



# **ORGANIZAÇÃO E TRATAMENTO DE DADOS**

**Maria Eugénia Graça Martins**

**João Pedro Ponte**

**Junho de 2010**



# Índice

Capítulo 1 – INTRODUÇÃO .....	3
Capítulo 2 – ORIENTAÇÕES CURRICULARES PARA O ENSINO DA ESTATÍSTICA .....	5
2.1 Introdução .....	7
2.2 Literacia estatística como objectivo curricular.....	7
2.3 Pensamento estatístico e raciocínio estatístico.....	9
2.4 A Estatística e a Matemática .....	10
2.5 Orientações metodológicas para o ensino da Estatística.....	11
2.5.1 Orientações gerais .....	11
2.5.2 Tarefas e recursos .....	13
2.5.3 Organização do trabalho dos alunos e avaliação .....	16
Capítulo 3 – A INVESTIGAÇÃO ESTATÍSTICA. DADOS E VARIÁVEIS.....	19
3.1 O que é uma investigação estatística?.....	21
3.2 Recenseamento e sondagem .....	22
3.3 População e amostra. Dados e variáveis.....	24
Tarefa – Classificação de variáveis .....	27
3.4 “Limpar” os dados.....	28
Tarefa – Vamos limpar estes dados .....	28
Tarefa – Não serão irmãos a mais?.....	30
Projecto – Os Censos vão às Escolas.....	31
Capítulo 4 – TABELAS E GRÁFICOS.....	41
4.1 Introdução .....	43
4.2 Diagramas de Venn e de Carroll .....	43
Tarefa – Número de letras do nome.....	44
Tarefa – Figuras geométricas.....	44
Tarefa – Números de 1 a 30.....	45
Tarefa – Múltiplos de 2 e 4, até 20 .....	45
Tarefa – Figuras e sólidos .....	46
Tarefa – Ajudas a lavar a loiça? .....	46
Tarefa – Preferes maçã, laranja ou banana? .....	47
4.3 Tabelas e gráficos para dados qualitativos .....	48
4.3.1 Esquemas de contagem gráfica (tally charts).....	48
Tarefa – O mês do aniversário .....	49
4.3.2 Tabela de frequências para dados qualitativos .....	49
Exemplo – Animal doméstico preferido das turmas A e B... 51	
4.3.3 Gráfico de pontos e gráfico de barras para dados qualitativos.....	52
4.3.3.1 Gráfico de pontos.....	52
4.3.3.2 Gráfico de barras .....	54
Exemplo – Seguro do agricultor .....	55
Tarefa – Prato preferido.....	56
4.3.4 Pictograma .....	57
Tarefa – Bolachas preferidas .....	57

Exemplo – Os passageiros de um navio.....	58
Exemplo – Campo de jogos .....	59
Exemplo – Seguro do agricultor (cont.).....	61
Tarefa – os animais do jardim.....	61
4.3.5 Gráfico circular .....	63
Exemplo – A piza preferida nas turmas A e B .....	63
Tarefa – Animal doméstico preferido.....	66
Tarefa –Meio de transporte utilizado .....	67
4.3.6 Nem sempre um gráfico com barras é um gráfico de barras.....	68
Exemplo – Número de queixas recebidas num hospital, por negligência médica .....	71
Exemplo – Quantidade de açúcar nos cereais para crianças	72
4.4 Tabelas e gráficos para dados quantitativos discretos .....	73
4.4.1 Tabela de frequências para dados quantitativos discretos .	73
4.4.2 Gráfico de pontos e gráfico de barras para dados quantitativos discretos.....	74
4.4.2.1 Gráfico de pontos.....	74
4.4.2.2 Gráfico de barras .....	75
Tarefa – Número de irmãos .....	76
Exemplo - Candidatos a algumas vagas .....	78
4.4.2.2 1 Gráfico de barras para comparar dois ou. mais conjuntos de dados .....	80
Tarefa – Alguns dados sobre o agregado familiar .....	80
Tarefa – Cinco montinhos de feijões .....	82
4.5 Tabelas e gráficos para dados quantitativos contínuos .....	83
Exemplo – Altura e peso dos alunos de uma escola do 1.º ciclo .....	84
4.5.1 Tabela de frequências para dados quantitativos contínuos	87
4.5.2 Histograma.....	88
Exemplo – Duração de chamadas telefónicas .....	90
4.5.3 Gráficos das frequências relativas acumuladas .....	91
4.6 Outras representações gráficas .....	93
4.6.1 Gráfico (ou diagrama) de caule-e-folhas .....	93
Tarefa – Quantos segundos se consegue estar sem respirar	94
Exemplo – O tempo de sono do Pedro e do David .....	98
Tarefa– Vamos comparar as idades dos nossos pais e das nossas mães .....	99
Tarefa – Qual a espécie de milho mais vantajosa? .....	101
4.6.2 Diagrama de extremos e quartis .....	103
4.7 Formas frequentes de distribuição de dados.....	105
Exemplo – Salários de trabalhadores .....	107
Exemplo – Qual o aspecto da distribuição? .....	108
4.8 Representações gráficas e tabelas de frequências para dados bivariados.....	110
4.8.1 Diagrama de dispersão .....	111

Tarefa – Vamos comer queijo, mas não exageremos.....	112
4.8.2 Gráfico de linhas .....	113
Tarefa – As vendas estão a correr bem?.....	114
Tarefa – Registo da temperatura máxima e mínima no Porto, Lisboa, Faro, Ponta Delgada e Funchal.....	116
Capítulo 5 – CARACTERÍSTICAS AMOSTRAIS .....	117
5.1 Introdução .....	119
5.2 Medidas de localização.....	119
5.2.1 Média .....	121
Exemplo da “pouca resistência” da média.....	123
Tarefa – Número de vogais e de consoantes do nome .....	126
Tarefa – Desvios entre os dados e a média.....	129
Tarefa – Quais as idades dos meus filhos? Qual a minha idade? Qual a idade da minha mulher?.....	131
Tarefa – Idade média dos finalistas do curso de Matemática de 1950 .....	133
5.2.2 Mediana .....	133
Tarefa – Notas no teste de Matemática das turmas 9.ºA e 9.ºB .....	138
Tarefa – Média ou mediana dos salários?.....	140
Tarefa – Número de cigarros fumados por dia .....	141
5.2.3 Moda .....	142
5.2.4 Quartis .....	145
Tarefa – Notas no teste de Matemática das turmas 9.ºA e 9.ºB (cont.) .....	146
5.2.5 Percentis .....	147
Exemplo – A obesidade é um problema.....	148
Exemplo – Conversa entre mãe e filho .....	148
5.3 Medidas de variabilidade.....	149
5.3.1 Amplitude.....	149
5.3.2 Amplitude interquartil .....	150
5.3.3 Outras medidas de variabilidade: O desvio médio absoluto e o desvio padrão .....	151
Tarefa – Desvios entre os dados e a média (cont.).....	151
Tarefa – Temperaturas de duas cidades .....	156
Tarefa – Vamos comparar os consumos dos carros .....	158
Tarefa – vamos comparar as notas no mesmo teste de Inglês de alunos de três escolas diferentes .....	158
Capítulo 6 – PROBABILIDADE .....	161
6.1 Introdução .....	163
6.2 Probabilidade empírica ou experimental de um acontecimento.....	164
Exemplo – Qual o animal doméstico preferido?.....	170
Exemplo – Qual a probabilidade .....	171
Tarefa – A escala de probabilidade .....	171
Tarefa – Serão os jogos justos (ou equilibrados)?.....	174

Tarefa – O que é mais provável? .....	175
Tarefa – Qual o meio de transporte utilizado pelos alunos para irem para a escola?.....	177
Tarefa – Como seleccionar uma de duas pessoas, de uma forma justa .....	178
Exemplo – Os sabores da regularidade a longo termo..	181
Tarefa – O jogo será justo? .....	181
Tarefa – Qual a composição do saco de berlindes?.....	182
Tarefa – Os sacos de berlindes .....	184
Tarefa – Qual o tipo de prato que os alunos preferem?....	185
Tarefa – Qual a probabilidade do próximo condutor utilizar a Via Verde? .....	187
6.3 Selecção de uma amostra de uma população com o objectivo de estimar uma probabilidade.....	188
6.3.1 Amostra aleatória simples sem reposição e com reposição.....	190
Exemplo – Como seleccionar alunos de uma turma.....	191
6.3.2 Amostra estratificada .....	192
Tarefa – O meio de transporte utilizado pelos alunos para irem para a escola .....	192
6.3.3 Amostragem sistemática.....	194
6.4 Experiência aleatória, Espaço de resultados, Acontecimentos .....	195
Exemplo – Espaços de resultados.....	197
Exemplo – Lançamento da moeda .....	197
Exemplo – Tempo de vida .....	198
Exemplo – Lançamento de dois dados.....	199
Exemplo – O saco de berlindes .....	203
6.4.1 Utilização de diagramas de Venn para representar acontecimentos .....	205
Exemplo – Família de 2 filhos .....	205
Exemplo – A caixa de disquetes .....	205
6.4.2 Utilização de diagramas em árvore para representar acontecimento.....	206
Tarefa – A caixa de bombons.....	208
Tarefa – Quantas toietes pode a Mariana vestir?.....	208
Tarefa – Quais os resultados que fazem com que o Pedro ganhe o jogo? .....	209
6.4.3 Operações com acontecimentos .....	210
Tarefa – Lançamento de duas moedas de 1 euro .....	213
Tarefa – Lançamento de três moedas de 1 euro .....	214
6.5 Modelo de Probabilidade para um fenómeno aleatório. Probabilidade de um acontecimento .....	215
6.5.1 Probabilidade de um acontecimento .....	215
6.5.2 Processos de construção de modelos de probabilidade ou Como atribuir probabilidades aos acontecimentos elementares.....	217

6.5.2.1 Situação de simetria – Regra de Laplace .....	217
6.5.2.2 Probabilidade experimental ou frequencista .....	220
Tarefa – Qual o jogo preferido .....	221
6.5.2.3 Probabilidade subjectiva .....	222
6.5.3 Exemplos de modelos de probabilidade .....	222
Exemplo – Lançamento da moeda de um euro .....	222
Exemplo – Lançamento do dado.....	223
Tarefa – Dados especiais .....	223
Tarefa – A caixa de pastilhas M&M .....	224
Tarefa – O Tipo sanguíneo .....	224
Exemplo - Se seleccionar, ao acaso, uma pessoa residente em Portugal, é mais provável que seja homem ou mulher? .....	226
Exemplo – A idade da população residente em Portugal...	226
Tarefa – Qual a cor preferida para pintar o pátio da escola?.....	227
Tarefa – A roleta de duas cores.....	228
Tarefa – O jogo com berlindes .....	228
Tarefa – As duas caixas de berlindes .....	229
Tarefa – Quantos lançamentos são necessários?.....	229
Tarefa – A caixa com drageias de chocolate .....	232
Tarefa – O jogo de andebol .....	233
Tarefa – Escolhe ao acaso uma letra do alfabeto .....	233
Tarefa – Escolhe ao acaso uma letra da palavra “palavra”	233
Tarefa – Será que o Pedro vai comer a sua fruta preferida?.....	234
Tarefa – A caixa de disquetes .....	235
Tarefa – Lançamento de um dado equilibrado .....	235
Tarefa – Uma escolha difícil entre 5 candidatos! .....	236
Tarefa – Um jogo desequilibrado!.....	238
Tarefa – Lançamento de dois dados. Será que o jogo é justo?.....	238
Tarefa – Ainda o lançamento de dois dados .....	241
Tarefa – O lançamento de duas moedas.....	242
Tarefa - Quem consegue dar primeiro a volta ao quadrado? .....	243
Tarefa – Quem é que ganha o jogo? .....	245
Tarefa – Moedas não equilibradas .....	249
Tarefa – Vamos estimar a área do círculo com raio 0,5 unidades .....	250
Tarefa – Vamos estimar o valor de $\pi$ .....	251
Tarefa – Estimar a área de figuras.....	251
Tarefa – Uma chuva de meteoritos!.....	252
Tarefa especial – Vamos construir alguns dos nossos materiais para fazer experiências .....	253
Capítulo 7 – SIMULAÇÃO.....	255

7.1	Introdução .....	257
7.2	Números aleatórios e números pseudo-aleatórios .....	257
7.3	Simular o lançamento de uma moeda .....	260
7.4	Cálculo da probabilidade de acontecimentos por simulação .....	263
	Tarefa – Calcular a probabilidade de sair 2 faces Euro, em 2 lançamentos de uma moeda de um euro, equilibrada ...	263
	Tarefa – Qual a probabilidade de numa família de quatro filhos, todos serem rapazes? .....	265
	Tarefa – Qual a probabilidade de numa família haver um “casalinho”, mas com três filhos no máximo! .....	268
	Tarefa – Estimar as probabilidades dos resultados da soma das pintas das faces viradas para cima, quando se lançam dois dados .....	270
	Tarefa – O jogo é justo? .....	272
	Tarefa – Vamos a uma aposta?.....	275
	Tarefa – Qual a probabilidade de no lançamento de uma moeda, se verificarem pelo menos 3 faces Euro seguidas?.....	276
	Tarefa – Quem é que recebe mais comida?.....	277
	Tarefa – Qual será a probabilidade de cada amigo ficar com o seu chapéu-de-chuva? .....	280
	Tarefa – Qual a probabilidade de passar no exame?.....	285
	Tarefa proposta – Pequena sondagem sobre o tipo sanguíneo.....	286
	Tarefas .....	289
	Bibliografia .....	325

# 1 Introdução

Neste capítulo fazemos uma breve introdução aos objectivos da presente brochura.

---



---

## 1. Introdução

Todos os dias nos deparamos com informação estatística sobre áreas tão diversas como a economia, a educação, o desporto, a medicina ou a política. A nossa vida é em larga medida governada por dados que, conscientemente ou não, utilizamos na tomada de decisões. Sendo a Estatística a ciência que trata dos dados, ela deve fazer parte da educação dos alunos desde os níveis de escolaridade mais elementares, para que estes possam vir a ser cidadãos informados, consumidores inteligentes e profissionais competentes. Estreitamente relacionada com a Estatística surge a Teoria da Probabilidade, teoria que serve de base à quantificação da incerteza – uma característica sempre presente na nossa vida de todos os dias.

A preocupação com a compreensão dos aspectos elementares da Estatística e da Probabilidade tem vindo a reflectir-se no currículo escolar dos mais diversos países. Em Portugal, isso acontece desde há várias décadas. O novo *Programa de Matemática do Ensino Básico* inclui o tema “Organização e tratamento de dados” nos três ciclos, numa perspectiva de valorização da literacia estatística e do processo de investigação estatística, aspectos em que vai bastante além do programa anterior. Reconhecendo o papel do tema no desenvolvimento social e pessoal do aluno, o programa refere que este deve adquirir, ao longo da escolaridade, conhecimento de conceitos e representações de modo a compreender e ser capaz de produzir informação estatística e de a utilizar para resolver problemas e tomar decisões informadas. O programa aponta, também, para o desenvolvimento da compreensão da noção de probabilidade, tanto no seu aspecto teórico, como experimental. Note-se que, para além dos objectivos gerais de aprendizagem da Organização e tratamento de dados, o trabalho neste tema visa igualmente as finalidades e os objectivos gerais de aprendizagem da disciplina de Matemática no seu todo, articulando-se com os outros temas do programa e com as capacidades transversais – Resolução de problemas, Raciocínio e Comunicação.

A presente brochura desenvolve as orientações metodológicas respeitantes ao tema Organização e tratamento de dados, e discute aspectos fundamentais dos conceitos trabalhados no ensino básico, sugerindo tarefas a propor aos alunos e indicando como podem ser concretizadas na aula. Num ou noutro ponto, vai-se além do programa, abordando assuntos de interesse para a formação do professor. Deste modo, o próximo capítulo analisa o que se entende por literacia estatística e pensamento estatístico, realçando a sua importância no exercício da cidadania, e apresenta as orientações metodológicas para o ensino da Estatística. O capítulo 3 indica como se desenvolve uma investigação estatística e discute os conceitos fundamentais de dados e variáveis. De seguida, o capítulo 4 trata da representação de dados em tabelas e gráficos, um processo fundamental na realização de qualquer estudo estatístico. No capítulo 5 surgem as medidas de localização e de dispersão, que permitem caracterizar de forma abreviada aspectos importantes de um certo conjunto de dados. No capítulo 6, apresentamos uma abordagem dos conceitos fundamentais relativos à probabilidade indicados no programa. Finalmente, o capítulo 7, mostra como os meios computacionais podem ser usados para simular o comportamento dos fenómenos aleatórios.

---



## 2 Orientações curriculares para o ensino da Estatística

A Estatística pode ser considerada a ciência dos “dados”, analisamos aqui o que se entende por literacia estatística, realçando o facto de que um cidadão que a possui é um cidadão melhor informado, podendo assim participar mais conscientemente na vida social. De seguida, apresentamos as orientações metodológicas gerais para o ensino da Estatística, dando especial atenção às tarefas, recursos e modo de trabalho dos alunos.

---



## 2.1 Introdução

Neste capítulo começamos por analisar o que se entende, hoje em dia, por literacia estatística, bem como por pensamento estatístico e raciocínio estatístico. De seguida, apresentamos as principais orientações metodológicas para o ensino da estatística, com especial atenção às tarefas e recursos, à organização do trabalho dos alunos e à avaliação.

## 2.2 Literacia estatística como objectivo curricular

O objectivo do ensino da Estatística, a nível elementar, é, antes de mais, promover a literacia estatística, ensinando os alunos a ler e interpretar dados. Tal como foi importante para os nossos avós aprenderem a ler e contar, hoje em dia, a educação para a cidadania inclui saber ler e interpretar os números e gráficos com que nos deparamos no dia-a-dia. Quantas mães (e pais...) não ficam perplexas ao ouvirem o pediatra do seu filho dizer que a cabeça da criança está no percentil 25%, a altura no 50%, etc.? E quantos não gostariam que todos estes números rondassem os 100%, sem se aperceberem que estariam a desejar o indesejável?

Na linha do que sugere Lyn Arthur Steen<sup>1</sup>, podemos dizer que a literacia estatística consiste num conjunto de conhecimentos, convicções, predisposições, hábitos mentais, capacidades de comunicação e habilidades que as pessoas precisam para lidar de maneira eficaz com situações envolvendo dados de natureza quantitativa e qualitativa que surgem na sua vida e na sua actividade profissional. O progressivo desenvolvimento da Estatística e a crescente necessidade de conhecimentos estatísticos para enfrentar situações do quotidiano conduziram a uma preocupação crescente com a literacia estatística, à semelhança do que aconteceu, inicialmente, com a literacia matemática, exigida por uma quantificação cada vez mais acentuada de muitos aspectos da sociedade.

A literacia estatística deve permitir a cada um de nós resolver com segurança muitos problemas que nos dizem directamente respeito ou que nos são frequentemente apresentados pelos meios de comunicação social e cuja resolução apela a conhecimentos e pensamento estatísticos. Interpretar tabelas e gráficos, entender disputas salariais, índices de preços, oscilações bolsistas, taxas de desemprego, taxas relativas à evolução de doenças, mecanismos e resultados eleitorais e de sondagens, comparar a qualidade e os custos de bens ou serviços são apenas alguns exemplos. Richard Sheaffer<sup>2</sup> considera mesmo que esta nova literacia constitui um dos objectivos principais da educação ao nível elementar. Na sua perspectiva, trata-se de um objectivo que ultrapassa a própria disciplina de Matemática, devendo ser igualmente assumido por outras disciplinas.

A produção intensiva de informação é um fenómeno que se tem vindo a intensificar desde as últimas décadas do século XX, provocado pelo desenvolvimento dos computadores e pela sua utilização crescente na sociedade. Os cálculos e gráficos automáticos tornaram-se fundamentais na prática da Estatística e na sua utilização em todas as esferas vida social. Esta transformação rápida é responsável pelo aparecimento do conceito de literacia estatística, correspondendo à necessidade do

---

<sup>1</sup> Ver Steen, 2001.

<sup>2</sup> Sheaffer, 2001

---

homem moderno se adaptar às novas condições de vida, compreendendo e usando com eficácia a informação que lhe chega diariamente.

Muitos dos utilizadores actuais de Estatística trabalham com diversos tipos de software especialmente concebidos para fazer análise estatística de dados. Um dos programas mais conhecidos é o SPSS (*Statistical Package for the Social Sciences*), largamente usado na investigação em Psicologia, Sociologia e Educação, que permite calcular as medidas de estatística descritiva e fazer todo o tipo de análises estatísticas inferenciais. Outra família de programas muito usados, principalmente na gestão de empresas, é o SAS (*Statistical Analysis System*), que permite estudar as questões estratégicas das organizações como o controlo de custos, o aumento de proveitos e a eficiência e transparência financeiras. Um dos desenvolvimentos mais recentes deste software, o JMP 7 apresenta gráficos para explorar, compreender e visualizar interactivamente os dados, sendo especialmente apropriado para quem quer visualizar as relações e identificar dados com informação potencialmente duvidosa ("outliers"). Em muitas empresas e instituições utiliza-se a folha de cálculo Excel para a realização do tratamento estatístico de dados. Embora a folha de cálculo não tenha sido originalmente concebida para isso, a verdade é que ela permite calcular medidas estatísticas e representar dados em gráficos de vários tipos, alguns visualmente bastante apelativos. Para a maior parte dos estudos estatísticos simples o Excel serve perfeitamente e, por isso, nesta brochura recorreremos a este programa para a realização de diversas tarefas.

Na verdade, a tecnologia pode ser muito útil para o ensino e aprendizagem dos conceitos estatísticos. Por exemplo, se pretendermos analisar a fraca *resistência* da média, isto é, o facto desta ser muito influenciada por alguns dados, mesmo em pequena quantidade, mas de grandeza muito diferente dos restantes, facilmente fazemos essa experiência numa folha de cálculo. Um processo de grande utilidade, igualmente ligado à tecnologia, é a simulação, que pode ser utilizada, nomeadamente, para a *visualização* dos conceitos estatísticos e para uma melhor compreensão da variabilidade, ideia chave em Estatística. Foi, aliás, o desenvolvimento destes meios tecnológicos que fez com que surgisse uma nova forma de ensinar Estatística, centrando a atenção nos conceitos e não na forma de os calcular.

A preocupação com o ensino da Estatística a nível básico e secundário conheceu um momento importante, nos anos 80, quando a American Statistical Association (ASA), em cooperação com o National Council of Teachers of Mathematics (NCTM), desenvolveram o projecto *Quantitative Literacy* tendo em vista introduzir nos programas mais análise de dados e Estatística elementar. Os proponentes do projecto escolheram o termo "literacia quantitativa" em parte porque antecipavam alguma ansiedade relativamente ao termo "Estatística". Segundo indica Steen (2001), "literacia quantitativa" (ou "numeracia", como também é referida), não deve ser entendida como Estatística nem tão pouco como Matemática:

Literacia quantitativa é mais uma maneira de estar, uma maneira de abordar os problemas que emprega e enaltece tanto a Estatística como a Matemática. Ao contrário da Estatística, que se centra fundamentalmente na incerteza, a numeracia é frequentemente acerca da lógica da certeza. Ao contrário da Matemática, que trata fundamentalmente com o reino platónico de estruturas abstractas, a numeracia está frequentemente ancorada em dados derivados e relacionados com o mundo empírico. (...) Um desafio típico de numeracia envolve dados reais e processos incertos, mas requer Matemática elementar. Em contraste, um problema típico de Matemática envolve números e proces-

---

... directos, mas requer sofisticados conceitos abstractos. O teste da numeracia, assim como para qualquer literacia, consiste em verificar se uma pessoa usa, naturalmente, práticas apropriadas em muitos contextos diferentes. (...) Para ser útil para o aluno, a numeracia deve ser aprendida e usada em múltiplos contextos – em História e Geografia, em Economia e Biologia, em Agricultura e Culinária. Numeracia não é um entre muitos assuntos, mas uma parte integrante de todos os assuntos (Steen, 2001-*The Quantitative Literacy Design Team*, p. 5).

Um aspecto fundamental na literacia estatística é a capacidade de compreender e usar o pensamento estatístico e o raciocínio estatístico, pelo que se impõe uma discussão destes conceitos.

### 2.3 Pensamento estatístico e raciocínio estatístico

Fala-se muitas vezes indistintamente de literacia estatística, pensamento estatístico e raciocínio estatístico. A este respeito, Beth Chance (2002), diz:

Poucos [autores] dão uma definição formal do que se entende por pensamento estatístico. Muitos utilizam indiferentemente os termos “pensamento”, “raciocínio” e “literacia” como um esforço para a compreensão dos conceitos estatísticos para que a sua utilização não seja meramente uma manipulação de fórmulas, como tem sido demasiadas vezes utilizada e ensinada. (p. 2)

Na verdade, o ensino da Estatística a nível pré-universitário começou por dar especial atenção ao raciocínio estatístico, isto é, aos modos de raciocinar e resolver problemas próprios da Estatística, enfatizando técnicas, representações e processos de inferência. Posteriormente, percebeu-se a necessidade de atender à natureza dos dados estatísticos, valorizando de modo determinante o papel do contexto, pelo que se alargou a atenção a processos de pensamento mais amplos e variados. Finalmente, as considerações sobre as finalidades do ensino da Estatística na sociedade actual trouxeram para primeiro plano as preocupações com a literacia estatística. Estes três conceitos estão estreitamente relacionados porque a literacia estatística apoia-se no pensamento estatístico e este, por sua vez, tem como núcleo fundamental o raciocínio estatístico.

Uma investigação estatística realiza-se em diversas etapas e envolve aspectos específicos de raciocínio ou pensamento em cada uma delas. A primeira etapa consiste na formulação de questões para investigar. Neste ponto, é preciso considerar se as questões são ou não apropriadas e têm ou não uma natureza estatística, isto é, envolvem ou não variabilidade nos dados. A segunda etapa evolui a recolha dos dados. É preciso então definir um plano apropriado e seleccionar técnicas de recolha de dados. A terceira etapa refere-se à análise de dados, começando pela escolha da representação mais adequada tendo em conta a natureza dos dados e os fins em vista. Calculam-se medidas estatísticas, de localização, dispersão e associação, por exemplo. Procura-se, deste modo, descrever a tendência central e a variabilidade dos dados, o que pode envolver a procura de regularidades, bem como a identificação de diferenças entre os dados e entre os dados e distribuições de probabilidade conhecidas. Finalmente, a quarta etapa diz respeito à interpretação dos resultados, tendo em conta a questão proposta. Neste ponto formulam-se conclusões referentes aos dados, possíveis generalizações para além dos dados e também possíveis questões que podem servir de base a novas investigações.

---

Como Shaughnessy e Pfannkuck (2002), podemos considerar que a forma estatística de pensar envolve aspectos como:

- Reconhecimento da necessidade de *dados*, de modo a poder fazer julgamentos sobre situações reais;
- Realização de certas *transformações numéricas* para facilitar a compreensão (representação em tabelas e gráficos, cálculo de medidas de localização e dispersão);
- Procura de causas e explicações e previsão de acontecimentos a partir da exploração da *variabilidade*, usando *modelos estatísticos*;
- Consideração do *contexto* como essencial não só para observar mas também interpretar as mensagens existentes nos dados.

Em resumo, podemos dizer que o raciocínio estatístico envolve um processo explícito onde se identificam factos, estabelecem relações e fazem inferências. O pensamento estatístico, pelo seu lado, tem um lado intuitivo, informal e implícito que suporta o nosso raciocínio. Finalmente, a literacia estatística é a capacidade que nos permite interpretar a informação, avaliar a sua credibilidade, e produzir nova informação, quando necessário.

## 2.4 A Estatística e a Matemática

Devemos ter em atenção que a Estatística e a Matemática são ciências diferentes. Moore et al (1997) discutem esta diferença nos seguintes termos:

A Estatística é uma disciplina metodológica. Ela existe não por si própria, mas antes com o objectivo de oferecer a outros campos de estudo um conjunto de ideias coerentes e de instrumentos para tratar os dados. A necessidade de uma tal disciplina acontece devido à omnipresença da variabilidade. Os indivíduos variam. Medidas repetidas do mesmo indivíduo variam (...) A Estatística fornece-nos os meios para tratar com dados que têm em linha de conta a presença da variabilidade. O foco na variabilidade dá naturalmente à Estatística um conteúdo que a torna diferente da própria Matemática e de outras Ciências Matemáticas, mas não é só o conteúdo que distingue o pensamento estatístico do matemático. A Estatística requer um tipo de pensamento diferente, *já que os dados são mais do que números, são números com um contexto* (...) Na análise de dados o contexto fornece o significado. (p. 801)

Deste modo, enquanto que o pensamento matemático se refere a relações entre conceitos abstractos, o pensamento estatístico tem sempre presente o contexto que dá origem aos dados, que, por sua vez, permitem (ou não) responder a certas questões. Para ilustrar esta ideia, consideremos a situação em que temos um conjunto de notas de alunos a uma determinada disciplina e calculamos a respectiva média e mediana. Sob o ponto de vista matemático aplicámos duas fórmulas que conduzem a dois números, a média e a mediana. Sob o ponto de vista estatístico, temos muito mais do que isso. A comparação daqueles dois números, um relativamente ao outro, permite-nos visualizar certos aspectos da estrutura dos dados, isto é, o comportamento global da turma, naquela disciplina.

Também David Vere-Jones (1995) sublinha que o raciocínio estatístico é diferente do matemático e que a educação estatística não se pode restringir a uma visão da Estatística como um simples ramo da Matemática. O raciocínio matemático é um raciocínio eminentemente lógico, em que uma proposição ou é verdadeira ou é falsa. Em contrapartida, no raciocínio estatístico, tratamos com afirmações em que

---

não podemos dizer que são verdadeiras nem tão pouco falsas. As situações estatísticas envolvem incerteza, que procuramos quantificar através da probabilidade:

**Verdadeiro → ? Incerteza ? ← Falso**

A incerteza acompanha-nos no dia-a-dia, em todas as situações.

A Estatística é muito diferente dos vários ramos da Matemática estudados no ensino básico e secundário, como a Geometria, a Álgebra e a Análise. Isso resulta do seu envolvimento directo com o estudo de outras áreas como as Ciências da Saúde, a Economia, as Ciências Políticas e outras ciências sociais. É importante ensinar um médico, um técnico da indústria farmacêutica, um sociólogo, um psicólogo e todo aquele que faz uso da Estatística a utilizá-la correctamente. A utilização incorrecta desta ciência pode levar a decisões erradas com consequências negativas quer para o desenvolvimento das outras ciências quer para a vida do cidadão comum. Como refere Chris Chatfield (1991), os não especialistas cometem erros estatísticos com grande frequência. Para o evitar, é preciso desenvolver a literacia estatística.

Em Estatística, preocupamo-nos sobretudo com as medidas que se podem utilizar para reduzir a informação contida num conjunto de dados, com as representações gráficas mais convenientes para representar esses dados e com o modo de utilizar a tecnologia para calcular essas medidas e fazer essas representações gráficas, de modo a responder a questões concretas, interpretando de modo adequado toda a informação disponível. Enquanto que na Matemática sobressaem os aspectos lógicos, na Estatística sobressai a pertinência das interpretações. É importante sublinhar esta diferença entre Estatística e Matemática, uma vez que a Estatística, nos níveis básico e secundário, está integrada na disciplina de Matemática e é ensinada por professores de Matemática, que precisam de ter a clara percepção de que se trata de assuntos diferentes – que devem ser encarados de modo diferente e, muitas vezes, ensinados de modo diferente.

## **2.5. Orientações metodológicas para o ensino da Estatística**

### **2.5.1 Orientações gerais**

A investigação sobre o ensino e aprendizagem da Estatística e os resultados de estudos internacionais (como o TIMSS) têm evidenciado que os alunos revelam, com frequência, dificuldades e ideias incorrectas tanto no campo conceptual como em aspectos computacionais. Algumas destas dificuldades relacionam-se com a natureza da Estatística, enquanto que outras derivam das estratégias de ensino utilizadas e do tipo de experiências de aprendizagem proporcionadas aos alunos.

A questão central que se coloca no ensino da Estatística tem a ver, antes de mais, com os seus objectivos. Até aqui, a prática de ensino tem valorizado sobretudo a aprendizagem da representação de dados em tabelas e gráficos e do cálculo de medidas estatísticas como médias e medianas. Ora, como já referimos no ponto anterior, o objectivo central é o desenvolvimento da literacia estatística, que inclui a capacidade de ler e interpretar dados organizados na forma de tabelas e gráficos e de os usar para responder às questões mais variadas. Num outro nível, o ensino da Estatística visa desenvolver nos alunos a capacidade de planear e executar uma investigação estatística, bem como a capacidade de interpretar e avaliar criticamente os resultados de um estudo estatístico já realizado.

---

---

O trabalho do professor tem de contemplar todos estes objectivos. Por um lado, ele tem de promover a capacidade dos alunos compreenderem e usarem conceitos e representações estatísticas na resolução de questões diversas – incluindo conceitos como dados, variável, população, amostra, distribuição e medidas estatísticas e representações como tabelas e gráficos. Por outro lado, o professor tem de promover nos alunos a compreensão do que é uma investigação estatística, nas suas etapas fundamentais de formulação de questões, recolha, análise e interpretação dos dados e habilitá-los para realizar estudos deste tipo. O *Programa de Matemática* assume que estes dois objectivos desenvolvem-se em paralelo e reforçam-se mutuamente. A aprendizagem dos conceitos e representações específicas é essencial para a realização de estudos estatísticos cada vez mais complexos e a compreensão do que é uma investigação estatística dá sentido aos diversos conceitos e representações.

Deste modo, o programa apresenta como propósito principal de ensino, no 1.º ciclo, desenvolver nos alunos a capacidade de ler e interpretar dados organizados na forma de tabelas e gráficos assim como de os recolher, organizar e representar, com o fim de resolver problemas em contextos variados relacionados com o seu quotidiano. Nos ciclos seguintes este propósito inclui também a produção da informação estatística e a capacidade de tomar decisões informadas e apresentar argumentos a apoiá-las. Nos objectivos gerais de aprendizagem dos 2.º e 3.º ciclos surge ainda o planeamento e a realização de estudos envolvendo procedimentos estatísticos, com interpretação dos resultados obtidos, incluindo a avaliação intuitiva da credibilidade de argumentos por parte dos alunos.

A ênfase do trabalho na Estatística é colocada na análise exploratória de dados e no envolvimento progressivo dos alunos em experiências de natureza investigativa, desde os primeiros anos de escolaridade. Ao longo dos três ciclos, dá-se realce a processos e capacidades que promovem a literacia estatística dos alunos. Deste modo, não se pretende que os alunos, quando acabam o ensino básico, sejam capazes de realizar estudos estatísticos sofisticados, mas sim que compreendam e saibam utilizar a linguagem básica e as ideias fundamentais da Estatística, desde a formulação de questões a investigar à interpretação dos resultados.

O GAISE College Report (2005), enuncia seis recomendações que reflectem esta preocupação com o desenvolvimento da literacia estatística:

- Salientar a literacia estatística e desenvolver o pensamento estatístico;
- Utilizar dados reais;
- Acentuar a compreensão dos conceitos, em vez de apenas teoria e procedimentos;
- Fomentar uma aprendizagem activa na sala de aula;
- Utilizar tecnologia para desenvolver a compreensão dos conceitos e a análise dos dados;
- Utilizar a avaliação para conhecer e melhorar a aprendizagem dos alunos.

São inúmeros os documentos para o ensino da Estatística que realçam o facto de esta ser a ciência dos “dados”. Nela procura-se responder a questões de natureza muito diversa, seja de outras ciências, seja dos mais diversos campos da actividade social, através de uma análise conveniente de dados e da sua interpretação. Ao compreenderem que os dados são mais do que números, e ao reconhecerem a

---

---

necessidade de dados para tomarem decisões fundamentadas, os alunos desenvolvem o seu pensamento estatístico.

A diferença entre a Estatística e a Matemática, já realçada no ponto anterior, arrasta consigo a preocupação da parte dos educadores estatísticos sobre a ênfase que se deve colocar nos conceitos e no trabalho com dados, em detrimento das fórmulas e cálculos. A célebre frase do poeta António Machado "*Caminante, no hay camino, se hace camino al andar*", traduz, de forma exemplar, a metodologia que deve ser seguida no ensino da Estatística, desde o seu início. Devem ser os alunos, sob a orientação do professor, a planear a recolha dos dados necessários, para dar resposta às suas questões, nomeadamente sob a forma de pequenos projectos de investigação. Estes dados serão depois tratados e esse tratamento dará algumas respostas e provocará, eventualmente, novas questões.

O que os alunos aprendem está relacionado com o modo como aprendem e, por consequência, com as experiências que lhes são proporcionadas pelos professores. Assim, há dimensões no ensino que são essenciais para uma aprendizagem em profundidade e extensão, nomeadamente o tipo de tarefa e os recursos mobilizados para a sua realização.

### **2.5.2 Tarefas e recursos**

No ensino da Estatística há lugar para os mais diversos tipos de tarefa, incluindo investigações, projectos, jogos, problemas e exercícios. O *Programa de Matemática* sublinha que os conceitos deste tema devam ser trabalhados em todos os ciclos, a partir de problemas variados e também de investigações e projectos. Há também necessidade de propor exercícios para aprofundar e relacionar o conhecimento e a compreensão de conceitos e processos e criar rotinas de utilização adequada de procedimentos. Os jogos tanto podem servir para a introdução de novos conceitos como para a consolidação de aprendizagens já realizadas.

Uma investigação estatística é uma tarefa em que se começa por definir uma área de interesse – suscitada por uma curiosidade ou por uma necessidade muito concreta – e que se desenvolve ao longo das quatro etapas já referidas no ponto 2.1.2: (i) formulação de questões e concepção de um plano de investigação, (ii) selecção das técnicas de recolha e recolha dos dados, (iii) representação e análise dos dados, e (iv) interpretação dos dados e formulação de conclusões. Quando a investigação estatística se reveste de complexidade significativa e se prolonga no tempo, refere-se muitas vezes que se trata de um projecto de investigação estatística ou simplesmente de um projecto. Deste modo "investigação" e "projecto" estatístico acabam por ser dois conceitos que se sobrepõem. Também, por vezes se fala em "exploração" estatística, em especial quando já temos um conjunto de dados reunido e procuramos descobrir ao mesmo tempo que regularidades encerram e que questões podemos formular a seu respeito. Podemos dizer que um "problema" estatístico é uma questão bem definida que se coloca numa das etapas de uma investigação estatística, para a qual não há uma resposta imediata, e que um "exercício" é também uma questão bem definida que se coloca num dado momento do processo estatístico para a qual o aluno já conhece um processo de resolução.

Nos primeiros anos de escolaridade, o programa advoga que a aprendizagem da Estatística tem por base actividades relativas a situações diversificadas e dia-a-dia dos alunos, ricas em informação. Os alunos registam os dados e representam-nos

---

---

em tabelas e gráficos simples. A classificação e contagem de objectos são então tarefas importantes, usando diagramas de Venn e de Carroll, que constituem formas simples de representação de dados. Assim, desde o 1.º ciclo, os alunos devem envolver-se em experiências de recolha e organização de dados qualitativos e quantitativos discretos, representando-os em tabelas de frequências absolutas e em gráficos de vários tipos, como pictogramas e gráficos de barras, identificando a sua moda. Progressivamente, o professor deve dar atenção à linguagem utilizada que, sendo familiar aos alunos, deve promover o sentido de rigor, de acordo com o seu nível de desenvolvimento. Deve procurar que os alunos clarifiquem o significado de expressões e o uso de termos menos precisos colocando questões do tipo *Podes explicar de uma outra maneira? Dás-me um exemplo para perceber o que significa? Por que dizes isso?*

As propostas de trabalho para os alunos do 2.º ciclo estendem o que já foi anteriormente feito em termos de análise, aprofundando a interpretação dos dados. A formulação de questões e problemas com significado para os alunos – quando possível em colaboração com outras disciplinas – e, de seguida, a recolha, análise e interpretação de dados, devem estar sempre presentes. Na realização de projectos e investigações envolvendo dados de natureza variada, incluindo dados quantitativos discretos e contínuos, os alunos representam-nos em tabelas de frequências absolutas e relativas, gráficos de barras, circulares e de linha e diagramas de caule-e-folhas. Para descrever, analisar e interpretar a informação que esses dados contêm recorrem à moda, média aritmética, extremos e amplitude. Quando os dados não permitem responder às perguntas formuladas, o professor deve discutir as opções tomadas e incentivar os alunos a fazer nova recolha dados. Compete-lhe providenciar que os alunos trabalhem com dados relevantes e façam eles próprios a sua recolha. Assim, não só a sua familiaridade com a informação aumenta, como pode ser maior a sua motivação para construírem a sua “história”, representando-a de maneiras diferentes, extraíndo conclusões e fazendo previsões. Para além disso, trabalhando com os seus próprios dados, os alunos aprendem muito mais facilmente o sentido de vários conceitos estatísticos como média, mediana e moda. Os problemas ambientais e sociais, em especial os que se manifestam na sua região, constituem, frequentemente, um terreno fértil para questões a investigar, possibilitando aos alunos abordar problemas que os afectam, propor soluções e sentir a utilidade do seu trabalho.

No 3.º ciclo, os alunos podem realizar investigações estatísticas que incluem a comparação de dois ou mais conjuntos de dados, com a identificação das suas semelhanças e diferenças. Neste ciclo alarga-se o reportório das medidas estatísticas – que passam a incluir a mediana, os quartis e a amplitude interquartil – e das formas de representação de dados – com os diagramas de extremos e quartis e os histogramas. É neste ciclo que os alunos estudam as noções de população e amostra, ponderando elementos que afectam a sua representatividade e realizando e discutindo predições baseadas em estudos com amostras. Assim, o professor pode propor a realização de investigações estatísticas baseadas em situações reais, onde se utilizem os conceitos estatísticos estudados. Para desenvolverem essas investigações, os alunos formulam questões, planeiam o estudo estatístico, seleccionam amostras adequadas, recolhem dados dessas amostras, representam-nos e interpretam-nos. A partir das propriedades verificadas nos dados recolhidos, espera-se que os alunos façam conjecturas e discutam a validade das conclusões para a população de onde a amostra foi seleccionada. Deve discutir-se que informação é

---

---

necessária para tomar uma decisão válida ou tirar uma conclusão fundamentada. Os alunos, ao terminar o ensino básico, devem ter consciência de que raramente se obtém toda a informação pretendida através de uma amostra, mesmo em condições óptimas de recolha de dados, permanecendo sempre alguma incerteza nas conclusões formuladas sobre a população. É importante que tenham oportunidade de analisar em que casos é adequado recorrer ao estudo de toda a população ou apenas de uma amostra desenvolvendo também o seu sentido crítico relativamente ao uso de amostras mal seleccionadas.

Deste modo, para além das capacidades transversais de Resolução de Problemas, Raciocínio e Comunicação, também as *representações* assumem um papel fundamental em Estatística. Lidar com dados estatísticos envolve necessariamente escolher uma forma de os representar (em tabelas, diagramas e gráficos). As tabelas e as representações gráficas a usar dependem da natureza dos dados recolhidos e dos aspectos que pretendemos analisar. Assim, ao longo do ensino básico, é importante que os alunos tenham oportunidade de comparar diversos tipos de representação para a mesma situação e verificar quais são os mais apropriados. Devem ainda desenvolver uma atitude crítica relativamente à utilização de gráficos enganadores.

Um outro aspecto que assume grande importância em Estatística é o das *conexões*. Uma das características principais das tarefas deste tema é o grande interesse que pode ter a utilização de dados reais, que muitas vezes proporcionam um envolvimento entusiástico dos alunos na aprendizagem. Estes dados são recolhidos para responder a questões e não para treinar fórmulas ou realizar representações gráficas. Frequentemente, originam a formulação de novas questões e a recolha de novos dados. Mas também há situações em que é útil o uso de dados hipotéticos, nomeadamente quando pretendemos explorar propriedades de alguns conceitos estatísticos.

Podemos falar de conexões entre a Estatística e diversos campos da Matemática, como a Geometria (gráficos circulares), os Números e operações (cálculo de médias e quartis), e a Álgebra (tabelas, gráficos de linha). E, sendo os dados estatísticos sempre referentes a algum contexto, existe margem para o estabelecimento de conexões com áreas de actividade e campos do conhecimento dos mais diversos. Assim, por exemplo, os frequentes estudos de opinião levados a cabo pelos diversos canais de televisão ou as sondagens políticas podem ser analisados. As investigações em ciência, sobre a alimentação, mudanças climáticas, ou variação da natalidade fornecem também oportunidade para os alunos compreenderem como a natureza da amostra está relacionada com a informação a ser recolhida e os objetivos do estudo, para serem discutidos os cuidados a ter na selecção da amostra e as generalizações que se podem fazer.

Um aspecto a considerar pelo professor quando planeia trabalho a realizar são os recursos necessários. Na aprendizagem da Estatística os recursos fundamentais são: (i) os dados – tanto quanto possível reais e recolhidos pelos próprios alunos – e os instrumentos que permitem a sua recolha; e (ii) as calculadoras e computadores que realizam cálculos e todo o tipo de representações de forma, rápida, exacta e eficaz. Já referimos a importância de usar dados reais de situações do quotidiano, com destaque natural para dados existentes em jornais e revistas e na Internet. Por outro lado, no ensino da Estatística a tecnologia tem um papel fundamental. A tecnologia serve não só para a realização de cálculos fastidiosos e a sua represen-

---

tação gráfica, mas também para visualizar os conceitos estatísticos. Por exemplo, a comparação do comportamento da média e da mediana, quando se modificam alguns dos dados de um certo conjunto, é impraticável sem a utilização de calculadoras ou do computador. O programa de Matemática não advoga, no ensino básico, a utilização de vários tipos de *software* de Estatística. Trata-se de produtos dispendiosos<sup>3</sup> que não se justificam para a concretização dos nossos objectivos, para os quais, como já dissemos, é suficiente o uso da calculadora e de uma folha de cálculo como o Excel. Também se recomenda a utilização de *applets* que permitam visualizar a representação e comportamento de alguns conceitos estatísticos.

### 2.5.3 Organização do trabalho dos alunos e avaliação

Seja qual for o tipo de tarefa, é fundamental que o professor dê indicações claras sobre o que espera do trabalho dos alunos e os apoie na sua realização. Em Estatística, particularmente na realização de investigações e projectos, o trabalho dos alunos em grupo constitui uma forma natural de organização. Estas tarefas têm de ser realizadas em diversas etapas, permitindo dividir responsabilidades entre os alunos. Em cada etapa, é necessário verificar se os objectivos pretendidos foram atingidos ou se é necessário fazer alguma correcção. O trabalho realizado em grupo é usualmente muito mais criativo, completo e estimulante do que o realizado individualmente. No entanto, para que este trabalho resulte, é necessária a definição de objectivos claros, a estruturação e calendarização das acções a realizar e verificação do seu cumprimento. Espera-se dos alunos a tomada de iniciativa e assunção de responsabilidades.

O trabalho dos diversos grupos tem de ser apresentado na turma e discutido por todos. Esta forma de organizar o trabalho permite desenvolver uma dinâmica em aula em que todos os alunos têm oportunidade de apresentar o seu trabalho, de o ver questionado pelos outros alunos e também de questionar o trabalho dos seus colegas. Este momento de discussão, para além de contribuir para desenvolver a capacidade de comunicação dos alunos, permite-lhes muitas vezes aprofundar a compreensão dos conceitos, negociar significados e reformular raciocínios incorrectos. A discussão em grande grupo é o momento privilegiado para a partilha e debate de ideias, a sistematização dos conceitos e a institucionalização de conhecimentos. O professor tem que garantir nestes momentos condições para uma efectiva participação e aprendizagem da generalidade dos alunos, investindo na gestão do espaço e do tempo e na qualidade das intervenções. Por vezes, os trabalhos realizados pelos diversos grupos podem ser expostos fora da sala de aula, noutras locais da escola ou mesmo fora desta, apresentado o trabalho dos alunos a uma comunidade mais alargada.

Na realização de outros tipos de tarefa na sala de aula, como problemas e exercícios, podem ser mais vantajosas outras formas de organização, por exemplo com os alunos a trabalharem individualmente ou em pares. As formas de organização do

---

<sup>3</sup> Um dos problemas mais sérios suscitados pela utilização destes tipos de software de Estatística, é a facilidade com que qualquer pessoa os utiliza, sem saber muito bem o que está a fazer. Por exemplo, se tiver um conjunto de dados qualitativos, codificados com números, basta carregar num botão para obter a média, que neste caso não tem qualquer sentido! Analogamente, pode chegar à conclusão que existe uma forte correlação entre duas variáveis, porque carregou num botão para o cálculo da correlação, mas esqueceu-se que anteriormente deveria ter feito uma representação gráfica dos dados, para se aperceber da existência de uma associação linear entre os dados! Estes problemas sugerem a necessidade de desenvolver também uma literacia no uso deste tipo de programas, o que terá de ser equacionado quando eles eventualmente se tornarem mais acessíveis.

---

trabalho dos alunos devem variar, sucedendo-se as oportunidades para trabalho autónomo, interacção no seio de pequenos grupos e discussões colectivas com a moderação do professor.

Finalmente, uma palavra para a avaliação. A realização de todo o projecto ou investigação deve ter, naturalmente, uma, avaliação própria, feita em função da apresentação oral e escrita e, eventualmente de informação recolhida durante o desenrolar do processo. Para além disso, as aprendizagens em Estatística podem ser avaliadas como quaisquer outras aprendizagens, recorrendo a uma variedade de fontes de informação escrita e oral (testes, registos escritos com resposta a questões colocadas na aula, questões orais, etc.). Mais do que os aspectos computacionais, a avaliação deve insistir sobre o conhecimento e a compreensão dos conceitos estatísticos. Para que os próprios alunos tenham ideia da forma como estão a progredir na compreensão destes conceitos é muito importante o *feedback* do professor relativamente ao seu trabalho. Isso deve ocorrer tanto a propósito da realização de pequenas investigações e dos respectivos relatórios como das restantes tarefas realizadas no dia-a-dia.

---



## 3 A investigação estatística. Dados e variáveis

Descrevemos as diversas componentes de uma investigação estatística, realçando a omnipresença da variabilidade, e discutimos diversos conceitos fundamentais, com destaque para as variáveis e os dados estatísticos.

---



### 3.1 Introdução

É usual dizer que, sob o ponto de vista estatístico, o indivíduo não tem interesse e só passa a ser interessante quando faz parte de um todo! Por exemplo, a questão “Qual é o peso da Maria” não tem interesse sob o ponto de vista estatístico. No entanto, o peso da Maria e dos colegas da turma da Maria já tem interesse estatístico, pois admitimos a existência de variabilidade nesses pesos e poderá ser uma questão interessante saber qual o comportamento da variável Peso, relativamente aos alunos da turma da Maria, nomeadamente para fazer um estudo sobre obesidade na turma. Em contrapartida, não faria sentido realizar uma investigação estatística para averiguar quantos alunos da turma da Maria têm nome, pois sabemos que todos os alunos têm nome, pelo que na resposta a esta questão não esperamos a existência de variabilidade.

Indicámos no capítulo 2 que é a ideia-chave de *variabilidade* que está na base de qualquer estudo estatístico. Como é referido em GAISE (2005, p. 11), “*A formulação de uma questão estatística requer uma compreensão da diferença entre uma questão que antecipa uma resposta determinista e uma resposta baseada em dados que variam*”.

Como já referimos no capítulo 2, uma investigação estatística envolve, de um modo geral, quatro fases:

- 1.<sup>a</sup> fase – Formulação do problema a investigar, na forma de questões que se procuram responder através de dados;
- 2.<sup>a</sup> fase – Planeamento adequado para recolher dados apropriados;
- 3.<sup>a</sup> fase – Organização e tratamento dos dados recolhidos, através de tabelas, gráficos e algumas medidas;
- 4.<sup>a</sup> fase – Interpretação dos resultados obtidos e formulação de conclusões

A profundidade com que estas fases são tratadas depende do nível de ensino dos alunos. Assim, nos dois primeiros anos do 1.º ciclo, as questões a tratar podem prender-se com questões relacionadas com a turma, precisando de ser muito orientadas pelo professor. A classificação e contagem de objectos são tarefas indicadas para início do trabalho em Estatística, começando os alunos por classificar os dados utilizando diagramas de Venn e de Carroll. Antes de solicitar aos alunos a realização de representações gráficas, deve pedir-se-lhes para interpretar gráficos já construídos.

À medida que o ano de escolaridade dos alunos aumenta, eles vão progressivamente colocando as suas questões, que já não se restringem ao ambiente da turma. A pouco e pouco os alunos utilizam instrumentos apropriados para medir a variabilidade existente entre os indivíduos de um grupo e a comparar grupos. Assim, no 3.º ciclo, começam a compreender que os resultados que obtêm, ao estudar determinados conjuntos de dados, são susceptíveis de generalização, em determinadas condições. É a altura de distinguir entre **população** e **amostra** e de ponderar os elementos que podem afectar a representatividade de uma amostra em relação à respectiva população.

Ao longo de todos os ciclos, os alunos começam, a pouco e pouco, a tomar contacto com a variabilidade devida ao acaso e a utilizar linguagem associada a este concei-

to, desenvolvendo a noção de probabilidade. No 3.º ciclo, aprendem alguns processos de, em determinadas situações, quantificar esta probabilidade.

### 3.2 Recenseamento e sondagem<sup>4</sup>

É natural que o termo “sondagem” já faça parte do vocabulário do aluno, pois estamos constantemente a ouvir os meios de comunicação social anunciarem os resultados de sondagens. Já o mesmo pode não acontecer com as palavras “recenseamento” ou “censo”, mas estes termos são certamente conhecidos dos seus pais, que foram recenseados em 2001, se nessa data viviam em Portugal.

O termo recenseamento está, regra geral, associado à contagem oficial e periódica dos indivíduos de um país ou região. Para a maioria das pessoas a palavra *recenseamento* ou *censo* limita-se a esse significado. Ela abrange, no entanto, um leque mais vasto de situações. Assim pode definir-se recenseamento do seguinte modo:

**Recenseamento** – Estudo científico de um universo de pessoas, instituições ou objectos físicos com o propósito de adquirir conhecimentos, observando todos os seus elementos, e fazer juízos quantitativos acerca de características importantes desse universo.

Assim, a realização de recenseamentos pode-se estender a outras situações, tais como, às habitações (recenseamento da habitação), às indústrias (recenseamento industrial), à Agricultura (recenseamento agrícola), etc. É importante que fique claro que a palavra recenseamento está associada à análise de todos os elementos da população em causa e que tem por objectivo não só a enumeração dos seus elementos, como também o estudo de características importantes.

O recenseamento geral de uma população é uma prática que remonta à antiguidade (Roma e Egipto), onde já há conhecimento de recenseamentos da população, feitos a intervalos regulares, com o objectivo principal de obter informação para a colecta de impostos, chamada para o serviço militar e outros assuntos governamentais. Apesar disso, a sua prática corrente, com carácter periódico, só teve lugar, na maioria dos países, a partir do sec XIX. Esses censos periódicos são feitos em geral de 10 em 10 anos e, em princípio, todos os países são encorajados a cumprir certas normas internacionais ao elaborar um recenseamento. Em Portugal, a primeira operação que se conhece deste género foi levada a cabo por D. João III em 1527 e ficou conhecida pelo “numerando dos vizinhos”, tendo permitido estabelecer uma estimativa da população portuguesa. Este apuramento estatístico constitui um motivo de orgulho para os portugueses visto que foi um dos primeiros estudos deste género conhecido na Europa.

O INE, Instituto Nacional de Estatística, tem a seu cargo fazer recenseamentos da população portuguesa, o último dos quais, o XIV Recenseamento Geral da População, foi realizado em 2001. Nessa altura, o INE desenvolveu um projecto junto das escolas para motivar os alunos e fazer com que estes fossem para casa e falassem com os pais da importância de um censo. No fim deste capítulo, apresentamos a forma como o projecto foi desenvolvido.

---

<sup>4</sup> Esta secção segue de perto a brochura de Estatística, 10.º ano (Graça Martins et al. (1999)).

---

Neste recenseamento de 2001 ficaram a conhecer-se variadas características do povo português como a situação civil, a habitacional, a população emigrante, etc. Os dados relativos aos censos são extremamente importantes pois têm influência directa na decisão em assuntos de interesse nacional e local – na educação, emprego, saúde, transportes, recursos naturais, etc., etc. Comparando resultados de recenseamentos sucessivos pode extrapolar-se e predizer padrões futuros da população. Podemos obter informação sobre, por exemplo, a estrutura da idade da população e o crescimento populacional, fundamental para o planeamento na construção de novas escolas, alojamento para idosos, etc.

A realização de um recenseamento geral da população, além de implicar gastos muito elevados, é extremamente difícil de conduzir. Há problemas associados com a recolha adequada da informação, seu armazenamento, tratamento, posterior divulgação, etc. Assim, principalmente quando o número dos elementos da população é muito elevado, não é viável inquirir todos os elementos da população sempre que se quer estudar uma ou mais características particulares dessa população. Muitas vezes isso nem é desejável, pela perturbação que provoca nos inquiridos. Surge então o conceito de *sondagem*, que se pode definir como *o estudo científico de uma parte de uma população com o objectivo de analisar atitudes, hábitos e preferências da população relativamente a acontecimentos, circunstâncias e assuntos de interesse comum*. A realização de sondagens é uma actividade da segunda metade do séc. XX. Embora antes de 1930 já se tenham realizado sondagens, estas eram feitas de um modo muito pouco científico. Foi necessário um desenvolvimento adequado de métodos e técnicas estatísticas para que os resultados das sondagens pudessem ser analisados cientificamente.

Só em 1973 é que, pela primeira vez, apareceu publicado nos órgãos de comunicação social o resultado de uma sondagem realizada em Portugal, nomeadamente, "63% dos Portugueses nunca votaram" (Paula Vicente et al., 1996). Embora as sondagens se tenham popularizado devido a questões políticas, elas não são apenas um importante instrumento político; acima de tudo constituem um instrumento de importância vital em estudos de natureza económica e social. Assim, se nos meios políticos as sondagens são usadas para obter informação acerca das atitudes dos eleitores, de modo a planear campanhas, etc., elas são importantes também em estudos de mercado, para testar as preferências dos consumidores, descobrir o que mais os atrai nos produtos existentes ou a comercializar, tendo como objectivo o de satisfazer os clientes e aumentar as vendas. Também na área das ciências sociais as sondagens são importantes para, por exemplo, estudar as condições de vida de certas camadas da população.

Devemos ter presente que, contrariamente ao recenseamento, as sondagens inquiram ou analisam apenas uma parte da população em estudo, isto é, restringem-se a uma amostra dessa população, mas com o objectivo de extrapolar para todos os elementos da população os resultados observados na amostra.

Uma sondagem realiza-se em várias fases: escolha da amostra, obtenção da informação, análise dos dados e relatório final. Para que os resultados de uma sondagem sejam válidos há necessidade de essa amostra ser representativa da população. O processo de recolha da amostra, a amostragem, tem de ser efectuada com os cuidados adequados. Quando são usadas técnicas apropriadas e a amostra é

---

suficientemente grande, os resultados obtidos encontram-se em geral perto dos resultados que se obteriam, se fosse estudada toda a população.

### 3.3 População e amostra. Dados e variáveis

É desde os primeiros anos que os alunos devem compreender que dados são mais do que números e que a Estatística permite transformar dados em informação. Uma característica dos dados estatísticos é a variabilidade e é esta variabilidade que é objecto do estudo da Estatística.

Perante uma colecção de dados, há duas formas possíveis de abordar a sua análise, consoante o nosso interesse seja:

- o Apenas explorar a colecção de dados e encontrar padrões – esta colecção de dados é, por assim dizer, a **população** em estudo.
- o Extrapolar para um universo mais vasto os padrões encontrados na colecção de dados, a qual é parte (ou **amostra**) desse universo (ou **população**)<sup>5</sup>.

Para dar dois exemplos da nossa vida corrente, pense-se nos resultados obtidos quando se pergunta aos alunos da turma quantos irmãos têm e nos resultados obtidos numa sondagem, encomendada por um candidato às próximas eleições presidenciais. No primeiro caso, a população é a turma e os dados que se têm referem-se a toda a população. É este o contexto que, de um modo geral deve ser utilizado para os alunos mais novos, em que não se procura generalizar os resultados obtidos na análise dos dados. No segundo caso, os dados referem-se a uma pequena parte da população de interesse e procura-se, após a sua análise, generalizar para um conjunto mais vasto. A grande maioria das situações onde é necessária a utilização de metodologias estatísticas, enquadra-se neste segundo caso.

**População** – Conjunto de unidades individuais, que podem ser pessoas, animais ou resultados experimentais, com uma ou mais características em comum, que se pretendem analisar.

**Amostra** – Parte da população que é observada com o objectivo de obter informação para estudar a característica pretendida.

Se se observar toda a população diz-se que se faz um **Censo**.

Como dissemos anteriormente, o objectivo da Estatística é o estudo de populações, isto é, conjuntos de indivíduos (não necessariamente pessoas) com características comuns, que se pretendem conhecer. A uma característica comum, que assume valores diferentes de indivíduo para indivíduo, chamamos **variável**. Em termos mais precisos, uma **variável** é uma característica de um indivíduo ou objecto à qual se possa atribuir um número ou uma categoria. O indivíduo ou coisa relativamente ao qual se recolhe a informação é designado por **unidade observacional** ou **caso**. Ao resultado da observação da variável num indivíduo ou objecto, chamamos **dado estatístico** ou simplesmente **dado**.

Sendo então o nosso objectivo o estudo de uma (ou mais) característica da população, *vamos identificar população com a variável (característica) que se está a estu-*

---

<sup>5</sup> Esta secção segue de perto Graça Martins et al (2007) e Graça Martins (2006).

*dar*, dizendo que a população é constituída por todos os valores que a variável pode assumir. Por exemplo, relativamente à população portuguesa, se o objectivo do nosso estudo for a característica altura, diremos que a população é constituída por todos os valores possíveis para a variável altura.

Vimos também que amostras são *conjuntos de dados*, que representem convenientemente as populações de onde foram recolhidos. Do mesmo modo *identificaremos amostra* com os valores observados para a variável em estudo, sobre alguns elementos da população. Assim, na continuação do exemplo referido, os valores 156cm, 171cm, 163cm, 168cm, 166cm, obtidos ao medir a altura de 5 portugueses, constituem uma amostra da população a estudar.

Neste momento vamos admitir que dispomos de um desses *conjuntos de dados*, sem nos preocuparmos como foram obtidos, e pretendemos desenvolver processos de análise que nos permitam responder a algumas questões, tais como:

- Serão os dados quase todos iguais?
- Serão muito diferentes, uns dos outros?
- De que modo é que são diferentes?
- Existe alguma estrutura subjacente ou alguma tendência?
- Existem alguns agrupamentos especiais?
- Existem alguns dados muito diferentes da maior parte?

Estas questões, de um modo geral, não podem ser respondidas rapidamente, olhando unicamente para o conjunto dos dados! No entanto, se estiverem organizados sob a forma de tabelas ou gráficos, já a resposta às questões anteriores se torna mais simples. A metodologia estatística utilizada depende das variáveis que se estão a estudar, pelo que é importante começar por classificá-las. Uma classificação possível é a que se apresenta a seguir.

Uma variável diz-se **quantitativa** (ou numérica) se se referir a uma característica que se possa contar ou medir. Por exemplo, o número de irmãos de um aluno escolhido ao acaso, na turma, é uma variável quantitativa de contagem, enquanto que a sua altura é uma variável quantitativa de medição.

Uma variável diz-se **qualitativa** (ou categórica) se não for susceptível de medição ou contagem, mas unicamente de uma classificação, podendo assumir várias modalidades ou categorias. Por exemplo, a cor dos olhos do aluno referido anteriormente, é uma variável qualitativa. Se só assumir duas categorias, diz-se **binária**. É o caso da variável sexo, que assume as categorias Feminino e Masculino.

As variáveis quantitativas de contagem, isto é, que se referem a características que só se podem contar e não se podem medir, designam-se também por variáveis quantitativas **discretas**; por sua vez, as variáveis quantitativas de medição, isto é, que se podem medir, também se designam por variáveis quantitativas **contínuas**. Estas designações são bastante importantes, pois as ferramentas estatísticas a utilizar dependem do tipo de variável em estudo.

Algumas variáveis qualitativas apresentam uma ordem subjacente – são designadas por **qualitativas ordinais**. São exemplos de variáveis qualitativas ordinais: o *nível social* (com as categorias “baixo”, “médio” e “elevado”), o *grau de satisfação*

com um produto (com as categorias “nada satisfeito”, “pouco satisfeito”, “satisfeito”, “bastante satisfeito” e “muito satisfeito”) e grande parte das variáveis utilizadas em inquéritos na área das ciências sociais onde se avalia o nível atingido em cada variável solicitando ao respondente que coloque uma cruz numa grelha numerada de 1 a 5 (escala de Lickert).

No 1.º ciclo, logo desde o 1.º e 2.º anos de escolaridade, os alunos trabalham com dados qualitativos e dados quantitativos discretos. Também podem ser trabalhados dados de tipo contínuo, devidamente discretizados. No 2.º e 3.º ciclo trabalha-se com todo o tipo de dados, qualitativos e quantitativos, discretos e contínuos.

### Turma de referência

Para exemplificar os conceitos à medida que vão sendo introduzidos, vamos utilizar um conjunto de dados relativos aos alunos da turma (considerada turma de referência) e que podem ser obtidos, pedindo-lhes para preencherem a seguinte ficha:

Nome: \_\_\_\_\_

Número de letras no nome: \_\_\_\_\_

Número de irmãos: \_\_\_\_\_

Cor dos olhos: \_\_\_\_\_

Transporte utilizado para ir para a escola: \_\_\_\_\_

Tempo que demora de casa à escola (em minutos): \_\_\_\_\_

Comprimento do palmo (em centímetros): \_\_\_\_\_

O professor pode fazer algumas recomendações relativamente ao preenchimento da ficha, como por exemplo:

- Se os alunos utilizarem mais de um meio de transporte, consideram só o que utilizam na maior parte do tempo que levam de casa à escola;
- Para darem o tempo que demora de casa à escola, dão um valor aproximado ou então, no dia seguinte, têm o cuidado de verificar quanto tempo demoraram;
- Para medirem o comprimento do palmo utilizam uma régua, em que colocam o polegar da mão direita junto ao zero da régua e vêem até quantos centímetros chega o dedo mindinho.

Uma tabela possível, obtida numa turma de 24 alunos, é a seguinte:

Dados da turma

Nome	Número de letras no nome	Número de irmãos	Cor dos olhos	Transporte utilizado para ir de casa à escola	Tempo de casa à escola (minutos)	Comprimento do palmo (cm)
Ana Godinho	10	1	Castanhos	Autocarro	15	165
Ana Sofia Silva	13	2	Pretos	A pé	5	150
Andreia Sousa	12	0	Castanhos	Metro	14	173
Carolina Martins	15	0	Azuis	Carro	8	189
Daniela Silva	12	3	Castanhos	Carro	12	187
David Leal	9	1	Castanhos	Carro	10	195
Diogo Oliveira	12	4	Castanhos	A pé	13	137
Filipa Duarte	12	1	Verdes	Autocarro	20	166
Helena Afonso	12	2	Azuis	Carro	10	186
Inês Martins	11	1	Pretos	Carro	15	153
Joana Manso	10	0	Castanhos	Metro	17	159
João Miguel Ribeiro	17	1	Castanhos	Metro	13	144

João Pedro Batista	16	1	Castanhos	Metro	18	142
Liliana Isabel Cruz	17	2	Castanhos	Autocarro	9	155
Margarida Cabral	20	0	Castanhos	Autocarro	21	158
Miguel Esteves	13	2	Azuis	A pé	7	138
Nuno Pestana	11	3	Pretos	A pé	6	161
Patrícia Santos	14	1	Castanhos	Carro	11	163
Pedro Pinheiro	13	1	Verdes	A pé	12	172
Raquel Loureiro	14	0	Azuis	Metro	19	164
Rita Martins	11	2	Castanhos	Autocarro	15	165
Simão Valente	12	1	Castanhos	A pé	6	164
Sofia Matias	11	0	Castanhos	A pé	7	190
Tiago Neves	10	3	Castanhos	A pé	16	168

A tabela anterior apresenta os valores observados, nos alunos da turma, para as variáveis *Número de letras do nome*, *Número de irmãos*, *Cor dos olhos*, *Transporte utilizado para ir de casa à escola*, *Tempo que leva de casa à escola* e *Comprimento do palmo*. A metodologia utilizada para obter os dados da tabela, permite facilmente apercebermo-nos da natureza desses dados.

Assim, ao preencherem as fichas, a partir das quais se construiu a tabela, os alunos:

- Contaram o *número de letras do nome*, pelo que os dados correspondentes a essa variável são discretos;
- Mediram, com uma régua, o *comprimento do palmo*, pelos que os dados correspondentes a essa variável são contínuos;
- Mediram, com um relógio, o *tempo que demoram de casa à escola*, pelo que os dados respeitantes a essa variável são contínuos;
- Não puderam medir nem contar a *cor dos olhos*, mas unicamente atribuir-lhe uma categoria, pelo que a variável correspondente é qualitativa.

Não se deve insistir, junto dos alunos, sobretudo dos mais novos, nas designações das variáveis. O que deve ser realçado é que estas têm natureza diferente, e as ferramentas utilizadas para tratar os dados resultantes das observações dessas variáveis, também terão que ser diferentes, em algumas situações. Por exemplo, podemos representar dados qualitativos ou quantitativos discretos utilizando um diagrama de barras, mas não tem sentido utilizar esta mesma representação para dados contínuos, caso em que podemos usar um histograma ou gráfico de linha.

### Tarefa – Classificação de variáveis

1. Para cada uma das variáveis a seguir consideradas indica se é de natureza qualitativa ou quantitativa e neste caso se é discreta ou contínua:

- a) Número de pastilhas numa caixa de Smarties
- b) Cor do cabelo do primeiro colega que encontrar quando chegas à escola
- c) Idade do colega da alínea anterior
- d) Número de livros que compraste no último mês
- e) Marca do primeiro carro que passa, quando vais à janela
- f) Velocidade do carro da alínea anterior
- g) Tempo que levas de casa à escola
- h) Rendimento mensal de uma família
- i) Tempo de duração de uma chamada telefónica
- j) Número de mensagens que recibes no telemóvel (se não tiveres telemóvel, pode ser o do teu Pai), por dia

- k) Número de moedas que tens no porta-moedas
- l) Número de televisões numa casa
- m) O teu peso
- n) A tua altura
- o) A tua nota num teste de Matemática
- p) O teu programa favorito na TV

2. Dá alguns exemplos de variáveis qualitativas e quantitativas discretas e contínuas (não uses os exemplos do exercício anterior).

3. Algumas questões que talvez te possam interessar, são as seguintes:

- a) Há muitos colegas meus com animais domésticos?
- b) Quantos colegas haverá com 2 irmãos, como eu?
- c) Haverá mais colegas com telemóvel, ou sem telemóvel?

Como é que procederias para obter resposta para estas questões?

4. Dá exemplo de outras questões, para as quais seja necessário recolher dados, se se quiser conhecer a resposta.

### 3.4 “Limpar” os dados

É comum, quando se procede a uma análise de dados recolhidos verificar que estes contêm erros, acidentais ou não acidentais. Assim, antes de se proceder ao tratamento dos dados através de tabelas, gráficos ou do cálculo de medidas, deve-se olhar criticamente para os dados recolhidos, com o objectivo de os “limpar” dos erros. Por exemplo, se ao recolher informação sobre o tamanho do pé, se obtiver a informação de 300cm, obviamente que este valor está errado. Este erro pode ser acidental, nomeadamente ao digitar no computador o zero, carregou-se 2 vezes e ficaram 2 zeros. Se numa resposta sobre o ano de escolaridade, aparecer 1,2, também está errado, pois o ano de escolaridade tem de se exprimir na forma de um número inteiro. Estes erros podem ser acidentais, mas há outros que podem resultar de respostas dadas com pouco cuidado ou por brincadeira.

**Tarefa – Vamos limpar estes dados**<sup>6</sup>. Na tabela que se apresenta a seguir, alguns alunos mais brincalhões entretiveram-se a alterar alguns dos dados de uma tabela que contém respostas de alunos do ensino básico. Procura detectar esses erros e quando possível, sugere alterações de forma a ter dados “limpos”:

Sexo	Data de nascimento	Ano de escolaridade	Naturalidade	Altura	Tamanho do pé	Disciplina ou actividade preferida	Distância de casa à escola
M	12-04-1991	5	Portugal	143	26	Educação musical	de 1 a 2 km
F	31/02/92	4	Portugal	132	22	Estudo do Meio	menos de 2 km
F	14-01-1991	5.00	Portugal	14.2	2.3	Educação Física	2.5423 km
M	07-09-1989	6	Portugal	136	25	Matemática	de 1 a 2 km
M	13-12-1991	4	Angola	128	24	Língua Portuguesa	de 1 a 2 km
M	14-03-2001	5	Portugal	140	67	Matemática	menos de 1 km
F	06-05-1989	7	Moçambique	142	24	Língua Portuguesa	de 3 a 5 km
F	15-08-1990	6	Portugal	138	21	Língua Portuguesa	85km
M	20-02-1990	6	Portugal	192	23	Matemática	de 1 a 2 km

<sup>6</sup> Adaptada de uma actividade do Censusatchool.

M	19-05-1990	6	Portugal	140	20	Educação Física	de 1 a 2 km
O	29-06-1992	7	Lua	48	21	Estudo do Meio	3000km
M	09-10-1991	4	Cabo Verde	128	21	Língua Portuguesa	menos de 1 km
F	18-12-1990	5	Angola	135	21	Matemática	menos de 1 km
F	18-07-1991	0.5	Portugal	13.7	20	Ir para casa	de 3 a 5 km
M	03-06-1934	4	Portugal	129	21	Informática	menos de 1 km
F	13-02-1989	7	Moçambique	148	23	Matemática	de 1 a 2 km
F	15-09-1988	7	Portugal	150	22.5	Educação Física	de 1 a 2 km
F	07-08-1989	7	Portugal	140	24	Informática	menos de 1 km
M	08-06-1989	7	Angola	142	24	Matemática	menos de 1 km
M	31/11/87	11	Marte	1520	22	Informática	de 5 a 10 km
F/M	16-07-1988	8	Portugal	142	26	Chinês	de 2 a 3 km
F	28-04-1988	8	Portugal	145	26.5	Educação Física	1 kg
M	25-03-1992	4.1	Portugal	132.1	2.4.5	Matemática	menos de 1 km
M	26-02-1992	4	Portugal	130	21	Educação Física	menos de 1 km
F	08-07-1999	6	Portugal	142	22	Língua Portuguesa	de 2 a 3 km
M	23-05-1990	6	Cabo Verde	151	25.5	Matemática	de 2 a 3 km
M	01-03-1987	9	Angola	162	25	Educação Física	menos de 1 km
F	07-08-1991	6	Portugal	150	23	Educação musical	2 saltos
F	03-03-1992	4	Portugal	135	21	Informática	menos de 1 km

No exemplo anterior, alguns dos erros foram provocados deliberadamente com o objectivo de ilustrar uma situação que ocorre com frequência, sem ser intencional. Efectivamente, é comum haver um dígito repetido, uma vírgula a assinalar a casa decimal mal colocada, uma data trocada, etc. Outros erros que foram introduzidos deliberadamente, retratam situações intencionais, como é o caso de responder Marte ou Lua à pergunta sobre a naturalidade. Assim, antes de começar a tratar um conjunto de dados, deve ter-se um olhar crítico para detectar este tipo de erros que podem destruir toda uma análise subsequente.

Por exemplo, referindo-nos ainda à tabela do exemplo anterior, se se proceder ao cálculo da média dos valores referentes à altura, obtém-se o valor aproximadamente igual a 178cm. No entanto se limparmos os dados de alguns erros óbvios, nomeadamente os assinalados na tabela seguinte

Antes	Depois
14,2	142
192	eliminar
48	148
13,7	137
1520	152
132,1	132

já o valor obtido para a média vem aproximadamente igual a 141 cm.

Já anteriormente chamámos a atenção para o facto de "...os dados são mais do que números, são números com um contexto". Ora foi precisamente esse contexto que nos guiou nas alterações a fazer. Se todas as alterações, com excepção de uma, são mais ou menos óbvias, não tínhamos alternativa para o valor 192, pelo que decidimos eliminá-lo, não sem custos, pois reduzimos a nossa amostra de uma unidade, passando a ter só 28 dados em vez dos 29 iniciais.

**Tarefa – Não serão irmãos a mais?** Registou-se numa tabela de frequências que se apresenta a seguir, o resultado de um inquérito feito junto de alunos do 1.º ciclo do ensino básico de várias escolas, em que se colocava, entre outras, a seguinte questão *Quantos irmãos tens?*

Tabela 1

Número de irmãos	Freq. Rel (%)
0	16
1	51
2	18
3	7
4 ou mais	8

A seguir apresentam-se as respostas de alguns alunos de uma escola sobre algumas questões, entre as quais também se pergunta quantos irmãos têm:

Tabela 2

Sexo	Data de nascimento	Código Postal	Nº de irmãos	Nº irmãos com menos 18 anos
M	12-04-1991	1050-027	0	0
F	27-02-1992	1200-013	1	1
F	14-01-1991	1150-110	2	2
M	07-09-1989	1100-115	1	1
M	13-12-1991	1070-031	1	1
M	14-03-2001	1121-025	0	0
F	06-05-1989	1150-043	2	1
F	15-08-1990	1070-014	1	1
M	20-02-1990	1050-006	2	2
M	19-05-1990	1075-100	1	1
M	29-06-1992	1180-121	5	4
M	09-10-1991	1210-121	3	3
F	18-12-1990	1170-114	1	1
F	18-07-1991	1032-045	1	1
M	03-06-1991	1180-121	5	4

Haverá alguma coisa de estranho nesta tabela?

De acordo com a Tabela 1, qual a percentagem de alunos com 4 ou mais irmãos? Esse resultado não nos fará pensar se algo de anormal não se passará com as respostas dadas na Tabela 2? (Observe-se que, de acordo com a tabela 1, a percentagem de alunos do 1º ciclo com 4 ou mais irmãos anda à volta de 8%. De acordo com a tabela 2, aparentemente em 15 alunos, 2 têm 5 irmãos! No entanto, se repararmos melhor, verificamos que a morada é a mesma, pelo que afinal os 2 alunos que responderam são, com elevada probabilidade, irmãos...)

PROJECTO\*

# OS CENSOS VÃO ÀS ESCOLAS

Ensino Básico - 1º Ciclo



INSTITUTO NACIONAL DE ESTATÍSTICA  
PORTUGAL

**GABINETE DOS CENSOS 2001**

Avenida António José de Almeida

1000-043 Lisboa

Telef.: 21 842 61 00 Fax: 21 842 63

58

O projecto "OS CENSOS VÃO ÀS ESCOLAS" foi desenvolvido pelo Gabinete dos Censos 2001 e teve como objectivos:

- Dar a conhecer aos alunos dos diversos graus de ensino: o que são, para que servem e como se fazem os Censos;
- Mobilizar os pais e familiares dos alunos para a participação nos CENSOS 2001.

Este projecto consistiu numa aula relativa aos Censos, que foi ministrada em todas as escolas do ensino oficial e particular na primeira quinzena de Março de 2001.

Foram desenvolvidos três tipos de aulas de acordo com o nível de ensino:

NÍVEL DE ENSINO	IDADES	LIGAÇÕES CURRICULARES
Ensino Básico - 1º Ciclo	6-10	Estudo do Meio
Ensino Básico - 2º e 3º Ciclos	10-15	História e Geografia de Portugal Geografia História Matemática
Ensino Secundário	15-18	Geografia Matemática História Economia Introdução ao Desenvolvimento Económico e Social

A seguir apresentamos a parte do projecto referente aos 1º, 2º e 3º ciclos.

## Ensino Básico - 1º Ciclo

### 1. O QUE SÃO OS CENSOS?

Os Censos são a contagem de todas as pessoas que vivem no nosso País e de todas as habitações onde as pessoas vivem.

Com os Censos ficamos a saber:

**Quantos somos** - o número total de pessoas que vivem em Portugal de norte a sul, Regiões Autónomas dos Açores e da Madeira, e em cada uma das nossas cidades, vilas e aldeias;

**Como somos** - a idade das pessoas, as profissões, os estudos que as pessoas têm;

**Onde vivemos** - os locais onde as pessoas vivem;

**Como vivemos** - as características das habitações existentes em Portugal.

### 2. PARA QUE SERVEM OS CENSOS?

É muito importante conhecermos quantas são e como são as pessoas que vivem em Portugal e em cada uma das nossas cidades ou bairros para sabermos:

- O número de escolas, creches, lares de idosos que são necessários;
- Onde se devem construir as vias de comunicação, os hospitais, as fábricas,....;
- O número de representantes que cada região tem na Assembleia da República;
- Como distribuir o dinheiro pelas Câmaras Municipais.

Os resultados dos Censos são muito importantes porque servem para conhecer melhor o presente e preparar o futuro.

### 3. QUEM FAZ OS CENSOS?

O Instituto Nacional de Estatística (INE) é o organismo encarregue da preparação, execução e apuramento dos dados dos Censos 2001.

Dada a complexidade da operação estatística "Censos 2001", o INE tem a colaboração das Câmaras Municipais - responsabilizam-se pela organização, coordenação e controlo das tarefas do recenseamento na área do município - e das Juntas de Freguesia que asseguram a execução das operações dos Censos 2001 nas respectivas áreas.

#### **4. COMO SE FAZEM OS CENSOS ?**

A contagem das pessoas e das habitações é feita através do preenchimento de questionários.

Os questionários são entregues em cada casa e depois de preenchidos pelas pessoas são recolhidos por um recenseador (pessoa que distribui e recolhe os questionários).

Para tudo isto ser possível é muito importante que nas nossas casas seja preenchido um questionário por cada pessoa que lá vive.

**Tu também contas!**

#### **ACTIVIDADES:**

- Dar a preencher aos alunos a ficha de trabalho " O Meu Censo".
  - Depois de preenchidas as fichas apurar os resultados da turma de modo a ser possível responder a algumas perguntas:
    - Quantos rapazes e quantas raparigas existem na turma?
    - Quantos alunos têm 6, 7 ou 8 anos?
    - Quantos alunos nasceram em determinado local?
    - Quantos alunos têm irmãos?
-

# O MEU CENSO

Nome: \_\_\_\_\_

Menino

Menina

Quantos anos tens?

Local onde moras? \_\_\_\_\_



Local onde nasceste? \_\_\_\_\_



Quantas pessoas vivem em tua casa?



Tens irmãos ou irmãs? Sim  Não



# Ensino Básico 2º e 3º Ciclo

## 1. O QUE SÃO OS CENSOS?

A forma mais antiga e também mais directa de conhecer o número de pessoas que habitam um determinado território, consiste em realizar uma contagem através da observação exaustiva dos indivíduos, a que se dá o nome de recenseamento ou de uma forma mais abreviada "censo".

Os Censos são tradicionalmente a contagem da população de um país a que, em época mais recente, se acrescentou a sua melhor caracterização e um levantamento do parque habitacional.

É exactamente através dos Censos que o país fica a saber :

**Quantos somos?** - O total de pessoas a viver em Portugal Continental, Regiões Autónomas dos Açores e da Madeira e em cada um dos seus concelhos, freguesias ou bairros.

**Como somos?** - As características da população: sexo, idade, nacionalidade, naturalidade, profissão, grau de instrução, estatuto socio-profissional, etc .

**Onde vivemos?** - Os locais onde vivemos.

**Como vivemos?** - As características dos alojamentos onde vivemos.

### Breve História dos Censos

Já antes da era de Cristo se faziam recenseamentos, geralmente com objectivos militares e de cobrança de impostos. Por isso, a norma era a de as populações se deslocarem aos seus locais de origem e se apresentarem às respectivas autoridades para o registo de pessoas e/ou bens.

O primeiro censo populacional conhecido no território que é hoje Portugal foi realizado no ano zero, por ordem do Imperador César Augusto e dizia respeito à então província romana da Lusitânia. Posteriormente, na Idade Média também os Árabes efectuaram vários recenseamentos durante a sua permanência na Península Ibérica.

Já após a fundação da nacionalidade foram realizadas várias contagens mais ou menos extensas tendo preocupações sobretudo de ordem militar. A primeira destas operações foi o Rol de Besteiros do Conto, de D. Afonso III (1260-1279).

Em 1864, realizou-se o I Recenseamento Geral da população portuguesa, que foi o primeiro a reger-se pelas orientações internacionais do Congresso Internacional de Estatística de Bruxelas em 1853, marcando o início dos recenseamentos da época moderna.

Embora estas orientações já indicassem que os recenseamentos deveriam ser realizados de 10 em 10 anos o censo seguinte apenas se realizou em 1878, ao qual se seguiria o Censo de 1890. A partir de então os recenseamentos da população têm vindo a realizar-se, com poucas excepções, regularmente em intervalos de 10 anos.

Outro marco importante ocorreu em 1970, quando em simultâneo com o Recenseamento da População se realizou o I Recenseamento da Habitação.

O último censo realizado em Portugal foi em 1991.

## **2. PARA QUE SERVEM OS CENSOS?**

Através dos Censos é possível obter, para cada nível de detalhe geográfico (regiões, concelhos, freguesias, lugares,...) uma "fotografia" de todos os indivíduos residentes em Portugal e das condições em que habitam.

Os censos são uma fonte única e renovável que, caracterizando a população e o parque habitacional, surge como valioso instrumento de diagnóstico, planeamento e intervenção, em vários domínios:

- Na definição de objectivos e prioridades para as políticas globais de desenvolvimento (investimentos em educação, saúde, habitação ou transportes, medidas de combate ao desemprego, melhoria das condições de habitação, distribuição de fundos a nível regional e local,...);
- No planeamento regional e urbano (localização de escolas, hospitais, vias de comunicação, fábricas, etc);
- Nos estudos de mercado e sondagens de opinião;
- Na investigação em ciências sociais e políticas (elaboração de estudos no domínio económico e social,...)

Assim, os dados recolhidos pelos censos, sobre a população e a habitação, são fundamentais para proporcionar, ao governo e às autarquias locais, informação básica indispensável à definição e execução das suas políticas. Revestem-se, por isso, do maior interesse para toda a sociedade.

Os resultados dos Censos são fundamentais para conhecer o presente e preparar melhor o futuro do País.

## **3. QUEM FAZ OS CENSOS?**

O Instituto Nacional de Estatística (INE) é o organismo encarregue da preparação, execução e apuramento dos dados dos Censos 2001.

Dada a complexidade da operação estatística "Censos 2001", o INE tem a colaboração das autarquias locais. Assim, as Câmaras Municipais responsabilizam-se pela organização, coordenação e controlo das tarefas do recenseamento na área da respectiva jurisdição; enquanto que as Juntas de Freguesia asseguram a execução das operações dos Censos 2001 nas respectivas áreas.

---

#### **4. COMO SE FAZEM OS CENSOS ?**

A recolha de dados é realizada através do preenchimento de vários tipos de questionários de acordo com a unidade estatística a caracterizar: edifício, alojamento, família e indivíduo.

Os questionários são distribuídos em cada alojamento e depois de devidamente preenchidos são recolhidos por um recenseador (pessoa que distribui e recolhe os questionários) que, em caso de dificuldade, ajuda no preenchimento.

Assim, todos os alojamentos serão observados e todas as pessoas residentes serão caracterizadas através de questionários.

A informação recolhida refere-se às 0 horas do dia 12 de Março de 2001 - "**momento censitário**", que corresponde ao dia e hora em relação aos quais se recolhem os dados.

Para tudo isto ser possível é muito importante que nas nossas casas seja preenchido um questionário por cada pessoa que lá reside ou esteja temporariamente presente.

**Tu também contas!**

---

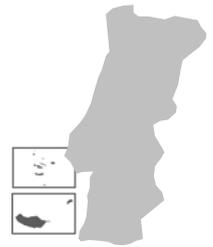
## ACTIVIDADES:

# Os Censos

- Ficha de Trabalho.

## Completa as seguintes frases:

1. Os Censos contam todos os residentes em \_\_\_\_\_ .



2. Os Censos realizam-se de \_\_\_\_ em \_\_\_\_ anos.

3. Os dados censitários são fundamentais para a analisar o P\_\_ S\_\_ T\_\_ e preparar o F\_\_ UR\_\_ do país.

4. Os indivíduos responsáveis pela distribuição e recolha dos questionários designam-se por R\_C\_\_S\_\_O\_ES.

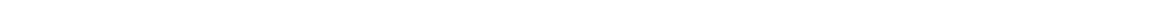
**Descobre as palavras relacionadas com os Censos - População, Habitação, Edifício, Alojamento, Família, Indivíduo.**

G A H E N O L I A P I A T E L R O R P H  
 E A R H T E T O E R N J F C Y T O B P A  
 H L O H S A P J O E D Ç E D I F Í C I O  
 A O S D A I U Ç O T I L Ã O K J D L A U  
 D J D I F B C I A D V O P A R P O T U P  
 R A S R A R I L P E I G M H A I R O Ç J  
 A M T P B H Ã T R I D P O P U L A Ç Ã O  
 P E E U A O A L A E U O L O F G O A M J  
 G N Ç R U T Ç E B Ç O N F U L A Ç E O T  
 R T A A L W A O B I Ã R L E Q Ç Q H B A  
 E O F D I G Í M I O P O F A M I L I A R  
 O A P O Ç A L Ã H A S O B Ç T O F Ã E O



## 4 Tabelas e gráficos

Apresentamos alguns processos, nomeadamente tabelas e gráficos, para organizar a informação contida nos dados, de forma a realçar as suas características mais importantes.





## 4.1 Introdução

Neste capítulo apresentamos formas de organizar os dados através de tabelas e gráficos. De acordo com o Programa de Matemática, logo nos 1.º e 2.º anos de escolaridade, os alunos aprendem a classificar dados utilizando diagramas de Venn e Carroll e aprendem a trabalhar com tabelas de frequências absolutas, gráficos de pontos e pictogramas. Pretende-se que os alunos saibam ler e interpretar informação registada por estes meios, e que saibam também construir representações destes tipos para um certo conjunto de dados. Nos 3.º e 4.º anos os processos de representação são alargados aos gráficos de barras. Mais tarde, no 2.º ciclo, os alunos aprendem a trabalhar com tabelas de frequências relativas e continuam a trabalhar com as representações anteriores, aprendendo também a construir e interpretar gráficos circulares e de linha, bem como diagramas de caule e folhas. Finalmente, no 3.º ciclo, os alunos trabalham com histogramas e diagramas de extremos e quartis.

As tabelas e os gráficos são instrumentos essenciais à representação e análise de dados, que os alunos devem aprender a usar com desembaraço. Isso tanto pode ocorrer na realização de investigações estatísticas, nas quais os alunos recolhem os seus próprios dados, usualmente organizados em grupos, como em tarefas mais estruturadas como as que apresentamos ao longo deste capítulo. O trabalho deve ter sempre como ponto de partida situações do dia-a-dia dos alunos ou situações com as quais eles sejam familiares. Toda a representação de dados em tabelas ou gráficos deve ser motivada por uma ou mais questões e depois dos dados representados deve indagar-se que outras questões seria ainda possível responder. É importante que os alunos ganhem sensibilidade para as potencialidades das diversas formas de representação dos dados e a sua adequação em função da natureza das variáveis em jogo (qualitativas, nominais ou ordinais, e quantitativas, discretas ou contínuas), e também para alguns aspectos que facilmente induzem em erro, como aqui apresentamos.

## 4.2 Diagramas de Venn e de Carroll

Começamos por referir algumas formas de classificar e organizar números ou objectos em listas ou tabelas simples, que não se podendo chamar propriamente instrumentos estatísticos, ajudam a organizar de uma forma simples alguns tipos de informação.

Os **diagramas de Venn** são representações gráficas particularmente adequadas para os alunos mais novos. Utilizam círculos ou rectângulos para uma classificação rápida de objectos ou números, que partilhem características comuns. Usualmente, considera-se um rectângulo que representa todo o conjunto a ser classificado, e dentro desse rectângulo consideram-se círculos que representam os elementos com as características de interesse.

Os **diagramas de Carroll** são tabelas rectangulares para organizar dados ou objectos segundo critérios de sim/não. O nome atribuído a estes diagramas, é uma homenagem a Lewis Carroll, matemático e escritor inglês, que gostava muito de problemas de lógica e de jogos matemáticos.

**Tarefa – Número de letras do nome.** Num primeiro momento, o professor desenha no quadro 2 círculos com a indicação de que num círculo se colocam nomes com 4 ou menos letras e no outro nomes com 4 ou mais letras. Depois, escolhe ao acaso 10 alunos, para irem ao quadro, colocar o seu nome no círculo adequado.

No fim, uma representação possível, em diagrama de Venn, pode ser a seguinte:



Algumas questões que podem ser colocadas:

1. Quantos alunos têm 4 letras no nome?
2. Quantos alunos têm mais de 4 letras no nome?

A seguir, o professor pede aos alunos que disponham os nomes no seguinte diagrama de Carroll, de acordo com os critérios indicados:

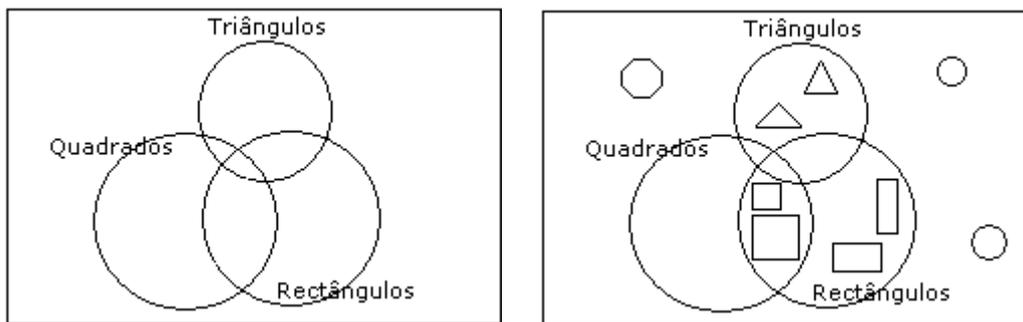
	Nome com 4 ou menos letras	Nome com mais de 4 letras
Rapariga		
Rapaz		

Depois da tabela completa, algumas questões que se podem colocar:

1. Quantos rapazes têm o nome na tabela?
2. Quantas raparigas têm o nome com 4 ou menos letras?

**Tarefa – Figuras geométricas.** O professor pede a cada aluno que desenhe no caderno uma figura geométrica. De seguida pede a vários alunos para irem ao quadro e representarem no seguinte diagrama de Venn, a figura que tinham desenhado no caderno.

Pode surgir, por exemplo:

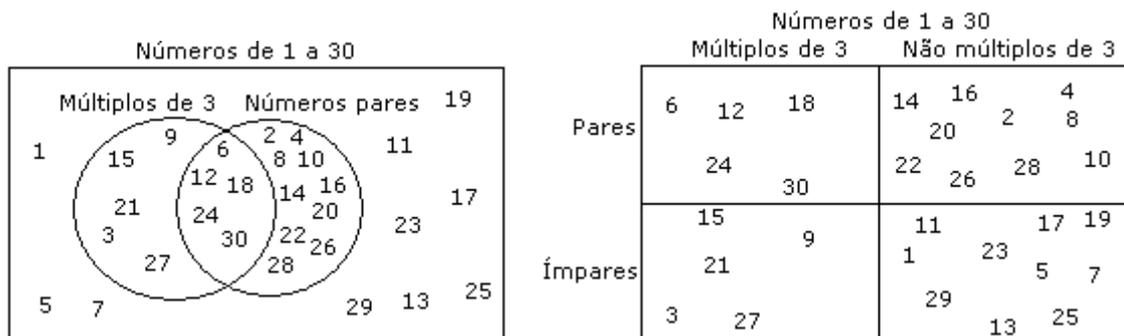


Algumas questões:

1. O que é um quadrado?
2. Um quadrado é um rectângulo?
3. E um rectângulo é um quadrado?
4. Quantos alunos desenharam figuras?
5. Quantas das figuras desenhadas não são nem Triângulos, nem Rectângulos?
6. E quantas das figuras desenhadas não são nem Triângulos, nem Rectângulos, nem Quadrados?
7. Como se explica que a resposta às questões 5. e 6. seja exactamente a mesma?

**Tarefa – Números de 1 a 30.** O professor pede aos alunos que classifiquem num diagrama de Venn e noutro de Carroll, os números de 1 a 30, segundo os seguintes critérios: ser ou não múltiplo de 3 e ser ou não par.

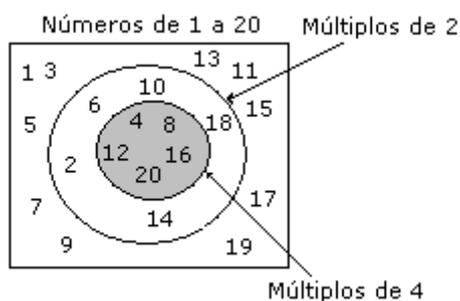
Duas representações possíveis são:



Algumas questões que se podem colocar, relativas a ambos os diagramas:

1. Quantos múltiplos de 3 são números pares?
2. Quais os números que não são pares nem múltiplos de 3?
3. E que números são simultaneamente pares e múltiplos de três?

**Tarefa – Múltiplos de 2 e 4, até 20.** Representar num diagrama de Venn, os números de 1 a 20 que sejam múltiplos de 2 e também múltiplos de 4.



Algumas questões: Dos números de 1 a 20,

1. Quantos são múltiplos de 4?
2. Quantos são múltiplos de 2?
3. Dê exemplo de um múltiplo de 2 que não seja múltiplo de 4.
4. Todos os múltiplos de 4 são múltiplos de 2?

**Tarefa – Figuras e sólidos.** O professor pede aos alunos para classificarem no diagrama de Carroll, que se apresenta, as seguintes figuras no plano e sólidos geométricos:

Cubo, quadrado, triângulo equilátero, pirâmide, cilindro, retângulo, trapézio, esfera, círculo, cone, prisma triangular, prisma com base quadrangular, paralelogramo, paralelepípedo.

		linhas perpendiculares	
		Sim	Não
Figuras planas			
Sólidos			

Algumas questões:

1. Quais as figuras planas sem linhas perpendiculares?
2. Quais os sólidos com linhas perpendiculares?
3. No conjunto indicado, há mais figuras planas ou sólidos geométricos?

**Tarefa – Ajudas a lavar a loiça?** Foi feito um inquérito numa escola onde se perguntava “em tua casa, ajudas a lavar a loiça?”. Responderam 258 alunos, dos quais 175 eram raparigas. Responderam afirmativamente à pergunta 118 raparigas e 51 rapazes. Preenche a tabela seguinte:

	Rapariga	Rapaz	Total
Ajuda			
Não ajuda			
Total			

A partir da tabela anterior, completa o seguinte diagrama de Venn:



### 4.3 Tabelas e gráficos para dados qualitativos

Como dissemos anteriormente, os dados qualitativos são os que resultam da observação de variáveis qualitativas. Representam a informação que identifica alguma qualidade não susceptível de medição ou contagem, mas unicamente de classificação, podendo assumir várias categorias ou modalidades. Por exemplo, o estado civil de um indivíduo pode assumir as categorias solteiro, casado, viúvo ou divorciado.

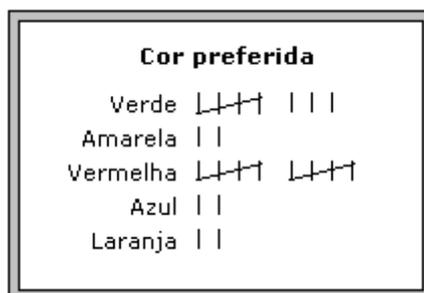
Por vezes codificam-se as variáveis qualitativas com números, como por exemplo, no caso da variável sexo, em que se pode representar o sexo masculino por 1 e o feminino por 2. No entanto, o facto de as categorias estarem representadas por números, não leva a variável a mudar de natureza, pelo que, por exemplo, não faz qualquer sentido calcular a média destas observações.

#### 4.3.1 Esquemas de contagem gráfica (tally charts)

Existem algumas representações muito simples que se podem construir directamente a partir do conjunto de dados ou durante o processo de recolha. Uma maneira possível de ir registando os dados, à medida que os vamos recolhendo, é utilizar o esquema de contagem gráfica (*tally chart*). Por exemplo, pretende-se averiguar, na turma, qual a cor preferida dos alunos. Então os alunos vão, um a um, ao quadro registar a sua cor preferida, do seguinte modo:

- O primeiro aluno, que prefere a cor verde, escreve Verde e à frente desenha um traço;
- O aluno seguinte que prefere a cor amarela, escreve Amarela e à frente um traço;
- A seguir vem outro aluno que prefere a cor verde e coloca um traço ao lado do que já lá estava;
- E assim sucessivamente, os alunos vão escrevendo as cores se é a primeira vez que aparecem ou colocando traços à frente das cores que já estão no quadro. O quinto traço coloca-se de forma oblíqua a cortar os 4 traços anteriores.

No fim obtém-se um esquema idêntico ao seguinte:



Da representação anterior, imediatamente se conclui que a cor preferida é a Vermelha, seguindo-se a Verde. Estes resultados podem sugerir ao professor que questione os alunos sobre qual o seu clube de futebol preferido. Será que as preferências de cor têm a ver com as preferências clubísticas?

Um esquema de contagem gráfica para a variável *Cor dos olhos* dos alunos da turma em referência é o seguinte:

Castanhos			
Pretos			
Verdes			
Azuis			

Como se verifica, predominam os olhos Castanhos, seguindo-se os Azuis. Só 2 alunos têm olhos Verdes.

Este esquema de contagem gráfica tem a grande vantagem de:

- permitir identificar as diferentes categorias ou modalidades que a variável qualitativa pode assumir no conjunto dos dados e
- permitir organizar os dados de tal maneira que facilmente se conta o número de elementos (frequências absolutas) em cada uma dessas categorias.

Não sendo um passo necessário para a construção das tabelas de frequência (que a seguir se apresentam), é um passo que, uma vez concluído, serve de base para a construção dessas tabelas.

**Tarefa – O mês de aniversário.** O professor propõe à turma averiguar qual o mês em que há mais alunos a fazer anos. Então distribui a seguinte folha, que passa de aluno para aluno, até todos terem assinalado com um traço o mês do seu aniversário:

Mês	
Janeiro	
Fevereiro	
Março	
Abril	
Maio	
Junho	
Julho	
Agosto	
Setembro	
Outubro	
Novembro	
Dezembro	

- Qual o mês em que há mais alunos a fazer anos?
- Qual o mês em que há menos alunos a fazer anos?
- Quantos alunos estavam na turma, no dia em que se realizou esta tarefa?

#### 4.3.2 Tabela de frequências para dados qualitativos

Os dados qualitativos são organizados na forma de **tabelas de frequências**, com duas ou mais colunas. Na primeira coluna, *coluna das categorias* ou classes, indicam-se todas as categorias presentes no conjunto de dados ou amostra a analisar; na coluna seguinte, *coluna das frequências absolutas*, regista-se o número de elementos da amostra, que pertencem a cada categoria (ou classe). É usual juntar

uma terceira coluna, *coluna das frequências relativas*, onde se regista, para cada categoria (ou classe) o valor que se obtém dividindo a frequência absoluta dessa categoria pela dimensão da amostra (número de elementos).

**Frequência absoluta** de uma categoria ou classe, é o número de elementos da amostra iguais a cada uma das categorias;

$$\text{Frequência relativa} = \frac{\text{frequência absoluta}}{\text{dimensão da amostra}}$$

Uma tabela de frequências reflecte a forma da *distribuição* da variável em estudo, na amostra considerada, isto é, quais as categorias ou modalidades que assume, assim como a frequência (absoluta e/ou relativa) com que assume essas modalidades. Enquanto os alunos não conhecerem fracções ou numerais decimais, utilizam-se unicamente frequências absolutas na construção das tabelas de frequências.

Para o conjunto de dados da turma, vamos construir a tabela de frequências respeitante às variáveis *Cor dos olhos* e *Transporte utilizado para ir de casa à escola*:

<i>Cor dos olhos</i>		<i>Transporte utilizado</i>	
<u>Categorias</u>	<u>Número de alunos</u>	<u>Categorias</u>	<u>Número de alunos</u>
Castanhos	15	Autocarro	5
Pretos	3	A pé	8
Verdes	2	Metro	5
Azuis	4	Carro	6

A tabela correspondente à variável *Cor dos olhos* foi obtida a partir do esquema de contagem gráfica construído para esta variável, na secção anterior.

A partir das tabelas construídas, pode dar-se resposta a algumas questões. Por exemplo, a partir da tabela respeitante à variável *Transporte utilizado*, algumas questões são:

- Qual o(s) meio(s) de transporte mais utilizado?
- Qual o(s) meio(s) de transporte menos utilizado?
- Quantos alunos vão de carro ou de metro?
- A partir dos dados apresentados, é de admitir que um número razoável de alunos mora perto da escola?
- Quantos alunos tem a turma? (admita que cada aluno preencheu uma das fichas a partir das quais se construiu o ficheiro Dados da Turma).

Se, ao organizar um conjunto de dados qualitativos, se verificar que existe uma categoria predominante dá-se-lhe o nome de **moda**. Assim, no que diz respeito à variável *Cor dos olhos*, a moda é a categoria "Olhos castanhos", enquanto no que diz respeito à variável *Transporte utilizado*, a moda é o "A pé".

Estas tabelas podem ser complementadas com a coluna das frequências relativas:

<i>Cor dos olhos</i>			<i>Transporte utilizado</i>		
<u>Categorias</u>	<u>Freq. abs.</u>	<u>Freq. rel.</u>	<u>Categorias</u>	<u>Freq. abs.</u>	<u>Freq. rel.</u>
Castanhos	15	0,625	Autocarro	5	0,208
Pretos	3	0,125	A pé	8	0,333
Verdes	2	0,083	Metro	5	0,208
Azuis	4	0,167	Carro	6	0,250
Total	24	1	Total	24	1

Nas tabelas anteriores introduzimos também uma linha com os totais das colunas. Esta metodologia é aconselhável, pois é um processo de verificação de que as frequências devem estar bem calculadas, já que:

- A soma das frequências absolutas é igual à dimensão da amostra;
- A soma das frequências relativas é igual a 1.

Esta última condição nem sempre se verifica devido ao facto de algumas frequências relativas serem dízimas infinitas, obrigando a arredondamentos. Por exemplo, se somarmos as frequências relativas na tabela que diz respeito à variável *Transporte utilizado*, obtemos o valor 0,999, que arredondado às unidades dá 1.

A utilização das frequências relativas é aconselhada quando se pretendem comparar duas amostras que digam respeito à mesma variável, mas que tenham dimensão diferente.

**Exemplo – Animal doméstico preferido dos alunos das turmas A e B.** Os professores de duas turmas da escola, A e B, pretendem averiguar se os alunos têm gostos idênticos relativamente ao animal doméstico preferido. Assim, em cada turma os alunos disseram qual o animal doméstico preferido e construíram as tabelas de frequência respectivas:

Animal doméstico preferido Turma A		Animal doméstico preferido Turma B	
Categoria	N.º de alunos	Categoria	N.º de alunos
Cão	11	Cão	14
Gato	5	Gato	6
Passarinho(s)	3	Passarinho(s)	4
Peixe(s)	1	Peixe(s)	2

As duas tabelas foram apresentadas nas duas turmas (em conjunto) e houve alguns alunos que, tendo em conta os dados apresentados, exprimiram as suas opiniões:

- Na turma B há mais alunos do que na turma A a preferirem o Cão;
- Na turma B há o dobro dos alunos da turma A que preferem os Peixes.

Será que estas conclusões estão correctas?

Na verdade, as conclusões não estão correctas pois estão baseadas nas frequências absolutas e as turmas não têm o mesmo número de alunos. Assim, devem-se calcular as frequências relativas, para se poderem tirar conclusões correctas, no que diz respeito à comparação das turmas. Adicionando uma coluna com as frequências relativas a cada uma das tabelas, temos:

Animal doméstico preferido Turma A			Animal doméstico preferido Turma B		
Categoria	N.º de alunos	Freq. relativa	Categoria	N.º de alunos	Freq. relativa
Cão	11	0,55	Cão	14	0,54
Gato	5	0,25	Gato	6	0,23
Passarinho(s)	3	0,15	Passarinho(s)	4	0,15
Peixe(s)	1	0,05	Peixe(s)	2	0,08
Total	20	1		26	1

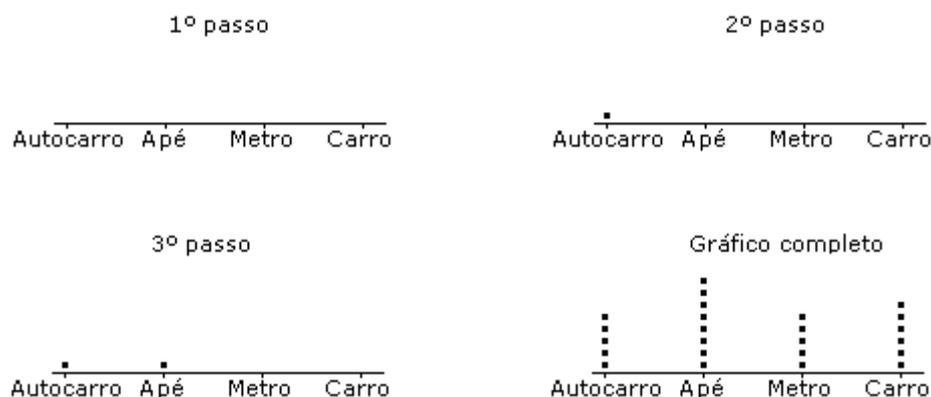
Ao compararmos as frequências relativas, verificamos que, afinal, na turma A há uma maior (embora pequena) frequência de alunos a preferirem o Cão. Verificamos também que, nas duas turmas, existe igual preferência pelos Passarinhos e que, embora haja 2 vezes mais alunos da turma B do que da turma A, a preferirem os peixes, não podemos dizer que na turma B existe o dobro dos alunos da turma A, a preferirem esse animal.

### 4.3.3 Gráfico de pontos e gráfico de barras para dados qualitativos

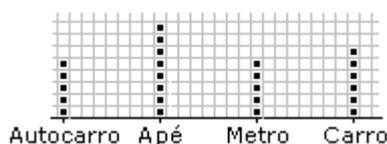
#### 4.3.3.1 Gráfico de pontos

A representação gráfica mais simples que se pode obter e que não necessita de nenhuma organização prévia dos dados, é o gráfico ou diagrama de pontos. Tal como o esquema de contagem gráfica, é uma representação que se pode ir construindo, no caso dos dados qualitativos, à medida que se recolhem os dados.

Começa-se por desenhar um eixo horizontal (ou vertical), onde se assinalam (igualmente espaçadas) as diferentes categorias ou modalidades que a variável assume no conjunto dos dados. Por cima de cada categoria (ou ao lado), marca-se um ponto sempre que ao recolher um dado ou ao percorrer o conjunto dos dados se encontrar um elemento da respectiva categoria. Por exemplo, para os dados da turma de referência, podemos distinguir os seguintes passos na construção do gráfico de pontos para a variável *Transporte utilizado*:



A construção do gráfico de pontos é facilitada se se utilizar papel quadriculado. Neste caso desenha-se um ponto por quadrícula, como se apresenta a seguir:



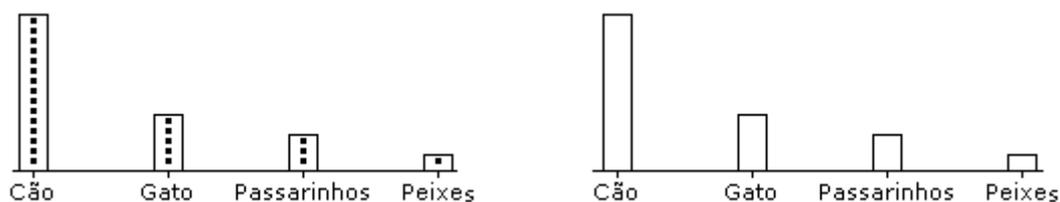
A organização dos dados num gráfico de pontos permite visualizar quais as categorias que predominam e quais as menos frequentes.

Ao investigarem qual o animal doméstico preferido, a metodologia a seguir para a recolha de dados pelos alunos pode ser a utilizada no esquema de contagem gráfica. O primeiro aluno a ir ao quadro, que já manifestou oralmente a sua preferência

pelo Cão, começa por desenhar um eixo horizontal ou vertical, onde assinala uma posição para situar a categoria Cão, escrevendo por baixo o nome Cão e, por cima, desenhando um ponto. O aluno seguinte, que prefere o gato, assinala uma posição para a categoria Gato e procede como o aluno anterior. Os outros alunos vão desenhando pontos em cima dos que já lá estão ou acrescentando categorias, em posições igualmente espaçadas umas das outras. Se esta investigação tivesse sido colocada à turma A da secção anterior, o resultado seria o seguinte:

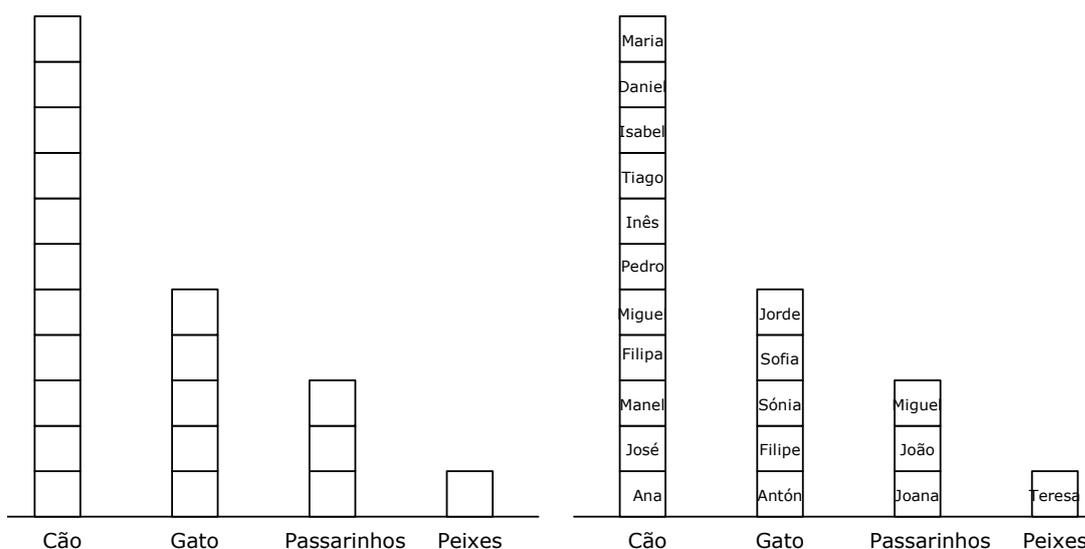


Se na representação gráfica anterior se envolverem os pontos com um rectângulo e a seguir se apagarem os pontos, obtemos um gráfico de barras, que é objecto de estudo da secção seguinte:



### Alternativa ao gráfico de pontos

Uma alternativa ao gráfico de pontos consiste em desenhar quadrados em vez de pontos. Enquanto que no gráfico de pontos a preferência de cada aluno é representado por um ponto, nesta representação alternativa, cada aluno desenha um quadrado, obtendo-se um gráfico como o que se apresenta a seguir:



Esta representação é muito aliciante para os alunos, pois cada um tem a oportunidade de colocar o seu nome no quadrado correspondente ao seu animal preferido.

Para se obter a frequência em cada animal doméstico, basta agora contar os quadrados respectivos

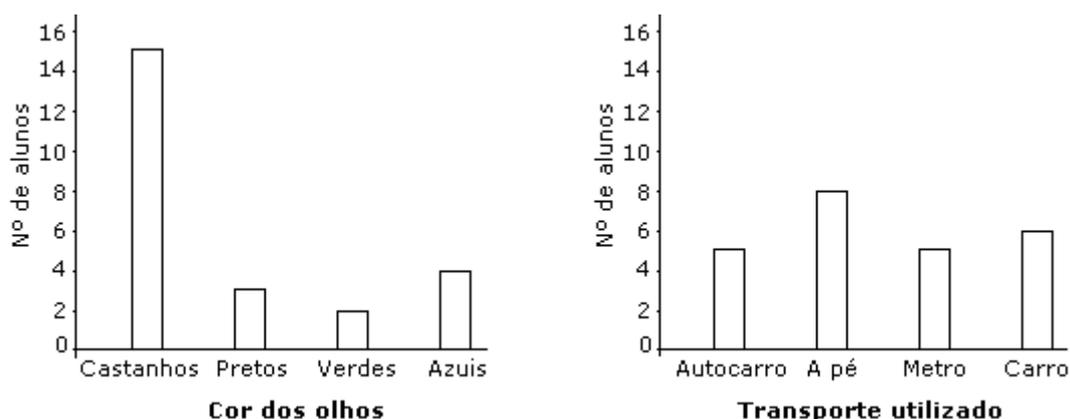
#### 4.3.3.2 Gráfico de barras

Uma forma de visualizar a informação de uma tabela de frequências é através do gráfico ou diagrama de barras. Para construir este gráfico, começa-se por desenhar um eixo horizontal (ou vertical), onde se assinalam (igualmente espaçadas) as diferentes categorias ou modalidades que a variável assume no conjunto dos dados. A ordem por que se colocam as categorias é arbitrária, a não ser que haja alguma ordem subjacente, como no caso dos dados qualitativos ordinais. Por cima de cada categoria (ou ao lado), desenha-se uma barra com altura proporcional ao número de casos observados nessa categoria. Desenha-se ainda um eixo vertical (horizontal), onde se marcam as frequências.

Ao contrário das alturas das barras, que dão uma mensagem muito precisa, a largura das barras não transmite qualquer informação. Deve, no entanto ter-se em atenção que, no mesmo gráfico, as barras devem ter todas a mesma largura, pois as barras mais largas podem chamar mais a atenção, induzindo em erro.

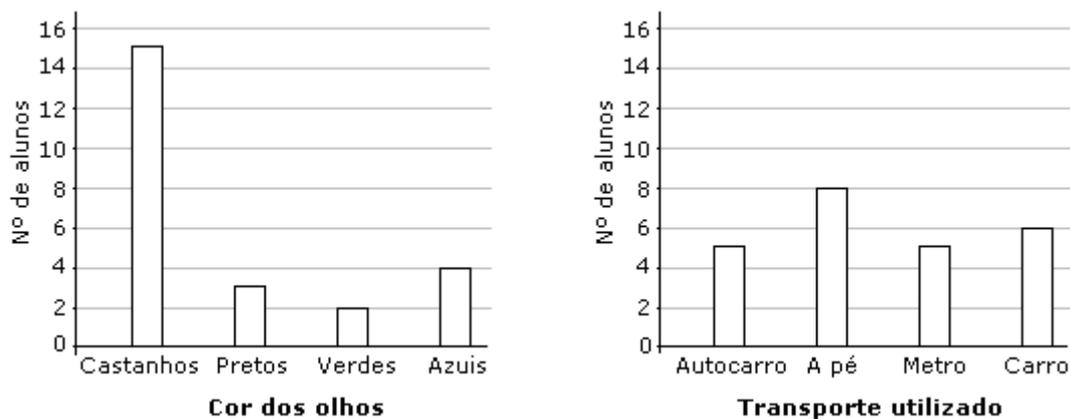
A observação que fizemos relativa às tabelas de frequência, quando se utilizam para comparar amostras que digam respeito à mesma variável, mas de dimensão diferente, tem aqui igual cabimento. Neste caso, as alturas das barras têm de ser iguais às frequências relativas das categorias para que a soma das alturas das barras em qualquer dos gráficos seja igual a 1, permitindo a comparação. Se não tivéssemos esta precaução e utilizássemos as frequências absolutas, a comparação entre os gráficos poderia induzir em erro.

Eis os gráficos de barras correspondentes às tabelas de frequência construídas na secção 4.3.2 para as variáveis *Cor dos olhos* e *Transporte utilizado*:



Ao contrário do gráfico de pontos, que não necessita de um eixo onde se marcam as frequências, no gráfico de barras ele faz parte integrante do gráfico e não pode ser omitido.

Por vezes, para facilitar a leitura das frequências associadas às diferentes categorias, desenham-se linhas paralelas ao eixo onde estão assinaladas as categorias:



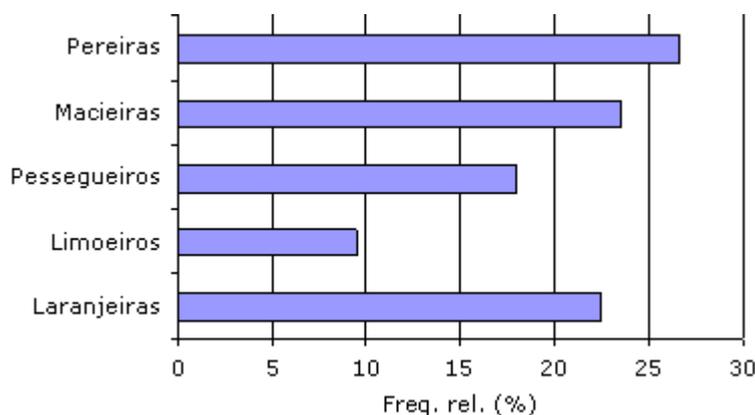
### Gráfico de barras horizontais

Os gráficos de barras horizontais utilizam o eixo vertical para marcar as classes e o eixo horizontal para marcar as frequências. São especialmente indicados no caso das variáveis quantitativas, em que, por vezes, o nome das classes é longo, sendo mais fácil colocá-los verticalmente do que ao lado uns dos outros.

**Exemplo – Seguro do agricultor<sup>1</sup>.** Com o objectivo de fazer um seguro, um agricultor teve de fazer o levantamento do número e tipo de árvores de fruto existentes no seu pomar. O resultado apresenta-se na tabela seguinte:

Classes	Freq. abs.	Freq. rel. (%)
Laranjeiras	320	22,4
Limoeiros	135	9,5
Pessegueiros	257	18,0
Macieiras	335	23,5
Pereiras	379	26,6
Total	1426	100,0

O gráfico de barras horizontais tem o seguinte aspecto:



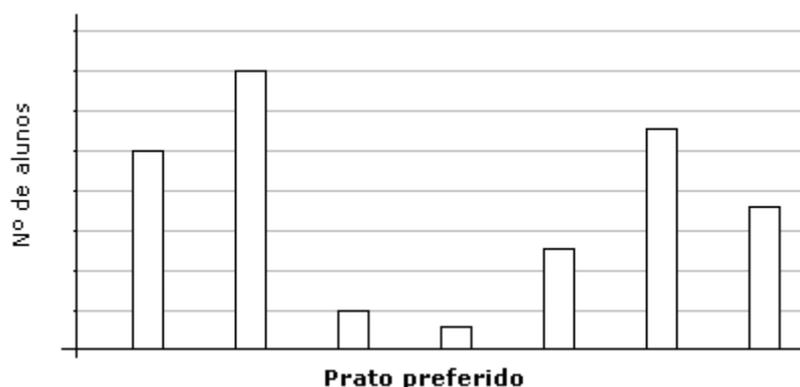
<sup>1</sup> Graça Martins et al (1999)

A principal vantagem dos gráficos, relativamente às tabelas, está na rapidez de leitura, pois permitem-nos ter uma percepção imediata de quais as categorias de maior e menor frequência, assim como a ordem de grandeza de cada categoria relativamente às restantes.

Para que um gráfico de barras transmita a informação que se pretende, sem ambiguidade, deve ter associado:

- o nome da variável que se está a estudar;
- os nomes das categorias que a variável assume, no eixo horizontal (ou vertical);
- uma escala no eixo vertical (ou horizontal). Nesta escala devem estar marcadas as frequências absolutas ou as frequências relativas das categorias que a variável assume no conjunto de dados considerados.

**Tarefa – Prato preferido** (Sugerida por uma actividade do CensusAtSchool). Na escola, o Director pretende averiguar os pratos preferidos dos alunos que comem na cantina, pelo que encarrega uma comissão de fazer um inquérito a alguns alunos. A metodologia utilizada para seleccionar estes alunos, foi a de interrogar os que se dirigiam à cantina, num dia escolhido ao acaso. A comissão encarregue do estudo apresentou ao Director um gráfico e um pequeno relatório com as conclusões:



Relatório: Os alunos interrogados apontaram 7 pratos distintos. Das respostas, pudemos tirar as seguintes conclusões:

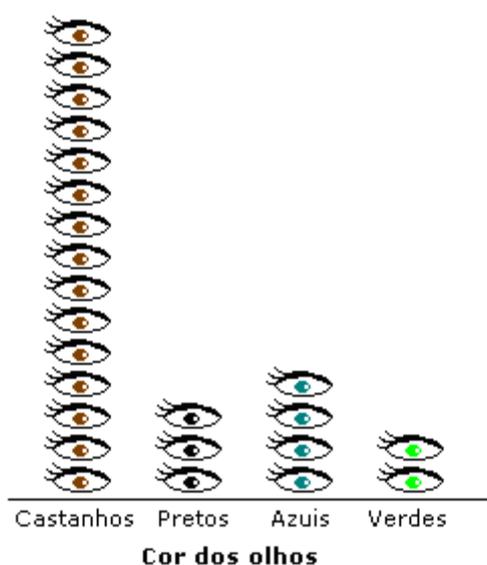
- a) O *Hambúrguer com batatas fritas* foi o prato mais votado
- b) O número de alunos que escolheu *Hambúrguer com batatas fritas*, foi o dobro dos que escolheram *Frango assado*
- c) Os *Filetes de peixe* receberam menos 4 votos do que o *Hambúrguer com batatas fritas*
- d) O *Esparquete à Bolonhesa* foi o segundo prato mais votado
- e) O *Bacalhau com natas* teve mais 4 votos do que o *Peixe assado*
- f) Houve quem votasse nas *Ervilhas com ovos*
- g) 5 alunos votaram no *Bacalhau com natas*

O Director recebeu este pequeno relatório e não ficou satisfeito, pois achou as conclusões muito confusas. Afinal, quantos alunos tinham votado? E quantos votaram em cada prato?

Podes ajudar a completar adequadamente o gráfico anterior? (Colocar as categorias e numerar a escala do eixo vertical).

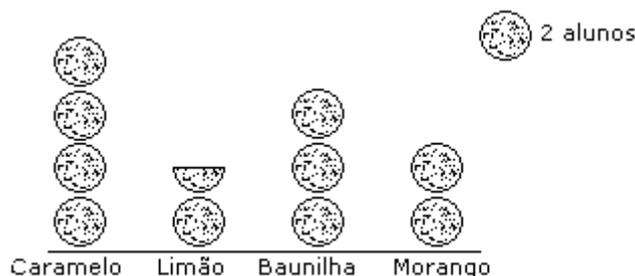
#### 4.3.4 Pictograma

Um pictograma é uma representação gráfica que usa símbolos alegóricos às variáveis que se estão a estudar. Por exemplo, se se estiver a estudar a variável cor dos olhos, é natural utilizar como símbolo um olho, enquanto que se o objecto do estudo for o sabor do gelado preferido, é natural utilizar como símbolo um gelado. A representação é idêntica ao gráfico de barras, com um eixo horizontal (ou vertical), mas onde se substitui a barra pelo número de símbolos correspondentes a cada categoria. Por exemplo, o pictograma correspondente à variável *Cor dos olhos* da turma de referência, pode ter o seguinte aspecto:



**Nota** – Por vezes uma figura representa mais do que um indivíduo. Nessa altura deve estar junto à representação gráfica o valor de cada figura.

**Tarefa – Bolachas preferidas.** No seguinte pictograma apresenta-se o resultado de um inquérito a uma turma, sobre qual o sabor preferido de um determinado tipo de bolachas:



Algumas questões:

- Quantos alunos responderam a esta questão?
- Quantos alunos disseram preferir sabor a Limão?

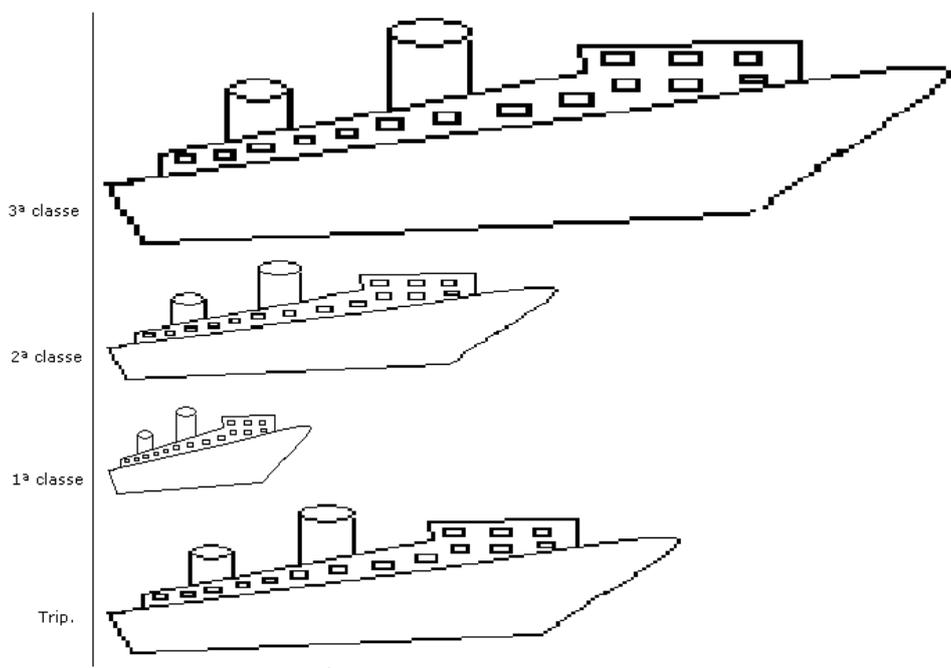
- c) Qual é o sabor mais popular?  
 d) Que nome se dá à categoria mais preferida?

A utilização de pictogramas exige vários cuidados, pois as figuras podem induzir em erro, como se ilustra nos exemplos seguintes.

**Exemplo – Os passageiros de um navio.** Considere um navio que transporta 525 pessoas, de acordo com a seguinte tabela de frequências, para a variável *Tipo de passageiro*:

Categorias	Freq. Abs.
Tripulação	141
1.ª classe	51
2.ª classe	115
3.ª classe	218
Total	525

A utilização do seguinte pictograma, para representar as frequências das categorias da variável em estudo, induz o leitor em erro:

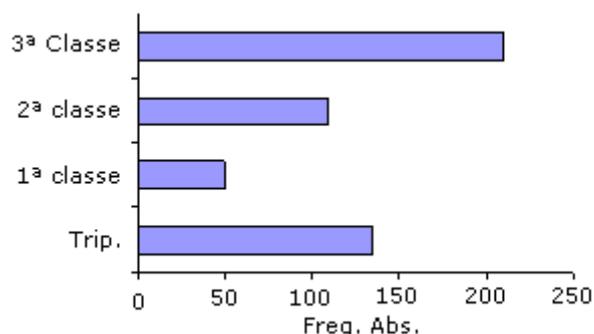


Há um princípio básico de uma boa representação gráfica, que neste caso foi quebrado – o princípio das áreas:

- *a área ocupada por parte de um gráfico, deve ser proporcional ao valor que essa parte representa.*

Ora, na figura anterior, a informação que se pretendia transmitir era a dada pelo comprimento do barco. Utilizaram-se figuras cujas áreas não são proporcionais aos valores das categorias, não dando uma informação correcta sobre as frequências correspondentes às diferentes categorias. Por exemplo, ao visualizar o gráfico anterior ficamos convencidos de que o número de passageiros viajando em 3.ª classe é mais do dobro dos que viajam em 2.ª, quando na verdade não chega ao dobro.

Uma representação gráfica correcta seria a seguinte, utilizando um gráfico de barras:



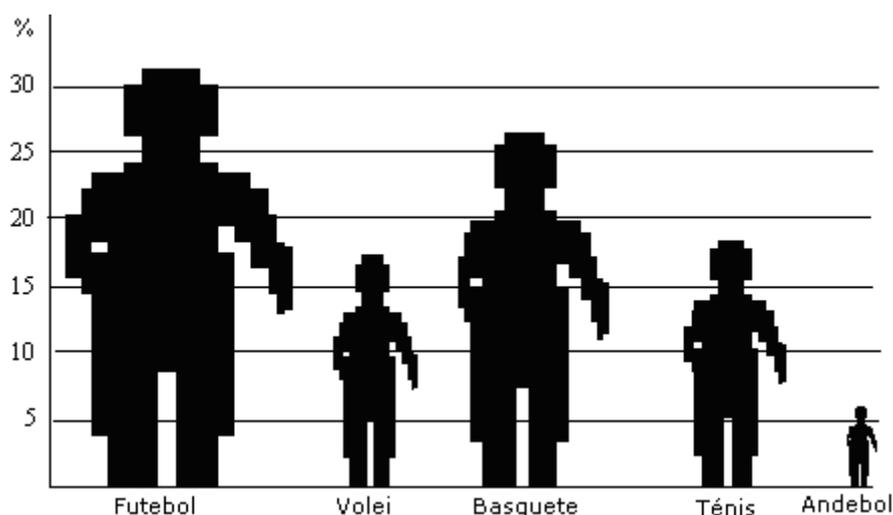
**Exemplo – Campo de jogos** (adaptado de Graça Martins et al. 1999). Numa escola o Director pretende construir um campo de jogos, pelo que gostaria de ter uma ideia de quais os jogos preferidos dos alunos. Encarregou um aluno de recolher a informação necessária, o qual utilizou a seguinte metodologia: elaborou uma lista de jogos possíveis e percorreu todas as turmas da escola, em número de 20, perguntando dentro de cada turma qual a opinião dos alunos cujo número fosse um múltiplo de 5. Em três turmas foram seleccionados 6 alunos e nas restantes 5. O resultado da recolha da informação tinha o seguinte aspecto

Futebol	x x
Volei	x x x x x x x x x x x x x x x x
Basquete	x x
Ténis	x x x x x x x x x x x x x x x x x x
Andebol	x x x x x x x

O aluno, ao perguntar a cada elemento da amostra a sua opinião, apontava o resultado com um x à frente da modalidade seleccionada. A forma como a informação foi recolhida permite imediatamente concluir que a modalidade preferida foi o futebol. A fim de transmitir verbalmente a informação ao Director, o aluno construiu a seguinte tabela de frequências

<i>Classes</i>	<i>Freq. abs.</i>	<i>Freq. rel (%)</i>
Futebol	32	31.1
Volei	18	17.5
Basquete	27	26.2
Ténis	19	18.4
Andebol	7	6.8
Total	103	100

Então o Director foi informado que as preferências dos alunos vão para o futebol seguindo-se o basquete. Depende agora das disponibilidades financeiras contemplar as diferentes modalidades, tendo em conta as preferências dos alunos. Procurando transmitir a informação graficamente, os alunos construíram o seguinte pictograma, onde se substituiu a barra por uma figura humana:



Na figura anterior a imagem correspondente à classe futebol é substancialmente maior que a que é utilizada para as outras modalidades ou classes. Daí dar uma ideia, errada, de que, por exemplo, a percentagem de alunos que preferem o futebol é várias vezes superior aos que preferem volei, quando nem sequer chega a ser o dobro. Este problema foi ocasionado pelo facto de se pretender que a figura humana ficasse proporcional, pelo que à medida que se aumentou a altura, também se aumentou a largura. O gráfico de barras correspondente tem o seguinte aspecto:

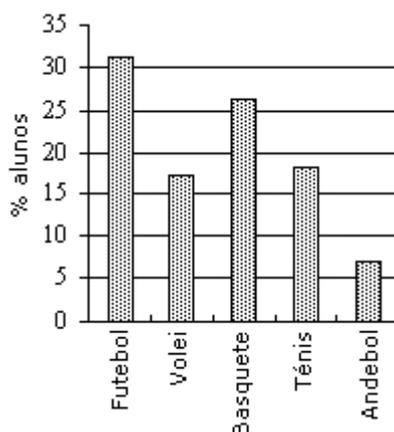
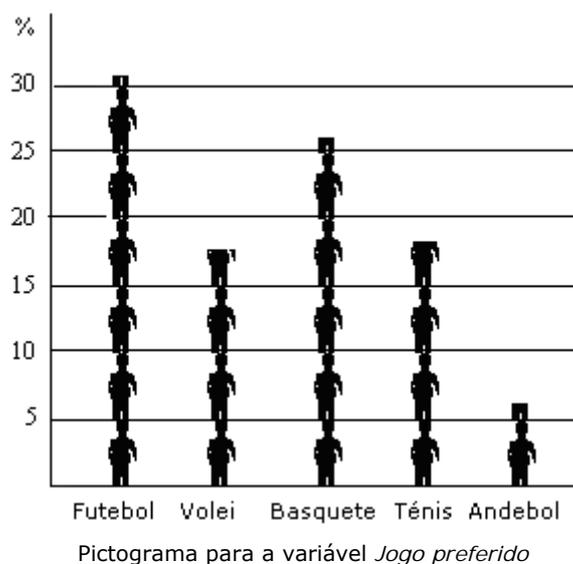
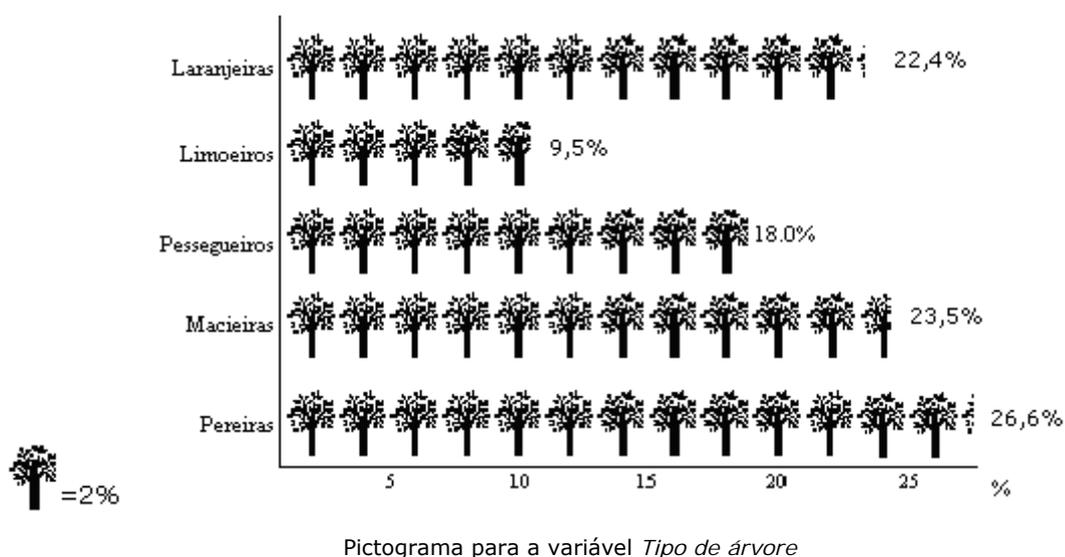


Gráfico de barras para a variável *Jogo preferido*

Na construção do gráfico de barras, como já dissemos nas indicações para a sua construção, deve ter-se em atenção que as barras devem ter a mesma largura, pois a mensagem que devem transmitir é a que está contida nas diferentes alturas das barras. Se umas barras forem mais largas do que outras, temos tendência a crer que as classes a que correspondem as barras mais largas têm maior frequência do que a que efectivamente têm. Este é um problema que não é tido em conta na construção de muitos *pictogramas*, em que as barras são substituídas por figuras, para tornar a representação gráfica mais atraente, como aconteceu no caso deste exemplo. Um pictograma possível, é o que se apresenta a seguir, em que a figura utilizada é uma figura humana, que corresponde a uma percentagem de 5%, que se replica o número de vezes que for necessário, sendo possível utilizar uma fracção da figura:

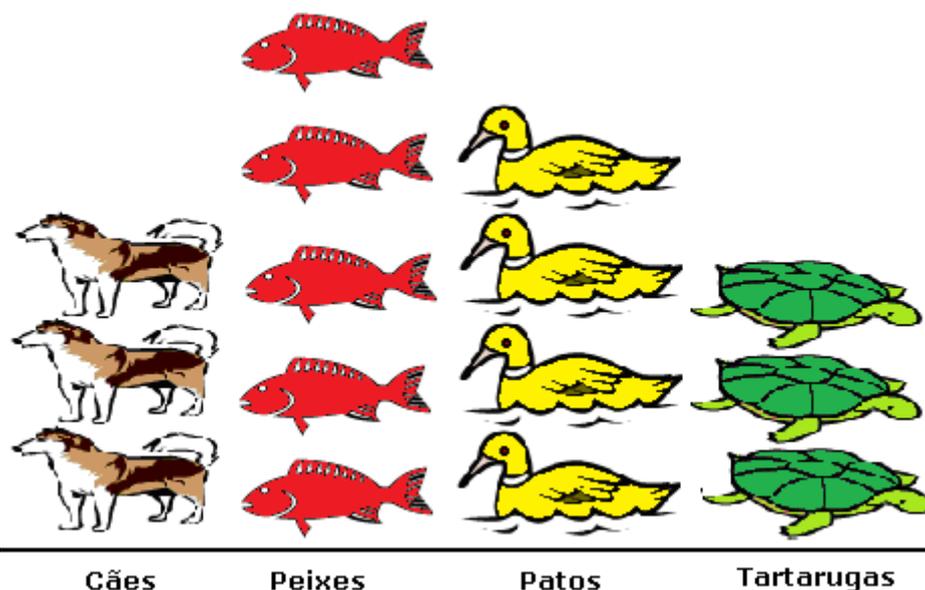


**Exemplo – Seguro do agricultor** (cont.). Uma representação gráfica possível para este exemplo, apresentado na secção 4.3.3.2 seria a seguinte, em que se considera uma figura sugestiva, mas sem incorrer no erro da representação inicial do exemplo anterior.



Embora seja comum dizer que uma imagem vale mais do que mil palavras, não podemos deixar de chamar a atenção para que esta frase tem sentido se a informação transmitida pela imagem for correcta, o que, como vimos, nem sempre acontece.

**Tarefa – Os animais do jardim** – No jardim da escola, que tem um lago muito bonito, o professor decidiu ir com os alunos verificar que tipo de animais é que havia no jardim. Verificaram que havia animais de 4 tipos: cães, peixes, patos e tartarugas, de acordo com o seguinte pictograma



Algumas questões que podem ser colocadas, são as seguintes:

- Quantos patos há no jardim?
- Quantos peixes há no jardim?
- Quantos animais vivem no jardim?
- Há alguns animais que existam na mesma quantidade? Se existirem, quais são?
- Quantos peixes há a mais do que patos?
- Ofereceram 2 tartarugas para o jardim da escola. Quantas tartarugas existem agora?

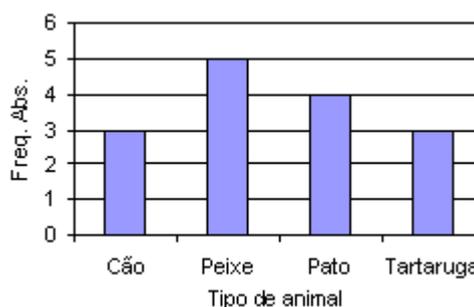
No problema anterior, o que é o *dado*? *Dado* é o resultado da observação do tipo de animal, pelo que o conjunto de *dados* observados foi:

*Cão, Cão, Cão, Peixe, Peixe, Peixe, Peixe, Peixe, Pato, Pato, Pato, Pato, Tartaruga, Tartaruga, Tartaruga*

Cada figura do pictograma representa cada dado, de uma forma sugestiva.

A partir do pictograma facilmente se constrói a tabela de frequências absolutas e o gráfico de barras associado:

Tipo de animal	Frequência absoluta
Cão	3
Peixe	5
Pato	4
Tartaruga	3
Total	15

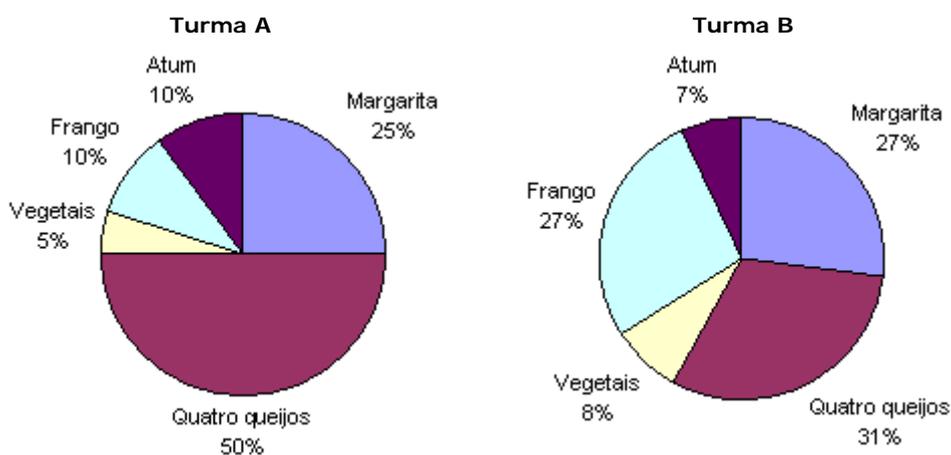


### 4.3.5 Gráfico circular

Uma representação gráfica muito utilizada, nomeadamente na comunicação social, é o **gráfico circular**. A base desta representação é um círculo que representa a forma como o total de um conjunto de dados se distribui pelas categorias. O círculo é dividido em sectores circulares, tantos quantas as categorias da variável em estudo, e o ângulo de cada sector é proporcional à frequência da categoria que representa. Assim, cada sector representa uma fracção do total de dados.

Os alunos devem começar por aprender a ler a informação transmitida por estas representações gráficas e só posteriormente procederem à sua construção. Para os alunos mais novos, a construção do gráfico circular deve ser feita através de dobragens do círculo em 2, 4 ou 8 partes, pelo que para estes alunos só se recomenda a sua utilização para representar frequências relativas próximas de  $\frac{1}{2}$ ,  $\frac{1}{4}$ ,  $\frac{1}{8}$ , ou para interpretar frequências relativas relacionadas com estas.

**Exemplo – A piza preferida nas turmas A e B.** Na turma, o professor apresentou os resultados de um inquérito, sobre qual a piza preferida nas turmas A e B, utilizando os seguintes gráficos circulares:



Como se verifica, num gráfico circular utilizam-se percentagens, já que é a forma indicada para representar a fracção de cada categoria como parte do todo, em que este "todo" é representado pelo círculo e equivale a 100%. Tivemos o cuidado de colocar estas percentagens, assim como os nomes das categorias, ao lado das "fatias" respectivas do círculo, para uma melhor leitura do gráfico.

Destas representações gráficas, imediatamente se conclui que:

- A maior parte dos alunos, tanto da turma A (50%), como da turma B (31%), preferem a piza Quatro queijos;
- A piza menos preferida, é a de Vegetais, tanto para os alunos da turma A, como da turma B;
- Metade dos alunos da turma A prefere a piza Quatro queijos e metade dos restantes, prefere a piza Margarita;
- Na turma A há igual número de alunos a preferirem a piza de Frango e a piza de Atum, enquanto que na turma B existe igual número de alunos a preferirem a piza Margarita e a de Frango.

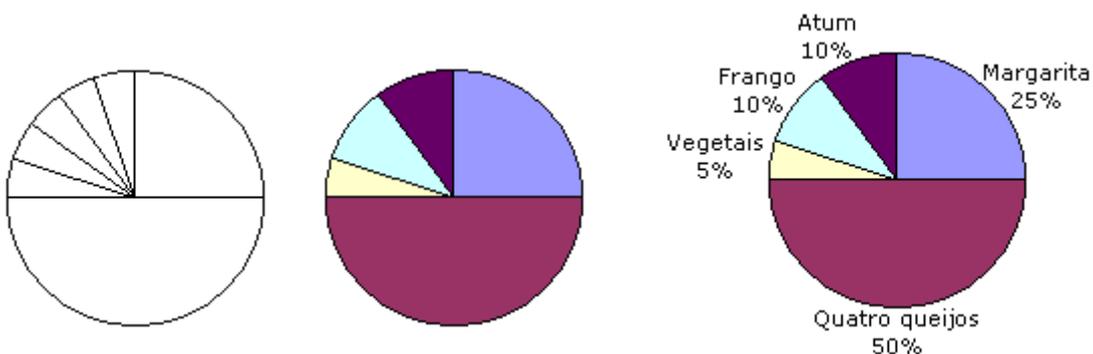
Exemplo de algumas questões suplementares, envolvendo um desafio mais forte, são as seguintes:

- Podemos afirmar que o número de alunos da turma A que prefere piza Quatro queijos, é superior ao número de alunos da turma B a preferir o mesmo tipo de piza?
- Sabe-se que na turma A, há 10 alunos a preferirem piza Quatro queijos. Quantos alunos tem a turma?
- Nas condições da alínea anterior, quantos alunos da turma A preferem piza de Vegetais?
- Se se duplicasse o número de alunos da turma A (situação pouco razoável, devido ao elevado número de alunos...) a preferirem cada tipo de piza, o que acontecia ao gráfico circular?

Admitindo agora que tínhamos a tabela de frequências correspondente à piza preferida da turma A, vejamos como proceder à construção do gráfico circular respectivo:

<b>Piza preferida</b>	Frequência Absoluta	Frequência Relativa (%)
Margarita	5	25
Quatro queijos	10	50
Vegetais	1	5
Frango	2	10
Atum	2	10
Total	20	100

- Como 50% dos alunos preferem a piza Quatro queijos, então metade do círculo corresponde a esta categoria;
- Como 25% dos alunos prefere piza Margarita, um quarto do círculo, corresponde à categoria Margarita;
- O quarto do círculo restante deve ser dividido em 5 sectores aproximadamente iguais, considerando-se uma das partes para a categoria Vegetais e duas partes para a categoria Frango e outras duas para a categoria Atum.
- Finalmente pintam-se os sectores e colocam-se as etiquetas e as percentagens correspondentes



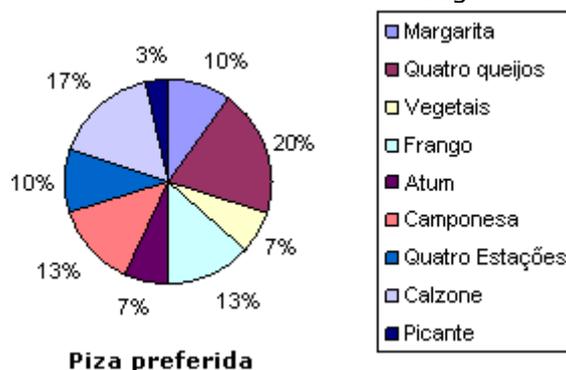
Nem sempre a construção do gráfico circular é tão simples como no caso anterior, em que as frequências relativas eram relativamente fáceis de marcar. Por exemplo, no caso da turma B, é mesmo necessário dividir a amplitude do ângulo de  $360^\circ$  em amplitudes proporcionais às frequências relativas das categorias para construir os sectores circulares. Estas amplitudes que se obtêm multiplicando  $360^\circ$  pelas fre-

quências relativas das categorias são acrescentadas à tabela de frequência e para desenhar os sectores circulares é necessário utilizar um transferidor:

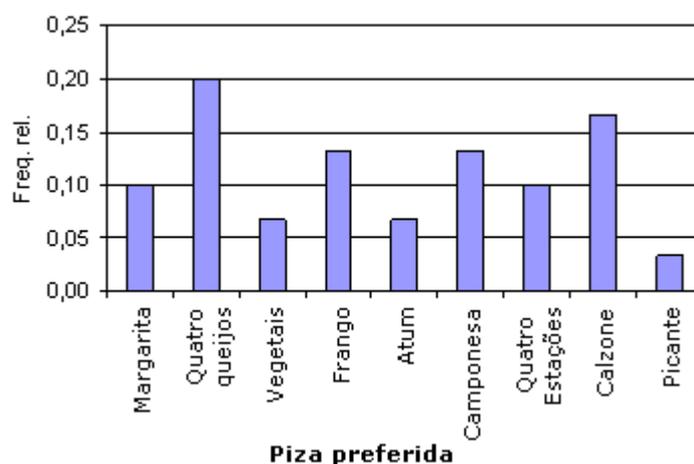
<b>Piza preferida</b>	Frequência Absoluta	Frequência Relativa (%)	Amplitude do ângulo
Margarita	7	27	97°
Quatro queijos	8	31	111°
Vegetais	2	8	27°
Frango	7	27	98°
Atum	2	7	27°
Total	26	100	360°

**Nota** – Em Graça Martins et al. (2007, p. 90) ensina-se a construir um gráfico circular a partir de papel quadriculado e cartolina.

A utilização dos gráficos circulares merece alguns cuidados, nomeadamente quando o número de categorias que a variável assume for demasiado grande, tornando confusa a informação que procura transmitir. Por exemplo, admitamos que as preferências dos alunos de uma outra turma foram as seguintes:



O gráfico está bem construído, com a legenda e as percentagens associadas às categorias indicadas, mas a mesma informação seria mais facilmente apreendida através de um gráfico de barras, como se apresenta a seguir, em que se torna mais fácil de visualizar as diferenças entre as frequências das diferentes categorias:

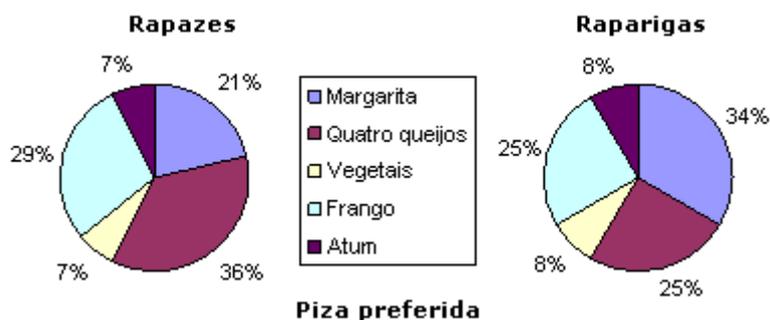


### **Nem sempre a utilização de gráficos circulares é a mais conveniente**

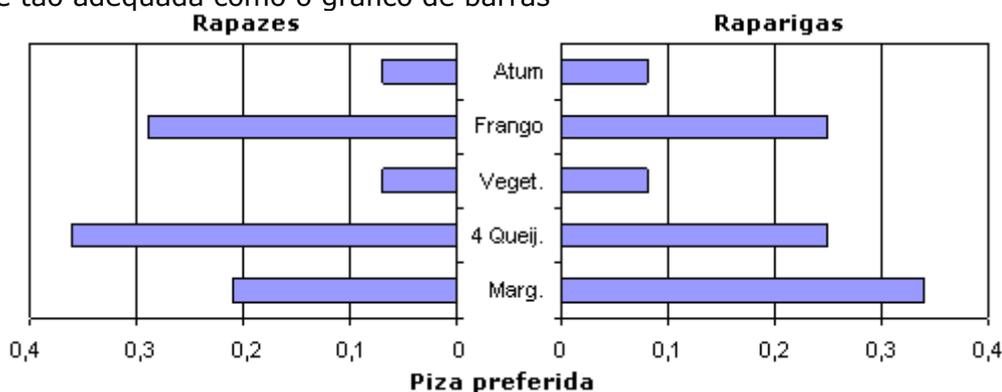
Embora no exemplo anterior tenhamos utilizado dois gráficos circulares para comparar os gostos de duas turmas, por vezes é preferível a utilização de gráficos de

barras dispostos de forma adequada. Esta observação é especialmente adequada para o caso em que nas distribuições em análise, algumas categorias tenham valores próximos, o que faz com que as áreas dos sectores circulares sejam difíceis de comparar.

Por exemplo, para distinguir os gostos dos rapazes dos gostos das raparigas da turma B, a utilização dos gráficos circulares



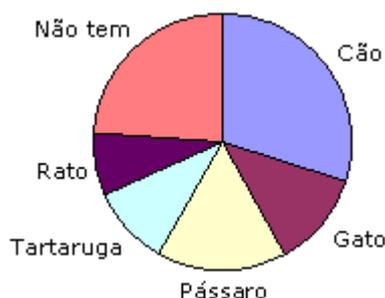
não é tão adequada como o gráfico de barras



pois nesta representação o comprimento das barras torna mais fácil comparar as frequências correspondentes às mesmas categorias.

Da representação gráfica anterior concluímos que a moda nos rapazes é a piza Quatro queijos, enquanto que nas raparigas é a Piza Margarita.

**Tarefa – Animal doméstico preferido.** O gráfico circular seguinte mostra o resultado de uma sondagem a 50 alunos de um escola, sobre qual o animal doméstico preferido:



Estima o número de alunos que:

- a) Têm cão
- b) Têm gato ou pássaro
- c) Não têm animal doméstico

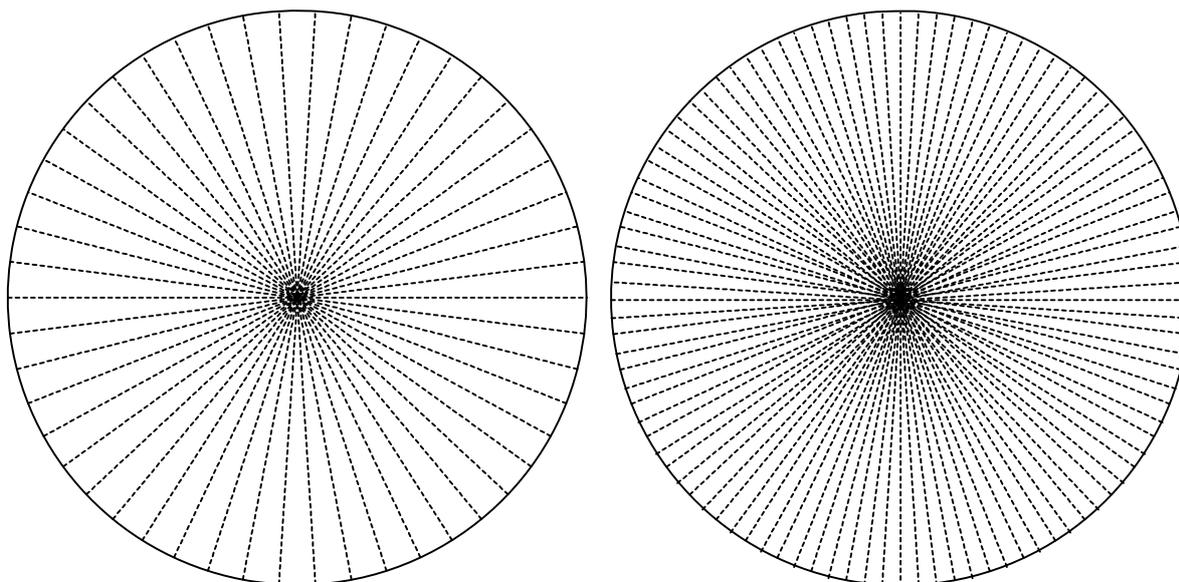
**Tarefa – Meio de transporte utilizado.** Aos mesmos alunos da tarefa anterior perguntou-se qual o meio de transporte que utilizavam para ir para a escola. Os resultados obtidos estão na seguinte tabela:

Transporte utilizado	N.º de alunos
Carro	8
Transportes públicos	21
A pé	18
Autocarro da Câmara	2
Outro	1

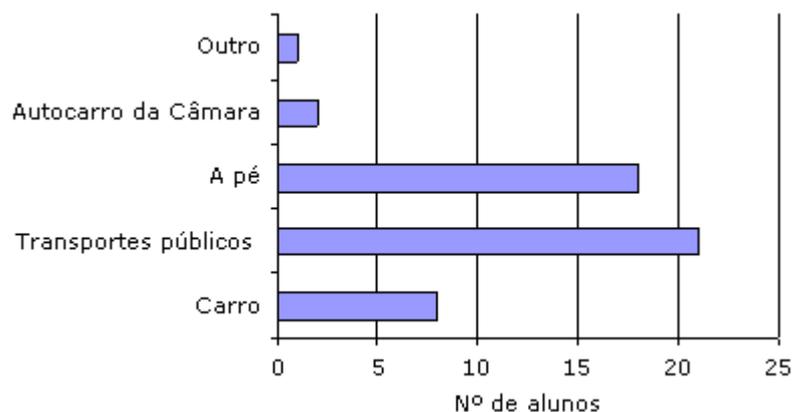
Completa a tabela seguinte com as frequências relativas, com 2 casas decimais e em percentagem:

Transporte utilizado	N.º de alunos	Fracção do todo (2 casas decimais)	Fracção do todo (percentagem)
Carro	8		
Transportes públicos	21		
A pé	18		
Autocarro da Câmara	2		
Outro	1		
Total	50		

Na figura seguinte apresentam-se 2 círculos, em que no primeiro estão marcadas 50 divisões iguais e no segundo 100 divisões iguais:



- Cada um dos círculos anteriores pode servir para construir gráficos circulares: num deles é mais fácil utilizar as frequências absolutas e no outro as frequências relativas (em percentagem). Explica porquê.
- Constrói os gráficos circulares utilizando quer as frequências absolutas, quer as frequências relativas. Compara as representações obtidas e descreve o que concluíste.
- A partir da tabela de frequências inicialmente dada, construiu-se o seguinte gráfico de barras para os mesmos dados:



- i) Qual das representações gráficas preferes? O gráfico circular ou o gráfico de barras?
- ii) Qual o tipo de informação que é realçada pelo gráfico circular?
- iii) Qual o tipo de informação que é realçada pelo gráfico de barras?

**Nota** – No texto anterior alertámos para o facto de ser necessário algum cuidado na utilização do gráfico circular, nomeadamente quando a distribuição a representar, apresenta muitas categorias (ou classes) ou quando os valores das frequências de algumas das categorias estão próximos. No entanto é uma representação por excelência, quando o que se procura realçar é a forma como os dados se distribuem pelas categorias, já que representa a fracção de cada categoria como parte do todo, em que este “todo” é representado pelo círculo e equivale a 100%.

Assim, a escolha da representação gráfica adequada para representar um conjunto de dados pode depender do que é que se procura realçar na distribuição desses dados.

#### 4.3.6 Nem sempre um gráfico com barras é um gráfico de barras...

É comum utilizarem-se gráficos com barras para representar os próprios dados e não as frequências com que as diferentes classes ou categorias surgem no conjunto de dados que se está a estudar.

Por exemplo, admitamos que se estava interessado em saber qual o número de alunos de cada uma das turmas do 7º. ano, de determinada escola. Depois de feita a contagem, chegou-se aos seguintes resultados:

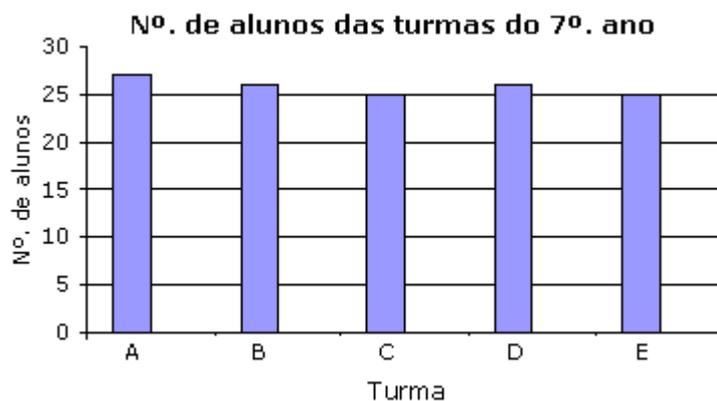
Turma	Nº. de alunos
A	27
B	26
C	25
D	26
E	25

A tabela anterior não é uma tabela de frequências, mas simplesmente uma tabela que apresenta os dados. Neste caso, a unidade observacional, isto é o objecto do nosso estudo, sobre o qual pretendemos recolher informação, é a *turma*, porque o nosso objectivo era saber quantos alunos tinha cada turma do 7º. ano. O dado é o

resultado da nossa observação! Assim, os nossos dados são o número de alunos das turmas A, B, C, D e E ou seja

27, 26, 25, 26, 25

Uma forma possível de representar a informação anterior, é utilizando um gráfico com barras, como o que se apresenta a seguir:



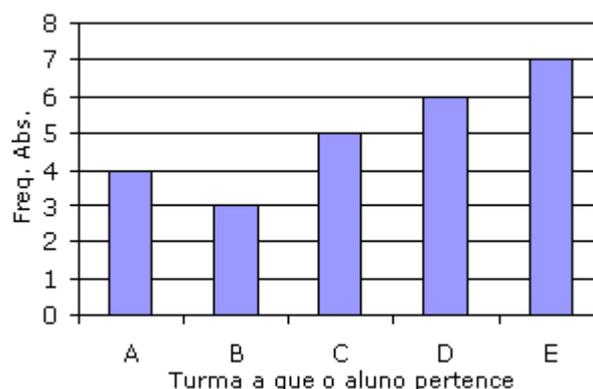
O gráfico anterior, embora seja um gráfico com barras, não é o que se chama, em Estatística, um *gráfico de barras*, pois é um gráfico onde estão representados os dados e não as frequências absolutas ou relativas de um conjunto de dados.

Admitamos agora que considerávamos um conjunto de 25 alunos do 7º. e decidíamos investigar a que turma pertencia cada aluno. Agora, a unidade observacional é o *aluno* e a característica que estamos a estudar é a turma a que pertence, pelo que os nossos dados serão do tipo

*A, C, B, A, E, C, ..., A*

Após resumir a informação contida no conjunto de dados anteriores, através de uma tabela de frequências, poderíamos construir o gráfico de barras associado. Um resultado possível poderia ser:

Turma	Freq. Absoluta
A	4
B	3
C	5
D	6
E	7
Total	25

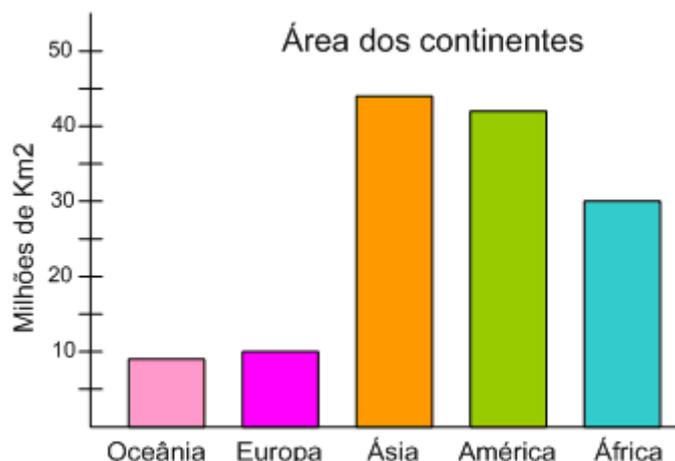


Este gráfico, embora idêntico ao apresentado no início desta secção, é um gráfico de barras onde estão representadas as **frequências** absolutas das categorias assumidas pela variável em estudo – *Turma a que o aluno pertence*, no conjunto dos dados. No outro gráfico estão representados os **dados** obtidos ao observar a variável – *Número de alunos por turma*.

### Confusão entre dados e frequência!

A situação em que se confundem **dados** com **frequências** é mais comum do que se poderia pensar. Mesmo em obras científicas aparecem situações em que se faz essa confusão. Reproduzimos a seguir um exemplo, de entre os vários que encontramos, em livros de texto de Matemática para o Ensino Básico:

"O gráfico representa as áreas dos continentes



1. Qual o continente que tem menor área?

2. Qual a área da Europa em ha?

3. Qual o continente com maior área?

4. Constrói uma tabela de frequências"

A resposta apresentada para a questão 4 é a seguinte:

Continente	Área milhões km <sup>2</sup>
Oceânia	9
Europa	10
Ásia	44
América	42
África	30

No gráfico anterior estão representadas as áreas dos cinco continentes e essas áreas são os dados resultantes da observação da variável *Área de cada continente*. Assim, não tem qualquer sentido a questão 4, onde se pede para construir uma tabela de frequências, pois nem o gráfico apresenta as frequências, nem a tabela é uma tabela de frequências. No entanto, tanto o gráfico como a tabela apresentam correctamente os dados observados.

Outro exemplo também encontrado num texto de Matemática é o seguinte:

"No jardim zoológico contou-se o número de visitantes que durante uma semana assistiram ao espectáculo dos golfinhos.

Os resultados obtidos foram:

Segunda – 0; Terça – 1000; Quarta – 1500; 5ª feira – 1250;  
Sexta – 1500; Sábado – 3000; Domingo – 2500

1. Elabora uma tabela de frequências absolutas
2. Constrói um gráfico de barras correspondente, considerando as frequências absolutas de 500 em 500
3. Indica os dois dias com maior número de visitantes. Porque será?"

A resolução apresentada é a seguinte:

Dia da semana	Nº. de visitantes
2ª	0
3ª	1000
4ª	1500
5ª	1250
6ª	1500
Sáb.	3000
Dom.	2500



Mais uma vez, nem a tabela anterior é uma tabela de frequências, nem o gráfico é um gráfico de barras.

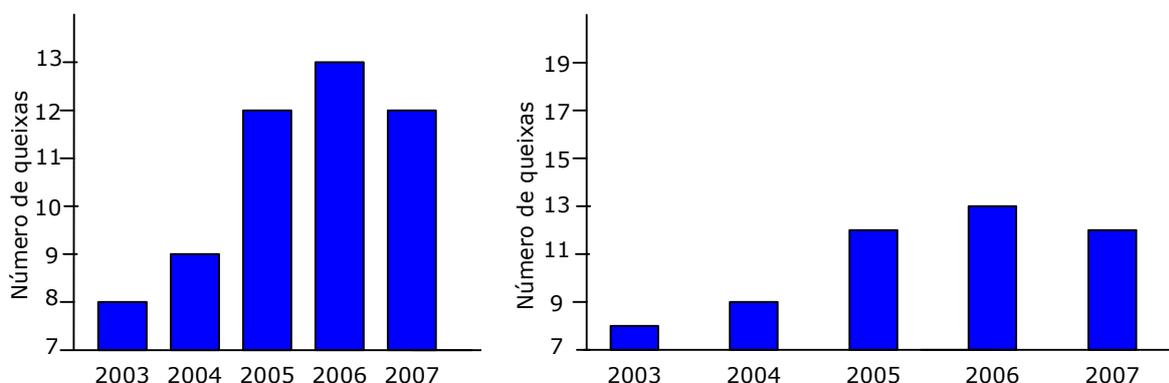
### Atenção às escalas!

A principal vantagem dos gráficos, relativamente às tabelas, está na rapidez de leitura, pois permitem-nos ter uma percepção imediata de quais as categorias de maior e menor frequência, assim como a ordem de grandeza de cada categoria relativamente às restantes.

Para que um gráfico com barras, quer represente os dados ou as frequências (gráfico de barras) transmita a informação que se pretende sem ambiguidade, deve ter uma escala onde devem estar marcados o valor dos dados ou das frequências (absolutas ou relativas).

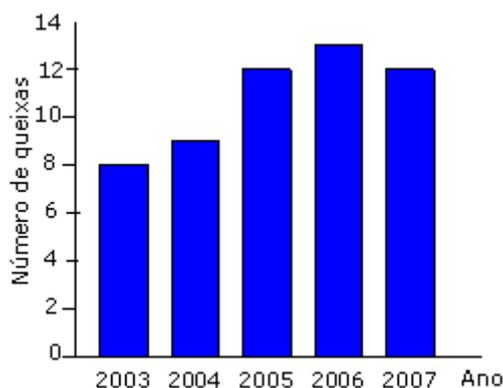
A manipulação das escalas, sobretudo a do eixo onde estão marcadas as frequências, pode ser usada com o intuito de transmitir informação incorrecta, como se verifica nos dois exemplos seguintes.

**Exemplo – Número de queixas recebidas num hospital, por negligência médica.** Suponha que num determinado hospital o número de queixas, no período de 2003 a 2007, foi o seguinte: 8, 9, 12, 13 e 12. Foram apresentadas as seguintes representações gráficas para transmitirem a informação anterior:



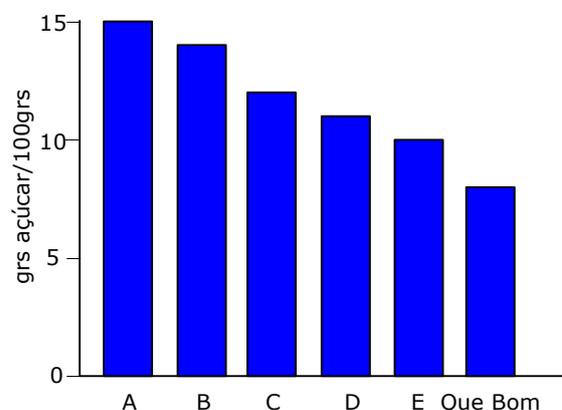
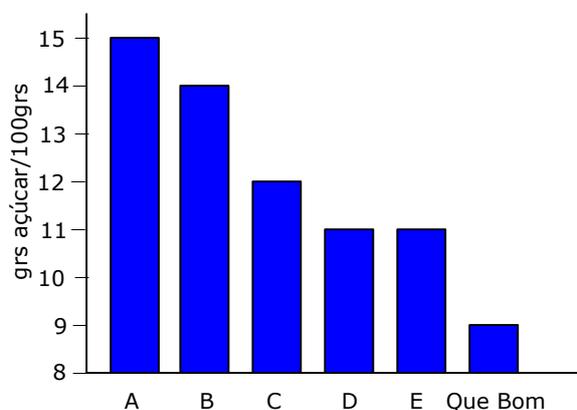
A representação gráfica da esquerda procura realçar o facto do número de queixas ter aumentado substancialmente, enquanto que a do lado direito procura desvalorizar

zar esse aumento. Em nenhum dos gráficos a escala se inicia no ponto 0, o que é um erro. Por outro lado, no gráfico do lado direito ainda se diminuiu a distância entre os incrementos do eixo vertical, ao mesmo tempo que se aumentou a distância entre as categorias no eixo horizontal. Uma representação correcta pode ser a seguinte:



Mais à frente veremos outra representação gráfica, o gráfico de linha, mais sugestivo e apropriado para representar este tipo de informação, em que se procura representar a evolução de uma variável, com o tempo.

**Exemplo – Quantidade de açúcar nos cereais para crianças.** Uma empresa que vende cereais para crianças faz publicidade aos seus cereais da marca “Que Bom”, alegando que têm menos açúcar do que os da concorrência. Para fundamentar a sua alegação apresenta o gráfico do lado esquerdo da figura seguinte, onde compara os 9grs de açúcar, por 100grs do cereal “Que Bom”, com os 15, 14, 12, 11 e 11 gramas, de açúcar, por 100grs, respectivamente dos cereais A, B, C, D e E:



## 4.4 Tabelas e gráficos para dados quantitativos discretos

Como já vimos na secção 3.3, as variáveis quantitativas de contagem, isto é, que se referem a características que só se podem contar e não se podem medir, designam-se também por variáveis quantitativas **discretas**. O resultado da observação destas variáveis são os dados quantitativos discretos. Estas variáveis só podem assumir um número finito ou infinito numerável de valores distintos.<sup>2</sup>

### 4.4.1 Tabela de frequências para dados quantitativos discretos

A construção da tabela de frequências para dados quantitativos discretos é idêntica à construída para dados qualitativos, considerando-se agora para **classes** os valores distintos que surgem no conjunto de dados.

Os dados discretos são organizados na forma de uma **tabela de frequências**, com três ou mais colunas. Na primeira coluna, *coluna das classes*, indicam-se todos os valores distintos,  $x_i^*$ , presentes na amostra a analisar; na coluna seguinte, *coluna das frequências absolutas*  $n_i$ , regista-se o número de vezes que cada valor  $x_i^*$ , surge na amostra. Numa terceira coluna, *coluna das frequências relativas* (ou percentagens)  $f_i$ , regista-se, para cada classe  $x_i^*$ , o valor que se obtém dividindo a frequência absoluta pela dimensão da amostra (número de elementos).

Pode ainda incluir-se na tabela de frequências mais duas colunas, a coluna das frequências absolutas acumuladas e a coluna das frequências relativas acumuladas, onde, para cada classe, se coloca a soma das frequências absolutas ou relativas, respectivamente.

As colunas das frequências acumuladas, nomeadamente a das frequências relativas é bastante útil no cálculo da mediana e dos quartis, medidas de localização de alguns pontos importantes da distribuição dos dados, como veremos mais à frente.

No exemplo da turma de referência, a variável *Número de irmãos* é de natureza discreta e a tabela de frequências construída a partir dos valores observados para os alunos da turma tem o seguinte aspecto:

N.º de irmãos $x_i^*$	Freq. Abs. $n_i$	Freq. Rel. $f_i$	Freq. Abs. Acum.	Freq. Rel. Acum.
0	6	0,250	6	0,250
1	9	0,375	15	0,625
2	5	0,208	20	0,833
3	3	0,125	23	0,958
4	1	0,042	24	1,000
Total	24	1,000		

Da tabela anterior podemos retirar algumas conclusões relativas a esta turma:

- Na turma predominam os alunos com um único irmão, pelo que o valor de 1 para a variável Número de irmãos, é a moda;
- 25% dos alunos não têm nenhum irmão;
- Não há alunos com mais de 4 irmãos;
- Mais de 95% dos alunos têm 3 ou menos irmãos.

<sup>2</sup> Recordemos que num conjunto infinito numerável pode estabelecer-se uma correspondência entre os seus elementos e o conjunto dos números naturais

No caso das variáveis qualitativas, ao construirmos as tabelas de frequência, não considerámos as frequências acumuladas. Efectivamente, só para alguns casos especiais das variáveis qualitativas, as variáveis **ordinais** – em que se pode estabelecer uma ordenação ou hierarquia entre as classes, é que tem sentido calcular as frequências acumuladas. Por exemplo, se os alunos da Turma de referência fizessem um teste de Língua Portuguesa e fossem classificados com as classificações de *Não satisfaz*, *Satisfaz pouco*, *Satisfaz*, *Satisfaz muito* e *Satisfaz plenamente*, uma tabela de frequências possível, seria a seguinte:

Classificação	Freq. Abs. $n_i$	Freq. Rel. $f_i$	Freq. Abs. Acum.	Freq. Rel. Acum.
Não satisfaz	1	0,042	1	0,042
Satisfaz pouco	4	0,167	5	0,208
Satisfaz	11	0,458	16	0,667
Satisfaz muito	6	0,250	22	0,917
Satisfaz plenamente	2	0,083	24	1,000
Total	24	1,000		

Tendo em consideração a tabela anterior, poderíamos concluir que:

- Só 5 alunos, ou seja, cerca de 21% dos alunos é que tiveram classificação negativa;
- Cerca de 79% (100%-21%) dos alunos tiveram nota positiva.

Embora a variável em estudo seja de tipo qualitativo, tem a particularidade de ser **ordinal**, pois pode-se estabelecer uma hierarquia entre as classes ou categorias que assume: *Não satisfaz* é menor que *Satisfaz pouco*, etc. No entanto, para os mesmos alunos, se estivermos a estudar a variável *Cor dos olhos*, para a qual construímos a tabela de frequências

Categorias	Freq. abs.	Freq. rel.
Castanhos	15	0,625
Pretos	3	0,125
Verdes	2	0,083
Azuis	4	0,167
Total		1

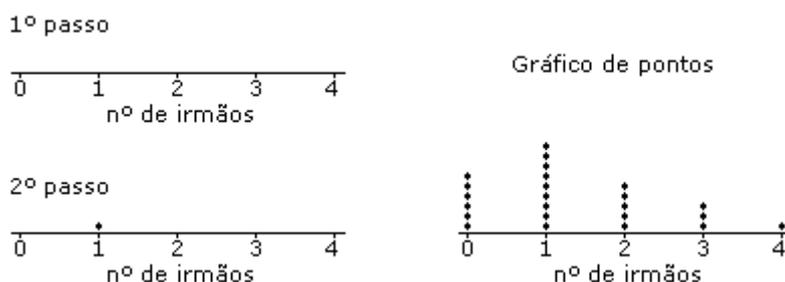
já pode dar origem a interpretações erradas juntarmos, à tabela, as colunas das frequências acumuladas. Não se pode dizer que 75% (62,5%+12,5%) dos alunos têm cor dos olhos menor ou igual a Pretos. Poderíamos eventualmente interpretar esse valor dizendo que 75% dos alunos têm olhos Castanhos ou Pretos, mas não é este o objectivo das frequências acumuladas.

#### 4.4.2 Gráfico de pontos e gráfico de barras para dados quantitativos discretos

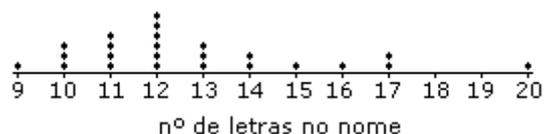
##### 4.4.2.1 Gráfico de pontos

Tal como para os dados qualitativos, a representação gráfica mais simples que se pode obter e que não necessita de nenhuma organização prévia dos dados, é o gráfico ou diagrama de pontos. É uma representação que se pode ir construindo à medida que se recolhem os dados. Começa-se por desenhar um eixo horizontal (ou vertical), onde se assinalam todos os valores que a variável assume no conjunto dos dados. Por cima de cada valor (ou ao lado), marca-se um ponto sempre que ao

recolher um dado ou ao percorrer o conjunto dos dados se encontrar um valor igual. Por exemplo, para os dados da turma de referência, podemos distinguir os seguintes passos na construção do gráfico de pontos para a variável *Número de irmãos*:



Se entre o mínimo e o máximo da amostra, houver alguns valores que não existam no conjunto dos dados a analisar, esses valores devem também ser assinalados no eixo, embora não se lhes associem quaisquer pontos. Por exemplo, o gráfico de pontos correspondente à variável *Número de letras no nome*, da turma de referência, tem o seguinte aspecto:



Da representação anterior, imediatamente se conclui que, na turma:

- Predominam os nomes com 12 letras;
- Não há nomes com 18 ou 19 letras;
- O nome maior tem 20 letras.

Para a construção do gráfico de pontos recomenda-se a utilização do papel quadriculado.

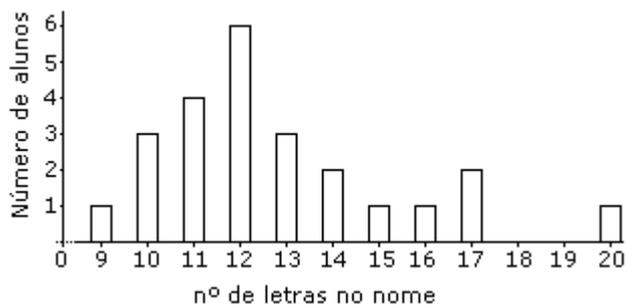
O gráfico de pontos dá uma informação muito semelhante à que é transmitida pelo gráfico de barras.

#### 4.4.2.2 Gráfico de barras

Dado um conjunto de dados de tipo quantitativo discreto, para o qual se construiu uma tabela de frequências, a representação gráfica mais utilizada é o gráfico de barras. Para este tipo de dados, a construção do gráfico de barras é semelhante à que fizemos para os dados de tipo qualitativo. Começa-se por desenhar um eixo (normalmente horizontal) e nesse eixo marcam-se os valores  $x_i^*$ , que constituem as classes. Nesses pontos marcam-se barras de altura igual à respectiva frequência absoluta ou relativa. Fazemos aqui uma observação idêntica à que já fizemos no caso da construção do gráfico de pontos: deve marcar-se no eixo a sequência completa dos valores entre o mínimo e o máximo observados, mesmo que alguns desses valores não constem da amostra.

Por exemplo, para o caso da variável *Número de letras no nome*, da turma de referência, tem-se:

N.º de letras no nome $x_i^*$	Freq. Abs. $n_i$	Freq. Rel. $f_i$
9	1	0,042
10	3	0,125
11	4	0,167
12	6	0,250
13	3	0,125
14	2	0,083
15	2	0,042
16	1	0,042
17	2	0,083
20	1	0,042
Total	24	1,000



As conclusões que tiramos a partir da representação gráfica anterior, são idênticas às obtidas a partir do gráfico de pontos.

**Tarefa – Número de irmãos.** O professor propôs na sua turma A averiguar o número de irmãos dos alunos (da turma) e posteriormente comparar com o número de irmãos dos alunos da turma de referência.

Para proceder à recolha dos dados, sugeriu que se construísse no quadro um “tally chart”, onde cada aluno ia assinalar quantos irmãos tinha. Admita que se obteve, como resultado, o seguinte esquema:

0	
1	
2	
3	
4	

Depois de todos os alunos presentes terem ido ao quadro, um diálogo possível pode ser o seguinte:

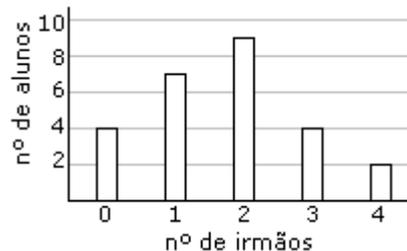
*Professora* – A partir da representação anterior pode-se concluir quantos alunos estão inscritos na turma A?

*Aluno* – Pode-se concluir que neste dia do estudo estão presentes 26 alunos, mas não quantos alunos estão inscritos na turma, já que alguns podem ter faltado.

A seguir, o professor propõe que um dos alunos vá ao quadro e, com a ajuda dos colegas, construa a tabela de frequências para a variável em estudo e o gráfico de barras:

N.º de irmãos	N.º de alunos
0	4
1	7
2	9
3	4
4	2
Total	26

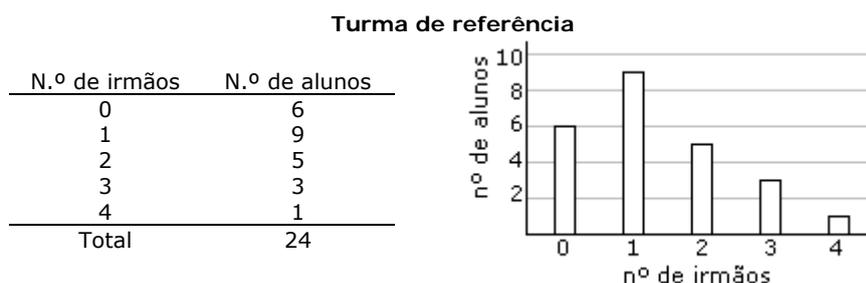
Turma A



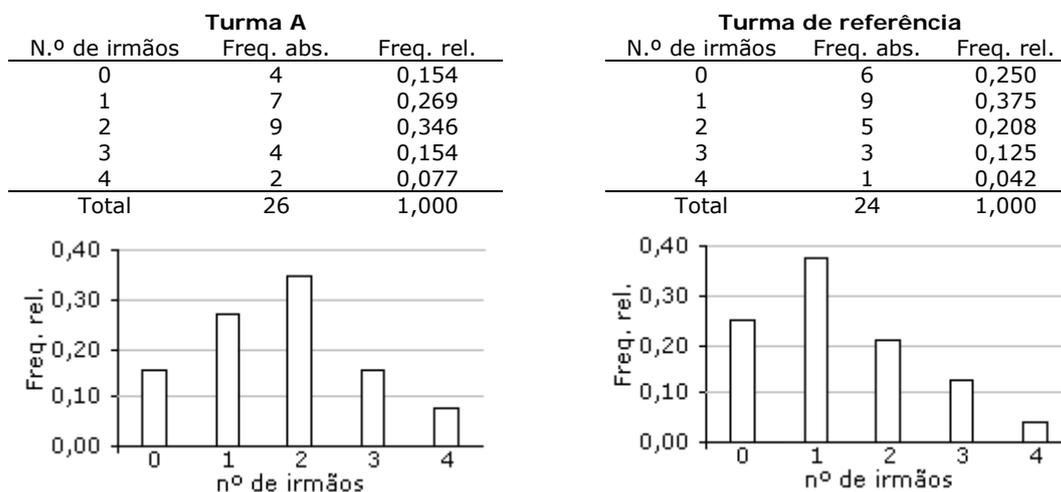
Os alunos chegaram a algumas conclusões interessantes, tais como:

- Predominam os alunos com 2 irmãos;
- Há quatro alunos sem irmãos;
- O número de alunos sem irmãos é igual ao número de alunos com 3 irmãos.

Para proceder à comparação solicitada, construíram o gráfico de barras para a variável *Número de irmãos* da turma de referência, tendo obtido o seguinte:



Os alunos queriam começar a fazer comparações entre os dois gráficos, mas o professor chamou a atenção para um ponto muito importante: é que as duas turmas não tinham o mesmo número de alunos e por isso não é correcto comparar representações gráficas em que as alturas das barras são as frequências absolutas. É necessário juntar às tabelas de frequências uma nova coluna com as frequências relativas e construir outros gráficos de barras em que as alturas das barras são as frequências relativas. Agora sim, já se podem fazer comparações, pois a soma das alturas das barras nos dois casos é igual a 1. Este estudo apresenta-se a seguir:



Algumas conclusões:

- De um modo geral, os alunos da turma A têm mais irmãos que os alunos da turma de referência;
- Enquanto que na turma A, cerca de 15% dos alunos não têm irmãos, na turma de referência esse valor aumenta para 25%;
- Na turma A predominam os alunos com 2 irmãos, enquanto que na turma de referência predominam os alunos com 1 irmão;
- Enquanto que na turma A, mais de 15% dos alunos têm 3 irmãos, na turma de referência esse valor não chega aos 13%.

Perante as conclusões anteriores, o professor lançou a seguinte questão: Não sabemos a idade dos alunos da turma de referência! Poderemos, no entanto, adiantar a hipótese de que são mais novos do que os alunos da turma A?

Porque é que o professor se lembrou de fazer esta suposição?

Ainda continuando com o mesmo tema, o professor colocou as seguintes questões aos alunos:

- Calcular a totalidade de irmãos dos alunos da turma;

- Se todos os alunos tivessem o mesmo número de irmãos, quantos irmãos teriam?

Facilmente os alunos indicaram um processo para calcular o número total de irmãos, pois bastou fazer

$$1 \times 7 + 2 \times 9 + 3 \times 4 + 4 \times 2 = 45$$

Para ver quantos irmãos teriam, se todos tivessem o mesmo número de irmãos, sugeriram que se dividisse o número total de irmãos, pelo número total de alunos

$$\frac{45}{26} \approx 1,7$$

Como interpretar este valor de 1,7, a que chamamos **média** do número de irmãos?

Se todos os alunos tivessem o mesmo número de irmãos, cada aluno teria mais do que 1 irmão, mas não chegaria a ter 2 irmãos, embora estivesse lá perto.

Quantos irmãos mais seriam necessários para dar os 2 irmãos para cada aluno? Seriam necessários 7 irmãos, para ter no total 52 irmãos, já que  $52/26=2$ . Neste caso diríamos que cada aluno tinha, em média, 2 irmãos.

Embora o gráfico de barras seja a representação mais utilizada para dados discretos, a sua utilização nem sempre é a mais conveniente, nomeadamente quando o número de valores distintos assumidos pelos dados é "muito" grande, dando origem a demasiadas classes.

**Exemplo - Candidatos a algumas vagas** (Adaptado de Freedman, 1991). No Distrito Sanitário de Chicago, a escolha dos técnicos é feita mediante um exame. Em 1966, havia 223 candidatos para 15 vagas. O exame teve lugar no dia 12 de Março e os resultados dos testes (inteiros numa escala de 0 a 100) apresentam-se a seguir:

26	27	27	27	27	29	30	30	30	30	31	31	31	32	32
33	33	33	33	33	34	34	34	35	35	36	36	36	37	37
37	37	37	37	37	39	39	39	39	39	39	39	40	41	42
42	42	42	42	43	43	43	43	43	43	43	43	44	44	44
44	44	44	45	45	45	45	45	45	45	46	46	46	46	46
46	47	47	47	47	47	47	48	48	48	48	48	48	48	48
49	49	49	49	50	50	51	51	51	51	51	52	52	52	52
52	53	53	53	53	53	54	54	54	54	54	55	55	55	56
56	56	56	56	57	57	57	57	58	58	58	58	58	58	58
58	59	59	59	59	60	60	60	60	60	60	61	61	61	61
61	61	62	62	62	63	63	64	65	66	66	66	67	67	67
67	68	68	68	69	69	69	69	69	69	69	71	71	72	73
74	74	74	75	75	76	76	78	80	80	80	80	81	81	81
82	82	83	83	83	83	84	84	84	84	84	84	84	90	90
90	91	91	91	92	92	92	93	93	93	93	95	95		

Neste caso, a construção da tabela de frequências, segundo a metodologia descrita para dados discretos, conduziria a uma tabela com demasiadas classes. Assim, resolvemos tomar como classes uma partição natural, para os dados considerados, que é a seguinte: considerar como classes os intervalos 20 a 29, 30 a 39, 40 a 49, 50 a 59, 60 a 69, 70 a 79, 80 a 89, 90 a 99.

Classes	Freq. abs.	Freq. rel.
20 a 29	6	0,027
30 a 39	36	0,161
40 a 49	52	0,233
50 a 59	46	0,206
60 a 69	36	0,161
70 a 79	12	0,054
80 a 89	20	0,090
90 a 99	15	0,067
Total	223	1

Tabela de frequências para os resultados dos testes

A representação gráfica para os dados organizados desta forma já não pode ser um diagrama de barras, pois não existe um ponto onde colocar a barra, uma vez que as classes são intervalos. Veremos, mais à frente, que a representação gráfica adequada é o histograma.

A organização dos dados na forma da tabela anterior permite realçar o facto de predominarem as classificações entre 40 e 49, diminuindo progressivamente para baixo e para cima desses valores. Temos, no entanto de estar conscientes de que ao fazer a redução de dados há informação que sobressai, como a estrutura subjacente aos dados, embora haja outra informação que possivelmente se perde. Vejamos qual o aspecto da tabela se tivéssemos considerado como classes todos os valores distintos da amostra, sem os agrupar:

Classe									
26	1	40	1	52	5	64	1	78	1
27	4	41	1	53	5	65	1	80	4
29	1	42	5	54	5	66	3	81	3
30	4	43	8	55	3	67	4	82	2
31	3	44	6	56	5	68	3	83	4
32	2	45	7	57	4	69	7	84	7
33	5	46	6	58	8	71	2	90	3
34	3	47	6	59	4	72	1	91	3
35	2	48	8	60	6	73	1	92	3
36	3	49	4	61	6	74	3	93	4
37	7	50	2	62	3	75	2	95	2
39	7	51	5	63	2	76	2		

Tabela de frequências para os dados sem estarem agrupados

O diagrama de barras correspondente tem o seguinte aspecto

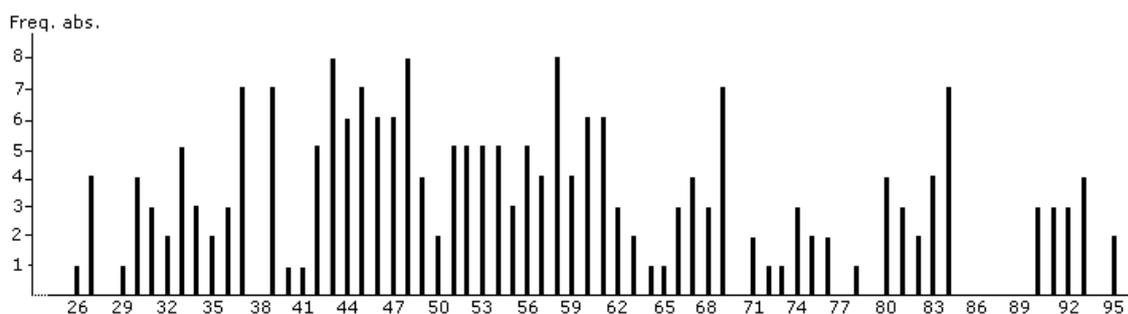


Diagrama de barras dos resultados nos testes

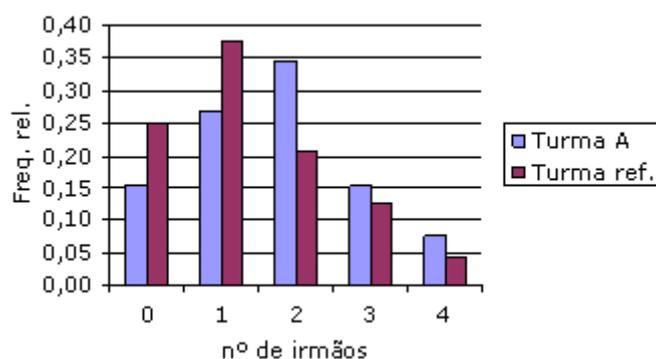
Da análise da tabela e do gráfico anterior verifica-se a existência de uma lacuna, não havendo classificações iguais a 85, 86, 87, 88 e 89 e o número de classificações iguais ou superiores a 90 ser de 15, precisamente igual ao número de vagas, para os 223 candidatos. Não terá havido batota da parte dos examinadores?

Chamamos a atenção para que esta representação, com tantas classes, não evidencia o padrão subjacente à distribuição dos dados, já que apresenta toda a variabilidade neles existente. Como dissemos anteriormente, embora estejamos perante um conjunto de dados discretos, o tratamento adequado será o mesmo dos dados contínuos, apresentado na próxima secção 4.5.

#### 4.4.2.2 1 Gráfico de barras para comparar dois ou mais conjuntos de dados

Quando pretendemos comparar dois (ou mais) conjuntos de dados relativos à mesma variável, uma representação gráfica adequada é o gráfico de barras, em que se apresentam, lado a lado, as distribuições das frequências para cada um dos conjuntos de dados. De preferência devem-se considerar sempre as frequências relativas, pois se os conjuntos de dados não tiverem o mesmo número de elementos, não é correcto utilizar as frequências absolutas.

No caso da tarefa **Número de irmãos** da secção anterior, o gráfico de barras utilizado para comparar as distribuições do número de irmãos na Turma A e na Turma de referência, tem o seguinte aspecto:

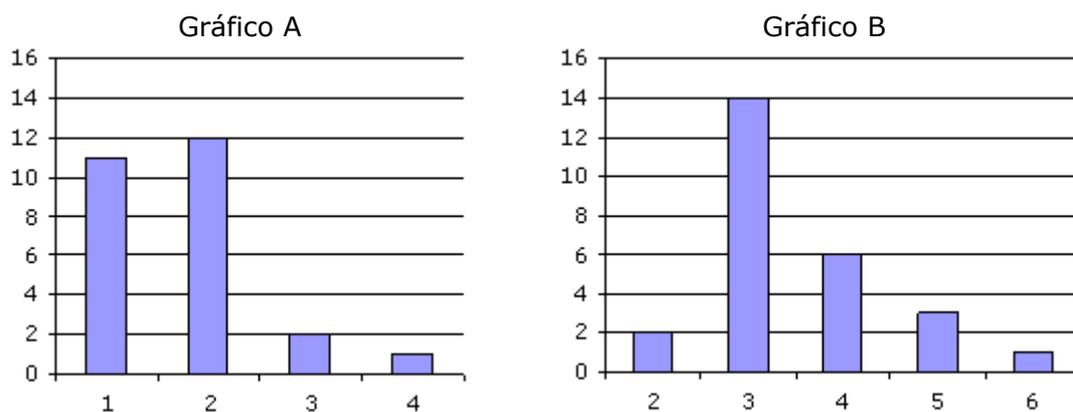


Repare-se que tivemos o cuidado de juntar uma legenda, onde se indica a que turma diz respeito cada cor das barras. Neste tipo de representação é mais fácil a comparação das frequências correspondentes às mesmas classes, uma vez que as barras estão adjacentes.

**Tarefa – Alguns dados sobre o agregado familiar.** Num inquérito realizado na escola, perguntou-se aos 26 alunos de uma turma do 6.º ano:

- Qual a dimensão do seu agregado familiar (quantas pessoas viviam em casa)?
- Quantos são crianças?
- Quantos aparelhos de televisão têm em casa?
- Quantos carros tem o agregado familiar.

A comissão encarregue do estudo apresentou os seguintes gráficos A e B, que procuram resumir a informação contida nas respostas às 2 primeiras questões

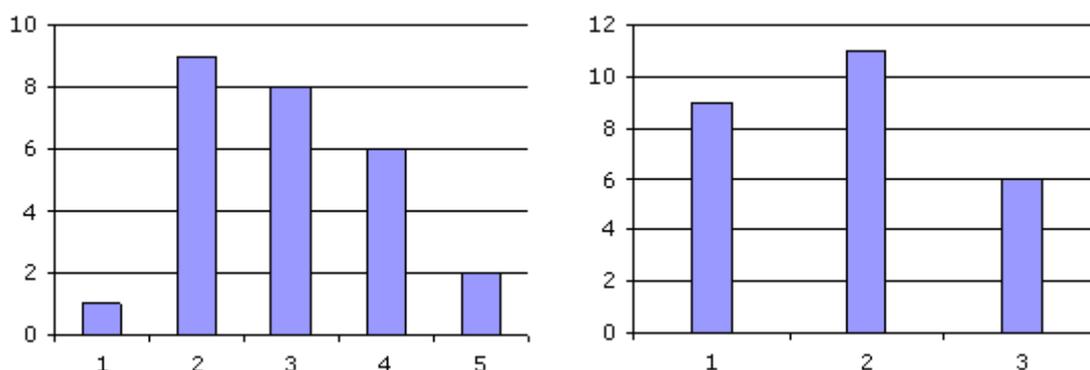


- Qual dos gráficos se refere à variável *número de pessoas do agregado familiar*? Porque é que o outro gráfico não pode representar o número de pessoas do agregado familiar dos 26 alunos a quem foi colocada a questão?
- Quantos agregados familiares têm 2 pessoas? Essas duas pessoas podem ser ambas adultas?
- Dos 26 agregados familiares, 14 são constituídos por quantas pessoas?
- Quantas pessoas tem o maior agregado familiar?
- Quantas pessoas têm os 26 agregados familiares?
- Se o número de pessoas a que chegaste na alínea anterior estivesse dividido igualmente por todos os agregados familiares, quantas pessoas tinha cada agregado familiar? Interpreta o número a que chegaste.

Considera agora também o outro gráfico que representa o número de crianças por agregado familiar.

- Quantos agregados familiares têm 1 criança?
- Qual o número de crianças que predomina nos agregados familiares?
- Pensas que os agregados familiares são fundamentalmente constituídos por um casal com um filho? Explica o teu raciocínio.

Com os dados obtidos nas respostas às outras duas questões, construíram-se os dois gráficos seguintes



- Comparando os dois gráficos, qual dos dois achas mais razoável para representar o *Número de televisões por agregado familiar*?
- No gráfico do lado esquerdo a classe 2 tem frequência absoluta igual a 9 e no gráfico da direita a classe 1 tem também frequência absoluta igual a 9.

---

No entanto as alturas das barras são diferentes. Como explicas esta situação?

- c) Completa os gráficos com as legendas adequadas.

**Tarefa – Cinco montinhos de feijões.** Num determinado dia o professor organizou os alunos em 4 grupos e pediu a cada grupo para levar no dia seguinte uma “mão cheia” de feijões. Neste dia, encarregou cada grupo de dividir todos os feijões em montinhos com 1, 2, 3, 4 ou 5 feijões. Seriam necessários fazer os montinhos necessários até esgotar todos os feijões. Depois dos montinhos feitos, cada grupo foi ao quadro apresentar graficamente o resultado da organização dos feijões pelos montinhos com os cinco tamanhos. Os resultados foram os seguintes:


- a) Comenta as representações utilizadas.  
b) Quantos feijões tinha cada grupo?
-

## 4.5 Tabelas e gráficos para dados quantitativos contínuos

Como vimos na secção 3.3, quando falamos das variáveis, uma variável quantitativa contínua é aquela que é passível de ser medida usando um certo instrumento. Por exemplo, relativamente a um aluno da turma, podemos utilizar uma fita métrica para medir a variável *altura*, uma balança para medir a variável *peso*, um termómetro para medir a *temperatura*, um relógio para medir o *tempo que demora de casa à escola*, um teste a Matemática para medir o *nível de conhecimentos* nessa disciplina, etc. Os valores que resultam destas medições são dados quantitativos contínuos.

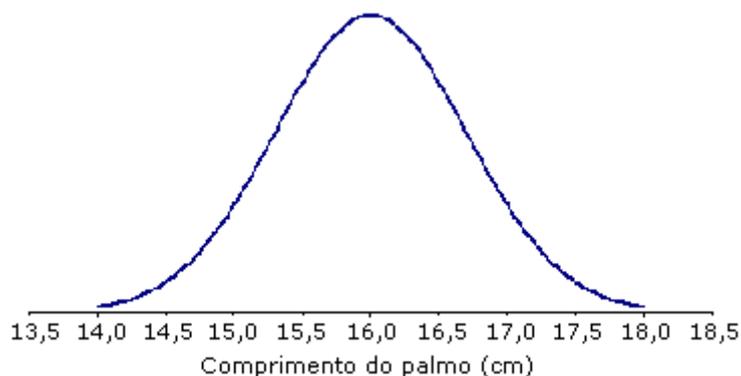
No exemplo da turma de referência as variáveis *tempo que demora de casa à escola* e *comprimento do palmo* são de natureza contínua. Estas variáveis são apresentadas com um arredondamento ao minuto e ao centímetro, respectivamente, mas os seus verdadeiros valores podem ser quaisquer números reais de um determinado intervalo.

Em linguagem corrente pode dizer-se que uma variável contínua não varia por "saltos", isto é, não passa de um valor a outro de um determinado intervalo, sem passar por todos os valores intermédios. Embora seja comum, quando encontramos um jovem que não vemos há algum tempo, exclamar: "Mas que salto que deste! Estás tão alto!", na realidade o jovem cresceu continuamente... Ao contrário da variável contínua, uma variável discreta varia por "saltos". Por exemplo, se uma família tem 2 filhos e teve um outro filho, obviamente que passou de 2 para 3, sem passar por valores intermédios.

Tendo em conta a própria definição de variável contínua, quando temos uma amostra de dados contínuos, estes podem ser todos diferentes, ou quando muito, existem apenas alguns valores iguais. A ocorrência de valores iguais com maior frequência do que a que se esperaria para dados contínuos, deriva do facto do instrumento de medida não ter uma grande precisão. Por exemplo, os valores apresentados para as variáveis contínuas *tempo que demora de casa à escola* e *comprimento do palmo* encontram-se "discretizados" por uma limitação do instrumento que se utilizou para as medir. Outro exemplo de uma variável contínua, que se apresenta "discretizada" é a *idade*. Quando se diz que um jovem tem 9 anos, significa que já fez os 9 anos, mas ainda não fez os dez, pelo que o 9 representa um intervalo de valores que se pode exprimir da seguinte forma:  $9 \leq \text{idade} < 10$ .

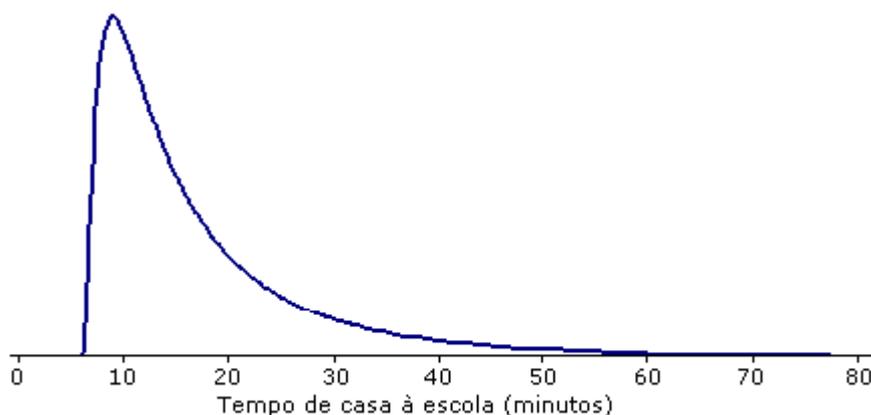
Mesmo existindo alguns valores iguais, o número de valores distintos pode ser tão grande que a metodologia utilizada para construir as tabelas de frequências de dados quantitativos discretos, em que se consideravam para classes os valores distintos nos dados, não pode ser aqui utilizada. Correríamos o risco de a frequência observada para cada valor distinto ser 1! Então, a alternativa é considerar classes na forma de intervalos.

Ao organizar os dados na forma de intervalos, o nosso objectivo é visualizar o padrão subjacente a esses dados. Por exemplo, é natural esperar que uma forma usual para a distribuição da variável *comprimento do palmo* dos alunos do 3.º ciclo tenha um aspecto simétrico, como o que se apresenta na figura seguinte,



com uma concentração de valores em volta dos 16cm, e cada vez menos valores à medida que o comprimento para o palmo diminui ou aumenta.

Já para a variável *tempo de casa à escola* em que, de um modo geral, predominam os tempos mais pequenos, em detrimento dos tempos maiores, esperamos uma distribuição com uma forma enviesada, como a que se apresenta a seguir,



### Como organizar os dados em classes?

Perante um conjunto de dados quantitativos contínuos, ao agrupá-los e ao representá-los graficamente, temos como objectivo que essa representação nos ajude a compreender os dados, fazendo sobressair algum padrão subjacente. Algumas questões que procuramos responder são, por exemplo:

- A distribuição é simétrica ou enviesada?
- Qual o centro da distribuição dos dados?
- Tem pequena ou grande variabilidade?

O primeiro passo no processo de agrupamento dos dados é saber em quantas classes vamos agrupar os dados. Muitas vezes o tipo da variável que se está a estudar pode dar indicação do número de classes e de como construir essas classes.

**Exemplo – Altura e peso dos alunos de uma escola do 1.º ciclo.** Pretendemos estudar as variáveis *altura* e *peso* dos alunos de uma escola do 1.º ciclo. Para isso, recolhemos a altura e o peso de 50 alunos dessa escola, obtendo os valores (em cm) para a *altura* e os valores (em kg) para o *peso* que se apresentam na seguinte tabela:

Altura	Peso								
132	26	135	29	146	40	142	32	143	35
145	39	145	35	141	33	143	34	147	40
150	45	136	30	144	35	146	40	147	40
149	45	143	32	159	57	151	46	135	29
130	26	137	30	157	49	135	30	132	28
135	30	141	30	158	58	143	38	140	30
145	40	135	29	134	30	140	31	138	30
130	28	141	32	146	40	146	43	154	47
148	40	145	35	145	34	156	45	150	45
150	47	136	30	148	43	133	29	130	28

Ao percorrer, na tabela, os dados referentes à variável *altura*, verificamos que o valor mínimo é 130cm e o valor máximo 159cm. Assim, é natural considerar como classes, para organizar os dados, as seguintes:

$$\begin{aligned}
 &130 \leq \text{altura} < 135 \\
 &135 \leq \text{altura} < 140 \\
 &140 \leq \text{altura} < 145 \\
 &145 \leq \text{altura} < 150 \\
 &150 \leq \text{altura} < 155 \\
 &155 \leq \text{altura} < 160
 \end{aligned}$$

As classes são todas disjuntas e a sua união contém todos os elementos da amostra, isto é, cada elemento da amostra só pode pertencer a uma das classes, mas pertence necessariamente a uma dessas classes.

Considerando agora os dados referentes à variável *peso*, verificamos que os valores máximos e mínimos são respectivamente 26kg e 58kg, pelo que uma escolha possível para as classes é:

$$\begin{aligned}
 &25 \leq \text{peso} < 30 \\
 &30 \leq \text{peso} < 35 \\
 &35 \leq \text{peso} < 40 \\
 &40 \leq \text{peso} < 45 \\
 &45 \leq \text{peso} < 50 \\
 &50 \leq \text{peso} < 55 \\
 &55 \leq \text{peso} < 60
 \end{aligned}$$

Do mesmo modo que anteriormente para a variável *altura*, também as classes anteriores foram construídas sem ambiguidade, na medida em que cada elemento da amostra pertence a alguma das classes e só a uma das classes.

### Regra de Sturges

Nos exemplos apresentados anteriormente, a formação de classes foi fácil de fazer de forma intuitiva. No entanto, isso nem sempre acontece. Nestes casos podemos usar a chamada *regra de Sturges*, que nos sugere o número de classes a usar para agrupar os dados:

*Regra de Sturges* – para organizar uma amostra, de dados contínuos, de dimensão  $n$ , pode considerar-se para número de classes o valor  $k$ , onde  $k$  é o menor inteiro tal que  $2^k > n$ .

Assim, se o número de elementos da amostra for 50, como nos exemplos apresentados anteriormente, o número aconselhado de classes é 6, já que  $2^5 < 50$  e  $2^6 > 50$ . Note-se que esta regra não tem que ser seguida “à letra” e deve ser entendida como uma ajuda, quando não se tem qualquer ideia de quantas classes construir,

para proceder ao agrupamento dos dados. É apresentada como informação para o professor, que, para este nível de ensino, não a deve ensinar aos alunos.

Para a formação das classes, na forma de intervalos, com a mesma amplitude, considera-se a seguinte metodologia:

Passo 1 – Toma-se como amplitude  $h$ , de cada intervalo, um valor arredondado por excesso, do quociente que se obtém dividindo a amplitude da amostra (máximo – mínimo) pelo número de classes,  $k$ .

Passo 2 – Formam-se as classes como intervalos fechados à esquerda e abertos à direita, ou vice-versa, isto é, abertos à esquerda e fechados à direita, sendo o extremo esquerdo do primeiro intervalo o mínimo da amostra, ou o extremo direito do  $k$ -ésimo intervalo o máximo da amostra, respectivamente.

### **Subdivisão em classes dos dados referentes à variável *Tempo que demora de casa à escola na Turma de referência***

Os valores observados são, depois de ordenados

5 6 6 7 7 8 9 10 10 11 12 12 13 13 14 15 15 15 16 17 18 19 20 21

Como a dimensão da nossa amostra é  $n=24$ , o menor inteiro  $k$  que satisfaz a condição  $2^k > 24$  é  $k=5$ . Para obter a amplitude de classe  $h$ , vamos dividir a amplitude da amostra, que é 16 ( $=21 - 5$ ), por 5. Este quociente vem igual a 3,2, pelo que um valor aproximado **por excesso** é, por exemplo, 3,25.

Para a construção das classes vamos convencionar que todos os intervalos são fechados à esquerda e abertos à direita, isto é, da forma  $[a, b[$ , onde o  $a$  pertence ao intervalo, mas o  $b$  já não pertence. Utilizando esta metodologia, temos os seguintes intervalos, para as classes:

1ª classe: $[5; 5+3,25[$	→	$[5; 8,25[$
2ª classe: $[8,25; 8,25+3,25[$	→	$[8,25; 11,50[$
3ª classe: $[11,50; 11,50+3,25[$	→	$[11,50; 14,75[$
4ª classe: $[14,75; 14,75+3,25[$	→	$[14,75; 18,00[$
5ª classe: $[18,00; 18,00+3,25[$	→	$[18,00; 21,25[$

O valor de 3,25 que utilizámos para a amplitude de classe, como aproximação por excesso do valor 3,2, é pouco natural. Mas o mesmo não acontece com 3 minutos e meio, pelo que outra alternativa possível para a amplitude de classe será  $h=3,5$ . Se se considerar este valor, o número de classes a usar é ainda de 5, como se pode ver facilmente, já que as classes que assim se obtêm

$[5; 8,5[$ ,  $[8,5; 12,0[$ ,  $[12,0; 15,5[$ ,  $[15,5; 19,0[$  e  $[19,0; 22,5[$

contêm todos os elementos da amostra.

Se se pretender construir intervalos em que os limites sejam números inteiros, podemos considerar como amplitude de classe 3 minutos ou 4 minutos, obtendo-se, respectivamente, as seguintes classes:

Amplitude de classe igual a 3 minutos	Amplitude de classe igual a 4 minutos
[5; 8[	[5; 9[
[8; 11[	[9; 13[
[11; 14[	[13; 17[
[14; 17[	[17; 21[
[17; 20[	[21; 25[
[20; 23[	

Repare-se que, quando se considerou como amplitude de classe 3 minutos, foi necessário construir 6 classes, de modo a cobrirem a totalidade dos dados; por outro lado, quando se considerou como amplitude de classe o valor 4 minutos, consideraram-se 5 classes, mas a última classe só tem um elemento. Chamamos a atenção para que não é correcto considerar a quarta classe na forma [17; 21], com o objectivo de evitar mais uma classe. A metodologia na construção dos intervalos de classe deve ser sempre a mesma: fechados à esquerda e abertos à direita, ou vice-versa.

Deste modo, existe uma grande maleabilidade na construção dos intervalos de classe. Em muitas situações, a regra básica a seguir é utilizar a informação disponível sobre a variável a estudar e o "bom senso" para a definição dos limites das classes.

A regra de *Sturges* pode ser usada como um primeiro passo na indicação de um número apropriado de classes. Na verdade, o que nós procuramos é um agrupamento dos dados em classes, para depois construirmos o histograma que, como veremos, deve evidenciar a estrutura subjacente aos dados. Assim, se se construírem muitas classes, essa representação apresentará muita da variabilidade presente nos dados, não conseguindo fazer sobressair o padrão que procuramos. Também um número muito pequeno de classes esconderá esse padrão.

#### 4.5.1 Tabela de frequências para dados quantitativos contínuos

Uma vez formadas as classes, a construção da tabela de frequências é idêntica à considerada para os dados discretos:

Os dados contínuos são organizados na forma de uma **tabela de frequências**, com três ou mais colunas. Na primeira coluna, *coluna das classes*, consideram-se os intervalos (classes) escolhidos para agrupar os dados; na coluna seguinte, *coluna das frequências absolutas*  $n_i$ , regista-se o total de elementos da amostra, que pertencem a cada classe. Numa terceira coluna, *coluna das frequências relativas* (ou percentagens)  $f_i$ , regista-se, para cada classe, o valor que se obtém dividindo a frequência absoluta pela dimensão da amostra.

De um modo geral é útil acrescentar à tabela de frequências ainda mais três colunas: coluna do *representante de classe* – em que se considera, usualmente, o ponto médio,  $x'_i$ , do intervalo de classe; coluna das *frequências absolutas acumuladas* – onde, para cada classe, se considera a soma da frequência absoluta dessa classe com as frequências absolutas das classes anteriores e coluna das *frequências rela-*

*tivas acumuladas*, que se calcula de forma idêntica à anterior, mas agora com as frequências relativas.

Vamos a seguir construir a tabela de frequências para os dados observados para a variável *altura* de um aluno da escola do 1.º ciclo, considerados na secção anterior.

Considerámos as 6 classes aí definidas, com intervalos de amplitude 5cm, fechados à esquerda e abertos à direita:

Classes	Representante da Classe $x'_i$	Freq. Abs. $n_i$	Freq. Rel. $f_i$	Freq. Abs. Acum	Freq. Rel. Acum.	Freq. Rel. Acum. (%)
[130, 135[	132,5	7	0,14	7	0,14	14
[135, 140[	137,5	9	0,18	16	0,32	32
[140, 145[	142,5	11	0,22	27	0,54	54
[145, 150[	147,5	14	0,28	41	0,82	82
[150, 155[	152,5	5	0,10	46	0,92	92
[155, 160[	157,5	4	0,08	50	1,00	100
Total		50	1,00			

Decidimos ainda acrescentar uma outra coluna, com as frequências relativas acumuladas, agora em percentagem.

A frequência absoluta da classe [130, 135[ é 7, porque existem nos dados 7 valores maiores ou iguais a 130 e menores que 135. Para as outras classes a metodologia é idêntica.

A soma das frequências absolutas é igual a 50, que é o número de dados, enquanto que a soma das frequências relativas é igual a 1. Por vezes, esta soma não dá exactamente 1, sendo esta situação devida ao facto dos valores das frequências relativas serem arredondados.

Como se verifica a partir da tabela predominam as alturas das classes centrais, havendo uma diminuição das frequências para as classes inferiores e superiores.

#### 4.5.2 Histograma

Uma vez os dados agrupados numa tabela de frequências, estamos aptos a construir o histograma, que é a representação gráfica mais utilizada para os dados quantitativos contínuos.

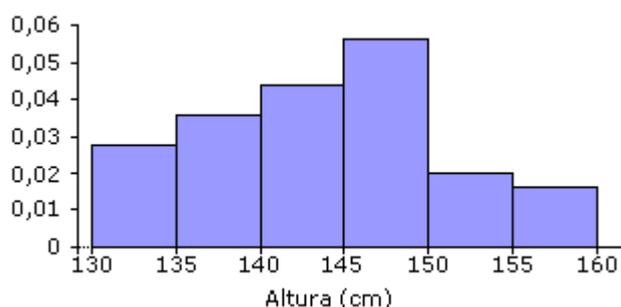
O histograma é um gráfico, formado por uma sucessão de rectângulos adjacentes, tendo cada um por base um intervalo de classe e com área igual (ou proporcional) à frequência relativa (ou absoluta) dessa classe.

Ao contrário do gráfico de barras, em que estas estão separadas e em que o que é relevante é a altura de cada uma, no histograma as barras (rectângulos) estão juntas e o que é importante é a área de cada uma.

Considerando então para áreas das barras as frequências relativas, vemos que a área total ocupada pelo histograma é igual a 1 ou 100%.

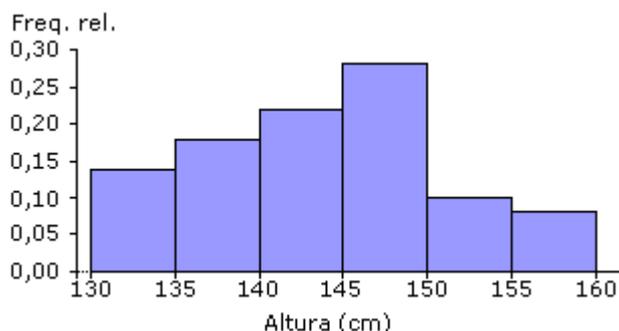
Tendo em conta a definição de histograma, para a sua construção é conveniente acrescentar uma nova coluna à tabela de frequências, com as frequências relativas a dividir pela amplitude de classe. Os valores desta coluna serão as alturas dos rectângulos com base nas classes respectivas:

Classes	Rep. Classe $x'_i$	Freq. Abs. $n_i$	Freq. Rel. $f_i$	Altura rectângulo classe $i=f_i/h$
[130, 135[	132,5	7	0,14	0,028
[135, 140[	137,5	9	0,18	0,036
[140, 145[	142,5	11	0,22	0,044
[145, 150[	147,5	14	0,28	0,056
[150, 155[	152,5	5	0,10	0,020
[155, 160[	157,5	4	0,08	0,016
Total		50	1,00	



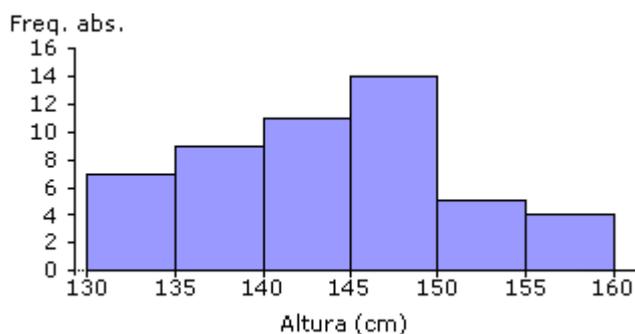
No histograma ao lado, a área do rectângulo mais à esquerda é igual a  $5 \times 0,028 = 0,14$ ; a área do rectângulo seguinte é  $5 \times 0,036 = 0,18$  e assim sucessivamente, donde a **área total do histograma é igual a 1** (soma das frequências relativas).

Suponhamos que em vez de construirmos o histograma como anteriormente, tínhamos considerado para alturas dos rectângulos as frequências relativas. Então, neste caso, as áreas dos rectângulos já não seriam iguais às frequências relativas, mas sim proporcionais e a área total ocupada pelo histograma seria igual a 5, em que 5 é a amplitude de classe:



No histograma ao lado, a área do rectângulo mais à esquerda é igual a  $5 \times 0,14$ ; a área do rectângulo seguinte é  $5 \times 0,18$  e assim sucessivamente, donde a **área total do histograma é igual a 5** ( $=5 \times 1$  onde 1 é a soma das frequências relativas).

Suponhamos ainda que agora se considerava para altura dos rectângulos as frequências absolutas. O resultado seria o seguinte:



No histograma ao lado, a área do rectângulo mais à esquerda é igual a  $5 \times 7$ ; a área do rectângulo seguinte é  $5 \times 9$  e assim sucessivamente, donde a **área total do histograma é igual a 250** ( $=5 \times 50$ , onde 50 é a soma das frequências absolutas).

Como se verifica, a imagem transmitida tem sempre o mesmo aspecto, já que as áreas dos rectângulos ou são iguais às frequências relativas, como é o caso do primeiro dos 3 histogramas anteriores, ou são proporcionais, com a mesma constante de proporcionalidade, que é igual à amplitude de classe no caso do segundo histograma ou à amplitude de classe vezes o número de dados, como é o caso do terceiro histograma. Assim, o eixo vertical só serve como auxílio para a construção dos rectângulos, não transmitindo, no caso do histograma, qualquer informação relevante:

Não devemos perder de vista que o histograma representa os dados através das áreas das barras e não das alturas, o que constitui uma grande diferença relativamente ao gráfico de barras. Outra grande diferença é que no histograma as barras estão juntas, para transmitir a ideia de continuidade da variável em estudo, enquanto que no gráfico de barras, estas são separadas.

De um modo geral, se tivermos  $n$  dados e estes tiverem sido organizados em  $k$  classes, todas com a mesma amplitude  $h$ , e representarmos por  $n_i$  e  $f_i$ , respectivamente as frequências absoluta e relativa da classe  $i$ , com  $i=1, \dots, k$ , a área total ocupada pelo histograma será igual a:

- $1$ , se se considerar para altura do rectângulo correspondente à classe  $i$ ,  $f_i/h$ , com  $i=1, \dots, k$ .
- $h$ , se se considerar para altura do rectângulo correspondente à classe  $i$ ,  $f_i$ , com  $i=1, \dots, k$ .
- $h \times n$ , se se considerar para altura do rectângulo correspondente à classe  $i$ ,  $n_i$ , com  $i=1, \dots, k$ .

Qualquer das formas anteriores pode ser utilizada para construir o histograma, excepto nas seguintes situações:

- As classes têm amplitudes diferentes, sendo, neste caso, necessário utilizar o primeiro procedimento;
- Pretende-se comparar histogramas de amostras com dimensão diferente, sendo, também necessário utilizar o primeiro procedimento, para compararmos figuras com a mesma área (igual a 1).

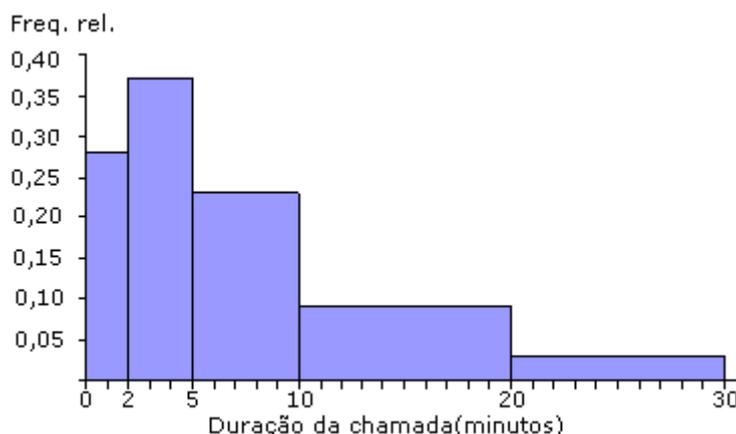
**Exemplo – Duração de chamadas telefónicas<sup>1</sup>.** Uma empresa, preocupada com os gastos em telefone, decidiu fazer um estudo sobre a *duração* (em minutos) das chamadas telefónicas. Assim, o departamento de controlo de qualidade recolheu uma amostra de dimensão 100, tendo construído a seguinte tabela de frequências, com os dados recolhidos:

Duração da chamada (em minutos)

Classes	Freq. absoluta	Freq. relativa
[0, 2[	28	0,28
[2, 5[	37	0,37
[5, 10[	23	0,23
[10, 20[	9	0,09
[20, 30[	3	0,03
Total	100	1,00

Elaborou depois o seguinte histograma, que apresentou à gerência:

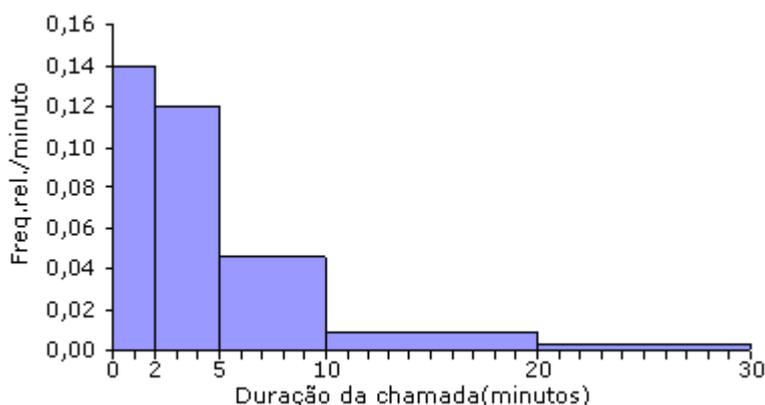
<sup>1</sup> Graça Martins et al. (2007).



Um dos gerentes, que sabia o que era um histograma, manifestou-se bastante preocupado com a percentagem de chamadas razoavelmente longas, já que a percentagem de chamadas com duração entre 5 e 10 minutos era um pouco superior às de duração entre 2 e 5 minutos e só um pouco inferior às de duração de 10 a 20 minutos, como se depreende pelas áreas dos rectângulos correspondentes às classes respectivas. Pediu para consultar a tabela de frequências e concluiu que aquela representação gráfica não estava correcta, pois as áreas dos rectângulos não eram proporcionais às frequências, induzindo em erro. Ele próprio acrescentou mais uma coluna à tabela de frequências, com as alturas correctas dos rectângulos e construiu o histograma correspondente:

Duração da chamada (em minutos)

Classes	Freq. absoluta	Freq. relativa	Freq. relativa/amplitude classe
[0, 2[	28	0,28	0,140
[2, 5[	37	0,37	0,122
[5, 10[	23	0,23	0,046
[10, 20[	9	0,09	0,009
[20, 30[	3	0,03	0,003
Total	100	1,00	



Repare-se que as duas representações são completamente diferentes. Agora, podemos concluir que predominam as chamadas com duração entre 2 e 5 minutos e que as chamadas com duração superior a 10 minutos são pouco frequentes.

#### 4.5.3 Gráficos das frequências relativas acumuladas

A partir da tabela de frequências, em que se consideraram as frequências relativas acumuladas, é possível construir gráficos que apresentam a evolução dessas fre-

quências ao longo das classes em que se organizaram os dados. Dois desses gráficos são o chamado **histograma cumulativo** ou histograma acumulado e a **função cumulativa**. Utilizam-se fundamentalmente na determinação gráfica de valores aproximados para a mediana e quartis, quando os dados estão agrupados. Estas medidas serão estudadas mais à frente, quando considerarmos as características amostrais, mas vamos indicar a forma de as obter devido ao facto de serem muito simples de compreender e de usar na construção de um diagrama de extremos e quartis, uma representação gráfica muito útil.

Como veremos, a **mediana**, representada por  $Me$ , é um valor que divide a amostra ordenada ao meio, isto é, 50% dos elementos da amostra são menores ou iguais à mediana e os restantes 50% são maiores ou iguais à mediana. Uma vez a amostra dividida em duas partes com igual número de elementos, cada uma destas partes ainda pode ser dividida ao meio. Às medianas da parte inferior e superior dos dados, chamamos respectivamente **1.º quartil** e **3.º quartil** e representamos por  $Q_1$  e  $Q_3$ . Assim, o 1.º quartil, a mediana e o 3.º quartil dividem a amostra (ordenada) em 4 partes iguais, cada uma contendo 25% dos dados.

Para obter graficamente estas medidas para os dados relativos à altura de 50 alunos de uma escola do 1.º ciclo, mas a partir dos dados agrupados, consideremos de novo a tabela da secção 4.5.1, que já contém as frequências relativas acumuladas:

Classes	Rep. Classe $x'_i$	Freq. Abs. $n_i$	Freq. Rel. $f_i$	Freq. Abs. Acum	Freq. Rel. Acum.	Freq. Rel. Acum. (%)
[130, 135[	132,5	7	0,14	7	0,14	14
[135, 140[	137,5	9	0,18	16	0,32	32
[140, 145[	142,5	11	0,22	27	0,54	54
[145, 150[	147,5	14	0,28	41	0,82	82
[150, 155[	152,5	5	0,10	46	0,92	92
[155, 160[	157,5	4	0,08	50	1,00	100
Total		50	1,00			

Repare-se que da tabela anterior ficamos a saber que a mediana se encontra na classe [140, 145[, já que antes desta classe se acumulam 32% dos dados e é nesta classe que se atinge os 50%. Admitindo que a frequência se distribui uniformemente sobre a amplitude de classe, isto é, a frequência de 22% (=54%-32%) se distribui uniformemente sobre o intervalo de amplitude 5cm, através da resolução de uma equação de proporcionalidade, obtém-se o incremento que devemos adicionar a 140cm, para obter a mediana:

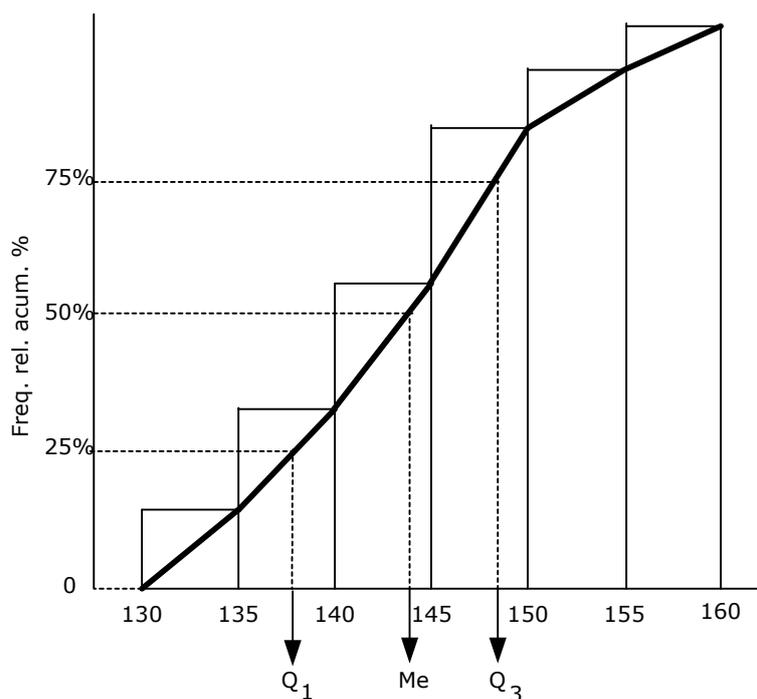
$$22 = 54 - 32 \quad 5 = 145 - 140$$

$$18 = 50 - 32 \quad x$$

$$x = \frac{18 \times 5}{22} \approx 3,6$$

Então o valor aproximado para a mediana será 143,6 (=140+3,6).

O processo que acabámos de descrever para a mediana pode também ser utilizado para obter valores aproximados para os quartis e é equivalente à seguinte resolução gráfica:



Os rectângulos do gráfico anterior têm por altura a frequência relativa acumulada e por essa razão se chama **histograma cumulativo** ao gráfico obtido. À linha poligonal que se desenhou na figura anterior chama-se **função cumulativa**. Pode dizer-se que é esta função que é útil na determinação dos quartis e da mediana, pelo que o histograma cumulativo só serviu como meio auxiliar para a sua construção, não tendo, neste momento, qualquer outra utilidade ou interpretação (observe-se que a função cumulativa se pode construir independentemente do histograma cumulativo).

Como a figura sugere o valor da mediana encontra-se próximo de 144. Do mesmo modo podemos avançar que o 1.º quartil deve andar próximo de 138, enquanto o 3.º quartil deve estar próximo de 148.

## 4.6 Outras representações gráficas

Para representar dados quantitativos usámos até aqui várias representações, de que destacamos o gráfico de barras e o histograma, de um modo geral utilizados para representar dados quantitativos discretos ou contínuos, respectivamente. Existem outras representações gráficas que podem ser utilizadas para dados quantitativos de qualquer tipo e que são o caule-e-folhas e o diagrama de extremos e quartis, e ainda o gráfico de linha, especialmente adequado para representar observações de variáveis que variam ao longo do tempo.

### 4.6.1 Gráfico (ou diagrama) de caule-e-folhas

Pode considerar-se que o gráfico ou diagrama em caule-e-folhas é um tipo de representação que se situa entre a tabela e o gráfico, uma vez que, de um modo geral, apresenta os verdadeiros valores da amostra, mas de uma forma sugestiva, que faz lembrar o histograma.

A base da construção de uma representação em caule-e-folhas está na escolha de um par de dígitos adjacentes nos dados, que vai permitir dividir cada dado do conjunto de dados em duas partes: o *caule* e a *folha*, que se dispõem para um e outro lado de um traço vertical, como exemplificamos a seguir.

**Tarefa – Quantos segundos se consegue estar sem respirar<sup>2</sup>.** Gostaríamos de ter uma ideia de quantos segundos conseguimos estar sem respirar. Suponha que um grupo de alunos fez esta experiência na turma e obteve os seguintes valores: 59, 38, 47, 23, 48, 55, 37, 48, 53, 37, 52, 39, 54, 57, 38, 46, 40, 41, 62, 63, 38, 65, 44, 68, 27, 35, 46, 60.

Podem ser feitas perguntas do tipo:

- Quantos segundos esteve sem respirar o aluno que aguentou menos tempo? E o aluno que aguentou mais tempo?
- O professor pode, com a ajuda dos alunos, organizar os dados num diagrama de caule-e-folhas.

Como o menor e o maior dos valores anteriores são, respectivamente, 23 e 68, para organizar os dados num gráfico de caule-e-folhas, vamos começar por considerar os seguintes caules (algarismos das dezenas dos valores iniciais):

2  
3  
4  
5  
6

Depois de considerar um segmento de linha vertical, ao lado dos caules, vamos pendurar as folhas, nos caules respectivos. Exemplificamos a seguir, um gráfico com a primeira folha, um outro com a primeira e a segunda folha e, finalmente, o gráfico com as folhas todas:

2		2		2	3	7						
3		3	8	3	8	7	7	9	8	8	5	
4		4		4	7	8	8	6	0	1	4	6
5	9	5	9	5	9	5	3	2	4	7		
6		6		6	2	3	5	8	0			

É costume ordenar as folhas correspondentes a cada caule, de modo que o gráfico final é o seguinte:

2	3	7
3	5	7
3	7	8
3	8	8
3	8	9
4	0	1
4	4	6
4	6	7
4	8	8
5	2	3
5	4	5
5	7	9
6	0	2
6	3	5
6	8	8

Repare-se que agora é muito fácil ordenar o conjunto de dados inicial, pois basta percorrer o gráfico de caule-e-folhas:

23, 27, 35, 37, 37, 38, 38, 38, 39, 40, 51, 44, 46, 46, 47, 48, 48, 52, 53, 54, 55, 57, 59, 60, 62, 63, 65 e 68.

<sup>2</sup> Graça Martins et al. (2007), p. 62.

**Sugestão** – Pode ser repetida a tarefa anterior, mas depois de ter inspirado e expirado, profundamente, 3 vezes. É interessante comparar os resultados agora obtidos, com os anteriores.

É de notar que a representação em caule-e-folhas:

- É, em geral, muito simples de fazer e torna-se, por isso, acessível, até a alunos do 1.º ciclo. É necessário ter algum cuidado na escolha do exemplo, para que não haja problemas na escolha do(s) dígito(s) que vão constituir os caules - denominados *dígito(s) dominantes*.
- Dá uma informação visual sobre a forma como os dados estão distribuídos.
- Permite identificar os dados iniciais, pelo que é muito útil para ordenar rapidamente a amostra.
- É uma representação muito sugestiva para comparar duas amostras.
- Facilita o cálculo da mediana e dos quartis.

### Escolha dos dígitos dominantes

Na construção de um gráfico de caule-e-folhas nem sempre é imediata a escolha dos dígitos dominantes. Se essa escolha conduzir a muitos caules o resultado final tem pouco de representação gráfica, pois será muito disperso. Se conduzir a poucos caules, para além de poder esconder padrões nos dados, torna-se de pouca utilidade na ordenação da amostra. Vamos ver o que acontece, por exemplo, com os dados da variável *Altura de um aluno de uma escola do 1.º ciclo*, de que se tem uma amostra de 50 dados, considerados no início da secção 4.5 e que repetimos a seguir:

132	135	135	141	146	158	142	143	143	140
145	145	145	135	141	134	143	140	147	138
150	130	136	141	144	146	146	146	147	154
149	148	143	145	159	145	151	156	135	150
130	150	137	136	157	148	135	133	132	130

As alturas variam entre 130cm e 159cm. Se tomarmos como dígito dominante o das centenas, ficaremos apenas com 1 caule. Se tomarmos os dois primeiros dígitos (até à classe das dezenas), ficaremos com 3 caules, o que também é pouco tendo em conta que a dimensão da amostra é  $n=50$ . Este problema pode ser resolvido subdividindo em dois, cada um dos 3 caules que se obtêm no segundo caso. No primeiro desses dois caules, identificado com um asterisco (\*), colocam-se as folhas de dígitos 0, 1, 2, 3, e 4 e no outro, identificado com um ponto (.), as folhas de dígitos 5, 6, 7, 8, e 9. Deste modo ficamos ao todo com 6 caules que é um número razoável para a dimensão de amostra que temos. Há ainda a possibilidade de subdividir cada caule em 5, um para as folhas 0 e 1, outro para as folhas 2 e 3, e assim por diante até ao último que terá as folhas 8 e 9, mas iríamos obter 15 caules que já é excessivo. A subdivisão de cada caule só pode ser feita em 2 ou 5 sub caules, para que cada um destes possa ficar com igual número de folhas associado e como na notação decimal temos 10 folhas possíveis, ou associamos 5 ou 2 folhas a cada caule.

Um gráfico de caule-e-folhas para representar os dados anteriores (onde a unidade de cada caule é a dezena de centímetros) é, então:

13*	2	0	0	4	3	2	0							
13.	5	5	6	7	5	6	5	5	8					
14*	3	1	1	1	4	2	3	3	0	3	0			
14.	5	9	5	8	5	5	6	6	5	8	6	6	7	7
15*	0	0	1	4	0									
15.	9	7	8	6										

Para construirmos o diagrama anterior percorremos os dados, coluna a coluna. O diagrama final deve apresentar-se com as folhas ordenadas:

13*	0	0	0	2	2	3	4							
13.	5	5	5	5	5	6	6	7	8					
14*	0	0	1	1	1	2	3	3	3	3	4			
14.	5	5	5	5	5	6	6	6	6	7	7	8	8	9
15*	0	0	0	1	4									
15.	6	7	8	9										

A partir do diagrama anterior facilmente se obtém a amostra ordenada:

130 130 130 132 132 133 134 135 135 135 135 135 136 136 137  
 138 140 140 141 141 141 142 143 143 143 143 144 145 145 145  
 145 145 146 146 146 146 147 147 148 148 149 150 150 150 151  
 154 156 157 158 159

A amostra ordenada ou a própria representação em caule-e-folhas com as folhas ordenadas, são utilizadas para o cálculo de algumas estatísticas ordinais, isto é, medidas que nos dão uma ideia da proporção ou percentagem de elementos da amostra menores ou maiores que determinado valor. Por exemplo, na amostra anterior, como o máximo é 159, obviamente que 100% dos elementos da amostra são menores ou iguais a 159. Também se podem calcular outras percentagens:

- 6% (3 elementos) dos elementos da amostra são menores ou iguais a 130;
- 10% (5 elementos) dos elementos da amostra são menores ou iguais a 132;
- 24% (12 elementos) dos elementos da amostra são menores ou iguais a 135;
- 28% (14 elementos) dos elementos da amostra são menores ou iguais a 136;
- 32% (16 elementos) dos elementos da amostra são menores ou iguais a 138;
- 44% (22 elementos) dos elementos da amostra são menores ou iguais a 142;
- 52% (26 elementos) dos elementos da amostra são menores ou iguais a 143;
- etc.

É evidente que para contar o número de elementos para calcular as percentagens anteriores não teríamos necessidade de dispor da amostra ordenada, se tivermos uma representação em caule-e-folhas, como é a situação presente. Efectivamente esta operação de contagem é mais fácil a partir do caule-e-folhas (com as folhas ordenadas) do que propriamente a partir da amostra ordenada.

Um problema inverso do anterior é, dada uma percentagem, procurar um valor que satisfaça determinadas condições. Por exemplo, qual o **valor** tal que 50% dos elementos da amostra são menores ou iguais a ele e os outros 50% são maiores ou iguais a ele, ou seja, com a notação introduzida na secção anterior, qual é o valor da mediana  $Me$ ?

Se a nossa amostra tivesse um número ímpar de dados, então seria o dado do meio, quando a amostra está ordenada. Como temos 50 elementos, então temos dois elementos no meio, pelo que qualquer valor entre esses dois estava em condições de satisfazer a nossa pretensão. No entanto, para evitar que nesta situação de termos um número par de dados, cada um escolha um valor diferente para a mediana, convencionou-se que esta é a semi-soma desses dois elementos do meio. A partir do caule-e-folhas facilmente verificamos que os elementos na posição 25 e 26 são iguais a 143. Então o valor procurado é 143  $(=(143+143)/2)$ . Se determinarmos, pelo mesmo processo, as medianas de cada uma das partes, em que os dados ficam divididos pela mediana, obtemos os quartis, respectivamente **1.º quartil** se for o da parte inferior e **3.º quartil** se for da parte superior. No caule-e-folhas seguinte assinalámos a negro a metade inferior:

```

13* | 0 0 0 2 2 3 4
13. | 5 5 5 5 5 6 6 7 8
14* | 0 0 1 1 1 2 3 3 3 3 4
14. | 5 5 5 5 5 6 6 6 6 7 7 8 8 9
15* | 0 0 0 1 4
15. | 6 7 8 9

```

Como temos agora 25 elementos, a mediana é o elemento do meio, ou seja o que está na posição 13, o valor 136. Então o **1.º quartil** é 136. Para determinar o **3.º quartil**, podemos ver qual o elemento que está na posição do meio da parte superior, a *itálico*, ou seja o que está na posição 13, mas agora a contar do fim, obtendo 147.

Chamamos a atenção para que na leitura do caule-e-folhas, a partir do fim, se começa sempre, em cada linha, pela folha mais afastada:

159 158 157 156 154 ...

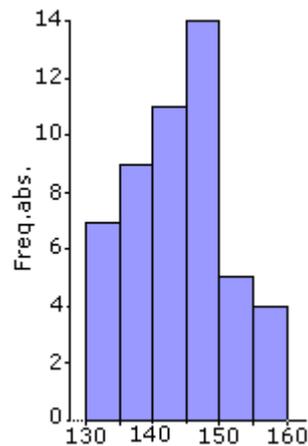
### Comparação entre o caule-e-folhas e o histograma

O caule-e-folhas dá uma imagem muito semelhante ao histograma. Vamos exemplificar, rodando de 90º, no sentido contrário ao ponteiro do relógio, o caule-e-folhas anterior e comparar com o histograma construído na secção anterior, para os mesmos dados:

```

13* | 0 0 0 2 2 3 4
13. | 5 5 5 5 5 6 6 7 8
14* | 0 0 1 1 1 2 3 3 3 3 4
14. | 5 5 5 5 5 6 6 6 6 7 7 8 8 9
15* | 0 0 0 1 4
15. | 6 7 8 9

```



No exemplo anterior, os caules correspondem às classes do histograma. Repare-se que pendurámos no caule 13\* as folhas correspondentes aos valores 130, 130, 130, 132, 132, 133 134, que são precisamente os elementos da amostra pertencentes à classe [130; 135[, e assim sucessivamente, para os outros caules e as outras classes.

A situação anterior, em que os caules correspondem às classes do histograma, só acontece quando estas têm amplitude 2, 5 ou 10 vezes uma potência de 10, em que os dois primeiros casos correspondem a dividir um caule em 5 sub caules ou 2 sub caules, como no exemplo anterior. De um modo geral, mesmo que não haja esta identificação entre os caules e as classes, o aspecto gráfico apresentado pelo caule-e-folhas e pelo histograma é idêntico, dando a mesma informação sobre a forma como se distribui a variável subjacente aos dados.

De um modo geral, existe maior flexibilidade na construção do histograma, na medida em que não temos as restrições na construção das classes, que temos no caule-e-folhas para a construção dos caules, sendo esta uma grande vantagem dos histogramas relativamente aos caule-e-folhas. Por outro lado, o caule-e-folhas tem a vantagem de manter a informação existente na amostra, não havendo necessidade de construir previamente as classes, que é uma operação que introduz sempre alguma subjectividade na construção do histograma.

### Utilização do caule-e-folhas para comparar 2 amostras

O gráfico caule-e-folhas é muito útil para comparar observações que digam respeito à mesma variável, mas que tenham sido recolhidas de grupos (populações) distintos.

**Exemplo – O tempo de sono do Pedro e do David<sup>3</sup>.** A seguir apresentam-se os tempos de sono (em horas), medidos durante 30 noites seguidas, de dois jovens. Compare-os.

Pedro			David		
8.7	9.3	8.7	7.1	9.5	7.1
9.4	5.3	7.4	8.3	7.1	7.4
6.6	7.3	6.3	7.1	7.5	7.4
6.0	6.7	5.9	7.9	7.9	7.8
6.9	5.8	10.0	7.5	6.4	6.2
9.9	4.7	6.5	6.2	6.2	8.6
6.3	5.6	8.6	8.2	7.5	8.4
8.9	5.9	7.7	8.7	7.7	6.6
10.1	9.4	9.0	8.5	7.6	8.1
9.6	7.6	7.9	7.6	8.8	7.1

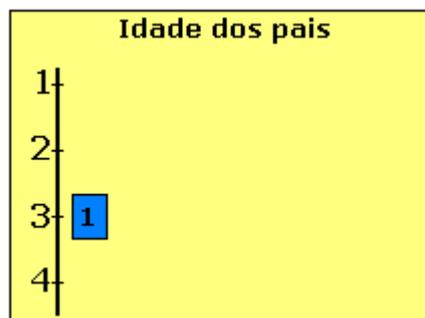
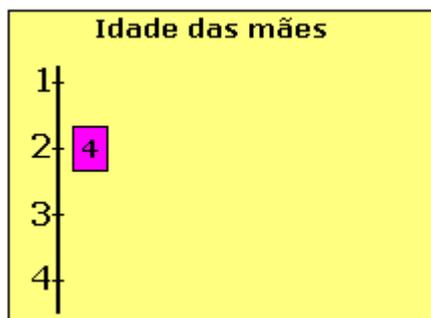
Para representar os caule-e-folhas paralelos, determinamos os caules (comuns) a partir da amostra de maior amplitude, ou seja, neste caso, dos dados correspondentes ao David.

<sup>3</sup> Adaptado de Graça Martins (2005), p. 55.

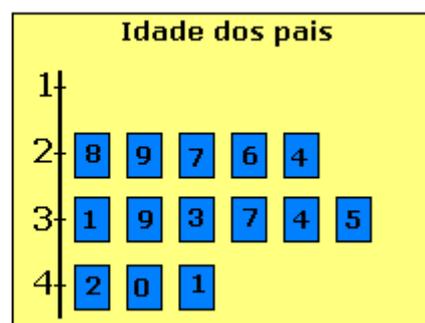
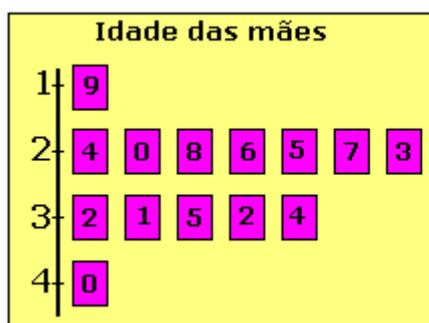


Em cada uma de duas cartolinas grandes, uma para colocar os rectângulos cor-de-rosa e outra os rectângulos azuis, desenha-se um eixo vertical e marcam-se, do lado esquerdo desse eixo os algarismos das dezenas (algarismos dominantes), que serão os caules.

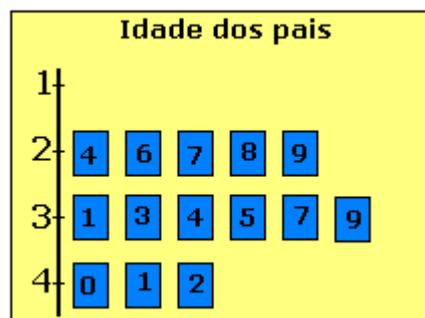
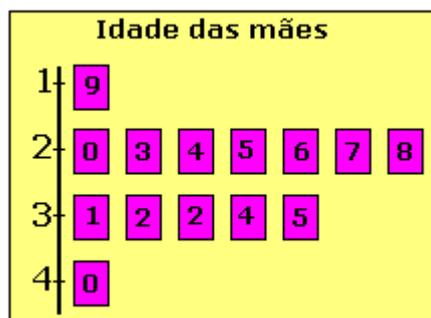
Agora cada aluno vai colocar os seus rectângulos de papel, que constituem as folhas, junto dos caules respectivos (o algarismo que ficou virado para baixo deve coincidir com o caule onde o aluno coloca o seu rectângulo de papel).



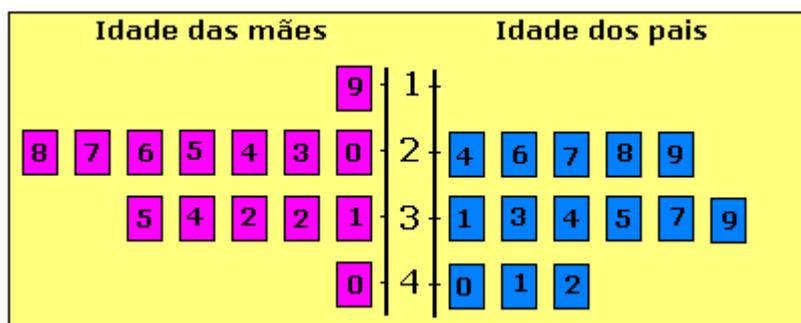
Depois de todos os alunos terem colocado os seus rectângulos de papel nos lugares devidos, obteve-se as seguintes representações em caule-e-folhas, uma com as idades das mães e outra com as idades dos pais:



Para finalizar, devem ordenar-se as folhas de cada caule:



O professor sugere agora aos alunos que em vez de andarem a colocar os rectângulos de papel em duas cartolinas separadas, juntem as cartolinas como se apresenta a seguir e coloquem os rectângulos com as idades das mães para o lado esquerdo e os rectângulos com as idades dos pais para o lado direito, sem esquecer que os menores valores são sempre os que estão mais perto dos caules:



Eis alguns exemplos de observações que podem evidenciar-se na discussão desta situação por toda a turma:

- O facto de terem escrito as idades dos pais e das mães em papéis de cor diferente, permite agora comparar facilmente estes dois conjuntos de dados;
- Se tivessem feito numa cartolina transparente o caule-e-folhas referente a um dos conjuntos de dados, por exemplo a que tem as idades das mães, bastava inverter essa cartolina transparente e fazer coincidir os caules das duas cartolinas, que se obtinha uma representação idêntica à anterior;
- As representações anteriores permitem concluir que nas mães predominam as idades à volta dos vinte e tal anos, enquanto que nos pais predominam as idades à volta dos trinta e tal anos, isto é, os pais são, de um modo geral, mais velhos que as mães;
- A mãe mais nova tem 19 anos, enquanto que a mais velha tem 40 anos;
- O pai mais novo tem 24 anos, enquanto que o mais velho tem 42 anos;
- A mediana para as idades das mães é igual a 27 anos e meio (como temos um número par de dados, não existe o elemento do meio, pelo que se somam as duas idades do meio dos dados ordenados e divide-se por 2), enquanto que a mediana para as idades dos pais é 33 anos e meio.

**Tarefa – Qual a espécie de milho mais vantajosa?**<sup>4</sup> Os cientistas desenvolveram uma nova variedade de milho, mais rica em certos aminoácidos, constituintes das proteínas, do que o milho normal. Para testar a eficácia desta nova espécie de milho na alimentação animal, foi dada uma ração com este milho, geneticamente modificado, a 20 pintainhos machos, com 1 dia de vida. A um grupo de controlo, constituído por outros tantos pintainhos, com o mesmo tempo de vida, foi dada uma ração idêntica, mas em que o milho modificado foi substituído pelo milho normal. Após 21 dias, o aumento de peso (em gramas) dos pintainhos foi:

Milho normal				Milho modificado			
380	321	366	356	361	447	401	375
283	349	402	462	434	403	393	426
356	410	329	399	406	318	467	407
350	384	316	272	427	420	477	392
345	455	360	431	430	339	410	326

Exemplo de observações que o professor pode partilhar com os alunos ao propor esta tarefa:

- Pretende-se com esta experiência testar a eficácia de um produto novo, relativamente a um produto habitual. Neste caso, pretende-se saber se o milho modificado é melhor ou pior que o milho normal;

<sup>4</sup> Adaptado de Moore (1997b), p. 234.

- A recolha de dados para este tipo de experiência pressupõe que se constituam dois grupos: a um grupo dá-se o produto que se quer testar e ao outro grupo, que se chama grupo de controlo, dá-se o produto habitual;
- Este tipo de experiência é muito utilizado quando se pretende testar a eficácia de um novo medicamento. Por vezes, ao grupo de controlo dá-se aquilo a que se chama um placebo. Um placebo é um produto aparentemente idêntico ao produto que se quer estudar, mas sem as mesmas propriedades químicas. Os indivíduos que tomam o placebo, pensam que estão a tomar o medicamento. Está provado que apesar de o placebo ser uma substância inerte, tem propriedades terapêuticas, pois o facto de um indivíduo estar a tomar uma substância que pensa que o vai curar, contribui para a cura – é o chamado “efeito placebo”. Quando se escolhem os indivíduos para os grupos que vão constituir o grupo do verdadeiro medicamento e o grupo do placebo, estes não sabem a que grupo pertencem.
- Para representar os dados anteriores, o professor deve analisar os dados com os alunos para os ajudar na escolha dos caules convenientes. Uma situação possível seria escolher para caules os dois primeiros dígitos, mas esta escolha originaria um número demasiado grande de caules: ficaríamos com mais caules do que dados! Uma outra possibilidade será escolher para caules o algarismo das centenas originando 3 caules. Mas se escolhermos para caules o algarismo das centenas, como escolher então as folhas, já que ficamos depois com dois algarismos? Tem-se duas alternativas: ou se truncam todos os dados, do último algarismo, ficando com 2 algarismos, ou se consideram folhas com 2 dígitos. A primeira alternativa é a mais utilizada, embora as duas representações finais sejam idênticas.

Algumas sugestões de representações:

Milho normal		Milho modificado
6 4 1 8 5 9 2 5 4 5 6 2 8	7 8 2 3 4	6 7 9 1 9 3 2 4 0 3 0 2 0 6 0 2 2 7 3 1

OU

Milho normal		Milho modificado
60 45 16 84 50 99 29 56 49 56 66 21 80 31 55 10 62 02	72 83 2 3 4	61 75 93 18 92 39 26 47 01 34 03 26 06 67 07 27 20 77 30 10

ou ainda, considerando na primeira das representações anteriores 2 sub caules para cada caule, já que se encontram penduradas um número exagerado de folhas num dos caules

Milho normal		Milho modificado
8 7 4 4 2 2 1 9 8 8 6 6 5 5 5 3 1 0 6 5	2. 3* 3. 4* 4.	1 2 3 6 7 9 9 0 0 0 0 1 2 2 2 3 3 4 6 7

O professor pode utilizar a última representação para pedir aos alunos que concluam da eficácia do milho modificado, no aumento do peso dos pintainhos. Pode ainda pedir que calculem a mediana dos dois grupos de dados e que comparem os valores obtidos.

#### 4.6.2 Diagrama de extremos e quartis

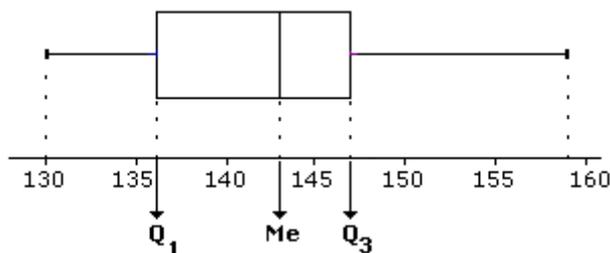
Utilizando a mediana e os quartis, que se aprendeu a determinar nas duas secções anteriores, juntamente com o mínimo e o máximo que se obtêm directamente a partir da amostra ordenada, pode construir-se uma representação gráfica muito simples, mas que evidencia de uma forma extremamente eficaz a forma como os dados se distribuem. Esta representação construída com base nestes 5 números, chama-se **diagrama de extremos e quartis** e constrói-se da seguinte forma:

1 – Desenha-se um rectângulo que tem de comprimento a amplitude entre os dois quartis, calculados a partir dos dados, e por altura um valor qualquer, que não tem qualquer interpretação;

2 – Do meio dos lados do rectângulo, perpendiculares à base, saem dois segmentos de recta que unem esses lados respectivamente com o mínimo e o máximo do conjunto dos dados.

3 – No interior do rectângulo desenha-se um traço que assinala a posição da mediana.

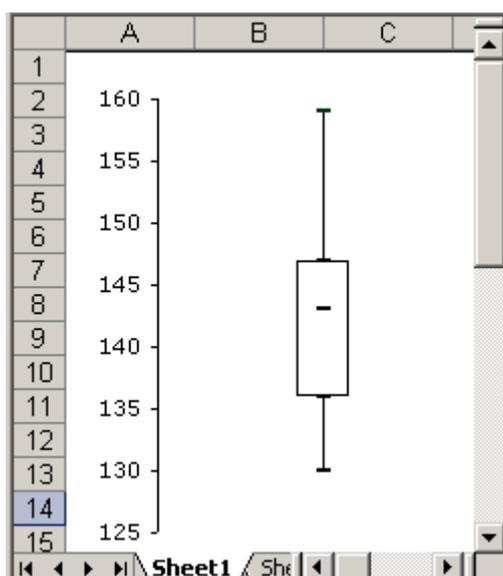
Na figura seguinte apresentamos o diagrama de extremos e quartis para o conjunto de dados da variável *Altura de um aluno de uma escola do 1.º ciclo*, de que se tem uma amostra de 50 dados, considerados no início da secção 4.5. Para estes dados já construímos, nas secções anteriores, o histograma, o gráfico de caule-e-folhas e obtivemos para a mediana e para o 1.º e 3.º quartis, respectivamente os valores 143, 136 e 147. Dos dados também se verifica que o mínimo é 130 e o máximo 159:



Da representação gráfica anterior sobressaem algumas características, nomeadamente:

- as alturas não se distribuírem de forma simétrica, tanto na parte central dos dados, como na parte mais afastada do centro;
- se os dados fossem simétricos, a mediana deveria situar-se a meio do rectângulo, o que não acontece;
- os 25% dos valores superiores também se encontram mais dispersos do que os 25% dos dados inferiores, isto é, existe uma maior variabilidade nas alturas dos alunos mais altos;

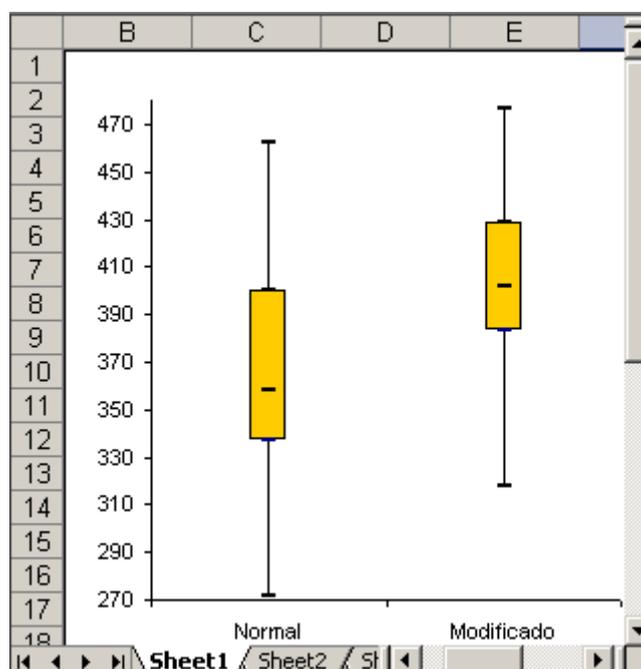
**Observação** - O diagrama de extremos e quartis apresentado anteriormente foi colocado na horizontal. No entanto também poderia ser apresentado na vertical, como na figura seguinte que se obtém quando se utiliza a folha de cálculo Excel:



### Utilização do diagrama de extremos e quartis para comparar várias amostras

Os diagramas de extremos e quartis, quando colocados em paralelo, são muito úteis para comparar 2 ou mais amostras. Fazem sobressair as semelhanças e diferenças entre a forma como os dados se distribuem, permitindo comparar a localização da mediana e dos quartis para as diferentes amostras, assim como a maior ou menor dispersão dos dados.

A seguir apresenta-se o diagrama de extremos e quartis paralelos para a tarefa proposta na secção anterior, sobre a eficácia do milho modificado:



A representação anterior torna evidente que o milho modificado produz, de um modo geral, maior aumento de peso nos pintainhos. Também se verifica que o aumento de peso apresenta menor variabilidade com este tipo de milho do que com o milho normal, o que significa que ao fim de um certo tempo a tomar a ração composta por milho modificado, os pintainhos têm um aspecto mais uniforme do

que se tivessem sido alimentados com ração composta com milho normal. Notamos ainda, sobretudo para os dados do milho normal, na parte central dos dados (nos 50% dos dados do meio da amostra ordenada), um ligeiro enviesamento para a direita, isto é, existe uma maior dispersão entre os dados compreendidos entre o 3.º quartil e a mediana, do que entre a mediana e o 2.º quartil. Se não houver inconveniente para a saúde, parece que o milho modificado é uma boa aposta!

Ao comparar várias distribuições de dados, devemos estar atentos à:

- Forma da distribuição;
- Simetria ou ausência de simetria;
- Variabilidade apresentada.

Os diagramas de extremos e quartis são particularmente úteis para compararmos a distribuição de vários conjuntos de dados, realçando aspectos particulares, como:

- Comparação das medianas;
- Comparação da dispersão entre os dados, utilizando as amplitudes entre os quartis;
- Identificação de possíveis "outliers" (valores muito grandes ou muito pequenos, relativamente aos restantes).

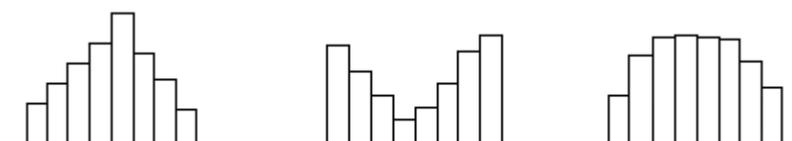
**Tarefa – Qual o número de letras do nosso nome?** Na escola os alunos vão organizar um passeio e o professor propôs que se mandassem fazer camisolas com o primeiro e último nome de cada aluno, no peito. O preço de cada camisola tem um valor fixo, ao qual é acrescido do preço do nome e este é tanto maior quanto mais letras tiver o nome. Por isso, o professor encarregou os alunos de fazerem um estudo sobre o número de letras do nome.

## 4.7 Formas frequentes de distribuição de dados

Dada uma amostra, o aspecto do histograma reflecte a forma da distribuição da população de onde os dados foram recolhidos, sugerindo-nos a escolha de um modelo teórico para essa distribuição. Ao agruparmos os dados, perdemos alguma informação contida nesses mesmos dados, mas em contrapartida obtemos informação sobre a estrutura da população que eles pretendem representar. Alguns histogramas apresentam formas que, pela frequência com que surgem, merecem referência especial e que referimos de seguida<sup>5</sup>.

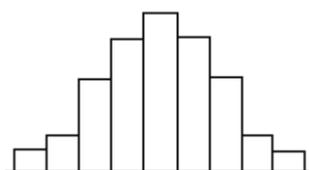
### Distribuições simétricas

A distribuição das frequências faz-se de forma aproximadamente simétrica, relativamente a uma classe média:



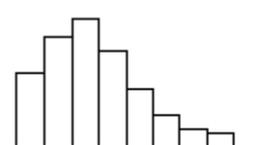
<sup>5</sup> Graça Martins (2005).

Um caso especial importante de uma distribuição simétrica é aquele que sugere a forma de um "sino". Esta distribuição surge em amostras provenientes de Populações *Normais*, termo cujo significado será explicado mais tarde, no âmbito das Probabilidades.

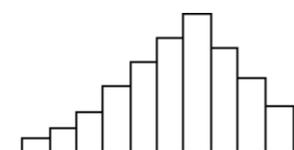


### Distribuições enviesadas

A distribuição das frequências faz-se de forma acentuadamente assimétrica, apresentando valores substancialmente mais pequenos num dos lados, relativamente ao outro:



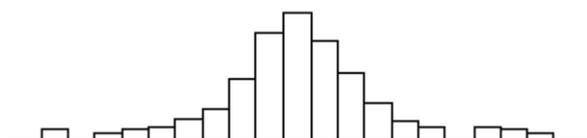
Enviesada para a direita



Enviesada para a esquerda

### Distribuições com caudas longas

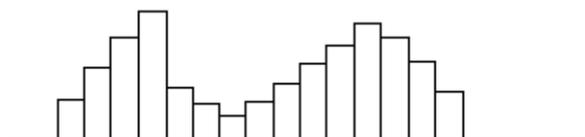
A distribuição das frequências faz-se de tal forma que existe um grande número de classes nos extremos, cujas frequências são pequenas, relativamente às classes centrais:



Uma distribuição deste tipo pode ser sintoma da existência de elementos que não pertencem à população ou que foram recolhidos de forma incorrecta (denominados "outliers").

### Distribuições com vários "picos" ou modas

A distribuição das frequências apresenta dois ou mais "picos" a que chamamos modas, sugerindo que os dados são constituídos por vários grupos distintos, ou que há uma mistura de populações com distribuições distintas:



Por exemplo, se ao representarmos os pesos de uma amostra de indivíduos, na forma de um histograma e obtivermos uma representação idêntica à anterior, temos sintomas da existência de uma mistura de duas populações distintas. Podemos, por exemplo, averiguar se não teremos uma amostra de indivíduos do sexo

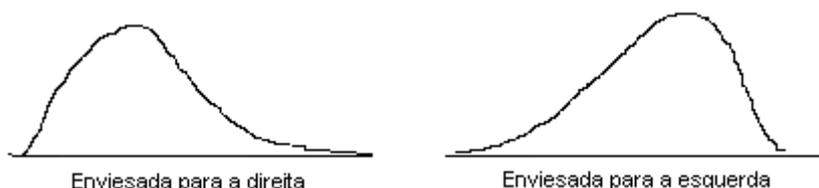
masculino ou feminino, ou até de indivíduos do mesmo sexo mas de classes etárias diferentes.

No caso das variáveis contínuas<sup>6</sup> os modelos teóricos são caracterizados pelas chamadas curvas de densidade. Estas são funções não negativas, que têm a particularidade de terem uma área unitária entre o eixo dos  $xx$  e o gráfico que as representa. Por exemplo, o seguinte gráfico



pode ser considerado a função densidade do modelo Normal, e a sua aplicação pode ser sugerida por um histograma com a forma de "sino", como o que apresentámos atrás como caso especial de uma distribuição simétrica.

Os histogramas enviesados apresentados anteriormente também sugerem, para as populações de onde as amostras foram seleccionadas, modelos com funções densidade com gráficos com o seguinte aspecto:



As distribuições com enviesamento para a direita são bem mais frequentes do que as que apresentam enviesamento para a esquerda. Uma variável que pode ser bem modelada por uma distribuição com enviesamento para a esquerda é a idade da reforma de um trabalhador. Efectivamente, a grande acumulação de idades das pessoas reformadas verifica-se à volta da classe etária dos 60 ao 70 anos. Para valores inferiores aos 60 anos há algumas pessoas que se reformam, mas com muito menor frequência. Outro exemplo de uma distribuição com um enviesamento para a esquerda é o que se obtém representando graficamente os resultados de um teste demasiado acessível para os alunos a que se destina. A maior parte dos alunos tem notas muito boas e só alguns alunos, os que não estudaram nada ou são muito fracos, é que têm uma nota reduzida. Ao contrário de um teste demasiado acessível, temos um teste demasiado difícil. Neste caso, os resultados apresentam uma distribuição com enviesamento para a direita. A representação gráfica de um teste adequado para os alunos a que se destina assemelha-se ao modelo Normal. Espera-se um número razoável de alunos com nota à volta da média das notas, com a frequência de alunos com nota alta ou baixa, a diminuir à medida que nos afastamos daquela média.

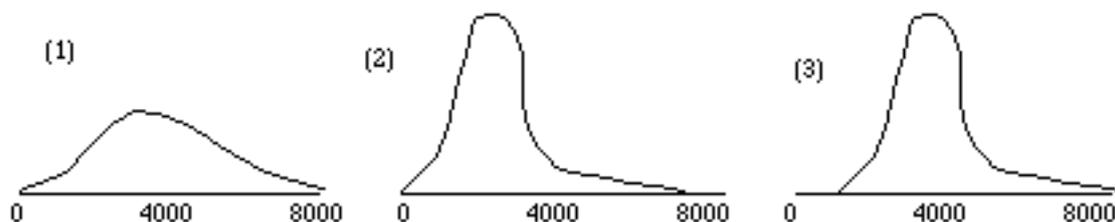
A seguir apresentamos alguns exemplos com esquemas de histogramas estilizados, que procuram traduzir a distribuição subjacente a várias variáveis quantitativas contínuas.

**Exemplo – Salários de trabalhadores**<sup>7</sup>. Recolheram-se os preços dos salários mensais de três tipos de trabalhadores. Os trabalhadores do grupo B ganham cerca de duas vezes mais do que os trabalhadores do grupo A; os trabalhadores do grupo

<sup>6</sup> Graça Martins et al. (2007).

<sup>7</sup> Adaptado de Freedman (1991).

C ganham mais 1500 euros por mês do que os do grupo A. Qual dos esquemas seguintes, de histogramas, se refere a cada um dos grupos?

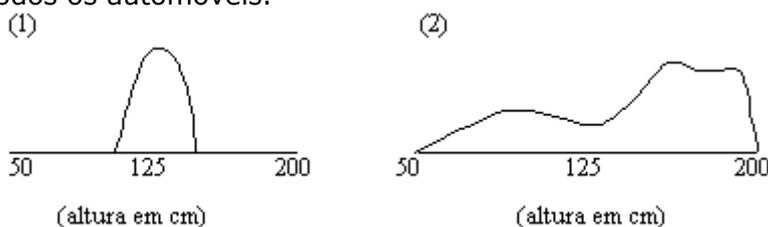


Para resolvermos esta questão, podemos pensar que se se diz que os trabalhadores do grupo B ganham o dobro dos trabalhadores do grupo A, isto significa, por exemplo, que enquanto a maior parte dos trabalhadores do grupo B auferem um salário à volta de 4000 euros, os do grupo A auferem um salário à volta de 2000 euros. Os trabalhadores do grupo C também têm um salário claramente superior aos do grupo A. Então é natural esperar que a figura (2) corresponde aos trabalhadores do grupo A, pois é a única cuja média é claramente inferior a 4000. Por outro lado, se os trabalhadores do grupo C ganham 1500 euros a mais do que os do grupo A, isto significa que a distribuição dos salários dos trabalhadores do grupo C terá um aspecto idêntico ao dos trabalhadores do grupo A, mas deslocada para a direita de 1500 euros. Então a figura (3) corresponderá aos salários dos trabalhadores do grupo C. Por exclusão de partes a figura (1) deve corresponder aos salários dos trabalhadores do grupo B, sendo de facto compatível com a condição dada do seu salário ser cerca de duas vezes maior que o dos trabalhadores do grupo A.

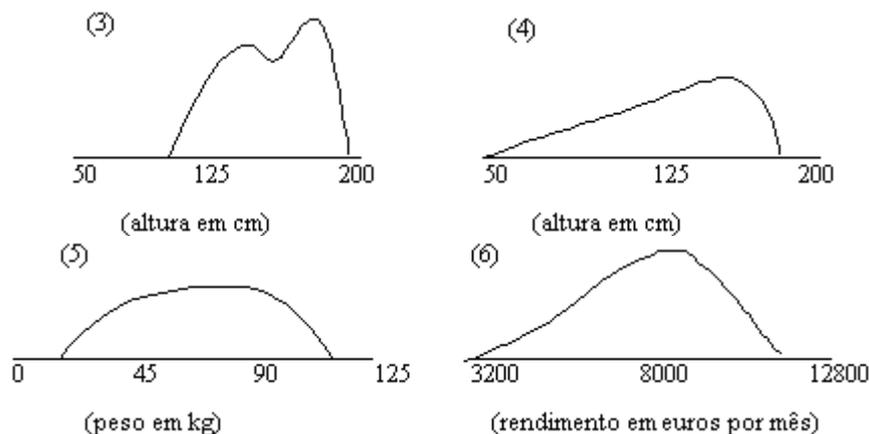
A distribuição com o aspecto (1) não é muito usual para representar salários, sendo mais usuais as distribuições com o aspecto (2) ou (3). Efectivamente, em geral, a distribuição dos salários tem um aspecto assimétrico, com um enviesamento para a direita. Isto deve-se ao facto de a maior parte dos salários se concentrarem numa determinada região, havendo alguns (poucos) salários que são substancialmente superiores aos restantes, provocando uma cauda da distribuição, alongada para a direita.

**Exemplo – Qual o aspecto da distribuição?**<sup>8</sup>. Seguidamente apresentam-se seis esquemas de histogramas, quatro dos quais apresentam os resultados do estudo, numa pequena cidade, das quatro características seguintes:

- Alturas de todos os elementos das famílias, em que os pais tenham idade inferior a 24 anos.
- Alturas dos casais (marido e mulher).
- Alturas de todos os indivíduos da cidade.
- Alturas de todos os automóveis.

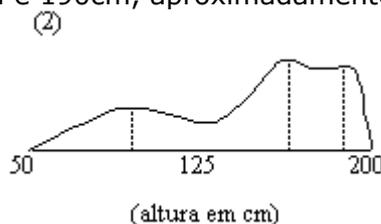


<sup>8</sup> Adaptado de Freedman (1991).



Quais dos esquemas de histogramas podem representar cada uma das variáveis anteriores?

Pensando na variável que representa a altura de um elemento, escolhido ao acaso, de uma família, em que os pais tenham idade inferior a 24 anos, esperamos obter um histograma com uma mancha idêntica à (2), onde se vislumbram 3 pontos, à volta dos quais se nota uma maior frequência, e que corresponderão à altura dos filhos – entre 80 e 90cm, que para casais com idades inferiores a 24 anos, ainda devem ser muito pequenos, e à altura dos membros do casal ou do marido, respectivamente à volta de 165cm e 190cm, aproximadamente:



Quando consideramos a distribuição das alturas dos elementos de um casal, é natural esperar um esquema idêntico ao da figura (3), com duas modas, reflectindo que de um modo geral, as alturas dos homens concentram-se em torno de um valor um pouco superior ao valor em torno do qual se concentram as alturas das mulheres.

Ao escolher um indivíduo ao acaso, na cidade, esperamos que a distribuição das alturas seja descrita pela figura (4) que apresenta um enviesamento para a esquerda, correspondente às alturas das crianças, que estão em minoria.

Finalmente, quando se estuda a variável que representa a altura de um carro, o histograma adequado é o que corresponde à mancha (1) que traduz o facto de os carros terem quase todos a mesma altura, andando à volta de 125cm.

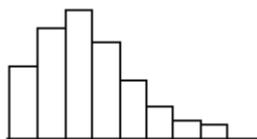
### Caule-e-folhas, histograma ou diagrama de extremos e quartis

Já na secção 4.6.1, quando apresentámos a representação dos dados em caule-e-folhas, frisámos o facto de a informação transmitida pelo histograma sobre o padrão da distribuição da população subjacente aos dados, ser idêntica à transmitida pelo gráfico de caule-e-folhas. Adiantamos que o mesmo se verifica com o diagrama de extremos e quartis. Por exemplo, as seguintes representações, obtidas para o mesmo conjunto de dados, dão o mesmo tipo de informação, sugerindo que a distribuição da população tem um enviesamento para a direita:

```

0 | 0 1 3 6 7 8
1 | 1 1 2 3 5 7 8 8 9 9
2 | 0 1 3 4 4 5 6 7 7 8 9
3 | 4 4 5 6 6 8 8 9
4 | 1 1 2 3 4 4 5
5 | 2 2 3 7
6 | 3 6 7
7 | 1 5
8 | 9
9 | 5

```



Como já referimos, quando se faz a representação dos dados, perde-se sempre alguma informação que eles contêm, mas em contrapartida obtemos informação sobre a estrutura da população de onde eles provêm. Das representações gráficas anteriores, aquela em que se perdeu mais informação foi o diagrama de extremos e quartis, mas também foi a mais simples de ser construída – bastou recolher, a partir dos dados, informação sobre cinco números (mínimo, máximo, 1.º quartil, 3.º quartil e mediana). Ao construir o histograma também perdemos alguma da informação contida nos dados, uma vez que os agrupámos em classes, mas em contrapartida ficámos com uma ideia do padrão da distribuição subjacente aos dados. A representação em que se perde menos informação é o caule-e-folhas, mas como também já vimos, devido à especificidade da escolha dos caules, a sua construção pode apresentar menos maleabilidade do que o histograma.

#### 4.8 Representações gráficas e tabelas de frequências para dados bivariados

Embora a representação de dados bivariados não faça parte do programa do ensino básico, pensamos ser oportuno a introdução de uma breve referência à representação deste tipo de dados. Uma representação gráfica frequentemente utilizada e com grande divulgação na comunicação social, o **gráfico de linha (ou série temporal)**, que abordaremos a seguir, não é mais do que uma representação gráfica da forma como duas variáveis se relacionam uma com a outra, mas numa situação especial, em que uma das variáveis é o tempo.

No nosso dia-a-dia temos muitas vezes necessidade de recolher informação sobre duas (ou mais) variáveis acerca do mesmo indivíduo. Por exemplo:

- quando os pais levam o filho ao pediatra ele recolhe informação sobre o peso e a altura da criança;
- quando um adulto vai ao médico, este recolhe informação sobre a idade, o peso, o nível de colesterol, a tensão máxima e mínima, etc.;
- quando alguém pretende comprar um carro, recolhe informação sobre a cilindrada e o consumo médio aos 100km;
- a um aluno candidato à Universidade pede-se a nota de candidatura e a nota na prova específica (esta prova varia de Faculdade para Faculdade);
- em estudos económicos estuda-se o salário de um indivíduo e o nível de escolaridade que detém;
- o governo interessa-se pela evolução da taxa de desemprego, ao longo dos últimos anos;

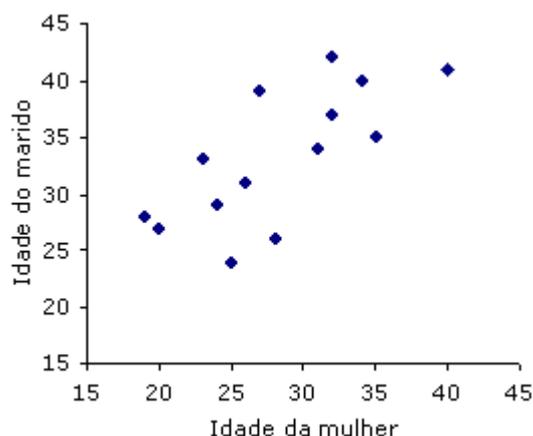
- para impor uma taxa máxima de álcool para os condutores, o governo encomendou um estudo sobre a associação entre a taxa de álcool e o tempo de reacção;
- quando troveja, associamos a distância a que se encontra a trovoadas, com o tempo que demora o trovão, após vermos o relâmpago;
- etc.

#### 4.8.1 Diagrama de dispersão

Nos casais, espera-se que a idade da mulher esteja relacionada com a idade do marido. Será verdade que quanto mais velha é a mulher, mais velho é o marido? Propomos fazer um estudo sobre esta questão, pelo que vamos utilizar a informação sobre as idades da mulher e do marido da **Tarefa Vamos comparar as idades dos nossos pais e das nossas mães**, da secção 4.6.1. Das representações aí consideradas, já concluímos que os homens tendem a ser mais velhos que as mulheres, mas na verdade nem todos os maridos são mais velhos do que as mulheres, nem tão pouco conseguimos saber qual a percentagem de homens que são mais velhos do que as respectivas mulheres. Efectivamente, com os dados relativos aos pais e às mães separados, nada podemos concluir. Uma solução que o professor arranjou, para em qualquer altura recuperar os pares de dados trazidos, pelos alunos, nos dois pedaços de papel, foi numerar (na parte de dentro do papel) com o mesmo número, o par de pedaços de papel dado a cada aluno. Assim, conseguiu-se reconstituir as idades dos casais, que se apresentam na tabela seguinte:

Idade da mulher	19	24	20	28	26	25	27	23	32	31	35	32	34	40
Idade do marido	28	29	27	26	31	24	39	33	37	34	35	42	40	41

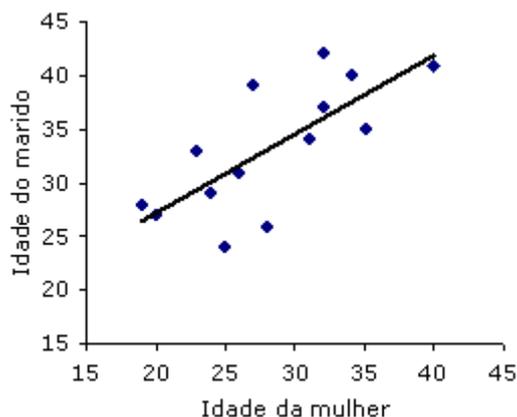
Representámos num sistema de eixos coordenados os pontos de coordenadas (Idade da mulher, Idade do marido), como se apresenta a seguir:



Na representação anterior, a que chamamos **diagrama de dispersão**, apercebemo-nos que, de um modo geral, à medida que a idade da mulher aumenta, também aumenta a idade do marido. Dizemos de um modo geral, porque nem sempre isso aconteça, mas existe *tendência* a que homens mais velhos estejam casados com mulheres mais velhas.

**Diagrama de dispersão** – é uma representação gráfica para dados bivariados (pares de dados) quantitativos, em que cada par de dados  $(x, y)$  é representado por um ponto de coordenadas  $(x, y)$ , num sistema de eixos coordenados.

Este tipo de representação é muito útil, pois permite realçar algumas propriedades entre os dados, nomeadamente no que diz respeito ao tipo de associação entre as variáveis representadas por  $x$  e  $y$ . No exemplo anterior, a nuvem de pontos, embora um pouco dispersa, apresenta uma forma alongada, que pode ser representada por uma recta com declive positivo:



Quanto mais perto os pontos se dispuserem ao longo de uma recta, maior será o grau de associação entre as duas variáveis. Essa associação diz-se *positiva*, se a recta tiver declive positivo. O exemplo anterior é um caso de uma associação positiva. A associação será *negativa*, se a recta tiver declive negativo. Neste caso, quanto maior for o valor de uma das variáveis, menor será, de um modo geral, o valor da outra variável.

**Tarefa – Vamos comer queijo, mas não exageremos...**<sup>9</sup>. O queijo, proveniente do leite, é um alimento rico em cálcio. No entanto, é necessário não abusar, já que, de um modo geral, é um alimento muito calórico e a maior parte das vezes rico em gordura. Na tabela seguinte apresentamos, para vários tipos de queijo, a quantidade de gordura e o número de calorias, por cada 100 gramas de queijo:

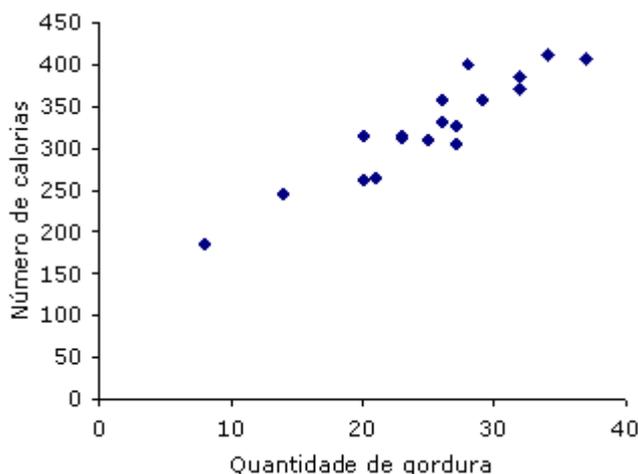
Alimento (100g)	Gordura (g)	Calorias
■ Queijo Brie	20	263
▲ Queijo Camembert	23	313
▲ Queijo da Ilha	26	357
▲ Queijo da Serra curado	32	385
▲ Queijo da Serra fresco	27	327
▲ Queijo de Azeitão	25	309
▲ Queijo de Évora	34	412
▲ Queijo de Serpa	26	330
▲ Queijo de Tomar	27	305
● Queijo flamengo 20%	8	185
■ Queijo flamengo 30%	14	246
▲ Queijo flamengo 45%	23	315
■ Queijo fresco	21	265

<sup>9</sup> Adaptado de Graça Martins et al. (2007).

▲ Queijo Gorgonzola	37	407
■ Queijo Gruyère	20	315
▲ Queijo Parmesão	28	401
▲ Queijo Roquefort	32	371
▲ Queijo Suíço	29	357

- - Alimento com baixo teor em gordura mas podendo ter um elevado conteúdo em calorias.
- - Alimento intermediário: consumir com moderação.
- ▲ - Alimento rico em gordura: comer pontualmente ou moderar o seu consumo.

A tabela anterior permite vários estudos no que diz respeito à quantidade de gordura e ao número de calorias dos diferentes tipos de queijo. Uma possível abordagem é começar por tentar relacionar as duas variáveis Quantidade de gordura (em gramas) e Número de calorias. O diagrama de dispersão para estas variáveis tem o seguinte aspecto:



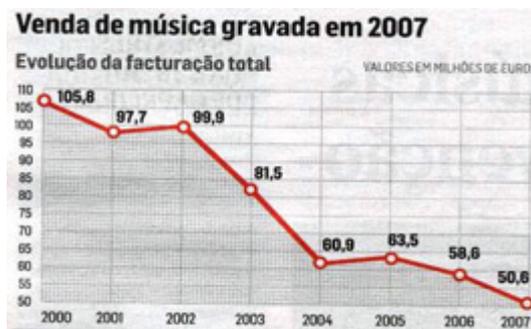
No gráfico anterior é nítida a forte associação linear positiva entre a *Quantidade de gordura* e o *Número de calorias*, como aliás seria de esperar. Este exemplo, que será abordado mais à frente, pode ser aproveitado pelo professor para discutir com os alunos sobre os malefícios de ingerir alimentos com grande quantidade de gordura e já que o queijo faz bem, pois tem cálcio, necessário para uma boa formação dos ossos, devemos escolher os tipos de queijo com menos gordura, que também têm, de um modo geral, menos calorias.

#### 4.8.2 Gráfico de linhas

Um gráfico de linhas é um caso especial de um diagrama de dispersão. É utilizado para representar, visualmente, a forma como uma variável evolui em relação a outra variável, sendo esta outra variável, quase sempre, o tempo. Por exemplo, repare no gráfico que fez parte do Desafio 23 do ALEA ([www.alea.pt](http://www.alea.pt)) e que foi retirado de uma notícia que apareceu no jornal *Diário de Notícias* do dia 5 de Abril de 2008.

Nesse desafio apresentou-se o gráfico tal e qual como apareceu na comunicação social e fazia parte do desafio responder às questões que se acrescentam a seguir:

O gráfico apresenta a evolução da facturação total, em milhões de euros, do mercado discográfico português de 2000 a 2007. A partir da informação contida no gráfico, pede-se que se responda a algumas questões.



Fonte: AFP

**Questão 1:** Embora a tendência da evolução da facturação seja nitidamente decrescente, houve alguns anos em que se verificou um ligeiro crescimento. **Entre que anos consecutivos** se registou esse crescimento na venda de música gravada? Qual o **valor do crescimento**, em **percentagem**? Apresenta o resultado aproximado às décimas.

**Questão 2:** Na notícia afirma-se que o mercado português da música gravada facturou, em 2007, menos 13,7% que em 2006. De acordo com o gráfico, esta afirmação é verdadeira? Justifica a tua resposta.

**Questão 3:** De 2000 para 2007, qual o decréscimo, em percentagem, verificado na facturação discográfica? Apresenta o resultado aproximado às décimas.

Este exemplo pode ser trabalhado com os alunos, com o auxílio do professor, quando aqueles estiverem a estudar e a interpretar a variação de uma função representada por um gráfico.

**Tarefa – As vendas estão a correr bem?** A proprietária de uma livraria, montou, num canto da sala, uma máquina de café, a título de experiência. Os lucros com a venda dos livros estavam a baixar, de modo que a venda do café talvez ajudasse a equilibrar o negócio. Passado algum tempo, decidiu averiguar se o negócio com a máquina de café era compensador. Assim, pediu a um grupo de alunos do 3.º ciclo, da escola ao lado da livraria, que costumavam passar por lá para folhear uns livros, para lhe fazerem um pequeno estudo sobre se valeria a pena continuar com o negócio. Os jovens decidiram ajudar a senhora. Depois de trocarem algumas impressões uns com os outros de como atacar o problema, uma coisa não tinham dúvidas: precisavam de dados! Felizmente a dona da livraria tinha registado as quantias auferidas com a venda do café, nos últimos 20 dias. Os jovens registaram os dados, tendo a senhora garantido que a ordem apresentada, era a ordem pela qual os dados tinham sido recolhidos:

300, 100, 200, 300, 100, 200, 200, 300, 300, 300, 200, 300, 400, 300, 300, 400, 500, 400, 400, 500

Na posse dos dados, começaram a pensar na metodologia a seguir, de forma a extrair alguma informação que pudesse ajudar a proprietária. Nessa discussão, alguns dos alunos decidiram calcular algumas medidas, fazer algumas representações gráficas, ou seja, tentar arranjar alguns processos úteis de conseguir que eles “falassem”, pois aquele conjunto de valores não lhes estava a dizer nada...

Já todos sabiam calcular a moda e a média, pelo que começaram por aí. No entanto, um deles alertou para o facto de reduzir um conjunto de dados a 2 medidas era talvez demasiado drástico, pelo que decidiram fazer duas representações gráficas, nomeadamente uma representação em caule-e-folhas e um diagrama de dispersão com os pares (dia, quantia auferida). Ao tomarem esta decisão, tentaram realçar alguns pontos, tais como:

- Alguns aspectos dos dados apresentados pelo caule-e-folhas, que não sobressaíam a partir do diagrama de dispersão;
- Alguns aspectos dos dados apresentados pelo diagrama de dispersão, que não sobressaíam a partir do caule-e-folhas;
- Qual das representações interessaria mais à proprietária da livraria.

Apresentamos a seguir um pequeno relatório com a análise dos dados fornecidos pela proprietária da livraria:

#### *Relatório*

*Questão* – Foi-nos pedido que elaborássemos um estudo, para averiguar se a venda de café na livraria estaria a resultar. A proprietária da livraria, forneceu-nos uma tabela com as vendas nos 20 últimos dias, não nos tendo fornecido mais nenhuma informação, nomeadamente os custos com a manutenção da máquina, os custos do grão de café, ou com a mão-de-obra envolvida neste pequeno negócio.

*Metodologia utilizada* – Cálculo de algumas estatísticas:

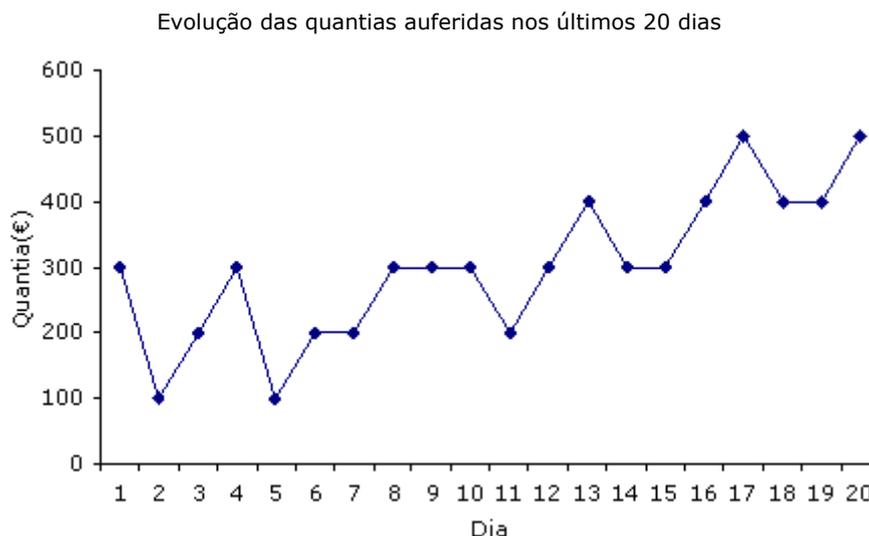
Moda	Média	Mediana
300 €	300€	300€

O facto de as 3 medidas anteriores serem iguais, levou-nos a desconfiar que a distribuição dos dados era simétrica, o que se veio a confirmar com o gráfico de caule-e-folhas, que apresentamos a seguir:

10		0	0					
20		0	0	0	0			
30		0	0	0	0	0	0	0
40		0	0	0	0			
50		0	0					

A distribuição dos dados é simétrica, razoavelmente concentrada em torno da média de 300 euros, não se tendo registado quantias muito baixas ou muito elevadas.

Como a representação anterior não nos diz nada sobre a forma como o negócio está a evoluir, construiu-se um gráfico de linhas, que se apresenta a seguir:



Este gráfico é bem elucidativo ao mostrar que o negócio evolui de forma positiva, com tendência para crescer. Esta é uma característica importante, que não era realçada no caule-e-folhas, mas que naturalmente vai interessar à dona da livraria.

*Conclusão* – Tendo em linha de conta a informação que nos foi facultada, concluímos que com a venda do café a proprietária auferes uma quantia média diária de 300 euros, havendo uma tendência para este valor aumentar.

**Tarefa – Registo da temperatura máxima e mínima no Porto, Lisboa, Faro, Ponta Delgada e Funchal.** Pretende-se averiguar se haverá indícios de diferença na temperatura nas três cidades apontadas de Portugal continental, uma do norte, outra do centro e outra do Sul e das duas cidades dos arquipélagos dos Açores e Madeira, respectivamente. Para poderem responder a esta questão, os alunos resolveram fazer uma recolha de dados durante 20 dias.

Para fazerem a tarefa, o professor arranhou cinco folhas quadriculadas, uma para cada cidade, como a que se apresenta a seguir. Pendurou na parede, com fita-cola e colocou ao pé das folhas dois lápis, um azul e outro encarnado. Os alunos foram divididos em grupos de dois e cada grupo ficou encarregue de ouvir no noticiário da manhã, num dia especificado pelo professor, sem esquecer os fins-de-semana, a previsão das temperaturas máxima e mínima para as cinco cidades. Quando os alunos chegavam à turma, apontavam com o lápis azul a temperatura mínima e com o lápis encarnado a temperatura máxima, de cada cidade, na folha respectiva.

Ao fim dos 20 dias, completam o gráfico de linha unindo os pontos azuis e os pontos encarnados. Algumas questões que podem ser abordadas:

- Qual o valor máximo e mínimo obtido para as temperaturas recolhidas, para cada cidade? No caso das cidades de Portugal Continental, estes valores estarão associados à localização geográfica das cidades?
- Para cada cidade calcula, para cada dia, a diferença entre a temperatura máxima e a temperatura mínima. Utilizando uma representação gráfica adequada, compara os cinco conjuntos de dados obtidos.

Se a escolha de dados tivesse sido feita noutra estação do ano, pensas que obterias dados com aspecto diferente? Haverá alguma(s) das cidades onde a estação do ano tenha menor influência do que noutra(s) cidades? Justifica a tua resposta.

## 5 Características amostrais

Uma descrição numérica das distribuições de dados é feita através de alguns números que realçam alguns aspectos específicos da distribuição dos dados, nomeadamente no que diz respeito à localização de alguns pontos importantes, como o centro da distribuição, ou à dispersão ou variabilidade apresentada pelos dados.

---



## 5.1 Introdução

Vimos, no capítulo anterior, alguns processos de resumir a informação contida nos dados, utilizando tabelas e gráficos. Veremos, neste capítulo, um outro processo de resumir essa informação utilizando determinadas **medidas**, na forma de números, calculadas a partir dos dados e que servem para os caracterizar, que se chamam **estatísticas**. Das medidas ou estatísticas que iremos definir, destacam-se as **medidas de localização**, nomeadamente as que localizam o centro da distribuição dos dados, também chamadas medidas de *tendência central*, e as **medidas de dispersão**, que medem a variabilidade dos dados.

Observemos que, ao resumir a informação contida nos dados na forma de alguns números, estamos a proceder a uma redução "drástica" desses dados. Assim, aquelas medidas devem ser convenientemente escolhidas, de modo a representarem o melhor possível o conjunto de dados que pretendem sumarizar. Definiremos várias medidas, mas notamos desde já que não se pode dizer que uma é melhor do que outra, de uma forma geral. Cada uma delas tem as suas vantagens e os seus inconvenientes e a escolha da medida a usar depende do contexto e da situação em causa.

No 1.º ciclo só se estuda uma destas medidas – a moda. No 2.º ciclo surgem os conceitos de média aritmética, extremos e amplitude. E, finalmente, no 3.º ciclo, os alunos aprendem a mediana, os quartis e amplitude interquartil, devendo ser capazes de escolher as medidas de localização mais adequadas para resumir a informação contida nos dados. O estudo destas medidas não deve redundar na realização de exercícios de cálculo repetitivos, em que a própria natureza dos dados nem chega a ser bem discutida, e, frequentemente, se perdem de vista as questões a que se queria responder. Pelo contrário, é na exploração de situações significativas para os alunos – com dados fornecidos pelo professor como os que se apresentam neste capítulo ou com dados recolhidos pelos próprios alunos – que estes conceitos devem ser considerados, aproveitando-se, para isso, as oportunidades relevantes. Como complemento de informação para o professor, para além das medidas de localização e dispersão incluídas no *Programa*, apresentamos ainda neste capítulo algumas medidas adicionais (percentis, desvio médio absoluto e desvio-padrão).

## 5.2 Medidas de localização

Será mesmo necessário utilizar os dois tipos de medidas, isto é de localização e de dispersão, para caracterizar um conjunto de dados? O exemplo seguinte procura responder a esta questão.

Suponha que dois alunos do 7.º ano obtiveram as seguintes notas no 3.º período:

Pedro	4	3	3	3	3	3	4	3	4	3
João	5	2	2	3	4	3	5	3	3	3

O Pedro e o João tiveram a mesma média de 3.3, mas o João não transitou de ano, pois teve duas negativas. Quer dizer que utilizámos uma medida de redução dos dados, a **média**, que não é suficiente para caracterizar e diferenciar os dois conjuntos de dados. Efectivamente, se representarmos num diagrama de caule-e-folhas os dois conjuntos, obtemos duas representações com aspecto diferente, já que na

segunda representação se verifica uma maior *variabilidade*, isto é, os dados estão mais dispersos:

3	3 3 3 3 3 3 3	2	2 2
4	4 4 4	3	3 3 3 3 3
		4	4
		5	5 5

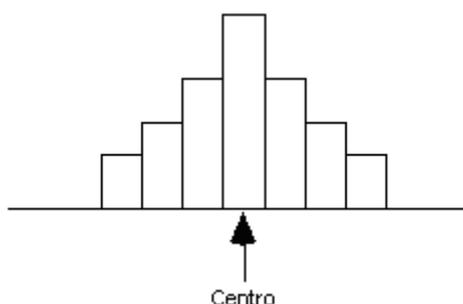
Para definir as medidas que vão ser utilizadas para resumir a informação contida nos dados, utilizamos a seguinte notação para representar os dados

$$x_1, x_2, x_3, \dots, x_n$$

onde  $x_1, x_2, \dots, x_n$ , representam, respectivamente, a 1.<sup>a</sup> observação ou 1.<sup>o</sup> dado, a 2.<sup>a</sup> observação ou 2.<sup>o</sup> dado, a n-ésima observação ou n-ésimo dado, a serem consideradas ou considerados para constituir a amostra de dimensão **n**. Esta notação não pressupõe uma ordenação.

*Uma medida de localização* é um número que nos dá informação sobre a ordem de grandeza dos dados da amostra, não só da parte central da distribuição dos dados, onde se concentram a maior parte dos dados, como também nas caudas.

De entre as medidas de localização, merecem destaque especial as que indicam o *centro de uma amostra*. Vimos no capítulo anterior que uma representação gráfica adequada para um conjunto de dados contínuos era, por exemplo, o histograma. Vimos também que um histograma pode ter vários aspectos, apresentando, nomeadamente, uma forma simétrica ou enviesada. No caso particular do histograma ser perfeitamente simétrico, não há dúvida em dizer qual o centro dessa distribuição:



No entanto, a situação anterior, a existir, é muito rara, pois devido à aleatoriedade presente nos dados, os histogramas não apresentam aquele aspecto. Por outro lado, quando o histograma é enviesado, a situação ainda se torna mais complicada, pois é difícil dizer o que é o centro. Existem então, vários processos para definir o centro, cujas medidas não dão necessariamente o mesmo resultado. Destas medidas destacamos a média e a mediana, a definir seguidamente.

Apresentaremos ainda outras medidas de localização, não para representarem o centro da distribuição dos dados, mas sim outros pontos representativos da aglomeração das caudas, de que destacamos os quartis.

### 5.2.1 Média

A média amostral, ou simplesmente *média*, é a medida de localização do centro da amostra mais vulgarmente utilizada. Representa-se por  $\bar{x}$  e calcula-se utilizando o seguinte processo:

- Somam-se todos os elementos da amostra;
- Divide-se o resultado da soma pelo número de elementos da amostra.

Por exemplo, para os dados referentes à variável *número de letras do nome*, da Turma de referência, vem que a média do número de letras dos nomes dos primeiros 10 alunos considerados na tabela,

Nome	Número de letras no nome
<b>Ana Godinho</b>	<b>10</b>
<b>Ana Sofia Silva</b>	<b>13</b>
<b>Andreia Sousa</b>	<b>12</b>
<b>Carolina Martins</b>	<b>15</b>
<b>Daniela Silva</b>	<b>12</b>
<b>David Leal</b>	<b>9</b>
<b>Diogo Oliveira</b>	<b>12</b>
<b>Filipa Duarte</b>	<b>12</b>
<b>Helena Afonso</b>	<b>12</b>
<b>Inês Martins</b>	<b>11</b>
Joana Manso	10
João Miguel Ribeiro	17
João Pedro Batista	16
Liliana Isabel Cruz	17
Maria Margarida Cabral	20
Miguel Esteves	13
Nuno Pestana	11
Patrícia Santos	14
Pedro Pinheiro	13
Raquel Loureiro	14
Rita Martins	11
Simão Valente	12
Sofia Matias	11
Tiago Neves	10

é igual a

$$\frac{10 + 13 + 12 + 15 + 12 + 9 + 12 + 12 + 12 + 11}{10} = 11,8$$

O que significa uma média de 11,8 letras para o número de letras dos nomes dos 10 alunos? Obviamente que não há 11,8 letras!

O que aquele valor significa é que 12 letras nos dão um valor que representa razoavelmente bem o número de letras dos nomes daqueles 10 alunos, isto é, se pretendêssemos distribuir equitativamente as 118 letras dos nomes dos 10 alunos, dando a cada um o mesmo número de letras, ou um número aproximado de letras, esse valor andaria à volta de 12 letras. Neste caso não poderíamos dar 12 letras a cada um dos 10 alunos, pois seriam necessárias 120 letras, mas poderíamos dar 12 letras a 8 dos alunos e 11 aos 2 restantes e ficavam todos com um número igual ou aproximado de letras.

Se em vez do número de letras dos nomes, aqueles valores significassem a quantidade, em euros, que cada aluno tinha no bolso, e pretendêssemos calcular a média

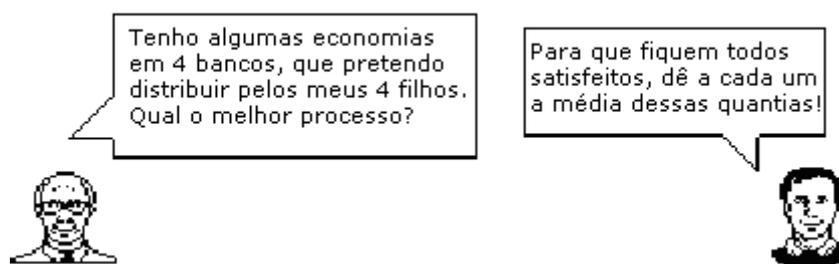
das quantias que os 10 alunos tinham nos bolsos, a interpretação do valor 11,8 euros já não traria qualquer problema, pois se quiséssemos distribuir os 118 euros equitativamente pelos 10 alunos, seria possível dar a cada um a mesma quantia, ou seja 11 euros e 80 cêntimos.

A média é uma estatística largamente utilizada no dia a dia. Quando um empregador pergunta a um candidato a um emprego, qual a média da sua licenciatura, quer avaliar, a partir de um único número, todo o percurso escolar desse candidato como aluno (estamos a admitir que a média da licenciatura é a média aritmética simples, como a que foi aqui apresentada, o que nem sempre acontece). A média é, aliás, uma estatística com que o aluno se confronta durante todo o percurso escolar!

Pode dizer-se que a média é o ponto de equilíbrio de todos os elementos da amostra, na medida em que equilibra os valores grandes com os pequenos. Esta característica pode-se tornar uma vantagem em certas situações:

- Quando o que se pretende representar é a *quantidade total expressa pelos dados*, utiliza-se a média.

Na realidade, ao multiplicar a média pelo número total de elementos, obtemos a quantidade pretendida!



Utilizando a notação introduzida anteriormente para representar a amostra, a média obtém-se a partir da expressão:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

### A média será sempre uma medida representativa dos dados?

Não, a média nem sempre é uma boa medida para representar os dados. O seguinte exemplo é elucidativo do que acabamos de dizer.

Suponha um aluno que ao longo do ano fez 6 testes, em que teve negativa em 5 deles e uma positiva, no primeiro teste, que era muito simples:

19    8,9    7,8    9,4    8,6    9,3

Pois este aluno teve uma média positiva, igual a 10,5, ou seja 11!

Embora todas as notas, menos uma, estejam no intervalo [7,8; 9,4], o valor obtido para a média não reflecte o conjunto das notas do aluno! Uma medida que se pretendia representativa dos dados, não está a conseguir esse objectivo, pois se nos disserem que um conjunto de dados tem média 10,5, imediatamente pensamos em

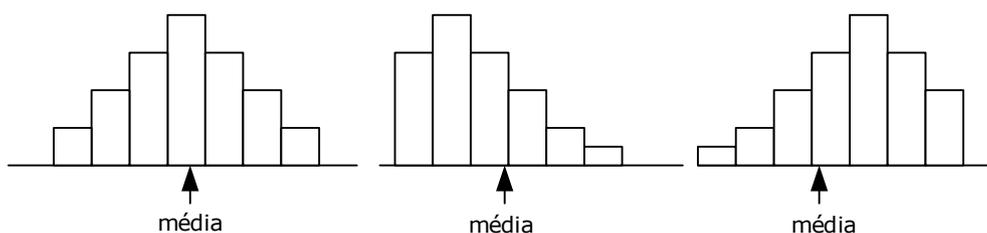
valores que não se afastam muito deste valor, uns menores e outros maiores, numa proporção aproximada.

O que acontece é que a *média é muito sensível a valores muito grandes ou muito pequenos*, vulgarmente chamados de “outliers”, dizendo-se por isso que é uma medida pouco *resistente*. A pouca resistência vem precisamente do facto de ser muito influenciada e “não resistir” a estes valores, mesmo que existam em pequena quantidade, quando comparados com todos os restantes valores.

No caso do exemplo foi o valor 19 que inflacionou a média. Além disso, temos alguma razão para pensar que o aluno efectivamente não deveria ter média positiva, pois só teve uma boa nota no primeiro teste, ainda por cima muito simples, como é afirmado.

- Sendo a média uma medida tão sensível aos dados, é preciso ter cuidado com a sua utilização, pois pode dar uma imagem distorcida dos dados que pretende representar!

Efectivamente a média constitui um bom resumo dos dados nos casos em que estes se distribuem de forma aproximadamente simétrica, com uma zona central de maior concentração e caudas que não se alonguem demasiado. Esquemáticamente podemos posicionar a média da forma que se segue, tendo em conta a representação gráfica na forma de histograma:

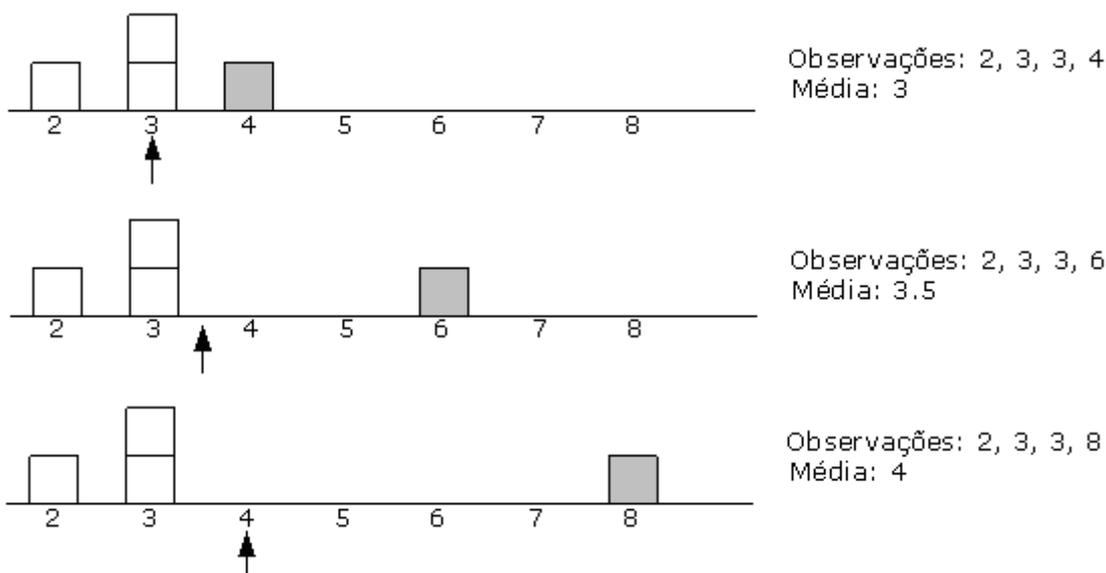


No histograma do lado esquerdo temos uma figura aproximadamente simétrica, pelo que o centro está bem definido. No histograma do centro o enviesamento para a direita provoca uma deslocação da média para a direita; finalmente no histograma da direita o enviesamento provoca uma deslocação da média para a esquerda.

Quando a distribuição dos dados apresenta um grande enviesamento, tem pouco interesse utilizar a média como centro da distribuição dos dados. Aliás, quando a distribuição dos dados não for aproximadamente simétrica é o próprio conceito de “centro da distribuição” que deixa de ter sentido, pelo que nenhuma medida de tendência central tem interesse como resumo da informação contida nos dados.

**Exemplo da “pouca resistência” da média** – Considerando os valores 2, 3, 3 e 4, construímos um diagrama de barras e posicionámos a média. De seguida, alterámos um desses valores para estudar o seu comportamento.

É interessante verificar que um diagrama de barras (ou histograma) se comporta como um balancé, em que o ponto de apoio é a média. Ao contrário da mediana, como se verá adiante, a percentagem de elementos para um e outro lado da média não é necessariamente igual a 50%.



Repare-se como varia a média, à medida que se altera um dos dados. Para restabelecer o equilíbrio entre o valor que está a aumentar e os restantes valores, a média também está a aumentar.

### Pode-se sempre calcular a média?

Não, a média só pode ser calculada para dados quantitativos!

Chamamos a atenção para que, com dados de tipo *qualitativo*, não tem sentido calcular a média, mesmo que os dados sejam números. Se, por exemplo, temos um conjunto de "1's" e "2's" para representar as categorias da variável sexo, em que se utilizou o 1 para representar o sexo masculino e o 2 para o sexo feminino (variável codificada), não tem qualquer significado calcular a média daquele conjunto de dados. A média só se pode calcular para dados quantitativos, quer discretos, quer contínuos.

### Cálculo da média para dados discretos agrupados

Em amostras de *dados quantitativos discretos* aparecem muitos valores repetidos e, em vez de se somarem separadamente todos os valores da amostra, podem agrupar-se os valores que se repetem, obtendo-se

$$\bar{x} = \frac{x_1^* n_1 + x_2^* n_2 + \dots + x_k^* n_k}{n},$$

onde  $x_1^*$ ,  $x_2^*$ , ...,  $x_k^*$  representam os  $k$  valores distintos que surgem na amostra e  $n_i$  representa a frequência absoluta com que surge  $x_i^*$ ,  $i=1, 2, \dots, k$ .

Por exemplo, para calcular a média do número de letras do nome dos 24 alunos da turma de referência, podemos considerar a tabela de frequências com os dados agrupados, construída na secção 4.4.2.2

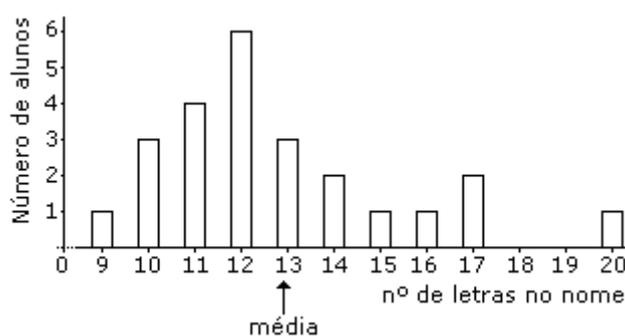
Nº de letras no nome $x_i^*$	Freq. Abs. $n_i$	Freq. Rel. $f_i$
9	1	0,042
10	3	0,125
11	4	0,167
12	6	0,250
13	3	0,125
14	2	0,083
15	1	0,042
16	1	0,042
17	2	0,083
20	1	0,042
Total	24	1,000

e utilizá-la para calcular a média do número de letras dos nomes dos 24 alunos:

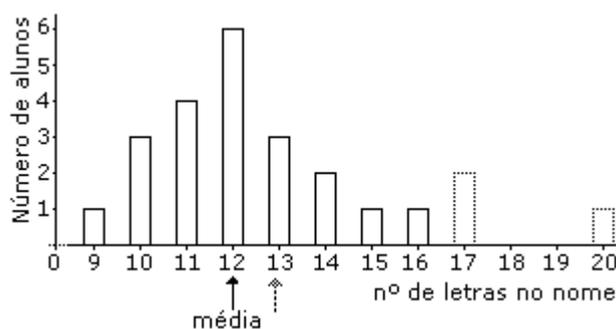
$$\bar{x} = \frac{9 \times 1 + 10 \times 3 + 11 \times 4 + 12 \times 6 + 13 \times 3 + 14 \times 2 + 15 \times 1 + 16 \times 1 + 17 \times 2 + 20 \times 1}{24}$$

obtendo-se para  $\bar{x} = 12,8$  letras.

Na figura seguinte apresentamos a posição da média:



O gráfico anterior mostra-nos um enviesamento para a direita na distribuição do número de letras do nome dos 24 alunos. Retirando os valores correspondentes aos 3 nomes com mais letras, esperamos que a média diminua. Efectivamente, calculando agora a média sem esses 3 valores obtemos o valor 12, diminuindo a média de cerca de uma unidade:



Sugestão – Verificar como é que se pode calcular a média, quando os dados estão agrupados, utilizando as frequências relativas, em vez de utilizar as frequências absolutas.

### Cálculo da média para dados contínuos agrupados

Para dados quantitativos contínuos, já organizados em classes, utiliza-se a fórmula anteriormente apresentada para calcular um valor aproximado para a média dos dados agrupados, sendo agora  $x_i^*$  o representante ou marca da  $i$ -ésima classe, que representámos por  $x'_i$ , e  $n_i$  a respectiva frequência absoluta. O valor que se obtém para a média, quando os dados estão agrupados é, neste caso, um valor aproximado, já que não estamos a calcular a média com os verdadeiros valores. Assim, utilizando a *tabela de frequências construída, na secção 4.5.1*, para a variável *altura de um aluno* da escola do 1º ciclo

Classes	Rep. Classe $x'_i$	Freq. Abs. $n_i$	Freq. Rel. $f_i$
[130, 135[	132,5	7	0,14
[135, 140[	137,5	9	0,18
[140, 145[	142,5	11	0,22
[145, 150[	147,5	14	0,28
[150, 155[	152,5	5	0,10
[155, 160[	157,5	4	0,08
Total		50	1,00

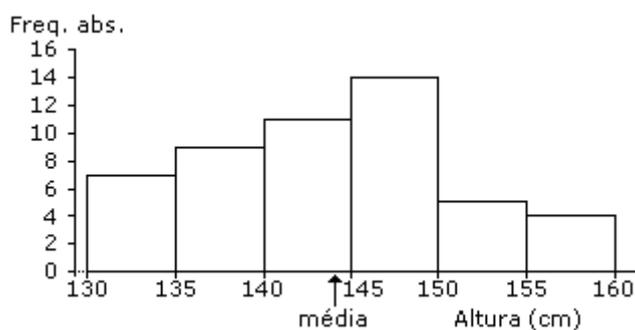
podemos obter um valor aproximado para a média das alturas:

$$\bar{x} \approx \frac{132,5 \times 7 + 137,5 \times 9 + 142,5 \times 11 + 147,5 \times 14 + 152,5 \times 5 + 157,5 \times 4}{50}$$

$$\approx 143,8\text{cm}$$

O valor obtido para a média, considerando os dados agrupados, é uma aproximação do valor obtido quando se consideram todos os dados.

Na figura seguinte apresenta-se a posição da média (aproximada) no histograma correspondente à tabela de frequências anterior:



A distribuição dos dados não apresenta um grande enviesamento, pelo que a posição da média reflecte razoavelmente bem o centro da distribuição dos dados.

Sugestão – Verificar que o valor obtido para a média quando se consideram todos os dados é igual a 142,7cm.

**Tarefa – Número de vogais e de consoantes do nome.** Na turma o professor desenhou no quadro uma tabela com 2 colunas e pediu a cada aluno que fosse preencher a tabela, indicando na primeira coluna o número de vogais do primeiro e último nome e na segunda coluna o número de consoantes.

Para os nomes dos alunos da turma de referência obter-se-ia a tabela seguinte:

Nº de vogais	Nº de consoantes
5	5
7	6
7	5
6	9
6	6
4	5
8	4
6	6
6	6
4	7
5	5
10	7
8	8
8	9
9	11
6	7
5	6
6	8
6	7
8	6
4	7
6	6
6	5
5	5

- a) O professor pediu aos alunos que organizassem os dados numa tabela de frequências e construíssem o diagrama de barras para cada um dos conjuntos de dados. Pediu ainda que respondessem às seguintes questões:
- Qual o aspecto apresentado pelos diagramas de barras construídos para os dois conjuntos de dados?
  - A forma apresentada pelos diagramas de barras permite estimar o valor aproximado para as médias dos dois conjuntos de dados? Utilizando as tabelas de frequência com os dados agrupados, calcule as médias dos conjuntos de dados e compare-as com os valores estimados a partir dos diagramas de barras.
  - Os nomes têm, de um modo geral, mais vogais ou mais consoantes? O que é que permitiu responder dessa maneira?
  - Quantos alunos têm no nome mais vogais do que a média? E mais consoantes do que a média? Este resultado era esperado a partir das representações gráficas dos dados?

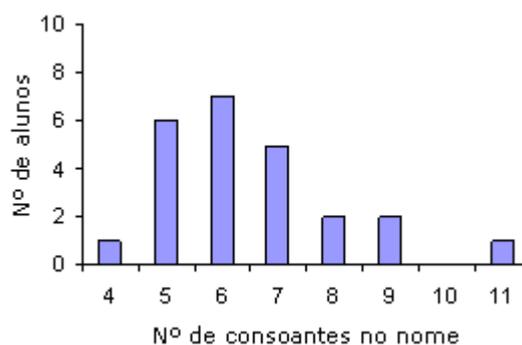
Para os dados da turma de referência, as tabelas de frequência e os diagramas de barras correspondentes são os seguintes:

**N.º de vogais no nome**

N.º de vogais	N.º de alunos
4	3
5	4
6	9
7	2
8	4
9	1
10	1
Total	24

**N.º de consoantes no nome**

N.º de consoantes	N.º de alunos
4	1
5	6
6	7
7	5
8	2
9	2
10	0
11	1
Total	24



O gráfico do lado esquerdo é bastante atípico, apresentando 2 modas, com uma das modas, no ponto 6, bastante mais frequente que a outra moda, no ponto 8. Estamos numa situação em que a média não reflecte o comportamento dos dados, mas esperamos que esteja entre o 6 e o 7, mas mais perto do 6. Relativamente ao gráfico do lado direito, temos um enviesamento para a direita, pelo que estimamos que a média do número de consoantes esteja entre o 6 e o 7, talvez mais perto do 7.

Utilizando as tabelas de frequências, calcularam-se as médias para o número de vogais e para o número de consoantes dos nomes dos 24 alunos:

$$\bar{x} = \frac{\text{Média do número de vogais}}{24} = \frac{4 \times 3 + 5 \times 4 + \dots + 10 \times 1}{24} = \frac{151}{24} \approx 6,3$$

$$\bar{x} = \frac{\text{Média do número de consoantes}}{24} = \frac{4 \times 1 + 6 \times 4 + \dots + 11 \times 1}{24} = \frac{156}{24} = 6,5$$

Os valores obtidos para as médias estão perfeitamente dentro do intervalo de valores que esperávamos obter a partir das representações gráficas. Estes valores levam-nos a concluir que os nomes têm em média, aproximadamente o mesmo número de vogais e consoantes, com alguma predominância de nomes com mais consoantes do que vogais.

No que diz respeito aos dados sobre o número de vogais e de consoantes no nome, temos, respectivamente:

- 16 nomes com um número de vogais inferior à média e 8 nomes com um número de vogais superior à média.
- 14 nomes com um número de consoantes inferior à média e 10 nomes com um número de consoantes superior à média.

Os valores anteriores são sintomáticos da falta de simetria da distribuição dos dados.

- b) Se à lista de nomes anteriores juntasse o nome da professora, que se chama Maria Natividade Almeida Jorge, espera-se que a média do número de vogais e do número consoantes, se mantenha, aumente ou diminua? Porquê? Como procederia para calcular a média de cada conjunto de dados a que acrescentou mais um elemento?

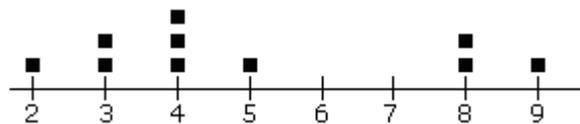
Considerando os dados da turma de referência, esperamos que as médias aumentem um pouco, pois estamos a acrescentar a cada um dos conjuntos de dados, um novo dado maior que os restantes, já que o novo nome tem 14 vogais e 13 consoantes. Para calcular as novas médias vamos aproveitar alguns dos resultados obtidos para calcular as médias anteriores:

$$\bar{x} = \frac{\text{Média do número de vogais}}{25} = \frac{151 + 14}{25} = 6,6$$

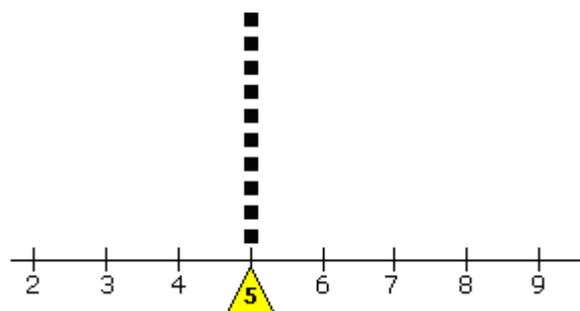
$$\bar{x} = \frac{\text{Média do número de consoantes}}{25} = \frac{156 + 13}{25} \approx 6,8$$

Apresentamos a seguir algumas tarefas para consolidar os conhecimentos sobre a média.

**Tarefa – Desvios entre os dados e a média.** Na turma o professor perguntou a 10 alunos que tinham o estojo de lápis em cima da mesa, quantos lápis (incluindo lápis de cor) tinham no estojo. As respostas obtidas 2, 3, 3, 4, 4, 4, 5, 8, 8, 9, encontram-se representadas no seguinte diagrama de pontos:

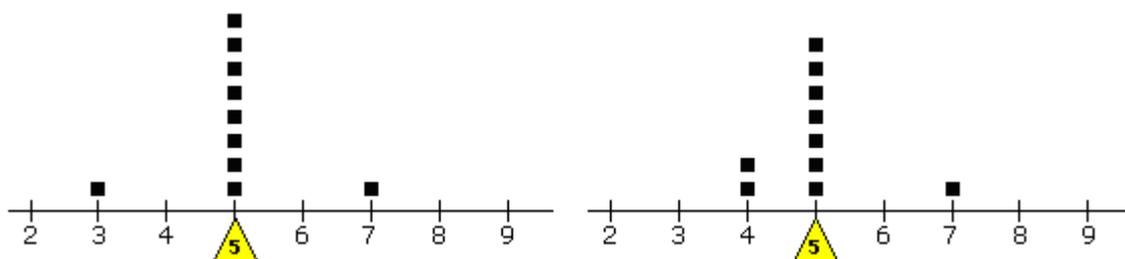


O professor pediu aos alunos para calcularem o número total de lápis existentes nos 10 estojos, tendo um dos alunos, o Miguel, chegado rapidamente à conclusão que tinham, ao todo, 50 lápis. De seguida o professor perguntou se algum dos alunos saberia dizer com quantos lápis ficaria cada um dos 10 alunos, se se distribuíssem os 50 lápis por todos, mas de forma igual. Todos responderam ao mesmo tempo que seriam 5 lápis! Cada um dos 10 alunos ficaria com 5 lápis e o gráfico de pontos que ilustra esta situação é o seguinte

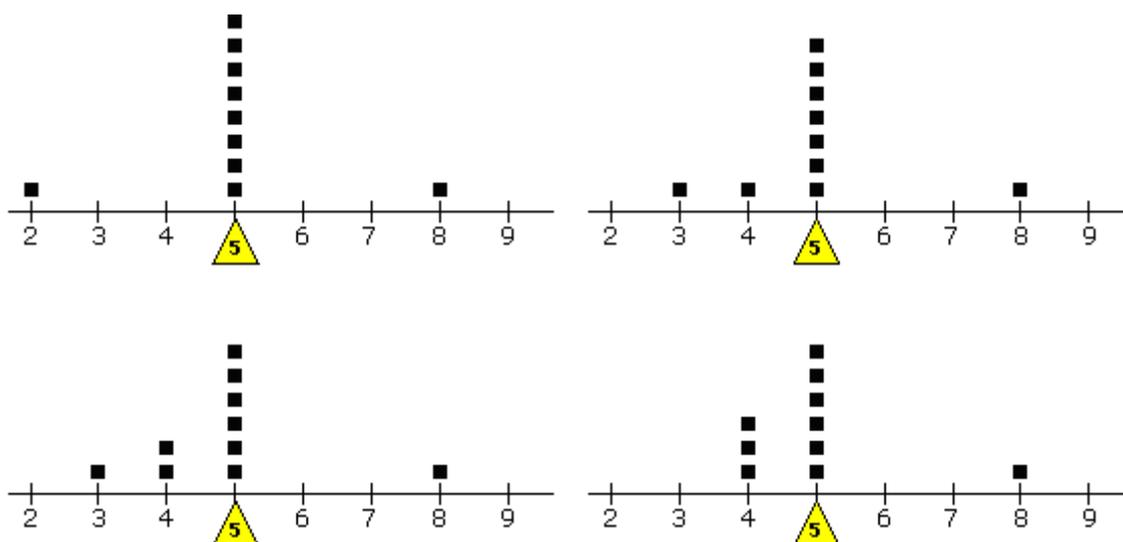


Imagine que a linha horizontal representa um balancé que está em equilíbrio no ponto de apoio assinalado no ponto 5. Existem agora vários processos de manter o balancé em equilíbrio, mas sem estarem todos os valores acumulados no ponto 5. Vejamos algumas situações:

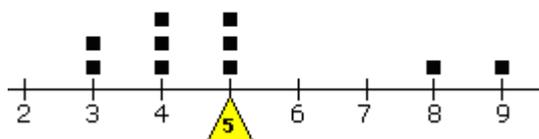
Deslocamos um dos pontos para a posição 7, mas para que o balancé mantenha o equilíbrio teremos de proceder de uma de duas maneiras: ou deslocamos um dos valores para a posição 3 ou 2 valores para a posição 4, como se apresenta a seguir



Se tivéssemos deslocado um dos pontos para a posição 8, em vez da posição 7, para que o balancé mantivesse o equilíbrio teríamos de proceder de uma das seguintes formas:

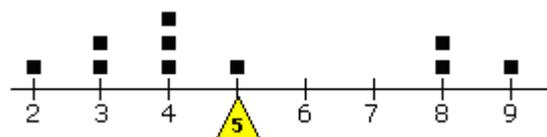


Peguemos na última situação e desloquemos um outro ponto da posição 5 para a posição 9. Como contrabalançar este ponto que afastámos da posição onde estava de 4 unidades ( $9-5=4$ )? Teremos de deslocar outro ou outros pontos da posição 5, mas agora para o lado esquerdo e de tal modo que as distâncias somem também 4 unidades. Uma situação possível é a que se apresenta a seguir (verifique se há mais situações possíveis):

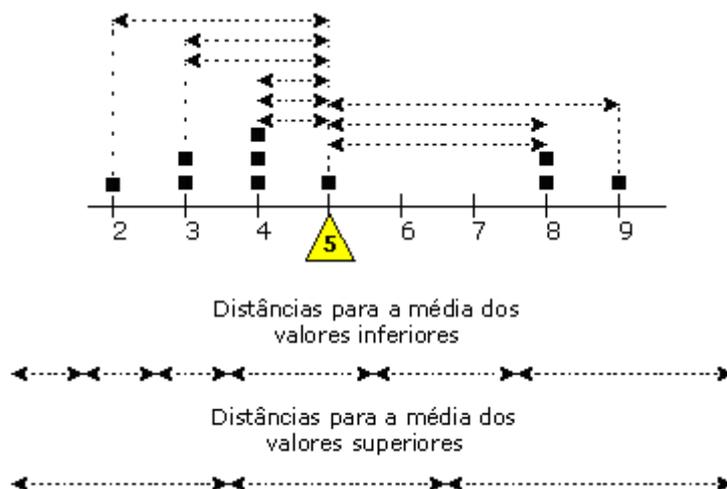


Colocámos 2 pontos na posição 3, pois a soma das suas distâncias à posição 5 é igual a 4 ( $(5-3)+(5-3)$ ), como se pretendia.

Retiremos ainda um outro ponto da posição 5 e coloquemos na posição 8. Para compensar este afastamento de 3 unidades, podemos deslocar para o outro lado, e a igual distância, um outro ponto:



Repare-se que a representação a que chegámos, foi a que nos foi dada inicialmente, representando a distribuição dos dados. Da forma como chegámos a esta representação, verificamos que distância total dos pontos superiores a 5, é igual à distância total dos pontos inferiores a 5:



Este ponto que goza desta propriedade é precisamente a média que, como já tínhamos afirmado anteriormente, é o ponto de equilíbrio da distribuição dos dados.

Se em vez de falarmos em distâncias, falarmos nas diferenças entre os valores e a média, obviamente que as diferenças entre os valores abaixo da média e a média têm sinal negativo. Como a soma dessas diferenças é igual, em valor absoluto, à soma das diferenças dos valores acima da média, para a média, vem que a soma das diferenças entre todos os valores da amostra e a média, é igual a zero.

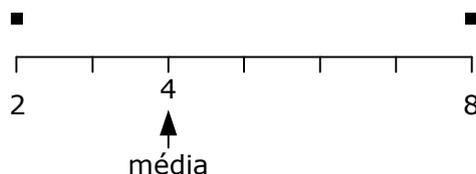
**Propriedade** – Se a todos os valores da amostra, subtrairmos a média, a soma das diferenças obtidas é igual a zero.

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) = 0$$

Repare-se que a propriedade anterior é uma consequência directa do facto da média ser o ponto de equilíbrio da distribuição dos dados. Para valores superiores à média, as diferenças são positivas mas para valores inferiores à média, as diferenças são negativas, sendo os totais das diferenças dos valores acima da média e abaixo da média iguais em valor absoluto.

**Tarefa – Quais as idades dos meus filhos? Qual a minha idade? Qual a idade da minha mulher?** – O professor chegou à turma e disse: a média das idades dos meus 4 filhos é 4 anos. O mais novo tem 2 e o mais velho 8. Que idades podem ter os meus dois outros filhos?

O professor desenhou no quadro o gráfico de pontos que ilustrava a situação que acabava de descrever:



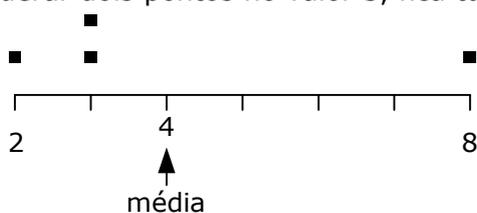
Entretanto desenrolou-se o seguinte diálogo:

*Miguel* (aluno): Oh professor, os outros dois filhos são gémeos?

*Professor*: Por acaso são!

*Miguel*: Então têm 3 anos, porque se a média é 4, a distância que vai do 8 ao 4 é 4, e a que vai do 2 ao 4 é 2, pelo que falta somar 2 unidades, para que o total das

distâncias dos valores superiores a 4 seja igual ao total das distâncias dos valores menores que 4. Se considerar dois pontos no valor 3, fica tudo certo!



As idades dos filhos do professor são 2, 3, 3 e 8 anos.

Comentário – Como a idade é uma variável, que embora seja contínua, é considerada em números inteiros, poder-se-ia dar o caso de os filhos não serem gémeos e terem ambos 3 anos. Também se poderia dar o caso de um dos filhos ter também 2 anos, como o irmão mais novo e o outro ter 4 anos.

Mas o diálogo não acabou aqui...

*Professor:* a média das idades dos meus filhos com a minha idade é igual a 9 anos. Que idade tenho eu?

*Tiago (aluno)* – Oh professor, eu vou fazer aqui umas contas rápidas e já lhe digo!

*Professor* – Explica essas contas que vais fazer, para todos ouvirmos.

*Tiago* – Sabemos que

$$\frac{2 + 3 + 3 + 8 + \text{idade professor}}{5} = 9$$

$$\frac{16 + \text{idade professor}}{5} = 9$$

Então

$$16 + \text{idade professor} = 45$$

e portanto

$$\text{Idade professor} = 29 \text{ anos}$$

*Professor* – Muito bem, Tiago. Conseguiste calcular a minha idade utilizando a definição da média.

*Miguel* - Oh professor, eu sei resolver isto de outra maneira!

*Professor* – Então explica-nos como é que fazes.

*Miguel* – Todas as idades dos seus filhos são inferiores à média, que é 9. Calculei as distâncias dessas idades à média

Idades	Distâncias para a média
2	9-2=7
3	9-3=6
3	9-3=6
8	9-1=1
Total	20

Pelas contas que acabei de fazer, sei que a idade do professor tem de ser superior à média de 20 unidades. Então a idade do professor é 29 anos, pois 20+9=29!

*Professor* – Muito bem Miguel. Mas agora ainda quero colocar outra questão. Qual é a média das idades da minha família, sabendo que a minha mulher tem 28 anos?

*André* – Vou responder eu! Se a soma das idades dos filhos do professor com a idade do professor é 45 anos, como vimos há pouco, então temos

$$\frac{45 + 28}{6} = \frac{73}{6} \approx 12,2 \text{ anos}$$

A média é aproximadamente 12 anos.

*Professor* – Muito bem André. Vejo que não te enganaste ao colocar no denominador o valor 6, pois estamos a fazer a média das idades de 6 pessoas. Estava com receio que colocasses no denominador 2, por no numerador só teres 2 parcelas.

*André* – Mas a primeira parcela já é o resultado da soma de 5 valores e eu para poupar tempo não os escrevi todos e pus logo o resultado.

**Tarefa – Idade média dos finalistas do curso de Matemática de 1950.** No jantar comemorativo dos 50 anos do curso, chegou-se à conclusão que a idade média dos matemáticos presentes era 71,5 anos. No ano seguinte os convivas resolveram comemorar de novo e, ao calcular a idade média dos presentes, chegou-se à conclusão que esta tinha descido e era 71,2 anos. Como é possível que, tendo a idade dos presentes aumentado de um ano, a média tenha baixado?

Vamos ver de seguida uma outra medida de localização do centro da amostra, alternativa à média – a mediana.

### 5.2.2 Mediana

A mediana é um valor que divide a amostra ao meio: metade dos valores da amostra são inferiores ou iguais (não superiores) à mediana e os restantes são maiores ou iguais (não inferiores) à mediana. Por outras palavras, até à mediana (inclusive) está, quanto muito, 50% da amostra; para lá da mediana (inclusive) está também, quanto muito, 50% da amostra.

Como obter a mediana?

Para determinar a mediana é fundamental, começar por **ordenar** os dados. Entretanto podem-se verificar duas situações, quanto à dimensão da amostra:

- Se a dimensão da amostra é ímpar, há um dos elementos da amostra ordenada que tem tantos elementos para a esquerda como para a direita e esse elemento *central* é a mediana.
- Se a dimensão da amostra é par, não há nenhum elemento que tenha a propriedade de a dividir ao meio. Há dois valores *centrais* e define-se a mediana como sendo a média aritmética desses dois valores.

Vejamos como calcular a mediana para alguns valores das dimensões das amostras:

**N ímpar**

Valor de $n$ – dimensão da amostra		A mediana é o elemen- to na posição:
3		2ª
5		3ª
7		4ª
9		5ª
11		6ª
...	...	...
27		14ª
...	...	...
$n$		$\frac{n+1}{2}$

**N par**

Valor de $n$ – dimensão da amostra		A mediana é a semi- soma dos elementos nas posições:
4		2ª e 4ª
6		3ª e 4ª
8		4ª e 5ª
10		5ª e 6ª
12		6ª e 7ª
...	...	...
26		13ª e 14ª
...	...	...
$n$		$\frac{n}{2}$ e $\frac{n}{2} + 1$

Repare-se que da forma como se calcula a mediana, quando a dimensão  $n$  da amostra é ímpar, a mediana é um elemento da amostra. Quando  $n$  é par, só será um elemento da amostra se os dois elementos centrais forem iguais.

Uma regra prática para obter a posição da mediana consiste em fazer o quociente  $\frac{n+1}{2}$ :

- Se este quociente for um número inteiro, o que se verifica quando  $n$  é **ímpar**, toma-se para mediana o elemento nessa posição;

- Se este quociente terminar em 0,5, o que se verifica quando **n é par**, considera-se a sua parte inteira e faz-se a semi-soma do elemento a que corresponde essa ordem, com o elemento da ordem seguinte.

Por exemplo, suponhamos que se pretende saber qual a mediana do número de letras do nome dos alunos da turma de referência:

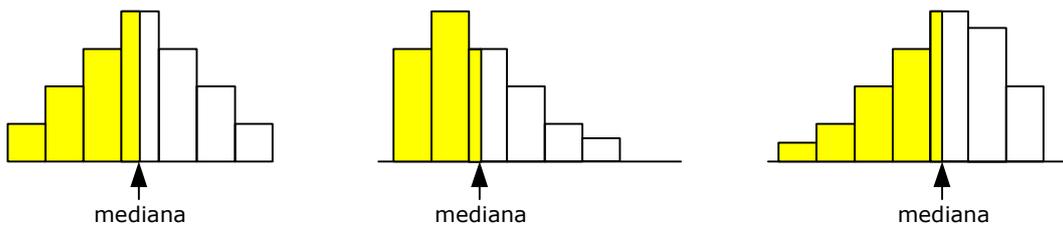
10, 13, 12, 15, 12, 9, 12, 12, 12, 11, 10, 17, 16, 17, 20, 13, 11, 14, 13, 14, 11, 12, 11, 10

Para calcular a mediana é necessário começar por ordenar a amostra:

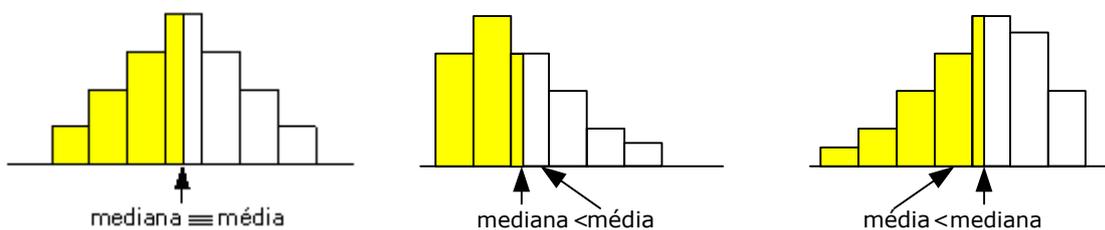
9, 10, 10, 10, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 13, 13, 13, 14, 14, 15, 16, 17, 17, 20

Temos 24 elementos, pelo que a mediana é a semi-soma dos elementos da 12.<sup>a</sup> posição ( $\frac{24 + 1}{2} = 12,5$ ) e da 13.<sup>a</sup> posição. Como os elementos nestas posições são iguais a 12, a mediana é 12.

Esquemáticamente podemos posicionar a mediana da forma que se segue, tendo em conta a representação gráfica na forma de histograma:

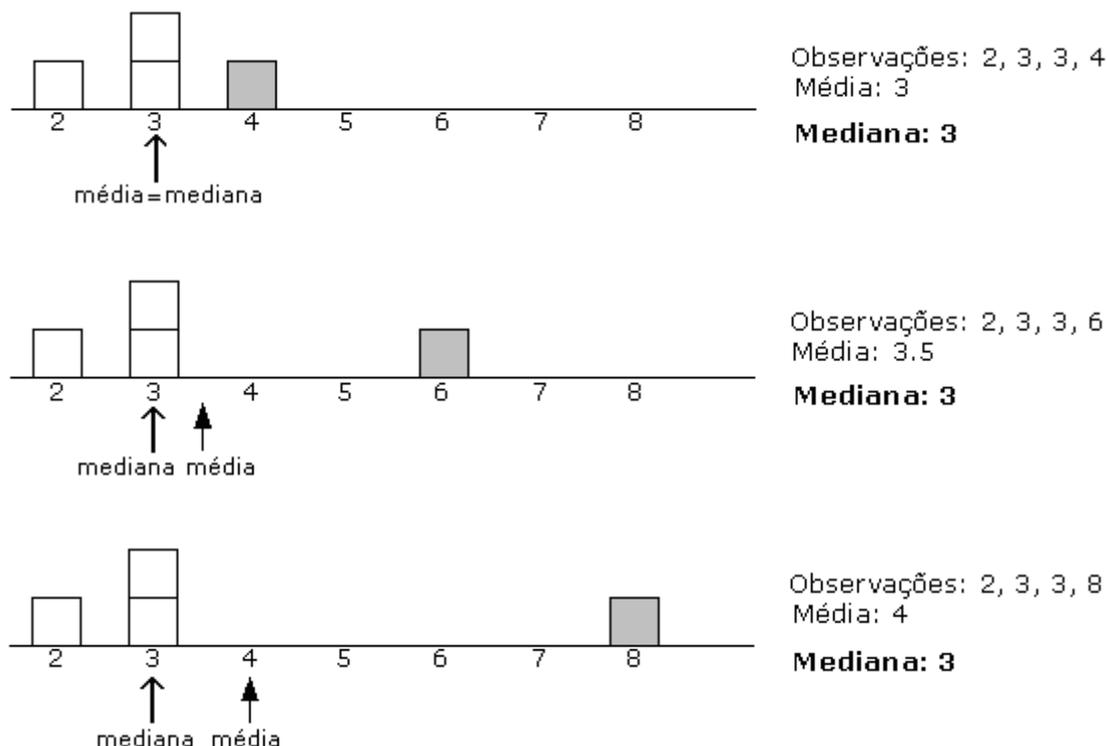


Ao contrário da mediana que "divide" o histograma em duas partes com áreas iguais, a média, como vimos, é o ponto de equilíbrio do histograma, em que se entra em linha de conta não só com a frequência das classes, mas também, com a distância a que estão do centro. Assim, na figura anterior, enquanto que no histograma do lado esquerdo, a média coincidirá com a mediana, no do centro, que apresenta um enviesamento para a direita, a média será "puxada" para a direita da mediana. Por outro lado, no histograma que apresenta o enviesamento para a esquerda, a média será "puxada" para a esquerda da mediana:



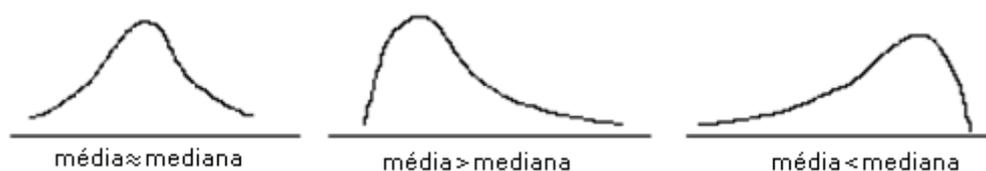
Como já referimos, a média, ao contrário da mediana, é uma medida muito pouco resistente, isto é, é muito influenciada por valores "muito grandes" ou "muito pequenos", mesmo que estes valores surjam em pequeno número na amostra. Estes valores, que chamamos de "outliers", são os responsáveis pela má utilização da média em muitas situações em que teria mais significado utilizar a mediana.

Por exemplo, na situação considerada para exemplificar a pouca resistência da média, e que voltamos a apresentar a seguir, verifica-se que, ao contrário da média, a mediana não se altera, quando se altera um dos dados:



A mediana tem como principal desvantagem o facto de, no seu cálculo, só fazer intervir 1 ou 2 valores da amostra. No entanto, esta desvantagem transforma-se em vantagem, por comparação com a média, quando a distribuição da amostra é muito enviesada. A mediana é muito resistente e não é afectada pelos valores extremos, como acabámos de ver no exemplo anterior, em que a mediana não de alterou.

Resumindo, como a média é influenciada quer por valores muito grandes, quer por valores muito pequenos, se a distribuição dos dados for enviesada para a direita (alguns valores grandes como *outliers*), a média tende a ser maior que a mediana; se for aproximadamente simétrica, a média aproxima-se da mediana e se for enviesada para a esquerda (alguns valores pequenos como *outliers*), a média tende a ser inferior à mediana. Representando as distribuições dos dados (esta observação é válida para as representações gráficas na forma de diagrama de barras ou de histograma) na forma de uma mancha, temos, de um modo geral (Graça Martins, 2005):



Observe-se que o simples cálculo da média e da mediana nos pode dar informação sobre a forma da distribuição dos dados.

### Cálculo da mediana quando os dados estão agrupados

No estudo de dados qualitativos ordinais (isto é, onde se pode considerar uma ordem subjacente às categorias) faz sentido indicar a categoria mediana. A categoria mediana é aquela onde, pela primeira vez, a frequência relativa acumulada atinge ou ultrapassa os 50%. Esta mesma definição serve para identificar a classe mediana no caso de se estar perante dados agrupados, quer sejam discretos, quer sejam contínuos.

Consideremos o exemplo apresentado na secção anterior sobre a variável *número de letras no nome* dos alunos da turma de referência, cuja tabela de frequências se apresenta a seguir:

N.º de letras no nome $x_i^*$	Freq. Abs. $n_i$	Freq. Rel. $f_i$
9	1	0,042
10	3	0,125
11	4	0,167
12	6	0,250
13	3	0,125
14	2	0,083
15	1	0,042
16	1	0,042
17	2	0,083
20	1	0,042
Total	24	1,000

Para podermos calcular a mediana, a partir da tabela dos dados agrupados, vamos juntar à tabela uma nova coluna com as frequências relativas acumuladas:

N.º de letras no nome $x_i^*$	Freq. Abs. $n_i$	Freq. Rel. $f_i$	Freq. Rel. Acum. %
9	1	0,042	4,2
10	3	0,125	16,7
11	4	0,167	33,4
12	6	0,250	<b>58,4</b>
13	3	0,125	70,9
14	2	0,083	79,2
15	1	0,042	83,4
16	1	0,042	87,6
17	2	0,083	95,9
20	1	0,042	≈1
Total	24	1,000	

Reparando na tabela, verifica-se que a frequência relativa acumulada atinge o valor 50% quando se soma a frequência relativa correspondente ao valor 12. Então a mediana é igual a 12.

Numa tabela idêntica à anterior, com os dados agrupados, pode-se verificar uma situação especial como a que se apresenta a seguir:

N.º de letras no nome $x_i^*$	Freq. Rel. $f_i$	Freq. Rel. Acum. %
9	0,042	4,2
10	0,125	16,7
11	0,200	36,7
12	0,133	<b>50,0</b>
13	0,125	62,5
14	...	...
15	...	...
16	...	...
17	...	...
20	...	≈1
Total	1,000	

Na tabela anterior verifica-se que 50% dos alunos têm nomes com 12 ou menos letras e os outros 50% têm nomes com 13 ou mais letras. Esta situação só se pode verificar se o número de elementos da amostra for par. Como vimos anteriormente, neste caso a mediana será a semi-soma dos dois elementos centrais, pelo que a mediana seria 12,5.

No caso de se tratar de uma variável contínua, como por exemplo a variável *altura de um aluno* da escola do 1.º ciclo, estudada na secção 4.5.1 e cuja tabela se apresenta a seguir, a classe mediana também se consegue identificar facilmente, pois basta verificar qual a classe a que corresponde uma frequência acumulada igual a 50%:

Classes	Representante da Classe $x'_i$	Freq. Abs. $n_i$	Freq. Rel. $f_i$	Freq. Abs. Acum.	Freq. Rel. Acum.	Freq. Rel. Acum. (%)
[130, 135[	132,5	7	0,14	7	0,14	14
[135, 140[	137,5	9	0,18	16	0,32	32
[140, 145[	142,5	11	0,22	27	0,54	54
[145, 150[	147,5	14	0,28	41	0,82	82
[150, 155[	152,5	5	0,10	46	0,92	92
[155, 160[	157,5	4	0,08	50	1,00	100
Total		50	1,00			

Da tabela anterior conclui-se que a classe mediana é a classe [140, 145[. No entanto, ao contrário do que se verifica com as variáveis discretas agrupadas, em que a mediana está bem determinada, no caso de dados contínuos a classe mediana depende do agrupamento que se fizer para os dados. Vimos na secção 4.5.2 um processo de obter um valor aproximado para a mediana a partir da tabela de frequências ou a partir do histograma acumulado.

Nota – Não existe uma notação única para representar a mediana. As notações mais usuais são  $m$ ,  $M$  ou  $Me$ .

**Tarefa – Notas no teste de Matemática das turmas 9.º A e 9.º B.** Na turma o professor apresentou as notas que os seus alunos das turmas 9.º A e do 9.º B tinham tido no mesmo teste de Matemática. Pretendia que os alunos lhe dissessem qual seria a turma que teve um melhor desempenho no teste.

Notas da Turma 9.º A

10,6 9,8 10,4 10,8 11,2 10,2 11,6 10,6 9,8 12,2 12,4 11,4 10,8 13,8 8,6  
10,4 11,2 11,8 10,6 11,6

Notas da Turma 9.º B

9,4 10,0 11,0 8,2 13,6 10,0 9,4 11,2 9,8 12,6 15,6 7,2 16,8 10,8 9,4 8,8  
11,2 7,4 12,4 15,0

Para comparar as duas turmas, o professor sugeriu que os alunos se organizassem em grupos e cada um dos grupos tentaria retirar alguma informação relevante a partir dos dados. Depois de alguma discussão, ficou decidido que um dos grupos iria fazer uma representação gráfica dos dados e um outro grupo iria apresentar os resultados utilizando a média e a mediana, para terem uma ideia do comportamento global das turmas. Os outros grupos não especificaram qual a forma como iriam pegar no problema, pelo que se houvesse alguma informação relevante que não tivesse sido apresentada pelos 2 grupos, também interviriam na apresentação final das conclusões.

Antes de começarem a organizar os dados, um aluno reparou que na turma B tinha havido uma nota muito boa, mas este facto não significava que a turma B tivesse tido um melhor desempenho. O comportamento individual não reflecte o comportamento da turma.

#### *Apresentação do 1.º grupo*

O professor pediu-nos para compararmos os resultados num teste a Matemática, dos alunos do 9.º A e do 9.º B. Para termos uma ideia do comportamento global das notas, decidimos representar os dois conjuntos de dados numa representação em caule-e-folhas, que tem a mais valia de permitir obter, de uma forma muito simples, os dados ordenados:

Notas da Turma 9.º A		Notas da Turma 9.º B
	7	2 4
6	8	2 8
8 8	9	4 4 8 4
6 4 8 6 2 8 4 6	10	0 0 8
6 8 2 4 6 2	11	0 2 2
4 2	12	6 4
8	13	6
	14	
	15	6 0
	16	8

Numa primeira análise, as representações gráficas anteriores permitem-nos concluir que os alunos da turma A foram mais regulares, havendo uma concentração de notas à volta dos valores 10 e 11, o que nos leva a sugerir que a média deve estar compreendida entre estes dois valores. Se se considerar que uma nota superior a 9,5 é positiva, podemos dizer que na turma A houve apenas uma negativa. No que diz respeito à turma B, houve alunos com notas mais baixas mas, em contrapartida, também houve alunos com notas mais altas. Como se pode ver, houve uma maior dispersão das notas. No que diz respeito à média, estimamos que deve estar perto de 11.

#### *Apresentação do 2.º grupo*

Temos dois conjuntos de dados, referentes às notas de duas turmas e pretendemos averiguar qual a turma que teve um melhor desempenho no teste. Vamos utilizar como medidas de comparação a média e a mediana. Na utilização da média temos que ter os cuidados devidos, já que esta medida é muito pouco resistente, sendo facilmente influenciada por valores grandes ou pequenos, ou seja, valores que saiam fora do âmbito da maior parte dos restantes.

Para calcular a mediana temos de ter os dados ordenados, que se apresentam a seguir:

Notas da Turma 9.º A

8,6 9,8 9,8 10,2 10,4 10,4 10,6 10,6 10,6 10,8 10,8 11,2 11,2 11,4 11,6  
11,6 11,8 12,2 12,4 13,8

Notas da Turma do 9.º B

7,2 7,4 8,2 8,8 9,4 9,4 9,4 9,8 10,0 10,0 10,8 11,0 11,2 11,2 12,4  
12,6 13,6 15,0 15,6 16,8

Utilizando a expressão que nos permite calcular a média, concluímos que os dois conjuntos de dados têm a mesma média, igual a 10,99. O cálculo da mediana

pode-nos trazer mais alguma informação interessante sobre as notas das duas turmas. Como temos um número par de dados, a mediana será a semi-soma dos elementos de ordem 10.<sup>a</sup> e 11.<sup>a</sup>, na amostra ordenada. Resumimos na seguinte tabela as características amostrais média e mediana dos dois conjuntos de dados:

	Média	Mediana
Notas turma A	10,99 $\approx$ 11	10,8 $\approx$ 11
Notas turma B	10,99 $\approx$ 11	10,4 $\approx$ 10

Os resultados anteriores são sintomáticos da falta de “regularidade” anunciada pelos nossos colegas, quando apresentaram a representação gráfica dos dados. Efectivamente, enquanto que na turma A a média e a mediana são muito próximas, o mesmo não acontece na turma B, sintoma da falta de simetria dos dados. Concluimos também que nesta turma, 50% dos alunos têm nota menor ou igual a 10,4. Uma investigação mais cuidada aos dados ordenados, permite-nos verificar que na turma A, só 30% dos alunos é que têm nota menor ou igual a 10,4. Se além disso nos preocuparmos com a percentagem de notas negativas, verificamos que na turma A só houve uma negativa, ou seja 5% de negativas, enquanto que na turma B houve 7 negativas (ou seja, 35%). Somos assim de opinião, que a turma A teve um melhor desempenho, embora na turma B tenha havido as duas melhores notas do teste.

**Tarefa – Média ou mediana dos salários?** O professor chegou à aula e colocou a seguinte situação que lhe tinha sido colocada por um amigo que trabalha numa empresa que emprega 160 trabalhadores. Na empresa, iam em breve discutir um aumento de salários e o administrador estava um pouco renitente em dar um aumento superior à inflação, pois alegava que a média dos salários naquela empresa, estava de acordo com a média dos salários auferidos pelos trabalhadores daquele ramo de actividade. O amigo do professor estava confuso com esta situação, pois tinha comparado salários de vários amigos seus, de outras empresas, e quase todos ganhavam mais que ele. Como explicar esta situação?

Dados fornecidos pelo amigo do professor, sobre os 6 níveis de salários existentes na empresa:

Salário (em euros)	400	450	600	700	1000	5000
Nº empregados	23	58	50	20	7	2

Quando o professor apresentou a tabela anterior, houve logo um aluno, o Miguel, que interpelou o professor, tendo-se estabelecido o seguinte diálogo:

*Miguel* – Oh professor, eu acho que a média não é uma boa medida para caracterizar esses dados!

*Professor* – Então porquê, Miguel?

*Miguel* – Porque estou a ver que há dois salários que são muito grandes, quando comparados com os restantes. Como nós aprendemos que a média não é uma medida boa quando há destes valores “esquisitos”, de certeza que esses dois salários vão fazer com que a média, venha “grande” e acaba por ser uma medida enganadora.

*Professor* – A esses valores esquisitos de que falas, dá-se o nome de “outliers”, que efectivamente quer dizer “estranhos”, fora do contexto dos restantes. Mas então o que é que propões?

*Miguel* – Penso que se deve calcular a mediana.

Professor – Vamos então calcular a mediana e a média e aproveitamos para comparar os dois resultados.

#### Cálculo da média:

$$\bar{x} = \frac{400 \times 23 + 450 \times 58 + 600 \times 50 + 700 \times 20 + 1000 \times 7 + 5000 \times 2}{160}$$

$$\bar{x} \approx 602 \text{ euros}$$

#### Cálculo da mediana:

Para calcular a mediana considera-se a tabela das frequências relativas acumuladas

Salário (em euros)	400	450	600	700	1000	5000
Nº empregados	23	58	50	20	7	2
Freq. Rel. (%)	14,38	36,25	31,25	12,50	4,38	1,25
Freq. Rel. acum. (%)	14,38	<b>50,63</b>	81,88	94,38	98,75	100,00

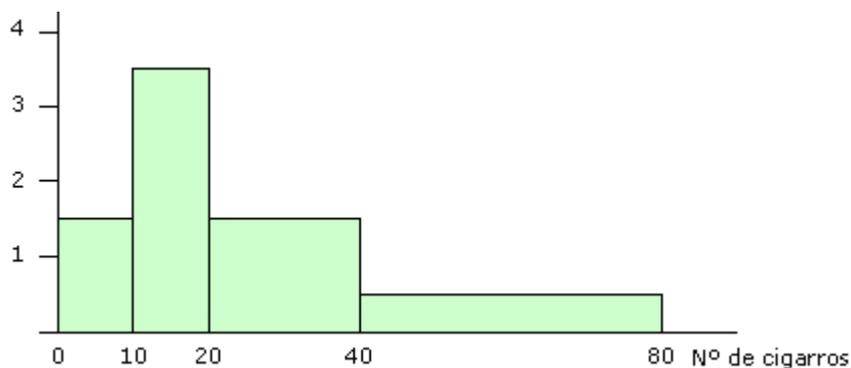
A partir da tabela anterior concluímos que a mediana é 450 euros, porque corresponde ao valor em que se atingiu a frequência relativa acumulada de 50%.

O Miguel tinha razão. A mediana dá-nos uma ideia mais correcta do nível dos salários, que são de um modo geral baixos. Na verdade 50% dos salários são menores ou iguais a 450 euros. A média é muito superior à mediana, o que acontece sobretudo devido aos 2 salários de 5000 euros, eventualmente dos administradores, que inflacionaram a média. Repare-se que, dos 160 trabalhadores, só 29 é que têm um salário superior à média.

O professor decidiu falar com o amigo e sugerir-lhe que na discussão sobre os aumentos dos salários invocassem o facto de a mediana ser tão baixa. Se os administradores ainda continuassem a querer utilizar a média, como medida de referência, então deviam sugerir-lhe que não contassem para a média os 9 salários mais altos. A distribuição ainda continha algum enviesamento para a direita, mas agora a média viria igual a 525, mais próxima da mediana.

*Uma situação caricata* – Num autocarro viajavam 25 trabalhadores da empresa X, que em média ganhavam 450 euros por mês. O nível de vida destes trabalhadores aumentou de um momento para o outro, quando entrou no autocarro o administrador da empresa, pois passaram a ganhar muito mais, em média!

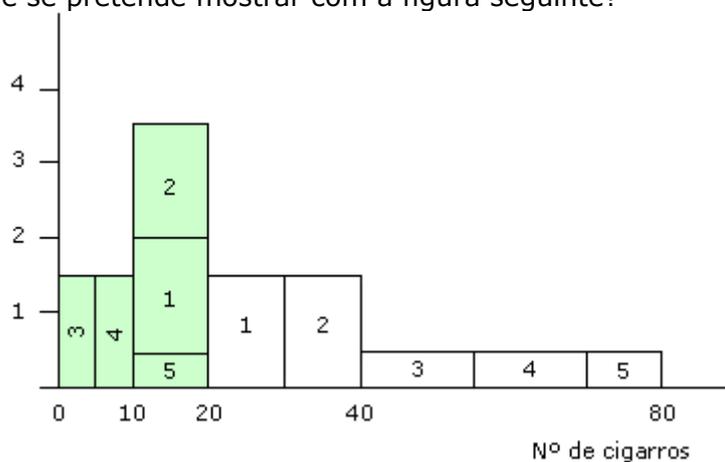
**Tarefa –Número de cigarros fumados por dia**<sup>1</sup>. O professor apresentou na aula um histograma que representa o resultado de um estudo sobre o Tabaco e a Saúde Pública e em que é apresentado o número de cigarros que é fumado por dia por indivíduos do sexo masculino:



<sup>1</sup> Adaptada de Freedman et al. (1991).

Nesse estudo era dito que os intervalos considerados para o número de cigarros por dia incluíam o limite superior e não o limite inferior e que a percentagem de homens que fumava 10 ou menos cigarros, por dia, era de 15%. Algumas questões relativamente ao estudo apresentado:

- Qual a percentagem de homens que fuma mais de meio maço, mas não mais de 1 maço, por dia?
- Qual a percentagem de homens que fuma mais de um maço, mas não mais de 2 maços, por dia?
- Estime a percentagem de homens que fuma mais de 3 maços por dia?
- Estime a percentagem de homens que fuma entre 2 e 3 maços por dia?
- Tendo em atenção o histograma anterior, espera que a mediana seja superior ou inferior à média? Estime valores para essas características.
- Obtenha valores aproximados para a média e a mediana e compare os valores obtidos com os valores estimados na alínea anterior.
- O que é que se pretende mostrar com a figura seguinte?



### 5.2.3 Moda<sup>2</sup>

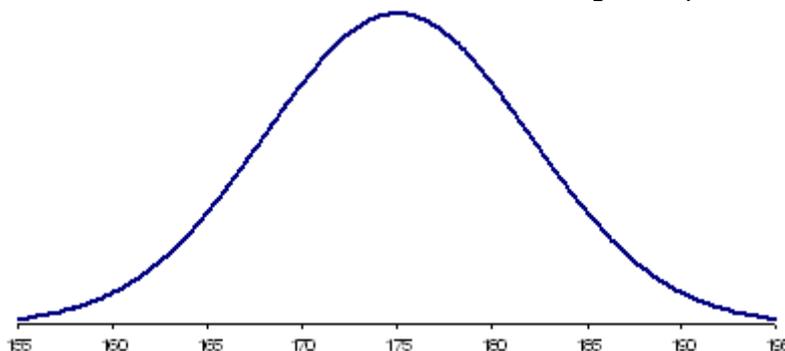
Uma outra medida que costuma ser apresentada como medida de tendência central é a moda. No entanto a moda é uma medida que, a este nível, tem pouco interesse, como medida de localização do centro da distribuição dos dados, e deve a sua importância ao facto de ser a única medida que pode ser calculada para dados qualitativos, para os quais não se possa estabelecer uma hierarquia entre as várias categorias que a variável pode assumir, não sendo possível, portanto, nem calcular a média nem a mediana.

Em amostras de dados qualitativos dá-se o nome de *moda* ou *categoria modal*, à categoria de maior frequência na amostra. Em amostras de dados quantitativos discretos, designa-se por *moda* qualquer valor que esteja ladeado por valores de menor frequência. As modas são, pois, "picos" na distribuição de frequências. Em amostras de dados quantitativos contínuos, após subdivisão em classes, ficam identificadas as *classes modais*, que são aquelas que estão ladeadas de classes de menor frequência.

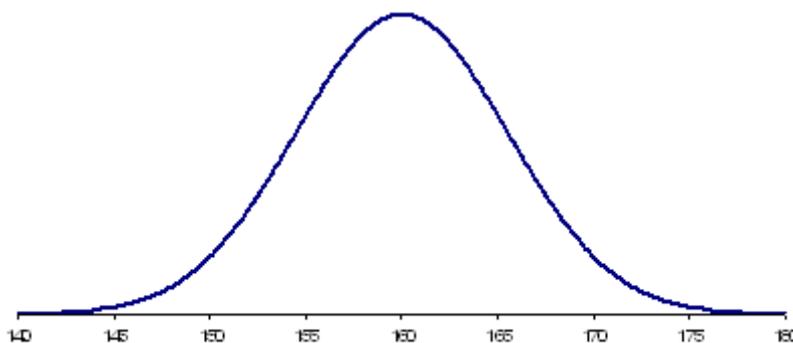
Em curvas que modelam situações da vida real, dá-se o nome de moda a qualquer máximo relativo da curva de densidade. Os modelos teóricos de interesse têm uma única moda e é usual dizer que o aparecimento de várias modas pode evidenciar

<sup>2</sup> Esta secção segue de perto a secção 3.2.5 de Graça Martins et al (2007).

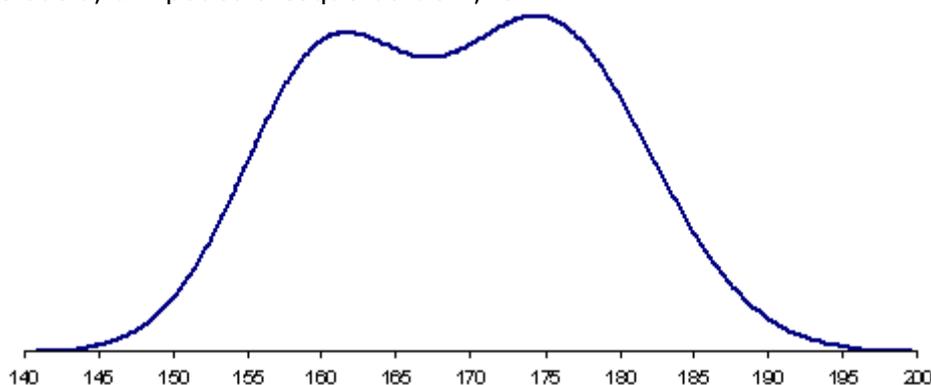
mistura de populações e é nesta análise que poderá ter interesse a identificação da moda ou modas. Para ilustrar esta ideia, tome-se o exemplo das alturas na população portuguesa. Se considerarmos somente a subpopulação dos homens, a distribuição das suas alturas não deve afastar-se muito do seguinte padrão:



Note-se que a zona de maior concentração ou densidade, está entre 1,70m e 1,80m, sendo a moda (máximo relativo da curva) igual a 1,75m. A forma da distribuição das alturas das mulheres deverá ser idêntica, mas localizada em torno de 1,60m:



Que aconteceria se considerássemos as duas subpopulações em conjunto? Onde ficaria a moda? Em 1,75m, em 1,60m ou algures entre estes dois valores? Na verdade o que acontece é que surgem duas modas!... Uma, um pouco à direita de 1,60m e outra, um pouco à esquerda de 1,75m:



A bimodalidade torna-se ainda mais evidente se a zona central de uma das distribuições se encontrar muito afastada da zona central da outra e se a percentagem de observações pertencentes a cada uma das duas subpopulações for idêntica. Retomando o exemplo das alturas, se numa amostra de 100 indivíduos tivermos 10 mulheres e 90 homens é muito pouco provável que o histograma apresente bimodalidade, contrariamente ao que deverá ocorrer em amostras com 50 homens e 50 mulheres.

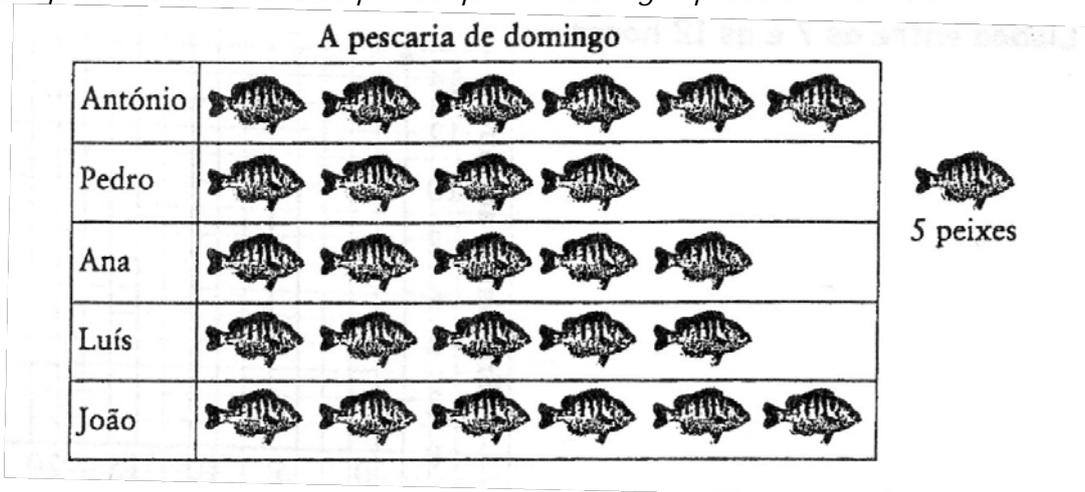
Ainda a propósito deste exemplo, chamamos mais uma vez a atenção, para o facto de o histograma ser uma representação gráfica que, para alguns conjuntos de dados, pode mudar sensivelmente de aspecto, quando se altera a amplitude de classe ou o ponto onde se começam a construir as classes. Assim, para o mesmo conjunto de dados pode acontecer haver representações gráficas diferentes, nomeadamente em termos do número de modas.

Sendo então a moda uma medida com uma aplicação relativamente restrita, tem algum interesse quando dispomos de uma amostra de grande dimensão, mas com um número restrito de valores distintos. Por exemplo, uma boa utilização da moda é na indicação do número de filhos de uma família "típica" portuguesa, ou no tamanho do pé de uma mulher. O dono de uma sapataria tem interesse em saber qual o tamanho mais vendido, pois será nesse tamanho que vai investir mais, no princípio de cada época.

### Confusão entre dado mais frequente e dado com o maior valor!

Por vezes verifica-se haver alguma confusão entre dado mais frequente e dado com o maior valor. Para justificar o que acabámos de dizer, apresentamos a seguir dois exemplos retirados de textos de Matemática do Ensino Básico.

"O gráfico representa o número de peixes que cinco amigos pescaram num dia:



- 1.1 – Como se chama este tipo de gráfico?
- 1.2 – Quantos peixes pescou o António mais que o Pedro?
- 1.3 – Quantos peixes pescaram ao todo os cinco amigos?
- 1.4 – Classifique a distribuição quanto à **moda**. Justifique".

O gráfico anterior apresenta os dados da variável que se está a estudar – Número de peixes pescados num dia, por cada um de cinco amigos. A partir desse gráfico verificamos que os dados são

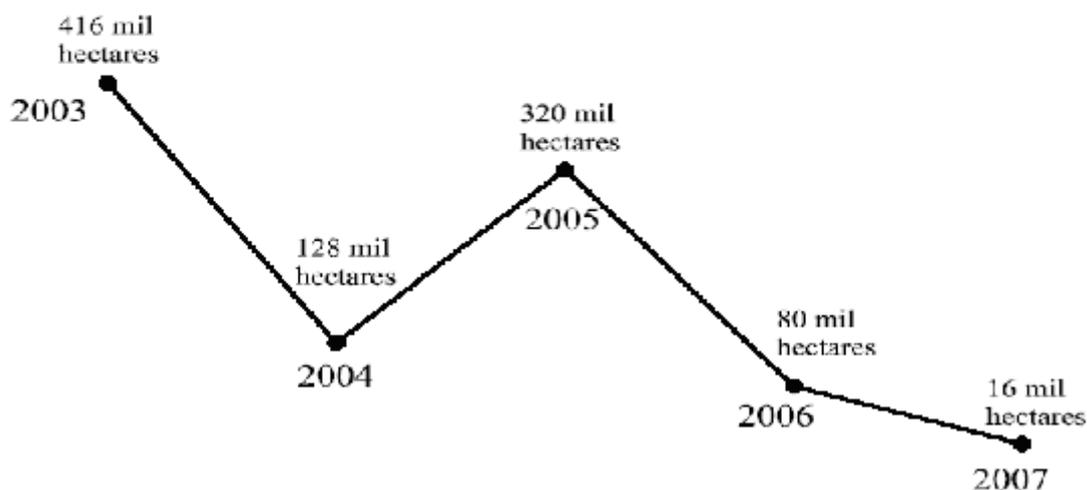
30, 20, 25, 25, 30

pelo que temos dois dados com a mesma frequência, ou seja, temos duas modas: o 25 e o 30 (Seguramente que não era esta a resposta que os autores do exemplo pretendiam para a questão 1.4!).

Ainda um segundo exemplo:

"O gráfico seguinte mostra o número de hectares de floresta ardida, em Portugal Continental, entre os anos de 2003 e 2007

## Incêndios florestais em Portugal



*Classifica a distribuição quanto à moda."*

No exemplo anterior a variável em estudo é – Área ardida (em milhares de ha) em cada um dos anos considerados, pelo que os dados são (em milhares de ha)

416, 128, 320, 80, 17

Como os dados são todos diferentes, não existe moda (Seguramente que o que os autores pretendiam não era a resposta anterior!).

Nota – A representação considerada anteriormente – gráfico de linha, é a adequada para representar este tipo de dados, em que se estuda a evolução da variável com o tempo.

**Observação** - Quando se pretende saber qual o *centro de uma distribuição de dados*, a resposta a esta pergunta é fácil se a distribuição for aproximadamente simétrica e unimodal (só com uma moda). Se a distribuição dos dados apresentar outras formas, nomeadamente enviesamento ou várias modas, já o conceito de centro da distribuição dos dados pode não fazer sentido, como já referimos anteriormente ao tratarmos das medidas de tendência central média e mediana.

### 5.2.4 Quartis

A média e a mediana dão-nos duas formas diferentes de localizarmos o centro da distribuição dos dados. Existem outras medidas, os **quartis**, que localizam outros pontos da distribuição dos dados, que não o centro, e que têm a mais valia de servirem para definir uma medida da variabilidade existente entre os dados.

Como vimos na definição de mediana, esta divide a amostra ordenada em duas partes com igual percentagem de elementos. Considerando cada uma destas partes e calculando a mediana, obteremos o 1.º e 3.º quartis, que já foram utilizados na construção do diagrama de extremos e quartis. A mediana, que também se poderia designar de 2.º quartil, e os 1.º e 3.º quartis localizam pontos que dividem a distribuição dos dados em quatro partes, com igual percentagem de elementos. Daí vem o nome de quartis!

Há vários processos para calcular os quartis, nem todos conducentes aos mesmos valores, mas a valores próximos, desde que a amostra tenha uma dimensão razoável, que é a situação de interesse em estatística, em que se procura reduzir a informação contida nesses dados, através de algumas medidas.

A metodologia que, a este nível, recomendamos para obter os quartis é a seguinte:

- Ordenar os dados e calcular a mediana  $Me$ ;
- O 1.º quartil,  $Q_1$ , é a mediana dos dados que ficam para a esquerda de  $Me$ ;
- O 3.º quartil,  $Q_3$ , é a mediana dos dados que ficam para a direita de  $Me$ .

Ao calcular os quartis pelo processo anterior, podem-se levantar algumas dúvidas, no caso em que a dimensão da amostra é ímpar. Efectivamente, neste caso a mediana coincide com um dos elementos da amostra e poderíamos optar por considerá-lo incluído nas duas metades em que fica dividida a amostra, ou não o considerar em nenhuma das metades. A nossa opção é considerá-lo pertencente às duas metades.

Por analogia com a definição que demos para a mediana, podemos dizer que até ao 1.º quartil (inclusive) está, pelo menos, 25% da amostra; para lá do 1.º quartil (inclusive) está, pelo menos, 75% da amostra. De forma análoga podemos dizer que até ao 3.º quartil (inclusive) está, pelo menos, 75% da amostra; para lá do 3.º quartil (inclusive) está, pelo menos 25% da amostra.

**Tarefa – Notas no teste de Matemática das turmas 9.º A e 9.º B (cont.).** Na comparação dos resultados das duas turmas, podemos também utilizar os quartis, na medida em que quanto maiores forem, melhor terá sido o comportamento da turma. Vejamos então quais os quartis para os dois conjuntos de dados considerados. Para o seu cálculo vamos considerar as amostras já ordenadas:

Notas da Turma 9.º A

8,6 9,8 9,8 10,2 **10,4 10,4** 10,6 10,6 10,6 10,8 10,8 11,2 11,2 11,4 **11,6**  
11,6 11,8 12,2 12,4 13,8

Notas da Turma 9.º B

7,2 7,4 8,2 8,8 **9,4 9,4** 9,4 9,8 10,0 10,0 10,8 11,0 11,2 11,2 **12,4**  
12,6 13,6 15,0 15,6 16,8

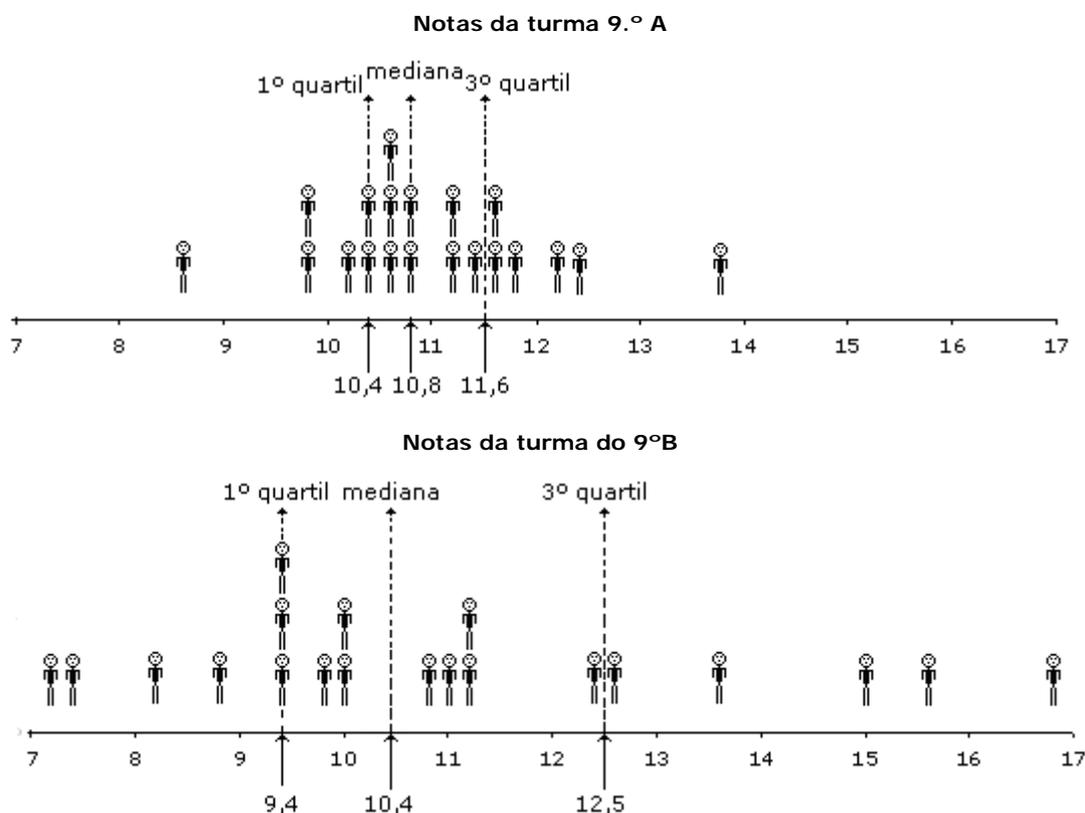
Como temos um número par de dados, a mediana, como já se viu, é a semi-soma dos dois elementos centrais. Cada uma das partes em que ficaram divididos os dados pela mediana, tem 10 elementos. A mediana de cada uma destas partes será ainda a semi-soma dos dois elementos centrais, assinalados a negro, obtendo-se para os quartis os valores seguintes:

	<b>1.º quartil</b>	<b>3.º quartil</b>
Notas turma 9.º A	$\frac{10,4 + 10,4}{2} = 10,4$	$\frac{11,6 + 11,6}{2} = 11,6$
Notas turma 9.º B	$\frac{9,4 + 9,4}{2} = 9,4$	$\frac{12,4 + 12,6}{2} = 12,5$

Da tabela anterior concluímos que pelo menos 25% dos alunos da turma 9.º B tiveram negativa, enquanto que na turma 9.º A foram menos de 25% a ter negativa, já que 25% tiveram nota menor ou igual a 10,4. Por outro lado, 25% dos alunos da turma 9.º B tiveram nota maior ou igual a 12,5. Mais uma vez se comprova que a turma 9.º A é mais regular, pois 50% das notas estão no intervalo de amplitude 1,2

( $=11,6-10,4$ ), enquanto que na turma 9.º B, 50% dos alunos estão num intervalo de amplitude 3,1 ( $=12,5-9,4$ ).

O comportamento das duas turmas é visível quando se comparam os dois diagramas seguintes:



É evidente a existência de um comportamento diferente das duas turmas: uma maior variabilidade na turma 9.º B, com alguns alunos a terem notas piores e também alguns alunos a terem notas melhores.

### 5.2.5 Percentis<sup>3</sup>

Os percentis de que a mediana e os quartis são casos particulares, são medidas de localização com grande interesse, nomeadamente para avaliar a posição relativa dos dados. Por exemplo, suponha que uma mãe vai, com o seu bebé de 6 meses, à consulta de rotina, do pediatra. Este, depois de pesar e medir a criança, consulta umas tabelas e só nessa altura comenta com a mãe, o estado de crescimento do seu filho. Pode acontecer que alguns dos seus comentários sejam desta forma:

-Minha senhora, o seu filho, no que diz respeito ao peso, está no *percentil* 90. Vamos ter que ter algum cuidado!

Afinal o que significa o percentil 90? Significa que 90% das crianças com 6 meses têm um peso menor ou igual ao do bebé e só 10% têm um peso maior ou igual!

De um modo geral define-se **percentil p** de um conjunto de dados, como sendo o valor que tem **p%** dos dados menores ou iguais a ele, e os restantes maiores ou

<sup>3</sup> Esta secção, que foi integralmente retirada de Graça Martins et al (2007), inclui-se a título informativo, com um cunho complementar, uma vez que os percentis não fazem parte do programa.

iguais. O 1.º e o 3.º quartis também são conhecidos como percentil 25% e 75%, respectivamente. Analogamente, a mediana é o percentil 50%.

**Exemplo – A obesidade é um problema.** A comunicação social tem alertado a opinião pública para o problema da obesidade, nomeadamente nas crianças. Então, como é que poderemos saber se o nosso filho está obeso? Como é que o médico, além da sua experiência, sossega a mãe sobre a saúde e bem estar do seu filho? Existem tabelas, que apresentam, para cada idade, os valores dos percentis para as variáveis “peso” e “altura”. A tabela seguinte, que se retirou da Internet, apresenta, para os vários meses de idade, valores adequados, entre os quais deve estar o peso (em kg) da criança. Estes valores poderiam ser, por exemplo, os percentis 25% e 75%<sup>4</sup>, considerando-se um “peso normal”, aquele que se encontra nos 50% dos pesos centrais, quando se considera o conjunto dos pesos dos bebés (da população que se está a estudar, quer seja portuguesa, italiana, inglesa, alemã, etc.) com determinada idade:

	Ao nascer	1 mês	2 meses	3 meses	4 meses	5 meses	6 meses	7 meses	8 meses
Mínimo	2.750	3.500	4.000	4.750	5.500	6.000	6.500	7.000	7.500
Máximo	4.000	5.000	6.000	7.000	7.800	8.500	9.000	9.700	10.000
	9 meses	10 meses	11 meses	1 ano	1 ano 1m	1 ano 2m	1 ano 3m	1 ano 4m	1 ano 5m
Mínimo	7.900	8.300	8.500	8.800	9.000	9.250	9.500	9.700	9.800
Máximo	10.500	10.900	11.250	11.500	11.800	12.000	12.400	12.600	12.800
	1 ano 6m	1 ano 7m	1 ano 8m	1 ano 9m	1 ano 10m	1 ano 11m	2 anos	2 anos 1m	2 anos 2m
Mínimo	10.000	10.150	10.300	10.500	10.600	10.700	10.900	11.000	11.200
Máximo	13.000	13.300	13.600	13.800	14.000	14.200	14.500	14.650	14.800
	2 anos 3m	2 anos 4m	2 anos 5m	2 anos 6m	2 anos 7m	2 anos 8m	2 anos 9m	2 anos 10m	2 anos 11m
Mínimo	11.300	11.500	11.600	11.750	11.900	12.000	12.100	12.250	12.400
Máximo	15.000	15.250	15.500	15.700	15.900	16.000	16.300	16.500	16.750
	3 anos	3 anos 4m	3 anos 8m	4 anos	4 anos 4m	4 anos 8m	5 anos		
Mínimo	12.600	13.200	13.750	14.300	15.000	15.500	16.000		
Máximo	17.000	17.700	18.500	19.300	20.200	21.000	21.800		

A partir da tabela anterior, concluímos que um peso razoável, nem muito magro, nem muito gordo, para um bebé de 2 anos e meio, será um peso compreendido no intervalo [11,750kg, 15,700kg].

**Exemplo – Conversa entre mãe e filho.** Imagine a seguinte conversa entre uma mãe e o seu filho de 15 anos.

*Filho* - Mãe, tive 14 no teste de Biologia!

*Mãe* - E então isso é bom ou nem por isso?

*Filho* - Como assim? Digo que tive 14 e ainda me perguntas se isso é bom?

*Mãe* - Pois, pergunto... E até pergunto a que percentil é que corresponde essa nota?

*Filho* - Mas o que é isso de percentil? Não sei do que estás a falar!

*Mãe* - Quantos alunos na tua escola fizeram esse teste?

*Filho* - Foram 100, porquê?

<sup>4</sup> Na apresentação da tabela não é indicado o que representam o mínimo e o máximo.

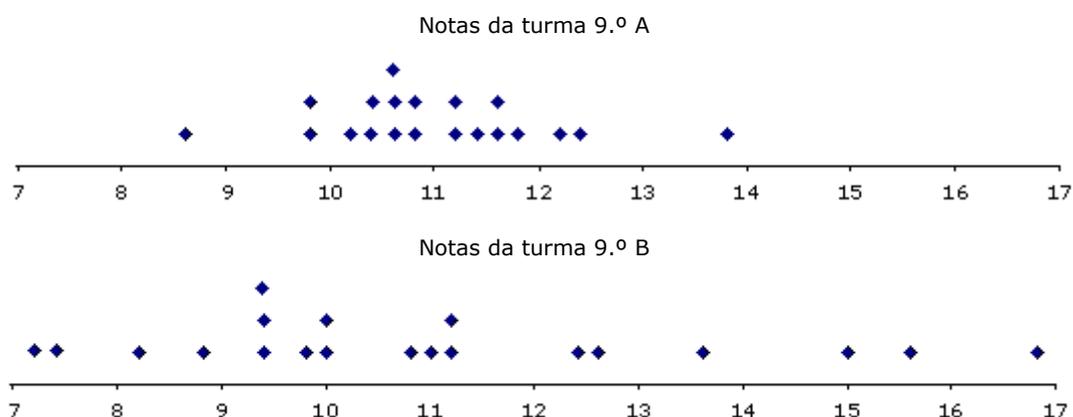
*Mãe* – E quantos tiveram nota maior que 14?

*Filho* – Bom, não vi bem, mas parece-me que foram uns 80!

*Mãe* – Afinal, não tens razão para estar tão satisfeito! Ficaste no percentil 20. Só 20% dos teus colegas tiveram nota menor ou igual à tua. Esse exame foi mesmo muito fácil.

### 5.3 Medidas de variabilidade

Consideremos de novo as notas dos alunos das turmas 9.º A e 9.º B, consideradas na secção 5.2.2 e de novo na secção 5.2.4. Nos estudos e discussões envolvendo estes dois conjuntos de dados, verificámos que apesar de apresentarem a mesma média, têm um comportamento bastante diferente, no que diz respeito à variabilidade. Como também adiantámos, as notas da turma 9.º B apresentam, uma maior variabilidade ou dispersão. Representando, de novo, os dois conjuntos na forma de diagramas de pontos



mais uma vez, é visível que a distribuição das notas da turma 9.º B apresenta uma maior variabilidade que a distribuição das notas da turma do 9.º A, pelo que se levanta o problema de arranjar uma medida que possa ser utilizada para medir essa maior ou menor variabilidade e que possa caracterizar os dois conjuntos de dados, com distribuições tão diferentes, mas com a mesma média.

Apresentamos a seguir as medidas de variabilidade mais vulgarmente utilizadas e que são a amplitude, a amplitude interquartil e o desvio padrão.

#### 5.3.1 Amplitude

A amplitude é a medida mais simples que pode ser utilizada para medir a variabilidade apresentada por um conjunto de dados. Obtém-se fazendo a diferença entre o máximo e o mínimo dos dados:

$$\text{Amplitude} = \text{máximo} - \text{mínimo}$$

No caso das notas das turmas 9.º A e 9.º B, temos que a

$$\begin{aligned} \text{Amplitude (notas da turma 9.º A)} &= 13,8 - 8,6 \\ &= 5,2 \end{aligned}$$

$$\begin{aligned} \text{Amplitude (notas da turma 9.º B)} &= 16,8 - 7,2 \\ &= 9,6 \end{aligned}$$

Como se esperava, a turma 9.º B apresenta uma amplitude maior.

Esta medida, muito simples de calcular, pode também ser muito enganadora. É baseada em dois únicos dados, que podem ser muito atípicos na distribuição de todos os dados do conjunto. É uma medida muito “pouco resistente”, pois depende muito da existência de valores muito “pequenos” ou muito “grandes”, a que demos o nome de “outliers”, no nosso conjunto de dados. É uma medida que normalmente não é utilizada, sobretudo se a distribuição dos dados apresentar enviesamento ou “outliers”.

### 5.3.2 Amplitude interquartil

Uma outra medida de variabilidade, alternativa à amplitude, é a amplitude interquartil. Esta medida, ao contrário da amplitude definida anteriormente, só entra em linha de conta com a parte central dos dados e calcula-se fazendo a diferença entre o 3.º e o 1.º quartis. A amplitude interquartil, que já foi utilizada na construção do diagrama de extremos e quartis, dá-nos informação sobre a amplitude do intervalo que contém 50% dos dados centrais. Esta informação, é visível no diagrama de extremos e quartis pelo comprimento da caixa.

**Amplitude interquartil** = Diferença entre o 3.º quartil e o 1.º quartil

Para as notas das turmas 9.º A e 9.º B, temos

$$\begin{aligned} \text{Amplitude interquartil (notas da turma 9.º A)} &= 11,6 - 10,4 \\ &= 1,2 \\ \text{Amplitude interquartil (notas da turma 9.º B)} &= 12,5 - 9,4 \\ &= 3,1 \end{aligned}$$

Como se verifica e era espectável, a amplitude interquartil é superior para as notas da turma 9.º B.

Na interpretação da amplitude interquartil tem que se ter em atenção que uma amplitude interquartil nula não significa, necessariamente, a não existência de variabilidade. Por exemplo os seguintes dados

10 11 13 14 14 14 14 14 14 14 14 14 14 16 17 18

apresentam variabilidade, mas, no entanto, a amplitude interquartil é nula. Efectivamente o 1.º e 3.º quartis são iguais a 14, fazendo com que a diferença entre os quartis venha igual a zero.

Propriedades da amplitude interquartil

- A amplitude interquartil será tanto maior, quanto maior for a variabilidade presente nos dados;
- Se não houver variabilidade, isto é, se os dados forem todos iguais, então a amplitude interquartil vem igual a zero;
- No entanto, se a amplitude interquartil de um conjunto de dados for nula, não significa necessariamente que não haja variabilidade;
- A amplitude interquartil é uma medida de variabilidade que se utiliza frequentemente, sobretudo se os dados apresentarem algum enviesamento ou “outliers”.

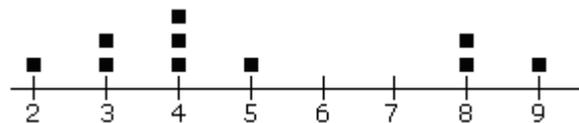
### 5.3.3 Outras medidas de variabilidade: O desvio médio absoluto e o desvio padrão

Quando estudámos as medidas de localização do centro da distribuição dos dados dissemos que as mais utilizadas são a média e a mediana. Então, ao pesquisarmos medidas de variabilidade, é natural que procuremos medidas que meçam a variabilidade relativamente a estas medidas de localização, que representam valores “típicos” da distribuição dos dados.

Se falarmos, por exemplo, na *altura média* de um conjunto de alunos, é natural querermos saber qual a variabilidade das alturas desses alunos relativamente a essa média. Terão os alunos alturas semelhantes? Terão alturas muito afastadas da média? Como medir essa variabilidade?

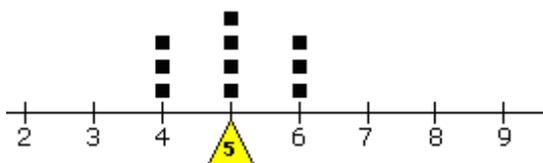
Para introduzir a noção de variabilidade, vamos retomar um exemplo da secção 5.2.1:

**Tarefa – Desvios entre os dados e a média (cont.).** Na turma o professor perguntou a 10 alunos que tinham o estojo de lápis em cima da mesa, quantos lápis (incluindo lápis de cor) tinham no estojo. As respostas obtidas 2, 3, 3, 4, 4, 4, 5, 8, 8, 9, encontram-se representadas no seguinte diagrama de pontos:

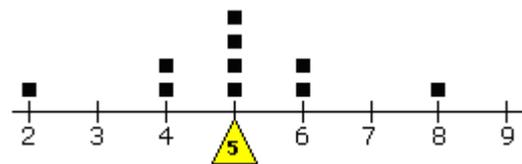


Vimos que a média dos valores considerados é igual a 5, mas também vimos que havia outras configurações, ou outros conjuntos de 10 dados que mantinham a mesma média. Alguns desses conjuntos apresentam-se a seguir:

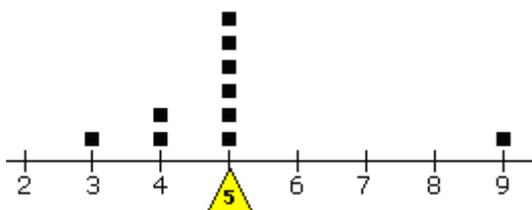
1.



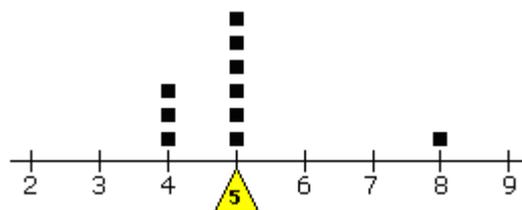
2.



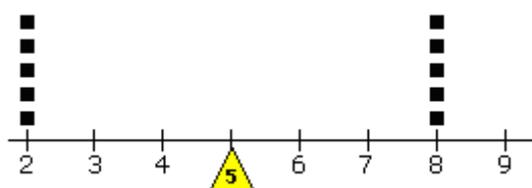
3.



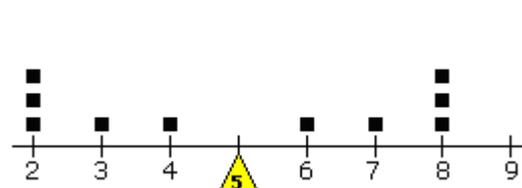
4.



5.



6.



Qualquer dos conjuntos representados anteriormente apresenta variabilidade. Qual o que apresenta maior variabilidade? E que tipo de variabilidade estamos a pretender medir? Se repararmos no conjunto de dados correspondente à representação 5 e 6, respectivamente os dados

2 2 2 2 2 8 8 8 8 8

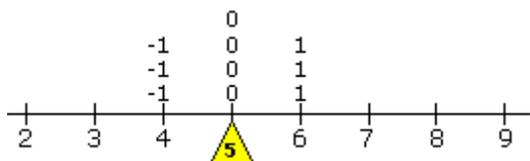
e os dados

2 2 2 3 4 6 7 8 8 8

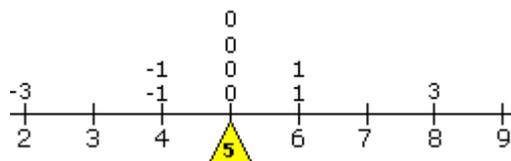
verificamos que no primeiro caso existem 5 valores iguais a 2 e 5 valores iguais a 8, enquanto que no segundo caso só dois dos valores é que se repetem e com menor frequência. Qual dos dois conjuntos apresenta maior variabilidade? Se pedirmos a um grupo de alunos para responderem a esta questão, dir-nos-ão que o primeiro conjunto apresenta menor variabilidade, pois intuitivamente estão a interpretar a variabilidade em termos de "mais ou menos iguais, uns relativamente aos outros", independentemente de considerarem um ponto padrão como referência, nomeadamente a média.

Vejamus então como medir a variabilidade de cada um dos conjuntos representados anteriormente, considerando como ponto de referência a média. Começemos por, em cada representação, substituir o ponto que representa o dado, pelo seu desvio para a média. Estes desvios serão positivos se os valores forem inferiores à média e negativos se forem superiores à média:

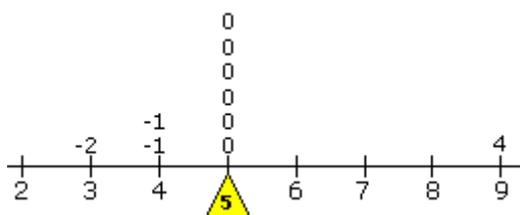
1.



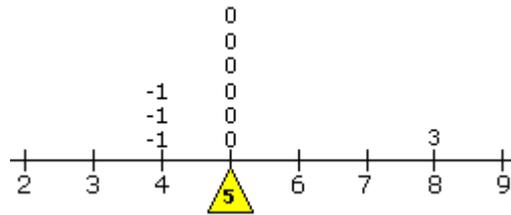
2.



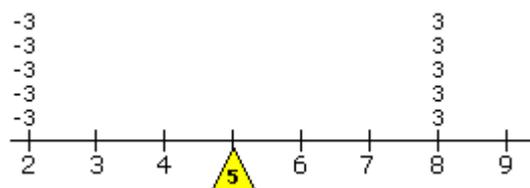
3.



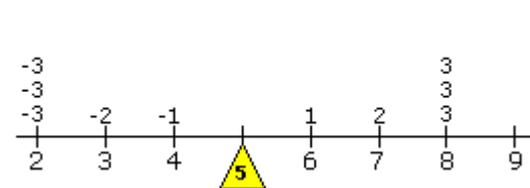
4.



5.



6.



Já sabemos que a soma dos desvios positivos é igual à soma dos desvios negativos, pelo que vamos considerar uma medida de variabilidade que entre em linha de con-

ta com a soma dos desvios em valor absoluto. Para cada um dos conjuntos de dados anteriores temos:

1.

Dados: 4, 4, 4, 5, 5, 5, 5, 6, 6, 6

Valor	Desvio para a média	Desvio para a média em valor absoluto
4	-1	1
4	-1	1
4	-1	1
5	0	0
5	0	0
5	0	0
5	0	0
6	1	1
6	1	1
6	1	1
Total	0	6

$$\text{Média dos desvios absolutos} = \frac{6}{10} = 0,6$$

2.

Dados: 2, 4, 4, 5, 5, 5, 5, 6, 6, 8

Valor	Desvio para a média	Desvio para a média em valor absoluto
2	-3	3
4	-1	1
4	-1	1
5	0	0
5	0	0
5	0	0
5	0	0
6	1	1
6	1	1
8	3	3
Total	0	10

$$\text{Média dos desvios absolutos} = \frac{10}{10} = 1,0$$

3.

Dados: 3, 4, 4, 5, 5, 5, 5, 5, 9

Valor	Desvio para a média	Desvio para a média em valor absoluto
3	-2	2
4	-1	1
4	-1	1
5	0	0
5	0	0
5	0	0
5	0	0
5	0	0
5	0	0
9	4	4
Total	0	8

$$\text{Média dos desvios absolutos} = \frac{8}{10} = 0,8$$

4.

Dados: 4, 4, 4, 5, 5, 5, 5, 5, 8

Valor	Desvio para a média	Desvio para a média em valor absoluto
4	-1	1
4	-1	1
4	-1	1
5	0	0
5	0	0
5	0	0
5	0	0
5	0	0
8	3	3
Total	0	6

$$\text{Média dos desvios absolutos} = \frac{6}{10} = 0,6$$

5.

Dados: 4, 4, 4, 5, 5, 5, 5, 6, 6, 6

Valor	Desvio para a média	Desvio para a média em valor absoluto
2	-3	3
2	-3	3
2	-3	3
2	-3	3
2	-3	3
8	3	3
8	3	3
8	3	3
8	3	3
8	3	3
Total	0	30

$$\text{Média dos desvios absolutos} = \frac{30}{10} = 3,0$$

6.

Dados: 2, 2, 2, 3, 4, 6, 7, 8, 8, 8

Valor	Desvio para a média	Desvio para a média em valor absoluto
2	-3	3
2	-3	3
2	-3	3
3	-2	2
4	-1	1
6	1	1
7	2	2
8	3	3
8	3	3
8	3	3
Total	0	10

$$\text{Média dos desvios absolutos} = \frac{24}{10} = 2,4$$

Para cada um dos conjuntos de dados calculamos a média dos desvios em valor absoluto e vamos considerar esta medida, a que chamamos **desvio médio absoluto**, como uma medida da variabilidade dos nossos dados. Comparando os resultados obtidos, verificamos que o conjunto de dados que apresenta maior variabili-

dade é o 5., enquanto que o 1. e o 4. apresentam a menor variabilidade, igual a 0,6. Este resultado a que chegámos é de certo modo esperado, pois reparando nas diferentes representações gráficas, verificamos que estes conjuntos são os que apresentam a distribuição com os dados mais perto da média.

**Desvio médio absoluto** - Consideremos uma amostra  $(x_1, x_2, \dots, x_n)$  com média  $\bar{x}$ . Para medir a variabilidade dos dados relativamente à média, começa-se por calcular, para cada dado, a diferença entre ele e a média, a que chamamos desvio:

$$x_1 - \bar{x}, x_2 - \bar{x}, x_3 - \bar{x}, \dots, x_n - \bar{x}$$

Para obter a variabilidade de todos os dados, seria natural somar todos os desvios. Acontece que, como já vimos, a soma destes desvios é sempre igual a zero, pelo que esta solução não serve. Então, vamos considerar não os próprios desvios, mas os seus valores absolutos:

$$|x_1 - \bar{x}|, |x_2 - \bar{x}|, |x_3 - \bar{x}|, \dots, |x_n - \bar{x}|$$

Define-se desvio médio absoluto como sendo a média destes desvios absolutos:

$$\text{Desvio médio absoluto} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$

Recordemos que a substituição dos desvios pelos seus valores absolutos foi devida ao facto de a soma dos desvios ser igual a zero, uma vez que a soma dos desvios positivos cancela com a soma dos desvios negativos. Uma alternativa a considerar os *módulos dos desvios*, consiste em considerar os *quadrados dos desvios* e em construir uma outra medida à custa de uma média destes quadrados. Intuitivamente esta medida, a que vamos chamar variância, não nos parece uma boa alternativa, pois resulta uma medida cujas unidades são o quadrado das unidades originais dos dados. Este inconveniente é ultrapassado se utilizarmos como medida de variabilidade a raiz quadrada da variância, a que damos o nome de **desvio padrão**.

**Desvio padrão** - Consideremos então a amostra  $(x_1, x_2, \dots, x_n)$  com média  $\bar{x}$ . Para medir a variabilidade dos dados relativamente à média, começa-se por calcular, para cada dado, a diferença entre ele e a média:

$$x_1 - \bar{x}, x_2 - \bar{x}, x_3 - \bar{x}, \dots, x_n - \bar{x}$$

Para obter a variabilidade de todos os dados, vamos considerar não os próprios desvios, mas os seus quadrados:

$$(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, (x_3 - \bar{x})^2, \dots, (x_n - \bar{x})^2$$

Define-se **variância** e representa-se por  $s^2$ , a medida que se obtém somando os quadrados dos desvios e dividindo pelo número de observações menos uma:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

Para que a medida da variabilidade venha na mesma unidade dos dados originais, a media que se considera é  $s$ , a raiz quadrada da variância, a que se dá o nome de **desvio padrão**

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

## Notas

1. No processo que leva à construção da variância, o motivo que nos leva a considerar os quadrados dos desvios, assim como anteriormente, na definição do desvio médio absoluto se consideraram os desvios absolutos, já tem uma explicação – a soma dos desvios das observações para a média, é sempre igual a zero, pelo que ou consideramos os desvios em valor absoluto, ou os seus quadrados.

2. Mas então porque é que não consideramos a média desses desvios ao quadrado, dividindo a sua soma por  $n$  em vez de  $(n-1)$ , como está proposto? A este nível, a resposta que pode ser dada prende-se de certo modo com o motivo que nos levou a considerar os quadrados, em vez dos próprios desvios: como a soma dos  $n$  desvios é igual zero, basta conhecer  $(n-1)$  desses desvios, para que o  $n$ -ésimo fique automaticamente determinado. Assim, como só temos  $(n-1)$  desvios independentes, dividimos por  $(n-1)$  em vez de  $n$ . Esta é uma forma “simplista” de abordar o problema, pois esta mesma razão levar-nos-ia a considerar para o desvio médio absoluto o quociente da soma dos desvios absolutos por  $n-1$ , em vez de ser por  $n$ , como fizemos. Vamos então adiantar um pouco mais a explicação, embora corramos o risco de a tornar mais complicada.... O que acontece é que em Estatística, normalmente o nosso objectivo é estudar populações a partir de amostras recolhidas dessas populações. Quando se recolhe uma amostra, procede-se ao seu estudo gráfico para tentar obter a estrutura ou padrão da distribuição da população de onde se retirou a amostra e arranjar um modelo para essa população, e também se calculam algumas características amostrais, que pretendem “estimar” características populacionais correspondentes, que são os parâmetros. Por exemplo para conhecer o parâmetro – *valor médio* das alturas ou *altura média* dos portugueses adultos (população constituída pelas alturas de todos os portugueses adultos), recolhe-se uma amostra, uma vez que é impraticável observar a altura de todos os portugueses e calcular a média. A partir da amostra recolhida, calculamos a média e este valor é um valor aproximado do parâmetro *altura média* pretendido. Dizemos que a *média* da amostra observada é uma *estimativa* do *valor médio* da população de onde se observou a amostra. Ora o mesmo se passa se pretendermos conhecer o parâmetro *variância* da população, também chamado de *variância populacional*. Para estimar este parâmetro, calculamos a variância da amostra observada e é agora que chegámos ao ponto crucial: pode-se mostrar que a variância que se calcula a partir da amostra, dividindo a soma dos quadrados dos desvios por  $n-1$  dá uma “melhor” estimativa da variância populacional, do que se dividirmos por  $n$ . Como em Estatística, de um modo geral o nosso objectivo é fazer Inferência Estatística, isto é, inferir propriedades da população, a partir das propriedades verificadas na amostra, convém que, no caso de estarmos a estimar parâmetros, que as estimativas sejam as melhores possíveis. Embora as “boas” propriedades manifestadas pela variância amostral, quando se pretende estimar a variância populacional, quando se considera  $n-1$ , já não se verifiquem quando calculamos o desvio padrão, mesmo assim, ainda é preferível considerar  $n-1$  em vez de  $n$ .

3. Alguma literatura apresenta o desvio padrão  $s'$ , considerando  $n$  em vez de  $n-1$ :

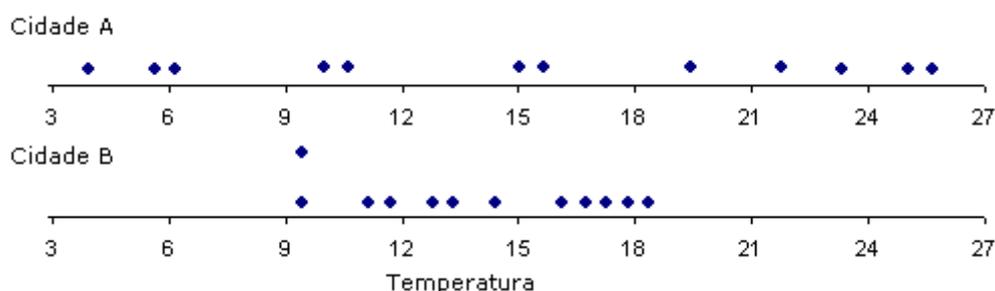
$$s' = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

Devemos acrescentar que este procedimento embora não esteja correcto, conduz a uma estimativa que também pode ser utilizada, desde que  $n$  seja suficientemente grande, ou seja, quando  $\frac{n}{n-1} \approx 1$ , pois neste caso  $s' \approx s$ .

No que diz respeito ao desvio médio absoluto, não nos vamos preocupar se se divide por  $n$  ou por  $n-1$ , pois a medida que efectivamente é utilizada em Estatística, como medida da variabilidade de uma amostra é o desvio padrão e não o desvio médio absoluto. Por isso, as razões invocadas para o desvio padrão e que se prendem com a Inferência Estatística, não têm aqui cabimento.

**Tarefa – Temperaturas de duas cidades**<sup>5</sup>. Na seguinte tabela são apresentadas as temperaturas ( $^{\circ}\text{C}$ ) médias mensais das cidades A e B:

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Cidade A	3,9	5,6	10,0	15,0	19,4	23,3	25,6	25,0	21,7	15,6	10,6	6,1
Cidade B	9,4	11,1	11,7	13,3	14,4	16,7	17,2	17,8	18,3	16,1	12,8	9,4



- A partir dos dados da tabela calcula a mediana das temperaturas médias mensais das duas cidades. Antes de calcular a mediana tens que ter alguma preocupação prévia com os dados? As medianas das temperaturas das duas cidades, são próximas?
- Uma vez que os centros das distribuições dos dois conjuntos de dados estão muito próximos, podemos concluir que não existe diferença entre estas duas cidades, no que diz respeito às temperaturas médias mensais?
- Qual das cidades apresenta maior variabilidade nas suas temperaturas mensais? (Responde observando os diagramas de pontos)
- Qual a temperatura mensal mais alta para a cidade A? E a mais baixa? Qual a diferença de temperaturas? Que nome dás a esta diferença?
- Calcula a amplitude das temperaturas mensais da cidade B.
- Considera as temperaturas da cidade A, inferiores à mediana e calcula a mediana do conjunto de temperaturas considerado. Que nome dás a este valor?
- Faz o mesmo que na alínea anterior, mas agora com as temperaturas superiores à mediana. Que nome dás ao valor obtido?
- Verifica que um quarto dos valores são inferiores ou iguais ao 1.º quartil, um quarto são superiores ou iguais ao 3.º quartil e metade estão entre o 1.º e o 3.º quartis.
- Calcula a amplitude interquartil da distribuição das temperaturas mensais da cidade A.
- A amplitude interquartil da distribuição das temperaturas mensais da cidade A é superior à amplitude interquartil da distribuição das temperaturas mensais da cidade B, que é  $5,55^{\circ}\text{C}$ ? A resposta que deste à alínea c) já previa a conclusão a que chegaste?

<sup>5</sup> Adaptado de Rossmann et al. (2001), p. 83.

- k) A tabela seguinte apresenta algumas estatísticas relativas à distribuição das temperaturas mensais da cidade B. Completa-a com as estatísticas correspondentes da distribuição das temperaturas mensais da cidade A:

	Mínimo	1º quartil	Mediana	3º quartil	máximo
Cidade A					
Cidade B	9	11,40	13,85	16,95	18

- l) Constrói diagramas de extremos e quartis paralelos para as temperaturas das duas cidades, e compara as duas cidades no que diz respeito às temperaturas médias mensais.
- m) A média das temperaturas mensais da cidade A é 15,15°C. Completa os dois valores em falta na coluna "desvio para a média" da tabela seguinte:

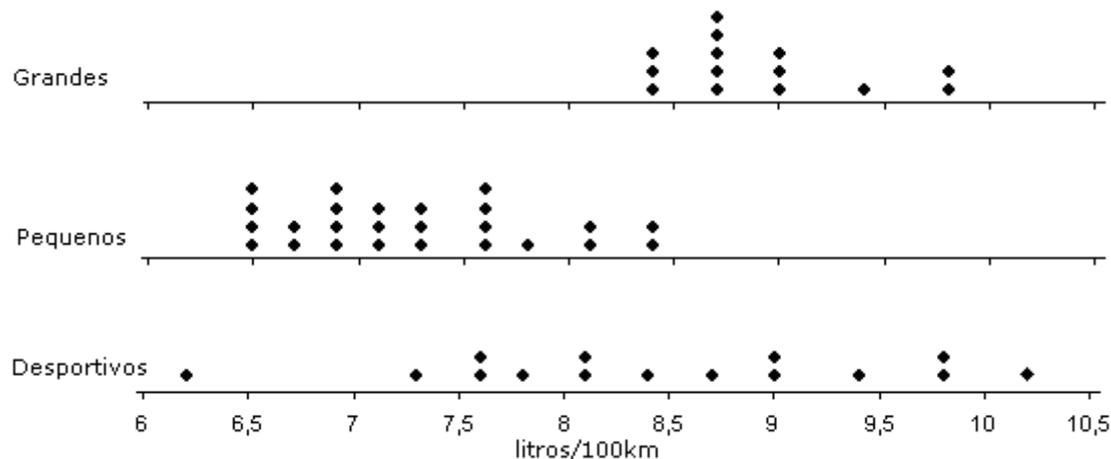
Mês	Temperatura	Desvio para a média	Valor absoluto do desvio	Quadrado do desvio
Jan	3,9	-11,25	11,25	126,5625
Fev	5,6	-9,55	9,55	91,2025
Mar	10,0			
Abr	15,0	-0,15	0,15	0,0225
Mai	19,4	4,25	4,25	18,0625
Jun	23,3	8,15	8,15	66,4225
Jul	25,6	10,45	10,45	109,2025
Ago	25,0			
Set	21,7	6,55	6,55	42,9025
Out	15,6	0,45	0,45	0,2025
Nov	10,6	-4,55	4,55	20,7025
Dez	6,1	-9,05	9,05	81,9025
Total	181,8			

Depois de preencheres a coluna "Desvio para a média" com os valores que faltam, calcula a soma dos 12 desvios. Ficaste admirado com o valor a que chegaste? Porquê?

- n) Para obter uma medida da variabilidade dos dados, vamos trabalhar com as distâncias dos valores relativamente à média, pelo que vamos considerar os desvios em valor absoluto. Completa a coluna dos desvios em valor absoluto e depois calcula a soma dos valores dessa coluna. Consegues dizer, só fazendo uma operação de dividir, a que é igual a soma dos desvios positivos?
- o) Calcula a média dos valores absolutos dos desvios. Que nome dás à medida obtida?
- p) Uma alternativa a calcular os desvios para a média em valor absoluto, é considerar os quadrados dos desvios. Completa a coluna dos quadrados dos desvios e depois calcula a soma desses quadrados.
- q) Divide a soma obtida na alínea anterior por 11 (dimensão da amostra menos 1). Que nome dás à medida obtida?
- r) Para converter a medida obtida na alínea anterior, às unidades originais dos dados, calcula a sua raiz quadrada. Que nome dás à medida obtida?
- s) Com o auxílio da máquina de calcular ou da folha de Excel do computador, calcula o desvio padrão das temperaturas mensais da cidade B. Compara com o desvio padrão obtido para as temperaturas mensais da cidade A. Qual o maior? Era o que esperavas?

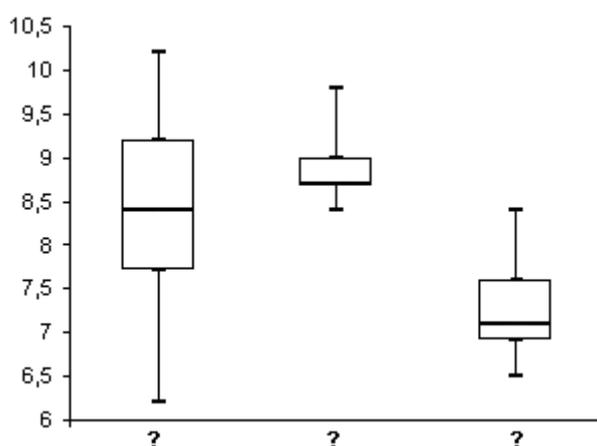
Sugestão – esta tarefa pode ser aproveitada para comparar as temperaturas registadas pelos alunos em duas cidades, como por exemplo, Lisboa e Porto.

**Tarefa – Vamos comparar os consumos dos carros.** Para comparar o consumo médio de três tipos de carros, nomeadamente carros de família (Grandes), carros utilitários (Pequenos) e carros Desportivos, recolheu-se informação junto de algumas marcas de carros, do consumo médio de gasolina, por 100km. Os resultados obtidos são apresentados no seguinte diagrama de pontos:



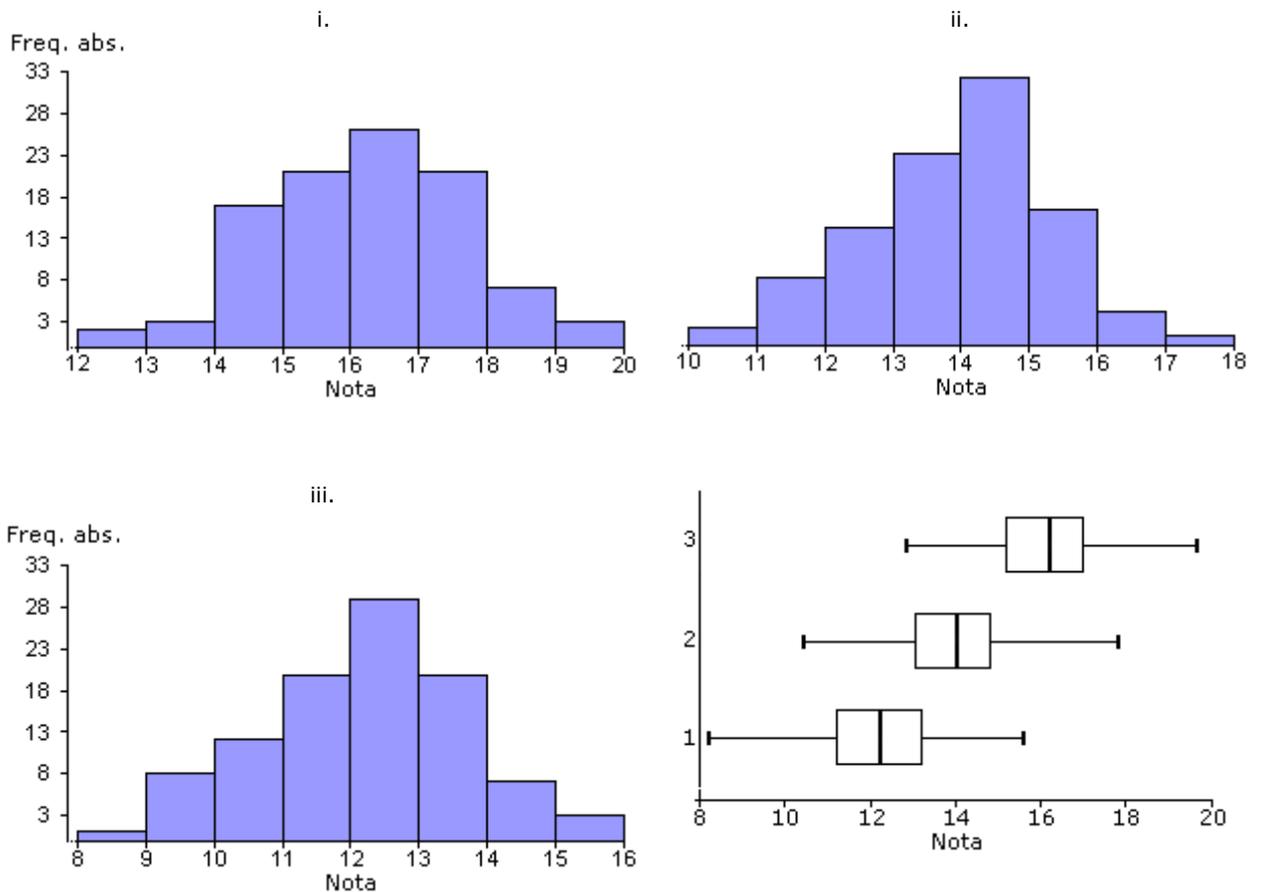
Para cada uma destas três amostras de carros, calculou-se a amplitude interquartil e o desvio padrão e construiu-se o diagrama de extremos e quartis. A partir das representações gráficas anteriores e sem fazeres quaisquer cálculos, associa cada par de estatísticas e cada diagrama de extremos e quartis a cada um dos tipos de carros considerados.

Amplitude interquartil	0,3	1,5	0,7
Desvio padrão	0,59	1,09	0,47
Tipo	?	?	?



**Tarefa – Vamos comparar as notas no mesmo teste de Inglês de alunos de três escolas diferentes.** De cada uma de três escolas da zona de Lisboa, A, B e C, foram seleccionados 100 alunos e registaram-se as notas obtidas no exame nacional de Língua Portuguesa. O professor pediu a três grupos de alunos que resumissem a informação contida nos dados. Um dos grupos calculou algumas características amostrais e os outros dois grupos fizeram representações gráficas:

	A	B	C
Média	16,1	12,1	13,9
Amplitude interquartil	1,85	2,05	1,8
Desvio padrão	1,4	1,51	1,4



a) Completa a seguinte tabela, estabelecendo as correspondências entre as características amostrais e as representações gráficas construídas

Características amostrais	A	B	C
Histograma			ii
Diagrama de extremos e quartis	3		

b) Qual o aspecto mais relevante sobre o comportamento a Inglês dos alunos das três escolas?



# 6 Probabilidade

Neste capítulo fazemos uma introdução à Probabilidade experimental de um acontecimento, obtida através da repetição de um número grande de vezes da experiência conducente à realização desse acontecimento. Considera-se também a Probabilidade teórica como modelo probabilístico em situações especiais de simetria. Ainda utilizando estes modelos, dão-se indicações sobre o cálculo de probabilidades de alguns acontecimentos.



## 6.1 Introdução<sup>1</sup>

A probabilidade, como acontece com muitas outras noções que usamos com frequência, não é fácil de definir, a menos que estejamos em condições de recorrer a conceitos matemáticos precisos. No entanto, sabemos usá-la com perícia, em muitas situações práticas, mesmo sem disso nos apercebermos. Qualquer um de nós, em face de um determinado acontecimento futuro, é capaz de fazer conjecturas sobre a probabilidade da sua realização. Quantas vezes nos ouvimos fazer afirmações do género “É muito provável que...”, “É pouco provável que...”, “É mais provável que...”. Embora os juízos probabilísticos que exprimimos sejam, a maior parte das vezes, em termos comparativos, há situações em que nos sentimos seguros em atribuir um valor numérico à possibilidade da realização de um determinado acontecimento. Por exemplo, se nos perguntarem qual a probabilidade de existir um homem com três metros de altura, respondemos certamente que essa probabilidade é zero, já que o nosso conhecimento nos faz acreditar que esse acontecimento é impossível. Por outro lado, se nos perguntarem qual a probabilidade de o Sol nascer amanhã, não temos dúvida em afirmar que é um. Se pretendermos decidir quem, entre duas pessoas deve fazer um determinado trabalho, podemos fazer a escolha atirando uma moeda ao ar. Neste caso assumimos implicitamente que, procedendo deste modo, estamos a ser justos já que atribuímos probabilidades iguais (na escala de 0 a 1 corresponderia a  $\frac{1}{2}$ ) a cada um de poder vir a realizar o dito trabalho.

O termo Probabilidade é utilizado todos os dias de forma mais ou menos intuitiva, pois nos mais variados aspectos da nossa vida, está presente a *incerteza*:

- Se não avistarmos nuvens, dizemos que é pouco provável que chova;
- Dizemos que a probabilidade do próximo bebé, de uma determinada família, ser do sexo masculino é aproximadamente 50%;
- Dizemos que a probabilidade de lançar uma moeda de 1 euro ao ar e sair a face com o 1, é 50%;
- Dizemos que a probabilidade de ganhar no Euromilhões é quase nula;
- O político interroga-se sobre qual a probabilidade de ganhar as próximas eleições;
- O aluno interroga-se sobre qual a probabilidade de obter nota positiva num teste de respostas múltiplas, para o qual não se preparou e responde sistematicamente ao acaso;
- Para tratar determinada doença, o médico pretende saber se um novo medicamento oferece maior probabilidade de cura que o medicamento habitual;
- O fabricante desejaria saber se um produto que pretende lançar no mercado, terá uma boa probabilidade de aceitação;
- Numa fábrica, o departamento de controlo de qualidade pretende averiguar a probabilidade de uma máquina não avariar no próximo ano;
- Um empresário têxtil precisa de saber qual a probabilidade de conseguir vender camisas de homem de tamanho maior ou igual que 45, para saber se deve fazer camisas destes números;
- etc.

---

<sup>1</sup> Nesta introdução seguimos de perto Graça Martins et al (1999) e Graça Martins et al (2007)

Todos estes exemplos têm uma característica comum, que é o facto de *não se conseguir prever com exactidão e de antemão*, qual o resultado da situação de incerteza. Perante as várias possibilidades que se nos apresentam, não sabemos qual a que se vai verificar. Ao emitirmos um juízo de valor, como fizemos em alguns dos exemplos considerados, não estamos mais do que a anunciar o nosso *grau de convicção* na realização de algum acontecimento. Para exprimir esta convicção estamos a recorrer, embora intuitivamente, à *frequência relativa* com que o acontecimento se pode repetir.

A probabilidade está presente sempre que estivermos perante um *fenómeno aleatório*, isto é, um fenómeno para o qual não sabemos de antemão qual o resultado que se vai verificar, na próxima repetição (admite-se que o fenómeno se pode repetir), mas para o qual é possível verificar uma certa *regularidade a longo termo*, ou seja, para um grande número de repetições do fenómeno. É esta última característica do fenómeno aleatório que o distingue de um processo caótico, já que ambos têm a característica comum de não se conseguir antecipar com exactidão qual o resultado que se vai obter quando se realizam.

**Fenómenos aleatórios** – São fenómenos para os quais os resultados das realizações individuais são incertos, mas em que se admite ser possível encontrar um padrão genérico de comportamento.

São exemplos de fenómenos aleatórios aqueles que têm como resultado observável:

- A chave do totoloto em cada semana;
- A resposta de uma doença a um tratamento feito com determinado medicamento;
- O estado do tempo no dia seguinte;
- O comportamento dos eleitores nas próximas eleições legislativas;
- O comportamento de um aluno no exame de resposta múltipla, para o qual não estudou;
- O comportamento do mercado perante um produto novo para lavar a roupa;
- O comprimento do próximo bebé a nascer na cidade;
- etc.

## 6.2 Probabilidade empírica ou experimental de um acontecimento

Nos fenómenos determinísticos, conseguimos dizer o que vai acontecer quando o fenómeno se realiza, mas o mesmo não acontece com os fenómenos aleatórios. Não temos dúvidas de que ao lançar ao ar uma moeda de um euro, ela cai, mas não sabemos qual a face que fica virada para cima quando ela assenta no sítio em que caiu, quer seja o chão, a nossa mão ou em cima de uma mesa. No exemplo anterior o resultado "A moeda cai", obtido como consequência do lançamento da moeda ao ar, não é um resultado incerto, pois temos a certeza que a moeda não fica no ar<sup>2</sup>! No entanto, já é um resultado incerto "A face Euro fica virada para

---

<sup>2</sup> Estamos a pensar numa experiência feita num ambiente normal, no planeta Terra...

cima”, pois não temos a certeza que isso aconteça. A nossa curiosidade leva-nos então a tentar antecipar o que vai acontecer quando lançamos a moeda ao ar e exprimimos essa necessidade utilizando a linguagem da probabilidade ao dizer “A probabilidade de lançar uma moeda de um euro ao ar e sair a face Euro é 50%”.

À realização do fenómeno aleatório chamamos *experiência aleatória*. Assim, no caso do lançamento da moeda a experiência aleatória consiste em lançar a moeda ao ar e verificar qual a face que fica virada para cima. A experiência é *aleatória* porque não sabemos se é a face Euro ou a face Nacional que vai ficar virada para cima. Em contrapartida, a experiência que consiste em lançar a moeda ao ar e ver se cai, já não é aleatória!

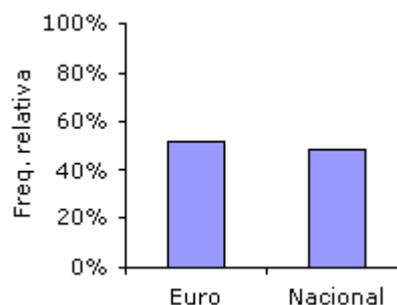
A repetição de experiências aleatórias associadas a determinado fenómeno aleatório é o processo utilizado para a *aquisição de dados*, que, uma vez analisados, nos permitem inferir propriedades do fenómeno aleatório em estudo.

Admitamos, por exemplo, que tínhamos uma moeda de um euro e que pretendíamos verificar se havia alguma razão para suspeitar que a moeda não era equilibrada, isto é, se seriam diferentes as possibilidades de sair a face Euro ou a face Nacional quando se lança a moeda ao ar. Para recolher dados que nos permitam responder à questão anterior, vamos repetir um grande número de vezes a experiência aleatória que consiste em lançar a moeda ao ar e verificar a face que fica voltada para cima. Suponhamos que após a repetição da experiência 50 vezes, se tinha observado a seguinte sequência, onde representamos por E a face Euro e por N a face Nacional:

E E E E E N N N E N E E N E E N E E E N E N E N N E  
E N N E N N N N E N E N N E N N E N N E E N E E N

Se resumirmos numa tabela de frequência os dados anteriores, obtemos o seguinte resultado:

Face virada para cima	Nº de vezes	Frequência relativa	Frequência relativa %
Euro	26	26/50	52%
Nacional	24	24/50	48%



Nestes 50 lançamentos, a face Euro ficou virada para cima 26 vezes, pelo que a proporção de vezes que se obteve a face Euro está próxima dos 50%. Intuitivamente somos levados a concluir que não temos razão para rejeitar o *modelo* que tínhamos idealizado, de que a moeda era equilibrada.

Repare-se que a situação descrita anteriormente é uma situação típica de uma *investigação estatística*:

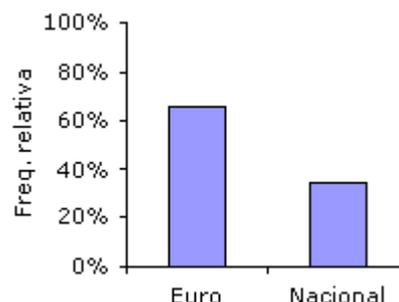
- ✓ Formula-se uma conjectura;
- ✓ Recolhem-se dados que permitam avaliar da veracidade dessa conjectura;
- ✓ Exprime-se uma posição sobre a veracidade ou não da conjectura.

Suponhamos, no entanto, que em vez da sequência anterior se tinham obtido os seguintes dados:

E E E E E N E N E E N E N E E E E E E E E N N E  
E N E E E E N N E N E N E E N N E N N E E N E E N

Ao resumir as observações anteriores numa tabela de frequência, obtemos o seguinte resultado:

Face virada para cima	Nº de vezes	Frequência relativa	Frequência relativa %
Euro	33	33/50	66%
Nacional	17	17/50	34%



Será que nestas circunstâncias ainda continuaríamos a dizer que "A probabilidade de lançar uma moeda de um euro ao ar e sair a face euro é 50%"?

A nossa intuição leva-nos a dizer que algo de errado se passa com a moeda e que deve ter algum defeito, pois se a moeda fosse *equilibrada* esperaríamos que a proporção de vezes que sai a face Euro fosse aproximadamente igual à proporção de vezes que sai a face Nacional, como se observou na primeira sequência. Como dissemos anteriormente, a *repetição* da experiência aleatória permitiu recolher dados que, uma vez analisados, nos permitem inferir propriedades do fenómeno aleatório em estudo e que, neste caso, se pode traduzir na seguinte conclusão sobre a conjectura "A moeda é equilibrada" e sobre o que acontece quando se lança a moeda ao ar:

- A moeda *não é equilibrada* e no próximo lançamento da moeda é mais provável sair a face Euro do que a face Nacional.

### Porque é que é necessário repetir a experiência um grande número de vezes?

Na definição de fenómeno aleatório diz-se que:

... quando o fenómeno se realiza, não se conhece o resultado que se vai obter,

**mas**, no entanto, verifica-se um *padrão genérico de comportamento* ou uma *regularidade a longo termo*.

Isto significa que à medida que formos repetindo a experiência aleatória associada a esse fenómeno, começamos a observar que o fenómeno tem um comportamento "previsível". É o que acontece no caso do lançamento da moeda "equilibrada". Não sabemos o que acontece em cada realização do fenómeno, mas ao fim de muitas realizações podemos afirmar que a proporção de vezes que se verifica a face Euro está próxima de 50%. Esta regularidade não se verifica quando repetimos a experiência um número pequeno de vezes.

Suponhamos, por exemplo, que só tínhamos realizado as primeiras 10 experiências cujos resultados foram:

E E E E E N N N E N

Repare-se que a proporção de vezes que se verificou a face Euro foi de 60%.  
 Vejamos ainda o que se passa com as restantes observações, mas consideradas em séries de 10:

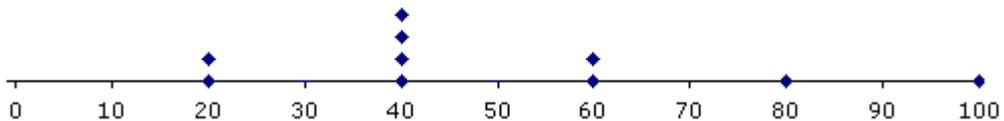
										Proporção de faces Euro
E	E	N	E	E	N	E	E	E	E	80%
N	E	N	N	E	E	N	N	E	N	40%
N	N	N	E	N	E	N	N	E	N	30%
N	E	N	N	E	E	N	E	E	N	50%

Como se verifica da tabela anterior, existe uma grande *variabilidade* na proporção de faces Euro nas sequências de 10 observações:



E se se tivessem considerado séries de 5 lançamentos? Vejamos o que acontece com a proporção de faces euro nas 10 sequências de 5 lançamentos:

Sequências	Proporção de faces Euro
E E E E E	100%
N N N E N	20%
E E N E E	60%
N E E E E	80%
N E N N E	40%
E N N E N	40%
N N N E N	20%
E N N E N	40%
N E N N E	40%
E N E E N	60%



Como se verifica do gráfico anterior, existe uma grande *variabilidade* nas proporções de vezes que surge a face Euro, no lançamento da moeda 5 vezes.

Voltemos novamente à situação dos 50 lançamentos da moeda. Se voltássemos a repetir a experiência outras 50 vezes iríamos obter a mesma percentagem de faces Euro? Não necessariamente, mas o melhor é confirmar. Obviamente que estar a repetir a experiência 50 vezes é bastante maçador, mas enquanto não soubermos como rodear este problema *simulando* a experiência, assunto a tratar mais à frente, vamos recolher várias *amostras* de dimensão 50 (os dados de cada amostra obtêm-se repetindo a experiência de lançar a moeda 50 vezes) e registar a percentagem de faces Euro, obtidas:

1ª amostra

E E E N E N N E N N N N N E E E N E N N N E N N N  
 E N N N E E N E N E E N E E E N E N E N E N E N E

2ª amostra

N E N E E N E N E N N N N N E N N E N E N E E N E E  
 N N N N E E N N N E E E N E E E E N N E E N E N N

3ª amostra

N E E E N N N N E N N E E E E E N N E N E N N E N  
 N N E E E E E N N E E N E N N N E N E N E N E E E

4ª amostra

N N N N E N N E N E E N N N E E E E N E N N N N N  
 E E E N N N E N N E N E E E E N E E N E N E E E N

5ª amostra

N E N E N N N E N E N N E N E E E N E N N E E E E  
 E N N N E E E N N N E E N N E E E E E N E N E E N

6ª amostra

N E E N E E E N E N N N N E N N E N N N N E E E E  
 N E E N E N E N E E E N E E N E E N N N N E N E N

7ª amostra

N N E E N N E E N N N N E N N N E E N N E E N N E  
 E N E N N E N E N E E N N E E E N N N E N N E E N

8ª amostra

N E N N N E E E E N E E E N N E N E N N E E N E E  
 E N N N E N E N E E N N N E N E N E E N E N E E N

9ª amostra

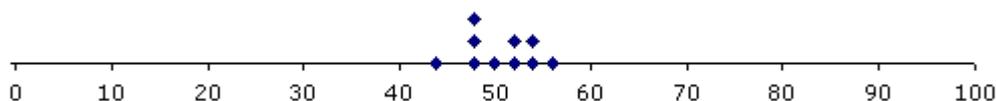
E N N E E E E N N E E E N E E N E E N E N N N N  
 N E E N N E N N E N E E N E E E N N N E E E E N E

10ª amostra

N N E E E E E N E E N E E E E N N E N N E E E E N  
 N N N E E E E N N N E E N N N N E N E E E N N N

Na seguinte tabela de frequências apresentam-se as percentagens de faces Euro das 10 amostras que decidimos recolher:

	Proporção de faces Euro
1ª amostra	48%
2ª amostra	48%
3ª amostra	52%
4ª amostra	48%
5ª amostra	54%
6ª amostra	50%
7ª amostra	44%
8ª amostra	52%
9ª amostra	56%
10ª amostra	54%



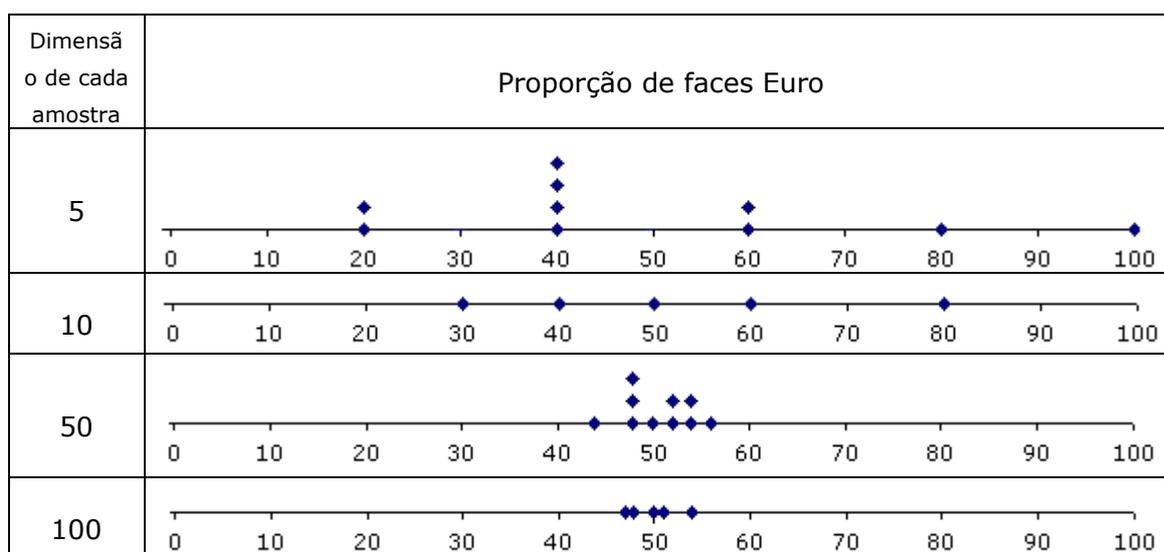
Como estávamos à espera, não obtivemos sempre a mesma percentagem de faces Euro, mas obtivemos valores razoavelmente próximos de 50%!

E se em vez de repetir a experiência 50 vezes, repetíssemos 100? Sem muito trabalho podemos verificar o que acontece, pois basta juntar as amostras anteriores 2 a 2 para obter amostras de dimensão 100:

	Proporção de faces Euro
1ª e 2ª amostras	48%
3ª e 4ª amostras	50%
5ª e 6ª amostras	51%
7ª e 8ª amostras	47%
9ª e 10ª amostras	54%



Não há dúvida! À medida que aumentamos o número de repetições da experiência aleatória, mais próximas umas das outras e de 50% ficam as percentagens de faces Euro obtidas:



Este exemplo dá-nos confiança que se aumentarmos ainda mais o número de repetições da experiência, a proporção de faces Euro *vai ter tendência a estabilizar à volta de 50%*. Para finalizar, verifique-se que se tivéssemos considerado conjuntamente as 500 repetições (das 10 amostras de dimensão 50) da experiência aleatória, a percentagem de faces Euro obtidas seria de 50,6%, valor bem próximo de 50%.

Esta regularidade estatística é utilizada para quantificar a probabilidade de um acontecimento, identificando-a com a frequência relativa com que esse acontecimento se observa, para um grande número de realizações da experiência. Em termos estatísticos "estimámos" a probabilidade (desconhecida) da realização de um acontecimento, pela frequência relativa ou percentagem de vezes com que esse acontecimento se verifica. É usual chamar a esta percentagem a *probabilidade empírica* ou *experimental*.

**Probabilidade empírica (ou frequentista)** – A **probabilidade** de um determinado acontecimento aleatório é a percentagem de vezes que se espera que ele aconteça, se se repetir a experiência, um grande número de vezes, nas mesmas condições.

### Exemplo – Qual o animal doméstico preferido



Consideremos a seguinte tabela que resultou de organizar a informação referente a uma sondagem feita a 50 alunos de uma escola, sobre qual o animal doméstico preferido:

Animal doméstico preferido	Nº de alunos	Freq. Relativa
Cão	35	0,70
Gato	8	0,16
Peixe	4	0,08
Passarinho	2	0,04
Outro qualquer	1	0,02
	50	1,00

Suponhamos que na altura em que se recolheu a informação da tabela, se pretendia recolher informação sobre a preferência de mais um aluno da escola, escolhido ao acaso. Algumas questões que se podem colocar sobre este aluno, no que diz respeito ao animal doméstico preferido, são as seguintes:

- Qual será o animal doméstico mais provável, da preferência deste aluno?
- Qual será um valor aproximado para a probabilidade deste aluno preferir o Cão?

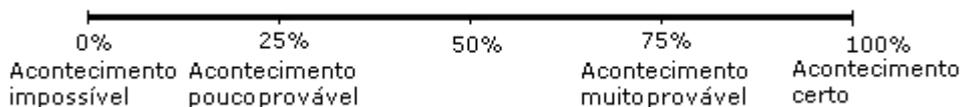
Na tabela anterior, verificamos que dos 50 alunos, 35 preferem o Cão. Então, é natural esperar que este outro aluno também prefira o Cão. Por outro lado, já que a frequência relativa do acontecimento "O animal doméstico preferido é o Cão" é de 0,70, esperamos que a probabilidade deste acontecimento esteja próxima de 0,70 ou 70%.

No exemplo anterior, a experiência consiste em seleccionar um aluno ao acaso e em averiguar qual o animal doméstico preferido. Existem várias respostas possíveis e é por essa razão que o resultado da experiência é *aleatório*: antes de registar a resposta do aluno, não temos informação suficiente para saber, de entre os acontecimentos "O animal doméstico preferido é o Cão", "O animal doméstico preferido é o Gato", "O animal doméstico preferido é o Peixe", "O animal doméstico preferido é o Passarinho", "O animal doméstico preferido é Outro qualquer", qual o acontecimento que se vai verificar.

As probabilidades assumem valores numa escala de 0% a 100% (ou 0 a 1). Se um acontecimento é *impossível*, atribui-se-lhe uma probabilidade de 0% (ou 0). Se temos a certeza que um acontecimento se vai verificar, ou seja, se é um acontecimento *certo*, então atribui-se-lhe a probabilidade de 100% (ou 1).

A probabilidade pode-se exprimir sob a forma de percentagem, decimal ou fracção.

Podemos utilizar uma escala de probabilidade para situar alguns acontecimentos, quanto à credibilidade atribuída à sua realização na próxima repetição da experiência aleatória:



A probabilidade de um acontecimento não se verificar é igual a 100% (ou 1) menos a probabilidade de se verificar.

Assim, como atribuímos anteriormente um valor aproximado de 70% ao acontecimento "O animal doméstico preferido é o Cão", podemos dizer que um valor aproximado para a probabilidade do acontecimento "O animal doméstico preferido *não* é o Cão" é cerca de 30%.

**Exemplo – Qual a probabilidade?**<sup>3</sup> Um computador está programado para calcular várias probabilidades. Associe as respostas numéricas com as descrições verbais seguintes:

- |          |   |
|----------|---|
| (a) -50% | (i) É tão provável acontecer, como não acontecer    |
| (b) 0%   | (ii) É muito provável que aconteça, mas não é certo |
| (c) 10%  | (iii) Isto não pode acontecer                       |
| (d) 50%  | (iv) Pode acontecer, mas é pouco provável           |
| (e) 90%  | (v) Isso acontecerá, de certeza                     |
| (f) 100% | (vi) Há um erro no programa                         |
| (g) 200% |   |

Nos valores numéricos, existem 2 que não podem ser probabilidades. Assim, só podem ser atribuídos a um erro no programa, donde (vi) corresponde a (a) e (g). Se um acontecimento é tão provável de acontecer, como de não acontecer, então temos que a sua probabilidade é cerca de 50% e, portanto, (i) corresponde a (d). As outras associações são (ii) a (e); (iii) a (b); (iv) a (c) e (v) a (f).

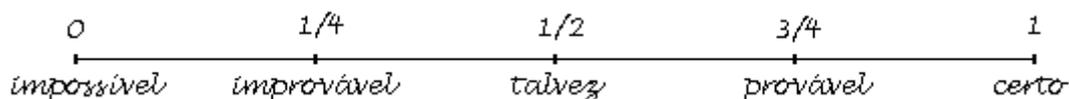
No ensino básico,, a probabilidade deve ser introduzida de uma forma intuitiva. De seguida descrevemos uma tarefa, que pode ser dividida em várias sessões, em que se descreve um processo que pode ser utilizado para introduzir a linguagem da probabilidade nos alunos do 1º ciclo.

**Tarefa – A escala de probabilidade.** Com o objectivo de promover o domínio da linguagem da probabilidade, o professor pede aos alunos que relatem acontecimentos do dia a dia em que introduzam os termos *impossível*, *certo*, *provável*, *muito provável*, *pouco provável*. Depois de algum tempo dedicado a esta discussão, o professor desenha uma linha no quadro e no início da linha marca um zero. Então o professor diz aos alunos que 0 será a probabilidade que se vai atribuir a um *acontecimento impossível* e que o maior valor possível para a probabilidade será 1 ou 100%, sendo o 1 marcado no fim da linha, a que corresponde o acontecimento *certo*:

<sup>3</sup> Freedman et al (1991)



Alguma discussão em volta de alguns termos utilizados pelos alunos, permite acrescentar à linha anterior mais algumas indicações, chegando à seguinte representação:

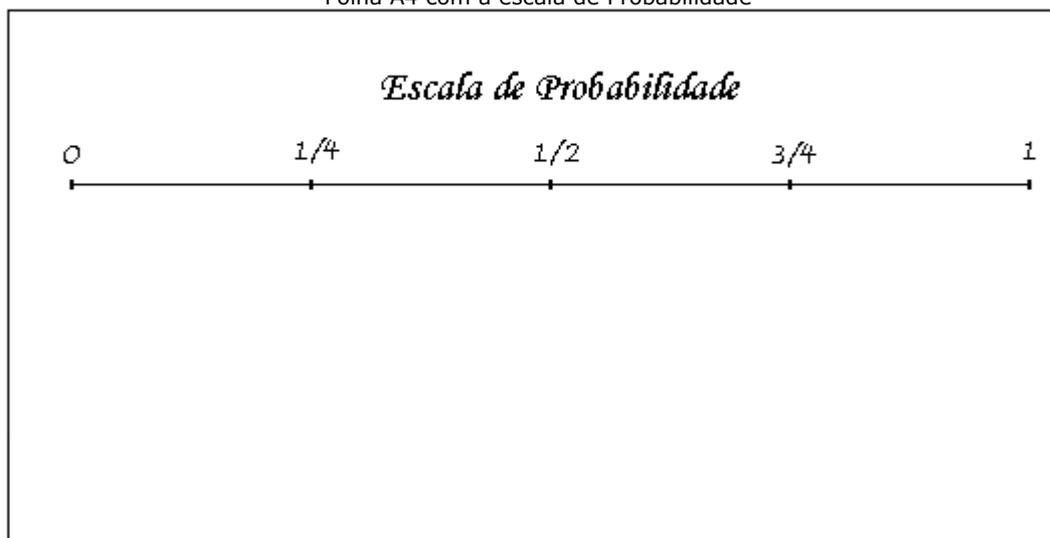


O professor pede aos alunos que vão sucessivamente ao quadro escrever algumas frases para depois, em conjunto as situarem no eixo anteriormente desenhado, de acordo com a *credibilidade* que lhes atribuem. Poderá haver alguma discussão sobre a posição onde colocar as frases, pois a credibilidade que se atribui a cada acontecimento, não é necessariamente igual para todos os alunos. Algumas das frases que os alunos escreveram, foram:

- A - Amanhã vai chover
- B - Hoje vou ver televisão quando chegar a casa
- C - Uma pedra de gelo num copo de água derrete
- D - O Benfica vai ganhar a taça
- E - Se lançar uma moeda de 1 Euro ao ar, vai aparecer a face Euro
- F - Se retirar uma carta ao acaso, de um baralho, obtenho um Rei
- G - Amanhã quando for passear, vou encontrar um dinossáurio vivo
- H - Amanhã quando for ao parque, vou ver passarinhos
- I - Amanhã o Sol vai nascer

Numa sessão seguinte o professor divide os alunos em grupos de 3 ou 4 e dá a cada grupo uma folha A4, onde está desenhada a Escala de Probabilidade, e uma folha com um conjunto de declarações (O professor pode pedir aos alunos para fazerem a sua própria folha de declarações).

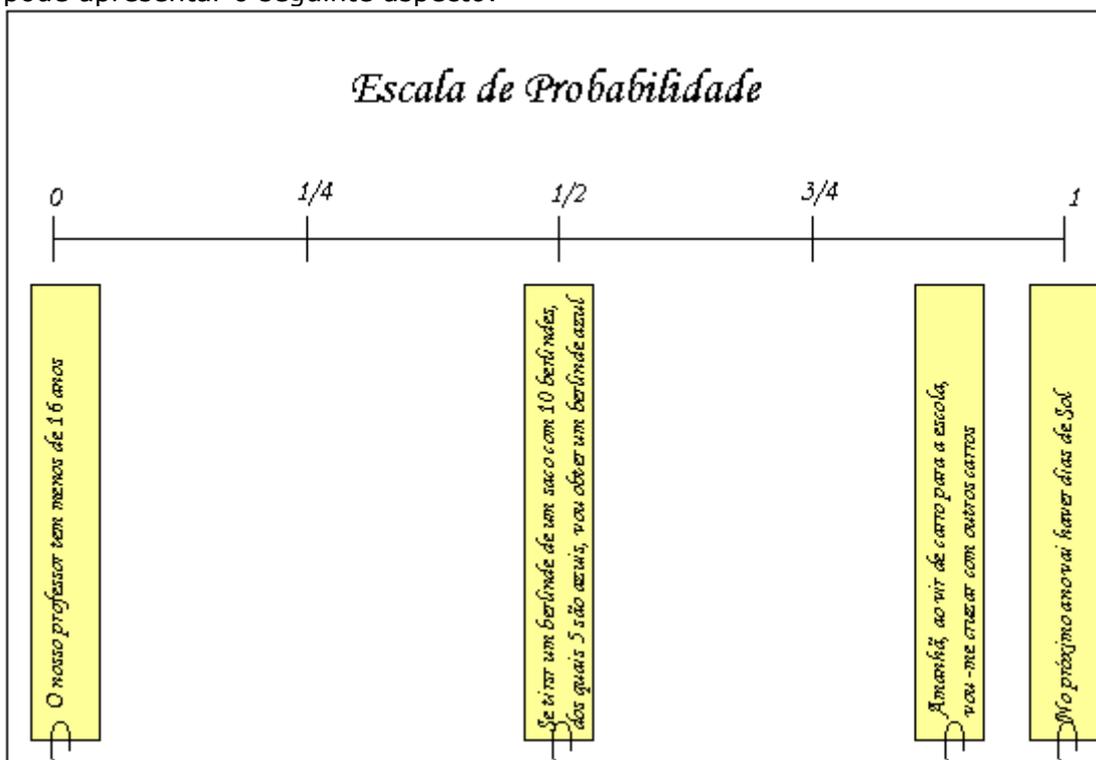
Folha A4 com a escala de Probabilidade



Folha com as declarações

<i>O próximo carro a passar na rua é preto</i>	<i>O nosso professor tem menos de 16 anos</i>
<i>Se tirar um berlinde de um saco com 10 berlines, dos quais 5 são azuis, vou obter um berlinde azul</i>	<i>O próximo bebé a nascer na cidade é rapaz</i>
<i>No ano que vem o Benfica vai ganhar a taça de Portugal</i>	<i>Amanhã, ao vir de carro para a escola, vou-me cruzar com outros carros</i>
<i>No ano que vem vai nevar em Lisboa</i>	<i>Se lançar um dado, vai-me sair um 6</i>
<i>No próximo ano vai haver dias de Sol em Portugal</i>	<i>Se lançar duas moedas vão-me sair duas faces Euro</i>
<i>No ano que vem vai nevar na Serra da Estrela</i>	<i>No próximo jogo do Sporting-Benfica, vai haver empate</i>

Os alunos, de cada grupo, devem começar por recortar esta folha, de modo a separarem as declarações. Depois, em conjunto, posicionam-nas na escala de probabilidade, de acordo com a credibilidade que atribuírem a cada uma dessas frases. Sugere-se que utilizem cliques. A meio da sessão, a folha de um dos grupos pode apresentar o seguinte aspecto:



A sugestão de prenderem as declarações à folha, com um clipe, com a escala de Probabilidade, prende-se com o facto de depois de alguma discussão conjunta, os grupos poderem rever as suas convicções na credibilidade atribuída à realização de alguns dos acontecimentos, atribuindo-lhes uma posição diferente na escala.

## Tarefa – Serão os jogos justos (ou equilibrados)?



Na turma o professor propõe alguns jogos para serem jogados com uma moeda ou com um dado por pares de alunos e pretende que no fim do jogo os alunos concluam se o jogo é justo ou não, isto é, se dará a mesma possibilidade de ganhar a ambos os jogadores. Para estes jogos o professor levou algumas moedas de 1 euro, alguns dados de 6 faces e um punhado de feijões.

**1º jogo** – Este jogo é jogado por dois alunos, por exemplo o Pedro e a Rita, que têm à partida uma caixa com 20 feijões e um dado. O jogo consiste em lançar um dado e se sair face com um *número par* de pintas, o Pedro retira um feijão da caixa e fica com ele. Se sair face com um *número ímpar* de pintas é a Rita que retira o feijão. Ganha o jogo quem tiver mais feijões quando se esgotar a caixa.

Algumas questões:

- À partida quais são as expectativas sobre quem vai ganhar o jogo?
- Será o jogo justo?
- Se jogarem 2 vezes o mesmo jogo, é de esperar que ganhe o mesmo jogador?

**2º jogo** – Este jogo é jogado por dois alunos, por exemplo a Maria e a Joana, que têm à partida uma caixa com 20 feijões e um dado. O jogo consiste em lançar um dado e se sair face em que o número de pintas é um *número primo*, a Maria retira um feijão da caixa e fica com ele. Se sair uma face com um número de pintas que *não seja número primo*, é a Joana que retira o feijão. Ganha o jogo quem tiver mais feijões quando se esgotar a caixa.

Algumas questões:

- À partida quais são as expectativas sobre quem vai ganhar o jogo?
- Será o jogo justo?
- Se jogarem 2 vezes o mesmo jogo, é de esperar que ganhe o mesmo jogador?

**3º jogo** – Este jogo é jogado por dois alunos, por exemplo o João e o Bernardo, que têm à partida uma caixa com 20 feijões e duas moedas de um Euro. O jogo consiste em lançar as moedas e se saírem duas *faces iguais*, o João retira um feijão da caixa e fica com ele. Se saírem duas *faces diferentes*, é o Bernardo que retira o feijão. Ganha o jogo quem tiver mais feijões quando se esgotar a caixa.

Algumas questões:

- À partida quais são as expectativas sobre quem vai ganhar o jogo?
- Será o jogo justo?
- Se jogarem 2 vezes o mesmo jogo, é de esperar que ganhe o mesmo jogador?

**4º jogo** – Este jogo é jogado por dois alunos, por exemplo a Sara e o Santiago, que têm à partida uma caixa com 20 feijões e dois dados. O jogo consiste em lançar os dois dados e se a soma das pintas *for menor ou igual 6* a Sara retira um feijão da caixa e fica com ele. Se a soma das pintas for maior ou igual a 8 é o Santiago que retira o feijão. Se a soma das pintas for 7, ninguém retira feijões. Ganha o jogo quem tiver mais feijões quando se esgotar a caixa.

Algumas questões:

- À partida quais são as expectativas sobre quem vai ganhar o jogo?
- Será o jogo equilibrado?
- Se jogarem 2 vezes o mesmo jogo, é de esperar que ganhe o mesmo jogador?

Todos estes jogos devem ser orientados pelo professor, que deve chamar a atenção para o facto de em qualquer um dos deles se pretender a repetição da experiência aleatória 20 vezes. Deve indicar que isso não pode ser considerado um número razoável de vezes, de forma a estabilizar as frequências relativas com que os acontecimentos se realizam. No entanto, a partir dos resultados obtidos com as 20 repetições, os alunos podem conjecturar sobre se o jogo será justo ou não, tendo em conta o número de feijões que cada aluno que compõe o par conseguiu ganhar. Se o número de feijões for *muito* diferente é natural que se ponha a hipótese de que o jogo não seja justo. Podem também os alunos colocar a hipótese de o número de vezes que jogaram o jogo não ser suficiente para decidirem sobre se o jogo será equilibrado ou não, e sugerirem que se façam mais algumas jogadas. Eventualmente poderão jogá-lo mais algumas vezes, registando o vencedor num esquema de contagem gráfica (*tally chart*).

### Tarefa – O que é mais provável?



Numa turma com 24 alunos, 16 são raparigas e 8 são rapazes. Dos 24 alunos, metade têm olhos castanhos e a outra metade, olhos de outra cor. Também se sabe que 8 dos alunos (rapazes ou raparigas) são louros. O professor que usa fichas, todos os dias selecciona uma ficha ao acaso, depois de ter baralhado as fichas como quem baralha um baralho de cartas, para que o aluno com o nome que consta na ficha seleccionada vá ao quadro resolver um problema. Na próxima ida ao quadro:

- É mais provável que seja seleccionado um rapaz ou uma rapariga?
- É mais provável que o aluno tenha olhos castanhos ou de outra cor?
- É mais provável que o aluno seja louro ou não seja louro?
- Quais as estimativas para as probabilidades dos acontecimentos anteriores?

Para responder às três primeiras questões, os alunos devem ter sensibilidade para verificar que quantos mais alunos houver pertencentes a determinada categoria, mais provável é ser seleccionado um aluno pertencente a essa categoria, se a *selecção for feita aleatoriamente* (ao acaso), como é pressuposto, ao exigir que as fichas sejam baralhadas. Assim, será mais provável ser seleccionada uma rapariga, do mesmo modo que é mais provável ser seleccionado um aluno que não seja louro e existe igual probabilidade de ser seleccionado um aluno de olhos castanhos e um que não tenha olhos castanhos.

Para responder à última questão, decidiram registar numa folha as características do aluno seleccionado (sempre seleccionando uma ficha ao acaso) durante 30 aulas consecutivas, tendo obtido os seguintes registos:

Dia	Rapariga	Olhos castanhos	Louro
1	0	0	0
2	1	1	1
3	0	0	0
4	1	1	1
5	1	1	1
6	0	0	0
7	1	1	1
8	1	1	0
9	0	0	0
10	1	1	0
11	0	0	0
12	0	0	0
13	1	1	1
14	1	0	0
15	0	0	0
16	1	1	1
17	1	1	0
18	1	0	0
19	1	1	1
20	1	1	1
21	1	1	1
22	1	1	1
23	1	1	1
24	1	0	0
25	1	1	0
26	0	0	0
27	0	0	0
28	0	0	0
29	1	0	0
30	1	1	1
Total	20	16	12
Freq. Rel.	20/30≈ 67%	16/30≈53%	12/30=40%

Na tabela anterior representou-se por um 1 sempre que se verificava o acontecimento de interesse, e por 0 caso contrário. Por exemplo, sempre que era seleccionada uma rapariga colocava-se um 1. Caso contrário escrevia-se um 0. Assim, no 1.º dia em que começaram a fazer os registos, verificou-se que foi ao quadro um rapaz que não tinha olhos castanhos e não era louro. Repare-se que com esta forma de registar as observações, para obter as frequências absolutas basta somar a coluna de 0's e 1's.

Da tabela anterior conclui-se que uma estimativa para a probabilidade:

- de ser seleccionada uma rapariga é aproximadamente 67%;
- de ser seleccionado um aluno de olhos castanhos é aproximadamente 53%;
- de ser seleccionado um aluno louro anda à volta de 40%.

Assim, numa próxima chamada ao quadro espera-se que o aluno seja rapariga e espera-se que o aluno não seja louro. Quanto ao facto de ter ou não olhos castanhos, espera-se que tanto possa acontecer uma coisa como outra.

O professor deve chamar a atenção para que a partir da tabela anterior ainda se podem estimar as probabilidades de outros acontecimentos. Como estimar, por exemplo, a probabilidade de seleccionar uma "rapariga de olhos castanhos e loura"? Basta contar o número de linhas em que há três 1's e dividir esse valor por 30,

para obter a frequência relativa com que se observou o acontecimento pretendido. No caso considerado a estimativa obtida é de 40%(=12/30).

O professor pode sugerir ainda que estimassem as probabilidades de outros acontecimentos, tais como:

- O próximo aluno a ser chamado ser "rapaz de olhos castanhos e não louro";
- O próximo aluno a ser chamado ser uma "rapariga loura";
- O próximo aluno a ser chamado ser "rapaz de olhos não castanhos e não louro";
- etc.

#### Tarefa – Qual o meio de transporte utilizado pelos alunos para irem para a escola?<sup>4</sup>



Num determinado dia o professor decidiu juntamente com os alunos fazerem um pequeno trabalho de investigação sobre o meio de transporte utilizado para irem para a escola. Decidiu escrever no quadro os meios de transporte que julgava serem os utilizados e pediu a cada um dos alunos que fosse ao quadro e assinalasse qual o meio de transporte que utilizou, à frente da categoria respectiva:

Autocarro
Carro
A pé
Comboio
Bicicleta

O primeiro aluno a ir ao quadro, e que utilizava o autocarro, em vez de colocar simplesmente um traço ou uma cruz à frente da categoria *Autocarro*, resolveu desenhar uma figura estilizada, que representava um rapaz:

Autocarro	
Carro	
A pé	
Comboio	
Bicicleta	

Os outros alunos acharam a ideia muito interessante e no fim obtiveram o seguinte pictograma, correspondente ao meio de transporte utilizado pelos alunos presentes na turma, naquele dia:

Autocarro									
Carro									
A pé									
Comboio									
Bicicleta									

Algumas questões, baseadas no pictograma anterior e a serem trabalhadas com os alunos, podem ser as seguintes:

<sup>4</sup> Esta tarefa foi sugerida por um exemplo de Watson (2006).

1. O que é que se ganhou em ter utilizado uma figura masculina ou feminina, em vez de utilizar simplesmente um traço ou uma cruz?
2. Quantos alunos estavam na turma naquele dia?
3. Se se tivesse feito o mesmo estudo noutra dia, ter-se-ia obtido um pictograma perfeitamente igual?
4. Um aluno que chegou atrasado à aula, já depois de se ter feito a representação gráfica anterior, disse que tinha vindo de Carro. Este aluno será rapaz ou rapariga? Justificar a resposta.
5. Admitindo que o Ricardo não tinha ido à escola naquele dia, por estar doente, representar numa escala de probabilidade, a "credibilidade" atribuída a cada uma das seguintes afirmações, relacionadas com o transporte utilizado pelo Ricardo para ir, no dia seguinte, para a escola:
  - a) O Ricardo utiliza o Comboio
  - b) O Ricardo utiliza a Bicicleta
  - c) O Ricardo utiliza o Autocarro
  - d) O Ricardo utiliza o Carro

As respostas às questões anteriores podem ser várias e o professor deve ter a preocupação de investigar quais os argumentos que conduzem às respostas dos alunos. Por exemplo, pode-se esperar que algum aluno responda à questão 4. dizendo que é rapaz. Ao dar esta resposta, o aluno estará eventualmente a utilizar um raciocínio baseado no padrão da sequência: 2 raparigas, um rapaz, 2 raparigas, 1 rapaz! Neste caso o professor deve chamar a atenção que a ordem pela qual estão representados os símbolos não tem interesse relevante. Orienta-se então a discussão para o argumento frequencista: dos 5 alunos que viajaram de carro, 4 eram raparigas. Então é "mais provável" que um outro aluno que viaje de carro, seja rapariga. Na resposta à questão 5, a "credibilidade" atribuída à opção c) pode ser superior à atribuída à opção b). Efectivamente dos alunos que viajaram de bicicleta, só 1 é que é rapariga, o que nos inclina para uma "forte" credibilidade a que o Ricardo tenha viajado de Bicicleta. No entanto, verifica-se que dos 27 alunos, 9 viajaram de autocarro, acabando por ser este o meio de transporte mais utilizado. Com tão poucos dados recolhidos, não podemos ter certeza sobre quaisquer afirmações que façamos, mas podemos formular algumas conjecturas e atribuir-lhes um certo grau de credibilidade.

**Tarefa – Como seleccionar uma de duas pessoas, de uma forma justa<sup>5</sup>.** Na turma, constituída por 2 rapazes – o Tiago e o Ricardo, e 16 raparigas, era necessário escolher um aluno rapaz para pertencer a uma comissão que tinha de integrar os dois sexos. Como só havia dois rapazes decidiram atirar uma moeda de 1 euro ao ar. Se saísse a face Euro (E) seria escolhido o Ricardo, caso contrário, se saísse a face Nacional (N) seria o Tiago. Antes de lançarem a moeda, o Tiago questionou o professor sobre se esse processo de selecção seria justo. Embora fosse um processo habitual de fazer uma escolha entre duas situações, quem é que lhe garantia que seria de 50% a possibilidade de ser ele o escolhido? Ou por outras palavras, o que ele desejava saber era se a moeda era equilibrada.

---

<sup>5</sup> Graça Martins et al. (2007), p.164

Decidiram fazer uma experiência que consistia em lançar a moeda algumas vezes e registar os resultados obtidos. A fim de 10 lançamentos, os resultados obtidos foram os seguintes:

N    E    N    N    N    E    E    E    E    E

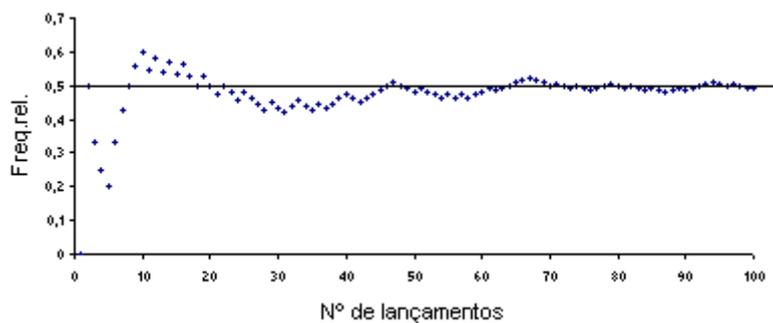
Estes resultados não sossegaram o Tiago, pois ele começou a pensar que só teria 40% de possibilidades de ser seleccionado, uma vez que em 10 vezes a moeda só lhe foi favorável 4 vezes!

O professor chamou então a atenção para o facto de se ter de realizar a experiência um grande número de vezes, pois com 10 lançamentos não podemos tirar qualquer conclusão. Fizeram então mais 90 lançamentos, tendo obtido os seguintes resultados:

N.º do lanç.	Result.	N.º de faces N	Freq. Rel. da face N	N.º do lanç.	Result.	N.º de faces N	Freq. Rel. da face N
1	N	1	1,000	51	E	26	0,510
2	E	1	0,500	52	N	27	0,519
3	N	2	0,667	53	N	28	0,528
4	N	3	0,750	54	N	29	0,537
5	N	4	0,800	55	E	29	0,527
6	E	4	0,667	56	N	30	0,536
7	E	4	0,571	57	E	30	0,526
8	E	4	0,500	58	N	31	0,534
9	E	4	0,444	59	E	31	0,525
10	E	4	0,400	60	E	31	0,517
11	N	5	0,455	61	E	31	0,508
12	E	5	0,417	62	N	32	0,516
13	N	6	0,462	63	E	32	0,508
14	E	6	0,429	64	E	32	0,500
15	N	7	0,467	65	E	32	0,492
16	E	7	0,438	66	E	32	0,485
17	N	8	0,471	67	E	32	0,478
18	N	9	0,500	68	N	33	0,485
19	E	9	0,474	69	N	34	0,493
20	N	10	0,500	70	N	35	0,500
21	N	11	0,524	71	E	35	0,493
22	E	11	0,500	72	N	36	0,500
23	N	12	0,522	73	N	37	0,507
24	N	13	0,542	74	E	37	0,500
25	E	13	0,520	75	N	38	0,507
26	N	14	0,538	76	N	39	0,513
27	N	15	0,556	77	E	39	0,506
28	N	16	0,571	78	E	39	0,500
29	E	16	0,552	79	E	39	0,494
30	N	17	0,567	80	N	40	0,500
31	N	18	0,581	81	N	41	0,506
32	E	18	0,563	82	E	41	0,500
33	E	18	0,545	83	N	42	0,506
34	N	19	0,559	84	N	43	0,512
35	N	20	0,571	85	E	43	0,506
36	E	20	0,556	86	N	44	0,512
37	N	21	0,568	87	N	45	0,517
38	E	21	0,553	88	E	45	0,511
39	E	21	0,538	89	E	45	0,506

40	E	21	0,525	90	N	46	0,511
41	N	22	0,537	91	E	46	0,505
42	N	23	0,548	92	E	46	0,500
43	E	23	0,535	93	E	46	0,495
44	E	23	0,523	94	E	46	0,489
45	E	23	0,511	95	N	47	0,495
46	E	23	0,500	96	N	48	0,500
47	E	23	0,489	97	E	48	0,495
48	N	24	0,500	98	N	49	0,500
49	N	25	0,510	99	N	50	0,505
50	N	26	0,520	100	N	51	0,510

O gráfico seguinte mostra a evolução da frequência relativa da saída da face N, à medida que se fazem os sucessivos lançamentos da moeda:



Tendo em conta os resultados anteriores, verifica-se que a frequência relativa da saída da face Nacional, tende a estabilizar à volta dos 50%. Assim, não temos razão para rejeitar a hipótese de a moeda ser equilibrada, dando 50% de probabilidade a cada face.

### Ainda a regularidade a longo termo...

Neste momento, em que já nos apercebemos do que é a *regularidade a longo termo*, é pertinente questionarmos:

*Será que o acaso pode ser governado?* Então não estamos a admitir que a longo termo é possível obter um padrão genérico de comportamento do fenómeno aleatório?

Efectivamente, quando observamos o fenómeno em estudo um número suficientemente grande de vezes (nas mesmas condições...), verifica-se um comportamento que pode ser modelado, isto é, podemos arranjar um modelo para exprimir a aleatoriedade. Mas atenção! Esta regularidade não existe a não ser a longo termo! E a longo termo significa que temos de repetir a experiência, nas mesmas condições, um número suficiente de vezes até verificarmos que a frequência relativa com que o acontecimento se realiza tem tendência a estabilizar.

Na situação comum do lançamento de uma moeda ou de um dado, não podemos dizer qual a face que sai no próximo lançamento. No entanto se lançarmos a moeda ou o dado (equilibrados) um número razoável de vezes, esperamos que aproximadamente metade das vezes saia a face Euro da moeda e aproximadamente um sexto das vezes saia a face 1 do dado, de modo que a frequência relativa com que se verifica a face Euro anda à volta de 50% e a frequência relativa com que se verifica a face 1 anda à volta de 17%. Suponha agora que lança a moeda 8 vezes e que obteve a seguinte sequência:

## E N E E N N N N

Se lançar novamente a moeda, o que é que espera que saia? Embora lhe apetecesse dizer que no próximo lançamento é mais provável que saia a face Euro, para equilibrar o número de faces Euro, com o número de faces Nacional, na verdade no próximo lançamento tanto pode sair a face Euro como a face Nacional, já que os sucessivos lançamentos são independentes uns dos outros – a moeda não tem memória...

**Exemplo – Os dissabores da regularidade a longo termo...**<sup>6</sup>. A regularidade a longo termo se não for bem compreendida, pode acarretar alguns dissabores! Foi o que aconteceu com aquele casal que tinha planeado ter 4 filhos. Depois de nascerem 4 raparigas, e na expectativa de terem um rapazinho, ainda tentaram mais 3 vezes e ficaram com uma linda equipa de 7 raparigas! Depois destas 7 raparigas o médico assegurou-lhes que era praticamente certo que o bebé seguinte seria rapaz. Infelizmente para este casal, os fenómenos aleatórios que consistem em ter mais uma criança ou lançar mais uma vez a moeda, são idênticos. Efectivamente 8 raparigas de seguida, é muito improvável, mas uma vez nascidas 7 raparigas, não é de todo improvável que o próximo bebé seja rapariga – e foi!

**Tarefa – O jogo será justo**<sup>7</sup>? O professor propõe aos alunos realizarem o seguinte jogo, para o qual é necessário algumas fichas ou berlindes ou até botões, desde que só difiram na cor: os alunos organizam-se em 5 grupos, e cada grupo escolhe uma ficha de cor diferente. As cinco fichas (amarela, verde, vermelha, azul, branca) são colocadas num copo de plástico opaco (ou numa caixa ou num saco...) e procede-se à extracção, com reposição, de 30 fichas. Sempre que se retira uma ficha, regista-se a cor e repõe-se novamente no copo, antes de retirar a seguinte. Ganha o grupo cuja cor tenha saído mais vezes.

- Será que o jogo é justo (equilibrado)?
- Na próxima vez que se jogar o jogo será que ganha o mesmo grupo? Se ganhar o mesmo grupo significa que o jogo não é justo?

O professor, sem que os alunos se tenham apercebido, retirou duas das fichas, por exemplo a amarela e a verde e colocou duas fichas vermelhas em sua substituição. Pediu aos alunos para jogarem novamente o jogo e registarem os resultados. Depois de o jogo ser jogado algumas vezes, o professor questiona os alunos sobre se:

- Haverá algum grupo que esteja a ganhar mais vezes que os outros grupos? Os resultados são diferentes, quando comparados com a situação anterior?

Depois de alguma discussão sobre a composição das fichas no copo de plástico, o professor mostra o seu conteúdo e pede aos alunos para fazerem uma previsão sobre qual será o próximo grupo a ganhar, se repetirem novamente o jogo.

Note-se que este tipo de actividade é importante para desenvolver conceitos básicos de probabilidade. Neste jogo, se cada grupo for representado por uma cor, o jogo será justo, já que cada grupo tem a mesma possibilidade de ganhar.

---

<sup>6</sup> Adaptado de Moore (1997)

<sup>7</sup> Sugerido por Way (1997)

Contudo, como as selecções são aleatórias, não temos a garantia que todas as cores saiam com igual frequência. Quando se alterou a composição das fichas no copo, espera-se que haja uma alteração nos resultados. Para já, há dois grupos que não podem ganhar! Esta actividade mostra também a forma como os *dados* nos podem ajudar a tomar decisões. O jogo permitiu gerar dados os quais são usados para tomar decisões acerca da composição das fichas no copo.

**Tarefa – Qual a composição do saco de berlindes?** O professor chega à aula com um saco que contém 10 berlindes de duas cores, de entre cinco cores possíveis e pretende que os alunos:

- digam quais as cores dos berlindes que estão no saco e que
- estimem quantos berlindes são de cada cor.



O professor começa por indicar qual a metodologia para a recolha de dados:

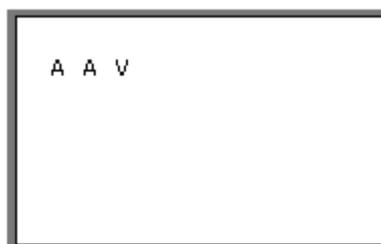
- Pede a um aluno que retire um berlinde do saco, mostre o berlinde aos colegas para se aperceberem de qual a cor, e reponha o berlinde no saco. Uma vez que a "memória é curta", a cor do berlinde é registada no quadro. Admitindo que o berlinde era azul, regista um A.



Esta forma de realizar a experiência consiste numa *extracção com reposição*.

Neste momento, associado à experiência anterior podemos dizer que o acontecimento "O saco tem berlinde(s) de cor azul" é um acontecimento *certo*, com probabilidade 100%.

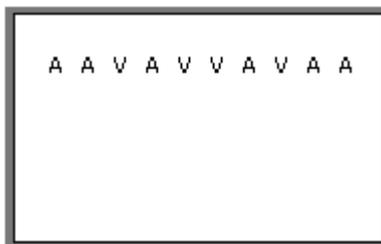
- O professor pede a outros alunos que procedam como o primeiro aluno. O segundo aluno a tirar o berlinde também tirou um berlinde azul, mas o terceiro aluno tirou um berlinde vermelho. Após estas três extracções com reposição, tem-se o seguinte registo no quadro (antes de cada extracção, os berlindes são baralhados dentro do saco)



Neste momento os alunos já têm a resposta à primeira questão: no saco existem berlindes de cores Azul e Vermelha. Ficámos também a saber que o acontecimento

“O saco contém berlindes de cor diferente de Azul ou Vermelha” é um acontecimento *impossível*, a que associamos a probabilidade de 0% (recordemos que o professor tinha dito que só havia duas cores distintas de berlindes no saco).

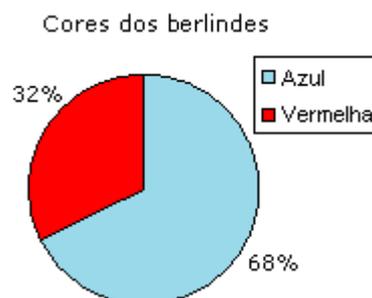
- O professor propõe que se continue a realizar a mesma experiência de retirar o berlinde, registar a cor e repor o berlinde no saco. Após 10 realizações da experiência, os resultados obtidos foram



Após as 10 extracções o professor lembra que o saco contém 10 berlindes e pergunta se neste momento podemos dizer que o saco tem 6 berlindes azuis e 4 vermelhos? É de esperar que se coloquem questões acerca desta situação, e que alguns alunos respondam que não, pois o mais natural é que ao retirar ao acaso os berlindes do saco, alguns berlindes tenham sido retirados mais do que uma vez e alguns nunca tenham chegado a ser retirados. O que fazer então, para estimar a proporção de berlindes de cada cor? Alguns alunos já alertados para o *raciocínio frequencista*, sugerem que se façam várias extracções (com reposição), sendo de esperar que ao fim de “muitas” extracções a frequência relativa com que se verificou a saída de cada cor, possa reflectir a composição do saco. É razoável admitir que se o saco contiver uma maior proporção de berlindes azuis, estes saiam com maior frequência.

Ao fim de 80 extracções os resultados encontram-se resumidos na seguinte tabela de frequências:

Cor do berlinde	Freq. Abs.	Freq. Rel.
Azul	54	0,675
Vermelha	26	0,325
	80	1,00



O que concluir da tabela e do diagrama circular anteriores? A frequência relativa da saída de berlindes de cor azul é 0,675. Então esperamos que aproximadamente 68% dos berlindes do saco sejam azuis. Respondendo à questão inicial, como o saco tem 10 berlindes, *estimamos* que 7 berlindes sejam azuis e 3 berlindes sejam vermelhos.

- Se procedermos a uma nova extracção de um berlinde, qual a cor que esperamos que o berlinde tenha?

Esperamos que seja Azul, já que a probabilidade de retirar um berlinde azul anda à volta de 68%, enquanto que a probabilidade de retirar um berlinde vermelho anda à volta de 32%.

Para fazer a inferência sobre a composição do saco, um factor importante com que se teve de entrar em linha de conta, foi o *número de repetições da experiência* aleatória. Intuitivamente estamos “confiantes” que quanto maior for o número de repetições da experiência, ou seja, *quanto maior for a dimensão da amostra recolhida*, melhor será o resultado da nossa inferência.

Na determinação da composição do saco de berlindes, pode acontecer que algum dos alunos sugira o seguinte raciocínio, para estimar o número de berlindes azuis: fazer várias sucessões de extracções de 10 berlindes (número de berlindes do saco) e contar em cada sucessão de 10 extracções, quantos berlindes azuis se obtiveram. Por exemplo, suponhamos que se tinha procedido a 8 sucessões de 10 extracções, com os seguintes resultados:

	Nº de berbindes azuis
V A A V A A A A V	7
A A A A V A A A V V	7
A A V A V A V A A A	7
A A V A A V A V A A	7
A V A A A V A V V A	6
V A A A V V A V A A	6
A A A A A V A A A	9
A A A V A A A A V	8

O número de berlindes azuis em 10 berlindes, variou entre 6 e 9. Qual o melhor valor para representar os 8 valores obtidos? Naturalmente a média! Assim, calculando a média dos dados anteriores

$$\frac{7 + 7 + 7 + 7 + 6 + 6 + 9 + 8}{8} = 7,125$$

obtemos 7,125, pelo que escolhemos o inteiro 7 (valor inteiro mais próximo de 7,125) como o número de berlindes azuis no saco.

Será que este raciocínio que acabámos de fazer é diferente do raciocínio frequentista que utilizámos inicialmente? Vejamos que não! Se contabilizarmos o total de berlindes azuis no total de extracções realizadas, então a proporção de berlindes de cor azul será  $57/80=0,7125$ . Admitindo que esta proporção reflecte a proporção de berlindes azuis no saco, consideramos que no saco de 10 berlindes, cerca de 70% são azuis, ou seja 7.

**Tarefa – Os sacos de berlindes.** O professor organiza os alunos em grupos de 3 ou 4. Dá a cada um dos grupos um saco mistério com 4 berlindes. Cada grupo conhece a cor dos berlindes que estão no seu saco, mas não pode dizer aos outros grupos. A experiência consiste em cada grupo fazer 30 extracções, com reposição, de berlindes do seu saco registando a cor dos berlindes que saíram antes de os repor novamente no saco. Sugere-se que cada grupo tenha uma folha idêntica à seguinte para o acompanhamento da experiência:

Folha

1. Verifica a composição do saco e regista, numa tabela, todos os *resultados* possíveis que poderás obter se retirares um berlinde, ao acaso, do saco.
2. Se realizares 30 experiências de retirar um berlinde e repô-lo no saco, quantas vezes esperas que se verifica cada um dos *resultados*? Preenche a tabela seguinte com as tuas predições:

Resultado	Número de vezes que se espera que se verifique o resultado

3. Realiza agora a experiência sugerida na alínea anterior e preenche a tabela seguinte:

Resultado	Frequência absoluta

4. Compara os resultados obtidos na tabela anterior, com os da tabela da alínea 2.

Depois de cada grupo realizar a experiência, regista no quadro a tabela com os resultados obtidos. A partir dessas tabelas, tentam descrever a composição dos sacos uns dos outros.

É claro que os berlindes podem ser substituídos por botões do mesmo tamanho e feitio, só diferindo na cor, ou “caricas” de refrigerantes também do mesmo tamanho, ou fichas feitas pelos próprios alunos, em cartolina grossa, etc.

**Tarefa – Qual o tipo de prato que os alunos preferem?** Numa escola o Director pretende saber como se distribui a preferência dos alunos, relativamente a um conjunto de pratos de referência, ou seja, qual a probabilidade de cada prato ser o preferido, para satisfazer o mais possível os alunos que vão comer à cantina. Encarregou um grupo de alunos de recolher a informação necessária, tendo estes utilizado a seguinte metodologia:

- Elaboraram uma lista com os pratos que a cantina indicou;
- Colocaram-se à entrada da cantina na hora do almoço;
- De entre os alunos que iam chegando para almoçar, só eram seleccionados para responder de cinco em cinco alunos. Por exemplo, seleccionava-se para responder o 5º aluno a chegar, o 10º, o 15º, etc.

O resultado da recolha da informação apresentou o seguinte aspecto:

Esparquete à Bolonhesa	
Peixe assado com batatas e legumes	
Filetes no forno com arroz e salada	
Bacalhau com natas e salada	
Rolo de carne com puré e salada	

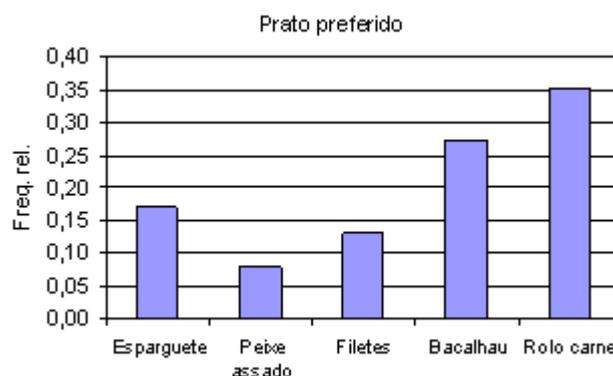
A partir dos dados anteriores os alunos construíram uma tabela de frequências e um gráfico de barras, tendo elaborado um pequeno relatório que entregaram ao Director:

### Relatório

*Questão:* Foi-nos pedido que investigássemos junto dos nossos colegas que vão comer à cantina, quais são os pratos preferidos, da seguinte lista de pratos: “Esparguete à Bolonhesa”, “Peixe assado com batatas e legumes”, “Filetes no forno com arroz e salada”, “Bacalhau com natas e salada” ou “Rolo de carne com puré e salada”.

*Metodologia:* Decidimos recolher uma amostra de alunos a quem fizemos a pergunta sobre qual destes pratos era o seu preferido. A selecção dos alunos foi feita da seguinte forma: só interrogávamos os alunos que chegavam à cantina, de cinco em cinco. Esta forma de seleccionar evitou que atrasássemos a fila, o que aconteceria de perguntássemos a todos os alunos, assim como evitou, quanto a nós, que nas respostas os alunos não fossem influenciados uns pelos outros, por estarem próximos e ouvirem as respostas uns dos outros. A partir das respostas recebidas construímos uma tabela de frequências e um gráfico de barras, que se apresentam a seguir

Prato preferido	Freq. abs	Freq. Rel.
Esparguete à Bolonhesa	13	0,1688
Peixe assado com batatas e legumes	6	0,0779
Filetes no forno com arroz e salada	10	0,1299
Bacalhau com natas e salada	21	0,2727
Rolo de carne com puré e salada	27	0,3506
Total	77	1



*Conclusões:* Da tabela e do gráfico anteriores sobressai que o prato preferido é o Rolo de carne com puré e salada, sendo o menos preferido o Peixe assado com batatas e legumes. Concluímos que aproximadamente 35% dos alunos inquiridos preferem o Rolo de carne com puré e salada, cerca de 27% preferem o Bacalhau com natas e salada, cerca de 17% o Esparguete à Bolonhesa, cerca de 13% os Filetes no forno com arroz e salada e finalmente só cerca de 8% dos alunos é que preferem o Peixe assado no forno com batatas e legumes. Como acreditamos que a nossa amostra foi bem seleccionada e é representativa dos alunos que almoçam na cantina, pensamos que estas proporções devem ser reflectidas para a população constituída por todos os alunos que almoçam na cantina. Assim, recomendamos que em 35% dos dias que na cantina decidirem fazer um destes pratos, se faça o Rolo de carne. Recomendamos ainda que em cerca de 27% desses dias se faça o Bacalhau com natas. Destes pratos, o peixe assado com batatas e legumes só deve ser feito cerca de 8% das vezes.

Nota: No processo de selecção dos alunos para responderem à questão de interesse, alguém do grupo sugeriu que se colocasse numa mesa, à entrada da cantina, um inquérito e se pedisse aos alunos para o preencherem e colocarem numa caixa, bem visível. Escusavam de estar a perder tempo à espera que os alunos fossem chegando para os interrogarem! No entanto depois de alguma troca de impressões abandonaram esta ideia, pois chegaram à conclusão que este processo de selecção da amostra, *por resposta voluntária*, conduz quase sempre a uma amostra enviesada, isto é uma amostra que não é representativa da população que se pretende estudar. Lembraram-se aliás, que este é um processo muitas vezes utilizado, erradamente, pela comunicação social para fazer sondagens, junto da população.

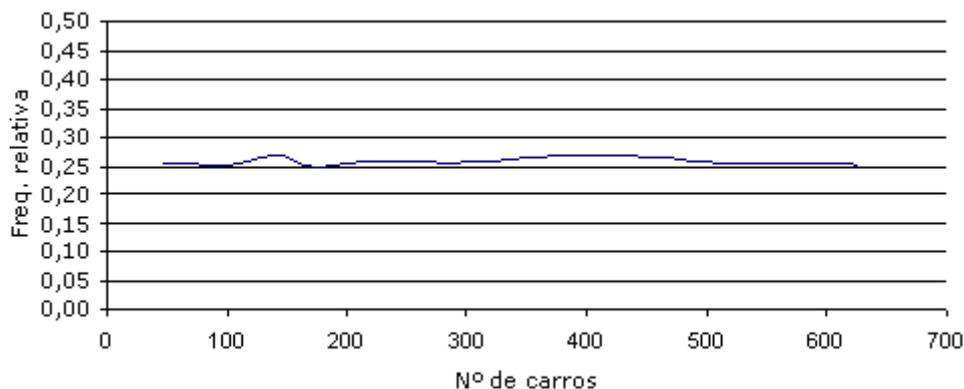
**Tarefa – Qual a probabilidade do próximo condutor utilizar a Via Verde?** O professor propôs aos alunos estimarem a probabilidade de um condutor, escolhido ao acaso de entre os que passam à frente da Escola, utilizar a Via Verde. Baseado numa notícia que tinha lido na comunicação social, o professor tinha formulado a conjectura de que 25% dos automobilistas utilizam a Via Verde. Serão os dados recolhidos pelos alunos consistentes com esta conjectura, ou pelo contrário, põem-na em causa?

A recolha de dados foi devidamente planeada entre o professor e os alunos, tendo estes sido distribuídos em grupos de dois alunos, em que cada grupo iria num dos intervalos, para a porta da escola, verificar quantos carros passavam e destes quantos tinham o identificador de Via Verde. Para uma recolha de dados mais eficiente, um dos elementos do grupo levava uma folha onde apontaria o que o colega lhe dissesse e que se traduzia em Sim ou Não. Por exemplo, suponhamos que o grupo 1 obteve o seguinte registo:

Via Verde	Registo das ocorrências	Nº de carros
Sim		11
Não		36
	Total	47

Ao fim de uma semana de recolha de dados, os 12 grupos de alunos juntaram os resultados obtidos na seguinte tabela, onde se registam também os valores acumulados e as frequências relativas correspondentes:

Grupo	Nº de carros com Via Verde	Nº de carros	Nº de carros com Via Verde acumulado	Nº de carros acumulado	Freq. rel. carros com Via Verde
1	12	47	12	47	0,2553
2	14	56	26	103	0,2524
3	12	38	38	141	0,2695
4	4	29	42	170	0,2471
5	15	49	57	219	0,2603
6	14	58	71	277	0,2563
7	19	65	90	342	0,2632
8	15	46	105	388	0,2706
9	17	73	122	461	0,2646
10	7	44	129	505	0,2554
11	14	57	143	562	0,2544
12	14	63	157	625	0,2512
Total	157	625			



O gráfico de linha anterior mostra a evolução da percentagem de carros com Via Verde, perto dos 25%, com tendência a aproximar-se deste valor, à medida que o número de carros vistos aumenta. Este resultado permite-nos inferir que a nossa conjectura estava certa e que efectivamente a percentagem de carros com Via Verde anda à volta de 25%. Assim, estimamos que a probabilidade de o próximo carro, que passa à frente da escola, ter Via Verde é de cerca de 25%.

### 6.3 Selecção de uma amostra de uma população com o objectivo de estimar uma probabilidade

Nas secções anteriores falamos frequentemente em seleccionar amostras de populações, com o objectivo de inferir para essas populações as propriedades verificadas na amostra. Assim, uma vez que se pretende estender à população as propriedades estudadas na amostra, esta tem que ser *representativa* da população, isto é, tem que reflectir a composição da população.

Uma amostra que não seja representativa da População diz-se *enviesada* e a sua utilização pode dar origem a interpretações erradas, como se sugere nos seguintes exemplos:

- ✓ utilizar uma amostra constituída por 10 benfiquistas, para prever o vencedor do próximo Benfica – Sporting!
- ✓ utilizar uma amostra constituída por leitores de determinada revista especializada, para tirar conclusões sobre a população em geral.

O planeamento de um estudo estatístico, que começa com a forma de seleccionar a amostra, deve ser feito de forma a evitar *amostras enviesadas*. Alguns processos que provocam quase sempre amostras enviesadas são, por exemplo, a *amostragem por conveniência* e a obtenção de uma amostra por *resposta voluntária*. Este último processo é usado, com muita frequência, pelas estações de televisão, com resultados por vezes contraditórios com os que se obtêm quando se utiliza um processo correcto de seleccionar a amostra. É comum, quando se está a debater um tema de interesse geral, como por exemplo a “Despenalização do aborto”, ou os “Touros de morte em Portugal”, os meios de comunicação social convidarem a população a ligar ou a enviar uma mensagem para um ou outro número conforme a opinião for de “Sim” ou “Não”. De um modo geral, é um segmento da população com muito interesse no tema que responde a esta solicitação, fazendo com que a amostra obtida seja enviesada.

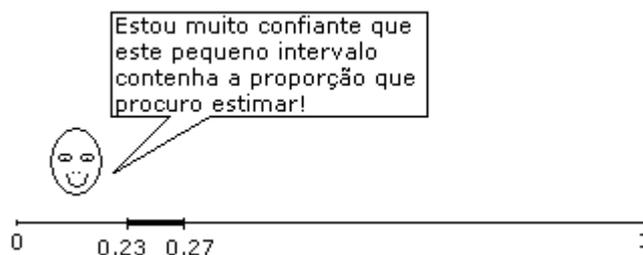
A utilização de uma amostragem por conveniência também se realiza frequentemente, quando se selecciona a amostra a partir de uma listagem dos elementos de determinado clube ou grupo, como por exemplo a Ordem dos Engenheiros ou a Associação dos Professores.

O problema da selecção da amostra é um problema para o qual, nesta fase, não é possível avançar nenhuma teoria, mas sobre o qual se podem tecer algumas considerações gerais, tanto no que diz respeito

- Ao número de elementos que devemos seleccionar, ou seja, à dimensão da amostra;
- Como à forma de seleccionar esses elementos da população, para constituírem a amostra.

No que diz respeito à dimensão da amostra:

- ✓ Esta dimensão depende muito da variabilidade da população subjacente. Por exemplo, se relativamente à população constituída pelos alunos do 10.º ano de uma escola secundária, estivermos interessados em estudar a média das suas idades, a dimensão da amostra a recolher não necessita de ser muito grande já que a variável idade apresenta valores muito semelhantes, numa classe etária muito restrita. No entanto se a característica a estudar for o tempo médio que os alunos levam a chegar de casa à escola, já a amostra terá de ter uma dimensão maior, uma vez que a variabilidade da população é muito maior. Cada aluno pode apresentar um valor diferente para esse tempo. Num caso extremo, se numa população a variável a estudar tiver o mesmo valor para todos os elementos, então bastaria recolher uma amostra de dimensão 1 para se ter informação completa sobre a população; se, no entanto, a variável assumir valores diferentes para todos os elementos, para se ter o mesmo tipo de informação, com a mesma precisão, seria necessário investigar todos os elementos. Assim, quanto maior for a variabilidade da característica que se está a estudar, maior terá que ser a dimensão da amostra a recolher.
- ✓ A dimensão da amostra terá de ser tanto maior, quanto maior for a *precisão* exigida. Existem técnicas que permitem obter valores mínimos para as dimensões das amostras a recolher e que garantem estimativas com uma determinada *precisão* exigida à partida. Uma vez garantida essa precisão, a opção por escolher uma amostra de maior dimensão, é uma questão a ponderar entre os *custos envolvidos* e o *ganho com o acréscimo de precisão*. Nos exemplos da secção anterior esta exigência já foi manifestada. Por exemplo, quando numa tarefa da secção anterior, procurámos estimar a proporção de condutores com Via Verde, obviamente que sabemos à partida que o intervalo (0, 1) contém essa proporção. Mas este saber não nos adianta nada! Nós precisamos de uma maior precisão, isto é, de dois valores  $a$  e  $b$ , entre 0 e 1, tal que o intervalo  $[a, b]$  tenha uma pequena amplitude (quanto menor for a amplitude do intervalo, maior é a precisão) e que com uma “grande confiança” contenha essa proporção:



Note-se que a confiança de que estamos a falar pode ser medida em termos de probabilidade (a estudar ao nível do ensino secundário).

- ✓ Convém ainda observar que a dimensão da amostra a recolher não é directamente proporcional à dimensão da população a estudar, isto é, se por exemplo para uma população de dimensão 1000 uma amostra de dimensão 100 for suficiente para o estudo de determinada característica, não se exige necessariamente uma amostra de dimensão 200 para estudar a mesma característica de uma população análoga, mas de dimensão 2000, quando se pretende obter a mesma precisão. Como dizia George Gallup, um dos pais da consulta da opinião pública (Tannenbaum, 1998, p. 438): "*Whether you poll the United States or New York State or Baton Rouge (Louisiana) ... you need ... the same number of interviews or samples. It's no mystery really – if a cook has two pots of soup on the stove, one far larger than the other, and thoroughly stirs them both, he doesn't have to take more spoonfuls from one than the other to sample the taste accurately*".
- ✓ Finalmente chama-se a atenção para o facto de que se o processo de amostragem originar uma amostra enviesada, aumentar a dimensão não resolve nada, antes pelo contrário! Por exemplo, quando pretendemos estimar a altura média dos alunos de uma escola, não é pelo facto de se perguntar a altura a todos os elementos da equipa de basquete da escola, em vez de só a alguns, que obtemos uma melhor estimativa para essa altura média de todos os alunos da escola!

### 6.3.1 Amostra aleatória simples sem reposição e com reposição

Tão importante como a dimensão da amostra é a forma como os elementos são seleccionados da população. Como podemos estar confiantes de que a amostra obtida é representativa? Um princípio fundamental que se tem de ter presente é o da *aleatoriedade*. Temos de utilizar um processo que garanta que qualquer elemento da população tenha alguma possibilidade de pertencer à amostra:

*Amostra aleatória ou probabilística e amostra não aleatória* – Dada uma população, uma amostra aleatória ou probabilística é uma amostra tal que qualquer elemento da população tem alguma probabilidade de ser seleccionado para a amostra. Numa amostra não aleatória, alguns elementos da população podem não ter qualquer possibilidade de serem seleccionados para a amostra.

Suponha o caso de um aluno que foi encarregue de seleccionar uma amostra de alunos da escola, para averiguar quantas horas passam por dia à frente da televisão. Este aluno decide só perguntar aos colegas do sexo feminino! Esta

amostra não é aleatória, pois há parte dos alunos que nunca poderão ser seleccionados para pertencer à amostra.

Existem algumas técnicas para obter amostras aleatórias. Exemplificamos duas dessas técnicas que conduzem às *amostras aleatórias simples* e *amostras estratificadas*.

**Amostra aleatória simples sem reposição** – Dada uma população, 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 seleccionado.

Uma amostra destas pode ser escolhida sequencialmente da população, escolhendo um elemento de cada vez, sem reposição, pelo que em cada selecção cada elemento tem a mesma probabilidade de ser seleccionado. Exemplificamos a seguir um processo de obter uma amostra aleatória simples.

**Exemplo – Como seleccionar alunos de uma turma.** Consideremos a população constituída pelos 18 alunos de uma turma do 10.º ano de uma determinada escola secundária, em que a característica de interesse a estudar é a altura média desses alunos. Uma maneira possível de recolher desta população uma amostra aleatória, seria escrever cada um dos indicadores ( $n.º$  do aluno, nome, ...) dos elementos da população num quadrado de papel, inserir todos esses bocados de papel numa caixa e depois seleccionar tantos quantos a dimensão da amostra desejada. A recolha tem de ser feita *sem reposição* pois quando se retira um papel (elemento da população), ele não é repostado enquanto a amostra não estiver completa (com a dimensão desejada). Qualquer conjunto de números recolhidos desta forma dará origem a uma amostra aleatória, constituída pelas alturas dos alunos seleccionados.

O processo que acabámos de descrever não é prático se a população a estudar tiver dimensão elevada. Neste caso, um processo de seleccionar uma amostra aleatória simples consiste em utilizar uma opção da calculadora, a função `randInt(i,j)`, que gera números aleatórios<sup>8</sup> inteiros dentro dos limites especificados  $i$  e  $j$ . Para seleccionar uma amostra de uma população utilizando a calculadora procede-se em duas etapas:

- ✓ atribui-se um número a cada elemento da população, sequencialmente, por exemplo de 1 até  $N$  (dimensão da população);
- ✓ utiliza-se a calculadora para gerar números inteiros entre 1 e  $N$ . Têm de se gerar, pelo menos, tantos números quantos os necessários para constituírem a amostra. Dizemos pelo menos, pois se durante o processo da geração se obtiver algum número igual a algum que já tenha saído, deita-se esse número fora e gera-se um outro.

**Exemplo (cont.)** - Considerando a população do exemplo anterior, constituída por 18 elementos, vamos numerá-los com os números 1, 2, 3, ..., 17, 18 (podia ser utilizado qualquer outro conjunto de 18 números sequenciais). Para seleccionar uma amostra de dimensão 4 geramos 4 números na calculadora, utilizando a opção `randInt(1,18)`. Como dissemos anteriormente, se se pretende uma amostra aleatória simples sem reposição, quando se geram os números, se se obtiverem

---

<sup>8</sup> Pseudo-aleatórios

números repetidos, tem de se proceder a novas gerações para se obterem números diferentes.

Uma alternativa ao tipo de amostragem descrito anteriormente é considerar a amostra com reposição:

**Amostra aleatória com reposição** – para obter uma amostra aleatória com reposição, quando um elemento é seleccionado da população, verifica-se a característica de interesse e repõe-se na população, antes de se seleccionar o próximo elemento.

Embora saia fora do âmbito deste nível, podemos adiantar que estes dois processos de amostragem não são equivalentes se a dimensão da população não for suficientemente grande. Uma inferência baseada numa amostra de dimensão  $n$ , recolhida sem reposição é, de um modo geral, mais correcta do que a baseada numa amostra da mesma dimensão, mas recolhida com reposição. No entanto, se a dimensão da população for muito grande, quando comparada com a dimensão da amostra que se recolhe, já podemos considerar os dois processos equivalentes. É aliás simples de explicar porque é que se verifica esta situação: se a população tiver uma grande dimensão, a probabilidade de seleccionar o mesmo elemento duas vezes é muito pequena, pelo que é praticamente indiferente que o elemento seleccionado para pertencer à amostra seja posteriormente repostado ou não, antes de seleccionar o próximo elemento.

Quando é que se costuma considerar que a população é muito grande, utilizando-se até o termo “população infinita”? É quando a sua dimensão é pelo menos 20 vezes maior que a dimensão da amostra que se pretende seleccionar.

### 6.3.2 Amostra estratificada

Por vezes sabemos que a nossa população é constituída por alguns grupos mais ou menos homogéneos entre si, relativamente à característica que se está a estudar. Neste caso dizemos que a população é constituída por estratos e a melhor forma de recolher uma amostra desta população é recolher uma *amostra estratificada*. Por exemplo se se pretende estimar a idade média dos alunos de uma escola secundária, é possível considerar três estratos mais ou menos homogéneos quanto à característica Idade e que são os anos de escolaridade – 10.º, 11.º e 12.º anos. Posteriormente, selecciona-se de cada um destes estratos uma percentagem de elementos que irão constituir a amostra, sendo esta percentagem, de um modo geral, proporcional à dimensão dos estratos.

**Amostra estratificada** - Uma vez identificados os estratos, extrai-se de cada um destes estratos uma amostra, de forma aleatória. O conjunto de todas estas amostras constitui a amostra pretendida.

**Tarefa – O meio de transporte utilizado pelos alunos para irem para a escola. Como seleccionar uma amostra para fazer este estudo?**



A Junta de Freguesia estava interessada em conhecer o meio de transporte utilizado pelos alunos para irem para a escola da sua zona, que só tinha o 3º ciclo, pois pretendia saber se seria necessário pedir um reforço nas carreiras de autocarros que servem a escola.

Assim, pediu ao Director que lhe fornecesse a informação desejada. Conhecedor deste pedido, um grupo de 3 alunos, que pretendia fazer um pequeno projecto sobre Estatística, no âmbito dos seus trabalhos escolares, ofereceu-se para realizar este estudo.

O primeiro passo para a realização do estudo, foi a planificação da recolha de dados. A amostra a recolher deveria reflectir a constituição da população, isto é, se um dos anos tivesse mais alunos seria necessário recolher para a amostra mais alunos desse ano. Então, terá de constituir-se a amostra com uma percentagem de alunos de cada ano de escolaridade, de modo a representar, o melhor possível, todos os alunos da escola. Esta percentagem teria de ser calculada em função do número de alunos que se pretendessem inquirir. Por exemplo, para simplificar, admitamos que a escola tem um total de 469 alunos, assim distribuídos pelos 3 anos:

7.º ano – 140 alunos; 8.º ano – 148 alunos; 9.º ano – 181 alunos

Admitamos que, depois de algumas considerações sobre o número de alunos a inquirir, se tinha decidido recolher uma amostra de dimensão 75. O problema da escolha da dimensão da amostra sai fora do âmbito deste nível de ensino. Deve-se, no entanto, referir que quanto maior for a dimensão da amostra, mais fiáveis serão os resultados, mas também mais tempo será necessário para os obter (neste caso não se põe a questão dos gastos em dinheiro, como é normalmente a situação deste tipo de estudos). Neste caso, vamos admitir que o grupo decidiu inquirir 75 colegas, pois acharam razoável que cada um fizesse cerca de 25 entrevistas.

O 1.º passo para a selecção da amostra é calcular a proporção de cada ano, na população:

Ano	Nº de alunos por ano	Proporção
7.º	140 (6 turmas)	0,298507
8.º	148 (6 turmas)	0,315565
9.º	181 (7 turmas)	0,385928
Total	469	

O 2.º passo é calcular quantos alunos se vão seleccionar de cada ano para pertencerem à amostra, o que será feito tendo em conta as proporções obtidas anteriormente:

Ano	Proporção	Nº de alunos da amostra	
7.º	0,298507	$0,298507 \times 75 = 22,39$	22
8.º	0,315565	$0,315565 \times 75 = 23,67$	24
9.º	0,385928	$0,385928 \times 75 = 28,94$	29
Total			75

Perante os resultados anteriores decidiram que cada um dos 3 alunos que constituía o grupo que estava a realizar o estudo ficaria encarregue de inquirir cada ano de escolaridade.

Como seleccionar agora os alunos de cada ano? Podem definir-se várias estratégias. Por exemplo, no caso do 7.º ano de escolaridade, com 6 turmas, de 4 das turmas escolhem aleatoriamente 4 alunos e das outras 2 escolhem 3 alunos. Esta escolha pode ser feita através do n.º do aluno, utilizando a função `randInt(1, n.º alunos da turma)`, da calculadora, ou então utilizando o seguinte processo para a escolha dos alunos de uma das turmas com 25 alunos: quando tocar para um dos

intervalos, são seleccionados, por exemplo, o 2.º, o 8.º, o 14.º e o 20.º alunos a saírem da sala.

Este processo que acabamos de descrever, de seleccionar o 2.º, o 8.º, o 14.º e o 20.º, tem a particularidade, como se nota imediatamente, de a diferença entre os números ser constante e neste caso igual a 6. Temos aqui um caso particular da amostragem sistemática, que desenvolveremos mais em pormenor na secção seguinte.

### 6.3.3 – Amostra sistemática

Uma alternativa à amostragem aleatória simples sem reposição, é a amostragem sistemática, particularmente indicada quando temos uma listagem dos indivíduos da População. Por exemplo, se pretendermos seleccionar uma amostra de 75 alunos de uma Escola com 1350 alunos, considera-se um ficheiro com o nome dos 1350 alunos ordenados por ordem alfabética (ou por qualquer outra ordem). Considera-se o quociente  $1350/75=18$  e dos primeiros 18 elementos da lista, selecciona-se um aleatoriamente. A partir deste elemento seleccionamos sistematicamente todos os elementos distanciados de 18 unidades. Assim, se o elemento seleccionado aleatoriamente de entre os primeiros 18, foi o 14, os outros elementos a serem seleccionados são 32 ( $=14+18$ ), 50 ( $=32+18$ ), 68 ( $=50+18$ ), etc. Obviamente que o quociente entre a dimensão da população e a da amostra não é necessariamente inteiro, como anteriormente, mas não há problema pois considera-se a parte inteira desse quociente.

Na secção anterior apresentou-se um exemplo de uma amostragem sistemática. Numa turma com 25 alunos, em que se pretendia seleccionar 4 alunos, utilizou-se o seguinte processo:

- 1) Fez-se o quociente  $25/4$ , tendo-se obtido a parte inteira igual a 6;
- 2) Dos alunos com os números de 1 a 6, escolheu-se aleatoriamente 1, tendo-se obtido, por exemplo, o número 2;
- 3) Os alunos seleccionados são os números 2, 8 ( $=2+6$ ), 14 ( $=8+6$ ) e 20 ( $=14+6$ ).

**Amostra aleatória sistemática** – Dada uma população de dimensão  $N$ , ordenada por algum critério, se se pretende uma amostra de dimensão  $n$ , escolhe-se aleatoriamente um elemento de entre os  $k$  primeiros, onde  $k$  é a parte inteira do quociente  $N/n$ . A partir desse elemento escolhido, escolhem-se todos os  $k$ -ésimos elementos da população para pertencerem à amostra.

## 6.4 Experiência aleatória, Espaço de resultados, Acontecimentos.

Como dissemos na secção 3.3, o objectivo da Estatística é o estudo de *Populações*, isto é conjunto de indivíduos (não necessariamente pessoas) com algumas características comuns que se pretendem estudar. Por exemplo, podemos estar interessados em estudar a característica *Número de irmãos* de cada aluno de uma escola. O fenómeno que consiste em observar esta característica é um *fenómeno aleatório*, pois não sabemos, antes de fazer a pergunta ao aluno, qual a resposta que ele vai dar. No entanto já não teria interesse averiguar a característica "O aluno tem nome?", pois sabemos que neste caso todos os alunos teriam de dar a mesma resposta e que é "Sim".

Assim, a observação de uma característica que tenhamos interesse em estudar e a que chamámos *Variável* (secção 3.3), não é mais que a observação de um fenómeno aleatório, caracterizado pela existência de *variabilidade* dos seus resultados e em que a preocupação permanente é a de construir um modelo que traduza essa variabilidade.

Um princípio fundamental da Estatística é compreender que:

*A variabilidade existe e pode ser modelada*

Existem situações, que são aliás as mais correntes em Estatística, em que não é possível observar a característica em estudo em todos os elementos da População estudando-se só uma parte da População - a Amostra.

Como o nosso objectivo é inferir propriedades para a população a partir do estudo dos dados da amostra, as amostras constituídas pelos dados recolhidos devem representar convenientemente as populações subjacentes. Como vimos na secção anterior, dizemos que neste caso as amostras são *representativas* das populações de onde foram seleccionadas.

Como também vimos na secção anterior, esta fase da selecção de uma amostra com o objectivo de tirar conclusões para a população é muito importante, pois se a amostra não for convenientemente seleccionada, as conclusões que depois retiramos para a população podem ser falsas. Por exemplo, não seria correcto, recolher informação junto de alguns rapazes da escola sobre quantas horas passam por semana a jogar no computador, se pretendermos saber quantas horas todos os alunos da escola gastam nessa actividade. De um modo geral, as raparigas não são tão entusiastas desta forma de passar o tempo...

Assim, ao admitir a representatividade de uma amostra, seleccionada de uma população com o objectivo de estudar determinada característica, estamos a admitir que a *proporção de indivíduos na população, com essa característica, é preservada na amostra*. Por outro lado, ao inferir para a população as propriedades verificadas na amostra, estamos também a admitir a preservação da proporção verificada na amostra, para a população, sendo esta a base do raciocínio inferencial em Estatística, isto é do raciocínio que nos permite a partir do conhecimento da "parte" conjecturar para o "todo" e posteriormente tomar decisões, quantificando o erro das decisões tomadas (questão que se situa fora do âmbito desta brochura).

Ao processo utilizado para a aquisição dos *dados* que constituem a amostra, damos o nome de *experiência aleatória*. Como já referimos anteriormente, **experiência aleatória** é o processo de observar um resultado de um fenómeno aleatório.

Quando se realiza uma experiência aleatória:

- obtém-se um *resultado*, de entre um conjunto de resultados conhecidos de antemão,
- mas *não se tem conhecimento suficiente* sobre o resultado que irá sair em cada realização da experiência.
- Admite-se ainda que a experiência se pode *repetir* e que as repetições são realizadas nas mesmas circunstâncias e são *independentes*.

A experiência tem de se repetir tantas vezes quantas as observações necessárias para constituir a nossa amostra. Na selecção de uma amostra temos de ter em consideração que se deve recolher um número razoável de dados para que as nossas inferências sejam mais precisas.

Um ponto importante a referir, é que, embora seleccionando o mesmo número de elementos da população, raramente se obtêm duas amostras com os mesmos dados. É precisamente esta aleatoriedade presente na repetição da experiência e por conseguinte, no processo de recolha de dados, ou dito ainda de outro modo, na *selecção da amostra*, que, ao produzir um determinado *padrão de comportamento*, nos vai permitir inferir para a População as propriedades verificadas na Amostra que entretanto se seleccionou.

Se, por exemplo, estivermos interessados em averiguar se uma moeda é equilibrada, temos de repetir a experiência de lançar a moeda um número “grande” de vezes. Se após um grande número de lançamentos da moeda se verifica que a percentagem de vezes que a face Euro surge, é superior à percentagem de vezes da face Nacional, podemos inferir que a moeda não é equilibrada. Não poderíamos tirar esta conclusão, mesmo que em 10 lançamentos da moeda a face Euro tivesse surgido 8 vezes, pois uma repetição de 10 experiências não permite visualizar nenhum padrão de comportamento da moeda.

A selecção aleatória de uma amostra, de *dimensão razoável*, permite que os resultados do estudo da amostra possam ser estendidos para a População (ver secção 6.3 sobre a selecção da amostra).

São exemplos de experiências aleatórias:

- Perguntar a uma pessoa ao acaso, da sua cidade, quantas pessoas constituem o seu agregado familiar;
- Perguntar a um aluno ao acaso, da escola, qual o animal doméstico preferido;
- Lançar uma moeda de 1 Euro ao ar e ver o resultado que sai;
- Lançar uma moeda de um euro ao ar 10 vezes e ver quantas vezes sai a face euro;
- Ao acordar, de manhã, ir à janela e num período de 5 minutos, ver quantos carros encarnados passam;
- Medir o tempo que de manhã se leva a chegar à escola;
- Perguntar a um aluno ao acaso, da escola, quantas mensagens de telemóvel enviou no dia anterior;
- Escolher ao acaso 3 alunos da turma (com 10 rapazes) e verificar quantos são rapazes.

As situações anteriores são exemplos de experiências aleatórias, porque além de envolverem aleatoriedade, o que se pretende observar está bem especificado. O mesmo não se passa com a seguinte situação: *ao acordar, de manhã, ir à janela*.

Efectivamente, na situação anterior não se especificou o que se pretende observar, ou seja, qual o fenómeno aleatório em estudo, de modo a termos uma experiência aleatória. No entanto, associado à situação anterior são experiências aleatórias (Graça Martins et al, 1999):

- *Ao acordar, de manhã, ir à janela e ver se chove;*
- *Ao acordar, de manhã, ir à janela e contar num período de 5 minutos, quantos carros encarnados passam.*

Ao conjunto de todos os resultados possíveis associados à realização de uma experiência aleatória, chamamos espaço de resultados ou espaço amostral.

**Espaço de resultados** – conjunto cujos resultados são os que consideramos como possíveis, ao modelar um fenómeno aleatório.

**Exemplo – Espaços de resultados.** Para cada uma das experiências aleatórias consideradas anteriormente construa o espaço de resultados associado.

- Nº de pessoas do agregado familiar  
{1, 2, 3, 4, ...}
- Perguntar a um aluno ao acaso, da escola, qual o animal doméstico preferido {cão, gato, peixe, passarinho, tartaruga, coelho, hámster, rato, tartaruga, ...}
- Lançar uma moeda de 1 Euro ao ar e ver o resultado que sai  
{face Euro, face Nacional}
- Lançar uma moeda de um euro ao ar 10 vezes e ver quantas vezes sai a face euro  
{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
- Medir o tempo que leva a chegar à escola de manhã (em minutos)  
{1, 2, 3, 4, 5, 6, 7, ... }
- Perguntar a um aluno da escola, escolhido ao acaso, quantas mensagens de telemóvel enviou no dia anterior  
{0, 1, 2, 3, 4, 5, 6, 7, ... }
- Escolher ao acaso 3 alunos da turma (com 10 rapazes) e verificar quantos são rapazes  
{0, 1, 2, 3}
- Ao acordar, de manhã, ir à janela e ver se chove  
{chove, não chove}
- Ao acordar, de manhã, ir à janela e contar num período de 5 minutos, quantos carros encarnados passam  
{0, 1, 2, 3, 4, ...}

A definição correcta do espaço de resultados associados a uma experiência é um passo fundamental para de seguida definirmos acontecimentos. No entanto, nem sempre esta definição é simples, sendo um processo que, por vezes, envolve alguma idealização no *modelo* utilizado para interpretar a realidade.

**Exemplo – Lançamento da moeda**<sup>9</sup>. Admita que tem uma moeda de um Euro, equilibrada. Mas o que é uma moeda equilibrada? É a moeda relativamente à qual

---

<sup>9</sup> Graça Martins (2005), p. 128

se admite, à partida, que existe igual possibilidade de sair face Euro ou face Nacional ou no próximo lançamento que façamos com ela – estamos a admitir o *princípio da simetria*, de que falaremos a seguir. Estamos, assim, a pensar num *modelo matemático* para traduzir o facto de que em qualquer lançamento da moeda, só temos dois resultados possíveis, face Euro e face Nacional e em que a probabilidade de sair a face Euro é igual à de sair a face Nacional e igual a 1/2:



Modelo para o resultado do lançamento da moeda equilibrada

Resultado	Face Euro	Face Nacional
Probabilidade	1/2	1/2

Não nos estamos a preocupar, por exemplo, com a força ou direcção com que atiramos a moeda, nem tão pouco com o desgaste acusado pela moeda após sucessivos lançamentos! Também não estamos a encarar a hipótese da moeda cair de pé! Se nos estivéssemos a preocupar em arranjar um modelo que traduzisse mais fielmente a realidade, estaríamos a arranjar um modelo matemático tão complicado que seria impossível de tratar e não nos serviria para nada. O estatístico George Box dizia:

*Todos os modelos são maus, alguns modelos são úteis.*

Assumindo então o *modelo* anterior, um pouco simplista, para o lançamento da moeda, se lançarmos a moeda repetidas vezes, esperamos que o número de faces Euro seja aproximadamente metade do número de lançamentos. Se, por outro lado, recolhermos uma amostra de dimensão 1, isto é, se fizermos um único lançamento, não sabemos qual o resultado que se vai verificar, se será face Euro ou face Nacional, mas dizemos que a probabilidade de sair face Euro é 1/2.

Como refere Bartholomew (1995) “*We all depend on models to interpret our everyday experiences. We interpret what we see in terms of mental models constructed on past experience and education. They are the constructs that we use to understand the pattern of our experiences*”.

O comportamento de grandes grupos de indivíduos, pode ser também considerado aleatório e o processo utilizado para definir um modelo, é o de verificar o que é que se passa com um grande conjunto de indivíduos (Graça Martins, 2005), como exemplificamos a seguir.

**Exemplo – Tempo de vida**<sup>10</sup>. Se nos perguntassem qual a probabilidade de uma determinada pessoa morrer no próximo ano, obviamente que não saberíamos dizer. No entanto, se observarmos milhões de pessoas, poderemos obter um padrão para o comportamento das mortes. É assim que poderemos dizer que a proporção de homens, com idades compreendidas entre 25 e 34 anos, que morrerão no próximo ano, anda à volta de 0,21%. Esta percentagem, verificada para um conjunto grande de indivíduos, será entendida como a *probabilidade* de que um homem jovem morra no próximo ano. Para as mulheres com aquela idade, a probabilidade de morrer será cerca de 0,07%. Estamos, a partir da observação de resultados verificados numa amostra, a inferir para toda a população constituída pelos indivíduos da classe etária considerada. Estes modelos têm muito interesse para as

<sup>10</sup> Moore (1997).

companhias de seguros, quando se trata nomeadamente de seguros de vida, já que lhes vai permitir definir uma política de preços para as apólices, sendo até natural que cobrem mais por um seguro de vida a um homem, do que a uma mulher.

Pode ainda acontecer que tenhamos de idealizar um modelo que não corresponde à realidade, mas para o qual não exista outra possibilidade de o definir. Por exemplo se pensarmos na experiência aleatória que consiste em averiguar o tempo de vida  $T$  de uma pessoa escolhida ao acaso, consideramos para espaço de resultados  $S = \{\text{Todos os valores de } T, \text{ tal que } T > 0\}$ . Será que uma pessoa pode ter 500 anos? E 400? E 200? Temos dificuldade em estabelecer um limite superior para o valor de  $T$ , pelo que temos de nos abstrair um pouco da realidade considerando aquele modelo para o espaço de resultados.

**Acontecimento** – É um resultado ou um conjunto de resultados do espaço de resultados.

Considerando a experiência aleatória que consiste em perguntar a uma pessoa residente na sua cidade, escolhida ao acaso, quantas pessoas constituem o seu agregado familiar, o espaço de resultados é constituído por todos os inteiros não negativos (excluído o zero). Alguns acontecimentos são:

- 3 pessoas, que podemos representar por  $\{3\}$
- Entre 2 e 4 pessoas (inclusive), que podemos representar por  $\{2, 3, 4\}$
- Mais de 3 pessoas, que podemos representar por  $\{4, 5, 6, \dots\}$
- Menos de 10 pessoas, que podemos representar por  $\{1, 2, 3, \dots, 9\}$

De um modo geral os acontecimentos são representados por letras maiúsculas  $A$ ,  $B$ , etc. Diz-se que se *realizou* o acontecimento  $A$ , quando o resultado da experiência pertence a  $A$ .

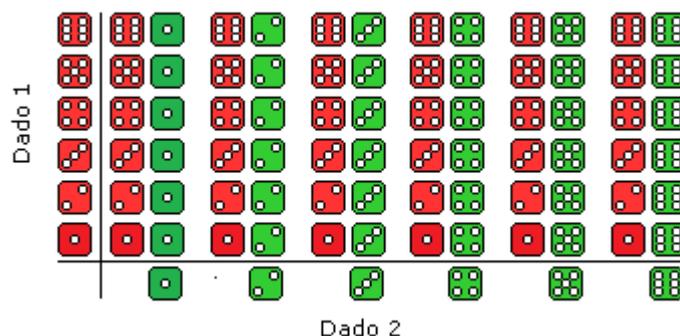
Quando os acontecimentos são constituídos por um único resultado, dizem-se **acontecimentos elementares**.

### Exemplo – Lançamento de dois dados



Considerando a experiência aleatória que consiste em lançar dois dados e verificar as faces que ficam voltadas para cima, pretende-se identificar o espaço de resultados e os acontecimentos “o número de pintas é igual nos dois dados” e “a soma das pintas é 7”.

Para descrever o espaço de resultados vamos considerar dois dados, um vermelho e um verde, para os distinguir. O espaço de resultados é constituído por todos os pares de dados considerados na figura a seguir. O número de elementos do espaço de resultados é  $36 = 6 \times 6$ .



Chamamos a atenção que, por exemplo, o par (1,3) não é o mesmo que o par (3,1). No par ordenado, o primeiro elemento refere-se a um dos dados (neste caso o dado vermelho) e o segundo elemento refere-se ao outro dado (o dado verde).

O acontecimento A, "o número de pintas é igual nos dois dados" é constituído pelos pares



ou em notação em termos dos pares ordenados, referindo-nos ao número de pintas

$$A = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$$

Finalmente o acontecimento "a soma das pintas é 7" é constituído pelos pares



ou em notação em termos dos pares ordenados

$$B = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$$

Qual a diferença entre o espaço de resultados associado à experiência aleatória do lançamento de dois dados e a experiência que consiste no lançamento do mesmo dado duas vezes? O espaço de resultados é idêntico nas duas experiências. Considerámos dados de cores distintas para justificar a nossa opção para descrever o espaço de resultados como um conjunto de pares ordenados, mas é óbvio que este mesmo espaço serve para modelar o lançamento de dois dados idênticos ou dois lançamentos de um mesmo dado.

Associado à experiência que acabámos de descrever no exemplo anterior, poderíamos ter considerado o seguinte espaço de resultados S:

$S = \{\text{saírem dois 1's, sair um 1 e um 2, sair um 1 e um 3, sair um 1 e um 4, sair um 1 e um 5, sair um 1 e um 6, saírem dois 2's, sair um 2 e um 3, sair um 2 e um 4, sair um 2 e um 5, sair um 2 e um 6, saírem dois 3's, sair um 3 e um 4, sair um 3 e um 5, sair um 3 e um 6, saírem dois 4's, sair um 4 e um 5, sair um 4 e um 6, saírem dois 5's, sair um 5 e um 6, saírem dois 6's}\}$ .

Qual a desvantagem em considerar este espaço de resultados? Como veremos mais à frente, se o espaço de resultados for constituído por resultados igualmente possíveis, o que não acontece nesta situação, podemos utilizar uma regra conhecida pela regra de Laplace, para atribuir probabilidades a acontecimentos associados ao fenómeno em estudo.

Ainda associado ao lançamento dos dois dados, suponhamos que o fenómeno aleatório que estávamos interessados em estudar, era o resultado da soma das pintas das faces que ficam voltadas para cima. Neste caso o espaço de resultados é

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

**Nota histórica** (Freedman et al. 1991) - No século XVII, os jogadores italianos costumavam fazer apostas sobre o número total de pintas obtidas no lançamento de 3 dados. Acreditavam que a possibilidade de obter um total de 9 era igual à possibilidade de obter um total de 10. Por exemplo, diziam que uma combinação possível para dar um total de 9 seria

1 pinta num dos dados, 2 pintas num outro dado, 6 pintas no terceiro dado

Abreviando o resultado anterior para "1 2 6", todas as combinações para dar o 9 são:

1 2 6    1 3 5    1 4 4    2 3 4    2 2 5    3 3 3

Analogamente, obtinham 6 combinações para o 10:

1 4 5    1 3 6    2 2 6    2 3 5    2 4 4    3 3 4

Assim, os jogadores argumentavam que o 9 e o 10 deveriam ter a mesma possibilidade de se verificarem. Contudo, a experiência mostrava que o 10 aparecia com uma frequência um pouco superior ao 9. Pediram a Galileu que os ajudasse nesta contradição, tendo este realizado o seguinte raciocínio: Pinte-se um dos dados de branco, o outro de cinzento e o outro de preto. De quantas maneiras se podem apresentar os três dados depois de lançados? O dado branco pode apresentar 6 possibilidades diferentes. Para cada uma destas possibilidades o dado cinzento pode apresentar 6 possibilidades, obtendo-se  $6 \times 6$  possibilidades para os dois dados. Correspondendo a cada uma destas possibilidades, o dado preto pode apresentar 6 possibilidades obtendo-se no total  $6 \times 6 \times 6 = 216$  possibilidades. Galileu listou todas as 216 maneiras de 3 dados se apresentarem depois de lançados. Depois percorreu a lista e verificou que havia 25 maneiras de obter um total de 9 e 27 maneiras de obter um total de 10. O raciocínio dos jogadores não entrava em linha de conta com as diferentes maneiras como os dados se podiam apresentar. Por exemplo o triplo "3 3 3", que dá o 9, corresponde unicamente a uma forma de os dados se apresentarem, mas o triplo "3 3 4" que dá o 10, corresponde a 3 maneiras diferentes:



pelo que o raciocínio dos jogadores deve ser corrigido de acordo com a tabela seguinte:

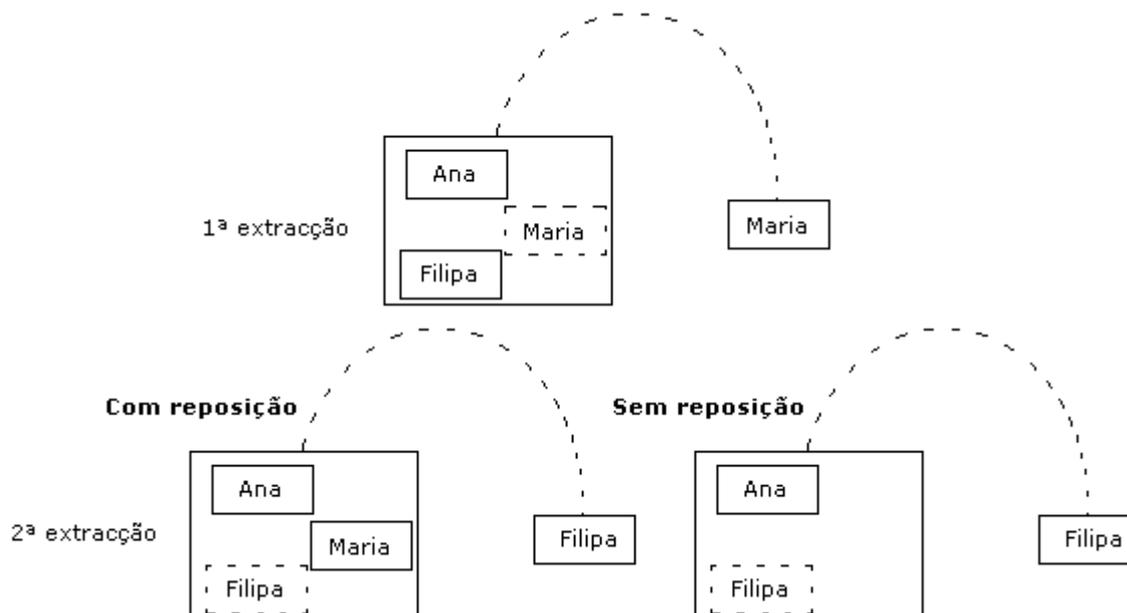
Tripos para o 9			Nº de maneiras de obter o triplo	Tripos para o 10			Nº de maneiras de obter o triplo
1	2	6	6	1	4	5	6
1	3	5	6	1	3	6	6
1	4	4	3	2	2	6	3
2	3	4	6	2	3	5	6
2	2	5	3	2	4	4	3
3	3	3	1	3	3	4	3
Total			25	Total			27

Por vezes para definirmos o espaço de resultados associados com determinadas experiências, é necessário acrescentar algo sobre a metodologia da realização da experiência. Por exemplo se pretendermos obter o espaço de resultados associado à experiência aleatória que consiste em retirar duas bolas de uma caixa contendo quatro bolas brancas e duas pretas, é necessário saber se após retirar a primeira bola ela é repostada ou não na caixa.

Na secção 6.3.1 já falámos no processo de amostragem sem reposição e com reposição. Vamos de seguida exemplificar esse processo, no contexto da construção do espaço de resultados.

### Extracções com reposição e sem reposição

Colocaram-se<sup>11</sup> numa caixa 3 papéis com o nome de 3 meninas: Ana, Maria e Filipa. Considere a experiência aleatória que consiste em retirar da caixa 2 papéis e verificar os nomes que saíram. Qual o espaço de resultados? Para responder a esta questão é necessário saber se a extracção se faz *com reposição*, isto é, se uma vez retirado um papel e verificado o nome se volta a colocar o papel na caixa, antes de proceder à extracção seguinte, ou se a extracção é feita *sem reposição*, isto é, uma vez retirado um papel, ele não é repostos antes de se proceder à próxima extracção. No esquema seguinte procuramos representar as duas situações:



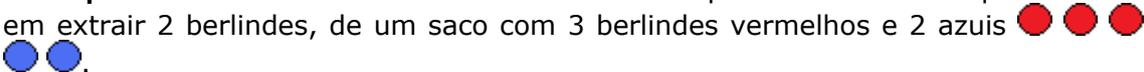
Admitimos que na 1.<sup>a</sup> extracção saiu o papel com o nome da Maria. Na 2.<sup>a</sup> extracção, saiu o nome da Filipa nos dois casos, mas *na extracção com reposição* havia uma possibilidade em três de ele sair, tal como na 1.<sup>a</sup> extracção, enquanto que na *extracção sem reposição* havia uma possibilidade em duas de ele sair. Quer dizer que neste caso havia uma maior probabilidade de sair o nome da Filipa. Os espaços de resultados correspondentes às duas situações com reposição e sem reposição, são:

Espaço de resultados com reposição	Espaço de resultados sem reposição
Ana, Ana	Ana, Maria
Ana, Maria	Ana, Filipa
Ana, Filipa	Maria, Ana
Maria, Ana	Maria, Maria
Maria, Maria	Maria, Filipa
Maria, Filipa	Filipa, Ana
Filipa, Ana	Filipa, Maria
Filipa, Maria	Filipa, Filipa
Filipa, Filipa	

<sup>11</sup> Graça Martins et al (1999).

O acontecimento "saiu o nome da Maria" é constituído pelos seguintes resultados, considerando a extracção com reposição e sem reposição, respectivamente:

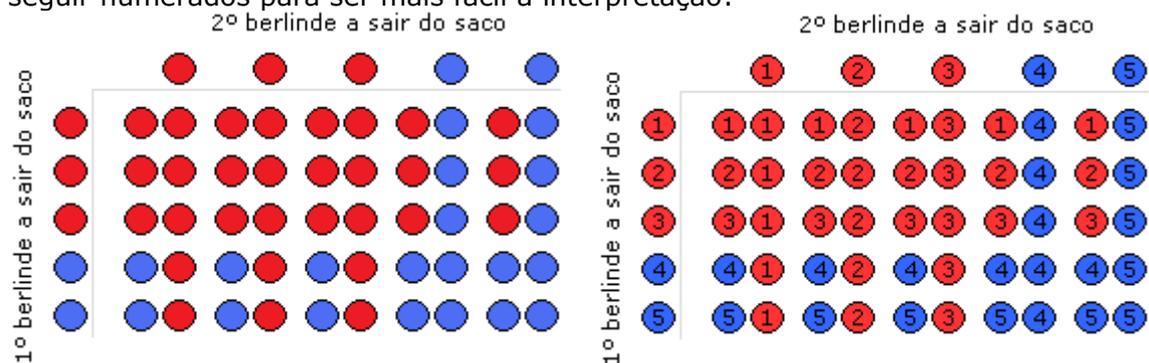
Acontecimento "Saiu o nome da Maria"	Resultados
Com reposição	Ana, Maria Maria, Ana Maria, Maria Maria, Filipa Filipa, Maria
Sem reposição	Ana, Maria Maria, Ana Maria, Filipa Filipa, Maria

**Exemplo – O saco de berlindes.** Considere a experiência aleatória que consiste em extrair 2 berlindes, de um saco com 3 berlindes vermelhos e 2 azuis .

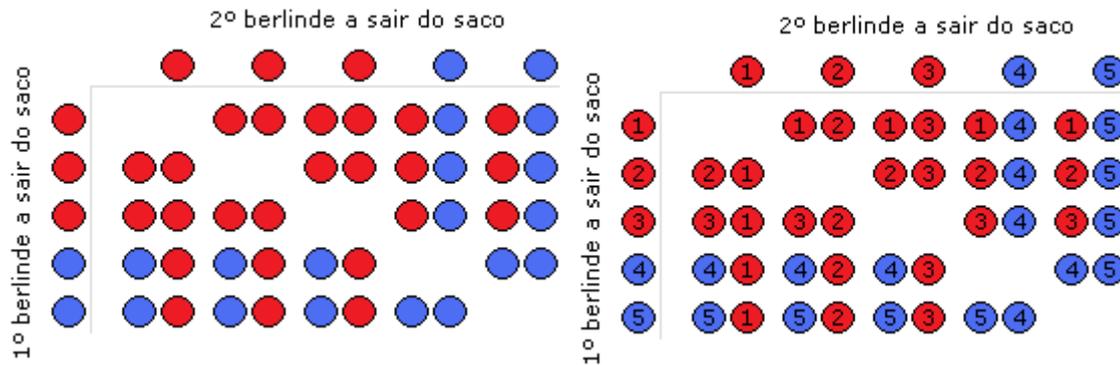
Que espaço de resultados se associa a esta experiência?

Para já é necessário saber se a extracção se faz com reposição ou sem reposição. Vamos considerar as duas situações. Para identificar o espaço de resultados será mais fácil numerar os berlindes, pelo que vamos numerar os berlindes vermelhos com 1, 2 e 3 e os azuis com 4 e 5.

**Com reposição** – Quando se retira um berlinde verifica-se a cor e torna-se a repor o berlinde no saco antes de extrair o próximo. um espaço de resultados pode ser constituído por todos os resultados, em número de 25, do esquema seguinte, em que se considera primeiro os berlindes como normalmente se apresentam, e a seguir numerados para ser mais fácil a interpretação:



**Sem reposição** – Neste caso o espaço de resultados é constituído por todos os resultados do espaço do esquema anterior, exceptuando os pares constituídos pelo mesmo berlinde:



O acontecimento "tirar 2 berlindes de cor diferente" é constituído pelos resultados



tanto no esquema com reposição, como sem reposição.

**Podemos considerar mais do que um espaço de resultados ao modelar um fenómeno aleatório?**

Sim!

A associação de um espaço de resultados a uma certa experiência aleatória nem sempre é única.

No caso do exemplo anterior, podemos assumir que o espaço de resultados associado à experiência que consiste em retirar 2 berlindes de um sacco com 3 berlindes vermelhos e 2 azuis é constituído pelos resultados elementares

$\{2 \text{ berlindes vermelhos, } 1 \text{ berlinde vermelho e } 1 \text{ berlinde azul, } 2 \text{ berlindes azuis}\}$

quer a extracção se faça com ou sem reposição. Este é aliás, o espaço de resultados mais intuitivo e que nos vem imediatamente ao pensamento quando idealizamos ou realizamos a experiência considerada. Normalmente é-nos indiferente qual o berlinde seleccionado em cada tiragem, porque estamos interessados unicamente na cor. No entanto, como veremos mais à frente, quando pretendermos associar probabilidades aos seus resultados, esta associação não é imediata como no espaço de resultados considerado inicialmente, já que os seus resultados não são todos igualmente possíveis.

Do mesmo modo, o espaço de resultados associado à experiência aleatória que consiste em lançar 2 moedas de 1 Euro e ver o resultado que sai, também pode ser interpretado como sendo

$\{(\text{Euro, Euro}), (\text{Euro, Nacional}), (\text{Nacional, Euro}), (\text{Nacional, Nacional})\}$

ou

$\{(2 \text{ faces Euro}), (2 \text{ faces Nacional}), (1 \text{ face Euro e } 1 \text{ face Nacional})\}$ .

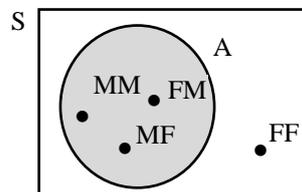
Também, do mesmo modo que no caso anterior, este último espaço de resultados não tem os resultados todos igualmente possíveis ao contrário do primeiro.

### 6.4.1 Utilização de diagramas de Venn para representar acontecimentos



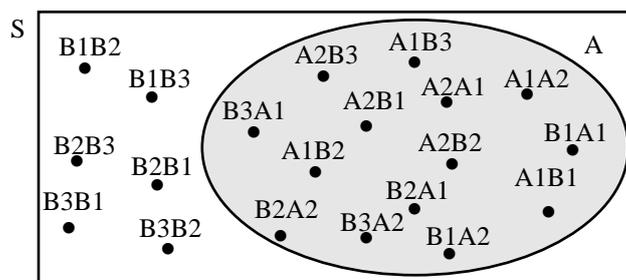
Uma técnica utilizada para visualizar o espaço de resultados e os acontecimentos associados a uma experiência aleatória, consiste em utilizar um rectângulo para representar o espaço de resultados e círculos para representar os acontecimentos. A essas representações chamamos diagramas de Venn.

**Exemplo – Famílias de 2 filhos.** Considere a experiência aleatória que consiste em verificar o sexo dos filhos das famílias de 2 filhos. O espaço de resultados é constituído pelos resultados  $S = \{MM, MF, FM, FF\}$ . Seja  $A$  o acontecimento “pelo menos um dos filhos é do sexo masculino”. Representando num diagrama de Venn, temos



**Exemplo – A caixa de disquetes.** Considere a experiência aleatória que consiste em retirar 2 disquetes, de uma caixa de 5 disquetes, em que 2 estão avariadas. Represente, através de um diagrama de Venn, o espaço de resultados e o acontecimento  $A = \{\text{pelo menos uma disquete está avariada}\}$ .

Representando as disquetes boas por  $B_1, B_2$  e  $B_3$  e as avariadas por  $A_1$  e  $A_2$ , temos



onde representamos, por exemplo, por  $B_1B_2$ , a saída das disquetes boas  $B_1$  e  $B_2$ .

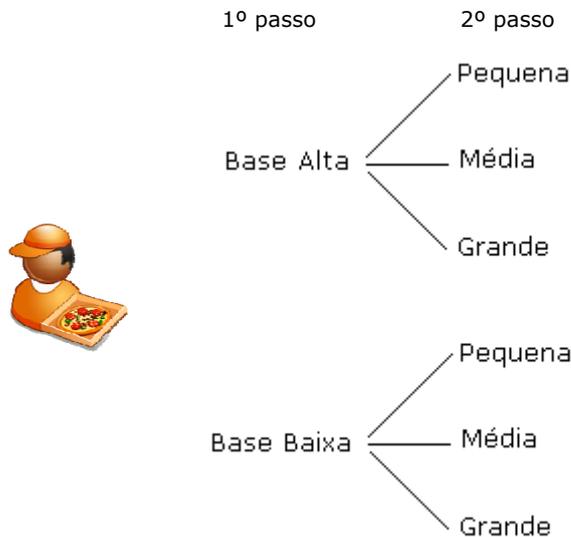
De um modo geral os diagramas de Venn não são construídos à escala, pelo que a área ocupada dentro do espaço de resultados com a figura utilizada para representar um acontecimento não é, por este motivo, necessariamente proporcional à probabilidade de esse acontecimento se realizar. No entanto, se a área ocupada pelo espaço de resultados fosse igual à unidade e os diagramas de Venn fossem construídos à escala, já as figuras utilizadas para representar os acontecimentos seriam construídas de forma a que as suas áreas fossem iguais às probabilidades dos acontecimentos, que representam, se realizarem.

### 6.4.2 Utilização de diagramas em árvore para representar acontecimentos



Um outro processo utilizado para visualizar acontecimentos é o diagrama em árvore. Este processo é especialmente adequado quando a experiência aleatória implica a ocorrência de uma sequência de passos. Por exemplo, admitamos que uma loja que vende piza, tem duas modalidades para a base da piza, nomeadamente Base Alta ou Base Baixa e tem 3 alternativas para o tamanho: Pequena, Média ou Grande. Uma pessoa que escolha uma destas pizzas, ao acaso, de quantas maneiras possíveis é que pode fazer a escolha, tendo em consideração o critério da base e do tamanho?

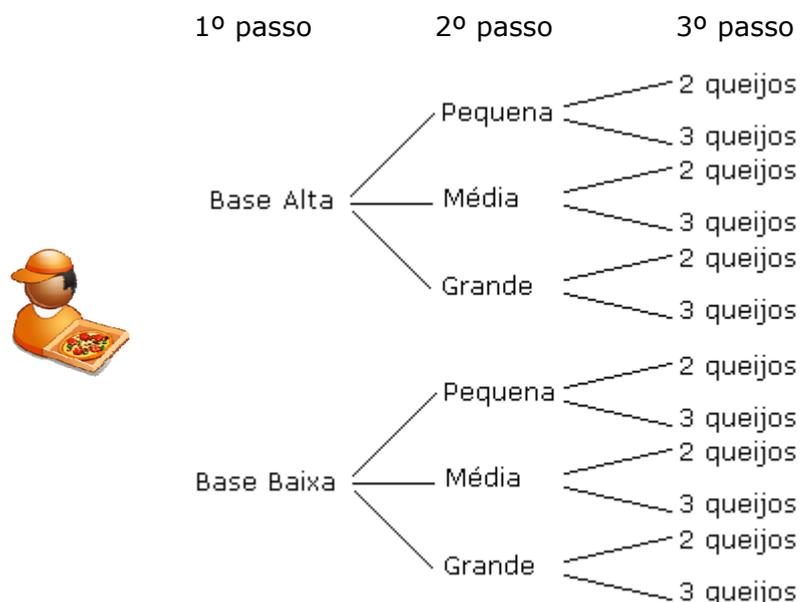
O primeiro passo será a escolha da base, seguindo-se o tamanho da piza:



O espaço de resultados associado a esta experiência aleatória, que consiste em seleccionar, ao acaso, uma piza, é constituído pelos seguintes resultados:

{(Base Alta, Pequena), (Base Alta, Média), (Base Alta, Grande), (Base Baixa, Pequena), (Base Baixa, Média), (Base Baixa, Grande)}

Suponhamos ainda que cada piza pode ter na cobertura 2 ou 3 tipos de queijo. Agora, de quantas maneiras diferentes se pode escolher a piza? Considerando agora mais um 3º passo, temos:



Assim, o espaço de resultados associado à experiência aleatória que consiste em seleccionar uma pizza ao acaso, tendo em consideração a base, o tamanho e a cobertura, é constituído pelos resultados:

{(Base Alta, Pequena, 2 queijos), (Base Alta, Pequena, 3 queijos), (Base Alta, Média, 2 queijos), (Base Alta, Média, 3 queijos), (Base Alta, Grande, 2 queijos), (Base Alta, Grande, 3 queijos), (Base Baixa, Pequena, 2 queijos), (Base Baixa, Pequena, 3 queijos), (Base Baixa, Média, 2 queijos), (Base Baixa, Média, 3 queijos), (Base Baixa, Grande, 2 queijos), (Base Baixa, Grande, 3 queijos)}

Associados ao espaço de resultados anterior, podemos considerar vários acontecimentos:

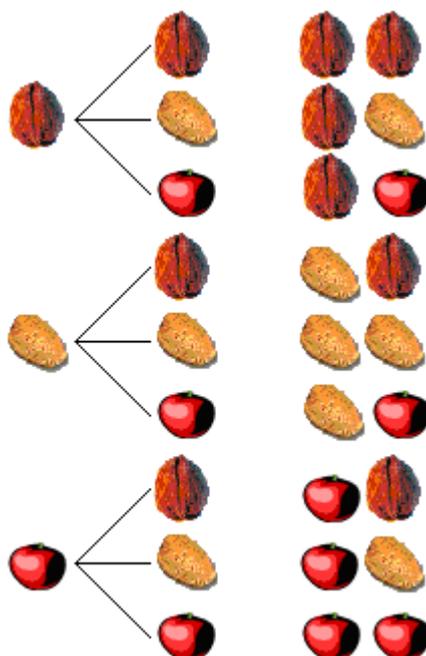
Acontecimento	Resultados para que o acontecimento se realize:
Seleccionar uma pizza ao acaso e ela ser Média e de 3 queijos	(Base Alta, Média, 3 queijos) (Base Baixa, Média, 3 queijos)
Seleccionar uma pizza ao acaso e ela ser Média	(Base Alta, Média, 2 queijos) (Base Alta, Média, 3 queijos) (Base Baixa, Média, 2 queijos) (Base Baixa, Média, 3 queijos)
Seleccionar uma pizza ao acaso e ela ter Base Alta e ser de 2 queijos	(Base Alta, Pequena, 2 queijos) (Base Alta, Média, 2 queijos) (Base Alta, Grande, 2 queijos)

Os resultados que compõem os acontecimentos anteriores obtêm-se muito facilmente a partir da árvore, percorrendo os ramos que satisfaçam as características desejadas. Por exemplo, ao pretender que a pizza seja Média, verificamos que existem dois ramos com a característica Média, e cada um destes ramos ainda se divide em dois raminhos. Assim, todos os raminhos que tenham o

nó na característica Média, serão resultados do acontecimento "Selecionar uma piza ao acaso e ela ser Média".

**Tarefa – A caixa de bombons.** O professor leva para a sala de aula uma grande caixa cheia de bombons de três tipos, em igual proporção. Cada bombom ou tinha uma amêndoa, ou uma noz, ou uma cereja. O professor pediu ao Pedro para retirar dois bombons ao acaso. De quantas (e quais) maneiras possíveis é que o Pedro pode tirar os dois bombons? O professor pediu ainda para descreverem o acontecimento "O Pedro retirou pelo menos um bombom com amêndoa", que vamos, para simplificar, representar pelo acontecimento A.

O que se pretende é o espaço de resultados associado à experiência aleatória que consiste em retirar 2 bombons e verificar o tipo de bombom. Utilizando ainda o diagrama em árvore temos:



$S = \{\text{noz noz, noz amêndoa, noz cereja, amêndoa noz, amêndoa amêndoa, amêndoa cereja, cereja noz, cereja amêndoa, cereja cereja}\}$

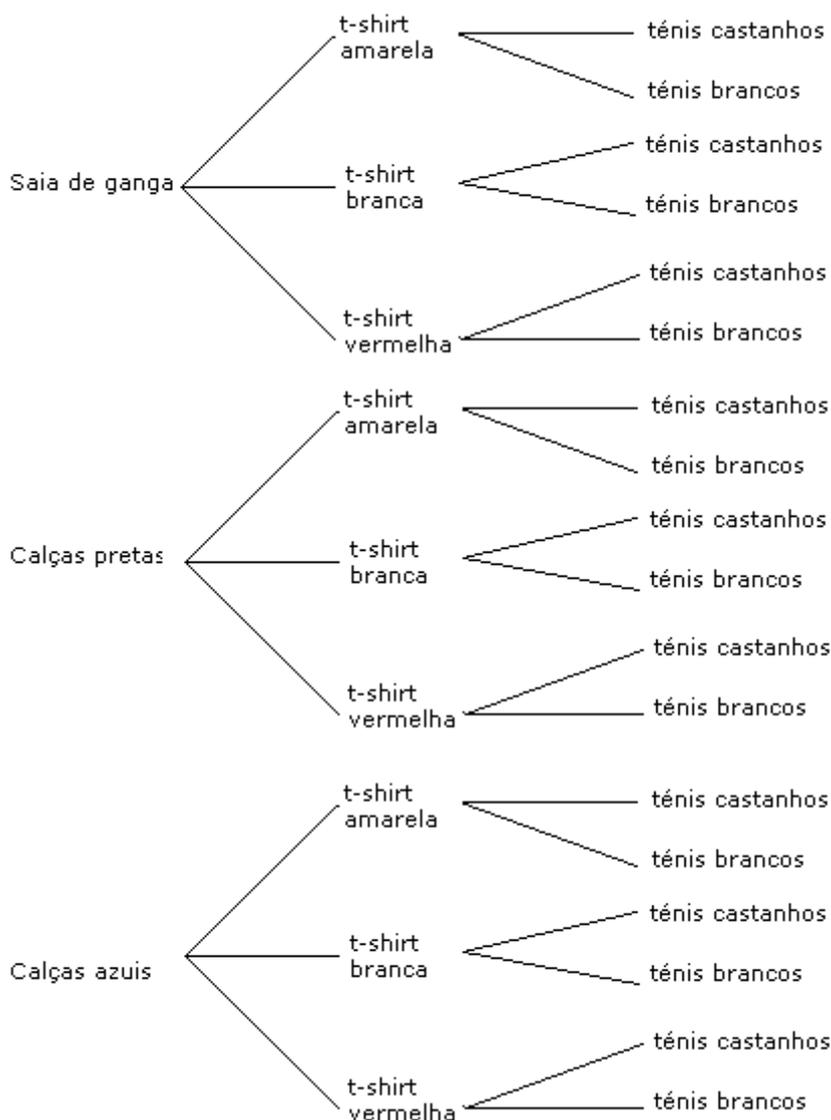
O acontecimento A é constituído pelos seguintes resultados:

$A = \{\text{noz amêndoa, amêndoa noz, amêndoa amêndoa, amêndoa cereja, cereja amêndoa}\}$

Os diagramas em árvore são uma boa opção para representar, de uma forma sistemática, todos os resultados possíveis quando estamos perante uma sequência de acontecimentos, como no caso anterior e na tarefa seguinte.

**Tarefa – Quantas "toiletas" pode a Mariana vestir?** Certa manhã a Mariana estava muito indecisa sobre o que devia vestir para ir para a escola. Tinha 1 saia de ganga e 2 pares de calças, um de ganga preta e outro de ganga azul, que ficavam muito bem com 3 t-shirts e que condiziam com dois pares de ténis, uns castanhos e outros brancos. Experimentou tantas toiletas, vestiu, despiu, vestiu..., que acabou

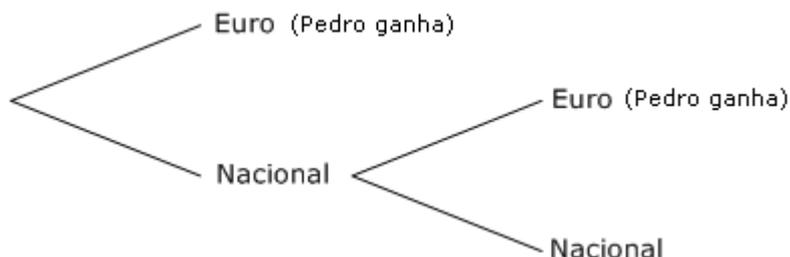
por chegar atrasada à escola! És capaz de dizer quantas toiletas diferentes a Mariana poderia ter experimentado?



Tendo em consideração o diagrama em árvore anterior, a Mariana se tivesse experimentado todas as combinações possíveis de peças de vestuário e calçado, teria experimentado 18 toiletas! Para obter cada uma das toiletas, basta seguir os ramos da árvore. De acordo com o diagrama anterior, a primeira e última toiletas são, respectivamente (Saia de ganga, t-shirt amarela, ténis castanhos) e (Calças azuis, t-shirt vermelha, ténis brancos).

**Tarefa – Quais os resultados que fazem com que o Pedro ganhe o jogo?** O professor propôs o seguinte jogo a ser jogado pelo Pedro e pelo Marco. Lança-se uma moeda de um Euro ao ar e se sair a face Euro, o Pedro ganha o jogo. Se sair a face Nacional, lança-se novamente a moeda ao ar e se sair a face Euro o Pedro ganha o jogo e se sair a face Nacional, ganha o Marco. Quais os resultados possíveis deste jogo e quais os resultados que fazem com que o Pedro ganhe o jogo?

O esquema seguinte traduz o jogo anterior: num primeiro lançamento da moeda, ou sai face Euro ou face Nacional. Se sair face Euro o jogo acaba e o Pedro ganha. Se sair face Nacional lança-se novamente a moeda e o Pedro ganha novamente se sair face Euro. Assim, o Pedro ganha se se verificar Euro ou (Nacional, Euro), enquanto que o Marco ganha se se verificar (Nacional, Nacional)

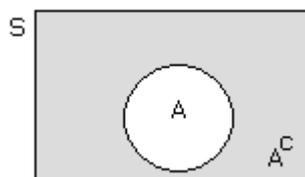


### 6.4.3 Operações com acontecimentos

O facto de interpretarmos acontecimentos como conjuntos, aliada à utilização dos diagramas de Venn para os visualizar, vai-nos permitir introduzir alguma terminologia utilizada quando falamos de acontecimentos. Assim, representando os acontecimentos associados a um determinado espaço de resultados  $S$ , por  $A$ ,  $B$ ,  $C$ , ..., temos:

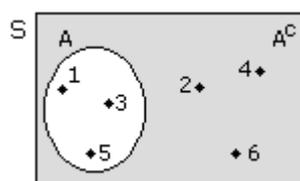
#### ❖ Acontecimento complementar

Acontecimento **complementar** do acontecimento  $A$ , é o acontecimento constituído por todos os resultados do espaço de resultados  $S$ , que não estão em  $A$ . Este acontecimento representa-se por  $\bar{A}$  ou  $A^c$



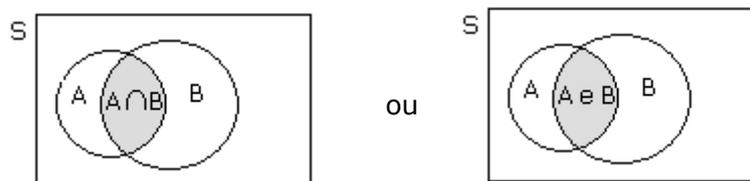
Quando um acontecimento se realiza, o seu complementar não se pode realizar.

**Exemplo – Lançamento do dado.** Considere a experiência que consiste em lançar um dado com 6 faces e em verificar qual o número de pintas da face que fica virada para cima. O espaço de resultados associado é  $S = \{1, 2, 3, 4, 5, 6\}$ . Se representar por  $A$  o acontecimento "saída de uma face com um número ímpar de pintas", temos  $A = \{1, 3, 5\}$ . O acontecimento complementar de  $A$  é o acontecimento "saída de uma face com um número par de pintas", que representamos por  $\bar{A} = \{2, 4, 6\}$



### ❖ Acontecimento intersecção

Acontecimento **intersecção** dos acontecimentos A e B, é o acontecimento constituído pelos resultados que pertencem simultaneamente a A e a B. Este acontecimento representa-se por  $A \cap B$  ou  $(A \text{ e } B)$



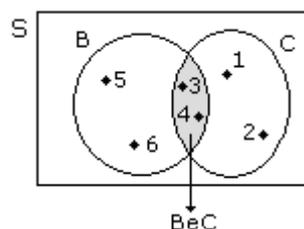
Para que o acontecimento intersecção de dois acontecimentos se realize, é necessário que os dois acontecimentos se realizem simultaneamente.

**Exemplo – Lançamento do dado** (cont.). Sejam B e C os seguintes acontecimentos:

Acontecimento B → "O número de pintas é maior ou igual a 3" ou seja  
 $B = \{3, 4, 5, 6\}$

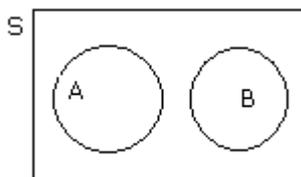
Acontecimento C → "O número de pintas é menor ou igual a 4" ou seja  
 $C = \{1, 2, 3, 4\}$

Acontecimento **intersecção**  $B \text{ e } C = \{3, 4\}$



### ❖ Acontecimentos disjuntos ou mutuamente exclusivos

Acontecimentos **disjuntos** ou **mutuamente exclusivos** são acontecimentos que não têm resultados comuns



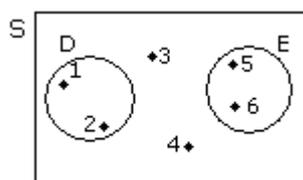
Quando dois acontecimentos são mutuamente exclusivos, a realização de um deles implica que o outro não se realize.

**Exemplo – Lançamento do dado** (cont.). Sejam D e E os seguintes acontecimentos

Acontecimento D → "O número de pintas é menor ou igual a 2" ou seja  $D = \{1, 2\}$

Acontecimento E → "O número de pintas é maior que 4" ou seja  $E = \{5, 6\}$

Os acontecimentos D e E são **mutuamente exclusivos**



#### ❖ Acontecimento impossível

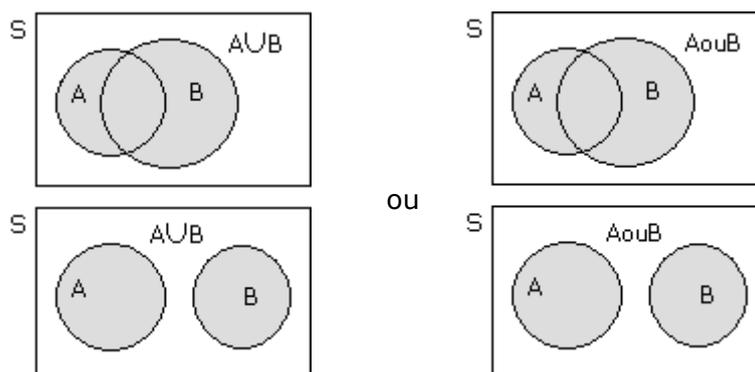
Acontecimento **impossível** é o acontecimento que resulta da intersecção de acontecimentos disjuntos ou mutuamente exclusivos, ou seja, é o acontecimento que não tem qualquer resultado do espaço de resultados. Representa-se pelo símbolo  $\emptyset$  (um zero cortado por um traço).

**Exemplo – Lançamento do dado** (cont.). Sejam D e E os acontecimentos considerados anteriormente. Então

$$D \cap E = \{\} = \emptyset$$

#### ❖ Acontecimento união

Acontecimento **união** dos acontecimentos A e B é o acontecimento constituído por todos os resultados de A ou de B. Representa-se por  $A \cup B$  ou  $(A \text{ ou } B)$



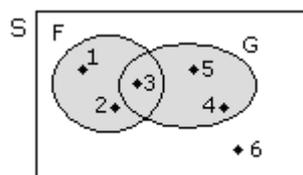
Para que a união de dois acontecimentos se realiza, basta que um dos acontecimentos se realize.

**Exemplo – Lançamento do dado** (cont.). Sejam F e G os seguintes acontecimentos

Acontecimento F → "O número de pintas é menor ou igual a 3" ou seja  $F = \{1, 2, 3\}$

Acontecimento G → "O número de pintas é maior que 2 e menor que 6" ou seja  $G = \{3, 4, 5\}$

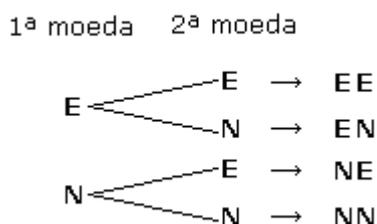
Acontecimento **união**  $F \cup G = \{1, 2, 3, 4, 5\}$



**Tarefa – Lançamento de duas moedas de 1 euro.** O professor propõe aos alunos realizar a experiência aleatória que consiste em lançar 2 moedas de 1 euro e verificar as faces que ficam viradas para cima.

- Quais os resultados que constituem o espaço de resultados?
- Representando por A o acontecimento "A face Euro verifica-se 1 vez" e por B o acontecimento "A face Nacional verifica-se pelo menos 1 vez". Quais os resultados que compõem os acontecimentos  $\bar{A}$ ,  $A \cup B$  e  $A \cap B$ ?

Resolução: Os resultados da experiência aleatória que consiste em lançar 2 moedas, e verificar as faces que ficam voltadas para cima, podem ser representados num diagrama em árvore, como se apresenta a seguir:



No esquema anterior representámos por E a face Euro e por N a face Nacional.

- Então o espaço de resultados é

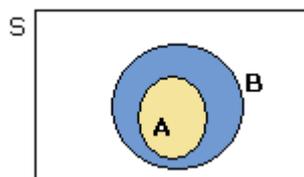
$$S = \{EE, EN, NE, NN\}$$

O espaço de resultados também poderia ser obtido a partir de uma tabela de dupla entrada:

2ª moeda	E	N
1ª moeda		
E	EE	EN
N	NE	NN

- $A = \{EN, NE\}$   
 $B = \{EN, NE, NN\}$   
 $\bar{A} = \{EE, NN\}$   
 $A \cup B = \{EN, NE, NN\}$   
 $A \cap B = \{EN, NE\}$

Observe-se que os acontecimentos A e B têm a particularidade de A estar contido em B, pois todos os resultados de A são resultados de B. Assim, como se visualiza no seguinte diagrama de Venn, quando dois acontecimentos A e B são tais que um está contido no outro, por exemplo A contido em B, a união destes dois acontecimentos é o acontecimento B, enquanto que a intersecção é o acontecimento A:

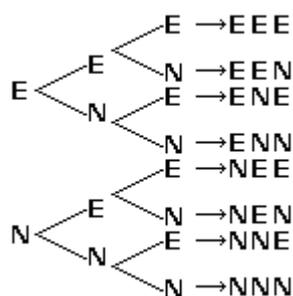


### Tarefa – Lançamento de três moedas de 1 euro

1ª versão – Considere-se a experiência aleatória que consiste em lançar 3 moedas de 1 euro e verificar as faces que ficam viradas para cima.

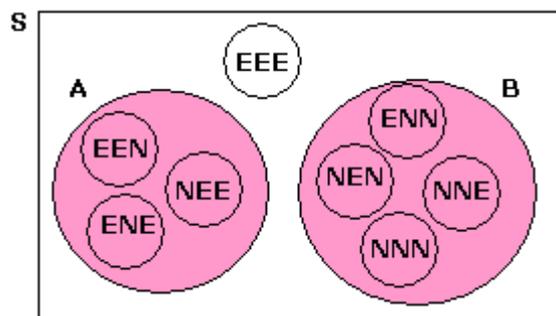
- Quais os resultados que constituem o espaço de resultados?
- Considere os seguintes acontecimentos:
  - A “a face Euro aparece 2 vezes”
  - B “a face Nacional aparece pelo menos 2 vezes”
- Quais os resultados que compõem os acontecimentos  $A \cup B$  e  $A \cap B$ ?

Considerando ainda um diagrama em árvore para representar os resultados da experiência aleatória que consiste em lançar 3 moedas e verificar as sequências das faces que ficam voltadas para cima, temos



- Do esquema anterior verifica-se que  
 $S = \{EEE, EEN, ENE, ENN, NEE, NEN, NNE, NNN\}$
- $A = \{EEN, ENE, NEE\}$   
 $B = \{ENN, NEN, NNE, NNN\}$
- $A \cup B = \{EEN, ENE, NEE, ENN, NEN, NNE, NNN\}$  e  $A \cap B = \{\} = \emptyset$

Numa representação em diagrama de Venn, temos



Repare-se que o acontecimento união de A e B é o acontecimento que se realiza sempre que não saírem 3 faces Euro no lançamento das 3 moedas.

2ª versão - Consideremos ainda a experiência aleatória que consiste em lançar as 3 moedas de 1 euro, mas agora pretende-se observar quantas vezes sai a face Euro. Qual o espaço de resultados associado? Quais os resultados que compõem A e B?

Neste caso o que se pretende registar é o nº de vezes que sai a face Euro e não as sequências de faces. Assim

$$S = \{0, 1, 2, 3\}$$

A partir deste novo espaço de resultados, os acontecimentos A e B são:

$A = \{2\}$  e  $B = \{0,1\}$  (Para que a face Nacional apareça pelo menos 2 vezes, a face Euro não pode aparecer nenhuma vez ou só pode aparecer 1 vez.

## 6.5 Modelo de Probabilidade para um fenómeno aleatório. Probabilidade de um acontecimento

No que se segue vamos admitir que os fenómenos aleatórios que se vão estudar são fenómenos cujos espaços de resultados são finitos. Assim, definir um *modelo de probabilidade*, para um fenómeno aleatório, implica:

- A identificação de um espaço de resultados;
- Uma forma de atribuir um número a cada um dos resultados, isto é, a cada acontecimento elementar, a que chamaremos *probabilidade*.

O processo de atribuir probabilidades deve ser tal, que algumas regras básicas devem ser satisfeitas para todos os modelos:

- Regra 1 – Uma probabilidade deve ser um número não negativo;
- Regra 2 – A soma das probabilidades dos acontecimentos elementares que compõem o espaço de resultados é igual a 1.

Observe-se que a definição anterior não exclui a possibilidade de um acontecimento elementar ter probabilidade zero. No entanto, em espaços finitos uma probabilidade igual a zero é interpretada, na prática, como uma *impossibilidade*, pelo que qualquer resultado do espaço de resultados, com probabilidade nula, pode ser eliminado do espaço de resultados (Feller, 1968, p. 22).

### 6.5.1 Probabilidade de um acontecimento

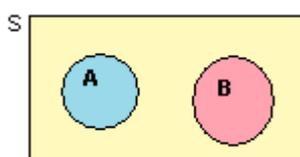
Admitamos, para já, que tenhamos um modelo de probabilidade associado a um fenómeno aleatório (veremos a seguir alguns processos para atribuir probabilidades aos resultados de um espaço de resultados). Uma vez definido esse modelo, como obter a probabilidade de um acontecimento?

**Probabilidade de um acontecimento** – A *probabilidade de um acontecimento A* representa-se por  $P(A)$  e define-se como sendo a soma das probabilidades dos acontecimentos elementares que compõem A.

### Regras para a Probabilidade

A interpretação que fizemos de Probabilidade de um acontecimento, no início deste capítulo, como a proporção de vezes que o acontecimento se verifica, quando repetimos a experiência um grande número de vezes, sugere-nos um conjunto de regras a que deve obedecer qualquer forma de atribuir probabilidades a acontecimentos de um mesmo espaço de resultados S:

- Regra 1 – Qualquer que seja o acontecimento A, tem-se que  $P(A) \geq 0$ ;
- Regra 2 – A probabilidade do espaço de resultados, S, é igual a 1,  $P(S)=1$ ;
- Regra 3 – Dados os acontecimentos A e B, disjuntos, então a probabilidade de A ou B se realizarem,  $P(A \cup B)$ , ou  $P(A \text{ ou } B)$ , é igual à soma das probabilidades de A e de B se realizarem,

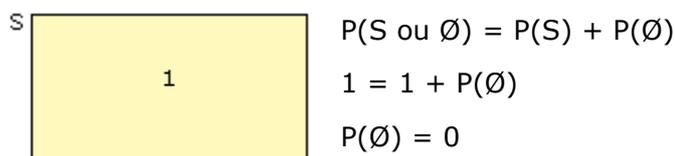


$$P(A \text{ ou } B) = P(A) + P(B)$$

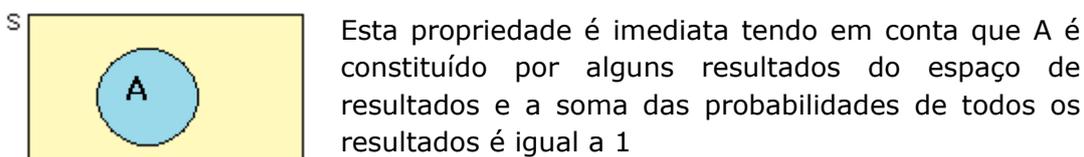
### Propriedades da Probabilidade

Tendo em conta as regras anteriores, e com a ajuda dos diagramas de Venn, podem-se deduzir as seguintes propriedades para a probabilidade de acontecimentos de um mesmo espaço de resultados  $S$ :

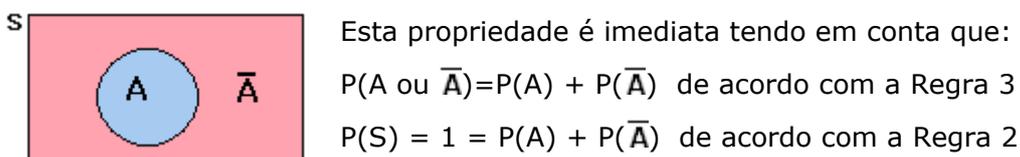
*Propriedade 1* – A probabilidade do acontecimento impossível é igual a zero,  $P(\emptyset)=0$



*Propriedade 2* – A probabilidade  $P(A)$  de qualquer acontecimento  $A$  é tal que  $0 \leq P(A) \leq 1$

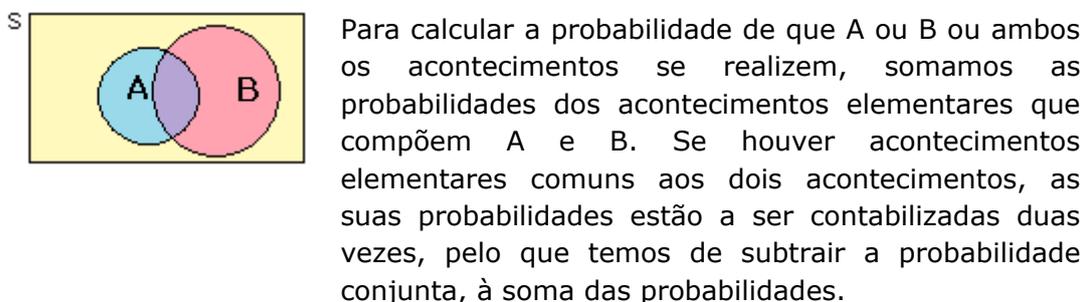


*Propriedade 3* – A probabilidade do acontecimento complementar  $\bar{A}$  do acontecimento  $A$ ,  $P(\bar{A})$  é igual a  $P(\bar{A}) = 1 - P(A)$



*Propriedade 4* – Dados dois acontecimentos  $A$  e  $B$ , a probabilidade de  $A$  ou  $B$  ou ambos se realizarem,  $P(A \cup B)$ , é igual à soma das probabilidades de  $A$  e de  $B$  se realizarem, menos a probabilidade de  $A$  e  $B$  se realizarem conjuntamente:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Nota – Quando perguntamos a alguém se quer fruta ou doce para a sobremesa, será que a pessoa tem de escolher uma de entre as duas alternativas, ou poderá escolher as duas? Esta ambiguidade não existe se perguntarmos “Amanhã vais para a escola de ténis ou de sandálias?” Normalmente quando utilizamos o termo “ou”

ele não tem um significado disjuntivo ou exclusivo, pelo que quando se pede a probabilidade de os acontecimentos A ou B ocorrerem, significa que pode ocorrer qualquer um dos dois ou ambos os acontecimentos. Se pretendermos obter a probabilidade de A ou B ocorrerem, mas sem que ambos possam ocorrer, então, como facilmente se verifica construindo o diagrama de Venn

$$P(A \cup B) = P(A) + P(B) - 2 \times P(A \cap B)$$

### 6.5.2 Processos de construção de modelos de probabilidade ou como atribuir probabilidades aos acontecimentos elementares

A probabilidade começou por ser estudada por matemáticos franceses que desenvolveram modelos matemáticos associados aos chamados jogos de azar. Neste caso, é quase sempre possível encontrar um espaço de resultados para cujos elementos, à partida, não se tem razão para admitir que não tenham igual probabilidade de ocorrer. É o que acontece com a moeda ou o dado, que admitimos serem equilibrados e portanto cada face tem igual possibilidade de sair, ou com o baralho de cartas, em que admitimos que cada uma das cartas tem a mesma possibilidade de ser extraída. Esta situação, embora bastante restritiva, é muito simples de ser tratada, como se descreve a seguir.

#### 6.5.2.1 Situação de simetria – Regra de Laplace

Admitamos então que estamos numa situação de simetria, em que damos igual possibilidade à realização de cada resultado de um espaço de resultados. Por exemplo ao lançar dois dados equilibrados, damos igual possibilidade a cada um dos 36 resultados possíveis constituídos pelos pares ordenados (em que os números indicam o nº de pintas de cada um dos dois dados):

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)  
 (2,1) (2,2) (2,3) (2,4) (2,5) (2,6)  
 (3,1) (3,2) (3,3) (3,4) (3,5) (3,6)  
 (4,1) (4,2) (4,3) (4,4) (4,5) (4,6)  
 (5,1) (5,2) (5,3) (5,4) (5,5) (5,6)  
 (6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

Então, se os 36 resultados são todos igualmente possíveis, tendo em conta as Regras 1 e 2 a satisfazer por qualquer Regras 1, 2 e 3, enunciadas para a Probabilidade, a cada um dos resultados atribui-se probabilidade  $1/36$ .

De um modo geral, se um espaço de resultados tem  $n$  resultados e todos são igualmente possíveis, então a cada acontecimento elementar atribuímos a probabilidade  $1/n$ .

O resultado anterior é imediato, tendo em consideração as regras que enunciámos para a Probabilidade. Efectivamente, se temos  $n$  resultados igualmente possíveis e a soma das probabilidades de todos os resultados, considerados como acontecimentos elementares, tem de ser igual a 1 (Regra 2 e Regra 3) e além disso a probabilidade de qualquer acontecimento elementar é maior que 0 (Regra 1 e o facto de em espaços finitos uma probabilidade igual a zero ser interpretada, na

prática, como uma *impossibilidade*, pelo que qualquer resultado do espaço de resultados com probabilidade nula pode ser eliminado do espaço de resultados (Feller, 1968, p. 22)), vem imediatamente que cada acontecimento elementar tem probabilidade  $1/n$ .

No que diz respeito à probabilidade de qualquer acontecimento  $A$ , associado a um espaço de resultados, em que os resultados são *igualmente possíveis*, a regra 3, permite-nos atribuir-lhe uma probabilidade da seguinte forma:

- Se o acontecimento  $A$  for constituído por  $m$  resultados, ou seja  $m$  acontecimentos elementares, e o espaço de resultados for constituído por  $n$  resultados, a probabilidade de  $A$  é igual à soma de  $m$  parcelas iguais a  $1/n$ , ou seja, é igual a  $m/n$ .

Se um espaço de resultados,  $S$ , tem  $n$  resultados e todos são igualmente possíveis, então a probabilidade de qualquer acontecimento  $A$ , é igual ao quociente entre o número de resultados de  $A$  e o número de resultados de  $S$ . Mais geralmente, temos o seguinte resultado conhecido como **Regra ou Lei de Laplace**

Define-se probabilidade do acontecimento  $A$  associado a um espaço de resultados  $S$ , com  $n$  resultados igualmente possíveis, como sendo a razão entre o número  $m$  de resultados **favoráveis** a  $A$  (resultados que compõem  $A$ ), e o número  $n$  de resultados **possíveis** (resultados que constituem  $S$ ):

$$P(A) = \frac{m}{n}$$

Este processo de atribuir probabilidades aos acontecimentos exige uma enumeração correcta do espaço de resultados e a conseqüente enumeração de quais os resultados elementares que compõem os acontecimentos para os quais pretendemos atribuir probabilidades.

Só aparentemente é que se tem uma tarefa simples! Senão vejamos: admitamos que se pretende calcular a probabilidade de no lançamento de duas moedas de um euro, se obter igual número de faces Euro (E) e faces Nacional (N). Não há dúvida de que o espaço de resultados é constituído pelos resultados  $S=\{EN, EE, NE, NN\}$ , todos igualmente possíveis, e sendo o acontecimento  $A$ , saída de 1 face Euro e 1 face nacional  $A=\{EN, NE\}$ , temos que  $P(A)=2/4$ , ou seja  $P(A)=50\%$ .

Suponhamos agora que lançamos 4 vezes a moeda e pretendemos obter a probabilidade de igual número de faces Euro e Nacional. Agora temos  $S= \{EEEE, EEEN, EENE, EENN, ENEE, ENEN, ENNE, ENNN, NEEE, NEEN, NENE, NENN, NNEE, NNEN, NNNE, NNNN\}$  e  $A=\{EENN, ENEN, ENNE, NEEN, NENE, NNEE\}$  e  $P(A)=6/16$ , ou seja  $P(A)=37,5\%$ .

E se lançarmos a moeda 6 vezes? O espaço de resultados é constituído por 64 resultados, todos igualmente possíveis, dos quais 20 constituem o acontecimento "Igual número de faces Euro e faces Nacional". Continuando a representar este acontecimento por  $A$ , vem  $P(A)=31,25\%$ .

Nota: Para concluir que o espaço de resultados anteriormente referido, é constituído por 64 resultados possíveis, basta pensar da seguinte forma:

Temos 6 espaços para preencher com as letras E ou N:

— — — — —

No primeiro espaço podemos colocar o E ou o N, pelo que temos 2 possibilidades:

E \_ \_ \_ \_ \_  
ou  
N \_ \_ \_ \_ \_

No espaço seguinte temos outras 2 possibilidades, que combinadas com as anteriores dão  $2 \times 2 = 2^2$  possibilidades:

E E \_ \_ \_ \_ \_  
ou  
E N \_ \_ \_ \_ \_  
ou  
N E \_ \_ \_ \_ \_  
ou  
N N \_ \_ \_ \_ \_

Repetindo o processo até à 6ª posição, o número de resultados diferentes é  $2^6$ . Para saber quantos destes resultados têm 3 faces Euro, podemos raciocinar da seguinte forma:

3 faces Euro seguidas Número de resultados

<u>E</u> <u>E</u> <u>E</u> _ _ _	4
_ <u>E</u> <u>E</u> <u>E</u> _ _	
_ _ <u>E</u> <u>E</u> <u>E</u> _	
_ _ _ <u>E</u> <u>E</u> <u>E</u>	

2 faces Euro seguidas

<u>E</u> <u>E</u> _ <u>E</u> _ _	12
<u>E</u> <u>E</u> _ _ <u>E</u> _	
<u>E</u> <u>E</u> _ _ _ <u>E</u>	
_ <u>E</u> <u>E</u> _ <u>E</u> _	
_ <u>E</u> <u>E</u> _ _ <u>E</u>	
_ _ <u>E</u> <u>E</u> _ <u>E</u>	
<u>E</u> _ <u>E</u> <u>E</u> _ _	
<u>E</u> _ _ <u>E</u> <u>E</u> _	
_ <u>E</u> _ <u>E</u> <u>E</u> _	
_ _ <u>E</u> _ <u>E</u> <u>E</u>	

Sem faces Euro seguidas

<u>E</u> _ <u>E</u> _ <u>E</u> _	4
<u>E</u> _ _ <u>E</u> _ <u>E</u>	
_ <u>E</u> _ <u>E</u> _ <u>E</u>	
<u>E</u> _ <u>E</u> _ _ <u>E</u>	

Assim, o número de resultados possíveis com 3 faces Euro e 3 faces Nacional é igual a 20.

E se o número de vezes que lançamos a moeda for 10? Generalizando o processo utilizado anteriormente para obter o número de resultados possíveis chegamos a 1024 ( $=2^{10}$ ) resultados, todos igualmente possíveis. Podemos adiantar que destes

resultados, 252 têm igual número de faces Euro e faces Nacional, mas a forma de chegar a este valor não é simples. Então a probabilidade de se realizar o acontecimento "Igual número de faces Euro e faces Nacional", representado por  $A$ , será  $P(A) = 252/1024$  ou seja  $P(A) \approx 25\%$ .

Como facilmente se depreende, à medida que o número de lançamentos da moeda aumenta, mais complicada é a tarefa de construir o espaço de resultados associado. A título de curiosidade, acrescentamos que se lançarmos a moeda 100 vezes, o número de resultados do espaço de resultados é igual a 1 267 650 600 228 229 401 496 703 205 376 e que a probabilidade de obter igual número de faces Euro e faces Nacional é aproximadamente igual a 8%. Existem instrumentos matemáticos de que a análise combinatória é um exemplo, que nos facilitam estes processos de contagem, mas que saem fora do âmbito deste curso.

*O mito dos 50%!*

Outra situação que merece reflexão é o mito dos 50%. Com frequência ao resultado de uma situação aleatória é erradamente atribuída a probabilidade de 50%, quando temos dois resultados em jogo. Por exemplo, na confecção de um artigo produzido por uma máquina, este pode sair defeituoso ou não defeituoso. Embora só consideremos estas duas situações, como os resultados possíveis da análise do artigo, não tem sentido considerar que estes acontecimentos são igualmente prováveis. Do mesmo modo não tem qualquer sentido admitir que a probabilidade de ganhar o Euromilhões seja de 50%, embora os resultados possíveis para um jogador sejam ganhar ou não ganhar. Quem sabe se muita gente não joga, a pensar que tem 50% de probabilidade de ganhar!

### 6.5.2.2 Probabilidade experimental ou frequencista

Quando não é possível utilizar o argumento da simetria, admitindo que os resultados do espaço de resultados são igualmente possíveis, então recorre-se à definição de fenómeno aleatório e como já anteriormente frisámos, temos em consideração a regularidade que se observa na sua repetição, para quantificarmos a percentagem de vezes que um acontecimento se realiza, como sendo a sua probabilidade.

Apresentamos a seguir, formalmente, a definição de probabilidade experimental ou frequencista, resultante de todo o desenvolvimento feito nas secções iniciais deste capítulo.

Define-se **probabilidade** (experimental ou frequencista) de um acontecimento  $A$  e representa-se por  $P(A)$  como sendo o valor à volta do qual tende a estabilizar a frequência relativa da realização de  $A$ , num grande número de repetições da experiência aleatória.

Mais uma vez chamamos a atenção para que esta regularidade tem que ser uma regularidade a *longo termo*. Esta regularidade não tem que existir, a não ser ao fim de um número muito grande de repetições do fenómeno aleatório.

Nem, tão pouco, existe a lei das compensações! Se, por exemplo, no lançamento de uma moeda de um euro, que admitimos ser equilibrada, obtivermos a seguinte sequência de resultados

Euro, Nacional, Nacional, Euro, Nacional

não podemos esperar que no próximo lançamento saia a face Euro, para tentar compensar com mais uma face Euro, as três faces Nacional.

Do mesmo modo, se obtivermos em seis lançamentos de uma moeda a sequência

Euro, Euro, Euro, Euro, Euro, Euro

será que é mais provável que no próximo lançamento se verifique a face Nacional? De modo nenhum, pois a moeda “não tem memória” e não é pelo facto de nos lançamentos anteriores ter saído a face Euro, que faz com que no próximo lançamento a face Nacional tenha maior possibilidade de sair. Os sucessivos lançamentos são independentes.

Outra situação, não intuitiva, resultante de uma interpretação abusiva da regularidade a longo termo é considerar as seguintes proposições como equivalentes (Moore, 1997b, p. 421):

- “Em muitos lançamentos de uma moeda equilibrada, a proporção de faces Euro é aproximadamente 50%”
- “Em muitos lançamentos de uma moeda equilibrada, o número de faces Euro é aproximadamente metade do número de lançamentos”

Na realidade, enquanto que a primeira asserção é verdadeira, o mesmo não se passa com a segunda.

A *regularidade a longo termo* significa que a *frequência relativa* da saída de face Euro tende a estabilizar à volta de 50%. Por exemplo, se lançarmos uma moeda 3000 vezes, em que se registou o número de faces Euro após 100, 500, 1000, 2000 e 3000 lançamentos, um resultado possível poderia ter sido o seguinte:

Nº lançamentos	Nº faces Euro obtidas x	Metade dos lançamentos y	y - x	Freq. Relativa da face Euro
100	49	50	1	0,49
500	253	250	3	0,51
1000	495	500	5	0,50
2000	993	1000	7	0,50
3000	1510	1500	10	0,50

Como se verifica, pode acontecer que o número de faces Euro obtidas, se afaste de metade do número de lançamentos, não impedindo que a frequência relativa tenha tendência a estabilizar à volta do valor 50%.

**Tarefa – Qual o jogo preferido?** Dão-lhe a escolher entre fazer 10 ou 100 jogadas com uma moeda de 1 euro, equilibrada, para ganhar uma viagem, nas seguintes situações:

- Ganha a viagem se a proporção de vezes que aparece a face Euro estiver entre 40% e 60%. Escolhe fazer 10 ou 100 jogadas?
- Ganha a viagem se o número de vezes que aparece a face Euro for igual a metade dos lançamentos. Escolhe fazer 10 ou 100 jogadas?

De acordo com o que foi dito anteriormente, a proporção ou frequência relativa da saída da face Euro tende a estabilizar à volta de 50% (moeda equilibrada) à medida

que o número de lançamentos da moeda aumenta, pelo que na primeira situação se deve escolher fazer 100 jogadas. Também de acordo com o que dissemos anteriormente, não é verdade que em muitos lançamentos da moeda equilibrada, o número de vezes que sai a face Euro seja igual a metade do número de lançamentos, pelo que na segunda situação escolheríamos só fazer 10 jogadas. Aliás, os resultados do exemplo apresentado na secção anterior permitem-nos inferir que à medida que se aumenta o número de lançamentos, diminua a probabilidade de obter igual número de faces Euro e faces Nacional.

### 6.5.2.3 Probabilidade subjectiva

A maior parte das vezes não se pode repetir a experiência as vezes que se quer, nem tão pouco assumir que os resultados da experiência são igualmente possíveis. Por exemplo, qual a probabilidade de um aluno obter uma nota superior a 14 na disciplina de Bioestatística, onde se encontra matriculado no 1º semestre, no curso de Biologia de determinada faculdade? Nem é desejável que a experiência se repita, nem devemos atribuir igual possibilidade aos acontecimentos nota superior a 14 e nota menor ou igual que 14. No entanto, se formos ver o currículo do aluno poderemos atribuir uma probabilidade elevada (ou baixa) ao acontecimento em causa. A probabilidade diz-se, neste caso, subjectiva, pois foi baseada em informação anterior e num julgamento subjectivo.

Por exemplo, se lhe perguntarem qual a probabilidade do próximo presidente da República ser homem, o que responde? Não é natural esperar que exista igual probabilidade de ser homem ou mulher e com a informação que se tem do passado, é natural atribuir a esse acontecimento uma probabilidade de 100%.

Uma vez que existe algo de arbitrário na atribuição de probabilidades a acontecimentos seguindo esta teoria, é de difícil aplicação, embora recentemente esteja a ter grande sucesso.

**Probabilidade subjectiva** – atribui-se a um acontecimento uma probabilidade com base na experiência e informação anteriores.

### 6.5.3 Exemplos de modelos de probabilidade

Nesta secção apresentaremos alguns modelos de probabilidade e trabalharemos algumas das propriedades da probabilidade, formalizadas na secção 6.5.1.

**Exemplo – Lançamento da moeda de um euro** - Consideremos a experiência aleatória que consiste em verificar qual a face que fica voltada para cima, quando se lança uma moeda de um euro. Um modelo de probabilidade que normalmente se mostra adequado para descrever o fenómeno associado é o seguinte

Resultados	Face Euro E	Face Nacional N
Probabilidade	1/2	1/2

Na tabela anterior apresentam-se os resultados possíveis e as probabilidades desses resultados ocorrerem. Como dissemos no início da secção 6.4, pode eventualmente acontecer que, ao lançar a moeda, ela fique em pé! No entanto este resultado é tão raro, que lhe atribuímos a probabilidade 0, não incluindo sequer o resultado "Ficar em pé", como um resultado possível. Ao atribuir igual probabilidade à saída da face Euro e da face Nacional, estamos a admitir que a moeda foi

construída de tal modo que se a lançarmos muitas vezes, aproximadamente 50% das vezes sairá a face Euro e as outras 50% das vezes a face Nacional. Se tivéssemos alguma razão para admitir que a moeda não era equilibrada, já o modelo anterior não servia.

**Exemplo – Lançamento do dado** – Consideremos a experiência aleatória que consiste em lançar um dado e verificar qual a face que fica voltada para cima. Também neste caso, um modelo que normalmente se mostra adequado para descrever o fenómeno aleatório associado é o seguinte

Face						
Probabilidade	1/6	1/6	1/6	1/6	1/6	1/6

Mais uma vez, ao considerar o modelo anterior, estamos a admitir que o dado foi construído de tal modo que qualquer face terá igual probabilidade de sair, o que nos leva a questionar se o modelo se adequa! Na verdade, admitindo que o cubo que vai dar origem ao dado tinha sido construído nas devidas condições de modo que cada face tinha igual probabilidade de sair, quando se desenham as pintas, não estaremos a alterar a estrutura do cubo? Naturalmente que uma face com 6 pintas será um pouco mais pesada que uma com 1 pinta devido ao peso da tinta! E já agora, também não terá influência a forma como se agarra o dado, assim como a forma como se lança? O facto é que se pretendermos arranjar um modelo ideal que entre em consideração com todos estes factores, não vamos conseguir arranjar modelo nenhum. Então vamos adoptar o modelo anterior, que é bastante satisfatório e traduz razoavelmente bem o fenómeno em estudo, ou seja, o fenómeno que consiste em verificar qual a face que fica virada para cima quando se lança um dado, aparentemente “normal”.

O facto de se admitir este modelo de probabilidade para o nº de pintas da face que fica virada para cima ao lançar um dado permite-nos agora construir modelos para experiências mais elaboradas, envolvendo vários lançamentos de um dado, ou o lançamento de vários dados.

*Mais uma vez chamamos a atenção para que os modelos de probabilidade, não são mais do que modelos! São idealizações “matemáticas” que tentam traduzir, o melhor possível, a realidade associada ao fenómeno que procuram descrever.*

**Tarefa – Dados especiais**<sup>12</sup>. O professor chegou à aula e apresentou quatro indicações para as probabilidades dos resultados do lançamento de um dado:

	Face						
1º dado	Probabilidade	1/7	1/7	1/7	1/7	1/7	1/7
2º dado	Probabilidade	1/6	1/6	1/6	1/12	1/12	1/3
3º dado	Probabilidade	1/6	1/6	1/6	1/6	1/6	1/3
4º dado	Probabilidade	1	1	2	1	1	2

<sup>12</sup> Adaptado de Moore (1997), p. 415

Quais destas indicações estão correctas, só se pode saber lançando o dado um grande número de vezes. No entanto alguns dos modelos propostos não são modelos de probabilidade legítimos. Porquê?

Nesta situação, em que se argumenta que os dados são especiais, não havendo nenhuma razão para admitir que os dados são equilibrados, só pela experimentação se poderia validar cada um dos modelos. No entanto, só o modelo respeitante ao dado 2 é que é legítimo, pois nenhum dos outros pode ser modelo de probabilidade, já que a soma das probabilidades de todos os resultados não é igual a 1. O modelo 4 ainda tem outro defeito, que é apresentar probabilidades superiores a 1.

**Tarefa – A caixa de pastilhas M&M.** O professor leva para a aula uma caixa de pastilhas M&M, com 100 pastilhas, e diz aos alunos que existem pastilhas de 6 cores. A composição das caixas é tal que a probabilidade de tirar uma pastilha ao acaso e ela ter cada uma das cores é dada pela seguinte tabela:

Cor	Vermelha	Amarela	Castanha	Laranja	Verde	Azul
Probabilidade	0,25	0,25	0,15	0,10	0,20	?

- a) O professor não disse qual a probabilidade de a pastilha ter cor azul. Qual é essa probabilidade?  
Como a soma das probabilidades tem de dar igual a 1, então a probabilidade da pastilha ter a cor azul é 0,05.
- b) O professor pediu a um aluno para tirar 1 pastilha da caixa.
- Qual a probabilidade da pastilha ser vermelha?
  - E qual a probabilidade de ser vermelha ou amarela?
  - E qual a probabilidade de não ser vermelha?
- A probabilidade de ser vermelha é 0,25 ou 25%. A probabilidade de ser vermelha ou amarela,  $P(\text{vermelha ou amarela}) = P(\text{vermelha}) + P(\text{amarela})$ , pois os acontecimentos "Ser vermelha" e "Ser amarela" são disjuntos. Então a probabilidade pretendida é 0,50 ou 50%. A probabilidade de não ser vermelha é igual a 0,75 ( $=1-0,25$ ) ou 75%.
- c) O professor decidiu distribuir, ao acaso, as pastilhas pelos 20 alunos da turma, cabendo a cada um 5 pastilhas. Qual a cor ou cores que se espera surjam com maior frequência? Quantas pastilhas se esperam de cor vermelha? E de cor azul?  
As pastilhas em maior proporção são as vermelhas e amarelas, numa proporção de 25% cada cor. Então nas 100 pastilhas espera-se que cerca de 25 sejam vermelhas e cerca de 25 sejam amarelas. A cor que existe em menor proporção é a azul e espera-se que sejam cerca de 5 pastilhas azuis nas 100 pastilhas.

**Tarefa – O tipo sanguíneo.** Como se sabe, é muito importante que um país tenha conhecimento da forma como se distribui o tipo de grupo sanguíneo, de entre os seus cidadãos. Esta importância reside em vários factores, nomeadamente na previsão de stocks. De acordo com informação disponível na página do Instituto Português do Sangue (<http://www.ipsangue.org/maxcontent-documento-231.html> e revista *ABO* n.º 29 de Janeiro/ Março de 2007) a distribuição, média, dos grupos sanguíneos na população portuguesa faz-se de acordo com o seguinte modelo de probabilidade

Tipo sanguíneo	O	A	B	AB
Probabilidade	42%	47%	8%	3%

Da tabela anterior concluímos que se seleccionarmos um indivíduo ao acaso na população portuguesa, o mais provável é que seja do grupo sanguíneo A e o menos provável é que seja do grupo sanguíneo AB. Quando referimos anteriormente que a distribuição é, em média, a que se apresenta, significa que há algumas alterações de região para região (ver referência anterior do Instituto Português do Sangue).

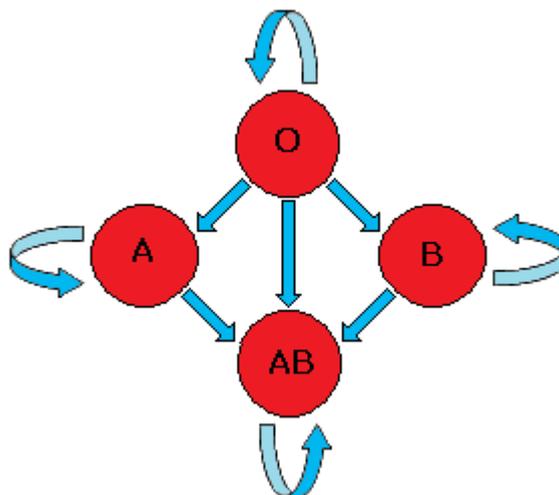
A distribuição do tipo de sangue não é idêntica para todas as populações. Por exemplo, para a população espanhola o modelo anterior não é adequado, já que a distribuição do tipo de sangue se faz (em média) de acordo com o modelo

Tipo sanguíneo	O	A	B	AB
Probabilidade	45%	42%	10%	3%

([http://pt.wikipedia.org/wiki/Grupo\\_sangu%C3%ADneo](http://pt.wikipedia.org/wiki/Grupo_sangu%C3%ADneo))

Como se verifica, ao contrário do que se passa com a população portuguesa, na população espanhola é mais provável que um espanhol, seleccionado ao acaso, tenha sangue de tipo O.

Esta tarefa pode ser adequada para o professor falar na turma sobre a compatibilidade entre os tipos sanguíneos que se apresenta no diagrama seguinte:



Como se exemplifica na figura anterior, o grupo sanguíneo O é dador universal, pois pode doar sangue a qualquer pessoa, independentemente do seu tipo de sangue. No entanto só pode receber sangue de pessoas com o mesmo tipo de sangue O. O grupo sanguíneo AB é receptor universal, pois pode receber sangue de qualquer tipo; no entanto, só pode doar sangue a pessoas do mesmo tipo AB. Os grupos A ou B podem receber sangue dos seus respectivos grupos ou do grupo O. É importante o conhecimento deste esquema, pois numa situação de aflição em que não se tenha tempo de averiguar o grupo sanguíneo de um indivíduo a necessitar de sangue, sabe-se que se pode dar o grupo O.

Tendo em consideração a tabela da distribuição do tipo de grupo sanguíneo da população portuguesa e o esquema anterior, responda às seguintes questões:

- a) O Ricardo tem sangue de tipo O. Se precisar de sangue, qual a probabilidade de escolhendo um indivíduo ao acaso de entre a população portuguesa, ele possa dar sangue ao Ricardo?  
O Ricardo só pode receber sangue de tipo O, pelo que a probabilidade de alguém lhe poder dar sangue, se for seleccionado ao acaso é de 0,42 ou 42%.
- b) A Rita tem sangue de tipo AB. Se precisar de sangue, qual a probabilidade de escolhendo um indivíduo ao acaso de entre a população portuguesa, ele possa dar sangue à Rita.  
Qualquer pessoa pode dar sangue à Rita, pelo que a probabilidade é igual a 1 ou 100%.
- c) A Joana tem sangue de tipo A. Se precisar de sangue, qual a probabilidade de escolhendo um indivíduo ao acaso de entre a população portuguesa, ele possa dar sangue à Joana.  
A Joana pode receber sangue tipo O ou tipo A. Assim a probabilidade que uma pessoa seleccionada ao acaso possa dar sangue à Joana é 0,89 (=0,42+0,47) ou 89%.

**Exemplo – Se seleccionar, ao acaso, uma pessoa residente em Portugal, é mais provável que seja homem ou mulher?** De acordo com o Censo 2001 ([www.ine.pt](http://www.ine.pt)), a população residente em Portugal, distribui-se da seguinte forma, quanto ao sexo:

Sexo	Masculino	Feminino
Nº de residentes	5000141	5355976

Qual a probabilidade de escolhendo um residente ao acaso, ele ser do sexo feminino?

De acordo com a tabela anterior, podemos definir o seguinte modelo de probabilidade para o fenómeno aleatório que consiste em averiguar o sexo de uma pessoa escolhida ao acaso, de entre a população residente:

Sexo	Masculino	Feminino
Probabilidade	0,48	0,52

Para obter o modelo anterior, considerámos como probabilidades dos acontecimentos Masculino e Feminino as suas frequências relativas na população considerada. Admitindo que os 10356117 residentes são igualmente possíveis de ser seleccionados, 5000141 e 5355976 são, respectivamente, o número de resultados favoráveis à ocorrência dos acontecimentos Masculino e Feminino.

Assim, respondendo à questão colocada:

- ✓ A probabilidade de seleccionar um residente ao acaso e ele ser do sexo Feminino é 0,52, ou

$$P(\text{sexo feminino}) = 0,52$$

**Exemplo – A idade da população residente em Portugal** - Consideremos a experiência que consiste em seleccionar uma pessoa ao acaso de entre a população residente em Portugal e verificar qual a classe etária a que pertence, tendo em conta a seguinte classificação: entre 0 e 14 anos, entre 15 e 24 anos, entre 25 e 64

anos e 65 ou mais anos. De acordo com o censo de 2001, temos a seguinte tabela para o número de residentes de cada classe etária:

Idade	Entre 0 e 14 anos	Entre 15 e 24 anos	Entre 25 e 64 anos	65 ou mais anos
Nº de residentes	1656602	1479587	5526435	1693493

Considerando como probabilidades, as frequências relativas, podemos considerar o seguinte modelo de probabilidade, para descrever o fenómeno aleatório que consiste em seleccionar, ao acaso, um indivíduo residente e verificar a que classe etária pertence:

Idade	Entre 0 e 14 anos	Entre 15 e 24 anos	Entre 25 e 64 anos	65 ou mais anos
Probabilidade	0,160	0,143	0,534	0,164

Repare-se que nos dois exemplos anteriores, para uma mesma população, considerámos dois fenómenos aleatórios diferentes, tendo para cada um desses fenómenos construído um modelo de probabilidade: num dos casos estávamos interessados em averiguar o sexo de uma pessoa escolhida ao acaso, enquanto que no outro caso o que nos interessava era saber a idade dessa pessoa. Esta situação serve para chamar a atenção para o facto de que quando estamos interessados em estudar um fenómeno aleatório, ele tem que ser bem descrito, identificando o que se pretende efectivamente estudar. Nos exemplos considerados, não basta dizer que se seleccionou ao acaso um indivíduo residente em Portugal! É necessário dizer o que se pretende observar relativamente a esse indivíduo seleccionado.

#### Tarefa – Qual a cor preferida para pintar o pátio da escola?

O director da escola decidiu investigar, junto dos alunos, qual a cor preferida para pintar o pátio da escola e pediu para votarem uma das seguintes cores: amarelo, verde ou azul. Como resultado da votação, 43% dos alunos escolheram amarelo, 35% azul e 12% verde. Qual a probabilidade de um aluno, escolhido ao acaso, preferir amarelo ou azul? E qual a probabilidade de não preferir o verde?

A atribuição de probabilidades aos acontecimentos pode ser feita de várias formas: ou se repete a experiência muitas vezes e se verifica a percentagem de vezes que o acontecimento se realizou, ou se deduz a partir de resultados igualmente possíveis, ou se tem em conta outro tipo de informação. Neste caso temos as frequências relativas dos acontecimentos "Preferir amarelo", "Preferir azul" ou "Preferir verde", que vamos representar simplesmente por Amarelo, Azul ou Verde. Será que o modelo seguinte pode constituir um modelo de probabilidade para o fenómeno em estudo, que é o de averiguar qual a cor preferida?

Cor preferida	Amarelo	Azul	Verde
Probabilidade	0,43	0,35	0,12

Repare-se que a soma das probabilidades anteriores não é igual a 1. Então não temos um modelo de probabilidade. No entanto basta considerar o resultado "Prefere outra cor diferente das 3 cores consideradas" com uma probabilidade de 0,10, para já termos um modelo de probabilidade.

Cor preferida	Amarelo	Azul	Verde	Outra cor
Probabilidade	0,43	0,35	0,12	0,10

$$\begin{aligned}
 P(\text{Amarelo ou azul}) &= P(\text{Amarelo}) + P(\text{Azul}) \quad (\text{porque os acontecimentos Amarelo e} \\
 &= 0,43 + 0,35 \quad \text{Azul são disjuntos, uma vez que um} \\
 &= 0,43 + 0,35 \quad \text{aluno não pode preferir duas cores)} \\
 &= 0,78
 \end{aligned}$$

A probabilidade de um aluno preferir amarelo ou azul é igual a 78%.

O acontecimento "Não preferir verde" é o complementar de "Preferir verde", pelo que

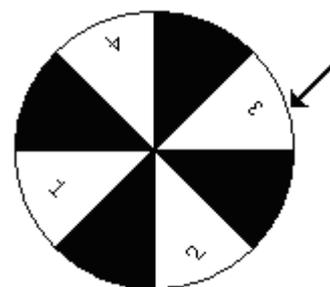
$$\begin{aligned}
 P(\text{Não preferir verde}) &= 1 - P(\text{Verde}) \\
 &= 1 - 0,12 \\
 &= 0,88
 \end{aligned}$$

A probabilidade de um aluno não preferir o verde é 88%.

### Tarefa – A roleta de duas cores

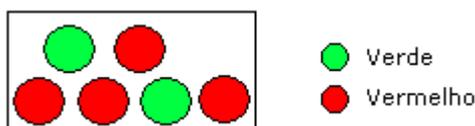
O professor pede a um aluno que faça rodar a roleta e observar a parte apontada pela seta.

- Qual a probabilidade da seta apontar uma parte preta?
- Qual a expectativa sobre o número de vezes que se espera que a seta aponte uma parte preta, se rolar a roleta 20 vezes? Se não obtiver 10 vezes a parte preta significa que a roleta está mal construída?
- Se rolar a roleta 100 vezes, qual a percentagem de vezes que espera que a seta aponte um número par?



A probabilidade da seta apontar uma parte preta é  $\frac{1}{2}$  ou 50%, já que das 8 partes em que está dividida a roleta, que se admitem igualmente possíveis de serem apontadas pela seta, 4 são pretas, donde a probabilidade pretendida é  $\frac{4}{8}$  ou 50%. Se a roleta rodar 20 vezes espera-se que a seta aponte uma parte preta cerca de metade das vezes, ou seja 10 vezes, mas se não apontar 10 vezes não significa que a roleta esteja mal construída. Se rolar a roleta 100 vezes, espera-se que a seta aponte um número par, 25% das vezes.

**Tarefa – O jogo com berlindes.** Numa caixa estão 6 berlindes, 2 verdes e 4 vermelhos. Quando retira o berlinde anota a cor e repõe outra vez na caixa.



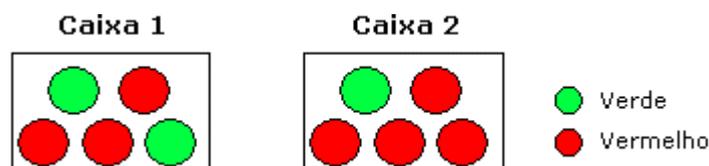
- a) Descreva um modelo de probabilidade para o fenómeno aleatório que consiste em retirar um berlinde e registar a cor.
- b) Ao fim de 300 extracções, com reposição, quantos euros espera ganhar:
  - i) Se por cada berlinde verde que sair, ganhar 1 euro?
  - ii) Se por cada berlinde verde ganhar 2 euros e por cada berlinde vermelho perder um euro?

Um modelo de probabilidade para o fenómeno aleatório que consiste em registar a cor do berlinde pode ser o seguinte:

Cor	Verde	Vermelha
Probabilidade	1/3	2/3

já que dos 6 berlines que estão na caixa, 2 são verdes e 4 são vermelhos. Como nas 300 extracções (com reposição) se espera que saia cerca de 100 vezes berline verde (1/3 das 300 vezes) e nas outras vezes berline vermelho, na primeira situação espera-se ganhar 100 euros, enquanto que na segunda situação se espera ganhar 200 euros e perder outros 200 euros, pelo que nesta segunda hipótese não é de esperar ganhar nem perder.

**Tarefa – As duas caixas de berlines.** Suponha que tem as seguintes caixas, cada uma com 5 berlines verdes e vermelhos. Quando se retira um berline, se ele for verde ganham-se 2 euros, se for vermelho ganha-se 1 euro:



Dão-lhe a possibilidade de escolher uma das 2 caixas para fazer 100 extracções, com reposição. Qual das caixas prefere?

Em cada extracção existem 2 possibilidades em 5 de sair um berline verde, se se fizer a extracção da caixa 1, enquanto que se for da caixa 2, essas possibilidades diminuem para metade. Assim, se se escolher a caixa 1, a probabilidade de extrair um berline verde é 40%, pelo que nas 100 extracções se espera que a cor verde saia cerca de 40 vezes e a cor vermelha 60 vezes, donde se espera ganhar aproximadamente 140 euros ( $=40 \times 2 + 60 \times 1$ ). Com a caixa 2, em que a probabilidade de extrair um berline verde é de 20%, espera-se extrair 20 berlines verdes e 80 vermelhos, donde o ganho seria aproximadamente de 120 euros ( $=20 \times 2 + 80 \times 1$ ). Portanto é preferível a caixa 1.

**Tarefa – Quantos lançamentos são necessários?** Qual a probabilidade de, no lançamento de uma moeda de 1 euro, repetidamente, sair pela primeira vez a face Euro ao fim do 3º lançamento?

Quando se lança uma moeda de 1 euro, a face Euro pode sair no 1º lançamento, ou pode sair pela 1ª vez no 2, ou 3º, ou 4º, etc. Lançamentos. O que se pretende é estimar a probabilidade do acontecimento "A face Euro saiu pela 1ª vez no 3º lançamento".

Para estimar esta probabilidade realizou-se 50 vezes a experiência de lançar a moeda até sair cara pela primeira vez, cujos resultados se apresentam a seguir:

Nº experiência	1º	2º	3º	4º	5º	6º	7º	Nº lançamentos
1	N	E						2
2	E							1
3	E							1
4	E							1
5	E							1
6	E							1
7	N	N	N	E				4

8	N	N	E					<b>3</b>
9	N	E						2
10	N	N	N	N	N	E		6
11	N	E						2
12	E							1
13	N	N	E					<b>3</b>
14	N	E						2
15	N	E						2
16	E							1
17	N	E						2
18	E							1
19	N	E						2
20	E							1
21	N	E						2
22	N	N	N	N	N	N	E	7
23	E							1
24	E							1
25	E							1
26	E							1
27	E							1
28	E							1
29	E							1
30	E							1
31	E							1
32	N	N	N	E				4
33	N	N	N	E				4
34	N	E						2
35	E							1
36	E							1
37	E							1
38	N	E						2
39	E							1
40	N	E						2
41	E							1
42	N	N	E					<b>3</b>
43	N	N	E					<b>3</b>
44	N	N	E					<b>3</b>
45	N	N	N	N	E			5
46	N	E						2
47	E							1
48	N	N	E					<b>3</b>
49	E							1
50	N	E						2

Nas 50 experiências, verificou-se a saída da face Euro pela primeira vez ao 3º lançamento, 6 vezes, pelo que uma estimativa para a probabilidade do acontecimento "A primeira vez que se regista a face Euro é ao fim do 3º lançamento" é  $6/50=0,12$ .

Vamos aproveitar as experiências anteriores para atribuir um modelo de probabilidade para o fenómeno aleatório que consiste em averiguar quantas vezes é necessário lançar a moeda até sair a face Euro. Quais os resultados possíveis? Nas experiências anteriores o valor máximo que obtivemos para o número de

lançamentos foi 7, mas ninguém nos garante que ao realizar outra experiência não sejam necessários mais lançamentos. Assim propomos o seguinte *modelo empírico*, em que consideramos como probabilidades as frequências relativas obtidas nas 50 experiências:

Nº de lançamentos	1	2	3	4	5	6	7 ou mais
Probabilidade	25/50	13/50	6/50	3/50	1/50	1/50	1/50

ou

Nº de lançamentos	1	2	3	4	5	6	7 ou mais
Probabilidade	0,50	0,26	0,12	0,06	0,02	0,02	0,02

Vamos admitir que a moeda era equilibrada e considerar um *modelo teórico* baseado nesta hipótese. Então, vejamos como calcular as probabilidades para os vários acontecimentos elementares:

1 lançamento, ou seja, sair a face E quando se lança a moeda:

Casos igualmente possíveis: E N

Casos favoráveis: E

$$P(1 \text{ lançamento}) = 1/2 \\ = 0,50$$

2 lançamentos, ou seja, não sair no 1º, mas no 2º:

Casos igualmente possíveis: EE, EN, NE, NN

Casos favoráveis: NE

$$P(2 \text{ lançamentos}) = 1/4 \\ = 0,25$$

3 lançamentos, ou seja, não sair no 1º nem no 2º e sair no 3º:

Casos igualmente possíveis: EEE, EEN, ENE, ENN, NEE, NEN, NNE, NNN

Casos favoráveis: NNE

$$P(3 \text{ lançamentos}) = 1/8 \\ = 0,125$$

4 lançamentos, ou seja, não sair no 1º, nem no 2º, nem no 3º e sair no 4º:

Casos igualmente possíveis: EEEE, EEEN, EENE, EENN, ENEE, ENEN, ENNE, ENNN, NEEE, NEEN, NENE, NENN; NNEE, NNEE, NNNE, NNNN

Casos favoráveis: NNNE

$$P(4 \text{ lançamentos}) = 1/16 \\ = 0,062$$

Repare-se que a metodologia aqui seguida para obter os resultados possíveis sempre que se faz mais um lançamento foi acrescentar E ou N a cada resultado possível do lançamento anterior. Assim, de lançamento para lançamento, os resultados possíveis duplicam, mas obtemos sempre um único resultado favorável. Uma vez explicado o mecanismo, temos:

5 lançamento, ou seja, não sair no 1º, nem no 2º, nem no 3º, nem no 4º e sair no 5º:

Casos igualmente possíveis: EEEEE, EEEEN, ..., em número de 32

Casos favoráveis: NNNNE

$$P(5 \text{ lançamentos}) = 1/32 \\ = 0,031$$

6 lançamentos, ou seja, não sair no 1º, nem no 2º, ..., nem no 5º e sair no 6º:

Casos igualmente possíveis: EEEEE, EEEEE, ..., em número de 64

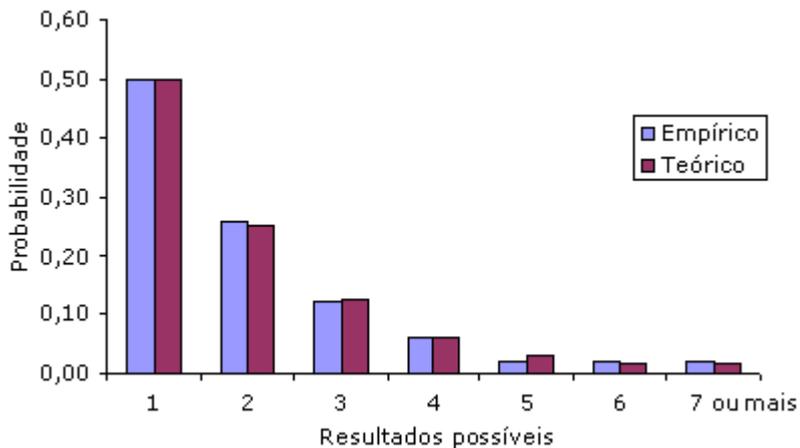
Casos favoráveis: NNNNE

$$P(6 \text{ lançamentos}) = 1/64 \\ = 0,016$$

7 ou mais lançamentos:

$$P(7 \text{ ou mais}) = 1 - P(1 \text{ ou } 2 \text{ ou } 3 \text{ ou } 4 \text{ ou } 5 \text{ ou } 6) \\ = 1 - \{P(1) + P(2) + P(3) + P(4) + P(5) + P(6)\} \text{ pois os acontecimentos} \\ \text{são disjuntos} \\ = 1 - 0,984 \\ = 0,016$$

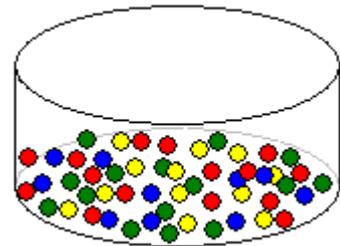
Comparando os dois modelos, verificamos que as probabilidades empíricas e teóricas estão muito próximas:



**Tarefa – A caixa com drageias de chocolate** - Uma caixa tem 52 drageias de chocolate, das quais 15 são vermelhas, 10 azuis, 12 amarelas e as restantes verdes.

Depois de abanar a caixa, para misturar as drageias, retira-se uma ao acaso, sem olhar. Qual a probabilidade de ser verde? Construir um modelo de probabilidade para o fenómeno aleatório que consiste em tirar uma drageia e verificar a cor.

$$\text{Número de drageias verdes} = 52 - (15 + 10 + 12) \\ = 52 - 37 \\ = 15$$



Como a drageia é retirada ao acaso, todas têm igual possibilidade de serem retiradas, pelo que a probabilidade pretendida é

$$P(\text{drageia verde}) = 15/52$$

As probabilidades de retirar uma drageia vermelha, amarela ou azul, calculam-se de forma análoga, pelo que temos o seguinte modelo de probabilidade:

Cor da drageia	Verde	Vermelha	Amarela	Azul
Probabilidade	15/52	15/52	12/52	10/52

**Tarefa – O jogo de andebol** - O estatístico da equipa de andebol de uma certa escola, com base no historial de jogos anteriores com o mesmo adversário, sugeriu o seguinte modelo probabilístico para o resultado final do próximo jogo:

Resultado	Vitória	Empate	Derrota
Probabilidade	0,4	0,1	0,5

O treinador, que acha que a equipa está a atravessar um bom momento de forma, é de opinião que a probabilidade de Vitória deverá ser igual a 0,6 e não 0,4. Admitindo que a probabilidade de Empate não se altera, qual é a probabilidade da equipa vir a ser derrotada?

A soma das probabilidades tem de ser igual a 1 (100%). Assim a probabilidade de derrota passará a ser igual a 0,3.

Seria possível manter a probabilidade de derrota alterando a probabilidade de empate? Não, pois  $0,6+0,5=1,1$  e, para a soma de todas as probabilidades ser igual a 1, a probabilidade de empate teria de ser negativa, o que não é possível num modelo probabilístico.

**Tarefa – Escolhe ao acaso uma letra do alfabeto.** O professor colocou a seguinte questão aos alunos: Suponham que escrevíamos cada letra do alfabeto num papel, colocávamos numa caixa e depois de baralhar, seleccionávamos um papel ao acaso. Qual a probabilidade de ser vogal? Se seleccionar 2 letras, qual a probabilidade de uma ser vogal e a outra consoante?

Como o alfabeto português tem 26 letras (as letras k, w e y já pertencem ao alfabeto português), das quais 5 são vogais, a probabilidade pretendida é

$$P(\text{vogal}) = 5/26 \\ = 0,192$$

Para obter a probabilidade de ao retirar 2 letras, uma ser vogal e a outra consoante, vamos contar o número de casos possíveis e de casos favoráveis.

- Casos possíveis: cada letra da 1.<sup>a</sup> extracção pode combinar-se com qualquer letra na 2.<sup>a</sup> extracção, excepto consigo própria, pelo que o número de casos possíveis é  $26 \times 25$ ;
- Casos favoráveis: se na 1.<sup>a</sup> extracção sair uma vogal, ela pode combinar-se com qualquer consoante na 2.<sup>a</sup> extracção e temos  $5 \times 21$  casos; se na 1.<sup>a</sup> extracção sair uma consoante, ela pode combinar-se com qualquer vogal e o número de casos é  $21 \times 5$ ; assim o número de casos favoráveis é  $5 \times 21 + 21 \times 5$ ;

$$P(\text{vogal e consoante}) = \frac{5 \times 21 + 21 \times 5}{26 \times 25} \\ = 0,323$$

**Tarefa – Escolhe ao acaso uma letra da palavra “palavra”.** Se escolheres ao acaso uma letra da palavra “palavra” quais os resultados que podes obter? São todos igualmente possíveis? Qual a probabilidade de seleccionares cada uma das letras?

Quando seleccionamos ao acaso uma letra da palavra "palavra", os resultados possíveis que podemos obter são

p   a   l   v   r

Estes resultados não são todos igualmente possíveis, pois a letra **a** tem maior probabilidade de ser seleccionada, pois das 7 letras que compõem a palavra, 3 são a's. Então temos 3 em 7 possibilidades de escolher o a e 1 em 7 possibilidades de escolher qualquer uma das outras letras. Assim, um modelo de probabilidade para a escolha de uma letra ao acaso de "palavra" é:

Letra	p	a	l	v	r
Probabilidade	1/7	3/7	1/7	1/7	1/7

Repare-se que a soma das probabilidades é igual a 1, como seria de esperar.

Pode haver da parte de alguns alunos alguma dificuldade na compreensão da atribuição de probabilidades aos acontecimentos elementares anteriores constituídos por cada uma das letras anteriores. O professor poderá abordar o problema da seguinte forma: escreve cada um dos a's de palavra com uma cor diferente e considera como resultados possíveis as 7 letras, em que distingue os 3 a's, mas em que agora os 7 resultados são igualmente possíveis

Letra	p	a	l	a	v	r	a
Probabilidade	1/7	1/7	1/7	1/7	1/7	1/7	1/7

O acontecimento que consiste em seleccionar um a é composto por 3 resultados favoráveis, que são os 3 a's de cores diferentes, de entre 7 possíveis, de onde

$$\begin{aligned}
 \text{Probabilidade de seleccionar a} &= \text{Probabilidade de seleccionar (a ou a ou a)} \\
 &= \text{Prob. de sel. a} + \text{Prob. de sel. a} + \text{Prob. de sel. a} \\
 &= 1/7 + 1/7 + 1/7 \\
 &= 3/7
 \end{aligned}$$

A probabilidade de seleccionar cada uma das outras letras é 1/7.

**Tarefa – Será que o Pedro vai comer a sua fruta preferida?** O professor pediu a cada um dos alunos que trouxessem no dia seguinte uma peça de fruta, que podia ser a fruta preferida. Nesse dia, à medida que cada aluno trazia a peça de fruta esta era colocada num saco, enquanto o professor apontava no quadro quais os tipos de fruta e quantas peças de cada tipo se tinham juntado no saco:

maçã	IIII
pera	IIIIII
banana	IIIIIIII
Laranja	III

Quando chegasse a hora do lanche, cada aluno retirava do saco, ao acaso, uma peça de fruta. Qual a probabilidade de ao André, que era o primeiro aluno a retirar a peça de fruta, calhar a sua fruta preferida, que é a banana?

Quando o professor acabou de explicar como iria ser feita a selecção de cada fruta, alguns alunos não estavam de acordo! Segundo eles, a escolha da peça de fruta

não poderia ser feita de forma completamente aleatória, pois se as maçãs se podiam confundir ligeiramente com as laranjas e eventualmente com as peras, já o mesmo não se passava com as bananas. Estas são facilmente distinguíveis ao tacto, mesmo de olhos fechados. Então decidiram utilizar a seguinte metodologia: escreviam o nome de cada peça de fruta num papel, metiam o papel no saco, e o que os alunos seleccionavam era o nome do fruto e não o fruto propriamente dito. Ao todo o saco continha 24 papéis com nomes de frutos, dos quais 9 eram de bananas. Então a probabilidade de ao Pedro calhar uma banana é  $9/24$ .

**Tarefa – A caixa de disquetes.** Uma caixa tem 5 disquetes, das quais 1 está avariada. Selecciono 2 disquetes ao acaso. Qual a probabilidade de me calhar a disquete avariada? Obtenha um modelo de probabilidade para o fenómeno aleatório que consiste em escolher 2 disquetes e verificar quantas estão avariadas.

Para facilitar, vamos representar as disquetes boas por B1, B2, B3 e B4 e a disquete avariada por A. Para obter os resultados possíveis associados a esta experiência vamos considerar a seguinte tabela (estamos perante uma extracção sem reposição):

		2ª disquete				
		B1	B2	B3	B4	A
1ª disquete	B1	-	B1 B2	B1 B3	B1 B4	B1 A
	B2	B2 B1	-	B2 B3	B2 B4	B2 A
	B3	B3 B1	B3 B2	-	B3 B4	B3 A
	B4	B4 B1	B4 B2	B4 B3	-	B4 A
	A	A B1	A B2	A B3	A B4	-

O espaço de resultados é constituído por 20 resultados igualmente possíveis, dos quais 8 fazem com que se verifique o acontecimento pretendido e que é "seleccionar a disquete avariada". Então a probabilidade pretendida é  $8/20$  ou 40%.

O espaço de resultados associado à experiência aleatória que consiste em extrair 2 disquetes e verificar quantas estão avariadas só pode ter como resultados possíveis o 0 e o 1. A probabilidade de 0 disquetes avariadas na amostra das 2 é  $12/20$ , donde podemos considerar o seguinte modelo de probabilidade para o número de disquetes avariadas na amostra das duas disquetes:

Número de disquetes avariadas	0	1
Probabilidade	0,60	0,40

**Tarefa – Lançamento de um dado equilibrado.** Considere-se a experiência aleatória que consiste em lançar um dado equilibrado e em verificar a face que fica voltada para cima.

- Qual o modelo de probabilidade associado ao fenómeno aleatório em estudo?
- Qual a probabilidade de se obter uma face com 2 ou menos pintas?
- Qual a probabilidade de se obter um número par?

Resolução: a) Ao dizermos que o dado é equilibrado estamos a colocar-nos numa situação de simetria ou equilíbrio, pelo que admitimos o seguinte modelo para a probabilidade de saída de cada uma das faces do dado:

Face	1	2	3	4	5	6
Probabilidade	1/6	1/6	1/6	1/6	1/6	1/6

b) O acontecimento A, "Face com um número de pintas menor ou igual a 2" tem 2 resultados favoráveis,  $A = \{1, 2\}$  donde  $P(A) = 2/6$

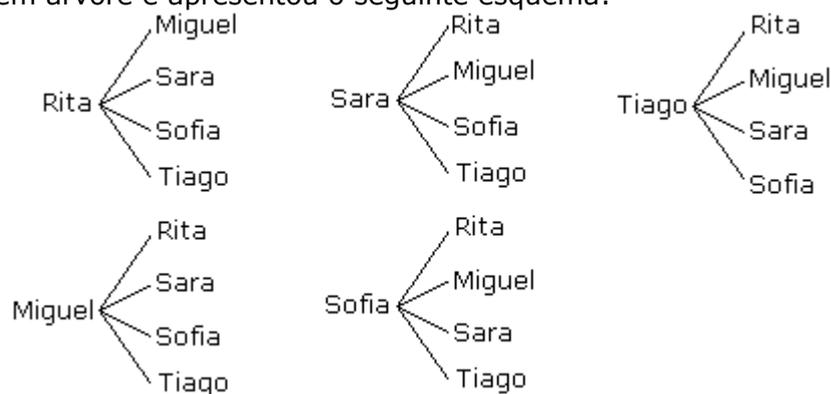
c) Quanto ao acontecimento B, "Saída de número par", como temos 3 resultados favoráveis,  $B = \{2, 4, 6\}$ , donde  $P(B) = 3/6$ .

**Tarefa – Uma escolha difícil entre 5 candidatos!** O professor pretendia eleger uma comissão constituída por 2 alunos para preparar a festa de finalistas. Como esta comissão tinha algumas regalias, nomeadamente estava dispensada de algumas aulas..., quando o professor acabou de falar, houve logo 5 alunos – a Rita, o Miguel, a Sara, a Sofia e o Tiago, que quase simultaneamente levantaram o braço desejosos de pertencer à comissão! Questões colocadas pelo professor:

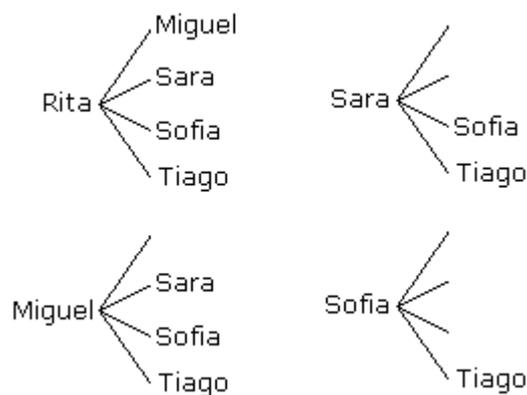
- Como escolher 2 alunos de entre os cinco candidatos?
- Qual a probabilidade de os dois rapazes fazerem parte da comissão?
- Qual a probabilidade de a comissão ser constituída por um aluno de cada sexo?

Como o professor não pretende privilegiar nenhum dos alunos, terá de proceder a uma selecção aleatória de uma amostra de 2 alunos de entre os 5 alunos candidatos. Depois de alguma discussão, os alunos recordaram o que o professor tinha ensinado sobre a selecção aleatória de amostras. Neste caso é evidente que se tem de proceder a uma selecção sem reposição, pois não se pode correr o risco de o mesmo aluno ser seleccionado duas vezes. Então decidiu-se proceder do seguinte modo: escrevem-se os 5 nomes em pedaços de papel de igual tamanho, dobram-se, colocam-se numa caixa, baralham-se e retiram-se 2 nomes de uma vez.

O professor tinha explicado que o processo anteriormente descrito de seleccionar uma amostra, dá a todas as amostras a mesma probabilidade de serem seleccionadas. Para calcular as probabilidades pretendidas, é necessário começar por construir o espaço de resultados. Um grupo de alunos propôs utilizar um diagrama em árvore e apresentou o seguinte esquema:



Contudo, ao analisarem o diagrama anterior, alguns alunos questionaram o facto de na escolha dos dois alunos não interessar a ordem pela qual os alunos são escolhidos, pois seleccionaram-se os dois alunos ao mesmo tempo, pelo que seleccionar a Rita e o Miguel é idêntico a seleccionar o Miguel e a Rita. Assim, os pares que interessam são os dos seguintes ramos, pois todos os outros estão repetidos:



Espaço de resultados:

$S = \{\text{Rita Miguel, Rita Sara, Rita Sofia, Rita Tiago, Miguel Sara, Miguel Sofia, Miguel Tiago, Sara Sofia, Sara Tiago, Sofia Tiago}\}$

Modelo de probabilidade:

Par	Rita Miguel	Rita Sara	Rita Sofia	Rita Tiago	Miguel Sara	Miguel Sofia	Miguel Tiago	Sara Sofia	Sara Tiago	Sofia Tiago
Prob.	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10

O acontecimento "Os dois rapazes fazem parte da comissão" é constituído por um único resultado, pelo que a probabilidade pretendida é igual a 1/10.

O acontecimento "A comissão tem alunos dos dois sexos" é constituído por 6 resultados, pelo que a probabilidade pretendida é igual a 6/10.

A tarefa não terminou aqui, pois o professor colocou a seguinte questão: o que é que acontece se a selecção dos dois nomes não for feita de uma única vez, mas sequencialmente (sem repor o primeiro elemento seleccionado), interessando a ordem pela qual os nomes são seleccionados? Esta situação poderia colocar-se se, por exemplo, o primeiro elemento a ser seleccionado fosse o "representante" da comissão.

Qual o espaço de resultados? Quais as probabilidades dos acontecimentos em jogo?

Os alunos que tinham apresentado o primeiro esquema em árvore sugeriram que agora o espaço de resultados seria constituído pelos vinte resultados, visualizados no 1º esquema:

$S = \{\text{Rita Miguel, Rita Sara, Rita Sofia, Rita Tiago, Miguel Rita, Miguel Sara, Miguel Sofia, Miguel Tiago, Sara Rita, Sara Miguel, Sara Sofia, Sara Tiago, Sofia Rita, Sofia Miguel, Sofia Sara, Sofia Tiago, Tiago Rita, Tiago Miguel, Tiago Sara, Tiago Sofia}\}$

Como os resultados também são todos igualmente possíveis, o modelo de probabilidade adequado é o seguinte:

Par	Rita Miguel	Rita Sara	Rita Sofia	Rita Tiago	Miguel Rita	Miguel Sara	Miguel Sofia	Miguel Tiago	Sara Rita	Sara Miguel
Prob.	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20
Par	Sara Sofia	Sara Tiago	Sofia Rita	Sofia Miguel	Sofia Sara	Sofia Tiago	Tiago Rita	Tiago Miguel	Tiago Sara	Tiago Sofia
Prob.	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20

O acontecimento "Os dois rapazes fazem parte da comissão" é constituído por dois resultados, pelo que a probabilidade pretendida é igual a 2/20 ou 1/10.

O acontecimento "A comissão tem alunos dos dois sexos" é constituído por 12 resultados, pelo que a probabilidade pretendida é igual a  $12/20$  ou  $6/10$ .

Concluimos que os acontecimentos anteriores têm probabilidades iguais. No entanto agora, com este outro espaço de resultados tem sentido considerar outros acontecimentos, como por exemplo "O Miguel foi seleccionado para representante de um grupo". Neste caso, de entre os 20 resultados possíveis só temos 4 favoráveis, pelo que a probabilidade pretendida é  $4/20$ . Repare-se que este acontecimento não é o mesmo que "O Miguel foi seleccionado para a comissão", cuja probabilidade é  $8/20$ .

**Tarefa – Um jogo desequilibrado!** O professor propôs o seguinte jogo ao Tiago e à Inês: lançam um dado e se sair uma face com menos de 3 pintas, o Tiago ganha um rebuçado. Se sair uma face com 3 ou mais pintas a Inês ganha um rebuçado.

- Ao fim de alguns lançamentos do dado, quem é que esperas que tenha mais rebuçados? Achas que o jogo é equilibrado?
- Ao fim de 30 lançamentos do dado, quantos rebuçados esperas que tenha cada um dos jogadores?
- Se concluíste que o jogo é desequilibrado, podes ajudar o professor a mudar as regras sobre o número de rebuçados que dá ao Tiago ou à Inês, em cada jogada, de modo que após 30 jogadas tenham aproximadamente o mesmo número de rebuçados?

No jogo anterior, o Tiago tem metade da probabilidade da Inês de ganhar um rebuçado, já que a Inês tem 4 possibilidades em 6 de o ganhar. Assim, ao fim de alguns lançamentos é de esperar que Inês tenha mais rebuçados e o jogo não é equilibrado.

Ao fim de 30 lançamentos, espera-se que  $4/6$  das vezes tenha saído uma face com 3 ou mais pintas, pelo que se espera que a Inês ganhe 20 das 30 jogadas e receba, por conseguinte, 20 rebuçados, enquanto que o Tiago só receberá 10.

Para tornar o jogo equilibrado, basta dar 2 rebuçados ao Tiago, sempre que sair uma face com 1 ou 2 pintas. Embora se espere que ele só ganhe 10 das 30 jogadas, receberá 20 rebuçados, já que agora cada jogada ganha dá direito a 2 rebuçados.

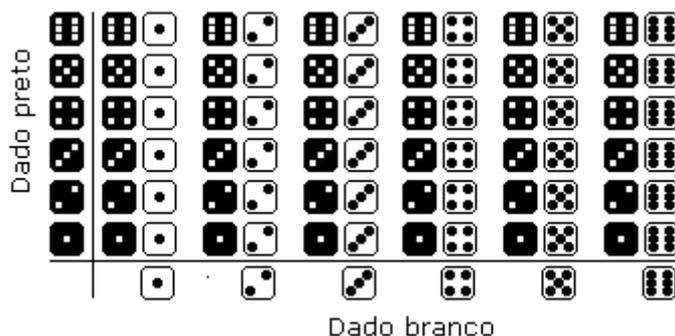
**Tarefa – Lançamento de dois dados. Será que o jogo é justo?** Na escola o professor propôs o seguinte jogo para ser jogado entre a Rita, o João e o Miguel: lançam-se 2 dados de 6 faces e verifica-se a soma das pintas dos dados, que pode ir de 2 a 12. Se a soma for 2, 3, 4 ou 5 o João ganha um ponto; se for 6, 7 ou 8 ganha a Rita um ponto; finalmente, se for 9, 10, 11 ou 12, ganha o Miguel. A Rita ficou muito zangada com o professor, dizendo que este a estava a desfavorecer, uma vez que aos outros colegas dava 4 possibilidades, enquanto que a ela só dava 3 e assim os colegas tinham uma maior probabilidade de ganharem. Será que ela tinha razão?

Para ser mais fácil de descrever a actividade, vamos considerar dois dados em que um é preto e o outro é branco. Vamos esquematizar todas as situações possíveis de se verificarem, quando se lançam os dois dados:

Suponhamos que no dado preto saiu 1 pinta. Então no dado branco pode ter saído qualquer valor de 1 a 6:



Repetindo o processo, mas agora com 2, 3, ..., 6 pintas no dado preto, obtemos a figura seguinte, onde temos esquematizado todos as situações possíveis, em número de 36 do espaço de resultados S:



Vamos considerar uma tabela com os números das pintas e a soma respectiva:

6+1=7	6+2=8	6+3=9	6+4=10	6+5=11	6+6=12
5+1=6	5+2=7	5+3=8	5+4=9	5+5=10	5+6=11
4+1=5	4+2=6	4+3=7	4+4=8	4+5=9	4+6=10
3+1=4	3+2=5	3+3=6	3+4=7	3+5=8	3+6=9
2+1=3	2+2=4	2+3=5	2+4=6	2+5=7	2+6=8
1+1=2	1+2=3	1+3=4	1+4=5	1+5=6	1+6=7

Analisando com cuidado a tabela anterior, verificamos que existem algumas somas que surgem com mais frequência do que outras. Por exemplo a soma 12 só aparece quando sair 6 pintas nos dois dados



enquanto que a soma 5 aparece nas seguintes situações



Então concluímos que enquanto a probabilidade de o 12 sair é de 1 em 36, o 5 tem uma probabilidade maior, de 5 em 36. A partir da tabela anterior podemos construir uma outra tabela, com o número de vezes que pode sair cada resultado para a soma das pintas, quando se lançam 2 dados:

Resultado da soma das pintas	Número de vezes que se pode verificar	Quem ganha
2	1	João
3	2	João
4	3	João
5	4	João
6	5	Rita
7	6	Rita
8	5	Rita
9	4	Miguel
10	3	Miguel
11	2	Miguel
12	1	Miguel

Então quando se lançam os dois dados, de acordo com as regras estipuladas para o jogo:

- o João tem 10 ( $1+2+3+4$ ) possibilidades de ganhar;
- a Rita tem 16 ( $5+6+5$ ) possibilidades de ganhar;
- o Miguel tem 10 ( $4+3+2+1$ ) possibilidades de ganhar.

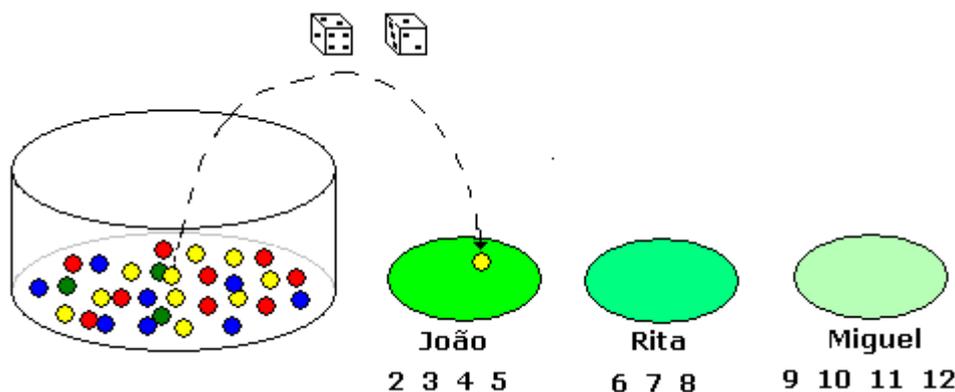
Afinal a Rita não tinha razão, pois estava a ser privilegiada neste jogo, que não era um jogo justo.

O professor então propôs que redistribuíssem os resultados possíveis pelos 3 colegas, de forma a transformarem um jogo que não era justo, num jogo justo. Depois de alguma discussão, propuseram a seguinte regra: se a soma for 2, 7 ou 8 o João ganha um ponto; se for 4, 5 ou 6 ganha a Rita um ponto; finalmente, se for 3, 9, 10, 11 ou 12, ganha o Miguel. Será que chegaram a uma boa solução?

Ainda aproveitando os resultados da tabela anterior, pode-se considerar o seguinte modelo de probabilidade associado à experiência que consiste em lançar dois dados e verificar a soma das pintas das faces que ficam viradas para cima:

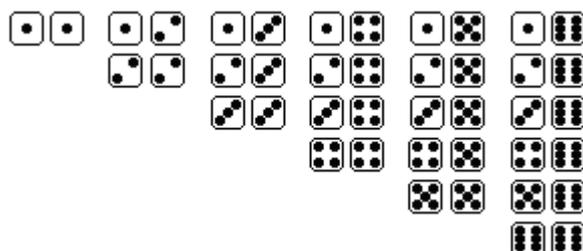
Resultado	2	3	4	5	6	7	8	9	10	11	12
Probabilidade	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

**Tarefa – Vamos lançar 2 dados (cont).** Uma versão desta tarefa pode ser realizada na sala de aula da seguinte forma: o professor coloca numa taça de plástico transparente alguns smarties (em número superior ao número de alunos da turma). A professora lança 2 dados e conforme o número que se verificar para a soma das pintas das faces que ficam voltadas para cima, retira um smartie da taça e coloca no prato do João, da Rita ou do Miguel (na figura, exemplificamos uma situação em que a soma é igual a 3, pelo que o smartie foi colocado no prato do João). Quando se esgotarem os smarties da taça, ganha aquele que tiver maior número de smarties no seu prato. Quem é que se espera que ganhe?



No fim do jogo todos os alunos têm direito a um smartie, ficando o aluno ganhador com os que sobram.

**Tarefa – Ainda o lançamento de dois dados.** Quando formalizámos a noção de espaço de resultados, vimos que associado à experiência aleatória que consiste em lançar dois dados e verificar as faces que ficam voltadas para cima, existem algumas dificuldades em especificar o espaço de resultados. Esta dificuldade prende-se com o facto de se poderem considerar os dois dados idênticos, como *indistinguíveis*. Neste caso, em que se consideram os dados indistinguíveis, o número de resultados do espaço de resultados não é 36, como na tarefa anterior, mas sim 21, como se apresenta a seguir, para o espaço  $S'$ :



Os dois espaços de resultados considerados respectivamente na tarefa anterior e nesta tarefa, são usados para modelar situações diferentes, embora ambas associadas à experiência que consiste em lançar dois dados e verificar as faces voltadas para cima. Enquanto que no primeiro caso, se utiliza o espaço de resultados  $S$  com 36 resultados, para modelar o lançamento de dois dados iguais ou dois dados diferentes lançados ao mesmo tempo, ou um dado lançado duas vezes, no segundo caso considera-se que os dois dados são indistinguíveis.

Como atribuir probabilidades aos resultados deste espaço de resultados  $S'$ ? Podemos, como na tarefa anterior, argumentar que os resultados são todos igualmente possíveis? Não! Se lançarmos os dados, há o dobro das possibilidades de se obter, por exemplo (1,2) do que (1,1)! Os resultados em que as faces são iguais têm metade da possibilidade de qualquer um dos outros resultados. Se representarmos a probabilidade de sair faces iguais por  $p$  temos para a probabilidade de sair qualquer resultado:

$$\begin{aligned}
 &P + 2 \times p + \\
 &P + 2 \times p + 2 \times p + 2 \times p + 2 \times p + \\
 &P + 2 \times p + 2 \times p + 2 \times p + \\
 &P + 2 \times p + 2 \times p + \\
 &P + 2 \times p + \\
 &P
 \end{aligned}$$

Como a soma das probabilidades dos acontecimentos elementares que constituem o espaço de resultados tem de ser igual a 1, e tendo em consideração que a soma anterior é igual a  $36 \times p$ , vem que

$$36 \times p = 1 \rightarrow p = 1/36$$

e o modelo de probabilidade associado é

Resultado	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
Probabilidade	1/36	2/36	2/36	2/36	2/36	2/36	1/36	2/36	2/36	2/36	2/36
Resultado	(3,3)	(3,4)	(3,5)	(3,6)	(4,4)	(4,5)	(4,6)	(5,5)	(5,6)	(6,6)	
Probabilidade	1/36	2/36	2/36	2/36	1/36	2/36	2/36	1/36	2/36	1/36	

Para atribuir as probabilidades anteriores aos acontecimentos elementares respectivos, poderíamos ter utilizado o modelo de probabilidades associado ao espaço de resultados  $S$ , em que cada acontecimento elementar tem probabilidade  $1/36$ . Bastaria considerar cada resultado do espaço  $S'$ , em que as faces são diferentes, como um acontecimento constituído por dois resultados elementares do espaço  $S$ . Por exemplo, o resultado  $(1,2)$  de  $S'$  é obtido se se verificar  $(1,2)$  ou  $(2,1)$  em  $S$ .

**Tarefa – O lançamento de duas moedas**<sup>13</sup>. O professor no início da aula colocou a seguinte questão: tenho aqui duas moedas de 1 euro. Se as lançar, qual a probabilidade de obter “uma face Euro e uma face Nacional”?

Os alunos dividiram-se na resposta a esta questão, que foi respondida da seguinte forma: Um pouco mais de 50% dos alunos respondeu  $1/2$ , cerca de 25% dos alunos respondeu  $1/3$  e os restantes responderam  $1/4$ . Qual o raciocínio que pode ter estado por trás destas respostas? Esses raciocínios estarão correctos ou não?

A resposta dada pela maior parte dos alunos é aquela que, de um modo geral, estaríamos à espera, já que os resultados possíveis do lançamento das duas moedas são

Euro Euro, Euro Nacional, Nacional Euro e Nacional Nacional

havendo, por conseguinte, 2 resultados favoráveis de entre 4 resultados igualmente possíveis. No entanto, com este mesmo espaço de resultados, alguns alunos interpretaram o acontecimento “uma face Euro e uma face Nacional”, como sendo uma face Euro, seguida de uma face Nacional, que embora não estivesse implícita na pergunta, poderia, efectivamente, ser interpretado dessa maneira, pelo que o raciocínio dos alunos que indicaram  $1/4$  está correcto. Mais complicado será interpretar o que levou alguns alunos a responderem  $1/3$ ! Eventualmente teriam pensado em 3 resultados possíveis, dos quais só 1 seria favorável... e uma possibilidade é que tenham considerado como resultados possíveis

2 faces Euro, 2 faces Nacional, 1 face de cada qualidade

esquecendo-se que estes resultados não são igualmente possíveis, pelo que não se pode aplicar a regra de Laplace. Efectivamente o resultado “uma face de cada qualidade” pode ser obtido de duas maneiras possíveis: “face Euro e face Nacional” ou “face Nacional e face Euro”.

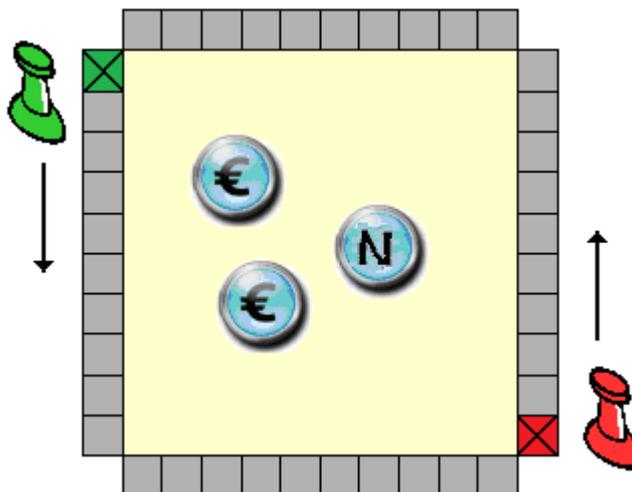
O professor para que os alunos compreendessem melhor esta situação, mostrou, por analogia, o que se passa com o lançamento de dois dados, em que o que interessa é o resultado da soma das pintas. Também neste caso, o espaço de resultados é constituído por 11 resultados, mas não são todos igualmente possíveis. Por exemplo, a probabilidade de obter 11 não é igual à probabilidade de obter 12, já que o 11 pode ser obtido de duas maneiras distintas, enquanto que o 12 só pode ser obtido de uma maneira.

---

<sup>13</sup> Sugerida por Burrill (2006), p. 50

**Tarefa – Quem consegue dar primeiro a volta ao quadrado?** Propomos a seguir um jogo<sup>14</sup> que consiste no seguinte.

Numa cartolina, constrói-se um quadrado com 12 unidades de lado e em cada lado constroem-se 10 quadradinhos com uma unidade de lado:



Este jogo é jogado por dois jogadores que movem o seu pino na direcção indicada sendo o objectivo do jogo e as regras, os seguintes:

*Objectivo do jogo:* Ser o primeiro jogador a dar uma volta completa ao quadrado, chegando ao ponto de onde partiu.

*Regra do jogo:* Depois de se escolher qual o jogador que começa o jogo, este lança 3 moedas de um euro e movimenta o seu pino um certo número de quadradinhos, de acordo com o tipo de faces que ficam voltadas para cima

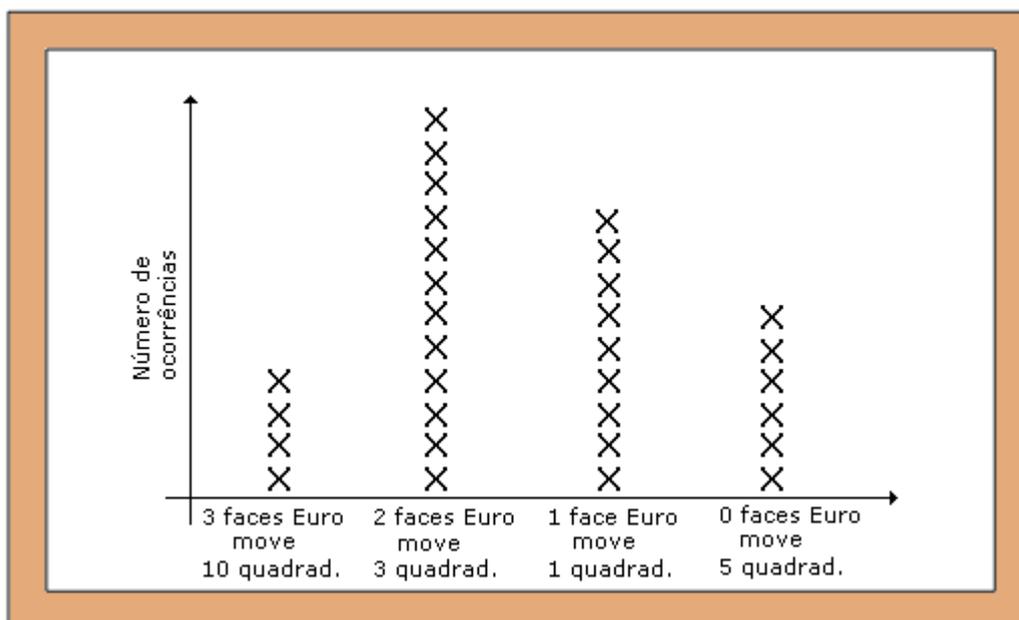
3 faces Euro	→ movimenta 10 quadradinhos
2 faces Euro e 1 Nacional	→ movimenta 3 quadradinhos
1 face Euro e 2 Nacional	→ movimenta 1 quadradinhos
3 faces Nacional	→ movimenta 5 quadradinhos

Quando o primeiro jogador acabar de mover o seu pino, será a vez do segundo jogador lançar as 3 moedas e movimentar o seu pino.

Existe ainda uma regra especial e que é a seguinte: sempre que o pino de um jogador atingir ou ultrapassar o pino do outro jogador, este jogador que foi ultrapassado volta ao ponto de onde partiu.

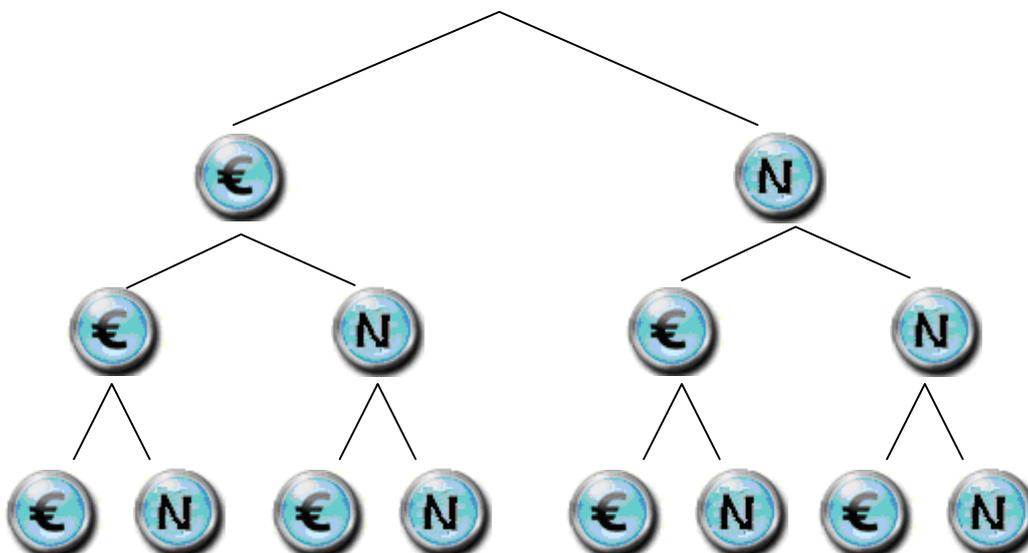
Esta tarefa, orientada pelo professor, vai ser utilizada para comparar os resultados experimentais do lançamento das 3 moedas, com alguns resultados teóricos. Para obter esses resultados experimentais, o jogo pode ser jogado por vários pares de alunos e o professor pede aos alunos para irem registando num gráfico, desenhado no quadro, os resultados dos sucessivos lançamentos das 3 moedas. Depois de alguns jogos, o gráfico pode apresentar o seguinte aspecto:

<sup>14</sup> Sugerido pelo Sticks and Stones, Illuminations, [www.illuminations.nctm.org](http://www.illuminations.nctm.org)



O professor pede aos alunos para compararem as alturas das barras do gráfico e os alunos podem concluir que a ocorrência de 2 faces Euro ou uma face Euro são mais prováveis que a ocorrência de 3 faces ou 0 faces Euro. A partir do gráfico o professor pede ainda para os alunos estimarem a probabilidade de saírem 3 faces Euro quando se lançam as 3 moedas. No total das 31 ocorrências, verificou-se 4 vezes a saída de 3 faces Euro, pelo que uma estimativa para a probabilidade pretendida é  $\frac{4}{31}$  ou seja aproximadamente 12,9%. De modo análogo obtém-se como estimativa para a probabilidade da ocorrência de 2 faces Euro um valor aproximado a 38,7%.

De seguida o professor propõe aos alunos que considerem o modelo de probabilidade associado à experiência que consiste em lançar 3 moedas de Euro e verificar as faces que ficam viradas para cima. A construção do modelo de probabilidade pressupõe que se obtenham todos os resultados possíveis e que depois se atribua uma probabilidade a cada um desses resultados, sem esquecer que a soma das probabilidades de todos os resultados tem que ser igual a 1. Para obter todos os resultados possíveis pode-se utilizar um diagrama em árvore, como o que se apresenta a seguir:



O espaço de resultados  $S$  é constituído pelos seguintes resultados

$$S = (EEE, EEN, ENE, ENN, NEE, NEN, NNE, NNN)$$

Admitindo que a moeda é equilibrada, todos os resultados anteriores têm igual possibilidade de se verificarem, pelo que a probabilidade de cada um é  $1/8$ :

Resultado	EEE	EEN	ENE	ENN	NEE	NEN	NNE	NNN
Probabilidade	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$

Agora os alunos podem comparar as probabilidades experimentais calculadas anteriormente, com as probabilidades teóricas calculadas a partir deste modelo. Para já, concluem que a probabilidade de sair 3 faces Euro é igual à probabilidade de sair 0 faces Euro e igual a  $1/8$  ou seja aproximadamente 12,5%. Quanto ao acontecimento "2 faces Euro", este é constituído por 3 resultados, pelo que a probabilidade de se realizar é  $3/8$  ou seja 37,5%. O acontecimento "1 face Euro" também tem probabilidade  $3/8$ , pois também é constituído por três resultados elementares.

Finalmente o professor ainda pode colocar a seguinte questão aos alunos: em média, quantas jogadas são necessárias para terminar um jogo?

Uma estimativa deste valor pode ser obtida fazendo vários jogos, registando os resultados e fazendo a média dos resultados obtidos. No entanto podemos ainda abordar esta questão através do seguinte raciocínio: tendo em conta o modelo de probabilidade anterior, espera-se que em 8 jogadas saia 1 vez 3 faces Euro, 3 vezes 2 faces Euro, 3 vezes 1 face Euro e finalmente 1 vez 0 faces Euro, o que faz com que em 8 jogadas (considera-se uma jogada o lançamento das 3 moedas e a movimentação do pino de acordo com o resultado do lançamento) se espera movimentar o seguinte número de quadradinhos:

$$1 \times 10 + 3 \times 3 + 3 \times 1 + 1 \times 5 = 27$$

Assim, se em 8 jogadas se espera movimentar 27 quadradinhos, em média por jogada espera-se movimentar  $27/8=3,375$  quadradinhos. Então, para dar a volta aos 40 quadrados espera-se fazer  $40/3,375 \approx 12$  jogadas. Como estão dois jogadores em jogo, espera-se terminar o jogo com um número de jogadas à volta de 24. Se algum jogador tiver de voltar atrás terá de haver mais jogadas para terminar o jogo!

**Tarefa – Quem é que ganha o jogo?**<sup>15</sup> Na escola o professor propôs o seguinte jogo a ser jogado por dois alunos, o David e o António: lançam-se duas moedas e em cada lançamento, se saírem faces diferentes, o David ganha um ponto; caso contrário ganha o António o ponto. Ganha o jogo, aquele que, ao fim de 50 jogadas tiver ganho mais pontos. Quem é que ganhará o jogo?

Apresentamos a seguir o resultado do lançamento de 2 moedas equilibradas:

---

<sup>15</sup> Graça Martins et al (2007).

Jogada	1ª moeda	2ª moeda	António ganha	David ganha	Pontos António	Pontos David
1	N	N	1	0	1	0
2	E	E	1	0	2	0
3	E	N	0	1	2	1
4	N	N	1	0	3	1
5	E	N	0	1	3	2
6	N	N	1	0	4	2
7	E	E	1	0	5	2
8	E	N	0	1	5	3
9	E	E	1	0	6	3
10	E	E	1	0	7	3
11	E	N	0	1	7	4
12	N	E	0	1	7	5
13	N	N	1	0	8	5
14	E	N	0	1	8	6
15	N	N	1	0	9	6
16	E	N	0	1	9	7
17	E	N	0	1	9	8
18	E	E	1	0	10	8
19	N	E	0	1	10	9
20	N	N	1	0	11	9
21	N	E	0	1	11	10
22	N	E	0	1	11	11
23	N	E	0	1	11	12
24	N	N	1	0	12	12
25	N	N	1	0	13	12
26	E	E	1	0	14	12
27	E	E	1	0	15	12
28	N	N	1	0	16	12
29	N	E	0	1	16	13
30	E	N	0	1	16	14
31	N	E	0	1	16	15
32	E	E	1	0	17	15
33	E	N	0	1	17	16
34	N	E	0	1	17	17
35	E	E	1	0	18	17
36	E	N	0	1	18	18
37	N	E	0	1	18	19
38	E	E	1	0	19	19
39	E	N	0	1	19	20
40	E	N	0	1	19	21
41	E	E	1	0	20	21
42	N	E	0	1	20	22
43	E	E	1	0	21	22
44	E	E	1	0	22	22
45	N	E	0	1	22	23
46	N	N	1	0	23	23
47	E	E	1	0	24	23
48	E	E	1	0	25	23
49	E	E	1	0	26	23
50	E	N	0	1	26	24

Neste jogo ganhou o António, pois ao fim de 50 jogadas tinha alcançado 26 pontos, enquanto que o David tinha 24 pontos. Resolveram jogar novamente o mesmo jogo, tendo obtido os resultados seguintes:

Jogada	1ª moeda	2ª moeda	António ganha	David ganha	Pontos António	Pontos David
1	E	E	1	0	1	0
2	E	E	1	0	2	0
3	N	N	1	0	3	0
4	N	N	1	0	4	0
5	E	N	0	1	4	1
6	N	E	0	1	4	2
7	E	N	0	1	4	3
8	E	N	0	1	4	4
9	N	E	0	1	4	5
10	N	E	0	1	4	6
11	N	N	1	0	5	6
12	N	E	0	1	5	7
13	N	E	0	1	5	8
14	E	E	1	0	6	8
15	E	N	0	1	6	9
16	N	N	1	0	7	9
17	N	N	1	0	8	9
18	E	N	0	1	8	10
19	N	E	0	1	8	11
20	N	E	0	1	8	12
21	N	N	1	0	9	12
22	N	N	1	0	10	12
23	N	E	0	1	10	13
24	E	E	1	0	11	13
25	E	E	1	0	12	13
26	N	N	1	0	13	13
27	N	E	0	1	13	14
28	N	N	1	0	14	14
29	N	E	0	1	14	15
30	E	E	1	0	15	15
31	E	E	1	0	16	15
32	N	N	1	0	17	15
33	E	E	1	0	18	15
34	N	N	1	0	19	15
35	N	E	0	1	19	16
36	E	E	1	0	20	16
37	N	E	0	1	20	17
38	E	E	1	0	21	17
39	N	E	0	1	21	18
40	E	N	0	1	21	19
41	E	N	0	1	21	20
42	E	N	0	1	21	21
43	E	N	0	1	21	22
44	N	E	0	1	21	23
45	E	E	1	0	22	23
46	E	N	0	1	22	24
47	N	E	0	1	22	25
48	N	E	0	1	22	26
49	N	N	1	0	23	26
50	E	E	1	0	24	26

Desta vez ganhou o David! Resolveram fazer ainda um 3º jogo para a desforra e obtiveram os seguintes resultados:

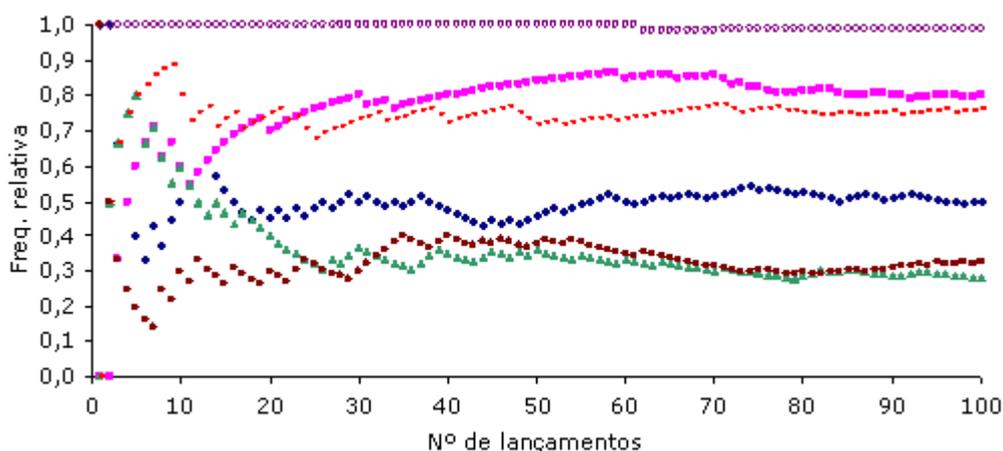
Jogada	1 <sup>a</sup> moeda	2 <sup>a</sup> moeda	António ganha	David ganha	Pontos António	Pontos David
1	N	N	1	0	1	0
2	E	N	0	1	1	1
3	E	N	0	1	1	2
4	E	N	0	1	1	3
5	E	N	0	1	1	4
6	N	E	0	1	1	5
7	E	E	1	0	2	5
8	E	E	1	0	3	5
9	N	E	0	1	3	6
10	N	N	1	0	4	6
11	E	N	0	1	4	7
12	N	N	1	0	5	7
13	N	E	0	1	5	8
14	N	N	1	0	6	8
15	E	N	0	1	6	9
16	E	N	0	1	6	10
17	N	N	1	0	7	10
18	E	N	0	1	7	11
19	N	N	1	0	8	11
20	N	N	1	0	9	11
21	N	E	0	1	9	12
22	N	E	0	1	9	13
23	N	N	1	0	10	13
24	N	N	1	0	11	13
25	E	N	0	1	11	14
26	E	E	1	0	12	14
27	E	E	1	0	13	14
28	N	E	0	1	13	15
29	N	N	1	0	14	15
30	E	E	1	0	15	15
31	E	N	0	1	15	16
32	E	E	1	0	16	16
33	N	N	1	0	17	16
34	N	N	1	0	18	16
35	N	E	0	1	18	17
36	N	E	0	1	18	18
37	E	E	1	0	19	18
38	E	N	0	1	19	19
39	E	N	0	1	19	20
40	E	E	1	0	20	20
41	E	N	0	1	20	21
42	N	N	1	0	21	21
43	E	N	0	1	21	22
44	N	N	1	0	22	22
45	E	E	1	0	23	22
46	E	N	0	1	23	23
47	E	N	0	1	23	24
48	N	E	0	1	23	25
49	N	N	1	0	24	25
50	N	N	1	0	25	25

Agora empataram! Afinal não se pode dizer à partida quem é que sairá o vencedor, pois qualquer um dos dois alunos tem igual “chance” de ganhar o jogo.

Uma versão deste exemplo pode ser realizado na sala de aula, em que o professor desenha no quadro um trajecto que será percorrido pelos alunos que estão a jogar,

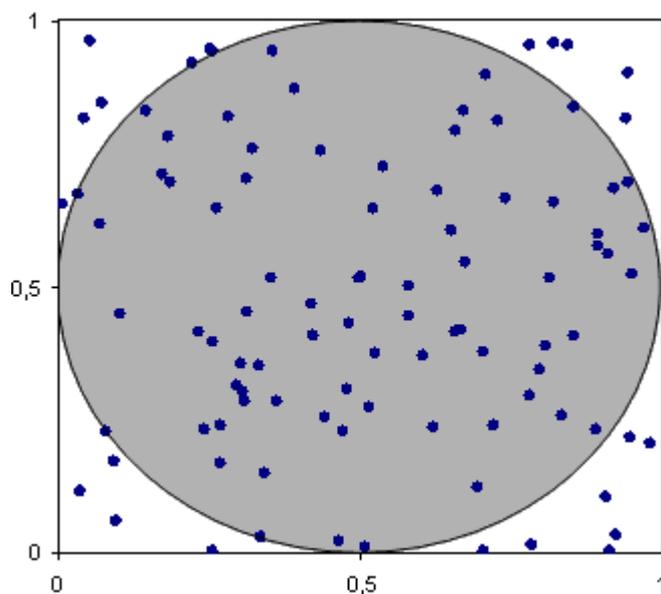


- e) Depois dos 50 lançamentos, estar-se-á razoavelmente seguro que as moedas estão correctamente identificadas? Explique porquê.
- f) O seguinte gráfico mostra a evolução da frequência relativa para as 6 moedas, à medida que o número de lançamentos aumenta:



Comente o que é que este gráfico revela sobre a probabilidade, como um conceito sobre o comportamento de um processo aleatório a longo-termo e não a curto-termo.

**Tarefa – Vamos estimar a área do círculo com raio 0,5 unidades.** O professor apresentou aos alunos um gráfico, onde estava desenhada uma circunferência dentro de um quadrado de lado 1. Para o espaço delimitado pelo quadrado um atirador muito “nabo”<sup>17</sup> atirou, ao acaso, 100 setas, que ficaram marcadas, como se vê na figura seguinte:



<sup>17</sup> O atirador era tão “nabo” e não tinha pontaria nenhuma, pelo que a seta podia acertar, ao acaso, em qualquer ponto do quadrado.

Algumas das setas caíram dentro da circunferência e outras fora. O professor propôs aos alunos estimarem a área do círculo com base na distribuição das marcas dentro do quadrado.

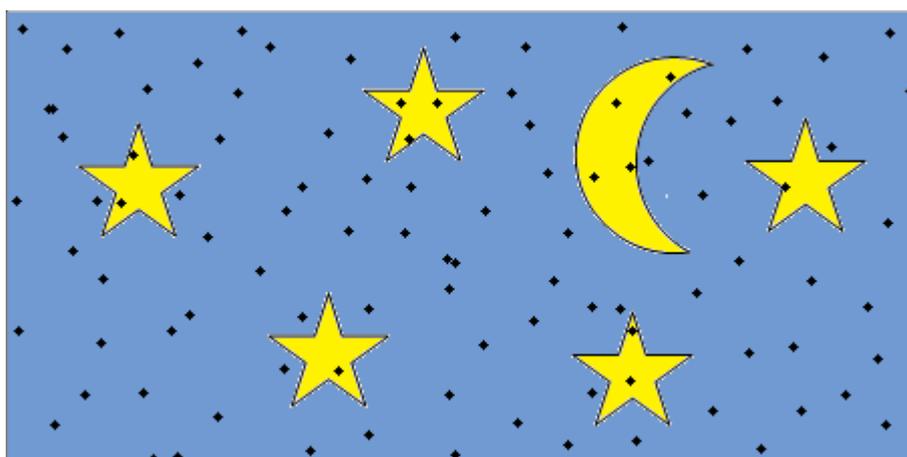
Se as marcas deixadas pelas setas se distribuem aleatoriamente pela superfície do quadrado, então a proporção de marcas dentro do quadrado dão-nos uma estimativa da área do círculo, já que a área do quadrado é 1. Assim, contando o número de marcas dentro da circunferência (neste caso é mais fácil contar as 22 marcas fora da circunferência), verificamos que essa proporção é  $(100-22)/100=0,78$ . Este é o valor estimado para a área do círculo.

Se os alunos já souberem calcular a área do círculo será interessante compararem o valor estimado com o valor calculado através da fórmula da área.

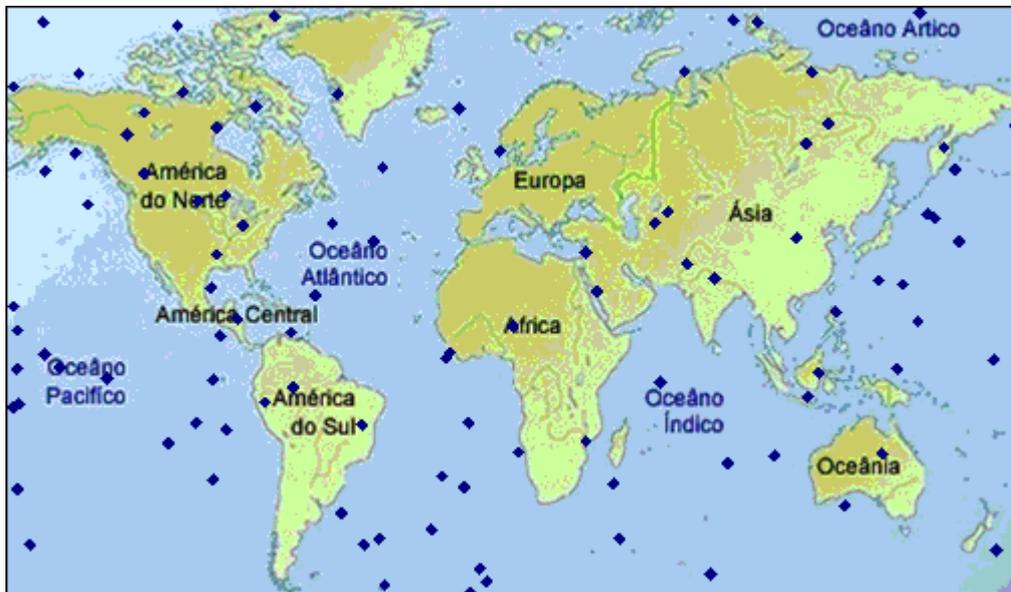
**Tarefa – Vamos estimar o valor de  $\pi$ .** O trabalho desenvolvido na tarefa anterior pode ser utilizado para estimar o valor de  $\pi$ . Para isso basta admitir que se conhece que a área de um círculo de raio  $R$  é igual a  $\pi R^2$ . Então, como vimos na tarefa anterior, uma estimativa para a área do círculo de raio 0,5 unidades é 0,78, pelo que uma estimativa para o valor de  $\pi$  será  $0,78/0,5^2=3,12$ .

Obter-se-ia uma estimativa mais precisa para o valor de  $\pi$  se em vez de 100 pontos, aumentássemos o número de pontos que preenchem a área do quadrado, pois neste caso também se obteria uma estimativa mais precisa para a área do círculo. No capítulo seguinte sobre simulação, será apresentado um processo de gerar aleatoriamente os pontos que simulam as marcas das setas do atirador.

**Tarefa – Estimar a área de figuras.** O processo utilizado para estimar a área do círculo, pode ser utilizado para estimar a área de uma figura para a qual não exista uma expressão simples que permita calcular a sua área. O professor apresentou aos alunos uma fotografia, com área igual a uma unidade, de um pedaço de “céu”, onde estão desenhadas algumas estrelas e a lua, em quarto minguante. Nessa fotografia foram desenhados, aleatoriamente 100 pontos. A proposta do professor é que os alunos estimem a área dos corpos celestes da fotografia.



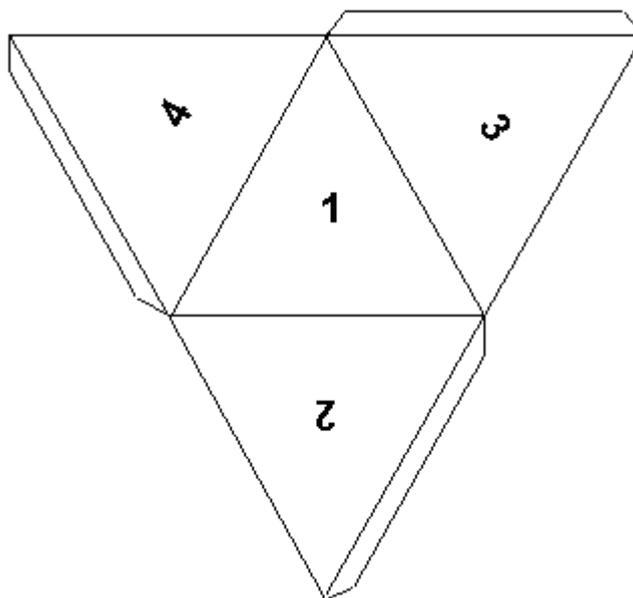
**Tarefa – Uma chuva de meteoritos!** O professor apresentou uma figura, que representa o planisfério, onde estão registadas as marcas de 100 meteoritos que alcançaram o planeta Terra. Alguns caíram no mar, outros em terra. Baseando-te nos 100 meteoritos que se registam na figura, és capaz de estimar a probabilidade de o próximo meteorito cair em terra? (A área do rectângulo é igual a 1 unidade de área)



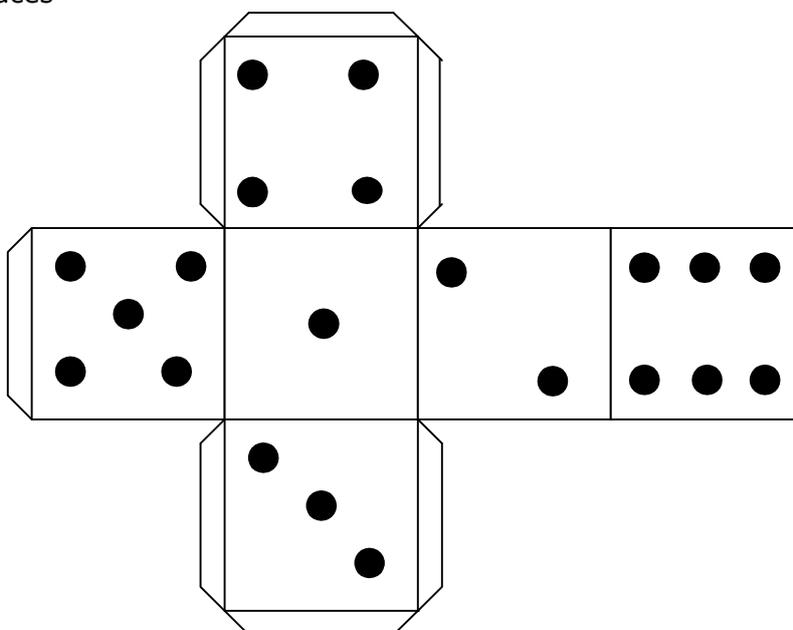
Dos 100 meteoritos, que se distribuem aleatoriamente sobre a superfície terrestre, cerca de 30 (a contagem não é muito fácil...) alcançaram algum continente, pelo que estimamos que com uma probabilidade de 30%, o próximo meteorito alcance algum continente.

**Tarefa especial – Vamos construir alguns dos nossos materiais para fazer experiências.** Propomos aqui uma tarefa especial, que o professor, juntamente com o professor de outra disciplina, nomeadamente de Educação Artística ou Visual, pode levar a cabo juntamente com os alunos: construir os seus próprios materiais. Como fomos indicando ao longo do texto, podem ser utilizados materiais como botões, tampas de refrigerantes, berlindes, etc. As moedas utilizadas nas experiências podem ser realizadas pelos alunos em cartolina dura, em que de um lado desenham um **N**, para indicar a face Nacional e do outro o símbolo do euro €. Propomos aqui a construção de um dado de 4 faces (tetraedro) e de um dado de 6 faces.

Dado de 4 faces



Dado de 6 faces





# 7 Simulação

Neste capítulo fazemos uma introdução à Simulação, instrumento poderoso, que sobretudo nas três últimas décadas, com o desenvolvimento e aperfeiçoamento dos meios computacionais, contribuiu de forma decisiva para o estudo de leis da probabilidade e cálculo de probabilidades associadas a determinados acontecimentos. No capítulo anterior, vimos que um processo de estimar a probabilidade de um acontecimento, seria repetir muitas vezes a experiência e contabilizar a proporção de vezes que o acontecimento se realiza nas sucessivas repetições. Veremos como, utilizando meios computacionais, quer seja a máquina de calcular, quer seja o computador, podemos imitar o comportamento da realização do fenómeno aleatório.



## 7.1 Introdução

Vimos no capítulo anterior que a definição de um modelo de probabilidade, associado a um fenómeno aleatório, pressupõe que sejam especificados todos os resultados possíveis e que a cada resultado possível seja atribuído uma probabilidade. Muitas vezes estas probabilidades são atribuídas tendo em conta a experiência que temos sobre a realização de fenómenos do mesmo género. É o que se passa com o modelo associado aos fenómenos aleatórios que consistem em lançar uma moeda ou um dado e verificar qual a face que fica virada para cima. Se não tivermos razões que nos levem a suspeitar que a moeda ou o dado não são equilibrados, consideramos os modelos, usuais, em que atribuímos igual probabilidade a cada uma das faces, quer da moeda, quer do dado.

No entanto, mesmo neste caso em que é fácil definir um modelo para o fenómeno aleatório, por vezes não é fácil calcular probabilidades de acontecimentos relacionados com esse fenómeno. Por exemplo, se lançarmos 10 vezes uma moeda de um euro, equilibrada, o cálculo teórico da probabilidade do acontecimento "obter 4 ou mais faces Euro ou Nacional seguidas" não é acessível a este nível. Então a única solução seria repetir muitas vezes a experiência de lançar a moeda 10 vezes e estimar a probabilidade do acontecimento, pela proporção de vezes em que a face Euro ou a face Nacional aparece 4 ou mais vezes seguidas, em sequências de 10 lançamentos.

Como é fácil de entender, estar a repetir a experiência de lançar a moeda 10 vezes, ver o que acontece, outras 10 vezes e ver o que acontece, e assim por diante..., não é uma tarefa simples. Esta situação não deixa os estatísticos muito preocupados, pois o comportamento aleatório do lançamento da moeda pode ser imitado, utilizando a tecnologia, e neste caso dizemos que estamos a simular a realização do fenómeno.

*Simulação* – processo artificial utilizado para imitar o comportamento de um fenómeno aleatório, utilizando, de um modo geral, números aleatórios.

Veremos na secção seguinte o que se entende por números aleatórios e veremos ainda que, hoje em dia, o que se utiliza são os números pseudo-aleatórios, gerados pela máquina de calcular ou pelo computador.

## 7.2 Números aleatórios e números pseudo-aleatórios

Considere um saco com 10 berlines, iguais ao tacto, e numerados de 0 a 9. Depois de baralhar os berlines dentro do saco, seleccione um ao acaso, verifique o número do berline seleccionado e reponha-o no saco. Se repetirmos o processo várias vezes, poderemos obter uma sequência de dígitos como a que se apresenta a seguir:

7 1 3 1 9 2 2 0 3 4 0 5 7 5 6 2 8 3 9 5 ...

Em cada selecção de um berline do saco, temos igual probabilidade de seleccionar cada um dos 10 dígitos 0, 1, ..., 9. O berline seleccionado em cada extracção não depende dos berlines seleccionados nas extracções anteriores, pelo que os resultados são independentes uns dos outros. Uma tabela com dígitos obtidos por este processo diz-se que é uma tabela de dígitos aleatórios.

**Dígitos aleatórios** - Uma tabela de dígitos aleatórios é uma listagem dos dígitos 0, 1, 2, 3, 4, 5, 6, 7, 8 ou 9 tal que:

- qualquer um dos dígitos considerados tem igual possibilidade de figurar em qualquer posição da lista;
- a posição em que figura cada dígito é independente das posições dos outros dígitos.

Apresenta-se a seguir um extracto de uma tabela de números aleatórios (Moore, 1997). O facto de os dígitos se apresentarem agrupados 5 a 5 é só para facilidade de leitura.

---

**Linha**

---

<b>101</b>	19223	95034	05756	28713	96409	12531	42544	82853
<b>102</b>	73676	47150	99400	01927	27754	42648	82425	36290
<b>103</b>	45467	71709	77558	00095	32863	29485	82226	90056
<b>104</b>	52711	38889	93074	60227	40011	85848	48767	52573
<b>105</b>	95592	94007	69971	91481	60779	53791	17297	59335
<b>106</b>	68417	35013	15529	72765	85089	57067	50211	47487
<b>107</b>	82739	57890	20807	47511	81676	55300	94383	14893
<b>108</b>	60940	72024	17868	24943	61790	90656	87964	18883
<b>109</b>	36009	19365	15412	39638	85453	46816	83485	41979

---

A partir da tabela de dígitos aleatórios podem-se obter números aleatórios de 2 dígitos - qualquer par dos 100 pares possíveis 00, 01, ...98, 99, tem igual probabilidade de ser seleccionado, de 3 dígitos - qualquer triplo dos 1000 triplos possíveis 000, 001, ...998, 999, tem igual probabilidade de ser seleccionado, etc., tomando os dígitos da tabela 2 a 2, 3 a 3, etc., a partir de uma linha qualquer e percorrendo-a da esquerda para a direita.

A apresentação de uma tabela como a anterior tem mais interesse histórico do que interesse real. Efectivamente, hoje em dia, existe a possibilidade de utilizar o computador ou uma simples máquina de calcular para gerar os números aleatórios. No entanto, convém ter presente que os números que se obtêm são *pseudo-aleatórios*, já que é um mecanismo determinista que lhes dá origem, embora se comportem como números aleatórios (passam numa bateria de testes destinados a confirmar a sua aleatoriedade). Não obstante hoje em dia ser mais comum a utilização da tecnologia para fazer simulações, vamos na secção seguinte exemplificar o uso de uma tabela e dígitos aleatórios num processo de simulação.

### **A função RAND do Excel ou da máquina de calcular**

Mais geralmente, quando falamos em números aleatórios, sem qualquer outra referência, não nos estamos a referir explicitamente a números inteiros, mas sim a números do intervalo [0, 1]. Os algoritmos de geração de números pseudo-aleatórios estão concebidos de modo a que ao considerar uma qualquer sequência de números gerados se obtenha aproximadamente a mesma proporção de observações em sub intervalos de igual amplitude do intervalo [0,1]. Assim, por exemplo, se se fizer correr o algoritmo 100 vezes, é de esperar que caiam 25 dos números gerados em cada quarto do intervalo [0,1]. Na tabela seguinte está listada

uma sequência de 100 números pseudo-aleatórios (NPA) obtida através do gerador RAND do Excel<sup>1</sup>:

1ª coluna	2ª coluna	3ª coluna	4ª coluna
0,842050	0,406320	0,848744	0,810469
0,965131	0,676239	0,722927	0,825587
0,761648	0,552387	0,079614	0,298300
0,359825	0,208420	0,098150	0,818893
0,054705	0,102768	0,147229	0,557920
0,466613	0,493374	0,150888	0,540352
0,814300	0,638416	0,086141	0,007840
0,449515	0,090759	0,197460	0,209145
0,901502	0,552418	0,466389	0,221584
0,862762	0,507097	0,613583	0,389183
0,395195	0,415666	0,210044	0,379011
0,420519	0,469764	0,053714	0,478208
0,124664	0,765629	0,737348	0,696311
0,537707	0,451921	0,702749	0,683382
0,033277	0,523063	0,908485	0,708764
0,024371	0,213326	0,442821	0,983754
0,558313	0,283191	0,153907	0,655705
0,087859	0,429387	0,735276	0,890680
0,069915	0,221549	0,358037	0,578713
0,774156	0,039495	0,490216	0,755072
0,789583	0,480287	0,302539	0,970551
0,702971	0,109918	0,444822	0,995760
0,087455	0,713230	0,806147	0,569285
0,103532	0,623757	0,377823	0,161851
0,996667	0,129629	0,196290	0,753139

Como se pode verificar por contagem, esta lista inclui 30 números no intervalo  $[0;0,25]$ , 24 números nos intervalos  $]0,25;0,5]$  e  $]0,5;0,75]$  e 22 números no intervalo  $]0,75;1]$ . Embora haja métodos estatísticos para avaliar se são ou não significativas as diferenças entre estas frequências observadas e as frequências esperadas ( $25 - 25 - 25 - 25$ ), facilmente a nossa sensibilidade aceita que estes resultados não contradizem o que se esperaria de uma escolha ao acaso de 100 números do intervalo  $[0,1]$ .

Uma tabela idêntica à anterior poderia ter sido obtida a partir da função RAND, na máquina de calcular.

### **A função RANDBETWEEN do Excel ou a função randINT da máquina de calcular**

No caso particular de pretendermos números inteiros, então podemos utilizar a função Randbetween( $m, n$ ) do Excel ou a função randINT( $m, n$ ) da máquina de calcular. Estas funções que têm como argumentos dois números inteiros  $m$  e  $n$ , com  $m < n$ , sempre que são utilizadas devolvem-nos um número inteiro entre  $m$  e  $n$ , tal que qualquer outro número inteiro entre esses limites tem igual probabilidade de surgir. Por exemplo, para simular o lançamento de um dado equilibrado,

<sup>1</sup> Graça Martins et al (2007) e Anexo para interpretação do programa de MACS, pag 93

utilizamos a função `randINT(1,6)` da máquina de calcular, ou a função `Randbetween(1, 6)` do Excel.

### 7.3 Simular o lançamento de uma moeda

A simulação do lançamento de uma moeda, embora seja um problema simples, está na base do cálculo da probabilidade, por simulação, de alguns acontecimentos, cujo cálculo analítico seria complicado. Permite ainda ilustrar alguns conceitos básicos de probabilidade que, por vezes, fogem à nossa intuição.

É exemplo do que dissemos anteriormente a situação que diz respeito à regularidade a longo termo e que já foi abordada na secção 6.2 e 6.5.2.2 e que abordaremos de novo nesta secção.

O processo de simulação pressupõe alguns passos que exemplificaremos de seguida:

*Passo 1 – Definição do modelo de probabilidade para o lançamento da moeda*

No modelo que vamos adoptar para o lançamento da moeda vamos admitir que:

- ✓ Em cada lançamento existe igual probabilidade de sair a face Euro e a face Nacional (admitimos que a moeda é equilibrada);
- ✓ Os lançamentos são independentes uns dos outros.

*Passo 2 – Atribuição de números aleatórios para representar os resultados do lançamento da moeda.*

Esta atribuição pode ser feita de várias maneiras, das quais vamos exemplificar duas, tendo em conta as tabelas de dígitos aleatórios e de números pseudo-aleatórios consideradas anteriormente.

Considerando a tabela de dígitos aleatórios, sabemos que cada um dos dígitos 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9, tem probabilidade igual a  $1/10$  ou 0,1 de surgir e que além disso os sucessivos dígitos da tabela são independentes. Então uma forma possível de atribuir dígitos ao resultado do lançamento da moeda, é considerar que:

- ✓ Cada dígito simula o resultado do lançamento da moeda;
- ✓ Dígitos ímpares representam a face Euro e dígitos pares a face Nacional (estamos a admitir que o zero é par).

Este processo de atribuir probabilidades está de acordo com o modelo proposto, já que esta atribuição dá à saída de face Euro uma probabilidade igual a  $5/10$  (5 favoráveis em 10 possíveis), e além disso os dígitos sucessivos da tabela simulam lançamentos independentes.

Nota: Outro processo possível seria considerar os dígitos menores que 5 (ou seja, 0, 1, 2, 3 e 4) para representarem a face Euro e os maiores ou iguais a 5 (ou seja 5, 6, 7, 8 e 9), para representarem a face Nacional.

*Passo 3 – Simular muitas repetições*

Cada dígito da tabela simula um lançamento da moeda, pelo que para simular os sucessivos lançamentos basta considerar os dígitos sucessivos da tabela.

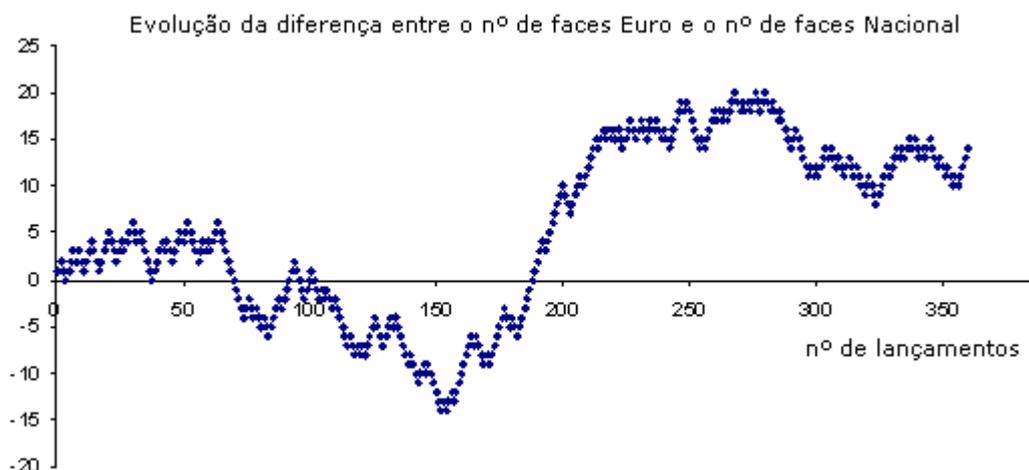
Na simulação que apresentamos, vamos considerar que se pretende estimar a probabilidade do acontecimento "sair face Euro no lançamento da moeda" e vamos ainda ver o que acontece à diferença entre o número de faces Euro e faces Nacional, à medida que aumentamos o número de lançamentos:

Nº de lançamento i	Dígito	Face	Nº de faces Euro até ao lançamento i	Proporção de faces Euro até ao lançamento i	Diferença entre nº faces Euro e nº faces Nacional
1	1	E	1	1	1
2	9	E	2	1	2
3	2	N	2	0,6666667	1
4	2	N	2	0,5	0
5	3	E	3	0,6	1
6	9	E	4	0,6666667	2
7	5	E	5	0,7142857	3
8	0	N	5	0,625	2
9	3	E	6	0,6666667	3
10	4	N	6	0,6	2
11	0	N	6	0,5454545	1
12	5	E	7	0,5833333	2
13	7	E	8	0,6153846	3
14	5	E	9	0,6428571	4
15	6	N	9	0,6	3
16	2	N	9	0,5625	2
17	8	N	9	0,5294118	1
18	7	E	10	0,5555556	2
19	1	E	11	0,5789474	3
20	3	E	12	0,6	4
21	9	E	13	0,6190476	5
22	6	N	13	0,5909091	4
23	4	N	13	0,5652174	3
24	0	N	13	0,5416667	2
25	9	E	14	0,56	3
26	1	E	15	0,5769231	4
27	2	N	15	0,5555556	3
28	5	E	16	0,5714286	4
29	3	E	17	0,5862069	5
30	1	E	18	0,6	6
31	4	N	18	0,5806452	5
32	2	N	18	0,5625	4
33	5	E	19	0,5757576	5
34	4	N	19	0,5588235	4
...					
346	4	N	180	0,5202312	14
347	6	N	180	0,518732	13
348	8	N	180	0,5172414	12
349	1	E	181	0,5186246	13
350	6	N	181	0,5171429	12
351	8	N	181	0,5156695	11
352	3	E	182	0,5170455	12
353	4	N	182	0,5155807	11
354	8	N	182	0,5141243	10
355	5	E	183	0,515493	11
356	4	N	183	0,5140449	10
357	1	E	184	0,5154062	11
358	9	E	185	0,5167598	12
359	7	E	186	0,5181058	13
360	9	E	187	0,5194444	14

Da tabela anterior e do gráfico seguinte, concluímos que a proporção de faces Euro tende a estabilizar à volta dos 52%, para o número de lançamentos realizados. Se tivéssemos levado a simulação mais longe, seria de esperar que a proporção de faces Euro se aproximasse mais de 50%.



No gráfico seguinte apresentamos a evolução da diferença entre o número de faces Euro e o número de faces Nacional. Ao contrário do que se passa com a proporção de faces Euro que se aproxima de 50%, à medida que o número de lançamentos aumenta, o número de faces Euro não se aproxima de metade dos lançamentos, e a diferença entre o número de faces Euro e faces Nacional não tende a estabilizar à volta de zero.



Nota – É importante ter presente que os números pseudo-aleatórios utilizados no processo de simulação, são gerados por um mecanismo determinístico, que imita razoavelmente bem o aleatório. Há vários processos para gerar estes números, nomeadamente a função Rand do Excel, de que já falámos anteriormente. Todos estes processos foram estudados de forma que os conjuntos de números que geram, passam num conjunto de testes estatísticos, que não rejeitam a hipótese desses conjuntos de números poderem ser considerados como aleatórios.

## 7.4 Cálculo da probabilidade de acontecimentos, por simulação

Nesta secção vamos exemplificar o processo de estimar a probabilidade de alguns acontecimentos, por simulação.

**Tarefa – Calcular a probabilidade de sair 2 faces Euro, em 2 lançamentos de uma moeda de um euro, equilibrada.** Os dois primeiros passos neste processo de simulação são idênticos aos considerados no lançamento da moeda equilibrada. Dois dígitos sucessivos da tabela simulam 2 lançamentos, pelo que considerando vários conjuntos de dois dígitos simulamos várias repetições. Iniciando a consulta da tabela na linha 102, temos:

Repetição	Dígitos	Euro/Nacional	Repetição	Dígitos	Euro/Nacional
1	73	<b>Euro Euro</b>	26	77	<b>Euro Euro</b>
2	67	Nacional Euro	27	55	<b>Euro Euro</b>
3	64	Nacional Nacional	28	80	Nacional Nacional
4	71	<b>Euro Euro</b>	29	00	Nacional Nacional
5	50	Euro Nacional	30	95	<b>Euro Euro</b>
6	99	<b>Euro Euro</b>	31	32	Nacional Nacional
7	40	Nacional Nacional	32	86	Nacional Nacional
8	00	Nacional Nacional	33	32	Nacional Nacional
9	19	<b>Euro Euro</b>	34	94	Euro Nacional
10	27	Nacioanl Euro	35	85	Nacional Euro
11	27	Nacioanl Euro	36	82	Nacional Nacional
12	75	<b>Euro Euro</b>	37	22	Nacional Nacional
13	44	Nacional Nacional	38	69	Nacional Euro
14	26	Nacional Nacional	39	00	Nacional Nacional
15	48	Nacional Nacional	40	56	Euro Nacional
16	82	Nacional Nacional	41	52	Euro Nacional
17	42	Naciona Nacional	42	71	<b>Euro Euro</b>
18	53	<b>Euro Euro</b>	43	13	<b>Euro Euro</b>
19	62	Nacional Nacional	44	88	Nacional Nacional
20	90	Euro Nacioal	45	89	Nacional Euro
21	45	Nacional Euro	46	93	<b>Euro Euro</b>
22	46	Nacional Nacional	47	07	Nacional Euro
23	77	<b>Euro Euro</b>	48	46	Naciona Nacional
24	17	<b>Euro Euro</b>	49	02	Naciona Nacional
25	09	Nacional Euro	50	27	Nacional Euro

Assinalámos a preto o acontecimento de interesse, que era a saída de 2 faces Euro. Uma estimativa para a probabilidade desse acontecimento é 14/50 ou seja 0,28. As 50 repetições realizadas não são suficientes para estarmos confiantes na precisão da estimativa obtida para a probabilidade do acontecimentos “duas faces Euro no lançamento de uma moeda duas vezes”. Um maior número de repetições conduzir-nos-ia a uma probabilidade mais próxima de 0,25, que é o resultado teórico para a probabilidade desse acontecimento.

A atribuição das probabilidades no passo 2 pode ser feita utilizando a tabela de números pseudo-aleatórios da página 253.

*Passo 2' – Considerando então essa tabela, vamos admitir que:*

- ✓ Cada número simula o resultado do lançamento da moeda;
- ✓ Um número  $\leq 0,5$  representa a face Euro e um número  $> 0,5$  representa a face Nacional.

Esta atribuição de probabilidades está de acordo como modelo proposto, já que os intervalos  $[0, 0,5]$  e  $]0,5, 1]$  têm igual amplitude, pelo que a probabilidade de obter números em cada um desses intervalos é 0,5.

*Passo 3'* – Para simular as várias repetições atribuímos os números pseudo-aleatórios aos resultados face Euro e face Nacional percorrendo as colunas 1 e 2 para as primeiras 25 repetições e de seguida as colunas 3 e 4 para as 25 repetições seguintes:

Repetição	Euro/Nacional	Repetição	Euro/Nacionalo
1	Nacional Euro	26	Nacional Nacional
2	Nacional Nacional	27	Nacional Nacional
3	Nacional Nacional	28	<b>Euro Euro</b>
4	<b>Euro Euro</b>	29	Euro Nacional
5	<b>Euro Euro</b>	30	Euro Nacional
6	<b>Euro Euro</b>	31	Euro Nacional
7	Nacional Nacional	32	<b>Euro Euro</b>
8	<b>Euro Euro</b>	33	<b>Euro Euro</b>
9	Nacional Nacional	34	<b>Euro Euro</b>
10	Nacional Nacional	35	Nacional Euro
11	<b>Euro Euro</b>	36	<b>Euro Euro</b>
12	<b>Euro Euro</b>	37	<b>Euro Euro</b>
13	Euro Nacional	38	Nacional Nacional
14	Nacional Euro	39	Nacional Nacional
15	Euro Nacional	40	Nacional Nacional
16	<b>Euro Euro</b>	41	Euro Nacional
17	Nacional Euro	42	Euro Nacional
18	<b>Euro Euro</b>	43	Nacional Nacional
19	<b>Euro Euro</b>	44	Euro Nacional
20	Nacional Euro	45	Euro Nacional
21	Nacional Euro	46	Euro Nacional
22	Nacional Euro	47	Euro Nacional
23	Euro Nacional	48	Nacional Nacional
24	Euro Nacional	49	<b>Euro Euro</b>
25	Nacional Euro	50	Euro Nacional

A simulação anterior conduziu-nos a uma estimativa da probabilidade de 0,32.

Quaisquer outros 50 números aleatórios ou pseudo-aleatórios conduzir-nos-iam a outras estimativas para a probabilidade do acontecimento de interesse. Como, de um modo geral, não sabemos qual o valor da probabilidade teórica, se calcularmos várias estimativas, não saberemos qual a mais precisa, ou seja, qual a que está mais perto da probabilidade teórica, caso fosse possível calculá-la, tendo em conta o modelo de probabilidade adoptado para o fenómeno aleatório em estudo e com o qual baseámos a nossa simulação. Se pretendermos aumentar a confiança na estimativa da probabilidade que estamos a calcular, temos uma solução que é aumentar o número de simulações do fenómeno em estudo.

### **Não se fazem omeletas sem ovos...**

Como se lê em Moore (1996, p. 429) *“Parece um pouco duvidoso iniciar um processo para obter probabilidades assumindo que já conhecemos algumas outras probabilidades, mas nem mesmo na matemática temos alguma coisa sem dar nada em troca. A ideia é estabelecer a estrutura básica do fenómeno aleatório e então utilizar a simulação para passarmos deste modelo para obter probabilidades de acontecimentos mais complicados. O modelo é baseado em informação ou experiência passada. Se ele não descrever correctamente o fenómeno aleatório,*

*então as probabilidades obtidas a partir dele por simulação, também estarão incorrectas” (tradução livre).*

Outra forma de exprimir a ideia do parágrafo anterior é dizer que “não há almoços grátis!” Como se exemplifica no caso anterior, ao estimarmos a probabilidade de obter 2 faces Euro nos 2 lançamentos da moeda, estamos a partir do princípio e a assumir que a probabilidade de sair face Euro é igual à de sair face Nacional e igual a  $\frac{1}{2}$ . Obviamente que não seria correcto utilizar o processo anterior para simular o lançamento de uma moeda enviesada, já que neste caso as duas faces não têm igual probabilidade de saírem, como foi assumido no modelo anterior. Nesta situação a única solução é proceder ao lançamento da moeda em estudo e ir registando os resultados obtidos, até conseguirmos que a frequência relativa, do acontecimento de que se pretende calcular a probabilidade, apresente uma certa estabilização.

**Tarefa – Qual a probabilidade de numa família de quatro filhos, todos serem rapazes?** O modelo para o nascimento de rapaz ou rapariga é idêntico ao do lançamento de uma moeda equilibrada, se assumirmos que existe igual probabilidade de nascer rapaz e rapariga e que os nascimentos são independentes uns dos outros (na realidade, de acordo com o nosso último censo, a probabilidade de nascer rapaz anda à volta de 51%). Assim, simular o nascimento de 4 crianças e ver o sexo, é idêntico a simular o lançamento de uma moeda de um Euro 4 vezes e verificar a face que fica voltada para cima.

*Passo 1 – Definição do modelo de probabilidade para o nascimento de rapaz*

O modelo que vamos adoptar para o nascimento de rapaz pressupõe que:

- ✓ Em cada nascimento existe igual probabilidade de ser rapaz ou rapariga;
- ✓ Os nascimentos são independentes uns dos outros.

*Passo 2 – Atribuição de números aleatórios para representar os resultados do nascimento.*

Vamos utilizar a função Rand do Excel, do seguinte modo:

- ✓ Cada número gerado simula o nascimento de uma criança;
- ✓ Um número  $\leq 0,5$  representa o nascimento de rapaz, e um número  $> 0,5$  representa o nascimento de rapariga.

Como vimos no exemplo anterior, esta atribuição de probabilidades está de acordo com o modelo proposto, já que os intervalos  $[0, 0,5]$  e  $]0,5, 1]$  têm igual amplitude, pelo que a probabilidade de obter números em cada um desses intervalos é 0,5.

*Passo 3 – Simular muitas repetições*

Vamos gerar muitas repetições de 4 números pseudo-aleatórios, para simular os 4 nascimentos:

1º filho	2º filho	3º filho	4º filho	1º filho rapaz?	2º filho rapaz?	3º filho rapaz?	4º filho rapaz?	Nº rapazes nos 4 filhos	4 rapazes ?
0,042293	0,832859	0,850776	0,415564	1	0	0	1	2	0
0,269279	0,881478	0,780493	0,102199	1	0	0	1	2	0
0,474479	0,528969	0,848406	0,660313	1	0	0	0	1	0
0,238322	0,804784	0,589457	0,413003	1	0	0	1	2	0
0,327561	0,785789	0,224090	0,626057	1	0	1	0	2	0
0,861051	0,018105	0,707192	0,016593	0	1	0	1	2	0

0,116059	0,553769	0,903991	0,378653	1	0	0	1	2	0
0,443542	0,586487	0,673580	0,278990	1	0	0	1	2	0
0,479738	0,498062	0,293571	0,465006	1	1	1	1	4	1
0,519160	0,836728	0,261501	0,709615	0	0	1	0	1	0
0,806497	0,986511	0,672229	0,206301	0	0	0	1	1	0
0,518291	0,922520	0,626536	0,738124	0	0	0	0	0	0
0,948191	0,176873	0,137029	0,291067	0	1	1	1	3	0
0,319196	0,034793	0,590844	0,564857	1	1	0	0	2	0
0,497591	0,035276	0,983357	0,613159	1	1	0	0	2	0
0,917203	0,195162	0,160754	0,571587	0	1	1	0	2	0
0,697895	0,876783	0,934845	0,620200	0	0	0	0	0	0
0,777700	0,048426	0,337187	0,179181	0	1	1	1	3	0
0,511748	0,767437	0,849113	0,344513	0	0	0	1	1	0
0,891218	0,307059	0,251028	0,616891	0	1	1	0	2	0
0,880875	0,865897	0,152129	0,109405	0	0	1	1	2	0
0,951888	0,908047	0,034045	0,863988	0	0	1	0	1	0
0,238392	0,082083	0,490909	0,833546	1	1	1	0	3	0
0,520766	0,698405	0,817629	0,333006	0	0	0	1	1	0
0,836980	0,148513	0,960522	0,880070	0	1	0	0	1	0
0,598769	0,606425	0,281036	0,831568	0	0	1	0	1	0
0,135802	0,722615	0,229893	0,230967	1	0	1	1	3	0
0,048072	0,493240	0,283100	0,083356	1	1	1	1	4	1
0,590230	0,316114	0,979242	0,060746	0	1	0	1	2	0
0,507317	0,237199	0,774544	0,043540	0	1	0	1	2	0
0,557046	0,632492	0,525553	0,493075	0	0	0	1	1	0
0,453047	0,003776	0,692767	0,223822	1	1	0	1	3	0
0,781730	0,146744	0,702416	0,335659	0	1	0	1	2	0
0,823291	0,370645	0,918915	0,884146	0	1	0	0	1	0
0,696797	0,234292	0,091790	0,292912	0	1	1	1	3	0
0,238624	0,979179	0,709441	0,435784	1	0	0	1	2	0
0,371197	0,458286	0,483376	0,845387	1	1	1	0	3	0
0,797437	0,994756	0,489628	0,648340	0	0	1	0	1	0
0,822567	0,405794	0,497899	0,794578	0	1	1	0	2	0
0,409935	0,784336	0,772683	0,789198	1	0	0	0	1	0
0,031521	0,634780	0,628484	0,776446	1	0	0	0	1	0
0,877285	0,818784	0,131070	0,377268	0	0	1	1	2	0
0,431466	0,779751	0,536371	0,544740	1	0	0	0	1	0
0,071561	0,715139	0,195817	0,236250	1	0	1	1	3	0
0,512900	0,636603	0,919911	0,200576	0	0	0	1	1	0
0,011701	0,826350	0,592355	0,106301	1	0	0	1	2	0
0,932498	0,230313	0,688338	0,940086	0	1	0	0	1	0
0,012260	0,731709	0,882629	0,291549	1	0	0	1	2	0
0,771633	0,282324	0,400765	0,751419	0	1	1	0	2	0
0,274627	0,720612	0,997822	0,886745	1	0	0	0	1	0
0,636812	0,257337	0,676606	0,341877	0	1	0	1	2	0
0,294693	0,785398	0,052230	0,067127	1	0	1	1	3	0
0,556242	0,331684	0,444164	0,391702	0	1	1	1	3	0
0,554021	0,404411	0,391814	0,007171	0	1	1	1	3	0
0,672498	0,442011	0,925602	0,873412	0	1	0	0	1	0
0,271859	0,075867	0,664250	0,033255	1	1	0	1	3	0
0,364026	0,845316	0,627482	0,856940	1	0	0	0	1	0
0,632611	0,553600	0,770940	0,274467	0	0	0	1	1	0
0,649012	0,029424	0,006652	0,975738	0	1	1	0	2	0
0,880217	0,464580	0,444413	0,229474	0	1	1	1	3	0
0,729499	0,862228	0,434498	0,712452	0	0	1	0	1	0
0,828743	0,473059	0,188661	0,362972	0	1	1	1	3	0
0,716120	0,509904	0,006310	0,038622	0	0	1	1	2	0
0,412185	0,497310	0,608060	0,849470	1	1	0	0	2	0
0,981724	0,115540	0,050374	0,286358	0	1	1	1	3	0
0,381995	0,448187	0,918297	0,122803	1	1	0	1	3	0
0,259900	0,155538	0,494559	0,196396	1	1	1	1	4	1
0,315719	0,364588	0,423351	0,503743	1	1	1	0	3	0
0,649735	0,399039	0,469560	0,555198	0	1	1	0	2	0
0,279561	0,400026	0,734911	0,884434	1	1	0	0	2	0
0,217361	0,710319	0,005094	0,087384	1	0	1	1	3	0
0,693183	0,111708	0,147777	0,169500	0	1	1	1	3	0

0,484144	0,427837	0,319137	0,725863	1	1	1	0	3	0
0,667289	0,471618	0,979072	0,318985	0	1	0	1	2	0
0,375134	0,258344	0,725386	0,251190	1	1	0	1	3	0
0,191600	0,062678	0,997449	0,221200	1	1	0	1	3	0
0,237928	0,570485	0,325694	0,095081	1	0	1	1	3	0
0,479030	0,416032	0,736089	0,454510	1	1	0	1	3	0
0,407433	0,253097	0,157676	0,636510	1	1	1	0	3	0
0,439608	0,986541	0,567047	0,105619	1	0	0	1	2	0
0,028231	0,344357	0,134230	0,159250	1	1	1	1	4	1
0,622533	0,397073	0,574875	0,553743	0	1	0	0	1	0
0,839669	0,023235	0,693584	0,578297	0	1	0	0	1	0
0,451561	0,152428	0,938211	0,987627	1	1	0	0	2	0
0,276442	0,775556	0,775507	0,796108	1	0	0	0	1	0
0,415808	0,177497	0,462652	0,300068	1	1	1	1	4	1
0,591290	0,217437	0,850252	0,584043	0	1	0	0	1	0
0,184162	0,337955	0,413556	0,824248	1	1	1	0	3	0
0,511650	0,734839	0,543550	0,167983	0	0	0	1	1	0
0,961578	0,127654	0,619978	0,597268	0	1	0	0	1	0
0,151112	0,759022	0,970418	0,348384	1	0	0	1	2	0
0,870196	0,884215	0,568004	0,733497	0	0	0	0	0	0
0,965750	0,997447	0,012327	0,269384	0	0	1	1	2	0
0,407247	0,134092	0,921995	0,633147	1	1	0	0	2	0
0,129992	0,026175	0,442862	0,157719	1	1	1	1	4	1
0,741260	0,046352	0,823044	0,740928	0	1	0	0	1	0
0,799936	0,900769	0,513758	0,075029	0	0	0	1	1	0
0,314404	0,518245	0,279696	0,640607	1	0	1	0	2	0
0,646709	0,492754	0,751594	0,791626	0	1	0	0	1	0
0,238706	0,194777	0,760938	0,320155	1	1	0	1	3	0

6

Na tabela anterior, as 4 primeiras colunas apresentam as 100 simulações dos nascimentos dos 4 filhos, utilizando a função RAND(), como se mostra a seguir:

	A	B	C	D
1	1º filho	2º filho	3º filho	4º filho
2	=RAND()	=RAND()	=RAND()	=RAND()
3	=RAND()	=RAND()	=RAND()	=RAND()
4	=RAND()	=RAND()	=RAND()	=RAND()
5	=RAND()	=RAND()	=RAND()	=RAND()
6	=RAND()	=RAND()	=RAND()	=RAND()

Nas colunas seguintes utilizámos a função IF, para verificar se cada nascimento era ou não rapaz. Sempre que fosse rapaz, assinalámos esse facto com um 1. Na coluna seguinte, ou seja a coluna I, colocámos a soma das 4 colunas anteriores, o que nos permite detectar facilmente se os 4 nascimentos são rapazes. Na coluna seguinte, a coluna J, ainda utilizando a função IF, verificámos se nos 4 nascimentos se tinha verificado os 4 rapazes. Caso afirmativo, assinalámos esse facto com um 1. Caso contrário com um 0. Este processo permite que a soma da coluna J nos dê o número de casos favoráveis à realização do acontecimento de que estamos a calcular a probabilidade. Na simulação anterior obtivemos 6 casos favoráveis, pelo que uma estimativa para a probabilidade pretendida é  $6/100 = 0,06$ .



sexo. Esta metodologia é mais simples de implementar, do que estar em cada repetição a verificar se nos dois primeiros filhos já existem os dois sexos.

Utilizando a folha de Excel, nas colunas B, C e D simulamos o nascimento de três crianças e nas colunas E, F e G, testamos se são rapazes ou raparigas:

	A	B	C	D	E	F	G
1	Nºrepetição	1ºfilho	2ºfilho	3ºfilho	Sexo	Sexo	Sexo
2	1	=RAND()	=RAND()	=RAND()	=IF(B2<0,5;"rapaz";"rapariga")	=IF(C2<	=IF(D2<
3	2	=RAND()	=RAND()	=RAND()	=IF(B3<0,5;"rapaz";"rapariga")	=IF(C3<	=IF(D3<
4	3	=RAND()	=RAND()	=RAND()	=IF(B4<0,5;"rapaz";"rapariga")	=IF(C4<	=IF(D4<
5	4	=RAND()	=RAND()	=RAND()	=IF(B5<0,5;"rapaz";"rapariga")	=IF(C5<	=IF(D5<
6	5	=RAND()	=RAND()	=RAND()	=IF(B6<0,5;"rapaz";"rapariga")	=IF(C6<	=IF(D6<
7	6	=RAND()	=RAND()	=RAND()	=IF(B7<0,5;"rapaz";"rapariga")	=IF(C7<	=IF(D7<
8	7	=RAND()	=RAND()	=RAND()	=IF(B8<0,5;"rapaz";"rapariga")	=IF(C8<	=IF(D8<
9	8	=RAND()	=RAND()	=RAND()	=IF(B9<0,5;"rapaz";"rapariga")	=IF(C9<	=IF(D9<

Na coluna H, testamos se os dois primeiros filhos são do mesmo sexo. Se forem de sexo diferente, consideramos um sucesso e assinalamos esse facto com um 1. Caso sejam do mesmo sexo, testa-se se o 3º filho é do mesmo sexo que os anteriores: se for de sexo diferente considera-se um sucesso que é assinalado, como no caso anterior, com um 1. Se forem todos do mesmo sexo esse acontecimento é um insucesso e é assinalado com um 0. A soma dos valores da coluna H dá-nos o número de sucessos. Nas colunas I e J consideramos, respectivamente, a frequência absoluta acumulada e a frequência relativa acumulada do acontecimento "Ter um casalinho":

	H	I	J
1	Um casalinho?	Freq. abs.acum.	Freq. rel. acum.
2	=IF(E2=F2;IF(G2=F2;0;1);1)	=H2	=I2/A2
3	=IF(E3=F3;IF(G3=F3;0;1);1)	=H3+I2	=I3/A3
4	=IF(E4=F4;IF(G4=F4;0;1);1)	=H4+I3	=I4/A4
5	=IF(E5=F5;IF(G5=F5;0;1);1)	=H5+I4	=I5/A5
6	=IF(E6=F6;IF(G6=F6;0;1);1)	=H6+I5	=I6/A6
7	=IF(E7=F7;IF(G7=F7;0;1);1)	=H7+I6	=I7/A7
8	=IF(E8=F8;IF(G8=F8;0;1);1)	=H8+I7	=I8/A8
9	=IF(E9=F9;IF(G9=F9;0;1);1)	=H9+I8	=I9/A9

Realizaram-se 1000 repetições de que se apresenta a seguir um pequeno extracto:

	A	B	C	D	E	F	G	H	I	J
1	Nº repetição	1ºfilho	2ºfilho	3ºfilho	Sexo	Sexo	Sexo	Um casalinho?	Freq. abs.acum.	Freq. rel. acum.
2	1	0,2087	0,9158	0,2881	rapaz	rapariga	rapaz	1	1	1,00
3	2	0,9050	0,2457	0,6099	rapariga	rapaz	rapariga	1	2	1,00
4	3	0,7717	0,7467	0,0112	rapariga	rapariga	rapaz	1	3	1,00
5	4	0,9694	0,9149	0,9705	rapariga	rapariga	rapariga	0	3	0,75
6	5	0,3542	0,1504	0,2805	rapaz	rapaz	rapaz	0	3	0,60
7	6	0,5413	0,3153	0,4417	rapariga	rapaz	rapaz	1	4	0,67
8	7	0,0666	0,3778	0,3209	rapaz	rapaz	rapaz	0	4	0,57
9	8	0,1367	0,6527	0,5899	rapaz	rapariga	rapariga	1	5	0,63

...

	A	B	C	D	E	F	G	H	I	J
992	991	0,1447	0,7676	0,9283	rapaz	rapariga	rapariga	1	751	0,76
993	992	0,7411	0,0844	0,4408	rapariga	rapaz	rapaz	1	752	0,76
994	993	0,8223	0,7175	0,3184	rapariga	rapariga	rapaz	1	753	0,76
995	994	0,5539	0,5720	0,3409	rapariga	rapariga	rapaz	1	754	0,76
996	995	0,0779	0,8164	0,9033	rapaz	rapariga	rapariga	1	755	0,76
997	996	0,7223	0,4818	0,4649	rapariga	rapaz	rapaz	1	756	0,76
998	997	0,1072	0,8329	0,9374	rapaz	rapariga	rapariga	1	757	0,76
999	998	0,7483	0,8733	0,0724	rapariga	rapariga	rapaz	1	758	0,76
1000	999	0,7939	0,7833	0,5022	rapariga	rapariga	rapariga	0	758	0,76
1001	1000	0,3394	0,8824	0,5790	rapaz	rapariga	rapariga	1	759	0,76

Como se verifica da tabela anterior, a frequência relativa estabilizou à volta do valor 0,76, pelo que uma estimativa para a probabilidade de o casal ter um casalinho, num máximo de 3 filhos, é 0,76.

E se a probabilidade de nascer rapaz for um pouco superior à de nascer rapariga, como indiciam os resultados do Censo 2001? Vamos então repetir a simulação anterior, mas agora considerando que a probabilidade de nascer rapaz é 0,51 e a de nascer rapariga 0,49. A única alteração que temos de fazer na simulação anterior, é nas colunas E, F e G, em que agora consideramos o nascimento de um rapaz sempre que o número pseudo-aleatório gerado for  $<0,51$ .

Neste caso o resultado da simulação deu como estimativa para a probabilidade de o casal ter um casalinho, o valor de 0,74, como se verifica no extracto da folha de Excel que se apresenta a seguir:

	A	B	C	D	E	F	G	H	I	J
994	993	0,1693	0,9280	0,3823	rapaz	rapariga	rapaz	1	736	0,74
995	994	0,6928	0,0437	0,7883	rapariga	rapaz	rapariga	1	737	0,74
996	995	0,9203	0,7212	0,5500	rapariga	rapariga	rapariga	0	737	0,74
997	996	0,8047	0,3149	0,4329	rapariga	rapaz	rapaz	1	738	0,74
998	997	0,4217	0,1473	0,5310	rapaz	rapaz	rapariga	1	739	0,74
999	998	0,5172	0,4897	0,8708	rapariga	rapaz	rapariga	1	740	0,74
1000	999	0,5610	0,9626	0,9590	rapariga	rapariga	rapariga	0	740	0,74
1001	1000	0,4188	0,9278	0,2715	rapaz	rapariga	rapaz	1	741	0,74

**Tarefa – Estimar as probabilidades dos resultados da soma das pintas das faces viradas para cima, quando se lançam dois dados.** Vimos no capítulo anterior um modelo de probabilidade para o resultado da soma das pintas das faces viradas para cima, quando se lançam dois dados. Vamos admitir que os dados são equilibrados

*Passo 1 – Definição do modelo de probabilidade para o lançamento dos dois dados*

O modelo que vamos adoptar para o lançamento de cada um dos dados pressupõe que:

- ✓ Em cada lançamento existe igual probabilidade de sair cada uma das faces;
- ✓ Os lançamentos são independentes uns dos outros e de dado para dado.

*Passo 2 – Atribuição de números aleatórios para representar os resultados do lançamento de dois dados.*

Vamos utilizar a função Randbetween(1;6) do Excel, do seguinte modo:

- ✓ Cada número gerado simula o número de pintas da face que fica voltada para cima quando se procede ao lançamento de um dado;

- ✓ A função anterior é utilizada duas vezes para representar os dois dados.

*Passo 3 – Simular muitas repetições*

Vamos gerar muitas repetições de 2 números pseudo-aleatórios, para simular os 2 lançamentos:

	A	B	C
1	Dado1	Dado2	Soma
2	=RANDBETWEEN(1;6)	=RANDBETWEEN(1;6)	=A2+B2
3	=RANDBETWEEN(1;6)	=RANDBETWEEN(1;6)	=A3+B3
4	=RANDBETWEEN(1;6)	=RANDBETWEEN(1;6)	=A4+B4
5	=RANDBETWEEN(1;6)	=RANDBETWEEN(1;6)	=A5+B5
6	=RANDBETWEEN(1;6)	=RANDBETWEEN(1;6)	=A6+B6
7	=RANDBETWEEN(1;6)	=RANDBETWEEN(1;6)	=A7+B7
8	=RANDBETWEEN(1;6)	=RANDBETWEEN(1;6)	=A8+B8
9	=RANDBETWEEN(1;6)	=RANDBETWEEN(1;6)	=A9+B9

	A	B	C
1	Dado1	Dado2	Soma
2	3	1	4
3	6	4	10
4	4	3	7
5	1	1	2
6	2	5	7
7	4	6	10
8	4	3	7
9	3	2	5

Na figura acima está um pequeno extracto da tabela que contém as 1000 repetições realizadas e os resultados da soma do número de pintas. Utilizando agora a função *Countif* do Excel, facilmente se obtém a frequência relativa com que se verificou cada resultado:

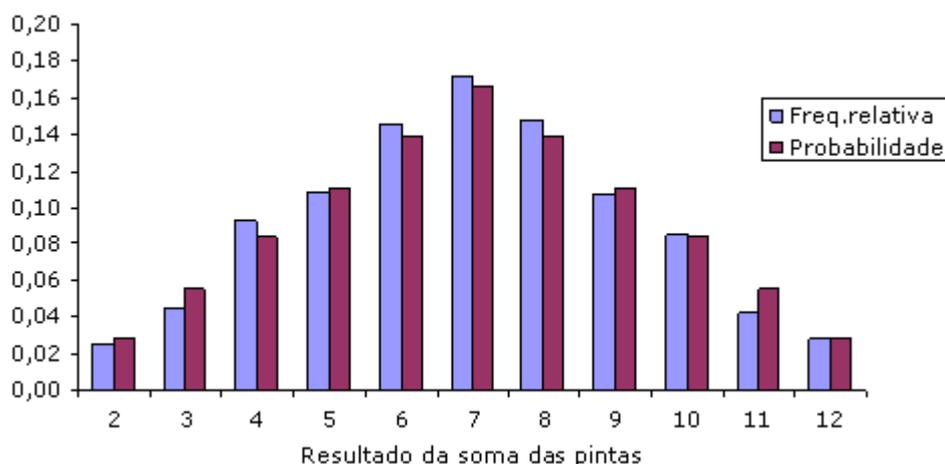
	E	F
1	Resultado	Freq. relativa
2	2	=COUNTIF(\$C\$2:\$C\$1001;E2)/1000
3	3	=COUNTIF(\$C\$2:\$C\$1001;E3)/1000
4	4	=COUNTIF(\$C\$2:\$C\$1001;E4)/1000
5	5	=COUNTIF(\$C\$2:\$C\$1001;E5)/1000
6	6	=COUNTIF(\$C\$2:\$C\$1001;E6)/1000
7	7	=COUNTIF(\$C\$2:\$C\$1001;E7)/1000
8	8	=COUNTIF(\$C\$2:\$C\$1001;E8)/1000
9	9	=COUNTIF(\$C\$2:\$C\$1001;E9)/1000
10	10	=COUNTIF(\$C\$2:\$C\$1001;E10)/1000
11	11	=COUNTIF(\$C\$2:\$C\$1001;E11)/1000
12	12	=COUNTIF(\$C\$2:\$C\$1001;E12)/1000

	E	F
1	Resultado	Freq. relativa
2	2	0,025
3	3	0,045
4	4	0,092
5	5	0,109
6	6	0,146
7	7	0,172
8	8	0,148
9	9	0,108
10	10	0,085
11	11	0,042
12	12	0,028

Na tabela anterior, as frequências relativas são estimativas das probabilidades dos resultados da soma das pintas das faces viradas para cima, quando se lançam dois dados. Na tabela e gráficos seguintes comparamos essas probabilidades experimentais com as probabilidades teóricas obtidas a partir do modelo de probabilidade da página 240:

	E	F	G
1	Resultado	Freq. relativa	Probabilidade
2	2	0,025	0,028
3	3	0,045	0,056
4	4	0,092	0,083
5	5	0,109	0,111
6	6	0,146	0,139
7	7	0,172	0,167
8	8	0,148	0,139
9	9	0,108	0,111
10	10	0,085	0,083
11	11	0,042	0,056
12	12	0,028	0,028



**Tarefa – O jogo é justo?** O Pedro e a Rita disputam entre si o seguinte jogo: lançam dois dados e calculam a diferença entre as pintas das faces que ficam viradas para cima. O Pedro ganha um rebuçado se essa diferença for 0, 1 ou 2, enquanto que a Rita ganha um rebuçado se a diferença for 3, 4 ou 5. Ao fim de 20 jogadas quem é que se espera que ganhe o jogo? O Pedro e a Rita estavam convencidos de que o jogo seria equilibrado, já que a cada um deles estavam atribuídos 3 dos 6 resultados possíveis. Começaram a desconfiar de que alguma coisa não estaria a correr bem, pois o Pedro estava a ganhar com mais frequência que a Rita. Afinal o jogo será justo? Se não for justo, consegues arranjar um processo de o tornar justo?

O processo de simulação é idêntico ao da tarefa anterior, mas agora em vez de se calcular a soma das pintas, calcula-se a diferença. Utilizámos um outro processo para simular o número de pintas do dado, com o auxílio da função Rand(). O processo de simulação apresenta-se a seguir:

	A	B	C
1	Dado 1	Dado2	Diferença
2	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A2-B2)
3	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A3-B3)
4	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A4-B4)
5	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A5-B5)
6	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A6-B6)
7	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A7-B7)
8	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A8-B8)
9	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A9-B9)
10	=INT(1+6*RAND())	=INT(1+6*RAND())	=ABS(A10-B10)

	A	B	C
1	Dado 1	Dado2	Diferença
2	5	5	0
3	1	5	4
4	3	6	3
5	3	3	0
6	5	3	2
7	6	6	0
8	1	5	4
9	3	3	0
10	3	4	1

Note-se que para obter a diferença entre as pintas dos dois dados considerámos a função ABS, que nos devolve o valor absoluto dessa diferença.

Simulámos o lançamento dos dois dados 20 vezes, tendo obtido os seguintes resultados:

	F	G	H	I	J	K	L
1	Diferença	Freq. Abs.	Freq. Rel.				
2	0	3	0,15		Pedro ganha	15	jogos
3	1	7	0,35		Rita ganha	5	jogos
4	2	5	0,25				
5	3	3	0,15				
6	4	2	0,10				
7	5	0	0,00				
8		20	1				

Repetimos mais vezes a simulação das 20 jogadas, tendo obtido os resultados da tabela seguinte, onde na primeira linha aparece o número de rebuçados ou jogadas ganhas pelo Pedro em cada 20 jogadas:

	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
9	Pedro	14	11	13	13	14	17	14	14	10	8	12	14	12	13	12	12	15	16	14	11	12	13	15	11	13
10	Rita	6	9	7	7	6	3	6	6	10	12	8	6	8	7	8	8	5	4	6	9	8	7	5	9	7

Dos 25 jogos (de 20 jogadas cada um dos jogos), o Pedro ganhou 23, a Rita 1 e empataram 1 jogo.

Como se pode ver pelo resultado da simulação, o jogo está longe de ser equilibrado! Os resultados 0, 1 ou 2 saem com uma frequência muito maior que os restantes. Para obter uma estimativa, com maior precisão, para a probabilidade do Pedro ganhar o jogo simulámos 100 jogadas e concluímos que o Pedro ganha aproximadamente 69% das jogadas ( $0,19+0,29+0,21=0,69$ ) enquanto que a Rita ganha cerca de 31%:

	F	G	H
10	Diferença	Freq. Abs.	Freq. Rel.
11	0	19	0,19
12	1	29	0,29
13	2	21	0,21
14	3	16	0,16
15	4	12	0,12
16	5	3	0,03
17		100	1

Uma forma de equilibrar o jogo seria atribuir ao Pedro os resultados 0, 1 e 5 e à Rita os resultados 2, 3 e 4.

Como complemento desta tarefa, vamos comparar as probabilidades experimentais obtidas para os resultados 0, 1, 2, 3, 4 e 5, com as probabilidades teóricas. Para obter estas probabilidades, basta fazer um raciocínio semelhante ao que foi feito para obter as probabilidades dos resultados da soma das pintas das faces obtidas quando se lançam dois dados. Considera-se uma tabela com os resultados dos lançamentos dos dois dados, a que chamámos Dado 1 e Dado 2, e preenche-se o interior da tabela com a diferença das pintas:

Dado 2 \ Dado 1	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Dos 36 pares, igualmente possíveis, que constituem o espaço de resultados associado ao lançamento dos dois dados:

- 6 pares fazem com que o resultado da diferença seja 0
- 10 pares fazem com que o resultado da diferença seja 1
- 8 pares fazem com que o resultado da diferença seja 2
- 6 pares fazem com que o resultado da diferença seja 3
- 4 pares fazem com que o resultado da diferença seja 4
- 2 pares fazem com que o resultado da diferença seja 5

Assim, tem-se o seguinte modelo de probabilidade para o resultado da diferença entre o número de pintas:

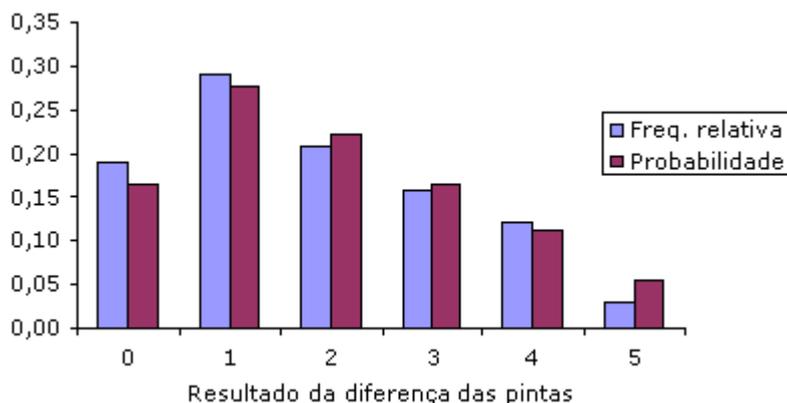
Resultado	0	1	2	3	4	5
Probabilidade	6/36	10/36	8/36	6/36	4/36	2/36

ou,

Resultado	0	1	2	3	4	5
Probabilidade	0,167	0,278	0,222	0,167	0,111	0,056

A partir do modelo anterior pode-se calcular a probabilidade (teórica) do Pedro ganhar a próxima jogada, bastando para isso somar as probabilidades de obter os resultados 0, 1 ou 2, que dá aproximadamente 67%. O valor da probabilidade experimental de aproximadamente 69%, obtido com as 100 jogadas, é uma estimativa razoavelmente próxima da probabilidade teórica.

Pode-se ainda utilizar o modelo anterior para comparar as frequências relativas obtidas na simulação das 100 jogadas, com as probabilidades teóricas dos resultados que se obtêm quando se faz a diferença entre os valores das pintas obtidas quando se lançam dois dados:



**Tarefa – Vamos a uma aposta?** Cinco amigas encontraram-se para comer um gelado. Falaram sobre vários temas, até que surgiu o assunto dos signos. Nessa altura uma delas, a Maria, propôs a seguinte aposta às restantes:

- Aposto um almoço em como duas de nós têm o mesmo signo! Alguém quer aceitar esta aposta?

Se fosses uma das amigas da Maria aceitavas a aposta?

Vamos ajudar-te a tomar uma decisão. Vejamos como simular a situação de em cinco pessoas, duas terem o mesmo signo. Sem a ajuda da tecnologia, este processo poderia ser simulado da forma seguinte: corta-se a figura seguinte de forma a separar os signos que se metem numa caixa.



Depois selecciona-se, da caixa, um dos pedaços de papel, verifica-se qual o signo e repõe-se novamente na caixa o papel retirado. Repete-se este processo 5 vezes e regista-se se houve repetição de algum signo. Se nos 5 signos retirados houver repetições, regista-se esse facto como um sucesso. Repete-se algumas vezes o processo de retirar os cinco signos, e regista-se o número de sucessos.

Este processo de simular o acontecimento de interesse, embora simples, é muito demorado e pouco interessante, pelo que rapidamente chegamos à conclusão que é preferível utilizar a tecnologia...

Para simular a saída de cada um dos 12 signos, utilizou-se a função `Randbetween(1;12)` do Excel, que permite obter cada um dos valores inteiros entre 1 e 12 com igual probabilidade. Não nos vamos preocupar com a correspondência entre os signos e os números, já que a nossa questão se prende com a repetição de qualquer dos signos e não com algum em particular.

	A	B	C	D	E
1	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
2	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
3	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
4	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
5	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
6	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
7	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
8	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
9	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE
10	=RANDBETWEEN(1;12)	=RANDBE	=RANDBE	=RANDBE	=RANDBE

Apresenta-se a seguir um extracto das 500 repetições em que em cada repetição se simulou a retirada, com reposição, de 5 signos.

Sempre que na simulação dos cinco números ocorrerem pelo menos dois iguais, então registamos esse facto como um sucesso. Caso contrário temos um insucesso.

	A	B	C	D	E	F	G
1	8	7	3	7	11		1
2	4	7	8	11	11		1
3	9	7	1	2	7		1
4	11	10	8	3	8		1
5	9	10	1	3	12		0
6	2	6	10	11	12		0
7	12	1	9	11	12		1
8	...	...	...	...	...		...
495	4	9	4	10	11		1
496	1	5	7	6	10		0
497	7	4	1	2	2		1
498	11	11	12	12	11		1
499	12	11	6	2	10		0
500	12	4	10	12	3		1
501							308

Na coluna G representou-se por 1 um sucesso e por 0 um insucesso. O número de sucessos nas 500 realizações da experiência foi de 308, pelo que uma estimativa para a probabilidade de pelo menos duas das cinco amigas terem signos repetidos é 0,616, ou seja é superior a 60%. Com este resultado, talvez fosse de apostar no almoço!

**Tarefa – Qual a probabilidade de no lançamento de uma moeda, se verificarem pelo menos 3 faces Euro seguidas?** O professor chegou à turma e escreveu no quadro várias sequências de E (Euro) e N (Nacional), que pretendiam representar o resultado do lançamento de uma moeda de um euro 6 vezes. Uma dessas sequências tinha sido inventada e o professor pretendia que os alunos lhe dissessem qual era a que a que eles achavam mais provável de ter sido inventada:

E E N E E N   E E E N E N   E N E N E N   N E E N N N   N N E N N N

Depois de algumas considerações sobre se a moeda seria ou não equilibrada, e de o professor responder que a moeda era equilibrada, a maior parte dos alunos exprimiu as suas conclusões, não em termos da "mais provável de ter sido inventada", mas da "menos provável de ter sido inventada".

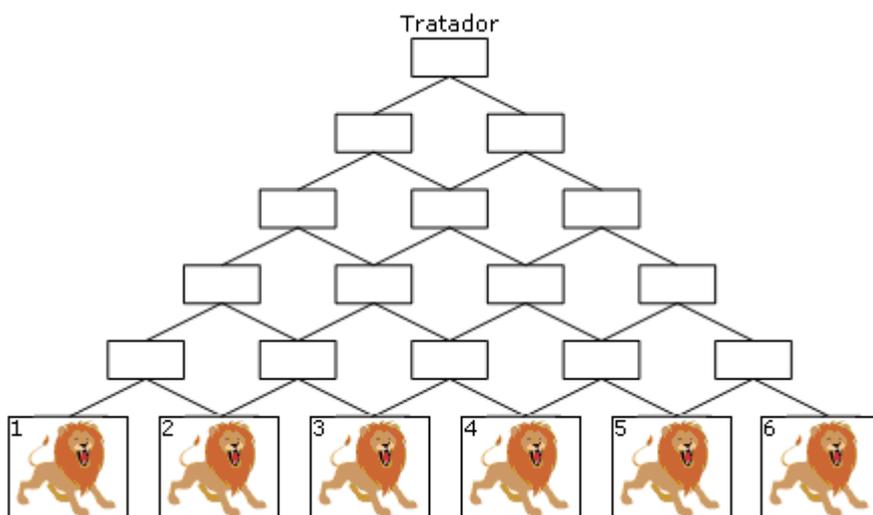
O seu raciocínio baseava-se no seguinte: se a moeda é equilibrada, então existe igual possibilidade de sair face Euro ou face Nacional, pelo que nos 6 lançamentos espera-se igual número de faces Euro e faces Nacional! Além disso 3 faces Euro ou

3 faces Nacional seguidas é pouco provável...! Concluía portanto, que o mais razoável é que a sequência não inventada fosse E N E N E N!

Acontece que o raciocínio anterior está longe de estar correcto! Numa sequência de 6 lançamentos, não se pode esperar a regularidade que se espera numa sequência de muitos lançamentos. O facto da moeda ser equilibrada significa que depois de muitas repetições a frequência com que se verifica a face Nacional é aproximadamente igual à frequência com que se verifica a face Euro. Chamamos a atenção para que quando nos referimos à frequência, estamos a referir-nos à frequência relativa, pois já vimos que à medida que o número de repetições aumenta, a frequência absoluta não obedece a nenhuma regularidade. Por outro lado, embora seja pouco intuitivo para os alunos, é bem mais frequente observar uma sequência de 6 lançamentos em que se verificam 3 faces Euro seguidas, do que uma sequência em que haja alternância da face Euro com a face Nacional nos 6 lançamentos. Para vermos como a nossa intuição nos engana com frequência, vamos estimar essas probabilidades. Para isso repetiu-se 500 vezes a simulação de seis lançamentos da moeda. O processo de simulação é idêntico ao realizado noutras tarefas anteriores, pelo que nos abstermos de o considerar aqui e apresentamos só os resultados.

Nas 500 simulações de 6 lançamentos da moeda, observou-se 150 vezes a sequência E E E e 13 vezes as sequências E N E N E N ou N E N E N E, pelo que uma estimativa para a probabilidade de 3 faces Euro seguidas é aproximadamente 30%, enquanto que uma estimativa para a probabilidade de obter uma sequência de 6 faces alternadas não chega a 3%!

**Tarefa – Quem é que recebe mais comida?** No jardim zoológico existem seis leões, cada um na sua jaula. O tratador resolveu arranjar um processo de dar a comida aos leões, em que cada pedaço de carne passa por cinco prateleiras até chegar a um leão. Em cada prateleira o pedaço de carne pode escorregar para a prateleira da direita ou da esquerda com igual probabilidade:



O tratador estava convencido que ao fim de vários pedaços de carne, todos os leões teriam mais ou menos a mesma quantidade de comida, já que para chegar do tratador a cada leão cada pedaço de carne tem de passar pelo mesmo número de prateleiras. Acontece que ao fim de alguns dias uns leões estavam mais gordos do

que outros e havia alguns que estavam mesmo a definir! És capaz de mostrar ao tratador que este processo de lançar a comida aos leões é capaz de não ser bom?

Vamos simular o lançamento de vários pedaços de carne e verificar se existem algumas posições onde seja mais provável de chegar a carne do que a outras. A simulação da saída do pedaço de carne de cada prateleira é idêntica à simulação do lançamento de uma moeda equilibrada, ou do nascimento de um rapaz, como fizemos numa das tarefas anteriores. Como cada pedaço de carne tem de percorrer 5 prateleiras:

- Se virar 5 vezes para a direita, vai parar ao Leão da jaula 1;
- Se virar 4 vezes para a direita e 1 vez para a esquerda em qualquer das prateleiras, vai parar ao Leão da jaula 2;
- Se virar 3 vezes para a direita e 2 vezes para a esquerda em quaisquer das prateleiras, vai parar ao Leão da jaula 3;
- Se virar 2 vezes para a direita e 3 vezes para a esquerda em quaisquer das prateleiras, vai parar ao Leão da jaula 4;
- Se virar 1 vez para a direita e 4 vezes para a esquerda em quaisquer das prateleiras, vai parar ao Leão da jaula 5;
- Se virar 0 vezes para a direita e 5 vezes para a esquerda, vai parar ao Leão da jaula 6.

Sintetizando os diferentes passos da simulação, vamos admitir que:

- ✓ Em cada prateleira existe igual probabilidade de virar à esquerda ou à direita;
- ✓ As prateleiras são independentes umas das outras.

Vamos utilizar a função Rand do Excel, do seguinte modo:

- ✓ Cada número gerado simula a saída de um pedaço de carne de uma prateleira;
- ✓ Um número  $\leq 0,5$  representa a saída para a direita, e um número  $> 0,5$  representa a saída para a esquerda.

Vamos gerar muitas repetições de 5 números pseudo-aleatórios, para simular a passagem de muitos pedaços de carne pelas 5 prateleiras:

	A	B	C	D	E
1	1ª prateleira	2ª prateleira	3ª prateleira	4ª prateleira	5ª prateleira
2	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
3	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
4	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
5	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
6	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
7	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
8	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
9	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
10	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()

	A	B	C	D	E
1	1ªprateleira	2ªprateleira	3ªprateleira	4ªprateleira	5ªprateleira
2	0,3325	0,5879	0,2907	0,9607	0,9695
3	0,4584	0,1177	0,2447	0,2675	0,6622
4	0,1339	0,0943	0,0424	0,1210	0,9784
5	0,5090	0,3252	0,8872	0,8102	0,9611
6	0,4801	0,7748	0,5773	0,7279	0,7316
7	0,1413	0,5362	0,0861	0,4083	0,1554
8	0,6725	0,7201	0,9466	0,9550	0,9560
9	0,9989	0,8299	0,4829	0,8987	0,7618
10	0,0175	0,1141	0,4216	0,3930	0,7758

	G	H	I	J	
1	1ªprateleira	2ªprateleira	3ªprateleira	4ªprateleira	5ªpra
2	=IF(A2<0,5;"D";"E")	=IF(B2<0,5;"D";"E")	=IF(C2<0,5;"D";"E")	=IF(D2<0,5;"D";"E")	=IF(E
3	=IF(A3<0,5;"D";"E")	=IF(B3<0,5;"D";"E")	=IF(C3<0,5;"D";"E")	=IF(D3<0,5;"D";"E")	=IF(E
4	=IF(A4<0,5;"D";"E")	=IF(B4<0,5;"D";"E")	=IF(C4<0,5;"D";"E")	=IF(D4<0,5;"D";"E")	=IF(E
5	=IF(A5<0,5;"D";"E")	=IF(B5<0,5;"D";"E")	=IF(C5<0,5;"D";"E")	=IF(D5<0,5;"D";"E")	=IF(E
6	=IF(A6<0,5;"D";"E")	=IF(B6<0,5;"D";"E")	=IF(C6<0,5;"D";"E")	=IF(D6<0,5;"D";"E")	=IF(E
7	=IF(A7<0,5;"D";"E")	=IF(B7<0,5;"D";"E")	=IF(C7<0,5;"D";"E")	=IF(D7<0,5;"D";"E")	=IF(E
8	=IF(A8<0,5;"D";"E")	=IF(B8<0,5;"D";"E")	=IF(C8<0,5;"D";"E")	=IF(D8<0,5;"D";"E")	=IF(E
9	=IF(A9<0,5;"D";"E")	=IF(B9<0,5;"D";"E")	=IF(C9<0,5;"D";"E")	=IF(D9<0,5;"D";"E")	=IF(E
10	=IF(A10<0,5;"D";"E")	=IF(B10<0,5;"D";"E")	=IF(C10<0,5;"D";"E")	=IF(D10<0,5;"D";"E")	=IF(E

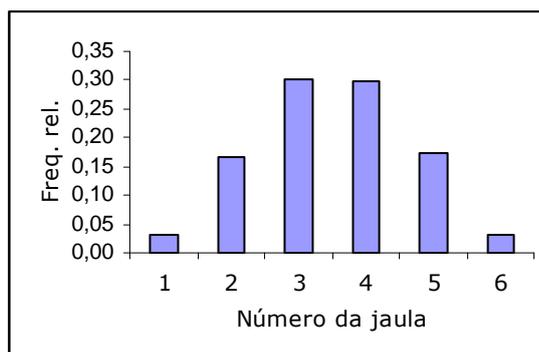
	G	H	I	J	K	M
1	1ªprateleira	2ªprateleira	3ªprateleira	4ªprateleira	5ªprateleira	Jaula
2	D	E	D	E	E	4
3	D	D	D	D	E	5
4	D	D	D	D	E	5
5	E	D	E	E	E	2
6	D	E	E	E	E	2
7	D	E	D	D	D	5
8	E	E	E	E	E	1
9	E	E	D	E	E	2
10	D	D	D	D	E	5

A contagem do número de vezes que em cada linha aparece o E foi feita utilizando a função Countif, e a atribuição do número da jaula é feita adicionando a esse resultado o 1:

	L	M
1		Jaula
2	=COUNTIF(G2:K2;"E")	=L2+1
3	=COUNTIF(G3:K3;"E")	=L3+1
4	=COUNTIF(G4:K4;"E")	=L4+1
5	=COUNTIF(G5:K5;"E")	=L5+1
6	=COUNTIF(G6:K6;"E")	=L6+1
7	=COUNTIF(G7:K7;"E")	=L7+1
8	=COUNTIF(G8:K8;"E")	=L8+1
9	=COUNTIF(G9:K9;"E")	=L9+1
10	=COUNTIF(G10:K10;"E")	=L10+1

Fizemos 500 repetições e obtivemos os seguintes resultados:

	O	P	Q
1	Jaula	Freq. Abs.	Freq. Rel.
2	1	15	0,030
3	2	83	0,166
4	3	151	0,302
5	4	149	0,298
6	5	86	0,172
7	6	16	0,032
8		500	1



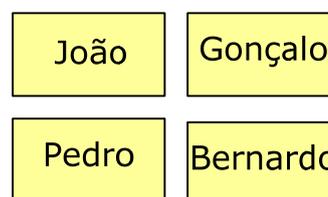
Os resultados anteriores comprovam que os leões não estavam a ser alimentados em igual proporção. Efectivamente, os leões das jaulas 3 e 4 recebiam 60% da comida, enquanto que os das jaulas 1 e 6 só recebiam cerca de 6%.

**Tarefa – Qual será a probabilidade de cada amigo ficar com o seu chapéu-de-chuva?**<sup>2</sup> Quatro amigos, o João, o Gonçalo, o Pedro e o Bernardo foram ao cinema e como estava a chover levaram chapéu-de-chuva, que tiveram de deixar no bengaleiro, à entrada da sala. Quando acabou o filme dirigiram-se ao bengaleiro e cada um colocou a sua ficha com o número do cabide onde estava o chapéu-de-chuva, em cima do balcão. A senhora que estava a tomar conta do bengaleiro pegou nas 4 fichas, retirou os 4 chapéus-de-chuva e entregou ao acaso um chapéu a cada um dos amigos!

Qual a probabilidade de cada um ter recebido o seu chapéu-de-chuva? Esta probabilidade será superior ou inferior à probabilidade de nenhum ter recebido o seu chapéu?

Vamos começar por descrever um processo de simulação da atribuição aleatória dos chapéus-de-chuva pelos 4 amigos, mas neste caso, em vez de usar a tecnologia, como nas tarefas anteriores, usamos um processo manual, com materiais simples e facilmente realizado na sala de aula: divide-se uma folha de papel A4 em 4 partes, e em cada uma das partes escrevemos um dos 4 nomes. Num pedaço de cartolina cortam-se 4 rectângulos iguais, do tamanho aproximado de uma carta de jogar e em cada um escreve-se também o nome de um dos 4 amigos. Estes cartões vão representar os chapéus-de-chuva.

João	Gonçalo
Bernardo	Pedro



Baralham-se os cartões, e com os nomes virados para baixo, distribuem-se aleatoriamente sobre a folha de papel, cada um em cada uma das partes em que a

<sup>2</sup> Adaptado de Rossman et al (2001), p. 301

folha foi dividida. De seguida viram-se os cartões e regista-se o número de chapéus que calharam aos respectivos donos:

João [Cartão amarelo]	Gonçalo [Cartão amarelo]	João Bernardo [Cartão amarelo]	Gonçalo João [Cartão amarelo]
Bernardo [Cartão amarelo]	Pedro [Cartão amarelo]	Bernardo Gonçalo [Cartão amarelo]	Pedro Pedro [Cartão amarelo]

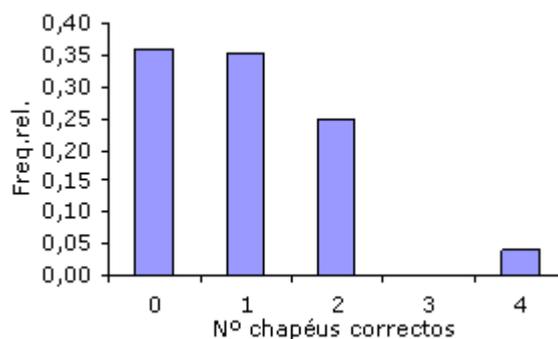
Na atribuição dos chapéus anteriormente considerada só houve uma correspondência correcta, uma vez que só o Pedro é que teve o seu chapéu.

Esta experiência pode-se realizar na sala de aula por vários grupos de alunos, registando-se numa tabela os resultados obtidos por todos os grupos, para o número de chapéus que foram correctamente atribuídos. Ao fim de 125 repetições obtiveram-se os seguintes resultados:

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
1	1	0	1	1
1	4	1	2	1
0	2	0	0	2
2	1	0	0	0
2	0	2	2	1
2	0	1	0	1
1	2	0	2	2
2	1	0	1	2
2	1	1	1	0
0	1	0	0	1
2	4	0	0	1
0	4	0	0	0
0	4	1	1	2
2	1	1	0	0
1	0	1	2	2
0	0	0	1	1
2	1	0	0	1
2	2	1	2	2
1	1	0	1	0
0	2	2	2	1
2	0	0	4	1
1	0	0	1	0
1	0	2	1	0
2	0	2	0	1
0	1	1	1	0

Os resultados anteriores podem ser resumidos na seguinte tabela de frequências e no diagrama de barras respectivo:

Nº chapéus correctos	Freq.abs.	Freq.rel.
0	45	0,36
1	44	0,35
2	31	0,25
3	0	0,00
4	5	0,04
Total	125	



Da tabela anterior verificamos que estimativas para a probabilidade de todos os amigos terem recebido o seu chapéu e de nenhum ter recebido o chapéu que lhe pertenciam são, respectivamente, 4% e 36%.

Verificamos também que não se verificou a possibilidade de três dos quatro amigos terem recebido os seus chapéus! Já seria um resultado esperado? Podemos adiantar que a probabilidade do acontecimento "Três dos quatro amigos recebem o seu chapéu" é igual a zero?

A implementação em Excel desta simulação pode ser feita da seguinte forma:

1. Representam-se os 4 amigos João, Gonçalo, Pedro e Bernardo por 1, 2, 3 e 4, respectivamente.
2. Gera-se um conjunto de 4 números pseudo-aleatórios e considera-se a ordem de cada um dentro desse conjunto (a probabilidade de haver dois números iguais é igual a zero). Estes números assim obtidos representam os chapéus-de-chuva. Cada uma destas ordenações é considerada como uma permutação dos 4 números 1, 2, 3 e 4. Sempre que houver coincidência de posição entre os números que representam os chapéus e os números que representam os amigos, temos um chapéu correctamente atribuído:

Na tabela seguinte simulamos várias atribuições dos chapéus aos 4 rapazes. Nas colunas A, B, C e D geramos os números pseudo-aleatórios, enquanto que nas colunas E, F, G e H se colocam as ordens dos números anteriores:

	A	B	C	D	E	F	G	H
1					1	2	3	4
2	0,533745	0,328636	0,941379	0,596829	2	1	4	3
3	0,798905	0,522238	0,648000	0,046260	4	2	3	1
4	0,184600	0,122697	0,142513	0,855135	3	1	2	4
5	0,386946	0,423667	0,868105	0,543366	1	2	4	3
6	0,993263	0,427269	0,254599	0,698081	4	2	1	3
7	0,470085	0,928101	0,435957	0,674889	2	4	1	3
8	0,372178	0,519314	0,383613	0,722340	1	3	2	4
9	0,140098	0,678730	0,947669	0,028423	2	3	4	1
10	0,776288	0,671346	0,898459	0,873389	2	1	4	3
11	0,756998	0,157176	0,924244	0,547016	3	1	4	2
12	0,010649	0,527832	0,418300	0,340001	1	4	3	2
13	0,090495	0,651239	0,373328	0,296977	1	4	3	2
14	0,349492	0,786104	0,356045	0,108264	2	4	3	1
15	0,323950	0,935830	0,235480	0,706960	2	4	1	3
16	0,850841	0,421778	0,870159	0,894493	2	1	3	4
17	0,119558	0,650625	0,082222	0,138328	2	4	1	3
18	0,241127	0,668011	0,682551	0,815034	1	2	3	4
19	0,993044	0,471330	0,516536	0,815036	4	1	2	3
20	0,866755	0,351051	0,666058	0,558450	4	1	3	2
21	0,605353	0,902571	0,565335	0,407095	3	4	2	1

Sempre que houver uma coincidência de posição, considera-se que se atribuiu correctamente o chapéu ao seu dono. Por exemplo, na segunda simulação realizada, o Gonçalo e o Pedro receberam os seus chapéus; na terceira simulação foi o Bernardo que recebeu o seu chapéu; na quarta simulação o João e o Gonçalo receberam os seus chapéus; etc. Repare-se que na simulação correspondente à linha 18, todos os chapéus foram atribuídos correctamente aos seus donos.

A determinação das ordens foi feita utilizando uma função do Excel, que é a função Rank, como se pode ver na figura seguinte:

	E	F	G	H
1				
2	=RANK(A2;\$A2:\$D2;1)	=RANK(B2;\$A2:\$D2;1)	=RANK(C2;\$A2:\$D2;1)	=RANK(D2;\$A2:\$D2;1)
3	=RANK(A3;\$A3:\$D3;1)	=RANK(B3;\$A3:\$D3;1)	=RANK(C3;\$A3:\$D3;1)	=RANK(D3;\$A3:\$D3;1)
4	=RANK(A4;\$A4:\$D4;1)	=RANK(B4;\$A4:\$D4;1)	=RANK(C4;\$A4:\$D4;1)	=RANK(D4;\$A4:\$D4;1)
5	=RANK(A5;\$A5:\$D5;1)	=RANK(B5;\$A5:\$D5;1)	=RANK(C5;\$A5:\$D5;1)	=RANK(D5;\$A5:\$D5;1)
6	=RANK(A6;\$A6:\$D6;1)	=RANK(B6;\$A6:\$D6;1)	=RANK(C6;\$A6:\$D6;1)	=RANK(D6;\$A6:\$D6;1)
7	=RANK(A7;\$A7:\$D7;1)	=RANK(B7;\$A7:\$D7;1)	=RANK(C7;\$A7:\$D7;1)	=RANK(D7;\$A7:\$D7;1)
8	=RANK(A8;\$A8:\$D8;1)	=RANK(B8;\$A8:\$D8;1)	=RANK(C8;\$A8:\$D8;1)	=RANK(D8;\$A8:\$D8;1)
9	=RANK(A9;\$A9:\$D9;1)	=RANK(B9;\$A9:\$D9;1)	=RANK(C9;\$A9:\$D9;1)	=RANK(D9;\$A9:\$D9;1)

A função Rank(a,b,c) tem o seguinte significado: a é o elemento de que pretendemos obter a ordem, quando inserido numa lista, que é representada por b. O argumento c, quando omissivo significa que a ordem que pretendemos é a ordem decrescente. Qualquer valor deste argumento, significa que a ordem é crescente.

Utilizando a função IF nas colunas I, J, K e L, verificou-se se havia atribuição correcta do chapéu ao dono.

Finalmente na coluna M, contabilizou-se o número de atribuições correctas, em cada simulação

	I	J	K	L	M
1					
2	=IF(E2=1;1;	=IF(F2=2;1;	=IF(G2=3;1;	=IF(H2=4;1;	=SUM(I2:L2)
3	=IF(E3=1;1;	=IF(F3=2;1;	=IF(G3=3;1;	=IF(H3=4;1;	=SUM(I3:L3)
4	=IF(E4=1;1;	=IF(F4=2;1;	=IF(G4=3;1;	=IF(H4=4;1;	=SUM(I4:L4)
5	=IF(E5=1;1;	=IF(F5=2;1;	=IF(G5=3;1;	=IF(H5=4;1;	=SUM(I5:L5)
6	=IF(E6=1;1;	=IF(F6=2;1;	=IF(G6=3;1;	=IF(H6=4;1;	=SUM(I6:L6)
7	=IF(E7=1;1;	=IF(F7=2;1;	=IF(G7=3;1;	=IF(H7=4;1;	=SUM(I7:L7)
8	=IF(E8=1;1;	=IF(F8=2;1;	=IF(G8=3;1;	=IF(H8=4;1;	=SUM(I8:L8)
9	=IF(E9=1;1;	=IF(F9=2;1;	=IF(G9=3;1;	=IF(H9=4;1;	=SUM(I9:L9)

	B	C	D	E	F	G	H	I	J	K	L	M
1				1	2	3	4					Nº atribuições correctas
2	0,328636	0,941379	0,596829	2	1	4	3	0	0	0	0	0
3	0,522238	0,648000	0,046260	4	2	3	1	0	1	1	0	2
4	0,122697	0,142513	0,855135	3	1	2	4	0	0	0	1	1
5	0,423667	0,868105	0,543366	1	2	4	3	1	1	0	0	2
6	0,427269	0,254599	0,698081	4	2	1	3	0	1	0	0	1
7	0,928101	0,435957	0,674889	2	4	1	3	0	0	0	0	0
8	0,519314	0,383613	0,722340	1	3	2	4	1	0	0	1	2
9	0,678730	0,947669	0,028423	2	3	4	1	0	0	0	0	0
10	0,671346	0,898459	0,873389	2	1	4	3	0	0	0	0	0
11	0,157176	0,924244	0,547016	3	1	4	2	0	0	0	0	0
12	0,527832	0,418300	0,340001	1	4	3	2	1	0	1	0	2
13	0,651239	0,373328	0,296977	1	4	3	2	1	0	1	0	2
14	0,786104	0,356045	0,108264	2	4	3	1	0	0	1	0	1
15	0,935830	0,235480	0,706960	2	4	1	3	0	0	0	0	0
16	0,421778	0,870159	0,894493	2	1	3	4	0	0	1	1	2
17	0,650625	0,082222	0,138328	2	4	1	3	0	0	0	0	0

Repetimos a simulação 500 vezes e obtivemos as seguintes estimativas (frequências relativas) para as probabilidades do número de chapéus correctamente atribuídos:

Nº chapéus correctos	Freq.abs.	Freq.rel.
0	191	0,382
1	162	0,324
2	129	0,258
3	0	0,000
4	18	0,036
Total	500	1,000

Para esta situação é relativamente simples obter o modelo de probabilidade teórico, considerando todas as 24 permutações possíveis de 1234:

Permutação	Nº chapéus correctos	Permutação	Nº chapéus correctos
<b>1234</b>	4	<b>1243</b>	2
<b>1324</b>	2	<b>1342</b>	1
<b>1423</b>	1	<b>1432</b>	2
<b>2134</b>	2	<b>2143</b>	0
<b>2314</b>	1	<b>2341</b>	0
<b>2413</b>	0	<b>2431</b>	1
<b>3124</b>	1	<b>3142</b>	0
<b>3214</b>	2	<b>3241</b>	1
<b>3412</b>	0	<b>3421</b>	0
<b>4123</b>	0	<b>4132</b>	1
<b>4213</b>	1	<b>4231</b>	2
<b>4312</b>	0	<b>4321</b>	0

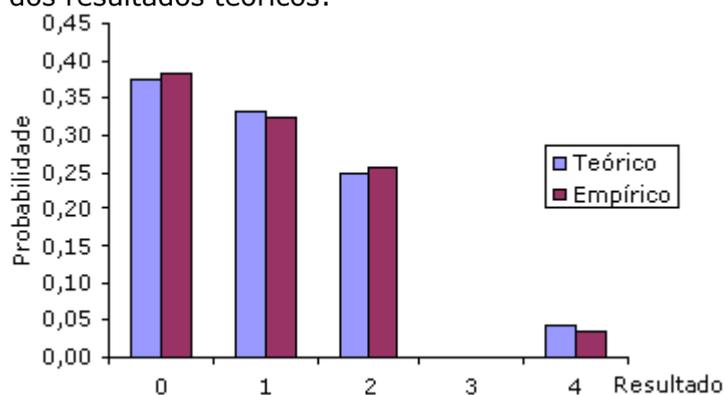
Então, considerando que todas as permutações são igualmente possíveis, a probabilidade de cada resultado do número de chapéus correctamente atribuídos é obtida através da lei de Laplace, como sendo o número de casos favoráveis sobre o número de casos possíveis:

Nº chapéus correctos	0	1	2	4
Probabilidade	9/24	8/24	6/24	1/24

ou

Nº chapéus correctos	0	1	2	4
Probabilidade	0,375	0,333	0,250	0,042

Comparando os resultados do modelo empírico anteriormente obtido, com o modelo teórico, verificamos que os resultados obtidos experimentalmente por simulação, estão próximos dos resultados teóricos:



**Tarefa – Qual a probabilidade de passar no exame?** O João é pouco estudioso e como não gosta da disciplina de *Estudos Gerais*, apresenta-se ao exame e responde a todas as questões ao acaso. O exame é constituído por 5 questões e cada questão é de resposta múltipla, tendo 6 respostas possíveis, das quais só uma é correcta. O João só passa no exame se responder bem a pelo menos 3 questões. Qual a probabilidade do João passar a *Estudos Gerais*?

- Qual a probabilidade do João responder certo a uma questão?
- Utilizar a tabela de dígitos aleatórios da página 258 para simular a realização de um exame à disciplina de Estudos Gerais.
- Simular várias repetições do exame.
- Qual a estimativa para a probabilidade do João passar ao exame?

Resolução:



Porque é que o João decidiu lançar um dado ao ar, para decidir qual a resposta que devia dar a cada questão?

- Como o João responde ao acaso, a probabilidade de responder bem a cada questão é  $1/6$ , já que cada questão tem 6 respostas possíveis, das quais só uma é a correcta. Assim, o João atira o dado ao ar e escolhe a resposta cujo número for o número de pintas do dado.
- Para utilizar a tabela de dígitos aleatórios da página 258, vamos proceder do seguinte modo:

Consideram-se números decimais de 3 dígitos e se um desses números

- For menor que 0,167 admitimos que se escolhe a resposta 1;
- Estiver entre 0,167 e 0,333 admitimos que se escolhe a resposta 2;
- Estiver entre 0,333 e 0,500 admitimos que se escolhe a resposta 3;
- Estiver entre 0,500 e 0,667 admitimos que se escolhe a resposta 4;
- Estiver entre 0,667 e 0,833 admitimos que se escolhe a resposta 5;
- Estiver entre 0,833 e 1 admitimos que se escolhe a resposta 6.

Repare-se que segundo esta metodologia, cada número tem igual probabilidade de sair, sendo essa probabilidade igual a  $1/6$ , pois dividimos o intervalo (0, 1) em 6 partes iguais.

Admitindo que o exame tem 5 questões e que as respostas certas são os números 2, 5, 4, 3, 1, respectivamente, vejamos o resultado da simulação de um exame:

	Questão 1	Questão 2	Questão 3	Questão 4	Questão 5
<b>Respostas certas</b>	<b>2</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>1</b>
Número aleatório	0,192	0,239	0,503	0,405	0,756
Resultado simulação	2	2	4	3	5

Na simulação anterior o João respondeu bem a 3 questões! Será que não vale a pena estudar?

- Apresentamos a seguir o resultado da simulação de 24 repetições do exame (tantas quantas a tabela disponibilizou):

						2	5	4	3	1	Nº respostas certas	Freq. rel. de 3 ou mais respostas certas.
1	0,192	0,239	0,503	0,405	0,756	2	2	4	3	5	3	1,000
2	0,287	0,139	0,640	0,912	0,531	2	1	4	6	4	2	0,500
3	0,425	0,448	0,285	0,373	0,676	3	3	2	3	5	1	0,333
4	0,471	0,509	0,940	0,001	0,927	3	4	6	1	6	0	0,250
5	0,277	0,544	0,264	0,882	0,425	2	4	2	6	3	1	0,200
6	0,362	0,904	0,546	0,771	0,709	3	6	4	5	5	1	0,167
7	0,775	0,580	0,009	0,532	0,863	5	4	1	4	6	0	0,143
8	0,294	0,858	0,222	0,690	0,056	2	6	2	5	1	2	0,125
9	0,527	0,113	0,888	0,993	0,074	4	1	6	6	1	1	0,111
10	0,602	0,274	0,001	0,185	0,848	4	2	1	2	6	0	0,100
11	0,487	0,675	0,257	0,395	0,592	3	5	2	3	4	1	0,091
12	0,940	0,076	0,997	0,191	0,481	6	1	6	2	3	0	0,083
13	0,607	0,795	0,379	0,117	0,297	4	5	3	1	2	1	0,077
14	0,593	0,356	0,841	0,735	0,013	4	3	6	5	1	1	0,071
15	0,155	0,297	0,276	0,585	0,089	1	2	2	4	1	1	0,067
16	0,570	0,675	0,021	0,147	0,487	4	5	1	1	3	1	0,063
17	0,827	0,395	0,789	0,020	0,807	5	3	5	1	5	0	0,059
18	0,475	0,118	0,167	0,655	0,300	3	1	2	4	2	0	0,056
19	0,943	0,831	0,489	0,360	0,940	6	5	3	3	6	1	0,053
20	0,720	0,241	0,786	0,824	0,943	5	2	5	5	6	0	0,050
21	0,617	0,909	0,065	0,687	0,964	4	6	1	5	6	0	0,048
22	0,188	0,833	0,600	0,919	0,365	2	6	4	6	3	2	0,045
23	0,154	0,123	0,963	0,885	0,453	1	1	6	6	3	0	0,043
24	0,468	0,168	0,348	0,541	0,979	3	2	3	4	6	0	0,042

d) Das 24 simulações do exame, o João só passaria em uma delas, pelo que uma estimativa para a probabilidade do João passar é 0,042.

O número de simulações realizadas anteriormente não é suficiente, como se comprova pelo facto da frequência relativa do acontecimento "3 ou mais respostas certas" não ter estabilizado.

Uma resolução alternativa à resolução anterior, sem utilizar a tabela de dígitos aleatórios, poderá ser a seguinte:

- Utilizar a função RAND() do Excel, ou da máquina de calcular, para gerar números pseudo-aleatórios entre 0 e 1;
- Se o número obtido for inferior a 0,167, considerar a resposta a uma questão como certa e atribuir o valor 1. Caso contrário atribuir o valor 0;
- Repetir o passo anterior 5 vezes, contabilizando o número de 1's (respostas certas) obtidos (as), para simular a realização de um exame;
- Repetir o procedimento anterior até que a frequência relativa do acontecimento "3 ou mais respostas certas" estabilize.

**Tarefa proposta – Pequena sondagem sobre o tipo sanguíneo.** De acordo com informação disponível na página do Instituto Português do Sangue (<http://www.ipsangue.org/maxcontent-documento-231.html> e revista *ABO* nº 29 de Janeiro/ Março de 2007), 42% da população tem grupo sanguíneo de tipo O. Na turma, o professor pediu a um grupo de alunos que investigassem junto de 10 colegas, escolhidos ao acaso, qual o seu grupo sanguíneo e que calculassem uma estimativa para a probabilidade de todos terem o grupo sanguíneo de tipo O.

Para obter a estimativa pretendida, o grupo de alunos decidiu utilizar a seguinte metodologia:

1º passo – Utilizando a função RAND() do Excel, simulou a resposta de cada aluno à questão “O teu grupo sanguíneo é de tipo O?”, da seguinte forma: Se o valor de RAND() $<0,42$ , considera-se a resposta Sim e representa-se por um 1; caso contrário representa-se por um 0.

2º passo – Repetiram o processo anterior 10 vezes, em 10 colunas do Excel, para simular a pergunta a 10 colegas;

3º passo – Consideraram uma coluna auxiliar onde em cada célula colocaram o produto das 10 células da mesma linha. Se este produto for igual a 1, significa que todos os 10 alunos têm grupo O;

4º passo – Repetiram o processo anterior várias vezes para simular as respostas de 10 colegas e calcularam a frequência relativa do número de 1's da coluna auxiliar, que dá uma estimativa da probabilidade pretendida.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	
2	1	0,79	0,49	0,47	0,43	0,47	0,52	0,46	0,85	0,94	0,90	0	0	0	0	0	0	0	0	0	0	0
3	2	0,60	0,40	0,57	0,50	0,74	0,75	0,82	0,96	0,09	0,93	0	1	0	0	0	0	0	0	0	1	0
4	3	0,54	0,03	0,64	0,39	0,13	0,40	0,01	0,50	0,60	0,79	0	1	0	1	1	1	1	0	0	0	0
5	4	0,71	0,88	0,41	0,86	0,77	0,66	0,57	0,63	1,00	0,36	0	0	1	0	0	0	0	0	0	0	1
6	5	0,09	0,28	0,51	0,55	0,79	0,63	0,04	0,19	0,33	0,80	1	1	0	0	0	0	1	1	1	0	0
7	6	0,98	0,00	0,91	0,33	0,02	0,63	0,96	0,34	0,35	0,83	0	1	0	1	1	0	0	1	1	0	0
8	7	0,31	0,27	0,31	0,66	0,86	0,80	0,07	0,06	0,46	0,18	1	1	1	0	0	0	1	1	0	1	0
9	...				...					...												
997	995	0,76	0,60	0,73	0,58	0,41	0,10	0,29	0,96	0,67	0,79	0	0	0	0	1	1	1	0	0	0	0
998	996	0,12	0,54	0,76	0,06	0,59	0,16	0,35	0,07	0,34	0,83	1	0	0	1	0	1	1	1	1	0	0
999	997	0,83	0,28	0,18	0,57	0,21	0,30	0,32	0,12	0,88	0,25	0	1	1	0	1	1	1	1	0	1	0
1000	998	0,84	0,51	0,91	0,46	0,54	0,79	0,82	0,23	0,72	0,54	0	0	0	0	0	0	0	1	0	0	0
1001	999	0,32	0,55	0,30	0,61	0,06	0,63	0,78	0,84	0,47	0,02	1	0	1	0	1	0	0	0	0	1	0
1002	###	0,94	0,38	0,46	0,72	0,45	0,72	0,62	0,98	0,31	0,79	0	1	0	0	0	0	0	0	1	0	0

A simulação das 10 respostas foi repetida 1000 vezes e em nenhuma das vezes se verificou 1 em todas as respostas. Assim, uma estimativa para a probabilidade pretendida é 0.

Nota – Assumindo que a probabilidade de um indivíduo, escolhido ao acaso, ter sangue de tipo O é 0,42, pode-se mostrar que a probabilidade de 10 indivíduos, escolhidos ao acaso, terem todos sangue de tipo O, é 0,00017.

Uma alternativa à utilização da folha de Excel, é a utilização da máquina de calcular. Nas tarefas do capítulo seguinte consideramos 2 tarefas em que ela é utilizada.



# Tarefas

Apresentam-se a seguir alguns exemplos de tarefas para a sala de aula.





## Animais selvagens

		Total
		
		
		
		
		
		
		
		
		

### Tarefa – O mês do nosso aniversário



O professor faz passar uma folha A4 com o esquema que se apresenta a seguir, onde pede a cada aluno que coloque um x à frente do mês em que faz anos.

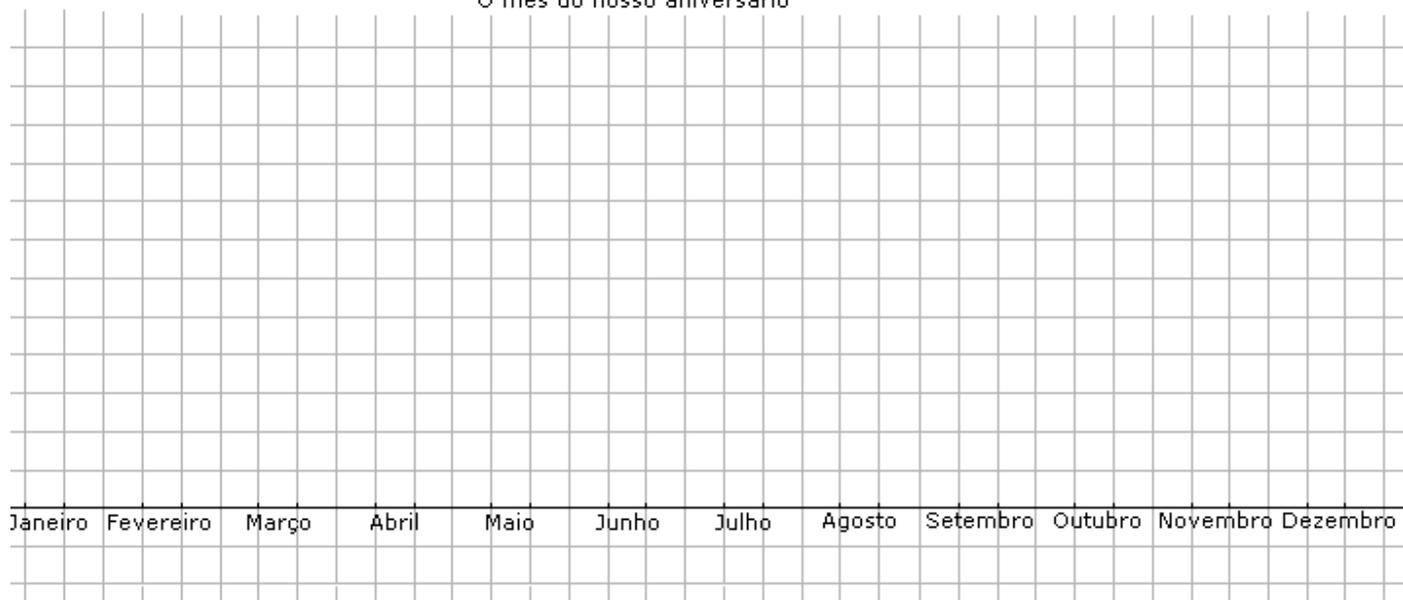
Mês		
Janeiro		
Fevereiro		
Março		
Abril		
Mai		
Junho		
Julho		
Agosto		
Setembro		
Outubro		
Novembro		
Dezembro		

Depois de recolher a folha pede aos alunos para responderem a algumas questões:

- Quantos alunos colocaram um x na folha?
- Quantos alunos estão na sala de aula?
- Houve algum aluno que não colocou um x na folha? Será que esse aluno não se lembra em que mês faz anos?
- Qual o mês em que mais alunos fazem anos?
- Qual o mês em que menos alunos fazem anos?

Outra forma de recolher a informação anterior é utilizando uma folha quadriculada e construir um gráfico de pontos:

O mês do nosso aniversário





## Tarefa – Qual o tipo de gelado preferido?



Na turma, decidiram recolher informação sobre o tipo de gelado preferido. Pretendiam ainda averiguar se o gosto das meninas seria idêntico ao dos rapazes.

Para responder à questão anterior, o professor escreveu no quadro os quatro tipos de gelados preferidos pelos alunos e pediu a cada aluno que apontasse qual o gelado preferido:

<i>Rapazes</i>		<i>Raparigas</i>
	Chocolate	
	Baunilha	
	Morango	
	Nata	

No esquema anterior podem-se distinguir os gelados preferidos dos rapazes e das raparigas, já que os rapazes registaram as suas preferências do lado esquerdo, enquanto que as raparigas o fizeram do lado direito.

Algumas questões que podem ser respondidas a partir do esquema anterior:

- Quantos alunos estavam na turma, quando responderam a esta questão?
- Estavam mais rapazes, ou mais raparigas?
- Qual o gelado preferido das raparigas?
- O gelado preferido dos rapazes é o mesmo que o das raparigas?
- Algum rapaz prefere gelado de nata?



## Tarefa – O nosso tipo de sandes favorita para o lanche



O professor e os alunos decidiram fazer um pequeno trabalho de investigação sobre o tipo de sandes que os alunos da turma preferem para o lanche. Depois de alguma troca de impressões, chegaram à conclusão que os alunos gostavam de sandes de queijo, de fiambre, mistas e de doce. Mas qual seria a preferida?

Cada aluno foi ao quadro escrever o nome da sandes preferida e o resultado foi o seguinte:

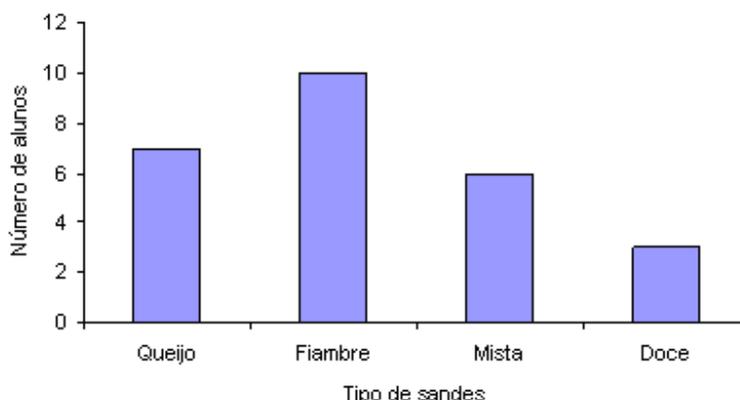
queijo, queijo, doce, fiambre, mista, mista, queijo, fiambre, fiambre, fiambre, mista, mista, doce, fiambre, queijo, queijo, mista, fiambre, fiambre, queijo, doce, fiambre, mista, fiambre, fiambre, queijo

Depois de todos terem escrito o nome da sandes preferida, obtiveram uma confusão de nomes, em que era muito difícil concluir alguma coisa com a informação recolhida desta forma. Seria necessário organizar os dados para saberem qual a sandes preferida. A maneira mais simples seria construir uma tabela (de frequências), onde colocavam o nome de todas as sandes e à frente de cada nome, o número de alunos que a preferiam:

Tipo de sandes	Número de alunos
Queijo	7
Fiambre	10
Mista	6
Doce	3

Depois da tabela construída foi fácil concluir que a sandes preferida era a de fiambre e que aquela que os alunos gostavam menos era a de doce.

Ainda sugeriram fazer uma representação gráfica muito simples, o *gráfico de barras*, que fazia sobressair melhor que uma tabela, a informação recolhida:

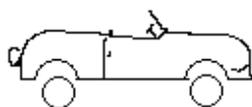




### Tarefa – Qual a cor de carros preferida pelos alunos da nossa escola?



Um comerciante de automóveis decidiu investigar quais as cores de carros preferidas pelos futuros compradores de carros. Assim, recolheu alguma informação junto dos alunos de uma escola que apresentou no seguinte pictograma onde



representa 4 carros

Cor	Rapaz	Rapariga
Preta		
Encarnada		
Prateada		

Responde às seguintes questões:

- Quantos rapazes preferem a cor preta?
- Qual a cor preferida das raparigas? É a mesma que a dos rapazes?
- Quantas raparigas manifestaram a sua opinião? E quantos rapazes?



## Tarefa – A temperatura que faz em algumas localidades



A professora propôs aos alunos da turma fazerem um trabalho de investigação sobre o tempo que faz. Os alunos são divididos em grupos e cada grupo tem como objectivo estudar a temperatura numa dada região. Poderão escolher algumas cidades do Norte, Centro e Sul do país, como por exemplo, Porto, Lisboa e Faro e ainda a cidade onde vivem ou a mais perto da região em que vivem.

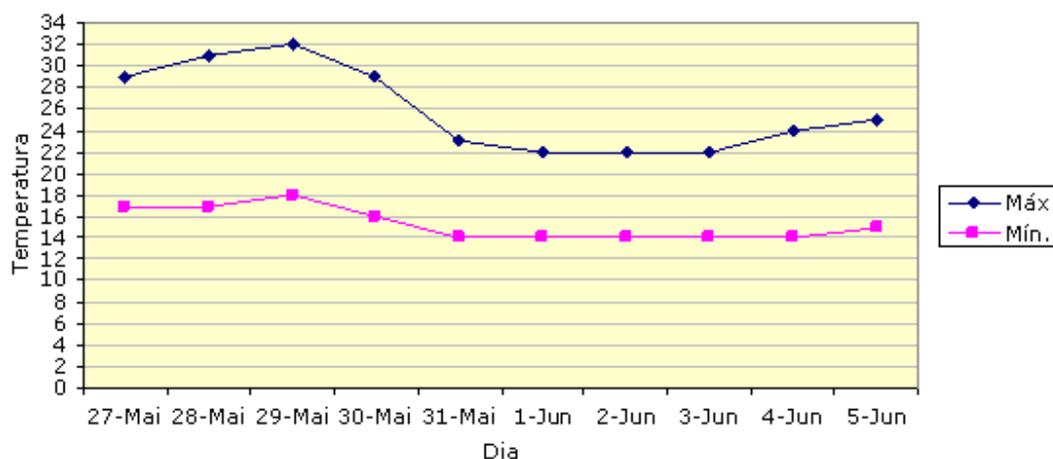
Cada dia e durante 30 dias seguidos, cada grupo recolhe a informação, através dos meios de comunicação social, da Internet, etc., sobre a temperatura máxima e mínima que fará nesse dia.

A planificação da recolha de dados deve ser feita com algum cuidado, tendo em conta os objectivos pretendidos. Por exemplo, se se pretender estudar a evolução da temperatura durante esses 30 dias, terá de se recolher a informação sobre o dia e sobre a temperatura máxima e mínima desse dia. Se se pretender unicamente calcular algumas estatísticas sobre as temperaturas, já não será necessário recolher informação sobre o dia em que as temperaturas foram recolhidas. A professora deve alertar para o facto de que este tipo de informação que se recolhe ao longo do tempo pode ser objecto de uma representação em gráfico de linha, pelo que temos uma informação mais rica se se registar também o dia.

Para exemplificar apresentamos o registo da temperatura na cidade de Lisboa durante 10 dias seguidos:

Dia	Temp. Máxima(°C)	Temp. Mínima (°C)
27-Mai	29	17
28-Mai	31	17
29-Mai	32	18
30-Mai	29	16
31-Mai	23	14
01-Jun	22	14
02-Jun	22	14
03-Jun	22	14
04-Jun	24	14
05-Jun	25	15

A partir dos dados anteriores constrói-se um gráfico de linhas, onde registamos a evolução das temperaturas máxima e mínima, entre as datas consideradas:



A partir do gráfico anterior podemos responder facilmente a algumas questões tais como:

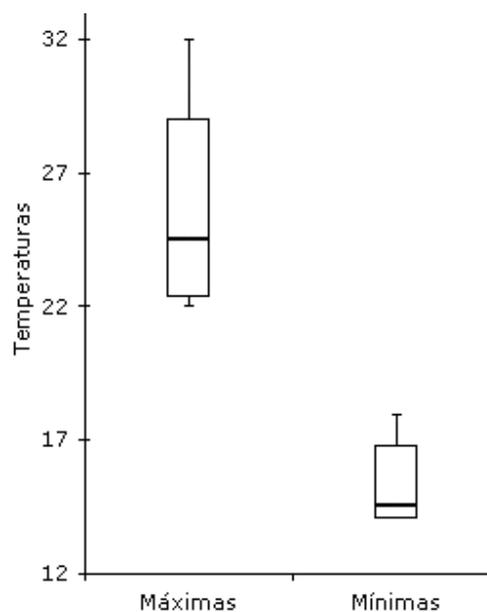
- Em que dia é que se verificou a temperatura máxima mais alta? Nesse dia a temperatura mínima também foi a mais alta?
- Qual o dia, ou dias, em que a temperatura mínima foi mais baixa?
- Houve maior variação entre as temperaturas máximas ou entre as temperaturas mínimas?

Outras questões que podem ter interesse é averiguar a temperatura média, tanto entre as máximas como as mínimas, o dia em foi maior a diferença entre a temperatura máxima e a temperatura mínima, a que damos o nome de amplitude térmica, etc.

Será também interessante comparar as temperaturas de várias cidades.

Uma representação que também pode ser utilizada para comparar as temperaturas máximas com as mínimas ou as temperaturas entre cidades, é o diagrama de extremos e quartis. Para isso é necessário calcular a mediana e os quartis. Vamos exemplificar com os dados anteriores:

	Máximas	Mínimas
1º quartil	22,25	14
Mínimo	22	14
Mediana	24,5	14,5
Máximo	32	18
3º quartil	29	16,75



A partir do diagrama de extremos e quartis anteriores apercebemo-nos que os dados apresentam um enviesamento para a direita, nomeadamente os dados referentes às temperaturas mínimas, que têm a particularidade de o mínimo e o 1º quartil coincidirem. Como seria de esperar, as temperaturas máximas são sensivelmente superiores às temperaturas mínimas.

Esta representação é adequada para comparar vários conjuntos de dados, pelo que seria interessante os vários grupos juntarem os dados observados referentes às várias cidades e compararem-nos através de um gráfico deste tipo.

## Tarefa – A corrida de automóveis

Este jogo vai ser jogado por 11 corredores, escolhidos de entre os alunos da turma, com os carros numerados de 2 a 12, que vão participar numa corrida, com a seguinte particularidade. O vencedor da corrida não depende da velocidade do carro, mas do resultado do lançamento de dois dados, de acordo com as seguintes regras:

- Lançam-se dois dados equilibrados e somam-se as pintas das faces que ficam viradas para cima;
- Faz-se andar de uma quadrícula o carro cujo número é o resultado da soma anterior; para simular esta deslocação pode-se marcar uma cruz na quadrícula respectiva.
- O jogo termina assim que algum dos carros percorrer as 15 quadrículas, que estão à sua frente ou seja assim que as 15 quadrículas correspondentes a algum carro estiverem preenchidas. Esse carro será o vencedor.

Algumas questões que o professor discutirá com os alunos, antes de começar a corrida:

- Porque é que os carros só estão numerados de 2 a 12? Porque é que não aparece o número 1?
- Quais são as expectativas para o(s) número(s) do(s) carro(s) vencedor(es)? Porquê?

À medida que a tabela com os carros for sendo preenchida com cruces, deve o professor juntamente com os alunos avaliar a representação que está a ser obtida e quais os carros que têm mais cruces à frente. Deve chamar a atenção para o facto de, depois de um número razoável de lançamento dos dados, a tabela obtida permitir estimar, experimentalmente, as probabilidades do resultado da soma das pintas dos dados ser 2, 3, ..., ou 12.

Além desta avaliação experimental, sobre as probabilidades do resultado da soma das pintas dos dados ser 2, 3, ..., ou 12, o professor orienta os alunos no sentido de obter teoricamente essas probabilidades, a partir de um modelo proposto para a experiência aleatória que consiste em lançar 2 dados e verificar a soma das pintas das faces viradas para cima.

O espaço de resultados associado a esta experiência aleatória é constituído pelos seguintes resultados:

$$S = (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)$$

Para obter as probabilidades dos acontecimentos elementares constituídos pelos resultados anteriores, constrói-se uma tabela de dupla entrada, em que na coluna do lado esquerdo se assinalam as pintas de um dos dados, por exemplo o Dado 1, e na primeira linha se assinalam as pintas do outro dado. As células do interior da tabela serão preenchidas com o resultado da soma dos algarismos que estiverem na mesma linha e na mesma coluna onde estão registadas o número de pintas:

		Dado 2					
		1	2	3	4	5	6
Dado 1	1						
	2						
	3						
	4						
	5						
	6						

		Dado 2					
		1	2	3	4	5	6
Dado 1	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Da tabela anterior verifica-se que quando se lançam os dois dados existem 36 pares, igualmente possíveis, que fazem com que a soma das pintas varie entre 2 e 12. Só o par (1, 1) dá o resultado 2, assim como só o par (6, 6) dá o resultado 12. Para obter o resultado 3 ou 11 existem 2 possibilidades, dadas respectivamente pelos pares (1, 2), (2, 1) e (5, 6), (6, 5). Verifica-se também que o resultado 7 é o que se obtém mais vezes, pois pode ser obtido por obtido pelos pares (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) e (6,1). Tendo em consideração a regra de Laplace, vem o seguinte modelo de probabilidade para o resultado da soma das pintas das faces dos dois dados:

Resultado	2	3	4	5	6	7	8	9	10	11	12
Probabilidade	1/36	2/36	3/36	4/36	5/36	6/36	1/36	2/36	3/36	4/36	5/36

Os alunos são incentivados a comparar as frequências relativas obtidas experimentalmente, com as probabilidades teóricas dadas pelo modelo anterior.



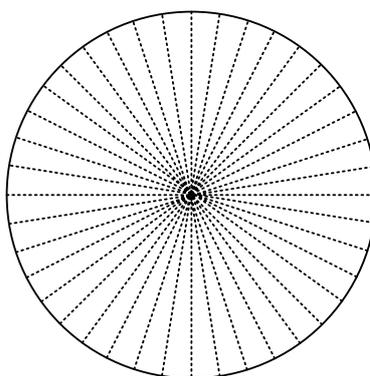
### Tarefa – A cor dos olhos<sup>1</sup>

Com esta tarefa pretende-se desenvolver as propriedades do gráfico circular<sup>2</sup>.

Os dados resultantes da observação da variável *Cor dos olhos* numa amostra de alunos, são apresentados, já organizados, na seguinte tabela de frequências:

Cor dos olhos	Pretos	Castanhos	Azuis	Verdes
Nº de alunos	12	18	7	3

1. A quantos alunos foi observada a cor dos olhos?
2. Representa graficamente os dados no seguinte gráfico circular e pinta com lápis de cor as zonas referentes a cada categoria.

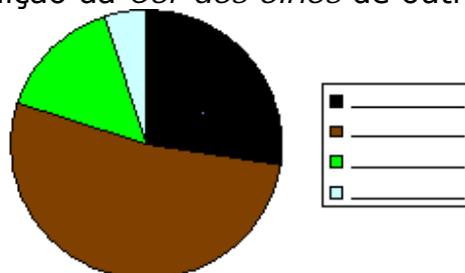


3. Supõe que duplicavas o número de alunos em cada categoria da variável *Cor de olhos*. Qual o aspecto do gráfico circular que representa estes novos dados? Justifica a tua resposta.
4. Supõe que duplicavas o número de alunos que tem olhos pretos, isto é, a tabela de frequências é agora a seguinte:

Cor dos olhos	Pretos	Castanhos	Azuis	Verdes
Nº de alunos	24	18	7	3

O ângulo do sector circular correspondente à categoria olhos Pretos, também duplicava? Justifica a resposta.

5. Supõe que o professor te apresentava o seguinte gráfico circular, que representa a distribuição da *Cor dos olhos* de outros 40 alunos:



Completa a legenda anterior, sabendo que:

- a moda é a cor Castanha;
- o número de alunos com olhos Azuis é 2;
- o número de alunos com olhos Verdes é o triplo do de olhos Azuis.

<sup>1</sup> Activalea 13 – [www.alea.pt](http://www.alea.pt)

<sup>2</sup> Esta actividade é especialmente adequada para ser resolvida no Excel, pois permite visualizar imediatamente no gráfico circular, qualquer alteração processada na tabela de frequências.



### Tarefa – Os jogos olímpicos de Atenas, em 2004

Esta tarefa tem como objectivo desenvolver a capacidade de resumir a informação contida em dados com distribuição enviesada

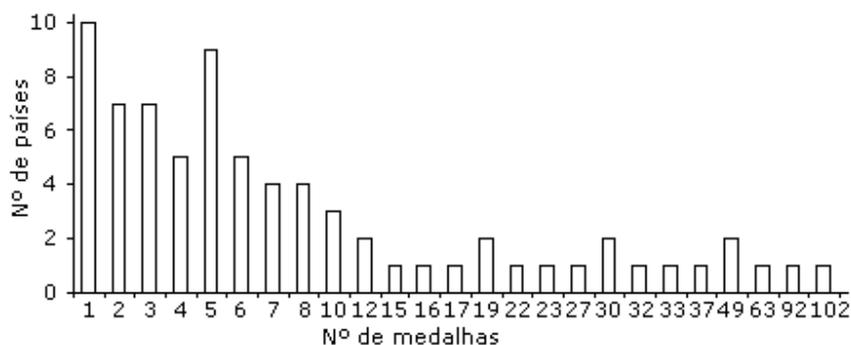
Nos jogos olímpicos de 2004, realizados em Atenas, foram distribuídas 929 medalhas, pelos 74 países medalhados. Nestes jogos Portugal ganhou 3 medalhas, 2 de prata e 1 de bronze.

- a) Sabes quem ganhou as 3 medalhas de Portugal e em que modalidades?  
(Se não souberes pesquisa em <http://www.olympic.org/uk/games>)
- b) Em média, quantas medalhas ganhou cada um dos países medalhados?
- c) Nestes jogos os países que mais se distinguiram foram os EUA com 102 medalhas, a Rússia com 92, a China com 63 e a Alemanha e Austrália com 49 medalhas, cada um.
  - i) Quantas medalhas ganharam ao todo estes 5 países? Quantas medalhas ganharam os 69 países restantes?
  - ii) Em média, quantas medalhas ganhou cada um dos 69 países referidos na alínea anterior?
- d) De acordo com as respostas que deste às alíneas anteriores, achas que a média é, neste caso, uma boa medida de localização do centro da distribuição dos dados?
- e) A variável *Número de medalhas ganhas por cada país medalhado* é discreta ou contínua?
- f) A seguir apresentamos a tabela de frequências do conjunto de dados em estudo:

Tabela 1

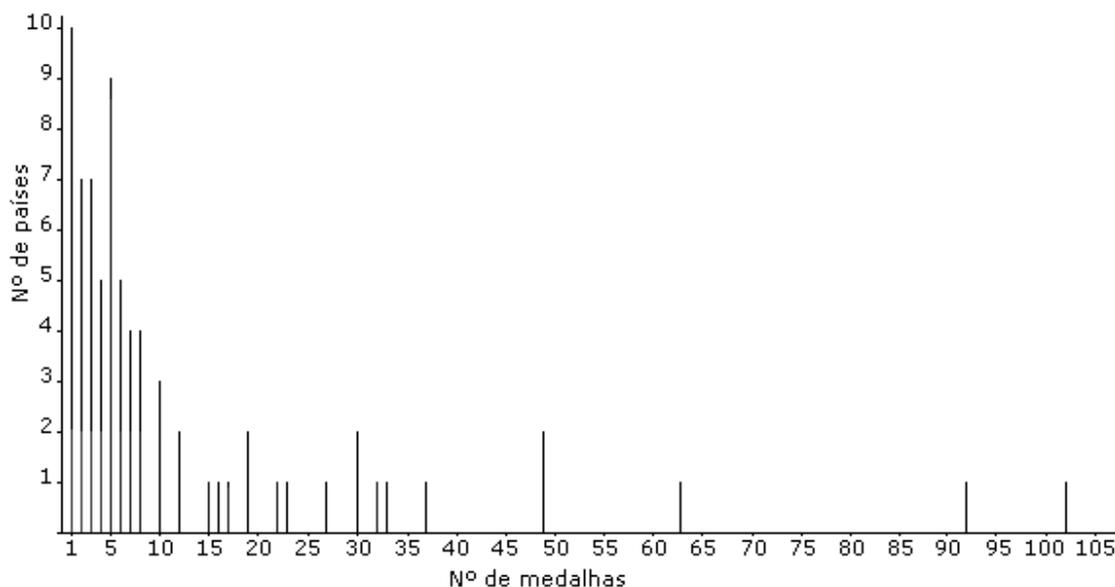
Nº de medalhas	Nº de países	Nº de medalhas	Nº de países
1	10	19	2
2	7	22	1
3	7	23	1
4	5	27	1
5	9	30	2
6	5	32	1
7	4	33	1
8	4	37	1
10	3	49	2
12	2	63	1
15	1	92	1
16	1	102	1
17	1		

- g) Com os dados da Tabela 1, alguém construiu a seguinte representação gráfica:



Achas que o gráfico anterior, que pretende representar a distribuição do *Número de medalhas ganhas por cada país medalhado* está correcta? Porquê?

h) Outra representação gráfica para os dados é a seguinte:



Que nome se dá à representação anterior? A distribuição dos dados é simétrica ou enviesada?

- i) Tendo em conta a representação anterior e sem fazeres quaisquer cálculos, qual a relação de grandeza que esperas que haja entre a média e a mediana? Serão aproximadamente iguais ou não? Neste caso qual é a medida que tu esperas que seja maior?
- j) Embora a representação mais comum para os dados discretos seja o diagrama de barras, existem outras representações que podem ser utilizadas, como por exemplo o caule e folhas, o diagrama de extremos e quartis e até o histograma, apesar desta representação ser mais adequada para dados de tipo contínuo. No entanto, como temos muitos dados e o número de valores distintos é razoavelmente grande, justifica-se aqui a construção de um histograma. Procede a um agrupamento dos dados de acordo com as seguintes classes e completa a tabela de frequências:

Classes	Freq.absoluta	Freq.relativa	Freq.rel.acumulada
[0, 10[			
[10, 20[			
[20, 30[			
[30, 40[			
[40, 50[			
[50, 60[			
[60, 70[			
[70, 80[			
[80, 90[			
[90, 100[			
[100, 110[			
Total			

- k) Constrói o histograma para os dados em estudo e interpreta o gráfico obtido.
- l) A partir da Tabela 2 consegues dizer, sem fazeres quaisquer cálculos, qual o intervalo onde estão a mediana e os 1º e 3º quartis? Justifica a tua resposta.
- m) Utilizando ainda a Tabela 2 calcula um valor aproximado para a média do *Número de medalhas ganhas por cada país medalhado*. Compara o valor obtido com o que obtiveste na alínea b). O que concluis? (*Não te esqueças que, efectivamente, quando temos os dados originais, e precisamos de conhecer a média, não se deve ir calculá-la a partir dos dados agrupados, pois o erro cometido pode ser razoável, como verificaste neste exemplo*).
- n) Constrói o diagrama de extremos e quartis e interpreta a representação obtida.
- o) Nas alíneas k) e n) construístes duas representações para os mesmos dados e obtiveste representações que te dão o mesmo tipo de informação sobre a estrutura subjacente aos dados.
- 1) Qual das duas representações foi mais simples de construir?
  - 2) Se te pedissem para organizar os dados originais em classes, escolhias necessariamente as classes que te foram apresentadas?
  - 3) Se pedisses a duas pessoas diferentes para, a partir da Tabela 1, organizarem os dados na forma de um histograma ou de um diagrama de extremos e quartis, qual das duas representações vinha necessariamente igual?
- Da alínea anterior podes concluir que a construção do histograma depende, em grande parte, da pessoa que o está a construir. Esta particularidade faz com que o histograma não seja aquilo a que se chama uma figura resistente, pois o aspecto resultante depende do número de classes e da amplitude de classe que se considera.*



### Tarefa – Fazes a cama quando te levantas?

Esta tarefa tem como objectivo desenvolver a compreensão da utilização do diagrama de Venn e de Carroll para resumir a informação e a utilização destas representações para facilitar o cálculo de probabilidades

Num inquérito a alunos de uma escola secundária, 490 de 564 raparigas interrogadas, responderam que **sim** à pergunta “Fazes a cama quando te levantas?”. Dos 476 rapazes interrogados, 425 responderam afirmativamente à mesma questão. Com a informação anterior completa o seguinte diagrama de Carroll:

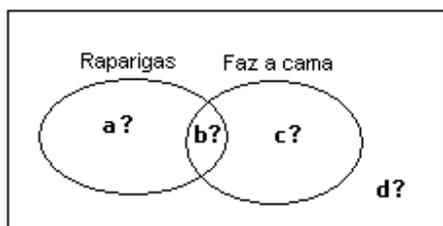
	Faz a cama	Não faz a cama
Rapariga		
Rapaz		

Depois de completares a tabela anterior, junta mais uma linha e uma coluna, para colocares os totais:

	Faz a cama	Não faz a cama	Totais
Rapariga			
Rapaz			
Totais			

Responde às seguintes questões:

- Quantos alunos participaram no inquérito?
- Quantos alunos fazem a cama?
- Quantos alunos não fazem a cama?
- Quantos rapazes não fazem a cama?
- Quantas raparigas não fazem a cama?
- Quem é que tu pensas que faz a cama com mais frequência? Os rapazes ou as raparigas?
- Calcula a percentagem de raparigas e de rapazes que fazem a cama quando se levantam. Qual a percentagem mais elevada? A resposta que deste à questão anterior, está de acordo com os valores que obtiveste nesta questão?
- Preenche o seguinte diagrama de Venn e responde às questões seguintes:



O que significam e quanto valem (em números inteiros) os espaços representados pelas letras (Toma como exemplo a letra a):

- a - Número de raparigas que não faz a cama = 74  
 b.-. \_\_\_\_\_  
 c.-. \_\_\_\_\_  
 d.-. \_\_\_\_\_

- Quantos alunos têm o hábito de fazer a cama?
- Se escolhesses um aluno ao acaso (de entre os inquiridos), seria mais provável que fosse rapaz ou rapariga?
- Se escolhesses um aluno ao acaso, seria mais provável que fosse um dos que faz a cama ou dos que não faz a cama?
- Se escolhesses um aluno de entre os que faz a cama, seria mais provável que fosse rapaz ou rapariga? Justifica a tua resposta.



## Tarefa – Quanto tempo governaram os presidentes da república portugueses?<sup>3</sup>

Esta tarefa tem como objectivo interpretar o diagrama de extremos e quartis – forma, simetria e variabilidade

Para responder à questão em estudo, recolheu-se a informação necessária na página oficial da Presidência da República Portuguesa<sup>4</sup>. Parte dessa informação é apresentada a seguir:

### Presidentes

2006/ ... - Cavaco Silva  
1996/2006 - Jorge Sampaio  
1986/1996 - Mário Soares  
1976/1986 - Ramalho Eanes  
1974/1976 - Costa Gomes  
1974/1974 - António de Spínola  
1958/1974 - Américo Tomás  
1951/1958 - Craveiro Lopes  
1926/1951 - Óscar Carmona  
1926/1926 - Gomes da Costa  
1926/1926 - Mendes Cabeçadas  
1925/1926 - Bernardino Machado  
1923/1925 -Teixeira Gomes  
1919/1923 - António José de Almeida  
1918/1919 - Canto e Castro  
1917/1918 - Sidónio Pais  
1915/1917 - Bernardino Machado  
1915/1915 - Teófilo Braga  
1911/1915 - Manuel de Arriaga

1. A tabela anterior mostra os sucessivos Presidentes da República de Portugal, desde a implantação da República. Sabes em que data foi implantada a República? Até essa data, qual o regime que vigorava em Portugal?

2. Consultando a fonte de informação indicada na introdução, investiga quanto tempo, em meses, esteve na presidência cada um dos presidentes. Por exemplo, o presidente Teófilo Braga esteve desde 29 de Maio a 5 de Outubro de 1915, o que perfaz 4 meses e 6 dias, ou seja 4,2 meses (6 dias é 0,2 (=6/30) meses). A partir dos dados obtidos, responde às seguintes questões:

2.1. Qual ou quais os presidentes que estiveram mais tempo na presidência?

2.2. Qual ou quais os presidentes que estiveram menos tempo na presidência?

2.3. Consegues detectar algum período bastante conturbado da vida política portuguesa? Justifica a tua resposta.

2.4. Constrói um diagrama de extremos e quartis para os tempos em que os presidentes estiveram na presidência da república. Interpreta a representação obtida.

<sup>3</sup> Activalea 21 - [www.alea.pt](http://www.alea.pt)

<sup>4</sup> <http://www.museu.presidencia.pt/presidentes.php>



### Tarefa – O lançamento da roleta para ajudar a compreender a aleatoriedade e a variabilidade<sup>5</sup>

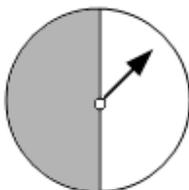
Esta tarefa tem como objectivo compreender a variabilidade e a aleatoriedade

Apresenta-se a seguir um exemplo de um teste de avaliação que o professor pode aplicar na aula aos seus alunos para averiguar a compreensão de alguns conceitos básicos associados à probabilidade e à estatística.

- Questão 1. a) O que entendes por *variabilidade*?  
 b) Utiliza a palavra *variabilidade* numa frase  
 c) Dá um exemplo de algo que *varie*

- Questão 2. a) O que entendes por *aleatório*?  
 b) Dá um exemplo de algo que aconteça de forma aleatória

O professor leva para a sala de aula uma roleta como a que se apresenta na figura

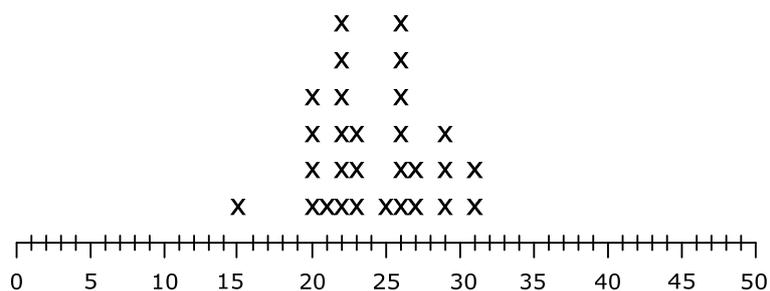


- Questão 3. Se rodares a roleta uma vez, qual a *probabilidade* de a seta cair na zona sombreada?
- Questão 4. Se rodares a roleta 50 vezes, quantas vezes esperas que a seta caia na zona sombreada? Porquê?
- Questão 5. Se rodares de novo a roleta 50 vezes, esperas obter o mesmo valor que indicaste na questão anterior, para o número de vezes que esperas que a seta caia na zona sombreada?
- Questão 6. Quais os valores que te surpreenderiam, se os obtivesses como resultado de rodares a roleta 50 vezes?
- Questão 7. Supõe que rodas a roleta 50 vezes e registas o número de vezes que a seta aponta a zona sombreada e que procedes da forma anterior 6 vezes. Aponta os valores que descrevam os resultados que poderias obter

\_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_

- Questão 8. Supõe que cada aluno da turma rodou a roleta 50 vezes e que os resultados do número de vezes que a seta apontou a zona sombreada se registaram no seguinte gráfico:

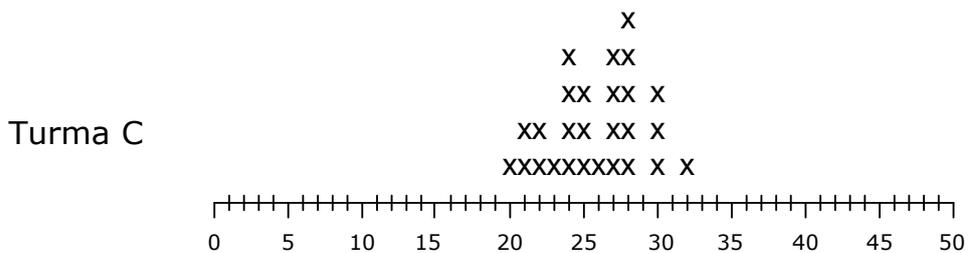
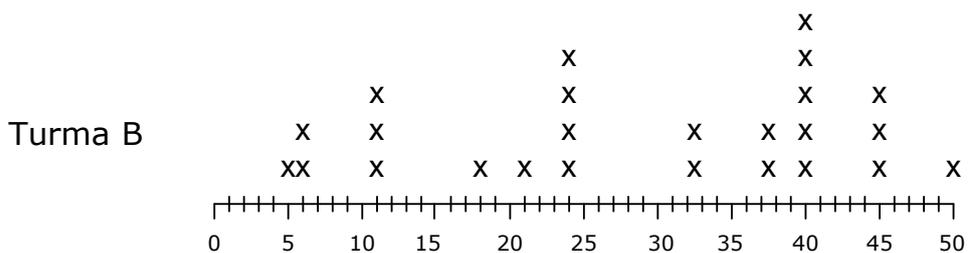
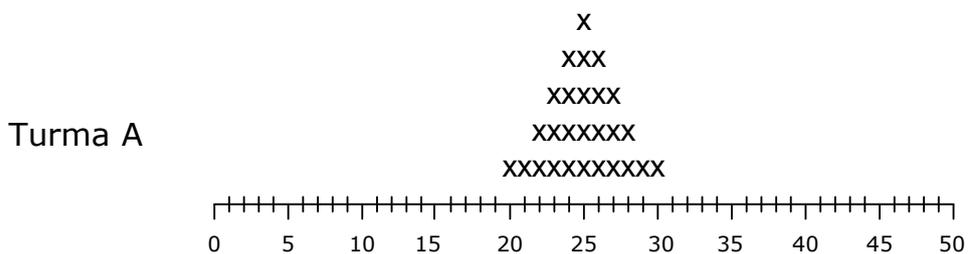
<sup>5</sup> Adaptado de Thinking and Reasoning with Data and Chance, NCTM, 2006, pag. 64



- a) Quantos alunos estavam na turma?
- b) Qual foi o menor valor obtido?
- c) Qual foi o maior valor obtido?
- d) Qual é a amplitude dos valores obtidos?
- e) Qual é a moda dos valores obtidos?

Questão 9. Imagina que outras três turmas apresentaram gráficos semelhantes ao anterior, em que era pressuposto estarem representados os resultados da experiência de rodar várias vezes a roleta 50 vezes. Desconfia-se que terá havido "batota" e algum(a) da(s) turma(s) em vez de realizarem a experiência, inventaram os resultados.

Alguns dos resultados não resultaram da realização da experiência? Porquê?



### Tarefa - Um jogo com dois dados<sup>6</sup>.

Esta tarefa tem como objectivo comparar a probabilidade teórica e a probabilidade experimental de um acontecimento.

Uma boa actividade introdutória ao estudo das probabilidades é apresentar este jogo aos alunos e perguntar-lhes se lhes parece que algum dos jogadores está em vantagem.

#### JOGO DOS DOIS DADOS

- Dois jogadores.
- Em cada jogada, cada jogador lança um dado e somam-se os pontos dos dois dados.
- O jogador **A** marca um ponto se a soma for 5, 6, 7 ou 8.
- O jogador **B** marca um ponto se a soma for 2, 3, 4, 9, 10, 11 ou 12.
- Ganha quem primeiro obtiver 20 pontos.

Depois de ouvir as opiniões dos alunos mas antes de as discutir, propor que eles façam alguns jogos. Para isso, devem organizar-se em grupos de dois, escolhendo entre si qual deles é o jogador A e qual é o B.

Uma boa parte dos alunos prefere ser o jogador B porque, das onze somas possíveis, há sete que fazem o jogador B ganhar e só quatro que o fazem perder. Um pouco apressadamente concluem que a probabilidade de ganhar seria  $\frac{7}{11}$ .

Depois de cada aluno receber um dado, cada grupo de alunos faz um jogo.

Se o professor não dispuser de dados suficientes, pode-se usar a calculadora gráfica para simular o lançamento dos dados.

Na TI-83 carregamos na tecla MATH e em PRB escolhemos 5:randInt(. Depois escrevemos, separados por vírgulas, os limites entre os quais queremos que a máquina escolha números inteiros ao acaso: 1 e 6. Como queremos o resultado de dois dados, acrescentamos mais uma vírgula e o número 2. Agora, cada vez que carregarmos em ENTER aparecem dois números correspondentes aos dois dados.

```
MATH NUM CPX PRB
1:rand
2:nPr
3:nCr
4:!
5:randInt(
6:randNorm(
7:randBin(
```

```
randInt(1,6,2)
(3 6)
(3 1)
(5 2)
(2 1)
(6 5)
(4 1)
```

Somando os dois números, vemos se foi o jogador A ou o jogador B a ganhar. Neste exemplo, o jogador A marcou pontos no 2º, 3º e 6º lançamentos.

Terminado o jogo, cada grupo vai ao quadro registar o seu resultado numa tabela com o seguinte aspecto.

	Jogador A	Jogador B
	20	14
	19	20
	20	16
	...	...
Total	274	223

<sup>6</sup> Graça Martins et al (1999), p.44

Normalmente, o jogador A ganhará a maior parte dos jogos. Isto faz-nos suspeitar que A está em vantagem. Além disso, a soma dos pontos de todos os jogos, é também maior para A. No exemplo que aqui apresentamos, vemos que A fez 274 pontos e B fez 223.

Houve  $274 + 223 = 497$  jogadas. Então, as frequências relativas das jogadas vitoriosas para cada jogador são:

$$f_A = \frac{274}{497} \approx 0.551 \quad f_B = \frac{223}{497} \approx 0.449$$

Em seguida, o professor pode propor aos alunos que procurem mostrar que realmente o jogador A está em vantagem. Se necessário, ir indicando pistas:

Será a soma "2" tão fácil de acontecer como a "7"? Só sai "2" se em ambos os dados sair 1, enquanto que "7" é possível de várias maneiras: 1+6 ou 2+5 ou 3+4 ou ...

Por outro lado, sair 3 num dado e 4 no outro é diferente de sair 4 no primeiro e 3 no segundo...

Pedir em seguida aos alunos que identifiquem os dados – por exemplo, dado azul e dado vermelho – e façam uma tabela de duas entradas com todos os casos possíveis.

		Dado Vermelho					
		1	2	3	4	5	6
Dado azul	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Vê-se então que há 36 casos elementares possíveis e organiza-se um quadro com o número de casos favoráveis para cada resultado.

Resultado	2	3	4	5	6	7	8	9	10	11	12
Casos favoráveis	1	2	3	4	5	6	5	4	3	2	1

Agora já podemos ver se algum jogador tem vantagem.

O jogador A ganha se sair 6, 7, 8 ou 9.

Os casos favoráveis a A são  $5+6+5+4 = 20$ .

O jogador B ganha saindo 2, 3, 4, 5, 10, 11 ou 12.

Os casos favoráveis a B são  $1+2+3+4+3+2+1 = 16$ .

Conclui-se então que o jogo é favorável ao jogador A, apesar de só lhe servirem quatro resultados. A probabilidade de ele ganhar uma jogada é  $\frac{20}{36}$  ou 55.6%.

Para o jogador B, a probabilidade de ganhar é  $\frac{16}{36}$  ou 44.4%.

Esta actividade pode ser formalmente apresentada da seguinte forma: Considere a experiência aleatória que consiste em lançar dois dados e em verificar a soma das pintas das faces que ficam viradas para cima. Qual a probabilidade de se obter um 6, 7, 8 ou 9? Como o espaço de resultados  $S$  associado a esta experiência é constituído por  $S = \{(1,1), (1,2), \dots, (1,6), (2,1), (2,2), \dots, (2,6), (3,1), (3,2), \dots, (3,6), (4,1), (4,2), \dots, (4,6), (5,1), (5,2) \dots, (6,6), (6,1), (6,2), (6,6)\}$ , todos eles igualmente possíveis, se os dados forem equilibrados, o acontecimento  $D$ , que faz com que a soma das pintas seja a pretendida, é constituído pelos resultados  $D = \{(1,5), (2,4), (3,3), (4,2), (5,1), (1,6), (2,5), (3,4), (4,3), (5,2), (6,1), (2,6), (3,5), (4,4), (5,3), (6,2), (3,6), (4,5), (5,4), (6,3)\}$ , pelo que a probabilidade pretendida é  $\frac{20}{36}$ .

## Tarefa – Qual a probabilidade de obter uma soma superior a 13, quando se lançam 3 dados?<sup>7</sup>

Esta tarefa tem como objectivo comparar a probabilidade teórica com a probabilidade experimental de um acontecimento.

Num certo jogo, lançam-se três dados normais e ganha-se quando a soma das pintas é maior que 13. Qual é a probabilidade de ganhar?

Há vários processos de descobrir esta probabilidade, uns experimentais, outros teóricos. Quando o cálculo teórico é muito trabalhoso, difícil ou mesmo impossível, recorre-se aos métodos experimentais para obter um valor aproximado. Vamos ver aqui vários desses processos e no fim determinaremos o valor exacto

### 1º Processo – Experimentação directa

Pegam-se em três dados, lançam-se muitas vezes e de cada vez regista-se o resultado da soma. Ao fim de muitas experiências (que podem ir sendo feitas simultaneamente por várias pessoas diferentes), calcula-se a frequência relativa dos resultados maiores que 13. Se o número de experiências for suficientemente grande, esta frequência é uma boa estimativa da probabilidade.

### 2º Processo – Simulação com a calculadora

Em vez de usar os dados, podemos fazer uma simulação com a calculadora, pedindo para ela gerar um conjunto de três números aleatórios entre 1 e 6, inclusive. Cada um destes números corresponde a um dado. Cada vez que carregarmos em ENTER aparece-nos um conjunto de três números que temos de somar para ver se o resultado é maior que 13.

```

MATH NUM CPX [MODE]
1:rand
2:nPr
3:nCr
4:!
5:randInt(
6:randNorm(
7:randBin(

randInt(1,6,3)
(6 5 3)
(3 3 3)
(6 2 4)
(4 4 1)
(2 2 4)

(3 4 4)
(6 5 2)
(4 2 3)
(4 1 5)
(1 4 3)

```

Podemos evitar o trabalho de somar os três números. Com a instrução `sum(`, a máquina efectua imediatamente a soma dos três números da lista, embora assim deixemos de saber que números saíram efectivamente nos dados.

```

y LIST
NAMES OPS [MODE]
1:min(
2:max(
3:mean(
4:median(
5:sum(
6:Prod(
7:stdDev(

sum(randInt(1,6,
3)
13
13
12
16
6

11
12
13

```

<sup>7</sup> Graça Martins et al (1999), p.48

Cada vez que carregamos em ENTER obtemos um número entre 3 e 18. Para evitar enganos e maior facilidade da contagem, é aconselhável fazer aparecer cinco resultados de cada vez. Depois de registrar os resultados, faz-se CLEAR, obtêm-se mais cinco resultados, e assim sucessivamente. Na figura anterior temos os resultados de 10 experiências, em que só uma vez a soma foi maior que 13.

Se houver um grupo de alunos a fazer isto simultaneamente, rapidamente se consegue um grande número de experiências.

### 3º Processo – Programa de simulação com a calculadora

É possível usar um programa muito simples que faça todo o trabalho anterior por nós. Em anexo neste livro está o programa DADOS3 que faz precisamente isto. Chamamos o programa, indicamos quantas experiências queremos fazer e passado uns momentos a máquina indica-nos o número de experiências e a frequência relativa de resultados maiores que 13

<pre>PrgrmDADOS3</pre>	<pre>NUMERO DE EXPERIENCIAS? ?100</pre>	<pre>EXPERIENCIAS 100 FREQ REL= .15</pre>
------------------------	---	---

Começámos com 100 experiências e a frequência é de 0,15. Mas este número de experiências é demasiado pequeno para podermos ter confiança no resultado. Então, carregando em ENTER, aparece um menu que permite continuar a simulação. Acrescentamos mais 900 experiências, para que o total passe a ser 1000.

<pre>SIMULACAO 1: PARAR 2: CONTINUAR 3: INICIAR</pre>	<pre>NUMERO DE EXPERIENCIAS? ?900</pre>	<pre>EXPERIENCIAS 1000 FREQ REL= .167</pre>
---	---	---

Nesta simulação, a frequência foi de 0,167. É de esperar que a probabilidade de ganhar neste jogo seja um valor bastante próximo deste.

É de referir que este programa faz cerca de 500 experiências num minuto.

Prolongámos a simulação até às 10000 experiências e a frequência foi de 0,1651.

### 4º Processo – Cálculo teórico

Os processos anteriores só nos dão valores aproximados da probabilidade pedida, valores esses tanto mais fiáveis quanto maior tiver sido o número de experiências feito.

No entanto, podemos obter o valor exacto da probabilidade fazendo o cálculo teórico. Para isso temos de calcular o número de casos possíveis quando se lançam três dados e o de casos favoráveis, que correspondem a somas maiores que 13.

$$\text{Casos possíveis} = 6^3 = 216$$

Antes de contabilizar os casos favoráveis, convém contar o número de maneiras diferentes com que pode aparecer um conjunto de três números:

- 1) Números todos iguais (por exemplo 5-5-5)  
só há uma maneira: 5-5-5.
- 2) Dois iguais e um diferente (por exemplo 6-6-5)  
três maneiras: 6-6-5, 6-5-6, 5-6-6.
- 3) Todos diferentes (por exemplo 6-5-4)  
seis maneiras: 6-5-4, 6-4-5, 5-6-4, 5-4-6, 4-6-5, 4-5-6.

Façamos um quadro para as várias somas maiores ou iguais a 14.

Soma	Tipo	Nº de casos
18	6 - 6 - 6	1
17	6 - 6 - 5	3
16	6 - 6 - 4	3
	6 - 5 - 5	3
15	6 - 6 - 3	3
	6 - 5 - 4	6
	5 - 5 - 5	1
14	6 - 6 - 2	3
	6 - 5 - 3	6
	6 - 4 - 4	3
	5 - 5 - 4	3
Total		35

Agora já podemos determinar a probabilidade:

$$P(\text{soma} > 13) = \frac{35}{216} \approx 0.162$$



# Bibliografía



## Bibliografia

Na preparação desta Brochura seguiu-se essencialmente a seguinte bibliografia:

Burrill G. ed. (2006) - *Thinking and Reasoning with Data and Chance*. NCTM, Inc. Reston

Chance B. (2002) -Components of Statistical Thinking and Implications for Instruction and Assessment, *Journal of Statistics Education*, Volume 10, Number 3.

Chatfield, C. (1991). Avoiding Statistical Pitfalls. *Statistical Science*, 6, 3, 240-268.

De Veaux, R. e Velleman, P. (2004) – *Intro Stats*, Pearson Education, Inc.

Feller, W. (1968) – *An Introduction to Probability Theory and its Applications*, John Wiley & Sons.

Freedman, D., Pisani, R., Purves, R., Adhikari, A. (1991). - *Statistics*. W. W. Norton & Company.

Gaise Report (2005) - *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) Report, A Pre-k-12 curriculum framework, August 2005 –American Statistical Association

Gal, I. (2002). Adult's Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, 70, 1, 1-51.

Graça Martins, M. E., Branco, J. (2000) – Literacia Estatística. Revista da APM

Graça Martins, M. E., Monteiro, C., Viana, J. P., Turkman, M. A. A. (1999a) – *Probabilidades e Combinatória*, Ministério da Educação, Departamento do Ensino Secundário.

Graça Martins, M. E., Cerveira, A. (1999b) – *Introdução às Probabilidades e à Estatística*, Universidade Aberta.

Graça Martins, M. E., Monteiro, C., Viana, J. P., Turkman, M. A. A. (1997) – *Estatística*, Ministério da Educação, Departamento do Ensino Secundário.

Jenni Way, (1997) - *Way in Reflections*, vol. 21, nº 1, May 97, Research in Probability and Statistics: Reflections and Directions. In D. Grouws (Ed.), Handbook of Research on Mathematics Teaching and Learning.

Mann, P. (1995) – *Introductory Statistics*. John Wiley & Sons.

Mendenhall. W., Beaver, R. (1994) – *Introduction to Probability and Statistics*. Duxbury Press.

Moore, D. (1997). New Pedagogy and New Content. The Case of Statistics. *International Statistical Review*, 65, 2, 123-165

Moore, D. (1997) – *Statistics – Concepts and Controversies*. Freeman.

Moore, D. (1996) – *The Basic Practice of Statistics*, Freeman.

Moore, D., McCabe, G. (1996) – *Introduction to the Practice of Statistics*, Freeman.

Rasfeld, P. (2001). The Role of Statistics in School Mathematics Teaching Today.

*International Journal for Mathematics Teaching and Learning*,  
<http://www.cimt.plymouth.ac.uk/journal/default.htm>.

Rossman, A. and Chance B. (2001) – *Workshop Statistics , Discovery with data*. Key College Publishing. Emeryville, CA.

Sheaffer, R. L. et al (2004) – *Activity-based statistics* – student guide. Key College Publishing, USA.

Sheaffer, R. (2001) – Quantitative Literacy and Statistics. *Amstat News* 293, Nov 2001, 3-4

Steen, L. A., ed (1997). *Why Numbers Count: Quantitative Literacy for Tomorrow America*. The College Board. New York.

Steen, L. A., ed (2001). *Mathematics and Democracy: The case for Quantitative Literacy*. Prepared by the National Council on Education and the Disciplines. Princeton.

Tannenbaum, P. and al. - *Excursions in modern Mathematics*, Prentice Hall, 1998.

Vere-Jones, D. (1995). The Coming of Age of Statistical Education. *International Statistical Review*, 63, 1, 3-23.

Vicente, P., Reis, E., Ferrão, F. – *Sondagens*, Edições Sílabo, Lda, 1996

Watson J. M. (2006) - Issues for Statistical Literacy in the Middle School, *ICOTS-7*

### **Páginas na Internet**

ALEA - <http://www.alea.pt>

Instituto Nacional de Estatística - [www.ine.pt](http://www.ine.pt)

(Tem informação sobre Portugal, ao nível da freguesia)

Eurostat – [europa.eu.int/comm/eurostat/](http://europa.eu.int/comm/eurostat/)

(Tem informação relativa aos diversos países da Europa)

World Health Organization – <http://www.who.int/research/en/>

(Tem informação sobre temas ligados à saúde, para todos os países do mundo)

World in figures – [http://.stat.fi/tup/maanum/index\\_en.html](http://.stat.fi/tup/maanum/index_en.html)

(Tem informação das mais diversas áreas, tais como população e estatísticas vitais, cultura, religiões, emprego, consumo, etc., relativa a todos os países do mundo)