Organização e tratamento de dados

no programa de

Matemática

Ensino básico

Maria Eugénia Graça Martins
(memartins@fc.ul.pt)

Dezembro 2015

"Data are numbers, but they are not "just numbers". Data are numbers with a context. …. Because data are numbers with a context, doing statistics means more than manipulating numbers"

In Moore, STATISTICS Concepts and Controversies, página XV

MAPA

Voltar

Apresentação

Este curso de *Organização e tratamento de dados* - **OTD**, aborda os conteúdos deste domínio no Programa de Matemática do ensino básico (homologado em 17 de Junho de 2013), transcritos nos slides seguintes.

Foi construído com o objectivo de colmatar alguma ausência de textos sobre o assunto, que possam ajudar o professor na sua prática lectiva.

Pode ser consultado pontualmente, sobre determinado assunto de interesse imediato, pelo que se decidiu incluir um slide a que se chamou MAPA, para uma melhor orientação na sua consulta.

Por muito cuidado que se tenha posto na sua elaboração, tem-se a consciência que há sempre gralhas que escapam, ou que alguns assuntos possam não estar tão claros quanto o necessário. Para o melhorar conto convosco, caros leitores, aos quais peço para me reportarem as situações apontadas anteriormente.

A autora

Maria Eugénia Graça Martins
(megm@ef.pt ou mematins@fc.ul.pt)

Dezembro 2015

OTD no programa de Matemática do Ensino Básico

OTD1	• Diagramas de Venn com conjuntos disjuntos. **Representação de dados** • Gráfico de pontos e pictograma em que cada figura representa uma unidade.

OTD2	• Diagramas de Venn e Carroll. **Representação de dados** • Tabelas de frequências absolutas, gráficos de pontos, de barras e pictogramas em diferentes escalas; • Esquemas de contagem (tally charts).

OTD no programa de Matemática do Ensino Básico

OTD3	**Representação e tratamento de dados** • Diagramas de caule-e-folhas; • Frequência absoluta; • Moda; • Mínimo, máximo e amplitude; • Problemas envolvendo análise e organização de dados, frequência absoluta, moda e amplitude.

OTD4	**Tratamento de dados** • Frequência relativa; • Noção de percentagem; • Problemas envolvendo o cálculo e a comparação de frequências relativas.

OTD no programa de Matemática do Ensino Básico

OTD5	**Representação e tratamento de dados** • Tabelas de frequências absolutas e relativas; • Gráficos de barras e de linhas; • Média aritmética; • Problemas envolvendo a média e a moda; • Problemas envolvendo dados em tabelas, diagramas e gráficos.

OTD6	**Representação e tratamento de dados** • População e unidade estatística; • Variáveis quantitativas e qualitativas; • Gráficos circulares; • Análise de conjuntos de dados a partir da média, moda e amplitude; • Problemas envolvendo dados representados de diferentes formas.

OTD no programa de Matemática do Ensino Básico

OTD7	**Medidas de localização** • Sequência ordenada dos dados; • Mediana de um conjunto de dados; definição e propriedades; • Problemas envolvendo tabelas, gráficos e medidas de localização.

OTD8	**Diagramas de extremos e quartis** • Noção de quartil; • Diagramas de extremos e quartis; • Amplitude interquartil; • Problemas envolvendo gráficos diversos e diagramas de extremos e quartis.

OTD9	**Histogramas** • Variáveis estatísticas discretas e contínuas; classes determinadas por intervalos numéricos; agrupamento de dados em classes da mesma amplitude; • Histogramas; propriedades; • Problemas envolvendo a representação de dados em tabelas de frequência e histogramas.

OTD no programa de Matemática do Ensino Básico

OTD9	**Probabilidade** • Experiências deterministas e aleatórias; universo dos resultados ou espaço amostral; casos possíveis; • Acontecimentos: casos favoráveis, acontecimento elementar, composto, certo, impossível; • Acontecimentos disjuntos ou incompatíveis e complementares; • Experiências aleatórias com acontecimentos elementares equiprováveis; • Definição de Laplace de probabilidade; propriedades e exemplos; • Problemas envolvendo a noção de probabilidade e a comparação de probabilidades de diferentes acontecimentos compostos, utilizando tabelas de dupla entrada e diagramas em árvore; • Comparação de probabilidades com frequências relativas em experiências aleatórias em que se presume equiprobabilidade dos casos possíveis

MAPA

5. Probabilidade

Fenómeno aleatório
Experiência aleatória
Espaço de resultados
Acontecimentos
Operações com acontecimentos
Modelo de probabilidade
Probabilidade de um acontecimento
Probabilidade frequencista
Probabilidade de Laplace
Regras para a probabilidade
Propriedades da probabilidade

Índice

1.Introdução

População. Variável. Dado
Amostra
Diagramas de Venn e de Carroll
Tipos de variáveis e de dados

OTD

4. Dados quantitativos

Diagrama de extremos e quartis

2. Tabelas de frequência. Representações gráficas

2.1 Dados qualitativos

Tabela de frequências
Gráfico de pontos
Pictograma
Esquema de contagem
Gráfico circular
Gráfico de barras
Moda

Dados quantitativos

2.2 Discretos

Tab. de frequências
Gráfico de pontos
Gráfico de barras
Caule-e-folhas

2.3 Contínuos

Tab.de frequências
Histograma
Caule-e-folhas

2.4 Dados quantitativos bivariados

Gráfico de dispersão
Gráfico de linhas
Tabela de frequências

3. Dados quantitativos – Medidas numéricas

3.1 Medidas de localização

Média
Mediana
Moda
Quartis

3.2 Medidas de dispersão

Amplitude
Amplitude interquartil

Índice

Índice

Índice

Índice

1 Introdução

- População. Unidade estatística. Variável. Dado
- Amostra. Dimensão da amostra. Conjunto de dados
- Diagramas de Venn e de Carroll
- Tipos de variáveis e Tipos de dados
- Ficheiro *Dados sobre casas*
- Distribuição

População. Unidade estatística. Variável. Dado.

Objectivo da Estatística

O objectivo da Estatística é o estudo de **Populações**.

População - conjunto de indivíduos (não necessariamente pessoas) com *características* comuns, que se pretendem estudar. Ao indivíduo ou coisa sobre o qual se estuda a *característica* de interesse dá-se o nome de *unidade observacional* ou ***unidade estatística***.

- A qualquer *característica* comum, à qual se possa atribuir um número ou uma categoria, podendo assumir valores (números ou categorias) diferentes de unidade estatística para unidade estatística, chamamos ***variável*.**
- O resultado da observação da variável, sobre a unidade estatística, é o **dado estatístico** ou simplesmente ***dado*.**

População. Unidade estatística. Variável. Dado.

Um princípio fundamental da Estatística é reconhecer a existência de **variabilidade.**

A variabilidade está presente em todas as situações do mundo que nos rodeia, pelo que as conclusões que tiramos a partir dos dados que se nos apresentam, têm inerente um certo grau de incerteza.

A Estatística trata e estuda esta variabilidade e o objectivo de qualquer estudo estatístico é a partir dos dados retirar conclusões e exprimir o grau de confiança que devemos ter nessas conclusões.

População. Unidade estatística. Variável. Dado.

Exemplo – Considere-se a seguinte população constituída pelos alunos de uma sala de aula, que são as *unidades estatísticas*. Algumas características comuns a todos os alunos são, por exemplo:

- Tipo de transporte utilizado para ir para a escola;
- Idade;
- Número de irmãos.

Todas estas características podem assumir valores diferentes de aluno para aluno, pelo que são designadas por *variáveis*. Se pensarmos na variável *Número de irmãos*, e perguntarmos a cada aluno quantos irmãos tem, obteremos os *dados*, que resultaram de observar a variável em estudo.

As características comuns *Ter nome* e *Ter cabeça* não são variáveis, pois todos os alunos têm nome e cabeça.

População. Unidade estatística. Variável. Dado.

Nota – É importante ter presente que para termos os **dados**, é pressuposto:

1) Termos uma população, cujos elementos denominados unidades estatísticas têm uma (ou mais) característica(s) comum(ns) que se pretende(m) estudar;
2) Esta característica pode assumir valores diferentes de unidade estatística para unidade estatística e chama-se variável;
3) Ao resultado da observação da variável sobre a unidade estatística chama-se dado.

População. Unidade estatística. Variável. Dado.

Exemplo - Consideremos a população constituída pelos livros existentes na biblioteca da escola e suponhamos que pretendemos estudar o número de páginas desses livros. Assim:

1) As **unidades estatísticas** são os livros;
2) Uma característica comum que interessa estudar é o número de páginas de cada livro, sendo esta, então, a **variável** em estudo;
3) Se o resultado da contagem do número de páginas de um destes livros for 54, então 54 é um **dado**.

(continua)

População. Unidade estatística. Variável. Dado.

Suponhamos agora que, no fim do ano lectivo, se perguntou aos alunos Pedro, Sara, Tiago, Maria, Vera, Rita, Manel, Miguel, Susana, Rute, José, Afonso e Ricardo quantos livros tinham requisitado na biblioteca, durante o ano.

Neste caso a população é constituída pelos 13 alunos:

1) As **unidades estatísticas** são os 13 alunos Pedro, Sara, Tiago, Maria, Vera, Rita, Manel, Miguel, Susana, Rute, José, Afonso e Ricardo;
2) Uma característica comum que estamos interessados em estudar é o número de livros requisitados por cada aluno, sendo esta a **variável** em estudo;
3) Os dados são o resultado da contagem do número livros requisitados por cada aluno. Por exemplo se o Pedro requisitou 4, então 4 é um **dado**.

(continua) MAPA

População. Unidade estatística. Variável. Dado.

Suponhamos ainda que o que se pretendia estudar era o número de livros requisitados na biblioteca, em cada mês do ano lectivo.

Neste caso já as unidades estatísticas são os meses, a variável de interesse será o número de livros requisitados em cada mês e os dados são os valores da tabela abaixo, que se obtiveram consultando as fichas de requisição:

Mês	**Nº livros**	**Mês**	**Nº livros**
Setembro	35	Março	85
Outubro	56	Abril	98
Novembro	77	Maio	63
Dezembro	54	Junho	52
Janeiro	65	Julho	27
Fevereiro	79		

Amostra. Dimensão da amostra. Conjunto de dados

Quando não é possível estudar todos os elementos da população recolhe-se uma amostra.

Amostra – conjunto constituído por algumas unidades estatísticas da população, que se observam com o objectivo de tirar conclusões para a população.

Dimensão da amostra – número de unidades estatísticas, que constituem a amostra.

À colecção ou conjunto constituído pelos *dados* que resultaram de observar todas as unidades estatísticas que compõem a amostra também se chama **amostra**.

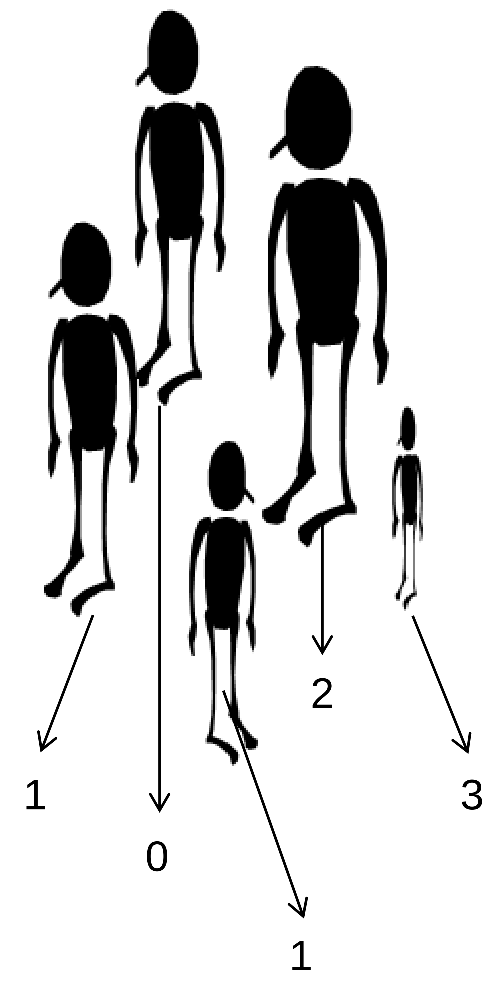
Assim, de um modo geral, quando falamos em **amostra**, estamos a referir-nos a um **conjunto de dados**.

Nota – Se observarmos todas as unidades estatísticas da população, também teremos um conjunto de dados.

Amostra. Dimensão da amostra. Conjunto de dados

Exemplo – Se a *população* considerada no exemplo do seguinte slide fosse muito grande, teríamos de seleccionar algumas *unidades estatísticas*, ou seja, alguns alunos, sobre os quais iríamos observar a *variável* em estudo, o "número de irmãos". Vamos admitir que se selecciona uma *amostra* de *dimensão* 5, ou seja, 5 alunos. Pergunta-se a cada um quantos irmãos tem e o resultado será um *conjunto de dados*, que poderia ser constituído pelos 5 dados seguintes

1, 0, 1, 2, 3



Amostra. Dimensão da amostra. Conjunto de dados

Uma vez que os *dados* resultam da observação de unidades estatísticas diferentes, dois *dados* podem ser numericamente iguais, mas como *dados* não são iguais uma vez que resultaram da observação de unidades estatísticas diferentes.

É importante reconhecer que dados são mais do que números, são números com um contexto, que se estudam com um determinado objectivo.

Nem todos os números são dados. Por exemplo, os números 1, 0, 1, 2, 3, considerados no exemplo anterior só são dados porque resultaram da observação da variável *Número de irmãos,* em 5 alunos.

Amostra. Dimensão da amostra. Conjunto de dados

Exemplo - Suponhamos que estamos interessados em estudar a variável "número de irmãos" de cada um dos alunos de uma turma (Neste caso, a nossa população é constituída pelos alunos da turma). Assim, perguntou-se a cada aluno quantos irmãos tinha e obtiveram-se os resultados registados na seguinte tabela

Maria	David	Joana	Pedro A.	Isabel	José	Miguel	Tomás	Ana	Inês
2	0	1	1	3	1	2	0	3	1
Ana F.	**Vera**	**Sara**	**Bernardo**	**Gui**	**Sofia**	**João**	**Tiago**	**Rita**	**Vicente**
0	2	1	1	4	2	1	0	2	1

Como em Estatística não nos interessa saber quem é que respondeu 0, 1, etc, deixamos "cair" os nomes e só ficamos com os resultados da observação, ou seja, os **dados**

2 0 1 1 3 1 2 0 3 1 0 2 1 1 4 2 1 0 2 1

Têm o mesmo valor numérico, mas são dados diferentes!!!!

Amostra. Dimensão da amostra. Conjunto de dados

Se em vez de dados, tivessemos simplesmente os números

2 0 1 1 3 1 2 0 3 1 0 2 1 1 4 2 1 0 2 1

então se representassemos por B o *conjunto* desses números, teríamos

$$B = \{0, 1, 2, 3, 4\}$$

Um ***conjunto de dados*** não é o mesmo que um ***conjunto de números***, pois não satisfaz, necessáriamente, uma das condições básicas quanto à enumeração dos elementos de um conjunto, nomeadamente a de não ter elementos repetidos. Em Estatística não existem elementos repetidos já que estamos a tratar com **dados** e **dados são mais que números, são números com um contexto**.

Amostra. Dimensão da amostra. Conjunto de dados

Exemplo - Caderno de apoio Matemática, 1º ciclo, **OTD2**, página 33

1.1
1.2

Exemplo*

Considera os seguintes conjuntos de números:

A = {2, 4, 6, 8, 10, 12, 14, 16, 18, 20}

B = {3, 6, 9, 12, 15, 18, 21, 24, 27, 30}

a. Determina o conjunto interseção de A com B.

b. Determina o conjunto reunião de A com B.

R.: a. O *conjunto interseção de A com B* é {6, 12, 18}; $A \cap B = \{6, 12, 18\}$

b. O *conjunto reunião de A com B* é

{2, 3, 4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 24, 27, 30};

$A \cup B = \{2, 3, 4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 24, 27, 30\}$

Como em Estatística trabalhamos com dados, vamos dar um contexto aos números anteriores.

Amostra. Dimensão da amostra. Conjunto de dados

Exemplo modificado - Considera os seguintes conjuntos de dados resultantes de observar quantos euros tinha no bolso cada um dos 10 alunos seleccionados nas turmas A e B:

A = {2, 4, 6, 8, 10, 12, 14, 16, 18, 20}

B = {3, 6, 9, 12, 15, 18, 21, 24, 27, 30}

- a. Quantos alunos da turma A tinham igual quantia aos da turma B?
- b. Considera o conjunto das quantias dos alunos das duas turmas. Quantos euros tinham os alunos no total?

R: a. 3 alunos da turma A tinham igual quantia a 3 alunos da turma B.

b. O conjunto das quantias dos alunos das duas turmas é o conjunto {2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30} e no total os alunos tinham 2+4+6+...+20+3+6+9+30=265 euros.

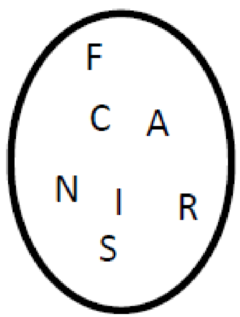
Amostra. Dimensão da amostra. Conjunto de dados

Exemplo - Caderno de apoio de Matemática, 1º ciclo, **OTD2**, página 33

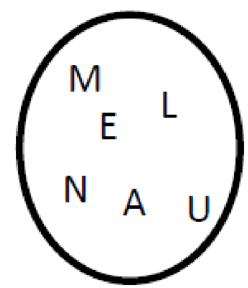
1.1 **Exemplo**

1.2 *Verifica quais são os elementos que pertencem a ambos os conjuntos.*

Conjunto das letras do nome **Francisca**



Conjunto das letras do nome **Manuel**



R.: Os elementos que pertencem aos dois conjuntos são o A e o N.

Vamos acrescentar a seguinte questão:

- *Qual a cardinalidade dos dois conjuntos anteriores*?

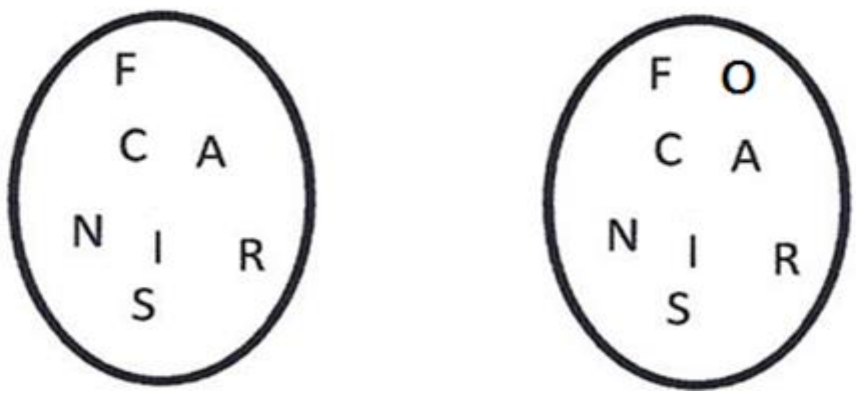
A resposta seria 7 e 6, respectivamente para o conjunto das letras do nome Francisca e para o conjunto das letras do nome Manuel

Amostra. Dimensão da amostra. Conjunto de dados

Suponhamos que em vez do nome Manuel tínhamos considerado o nome Francisco e colocavamos a seguinte questão:

Verifica qual dos conjuntos é mais numeroso

Conjunto das letras do nome **Francisca** | Conjunto das letras do nome **FranciscO**



O conjunto das letras do nome Francisco é mais numeroso que o conjunto das letras do nome Francisca...

Amostra. Dimensão da amostra. Conjunto de dados

Continuando ainda com o exemplo anterior, suponhamos que na sala de aula, o professor em conjunto com os alunos tinham decidido estudar a variável *tamanho dos nomes*, pelo que seria necessário recolher a informação sobre o número de letras do nome de cada aluno.

Dois dos alunos são o Francisco e a Francisca...

Então, concluíram, com alguma surpresa, que:

- O número de letras do nome Francisco é igual ao número de letras do nome Francisca.

- O **conjunto** das letras do nome Francisco é mais numeroso que o **conjunto** das letras do nome Francisca...

E ficaram um tanto perplexos...

Amostra. Dimensão da amostra. Conjunto de dados

NOTA

Pensamos que, ao contrário do que preconiza o programa, não é aconselhável dar a definição formal de conjunto no âmbito de OTD, já que pode baralhar os alunos, ao trabalharem com conjuntos de dados. Os autores do programa sentiram a necessidade de introduzir a noção do **conjunto**, para poderem utilizar os diagramas de Venn.

No entanto, em OTD não se tem necessidade de uma definição formal de conjunto, para trabalhar com os diagramas de Venn.

Vejamos que assim é... nas fichas seguintes.

Diagramas de Venn e de Carroll

Os diagramas de Venn e de Carroll, não sendo propriamente processos de organização de dados, são, no entanto, instrumentos particularmente adequados para os alunos mais novos procederem a uma classificação rápida de *dados*, *números* ou *objectos*.

Diagramas de Venn e de Carroll

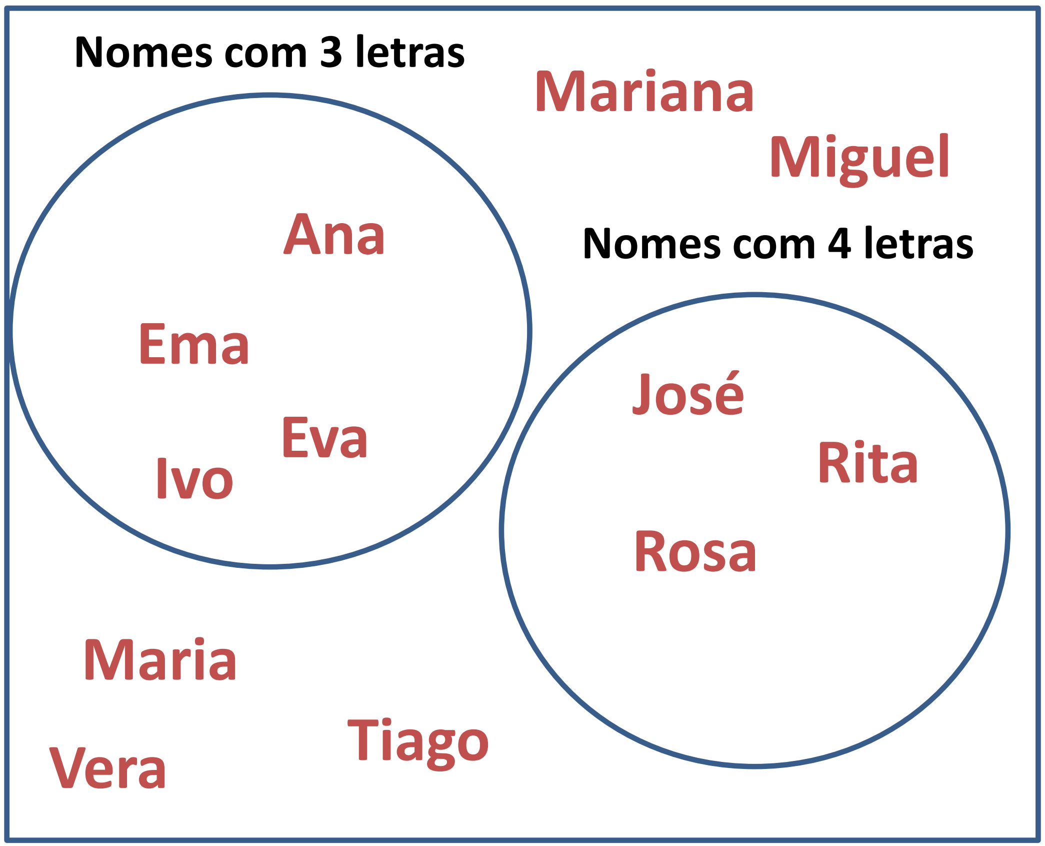
Os **diagramas de Venn** são representações gráficas que utilizam círculos ou rectângulos para uma classificação rápida de dados, números ou objectos, que partilhem características comuns. Usualmente, considera-se um rectângulo que representa todo o conjunto a ser classificado, e dentro desse rectângulo consideram-se círculos que representam os elementos com as características de interesse (Graça Martins, M. E. et al, Organização e tratamento de dados, página 43)(1).

(1) http://area.dgidc.min-edu.pt/materiais_npmeb/matematicaOTD_Final.pdf

Diagramas de Venn e de Carroll

1. Na sala de aula o tema de estudo era o número de letras no nome (próprio). O professor pediu a alguns alunos para irem ao quadro escrever o nome no sítio correcto.

Número de letras do nome



a) A quantos alunos é que o professor pediu para irem ao quadro escrever o nome?

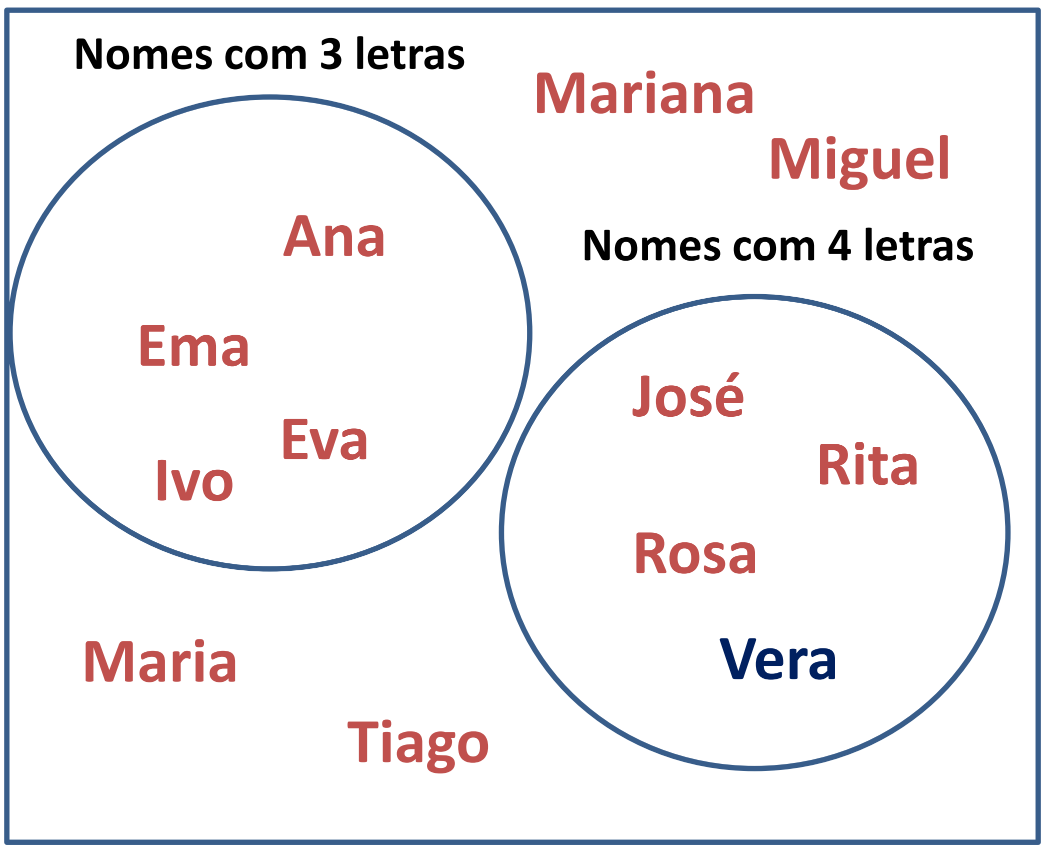
b) Quantos desses alunos têm 3 letras no nome?

c) Algum aluno se enganou a escrever o nome no sítio adequado? Se sim, corrige-o.

Diagramas de Venn e de Carroll

1. Na sala de aula o tema de estudo era o número de letras no nome (próprio). O professor pediu a alguns alunos para irem ao quadro escrever o nome no sítio correcto.

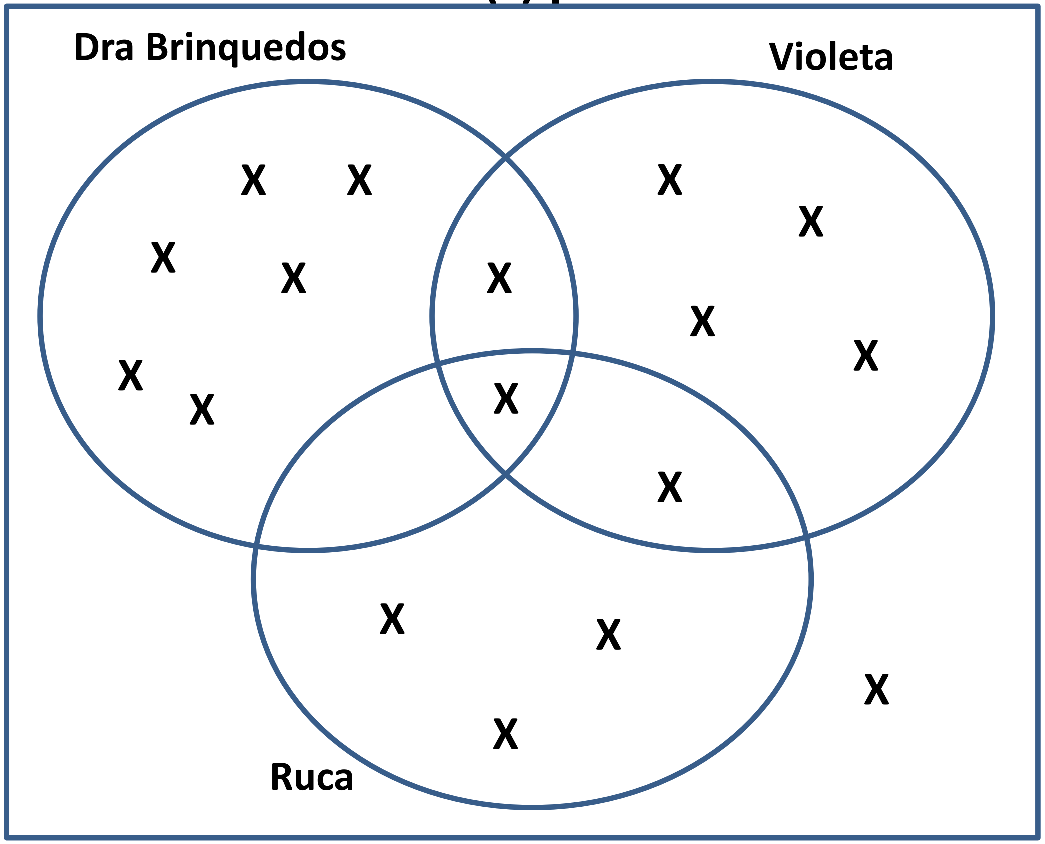
Número de letras do nome



a) Pediu a 12 alunos.
b) São 4 alunos que têm três letras no nome.
c) A Vera enganou-se a escrever o nome no sítio adequado. Deveria ter escrito dentro do círculo destinado aos nomes com 4 letras

2. Perguntou-se aos alunos de uma sala de aula qual a sua série ou séries preferidas. Obteve-se o seguinte diagrama de Venn com as preferências:

Série(s) preferida



a) Quantos alunos preferem a Dra. Brinquedos?

b) Quantos alunos preferem a Dra. Brinquedos e o Ruca?

c) Quantos alunos gostam da Dra. Brinquedos, mas não gostam do Ruca?

d) Quantos alunos só preferem a Dra. Brinquedos?

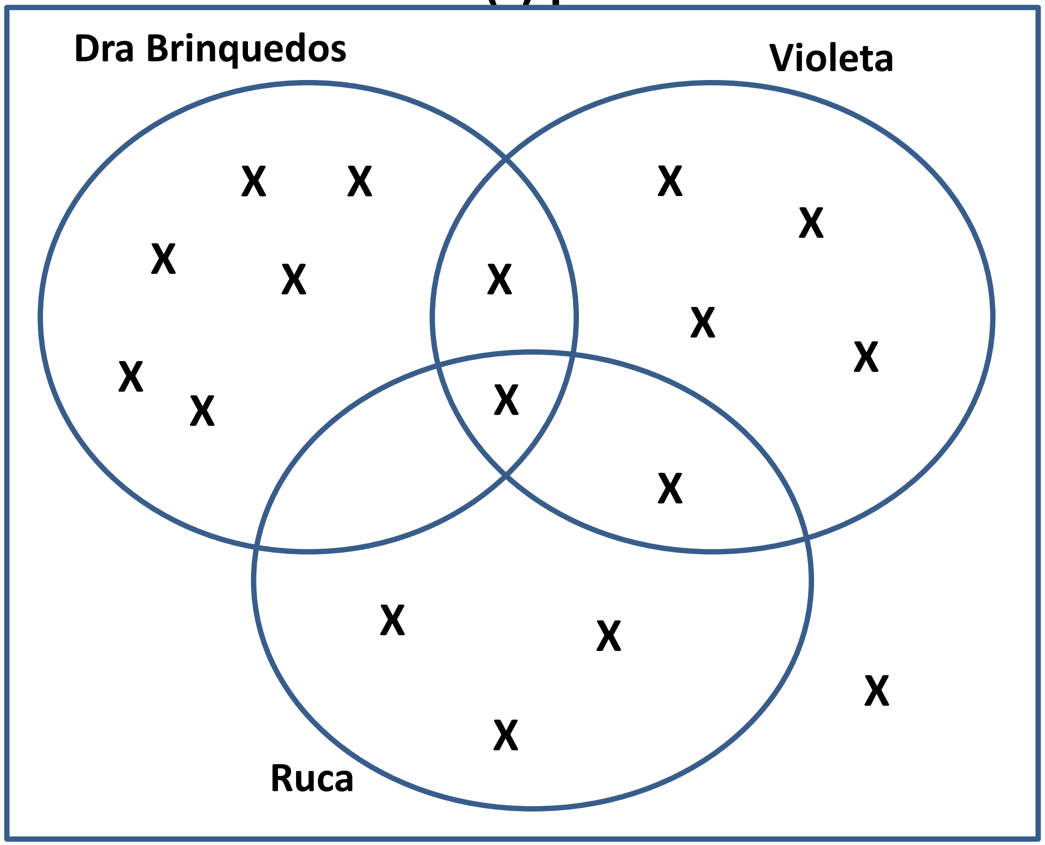
e) Algum ou alguns alunos preferem as três séries?

f) Há algum aluno que não prefira nenhuma das três séries?

Diagramas de Venn e de Carroll

2. Perguntou-se aos alunos de uma sala de aula qual a sua série ou séries preferidas. Obteve-se o seguinte diagrama de Venn com as preferências:

Série(s) preferida



a) Quantos alunos preferem a Dra. Brinquedos? 8

b) Quantos alunos preferem a Dra. Brinquedos e o Ruca? 1

c) Quantos alunos gostam da Dra. Brinquedos, mas não gostam do Ruca? 7

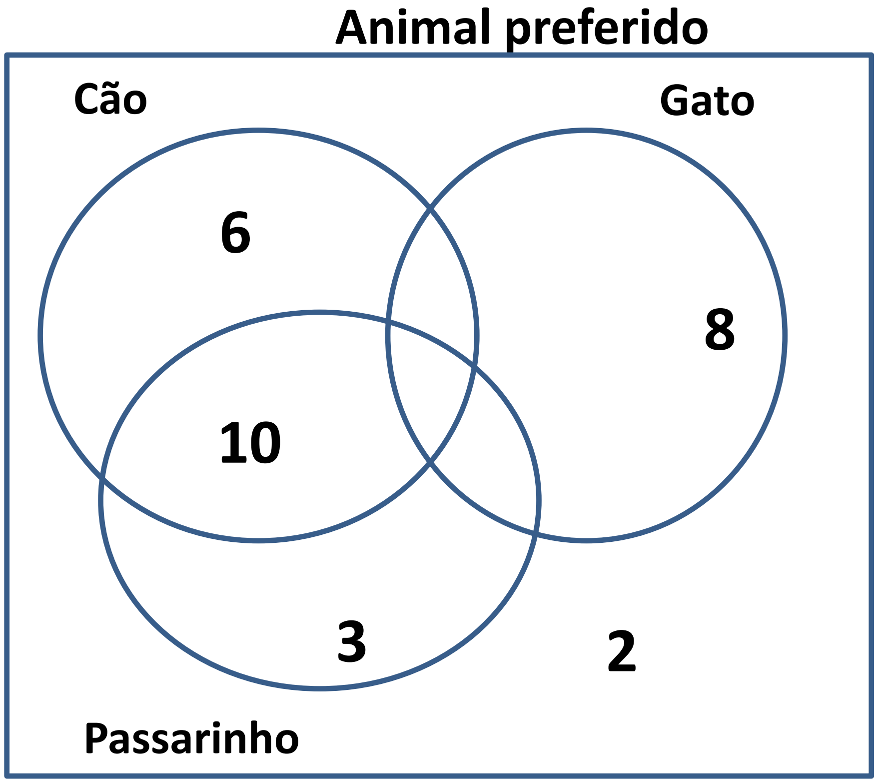
d) Quantos alunos só preferem a Dra. Brinquedos? 6

e) Algum ou alguns alunos preferem as três séries? 1

f) Há algum aluno que não prefira nenhuma das três séries? 1

Diagramas de Venn e de Carroll

3. Perguntou-se a algumas crianças qual o seu animal favorito tendo-se obtido o resultado apresentado no seguinte diagrama de Venn:

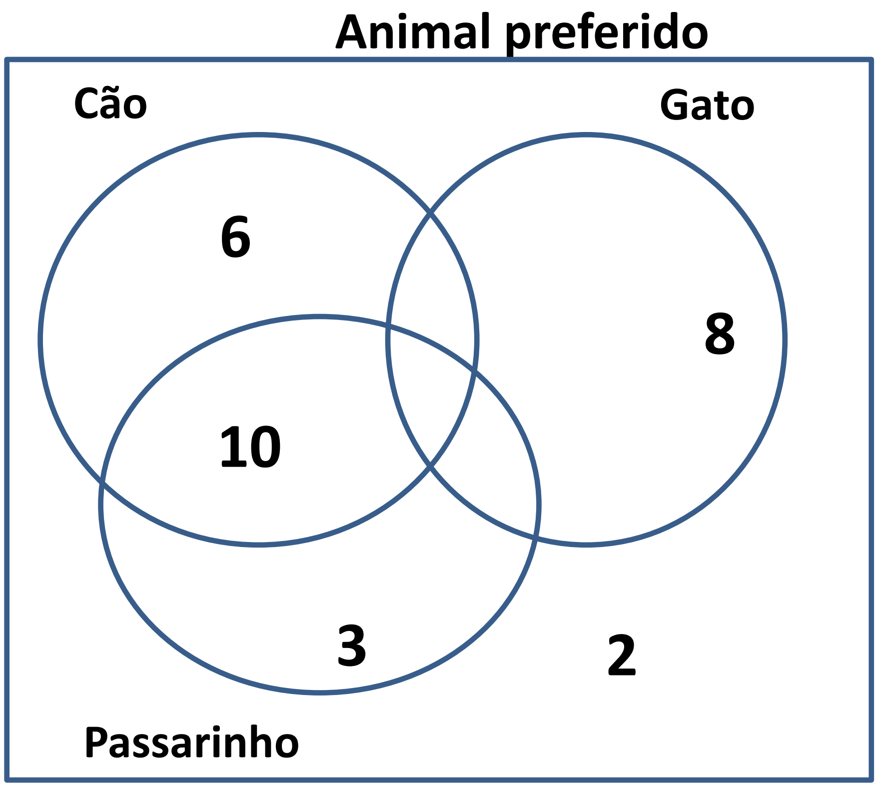


Quantas crianças:

a) Disseram preferir o cão?
b) Disseram preferir o gato?
c) Disseram preferir o passarinho?
d) Disseram preferir o cão e o passarinho?
e) Disseram preferir o cão e o gato?
f) Disseram que nenhum dos animais anteriores era o seu animal favorito?
g) Responderam à questão sobre o animal favorito?

Diagramas de Venn e de Carroll

3. Perguntou-se a algumas crianças qual o seu animal favorito tendo-se obtido o resultado apresentado no seguinte diagrama de Venn:

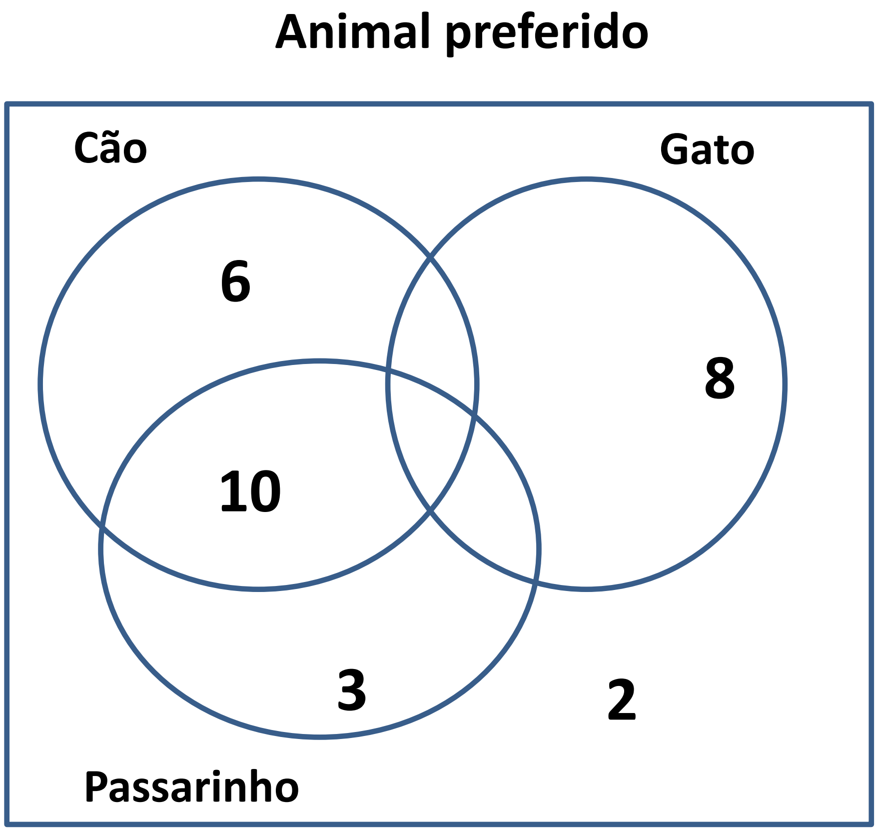


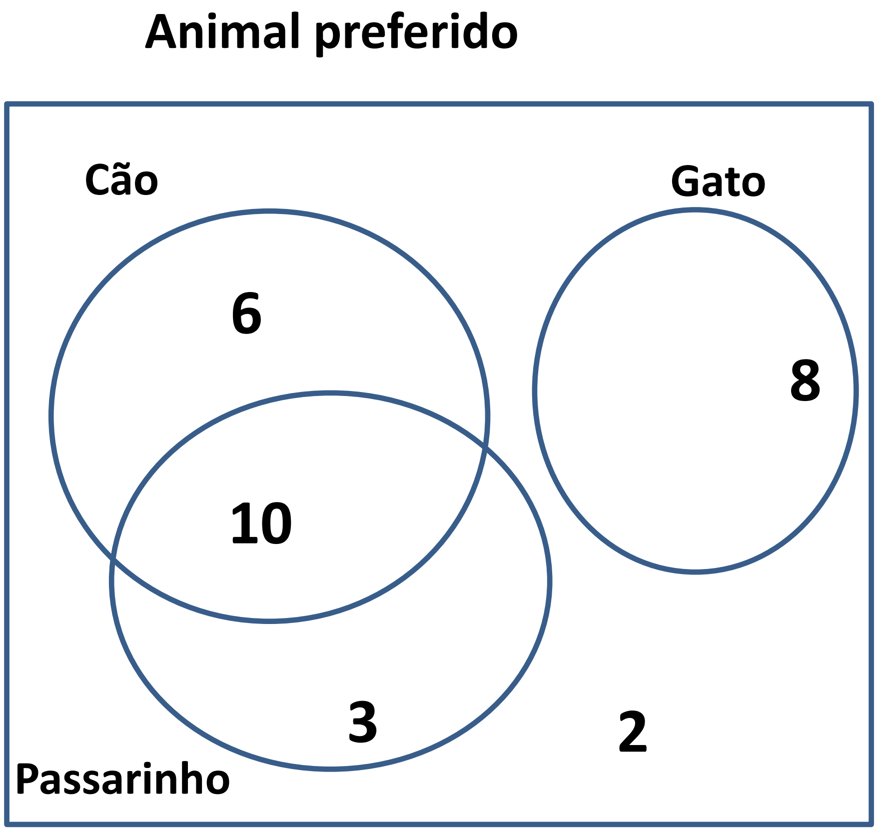
Quantas crianças:

a) Disseram preferir o cão? 16
b) Disseram preferir o gato? 8
c) Disseram preferir o passarinho? 13
d) Disseram preferir o cão e o passarinho? 10
e) Disseram preferir o cão e o gato? 0
f) Disseram que nenhum dos animais anteriores era o seu animal favorito? 2
g) Responderam à questão sobre o animal favorito? 29

Diagramas de Venn e de Carroll

4. Pensas que os dois diagramas seguintes representam a mesma informação?

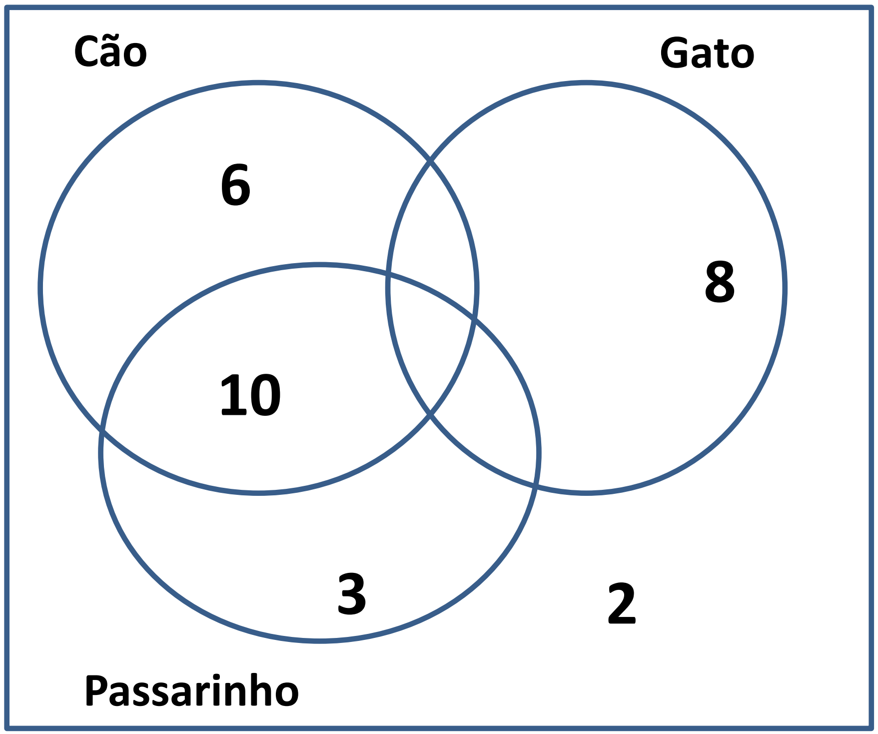




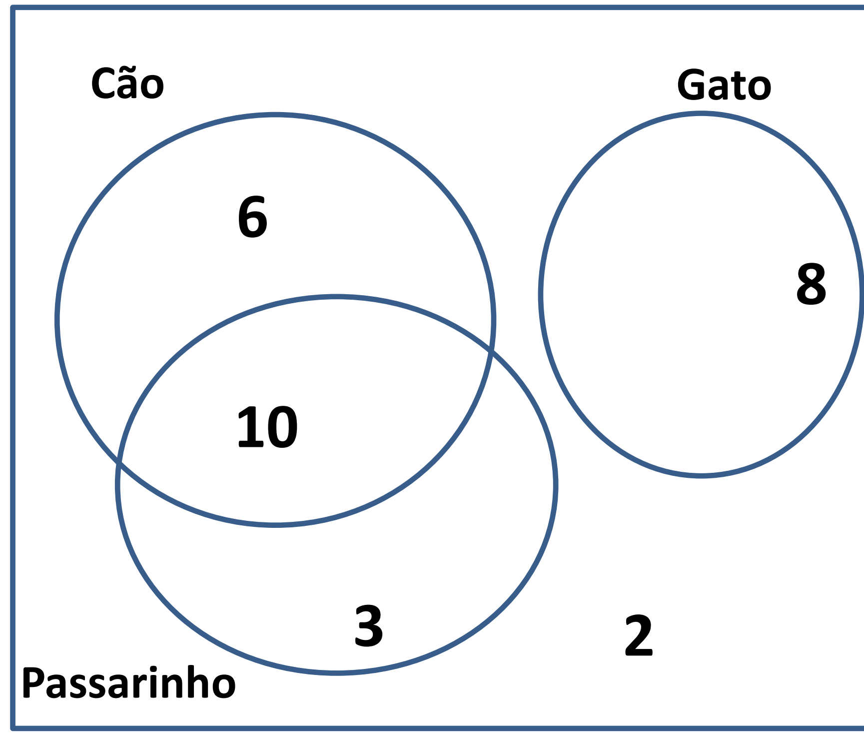
Diagramas de Venn e de Carroll

4. Representam a mesma informação, pois quem prefere o Gato, não prefere nem o cão nem o passarinho e o círculo que o representa não tem que intersectar os outros dois.

Animal preferido

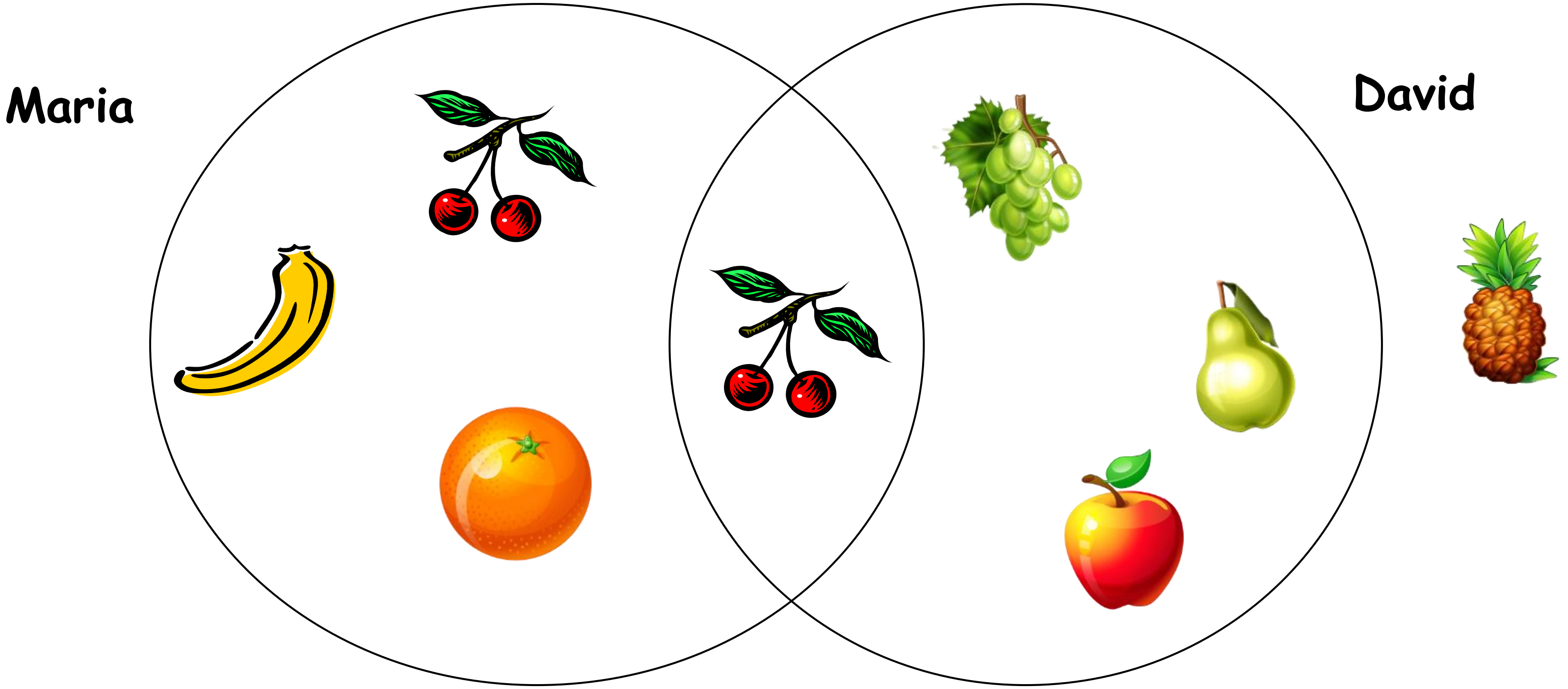


Animal preferido



Diagramas de Venn e de Carroll

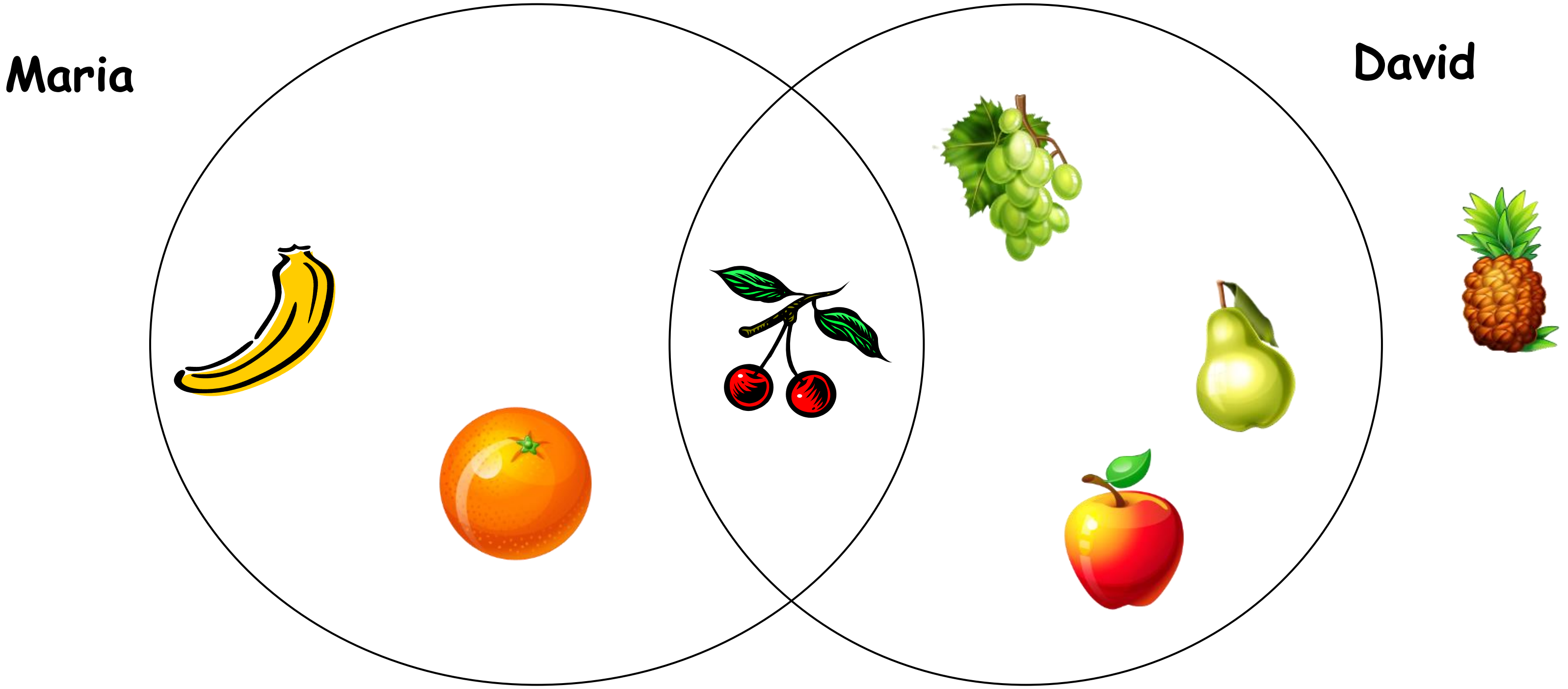
5. Frutos preferidos da Maria e do David



1. Na representação anterior há alguma coisa que consideras errada? Se sim, corrige.
2. Quais os frutos preferidos da Maria?
3. Há algum fruto que seja da preferência dos dois amigos?
4. Há algum fruto que não seja da preferência dos dois amigos?
5. Quantos são os tipos de frutos preferidos da Maria?
6. Quais são os frutos que a Maria prefere, mas não são da preferência do David?

Diagramas de Venn e de Carroll

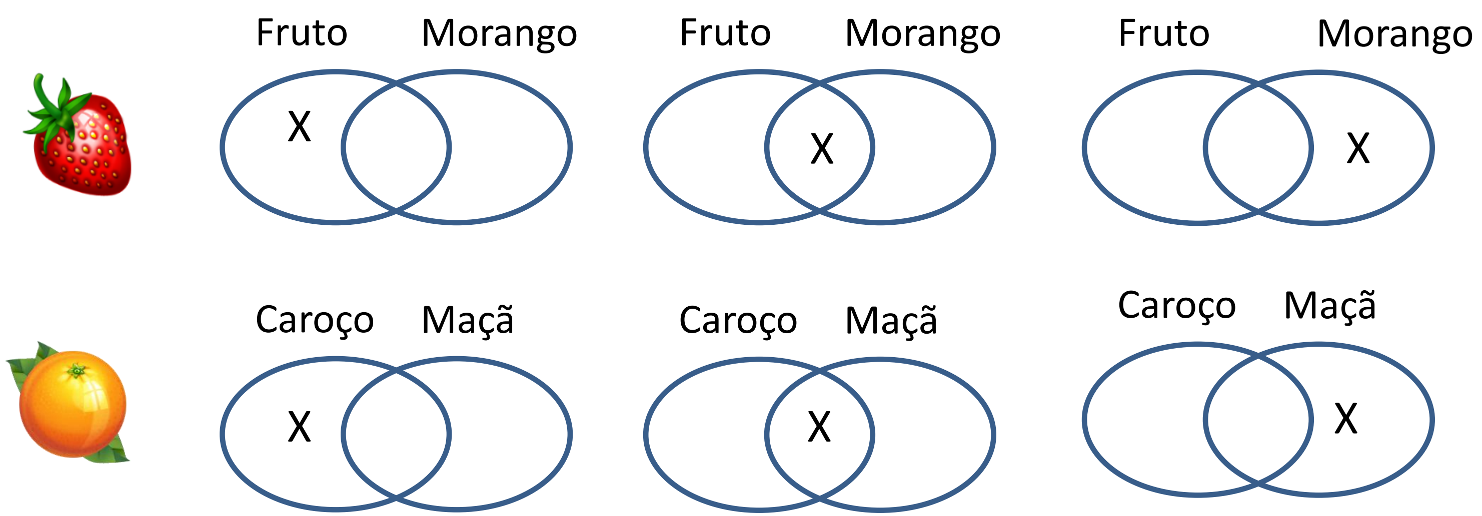
5. Frutos preferidos da Maria e do David

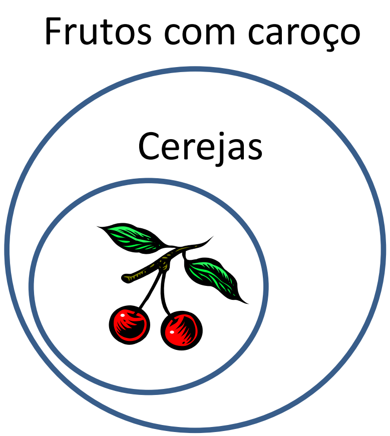


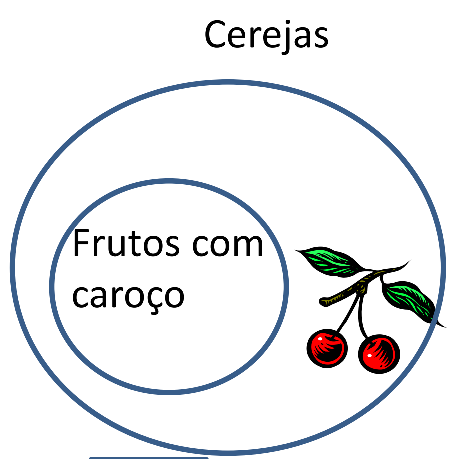
1. Sim. As cerejas estavam repetidas.
2. Banana, laranja e cerejas.
3. As cerejas.
4. O ananás.
5. São 3.
6. A banana e a laranja.

Diagramas de Venn e de Carroll

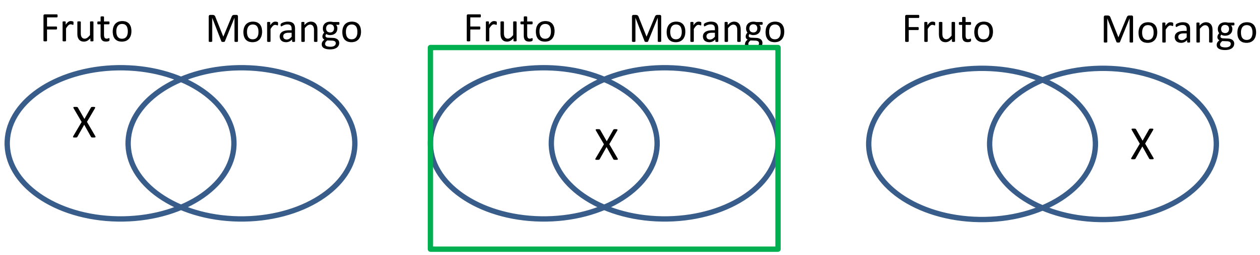
6. Assinala a opção correcta

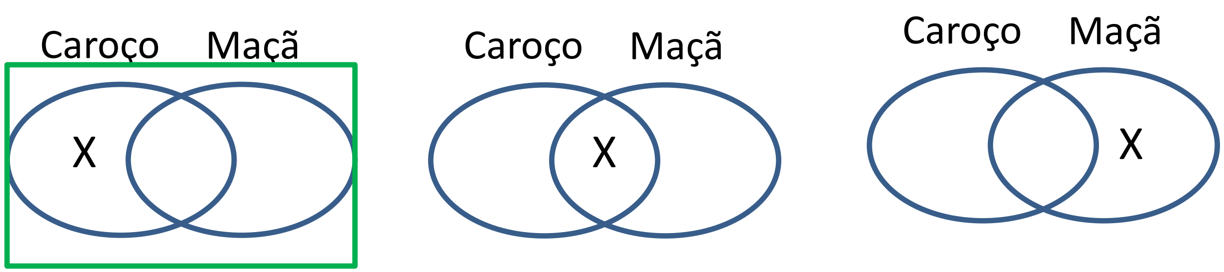


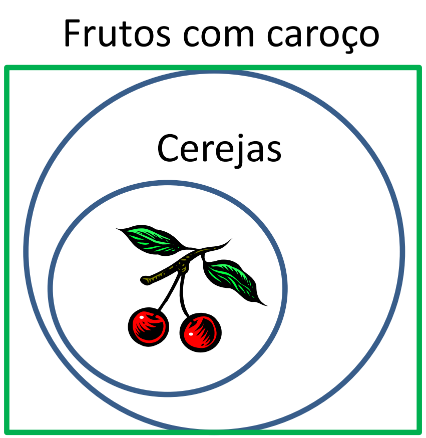


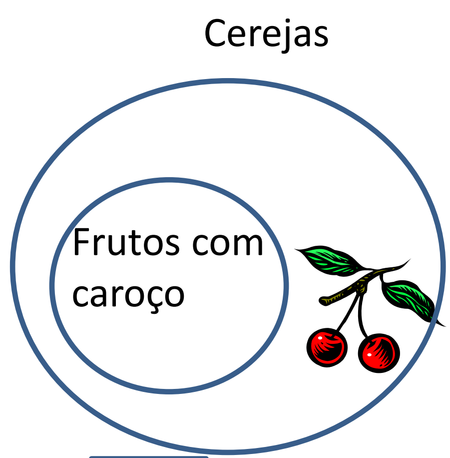


6. Assinala a opção correcta





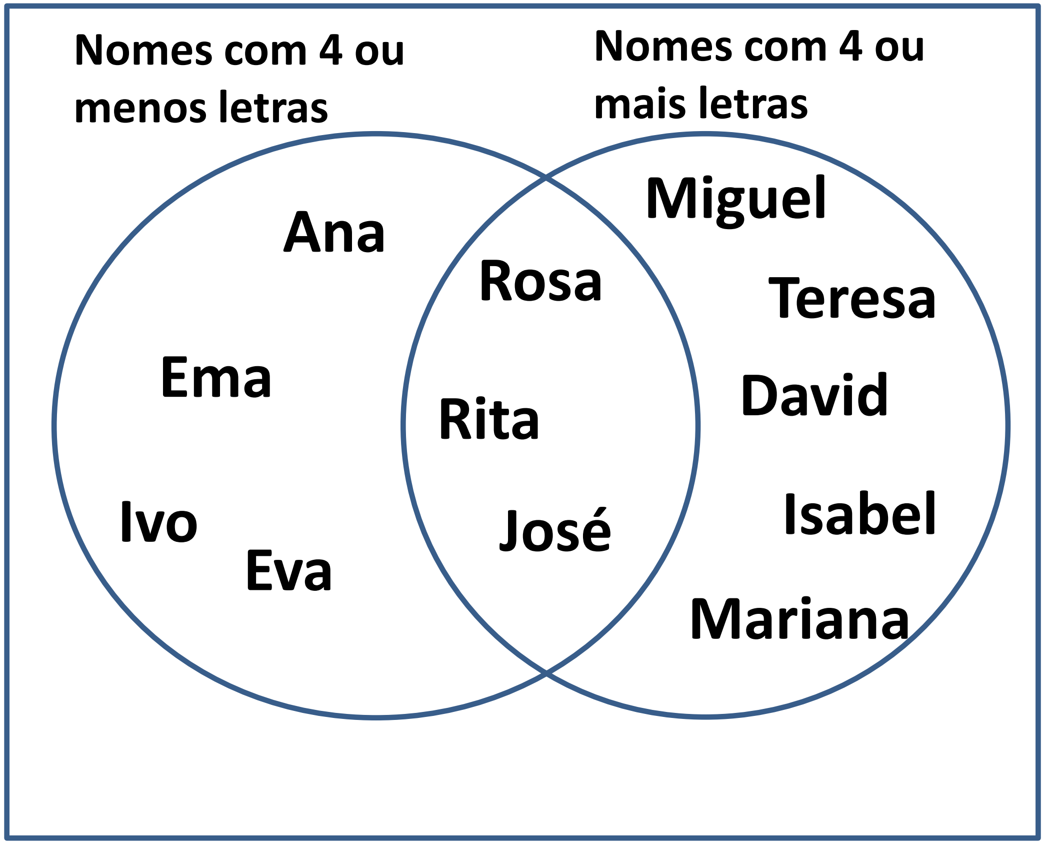




Diagramas de Venn e de Carroll

7. O professor pediu a alguns alunos para irem ao quadro escrever o nome no sítio correcto:

Número de letras do nome



a) A quantos alunos é que o professor pediu para irem ao quadro escrever o nome?

b) Quantos desses alunos têm 4 letras no nome?

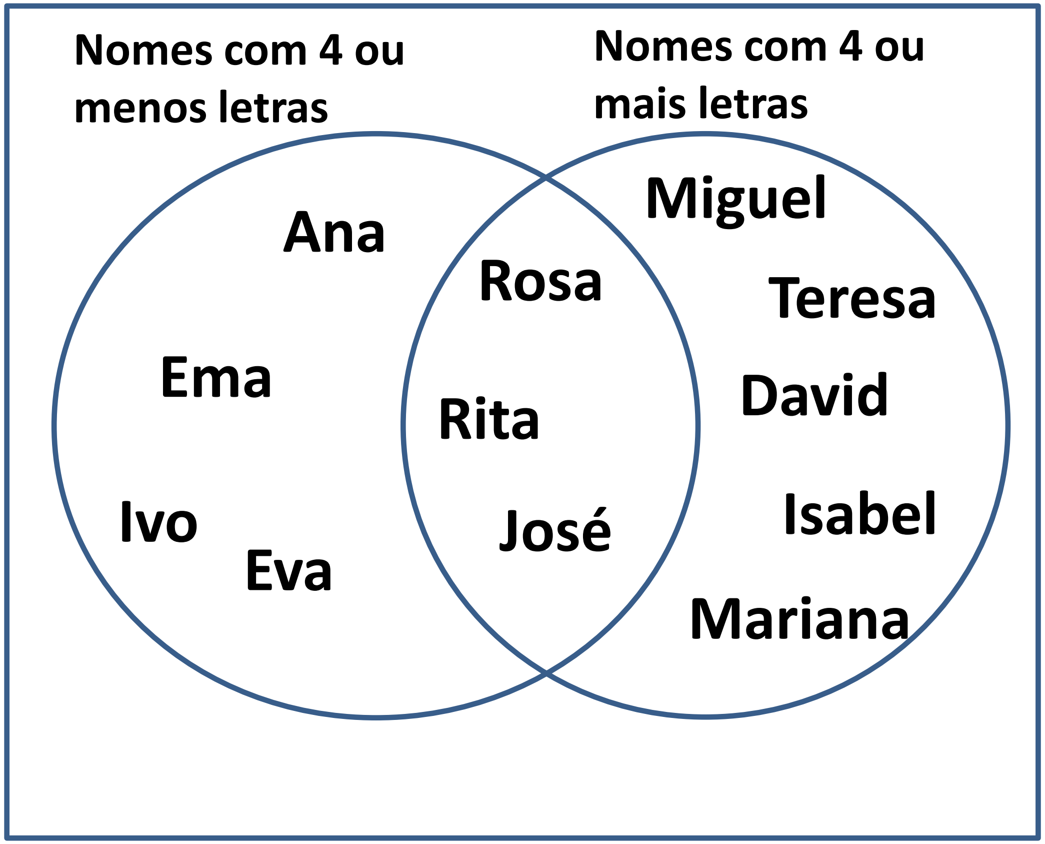
c) Algum aluno poderia escrever o nome dentro do rectângulo, mas fora dos círculos? Porquê?

Consideremos um diagrama de Carroll para representar esta situação.

Diagramas de Venn e de Carroll

7. O professor pediu a alguns alunos para irem ao quadro escrever o nome no sítio correcto:

Número de letras do nome



a) A quantos alunos é que o professor pediu para irem ao quadro escrever o nome? 12

b) Quantos desses alunos têm 4 letras no nome? 3

c) Algum aluno poderia escrever o nome dentro do rectângulo, mas fora dos círculos? Porquê? Não, porque um nome tem 4 letras, ou menos de 4 letras ou mais de 4 letras.

Consideremos um diagrama de Carroll para representar esta situação.

Diagramas de Venn e de Carroll

Os **diagramas de Carroll** são tabelas rectangulares para organizar dados, números 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.

Vejamos como organizar na forma de um diagrama de Carroll a informação referente ao número de letras do nome do exemplo 7, do slide anterior.

Diagramas de Venn e de Carroll

1. **Número de letras no nome**

	Nomes com **4 ou menos** letras	Nomes com **mais de 4** letras
Sexo feminino	**Ana Ema Rosa Rita Eva**	**Teresa Isabel Mariana**
Sexo masculino	**Ivo José**	**Miguel David**

1. A quantos alunos é que o professor pediu para irem ao quadro escrever o nome?
2. Quantos alunos são do sexo masculino?
3. Quantas alunas têm o nome com 4 ou menos letras?

Nota

Repare-se que ao contrário do que se fez no diagrama de Venn em que se considerou nos dois círculos nomes com 4 letras, agora para construir a tabela – diagrama de Carroll, considerámos classes disjuntas.

Diagramas de Venn e de Carroll

1. **Número de letras no nome**

	Nomes com **4 ou menos** letras	Nomes com **mais de 4** letras
Sexo feminino	**Ana Ema Rosa Rita Eva**	**Teresa Isabel Mariana**
Sexo masculino	**Ivo José**	**Miguel David**

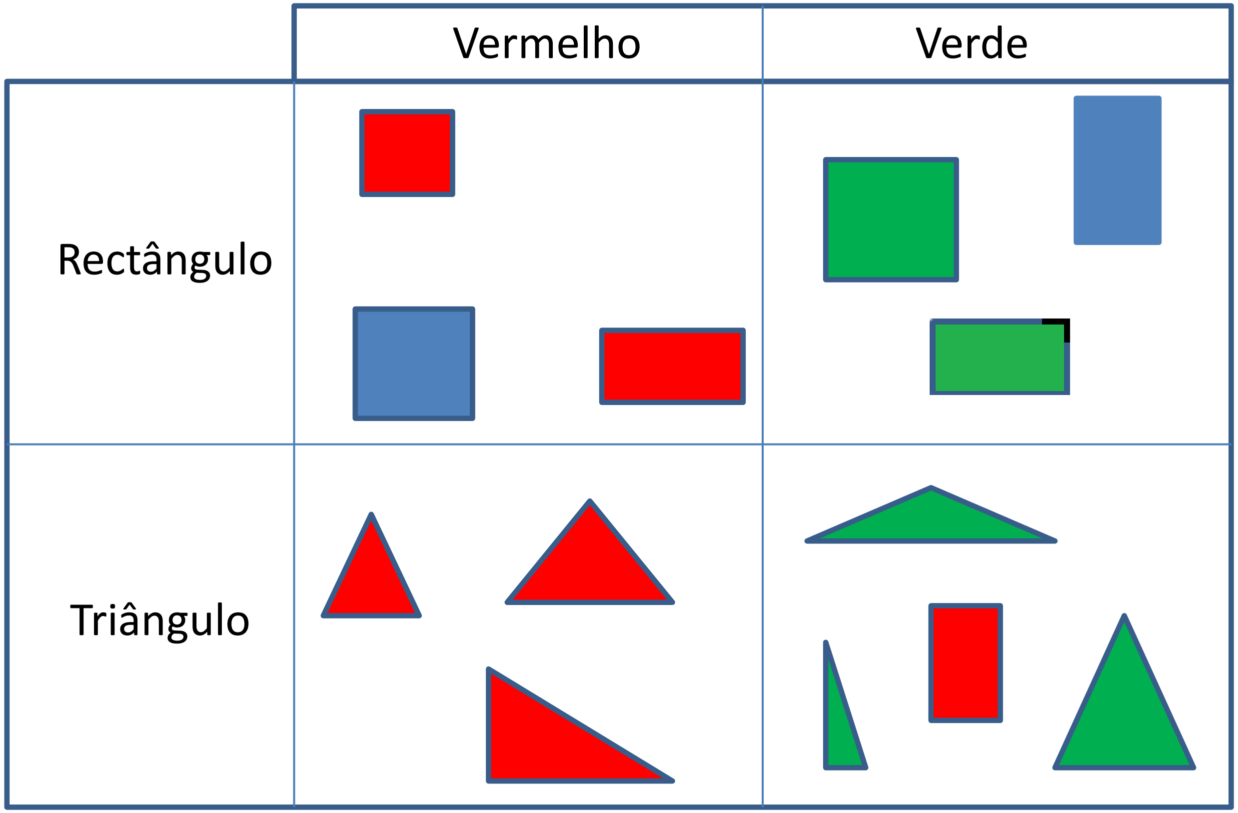
1. A quantos alunos é que o professor pediu para irem ao quadro escrever o nome? 12
2. Quantos alunos são do sexo masculino? 4
3. Quantas alunas têm o nome com 4 ou menos letras? 5

Nota

Repare-se que ao contrário do que se fez no diagrama de Venn em que se considerou nos dois círculos nomes com 4 letras, agora para construir a tabela considerámos classes disjuntas.

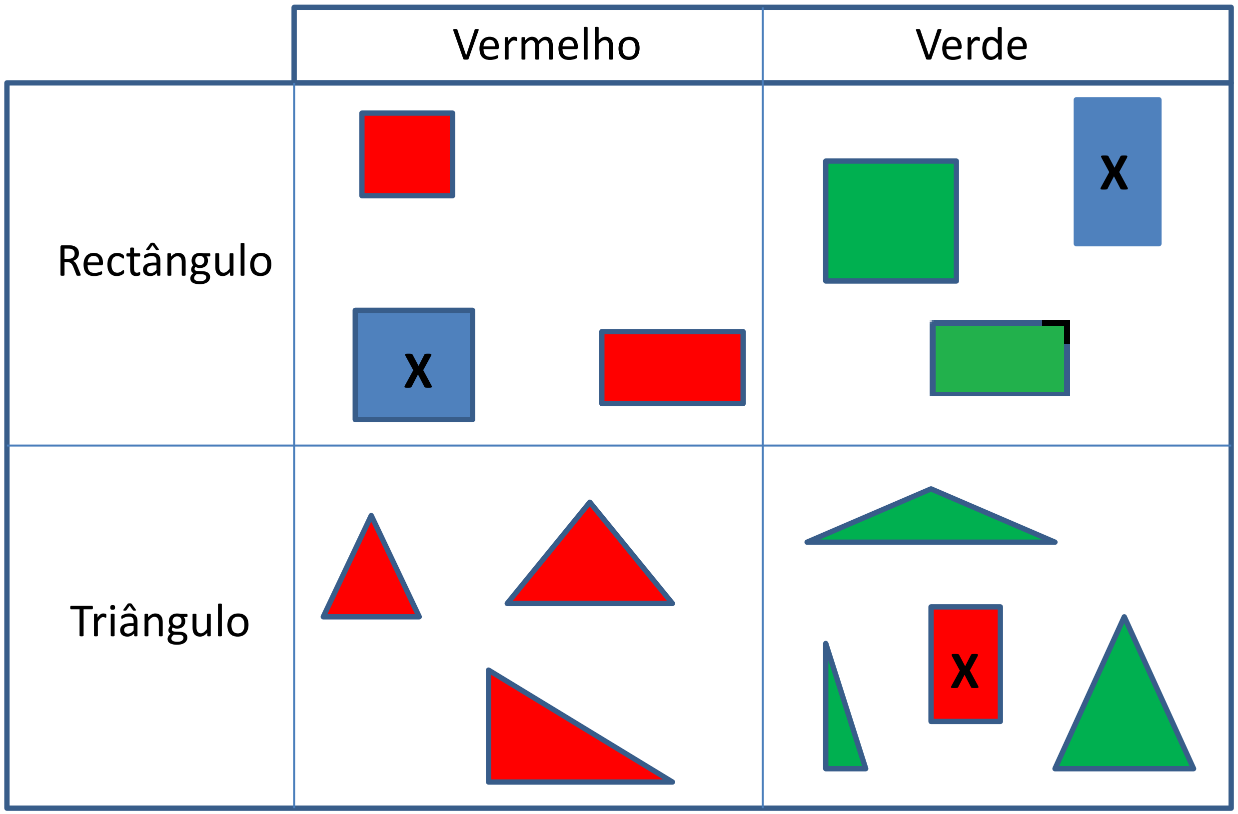
Diagramas de Venn e de Carroll

2. O professor apresentou os seguintes diagramas de Carroll. Os diagramas apresentam alguma(s) situação(ões) errada(s)? Se sim assinala com um **X** ou um **O** os elementos mal classificados.



Diagramas de Venn e de Carroll

2. O professor apresentou os seguintes diagramas de Carroll. Os diagramas apresentam alguma(s) situação(ões) errada(s)? Se sim assinala com um **X** ou um O os elementos mal classificados.



Diagramas de Venn e de Carroll

3.

	Múltiplos de 3	Não múltiplos de 3
Menor que 20	3 6 9 12 18 15	1 5 7 8 13 17
Maior ou igual a 20 e menor que 30	20 27 21 33 30	22 24 28 26 29

Diagramas de Venn e de Carroll

3.

	Múltiplos de 3	Não múltiplos de 3
Menor que 20	3 6 9 12 15 18	1 5 7 8 13 17
Maior ou igual a 20 e menor que 30	20 27 21 30 33	22 24 28 26 29

Diagramas de Venn e de Carroll

Como vimos dos exemplos anteriores e como já dissemos anteriormente, pensamos que **não se justifica**, em OTD, o estudo formal de **Teoria dos conjuntos**, pois em vez de facilitar a classificação de dados, pode ser motivo para, a este nível, originar confusão com a definição formal e matemática de conjunto.

Tipos de variáveis

Na determinação da análise estatística apropriada para um conjunto de dados, é importante classificar as variáveis quanto ao tipo:

Uma variável diz-se **quantitativa** (ou numérica) se se referir a uma característica, da unidade estatística, que se possa *contar* ou *medir.*

- Uma variável quantitativa que se refere a uma característica que se possa contar, designa-se por **quantitativa discreta**.
- Uma variável quantitativa que se refere a uma característica que se possa medir, designa-se por **quantitativa contínua**. Ao contrário da variável discreta, uma variável contínua pode assumir um número infinito de valores reais, entre dois quaisquer valores que assuma.

Como resultado da observação de uma variável discreta ou contínua têm-se, respectivamente, os **dados discretos** ou os **dados contínuos**.

Tipos de variáveis (cont)

Como exemplo de variáveis discretas que podem ser observadas sobre cada aluno (unidade estatística) de uma turma temos o “número de irmãos” e o “número de mensagens de telemóvel recebidas por dia”.

Como exemplos de variáveis contínuas temos a “altura” e o “tempo que leva de casa à escola”.

Deve ter-se presente que, devido à limitação dos instrumentos de medida, uma variável contínua apresenta-se, normalmente, discretizada. Por exemplo, no caso da variável altura, um aluno pode medir 1,6431456... metros, embora se considere a altura igual a 1,64m ou 164cm.

Tipos de variáveis (cont)

- 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. Os dados resultantes são denominados **dados qualitativos** ou **dados categóricos**. Algumas variáveis qualitativas têm subjacente uma ordem, sendo denominadas de **qualitativas ordinais**.

Como exemplos de variáveis qualitativas que podem ser observadas sobre cada aluno de uma turma (unidade estatística) temos "cor dos olhos", "meio de transporte utilizado para ir para a escola", "ter telemóvel" e "ano de escolaridade", sendo esta última *qualitativa ordinal*.

Tipos de variáveis e tipos de dados

Atenção

Ao contrário do que acontece com os dados quantitativos, com os dados qualitativos não tem sentido fazer operações aritméticas, ainda que estes dados estejam representados por números.

Tipos de variáveis e tipos de dados

Apresenta-se a seguir o ficheiro “Dados sobre casas” da brochura *Análise de dados* (Maria Eugénia Graça Martins, Luísa Canto e Castro Loura e Maria de Fátima Mendes, DGIDC, 2007), que será utilizado para exemplificar alguns dos conceitos introduzidos.

Dados sobre casas 1

Ident.	Nº de assoalhadas	Área (m2)	Estado (novo-1, usado-0	Garagem (sim-1, não-0)	Zona	Preço(milhares euros)
1	3	99,01	0	0	C	138,50
2	3	90,49	0	0	B	190,30
3	3	109,01	0	0	B	179,26
4	3	104,75	0	0	B	162,74
5	5	138,7	1	1	A	357,32
6	2	87,26	0	0	B	157,39
7	2	93,74	0	0	B	138,34
8	4	118,47	0	0	B	209,46
9	2	88,86	0	1	A	169,60
10	2	95,63	0	0	B	153,56
11	3	104,28	0	0	C	149,00
12	3	126,53	1	0	A	299,33
13	4	118,46	0	0	B	207,66
14	3	98,85	0	1	B	182,86
15	3	100,28	1	1	A	236,27
16	3	94,66	0	0	B	188,17
17	2	87,95	0	0	C	122,84
18	2	92,35	0	1	B	149,20
19	2	101,07	0	0	A	160,13
20	1	66,32	0	1	A	147,89

Dados sobre casas 2

21	2	96,81	1	0	A	202,63
22	3	103,17	0	0	A	205,92
23	2	109,03	0	1	A	185,66
24	3	119,03	0	1	A	210,21
25	2	100,82	0	1	A	208,88
26	1	79,5	1	0	A	186,09
27	3	114,63	0	0	B	183,49
28	2	91,14	0	0	C	126,80
29	2	94,94	0	0	A	165,69
30	2	98,08	1	1	A	290,00
31	3	94,85	0	1	B	170,18
32	3	102,89	0	1	B	189,22
33	2	104,38	1	0	A	255,90
34	3	112,87	1	0	A	281,25
35	2	87,57	0	0	C	121,47
36	2	76,59	1	1	A	210,24
37	5	163,34	0	0	B	295,98
38	3	154,21	0	0	A	255,03
39	1	75,85	0	0	A	135,69
40	2	90,19	0	0	B	151,26

Tipos de variáveis e tipos de dados

No ficheiro "Dados sobre casas" tem-se uma população em que as unidades estatísticas são as casas. Algumas variáveis de interesse são:

- **Variáveis qualitativas** - *Estado da casa*, *Ter garagem* e *Zona*
- **Variável quantitativa discreta** – *Número de assoalhadas*
- **Variáveis quantitativas contínuas** - *Área* e *Preço*

Tipos de variáveis e tipos de dados

Notas

1. O facto de os dados resultantes da observação de uma variável se representarem por números, não significa que sejam de tipo quantitativo ou que a variável seja quantitativa. No ficheiro anterior as variáveis *Estado da casa* e *Ter garagem* são qualitativas, embora se tenham codificado os dados resultantes da sua observação por números.
2. Nem todos os dados acarretam consigo o mesmo tipo de informação. Por exemplo, ao estudar uma população em que as unidades estatísticas são pessoas, pode-se classificar a variável "sexo" associando o número 1 à categoria masculino e o número 2 à categoria feminino. Se observarmos um conjunto de pessoas quanto ao sexo e obtivermos um conjunto de 1's e 2's, não tem qualquer sentido calcular a média destes dados, nem realizar

(continua)

Tipos de variáveis e tipos de dados

quaisquer operações numéricas. No entanto, se aquele conjunto de 1's e 2's for o resultado de observar a variável "número de filhos" sobre as mesmas unidades estatísticas, então já faz sentido calcular a média.

3. Uma variável discreta pode ser apresentada sem ser na forma de inteiro. Por exemplo, a pontuação dada pelo júri numa competição de patinagem artística pode assumir qualquer valor entre 0 e 10, com uma casa decimal (é idêntico a considerar os inteiros entre 0 e 100).

Tipos de variáveis e tipos de dados

Atenção

O tipo de dados que resultam da observação de uma característica dependem do processo como é feita essa observação, ou seja da escala de medida, e não da característica que está a ser observada. Pode uma característica ser susceptível de uma medição em várias escalas de medida, pelo que os dados resultantes dessa medição dependerão da escala utilizada.

Por exemplo, se pretendermos *contabilizar* a "classificação obtida num teste de Matemática", temos duas escalas possíveis, que podem ser as seguintes: ou consideramos os números inteiros entre 0 e 100, ou consideramos as categorias Mau, Medíocre, Suficiente, Bom e Muito Bom. Assim, os dados da variável em estudo serão, respectivamente, quantitativos discretos ou qualitativos ordinais.

Tipos de variáveis e tipos de dados

Exemplo - Considere-se um lote de camisolas de lã, de uma certa marca, à venda numa determinada loja. O *Tamanho da camisola* é uma variável? Se sim, de que tipo?

Neste caso as camisolas são as unidades estatísticas sobre as quais vamos fazer uma "medição" e o *Tamanho de uma camisola* é uma variável, uma vez que pode variar de camisola para camisola. Já que o tamanho está ligado ao comprimento, poderíamos ser levados a dizer que a variável é quantitativa contínua e os dados se fossem o resultado de medir o comprimento das camisolas com uma fita métrica seriam contínuos.

No entanto, normalmente as camisolas apresentam vários tamanhos, nomeadamente XS, S, M, L e XL. Assim, os dados resultantes da medição desta variável *Tamanho da camisola* são de tipo qualitativo, ordinal, apresentando 5 categorias possíveis.

Tipos de variáveis e tipos de dados

Exemplo - Se o ficheiro "Dados sobre casas", apresentasse a ***Idade*** das casas, como se classificaria esta variável?

Para já adianta-se que a *Idade* é uma variável, à partida, de tipo quantitativo contínuo, pois é uma variável que é medida – o tempo mede-se. No entanto, quando falamos na idade, falamos em números inteiros, como se a variável fosse quantitativa discreta.

O que acontece é que quando referimos a idade, nos estamos a referir a classes, pois quando dizemos que uma casa tem, por exemplo, 15 ou 21 anos, significa que a sua idade está, respectivamente, na classe [15, 16[ou na classe [21, 22[. Assim, a variável *Idade,* na sua essência uma variável contínua, utiliza-se "discretizada" em classes representadas por inteiros (positivos).

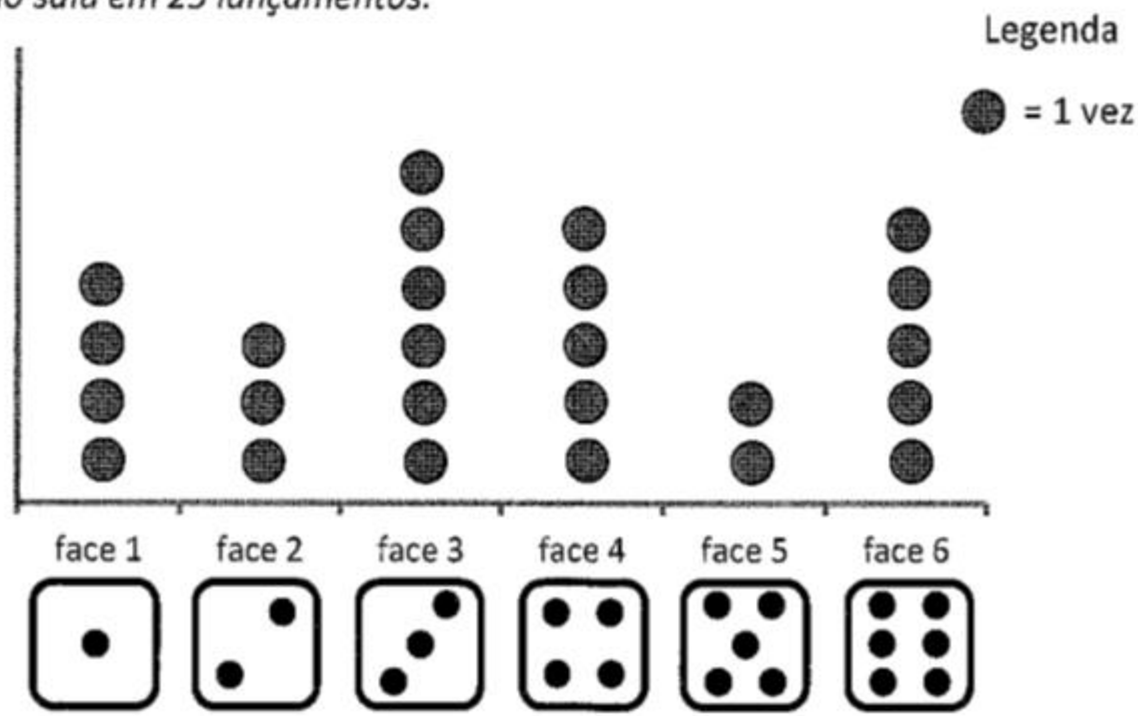
Nota - Na prática qualquer variável contínua é apresentada discretizada, pois a sua medição está condicionada à precisão do instrumento de medida...

Tipos de variáveis e tipos de dados

Considere-se o seguinte exemplo do caderno de apoio do 1º ciclo, OTD1, página 15

Exemplo

Observa o gráfico que representa o número de vezes que cada uma das faces de um dado saiu em 25 lançamentos.



a. *Quantas vezes saiu a face 4?*
b. *Qual foi a face que saiu menos vezes?*

Algumas questões:

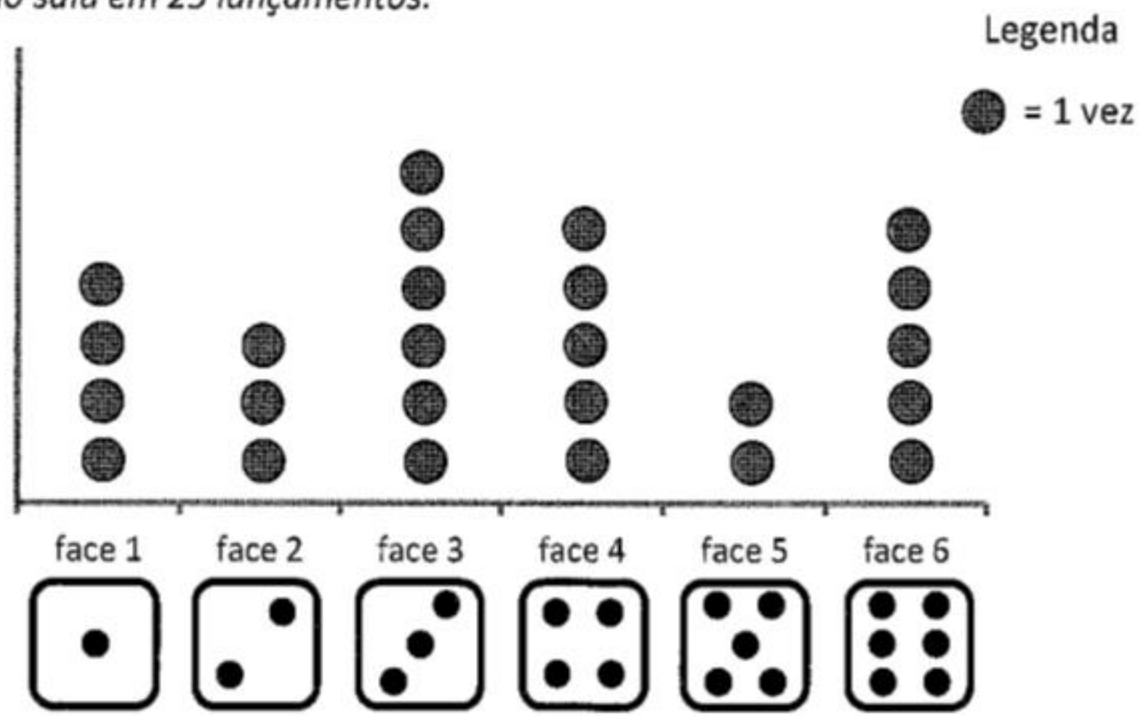
1. Qual a variável que se está a estudar?
2. Essa variável é de tipo qualitativo ou quantitativo?
3. Escreve 3 *dados* que pudesses ter obtido quando lanças o dado e registas a face que ficou virada para cima.

Tipos de variáveis e tipos de dados

Considere-se o seguinte exemplo do caderno de apoio do 1º ciclo, OTD1, página 15

Exemplo

Observa o gráfico que representa o número de vezes que cada uma das faces de um dado saiu em 25 lançamentos.



a. *Quantas vezes saiu a face 4?*
b. *Qual foi a face que saiu menos vezes?*

Algumas questões:

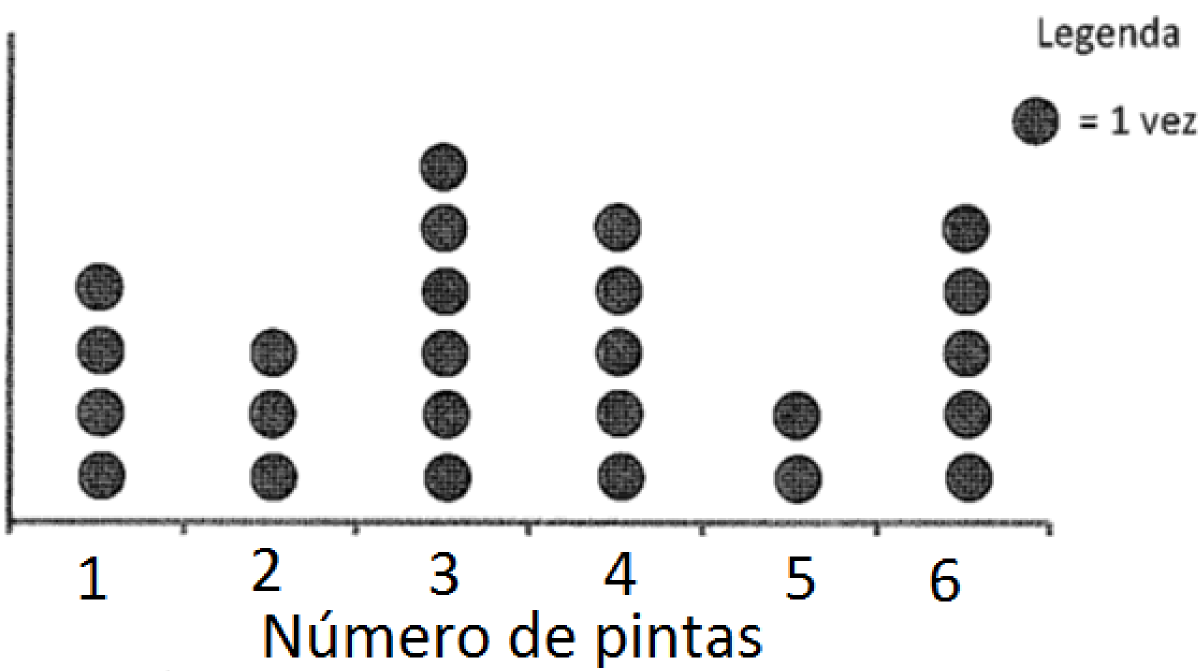
1. A variável que se está a estudar é a "etiqueta" da face que fica voltada para cima. Esta variável assume as categorias face 1, face 2, ..., face 6.
2. A variável é de tipo qualitativo.
3. face 2, face 5, face 1.

Tipos de variáveis e tipos de dados

Considere-se agora o seguinte:

Exemplo

Lança um dado e regista o número de pintas da face que ficou voltada para cima. Observa o gráfico que representa os resultados observados em 25 lançamentos.



a. *Quantas vezes saíu a face com 4 pintas?*
b. *Qual foi a face que saiu menos vezes?*

Algumas questões:

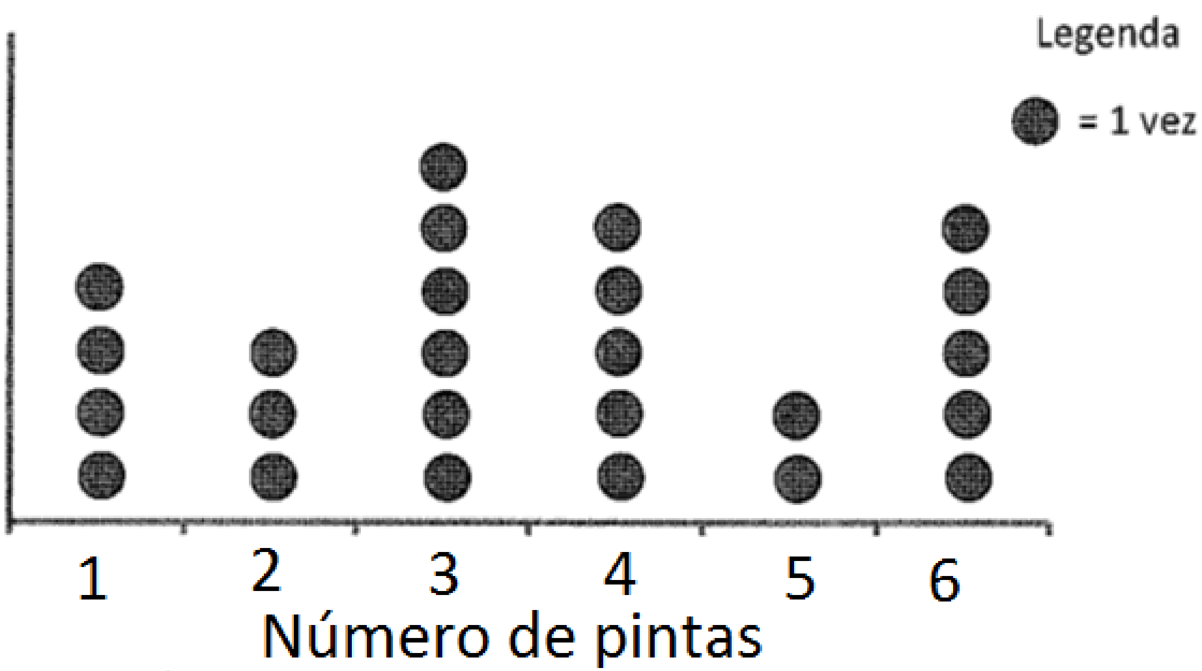
1. Qual a variável que se está a estudar?
2. Essa variável é de tipo qualitativo ou quantitativo?
3. Escreve 3 *dados* que pudesses ter obtido quando lanças o dado e registas a face que ficou virada para cima.

Tipos de variáveis e tipos de dados

Considere-se agora o seguinte:

Exemplo

Lança um dado e regista o número de pintas da face que ficou voltada para cima. Observa o gráfico que representa os resultados observados em 25 lançamentos.



a. *Quantas vezes saíu a face com 4 pintas?*
b. *Qual foi a face que saiu menos vezes?*

Algumas questões:

1. A variável que se está a estudar é o "Número de pintas da face que fica voltada para cima".
2. A variável é quanti-tativa discreta.
3. 2, 5, 1.

Tipos de variáveis e tipos de dados

Considere-se ainda o seguinte exemplo do **Caderno de apoio 2º ciclo, OTD6, página 57.**

Descritor	Texto de apoio
1.3	**Exemplo** *Das seguintes variáveis estatísticas indica as que são qualitativas:* • *Idade;* • *Ano de escolaridade;* • *Altura;* • *Nacionalidade;* • *Meio de transporte utilizado no percurso casa-escola;* • *Tempo médio gasto no percurso casa-escola;* • *Código postal.* No que diz respeito ao último exemplo, é importante chamar a atenção dos alunos para o facto de que nem todas as variáveis que tomam valores envolvendo representações com algarismos são quantitativas. O código postal é uma variável estatística qualitativa, uma vez que não resulta de uma contagem, ou, mais geralmente da medida de uma dada grandeza.

Ver na página seguinte comentários a este exemplo

Comentários

- Para saber se estamos perante uma variável, temos de saber quais as unidades estatísticas que estão a ser estudadas. Admitamos que as unidades estatísticas são os alunos de uma escola. Sendo assim, poderá ter sentido considerar como variáveis a *Idade*, *Ano de escolaridade*, *Altura*, *Nacionalidade*, *Meio de transporte utilizado no percurso casa-escola* e eventualmente (ver comentário) *Código postal*.
- Para saber de que tipo é a variável que se pretende estudar, tendo em consideração a chamada de atenção do slide, pode ser necessário indicar qual a escala de medida.

Comentários (continuação)

- Entre o conjunto de sugestões de variáveis, é apresentada o *Tempo **médio** gasto no percurso casa-escola*, que não pode ser considerada uma variável, uma vez que não é susceptível de ser observada sobre o aluno. Quando se fala em *tempo médio*, é porque, de qualquer modo, se procedeu a uma média de tempos. Assim, poderemos estar a referir-nos ao tempo que resulta de fazer a média dos dados (tempos) que se obtêm quando se registam os valores que a variável *Tempo gasto no percurso casa-escola* assume sobre todos os alunos da escola (que são as unidades estatísticas) num determinado dia ou, para cada aluno, a média dos tempos observados no percurso casa-escola, durante vários dias, por esse aluno.

Tipos de variáveis e tipos de dados

Comentários (continuação)

- Será que o *código postal*, que identifica um endereço ou localização geográfica, é uma variável (estatística)? Se for variável, será qualitativa, sem dúvida. Mas, a pergunta que se faz é: quais as categorias ou modalidades que esta (pressuposta) variável pode assumir? Se as unidades estatísticas em estudo forem os endereços de uma rua, então não temos variabilidade, pois os endereços terão o mesmo código postal, e só existirá uma categoria. Por outro lado, se os endereços forem de ruas diferentes, então todos os códigos serão diferentes e teremos tantas categorias quantos os endereços, pelo que não há possibilidade de proceder a qualquer organização de dados! Só terá sentido falar no código postal como variável estatística, se houver algum contexto que o justifique. Somos de opinião que não tem sentido considerar o código postal como variável, pela mesma razão que também não se deve considerar como variável a ser observada sobre uma pessoa o seu *número de BI*. <u>Em Estatística os dados são recolhidos para responderem a questões</u>.

Tipos de variáveis e tipos de dados

Exemplo – Por vezes vemos colocada a seguinte questão.
Considera o *Número de alunos do sexo masculino*. É uma variável?

Não e sim! Sem termos o contexto da situação, não podemos dar uma resposta. Se não são indicadas as unidades estatísticas em estudo, não podemos dizer se o *Número de alunos do sexo masculino* é uma variável. Por exemplo, se considerarmos que as unidades estatísticas são os alunos de uma sala de aulas, não temos uma variável, mas a frequência absoluta da categoria sexo masculino, da variável *Sexo*, variável qualitativa.

Por outro lado, se considerarmos como unidades estatísticas as salas de aulas de uma escola, então já o *Número de alunos do sexo masculino* em cada sala é uma variável, nomeadamente uma variável quantitativa discreta.

Tipos de variáveis e tipos de dados

Exemplo – Pretende-se comparar o "tamanho" de vários livros.

a) Indica duas variáveis quantitativas que possam ser utilizadas para descrever o tamanho de um livro.

b) Qual das variáveis que indicaste, consideras mais adequada para estimar o tempo que demorarias a ler um livro?

Resposta

a) Duas variáveis que podem ser indicadas para medir a característica "tamanho de um livro" podem ser o seu *número de páginas* ou o seu *peso,* em gramas. A primeira será quantitativa discreta, enquanto que a segunda será quantitativa contínua.

b) A variável mais adequada será o *número de páginas* do livro.

1. Supõe que a tua mãe pretende fazer um bolo e não tem ovos. Vais com ela ao supermercado, à prateleira dos ovos e para a mesma marca, encontras vários tamanhos de ovos.
Consegues explicar à tua mãe, que tipo de variável é a que é utilizada para medir o tamanho dos ovos?
2. Ainda no supermercado, decidem comprar uma caixa de Camarões. És capaz de indicar uma variável que sirva para descrever a "quantidade" de camarões que a caixa contém?
3. Os camarões vão servir para fazer um arroz á valenciana, para o qual é necessário ervilhas. Pegam num pacote de ervilhas que tem indicado o peso de mais ou menos 500 gramas, o que significa que a quantidade de ervilhas está a ser medida pela variável peso. Que outra variável quantitativa pode ser utilizada para medir a quantidade de ervilhas do pacote? Achas que seria prático utilizar esta medida, em vez do peso, para medir a quantidade de ervilhas?

Distribuição

Dado um conjunto de **dados**, resultantes da observação de uma dada **variável**, pretendemos desenvolver processos de análise que nos permitam responder a algumas questões, tais como:

- Serão os dados quase todos iguais, numericamente?
- Serão muito diferentes, uns dos outros?
- De que modo é que são diferentes?
- Existe alguma estrutura subjacente ou alguma tendência? Dispõem-se de forma simétrica, ou apresentam enviesamento?
- 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, mas se estiverem organizados sob a forma **de tabelas** e **gráficos**, adequados, já a resposta às questões anteriores se torna mais simples.

Distribuição

A **distribuição** de uma variável é o padrão apresentado por um conjunto de dados, resultantes da observação da variável, organizados na forma de

- **Tabelas de frequência**

  e

- **Gráficos**

Esta descrição é complementada, no caso de dados quantitativos, por

- **Medidas numéricas – de localização e dispersão**

Existem alguns processos de organização de dados mais adequados conforme o tipo de dados a organizar, pelo que nos slides seguintes analisaremos separadamente os dados qualitativos dos dados quantitativos e nestes, distinguiremos os dados discretos dos dados contínuos.

2.1 Dados qualitativos ou categóricos

- Tabela de frequências
- Gráfico ou diagrama de pontos
- Pictograma
- Esquema de contagem (gráfica)
- Gráfico ou diagrama circular
- Gráfico ou diagrama de barras
- Moda

2.1 Dados qualitativos **Tabela de frequências**

Normalmente, o primeiro passo na organização de um conjunto de dados qualitativos ou categóricos, é contar quantas vezes aparece a mesma categoria ou classe(1), ou o mesmo valor, se as categorias estiverem codificadas.

Frequência absoluta – de uma categoria (ou classe) é o número de vezes que essa categoria surge no conjunto de dados.

Os dados são organizados numa **tabela de frequências** onde se consideram as diferentes categorias e as respectivas frequências absolutas. Além das frequências absolutas, também se apresentam as frequências relativas, onde

Frequência relativa – de uma categoria (ou classe) é a frequência absoluta a dividir pelo número de dados.

(1) À categoria também se pode dar o nome de classe.

2.1 Dados qualitativos **Tabela de frequências**

Do ficheiro "Dados sobre casas" considerem-se os dados referentes à variável Zona.

Tabela de frequências – apresenta as diferentes categorias, a *frequência absoluta* com que cada categoria surge no conjunto de dados e ainda as *frequências relativas*.

Zona	Frequência Absoluta (n_i)	Frequência Relativa (f_i)
A	19	0,475
B	16	0,400
C	5	0,125
Total	40	1,000

=19/40

Nº de dados

Fequência absoluta

de uma categoria (ou classe) é o total de elementos do conjunto de dados que pertence a essa categoria (ou classe).

Frequência relativa

É a frequência absoluta a dividir pelo número de dados.

2.1 Dados qualitativos **Tabela de frequências**

Na tabela de frequências é usual representar as frequências relativas em percentagem. Para representar as frequências relativas em **percentagem**, basta multiplicar por 100 e considerar o símbolo %. Ao considerarmos as frequências relativas em termos de percentagem, é mais simples de visualizar a forma relativa como as categorias se distribuem umas relativamente às outras, admitindo que o total é igual a 100%.

Zona	Frequência Absoluta	Frequência Relativa %
A	19	47,5
B	16	40,0
C	5	12,5
Total	40	100,0

Da tabela ao lado concluímos que predominam as casas da zona A, com uma percentagem de 47,5%.

2.1 Dados qualitativos **Tabela de frequências**

De um modo geral as frequências relativas dão mais informação do que as frequências absolutas.

Exemplo – A tabela seguinte mostra a distribuição do número de alunos do ensino básico pelos diferentes ciclos , no ano de 2013 (*Fonte – Pordata*)

Nível	Frequência Absoluta	Frequência Relativa %
1º ciclo	440.378	40,3
2º ciclo	252.667	23,1
3º ciclo	400.478	36,6
Total	1.093.523	100,0

Da tabela anterior concluímos que a percentagem de alunos no 1º ciclo é bastante superior à percentagem de alunos no 2º ciclo e um pouco superior à percentagem de alunos do 3º ciclo.

Nota

Na distribuição de frequências de um conjunto de dados qualitativos, devem considerar-se tantas categorias ou classes, quantas as diferentes categorias existentes nos dados. No entanto, se houver um número muito grande de categorias com pequenas frequências, podem-se agrupar algumas dessas categorias numa classe a que se dá o nome de *Outros* e cuja frequência absoluta será a soma das frequências absolutas das categorias consideradas em *Outros*.

2.1 Dados qualitativos **Tabela de frequências**

Se os dados que estão a ser representados na **tabela de frequências** forem *qualitativos ordinais*, podem-se acrescentar na tabela duas colunas com as *frequências acumuladas*. Para cada categoria, obtém-se a frequência acumulada adicionando à frequência dessa categoria, as frequências das categorias anteriores.

Exemplo – Classificações num teste dos 20 alunos da turma A

Classificação	Freq abs	Freq rel	Freq abs acum	Freq rel acum %
Insuficiente	3	0,15	3	15%
Suficiente	5	0,25	8	40%
Médio	8	0,40	16	80%
Bom	3	0,15	19	95%
Muito bom	1	0,05	20	100%
Total	20	1,00		

=(0,15+0,25)*100

A partir da tabela podemos concluir, por exemplo, que 80% dos alunos tiveram classificação igual ou inferior a médio.

Numa **tabela de frequências**:

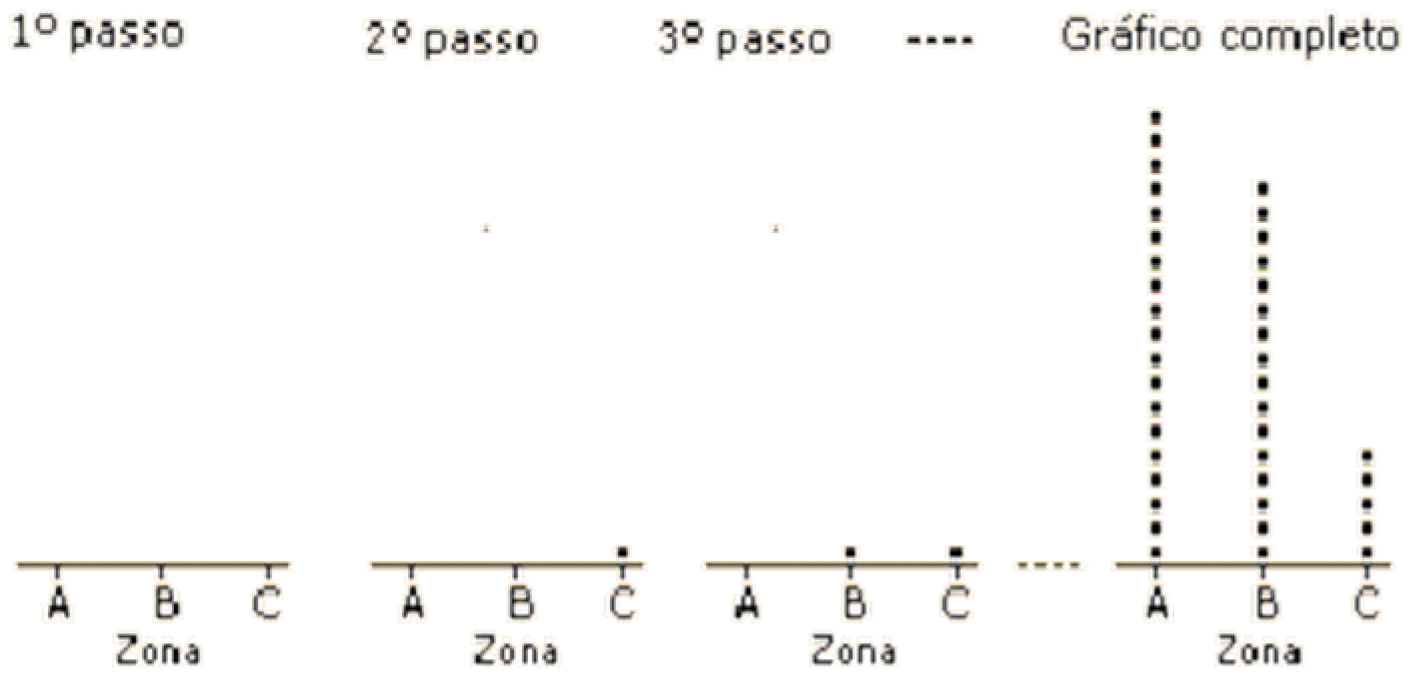
- A soma das frequências absolutas deve ser igual ao número de dados;
- É usual apresentar as frequências relativas em **percentagem**, multiplicando por 100, o valor que se obtém quando se divide a frequência absoluta pelo total de dados;
- A soma das frequências relativas deve ser igual a 1 (ou 100% se estiverem representadas em percentagem).

Nota

Em muitas situações as frequências relativas são dízimas infinitas obrigando, por isso, a arredondamentos. Estes têm de ser feitos com algum cuidado, de modo a que o total seja igual a 1 ou 100%.

2.1 Dados qualitativos **Gráfico ou diagrama de pontos**

Gráfico ou **diagrama de pontos** - desenha-se um eixo horizontal ou vertical, onde se assinalam as diferentes modalidades ou categorias da variável em estudo e, por cima de cada categoria, representa-se um ponto, sempre que ao percorrer o conjunto de dados se encontrar a respectiva categoria. Considerando os dados da variável Zona do ficheiro "Dados sobre casas"



2.1 Dados qualitativos **Gráfico ou diagrama de pontos**

O **gráfico** ou **diagrama de pontos** é uma representação muito simples de utilizar quando não se têm muitos dados.

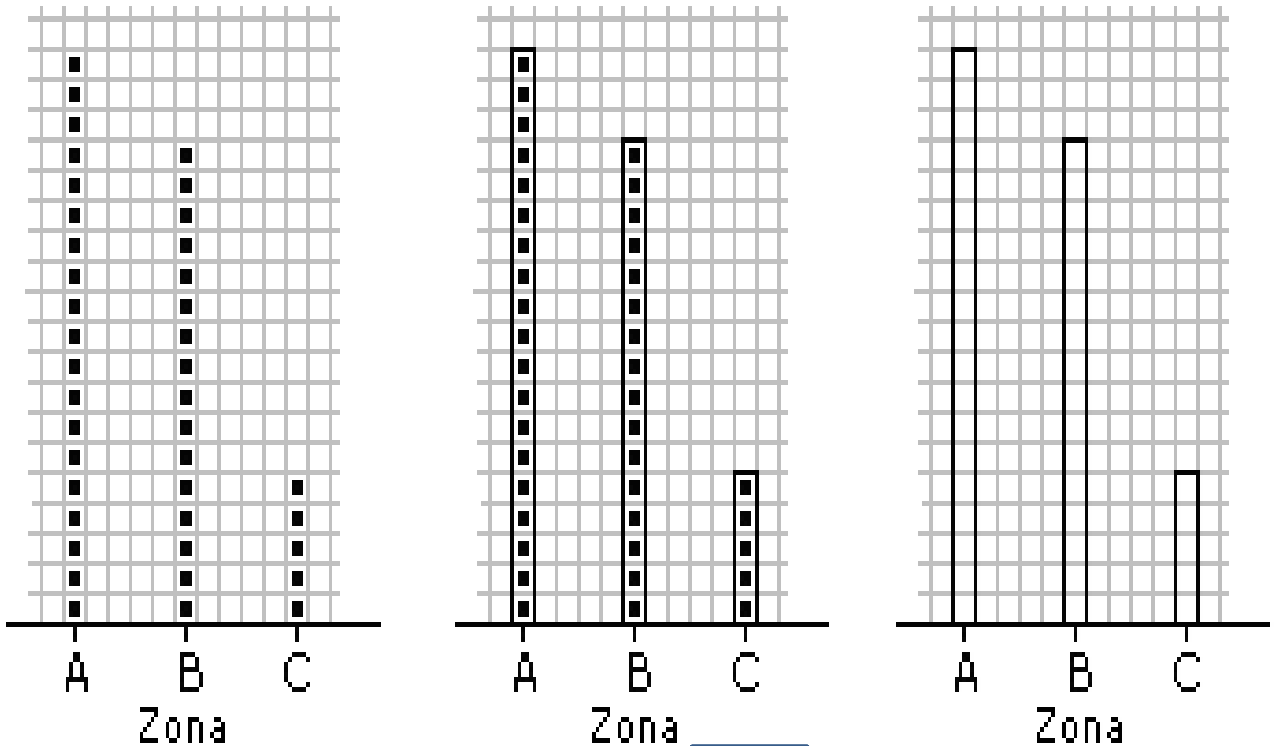
Para a sua construção não é necessário a construção prévia de uma tabela de frequências; pelo contrário, a partir de um diagrama de pontos, é mais fácil construir uma tabela de frequências, pois basta contar o número de pontos associados a cada categoria.

Este gráfico pode-se construir à medida que se vão recolhendo os dados. Aconselha-se a utilização de papel quadriculado para a sua construção.

Da representação gráfica anterior conclui-se imediatamente que no nosso conjunto de dados predominam as casas da zona A, sendo esta categoria a moda.

2.1 Dados qualitativos **Gráfico ou diagrama de pontos**

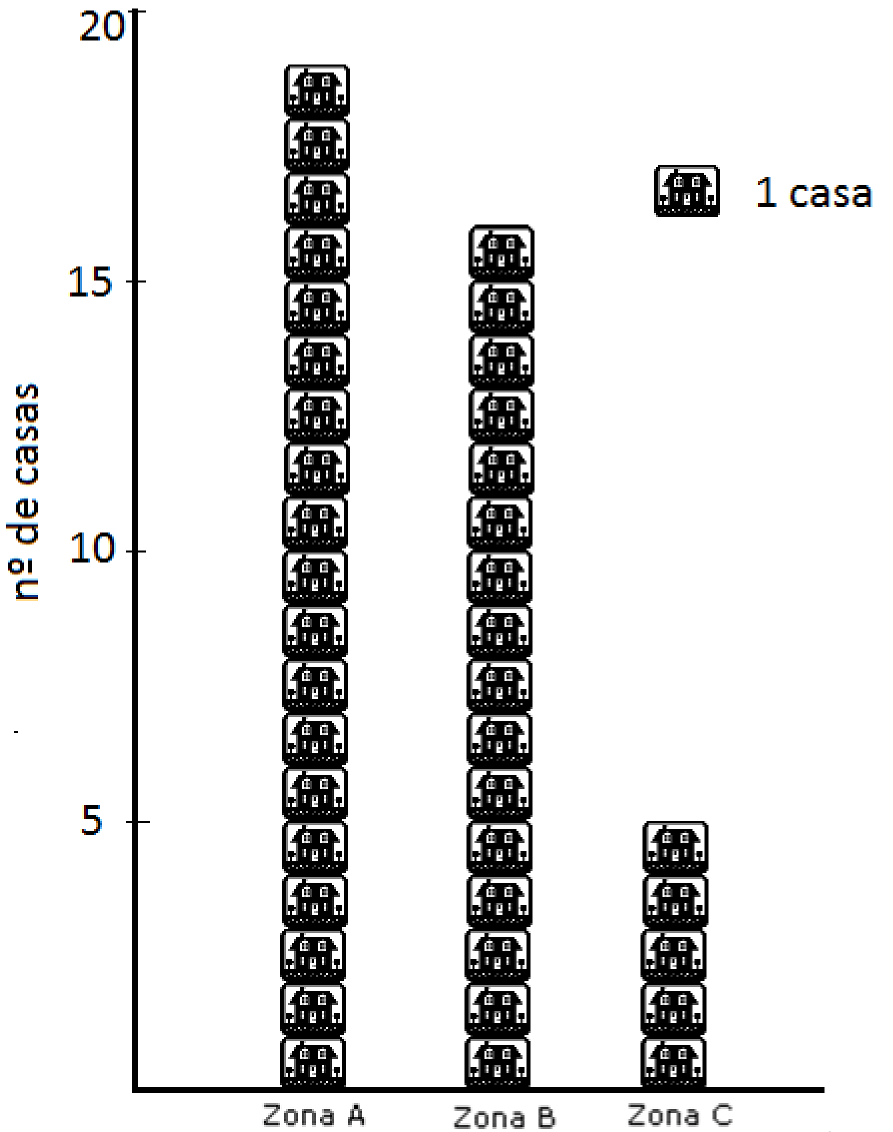
O gráfico de pontos evolui, como se apresenta a seguir, para um outro tipo de representação gráfica que se chama gráfico de barras. Como se verá à frente, para ter esta representação será necessário acrescentar um eixo vertical onde se marcam as frequências absolutas ou relativas



2.1 Dados qualitativos **Pictograma**

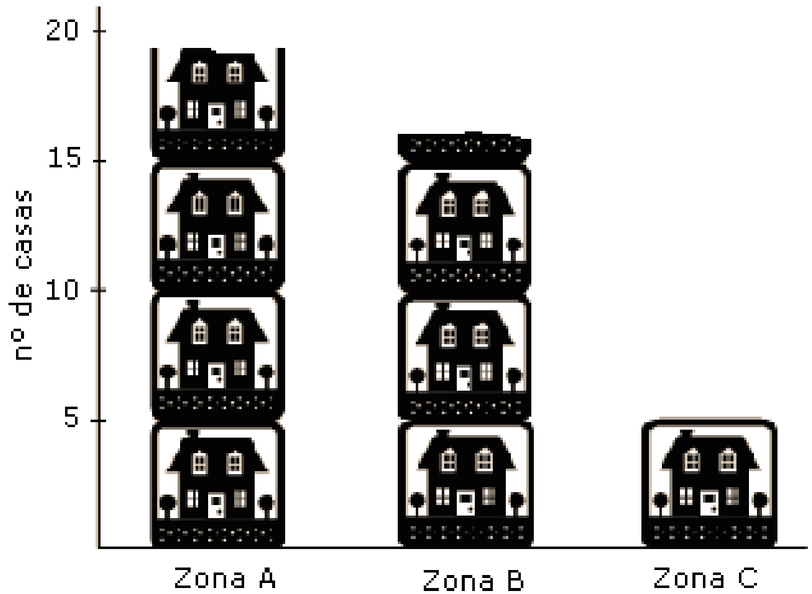
Pictograma - Começa-se por escolher uma figura ilustrativa da unidade estatística. Cada figura pode representar uma ou mais unidades estatísticas. De seguida, empilham-se as figuras que representam as unidades estatísticas até perfazer a frequência absoluta observada em cada categoria.

Ao lado temos o pictograma que aprresenta a forma como se distribuem as casas pelas categorias Zona A, Zona B e Zona C.



2.1 Dados qualitativos **Pictograma**

Construimos a seguir um pictograma com a mesma informação que o pictograma anterior, mas agora cada figura representa 5 casas



5 casas

No pictograma foi necessário cortar a figura para representar o 4 e o 1, pois 19 e 16 não são múltiplos de 5.

2 Tabela de frequências. Representações gráficas
2.1 Dados qualitativos **Pictograma**

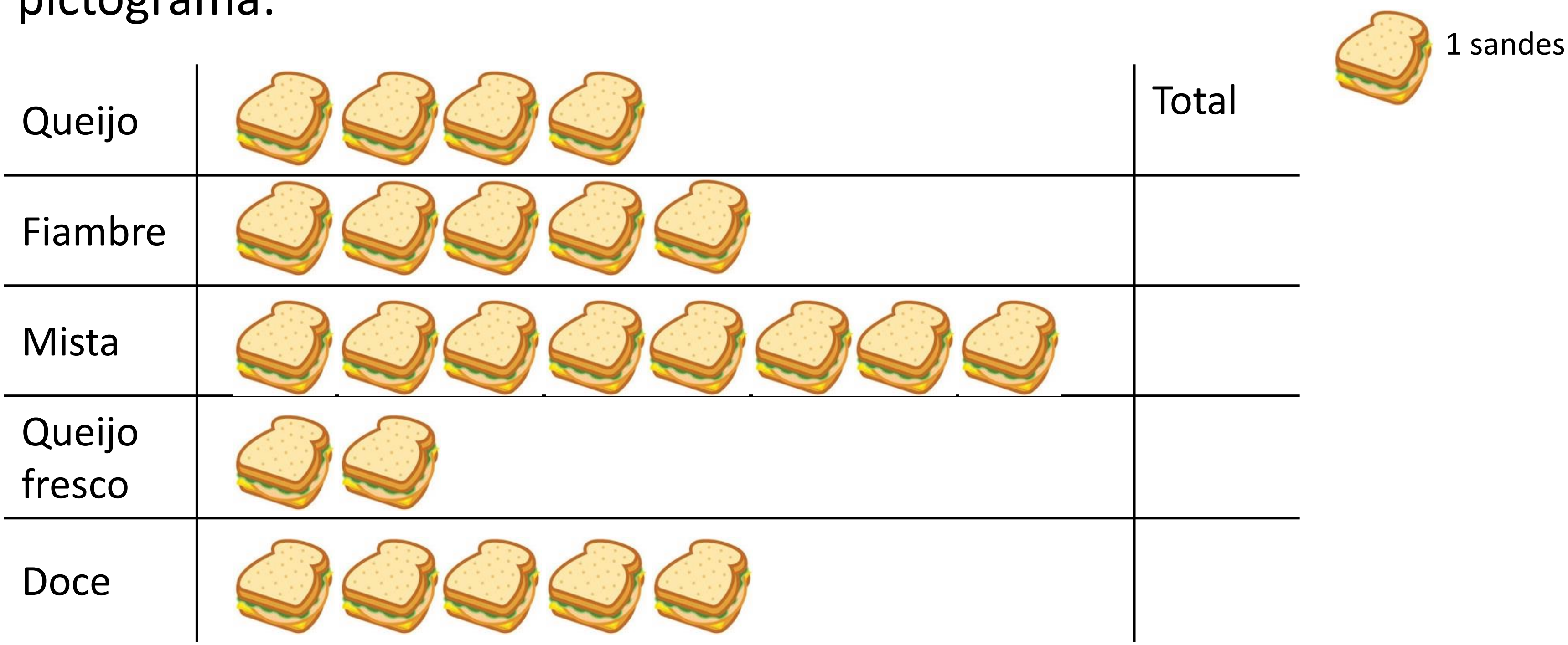
Exemplo - A sandes preferida – A professora de Matemática pediu aos alunos que fizessem um estudo sobre qual a sua sandes preferida. Assim, depois de averiguar quais os tipos de sandes que os alunos gostavam, deu um autocolante a cada um com o desenho de uma sandes  e pediu que fossem colocar no local adequado do esquema seguinte :

Queijo		Total
Fiambre		
Mista		
Queijo fresco		
Doce		

(continua)

2.1 Dados qualitativos **Pictograma**

Depois de todos terem colocado o seu autocolante, tinham o seguinte pictograma:



Algumas questões:

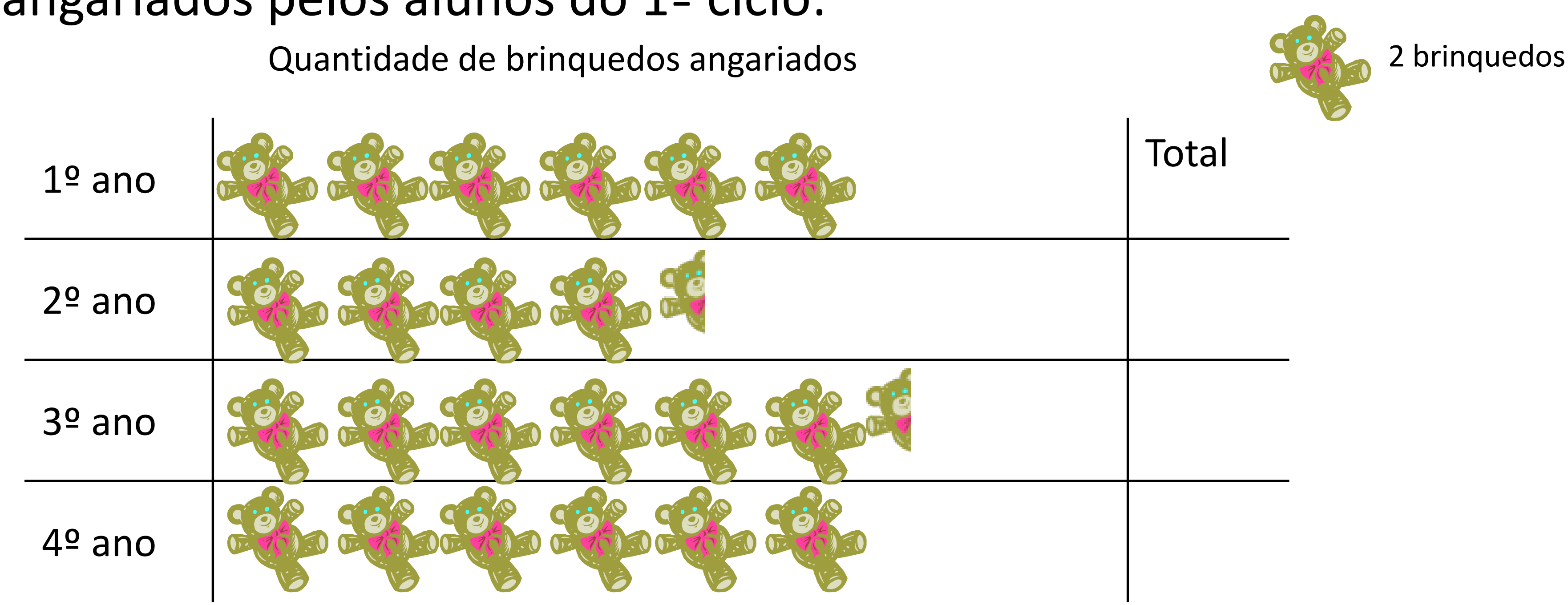
a) Quantos alunos estavam na sala de aula?

b) Qual a sandes preferida?

c) Há um tipo de sandes que tem o dobro das preferências, relativamente a outro tipo. Quais são estas sandes?

2.1 Dados qualitativos **Pictograma**

Exemplo – Recolha de brinquedos - Na escola pediu-se que cada aluno trouxesse um brinquedo que já não usasse, mas ainda em bom estado, para no Natal distribuir a algumas cianças. O pictograma seguinte mostra a quantidade de brinquedos angariados pelos alunos do 1º ciclo:



Responde às seguintes questões:

2.1 Dados qualitativos **Pictograma**

a) Quantos brinquedos conseguiram os alunos do

1º ano___2º ano___ 3º ano___ 4º ano___

a) Qual o ano que conseguiu mais brinquedos?

b) Qual o ano que conseguiu menos brinquedos?

c) Quantos brinquedos a mais conseguiu o 4º ano relativamente ao 3º ano?

d) A professora do 1º ano resolveu trazer 3 brinquedos e juntar aos brinquedos dos seus alunos. Como fica o pictograma?

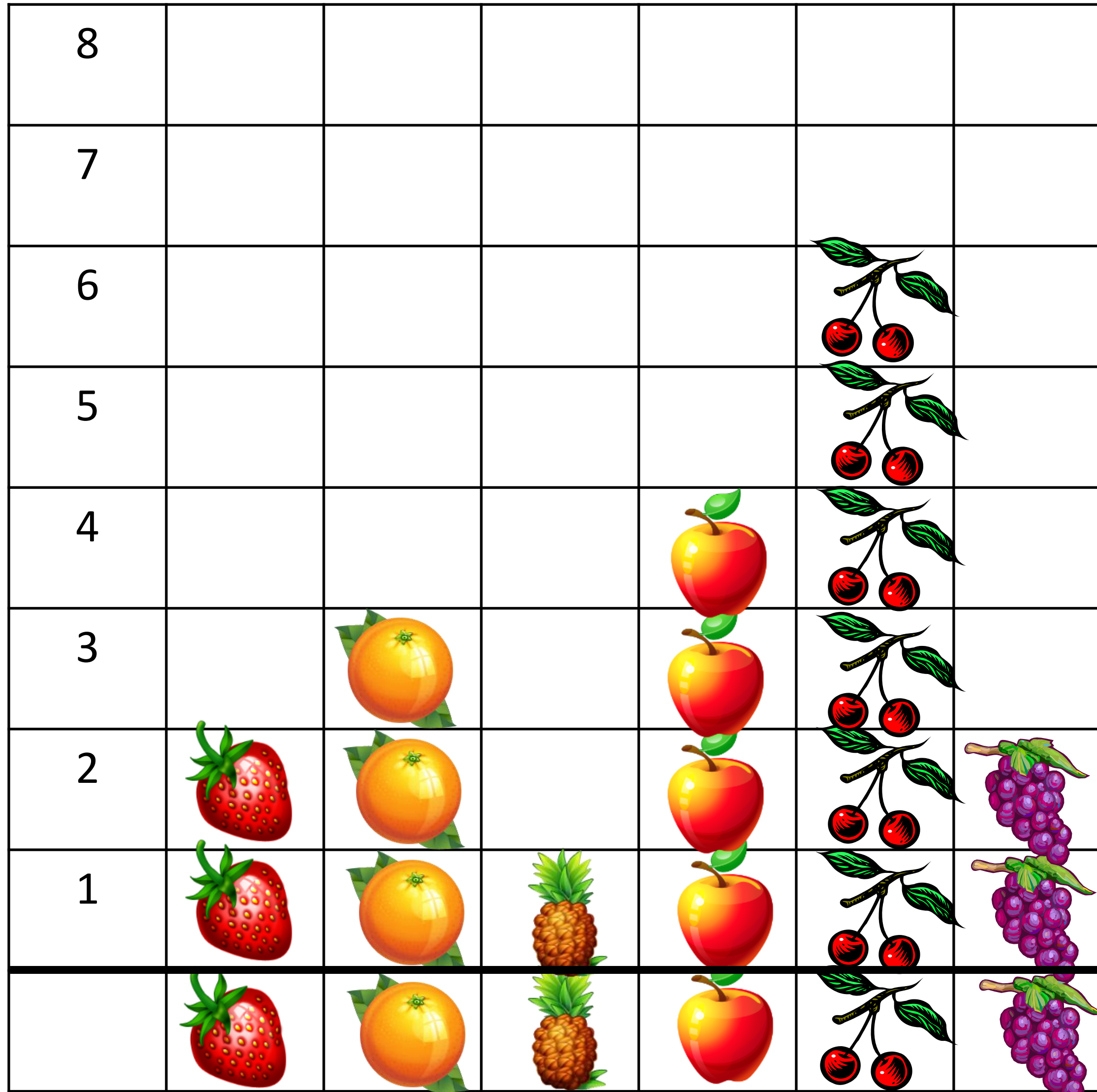
e) Quantos brinquedos conseguiram no total dos anos do 1º ciclo?

f) Uma loja de brinquedos decidiu juntar-se à campanha de angariação de brinquedos e deu a cada classe o dobro dos brinquedos que tinha. Com quantos brinquedos ficaram no total?

2.1 Dados qualitativos **Pictograma**

Exemplo – Observa o pictograma ao lado.

1. Dá um título ao pictograma.
2. Tendo em consideração o pictograma e o título que lhe deste, coloca algumas questões e responde a essas questões.



2.1 Dados qualitativos **Pictograma**

Nota

Um pictograma nem sempre transmite o mesmo tipo de informação.

Repare-se que, enquanto o pictograma considerado no exemplo da *Sandes preferida* apresenta, para cada tipo de sandes, a frequência absoluta de cada categoria, no pictograma do exemplo *Recolha de brinquedos*, o que está representado são os próprios dados. Para cada ano de escolaridade (do 1º ciclo) foi-se "observar" quantos brinquedos foram recolhidos, nomeadamente 12, 9, 13 e 12 respectivamente para o 1º, 2º, 3º e 4º ano. O nosso conjunto de dados é constituído por 4 dados, verificando-se que o 12 se verificou 2 vezes, sendo o dado mais frequente (moda).

Na interpretação do pictograma, deve-se ter em atenção sobre o que é que está representado.

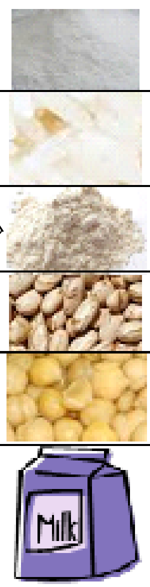
2.1 Dados qualitativos **Esquema de contagem**

Esquema de contagem - 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 (Graça Martins, M. E. et al, OTD, página 48):

- 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.

2.1 Dados qualitativos **Esquema de contagem**

Exemplo – O Banco Alimentar - Durante a campanha do Banco Alimentar Contra a Fome a escola decidiu participar e pediu aos alunos que pedissem aos pais 1 pacote (só um pacote) de qualquer um dos alimentos seguintes: açucar, arroz, farinha, feijão, grão e leite. À medida que os produtos iam chegando, a pessoa encarregada de os receber ia fazendo o seu registo e no fim obteve o seguinte esquema de contagem:



Açucar	卌 卌 ||	
Arroz	卌 卌 卌 |	
Farinha	卌 卌 卌 ||||	
Feijão	卌 卌 卌 ||	
Grão	卌 卌	
Leite	卌 卌 卌 |	

a) Quantos pacotes se receberam de açúcar?

b) Quantos pacotes se receberam de grão?

c) Se cada pacote de farinha tiver 1 quilograma, quantos quilograma se receberam de farinha?

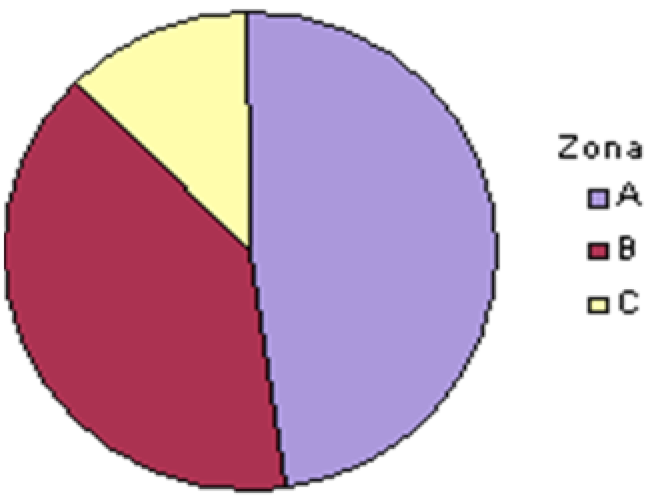
d) Se um pacote de grão tiver meio quilograma, quantos quilograma se receberam de grão?

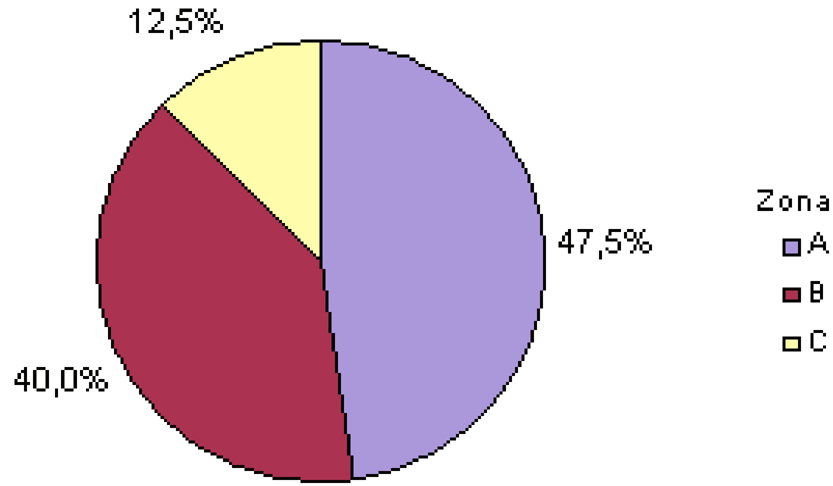
e) Completa a tabela anterior escrevendo na última coluna o total de pacotes recebidos de cada produto.

f) Qual o alimento que foi recebido em maior quantidade?

2.1 Dados qualitativos **Gráfico circular**

Gráfico ou **diagrama circular** - representação constituída por um círculo, em que se apresentam vários sectores circulares, tantos quantas as categorias consideradas na tabela de frequências do conjunto de dados em estudo. O ângulo de cada sector circular é proporcional à frequência observada na categoria que lhe corresponde. Para os dados da variável Zona, o sector circular correspondente à Zona A terá um ângulo de 360°x0,475=171°, o da Zona B terá um ângulo de 360°x0.400=144°, enquanto que o da Zona C terá 360°x0.125=45°. A soma dos 3 ângulos é 360°.





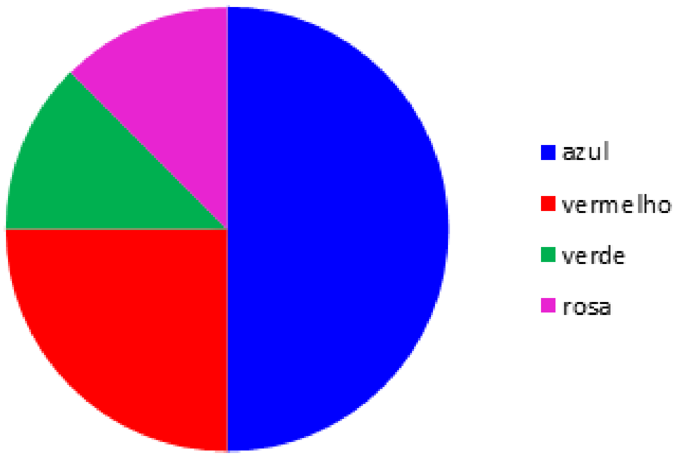
2.1 Dados qualitativos **Gráfico circular**

Para os alunos mais novos sugere-se a construção de gráficos circulares através de dobragens do círculo em partes iguais, considerando unicamente as situações em que seja adequado a divisão do círculo em 2, 4 ou 8 partes.

Também para os alunos mais novos sugere-se a consulta do processo descrito na página 90 da brochura Análise de dados (Maria Eugénia Graça Martins, Luísa Canto e Castro Loura e Maria de Fátima Mendes, DGIDC, 2007) para a construção do gráfico circular.

2.1 Dados qualitativos **Gráfico circular**

Exemplo - Cor favorita - Perguntou-se a algumas crianças qual a cor favorita. Os resultados estão apresentados no seguinte diagrama circular



Sabendo que o número de crianças que escolheu o azul é 20:

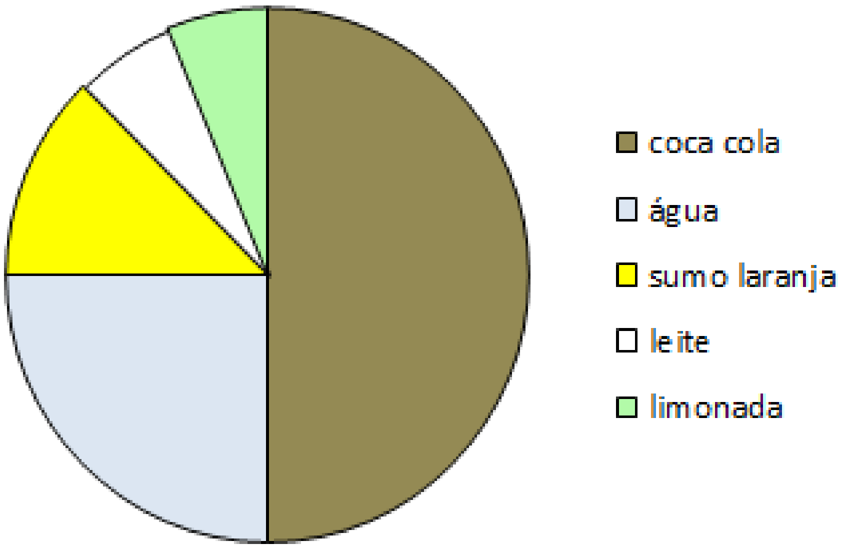
- Quantas crianças escolheram o vermelho?
- Quantas crianças escolheram o verde?
- Quantas crianças escolheram o rosa?
- Qual a cor mais popular?
- Quais as cores menos populares?
- A quantas crianças se colocou a questão?

Perguntou-se a 48 crianças qual a cor favorita e o gráfico circular obtido é exactamente igual ao anterior. Desta vez quantas crianças preferiram cada uma das cores?

Azul_____ Vermelho______ Verde_____ Rosa_____

2.1 Dados qualitativos **Gráfico circular**

Exemplo - Bebida favorita - Perguntou-se a algumas crianças qual a bebida favorita. Os resultados estão apresentados no seguinte diagrama circular



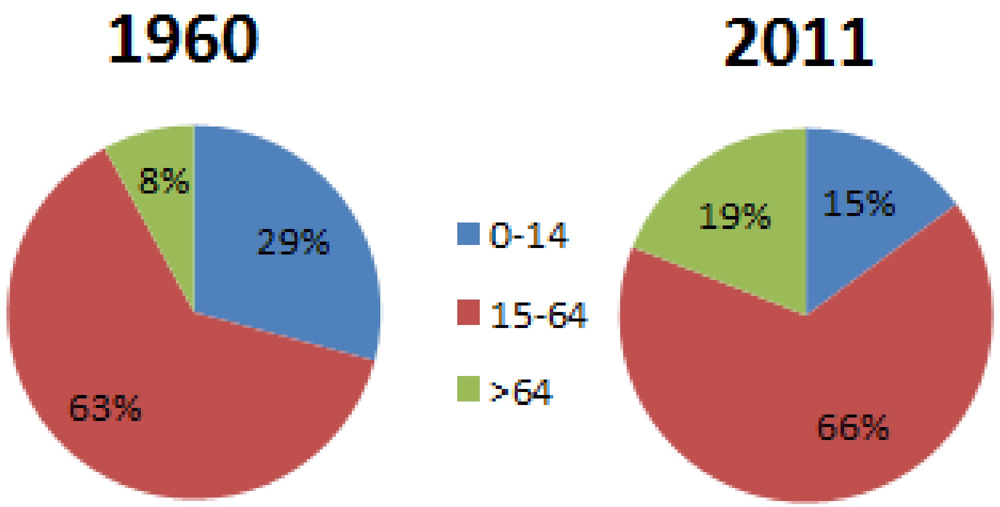
a) Qual a percentagem de crianças que prefere:

- A coca cola?
- A água?
- O sumo de laranja?
- O leite?
- A limonada?

b) Se 2 crianças preferiram a limonada, quantas crianças responderam à questão sobre a bebida favorita?

2.1 Dados qualitativos **Gráfico circular**

Exemplo – População residente - Segundo os censos de 1960 e 2011, a população residente em Portugal distribuia-se por grandes grupos etários da forma seguinte (www.pordata.pt):

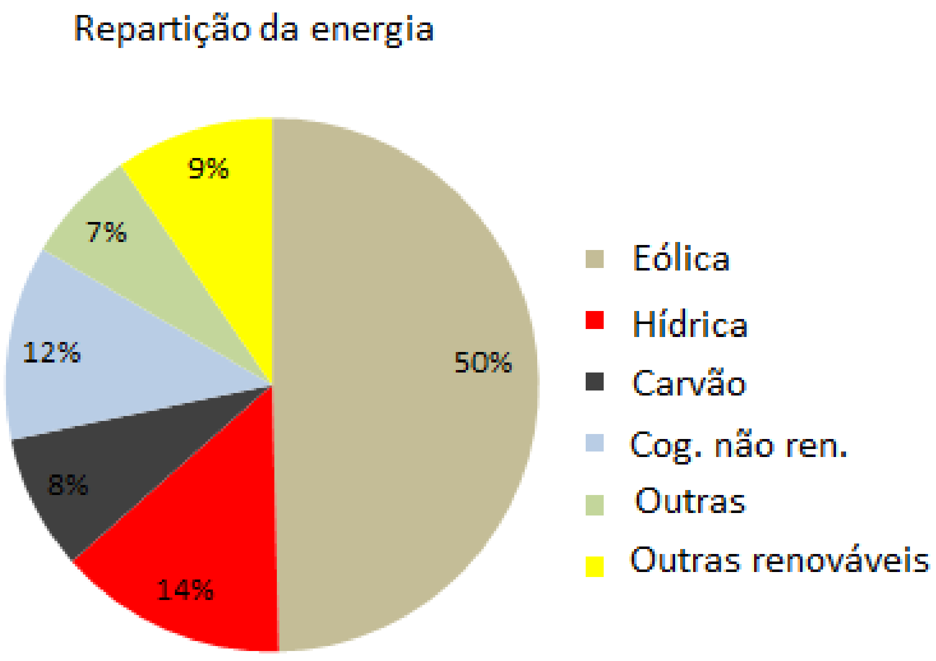


a) Tendo em conta os gráficos ao lado, quais as diferenças mais importantes na forma como a população residente se distribuia em 1960 e 2011?

b) Sabendo que segundo os censos 2011, o número de residentes com idade inferior ou igual a 14 anos era igual a 1.572.329, diz qual a população total residente em 2011.

c) Sabendo que de 1960 para 2011 o aumento da população foi de aproximadamente 18,8%, és capaz de obter um valor aproximado para a população residente em 1960?

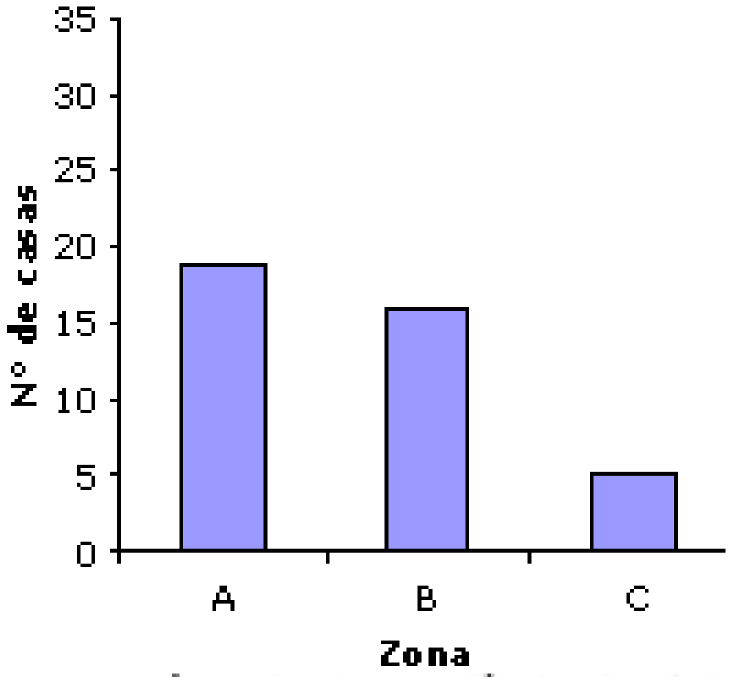
Exemplo – As fontes de energia - Num folheto distribuído pela *EDP Serviço Universal*, diz-se que a repartição da energia comercializada em 2013 se faz de acordo com o seguinte gráfico (www.edpsu.pt):



a) Qual a principal fonte de energia?
b) A energia obtida através do carvão é considerada energia renovável?
c) Dá um exemplo de uma energia renovável.
d) De acordo com o gráfico ao lado, qual a percentagem aproximada de energia considerada renovável distribuída pela EDP Serviço Universal?
e) É bom para o ambiente que haja uma grande percentagem de energias renováveis?

2.1 Dados qualitativos **Gráfico de barras**

Gráfico ou **diagrama de barras** – Considera-se um eixo horizontal (ou vertical), onde se marcam as categorias, igualmente espaçadas. Desenha-se uma barra para cada categoria, sendo a altura (comprimento) da barra igual à frequência absoluta ou frequência relativa observada nessa categoria. As barras devem ter a mesma largura.



Observando o gráfico de barras construído para os dados da variável Zona do ficheiro "Dados sobre casas" verifica-se que, no conjunto de dados representado, predominam as casas da zona A.

Nota – As barras devem ter a mesma largura e devem estar igualmente espaçadas (deve-se deixar igual espaço entre elas).

2.1 Dados qualitativos **Gráfico de barras**

A construção de um gráfico de barras pressupõe a construção prévia de uma tabela de frequências, para se saber qual a altura da barra associada a cada categoria (ou classe). De preferência devem-se utilizar as *frequências relativas*, para que a soma das alturas das barras seja igual a 1. Este cuidado permite a utilização do gráfico de barras para comparar conjuntos de dados de dimensão diferente. Por exemplo, suponha que na turma se recolheu informação sobre o tipo de transporte utilizado para ir para a escola, tendo-se obtido os seguintes resultados:

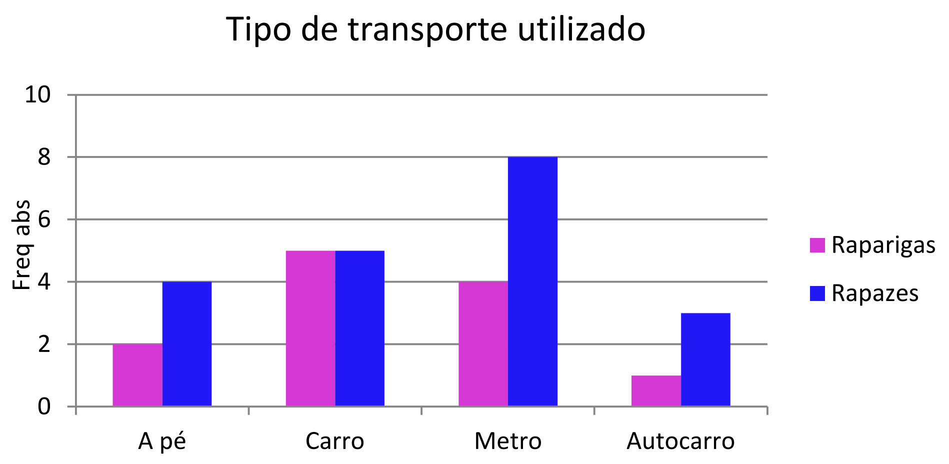
Raparigas

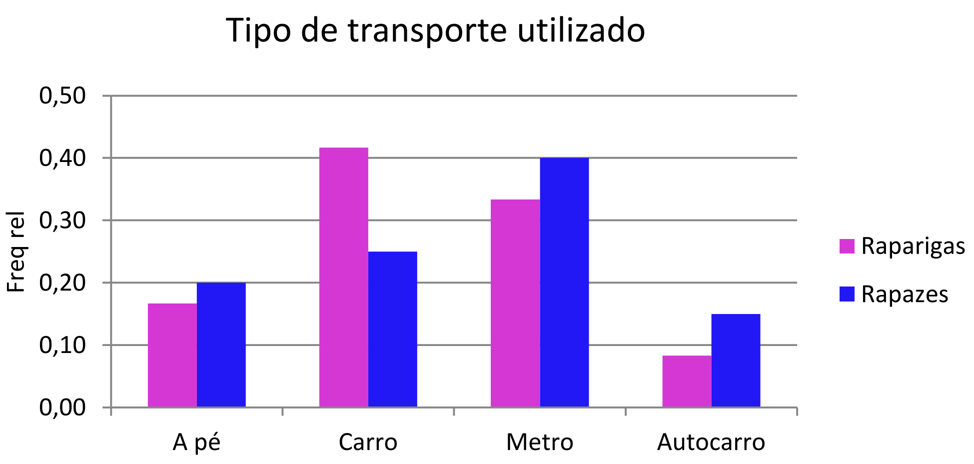
Categoria	Freq. abs	Freq. rel
A pé	2	0,17
Carro	5	0,42
Metro	4	0,33
Autocarro	1	0,08
Total	12	1,00

Rapazes

Categoria	Freq. abs	Freq. rel
A pé	4	0,20
Carro	5	0,25
Metro	8	0,40
Autocarro	3	0,15
Total	20	1,00

2.1 Dados qualitativos **Gráfico de barras**



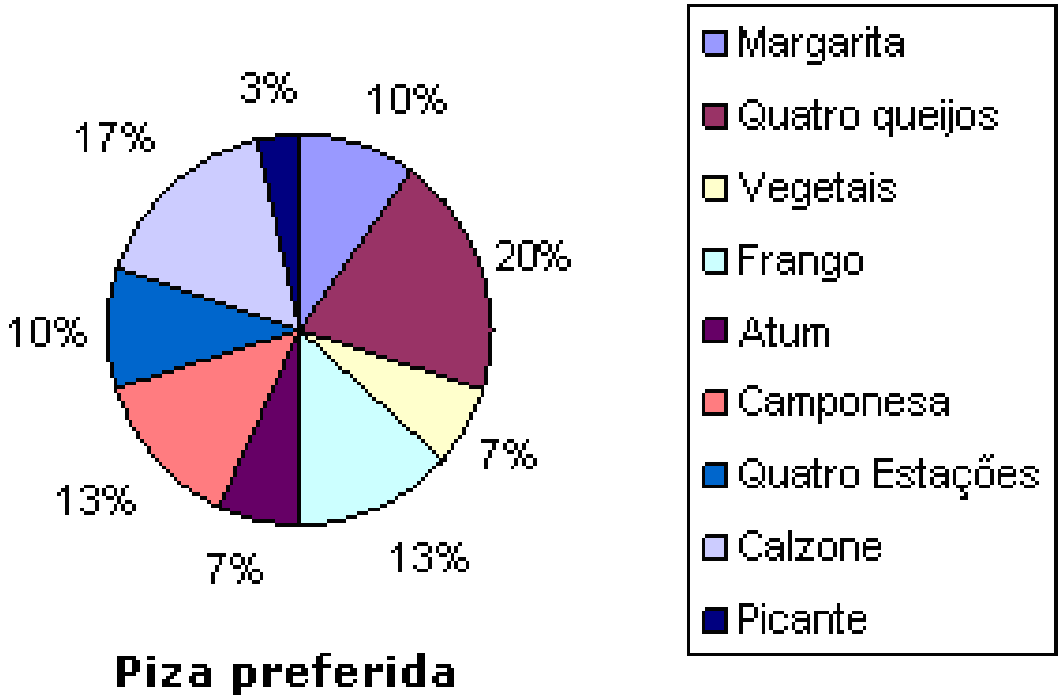


Repare-se que os dois gráficos ao lado transmitem informação diferente: o de cima indica que as raparigas andam tanto de carro como os rapazes e que há o dobro os rapazes que andam de metro, relativamente às raparigas! Mas não é verdade....

O gráfico de baixo, construído depois de se calcularem as frequências relativas, mostra que a percentagem de raparigas que anda de carro (42%) é bem superior à dos rapazes (cerca de 16%) e no que respeita a andar de metro existe uma diferença de apenas 7% entre a percentagem rapazes que o utiliza (40%) e a percentagem das raparigas (33%) que também o utiliza.

2.1 Dados qualitativos **Gráfico de barras e gráfico circular**

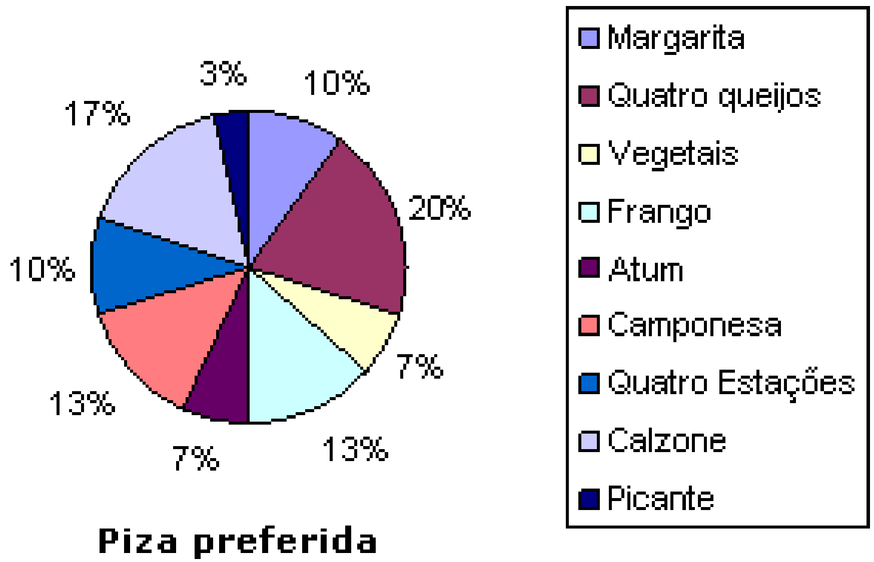
A utilização dos gráficos circulares merece alguns cuidados, nomeadamente quando o número de categorias que a variável assume for demasiado grande, tornando confusa a informação que procura transmitir. Por exemplo, admitamos que se perguntou aos alunos de uma turma qual a piza preferida, tendo-se construído o seguinte gráfico circular com os resultados obtidos (Graça Martins, M.E. et al, OTD, página 66):

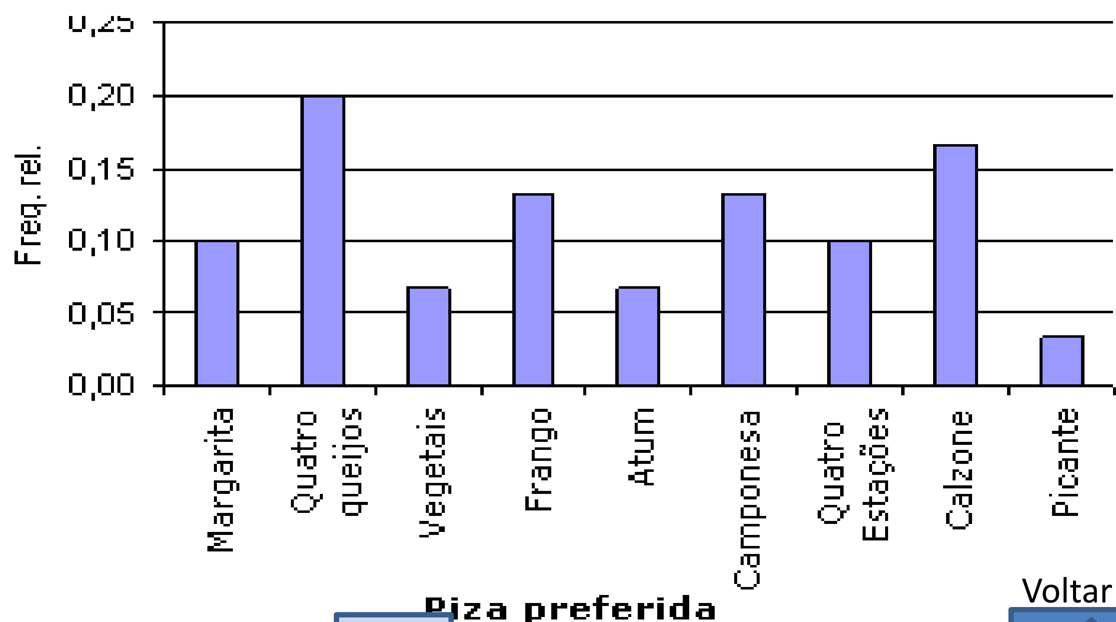


O gráfico está bem construído, com a legenda e as percentagens associadas às categorias indicadas, mas a mesma informação seria mais facilmente apreendida através de um gráfico de barras, como se apresenta a seguir, em que se torna mais fácil de visualizar as diferenças entre as frequências das diferentes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

(continua)

2.1 Dados qualitativos **Gráfico de barras e gráfico circular**





2.1 Dados qualitativos **Gráfico de barras e gráfico circular**

Aos mesmos alunos do exemplo anterior perguntou-se qual o meio de transporte que utilizavam para ir para a escola. Os resultados obtidos estão na tabela ao lado

(Graça Martins, M. E. OTD, página 68)

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

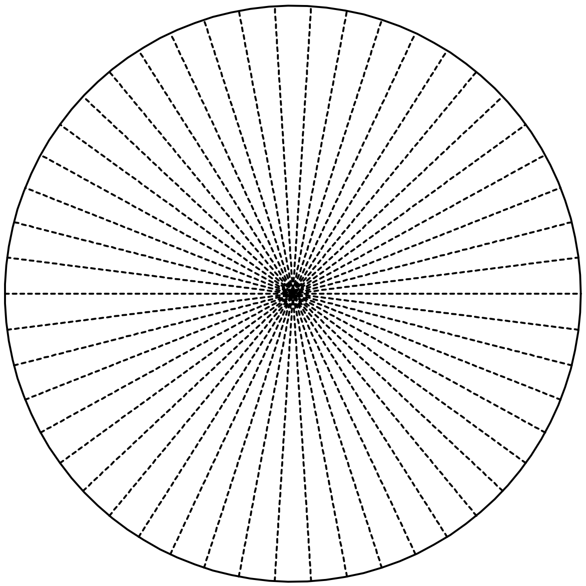
Completa a tabela abaixo com as frequências relativas, com 2 casas decimais e em percentagem:

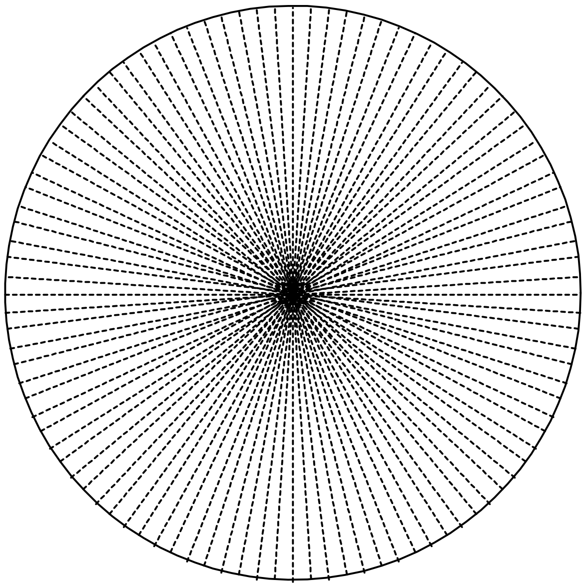
Transporte utilizado	N.º de alunos	Fracção do todo (2 casas decimais)	Fracção do todo (percentagem)
Carro	8		
Transportes públicos	21		
A pé	18		
Autocarro da Câmara	2		
Outro	1		
Total	50		

Continua

2.1 Dados qualitativos **Gráfico de barras e gráfico circular**

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

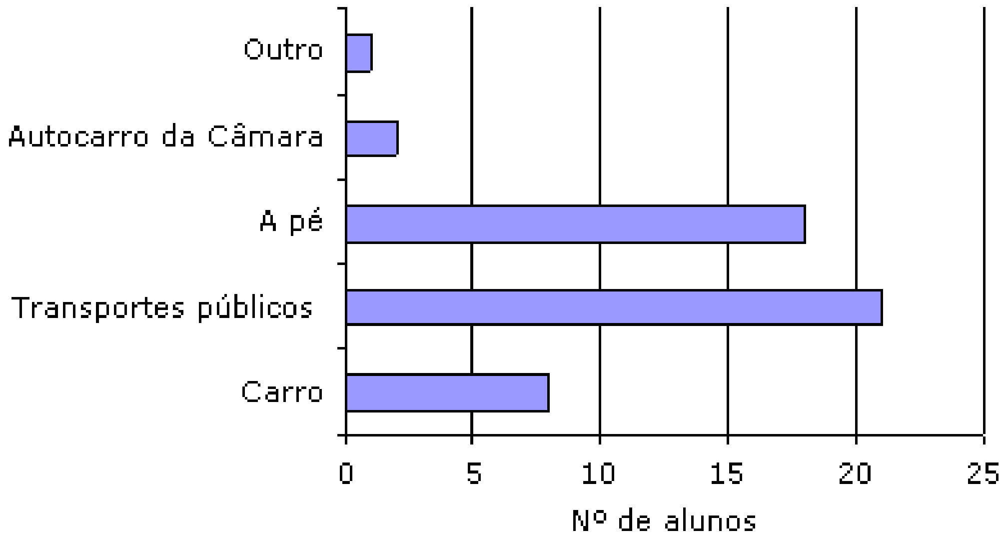




- Cada um dos círculos anteriores pode servir para construir gráficos circulares: num deles é mais fácil utilizar as frequências absolutas e no outro as frequências relativas (em percentagem) da tabela do slide anterior . Explica porquê.
- Constrói os gráficos circulares utilizando quer as frequências absolutas, quer as frequências relativas. Compara as representações obtidas e descreve o que concluíste.

2.1 Dados qualitativos **Gráfico de barras e gráfico circular**

- 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?

Gráfico circular ou Gráfico de barras?

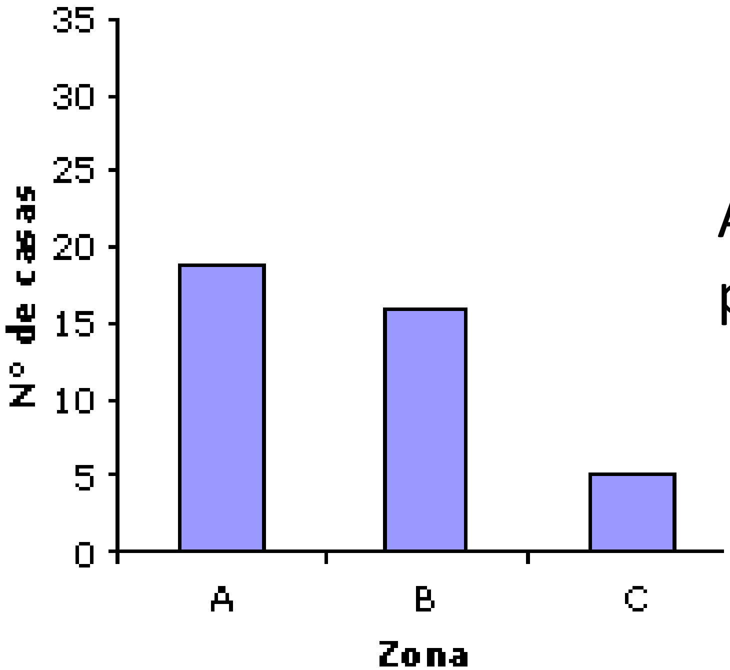
Anteriormente alertámos para o facto de ser necessário algum cuidado na utilização do gráfico circular, nomeadamente quando a distribuição a representar, apresenta muitas categorias (ou classes) ou quando os valores das frequências de algumas das categorias estão próximos.

No entanto, o gráfico circular é uma representação por excelência, quando o que se procura realçar é a forma como a totalidade dos 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%. Por outro lado o gráfico de barras realça a relação entre as frequências das categorias.

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.

2.1 Dados qualitativos **Moda**

Moda – categoria que ocorre com maior frequência no conjunto dos dados.

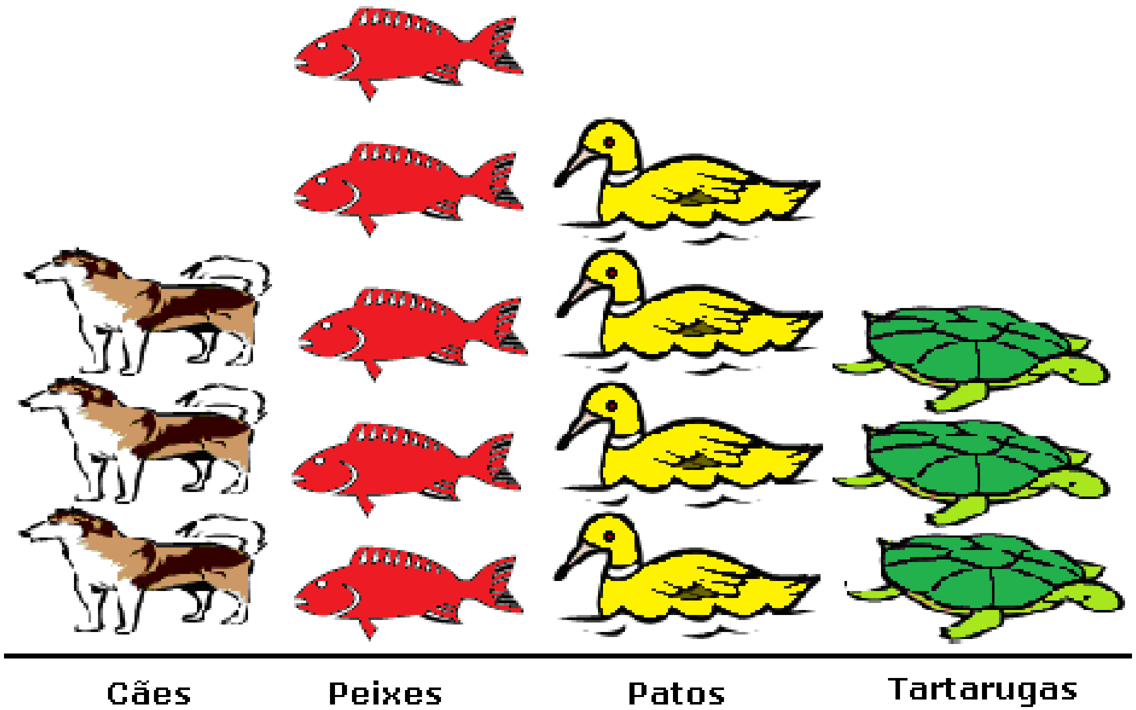


A **moda** é a categoria A, ou seja, predominam as casas na zona A

Nota – De um modo geral, para os dados qualitativos a única medida que se pode calcular é a **moda**. Existe uma excepção que é o caso dos dados serem *qualitativos ordinais*, em que também se pode calcular a *mediana*, como se verá, mais à frente.

2.1 Dados qualitativos **Exemplos**

Exemplo - 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:



Responde às seguintes questões:

- 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?

2.1 Dados qualitativos **Exemplos**

Exemplo - Animais do jardim (continuação) Neste exemplo, o que é o **dado**?

Neste caso, **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**:

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

Resumindo:

Unidade observacional ou estatística (elemento sobre o qual vai incidir o nosso estudo) – animal do jardim

Variável (ou característica de interesse) a estudar – tipo de animal

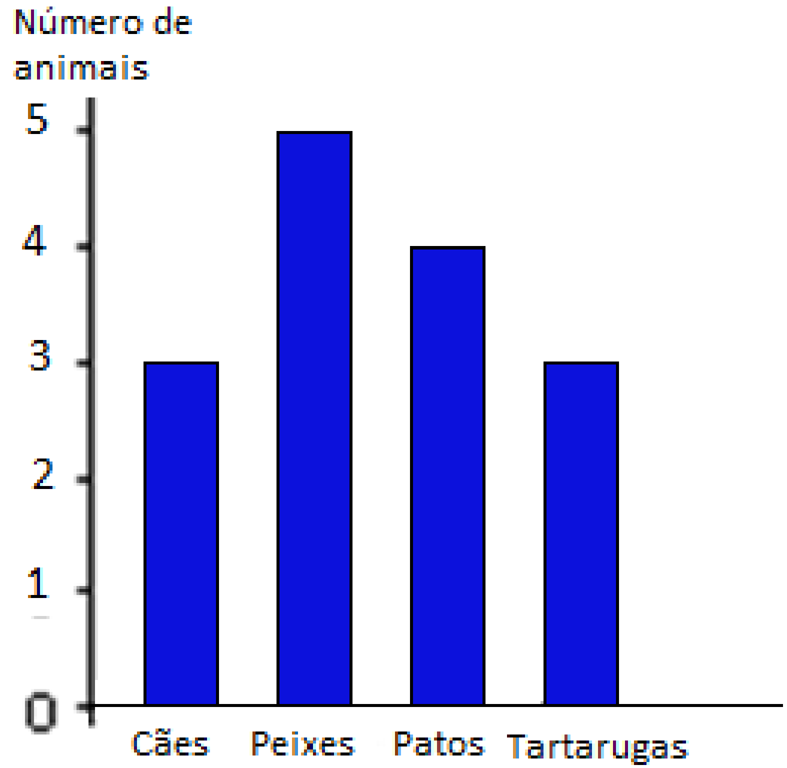
Dado – Resultado da observação do tipo de animal sobre a unidade estatística. Por exemplo, *Cão*, *Peixe*, etc.

Frequência absoluta de cada categoria – o número de vezes que cada dado surge no conjunto de dados. Por exemplo o Cão surge 3 vezes, o Peixe surge 5 vezes, etc.

2.1 Dados qualitativos **Exemplos**

Exemplo - Animais do jardim (continuação) Outra forma de representar a distribuição dos dados é utilizando um gráfico de barras, como se apresenta a seguir:

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



A informação transmitida pelo gráfico de barras é idêntica à do pictograma, embora esta seja mais aliciante para os alunos mais novos.

A moda é a categoria peixes, já que é a que apresenta maior frequência.

Exemplo - A sopa preferida – O sr. Rogério, gerente da cantina da escola, pediu à professora de Matemática do 4º ano que o ajudasse a saber qual a sopa preferida dos alunos que frequentam a cantina.

A professora aceitou ajudar o sr. Rogério e colocou este desafio aos alunos, que ficaram entusiasmados por poderem colaborar numa tarefa importante para a cantina da escola.

Como planear a forma de recolher a informação para responder à questão colocada pelo sr. Rogério? Foram seleccionadaos ao acaso 2 alunos da turma que ficaram encarregados de perguntar aos colegas qual a sopa preferida e a metodologia utilizada foi a seguinte: num dia destinado à realização da tarefa, assim que abriu a cantina, colocaram-se à porta e de 3 em 3 alunos colocavam a questão: - *Qual a tua sopa preferida*?

2.1 Dados qualitativos **Exemplos**

Exemplo - A sopa preferida (continuação) A resposta era anotada numa folha com o seguinte aspecto:

...	… … …

Os dois primeiros alunos a quem colocaram a questão responderam que a sua sopa preferida era sopa de agrião e canja, respectivamente. Assim, depois de anotadas estas respostas, tinham a seguinte tabela:

Agrião	I
Canja	I
…	… … …

2.1 Dados qualitativos **Exemplos**

Exemplo - A sopa preferida (continuação) Depois de ter terminado o intervalo de almoço, a tabela onde tinham registado as respostas tinha o seguinte aspecto:

Agrião	III
Canja	~~IIII~~ ~~IIII~~ ~~IIII~~ II
Caldo verde	~~IIII~~ III
Grão	IIII
Cenoura	~~IIII~~ ~~IIII~~
Legumes	~~IIII~~ I
Ervilhas	~~IIII~~

Tinham recolhido 53 respostas cujos resultados resumiram na seguinte tabela de frequências:

Sopa	Freq.abs.	Freq.rel.
Agrião	3	5,7%
Canja	17	32,1%
Cal. verde	8	15,1%
Grão	4	7,5%
Cenoura	10	18,9%
Legumes	6	11,3%
Ervilhas	5	9,4%
Total	53	100,0%

Conclusão, a ser transmitida ao Sr. Rogério:

- Sr. Rogério, colocámos a questão de qual a sopa preferida a 53 colegas. Das respostas obtidas, temos a indicação que a sopa preferida é a canja, com cerca de 32% de preferência, seguindo-se a sopa de cenoura, com Cerca de 19% de preferência.

 Propostas de exercícios

1. Cor dos olhos - Considera a seguinte tabela de frequências que apresenta os dados obtidos quando se observou a cor dos olhos de 20 alunos de uma sala de aulas:

Cor dos olhos	Frequência absoluta	Frequência relativa %
Castanhos	9	45
Pretos	5	25
Azuis	3	15
Verdes		
Cinzentos	1	5
Total	**20**	**100**

a) Quantos alunos têm olhos verdes?

b) Qual a cor de olhos que predomina?

c) Existem duas cores em que o número de alunos que tem uma dessas cores é o triplo do número de alunos que tem a outra. Quais são?

 Propostas de exercícios

2. Classificações - Considera a seguinte tabela de frequências que apresenta as classificações obtidas num teste de Português por um conjunto de alunos. As classificações estão numa escala de A a E, onde A é a mais baixa e E a mais alta:

Classificação	Frequência absoluta	Frequência relativa %
A	3	
B	7	
C	10	
D	11	
E	4	
Total		**100**

a) Quantos alunos fizeram o teste de Português?

b) Quantos alunos tiveram a classificação mais baixa?

c) Qual a percentagem de alunos que teve classificação menor ou igual a C?

 Propostas de exercícios

3. Animais - Considera a seguinte tabela de frequências absolutas.

Animal	Frequência absoluta
Cão	10
Gato	7
Peixe	8
Passarinho	6
Tartaruga	4
Total	

Inventa uma situação que possa ser respondida pela tabela anterior.

Coloca algumas questões que possam ser respondidas a partir da tabela.

Propostas de exercícios

4. Sanduiche preferida - O seguinte diagrama de pontos apresenta a resposta dada por alguns alunos da escola quando se lhes perguntou qual a sanduiche preferida:

a) Quantos alunos responderam sobre a sanduíche preferida?

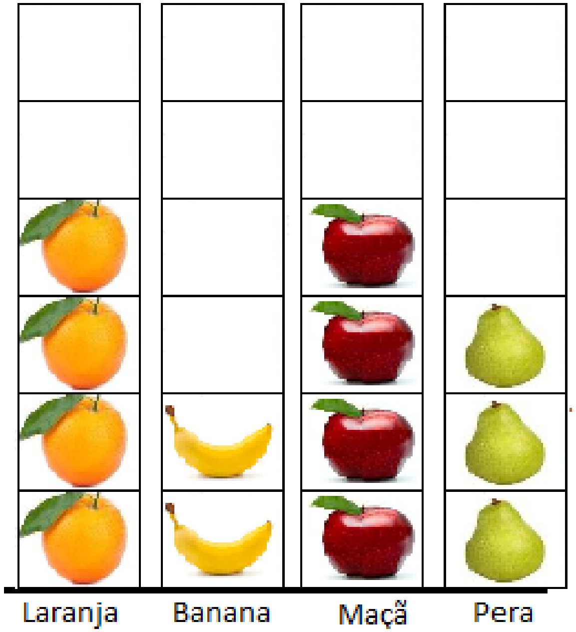
b) Qual a sanduiche preferida por estes alunos?

c) Qual a sanduiche que os alunos gostam menos?

d) O número de alunos que prefere a sanduiche mista é o dobro dos alunos que preferem qual sanduiche?

 Propostas de exercícios

5. Os frutos - A mãe da Maria foi ao supermercado e comprou 3 quilos de fruta variada. Chegou a casa e verificou que tinha comprado 6 bananas, 4 laranjas, 3 peras e 5 maçãs. Completa o pictograma seguinte de forma a representar a fruta que a mãe da Maria comprou.



a) Quantos frutos comprou a mãe da Maria?

b) Quantas bananas comprou a mais do que peras?

c) Completa a tabela de frequências seguinte, que representa o tipo de fruta que a mãe da Maria comprou:

Tipo de fruta	**Nº de frutos**
Laranja	4
Banana	6
	5
Pera	
Total	

6. O passeio - A professora queria levar os seus alunos a dar um passeio, pelo que decidiu perguntar a cada um qual a sua preferência e iriam onde fosse do agrado da maioria. As alternativas eram as seguintes: praia, jardim zoológico, parque, andar de bicicleta ou andar de skate. Estes nomes foram escritos no quadro e depois de todos os alunos terem manifestado a sua preferência tinha-se o seguinte esquema:

<table>
<tr><th></th><th></th><th>Total</th></tr>
<tr><td>Praia</td><td>| | |</td><td></td></tr>
<tr><td>Jardim zoológico</td><td>~~| | | |~~ | | |</td><td></td></tr>
<tr><td>Parque</td><td>| |</td><td></td></tr>
<tr><td>Bicicleta</td><td>| | | |</td><td></td></tr>
<tr><td>Skate</td><td>| | |</td><td></td></tr>
</table>

Completa o esquema anterior com os totais.

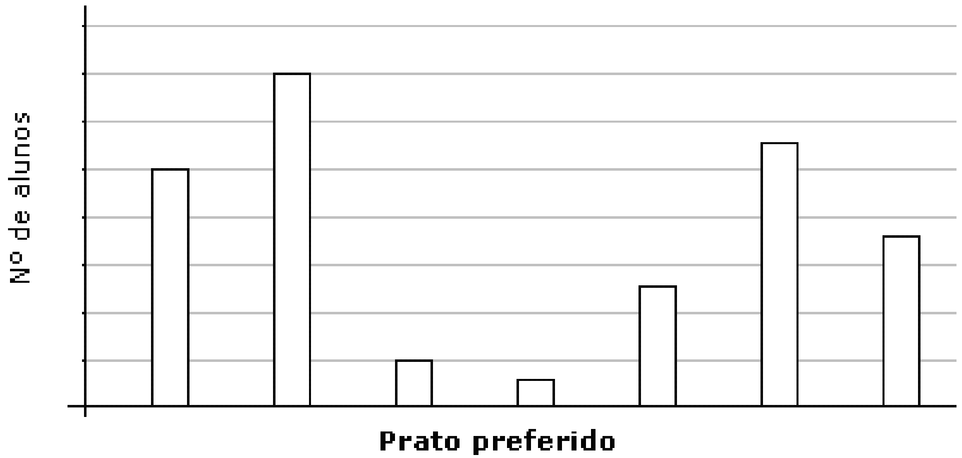
a) Onde é que a professora decidiu levar os alunos?

b) Quantos alunos preferiram ir ao jardim zoológico?

c) Qual o sítio menos votado?

d) Quantos alunos manifestaram a sua preferênca?

7. Prato preferido (Sugerida por uma actividade do CensusAtSchool)(Graça Martins, M.E. et al, 2011, página 56) – 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:

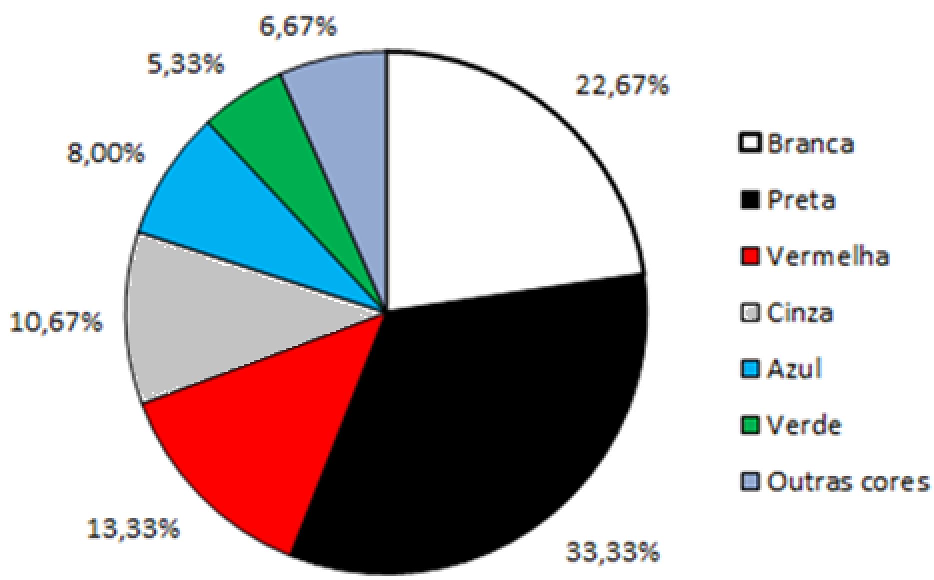
Prato preferido (Continuação)

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 *Esparguete 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).

8 . A cor dos carros – Estava-se a falar na sala de aula sobre a cor dos carros e alguém afirmou que a cor branca estava na "moda"! Será verdade que é uma cor muito frequente? Vamos averiguar...

Formaram-se grupos de 2 alunos, em que em cada intervalo, cada grupo se iria colocar à porta da escola e registar as cores dos carros que passavam. Ao fim de 1 dia de recolha de dados, os alunos juntaram-se e apresentaram a informação recolhida no seguinte gráfico:



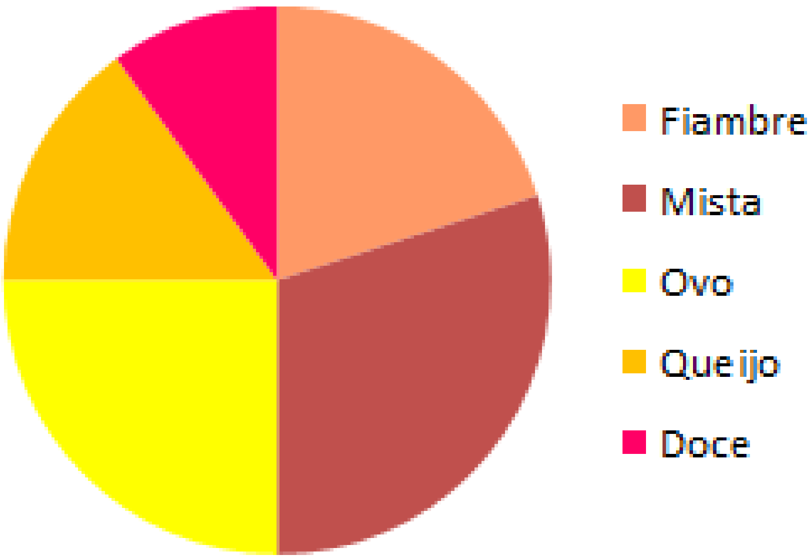
A partir do gráfico, responde às seguintes questões:

1. Qual a cor mais frequente?
2. Qual a moda?
3. Achas que a cor branca está na "moda"(1)?
4. Se o número de carros de cor azul é igual a 6, quantos carros foram observados?

(1) Em linguagem comum estar "na moda" não significa que seja necessariamente a situação mais frequente...

Propostas de exercícios

9 . Sanduiche preferida - O seguinte gráfico circular apresenta a resposta dada por alguns alunos da escola quando se lhes perguntou qual a sanduiche preferida:



A partir do gráfico, responde às seguintes questões:

1. Qual a sanduiche preferida?
2. Qual a sanduiche que os alunos gostam menos?
3. Se 5 alunos responderam que preferiam sanduiche de ovo, quantos alunos responderam a esta questão?

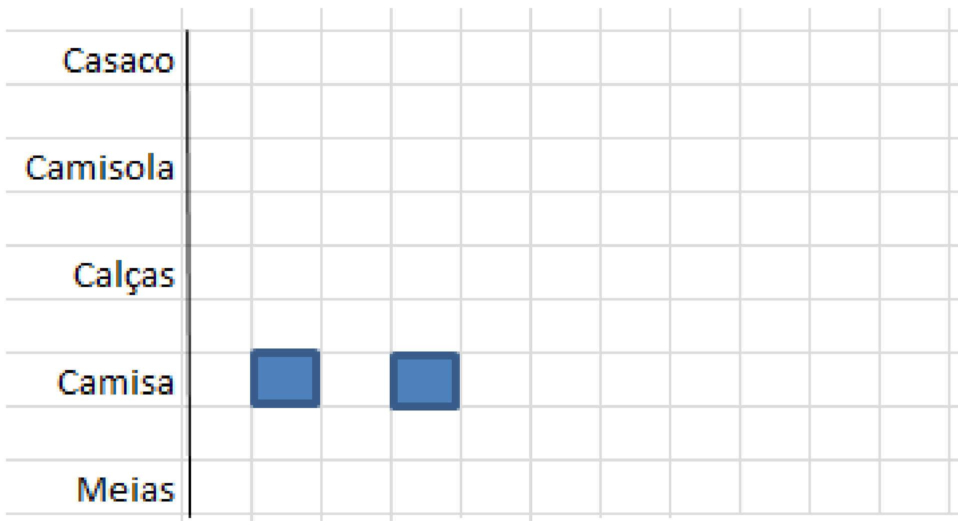
4. Se tivesse havido o dobro dos alunos a escolher cada uma das sanduiches, o que é que se alterava no gráfico circular anterior? Explica porquê.

2.1 Dados qualitativos **Propostas de exercícios**

10 . Recolha de roupa – Na campanha de recolha de roupa para distribuir no Natal, uma determinada instituição apresentou a seguinte tabela de frequências sobre o tipo e quantidade de roupa recolhida:

Tipo	Casacos	Camisolas	Calças	Camisas	Meias
Quantidade (em kg)	25	20	15	10	5

Completa o seguinte pictograma



= ____ kg

2.2 Dados quantitativos discretos

- Tabela de frequências
- Gráfico ou diagrama de pontos
- Gráfico ou diagrama de barras
- Caule-e-folhas
- Confusão entre dado discreto e frequência absoluta

2.2 Dados quantitativos discretos **Tabela de frequências**

Normalmente, o primeiro passo na organização de um conjunto de dados quantitativos discretos, é organizá-los numa tabela de frequências.

Tabela de frequências – dado um conjunto de dados discretos, estes são organizados, no mínimo, em 3 colunas: coluna das *classes* – onde se indicam todos os valores distintos que surgem no conjunto de dados, que representamos por x^*_i; coluna das *frequências absolutas* – onde se regista o total de elementos do conjunto de dados que pertencem a cada classe (ou número de vezes que cada valor x^*_i surge no conjunto de dados) e coluna das *frequências relativas* (ou percentagens) – onde se coloca, para cada classe, o valor que se obtém dividindo a respectiva frequência absoluta pelo número de dados. A tabela de frequências pode ainda incluir mais 2 colunas: a coluna das *frequências absolutas acumuladas* e a coluna das *frequências relativas acumuladas.*

2.2 Dados quantitativos discretos **Tabela de frequências**

Para a variável *Número de assoalhadas*, do ficheiro “Dados sobre casas”, temos a seguinte tabela de frequências:

Nº de Assoalhadas x^*_i	Freq. Abs. n_i	Freq. Rel. f_i	Freq. Abs. Acum.	Freq. Rel. Acum.
1	3	0,075	3	0,075
2	17	0,425	20	0,500
3	16	0,400	36	0,900
4	2	0,050	38	0,950
5	2	0,050	40	1,000
Total	**40**	**1,000**		

= 3+17

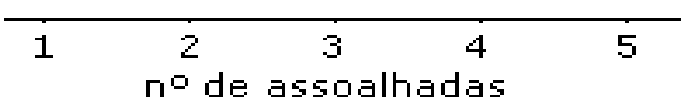
= 0,075+0,425

Da tabela anterior conclui-se, por exemplo, que 90% das casas têm 3 ou menos assoalhadas e que só 10% das casas têm 4 ou 5 assoalhdas.

2.2 Dados quantitativos discretos **Gráfico de pontos**

Gráfico ou **diagrama de pontos** - Tal como no caso de dados qualitativos ou categóricos, a representação gráfica mais simples é o gráfico ou diagrama de pontos. Para obter essa representação, basta traçar um eixo horizontal (ou vertical), onde se assinalam os diferentes valores que surgem no conjunto de dados ou mais correctamente, todos os valores entre o mínimo e o máximo, incluindo estes. Por cima de cada valor marca-se um ponto, sempre que se encontar um valor igual, ao percorrer o conjunto de dados. Para a variável *Número de assoalhadas*, temos:

1º passo

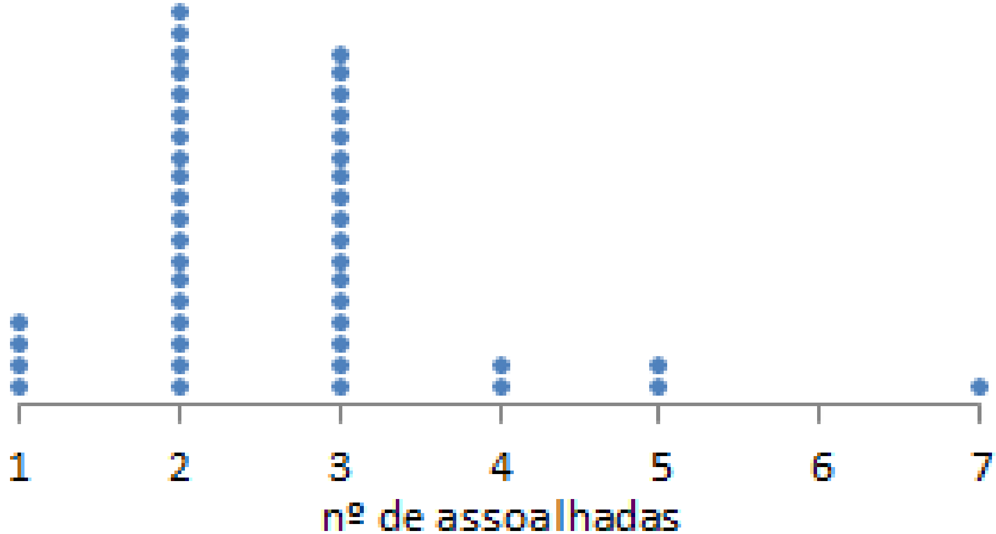


2º passo

2.2 Dados quantitativos discretos **Gráfico de pontos**

Da representação anterior conclui-se que predominam as casas com 2 assoalhadas, seguindo-se de perto as casas com 3 assoalhadas. Esta representação gráfica tem a vantagem de se poder construir à medida que se vão recolhendo os dados.

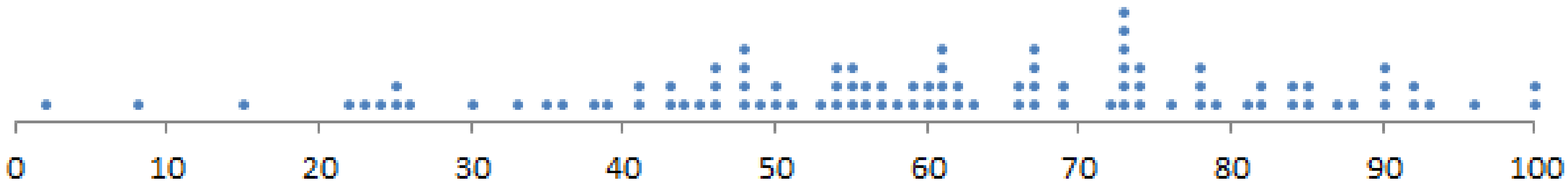
Suponhamos que além dos dados considerados anteriormente, ainda se tinham observado mais 5 casas, com o seguinte número de assoalhadas: 1, 2, 2, 7, 3. A representação anterior facilmente se adapta considerando mais estes dados:

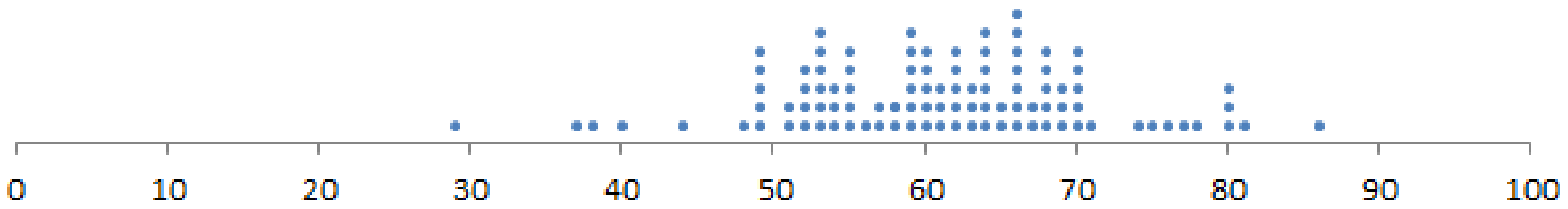


O gráfico ao lado mostra que existe um dado que é o 7, que pode ser considerado um pouco discrepante relativamente aos restantes.

2.2 Dados quantitativos discretos **Gráfico de pontos**

Exemplo – As notas das turmas A e B - Considerem-se os dois gráficos seguintes que representam as notas obtidas pelas turmas A e B no mesmo teste de matemática



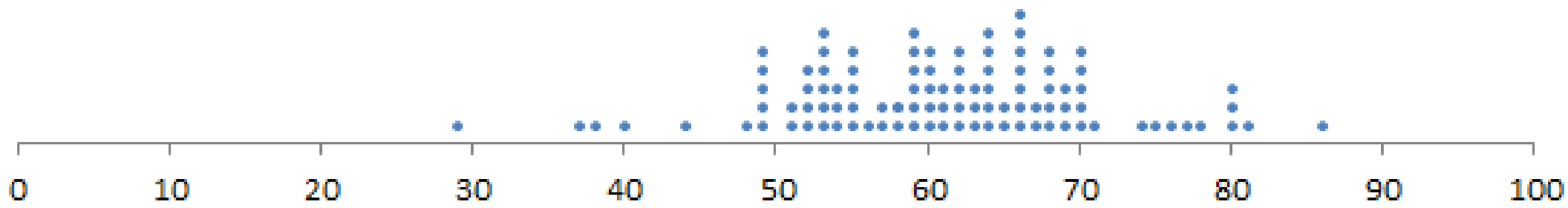


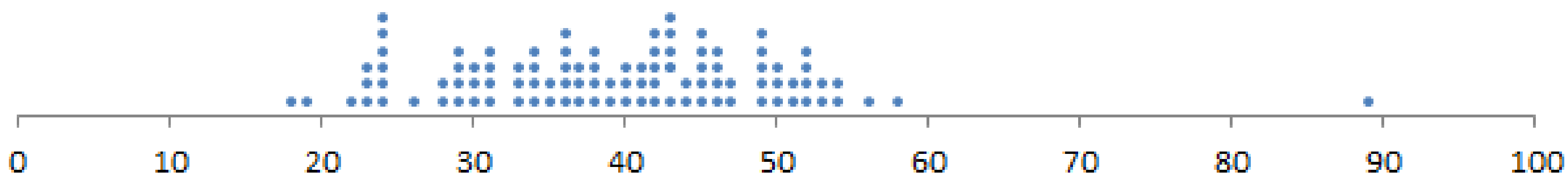
Comenta as principais diferenças apresentadas na distribuição das notas das duas turmas, referindo:

a) A simetria
b) A variabilidade ou dispersão das notas
c) O centro da distribuição
d) A existência de algumas notas discrepantes relativamente às restantes.

2.2 Dados quantitativos discretos **Gráfico de pontos**

Exemplo – As notas das turmas B e C - Considerem-se os dois gráficos seguintes que representam as notas obtidas pelas turmas B e C no mesmo teste de matemática





Comenta as principais diferenças apresentadas na distribuição das notas das duas turmas, referindo:

a) A simetria

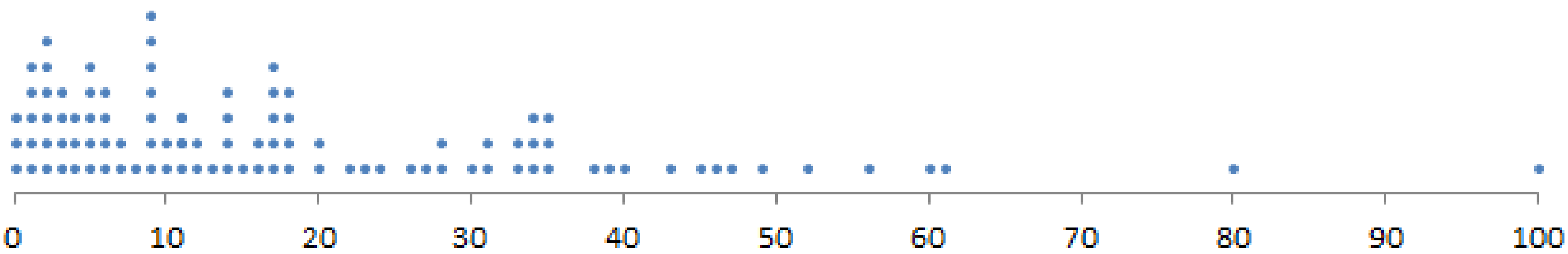
b) A variabilidade ou dispersão das notas

c) O centro da distribuição

d) A existência de algumas notas discrepantes relativamente às restantes.

2.2 Dados quantitativos discretos **Gráfico de pontos**

Exemplo – As notas a Língua Portuguesa - O gráfico seguinte apresenta a distribuição dos resultados obtidos num teste de Língua Portuguesa numa turma do 6º ano.



Os alunos quando acabaram o teste disseram ao professor que o teste tinha sido muito difícil. O professor tinha mandado os alunos ler um livro nas férias do Natal e o teste, depois de férias, continha essencialmente questões relacionadas com o conteúdo do livro. Tendo em consideração os resultados apresentados no gráfico anterior, responde às seguintes questões:

1) Os resultados do teste foram bons ou maus?

2) O que é que achas que aconteceu? Será que o teste era mesmo difícil ou poderá ter havido outra razão para os resultados obtidos?

2.2 Dados quantitativos discretos **Gráfico de barras**

Diagrama ou **gráfico de barras** – Depois dos dados organizados na forma de uma tabela de frequências, o gráfico ou diagrama de barras é uma representação gráfica que consiste em marcar num sistema de eixos coordenados, no eixo horizontal, o valor correspondente a cada classe e, nesses pontos, barras verticais de altura igual (ou proporcional) à respectiva frequência absoluta ou relativa. É preferível utilizar as frequências relativas sempre que se pretenda comparar conjuntos de dados de diferente dimensão. Para a variável *Número de assoalhadas* do ficheiro "Dados sobre casas", temos:

Do gráfico ao lado conclui-se que predominam as casas com 2 assoalhadas, seguindo-se, de perto, as de 3 assoalhadas.

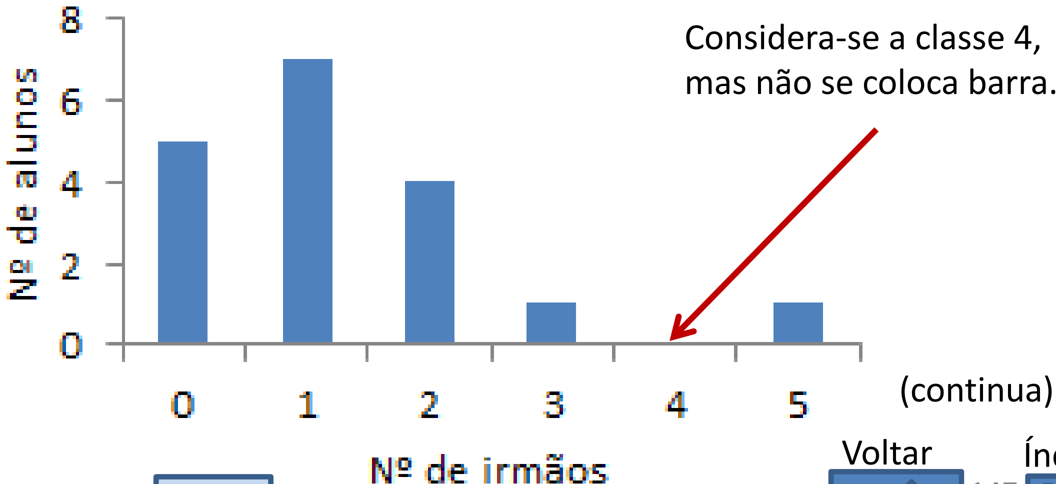
2.2 Dados quantitativos discretos **Gráfico de barras**

Na construção do **gráfico de barras** deve ter-se especial atenção ao seguinte: no eixo horizontal, deve ser marcada a sequência completa dos valores, entre o mínimo e o máximo observados, mesmo que algum esteja em falta no conjunto dos dados. Nesse caso não haverá qualquer barra vertical nesse ponto.

Exemplo – Número de irmãos - Recolheu-se a informação sobre o número de irmãos de cada aluno da turma, tendo-se obtido os dados resumidos na tabela abaixo.

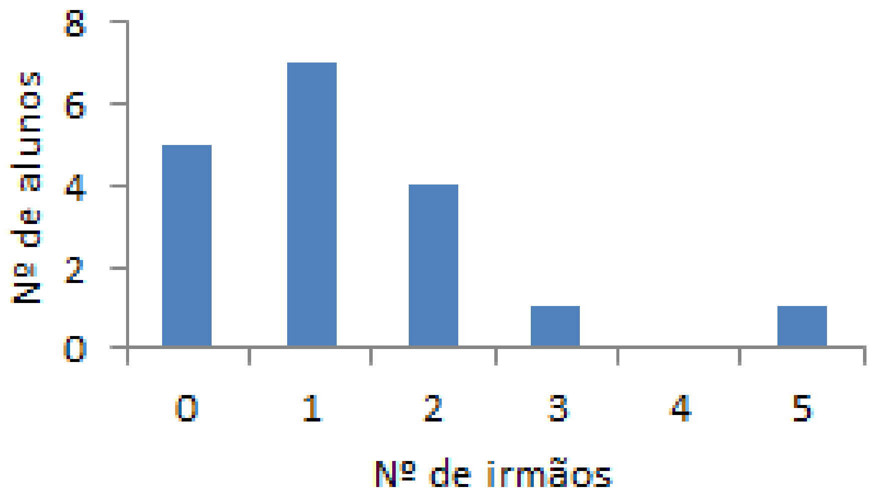
Embora não exista a classe 4, ela terá de ser representada no gráfico

Nº irmãos	Nº alunos
0	5
1	7
2	4
3	1
5	1



(continua)

2.2 Dados quantitativos discretos **Gráfico de barras**



A partir do gráfico anterior, em que se representa o número de irmãos de todos os alunos de uma turma, diz:

a) Quantos alunos tem a turma?

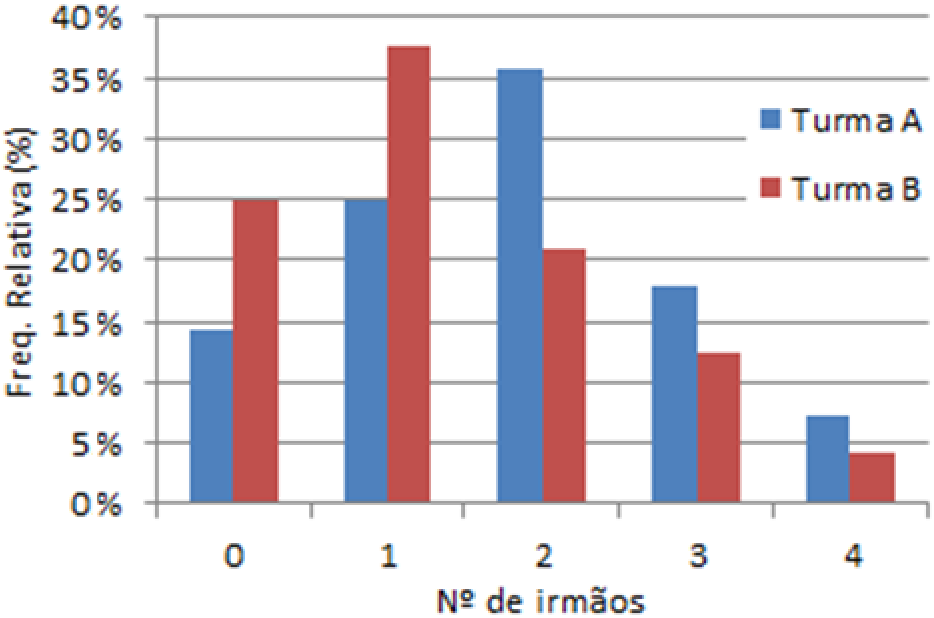
b) Nessa turma quantos alunos têm irmãos?

c) Algum aluno tem 4 irmãos?

d) Nessa turma predominam os alunos de famílias de quantos filhos?

2.2 Dados quantitativos discretos **Gráfico de barras**

O gráfico de barras pode ser utilizado para comparar as distribuições de 2 conjuntos de dados referentes à mesma variável. Por exemplo, na escola, perguntou-se aos alunos de 2 turmas, quantos irmãos tinham, tendo-se obtido os resultados apresentados no seguinte diagrama de barras:



Como se verifica a partir do gráfico ao lado, pode-se dizer que, de um modo geral, os alunos da turma A têm mais irmãos que os alunos da turma B.

2 Tabela de frequências. Representações gráficas
2.2 Dados quantitativos discretos **Gráfico de barras**

Exemplo - Alguns dados sobre o agregado familiar (Graça Martins, M.E. et al, OTD) - Num inquérito realizado na escola, perguntou-se aos 26 alunos de uma turma do 6.º ano:

1. Qual a dimensão do seu agregado familiar (quantas pessoas viviam em casa)?
2. Quantos são crianças?
3. Quantos aparelhos de televisão têm em casa?
4. 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 questões 1. e 2.:

Gráfico A

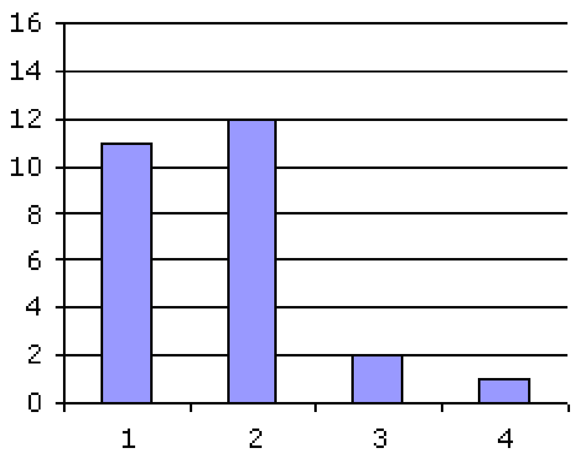
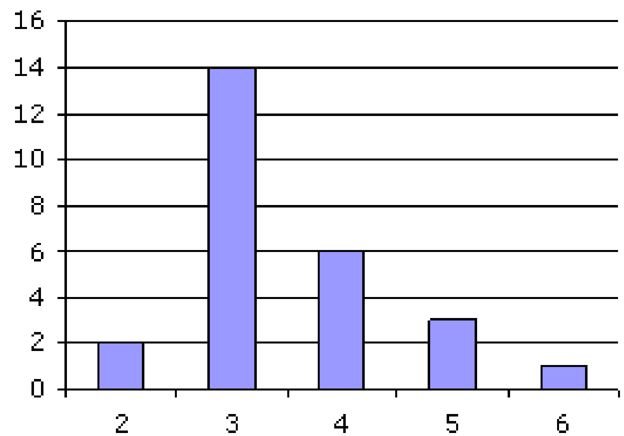


Gráfico B



a) 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?

b) Quantos agregados familiares têm 2 pessoas? Essas duas pessoas podem ser ambas adultas?

c) Dos 26 agregados familiares, 14 são constituídos por quantas pessoas?

d) Quantas pessoas tem o maior agregado familiar?

e) Quantas pessoas têm os 26 agregados familiares?

2.2 Dados quantitativos discretos **Gráfico de barras**

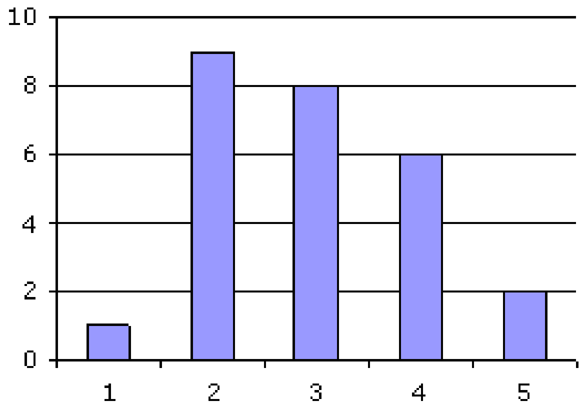
Considera agora também o outro gráfico que representa o número de crianças por agregado familiar.

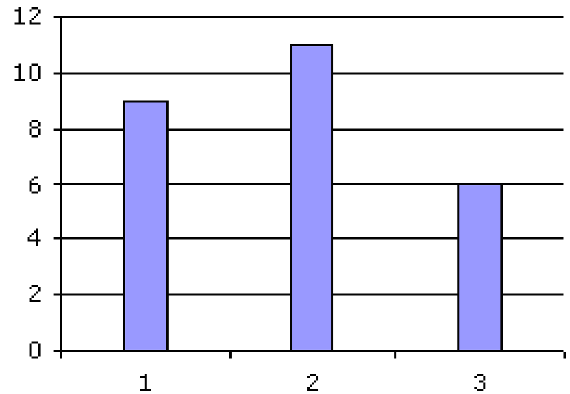
a) Quantos agregados familiares têm 1 criança?

b) Qual o número de crianças que predomina nos agregados familiares?

c) 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, a 3. e a 4., construíram-se os dois gráficos seguintes





a) Comparando os dois gráficos, qual dos dois achas mais razoável para representar o *Número de televisões por agregado familiar*?

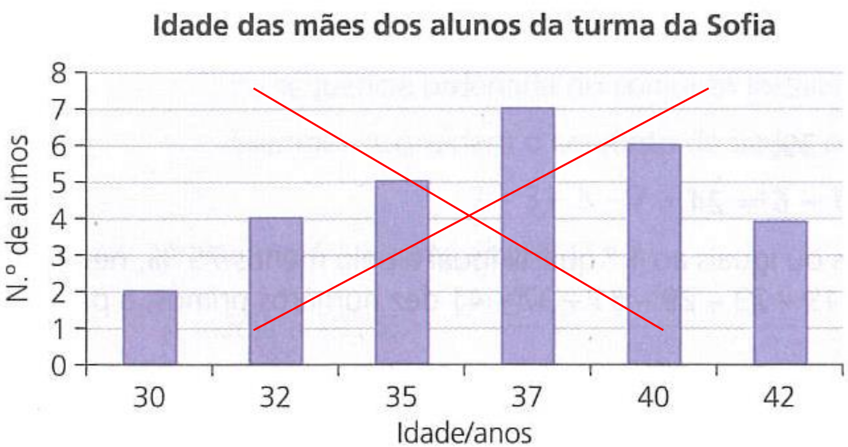
b) No gráfico do lado esquerdo a classe 2 tem frequência absoluta igual a 9 e no gráfico da direita a classe 1 tem também frequência absoluta igual a 9. No entanto as alturas das barras são diferentes. Como explicas esta situação?

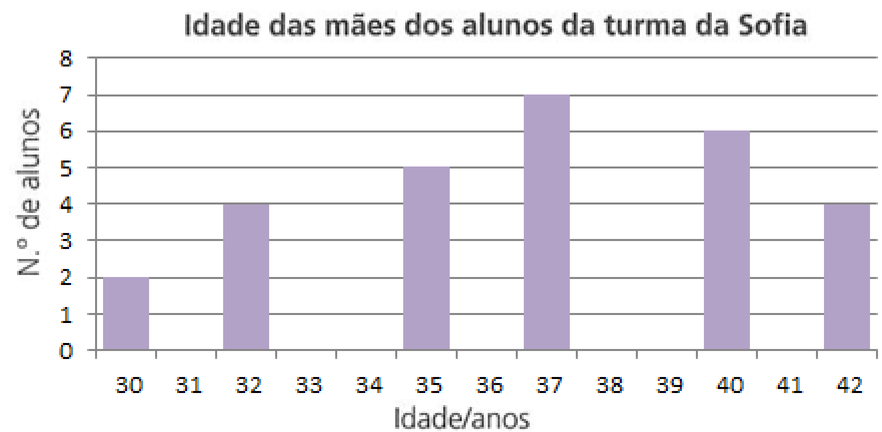
c) Completa os gráficos com as legendas adequadas.

2.2 Dados quantitativos discretos **Gráfico de barras**

Nota

Na construção de um gráfico de barras de dados discretos devem-se registar as classes igualmente espaçadas, não esquecendo classes intermédias em que a frequência absoluta é nula (ver observação). Veja-se o seguinte gráfico, em baixo à esquerda, retirado de um texto de Matemática do ensino básico. Embora a característica Idade de uma pessoa seja uma variável de tipo contínuo, pode-se utilizar discretizada, na forma de números inteiros. No entanto a representação não está correcta, pois não dá a ideia do que se pretende com uma representação gráfica, nomeadamente a existência de simetria, de lacunas, de variabilidade, etc. O gráfico correcto é o que se apresenta do lado direito.





2.2 Dados quantitativos discretos **Caule-e-folhas**

Caule-e-folhas - É um tipo de representação que se pode considerar entre a tabela e o gráfico. É com os próprios dados que se vai construindo a representação gráfica. Cada dado é separado em duas partes: o "caule" e a "folha".

Exemplo – Tempo sem respirar - 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, 27, 35, 46, 60

Vamos considerar como caules os algarismos das dezenas e como folhas os algarismos das unidades.

Começa por se desenhar um traço vertical e do lado esquerdo desse traço colocam-se os caules, tendo a preocupação de considerar todos os caules entre o mínimo e o máximo (veremos mais à frente que pode haver situações em que haja dados "discrepantes" considerados "outliers", que inviabilizam a representação de todos os caules). (continua)

2.2 Dados quantitativos discretos **Caule-e-folhas**

59, 38, 47, 23, 48, 55, 37, 48, 53, 37, 52, 39, 54, 57, 38, 46, 40, 41, 62, 63, 38, 65, 44, 27, 35, 46, 60

De seguida começa-se a percorrer o conjunto dos dados e vão-se colocando as folhas junto dos caules respectivos. No fim é costume ordenar as folhas penduradas em cada caule, embora não seja necessário para visualizar o aspecto da distribuição.

1º passo

2	
3	
4	
5	9
6	

2º passo

2	
3	8
4	
5	9
6	

3º passo

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

4º passo

2|3 lê-se 23

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

(Continua)

2.2 Dados quantitativos discretos **Caule-e-folhas**

59, 38, 47, 23, 48, 55, 37, 48, 53, 37, 52, 39, 54, 57, 38, 46, 40, 41, 62, 63, 38, 65, 44, 27, 35, 46, 60

Suponha que aos dados anteriores se juntou o tempo que o professor conseguiu estar sem respirar que foi de 93 segundos. Incluir esse dado na representação de caule e folha.

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

2|3 lê-se 23

Da representação em caule-e-folhas verifica-se que existe um dado "discrepante" relativamente aos outros, que se distribuem de forma aproximadamente simétrica. O dado bastante maior que os restantes provoca um enviesa-mento para a direita na distribuição dos dados. (Continua)

2.2 Dados quantitativos discretos **Caule-e-folhas**

59, 38, 47, 23, 48, 55, 37, 48, 53, 37, 52, 39, 54, 57, 38, 46, 40, 41, 62, 63, 38, 65, 44, 27, 35, 46, 60

Suponha agora que aos dados anteriores se juntou o tempo que um mergulhador profissional conseguiu estar sem respirar que foi de 600 segundos. Incluir esse dado na representação de caule e folha.

2|3 lê-se 23

2	3	7						
3	5	7	7	8	8	8	9	
4	0	1	4	6	6	7	8	8
5	2	3	4	5	7	9		
6	0	2	3	5				
...								
60	0							

Este novo dado implica que se considere o caule 60 e a folha 0. No entanto não é viável considerar todos os caules até ao 60, já que o valor 600 é um valor muito maior que os restantes, sendo considerado um "outlier". Então um processo de representar estes valores é interromper o traço, como se indica ao lado.

Nota

Junto do caule-e-folhas deve colocar-se uma indicação de como ler o caule e a folha, para se saber a ordem de grandeza dos dados, se entretanto não tivermos presente os dados originais. Se os nossos dados fossem o resultado de observar uma variável contínua, por exemplo o tempo em horas de realização de uma dada tarefa por 27 pessoas, assumindo os valores 5,9 3,8 4,7 2,3 4,8 5,5 3,7 4,8 5,3 3,7 5,2 3,9 5,4 5,7 3,8 4,6 4, 0 4,1 6,2 6,3 3,8 6,5 4,4 2,7 3,5 4,6 6,0, a representação em caule-e-folhas seria igual à anterior, mas teríamos

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

2|3 lê-se 2,3

2.2 Dados quantitativos discretos **Caule-e-folhas**

Exemplo – Notas a Matemática - Considerem-se as seguintes notas obtidas por 62 alunos num teste de Matemática: 30, 41, 21, 23, 11, 5, 38, 38, 61, 70, 52, 58, 42, 35, 35, 15, 21, 48, 79, 69, 69, 50, 59, 42, 64, 66, 66, 27, 78, 78, 68, 69, 44, 67, 73, 76, 76, 68, 52, 75, 77, 53, 69, 54, 45, 77, 46, 56, 59, 67, 76, 77, 68, 57, 50, 47, 58, 61, 65, 51, 59, 31

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

Do enviesamento apresentado pela representação em caule-e-folhas conclui-se que o teste de Matemática era bastante acessível ou que a maioria dos alunos tinha estudado bem a matéria, ou se verificaram as duas situações. A classe modal é o intervalo [60, 70 [.

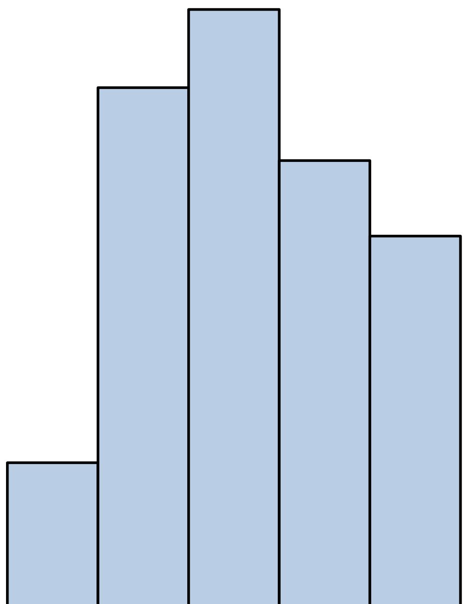
2.2 Dados quantitativos discretos **Caule-e-folhas**

Caule-e-folhas

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

2|3 lê-se 23

O caule e folhas sugere uma outra representação gráfica chamada histograma



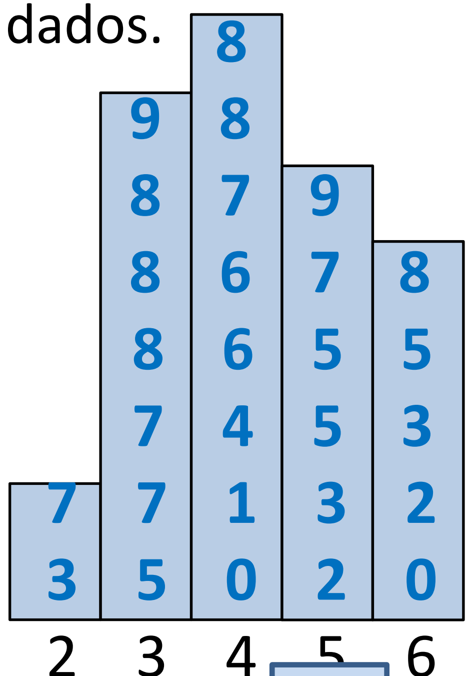
Classes	Freq.abs.
[20,30[	**2**
[30,40[	**7**
[40,50[	**8**
[50,60[	**6**
[60,70[	**5**

(Continua)

2.2 Dados quantitativos discretos **Caule-e-folhas e histograma**

O histograma é um tipo de representação usado, essencialmente, para dados quantitativos contínuos, em que se organizam os dados em classes, na forma de intervalos. Utiliza-se também para dados discretos quando houver muitos dados, com muitos valores distintos. É um **diagrama de áreas**, formado por uma sucessão de rectângulos adjacentes, tendo cada um por base um intervalo de classe e por área a frequência relativa ou absoluta dessa classe. Assim, a área total ocupada pelo histograma será igual a 1 ou igual ao nº de dados.

Quando as classes tiverem a mesma amplitude, pode-se considerar para altura dos rectângulos as frequências absolutas (como é o caso ao lado) ou as relativas.



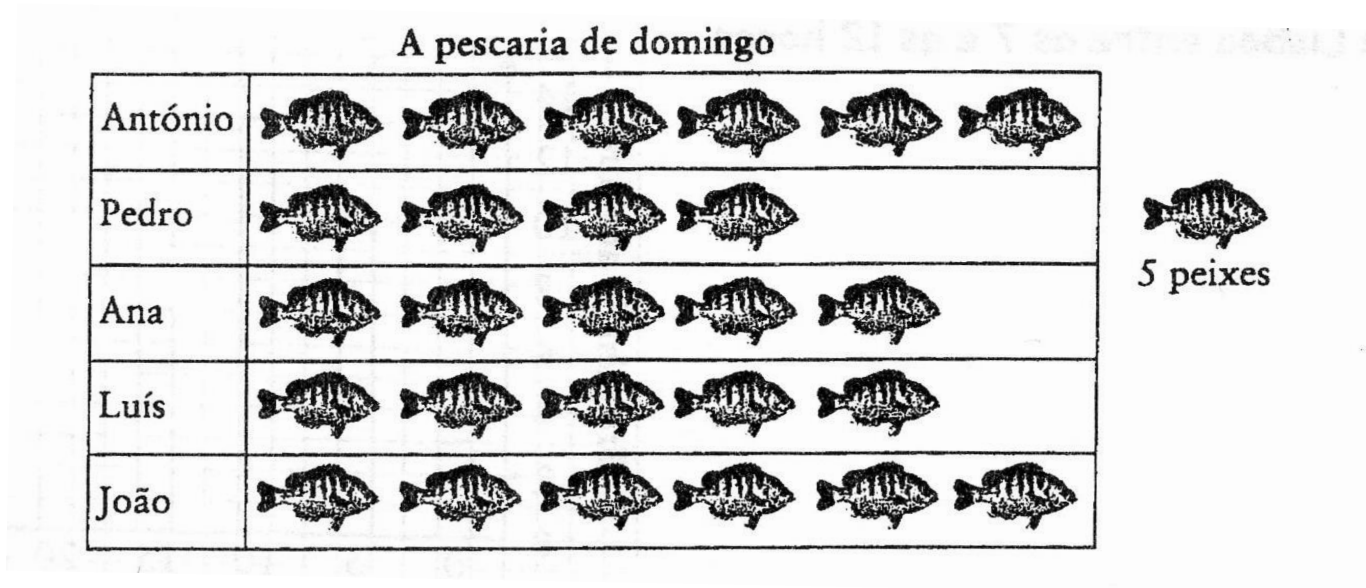
Classes	Freq.abs.
[20,30[	**2**
[30,40[	**7**
[40,50[	**8**
[50,60[	**6**
[60,70[	**5**

Observação

Em alguns manuais de Matemática encontram-se, com alguma frequência, exemplos em que se faz confusão entre dado (discreto) e frequência absoluta. Por esta razão achamos conveniente fazer este alerta com alguns exemplos encontrados nesses manuais.

Exemplo 1 (Retirado de um manual de Matemática)

"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*."

Comentário - Na figura anterior, *aparentemente* idêntica ao pictograma do exemplo da ficha 115, o que é o **dado**? O **dado** é o resultado da observação do "número de peixes" que cada um dos 5 amigos apanhou, ou seja, o conjunto de dados observados é

30, 20, 25, 25, 30

Assim, ao contrário do exemplo referido anteriormente, o que a figura aqui está a representar são os **dados** e não as **frequências absolutas** desses **dados**. Temos aqui dados quantitativos discretos. Poderíamos, a partir do pictograma construir a seguinte tabela:

Nome	Nº de peixes pescados
António	30
Pedro	20
Ana	25
Luís	25
João	30

Atenção

A tabela ao lado **não é uma tabela de frequências**, já que é uma simples listagem com os dados observados. Não se observou o António 30 vezes, etc.

O que é que se pretendia quando se pedia para classificar a representação quanto à **moda**? **Está-se a confundir**

dado com frequência

e

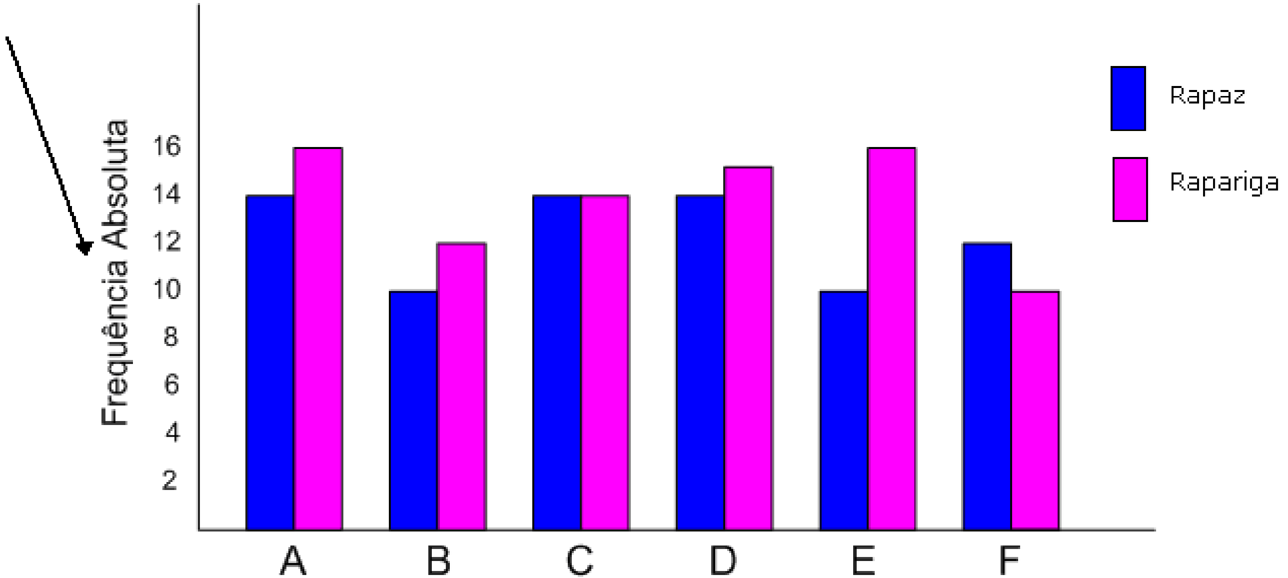
valor máximo com dado mais frequente.

Resumindo, no exemplo apresentado, tem-se:

- **Unidade estatística** – cada um dos 5 amigos
- **Variável** a estudar – nº de peixes pescados no Domingo por cada um dos 5 amigos
- **Dado** – resuldado da observação. Por exemplo 20, 25, etc. Estes dados são quantitativos discretos.
- **Frequência absoluta** de cada classe – o nº de vezes que cada dado surge no conjunto dos dados. A frequência absoluta do 20 é 1 e a do 25 e 30 é 2.

Exemplo 2 (Retirado de um manual de Matemática)

"Alunos das turmas do 5º Ano"



Comentário - Que leitura é que se pode fazer deste gráfico? **Nenhuma interes-sante! Algo está errado!**

Se no eixo das abcissas estão as categorias A, B, C, D e E e no eixo das ordenadas está a indicação de frequência absoluta, aparentemente significaria que as categorias A, B, ... têm as frequências indicadas. Mas 2 barras para cada categoria? E a barra azul a significar Rapaz e a rosa a significar Rapariga? E o título do gráfico a indicar "Alunos das turmas do 5º ano"?

O que está errado é a indicação de frequência absoluta no eixo das ordenadas. Num conjunto de dados, frequência absoluta de uma categoria é o número de vezes que essa categoria surge nesse conjunto de dados.

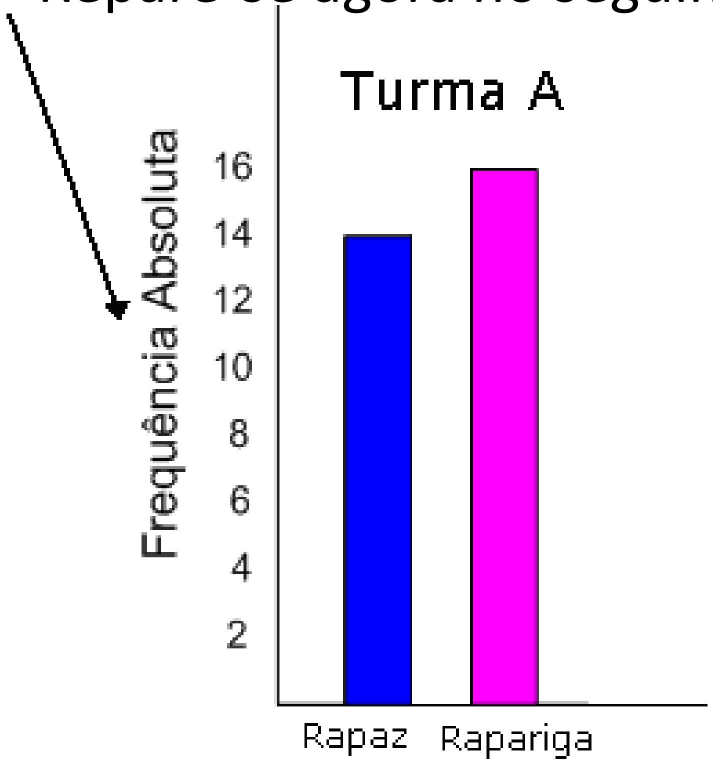
Neste caso o que é o dado? A unidade estatística é a **turma**, porque o nosso objectivo era saber quantos alunos tinha cada turma do 5º ano. O dado é o resultado da nossa observação! Assim, os nossos dados são o número de alunos das turmas A, B, ..., E e F ou seja 30, 22, ...22.

Resumindo:

- **Unidade estatística** (sobre a qual vai incidir o nosso estudo) – a turma
- **Variável a observar** – nº de alunos da turma
- **Dado** – resultado da observação do número de alunos da turma. Por exemplo 30, 22, ...etc. São dados quantitativos discretos.
- **Frequência absoluta** de uma classe – nº de vezes que essa classe surge no conjunto dos dados.

Assim, o gráfico anterior apesar de ser um gráfico **com** barras, **não** é um gráfico **de** barras, pois o que nos está a indicar é o valor dos dados e não as frequências com que esses dados surgem.

Repare-se agora no seguinte gráfico:



Neste gráfico, que é um **gráfico de barras**, está representada a informação que resultou de termos observado cada aluno da turma A, relativamente à característica sexo: *Rapaz* ou *Rapariga*. Temos uma variável de tipo qualitativo, que pode assumir duas categorias distintas. Relativamente aos alunos da turma A, verificámos que:

Categoria	Frequência absoluta
Rapaz	14
Rapariga	16

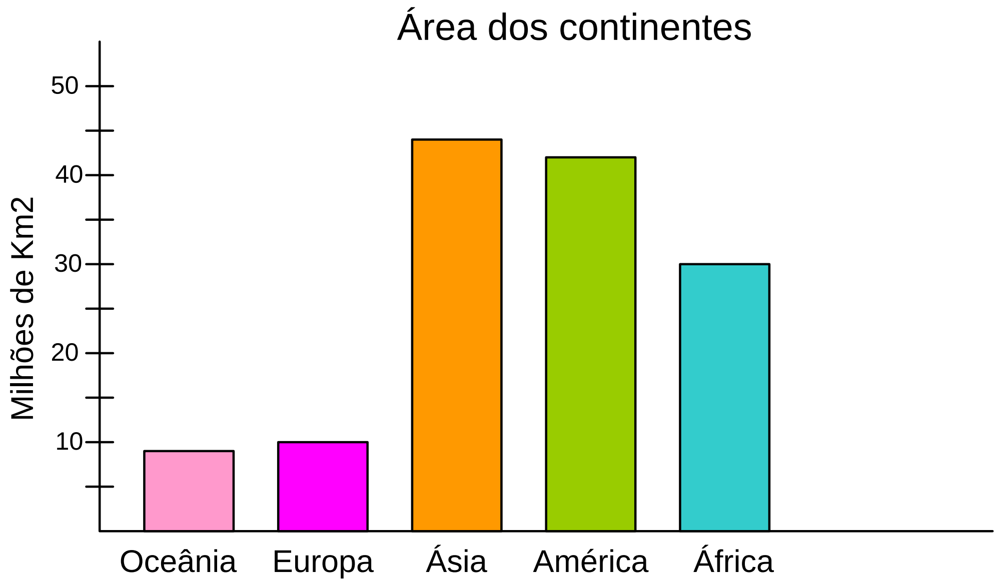
Resumindo:

- **Unidade estatística** – aluno da turma A
- **Variável** a estudar – sexo do aluno
- **Dado** – resultado da observação do sexo (rapaz ou rapariga)
- **Frequência absoluta** de cada categoria – nº de vezes que a categoria aparece no conjunto dos dados

Exemplo 3 (Retirado de um manual de Matemática)

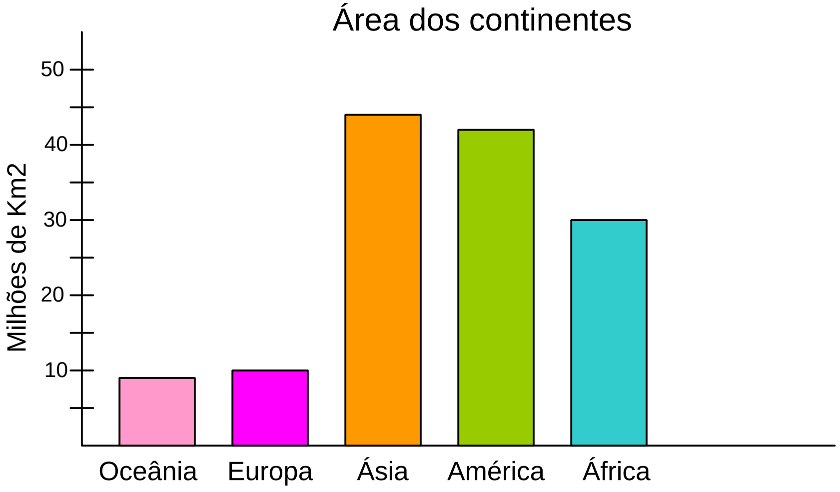
FREQUÊNCIAS ABSOLUTAS E TABELAS

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 absolutas.**

Comentário - A resposta apresentada no livro para a questão 4 em que se pede uma tabela de frequências absolutas é a seguinte



Continente	**Área milhões km²**
Oceânia	9
Europa	10
Ásia	44
América	42
África	30

Não é uma tabela de frequências!!

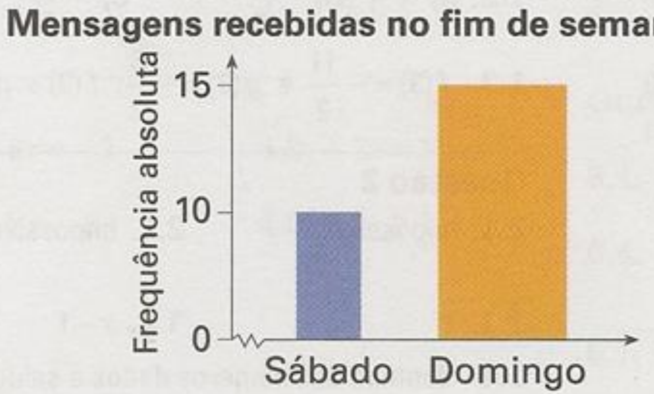
Como no exemplo 2, estão a confundir *dados* com *frequências*!!

Nem o gráfico é um gráfico de barras, nem a tabela é uma tabela de frequências! No entanto, tanto o gráfico como a tabela dão informação correcta na medida em que se limitem a representar os **dados**.

Exemplo 4 (Retirado de um manual de Matemática)

Gráfico de barras

A Ana construiu o seguinte gráfico:



Explica por que razão o gráfico está mal construído.

Comentário

Quando se pede para explicar porque razão o gráfico está mal construído, o objectivo é que a resposta seja a seguinte: a barra referente a Domingo não pode ter uma largura superior à barra referente a Sábado, uma vez que transmite uma ideia errada, dando a entender que o número de mensagens recebidas no Domingo é muito superior ao número de mensagens recebidas no Sábado, já que a área da barra referente a Domingo é várias vezes superior à área da barra referente ao sábado ...(cont)

Exemplo 4 (Retirado de um manual de Matemática)

No entanto, ao contrário do que é indicado, **não** temos um gráfico de barras, nem os valores registados no eixo vertical são frequências absolutas. Se fossem frequências absolutas, a leitura do gráfico seria:

O Sábado tem uma frequência absoluta igual a 10 e o Domingo uma frequência absoluta igual a 15, o que não tem sentido, já que não estamos a estudar nenhuma variável cujas categorias sejam o Sábado e o Domingo. O que está representado no gráfico *com* barras são os dados!

Exemplo 5 (Retirado de um manual de Matemática)

O gráfico de barras da figura 1 mostra o número de automóveis da marca *X* vendidos em 2012.

Figura 1

5.1. Qual foi o número de automóveis vendidos no primeiro quadrimestre do ano?

5.2. Em que mês foram vendidos mais 50% dos automóveis do que em junho?

5.3. Determina a moda do número de automóveis vendidos.
Qual o significado do valor que encontraste?

Comentário – Mais uma vez temos um gráfico com barras e não um gráfico de barras. No gráfico estão representados os dados que são o número de carros vendidos nos vários meses do ano de 2012.

A resposta que é dada à questão 5.3, no livro de onde se retirou o exemplo, é Fev, fazendo-se confusão entre dado mais frequente e dado com maior valor.

Os dados em estudo são os seguintes: 75, 90, 49, 40, 35, **50**, 65, 80, 30, 45, **50** e 71. Neste conjunto de dados existem dois dados iguais e os outros são todos diferentes. Assim, a moda é 50.

 Propostas de exercícios

1. Número de páginas lidas (Adaptado do exame de 6º ano de 2014) A professora pediu a todos os alunos que frequentavam o Clube de Leitura que registassem o número de páginas lidas durante o fim de semana. Na lista seguinte apresentam-se os dados relativos ao número de páginas lidas por cada um desses alunos:

27	39	45	40	55	31	56	50	42	43
50	47	56	50	84	35	45	48	31	47

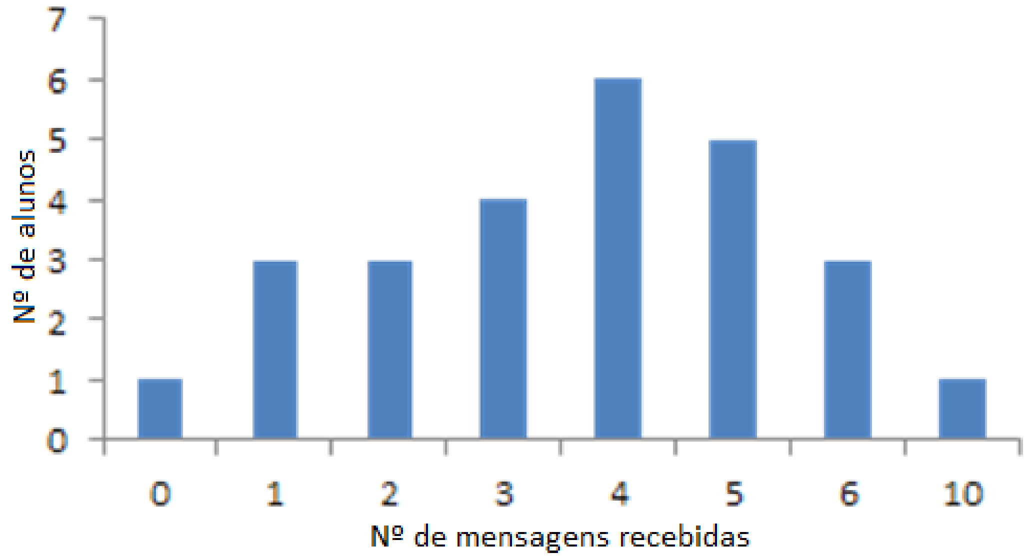
1. No diagrama de caule-e-folhas ao lado já estão colocadas as folhas do primeiro caule e do último caule. Completa o diagrama, tendo em conta o número de páginas lidas por cada um dos alunos. Apresenta as folhas por ordem crescente.

2	7
3	
4	
5	
6	
7	
8	4

2. Assinala com X a opção que indica a frequência relativa dos alunos que leram exatamente 50 páginas durante o fim de semana

☐ 5% ☐ 10% ☐ 15% ☐ 20%

2. Número de mensagens recebidas A professora propôs aos alunos realizarem um estudo sobre o número de mensagens recebidas, no telemóvel, ao fim de semana. Assim, pediu para cada aluno ir registar no quadro quantas mensagens tinha recebido no último fim de semana. Com os dados obtidos, um grupo de alunos apresentou o seguinte gráfico de barras:



A professora assim que viu o gráfico disse que tinha um erro, de modo que deveria ser corrigido. Os alunos que o fizeram não conseguiram perceber o que é que não estava correcto. És capaz de os ajudar a emendar o gráfico?

2.3 Dados quantitativos contínuos

- Tabela de frequências
- Histograma
- Caule-e-folhas

Enquanto que no caso dos dados discretos a construção da tabela de frequências é, de um modo geral, muito simples, já que normalmente se consideram como classes os valores distintos que surgem no conjunto dos dados, no caso de dados contínuos o processo de resumir os dados é um pouco mais elaborado, já que a definição das classes não é tão imediata. Efectivamente não tem sentido considerar, para classes, os diferentes valores que surgem no conjunto de dados, pois eventualmente eles são todos diferentes.

Tabela de frequências - Se os dados são quantitativos contínuos, organizam-se em classes na forma de intervalos. Sempre que possível estes intervalos devem ter a mesma amplitude. A informação é organizada, no mínimo, em 3 colunas: coluna das *classes* – onde se identificam as classes, na forma de intervalos, em que se subdividiu o conjunto de dados; coluna das *frequências absolutas* e coluna das *frequências relativas* (ou percentagens). A *tabela de frequências* pode ainda incluir mais 3 colunas: coluna do *representante da classe* – onde se indica o ponto médio de cada intervalo de classe (usualmente escolhe-se o ponto médio do intervalo para representante da classe); coluna da *frequências absolutas acumuladas* e coluna da *frequências relativas acumuladas* .

2.3 Dados quantitativos contínuos **Tabela de frequências**

Perante um conjunto de dados de dados de tipo contínuo, o que se pretende com a subdivisão em classes é tornar patente a forma como esses dados se distribuem. Em muitos casos o "bom senso" preside à escolha das classes (principalmente em conjuntos de dados que se distribuam de forma muito enviesada). No entanto, para dados que se distribuem de forma aproximadamente simétrica, é usual construir os intervalos, que constituem as classes, de igual comprimento (amplitude) e há uma regra relativamente simples para a determinação do número de classes a considerar. Chama-se *regra de Sturges*, e é uma regra que nos sugere um valor para o número de classes em que se pode organizar um conjunto de dados. Apresenta-se a seguir uma versão desta regra:

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

2.3 Dados quantitativos contínuos **Tabela de frequências**

Uma vez obtido o valor de k para o número de classes, considera-se como amplitude h, de cada intervalo, um valor arredondado por excesso daquele que se obtém dividindo a amplitude (máximo-mínimo) do conjunto de dados, pelo número k de classes, que devem ser disjuntas e tal que a união tem de conter todos os dados do conjunto de dados a organizar.

As classes devem ser construídas segundo a mesma metodologia: fechadas à esquerda e abertas à direita, ou viceversa.

Exemplo - Subdivisão em classes dos dados referentes à variável *Área* do ficheiro "Dados sobre casas".

Uma vez que temos 40 dados, o menor inteiro k tal que $2^k>40$ é 6. Vamos então subdividir o conjunto dos dados em 6 intervalos de igual amplitude h. Para escolher os intervalos temos de começar por calcular a amplitude do conjunto de dados. Verifica-se que a área mínima é 66,3 e a área máxima é 163,3. Assim, uma possibilidade para a escolha de h seria considerar um valor aproximado por excesso do quociente (163,3-66,3)/6=16,16666667...

2.3 Dados quantitativos contínuos **Tabela de frequências**

Em alternativa, também se pode escolher um intervalo com uma amplitude múltipla de 6 (de 64m^2 a 166m^2, por exemplo) o que conduz a um valor inteiro para h (h=17) e, consequentemente, a intervalos de classe cujos extremos são também números inteiros. Vamos optar por esta segunda hipótese, por ser a de mais fácil leitura. Antes de apresentar a tabela convém ainda estabelecer uma convenção quanto à inclusão ou não de cada extremo dos intervalos de classe. Assim, vamos convencionar que todos os intervalos são fechados à esquerda e abertos à direita, isto é, da forma [a, b[, onde o número que surge no extremo esquerdo (a) pertence ao intervalo, mas o número que surge no extremo direito (b) já não pertence.

2.3 Dados quantitativos contínuos **Tabela de frequências**

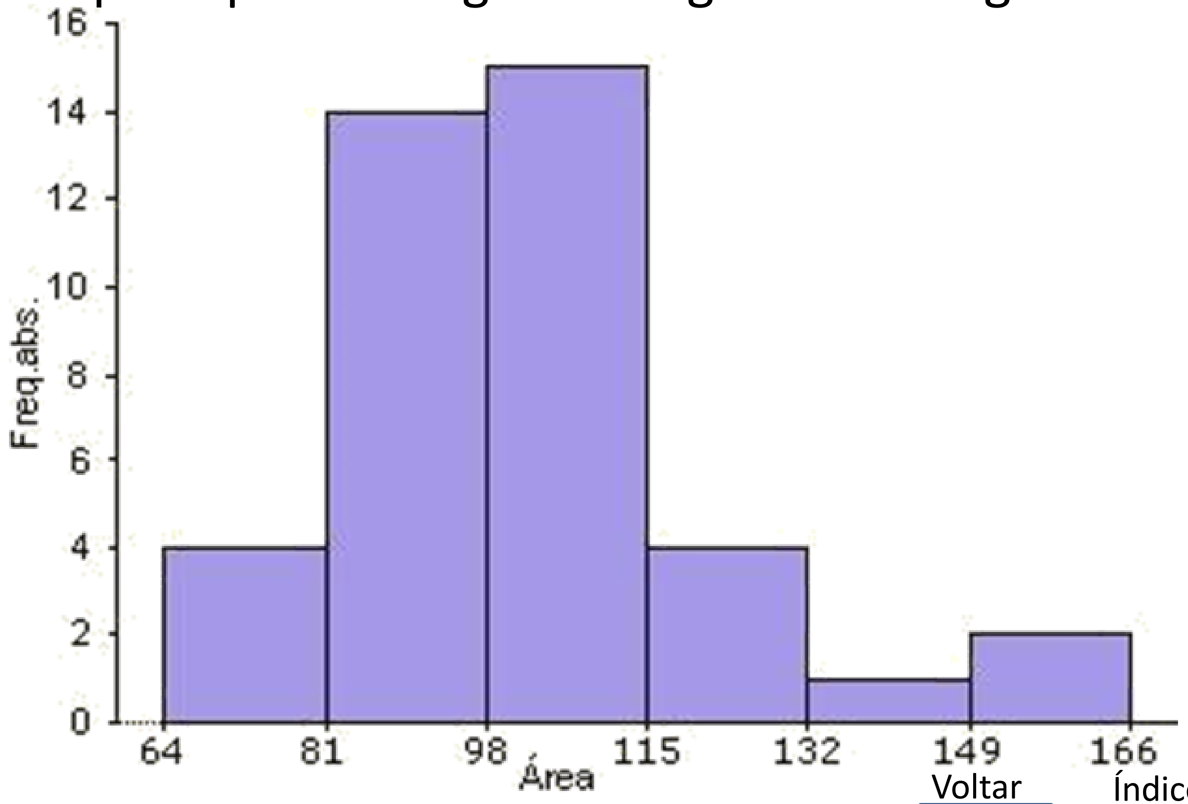
Tabela de frequências para os dados referentes à variável *Área*

Classes	Rep. classe x^*_i	Freq. Abs. n_i	Freq. Rel. f_i	Freq. Abs. Acum.	Freq. Rel. Acum.
[64, 81[	72,5	4	0,100	4	0,100
[81, 98[	89,5	14	0,350	18	0,450
[98, 115[	106,5	15	0,375	33	0,825
[115, 132[	123,5	4	0,100	37	0,925
[132, 149[	140,5	1	0,025	38	0,950
[149, 166[	157,5	2	0,050	40	1,000
Total		40	1,000		

2.3 Dados quantitativos contínuos **Histograma**

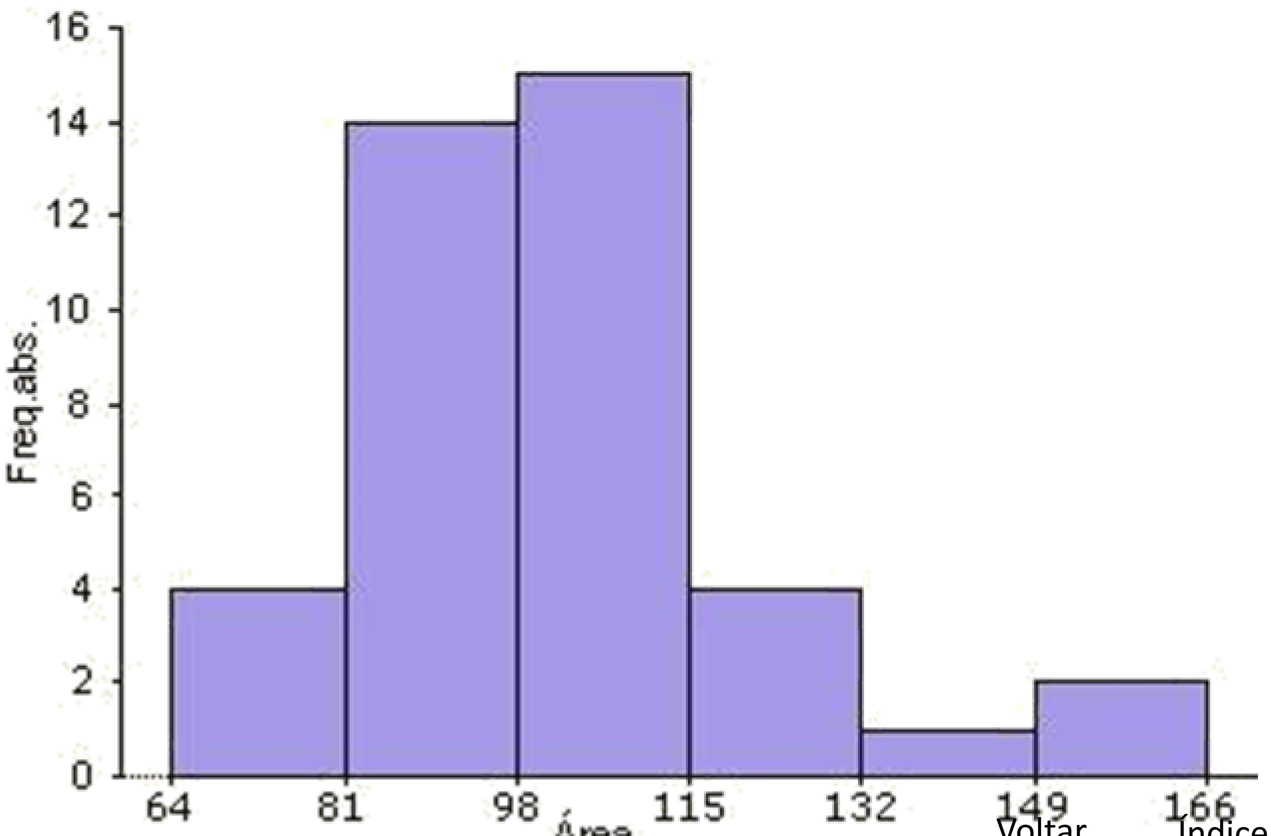
Histograma - O histograma é uma representação usada, essencialmente, para dados quantitativos contínuos. É um **diagrama de áreas**, formado por uma sucessão de rectângulos adjacentes, tendo cada um por base um intervalo de classe e por área a frequência relativa ou absoluta dessa classe. Assim, a área total ocupada pelo histograma é igual a 1 ou igual ao número de dados.

Quando as classes têm a mesma amplitude h pode considerar-se para altura dos rectângulos as frequências relativas ou absolutas (caso do histograma ao lado), pelo que a área total será igual a h ou hxn, onde n é o número de dados.



2.3 Dados quantitativos contínuos **Histograma**

Da análise do histograma dos dados da variável *Área* concluímos que predominam as casas com área aproximada entre 80 e 115 metros quadrados. Verifica-se também que a distribuição da área das casas apresenta um pequeno enviesamento para a direita, resultado da existência de duas casas com área um pouco superior às restantes.



Nota

Quando os dados em estudo são dicretos, mas o conjunto de dados for bastante numeroso e houver muitos dados distintos, torna-se impraticável a sua representação num gráfico de barras (já que para a construção da tabela de frequências se consideram tantas classes quantos os valores distintos existentes no conjunto de dados...), pelo que se pode utilizar o histograma para os representar.

2.3 Dados quantitativos contínuos **Caule-e-folhas**

Caule-e-folhas – Constrói-se de forma idêntica à que se considerou para os dados discretos. Vamos exemplificar com os dados da variável *Área* do ficheiro "Dados sobre casas". Para construir o caule-e-folhas vamos desprezar a parte decimal dos dados e considerar como caules os algarimos das dezenas e centenas e como folhas os algarismos das unidades. Seguinto esta metodologia obtém-se o seguinte diagrama:

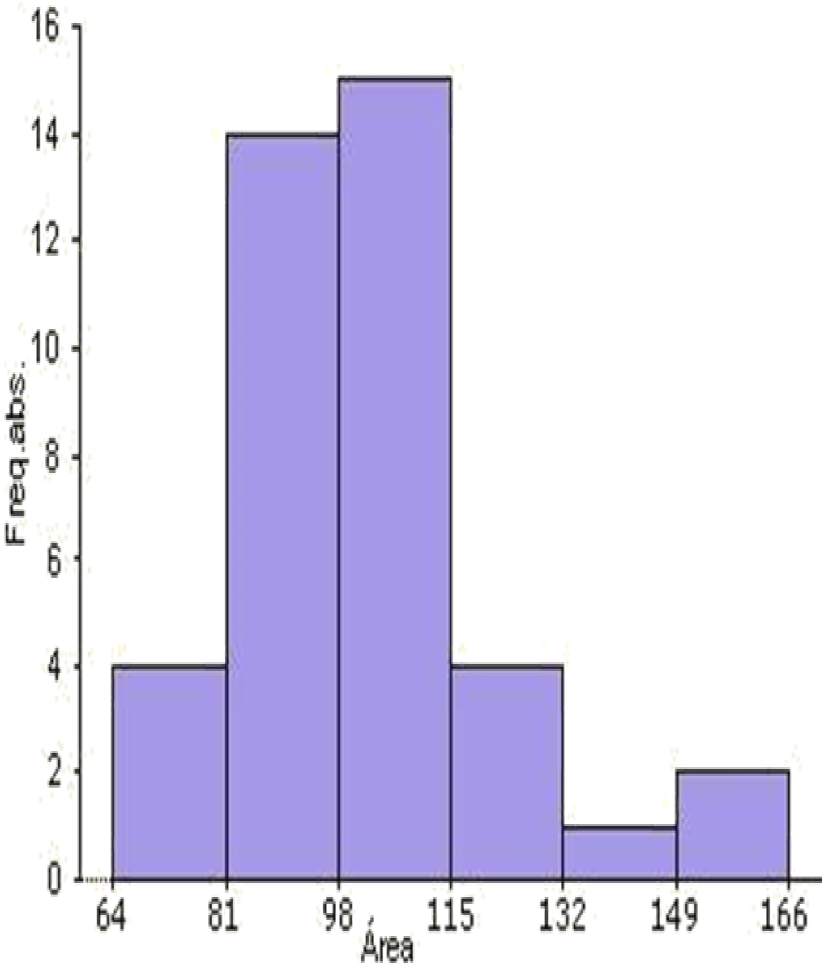
6	6
7	5 6 9
8	7 7 8 8
9	0 0 1 2 3 4 4 4 5 6 8 8 9
10	0 0 1 3 3 4 4 4 9 9
11	2 4 8 8 9
12	6
13	8
14	
15	4
16	3

6|6 lê-se 66

2.3 Dados quantitativos contínuos **Caule-e-folhas**

Repare-se que a informação transmitida pelo histograma é semelhante à que é transmitida pelo caule-e-folhas

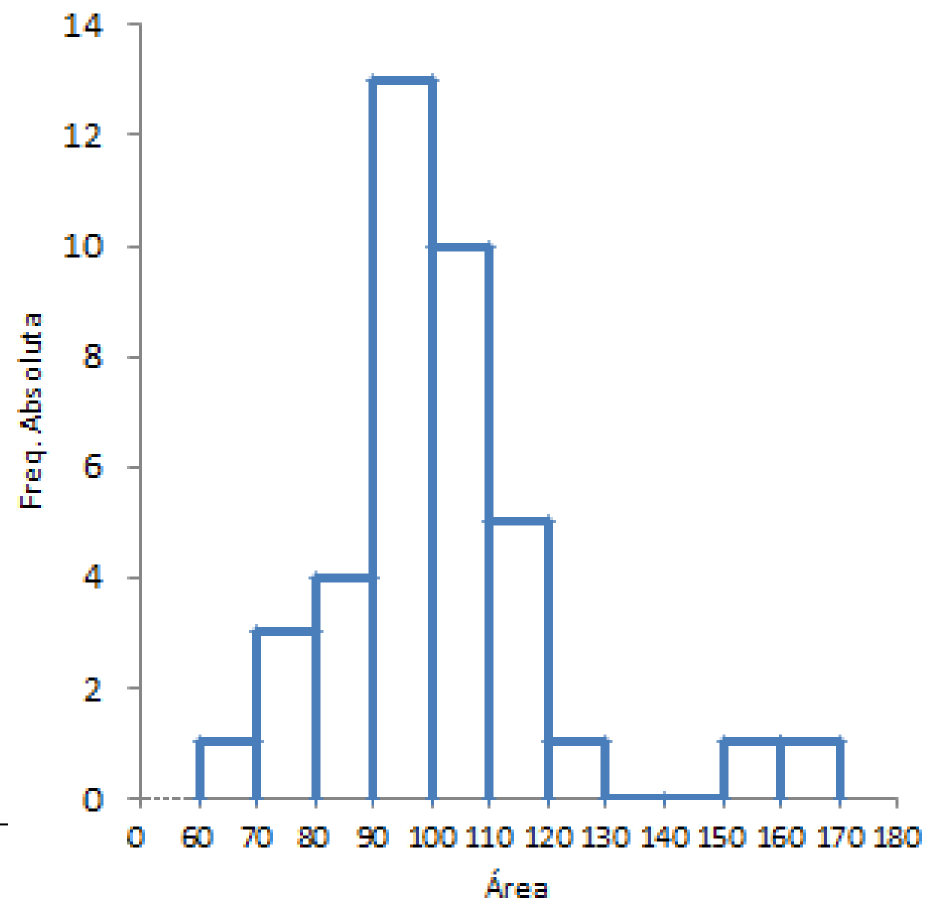
			9							
			8							
			8							
			6	9						
			5	9						
			4	4						
			4	4						
			4	4						
			3	3	9					
		8	2	3	8					
	9	8	1	1	8					
	6	7	0	0	4					
6	5	7	0	0	2	6	8		4	3
6	7	8	9	10	11	12	13	14	15	16



2.3 Dados quantitativos contínuos **Caule-e-folhas**

As duas imagens seriam idênticas se as classes consideradas para construir o histograma fossem [60, 70 [, [70, 80[, [80, 90[, [90, 100[, [100, 110[, [110, 120[, [120, 130[, [130, 140[, [140, 150[, [150, 160[e [160, 170[

			9							
			8							
			8							
			6	9						
			5	9						
			4	4						
			4	4						
			4	4						
			3	3	9					
		8	2	3	8					
	9	8	1	1	8					
	6	7	0	0	4					
6	5	7	0	0	2	6	8		4	3
6	7	8	9	10	11	12	13	14	15	16



2.3 Dados quantitativos contínuos **Caule-e-folhas**

No caule-e-folhas dos dados da variável *Área* ainda se torna mais evidente o que já tínhamos visto no histograma – a existência de alguns dados um pouco superiores aos restantes, nomeadamente pelo facto de não existirem folhas penduradas no caule 14. Existem procedimentos que nos permitem verificar se alguns dados podem ser considerados “muito diferentes” dos restantes, normalmente chamados de “outliers”. Aqui, não entraremos em detalhe sobre esses procedimentos. Suponhamos, no entanto que aparecia uma casa com uma área de 300 metros quadrados. Como já vimos no caso dos dados discretos, uma metodologia possível para representar esse dado seria quebrar o traço vertical como se indica ao lado, já que é inviável estar a considerar todos os caules intermédios. A representação obtida permite concluir que há 2 casas cuja área é razoavelmente superior às restantes e 1 casa com área muito maior que as restantes.

6	6
7	5 6 9
8	7 7 8 8
9	0 0 1 2 3 4 4 4 5 6 8 8 9
10	0 0 1 3 3 4 4 4 9 9
11	2 4 8 8 9
12	6
13	8
14	
15	4
16	3
....	..
30	0

2.3 Dados quantitativos contínuos **Caule-e-folhas**

Exemplo – Taxa de fertilidade - Para vários países considerou-se a taxa de fertilidade, cujos valores se apresentam na tabela.

Áustria	1,4
Bélgica	1,8
Brasil	1,8
Cabo Verde	2,3
Canadá	1,6
Dinamarca	1,7
Finlândia	1,8
França	2
Alemanha	1,4
Grécia	1,3
Índia	2,5
Irlanda	2
Itália	1,4
Luxemburgo	1,6
México	2,2
Holanda	1,7
Noruega	1,9
Peru	2,4
Portugal	1,3
Espanha	1,3
Suécia	1,9
Suiça	1,5
Turquia	2,1
Reino Unido	1,9
Estados Unidos	1,9

Como se verifica, a representação em caule-e-folhas apresenta demasiadas folhas nos dois caules:

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

A solução é dividir cada caule em dois subcaules: no 1º subcaule, anotado com um *, colocam-se as folhas 0, 1, 2, 3 e 4 e no segundo subcaule, a que se junta um ., as folhas 5, 6, 7, 8 e 9.

```
1* | 4 4 3 4 3 3
1. | 8 8 6 7 8 6 7 9 9 5 9 9
2* | 3 0 0 2 4 1
2. | 5
```

1|4 lê-se 1,4

2.3 Dados quantitativos contínuos **Caule-e-folhas**

O **Caule-e-folhas** é uma construção adequada para comparar dois conjuntos de dados.

Exemplo – Tempos de sono - Apresentam-se, a seguir, os tempos de sono (em horas), medidos durante 30 noites seguidas, do Pedro e do David

	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

Caule-e-folhas

Pedro					David								
			7	4.									
			3	5*									
9	9	8	6	5.									
	3	3	0	6*	2	2	2	4					
9	7	6	5	6.	6								
		4	3	7*	1	1	1	1	1	4	4		
	9	7	6	7.	5	5	5	6	6	7	8	9	9
				8*	1	2	3	4					
9	7	7	6	8.	5	6	7	8					
4	4	3	0	9*									
		9	6	9.	5								
		1	0	10*									

4|7 lê-se 4,7

Como se verifica na representação ao lado, conclui-se que o Pedro, ao contrário do David, tem um sono mais irregular, apresentado uma grande variabilidade nos tempos de sono.

1. **Dossiê XVI – Censos 2011 - Tu também contas!** www.alea.pt

Durante 2011, no âmbito dos Censos 2011, o Gabinete dos Censos do INE, em colaboração com o ALEA, desenvolveu a iniciativa "Os Censos vão às Escolas". Esta iniciativa teve por objetivos:

- 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 2011;
- incentivar os alunos para ajudar os pais na resposta aos Censos 2011 pela Internet.

Um dos resultados mais importantes desta iniciativa é o do Mini-Censos. ... O principal objetivo dos Mini-Censos foi dar a conhecer aos alunos do 1º ciclo o que são, para que servem e como se fazem os Censos. ... A este inquérito responderam 199 escolas perfazendo um total de 23054 alunos do 1º ciclo, com respostas consideradas válidas. Pode aceder aos dados indo a

www.alea.pt
conjuntos de dados
Mini-Censos
Importação

Na ficha seguinte é apresentada uma tabela de frequências com os dados, medidos em cm, referentes à variável Altura:

1. Dossiê XVI – Censos 2011 - Tu também contas! (continuação)

Tabela de frequências para a Altura

Classe	Frequência absoluta	Frequência relativa %
[80-90[	14	0,06
[90-100[	61	0,26
[100-110[	266	1,15
[110-120[	1912	8,29
[120-130[	6826	29,61
[130-140[	8436	36,59
[140-150[	4392	19,05
[150-160[	1062	4,61
[160-170[	82	0,36
[170-180[	3	0,01
Total	23054	100,00

A partir da tabela ao lado responde às seguintes questões:

a) Em quantas classes foram agrupados os dados referentes à Altura?

b) Qual a classe que tem menor frequência?

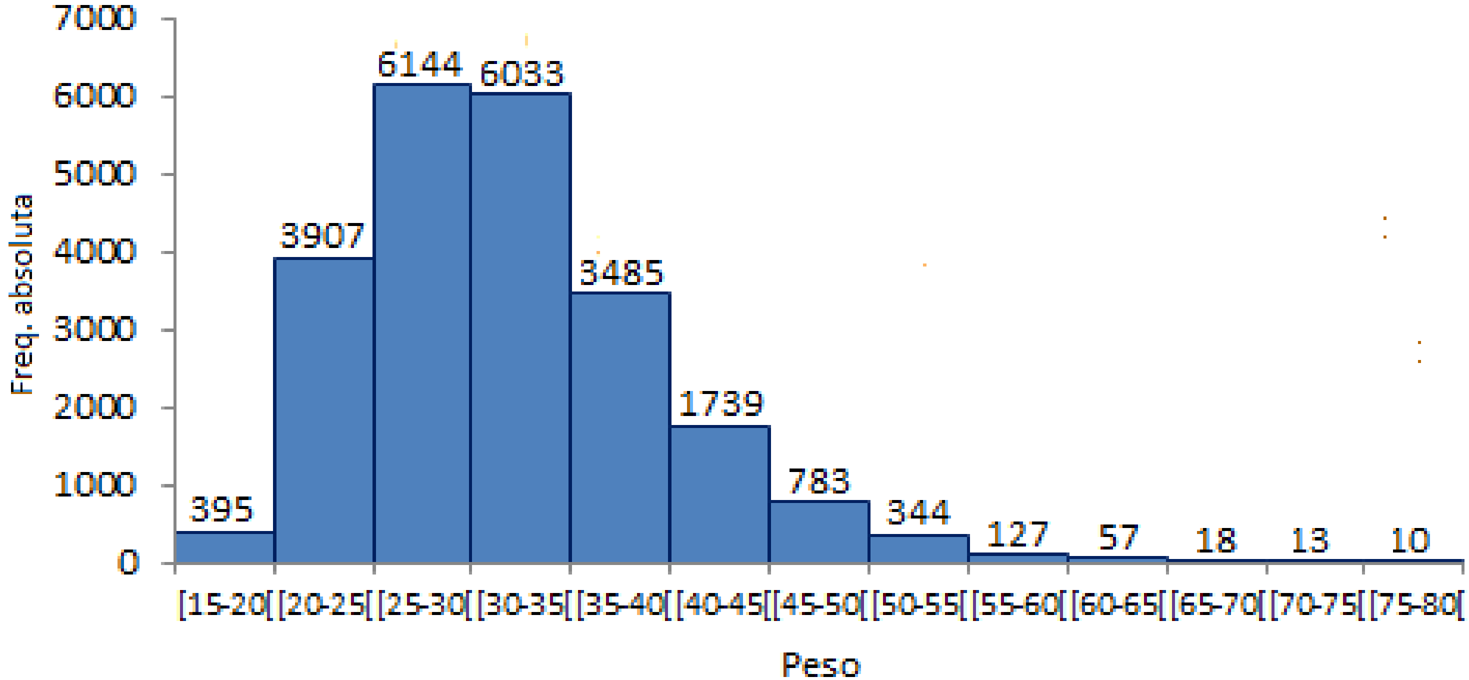
c) Qual a percentagem de alunos que mede menos de 140 cm?

d) Quantos alunos medem 160 cm ou mais?

e) Constrói o histograma adequado para representar os dados.

1. Dossiê XVI – Censos 2011 - Tu também contas! (continuação)

Os dados referentes à variável Peso (kg) encontram-se organizados no seguinte histograma:



A partir do histograma responde às seguintes questões:

a) Quantos alunos pesam menos de 30 kg?

b) Qual a percentagem de alunos que pesa 55 kg ou mais?

c) O histograma anterior realça um aspecto um pouco "preocupante" no que diz respeito ao peso dos alunos do 1º ciclo. Comenta esta afirmação.

2.4 Dados quantitativos bivariados

- **Gráfico ou Diagrama de dispersão**
- **Gráfico de linha** (caso particular do gráfico de dispersão)
- **Tabela de frequências**

2.4 Dados quantitativos bivariados

Por vezes os dados que se pretendem estudar aparecem sobre a forma de pares de valores, isto é cada unidade estatística da população contribui com dois valores. É o que acontece, por exemplo quando se considera para cada aluno a sua altura e o seu peso, ou a idade da mãe e a idade do pai, etc.

Dados bivariados - são o resultado da observação de duas variáveis sobre a mesma unidade estatística.

Consideremos a população constituída pelos alunos de uma escola. Se estivermos interessados em estudar as variáveis *altura* (em cm) e o *peso* (em kg) sobre a unidade estatística que é o aluno, o resultado de cada observação é um dado bivariado constituído por uma altura e um peso, por exemplo (120 cm, 23kg).

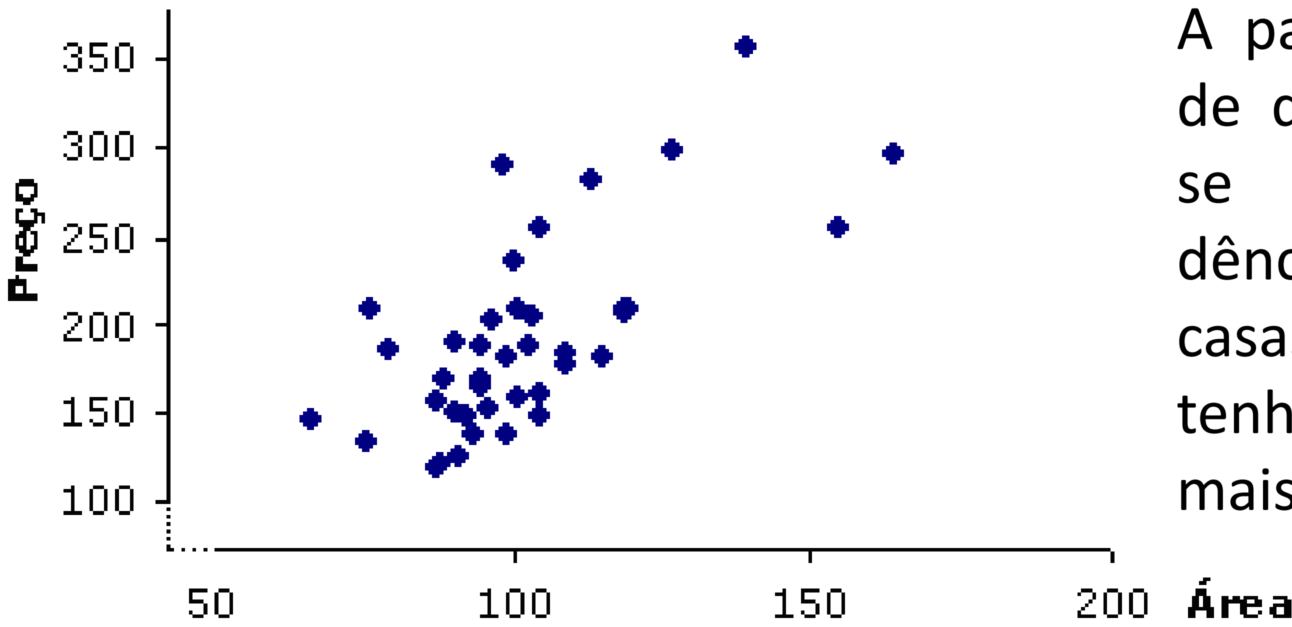
A representação gráfica utilizada para representar este tipo de dados é o diagrama ou gráfico de dispersão.

Gráfico ou **diagrama de dispersão** - é uma representação gráfica utilizada para representar dados resultantes de se observarem 2 variáveis quantitativas sobre a mesma unidade estatística. Cada par de dados (x_i,y_i) é representado, num sistema de eixos ortogonais, por um ponto de coordenadas (x_i,y_i). Obtém-se assim uma nuvem de pontos que nos permite avaliar de imediato se há ou não uma relação entre as duas variáveis.

Do ficheiro Dados sobre casas temos para as variáveis (*Área*, *Preço*):

(continua)

Diagrama de dispersão



A partir do diagrama de dispersão verifica-se que existe tendência para que as casas com maior área tenham um preço mais elevado

2.4 Dados quantitativos bivariados Gráfico de linhas

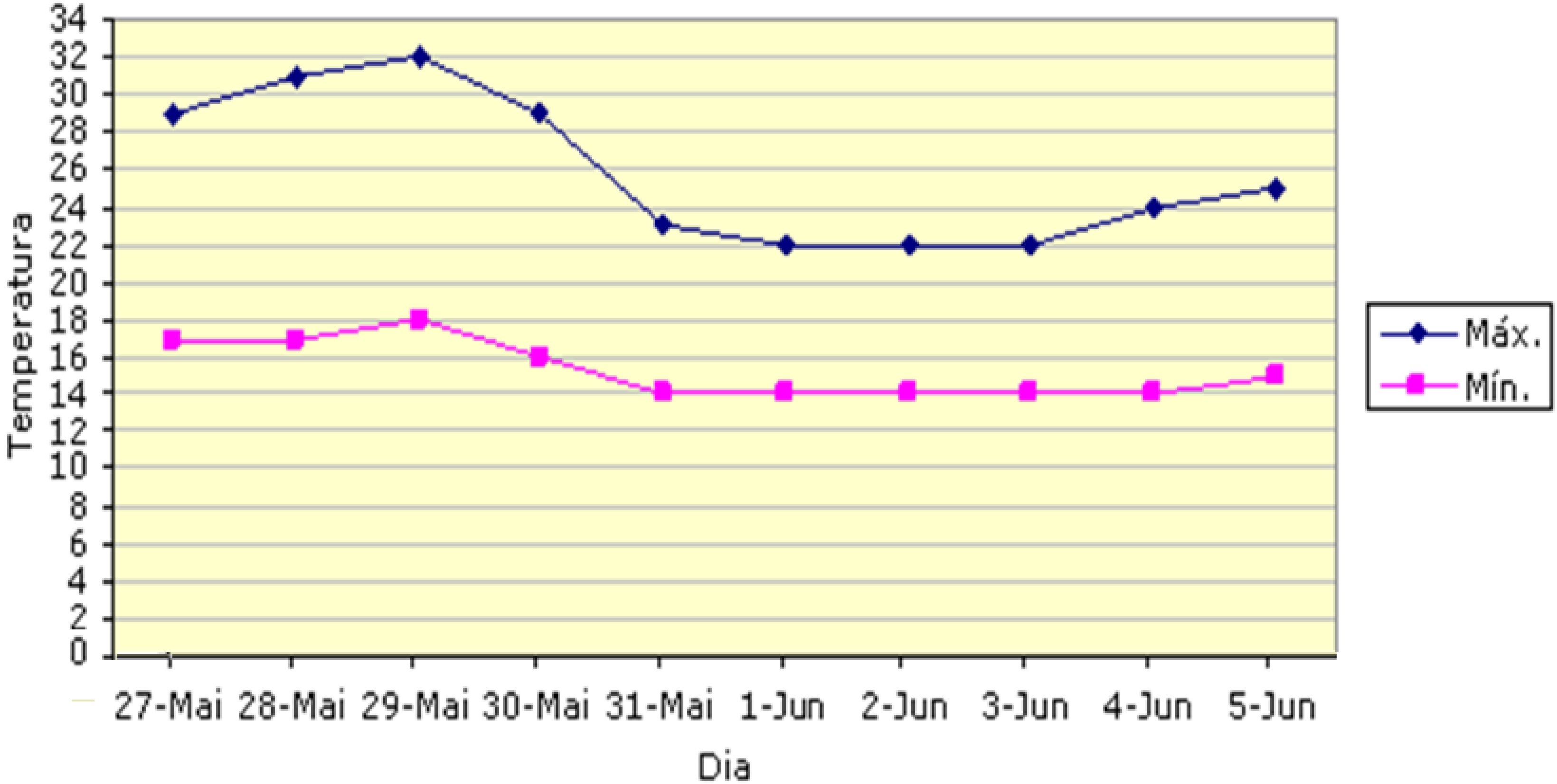
Gráfico de linhas - é um gráfico de dispersão especial, em que uma das variáveis (a variável que se coloca no eixo dos xx) é o tempo t. Uma vez representados os pontos (ti, yi), estes são unidos por linhas, dando origem ao gráfico de linhas.

Exemplo – Temperatura em Lisboa – Durante 10 dias seguidos registou-se a temperatura na cidade de Lisboa:

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

Com os dados anteriores construíu-se um gráfico de linhas, onde se registou a evolução da temperatura máxima e mínima em Lisboa, entre os dias 27 Maio e 5 de Junho (continua)

Gráfico de linhas



A partir do gráfico de linhas verifica-se que a temperatura máxima foi atingida no dia 29 de maio. Verifica-se também que foi nesse dia que se verificou a maior amplitude térmica – diferença entre a temperatura máxima e a mínima.

Nota

Para os alunos mais novos não se preconiza a representação gráfica de dados bivariados, o que não impede que se trabalhe com situações muito simples de registo destes dados, em que uma das variáveis é o **tempo**. Pode-se começar por pedir aos alunos para fazerem a leitura de gráficos simples como os que se apresentam nas propostas de exercícios.

Para construir o gráfico de linhas que faz parte do programa do 5º ano do ensino básico, basta os conhecimentos que os alunos obtiveram em Geometria no 3º ano, no tópico **Localização e orientação no espaço,** onde é pressuposto "*Reconhecer, numa grelha quadriculada na qual cada linha "horizontal" e cada coluna "vertical" está identificada por um símbolo, que qualquer quadrícula pode ser localizada através de um par de coordenadas. Identificar quadrículas de uma grelha quadriculada através das respetivas coordenadas".* Estes conhecimentos são suficientes para construir os gráficos de linha.

Tabela de frequências - Um outro processo de organizar a informação correspondente a dados bivariados é utilizando tabelas de frequências de dupla entrada. Estas tabelas permitem organizar dados, quer de tipo qualitativo, quer de tipo quantitativo, quando são de tipo bivariado, isto é, podem ser classificados segundo dois critérios. O seu aspecto é o de uma tabela com linhas, correspondentes a um dos critérios, e colunas correspondente ao outro critério.

Exemplo – Relativamente ao ficheiro Dados sobre casas considerem-se os dados relativamente ao número de assoalhadas e à Zona. Vamos organizar os dados numa tabela de frequências segundo os dois critérios Número de assoalhadas e Zona, considerando as linhas da tabela para o critério Zona e as colunas para o critério Número de assoalhadas:

(Continua)

2.4 Dados quantitativos bivariados Tabela de frequências

Tabela de frequências dos dados (Zona, Nº assoalhadas)

Zona \ Nº ass	1	2	3	4	5	Total
A	3	9	6	0	1	19
B	0	5	8	2	1	16
C	0	3	2	0	0	5
Total	3	17	16	2	2	40

Da tabela anterior pode-se concluir, por exemplo:

a) Predominam as casas da zona A, num total de 19;

b) Predominam as casas com 2 assoalhadas, num total de 17;

c) Predominam as casas da zona A e com 2 assoalhadas, num total de 9, seguindo-se as casas da zona B com 3 assoalhadas, num total de 8;

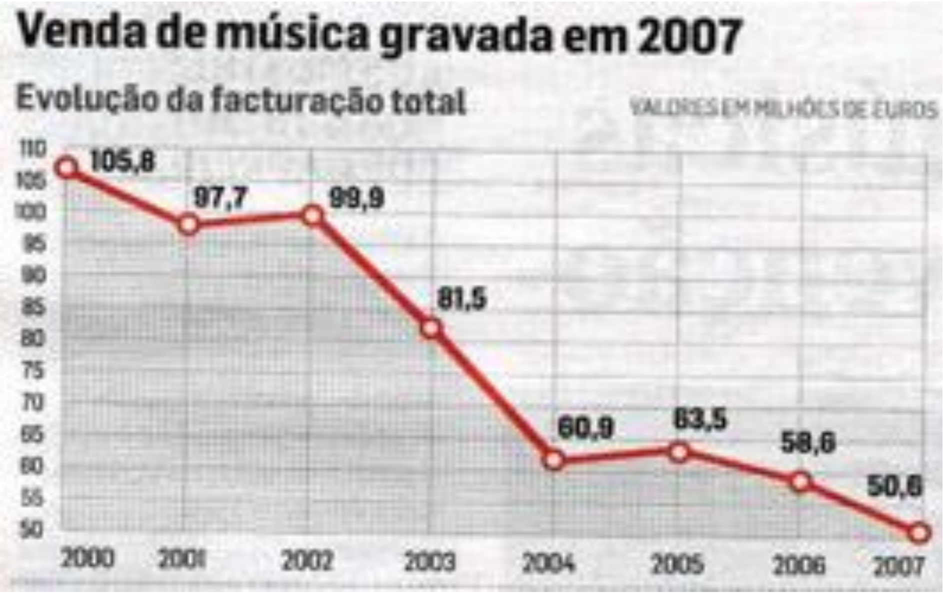
d) Na zona B não há casas com 1 assoalhadas;

e) Na zona C não há casas com 1, 4 ou 5 assoalhadas;

f) Só há casas com 4 assoalhadas na zona B.

1. (www.alea.pt – **Desafio 23**)

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, responde às seguintes questões

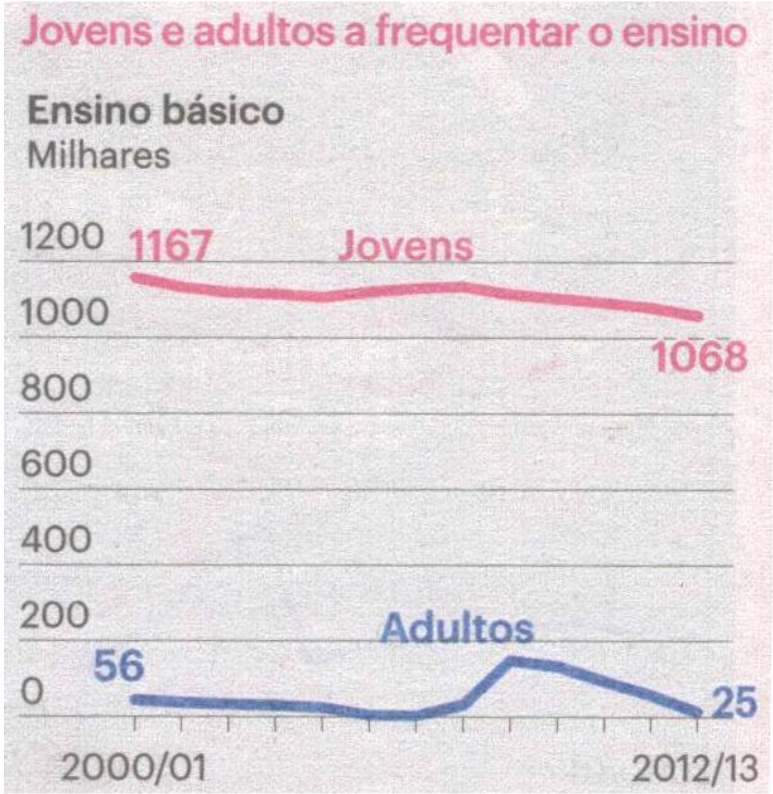


Questão 1: Embora a tendência da evolução da facturação seja nitidamente decrescente, houve alguns anos em que se verificou um ligeiro crescimento. **Entre que anos consecutivos** se registou esse crescimento na venda de música gravada? Qual o **valor do crescimento**, em **percentagem**? Apresenta o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

2. (www.alea.pt – Desafio 40) No gráfico é apresentada a evolução do número de alunos jovens e adultos a frequentar o ensino básico desde o ano letivo 2000/01 até 2012/13.



Questão 1.

Que nome se dá a uma representação gráfica deste tipo?

Questão 2.

Diz, justificando a tua escolha, se as afirmações seguintes são verdadeiras ou falsas:

(A) De 2000/01 a 2012/13, a evolução do número de alunos jovens a frequentar o ensino básico foi sempre decrescente.

(B) Em 2012/13 frequentavam o ensino básico cerca de um milhão e noventa e três mil alunos.

(C) O número de alunos adultos a frequentar o ensino básico em 2012/13 baixou mais de 50%, em relação ao ano letivo 2000/01.

(D) Em 2012/13, 2,34% dos alunos a frequentar o ensino básico eram adultos.

3. Considere-se de novo o dossiê XVI referido numa ficha anterior e a tabela de frequências seguinte, que representa os dados bivariados (Sexo, Idade):

Tabela de frequências dos dados (Sexo, Idade)

Sexo \ Idade	6	7	8	9	10	≥11	Total
Masculino	6,48%	10,74%	12,44%	13,79%	6,05%	1,11%	50,62%
Feminino	6,78%	10,81%	12,21%	13,53%	5,30%	0,75%	49,38%
Total	13,26%	21,55%	24,65%	27,33%	11,35%	1,87%	100,00%

Considerando a tabela anterior responde às seguintes questões:

a) De entre os alunos que responderam ao inquérito há mais alunos do sexo masculino ou feminino?
b) Pensas que no 1º ano de escolaridade há mais meninos ou meninas? Porquê?
c) Qual a percentagem de alunos com 9 ou mais anos?
d) A partir da resposta anterior, és capaz de sugerir qual o ano de escolaridade que predomina nos alunos que responderam ao inquérito?

3.1 Medidas de localização

- Média
- Mediana
- Moda

3.2 Medidas de dispersão

- Amplitude
- Amplitude interquartil

Medidas numéricas

Além das tabelas e dos gráficos utilizados para descrever os dados, para os dados quantitativos também se utilizam medidas numéricas, calculadas a partir dos dados. Destas medidas distinguiremos as **medidas de localização** que localizam alguns pontos importantes dos dados, nomeadamente o centro da distribuição dos dados, e as **medidas de dispersão**, que medem a variabilidade ou dispersão dos dados.

3.1 Medidas de localização

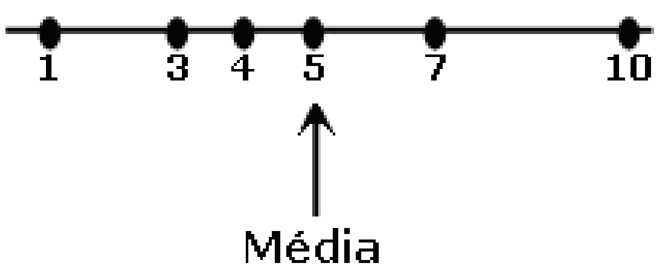
Uma medida de localização é um número que dá informação sobre alguns pontos da distribuição dos dados, tanto no que respeita na parte central, como nas caudas. De entre as medidas consideradas para localizarem o centro da distribuição dos dados destacam-se a **média** e a **mediana**. Outras medidas de localização que se referem a pontos nas caudas da distribuição, são o 1º e 3º **quartis**.

3.1 Medidas de localização Média

Média – é a medida de localização do centro da distribuição dos dados mais conhecida. Obtém-se fazendo a soma de todos os dados a dividir pelo número de dados.

A média é um número que "equilibra" os grandes valores com os pequenos valores (Graça Martins, M. E. et al, 2007, página 114).

A **média** dos valores (1,3,4,5,7,10) é **5**



$$\frac{1+3+4+5+7+10}{6} = 5$$

3.1 Medidas de localização Média

Exemplo – Número de letras do nome - Recolheu-se a informação sobre o número de letras do nome de 10 alunos (Graça Martins, M. E. et al, OTD – página 121)

Nome	Número de letras no nome
Ana Godinho	10
Ana Sofia Silva	13
Andreia Sousa	12
Carolina Martins	15
Daniela Silva	12
David Leal	9
Diogo Oliveira	12
Filipa Duarte	12
Helena Afonso	12
Inês Martins	11

$$\text{média} = \frac{10+13+12+\ldots+12+11}{10}$$

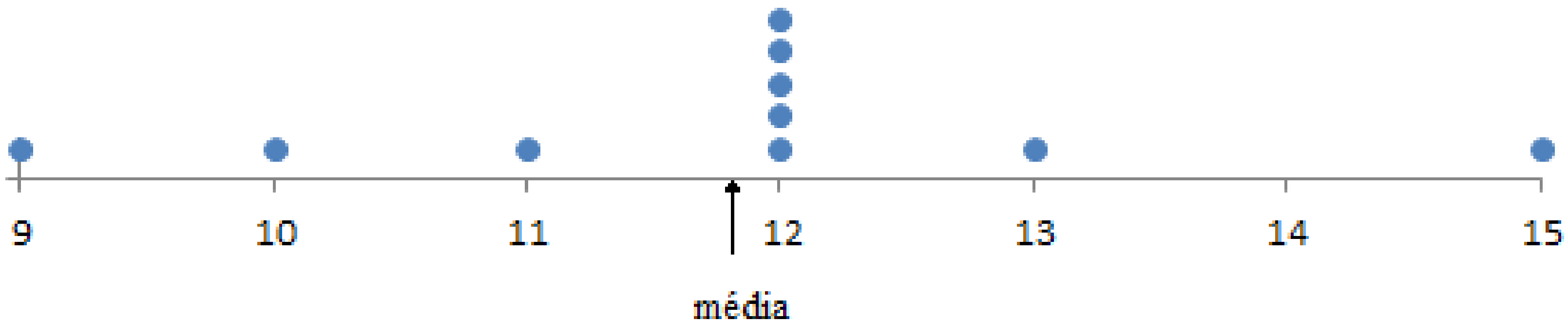
$$\text{média} = \frac{118}{10}$$

média=11,8 letras

3.1 Medidas de localização **Média**

A **média** do número de letras no nome é igual a 11,8.

O que significa? Consideremos os dados num diagrama de pontos



O que aquele valor significa é que 12 letras nos dá 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 quantia, 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.

Fórmula para o cálculo da média

Se representarmos um conjunto de n dados por $x_1, x_2, x_3, \ldots, x_n$, então a fórmula para o cálculo da média, que se representa com a notação $\bar{x}$, será:

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

No exemplo anterior, em que temos 10 dados tem-se, utilizando a notação anterior:

$$n=10 \text{ e } x_1=10,\ x_2=13,\ x_3=12,\ \ldots,\ x_{10}=11$$

Nota

Quando se representam os dados de um conjunto de n dados por $x_1, x_2, x_3, \ldots, x_n$, nesta notação não está implícita qualquer ordem de grandeza entre os dados. Podemos considerar que se representa por x_1 o primeiro dado a ser recolhido, por x_2 o segundo dado a ser recolhido, etc.

Exemplo – Alguns dados sobre os alunos da turma - Naquele dia o professor decidiu pedir a cada aluno alguns dados pessoais como o número de irmãos, a idade e a altura. Os 18 alunos foram organizados em alguns grupos para analisarem os dados recolhidos e 3 desses grupos foram encarregues de calcular, respectivamente, a média do número de irmãos, a média das idades e a média das alturas. Os resultados apresentados foram os seguintes:

Média do número de irmãos dos alunos =1,11

Média das idades dos alunos= 9,72 anos

Média das alturas dos alunos = 146,72 cm

Depois dos resultados anteriores apresentados o professor decidiu incluir os seus dados juntamente com os dados recolhidos dos alunos e colocou as seguintes questões: - Digo-vos que não tenho irmãos, tenho 45 anos e a minha altura é 169cm. Respondam às seguintes questões:

Exemplo – Alguns dados sobre os alunos da turma (continuação)

Juntando os dados referentes a mim:

1) Sem fazer qualquer conta, a média do número de irmãos baixa ou aumenta?
2) A média das idades pode vir superior a 11, mesmo sabendo que nenhum aluno tem idade superior a 11 anos? Calculem a nova média das idades.
3) Calculem a média das alturas dos alunos e incluam a minha.

Respostas:

1) A média baixa, pois juntou-se mais um dado igual a zero e a soma dos dados agora é a dividir por 19, em vez de ser a dividir por 18.
2) Pode vir superior a 11 pois juntou-se mais um dado que é muito maior que todos os outros. A média passou a ser igual a 11,58 anos. Este valor é razoavelmente superior à média das idades dos alunos porque a idade do professor é bastante superior às idades dos alunos.
3) Tendo em consideração as médias das alturas dos alunos, temos:

   Soma das alturas dos alunos =146,72cm$\times$18=2640,96cm

   Média das alturas= $\frac{2640{,}96+169}{19}$ = 147,89cm

Exemplo – Número de letras do nome - Suponha que à informação sobre o número de letras do nome dos 10 alunos considerados no exemplo da ficha, se juntaram mais 14 nomes, dos restantes alunos da turma de que se estava a estudar o número de letras do nome (Graça Martins, M. E. et al, OTD – página 121). Vamos admitir que os dados nos foram apresentados organizados numa tabela de frequências.

Como calcular a média para os dados agrupados?

Nome	Número de letras no nome
Ana Godinho	10
Ana Sofia Silva	13
Andreia Sousa	12
Carolina Martins	15
Daniela Silva	12
David Leal	9
Diogo Oliveira	12
Filipa Duarte	12
Helena Afonso	12
Inês Martins	11
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

Exemplo (continuação)

Uma vez que os dados estão agrupados, para calcular a média a soma de todos os dados é substituída pela soma dos produtos de cada um dos dados com valor numérico diferente pela sua frequência absoluta e divide-se o resultado obtido pelo número total de dados:

Nº de letras no nome	Freq. Abs.	Freq. Rel.
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

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

média= 12,8

A média do número de letras por nome é de 12,8 letras.

Fórmula para o cálculo da média para dados discretos agrupados

Se os dados se apresentarem agrupados e representarmos por

$$\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots, \tilde{x}_m$$

os m diferentes valores do conjunto de n dados, em que cada $\tilde{x}_i$ tem frequência absoluta n_i e frequência relativa f_i, com i=1, ..., m, então a fórmula para o cálculo da média será:

$$\bar{x} = \frac{n_1 \times \tilde{x}_1 + n_2 \times \tilde{x}_2 + n_3 \times \tilde{x}_3 + \ldots + n_m \times \tilde{x}_m}{n_1 + n_2 + n_3 + \ldots + n_m}$$

ou

$$\bar{x} = f_1 \times \tilde{x}_1 + f_2 \times \tilde{x}_2 + f_3 \times \tilde{x}_3 + \ldots + f_m \times \tilde{x}_m$$

Nota – Recorda-se que $n_1+n_2+n_3+\ldots+n_m=n$ e $f_i=\frac{n_i}{n}$

Fórmula para o cálculo da média para dados contínuos agrupados

Se os dados se apresentarem agrupados em k classes e representarmos por

$$x_1^*, x_2^*, x_3^*, \ldots, x_k^*$$

os representantes das k classes em que os n dados estão organizados, em que cada classe tem frequência absoluta n_i e frequência relativa f_i, com i=1, ..., k, então a fórmula para o cálculo de um valor aproximado (1) da média será:

$$\bar{x} \approx \frac{n_1 \times x_1^* + n_2 \times x_2^* + n_3 \times x_3^* + \ldots + n_k \times x_k^*}{n_1 + n_2 + n_3 + \ldots + n_k}$$

ou

$$\bar{x} \approx f_1 \times x_1^* + f_2 \times x_2^* + f_3 \times x_3^* + \ldots + f_k \times x_k^*$$

(1) Uma vez que os dados pertencentes a cada classe são substituídos por um representante da classe, obtém-se um valor aproximado para a média.

Pode-se sempre calcular a média?

A média só pode ser calculada para dados quantitativos!

Quando a natureza da variável em estudo é qualitativa, acontece, por vezes, atribuir códigos numéricos às diferentes categorias. O cálculo da média desses códigos não tem, obviamente, qualquer sentido. Por exemplo, no caso dos dados do ficheiro *Dados sobre casas*, não tem qualquer sentido calcular a média das observações respeitantes à variável qualitativa *Estado*, que assume as categorias usada e nova, representadas respectivamente por 0 e 1 (Graça Martins, M. E. et al, 2007, página 115).

3.1 Medidas de localização Média

Atenção à utilização da média

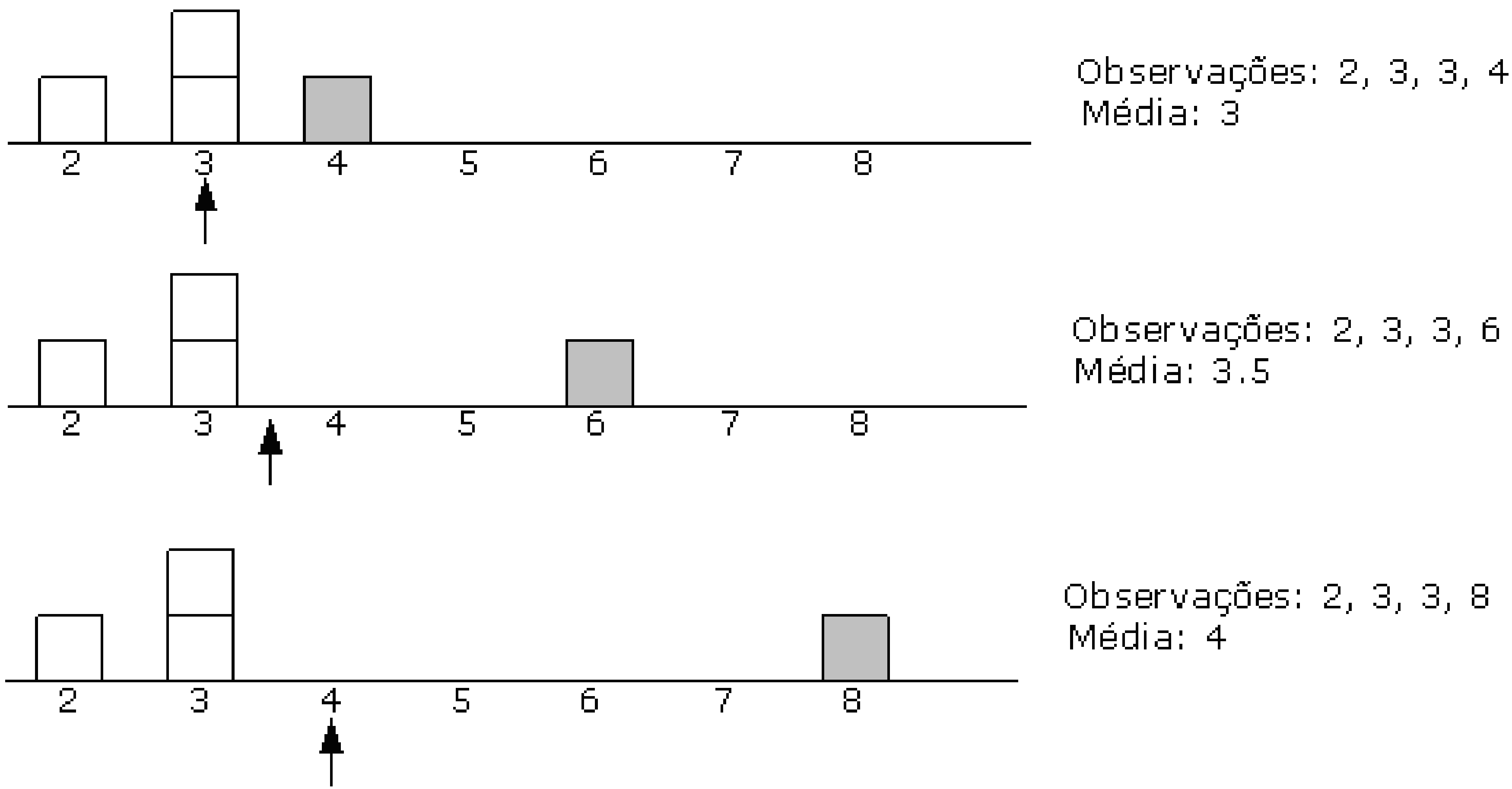
A média só é uma boa medida de localização do centro da distribuição dos dados se a distribuição for aproximadamente simétrica!!!

Caso contrário tem de se ter cuidado... Veja-se o que acontece no exemplo do slide se substituirmos o valor numérico do dado 10 por 25:

Se nos disserem que um conjunto de valores tem média 7,5, imaginamos que os valores se distribuem à volta do 7,5, aproximadamente metade de cada lado. Não pensamos num conjunto de valores em que todos, à excepção de um deles, são inferiores à média!

Veja-se ainda o slide seguinte.

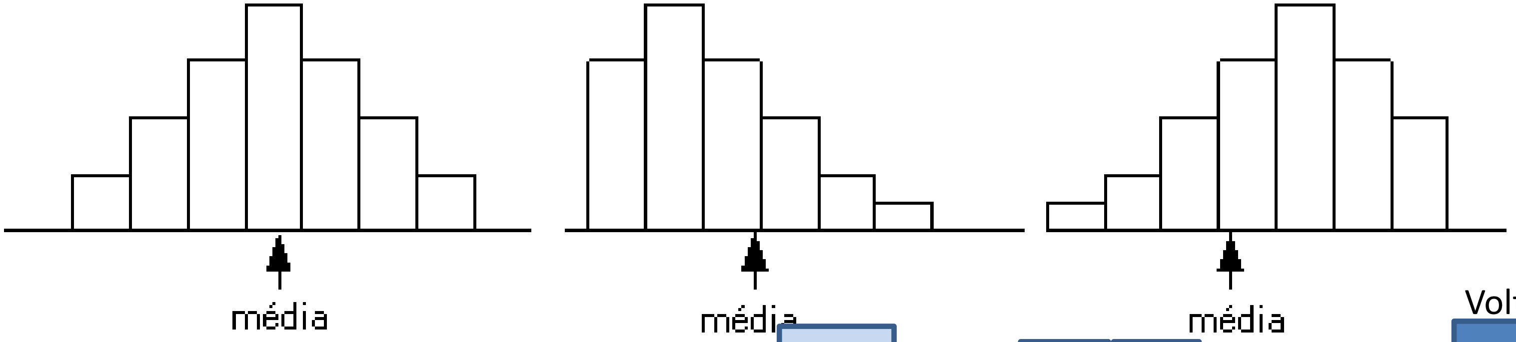
3.1 Medidas de localização Média



Repare-se nas alterações no valor da média, provocadas pela alteração de um dos 4 valores. Nos dois últimos casos 3 dos valores são menores que a média e só um é que é superior, pelo que a média não dá uma boa ideia da distribuição dos dados, nestes casos.

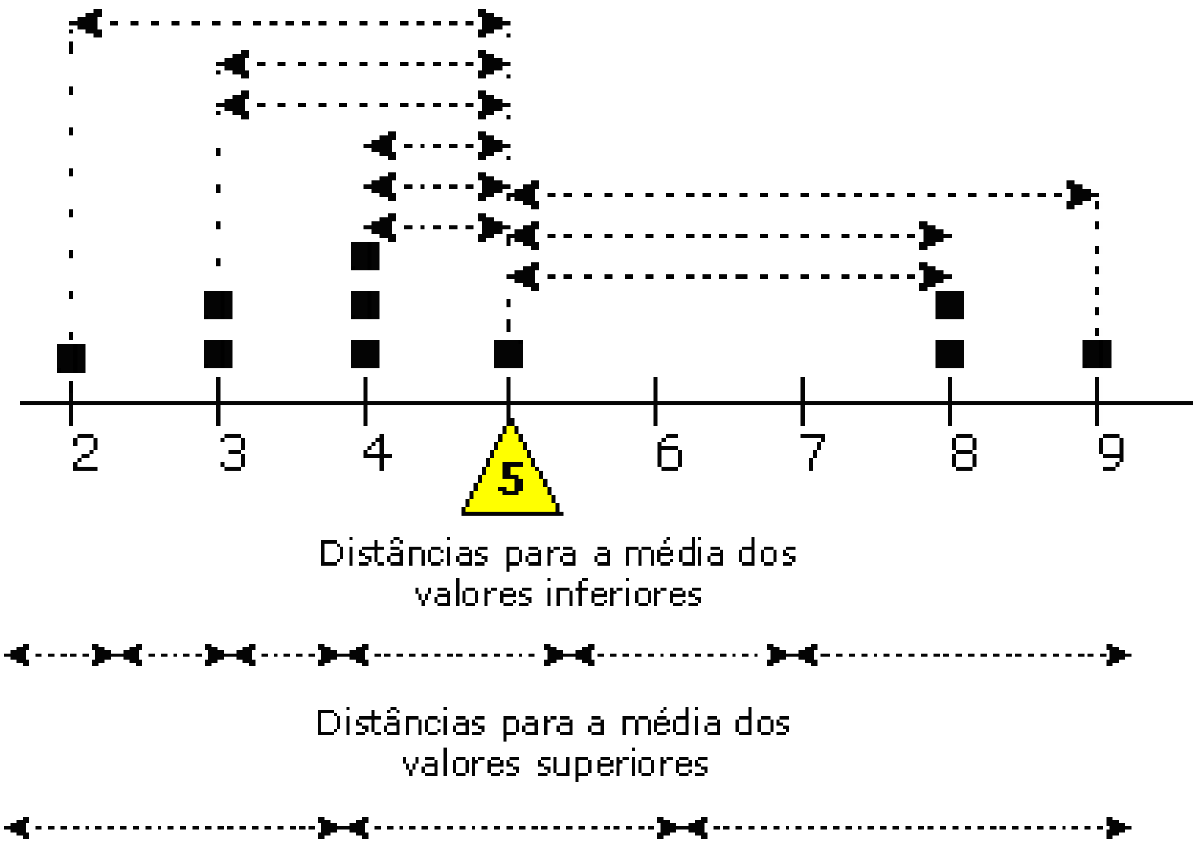
3.1 Medidas de localização Média

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. Na figura seguinte apresenta-se a posição da média em distribuições simétrias e com enviesamento, respectivamente para a direita e para a esquerda



3.1 Medidas de localização Média

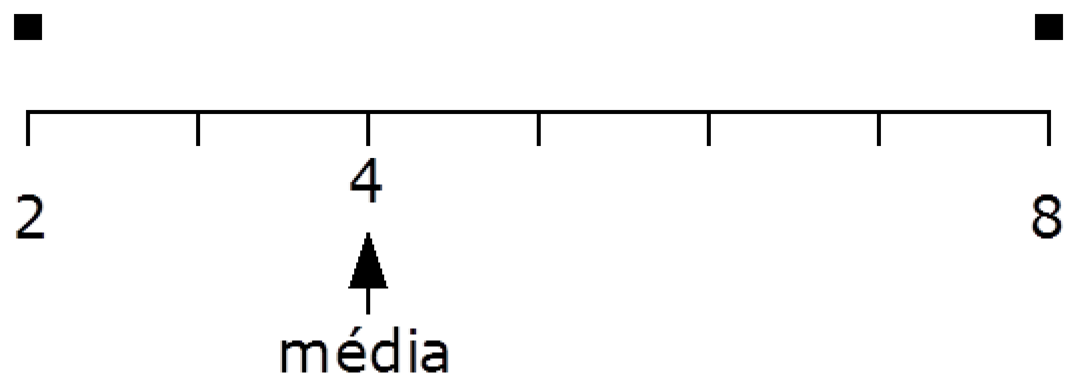
Propriedade da média – num conjunto de dados, a soma das distâncias para a média dos dados que lhe são inferiores é igual à soma das distâncias para a média dos valores que lhe são superiores



3.1 Medidas de localização Média

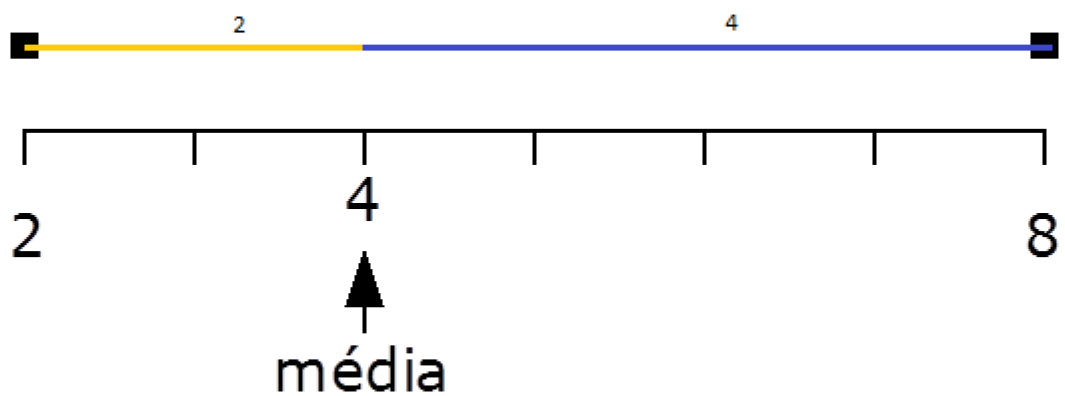
Exemplo - Quais as idades dos meus filhos? Qual a minha idade? O professor chegou à turma e disse: a média das idades dos meus 4 filhos é 4 anos. O mais novo tem 2 e o mais velho 8. Que idades podem ter os meus dois outros filhos? (Graça Martins, M.E. et al OTD – página 131)

O professor desenhou no quadro o gráfico de pontos que ilustrava a situação que acabava de descrever:

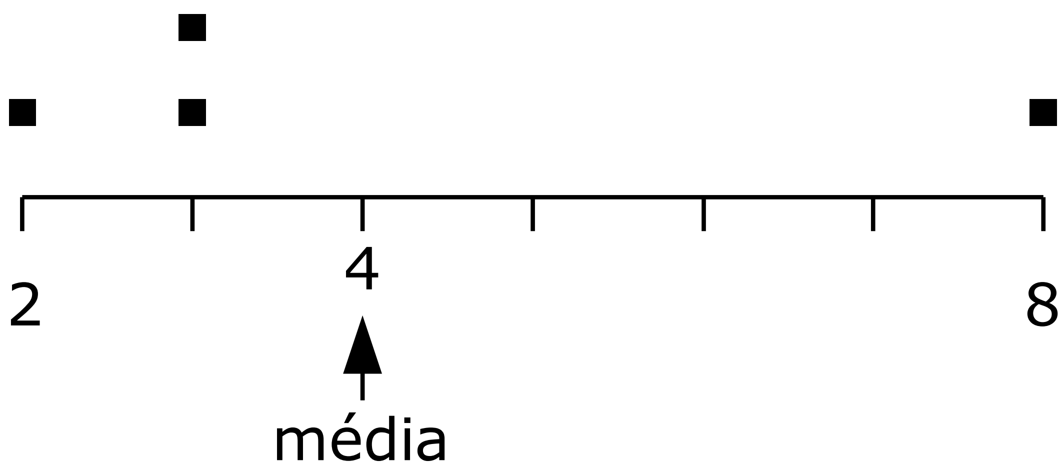


(continua)

3.1 Medidas de localização Média



Para que a soma das distâncias dos valores superiores seja igual à dos valores inferiores, os outros dois têm 3 anos e são gémeos....



3.1 Medidas de localização Média

Quais as idades dos meus filhos? Qual a minha idade? – E se a média da minha idade com as idades dos meus filhos for 9 anos, que idade tenho eu?

$$\frac{2+3+3+8+\text{idadeprofessor}}{5}=9$$

$$\frac{16+\text{idadeprofessor}}{5}=9$$

16 + idade do professor = 45

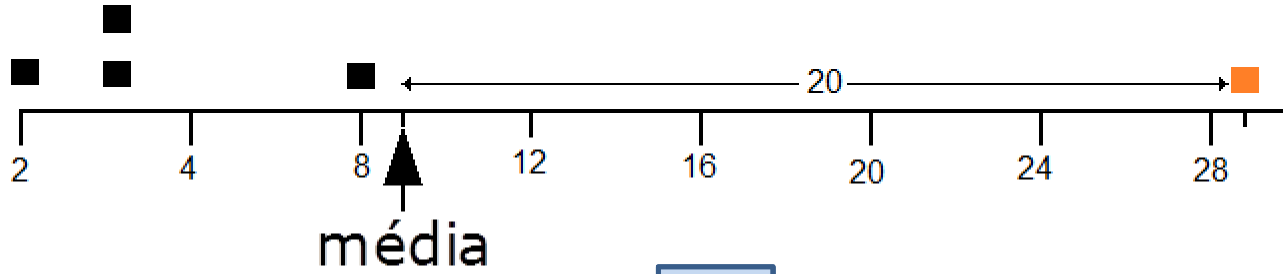
idade do professor **= 29**

Quais as idades dos meus filhos? Qual a minha idade? – E se a média da minha idade com as idades dos meus filhos for 9, que idade tenho eu?

Uma alternativa à resolução anterior é considerar o seguinte raciocínio:

Como os dados abaixo da média distam 20 unidades da média, o dado acima da média também tem de distar 20 unidades pelo que tem de ser igual a **20+9=29**

Idades	Distâncias para a média
2	**9-2=7**
3	**9-3=6**
3	**9-3=6**
8	**9-8=1**
Soma das distâncias	**20**



Nota

A propriedade da ficha pode ser enunciada da seguinte forma:

Propriedade - a soma das distâncias dos dados para a sua média é nula, ou seja, se representarmos um conjunto de n dados por x_1, x_2, x_3, ..., x_n, e a sua média por $\overline{x}$, então

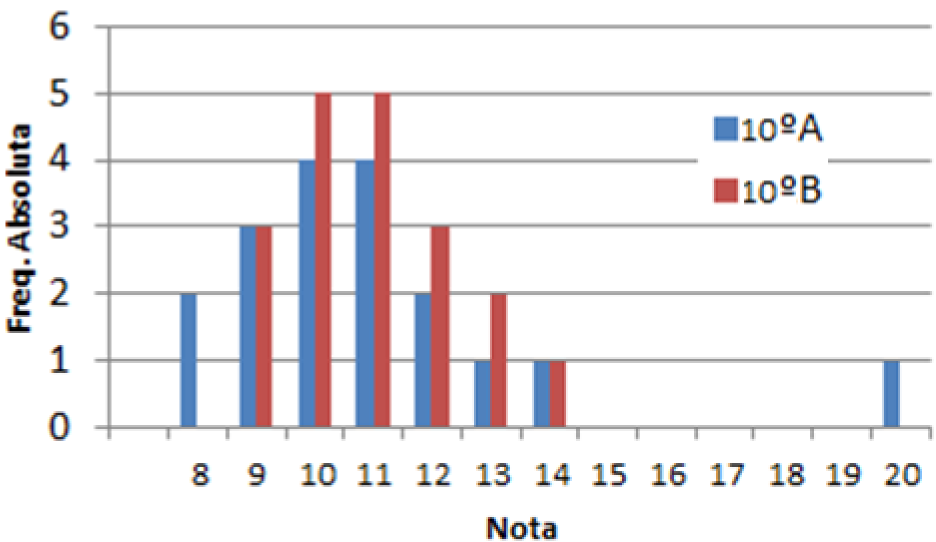
$$(x_1-\overline{x})+(x_2-\overline{x})+(x_3-\overline{x})+\ldots+(x_n-\overline{x})=0$$

Demonstração

$$\begin{aligned}(x_1-\overline{x})+(x_2-\overline{x})+(x_3-\overline{x})+\ldots+(x_n-\overline{x}) &= (x_1+x_2+x_3+\ldots+x_n)-\overline{x}-\overline{x}-\overline{x}-\ldots-\overline{x}\\ &= n\times\overline{x}-n\times\overline{x}\\ &= 0\end{aligned}$$

3.1 Medidas de localização Média

Exemplo – Notas a Português - A professora de Português quando entregou os resultados do último teste, que tinha sido igual para as turmas do 10ºA e 10ºB, comentou que os alunos das duas turmas tinham tido um comportamento muito semelhante no teste, pois as médias das duas turmas era aproximadamente igual. A Maria que sabia que a média nem sempre era uma boa medida para resumir os dados, pediu à professora para lhe facultar a distribuição das notas das duas turmas, pois gostaria de averiguar se efectivamente o comentário da professora estava correcto.



Perante a representação gráfica com a distribuição das notas das duas turmas, a Maria disse à professora:

- Professora, penso que não tem razão quando diz que os alunos das duas turmas tiveram resultados semelhantes! Repare que, de um modo geral a turma B teve resultados melhores e só houve 2 negativas, enquanto que na turma A houve 5 negativas.

O que provocou que a média da turma A seja idêntica à da turma B, apesar da maioria das notas ser inferior, foi um dos alunos ter tido um 20!

3.1 Medidas de localização Média

Exemplo – As idades dos alunos - A turma do 8º A tem 20 alunos, dos quais 12 são do sexo feminino. A média das idades das meninas é 13,4 anos e a média das idades dos rapazes é 13,9 anos. Qual a média das idades dos alunos da turma?

Resolução:
Se a média das idades das meninas é 13,4 anos, então a soma das idades das 12 meninas é 12x13,4 anos =160,8 anos;

Se a média das idades dos rapazes é 13,9 anos, então a soma das idades dos 8 rapazes é 8x13,9 anos =111,2 anos;

$$\text{Média das idades dos alunos da turma} = \frac{\text{soma das idades dos alunos}}{\text{número de alunos}}$$

$$\text{Média das idades dos alunos da turma} = \frac{160{,}8 + 111{,}2}{20}$$

Média das idades dos alunos da turma=13,6 anos

3.1 Medidas de localização Média

Propriedade da média – se num conjunto de dados cuja média é $\overline{x}$ adicionarmos o mesmo valor k a cada um dos dados, então a média dos dados transformados será igual a $\overline{x}$ + k. Demonstração

Exemplo – As notas a Filosofia - O exame de Filosofia tinha sido demasiado difícil e as notas foram um desastre... Nunca tinha acontecido uma média tão baixa de 7 valores (numa escala de 0 a 20). De quanto é que o professor de Filosofia deveria aumentar a nota de cada aluno (o mesmo a todos...), de modo a obter uma média positiva?

Propriedade da média – se num conjunto de dados cuja média é $\overline{x}$ multiplicarmos o mesmo valor k por cada um dos dados, então a média dos dados transformados será igual a $k\times\overline{x}$. Demonstração

Exemplo – O avô e os netos– O avô deu a cada um dos 5 netos uma quantia em dinheiro, em que cada um recebeu em euros um valor igual à sua idade. A média das idades dos netos é 10 anos. Se o avô decidir dar o dobro do que deu a cada um, que quantia é que terá de ter disponível para dar aos netos?

Propriedade da média – se num conjunto de dados cuja média é $\overline{x}$ adicionarmos o mesmo valor k a cada um dos dados, então a média dos dados transformados será igual a $\overline{x}$ + k.

Se representarmos os dados por x_1, x_2, x_3, ..., x_n e a média por $\overline{x}$ então os dados transformados serão x_1+k, x_2+k, x_3+k, ..., x_n+k, com média

$$\frac{x_1+k+x_2+k+x_3+k+\ldots+x_n+k}{n}=\frac{x_1+x_2+x_3+\ldots+x_n+n\times k}{n}=\frac{x_1+x_2+x_3+\ldots+x_n}{n}+\frac{n\times k}{n}=\overline{x}+k$$

Propriedade da média – se num conjunto de dados cuja média é $\overline{x}$ multiplicarmos o mesmo valor k por cada um dos dados, então a média dos dados transformados será igual a $k\times\overline{x}$.

Se representarmos os dados por x_1, x_2, x_3, ..., x_n e a média por $\overline{x}$ então os dados transformados serão $x_1\times k$, $x_2 \times k$, $x_3 \times k$, ..., $x_n \times k$, com média

$$\frac{x_1\times k+x_2\times k+x_3\times k+\ldots+x_n\times k}{n}=\frac{k\times(x_1+x_2+x_3+\ldots+x_n)}{n}=k\times\overline{x}$$

Mediana – A mediana é um valor que divide o conjunto de dados quantitativos, ordenados, ao meio: metade dos dados são não superiores (menores ou iguais) à mediana e os restantes são não inferiores (maiores ou iguais) à mediana. Por outras palavras, até à mediana (inclusivé) estão, pelo menos, 50% dos dados; para lá da mediana (inclusivé) estão também, pelo menos, 50% dos dados (Graça Martins, M. E. et al, 2007, página 117).

Como obter a mediana?

1) Começar por *ordenar* os dados do conjunto de dados;
2) Uma vez os dados ordenados:

- Se o número de dados for ímpar, a mediana é o elemento do meio;
- Se o número de dados for par, a mediana é a semi-soma dos dois elementos do meio.

Nota

Como já se disse anteriormente (ver), quando se representam os dados de um conjunto de n dados por x_1, x_2, x_3, ..., x_n, nesta notação não está implícita qualquer ordem entre os dados. Podemos considerar que se representa por x_1 o primeiro dado a ser recolhido, por x_2 o segundo dado a ser recolhido, etc. Uma vez os dados ordenados, é usual representá-los por $x_{1:n}$, $x_{2:n}$, , ..., $x_{n:n}$ onde $x_{1:n}$ representa o mínimo ou seja o dado de ordem 1 ou na posição 1, $x_{2:n}$ representa o segundo mínimo, ou seja o dado de ordem 2 ou na posição 2, etc. e finalmente $x_{n:n}$ representa o máximo, ou seja o dado de ordem n ou na posição n, ou seja:

$$x_{1:n} \leq x_{2:n} \leq x_{3:n} \leq \ldots \leq x_{n:n}$$

Uma outra notação utilizada para os dados ordenados é considerar $x_{(1)}$, $x_{(2)}$, $x_{(3)}$, ..., $x_{(n)}$, com $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \ldots \leq x_{(n)}$

Se o número de dados for n, uma regra para obter a posição da mediana é fazer o quociente $\frac{n+1}{2}$

Uma vez os dados ordenados

- Se aquele quociente for um número inteiro, o que se verifica quando n é ímpar, toma-se para mediana o elemento nessa posição;
- Se aquele 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.

Nota A mediana não é necessariamente um dos dados. Sê-lo-á, sempre, se o número de dados for ímpar. No caso de ser par (e os dados forem quantitativos) só o será se os dois elementos do meio forem iguais.

3.1 Medidas de localização Mediana

Considerando o exemplo do slide, temos

10 13 12 15 12 9 12 12 12 11

Número de letras no nome
10
13
12
15
12
9
12
12
12
11
28

Ordenar

Dados ordenados

9 10 11 12 12 ↑ 12 12 12 13 15

Mediana =12 (=(12+12)/2) letras

Média =11,8 letras

Se houvesse mais um aluno com 28 letras no nome

9 10 11 12 12 12 12 12 13 15 28
↑

então

Mediana = 12 letras

Média = **13,3** letras

Repare-se que a mediana se manteve, mas a média veio bastante alterada!

NOTA Cálculo da Mediana para dados qualitativos ordinais

Como foi dito no estudo da média, esta só pode ser calculada quando os dados forem quantitativos. E no que respeita a mediana? Pode-se calcular a mediana se os dados forem qualitativos? Sim, se forem qualitativos ordinais.

Vejamos o seguinte exemplo:

A professora ao entregar os testes disse:

- Meninos, tenho aqui os testes e digo-vos que houve 4 insuficientes (I), 6 suficientes (S), 5 bons (B) e 2 muito bons (MB). Alguém é capaz de me dizer qual a mediana dos resultados do teste?

O Pedro respondeu imediatamente:

- Professora, temos 17 notas, que são os nossos dados, de modo que a mediana está na posição *nona* dos dados ordenados. Como os dados ordenados são

  I I I I S S S S **S** S B B B B B MB MB

  então, na *nona* posição temos a nota S, suficiente, que é, portanto, a mediana.

(continua)

3.1 Medidas de localização Mediana

Consideremos agora uma turma com 22 alunos, cujos resultados num determinado teste foram

I I I I S S S S S S S B B B B B B B MB MB MB MB

Como obter a mediana?

Se os dados fossem quantitativos, como temos um número par de dados a mediana seria a semi-soma dos dados nas duas posições do meio, ou sejam, nas posições 11 e 12. No entanto, como temos dados qualitativos, não tem sentido fazer a semi-soma de dados. Então, escolhe-se para mediana o dado na posição 11.

De um modo geral, se tivermos n dados qualitativos ordinais, com n par, considera-se como mediana o elemento na posição n/2, quando os dados estiverem ordenados (posição em que se atinge 50% dos dados menores ou iguais à mediana).

Cálculo da mediana para dados discretos, agrupados

Exemplo - O número de mensagens - Para a disciplina de Matemática, o professor pediu a um grupo de alunos que fizesse um estudo sobre o número de mensagens que os alunos enviam por dia. O grupo de alunos, depois de algumas trocas de impressões com a professora, decidiu perguntar a 100 alunos da escola, escolhidos ao acaso e possuidores de telemóvel, quantas mensagens tinham enviado no dia anterior. Os resultados são apresentados na seguinte tabela de frequências:

Nº mens.	Freq. Abs.	Freq. Abs. Acum.	Freq. Rel.	Freq. Rel. Acum.
0	5	5	0,05	0,05
1	10	15	0,1	0,15
2	14	29	0,14	0,29
3	15	44	0,15	0,44
4	25	69	0,25	0,69
5	18	87	0,18	0,87
6	9	96	0,09	0,96
7	4	100	0,04	1

Vejamos como calcular a mediana a partir da tabela com os dados agrupados.

(continua)

Cálculo da mediana para dados discretos, agrupados

Nº mens.	Freq. Abs.	Freq. Abs. Acum.	Freq. Rel.	Freq. Rel. Acum.%
0	5	5	4,55%	4,55%
1	10	15	9,09%	13,64%
2	15	30	13,64%	27,27%
3	18	48	16,36%	43,64%
4	25	73	22,73%	66,36%
5	20	93	18,18%	84,55%
6	12	105	10,91%	95,45%
7	5	110	4,55%	100,00%

Pela definição de mediana, ela será um valor tal que 50% dos dados ordenados são menores ou iguais a ela e os outros 50% são maiores ou iguais a ela. A partir da tabela e considerando a coluna das frequências relativas acumuladas, verificamos que 43,64% dos dados são menores ou iguais a 3 e 22,73% dos dados são iguais a 4.

Isto significa que o valor de 50% é atingido no dado 4, pelo que a mediana pretendida é 4 mensagens.

De um modo geral se os dados estiverem agrupados, a categoria mediana é aquela onde, pela primeira vez, a frequência relativa acumulada atinge ou ultrapassa 50%. No entanto, se o número de dados for par e em alguma das classes se atingir a frequência relativa acumulada de 50%, temos de fazer a semi-soma com a classe seguinte, como se apresenta a seguir:

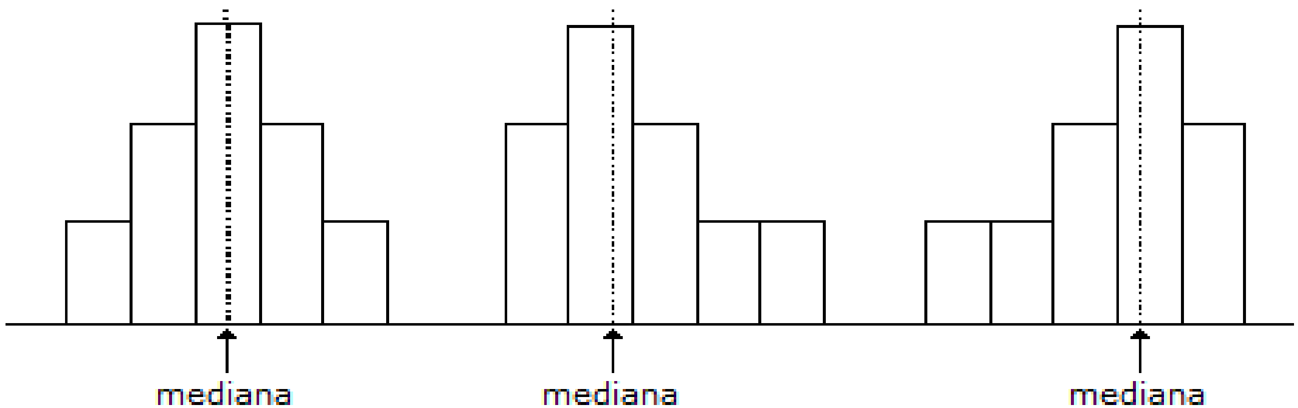
Cálculo da mediana para dados discretos, agrupados

Nº mens.	Freq. Abs.	Freq. Abs. Acum.	Freq. Rel.	Freq. Rel. Acum.%
0	5	5	4,55%	4,55%
1	10	15	9,09%	13,64%
2	15	30	13,64%	27,27%
3	25	55	22,73%	**50,00%**
4	20	75	18,18%	68,18%
5	18	93	16,36%	84,55%
6	12	105	10,91%	95,45%
7	5	110	4,55%	100,00%

Como se verifica na tabela ao lado, a classe 3 corresponde exactamente a 50% de frequência acumulada! Isto é, 50% dos dados são menores ou iguais a 3 e os outros 50% são maiores ou iguais a 4. Esta situação só pode ocorrer quando o número de dados é par, sendo que, neste caso, a mediana é a semi-soma dos dois elementos centrais, ou seja, a mediana será 3,5 mensagens.

Cálculo da mediana para dados agrupados, na forma de um histograma

Se os dados estiverem agrupados na forma de um histograma a posição da mediana é a posição em que passando uma linha vertical por esse ponto o histograma fica dividido em duas partes com áreas iguais

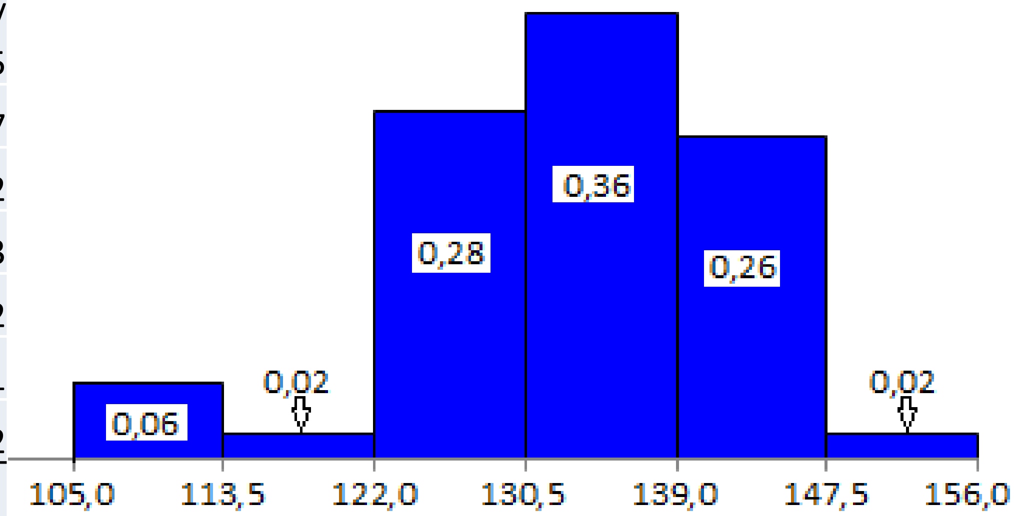


Se os dados nos forem apresentados já organizados na forma de um histograma ou de uma tabela de frequências em que as classes são intervalos, então não temos possibilidade de obter o valor exacto da mediana, mas só um valor aproximado. Vejamos o exemplo a seguir:

Cálculo da mediana para dados agrupados, na forma de um histograma

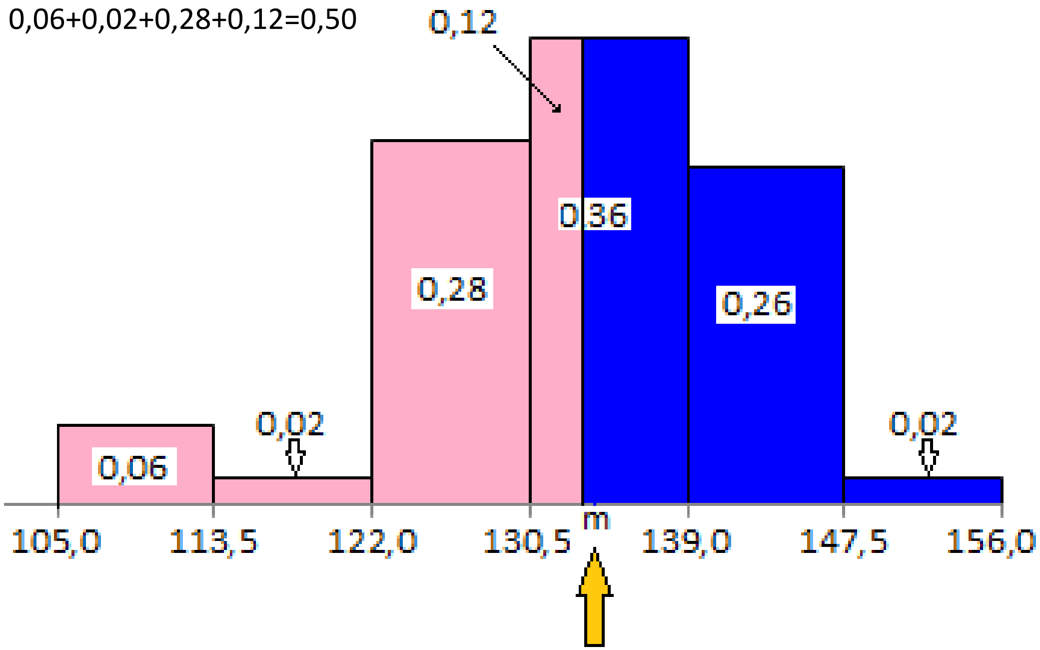
Considerem-se as alturas (em cm) dos primeiros 50 alunos do 1º ciclo, do ficheiro *RespostasMiniCensosNovo* considerado em www.alea.pt→ *conjuntos de dados → Mini-Censos*, e organizem-se na seguinte tabela de frequências, a partir do qual se construíu o histograma (a área do histograma é 1!) ao lado:

Classes	Freq. Abs.	Freq. Rel.	Freq.rel. Acum.	Freq.rel/ 8,5
[105,0;113,5[	3	0,060	0,060	0,007
[113,5;122,0[	1	0,020	0,080	0,002
[122,0;130,5[	14	0,280	0,360	0,033
[130,5;139,0[	18	0,360	**0,720**	0,042
[139,0;147,5[	13	0,260	0,980	0,031
[147,5;156,0[	1	0,020	1,000	0,002
Total	50	1,000		



A classe onde se encontra a mediana é a classe [130,5;139,0[, pois é nesta classe que a frequência relativa acumulada atinge 0,5 ou 50%. (continua)

Cálculo da mediana para dados agrupados, na forma de um histograma



A obtenção de um valor aproximado para a mediana, consiste em obter o ponto m do intervalo [130,5;139,0[tal que se se passar uma linha vertical por ele, a área do histograma à esquerda dessa linha é igual a 0,5.

Admitindo que a área de 0,36 se distribui uniformemente no intervalo assinalado, utilizando uma regra de 3 simples, obtém-se o ponto m:

$$\frac{8{,}5}{0{,}36} = \frac{m - 130{,}5}{0{,}12} \rightarrow \quad m \approx 133{,}3 \text{ cm}$$

O valor aproximado da mediana é 133,3 centímetros.

Propriedade da mediana – se num conjunto de dados cuja mediana é m adicionarmos o mesmo valor k a cada um dos dados, então a mediana dos dados transformados será igual a m+k. Demonstração

Exemplo – No concurso de acesso a determinadom curso do ensino superior só entram os alunos com nota de candidatura positiva e igual ou superior à mediana. No ano passado a mediana das notas de candidatura foi 85 (numa escala de 0 a 100). Indica uma forma de alterar as notas, de forma a que pelo menos 50% dos candidatos possam entrar ao curso em causa?

Propriedade da mediana – se num conjunto de dados cuja mediana é m multiplicarmos o mesmo valor k por cada um dos dados, então a mediana dos dados transformados será igual a k× m. Demonstração

Exemplo – A mediana dos pesos dos bebés de 3 meses da creche ABC é de 5,5kg. Habitualmente as crianças aos 18 meses pesam o dobro do peso que tinham aos 3 meses. Qual o valor que se espera para a mediana dos pesos destas crianças, daqui a 15 meses?

3.1 Medidas de localização Mediana

Propriedade da mediana – se num conjunto de dados cuja mediana é m adicionarmos o mesmo valor k a cada um dos dados, então a mediana dos dados transformados será igual a m+k.

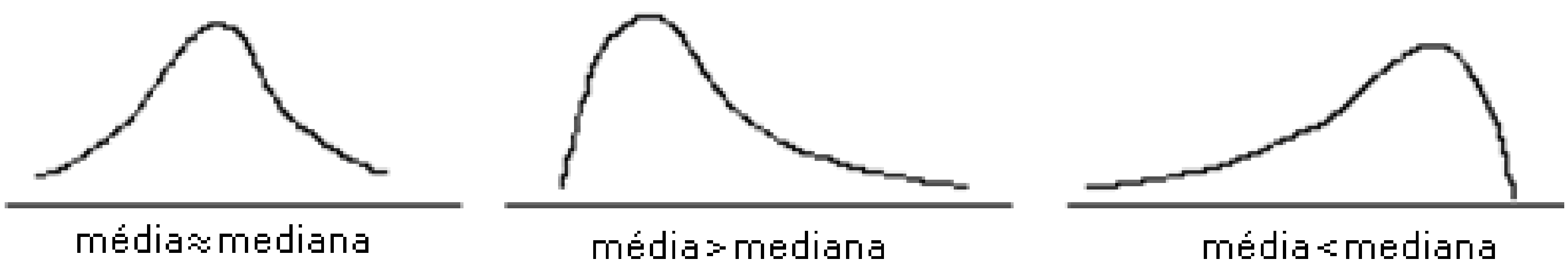
Se representarmos os dados por $x_1, x_2, x_3, \ldots, x_n$ e a mediana por m então os dados transformados serão $x_1+k, x_2+k, x_3+k, \ldots, x_n+k$, e a posição relativa que os dados têm após a transformação é a mesma que antes de transformados, pelo que basta adicionar o valor k à mediana dos dados originais para obter a mediana dos dados transformados.

Propriedade da mediana – se num conjunto de dados cuja mediana é m multiplicarmos o mesmo valor k por cada um dos dados, então a mediana dos dados transformados será igual a $k \times m$.

A demonstração desta propriedade é idêntica à demonstração da propriedade anterior.

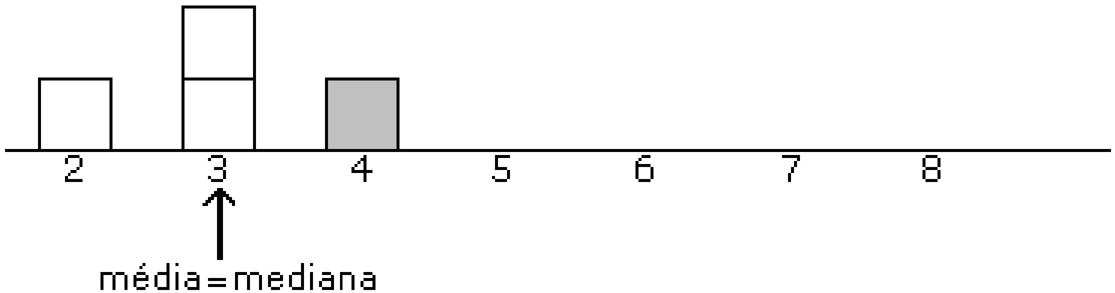
Média ou Mediana?

- Se a distribuição dos dados for aproximadamente simétrica, é preferível a média;
- Se a distribuição dos dados for enviesada, é preferível utilizar a mediana que é, ao contrário da média, uma *medida resistente*.



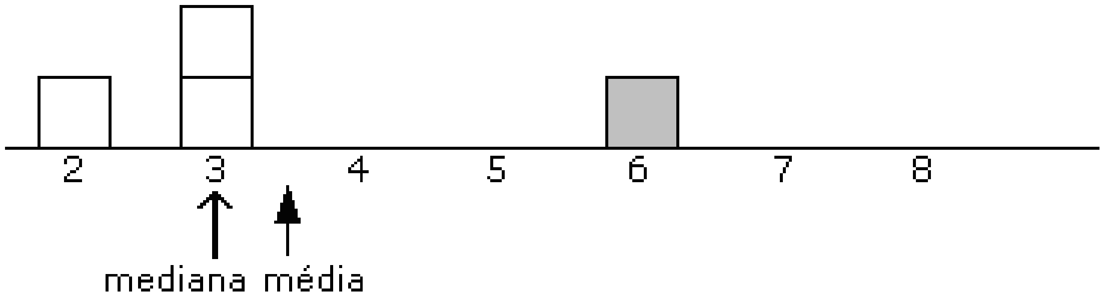
- Se a distribuição é aproximadamente simétrica, a média e a mediana são aproximadamente iguais;
- Quando a distribuição é enviesada para a direita, a média é superior à mediana;
- Quando a distribuição é enviesada para a esquerda, a média é inferior à mediana.

Média ou Mediana?



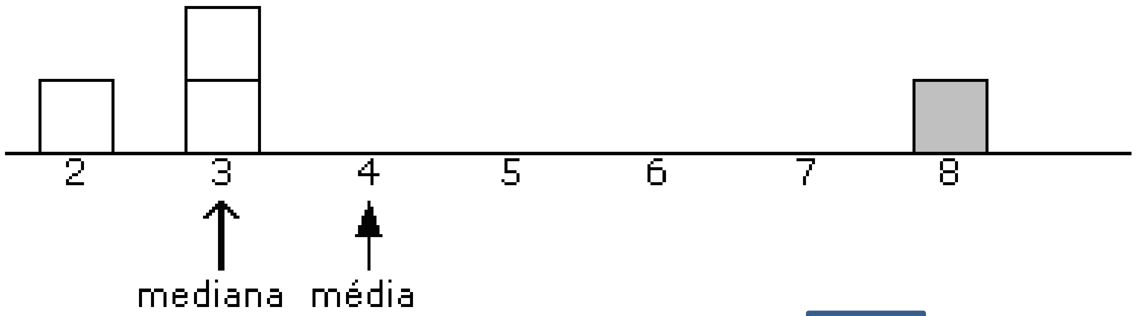
Observações: 2, 3, 3, 4
Média: 3

Mediana: 3



Observações: 2, 3, 3, 6
Média: 3.5

Mediana: 3



Observações: 2, 3, 3, 8
Média: 4

Mediana: 3

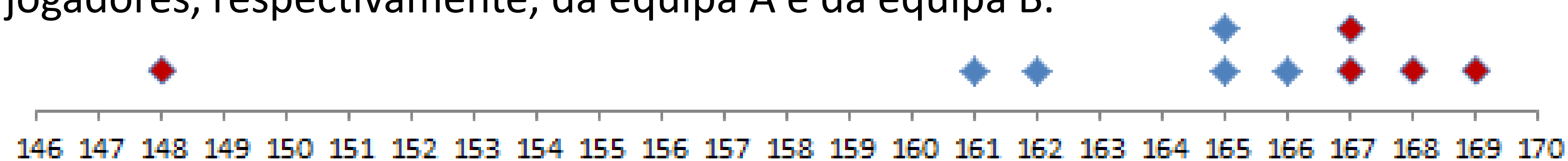
Exemplo – Consultar a Activalea 21 em

(http://www.alea.pt/html/statofic/html/dossier/doc/ActivAlea21.pdf) e

1) Actualizar o tempo de governação do actual presidente.
2) Calcular a média e a mediana
3) Qual das medidas anteriores é a mais conveniente para localizar o centro da distribuição dos dados?

3.1 Medidas de localização Mediana

Exemplo – O jogo de basquetebol - Um grupo de 10 amigos decidiram jogar basquetebol. O primeiro passo consistia em escolher as duas equipas A e B, o que foi feito por sorteio. Por curiosidade, uma vez as equipas constituídas, decidiram calcular a média das alturas dos jogadores de cada equipa. Concluíram que as médias das alturas dos jogadores das duas equipas eram iguais, o que os deixou satisfeitos. Mas afinal, quando se prepararam para jogar, verificaram, com surpresa, que uma das equipas tinha quase todos os jogadores mais altos que a outra, o que a deixava em vantagem! Representaram as alturas num gráfico de pontos, considerando o azul e o vermelho para assinalar os jogadores, respectivamente, da equipa A e da equipa B.



	Equipa A	Equipa B
Média	163,8	163,8
Mediana	165	167

Apesar dos jogadores das duas equipas terem, em média, a mesma altura, são bastante diferentes como se verifica. A mediana é um melhor indicador das alturas dos jogadores das equipas.

3.1 Medidas de localização Média ou Mediana

Exemplo – Os salários na empresa - Numa pequena empresa com oito pessoas, cinco dos trabalhadores ganham, cada um, 1.500€ mensais, dois dos trabalhadores ganham 4.000€ e o gerente ganha 17.500€ mensais. Qual é o salário médio e o salário mediano pago nesta empresa? Quantos dos trabalhadores ganham menos do que a média? Interpreta os resultados obtidos.

Resolução:

$$\text{Média}=\frac{5\text{x}1.500+2\text{x}4.000+17.500}{8}$$

$$= 4.125$$

A média dos salários é 4.125€, enquanto a mediana é 1.500€.

Sete dos trabalhadores ganham menos do que a média e só um ganha mais do que a média.

A média é muito superior à mediana devido à existência de um salário muito superior aos restantes.

3.1 Medidas de localização Moda

Moda – Para os dados qualitativos ou categóricos vimos que se designa por moda a categoria de maior frequência. Para um conjunto de dados quantitativos define-se **moda** como sendo o valor que surge com maior frequência, se os dados são discretos, ou o intervalo de classe com maior frequência, se os dados são contínuos. No entanto, para os dados quantitativos generaliza-se a noção de moda. Assim, designa-se por **moda** ou **classe modal** qualquer classe (valor no caso discreto(1) ou intervalo no caso contínuo) que esteja ladeada por classes de menor frequência. As **modas** serão os "picos" na distribuição de frequências.

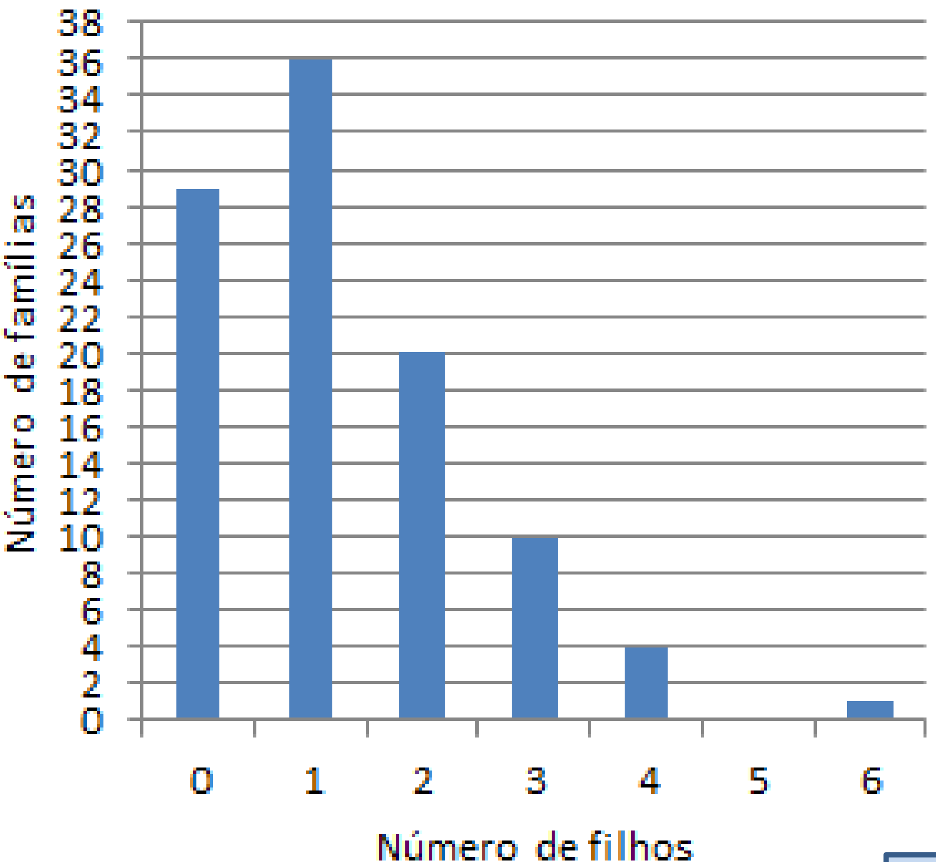
(1) Se os dados discretos estiverem agrupados na forma de intervalos, também se considera a classe modal na forma de intervalo. Veja-se o exemplo da ficha .

3.1 Medidas de localização Moda

Não sendo propriamente uma medida de localização, embora seja apresentada como tal, a moda deve a sua importância ao facto de ser a única medida que pode ser calculada para dados qualitativos, não susceptíveis de ordenação. No caso de dados quantitativos discretos, tem também algum interesse quando dispomos de um conjunto de dados 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 usual do pé da criança de determinada classe etária. O dono de uma sapataria tem interesse em saber qual o tamanho mais vendido, pois será nesse tamanho que vai investir, no princípio de cada época.

Para saber mais consultar a ActivAlea número 17 em http://www.alea.pt/html/statofic/html/dossier/doc/ActivAlea17.pdf

Exemplo - Perguntou-se a 100 famílias portuguesas, escolhidas ao acaso, quantos filhos tinham. Apresenta-se o resultado no seguinte gráfico de barras:



- Quantas famílias disseram não ter filhos?
- Quantas famílias têm 5 filhos?
- Qual a moda da distribuição do número de filhos das 100 famílias portuguesas?

Quartis - Os **quartis** são medidas de localização que dividem o conjunto de dados de tipo quantitativo (ou qualitativo ordinal) ordenados, em quatro partes, cada uma com uma percentagem de dados aproximadamente igual. O 1º quartil (ou *quartil inferior)*, representado por Q_1 (ou $Q_{0,25}$) e o 3º quartil (ou *quartil superior)*, representado por Q_3 (ou $Q_{0,75}$) são medidas que localizam alguns pontos da distribuição dos dados de tal forma que aproximadamente 25% dos dados são inferiores ou iguais a Q_1, aproximadamente 25% dos dados são superiores ou iguais a Q_3 e os restantes dados, aproximadamente 50%, situam-se entre Q_1 e Q_3. O 2º quartil é o mesmo que mediana. De um modo geral, quando nos referimos aos quartis, estamos a referir-nos ao 1º e 3º quartis, uma vez que o 2º quartil é designado por mediana (Graça Martins, E. (2013), WikiCiências, 4(07):0777).

3.1 Medidas de localização Quartis

Dado um conjunto de dados (Graça Martins, E. (2013), WikiCiências, 4(07):0777), considera-se que o **1º quartil** é um valor tal que pelo menos 25% dos dados são não maiores (menores ou iguais) do que ele e pelo menos 75% dos dados são não menores (maiores ou iguais) do que ele e o **3º quartil** é um valor tal que pelo menos 75% dos dados são não maiores do que ele e pelo menos 25% dos dados são não menores do que ele.

Há vários processos para calcular os quartis, que não conduzem necessariamente 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 resumir a informação contida nos dados, através de algumas medidas.

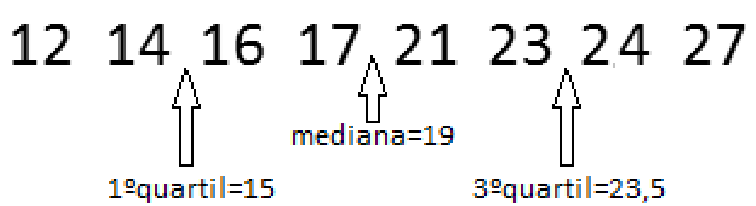
3.1 Medidas de localização Quartis

Uma metodologia que pode ser seguida para obter os quartis é a seguinte:

- Ordenar os dados por ordem crescente e calcular a mediana;
- O 1.º quartil, Q_1, é a mediana dos dados que ficam para a esquerda da mediana;
- O 3.º quartil, Q_3, é a mediana dos dados que ficam para a direita da mediana.

Quando o número de dados é par, por exemplo n=2×m, não há ambiguidade na definição dos quartis, já que serão sempre as medianas de conjuntos de m dados.

Por exemplo, no caso dos dados 12, 14, 16, 17, 21, 23, 24, 27, temos



3.1 Medidas de localização Quartis

O mesmo não acontece quando n é ímpar, por exemplo n=2×m+1, em que a mediana é o dado na posição m+1, quando os dados estão ordenados e em que para o cálculo do 1º e 3º quartis duas metodologias podem ser seguidas:

- Calcular as medianas do conjunto dos dados da posição 1 até à posição m+1 e do conjunto dos dados da posição m+1 até à posição 2×m+1 (a mediana está incluída nos dois conjuntos de dados), ou
- Calcular as medianas do conjunto de dados da posição 1 até à posição m e do conjunto dos dados da posição m+2 até à posição 2×m+1.

Por exemplo, no caso dos dados 12, 14, 16, 17, 21, 23, 24, 27, 30, temos, respectivamente:

12 14 16 17 21 23 24 27 30

mediana=21

1ºquartil=16

3ºquartil=24

ou

12 14 16 17 21 23 24 27 30

mediana=21

1ºquartil=15

3ºquartil=25,5

Atenção

É frequente confundir , num conjunto de dados, o **dado mais frequente** com o **dado de maior valor**, ou seja, confundir a moda com o dado de maior valor. Veja-se, por exemplo, a ficha .

Nota

Nenhum dos processos anteriores conduz a que a definição que foi dada para os quartis se cumpra estritamente, no que respeita as percentagens consideradas de 25% e 75%. Há estatísticos que utilizam o primeiro processo e outros que utilizam o segundo. A metodologia utilizada pelos autores do programa é a segunda em que, para o cálculo dos quartis, não se considera a mediana como fazendo parte dos dois conjuntos em que os dados ficam divididos. Chamamos, no entanto, a atenção que, quando o número de dados é ímpar, o processo seguido pelo Excel é o que conduz aos valores dos quartis em que se considera a mediana como pertencente aos dois conjuntos de dados, ou seja, a metodologia indicada em primeiro lugar.

Vejam-se os seguintes exemplos, em que se juntou o valor do 1º quartil calculado na folha de Excel:

3.1 Medidas de localização Quartis

No exemplo considerado anteriormente, em que n=9, tem-se a seguinte tabela:

N=9	Q_1	% dados $\leq Q_1$	% dados $\geq Q_1$
1º processo	16	33%	78%
2º processo	15	22%	78%
Excel	16	33%	78%

Se aos dados se juntarem os valores 31 e 35, tem-se:

12 14 16 17 21 23 24 27 30 31 35

1ºquartil= 16,5 mediana= 23 3ºquartil= 28,5

12 14 16 17 21 23 24 27 30 31 35

1ºquartil= 16 mediana= 23 3ºquartil= 30

N=11	Q_1	% dados $\leq Q_1$	% dados $\geq Q_1$
1º processo	16,5	27%	73%
2º processo	16	27%	82%
Excel	16,5	27%	73%

3.1 Medidas de localização Propostas de exercícios

1. O queijo e as calorias (Adaptado de Graça Martins, M.E. et al 2007, página 135) – 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 ao lado apresenta-se, para vários tipos de queijo, a quantidade de gordura e o número de calorias, por cada 100 gramas de queijo.

1) Constrói um gráfico de caule-e-folhas para os dados referentes às calorias. Notas algum enviesamento na distribuição dos dados? O que é que esperas quanto à relação de grandeza entre a média e a mediana?

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
Queijo Gorgonzola	37	407
Queijo Gruyère	20	315
Queijo Parmesão	28	401
Queijo Roquefort	32	371
Queijo Suiço	29	357

3.1 Medidas de localização Propostas de exercícios

b) Ordena os queijos pelo seu valor calórico. Qual o tipo de queijo que é mais calórico? Esse queijo é também, o que tem mais gordura?

c) Calcula a mediana do número de calorias.

d) Calcula a média do número de calorias e compara com a mediana. Dessa comparação resultou o que estavas à espera?

e) Constrói um diagrama de dispersão dos dados (Gordura, Calorias). Comenta a relação existente entre a variável Gordura e a variável Calorias.

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
Queijo Gorgonzola	37	407
Queijo Gruyère	20	315
Queijo Parmesão	28	401
Queijo Roquefort	32	371
Queijo Suiço	29	357

2. Os golos do Cristiano Ronaldo

Na tabela ao lado estão registados, por época, o número de jogos que o Cristiano Ronaldo jogos e o número de golos marcados.

a) Calcula quantos golos, em média, o Ronaldo marca por jogo e acrescenta uma nova coluna à tabela ao lado com os resultados obtidos.

b) Considera os resultados obtidos na alínea anterior como os teus novos dados. Calcula a média e a mediana. Qual destas medidas achas que representa melhor a actual "perfomance" do Cristiano Ronaldo (estamos em 2014)?

Época	Nº jogos	Nº golos	Nº golos, em média, por jogo
2003/2004	40	6	
2004/2005	50	9	
2005/2006	47	12	
2006/2007	53	23	
2007/2008	49	42	
2008/2009	53	26	
2009/2010	35	33	
2010/2011	54	53	
2011/2012	55	60	
2012/2013	55	55	
2013/2014	47	51	

3. Os resultados do teste

Num ponto de Matemática com 5 questões, cada uma valendo 4 valores, verificaram-se os seguintes resultados:

a) Se o teste foi realizado por 100 alunos, qual a pontuação média obtida?

b) Se o teste foi realizado por 200 alunos, qual a pontuação média obtida?

c) Será que se pode calcular a média sem saber o número de alunos? Deduz uma expressão para o cálculo da média, quando os dados estão agrupados e se tem a frequência relativa de cada valor.

d) Qual o valor da mediana? Compara com a média e interpreta os resultados obtidos.

5% dos alunos tiveram			0
10%	“	“	4
25%	“	“	8
40%	“	“	12
15%	“	“	16
5%	“	“	20

4. As idades dos alunos da turma A

Na tabela seguinte estão indicadas as idades dos alunos da turma A

Idade	13	14	15	m
Nº alunos	1	10	12	2

a) Sabe-se que a média das idades é 14,68 anos. Qual o valor de m?

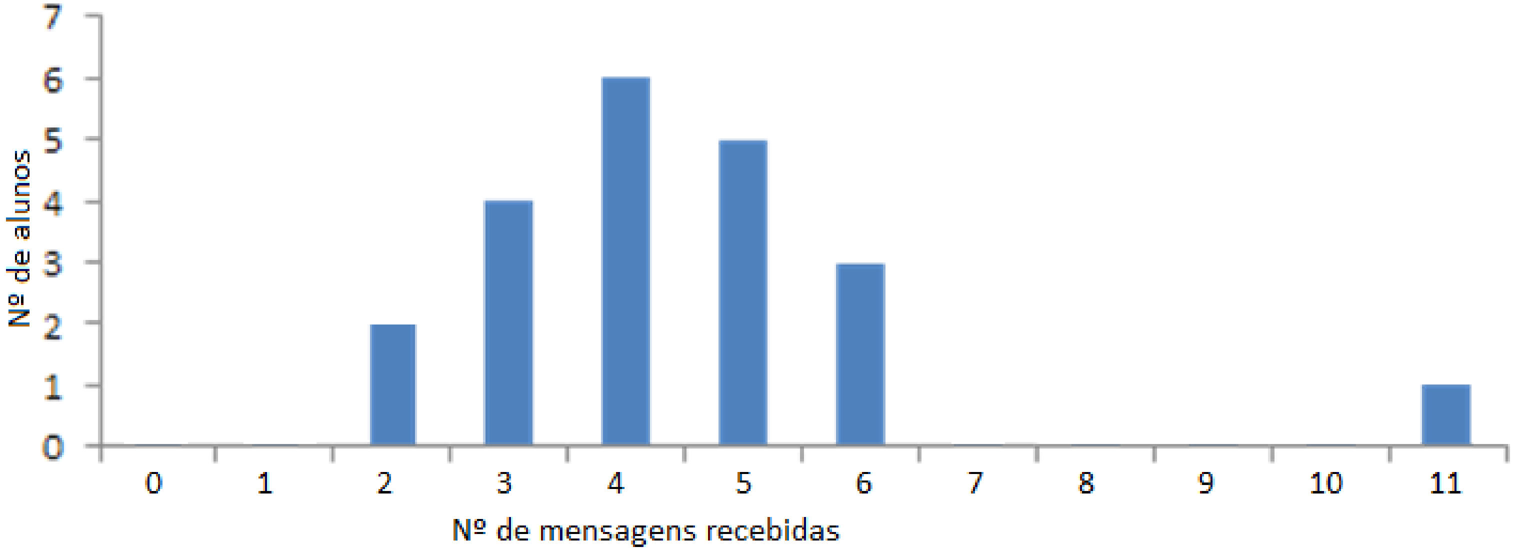
b) Daqui um ano, qual a média das idades destes alunos?

c) Qual o valor da mediana das idades?

d) Antes de construir um gráfico com a distribuição dos dados, já se pode esperar que ela seja aproximadamente simétrica. Comenta esta afirmação.

5. Número de mensagens recebidas

O gráfico seguinte representa o número de mensagens recebidas no fim de semana pelos alunos de uma turma.



a) Quantos alunos estavam na turma, no dia em que se recolheu a informação sobre o número de mensagens recebidas?

b) Qual a mediana do número de mensagens recebidas?

c) Tendo em consideração o aspecto do gráfico, esperas que a média do número de mensagens recebidas seja superior ou inferior à mediana? Porquê? Verifica se a tua suposição está correcta.

6. As poupanças

1) Nos últimos 5 meses a Maria conseguiu poupar em média 24€. Vem aí o mês de Dezembro e no Natal a Maria costuma receber algum dinheiro. Quanto é que ela recebeu, sabendo que a média das suas poupanças subiu para 50€?

2) O avô do João costuma dar aos 5 netos uma média de 15€ por mês. No mês de natal ele decidiu aumentar da mesma quantia cada um dos netos e vai dar no total 250€. De quanto é que ele aumentou cada neto?

3) O avô do Manel deu no último natal uma média de 100€ aos seus 5 netos. Se ele no próximo natal quiser duplicar o que dá a cada um, quanto é que vai ter de gastar?

7. As turmas A e B fizeram o mesmo teste e a partir dos resultados obtidos calcularam-se as seguintes estatísticas:

	Turma A	Turma B
Média	11	13
1º quartil	8	6
3º quartil	10	10
Mediana	9	8

A partir da tabela anterior, podes comentar, a teu ver, qual a turma que se portou melhor neste teste?

Explica o teu raciocínio.

8. Na consulta dos 6 meses o pediatra disse à mãe do bebé: - *Minha senhora, o seu filho está acima do 3º quartil, no que diz respeito ao peso*.

A mãe terá que ter alguns cuidados com a alimentação do bebé ou pode continuar a proceder como até à data?

3.2 Medidas de dispersão

As medidas de localização não são suficientes para caracterizar completamente um conjunto de dados.

Dois países podem ter o mesmo rendimento médio e o mesmo rendimento mediano per capita e no entanto serem muito diferentes. Esta situação verificar-se-á, por exemplo, se num dos países houver grandes diferenças de rendimento, com uma grande percentagem de pessoas com rendimentos baixos e uma grande percentagem de pessoas com rendimentos altos, enquanto que no outro país os rendimentos não diferirem muito entre as pessoas.

Do mesmo modo, duas turmas podem ter um comportamento muito diferente na realização do mesmo teste, embora apresentando uma média e mediana iguais.

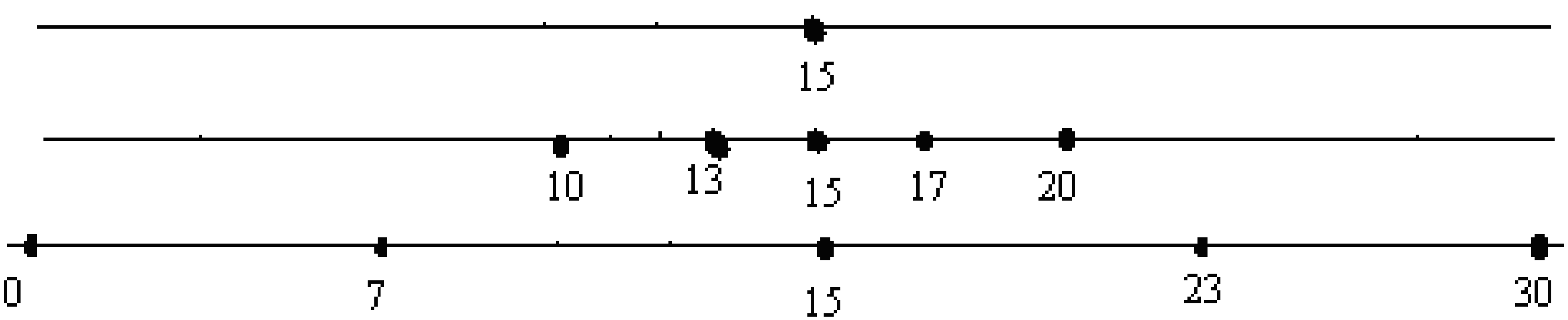
3.2 Medidas de dispersão (ou variabilidade)

Para caracterizar um conjunto de dados, é necessário acrescentar às medidas de localização, medidas que deem indicação da variabilidade presente nos dados.

Considerem-se, por exemplo, os três conjuntos de dados:

Conjunto 1	15	15	15	15	15
Conjunto 2	10	13	15	17	20
Conjunto 3	0	7	15	23	30

Embora tenham a mesma média e mediana, têm um aspecto bem diferente no que diz respeito à variabilidade ou dispersão



3.2 Medidas de dispersão (ou variabilidade)

Um aspecto importante na caracterização de um conjunto de dados é o da determinação da variabilidade ou dispersão desses dados, entre eles, ou relativamente à medida de localização do centro da distribuição dos dados. Como a medida de localização mais utilizada é a média, é relativamente a ela que se define a principal medida de dispersão - o desvio padrão. No entanto, por não fazer parte do programa em vigor para o ensino básico, o desvio padrão não será, neste curso, objecto de estudo.

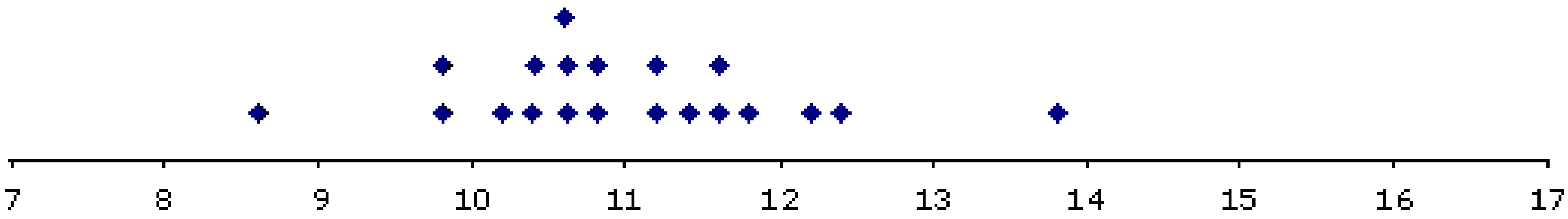
Assim, as medidas de dispersão que serão definidas são a **amplitude** e a **amplitude interquartil**, que dão uma indicação, para cada conjunto de dados, da variabilidade entre os dados.

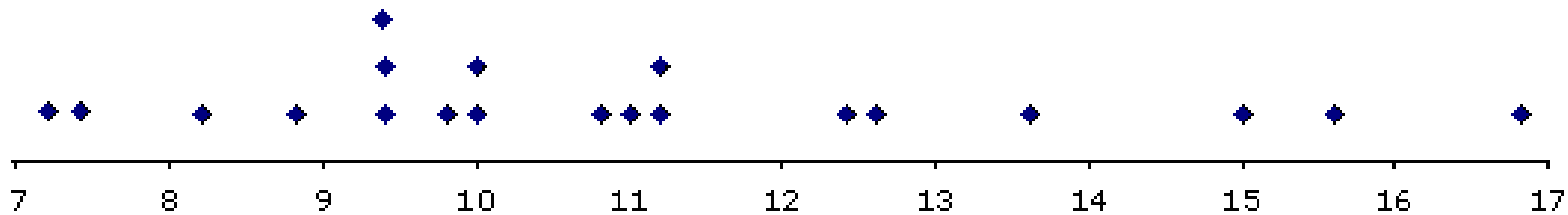
3.2 Medidas de dispersão **Amplitude**

Amplitude – é a medida de dispersão ou variabilidade que se obtém fazendo a diferença entre o máximo e o mínimo do conjunto de dados

Amplitude = máximo – mínimo

Por exemplo, os dois conjuntos de dados representados nos gráficos de pontos seguintes, têm a mesma média,





mas no primeiro caso temos uma amplitude à volta de 5, enquanto que no segundo caso a amplitude anda à volta de 10, denotando uma maior dispersão entre os dados.

A **Amplitude** como medida de variabilidade é muito pobre, pois é muito pouco resistente. Basta haver no conjunto de dados um valor discrepante, normalmente designado de "outlier" para que a amplitude não transmita a variabilidade dos dados, por ser muito influenciada por esse valor discrepante!

Uma alternativa à amplitude, é a **amplitude interquartil**, já que o desvio padrão, a medida de dispersão ou variabilidade mais utilizada, está fora do âmbito deste curso, como se disse anteriormente.

Amplitude interquartil - é a medida de dispersão ou variabilidade que se obtém fazendo a diferença entre o 3º e 1º quartis

Amplitude interquartil = 3º quartil – 1º quartil

A amplitude interquartil dá informação sobre a amplitude do intervalo que contém os 50% dos dados centrais do conjunto de dados.

Esta medida é *não negativa* e *será tanto maior quanto maior for a variabilidade* nos dados. No entanto, uma amplitude interquartil nula, não significa necessariamente, que os dados não apresentem variabilidade. Por exemplo, o seguinte conjunto de dados

10 20 30 30 30 30 30 30 40 50

tem amplitude interquartil igual a zero, mas apresenta variabilidade.

4 Dados quantitativos – outra representação gráfica

Algumas das medidas definidas anteriormente, nomeadamente a mediana e os quartis, permitem construir outras representações gráficas tanto para dados quantitativos discretos como contínuos – o **diagrama de extremos e quartis** e a **caixa dos bigodes**.

Só será abordada a construção do diagrama de extremos e quartis já que a caixa dos bigodes, uma versão do diagrama de extremos e quartis mais elaborada e que permite detectar os valores do conjunto dos dados considerados *outliers*, não pertence ao programa.

Diagrama de extremos e quartis

Diagrama de extremos e quartis - representação gráfica muito simples, mas que evidencia de uma forma bastante eficaz a maneira como os dados se distribuem. É construída à custa de 5 números: a mediana, os 1º e 3º quartis e os extremos, ou seja, o mínimo e o máximo dos dados.

Para construir o **diagrama de extremos e quartis**

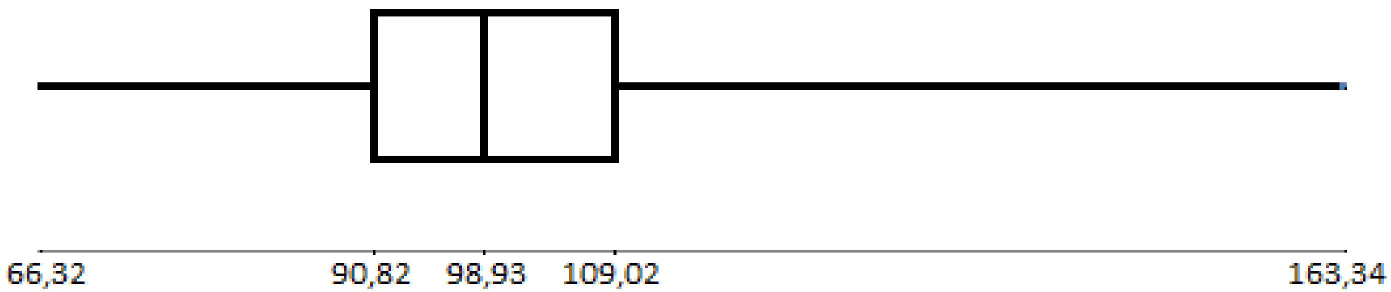
- começa por se desenhar um rectângulo com base de comprimento igual à amplitude interquartil e de altura qualquer valor;
- de seguida, do meio dos lados perpendiculares à base desenham-se segmentos de recta que unem os lados, respectivamente, com o mínimo e o máximo;
- Finalmente, no interior do rectângulo desenha-se um segmento de recta , perpendicular à base e que assinala a posição da mediana.

Diagrama de extremos e quartis

Exemplo - Do ficheiro Dados sobre casas, considerem-se os dados relativos às áreas das casas. Ordenados os dados, obteve-se o seguinte conjunto de 5 números

Mediana	1ºquartil	3ºquartil	Mínimo	Máximo
98,93m^2	90,82m^2	109,02m^2	66,32m^2	163,34m^2

com o qual se desenhou o diagrama de extremos e quartis seguinte



Da representação anterior conclui-se que os dados se distribuem com assimetria positiva, o que significa que há algumas casas cuja área é razoavelmente superior à área da maior parte das casas (cont)

Diagrama de extremos e quartis

Repare-se que a conclusão que se retirou anteriormente já tinha sido evidenciada por outras representações gráficas como o caule-e-folhas e o histograma (ver).

Pelo comprimento dos segmentos de recta que saem dos lados do rectângulo conclui-se que há uma grande variabilidade entre as áreas das 25% das casas com maior área, ou seja, com área superior ao 3º quartil, sendo essa variabilidade bastante superior à que existe entre as áreas das 25% das casas com menor área. Por outro lado, a parte central dos dados distribui-se de forma aproximadamente simétrica, já que a mediana está sensivelmente a meio do rectângulo.

O diagrama de extremos e quartis é muito útil para comparar vários conjuntos de dados, como se exemplifica a seguir.

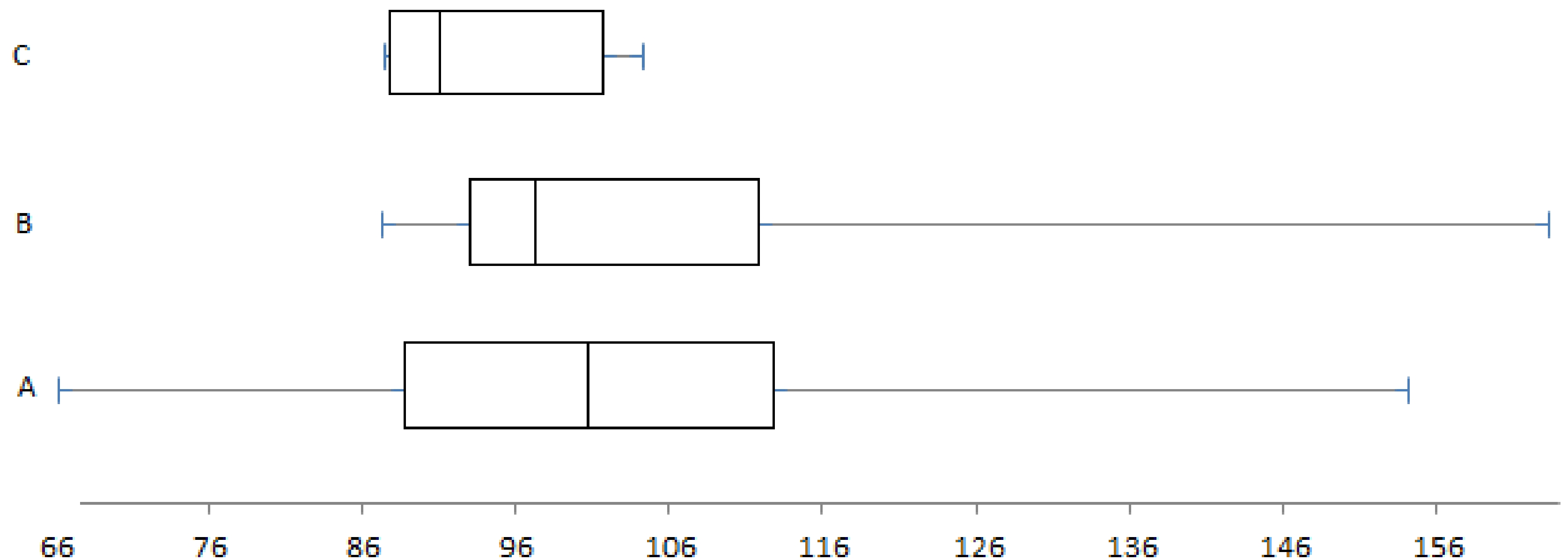
Diagrama de extremos e quartis

Exemplo – Considerando ainda os dados referentes às casas do ficheiro Dados sobre casas, vamos comparar as áreas das casas das diferentes zonas. Os números necessários para construir os diagramas de extremos e quartis paralelos, foram calculados a partir dos dados referentes a cada uma das zonas e estão na tabela seguinte:

Zona	Mediana	1ºquartil	3ºquartil	mínimo	máximo
A	100,82	88,86	112,87	66,32	154,21
B	97,24	93,04	111,82	87,26	163,34
C	91,14	87,76	101,64	87,57	104,28

Para cada zona constrói-se um diagrama de extremos e quartis e apresentam-se em paralelo para mais fácil ciomparação dos e conjuntos de dados:

Diagrama de extremos e quartis



Da representação anterior verifica-se que a distribuição dos dados referentes à zona A é, na parte central, aproximadamente simétrica, ao contrário do que se verifica nas zonas B e C. Verificamos ainda que as distribuições das áreas das casas das zona B e C apresentam um acentuado enviesamento para a direita, isto é, a variabilidade entre as áreas das 50% das casas maiores é bastante superior à variabiliadde entre as áreas das 50% das casas de menor área. De referir ainda que a variabilidade apresentada pelos dados referentes à zona C é substancialmente mais pequena do que os das outras zonas, mas não podemos deixar de referir que na zona C só tínhamos 5 casas.

Diagrama de extremos e quartis

Os diagramas de extremos e quartis são particularmente úteis para comparar a distribuição de vários conjuntos de dados, no que diz respeito à:

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

realçando aspectos particulares, como:

- Comparação das medianas;
- Comparação da dispersão entre os dados, utilizando as amplitudes interquartis;

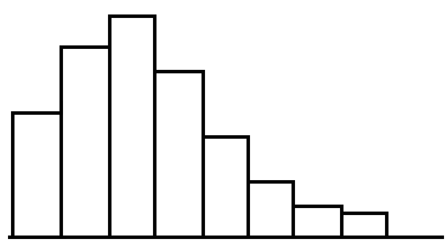
Por outro lado, o comprimento dos traços qua saem dos lados do rectângulo também nos podem alertar para a existência de possíveis "outliers" (valores muito grandes ou muito pequenos, relativamente aos restantes).

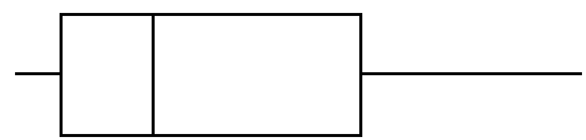
Diagrama de extremos e quartis

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

Para os dados quantitativos contínuos vimos 3 formas de os representar: caule-e-folhas, histograma e diagrama de extremos e quartis. Anteriormente frisámos o facto de a informação transmitida pelo histograma sobre o padrão da distribuição dos dados, ser idêntica à transmitida pelo gráfico de caule-e-folhas (ver). 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 dos dados 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
```





(continua)

Diagrama de extremos e quartis

Quando se faz a representação dos dados, perde-se sempre alguma da 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 no 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 (Graça Martins, M. E. et al, 2011, página 110). A representação em que se perde menos informação é o caule-e-folhas, mas a sua construção pode apresentar menor maleabilidade do que o histograma devido à especificidade da escolha dos caules e pode ser pouco prática se o número de dados for razoavelmente grande.

Diagrama de extremos e quartis Propostas de exercícios

1. A professora chegou à aula e disse: - Cerca de metade dos alunos tiveram nota positiva no teste e cerca de um quarto dos alunos tiveram nota menor ou igual a 6 (numa escala de 0 a 20). Também vos digo que a amplitude interquartil é 7.

Respondam às seguintes questões:

a) Cerca de 50% dos alunos tiveram nota entre ____ e _____

b) Cerca de 25% dos melhores resultados foram superiores a ____

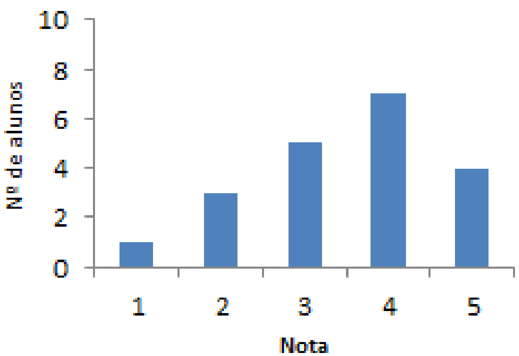
c) Cerca de 25% dos piores resultados foram inferiores a _____

d) Façam uma representação gráfica adequada dos resultados obtidos, sabendo que a nota mínima foi 3 e a máxima 18.

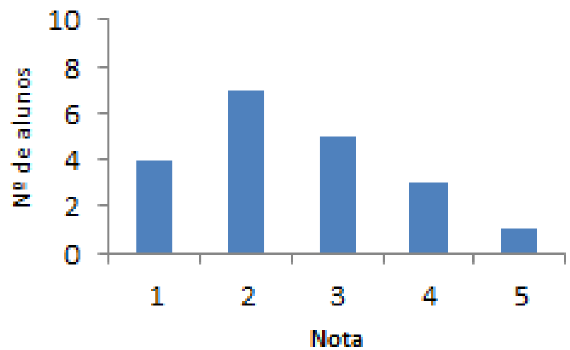
Diagrama de extremos e quartis Propostas de exercícios

2. Considera os seguintes diagramas de barras que representam os resultados no mesmo teste de 3 turmas:

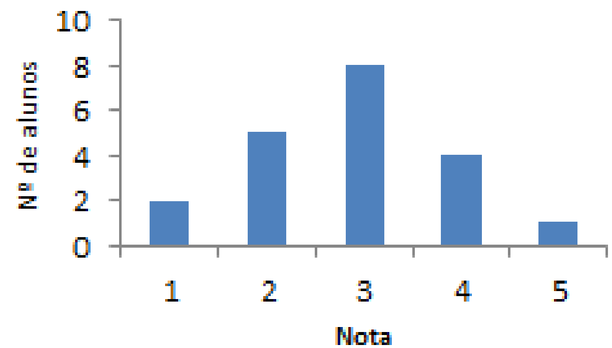
Turma A



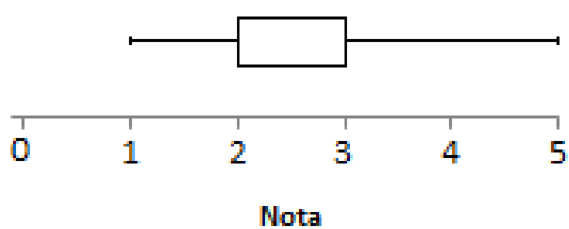
Turma B



Turma C



Para uma dessas turmas, os resultados também foram apresentados com o seguinte diagrama de extremos e quartis:



Qual foi a turma que foi representada?

5 Probabilidade

- Introdução
- Fenómeno aleatório
- Experiência aleatória
- Espaço de resultados ou espaço amostral
- Acontecimentos
  - Diagramas de Venn para representar acontecimentos
  - Diagrama em árvore para representar acontecimentos
- Operações com acontecimentos
- Modelo de probabilidade
- Probabilidade frequencista ou empírica
- Probabilidade de Laplace ou laplaciana
- Regras para a probabilidade
- Propriedades da probabilidade

Todos os dias somos confrontados com situações, que nos conduzem a utilizar, intuitivamente, a noção de probabilidade:

- Dizemos que existe uma pequena probabilidade de ganhar o totoloto;
- Dizemos que existe uma grande probabilidade de não chover num dia em que não se avistam nuvens;
- O médico interroga-se sobre qual a probabilidade de um doente, tratado com um novo medicamento sobreviver;
- Se lançarmos uma moeda de euro, dizemos que temos uma probabilidade de 50% de sair a face nacional;
- Dizemos que existe uma probabilidade nula de encontrar uma pessoa com 3 metros de altura;

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.

Fenómeno aleatório

Fenómeno aleatório - Um fenómeno diz-se que é um fenómeno aleatório quando o resultado de cada realização é incerto, mas admite-se ser possível encontrar um padrão de comportamento, depois de muitas repetições (Graça Martins, M.E. http://wikiciencias.casadasciencias.org/wiki/index.php/Fenómeno_aleatório)

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 comportamento dos eleitores nas próximas eleições legislativas;
- O comportamento de um aluno no exame de resposta múltipla, para o qual não estudou;
- O comprimento do próximo bebé a nascer no hospital Bom Sucesso;
- etc

Fenómeno aleatório

A probabilidade desenvolveu-se admitindo que 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 exatidão, qual o resultado que se vai obter quando se realizam.

Por oposição a fenómeno aleatório temos o **fenómeno determinista**. Por exemplo, o fenómeno que consiste em largar uma pedra que temos na mão e ver se "a pedra cai" não é um fenómeno aleatório, já que o resultado da realização do fenómeno não é incerto – a pedra cai (as leis da Física até permitem saber quanto tempo leva a chegar ao chão http://wikiciencias.casadasciencias.org/wiki/index.php/Fenómeno_aleatório)

À realização de um fenómeno aleatório, ou seja, ao processo de observar um dos seus resultados, chamamos **experiência aleatória**.

Experiência aleatória

O objectivo da Teoria da Probabilidade é o estudo de **fenómenos aleatórios**, a partir de modelos matemáticos, a que chamamos ***modelos de probabilidade***, que os possam descrever convenientemente. Para formalizarmos a noção de probabilidade, vamos introduzir alguma terminologia própria, sendo a noção mais básica a de experiência aleatória.

Experiência aleatória - é a realização de um fenómeno aleatório, ou seja, é o processo de observar um resultado do fenómeno.

Se, por exemplo, estivermos interessados em estudar o fenómeno aleatório "comportamento de uma moeda de euro ao ser lançada", realizamos a experiência aleatória que consiste em lançar a moeda e ver a face que fica voltada para cima. Neste caso admitimos que os resultados possíveis são face Euro e face Nacional.

Experiência aleatória

Numa experiência aleatória:

- obtém-se um *resultado*, de entre um conjunto de resultados, que admitimos como conceptualmente possíveis, conhecidos de antemão, a que se dá o nome de **espaço (ou universo) de resultados** ou **espaço-amostra** (ou amostral), mas *não se tem conhecimento exacto* de qual o resultado que sai em cada realização da experiência;
- *admite-se* que a experiência se pode *repetir* e que as repetições são realizadas nas mesmas circunstâncias e não se influenciam umas às outras, verificando-se uma regularidade estatística, isto é, é possível encontrar números entre 0 e 1, que representam a frequência relativa com que se verificam os resultados individuais, num grande número de realizações, independentes, da experiência.

(Graça Martins, M. E. http://wikiciencias.casadasciencias.org/wiki/index.php/Experiência_aleatória)

Experiência aleatória

Esta definição de experiência aleatória, em que se admite que a experiência se pode repetir o número de vezes que se quiser, independentemente umas das outras e sempre nas mesmas circunstâncias, apresentando uma *regularidade estatística*, prepara-nos para a definição de *probabilidade*, segundo *a teoria frequencista*, como veremos mais à frente.

Experiência aleatória

São exemplos de **experiências aleatórias** (Graça Martins, M.E. 2011, página 196):

- 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.

Probabilidade **Experiência aleatória**

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.

Espaço de resultados

Espaço de resultados ou **espaço amostral** – conjunto dos resultados que consideramos como possíveis, quando se realiza a experiência aleatória.

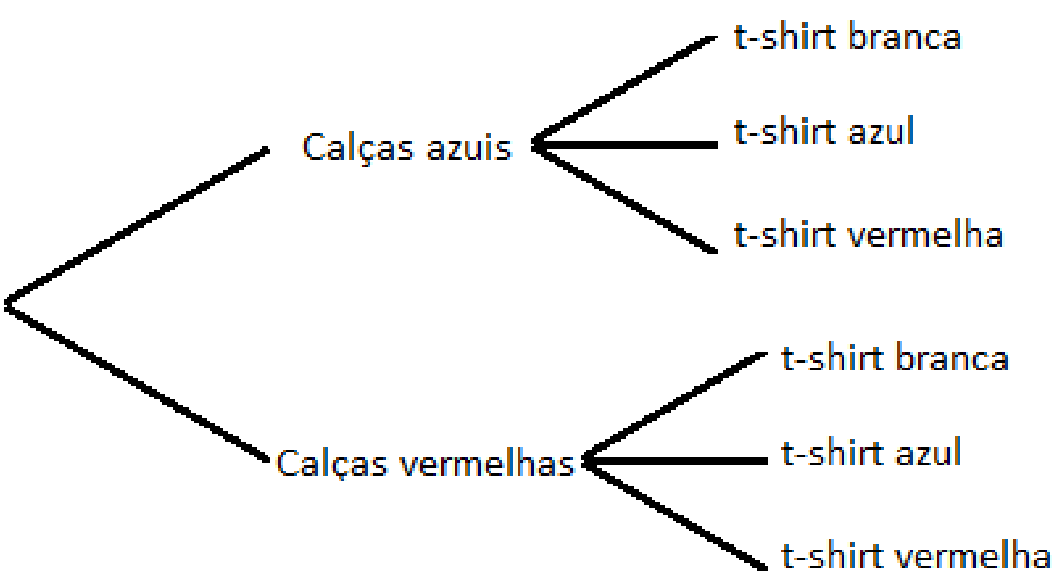
Para as experiências aleatórias consideradas anteriormente temos

Experiência aleatória	Espaço de resultados
....nº de pessoas do agregado familiar	{1, 2, 3, 4, ...}
... animal doméstico preferido	{cão, gato, peixe, passarinho, tartaruga, coelho, hámster, rato, tartaruga, ...}
... resultado que sai quando lança a moeda de 1 euro	{face Euro, face Nacional}
... nº de faces Euro em 10 lançamentos da moeda	{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
... tempo de casa à escola (em minutos)	{1, 2, 3, 4, 5, 6, 7, ... }
... número de mensagens enviadas	{0, 1, 2, 3, 4, 5, 6, 7, ... }
... quantos rapazes em 3 alunos	{0, 1, 2, 3}
... ver se chove	{chove, não chove}
... nº carros encarnados em 5 minutos	{0, 1, 2, 3, 4, ...}

Probabilidade Espaço de resultados Diagrama em árvore

Quando a experiência aleatória implica uma sucessão de passos, um processo especialmente adequado para representar os resultados do espaço de resultados é o **diagrama** ou **esquema em árvore**.

Exemplo – A Mariana está indecisa na escolha de qual a "toilete" que vai levar para a escola. Tem como opção de escolha calças de ganga azul ou vermelha e t-shirts branca, azul ou vermelha. Quantas "toiletes" pode a Mariana escolher?

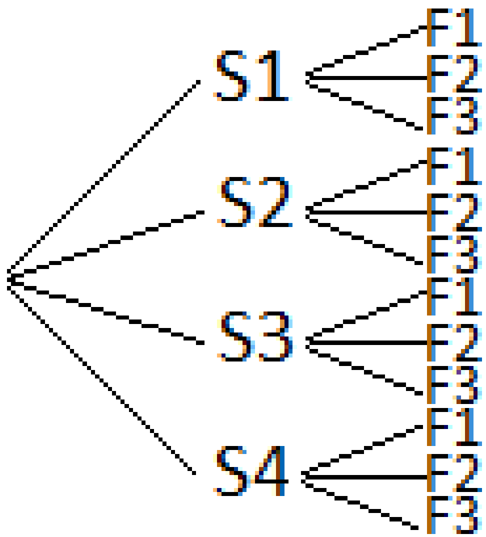


A Mariana pode escolher 6 “toiletes”, ou seja, o espaço de resultados é constituído por 6 resultados, que são:

S={Calças azuis,t-shirt branca; Calças azuis,t-shirt azul; Calças azuis,t-shirt vermelha; Calças vermelhas,t-shirt branca; Calças vermelhas,t-shirt azul; Calças vermelhas,t-shirt vermelha}

Exemplo – No bar da escola há 4 tipos de sanduiche e 3 tipos de fruta. Um aluno vai ao bar e escolhe ao acaso uma sanduiche e uma fruta para o lanche. Que lanches constituídos por uma sanduiche e uma peça de fruta pode calhar ao aluno?

Se representarmos os 4 tipos de sanduiche por S1, S2, S3 e S4 e os 3 tipos de fruta por F1, F2 e F3, temos para o espaço de resultados:



S={(S1,F1); (S1,F2); (S1,F3); (S2,F1); (S2,F2); (S2,F3); (S3,F1); (S3,F2); (S3,F3); (S4,F1); (S4,F2); (S4,F3)}

O espaço de resultados é constituído por 12 resultados possíveis.

Uma alternativa ao diagrama em árvore é a **tabela de dupla entrada**, que vamos utilizar para obter o espaço de resultados considerado no exemplo anterior.

Utilizando a mesma notação para as os 4 tipos de sanduiche e os 3 tipos de fruta, temos a seguinte tabela:

Sanduiche \ Fruta	F1	F2	F3
S1	(S1,F1)	(S1,F2)	(S1,F3)
S2	(S2,F1)	(S2,F2)	(S2,F3)
S3	(S3,F1)	(S3,F2)	(S3,F3)
S4	(S4,F1)	(S4,F2)	(S4,F3)

Espaço de resultados

S={S1,F1; S1,F2; S1,F3; S2,F1; S2,F2; S2,F3; S3,F1; S3,F2; S3,F3; S4,F1; S4,F2; S4,F3}

Acontecimento – é um subconjunto do espaço de resultados.

Acontecimento elementar - é o acontecimento constituído por um único resultado do espaço de resultados.

Acontecimento composto - é o acontecimento constituído por mais do que um resultado do espaço de resultados.

Os acontecimentos representam-se, de um modo geral, por letras maiúsculas A, B, C, D, ...

Os elementos de um acontecimento desigam-se por **casos favoráveis** a esse acontecimento. Diz-se que o acontecimento A **se realizou** ou **ocorreu** se o resultado da experiência aleatória pertence a A.

O espaço de resultados é um acontecimento a que se dá o nome de **acontecimento certo**.

O conjunto vazio, sem elementos, também se considera um acontecimento a que se dá o nome de **acontecimento impossível** e, tal como o conjunto vazio, representa-se por Ø.

Exemplo - Considerando a experiência aleatória que consiste em perguntar a uma pessoa, escolhida ao acaso, quantas pessoas constituem o seu agregado familiar, o espaço de resultados, que vamos representar por S, é constituído por todos os números inteiros não negativos (excluindo o zero):

$$S=\{1, 2, 3, \ldots\}$$

Alguns acontecimentos são:

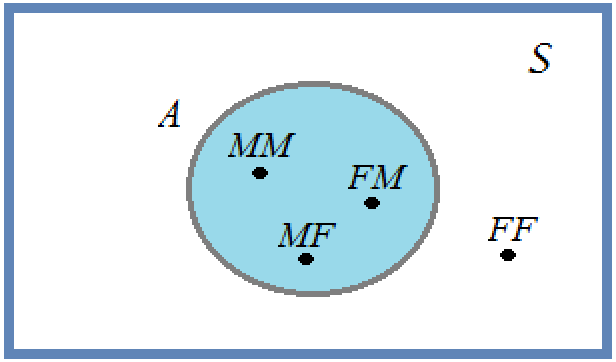
- 3 pessoas, que podemos representar por A={3}
- entre 2 e 4 pessoas (inclusive) que representamos por B={2, 3, 4)
- mais de 3 pessoas que representamos por C={4, 5, 6, …}
- menos de 10 pessoas que representamos por D={1, 2, …, 9}

Se a resposta à questão colocada, sobre a dimensão do agregado familiar, for 3, os acontecimetos A e B realizar-se-ão.

Para visualizar o espaço de resultados e os acontecimentos associados à realização de um fenómeno aleatório utilizam-se diagramas de Venn. Para representar o espaço de resultados S utiliza-se um rectângulo e no seu interior regiões fechadas, normalmente círculos, para representar os acontecimentos A, B, ….

Exemplo (http://wikiciencias.casadasciencias.org/wiki/index.php/Diagrama_de_Venn) - Consideremos a experiência aleatória que consiste em verificar o sexo dos filhos das famílias com 2 filhos. O espaço de resultados é constituído pelos resultados S ={MM,MF,FM,FF}.

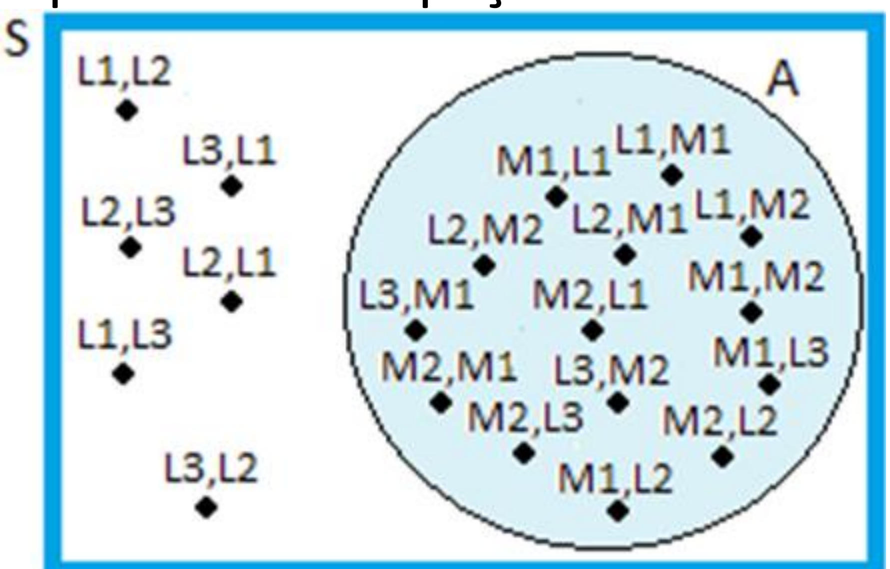
Seja *A* o acontecimento "Pelo menos um dos filhos é do sexo masculino". Representando num diagrama de Venn, temos



Exemplo – Consideremos a experiência aleatória que consiste em retirar 2 rebuçados de uma caixa com 5 rebuçados, em que 2 têm sabor a morango e os outros sabor a laranja. Representar através de um diagrama de Venn o espaço de resultados e o acontecimento A = {Pelo menos um dos rebuçados tem sabor a morango}.

Representando por L1, L2, L3, M1, M2 os 5 rebuçados, tem-se o seguinte diagrama de Venn a representar o espaço de resultados S,

constituído por 20 resultados e o acontecimento A constituído por 14 resultados.

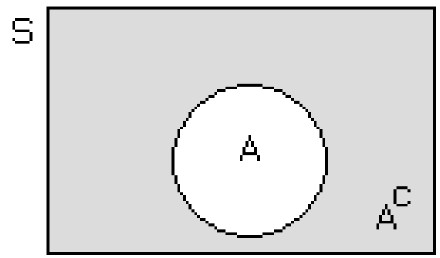


Probabilidade Operações com acontecimentos

Vamos utilizar os diagramas de Venn para apresentar a terminologia utilizada quando falamos de acontecimentos. Assim, representando o espaço de resultados por S e os acontecimentos por A, B, C, ..., associados a S, temos:

- Acontecimento **complementar** (ou **contrário**) do acontecimento A

É o acontecimento constituído por todos os resultados de S, que não estão em A e representa-se por $\bar{\mathbf{A}}$ ou $\mathbf{A}^C$.



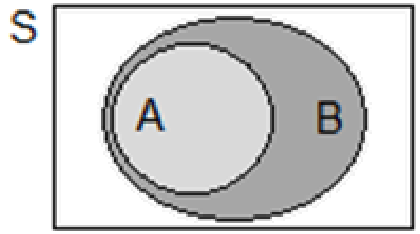
Quando um acontecimento se realiza, o seu complementar não se pode realizar.

(Ver http://www.alea.pt/html/probabil/html/cap_02/html/cap2_1_6.html)

Operações com acontecimentos

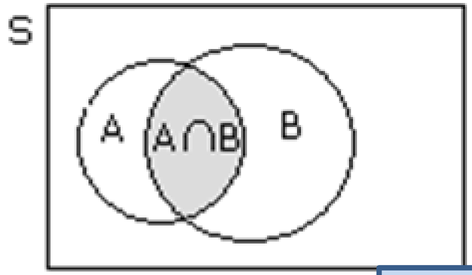
- Acontecimento A **implica** acontecimento B

O acontecimento A **implica** a realização do acontecimento B, quando todo o resultado de A é um resultado de B; indica-se este facto escrevendo **A⊂B**.



- Acontecimento **intersecção** entre os acontecimentos A e B

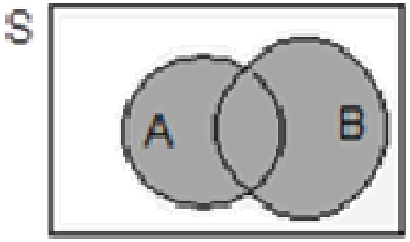
É o acontecimento que se realiza se e só se A e B se realizam simultaneamente, sendo constituído pelos resultados comuns a A e B. Representa-se por A∩B ou (AeB)



Operações com acontecimentos

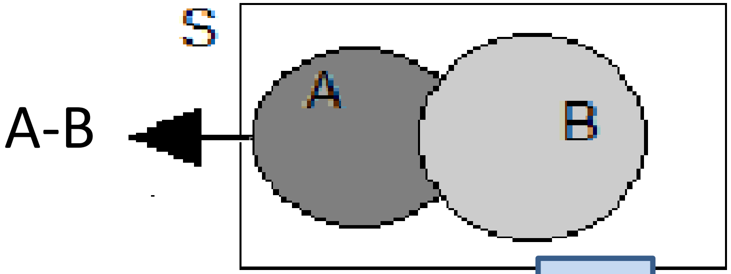
- Acontecimento **união** entre os acontecimentos A e B

É o acontecimento que se realiza se e só se A ou B se realizam, sendo constituído pelos resultados que pertencem a pelo menos um dos acontecimentos A ou B. Representa-se por $A \cup B$ ou (AouB)



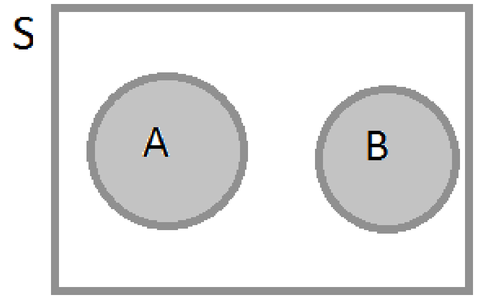
- Acontecimento **diferença** entre os acontecimentos A e B

É o acontecimento que se realiza se e só se A se realiza sem que B se realize, sendo constituído pelos resultados de A que não pertencem a B. Representa-se por A-B



- Acontecimentos **disjuntos** ou **mutuamente exclusivos** ou **incompatíveis**

São acontecimentos que não têm resultados comuns, pelo que a realização de um deles implica a não realização do outro



O acontecimento **impossível** é o que resulta da intersecção de acontecimentos disjuntos. De forma análoga ao que se passa na teoria dos conjuntos representa-se por Ø, símbolo do conjunto vazio, mas aqui lê-se acontecimento impossível e não acontecimento vazio.

Exemplo – Considere a experiência aleatória que consiste em lançar um dado de 6 faces e em verificar o número de pintas da face que fica virada para cima. Representar com a notação que achar conveniente:

a) O espaço de resultados;

b) O acontecimento A, que consiste em sair uma face par (com um número de pintas par);

c) O acontecimento B que consiste em sair uma face ímpar;

d) O acontecimento C que consiste em sair uma face com um número de pintas menor que 3;

e) O acontecimento D intersecção de A com B. O que conclui acerca dos acontecimentos A e B?

f) O acontecimento E união de A com B. O que conclui acerca dessa união?

g) O acontecimento F intersecção de B e C

(Resolução)

Operações com acontecimentos

Resolução

a) Espaço de resultados S ={1, 2, 3, 4, 5, 6};

b) A = {2, 4, 6};

c) B = {1, 3, 5};

d) C= {1, 2};

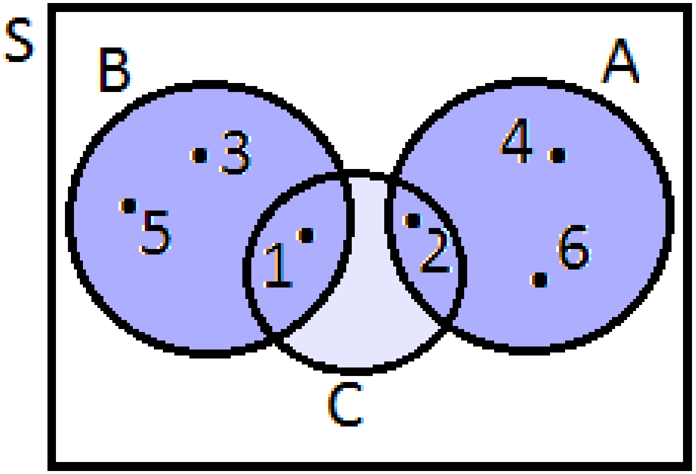
e) $D=A\cap B=\emptyset$, acontecimento impossível;

f) $E = A\cup B=$ {1, 2, 3, 4, 5, 6}

Os acontecimentos A e B dizem-se complementares, pois são disjuntos e a sua união é o espaço de resultados;

g) $F=B\cap C=$ {1}

Os acontecimentos anteriores podem ter a seguinte representação em diagrama de Venn



Exemplo - Considere a experiência aleatória que consiste em retirar dois cartões de uma caixa com 4 cartões (fisicamente iguais) numerados de 1 a 4 e em verificar os números aí inscritos.

Considere as seguintes situações:

i) Assim que retira o primeiro cartão, ele é reposto na caixa, antes de retirar o segundo cartão;

ii) Depois de retirar o primeiro cartão, retira o segundo sem repor o primeiro (Equivalente a retirar os dois cartões simultaneamente).

a) Qual o espaço de resultados associado à experiência anterior?

b) Considere os acontecimentos A – um dos cartões (e só um) tem um número par; B – os dois cartões têm número par; C – pelo menos um dos cartões tem número par; D – os dois cartões têm o mesmo número

(Resolução)

Resolução

i) *Extracção com reposição*

Para obter o espaço de resultados considere-se a seguinte tabela de dupla entrada

1ºcartão \ 2ºcartão	1	2	3	4
1	(1,1)	(1,2)	(1,3)	(1,4)
2	(2,1)	(2,2)	(2,3)	(2,4)
3	(3,1)	(3,2)	(3,3)	(3,4)
4	(4,1)	(4,2)	(4,3)	(4,4)

a) S={(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)}

b) A= {(1,2), (1,4), (2,1), (2,3), (3,2), (3,4), (4,1), (4,3)}

B= {(2,2), (2,4), (4,2), (4,4)}

C={(1,2),(1,4),(2,1),(2,2) ,(2,3),(2,4), (3,2), (3,4), (4,1), (4,2), (4,3), (4,4)}

D={(1,1), (2,2), (3,3), (4,4)}

ii) *Extracção sem reposição*

Na tabela anterior os resultados a sombreado não se podem verificar

a) S={(1,2), (1,3), (1,4), (2,1), (2,3), (2,4), (3,1), (3,2), (3,4), (4,1), (4,2), (4,3)}

b) A= {(1,2), (1,4), (2,1), (2,3), (3,2), (3,4), (4,1), (4,3)}

B= {(2,4), (4,2)}

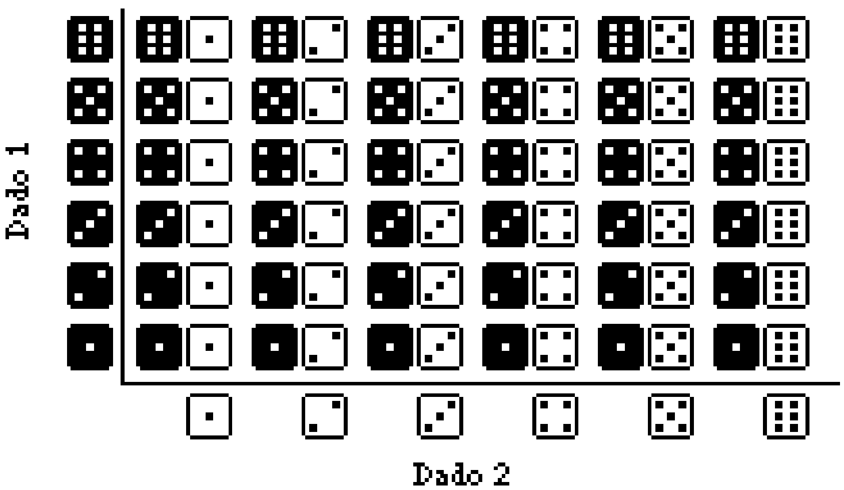
C={(1,2), (1,4), (2,1),(2,3),(2,4), (3,2), (3,4), (4,1), (4,2), (4,3)}

D={ }=Ø

Probabilidade **Operações com acontecimentos**

Exemplo (Graça Martins, M.E. 2005, página 131) **–** Considere a experiência aleatória que consiste no lançamento de 2 dados e em verificar o número de pintas das faces que ficam viradas para cima. Identificar o espaço de resultados e os acontecimentos A e B, respectivamente,"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 preto e um branco, para os distinguir. O espaço de resultados é constituído por todos os pares de dados considerados na figura a seguir



S={(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)}

A= {(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)}

B= {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}

(Continua)

Nota 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 S 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.

Nota 2 - Associado à experiência que acabámos de descrever no exemplo anterior, poderíamos ter considerado o seguinte espaço de resultados:

S = { 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 a regra de Laplace, para atribuir probabilidades a acontecimentos associados ao fenómeno em estudo.

Modelo de probabilidade

Como dissemos anteriormente o nosso objectivo é definir modelos de probabilidade para fenómenos aleatórios que estejamos interessados em estudar. Vamos admitir que estes fenómenos aleatórios têm espaços de resultados finitos. Assim, em espaços finitos, a definição de um **modelo de probabilidade** para um fenómeno aleatório implica:

- A identificação de um espaço de resultados $S=\{s_1, s_2, ..., s_n\}$;
- Uma forma de atribuir um número a cada um dos resultados, isto é, a cada acontecimento elementar, a que chamaremos **probabilidade**.

Acontecimento	$\{s_1\}$	$\{s_2\}$	$\{s_3\}$	...	$\{s_n\}$
Probabilidade	p_1	p_2	p_3		p_n

Probabilidade Modelo de probabilidade

O processo de atribuir probabilidades deve ser tal, que algumas regras básicas devam ser satisfeitas para todos os modelos. Vamos então considerar as seguintes regras, que são intuitivas:

- **Regra 1** – Uma probabilidade deve ser um número entre 0 e 1 (ou 100%).
- **Regra 2** - A soma das probabilidades dos acontecimentos elementares que compõem o espaço de resultados é igual a 1 (ou 100%).

NOTA As regras anteriores não excluem 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), página 22).
(http://wikiciencias.casadasciencias.org/wiki/index.php/Modelo_de_probabilidade)

Modelo de probabilidade

Nota Nem sempre a definição do espaço de resultados associado a uma experiência aleatória é uma tarefa simples. Vejamos o seguinte exemplo.

Exemplo (adaptado de Graça Martins, M.E. 2005, página 128) Lançamento de uma moeda - Admita que tem uma moeda de euro equilibrada. Mas o que é uma moeda equilibrada? É aquela em que estamos a admitir, à partida, que existe igual possibilidade de sair face euro ou face nacional no próximo lançamento que façamos com ela – estamos a admitir o *princípio da simetria*, isto é, estamos a admitir, na nossa cabeça, um *modelo matemático* em que assumimos que em qualquer lançamento da moeda, existe igual possibilidade de sair a face euro ou 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, considerando 3 resultados possíveis para o espaço de resultados! 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. (continua)

Nota (continuação)

É neste sentido que deve ser interpretada a frase do estatístico George Box, que 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, ou seja, que em, aproximadamente, 50% dos lançamentos se observe a face euro.

Se, por outro lado, recolhermos uma amostra de dimensão 1, isto é, 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.

Probabilidade de um acontecimento

Admitamos, para já, que tínhamos um processo de definir um modelo de probabilidade (veremos a seguir processos de atribuir probabilidades aos acontecimentos elementares). Uma vez definido esse modelo de probabilidade, como obter a probabilidade de acontecimentos?

Uma vez que um acontecimento é um conjunto de resultados, vamos definir a probabilidade do acontecimento A, que representamos por P(A), à custa das probabilidades dos resultados que compõem A:

Probabilidade de um acontecimento A - Em espaços finitos, a probabilidade de um acontecimento A é a soma das probabilidades dos acontecimentos elementares que compõem A e representa-se por P(A).

Probabilidade de um acontecimento

De acordo com o que ficou dito anteriormente, se tivermos processos de atribuir probabilidades aos acontecimentos elementares, então temos o problema de obter a probabilidade de um acontecimento resolvido!

Veremos a seguir duas formas de atribuir probabilidades aos acontecimentos elementares:

- Probabilidade frequencista ou empírica
- Probabilidade laplaciana ou de Laplace

Probabilidade frequencista ou empírica

A definição de experiência aleatória, segundo a qual se *admite* que a experiência se pode repetir o número de vezes que se quiser, independentemente umas das outras e sempre nas mesmas circunstâncias, apresentando uma regularidade estatística, prepara-nos para a definição de probabilidade segundo o *conceito ou interpretação frequencista de probabilidade.*

Consideremos de novo o ficheiro Dados sobre casas, resultado do registo das características de 40 casas de uma dada região e suponhamos que nessa região se pretendia recolher informação sobre uma outra casa, escolhida ao acaso (Graça Martins, M.E. et al, 2007, página 155). Algumas questões que se podem colocar sobre essa outra casa são as seguintes:

- Será mais provável que essa casa seja nova ou usada?
- Qual será um valor aproximado para a probabilidade de a casa ser usada?

(continua)

Uma vez que das 40 casas, 31 são usadas e 9 são novas, é natural esperar que seja mais provável que esta outra casa seja usada. Por outro lado, esperamos que a probabilidade de, na dita região, encontrar à venda uma casa usada, esteja próxima de 80%, já que a frequência relativa obtida para o acontecimento “Casa usada” foi 77,5%.

Neste exemplo, a experiência consiste em seleccionar uma casa ao acaso e em verificar se é usada ou nova. Existem dois casos possíveis para o estado da casa e é por essa razão que o resultado da experiência é aleatório: antes de verificar a casa, não temos informação suficiente para saber qual dos resultados se vai verificar, se é usada ou nova.

À medida que vamos repetindo a experiência aleatória a frequência relativa do acontecimento elementar “Casa usada” tende a fixar-se num valor que estará próximo de 0,80.

Probabilidade empírica ou frequencista - de um acontecimento A, representada por P(A), é o valor obtido para a frequência relativa da realização de A, num grande número de repetições da experiência aleatória, nas mesmas condições ou em condições idênticas. À medida que o número de repetições da experiência aleatória aumenta, a frequência relativa com que se realiza A tende a *estabilizar* para um valor entre 0 e 1. Este valor é *interpretado* como sendo a probabilidade do acontecimento A se realizar.

Probabilidade frequencista ou empírica

Considerando que o espaço de resultados associado ao fenómeno aleatório que consiste em averiguar o estado da casa é constituído pelos resultados "nova" e "usada", um modelo de probabilidade que tem sentido considerar é o seguinte:

Estado da casa	Nova	Usada
Probabilidade	0,20 ou 20%	0,80 ou 80%

Representamos por Nova e Usada os acontecimentos elementares.

P(Nova) = 0,20

P(Usada) = 0,80

A probabilidade pode ser descrita na forma de fracção, decimal ou percentagem.

Atenção

A *regularidade estatística* que admitimos verificar-se quando se realiza a experiência muitas vezes, 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 probabilidade de sair. Os sucessivos lançamentos são independentes.

Probabilidade frequencista ou empírica

Exemplo - Consideremos a seguinte tabela que resultou de averiguar junto de 50 alunos de uma escola, qual a sua fruta preferida:

Fruta preferida	Nº de alunos	Freq. Relativa (%)
Maçã	3	6
Pera	4	8
Cerejas	32	64
Uvas	5	10
Morangos	6	12
	50	100

Se se tivesse averiguado a preferência de mais um aluno:

- Qual a fruta que se espera que ele responda?
- Qual será um valor aproximado para a probabilidade de este aluno preferir as cerejas?

Da tabela verificamos que 64% dos alunos preferem Cerejas, pelo que esperamos que o aluno responda Cerejas e um valor aproximado para a probabilidade de um aluno, escolhido ao acaso, preferir Cerejas está próxima de 64%.

Probabilidade frequencista ou empírica

Exemplo – Suponha que lança um dado 1000 vezes e verifica a face que ficou voltada para cima, tendo obtido os seguintes resultados:

Face	Freq. abs.	Freq. rel.(%)
1	159	15.9%
2	163	16.3%
3	160	16.0%
4	161	16.1%
5	86	8.6%
6	271	27.1%

Perante os resultados da tabela somos levados a sugerir o seguinte modelo de probabilidade para o fenómeno aleatório que consiste em verificar qual a face que sai no lançamento de um dado:

Face	Probabilidade
1	16%
2	16%
3	16%
4	16%
5	9%
6	27%

O que este modelo indicia é que o dado não é equilibrado, pois as faces não têm todas a mesma probabilidade de saírem.

Probabilidade frequencista ou empírica

Exemplo – Suponha que o director da escola pretende averiguar qual a cor preferida dos alunos para pintar o pátio, pelo que decidiu proceder a uma sondagem. Foram questionados 152 alunos, cujas respostas foram resumidas na seguinte tabela:

Cor preferida	Azul	Branca	Verde	Amarela
Nºalunos	68	10	45	29

A partir dos resultados da tabela anterior foi considerado o seguinte modelo de probabilidade para o fenómeno aleatório em estudo "cor preferida dos alunos para pintar o pátio da escola" e foram calculadas as probabilidades de alguns acontecimentos

≈29/152

Cor preferida	Azul(1)	Branca	Verde	Amarela
Nºalunos	45%	6%	30%	19%

(1) "Azul" significa "Um aluno da escola, escolhido ao acaso, preferir a cor azul". Cada aluno só podia indicar uma cor.

P(Azul) = 45%

P(Azul ou Verde) = 0,45+0,30= 0,75 ou 75%

P(Azul e Verde) = 0 (O acontecimento Azul e Verde é impossível)

Nota - A frequência relativa ou proporção de vezes com que um acontecimento se realiza não é a probabilidade, mas sim um valor aproximado para a probabilidade. O que se pode dizer é que a frequência relativa com que um acontecimento se verifica na repetição de uma experiência aleatória aproxima-se da probabilidade (teórica) desse acontecimento, à medida que o número de repetições aumenta (Lei dos grandes números).

Nota – As sondagens feitas pelos partidos para averiguar as preferências do eleitorado têm por base a noção de probabilidade frequencista.

Probabilidade de Laplace ou laplaciana

Em algumas situações especiais, de espaços de resultados com **n** resultados, podemos considerar que os **n** acontecimentos elementares são igualmente possíveis - situação que denominaremos como *equiprobabilidade* ou *simetria*. Assim, a probabilidade de cada acontecimento elementar é **1/n**, já que segundo a regra 2, a soma das probabilidades dos **n** acontecimentos elementares é igual a **1**.

Definida a probabilidade de um acontecimento elementar, definimos a probabilidade de um acontecimento A , **P(A)**, como sendo a soma das probabilidades dos acontecimentos elementares que compõem A. Somos assim conduzidos à interpretação laplaciana ou de Laplace de Probabilidade, que se enuncia da seguinte forma:

Probabilidade laplaciana ou **de Laplace** - Dado o espaço de resultados S, constituído por um número finito n de elementos, todos eles igualmente possíveis, define-se probabilidade do acontecimento A, associado ao espaço de resultados S, e representa-se por P(A), como sendo a razão entre o número de resultados favoráveis a A (resultados que compõem A) e o número de resultados possíveis (resultados que compõem S)

$$P(A) = \frac{\text{número de resultados favoráveis a } A}{\text{número de resultados possíveis}}$$

Esta definição também é conhecida como **regra de Laplace** ou definição clássica de probabilidade.

Probabilidade de Laplace ou laplaciana

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.

Probabilidade de Laplace ou laplaciana

Outras situações em que pode ter sentido utilizar a probabilidade de Laplace é em estudos de fenómenos aleatórios associados a populações pequenas, por exemplo ao nível dos alunos da sala de aula, como se exemplificará mais à frente. No entanto, estas situações, muito simples de serem tratadas considerando a probabilidade laplaciana, são bastante restritivas.

Efectivamente, na vida real, quase nunca podemos assumir que os resultados do espaço de resultados associado a um fenómeno aleatório, que tenha interesse estudar, sejam equiprováveis.

Probabilidade de Laplace ou laplaciana

Por exemplo, consideremos os seguintes fenómenos aleatórios:

Fenómeno aleatório	Resultados possíveis	Observação
Estado de um parafuso, escolhido ao acaso, de um lote de parafusos	Não tem defeito, Tem defeito	A probabilidade de não ter defeito é muito inferior à de ter defeito
Número de acidentes verificados num determinado cruzamento, num dia escolhido ao acaso	0, 1, 2, 3, ...	Não tem sentido considerar que os acontecimentos elementares tenham igual probabilidade
Tamanho de uma camisola a ser fabricada por uma empresa textil	XS, S, M, L, XL	Existem alguns tamanhos que são os mais procurados, por exemplo o M e o L, pelo que a probabilidade de se-rem estes tamanhos a serem produ-zidos deve ser maior

Probabilidade de Laplace ou laplaciana

Exemplo – Para eleger o delegado de turma a professora pediu a cada um dos 24 alunos da turma que escrevesse o nome num bocado de papel (a professora distribuiu papéis iguais a todos os alunos) e o metesse dentro de uma caixa. Depois retiraria um papel ao acaso.

a) Qual a probabilidade de o Francisco ser eleito?

b) A probabilidade de ser eleito um rapaz é 1/3. Quantos rapazes tem a turma?

Resolução

a) A probabilidade da cada aluno ser eleito é 1/24, pelo que a probabilidade de ser o Francisco é 1/24.

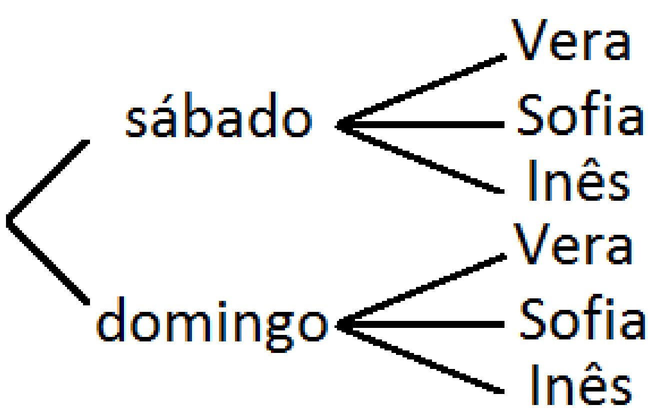
b) A probabilidade do acontecimento "ser rapaz", acontecimento composto por x casos favoráveis, é 1/3. Sabendo que o número de casos possíveis é 24, pela regra de Laplace vem

$\frac{x}{24} = \frac{1}{3}$ de onde x=8, ou seja, a turma tem 8 rapazes.

Probabilidade de Laplace ou laplaciana

Exemplo – A Maria está indecisa entre ir ao cinema no sábado ou no domingo. Também está na dúvida se convida a Vera, a Sofia ou a Inês para a acompanhar. Para tomar a decisão relativamentre ao dia, lançou uma moeda de um euro ao ar e se saísse a face nacional iria no sábado. Para decidir qual a amiga a convidar lançou um dado de seis faces ao ar e se saísse 1 ou 2, convidaria a Vera, se saísse 3 ou 4, seria a Sofia e se fosse 5 ou 6 seria a Inês. Qual a probabilidade de a Maria ir ao cinema num sábado com a Vera?

Resolução



O espaço de resultados associado ao fenómeno aleatório em estudo é:

S={(sábado, Vera), (sábado, Sofia), (sábado, Inês), (domingo, Vera), (domingo, Sofia), (domingo, Inês)}

Da forma como foi tomada cada uma das opções, os acontecimentos elementares são equipro-váveis, pelo que a probabilidade de a Maria ir ao cinema no sábado com a Vera, é igual a 1/6.

Probabilidade de Laplace ou laplaciana

Exemplo – A família Feliz, composta por mãe, pai e 3 filhos, está a pensar fazer uma viagem, mas não se entendem sobre o destino da viagem, pois cada um quer dar o seu "palpite".... Então decidiram escolher ao acaso quem vai decidir onde ir. Só que o pai e a mãe têm o dobro das possibilidades de cada filho de serem escolhidos para tomar a decisão! Qual a probabilidade de irem viajar para o destino escolhido pela mãe?

Resolução

Uma vez que a mãe e o pai têm o dobro das possibilidades de serem escolhidos, é como se o espaço de resultados fosse constituído por 7 resultados, todos igualmente possíveis:

$$S=\{\text{mãe, mãe, pai, pai, filho1, filho2, filho3}\}$$

Assim, utilizando um processo de seleccionar ao acaso cada um dos resultados anteriores (por exemplo colocando numa caixa 7 bocados de papel com cada um dos resultados), a probabilidade de ser seleccionada a mãe é igual a 2/7.

Exemplo – A comissão de curso da licenciatura em Estatística é constituída pela Teresa, Isabel, Leonor, António e Francisco. Pretende-se seleccionar, ao acaso, 2 destes alunos para organizarem a festa final de ano. Qual a probabilidade de serem seleccionados 2 alunos de sexo diferente? E do mesmo sexo?

Resolução

Para obter o espaço de resultados pode considerar-se uma tabela de 2 entradas onde estão representados os resultados possíveis

	Teresa	**Isabel**	**Leonor**	**António**	**Francisco**
Teresa	-	(T,I)	(T,L)	(T,A)	(T,F)
Isabel	(I,T)	-	(I,L)	(I,A)	(I,F)
Leonor	(L,T)	(L,I)	-	(L,A)	(L,F)
António	(A,T)	(A,I)	(A,L)	-	(A,F)
Francisco	(F,T)	(F,I)	(F,L)	(F,A)	-

A probabilidade de obter alunos de sexo diferente é 12/20, ou seja 60%. A probabilidade de obter alunos do mesmo sexo é 8/20, ou seja 40%. Estes acontecimentos são complementares.

Probabilidade **Probabilidade de Laplace ou laplaciana**

Exemplo – Considere a experiência aleatória que consiste em lançar um dado 2 vezes e observar a soma do número de pintas das faces que ficam viradas para cima.

a) Qual o espaço de resultados associado à experiência anterior? Os resultados deste espaço de resultados são equiprováveis?

b) Qual a probabilidade de obter um número par? E um número ímpar? Qual destes acontecimentos é mais provável?

c) Qual a probabilidade de obter um número múltiplo de 2 e de 5 simultaneamente?

d) Se representar por A o acontecimento "obter um número maior ou igual a 3" e por B "obter um número menor ou igual a 3" $P(A \cup B)=P(A)+P(B)$?

(Resolução)

Probabilidade de Laplace ou laplaciana

Resolução

a) O espaço de resultados S é constituído pelos seguintes resultados

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

Os resultados não são equiprováveis. Basta pensar, por exemplo, que o 2 só pode ser obtido a partir do par de resultados (1,1), enquanto o 3 pode ser obtido a partir dos pares (1,2) e (2,1). Cada par representa os dois lançamentos do dado.

b) Para calcular as probabilidades pedidas vamos considerar um outro espaço de resultados, em que se consideram os resultados dos 2 lançamentos de um dado:

2º lançamento

1º lançamento

	1	**2**	**3**	**4**	**5**	**6**
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Para obter os resultados de S, basta somar os valores de cada par, obtendo-se uma nova tabela que se deduz da tabela ao lado:

Probabilidade de Laplace ou laplaciana

Resolução (continuação)

2º lançamento

1º lançamento	1	2	3	4	5	6
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

Temos 36 resultados todos igualmente prováveis, pelo que agora já podemos aplicar a probabilidade de Laplace:

$P(\{2\})=1/36$
$P(\{3\})=2/36$
$P(\{4\})=3/36$
$P(\{5\})=4/36$
$P(\{6\})=5/36$
$P(\{7\})=6/36$
$P(\{8\})=5/36$
$P(\{9\})=4/36$
$P(\{10\})=3/36$
$P(\{11\})=2/36$
$P(\{12\})=1/36$

Modelo de probabilidade para fenómeno aleatório em estudo “soma das pintas das faces que ficam viradas para cima quando se lança um dado 2 vezes”:

Acontecimento	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

Probabilidade Probabilidade de Laplace ou laplaciana

Resolução (continuação)

Uma vez o modelo de probabilidade construído, temos

$P(\text{par}) = P(\{2\} \text{ ou } \{4\} \text{ ou } \{6\} \text{ ou } \{8\} \text{ ou } \{10\} \text{ ou } \{12\})$

$= P(\{2\})+P(\{4\})+P(\{6\})+P(\{8\})+P(\{10\})+P(\{12\})$ (Prob. de um acontecimento)

$= 1/36+3/36+5/36+5/36+3/36+1/36$

$= 18/36$ (Ver Nota slide seguinte)

$= 1/2$

$P(\text{ímpar})=1/2$

Os dois acontecimentos são igualmente prováveis.

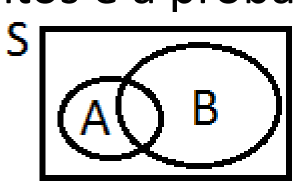
c) $P(\{10\})=3/36$

$=1/12$

d) $P(A)=35/36$

$P(B)=3/36$

Não é verdade que $P(A \cup B)=P(A)+P(B)$, pois os acontecimentos A e B não são disjuntos e a probabilidade de obter o 3 está a ser contabilizada 2 vezes



$P(A \cup B)=35/36+3/36-2/36$

$=36/36=1$ (Repare-se que $(A \cup B)=S$)

Probabilidade de Laplace ou laplaciana

Nota

Outra forma de calcular a probabilidade do acontecimento "par" é verificar quantos dos 36 resultados do espaço de resultados conduzem a uma soma par nas pintas do dado nos dois lançamentos:

2º lançamento

1º lançamento

	1	2	3	4	5	6
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

Como se verifica da tabela ao lado, dos 36 resultados possíveis, 18 são favoráveis à realização do acontecimento "par", pelo que pela probabilidade de Laplace

$$P(par)=1/2$$

Regras para a probabilidade

Dado um espaço de resultados S, com um número finito de elementos, tanto a probabilidade frequencista, como a probabilidade laplaciana satisfazem o seguinte conjunto de regras ou **axiomas**:

A1 $P(A) \geq 0$, qualquer que seja o acontecimento A associado a S.

A2 $P(S)=1$.

A3 Se os acontecimentos A e B, associados a S, forem disjuntos, então $P(A \cup B)=P(A)+P(B)$.

Probabilidade frequencista Probabilidade laplaciana

A probabilidade frequencista satisfaz os axiomas da Probabilidade

1. Dado um acontecimento A, como P(A) é o valor para que estabiliza a frequência relativa com que A ocorreu nas sucessivas repetições da experiência aleatória, uma vez que a frequência relativa é não negativa, o mesmo acontece a P(A), pelo que A1 se verifica.
2. Em cada realização da experência S ocorreu sempre, já que a ocorrência de qualquer resultado de S implica a realização de S, pelo que S é um acontecimento certo, P(S)=1, verificando-se A2.
3. Dados os acontecimentos A e B, disjuntos, quando se realiza a experiência se $(A \cup B)$ ocorre é porque o resultado da experiência pertence a A ou a B e não pode pertencer aos dois acontecimentos ao mesmo tempo. Assim, o número de vezes que $(A \cup B)$ ocorre é igual ao número de vezes que A ocorre mais o número de vezes que B ocorre, pelo que o valor para que estabiliza a frequência relativa da ocorrência de $(A \cup B)$ é igual à soma dos valores para que estabilizam as frequências relativas da ocorrência de A e de B, como se pretendia mostrar.

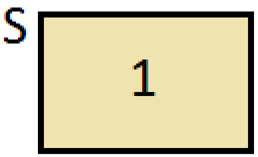
A probabilidade laplaciana satisfaz os axiomas da Probabilidade

1. Se representarmos por *n* o número de resultados possíveis, ou seja, o número de elementos de S, dado um acontecimento A, a probabilidade de A, P(A), é o quociente entre o número n_A de resultados favoráveis a A, ou seja, o número de elementos de A, e n. Como $n_A \geq 0$, vem $P(A)=\frac{n_A}{n} \geq 0$.
2. Como o número de resultados favoráveis a S é n, temos que $P(S)=\frac{n}{n}=1$.
3. Se A e B forem disjuntos, o número de resultados favoráveis a $(A \cup B)$ é igual ao número de resultados favoráveis a A, mais o número de resultados favoráveis a B. Assim,
4. $P(A \cup B)=\frac{n_{(A \cup B)}}{n}=\frac{n_A+nB}{n}=\frac{n_A}{n}+\frac{n_B}{n}=P(A)+P(B)$

Propriedades da probabilidade

Com a ajuda dos diagramas de Venn e tendo em conta as regras anteriores, podem-se deduzir as seguintes propriedades para a probabilidade de acontecimentos de um mesmo espaço de resultados S.

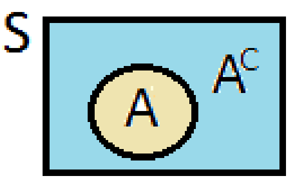
1. A probabilidade do acontecimento impossível é igual a 0.



$P(S \cup \emptyset)=P(S)+P(\emptyset)$ pela regra 3

$1=1+ P(\emptyset)$ pela regra 2

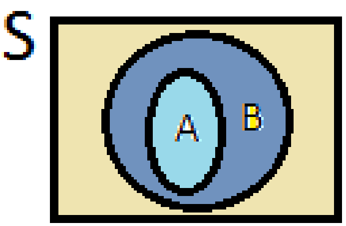
2. A probabilidade do acontecimento A^C, complementar do acontecimento A, é igual a $P(A^C)=1-P(A)$



$P(A \cup A^c)=P(A)+P(A^c)$ pela regra 3

$P(S)=1= P(A)+P(A^c)$ pela regra 2

3. Dados os acontecimentos A e B, se A implica B, $A \subset B$ então $P(A) \leq P(B)$

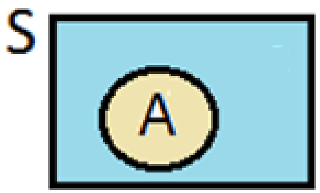


$P(B)=P(A)+P(B-A)$ pela regra 3

$P(B) \geq P(A)$ pois $P(B-A) \geq 0$ pela regra 2

Propriedades da probabilidade

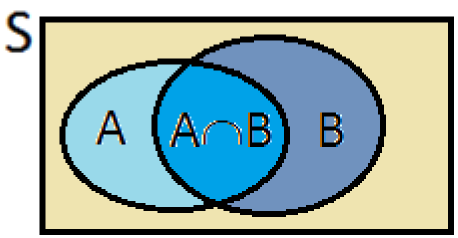
4. Qualquer que seja o acontecimento A, $0 \leq P(A) \leq 1$



Como $A \subset S$, pela regra 1 e pela propriedade 3 vem que $0 \leq P(A) \leq 1$

5. Quaisquer que sejam os acontecimentos A e B, tem-se

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



$P(A \cup B) = P(A-B) + P(A \cap B) + P(B-A)$ pela regra 3

$P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

1. Os alunos da turma distribuem-se por sexo e idade de acordo com a seguinte tabela:

Sexo \ Idade	**11**	**12**	**13**	**Total**
Feminino	4	5	3	12
Masculino	3	6	4	13
Total	7	11	7	25

O professor tem os nomes dos alunos inscritos em fichas todas iguais ao tacto, mas de cor diferente – rosa para o sexo feminino e azul para o sexo masculino. Sem olhar, selecciona uma ficha ao acaso:

a) Qual a probabiliadde de ser uma ficha de um rapaz?
b) Qual a probabilidade de ser uma ficha de uma rapariga?
c) Os acontecimentos "ser rapaz" e "ser rapariga" são complementares? Porquê?
d) O professor selecciona uma ficha azul. Qual a probabilidade de ser de um aluno de 11 anos? (Resolução)

1. Resolução

a) Como nos 25 alunos há 13 rapazes, há 13 casos favoráveis de entre 25 possíveis, pelo que a probabilidade do acontecimento “ser rapaz” é 13/25.

b) A probabilidade de “ser rapariga” é 12/25.

c) Quando se selecciona uma ficha não se podem verificar simultaneamente os acontecimentos “ser rapaz” e “ser rapariga”, mas um deles tem de se verificar necessariamente, pelo que são complementares.

d) Uma vez que se sabe que a ficha seleccionada é azul, agora os casos possíveis são 13, enquanto que os casos favoráveis são 3. Assim, a probabilidade pretendida é 3/13.

2. Para angariarem fundos para a viagem de finalistas do 3º ciclo, os alunos do 9º ano decidiram fazer rifas para venderem aos pais.

Num total de 250 rifas, 50 rifas têm prémio.

a) Qual a probabilidade da primeira rifa a ser vendida ter prémio?

b) Sabe-se que o pai do João foi o primeiro a comprar uma rifa e teve um prémio. Qual a probabilidade de a segunda rifa a ser vendida ter prémio?

c) O resultado obtido em b) é igual ao resultado obtido em a)? Estavas à espera que fossem iguais ou diferentes? Porquê?

(Resolução)

3. O João está a lançar uma moeda de 1 euro ao ar, equilibrada, e nos 6 primeiros lançamentos obteve

Euro, Nacional, Nacional, Nacional, Nacional, Nacional

No próximo lançamento que fizer é mais provável obter a face Euro ou a face Nacional? Porquê? (Resolução)

Exemplos

2. Resolução

a) Uma vez que nas 250 rifas, 50 têm prémio, há 50 casos favoráveis de entre 250 possíveis, pelo que a probabilidade pretendida é 50/250=0,20 ou 20%

b) Como o pai do João comprou uma rifa que se sabe ter prémio, neste momento temos 249 rifas, das quais 49 têm prémio. Assim, a probabilidade da próxima rifa a ser vendida ter prémio é 49/249.

c) Os resultados não são iguais, pois o número de casos possíveis vem diminuído de uma unidade, assim como o número de casos favoráveis e quando se tem uma fracção $\frac{a}{b} \neq \frac{a-1}{b-1}$

3. Resolução

A probabilidade de obter a face Euro é igual à probabilidade de obter a face Nacional e igual a ½, independentemente do que se passou. A moeda não tem memória...

4. A Ana, a Beatriz e a Carolina decidem ir ao cinema. São-lhes vendidos 3 bilhetes para 3 lugares seguidos, que elas distribuem, ao acaso, entre si. Qual a probabilidade de a Ana e a Beatriz ficarem lado a lado? (Resolução)

5. Numa caixa estão 5 rebuçados, 3 com sabor a laranja e 2 com sabor a morango. O João tira dois rebuçados, um com cada mão. Qual a probabilidade de retirar um rebuçado com sabor a morango com a mão esquerda? (Resolução)

6. Numa caixa de Smarties, as drageias têm várias cores, distribuídas de acordo com a seguinte tabela:

Cor	Vermelha	Azul	Amarela	Verde	Roxa
Probabilidade	0,25	0,30	0,20	0,15	r

a) Qual a probabilidade de retirar uma drageia que não tenha cor roxa?

b) Qual a probabilidade de retirar uma drageia de cor roxa?

c) Qual a probabilidade de retirar uma drageia de cor vermelha ou azul?

d) Qual a probabilidade de retirar uma drageia de cor vermelha e roxa?
(Resolução)

4. Resolução Representando as 3 amigas por A, B e C, elas podem ficar colocadas nas seguintes posições em que se assinalou a cor os casos favoráveis. Assim, sendo 4 casos favoráveis de entre 6 possíveis, a probabilidade pretendida é 4/6 ou seja, 2/3.

A	B	C		B	C	A
A	C	B		C	A	B
B	A	C		C	B	A

5. Resolução Representando os rebuçados por L1, L2, L3, M1 e M2, o espaço de resultados associado à experiência aleatória que consiste em seleccionar 2 rebuçados pode ser obtida a partir da seguinte tabela de dupla entrada:

Rebuçado mão direita

Rebuçado mão esquerda

	L1	L2	L3	M1	M2
L1		L1, L2	L1, L3	L1, M1	L1, M2
L2	L2, L1		L2, L3	L2, M1	L2, M2
L3	L3, L1	L3, L2		L3, M1	L3, M2
M1	M1, L1	M1, L2	M1, L3		M1, M2
M2	M2, L1	M2, L2	M2, L3	M2, M1	

Dos 20 resultados possíveis que constituem o espaço de resultados, 8 são favoráveis (assinalados a azul). Assim, a probabilidade pretendida é 8/20 ou seja 40%.

6. Resolução

a) A probabilidade de seleccionar uma drageia que não tenha cor roxa é

0,25+0,30+0,20+0,15=0,90 ou 90%

b) A probabilidade de seleccionar uma drageia de cor roxa é igual a

1-0,90=0,10 ou 10%

c) P("vermelha" ou "azul")=P("vermelha")+P("azul")

= 0,25+0,30

= 0,55 ou 55%

d) O acontecimento ("vermelho" e "azul") é impossível, pois não se pode retirar ao mesmo tempo uma drageia vermelha e azul. Assim,

P("vermelha" e "azul")=0, pois P(Ø)=0.

1. Serão os jogos justos (ou equilibrados)? (Adaptado de Graça Martins, M. E. et al 2011, página 174) Na turma o professor propõe quatro 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. Depois dessa discussão pede-se para calcular a probabilidade teórica (de Laplace) de um dos alunos ganhar. 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?

Admitindo que o dado é equilibrado, qual a probabilidade de ganhar a Rita?

(Ver OBSERVAÇÃO)

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?

Admitindo que o dado é equilibrado, qual a probabilidade de ganhar a Maria?

(Ver OBSERVAÇÃO)

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?

Admitindo que as moedas são equilibradas, qual a probabilidade de ganhar o João?

(Ver OBSERVAÇÃO)

Tarefas propostas

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?

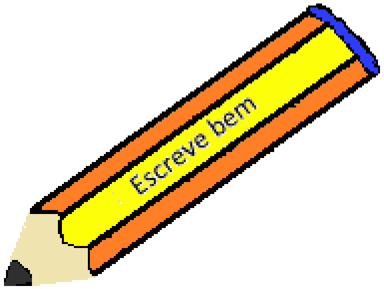
Admitindo que os dados são equilibradas, qual a probabilidade de ganhar a Sara?

(Ver OBSERVAÇÃO)

Observação – 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.

Ver outros exemplos em Graça, Martins, M.E. 2011, página 175.

2. A experiência com o lápis



Considera um lápis com seis faces, idêntico ao da imagem ao lado. Considera a experiência aleatória que consiste em rodar o lápis em cima de uma mesa e verificar se a face que ficou virada para cima é a que tem a marca do lápis. Repete a experiência 30 vezes.

Nº experiência	Sim-1; Não-0	Freq. Rel.
1	0	0
2	1	0,50
3	0	0,33
4	0	0,25
5	1	0,40
...	...	...
29	...	...
30	...	...

a) Ao fim das 30 repetições calcula a frequência relativa com que se verificou a face com a marca virada para cima, como se exemplifica na tabela ao lado.

b) Pensas que o valor que obtiveste para a frequência relativa é uma boa estimativa para a probabilidade de a face que fica virada para cima ser a que tem a marca? Explica porquê.

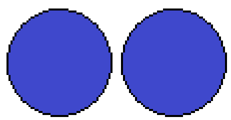
c) Admitindo que as 6 faces do dado têm igual probabilidade de ficarem viradas para cima, qual a probabilidade de obteres a face com a marca?

d) Compara os valores obtidos em a) e em c).

3. Um jogo com duas fichas

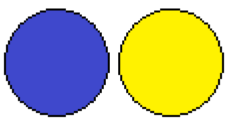
Considera duas fichas azuis, por exemplo, das que se colocam nos carrinhos de supermercado. A uma das fichas pinta de amarelo uma das faces. O Pedro e o Francisco decidiram jogar o seguinte jogo. Colocam as fichas num copo, agitam o copo, lançam as fichas sobre a mesa e verificam as cores:

Se sair a mesma cor nas 2 fichas



O Pedro ganha 1 ponto

Se as duas fichas não tiverem a mesma cor



O Francisco ganha 1 ponto

Ganha o jogo quem conseguir primeiro 10 pontos.

a) Achas que o jogo é equilibrado?

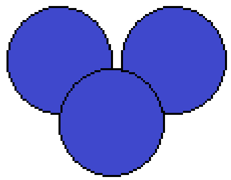
b) Simla o jogo do Pedro e do Francisco e regista as frequências relativas do número de pontos de cada jogador em 20 lançamentos.

c) Considera o espaço de resultados associado ao fenómeno aleatório em estudo. Atribui probabilidades aos acontecimentos elementares e justifica que o jogo é equilibrado.

4. Um jogo com três fichas

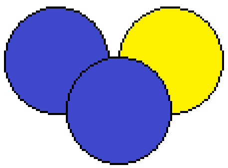
Acrescenta uma nova ficha azul (nas duas faces) às fichas do jogo anterior.

Se sair a mesma cor nas 3 fichas



O Pedro ganha 1 ponto

Se as 3 fichas não tiverem a mesma cor



O Francisco ganha 1 ponto

Ganha o jogo quem conseguir primeiro 10 pontos.

a) Achas que o jogo é equilibrado?

b) Simula o jogo do Pedro e do Francisco e regista as frequências relativas do número de pontos de cada um em 20 lançamentos.

c) Considera o espaço de resultados associado ao fenómeno aleatório em estudo. Atribui probabilidades aos acontecimentos elementares e justifica se o jogo é equilibrado.

5. Ainda um jogo com três fichas ...

Considera o jogo anterior, mas em que a terceira ficha tem uma face azul e outra amarela, ou seja, neste momento tens 3 fichas em que uma é azul nas duas faces e as outras duas fichas têm uma face azul e a outra amarela.

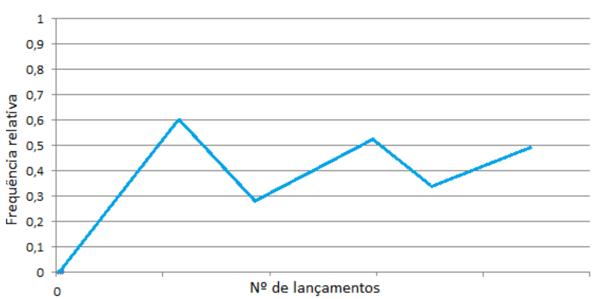
Ganha o jogo quem conseguir primeiro 10 pontos.

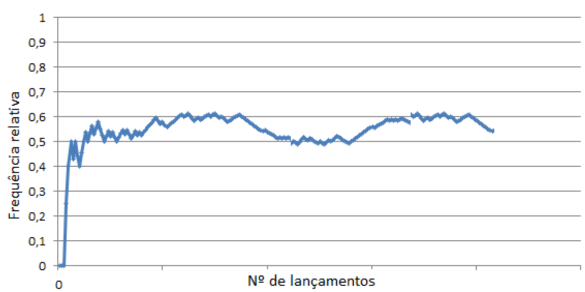
a) Achas que o jogo é equilibrado?
b) Regista as frequências relativas do número de pontos de cada um em 20 lançamentos.
c) Considera o espaço de resultados associado ao fenómeno aleatório em estudo. Atribui probabilidades aos acontecimentos elementares e justifica se o jogo é equilibrado.

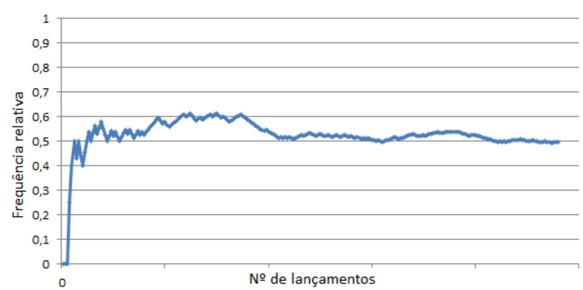
6. A moeda será equilibrada?

Supõe que tens uma moeda de um euro que não sabes se é equilibrada, isto é se as duas faces têm a mesma probabilidade de sair quando a lanças. Descreve uma experiência que te permita adiantar a hipótese de a moeda ser efectivamente equilibrada.

Se fizeres lançamentos sucessivos da moeda e representares num gráfico de linhas a evolução da frequência relativa da saída da face Euro ao longo dos lançamentos, qual dos gráficos abaixo é que te permitiria concluir que uma estimativa para a probabilidade teórica da saída da face Euro é ½, ou seja, que a moeda é equilibrada?







Bibliografia

Neste texto utilizou-se, essencialmente, a seguinte bibliografia:

Bibliografia

Feller, W. (1968) – *An Introduction to Probability Theory and its Applications*, John Wiley & Sons.

Graça Martins, M.E., Ponte, J. P. (2011) – Organização e tratamento de dados, DGIDC.

Graça Martins, M.E., Loura, L., Mendes, M. F. (2007) – Análise de dados, DGIDC.

Graça Martins, M.E. (2005) - Introdução à Probabilidade e à Estatística – com complementos de Excel – SPE, em http://www.arquivoescolar.org/bitstream/arquivo-e/97/1/IPE%202005.pdf.

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.

Moore, D. (1997) – *Statistics* – Concepts and Controversies. Freeman.

Páginas na Internet

ALEA - http://www.alea.pt

PORDATA - http://www.pordata.pt/

Fim

Para alguma dúvida contactar

Maria Eugénia Graça Martins

memartins@fc.ul.pt ou

megm@ef.pt