

# ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO

P. PORTO

## CAP.1 – Estatística Descritiva



UC: Oficina de Estatística  
2º Ano, 1º Semestre

Docente: Rosa Silveira  
Contacto: [rmss@estg.ipp.pt](mailto:rmss@estg.ipp.pt)

## 1. Estatística Descritiva

Porquê “Oficina de Estatística”?

*“Saber o que medir e como medir transforma um mundo complicado num mundo muito menos complicado. Se aprendermos a olhar para os dados de uma forma certa, conseguimos explicar enigmas que, caso contrário, podiam parecer impossíveis de resolver. Porque não há nada como o poder cristalino dos números para limpar camadas de confusão e contradições.”*

Levitt & Dubner (2011) in *Freakonomics*



## 1. Estatística Descritiva

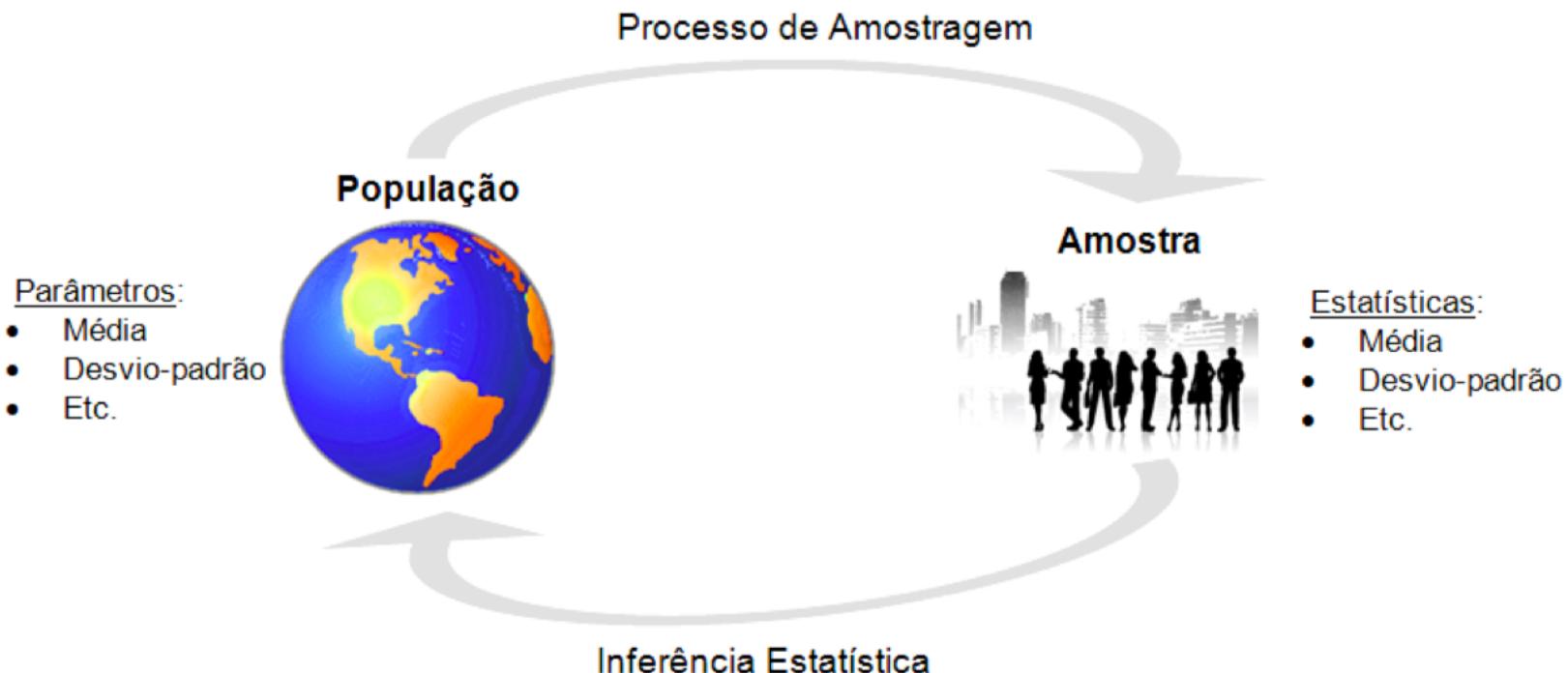
Conceito e Importância da Estatística

- ✓ Ramo da Matemática Aplicada **não determinística mas estocástica** (baseada no acaso)
- ✓ Na economia, biologia, medicina, meteorologia... a experimentação não é efetuada de forma rigorosa porque os objetos de estudo estão em constante mudança;
- ✓ A **Estatística** procura estudar uma ou várias características de uma população tendo por base um conjunto de métodos que permitem a recolha, classificação, apresentação e interpretação dos dados sobre o fenómeno aleatório em estudo.



## 1. Estatística Descritiva

### População e Amostra





## 1. Estatística Descritiva

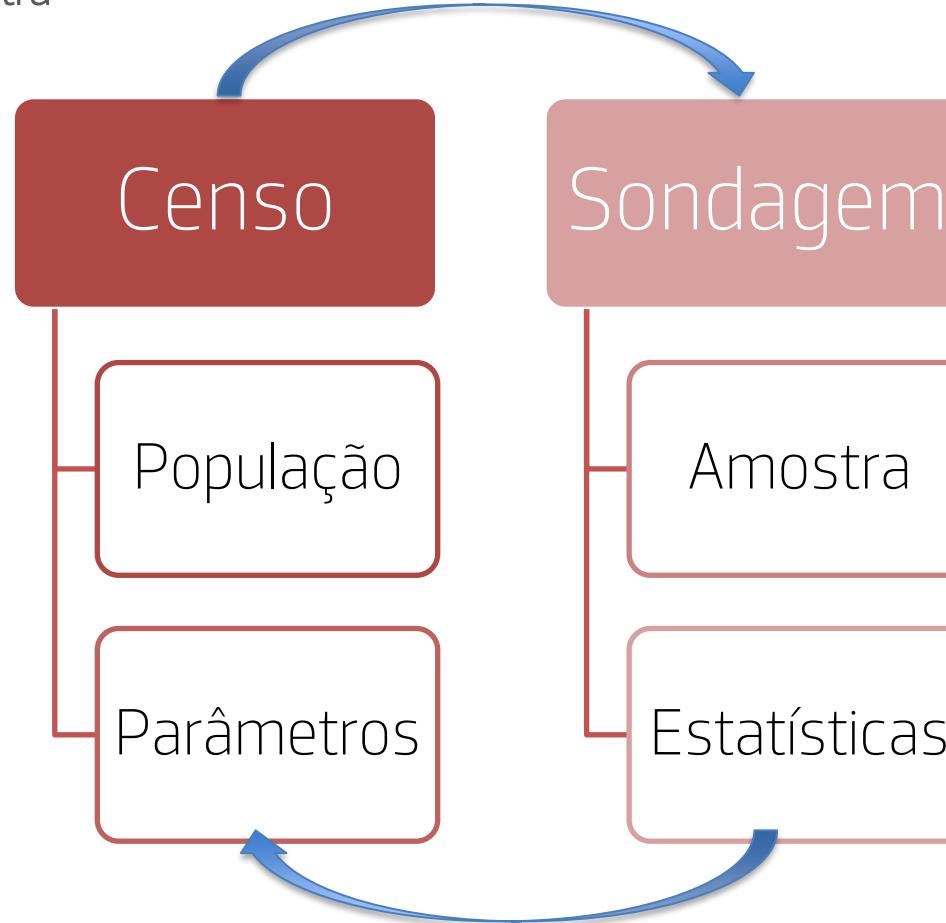
### População e Amostra

- ✓ Ao grupo de todos os elementos que se pretende estudar e que possuem uma característica (ou mais) em comum designamos por **população**.
- ✓ No entanto, nem sempre é viável inquirir todos os elementos de uma população por diversas razões: i) população de grande dimensão; ii) custo excessivo do processo de recolha e tratamento dos dados; iii) inacessibilidade a alguns dos elementos da população. Como tal, seleciona-se um subgrupo da população para análise, designado por **amostra**.
- ✓ As medidas relativas à **amostra** designam-se por **estatísticas**. O valor dessas estatísticas varia de amostra para amostra, ou seja, tratam-se de uma variável aleatória (v.a.). As medidas relativas a uma **população** designam-se por **parâmetros**.

## 1. Estatística Descritiva

### População e Amostra

Se for possível inquirir toda a População , os parâmetros tornam-se conhecidos pela Estatística Descritiva



Escolher uma amostra é um processo que implica erro no cálculo de parâmetros por Inferência Estatística. A amostra deve ser grande e representativa

## 1. Estatística Descritiva

Subáreas da Estatística

Podemos dividir a estatística, essencialmente, em duas áreas:

- > Estatística Descritiva.
- > Inferência Estatística/ Estatística Indutiva.

### Estatística Descritiva

- ✓ Conjunto de técnicas apropriadas para recolher, organizar, reduzir e apresentar dados estatísticos. No âmbito da estatística descritiva, procura-se sintetizar e representar de forma compreensível a informação contida num conjunto de dados.
- ✓ Essa tarefa materializa-se na construção de tabelas, gráficos ou no cálculo de medidas que representem convenientemente a informação contida nos dados – as estatísticas.

## 1. Estatística Descritiva

Subáreas da Estatística

### Inferência Estatística/ Inferência Estatística

- ✓ Procura retirar conclusões para a população em geral, com base na análise dos resultados obtidos para um ou mais subconjuntos (amostras). Trata-se de um conjunto de técnicas que permitem caracterizar uma população, requerendo o conhecimento das probabilidades.
  
- ✓ As principais técnicas utilizadas são a estimação – que visa determinar o valor dos parâmetros desconhecidos (da qual se pode destacar, por exemplo, os intervalos de confiança) – e os testes de hipóteses – que visam testar suposições acerca das características de uma certa população.

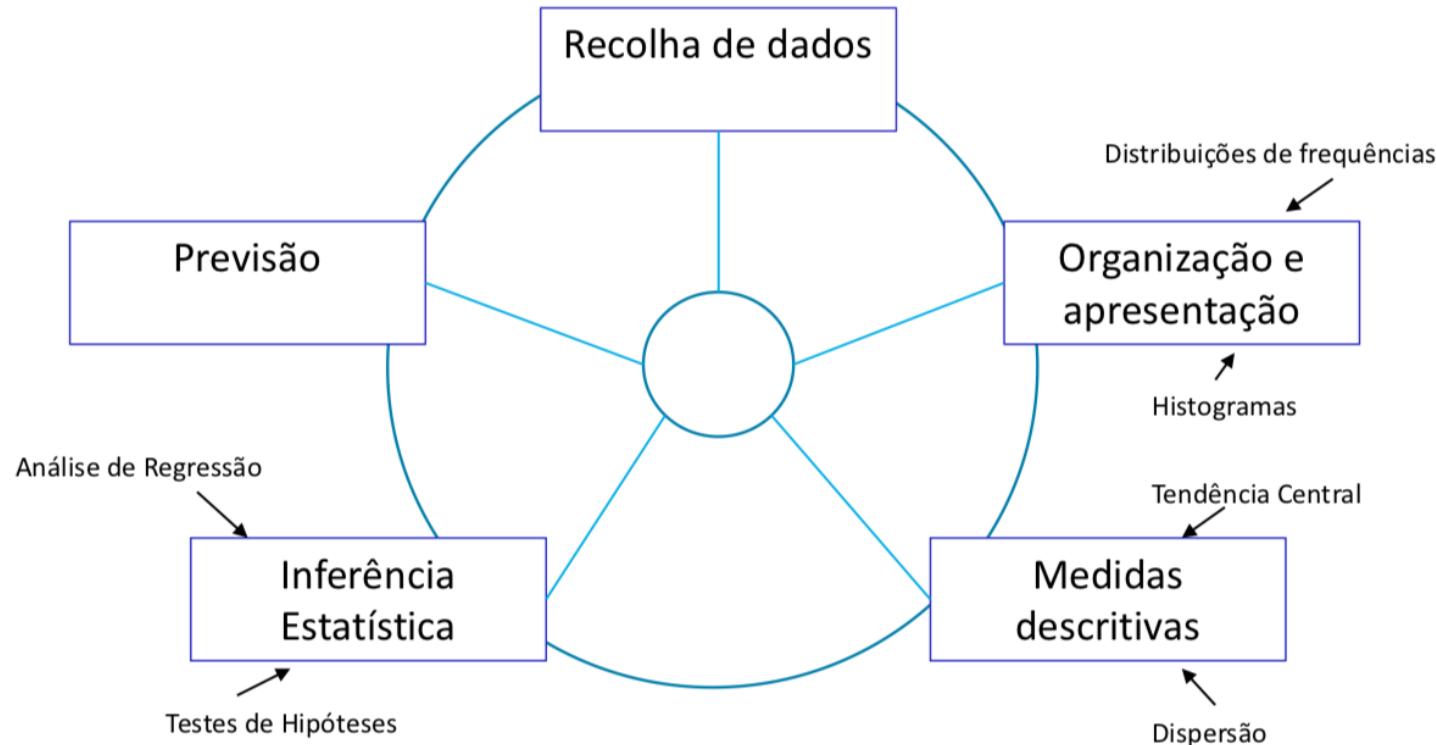
## 1. Estatística Descritiva

Resumindo: Estatística Descritiva *versus* Inferência Estatística

Estatística Descritiva	Inferência Estatística
<ul style="list-style-type: none"><li>➤ Recolher</li><li>➤ Organizar</li><li>➤ Sumariar</li><li>➤ Analisar</li><li>➤ Representar</li></ul>	<ul style="list-style-type: none"><li>➤ Prever valores dos parâmetros da população</li><li>➤ Testar hipóteses sobre os valores dos parâmetros da população</li><li>➤ Tomar decisões</li></ul>

## 1. Estatística Descritiva

Método Estatístico



## 1. Estatística Descritiva

### Conceito de Variável

- ✓ Em qualquer estudo estatístico é necessário definir a **unidade estatística** (qualquer indivíduo, objeto ou fato que é objeto da observação ou das conclusões) e o que se pretende estudar sobre ela (**características**) – **Variáveis**
  - > **Dado Estatístico:** resultado da observação das unidades estatísticas que compõem um determinado conjunto.
  - > **Variável:** característica relativa a todos os indivíduos (ou unidades estatísticas) observados. O valor desta característica varia com as observações.
- ✓ Os dados são, geralmente, discriminados em **listas ou em tabela** (vetores ou matrizes)

## 1. Estatística Descritiva

Dados Qualitativos *versus* Dados Quantitativos



- ✓ A maioria dos estudos estatísticos são quantitativos mas incompletos;
- ✓ O objetivo dos estudos qualitativos passa por ganhar um profundo conhecimento do acontecimento (Carson *et al.*, 2001) que é conseguido pela imersão do investigador no contexto do fenómeno a ser estudado (Patton, 1990).
- ✓ As variáveis qualitativas acrescem a possibilidade de se estudarem grupos e escalas de opiniões, concordâncias, sentimentos, dor... que de outra forma seria impossível considerar.

## 1. Estatística Descritiva

Dados Qualitativos *versus* Dados Quantitativos

Características	Quantitativa	Qualitativa
Foco	"Porquê": preocupa-se com as causas.	"Como": preocupa-se em compreender os fenómenos.
Objeto de Estudo	Factos naturais descritos.	Fenómenos humanos (significados).
Investigador	Distancia-se do facto pesquisado.	Olha à luz da sua subjetividade.
Objetivos de Pesquisa	Teste de hipóteses: descrição e estabelecimento de correlações matemáticas (estatísticas) e causais entre factos.	Compreensão; explanação. Apreensão e interpretação da relação de significados de fenómenos para os indivíduos e a sociedade.
Tratamento/ Análise dos Dados	Técnicas estatísticas.	Análise de conteúdo.
Instrumentos de Pesquisa	Experimentações, questionários fechados e escalas.	Observação, entrevistas, diário de campo e análise documental.

Fonte: Pasqualotti & Portella (2003)

## 1. Estatística Descritiva

### Classificação das Variáveis

#### 1. Variáveis Qualitativas (ou Categóricas)

Os dados podem ser separados em diferentes categorias que se distinguem por características não numéricas, podendo ser expressos em duas escalas distintas: nominal e ordinal.

- ✓ **Escala Nominal:** quando os dados estão divididos por categorias que não possuem ordem, ou seja, quando cada um dos dados for identificado pela atribuição de um nome que designa uma classe. As classes também podem ser designadas por números. Contudo, para que a escala seja nominal não se poderá estabelecer qualquer relação de ordem entre tais números. Exemplos:
  - > Classificação de pessoas pela cor de cabelo: preto, castanho, branco, loiro, entre outros.
  - > Classificação dos consumidores de bens de primeira necessidade pelo sexo: feminino ou masculino.
  - > Classificação dos consumidores pelo género no âmbito de um programa estatístico .

## 1. Estatística Descritiva

Classificação das Variáveis

### 1. Variáveis Qualitativas (ou Categóricas)

✓ **Escala Ordinal:** quando os dados estão divididos por categorias que obedecem a uma sequência com significado, isto é, a ordem das classes é relevante. Exemplos:

- > Classificações obtidas num teste de estatística: mau, medíocre, suficiente, bom ou muito bom.
- > Classificação dos clientes : clientes A (muito importantes) a C (menos importantes).

Para além desta designação, as variáveis de natureza qualitativa podem ainda ser classificadas como:

- ✓ **Dicotómicas:** se existem apenas duas classes. Exemplo: género (masculino ou feminino).
- ✓ **Polítómicas:** se existem três ou mais classes. Exemplo: Estado Civil (solteiro, casado, divorciado ou viúvo).

## 1. Estatística Descritiva

Classificação das Variáveis

## 2. Variáveis Quantitativas (ou Numéricas)

As variáveis de natureza quantitativa podem ainda ser classificadas em:

- ✓ **Variáveis Discretas:** podem assumir um número finito ou uma infinidade numerável de valores.

Exemplos:

- > Resultados de 150 lançamentos de um dado.
- > Número de pessoas que visita um museu numa hora.

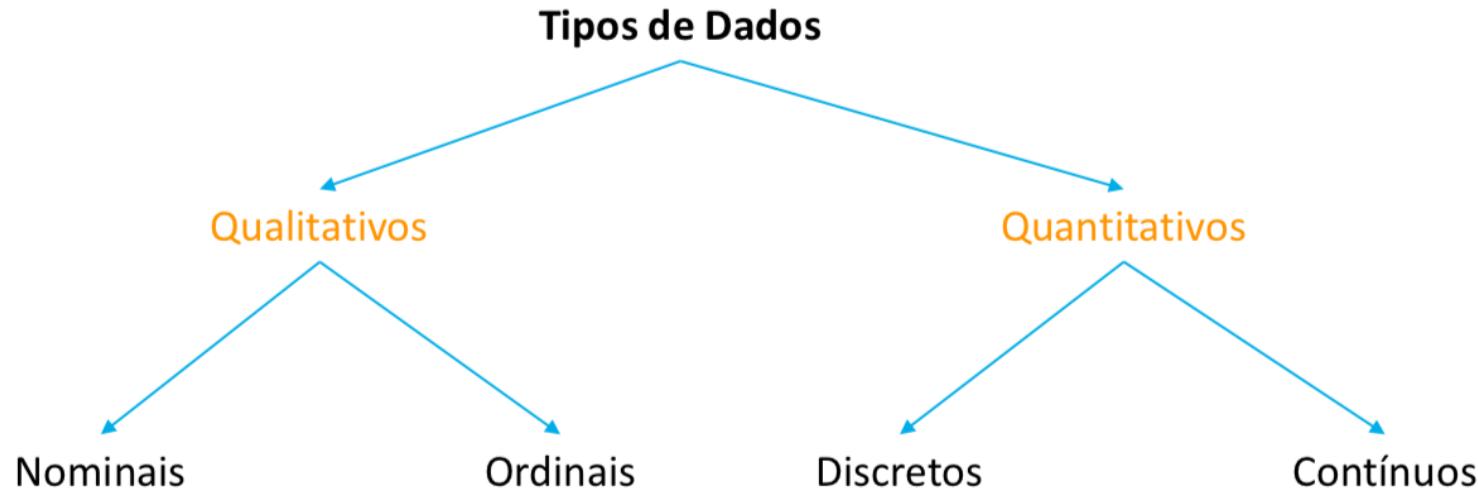
- ✓ **Variáveis Contínuas:** podem tomar um número infinito não numerável de valores, ou seja, podem assumir qualquer valor dentro de um intervalo de números reais. Exemplos:

- > Temperatura do ar expressa em graus Celsius ( $^{\circ}\text{C}$ ).
- > Volumes de investimento traduzidos em milhares de euros.



## 1. Estatística Descritiva

Classificação das Variáveis



- ✓ Os estudos envolvem os cruzamentos de várias variáveis;
- ✓ Nesta UC apenas pretendemos analisar cruzamentos duas a duas
- ✓ As variáveis qualitativas definem os grupos (variável independente)

## 1. Estatística Descritiva

### Representações Tabulares de Dados Qualitativos ou Quantitativos Discretos

**Tabela de Frequências:** disponibiliza um acesso rápido ao número, à percentagem ou proporção de elementos observados com uma determinada característica, valor ou intervalo de valores (denominadas classes de valores). Relaciona as  $K$  categorias/classes de valores com o número de ocorrências/frequências que pertencem a cada categoria ou classe.

Categorias ( $x_i$ )	Frequência absoluta ( $n_i$ )	Frequência relativa ( $f_i$ )	Frequência absoluta acumulada ( $N_i$ )	Frequência relativa acumulada ( $F_i$ )
$x_1$	$n_1$	$f_1 = n_1/n$	$N_1 = n_1$	$F_1 = f_1$
$x_2$	$n_2$	$f_2 = n_2/n$	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
$x_3$	$n_3$	$f_3 = n_3/n$	$N_3 = n_1 + n_2 + n_3$	$F_3 = f_1 + f_2 + f_3$
...	...	...	...	...
$x_k$	$n_k$	$f_k = n_k/n$	$N_k = \sum_{i=1}^k n_k = n$	$N_k = \sum_{i=1}^k f_k = 1$
Total	$n$	1		

Nota: Para **dados qualitativos** não se calculam as frequências (absolutas e relativas) acumuladas que correspondem às 4<sup>a</sup> e 5<sup>a</sup> colunas.

## 1. Estatística Descritiva

### Representações Tabulares de Dados Qualitativos ou Quantitativos Discretos

- ✓ As **frequências acumuladas (absolutas ou relativas)** incorporadas na tabela têm interesse quando se pretende determinar o número ou a percentagem de observações que são menores ou iguais a um determinado valor.
- ✓ Dicas para a construção das classes:
  - > As classes deverão ser mutuamente exclusivas, ou seja, um valor observado apenas poderá pertencer a uma e a uma só classe.
  - > As classes deverão ser exaustivas, isto é, deverão compreender todos os valores observados.



## 1. Estatística Descritiva

### Representações Tabulares de Dados Qualitativos ou Quantitativos Discretos

**Exemplo 1:** Realizou-se um inquérito a 50 trabalhadores de uma empresa de construção para analisar a utilidade que davam aos capacetes de proteção. Verificou-se que 4 trabalhadores afirmaram usar SEMPRE o capacete de proteção, 23 usam QUASE SEMPRE, 16 que RARAMENTE usam e 7 assumiram que NUNCA usam o capacete de proteção. Tratam-se de dados qualitativos, de natureza ordinal (categorias ordenadas).

Tipo de Utilidade	Número de trabalhadores	Frequência relativa
Sempre	4	$\frac{4}{50} = 8\%$
Quase Sempre	23	$\frac{23}{50} = 46\%$
Raramente	16	$\frac{16}{50} = 32\%$
Nunca	7	$\frac{7}{50} = 14\%$
<b>Total</b>	50	$\frac{50}{50} = 100\%$

## 1. Estatística Descritiva

### Representações Tabulares de Dados Qualitativos ou Quantitativos Discretos

**Exemplo 2:** Numa amostra constituída por 120 peças constatou-se que 100 não tinham qualquer defeito, 15 tinham defeitos mas eram recuperáveis e 5 eram irrecuperáveis (ou seja, constituíam sucata). Uma proposta de representação tabular destes dados qualitativos, de natureza nominal (categorias não ordenadas) é a seguinte:

Categorias de peças ( $x_i$ )	Frequência absoluta ( $n_i$ )	Frequência relativa % ( $f_i$ )
Sem defeito	100	83,3
Recuperáveis	15	12,5
Sucata	5	4,2
Total	120	100

## 1. Estatística Descritiva

### Representações Tabulares de Dados Qualitativos ou Quantitativos Discretos

**Exemplo 3:** Num estudo para analisar a ocorrência de acidentes de trabalho num determinado hospital, em 397 profissionais de saúde, verificou-se que 16 não sofreram qualquer acidente, 32 tiveram 1 acidente, 89 reportaram 2 acidentes, 137 sofreram 3 acidentes, 98 tiveram 4 acidentes e 25 profissionais reportaram 5 acidentes. Uma proposta de representação tabular destes **dados quantitativos discretos** é a seguinte:

Número de acidente por profissional ( $x_i$ )	Número de profissionais ( $n_i$ )	Proporção de profissionais ( $f_i$ )	N.º acumulado de profissionais ( $N_i$ )	Proporção acumulada de profissionais ( $F_i$ )
0	16	0,0403	16	0,0403
1	32	0,0806	48	0,1209
2	89	0,2242	137	0,3451
3	137	0,3451	274	0,6902
4	98	0,2469	372	0,9370
5	25	0,0630	397	1,0000
<b>Total</b>	<b>397</b>	<b>1</b>		

## 1. Estatística Descritiva

### Representações Tabulares de Dados Quantitativos Contínuos

Classes (C <sub>i</sub> )	Ponto médio (x <sub>i</sub> )	Frequência absoluta (n <sub>i</sub> )	Frequência relativa (f <sub>i</sub> )	Frequência absoluta acumulada (N <sub>i</sub> )	Frequência relativa acumulada (F <sub>i</sub> )
[L <sub>I</sub> <sub>1</sub> ; L <sub>S</sub> <sub>1</sub> [	x <sub>1</sub>	n <sub>1</sub>	f <sub>1</sub> = n <sub>1</sub> /n	N <sub>1</sub> = n <sub>1</sub>	F <sub>1</sub> = f <sub>1</sub>
[L <sub>I</sub> <sub>2</sub> ; L <sub>S</sub> <sub>2</sub> [	x <sub>2</sub>	n <sub>2</sub>	f <sub>2</sub> = n <sub>2</sub> /n	N <sub>2</sub> = n <sub>1</sub> + n <sub>2</sub>	F <sub>2</sub> = f <sub>1</sub> + f <sub>2</sub>
[L <sub>I</sub> <sub>3</sub> ; L <sub>S</sub> <sub>3</sub> [	x <sub>3</sub>	n <sub>3</sub>	f <sub>3</sub> = n <sub>3</sub> /n	N <sub>3</sub> = n <sub>1</sub> + n <sub>2</sub> + n <sub>3</sub>	F <sub>3</sub> = f <sub>1</sub> + f <sub>2</sub> + f <sub>3</sub>
...	...	...	...	...	...
[L <sub>I</sub> <sub>k</sub> ; L <sub>S</sub> <sub>k</sub> ]	x <sub>k</sub>	n <sub>k</sub>	f <sub>k</sub> = n <sub>k</sub> /n	N <sub>k</sub> = $\sum_{i=1}^k n_k = n$	N <sub>k</sub> = $\sum_{i=1}^k f_k = 1$
Total		n	1		

## 1. Estatística Descritiva

### Representações Tabulares de Dados Quantitativos Contínuos

**Exemplo:** O quadro seguinte apresenta a duração média de processos (em anos) relacionados com um determinado tipo de crime.

2.25	0.63	2.10	2.37	0.80	3.44
8.55	2.07	3.22	2.48	0.54	4.19
7.40	4.60	3.45	6.20	1.88	0.82
1.91	0.10	6.31	9.56	2.10	2.77
1.57	4.21	5.14	5.42	0.65	4.30

- Construção das classes

$$n = 30$$

$$k = \sqrt{30} \approx 5$$

$$X_{\max} = 9.56 \quad X_{\min} = 0.1$$

$$a = \frac{9.56 - 0.1}{5} \approx 2$$

Duração (em anos)	Nº de processos	Percentagem de processos	Frequência absoluta acumulada	Frequência relativa acumulada
Classes	$n_i$	$f_i$	$N_i$	$F_i$
[0, 2]	9	0.30	9	0.30
]2, 4]	10	0.33	19	0.63
]4, 6]	6	0.20	25	0.83
]6, 8]	3	0.10	28	0.93
]8, 10]	2	0.07	30	1.00
Total	30	1		

## **1. Estatística Descritiva**

### Representações Gráficas de Dados Univariados

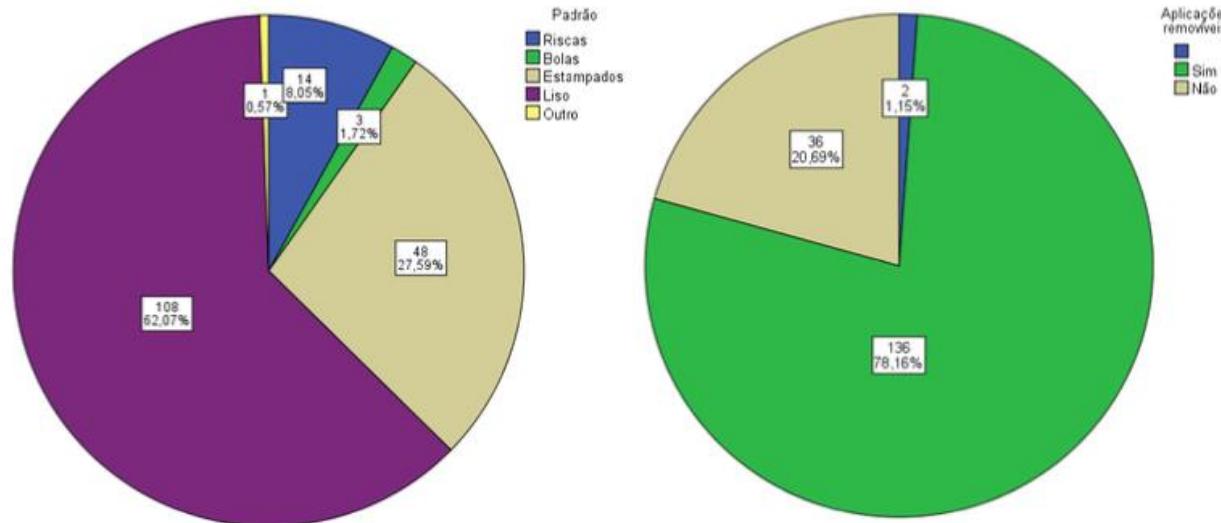
Os gráficos mais utilizados para a representação e melhor visualização dos dados são:

Gráfico Circular	Dados qualitativos.
Gráfico de Barras	Dados qualitativos ou quantitativos discretos.
Gráfico de Frequências Acumuladas	Dados qualitativos na escala ordinal ou quantitativos discretos.
Histograma	Dados quantitativos contínuos.
Polígono de Frequências	Dados quantitativos.
Polígono de Frequências Acumuladas	Dados quantitativos contínuos.
Diagrama de Caule e Folhas	Dados não agrupados quantitativos.
Caixa de Bigodes	Dados não agrupados quantitativos.

## 1. Estatística Descritiva

### Representações Gráficas de Dados Univariados

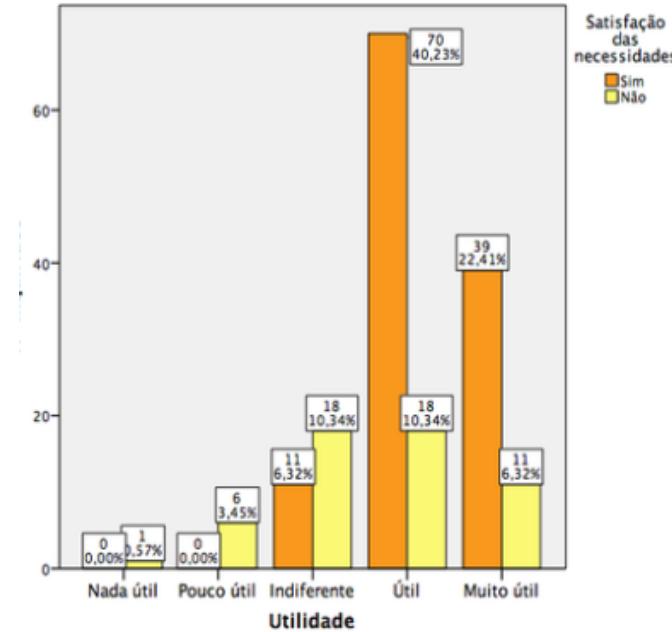
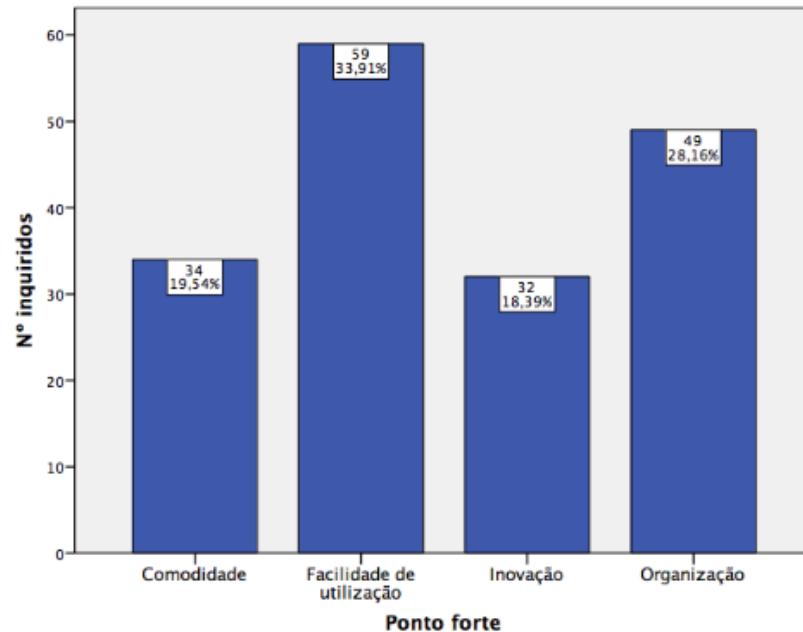
**Gráfico Circular:** constituído por um círculo dividido em tantas fatias quantas as categorias da variável. O tamanho das fatias é determinado pelo número ou percentagem/proporção de observações nas categorias e pelas frequências absolutas ou relativas.



## 1. Estatística Descritiva

### Representações Gráficas de Dados Univariados

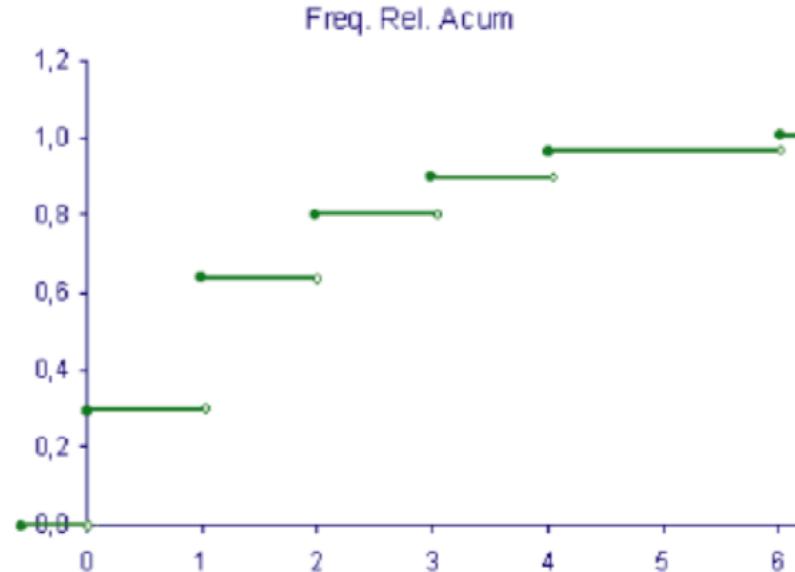
**Gráfico de Barras:** diagrama de barras, sendo cada barra associada a cada um das categorias da variável. A altura das barras é determinada pelas frequências absolutas ou relativas.



## 1. Estatística Descritiva

### Representações Gráficas de Dados Univariados

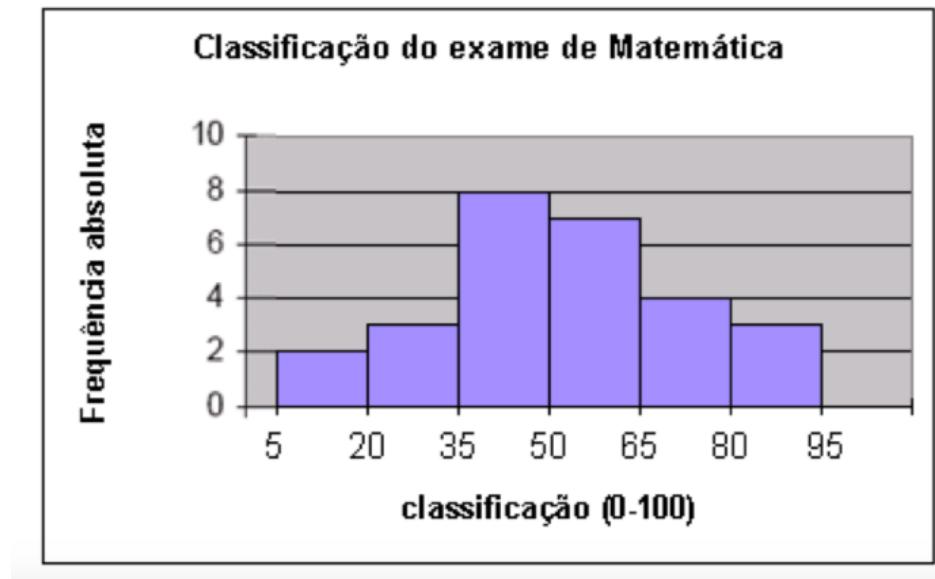
**Gráfico de Frequências Acumuladas:** gráfico de linhas onde são representadas as frequências absolutas ou relativas acumuladas.



## 1. Estatística Descritiva

### Representações Gráficas de Dados Univariados

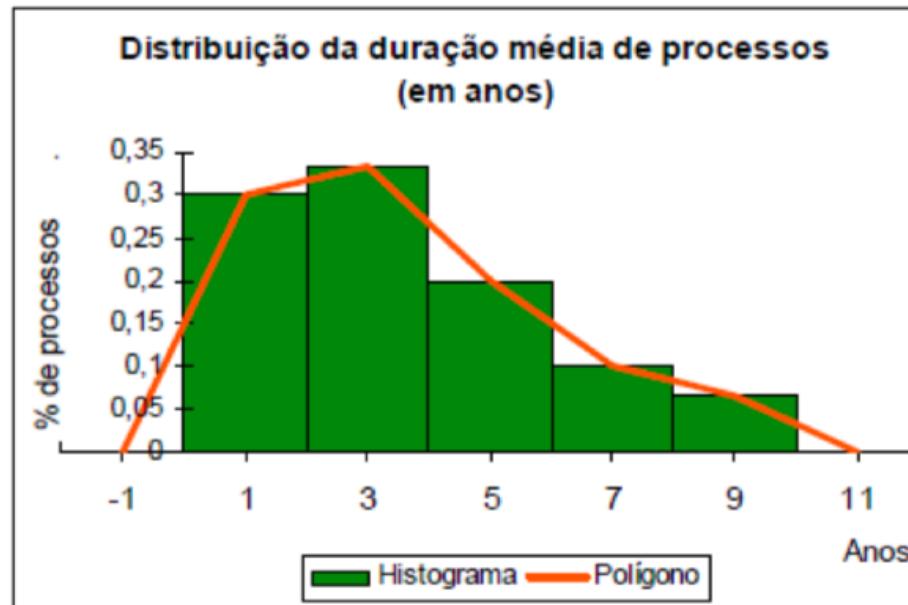
**Histograma:** gráfico de barras adjacentes com uma barra associada a cada uma das classes da variável. A base de cada barra é proporcional à amplitude da respetiva classe e a área é proporcional às frequências absolutas ou relativas.



## 1. Estatística Descritiva

### Representações Gráficas de Dados Univariados

**Polígono de Frequências:** gráfico de linhas onde são representadas as frequências absolutas ou relativas nos pontos médios das classes.

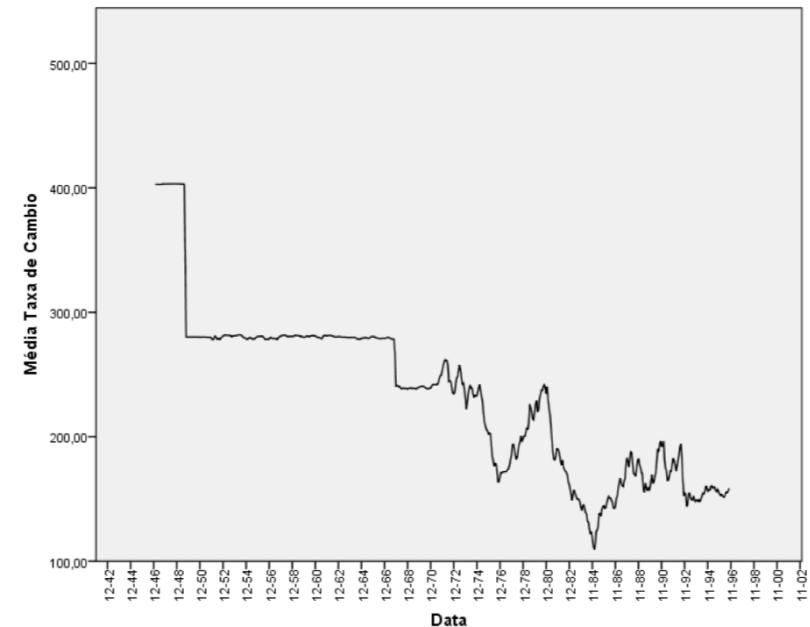


## 1. Estatística Descritiva

### Representações Gráficas de Dados Univariados

**Gráfico de Série Temporal:** gráfico que apresenta como a variável evolui ao longo do tempo.

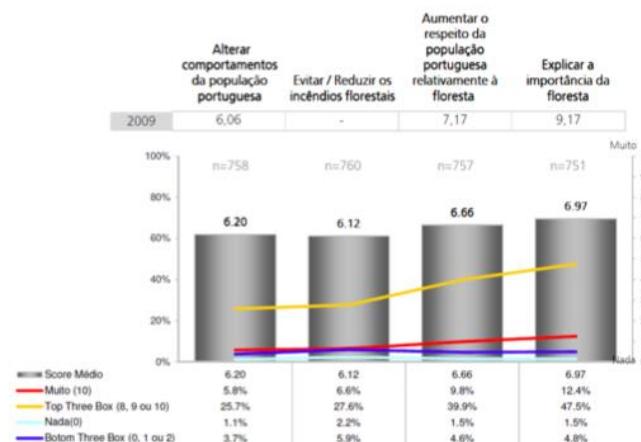
- ✓ O gráfico seguinte representa os dados de uma série temporal (mensal), desde janeiro de 1947 a outubro de 1996, relativos à taxa de câmbio libra esterlina/dólar americano. O conjunto de dados contém 598 observações.
- ✓ Possível interpretação da série temporal: é possível observar-se uma tentativa do Governo britânico para manter a taxa de câmbio fixa até ao final de 1971 e a depreciação gradual da libra sendo que esta flutuou de forma descendente em meados da década de 70.



## 1. Estatística Descritiva

### Representações Gráficas de Dados Univariados

Fonte: ICNF, 2013. IFN6 – Áreas dos usos do solo e das espécies florestais de Portugal continental. Resultados preliminares. [pdf], 34 pp, Instituto da Conservação da Natureza e das Florestas. Lisboa.



**Ficha técnica**  
**Pergunta 13:** Em que medida considera que a campanha "Portugal sem fogos depende de todos" contribui para...? Utilize uma escala de 0 a 10, em que 0 significa Nada e 10 Muito.  
**Tratamento estatístico:** média e frequência **Tipo de resposta:** escala

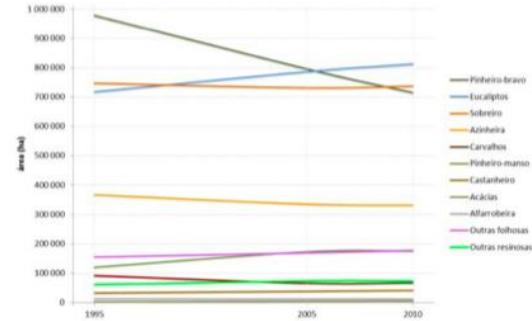
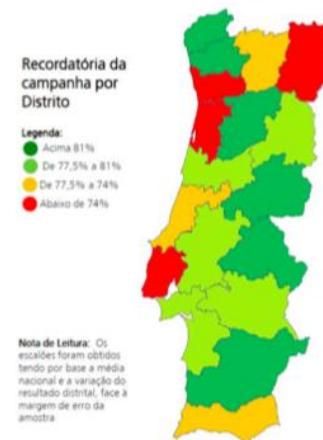


Figura 6 – Evolução das áreas totais por espécie

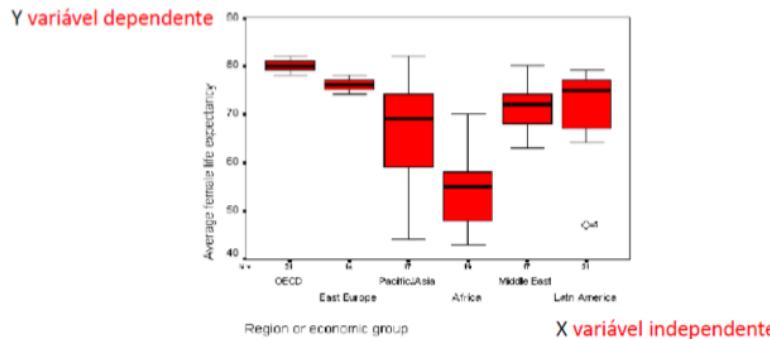


Impacto das campanhas de sensibilização 4ª Vaga.

In: Percepção da população portuguesa sobre os incêndios florestais e respectivas causas e impacto das campanhas de sensibilização. 4.º Vaga. Spirituc – Investigação Aplicada.

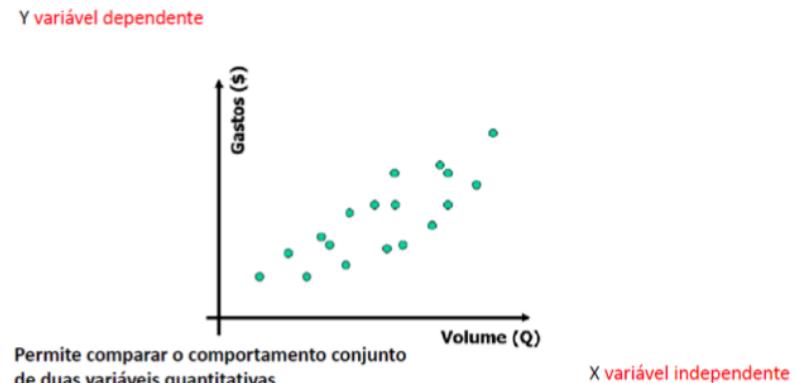
## 1. Estatística Descritiva

### Representações Gráficas de Dados Bivariados



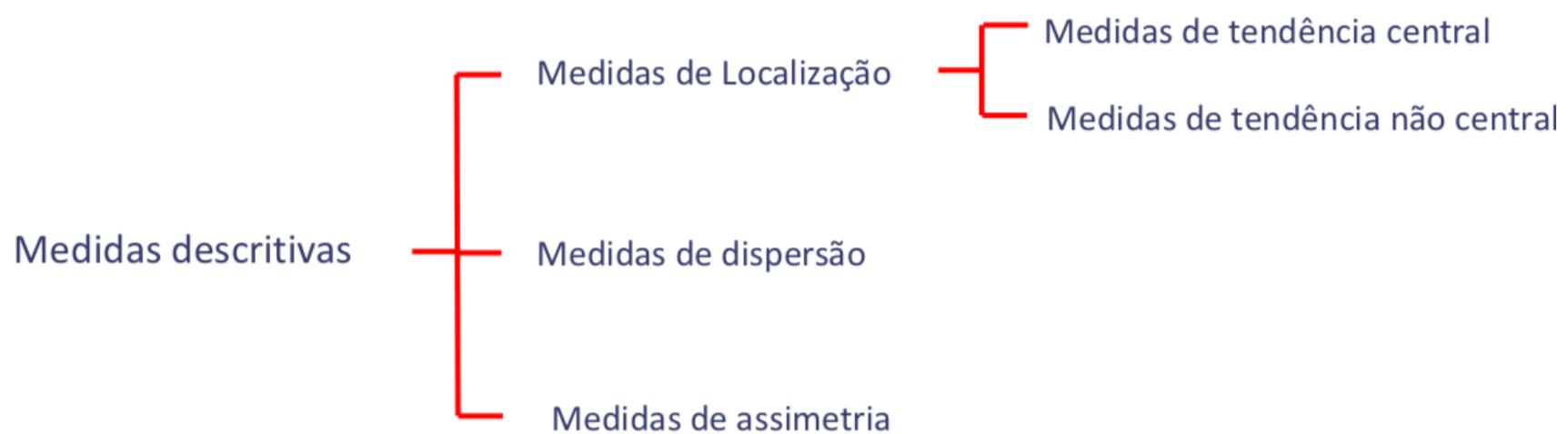
Permite comparar uma variável quantitativa (variável dependente) com uma variável que toma um número discreto de valores (variável independente)

Diagrama de dispersão (scatterplot)



## 1. Estatística Descritiva

### Medidas Descritivas



## Medidas de Localização – Tendência Central

## 1. Estatística Descritiva

### Medidas de Localização

As **Medidas de Tendência Central** informam sobre a localização de alguns valores importantes da distribuição e representam os fenómenos pelos seus valores médios, em torno dos quais tendem a concentrar-se os valores observados.



## 1. Estatística Descritiva

### Média Aritmética/Média Amostral

**Dados Discretos/Não Agrupados:** a média é a soma das observações dividida pelo seu número total.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Onde  $x_1, x_2, \dots, x_n$  são os  $n$  dados/observações que constituem a amostra.

**Dados Agrupados:** a média é a soma da marca da classe multiplicada pela frequência, dividida pelo número total de observações.

$$\bar{x} = \frac{\sum_{i=1}^n n_i \cdot x_i}{n}$$

Marca da classe

Onde  $n$  é o número de classes,  $n_i$  é a frequência absoluta da classe  $i$  e  $x_i$  é a marca da classe ou ponto médio da mesma.

## 1. Estatística Descritiva

### Média Aritmética/Média Amostral – Propriedades

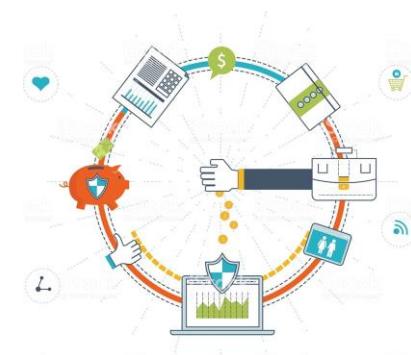
#### Propriedades da Média

✓ Vantagens:

- > Facilidade de interpretação e cálculo.
- > Utiliza toda a informação disponível e pode ser calculada com precisão matemática.

✓ Desvantagens:

- > Influenciada por valores extremos (*outliers*) que têm um peso significativo no cálculo da média.
- > Pode não corresponder a um valor concreto da variável.



## 1. Estatística Descritiva

Determinação da Média Aritmética/Média Amostral – Exemplos

Cálculo da média aritmética de dados não agrupados:

Dados: 4 5 7 11 8 7

Média aritmética:  $\bar{x} = \frac{4 + 5 + 7 + 11 + 8 + 7}{6} = 7$

Cálculo da média aritmética de dados agrupados por classes intervalares:

Classes	Marca das classes (x)	Frequência absoluta (n <sub>i</sub> )
[2; 4,5[	$\frac{2 + 4,5}{2} = 3,25$	3
[4,5; 7[	$\frac{4,5 + 7}{2} = 5,75$	2
[7; 9,5[	$\frac{7 + 9,5}{2} = 8,25$	5
[9,5; 12[	$\frac{9,5 + 12}{2} = 10,75$	4
		n = 14

Média aritmética:

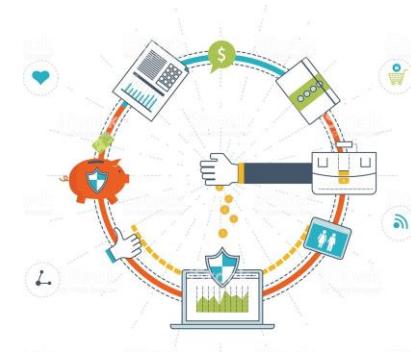
$$\bar{x} = \frac{3 \times 3,25 + 2 \times 5,75 + 5 \times 8,25 + 4 \times 10,75}{14} \approx 7,55$$

## 1. Estatística Descritiva

### Moda

A **moda ( $M_0$ )** é o valor da variável ao qual corresponde uma maior frequência (absoluta ou relativa), ou seja, trata-se do valor mais comum.

- ✓ **Dados Discretos/Não Agrupados:** é o valor que surge com mais frequência.
  - > Amostra Bimodal: existem dois valores com a maior frequência.
  - > Amostra Amodal: todos os valores da variável têm a mesma frequência.
  - > Amostra Plurimodal: existem vários valores com a frequência mais alta.
- ✓ **Dados Agrupados:** é o intervalo da classe com maior frequência que assume a designação de classe modal.



## 1. Estatística Descritiva

### Moda – Propriedades

#### Propriedades da Moda

✓ Vantagens:

- > É pouco sensível a valores extremos (*outliers*). Sendo assim, é mais indicativa de valores típicos do que a média.

✓ Desvantagens:

- > Tem pouco (ou nenhum) significado para um pequeno número de observações. Neste caso, muito frequentemente, não existe moda uma vez que os valores não se repetem.
- > No caso de dados quantitativos contínuos, o cálculo da moda torna-se um processo computacionalmente complexo.
- > O valor exato pode não corresponder a um dado observado.
- > Pode haver mais do que uma moda.

## 1. Estatística Descritiva

### Cálculo da Moda para Dados Agrupados por Classes Intervalares

1º) Identificar a **classe modal** (a classe com maior frequência).

2º) O valor da moda é dado pela seguinte expressão:

$$M_0 = LI_{Mo} + ac_{Mo} \times \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

$$\Delta_1 = n_{Mo} - n_{Mo-1} \quad \Delta_2 = n_{Mo} - n_{Mo+1}$$

Onde:

$LI_{Mo}$  – limite inferior da classe modal;

$ac_{Mo}$  – amplitude da classe modal;

$n_{Mo}$  – freq. abs. da classe modal;

$n_{Mo-1}$  – freq. abs. da classe anterior à classe modal;

$n_{Mo+1}$  – freq. abs. da classe posterior à classe modal;

## 1. Estatística Descritiva

Determinação da Moda – Exemplos

Cálculo da moda para dados não classificados:

Dados: 4 5 7 11 8 7

Moda:  $M_o = 7$

Cálculo da moda para dados agrupados por classes intervalares:

Classe modal

Classes	Marca das classes ( $x_i$ )	Frequência absoluta ( $n_i$ )
[2; 4,5[	$\frac{2 + 4,5}{2} = 3,25$	3
[4,5; 7[	$\frac{4,5 + 7}{2} = 5,75$	2
[7; 9,5[	$\frac{7 + 9,5}{2} = 8,25$	5
[9,5; 12[	$\frac{9,5 + 12}{2} = 10,75$	4
		$n = 14$

## 1. Estatística Descritiva

### Mediana

A **mediana** é o ponto central das observações quando estas são colocadas por ordem crescente. Assim, é o valor que divide a amostra a meio, i.e., 50% dos elementos da amostra são menores ou iguais à mediana e os outros 50% são maiores ou iguais.

- ✓ Dados Discretos/Não Agrupados:

$$\tilde{x} = \begin{cases} \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ par} \\ x_{(\frac{n}{2}+1)}, & \text{se } n \text{ ímpar} \end{cases}$$

## 1. Estatística Descritiva

Cálculo da Mediana para Dados Discretos/Não Agrupados – Exemplo

*Considere os seguintes conjuntos de observações. Determine, em cada caso, a mediana.*

- a)  $\{5, 5, 7, 15, 16, 17, 24\}$
- b)  $\{18, 7, 6, 1, -6, -30\}$

*Resolução:*

- a) As 7 observações deste conjunto estão em ordem crescente. A mediana é  $\tilde{x} = x_{(4)} = 15$

*Ou seja, o valor da observação do meio (a 4.º das 7 observações).*

- a) As 6 observações deste conjunto estão em ordem decrescente. A mediana é  $\tilde{x} = \frac{x_{(3)} + x_{(4)}}{2} = \frac{1 + 6}{2} = 3,5$

## 1. Estatística Descritiva

### Cálculo da Mediana para Dados Agrupados por Classes Intervalares

1º) Identificar a **classe mediana ( $M_e$ )** cujo  $F_i \geq 0,5$  (frequência relativa acumulada).

2º) O valor da mediana é dado pela seguinte expressão:

$$\tilde{x} = LI_{Me} + ac_{Me} \times \frac{0,5 - F_{Me-1}}{f_{Me}}$$

Onde,

$LI_{Me}$  – limite inferior da classe mediana;

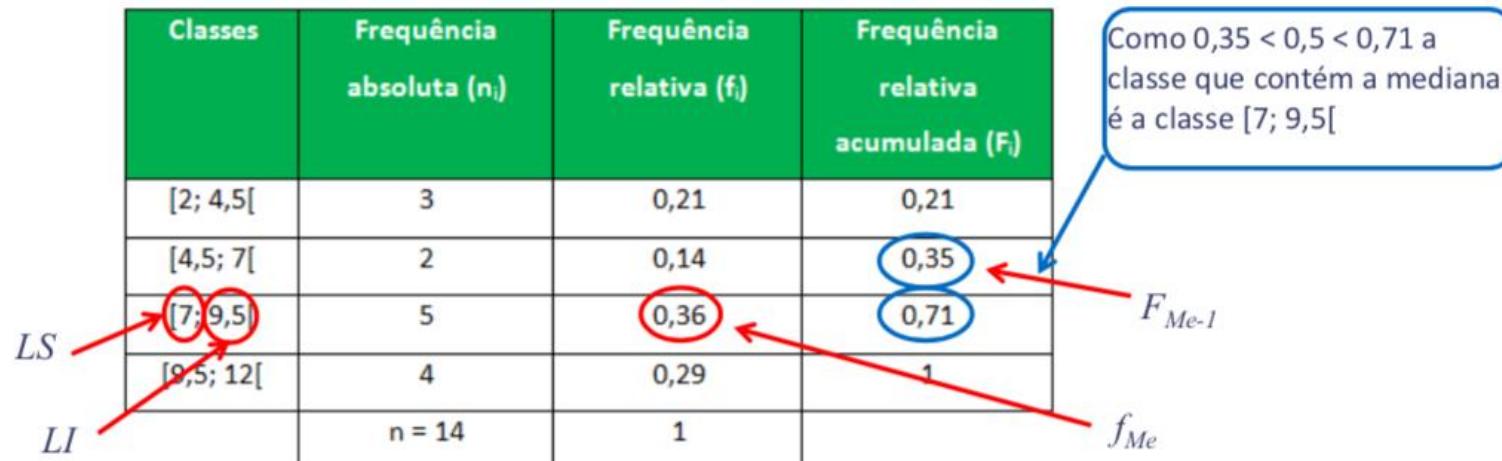
$ac_{Me}$  – amplitude da classe mediana;

$F_{Me-1}$  – freq. rel. acumulada da classe anterior à classe mediana;

$f_{Me}$  – freq. rel. da classe mediana.

## 1. Estatística Descritiva

Cálculo da Mediana para Dados Agrupados por Classes Intervalares – Exemplo



$$\text{Mediana: } \tilde{x} = 7 + \frac{0,5 - 0,35}{0,36} \times (9,5 - 7) \approx 8,04$$

## 1. Estatística Descritiva

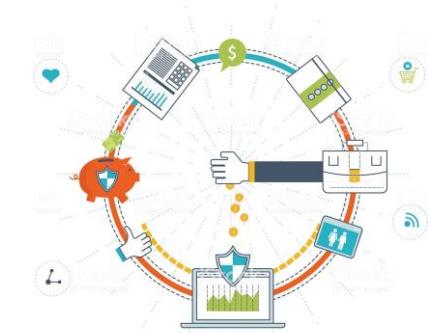
### Propriedades da Mediana

✓ Vantagens:

- > É pouco sensível a valores extremos (*outliers*). Sendo assim, é mais indicativa de valores típicos do que a média.

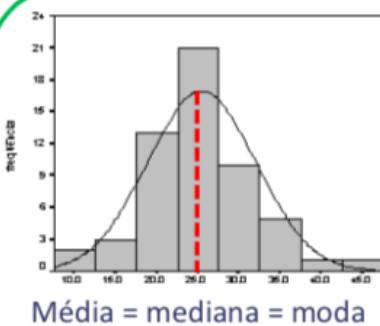
✓ Desvantagens:

- > Para calcular a mediana é necessário ordenar as observações, o que se pode tornar numa operação bastante complexa.
- > Tem pouco significado para um número pequeno de observações.



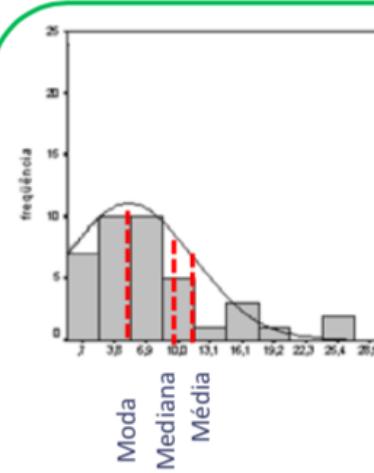
## 1. Estatística Descritiva

### Comparação de Distribuições – Média, Moda e Mediana



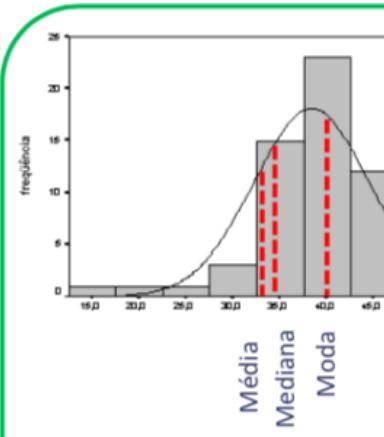
Distribuição simétrica

A média aproxima-se da mediana.



Distribuição assimétrica positiva

Moda < Mediana < Média.



Distribuição assimétrica negativa

Média < Mediana < Moda.

## Medidas de Localização – Tendência Não Central

## 1. Estatística Descritiva

### Quartis

Os **quartis** dividem um conjunto de dados dispostos por ordem crescente em quatro partes iguais.

- ✓ O **1º Quartil ( $Q_1$ )** é o valor que divide a amostra em duas partes de tal modo que 25% das observações sejam iguais ou inferiores a esse valor e 75% das observações sejam iguais ou superiores a esse valor.
- ✓ O **2º Quartil ( $Q_2$ )** é o valor que divide a amostra em duas partes de tal forma que 50% das observações sejam iguais ou inferiores a esse valor e 50% sejam iguais ou superiores ao valor estipulado.
- ✓ O **3º Quartil ( $Q_3$ )** é o valor que divide a amostra em duas partes de tal modo que 75% das observações sejam iguais ou inferiores a esse valor e 25% sejam iguais ou superiores a esse valor.
- ✓ O **4º Quartil ( $Q_4$ )** representa 100% das observações.

## 1. Estatística Descritiva

### Quartis – Exemplo

Consideremos um estudo feito numa portagem em que durante uma hora se contaram o número de pessoas que iam dentro de cada automóvel.

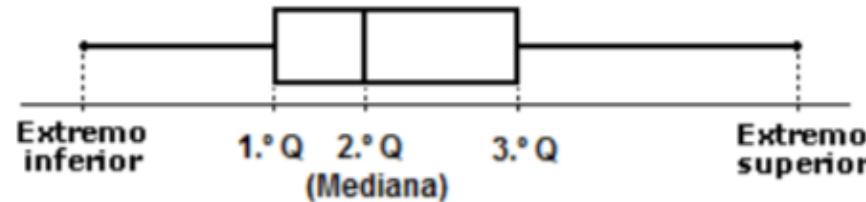
N.º de Pessoas Classes	Frequência absoluta ( $n_i$ )	Frequência relativa ( $f_i\%$ )	Frequência relativa acumulada ( $F_i\%$ )
1	20	33,33	33,33
2	15	25,00	58,33
3	10	16,67	75,00
4	10	16,67	91,67
5	5	8,33	100

← 1.º Quartil  
← 2.º Quartil  
← 3.º Quartil

## 1. Estatística Descritiva

### Diagrama de Extremos e Quartis (*Boxplot*)

- ✓ Para facilitar a comparação entre distribuições recorre-se, frequentemente, a um **diagrama de extremos e quartis**. Este diagrama corresponde a uma representação gráfica que permite avaliar a tendência central, dispersão e padrão de assimetria de um conjunto de dados com base em medidas resistentes.
- ✓ Neste sentido, são representados os valores máximo e mínimo da distribuição e os valores dos respetivos quartis. Começa-se por desenhar um eixo horizontal ou vertical e, com uma escala adequada aos dados, marcam-se os extremos e os quartis. Seguidamente, os retângulos entre quartis são construídos.



## 1. Estatística Descritiva

### Diagrama de Extremos e Quartis (*Boxplot*)

- ✓ O **diagrama de extremos e quartis** representa:

- > Uma medida de tendência central (ex: a mediana).
- > Um retângulo central que permite analisar a dispersão das observações típicas (ex: observações que se encontram compreendidas entre o 1º e 3º quartis).
- > Duas barras verticais que demonstram a totalidade ou maioria das observações (ex: amplitude total ou amplitude das observações não identificadas como *outliers*).

Atenção

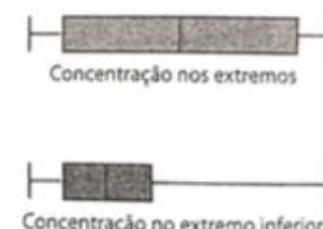
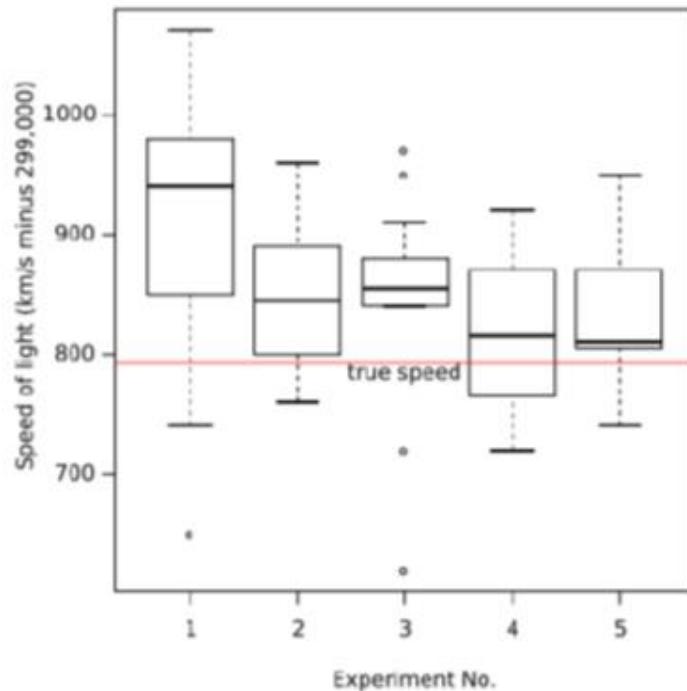


O padrão de assimetria para as observações típicas pode avaliar-se comparando as distâncias dos dois limites do retângulo central à medida de localização.

Valores "demasiado distantes" (ex: valores com uma distância à mediana superior a  $2,5 * \text{AIQ}$ ), por vezes, são considerados *outliers* e representados isoladamente.

## 1. Estatística Descritiva

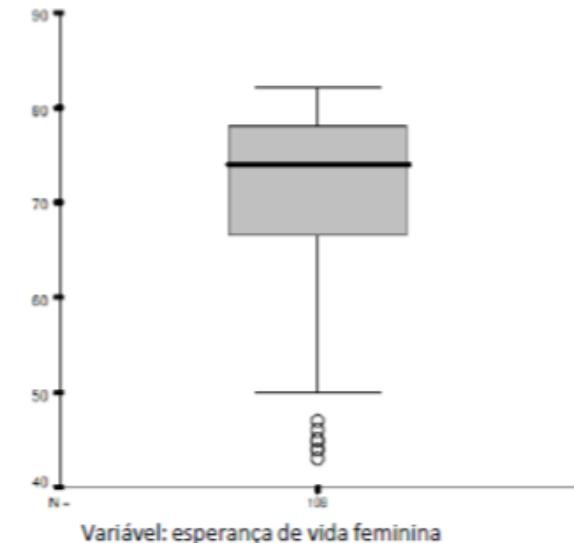
Diagrama de Extremos e Quartis (*Boxplot*)



## 1. Estatística Descritiva

### Diagrama de Extremos e Quartis (*Boxplot*) – Exemplo

- ✓ A figura contém a caixa de bigodes representativa da variável esperança média de vida feminina em 108 países do mundo em 1995.
- ✓ Podemos observar a existência de vários países com valores de esperança média de vida bastante reduzidos (inferiores a 50 anos de idade) e identificados com *outliers*.
- ✓ Uma análise posterior a estes países poderá facilitar a deteção de causas para uma esperança média de vida tão reduzida – são países onde os níveis de pobreza são muito elevados.
- ✓ Verifica-se, também, que a mediana da esperança média de vida feminina é de 74 anos pelo que metade dos países considerados apresentam esperanças de vida superiores a 74 anos.
- ✓ Por último, constata-se que a esperança de vida não vai além dos 82 anos.



## 1. Estatística Descritiva

### O que é um *Outlier*?

- ✓ Um *outlier* é uma observação anómala cujo valor está claramente em contradição com o padrão dominante dos restantes dados. Pode ser o resultado de erros na recolha dos dados podendo distorcer a sua interpretação e exercer uma influência excessiva sobre várias medidas sumárias, tais como a média e o desvio padrão.
  
- ✓ Quando um *outlier* corresponde a um valor corretamente observado pode ser particularmente importante e revelador, uma vez que sugere um comportamento extremo ou uma alteração no processo gerador de dados. Por essa razão os *outliers* devem ser cuidadosamente examinados antes de se recorrer à análise do conjunto completo de dados.

Os *outliers* não deverão ser eliminados automaticamente sem qualquer justificação!

## 1. Estatística Descritiva

### Percentis

O **percentil** corresponde aos valores da variável que dividem a distribuição em cem partes iguais.

✓ Interpretação:

- > O percentil ( $P_p$ ) divide a amostra em duas partes iguais.
- > Na primeira parte  $100 \times p\%$  dos elementos são menores ou iguais ao percentil.
- > Na segunda parte  $100 \times (p - 1)\%$  dos elementos são maiores ou iguais ao percentil.

✓ Observações:

- > 1º Quartil = Percentil 25
- > 2º Quartil = Percentil 50 = Mediana
- > 3º Quartil = Percentil 75

## 1. Estatística Descritiva

### Cálculo dos Percentis para Dados Agrupados por Classes Intervalares

- 1º) Identificar a classe cuja  $F_i = \text{Percentil}$ . A classe identificada será a **classe percentil P**, ou seja, é a classe que contém o percentil.
- 2º) O valor do percentil é dado pela seguinte expressão:

$$P_p = LI_p + ac_p \times \frac{p/100 - F_{p-1}}{f_p}$$

Onde,

$LI_p$  – limite inferior da classe percentil;

$ac_p$  – amplitude da classe percentil;

$F_{p-1}$  – frequência relativa acumulada da classe anterior à classe percentil;

$f_p$  – frequencia relativa da classe percentil.

## Medidas de Dispersão

## 1. Estatística Descritiva

### Medidas de Dispersão

- ✓ As **medidas de tendência central** são importantes mas não fornecem informação completa sobre o conjunto de valores. Deste modo, falta efetuar uma análise que nos dê informações acerca da dispersão desses valores.
- ✓ A forma como as observações se concentram ou dispersam em torno das medidas de tendência central fornecem algumas indicações acerca da distribuição dos dados. Assim, para melhor caracterizarmos uma distribuição devemos considerar medidas que exprimam o grau de dispersão ou variabilidade dos dados – **medidas de dispersão**.

As **Medidas de dispersão** descrevem a dispersão/concentração dos dados da amostra em torno dos valores centrais.

#### Medidas de dispersão

- Amplitude total
- Amplitude interquartil
- Variância
- Desvio padrão
- Coeficiente de dispersão
- Coeficiente de variação

## 1. Estatística Descritiva

### Medidas de Dispersão

A **amplitude total (A)** de um conjunto de valores é definida como a diferença entre a maior e a menor das observações medindo a dispersão total dos valores do conjunto.

$$A = \text{máximo} - \text{mínimo}$$

A **amplitude interquartil (AIQ)** é a diferença entre o 3º e 1º quartis. Fornece indicação quanto à forma como as observações se distribuem em torno da mediana. Assim, esta medida corresponde à amplitude do intervalo que engloba 50% das observações centrais.

$$AIQ = Q_3 - Q_1$$

- ✓ Se o valor da amplitude interquartil for pequena, os dados estão concentrados em torno dos valores centrais.
- ✓ Se o valor da amplitude interquartil for grande, a dispersão em torno dos valores centrais é grande.

## 1. Estatística Descritiva

### Medidas de Dispersão

A **variância ( $s^2$ )** corresponde à média dos quadrados dos desvios das observações em relação à média.

- ✓ Dados Discretos/ Não Agrupados:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Variância não corrigida

$$s'^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Variância corrigida

- ✓ Dados Agrupados:

$$s^2 = \frac{\sum_{i=1}^n n_i \cdot (x_i - \bar{x})^2}{n}$$

Variância não corrigida

$$s'^2 = \frac{\sum_{i=1}^n n_i \cdot (x_i - \bar{x})^2}{n-1}$$

Variância corrigida

Nota: Quando a amostra é de pequena dimensão deve calcular-se a **variância corrigida**.

## 1. Estatística Descritiva

### Medidas de Dispersão

O **desvio padrão (s)** indica a proximidade com que os valores estão agrupados à volta da média e, ao contrário da variância, está expresso nas mesmas unidades da variável.

- > Um valor pequeno do desvio padrão significa que as observações estão pouco espalhadas (pouco dispersas) à volta da média.
- > O desvio padrão é a raiz quadrada positiva da variância.

$$s = \sqrt{s^2}$$

Desvio padrão não corrigido

$$s' = \sqrt{s'^2}$$

Desvio padrão corrigido

O **coeficiente de dispersão (CD)** mede o grau de concentração de valores em torno da média.

$$CD = \frac{s}{\bar{x}}$$

## 1. Estatística Descritiva

### Medidas de Dispersão

O **coeficiente de variação (CV)** mede o grau de concentração de valores em torno da média (em percentagem).

$$CV = \frac{s}{\bar{x}} \times 100$$

- ✓ Estes coeficientes não dependem das unidades em que a variável é expressa pelo que são úteis para comparar duas ou mais distribuições relativamente à dispersão.
- ✓ Em termos práticos, é usual considerar-se que um **coeficiente de variação** superior a 50% indica alto grau de dispersão relativa e, consequentemente, uma pequena representatividade da média como medida estatística.
- ✓ Para valores do **coeficiente de variação** inferiores a 50% a média será tanto mais representativa quanto menor o valor desse coeficiente.

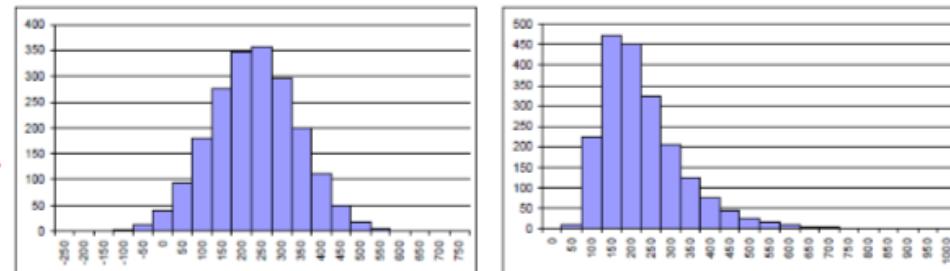
## Medidas de Assimetria

## 1. Estatística Descritiva

### Medidas de Assimetria

- ✓ As **medidas de tendência central** e as **medidas de dispersão**, embora forneçam informação importante, são insuficientes para uma caracterização da distribuição de frequências.
- ✓ Neste sentido, surgem as **medidas de assimetria** que permitem verificar até que ponto uma distribuição de frequências é enviesada ou assimétrica (sem necessitar de calcular a média, a moda e a mediana) permitindo comparar distribuições diferentes.

Duas Distribuições com a mesma média e desvio padrão, mas com formatos diferentes.

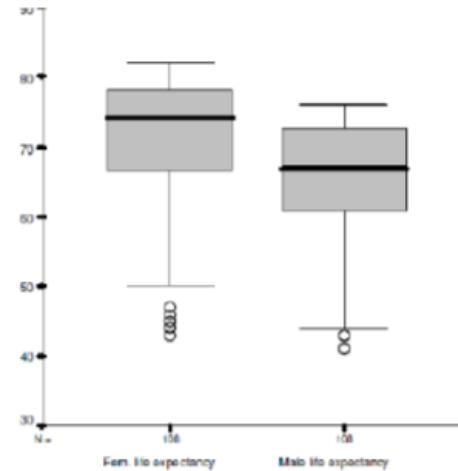


Média: 205

Desvio-Padrão: 109

## 1. Estatística Descritiva

### Medidas de Assimetria – Interpretação Gráfica



- ✓ A figura contém a caixa de bigodes paralelas (ou comparativas) das variáveis referentes à esperança média de vida masculina e feminina em 108 países em 1995.
- ✓ Verifica-se que a esperança média de vida masculina tende a ser inferior à esperança média de vida feminina (a caixa de bigodes encontra-se abaixo da feminina).
- ✓ Em ambos os casos deteta-se uma **assimetria negativa** havendo, desta forma, uma maior concentração de valores elevados.

## 1. Estatística Descritiva

### Medidas de Assimetria

O **coeficiente de assimetria de Pearson (gP)** é dado pela seguinte expressão:

$$g_P = \frac{\bar{x} - M_o}{s}$$

- ✓ Para  $gP \approx 0$  a distribuição é simétrica.
- ✓ Para  $gP \approx 3$  a distribuição é assimétrica positiva.
- ✓ Para  $gP \approx -3$  a distribuição é assimétrica negativa.

Nota: O grau de assimetria de Pearson só pode ser utilizado quando a **distribuição é unimodal**, ou seja, só tem uma moda.

O **SPSS** calcula outro coeficiente de assimetria que pode ser traduzido da seguinte forma:

$$G = \frac{Skewness}{Std. Error of Skewness}$$

Este coeficiente é designado por **Skewness (G)** e caracteriza a assimetria da função probabilidade tendo a seguinte interpretação:

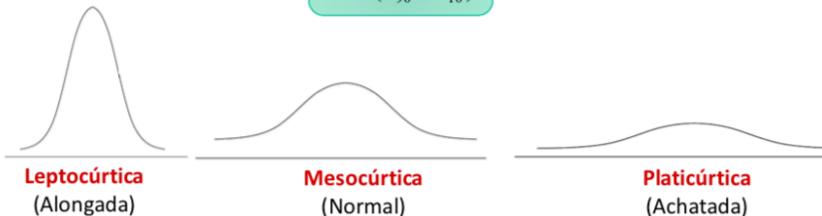
- ✓  $-1,96 < G < 1,96$  a distribuição é simétrica.
- ✓  $G > 1,96$  a distribuição é assimétrica positiva.
- ✓  $G < -1,96$  a distribuição é assimétrica negativa.

## 1. Estatística Descritiva

### Medidas de Assimetria

A **curtose (k)** indica o grau de achatamento de uma distribuição e pode ser calculada da seguinte forma:

$$k = \frac{Q_3 - Q_1}{2.(P_{90} - P_{10})}$$



- ✓ Para  $K < 0,263$  a distribuição é leptocúrtica.
- ✓ Para  $K = 0,263$  a distribuição é mesocúrtica.
- ✓ Para  $K > 0,263$  a distribuição é platicúrtica.

O **SPSS** calcula outro coeficiente de curtose (K) que pode ser traduzido da seguinte forma:

$$K = \frac{\text{Kurtosis}}{\text{Std.Error of Kurtosis}}$$

Este coeficiente é fornecido pelo *software* estatístico e caracteriza o grau de achatamento da função probabilidade tendo a seguinte interpretação:

- ✓  $K > 1,96$  a distribuição é leptocúrtica.
- ✓  $-1,96 < K < 1,96$  a distribuição é mesocúrtica.
- ✓  $K < -1,96$  a distribuição é platicúrtica.

## 1. Estatística Descritiva

Assimetria e Curtose – Exemplo do *Output* do SPSS

Statistics		
idade		
N	Valid	9
	Missing	1
Skewness		,590
Std. Error of Skewness		,717
Kurtosis		,569
Std. Error of Kurtosis		1,400

### Coeficiente de assimetria

$$G = \frac{Skewness}{Std. Error of Skewness}$$
$$= \frac{0,590}{0,717} \approx 0,823$$

Como  $-1,96 < G < 1,96$  a distribuição é simétrica.

### Coeficiente de Curtose

$$K = \frac{Kurtosis}{Std. Error of Kurtosis}$$
$$= \frac{0,569}{1,4} \approx 0,41$$

Como  $-1,96 < K < 1,96$  a distribuição é mesocúrtica.

## 1. Estatística Descritiva

Notação das Medidas Estatísticas

Medida Estatística	População (parâmetro)	Amostra (valor observado – estatística)
Dimensão	$N$	$n$
Média	$\mu$	$\bar{x}$
Proporção	$p$	$\hat{p}$
Variância	$\sigma^2$	$s^2$
Desvio padrão	$\sigma$	$s$
Coeficiente de correlação	$\rho$	$r$

Tabela 1 – Notação de algumas medidas estatísticas.

## Medidas de Associação

## 1. Estatística Descritiva

### Medidas de Associação de Variáveis

- ✓ As **medidas de associação** quantificam a intensidade e a direção da associação entre duas variáveis.
- ✓ Estas medidas, também designadas por **coeficientes de correlação**, medem pura e simplesmente a associação entre variáveis sem qualquer implicação de causa e efeito entre ambas.
- ✓ As correlações podem ser:
  - > Bivariadas: quando envolvem apenas duas variáveis.
  - > Multivariadas: se envolverem mais de duas variáveis.
- ✓ Depois de terem sido estudadas as variáveis estatísticas isoladamente (**distribuições unidimensionais**) agora vamos analisar duas variáveis em conjunto e verificar se existe ou não alguma relação entre elas. São as chamadas **distribuições bidimensionais**.
- ✓ Dependendo da natureza das variáveis existem diferentes **coeficientes de correlação** que se podem determinar:
  - > Pearson: ambas as variáveis são quantitativas.
  - > Spearman: as variáveis podem ser quantitativas ou ordinais.
  - > V de Cramer: ambas as variáveis são nominais.

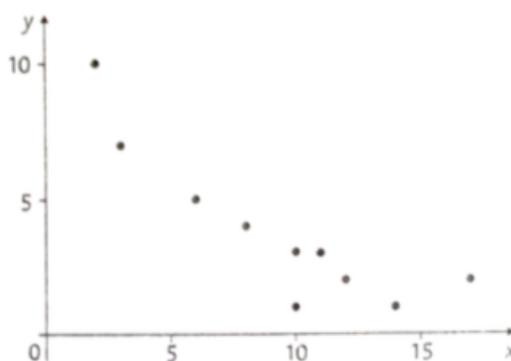
## 1. Estatística Descritiva

### Medidas de Associação de Variáveis – Exemplo

Num estudo efetuado pelo gabinete de qualidade de uma empresa, realizou-se um estudo com o objetivo de determinar a relação entre o número de anos de trabalho na empresa ( $x$ ) e o número de conformidades ( $y$ ) registadas no ano de 2014. Para esse efeito, foram selecionados, aleatoriamente, 10 funcionários e obtiveram-se os seguintes resultados:

x	10	3	12	11	6	8	14	17	10	2
y	1	7	2	3	5	4	1	2	3	10

O primeiro passo será representar graficamente os pares ordenados ( $x,y$ ) para visualizar a relação entre eles:



À representação das variáveis  $x$  e  $y$  num referencial cartesiano chamamos **diagrama de dispersão** ou **nuvem de pontos**. Este diagrama permite-nos visualizar se existe, ou não, relação entre as variáveis e identificar qual a condição mais adequada para descrever esta relação ou correlação.

Neste caso, como o número de não conformidades diminui à medida que o número de anos de trabalho aumenta estamos perante uma **correlação linear negativa**.

## 1. Estatística Descritiva

### Coeficiente de Correlação Linear

- ✓ Se todos os pontos do diagrama se localizarem na proximidade de uma reta, a correlação denomina-se **linear**.
- ✓ A correlação é **positiva** se ambas as variáveis variam no mesmo sentido (ou seja, se  $y$  aumentar a variável  $x$  também aumenta).
- ✓ A correlação é **negativa** se ambas as variáveis variam em sentidos opostos (isto é, se  $y$  diminuir,  $x$  deverá aumentar).



## 1. Estatística Descritiva

### Coeficiente de Correlação Linear

- ✓ Para medir o grau de associação linear entre duas variáveis, utiliza-se o **coeficiente de correlação linear** ou **coeficiente de correlação de Pearson**. Este valor, que se representa por  $r$ , varia entre -1 e 1.

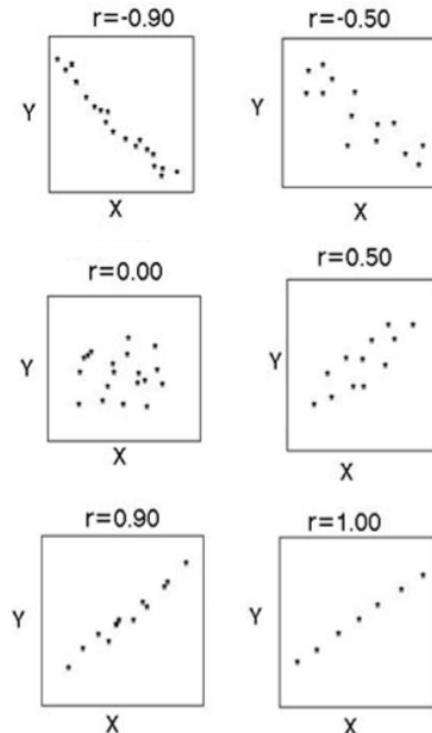
$$-1 \leq r \leq 1$$

- ✓ Quando a correlação é perfeita e negativa, o coeficiente toma o valor de -1.
- ✓ Se a correlação é perfeita e positiva, o coeficiente toma o valor de 1.
- ✓ A valores próximos de zero corresponde uma correlação fraca.

Em termos amostrais utiliza-se o símbolo  $r$ , em termos populacionais utiliza-se a letra grega  $\rho$  (rho).

## 1. Estatística Descritiva

### Coeficiente de Correlação Linear



Coeficiente de Correlação	Correlação
$r = 1$	Perfeita positiva
$0.8 \leq r < 1$	Forte positiva
$0.5 \leq r < 0.8$	Moderada positiva
$0.1 \leq r < 0.5$	Fraca positiva
$0 \leq r < 0.1$	Ínfima positiva
$r = 0$	Nula
$-0.1 < r < 0$	Ínfima negativa
$-0.5 < r \leq -0.1$	Fraca negativa
$-0.8 < r \leq -0.5$	Moderada negativa
$-1 < r \leq -0.8$	Forte negativa
$r = -1$	Perfeita negativa

## 1. Estatística Descritiva

### Coeficiente de Correlação Linear e Covariância

- ✓ O **coeficiente de correlação de Pearson** é calculado a partir da covariância sendo dado por:

$$r = \frac{s_{xy}}{s_x s_y},$$

- ✓ Onde:

>  $s_{xy}$  é a **covariância amostral** definida por:

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left( \left[ \sum_{i=1}^n x_i y_i \right] - n \bar{x} \bar{y} \right),$$

onde  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  e  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ .

>  $s_x$  e  $s_y$  representam, respetivamente, o **desvio padrão** das variáveis x e y definidos por:

Falta a barra nas médias!

Exemplo:

$x=(2,2,3,5)$

$s=1,5$  por ambas as fórmulas!

$$s_x = \sqrt{\frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 \right) - n \bar{x}^2}$$

$$s_y = \sqrt{\frac{1}{n-1} \left( \sum_{i=1}^n y_i^2 \right) - n \bar{y}^2}$$

## 1. Estatística Descritiva

### Coeficiente de Correlação Linear e Covariância – Exemplo

Considerando o exemplo evidenciado no slide 78 e sabendo que existe uma correlação linear negativa entre as variáveis x e y, vamos calcular o valor do coeficiente de correlação linear ( $r$ ).

x	10	3	12	11	6	8	14	17	10	2
y	1	7	2	3	5	4	1	2	3	10

Por colunas:

X, Y (somas);

Produtos  $xi \cdot yi$  (Soma)

Quadrados  $xi$  (Soma)

Quadrados  $yi$  (Soma)

média X

média Y

SX

SY

SXY

Classificação

Resolução:

$$\sum_{i=1}^{10} x_i = 93; \quad \sum_{i=1}^{10} y_i = 38; \quad \sum_{i=1}^{10} x_i y_i = 248; \quad \sum_{i=1}^{10} x_i^2 = 1063; \quad \sum_{i=1}^{10} y_i^2 = 218;$$

$$\bar{x} = \frac{93}{10} = 9.3; \quad \bar{y} = \frac{38}{10} = 3.8; \quad s_x \approx 4.692; \quad s_y \approx 2.86;$$

$$s_{xy} = \frac{1}{n-1} \left( \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right) = \frac{1}{10-1} (248 - 10 \times 9.3 \times 3.8) \approx -11.711;$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{-11.711}{4.692 \times 2.86} \approx -0.873 (\text{Forte correlação negativa})$$

Variáveis	Tipo	Valores	Exemplos	Gráficos	Estatísticas
Qualitativas	Nominais	Conjunto de nomes (sem ordem)	Cor de olhos; Concelho; Género; Situação profissional; (...)	Circular; Barras;	Frequências Absolutas ou relativas; Moda;
	Ordinais	Características mensuráveis (com ordem)	Escalas de Likert; Concordância Preferências; Intensidade da dor; (...)	Barras; Diagrama de extremos e quartis;	Frequências Absolutas ou relativas; Moda; Mediana; Quantis; Amplitude amostral; Amplitude Interquartil;
Quantitativas	Discretas	Conjunto numérico finito ou infinito numerável	Número de pessoas; Quantidades; (...)	Barras; Diagrama de extremos e quartis;	Frequências Absolutas ou relativas; Moda; Média,
	Contínuas	Conjunto numérico dentro de um determinado intervalo ou conjunto infinito não numerável	Tempo; Peso; Comprimento; (...)	Histogramas; Diagrama de extremos e quartis;	Mediana; Quantis; Amplitude amostral; Amplitude Interquartil; Variânci;a, Dispersão; Assimetria; Achatamento; Outliers

# Estatísticas em R

Contagens	Frequências Absolutas (unidades); Frequências Relativas (percentagens);	table(variavel) prop.table(table(variavel))
Localização Central:	Moda - mais frequente; Média - valor em torno do qual a distribuição se encontra (menos resistente)  Mediana - por ordem crescente é valor do meio localiza 50% (mais resistente à influência de <b>outliers</b> );	table(variavel) mean(variavel)  median(variavel) Ou quantile(variavel, 0.5)
Localização Não central:	Quantis - dividem a amostra em n partes iguais;  Quartis - 4 partes iguais a 25%: Ex.: O 2º quartil é a mediana, o terceiro acumula 75%  Decis - 10 partes iguais a 10%;  Percentis - 100 partes iguais a 1%;	summary(variavel) (quartis e média)  quantile(variavel) (quartis )  quantile(variavel, 0.85) (Percentil 85)

# Medidas de Dispersão e Forma

Tipo		Comandos
Dispersão:	Amplitude amostral: Max-Min (100%); Amplitude Interquartil: Q3-Q1 (50%);  Variância; Desvio-padrão; ...	<code>max(variavel) -min(variavel)</code>  <code>var(variavel)</code> <code>sd(variavel)</code>
Assimetria	Simetria estuda Viés e caudas; Positiva: média>moda Negativa: média<moda	<code>skewness(variavel)</code>
Curtose	Compara o achatamento provocado pela variância Mesocúrtica: igual à Normal Leptocúrtica: ponteaguda Platicúrtica: mais plana	<code>kurtosis(variavel)</code>
Outliers	Moderados; Severos;	<code>boxplot.stats(variavel)</code>

# Gráficos em R

Gráficos:		Comandos R
Circular	Variáveis qualitativas nominais (grupos!!!)	<code>pie(table(variavel))</code>
Barras	Variáveis qualitativas nominais Variáveis qualitativas ordinais Variáveis quantitativas discretas	<code>barplot(table(variavel))</code> <code>boxplot(variavel)</code>
histograma	Variáveis quantitativas contínuas	<code>hist(variavel)</code> <code>boxplot(variavel)</code>
Comparação entre grupos	variáveis numéricas por grupos	<code>boxplot(variavel_num ~ var_grupos)</code> <code>tapply(var_num, var_grupos, summary)</code> <code>table(genero, idade)</code>

# Cruzamentos

	Nominal	Ordinal	Discreta	Contínua
Nominal	Tabela Dupla entrada Gráficos empilhados			Boxplot Múltiplo
Ordinal	Tabela Dupla entrada Boxplot Múltiplo			
Discreta		Boxplot Múltiplo		Nuvem de Pontos
Contínua				