



Lisbon School  
of Economics  
& Management  
Universidade de Lisboa

# Estatística I

Licenciatura em Gestão do Desporto (LGD)

2.º Ano/1.º Semestre

2023/2024

# Aulas Teórico-Práticas N.ºs 1 e 2 (Semana 1)

**Docente:** Elisabete Fernandes

**E-mail:** efernandes@iseg.ulisboa.pt



<https://doity.com.br/estatistica-aplicada-a-nutricao>



<https://basiccode.com.br/produto/informatica-basica/>

# Conteúdos Programáticos

Aulas TP  
(Semanas 1 e 3)

- **Capítulo 1:** Análise Descritiva
- **Capítulo 2:** Probabilidades

Aulas TP  
(Semanas 3 a 6)

- **Capítulo 3:** Variáveis Aleatórias Unidimensionais

Aulas TP  
(Semanas 7 a 9)

- **Capítulo 4:** Variáveis Aleatórias Multidimensionais

Aulas TP  
(Semanas 10 a 12)

- **Capítulo 5:** Variáveis Aleatórias Especiais

**Material didático:** Exercícios do Livro Murteira et al (2015), Formulário e Tabelas Estatísticas

**Bibliografia:** B. Murteira, C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta e F. Pimenta; *Introdução à Estatística*, 2ª ed., Escolar Editora, 2015.

<https://cas.iseg.ulisboa.pt>

## **1. Introdução à Estatística**

### 1.1. Estatística Descritiva

## **2. Probabilidades**

### 2.1. Introdução

### 2.2. Experiência aleatória. Espaço de resultados. Acontecimentos

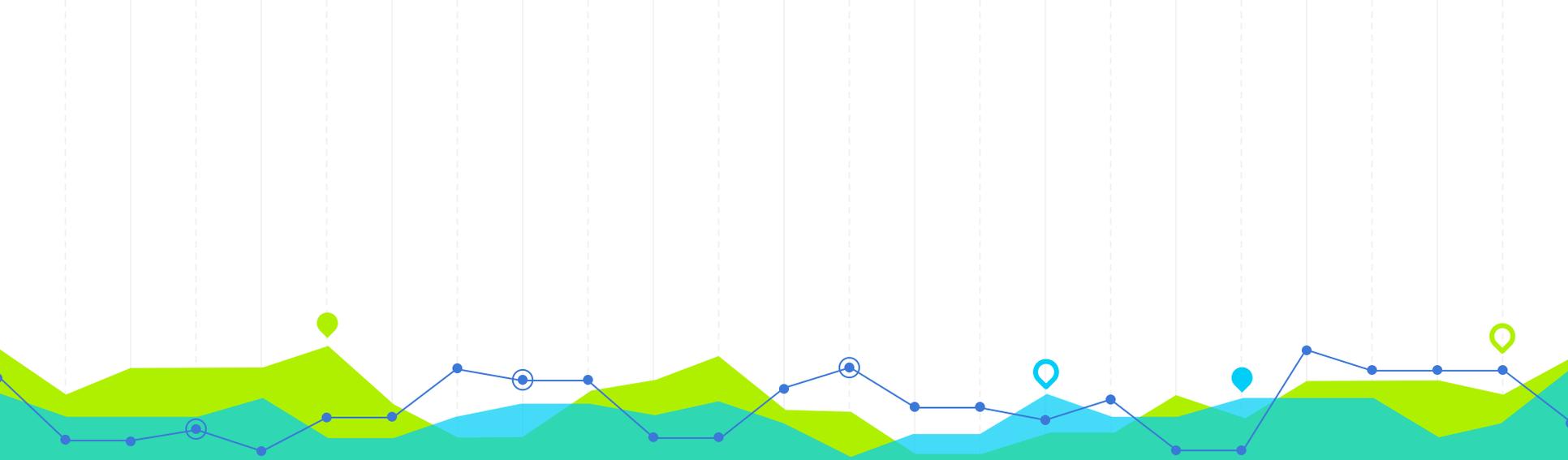
### 2.3. Medida de probabilidade. Axiomática de Kolmogorov

### 2.4. Interpretações do conceito de probabilidade

### 2.5. Probabilidade condicionada.

### 2.6. Teorema da probabilidade total e teorema de Bayes

### 2.7. Acontecimentos independentes



# O Que é a Estatística?

1



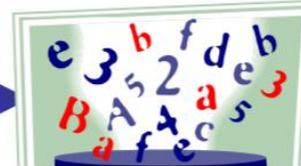
A **Estatística** pode ser definida  
como um conjunto de métodos  
para **Recolha, Análise** e  
**Interpretação** de DADOS.



Perguntas



Estudos



Dados



Informação



Respostas

**Estatística**

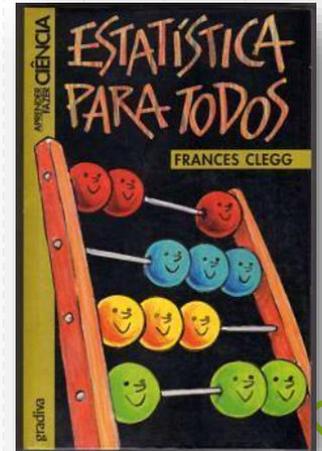
<http://www.est.ufmg.br/~edna/bionutri/NUT-Aula01.pdf>



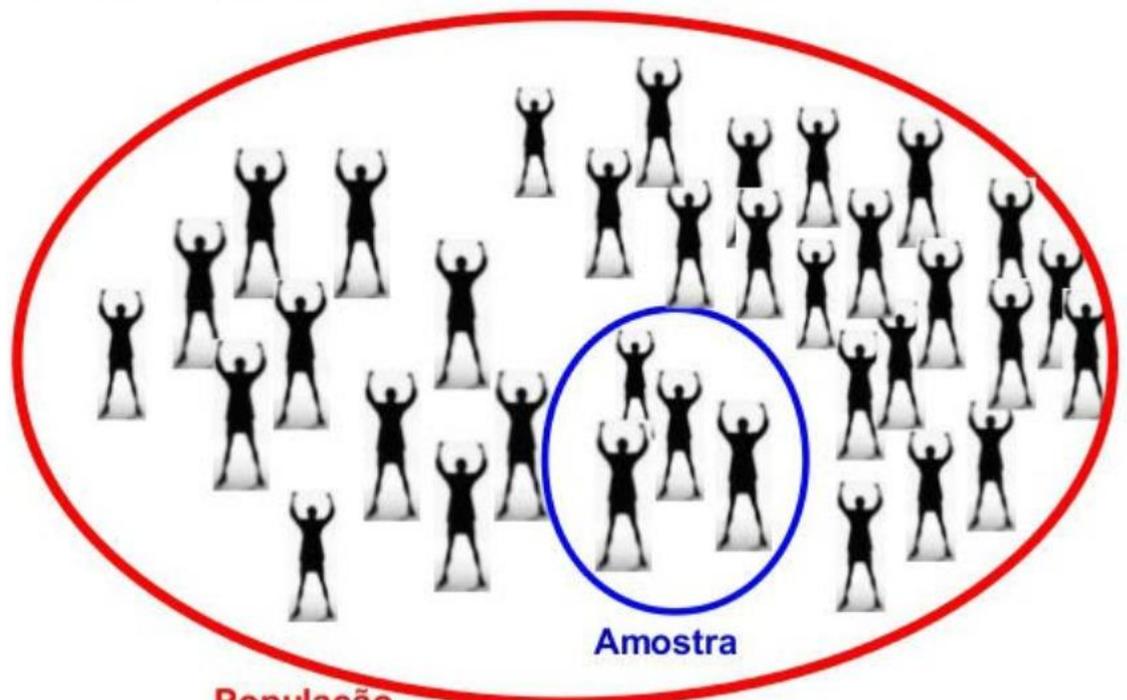


Os *símbolos estatísticos* são semelhantes aos símbolos usados em qualquer outro idioma. Para adquirir fluência num idioma necessitamos apenas de Tempo, Esforço e Prática.

A Estatística é para todos!



<https://www.custojusto.pt>



População

Amostra

<https://sites.google.com/site/estatisticabasicacc/conteudo/>

➤ **POPULAÇÃO:**

• é uma coleção completa de todos os elementos a serem estudados

➤ **AMOSTRA:**

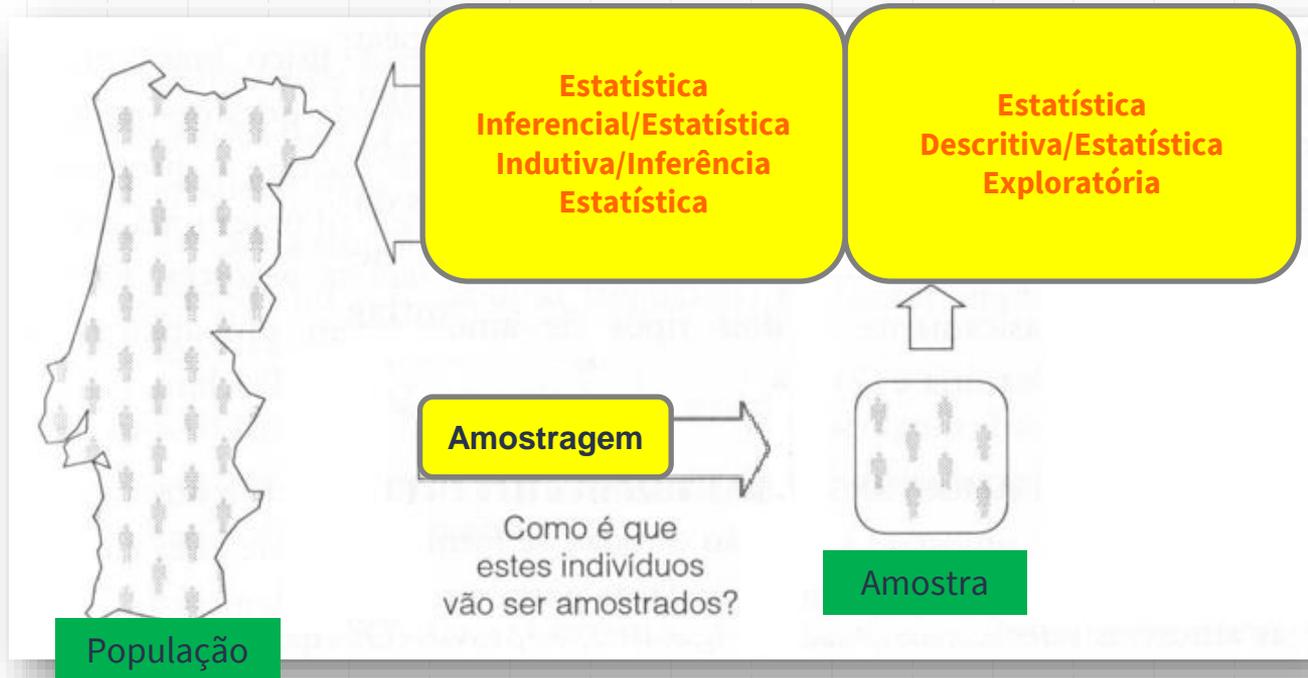
• é um subconjunto da população

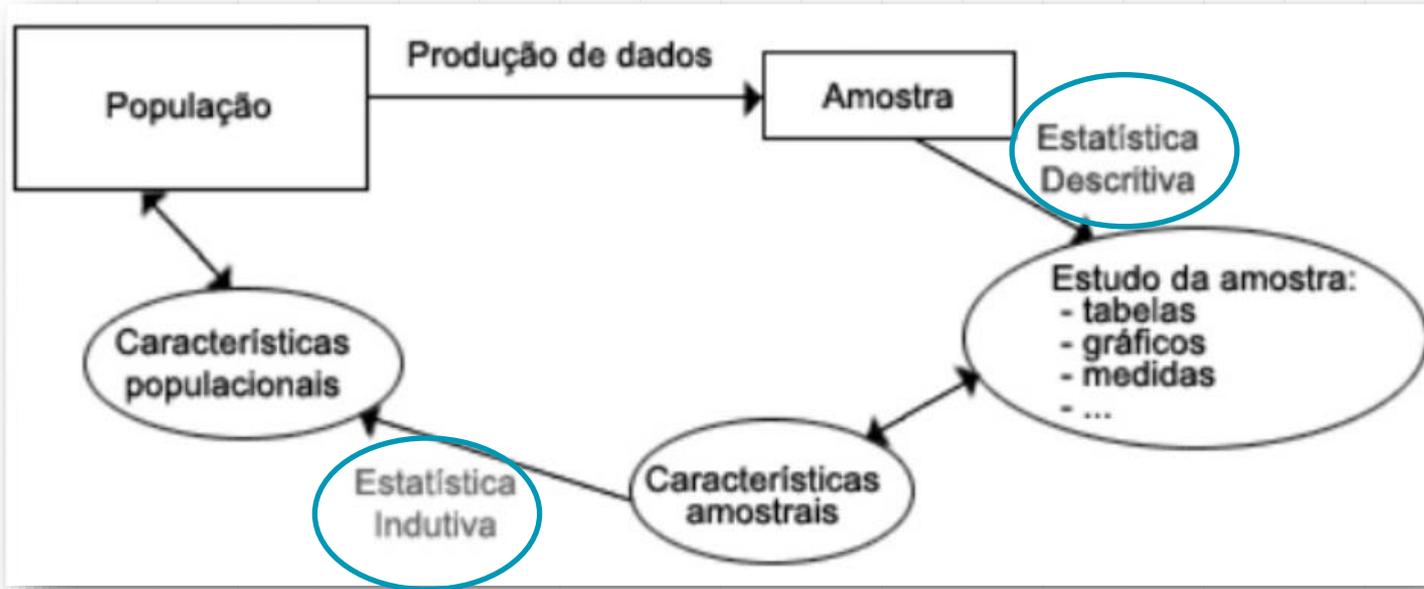
➤ **CENSO:**

• é uma coleção de dados relativos a todos os elementos de uma população:

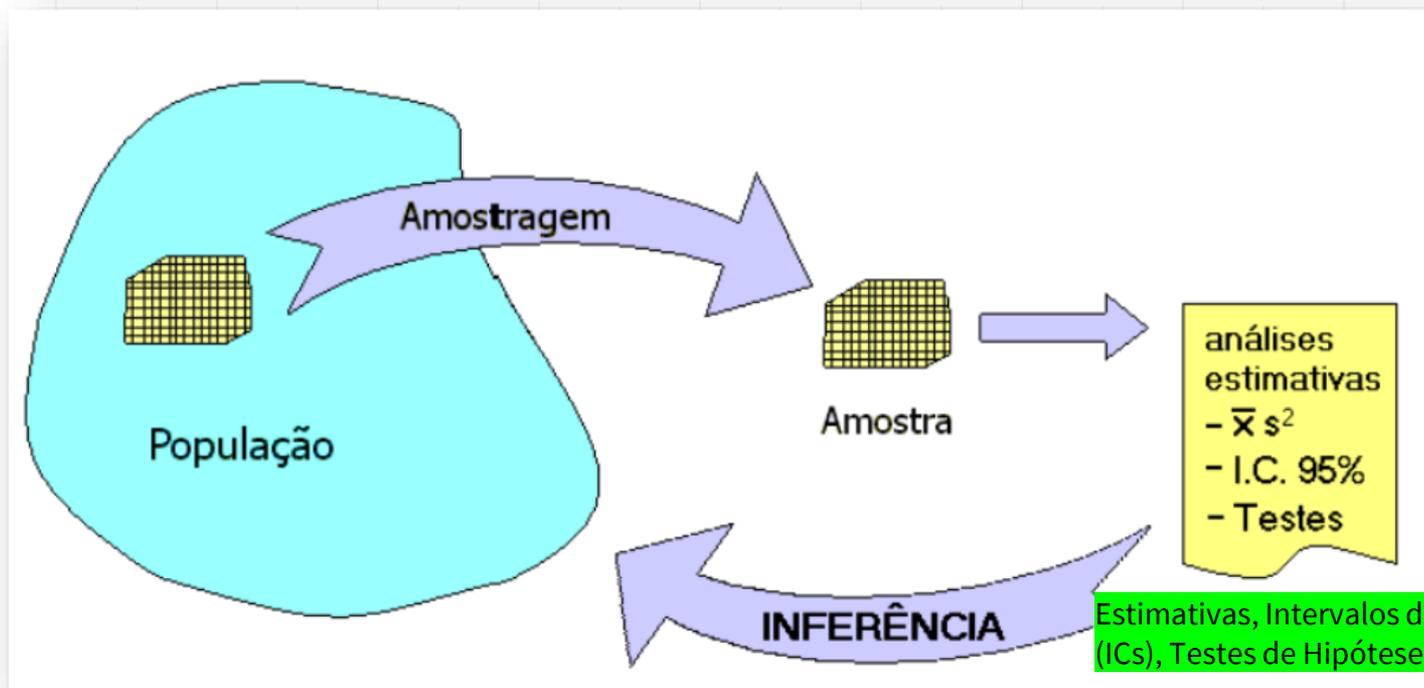
<https://slideplayer.com.br/slide/2627398>

# Estadística Exploratória vs. Estatística Inferencial





<https://umolharmatematico.weebly.com/descritivaindutiva.html>



Estimativas, Intervalos de Confiança (ICs), Testes de Hipóteses...

<https://sites.google.com/site/estatisticabasicacc>

# Análise Exploratória de Dados Univariada e Bivariada

## ▶ Análise Univariada

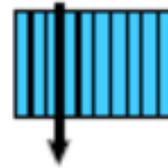
- Cada variável é tratada isoladamente e deve ser o primeiro passo de exploração dos dados.
- **Exemplo:** Estudo sobre a opinião dos alunos do ISEG sobre o package R.

## ▶ Análise Bivariada

- Estabelecem-se relações entre duas variáveis.
- **Exemplo:** A opinião dos alunos do ISEG sobre o package R consoante o género.

Análise univariada, tab. simples

*Descrever uma variável por vez*

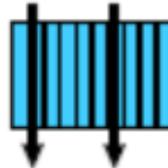


Quem?



Análise bivariada, tab. cruzada

*Relacionar 2 variáveis,  
explicar*

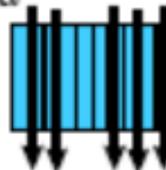


Quem?  
O quê?



Análise multivariada, tab. múltipla

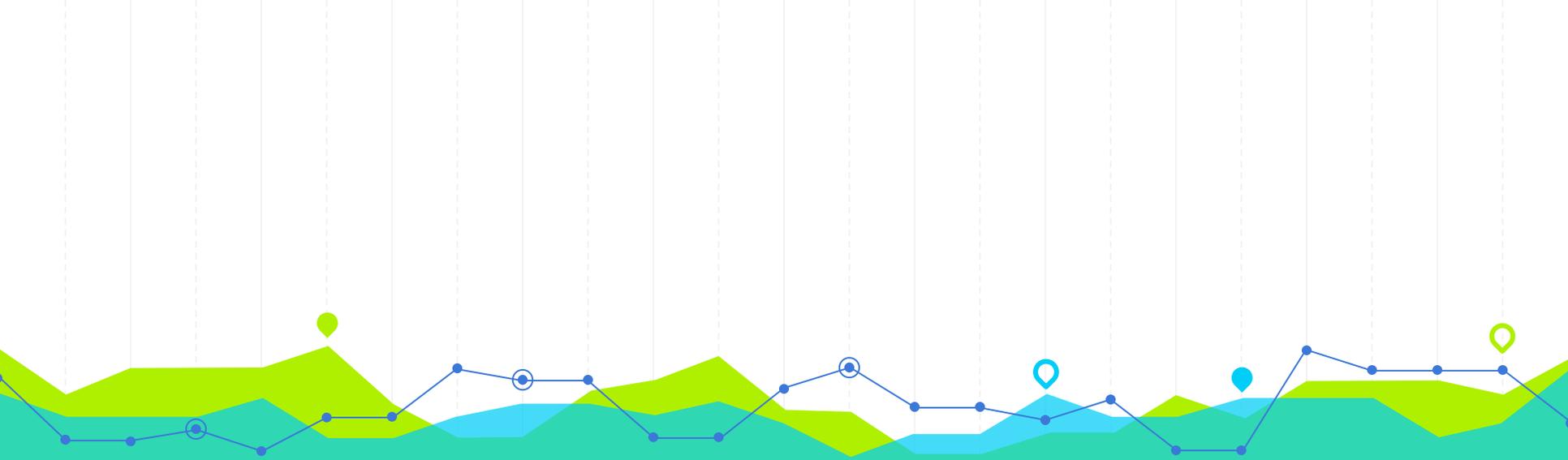
*Analisar simultaneamente  
diversas variáveis, sintetizar*



Quem?  
O quê?  
Quando?  
Por quê?



<https://www.scielo.br>



# Análise Descritiva de Dados

# 2

# Classificação de Variáveis Aleatórias



## Exemplo: Escalas

a) Consome doces na última semana antes de uma prova?

Sim

Não

b) Se a sua resposta à alínea a) é afirmativa, com que frequência consome doces?

1 a 2 dias

3 a 4 dias

Mais de 4 dias

c) Se consome doces na última semana antes de uma prova, com que frequência os consome (em dias)?



# Exemplo: Classificação de Variáveis

Altura de um aluno

Contínua

N.º de respostas incorretas num exame de Estatística

Discreta

Tempo de espera para uma consulta de nutrição

Contínua

Classificação obtida num trabalho de Estatística

Discreta/Contínua

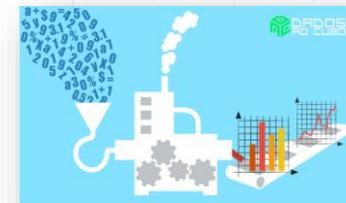
Nota final do aluno na disciplina de Estatística

Discreta/Contínua

▶ N.º de exames médicos realizados durante um ano por uma pessoa

• Discreta

# Análise Exploratória Univariada



## ☐ Variáveis medidas numa escala nominal ou ordinal

- **Estatísticas:** Moda, Estatísticas de Ordem ou Quantis (quartis, decis, percentis), Amplitude Total ou Range e Amplitude interquartil ou Intervalo Interquartil (estas três últimas apenas para as variáveis ordinais e tendo pelo menos 5 classes)
- **Gráficos:** Gráfico de Barras, Diagrama Circular e Boxplot ou Gráfico de Caixa-de-Bigodes com a identificação de outliers (este último apenas para as variáveis ordinais com pelo menos 5 classes)
- **Tabelas de Frequências**

## ☐ Variáveis medidas numa escala métrica

- **Estatísticas:** Média, Moda, Estatísticas de Ordem, Amplitude Total ou Range, Amplitude ou Intervalo Interquartil, Variância ( $s^2$ ), Desvio Padrão ( $s$ ), Coeficiente de Variação ( $CV=s/Média$ ), Medida de Assimetria do SPSS (“Skewness/Std error”), Coeficiente de Assimetria de Pearson e Medida de Achatamento do SPSS (“Kurtosis/Std error”)
- **Gráficos:** Histograma (apenas para as variáveis contínuas), Gráfico de Barras (apenas para as variáveis discretas), Boxplot com a identificação de outliers e Diagrama de caule-e-folhas
- **Tabelas de Frequências** (com classes)

# Medidas Amostrais

As Medidas Amostrais ou estatísticas permitem resumir os dados através de um só valor.

## Medidas de Localização/Tendência

- **Central:** Moda, Mediana, Média
- **Não central ou relativa:** Alguns Quantis (Quartis, Decis, Percentis, Mínimo, Máximo)

## Medidas de Dispersão

- Amplitude Total ou Amostral, Amplitude Interquartil, Variância, Desvio Padrão, Coeficiente de Variação

## Medidas de Assimetria

- Coeficientes de Assimetria

## Medidas de Achatamento/Curtose/Forma

- Coeficientes de Achatamento

O **quantil**  $Q_p^*$  divide a amostra em duas partes tais que:

- Na 1ª parte  $100 \times p\%$  dos elementos são menores ou iguais a  $Q_p^*$ ;
- Na 2ª parte  $100 \times (1 - p)\%$  dos elementos são maiores ou iguais a  $Q_p^*$ ;

$$Q_p^* = \begin{cases} \text{Quartis} = Q_i, & p = \frac{i}{4}, i = 1, 2, 3, 4 \\ \text{Decis} = D_i, & p = \frac{i}{10}, i = 1, 2, \dots, 9, 10 \\ \text{Percentis} = P_i, & p = \frac{i}{100}, i = 1, 2, \dots, 99, 100 \end{cases}$$

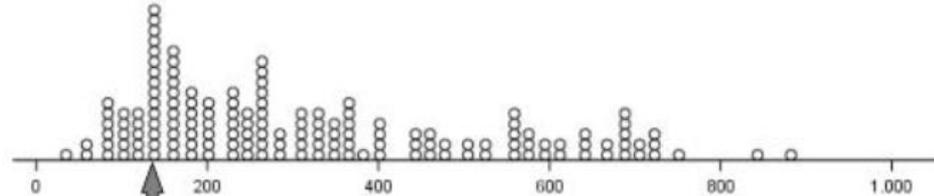


$$q_\alpha = \begin{cases} x_{[n\alpha]+1:n} & \text{se } n\alpha \text{ não inteiro} \\ \frac{x_{n\alpha:n} + x_{n\alpha+1:n}}{2} & \text{se } n\alpha \text{ inteiro} \end{cases}$$

# Medidas de Localização

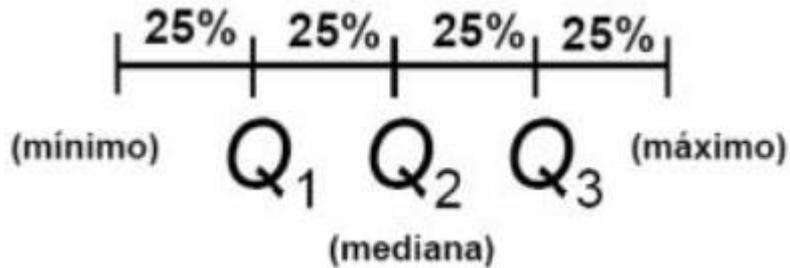
MEDIA:

$$\sum_{i=1}^n \frac{x_i}{n}$$



Moda

<https://www.slideserve.com/mort/estat-stica-descritiva>



**Quartis** – são os valores ( $Q_1$ ,  $Q_2$  e  $Q_3$ ) que dividem a amostra, depois de ordenada, em quatro partes iguais (ou o mais iguais possível).  $Q_2$  coincide com a mediana.

# Medidas de Dispersão

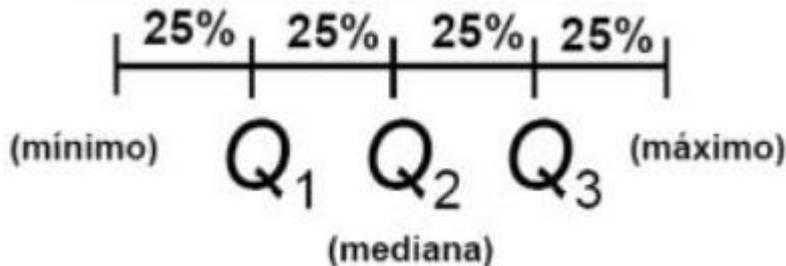
DESVIO PADRÃO

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

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

Amplitude Amostral = Máximo - Mínimo

<https://www.slideserve.com/mort/estat-stica-descritiva>



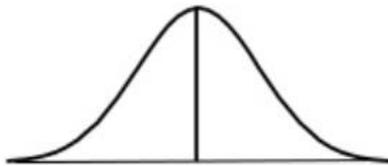
Intervalo interquartil é  $IQ = (Q_3 - Q_1)$

COEFICIENTE de  
VARIAÇÃO

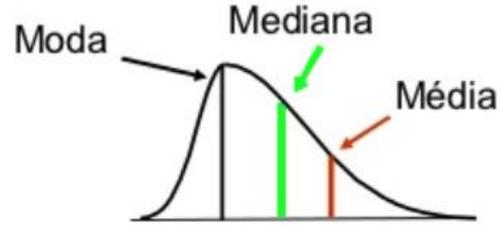
$$\frac{s}{\bar{x}} \times 100$$

# Assimetria/Enviesamento

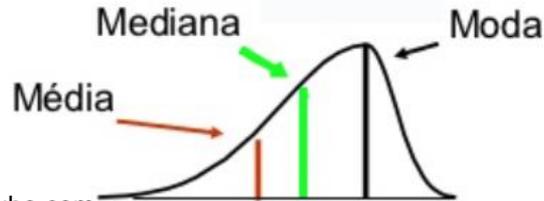
**Distribuição Simétrica**  
Média = Mediana = Moda



**Assimetria à direita ou positiva**



**Assimetria à esquerda ou negativa**



<https://dadosaocubo.com>

Curiosidade

**Coefficiente de Assimetria do SPSS e Jamovi:**

- ❖  $\text{Skewness/Std error} < -2 \Rightarrow$  **Assimétrica negativa**
- ❖  $-2 < \text{Skewness/Std error} < 2 \Rightarrow$  **Simétrica**
- ❖  $\text{Skewness/Std error} > 2 \Rightarrow$  **Assimétrica positiva**

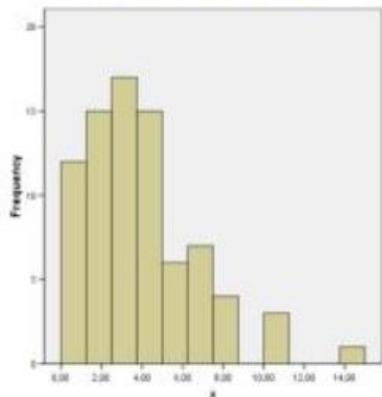
# Coeficiente de Assimetria/Enviesamento

COEFICIENTE de ASSIMETRIA  
de PEARSON

$$\frac{\bar{x} - M_0}{s}$$

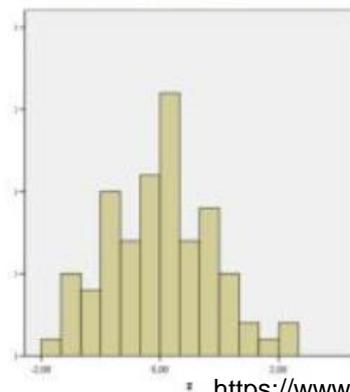
Assimetria positiva

Coef.ass. >0



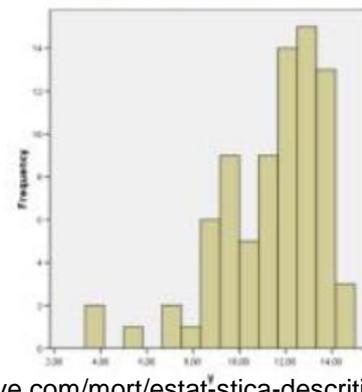
Quase simetria

Coef.ass. ~ 0



Assimetria negativa

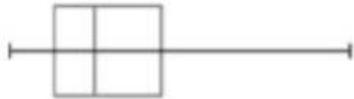
Coef.ass. <0



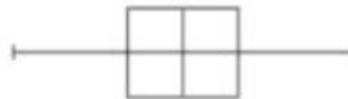
<https://www.slideserve.com/mort/estat-stica-descritiva>

# Assimetria/Enviesamento

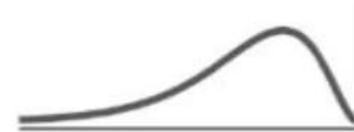
Assimetria positiva



Simetria



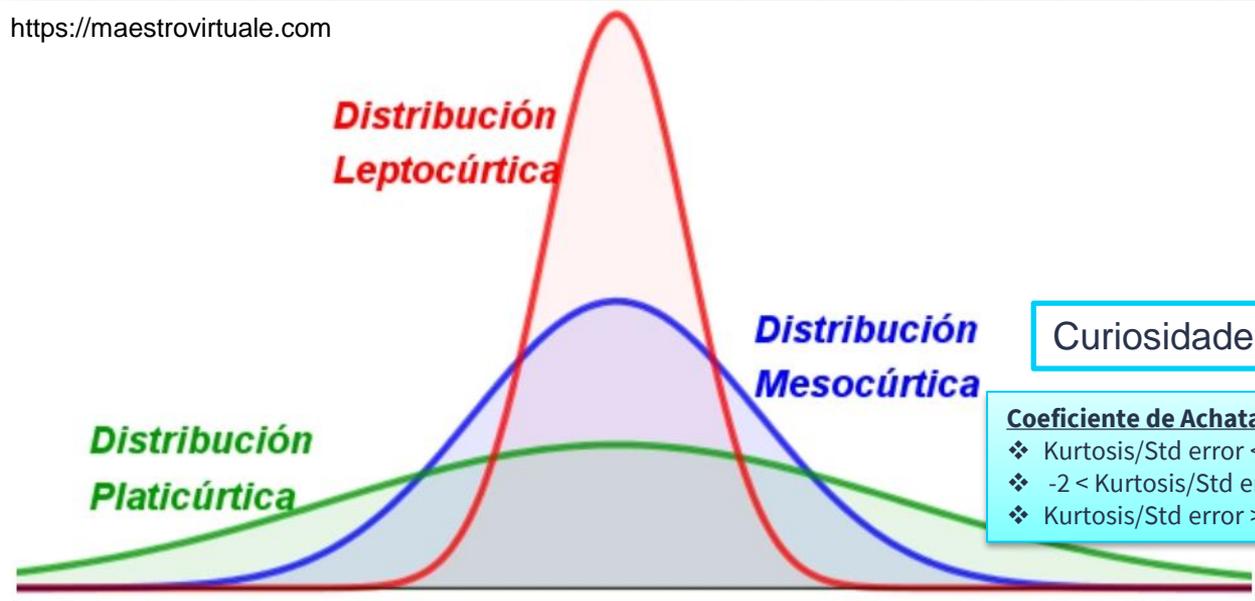
Assimetria negativa



<https://www.slideserve.com/mort/estat-stica-descriptiva>

# Achatamento/Curtose/Forma

<https://maestrovirtuale.com>



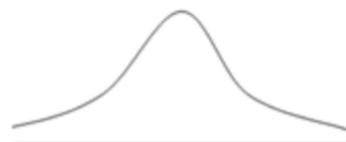
Curiosidade

**Coefficiente de Achatamento do SPSS e Jamovi:**

- ❖  $Kurtosis/Std\ error < -2 \Rightarrow$  **Platicurtica**
- ❖  $-2 < Kurtosis/Std\ error < 2 \Rightarrow$  **Mesocurtica**
- ❖  $Kurtosis/Std\ error > 2 \Rightarrow$  **Leptocurtica**

# Coefficiente de Achatamento/Curtose/Forma

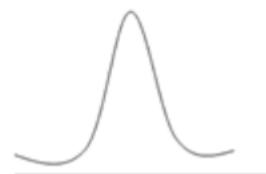
$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$



Mesocúrtica  $C = 0,263$



Platicúrtica  $C > 0,263$



Leptocúrtica  $C < 0,263$

# Gráfico de Caixa-de-Bigodes ou Boxpot

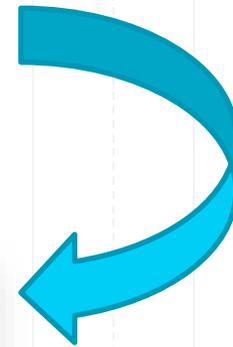
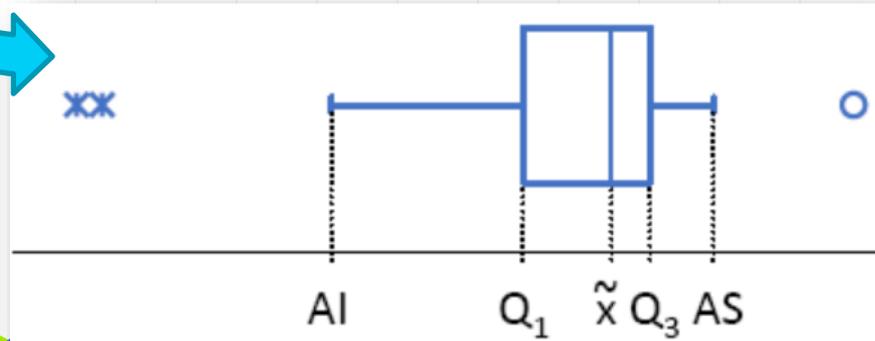
Intervalo interquartil é  $IQ=(Q_3-Q_1)$

- Outliers moderados (marcados com um círculo)

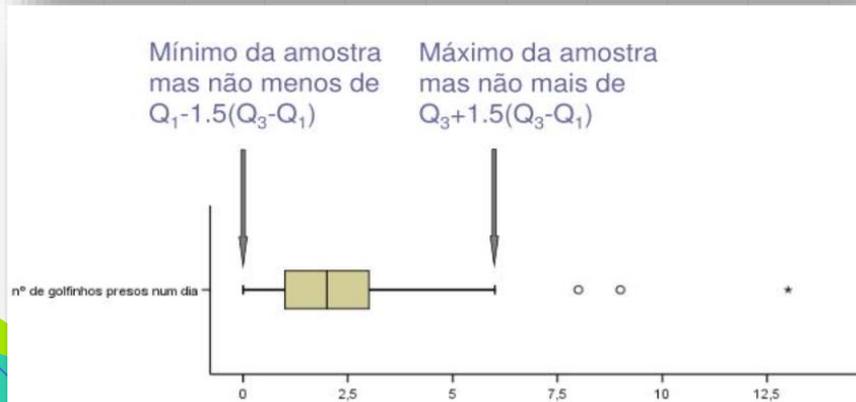
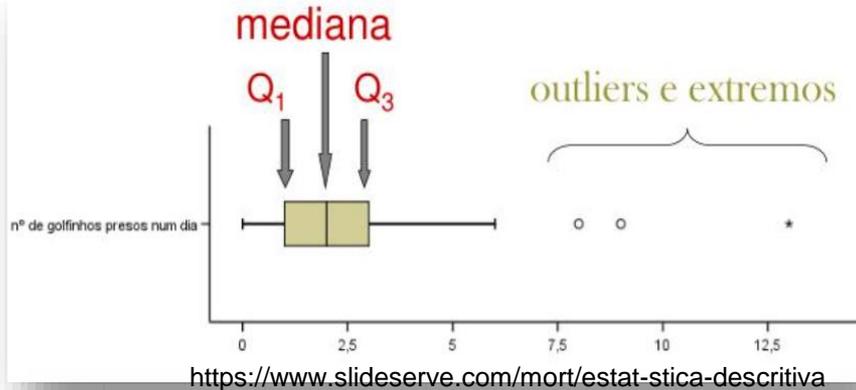
BI:  $(Q_1-1,5 \times IQ; Q_3+1,5 \times IQ)$  Barreiras interiores

- Outliers severos (marcados com um asterisco)

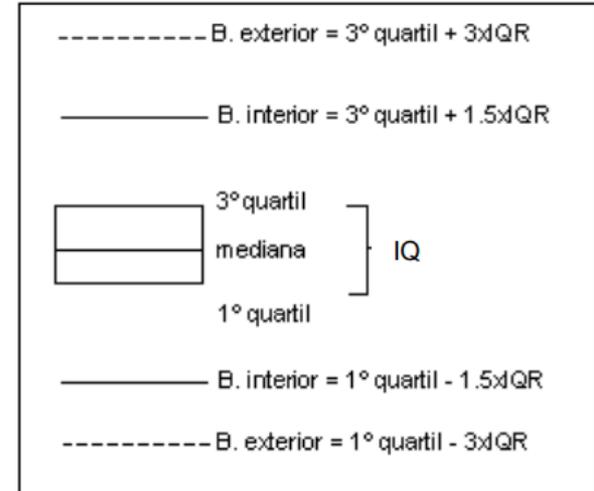
BE:  $(Q_1-3 \times IQ; Q_3+3 \times IQ)$  Barreiras exteriores



# Boxplots e Outliers



## As barreiras



barreiras interiores: 1º Quartil - 1.5 x IQR  
3º Quartil + 1.5 x IQR.

barreiras exteriores: 1º Quartil - 3 x IQR  
3º Quartil + 3 x IQR.

# Análise Exploratória Bivariada



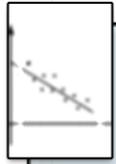
- ❑ **Duas variáveis categóricas (duas nominais, duas ordinais ou uma nominal e a outra ordinal)**
  - Tabela de Contingência (é utilizada para estudar a relação entre duas variáveis qualitativas; descreve a frequência das categorias de uma das variáveis relativamente às categorias da outra)
  - Medidas de Associação
- ❑ **Uma variável contínua em função de uma variável categórica**
  - Gráfico da variável contínua para cada classe da variável categórica (sendo esse gráfico, por exemplo um Boxplot, Histograma, etc)
  - Estatísticas da variável contínua para cada classe da variável categórica
- ❑ **Duas variáveis contínuas**
  - Diagrama de Dispersão (é utilizado para estudar a relação entre duas variáveis contínuas; descreve a posição de cada caso num referencial onde as coordenadas são definidas pelos valores de cada uma das variáveis)
  - Coeficiente de Correlação de Pearson e Coeficiente de Determinação

# Coeficiente de Correlação de Pearson

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[ \sum_{i=1}^n (y_i - \bar{y})^2 \right]}}$$



$r = -1$   
• Correlação linear negativa perfeita



$-1 < r \leq -0,8$   
• Correlação linear negativa forte



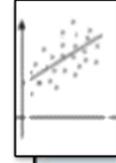
$-0,8 < r \leq -0,5$   
• Correlação linear negativa moderada



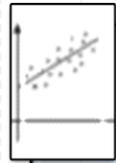
$-0,5 < r < 0$   
• Correlação linear negativa fraca



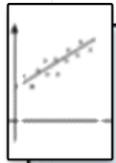
$r = 0$   
• Não existe correlação linear



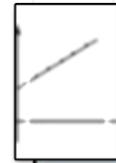
$0 < r < 0,5$   
• Correlação linear positiva fraca



$0,5 \leq r < 0,8$   
• Correlação linear positiva moderada



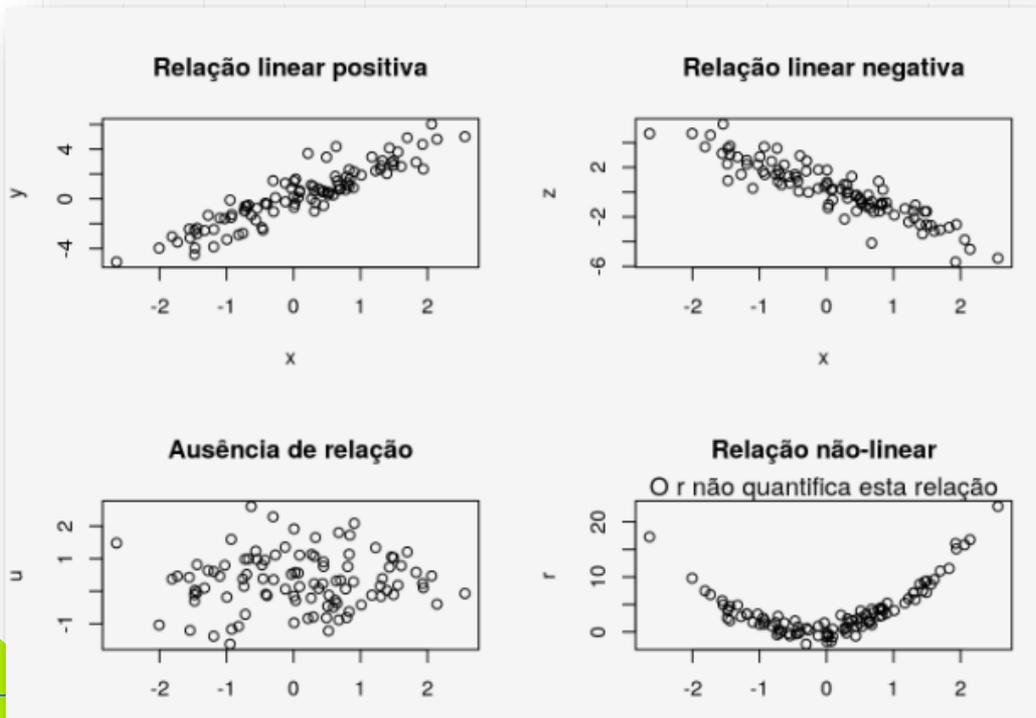
$0,8 \leq r < 1$   
• Correlação linear positiva forte

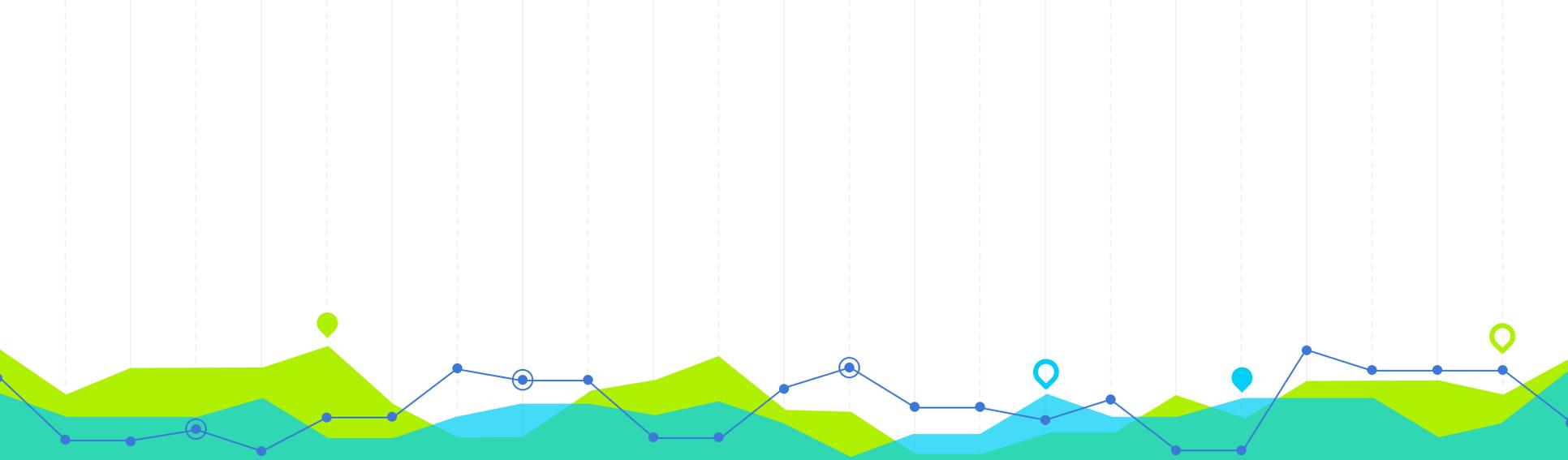


$r = 1$   
• Correlação linear positiva perfeita

# Associação/Correlação Linear?

<https://lec.pro.br//2-nao-categorizado/150-cor-pearson>





# Análise Exploratória de Dados no Excel

# 3

# Fazer uma Análise Descritiva no Excel

**Passo 1:** Organizar os dados

**Passo 2:** Calcular Medidas Amostrais

**Passo 3:** Construir Gráficos

**Passo 4:** Interpretar os Resultados

# Análise Descritiva Univariada

A	B	C	D	E
Id	Idade	Peso (Kg)	Altura (m)	Sexo (1=Mas, 2=Fem)
Aluno_1	20	79	1,8	1
Aluno_2	18	77	1,74	1
Aluno_3	23	80	1,82	1
Aluno_4	18	73	1,7	1
Aluno_5	19	69	1,72	2
Aluno_6	21	63	1,6	2
Aluno_7	20	60	1,62	2
Aluno_8	18	65	1,55	2
Aluno_9	23	55	1,58	2
Aluno_10	20	59	1,63	2

<b>MÉDIA</b> (valor1; valor2;...)	Média aritmética
<b>MED</b> (valor1; valor2;...)	Mediana
<b>MODO</b> (valor1; valor2;...)	Moda
<b>DESVPAD</b> (...; ...; ...)	Desvio-padrão
<b>VAR</b> (...; ...; ...)	Variância
<b>QUARTIL</b> (...; k)	Quartil k
<b>PERCENTIL</b> (...; k/100)	Percentil k

Medidas	Idade	Peso	Altura	Sexo
Média	20	1,6	68	-
Mediana	20	67	1,665	-
Moda	20	#N/D	#N/D	2
Mínimo	18	55	1,55	-
Máximo	23	80	1,82	-
Variância	3,5556	80	0,00876	-
Desvio-Padrão	32	720	0,07884	-
Amplitude Total	5	25	0,27	-
Quartil 1	18	59,75	1,595	-
Quartil 3	21,5	77,5	1,755	-
Percentil 50	20	67	1,665	-
Percentil 10	18	55,4	1,553	-
Percentil 90	23	79,9	1,818	-
Skewness	0	-	-	-
Kurtosis	0,35	0,362245	0,301887	-

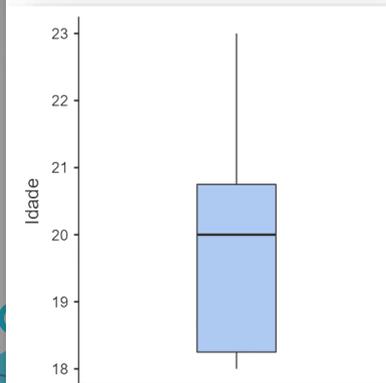
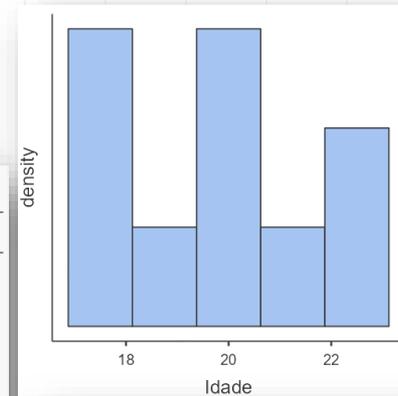
# Análise Descritiva Univariada

Id	Idade	Peso	Altura	Sexo
Aluno_1	20	79	1.80	1
Aluno_2	18	77	1.74	1
Aluno_3	23	80		
Aluno_4	18	73		
Aluno_5	19	69		
Aluno_6	21	63		
Aluno_7	20	60		
Aluno_8	18	65		
Aluno_9	23	55		
Aluno_10	20	59		

## Estatística Descritiva

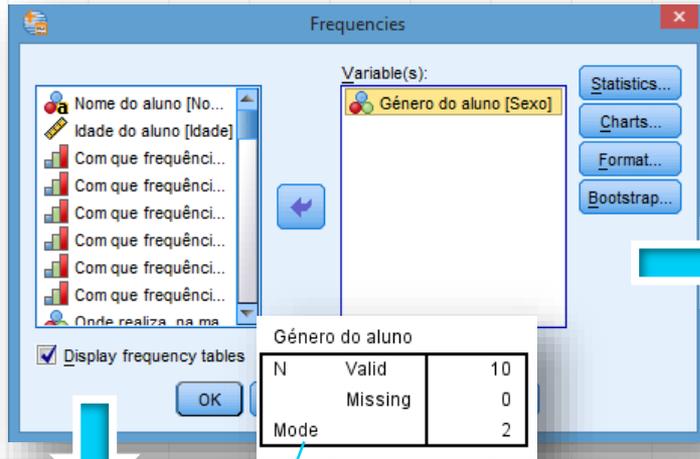
	Idade	Peso	Altura
N	10	10	10
Omisso	0	0	0
Média	20.0	68.0	1.68
Mediana	20.0	67.0	1.67
Moda	18.0 <sup>a</sup>	55.0 <sup>a</sup>	1.55 <sup>a</sup>
Desvio-padrão	1.89	8.94	0.0936
Variância	3.56	80.0	0.00876
Amplitude	5.00	25.0	0.270
Mínimo	18.0	55.0	1.55
Máximo	23.0	80.0	1.82
Assimetria	0.621	0.0629	0.281
Erro-padrão da Assimetria	0.687	0.687	0.687
Curtose	-0.679	-1.53	-1.29
Erro-padrão da Curtose	1.33	1.33	1.33

<sup>a</sup> Existe mais de uma moda, apenas a primeira é apresentada

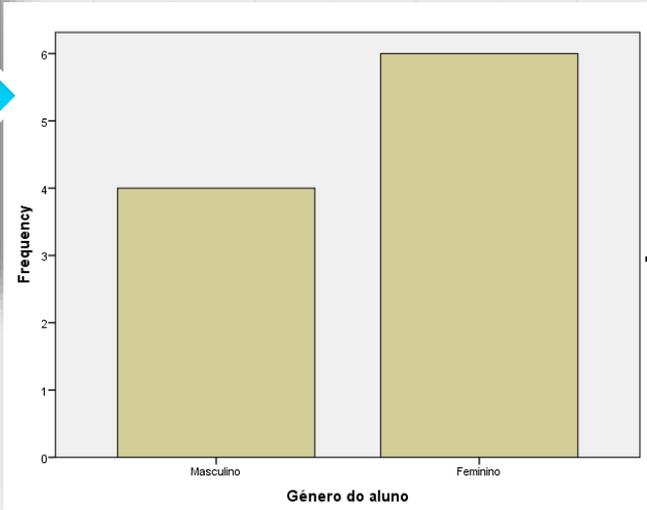


# Análise da Variável Sexo

SPSS:  Analyze  Descriptive Statistics  Frequencies  
[ Statistics /  Mode /  Charts /  Bar charts /  Frequencies ...]



Função Frequencies:  
Estatísticas e Gráficos



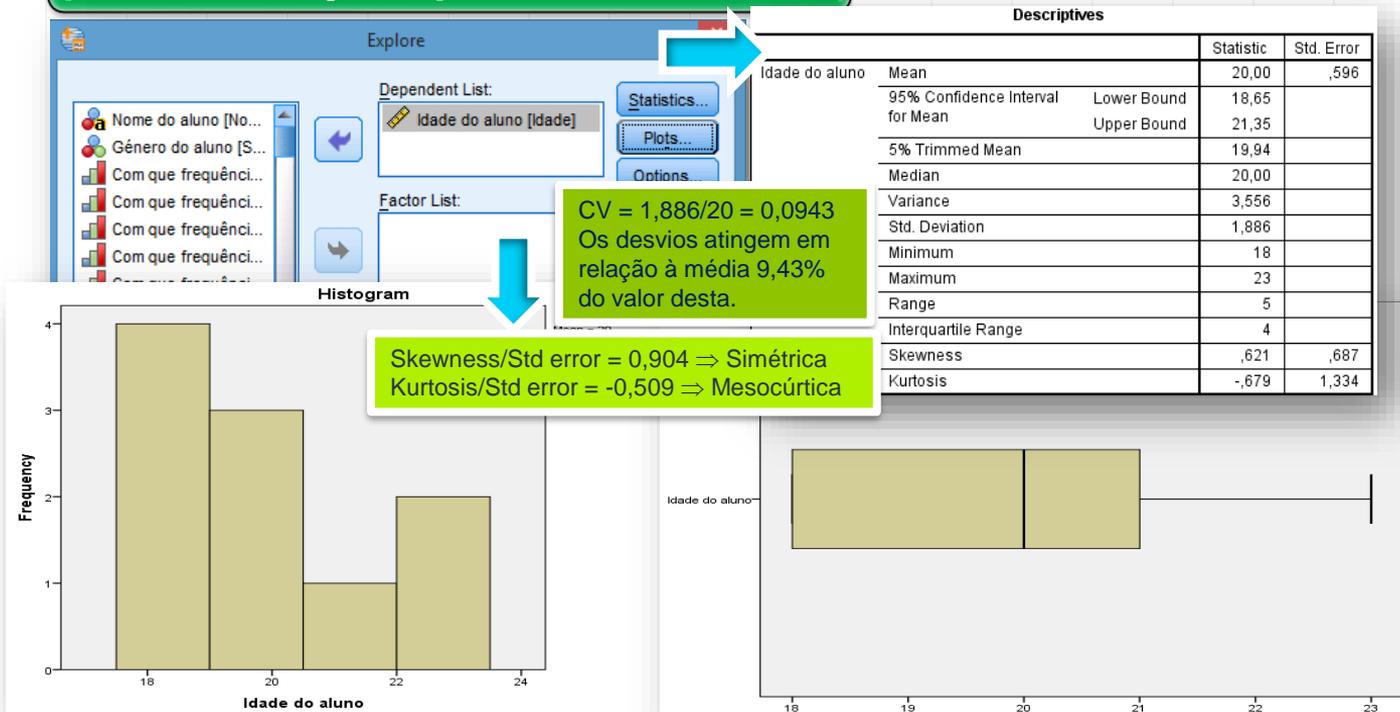
**Gênero do aluno**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Masculino	4	40,0	40,0	40,0
Feminino	6	60,0	60,0	100,0
Total	10	100,0	100,0	

# Análise da Variável Idade

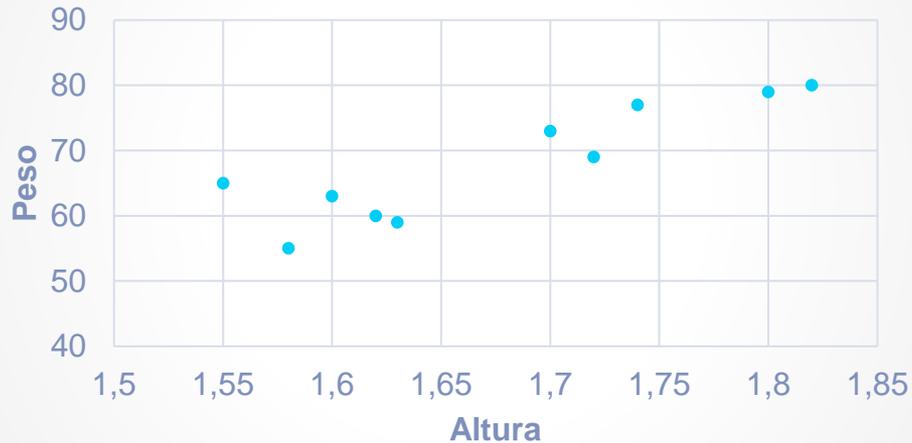
SPSS: Analyze Descriptive Statistics Explore  
[ Plots / Histogram ...]

Função Explore:  
Estatísticas e Gráficos



# Análise Descritiva Bivariada

## Diagrama de Dispersão

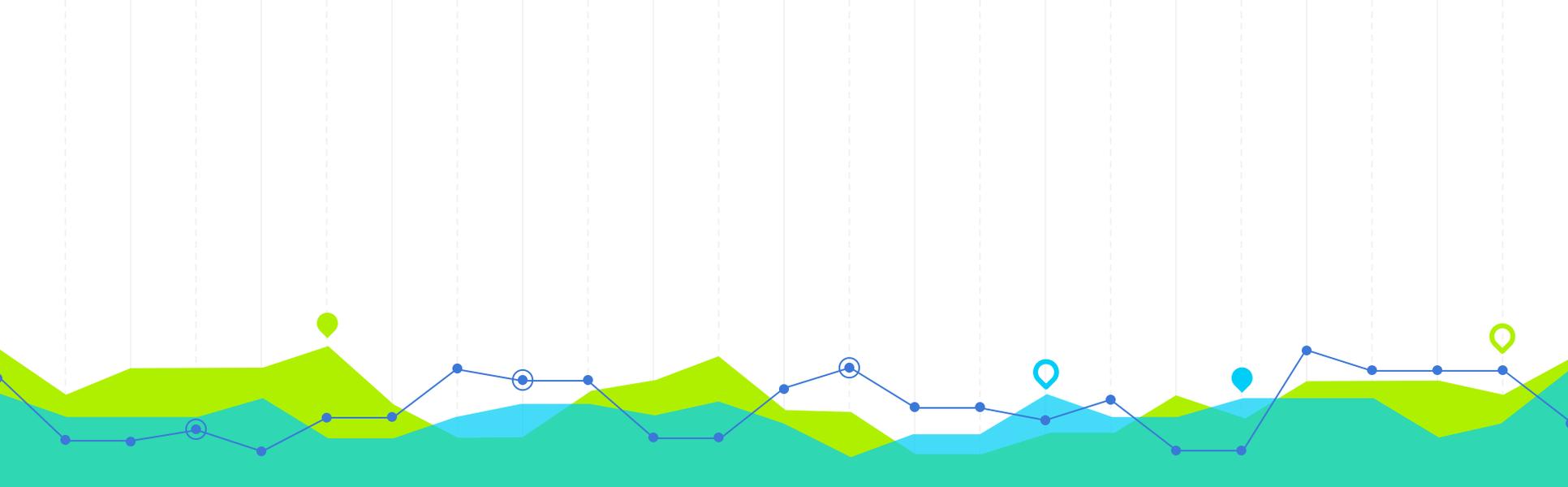


A	B	C	D	E
Id	Idade	Peso (Kg)	Altura (m)	Sexo (1=Mas, 2=Fem)
Aluno_1	20	79	1,8	1
Aluno_2	18	77	1,74	1
Aluno_3	23	80	1,82	1
Aluno_4	18	73	1,7	1
Aluno_5	19	69	1,72	2
Aluno_6	21	63	1,6	2
Aluno_7	20	60	1,62	2
Aluno_8	18	65	1,55	2
Aluno_9	23	55	1,58	2
Aluno_10	20	59	1,63	2

**Coefficiente de Correlação de Pearson**

$r = 0,889273$

Associação Linear positiva Forte

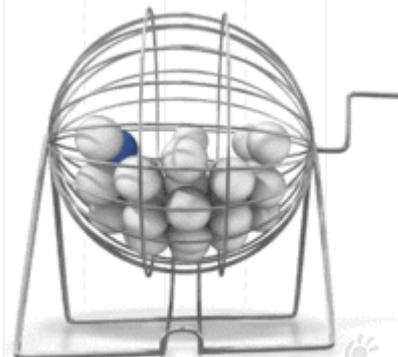
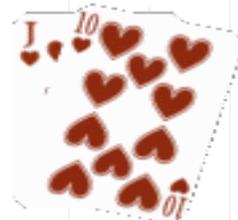
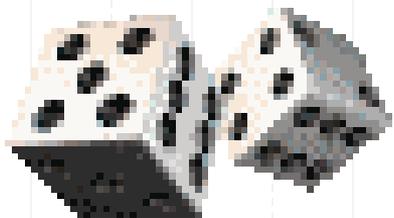


# Experiência Aleatória, Espaço de Resultados e Acontecimentos

# 4

# Experiência Aleatória

É toda a experiência que, mesmo repetida várias vezes, sob condições semelhantes, apresenta resultados imprevisíveis, dentro dos resultados possíveis.



# Espaço de Resultados / Espaço Amostral / Universo

- É o conjunto de todos os resultados possíveis da experiência.
- Representa-se por  $S$  ou  $\Omega$ .

# Espaço de Resultados

Qual é o espaço de resultados do lançamento de uma moeda?

$$\Omega = \{\text{Cara, Coroa}\}$$

Qual é o espaço de resultados do lançamento de um dado?

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$



# Acontecimento ou Evento

É todo subconjunto de um espaço de resultados  $\Omega$  de uma experiência aleatória.

# Acontecimento ou Evento

No lançamento de um dado, deseja-se que “saíam” valores menores que 4. Qual o evento desta experiência?

$$A = \{1, 2, 3\}$$



# Tipos de Acontecimentos

- Certo;
- Impossível;
- Muito provável;
- Pouco provável;
- Elementar;
- Composto.

# Tipos de Acontecimentos

Obter um número menor que 7 no lançamento de um dado representa que tipo de evento?

E um resultado igual a zero?

E um resultado maior que 5?



# Resumindo....

## Experiência Aleatória, Espaço de Resultados e Acontecimentos

**Experiência Aleatória ou casual** caracteriza-se pela impossibilidade de prever o resultado que se obterá, apesar de todos os seus possíveis resultados serem conhecidos à partida.

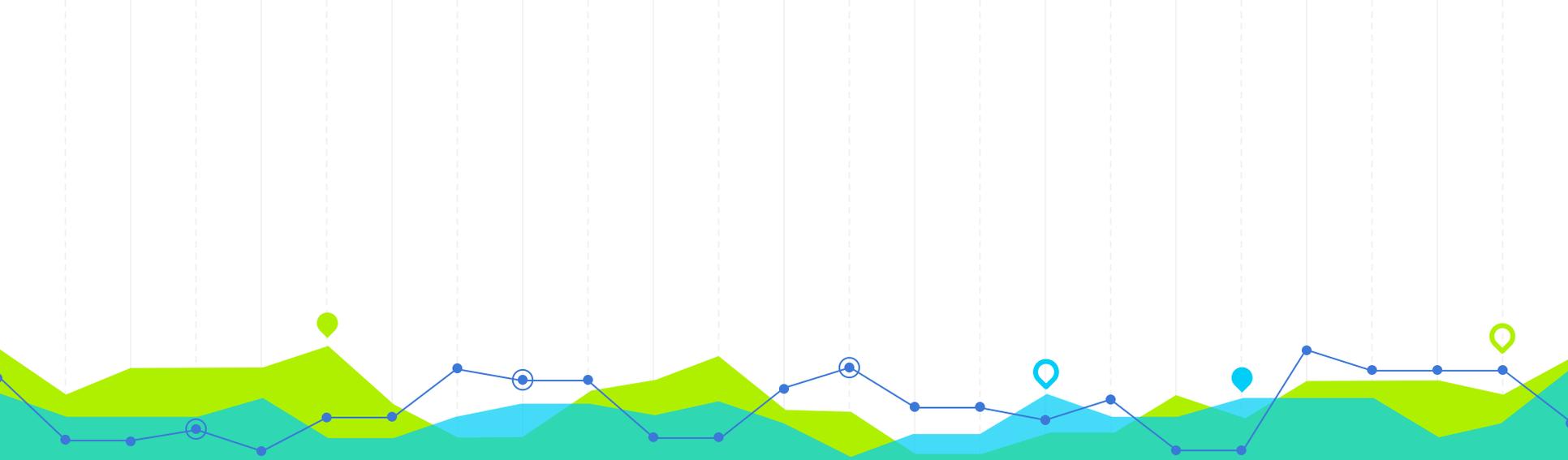
**Universo, espaço de resultados ou espaço amostral** de uma experiência aleatória é o conjunto de todos os seus resultados possíveis, denotado por  $\Omega$ .

**Acontecimento** de uma experiência aleatória  $E$  com universo  $\Omega$  é um subconjunto de  $\Omega$ .

### Exemplo:

- Experiência  $E$ : Lançamento de um dado
- $\Omega = \{1, 2, 3, 4, 5, 6\}$ ,
- Acontecimentos elementares  $A = \{1\}$ ,  $B = \{2\}$ , ...
- Acontecimentos compostos  $A = \{1, 2\}$ ,  $B = \{2, 4, 6\}$ , ...
- Acontecimento certo  $A = \Omega$       Acontecimento impossível =  $\{\}$

Espaço de acontecimentos  
=  $\{\{\}, \Omega, \{1\}, \{2\}, \{3\}, \{4\},$   
 $\{5\}, \{6\}, \{1,2\}, \{1,3\}, \{1,4\},$   
 $\{1,5\}, \{1,6\}, \{2,3\}, \{2,4\},$   
 $\{2,5\}, \{2,6\}, \{3,4\}, \{3,5\},$   
 $\{3,6\}, \{4,5\}, \{4,6\}, \{5,6\},$   
 $\{1,2,3\}, \dots \}$



# Experiência Aleatória, Espaço de Resultados e Acontecimentos: Exercícios

# 5

2. Duas lâmpadas vão ser submetidas a um teste que consiste em mantê-las ligadas até que ambas falhem, registrando-se a duração de cada uma delas. Sabe-se que nenhuma das lâmpadas dura mais de 1600 horas.

Represente o espaço de resultados e os seguintes acontecimentos:

*A* – «Nenhuma das lâmpadas tem duração superior a 1000 horas.»

*B* – «Só uma das lâmpadas tem duração superior a 1000 horas.»

*C* – «Uma das lâmpadas dura pelo menos o dobro da outra.»

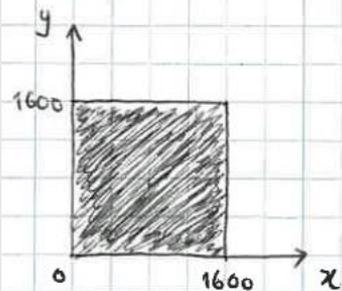
*D* – «A soma da duração das duas lâmpadas é superior a 2000 horas.»



## Exercício 2

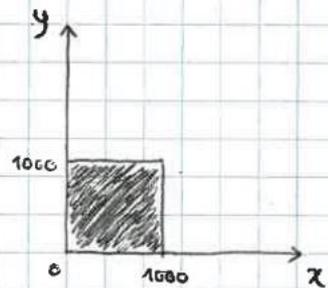
Experiência aleatória - observação da duração das duas lâmpadas.

$$\Omega = \{(x, y) : 0 \leq x \leq 1600 \wedge 0 \leq y \leq 1600\}$$

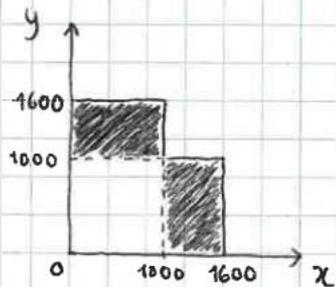


## Exercício 2

$$A = \{(x,y) : 0 \leq x \leq 1000 \wedge 0 \leq y \leq 1000\}$$

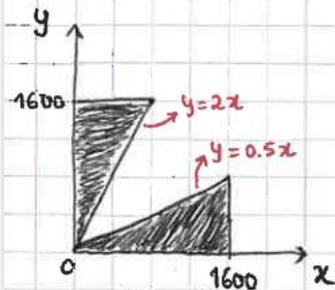


$$B = \{(x,y) : (1000 < x \leq 1600 \wedge 0 \leq y \leq 1000) \vee (0 \leq x \leq 1000 \wedge 1000 < y \leq 1600)\}$$

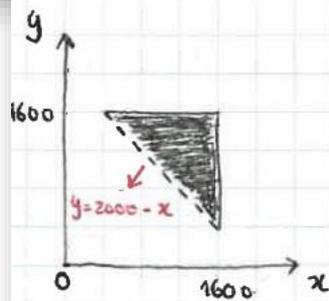


## Exercício 2

$$C = \{(x, y) : (x \geq 2y \vee y \geq 2x) \wedge (0 \leq x \leq 1600 \wedge 0 \leq y \leq 1600)\}$$



$$D = \{(x, y) : x + y > 2000 \wedge 0 \leq x \leq 1600 \wedge 0 \leq y \leq 1600\}$$



# Obrigada!

Questões?

