

# Estatística Aplicada à Gestão

## Licenciatura em Gestão de Empresas

Tiago Dias Domingues

Escola Superior de Gestão e Tecnologia de Santarém



## ■ **ESTATÍSTICA DESCRITIVA**

Organização e apresentação de dados; Medidas de localização de tendência central e não central; Medidas de dispersão; O coeficiente de variação

## ■ **PROBABILIDADES**

Experiência aleatória, espaço de resultados e acontecimentos; Princípio fundamental da contagem; Álgebra dos acontecimentos; Axiomática da probabilidade; Probabilidade condicional e independência; Teorema da probabilidade total; Teorema de Bayes

## ■ VARIÁVEIS ALEATÓRIAS

Variáveis aleatórias discretas: a função massa de probabilidade e a função de distribuição. Principais modelos discretos: Bernoulli, Binomial, Hipergeométrica e Poisson. Momentos de uma variável aleatória discreta. Variáveis aleatórias contínuas: a função densidade de probabilidade e a função de distribuição. Principais modelos contínuos: a distribuição uniforme contínua, distribuição exponencial e a distribuição Normal. O teorema da aditividade da normal. O teorema limite central e aplicações

## ■ INFERÊNCIA ESTATÍSTICA

Estimação pontual e intervalar. Testes de hipóteses para uma e duas populações

## Avaliação contínua

- Realização de dois testes ao longo do semestre (Cotado de 0 a 20 valores cada)

**Data do 1.º teste: 29 novembro 2022**

**Data do 2.º teste: 24 janeiro 2023**

## Avaliação contínua

### Condições de aprovação:

- $T1 \geq 8$  valores e  $T2 \geq 8$  valores (nota mínima de 8 valores em cada momento)
- $(T1 + T2)/2 \geq 9,5$  valores

### Material de consulta:

- Calculadora gráfica
- Formulário a disponibilizar pelo docente na página da disciplina

## Avaliação por exame final

- Caso o aluno não obtenha aprovação na avaliação contínua dispõe da 1.<sup>a</sup> e 2.<sup>a</sup> épocas de exame;
- As melhorias de nota são realizadas na 2.<sup>a</sup> época de exame;
- Data dos exames: mediante consulta do calendário de exames a disponibilizar na página da ESGTS

### Condições de aprovação:

- $E1 \geq 9,5$  valores ou  $E2 \geq 9,5$  valores

### Material de consulta:

- Calculadora científica/ gráfica
- Formulário a disponibilizar pelo docente na página da disciplina

# Bibliografia

- Murteira, B., Ribeiro, CS., Silva, JA., Pimenta, C., Pimenta, F. (2014). Introdução à Estatística, 3ªEd. Escolar Editora, Lisboa.
- Murteira, B. e Antunes, M. (2012). Probabilidade e Estatística, volumes I e II. Escolar Editora, Lisboa.
- Pestana, D. D. e Velosa, S. F. (2010). Introdução à Probabilidade e à Estatística, vol. I, 4ª ed., Fundação Calouste Gulbenkian, Lisboa.
- Dalgaard, P. (2008). Introductory Statistics with R, Springer.
- Douguerty, E. (1990). Probability and Statistics For The Engineering, Computing and Physical Sciences, Prentice-Hall.
- Montgomery, D. and Runger, G. (2010). Applied Statistics and Probability for Engineering, Wiley.

# CAPÍTULO 1

## Estatística Descritiva e Análise Exploratória de Dados

# O que é a Estatística?

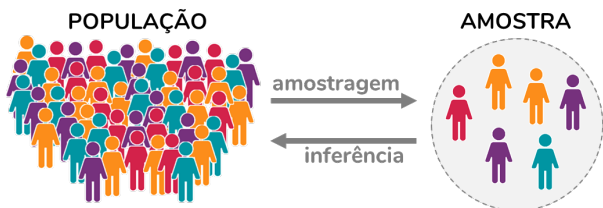
- é uma palavra derivada do latim «status» (estado)
- é a ciência que se ocupa da recolha, análise e interpretação de determinado tipo de informação, com o objetivo de descrever, modelar e a partir daí inferir e prever
- é possível desta forma distinguir o ramo da **estatística descritiva** (que se ocupa da descrição dos dados) e a **inferência estatística** (onde se procede à inferência e predição)

# População e amostra

- Uma **população** é um conjunto vasto de observações ou unidades experimentais com características próprias (a que chamamos **características populacionais**) que se pretendem estudar. Formalmente, a população é o conjunto de todas as observações sobre as quais recai o nosso estudo/interesse e será denotada ao longo desta disciplina por letras maiúsculas do alfabeto (p.ex.  $X$ )
  
- **Exemplo:** Supondo que se pretende estudar a altura (em cm) dos alunos da ESGT, tem-se que a **população alvo são todos os alunos inscritos na ESGT**

# População e amostra

- De modo a conhecerem-se as características da população procede-se à recolha de uma **amostra representativa** dessa população, de modo a inferir-se das características amostrais para as características populacionais. Ao longo desta disciplina irão utilizar-se letras minúsculas do alfabeto para denotar uma amostra



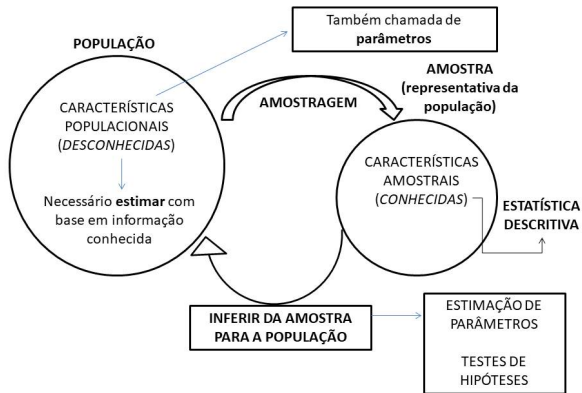
<sup>1</sup>Retirado de <https://fernandafperes.com.br/intervalo-de-confianca>

# População e amostra

- O objetivo passa por conhecer características da população, também chamados de **parâmetros** e que são **estimados** através de características (conhecidas) da amostra, chamadas de **estatísticas**. Os parâmetros da população são quantidades desconhecidas mas fixas na população em estudo
- A amostra recolhida da População é obtida com base em **técnicas de amostragem** de maneira a obter-se **representatividade**
- O conhecimento das características amostrais é obtido através da aplicação das técnicas de descrição de dados, ou seja, da **estatística descritiva**

# População e amostra

Em resumo:



# População e amostra - notação

Tal como já foi referido anteriormente, iremos reservar as letras maiúsculas para a população e as minúsculas para a amostra. De uma forma geral uma característica da população será designada por  $X$ .

A amostra observada, recolhida da População em estudo de modo a inferir sobre a característica  $X$ , será denotada por  $(x_1, x_2, \dots, x_n)$ , onde  $n$  é a dimensão da amostra, ou seja, o número de observações. Podemos também denotar a amostra observada como sendo  $(x_i), i = 1, \dots, n$ . Uma amostra observada é realização de uma amostra aleatória  $(X_1, X_2, \dots, X_n)$ .

# Obtenção dos dados - o processo de amostragem

A constituição da amostra envolve várias etapas, sendo a primeira o chamado **plano de amostragem**:

- definição da população alvo/interesse
- definição do processo utilizado para a seleção dos elementos da população a incluir na amostra
- estabelecimento da dimensão da amostra

As operações de amostragem são, em geral, caras optando-se muitas vezes por obter amostras auto-seleccionadas ou amostras por "conveniência" em que o investigador selecciona os elementos da população a incluir de forma propositada.

# Obtenção dos dados - o processo de amostragem

Na situação ideal estamos interessados em que as amostras sejam obtidas por processos de amostragem aleatória simples sem reposição, em que qualquer das possíveis subamostras de dimensão  $n$  que se podem retirar de uma população têm igual probabilidade de ser seleccionada. Contudo, este processo é bastante dispendioso.

O processo de amostragem que acabámos de descrever é então o chamado processo de **amostragem aleatória (probabilística)**. Já no caso de se considerarem amostras por conveniência, o processo que dá origem a estas amostras é chamado de **amostragem não aleatória (não probabilística)**.

# Obtenção dos dados - o processo de amostragem

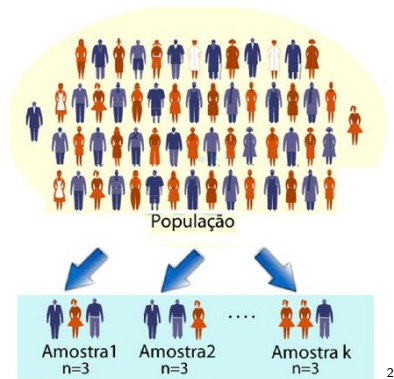
## Características do processo de **amostragem probabilística**:

- cada um dos elementos da população tem uma probabilidade (conhecida *a priori*) de ser seleccionado para a amostra;
- baseados total ou parcialmente na intervenção do acaso;
- tendo em conta a representatividade da amostra, é possível avaliar os correspondentes parâmetros populacionais com a precisão desejada
- conduzem às chamadas amostras aleatórias

# Obtenção dos dados - o processo de amostragem

Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



<sup>2</sup>Retirado de <https://slideplayer.com.br/slide/14293759/>

## Características do processo de **amostragem não probabilística**:

- os elementos da população a incluir na amostra são escolhidos pelo investigador de forma propositada;
- são utilizados em alternativa ao processo de amostragem probabilístico devido a razões de custo.

Seleccionada a amostra que vai servir para estudar uma característica da População, estamos então em condições de descrevê-la através de técnicas de Estatística Descritiva.

# Tipo de dados e variáveis

Vimos que o principal objetivo da Estatística é o de dada uma População, recolher uma amostra, organizá-la e resumí-la através de características amostrais apropriadas e extrapolar a informação da amostra para a População em estudo. A amostra observada é constituída por dados que podem ser classificados em dois tipos: **qualitativos** e **quantitativos**.

- **Tipos de dados e a escala de Steven**

- **qualitativos:** os dados do tipo qualitativo não são passíveis de mensuração. Nestes casos, apenas são atribuídas categorias.

Os dados do tipo qualitativo podem ainda ser subdivididos em **nominais** e **ordinais**.

# Tipo de dados e variáveis

Um exemplo de dados qualitativos **nominais** é o sexo de um indivíduo, cujas categorias são *Masculino* e *Feminino*.

Já para os dados qualitativos **ordinais** são exemplos: o número de estrelas com que se classifica um hotel ou a qualidade de um seguro de saúde (*Fraco, Bom ou Muito Bom*).

**Pergunta: O que diferencia os dados do tipo qualitativo nominais dos ordinais?** *Os dados do tipo qualitativo ordinal permitem uma ordenação das categorias ao contrário dos nominais.*

# Tipo de dados e variáveis

- **quantitativos:** os dados do tipo quantitativo são mensuráveis, ou seja, podemos contar ou medir através deles.

Os dados do tipo quantitativo também permitem uma subdivisão. Estes podem ser de **escala intervalar/razão** ou ainda **percentual/absoluto**.

**Pergunta: O que diferencia os dados do tipo quantitativo intervalares/razão dos percentuais/absolutos?** *Os primeiros não possuem um zero absoluto, ao contrário dos da segunda subclassificação.*

# Tipo de dados e variáveis

**Exemplo:** a temperatura (graus centígrados) é uma **variável** do tipo quantitativo, mas que não possui um zero absoluto, uma vez que  $0^{\circ}\text{C}$  não representa ausência de temperatura, remetendo-nos assim para as variáveis de tipo quantitativo intervalar.

A classificação dos dados em quantitativos e qualitativos e respetivas subclassificações deve-se a Steven (1946), denominando-se por **escala de Steven**.

Os dados resultam da observação de **variáveis**.

# Tipo de dados e variáveis

As variáveis podem ainda ser classificadas em dois tipos:

- **discretas:** a variável assume valores num conjunto finito ou infinito numerável. **Exemplos:** número de filhos, número de lançamentos de um dado, número de lançamentos de uma moeda, etc.
- **contínuas:** a variável toma valores num conjunto infinito não numerável. **Exemplos:** peso (kg), altura (cm), etc.

**Observação:** os dados qualitativos dão sempre origem a variáveis discretas e os dados quantitativos podem dar origem a variáveis discretas ou contínuas.

# Organizar os dados e resumir a informação - dados do tipo qualitativo

A organização de dados do tipo qualitativo passa pela construção de uma tabela, chamada de **tabela de frequências**.

Esta tabela é constituída por duas colunas que dizem respeito à **frequência absoluta** e **relativa** de cada uma das categorias. A **frequência absoluta** de uma categoria  $C_i$  denota-se por  $n_i$  e corresponde ao número de elementos da amostra que pertence a essa categoria.

A **frequência relativa** de uma categoria  $C_i$  denota-se por  $f_i$  e corresponde à proporção de elementos da amostra que pertence a essa categoria.

# Organizar os dados e resumir a informação - dados do tipo qualitativo

Ou seja,

$$f_i = \frac{n_i}{n},$$

em que  $n$  é a dimensão da amostra.

A soma de todas as frequências absolutas é igual a  $n$  e a soma das frequências relativas é a igual a 1 (ou 100%).

# Organizar os dados e resumir a informação - dados do tipo qualitativo

**Exemplo:**<sup>3</sup> Os dados que se seguem, dizem respeito à situação profissional por nível de escolaridade de indivíduos dos Estados Unidos no ano de 2003.

	Ensino Superior	Ensino secundário	Outras habilitações	Sem escolaridade
A trabalhar	8224	5654	11 167	2583
Desempregado	217	231	693	303
Reformado	956	1354	3107	2549
Total ( <i>n</i> )	9397	7239	14967	5435

A informação apresentada nesta tabela diz respeito às frequências absolutas.

<sup>3</sup>Fonte: Barrow, M. Statistics for Economics, Accounting and Business Studies (2017)

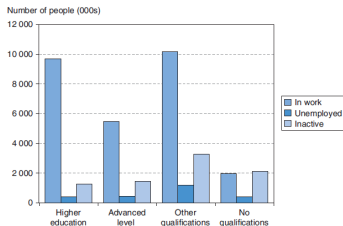
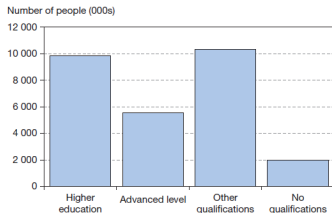
# Organizar os dados e resumir a informação - dados do tipo qualitativo

Caso pretendêssemos as frequências relativas, bastaria então calcular o quociente entre a frequência absoluta e o número de indivíduos observados em cada categoria. Ou seja,

	Ensino Superior	Ensino secundário	Outras habilitações	Sem escolaridade
A trabalhar	$8224/9397 = 88\%$	78%	74%	48%
Desempregado	$217/9397 = 2\%$	3%	5%	6%
Reformado	$956/9397 = 10\%$	19%	21%	47%

# Representações gráficas - dados do tipo qualitativo

## • Gráfico de barras (I)

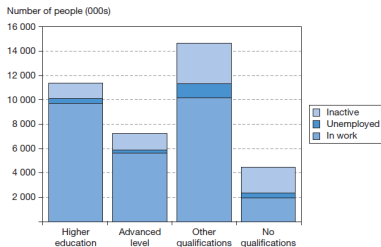


4

<sup>4</sup>Fonte: Barrow, M. Statistics for Economics, Accounting and Business Studies (2017)

# Representações gráficas - dados do tipo qualitativo

## • Gráfico de barras (II)

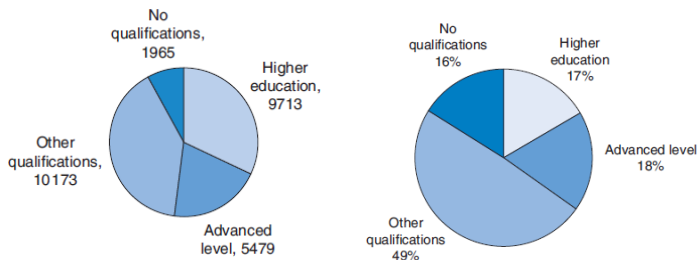


5

<sup>5</sup>Fonte: Barrow, M. Statistics for Economics, Accounting and Business Studies (2017)

# Representações gráficas - dados do tipo qualitativo

## • Diagrama circular



6

<sup>6</sup>Fonte: Barrow, M. Statistics for Economics, Accounting and Business Studies (2017)

# Organizar os dados e resumir a informação - dados do tipo quantitativo

O modo como conseguimos resumir a informação de dados quantitativos é através do cálculo de medidas, que se podem dividir em **medidas de localização** e **medidas de dispersão**. As primeiras permitem perceber a ordem de grandeza dos dados, ou seja, onde se localizam na reta real; as segundas indicam como os dados se distribuem na reta real.

## ★ Medidas de localização

Começamos por abordar a medida clássica e conhecida de todos, a **média**.

# Organizar os dados e resumir a informação - dados do tipo quantitativo

- A média representa-se por  $\bar{x}$  e consiste no quociente entre a soma de todas as observações pela dimensão da amostra. Ou seja,

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

Trata-se portanto de uma medida de tendência central.

- Uma outra medida de localização é a **moda** ( $m$ ), que corresponde ao elemento mais frequente na amostra.
- A **mediana** ( $M$ ), é também uma medida de localização muito utilizada em que para o seu cálculo é necessário proceder à **ordenação da amostra**.

# Organizar os dados e resumir a informação - dados do tipo quantitativo

A mediana é então interpretada como sendo o valor tal que 50% das observações são menores ou iguais que a mediana e 50% das observações são maiores ou iguais que a mediana.

**Notação:** denota-se **amostra ordenada** por  $(x_{(1)}, \dots, x_{(n)})$ , onde  $x_{(i)}$  é a  $i$ -ésima estatística ordinal, verificando-se

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

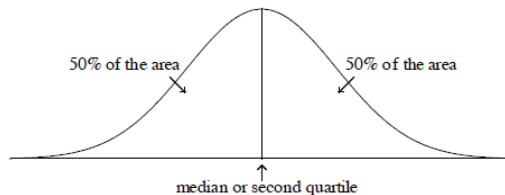
Tem-se então que a mediana é dada por

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

# Organizar os dados e resumir a informação - dados do tipo quantitativo

Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



7

---

<sup>7</sup>Fonte: Nicholas, J. Introduction to Descriptive Statistics (2006) 

# Organizar os dados e resumir a informação - dados do tipo quantitativo

- **Quantil empírico de ordem  $p$  ( $Q_p$ ).** Esta medida indica-nos que  $100p\%$  dos valores do conjunto de observações são menores ou iguais a  $Q_p$  e  $(1 - p)100\%$  são maiores ou iguais a  $Q_p$ . O quantil empírico pode ou não ser um elemento do conjunto de observações, tal como se observou com a mediana. Formalmente,

$$Q_p = \begin{cases} x_{([np]+1)} & ,\text{se } np \text{ não é inteiro} \\ \frac{x_{(np)} + x_{(np+1)}}{2} & ,\text{se } np \text{ inteiro,} \end{cases}$$

em que  $[np]$  representa a parte inteira de  $np$ .

# Organizar os dados e resumir a informação - dados do tipo quantitativo

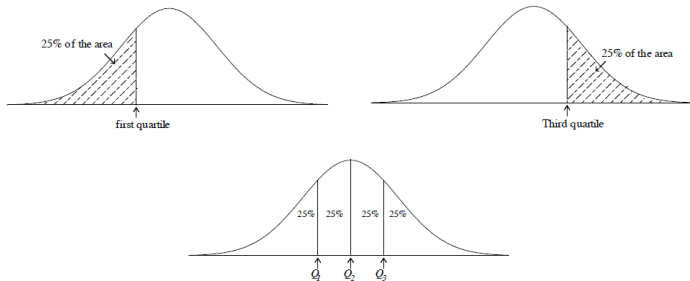
Um exemplo clássico de quantis são o 1.º quartil e o 3.º quartil. Sabendo que para o primeiro caso se tem  $p = \frac{1}{4}$  e para o segundo,  $p = \frac{3}{4}$  e de acordo com a definição exposta atrás, tem-se o seguinte:

- ▷ No caso do 1.º quartil ( $Q_{1/4}$  ou  $Q_{0.25}$ ), **25%** das observações são menores ou iguais ao valor do 1.º quartil;
- ▷ No caso do 3.º quartil ( $Q_{3/4}$  ou  $Q_{0.75}$ ), **75%** das observações são menores ou iguais ao valor do 3.º quartil.

# Organizar os dados e resumir a informação - dados do tipo quantitativo

Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



8

<sup>8</sup>Fonte: Nicholas, J. Introduction to Descriptive Statistics (2006)

# Organizar os dados e resumir a informação - dados do tipo quantitativo

## ★ Medidas de dispersão

As medidas de dispersão dizem respeito ao comportamento dos dados, no que se refere à sua variabilidade.

Introduzimos o conceito de **variância** que não é mais do que a média das distâncias das observações relativamente à média.

Ou seja,

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

Desenvolvendo a soma de quadrados, resulta ainda que

# Organizar os dados e resumir a informação

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

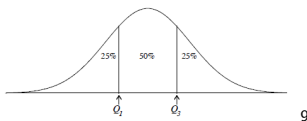
Como resultado da variância, surge o **desvio padrão**. Esta medida calcula-se fazendo a raíz quadrada da variância. Ou seja,

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

# Organizar os dados e resumir a informação - dados do tipo quantitativo

- **Amplitude interquartil:** corresponde à diferença entre o 3.º quartil e o 1.º quartil. Denota-se por  $AIQ$ .

$$AIQ = Q_{3/4} - Q_{1/4}.$$



- **Amplitude amostral:** corresponde à diferença entre o máximo e o mínimo da amostra.

$$R = x_{(n)} - x_{(1)}.$$

<sup>9</sup>Fonte: Nicholas, J. Introduction to Descriptive Statistics (2006)   

# Organizar os dados e resumir a informação - dados do tipo quantitativo

- **Coeficiente de variação:** trata-se de uma medida de dispersão relativa que indica a ordem de grandeza do desvio padrão quando comparado com a média. Trata-se de uma medida muito utilizada quando se pretende comparar a variabilidade de amostras distintas.

$$cv = \frac{s}{\bar{x}} \times 100\%.$$

# Representações gráficas - dados do tipo quantitativo

## • Histograma

Suponhamos que foi retirado da produção de uma fábrica um conjunto de 30 lâmpadas de 60Watts. Foi realizado um teste à durabilidade (em horas) das lâmpadas e os resultados obtidos foram os seguintes:

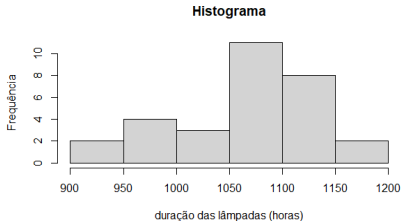
963.4	1175.9	1001.7	1198.2	1078.3	1065.4	1124.8	1083.8
1092.7	1143.8	1087.3	1114.1	1089.5	1133.5	1072.8	1003.4
922.0	1121.8	1142.0	950.1	1021.7	1052.1	987.2	1108.4
1099.3	1075.6	901.3	988.8	1074.1	1109.8		

O histograma é uma representação gráfica utilizada para representar dados do tipo quantitativo, para os quais se realizou algum tipo de agrupamento.

# Representações gráficas - dados do tipo quantitativo

- **Histograma**

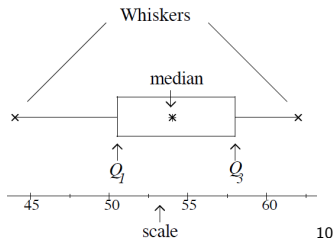
Trata-se de um gráfico de áreas, em que a soma da área de cada um dos rectângulos que formam o histograma é igual a 1 ou 100%.



# Representações gráficas - dados do tipo quantitativo

## • Diagrama em caixa-de-bigodes

Além do histograma, uma outra representação gráfica possível para dados quantitativos é o diagrama em caixa-de-bigodes. Trata-se de uma representação bastante útil para comparar amostras, bem como obter informação sobre a simetria (ou falta de simetria) dos dados. A sua construção envolve as seguintes medidas

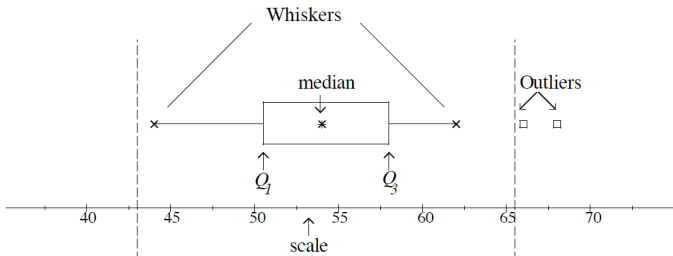


<sup>10</sup>Fonte: Nicholas, J. Introduction to Descriptive Statistics (2006)   

# Representações gráficas - dados do tipo quantitativo

## • Diagrama em caixa-de-bigodes

Através do diagrama em caixa-de-bigodes é possível observarmos observações extremas, que se afastam das restantes observações da amostra. A essas observações extremas damos o nome de **outliers**.



11

<sup>11</sup>Fonte: Nicholas, J. Introduction to Descriptive Statistics (2006)   

# Representações gráficas - dados do tipo quantitativo

- **Diagrama em caixa-de-bigodes**

Como saber se uma observação é um outlier?

- ★ **A barreira de outliers**

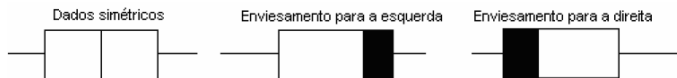
$$B.I = Q_{1/4} - 1.5 \times AIQ$$

$$B.S = Q_{3/4} + 1.5 \times AIQ$$

Todas as observações que estiverem fora do intervalo  $[B.I, B.S]$  são consideradas outliers.

# Representações gráficas - dados do tipo quantitativo

- **Análise da simetria dos dados através do diagrama em caixa-de-bigodes**



12

<sup>12</sup>Fonte: Martins, EG. Introdução à Probabilidade e à Estatística (2005)  

# Representações gráficas - dados do tipo quantitativo

Relativamente à análise da simetria dos dados, podemos então concluir que

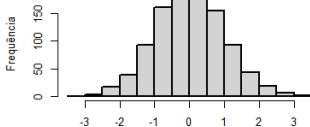
- **Distribuição simétrica:**  $\text{média} = \text{mediana} = \text{moda}$
- **Distribuição assimétrica positiva ou enviesada à direita:**  $\text{moda} < \text{mediana} < \text{média}$
- **Distribuição assimétrica negativa ou enviesada à esquerda:**  $\text{média} < \text{mediana} < \text{moda}$

# Representações gráficas - dados do tipo quantitativo

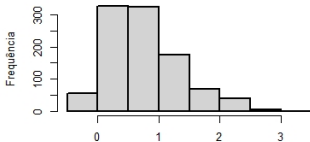
Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues

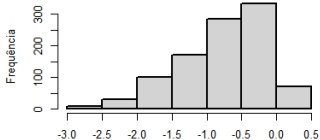
**Distribuição simétrica**



**Distribuição assimétrica positiva**



**Distribuição assimétrica negativa**



# Representações gráficas - dados do tipo quantitativo

## Algumas observações:

- a média é uma medida pouco resistente à presença de *outliers*. Nesse sentido, deve usar-se a mediana como medida de localização (mais resistente)
- o desvio padrão é uma medida pouco resistente à presença de *outliers*. Nesse sentido, deve usar-se a amplitude inter-quartil como medida de dispersão (mais resistente)

# Representações gráficas - dados do tipo quantitativo

## Exemplo:

**Amostra 1:** 5 6 6 7 7 8 10

■ Média=7; Mediana=7

**Amostra 2:** 5 6 6 7 7 8 50

■ Média=8.43; Mediana=7

# Exemplo de aplicação

Os dados seguintes dizem respeito ao tempo (em minutos) que 12 funcionários demoraram a chegar ao seu local de trabalho num determinado dia <sup>13</sup>.

18	34	68	22	10	92	46	52	38	29	45	37
----	----	----	----	----	----	----	----	----	----	----	----

Repare-se que a amostra não se encontra ordenada e utilizando a notação estabelecida anteriormente para uma amostra observada, tem-se que

$x_1 = 18, x_2 = 34, x_3 = 68, x_4 = 22, x_5 = 10, x_6 = 92, x_7 = 46, x_8 = 52, x_9 = 38, x_{10} = 29, x_{11} = 45, x_{12} = 37$ , em que  $n = 12$ .

<sup>13</sup>Fonte: Nicholas, J. Introduction to Descriptive Statistics (2006) 

# Exemplo de aplicação

Procedendo à ordenação da amostra,

$$x_{(1)} = 10, x_{(2)} = 18, x_{(3)} = 22, x_{(4)} = 29, x_{(5)} = 34, x_{(6)} = 37, x_{(7)} = 38, x_{(8)} = 45, x_{(9)} = 46, x_{(10)} = 52, x_{(11)} = 68, x_{(12)} = 92.$$

Estando a amostra ordenada, é possível identificar o mínimo e o máximo da amostra:  $x_{(1)} = 10$  e  $x_{(12)} = 92$ , respetivamente. Vamos agora proceder ao cálculo das medidas de localização.

- **Média:** Recorde-se que a média ( $\bar{x}$ ) é definida como sendo o quociente da soma de todas as observações pela dimensão da amostra. ou seja,

$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{12} = \frac{491}{12} = 40.917\text{min.}$$

# Exemplo de aplicação

No que diz respeito à moda, repare-se que não há nenhum elemento mais frequente, pelo que todos os elementos da amostra são modas.

- **Mediana:** para o cálculo da mediana, o primeiro passo a verificar é se a dimensão da amostra é par ou ímpar. Neste caso, como  $n = 12$  (é par), a mediana vem dada por

$$M = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}.$$

Assim,

$$M = \frac{x_{(\frac{12}{2})} + x_{(\frac{12}{2}+1)}}{2} = \frac{x_{(6)} + x_{(7)}}{2} = \frac{37 + 38}{2} = 37.5 \text{ min.}$$

# Exemplo de aplicação

- **Quartis:** vamos proceder ao cálculo do 1.º e 3.º quartis, respetivamente. Tem-se que

**1.º quartil ( $Q_{0.25}$ ):**  $np = 12 \times 0.25 = 3$  (é inteiro), logo,

$$Q_{0.25} = \frac{x_{(np)} + x_{(np+1)}}{2} = \frac{x_{(3)} + x_{(4)}}{2} = \frac{22 + 29}{2} = 25.5\text{min.}$$

**3.º quartil ( $Q_{0.75}$ ):**  $np = 12 \times 0.75 = 9$  (é inteiro), logo,

$$Q_{0.75} = \frac{x_{(np)} + x_{(np+1)}}{2} = \frac{x_{(9)} + x_{(10)}}{2} = \frac{46 + 52}{2} = 49\text{min.}$$

# Exemplo de aplicação

Calculadas as medidas de localização, vamos agora proceder ao cálculo das medidas de dispersão.

- **Variância:** sabemos que a variância corresponde à média do quadrado das distâncias de cada uma das observações relativamente à média. Em particular, vimos que poderia ser útil considerar

$$s^2 = \frac{\sum_{i=1}^{12} x_i^2}{n-1} - \frac{n}{n-1} \bar{x}^2.$$

Vamos então proceder ao cálculo de  $\sum_{i=1}^{12} x_i^2$ .

# Exemplo de aplicação

$x_i$	$x_i^2$
10	100
18	324
22	484
29	841
34	1156
37	1369
38	1444
45	2025
46	2116
52	2704
68	4624
92	8464
$\sum_{i=1}^{12} x_i = 491$	$\sum_{i=1}^{12} x_i^2 = 25651$

Sabendo que a média (já calculada anteriormente) é de 40.917, então  $(\bar{x})^2 = (40.917)^2 = 1674.174$ , tem-se que

$$s^2 = \frac{25651}{11} - \left(\frac{12}{11} \cdot 1674.174\right) = 505.5379 \text{ min.}^2$$

# Exemplo de aplicação

Assim,

$$s = \sqrt{s^2} = \sqrt{505.5379} = 22.48417 \text{ min.}$$

Temos ainda que  $x_{(1)} = 10$  e  $x_{(12)} = 92$ , pelo que a amplitude amostral é  $R = 92 - 10 = 82$ . Tendo em conta o primeiro e terceiro quartis (calculados anteriormente), a amplitude interquartil vem dada por

$$AIQ = Q_{3/4} - Q_{1/4} = 49 - 25.5 = 23.5.$$

Relativamente ao coeficiente de variação, tem-se que

$$cv = \frac{s}{\bar{x}} \times 100\% = \frac{22.48417}{40.917} \times 100 = 54.95\%.$$

# Exemplo de aplicação

Vamos representar os dados com recurso a um diagrama em caixa-de-bigodes. Para tal, necessitamos das seguintes medidas:  $Q_{1/4}$ ,  $M$ ,  $Q_{3/4}$ ,  $AIQ$ ,  $min.$  e  $max.$ . Vimos anteriormente que  $Q_{1/4} = 25.5$ ,  $M = 37.5$ ,  $Q_{3/4} = 49$ ,  $min. = 10$  e  $max. = 92$ . Assim,

$$B.I = 25.5 - (1.5 \times 23.5) = -9.75$$

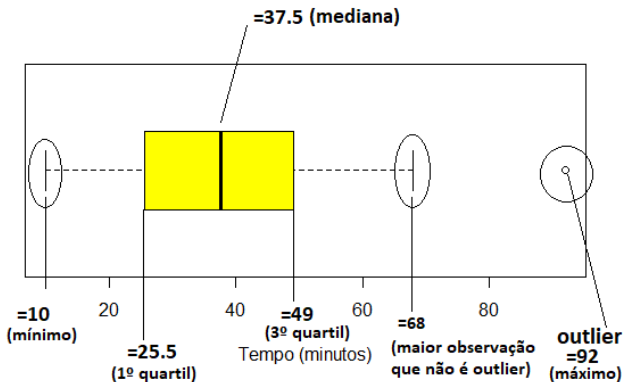
$$B.S = 49 + (1.5 \times 23.5) = 84.25$$

De acordo com a barreira de outliers, verificamos que  $92 \notin [-9.75, 84.25]$ , pelo que é considerado um outlier. Não se verificando a existência de outliers à esquerda, temos que o bigode à esquerda vai até ao mínimo da amostra, enquanto que o bigode à direita vai até à observação da amostra imediatamente anterior que não é considerada um outlier, ou seja, até ao valor 68.

# Exemplo de aplicação

Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



# CAPÍTULO 2

## Probabilidade

# Introdução

- **Experiência aleatória** é uma experiência que pode ser repetida nas mesmas condições, em que não se sabe qual o resultado que se irá observar, contudo conhece-se o conjunto de resultados possíveis (ou espaço de resultados), ou seja, o seu universo. Designaremos o conjunto de resultados possíveis por  $\Omega$ .
- **Acontecimento**: objeto(s) ao(s) qual(is) atribuímos probabilidade. Podemos distinguir dois tipos de acontecimentos: **acontecimentos simples ou elementares** e os **acontecimentos compostos**. O exemplo seguinte permite diferenciar os dois tipos de acontecimentos.

# Introdução

**Exemplo:** Consideremos a experiência aleatória que consiste no lançamento de um dado. O conjunto de resultados possíveis é  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ; sair a face 2 ( $\{2\}$ ) é um acontecimento simples ou elementar e sair face par ( $\{2, 4, 6\}$ ) é um acontecimento composto. Se, em vez de lançarmos um dado e observarmos a face voltada para cima, forem lançados dois dados, então o espaço de resultados é dado por  $\Omega =$   
 $\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), \dots, (6, 5), (6, 6)\}$ ,  
ou seja,

$$\Omega = \{(i, j) : i = 1, \dots, 6; j = 1, \dots, 6\}.$$



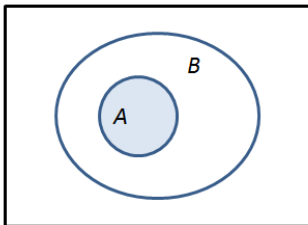
# Introdução

**Nota:** ao conjunto vazio,  $\emptyset$ , chamamos de acontecimento impossível e a  $\Omega$  acontecimento certo. Adicionalmente, o cardinal de um conjunto corresponde ao número de elementos desse conjunto. No exemplo anterior,  $\#\Omega = 36$ , em que  $\#$  representa a cardinalidade do conjunto.

Ainda relativamente ao espaço de resultados,  $\Omega$ , diz-se que é um espaço **discreto** se toma valores num conjunto finito ou infinito numerável, ou seja, podemos contar cada um dos seus elementos; caso contrário diz-se **contínuo** se contém um intervalo de números reais. Deixa-se como observação que qualquer elemento de  $\Omega$  é um subconjunto do mesmo.

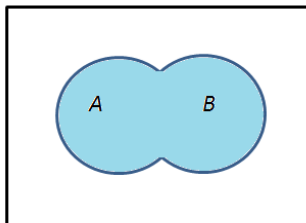
## Algumas notas sobre teoria de conjuntos:

- **Subconjunto:** dados dois acontecimentos  $A$  e  $B$ , diz-se que  $A$  é um subconjunto de  $B$ , e escreve-se  $A \subset B$  ( $\subseteq$ ), se e só se a realização de  $A$  implica a realização de  $B$ .



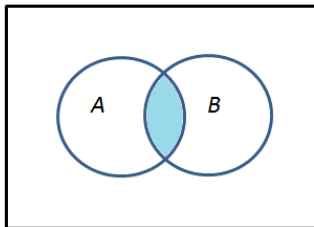
# Introdução

- dados dois acontecimentos  $A$  e  $B$ , a união dos dois acontecimentos denota-se por  $A \cup B$ .



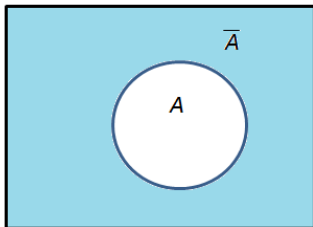
# Introdução

- dados dois acontecimentos  $A$  e  $B$ , a interseção dos dois acontecimentos denota-se por  $A \cap B$ .



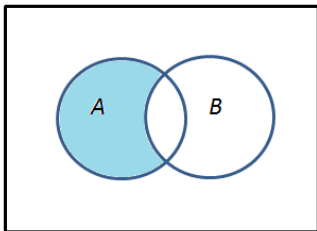
# Introdução

- dado um acontecimento  $A$ , para denotar o complementar de  $A$ , ou seja, dizer que  $A$  não ocorre, utiliza-se  $\bar{A}$ .



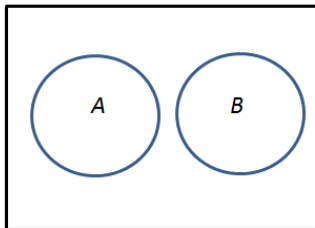
# Introdução

- dados dois acontecimentos  $A$  e  $B$ , a diferença dos acontecimentos  $A$  e  $B$  representa-se por  $A - B = (A \cap \bar{B})$ .



# Introdução

- dados dois acontecimentos  $A$  e  $B$ , diz-se que  $A$  e  $B$  são mutuamente exclusivos ou disjuntos, se  $A \cap B = \emptyset$ .



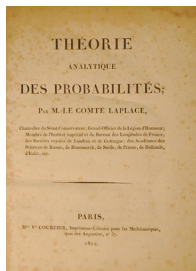
# Introdução

Ainda algumas propriedades importantes:

1. **Distributiva:**  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$  e  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ ;
2. **Leis de De Morgan:**  $\overline{A \cap B} = \overline{A} \cup \overline{B}$  e  $\overline{A \cup B} = \overline{A} \cap \overline{B}$ .

# Introdução

- Pierre-Simon Laplace foi um dos primeiros matemáticos a estabelecer uma teoria rigorosa da probabilidade, tendo publicado o célebre "Théorie Analytique des Probabilités (1812)" onde surge a definição clássica de probabilidade.



- **Definição Clássica ou de Laplace:** a probabilidade de um acontecimento  $A$  é dada pelo quociente entre o número de casos favoráveis e o número de casos possíveis, desde que estes sejam **equiprováveis**, ou seja,

$$P(A) = \frac{\textit{número de casos favoráveis a } A}{\textit{número de casos possíveis}}.$$

**Exemplo:** Considerando um dado equilibrado e o acontecimento  $A$  : "Sair face ímpar" , tem-se  $P(A) = \frac{3}{6} = \frac{1}{2}$ .

- **Definição frequencista de probabilidade:** a probabilidade de um acontecimento  $A$  é dada pelo limite da frequência relativa com que se observou  $A$ , isto é,

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n},$$

em que  $n_A$  representa o número de observações de  $A$ , e  $n$  o número de realizações da experiência aleatória. Para valores elevados de  $n$ , podemos assumir que  $P(A) \approx \frac{n_A}{n}$ .

# Introdução

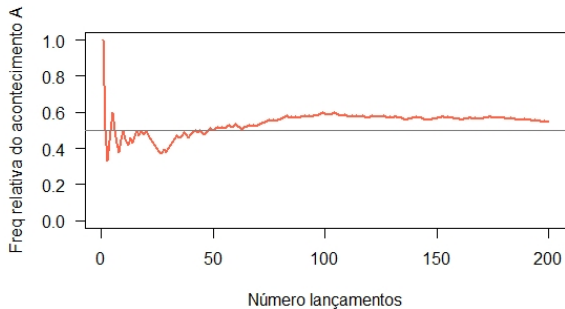
**Exemplo:** Lançou-se 200 vezes uma moeda regular e registou-se ao fim de cada lançamento a frequência relativa do acontecimento A: "sair coroa". Na tabela seguinte encontram-se os primeiros 100 lançamentos.

$n$	$f_n$	$n$	$f_n$	$n$	$f_n$	$n$	$f_n$	$n$	$f_n$	$n$	$f_n$
1	1.00	19	0.421	37	0.486	55	0.473	73	0.479	91	0.495
2	0.500	20	0.450	38	0.474	56	0.482	74	0.473	92	0.489
3	0.667	21	0.476	39	0.462	57	0.491	75	0.480	93	0.495
4	0.750	22	0.455	40	0.450	58	0.500	76	0.487	94	0.500
5	0.600	23	0.435	41	0.439	59	0.508	77	0.494	95	0.505
6	0.667	24	0.417	42	0.452	60	0.500	78	0.487	96	0.500
7	0.571	25	0.400	43	0.465	61	0.492	79	0.494	97	0.495
8	0.625	26	0.385	44	0.455	62	0.484	80	0.488	98	0.490
9	0.556	27	0.407	45	0.467	63	0.476	81	0.481	99	0.495
10	0.500	28	0.429	46	0.457	64	0.469	82	0.488	100	0.500
11	0.455	29	0.448	47	0.447	65	0.462	83	0.494		
12	0.500	30	0.433	48	0.458	66	0.470	84	0.500		
13	0.538	31	0.452	49	0.469	67	0.478	85	0.494		
14	0.500	32	0.438	50	0.480	68	0.485	86	0.488		
15	0.467	33	0.424	51	0.490	69	0.478	87	0.483		
16	0.438	34	0.441	52	0.481	70	0.471	88	0.489		
17	0.412	35	0.457	53	0.491	71	0.465	89	0.483		
18	0.389	36	0.472	54	0.481	72	0.472	90	0.489		

# Introdução

Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



- **Lei dos grandes números:** quando uma experiência aleatória pode repetir-se um grande número de vezes em condições semelhantes, definindo um acontecimento  $A$  no respetivo espaço de resultados e calculada a respetiva frequência relativa da sua realização em  $n$  provas, verifica-se (empiricamente) que quando  $n$  aumenta, existe uma tendência para a **estabilização da frequência relativa** em torno de um número que corresponde de forma aproximada a  $P(A)$ . Assim, qualquer que seja  $\varepsilon > 0$ , existe um número de provas,  $n_0$ , a partir do qual

$$P(A) - \varepsilon < f_n(A) < P(A) + \varepsilon,$$

ou seja,  $\lim_{n \rightarrow +\infty} P(|f_n(A) - P(A)| < \varepsilon) = 1$ .

# Axiomática de Kolmogorov

Não existe uma definição do que é a probabilidade, existem é conjunto de "regras" ou **axiomas** que permitem descrever o seu comportamento. Assim, probabilidade é uma medida que verifica os seguintes axiomas:

1.  $P(A) \geq 0$ , qualquer que seja o acontecimento  $A$ ;
2.  $P(\Omega) = 1$ ;
3. Se  $A$  e  $B$  são acontecimentos disjuntos,  
 $P(A \cup B) = P(A) + P(B)$ ;

Este último axioma pode ser ainda generalizado da seguinte forma:

- 3\* Se  $A_1, A_2, \dots$  são acontecimentos disjuntos dois a dois, então

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

# Axiomática de Kolmogorov

Como consequências imediatas da axiomática de Kolmogorov, tem-se:

1.  $P(\emptyset) = 0$ ;
2. Se  $A \subseteq B$ , então  $P(A) \leq P(B)$ ;
3.  $P(\overline{A}) = 1 - P(A)$ ;
4.  $P(A) \in [0, 1]$ ;
5.  $P(A - B) = P(A \cap \overline{B}) = P(A) - P(A \cap B)$ ;
6.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

# Axiomática de Kolmogorov

- **Exemplo:** Em determinada população, 9.8% das pessoas adquirem a revista  $A$ , 22.9% a revista  $B$  e 5.1% ambas as revistas. Podem então definir-se os acontecimentos  $A$ : "adquirir a revista  $A$ " e  $B$ : "adquirir a revista  $B$ ".
- (a) A probabilidade de adquirir somente a revista  $A$ , acontecimento  $A \cap \overline{B} = A - B$ , é dada por

$$P(A \cap \overline{B}) = P(A - B) = P(A) - P(A \cap B) = 0.098 - 0.051 = 0.047.$$

# Axiomática de Kolmogorov

- (b) A probabilidade de uma pessoa escolhida ao acaso adquirir pelo menos uma das revistas, acontecimento  $A \cup B$ , é dada por

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.098 + 0.229 - 0.051 = 0.276.$$

- (c) A probabilidade de não adquirir nem a revista  $A$ , nem a revista  $B$ , acontecimento  $\overline{A \cap B}$ , é dada por

$$P(\overline{A \cap B}) = 1 - P(A \cap B) = 1 - 0.051 = 0.949.$$

# Análise combinatória - algumas revisões

- **Arranjos (sem repetição):** considere-se um conjunto com  $n$  elementos distintos e suponha-se que a partir deles se formam grupos com um número fixo de  $k$  elementos ( $k = 1, 2, \dots, n$ ), não repetidos, **mas onde os grupos diferem pela ordem em que os elementos são seleccionados**. Os grupos que assim se obtêm chamam-se de arranjos de  $n$  elementos em grupos de  $k$  elementos. Representa-se por  $A_k^n$ .

**Exemplo:** Dispondo de quatro peças de pano de cores diferentes (por exemplo: azul, vermelho, branco e amarelo), quantas bandeiras tricolores de faixas de pano verticais podem elaborar-se sem repetir as cores?

Neste caso, uma vez que a ordem das cores produz bandeiras diferentes, pretende-se:  $A_3^4 = 4 \times 3 \times 2 = 24$ .

# Análise combinatória - algumas revisões

- **Arranjos (com repetição):** quando nos arranjos **entram elementos repetidos do conjunto**, tem-se que o número de arranjos de  $n$  elementos em grupos de  $k$  elementos é dado por  $\alpha_k^n = n^k$ .
- **Permutações:** as permutações de um conjunto de  $n$  elementos distintos são arranjos de  $n$  elementos em grupos de  $n$  elementos, isto é, as permutações são grupos formados por todos os elementos do conjunto. O número de permutações de  $n$  elementos é então dado por

$$P_n = A_n^n = n(n-1)(n-2)\dots 2 \times 1 = n!$$

# Análise combinatória - algumas revisões

- **Combinações:** considere-se um conjunto com  $n$  elementos distintos e suponha-se que a partir deles se formam grupos com um número fixo de  $k$  elementos ( $k = 1, 2, \dots, n$ ), não repetidos, **mas onde os grupos não diferem pela ordem em que os elementos são seleccionados**. Os grupos que assim se obtêm chamam-se de combinações de  $n$  elementos em grupos de  $k$  elementos. Representa-se por  $C_k^n$  ou  $\binom{n}{k}$ .

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

# Probabilidade condicionada e independência de acontecimentos

Para introduzir o conceito de probabilidade condicional, consideremos o exemplo de Pestana e Velosa (2010) que consiste no seguinte: se numa urna houver 5 bolas brancas e 5 bolas pretas, tirando uma bola ao acaso (e podemos assim considerar que os requisitos laplacianos de equiprobabilidade dos acontecimentos elementares estão a ser respeitados), a probabilidade de sair bola branca é

$$P(B) = \frac{5}{10}.$$

# Probabilidade condicionada e independência de acontecimentos

Suponhamos que a bola extraída é de facto branca, e que a removemos (não é repostada na urna); passou a haver na urna apenas 9 bolas, das quais 4 são brancas e 5 são pretas. Se tirarmos agora ao acaso uma das restantes, a probabilidade de extrair bola branca, sabendo que antes foi retirada uma bola branca, é

$$P(B_2|B_1) = \frac{4}{9}.$$

A notação indicada anteriormente lê-se como "a probabilidade de sair bola branca na segunda extração condicional a ter saído bola branca na primeira extração".

# Probabilidade condicionada e independência de acontecimentos

Por outro lado, podemos pensar na probabilidade de sair bola branca na primeira extração e bola branca na segunda extração. Ora, para a primeira extração existem 5 casos favoráveis, enquanto que na segunda já existem apenas 4. Relativamente aos casos possíveis, existem 10 na primeira extração e 9 na segunda. Então a probabilidade pedida é dada por

$$P(B_1 \cap B_2) = \frac{5 \times 4}{10 \times 9},$$

que não é mais do que o produto da probabilidade de  $B_1$  pela probabilidade condicional de  $B_2$ , sabendo que  $B_1$  ocorreu.

# Probabilidade condicionada e independência de acontecimentos

Desta forma, temos então a seguinte definição:

- **Probabilidade Condicionada:** dados dois acontecimentos  $A$  e  $B$ , a probabilidade condicional de  $A$  dado  $B$ , ou sabendo  $B$  é dada por

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

se  $P(B) > 0$ .

# Probabilidade condicionada e independência de acontecimentos

Da definição de probabilidade condicionada, obtém-se facilmente a regra do produto:

$$P(A \cap B) = P(A|B) \times P(B).$$

No caso em que a ocorrência do acontecimento  $B$  não tem qualquer influência sobre a ocorrência de  $A$ , tem-se  $P(A|B) = P(A)$ . Analogamente se coloca a questão quando se tem  $P(B|A)$ . Desta forma, utilizando novamente na regra do produto, tem-se

$$P(A \cap B) = P(A) \times P(B).$$

# Probabilidade condicionada e independência de acontecimentos

- **Exemplo:** Depois de inúmeras experiências com peças de vulcanite moldada obtiveram-se os resultados que se encontram na tabela seguinte, onde se estabelece a classificação das peças segundo dois critérios - porosidade e dimensão - cada um dos quais revestindo duas modalidades:  $A$ : "peças porosas";  $\bar{A}$ : "peças não porosas";  $B$ : "peças de dimensionamento defeituoso";  $\bar{B}$ : "peças de dimensionamento não defeituoso".

Dimensionamento	Porosas	Não porosas	Total
Defeituoso	2.1%	4.9%	7%
Não defeituoso	18.1%	74.9%	93%
Total	20.2%	79.8%	100%

# Probabilidade condicionada e independência de acontecimentos

Do exemplo anterior podemos obter:

- Probabilidade de obter uma peça porosa:  $P(A) = 0.202$
- Probabilidade de obter uma peça não porosa:  
 $P(\bar{A}) = 0.798$
- Probabilidade de obter uma peça de dimensionamento defeituoso:  $P(B) = 0.070$
- Probabilidade de obter uma peça de dimensionamento não defeituoso:  $P(\bar{B}) = 0.930$
- Probabilidade de obter uma peça porosa e dimensionamento defeituoso:  $P(A \cap B) = 0.021$
- Probabilidade de obter uma peça porosa e de dimensionamento não defeituoso:  $P(A \cap \bar{B}) = 0.181$

# Probabilidade condicionada e independência de acontecimentos

- Tendo-se verificado que uma peça tinha dimensionamento defeituoso, qual a probabilidade de que seja porosa?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.021}{0.070} = 0.3$$

- Se uma peça tem dimensionamento defeituoso, qual a probabilidade de ser não porosa?

$$P(\bar{A}|B) = \frac{\bar{A} \cap B}{P(B)} = \frac{0.049}{0.070} = 0.7$$

- Se a peça é porosa, qual a probabilidade de ter dimensionamento defeituoso?

$$P(B|A) = \frac{B \cap A}{P(A)} = \frac{0.021}{0.202} = 0.104$$

# Probabilidade condicionada e independência de acontecimentos

- **Acontecimentos independentes:** dois acontecimentos  $A$  e  $B$  dizem-se independentes se e só se,

$$P(A \cap B) = P(A) \times P(B).$$

Aliada à definição de probabilidade condicionada e independência, surge um dos teoremas mais fecundos em teoria de probabilidades, o **teorema da probabilidade total**. Para tal, começamos por relembrar o conceito de partição, fundamental para a compreensão do que vem a seguir

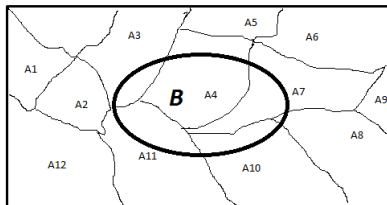
# Teorema da Probabilidade Total e Teorema de Bayes

■ Uma família  $\{A_n\}_{n \in \mathbb{N}}$  é uma **partição** do conjunto  $\Omega$  se e só se:

1.  $A_i \cap A_j = \emptyset, \forall i \neq j$ ;
2.  $\cup_{n \in \mathbb{N}} A_n = \Omega$ ;
3.  $P(A_i) > 0, \forall i$ .

# Teorema da Probabilidade Total e Teorema de Bayes

Consideremos uma partição de um acontecimento  $B$  induzida por uma partição de  $\Omega$ .



$\Omega$

# Teorema da Probabilidade Total e Teorema de Bayes

**Teorema da Probabilidade Total:** Seja  $B$  um acontecimento e  $\{A_n\}_{n \in \mathbb{N}}$  uma partição do universo  $\Omega$ . Então,

$$P(B) = \sum_{n \in \mathbb{N}} P(B|A_n)P(A_n).$$

Como consequência deste teorema, tem-se o seguinte:

*Quaisquer que sejam os acontecimentos  $A$  e  $B$ ,*

$$P(B) = P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A}).$$

Neste caso, foi considerada a partição  $\{A, \bar{A}\}$  de  $\Omega$ .

# Teorema da Probabilidade Total e Teorema de Bayes

- **Exemplo:** Considerem-se os acontecimentos A: "chove num determinado dia" e B: "dia nublado". Suponhamos que nunca chove num dia não nublado, que a probabilidade de chover num dia nublado é 0.35 e que a probabilidade de amanhã estar um dia nublado é 0.20. Qual a probabilidade de chover amanhã? Tem-se que

$$P(A \cap \bar{B}) = 0; P(A|B) = 0.35; P(B) = 0.20.$$

Pelo teorema da probabilidade total,

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A \cap B).$$

# Teorema da Probabilidade Total e Teorema de Bayes

Assim,

$$P(A) = P(A \cap B) = P(A|B)P(B) = 0.20 \times 0.35 = 0.07.$$

## Teorema de Bayes.

- Se  $\{A_n\}_{n \in \mathbb{N}}$  é uma partição de  $\Omega$  e  $P(A_n) > 0$ , então para qualquer acontecimento  $B$  de  $\Omega$  tal que  $P(B) > 0$ , verifica-se

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{n \in \mathbb{N}} P(B|A_n)P(A_n)}, \forall i.$$

# Teorema da Probabilidade Total e Teorema de Bayes

- Exemplo: Um falso baralho de cartas constituído por 52 reis é misturado num saco com três baralhos honestos de 52 cartas. Uma carta é escolhida ao acaso do saco. A probabilidade de a carta pertencer ao baralho falso é  $1/4$ . Se a carta extraída for um rei, qual a probabilidade de a carta pertencer ao baralho falso?

$$P(\text{falso}|\text{rei}) = \frac{P(\text{falso})P(\text{rei}|\text{falso})}{P(\text{falso})P(\text{rei}|\text{falso}) + P(\text{honesto})P(\text{rei}|\text{honesto})} =$$
$$= \frac{(1/4) \times 1}{(1/4) \times 1 + (3/4) \times (1/13)} = 0.81.$$

## CAPÍTULO 3

### Variáveis aleatórias e modelos discretos

- **Variável aleatória** é uma função que associa um valor real a cada possível resultado de uma experiência aleatória, isto é, é uma função definida no espaço de resultados  $\Omega$  e que toma valores em  $\mathbb{R}$ .

$$\begin{aligned} X &: \Omega \longrightarrow \mathbb{R} \\ &\omega \longrightarrow x(\omega) \end{aligned}$$

**Exemplo:** Considerando a experiência aleatória que consiste no lançamento de um dado e observar o número de pintas da face que fica volta para cima é uma variável aleatória.

# Introdução

Tem-se então que à variável aleatória  $X$ : "Número de pintas da face que fica voltada para cima no lançamento de um dado" é atribuído um valor do conjunto  $\{1, 2, 3, 4, 5, 6\}$ .

Se considerarmos agora a experiência aleatória que consiste no registo das alturas (cm) de 20 alunos e definirmos  $Y$ : "valor da altura de um aluno em cm", então é possível associar a  $Y$  um valor real correspondente à altura do aluno. Ou seja,  $Y$  é uma variável aleatória.

# Tipos de variáveis aleatórias

Existem dois tipos de variáveis aleatórias: as **discretas** e as **contínuas**. As variáveis aleatórias discretas tomam valores num conjunto finito ou infinito numerável. Caso contrário, a variável diz-se contínua, ou seja, toma valores num conjunto infinito (ex.  $\mathbb{R}$ ).

- **Variáveis aleatórias discretas**

Vimos que uma variável aleatória diz-se discreta se toma valores num conjunto finito ou infinito numerável. Todas as variáveis aleatórias discretas têm associada uma distribuição de probabilidade a que chamamos **função massa de probabilidade (f.m.p)**.

# Tipos de variáveis aleatórias

Trata-se de uma função que a cada valor que a variável toma esteja associada a respetiva probabilidade de se observar tal valor, ou seja,

$$\begin{array}{c|cccccc} x_i & x_1 & x_2 & x_3 & \dots & x_n \\ \hline p_i = P(X = x_i) & p_1 & p_2 & p_3 & \dots & p_n \end{array}$$

Contudo, para que a função acima seja um função massa de probabilidade é necessário que sejam verificadas as seguintes condições

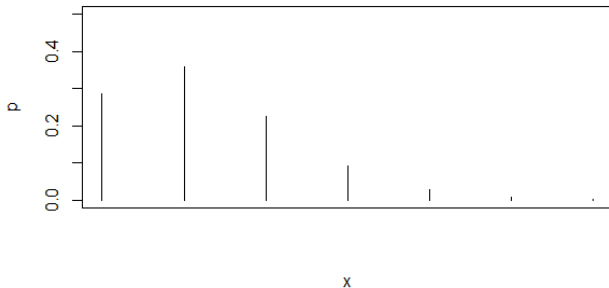
$$\sum_i p_i = P(X = x_i) = 1$$

e,

$$P(X = x_i) \geq 0, \forall i.$$

# Tipos de variáveis aleatórias

A representação gráfica da f.m.p. é um simples gráfico de barras



Para além do interesse em conhecer a probabilidade associada a cada um dos valores da variável aleatória, interessa muitas vezes conhecer o valor associado a acontecimentos que podem ser descritos como

- $X < a, X \leq a, X > a, X \geq a$  ou ainda
- $a < X < b, a \leq X < b, a < X \leq b, a \leq X \leq b,$

em que  $a$  e  $b$  são valores reais quaisquer ( $a \leq b$ ).

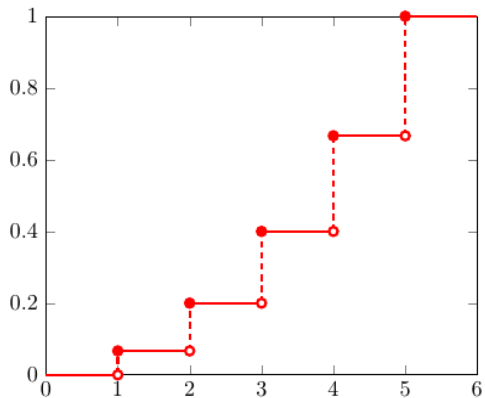
Para dar resposta a este tipo de questões, surge a necessidade de definir a chamada **função de distribuição (f.d.)**.

# Tipos de variáveis aleatórias

Define-se a função de distribuição de uma variável aleatória  $X$  como

$$F_X(x) = P(X \leq x), \forall x \in \mathbb{R}.$$

Trata-se portanto de uma função cumulativa, sendo a sua representação gráfica uma função em escada.



# Tipos de variáveis aleatórias

**Exemplo:** Seja  $X$  a variável aleatória que representa o número de acidentes observados numa determinada linha férrea. Considerando que se observam no máximo 6 acidentes, o suporte da variável é  $\{0, 1, 2, 3, 4, 5, 6\}$ . A distribuição de probabilidade é dada por:

$x_i$	0	1	2	3	4	5	6
$p_i = P(X = x_i)$	0.287	0.358	0.224	0.093	0.029	0.007	0.002

Podemos dizer que se trata de uma f.m.p.? De facto, constata-se que todas as probabilidades são maiores ou iguais que 0 e que a sua soma é a unidade, pelo que podemos concluir que se trata de uma f.m.p..

Considerando novamente o exemplo do número de acidentes, a função de distribuição é definida por

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.287 & 0 \leq x < 1 \\ 0.645 & 1 \leq x < 2 \\ 0.869 & 2 \leq x < 3 \\ 0.962 & 3 \leq x < 4 \\ 0.991 & 4 \leq x < 5 \\ 0.998 & 5 \leq x < 6 \\ 1 & x \geq 6 \end{cases}$$

Se pensarmos por exemplo, na probabilidade de se observarem mais do que 2 acidentes, tal probabilidade pode ser facilmente calculada com recurso à função massa de probabilidade

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = 0.355,$$

ou com recurso à função de distribuição

$$P(X \geq 2) = 1 - P(X < 2) = 1 - F_X(1) = 1 - 0.645 = 0.355.$$

Este último caso, torna-se bastante útil quando a variável pode tomar inúmeros valores.

## Mais geralmente, tem-se que:

- $P(X \leq a) = F_X(a)$ ;
- $P(X < a) = P(X \leq a) - P(X = a) = F_X(a) - P(X = a)$ ;
- $P(X > a) = 1 - P(X \leq a) = 1 - F_X(a)$ ;
- $P(X \geq a) = 1 - P(X < a) = 1 - F_X(a) + P(X = a)$ ;
- $P(a < X \leq b) = F_X(b) - F_X(a)$ ;
- $P(a < X < b) = P(a < X \leq b) - P(X = b) = F_X(b) - F_X(a) - P(X = b)$ ;
- $P(a \leq X < b) = P(X = a) + P(a < X \leq b) - P(X = b) =$   
 $= P(X = a) + F_X(b) - F_X(a) - P(X = b)$ ;
- $P(a \leq X \leq b) = P(X = a) + P(a < X \leq b) = P(X = a) + F_X(b) - F_X(a)$ .

# Momentos de uma variável aleatória discreta

Depois de definidas importantes propriedades sobre o cálculo de probabilidades em variáveis aleatórias discretas, importa ter em consideração algumas medidas caracterizadoras destas variáveis.

Seja  $X$  uma variável aleatória discreta com função massa de probabilidade dada por

$$X = \begin{cases} x_k & k \in \kappa \\ p_k = P(X = x_k) \end{cases}$$

O **valor médio** (ou valor esperado, ou esperança matemática, ou momento de 1ª ordem) de  $X$  é

$$E(X) = \sum_{k \in \kappa} x_k p_k,$$

# Momentos de uma variável aleatória discreta

exigindo-se convergência absoluta da série, isto é

$$\sum_{k \in \kappa} |x_k| p_k < \infty.$$

Consideremos o exemplo do número de acidentes numa linha férrea. Vimos que a função massa de probabilidade da variável aleatória  $X$  era definida por

$x_i$	0	1	2	3	4	5	6
$p_i = P(X = x_i)$	0.287	0.358	0.224	0.093	0.029	0.007	0.002

pelo que o seu valor médio é dado por

$$E(X) = 0 \times 0.287 + 1 \times 0.358 + 2 \times 0.224 + 3 \times 0.093 + 4 \times 0.029 + 5 \times 0.007 + 6 \times 0.002 = 1.2$$

Conclui-se que o número médio de acidentes é de aproximadamente 1.

## ● Propriedades do valor médio

1. O valor médio de uma constante é essa constante.  
Considerando,

$$K = \begin{cases} K \\ 1 \end{cases}$$

que se trata de uma variável aleatória degenerada, tem-se que o seu valor médio é dado por  $E(K) = K \times 1 = K$ .  
Em particular,

$$E(E(X)) = E(X).$$

# Momentos de uma variável aleatória discreta

2. O valor médio de  $X^n$ , chamado momento de ordem  $n$ , é

$$E(X^n) = \sum_{k \in \kappa} x_k^n p_k,$$

se  $\sum_{k \in \kappa} |x_k^n| p_k < \infty$ .

Considerando a transformação de uma variável aleatória,  $Y = g(X)$ , então

$$E(Y) = E[g(X)] = \sum_{k \in \kappa} g(x_k) p_k,$$

se  $\sum_{k \in \kappa} |g(x_k)| p_k < \infty$ .

# Momentos de uma variável aleatória discreta

3. Dados  $a, b \in \mathbb{R}$  quaisquer, se existir  $E(X)$ , então,

$$E(aX + b) = aE(X) + b.$$

4. O valor médio da soma de duas variáveis aleatórias é a soma dos repetivos valores médios. Isto é,

$$E(X + Y) = E(X) + E(Y).$$

Deixa-se a prova deste resultado como exercício.

5. Se  $X$  e  $Y$  são variáveis aleatórias independentes, então

$$E(X.Y) = E(X).E(Y).$$

# Momentos de uma variável aleatória discreta

Mais geralmente, sendo  $X_1, X_2, \dots, X_n$  variáveis aleatórias quaisquer,

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i).$$

Se as variáveis aleatórias forem **i.i.d.** (independentes e idênticamente distribuídas), então

$$E(X_i) = E(X) = \mu, \forall i \in \{1, \dots, n\},$$

donde,

$$E\left(\sum_{i=1}^n X_i\right) = n.E(X) = n\mu.$$

# Momentos de uma variável aleatória discreta

Dada uma amostra aleatória, a sua média é uma medida de localização de bastante importância, uma vez que é a medida utilizada para fazer inferência sobre o valor médio populacional. Define-se como

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

O seu valor esperado é dado por

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \cdot n\mu = \mu.$$

A variância de uma variável aleatória é uma medida de dispersão da variável em torno do seu valor médio, sendo definida por

# Momentos de uma variável aleatória discreta

$$V(X) = E\{[X - E(X)]^2\},$$

se  $E(X^2) < \infty$ .

Da expressão anterior resulta ainda que,

$$\begin{aligned}V(X) &= E\{[X - E(X)]^2\} \\&= E\{X^2 - 2X \cdot E(X) + [E(X)]^2\} \\&= E(X^2) - 2E(X)E(X) + [E(X)]^2 \\&= E(X^2) - E(X)^2\end{aligned}$$

Este resultado é também conhecido como *Teorema de König*, tendo bastante utilidade prática.

# Momentos de uma variável aleatória discreta

Consideremos novamente o exemplo do número de acidentes numa linha férrea. Sabemos que  $E(X) = 1.248$ , pelo que  $E(X)^2 = (1.248)^2 = 1.56$ . Para o cálculo da variância resta apenas saber  $E(X^2)$ . Repare-se que  $E(X^2)$  corresponde ao momento de ordem 2, pelo que é calculado da seguinte forma,

$$E(X^2) = \sum_x x^2 p_x.$$

Temos então que

$$E(X) = 0^2 \times 0.287 + 1^2 \times 0.358 + 2^2 \times 0.224 + 3^2 \times 0.093 + 4^2 \times 0.029 + 5^2 \times 0.007 + 6^2 \times 0.002 = 2.802$$

## ● Propriedades da variância

1. A variância de uma constante é 0;
2. Se  $Y = \alpha X + \beta$ ,  $V(Y) = \alpha^2 V(X)$ . Sendo  $\beta$  uma constante, naturalmente que não tem qualquer influência na variância de  $Y$ .
3.  $V(X + Y) = V(X) + V(Y) + 2cov(X, Y)$ ;
4.  $V(X - Y) = V(X) + V(Y) - 2cov(X, Y)$ ;
5.  $V(X + Y) = V(X - Y) = V(X) + V(Y)$  se  $X$  e  $Y$  forem variáveis aleatórias independentes;

# Momentos de uma variável aleatória discreta

6. Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias i.i.d., em que  $V(X) = \sigma^2$ . Então,

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = \sum_{i=1}^n \sigma^2 = n\sigma^2.$$

7. A variância da média é dada por

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

# Modelos discretos

**Modelar** é estabelecer uma aproximação à realidade. Tendo em conta a característica que se observa, é estabelecer um modelo probabilístico que caracteriza o comportamento dessa característica com o objetivo de estabelecer inferências.

As variáveis aleatórias discretas resultam essencialmente de **contagens**. De seguida são apresentados os principais modelos de probabilidade utilizados quando estamos na presença de variáveis aleatórias discretas.

## ● Modelo de Bernoulli

O modelo de Bernoulli é dos modelos de probabilidade mais simples, uma vez que existem apenas duas possibilidades (0 ou 1), genericamente chamadas de "sucesso" e "insucesso", em que 1 representa "sucesso" e 0 "insucesso". Consideremos o seguinte exemplo que ilustra bem a aplicabilidade deste modelo de probabilidade.

**Exemplo:** Foi levado a cabo um estudo por parte da seguradora "Seguríssima" sobre a satisfação dos seus clientes relativamente ao seguro que possuem atualmente (trata-se apenas da satisfação de um único seguro).

# Modelos discretos

A questão colocada é muito simples e possui apenas duas possibilidades de resposta: "Estou satisfeito", "Não estou satisfeito". De anos anteriores, a seguradora sabe que a probabilidade de sucesso ("Estar satisfeito") com o seguro em estudo é de 0.43. Procedeu-se então à seleção aleatória de um cliente e colocou-se a questão. Seja  $X$  a variável aleatória que indica o grau de satisfação do cliente. A função massa de probabilidade é definida por

$$\begin{array}{c|cc} x_i & 0 & 1 \\ \hline p_i = P(X = x_i) & 1-0.43 & 0.43 \end{array}$$

# Modelos discretos

Os modelos de probabilidade têm a particularidade de serem constituídos por parâmetros (desconhecidos) e que normalmente são estimados com recurso a uma amostra representativa da população. Caso não existissem parâmetros, não faria sentido inferirmos nada sobre aquela que acreditamos ser a distribuição da população que estamos a estudar.

No caso do modelo de Bernoulli, este é constituído por um único parâmetro,  $p$ , que designa a probabilidade de sucesso, ou seja,  $P(X = 1)$ , representando-se por  $X \sim \text{Bernoulli}(p)$ .

O complementar da probabilidade de sucesso,  $1 - p$ , naturalmente designará a probabilidade de insucesso, ou seja,  $P(X = 0)$ .

# Modelos discretos

Considerando que  $X \sim \text{Bernoulli}(p)$ , a sua função massa de probabilidade é dada por

$$\begin{array}{c|cc} x_i & 0 & 1 \\ \hline p_i = P(X = x_i) & 1 - p & p \end{array}$$

O valor médio e a variância de uma variável aleatória com distribuição  $\text{Bernoulli}(p)$  são dados por:

- $E(X) = p$ ;
- $V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p)$ .

# Modelos discretos

Uma sucessão de provas de *Bernoulli* é uma sequência de experiências aleatórias em que:

- em cada prova apenas interessa se o resultado é um sucesso (1) ou um insucesso (0);
- a probabilidade de sucesso é constante (mantém-se de prova para prova);
- o resultado de cada prova é independente do resultado de qualquer das remanescentes.

## • Modelo Binomial

Considerando que numa determinada experiência estamos a realizar um conjunto de  $n$  provas de Bernoulli, naturalmente que o resultado será uma sequência de realizações de  $n$  provas de *Bernoulli*, ou seja, uma sequência de zeros e uns cujo significado já foi exposto anteriormente. Nestes casos, não estamos interessados na ordem pela qual os sucessos ocorrem, mas sim, no **número de sucessos** entre as  $n$  realizações.

Assim, diz-se que uma variável aleatória  $X$  segue uma distribuição Binomial de parâmetros  $n$  e  $p$ , em que  $n$  designa o número de provas de Bernoulli e  $p$  a probabilidade de sucesso

# Modelos discretos

$$X \sim \text{Binomial}(n, p),$$

com f.m.p. dada por

$$X : \begin{cases} k \\ \binom{n}{k} p^k (1-p)^{n-k} \end{cases} \quad k = 0, 1, \dots, n$$

Prova-se facilmente que

- $E(X) = np$ ;
- $V(X) = np(1-p)$ .

**Exemplo:** Retomemos o exemplo abordado no modelo de *Bernoulli*. Consideremos que foram selecionados de forma aleatória 10 clientes. Para estes 10 clientes pretende-se saber quantos estão satisfeitos com o seguro. Seja então  $X$  a variável aleatória que representa o número de indivíduos satisfeitos, em 10.

$$X \sim \text{Binomial}(10, 0.43),$$

pelo que a correspondente f.m.p. é dada por

$$P(X = x) = \binom{n}{x} (0.43)^x (1 - 0.43)^{10-x}, x = 0, 1, \dots, 10.$$

# Modelos discretos

Estando completamente especificada a distribuição de probabilidade, podemos então responder a diversas questões, como por exemplo, dos 10 indivíduos, qual a probabilidade de 2 estarem satisfeitos? Tal resulta em

$$P(X = 2) = \binom{10}{2} 0.43^2 (1 - 0.43)^{10-2} = 0.0927.$$

- **Modelo Hipergeométrico**

Numa dada uma população com dimensão  $N$ , sabe-se que  $M$  dos seus elementos possuem um atributo  $A$  (e  $N - M$  não o possuem). Retiram-se  $n$  elementos dessa população, ao acaso e **sem reposição**.

A v.a.  $X$ : “n<sup>o</sup> de elementos, de entre os  $n$ , que possuem o atributo  $A$ ” tem distribuição Hipergeométrica, ou seja,

$$X \sim HG(N, n, M).$$

# Modelos discretos

A f.m.p. da v.a.  $X$  é dada por

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, k = 0, 1, 2, \dots, \min(n, M) \text{ e } n-k \leq N-M$$

O valor médio e a variância da v.a.  $X$  são dados por

$$E(X) = n \cdot \frac{M}{N}; \quad \text{Var}(X) = \frac{n \cdot M \cdot (N - M) \cdot (N - n)}{N^2 \cdot (N - 1)}.$$

# Modelos discretos

**Exemplo:** Qual a probabilidade de, em 5 das lâmpadas, retiradas sem reposição de uma caixa que contém 8 lâmpadas azuis e 2 brancas, haver apenas 1 branca?

Seja  $X$  a v.a. que representa o número de lâmpadas brancas de entre as 5 retiradas. Tem-se que

$$X \sim HG(10, 5, 2).$$

Pretende-se

$$P(X = 1) = \frac{\binom{2}{1} \binom{10-2}{5-1}}{\binom{10}{5}} = 0.556.$$

## ● Modelo de Poisson

O modelo de Poisson é adequado quando se pretende estudar o número de ocorrências de um certo acontecimento num determinado intervalo de tempo ou região do espaço.

Suponhamos que se verificam as seguintes hipóteses:

1. a ocorrência do acontecimento num determinado intervalo é independente da ocorrência do acontecimento em qualquer outro intervalo distinto;
2. a probabilidade de exatamente uma ocorrência do acontecimento em qualquer intervalo de amplitude  $h$  arbitrariamente pequena é aproximadamente  $\lambda h$ ;

# Modelos discretos

3. a probabilidade de duas ou mais ocorrências do acontecimento em qualquer intervalo de amplitude  $h$  arbitrariamente pequena é aproximadamente 0;

dizemos que a variável aleatória,  $X$ , que representa o número de ocorrências do acontecimento, num intervalo unitário tem distribuição de Poisson de parâmetro  $\lambda$ , ( $X \sim P(\lambda)$ ), com f.m.p. dada por

$$P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}; \quad k = 0, 1, 2, \dots; \quad \lambda > 0.$$

## Exemplos:

- 1) número de consultas a uma base de dados por minuto;
- 2) número de novos casos de hepatite numa certa região do país;
- 3) número de bactérias num determinado volume de água.

O valor médio e a variância de  $X$  são dados por:

- $E(X) = \lambda$ ;
- $V(X) = \lambda$ .

# Modelos discretos

**Exemplo:** O número de alunos que chegam à caixa do bar num intervalo de 5 minutos é uma v.a. com distribuição de Poisson. Sabendo que em cada 5 minutos chegam em média 3 alunos, determine a probabilidade de:

- (a) chegarem 4 alunos
- (b) não chegarem mais de 2 alunos
- (c) chegar nenhum aluno

# Modelos discretos

**Resolução:** Seja  $X$  a v.a. que representa o número de alunos que chegam à caixa de um bar no intervalo de 5 minutos. Tem-se que

$$X \sim Pois(3).$$

(a) : Pretende-se  $P(X = 4)$ , ou seja,

$$P(X = 4) = e^{-3} \cdot \frac{3^4}{4!} = 0.168.$$

(b) : Pretende-se  $P(X \leq 2)$ , ou seja,

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-3} \frac{3^0}{0!} + e^{-3} \frac{3^1}{1!} + e^{-3} \frac{3^2}{2!} = 0.4232$$

# Modelos discretos

(c) Não chegar nenhum aluno, significa que  $X = 0$ , ou seja,

$$P(X = 0) = e^{-3} \cdot \frac{3^0}{0!} = 0.0498.$$

## CAPÍTULO 4

### Variáveis aleatórias e modelos contínuos

# Variáveis aleatórias contínuas

Até agora estudaram-se variáveis que tomam valores num conjunto finito ou infinito numerável, as chamadas variáveis aleatórias discretas. Para estas variáveis vimos como estabelecer a sua função de probabilidade, ou seja, a **função massa de probabilidade** que a cada valor que a variável toma **associa a respetiva probabilidade da variável tomar esse valor**, isto é,

$$\begin{array}{c|cccccc} x_i & x_1 & x_2 & x_3 & \dots & x_n \\ \hline p_i = P(X = x_i) & p_1 & p_2 & p_3 & \dots & p_n \end{array}$$

Contudo, existem diversas situações em que uma variável aleatória pode tomar valores num conjunto infinito não numerável, remetendo-nos para as variáveis aleatórias absolutamente contínuas.

# Variáveis aleatórias contínuas

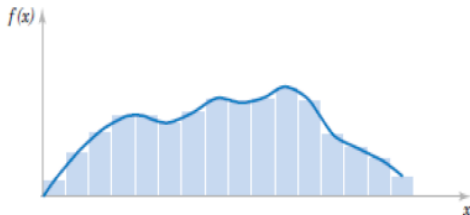
O suporte de uma variável aleatória contínua não contém pontos com massa de probabilidade estritamente positiva, nem contém pontos isolados, ou seja,

$$P(X = x) = 0, \forall x \in \mathbb{R}.$$

Assim, para as variáveis aleatórias contínuas, pretende-se saber qual a probabilidade de a variável tomar valores num determinado intervalo.

No primeiro capítulo vimos que poderíamos representar graficamente variáveis aleatórias contínuas através de um histograma. A frequência relativa corresponde a uma estimativa da probabilidade de um certo valor se situar num dado intervalo.

# Variáveis aleatórias contínuas



A função que permite calcular a probabilidade da variável aleatória  $X$  pertencer a um dado intervalo é chamada de **função densidade de probabilidade** e será denotada por  $f(x)$ . Podemos então dizer que o histograma é uma aproximação da função densidade.

# Variáveis aleatórias contínuas

Tem-se então que

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

Uma função densidade de probabilidade verifica as seguintes condições:

1. uma função não negativa, pois é a derivada de uma função não decrescente;
2.  $\int_{-\infty}^{+\infty} f(x)dx = 1$ .

# Variáveis aleatórias contínuas

Recorde-se que a função de distribuição de uma v.a.  $X$  é dada por

$$F_X(x) = P(X \leq x), \forall x \in \mathbb{R},$$

que no caso contínuo faz corresponder a cada valor de  $x$  a probabilidade associada à semi-reta  $(-\infty, x]$ . Ou seja,

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt, \forall x \in \mathbb{R},$$

sendo  $f(\cdot)$  a função densidade.

# Variáveis aleatórias contínuas

- Propriedades da função de distribuição:

(a)  $0 \leq F(x) \leq 1$ ;

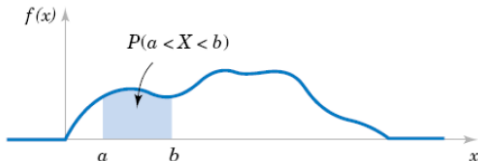
(b)  $\lim_{x \rightarrow -\infty} F(x) = 0; \lim_{x \rightarrow +\infty} F(x) = 1$ ;

(c)  $F(x)$  é uma função monótona não decrescente;

(d)  $F(x)$  é uma função contínua em  $\mathbb{R}$ .

# Variáveis aleatórias contínuas

Consideremos uma v.a.  $X$  e o intervalo  $[a, b]$ , com  $a < b$ .  
Tem-se que



Mais concretamente, uma vez que no caso contínuo

$P(X = x) = 0, \forall x \in \mathbb{R}$ , tem-se que

$$P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f(x)dx.$$

# Variáveis aleatórias contínuas

A função densidade de probabilidade é utilizada para descrever o modelo de probabilidade da v.a.  $X$ .

Note-se que, por definição, a função de distribuição é dada por

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x)dx,$$

pelo que é possível concluir que a função de distribuição é uma primitiva da função densidade de probabilidade.

# Variáveis aleatórias contínuas

**Exemplo:** A função densidade de probabilidade de uma variável aleatória é dada por

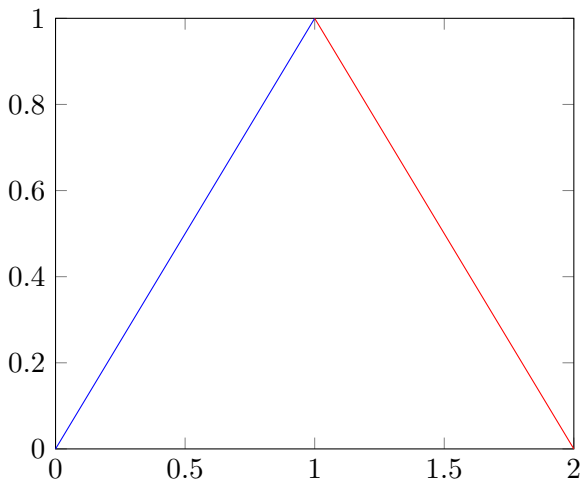
$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2 - x & 1 < x < 2 \\ 0 & \text{c.c} \end{cases}$$

Determine  $P(X \leq 1.8)$ . **Resolução:** Repare que a função densidade pode ser representada graficamente como

# Variáveis aleatórias contínuas

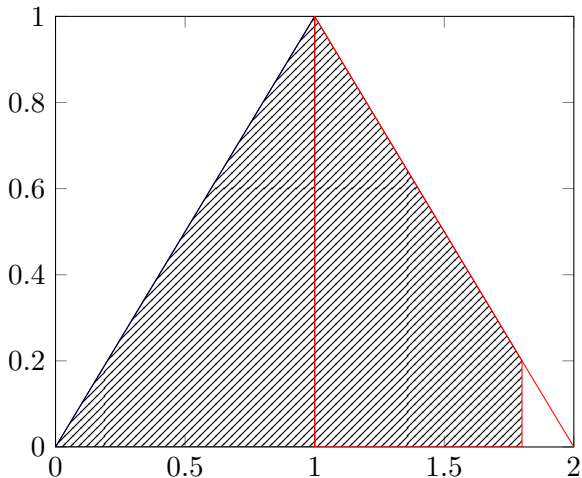
Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



# Variáveis aleatórias contínuas

Tem-se que  $P(X \leq 1.8)$  corresponde a calcular a seguinte área



# Variáveis aleatórias contínuas

Temos então que

$$\begin{aligned}P(X \leq 1.8) &= \int_{-\infty}^{1.8} f(x)dx = \int_0^{1.8} f(x)dx \\&= \int_0^1 f(x)dx + \int_1^{1.8} f(x)dx \\&= \int_0^1 xdx + \int_1^{1.8} 2 - xdx \\&= \left[\frac{x^2}{2}\right]_0^1 + \left[2x - \frac{x^2}{2}\right]_1^{1.8} = \frac{1}{2} + 0.48 = 0.98\end{aligned}$$

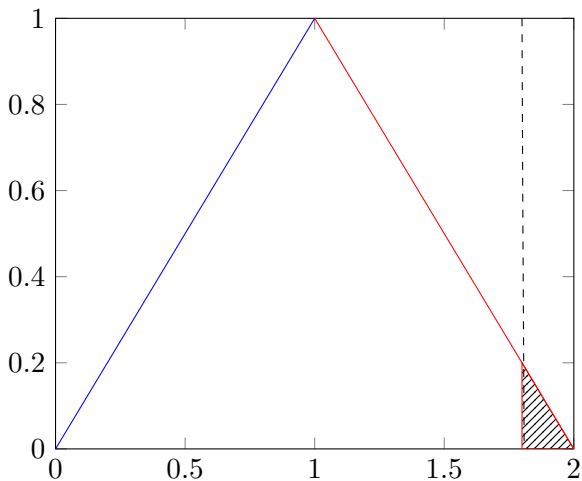
Note que a probabilidade anterior também poderia ser calculada utilizando o complementar, ou seja,

$P(X \leq 1.8) = 1 - P(X > 1.8)$ , pelo que a área a calcular seria

# Variáveis aleatórias contínuas

Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



# Momentos de uma variável aleatória contínua

- **Valor médio,  $E(X)$ :**

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx.$$

- **Variância:**

$$V(X) = E(X^2) - E^2(X),$$

em que  $E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx.$

# Momentos de uma variável aleatória contínua

## • Propriedades do valor médio e da variância

O valor médio e a variância de uma v.a. contínua gozam das mesmas propriedades que o valor médio e variância de uma v.a. discreta. Assim,

### -Valor médio:

(a)  $E(k) = k$

(b)  $E(kX) = kE(X)$

(c)  $E(aX \pm bY) = aE(X) \pm bE(Y)$

(d) Se  $X$  e  $Y$  são v.a.'s independentes então  
 $E(XY) = E(X)E(Y)$

# Momentos de uma variável aleatória contínua

## -Variância:

(a)  $V(k) = 0$

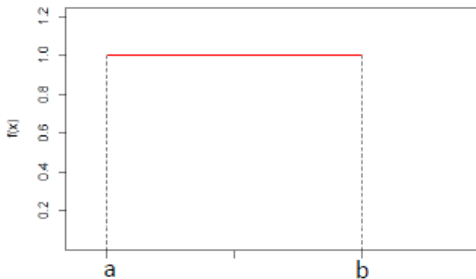
(b)  $V(kX) = k^2V(X)$

(c) Se  $X$  e  $Y$  são independentes então  
 $V(X + Y) = V(X) + V(Y)$

# Modelos contínuos

- **Modelo Uniforme:**

Uma v.a.,  $X$ , tem distribuição uniforme num intervalo  $[a, b]$  se a probabilidade de  $X$  pertencer a qualquer subintervalo é proporcional ao comprimento do mesmo. Escreve-se  $X \sim U(a, b)$ .



# Modelo Uniforme

A função densidade de probabilidade é dada por

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{se } x \notin [a, b] \end{cases}$$

A função de distribuição é dada por

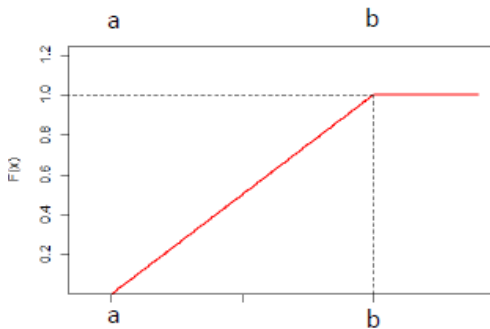
$$F(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x < b \\ 1 & \text{se } x \geq b \end{cases}$$

Graficamente a função de distribuição é dada por

# Modelo Uniforme

Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



# Modelo Uniforme

O valor médio e variância de uma v.a. com distribuição Uniforme no intervalo  $[a,b]$  são dados por

$$E(X) = \frac{a + b}{2}, \quad V(X) = \frac{(b - a)^2}{12}.$$

# Modelo Uniforme

**Exemplo:** O tempo de espera (em minutos) num consultório,  $X$ , é uma variável aleatória com distribuição Uniforme no intervalo  $(0,10)$ . Determine  $P(X \leq 3)$ . **Resolução:** De acordo com as informações disponibilizadas, tem-se que  $X \sim U(0, 10)$ , pelo que  $a = 0$  e  $b = 10$ . A função densidade de probabilidade é dada por

$$f(x) = \begin{cases} \frac{1}{10} & \text{se } x \in [0, 10] \\ 0 & \text{se } x \notin [0, 10] \end{cases}$$

A função de distribuição é dada por

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{x}{10} & \text{se } 0 \leq x < 10 \\ 1 & \text{se } x \geq 10 \end{cases}$$

# Modelo Uniforme

Assim,

$$P(X \leq 3) = F(3) = \frac{3}{10} = 0.3.$$

# Modelo Exponencial

Uma v.a  $X$  tem distribuição Exponencial com parâmetro  $\lambda$  se a sua função densidade de probabilidade é dada por

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \lambda e^{-\lambda x} & \text{se } x \geq 0 \text{ e } \lambda > 0 \end{cases}$$

A correspondente função de distribuição é dada por

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-\lambda x} & \text{se } x \geq 0 \text{ e } \lambda > 0 \end{cases}$$

Escreve-se  $X \sim Exp(\lambda)$ .

# Modelo Exponencial

O valor médio e a variância de uma v.a.,  $X$ , com distribuição Exponencial( $\lambda$ ) são dados respetivamente por

$$E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}.$$

**A distribuição Exponencial é utilizada quando se pretende modelar o tempo que decorre entre dois acontecimentos consecutivos de Poisson( $\lambda$ ) num intervalo de tempo  $T$ . O intervalo de tempo é o considerado na distribuição de Poisson.**

# Modelo Exponencial

## • Relação entre o modelo exponencial e o modelo de Poisson:

Repare que o tempo entre ocorrências consecutivas pode ser expresso em função do número de ocorrências. Considere novamente o exemplo aplicado ao caso do modelo de Poisson (número de alunos que chegam a uma caixa de bar). A v.a. que representa o número de chegadas de alunos à caixa do bar num intervalo de 5 minutos tem distribuição  $X \sim Pois(3)$ .

Se considerarmos agora a variável aleatória  $T$  que representa o tempo (em minutos) entre chegadas consecutivas de alunos à caixa do bar, tem-se que

$$T \sim Exp(3/5).$$

# Modelo Exponencial

Podemos então colocar a questão sobre o qual a probabilidade de que o tempo entre chegadas consecutivas seja superior a 2 minutos?

Temos então que a probabilidade pretendida é

$$P(T > 2) = 1 - P(T \leq 2) = 1 - (1 - e^{-\frac{3}{5} \times 2}) = 0.301.$$

Ora, mas dizer que o tempo entre chegadas consecutivas excede 2 minutos é o mesmo que dizer que em 2 minutos não existiram chegadas. Então,

$$P(T > 2) = P(X = 0) = e^{-\frac{6}{5}} \frac{(6/5)^0}{0!} = 0.301.$$

# Modelo Exponencial

Suponhamos agora que é sabido que não chegaram alunos à caixa do bar nos primeiros 3 minutos. Qual a probabilidade de haver chegadas nos 3 minutos seguintes?

Neste caso, a probabilidade pedida é expressa através de uma probabilidade condicional, isto é,

$$P(T < 6 | T > 3) = \frac{P(3 < T < 6)}{P(T < 6)} = \frac{F(6) - F(3)}{1 - F(3)} = 0.834701 = P(T < 3).$$

**Quer isto dizer que a distribuição exponencial não tem memória.**

# Modelo Normal

O modelo Normal ou Gaussiano é amplamente utilizado na prática dadas as suas boas propriedades. É constituído por dois parâmetros (um de localização e outro de dispersão), nomeadamente o valor médio que denotaremos por  $\mu$  e o desvio padrão,  $\sigma$ . Desta forma, uma variável aleatória  $X$  que tenha distribuição gaussiana ou normal é representada por

$$X \sim N(\mu, \sigma),$$

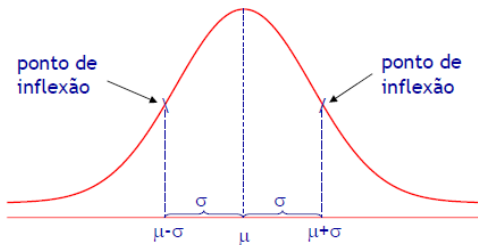
em que  $E(X) = \mu$  e  $V(X) = \sigma^2$ .

# Modelo Normal

A sua função densidade de probabilidade é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+.$$

A representação gráfica da função densidade de probabilidade da distribuição Normal é a seguinte



# Modelo Normal

- **Algumas propriedades:**

- trata-se de uma distribuição simétrica em torno do seu valor médio;
- tem dois pontos de inflexão de abcissas  $\mu - \sigma$  e  $\mu + \sigma$

A função de distribuição de uma v.a.  $X \sim N(\mu, \sigma)$  é dada por

$$P(X \leq x) = \int_{-\infty}^x f(x)dx = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}u^2} du, \forall x \in \mathbb{R}.$$

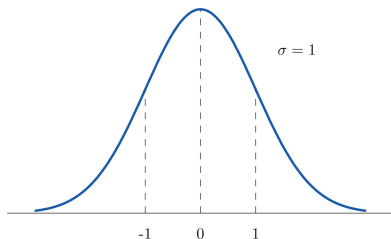
# Modelo Normal

Em muitas situações da vida prática, trabalhar com a função densidade de probabilidade tal como foi apresentada pode não ser uma boa estratégia, uma vez que o tratamento analítico dessa função pode ser algo complexo. Desta forma, existe uma forma de centrar e reduzir uma variável aleatória normal, denominada de **standardização**.

Se  $X \sim N(\mu, \sigma)$ , então  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ , a que chamamos de normal padrão, ou normal standard.

A vantagem de trabalhar com uma variável aleatória normal standardizada reside no facto de existirem os quantis tabelados para esta distribuição. A representação da f.d.p. da normal padrão é

# Modelo Normal



Trata-se, portanto, de uma distribuição simétrica em torno da origem.

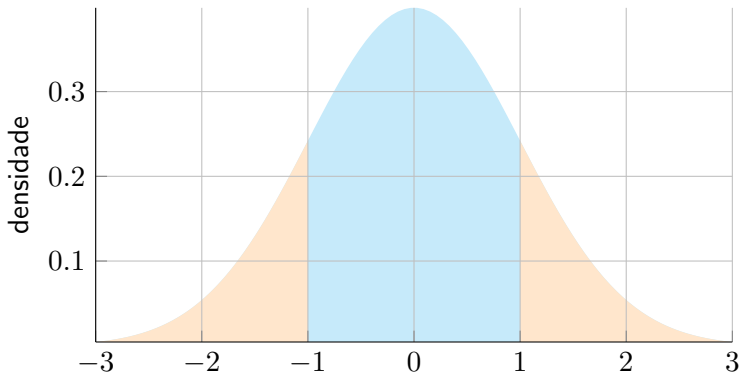
Temos então que, se  $X \sim N(\mu, \sigma)$ ,

- $P(X \leq a) = P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$ ;
- $P(a < X \leq b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$ .

# Modelo Normal

Consideremos  $Z \sim N(0, 1)$ ,  $a = -1$  e  $b = 1$ , pela propriedade da simetria temos que

$$P(Z \leq -a) = P(Z \geq a) = 1 - P(Z \leq a), \text{ ou seja,}$$



# Modelo Normal

Tendo em conta que  $P(X \leq a) = \Phi(a)$ , podemos então concluir do que foi referido anteriormente que  $\Phi(-a) = 1 - \Phi(a), \forall a \in \mathbb{R}$ .

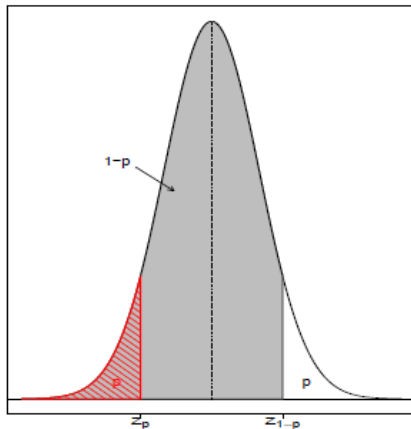
Seja  $Z \sim N(0, 1)$ . Denota-se o quantil de probabilidade  $p$  da distribuição normal standard como sendo  $z_p$ , tendo-se pela definição de quantil

$$z_p \Leftrightarrow P(Z \leq z_p) = p.$$

# Modelo Normal

Estadística  
Aplicada à  
Gestão

Tiago Dias  
Domingues



# Modelo Normal

## Tabela da distribuição Normal

Standard Normal Table —  $P(Z \leq z)$

$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706

$P(Z < 1.1) = 0.8643$

$z_{0.9236} = 1.43$

# Modelo Normal

**Exemplo:** Sabe-se que a concentração de estrôncio (em mg/ml) na água de um determinado rio é uma variável aleatória com valor médio  $\mu = 8$  mg/ml e desvio padrão  $\sigma = 1.5$  mg/ml.

Qual a probabilidade de se encontrar uma concentração

- (a) inferior a 9 mg/ml?
- (b) superior a 5 mg/ml?
- (c) entre 4 mg/ml e 10 mg/ml?
- (d) qual o valor do quantil de probabilidade 0.975 da variável aleatória em causa?

# Modelo Normal

**Resolução:** Seja  $X$  a v.a. que representa a concentração de estrôncio (em mg/ml) na água de um rio. Sabe-se que

$$X \sim N(8, 1.5).$$

(a) Pretende-se

$$P(X < 9) = P\left(\frac{X - 8}{1.5} < \frac{9 - 8}{1.5}\right) = P(Z < 0.67) = 0.7486.$$

$$\begin{aligned} \text{(b)} \quad P(X > 5) &= 1 - P(X \leq 5) = 1 - P\left(\frac{X - 8}{1.5} \leq \frac{5 - 8}{1.5}\right) = \\ &= 1 - P(Z \leq -2) = 1 - \Phi(-2) = 1 - (1 - \Phi(2)) = \Phi(2) = \\ &= 0.9772. \end{aligned}$$

# Modelo Normal

(c) Neste caso,

$$\begin{aligned}P(4 \leq X \leq 10) &= P\left(\frac{4-8}{1.5} \leq \frac{X-8}{1.5} \leq \frac{10-8}{1.5}\right) \\&= P(-2.67 \leq Z \leq 1.33) = \Phi(1.33) - \Phi(-2.67) \\&= \Phi(1.33) - (1 - \Phi(2.67)) = 0.9082 - (1 - 0.9962) = 0.9044.\end{aligned}$$

# Modelo Normal

- (d) Pretende-se saber qual o valor de  $z_{0.975}$ . Tem-se então  $P(Z \leq z_{0.975}) = 0.975$ . Consultando a tabela, tem-se

$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

# Modelo Normal

Estatística  
Aplicada à  
Gestão

Tiago Dias  
Domingues

Assim,  $z_{0.975} = 1.96$ .

# Teorema da Aditividade da Normal

Sejam  $X_1, X_2, \dots, X_n$  v.a.'s independentes e idênticamente distribuídas (i.i.d.) em que  $X_i \sim N(\mu_i, \sigma_i), i = 1, \dots, n$ ; considerando  $a_1, a_2, \dots, a_n$  números reais quaisquer, tem-se que

$$T = \sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sqrt{\sum_{i=1}^n a_i^2 \sigma_i^2}\right).$$

**Corolário:** Se  $X_1, X_2, \dots, X_n$  são v.a.'s independentes e idênticamente distribuídas (i.i.d.) em que  $X_i \sim N(\mu, \sigma), i = 1, \dots, n$ ; considerando, tem-se que

$$T = \sum_{i=1}^n X_i \sim N\left(n\mu, \sigma\sqrt{n}\right).$$

# Distribuição da média de variáveis aleatórias

Consideremos uma amostra aleatória de dimensão  $n$ , isto é,  $(X_1, X_2, \dots, X_n)$  extraída de uma população  $X \sim N(\mu, \sigma)$ . A média da amostra aleatória é dada por

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Repare que  $\bar{X}$  é uma função das v.a.'s  $X_i$ , pelo que será também uma v.a. Em particular, a média é uma combinação linear de v.a.'s, uma vez que

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \sum_{i=1}^n \frac{1}{n} X_i.$$

# Distribuição da média de variáveis aleatórias

Pelo Teorema da Aditividade da Normal, tem-se que a média de uma amostra aleatória extraída de uma população  $X \sim N(\mu, \sigma)$  ainda tem distribuição Normal com parâmetros,

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Para a obtenção do valor médio da média e da respetiva variância, basta notar que

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu. \end{aligned}$$

Deixa-se como exercício o resultado da variância.

# Teorema Limite Central (T.L.C.)

**Caso I:** Sejam  $X_1, X_2, \dots, X_n$  v.a.'s independentes e idênticamente distribuídas (i.i.d.) com valor médio e variância finitos, isto é,  $E(X_i) = \mu$  e  $V(X) = \sigma^2$ , então para  $n$  suficientemente grande ( $n \geq 30$ ),

$$T = \sum_{i=1}^n X_i \curvearrowright^a N\left(n\mu, \sigma\sqrt{n}\right).$$

**Caso II:** Sejam  $X_1, X_2, \dots, X_n$  v.a.'s independentes e idênticamente distribuídas (i.i.d.) com valor médio e variância finitos, isto é,  $E(X_i) = \mu$  e  $V(X) = \sigma^2$ ; seja  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , então para  $n$  suficientemente grande ( $n \geq 30$ ),

$$\bar{X} \curvearrowright^a N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

## CAPÍTULO 5

### Inferência Estatística: estimação pontual intervalar

# Estimação

Recordamos que o objetivo principal da Estatística é o de dado uma amostra recolhida de uma População, organizar e resumir a informação amostral e inferir/extrapolar a informação resumida para a População.

O **processo de estimação**, no qual assenta a Inferência Estatística (ramo que nos permite tirar conclusões de uma amostra para a População) é utilizado quando determinado parâmetro de interesse é desconhecido, tentando-se obter a partir dos dados informação que nos permita ter uma ideia do valor (desconhecido) do parâmetro que está a ser alvo de estudo. Caso a informação seja traduzida num único valor, diz-se que o processo de estimação é **pontual**.

# Estimação

Quando obtemos um valor para o parâmetro (desconhecido) da População a partir de uma amostra, dizemos que obtemos uma **estimativa desse parâmetro**. Essa estimativa é produzida por um **estimador**, que é uma função da amostra.

Contudo, pode haver interesse, para além da produção de uma estimativa pontual, obter um intervalo que envolva a estimativa pontual e ao qual se pode atribuir uma certa confiança de conter o verdadeiro valor do parâmetro. A este processo chama-se de **estimação intervalar**.

**Exemplo:** Suponhamos que uma v.a.  $X$  tem distribuição Normal, sendo o valor médio,  $\mu$ , e o desvio padrão,  $\sigma$ , desconhecidos.

Naturalmente pretendemos que os estimadores a utilizar para a estimação do valor médio e do desvio padrão sejam o mais precisos possível, sendo por isso necessário que os estimadores possuam algumas propriedades:

# Estimação

1. deve ser **centrado**, ou seja, o seu valor médio deve ser igual ao parâmetro que se pretende estimar. Por exemplo, se o parâmetro de interesse é o valor médio,  $\mu$ , o estimador indicado é a média amostral da amostra

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

uma vez que  $E(\bar{X}) = \mu$ , qualquer que seja a população.

# Estimação

- além de ser centrado, deve apresentar a menor variância possível. Quanto menor for a variância do estimador, mais eficiente este será. No caso do valor médio, a média amostral é um estimador centrado e consistente:

$$V(\bar{X}) = \frac{V(X)}{n} \xrightarrow{n \rightarrow +\infty} 0.$$

A escolha do melhor estimador requer a aplicação dos chamados métodos de estimação, sendo o mais utilizado, o método da máxima verosimilhança.

# Estimação

O que acabámos de referir é que  $\bar{X}$  é um **estimador** para o valor médio  $\mu$ , pelo que o valor de  $\bar{x}$  obtido a partir de uma amostra será uma **estimativa** para  $\mu$ .

**Exemplo:** Com o objetivo de estudar as emissões de monóxido de carbono (g/km) de um certo tipo de veículo (que se supõe circular a uma velocidade de 90km/h), foi recolhida uma amostra, tendo-se obtido os seguintes resultados:

17.3   17.8   18.0   17.7   18.2   17.4   17.6   18.1

# Estimação

A amostra foi retirada de uma população que se assume como tendo distribuição Normal.

Neste contexto, uma **estimativa pontual** para o valor médio  $\mu$  da população da qual os dados foram recolhidos será:

$$\bar{x} = \frac{\sum_{i=1}^8 x_i}{8} = \frac{142.1}{8} = 17.7625g/km.$$

Caso nos fosse pedido para estimar a variância populacional,  $\sigma^2$ , um estimador natural seria a variância amostral,  $s^2$ .

# Estimação

## Em resumo:

Parâmetro	Estimador	Estimativa
Valor médio ( $\mu$ )	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Variância ( $\sigma^2$ )	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
Proporção ( $p$ )	$p = \frac{1}{n} \sum_{i=1}^n X_i$	$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$

# Estimação intervalar

Tal como referimos anteriormente, além da produção de estimativas pontuais para um determinado parâmetro desconhecido, é possível contruir um intervalo de valores que com um certo grau de confiança irá conter o verdadeiro valor do parâmetro a que chamamos **intervalo de confiança**. Denotamos o grau de confiança de um intervalo por  $(1 - \alpha)$ .

Note-se que a construção de um intervalo de confiança fornece uma informação mais fina sobre os estimadores escolhidos, nomeadamente sobre os rigor das estimativas produzidas.

# Estimação intervalar

## Construção de um intervalo de confiança:

Dada uma amostra aleatória  $X_1, X_2, \dots, X_n$ , seja  $\theta$  o parâmetro a estimar.

Considere-se uma estatística  $T$ , função da amostra e de  $\theta$ , cuja distribuição é conhecida e não depende de  $\theta$ , isto é, trata-se de uma **variável fulcral**.

Determinam-se os valores de  $a$  e  $b$  tais que  
 $P(a < T < b) = 1 - \alpha$ .

Resolve-se a dupla desigualdade,  $a < T < b$ , em ordem a  $\theta$ , o que permite obter

$$P(L_1 < \theta < L_2) = 1 - \alpha.$$

# Estimação intervalar

Assim, o intervalo de confiança é  $(L_1, L_2)$ .

- **Exemplo (Construção do intervalo de confiança para o valor médio no caso de populações normais)**

Seja  $X_1, X_2, \dots, X_n$  uma amostra aleatória de uma população com distribuição normal,  $X_i \sim N(\mu, \sigma)$ . Considere-se  $\sigma$  conhecido. Como pretendemos um intervalo de confiança para  $\mu$ , temos que o estimador é  $\bar{X}$ .

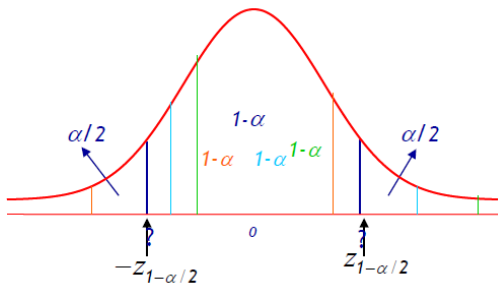
Sabe-se que

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right),$$

pelo que  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$  é a nossa variável fulcral.

# Estimação intervalar

Em particular, pretendemos  $P(a < Z < b) = 1 - \alpha$ , sendo  $1 - \alpha$  o grau de confiança do intervalo. Temos então que



e portanto  $a = -z_{1-\alpha/2}$  e  $b = z_{1-\alpha/2}$ .

# Estimação intervalar

Assim,

$$\begin{aligned}P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) &= 1 - \alpha \Leftrightarrow \\ \Leftrightarrow P(-z_{1-\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}) &= 1 - \alpha \Leftrightarrow \\ \Leftrightarrow P(-z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \Leftrightarrow \\ \Leftrightarrow P(-\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < -\mu < \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \Leftrightarrow\end{aligned}$$

# Estimação intervalar

$$\Leftrightarrow P\left(\bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \Leftrightarrow$$

$$\Leftrightarrow P\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha,$$

e portanto, um intervalo de confiança para  $\mu$  é dado por

$$\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right).$$

# Estimação intervalar

**Interpretação do intervalo de confiança:** Se recolhêssemos 100 amostras aleatórias todas da mesma dimensão e se para cada uma delas fosse calculado o intervalo acima referido, então aproximadamente 95 desses intervalos iriam conter o verdadeiro valor do parâmetro, enquanto que 5 deles não iriam conter.

Podemos assim concluir que a construção do intervalo de confiança para um parâmetro depende da distribuição da variável fulcral, havendo os seguintes cenários para o caso do valor médio ( $\mu$ ):

## Intervalos de confiança para $\mu$

- Populações normais,  $\sigma$  conhecido

$$\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right).$$

- Populações normais,  $\sigma$  desconhecido

$$\left(\bar{X} - t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}\right).$$

- Quaisquer populações, grandes amostras

$$\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right).$$

# Estimação intervalar

Consideremos novamente o exemplo das emissões de monóxido de carbono, isto é:

**Exemplo:** Com o objetivo de estudar as emissões de monóxido de carbono (g/km) de um certo tipo de veículo (que se supõe circular a uma velocidade de 90km/h), foi recolhida uma amostra, tendo-se obtido os seguintes resultados:

17.3   17.8   18.0   17.7   18.2   17.4   17.6   18.1

A amostra foi retirada de uma população que se assume como tendo distribuição Normal.

**Pergunta:** encontre um intervalo de confiança a 95% para  $\mu$ .

# Estimação intervalar

**Parâmetro desconhecido:**  $\mu$                       **Estimador:**  $\bar{X}$

**Pressupostos:** População Normal,  $\sigma$  desconhecido.

**Grau de confiança:**  $95\% \Rightarrow 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$ .

Nestas condições, um intervalo de confiança para  $\mu$  é dado por

$$\left( \bar{X} - t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right).$$

Precisamos de estimar o desvio padrão da amostra  $s$  e o quantil de probabilidade  $t_{1-\alpha/2}$

# Estimação intervalar

Como  $\alpha = 0.05$ , pretende-se

$$t_{1-0.05/2}(8-1) = t_{0.975}(7) = 2.364624.$$

A média e desvio padrão da amostra são dados por:

$\bar{x} = 17.7625$ ;  $s = 0.3248626$ , pelo que

$$\left(17.7625 - 2.364624 \times \frac{0.3248626}{\sqrt{8}}, 17.7625 + 2.364624 \times \frac{0.3248626}{\sqrt{8}}\right). \\ = (17.49091, 18.03409).$$

## Intervalo de confiança para $\sigma^2$ - populações normais

- Variável fulcral

$$X^2 = (n - 1) \frac{S^2}{\sigma^2} \sim \chi^2(n - 1).$$

Pelo que um intervalo de confiança para  $\sigma^2$  a  $(1 - \alpha) \times 100\%$  é dado por

$$\left( \frac{(n - 1)S^2}{\chi_{1-\alpha/2}^2(n - 1)}; \frac{(n - 1)S^2}{\chi_{\alpha/2}^2(n - 1)} \right).$$

## Intervalo de confiança para $p$ - populações de Bernoulli

- Variável fulcral

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{a}{\sim} N(0, 1).$$

Pelo que um intervalo de confiança para  $p$  a  $(1 - \alpha) \times 100\%$  é dado por

$$\left( \hat{p} - z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right).$$

## CAPÍTULO 6

### Inferência Estatística: testes de hipóteses (uma população)

# Testes de hipóteses

Os testes de hipóteses foram uma poderosa ferramenta desenvolvida para a tomada de decisões e consequentes conclusões. Em Estatística nada é certo até prova em contrário, e portanto durante todo o processo, estamos a lidar com o acaso, para o qual pretendemos minimizar todas as decisões devidas ao mesmo. Considere-se o seguinte caso que permite introduzir de uma forma bastante clara o conceito dos testes de hipóteses:

"Culpado ou inocente?"

# Testes de hipóteses

Quando um indivíduo é notificado para ir a tribunal, este é inocente até prova em contrário. Cabe ao juiz, com base nas provas recolhidas (amostra) tomar a decisão de condenar ou absolver o indivíduo.

Temos portanto duas hipóteses em cima da mesa: 1) o indivíduo é inocente; 2) o indivíduo é culpado. A informação que se assume como sendo verdadeira (o indivíduo é inocente) constitui em termos estatísticos a hipótese nula que denotamos por  $H_0$ . O complementar desta hipótese constitui a hipótese alternativa que denotamos por  $H_1$  e cuja informação se traduz na condenação do indivíduo.

# Testes de hipóteses

Contudo, as decisões que tomamos não estão isentas de erros, pelo que é necessário controlá-los. Consideremos o quadro seguinte que ilustra os tipos de erro que podem ocorrer num teste de hipóteses.

<b>Hipótese/Decisão tomada</b>	<b>Rejeitar <math>H_0</math></b>	<b>Não rejeitar <math>H_0</math></b>
$H_0$ verdadeira	erro	decisão correta
$H_0$ falsa	decisão correta	erro

# Testes de hipóteses

De acordo com o quadro anterior, podemos constatar a existência de dois tipos de erro. O primeiro é o erro que resulta de rejeitar  $H_0$  quando esta é verdadeira. Este é também chamado de **erro de 1<sup>a</sup> espécie ou nível de significância do teste**, que iremos denotar por  $\alpha$ . Podemos ainda escrever que

$$\alpha = P(\text{Rejeitar } H_0 | H_0 \text{ verdadeira}).$$

O segundo tipo de erro que resulta de não rejeitar  $H_0$  sabendo que  $H_0$  é falsa, é chamado de erro de 2<sup>a</sup> espécie e denota-se por  $\beta$ . Repare-se que cometer um erro de 1<sup>a</sup> espécie é muito mais grave do que cometer um erro de 2<sup>a</sup> espécie. Basta notar que no primeiro caso e segundo o nosso exemplo, estamos a condenar um inocente.

# Testes de hipóteses

Relativamente ao nível de significância ( $\alpha$ ) consideram-se como níveis usuais: 1%, 5% e 10%. Por exemplo,  $\alpha = 1\%$  significa que é possível rejeitar  $H_0$  sabendo que esta é verdadeira em 1% das vezes. Trata-se de um erro controlado e que é fixado *a priori*.

Definimos ainda a **potência do teste** como sendo o complementar do erro de 2<sup>a</sup> espécie, ou seja,

$$1 - \beta = P(\text{Rejeitar } H_0 | H_0 \text{ falsa}).$$

**Note que a forma de minimizar simultaneamente o erro de 1<sup>a</sup> e 2<sup>a</sup> espécie é aumentar a dimensão da amostra.**

# Testes de hipóteses

Com a realização de um teste de hipóteses, pretende-se então conjecturar sobre o valor de um determinado parâmetro (populacional) de interesse. Suponhamos que o parâmetro que se pretende estudar é o valor médio,  $\mu$ . Relativamente ao valor que conjecturamos para o parâmetro e de acordo com o objetivo da construção do teste de hipóteses, podemos ter três situações:

- 1) conjecturar se o valor médio é superior a um determinado valor,  $\mu_0$ ;
- 2) conjecturar se o valor médio seja inferior a um determinado valor  $\mu_0$ ;
- 3) conjecturar se o valor médio é diferente de um determinado valor,  $\mu_0$ .

# Testes de hipóteses

Os dois primeiros casos são chamados de testes de hipóteses **unilaterias direito e esquerdo**, respetivamente. O terceiro caso diz respeito aos testes **bilaterais**. Formalmente,

- 1)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu > \mu_0$ ;
- 2)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu < \mu_0$ ;
- 3)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu \neq \mu_0$ .

O valor que se assume para  $\mu_0$  consiste em experiência passada, ou seja, conhecimento prévio sobre o problema em estudo. O que pretendemos saber é se com o conjunto de dados que estamos a trabalhar, existe evidência (ou não) para rejeitarmos  $H_0$  a favor de  $H_1$ , sob a veracidade da hipótese nula.

# Testes de hipóteses

A questão que surge neste momento é, como quantificar esta evidência?

Repare que o nosso objetivo continua a ser o de realizar inferências sobre os parâmetros de uma dada população e que estas inferências dependem da amostra que é recolhida dessa mesma população.

Considerando, por exemplo, o valor médio  $\mu$ , vimos que para realizar inferências sobre este parâmetro ao nível da estimação intervalar, era necessário definir a variável fulcral que dependia apenas do parâmetro de interesse e tinha distribuição de amostragem completamente especificada.

# Testes de hipóteses

Para a realização de um teste de hipóteses vamos continuar a necessitar dessas variáveis, que neste contexto serão chamadas de **estatística de teste**.

## Exemplo - estatísticas de teste para o valor médio

- Populações normais,  $\sigma$  conhecido

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

- Populações normais,  $\sigma$  desconhecido

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n - 1).$$

# Testes de hipóteses

- Quaisquer populações, grandes amostras ( $n \geq 30$ )

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \text{ou} \quad Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim^a N(0, 1)$$

Depois de definidas as estatísticas de teste, é necessário proceder à tomada de decisão do teste, ou seja, **de rejeitar ou não rejeitar** a hipótese nula (uma vez que é esta que é assumida como sendo verdadeira).

# Testes de hipóteses

**Exemplo:** Com o objetivo de estudar as emissões de monóxido de carbono (g/km) de um certo tipo de veículo (que se supõe circular a uma velocidade de 90km/h), foi recolhida uma amostra, tendo-se obtido os seguintes resultados:

17.3   17.8   18.0   17.7   18.2   17.4   17.6   18.1

A amostra foi retirada de uma população que se assume como tendo distribuição Normal. Normas ambientais estabelecem que o valor médio das emissões deve ser no máximo de 18 g/km. Considerando a amostra recolhida acha que estas normas estão a ser cumpridas? Considere um nível de significância de 2.5%. Para o teste realizado indique qual o tipo de erro em que pode incorrer.

# Testes de hipóteses

Neste exmplo, temos que o parâmetro em estudo é o valor médio  $\mu$  que representa o as emissões médias de monóxido de carbono.

O segundo passo consiste em estabelecer as hipóteses nula e alternativa. Neste caso tem-se que

$$H_0 : \mu \leq 18 \quad \text{vs.} \quad H_1 : \mu > 18.$$

Considerando que estamos na presença de uma população normal com  $\sigma$  desconhecido, tem-se que a estatística de teste é dada por

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n - 1).$$

# Testes de hipóteses

Depois de definida a estatística de teste pretende-se então averiguar se os dados são consentâneos com as hipóteses estabelecidas.

Note-se que se  $H_1$  for verdadeira, então esperamos que tal se traduza nos dados, isto é, esperamos que a média da amostra seja, efectivamente, maior do que 18. Ou seja, se  $H_1$  for verdadeira, o valor observado da estatística de teste deverá ser um valor positivo, grande em valor absoluto. Quanto mais longe de 0, para a direita, este valor estiver, mais apontam dos dados no sentido de  $H_1$  ser mais plausível do que  $H_0$ . Calculemos então o valor observado de  $T$ :

# Testes de hipóteses

$$T_{obs} = \frac{17.7625 - 18}{0.3248626/\sqrt{8}} = -2.0678.$$

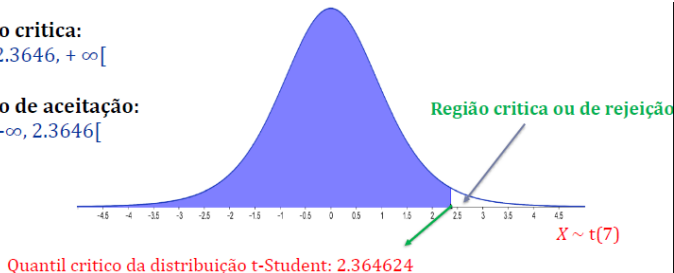
Fixado o erro de 1ª espécie, tem-se que é possível definir a região de rejeição e de aceitação do teste.

**Região crítica:**

$$[2.3646, +\infty[$$

**Região de aceitação:**

$$]-\infty, 2.3646[$$



# Testes de hipóteses

Posto isto, vamos verificar se o valor observado da estatística pertence à região crítica ou à região de aceitação. Note que se

$$T_{obs} > t_{1-\alpha}(7),$$

então rejeita-se  $H_0$ , caso contrário a hipótese nula não é rejeitada.

Como  $T_{obs} = -2.0678 < t_{0.975}(7) = 2.3646$ , não se rejeita  $H_0$ . Concluimos que não existe evidência para afirmar que as emissões médias de monóxido de carbono sejam superiores a 18 g/km.

# Testes de hipóteses

Uma forma alternativa de tomar a decisão num teste de hipóteses é através do cálculo do **p-value**. O p-value não é mais do que a probabilidade da estatística de teste produzir um valor tão ou mais extremo no sentido da hipótese alternativa do que o valor produzido com base nos dados observados.

Neste caso,

$$p - value = P(T \geq T_{obs})$$

Ou seja,

$$p - value = P(T \geq T_{obs}) = P(T \geq -2.0678) =$$

$$1 - P(T < -2.0678) = 1 - F(-2.0678) = 1 - (1 - F(2.0678)) =$$

# Testes de hipóteses

$$= 1 - (1 - 0.961266) = 0.961266.$$

De seguida procede-se à comparação do p-value com o nível de significância utilizado e conclui-se que

- se  $\text{p-value} \leq \alpha$ , então rejeita-se  $H_0$ ;
- se  $\text{p-value} > \alpha$ , então não se rejeita  $H_0$ .

Neste exemplo, como  $\text{p-value} = 0.961266 > 0.05$ , não se rejeita  $H_0$ , tal como seria de esperar pela decisão tomada com base no valor observado da estatística de teste.

## Em resumo (regras de decisão):

■ População normal,  $\sigma$  desconhecido:

1)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu > \mu_0$ ;

2)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu < \mu_0$ ;

3)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu \neq \mu_0$ .

1) rejeita-se  $H_0$  se  $T_{obs} \geq t_{1-\alpha}(n-1)$ ;

2) rejeita-se  $H_0$  se  $T_{obs} \leq t_{\alpha}(n-1)$ ;

3) rejeita-se  $H_0$  se  $|T_{obs}| \geq t_{1-\alpha/2}(n-1)$ ;

# Testes de hipóteses

## ■ População normal, $\sigma$ conhecido:

1)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu > \mu_0$ ;

2)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu < \mu_0$ ;

3)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu \neq \mu_0$ .

1) rejeita-se  $H_0$  se  $Z_{obs} \geq z_{1-\alpha}$ ;

2) rejeita-se  $H_0$  se  $Z_{obs} \leq z_{\alpha}$ ;

3) rejeita-se  $H_0$  se  $|Z_{obs}| \geq z_{1-\alpha/2}$ ;

# Testes de hipóteses

■ Qualquer população, grandes amostras ( $n \geq 30$ ):

1)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu > \mu_0$ ;

2)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu < \mu_0$ ;

3)  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu \neq \mu_0$ .

1) rejeita-se  $H_0$  se  $Z_{obs} \geq z_{1-\alpha}$ ;

2) rejeita-se  $H_0$  se  $Z_{obs} \leq z_{\alpha}$ ;

3) rejeita-se  $H_0$  se  $|Z_{obs}| \geq z_{1-\alpha/2}$ ;

## Teste de hipóteses para a variância ( $\sigma^2$ ):

- 1)  $H_0 : \sigma^2 = \sigma_0^2$  vs.  $H_1 : \sigma^2 > \sigma_0^2$ ;
- 2)  $H_0 : \sigma^2 = \sigma_0^2$  vs.  $H_1 : \sigma^2 < \sigma_0^2$ ;
- 3)  $H_0 : \sigma^2 = \sigma_0^2$  vs.  $H_1 : \sigma^2 \neq \sigma_0^2$ .

**Estatística de Teste:**  $X^2 = (n - 1) \frac{S^2}{\sigma_0^2} \sim \chi^2(n - 1)$ .

- 1) rejeita-se  $H_0$  se  $X_{obs}^2 \geq \chi_{1-\alpha}^2(n - 1)$ ;
- 2) rejeita-se  $H_0$  se  $X_{obs}^2 \leq \chi_{\alpha}^2(n - 1)$ ;
- 3) rejeita-se  $H_0$  se  $|X_{obs}^2| \geq \chi_{1-\alpha/2}^2(n - 1)$ ;

# Testes de hipóteses

## Teste de hipóteses para a proporção ( $p$ ):

- 1)  $H_0 : p = p_0$  vs.  $H_1 : p > p_0$ ;
- 2)  $H_0 : p = p_0$  vs.  $H_1 : p < p_0$ ;
- 3)  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$ .

**Estatística de Teste:**  $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{a}{\sim} N(0, 1)$ .

- 1) rejeita-se  $H_0$  se  $Z_{obs} \geq z_{1-\alpha}$ ;
- 2) rejeita-se  $H_0$  se  $Z_{obs} \leq z_{\alpha}$ ;
- 3) rejeita-se  $H_0$  se  $|Z_{obs}| \geq z_{1-\alpha/2}$ ;

## CAPÍTULO 7

### Inferência Estatística: testes de hipóteses (duas populações)

# Testes de hipóteses

Até aqui estivemos interessados em realizar inferências sobre os parâmetros de uma população. Contudo, o interesse pode ir mais além, no sentido em que é possível comparar populações quanto aos seus parâmetros. Neste capítulo vamos apenas considerar os casos de duas populações, iniciando com os testes de hipóteses para comparação de valores médios.

**Exemplo:** Suponhamos que estamos interessados em comparar o tempo médio de vida de pilhas da marca A e da marca B. Para tal, foram observadas 10 pilhas de cada uma das marcas, tendo-se obtido os seguintes resultados:

<b>marca A:</b>	85.6	77.4	78.5	80.2	96.3	88.3	82.1	80.9	79.5	79.2
<b>marca B:</b>	64.8	65.4	68.7	72.2	74.5	75.4	76.4	76.9	78.5	79.0

# Testes de hipóteses para comparação de valores médios

De acordo com o exemplo anterior temos duas características em estudo: o tempo de vida das pilhas e a respetiva marca ( $A$  ou  $B$ ). Será natural portanto, considerar duas populações  $X_A$  e  $X_B$  que dizem respeito ao tempo de vida das pilhas da marca  $A$  e da marca  $B$ , respetivamente, dando origem a duas amostras consoante a marca das pilhas.

As amostras que se podem retirar de duas populações podem ser classificadas em **independentes** ou **emparelhadas**.

- **amostras independentes:** os indivíduos ou unidades experimentais são distribuídos pelos grupos de forma aleatória, sendo os grupos **mutuamente exclusivos**.

# Testes de hipóteses para comparação de valores médios

- **amostras emparelhadas:** os indivíduos ou unidades experimentais são observados aos pares (por exemplo em diferentes momentos no tempo).

De acordo com o exemplo anterior podemos então classificar as duas amostras como sendo independentes. Por uma questão de simplicidade, consideremos que as duas populações  $X_A$  e  $X_B$  do exemplo anterior têm distribuição normal com valores médios  $\mu_A$ ,  $\mu_B$  e desvios padrões  $\sigma_A$  e  $\sigma_B$ , respetivamente, isto é

$$X_A \sim N(\mu_A, \sigma_A), \quad X_B \sim N(\mu_B, \sigma_B).$$

# Testes de hipóteses para comparação de valores médios

Pretende-se então saber se existe evidência para afirmar se o tempo médio de vidas das pilhas das duas marcas é ou não igual. Podemos então formalizar as seguintes hipóteses:

$$H_0 : \mu_A = \mu_B \quad vs. \quad H_1 : \mu_A \neq \mu_B.$$

# Testes de hipóteses para comparação de valores médios

Repare-se que testar a hipótese de igualdade dos valores médios é equivalente a testar a hipótese de que a diferença seja nula, isto é

$$H_0 : \mu_A - \mu_B = 0 \quad vs. \quad H_1 : \mu_A - \mu_B \neq 0,$$

pelo que o parâmetro em estudo será a diferença dos valores médios ( $\mu_A - \mu_B$ ). Um bom estimador da diferença dos valores médios será a diferença das médias, ou seja,  $\bar{X}_A - \bar{X}_B$ . Em particular,

$$E(\bar{X}_A - \bar{X}_B) = \mu_A - \mu_B \quad e \quad var(\bar{X}_A - \bar{X}_B) = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}.$$

# Testes de hipóteses para comparação de valores médios

Facilmente compreendemos que as reais diferenças (em termos médios) entre duas populações podem ser influenciadas pela variabilidade inerente a cada uma das populações. Nesse sentido, o parâmetro de dispersão das populações  $A$  e  $B$  são encarados como **parâmetros perturbadores**.

Assim sendo, quando se pretende comparar valores médios de duas **populações normais para as quais as variâncias são desconhecidas**, é necessário ter consideração se as **variâncias se podem considerar iguais ou diferentes**.

# Testes de hipóteses para comparação de valores médios

## Amostras independentes

### 1. Populações normais, variâncias conhecidas

#### Hipóteses a testar:

- 1)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B > \delta_0$
- 2)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B < \delta_0$
- 3)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B \neq \delta_0$

#### Estatística de teste:

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)_0}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \underset{H_0}{\sim} N(0, 1)$$

# Testes de hipóteses para comparação de valores médios

Rejeita-se  $H_0$  se:

- 1)  $Z_{obs} \geq z_{1-\alpha}$
- 2)  $Z_{obs} \leq z_{\alpha}$
- 3)  $|Z_{obs}| \geq z_{1-\alpha/2}$

**Intervalo de confiança a  $(1 - \alpha) \times 100\%$  para  $\mu_A - \mu_B$**

$$\left( \bar{x}_A - \bar{x}_B \pm z_{1-\alpha/2} \times \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \right)$$

# Testes de hipóteses para comparação de valores médios

## 2. Populações normais, variâncias desconhecidas e iguais ( $\sigma_A^2 = \sigma_B^2 = \sigma^2$ )

### Hipóteses a testar:

- 1)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B > \delta_0$
- 2)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B < \delta_0$
- 3)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B \neq \delta_0$

Como as variâncias são desconhecidas e são consideradas iguais, o parâmetro  $\sigma^2$  precisa de ser estimado a partir da amostra, sendo o melhor estimador a variância,  $s^2$ . Contudo, como temos a informação de duas amostras independentes a ter em conta aquando da estimação, o que se vai estimar é a chamada **variância ponderada** (pelas duas amostras), isto é,

# Testes de hipóteses para comparação de valores médios

$$S_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$$

Assim, a estatística de teste vem dada por

$$T = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)_0}{S_p \times \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \underset{H_0}{\sim} t(n_A + n_B - 2)$$

Rejeita-se  $H_0$  se:

- 1)  $T_{obs} \geq t_{1-\alpha}(n_A + n_B - 2)$
- 2)  $T_{obs} \leq t_{\alpha}(n_A + n_B - 2)$
- 3)  $|T_{obs}| \geq t_{1-\alpha/2}(n_A + n_B - 2)$

# Testes de hipóteses para comparação de valores médios

**Intervalo de confiança a  $(1 - \alpha) \times 100\%$  para  $\mu_A - \mu_B$**

$$\left( \bar{x}_A - \bar{x}_B \pm t_{1-\alpha/2}(n_A + n_B - 2) \times s_p \times \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \right)$$

# Testes de hipóteses para comparação de valores médios

## 3. Populações normais, variâncias desconhecidas e diferentes

### Hipóteses a testar:

- 1)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B > \delta_0$
- 2)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B < \delta_0$
- 3)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B \neq \delta_0$

### Estatística de teste:

$$T = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)_0}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \underset{H_0}{\sim} t_v,$$

# Testes de hipóteses para comparação de valores médios

onde  $v = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{s_A^2/n_A}{n_A-1} + \frac{s_B^2/n_B}{n_B-1}}$ . Quando  $v$  não é inteiro, considera-se a parte inteira de  $v$ .

**Como saber se as variâncias se podem considerar iguais ou diferentes? - o teste F (caso de populações normais)**

Pretende-se testar a hipótese:

$$H_0 : \sigma_A^2 = \sigma_B^2 \quad vs. \quad H_1 : \sigma_A^2 \neq \sigma_B^2.$$

# Testes de hipóteses para comparação de valores médios

Repare-se que as hipóteses relativas à variância podem ainda ser escritas como

$$H_0 : \frac{\sigma_A^2}{\sigma_B^2} = 1 \quad \text{vs.} \quad H_1 : \frac{\sigma_A^2}{\sigma_B^2} \neq 1.$$

A estatística de teste é dada por

$$F = \frac{S_A^2}{S_B^2} \left( \frac{\sigma_A^2}{\sigma_B^2} \right)_0 \underset{H_0}{\sim} F(n_A - 1, n_B - 1),$$

onde  $F(n_A - 1, n_B - 1)$  designa a distribuição F-Snedecor com  $(n_A - 1, n_B - 1)$  graus de liberdade. Deixa-se como observação que  $F_\alpha(m, n) = \frac{1}{F_{1-\alpha}(n, m)}$ .

# Testes de hipóteses para comparação de valores médios

Em geral, é ainda possível estabelecer as seguintes hipóteses

- 1)  $H_0 : \sigma_A^2 = \sigma_B^2$  vs.  $H_1 : \sigma^2 > \sigma_B^2$
- 2)  $H_0 : \sigma_A^2 = \sigma_B^2$  vs.  $H_1 : \sigma^2 < \sigma_B^2$
- 3)  $H_0 : \sigma_A^2 = \sigma_B^2$  vs.  $H_1 : \sigma^2 \neq \sigma_B^2$

Rejeita-se  $H_0$  se:

- 1)  $F_{obs} \geq F_{1-\alpha}(n_A - 1, n_B - 1)$
- 2)  $F_{obs} \leq F_{\alpha}(n_A - 1, n_B - 1)$
- 3)  $|F_{obs}| \geq F_{1-\alpha/2}(n_A - 1, n_B - 1)$

# Testes de hipóteses para comparação de valores médios

No caso de afastamentos à distribuição normal, utiliza-se o **teste de Levene** para testar a hipótese de igualdade das variâncias.

**Intervalo de confiança a  $(1 - \alpha) \times 100\%$  para  $\frac{\sigma_A^2}{\sigma_B^2}$**

$$\left( \frac{s_A^2/s_B^2}{F_{1-\alpha/2}(n_A - 1, n_B - 1)}; \frac{s_A^2/s_B^2}{F_{\alpha/2}(n_A - 1, n_B - 1)} \right)$$

# Testes de hipóteses para comparação de valores médios

## 4. Quaisquer populações, grandes amostras

### Hipóteses a testar:

- 1)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B > \delta_0$
- 2)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B < \delta_0$
- 3)  $H_0 : \mu_A - \mu_B = \delta_0$  vs.  $H_1 : \mu_A - \mu_B \neq \delta_0$

### (i) variâncias conhecidas

#### Estatística de teste:

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)_0}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \underset{H_0}{\sim} N(0, 1),$$

# Testes de hipóteses para comparação de valores médios

## (ii) variâncias desconhecidas (iguais ou diferentes)

### Estatística de teste:

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)_0}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \underset{H_0}{\sim} N(0, 1),$$

Em ambos os casos (i) e (ii)), rejeita-se  $H_0$  se

- 1)  $Z_{obs} \geq z_{1-\alpha}$
- 2)  $Z_{obs} \leq z_{\alpha}$
- 3)  $|Z_{obs}| \geq z_{1-\alpha/2}$

# Testes de hipóteses para comparação de valores médios

Consideremos novamente os dados de tempo de vida das pilhas da marca  $A$  e da marca  $B$ .

**marca A:** 85.6 77.4 78.5 80.2 96.3 88.3 82.1 80.9 79.5 79.2

**marca B:** 64.8 65.4 68.7 72.2 74.5 75.4 76.4 76.9 78.5 79.0

Pretende-se então saber se existe evidência para afirmar que os tempos de vida sejam diferentes consoante a marca utilizada.

As hipóteses em teste são

$$H_0 : \mu_A - \mu_B = 0 \quad vs. \quad H_1 : \mu_A - \mu_B \neq 0.$$

# Testes de hipóteses para comparação de valores médios

Admitindo que as populações são normais (porque nos é dito), estamos nas condições de **populações normais e variâncias desconhecidas**. Resta saber se as variâncias de podem considerar iguais ou diferentes. Para já vamos começar por apresentar algumas características amostrais para as duas amostras, nomeadamente a média, variância e desvio padrão.

	Marca $A$ ( $n = 10$ )	Marca $B$ ( $n = 10$ )
Média	82.8	73.18
Variância	33.63	27.20
Desvio padrão	5.80	5.22

# Testes de hipóteses para comparação de valores médios

- **Testar a hipótese de igualdade das variâncias - teste F**  
Pretende-se testar

$$H_0 : \sigma_A^2 = \sigma_B^2 \quad vs. \quad H_1 : \sigma_A^2 \neq \sigma_B^2$$

Tem-se então que o valor observado da estatística de teste é

$$F_{obs} = \frac{33.63}{27.20} \times 1 = 1.24$$

Como se trata de um teste bilateral, rejeita-se  $H_0$  se  $F_{obs} \geq F_{1-\alpha/2}(10-1, 10-1)$ . Considerando  $\alpha = 0.05$ , tem-se  $F_{0.975}(9, 9) = 4.0260$ . Como  $1.24 < 4.0260$ , não se rejeita  $H_0$ . Assim, existe evidência para afirmar que as variâncias são iguais.

# Testes de hipóteses para comparação de valores médios

Estamos então em condições de populações normais, variâncias desconhecidas e iguais, pelo que a estatística de teste para o teste de comparação de valores médios é dada por

$$T = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)_0}{S_p \times \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \underset{H_0}{\sim} t(n_A + n_B - 2)$$

Vamos então calcular a variância ponderada:

$$s_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2} = \frac{(9 \times 33.63) + (9 \times 27.20)}{10 + 10 - 2} = 30.415$$

# Testes de hipóteses para comparação de valores médios

$$\text{Assim, } s_p = \sqrt{s_p^2} = \sqrt{30.415} = 5.51 \text{ e,}$$

$$T_{obs} = \frac{(82.8 - 73.18) - 0}{5.51 \times \sqrt{\frac{1}{10} + \frac{1}{10}}} = 3.90$$

Tratando-se de um teste bilateral, rejeita-se  $H_0$  se  $T_{obs} \geq t_{1-\alpha/2}(10 + 10 - 2)$ . Considerando  $\alpha = 0.05$ , tem-se  $t_{0.975}(18) = 2.10$ . Como  $3.90 > 2.10$ , rejeita-se  $H_0$ . Existe evidência para afirmar que o tempo médio de vida das pilhas difere consoante a marca.

# Testes de hipóteses para comparação de valores médios

## Amostras emparelhadas

- **Populações normais**

Seja  $(X_1, Y_1), \dots, (X_n, Y_n)$  uma amostra composta por pares de observações de duas populações normais  $X \sim N(\mu_X, \sigma_X)$  e  $Y \sim N(\mu_Y, \sigma_Y)$ . Falamos por exemplo de observações que dizem respeito a diferentes instantes no tempo e o que se pretende saber é se existem diferenças entre esses instantes. Nesse sentido, considera-se a amostra das diferenças

$D = X_i - Y_i, \quad i = 1, \dots, n.$  Prova-se que

$$D = X - Y \sim N(\mu_D, \sigma_D),$$

em que  $\mu_D = \mu_X - \mu_Y$  e  $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2cov(X, Y).$

# Testes de hipóteses para comparação de valores médios

A partir da nova amostra  $D$  é então possível calcular a respetiva média e variância, isto é,

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} \quad \text{e} \quad s_D^2 = \frac{\sum_{i=1}^n d_i^2}{n-1} - \frac{n}{n-1} \bar{d}^2.$$

Hipóteses a testar:

- 1)  $H_0 : \mu_D = \delta_0$  vs.  $H_1 : \mu_D > \delta_0$
- 2)  $H_0 : \mu_D = \delta_0$  vs.  $H_1 : \mu_D < \delta_0$
- 3)  $H_0 : \mu_D = \delta_0$  vs.  $H_1 : \mu_D \neq \delta_0$

# Testes de hipóteses para comparação de valores médios

A estatística de teste é dada por:

$$T = \frac{\bar{D} - \delta_0}{S_D/\sqrt{n}} \underset{H_0}{\sim} t(n-1)$$

Rejeita-se  $H_0$  se:

- 1)  $T_{obs} \geq t_{1-\alpha}(n-1)$
- 2)  $T_{obs} \leq t_{\alpha}(n-1)$
- 3)  $|T_{obs}| \geq t_{1-\alpha/2}(n-1)$

# Testes de hipóteses para comparação de valores médios

- **Quaisquer populações, grandes amostras**

Hipóteses a testar:

1)  $H_0 : \mu_D = \delta_0$  vs.  $H_1 : \mu_D > \delta_0$

2)  $H_0 : \mu_D = \delta_0$  vs.  $H_1 : \mu_D < \delta_0$

3)  $H_0 : \mu_D = \delta_0$  vs.  $H_1 : \mu_D \neq \delta_0$

A estatística de teste é dada por:

$$Z = \frac{\bar{D} - \delta_0}{S_D / \sqrt{n}} \underset{a}{\overset{H_0}{\rightsquigarrow}} N(0, 1)$$

# Testes de hipóteses para comparação de valores médios

Rejeita-se  $H_0$  se:

- 1)  $Z_{obs} \geq z_{1-\alpha}$
- 2)  $Z_{obs} \leq z_{\alpha}$
- 3)  $|Z_{obs}| \geq z_{1-\alpha/2}$

**Exemplo:** Uma companhia de táxis está interessada em saber se o uso de pneus radiais em vez de pneus regulares tem influência no consumo de combustível. Para tal foi fixado um troço a ser percorrido por 12 carros para os quais foram colocados pneus radiais, tendo-se registado o consumo de combustível em  $km/l$ . Posteriormente, aos mesmos carros foram colocados pneus regulares, tendo-se registado novamente o consumo de combustível após terem percorrido o mesmo troço. Os resultados foram os seguintes:

# Testes de hipóteses para comparação de valores médios

Consumo de combustível ( $km/l$ ):

P.radiais	4.2	4.7	6.6	7.0	6.7	4.5	5.7	6.0	7.4	4.9	6.1
P.regulares	4.1	4.9	6.2	6.9	6.8	4.4	5.7	5.8	6.9	4.7	6.0
Diferenças	0.1	-0.2	0.4	0.1	-0.1	0.1	0.0	0.2	0.5	0.2	0.1

De acordo com a amostra das diferenças, obteve-se

$$\bar{d} = 0.142 \quad e \quad s_D = 0.198.$$

Pretende-se testar:

$$H_0 : \mu_D = 0 \quad vs. \quad H_1 : \mu_D \neq 0$$

# Testes de hipóteses para comparação de valores médios

Tratando-se de uma amostra pequena, a estatística de teste é dada por

$$T = \frac{\bar{D} - \delta_0}{S_D/\sqrt{n}} \underset{H_0}{\sim} t(n-1)$$

pelo que o valor observado é  $T_{obs} = \frac{0.142-0}{0.198/\sqrt{12}} = 2.48$ .

Tratando-se de um teste bilateral, rejeita-se  $H_0$  se  $|T_{obs}| \geq t_{1-\alpha/2}(n-1)$ . Assumindo  $\alpha = 0.05$ ,  $t_{0.975}(11) = 2.20$ . Como  $2.48 > 2.20$ , rejeita-se  $H_0$ . Existe evidência para afirmar que o tipo de pneus tem influência no consumo de combustível.

# Testes de hipóteses para comparação de proporções

Sejam  $X_1$  e  $X_2$  duas populações de Bernoulli independentes com parâmetros  $p_1$  e  $p_2$ , respetivamente.

Para comparar estas populações é possível estabelecer as seguintes hipóteses:

- 1)  $H_0 : p_1 - p_2 = \delta_0$  vs.  $H_1 : p_1 - p_2 > \delta_0$
- 2)  $H_0 : p_1 - p_2 = \delta_0$  vs.  $H_1 : p_1 - p_2 < \delta_0$
- 3)  $H_0 : p_1 - p_2 = \delta_0$  vs.  $H_1 : p_1 - p_2 \neq \delta_0$

Um estimador de  $p_1 - p_2$  é  $\hat{p}_1 - \hat{p}_2$ , onde  $\hat{p}_1 = \frac{X_1}{n_1}$  e  $\hat{p}_2 = \frac{X_2}{n_2}$ , em que  $X_1$  e  $X_2$  representam o número de elementos da amostra que pertencem à categoria em estudo.

# Testes de hipóteses para comparação de proporções

## Estatística de teste:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \underset{H_0}{\sim} N(0, 1)$$

Rejeita-se  $H_0$  se:

- 1)  $Z_{obs} \geq z_{1-\alpha}$
- 2)  $Z_{obs} \leq z_{\alpha}$
- 3)  $|Z_{obs}| \geq z_{1-\alpha/2}$

# Testes de hipóteses para comparação de proporções

**Caso particular:** Quando  $p_1 - p_2 = 0 \Leftrightarrow p_1 = p_2 = p$ , tem-se que o estimador comum é dado por

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}.$$

Nesse caso a estatística de teste vem dada por

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\hat{p}(1 - \hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \underset{H_0}{\sim} N(0, 1)$$

# Testes de hipóteses para comparação de proporções

**Exemplo:** Em dada altura, de acordo com um inquérito por amostragem a 1500 cidadãos de um país, a popularidade do governo em funções era de 32.5%. Posteriormente, o governo anunciou um pacote de medidas de austeridade. Algum tempo após este anúncio foi efectuada nova sondagem a 2000 cidadãos, na qual a popularidade do governo desceu 3%. Verifique se a quebra de popularidade do governo é significativa (use  $\alpha = 0.05$ ).

# Testes de hipóteses para comparação de proporções

Seja  $\hat{p}_1$  a percentagem de popularidade inicial do governo e  $\hat{p}_2$  a percentagem de popularidade do governo após o anúncio de medidas de austeridade. Tem-se que  $\hat{p}_1 = 0.325$ ,  $\hat{p}_2 = 0.295$  (quebra de 3% relativamente à popularidade inicial),  $n_1 = 1500$  e  $n_2 = 2000$ . Pretende-se testar:

$$H_0 : p_1 = p_2 \quad vs. \quad H_1 : p_1 > p_2 \Leftrightarrow$$

$$H_0 : p_1 - p_2 = 0 \quad vs. \quad H_1 : p_1 - p_2 > 0$$

O valor observado da estatística é

# Testes de hipóteses para comparação de proporções

$$Z = \frac{(0.395 - 0.295) - 0}{\sqrt{0.308 \times (1 - 0.308) \times \left(\frac{1}{1500} + \frac{1}{2000}\right)}} = 1.902$$

em que

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{(1500 \times 0.325) + (2000 \times 0.295)}{1500 + 2000} = 0.308.$$

Tratando-se de um teste unilateral direito, rejeita-se  $H_0$  se  $Z_{obs} \geq z_{1-\alpha}$ . Assumindo  $\alpha = 0.05$ , tem-se  $z_{0.95} = 1.64$ . Como  $1.902 > 1.64$ , rejeita-se  $H_0$ . Existe evidência para afirmar que a quebra de popularidade é significativa.