

Caro Aluno,

Este material compõe o acervo de recursos educacionais do Sistema Universidade Aberta do Brasil (UAB), disponível no Portal EduCapes, e será adotado no Curso Nacional de Bacharelado em Biblioteconomia na modalidade EaD (BibEAD).O conteúdo e formato originais foram preservados, conforme indicado nas páginas que se seguem. "





# Computação

## Estatística e Probabilidade

Jorge Luiz de Castro e Silva  
Maria Wilda Fernandes  
Rosa Lívia Freitas de Almeida

3ª edição  
Fortaleza - Ceará



2015





Esta obra é licenciada por uma Licença *Creative Commons*: Atribuição – Uso Não Comercial – Não a Obras Derivadas (by-nc-nd). Os termos desta licença estão disponíveis em: <<http://creativecommons.org/licenses/by-nc-nd/3.0/>>. Direitos para esta edição compartilhados entre os autores e a editora EdUECE. Qualquer parte ou a totalidade do conteúdo desta publicação pode ser reproduzida ou compartilhada. Obra sem fins lucrativos e com distribuição gratuita. O conteúdo do livro publicado é de inteira responsabilidade de seus autores, não representando a posição oficial da EdUECE.

Editora Filiada à



**Presidenta da República**

Dilma Vana Rouseff

**Ministro da Educação**

Renato Janine Ribeiro

**Presidente da CAPES**

Carlos Afonso Nobre

**Diretor de Educação a Distância da CAPES**

Jean Marc Georges Mutzig

**Governador do Estado do Ceará**

Camilo Sobreira de Santana

**Reitor da Universidade Estadual do Ceará**

José Jackson Coelho Sampaio

**Vice-Reitor**

Hidelbrando dos Santos Soares

**Pró-Reitor de Pós-Graduação**

Jefferson Teixeira de Souza

**Coordenador da SATE e UAB/UECE**

Francisco Fábio Castelo Branco

**Coordenadora Adjunta UAB/UECE**

Eloísa Maia Vidal

**Diretor do CCT/UECE**

Luciano Moura Cavalcante

**Coordenador da Licenciatura**

**em Computação**

Francisco Assis Amaral Bastos

**Coordenadora de Tutoria e Docência**

**em Computação**

Maria Wilda Fernandes

**Editor da UECE**

Erasmus Miessa Ruiz

**Coordenadora Editorial**

Rocylândia Isidoro de Oliveira

**Projeto Gráfico e Capa**

Roberto Santos

**Diagramador**

Francisco José da Silva Saraiva

**Conselho Editorial**

Antônio Luciano Pontes

Eduardo Diatáhy Bezerra de Menezes

Emanuel Ângelo da Rocha Fragoso

Francisco Horácio da Silva Frota

Francisco Josênio Camelo Parente

Gisafran Nazareno Mota Jucá

José Ferreira Nunes

Liduína Farias Almeida da Costa

Lucili Grangeiro Cortez

Luiz Cruz Lima

Manfredo Ramos

Marcelo Gurgel Carlos da Silva

Marcony Silva Cunha

Maria do Socorro Ferreira Osterne

Maria Salete Bessa Jorge

Silvia Maria Nóbrega-Therrien

**Conselho Consultivo**

Antônio Torres Montenegro (UFPE)

Eliane P. Zamith Brito (FGV)

Homero Santiago (USP)

Ieda Maria Alves (USP)

Manuel Domingos Neto (UFF)

Maria do Socorro Silva Aragão (UFC)

Maria Lírida Callou de Araújo e Mendonça (UNIFOR)

Pierre Salama (Universidade de Paris VIII)

Romeu Gomes (FIOCRUZ)

Túlio Batista Franco (UFF)

E529j Jorge Luiz de Castro e Silva, Maria Wilda Fernandes  
Rosa Livia Freitas de Almeida

Estatística e Probabilidade  
Fortaleza - 2015 – 3ª Edição  
125p.

Publicação do Sistema UAB/UECE, 2010

ISBN: 979-85-89427-32-5  
Estatística e Probabilidade

CDD 668

Editora da Universidade Estadual do Ceará – EdUECE  
Av. Dr. Silas Munguba, 1700 – Campus do Itaperi – Reitoria – Fortaleza – Ceará  
CEP: 60714-903 – Fone: (85) 3101-9893  
Internet: [www.uece.br](http://www.uece.br) – E-mail: [eduece@uece.br](mailto:eduece@uece.br)  
Secretaria de Apoio às Tecnologias Educacionais  
Fone: (85) 3101-9962

# Sumário

<b>Apresentação.....</b>	<b>5</b>
<b>Capítulo 1 – Introdução à Estatística.....</b>	<b>7</b>
1. A Estatística como campo de conhecimento .....	9
2. Porque estudar Estatística .....	10
3. Método estatístico .....	11
<b>Capítulo 2 – População e Amostra .....</b>	<b>13</b>
1. Variáveis .....	15
2. Escala de medidas .....	15
3. População .....	16
4. Censo .....	17
5. Amostra .....	17
<b>Capítulo 3 – Gráficos Estatísticos .....</b>	<b>21</b>
1. Apresentação gráfica .....	23
2. Diagramas .....	23
<b>Capítulo 4 – Tabelas e Séries Estatísticas .....</b>	<b>29</b>
1. Tabelas .....	31
2. Séries estatísticas .....	34
<b>Capítulo 5 – Distribuição de frequência .....</b>	<b>37</b>
1. Sintetizando dados qualitativos .....	39
2. Sintetizando dados quantitativos.....	41
<b>Capítulo 6 – Medidas de Posição.....</b>	<b>57</b>
1. Introdução .....	59
2. Média aritmética ( $\bar{x}$ ).....	59
3. Mediana (Md) .....	63
4. Moda (Mo).....	68
5. Aplicação das medidas de posição.....	72
6. Separatrizes .....	72
<b>Capítulo 7 – Medidas de Dispensão ou de Variabilidade.....</b>	<b>79</b>
1. Dispensão ou variabilidade .....	81
2. Assimetria.....	87
3. Curtose.....	89
<b>Capítulo 8 – Fundamentos de Probabilidade.....</b>	<b>93</b>
1. Introdução .....	95
2. Espaço Amostral e Eventos .....	97

3. Definição de probabilidade.....	98
4. Espaços amostrais finitos e infinitos .....	100
5. Probabilidade condicional .....	101
6. Variáveis aleatórias, funções densidade de probabilidade .....	105
<b>Capítulo 9 – Distribuições estatísticas .....</b>	<b>111</b>
1. Distribuições discretas .....	113
2. Distribuições contínuas .....	116
3. Momentos de uma distribuição de probabilidade.....	121
<b>Sobre os autores .....</b>	<b>125</b>

# Apresentação

Esta publicação constitui-se como um importante recurso posto à disposição dos alunos do curso de licenciatura plena em Computação, na modalidade educação a distância da Universidade Estadual do Ceará, tendo como finalidade apresentar uma introdução aos princípios gerais da Estatística (seu campo de atuação) e aprofundar alguns conceitos importantes ao entendimento do seu campo de estudo.

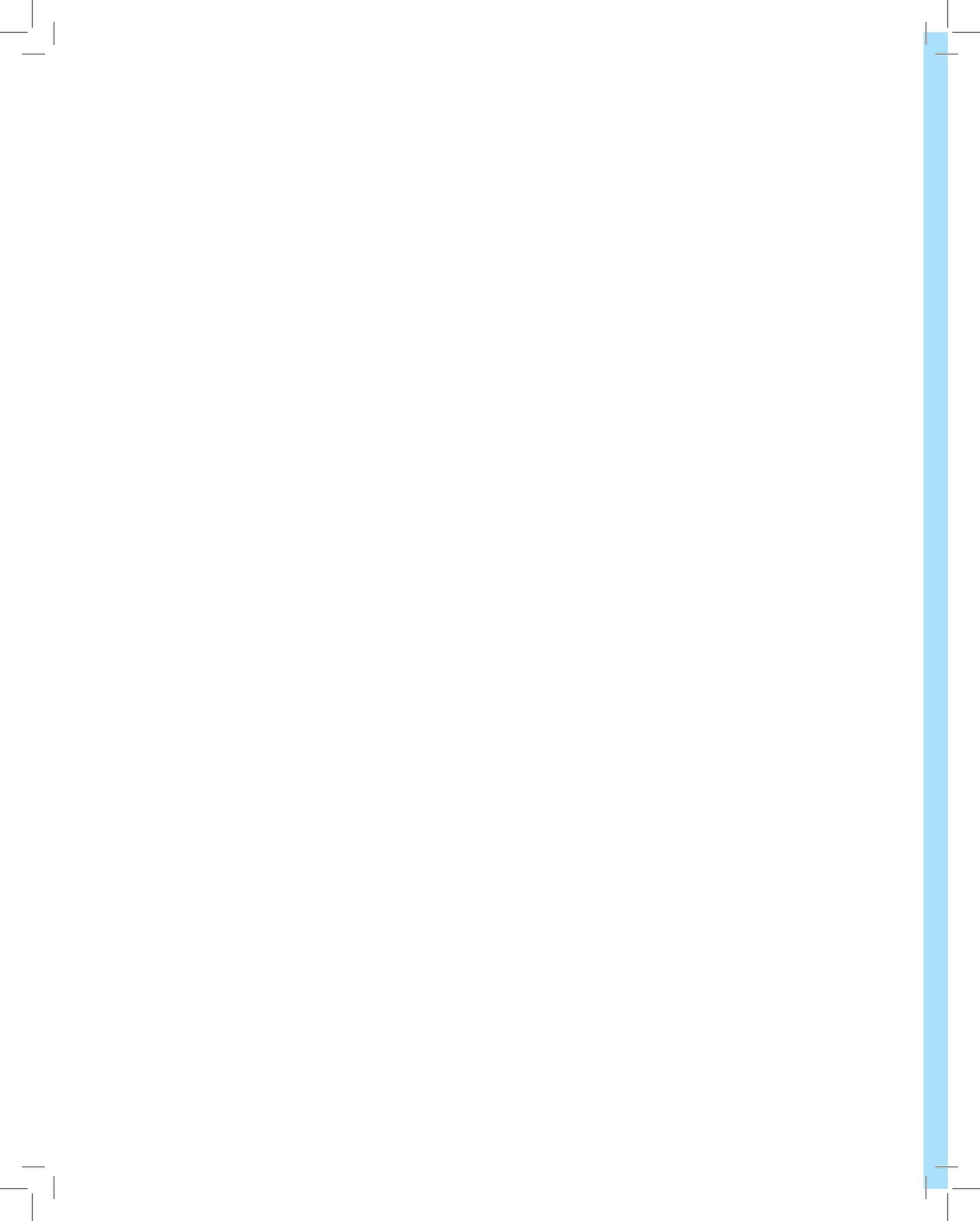
O livro está organizado em 9 capítulos. Os capítulos 1 a 5 tratam dos tópicos referentes à Estatística Descritiva, iniciando com a apresentação da Estatística como campo de conhecimento; explicando os significados de população e amostra; apresentando os diversos tipos de gráficos estatísticos, tabelas e séries. O capítulo 5 é dedicado a distribuição de frequência de dados qualitativos e quantitativos.

Os capítulos 6 e 7 encerram a parte dedicada a Estatística Descritiva, apresentando medidas de posição e medidas de dispersão ou de variabilidade.

Os capítulos 8 e 9 são dedicados ao estudo de probabilidade, abordando a definição conceitual, espaços amostrais finitos e infinitos, probabilidade condicional e distribuições estatísticas.

Ao final de cada capítulo, são apresentadas atividades de avaliação que podem ser utilizadas como verificação de aprendizagem do aluno.

**Os autores**



**Capítulo**

**1**

# **Introdução à Estatística**



## Objetivos

- Apresentar os conceitos básicos de Estatística, o seu campo de aplicação e as fases do método estatístico descritivo.
- Apresentar os conceitos de variáveis, população e amostra além dos motivos para se fazer uso de amostragem e quais os seus tipos principais.

## 1. A Estatística como campo de conhecimento

Em nosso dia a dia, frequentemente estamos fazendo observações de fenômenos e gerando dados. Os professores analisam dados de alunos; analistas de sistemas analisam dados de desempenho de sistemas computacionais; médicos analisam resposta do paciente a tratamentos, e todos nós, ao lermos jornais e revistas, estamos vendo resultados estatísticos provenientes do censo demográfico, de pesquisas eleitorais, da bolsa de valores etc.

Os dados podem provir de estudos observacionais ou de experimentos planejados. Ao acompanharmos o desempenho de um processo produtivo em sua forma natural, estamos fazendo um estudo observacional; ao alterar de forma proposital alguma variável do processo para verificar seus resultados, estamos realizando um experimento.

Estatística é um campo do estudo centrado na produção de metodologia para coleta, organização, descrição, análise e interpretação de dados bem como na obtenção de conclusões válidas e na tomada de decisões razoáveis baseadas em tais análises.

- Decidir qual é o melhor plano experimental e amostral para a realização da Pesquisa.
- Organizar e sumarizar dados obtidos por classificação, por contagem ou por mensuração.
- Fazer inferência sobre populações de unidades (indivíduos, objetos, animais) quando apenas uma parte (amostra) é estudada (classificada, contada ou medida).

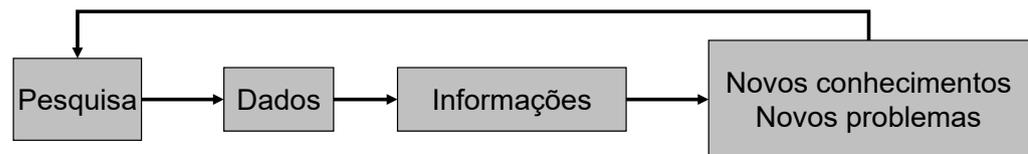
A Estatística pode ser aplicada em praticamente todas as áreas do conhecimento humano, tais como administração, Economia, Farmácia, Educação, Agricultura, Informática, Psicologia, indústria, comércio, Medicina e várias outras.

Podemos dividir a estatística em dois grupos: descritiva e indutiva.

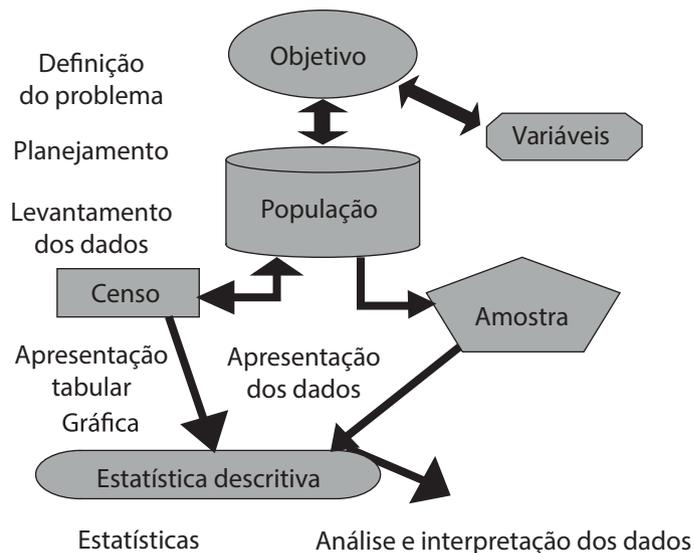
A coleta, a organização e a descrição dos dados estão a cargo da **Estatística Descritiva**, enquanto a análise e a interpretação desses dados ficam a cargo da **Estatística Indutiva** ou **Inferencial**.

A estatística envolve técnicas para coletar, organizar, descrever, analisar e interpretar dados provenientes de estudos experimentais e de estudos observacionais.

A análise estatística de dados geralmente tem por objetivo a tomada de decisões, resolução de problemas ou produção de conhecimentos. Novos conhecimentos em geral nos levam a novos problemas, resultando em um processo iterativo.



Visão geral do processo estatístico



## 2. Porque estudar Estatística

O interesse pelo estudo da estatística se justifica porque

- a natureza apresenta variabilidade.
- ocorrem variações de indivíduo para indivíduo.
- ocorrem variações no mesmo indivíduo.

- a Estatística estuda como controlar, minimizar e observar a variabilidade inevitável de todas as medidas e observações.
- sem métodos estatísticos, a validade científica ficaria comprometida.

## 3. Método estatístico

### 3.1 O método científico

Podemos entender método científico como um conjunto de regras básicas para desenvolver uma experiência a fim de produzir um novo conhecimento, bem como corrigir e integrar conhecimentos pré-existentes.

Dos métodos científicos, podemos destacar o método experimental e o estatístico.

### 3.2 O método experimental

O método<sup>1</sup> experimental baseia-se em manter constantes todas as causas (fatores), menos uma, e variar essa causa de modo que o pesquisador possa descobrir seus efeitos, caso existam. É muito utilizado no estudo da Física, da Química etc.

<sup>1</sup> Método: é um conjunto de meios dispostos convenientemente para se chegar a um fim que se deseja.

### 3.3 O método estatístico

É utilizado quando precisamos descobrir fatos em um campo em que o método experimental não se aplica (como, nas Ciências Sociais), já que os vários fatores que afetam o fenômeno em estudo não podem permanecer constantes enquanto fazemos variar a causa que, naquele momento, nos interessa. A determinação das causas que definem o preço de uma mercadoria seria um bom exemplo: para aplicarmos o método experimental, teríamos que variar a quantidade da mercadoria para saber se tal fato iria influenciar ou não no seu preço.

Nesses casos, lançamos mão de outro método, o método estatístico. O método estatístico, diante da impossibilidade de manter as causas constantes, admite todas essas causas presentes variando-as, registrando essas variações e procurando determinar, no resultado final, que influências cabem a cada uma delas.

<sup>2</sup> Dado: é uma descrição limitada do real, desvinculada de um referencial explicativo e difícil de ser utilizado como informação por ser ininteligível

<sup>3</sup> Informação: é uma descrição mais completa do real associada a um referencial explicativo sistemático. Assim a informação é o dado, cuja forma e conteúdo são apropriados para um uso específico

### 3.4 Fases do método estatístico descritivo

#### a) Coleta de dados

A coleta de dados<sup>2</sup> é o meio pelo qual a informação<sup>3</sup> sobre as variáveis é coletada. A coleta de dados pode ser direta, quando ela é feita diretamente na fonte, e indireta, quando é feita através de outras fontes.

Em relação ao fator tempo, a coleta de dados pode ser classificada em contínua, periódica ou ocasional.

- **Coleta de dados contínua:** quando os eventos que acontecem durante determinado estudo são registrados à medida que ocorrem.
- **Coleta de dados periódica:** acontece em intervalos constantes de tempo, como nos censos.
- **Coleta de dados ocasional:** são aqueles realizados sem a preocupação de continuidade ou periodicidade, com o objetivo de atender a uma conjuntura ou emergência.

#### b) Crítica dos dados

A crítica dos dados é um processo de detecção de erros por inspeção cuidadosa dos dados coletados. Em geral, é tida com a responsável por retardar a conclusão dos resultados da pesquisa. No entanto, trata-se de uma etapa fundamental para garantir a qualidade dos dados.

#### c) Apuração dos dados

Consiste em resumir os dados através de uma contagem e de um agrupamento. É um trabalho de coordenação e de tabulação. Pode ser manual, eletromecânica ou eletrônica.

#### d) Exposição ou apresentação dos dados

A apresentação dos dados é de fundamental importância para uma pesquisa. Estes devem ser apresentados de forma adequada, por meio de tabelas e/ou gráficos que permitam sintetizar grandes quantidades de dados, tornando mais fácil a compreensão do atributo em estudo e permitindo uma futura análise.

#### e) Análise dos resultados

O objetivo de uma análise estatística é tirar conclusões que ajudem o pesquisador a resolver o problema proposto. O significado exato de cada um dos valores obtidos através do cálculo das várias medidas estatísticas disponíveis deve ser bem interpretado.

Capítulo

2

# População e Amostra



## Objetivo

- Definir o que são variáveis, população e amostra.

## 1. Variáveis

Chamamos de variável o conjunto de resultados possíveis de um fenômeno. Os símbolos utilizados para representar as variáveis são as letras maiúsculas do alfabeto, como X, Y, Z, ... que podem assumir qualquer valor de um conjunto de dados. Podemos citar como exemplo: idade, sexo, estado civil etc. A escolha da variável dependerá dos objetivos do estudo estatístico.

As variáveis podem ser classificadas dos seguintes modos.

**a) Qualitativas (ou atributos):** são características de uma população que não podem ser medidas, não têm ordenamento nem hierarquia. Essas variáveis e podem ser:

- **Nominais:** quando os valores são expressos por atributos. Ex: sexo, cor da pele, curso de graduação, nacionalidade etc.
- **Ordinais ou por postos:** quando a variável segue uma ordem, mesmo não podendo ser medida. Ex: escolaridade, cargos em uma empresa, patente militar, etc.

**b) Quantitativa:** quando os valores da variável forem expressos em números, podendo ser:

- **Discreta:** assume apenas valores pertencentes a um conjunto enumerável e resultam de uma contagem. Ex: número de filhos, quantidade de cursos etc.
- **Contínua:** pode assumir qualquer valor num intervalo razoável de variação. Ex: peso, altura, faixa etária etc.

## 2. Escala de medidas

Escala é um conjunto de símbolos ou números<sup>4</sup>, construído com base numa regra e aplica-se a indivíduos ou aos seus comportamentos ou atitudes. A posição de um indivíduo na escala é baseada na posse pelo indivíduo do atributo que a escala deve medir. As principais escalas de medida são: nominal, ordinal, intervalar e razão.

<sup>4</sup> Observe que algumas vezes são atribuídos números aos dados para serem inseridas no computador: 0 - sim; 1 - não, 2 - indeciso. Neste caso são apenas rótulos e não podem ser efetuados cálculos com estes números.

<sup>5</sup> Observe que na escala ordinal não é possível quantificar o quanto o nível 3 é melhor do que 2 ou o 4 é melhor do que 3.

- **Escala nominal**

Dá nome a uma categoria ou a uma classe. Os dados não podem ser dispostos em um esquema ordenado. Exemplos: respostas do tipo “sim”, “não” ou “indeciso”

- **Escala ordinal<sup>5</sup>**

Dá nome e uma ordem a uma categoria ou a uma classe. A diferença entre os valores dos dados não pode ser determinada ou não faz para a pesquisa sentido. Exemplos: grau de instrução: 1 = sem instrução; 2 = primeiro grau; 3 = segundo grau, 4 = superior; 5 = Mestre; 6 = Doutor.

- **Escala intervalar**

É verdadeiramente quantitativa. A mensuração é feita diretamente em números reais, obtidos mediante a comparação com um determinado valor fixo, denominado unidade. O nome "intervalar" está ligado aos intervalos entre as categorias da variável e aqui se sabe exatamente o quanto uma categoria é menor ou maior que outra ou, ainda, se há igualdade entre elas. As operações aritméticas comuns (soma, subtração, multiplicação e divisão) são aplicáveis. Ex: os valores de idade, altura, peso, pressão arterial, frequência cardíaca, exames laboratoriais, medidas diversas etc.

- **Escala proporcional ou nível de razão**

Tem todas as características das escalas apresentadas anteriormente e ainda fornece um zero absoluto ou uma origem significativa. Por haver um acordo universal acerca das localizações do ponto zero, as comparações entre magnitudes de valores na escala de razão são aceitáveis. Uma escala de razões reflete a quantidade real de uma variável. Todas as operações aritméticas são possíveis. Ex: peso. Peso Zero = ausência de peso. 60 kg é o dobro de 30 kg.

### 3. População

É um conjunto de elementos com pelo menos uma característica em comum, que deve delimitar inequivocamente quais os elementos pertencem à população e quais não pertencem. Exemplos: os alunos de uma universidade, os clientes de um banco.

- Como definir uma população?
- A quem interessa esse resultado?

- Se o analista dos resultados for o responsável pelos cursos de educação a distância de uma universidade, será que interessa a ele o desempenho dos alunos dos cursos presenciais?
- Devemos procurar as características que interessam ao analista dos resultados.
- Os alunos de uma universidade em 2010.
- Os alunos dos cursos a distância da universidade em 2010.

Perceba que a cada item, estamos especificando cada vez mais as características das pessoas a serem observadas, restringindo a população objeto de nossos estudos.

A população pode ser

- **Finita:** quando o número de unidades a observar pode ser contado e é limitado. Ex: alunos matriculados nas escolas públicas, pessoas que possuem aparelho telefone celular, número de alunos que se matricularam na disciplina “Estatística” na universidade em 2010 etc.
- **Infinita:** quando a quantidade de observação é ilimitada ou quando as unidades da população não podem ser contadas. Ex: conjunto de medidas de determinado comprimento, gases, líquidos, em que as unidades não podem ser identificadas ou contadas.

<sup>6</sup> Parâmetro: é uma medida numérica que descreve uma característica da população

<sup>7</sup> Estimativa: é uma medida numérica que descreve uma característica da amostra

## 4. Censo

É uma coleção de dados relativos a todos os elementos de uma população.

## 5. Amostra

Definida as características da população, o passo seguinte é o levantamento de dados acerca das características do objeto em estudo. Mas será que sempre é possível o levantamento de dados de toda a população que devemos analisar?

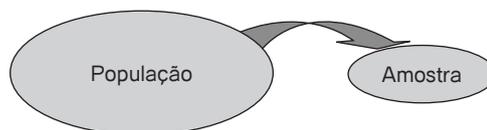
A maioria das vezes não é conveniente e em algumas é impossível devido aos seguintes fatores.

- **Tempo:** as informações devem ser obtidas com rapidez
- **Precisão:** as informações devem ser corretas
- **Custo:** no processo de coleta, sistematização, análise e interpretação, o custo deve ser o menor possível.

<sup>8</sup> Amostra: é um subconjunto de uma população, necessariamente finito, pois todos os seus elementos serão examinados para efeito da realização do estudo estatístico desejado.

<sup>9</sup> Tabelas de números aleatórios: você pode gerar uma tabela de números aleatórios fazendo uso do Excel ou do Br.Office Calc por intermédio da função ALEATÓRIO()

Por impossibilidade ou inviabilidade econômica ou temporal, devemos, então, delimitar nossas observações a uma parte da população, isto é, a uma amostra<sup>8</sup> proveniente dessa população.



Amostragem é uma técnica especial usada para recolher amostras que garante o acaso na escolha de modo a garantir à amostra o caráter de representatividade.

Vejam os três dos principais tipos de amostragem.

- **Amostragem casual simples:** composta de elementos retirados ao acaso da população, ou seja, consiste em selecionar a amostra através de um sorteio. Dessa maneira, todos os elementos da população terão igual probabilidade de serem escolhidos. Para realizar esse sorteio, podemos utilizar urnas, tabelas de números aleatórios<sup>9</sup> ou algum *software* que gere números aleatórios.
- **Amostragem sistemática:** É utilizada quando a população está naturalmente ordenada, como listas telefônicas, fichas de cadastramento etc.
- **Amostragem estratificada:** composta por elementos provenientes da divisão da população em subgrupos denominados estratos (por exemplo, por sexo, renda, bairro etc.)

### Atividades de avaliação



1. Faça uma pesquisa sobre a evolução da Estatística.
2. Cite algumas áreas em que a Estatística é aplicada.
3. Como podemos classificar a Estatística?
4. Qual a diferença existente entre Estatística Descritiva e Estatística Indutiva?
5. Quais as fases do método estatístico?
6. Com o objetivo de fazer um estudo sobre o número de irmãos dos alunos de uma escola, foi feita uma pesquisa em que responderam 60 alunos.

Identifique

a) a população em estudo.

b) a amostra escolhida.

c) a variável em estudo classificando-a.

7. Foi feito um estudo em uma universidade e recolheram-se dados referentes às variáveis idade, sexo, curso, ano de ingresso.

a) Das variáveis indicadas, quais são as quantitativas e quais são as qualitativas?

b) Das variáveis quantitativas, quais são contínuas?

## Referências



COSTA, F. S. **Introdução Ilustrada à Estatística**. São Paulo: Harbra, 1998.

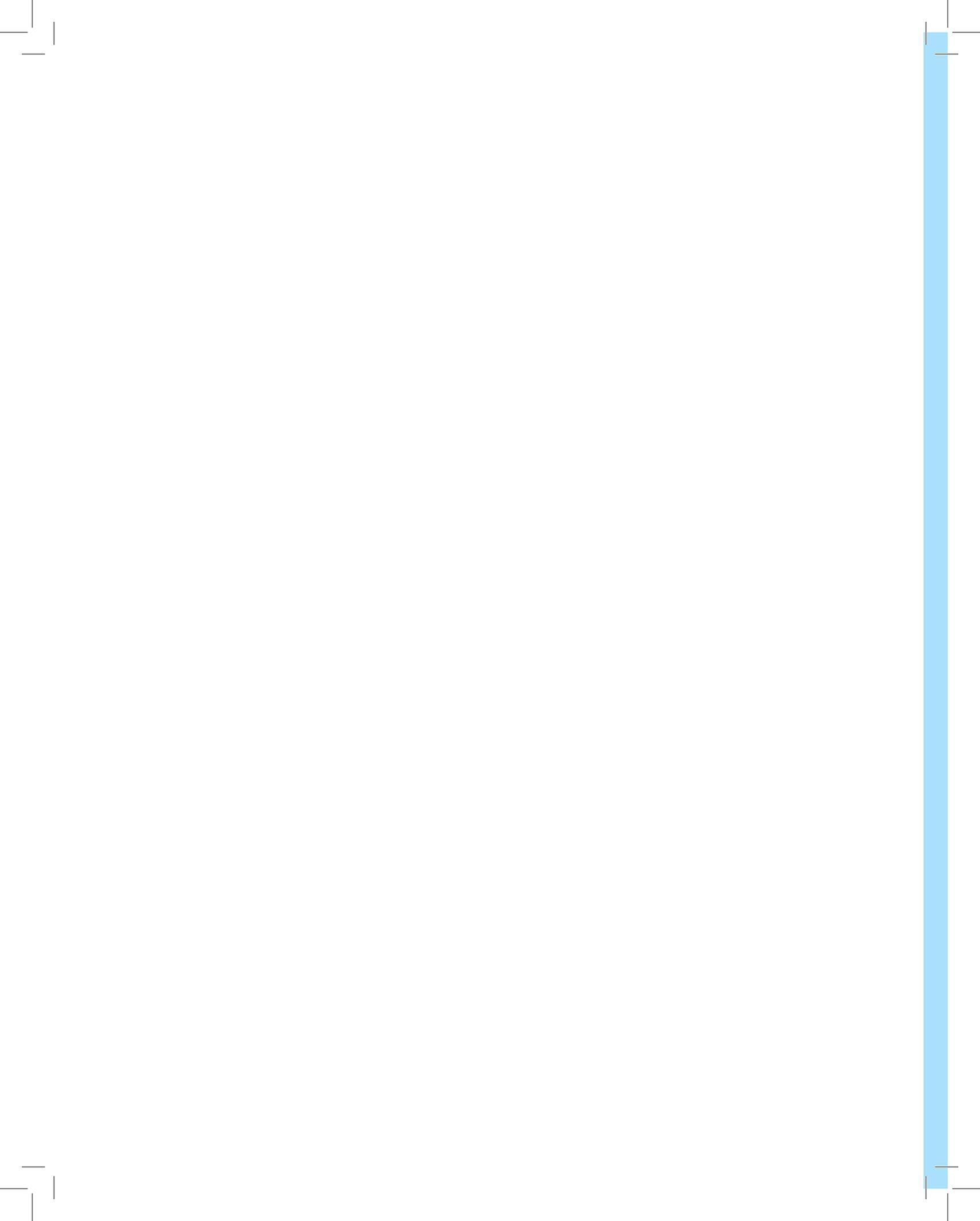
CRESPO, A. **Estatística Fácil**. Rio de Janeiro: Saraiva, 1996.

MARTINS G. A. **Estatística Geral e Aplicada**. São Paulo: Atlas, 2001.

MORETTIN, L. G. **Estatística Básica – Probabilidade**. São Paulo: Makron Books, 1993.

TRIOLA M. F. – **Introdução à Estatística e Probabilidade** – Exercícios Resolvidos e Propostos. Rio de Janeiro: LTC, 1999.

VIEIRA, S. **Estatística Experimental**. São Paulo: Atlas, 1999.



Capítulo

3

# Gráficos Estadísticos



## Objetivos

- Descrever a importância de gráficos e de tabelas para compreensão da informação.
- Apresentar as diversas maneiras de representação dos dados.
- Construir gráficos e tabelas usando as técnicas adequadas e seguindo normas vigentes e boas práticas.

## 1. Apresentação gráfica

Os dados podem ser apresentados em gráficos, com a finalidade de proporcionar ao interessado uma visão rápida do comportamento do fenômeno.

Representa qualquer tabela de maneira simples, legível e interessante, tornando claras as informações que poderiam passar despercebidas em dados apenas tabulados.

É importante que os gráficos sejam simples; as informações contidas devem ser diretas, e detalhes secundários, omitidos, devem ser claros para possibilitar uma correta interpretação e devem expressar a verdade sobre o caso em estudo.

Os principais tipos de gráficos são os diagramas, os cartogramas e os pictogramas.

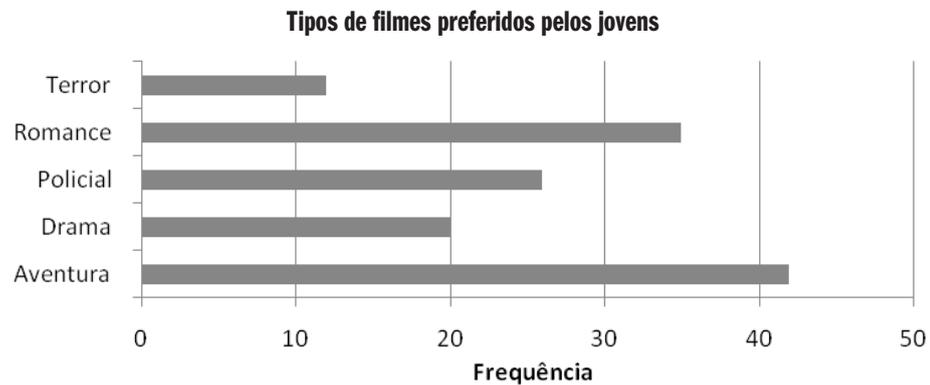
## 2. Diagramas

Os diagramas são gráficos geométricos que possuem no máximo duas dimensões. O sistema cartesiano é utilizado na sua construção.

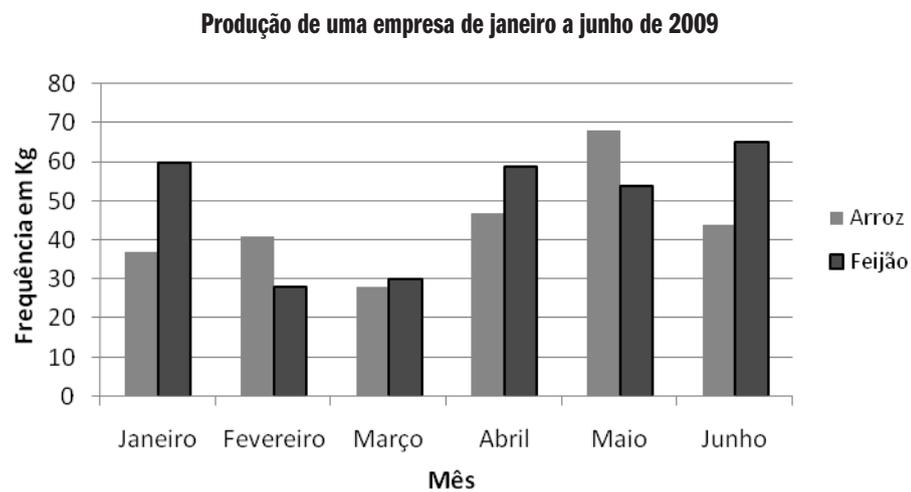
### 2.1 Gráfico em barras ou em colunas

Um gráfico em barras horizontais representa a série de dados através de retângulos dispostos horizontalmente com mesma altura e comprimentos proporcionais à frequência de cada dado. Esse gráfico é muito apropriado para representar graficamente os dados qualitativos, porém pode, ser utilizado também para representar dados quantitativos discretos.

Veja o exemplo a seguir.



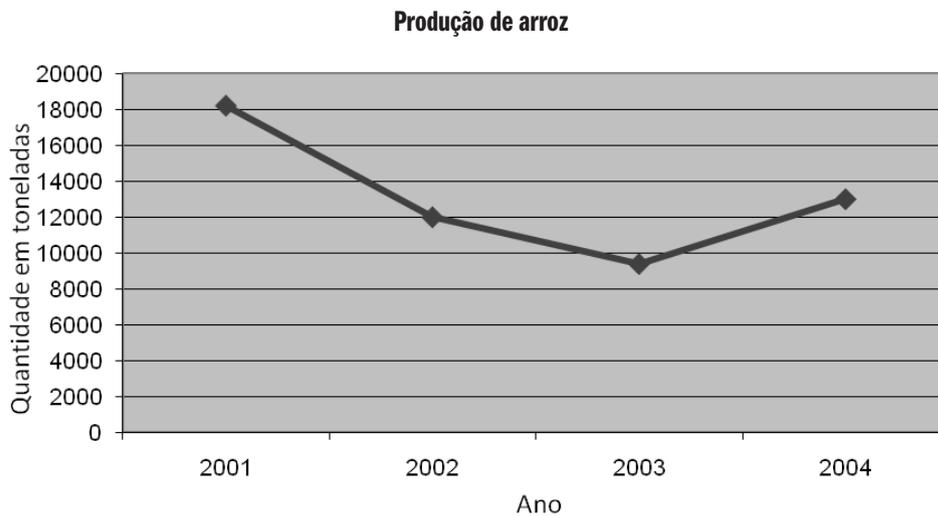
No gráfico em colunas, os retângulos são dispostos verticalmente com a mesma base e alturas proporcionais à frequência de cada dado. Os valores da variável são colocados no eixo horizontal, e a frequência no eixo vertical.



## 2.2 Gráfico em linhas

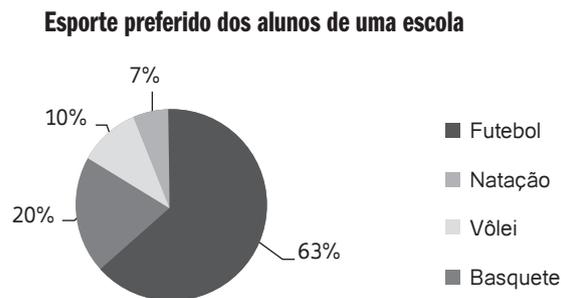
Neste gráfico é usada uma linha para representar a série estatística. Seu principal objetivo é evidenciar a tendência do fenômeno ou a forma como o ele está crescendo ou decrescendo através de um período de tempo. Seu traçado deve ser realizado considerando o eixo "x", horizontal, que representa a escala de tempo e o eixo "y", vertical, que representa a frequência observada dos valores.

Veja o exemplo a seguir



### 2.3 Gráfico em setores

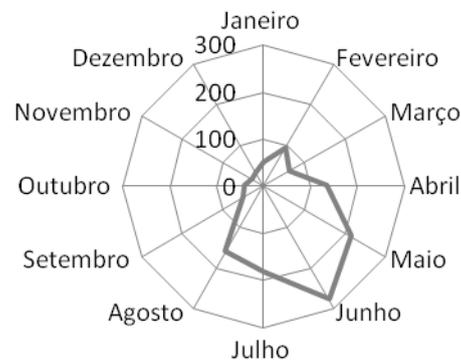
O gráfico de setor ou "pizza" como também é chamado, mostra o tamanho proporcional de itens que constituem uma série de dados para a sua soma. É utilizado principalmente quando se pretende comparar cada valor da série com o total.



### 2.4 Gráfico polar

Neste tipo de gráfico, a série de dados é representada por meio de um polígono. Ideal para representar séries temporais cíclicas, como a variação da precipitação pluviométrica ao longo do ano. Para sua construção, divide-se uma circunferência em tantos arcos quantos forem os dados a representar. Pelos pontos de divisas traçam-se raios. Em cada raio é representado um valor da série, marcando-se um ponto cuja distância ao centro é diretamente proporcional a esse valor e, em seguida, unem-se os pontos. No Excel, o nome da ferramenta a ser utilizada é RADAR e o traçado do gráfico é interno à circunferência.

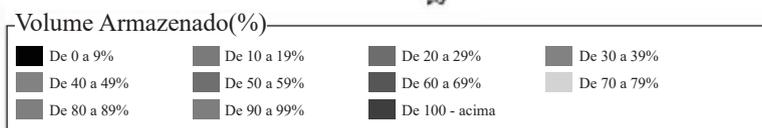
### Precipitação pluviométrica em uma cidade - 2008



Fonte: Dados Hipotéticos

## 2.5 Cartograma

O cartograma é a representação sobre uma carta geográfica (mapa). Este gráfico é empregado quando o objetivo é o de figurar os dados estatísticos diretamente relacionados a áreas geográficas ou políticas.

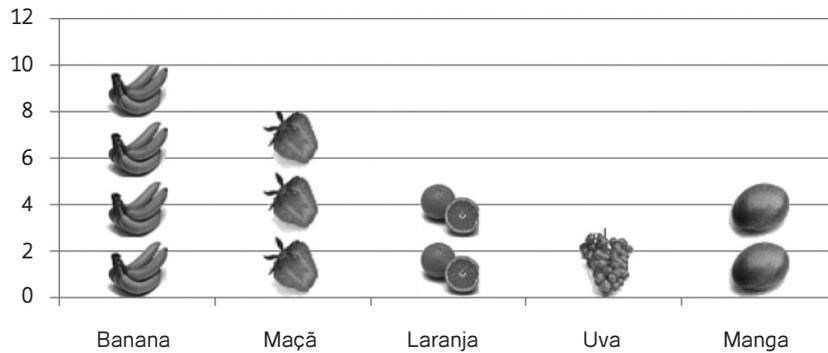


## 2.6 Pictogramas

É a apresentação de uma série estatística por meio de símbolos representativos de um fenômeno. A representação gráfica consta de figuras.

Veja o exemplo a seguir

Frutas preferidas pelas crianças de uma escola



### Saiba Mais



**Importante.** Ao construir um gráfico, evitar

- muitas linhas, para não ficar muito poluído;
- legendas e símbolos dos eixos muito pequenos;
- símbolos em tamanhos diferentes;
- desperdício de espaço;
- muitas marcas de escala.

### Atividades de avaliação



1. Segundo a Revista Y, o número de supermercados (em 100 unidades) em algumas cidades brasileiras em 2005 foi o seguinte: 80 em Belo Horizonte, 47 em Brasília, 69 em Curitiba, 72 em Porto Alegre, 144 no Rio de Janeiro e 399 em São Paulo. Organize os dados numa tabela e represente-os graficamente com um gráfico de colunas.
2. O quadro a seguir é o resultado de uma pesquisa sobre a preferência de atividade física em uma academia. Construa um gráfico de setores para representar o resultado.

ATIVIDADE FÍSICA DA ACADEMIA	
Atividade física	Quantidade de alunos
Musculação	30
Futebol	20
Natação	25
Hidroginástica	10
Ginástica	15

Dados Hipotéticos

3. Usando gráfico em barras, represente as tabelas a seguir

PRODUÇÃO DE ARROZ POR REGIÃO 2009	
Regiões	Quantidade (em toneladas)
Norte	946,2
Nordeste	1145,8
Centro-Oeste	1157,1
Sudeste	195,7
Sudeste	8571,5

Dados Hipotéticos

PRODUÇÃO DE PETRÓLEO BRUTO DO BRASIL 2000 - 2004	
Anos	Quantidade (em 1000m <sup>3</sup> )
2000	36180
2001	36410
2002	37164
2003	38011
2004	38200

Dados Hipotéticos

- Em um Restaurante X, o número de clientes no segundo semestre de 2009 foi o seguinte, (respectivamente, de julho a dezembro): 930, 820, 1080, 1230, 1190, 1740. Represente os dados através do gráfico em linhas.
- Pesquise em jornais e em revistas dois exemplos de pictograma e cartograma.

Capítulo

4

# Tabelas e Séries Estatísticas



## Objetivos

- Explicitar a importância das tabelas e os seus principais elementos.
- Apresentar as principais séries estatísticas.

### 1. Tabelas

Um dos objetivos da estatística é sintetizar os valores que uma ou mais variáveis podem assumir para que possamos ter uma visão global de um variação. Para isso, a Estatística apresenta esses valores em forma de tabelas e gráficos, que irão nos fornecer rápidas e seguras informações a respeito das variáveis em estudo, o que permite decisões administrativas e pedagógicas mais coerentes e científicas.

Os elementos de uma tabela são os seguintes.

- **Título:** deve responder às seguintes questões:
  - O que? (assunto ou fato a ser representado).
  - Onde? (o local onde ocorreu o fenômeno).
  - Quando? (a época ou tempo em que se verificou o fenômeno).
- **Cabeçalho:** parte da tabela na qual é designada a natureza do conteúdo de cada coluna.
- **Corpo:** parte da tabela composta por linhas e colunas.
- **Linhas:** parte do corpo que contém uma sequência horizontal de informações.
- **Colunas:** parte do corpo que contém uma sequência vertical de informações.
- **Coluna Indicadora:** coluna que contém as discriminações correspondentes aos valores distribuídos pelas colunas numéricas
- **Casa ou célula:** parte da tabela formada pelo cruzamento de uma linha com uma coluna.
- **Rodapé:** espaço aproveitado em seguida ao fecho da tabela, em que são colocadas as notas de natureza informativa (fonte, notas e chamadas).
- **Fonte:** refere-se à entidade que organizou ou forneceu os dados expostos.

- **Notas e Chamadas:** são esclarecimentos contidos na tabela (nota - conceituação geral; chamada - esclarecer minúcias em relação a uma célula).

Exemplo:

Tabela 01 – Total do acervo das bibliotecas, impressos e multimídia por área de conhecimento do CNPQ - 2009

Área (CNPq)	Livros	
	Títulos	Volumes
Ciências exatas e da terra	216	1133
Ciências biológicas	87	171
Engenharia/tecnologia	-	-
Ciências da saúde	48	63
Ciências agrárias	05	06
Ciências social aplicadas	2694	1581
Ciências humanas	460	1139
<b>Total</b>	<b>3510</b>	<b>4093</b>

Dados Hipotéticos

A apresentação de quadros e tabelas está regida pelas Normas de Apresentação Tabular (IBGE, 1979) e pelas Normas de Apresentação Tabular (Conselho Nacional de Estatística, 1958).

### 1.1 Quadros

Denomina-se **quadro** a apresentação de dados de forma organizada, para cuja compreensão não seria necessária qualquer elaboração matemático-estatística.

Qualquer que seja seu tipo, sua identificação aparece na parte inferior precedida da palavra “Quadro”, seguida de seu número de ordem, de ocorrência de algarismos arábicos, do respectivo título, da legenda explicativa e da fonte, se necessário.

### 1.2 Tabelas

São conjuntos de dados estatísticos, associados a um fenômeno, dispostos numa determinada ordem de classificação. Expressam as variações qualitativas e quantitativas de um fenômeno.

A finalidade básica da tabela é resumir ou sintetizar dados de maneira a fornecer o máximo de informação num mínimo de espaço.

São ainda características das tabelas

- Toda tabela deve ter significado próprio, dispensando consultas ao texto.
- Não devem ser apresentadas tabelas nas quais a maior parte dos casos indiquem inexistência do fenômeno.
- Caso sejam utilizadas tabelas reproduzidas de outros documentos, a pré-  
via autorização do autor se faz necessária, não sendo mencionada na  
mesma.
- O título é colocado na parte superior precedido da palavra “Tabela” e de  
seu número de ordem em algarismos arábicos, bem como do respectivo  
título, centrados na largura útil das páginas.
- As fontes citadas na construção de tabelas e notas eventuais aparecem  
no rodapé após a linha de fechamento; utilizam-se traços horizontais para  
fechar as linhas externas, separando os cabeçalhos do conteúdo da ta-  
bela e fechando a tabela.
- Evitam-se traços verticais para separar as colunas e traços horizontais  
para separar as linhas.
- A tabela deve ser colocada em posição vertical, para facilitar a leitura  
dos dados.
- No caso em que isso seja impossível, deve ser colocada em posição ho-  
rizontal, com o título voltado para a margem esquerda da folha.
- Se a tabela (ou quadro) não couber em uma página, deve ser continuado  
na página seguinte; nesse caso, o final não será delimitado por traço ho-  
rizontal na parte inferior, e o cabeçalho será repetido na página seguinte.

Tabela 1

<b>TABELA FACULDADES</b>			
<b>Faculdade</b>	<b>Novos alunos</b>	<b>Alunos de graduação</b>	<b>Alteração</b>
Universidade Cedar	110	103	+7
Faculdade Elm	223	214	+9
Academia Maple	197	120	+77
Faculdade Pine	134	121	+13
Instituto Oak	202	210	-8
Universidade xxx	24	20	+4
Faculdade Eee	43	53	-10
Academia Mmm	3	11	-8
Faculdade Ppp	9	4	+5
Instituto Okk	53	52	+1

Dados hipotéticos

## 2. Séries estatísticas

É toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local e da espécie e que determina o surgimento de quatro tipos fundamentais de séries estatísticas:

- **Série temporal ou cronológica:** é a série cujos dados estão dispostos em correspondência com o tempo, ou seja, varia o tempo e permanece constante o fato e o local.

### Exemplo:

Tabela 2

ALUNOS MATRICULADOS NOS CURSOS DE GRADUAÇÃO DA UNIVERSIDADE A	
Ano	Alunos matriculados
2007	17.300
2008	17.500
2009	16.837
2010	18.200

Dados Hipotéticos

- **Série geográfica ou territorial:** é a série cujos dados estão dispostos em correspondência com o local, ou seja, varia o local e permanecem constantes a época e o fato.

Tabela 3

ALUNOS MATRICULADOS NAS UNIDADES DO INTERIOR NOS CURSOS DE GRADUAÇÃO DA UNIVERSIDADE A - 2009	
Cidades	Alunos matriculados
Cidade A	550
Cidade B	1.605
Cidade C	610
Cidade D	689
Cidade E	1.288
Cidade F	340

Fonte: Dados Hipotéticos

- **Série específica ou qualitativa:** é a série cujos dados estão dispostos em correspondência com a espécie ou qualidade, ou seja, varia o fato e permanecem constantes a época e o local.

Veja o exemplo a seguir.

Tabela 4

ALUNOS MATRICULADOS NA UNIVERSIDADE A - 2009	
Alunos	Alunos matriculados
Graduação	16.837
Pós-graduação	1.048

Dados Hipotéticos

- **Série mista ou composta:** a combinação entre duas ou mais séries constituem novas séries denominadas "compostas" e apresentadas em tabelas de dupla entrada. O nome da série mista surge de acordo com a combinação de pelo menos dois elementos.

Tabela 5

EVOLUÇÃO DAS MATRICULAS DOS ALUNOS DA UNIVERSIDADE A - 2007 - 2009			
Alunos	Anos		
	2007	2008	2009
Graduação	17.300	17.500	16.837
Pós-graduação	997	1.010	1.048

Dados Hipotéticos

## Referências



- BARBETA, P.A., REIS, M.M., BORNIA, A.C. **Estatística para Cursos de Engenharia e Informática**, São Paulo: Atlas, 2004.
- BOLFARINE, Heleno & BUSSAB, Wilton O. **Elementos de Amostragem**. 1. ed. São Paulo: Edgard Blücher, 2005.
- BUSSAB, Wilton de Ol., MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva, 2002.
- COSTA, F.S. **Introdução Ilustrada à Estatística**. São Paulo: Harbra, 1998.
- CRESPO, A. **Estatística Fácil**. Rio de Janeiro: Saraiva, 1996.
- FONSECA, Jairo Simon da. **Introdução a Estatística**. Rio de Janeiro: LTC, 1993.
- JAIRO, Simon da Fonseca, MARTINS, Gilberto de Andrade. **Curso de Estatística**, São Paulo: Atlas, 1996.
- MARTINS G.A. **Estatística Geral e Aplicada**. São Paulo: Atlas, 2001.
- MEYER, Paul L. **Aplicações a Estatística**. Rio de Janeiro: LTC, 1992.

MIRSHAWKA, V. **Probabilidade e Estatística para Engenharia**, São Paulo: Nobel, 1978.

MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva. 2008.

PEREIRA, Wilson., TANAKA, Oswaldo K. **Estatística - Conceitos Básicos**, São Paulo: Makron Books, 1990.

ROSS, Sheldon. **A First Course in Probability**. 7. ed. New Jersey: Prentice Hall, 2005.

SPIEGEL M. R. **Estatística**. (Coleção Schaum). São Paulo: Editora Afiliada 1993.

TRIOLA M. F. **Introdução à Estatística e Probabilidade Exercícios Resolvidos e Propostos**. Rio de Janeiro: LTC, 1999.

TRIOLA, Mario F. **Introdução a Estatística**. Rio de Janeiro: LTC. 2005.

VIEIRA, S. **Estatística Experimental**. São Paulo: Atlas, 1999.

Capítulo

5

# Distribuição de frequência



## Objetivos

- Apresentar a forma de sintetizar dados qualitativos quantitativos.
- Fornecer o passo a passo da construção das tabelas de distribuição de frequência simples e para dados agrupados.
- Mostrar representações gráficas: histograma, polígono de frequência e ogiva.

### 1. Sintetizando dados qualitativos

Os dados qualitativos são o resultado da análise de variáveis qualitativas. A análise estatística desse tipo de dados resume-se à contagem do número de indivíduos em cada categoria e ao cálculo das respectivas porcentagens.

Quadro 1

Coca-cola	Sprite	Coca-cola	Pepsi-cola
Coca-cola zero	Coca-cola	Coca-cola	Coca-cola
Pepsi-cola	Sprite	Pepsi-cola	Pepsi-cola
Coca-cola	Pepsi-cola	Coca-cola	Sprite
Coca-cola zero	Coca-cola	Sprite	Coca-cola zero
Pepsi-cola	Pepsi-cola	Sprite	Coca-cola zero

#### 1.1 Distribuição de frequência para dados qualitativos

Uma distribuição de frequência agrupa os dados por classes de ocorrência, resumindo a análise de conjunto de dados grandes. Vejamos um exemplo. Adotemos o conjunto de dados do quadro anterior que representa o consumo de refrigerantes de um grupo de amigos.

Para construirmos a tabela de frequências, organizamos os dados em 3 colunas: coluna das categorias ou classes – onde se indicam todas as categorias da variável em estudo; coluna das frequências absolutas – onde se registra o total de elementos da amostra que pertencem a cada categoria – e coluna das frequências relativas (ou porcentagens) – onde se coloca, para cada categoria, o valor que se obtém dividindo a respectiva frequência absoluta pela dimensão da amostra.

Tabela 1

DISTRIBUIÇÃO DE FREQUÊNCIA DAS COMPRAS DE REFRIGERANTES		
Refrigerantes	Frequência Absoluta	Frequência Relativa
Coca-cola	8	0,3333
Coca-cola zero	4	0,1667
Pepsi-cola	7	0,2917
Sprite	5	0,2083
Total	24	1

Dados Hipotéticos

<sup>10</sup> Em alguns casos, as frequências relativas são dízimas infinitas obrigando, por isso, a arredondamentos. Estes podem ser feitos desde que o total seja igual a 1.

Ao organizarmos os dados de uma amostra numa tabela de frequências, é bom verificarmos se as frequências estão bem calculadas. Para isso é suficiente somá-las para todas as classes e verificar que a soma das frequências absolutas é igual à dimensão da amostra e que a soma das frequências relativas<sup>10</sup> é igual a 1.

## Saiba Mais



### Arredondamento de Dados

Regras: Portaria 36 de 06/07/1965 - INPM - Instituto Nacional de Pesos e Medidas.

- Se o primeiro algarismo após aquele que formos arredondar for de 0 a 4, conservamos o algarismo a ser arredondado e desprezamos os seguintes. Ex.: 7,34856 (para décimos) 7,3
- Se o primeiro algarismo após aquele que formos arredondar for de 6 a 9, acrescenta-se uma unidade no algarismo a ser arredondado e desprezamos os seguintes. Ex.: 1,2734 (para décimos) 1,3
- Se o primeiro algarismo após aquele que formos arredondar for 5, seguido apenas de zeros, conservamos o algarismo se ele for par ou aumentamos uma unidade se ele for ímpar, desprezando os seguintes. Ex.: 6,2500 (para décimos) 6,2 12,350 (para décimos) 12,4
- Se o 5 for seguido de outros algarismos dos quais, pelo menos um é diferente de zero, aumentamos uma unidade no algarismo e desprezamos os seguintes. Ex.: 8,2502 (para décimos) 8,3 8,4503 (para décimos) 8,5

## 1.2 Representação gráfica para dados qualitativos

O gráfico estatístico<sup>11</sup> é uma forma de apresentação/visualização dos dados estatísticos, que permite que a pessoa que está analisando esses dados tenha uma impressão mais rápida e mais viva do fenômeno em estudo, uma vez que os gráficos falam mais à compreensão do que as séries.

A grande vantagem dos gráficos em relação às tabelas de frequências está na rapidez da leitura, em que **não só há uma percepção imediata de qual a categoria de maior frequência**, como também se evidencia com mais clareza a ordem de grandeza de cada categoria relativamente às restantes.

Vejamos um exemplo de gráfico em colunas usando como base a nossa tabela de distribuição de frequência das compras de refrigerantes

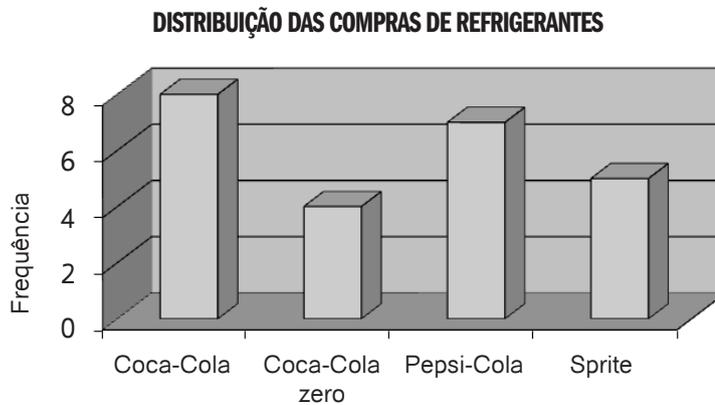


Gráfico 1

<sup>11</sup> Requisitos fundamentais para a construção de um gráfico:

**Simplicidade:** o gráfico deve ser destituído de detalhes de importância secundária, evitando-se, também, traços desnecessários que possam levar a pessoa que está analisando a uma interpretação equivocada dos dados.

**Clareza:** o gráfico deve possibilitar uma correta interpretação dos valores representativos do caso em estudo.

**Veracidade:** o gráfico deve ser a verdadeira expressão do fenômeno em estudo.

## 2. Sintetizando dados quantitativos

O dado quantitativo faz referência aos números no sentido de quantidade e podem se dividir em duas categorias.

- **Quantitativo discreto:** aquele que pode assumir apenas valores pertencentes a um conjunto enumerável.

Ex: população: habitações de uma cidade (variável: número de cômodos); população: casais em uma cidade (variável: número de filhos)

- **Quantitativo contínuo:** pode assumir qualquer valor em certo intervalo de variação.

**Exemplo:** população: pessoas residentes em uma cidade (variável: idade); estudantes: alunos de uma universidade (variável: estatura dos alunos)

## 2.1 Distribuição de frequência

Vamos imaginar uma pesquisa referente às estaturas de quarenta alunos que compõem uma amostra dos alunos de uma universidade, o que resultou na tabela de valores a seguir:

Tabela 2

ESTATURA DOS ALUNOS DE UMA UNIVERSIDADE									
166	160	161	150	162	160	165	167	164	160
162	168	161	163	156	173	160	155	164	168
155	152	163	160	155	155	169	151	170	164
154	161	156	172	153	157	156	158	158	161

Fonte: Dados Hipotéticos

A tabela anterior é denominada de "tabela primitiva" por considerar que os elementos ainda não foram numericamente organizados.

Visualizando a tabela anterior percebemos que fica difícil saber em torno de que valor tende a se concentrar as estaturas dos alunos, qual a menor ou qual a maior estatura ou, ainda, quantos alunos se acham abaixo ou acima de uma determinada estatura. Logo, é importante organizar os dados, e a maneira mais simples de fazer isso é através de uma certa ordenação (crescente ou decrescente). A tabela obtida através da ordenação dos dados recebe o nome de rol.

Tabela 3

ESTATURA DOS ALUNOS DE UMA UNIVERSIDADE									
150	154	155	157	160	161	162	164	166	169
151	155	156	158	160	161	162	164	167	170
152	155	156	158	160	161	163	164	168	172
153	155	156	160	160	161	163	165	168	173

Dados Hipotéticos

Depois de feita esta ordenação, ficou fácil visualizar que a menor estatura é 150 cm, que a amplitude de variação foi de  $173 - 150 = 23$  cm e, ainda, a ordem que um valor particular da variável ocupa no conjunto. Com um exame mais minucioso, vemos que há uma concentração das estaturas em algum valor entre 160 cm e 165 cm e, mais ainda, que há poucos valores abaixo de 155 cm e acima de 170 cm.

Construindo a tabela de frequência absoluta teremos:

Tabela 4

Distribuição de frequência da estatura de 40 alunos de uma universidade	
Estatura (cm)	Frequência
150	1
151	1
152	1
153	1
154	1
155	4
156	3
157	1
158	2
160	5
161	4
162	2
163	2
164	3
165	1
166	1
167	1
168	2
169	1
170	1
172	1
173	1
<b>Total</b>	<b>40</b>

Dados Hipotéticos

Observando a tabela anterior, percebemos que a sua construção não é adequada quando o número de valores da variável ( $n$ ) é de tamanho razoável. Então, é importante fazermos o agrupamento dos valores em vários intervalos que, em Estatística, chamamos de classes.

A tabela a seguir foi construída considerando a frequência de uma classe o número de valores da variável pertencente à classe. Essa tabela é denominada "distribuição de frequência com intervalos de classe".

<sup>12</sup>  $154 \vdash 158$  (é um intervalo fechado à esquerda e aberto à direita, tal que:  $154 \leq x < 158$ ), em vez de dizermos que a estatura de 1 aluno é de 154 cm; de 4 alunos, 155 cm; de 3 alunos, 156 cm; e de 1 aluno, 157 cm, dizemos que 9 alunos têm estaturas entre 154, inclusive, e 158 cm.

Tabela 5

DISTRIBUIÇÃO DE FREQUÊNCIA DA ESTATURA DE 40 ALUNOS DE UMA UNIVERSIDADE	
Classes	Frequência
150 $\vdash$ 154	4
154 $\vdash$ 158 <sup>12</sup>	9
158 $\vdash$ 162	11
162 $\vdash$ 166	8
166 $\vdash$ 170	5
170 $\vdash$ 174	3

Dados Hipotéticos

É importante salientar que, ao agruparmos os valores da variável em classes, ganhamos em simplicidade para perdermos os detalhes. Perceba que, na Tabela 3, é fácil verificar que quatro alunos têm 161 cm de altura e que na Tabela 4, não podemos ver se algum aluno tem a estatura de 159 cm. No entanto, sabemos com segurança que onze alunos têm estatura compreendida entre 158 e 162 cm. Porém a nova tabela realça o que há de essencial nos dados e, também, torna possível o uso de técnicas analíticas para sua total descrição, o que vem ao encontro da finalidade da estatística, que é especificar e analisar o conjunto de valores, desinteressando-se por casos isolados.

<sup>13</sup> Na distribuição de frequência com classe, o  $h_i$  será igual em todas as classes. A amplitude total da distribuição jamais coincide com a amplitude amostral, AT é sempre maior que AA.

## 2.2 Elementos de uma distribuição de frequência<sup>13</sup>

- **Classe:** são os intervalos de variação da variável; é simbolizada por  $i$  e o número total de classes, simbolizada por  $k$ . Ex: na tabela anterior,  $k = 6$  e  $158 \vdash 162$  é a 3ª classe, onde  $i = 3$ .
- **Limites de classe:** são os extremos de cada classe. O menor número é o limite inferior de classe ( $I_i$ ) e o maior número, limite superior de classe ( $L_i$ ). Ex: em  $158 \vdash 162$ ,  $I_3 = 158$  e  $L_3 = 162$ .
- **Amplitude do intervalo de classe:** é obtida através da diferença entre o limite superior e inferior da classe e é simbolizada por  $h_i = L_i - I_i$ . Ex: na tabela anterior,  $h_1 = 162 - 158 = 4$ .
- **Amplitude total da distribuição:** é a diferença entre o limite superior da última classe e o limite inferior da primeira classe.  $AT = L(\max) - I(\min)$ . No nosso exemplo,  $AT = 174 - 150 = 24$ .
- **Amplitude total da amostra (ROL):** é a diferença entre o valor máximo e o valor mínimo da amostra (ROL), onde  $AA = X_{\max} - X_{\min}$ . Em nosso exemplo,  $AA = 173 - 150 = 23$ .

- **Ponto médio de classe:** é o ponto que divide o intervalo da classe em duas partes iguais. No nosso exemplo, em  $158 \vdash 162$ ,  $x_i = (l_i + L_i)/2 \Rightarrow x_2 = (158 + 162)/2 = 160$

### 2.3 Passos para construção das categorias em classes

1. Organize os dados brutos em um ROL.
2. Calcule a amplitude amostral AA (no nosso exemplo,  $AA = 173 - 150 = 23$ )
3. Calcule o número de classes através da “Regra de Sturges<sup>14</sup>” (no nosso exemplo:  $v = 40$  dados, então, a princípio, a regra sugere a adoção de 6 classes)
4. Decidido o número de classes, calcule a amplitude do intervalo de classe dividindo a amplitude total da amostra pelo número de classes  $h \approx AA / k$  (no nosso exemplo  $h \approx 23/6 = 3,8 \approx 4$ )
5. Temos então o menor número da amostra, o número de classes e a amplitude do intervalo; com isso, é possível montar a tabela, com o cuidado para não aparecer em classes com frequência = 0 (zero). No nosso exemplo, o menor número da amostra =  $150 + 4 = 154$ , logo a primeira classe será representada por  $150 \mid - 154$ . As classes seguintes respeitarão o mesmo procedimento.

<sup>14</sup> A Regra de Sturges nos dá o número de classes em função do tamanho da amostra  
 $k \approx 1 + 3,3 \cdot \log n$   
 Onde:  
 $k$  é o número de classe;  
 $n$  é o tamanho da amostra.  
 No cálculo da amplitude do intervalo de classe, quando o resultado não é exato, devemos arredondá-lo para mais.

### 2.4 Tipos de frequências

- **Frequência simples ou absoluta ( $f_i$ ):** são os valores que realmente representam o número de dados de cada classe. A soma das frequências simples é igual ao número total dos dados da distribuição.

$$f_i = n$$

- **Frequência relativa ( $fr_i$ ):** são os valores das razões entre as frequências absolutas de cada classe e a frequência total da distribuição. A soma das frequências relativas é igual a 1 (100 %).

$$fr_i = f_i / f_i$$

- **Frequência acumulada ( $F_i$ ):** é o total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe.

$$F_k = f_1 + f_2 + \dots + f_k \quad \text{ou} \quad F_k = f_i \quad (i = 1, 2, \dots, k)$$

- **Frequência acumulada relativa ( $Fr_i$ ):** é a frequência acumulada da classe dividida pela frequência total da distribuição.

$$Fr_i = F_i / f_i$$

Podemos agora construir a tabela a seguir com as frequências estudadas considerando o nosso exemplo.

Tabela 11

DISTRIBUIÇÃO DE FREQUÊNCIA DA ESTATURA DE 40 ALUNOS DE UMA UNIVERSIDADE						
i	Estaturas (cm)	$f_i$	$x_i$	$fr_i$	$F_i$	$Fr_i$
1	150  – 154	4	152	0,100	4	0,100
2	154  – 158	9	156	0,225	13	0,325
3	158  – 162	11	160	0,275	24	0,600
4	162  – 166	8	164	0,200	32	0,800
5	166  – 170	5	168	0,125	37	0,925
6	170  – 174	3	172	0,075	40	1,000
		$\Sigma = 40$		$\Sigma = 1,000$		

Dados Hipotéticos

De acordo com o estudo dessas frequências, podemos responder às questões a seguir:

1. Quantos alunos têm estatura entre 158 cm, inclusive, e 156 cm?

**Resposta:** 11 – esses são os valores que formam a terceira classe ( $f_3 = 11$ )

2. Qual a porcentagem de alunos cujas estaturas são inferiores a 154 cm?

**Resposta:** 10% – esses valores são os que formam a primeira classe ( $fr_1 = 0,100$ ). Obtemos a resposta multiplicando a frequência relativa por 100:  $0,100 \times 100 = 10$

3. Quantos alunos têm estatura abaixo de 162?

**Resposta:** 24 – as estaturas abaixo de 162 são aquelas que formam as classes de ordem 1, 2 e 3. Assim, o número de alunos é dado por  $F_3 = \sum_{(i=1 \rightarrow 3)} f_i = f_1 + f_2 + f_3 \Rightarrow F_3 = 24$

4. Quantos alunos têm estatura maior ou igual a 158 cm?

**Resposta:** 27 – as estaturas maior ou igual a 158 são aquelas que formam as classes de ordem 3, 4, 5 e 6. Assim, o número de alunos é dado por  $\sum_{(i=3 \rightarrow 6)} f_i = f_3 + f_4 + f_5 + f_6 = 11 + 8 + 5 + 3 = 27$

## 2.5 Distribuição de frequências e sua representação gráfica para variáveis quantitativas contínuas

As variáveis quantitativas contínuas diferem um pouco das discretas na sua forma de representação gráfica. A diferença mais importante é que as frequências são associadas a intervalos de valores (classes de frequências) e não a valores individuais da variável em estudo. Graficamente, podemos representar pelo histograma, pelo polígono de frequência e pelo polígono de frequência acumulada.

## a) Histograma

O histograma é um gráfico formado por um conjunto de colunas retangulares. No eixo das abscissas, marcamos as classes, cujas amplitudes correspondem às bases dos retângulos. No eixo das ordenadas, marcamos as frequências absolutas, que correspondem às alturas dos retângulos. Os pontos médios das bases dos retângulos coincidem com os pontos médios dos intervalos de classes.

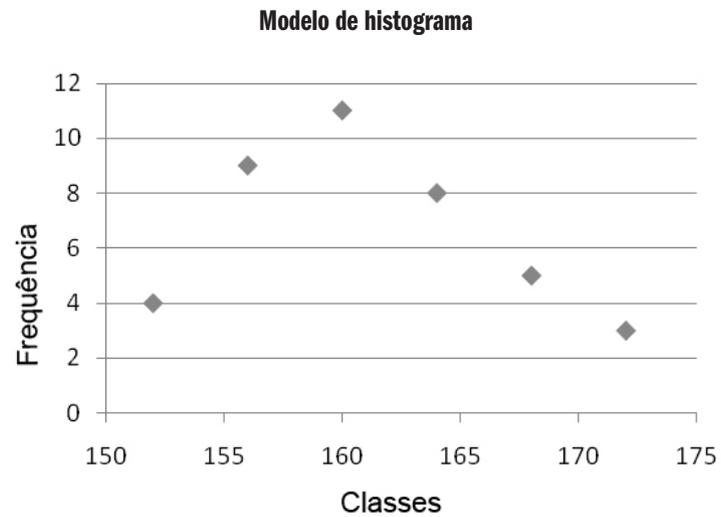
Roteiro para construção do histograma:

1. Obtenha a tabela de frequência a partir dos dados, agrupando-os em classes.
2. Desenhe dois eixos ortogonais de tamanho médio.
3. Divida o eixo horizontal em tantas partes quanto for o número de classes mais dois (considere uma classe à esquerda da primeira classe e outra à direita da última classe para deixar espaço suficiente para traçar o polígono de frequência, que veremos mais adiante) e marque os números correspondentes aos limites inferior e superior de cada classe.
4. Identifique a maior frequência da classe na tabela de frequência; escolha um número adequado, maior ou igual àquela frequência; marque esse número na extremidade do eixo vertical; divida o eixo vertical em algumas partes e marque os números correspondentes.
5. Para cada classe, desenhe um retângulo com largura igual à amplitude da classe com altura igual à frequência da classe.

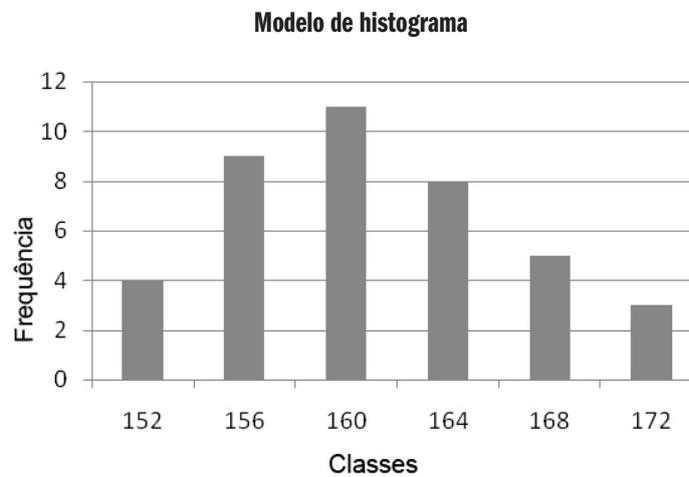
Para construir o histograma com o software Excel, depois de feita a tabela de distribuição de frequência em classes, adicione uma nova coluna com os pontos médios dos intervalos de classe, que serão as abscissas de um gráfico de dispersão, com as frequências correspondendo às alturas.

Classes	Frequência	Ponto médio
150   154	4	152
154   158	9	156
158   162	11	160
162   166	8	164
166   170	5	168
170   174	3	172

Veja o exemplo a seguir.

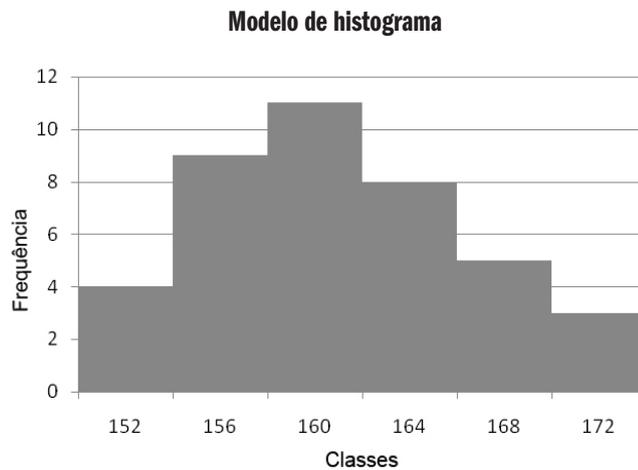


Com esse gráfico construído, clique com o botão direito do *mouse* sobre ele, selecione a opção ALTERAR TIPO DE GRÁFICO escolhendo colunas.



Sobre uma das colunas, com o botão direito do *mouse*, escolha a opção FORMATAR SEQUÊNCIA DE DADOS, em seguida ESPAÇAMENTO e zere-o.

Veja mais um modelo de histograma a seguir.

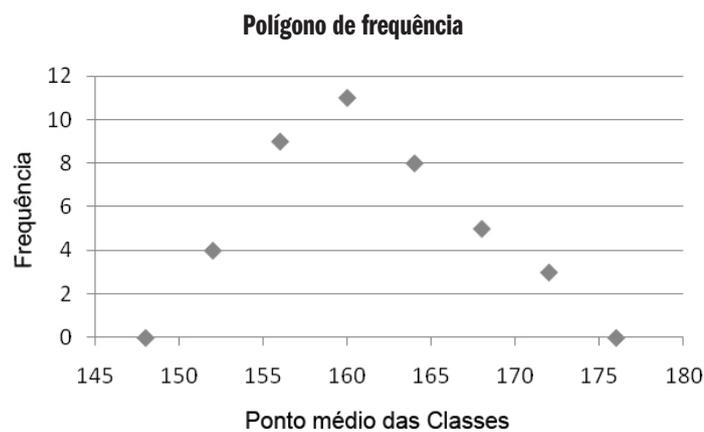


### b) Polígono de frequência

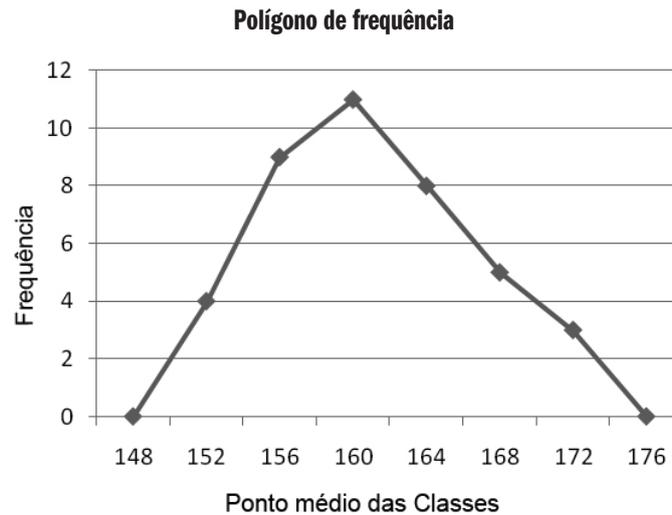
É a representação gráfica de uma distribuição por meio de um polígono. É obtido através da ligação dos pontos médios dos intervalos de classe em um histograma de frequências.

Para construir o polígono de frequência com o Excel, inclua uma classe anterior à primeira e uma posterior à segunda, ambas com frequência zero.

Classes	Frequência	Ponto médio
146   150	0	148
150   154	4	152
154   158	9	156
158   162	11	160
162   166	8	164
166   170	5	168
170   174	3	172
174   178	0	176



Uma vez construído esse gráfico, clique com o botão direito do *mouse* sobre o gráfico selecionando a opção TIPO DE GRÁFICO escolhendo GRÁFICOS EM LINHA.



### c) Ogiva

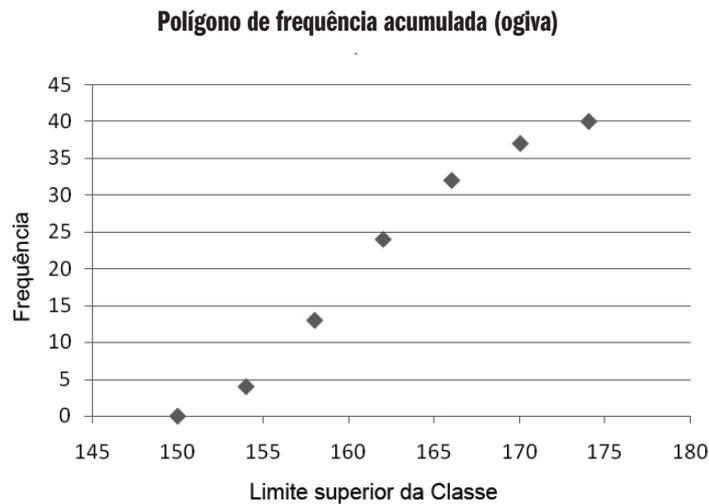
É um polígono de frequências acumuladas no qual elas são localizadas sobre perpendiculares levantadas nos limites inferiores ou superiores das classes, dependendo se a ogiva representar as frequências acumuladas abaixo ou acima, respectivamente.

Para construir o polígono de frequência acumulada (ogiva) com o Excel, inclua uma classe anterior à primeira com frequência nula. O gráfico será construído com os limites superiores de cada classe e as respectivas frequências acumuladas.

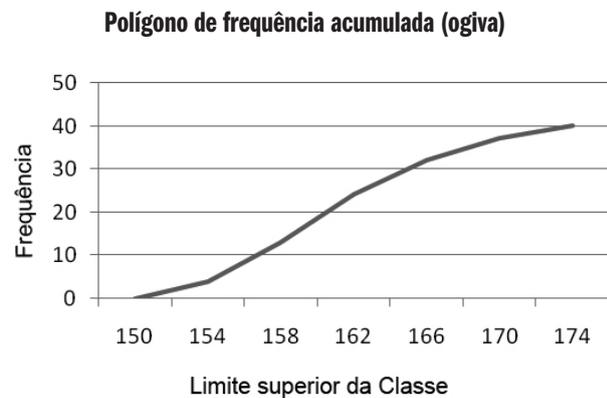
Classes	Frequência	Frequência acumulada
146   150	0	0
150   154	4	4
154   158	9	13
158   162	11	24
162   166	8	32
166   170	5	37
170   174	3	40

Com o Excel aberto, opção ASSISTENTE DE GRÁFICO, escolha tipo de gráfico (Dispersão), complete com o título do Gráfico, e com os valores dos eixos X e Y.

Veja o exemplo a seguir.



Uma vez construído esse gráfico, clique com o botão direito do *mouse* sobre o gráfico selecionando a opção tipo de gráfico escolhendo GRÁFICOS EM LINHA.



## 2.6 Distribuição de frequência sem intervalo de classes

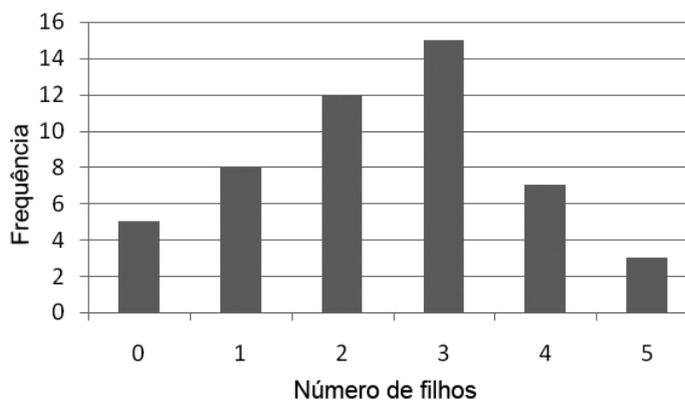
É a simples condensação dos dados conforme as repetições de seus valores. Cada valor é tomado como um intervalo de classe tratando-se de variável discreta de variação relativamente pequena.

Veja o exemplo a seguir.

Considere  $x$  a variável “número de filhos de 50 famílias”

$i$	$x_i$	$f_i$
1	0	5
2	1	8
3	2	12
4	3	15
5	4	7
6	5	3
		$\Sigma = 50$

A representação gráfica de uma distribuição de frequência de uma variável quantitativa discreta é denominada **gráfico de frequências**. Utilizando a tabela de frequência anterior obtemos o gráfico a seguir.



Outra representação utilizada é a do gráfico das frequências acumuladas e das frequências relativas acumuladas. Tomando-se os dados do nosso exemplo, podemos calcular as frequências, frequências acumuladas e frequências relativas acumuladas dos diversos valores. Esse cálculo está ilustrado na tabela a seguir.

$i$	$x_i$	$f_i$	$fr_i$	$F_i$	$Fr_i$
1	0	5	0,10	5	0,10
2	1	8	0,16	13	0,26
3	2	12	0,24	25	0,50
4	3	15	0,30	40	0,80
5	4	7	0,14	47	0,94
6	5	3	0,06	50	1,00
		$\Sigma = 50$	$\Sigma = 1,00$		

Veja a representação nos gráficos a seguir.

Gráfico de frequências relativas

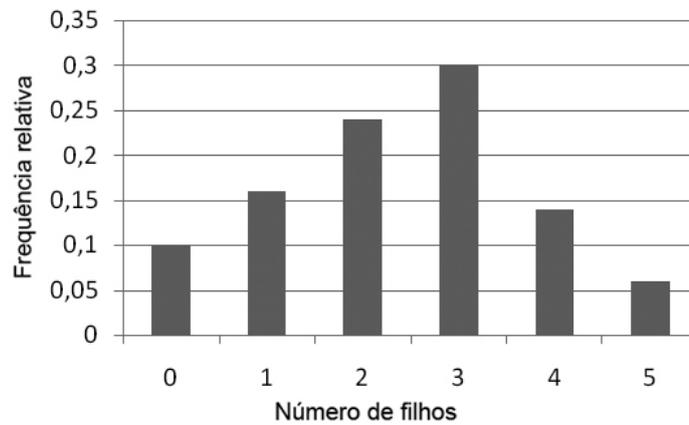
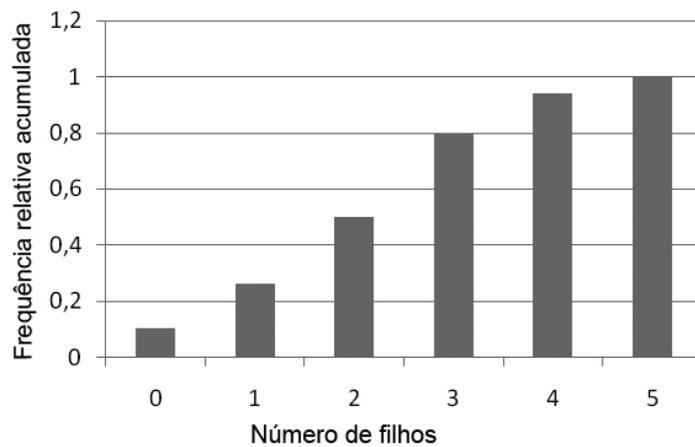


Gráfico de frequências relativas acumuladas



### Atividades de avaliação



1. Conhecidas as notas dos alunos em uma disciplina, de acordo com os dados a seguir, obtenha a distribuição de frequência, considerando 30 para limite inferior da primeira classe e, 10 para intervalo de classe.

84 68 33 52 47 73 68 61 73 77

74 71 81 91 65 55 57 35 85 88

59 80 41 50 53 65 76 85 73 60

67 41 78 56 94 35 45 55 64 74

65 94 66 48 39 69 89 86 42 54

2. Complete a tabela a seguir.

i	Classes	$f_i$	$x_i$	$fr_i$	$F_i$	$Fr_i$
1	0  – 8	4				
2	08  – 16	10				
3	16  – 24	14				
4	24  – 32	9				
5	32  – 40	3				
		$\Sigma = 40$				

3. Considere a seguinte tabela

i	Classes	fi
1	2,75  – 2,80	2
2	2,80  – 2,85	3
3	2,85  – 2,90	10
4	2,90  – 2,95	11
5	2,95  – 3,00	24
6	3,00  – 3,05	14
7	3,05  – 3,10	9
8	3,10  – 3,15	8
9	3,15  – 3,20	6
10	3,20  – 3,25	3
	<b>Total</b>	<b>90</b>

Identifique os seguintes elementos da tabela.

- Frequência simples absoluta da quinta classe.
- Frequência total.
- Limite inferior da sexta classe.
- Limite superior da quarta classe.
- Amplitude do intervalo de classe.
- Amplitude total.
- Ponto médio da terceira classe.

4. Construa o histograma da distribuição a seguir.

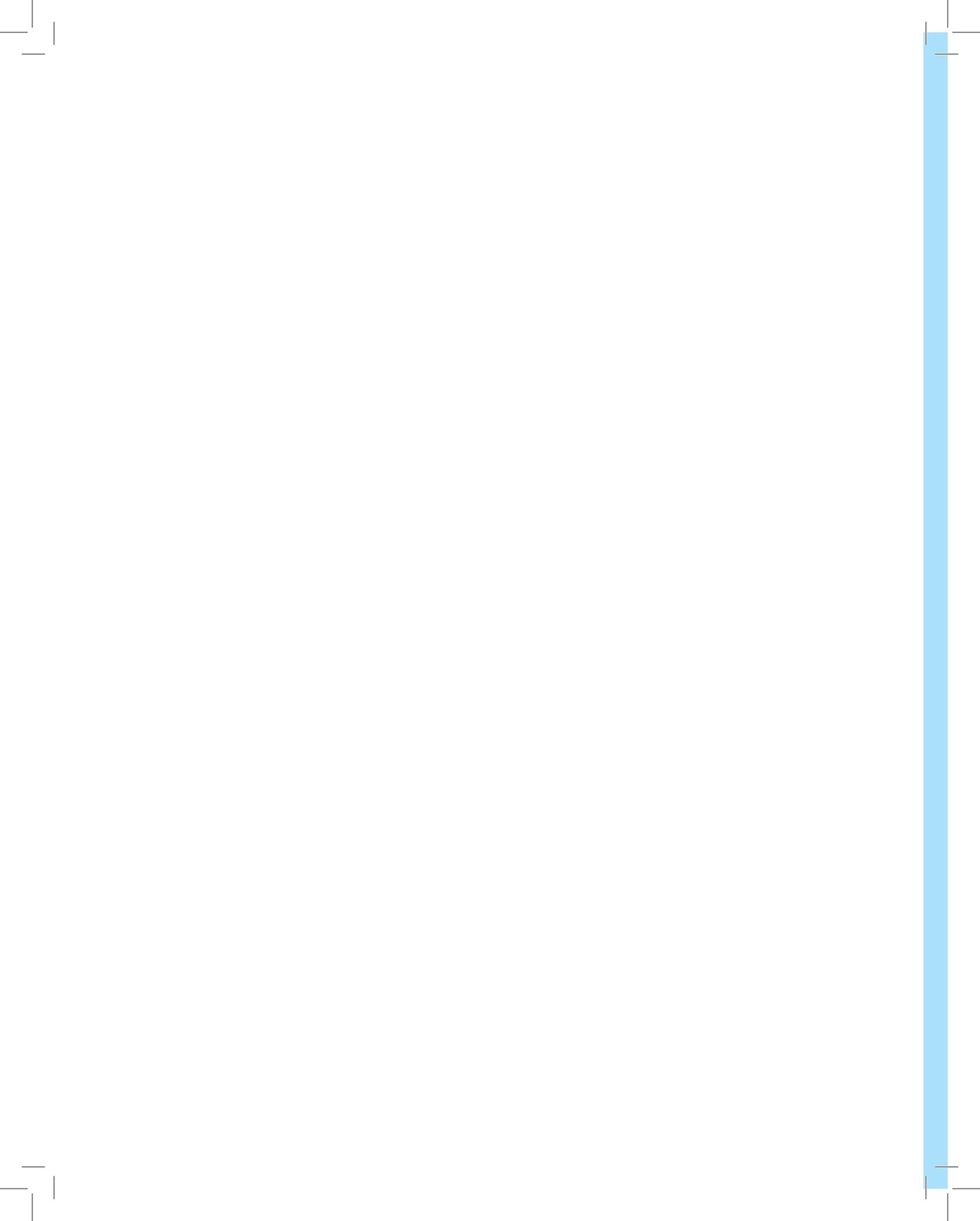
i	Áreas (em m <sup>2</sup> )	Número de lotes
1	300  – 400	14
2	400  – 500	46
3	500  – 600	58
4	600  – 700	76

i	Áreas (em m <sup>2</sup> )	Número de lotes
5	700  – 800	68
6	800  – 900	62
7	900  – 1000	48
	Total	<b>372</b>

## Referências



- BARBETA, P. A., REIS, M. M., BORNIA, A. C. **Estatística para Cursos de Engenharia e Informática**, São Paulo: Atlas, 2004.
- BOLFARINE, Heleno & BUSSAB, Wilton O. **Elementos de Amostragem**. 1. ed. São Paulo: Edgard Blücher, 2005.
- BUSSAB, Wilton de OI., MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva, 2002.
- COSTA, F. S. **Introdução Ilustrada à Estatística**. São Paulo: Harbra, 1998.
- CRESPO, A. **Estatística Fácil**. Rio de Janeiro: Saraiva, 1996.
- FONSECA, Jairo Simon da. **Introdução à Estatística**. Rio de Janeiro: LTC. 1993.
- JAIRO, Simon da Fonseca, MARTINS, Gilberto de Andrade. **Curso de Estatística**, São Paulo: Atlas, 1996.
- MARTINS G. A. **Estatística Geral e Aplicada**. São Paulo: Atlas, 2001.
- MEYER, Paul L. **Aplicações a Estatística**. Rio de Janeiro: LTC. 1996.
- MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva. 2013.
- PEREIRA, Wilson., TANAKA, Oswaldo K. **Estatística – Conceitos Básicos**, São Paulo: Makron Books, 1990.
- SPIEGEL M. R. **Estatística. Coleção Schaum**. São Paulo: Editora Afiliada 1993.
- TRIOLA. M. F. **Introdução à Estatística e Probabilidade - Exercícios Resolvidos e Propostos**. Rio de Janeiro: LTC, 1999.
- TRIOLA, Mario F. **Introdução a Estatística**. Rio de Janeiro: LTC. 2005.
- VIEIRA, S. **Estatística Experimental**. São Paulo: Atlas, 1999.



Capítulo

6

# Medidas de Posição



## Objetivos

- Apresentar as medidas de posição: média aritmética, mediana e moda, tanto para dados isolados como para dados agrupados e mostrar que elas servem para resumir informações sobre um conjunto de dados.
- Apresentar as medidas de dispersão: amplitude total, desvio médio absoluto, variância, desvio padrão e coeficiente de variação como medidas de assimetria e medidas de curtoses.

## 1. Introdução

A sintetização dos dados sob a forma de tabelas, gráficos e distribuições de frequências nos permite localizar a maior concentração de valores de uma dada distribuição. Agora, vamos ressaltar as tendências características de cada distribuição.

Estudaremos as medidas de posição, também denominadas de medidas de **tendência central**<sup>15</sup>, que são constituídas por médias, mediana e moda. Além dessas, existem as separatrizes, as quais são apenas medidas de posição e não medidas de tendência central.

<sup>15</sup> As medidas de tendência central representam os fenômenos pelos seus valores médios, em torno dos quais tendem a se concentrar em os dados.

## 2. Média aritmética ( $\bar{x}$ )

É a mais simples das médias e de fácil cálculo. A sua grande desvantagem é ser fortemente influenciada pelos valores extremos.

### 2.1 Dados não agrupados

Quando desejamos conhecer a média dos dados que não estão agrupados, determinamos a **média aritmética simples**. Para isso basta somar todos os valores e dividir o total pelo número deles.

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

Onde:  $\bar{x}$  é a média aritmética.

$x_i$  são os valores da variável.

$n$  é o número de valores.

**Exemplo:** Seja o conjunto de dados: 2, 3, 9, 5, 8, 10 e 19.

Então,

$$\bar{x} = \frac{2 + 3 + 9 + 5 + 8 + 10 + 19}{7} = \frac{56}{7} = 8$$

Pode ocorrer que a média seja um número diferente de todos os da série de dados que ela representa. Nesse caso, costumamos dizer que a média não tem existência concreta.

**Exemplo:** para os valores 2, 4, 6 e 8, a média aritmética tem valor 5.

## 2.2 Desvio em relação à média

O desvio em relação à média é definido, para cada uma das medidas, pela diferença entre a medida e a média. Ou seja:

$d_i = x_i - \bar{x}$  O índice  $i$  representa a posição da medida na tabela. Assim,  $d_1$  significa o desvio da primeira medida ( $x_1$ ) e  $d_2$  significa o desvio da segunda medida ( $x_2$ ) etc.

Em relação ao exemplo anterior, teremos

$$d_1 = x_1 - \bar{x} = 2 - 5 = -3$$

$$d_2 = x_2 - \bar{x} = 4 - 5 = -1$$

$$d_3 = x_3 - \bar{x} = 6 - 5 = 1$$

$$d_4 = x_4 - \bar{x} = 8 - 5 = 3$$

## 2.3 Propriedades da Média Aritmética

### a) Primeira Propriedade

- A soma algébrica dos desvios em relação à média é nula.

$$\sum_{i=1}^k d_i = 0$$

Então, para o nosso exemplo teremos

$$\sum_{i=1}^4 d_i = (-3) + (-1) + 1 + 3 = 0$$

### b) Segunda Propriedade

- Somando-se (ou subtraindo-se) uma constante (c) a todos os valores de uma variável, a média do conjunto fica aumentada (ou diminuída) dessa constante.

Se, no nosso exemplo, somarmos a constante 2 a cada um dos valores da variável teremos

$$y_1 = 4, y_2 = 6, y_3 = 8, y_4 = 10$$

Calculando a média de y,

$$\sum_{i=1}^4 d_i = (-3) + (-1) + 1 + 3 = 0$$

$$\text{Perceba que } \bar{y} = 7 = 5 + 2 \text{ logo } \bar{y} = \bar{x} + 2$$

### c) Terceira Propriedade

- Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante (c), a média do conjunto fica multiplicada (ou dividida) por essa constante.

Multiplicando por 3 cada um dos valores do nosso exemplo, teremos

$$y_1 = 6, y_2 = 8, y_3 = 18, y_4 = 24$$

Calculando a média de

$$\bar{y} = \frac{6 + 12 + 18 + 24}{4} = \frac{60}{4} = 15$$

$$\text{Perceba que } \bar{y} = 15 = 5 * 3 \text{ logo } \bar{y} = \bar{x} * 3$$

## 2.4 Dados agrupados

### a) Sem intervalo de classe

Como as frequências são números indicadores da intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva a calcular a **média aritmética ponderada**, pela seguinte fórmula

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \quad \text{Somatória da multiplicação de cada valor pela respectiva frequência e divisão pelo total de valores}$$

Vejamos o exemplo a seguir.

Considerando a distribuição relativa a um grupo de alunos de uma escola e tomando para variável o número de reprovações em uma disciplina, considere os dados a seguir.

Tabela 1

Número de reprovações em uma disciplina	$f_i$	$x_i f_i$
1	1	1
2	3	6
3	5	15
4	1	4
<b>Total</b>	<b><math>\Sigma = 10</math></b>	<b><math>\Sigma = 26</math></b>

Dados Hipotéticos

Temos, então

$$\sum x_i f_i = 26 \text{ e } \sum f_i = 10$$

Logo,

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{26}{10} = 2,6$$

Isto é

$$\bar{x} = 2,6 \text{ reprovações}$$

### b) Com intervalo de classe

Convencionamos que todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio determinamos, assim, a média aritmética ponderada pela seguinte fórmula.

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \quad \text{Onde } x_i \text{ é o ponto médio da classe}$$

Considerando a seguinte distribuição.

Tabela 2

Pesos	$x_i$	$f_i$	$x_i f_i$
40   43	41,5	3	124,5
43   46	44,5	4	178
46   49	47,5	10	475
49   52	50,5	13	656,5
52   55	53,5	10	535
55   58	56,5	6	339
58   61	59,5	4	238
<b>Total</b>		<b><math>\Sigma = 50</math></b>	<b><math>\Sigma = 2546</math></b>

Dados Hipotéticos

Temos, então

$$\sum x_i f_i = 2546 \text{ e } \sum f_i = 50$$

Logo,

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{2546}{50} = 50,92$$

Isto é  $\bar{x} = 50,92$

### Saiba Mais



#### Generalidades sobre a média aritmética

- É facilmente calculável.
- É rigorosamente definida e exata.
- Descreve todos os dados de uma série.
- É a medida de posição mais utilizada.
- Depende de cada valor da série.
- É influenciada por valores extremos.
- Não é utilizada para dados qualitativos.
- Pode não pertencer ao conjunto.

## 3. Mediana (Md)

A mediana de um conjunto de valores, dispostos segundo uma ordem (crescente ou decrescente), é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos.

### 3.1 Dados não agrupados

O cálculo da mediana envolve um passo prévio de ordenação da amostra. Em seguida, é tomado o valor central que apresenta o mesmo número de elementos à direita e à esquerda.

**Exemplo:** Dada uma série de valores: 5, 2, 6, 13, 9, 15, 10

De acordo com a definição de mediana, o primeiro passo a ser dado é o da ordenação (crescente ou decrescente) dos valores. Então teremos

2, 5, 6, 9, 10, 13, 15

O valor que divide a série anterior em duas partes iguais é igual a 9, logo a  $Md = 9$ .

Entretanto, se série de dados analisada tiver um número par de termos, a mediana será, por definição, qualquer dos números compreendidos entre os dois valores centrais da série. Convencionou-se utilizar o ponto médio.

No exemplo anterior, se acrescentarmos o valor 4 na série

2, 4, 5, 6, 9, 10, 13, 15,

teremos para mediana a média aritmética entre 6 e 9.

Portanto,

$$Md = \frac{6+9}{2} = \frac{15}{2} = 7,5$$

### 3.2 Método prático para o cálculo da mediana

- Se a série dada tiver número ímpar de termos, O valor mediano será o termo de ordem dado pela fórmula

$$\frac{(n+1)}{2} \quad n \text{ é o número de elementos da série}$$

**Exemplo:** Calcule a mediana da série { 1, 3, 0, 0, 2, 4, 1, 2, 5 }

Primeiro precisamos ordenar a série { 0, 0, 1, 1, 2, 2, 3, 4, 5 }

$n = 9$  logo,

$$\frac{(9+1)}{2} = 5, \text{ então, o } 5^{\text{o}} \text{ elemento da série ordenada será a mediana}$$

Portanto,  $Md = 2$

- Se a série dada tiver número par de termos, o valor mediano será a média aritmética dos termos de ordem

$$\frac{n}{2} \text{ e } \frac{n}{2} + 1$$

Na série 2, 4, 5, 6, 9, 10, 13, 15, teremos

$n = 8$ , então  $\frac{8}{2} = 4$  e  $\frac{8}{2} + 1 = 5$ . Logo, a mediana é a média aritmética

do 4<sup>o</sup> e 5<sup>o</sup> termos da série, isto é

$$Md = \frac{6+9}{2} = \frac{15}{2} = 7,5$$

### 3.3 Dados agrupados

#### a) Sem intervalo de classes

Para conhecer a mediana de uma série de valores agrupados sem intervalo de classe, seguiremos alguns passos. Consideremos a tabela a seguir.

Tabela 3

Número de reprovações em uma disciplina	$f_i$	$F_i$
1	1	1
2	3	4
3	4	8
4	1	9
<b>Total</b>	<b><math>\Sigma = 9</math></b>	

Dados Hipotéticos

O primeiro passo será descobrir o número de elementos<sup>16</sup> do conjunto, ou seja, somar a coluna da frequência ( $f_i$ ); identificaremos se  $n$  é par ou ímpar.

Caso o  $n$  seja ímpar, o conjunto terá apenas uma Posição Central, caso contrário teremos duas Posições Centrais.

Na tabela anterior, temos  $n=9$ , logo temos uma posição central que pode ser determinada pela fórmula a seguir.

$$\text{Posição central} = \frac{n + 1}{2}$$

Então, calculamos

$$\text{Posição central} = \frac{9 + 1}{2} = 5$$

Assim a nossa posição central = 5ª posição

Agora que sabemos a nossa posição central, o nosso próximo passo será comparar o valor encontrado da posição central com os valores da coluna da frequência acumulada ( $F_i$ ), até que esse valor seja maior ou igual ao valor da posição central.

No nosso exemplo, temos a posição central igual a 5, então, iniciando a comparação com os valores da nossa frequência acumulada.

- 1 é maior ou igual a 5? Não
- 4 é maior ou igual a 5? Não
- 8 é maior ou igual a 5? Sim

<sup>16</sup> Quando o número de elementos da série estatística for ímpar, haverá coincidência da mediana com um dos elementos da série. Quando o número de elementos da série estatística for par, nunca haverá coincidência da mediana com um dos elementos da série. A mediana será sempre a média aritmética dos dois elementos centrais da série.

Paramos então a nossa comparação, verificando a nossa tabela comprovamos que o elemento que corresponde a essa  $F_i$  (8) é o 3. Logo,

$$Md = 3$$

Agora vamos fazer um exemplo para a situação em que  $n$  seja par, considerando a tabela a seguir:

Tabela 4

$x_i$	$f_i$	$F_i$
1	1	1
2	3	4
3	5	9
4	1	10
<b>Total</b>	<b><math>\Sigma = 10</math></b>	

Dados Hipotéticos

Na tabela anterior, temos  $n = 10$ . Logo, teremos duas posições centrais que podem ser determinadas pela fórmula a seguir:

$$1^{\text{a}} \text{ posição central} = \frac{n}{2}$$

A  $2^{\text{a}}$  posição central sucede a  $1^{\text{a}}$  posição

Portanto,

- $1^{\text{a}}$  posição central = 5 e
- $2^{\text{a}}$  posição central = 6

Semelhante ao exemplo anterior, iremos comparar o valor encontrado da posição central com os valores da coluna da frequência acumulada ( $f_{ac}$ ), fazendo a comparação para a  $1^{\text{a}}$  posição central.

- 1 é maior ou igual a 5? Não
- 4 é maior ou igual a 5? Não
- 9 é maior ou igual a 5? Sim

Então, paramos e constatamos que o elemento correspondente a essa posição é o 3. Esse valor ficará guardado para o final da situação.

Passamos agora a trabalhar com a segunda posição central do conjunto, que é a  $6^{\text{a}}$  posição. Faremos novamente as perguntas, agora usando este valor como referência. Daí, teremos

- 1 é maior ou igual a 6? Não
- 4 é maior ou igual a 6? Não
- 9 é maior ou igual a 6? Sim

Neste momento, então, paramos e verificamos quem é o  $X_i$  correspondente, que é exatamente o  $X_i=3$ .

Descobertos os dois elementos que ocupam as posições centrais, teremos que calcular a sua média, para chegarmos à **mediana**<sup>17</sup> do conjunto. Vejamos

$$\text{Então: Md} = \frac{(3 + 3)}{2} = 3$$

### b) Com intervalo de classes

Para calcular a mediana onde os dados estão agrupados em uma distribuição de frequência com intervalo de classes, precisamos inicialmente identificar a classe na qual se acha a mediana, a chamada **classe mediana**, que corresponde à frequência acumulada imediatamente superior ao resultado de  $\frac{\sum f_i}{2}$ .

Considerando a distribuição de frequência da tabela 5 acrescida das frequências acumuladas, analise os dados a seguir.

Tabela 5

Pesos	$f_i$	$F_i$
40   43	3	3
43   46	4	7
46   49	10	17
49   52	13	30
52   55	10	40
55   58	6	46
58   61	4	50
<b>Total</b>	<b><math>\Sigma = 50</math></b>	

Dados Hipotéticos

Teremos

$$\frac{\sum f_i}{2} = \frac{50}{2} = 25$$

Deveremos fazer a comparação dos valores da **frequência acumulada**<sup>18</sup> com o valor encontrado, da mesma maneira que fizemos para o item anterior. Então teremos

- 3 é maior ou igual a 25? Não
- 7 é maior ou igual a 25? Não
- 17 é maior ou igual a 25? Não
- 30 é maior ou igual a 25? Sim

<sup>17</sup> Em uma série, a mediana, a média e a moda não têm, necessariamente, o mesmo valor.

A mediana depende da posição e não dos valores dos elementos na série ordenada. Essa é uma das diferenças marcantes entre mediana e média (que se deixa influenciar, muito pelos valores extremos). Vejamos. Em 5, 7, 10, 13, 15, a média = 10, e a mediana = 10.

Em 5, 7, 10, 13, 65 a média = 20 e a mediana = 10 portanto a média do segundo conjunto de valores é maior do que a do primeiro, por influência dos valores extremos, ao passo que a mediana permanece a mesma.

<sup>18</sup> Frequência acumulada ( $F_i = \sum f_i$ )

Paramos, então, a nossa comparação e verificamos que a quarta classe (49 | 52) será a nossa **Classe Mediana**.

Conhecendo a Classe Mediana da Distribuição de Frequências, aplicaremos a fórmula da mediana, a seguir:

$$Md = \text{linf} + \left[ \frac{\left( \frac{\sum f_i}{2} \right) - F(\text{ant})}{f} \right] \cdot h \quad \text{Classe Mediana}$$

Onde

- **linf** é o limite inferior da classe mediana;
- **f** é a frequência simples da classe mediana
- **F(ant)** é a frequência acumulada classe anterior à classe mediana;
- **h** é a amplitude<sup>19</sup> do intervalo da classe mediana

Então, para o nosso exemplo teremos

$$\frac{\sum f_i}{2} = \frac{50}{2} = 25$$

- $\text{linf} = 49$ ,
- $F(\text{ant}) = 17$
- $f = 13$
- $h = 3$

$$Md = 49 + \left[ \frac{\left( \frac{50}{2} \right) - 17}{13} \right] \cdot 3 = 50,84$$

## 4. Moda (Mo)

Define-se moda (ou modas) de um conjunto de valores como o valor (ou valores) de máxima frequência.

### 4.1 Dados não agrupados

Para os dados não agrupados, simplesmente se observa o elemento (ou elementos) de maior frequência.

A moda<sup>20</sup> em um conjunto de valores, diferentemente das outras medidas de tendência central, pode nem existir, bem como pode haver uma, ou duas, ou várias modas no mesmo conjunto.

<sup>19</sup> Lembrando que a amplitude do intervalo de uma classe é obtida através da diferença entre o limite superior e o inferior da classe.

<sup>20</sup> A moda é o valor na sequência que mais se repete, e não o número de vezes que ele aparece.

Vejamos os exemplos considerando os conjuntos de valores a seguir:

- {1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 5, 5}

Veja que o valor que mais se repete é o 2, então vamos ter um conjunto **unimodal**.

logo a  $Mo = 2$

- {1, 2, 3, 5, 6, 8, 10}

Observando o conjunto de valores anterior, percebemos que não há nenhum elemento que se repete; todos aparecem uma única vez, então, nesse caso, dizemos que se trata de um conjunto **amodal**.

No conjunto de valores anterior, percebemos que dois elementos – o 3 e o 7 – se repetem três vezes. Logo, nesse caso, vamos ter um conjunto chamado **bimodal**, ou seja, com dois valores modais. Logo,  $Mo = 3$ , e  $Mo = 7$

- {1, 2, 2, 2, 2, 3, 4, 5, 5, 5, 5, 7, 7, 8, 8, 8, 8, 9, 13, 15}

Nesse caso, vamos ter três valores com mesma frequência, 2, 5 e 8. Assim teremos um conjunto chamado **multimodal**.

## 4.2. Dados agrupados

### a) Sem intervalo de classe

Quando os dados estão agrupados, para determinarmos a moda, só teremos que observar o valor da variável que tem a maior frequência. Vejamos na tabela a seguir.

Tabela 6

$x_i$	$f_i$
1	1
2	3
3	5
4	1
<b>Total</b>	<b><math>\Sigma = 10</math></b>

Dados Hipotéticos

Verificamos que a maior frequência é  $f_i = 5$ , que corresponde ao elemento  $X_i = 3$ . Logo,

$Mo = 3$

### b) Com intervalo de classe

No caso em que os dados estão agrupados com intervalo de classe, a moda é o valor dominante da classe que apresenta a maior frequência que é denominada **classe modal**. A maneira mais simples para calcular a moda é tomar o ponto médio da classe modal. A esse valor denominamos de **moda bruta**.

Vamos determinar a moda para a tabela de distribuição a seguir.

Tabela 7

i	Pesos	f <sub>i</sub>
1	40   43	3
2	43   46	4
3	46   49	10
4	49   52	13
5	52   55	10
6	55   58	6
7	58   61	4
		Σ = 50

Maior f<sub>i</sub> = 13  
Classe Modal

Dados Hipotéticos

Identificamos na tabela anterior que a quarta classe é a que tem a maior frequência, logo.

$$Mo = \frac{49 + 52}{2} = 50,5$$

Existem outras formas de se calcular a moda de uma distribuição de frequências, uma delas é a utilização do **Método de Czuber**, que leva em consideração a **frequência**<sup>21</sup> anterior e posterior à classe modal e faz uso da fórmula a seguir para o seu cálculo.

$$Mo = l_i + \frac{d_1}{(d_1 + d_2)} \times h$$

Onde

- $l_i \Rightarrow$  limite inferior da classe modal
- $d_1 \Rightarrow$  diferença entre a  $f_i$  da classe modal e a  $f_i$  da classe anterior
- $d_2 \Rightarrow$  diferença entre a  $f_i$  da classe modal e a  $f_i$  da classe posterior (aquela que vem logo após a classe modal).
- $h \Rightarrow$  amplitude da classe modal.

<sup>21</sup> Classe anterior é a que precede a classe modal, e classe posterior é a que sucede a classe modal.

Como exemplo para aplicação da fórmula, vamos utilizar a tabela 7.

Já sabemos que a nossa classe modal é a quarta classe cuja  $f_i = 13$ .

Logo teremos

$$l_i = 49, d_1 = 3, d_2 = 3, h = 3$$

$$Mo = 49 + \frac{3}{(3 + 3)} \times 3 = 50,5$$

Outra maneira de calcular a moda é pelo método de King, que utiliza a fórmula a seguir:

$$Mo = l_i + \frac{d_1}{(d_1 + d_2)} \times h$$

Onde

- $l_{inf} \Rightarrow$  limite inferior da classe modal.
- $F_{post} \Rightarrow$   $f_i$  da classe posterior à classe modal;
- $f_{ant} \Rightarrow$   $f_i$  da classe anterior à classe modal;
- $h \Rightarrow$  amplitude da classe modal

Para o exemplo anterior, teremos

$$Mo = 49 + \frac{10}{10 + 10} \times 3 = 50,5$$

## Saiba Mais



### Generalidades sobre a moda

- É de fácil compreensão.
- Pode não existir em uma série ou ocorrer mais de uma vez em outras.
- Não é rigorosamente definida e exata.
- Seu cálculo pode depender de alguns valores da série.
- Não é influenciada por todos os valores de uma série.
- É muito utilizada quando há valores extremos.

## 5. Aplicação das medidas de posição

Foram apresentadas três medidas estatísticas conhecidas como Medidas de Tendência Central ou Medidas de Posição. Elas têm a finalidade de sintetizar as informações de um conjunto de dados resumindo-as em um único valor. Uma vez que o objetivo das três é semelhante, quando se deve usar a média, a moda e mediana?

Se estivermos diante de uma situação na qual essas três medidas apresentem o mesmo valor, a distribuição dos dados é simétrica; quando resultam em valores diferentes, porém muito próximos, indica que a forma dessa distribuição é aproximadamente simétrica. Nesses casos, optaremos por qualquer uma das três: média, moda ou mediana. Nos demais casos, devemos analisar as especificidades da situação estudada e escolher, dentre elas, a mais adequada. Veja o resumo abaixo que irá ajudá-lo a optar por uma das três, embora nada o impeça de calcular todas elas.

**Média:** quando a distribuição dos dados é aproximadamente simétrica e não apresenta valores extremos, devemos escolher a média, pois essa medida possui propriedades matemáticas mais fortes e é muito usada para estimar a média da população quando se faz inferências. Além disso, é fácil de ser calculada e é a mais popular dentre essas medidas

**Mediana:** Quando há valores discrepantes no conjunto de dados, devemos preferir a mediana, pois ela é uma medida que não é afetada por valores extremos, podendo, assim, representar bem esses valores.

**Moda:** quando trabalhamos com variáveis qualitativas nominais, a moda é a única medida de tendência central que podemos obter. Além disso, quando queremos evidenciar o valor que mais apareceu (se repetiu) em um conjunto de dados, também usamos a moda.

## 6. Separatrizes

A mediana caracteriza uma série de valores devido à sua posição central. No entanto, ela apresenta outra característica, tão importante quanto a primeira: **ela separa a série em dois grupos que apresentam o mesmo número de valores.**

As separatrizes são aquelas medidas que separam ou que dividem o conjunto em certo número de partes iguais. No caso da mediana, vimos que ela divide o conjunto em duas metades. Já o quartil, separa o conjunto em quatro partes iguais; o decil, em dez partes e, o centil (ou percentil), em cem partes iguais.

## 6.1 Os quartis

Chamamos de quartis os valores de uma série que a dividem em quatro partes iguais. Então deveremos ter 3 quartis ( $Q_1$ ,  $Q_2$  e  $Q_3$ ) para dividir uma série em quatro partes iguais.

- O primeiro quartil ( $Q_1$ ) é um valor tal que 75% dos dados ficam acima dele, e apenas 25% abaixo.
- No segundo quartil<sup>22</sup> ( $Q_2$ ) metade dos dados estão acima e metade abaixo, é equivalente a mediana.
- O terceiro quartil é o valor tal que 25% dos dados ficam acima, e 75% abaixo.

<sup>22</sup> O quartil 2 ( $Q_2$ ) sempre será igual a mediana da série.

Quando os dados não estão agrupados, para determinar os quartis o método mais prático é utilizar o princípio do cálculo da mediana para os 3 quartis.

**Exemplo 1:** Calcule os quartis da série { 5, 2, 6, 9, 10, 13, 15 }

Primeiramente, deveremos ordenar a série { 2, 5, 6, 9, 10, 13, 15 }

O valor que divide a série acima em duas partes iguais é igual a 9. Logo, a  $Md = 9$ , que será =  $Q_2$ .

Temos agora {2, 5, 6} e {10, 13, 15} como sendo os dois grupos de valores iguais proporcionados pela mediana (quartil 2). Para o cálculo dos quartis 1 e 3, basta calcular as medianas das partes iguais provenientes da verdadeira Mediana da série (quartil 2). Portanto, para o grupo {2, 5, 6}, teremos  $Q_1 = 5$  e para {10, 13, 15} teremos  $Q_3 = 13$

**Exemplo 2:** Calcule os quartis da série: { 1, 1, 2, 3, 5, 5, 6, 7, 9, 9, 10, 13 }

$$Q_2 = Md = (5+6)/2 = 5,5$$

O quartil 1 será a mediana da série à esquerda de  $Md$ : { 1, 1, 2, 3, 5, 5 }

$$Q_1 = Md = (2+3)/2 = 2,5$$

O quartil 3 será a mediana da série à direita de  $Md$ : {6, 7, 9, 9, 10, 13 }

$$Q_3 = Md = (9+9)/2 = 9$$

Quando os dados são agrupados, para determinar os quartis usamos a mesma técnica do cálculo da mediana, bastando substituir, na fórmula da mediana,  $\frac{\sum f_i}{2}$  por

$$\frac{k \sum f_i}{4}, \text{ sendo } k \text{ o número de ordem do quartil.}$$

Assim, teremos

$$Q_1 = l + \left[ \frac{\left(\frac{\sum f}{4}\right) - F(\text{ant})}{f} \right] \cdot h$$

e

$$Q_3 = l + \left[ \frac{\left(\frac{3 \sum f}{4}\right) - F(\text{ant})}{f} \right] \cdot h$$

$l$  é o limite inferior da classe  
 $f$  é a frequência simples da classe  
 $F(\text{ant})$  é a frequência acumulada da classe anterior,  
 $h$  é a amplitude do intervalo da classe mediana.

**Exemplo:** Considere a distribuição da tabela abaixo:

Tabela 8

Pesos	$f_i$	$F_i$
40   43	3	3
43   46	4	7
46   49	10	17
49   52	13	30
52   55	10	40
55   58	6	46
58   61	4	50
<b>Total</b>	<b><math>\Sigma = 50</math></b>	

$Q_1$

$Q_3$

- Primeiro quartil

Temos

$\frac{\sum f_i}{4} = \frac{50}{4} = 12,5$  (como  $17 \geq 12,5$  a classe do nosso primeiro quartil será 46 | 49)

$$Q_1 = 46 + \frac{(12,5 - 7)3}{10} = 46 + \frac{16,5}{10} = 46 + 1,65 = 47,65.$$

- Terceiro quartil

Temos

$\frac{3 \sum f_i}{4} = \frac{3 \times 50}{4} = 37,5$  (como  $40 \geq 37,5$  a classe do nosso terceiro quartil será 52 | 55)

$$Q_3 = 52 + \frac{(37,5 - 30)3}{10} = 52 + \frac{22,5}{10} = 52 + 2,25 = 54,25.$$

## 6.2 Os percentis

Chamamos de percentis as medidas que dividem a série em 100 partes iguais. ( $P_1, P_2, P_3, \dots, P_{99}$ ).

O cálculo de um percentil segue a mesma técnica do cálculo da mediana. A fórmula  $\frac{\sum f_i}{2}$  será substituída por

$$\frac{k \sum f_i}{100}, \text{ sendo } k \text{ o número de ordem do percentil.}$$

Assim, para o 27º percentil, temos

$$P_{27} = l + \left[ \frac{\left( \frac{27 \in f_i}{100} \right) - F(\text{ant})}{f} \right] \cdot h$$

**Exemplo:** Considerando a tabela anterior, temos para o oitavo percentil

$$\frac{8 \sum f_i}{100} = \frac{8 \times 50}{100} = 4$$

Logo,

$$P_8 = 43 + \frac{(4 - 3)3}{4} = 43 + 0,75 = 43,75$$

### Atividades de avaliação



1. Qual é a média de uma sala de 50 alunos, cujas notas obtidas formaram a seguinte distribuição?

Número de alunos	$f_i$
1	2
3	3
6	4
10	5
13	6
8	7
5	8
3	9
1	10

2. Calcule a média aritmética das distribuições de freqüência a seguir:

Notas	$f_i$
0   2	5
2   4	8
4   6	14
6   8	10
8   10	7

3. Considerando os conjuntos de dados a seguir, calcule a média, a mediana e a moda.

3, 5, 2, 6, 5, 9, 5, 2, 8, 6

20, 9, 7, 2, 12, 7, 2, 15, 7

51,6; 48,7; 50,3; 49,5; 48,9

15, 18, 20, 13, 10, 16, 14

4. Determine o valor da mediana da distribuição a seguir:

$x_i$	$f_i$
0   10	3
10   20	5
20   30	8
30   40	4
40   50	2

5. Os salários dos empregados de uma empresa estão distribuídos conforme tabela a seguir:

Faixa salarial (salários mínimos)	Número de empregados
01   5	15
09   40	40
09   13	10
13   17	5

Qual o salário mediano da empresa?

- 7 salários mínimos.
- 40 salários mínimos.
- 6,82 salários mínimos.
- 9 salários mínimos

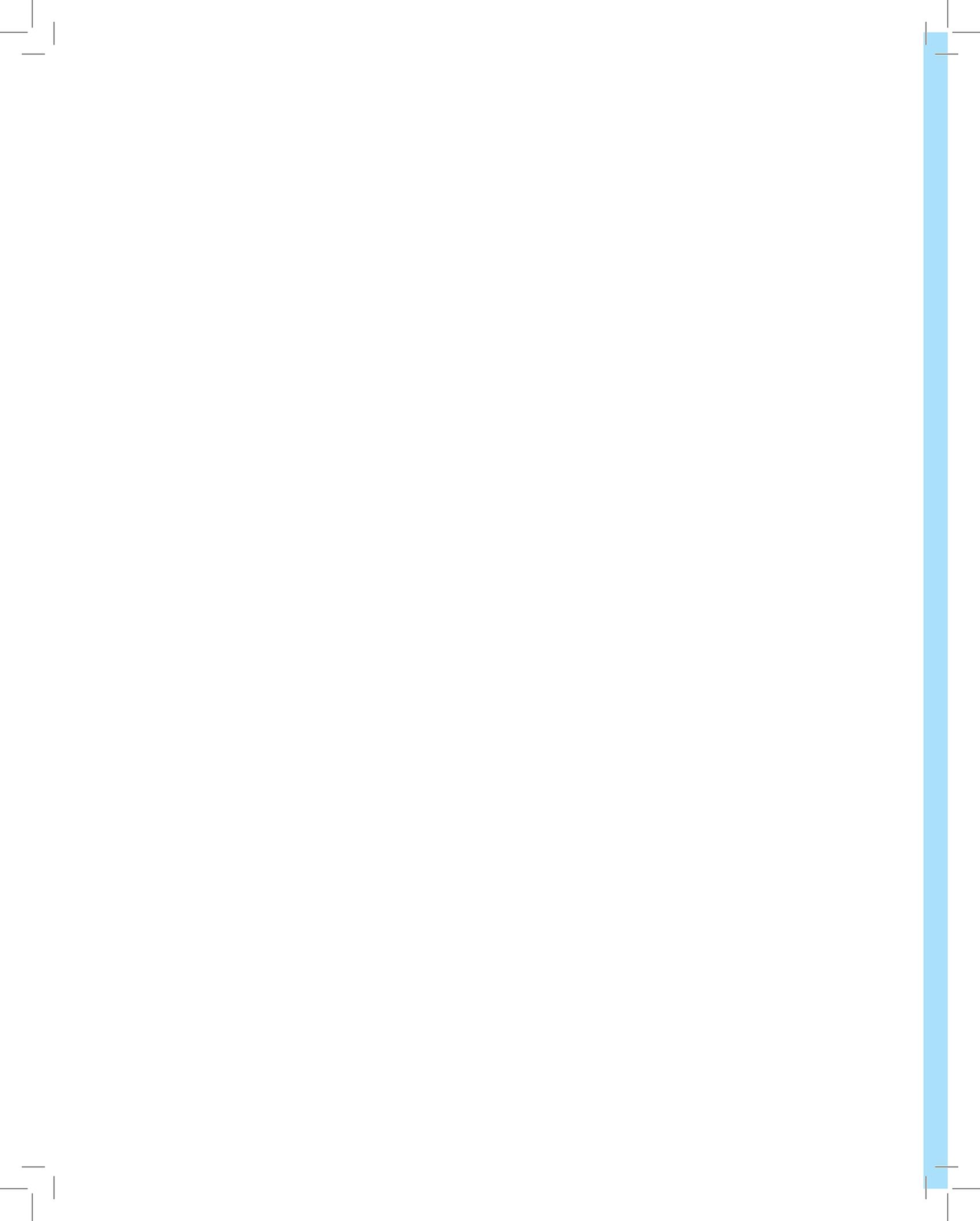
6. A série (40, 60, 70, 80, 90, 40, 70) é

- a) amodal
- b) bimodal
- c) unimodal
- d) multimodal

7. Calcule a média aritmética, a mediana, a moda, o primeiro e o terceiro quartis e o 10<sup>o</sup> percentis da distribuição a seguir:

Notas	f
0   2	5
2   4	8
4   6	14
6   8	10
8   10	7
	$\Sigma = 44$

## Objetivos



# 7

## Capítulo

# Medidas de Dispersão ou de Variabilidade



- Apresentar as principais medidas de dispersão ou de variabilidade.
- Conceituar as medidas de assimetria e de curtose.

## 1. Dispersão ou variabilidade

Sabemos que as **medidas de posição**<sup>23</sup> apresentam apenas uma das características dos valores numéricos de um conjunto de observações, o da tendência central. Nenhuma delas informa sobre o grau de variação ou de dispersão dos valores observados. Em qualquer grupo de dados, os valores numéricos não são semelhantes e apresentam desvios variáveis em relação à tendência geral de média.

<sup>23</sup> Medidas de posição: média, mediana e moda.

As medidas de dispersão ou de variabilidade têm como objetivo avaliar o quanto estão dispersos os valores de uma distribuição de frequência, ou seja, o grau de afastamento ou de concentração entre os valores.

A média que é considerada como um número que representa uma série de valores não pode, por si mesma, destacar o grau de homogeneidade ou de heterogeneidade que há entre eles.

Analisemos, por exemplo, os conjuntos de valores a seguir

- A = {60, 60, 60, 60, 60}
- B = {58, 59, 60, 61, 62}
- C = {5, 10, 40, 100, 145}

Ao calcularmos a mesma média aritmética desses conjuntos:

$$\bar{a} = \frac{\sum a_i}{n} = \frac{300}{5} = 60$$

$$\bar{b} = \frac{\sum b_i}{n} = \frac{300}{5} = 60$$

$$\bar{c} = \frac{\sum c_i}{n} = \frac{300}{5} = 60$$

Verificamos que os três conjuntos apresentam a mesma média aritmética: 60.

Chegamos à conclusão que o conjunto A é mais homogêneo que os conjuntos B e C, visto que todos os valores são iguais à média. O conjunto B, por sua vez, é mais homogêneo que o conjunto C, pois há menor diversificação entre cada um de seus valores e a média representativa. Logo, o conjunto A apresenta dispersão nula, e o conjunto B apresenta uma dispersão menor que C.

As principais medidas de dispersão são

1. Amplitude total.
2. Desvio médio absoluto.
3. Variância.
4. Desvio padrão.
5. Coeficiente de Variação

### 1.1 Amplitude total

A amplitude total em dados não agrupados é a diferença entre o maior e o menor valor da série de dados, ou seja,

$$AT = X_{\text{máximo}} - X_{\text{mínimo}}$$

Considerando os valores 30, 45, 48, 62 e 72 teremos

$$AT = 72 - 30 = 42$$

A amplitude total tem o inconveniente de só levar em conta os dois valores extremos da série, descuidando do conjunto de valores intermediários, o que quase sempre invalida a idoneidade do resultado. Ela é apenas uma indicação aproximada da dispersão ou da variabilidade.

No caso em que os dados estejam agrupados sem intervalos de classe, ainda teremos

$$AT = X_{\text{máximo}} - X_{\text{mínimo}}$$

Vejam os exemplos a seguir.

Tabela 1

$x_i$	$f_i$
0	2
1	6
3	5
4	3

Dados Hipotéticos

Calculando a amplitude,

$$AT = X_{\text{máximo}} - X_{\text{mínimo}} = 4 - 0 = 4$$

Com intervalos de classe a amplitude total é a diferença entre o limite superior da última classe e o limite inferior da primeira classe.

Logo,

$$AT = L \text{ máximo} - L \text{ mínimo.}$$

Tabela 2

Classes	Frequência
150   154	4
154   158	9
158   162	11
162   166	8
166   170	5
170   174	3

Dados Hipotéticos

Então,

$$AT = 174 - 150 = 24$$

Quanto maior a amplitude total, maior a dispersão ou variabilidade dos valores da variável.

A amplitude total é muito utilizada quando se deseja determinar a amplitude da temperatura em um dia ou no ano, no controle de qualidade ou uma medida de cálculo rápido sem muita exatidão.

## 1.2 Desvio médio absoluto

Como a amplitude total não leva em consideração todos os valores da série de dados, é preferível se trabalhar com medidas que utilizam toda a informação disponível do conjunto de dados. Uma dessas medidas é o desvio médio, que é a média aritmética dos desvios absolutos dos elementos da série, tomados em relação à sua média aritmética, que é representada por DMA e calculada pela fórmula a seguir.

$$DMA = \frac{\sum |x_i - \bar{x}|}{n}$$

Considerando a série de dados 1, 2, 3, 4, 5, calculamos o desvio médio absoluto, ou simplesmente desvio médio.

<sup>24</sup> As barras verticais indicam que são tomados os valores absolutos.

Calculando inicialmente a média, teremos

$$x = \frac{(1 + 2 + 3 + 4 + 5)}{5} = \frac{15}{5} = 3$$

Calculando o numerador da nossa fórmula,

$$\sum |x_i - \bar{x}| = |1-3| + |2-3| + |3-3| + |4-3| + |5-3| = 2 + 1 + 0 + 1 + 2 = 6$$

Logo,

$$DMA = \frac{6}{5} = 1,2$$

### 1.3 Variância

A variância mede a dispersão dos dados em torno de sua média, levando em consideração a totalidade dos valores da variável em estudo, o que a torna um índice de variabilidade bastante estável. A variância é representada por  $s^2$  e definida como sendo a média dos quadrados dos desvios em relação à média aritmética.

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

Como exemplo, consideremos que foi aplicado um teste a dois grupos com cinco alunos cada, o grupo A obteve os seguintes pontos 6, 8, 7, 4, 10 e o grupo B 9, 7, 8, 5, 6. Utilizando a variância vamos determinar o grupo mais regular.

Cálculo da variância do grupo A,

$$\bar{x} = \frac{(6 + 8 + 7 + 4 + 10)}{5} = \frac{35}{5} = 7$$

$$s^2 = \frac{(6-7)^2 + (8-7)^2 + (7-7)^2 + (4-7)^2 + (10-7)^2}{5} =$$

$$\frac{(-1)^2 + (1)^2 + (0)^2 + (-3)^2 + (3)^2}{5} = 4$$

Cálculo da variância do grupo B:

$$\bar{x} = \frac{(9 + 7 + 8 + 5 + 6)}{5} = \frac{35}{5} = 7$$

$$s^2 = \frac{+(7-7)^2 + (8-7)^2 + (5-7)^2 + (6-7)^2}{5} =$$

$$\frac{(2)^2 + (0)^2 + (1)^2 + (-2)^2 + (-1)^2}{5} = \frac{10}{5} = 2$$

Assim, concluímos que o grupo B demonstra maior regularidade, tendo em vista que a variância foi menor do que a do grupo A. Isso quer dizer que seus valores estão mais próximos da média do grupo.

#### • Fórmula alternativa para o cálculo da variância

Quando a média é um valor decimal não é exato, a fórmula da variância apresentada anteriormente não é muito prática, uma vez que entrará no cálculo “n” vezes aumentando os erros de arredondamento que ocorrem. Então, é melhor utilizar uma expressão alternativa que é obtida através de algumas manipulações algébricas na fórmula anterior.

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

### 1.4 Desvio padrão

É a medida de dispersão geralmente mais empregada, pois leva em consideração a totalidade dos valores da variável em estudo. O desvio padrão é uma medida de dispersão usada com a média. Mede a variabilidade dos valores à volta da média. O valor mínimo do desvio padrão é 0, indicando que não há variabilidade, ou seja, que todos os valores são iguais à média. O símbolo para o desvio padrão em um conjunto de dados observados é s, e a fórmula para obter o desvio padrão é a seguinte.

$$s = \sqrt{\frac{\sum (xi - \bar{x})^2}{n}}$$

A fórmula anterior é utilizada quando estamos trabalhando com uma distribuição de dados não agrupados.

Considerando os valores 30, 45, 48, 62 e 72, vamos construir a tabela a seguir para facilitar o cálculo do desvio padrão<sup>25</sup>:

Tabela 3

$X_i$	$\bar{X}_i$	$X_i - \bar{X}_i$	$(X_i - \bar{X}_i)^2$
30	51,4	-21,4	457,96
45	51,4	-6,4	40,96
48	51,4	-3,4	11,56
62	51,4	10,6	112,36
72	51,4	20,6	424,36
Total	-	-	1047,2

<sup>25</sup> O desvio padrão é a raiz quadrada da variância.

Como  $n = 5$ , temos

$$s = \sqrt{\frac{1047,2}{5}} = 14,47$$

### Propriedades do desvio padrão

- Ao adicionarmos ou subtrairmos uma constante a todos os valores de uma variável, o desvio padrão<sup>26</sup> não se altera.
- Ao multiplicarmos ou dividirmos todos os valores de uma variável por uma constante (diferente de zero), o desvio padrão fica multiplicado (ou dividido) por essa constante.

Quando os dados estão agrupados, a fórmula do desvio padrão será

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$$

Consideremos como exemplo a distribuição da tabela a seguir.

Tabela 4

$x_i$	$f_i$	$f_i x_i$	$f_i x_i^2$
0	2	0	0
1	6	6	6
3	5	15	45
4	3	12	48
	<b><math>\Sigma = 16</math></b>	<b><math>\Sigma = 33</math></b>	<b><math>\Sigma = 99</math></b>

Logo,

$$s = \sqrt{\frac{99}{16} - \left(\frac{33}{16}\right)^2} = \sqrt{6,18 - 2,06} = 2,03$$

### 1.5 Coeficiente de variação

O desvio padrão é uma medida limitada se utilizada isoladamente. Por exemplo, um desvio padrão de 2 unidades pode ser considerado pequeno para uma série de valores cujo valor médio é 200; no entanto, se a média for igual a 20, o mesmo não pode ser dito.

Quando desejamos comparar duas ou mais unidades relativamente à sua dispersão ou variabilidade, o desvio padrão não é a medida mais indicada, visto que ele se encontra na mesma unidade dos dados.

<sup>26</sup> Ao aplicar a fórmula do cálculo do desvio padrão para dados agrupados com intervalo de classe, lembre-se de que o valor de  $x_i$  é o ponto médio de cada classe.

Assim, contornamos este problema caracterizando a dispersão em termos relativos ao seu valor médio. Essa medida é denominada de coeficiente de variação de Pearson.

O coeficiente de variação de Pearson é a razão entre o desvio padrão e a média referentes aos dados de uma mesma série.

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

Consideremos o exemplo da tabela a seguir:

Tabela 5

Discriminação	Média	Desvio padrão
Estaturas	165 cm	3,0 cm
Pesos	55 kg	2,5 kg

Calculando os coeficientes de variação, temos

$$CV_e = \frac{3}{165} * 100 = 1,81\%$$

$$CV_p = \frac{2,5}{55} * 100 = 4,54\%$$

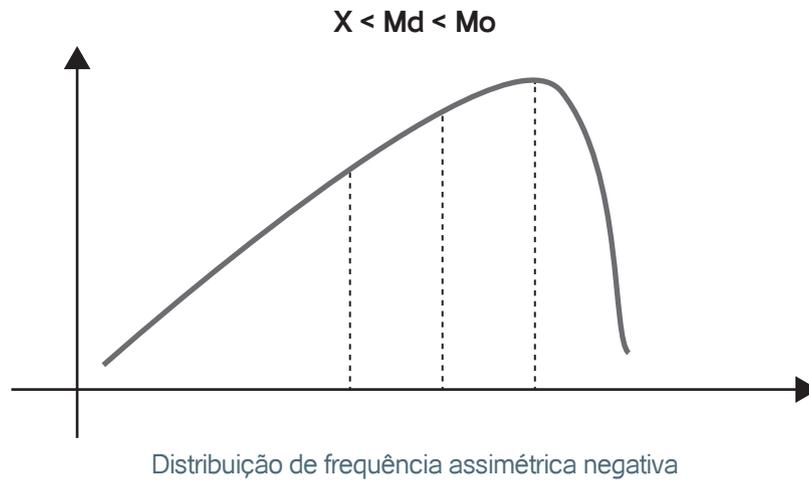
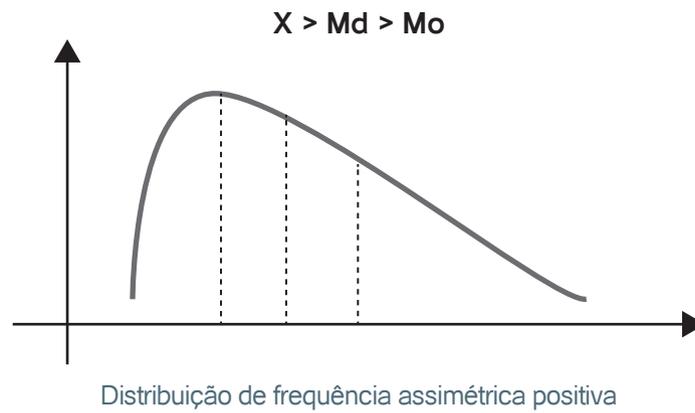
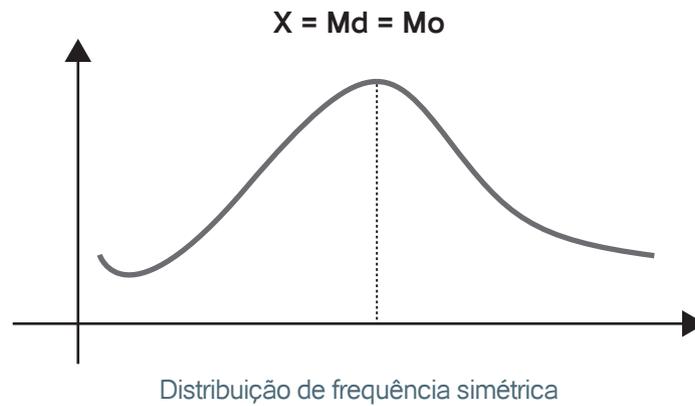
Logo, nesse grupo de indivíduos, os pesos apresentaram maior grau de dispersão em relação às estaturas.

### Para Refletir

1. Considere o seguinte conjunto de dados: 2, 3, 5, 7, 10. Utilize a fórmula alternativa para calcular a variância, sabendo que a média é 5,4.
2. O desvio padrão e a variância podem ser negativos?
3. Em que situação o desvio padrão e a variância são nulos? Qual e a amplitude neste caso?
4. Calcule o desvio padrão da idade de 5 pessoas  $I = \{10, 13, 24, 47, 50\}$
5. Sabendo que um conjunto de dados apresenta para média aritmética e para desvio padrão, respectivamente, 18,3 e 1,47, calcule o coeficiente de variação.
6. Um grupo de 100 estudantes tem uma estatura média de 163,8 cm, com um coeficiente de variação de 3,3%. Qual o desvio padrão desse grupo?

## 2. Assimetria

Uma distribuição de frequência pode ser caracterizada como simétrica ou assimétrica. Essa característica está associada à distribuição relativa da suas medidas de posição central: média, moda e mediana.



Baseando-se nas relações entre a média e a moda, podemos verificar o tipo de assimetria calculando o valor de suas diferenças.

$$x - Mo$$

Se

- $x - Mo = 0 \Rightarrow$  assimetria nula ou distribuição simétrica;
- $x - Mo < 0 \Rightarrow$  assimetria negativa ou à esquerda;
- $x - Mo > 0 \Rightarrow$  assimetria positiva ou à direita.

## 2.1 Coeficiente de assimetria de Person - $A_s$

A medida da diferença entre a média e a moda, por ser absoluta, sofre das mesmas restrições do desvio padrão, isto é, não permite a possibilidade de comparação entre as medidas de duas distribuições. Portanto, deve ser dada preferência ao coeficiente de assimetria de Person.

$$A_s = \frac{3(\bar{x} - Md)}{s}$$

Se  $0,15 < |A_s| < 1$ , a assimetria é considerada **moderada**.

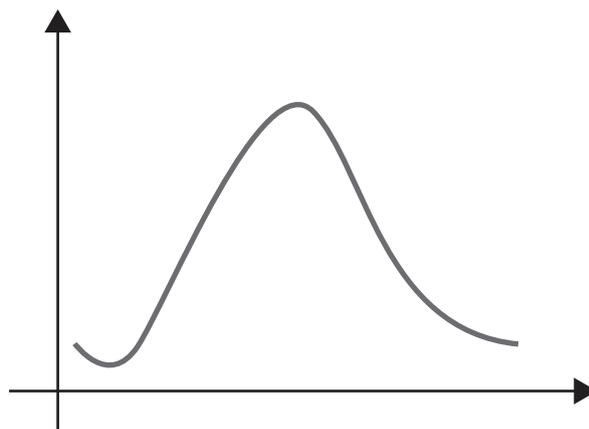
Se  $|A_s| > 1$ , a assimetria é considerada **forte**.

## 3. Curtose

A curtose exprime o grau de “achatamento” de uma distribuição de frequência em relação a uma distribuição padrão denominada **curva normal**<sup>27</sup>.

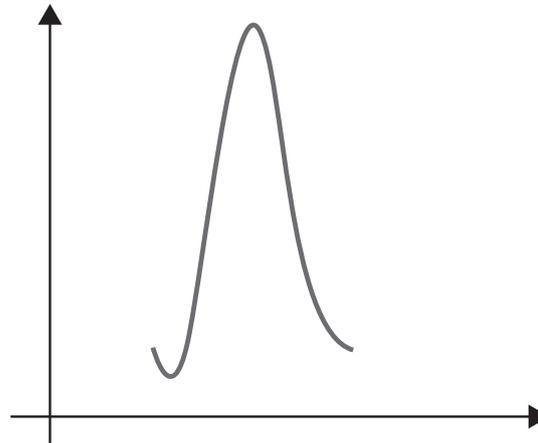
Essa característica está associada à concentração dos resultados. Assim, quanto à curtose, a distribuição de frequência pode ser: leptocúrtica, mesocúrtica e platicúrtica.

<sup>27</sup> Curva normal: curva correspondente a uma distribuição teórica de probabilidade.



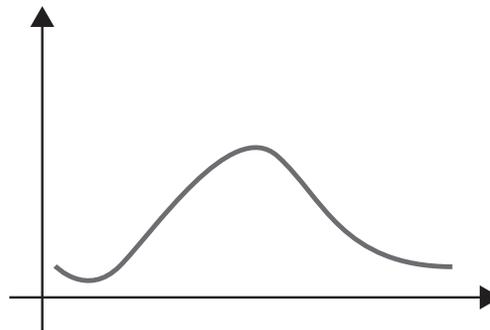
Distribuição de frequência mesocúrtica ou normal

A curva normal é a curva de base referencial recebe o nome de mesocúrtica.



Distribuição de frequência leptocúrtica

Quando a distribuição apresenta uma curva de frequência mais fechada que a normal ou mais aguda em sua parte superior, ela recebe o nome de **leptocúrtica**.



Distribuição de frequência platicúrtica

Quando a distribuição apresenta uma curva de frequência mais aberta que a normal ou mais achatada na sua parte superior ela recebe o nome de **platicúrtica**.

### 3.1. Coeficiente de Curtose – K

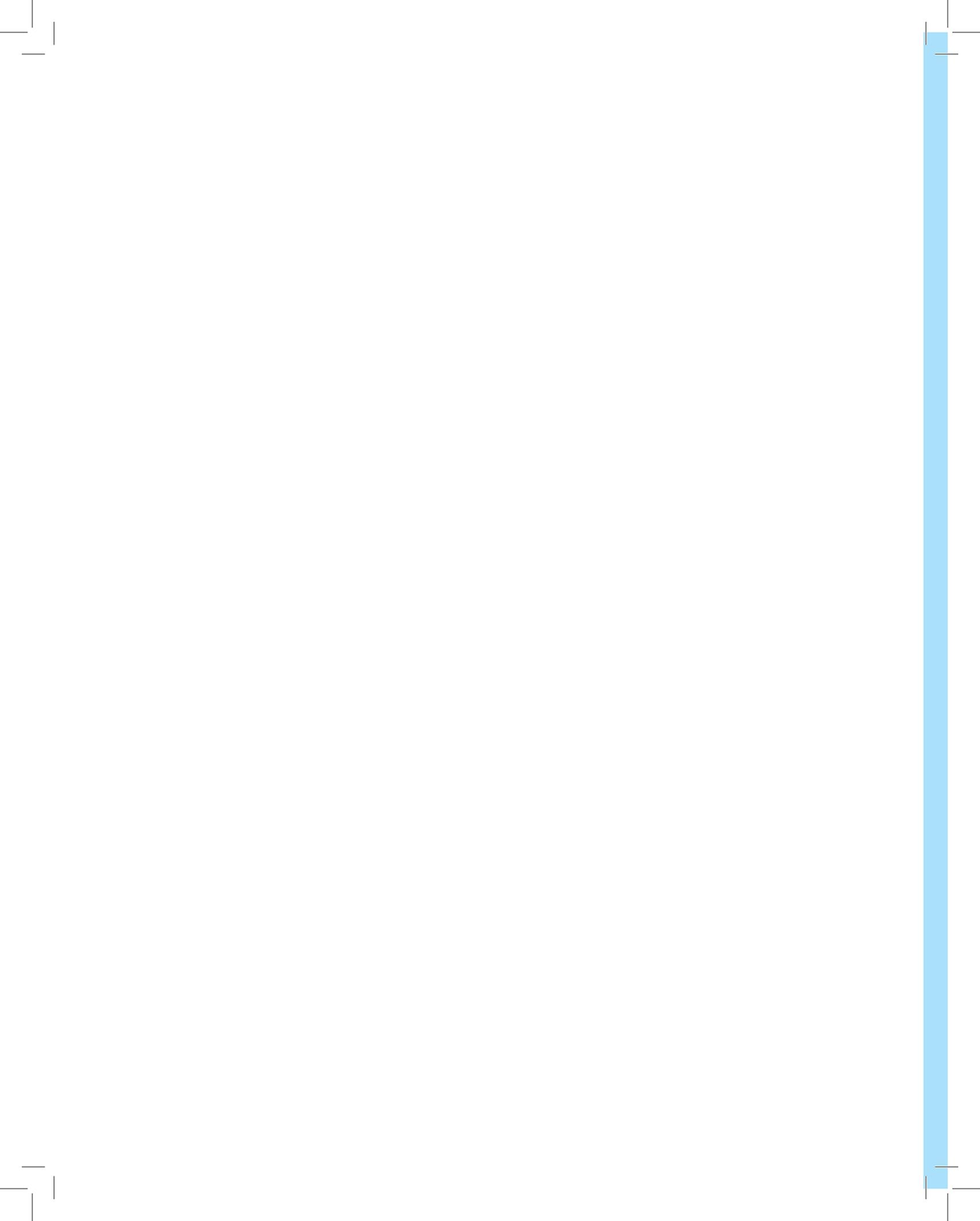
$$k = \frac{\sum (x_i - \bar{x})^4 * f_i}{n * s^4} - 3$$

- Se  $k = 0 \Rightarrow$  curva mesocúrtica.
- Se  $k > 3 \Rightarrow$  curva leptocúrtica.
- Se  $k < 3 \Rightarrow$  curva platicúrtica.

## Referências



- BARBETA, P. A., REIS, M. M., BORNIA, A. C. **Estatística para Cursos de Engenharia e Informática**, São Paulo: Atlas, 2004.
- BOLFARINE, Heleno & BUSSAB, Wilton O. **Elementos de Amostragem**. 1. ed. São Paulo: Edgard Blücher, 2005.
- BUSSAB, Wilton de Ol., MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva, 2002.
- COSTA, F. S. **Introdução Ilustrada à Estatística**. São Paulo: Harbra, 1998.
- CRESPO, A. **Estatística Fácil**. Rio de Janeiro: Saraiva, 1996.
- FONSECA, Jairo Simon da. **Introdução a Estatística**. Rio de Janeiro: LTC, 1993.
- JAIRO, Simon da Fonseca, MARTINS, Gilberto de Andrade. **Curso de Estatística**, São Paulo: Atlas, 1996.
- MARTINS G. A. **Estatística Geral e Aplicada**. São Paulo: Atlas, 2001.
- MEYER, Paul L. **Aplicações a Estatística**. Rio de Janeiro: LTC, 1996.
- MIRSHAWKA, V. **Probabilidade e Estatística para Engenharia**, São Paulo: Nobel, 1978.
- MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva, 2013.
- PEREIRA, Wilson., TANAKA, Oswaldo K. **Estatística – Conceitos Básicos**, Makron Books, 1990.
- ROSS, Sheldon. **A First Course in Probability**. 7. ed. New Jersey: Prentice Hall, 2005.
- SPIEGEL M. R. **Estatística**. Coleção Schaum. São Paulo: Editora Afiliada, 1993.
- TRIOLA M. F. **Introdução à Estatística e Probabilidade - Exercícios Resolvidos e Propostos**. Rio de Janeiro: LTC, 1999.
- TRIOLA, Mario F. **Introdução a Estatística**. Rio de Janeiro: LTC, 2005.
- VIEIRA, S. **Estatística Experimental**. São Paulo: Atlas, 1999.



Capítulo

8

# Fundamentos de Probabilidade



## Objetivos

- Apresentar e discutir os principais conceitos de probabilidade.
- Apresentar as variáveis aleatórias discretas contínuas e as funções densidade de probabilidade além das distribuições amostrais, estimação de parâmetros e os testes de hipóteses.
- Compreender e adotar corretamente os conceitos e atributos fundamentais para o uso da probabilidade.

## 1. Introdução

Probabilidade é o estudo de experimentos aleatórios ou não determinísticos. Mas o que é experimento? Um experimento é qualquer processo de observação. Um experimento pode ser, por exemplo

- Uma observação meteorológica ou sísmica. Nesse caso temos observações de experimentos naturais.
- Uma pesquisa de opinião para saber quantos eleitores votarão no candidato  $x$  ou  $y$  na próxima eleição ou para saber quantos alunos almoçam no RU (restaurante universitário) da UECE.
- Uma verificação de um exame de sangue ou o teste de fadiga de determinado material da construção civil são observações de experimentos controlados.

Nos experimentos mencionados, pode-se notar que a incerteza sempre está presente, o que quer dizer que, se esses experimentos forem repetidos em condições idênticas, não se pode determinar qual o resultado que ocorrerá. Tais experimentos são conhecidos como **experimentos aleatórios**. A incerteza esta associada à chance de ocorrência que atribuímos ao resultado de interesse.

**Exemplo 1:** Vai chover neste final de semana na Praia do Futuro, em Fortaleza?

Conjunto de possibilidades:  $S = \{\text{chove, não chove}\}$

Para calcular a probabilidade de chover, podemos ou usar a intuição (subjetivo) ou usar a frequência relativa dos últimos dez fins de semana em que choveu (objetivo).

**Exemplo 2:** Lançamento de um dado: você ganha se sair uma face par.

Conjunto de possibilidades:  $S = \{1, 2, 3, 4, 5, 6\}$

Conjunto de possibilidades favoráveis:  $F = \{2, 4, 6\}$

Probabilidade de você ganhar = ?

Supondo que um dado é honesto e equilibrado, será natural atribuímos a probabilidade

$$\frac{3}{6} = \frac{\text{número de casos favoráveis}}{\text{número de casos possíveis}}$$

A teoria da probabilidade está baseada na estabilidade da frequência relativa após  $N$  repetições de um experimento. Essa estabilidade (aproxima-se de um limite) é base da Teoria da Probabilidade. A definição clássica ou histórica de probabilidade é dada por

$$P(A) = \frac{n(A)}{n(S)},$$

onde,  $P(A)$  = probabilidade de ocorrer o evento  $A$ ,  $n(A)$  = número de casos favoráveis à ocorrência do evento  $A$  e  $n(S)$  = números de casos possíveis do experimento, desde que sejam igualmente prováveis. Essa definição pode ser aplicada apenas a uma classe limitada de problemas, isto é, aqueles em que é possível contar os elementos do espaço amostral,  $S$ , e do evento  $A$ . Nessa contagem, a técnica usada é análise combinatória.

Historicamente, a Teoria da Probabilidade começou com o estudo de jogos de azar (roleta, cartas etc.). Vejamos um breve histórico a partir do século XVI.

#### A era dos jogos de azar

- Cardano (1501-1576) → 1º matemático a calcular uma probabilidade correta.
- Fermat (1601-1655), Pascal (1623-1662) e Huygens (1629-1695) → envolveram-se na solução de um dos primeiros problemas de probabilidade.

#### O começo

- Bernoulli (1654-1705) → provou a Lei dos Grandes Números.
- Laplace (1749-1827) → publicou a *Théorie Analytique des Probabilités*.
- Poisson (1781-1840) → distribuição de probabilidade com seu nome.
- Gauss (1777-1855) → originou a Teoria dos Erros, particularmente, a dos Mínimos Quadrados.

#### Estamos chegando

- Chebyshev (1822-1894) → explorou as relações entre variáveis aleatórias e suas esperanças.

- Markov (1856-1922) → criou um novo ramo que é a Teoria das Variáveis Aleatórias Dependentes, conhecida como Cadeias de Markov.

### Axiomatização

- Borel (1871-1956) → criou a álgebra de Borel.
- Kolmogorov → publicou em 1933 um livro com a axiomática usada até hoje.

### Mente brilhante

- John von Neumann (1903-1957) → criou a Teoria dos Jogos (1928) e contribuiu para a descoberta da mecânica quântica e para o desenvolvimento da 1ª bomba atômica. É o pai da arquitetura digital dos computadores.

### Hoje

- Idéias recentes → Probabilidade Intervalar, Probabilidade Imprecisa, Teoria dos Fractais e a Teoria da Complexidade.

## 2. Espaço Amostral e Eventos

**Espaço amostral**, denotado por  $S$ , é o conjunto de todos os resultados possíveis de um experimento. Um resultado particular de  $S$  é um ponto amostral.

**Evento** é qualquer subconjunto do espaço amostral  $S$ , ou seja, é um conjunto de resultados possíveis do experimento aleatório. Ao se realizar um experimento aleatório, se o resultado pertence a um dado evento  $A$ , diz-se que  $A$  ocorreu. O evento  $A = \{a\}$ , em que  $a \in S$ , consistindo do único ponto amostral é o evento elementar.

Evento impossível: é o evento igual ao conjunto vazio ( $\emptyset$ ).

Evento certo: é o evento igual ao espaço amostral  $S$ .

**Exemplo 3:**  $S = \{\text{chove, não chove}\}$

Em geral, temos interesse em eventos particulares do experimento.

- **Evento A:** chove

Então,  $A = \{\text{chove}\} \subset S$ . O evento  $A$  é um subconjunto de  $S$ .

**Exemplo 4:**  $S = \{1, 2, 3, 4, 5, 6\}$

- **Evento B:** sair face par.

Então,  $B = \{\text{sair face par}\} = \{2, 4, 6\} \subset S$ . O evento  $B$  é um subconjunto de  $S$ .

**Exemplo 5:** Ainda considerando  $S = \{1, 2, 3, 4, 5, 6\}$ , podemos definir outros eventos como

- **Evento C:** sair uma face ímpar.

Então,  $C = \{1, 3, 5\}$

A um experimento aleatório está associado um espaço amostral  $S$ . O evento  $A$  ocorre se o resultado do experimento pertence a  $A$ . Os conjuntos  $S$  (evento certo) e  $\emptyset$  também são eventos (evento impossível).

- **Evento D:** sair uma face maior que 3.

$$\text{Então, } D = \{4, 5, 6\}$$

- **Evento E:** sair face 1.

$$\text{Então, } E = \{1\}$$

## 2.1 Operações com eventos

- **União:**  $A \cup B$  é o evento que ocorre se e somente se  $A$  ocorre ou  $B$  ocorre ou ambos ocorrem simultaneamente.
- **Intersecção:**  $A \cap B$  é o evento que ocorre se e somente se  $A$  e  $B$  ocorrem simultaneamente. Obs: se  $A \cap B = \emptyset$ , então  $A$  e  $B$  são ditos mutuamente exclusivos ou disjuntos.
- **Complementar:**  $A^c$  ou  $\bar{A}$  é o evento que ocorre se e somente se  $A$  não ocorrer.

**Exemplo 6:** Considerando o lançamento de um dado e os eventos  $A, B, C, D$  e  $E$  definidos nos exemplos anteriores, temos

$$B \rightarrow D = \text{sair uma face par e maior que 3}$$

$$\text{Então, } B \rightarrow D = \{2, 4, 6\} \rightarrow \{4, 5, 6\} = \{4, 6\}$$

$$B \rightarrow C = \text{sair uma face par e ímpar}$$

$$\text{Então, } B \rightarrow C = \{2, 4, 6\} \rightarrow \{1, 3, 5\} = \emptyset \text{ (B e C são disjuntos)}$$

$$C = \bar{B}$$

$$B = \bar{C}$$

$$B \cup D = \text{sair uma face par ou maior que 3}$$

$$\text{Então, } B \cup D = \{2, 4, 6\} \cup \{4, 5, 6\} = \{2, 4, 5, 6\}$$

$$B \cup C = \text{sair uma face par ou ímpar}$$

$$\text{Então, } B \cup C = \{2, 4, 6\} \cup \{1, 3, 5\} = \{1, 2, 3, 4, 5, 6\}$$

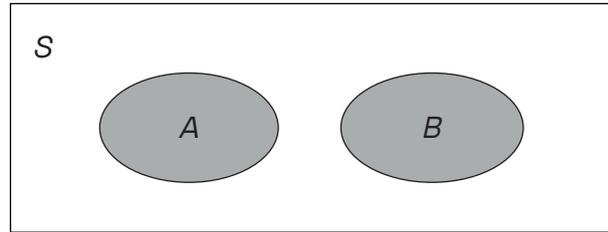
## 3. Definição de probabilidade

A probabilidade é uma função que atribui um número aos eventos de  $S$  (se  $A$  é um evento de  $S$ , então  $P(A)$  é a probabilidade de  $A$ ), que satisfaz os seguintes axiomas

- **Axioma 1:**  $\forall A, 0 \leq P(A) \leq 1$ .
- **Axioma 2:**  $P(S) = 1$ .
- **Axioma 3:** Se  $A$  e  $B$  são mutuamente exclusivos ou disjuntos, então

$$P(A \cup B) = P(A) + P(B)$$

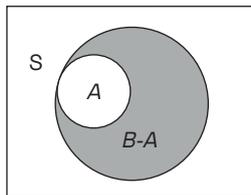
- **Axioma 4:**  $A_1, A_2, \dots$  são mutuamente exclusivos, então  $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$



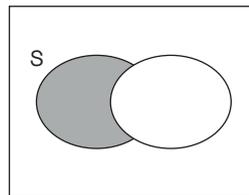
$$A \cup B = A + B$$

### Teoremas:

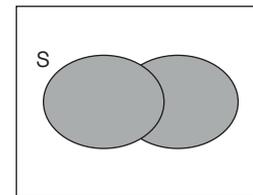
- **T1:** Se  $\emptyset$  é o conjunto vazio, então  $P(\emptyset) = 0$ .
- **T2:** Se  $A^c$  é o complemento de  $A$ , então  $P(A^c) = 1 - P(A)$ .
- **T3:** Se  $A \subset B$ , então  $P(A) \leq P(B)$ .
- **T4:** Se  $A$  e  $B$  são dois eventos quaisquer, então  $P(A - B) = P(A) - P(A \cap B)$ .
- **T5:** Se  $A$  e  $B$  são dois eventos quaisquer, então  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .
- **T6:** Se  $A$ ,  $B$  e  $C$  são eventos quaisquer, então  $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$



$$B - A$$



$$A - B$$



$$A \cup B$$

**Exemplo 7:** Dados do Censo Demográfico de 1991 publicado pelo IBGE relativos aos habitantes do Ceará, na faixa etária entre 20 a 24 anos com relação às variáveis sexo e leitura.

Sexo	Lê	Não lê	Total
Masculino	39.577	8.672	48.249
Feminino	46.304	7.297	53.601
Total	85.881	15.969	101.850

Um jovem entre 20 e 24 anos é escolhido ao acaso no estado do Ceará.  
 $S$  = conjunto de 101.850 jovens do Ceará, com idade entre 20 e 24 anos.

Eventos de interesse:

- $M$  = jovem sorteado é do sexo masculino = jovens do sexo masculino de  $S$
- $F$  = jovem sorteado é do sexo feminino
- $L$  = jovem sorteado sabe ler
- $M \cap L$  = jovem sorteado é do sexo masculino e sabe ler
- $M \cup L$  = jovem sorteado é do sexo masculino ou sabe ler

Podemos obter algumas probabilidades:

$$P(L) = \frac{\text{n}^\circ \text{ de jovens que sabem ler de } S}{\text{n}^\circ \text{ de jovens de } S} = \frac{85.881}{101.850} = 0,843$$

$$P(M) = \frac{\text{n}^\circ \text{ de jovens do sexo masculino de } S}{\text{n}^\circ \text{ de jovens de } S} = \frac{48.245}{101.850} = 0,473$$

$$F = \bar{M} \Rightarrow P(F) = P(\bar{M}) = 1 - P(M) = 1 - 0,473 = 0,527$$

$$P(M \cap L) = \frac{\text{n}^\circ \text{ de jovens do sexo masculino e que sabem ler de } S}{\text{n}^\circ \text{ total de jovens (} S \text{)}} = \frac{39.557}{101.850} = 0,388$$

$$P(M \cup L) = P(M) + P(L) - P(M \cap L) = 0,473 + 0,843 - 0,388 = 0,928$$

#### 4. Espaços amostrais finitos e infinitos

Seja  $S = \{a_1, a_2, \dots, a_n\}$  um espaço amostral finito. Obtemos um espaço de probabilidade finito se a cada ponto  $a_i \in S$  associarmos a um número real  $p_i$ , chamado de probabilidade de  $a_i$  ou  $P(a_i)$ . As seguintes propriedades são válidas para esses espaços.

Cada  $p_i$  é não negativo, ou seja,  $p_i \geq 0$

A soma dos  $p_i$  é 1, ou seja,  $\sum_{i=1}^n p_i = 1$

Observação: a probabilidade de um evento  $A$  passa a ser a soma das probabilidades dos pontos amostrais que o compõem.

**Exemplo 8:** Lance 3 moedas e observe o número de caras; então  $S = \{0, 1, 2, 3\}$ . Obtém-se o espaço de probabilidade da seguinte forma.

$$P(\{0\}) = \frac{1}{8}, P(\{1\}) = \frac{3}{8}, P(\{2\}) = \frac{3}{8} \text{ e } P(\{3\}) = \frac{1}{8}.$$

Como a probabilidade é não negativa, a soma delas é igual a 1.

Descreva os seguintes eventos e suas probabilidades:  $A$  = “pelo menos uma cara aparece” e  $B$  = “todas caras ou todas coroas aparecem”.

$$A = \{1, 2, 3\} \text{ e } B = \{0, 3\}$$

$$P(A) = P(\{1\}) + P(\{2\}) + P(\{3\}) = \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8}.$$

$$P(B) = P(\{0\}) + P(\{3\}) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Seja  $S$  um espaço amostral infinito,  $S = \{a_1, a_2, \dots\}$ . Obtém-se um espaço de probabilidade, associando a cada  $a_i \in S$  um número real  $p_i$ , chamado de probabilidade, tal que

$$\text{i) } p_i \geq 0$$

$$\text{ii) } \sum_{i=1}^{\infty} p_i = 1$$

**Exemplo 9:** Considere o espaço amostral  $S = \{1, 2, 3, \dots, \infty\}$  de um experimento de lançamento de uma moeda até que apareça uma cara. Chame de  $n$  o número de vezes que a moeda é lançada. Um espaço de probabilidade é obtido fazendo-se

$$P(\{1\}) = \frac{1}{2}, P(\{2\}) = \frac{1}{4}, \dots, P(\{n\}) = \frac{1}{2^n}, \dots, P(\{\infty\}) = 0.$$

**Em resumo:**

Relembrando a interpretação da probabilidade, seja  $A$  um evento de um experimento aleatório de um espaço amostral. Consideramos duas formas de se atribuir probabilidades aos eventos de um espaço amostral.

- $P(A)$  é uma crença (subjetiva) que se deposita na ocorrência de  $A$ .
- Interpretação frequentista ou de frequência relativa (objetiva)

$$f_n(A) = \frac{\text{n}^\circ \text{ de ocorrências de } A \text{ em } n \text{ ensaios independentes do experimento}}{n}$$

Quando  $n$  cresce:  $f_n(A) \approx P(A)$ , isto é, quando  $n$  cresce, a probabilidade é aproximada pelo valor da frequência relativa.

Há uma terceira definição ou interpretação, chamada de probabilidade geométrica, que não será abordada neste texto.

## 5. Probabilidade condicional

No exemplo anterior, se soubermos que o jovem sorteado é do sexo masculino, qual é a probabilidade de que saiba ler? Temos uma informação parcial: o jovem é do sexo masculino. Vamos designar a probabilidade de  $L$  quando se sabe que o jovem é do sexo masculino por  $P(L|M)$  e denominá-la probabilidade condicional de  $L$  dado  $M$ .

É natural atribuímos

$$P(L/M) = \frac{\text{n}^\circ \text{ de jovens que sabem ler dentre aqueles do sexo masculino}}{\text{n}^\circ \text{ total de jovens do sexo masculino}}$$

$$= \frac{39.577}{48.249} = 0,820$$

Note que

$$P(L/M) = \frac{\frac{\text{n}^\circ \text{ jovens do sexo masculino e que sabem ler}}{\text{no total de jovens}}}{\frac{\text{n}^\circ \text{ jovens do sexo masculino}}{\text{n}^\circ \text{ total de jovens}}}$$

Sejam  $A$  e  $B$  eventos de um experimento aleatório qualquer. Imitando a equação. 1, podemos dizer que a probabilidade condicional de  $A$  dado  $B$  (nota-se por  $P(A/B)$ ) é definida como

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

Por exemplo, a probabilidade de ser do sexo masculino dado que lê é dada por

$$P(M/L) = \frac{P(M \cap L)}{P(L)} = \frac{0,388}{0,843} = 0,460$$

### 5.1 Regra do produto

Da equação equação. 2 obtemos a **regra do produto** para a probabilidade da interseção de dois conjuntos:

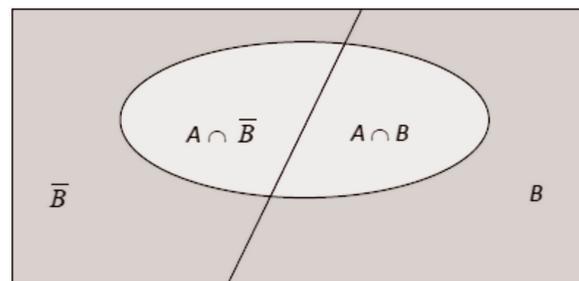
$$P(A \cap B) = P(A/B) P(B)$$

válida para quaisquer eventos  $A$  e  $B$  de  $S$ .

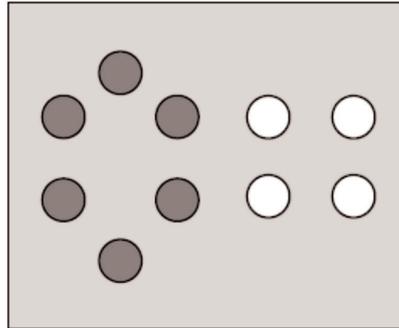
### 5.2 Regra da probabilidade total

Sejam  $A$  e  $B$  dois eventos. Há duas maneiras de  $A$  ocorrer: ou  $A$  e  $B$  ocorrem ( $A \cap B$ ) ou  $A$  e  $\bar{B}$  ocorrem ( $A \cap \bar{B}$ ).

Desse modo,  $A = \{A \cap B\} \cup \{A \cap \bar{B}\}$ , onde  $A \cap B$  e  $A \cap \bar{B}$  são conjuntos disjuntos. Pela regra da soma,  $P(A) = P(A \cap B) + P(A \cap \bar{B})$ . Pela regra do produto,  $P(A) = P(B) \cdot P(A/B) + P(\bar{B}) \cdot P(A/\bar{B})$  (regra da probabilidade total).



**Exemplo 10:** Em uma urna, há 10 bolas, sendo 4 brancas e 6 vermelhas. Duas bolas são sorteadas sucessivamente, sem reposição. Qual é a probabilidade da 2ª bola ser vermelha?

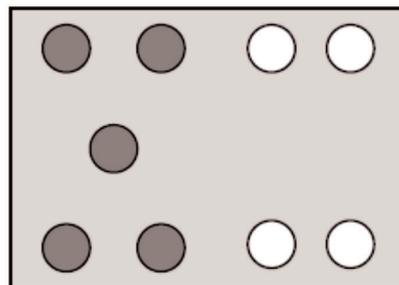


Se  $A$  é a probabilidade de a segunda bola sorteada ser vermelha, queremos calcular  $P(A)$ . Seja  $B$  a probabilidade de a primeira bola sorteada ser vermelha.

$$P(B) = \frac{6}{10}$$

$$P(\bar{B}) = 1 - P(B) = 1 - \frac{6}{10} = \frac{4}{10}$$

Se  $B$  ocorreu, isto é, saiu vermelha na primeira retirada, então



$P\{A/B\} = P\{\text{sortear 1 bola vermelha dentre 5 vermelhas e 4 brancas}\}$

$$P\{A/B\} = \frac{5}{9}$$

$P\{A/\bar{B}\} = P\{\text{sortear 1 bola vermelha dentre 6 vermelhas e 3 brancas}\} = 6/10$

Portanto,

$$P\{A\} = P\{B\} \cdot P\{A/B\} + P\{\bar{B}\} \cdot P\{A/\bar{B}\}$$

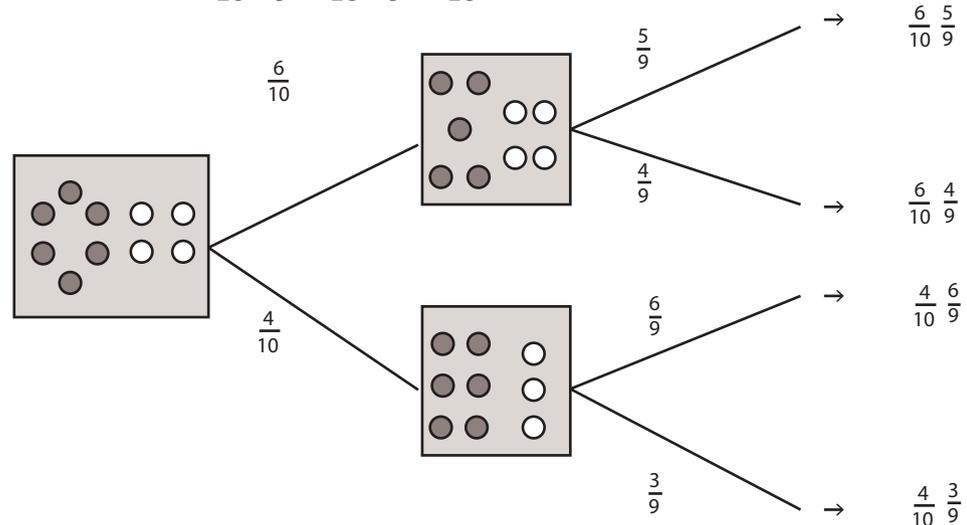
$$P\{A\} = \frac{5}{9} \cdot \frac{6}{10} + \frac{6}{9} \cdot \frac{4}{10}$$

$$P\{A\} = \frac{6}{10} \left( \frac{5}{9} + \frac{4}{9} \right)$$

$$P\{A\} = \frac{6}{10}$$

Podemos fazer o diagrama em árvore ou árvore de probabilidades da situação descrita neste exercício.

$$P\{A\} = \frac{6}{10} \cdot \frac{5}{9} + \frac{4}{10} \cdot \frac{6}{9} = \frac{6}{10}$$



### 5.3 Independência

Dois eventos  $A$  e  $B$  são independentes se  $P\{A/B\} = P\{A\}$  ou  $P\{B/A\} = P\{B\}$ .

No exemplo anterior,  $P\{A/B\} = \frac{5}{9} \neq \frac{6}{10} = P\{A\}$ , ou seja,  $A$  e  $B$  não são independentes.

Se o sorteio da 2ª bola for com reposição,

$P\{A/B\} = P\{\text{sortear 1 bola vermelha dentre 6 vermelhas e 4 brancas}\}$

$$= \frac{6}{10} = P\{A\}, \text{ então os eventos são independentes.}$$

### 5.4 Regra do produto para eventos independentes

Se  $A$  e  $B$  são eventos independentes,

$$P\{A \cap B\} = P\{B\} \cdot P\{A/B\} = P\{B\} \cdot P\{A\}$$

**Exemplo 11:** Considerando o exemplo anterior, qual a probabilidade de serem retiradas 2 bolas vermelhas (no sorteio com reposição)?

Calculando a probabilidade de  $A$ , temos

$$\begin{aligned} P\{A \cap B\} &= P\{A\} \cdot P\{B\} \\ &= \frac{6}{10} \cdot \frac{6}{10} \\ &= 0,36 \end{aligned}$$

## 6. Variáveis aleatórias, funções densidade de probabilidade

O objetivo do cálculo das probabilidades é criar modelos matemáticos capazes de representar os experimentos aleatórios.

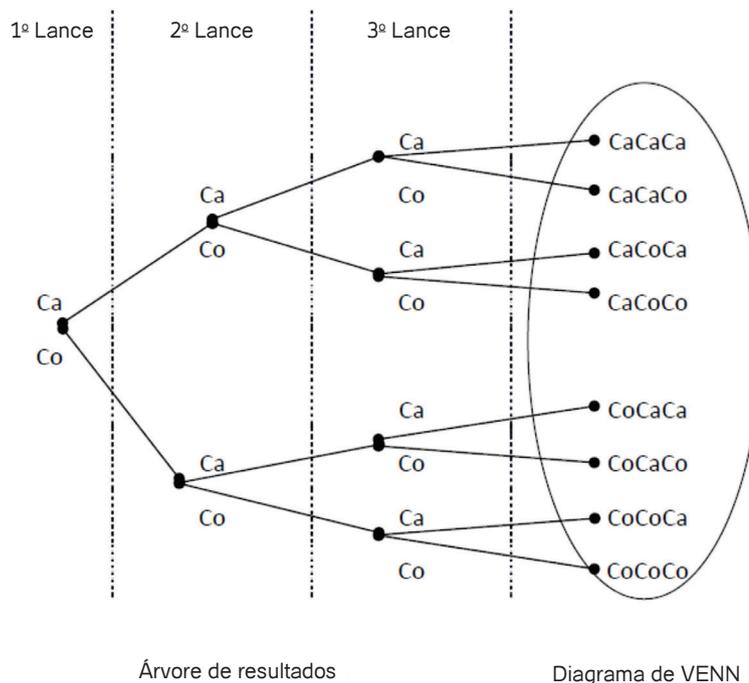
Experiência aleatória designa uma situação à qual estejam associados, de forma não controlada, dois ou mais resultados possíveis. O conjunto de todos os resultados possíveis de uma experiência aleatória é chamado de **espaço amostral**.

Os espaços amostrais podem ser discretos ou contínuos consoante os seus elementos sejam numeráveis ou não. Os espaços discretos podem ser finitos ou infinitos. O espaço amostral associado a uma experiência aleatória depende da forma como a experiência é avaliada, isto é, depende daquilo que está sendo observado.

**Exemplo 12:** Considere uma experiência constituída pelo lançamento de uma moeda ao ar por três vezes consecutivas.

Se o resultado for avaliado pelo número de “Ca” {caras} obtido, o espaço amostral é constituído pelo conjunto  $\{0, 1, 2, 3\}$ .

Se o resultado for avaliado pela sequência de “Ca”(caras) e “Co” (coroas), o espaço amostral é constituído por oito resultados possíveis.



O diagrama de Venn é utilizado na representação de resultados sequenciais. No entanto, é difícil lidar com espaços amostrais dessa maneira, é muito

mais fácil trabalhar com quantidades numéricas. Neste exemplo específico poderíamos estar interessados no número de “caras” nas 3 jogadas, e seria interessante definir uma função que associasse um número a cada resultado no **espaço amostral**.

Seja  $S$  o espaço amostral e  $X$  uma função que “pega” elementos deste espaço (resultados da experiência) e os leva num subconjunto de números reais. Essa função  $X$  é chamada de variável aleatória.

### 6.1 Definição (variável aleatória)

Considere uma experiência aleatória com espaço amostral  $S$ . Seja  $c$  um elemento de  $S$ . Uma variável aleatória  $X$  é uma função que associa um único número real  $X(c) = x$  a cada elemento do espaço amostral  $S$ . O espaço de  $X$  é o conjunto de números reais  $\mathfrak{X} = \{x : x = X(c) \forall c \in S\}$ .

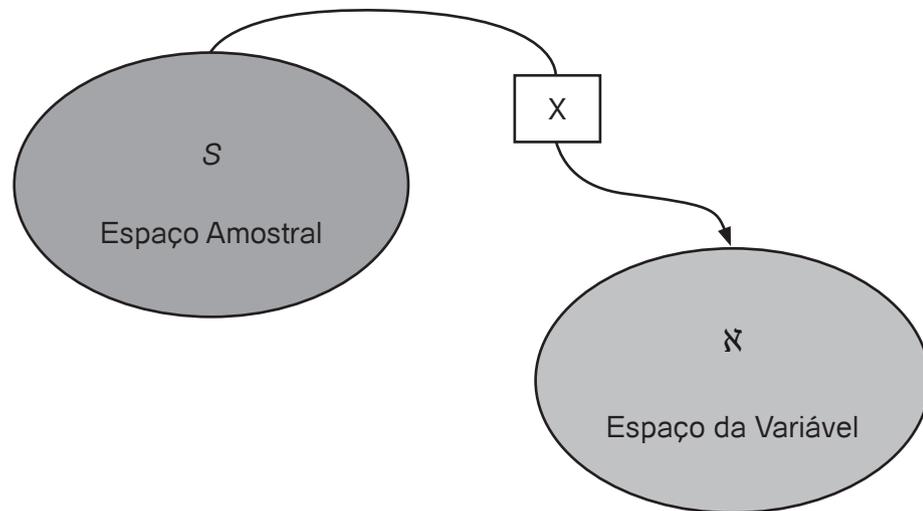


Figura 4 - Espaço amostral espaço da variável aleatória

Seja  $X$  uma variável aleatória definida num espaço amostral  $S$  e seja  $\rightarrow$  o espaço de  $X$ . Seja  $A$  um subconjunto de  $\rightarrow$  e  $s$  um subconjunto de  $S$ .

Já definimos a probabilidade de um evento  $s \rightarrow S$  e, agora, gostaríamos de estender essa definição e falar da probabilidade de um evento  $A \rightarrow \rightarrow$ . Ou seja, o nosso objetivo agora é definir probabilidades a partir de valores possíveis da variável aleatória, sem referência explícita aos pontos do espaço amostral que deram origem àqueles valores da variável aleatória.

Na verdade, o estudo de probabilidades começa a se aprofundar com a definição de variável aleatória, que será o objeto principal das nossas atenções a partir de agora. A noção de variável aleatória é tão importante que,

frequentemente, nem nos preocupamos com o que está “por trás” delas, isto é, na prática, muitas vezes, ignoramos o espaço amostral.

Como definir  $P(X \rightarrow A)$ ?

A maneira mais natural de fazer isso é associar a probabilidade do evento  $X \rightarrow A$  à probabilidade do evento  $S$  no espaço amostral  $S$ .

Ou seja, se  $A \rightarrow \rightarrow$ , definimos

$$P(X \rightarrow A) = P(S), \text{ onde } S = \{c \rightarrow S: X(c) \rightarrow A\}$$

Assim, a variável aleatória  $X$  é uma função que “transporta” a probabilidade de um espaço amostral  $S$  para um espaço  $\rightarrow$  de números reais.

**Exemplo 13:** Jogamos uma moeda duas vezes e estamos interessados no número de “caras” observado. O espaço amostral é

$$S = \{\text{onde } c = \text{CaCa, CaCo, CoCa, CoCo}\}.$$

Podemos definir uma variável aleatória  $X$  como

$$X(c) = \begin{cases} 0 & \text{sec} = \text{CaCa} \\ 1 & \text{sec} = \text{CaCo ou CoCa} \\ 2 & \text{sec} = \text{CoCo} \end{cases}$$

Ou seja,  $X$  é o número de caras nas duas jogadas. O espaço da variável aleatória  $X$  é  $\rightarrow = \{0, 1, 2\}$ , um subconjunto dos inteiros.

Seja  $A = \{x \rightarrow \rightarrow: x = 1\}$ . Como definir a probabilidade do evento  $A$ ? É só olhar para o subconjunto  $S$  do espaço amostral cujos elementos  $c$  são tais que  $X(c) \rightarrow A$ , ou seja,  $X(c) = 1$  aqui.

Nesse caso,  $S = \{\text{CaCo, CoCa}\}$ , o evento “uma cara em duas jogadas”, pois  $X\{\text{CaCo}\} = 1$  e  $X\{\text{CoCa}\} = 1$ . Assim,

$$P\{A\} = P\{X \rightarrow A\} = P\{S\} = P\{X=1\}$$

## 6.2 Definição de variável aleatória discreta e densidade de probabilidade discreta

Seja  $X$  uma variável aleatória cujo espaço é o conjunto unidimensional  $\rightarrow$ . Dizemos que  $X$  é uma variável aleatória discreta se o número de valores possíveis de  $X$  é finito ou contável.

Por exemplo, se os valores possíveis de  $X$  são  $\{0, 1, 2, \dots\}$  ou  $\{2, 4, 6, 8, \dots\}$ , ou ainda  $\{1/3, 1/4, 1/5, \dots, 1/n, \dots\}$ ,  $X$  é uma variável aleatória discreta. A partir da definição de variável aleatória discreta, podemos definir a densidade de probabilidade discreta, que é uma função da variável aleatória  $X$  que nos permite calcular probabilidades para todos os valores de  $X$ .

Seja  $f(x)$  uma função tal que

$$i) f(x) \geq 0 \forall x \in \mathcal{X}$$

$$\text{ii) } \sum_{x \in \mathcal{X}} f(x) = 1$$

$$\text{iii) } \forall A \subseteq \mathcal{X}, P(A) = P(X \in A) = \sum_{x \in A} f(x)$$

Ou seja, esta última propriedade nos permite calcular a probabilidade de qualquer evento envolvendo a variável aleatória  $X$ . Para qualquer evento  $A$  definido no espaço da variável aleatória  $X$ , a probabilidade de  $A$  é apenas o somatório de todos aqueles valores de  $x$  que compõem  $A$ .

$f(x)$  é chamada densidade de probabilidade da variável aleatória  $X$ , e dizemos que  $X$  é uma variável aleatória discreta.

### 6.3 Definição de variável aleatória contínua e densidade de probabilidade contínua

Seja  $X$  uma variável aleatória com espaço  $A$  tal que:

$$\text{i) } f(x) \geq 0 \forall x \in \mathcal{X}$$

ii)  $f(x)$  tem no máximo um número finito de descontinuidades em qualquer subintervalo finito de  $\rightarrow$ .

iii) Seja  $A \subseteq \rightarrow$ . A probabilidade de  $X \rightarrow A$  é:

$$P(X \in A) = P(A) = \int_A f(x) dx$$

A variável aleatória  $X$  é uma variável aleatória contínua e  $f(x)$  é sua densidade de probabilidade.

Seja  $X$  uma variável aleatória qualquer (contínua ou discreta). A função de distribuição da variável aleatória  $X$  é definida como

$$F(x) = P(X \leq x)$$

Seja  $f(x)$  a densidade de probabilidade de  $X$ , então

$$F(x) = \sum_{a \leq x} f(x)$$

Se  $X$  é variável aleatória discreta ou

$$F(x) = \int_{-\infty}^x f(x) dx$$

Se  $X$  é variável aleatória contínua.

### Propriedades da Função de Distribuição

- i)  $0 \leq F(x) \leq 1$  pois  $0 \leq \Pr(X \leq x) \leq 1$ .
- ii)  $F(x)$  é uma função não decrescente.
- iii)  $\lim_{x \rightarrow -\infty} F(x) = 0$
- iv)  $\lim_{x \rightarrow \infty} F(x) = 1$
- v) Se  $X$  é uma variável aleatória contínua, sua função de distribuição é contínua. Se  $X$  é discreta,  $F(x)$  é uma função contínua à direita, isto é, a função de distribuição apresenta “pulos” (descontinuidades) que só são “sentidos” quando nos aproximamos do ponto onde existe o “pulo” pela esquerda.

## 6.4 Relação entre a função densidade e a função de distribuição

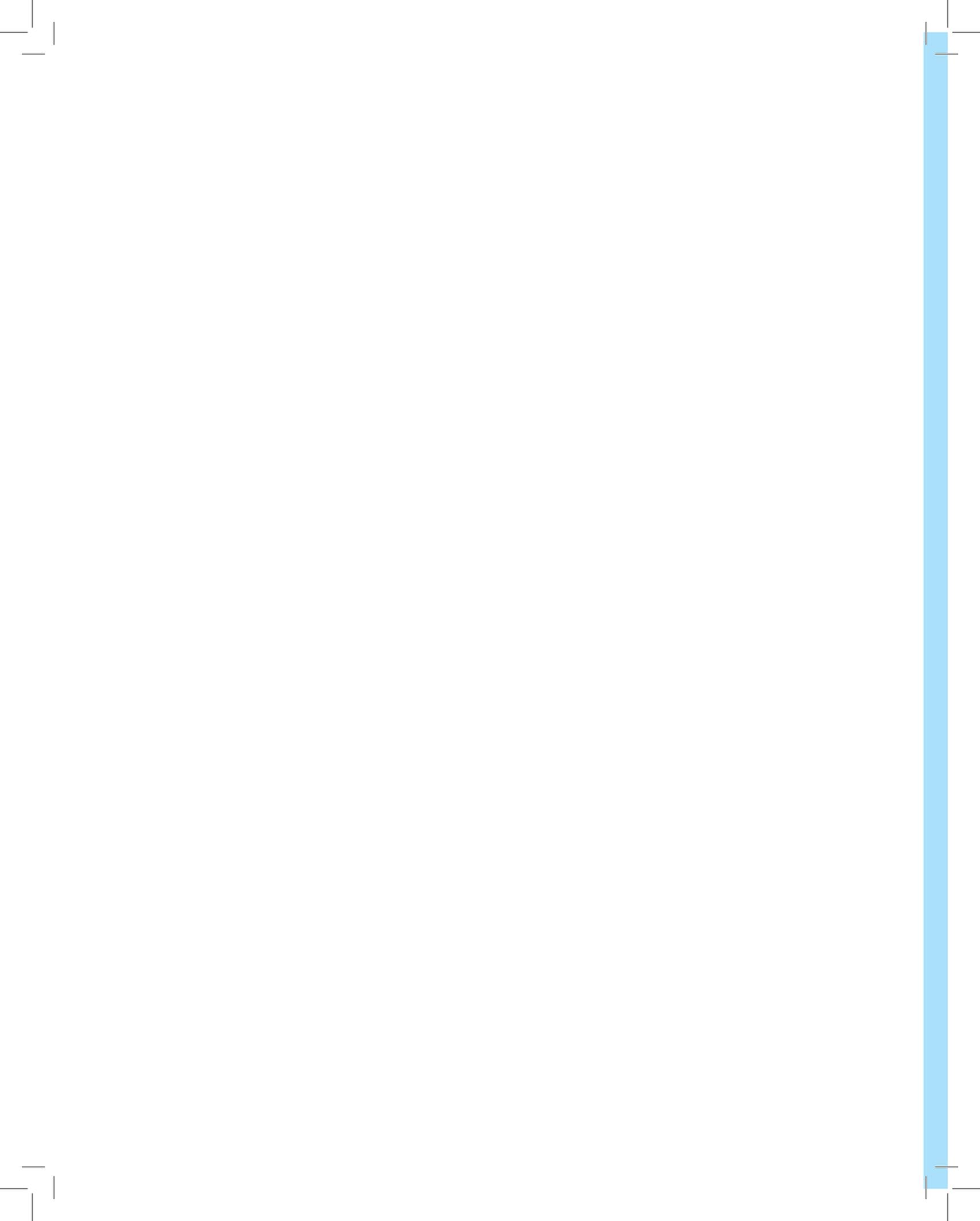
Considere uma variável aleatória contínua com densidade  $f(x)$  e função de distribuição  $F(x)$ . Então

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

$$\int_{-\infty}^b f(x) dx = F(x)$$

$$\frac{d}{dx} F(x) = f(x)$$

Ou seja, a densidade é a derivada da função de distribuição.



Capítulo

9

# Distribuições estatísticas



## Objetivo

- Apresentar as principais distribuições estatísticas discretas e contínuas.

## 1. Distribuições discretas

### 1.1 Distribuição binomial

A densidade binomial é uma das mais importantes em teoria da probabilidade. Ela surge como uma idealização matemática de diversas situações comuns na “vida real” e está intimamente ligada à amostragem com reposição. A situação clássica em que usamos uma densidade binomial é a seguinte.

Uma experiência tem apenas 2 resultados possíveis : “sucesso” e “falha”, em que a probabilidade de “sucesso” é  $p$  e a probabilidade de “falha” é  $q = 1 - p$ .

A experiência é repetida um número fixo ( $n$ ) de vezes, sempre nas mesmas condições, de tal forma que as probabilidades de “sucesso” ( $p$ ) e “falha” ( $q = 1 - p$ ) se mantêm inalteradas a cada repetição. As diversas repetições da experiência são feitas de maneira independente, ou seja, o resultado de uma repetição não afeta o resultado das outras.

A variável aleatória  $X$  que mede o número de “sucessos” nas  $n$  repetições da experiência é uma variável discreta, com valores possíveis  $0, 1, 2, \dots, n$ . Dizemos que esta variável tem densidade binomial com parâmetros  $n$  e  $p$ , e escrevemos que  $X \sim \text{Bin}(n, p)$ .

Nota : a escolha de um tipo de resultado como “sucesso” ou “falha” não implica em qualquer julgamento sobre o resultado ser “bom” ou “ruim”, é apenas uma questão de nomenclatura. Na verdade, a escolha do que é um “sucesso” ou “falha” depende da questão de interesse ao analisar o problema, de forma que o que é “sucesso” numa situação pode ser a “falha” num problema semelhante.

A equação da distribuição binomial é a seguinte.

$$P(X = x) = f(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

## 1.2 Distribuição hipergeométrica

Suponha que, na população, existem  $r$  objetos do tipo A (“sucessos”) e  $N - r$  objetos do tipo B (“falhas”).

Seja  $X$  o número de objetos do tipo A na amostra. Então as probabilidades dos diversos valores de  $X$  são dadas pela seguinte fórmula.

$$P(X = x) = f(x) = \frac{\binom{r}{x} \cdot \binom{N-r}{n-x}}{\binom{N}{n}},$$

onde  $x = 0, 1, 2, \dots, \min(n, r)$ .

Esta é a densidade hipergeométrica, usada para calcular probabilidades no caso de amostragem sem reposição.

Pode-se provar que  $f(x)$  acima definida integra a 1 (não é muito fácil).

## 1.3 Densidade Poisson

Esta densidade é usada principalmente para modelar o número de ocorrências de um evento “raro” (de probabilidade baixa) durante um intervalo de tempo especificado. Por exemplo, o número de acidentes numa estrada durante um fim de semana; o número de bactérias presentes numa solução após um certo período são, entre outros, eventos modelados pela distribuição de Poisson. Além disso, a distribuição de Poisson surge como um caso limite da distribuição  $\text{Bin}(n, p)$  quando  $n$  é grande, e  $p$  é pequeno (próximo de zero) e, neste contexto, é muito útil em aproximações numéricas.

A derivação da densidade Poisson pode ser feita de duas formas: a primeira está relacionada com o “processo de Poisson”, e a segunda surge como uma aproximação da densidade binomial.

## 1.4 Processo de Poisson

Considere uma sequência de eventos que ocorrem ao longo do tempo, como o número de carros vermelhos que param num sinal, o número de chamadas telefônicas que chegam a uma estação durante um certo intervalo de tempo.

Seja  $X_t$  o número de ocorrência no intervalo de tempo  $[0, t]$ . Claramente,  $X_t$  é uma variável aleatória discreta com valores possíveis  $0, 1, 2, \dots$ . Para derivar a densidade de  $X_t$ , partimos das seguintes premissas.

Seja  $\Delta t$  um intervalo de tempo pequeno. Então.

- A probabilidade de exatamente uma ocorrência em um intervalo de tempo  $\Delta t$  é aproximadamente  $k \Delta t$ .
- A probabilidade de exatamente zero ocorrências em um intervalo de tempo  $\Delta t$  é aproximadamente  $(1 - k)\Delta t$ .
- A probabilidade de duas ou mais ocorrências em um intervalo de tempo  $\Delta t$  é igual a um certo  $o(\Delta t)$ , onde  $o(\Delta t)/\Delta t$  tende a zero à medida que  $\Delta t$  tende a zero. Em outras palavras, a probabilidade de duas ou mais ocorrências em um intervalo de tempo  $\Delta t$  é um valor muito pequeno, e esse valor decresce a zero mais rapidamente que o comprimento do intervalo  $\Delta t$ .

Essas três premissas definem o tipo de processo que pode ser chamado de um processo de Poisson.

O parâmetro  $k$  acima é um número real  $> 0$ , chamado de **taxa média de ocorrência**.

Para cada instante  $t > 0$ , seja

$$P(X_t = x) = p_x(t), \text{ onde } x = 0, 1, 2, \dots$$

Fixando um instante qualquer  $t$  e aplicando a segunda premissa, nos dá

$$p_1(t + \Delta t) \cong [1 - k\Delta t]p_0(t)$$

Subtraindo  $p_0(t)$  de ambos os lados e dividindo o resultado por  $\Delta t$ , leva a

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} \cong kp_0(t)$$

Tomando-se o limite desta última expressão quando  $\Delta t$  tende a zero, encontramos, do lado esquerdo, a derivada de  $p_0(t)$ . Isso nos dá a equação diferencial

$$p_0'(t) = -kp_0(t)$$

Para  $x > 0$ , pode-se provar que as premissas resultam no seguinte sistema de equações diferenciais

$$p_x'(t) = -kp_x(t) + kp_{x-1}(t), \text{ onde } x = 1, 2, 3, \dots$$

A solução do sistema dado por (4) e (5) é

$$p_x(t) = \frac{(kt)^x e^{-kt}}{x!}, \text{ onde } x = 1, 2, 3, \dots$$

Para qualquer intervalo  $[0, t]$ , se fixarmos  $t$  e fizermos  $\lambda = kt$ , a equação acima reduz-se a

$$P(X = x) = f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ onde } x = 1, 2, 3, \dots$$

A densidade anterior é a densidade Poisson com parâmetro  $\lambda > 0$  e escrevemos

$$X \sim \text{Poisson}(\lambda).$$

### 1.5 Como usar a densidade Poisson na prática

Selecione um intervalo de tempo fixo. Conte o número de ocorrência de um certo evento de interesse neste intervalo. Esse número de ocorrências é uma variável discreta com valores possíveis 0, 1, 2, .... Se o evento é tal que a probabilidade do número de ocorrências no intervalo ser 0 ou 1 é “grande”, então o evento pode ser na prática modelado pela distribuição de Poisson. Uma densidade Poisson modela bem eventos “raros”, isto é, que não acontecem com grande frequência para qualquer intervalo de tempo fixo.

Por exemplo, o número de automóveis Corsa que entram num estacionamento no Rio de Janeiro num intervalo de 1 hora certamente não é uma variável Poisson, mas o número de Ferraris que entram no estacionamento no mesmo período de tempo deve ser Poisson.

## 2. Distribuições contínuas

### 2.1 Densidade Uniforme

A densidade uniforme serve para modelar o seguinte fenômeno: “escolhe-se um número aleatoriamente num intervalo dado”, por exemplo, o intervalo (0,1). A função “random”, presente na maioria das linguagens de computador, nada mais é do que um mecanismo para gerar números com distribuição uniforme no intervalo (0,1).

Uma variável aleatória  $X$  tem densidade uniforme no intervalo  $(a,b)$ , e escrevemos  $X \sim \text{Unif}(a,b)$  se a sua densidade é

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

Uma variável aleatória  $X$  com densidade uniforme no intervalo  $(a,b)$  tem a seguinte propriedade: qualquer subintervalo de comprimento  $d$  localizado dentro do intervalo  $(a,b)$  tem a mesma probabilidade.

A função de distribuição de uma variável aleatória  $\text{Unif}(a,b)$  é:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}$$

## 2.2 Densidade Exponencial

Uma variável aleatória com densidade exponencial é usada para modelar tempos de duração de equipamentos. Na verdade, existem densidades mais apropriadas para modelar esse fenômeno, pois, como veremos mais tarde, a densidade exponencial não leva em conta o desgaste do equipamento ao longo do tempo. A densidade exponencial é definida para variáveis contínuas e maiores que zero e depende de um parâmetro positivo,  $\lambda$ .

Notação:  $X \sim \text{Expo}(\lambda)$

A densidade exponencial é dada pela fórmula:

$$f(x) = \lambda e^{-\lambda x}, x \geq 0$$

E  $f(x) = 0$  se  $x < 0$ . Note que  $\lambda$  é  $> 0$  sempre.

A função de distribuição é dada por

$$F(x) = \begin{cases} 0, & x < 0 \\ \int_0^x \lambda e^{-\lambda u} du = -e^{-\lambda u} \Big|_0^x = 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

Note que o limite da função de distribuição quando  $x$  tende a  $+\infty$  é um.

O próximo gráfico apresenta as densidades exponenciais com parâmetros  $\lambda = 2, 4$  e  $8$ . Note que a densidade decai mais rápido quando  $\lambda$  é grande.

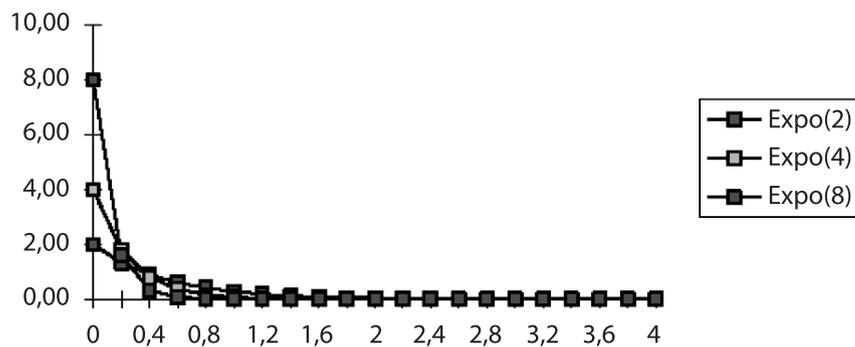


Figura 5 - Densidades exponenciais

O próximo gráfico exibe a função de distribuição de uma variável aleatória com parâmetros  $\lambda = 1$  e  $\lambda = 2$ .

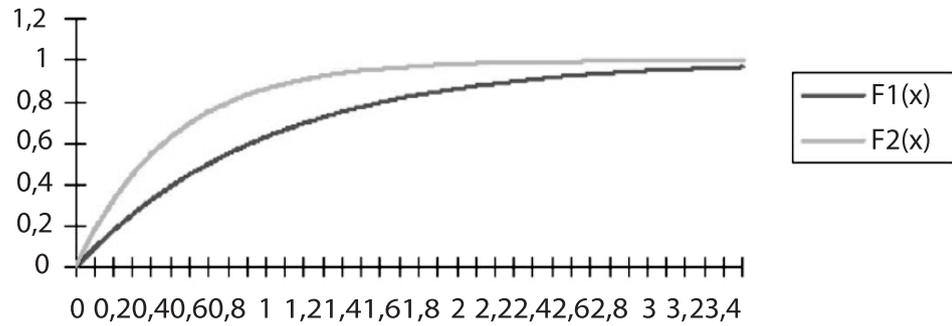


Figura 6 - Funções de distribuição densidades Expo(1) e Expo(2)

### 2.3 Função Gama

Seja  $a$  um número real maior que zero, não necessariamente inteiro. A função Gama com argumento  $a$  é definida por

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt$$

#### a) Propriedades da Função Gama

$$1) \Gamma(n) = (n-1)\Gamma(n-1) \text{ para } n > 1$$

A demonstração deste fato usa integração por partes.

$$2) \Gamma(n) = (n-1)! \text{ se } n \text{ é inteiro } > 1$$

$$3) \Gamma(1) = 0! = 1$$

$$4) \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

#### b) Densidade Gama

Seja  $X$  uma variável aleatória contínua definida no intervalo  $(0, \infty)$ . Dizemos que  $X$  tem densidade Gama com parâmetros  $\alpha$  e  $\beta$ , e escrevemos  $X \sim \text{Gama}(\alpha, \beta)$  se a densidade de  $X$  é

$$f(x) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}, & \text{onde } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

Os parâmetros  $a$  e  $\beta$  são números reais positivos  $a$  é conhecido como parâmetro de forma, e  $\beta$  é o parâmetro de escala.

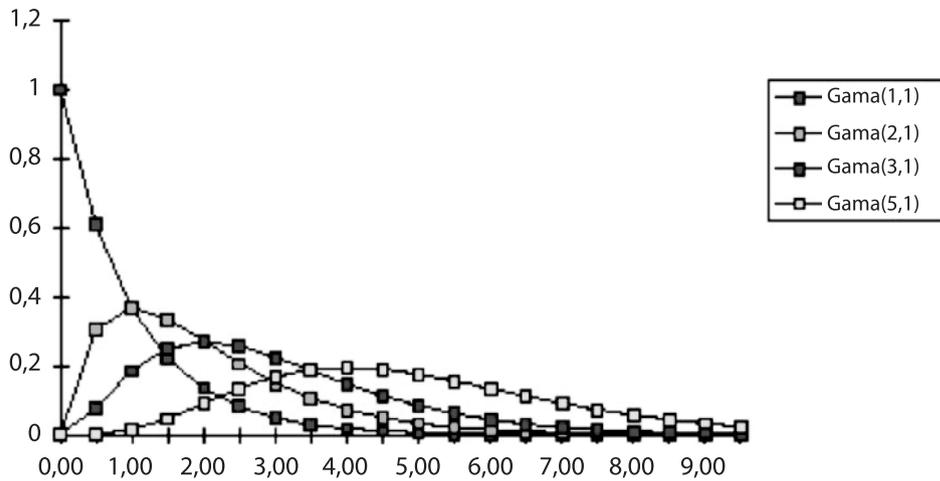


Figura 7 - Densidades Gama: gama(1,1), Gama(1,2), Gama(1,3) e Gama(1,5)

## 2.4 Densidade quiquadrado com k graus de liberdade)

Seja  $X$  uma variável aleatória contínua e positiva com densidade dada por

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{k}{2}\right)} x^{\frac{n-2}{2}} e^{-\frac{x}{2}} \text{ onde } x > 0$$

Tem densidade quiquadrado com  $(n - 1)$  graus de liberdade, e escrevemos:

$$X \sim \chi_k^2$$

A densidade quiquadrado com  $k$  graus de liberdade é apenas um caso particular da densidade gama. Na verdade,

$$\chi_k^2 = \text{Gama}\left(\alpha = \frac{k}{2}, \beta = \frac{1}{2}\right)$$

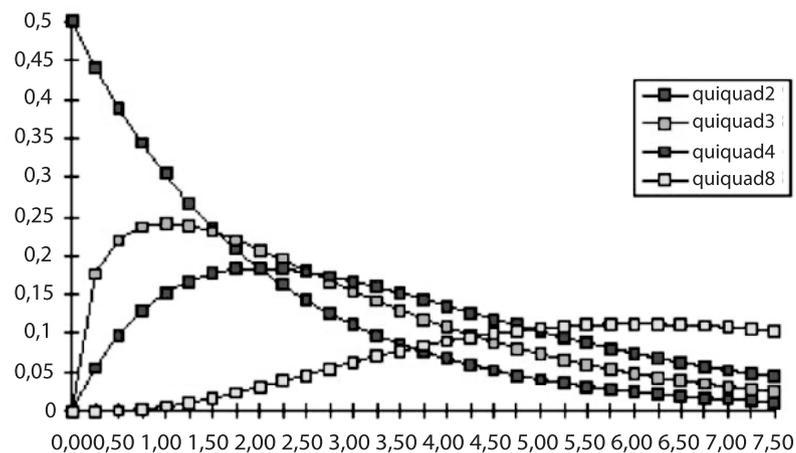


Figura 8 - Distribuições qui-quadrado com 2, 3, 4 e 8 graus de liberdade.

## 2.5 Distribuição Normal (Gaussiana)

A distribuição normal é, talvez, a mais importante das distribuições de probabilidade. Erros de mensuração de fenômenos físicos ou econômicos são frequentemente modelados pela distribuição normal, mas esta não é a única aplicação desta densidade. Por exemplo, a distribuição dos pesos, alturas e QI's das pessoas numa população também já foram modelados com sucesso por esta distribuição. A distribuição normal tem a forma de um sino e possui dois parâmetros,  $\mu$  e  $\sigma^2$ .

A distribuição normal é também chamada de gaussiana em homenagem ao matemático Carl Friedrich Gauss (1777-1855), que a utilizou pela primeira vez na modelagem de erros de medida. A distribuição normal também funciona como uma boa aproximação para outras densidades. Por exemplo, sob algumas condições pode-se provar que a densidade binomial pode ser aproximada pela normal.

### a) Densidade Normal com média $\mu$ e variância $\sigma^2$

Seja  $X$  uma variável aleatória contínua definida nos números reais. Dizemos que  $X$  tem densidade normal com média  $\mu$  e variância  $\sigma^2$  se a densidade de  $X$  é

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

#### Notação: $X \sim N(\mu, \sigma^2)$

Note que o segundo parâmetro ( $\sigma^2$ ) nesta notação é a variância de  $X$ . A seguir, exibimos gráfico das distribuições normais com média zero e variâncias 1, 2 e 4.

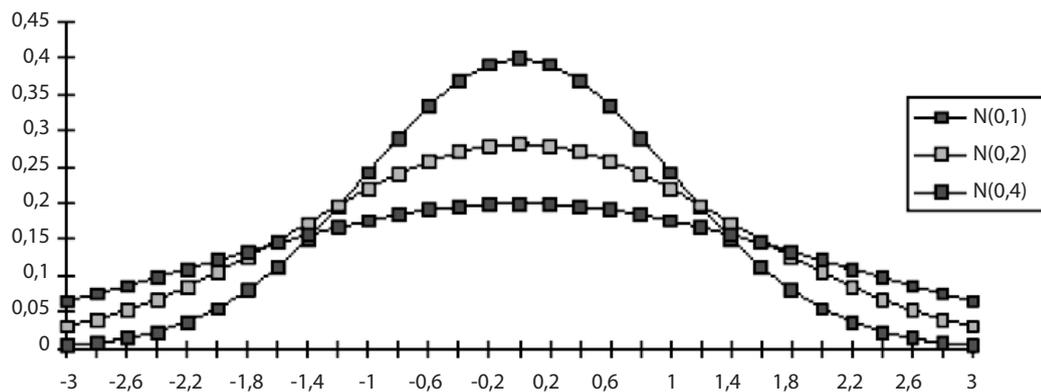


Figura 9 - Distribuições normais com média zero e variâncias 1, 2 e 4

Note que o máximo das densidades é encontrado quando  $x = 0$ , isto é, quando  $x$  é igual à média da distribuição. Isto vale para qualquer distribuição normal; o máximo de  $f(x)$  é obtido fazendo-se  $x = \mu$ , onde  $m$  é a média da normal. Também, quanto maior o valor da variância  $\sigma^2$ , mais "espalhada" é a distribuição.

## b) Propriedades da Distribuição Normal

1.  $f(x)$  dada pela expressão acima, integra a 1.
2.  $f(x) \geq 0$  sempre.
3. Os limites de  $f(x)$  quando  $x$  tende a  $+\infty$  e  $-\infty$ , são iguais a zero.
4. A densidade  $N(\mu, \sigma^2)$  é **simétrica em torno de  $\mu$** , ou seja,  
 $f(\mu + x) = f(\mu - x)$ .
5. O valor máximo de  $f(x)$  ocorre em  $x = \mu$ .
6. Os pontos de inflexão de  $f(x)$  são  $x = \mu + \sigma$  e  $x = \mu - \sigma$ .

## 3. Momentos de uma distribuição de probabilidade

A seguir, definimos alguns dos momentos de distribuições de probabilidade contínuas e discretas. Momentos são quantidades que nos dão uma idéia da tendência central, dispersão e assimetria de uma densidade de probabilidades.

### 3.1 Definição (média e variância)

A média (ou valor esperado, primeiro momento de uma variável aleatória) é definida como

$$\mu = E(X) = \begin{cases} \sum_x xf(x), \text{ caso discreto} \\ \int_{-\infty}^{+\infty} xf(x)dx, \text{ caso contínuo} \end{cases},$$

onde o somatório refere-se a todos os valores de  $X$  quando  $X$  é uma variável discreta. Quando  $X$  é uma variável contínua a média é calculada pela integral anterior, onde  $f(x)$  representa a densidade de probabilidade da variável  $X$ .

A média de uma variável aleatória representa uma medida de tendência central da distribuição de probabilidade dessa variável aleatória.

A variância de uma variável aleatória é uma medida da dispersão da distribuição de probabilidade, definida como

$$\sigma^2 = V(X) = \begin{cases} \sum_x (x - \mu)^2 f(x), \text{ caso discreto} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx, \text{ caso contínuo} \end{cases},$$

onde novamente  $f(x)$  representa a densidade de probabilidade (discreta ou contínua) da variável aleatória  $X$  e  $\mu$  é a média da variável aleatória. A variância é o segundo momento em torno da média, e corresponde ao momento de inércia em Mecânica.

Da própria definição segue que a variância é uma quantidade sempre maior ou igual a zero.

### 3.2 Desvio padrão

O desvio padrão de uma variável aleatória é a raiz quadrada positiva da sua variância, e denotado por  $\sigma$ .

O desvio padrão é expresso nas mesmas unidades que a variável aleatória, e a variância é dada nas unidades da variável aleatória ao quadrado. Logo, se a variável aleatória é medida em metros, o desvio padrão também está em metros, e a variância, em metros quadrados. Um valor pequeno do desvio padrão indica que existe pouca dispersão em torno da média. Se o desvio padrão é grande, os valores da variável aleatória estão muito dispersos em torno da média.

A média e a variância são casos particulares do que chamamos de "momentos" de uma distribuição de probabilidade. Os momentos de uma distribuição servem para caracterizar esta distribuição não apenas no que se refere à sua centralidade e sua dispersão, mas também com relação a outras características, como a simetria ou assimetria da densidade de probabilidade.

### 3.3 Késimo momento

O  $k$ ésimo momento da variável aleatória  $X$  é definido como

$$E(X^k) = \begin{cases} \sum_x x^k f(x), \text{ caso discreto} \\ \int_{-\infty}^{+\infty} x^k f(x) dx, \text{ caso contínuo} \end{cases}$$

Onde  $k = 1, 2, 3, \dots$

Obviamente, a definição não se aplica se algum dos  $E(X^k)$  é infinito.

**Definição:**  $k$ ésimo momento central ou  $k$ ésimo momento em torno da média)

O  $k$ ésimo momento central da variável aleatória  $X$  é definido como

$$E(X^k) = \begin{cases} \sum_x (x - \mu)^k f(x), \text{ caso discreto} \\ \int_{-\infty}^{+\infty} (x - \mu)^k f(x) dx, \text{ caso contínuo} \end{cases},$$

onde  $k = 1, 2, 3, \dots$

Logo, a média e a variância são apenas casos particulares de momentos. A média é o primeiro momento (isto é,  $\mu = E(X)$ ) e a variância é o segundo momento central, ou seja,  $E(X - \mu)^2$ .

A notação  $E(\dots)$  indica um valor esperado e pode ser estendida para funções mais gerais que  $X^k$  ou  $(X - \mu)^k$ .

### 3.4 Valor esperado de uma função de uma variável aleatória

Seja  $X$  uma variável aleatória com densidade  $f(x)$ , e seja  $u(X)$  uma função qualquer tal que

$$E[u(X)] = \begin{cases} \sum_x u(X)f(x), & \text{caso discreto} \\ \int_{-\infty}^{+\infty} u(x)dx, & \text{caso contínuo} \end{cases}$$

#### Formula alternativa para o cálculo da variância

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

Essa fórmula é válida para qualquer variável aleatória  $X$  (contínua ou discreta), desde que a média de  $X$  seja finita.

## Referências



BARBETA, P. A., REIS, M. M., BORNIA, A. C. **Estatística para Cursos de Engenharia e Informática**, São Paulo: Atlas, 2004.

BOLFARINE, Heleno & BUSSAB, Wilton O. **Elementos de Amostragem**. 1. ed. São Paulo: Edgard Blücher, 2005.

CONOVER, W J. **Practical Nonparametric Statistics**. 3. ed. New York: John Wiley, 1998.

FELLER, W. **Teoria das probabilidades e suas aplicações**. São Paulo, Edgard Blucher, 1976

MENDENHALL, William. **Probabilidade e Estatística**. São Paulo: Campus. 1985.

MIRSHAWKA, V. **Probabilidade e Estatística para Engenharia**, São Paulo: Nobel, 1978.

MORETTIN, L. G. **Estatística Básica Probabilidade**. São Paulo: Makron Books, 1993.

SPIEGEL MR. **Estatística**. Coleção Schaum. São Paulo: Editora Afiliada, 1993.

## Sobre os autores

**Jorge Luiz de Castro e Silva:** doutor em Ciência da Computação pela Universidade Federal de Pernambuco (2004), mestre em Ciência da Computação pela Universidade Federal do Ceará (1997), graduado em Tecnologia em Processamento de Dados pela Universidade Federal do Ceará (1978) e Bacharel em Administração pela Universidade Estadual do Ceará (1981). Professor Adjunto da Universidade Estadual do Ceará e Professor Orientador do Programa de Pós-Graduação (mestrado) em Ciências da Computação. Tem experiência na área de computação, com ênfase em Modelagem Analítica e Simulação, atuando principalmente nos seguintes temas: engenharia de tráfego de redes, tráfego auto-similar, metaheurísticas para roteamento de redes, modelagem matemática e reconhecimento de padrões em tráfego de redes.

**Maria Wilda Fernandes:** possui especialização em Administração de Sistemas para a Internet. Graduada em Licenciatura em Matemática pela Universidade Estadual do Ceará (1987) e Bacharel em Ciências da Computação pela Universidade Estadual do Ceará (1991). Atualmente, é servidora técnico-administrativa da Universidade Estadual do Ceará, exercendo a função de gerente da área de desenvolvimento de TI da UECE e tutora a distância do curso de Licenciatura Plena em Computação da Universidade Estadual do Ceará.

**Rosa Lívia Freitas de Almeida:** possui graduação em Engenharia Elétrica pela Universidade de Fortaleza (1981), especialização em Análise de Sistemas pela PUC - Rio de Janeiro (1989), mestre em Saúde Pública pela Universidade Federal do Ceará (2001). Foi professora adjunta na Faculdade Integrada do Ceará durante sete anos nos cursos de Graduação de Sistemas de Informação, Graduação em Fisioterapia, Tecnólogo em Gestão Hospitalar e nos cursos de especialização. Entre outras disciplinas, ministrou Probabilidade e Estatística e Planejamento Estatístico de Experimentos. Atua como Consultora em Tecnologia da Informação em Saúde, gerência de projetos de informática, operacionalização e suporte em Ambientes Virtuais de Aprendizagens (AVA) e tutoria em cursos de Ensino a Distância. É especialista em georeferenciamento e estatística espacial, análise de sistemas, desenvolvimento e implantação de sistema de informação.