



Magno Alves de Oliveira
(Taioabeiras-MG), matemático probabilista, licenciado pela Universidade Federal de Viçosa (UFV) e pós-graduado pela Universidade de Brasília (UnB), tem interesse pelas questões humanas, especialmente as da educação.

Acredita que a transformação de uma vida como a dele, por meio de uma educação de qualidade, não deve ser a exceção mas a regra. Por isso, atua como docente do Instituto Federal de Brasília (IFB).



Magno Alves de Oliveira

Probabilidade e estatística:
um curso introdutório

**Série Novos Autores
da Educação Profissional
e Tecnológica**

PROBABILIDADE E ESTATÍSTICA: UM CURSO INTRODUTÓRIO

Magno Alves de Oliveira

Este livro resulta de um trabalho desenvolvido nos cursos técnicos de Agropecuária e Agroindústria no Instituto Federal de Brasília (IFB)

– Campus Planaltina, durante o período de 2008 a 2010. É o resultado de um esforço para dotar a disciplina Matemática e Estatística Instrumental de um caráter técnico, mais ajustado às necessidades específicas daqueles cursos, e mais libertador para os interessados em aprofundar seus estudos em cursos de níveis mais elevados.

Ao mesmo tempo que não são exigidos maiores pré-requisitos do leitor, a obra busca contribuir para o desenvolvimento de competências gerais, como leitura, interpretação e representação esquemática de situações, organização lógica de ideias e comunicação formal, bem como para o desenvolvimento de habilidades e técnicas específicas, como procedimentos para levantamento, classificação e tratamento de informações, com vistas a auxiliar o futuro profissional na tomada de decisões, nos mais variados contextos, a partir de critérios objetivos e imparciais.

Por seu caráter geral, apesar de técnico, acredito que seu conteúdo possa ser utilizado em diversos cursos técnicos na modalidade subsequente, já no primeiro semestre, e também em cursos superiores, especialmente os tecnológicos, como introdução a disciplinas relacionadas com as áreas de probabilidade e estatística.



**PROBABILIDADE
E ESTATÍSTICA:
UM CURSO
INTRODUTÓRIO**

Presidência da República Federativa do Brasil

Ministério da Educação

Secretaria de Educação Profissional e Tecnológica

Instituto Federal de Brasília

Secretaria de Educação Profissional e Tecnológica (Setec)

Luiz Augusto Caldas Pereira Ex- Diretor da Diretoria de Formulação de Políticas da Educação Profissional e Tecnológica

Mariângela de Araujo Póvoas Pereira Assessora da Diretoria de Formulação de Políticas da Educação Profissional e Tecnológica

Viviani Pereira A. Guimarães Especialista em Gestão e Planejamento Educacional

Thais Silva Almeida Agente Administrativo

Flávia Caetano Técnica em Assuntos Educacionais

Série Novos Autores da Educação Profissional e Tecnológica

Magno Alves de Oliveira

**Probabilidade
e estatística:
um curso
introdutório**

Brasília-DF
2011

Ministério da
Educação



Avaliadores da Série "Novos Autores da Educação Profissional e Tecnológica"

Fabiano Oliveira Antonini
Fátima Lucília Vidal Rodrigues
Francisco de Assis Póvoas Pereira
Glória Marinho
Lilian Páscoa
Marco Antônio Botelho Soares
Mônica Braga Marçal
Nelma Mirim Chagas de Araújo
Otávio Fernandes Lima da Rocha
Patrícia Silva Santiago
Paulo Maria Ferreira
Robson Bastos Roen
Rogério José Câmara
Rossana Barros Silveira



Conselho Editorial
Edilsa Rosa da Silva
Elisa Raquel Gomes de Sousa
Josué de Sousa Mendes
Paulo Henrique de Azevedo Leão
Philippe Tshimanga Kabutakapua
Veruska Ribeiro Machado

Presidente do Conselho Editorial Paulo Henrique de Azevedo Leão
Secretária Executiva Veruska Ribeiro Machado
Coordenador de Publicações Paulo Henrique de Azevedo Leão
Editores Paulo Henrique de Azevedo Leão e Sandra Maria Branchine
Projeto gráfico e capa Marcos Hartwich
Assessoria editorial Jair Santana Moraes
Diagramação Raphael Caron Freitas
Tiragem 2.000 exemplares

O livro Probabilidade e estatística: um curso introdutório é uma publicação da Editora IFB
e compõe a série "Novos Autores da Educação Profissional e Tecnológica", da Secretaria de Educação Profissional
e Tecnológica (Setec) do Ministério da Educação.

Dados internacionais de Catalogação na Publicação (CIP)

048p

Oliveira, Magno Alves de.
Probabilidade e estatística: um curso introdutório / Magno
Alves de Oliveira – Brasília: Editora IFB, 2011.
166 p. : il. ; 24 cm – (Série Novos Autores da Educação
Profissional e Tecnológica)

ISBN 978-85-64124-07-3
Inclui bibliografia.

1. Probabilidade. 2. Estatística. I. Título.

CDU 519.21



Agradecimentos

Agradeço o incentivo da professora Ivone Moreyra e dedico este trabalho aos estudantes do Campus Planaltina do Instituto Federal de Brasília (IFB), onde tudo começou.



Sumário

Apresentação da Setec/MEC 11
Introdução 13

■ **CAPÍTULO I 17**

UMA INTRODUÇÃO À TEORIA DE CONJUNTOS 19

1. O conjunto das partes de um conjunto 19
2. Operações com conjuntos 25
3. Conjuntos numéricos 28
 - 3.1. Uma interpretação para frações 28
 - 3.2. Dízimas periódicas 31
 - 3.3. Proporcionalidade 33
 - 3.4. Regra de três 37

Exercícios 41

■ **CAPÍTULO II 55**

NOÇÕES DE CONTAGEM 57

1. Os princípios fundamentais da contagem 58

2. Contando elementos no conjunto das partes de um conjunto com cardinalidade finita 65
 3. Números combinatórios 69
- Exercícios 73

■ CAPÍTULO III 77**ANÁLISE EXPLORATÓRIA DE DADOS 79**

1. Medidas resumo 80
 - 1.1. Medidas de posição 80
 - 1.2. Medidas de dispersão 85
 2. Gráficos 91
- Exercícios 96

■ CAPÍTULO IV 103**ESTATÍSTICA DESCRIPTIVA 105**

1. Conceitos básicos 106
 2. Amostragem 109
 - 2.1. Técnicas de amostragem 110
 3. Aleatoriedade 114
 - 3.1. Noção intuitiva de variável aleatória 114
 4. Técnicas da Estatística descritiva 116
 - 4.1. Análise da frequência 116
 - 4.2. Gráficos 120
 - 4.3. Histogramas e distribuição da frequência 122
 - 4.4. Outros tipos de gráficos 125
 5. Resumo das etapas do trabalho estatístico 129
- Exercícios 131

■ CAPÍTULO V 135**NOÇÕES DE PROBABILIDADE 137**

1. Alguns conceitos básicos 138
2. Medidas de probabilidade 139
 - 2.1. O método clássico 140
 - 2.2. O método frequentista 141
 - 2.3. O método bayesiano 142
 - 2.4. Definição de probabilidade 142
3. Propriedades operatórias 145
 - 3.1. Monotonicidade da probabilidade 145

- 3.2. Probabilidade de não ocorrência de um evento 145
- 3.3. Probabilidade da união qualquer de eventos 147
- 4. Probabilidade conjunta e dependência 149
 - 4.1. Dependência 150
 - 4.2. Probabilidade condicional 151
 - 4.3. Teorema de Bayes 154
- Exercícios 156

■ REFERÊNCIAS BIBLIOGRÁFICAS 163



Apresentação da Setec/MEC

A série de publicações "Novos Autores da Educação Profissional e Tecnológica" é parte de um conjunto extenso de ações encaminhadas pela Secretaria de Educação Profissional e Tecnológica (Setec/MEC) que demarcam um período de valorização da Educação Profissional e Tecnológica (EPT). A concepção das atuais políticas afirma-se na defesa de uma EPT que não encerre o sentido de qualificação para o trabalho no limite do que é a atividade, mas sim do que se deve explicar pelas suas finalidades e pelos seus valores, ou seja, não no que se circunscreve ao técnico, mas sim à existência humana e à vida.

A extensão e a diversidade das medidas em curso, que se caracterizam por sua natureza sistêmica, são reveladoras da face deliberada de uma política que visa superar a compreensão reducionista e fragmentada de Educação Profissional e Tecnológica, em prol de uma visão em que parte e todo são inseparáveis. Nesse sentido, é fundamental observar que a expansão da oferta, com a criação de novas unidades públicas (Expansão das Redes Federal e Estaduais), e o estabelecimento de novos referenciais legais, normativos e pedagógicos só se configurarão em

objetivos verdadeiramente alcançados quando o nexo com a prática docente e a “sala de aula” traduzir-se em uma qualificação para o trabalho pautada pelo compromisso com a formação integral e cidadã.

Desse modo, é importante que os processos de formação para o exercício de profissões técnicas e tecnológicas não adotem abordagens que descoleem os “conhecimentos específicos” de seus fundamentos científicos e teóricos e não descontextualizem os saberes aplicados ao trabalho negando-lhes os determinantes sociais e humanos associados a sua produção, sob pena de uma formação meramente funcionalista e instrumentalizadora. Assim, nesse contexto – apesar do desenvolvimento das novas tecnologias e de textos digitais, e das contradições e dos debates que o cercam –, é precipitado negar a importância do livro como suporte para o desenvolvimento dos processos de ensino-aprendizagem, sem, contudo, negar as alternativas de que o professor dispõe para o exercício da autonomia e da liberdade no exercício da docência.

Por fim, a série “Novos Autores da Educação Profissional e Tecnológica” contempla entre seus objetivos a preocupação da Setec/MEC em inventariar e democratizar a reconhecida produção de material destinado ao conteúdo específico dos cursos de formação profissional produzidos por professores da Rede Federal de Educação Profissional Científica e Tecnológica.

Desejo aos professores e alunos proveitoso uso desta publicação.

Eliezer Moreira Pacheco
Secretário da Setec/MEC



Introdução

Fenômenos importantes relacionados às dinâmicas tectônica e climática do planeta Terra voltaram a ocupar lugar central na preocupação da população mundial. Para citar dois deles, o terremoto que assolou o Haiti, em janeiro de 2010, e os recentes recordes pluviométricos que assombraram a população brasileira no verão do mesmo ano. Além das perdas humanas, esses acontecimentos trazem enormes perdas materiais, seja na estrutura das cidades ou na produção agropecuária.

Fenômenos como esses, ou como o sobe e desce das bolsas de valores, que funcionam como termômetros da economia global, fazem crescer o papel da matemática como disciplina acadêmica e, sobretudo, vêm ressignificá-la. Isso porque é evidente que a sociedade precisa aprimorar os seus métodos de previsão de fenômenos tão importantes, para tentar minorar os seus impactos e, nesse aspecto, a matemática é uma importante ferramenta para o desenvolvimento científico e tecnológico, e o seu desenvolvimento implica o desenvolvimento da ciência e da tecnologia como um todo.

Todavia, o caráter aleatório de fenômenos como esses exige o desenvolvimento e o aprofundamento em áreas da matemática que se dedicam ao estudo de padrões de comportamento, registros, armazenamento e processamento de informações e de cálculos de probabilidades.

A sociedade já possui um conhecimento acumulado considerável, e novas teorias, técnicas e tecnologias têm sido desenvolvidas continuamente. É preciso garantir o crescimento desse conhecimento além da sua sustentação, para que possamos obter previsões cada vez mais razoáveis e vejamos capazes de nos prevenir contra possíveis danos.

Com vistas a esse futuro, reunimos neste livro/texto informações essenciais para que o estudante dê os primeiros passos no estudo da matemática, em especial, no que diz respeito à probabilidade e à estatística.

Ao longo deste trabalho, alguns conceitos serão tratados de forma bem intuitiva, sem tratamento matemático formal. Lembramos que o principal objetivo deste livro/texto é familiarizar o leitor com alguns aspectos importantes que fundamentam o estudo da estatística e da probabilidade e que, muitas vezes, são tratados de forma rasa na educação básica. Como consequência desse tratamento, surgem entraves que impedem a compreensão de conceitos mais robustos.

Em negrito, encontram-se destacados os termos que possuem significado matemático preciso. O formalismo do significado de muitos desses termos não será tratado aqui, mas sempre será apresentada uma idéia intuitiva para o leitor. Todavia, essa imprecisão não deve comprometer a leitura. Ao contrário, pretende ampliar a capacidade do leitor de situar os objetos aqui estudados na teoria aprendida no Ensino Médio, além de preparar o terreno para estudos mais aprofundados em estatística e probabilidade.

No Capítulo 1, apresentamos uma discussão sobre conjuntos e as principais relações que eles travam entre si. Como exemplo, introduzimos o conjunto dos números reais e damos atenção especial ao conjunto dos números racionais.

No Capítulo 2, ocupamo-nos da tarefa de contar o número de elementos de alguns conjuntos. Essa tarefa não é tão simples, uma vez que envolve a habilidade de classificação.

No Capítulo 3, associamos aos conjuntos diversas medidas descritoras do comportamento dos seus elementos. Essas medidas servem de critério para a comparação de conjuntos distintos. Em particular, introduzimos o conceito de **gráfico** como uma dessas medidas.

No Capítulo 4, discutimos sobre a aleatoriedade e apresentamos a Estatística como a ciência que trata dos fenômenos aleatórios. Técnicas do trabalho estatístico são apresentadas, desde os cuidados para se efetuar uma coleta de dados até a comunicação dessas informações, devidamente tratadas.

No Capítulo 5, apresentamos um tipo de medida que, de alguma maneira, quantifica a incerteza inerente aos fenômenos aleatórios: a medida de probabilidade. Discutimos sua definição formal, suas principais propriedades e alguns teoremas que nos ajudarão a resolver alguns problemas recorrentes do cotidiano.

Ao final de cada capítulo, selecionamos alguns problemas, com graus de dificuldade variados, relacionados à teoria ali discutida. Muitos desses exercícios apareceram em processos seletivos ou concursos públicos recentes de alguns institutos federais.

Ao elaborar este livro/texto, nossa preocupação foi fundamentar os conceitos de **estatística** e **probabilidade**, em geral vistos de forma superficial e descolada dos demais conteúdos no Ensino Médio, adaptando-os a uma linguagem mais objetiva, própria para discussões em cursos de nível técnico ou tecnológico.

Ressaltamos o caráter introdutório dessas ideias e esperamos que a linguagem utilizada seja libertadora o suficiente para que o leitor ouse alçar novos voos por essa teoria.

CAPÍTULO I

UMA INTRODUÇÃO À TEORIA DE CONJUNTOS



Uma introdução à teoria de conjuntos

1 Uma introdução à teoria de conjuntos

Uma noção fundamental para o estudo que desenvolveremos ao longo deste trabalho é a noção de **conjunto**. Na verdade, é impossível desenvolver a ideia do que é um conjunto dissociada daquela do que é **elemento**. Chamamos de conjunto a qualquer coleção de objetos, coisas ou seres, ainda que vazia. Cada um desses objetos, coisas ou seres é um elemento do conjunto.

Dois conjuntos serão **iguais** se possuírem exatamente os mesmos elementos, e serão distintos se diferirem um do outro pela quantidade de elementos ou pela presença de, pelo menos, um elemento num deles que difere de todos os elementos do outro. Esse conceito será formalizado mais adiante.

Note que até mesmo pra falar de **conjunto vazio** precisamos falar de elemento, ou melhor, da sua ausência. Quando um conjunto é vazio, costumamos representá-lo como \emptyset ou $\{\}$.

Representaremos um conjunto qualquer ora por letras maiúsculas do alfabeto (A, B, C, \dots) ou por letras maiúsculas do alfabeto grego ($\Gamma, \Omega, \Delta, \dots$), ora listando seus elementos, quando isso for possível, ora por uma descrição inequívoca de seus elementos, ou simplesmente por meio de um esquema pictográfico. Já os elementos serão representados por letras minúsculas do nosso alfabeto (a, b, c, \dots), sempre que necessário.

Exemplo 1.1. A expressão

$$\{0, 1, 2, \dots\}$$

representa o **conjunto dos números naturais**. Outras maneiras de representar esse conjunto são

$$\mathbb{N}$$

ou

$$\{x; x \text{ é um número natural}\}.$$

Exemplo 1.2. O conjunto

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

representa o **conjunto dos números inteiros**.

Exemplo 1.3. O conjunto

$$\mathbb{Q} = \{x = \frac{m}{n}; m \in \mathbb{Z} \text{ e } n \in \mathbb{Z}^*\}$$

representa o **conjunto dos números racionais**.¹ Cada elemento x de \mathbb{Q} é dito ser uma fração, e toda fração admite uma representação decimal, resultado da divisão do seu numerador pelo seu denominador. Por exemplo,

¹ O símbolo * sobreescrito ao conjunto \mathbb{Z} significa que estamos excluindo o zero do conjunto, ou seja, $\mathbb{Z}^* = \{\pm 1, \pm 2, \pm 3, \dots\}$. Isso é devido ao fato de não se definir frações com denominador nulo.

$$\frac{1}{2} = 0,5 \text{ (representação decimal infinita)}$$

e

$$\frac{1}{3} = 0,333\dots \text{ (dízima periódica).}$$

Exemplo 1.4. Chamamos de **conjunto dos números reais** e denotamos por \mathbb{R} o conjunto dos números que possuem representação decimal finita ou infinita. Observe que, pelo exposto no Exemplo 1.3, todo número racional é um número real. No entanto, existem números com representação decimal infinita que não admitem representação na forma de fração. É o caso, por exemplo, do número $0,30300300030000\dots$. Números como esse também são chamados de **dízimas não periódicas** ou **números irracionais**.²

Se um conjunto A possui um elemento, digamos a , então dizemos que a pertence a A e denotamos essa condição por

$$a \in A. \quad (1.1)$$

Quando um elemento qualquer b não faz parte da coleção A , dizemos que b não pertence a A , e denotamos essa condição por

$$b \notin A. \quad (1.2)$$

As relações (1.1) e (1.2) são chamadas de **relações de pertinência**, pois dizem respeito à posição relativa entre um elemento e um conjunto.

Podemos, também, considerar a posição relativa entre dois conjuntos quaisquer. Para isso, introduzimos a noção de **subconjunto**.

² Existem mais números irracionais do que racionais. Para se convencer disso, para cada número racional existente, podemos gerar mais de um número irracional correspondente. Por exemplo, para o número racional $0,333\dots$, podemos gerar os números irracionais $0,30300300030000\dots$ e $0,31311311131111\dots$ e para o número racional $0,5$, podemos gerar $0,530300300030000\dots$ e $0,531311311131111\dots$

Definição 1.1. Dados dois conjuntos A e B , dizemos que A é subconjunto de B , se todo elemento de A é também elemento de B . Denotamos essa condição assim: $A \subset B$.

Mais formalmente,³ podemos reescrever a Definição 1.1 assim:

$$A \subset B \Leftrightarrow a \in A \Rightarrow a \in B. \quad (1.3)$$

A expressão $A \subset B$ pode ser lida de três maneiras distintas: *A é parte de B* ou *A está contido em B*. De qualquer maneira, o seu significado preciso está descrito pela sentença lógica (3).

Explorando um pouco mais a Definição 1.1, podemos concluir que para que um conjunto A não seja um subconjunto de um conjunto B basta que ele possua um elemento que não pertença a B . Denotaremos esta situação assim:

$$A \not\subset B. \quad (1.4)$$

Podemos, então, concluir que o conjunto vazio é subconjunto de qualquer⁴ conjunto, isto é,

$$\emptyset \subset A, \forall \text{ conjunto } A.$$

As relações (1.3) e (1.4), que definem as posições relativas entre dois conjuntos quaisquer, são também chamadas de **relações de inclusão**.

Neste ponto, podemos formalizar o conceito de **igualdade de conjuntos**:

Definição 1.2. Dados dois conjuntos A e B , dizemos que $A = B$ se, e somente se, $A \subset B$ e $B \subset A$.

Pela definição acima, é fácil perceber que $\{1,2\} = \{2,1\} = \{1,1,1,2,2\}$. Fica claro, também, que a terminologia que estamos usando para conjuntos visa representar a variedade da informação presente em uma coleção.

³ Proposições lógicas do tipo " $p \Rightarrow q$ " podem ser lidas assim: se p então q , ou ainda, p implica q . Já proposições do tipo " $p \Leftrightarrow q$ " podem ser lidas como p se, e somente se, q , ou ainda, p é equivalente a q , e significam que $p \Rightarrow q$ e $q \Rightarrow p$.

⁴ O símbolo \forall pode ser lido assim: *para todo* ou *qualquer que seja*.

Exemplo 1.5. Considere o conjunto

$$B = \{1, \{1,2\}, 3\}.$$

Note que B é uma coleção de 3 elementos. Em particular, temos que $\{1,2\} \in B$. No entanto, $\{1,2\} \not\subset B$, já que $2 \notin B$.

De volta aos exemplos 1.1, 1.2, 1.3 e 1.4, concluimos que $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

1.1 O conjunto das partes de um conjunto

Dado um conjunto qualquer B , vimos que ele admite, pelo menos, um subconjunto, que é o conjunto vazio. Se substituirmos A na Definição 1.1 pelo próprio B , concluímos que $B \subset B$. Assim, o conjunto vazio e o próprio conjunto B são subconjuntos de B , também referidos como sendo os seus **subconjuntos triviais**.

Um conjunto pode admitir **subconjuntos não triviais** (ou próprios). Por exemplo, considere os conjuntos

$$C = \{\text{estudantes do IFB}\}$$

e

$$D = \{\text{estudantes do curso Tecnólogo em Agroecologia do IFB}\}.$$

Temos que $D \neq \emptyset$ e $D \neq C$. Como $D \subset C$, temos que D é um subconjunto próprio de C .

Voltemos ao conjunto arbitrário B e pensemos em todos os seus subconjuntos. À coleção de todos esses objetos damos o nome de **Conjunto das partes** de B , que denotaremos por $\wp(B)$ ou 2^B .

Exemplo 1.6. Considere o conjunto $E = \{1,2\}$. Então $\wp(E) = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$.

No último exemplo, vemos que o conjunto E possui 2 elementos, enquanto o conjunto $\wp(E)$ possui 4 elementos. À quantidade de elementos de um dado conjunto B damos o nome de **cardinalidade** e, neste trabalho, será denotada assim: $\#B$. De volta ao Exemplo 1.6, temos que $\#E = 2$ e $\#\wp(E) = 4$.

Neste momento, podemos observar que existem conjuntos com cardinalidade infinita, isto é, conjuntos com uma quantidade infinita de elementos. Volte aos exemplos 1.1 e 1.3 e note que $\#N = \infty$ e $\#Q = \infty$. Isso não significa que N e Q tenham a mesma quantidade de elementos. De fato, é fácil observar que $N \subset Q$ e que $1/2 \in Q$ mas $1/2 \notin N$, o que mostra que existem mais números racionais do que naturais.

A cardinalidade de um conjunto é que o define como um conjunto finito ou como um **conjunto infinito**.

Exemplo 1.7. *O intervalo limitado real $I = (0, 1]$, isto é, o conjunto de todos os números reais maiores do que 0 e menores ou iguais a 1 é um conjunto infinito.*

No exemplo anterior, apresentamos o que chamamos de intervalo limitado real, que são subconjuntos do conjunto dos números reais R , que apresentam um dos seguintes formatos

$$(a, b) = \{x \in R; a < x < b\},$$

$$[a, b) = \{x \in R; a \leq x < b\},$$

$$(a, b] = \{x \in R; a < x \leq b\}$$

ou

$$[a, b] = \{x \in R; a \leq x \leq b\}.$$

Dado um intervalo limitado real, podemos imaginar infinitos números que satisfaçam a relação que o define. No entanto, diferentemente do caso de N , o conjunto dos números naturais, não podemos listar os seus elementos. Em outras palavras, podemos entender o conjunto N como uma lista ordenada de números, razão pela qual dizemos que todo conjunto que apresenta essa

característica é **enumerável**. Pense um pouco e veja que o conjunto I do Exemplo 1.7 é **não enumerável**, isto é, não apresenta essa característica.

Um intervalo real pode não ser limitado. Nesse caso, a relação que o define assume um dos formatos abaixo:

$$(a, \infty) = \{x \in \mathbb{R}; x > a\},$$

$$(-\infty, b) = \{x \in \mathbb{R}; x < b\},$$

$$[a, \infty) = \{x \in \mathbb{R}; x \geq a\}$$

ou

$$(-\infty, b] = \{x \in \mathbb{R}; x \leq b\}.$$

Podemos representar $\mathbb{R} = (-\infty, \infty)$ e $\{a\} = [a, a]$ como casos degenerados da caracterização acima. Além disso, $\emptyset = (a, a)$, $(a, a]$ ou $[a, a)$.

Obviamente, se um conjunto B é infinito, então o conjunto $\wp(B)$ também o será. Como veremos no Capítulo 2, a razão de usarmos a notação 2^B para representar o conjunto das partes do conjunto B é porque, quando $\#B$ é finita, então $\#\wp(B) = 2^{\#B}$.

Reexaminando a situação proposta no Exemplo 1.6, vemos que

$$\#\wp(E) = 4 = 2^2 = 2^{\#E}.$$

2 Operações com conjuntos

Dados conjuntos quaisquer, podemos definir sobre eles diversas operações. As mais usuais são **união**, **interseção**, **diferença** e **produto cartesiano**.

Antes de partir para as definições propriamente ditas, convém esclarecer o significado de dois importantes conectivos lógicos: o e (denotado por \wedge) e o ou (denotado por \vee).

O conectivo \wedge remete à simultaneidade. Por exemplo, dizemos “essa propriedade possui rebanhos bovino e suíno”, Para informar a presença das duas espécies de rebanho na propriedade, simultaneamente. É bem diferente dizer “essa propriedade possui rebanho bovino ou suíno”, proposição que informa, na verdade, a presença de, pelo menos, um dos tipos de rebanho.

Em outros termos, considere as proposições:

p : essa propriedade possui rebanho bovino

e

q : essa propriedade possui rebanho bovino.

As proposições compostas $p \wedge q$ e $p \vee q$ representam, respectivamente, as duas frases do exemplo do parágrafo anterior. O conectivo \wedge diz que ambas as proposições p e q têm valor lógico verdadeiro, enquanto o conectivo \vee garante que, pelo menos, uma das proposições p e q tem valor lógico verdadeiro.

Compreendido o significado exato desses conectivos, estamos aptos a definir as principais operações para conjuntos.

Definição 1.3. Dados dois conjuntos, A e B , dizemos que a interseção de A e B , denotada por $A \cap B$, é o conjunto

$$A \cap B = \{x; x \in A \wedge x \in B\}.$$

Definição 1.4. Dados dois conjuntos, A e B , dizemos que a união de A e B , denotada por $A \cup B$, é o conjunto

$$A \cup B = \{x; x \in A \vee x \in B\}.$$

Definição 1.5. Dados dois conjuntos, A e B , dizemos que a diferença entre A e B , denotada por $A - B$, é o conjunto

$$A - B = \{x; x \in A \wedge x \notin B\}.$$

Exemplo 1.8. Considere os conjuntos $M = \{1, 2, 3\}$ e $N = \{2, 3, 4\}$. Temos que $M \cap N = \{2, 3\}$, $M \cup N = \{1, 2, 3, 4\}$, $M - N = \{1\}$ e $N - M = \{4\}$.

Exemplo 1.9. Considere os intervalos $I = (0, 1]$ e $J = [\frac{1}{2}, \frac{3}{2})$. Temos que $I \cap J = [2, 3]$, $I \cup J = (0, \frac{3}{2}]$, $I - J = (0, \frac{1}{2})$ e $J - I = (\frac{1}{2}, \frac{3}{2})$.

Em particular, quando $A \subset B$, o resultado das operações definidas acima para esses conjuntos fica: $A \cap B = A$, $A \cup B = B$ e $A - B = \emptyset$. Neste caso, costuma-se chamar a diferença $B - A$ como o **complemento** de A em relação a B e denotá-la por $C_B(A)$ ou C_B^A .

Definição 1.6. Dados dois conjuntos, A e B , dizemos que o produto cartesiano entre A e B , denotado por $A \times B$, é o conjunto

$$A \times B = \{(x, y); x \in A \wedge y \in B\}.$$

Observe que os elementos do conjunto $A \times B$ não são elementos de A nem de B . Eles possuem uma natureza híbrida e são chamados de **vetores**. Cada entrada do vetor carrega informação relacionada ao respectivo conjunto no produto cartesiano.

Exemplo 1.10. Sejam M e N como no Exemplo 1.8. Temos que o vetor $(1, 2) \in M \times N$, isto porque $1 \in M$ e $2 \in N$. No entanto, $(1, 2) \notin N \times M$.

Dados dois conjuntos finitos A e B , a cardinalidade do conjunto $A \times B$ é calculada assim:

$$\#(A \times B) = \#A \cdot \#B,$$

isto é, o número de vetores da forma (x, y) que conseguimos formar com elementos do conjunto A na primeira entrada e elementos do conjunto B na segunda entrada correspondem ao produto da cardinalidade dos dois conjuntos envolvidos. Quando o produto cartesiano apresenta, pelo menos, um conjunto infinito, então sua cardinalidade será infinita.

Exemplo 1.11. Sejam I e J os conjuntos apresentados no Exemplo 1.9. Temos que o vetor $(\frac{1}{2}, 1) \in I \times J$, enquanto o vetor $(1, \frac{1}{2}) \notin I \times J$. Além disso, $\#(I \times J) = \infty$, isto é, o conjunto $I \times J$ contém infinitos vetores.

3 Conjuntos numéricos

Nas seções anteriores, vimos exemplos de importantes conjuntos constituídos por números: N, Z, Q e R. Referiremo-nos a esses conjuntos, ou a seus subconjuntos, como sendo **conjuntos numéricos**.

O conjunto dos números racionais Q foi apresentado no Exemplo 1.3 como sendo o conjunto dos números que possuem uma representação no formato de fração. Mas o que quer dizer uma fração?

3.1 Uma interpretação para frações

Inicialmente, observe que todo número inteiro pode ser escrito no formato de fração. Por exemplo,

$$2 = \frac{2}{1} = \frac{4}{2} = \frac{200}{100} = \frac{-2.000}{-1.000}. \quad (1.5)$$

É por isso que dizemos que $Z \subset Q$.

Considere, agora, a fração

$$\frac{3}{2} = 1,5. \quad (1.6)$$

Note que 1,5 é a representação decimal da fração $3/2$, obtida pela aplicação do algoritmo da divisão. Podemos interpretar a Equação (6) assim:

o tamanho do número 3 é 1,5 vez o tamanho do número 2.

Analogamente, se considerarmos a fração

$$\frac{4}{5} = 0,8. \quad (1.7)$$

podemos interpretá-la como:

o tamanho do número 4 é 0,8 vez o tamanho do número 5.

Quando consideramos frações com denominador 100, dizemos ter aí um **número percentual**⁵ e o símbolo % é usado para denotar essa situação. Por exemplo,

$$\frac{20}{100} = 0,2 = 20\%.$$

Para se calcular um percentual qualquer de um dado número, basta multiplicar o número percentual pelo número dado.

Exemplo 1.12. Numa fazenda, o rebanho bovino é constituído de 500 cabeças, das quais 20% são vacas leiteiras. Então, o número de vacas leiteiras da fazenda é:

$$20\% \text{ de } 500 = 0,2 \cdot 500 = 100.$$

Assim, podemos voltar o olhar para a Equação (6) e propor a seguinte interpretação para ela:

$$\begin{aligned} \frac{3}{2} &= 1,5 \\ \Rightarrow 3 &= 1,5 \cdot 2 \\ \Rightarrow 3 &= (1 + 0,5) \cdot 2 \\ \Rightarrow 3 &= 1 \cdot 2 + 0,5 \cdot 2 \\ \Rightarrow 3 &= 2 + 50\% \cdot 2, \end{aligned}$$

isto é,

3 corresponde ao número 2 aumentado de 50%.

Analogamente, podemos interpretar a Equação (7) assim:

$$\begin{aligned} \frac{4}{5} &= 0,8 \\ \Rightarrow 4 &= 0,8 \cdot 5 \\ \Rightarrow 4 &= (1 - 0,2) \cdot 5 \\ \Rightarrow 4 &= 1 \cdot 5 - 0,2 \cdot 5 \\ \Rightarrow 4 &= 5 - 20\% \cdot 5, \end{aligned}$$

⁵ 20% lê-se: vinte por cento.

isto é,

4 corresponde ao número 5 aumentado de 20%.

Em suma, quando trabalhamos com frações, não podemos esquecer que estamos comparando dois números, vendo quanto o numerador é maior ou menor do que o denominador. Esse raciocínio pode ser usado para calcular rapidamente aumentos ou descontos percentuais.

Exemplo 1.13. Numa fazenda, o rebanho bovino é constituído de 500 cabeças. Entre elas, há um touro recentemente adquirido com a intenção de aumentar o rebanho, ao final de 2011, em 20%. Assim, ao final do ano, a previsão é de que o rebanho passe a ser de

$$\begin{aligned} 500 \text{ aumentando de } 20\% &= 500 + 0,2 \cdot 500 \\ &= 500(1 + 0,2) \\ &= 500 \cdot 1,2 \\ &= 600 \text{ cabeças.} \end{aligned}$$

Exemplo 1.14. Numa fazenda, o rebanho bovino é constituído de 500 cabeças. Uma doença atingiu o rebanho e dizimou 20% daquela população. Assim, restam

$$\begin{aligned} 500 \text{ reduzidos de } 20\% &= 500 - 0,2 \cdot 500 \\ &= 500(1 - 0,2) \\ &= 500 \cdot 0,8 \\ &= 400 \text{ cabeças.} \end{aligned}$$

Generalizando as idéias expostas nos exemplos 1.12, 1.13 e 1.14, considere uma quantidade qualquer C e um número percentual i . Para calcular o percentual i de C , basta multiplicar $i \cdot C$; para calcular a quantidade C com um aumento percentual i , basta multiplicar $(1 + i) \cdot C$; e, finalmente, para se calcular o valor de C , submetido a um desconto percentual i , basta multiplicar $(1 - i) \cdot C$.

Exemplo 1.15. (juros compostos) Um capital de R\$5.000,00 foi aplicado na poupança, que rende juros mensais fixos de 0,5% ao mês. Vamos acompanhar a evolução desse capital ao longo do tempo, denotando por $M(t)$ o montante t meses depois da aplicação. Observe que

$$\begin{aligned}
 M(0) &= 5.000, \\
 M(1) &= 5.000 \cdot (1 + 0,005) = 5.000 \cdot 1,005, \\
 M(2) &= M(1) \cdot 1,005 = (5.000 \cdot 1,005) \cdot 1,005 = 5.000 \cdot (1,005)^2, \\
 M(3) &= M(2) \cdot 1,005 = [5.000 \cdot (1,005)^2] \cdot 1,005 = 5.000 \cdot (1,005)^3, \\
 M(4) &= M(3) \cdot 1,005 = [5.000 \cdot (1,005)^3] \cdot 1,005 = 5.000 \cdot (1,005)^4
 \end{aligned}$$

e, em geral, passados t meses depois da aplicação, teremos um montante igual a

$$M(t) = M(t - 1) \cdot 1,005 = [5.000 \cdot (1,005)^{t-1}] \cdot 1,005 = 5.000 \cdot (1,005)^t.$$

Se considerarmos um capital qualquer C sendo investido numa aplicação que tem uma taxa fixa de retorno mensal i qualquer, como particularizado no Exemplo 1.15, obteremos que o montante $M(t)$ da aplicação ao longo do tempo t obedece à lei

$$M(t) = C \cdot (1 + i)^t$$

que é a famosa **função exponencial**, que regula os chamados **juros compostos**.

3.2 Dízimas periódicas

Dízimas periódicas são frações que não possuem representação decimal finita, mas apresentam um comportamento decimal bem definido, estruturado, previsível. Por exemplo, os números

$$\frac{1}{3} = 0,3333\dots, \quad -\frac{7}{6} = -1,1666\dots \quad \text{e} \quad \frac{21}{99} = 0,21212\dots$$

são dízimas periódicas, pois conseguimos advinhar quais os próximos números nas suas representações decimais infinitas. Já para o número $32,785689102\dots$ essa tarefa é impossível. Trata-se, portanto, de um número irracional.

No dia a dia, fazemos aproximações quando lidamos com números com representação decimal muito extensa, que podem ser do tipo truncamento ou

arredondamento. Fazer uma aproximação de um número decimal consiste em escolher uma quantidade de casas decimais tal que, para além delas, aquela informação não tem sentido prático. No caso do truncamento, as demais casas decimais são desprezadas.

Exemplo 1.16. Quando falamos de dinheiro, $1/3$ de real é, aproximadamente, 0,33 real. Representaremos essa situação assim:

$$\frac{1}{3} \approx 0,33.$$

Essa situação pode ser vista como um truncamento.

Já no caso do arredondamento, deve haver um critério que nos faça decidir se as demais casas decimais serão desprezadas ou se acrescerão uma unidade à última casa decimal que será preservada no número.

Exemplo 1.17. O sistema de registro de notas do IFB em 2009 admitia apenas as notas 0, $1/2$, $3/2$, ..., $19/2$, 10 como registro do aproveitamento final de cada etapa. Se a soma das notas parciais relativas aos instrumentos avaliativos de uma etapa fosse qualquer número diferente dos admitidos, o critério era arredondá-lo para a nota maior e mais próxima entre as admitidas. Em outras palavras, um estudante cuja soma de pontos for 5,333... no 1º bimestre terá sua nota lançada como 5,5.

No entanto, em muitas situações, essa praticidade pode não ser suficiente e, dependendo da grosseria da aproximação, erros podem se propagar e gerar grandes prejuízos materiais ou levar a conclusões não razoáveis. É o que acontece, por exemplo, tratando-se de programação computacional.

Quando lidamos com dízimas periódicas, sempre é possível saber qual é a fração que a gerou, também conhecida como **fração geratriz**. Para isso, basta considerar um múltiplo à potência de 10 adequada da dízima periódica e subtrair desse múltiplo a dízima, obtendo, assim, sua fração geratriz.

Exemplo 1.18. Sabemos que $y = 0,4444\dots$ é uma dízima periódica. Nesse caso, se fizermos a conta $10y - y$, obteremos uma representação em forma de fração para y . De fato,

$$10y - y = 4 \Rightarrow 9y = 4 \Rightarrow y = \frac{4}{9}.$$

No exemplo anterior, o múltiplo $10y$ foi escolhido por conveniência, pois a parte periódica na subtração se alinha, tornando a conta possível de ser feita. O exemplo abaixo apresenta uma situação diferente.

Exemplo 1.19. Sabemos que $z = 1,1232323\dots$ é uma dízima periódica. Nesse caso, a fração geratriz de z será obtida efetuando-se a conta $100z - z$. De fato,

$$100z - z = 112,323232\dots - 1,123232\dots = 111,2 \Rightarrow 99z = 111,2 \Rightarrow z = \frac{111,2}{99} \Rightarrow z = \frac{1.112}{990}.$$

Nesse último exemplo, a conveniência exigiu que o múltiplo 10^2y fosse o escolhido, pois o **período**⁶ da dízima periódica possuía dois números.

3.3 Proporcionalidade

Vimos, como na Equação 6, que um mesmo número racional pode ser escrito sob forma de fração de maneiras diferentes. Vimos também que a divisão de (razão entre) dois números expressa uma comparação entre eles, ou seja, o quanto o numerador é maior ou menor do que o denominador da fração.

Exemplo 1.20. “Nesta primeira avaliação da safra nacional dos cereais, leguminosas e oleaginosas para 2010, estima-se uma produção de 143,4 milhões de toneladas, superior 7,2% à obtida no ano passado. A área a ser colhida é 2,1% maior que a da safra de 2009, que foi de 47,2 milhões de hectares.” (IBGE, 2010, adaptado).

Observe nesse exemplo que foram comparados números que expressam quantidades de uma mesma **grandeza**,⁷ no caso, as grandeszas tonelada e área. O exemplo a seguir é um pouco diferente.

⁶ O período de uma dízima periódica corresponde à quantidade de números que tem seu comportamento reproduzido ao infinito.

⁷ Entendemos grandeza como sendo o atributo físico de um corpo que pode ser qualitativamente distinguido e quantitativamente determinado. Como exemplos de grandeszas, temos a distância percorrida por um carro, a altura de um prédio, o tempo de vida de um animal, a espessura de uma camada de asfalto, a cardinalidade de um conjunto finito, entre outras.

Exemplo 1.21. Na frase “O avião foi perdendo altura, à razão de 7 mil pés por minuto e se despedaçou”,⁸ a razão empregada expressa a velocidade de queda do avião, que, de forma equivalente, poderia ser informada como 14 mil pés por 2 minutos, ou 420 mil pés por hora.

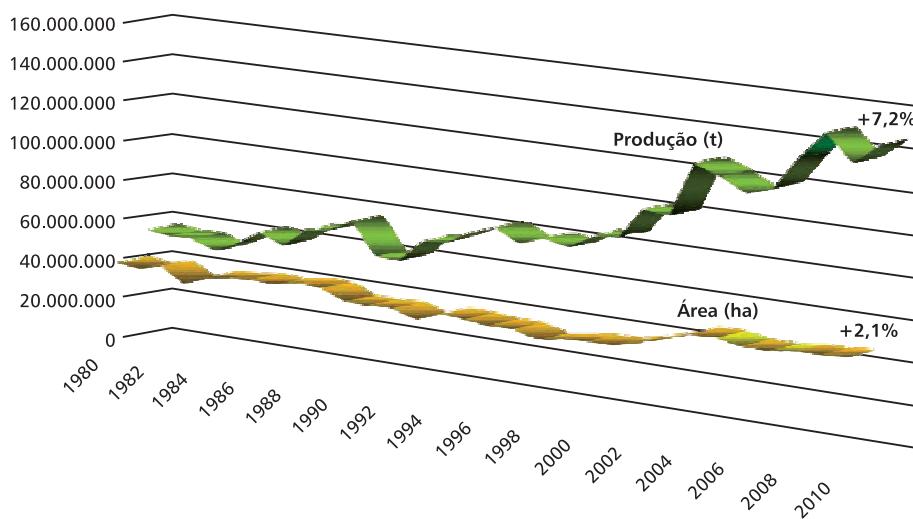


FIGURA 1: CEREAIS, LEGUMINOSAS E OLEAGINOSAS: ÁREA E PRODUÇÃO
BRASIL 1980 A 2010
Fonte: IBGE, 2010

Note que o numerador e o denominador da razão desse exemplo retratam quantidades de grandezas diferentes (comprimento e tempo). Todavia, se pensarmos que

$$\frac{7 \text{ mil pés}}{1 \text{ minuto}} = \frac{7 \cdot \text{mil pés}}{1 \cdot \text{minuto}} = 7 \cdot \frac{\text{mil pés}}{\text{minuto}},$$

concluímos que a razão entre as duas grandezas determina outra grandeza, a velocidade de queda do avião.

Em nosso cotidiano, as grandezas se relacionam o tempo todo. Por exemplo, quando vamos ao supermercado comprar um determinado produto,

⁸ Trecho extraído de matéria publicada no site <http://pt.wikinews.org>, em 16 de agosto de 2005.

o valor a ser pago no caixa se relaciona com a quantidade do produto; quando pegamos um táxi, as grandezas preço da corrida e distância percorrida se relacionam; quando um bioquímico administra uma substância a fim de inibir o crescimento de uma colônia de bactérias, o tempo de reação do medicamento se relaciona com o tamanho da população residual de bactérias; quando um agrônomo suplementa a alimentação de um rebanho, as grandezas quantidade de alimento consumido e massa do animal podem estar relacionadas.

O fato é que duas ou mais grandezas podem se relacionar das mais diferentes formas. Trataremos, aqui, de uma forma muito particular de relacionamento entre grandezas, mas de grande utilidade para resolver problemas simples do nosso dia a dia: a proporcionalidade.

Intuitivamente, observamos a proporcionalidade entre duas grandezas incrementando uma delas e observando se essa ação gera um incremento relativamente igual na outra grandeza ou inverso. No primeiro caso, diremos que as grandezas são **diretamente proporcionais** e, no segundo caso, diremos que elas são **inversamente proporcionais**.

Os exemplos abaixo nos ajudarão a melhorar a nossa intuição a respeito desse conceito.

Exemplo 1.22. Um cliente vai a um supermercado comprar leite. Lá, um litro de leite de caixinha é vendido a R\$1,75. Assim, a grandeza “valor da conta” será diretamente proporcional à grandeza “quantidade de caixinhas”. Isso porque, se o cliente dobrar a quantidade de caixinhas, terá sua conta dobrada; se triplicá-la, terá sua conta triplicada, e assim por diante.

Exemplo 1.23. Num determinado horário, uma empresa de táxi cobra, por corrida, uma bandeirada no valor de R\$4,00, mais R\$0,50 cada 200 metros rodados. Se um cliente desejar percorrer 2 quilômetros, pagará R\$9,00 pela corrida, enquanto, se desejar percorrer 4 quilômetros, pagará R\$14,00. Note que as grandezas “tamanho da corrida” e “preço da corrida” não são diretamente proporcionais, uma vez que o dobro do tamanho da corrida não implica o dobro do preço dela.

Exemplo 1.24. Um carro faz um percurso fixo a uma velocidade média de 30 km/h e gasta uma hora. Se, na volta, ele dobrar a velocidade média, ele reduzirá o tempo de viagem à metade. Se ele triplicar a velocidade média, seu tempo de viagem será reduzido a um terço do inicial. Assim, as grandezas “velocidade média” e “tempo de viagem” são inversamente proporcionais.

Formalmente, temos a seguinte definição:

Definição 1.7. Duas grandezas são ditas ser diretamente proporcionais quando apresentam **variação linear**, isto é, quando a razão entre as duas grandezas é constante,⁹ e são ditas ser inversamente proporcionais quando apresentam **variação hiperbólica**, isto é, quando o produto das duas é constante.¹⁰

Para associar a definição formal de proporcionalidade com a noção intuitiva construída anteriormente, sejam $X(\cdot)$ e $Y(\cdot)$ duas grandezas relacionadas. Essas grandezas apresentarão variação linear se

$$\frac{X(t)}{Y(t)} = c, \text{ qualquer que seja } o t$$

e apresentarão variação hiperbólica se

$$X(t) \cdot Y(t) = c, \text{ qualquer que seja } o t$$

onde c é uma constante qualquer não nula.

Assim, se no Exemplo 1.22, $X(\cdot)$ representa a “quantidade de caixinhas” e $Y(\cdot)$ representa o “valor da conta”, então $X(2)$ representará duas caixas de leite e $Y(2)$ representará o preço pago por duas caixas de leite. Vemos,¹¹ facilmente, que

$$\frac{Y(2)}{X(2)} = \frac{3,50}{2} = 1,75,$$

$$\frac{Y(3)}{X(3)} = \frac{5,25}{3} = 1,75$$

⁹ Conhecida como constante de proporcionalidade direta.

¹⁰ Conhecida como constante de proporcionalidade inversa.

¹¹ As unidades referente a cada grandeza foram, por simplicidade, omitidas, mas devem ser carregadas na interpretação dos problemas.

e, em geral,

$$\frac{Y(k)}{X(k)} = \text{constante}, \text{ seja qual for } k \in N.$$

Analogamente, se no Exemplo 1.24 considerarmos as grandezas $X(.)$ e $Y(.)$ como sendo, respectivamente, “velocidade média” e “tempo de viagem”, teremos satisfeita a equação

$$X(k) \cdot Y(k) = \text{constante}, \text{ seja qual for } k$$

As constantes 1,75 R\$/caixa e 30km obtidas acima são conhecidas como **constantes de proporcionalidade**.

3.4 Regra de três

Apesar de a proporcionalidade ser um tipo de relação muito particular entre grandezas, ela aparece com **frequência**, sobretudo nas situações cotidianas. Entretanto, é importante frisar a necessidade de analisarmos bem a situação para ver se, de fato, cabe ali raciocínio deste tipo.

Nesta seção, apresentaremos um algoritmo eficiente para resolver problemas envolvendo grandezas proporcionais, também conhecido por **regra de três**. Esses problemas se caracterizam por apresentar diversas informações a respeito de duas ou mais grandezas, sendo que para uma das grandezas apresentadas falta uma informação que deverá ser descoberta a partir das demais. Diremos que a **grandeza de interesse** é a grandeza para a qual falta uma informação e identificaremos o problema como um **problema de regra de três**, se cada uma das demais grandezas for, de alguma maneira, proporcional à grandeza de interesse.

Uma vez identificado que se trata de um problema de regra de três, executaremos as seguinte etapas:

1. Construir uma tabela informando todas as grandezas envolvidas no problema. Nesse exemplo genérico, X será a grandeza de interesse e G_1, G_2, \dots, G_n serão as demais grandezas envolvidas no problema.

GRANDEZAS	X	G_1	G_2	\dots	G_n
medida 1				\dots	
medida 2				\dots	

2. Escrever na tabela as medidas informadas no problema para cada grandeza, usando uma letra qualquer para representar a medida que se deseja obter (que será a **variável** do problema) e identificar a grandeza de interesse com uma seta apontada para uma direção qualquer. Nesse exemplo genérico, $x_1, a_1, a_2, b_1, b_2, \dots, \gamma_1, \gamma_2$ são valores conhecidos e y é a variável que se deseja descobrir.

GRANDEZAS	$\uparrow X$	G_1	G_2	\dots	G_n
medida 1	x_1	a_1	b_1	\dots	γ_1
medida 2	y	a_2	b_2	\dots	γ_2

3. Comparar¹² cada uma das outras grandezas com a grandeza de interesse, verificando o tipo de proporcionalidade. Se for direta, identificar a outra grandeza com uma seta apontada na direção da seta que identifica a grandeza de interesse; se for inversa, identificá-la com uma seta na direção oposta à seta que identifica a grandeza de interesse.

GRANDEZAS	$\uparrow X$	$\uparrow G_1$	G_2	\dots	G_n
medida 1	x_1	a_1	b_1	\dots	γ_1
medida 2	y	a_2	b_2	\dots	γ_2

GRANDEZAS	$\uparrow X$	$\uparrow G_1$	$\downarrow G_2$	\dots	G_n
medida 1	x_1	a_1	b_1	\dots	γ_1
medida 2	y	a_2	b_2	\dots	γ_2

GRANDEZAS	$\uparrow X$	$\uparrow G_1$	$\downarrow G_2$	\dots	$\uparrow G_n$
medida 1	x_1	a_1	b_1	\dots	γ_1
medida 2	y	a_2	b_2	\dots	γ_2

¹² Essa tarefa é mais delicada quando se tem muitas grandezas, pois exige que simpliquemos o problema a uma situação parecida, na qual apenas duas grandezas variam enquanto as outras permanecem fixas.

4. A partir dos elementos da tabela, montar a equação

$$\frac{x_1}{y} = \underbrace{\frac{a_1}{a_2}}_{\text{refere-se à grandeza } G_1} \cdot \underbrace{\frac{b_2}{b_1}}_{\text{refere-se à grandeza } G_2} \cdot \dots \cdot \underbrace{\frac{\gamma_1}{\gamma_2}}_{\text{refere-se à grandeza } G_n},$$

obedecendo ao que chamaremos de **critério de inversão**.

5. Resolva a equação, isolando a variável do problema.

O que chamamos de critério de inversão, na verdade, é uma regra muito simples que consiste em inverter a razão com as medidas da grandeza que tiver sido identificada com uma seta apontada para a direção oposta à seta da variável de interesse.

Executando esse algoritmo, conseguiremos resolver qualquer problema de regra de três!

Exemplo 1.25. (Regra de três simples) *Um homem percorre um trajeto de bicicleta. Se, pedalando à velocidade de 5 km/h, ele demora 6 horas, quanto tempo o homem demorará para percorrer esse mesmo trajeto a uma velocidade 3 Km/h?*

Primeiramente, montemos a tabela com as respectivas medidas informadas pelo problema.

VELOCIDADE	\uparrow TEMPO
5	6
3	t

Observe que estamos interessados na grandeza tempo. Note também que velocidade e tempo são grandezas inversamente proporcionais, isto é, quanto menor for a velocidade desenvolvida pelo ciclista, maior será o tempo gasto no trajeto. Assim:

\downarrow VELOCIDADE	\uparrow TEMPO
5	6
3	t

Finalmente, montamos a proporção, obedecendo ao critério de inversão, e efetuamos as contas:

$$\begin{aligned}\frac{6}{t} &= \frac{3}{5} \\ \Rightarrow 3 \cdot t &= 6 \cdot 5 \\ \Rightarrow t &= \frac{6 \cdot 5}{3} = 10 \text{ horas.}\end{aligned}$$

Problemas como o resolvido no exemplo acima são ditos ser de **regra de três simples** e recebem esse nome por envolverem apenas duas grandezas. Quando envolvem mais de duas grandezas, costumam ser chamados de problemas de **regra de três composta**.

Exemplo 1.26. (Regra de três composta) *Uma família subsiste da criação de 2 bois e, alimentando-os durante 8 dias, são consumidas 480 gramas de sal mineral. Se mais 2 bois são comprados, qual deverá ser a quantidade de sal mineral para a alimentação de todos os animais durante 12 dias?*

Primeiramente, montemos a tabela com as respectivas medidas informadas pelo problema.

BOIS	↑ GRAMAS	TEMPO
2	480	8
4	x	12

Observe que estamos interessados na grandeza gramas e, para descobrir o valor de x , deveremos verificar a relação de proporcionalidade que cada uma das outras grandezas possui com a grandeza de interesse. Note que, quanto maior a quantidade de animais, maior deverá ser a quantidade de sal mineral para alimentá-los. Logo, as grandezas bois e gramas são diretamente proporcionais.

↑ BOIS	↑ GRAMAS	TEMPO
2	480	8
4	x	12

Por outro lado, quanto maior o tempo de alimentação dos animais, maior deverá ser a quantidade de alimento disponível. Logo, as grandezas gramas e tempo também são diretamente proporcionais.

\uparrow BOIS	\uparrow GRAMAS	\uparrow TEMPO
2	480	8
4	x	12

Finalmente, montamos a proporção e efetuamos as contas. Neste caso, não precisaremos inverter nenhuma das razões.

$$\begin{aligned}\frac{480}{x} &= \frac{2}{4} \cdot \frac{8}{12} \\ \Rightarrow x \cdot 2 \cdot 8 &= 480 \cdot 4 \cdot 12 \\ \Rightarrow x = \frac{480 \cdot 4 \cdot 12}{2 \cdot 8} &= 1 \cdot 440 \text{ gramas}.\end{aligned}$$

Resumindo as informações desta seção, vimos na subseção 3.1 uma interpretação para a razão de dois números pertencentes a um mesmo conjunto, sem nos preocuparmos com o que eles representam. Na subseção 3.3, vimos uma interpretação para essas mesmas razões, mas quando numerador e denominador representam coisas diferentes, isto é, números que medem grandezas diferentes.

Quando lidamos com a razão de dois números que representam medidas de uma mesma grandeza retornamos, na verdade, à situação abordada na Subseção 3.1. De fato, como sugere o exemplo abaixo,

$$\frac{3 \text{ unidades da grandeza } A}{2 \text{ unidades da grandeza } A} = \frac{3}{2} \cdot \frac{\text{unidade da grandeza } A}{\text{unidade da grandeza } A} = \frac{3}{2}.$$

Exercícios

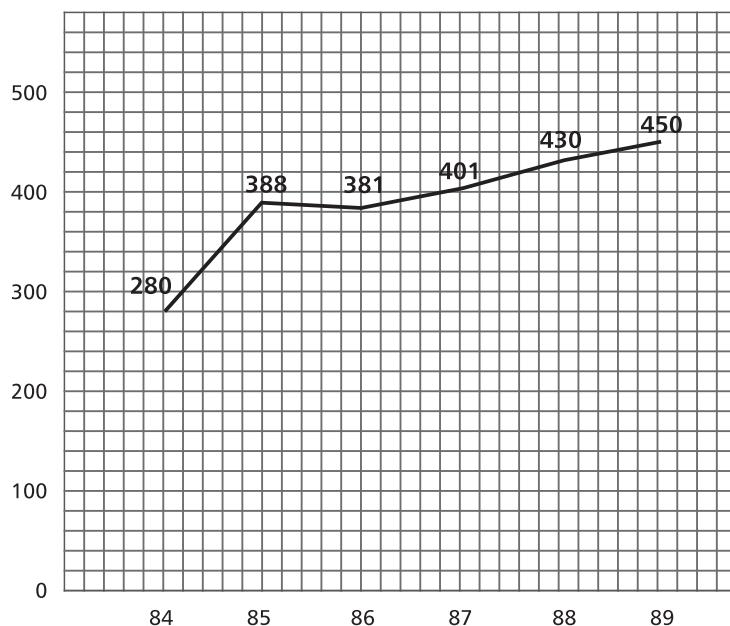
- Numa pesquisa de opinião pública a respeito do consumo de duas marcas de um mesmo produto, 50 pessoas foram entrevistadas, entre as quais 30 declararam consumir o produto A, 35 declararam consumir o produto B, e 5 declararam não consumir nenhum dos produtos. Quantos desses consumidores consomem ambos os produtos?
 - 20.
 - 25.
 - 15.
 - 5.
 - 10.

2. Em uma escola, 40% dos estudantes gostam de inglês, 60% gostam de espanhol e 10% não gostam de nenhum dos dois idiomas. Qual o percentual de estudantes que gosta apenas de espanhol?
- a) 20%.
 - b) 30%.
 - c) 40%.
 - d) 50%.
 - e) 60%.
3. (IFB) Em certo ano, uma escola com 500 estudantes oferecia oficinas e a matrícula era obrigatória em, pelo menos, uma delas, e o estudante poderia se matricular em, no máximo, duas. Feitas as matrículas, apurou-se que 200 estudantes fariam música e 200 fariam judô. Sabendo-se que o número de estudantes que farão ambas as atividades corresponde à metade do numero de estudantes que farão outras oficinas, assinale a única alternativa INCORRETA:
- a) 100 estudantes não farão música nem judô.
 - b) 100 estudantes matricularam-se apenas música.
 - c) 100 estudantes matricularam-se em música e judô.
 - d) 100 estudantes matricularam-se apenas em judô.
 - e) O numero de estudantes que se matricularam só em música ou só em judô é igual ao número de estudantes que não farão nem música nem judô.
4. (IFB) Num concurso público de nível médio, verificou-se que 60% dos candidatos tinham nível superior, entre os quais 35% eram homens. Sabendo que 60% dos candidatos eram homens, assinale a única alternativa CORRETA:
- a) 1% dos candidatos são mulheres que não possuem curso superior.
 - b) 21% dos candidatos são mulheres que possuem curso superior.
 - c) 1% das candidatas mulheres não possuem curso superior.
 - d) 35% dos candidatos são homens que possuem curso superior.
 - e) Há mais candidatos homens sem curso superior do que candidatas mulheres com curso superior.

5. (IFB) Numa escola há 500 estudantes. Sabendo-se que a quantidade de meninos é três vezes maior que a quantidade de meninas, então estudam nessa escola:
- a) 375 meninos e 125 meninas.
 - b) 125 meninos e 375 meninas.
 - c) 400 meninos e 100 meninas.
 - d) 100 meninos e 400 meninas.
 - e) 333 meninos e 111 meninas.
6. (IFB) Considere que, dos 50 hóspedes de um hotel, 35 tomaram café pela manhã e 35 tomaram leite. Sabendo que 10 deles optaram por não tomar café nem leite, assinale a opção INCORRETA.
- a) 5 hóspedes tomaram leite, mas não tomaram café.
 - b) 5 hóspedes tomaram café, mas não tomaram leite.
 - c) 10 hóspedes tomaram só café ou só leite.
 - d) 40 hóspedes tomaram café ou tomaram leite.
 - e) 40 hóspedes tomaram café e leite.
7. (IFG) O valor da expressão numérica
- $$\frac{0,001 \times (0,001)^2 \times 100}{0,00001}$$
- é
- a) 0,01.
 - b) 0,1.
 - c) 0,001.
 - d) 0,0001.
 - e) 1.
8. (IFG) Dos 350 candidatos que se inscreveram para um processo seletivo, apenas 280 compareceram. Então, pode-se afirmar que o índice percentual de não comparecimento foi de:
- a) 15%.
 - b) 18%.
 - c) 19%.
 - d) 2%.
 - e) 20%.

9. (IFB) Comprei um computador por 80% do que ele valia há um mês e, com isso, economizei R\$300,00. Assim, esse computador me custou:
- a) R\$1.500,00.
 - b) R\$1.200,00.
 - c) R\$1.800,00.
 - d) R\$2.100,00.
 - e) R\$2.400,00.
10. (IFG) Na compra de uma camisa, obteve-se um desconto de 15%, o que proporcionou uma economia de R\$6,00. Quanto foi pago pela camisa?
- a) R\$40,00.
 - b) R\$30,00.
 - c) R\$35,00.
 - d) R\$32,00.
 - e) R\$34,00.
11. (IFG) Qual a taxa final de aumento de um produto que sofreu um reajuste de % e, logo em seguida, foi reajustado em 5% sobre o valor anterior?
- a) 11%.
 - b) 12%.
 - c) 11,5%.
 - d) 11,3%.
 - e) 12,5%.
12. (IFG) Um livro que custava R\$70,00 foi vendido por R\$56,00. Qual foi a taxa percentual de desconto ?
- a) 10%.
 - b) 15%.
 - c) 20%.
 - d) 25%.
 - e) 17%.

13. (IFB) O gráfico abaixo representa a evolução do patrimônio de uma empresa, entre 1984 e 1989. No eixo horizontal, estão marcados os anos em que foram realizados os balanços. No eixo vertical, estão assinalados os valores (em milhares de reais) do patrimônio. Com base no gráfico, é CORRETO afirmar que:

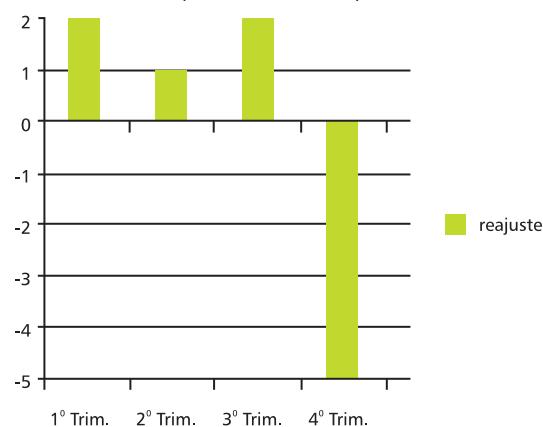


- a) comparado com o início do período observado, a empresa apresentou um aumento do patrimônio superior a 60%.
- b) de 1985 para 1986, houve uma queda de 2% no patrimônio dessa empresa.
- c) com o crescimento percentual no patrimônio observado entre os anos 1986 e 1987 foi igual ao observado entre os anos 1988 e 1989.
- d) comparado com o ano anterior, o maior crescimento do patrimônio da empresa se deu em 1989.
- e) o patrimônio da empresa apresentou desempenho crescente no período observado.

14. (IFB) Uma loja revende produtos importados. O preço de venda desseus produtos é estabelecido considerando uma margem de lucro de 20% sobre o preço da venda. Assim, o lucro gerado por um produto importado a um valor correspondente a R\$7.000,00 e revendido nessa loja será de:

- a) R\$1.750,00.
- b) R\$1.400,00.
- c) R\$1.500,00.
- d) R\$1.650,00.
- e) R\$1.900,00.

15. (IFB) O preço de uma mercadoria foi submetido sucessivamente às seguintes variações percentuais de preço durante o ano passado.



Com base nessas informações, é correto afirmar que:

- a) Ao final do ano passado, o preço da mercadoria apresentava um desconto inferior a 0,2%, comparado com o preço da mercadoria no início do ano.
- b) Ao final do ano passado, o valor da mercadoria era o mesmo apresentado no início daquele ano.
- c) Ao final do ano passado, o valor da mercadoria estava 1% maior do que no início daquele ano.
- d) Ao final do ano passado, o valor da mercadoria estava 0,1% maior do que no início daquele ano.
- e) Ao final do ano passado, o valor da mercadoria estava 1% menor do que no início daquele ano.

16. (IFB) Considere que uma pessoa tenha recebido R\$10.000,00 de um prêmio da loteria e tenha usado 20% deste prêmio para pagar uma dívida. Além disso, gastou 25% do que restou com uma festa. Que quantia sobrou do prêmio?
- a) R\$2.000,00.
 - b) R\$3.000,00.
 - c) R\$4.000,00.
 - d) R\$5.000,00.
 - e) R\$6.000,00.
17. (IFB) Uma loja oferecia suas mercadorias em dois planos de pagamento:
- PLANO 1: à vista, com desconto de 5%.
- PLANO 2: em duas parcelas iguais, sendo uma entrada e a outra a ser paga em 30 dias.
- Considerando essa situação, é correto afirmar que:
- a) a taxa de juros mensal da loja é superior a 10%.
 - b) a taxa de juros mensal da loja é de 5%.
 - c) a taxa de juros mensal da loja é de 10%.
 - d) a taxa de juros mensal da loja é inferior a 5%.
 - e) a taxa de juros mensal da loja é maior do que 5% e menor do que 10%.
18. (IFB) Há algum tempo, um investidor comprou um apartamento em Brasília e hoje deseja revendê-lo. O possível comprador dispõe de k reais para a entrada, quantia essa que é 20% menor do que o valor gasto pelo investidor à época da compra do imóvel. Se o investidor deseja ter um lucro de, pelo menos, 140% sobre o valor que gastou para adquirir o imóvel, qual deve ser o preço mínimo de venda?
- a) $3k$ reais.
 - b) $4k$ reais.
 - c) $2k$ reais.
 - d) $1,75k$ reais.
 - e) $1,92k$ reais.

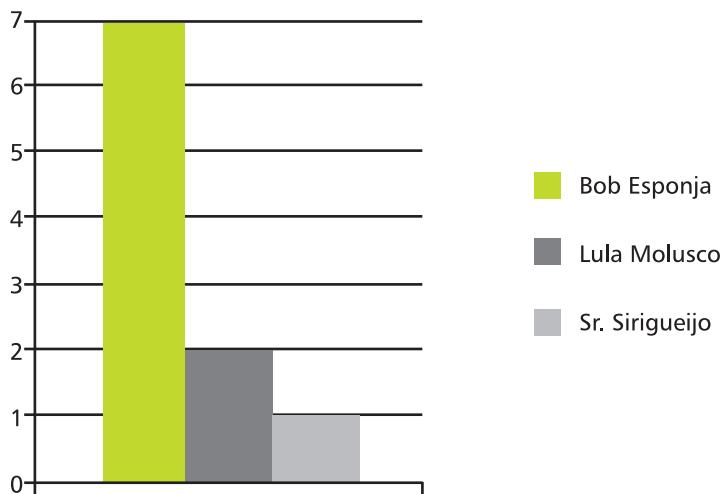
19. (IFB) O texto a seguir foi retirado do *sítio* oficial do IBGE, e trata da produtividade de alguns produtos agrícolas estimada para 2010, comparada com o ano anterior.

Nesta primeira avaliação da safra nacional dos cereais, leguminosas e oleaginosas para 2010, estima-se uma produção de 143,4 milhões de toneladas, superior 7,2% à obtida no ano passado. A área a ser colhida será maior 2,1% que a da safra de 2009, que foi de 47,2 milhões de hectares (IBGE, com adaptações).

Com base nestas informações, estima-se que

- a) a produtividade de 2010 superará a produtividade de 2009 em um percentual que está entre 0% e 5%.
 - b) a produtividade de 2009 superará a produtividade de 2010 em um percentual que está entre 5% e 10%.
 - c) a produtividade de 2010 superará a produtividade de 2009 em um percentual que está entre 10% e 15%.
 - d) a produtividade de 2009 superará a produtividade de 2010 em um percentual que está entre 10% e 15%.
 - e) a produtividade de 2010 superará a produtividade de 2009 em um percentual que está entre 5% e 10%.
20. (IFB) Até o mês passado, o preço de venda de um produto era calculado a partir do seu custo de fabricação, acrescido de uma margem de lucro de 20%. Neste mês, o custo de fabricação do produto reduziu-se em 10%, e o administrador resolveu aplicar um reajuste de 10% sobre o preço de venda desse produto. Essa medida fez com que, neste mês,
- a) o preço de venda do produto diminuisse 1%.
 - b) o preço de venda do produto aumentasse 1%.
 - c) o preço de venda do produto permanesse o mesmo.
 - d) a margem de lucro sobre o produto permanecesse a mesma.
 - e) a margem de lucro sobre o produto dobrasse.

21. (IFB) Uma pesquisa com 400 crianças mostrou a preferência delas por três personagens de desenho animado. O resultado foi o seguinte:



Então é CORRETO afirmar que:

- a) 210 crianças preferem o Bob Esponja.
 - b) apenas 70 crianças preferem o Bob Esponja.
 - c) as crianças preferem o Sr. Sirigueijo ao Lula Molusco.
 - d) 280 crianças preferem o Bob Esponja.
 - e) apenas 40 crianças preferem o Lula Molusco.
22. (IFB) Se 10 homens constroem um muro em 10 horas, quantos homens são necessários para construir o mesmo muro em 5 horas?
- a) 5.
 - b) 10.
 - c) 15.
 - d) 20.
 - e) 30.

23. (IFB) Um funcionário público do Instituto Federal de Educação, Ciência e Tecnologia de Brasília (IFB), viajando de carro a uma velocidade média de 80 km/h, gastou meia hora para se deslocar do *campus* Plano Piloto ao *campus* Planaltina. Na volta, no entanto, gastou 25 minutos. Assim, a velocidade média desenvolvida pelo carro do funcionário na volta do percurso foi de
- 96 km/h.
 - 66,6 km/h.
 - 85 km/h.
 - 90 km/h.
 - 100 km/h.
24. (EPCAr) Um trem com a velocidade de 45 km/h, percorre certa distância em três horas e meia. Nas mesmas condições e com a velocidade de 60 km/h, quanto tempo gastará para percorrer a mesma distância?
- 2h30min18s.
 - 2h37min8s.
 - 2h37min30s.
 - 2h30min30s.
 - 2h29min28s.
25. Em 6 dias de trabalho, 12 confeiteiros fazem 960 tortas. Em quantos dias 4 confeiteiros poderão fazer 320 tortas?
26. Em 6 dias, 6 galinhas botam 6 ovos. Quantos ovos 12 galinhas botam em 12 dias?
27. (IFB) Os nutricionistas que trabalham no restaurante do *campus* Planaltina informaram que os 320 comensais usuários do restaurante consomem 1.440 litros de leite a cada 18 dias. No entanto, eles estão preocupados com o aumento de comensais previsto para o próximo semestre, que passará a ser de 400, enquanto que a previsão no fornecimento de leite será de 1.500

litros para o mesmo período. Considerando essa nova situação, é correto afirmar que:

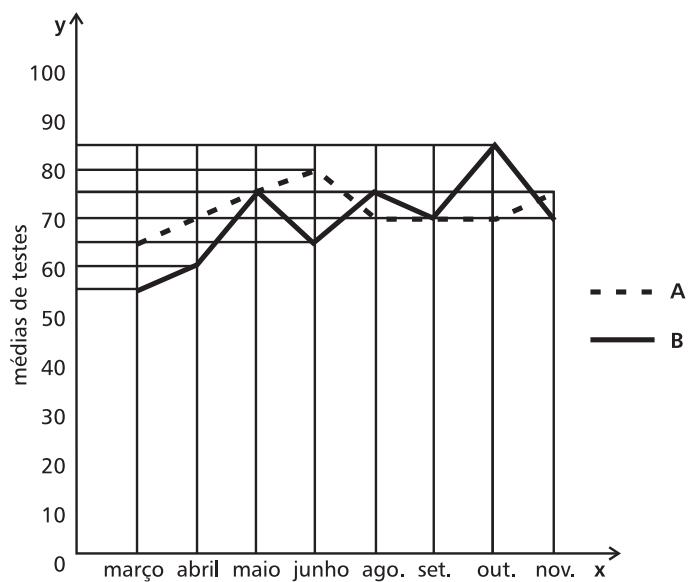
- a) havendo um aumento proporcional do consumo, o suprimento de leite será suficiente para 15 dias.
 - b) não há razão para preocupação, uma vez que o suprimento de leite aumentou proporcionalmente ao número de comensais.
 - c) havendo um aumento proporcional do consumo, o suprimento de leite será suficiente para, no máximo, 14 dias.
 - d) havendo um aumento proporcional do consumo, o suprimento de leite será suficiente para, pelo menos, 21 dias.
 - e) havendo um aumento proporcional do consumo, o suprimento de leite será suficiente para 16 dias.
28. (IFB) Uma empresa de refrigerantes gasta R\$2.000,00 em propaganda mensalmente. A tabela a seguir simula os gastos (em reais) dessa empresa em determinado mês, em função da produção de refrigerantes (em litros).

Produção (em litros)	Custo de produção (em reais)
0	2000
500	2100
1000	2200
1500	2300
2000	2400

Com base nesses dados, é CORRETO afirmar:

- a) que, ao se produzir 2.000 litros de refrigerante, o custo do litro é inferior a R\$1,00.
- b) que o custo total de fabricação é diretamente proporcional à quantidade de refrigerante produzida.
- c) que o custo total de fabricação é inversamente proporcional à quantidade de refrigerante produzida.
- d) que, ao se produzir 1.000 litros de refrigerante, o custo do litro é inferior a R\$2,00.
- e) que R\$2.000,00 representa um custo fixo de produção.

29. (IFB) Se duas torneiras enchem um tanque em 40 minutos, para encher o mesmo tanque em 16 minutos, são necessárias
- 5 torneiras.
 - 6 torneiras.
 - 8 torneiras.
 - 10 torneiras.
 - 12 torneiras.
30. (IFB) Sabendo-se que dois padeiros produzem 20 kg de pão em duas horas, é correto afirmar que, para se produzir os mesmos 20 kg de pão em 20 minutos, são necessários
- 6 padeiros.
 - 8 padeiros.
 - 10 padeiros.
 - 12 padeiros.
 - 14 padeiros.
31. (IFB) O gráfico abaixo mostra o perfil de desempenho das turmas A e B em Matemática, no ano passado.



Com base no gráfico, marque a única opção CORRETA.

- a) Nos meses de maio e setembro, ambas as turmas apresentaram o mesmo desempenho.
- b) No mês de maio, a turma A demonstrou melhor desempenho que a turma B.
- c) Comparado ao mês de abril, o desempenho da turma A cresceu mais que o desempenho da turma B no mês de maio.
- d) A turma B apresentou desempenho mais equilibrado que a turma A ao longo do ano.
- e) Comparado ao mês de março, o desempenho da turma A cresceu mais que o desempenho da turma B no mês de abril.

CAPÍTULO II

NOÇÕES DE CONTAGEM



Noções de contagem

Ao lidarmos com um conjunto discreto finito, seja de coisas, de informações ou mesmo de possibilidades de realização de um determinado experimento, em muitas situações cotidianas precisamos separar, ou mesmo contar, determinados elementos que comungam de uma determinada característica.

Exemplo 2.1. *Na abertura de uma conta corrente, o banco exige que o futuro cliente crie uma senha de 7 dígitos, a partir das 26 letras do alfabeto brasileiro e dos algarismos 0, 1, ..., 9. Porém, as senhas devem ter os 3 primeiros dígitos formados por letras e os 4 últimos por números.*

Exemplo 2.2. *Para representar o Instituto Federal de Brasília (IFB) num evento tecnológico em Santa Catarina, será enviado um grupo de três estudantes da turma do curso de Agropecuária subsequente ApA1, que possui 40 estudantes matriculados.*

Na situação do *Exemplo 2.1*, é possível criar muitas senhas diferentes. Por exemplo, é certo que a proposta de senha AFE3538 seja aceita pelo sistema, enquanto que seguramente a proposta de senha 7MN487P será rejeitada. Assim, podemos concluir que o conjunto de senhas admitidas pelo sistema do banco está contido no universo de senhas alfanuméricas de 7 dígitos.

Imagine, agora, como formar subconjuntos, de qualquer cardinalidade possível, a partir de um conjunto de 40 pessoas. Note que estamos falando do conjunto das partes desse conjunto de 40 pessoas. A situação do *Exemplo 2.2* se refere a subconjuntos específicos daquela turma de estudantes, ou seja, aqueles subconjuntos que possuem cardinalidade igual a 3.

As situações apresentadas induzem as questões naturais:

- Qual é a quantidade total de senhas aceitas pelos sistema do banco?
- De quantas maneiras distintas podemos escolher o grupo de 3 pessoas entre os 40 estudantes do ApA1 para participar do evento?

Essas questões sugerem, respectivamente, dois questionamentos mais gerais:

- Dado um conjunto Ω qualquer, qual a cardinalidade do conjunto $A \subset \Omega$ que reúne elementos com uma dada característica?
- Dado um conjunto Ω qualquer, quantos subconjuntos de Ω possuem uma dada cardinalidade?

Nos debruçaremos, agora, sobre a tarefa de responder a essas questões gerais.

1 Os princípios fundamentais da contagem

Nas situações listadas no início deste capítulo, precisamos resgatar a noção de cardinalidade de um conjunto, que é uma espécie de **medida**, quando

estamos trabalhando num universo de conjuntos discretos. Medir um conjunto é atribuir-lhe determinado peso, comparado a outros conjuntos similares. Por exemplo, tratando-se de conjuntos discretos, um conjunto de cardinalidade (ou medida) igual 3 significa que ele possui 3 elementos, enquanto que um conjunto de cardinalidade (ou medida) zero significa que é um conjunto vazio.

Outra noção importante que precisamos resgastar é a de produto cartesiano, apresentada na Definição 1.6. A próxima definição é uma generalização desse conceito.

Definição 2.1. Dado um conjunto A qualquer, definimos A^n como sendo o conjunto

$$A^n = \underbrace{A \times A \times \cdots \times A}_{n \text{ vezes}} = \{(a_1, a_2, \dots, a_n); a_i \in A, i = 1, 2, \dots, n\}.$$

Cada elemento de A^n é dito ser um **vetor n-dimensional**.

Se A for um conjunto discreto, então A^n também será discreto, e se, além disso, $\#A$ for finita, então também será finita a $\#(A^n)$.

Exemplo 2.3. Se $B = \{0, 1\}$, então os vetores

$$B^2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

e

$$B^3 = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}.$$

No *Exemplo 2.3*, perceba que $\#B = 2$, enquanto que $\#(B^2) = 4$ e $\#(B^3) = 8$. Descobrimos isso listando todas as possibilidades de vetores com dimensão 2 e 3, respectivamente. Mas qual será a cardinalidade de B^4 , B^5 ou B^{40} ? No que segue, tentaremos responder a essa questão.

Antes disso, contudo, voltemos a analisar a situação do *Exemplo 2.1* e nos concentremos em responder qual é a quantidade de senhas de 7 dígitos que podemos formar a partir dos elementos do conjunto

$$A = \{0, 1, \dots, 9, a, b, \dots, z\}.$$

Claramente, o trabalho de se criar uma senha de 7 dígitos deve ser desenvolvido nas seguintes etapas:

- (E_1) escolha do primeiro dígito;
- (E_2) escolha do segundo dígito;
- \vdots
- (E_7) escolha do sétimo dígito.

Todas as etapas devem ser executadas e, além disso, elas são independentes umas das outras. Isso significa que a execução da etapa E_i não interfere na execução da etapa E_j , para qualquer $1 \leq i, j \leq 7, i \neq j$.

Em situações como essa, podemos usar o

Definição 2.2. (PRINCÍPIO MULTIPLICATIVO DE CONTAGEM) Se um problema pode ser resolvido em n etapas E_1, E_2, \dots, E_n independentes, onde $\#(E_i)$ representa o número de maneiras distintas de se resolver a Etapa E_i , então o conjunto Ω de todas maneiras distintas de se resolver o problema é tal que $\#\Omega = \#(E_1) \cdot \#(E_2) \cdot \dots \cdot \#(E_n)$.

No caso particular que estamos tratando, temos que $\#(E_i) = 36$, qualquer que seja $i = 1, 2, \dots, 7$. Logo, se chamarmos de Ω o conjunto de todas as senhas alfanuméricas de 7 dígitos, temos que

$$\#\Omega = \#(E_1) \cdot \#(E_2) \cdot \dots \cdot \#(E_7) = 36 \cdot \dots \cdot 36 = 36^7 \quad (2.1)$$

Isto é, o número máximo de senhas alfanuméricas que podem ser geradas com 7 dígitos é 36^7 . A senha 7MN487P é uma dessas possibilidades.

Se chamarmos de Λ o conjunto contendo todas as senhas que o sistema bancário aceita como válidas, teremos que $\Lambda \subset \Omega$ e que a senha 7MN487P $\notin \Lambda$,

isto é, a inclusão $\Lambda \subset \Omega$ é **própria**.¹ Resta-nos responder qual é a cardinalidade de Λ , sabendo que um elemento de Λ possui letras nos três primeiros dígitos e números nos demais. Como antes, procedamos por etapas:

- (E_1) escolha do primeiro dígito;
- (E_2) escolha do segundo dígito;
- .
- .
- (E_7) escolha do sétimo dígito.

Como na situação anterior, as etapas E_i , $i = 1, \dots, 7$ são **independentes**. Logo, pelo princípio multiplicativo de contagem, teremos

$$\#\Lambda = \#(E_1) \cdot \#(E_2) \cdot \dots \cdot \#(E_7) = 26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 26^3 \cdot 10^4. \quad (2.2)$$

Neste momento, podemos voltar ao questionamento gerado logo depois da apresentação do *Exemplo 2.3*, ou seja, se considerarmos $B = \{0, 1\}$ descobrir, sem ter que listar (isso seria muito despendioso), qual será a cardinalidade de B^n . Esse problema pode ser resolvido de maneira similar ao problema das senhas, uma vez que cada vetor do conjunto B^n pode ser encarado como um senha de n dígitos formada a partir do conjunto B . Isso significa que o problema pode ser resolvido em n etapas independentes, em que cada etapa envolve a escolha de um número para ocupar uma entrada do vetor. Assim, pelo princípio multiplicativo de contagem,

$$\#(\{0, 1\}^n) = \underbrace{2 \cdot 2 \cdot \dots \cdot 2}_{n \text{ vezes}} = 2^n. \quad (2.3)$$

Em particular, $\#(B^4) = 2^4$, $\#(B^5) = 2^5$ e $\#(B^{40}) = 2^{40}$.

¹ Dados dois conjuntos A e B , tais que $A \subset B$, dizemos que essa inclusão é própria quando $A \neq B$, isto é, quando existe pelo menos um elemento de B que não pertence a A .

Outra situação bastante comum na modelagem de problemas de contagem é quando o problema deve ser resolvido em **etapas mutuamente exclusivas**. Para facilitar a compreensão, acompanhe o exemplo abaixo.

Exemplo 2.4. *Ao final de uma corrida, um ciclista está sedento. Resolve, então, passar numa lanchonete para matar a sede. A lanchonete oferece 3 marcas de água, 5 tipos de suco e 4 marcas de refrigerante, e o dinheiro que o ciclista possui lhe permite fazer uma única escolha.*

Analizando o *Exemplo 2.4*, é fácil concluir que o ciclista tem 12 maneiras distintas de resolver o seu problema. Apesar da simplicidade do exemplo, podemos explorá-lo a fim de refinar nossa maneira de planejar a contagem de elementos de um conjunto qualquer em situações mais gerais.

Vamos pensar por etapas. Resolver o problema do ciclista significa:

- (E_1) escolher um tipo de água
 - (E_2) escolher um tipo de suco
- ou
- (E_3) escolher um tipo de refrigerante.

No entanto, essas etapas são mutuamente exclusivas, o que significa dizer que a realização de uma exclui a realização das outras. Em situações como essa, para efetuar a contagem do número de possibilidades, utilizamos o **princípio aditivo de contagem**, enunciado abaixo:

Definição 2.3. (PRINCÍPIO ADITIVO DE CONTAGEM) Se um problema pode ser resolvido em n etapas E_1, E_2, \dots, E_n mutuamente exclusivas, em que $\#(E_i)$ representa o número de maneiras distintas de se resolver a Etapa E_i , $i = 1, 2, \dots, n$, então o conjunto Ω de todas maneiras distintas de se resolver o problema é tal que $\#\Omega = \#(E_1) + \#(E_2) + \dots + \#(E_n)$.

Nesse ponto da discussão, chamamos a atenção para a relação dos conectivos “e” e “ou”, que aparecem naturalmente na interpretação dos problemas e na esquematização das etapas de contagem, com os princípios fundamentais de contagem. O primeiro está relacionado ao princípio multiplicativo, enquanto que o segundo, ao princípio aditivo.

O segredo de se realizar corretamente a contagem dos elementos de um conjunto está na compreensão do formato do elemento que desejamos contar e na elaboração de um roteiro que se limite a aplicar os princípios fundamentais de contagem acima enunciados.

Exemplo 2.5. *A partir do conjunto*

$$M = \{0, 1, \dots, 9\},$$

quantas senhas de 4 dígitos podemos formar, de modo que comecem com número ímpar ou terminem com um número par?

No *Exemplo 2.5*, podemos considerar os conjuntos

$$P = \{\text{senhas de quatro dígitos que começam com número ímpar}\}$$

e

$$Q = \{\text{senhas de quatro dígitos que terminam com número par}\}.$$

Observe que queremos descobrir a cardinalidade do conjunto $P \cup Q$, e o problema estará resolvido quando respondermos às seguintes questões:

1. Quantas senhas de 4 dígitos podemos formar a partir dos elementos de M que começam com número ímpar?
2. Quantas senhas de 4 dígitos podemos formar a partir dos elementos de M que terminam com número par?
3. Quantas senhas de 4 dígitos podemos formar a partir dos elementos de M que começam com número ímpar e terminam com número par?

Isto porque os conjuntos P e Q não são disjuntos (eventos mutuamente exclusivos), logo, devemos fazer a seguinte conta:

$$\#(P \cup Q) = \#P + \#Q - \#(P \cap Q). \quad (2.4)$$

Perceba que diminuir a quantidade $\#(P \cap Q)$ na *Equação 2.4* corresponde a descontar os elementos que foram contados duas vezes.

Procedamos, então, ao cálculo das quantidades $\#P$, $\#Q$ e $\#(P \cap Q)$. Lembremo-nos de que estamos trabalhando no universo de senhas (vetores) com 4 entradas, isto é, estamos olhando para elementos da forma (x_1, x_2, x_3, x_4) , com $x_i \in M$, $i = 1, \dots, 4$. Para o cálculo de $\#P$, devemos percorrer as seguintes etapas:

- (E_1) escolha de um elemento x_1 para a ocupação da primeira entrada da senha, entre os elementos de $\{1, 3, 5, 7, 9\}$;
- (E_2) escolha de um elemento x_2 para a ocupação da segunda entrada da senha, entre os elementos de M ;
- (E_3) escolha de um elemento x_3 para a ocupação da terceira entrada da senha, entre os elementos de M ;
- (E_4) escolha de um elemento x_4 para a ocupação da quarta entrada da senha, entre os elementos de M ;

Como as etapas E_1, \dots, E_4 são independentes, pelo princípio multiplicativo de contagem, temos que

$$\#P = 5 \cdot 10^3.$$

De forma análoga,

$$\#Q = 5 \cdot 10^3 \text{ e } \#(P \cap Q) = 5^2 \cdot 10^2.$$

Finalmente, pela *Equação 2.4*, temos que:

$$\#(P \cup Q) = 5 \cdot 10^3 + 5 \cdot 10^3 - 5^2 \cdot 10^2 = 7.500.$$

Portanto, dado o conjunto

$$\Omega = \{\text{senhas de quatro dígitos formadas a partir do conjunto } M\},$$

acabamos de contar os elementos do subconjunto $\Lambda \subset \Omega$, cujos elementos apresentam a característica descrita abaixo:

$$\Lambda = \{s \in \Omega; s \text{ começa por número ímpar ou termina por número par}\}.$$

A próxima seção dedica-se a responder a outra questão geral feita no início deste capítulo.

2 Contando elementos no conjunto das partes de um conjunto com cardinalidade finita

No início deste capítulo, levantamos um questionamento geral motivados pela situação apresentada no *Exemplo 2.2*: como contar a quantidade de subconjuntos de um conjunto dado que possuem uma determinada cardinalidade.

Considere, inicialmente, o conjunto $\Omega_0 = \{a, b, c\}$. Temos que $\#(\Omega_0) = 3$. Nesse caso, conseguimos listar todos os elementos de $\wp(\Omega_0)$, o conjunto dos subconjuntos de Ω_0 :

$$\wp(\Omega_0) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}.$$

Note que os elementos de $\wp(\Omega_0)$ são conjuntos e que $\#(\wp(\Omega_0)) = 8$, isto é, o número total de conjuntos elementos de $\wp(\Omega_0)$ é 8. Além disso, há em $\wp(\Omega_0)$ um conjunto vazio, 3 conjuntos com 1 elemento, 3 conjuntos com 2 elementos e 1 conjunto com 3 elementos.

A partir de agora, considere um conjunto finito Ω qualquer com n elementos. Denotaremos por $\binom{n}{p}$ o número de subconjuntos que podemos formar a partir dos elementos de Ω com exatamente p elementos, em que p é um número inteiro maior ou igual a 0 e menor ou igual a n .

Exemplo 2.6. Se considerarmos o conjunto Ω_0 do início desta seção, então o número $\binom{3}{2}$ é a quantidade de subconjuntos de Ω_0 que possuem exatamente 2 elementos, isto é, $\binom{3}{2} = 3$.

Estabelecida essa notação, é preciso agora descobrir que número natural está associado a cada $\binom{n}{p}$. Nesse sentido, precisamos de uma noção ainda não apresentada até agora, a de factorial de um número natural.

Definição 2.4. Dado um número natural n , definimos

$$0! = 1! = 1$$

e

$$n! = 1 \cdot 2 \cdot \dots \cdot (n-1) \cdot n, \forall n \geq 2.$$

A expressão “ $n!$ ” lê-se “ n -fatorial” ou “fatorial de n ”.²

Essa definição nos permitirá acelerar algumas contas nos processos de contagem. Dela, é possível concluir imediatamente que

$$\begin{aligned} n! &= n \cdot (n-1)! \\ &= n \cdot (n-1) \cdot (n-2)! \\ &\quad \vdots \\ &= n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1 \end{aligned} \tag{2.5}$$

e que,

$$\frac{n!}{m!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-(m-1)) \cdot m!}{m!} = n \cdot (n-1) \cdot \dots \cdot (m+1), \tag{2.6}$$

Sempre que $m < n$.

O próximo exemplo ilustra o uso dessas propriedades.

Exemplo 2.7. Pela Propriedade (2.5), temos que

$$\frac{2010!}{2010} = \frac{2010 \cdot 2009!}{2010} = 2009!$$

e, pela Propriedade (2.6),

² Observe que, para qualquer n dado, $n!$ é um número natural.

$$\frac{5!}{3!} = 5 \cdot 4 = 20.$$

É comum na literatura que trata de problemas de contagem a denominação de arranjo para designar vetores sem elementos repetidos. Não faremos aqui uso desta terminologia. A proposição abaixo apresenta uma contagem para vetores que apresentam essa característica.

Proposição 2.1. *Seja Ω um conjunto tal que $\#\Omega = n$. Então a quantidade de vetores de Ω^p , $0 \leq p \leq n$, que possui todas as coordenadas distintas, é $\frac{n!}{(n-p)!}$.*

Demonstração: De fato, queremos contar vetores que tenham o seguinte formato:

$$(w_1, w_2, \dots, w_p), \text{ onde } w_i \neq w_j, \forall i, j = 1, 2, \dots, p.$$

Assim, para a escolha de w_1 temos n possibilidades, para a escolha de w_2 temos $n - p + 1$ possibilidades, e assim por adiante, até a escolha de w_p , em que teremos $n - 1$ possibilidades. Aplicando o princípio multiplicativo de contagem, teremos

$$n \cdot (n - 1) \cdot \dots \cdot (n - p + 1) = \frac{n!}{(n - p)!}$$

possibilidades de vetores com as especificações desejadas. ♦

A questão que se coloca é a de como determinar a quantidade de subconjuntos de Ω que possuem exatamente p elementos. Isto é, dispomos de n elementos e queremos formar agrupamentos com exatamente p deles, todos distintos. Se pensarmos em vetores, já vimos que conseguiríamos criar $\frac{n!}{(n - p)!}$. Mas isso não responde ao nosso problema, pois estamos falando de elementos que são conjuntos e não vetores.

Como vimos, os vetores $(x_1, x_2, x_3, \dots, x_p)$ e $(x_2, x_1, x_3, \dots, x_p)$ são distintos, enquanto que os conjuntos $\{x_1, x_2, x_3, \dots, x_p\}$ e $\{x_2, x_1, x_3, \dots, x_p\}$ são iguais. Por essa razão, o número $\frac{n!}{(n - p)!}$ é maior do que o que queremos, uma vez que corresponde à quantidade de vetores com o formato desejado. Para isso, precisamos descontar

todas as possíveis **permutações**. Fazendo isso, é como se estivéssemos eliminando os conjuntos iguais que foram contados como diferentes.

Definição 2.5. Dados dois vetores $v = (v_1, v_2, \dots, v_k)$ e $w = (w_1, w_2, \dots, w_k)$ pertencentes a um conjunto Ω^k , diremos que v é uma permutação de w , ou vice-versa, se as entradas do vetor v forem as mesmas entradas do vetor w , a menos de ordem.

Como exemplo, o vetor $(1, 3, 5)$ é uma permutação do vetor $(3, 1, 5)$. Mais geralmente, podemos considerar a seguinte proposição:

Proposição 2.2. *Considere o conjunto Ω^p . Se $w \in \Omega^p$ é um vetor com todas as entradas distintas, então w admite $p!$ permutações.*

Demonstração: De fato, a partir dos elementos do conjunto $A = (w_1, w_2, \dots, w_p)$, formado pelas entradas do vetor w , desejamos construir uma permutação de w , isto é, um novo vetor $v = (v_1, v_2, \dots, v_p)$ que resulte do embaralhamento das entradas de w . Temos, assim, p maneiras distintas de escolher v_1 , a primeira entrada de v , $p - 1$ maneiras distintas de escolher v_2 , a segunda entrada de v , e assim por diante, até que, para a escolha de v_p , a p -ésima entrada de v , teremos uma única maneira de fazê-lo. Utilizando o princípio multiplicativo de contagem, concluímos que há $p!$ maneiras distintas de fazer o serviço. ♦

Note que a Proposição 2.2 é um caso particular da Proposição 2.1.

Retornando à questão central desta seção, qual seja, a contagem da quantidade de subconjuntos de um conjunto que possuem um dada cardinalidade, basta dividir a quantidade de vetores pela quantidade de cópias permutadas de cada vetor. O resultado pode ser interpretado como a quantidade desejada, isto é,

$$\binom{n}{p} = \frac{\frac{n!}{n-p}}{p!} = \frac{n!}{(n-p)! p!}.$$

Acabamos, assim, de demonstrar a

Proposição 2.3. Dado um conjunto Ω com n elementos, o número $\binom{n}{p}$ de subconjuntos de Ω que possuem exatamente p elementos é dado pela expressão

$$\binom{n}{p} = \frac{n!}{(n-p)! p!}. \quad (2.7)$$

Essa proposição responde à questão levantada no início desta seção. Em particular, a resposta para a questão posta no Exemplo 2.2 é o número

$$\binom{40}{3} = \frac{40!}{(40 - 3)!3!} = \frac{40 \cdot 39 \cdot 38 \cdot 37!}{37!3 \cdot 2 \cdot 1} = 9.880.$$

3 Números combinatórios

Na seção anterior, aprendemos a calcular o valor do número $\binom{n}{p}$, que, até então, nada mais era que uma notação para o número de subconjuntos com p elementos formados a partir de um conjunto de n elementos. Números assim, definidos através da Equação 2.7, são chamados de **números combinatórios** ou **combinações**.

Os números combinatórios apresentam entre si importantes propriedades. A seguir, listamos as principais delas:

Propriedade 2.1. $\binom{n}{p} = \binom{n}{n-p}$, para todo $n \in N$.

Demonstração: De fato,

$$\binom{n}{p} = \frac{n!}{(n-p)! p!} = \frac{n!}{(n-(n-p))! (n-p)!} = \binom{n}{n-p},$$

para todo $0 \leq p \leq n$. ♦

Como casos particulares da Propriedade 2.1, temos

$$\binom{n}{0} = \binom{n}{n} = 1, \text{ para todo } n \in N,$$

e, de volta ao conjunto Ω_0 apresentado no início desta Seção,

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

Isto é, Ω_0 possui três subconjuntos com exatamente dois elementos e três subconjuntos com exatamente um elemento.

Propriedade 2.2. (Relação de Stiffel) Para todo $n \in \mathbb{N}$, temos

$$\binom{n}{p} + \binom{n}{p+1} = \binom{n+1}{p+1}, \text{ para todo } 0 \leq p < n.$$

Demonstração: De fato,

$$\begin{aligned} \binom{n}{p} + \binom{n}{p+1} &= \frac{n!}{(n-p)! p!} + \frac{n!}{(n-(p+1))! (p+1)!} \\ &= \frac{n!}{(n-p)(n-p-1)! p!} + \frac{n!}{(n-p-1)! (p+1) p!} \\ &= \frac{n! [p+1+n-p]}{(n-p)(n-p-1)! (p+1) p!} \\ &= \frac{n!(n+1)}{(n-p)! (p+1)!} \\ &= \frac{(n+1)!}{[(n+1)-(p+1)]! (p+1)!} \\ &= \binom{n+1}{p+1}, \end{aligned}$$

para todo $0 \leq p \leq n$. ♦

A Propriedade 2.3 a seguir será de fundamental importância para justificarmos a afirmação que fizemos na Seção 1, sobre a cardinalidade do conjunto das partes de um conjunto finito.

Para bem compreendê-la, entretanto, torna-se útil introduzir a noção de somatório, que nada mais é que uma notação para representar somas envolvendo uma quantidade inumerável de parcelas, particularmente útil quando lidamos com somas envolvendo muitas parcelas.

Para isso, considere $x = (x_1, x_2, \dots, x_n)$ um vetor carregando n informações numéricas quaisquer, não necessariamente distintas. Representaremos a soma

$$x_1 + x_2 + \dots + x_n = : \sum_{i=1}^n x_i. \quad (2.8)$$

Na Expressão (2.8), usamos a letra grega maiúscula *sigma* (Σ) para representar, de forma compacta, a soma. Na parte inferior de Σ , encontra-se um contador (no caso, representado pela letra i) que varia de 1 até n , indicando que todos os números, a começar daquele que se encontra na entrada 1 até o que se encontra na entrada n do vetor x , devem ser somados. A expressão x_i é uma representação genérica da informação constante da entrada i , $1 \leq i \leq n$, do vetor x . Dizemos que essa expressão é o **somando** da soma.³

Exemplo 2.8. Considere um vetor $y = (y_1, y_2, \dots, y_{10})$, tal que $y_j = 2^j$, para todo $j = 1, 2, \dots, 10$. Então,

$$\sum_{j=1}^{10} y_j = \sum_{j=1}^{10} 2^j = 2^1 + 2^2 + \dots + 2^{10}$$

e

$$\sum_{j=4}^7 y_j = \sum_{j=4}^7 2^j = 2^4 + 2^5 + 2^6 + 2^7.$$

Propriedade 2.3. Para todo $n \in \mathbb{N}$,

$$\sum_{p=0}^n \binom{n}{p} = 2^n.$$

³ Essa expressão compacta da soma costuma ser denominada de somatório. Lê-se a expressão assim: $\sum_{i=1}^n x_i$ somatório de x_i , i variando de 1 até n .

Demonstração: Para a demonstração dessa propriedade, usaremos um argumento chamado de indução. Para $n = 0$, temos que:

$$\sum_{p=0}^n \binom{n}{p} = \binom{0}{0} = 1 = 2^0.$$

Suponha que a propriedade seja verdadeira para $n = k$, isto é,

$$\sum_{p=0}^k \binom{k}{p} = 2^k.$$

Mostraremos que, sob essa hipótese,⁴ a propriedade será válida para o caso $n = k + 1$. De fato,

$$\begin{aligned} \sum_{p=0}^{k+1} \binom{k+1}{p} &= \binom{k+1}{0} + \sum_{p=1}^k \binom{k+1}{p} + \binom{k+1}{k+1} \\ &= \binom{k+1}{0} + \sum_{p=1}^k \left[\binom{k}{p-1} + \binom{k}{p} \right] + \binom{k+1}{k+1} \\ &= 1 + \sum_{p=1}^k \binom{k}{p-1} + \sum_{p=1}^k \binom{k}{p} + 1 \\ &= 1 + \sum_{p=0}^{k-1} \binom{k}{p} + (2^k - 1) + 1 \\ &= 1 + (2^k - 1) + (2^k - 1) + 1 \\ &= 2 \cdot 2^k \\ &= 2^{k+1}. \end{aligned}$$

Fica, assim, provado que a propriedade é válida para todo $n \in \mathbb{N}$. ♦

⁴ Essa hipótese é também conhecida como **hipótese de indução**.

Na demonstração da Propriedade 2.3, foi necessário o uso das propriedades 2.1 e 2.2, além de algumas manobras algébricas, para que ficasse claro o uso da hipótese de indução. Como consequência imediata dessas propriedades, temos:

Proposição 2.4. *Se Ω é um conjunto de cardinalidade n , então a cardinalidade do conjunto $\wp(\Omega)$ é 2^n .*

Demonstração: Como o conjunto $\wp(\Omega)$ é formado por todos os subconjuntos de Ω , então basta calcularmos a quantidade de subconjuntos com cardinalidades 0, 1, ..., n e a soma dessas quantidades será o número de elementos do conjunto $\wp(\Omega)$. Assim, pela Propriedade 2.3, temos que:

$$\#\wp(\Omega) = \sum_{p=0}^n \binom{n}{p} = 2^n.$$

E é por essa razão que também usamos a expressão 2^Ω para designar o conjunto $\wp(\Omega)$, quando Ω possui cardinalidade finita. ◆

Exercícios

1. Um conjunto Ω possui 10 elementos. Quantos subconjuntos de Ω possuem exatamente 6 elementos?
 - a) 210.
 - b) 420.
 - c) 630.
 - d) 60.
 - e) 360.

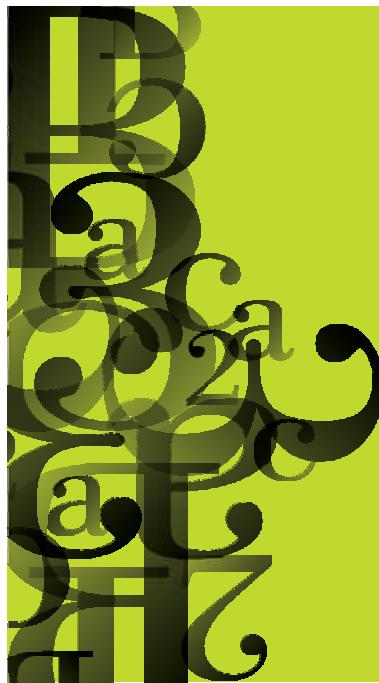
2. (IFG) Quantos números de quatro algarismos distintos maiores que 3.000 se pode formar com os algarismos 0, 1, 2, 3, 4 e 5 ?
 - a) 180.
 - b) 60.
 - c) 150.
 - d) 120.
 - e) 160.

3. (Cespe/UnB) A quantidade de anagramas que podem ser formados com a palavra CUTIA e que começam e terminam com consoante é igual a
- a) 6.
 - b) 10.
 - c) 12.
 - d) 18.
4. (Cespe/UnB) Uma escola possui 10 professores de matemática, 7 de ciências e 5 de português. A direção da escola pretende colocar o comando de uma excursão para seus alunos a um grupo formado por 2 professores de matemática, 3 de ciências e 2 de português. Neste caso, a quantidade de grupos distintos de professores que poderão ser formados para comandar a excursão é igual a
- a) 350.
 - b) 4.200.
 - c) 9.450.
 - d) 15.750.
5. (Cespe/UnB) Sabendo que $2^{14} = 16384$, então o somatório é
- $$\sum_{n=3}^{14} \binom{14}{n}$$
- a) 8.192.
 - b) 16.278.
 - c) 16.369.
 - d) 32.768.
6. A quantidade de anagramas distintos que se pode formar com a palavra AMBIENTE é igual a
- a) 20.160.
 - b) 25.200.
 - c) 32.400.
 - d) 40.320.

7. O número de anagramas da palavra MATEMÁTICA que começam por consoante é
- a) 75.600.
 - b) 90.720.
 - c) $10!$.
 - d) $5!$.
 - e) 30.240.

CAPÍTULO III

ANÁLISE EXPLORATÓRIA DE DADOS



Análise exploratória de dados

Estudaremos, neste capítulo, algumas medidas que podem ser associadas a conjuntos, ou a vetores que carregam informações, que, de alguma maneira, dão ideia do seu comportamento global. Essa ideia de comportamento também pode ser capturada a partir de um olhar atencioso para os instrumentos gráficos comumente utilizados para representar e comunicar essas informações.

Recordemos que, conforme definido no Capítulo 2, se $x = (x_1, x_2, \dots, x_k)$ um vetor carregando k dados numéricos, então

$$\sum_{i=1}^k x_i = x_1 + x_2 + \dots + x_k, \quad (3.1)$$

e podemos adaptar todas as propriedades usuais da soma para essa notação.

Essa notação será particularmente útil para que possamos nos expressar de forma compacta.

A seguir, introduziremos algumas medidas que apresentam informações resumidas de um conjunto ou de um vetor de dados. Essas medidas isoladamente não dizem muito, mas podem se tornar valiosos instrumentos para o trabalho estatístico que introduziremos no Capítulo 4.

1 Medidas resumo

Quando lidamos com um conjunto finito de informações, muitas vezes necessitamos de uma medida que as resuma. Numa escola, por exemplo, é uma prática comum, às vezes uma necessidade, procurar saber como está o aproveitamento dos seus estudantes. O gerente de uma padaria, num outro exemplo, pode estar em dúvida sobre quais as quantidades que devem ser produzidas de cada pão produzido e comercializado ali.

No capítulo anterior, vimos que a cardinalidade é uma medida que pode ser associada a conjuntos discretos e que conta a variedade de informação ali presente. Dependendo da estrutura do conjunto, outras medidas podem ser associadas a ele, por exemplo, as medidas de comprimento, de área e de volume. Todas essas medidas permitem a comparação de um conjunto com outros de mesma natureza.

No que segue, apresentamos algumas medidas estatísticas clássicas que possibilitam a comparação entre conjuntos de informações, as chamadas **medidas resumo**. Trata-se de números que associamos aos conjuntos e que, de alguma maneira, descrevem o comportamento global de seus elementos.

Classificamos as medidas resumo em dois grupos: as medidas de **posição** e as medidas de **dispersão**. Nas próximas subseções, detalharemos as medidas mais conhecidas de cada um desses grupos.

1.1 Medidas de posição

Das medidas de posição, a mais conhecida de todas é a **média**.

Definição 3.1. Seja X um conjunto de cardinalidade finita, isto é, $X = \{x_1, x_2, \dots, x_n\}$. Dizemos que a média¹ dos elementos de X é o número \bar{x} definido por

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}. \quad (3.2)$$

Em outras palavras, calcular o número médio de um conjunto numérico finito significa colher toda a informação trazida pelos elementos do conjunto e distribuir uniformemente essa informação para todos os elementos do conjunto. Essa distribuição uniforme fica caracterizada na Equação 17 pela divisão do conjunto da informação pela cardinalidade do conjunto X .

No entanto, quando trabalhamos com um conjunto de dados que eventualmente se repetem, a cardinalidade do conjunto tende a perder o sentido no cálculo da média, uma vez que cada elemento assume uma importância ou peso diferenciado diante dos demais. Nessa situação, é fundamental que, em vez de trabalhar com o conjunto que representa os dados disponíveis, trabalhemos com o **rol** desses dados.

Definição 3.2. Considere o vetor $x = (x_1, x_2, \dots, x_k)$, que carrega os dados resultantes da observação de algum fenômeno. Diremos que o rol desses dados é o vetor $\tilde{x} = (x_{(1)}, x_{(2)}, \dots, x_{(k)})$ tal que, para todo $1 \leq i \leq k$, tem-se $x_{(i)} = x_j$ para algum $1 \leq j \leq k$ e, além disso, ocorre

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(k)}$$

Ou seja, um rol é uma lista ordenada de todos os resultados da observação de algum fenômeno, ainda que repetidos.

Exemplo 3.1. A fim de melhorar a produção de leite bovino, um fazendeiro deseja suplementar a alimentação das suas matrizes leiteiras. Antes, contudo, observou a quantidade produzida por suas 9 matrizes num determinado dia. O vetor $m = (m_1, m_2, \dots, m_9) = (8, 5, 7, 12, 8, 5, 5, 7, 10)$ é o resultado do

¹ A média apresenta nesta definição também é conhecida como **média aritmética**.

processo de coleta, em que m_i representa a quantidade de litros colhidos na matriz do piquete i . O correspondente rol dessa situação será o vetor ordenado $\tilde{m} = (\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_9) = (m_{(1)}, m_{(2)}, \dots, m_{(9)}) = (5, 5, 5, 7, 7, 8, 10, 12)$. Qual a produção média de leite daquele dia?

No exemplo citado, poderíamos dizer que o conjunto das quantidades produzidas pelas matrizes leiteiras do fazendeiro naquele dia seria $\{5, 7, 8, 10, 12\}$, em que as quantidades são dadas em litros. Esse conjunto mostra a diversidade de resultados das observações feitas. A fim de calcular a média da produção, não poderíamos simplesmente utilizar a fórmula 17, uma vez que temos informações adicionais sobre a **frequência**² com que cada um desse elementos apareceu no processo de coleta de dados. Para resolver esse problema, é sensato aplicar o mesmo raciocínio da fórmula 17, porém considerando os elementos do rol. Procedendo assim, a média \bar{m} será

$$\bar{m} = \frac{\sum_{i=1}^9 m_i}{9} = \frac{\sum_{i=1}^9 \tilde{m}_i}{9} = \frac{5 \cdot 3 + 7 \cdot 2 + 8 \cdot 2 + 10 \cdot 1 + 12 \cdot 1}{9} = 7,444\dots$$

Assim, podemos refinar nosso conceito de média.

Definição 3.3. Seja X um conjunto finito que representa a variedade de informação num processo de coleta de dados, isto é, $X = \{x_1, x_2, \dots, x_n\}$, em que cada elemento x_i possui uma frequência f_i no rol que contém as informações do processo. Então, a média \bar{x} é dada por

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i}. \quad (3.3)$$

A Definição 3.3 é mais geral que a Definição 3.1, pois leva em consideração o peso de cada elemento na totalidade da informação. Por essa razão, a média obtida por esse processo é também conhecida como **média ponderada**.

² Entenderemos por **frequência** o número de vezes que um dado aparece num rol de informações.

A fim de evitar confusões relacionadas ao conceito de média, podemos defini-la simplesmente assim:

Definição 3.4. Seja $x = (x_1, x_2, \dots, x_n)$ um vetor de informações obtidas de um processo de coleta de dados. Define-se a média \bar{x} como sendo

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

Tal como na Definição 3.3, a média na Definição 3.4 leva em consideração o peso de cada informação no cálculo da média, na medida em que considera no somando todas as entradas do vetor, e isso é equivalente a considerar o conjunto que representa a variedade de informações e as frequências de seus respectivos elementos.

No *Exemplo 3.1*, vimos que essa medida resumo pode ser obtida analisando o rol das observações do processo. E é também observando o rol que definiremos outras importantes medidas descritoras do comportamento conjunto das observações: a **mediana** e a **moda**.

Definição 3.5. Dizemos que a mediana de um rol de dados $x = (x_{(1)}, x_{(2)}, \dots, x_{(k)})$ é a medida que centraliza o rol, isto é, qualquer número x_{med} tal que

$$x_{med} \leq 50\% \text{ dos dados} \text{ e } x_{med} \geq 50\% \text{ dos dados.}$$

Como consequência dessa definição, podemos ter

$$x_{med} = \begin{cases} x_{\left(\frac{k+1}{2}\right)}, & \text{se } k \text{ é ímpar} \\ \frac{x_{\left(\frac{k}{2}\right)} + x_{\left(\frac{k}{2}+1\right)}}{2}, & \text{se } k \text{ é par} \end{cases}$$

como uma mediana para o rol x .

Exemplo 3.2. No *Exemplo 3.1*, o número $\tilde{m}_{med} = \tilde{m}_{(5)} = 7$ é o valor mediano do vetor \tilde{m} , que representa a quantidade de leite produzido por matriz num dia fixado. Isso significa que metade das observações são de quantidades menores ou iguais a 7 e metade são de quantidades maiores ou iguais a 7.

Frequentemente, em vez de dividir o rol em dois “pedaços” de mesmo tamanho, há o interesse em dividi-lo em quatro, ou em dez, ou em quantos pedaços forem.

Exemplo 3.3. Comunicadas as notas de 5 jurados sobre um dado quesito, o regulamento da apuração do desfile das escolas de samba de São Paulo em 2010 ordenava o descarte da menor e da maior nota. Pelas regras, um quesito que recebeu um rol de notas $z = (9,5, 9,75, 10, 10, 10)$ terá as notas $z_{(1)} = 9,5$ e $z_{(5)} = 10$ descartadas. Isto é, a informação carregada pelo vetor ordenado foi dividida em 5 partes e, só depois do refinamento das notas, foi considerada na apuração do resultado final.

Medidas que, como a mediana, dividem a informação do rol em partes iguais são chamadas de **separatrizes**.

Definição 3.6. Seja $x = (x_{(1)}, x_{(2)}, \dots, x_{(k)})$ um rol qualquer de informações. As medidas Q_1, Q_2, \dots, Q_n são chamadas de separatrizes de x se

$$Q_i \geq \frac{i}{n+1} \text{ dos dados e } Q_i \leq \frac{n-i+1}{n+1} \text{ dos dados, para todo } i = 1, 2, \dots, n.$$

Em particular, os **quartis** Q_1, Q_2 e Q_3 são separatrizes que, como o próprio nome sugere, dividem as informações em 4 blocos de mesmo tamanho. Para fazer isso, procedemos de maneira análoga ao que fizemos na determinação da mediana.

Exemplo 3.4. O vetor $d = (3, 2,5, 3,2, 1,7, 1,9, 3,1, 3,2, 1,8, 1,9, 2,7)$, que carrega a informação sobre os diâmetros da cabeça de um parafuso numa determinada linha de produção, possui como rol o vetor $\tilde{d} = (1,7, 1,8, 1,9, 1,9, 2,5, 2,7, 3, 3,1, 3,2, 3,2)$. Os três quartis Q_1, Q_2 e Q_3 são dados por:

$$Q_2 = \tilde{d}_{med} = 2,6,$$

$$Q_1 = \tilde{d}_3 = d_{(3)} = 1,9$$

e

$$Q_3 = \tilde{d}_7 = d_{(7)} = 3,1.$$

Finalmente, apresentamos a definição de moda.

Definição 3.7. Dado um vetor de informações $x = (x_1, x_2, \dots, x_k)$, dizemos que a moda x_{mo} de x é a informação de maior frequência, isto é, a informação que mais se repete.

Exemplo 3.5. No Exemplo 3.1, a moda é o número $\tilde{m}_{mo} = 5$.

1.2 Medidas de dispersão

Embora as medidas de posição sejam os primeiros descritores numéricos aplicados a um vetor que carrega determinada informação, elas nem sempre garantem uma boa representatividade dele. Isso porque a boa representatividade de um vetor está diretamente relacionada à **variabilidade** da informação que ele carrega. Para maior clareza, acompanhe o exemplo seguinte.

Exemplo 3.6. Numa escola, dois grupos de 4 estudantes cada um foram submetidos a metodologias distintas de apresentação de um conteúdo. Ao final do período, todos os estudantes fizeram a mesma prova sobre o assunto. Os resultados foram os seguintes:

TABLE 1 – NOTAS DO GRUPO A

Estudante	1	2	3	4
Nota	5	5	6	4

TABLE 2 – NOTAS DO GRUPO B

Estudante	1	2	3	4
Nota	10	0	10	0

Se calcularmos a média e a mediana das duas turmas, essas medidas coincidirão (faça as contas!). Isso significa que o efeito das duas metodologias foi o mesmo?

O conhecimento das medidas de posição, apenas, não nos oferece subsídio para fazermos uma avaliação séria sobre os efeitos das metodologias apresentadas. Há, pois, claramente, na Turma B uma heterogeneidade que essas medidas não foram capazes de captar, nem mesmo a comparação entre elas. Surge, assim, a necessidade de definir novas medidas capazes de melhorar nossa compreensão a respeito de situações como essa. Nesse sentido, estudaremos um conjunto de novas medidas, chamadas de medidas de dispersão: a **amplitude total**, o **desvio médio absoluto**, a **variância**, o **desvio padrão** e o **coeficiente de variação**.

Definição 3.8. A amplitude total $\Delta(x)$ de um vetor x é a diferença entre a maior e a menor de todas as informação contidas em suas entradas.

Em particular, se o vetor for um rol $x = (x_1, x_2, \dots, x_k)$, então sua amplitude será

$$\Delta(x) = x_{(k)} - x_{(1)}$$

Exemplo 3.7. No Exemplo 3.1, a amplitude total do vetor m é

$$\Delta(m) = \Delta(\tilde{m}) = \tilde{m}_{(k)} - \tilde{m}_{(1)} = 12 - 5 = 7$$

Isso significa que todos as informações do vetor m , consequentemente do vetor \tilde{m} , não escapam de um intervalo de comprimento 7.

Embora dê uma ideia da dispersão das informações contidas no vetor, a amplitude em si é pouco informativa e muito afetada por **outliers** ou valores extremos. Isso porque foi uma medida construída com base apenas em duas informações carregadas pelo vetor.

Uma maneira de contornar esse problema é pensar numa medida para quantificar a dispersão, mas que envolva toda a informação para defini-la. E isso acontece no desvio médio absoluto. Para comprehendê-lo, iniciemos pelo significado da palavra desvio.

Consideremos um vetor de informações $x = (x_1, x_2, \dots, x_k)$ qualquer, cuja média é \bar{x} . Diremos que o desvio d_i que a informação x_i apresenta em relação à média \bar{x} é

$$d_i = x_i - \bar{x}$$

Claramente, vemos que o desvio em relação à média será negativo se a informação tiver valor numérico menor que a média e será positivo no caso contrário.

Um fato importante sobre os desvios em relação à média é o apresentado a seguir.

Proposição 3.1. *A soma de todos os desvios em relação à média d_i de um vetor de informações $x = (x_1, x_2, \dots, x_k)$ é zero. Isto é,*

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

Demonstração: De fato, usando a definição de média e propriedades do somatório, temos que

$$\begin{aligned} \sum_{i=1}^k d_i &= \sum_{i=1}^k (x_i - \bar{x}) = \sum_{i=1}^k x_i - k \bar{x} \\ &= \sum_{i=1}^k x_i - k \cdot \frac{\sum_{i=1}^k x_i}{k} = \sum_{i=1}^k x_i - \sum_{i=1}^k x_i = 0. \end{aligned}$$



Como consequência da propriedade provada acima, o desvio em relação à média não dá informação relevante quanto ao comportamento conjunto das informações do vetor. Apesar de sua utilidade com relação ao posicionamento de cada dado particular com respeito à média, é preciso definir uma medida de dispersão que traduza conjuntamente esse posicionamento para fins de comparação com outros vetores. Para tanto, se em vez do desvio em relação à média considerarmos a distância da informação à informação média, isto é, o desvio absoluto em relação à média, dado por

$$|d_i| = |x_i - \bar{x}|,$$

poderemos definir o desvio médio absoluto como a seguir.

Definição 3.9. Dado um vetor de informações $x = (x_1, x_2, \dots, x_k)$, o desvio médio absoluto $\delta(x)$ é definido como a média das distâncias que cada informação apresenta da informação média. Isto é,

$$\delta(x) = \frac{\sum_{i=1}^k |x_i - \bar{x}|}{k}.$$

Exemplo 3.8. Retornemos à situação abordada no Exemplo 3.6. Se $a = (5, 5, 6, 4)$ e $b = (10, 0, 10, 0)$ são os vetores que carregam as informações de rendimento dos estudantes dos grupos A e B, respectivamente, então, desde que $\bar{a} = \bar{b} = 5$, temos que

$$\delta(a) = \frac{|5 - 5| + |5 - 5| + |6 - 5| + |4 - 5|}{4} = \frac{1}{2}$$

e

$$\delta(b) = \frac{|10 - 5| + |0 - 5| + |10 - 5| + |0 - 5|}{4} = \frac{5 + 5 + 5 + 5}{4} = 5.$$

Repare que as notas do grupo A apresentaram menor oscilação em torno da média, ao contrário das notas do grupo B. O desvio médio absoluto capta esse comportamento à medida que se afasta de zero. Assim, quanto mais próximo de zero for o desvio médio absoluto, menor a oscilação dos dados em torno da média.

Apesar dessa importante interpretação, o desvio médio absoluto apresenta dificuldades de manipulação matemática devido à presença do módulo. Esse fato dificulta sua utilização em questões mais elaboradas da Estatística.

Em busca de uma medida capaz de captar a dispersão dos dados em torno da média com a mesma eficiência do desvio médio absoluto, mas que possa oferecer maiores possibilidades de manipulação matemática, apresentamos a variância, a mais conhecida das medidas de dispersão.

Definição 3.10. Dado um vetor de informações $x = (x_1, x_2, \dots, x_k)$, a variância $\sigma^2(x)$ é definida como a média dos quadrados da distância que cada informação apresenta da informação média. Isto é,

$$\sigma^2(x) = \frac{\sum_{i=1}^k |x_i - \bar{x}|^2}{k} = \frac{\sum_{i=1}^k (x_i - \bar{x})^2}{k}.$$

Exemplo 3.9. Ainda explorando a situação apresentada no Exemplo 3.6, sendo vetores $a = (5, 5, 6, 4)$ e $b = (10, 0, 10, 0)$ como no Exemplo 3.8, temos que

$$\sigma^2(a) = \frac{(5 - 5)^2 + (5 - 5)^2 + (6 - 5)^2 + (4 - 5)^2}{4} = \frac{1}{2}$$

e

$$\sigma^2(b) = \frac{(10 - 5)^2 + (0 - 5)^2 + (10 - 5)^2 + (0 - 5)^2}{4} = \frac{100}{4} = 25.$$

Vemos que é possível interpretar a variância analogamente ao desvio médio absoluto, já que a proximidade do zero implica menor variabilidade. Todavia, a escala dessas medidas fica alterada. De fato, podemos notar que $\delta(b) = 5$ pontos, enquanto que $\sigma^2(b)$ pontos ao quadrado. Isso, por si só, não representa problemas, uma vez que essas medidas são significativas em comparação com outro vetor de informações similares. No entanto, o uso da variância, apesar de muito comum, gera um desconforto por distorcer as escalas. No caso em tela, é complicado atribuir um sentido à escala “ponto ao quadrado”.

Pelas razões apresentadas, definimos o desvio-padrão como sendo a raiz quadrada da variância. Dessa forma, o desconforto gerado pela escala distorcida desaparece e a tratabilidade matemática permanece, junto com o principal, que é quantificar a variação do conjunto de dados em torno da média.

Definição 3.11. Dado um vetor de informações $x = (x_1, x_2, \dots, x_k)$, o desvio-padrão $\sigma(x)$ é definido como sendo a raiz quadrada da variância. Isto é,

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

Exemplo 3.10. Na situação apresentada no Exemplo 3.6, temos que

$$\sigma(a) = \sqrt{\sigma^2(a)} = \sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}} \approx 0,71 \text{ ponto}$$

e

$$\sigma(b) = \sqrt{\sigma^2(b)} = \sqrt{25} = 5 \text{ pontos.}$$

Para finalizar, apresentamos o coeficiente de variação.

Definição 3.12. Seja $x = (x_1, x_2, \dots, x_k)$ um vetor de informações, com média \bar{x} e desvio padrão $\sigma(x)$. Define-se o coeficiente de variação $CV(x)$ como sendo a razão

$$CV(x) = \frac{\sigma(x)}{\bar{x}}.$$

Essa é uma medida relativa que avalia o percentual de dispersão comparada com a média. Esse índice é denominado **variabilidade**. Uma das grandes vantagens dessa medida é a possibilidade de comparação de vetores de diferentes dimensões ou que carregam informações de naturezas diferentes, uma vez que ela é adimensional e não possui escala.

Exemplo 3.11. A equipe de nutricionistas de uma escola está incumbida de acompanhar a nutrição de 6 estudantes e, para isso, realizou um exame diagnóstico de dois diferentes aspectos: massa corporal e altura. As informações coletadas foram registradas nos vetores $m = (m_1, m_2, \dots, m_6)$ e $a = (a_1, a_2, \dots, a_6)$, em que os números m_i e a_i expressam, respectivamente, os dados de massa (em quilos) e altura (em centímetros) do estudante i , $i = 1, 2, \dots, 6$. Se $m = (68, 70, 86, 55, 75, 90)$ e $a = (170, 160, 164, 164, 170, 180)$, os estudantes apresentaram maior variabilidade na massa ou na altura?

Para responder à questão apresentada no Exemplo 3.11, precisamos comparar vetores que trazem informações de naturezas distintas. Para isso, usaremos o coeficiente de variação, que é uma medida independente de escala.

Fazendo as contas,³ é fácil ver que $\bar{m} = 74$ kg, $\bar{a} = 168$ cm, $\sigma(m) = 11,65$ kg e $\sigma(a) = 6,43$ cm. Logo,

$$CV(m) = \frac{\sigma(m)}{\bar{m}} = \frac{11,65 \text{ kg}}{74 \text{ kg}} = 0,157$$

³ Aqui, consideramos arredondamentos para duas casas decimais.

e

$$CV(a) = \frac{\sigma(a)}{\bar{a}} = \frac{6,43\text{ cm}}{168\text{ cm}} = 0,0383.$$

Portanto, o vetor das alturas apresentou uma variabilidade de 3,83%, que é significativamente inferior à variabilidade de 15,7% apresentada pelo vetor de massas corporais, e isso responde ao questionamento gerado no Exemplo 3.11.

Na próxima seção, estudaremos formas menos analíticas, mas não menos importantes, na busca de compreensão do comportamento coletivo de um conjunto ou vetor de informações.

2 Gráficos

Nesta seção, veremos alguns dos principais tipos de gráficos que resumem e comunicam informações. Alguns deles são bem conhecidos, devido a sua vasta utilização nos meios de comunicação. Outros, mais técnicos, são menos populares mas muito informativos do ponto de vista científico.

Comecemos por definir o que é esse objeto matemático tão valioso.

Definição 3.13. Dados dois conjuntos quaisquer A e B , chamamos de gráfico (ou relação) qualquer subconjunto não vazio de $A \times B$.

A princípio, essa definição parece estranha, pois, sempre que falamos de gráfico, nos vem à mente uma figura. Não estamos equivocados. É que alguns gráficos admitem uma representação geométrica num sistema de coordenadas, como é o caso de quando estabelecemos uma relação em duas dimensões no plano cartesiano. Gráfico, portanto, nada mais é que um conjunto de vetores que podem ser representados como um conjunto de pontos num sistema de coordenadas.

Exemplo 3.12. Considere o produto cartesiano $R \times R = R^2$ e o conjunto $M = \{(x, y) \in R^2; y = x\}$. Claramente, $M \subset R^2$. Uma apresentação gráfica para esse conjunto seria a Figura 3.1.

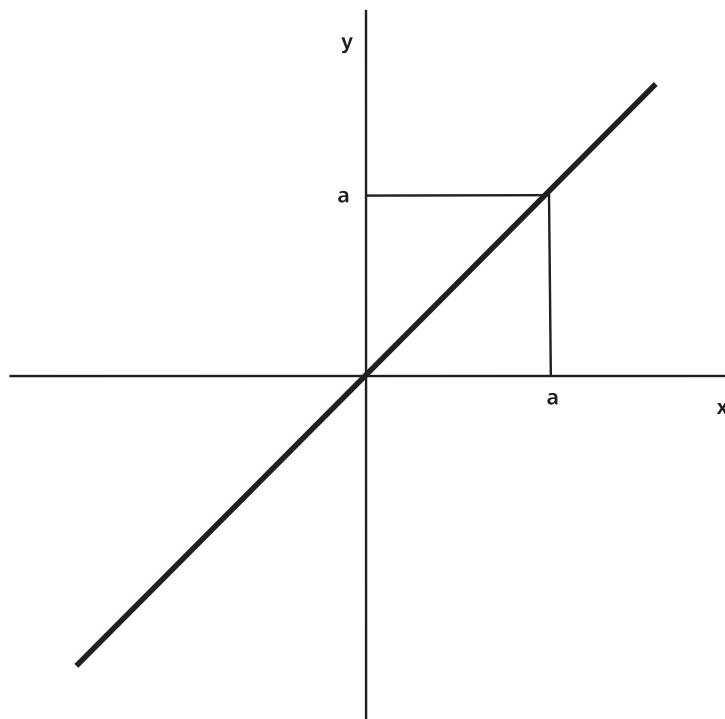


FIGURA 3.1 – GRÁFICO DA RELAÇÃO M

A representação de gráficos no plano cartesiano, como mostra a Figura 3.1, é bastante limitada. Na maioria das vezes, necessitamos visualizar relações entre conjuntos que apresentam características adicionais além do simples cruzamento ortogonal de seus elementos. Todavia, graças à proliferação dos recursos computacionais, existe hoje uma infinidade de tipos de gráficos à nossa disposição.

A utilização de um ou outro tipo de gráfico depende de múltiplos fatores, que vão desde as características dos fenômenos que originaram aqueles dados até o tipo de público que se deseja informar, passando pela experiência e domínio de técnicas apropriadas por parte do elaborador do gráfico.

No entanto, a utilização de recursos visuais deve ser feita cuidadosamente. Afinal, um gráfico desproporcional em suas medidas pode conduzir a conclusões equivocadas.

A Figura 3.2 apresenta um gráfico conhecido por gráfico de linhas. Esse nome se justifica por seu aspecto. O gráfico apresenta uma relação tempo versus número de habitantes (em milhões). Tecnicamente, foram apuradas apenas as medidas da população referentes aos anos de realização do censo demográfico. No entanto, os pontos do gráfico foram ligados com um segmento de reta para comunicar a ideia de crescimento linear dessa população nos períodos em que não foi possível sua contagem. Gráficos de linha como esse, em que uma das grandezas observadas é o tempo, também são conhecidos na literatura como **séries temporais**.

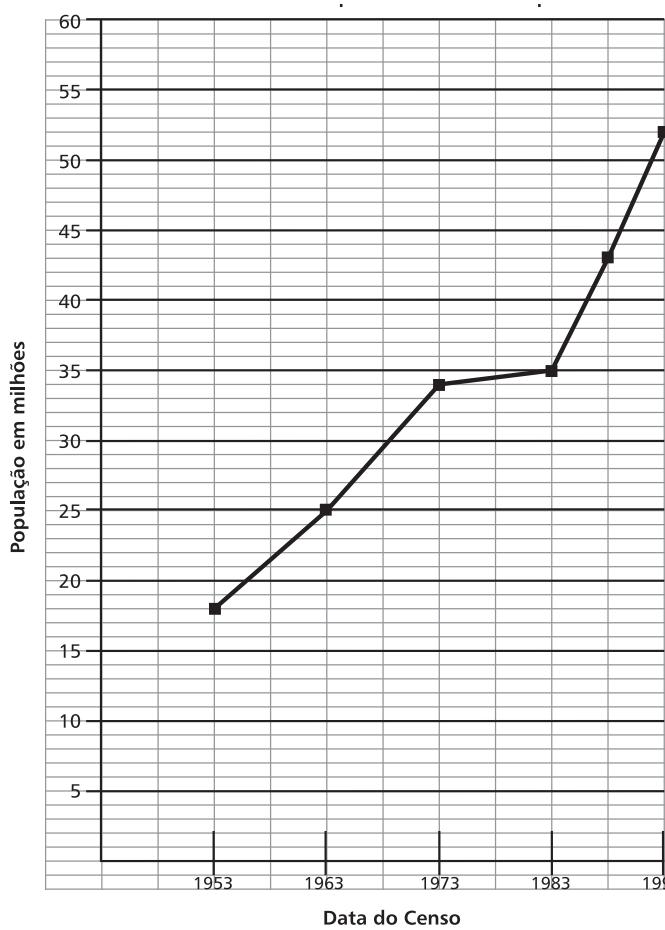


FIGURA 3.2 – CRESCIMENTO POPULACIONAL DO MUNICÍPIO A
Fonte: fictícia.

A Figura 3.3 exibe um tipo de gráfico muito popular, conhecido por **gráfico de setores**. Em geral, gráficos de setores comunicam bem informações que se encontram divididas por categorias.

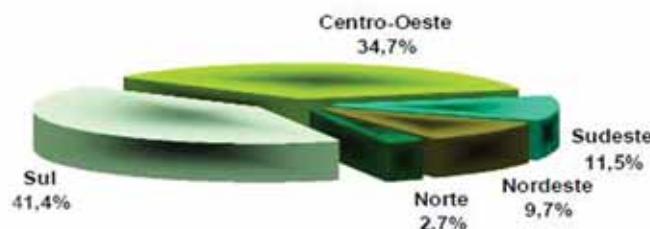


FIGURA 3.3 – CEREAIS, OLEAGINOSAS E LEGUMINOSAS. PARTICIPAÇÃO NA PRODUÇÃO SEGUNDO AS GRANDES REGIÕES.

Fonte: IBGE. *Primeiras estimativas da safra 2010.*

Hoje em dia, os programas de computadores fazem gráficos automaticamente, sem que nos preocupemos com nada, a não ser lançar os dados geradores. Mas é possível construir gráficos de setores manualmente, com o auxílio de compasso e transferidor. No caso da Figura 3.3, por exemplo, para definir o ângulo correspondente ao setor que irá representar cada categoria, basta fazer um cálculo simples de regra de três, já que toda a produção (100%) deve ser representada num círculo que possui 360° de ângulo interno, e as grandezas “ângulo do setor circular” e “percentual de participação na produção” são diretamente proporcionais.

O gráfico da Figura 3.4 é dito ser um **gráfico de barras**. Ele mostra como a população brasileira encontrava-se concentrada, em 2000, nos diferentes Estados. É possível observar que mais de 20% da população brasileira encontrava-se concentrada no Estado de São Paulo. Observe, também, que há barras de cores diferenciadas, o que permite comparar a distribuição da população pelo território em dois momentos distintos.

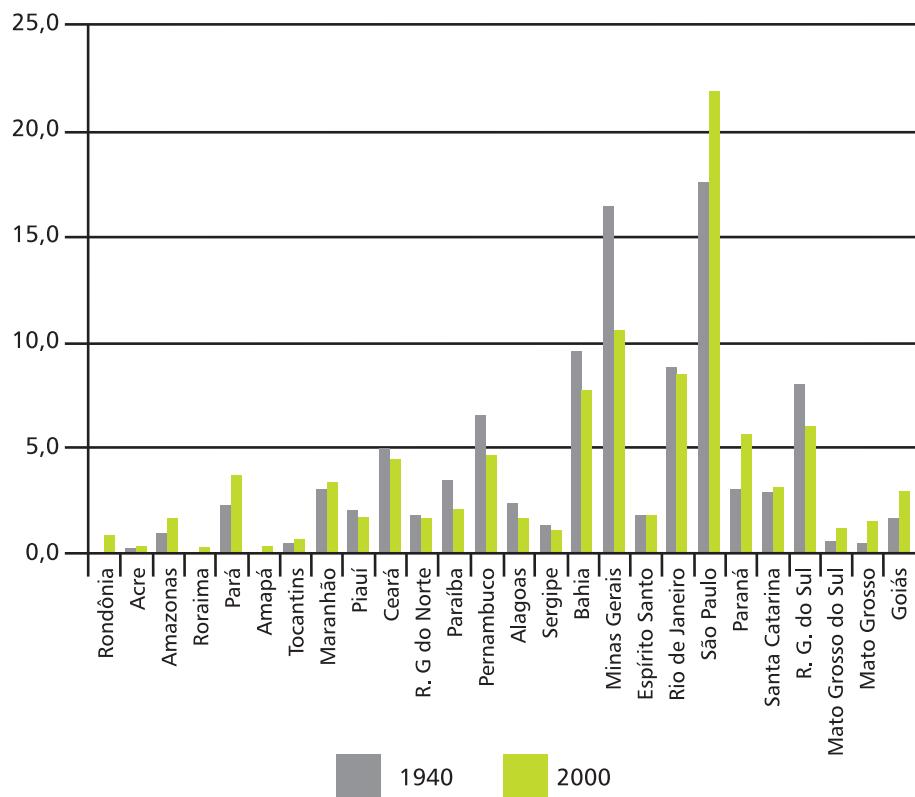


FIGURA 3.4 – DISTRIBUIÇÃO PERCENTUAL DA POPULAÇÃO, SEGUNDO AS UNIDADES DA FEDERAÇÃO - 1940/2000.
Fonte: IBGE. *Censo demográfico 1940/2000.*

Como dito anteriormente, existem vários tipos de gráficos, entre os quais podemos citar: cartogramas, pictogramas, climogramas, gráficos polares, tridimensionais, etc. A mídia explora frequentemente todos esses tipos de gráficos. Existem outros, mais técnicos, que costumam acompanhar trabalhos científicos. Veremos alguns destes últimos no Capítulo 4.

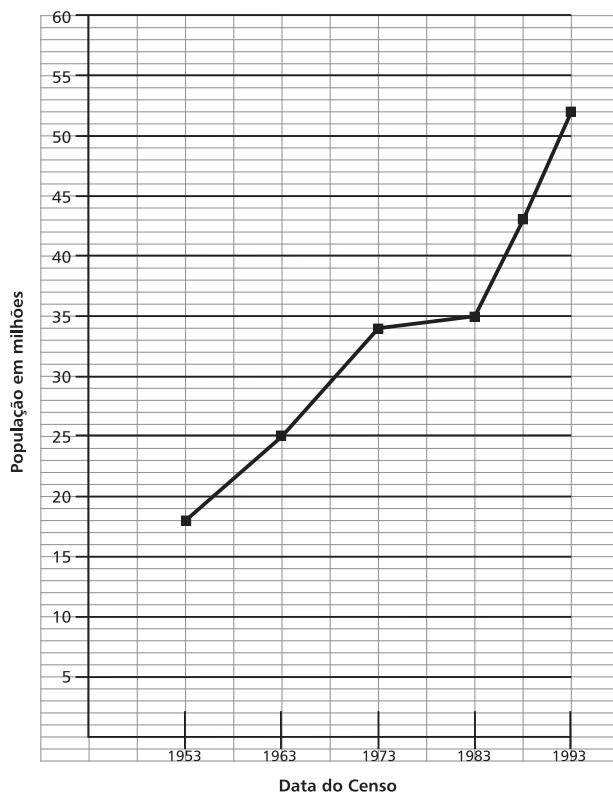
Exercícios

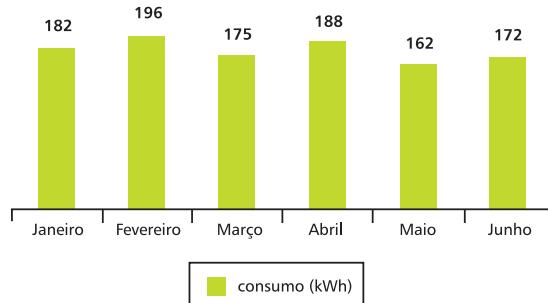
1. (IFB) Considere o gráfico a seguir, que se refere à população de um município, no período de 1953 a 1993.

O crescimento da população nesse município, entre 1983 e 1993, foi de

- a) 10 milhões.
- b) 17 milhões.
- c) 30 milhões.
- d) 35 milhões.
- e) 40 milhões.

2. (IFG) O gráfico a seguir mostra o consumo de energia elétrica, em kwh, numa residência, no período de janeiro a junho de 2008.





Com base nos dados do gráfico, assinale a alternativa correta.

- A variação percentual do consumo, no período de janeiro a fevereiro de 2008, foi de 6,19%.
 - A variação percentual do consumo, no período de maio a junho de 2008, foi de 6,17%.
 - A variação percentual do consumo, no período de fevereiro a março de 2008, foi de -9,71%.
 - A variação percentual do consumo, no período de março a abril de 2008, foi de 6,42%.
 - A variação percentual do consumo, no período de abril a maio de 2008, foi de -12,83%.
3. A tabela a seguir mostra a distribuição do consumo de energia elétrica em kWh nos meses de janeiro a junho de 2008 numa residência.

Meses	Consumo (KWH)
janeiro	240
fevereiro	154
março	101
abril	123
maio	146
Junho	157

Com base nos dados da tabela, assinale a alternativa incorreta:

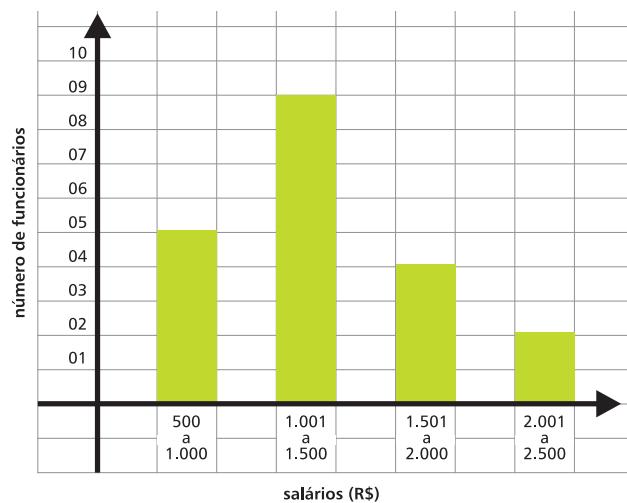
- a) O consumo médio mensal no primeiro trimestre de 2008 foi de 165 kwh.
- b) O consumo médio mensal no segundo trimestre de 2008 foi de 142 kwh.
- c) O consumo mediano de energia elétrica foi de 154 kwh.
- d) No primeiro trimestre de 2008, o consumo de energia elétrica em kwh foi decrescente.
- e) No segundo trimestre de 2008, o consumo de energia elétrica em kwh foi crescente.
4. A média de idade de um grupo de adolescentes é de 15 anos. Se dois desses jovens, com as idades de 14 e 17 anos, respectivamente, são substituídos por outros dois, de 15 e 16 anos de idade, então podemos afirmar que:
- a) A média de idade do grupo não se altera.
- b) A média de idade do grupo passa a ser de 15,5 anos.
- c) A média de idade do grupo passa a ser de 16 anos.
- d) A média de idade do grupo passa a ser de 14,5.
- e) Impossível calcular, uma vez que não foi informado o número de adolescentes do grupo.
5. (IFG) A tabela a seguir mostra a distribuição do consumo de água, em m³, nos meses de julho a dezembro de 2007 numa residência.

Meses	Consumo (m ³)
Julho	18
agosto	20
setembro	22
outubro	21
novembro	18
dezembro	24

Com base na tabela, assinale a alternativa incorreta.

- a) O aumento percentual do consumo no mês de agosto, comparado com o mês de julho, foi de 11,11%.
 - b) O consumo mediano mensal em litros foi de 20.000.
 - c) A taxa percentual de crescimento do maior consumo em relação ao menor consumo foi de 33,33%.
 - d) O consumo médio mensal em litros, nos meses de julho, agosto e setembro, foi de 20.000.
 - e) O consumo médio mensal em litros, nos meses de outubro, novembro e dezembro, foi de 21.000.
6. A média anual de uma escola para aprovação é 6,0. Sabendo que João tirou 7,0 no primeiro bimestre, 5,0 no segundo bimestre e 4,5 no terceiro bimestre, qual a nota mínima a ser alcançada por ele no quarto e último bimestre?
- a) 7,5.
 - b) 8,5.
 - c) 8,0.
 - d) 8,5.
 - e) 9,0.
7. Sejam $x = (x_1, x_2, \dots, x_n)$ e $y = (x_1 + c, x_2 + c, \dots, x_n + c)$, em que c é uma constante qualquer. Mostre que $\bar{y} = \bar{x} + c$.
8. Sejam $x = (x_1, x_2, \dots, x_n)$ e $y = (c \cdot x_1, c \cdot x_2, \dots, c \cdot x_n)$, em que c é uma constante qualquer diferente de zero. Mostre que $\bar{y} = c \cdot \bar{x}$.

9. (IFB) Uma balança encontra-se desregulada e fornece as medidas com um erro sistemático de 100 gramas a menos do que a medição real. Ao ser utilizada para medir n amostras de pacotes de 5 kg de arroz num supermercado, a diferença entre a média amostral produzida pela balança e a média amostral real será de
- 100 gramas.
 - $\frac{100}{n}$ gramas.
 - $100 \cdot n$ gramas.
 - 500 gramas.
 - $\frac{500}{n}$ gramas.
10. (IFB) O gráfico abaixo mostra os salários, em reais, dos 20 funcionários de uma loja.



Com base nesses dados, assinale a opção correta.

- A maioria dos funcionários recebe salários entre R\$1.000,00 e R\$1.500,00.
- Cinco funcionários recebem salários de R\$500,00.
- O maior salário recebido na loja é de R\$2.000,00.
- Há na loja um funcionário que recebe R\$1.250,00.
- Apenas 30% dos funcionários recebem salários superiores a R\$1.500,00.

11. (Cespe/UnB) Em estatística, duas medidas são de grande importância na análise de dados: medidas de tendência central e medidas de dispersão. Entre as medidas indicadas abaixo, são, respectivamente, medidas de dispersão e de tendência central:
- Mediana e média.
 - Média e moda.
 - Variância e moda.
 - Desvio-padrão e variância.
 - Desvio-padrão e desvio médio.
12. (Cespe/UnB) Em estatística, quando trabalhamos com medidas de tendência central, a medida que tem o mesmo número de valores abaixo e acima dela é:
- a média.
 - a mediana.
 - a média aritmética.
 - a média ponderada.
13. O cálculo do lucro unitário (L) de certo produto é dado por $L = 1,2.V - 0,8.C - 8,5$. Sabendo-se que o preço unitário de venda (V) desse produto é em média, R\$40,00, com desvio-padrão de R\$5,00, e que o custo (C) unitário é de R\$30,00, com desvio-padrão de R\$4,50, conclui-se que a média e o desvio-padrão do lucro unitário desse produto são, respectivamente,
- R\$12,50 e R\$4,80.
 - R\$15,50 e R\$7,00.
 - R\$12,50 e R\$5,90.
 - R\$15,50 e R\$4,80.
14. A tabela a seguir fornece o número mundial de acidentes fatais de aeronaves comerciais no período de 1985 a 1993.

Ano	1985	1986	1987	1988	1989	1990	1991	1992	1993
Acidentes	22	22	26	28	27	25	30	29	24

Com base nessas informações, determine a variância que esses dados apresentam.

15. (IFG) Dados os conjuntos numéricos $A = \{1000, 1001, 1002, 1003, 1004, 1005\}$ e $B = \{0,1, 2, 3, 4, 5\}$, podemos afirmar que:
 - a) O desvio-padrão de A é igual a mil vez o desvio-padrão de B .
 - b) O desvio-padrão de A é igual ao desvio-padrão de B .
 - c) O desvio-padrão de A é igual ao desvio-padrão de B , multiplicado pelo quadrado de 1.000.
 - d) O desvio-padrão de A é igual ao desvio-padrão de B , dividido por 1.000.
 - e) O desvio-padrão de A é igual ao quadrado do desvio-padrão de B .
16. (IFG) Os salários dos empregados da empresa X são 2% maiores dos que os da empresa Y para todos os empregados comparados individualmente. Com base nessas informações, é correto afirmar que:
 - a) O desvio-padrão dos salários dos empregados da empresa X é 2% menor do que o dos salários dos empregados da empresa Y .
 - b) O desvio-padrão dos salários dos empregados é o mesmo para ambas as empresas.
 - c) O desvio-padrão dos salários dos empregados da empresa X é 2% maior do que o dos salários dos empregados da empresa Y .
 - d) O desvio-padrão dos salários dos empregados da empresa X é igual ao desvio-padrão dos salários dos empregados da empresa Y , multiplicado por $(1,02)^2$.
 - e) Não há elementos para se compararem os desvios-padrão dos salários dessas empresas.

CAPÍTULO IV

ESTATÍSTICA DESCRIPTIVA



Estatística descritiva

Atualmente, grande parte da informação difundida, pelos meios de comunicação, e mesmo as produzidas no meio acadêmico, provém da análise de dados que foram coletados e, de alguma maneira, relacionados a outras informações conhecidas.

O Instituto Brasileiro de Geografia e Estatística (IBGE) é um órgão do governo brasileiro que realiza pesquisas intensivas e periódicas sobre o crescimento, comportamento, distribuição espacial e organização social da população brasileira, entre outras características desse conjunto de pessoas. As informações são coletadas e sistematizadas segundo critérios científicos robustos, e as conclusões advindas da análise dessas informações, quando comparadas com outras obtidas em um outro período de tempo, constituem valiosos instrumentos para o desenvolvimento das ciências humanas e sociais e, inclusive, para efeito de definição de políticas públicas.

O exemplo da Figura 4.1 a seguir apresenta uma comparação de informações colhidas em diferentes épocas.

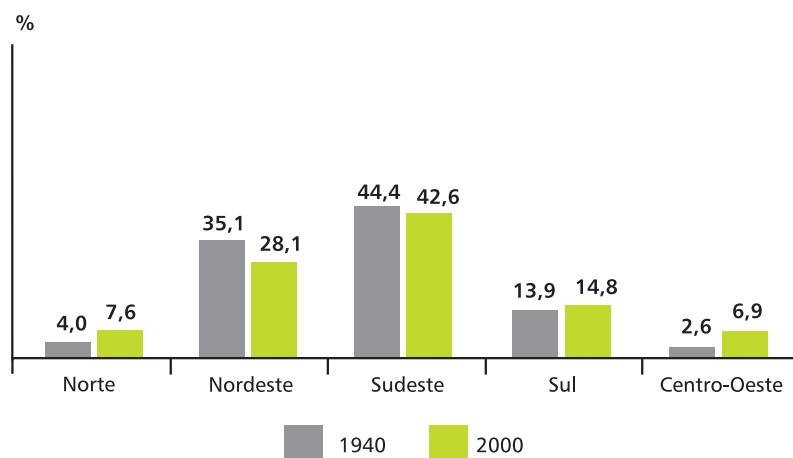


FIGURA 4.1 – DISTRIBUIÇÃO PERCENTUAL DA POPULAÇÃO, SEGUNDO AS GRANDES REGIÕES - 1940-2000

Fonte: IBGE. *Censo demográfico 1940-2000*.

Da análise dessas informações, podemos observar que as Regiões Sudeste e Nordeste apresentam as maiores participações da população relativa do País, ainda que passados 60 anos entre os períodos observados. Pode-se, todavia, inferir que no período de 1940 a 2000, as maiores taxas de crescimento ocorreram nas Regiões Centro-Oeste e Norte.

Neste capítulo, veremos que é a Estatística a ciência que se ocupa desse tipo de trabalho. Aprenderemos quais são as etapas do trabalho estatístico e os cuidados que devemos ter na coleta e no tratamento de informações para viabilizar a tomada de decisão. Veremos também a importância deste trabalho na comunicação de informações.

1 Conceitos básicos

De origem muito antiga, a Estatística teve suas primeiras atividades relacionadas com o recenseamento das populações agrícolas chinesas, com registros que datam de 2000 a.C.. Durante muito tempo, teve um caráter

meramente descritivo, mas, atualmente, impulsionada pelos avanços da tecnologia, ganhou *status* de ciência, e suas técnicas são cada vez mais difundidas e servem de suporte para o desenvolvimento de outras ciências.

Muitas vezes tratada como um ramo da matemática, a Estatística moderna possui métodos e técnicas próprias, daí o seu *status* de ciência, embora as ferramentas matemáticas lhe sejam essenciais.

Para nossos propósitos, entenderemos Estatística como um conjunto de métodos e técnicas que auxiliam a tomada de decisão, mediante a presença de incerteza.

A incerteza está, de alguma maneira, presente em todo o trabalho estatístico. Isso porque a maioria dos fenômenos que analisamos está permeada por ela, sejam fenômenos naturais, sociais, econômicos ou do comportamento humano. O tratamento quantitativo adequado à incerteza é obtido por meio do estudo de probabilidades, que introduzimos no Capítulo 5.

Mesmo mediante incertezas, o fato é que o tempo todo decisões devem ser tomadas e o estudo das técnicas da Estatística e de probabilidades auxiliam a minimizar seus riscos.

A Estatística pode ser dividida em duas partes distintas: a **Estatística Descritiva** e a **Estatística Inferencial**. O objetivo da Estatística Descritiva é resumir as principais características de um vetor de informações por meio de tabelas, gráficos e resumos numéricos. Já a tomada de decisão se apoia no uso da Estatística Inferencial, que vai além da mera descrição das informações e é resultado de uma análise mais profunda.

A Estatística Inferencial visa à busca por conclusões satisfatórias, mesmo diante de informações incompletas. O alicerce das técnicas da Estatística Inferencial está no estudo de **probabilidades**. Duas das importantes técnicas utilizadas nesse campo da Estatística são a **estimação de parâmetros**¹ e os **testes de hipóteses**.²

¹ A estimativa de parâmetros consiste em utilizar um conjunto de dados incompletos, ao qual iremos chamar de amostra, e nele calcular estimativas de quantidades de interesse. Essas estimativas podem ser pontuais (representadas por um único valor) ou intervalares.

² O fundamento dos testes de hipóteses é levantar suposições acerca de uma quantidade não conhecida e utilizar, também, dados incompletos para criar um critério de escolha.

Para se ter uma ideia, considere a situação seguinte:

Exemplo 4.1. *Um laboratório deseja verificar se uma nova droga aumenta a produção de testosterona em homens com idade acima de 35 anos. Ao aplicá-la a um grupo de 40 indivíduos, constatou-se que, depois de um período de tempo, a droga aumentou significativamente a quantidade do referido hormônio. Nessas condições, a nova droga deve ou não ser lançada no mercado?*

A questãoposta no Exemplo 4.1, que envolve a tomada de uma decisão, pode ser resolvida através da Estatística Inferencial. Uma outra situação similar é apresentada no Exemplo 4.2.

Exemplo 4.2. *Em uma fábrica de parafusos, a peça é considerada dentro da especificação, caso seu comprimento esteja no intervalo entre 4,8 cm e 5,2 cm. Os técnicos de controle de qualidade selecionam diariamente 100 parafusos fabricados e calculam o comprimento médio. Conhecendo a variabilidade nos tamanhos dos parafusos fabricados, caso o comprimento médio esteja abaixo de 4,99 cm ou acima de 5,01 cm, o processo será interrompido.*

Neste exemplo, espera-se que o comprimento médio de um subconjunto dos parafusos esteja dentro de um intervalo. Caso isso não ocorra, o processo de produção sofre interrupção. Neste caso, a Estatística inferencial é utilizada para criar uma regra de decisão com base em observações de um subconjunto de 100 peças.

Apesar do interesse prático da inferência, não abordaremos essas técnicas aqui devido ao caráter introdutório deste trabalho. Com vistas a emergir neste universo, apresentamos algumas definições básicas que deverão ser, naturalmente, associadas à teoria que vem sendo desenvolvida até aqui.

Definição 4.1. Dizemos que **população** é o conjunto de todas as unidades³ sobre as quais há o interesse de investigar uma ou mais características.

O conceito de população em Estatística é bem mais amplo que o uso comum dessa palavra. A população pode ser formada por pessoas, domicílios, peças de produção, cobaias ou qualquer outro elemento a ser investigado. A

³ Cada elemento da população é também denominado de **unidade estatística**.

caracterização da população é feita em função do problema a ser estudado. Para que haja uma clara definição das unidades que formam a população, é necessária a especificação de três elementos: uma característica em comum, localização temporal e localização geográfica.

Exemplo 4.3. *Estudo da obesidade em estudantes do Instituto Federal de Brasília (IFB) em 2010 por intermédio da medida do índice de massa corpórea.*

Na situação apresentada no Exemplo 4.3, cada unidade estatística que integra a população possui as características de possuir massa (característica comum), ser estudante em 2010 (localização temporal) e ser estudante do IFB (localização geográfica). Logo, a população é o conjunto dos estudantes do IFB em 2010.

Sendo um conjunto, uma população pode ser finita ou infinita. Mesmo quando finita, muitas vezes torna-se inviável, devido a razões de tempo ou por economia de recursos, a observação da característica desejada sobre todas as suas unidades. Neste caso, coletamos uma **amostra** dessa população.

Definição 4.2. Dizemos que uma amostra é um subconjunto não vazio da população.

Em particular, quando a amostra é a própria população, chamamos o processo de coleta de informações de **censo**. Quando o censo torna-se inviável, lançamos mão de uma amostra, que, em outras palavras, é uma **janela** para o universo populacional. Dependendo da escolha dessa janela, podemos ter um ângulo favorável ou não para a percepção do comportamento de toda a população quanto à característica desejada.

Existem técnicas que nos auxiliam no trabalho de escolher uma amostra que seja suficientemente representativa da população. Essas técnicas são chamadas de **técnicas de amostragem**.

2 Amostragem

Quando dizemos que uma pessoa é “honesta e trabalhadora”, estamos emitindo um juízo de valor que temos em razão de algumas ações praticadas por aquela pessoa, muitas vezes sem conhecer profundamente seus atos e

atividades. Quando procuramos um restaurante para jantar, norteamos a nossa escolha por algumas informações coletadas entre amigos, ou em experiências anteriores, ou mesmo nos baseamos em informações oferecidas pela internet, que são fornecidas baseadas em critérios desconhecidos.

O fato é que nem sempre tiramos conclusões corretas sobre as pessoas ou nem sempre escolhemos um bom restaurante para jantar. Ainda assim, o tempo todo escolhemos amostras para subsidiar nossas decisões e ansiamos por decisões sensatas.

Definição 4.3. A **amostragem** é o ramo da Estatística que estuda as relações entre uma amostra e a população da qual ela foi extraída.

Para que um levantamento por amostragem seja bem-sucedido é necessário que se conheça profundamente a população. Por exemplo: para se verificar a qualidade da água de um reservatório basta a coleta de uma pequena parcela de seu volume. Os exames laboratoriais, neste caso, são conclusivos devido ao fato de a população estudada ser homogênea. No caso geral, nem sempre essa homogeneidade ocorre, e esse é um detalhe importante que deve ser levado em consideração pelo pesquisador para a execução, com êxito, de um trabalho de amostragem.

Quando bem realizado, um estudo por amostragem apresenta diversas vantagens em relação ao censo tais como: custo reduzido, devido à necessidade de um menor volume de dados; maior rapidez, decorrente da velocidade de tabulação das informações; maior abrangência, uma vez que pode ser realizado em situações em que a realização do censo é impossível; maior exatidão, na medida em que possibilita trabalhar com uma equipe mais bem preparada, aumentando a qualidade das informações coletadas.

2.1 Técnicas de amostragem

O processo de amostragem pode ser aleatório⁴ ou não aleatório. Na Seção 3, introduzimos uma discussão sobre aleatoriedade. Para o momento,

⁴ Um processo aleatório de amostragem é também denominado de **probabilístico**, pois permite a aplicação de métodos de inferência e da teoria de probabilidades.

entendamos por aleatório aquilo que não é determinístico, isto é, aquilo que é impossível de se prever ou determinar.

Afirmamos que a aleatoriedade na escolha da amostra é fundamental para garantir a representatividade da população. Entretanto, em muitas situações práticas, existem restrições que impedem que a seleção da amostra seja totalmente ao acaso. Listamos abaixo alguns tipos de amostragem não aleatória, juntamente com algumas situações em que sua aplicação é possível.

- **Amostragem por acessibilidade:** a amostra atinge apenas a parte acessível da população.

Exemplo 4.4. Num vagão carregado de minério, por dificuldades de atingir todos os pontos, uma amostra é coletada apenas na camada superficial de 20 cm de minério.

- **Amostragem a esmo:** a amostra é constituída pelos elementos que se consegue capturar de uma população.

Exemplo 4.5. Num galpão de aves, uma amostra pode ser constituída por aves que foram tomadas no instante da coleta, sem, entretanto, ter havido sorteio.

- **Amostragem intencional:** o pesquisador escolhe, a seu juízo, os elementos da população que julga significativos para constituírem sua amostra.

Exemplo 4.6. Um historiador, por exemplo, pode optar por ouvir moradores de uma pequena cidade com mais de 90 anos a fim de reconstruir um aspecto histórico importante.

- **Amostragem por voluntariado:** caso em que o processo de obtenção dos dados é desagradável e a amostra é construída a partir da iniciativa das unidades estatísticas.

Exemplo 4.7. Uma pesquisa sobre alcoolismo pode ser realizada a partir de uma amostra formada por alcoólatras voluntários.

De acordo com a especificidade do que se pretende pesquisar, é possível fazer uma amostragem aleatória. Sendo assim, teremos toda a teoria probabilística

à nossa disposição e a inferência torna-se viável. Entre as possibilidades de amostragem aleatória, destacamos:

- **Amostragem aleatória simples:** consiste em selecionar aleatoriamente uma amostra de medida n em uma população de medida N . Pressupõe-se que a população é homogênea, isto é, todas as unidades estatísticas têm a mesma chance de ocorrer. Se a população for discreta, a escolha dos elementos da amostra pode se dar com ou sem reposição.

Exemplo 4.8. Numa população de cardinalidade $N = 50$, a escolha de uma amostra aleatória simples com cardinalidade $N = 10$ pode ser feita representando-se cada unidade num papel numerado de 1 a 50 e efetuando-se um sorteio, com ou sem reposição, dessas 10 unidades.

- **Amostragem sistemática:**⁵ caso exista uma lista das unidades populacionais, a amostragem sistemática é uma técnica simples que consiste na escolha de uma unidade para compor a amostra a cada bloco regular de tamanho $\frac{N}{n}$. Procede-se assim: primeiramente, considera-se $k = \left\lfloor \frac{N}{n} \right\rfloor$ ⁶ a seguir, sorteia-se um número inteiro no intervalo $[1, \frac{N}{n}]$, que serve como ponto de partida para a escolha do primeiro elemento a ser incluído na amostra; descartando os $k - 1$ próximos elementos, seleciona-se o segundo, e assim por diante. Tal como na amostragem aleatória simples, é necessária a existência de um cadastro. Uma das grandes vantagens da amostragem sistemática, em relação à amostragem aleatória simples, é a praticidade na seleção dos elementos. Problemas com a amostragem sistemática podem surgir quando a sequência dos elementos no cadastro induz a um comportamento periódico ou cíclico na característica a ser investigada.

Exemplo 4.9. Considere uma vila com 20 casas numeradas de 1 a 20. Se todas as casas cujos números são múltiplos de 4 estiverem mais perto da linha de trem e o intuito é medir poluição sonora, a amostragem sistemática não será adequada.

⁵ Como nem todas as unidades são passíveis de seleção, esse procedimento é classificado como quase aleatório.

⁶ $\lfloor x \rfloor$ é uma notação para representar o maior número inteiro menor ou igual a x . Por exemplo, se $y = 17,32$, então $\lfloor y \rfloor = 17$ e se $z = 17$, então $\lfloor z \rfloor = 17$.

- **Amostragem aleatória estratificada:** quando a população é heterogênea e for possível a formação de h estratos homogêneos em relação à característica estudada, esse procedimento consiste na aplicação de amostragem aleatória simples ou amostragem sistemática dentro de cada estrato. A amostragem estratificada é vantajosa quando há o conhecimento prévio de grupos que sejam mais homogêneos internamente e heterogêneos entre si, em relação à característica investigada. Nessas situações, há um ganho em relação à amostragem aleatória simples, pois a seleção dentro dos estratos leva à diminuição do tamanho da amostra, mantendo a precisão das estimativas. Uma etapa importante da amostragem aleatória estratificada é a alocação da amostra pelos estratos, ou seja, quantos elementos da amostra pertencerão ao estrato 1, ao estrato 2, ..., ao estrato h. Dois tipos de alocação são comumente aplicados: alocação uniforme (mesmo número de elementos nos estratos) e alocação proporcional (número de elementos proporcional ao tamanho do estrato).

Exemplo 4.10. No Instituto Federal de Brasília (IFB), campus Planaltina, há estudantes de educação técnica matriculados em duas modalidades: integrada e subsequente. Um projeto de extensão consiste em abrir turmas de Inglês para atendimento conjunto desses estudantes. Para tanto, a proficiência dos estudantes nessa língua será avaliada numa amostra de 10% da população, para definir o ponto de partida do curso. Se no IFB, atualmente, 60% das matrículas são de estudantes na modalidade subsequente, então, para um melhor diagnóstico, pode ser viável uma amostragem aleatória estratificada por modalidade de curso.

- **Amostragem por clusters:** neste método, em vez da seleção de unidades da população, são selecionados conglomerados (*clusters*) dessas unidades e, já dentro do conglomerado, procede-se à seleção de algumas unidades que comporão a amostra. Trata-se de uma alternativa interessante quando as unidades estão muito afastadas geograficamente ou quando não dispomos de uma lista das unidades. Esse tipo de amostragem induz indiretamente à aleatoriedade na seleção das unidades que formarão a amostra e tem a grande vantagem de facilitar a coleta de dados.

Exemplo 4.11. Se a unidade de interesse, por exemplo, for um estudante, pode ser que não exista um cadastro de estudantes, mas sim de escolas. Portanto, pode-se selecionar escolas e nelas investigar todos os alunos, ou alguns deles.

3 Aleatoriedade

Iniciamos esta seção com uma discussão sobre aleatoriedade. Na Seção 2, apresentamos o conceito de aleatório como aquilo que não é determinístico, isto é, aquilo que é impossível de se prever ou determinar.

No tocante à aleatoriedade na natureza, sua existência ou não, os filósofos gregos clássicos a compreendiam sob duas concepções diferentes: a aleatoriedade subjetiva de Demokritos,⁷ que a concebia apenas como outro nome para a ignorância humana acerca das causas determinantes de uma dada estrutura ou de um dado fenômeno e, consequentemente, ela seria apenas a expressão da nossa incapacidade de descrever, prever ou controlar, sendo, portanto, um determinismo disfarçado; e a aleatoriedade objetiva de Epikuros,⁸ que a concebia como a ausência de causas, ou seja, um processo aleatório objetivo teria comportamento impossível de se prever e controlar e, se o repetirmos a partir de estados iniciais e causas idênticas, ele produzirá efeitos diferentes, que são determinados totalmente ao acaso.

Entre os mais importantes seguidores de Demokritos, podemos citar o matemático Laplace e o físico Einstein. Este último é o autor da célebre frase “Deus não joga dados!”, mas também é um dos propositores de um experimento, em 1935, que teria o intuito de comprovar a não existência da aleatoriedade objetiva de Epikurus. Entretanto, nos anos 1980, a realização desse mesmo experimento veio provar exatamente o contrário do que o motivou: provou que o indeterminismo e a aleatoriedade objetiva são qualidades intrínsecas dos fenômenos naturais microscópicos, isto é, a aleatoriedade objetiva, de fato, existe na natureza.

3.1 Noção intuitiva de variável aleatória

Em um levantamento de dados, censitário ou por amostragem, investiga-se uma ou mais características de interesse que supostamente variam de uma

⁷ Filósofo grego pré-socrático que viveu entre os anos 460 e 370 a.C.. Discípulo de Leucippus, por sua teoria atômica para o Cosmos, muitos o consideram o “pai da ciência moderna”.

⁸ Filósofo grego que viveu entre 341 a.C e 270 a.C., fundador da escola filosófica chamada Epicurismo. Fazia parte dos seus ensinamentos que o universo é infinito e eterno, e que eventos no mundo são, em última análise, baseados nos movimentos e interações de átomos que se deslocam no espaço vazio.

unidade para outra. Essas características serão chamadas a partir de agora de **variáveis aleatórias**.

Exemplo 4.12. *Um pesquisa eleitoral investiga a intenção de votos da população em três candidatos, A, B ou C. Ao se entrevistar um indivíduo numa amostra aleatória da população, é imprevisível saber, de antemão, a intenção de voto desse eleitor. Por esse caráter imprevisível, denominamos a característica “intenção de voto” como variável aleatória.*

As variáveis aleatórias são classificadas de acordo com o conjunto em que assumem “valor”. Uma variável aleatória pode ser uma quantidade sobre a qual podem ser realizadas operações aritméticas, ou pode ser um atributo, como uma característica genética, uma zona de moradia ou uma classe social. No primeiro caso, a variável é classificada como **quantitativa** e na outra situação como **qualitativa**.

A classificação da variável vai ser determinante para o tipo de análise estatística a ser conduzida. Sobre uma variável qualitativa, não podemos calcular muitos dos resumos numéricos, tais como a média aritmética, a variância e o desvio-padrão. Por outro lado, o gráfico de setores é mais adequado para representar variáveis qualitativas do que quantitativas.

As variáveis aleatórias qualitativas podem ser classificadas como **ordinais** ou **nominais**. A variável aleatória do Exemplo 4.12 é qualitativa nominal, pois assume valores num conjunto que não admite ordenação. Se naquela mesma amostra do Exemplo 4.12 tivéssemos interesse em investigar “o grau de escolaridade dos eleitores”, e estes tivessem como possibilidades de resposta as opções nenhum, fundamental, médio ou superior, então a variável seria classificada como qualitativa ordinal.

Por sua vez, as variáveis quantitativas podem ser **discretas**, quando assumem valores num conjunto discreto; **contínuas**, quando assumem valores num conjunto contínuo; ou mesmo **nem discretas nem contínuas**, quando assumem valores em um conjunto que não é discreto nem contínuo.

Como exemplo de variável aleatória quantitativa discreta, temos “o número de acidentes na BR-020 numa dada semana”. Já a variável “estatura de um estudante do IFB” é uma variável aleatória quantitativa contínua, uma vez que há um intervalo, digamos de 0 a 2,5 metros, em que ela pode tomar valores.

4 Técnicas da Estatística Descritiva

A principal função da Estatística Descritiva é resumir as informações contidas em um vetor que as carregam por meio de tabelas, gráficos e medidas (resumos numéricos).

A descrição dos dados deve ser objetiva, ter precisão de significado e simplicidade no cálculo, para que as pessoas possam compreender e, eventualmente, reproduzir os resultados.

4.1 Análise da frequência

Considere $x = (x_1, x_2, \dots, x_n)$ um vetor de informações. Sem perda de generalidade, podemos supor que se trata de um rol, conforme nomenclatura utilizada no Capítulo 3, caso trate-se de uma variável aleatória quantitativa. Com base no rol, fica fácil visualizar a frequência de cada informação. Podemos, assim, construir a tabela de frequências.

Exemplo 4.13. A tabela a seguir resume as informações da variável estado civil presente em um questionário aplicado a moradores de comunidades de baixa renda em São Paulo.

TABELA 4.1 – ESTADO CIVIL EM UMA AMOSTRA DE 385 INDIVÍDUOS

<i>i</i>	Estado civil	<i>n_i</i>	<i>f_i</i>
1	solteiro	165	0,4286
2	casado	166	0,4312
3	divorciado	10	0,0260
4	viúvo	12	0,0312
5	outro	32	0,8310
Total		$\sum_i^5 n_i = 385$	$\sum_i^5 f_i = 1$

Fonte: Magalhães e Lima, 2004.

Na Tabela 4.1, n_i representa a **frequência absoluta** do estrato i , isto é, o número de vezes que o estrato i aparece no rol, e f_i representa sua **frequência relativa**, isto é, o percentual da frequência absoluta, quando comparada com o número total de observações.

As tabelas de frequência, além de possibilitarem rápida visualização dos dados, nos permitem tirar algumas conclusões simples, mas que prescindem de organização.

Exemplo 4.14. A tabela a seguir mostra a frequência de idades em uma amostra de 50 estudantes que responderam a um questionário sobre hábitos de lazer.

TABELA 4.2 – IDADE

<i>i</i>	Idade	<i>n_i</i>	<i>f_i</i>	<i>F_i</i>
1	17	9	0,18	0,18
2	18	22	0,44	0,62
3	19	7	0,14	0,76
4	20	4	0,08	0,84
5	21	3	0,06	0,9
6	22	0	0	0,9
7	23	2	0,04	0,94
8	24	1	0,02	0,96
9	25	2	0,04	1
Total		$\sum_{i=1}^9 n_i = 50$	$\sum_{i=1}^9 f_i = 1$	

Fonte: Magalhães e Lima, 2004.

Observe que a variável “idade” é quantitativa discreta e que a Tabela 4.15 é mais completa que a Tabela 4.13, uma vez que traz uma coluna a mais, correspondente à frequência relativa acumulada até a classe i , denotada por F_i .

Em geral, numa tabela organizada em k classes, a frequência relativa acumulada até a classe i é calculada assim:

$$F_i = \sum_{j=1}^i f_j, \forall i = 1, 2, \dots, k.$$

De posse da Tabela 4.15 de frequências, podemos concluir, imediatamente, que 90% dos estudantes têm idade menor ou igual a 21 anos.

Neste momento, cabe ressaltar que uma tabela de frequência é um instrumento técnico e que, portanto, segue algumas especificações. Toda tabela de frequência deve ter um título, que geralmente é a variável que estamos observando, e deve vir acompanhada da fonte de onde aquelas informações foram extraídas. Ela, em si mesma, deve comunicar todas as informações sem necessidade de complementação.

Exemplo 4.15. O grupo de 50 estudantes do Exemplo 4.15, quando questionados sobre a quantidade de horas semanais dedicadas à atividade física, apresentou a seguinte distribuição de frequências.

TABELA 4.3 – QUANTIDADE DE HORAS SEMANAIS DEDICADAS À ATIVIDADE FÍSICA

Classe <i>i</i>	Horas por semana	<i>n_i</i>	<i>f_i</i>	<i>F_i</i>
1	0 — 2	11	0,22	0,22
2	2 — 4	14	0,28	0,5
3	4 — 6	12	0,24	0,74
4	6 — 8	8	0,16	0,90
5	8 — 10	3	0,06	0,96
6	10 — 12	2	0,04	1,00
Total		$\sum_{i=1}^6 n_i = 50$	$\sum_{i=1}^6 f_i = 1$	

Fonte: Magalhães e Lima, 2004.

A variável “Quantidade de horas semanais dedicadas à atividade física” é quantitativa contínua. Devido à variedade de possibilidades de informação, os dados foram agrupados em classes que são intervalos de tempo. A notação $a |— b$ é equivalente à notação de intervalo $[a, b)$ apresentada no Capítulo 1.

Note que a apresentação das informações em classes intervalares torna o trabalho mais organizado e informativo. Isso sempre é possível quando

trabalhamos com variáveis quantitativas contínuas, ou ainda com variáveis discretas que tomam muitos valores distintos.

Vamos apresentar agora um critério para definir quantas classes distintas utilizar numa tabela de distribuição de frequências. É importante ressaltar que não há consenso neste assunto, daí a importância do estilo e do bom senso do pesquisador.

Os pacotes computacionais, entretanto, valem-se frequentemente dos dois critérios apresentados a seguir. Para compreendê-los, considere $x = (x_1, x_2, \dots, x_n)$ um rol contendo as informações observadas. Considere $\Delta(x)$ a amplitude de x , conforme a Definição 3.8, em que k é o número de classes que existirão na tabela de frequência e h o comprimento de cada classe. Teoricamente, esses elementos da estrutura de uma tabela de frequência se relacionam assim:

$$k = \left\lceil \frac{\Delta(x)}{h} \right\rceil + 1.$$

1. Critério de Sturges.

O critério de Sturges⁹ diz que

$$k \approx 1 + 3,3 \log n.$$

Sendo k um número inteiro, deveremos fazer um arredondamento ou truncamento, conforme a conveniência.

2. Critério da raiz quadrada.

O critério da raiz quadrada diz que

$$k \approx \sqrt{n}.$$

Sendo k um número inteiro, deveremos fazer um arredondamento ou truncamento, conforme a conveniência.

⁹ Quando lidamos com números da forma $y = 10^x$, em que x é um número real qualquer, dizemos que o expoente x é o logaritmo de y na base 10 e denotamos essa situação assim: $x = \log y$.

De uma forma ou de outra, é possível encontrar uma aproximação razoável para k e h , que se ajustam bem ao rol de informações. Ambos os critérios determinam o número de classes, em função¹⁰ do número de informações disponíveis. Todavia, o critério de Sturges é mais indicado quando a dimensão do vetor x for muito grande, isto é, quando no rol constar um número muito grande de informações.

4.2 Gráficos

Na subseção anterior, vimos um importante tipo de tabela que aparece frequentemente em trabalhos técnicos, as chamadas tabelas de frequência. Ao lidarmos com tabelas, é importante conhecer bem sua estrutura, para poder interpretá-la aproveitando todo seu potencial informativo.

Uma tabela simples de frequência de valores assumidos por uma variável aleatória qualitativa pode gerar um gráfico de setores, como o mostrado na Figura 4.2.

TABELA 4.4 – CEREAIS, OLEAGINOSAS E LEGUMINOSAS. PARTICIPAÇÃO DE PRODUÇÃO SEGUNDO AS GRANDES REGIÕES.

região	percentual
Centro-Oeste	34,7%
Nordeste	9,7%
Norte	2,7%
Sudeste	11,5%
Sul	41,4%
total	100%

Fonte: IBGE. *Primeiras estimativas da safra 2010*.

¹⁰ O critério de Sturges é mais indicado quando a dimensão de x for muito grande, devido ao crescimento da função logarítmica ser mais lento que a função raiz quadrada. Isso se reflete num menor número de classes para a tabela de frequência.

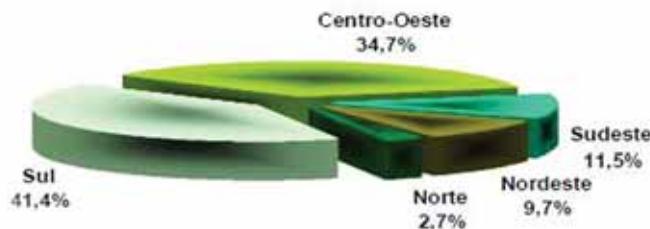


FIGURA 4.2 – CEREAIS, OLEAGINOSAS E LEGUMINOSAS. PARTICIPAÇÃO DE PRODUÇÃO SEGUNDO AS GRANDES REGIÕES.

Fonte: Fonte: IBGE. *Primeiras estimativas da safra 2010.*

Outros tipos de tabela aparecem em trabalhos técnicos, mas a interpretação de suas informações se processa de forma similar. A partir de uma tabela, podemos produzir diversos tipos de gráficos. Por exemplo, considere a tabela da Figura 4.3 a seguir, que apresenta frequências relativas de características conjugadas.

Grande regiões	Percentual das pessoas de 15 anos ou mais de idade usuárias de tabaco fumado (%)		
	Total	Sexo	
		Homem	Mulher
Brasil	17,2	21,6	13,1
Norte	16,8	22,0	11,7
Nordeste	17,2	22,9	11,9
Sudeste	16,7	20,4	13,3
Sul	19,0	22,5	15,9
Centro-Oeste	16,6	21,4	12,1

FIGURA 4.3 – PERCENTUAL DE PESSOAS DE 15 ANOS OU MAIS DE IDADE, USUÁRIAS DE TABACO, FUMADO OU NÃO FUMADO, POR SEXO, SEGUNDO AS GRANDES REGIÕES - 2008

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento. *Pesquisa Nacional por Amostra de Domicílios - 2008.*

Por meio dela, podemos construir um gráfico de barras, conforme mostra a Figura 4.4.

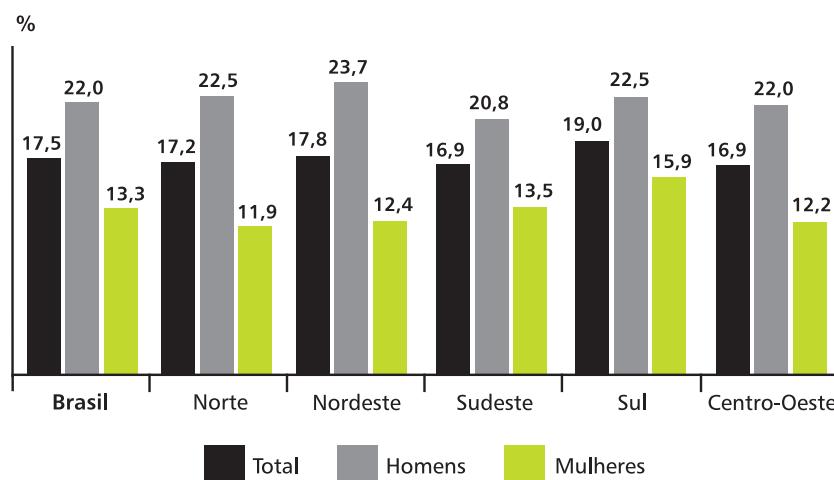


FIGURA 4.4 – PERCENTUAL DE PESSOAS DE 15 ANOS OU MAIS DE IDADE, USUÁRIAS DE TABACO, FUMADO OU NÃO FUMADO, POR SEXO, SEGUNDO AS GRANDES REGIÕES – 2008
Fonte: IBGE. Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento. *Pesquisa Nacional por Amostra de Domicílios - 2008*.

Tabelas e gráficos com características conjugadas são importantes pois ampliam nosso olhar sobre o comportamento conjunto daquelas informações, possibilitando, inclusive, a observação de diferenças de comportamento por diferentes estratos de população ou amostra heterogênea.

4.3 Histogramas e distribuição da frequência

O histograma é um gráfico composto por retângulos justapostos que representam a frequência (ou a frequência relativa) dos valores assumidos, geralmente,¹¹ por uma variável quantitativa contínua. Para uma variável com tais especificações, sabemos que a informação apresentada na tabela de frequência se encontra aglomerada em classes intervalares. A base de cada um dos retângulos que compõem o histograma de uma variável corresponde ao intervalo de classe e a sua altura à respectiva frequência (ou frequência relativa).

¹¹ Muitas vezes, construimos histogramas de variáveis quantitativas discretas, sobretudo quando é muito grande a quantidade de dados.

Quando organizamos as informações provenientes de uma variável quantitativa contínua em uma tabela de frequência, e fazemos uma distribuição intervalar dessa informação, estamos, na verdade, assumindo a hipótese de que ela se distribui uniformemente pelo intervalo.

Exemplo 4.16. 150 peixes mortos foram vítimas de contaminação de um rio e seu comprimento foi medido em milímetros. As medidas foram expressas na forma de tabela de frequência.

TABELA 4.5 – COMPRIMENTO (MM) DE PEIXES ENCONTRADOS MORTOS NO RIO

Classe i	Comprimento do peixe (mm)	n_i
1	100 — 110	7
2	110 — 120	16
3	120 — 130	19
4	130 — 140	31
5	140 — 150	41
6	150 — 160	23
7	160 — 170	10
8	170 — 180	3
Total		$\sum_{i=1}^8 n_i = 100$

Fonte: <http://leg.ufpr.br/~paulojus>

Essas informações geram o seguinte histograma:

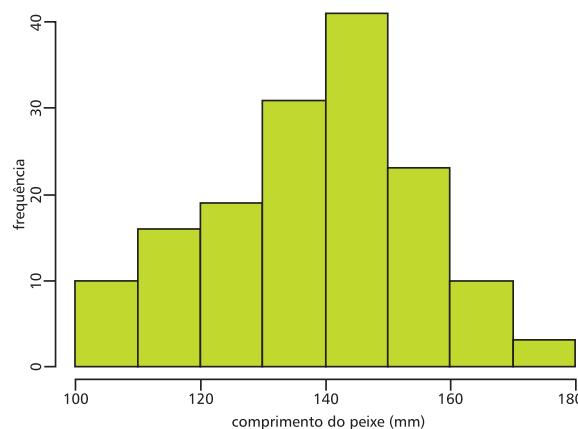


FIGURA 4.5 – COMPRIMENTO (MM) DE PEIXES ENCONTRADOS MORTOS NO RIO

Fonte: <http://leg.ufpr.br/~paulojus>

Considere agora um histograma qualquer, construído a partir de um vetor $x = (x_1, x_2, \dots, x_n)$ de informações de uma variável quantitativa contínua. Sejam

$$a_1 \leftarrow a_2, a_2 \leftarrow a_3, \dots, a_k \leftarrow a_{k+1}$$

os k intervalos de classe em que a informação contida em x foi distribuída no momento da construção da tabela de frequência. Se considerarmos m_i o valor médio do intervalo de classe i e marcarmos no histograma a frequência teórica n_i relacionada a esse número, geraremos, ao unir todos os pontos

$$(a_1, 0), (m_1, n_1), \dots, (m_k, n_k), (a_{k+1}, 0),$$

uma figura denominada **polígono de frequência**, cuja área que ela determina com o eixo horizontal é, aproximadamente, igual ao somatório das áreas dos retângulos que formam o histograma. Quando o número de dados da amostra aumenta indefinidamente e o intervalo de classe tende a zero, a frequência tende a se concentrar nos pontos médios das classes e a distribuição da frequência passa a ser uma função de **densidade de probabilidades**.¹²

A construção de histogramas tem caráter preliminar em qualquer estudo, e é um importante indicador da distribuição de dados. Um outro gráfico muito utilizado para capturar o comportamento conjunto dos dados é o gráfico de **distribuição de frequências** ou **ogiva**. A técnica utilizada para esboçá-lo é a mesma utilizada para construir o histograma. A única diferença é que, em vez de plotarmos as informações de frequência absoluta ou de frequência relativa de cada classe, informamos a frequência absoluta acumulada ou a frequência relativa acumulada em cada classe. No segundo caso, denominamos a ogiva de **função de distribuição**.

Mediante o estudo do histograma ou da função de distribuição de um vetor, podemos reforçar ou reduzir a hipótese sobre a **normalidade** do comportamento desses dados, ou mesmo se esse comportamento é o resultado da **mistura** de informações que possuem comportamento normal.

¹² Não faremos essa discussão aqui.

4.4 Outros tipos de gráficos

Nesta seção, apresentaremos dois tipos de gráficos bastante técnicos: o **gráfico de ramos e folhas** e o **box plot**. Trata-se de gráficos de construção simples, mas que possuem grande capacidade de revelar aspectos importantes de um conjunto de dados.

Apresentaremos como se dá a construção de um gráfico de ramos e folhas através de um exemplo.

Exemplo 4.17. Suponha que as informações a seguir sejam o resultado de alguma característica numérica observada em uma amostra de uma população qualquer.

14	37	40	40	42	30	38	40
40	42	30	38	40	41	43	33
45	35	39	40	41	46	36	39
40	41	47	36	39	40	41	53

Essas informações podem ser representadas assim:

Profundidade			$n = 40$
			Unidade: 1
1	1	4	
	2		
16	3	003556678889999	
24	4	00000000001111112234567	
1	5	3	

No gráfico, os números que estão imediatamente à esquerda da linha vertical são denominados **ramos** e os imediatamente à direita, **folhas**. Cada ramo tem uma **profundidade**, indicada na extremidade esquerda do gráfico, que corresponde ao número de folhas que ele contém.

São os valores numéricos de cada informação que compõem o gráfico de ramos e folhas. A unidade, localizada no canto superior direito do gráfico,

determina como serão os ramos, uma vez que as folhas só podem ser representadas por um número natural. Em nosso exemplo, como todos os dados são números na casa das dezenas, a dezena indicará o ramo e a unidade indicará a folha onde será lançada essa informação. Assim, a unidade do ramo é a dezena. Por exemplo, o dado 14 está representado pela folha 4 no ramo 1.

Dizemos que a **profundidade do dado** é a posição em que a folha que o representa se encontra em relação à extremidade mais próxima do gráfico. Por exemplo, a informação 14 tem profundidade 1.

O gráfico de ramos e folhas permite observar características como a presença de dados estranhos, a existência de simetria no conjunto de dados, a dispersão dos dados e a existência de concentrações de informações ou de lacunas entre eles.

Já o *box plot* é uma ferramenta gráfica que revela aspectos importantes como tendência central, variabilidade, simetria e existência de valores atípicos (*outliers*).

Dado o rol $x = (x_1, x_2, \dots, x_n)$, a construção do *box plot* para esse vetor de informações baseia-se no registro, em escala, das seguintes medidas: x_{\min} , Q_1 , x_{med} , Q_3 e x_{\max} , em que Q_1 , x_{med} e Q_3 são os quartis discutidos no Capítulo 3.

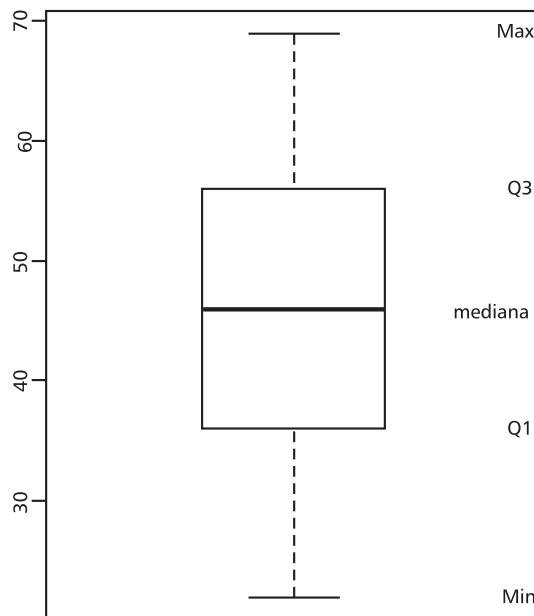
A Figura 4.6 mostra o desenho esquemático de um *box plot*. A parte central do gráfico é composta de uma caixa com o nível superior dado por Q_3 e o nível inferior por Q_1 . O tamanho da caixa é a medida de dispersão

$$A_{IQ} = |Q_3 - Q_1|,$$

denominada **amplitude interquartílica**. A mediana é representada por um traço no interior da caixa e segmentos de reta são colocados da caixa até os valores máximo e mínimo observados.

A medida da amplitude interquartílica pode ser usada para a detecção de valores atípicos. Um critério para a determinação se uma informação é ou não atípica é o seguinte:

$$x_i \text{ é outlier se } x_i > Q_3 + 1,5 \cdot A_{IQ} \text{ ou se } x_j < Q_1 - 1,5 \cdot A_{IQ}.$$

**FIGURA 4.6 – BOX PLOT COM BASE EM CINCO NÚMEROS**

Fonte: fictícia.

Costuma-se indicar a presença de *outliers* no *box plot* por meio de pontos acima das informações extremas. Assim, *outliers* são informações que se diferenciam tanto que são desprezadas para efeito de análise.

Exemplo 4.18. A tabela a seguir resume informações a respeito das alturas h de um grupo de 200 crianças de uma determinada escola, sendo 80 do sexo feminino e 120 do sexo masculino.

TABELA 4.6 – MEDIDAS RESUMO

Medida	Feminino	Masculino
h_{min}	78	78
h_{max}	136	148
Q_3	114	119
h_{med}	106	108
Q_1	94	96

Fonte: fictícia.

A comparação do comportamento conjunto das alturas desses dois grupos torna-se mais fácil através do *box plot* a seguir.

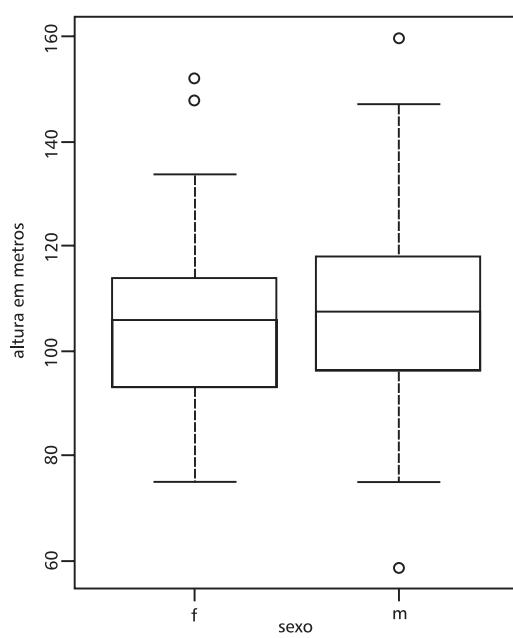


FIGURA 4.7 – ALTURA DE CRIANÇAS DO SEXO FEMININO (F) E MASCULINO (M)
Fonte: Tabela 4.18.

O *Exemplo 4.18* nos mostra como identificar num *box plot* a existência de *outliers*. Tanto no grupo do sexo masculino como no do sexo feminino, há a presença de dois valores atípicos. No caso da meninas, duas alturas divergiram muito para maior do que era considerado normal. No caso dos meninos, uma altura apresentou comportamento anormal para maior e outra para menor.

Esses pontos discrepantes dificilmente seriam detectados simplesmente observando-se os dados agregados, pois eles são identificados de acordo com a tendência central e a variabilidade do grupo ao qual pertencem. É bem provável que uma menina cuja altura é discrepante em relação às outras meninas não se destaque em relação à altura de todas as crianças, pois, ao incluir meninos, a medida de tendência central é aumentada.

Esse exemplo vem mostrar também a utilidade do *box plot* como ferramenta de análise multivariada, permitindo comparar populações de tamanhos distintos.

5 Resumo das etapas do trabalho estatístico

Diante da necessidade de investigação de alguma característica ou variável definida em uma população qualquer, devemos dar início a um rigoroso trabalho que envolve rotinas preestabelecidas: o trabalho estatístico.

Nesta seção, resumiremos as fases do trabalho estatístico em:

1. Planejamento
2. Coleta de dados
3. Tratamento dos dados
4. Análise dos resultados

Na fase do planejamento, precisamos:

- definir o problema, ou seja, saber exatamente o que pretendemos pesquisar;
- definir como levantar as informações e que dados esperamos encontrar;
- definir que tipo de levantamento faremos, se um censo ou um levantamento por amostragem;
- definir, no caso de levantamento por amostragem, qual o tipo de amostragem é mais apropriado para o que pretendemos pesquisar;
- definir um cronograma de atividades;
- definir os custos da operação.

Na segunda fase, precisamos ir a campo e coletar as informações. Essa é uma etapa operacional e deve ser feita de forma muito sistemática e padronizada, segundo critérios de abordagem das unidades previamente definidos na etapa do planejamento.

Os dados coletados são classificados de acordo com a distância em que se encontram da fonte geradora. São chamados de **primários**, quando obtidos diretamente da própria empresa ou organização que os produziu, e **secundários**, quando obtidos de empresa ou organização que não aquela que os produziu. Para ficar mais claro, quando consultamos o *site* oficial do IBGE e acessamos informações resultantes de uma pesquisa realizada e divulgada por essa instituição, estamos, na verdade, utilizando dados primários. Quando, em vez disso, acessamos uma informação dessa mesma pesquisa realizada pelo IBGE, em uma revista x, estamos utilizando dados secundários.

A utilização de dados primários é mais segura, pois diminui os riscos de erros devidos à transcrição das informações.

O tipo mais comum de dado é aquele obtido por **coleta direta**. Os dados fornecidos pelo IBGE são, em geral, obtidos assim. São comuns, no meio empresarial, pesquisas envolvendo a coleta direta de informações junto aos consumidores, sobre suas preferências de consumo. Ou em pesquisas eleitorais, em que a opinião dos eleitores é coletada para fins de previsão dos rumos da eleição.

Uma vez realizada a coleta de dados, é preciso tratar as informações, organizando-as mediante contagem e agrupamento. Nessa fase, é necessário que se faça uma crítica aos dados, com a finalidade de evitar o que se chama de **contaminação**. Essa crítica é feita descartando-se as informações obscuras, sejam por falhas no processo de coleta, de registro ou de equívocos gerados por má interpretação.

Normalmente, é feita uma condensação dessas informações em planilhas ou tabelas. Estatísticas (medidas) são calculadas, com o objetivo de caracterizar aquele conjunto de informações, e gráficos são produzidos.

Essas três primeiras fases são o que denominamos de Estatística Descritiva. A quarta fase, que envolve técnicas mais elaboradas, é a fase mais refinada, delicada e importante do trabalho estatístico. Trata-se da análise e interpretação dos dados, que é o objetivo da Estatística Inferencial. Através dos seus métodos, a Estatística Inferencial permite induzir conclusões ou fazer previsões, com erro controlado.

Por si só, a análise gráfica é um instrumento que possibilita tirar algumas conclusões sobre o fenômeno ou a variável estudados. Mas existem muitas outras técnicas elaboradas pela Estatística Inferencial que, conforme dito anteriormente, não são objetos deste curso.

Exercícios

1. (IFG) Ao nascerem, os bebês são pesados e medidos para se saber se estão dentro dos padrões de peso e altura esperados. Essas duas variáveis
 - a) são, ambas, qualitativas.
 - b) são contínua e discreta, respectivamente.
 - c) são discreta e contínua, respectivamente.
 - d) são ambas, discretas.
 - e) são ambas, contínuas.
2. (IFG) Numa determinada empresa, foi realizada uma pesquisa a respeito dos salários dos funcionários, em comparação com o salário mínimo atual, e foram obtidos os dados seguintes:

Classes de salário	Frequência
4 — 8	10
8 — 12	12
12 — 16	8
16 — 20	5
20 — 24	1

Com base nesses dados, pode-se afirmar que:

- a) uma mediana é obtida através da média aritmética entre os elementos 18° e 19°.
- b) 50% dos funcionários ganham, em média, 16 salários mínimos.
- c) a média salarial é de 9 salários mínimos.
- d) 22 é o primeiro quartil.
- e) o número de funcionários com salário inferior a 10 salários mínimos é de 22%.

3. Suponhamos que uma empresa deseja avaliar a distribuição dos salários pagos por hora a seus 30 funcionários.

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

- a) Construa uma tabela de frequências completa, contendo colunas para a frequência absoluta, a frequência absoluta acumulada, a frequência relativa e a frequência relativa acumulada.
- b) Construa um histograma a partir da frequência absoluta e outro histograma a partir da frequência relativa. Qual a relação existente entre esses dois gráficos?
- c) Usando a mesma técnica de construção do histograma, construa um gráfico para a frequência absoluta acumulada e outro para a frequência relativa acumulada. Qual a relação existente entre esses dois gráficos?
4. O barulho é medido em decibéis, representados por dB. Um decibel corresponde ao nível de som mais fraco que pode ser ouvido por uma pessoa de boa audição num ambiente silencioso. Um susurro corresponde a 30 dB; a voz humana em conversação normal produz cerca de 70 dB; um rádio com volume alto produz cerca de 100 dB. O desconforto para os ouvidos ocorre geralmente a partir de 120 dB. Os dados a seguir correspondem a 36 medições do barulho, em diferentes horários, num determinado local.

82	90	97	115	120	77
100	65	67	124	114	102
125	94	96	97	100	99
104	121	103	72	85	97
88	103	88	83	73	102
85	98	102	123	122	110

- a) Construa o diagrama de ramos e folhas para esses dados.
- b) Construa um *box plot* para esse conjunto de dados.

5. Suponha que um produtor de laranjas, que costuma guardar as frutas em caixas, está interessado em estudar o número de laranjas por caixa. Depois de um dia de colheita, 20 caixas foram contadas. Os resultados foram armazenados no vetor

$$y = (22, 29, 33, 35, 35, 37, 38, 43, 43, 44, 48, 48, 52, 53, 55, 57, 61, 62, 67, 69).$$

Esboce um *box plot* para o vetor y .

6. (FGV) As informações na tabela a seguir referem-se ao número de reclamações na Anatel por 1000 acessos em serviço, referentes ao período de 1º a 31 de agosto de 2008.

Holding	Reclamações por 100 acessos em serviço
Telefônica	1,163
Embratel	1,136
Brasil Telecom	1,053
GVT	0,848
Telemar	0,600
CTBC Telecom	0,203
Sercomtel	0,120
Intelig	0,030

Fonte: Anatel.

O diagrama de caixa (box plot) que melhor representa esses dados é:





CAPÍTULO V

NOÇÕES DE PROBABILIDADE



Noções de probabilidade

No capítulo anterior, foram mostrados alguns conceitos relacionados à Estatística Descritiva. Neste capítulo, apresentamos a base teórica para o desenvolvimento de técnicas da Estatística Inferencial, a serem desenvolvidas em estudos posteriores.

Vamos considerar as seguintes questões: Como saber se um determinado produto está sendo produzido dentro dos padrões de qualidade? Como avaliar a capacidade de um determinado exame acertar o verdadeiro diagnóstico? Questões como essas envolvem algum tipo de variabilidade ou incerteza, e as decisões podem ser tomadas por meio da teoria de probabilidades, que permite a quantificação da incerteza.

A seguir, veremos alguns conceitos básicos de probabilidade.

1 Alguns conceitos básicos

Estamos interessados em estudar **fenômenos aleatórios**, isto é, aqueles que, mesmo realizados sob as mesmas condições e circunstâncias, podem resultar em observações distintas.

Exemplo 5.1. *Um dado será lançado sucessivas vezes e a face de cima será observada.*

É claro que os possíveis resultados de cada lançamento no Exemplo 5.1 são 1, 2, ..., 6, mas não podemos afirmar qual será o resultado de um lançamento qualquer. Assim, o fenômeno “lançamento de um dado sucessivas vezes” é um fenômeno aleatório.

O conjunto de todos os resultados possíveis de um fenômeno aleatório é chamado de **espaço amostral**. No caso particular do Exemplo 5.1, o espaço amostral Ω daquele fenômeno aleatório é

$$\Omega = \{(w_1, w_2, \dots), w_i \in \{1, 2, \dots, 6\}, \text{ para todo } i \in \mathbb{N}\}$$

Cada vetor $w = (w_1, w_2, \dots) \in \Omega$ corresponde a uma possível **realização** do fenômeno ou **experimento**.

Exemplo 5.2. *O tempo de reação de um antibiótico aplicado em um animal doente é um fenômeno aleatório cujo espaço amostral $\Omega = \{t \in \mathbb{R}, t \geq 0\}$.*

Um subconjunto qualquer do espaço amostral Ω é dito ser um **evento**¹ desse espaço. Geralmente, representamos eventos por letras maiúsculas. Em particular, se A e B são eventos de Ω , então $A \cup B$, $A \cap B$, $A - B$ e A^c também o são. Dois eventos de Ω são **mutuamente exclusivos** quando a ocorrência de um impede a ocorrência do outro. Neste caso, $A \cap B = \emptyset$.

¹ Se A é um subconjunto de Ω , então representamos a diferença $\Omega - A$ por A^c , e dizemos que A^c é o complemento de A relativo à Ω .

Exemplo 5.3 Uma moeda será lançada duas vezes e a face de cima observada. O espaço amostral referente a esse fenômeno aleatório é

$$\Omega_0 = \{(k, k), (k, c), (c, k), (c, c)\},$$

onde k representa cara e c , coroa. Os conjuntos

$$A = \{\text{cara na primeira entrada}\} = \{(k, k), (k, c)\}$$

e

$$B = \{\text{cara na segunda entrada}\} = \{(k, k), (c, k)\}$$

são eventos de Ω_0 .

Um questionamento natural a essa altura seria: diante de um experimento aleatório, qual é a chance de ocorrer determinado evento? Na próxima seção, nos ocuparemos em encontrar uma maneira de medir essa chance.

2 Medidas de probabilidade

Introduzimos esta seção com uma rápida discussão sobre as concepções filosóficas em torno do conceito de **probabilidade**. Intuitivamente, entenderemos por probabilidade uma maneira de se medir a chance de um evento acontecer.

Historicamente, há três maneiras distintas de se conceber probabilidade: através dos métodos **clássico**, **frequentista** e **bayesiano**. Os dois primeiros são também chamados de métodos objetivos, por se basearem, respectivamente, em resultados igualmente prováveis e dados históricos; o último é também chamado de método subjetivo por se valer de crenças ou da habilidade do pesquisador para ser estabelecido.

O enfoque do método a ser utilizado varia de acordo com a situação. Apresentaremos, a seguir, a definição de probabilidade segundo cada um desses

enfoques, e assumiremos que as noções convergem em algum sentido para um objeto matemático bem definido, apresentado mais adiante.

Em cada uma das subseções seguintes, considere Ω um espaço amostral.

2.1 O método clássico

O método clássico baseia-se num fato controverso, de que cada elemento do espaço amostral Ω é **equiprovável**, isto é, tem a mesma chance de ocorrência.

Exemplo 5.4. *Um dado é não viciado quando, ao ser lançado, as chances de sair qualquer uma das faces são iguais. Portanto, ao considerar um dado não viciado, estamos, na verdade, adotando um método clássico para definir as chances de os eventos relacionados ao experimento “lançamento de um dado e observação da face de cima” ocorrerem.*

De acordo com o método clássico, a probabilidade de ocorrência de um evento $A \subset \Omega$, que denotaremos por $P(A)$, é

$$P(A) = \frac{\# A}{\# \Omega}.$$

Essa concepção de probabilidade é controversa, na medida em que ela utiliza como pressuposto exatamente o que ela precisa definir, mas funciona bem quando lidamos com conjuntos finitos.

De qualquer maneira, é fácil ver que o conceito de probabilidade clássico exprime uma comparação entre o evento e o espaço amostral, isto é, dá uma idéia de peso (porcentagem) do evento no espaço amostral.

Exemplo 5.5. *Considere os eventos A e B do espaço amostral Ω_0 do Exemplo 5.3. Na hipótese de a moeda ser equilibrada, isto é, não favorecer nenhum dos resultados, teremos que*

$$P(A) = P(B) = \frac{2}{4} = \frac{1}{2}.$$

2.2 O método frequentista

A situação apresentada no Exemplo 5.5, na prática, é rara. Isso porque a moeda pode ter sido intencionalmente desgastada, favorecendo um resultado. Ou mesmo o equipamento que a produziu pode estar descalibrado. É nesse aspecto que se insere o método frequentista, que se baseia no histórico de resultados do experimento para determinar a probabilidade $P(A)$ do evento A :

$$P(A) \approx \frac{\text{número de ocorrência de } A}{\text{número total de provas ou observações}}.$$

Exemplo 5.6. Um jogo praticado numa banca é tal que, nos 10 lançamentos da moeda, saíram 8 caras. Então, pela abordagem frequentista, a probabilidade de sair cara no próximo lançamento é de aproximadamente 4/5.

Para sermos precisos, segundo o conceito frequentista,

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

em que n_A representa o número de ensaios (repetições do experimento) nos quais o evento A foi observado em n realizações do experimento e a expressão $\lim_{k \rightarrow \infty} x_k$ representa, intuitivamente, um comportamento **limite**² para a expressão x_k , quando o k cresce indefinidamente.

Assim, o cálculo de $P(A)$ é baseado na repetição do experimento ou histórico de repetições. A credibilidade desse resultado dependerá do quanto o passado do processo aleatório é representativo do futuro, e essa dependência nem sempre é alta ou sequer pode existir.

Limitações dessa abordagem ocorrem também em situações nas quais não é possível realizar infinitos ensaios do processo. Na prática, isso nunca é possível!

² A definição formal de limite pode ser encontrada em [3].

2.3 O método bayesiano

O método bayesiano está assentado num enfoque subjetivo da probabilidade. Nele há a formulação de modelos cuja verificação é baseada na combinação de evidências experimentais com a intuição do pesquisador. Por esse método, define-se a probabilidade subjetiva como sendo uma medida do grau de confiança de uma pessoa em relação a uma proposição [8]. Ela é função da quantidade de informação de que a pessoa dispõe e possui a restrição de que deve obdecer a critérios de consistência e aos axiomas da probabilidade.

Como dissemos, apesar das diversas concepções filosóficas, em algum sentido todas elas convergem para um objeto matemático bem definido e que apresentaremos na próxima seção.

2.4 Definição de probabilidade

Formalmente, uma **medida de probabilidade**³ é uma função $P(\cdot)$ que associa números reais a eventos do espaço amostral e satisfaz as seguintes condições, denominadas axiomas da probabilidade:

1. $0 \leq P(A) \leq 1$, para qualquer evento A ;⁴
2. $P(\Omega) = 1$ e $P(\emptyset) = 0$;⁵
3. $P(\cup_{j=1}^n A_j) = \sum_{j=1}^n P(A_j)$, para quaisquer A_1, \dots, A_n disjuntos.

³ Os axiomas apresentados nessa definição referem-se ao que é chamado de probabilidade finitamente aditiva. Para um estudo matematicamente mais aprofundado, o Axioma 3, que diz respeito à **aditividade finita**, deve ser substituído por um similar, qual seja:

$P(\cup_{j=1}^n A_j) = \sum_{j=1}^{\infty} P(A_j)$
para quaisquer A_1, A_2, \dots disjuntos.

⁴ Esse axioma poderia ser reduzido a apenas $P(A) \geq 0$, que o fato de $P(A) \leq 1$ surgiria como consequência.

⁵ Na verdade, $P(\emptyset) = 0$ decorre do fato de $P(\Omega) = 1$ e dos outros dois axiomas.

Exemplo 5.7. Considere Ω um espaço amostral de cardinalidade finita e defina a função

$$\begin{aligned} P_0 &: \wp(\Omega) \rightarrow [0,1] \\ P_0(A) &= \frac{\#A}{\#\Omega}. \end{aligned}$$

Observe que para a função P_0 são satisfeitos os três axiomas da probabilidade. De fato, note primeiramente que, pela definição de $P_0(\cdot)$, temos que

$$P_0(\Omega) = \frac{\#\Omega}{\#\Omega} = 1$$

e

$$P_0(\emptyset) = \frac{\#\emptyset}{\#\Omega} = 0,$$

o que verifica o Axioma 2 na definição de probabilidade; se A é um evento qualquer de Ω , então $\emptyset \subset A \subset \Omega$. Logo, $\#\emptyset \leq \#A \leq \#\Omega$. Dividindo todos os membros dessa desigualdade por $\#\Omega$, teremos $0 \leq P_0(A) \leq 1$, qualquer que seja A . Isso verifica o Axioma 1. Por fim, se considerarmos A_1, A_2, \dots, A_n eventos de Ω , tais que $A_1 \cap A_2 \cap \dots \cap A_n = \emptyset$, então

$$\#(A_1 \cup A_2 \cup \dots \cup A_n) = \#(A_1) + \#(A_2) + \dots + \#(A_n).$$

Assim,

$$\begin{aligned} P_0(\bigcup_{j=1}^n A_j) &= \frac{\#(\bigcup_{j=1}^n A_j)}{\#\Omega} \\ &= \frac{\sum_{j=1}^n \#(A_j)}{\#\Omega} \\ &= \sum_{j=1}^n \frac{\#(A_j)}{\#\Omega} \\ &= \sum_{j=1}^n P_0(A_j), \end{aligned}$$

o que mostra que $P_0(\cdot)$ satisfaz o Axioma 3. Logo, $P_0(\cdot)$ é uma medida de probabilidade.

Exemplo 5.8. Considere Ω um espaço amostral que possa ser identificado como uma região de área finita e defina a função

$$P_1 : \mathcal{P}(\Omega) \rightarrow [0,1]$$

$$P_1(A) = \frac{\text{área de } A}{\text{área de } \Omega}.$$

Observe que a função P_1 também satisfaz os quatro axiomas da probabilidade. De fato, pela definição de $P_1(\cdot)$,

$$P_1(\Omega) = \frac{\text{área de } \Omega}{\text{área de } \Omega} = 1$$

e

$$P_1(\emptyset) = \frac{\text{área de } \emptyset}{\text{área de } \Omega} = 0,$$

já que um conjunto vazio não tem área,⁶ o que verifica o Axioma 2 na definição de probabilidade; analogamente ao Exemplo 5.7, é fácil ver que $0 \leq P_1(A) \leq 1$, qualquer que seja A . Isso verifica o Axioma 1. Por fim, se considerarmos A_1, A_2, \dots, A_n eventos de Ω , tais que $A_1 \cap A_2 \cap \dots \cap A_n = \emptyset$, então

$$\text{área de } (A_1 \cup A_2 \cup \dots \cup A_n) = \text{área de } A_1 + \text{área de } A_2 + \dots + \text{área de } A_n.$$

Assim,

$$\begin{aligned} P_0(\bigcup_{j=1}^n A_j) &= \frac{\text{área de } (\bigcup_{j=1}^n A_j)}{\text{área de } \Omega} \\ &= \frac{\sum_{j=1}^n \text{área de } A_j}{\text{área de } \Omega} \\ &= \sum_{j=1}^n \frac{\text{área de } A_j}{\text{área de } \Omega} \\ &= \sum_{j=1}^n P_0(A_j), \end{aligned}$$

o que mostra que $P_1(\cdot)$ satisfaz o Axioma 3. Logo, $P_1(\cdot)$ é também uma medida de probabilidade.

⁶ É claro que a área de um conjunto vazio é zero. No entanto, diferentemente da medida $\#(\cdot)$, na qual o único conjunto com medida vazia era o \emptyset , para a medida área de (\cdot) há infinitos outros subconjuntos não vazios cuja medida é zero. Por exemplo, qualquer subconjunto discreto de Ω ou mesmo segmentos de reta, linhas, curvas, isto é, subconjuntos que não delimitam área alguma.

Assim, fica provado que medidas de probabilidade existem, ou seja, trata-se de um objeto matemático bem definido.

A seguir, discutiremos algumas propriedades importantes das medidas de probabilidade que decorrem da definição apresentada nesta seção. Assumiremos que o espaço amostral Ω se encontra munido de uma medida de probabilidade $P(\cdot)$, definida sobre o conjunto $\wp(\Omega)$.⁷

3 Propriedades operatórias

3.1 Monotonicidade da probabilidade

A probabilidade é uma função monótona não decrescente, no sentido que se $A_1 \subset \Omega$, então $P(A_1) \leq P(A_2)$. Isso pode ser facilmente visto a partir dos axiomas 1 e 3, que definem a probabilidade.

De fato, note que podemos reescrever o conjunto A_2 como a união disjunta

$$A_1 = (A_2 - A_1) \cup A_1$$

dai,

$$P(A_1) = P((A_2 - A_1) \cup A_1) = P(A_2 - A_1) + P(A_1),$$

em que a última igualdade é garantida pelo Axioma 3, por termos no argumento da função probabilidade uma união disjunta de eventos. Pelo Axioma 1, $P(A_2 - A_1) \geq 0$, o que faz com que, de fato, $P(A_1) \leq P(A_2)$.

3.2 Probabilidade de não ocorrência de um evento

Seja Ω um espaço amostral e considere A um evento qualquer. Como $A^c \subset \Omega$, temos que A^c é também evento de Ω , dito ser o **complementar** do evento A .

⁷ Nessas condições, dizemos que o espaço $(\Omega, \wp(\Omega), P(\cdot))$ é um **espaço de probabilidade**. Por simplicidade, estamos considerando $\Omega := (\Omega, \wp(\Omega), P(\cdot))$.

É comum estarmos interessados em analisar a não ocorrência do evento A , isto é, a $P(A^c)$. Como

$$\Omega = A \cup A^c$$

e A e A^c eventos disjuntos, temos, pelos axiomas 2 e 4 da definição de probabilidade, que

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c).$$

Assim, demonstramos a seguinte proposição:

Proposição 5.1. *Se A^c for o evento complementar de A , então $P(A) = 1 - P(A^c)$.*

Essa proposição pode ser enunciada mais geralmente, considerando uma **partição** qualquer do conjunto Ω . Por partição de um conjunto entendemos qualquer coleção finita de conjuntos disjuntos cuja união resulta no próprio conjunto. Para sermos precisos, apresentamos a seguinte definição.

Definição 5.1. Seja Ω um conjunto não vazio. Se $P = \{\Omega_1, \Omega_2, \dots, \Omega_n\}$ é formado por subconjuntos de Ω tais que

$$\Omega_1 \cap \Omega_2 \cap \dots \cap \Omega_n = \emptyset$$

e

$$\bigcup_{j=1}^{n=1} \Omega_j = \Omega,$$

então dizemos que P é uma partição de Ω .

A seguir, apresentamos uma generalização para a Proposição 5.1.

Proposição 5.2. *Seja $P = \{\Omega_1, \Omega_2, \dots, \Omega_n\}$ uma partição de um espaço amostral Ω . Então, para qualquer $1 \leq i \leq n$,*

$$P(\Omega_i) = 1 - \sum_{j \neq i} P(\Omega_j)$$

Se $\Omega = \bigcup_{j=1}^n \Omega_j$ é uma união disjunta, então

$$A \cap \Omega = A \cap (\cup_{j=1}^n \Omega_j) = \cup_{j=1}^n (A \cap \Omega_j)$$

também será uma união disjunta. Daí, segue que

$$P(A) = P(A \cap \Omega) = P(\cup_{j=1}^n (A \cap \Omega_j)) = \sum_{j=1}^n P(A \cap \Omega_j).$$

Isso prova um importante resultado chamado de **Teorema da probabilidade total**.

Teorema 5.1 (Teorema da probabilidade total). *Seja $P = \{\Omega_1, \Omega_2, \dots, \Omega_n\}$ uma partição de um espaço amostral Ω e A um evento qualquer de Ω . Então,*

$$P(A) = \sum_{j=1}^n P(A \cap \Omega_j).$$

3.3 Probabilidade da união qualquer de eventos

O Axioma 3 da definição de probabilidade diz como proceder para se calcular a probabilidade da soma de eventos disjuntos. Já a Proposição 5.2 diz como calcular a probabilidade de um evento que pertence a uma partição de Ω . Agora, dados dois eventos quaisquer, não necessariamente disjuntos, como calcular a probabilidade da ocorrência da união desses eventos? O próximo resultado responde a essa questão.

Antes de estabelecê-lo, porém, é útil observar que um dado conjunto M sempre pode ser escrito como uma união disjunta em função de outro conjunto n qualquer, da seguinte forma:

$$M = (M \cap N) \cup (M \cap N^c). \quad (5.1)$$

Teorema 5.2. *Dado dois eventos A e B , a probabilidade de, pelo menos, um deles ocorrer é igual a soma das probabilidades de cada um menos a probabilidade de ambos ocorrerem simultaneamente, ou seja:*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Demonstração: Basta notar que

$$A \cup B = (A \cap B^c) \cup (A \cap B) \cup (B \cap A^c), \quad (5.2)$$

e que essa união é disjunta. De fato, pela Equação 19 e pelo Axioma 3 da probabilidade, temos que

$$P(A) = P(A \cap B) + P(A \cap B^c) \quad (5.3)$$

e

$$P(B) = P(A \cap B) + P(A^c \cap B). \quad (5.4)$$

Somando as equações 5.3 e 5.4, obtemos

$$P(A) + P(B) = 2P(A \cap B) + P(A \cap B^c) + P(A^c \cap B).$$

Daí,

$$\begin{aligned} P(A) + P(B) - P(A \cap B) &= P(A \cap B) + P(A \cap B^c) + P(A^c \cap B) \\ &= P(A \cap B) \cup (A \cap B^c) \cup (A^c \cap B) \\ &= P(A \cup B), \end{aligned}$$

sendo que a última igualdade ocorre por 5.2. Assim, fica estabelecido o resultado. ♦

Em particular, se A e B são mutuamente exclusivos, então $P(A \cap B) = 0$. Logo,

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

Exemplo 5.9. Um estudo realizado por uma empresa de recursos humanos mostrou que 45% dos funcionários de uma multinacional saíram da empresa porque estavam insatisfeitos com seus salários; 28% porque consideraram que a empresa não possibilitava o crescimento profissional; e 8% indicaram insatisfação tanto com o salário como com sua impossibilidade de crescimento profissional. Considere o evento S : “o funcionário sai da empresa em razão do salário” e o evento I : “o funcionário sai da empresa em razão da impossibilidade de

crescimento profissional". Nessas condições, a probabilidade de um funcionário sair da empresa devido à insatisfação com o salário ou à insatisfação com sua impossibilidade de crescimento profissional é de

$$P(S \cup I) = P(S) + P(I) - P(S \cap I) = 0,45 + 0,28 - 0,08 = 0,65.$$

Esse resultado pode ser generalizado. Para A_1, A_2, A_3 eventos de Ω , tem-se que

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

Para n eventos, tem-se

$$\begin{aligned} P(\bigcup_{j=1}^n A_j) &= \sum_{j=1}^n P(A_j) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \\ &\quad - \sum_{i < j < k < l} P(A_i \cap A_j \cap A_k \cap A_l) + \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

A prova desse resultado é baseada no **princípio da indução finita** e não será feita aqui.

4 Probabilidade conjunta e dependência

Em muitas situações, precisamos analisar a probabilidade de ocorrência simultânea de dois ou mais eventos de um espaço amostral. Na seção anterior, essa necessidade foi requerida no Exemplo 5.9. Entenderemos como **probabilidade conjunta** de dois ou mais eventos a probabilidade da interseção deles. Isto é, se A_1, A_2, \dots, A_n são eventos de Ω , então

$$P(A_1, A_2, \dots, A_n) := P(A_1 \cap A_2 \cap \dots \cap A_n)$$

é a probabilidade conjunta deles.

No caso particular do Exemplo 5.9, tínhamos a informação da probabilidade conjunta dos eventos S e I , isto é, foi dado que $P(S \cap I) = 0,08$, o que possibilitou a resolução do problema. A questão que surge é: como proceder no caso em que essa informação não estiver explícita?

Para simplificar a discussão, vamos trabalhar com dois eventos, digamos A e B . Já sabemos o que acontece quando os eventos A e B são mutuamente exclusivos, isto é, quando $A \cap B = \emptyset$. Nesse caso, $P(A, B) = P(A \cap B) = 0$.

Todavia, para responder a essa questão mais geralmente, precisaremos iniciar uma discussão sobre **dependência** de eventos.

4.1 Dependência

Introduzimos esta discussão a partir de um exemplo simples, apresentado no início deste capítulo: o Exemplo 5.3.

Exemplo 5.10. Suponhamos que, no Exemplo 5.3, a moeda lançada seja honesta. Então, cada vetor do espaço amostral Ω_0 é equiprovável, isto é, ocorre com a mesma probabilidade.⁸ No caso, como $\#(\Omega_0)$, temos que

$$P(\{(k, k)\}) = P(\{(k, c)\}) = P(\{(c, k)\}) = P(\{(c, c)\}) = \frac{1}{4}$$

Logo,

$$P(A) = P(B) = \frac{1}{2}$$

Como o evento

$$A \cap B = \{\text{cara nas duas entradas}\} = \{(k, k)\},$$

vemos que a probabilidade conjunta dos vetores A e B é dada por

⁸ Aqui, a medida de probabilidade natural é a definida no Exemplo 5.7.

$$P(A, B) = P(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A) \cdot P(B).$$

Analisando esse exemplo, podemos notar que a ocorrência ou não de cara na primeira entrada do vetor não interfere na ocorrência de cara na segunda entrada do vetor, isto é, a ocorrência ou não do evento A não interfere na ocorrência do evento B . Sempre que isso acontece, dizemos que os eventos A e B são **independentes**.

Matematicamente, a independência de eventos se caracteriza pela Equação (5.5), na definição abaixo.

Definição 5.2. Dois eventos A e B de um espaço amostral Ω são ditos independentes se

$$P(A, B) = P(A) \cdot P(B). \quad (5.5)$$

Pela definição, vemos que, de fato, os eventos A e B do Exemplo 5.10 são independentes.

Uma observação importante é que se A e B são mutuamente exclusivos, isto é, se $A \cap B = \emptyset$, não significa que eles sejam independentes. Pelo contrário, a independência ocorrerá apenas caso um desses eventos tenha probabilidade zero.

Quando a Equação 5.5 não ocorre, dizemos que os eventos são **dependentes**. Uma interpretação heurística para a situação de dependência entre eventos é de que a ocorrência ou não ocorrência de um interfere na ocorrência do outro.

Com essa noção de dependência, faz sentido pensar no quanto a ocorrência ou não de um evento interfere na probabilidade de outro ocorrer.

4.2 Probabilidade condicional

No Exemplo 5.10, vimos que o evento A não interfere na ocorrência de B , pois eles são independentes. O mesmo não ocorre quando os eventos

são dependentes. A fim de quantificar essa interferência, definimos o que chamaremos de **probabilidade condicional**, dado que um evento ocorre.

Definição 5.3. Sejam Ω um espaço amostral e $B \in \wp(\Omega)$, tal que $P(B) > 0$. A probabilidade condicional de A dado B é definida por

$$P(A|B) = \frac{P(A, B)}{P(B)}, \text{ para qualquer } A \in \wp(\Omega).$$

A fim de melhorar nossa intuição a respeito da probabilidade condicional, apresentamos o seguinte exemplo:

Exemplo 5.11. Suponha que uma caixa contenha 3 moedas: 2 honestas e uma de 2 caras. Ao retirar da caixa uma moeda ao acaso e jogá-la, qual é a probabilidade de a moeda ter sido a de 2 caras, dado que o resultado final foi cara?

Podemos analisar o problema como sendo um **experimento composto**, isto é, teremos de analisar mais de um experimento para tirar conclusões a respeito de um evento que, de alguma forma, relaciona esses experimentos. No caso, podemos considerar os seguintes experimentos:

- Experimento 1: “escolha de uma moeda da caixa”;
- Experimento 2: “lançamento de uma moeda da caixa e observação da face”.

O conjunto $\Omega_1 = \{m_H, m_V\}$ corresponde ao espaço amostral do Experimento 1, em que m_H e m_V significam, respectivamente, moeda honesta e moeda viciada. Pelas hipóteses do problema, na caixa há uma moeda viciada, no sentido de que possui 2 faces, e 2 honestas. Isso significa que, ao escolher aleatoriamente uma moeda entre as 3 disponíveis na caixa, temos que a $P_1(\{\text{sair moeda honesta}\}) = 2/3$ e $P_1(\{\text{sair moeda viciada}\}) = 1/3$.

Por outro lado, ao considerarmos o Experimento 2, o conjunto $\Omega_2 = \{k, c\}$ é seu espaço amostral. A presença da moeda viciada na caixa faz com que a quantidade de caras e coroas dentro da caixa esteja desequilibrada, isto é, a $P_2(\{\text{sair coroa}\}) = 1/3$ e $P_2(\{\text{sair cara}\}) = 2/3$.

O experimento composto, nesse caso, seria:

- Experimento composto: “escolha e lançamento de uma moeda e observação do resultado em 2 dimensões: quanto ao tipo de moeda e quanto a sua face”.

Podemos pensar no conjunto

$$\tilde{\Omega} = \{(m_{H'}, k), (m_{H'}, c), (m_{V'}, k), (m_{V'}, c)\}$$

como o espaço amostral para o Experimento composto, sobre o qual está definida uma medida de probabilidade

$$\tilde{P}(.) = \frac{\#(.)}{\#(\tilde{\Omega})} .$$

Consideraremos os eventos A e B contidos em $\tilde{\Omega}$, tais que $A = \{\text{moeda é viciada}\}$ e $B = \{\text{sair cara}\}$, e a partição de $\tilde{\Omega}$ pelo evento B , isto é,

$$\tilde{\Omega} = \tilde{\Omega}_B \cup \tilde{\Omega}_{B^c},$$

em que os conjuntos

$$\tilde{\Omega}_B = \{(w, \text{sair cara}); w \in \Omega_1\} = \{(m_{H'}, k), (m_{V'}, k)\}.$$

e

$$\tilde{\Omega}_{B^c} = \{(w, \text{não sair cara}); w \in \Omega_1\} = \{(m_{H'}, c), (m_{V'}, c)\}.$$

Se considerarmos o evento

$$\tilde{A}_B = \{\text{(moeda viciada, sair cara)}\} = \{(m_{V'}, k)\}.$$

então podemos calcular a probabilidade condicional $P(A/B)$ de a moeda ser viciada, considerando a informação de que saiu cara assim:

$$P(A/B) = \frac{\tilde{P}(\tilde{A}_B)}{\tilde{P}(\tilde{\Omega}_B)} = \frac{1}{2} .$$

Por outro lado, fazendo algumas manipulações algébricas, obtemos

$$\begin{aligned}
 P(A/B) &= \frac{\tilde{P}(\tilde{A}_B)}{\tilde{P}(\tilde{\Omega}_B)} = \frac{\#(\tilde{A}_B)}{\#(\tilde{\Omega})} = \\
 &= \frac{\#(\tilde{A}_B)}{\#(\tilde{\Omega}_B)} = \frac{\#(\tilde{A} \cap B)}{\#(\tilde{\Omega} \cap B)} = \frac{\#(A \cap B)}{\#B} = \frac{\tilde{P}(A \cap B)}{\tilde{P}(B)},
 \end{aligned} \tag{5.6}$$

em que \tilde{P} é uma medida de probabilidade conjunta definida sobre $\tilde{\Omega}$.

O que a Equação (5.6) sugere é que a probabilidade condicional de um evento A , dado um evento fixo B , sendo $A, B \subset \tilde{\Omega}$, é igual à razão entre a medida de probabilidade \tilde{P} associada à $\tilde{\Omega}$ avaliada nos eventos $A \cap B$ e B , respectivamente. Essa interpretação serve de motivação para a Definição 5.3.

Um fato importante é que a probabilidade condicional $P(.|B)$ é uma medida de probabilidade. Esse fato pode ser facilmente verificado e será deixado como exercício. Assim, todas as propriedades operatórias envolvendo medidas de probabilidade naturalmente serão válidas para a probabilidade condicional.

4.3 Teorema de Bayes

Nesta subseção, apresentaremos um teorema que pode ser visto como uma generalização do conceito de probabilidade condicional. Para isso, considere um espaço amostral Ω , uma partição $P = \{A_1, A_2, \dots, A_n\}$ de Ω e um evento B qualquer de Ω .

Teorema 5.3. (Teorema de Bayes)

$$P(A_j/B) = \frac{P(B/A_j)P(A_j)}{\sum_{j=1}^n P(B/A_j)P(A_j)}, \text{ qualquer que seja } j = 1, 2, \dots, n.$$

Demonstração: Pelo Teorema 5.1, temos que

$$P(B) = \sum_{j=1}^n P(B \cap A_j). \quad (5.7)$$

Da definição de probabilidade condicional, temos que

$$P(B \cap A_j) = P(B/A_j)P(A_j), \quad j = 1, 2, \dots, n. \quad (5.8)$$

Assim, se substituirmos a Equação 26 na Equação 25, obtemos, por um lado, que

$$P(B) = \sum_{j=1}^n P(B/A_j)P(A_j). \quad (5.9)$$

Por outro lado, também pela definição de probabilidade condicional,

$$P(A_j/B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B/A_j)P(A_j)}{\sum_{j=1}^n P(B/A_j)P(A_j)}. \quad (5.10)$$

A última igualdade na Equação 28 foi obtida utilizando-se a Equação 5.9. ◆

A fórmula de Bayes⁹ é útil quando conhecemos as probabilidades dos eventos da partição e a probabilidade condicional de B , dado A_j , mas não conhecemos diretamente a probabilidade de B .

Como aplicação do Teorema 5.3, vamos resolver o Exemplo 5.11, resolvido na subseção anterior.

Exemplo 5.12. Considere a situação do Exemplo 5.11 e os eventos listados abaixo, pertencentes ao espaço amostral $\bar{\Omega}$:

$$A_1 = \{\text{a moeda retirada é honesta}\},$$

$$A_2 = \{\text{a moeda retirada é a de duas caras}\}$$

e

$$B = \{\text{o resultado final é cara}\}.$$

⁹ A literatura, às vezes, se refere à fórmula de Bayes como sendo a “fórmula de probabilidades *a posteriori*”, $P(A_j/B)$, e as probabilidades dos eventos da partição, $P(A_j)$, são chamadas de probabilidades *a priori*.

Aplicando a fórmula de Bayes, teremos

$$P(A_2/B) = \frac{P(A_2)P(B/A_2)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2)} = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{1}{2}.$$

Exercícios

1. Determine a probabilidade de cada evento:
 - a) um número par aparece no lançamento de um dado não viciado.
 - b) um rei aparece ao extraír-se uma carta de um baralho.
 - c) pelo menos uma cara aparece no lançamento de 3 moedas.
 - d) pelo menos uma cara aparece no lançamento de n moedas.
 - e) duas copas aparecem ao se retirarem duas cartas de um baralho.
 - f) uma carta de copas e uma de ouros aparecem ao se extraírem duas cartas de um baralho.

2. Um ponto é escolhido ao acaso no intervalo $[1, 4]$. A probabilidade de que o ponto escolhido esteja entre 1 e 3 é de:
 - a) $\frac{1}{4}$.
 - b) $\frac{1}{3}$.
 - c) $\frac{1}{2}$.
 - d) $\frac{2}{3}$.
 - e) $\frac{3}{4}$.

3. A probabilidade de um atirador acertar um alvo é $1/4$. Se ele atirar 5 vezes, a probabilidade de ele acertar 2 tiros é de:

- a) $\frac{127}{512}$.
- b) $\frac{291}{512}$.
- c) $\frac{131}{512}$.
- d) $\frac{133}{512}$.
- e) $\frac{135}{512}$.

4. (IFG) No lançamento de dois dados perfeitos, qual é a probabilidade de o produto dos dois números obtidos ser par?

- a) $\frac{3}{4}$.
- b) $\frac{1}{2}$.
- c) $\frac{1}{3}$.
- d) $\frac{2}{3}$.
- e) $\frac{3}{5}$.

5. Num lote de 12 peças, há 4 defeituosas. Se 2 peças são retiradas aleatoriamente, determine:

- a) a probabilidade de ambas serem defeituosas.
- b) a probabilidade de ambas não serem defeituosas.
- c) a probabilidade de, pelo menos, uma ser defeituosa.

6. Uma urna contém 5 bolas verdes e 4 vermelhas. Se 3 delas são retiradas ao acaso, determine:
- a probabilidade de todas serem verdes.
 - a probabilidade de uma ser vermelha.
 - a probabilidade de, pelo menos, uma ser verde.
7. Três máquinas, A, B e C, produzem, respectivamente, 50%, 40% e 10% do total de peças de uma fábrica. As porcentagens de peças defeituosas nas respectivas máquinas são 3%, 5% e 2%. Uma peça é sorteada ao acaso e verifica-se que é defeituosa. Qual a probabilidade de que a peça tenha vindo da máquina B?
8. (IFG) Sabendo que $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$ e $P(A \cap B) = \frac{1}{6}$, podemos afirmar que:
- $P(B/A) = \frac{1}{3}$.
 - $P(A \cup B) = \frac{5}{6}$.
 - $P(A/B) = \frac{1}{2}$.
 - $P(A \cup B) = \frac{2}{3}$.
 - $P(A \cap B) = \frac{5}{6}$.
9. (IFG) Numa bolsa, há 5 moedas de R\$1,00 e 4 moedas de R\$0,50. A probabilidade de, ao retirarmos da bolsa 2 moedas aleatoriamente, obtermos R\$1,50 é de:
- $\frac{4}{9}$.
 - $\frac{2}{5}$.
 - $\frac{4}{5}$.
 - $\frac{1}{4}$.
 - $\frac{5}{9}$.

10. (IFG) Uma companhia de seguros vendeu apólices a 5 pessoas, todas da mesma idade e nas mesmas condições de saúde. A companhia sabe, através de estatísticas atuariais, que a probabilidade de cada uma dessas pessoas estar viva daí a 30 anos é de $\frac{2}{3}$. Qual é a probabilidade de, pelo menos, 3 dessas pessoas estarem vivas daqui a 30 anos?
- a) $\frac{32}{243}$.
b) $\frac{64}{243}$.
c) $\frac{192}{243}$.
d) $\frac{152}{243}$.
e) $\frac{132}{243}$.
11. (IFG) De acordo com certa tábua de mortalidade, a probabilidade de José e Manuel estarem vivos daqui a 20 anos é de 0,6 e 0,9, respectivamente. A probabilidade de um deles estar vivo e o outro estar morto daqui a 20 anos é de:
- a) 0,42.
b) 0,54.
c) 0,46.
d) 0,04.
e) 0,36.
12. (IFG) Os seguintes grupos de pessoas estão numa sala: 5 homens maiores de 21 anos; 4 homens com menos de 21 anos; 6 mulheres maiores de 21 anos e 3 mulheres menores de 21 anos de idade. Uma pessoa é escolhida ao acaso. Definem-se os seguintes eventos:

$$A = \{\text{a pessoa é maior de 21 anos}\};$$

$$B = \{\text{a pessoa é menor de 21 anos}\};$$

$$C = \{\text{a pessoa é homem}\};$$

e

$$D = \{\text{a pessoa é mulher}\}.$$

Com base nessas informações, marque a alternativa correta:

a) $P(B \cap D) = \frac{7}{36}$.

b) $P(A^c \cap B^c) = \frac{7}{36}$.

c) $P(A \cup C) = \frac{11}{36}$.

d) $P(B \cup C) = \frac{2}{18}$.

e) $P(B \cup D) = \frac{13}{18}$.

13. (IFG) Os algarismos 1, 2, 3, 4 e 5 são escritos em 5 cartões diferentes. Esses cartões são escolhidos (sem reposição) aleatoriamente, e os algarismos que vão aparecendo são escritos da esquerda para a direita, formando um número de 5 algarismos. Assim,

a) $\frac{2}{3}$ é a probabilidade de o número formado ser par.

b) $\frac{1}{20}$ é a probabilidade de o número formado ser ímpar.

c) $\frac{1}{20}$ é a probabilidade de o número formado ser ímpar, se considerarmos a escolha com reposição.

d) $\frac{2}{5}$ é a probabilidade de o número formado ser divisível por 5.

e) $\frac{2}{10}$ é a probabilidade de o número formado ser par, se considerarmos a escolha com reposição.

14. (IFG) Um par de dados é atirado. Encontre a probabilidade de a soma ser maior ou igual a 10, se um 5 aparece no primeiro dado.

a) $\frac{1}{6}$.

b) $\frac{1}{12}$.

c) $\frac{5}{12}$.

d) $\frac{5}{18}$.

e) $\frac{1}{18}$.

15. (IFG) Sejam A e B eventos de um espaço amostral Ω . Podemos afirmar que:

- a) $P(A) > 1$, se $A \subset B$.
- b) Se $A \subset B$, então $P(B) \leq P(A)$.
- c) Se $A \subset B$, então $P(A) \leq P(B)$.
- d) $P(\Omega) = 0$ e $P(\emptyset) = 1$.
- e) $\Omega = (A \cup A^c) - (A \cap A^c)$.

16. (IFG) Na urna u_1 , há 3 bolas pretas, 3 amarelas e 6 brancas. Na urna u_2 , há uma bola preta, 5 amarelas e 2 brancas. Escolhe-se uma urna ao acaso e dela extraí-se uma bola ao acaso. Assim,

- a) se a bola for preta, então $\frac{2}{3}$ é a probabilidade de ela ter vindo da urna u_2 .
- b) se a bola for amarela, então $\frac{2}{7}$ é a probabilidade de ela ter vindo da urna u_2 .
- c) se a bola for branca, então $\frac{2}{3}$ é a probabilidade de ela ter vindo da urna u_1 .
- d) se a bola for branca, então $\frac{1}{3}$ é a probabilidade de ela ter vindo da urna u_1 .
- e) se a bola for preta, então $\frac{5}{7}$ é a probabilidade de ela ter vindo da urna u_1 .

17. (IFB) Uma porca irá parir 10 leitões. Suponha que todos nascerão vivos. Considerando que existe 50% de chance de cada leitão nascer macho, então a probabilidade de que nasçam exatamente 4 machos é de:

- a) $210.(0,5)^{10}$.
- b) $15.(0,5)^{10}$.
- c) $210.(0,5)^6$.
- d) $210.(0,5)^4$.
- e) $1 - 15.(0,5)^6$.

REFERÊNCIAS BIBLIOGRÁFICAS



Referências bibliográficas

JAMES, B. *Probabilidade*: um curso de nível intermediário. 3. ed. Rio de Janeiro: Impa, 2006.

MAGALHÃES, M. N.; LIMA, A. C. P. *Noções de probabilidade e estatística*. São Paulo: Edusp, 2004.

LIMA, E. L. *Curso de análise*. V. 1. 10. ed. Rio de Janeiro: Associação Instituto de Matemática Pura e Aplicada, 2002.

BEARZOTI, E.; BUENO FILHO, J. S. S.; *Introdução à inferência estatística*. Lavras-MG: Ufla/Faepe, 2000.

MUNIZ, J. A.; BUENO FILHO, J. S. S.; VEIGA, R. D. *Técnicas de amostragem*. Lavras-MG: Ufla/Faepe, 1999.

LIMA, P. C.; ABREU, A. R. *Técnicas de amostragem*. Lavras-MG: Ufla/Faepe, 1999.

FONSECA, J. S.; MARTINS, G. A. *Curso de estatística*. 6. ed. São Paulo: Atlas, 1996.

O'HAGAN, A. *Bayesian inference*. Kendall's advanced theory of statistics. London: Cambridge University Press, 1994.

