

# 1

## A Natureza da Estatística

### O QUÊ?

Especificidade do pensamento estatístico a partir de problemas. Conceitos: população e amostra, parâmetro e estimador. Variáveis estatísticas e suas classificações. Organização dos dados em tabelas de frequências. Representações gráficas adequadas para os diferentes tipos de variáveis. Noções básicas de amostragem.

### POR QUÊ?

A Estatística está presente no mundo contemporâneo e chega aos cidadãos em todos os meios de comunicação. Diariamente somos confrontados com informações estatísticas sobre temas como Economia, Educação, Esportes, Saúde, Meio-Ambiente, entre outros. Tais informações orientam decisões em nossas vidas pessoais e permitem-nos exercer nossas responsabilidades como cidadãos. Um conhecimento básico de Estatística é fundamental na formação do cidadão para que este possa, de forma competente, apreciar e criticar argumentos baseados em dados.





Cadastre-se como colaborador no site do projeto: [umlivroaberto.com](http://umlivroaberto.com)

Título: A Natureza da Estatística

Ano/ Versão: 2021 / versão 1.3 de 26 de julho de 2021

Editora Instituto Nacional de Matemática Pura e Aplicada (IMPA-OS)

Realização: Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP)

Produção: Associação Livro Aberto

Coordenação: Fabio Simas e Augusto Teixeira ([livroaberto@impa.br](mailto:livroaberto@impa.br))

Autores: Flávia Landim (coordenadora da equipe - UFRJ)

José Ezequiel Soto Sanches

Nei Rocha (UFRJ)

Vanessa Matos (SEduc Angra dos Reis e Mesquita)

Letícia Rangel (Colégio de Aplicação da UFRJ)

Revisão: Cydara Ripoll

Letícia Rangel

Luciane Velasque

Design: Andreza Moreira (Tangentes Design)

Diagramação: Tarso Caldas

Ilustrações: Miller Guglielmo

Gráficos: Beatriz Cabral e Tarso Caldas

Capa: Foto de Chuttersnap no Unsplash  
<https://unsplash.com/photos/8I423fRMwjM>

Desenvolvido por

Licença:



Patrocínio:





# Introdução

## Habilidades da BNCC trabalhadas no Capítulo

### EM13MAT102

Analisar tabelas, gráficos e amostras de pesquisas estatísticas em relatórios divulgados por diferentes meios de comunicação, identificando, quando for o caso, inadequações que possam induzir a erros de interpretação, como escalas e amostras não apropriadas.

### EM13MAT106

Identificar situações da vida cotidiana nas quais seja necessário fazer escolhas levando-se em conta os riscos probabilísticos (usar este ou aquele método contraceptivo, optar por um tratamento médico em detrimento de outro etc.)

### EM13MAT406

Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.

## Pré-requisitos

### EF08MA23

Avaliar a adequação de diferentes tipos de gráficos para representar um conjunto de dados de uma pesquisa.

### EF08MA24

Classificar as frequências de uma variável contínua de uma pesquisa em classes, de modo que resumam os dados de maneira adequada para a tomada de decisões.

### EF08MA25

Obter os valores de medidas de tendência central de uma pesquisa estatística (média, moda e mediana) com a compreensão de seus significados e relacioná-los com a dispersão de dados, indicada pela amplitude.

### EF08MA26

Selecionar razões, de diferentes naturezas (física, ética ou econômica), que justificam a realização de pesquisas amostrais e não censitárias, e reconhecer que a seleção da amostra pode ser feita de diferentes maneiras (amostra casual simples, sistemática e estratificada).

### EF08MA27

Planejar e executar pesquisa amostral, selecionando uma técnica de amostragem adequada, e escrever relatório que contenha os gráficos apropriados para representar os conjuntos de dados, destacando aspectos como as medidas de tendência central, a amplitude e as conclusões.

### EF09MA21

Analizar e identificar, em gráficos divulgados pela mídia, os elementos que podem induzir, às vezes propositalmente, erros de leitura, como escalas inapropriadas, legendas não explicitadas corretamente, omissão de informações importantes (fontes e datas), entre outros.

### EF09MA22

Escolher e construir o gráfico mais adequado (colunas, setores, linhas), com ou sem uso de planilhas eletrônicas, para apresentar um determinado conjunto de dados, destacando aspectos como as medidas de tendência central.

### EF09MA23

Planejar e executar pesquisa amostral envolvendo tema da realidade social e comunicar os resultados por meio de relatório contendo avaliação de medidas de tendência central e da amplitude, tabelas e gráficos adequados, construídos com o apoio de planilhas eletrônicas.

## Observações

a) Os pré-requisitos não necessariamente foram contemplados no Ensino Fundamental. Por essa razão, muitos deles serão abordados nesse capítulo e no capítulo de "Medidas de Posição e Dispersão" que dá sequência a esse capítulo.

b) Na seção "Para saber +", ao final desse capítulo, utilizamos o termo *progressão aritmética*. O conhecimento deste tópico não é um pré-requisito. A conexão entre conceitos da Matemática é favorável para a visão do estudante sobre a Matemática como um todo.

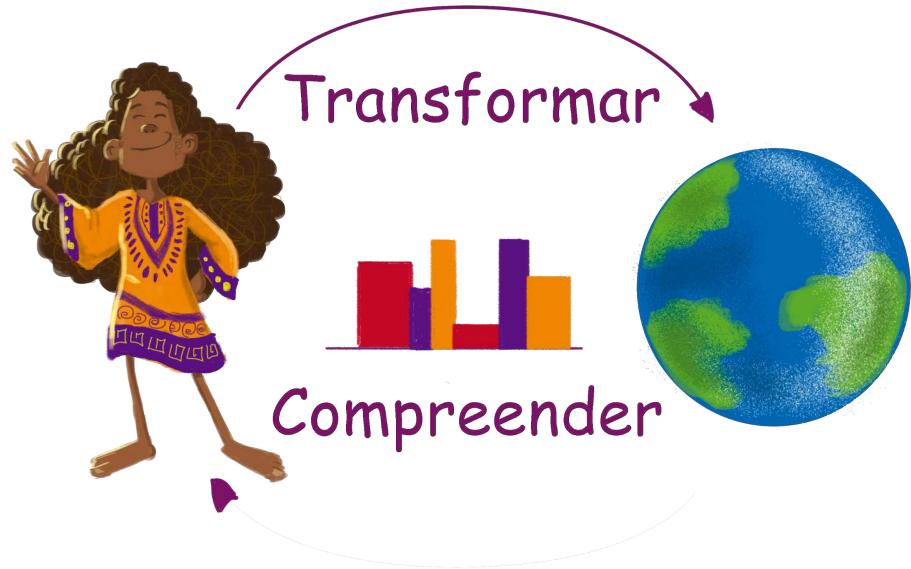
A Estatística está presente no mundo contemporâneo e chega aos cidadãos em todos os meios de comunicação e é a ferramenta por excelência no tratamento de modelagem de fenômenos aleatórios (não-determinísticos). Diariamente somos confrontados com informações estatísticas sobre temas que variam de Economia à Educação, de filmes a esportes, de comida à medicina, e de pesquisas de opinião a comportamento social. Tais informações orientam decisões em nossas vidas pessoais e permitem-nos exercer nossas responsabilidades como cidadãos. (Franklin et al., 2007).

A produção de conhecimento - sempre em constante evolução e reavaliação - nas mais variadas áreas muitas vezes requer o conhecimento estatístico.

A relevância do raciocínio estatístico e do conhecimento para o efetivo funcionamento na sociedade da informação levou à introdução do termo **Letramento Estatístico**: "A capacidade de compreender e avaliar criticamente resultados estatísticos que permitem a vida diária, acompanhada da capacidade de apreciar como o pensamento estatístico pode contribuir em decisões públicas e privadas, profissionais e pessoais." (Batanero e Borovcnik, 2016).

De acordo com Veaux e Velleman (2008), o desafio para o estudante (e o professor)

de Estatística introdutória é que, como na literatura e na arte, navegar por e dar sentido a exige não somente um conjunto de regras e axiomas, mas experiência de vida e "senso comum". São várias habilidades a serem trabalhadas e a maior parte delas exige capacidade de avaliação crítica em adição à manipulação matemática. A capacidade de avaliação crítica é adquirida com exemplos e experiência e isso demanda mais tempo.



## Objetivo geral

Identificar o pensamento estatístico a partir de suas ideias fundamentais, a saber, população e amostra, parâmetro e estimador, distribuições empíricas de dados.

Este capítulo trabalha as competências gerais da Educação Básica. Destacamos, de modo especial, as competências

- Exercitar a curiosidade intelectual e recorrer à abordagem própria das ciências, incluindo a investigação, a reflexão, a análise crítica, a imaginação e a criatividade, para investigar causas, elaborar e testar hipóteses, formular e resolver problemas e criar soluções (inclusive tecnológicas) com base nos conhecimentos das diferentes áreas.
- Argumentar com base em fatos, dados e informações confiáveis, para formular, negociar e defender ideias, pontos de vista e decisões comuns que respeitem e promovam os direitos humanos, a consciência socioambiental e o consumo responsável em âmbito local, regional e global, com posicionamento ético em relação ao cuidado de si mesmo, dos outros e do planeta.

Este capítulo aborda os conteúdos de organização e representação dos dados (incluindo agrupamentos de dados em classes), a construção de gráficos apropriados (incluindo o histograma), a interpretação e a análise crítica apresentadas em relatórios descritivos destacados na habilidade. Os conteúdos:

- a) realização de pesquisas considerando o planejamento, a discussão (se será censitária ou por amostra),
- b) seleção de amostras,
- c) elaboração e aplicação de instrumentos de coleta

serão trabalhados de forma transversal ao longo dos capítulos que tratam de Estatística, revisitando pré-requisitos previstos pela BNCC para o Ensino Fundamental.

As atividades propostas envolvem o uso da Estatística em diferentes situações, motivando o pensamento estatístico a partir de suas ideias fundamentais, a saber, população e amostra, parâmetro e estimador, distribuição e caracterizações da distribuição (posição e dispersão). Essas atividades não têm como objetivo o cálculo das medidas, mas a sua compreensão estrutural. Também serão trabalhados alguns distratores nessas atividades tais como:

- a) explorar a diferença entre um gráfico de barras e um histograma;
- b) destacar que a informação importante no gráfico de barras, adequado para variáveis qualitativas ou quantitativas discretas que assumem um conjunto moderado de valores, é a frequência na qual cada resposta ocorre,
- c) destacar ainda que, para efeito de comparações múltiplas, a frequência deve ser relativa ou porcentagem, dado que diferentes conjuntos podem ter tamanhos diferentes.

Neste capítulo serão apresentadas algumas atividades envolvendo a realização de pesquisas e coleta de dados e, no final do capítulo, será sugerida a realização de um projeto que deverá ser realizado ao longo de pelo menos três meses paralelamente às aulas. O projeto envolverá a formulação de um problema a ser investigado (de preferência envolvendo outra disciplina), a definição da população, a construção de um questionário, a coleta de dados (amostra ou censo), a análise dos resultados obtidos construindo gráficos e calculando medidas-resumo e a confecção de relatório final. Na conclusão do projeto, o capítulo "Medidas de Posição e Dispersão" já terá sido trabalhado. Recomenda-se que essa atividade seja preferencialmente trabalhada no primeiro ou segundo ano do Ensino Médio, pois no último ano há maior limitação de tempo em razão dos vários exames a serem realizados pelos estudantes.

Neste capítulo incluem-se:

- a) apresentação do diagrama de pontos introduzindo o conceito de distribuição empírica tanto em seu aspecto morfológico quanto variacional logo na primeira atividade revelando com isso a essência da Estatística;
- b) reflexão sobre possíveis equivalências, do ponto de vista estatístico, de medidas-resumo com a finalidade de tomada de decisão sob incerteza;
- c) utilização de uma base de dados reais de uma pesquisa já realizada;
- d) discussão sobre a adequação entre tipo de variável e tipo de gráfico;
- e) uso de tecnologia para a construção de gráficos;
- f) conceituação de parâmetro e estimador, elementos cruciais na Estatística.

g) abordagem da estatística e seus problemas, privilegiando o pensamento estatístico para interpretação dos resultados, ao invés de um puro tratamento matemático dos cálculos que levam aos resultados.

De acordo com [Batanero e Borovcnik \(2016\)](#), mesmo que os métodos de análise de dados nessa fase do ensino envolvam somente calcular e interpretar porcentagens ou medidas estatísticas simples, bem como interpretar vários tipos de gráficos, os autores sugerem que os estudantes apresentem problemas na compreensão dos conceitos e na relação desses conceitos para o contexto de modo a ter algum significado. Uma razão para essas dificuldades é que o ensino, em geral, foca sobre a aplicação de métodos em detrimento à interpretação de resultados em um dado contexto, buscando, assim, de forma equivocada, dar a estes um caráter determinístico.

Os distratores apresentados a seguir refletem a experiência dos envolvidos com o desenvolvimento desse capítulo.

- a) Confundir o valor da variável com o da frequência.
- b) Em caso de variável quantitativa discreta, considerar apenas os valores da variável apresentados na tabela ignorando as frequências.
- c) Confundir gráfico de barras com o histograma.
- d) Dificuldade de interpretar um resultado obtido via procedimento de inferência estatística.

Apesar de variáveis e variação também aparecerem em muitas áreas da Matemática, a Matemática lida com variação funcional (determinística) enquanto que a Estatística lida com variação aleatória. Portanto, um objetivo da Educação Estatística é capacitar os estudantes a raciocinar sobre dados em contextos sob condições de incerteza, e distinguir entre raciocínio estatístico e raciocínio matemático. Além disso, a Estatística fornece métodos para identificar, quantificar, explicar, controlar e reduzir variação.

Para evitar o uso de vários termos com o mesmo significado: variação, variabilidade e dispersão, optamos por usar a palavra dispersão no livro.

Como estratégia pedagógica propomos usar um processo reflexivo baseado no pensamento estatístico.

- a) Cálculos não serão valorizados, o mais importante neste capítulo é a compreensão dos conceitos.
- b) As atividades deverão estar sempre bem caracterizadas a um problema a ser resolvido em um contexto específico.
- c) O uso de recursos tecnológicos para a realização de cálculos e para a construção de gráficos é recomendado. Como recurso tecnológico, fez-se a opção pelo GeoGebra, por supor que os professores, em especial os Licenciados em Matemática têm familiaridade com o programa. Além disso, o GeoGebra atende satisfatoriamente as demandas da abordagem e das atividades propostas em todo o livro, é o recurso digital que ampara o texto em outros eixos temáticos e é um programa livre e disponível em: <https://www.geogebra.org/?lang=pt>.

Como alternativa ao GeoGebra pode ser utilizado o programa computacional R, livre e de código aberto, que pode ser baixado gratuitamente em: <https://www.r-project.org/>. O R, muito popular entre os Estatísticos, permite desde cálculos simples até

análises sofisticadas. Para utilização com estudantes do Ensino Médio recomendamos a Interface Rcmdr, por ser mais amigável e não exigir conhecimento prévio dos comandos. O link (<http://gae.uniriotec.br/7/material.html>) apresenta uma apostila com os passos iniciais para utilização do Rcmdr.

## Conexões com as outras áreas do conhecimento

A estatística é uma disciplina metodológica. Ela não existe para ela própria, mas para oferecer às outras áreas de conhecimento um conjunto coerente de ideias e ferramentas para lidar com dados. Nesse sentido a estatística é naturalmente interdisciplinar de modo que é possível trabalhar com as outras áreas do conhecimento. A seguir, destacamos habilidades das outras áreas que podem ser trabalhadas conjuntamente.

### a) Linguagem e suas tecnologias

#### EM13LP32

Selecionar informações e dados necessários para uma dada pesquisa (sem excedê-los) em diferentes fontes (orais, impressas, digitais etc.) e comparar autonomamente esses conteúdos, levando em conta seus contextos de produção, referências e índices de confiabilidade, e percebendo coincidências, complementaridades, contradições, erros ou imprecisões conceituais e de dados, de forma a compreender e posicionar-se criticamente sobre esses conteúdos e estabelecer recortes precisos.

#### EM13LP33

Selecionar, elaborar e utilizar instrumentos de coleta de dados e informações (questionários, enquetes, mapeamentos, opinários) e de tratamento e análise dos conteúdos obtidos, que atendam adequadamente a diferentes objetivos de pesquisa.

#### EM13LP39

Usar procedimentos de checagem de fatos noticiados e fotos publicadas (verificar/avaliar veículo, fonte, data e local da publicação, autoria, URL, formatação; comparar diferentes fontes; consultar ferramentas e sites checadores etc.), de forma a combater a proliferação de notícias falsas (fake news).

#### EM13LP40

Analizar o fenômeno da pós-verdade – discutindo as condições e os mecanismos de disseminação de fake news e também exemplos, causas e consequências desse fenômeno e da prevalência de crenças e opiniões sobre fatos –, de forma a adotar atitude crítica em relação ao fenômeno e desenvolver uma postura flexível que permita rever crenças e opiniões quando fatos apurados as contradisserem.

### b) Ciências Humanas e Sociais Aplicadas

#### EM13CHS103

Elaborar hipóteses, selecionar evidências e compor argumentos relativos a processos políticos, econômicos, sociais, ambientais, culturais e epistemológicos, com base na sistematização de dados e informações de diversas naturezas (ex-

pressões artísticas, textos filosóficos e sociológicos, documentos históricos e geográficos, gráficos, mapas, tabelas, tradições orais, entre outros).

#### EM13CHS106

Utilizar as linguagens cartográfica, gráfica e iconográfica, diferentes gêneros textuais e tecnologias digitais de informação e comunicação de forma crítica, significativa, reflexiva e ética nas diversas práticas sociais, incluindo as escolares, para se comunicar, acessar e difundir informações, produzir conhecimentos, resolver problemas e exercer protagonismo e autoria na vida pessoal e coletiva.

#### EM13CHS301

Problematizar hábitos e práticas individuais e coletivos de produção, reaproveitamento e descarte de resíduos em metrópoles, áreas urbanas e rurais, e comunidades com diferentes características socioeconômicas, e elaborar e/ou selecionar propostas de ação que promovam a sustentabilidade socioambiental, o combate à poluição sistêmica e o consumo responsável.

#### EM13CHS402

Analizar e comparar indicadores de emprego, trabalho e renda em diferentes espaços, escalas e tempos, associando-os a processos de estratificação e desigualdade socioeconômica.

### c) Ciências da Natureza e suas tecnologias

#### EM13CNT205

Interpretar resultados e realizar previsões sobre atividades experimentais, fenômenos naturais e processos tecnológicos, com base nas noções de probabilidade e incerteza, reconhecendo os limites explicativos das ciências.

#### EM13CNT206

Discutir a importância da preservação e conservação da biodiversidade, considerando parâmetros qualitativos e quantitativos, e avaliar os efeitos da ação humana e das políticas ambientais para a garantia da sustentabilidade do planeta.

#### EM13CNT301

Construir questões, elaborar hipóteses, previsões e estimativas, empregar instrumentos de medição e representar e interpretar modelos explicativos, dados e/ou resultados experimentais para construir, avaliar e justificar conclusões no enfrentamento de situações-problema sob uma perspectiva científica.

#### EM13CNT302

Comunicar, para públicos variados, em diversos contextos, resultados de análises, pesquisas e/ou experimentos, elaborando e/ou interpretando textos, gráficos, tabelas, símbolos, códigos, sistemas de classificação e equações, por meio de diferentes linguagens, mídias, tecnologias digitais de informação e comunicação (TDIC), de modo a participar e/ou promover debates em torno de temas científicos e/ou tecnológicos de relevância sociocultural e ambiental.

#### EM13CNT303

Interpretar textos de divulgação científica que tratem de temáticas das Ciências da Natureza, disponíveis em diferentes mídias, considerando a apresentação dos dados, tanto na forma de textos como em equações, gráficos e/ou tabelas, a consistência dos argumentos e a coerência das conclusões, visando construir estratégias de seleção de fontes confiáveis de informações.

## Organização do Capítulo

O capítulo está estruturado em dois blocos principais "1) A Natureza da Estatística" e "2) Gráficos para Variáveis Quantitativas", compostos por seções explorando, organizando e praticando cada um. Na parte final do capítulo há uma seção "Para saber mais", um "Projeto Aplicado" e uma seção com 20 exercícios.

A duração prevista para trabalhar esse capítulo de forma completa, incluindo a parte final do capítulo é de 6 a 8 tempos de 50 minutos.

### Bloco 1: A Natureza da Estatística

#### Duração prevista: 3 a 4 tempos de 50 minutos

##### a) Explorando

Inclui quatro atividades para trabalhar o objetivo geral.

##### b) Organizando

Apresenta conceitos básicos de Estatística como População e Amostra, Estatística descritiva e Estatística inferencial e a classificação de variáveis. A seção é encerrada, apresentando-se gráficos adequados para variáveis qualitativas, também chamadas de variáveis categóricas.

##### c) Praticando

Inclui 4 atividades complementares desse bloco, envolvendo um levantamento de dados da turma, classificação de variáveis, construção e análise de gráficos.

### Bloco 2: Gráficos para Variáveis Quantitativas

#### Duração prevista: 2 tempos de 50 minutos

##### a) Explorando 2

Inclui atividades que envolvem construir gráficos para representar variáveis quantitativas contínuas: uma cujo objetivo é estudar a distribuição de frequências dos valores observados (histograma) e, a outra, cujo objetivo é estudar seu comportamento ao longo do tempo (gráfico de linha).

##### b) Organizando 2

Apresenta a definição dos gráficos trabalhados nas atividades do explorando e discute a necessidade de se organizar variáveis quantitativas contínuas em intervalos de classe. Inclui dois exemplos de construção de histograma. O segundo exemplo ilustra o caso de intervalos de classe de comprimentos desiguais e apresenta a definição de densidade de frequência: absoluta e relativa.

##### c) Praticando 2

Inclui duas atividades de construção de histogramas, a primeira delas incluindo um aplicativo do GeoGebra de modo a automatizar a construção do his-

tograma e poder explorar diferentes quantidades de intervalos de classe de maneira rápida e simples.

## Parte final do capítulo

### Duração prevista: 2 tempos de 50 minutos

Nessa parte final do capítulo foram incluídas seções Para saber mais, Projeto aplicado e Exercícios, além de uma lista de sugestões de vídeos sobre Estatística e páginas da internet para trabalhar com dados reais.

Para a realização do Projeto Aplicado recomenda-se um planejamento de forma transversal ao longo do ano letivo, pois, dependendo da investigação a ser conduzida pelos estudantes, o tempo demandado para a coleta dos dados e sua organização pode ser maior do que um período sequencial de 6 a 8 aulas de Matemática. Dessa forma, o ideal é propor o projeto no início de um período letivo, ano, semestre, trimestre ou bimestre, retomando-o à medida que as etapas do projeto forem sendo concluídas.



# Bloco I: A Natureza da Estatística

## Objetivo geral

Identificar o pensamento estatístico a partir de suas ideias fundamentais, a saber, população e amostra, parâmetro e estimador, distribuições empíricas de dados.

### EM13MAT102

Analizar tabelas, gráficos e amostras de pesquisas estatísticas em relatórios divulgados por diferentes meios de comunicação, identificando, quando for o caso, inadequações que possam induzir a erros de interpretação, como escalas e amostras não apropriadas.

### EM13MAT106

Identificar situações da vida cotidiana nas quais seja necessário fazer escolhas levando-se em conta os riscos probabilísticos (usar este ou aquele método contraceptivo, optar por um tratamento médico em detrimento de outro etc.)

### EM13MAT406

Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.

A duração prevista para trabalhar esse primeiro bloco é de três a quatro tempos de aula de 50 minutos.

Este bloco foi dividido em três seções: explorando, organizando e praticando que serão brevemente descritas a seguir.

#### a) Explorando:

Essa seção começa com um texto introdutório para o alunos. A caixa para refletir, logo após o texto inicial, apresenta uma importante reflexão sobre a diferença entre proposições matemáticas e proposições estatísticas. As proposições matemáticas são classificadas em verdadeiras ou falsas, mas isso não é possível para as proposições estatísticas cuja validade envolve uma escala de incerteza tal como é muito provável, é pouco provável etc. Em seguida são sugeridas quatro atividades.

Na primeira atividade será trabalhada a noção de distribuição empírica, conceito chave para a construção de modelos de probabilidade. Em Estatística e Probabilidade, distribuição é uma coleção de propriedades de um conjunto de dados como um todo, não de um particular valor do conjunto. Uma distribuição consiste de todos os valores diferentes nos dados incluindo as frequências (ou probabilidades) associadas com cada valor. Variação e distribuição estão relacionadas a outras noções estatísticas fundamentais tais como "centro" ou "posição" (modeladas pela média, mediana, ou moda), dispersão (modeladas pelo desvio-padrão, ou variância, etc) e forma (por exemplo, bi-modal, uniforme, simétrica, assimétrica à direita, etc). Medidas de "centro" ou "posição" resumem a informação sobre uma distribuição, enquanto medidas de dispersão resumem a variabilidade no conjunto de dados. Cada valor de uma variável mostra algum desvio do "centro". Tais medidas serão trabalhadas no capítulo que dá sequência ao Capítulo "A Natureza da Estatística" ("Medidas de Posição e Dis-

persão"), mas elas já ocorrem nas atividades propostas nesse capítulo, pois média, mediana e moda são trabalhadas no Ensino Fundamental. Por essa razão optamos por começar com a atividade da escolha do fornecedor de parafusos, que ilustra diferentes distribuições e suas características. Essa atividade foi inspirada em Rossman e Chance (1998) e, embora não seja a recomendação, optamos por usar dados fictícios.

A segunda atividade envolve a comparação de medicamentos, usando dados empíricos e foi proposta para ilustrar situações que observar médias amostrais numericamente diferentes pode não levar a conclusão de que as médias na população sejam de fato diferentes, há que se olhar também a dispersão dos dados desses dois conjuntos. Nessa atividade há duas distribuições empíricas (resultantes de amostras da população) com médias numericamente diferentes. No entanto, a dispersão em torno dessas médias é grande de modo que comparando os dois diagramas de pontos, percebe-se grande interseção desses gráficos nas partes de maior frequência de observações, tornando-se razoável supor que as diferenças não são importantes ou, com se diz em estatística, estatisticamente significativas. Novamente, para facilitar a compreensão e ilustrar essa condição, foram considerados dados fictícios.

As terceira e quarta atividade envolvem o suplemento especial da Pesquisa Nacional por Amostra de Domicílios (PNAD) do IBGE realizada em 2015 que incluiu uma investigação sobre a prática de atividades físicas no último ano pela população brasileira de 15 anos ou mais. Esse contexto foi desdobrado em duas atividades em que a primeira, mais conceitual, envolve apenas classificar variáveis em numéricas ou não e, a segunda, envolve analisar infográficos.

#### b) Organizando:

Nessa seção introduzimos os conceitos de população e amostra, parâmetro e estimador, estatística descritiva e estatística inferencial, ilustrando-os por meio de exemplos e também usando como gancho as atividades da seção explorando. Em seguida, apresentamos a classificação de variáveis em quantitativas e qualitativas. Essa seção é encerrada com uma discussão sobre gráficos adequados para representar variáveis qualitativas (categóricas) cuja única escala é a da frequência. Define-se então frequência absoluta e frequência relativa e trabalha-se com gráficos de distribuições de frequências para variáveis qualitativas, em especial o gráfico de barras e o gráfico de setores. Um exemplo foi incluído nessa seção para discutir que tipo de frequência considerar: relativa ou absoluta.

#### c) Praticando:

Nessa seção são apresentadas quatro atividades para complementar a discussão sobre o bloco A Natureza da Estatística. Na primeira atividade solicita-se a condução de uma investigação entre os alunos da turma sobre a prática de atividade física, para depois comparar, por exemplo, o percentual da turma com o percentual correspondente a faixa etária da turma. É claro que não se espera que sejam similares, pois trata-se de uma amostra de conveniência (a turma) muito particular e de tamanho bastante resuzido, comparando-se ao número de jovens brasileiros nessa faixa etária. Nesse caso será um mero exercício de comparar o percentual da turma que pratica atividade física com o percentual indicado pela pesquisa realizada em 2015.

A segunda atividade envolve apenas classificar diferentes tipos de variáveis em quantitativa ou qualitativa.

A terceira atividade solicita ao estudante refazer o infográfico 4, apresentado na seção explorando, transformando o gráfico de setores em gráfico de retângulos e os dois gráficos de retângulos em gráfico de setores.

A quarta atividade ilustra um gráfico de barras que não representa uma distribuição de frequências, pois nesse gráfico o comprimento da barra corresponde à evolução de uma medida de inflação acumulada nos últimos 12 meses, considerando o período que vai de agosto de 2016 até agosto de 2017. Nessa atividade pedimos, indicando uma sugestão, representar esses dados, usando-se um gráfico de linha, comum para variáveis quantitativas observadas ao longo do tempo.

Os exercícios de número 1 a 10, 13, 15, 16 e 18, no final desse capítulo, referem-se ao conteúdo desse bloco.

## EXPLORANDO A NATUREZA DA ESTATÍSTICA

Vivemos cercados de incertezas. A todo momento somos bombardeados por informações sobre pesquisas científicas comprovando (estatisticamente) que tal substância causa uma patologia, ou sobre pesquisas de opinião, índices de pobreza, características sobre o envelhecimento da população e outros temas de natureza incerta. Num mundo assim, é importante ter espírito crítico para informações sujeitas à incerteza a fim de poder interpretá-las e, quando necessário, poder escolher, entre diferentes opções, aquela que parece melhor diante da incerteza. Nesse sentido, a Estatística é uma disciplina fundamental para todos os estudantes e, certamente, com grande responsabilidade para a formação crítica do cidadão, pois ela é usada nas mais variadas áreas do conhecimento tais como: Medicina, Economia, Política, Direito, Psicologia, Engenharia, Educação, entre outras.

Mas afinal o que é Estatística?

### Estatística

Arte e ciência de coletar, analisar, apresentar e interpretar dados, para que se tomem decisões sob incerteza.

### PARA REFLETIR

Proposições são elementos importantes na construção de toda a ciência. No que se refere à natureza da Estatística, em contraponto à natureza da Matemática, podemos destacar dois tipos de proposições.

Uma proposição é dita matemática se é possível classificá-la em *verdadeira ou falsa*, ainda que essa afirmação seja uma conjectura não provada. Assim, a proposição

*"O quadrado de um número par é par".*

é uma proposição matemática, pois sabemos que ela é verdadeira. Da mesma forma, a proposição

*"O triângulo de lados 6, 4 e 3 é um triângulo retângulo."*

é uma proposição matemática, pois sabemos que é falsa.

Por outro lado, uma proposição estatística é uma afirmação sobre a qual nunca teremos condição de afirmar se é verdadeira ou falsa, mas apenas aferir um nível de confiança para ela. A proposição

*"Uma moeda, que ao ser lançada 10 vezes, resulta em 10 coroas, não é uma moeda equilibrada."*

é uma proposição estatística, pois existe a possibilidade de em 10 lançamentos de uma moeda equilibrada obtermos 10 coroas, embora isso seja pouco provável de ocorrer.

**Observação:** Uma moeda é dita ser equilibrada se as probabilidades de se obter cara e coroa são iguais. Caso contrário, a moeda é dita ser não-equilibrada.

Se lançarmos 100 vezes essa mesma moeda e obtivermos 8 caras, teremos mais evidências para aceitar a proposição de que não seja equilibrada, mas ainda assim não poderemos afirmar que a proposição seja verdadeira. Proposições desse tipo que envolvem um nível de confiança sobre sua veracidade são proposições de natureza estatística.

## Objetivos Específicos

### Escolha do melhor fornecedor - Tomada de decisão

- Comparar distribuições empíricas de dados, considerando a necessidade de resumir a informação a partir de medidas de posição e de dispersão, tais como moda e amplitude, que auxiliam na descrição das distribuições.

## Sugestões e discussões

### Escolha do melhor fornecedor - Tomada de decisão

Pretende-se trabalhar nessa atividade vários conceitos importantes na Estatística tais como distribuição empírica, medidas de posição, medidas de dispersão, forma da distribuição, sem se preocupar com formalizações.

No item a) a resposta esperada é "diâmetros dos parafusos". No entanto os alunos podem achar que a frequência com que cada valor de diâmetro ocorre também é necessária. Esse tipo de gráfico, diagrama de pontos, reflete exatamente a tabela de frequências absolutas. No entanto, ele permite perceber por simples visualização a forma da distribuição e suas propriedades.

No item b), como todos os parafusos estão fora da especificação, a resposta é zero.

Item c): Fornecedor A: 14,5 mm; fornecedor B: 15,0 mm; fornecedor C: 15,0 mm e fornecedor D: 14,74 mm.

Para o item d) é necessário perceber que os intervalos assinalados no eixo horizontal correspondentes a 0,1 mm estão subdivididos em 5 partes de medida 0,02 mm. Portanto, a resposta a esse item é

Fornecedor	Valor Mínimo	Valor Máximo
A	14,42	14,58
B	14,60	15,24
C	14,58	15,60
D	14,56	14,12

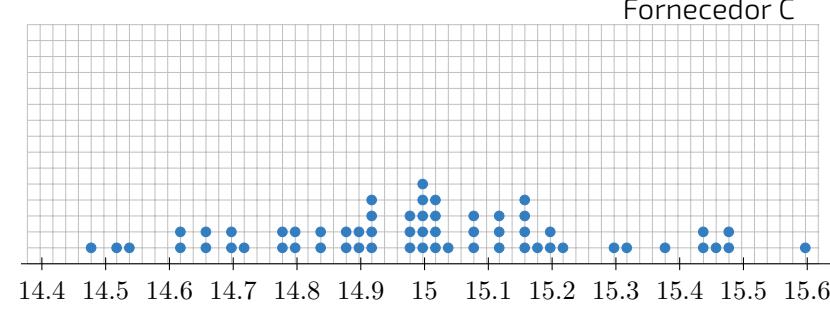
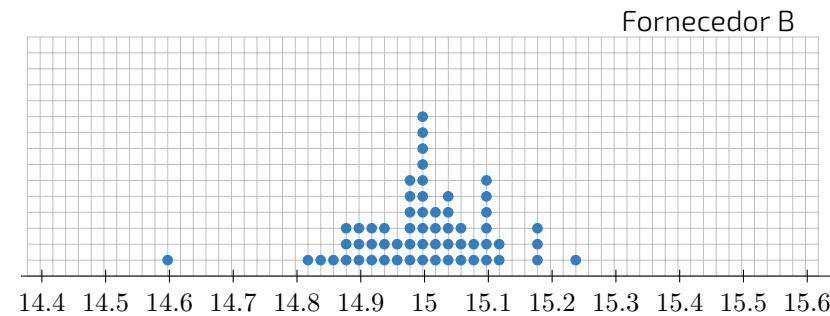
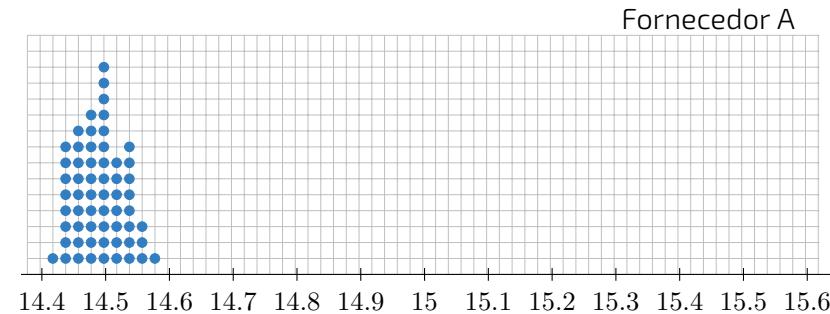
A reflexão tem o intuito de provocar um debate sobre estratégias de amostragem e representatividade das amostras, mesmo sem formalizar tais conceitos. No último item, observe que não é para resolver o problema proposto e sim, pensar em situações semelhantes que levariam a uma análise similar à análise feita nessa atividade, como por exemplo, estudar a vida de baterias de diferentes marcas, ou de uma mesma marca, porém fabricada em países diferentes etc.

## Escolha do melhor fornecedor - Tomada de decisão

## Atividade 1

*Controle de Qualidade na Produção de Parafusos (Inspirada em Rossman e Chance (1998).)*  
Rossman e Chance 1998

Uma indústria precisa comprar parafusos de diâmetro 15 mm cuja variação aceitável é 15,0 mm "mais ou menos" 0,2 mm. Há quatro empresas, A, B, C e D, fornecedoras desses parafusos, que são vendidos em caixas com 60 unidades. Para decidir de qual fornecedor passará a comprar os parafusos, a empresa resolveu comprar e analisar uma caixa de cada um dos fornecedores. Os diâmetros das peças foram medidos com instrumento de alta precisão e os valores obtidos estão representados nos gráficos a seguir, em que cada círculo representa um parafuso posicionado sobre a abscissa correspondente à medida do seu diâmetro, medido em precisão de 0,02 mm.



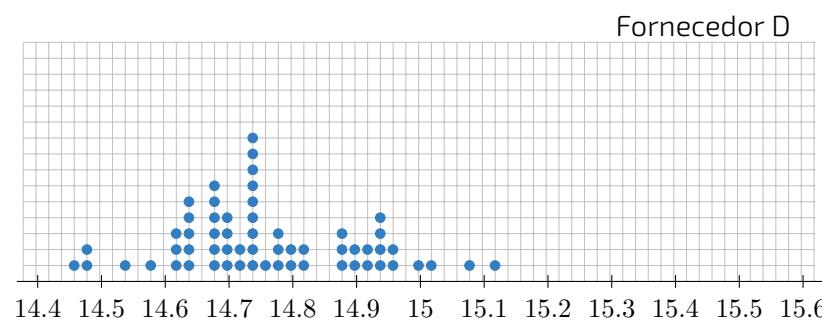


Figura 1.1: Diâmetro das peças dos fornecedores (Fonte: dados fictícios)

- Qual é a observação usada na construção desses gráficos?
- Quantos parafusos da caixa do fornecedor A atendem a especificação do comprador?
- Para cada fornecedor, identifique a medida do diâmetro de maior frequência.
- Considerando cada um dos fornecedores, identifique o menor e o maior diâmetros observados.

**Amplitude** Em Estatística, a amplitude é definida como a diferença entre o maior e o menor valores observados.

- Com base na sua resposta anterior, identifique os fornecedores cujos diâmetros dos parafusos observados variaram nos intervalos de menor amplitude e de maior amplitude.

**Dispersão** Segundo o dicionário Aurélio, dispersão significa (1) ato ou efeito de dispersar; (2) separação (de pessoas ou coisas) para diferentes partes. Em Estatística, existem diferentes medidas de dispersão, dentre as quais, a amplitude.

- De qual fornecedor você classifica o comportamento dos diâmetros dos parafusos como o de maior dispersão? E o de menor dispersão?
- Com base nesses dados, a(s) caixa(s) de qual(is) fornecedor(es) apresenta(m) pelo menos um parafuso dentro das especificações do comprador?
- Supondo que, para cada fornecedor, os comportamentos dos diâmetros dos parafusos sejam similares para as outras caixas, que fornecedor, com base nas especificações do comprador, você recomendaria ao comprador? Por quê?
- Todos os parafusos da caixa do fornecedor escolhido no item anterior seriam aproveitados?

### Solução: Escolha do melhor fornecedor - Tomada de decisão

- Apenas as medidas dos diâmetros dos parafusos.
- Nenhum, pois todos apresentam diâmetro inferior ao mínimo aceitável 14,8 mm.
- Fornecedor A: 14,5 mm; fornecedor B: 15,0 mm; fornecedor C: 15,0 mm e fornecedor D: 14,74 mm.

d)	Fornecedor	Valor Mínimo	Valor Máximo
	A	14,42	14,58
	B	14,60	15,24
	C	14,58	15,60
	D	14,56	14,74

- Menor amplitude: fornecedor A e maior amplitude: fornecedor C
- Em relação à amplitude, menor dispersão: fornecedor A e maior dispersão: fornecedor C.
- Fornecedores B, C e D.
- Fornecedor B, pois é o que tem maior número de parafusos dentro das especificações.
- Não, dois seriam descartados.

## Objetivos Específicos

### Comparação de medicamentos

- Construir diagrama de pontos
- Analisar distribuições empíricas, ou seja, construídas a partir de dados experimentais, usando diagrama de pontos para comparar médias; mais especificamente, para comparar médias populacionais, verificando que nem sempre é possível concluir que estas são iguais quando as médias amostrais são diferentes.

## Sugestões e discussões

### Comparação de medicamentos

O objetivo principal dessa atividade é mostrar situações distintas nas quais ao comparar duas médias diferentes (resultantes de amostras), não é possível afirmar que na população, os parâmetros correspondentes sejam diferentes. Por exemplo, situações nas quais apesar de as médias amostrais serem diferentes, não podemos rejeitar a hipótese de que as médias populacionais são iguais, devido à dispersão resultante da amostra.

As respostas possíveis a serem relatadas no campo para pesquisar devem estar contidas nos campos sobre observações referentes a reações adversas, interações medicamentosas etc. Em geral, as bulas sempre relatam situações que envolvem a observação de dados nesses casos e, algumas, apresentam a frequência na qual essas interações ou reações ocorrem. No entanto, pode ocorrer que uma particular bula não contenha informações do tipo solicitado.

## Nota 1

## PARA REFLETIR

- Comente a estratégia usada para a obtenção dos dados dos fornecedores: as medidas obtidas refletem o comportamento das medidas de todos os parafusos produzidos pelo fornecedor? Seria razoável medir todos os parafusos fabricados por um fornecedor?
- Que procedimento você usaria para confirmar a sua escolha inicial?
- Em Controle de Qualidade, área de aplicação da Estatística na Indústria, é muito comum realizar comparações de diferentes produtos para fazer uma escolha ou verificar se os mesmos atendem às especificações apresentadas. Proponha um problema desse tipo com algum produto e indique a estratégia a ser usada e que medidas deveriam ser observadas.

### Comparação de medicamentos

Atividade 2

Deseja-se comparar três medicamentos, X, Y e Z, no tratamento da dor de cabeça. Para isso 60 pacientes com perfis similares foram separados aleatoriamente em três grupos de 20 cada. Para cada grupo, será ministrado um dos medicamentos e observado o tempo de cura da dor de cabeça (em minutos). No quadro a seguir estão dispostos os dados obtidos (Fonte: dados fictícios).

Medicamentos	tempo em minutos														Soma						
	X	7	8	8	9	9	9	9	10	10	10	10	10	11	11	11	11	12	12	13	
Y	7	8	9	9	10	10	11	11	11	12	12	12	13	13	14	14	15	15	16	18	240
Z	11	11	11	11	11	12	12	12	12	12	12	12	12	12	12	13	13	13	13	13	240

- Organize as informações apresentadas no quadro acima em diagramas de pontos. Utilize uma folha de papel quadriculado, usando a mesma escala.
- A partir dos diagramas construídos, identifique o grupo que apresentou maior dispersão dos tempos de cura com base na amplitude.
- Determine os tempos médios de cura da dor de cabeça para cada substância.
- A partir dos diagramas construídos e das médias calculadas, responda:
  - Entre X e Y, qual medicamento você escolheria? Por quê?
  - Entre X e Z, qual medicamento você escolheria? Por quê?
  - Entre Y e Z, qual medicamento você escolheria? Por quê?
  - A partir dos dados disponíveis, é possível garantir que algum medicamento é melhor que os outros? Por quê?

## PARA REFLETIR

Em casa, procure algum remédio e leia a sua bula. Em seguida, identifique informações que você considera como resultantes de estudos que envolvam Estatística e anote-as em seu caderno.

### Pesquisa sobre a prática de esportes e atividade física

#### Atividade 3

Fonte: IBGE, Suplemento da PNAD/2015

A Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada pelo [IBGE](#), obtém informações anuais sobre características demográficas e socioeconômicas da população, como sexo, idade, educação, trabalho e rendimento, e características dos domicílios. Com periodicidade variável, a PNAD obtém informações sobre migração, fecundidade, entre outras, tendo os domicílios como unidade de coleta da informação. Temas específicos abrangendo aspectos demográficos, sociais e econômicos também são investigados.

Um aspecto fundamental da Estatística praticado nessa pesquisa é a forma na qual a amostra, subconjunto da população, é selecionada. Essa seleção é cuidadosamente planejada de modo que seja adequado estender os resultados obtidos na amostra para a população.

Para que os resultados de uma amostra possam ser estendidos para a população, é necessário planejar com cuidado como a amostra será selecionada, pois o critério de seleção da amostra depende da estrutura da população. Por exemplo, para saber se o feijão cozinhando na panela está bem temperado, basta provar uma pequena colherada. Por quê? Partimos do pressuposto de que todos os ingredientes foram bem misturados e, assim, a mistura é homogênea.

Quando dispomos de dados provenientes de um subconjunto da população sempre podemos descrever os dados nos restringindo apenas ao subconjunto. Se quisermos estender nossas conclusões para a população, será necessário o uso de outras tecnologias que permitam calcular as incertezas associadas a essas extensões.

Na PNAD 2015 foi realizada a investigação de um tema específico chamado "Suplemento de Práticas de Esporte e Atividade Física" no qual foram investigadas as pessoas moradoras de 15 anos ou mais de idade, **em seu tempo livre**, no período de referência de 365 dias, com o objetivo de quantificar aquelas que praticaram algum esporte ou atividade física no período considerado bem como a sua percepção quanto a isso. As informações levantadas nessa pesquisa foram obtidas por meio de um questionário no qual se perguntou:

- Se a pessoa moradora havia praticado esporte, e em caso afirmativo, a respectiva modalidade.
- Independente da resposta anterior, também se perguntou se a pessoa praticava alguma atividade física que não considerava como esporte, informando, em caso positivo, também a modalidade.

## Objetivos Específicos

### Pesquisa sobre a prática de esportes e atividade física

- Apresentar os conceitos de população e amostra.
- Comparar os diferentes tipos de variáveis analisados em uma pesquisa para adiante identificar variáveis qualitativas e quantitativas.

## Sugestões e discussões

### Pesquisa sobre a prática de esportes e atividade física

- No item **a**), espera-se que sejam indicadas algumas entre as seguintes variáveis: idade, sexo, educação, trabalho, rendimento, se pratica ou não atividade física, modalidade da atividade para quem pratica, motivação para a prática de atividade física, local da prática, frequência da prática, duração da atividade, participação em competições etc.
- No item **b**) deve-se informar as variáveis que assumem atributos (respostas não-numéricas) tais como sexo, prática de atividade física (sim ou não), modalidade da atividade física praticada etc.
- No item **c**) deve-se informar as variáveis que assumem valores numéricos tais como idade, rendimento, duração da atividade física etc.

### Solução: Pesquisa sobre a prática de esportes e atividade física

- a) Sexo. Idade. Educação. Trabalho. Rendimento. Prática de Atividade Física(AF). Modalidade da AF para quem pratica. Motivação para a AF. Local da Prática da AF. Duração da Prática da AF etc.
- b) Sexo. Educação. Trabalho. Prática de AF. Modalidade de AF. Motivação da Prática de AF. Local da Prática da AF.
- c) Idade. Rendimento. Duração da Prática de AF.

### Objetivos Específicos

#### Análise de Infográficos

- Analisar infográficos. Mais especificamente, analisar infográficos construídos pelo IBGE com os resultados da pesquisa PNAD/2015 referente ao suplemento especial de Prática de Atividades Físicas.
- Explorar possíveis associações sobre a prática de atividades físicas com outras variáveis envolvidas na pesquisa, tais como sexo, nível de instrução e rendimento.

- Outras informações levantadas nessa pesquisa foram: motivação para a prática da atividade física, local onde é praticada a atividade, frequência na qual a atividade é praticada, duração da atividade; e a participação em competições.
  - Também foram levantadas informações sobre as pessoas que responderam que não praticavam atividade física. Perguntou-se o motivo de não o fazerem e se haviam praticado anteriormente, caso em que se perguntou a modalidade praticada, a idade em que parou de praticar e a causa da interrupção.
  - Além dessas informações, a pesquisa investigou também a avaliação da população sobre a opção de o poder público investir no desenvolvimento de atividades físicas e esportivas ou em outra área (saúde, educação etc.) na vizinhança de seu domicílio.
- a) Liste pelo menos oito variáveis investigadas na PNAD e no "Suplemento de Práticas de Esporte e Atividade Física" da PNAD 2015, baseando-se no texto apresentado.
  - b) Das variáveis citadas no item anterior, quais delas apresentam respostas não numéricas?
  - c) Das variáveis citadas no item a), quais delas apresentam respostas numéricas?

Cada uma das unidades investigadas em um estudo estatístico é denominada um elemento. Assim, cada parafuso investigado é um elemento na atividade "Escolha do fornecedor"; cada paciente observado é um elemento na atividade "Comparação de medicamentos"; e cada domicílio e seus residentes são elementos na atividade da PNAD.

Cada característica observada de um elemento é uma variável estatística. Assim, a medida do diâmetro do parafuso é uma variável na atividade "Escolha do fornecedor", o tempo de cura da dor de cabeça é uma variável na atividade "Comparação de medicamentos" e, na atividade da PNAD, estão presentes várias variáveis estatísticas de interesse do domicílio e de seus residentes tais como local, número de cômodos, número de residentes; sexo, idade e rendimento dos residentes etc.

#### Análise de infográficos

Atividade 4

A seguir apresentaremos quatro infográficos (figuras 1.2, 1.3, 1.4, 1.5), produzidos pelo IBGE ([vamoscontar.ibge.gov.br](http://vamoscontar.ibge.gov.br)) usando os dados do Suplemento Prática de Esporte e Atividade Física da PNAD 2015.

Um infográfico é uma apresentação de informações integradas em textos sintéticos com dados numéricos e elementos gráficos e visuais tais como fotografias, desenhos, diagramas estatísticos, gráficos etc.

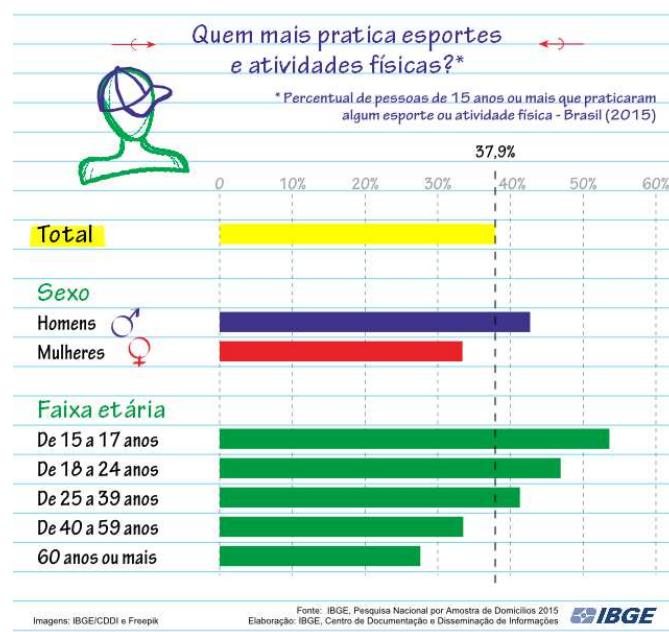


Figura 1.2: PNAD - Infográfico 1

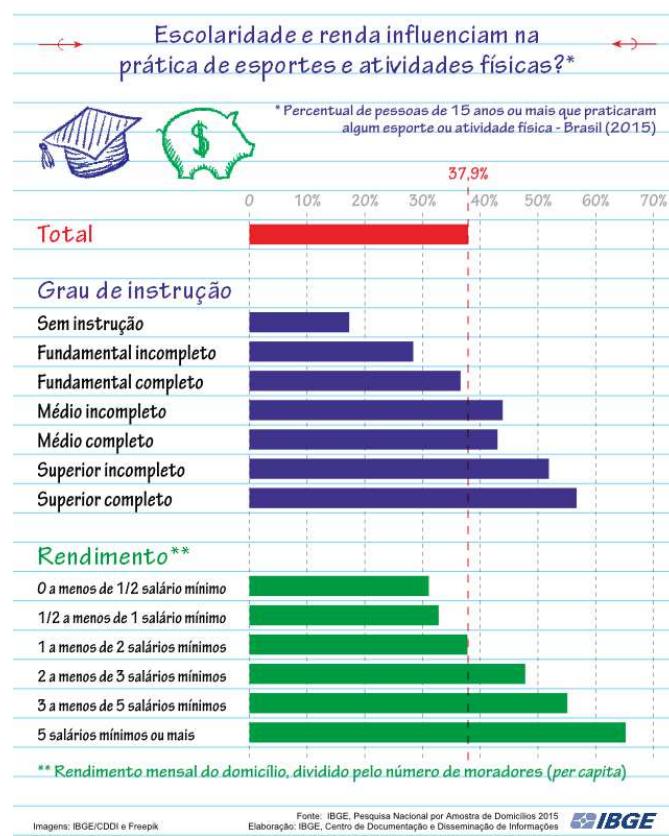


Figura 1.3: PNAD - Infográfico 2

## Sugestões e discussões

## Infográfico 1

O item **b)** pretende estimular a reflexão sobre o papel da inferência estatística. De fato, foi observada uma amostra de domicílios de algumas cidades brasileiras, mas como a amostra foi cuidadosamente planejada e a estrutura da população brasileira é conhecida, foi possível dar um passo maior e calcular uma estimativa da proporção das pessoas de 15 anos ou mais que praticam atividades físicas no Brasil. A porcentagem 37,9%, realização numérica de um estimador, representa uma estimativa da proporção das pessoas de 15 anos ou mais que praticaram atividades físicas no Brasil (2015) (parâmetro). Observe que não foi realizado um censo para obter essa informação. Portanto, associada a essa estimativa existe uma margem de erro (valor correspondente à oscilação em torno da estimativa pontual) e um nível de confiança. Por exemplo, se o nível de confiança for 95% isso implica que para cada 100 amostras de mesmo tamanho, em 95% delas o parâmetro se situa no intervalo considerando a margem de erro. Claro que a margem de erro deve ser pequena e o nível de confiança alto na PNAD. Esses conceitos, margem de erro e nível de confiança, têm sido bem divulgados nas pesquisas eleitorais para o público em geral. Se for um ano de eleição, peça aos alunos para trazer resultados de pesquisas eleitorais incluindo a margem de erro e o nível de confiança. Cabe também destacar que todas as proporções apresentadas na pesquisa são estimativas que devem ter pequena margem de erro com nível de confiança alto. Assim, pequenas diferenças nessas proporções devem ser olhadas com cuidado, não sendo possível afirmar que elas são diferentes.

O item **c)** visa levar a uma reflexão sobre hábitos saudáveis. Por que achamos que a prática de atividades físicas é importante para a saúde de uma pessoa? Como essa conclusão foi obtida?

Os itens **d)** e **e)** têm como objetivo estudar possíveis associações entre duas variáveis qualitativas, a saber, sexo e prática de atividade física **d)** e faixa etária e prática de atividade física **e)**. Observe que embora a idade seja uma variável quantitativa, quando ela é representada por faixas etárias ela se torna qualitativa.

É importante destacar, na análise desses gráficos, que o que se fez foi separar o conjunto de dados em subconjuntos como por exemplo, sexo feminino e sexo masculino e depois, observou-se a resposta sobre a prática de atividade física em cada subgrupo. Para efeito de comparação de grupos distintos, é importante trabalhar com a frequência relativa (ou porcentagem), pois os grupos podem ser de tamanhos diferentes e se os gráficos forem construídos com as frequências absolutas não será possível visualizar as relações entre as variáveis analisadas.

## Nota 2

## Sugestões e discussões

### Infográfico 3

Na análise do infográfico 3, cabe destacar que trata-se de um gráfico de barras típico representando a distribuição de frequências de uma variável qualitativa. É importante levar os alunos a perceber que para a variável modalidade, considerando o conjunto de todas as pessoas que responderam essa questão, calculou-se as porcentagens para cada tipo de atividade indicada. Discuta sobre a categoria outras atividades indicando que foram respostas com frequência muito pequena e, de fato, não faria sentido ir listando uma a uma essas modalidades. Em geral, nesses casos, o que se faz é agregar as respostas com frequência muito pequena na categoria outras. Sugira ao aluno pesquisar no link dessa pesquisa para verificar se, no instrumento de coleta de dados, essa questão era aberta (resposta livre) ou fechada (com opções a serem assinaladas).

Na análise desse gráfico, deve-se destacar que a altura das barras correspondem às porcentagens (frequências relativas) na qual ocorreram e que a soma dessas porcentagens será 100%. Também cabe comentar que as barras devem ter larguras iguais, mas não existe nenhum lugar geométrico definido ao longo do eixo horizontal para as respostas da variável modalidade de prática neste gráfico, ou seja, podemos mudar a posição das diferentes modalidades. As barras, separadas, são equidistantes e foram organizadas por ordem de decrescente de frequência. Como só há um eixo numérico (frequência), comente que as barras podem ser tanto verticais, como horizontais e essa orientação determinará a orientação do eixo que representa as frequências no gráfico.

## Sugestões e discussões

### Infográfico 4

Na análise do infográfico 4, é importante destacar que foram usados dois tipos de gráficos diferentes para representar variáveis qualitativas, mas ambos usam a mesma ideia, a saber, uma região é subdividida de maneira harmônica em sub-regiões (o círculo em setores circulares e o retângulo em retângulos menores de mesma largura contidos nele) cujas áreas em relação à área da região correspondem exatamente à frequência relativa (ou porcentagem) da categoria de resposta que a sub-região representa. Por exemplo, a área do setor em vermelho dividida pela área do círculo é 0,147 (ou 14,7% da área do círculo). A área do retângulo verde dividida pela área do retângulo inteiro é 0,578 (ou 57,8% da área do retângulo inteiro). São duas formas de olhar como cada categoria de resposta aparece em relação ao todo.

## Quais os principais esportes praticados?\*

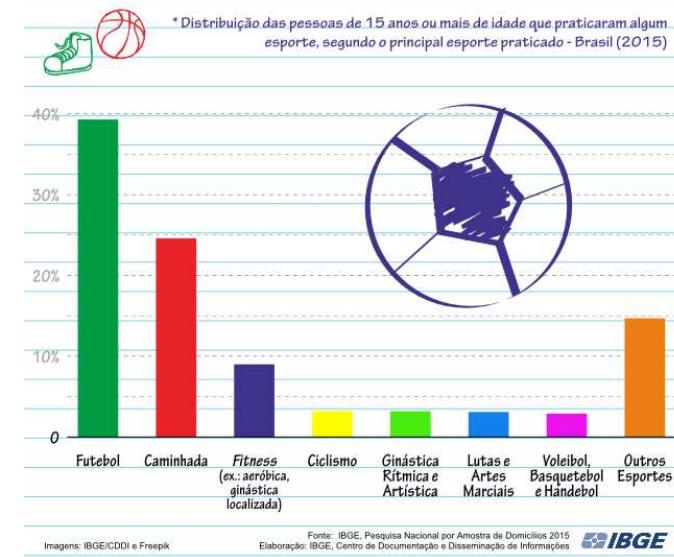


Figura 1.4: PNAD - Infográfico 3

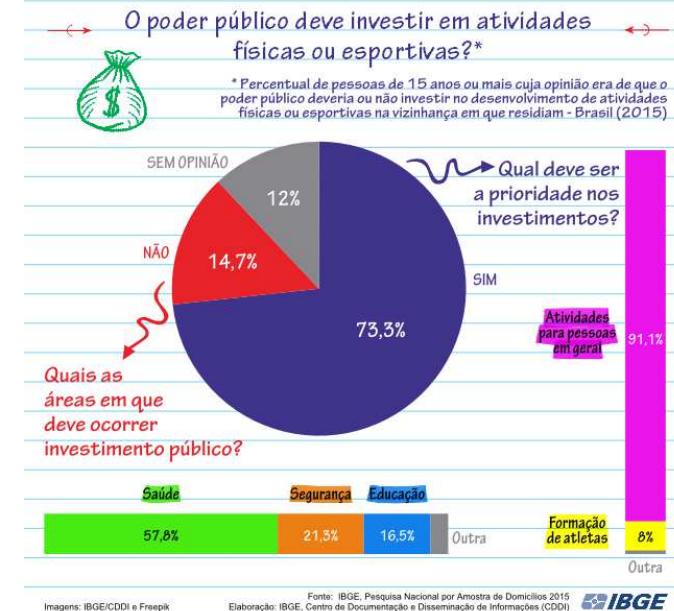


Figura 1.5: PNAD - Infográfico 4

Fonte: Pesquisa Nacional por Amostra de Domicílios 2015. Centro de Documentação e Disseminação de Informações (CDDI). IBGE

a) Segundo os resultados da pesquisa (veja a figura 1.2), qual a porcentagem de pessoas

de 15 anos ou mais que praticaram algum esporte ou atividade física no período de um ano?

- b)** O título genérico do infográfico da [figura 1.2](#), a saber, "Quem mais pratica esportes e atividades físicas? - Percentual de pessoas de 15 anos ou mais que praticaram algum esporte ou atividade física-Brasil (2015)", diz respeito à população brasileira de 15 anos ou mais ou à amostra coletada?
- c)** Com base nas recomendações médicas sobre a prática de atividades físicas para se ter boa saúde, como você avalia o resultado obtido na pesquisa para a população brasileira de 15 anos ou mais?
- d)** Considerando homens e mulheres separadamente, percebe-se alguma diferença com relação à prática de atividades físicas? Em caso afirmativo, descreva a(s) diferença(s) observada(s).
- e)** Considerando as faixas etárias discriminadas no infográfico, percebe-se alguma diferença com relação à prática de atividades físicas? Em caso afirmativo, descreva a(s) diferença(s) observada(s).
- f)** Considerando os diferentes graus de instrução ([figura 1.3](#)), percebe-se alguma diferença com relação à prática de atividades físicas? Em caso afirmativo, descreva a(s) diferença(s) observada(s).
- g)** Considerando as faixas de rendimento mensal per capita do domicílio, percebe-se alguma diferença com relação à prática de atividades físicas? Em caso afirmativo, descreva a(s) diferença(s) observada(s).
- h)** Qual foi a variável estudada no gráfico da [figura 1.4](#)?
- i)** A variável estudada no gráfico da [figura 1.4](#) tem respostas de que tipo: numéricas ou não-numéricas?
- j)** Qual foi a resposta que apresentou a maior frequência no gráfico da [figura 1.4](#)?
- k)** O que você acha que representa a resposta "Outros Esportes" no gráfico da [figura 1.4](#)?
- l)** Qual a porcentagem de pessoas de 15 anos ou mais que concorda com que o poder público deva investir em atividades físicas ou desportivas ([figura 1.5](#))?
- m)** Qual a opinião das pessoas de 15 anos ou mais que concordam que o poder público deve investir em atividades físicas ou esportivas com relação à prioridade de investimentos?
- n)** Entre as pessoas de 15 anos ou mais que não concordam que o poder público deve investir em atividades físicas ou esportivas, que área elas entendem como prioritária?
- o)** Podemos afirmar que 57,8% das pessoas de 15 anos ou mais defendem que o poder público deve investir em Saúde?"

### Solução: Análise de Infográficos

- a)** 37,9%
- b)** População brasileira de 15 anos ou mais.
- c)** Não parece satisfatório. Vários estudos têm demonstrado que a prática de atividades físicas é fundamental para se ter boa saúde.
- d)** Sim. Entre os homens brasileiros de 15 anos ou mais, pouco mais de 40 praticam atividade física; enquanto esse percentual para mulheres brasileiras de 15 anos ou mais é pouco menor do que 35%.
- e)** Sim. Percebe-se uma diminuição dos percentuais de pessoas que praticam atividade física, conforme a idade aumenta. Na faixa de 15 a 17 anos temos mais de 50%, na faixa de 18 a 24 anos temos um pouco menos do que 50% na faixa de 25 a 39 anos temos pouco mais de 40%, na faixa de 40 a 59 anos temos mais de 30% e na faixa 60 anos ou mais temos menos de 30%.
- f)** Sim, a porcentagem de pessoas de 15 anos ou mais que pratica atividade física cresce conforme o grau de instrução é maior.
- g)** Sim, a porcentagem de pessoas de 15 anos ou mais que pratica atividade física cresce conforme a faixa de rendimento per capita é maior.
- h)** Modalidade de atividade física praticada.
- i)** Não-numéricas: futebol, natação etc.
- j)** Futebol
- k)** Como as últimas modalidades discriminadas no gráfico apresentaram porcentagens muito pequenas ("ciclismo", "ginástica rítmica e artística", "lutas e artes marciais", "voleibol, basquetebol e handebol"), cerca de 2%, a categoria outros esportes reuniu modalidades que ocorreram com porcentagens muito pequenas, não cabendo representá-las separadamente no gráfico. Observe que a última modalidade, antes de "outros esportes" já está reunida em mais de uma modalidade, a saber, "voleibol, basquetebol e handebol".
- l)** 73,3%
- m)** Entre as pessoas que acham que se deva priorizar investimentos em atividades físicas, 91,1% acha que o investimento deve ser para atividades físicas para as pessoas em geral, 8% acha que deve ser para a formação de atletas e, o restante (0,9%) respondeu outro tipo de prioridade.
- n)** Entre as pessoas que não concordam que o poder público deve investir em atividades físicas, 57,8% acham que a prioridade deve ser Saúde, 21,3% acham que a prioridade deve ser Segurança, 16,5%, acham que a prioridade deve ser Educação e, o restante (4,4%) respondeu outros tipos de prioridade.
- o)** Não, de fato, são 57,8% de 14,7% o que dá cerca de ,5% das pessoas de 15 anos ou mais.

## ORGANIZANDO A NATUREZA DA ESTATÍSTICA

### EXEMPLO 1 Opinião dos estudantes sobre lanches na cantina (1)

Suponha que deseja-se investigar a opinião dos estudantes de um colégio quanto à modificação da lista de produtos vendidos na cantina para outros mais saudáveis, trocando refrigerantes por sucos naturais entre outros. Para isso, a direção da escola irá entrevistar cinco alunos sorteados de cada uma de suas 40 turmas.

### Conceitos Básicos

Em geral, a palavra população representa um conjunto de habitantes de um determinado lugar. No entanto, em Estatística, população tem um sentido mais amplo e pode ser definida como o conjunto de todos os elementos com pelo menos uma característica em comum. Observe que é exatamente essa característica em comum que vai definir o universo (população) de uma pesquisa.

Assim, em Estatística, a população não precisa ser um conjunto de pessoas, pode ser o conjunto de parafusos fabricados por uma indústria, o conjunto de animais de certa espécie que vivem em uma região, todos os estudantes universitários de um país etc.

**Amostra** É um subconjunto não-vazio da população.

Cada uma das unidades investigadas em um estudo estatístico é denominada um elemento. Assim, cada parafuso investigado é um elemento na atividade "Escolha do fornecedor"; cada paciente observado é um elemento na atividade "Comparação de medicamentos"; e cada domicílio e seus residentes são elementos na atividade da PNAD.

Cada característica observada de um elemento é uma variável estatística. Assim, a medida do diâmetro do parafuso é uma variável na atividade "Escolha do fornecedor", o tempo de cura da dor de cabeça é uma variável na atividade "Comparação de medicamentos" e, na atividade da PNAD, estão presentes várias variáveis estatísticas de interesse do domicílio e de seus residentes tais como local, número de cômodos, número de residentes; sexo, idade e rendimento dos residentes etc.

**Parâmetro** Característica numérica da população.

**Estimador** Função que produz estimativas de parâmetros usando os dados da amostra.

### EXEMPLO 2 Opinião dos estudantes sobre lanches na cantina (2)

Considere o exemplo inicial. Nesse caso tem-se:

- população, correspondendo a todos os estudantes do colégio;
- amostra, correspondendo aos 200 estudantes entrevistados (5 de cada turma);
- elemento, correspondendo a cada estudante entrevistado
- variável de interesse, correspondendo à opinião do estudante: "a favor" ou "contra" à mudança;

- parâmetro de interesse, correspondendo à proporção dos estudantes desse colégio que são favoráveis à mudança;
- estimativa do parâmetro de interesse, correspondendo à proporção entre os 200 estudantes entrevistados favoráveis à mudança.

Observação: Em um estudo desse tipo também costumam ser registradas outras variáveis tais como sexo, turno, idade, série etc.

As etapas da análise estatística podem ser divididas em duas estruturas básicas: Estatística Descritiva e Estatística Inferencial. A primeira corresponde a uma exploração das informações que podem ser retiradas dos dados amostrais de modo a reconhecer estruturas que possibilitem futuramente inferir sobre parâmetros de interesse. A segunda consiste em estabelecer modelos probabilísticos para que se possa fazer afirmações sobre a população com algum nível de confiança. Leia a caixa **Para Refletir** do início do capítulo para uma explicação.

Em resumo, a Estatística Descritiva é uma espécie de arqueologia dos dados observados e, a Estatística Inferencial, a indução das informações obtidas da amostra para características da população não observada em sua totalidade.

A PNAD faz uso da inferência estatística, pois ela investiga uma amostra de domicílios em algumas cidades brasileiras, mas propõe estimativas para as características da população brasileira.

Quando se realiza um censo - levantamento de dados de toda a população -, não existe a necessidade de fazer uma inferência estatística. No entanto, muitas vezes a realização de um censo é inviável, por várias razões como custo muito alto, tempo muito longo, entre outras.

## Classificação de Variáveis

A classificação das variáveis estudadas é importante, pois as técnicas e procedimentos estatísticos de análise de dados dependem do tipo de variável investigado. Nesse sentido é importante reconhecer a natureza de cada variável investigada para posterior tratamento da informação obtida. Por exemplo, se estamos estudando a modalidade de atividades físicas praticadas pelos brasileiros de 15 anos ou mais, não faz sentido calcular média, pois ela não assume valores numéricos.

Existem dois tipos principais de variáveis (qualitativas e quantitativas), que se subdividem, por sua vez, em duas categorias, conforme a [figura 1.6](#).

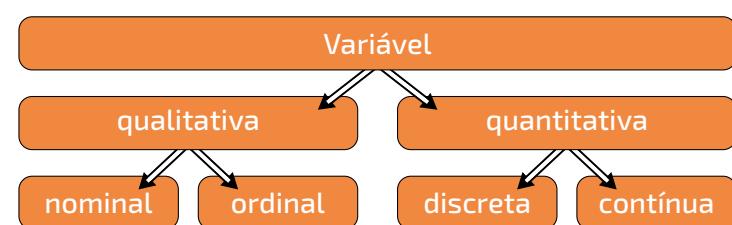


Figura 1.6: Classificação dos tipos de variáveis

**Variável qualitativa** Uma variável estatística é qualitativa se as possíveis respostas para ela

são atributos não-numéricos. A maior parte das variáveis identificadas no "Suplemento de Práticas de Esporte e Atividade Física" da PNAD/2015, representa variáveis qualitativas.

É comum encontrar o termo **variável categórica** para se referir a uma variável qualitativa. As variáveis qualitativas são classificadas como nominal ou ordinal.

**Variável qualitativa nominal** Uma variável qualitativa é nominal quando não existe nenhuma ordenação natural das respostas associadas à variável. Exemplos de variáveis nominais: bairro de residência, tipo sanguíneo, modalidade de atividade física que pratica etc.

**Variável qualitativa ordinal** A variável qualitativa é ordinal quando é possível estabelecer uma relação de ordem entre as respostas associadas a ela. Por exemplo, nível de instrução da mãe com as respostas possíveis: Ensino Fundamental completo, Ensino Médio completo, Ensino Superior incompleto e Ensino Superior completo. Podemos perceber que quem tem Ensino Médio completo tem maior nível de instrução de quem tem Ensino Fundamental completo.

**Variável quantitativa** Uma variável é quantitativa se as respostas para ela são numéricas. Exemplos de variáveis quantitativas são idade, peso, altura, temperatura, número de irmãos, número de horas semanais dedicadas à prática de atividade física.

Uma variável quantitativa é classificada em discreta ou contínua.

**Variável quantitativa discreta** As variáveis discretas resultam de uma contagem ou são variáveis cuja quantidade de valores possíveis é finita. Por exemplo, o número de atendimentos em um Pronto-Socorro nos finais de semana, o número de erros de impressão na página de um livro, número de irmãos etc.

**Variável quantitativa contínua** As variáveis quantitativas contínuas em geral resultam de uma medição. Por exemplo, altura, peso, temperatura etc.

Na análise dos infográficos vimos que uma variável quantitativa pode ser tratada como qualitativa: a idade trabalhada em faixas etárias torna-se uma variável qualitativa ordinal. No entanto, se considerarmos a idade em anos completos temos uma variável quantitativa. Por outro lado, também podemos transformar uma variável qualitativa em quantitativa. Considere a variável "prática de atividades físicas" que tem como respostas "Sim" ou "Não". Esse tipo de variável com apenas duas respostas é chamado variável binária e tem uma representação numérica natural. Podemos atribuir o número 1 para a resposta "Sim" e o número 0 para a resposta "Não". Essa codificação permite somar todas as respostas. A soma representará o número de pessoas na amostra que praticam atividade física e, a "média", representará a proporção de pessoas na amostra que praticam atividade física. Mas, deve-se ter consciência de que quando usamos números como códigos para representar variáveis qualitativas, em geral, não será possível realizar as quatro operações, pois o resultado não terá um significado.

**VOCÊ SABIA?**

A escolha do gráfico para representar distribuições de frequências observadas das variáveis estudadas está associada à classificação da variável. Para variáveis qualitativas, usa-se em geral os gráficos de barras e os gráficos de setores. Para variáveis quantitativas, usa-se em geral histogramas, ramo-e-folhas e boxplots.

## Tabelas de frequências e gráficos para variáveis qualitativas

### EXEMPLO 3 Frequência absoluta e frequência relativa

Numa turma de um colégio foram observados os tipos sanguíneos de seus 40 alunos. Verificou-se que 18 alunos têm sangue tipo "O", 12, tipo "A", 6, tipo "AB" e 4, tipo "B". Nesse exemplo, temos que as frequências absolutas para os tipos sanguíneos "O", "A", "AB" e "B" foram, respectivamente, 18, 12, 6 e 4. Em geral, quando queremos comparar grupos diferentes, usamos a frequência relativa em vez da frequência absoluta. A frequência relativa é dada pela razão entre a frequência absoluta e o número total de observações. Nesse exemplo, temos que as frequências relativas para os tipos sanguíneos "O", "A", "AB" e "B" foram, respectivamente, 0,45; 0,30; 0,15 e 0,10. Observe que em termos percentuais as frequências relativas observadas equivalem a, respectivamente, 45%, 30%, 15% e 10%.

É comum resumir esse tipo de informação, usando uma tabela, informando as respostas da variável e suas frequências. Nesse exemplo a variável é tipo sanguíneo e sua classificação é qualitativa nominal, pois assume respostas não numéricas "A", "B", "AB" e "O", sem uma ordenação natural. Em geral dispomos os valores dessa variável em ordem decrescente de frequência.

Tipo Sanguíneo	Frequência Absoluta	Frequência Relativa	Porcentagem (%)
O	18	0,45	45
A	12	0,30	30
AB	6	0,15	15
B	4	0,10	10
total	0	1,00	100

A **frequência absoluta** de cada resposta da variável qualitativa representa o número de observações dessa resposta. Assim, para a variável tipo sanguíneo e o conjunto de alunos dessa turma, as frequências absolutas dos tipos "O", "A", "AB" e "B" são, respectivamente, 18, 12, 6 e 4. Ou seja, ao todo observaram-se 18 alunos com sangue tipo "O", 12 alunos com sangue tipo "A", 6 alunos com sangue tipo "AB" e 4 alunos com sangue tipo "B".

A **frequência relativa** de cada resposta da variável qualitativa representa o quociente da frequência absoluta da resposta sobre o número total de observações. Observe que a frequência relativa sempre será um número entre 0 e 1 e que a soma das frequências relativas será sempre igual a 1. Veja na tabela desse exemplo.

Uma forma equivalente à frequência relativa é a porcentagem de cada resposta que é dada pelo produto da frequência relativa por 100. Veja na tabela desse exemplo.

Nas análises dos infográficos, trabalhamos com alguns tipos de gráficos para representar a distribuição de frequências de variáveis qualitativas. No [infográfico 3 \(Figura 1.4\)](#), tem-se um

gráfico de barras. Nesse gráfico, cada barra, de mesma largura, representa uma resposta e seu comprimento corresponde à frequência na qual a resposta ocorreu. Observe também que, nesse gráfico, se estivermos trabalhando com as porcentagens de cada resposta, a soma das porcentagens deve ser 100%.

Em geral, se a variável qualitativa for ordinal dispomos as respostas em ordem crescente, considerando as respostas. Se a variável é nominal, podemos dispor as respostas em ordem decrescente de frequência.

Alguns gráficos apresentados na atividade "Análise de infográficos" como o [infográfico 1 \(Figura 1.2\)](#) e [infográfico 2 \(Figura 1.2\)](#) usam barras para representar as frequências em subgrupos diferentes do conjunto observado. Mas eles não se encaixam na apresentação anterior. Verifique que se somarmos as porcentagens elas não resultarão em 100%. De fato, são gráficos de barras múltiplas, úteis para comparar distribuições de frequências em subgrupos distintos. Observe que, em cada um desses gráficos, a variável sob investigação é se a pessoa pratica ou não atividade física. No entanto, em vez de apresentar as porcentagens das respostas *Sim* e *Não* no subgrupo considerado, como a variável é binária, só foram apresentadas as porcentagens de *Sim* em cada subgrupo, pois nesse caso, as correspondentes porcentagens de *Não* são dadas pelo complementar em cada subgrupo considerado. Observe que os subgrupos comparados foram:

- homens e mulheres;
- faixas etárias: 15 a 17; 18 a 24; 25 a 39; 40 a 59 e 60 anos ou mais;
- graus de instrução: sem instrução; fundamental incompleto; fundamental completo; médio incompleto; médio completo; superior incompleto e superior completo;
- rendimento: menos de 1/2; de 1/2 a 1; de 1 a 2; de 2 a 3; de 3 a 5 e 5 salários mínimos ou mais.

Na [figura 1.7](#) foi realçado o caso dos subgrupos homens e mulheres da pesquisa.

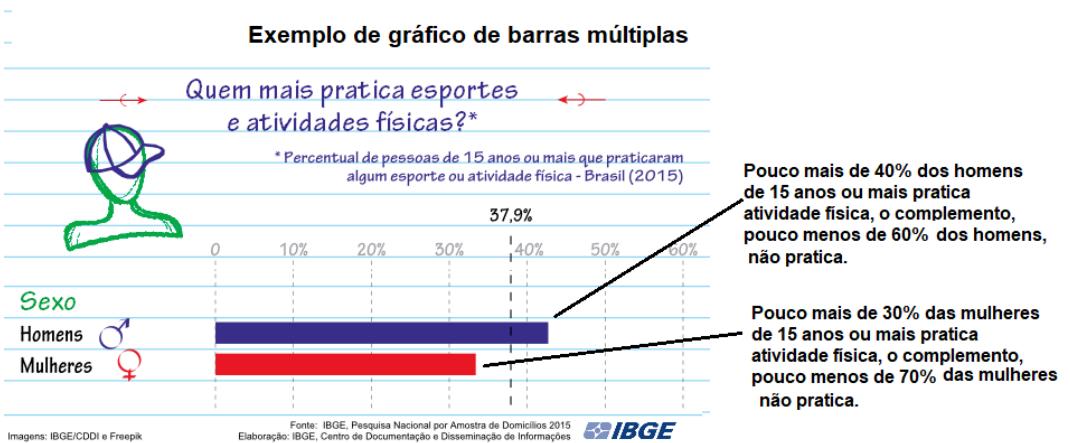


Figura 1.7: Detalhe legendado do [infográfico 2](#)

O mesmo ocorre quando analisamos os gráficos para faixa etária, grau de instrução e rendimento. Todos são gráficos de barras múltiplas que nos apoiaram em nossas análises sobre a

associação entre a prática de atividades físicas e a outra variável (sexo, faixa etária, grau de instrução, rendimento).

No [infográfico 4 \(Figura 1.5\)](#), temos um gráfico de setores e dois gráficos de retângulos. A ideia por trás desses gráficos é subdividir de maneira proporcional a figura maior em partes cujas áreas em relação à figura maior correspondam à frequência de cada resposta.

No gráfico de setores, subdividimos o círculo em setores circulares de tal modo que o quociente da área de cada setor em relação à área do círculo é igual à frequência relativa (ou porcentagem) da resposta que ele representa. Lembrando que o setor circular é definido pelo ângulo central, para traçar os setores em um gráfico desse tipo será necessário calcular as medidas dos ângulos centrais dos setores correspondentes a cada resposta da variável qualitativa.

Por exemplo, se numa distribuição a frequência relativa de uma resposta corresponde a 0,40 (40%) de todo o conjunto observado, no gráfico de setores essa resposta será representada por um setor circular cujo ângulo central é dado por  $0,4 \times 360^\circ = 144^\circ$  como ilustra a [figura 1.8](#).

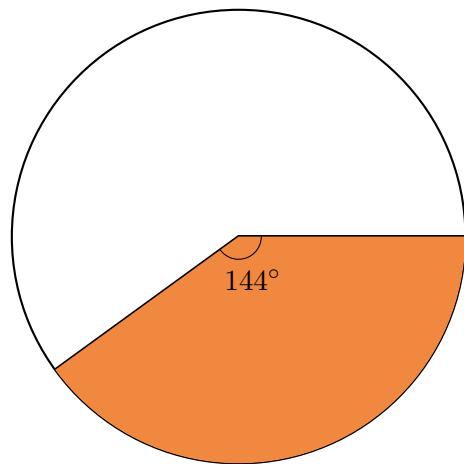


Figura 1.8: Setor circular de 144°

A medida da área da região em azul no círculo da [figura 1.8](#) corresponde à 40% da área do círculo.

Nos gráficos de retângulos essa mesma ideia é usada: o retângulo maior é subdividido em retângulos cujas áreas relativas correspondem às porcentagens das respostas que eles representam. Esses gráficos foram construídos para representar as respostas à pergunta "Em quais áreas em que deve ocorrer investimento público?" para quem respondeu *Não* à pergunta "O poder público deve investir em atividades físicas ou desportivas?" e também para representar as respostas à pergunta "Qual deve ser a prioridade nos investimentos?" para quem respondeu "Sim" à pergunta "O poder público deve investir em atividades físicas ou desportivas?".

#### EXEMPLO 4 Frequência absoluta ou frequência relativa?

Suponha o exemplo com os dados de tipos sanguíneos dos 40 alunos de uma turma. Agora desejamos comparar as respostas obtidas com um conjunto de 120 observações para as quais 30 são tipo A, 12, do tipo AB, 18, do tipo B 60, do tipo O. Os gráficos de barras na mesma escala, usando a frequência absoluta parecem bem diferentes, como mostra a figura a seguir.

Distribuições de frequências de tipo sanguíneo

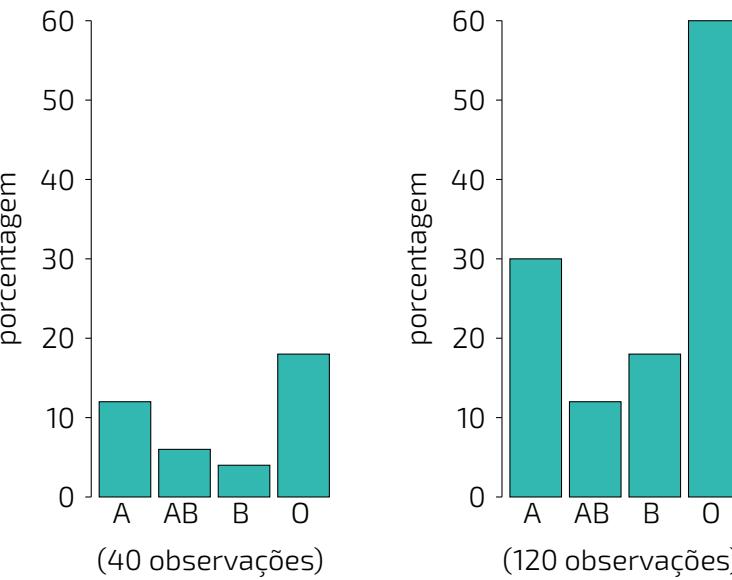
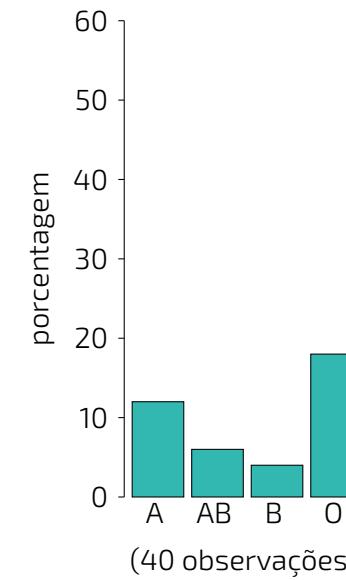


Figura 1.9: Gráficos de barras da distribuição na escala da frequência absoluta

Porém, os gráficos construídos, usando a escala da porcentagem, não parecem tão diferentes, como mostra a figura a seguir.

Distribuições de frequências de tipo sanguíneo

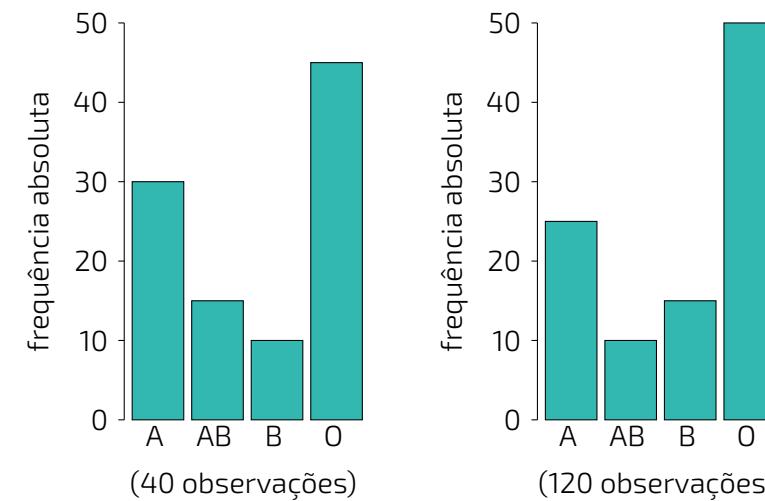
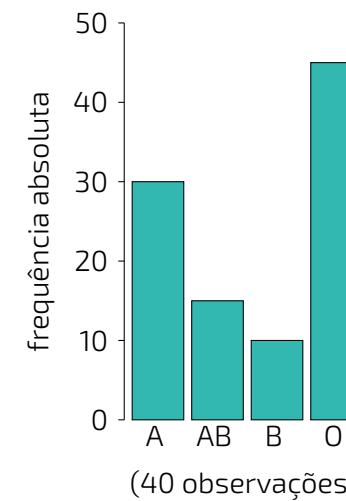


Figura 1.10: Gráficos de barras na escala da porcentagem

Comparando os dois gráficos, percebem-se apenas pequenas diferenças quanto às porcentagens dos sangues tipo "AB" e tipo "B".

Quando estamos trabalhando com variáveis qualitativas usamos a escala da frequência (absoluta, relativa ou porcentagem de cada resposta) na construção de gráficos para representar a distribuição de frequências das respostas dadas à variável sob investigação. As representações gráficas mais comuns são gráficos de barras ou colunas e gráficos de setores. Para comparações da mesma variável em grupos diferentes é comum usar o gráfico de barras múltiplas com frequências relativas ou porcentagens.

Como escolher entre o gráfico de setores ou o gráfico de barras para representar a distribuição de frequências de uma variável qualitativa?

Se o número de respostas diferentes é grande, maior que 4, ou se as diferenças nas frequências das respostas são pequenas, por exemplo uma tem porcentagem 20% e a outra tem porcentagem 23%, o gráfico de setores não será adequado, pois pequenas diferenças de ângulos não são perceptíveis, enquanto que com o gráfico de barras é fácil perceber pequenas diferenças.

Também, quando deseja-se fazer comparações múltiplas o gráfico de setores não é adequado. Observe que todos infográficos da atividade "Análise de infográficos" para comparar diferentes grupos quanto à prática de atividades físicas são gráficos de barras múltiplas.

## Objetivos Específicos

### Prática de atividade física na turma

- Comparar os resultados do suplemento especial da PNAD 2015 sobre prática de atividades físicas, realizando uma pesquisa na turma, incluindo algumas informações desse suplemento.

## Sugestões e discussões

### Prática de atividade física na turma

- Preparar uma tabela a ser preenchida pela turma com as informações: sexo, idade, prática ou não de atividade física em seu tempo livre, e a modalidade, de maneira a viabilizar a comparação dos dados obtidos com os resultados da PNAD/2015. A tabela poderá conter outras variáveis se forem julgadas de interesse pela turma como por exemplo, local da prática, duração da prática entre outras. Mas, para efeito de comparação com os infográficos, sexo e idade serão as variáveis necessárias nesse levantamento. Comente com os alunos que essa será uma amostra de conveniência, pois o interesse é estudar o perfil da turma quanto à prática de atividades físicas e por isso, as respostas da turma podem não ser similares às da pesquisa.
- Um padrão de itens a serem investigados é apresentado na figura a seguir, lembrando a necessidade de incluir a identificação de turma, caso a pesquisa seja realizada em mais de uma turma.

#### Nota 3

- Com base nas respostas obtidas, resumir a informação em tabelas de frequências, contar quantas respostas foram sim, calcular a porcentagem da turma que pratica atividade física e comparar com o resultado geral das pessoas de 15 anos ou mais, o percentual correspondente a essa faixa etária e o percentual correspondente a esse grau de instrução. Construir uma tabela de frequências com as modalidades esportivas incluindo as categorias apresentadas no infográfico do IBGE. Construir gráficos para representar as distribuições de frequências das variáveis investigadas nessa pesquisa. Construir gráficos de barras múltiplas, isto é, gráficos de barras separados por grupos diferentes, como por exemplo, sexo.

## Objetivos Específicos

### Classificação de variáveis

- Diferenciar variável qualitativa e variável quantitativa.
- Identificar variáveis qualitativas binárias.

#### Nota 4

## PRATICANDO A NATUREZA DA ESTATÍSTICA

### Prática de atividade física na turma

Atividade 5

Deseja-se comparar os hábitos de atividade física em tempo livre dos alunos da turma com os dados obtidos da PNAD/2015. Para isso preencha o formulário de dados fornecido pelo professor. Construa tabelas e gráficos resumindo a informação obtida.

### Classificação de variáveis

Atividade 6

Suponha que cada uma das variáveis a seguir foi observada para todos os alunos de sua turma. Indique se cada uma delas é uma variável qualitativa ou quantitativa. Se for uma variável qualitativa, indique se ela é binária (apenas duas respostas possíveis) ou não.

- Altura (em metros).
- Peso (em quilos).
- Índice de massa corporal (IMC) dado pelo quociente entre o peso (em quilos) e o quadrado da medida da altura (em metros).
- Tempo de sono na noite anterior.
- Se foi dormir na noite anterior antes ou depois da meia-noite.
- Mês de nascimento.
- Número de irmãos.
- Nota obtida na última avaliação de Matemática.
- Se tirou nota maior ou igual a 6,0 ou menor do que 6,0 na última avaliação de Matemática.
- Distância da casa à escola.
- Se o indivíduo possui cartão de crédito ou não.
- Modo de locomoção para a escola.

## Construção de gráficos para variáveis qualitativas

### Atividade 7

Considerando o [infográfico 4](#), transforme o gráfico de setores em gráfico de retângulos e os gráficos de retângulos em gráficos de setores.

## Análise de gráfico

### Atividade 8

Observe o gráfico a seguir publicado em um jornal.

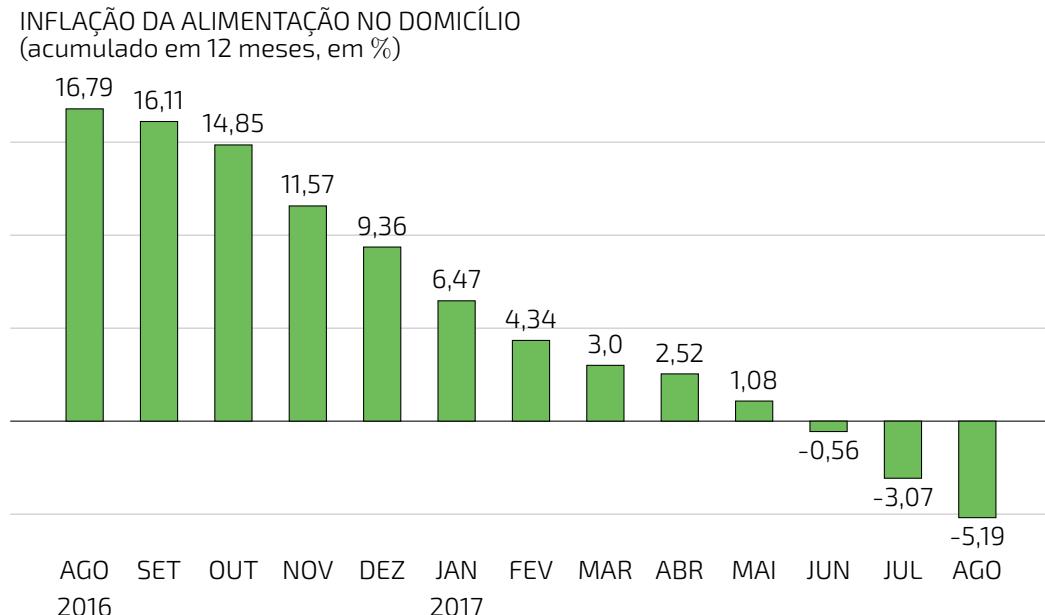


Figura 1.11: Inflação da alimentação acumulada nos últimos 12 meses (Fonte: IBGE)

- Como você classificaria esse gráfico?
- Qual é a informação representada pelo comprimento da barra nesse gráfico?
- Que tipo(s) de variável(is) ele está representando?
- Construa um gráfico diferente para representar a mesma informação, marcando num plano Cartesiano os pontos  $(x, y)$  em que  $x$  corresponde ao tempo e  $y$  corresponde à inflação acumulada no domicílio, unindo os pontos consecutivos por segmentos. É possível perceber a partir desse gráfico algum tipo de comportamento no período observado?

## Objetivos Específicos

### Construção de gráficos para variáveis qualitativas

- Construir gráficos de distribuições de frequências para variáveis qualitativas.

## Sugestões e discussões

### Construção de gráficos para variáveis qualitativas

Apesar de ser possível construir os gráficos solicitados nesta atividade, usando régua, transferidor e compasso, recomenda-se sugerir aos alunos usar algum recurso tecnológico para a construção dos mesmos, tais como GeoGebra ou planilha eletrônica (LibreOffice).

### Nota 5

## Objetivos Específicos

### Análise de gráfico

- Reconhecer que podem existir diversas formas de usar barras para representar algum tipo de dado, mas que nem todos os gráficos que usam barras são gráficos de barras no sentido da representação de uma distribuição de frequências.

## Sugestões e discussões

### Análise de gráfico

O gráfico desse exemplo é “um gráfico de barras”, mas as barras representam o valor da inflação da alimentação acumulado nos últimos 12 meses em função do tempo: de agosto de 2016 até agosto de 2017. Na seção “Explorando 2”, veremos que, para esse tipo de informação - valores de uma variável quantitativa ao longo do tempo -, é mais comum usar um gráfico de linhas unindo por segmentos os pontos consecutivos dados (tempo, valor da variável).

### Nota 6



## Bloco II: Gráficos para Variáveis Quantitativas

### Objetivo geral

Identificar o pensamento estatístico a partir de suas ideias fundamentais, a saber, população e amostra, parâmetro e estimador, distribuições empíricas de dados.

#### EM13MAT102

Analisar tabelas, gráficos e amostras de pesquisas estatísticas em relatórios divulgados por diferentes meios de comunicação, identificando, quando for o caso, inadequações que possam induzir a erros de interpretação, como escalas e amostras não apropriadas.

#### EM13MAT406

Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.

A duração prevista para trabalhar esse segundo bloco é de dois tempos de aula de 50 minutos.

Este bloco foi dividido em três seções: explorando, organizando e praticando que serão brevemente descritas a seguir.

#### a) Explorando:

Essa seção contém duas atividades, explorando Gráficos para Variáveis Quantitativas. A primeira atividade envolve a construção de um histograma de uma variável quantitativa contínua. A segunda atividade envolve a representação de uma série temporal, ou seja, de uma variável cujos valores são medidos ao longo do tempo.

#### b) Organizando:

Nessa seção, definimos os dois tipos de gráficos trabalhados nas atividades anteriores. Em seguida, discutimos a necessidade de se agrupar variáveis quantitativas contínuas em intervalos de classe, apresentando dois exemplos. O segundo exemplo, apresenta as definições de densidade de frequência relativa e densidade de frequência absoluta, fundamentais para a construção de histogramas quando os comprimentos dos intervalos de classe não são todos iguais. A seção termina com uma nota, ilustrando as diferenças entre um gráfico de barras e um histograma.

#### c) Praticando:

Inclui duas atividades de construção de histogramas. A primeira delas pede para considerar diferentes números de intervalos de classe. Para essa atividade é sugerido um aplicativo do GeoGebra que automatiza essas construções. O objetivo principal é perceber que a forma do histograma muda pouco quando aumentamos ou diminuímos o número de intervalos, considerando um total de 5 a 12 intervalos. A segunda atividade trabalha com a construção de um histograma para a situação em que os intervalos de classe apresentam comprimentos diferentes, levando a necessidade de se usar a escala da densidade de frequência.

Os exercícios de número 11, 12, 17 e 19, no final desse capítulo, referem-se ao conteúdo desse bloco.



## EXPLORANDO GRÁFICOS PARA VARIÁVEIS QUANTITATIVAS

### Construção do histograma

#### Atividade 9



Figura 1.13: Arranjo de radiotelescópios - Very Large Array (VLA), New Mexico, EUA. [Foto: Hajor CC-by-sa](#)

Os radiotelescópios são instrumentos de observação astronômica capazes de captar ondas eletromagnéticas não visíveis a olho nu: as ondas de rádio.

Um arranjo de oito radiotelescópios (A, B, C, D, E, F, G e H) como ilustrado na [Figura 1.13](#) detectou sinais cujos oito registros de tempo para cada radiotelescópio se encontram na tabela a seguir. (Fonte: dados fictícios)

A	B	C	D	E	F	G	H
3,03	4,37	5,04	5,73	4,03	5,37	6,04	6,74
3,38	4,46	5,11	5,84	4,38	5,46	6,11	6,84
3,60	4,55	5,19	5,95	4,60	5,55	6,19	6,96
3,78	4,63	5,29	6,08	4,78	5,64	6,29	7,08
3,92	4,71	5,36	6,23	4,92	5,72	6,36	7,23
4,04	4,79	5,45	6,41	5,04	5,79	6,45	7,40
4,16	4,87	5,54	6,62	5,16	5,87	6,54	7,63
4,27	4,95	5,64	6,97	5,26	5,95	6,64	7,97

### PARA REFLETIR

- Como construir uma tabela de frequências desses dados uma vez que os registros de tempo são todos distintos?
- Como você faria para visualizar o comportamento de uma variável com estas características?

A natureza quantitativa de uma variável contínua pode muitas vezes levar a resultados que praticamente não se repetem. Eles podem ser todos diferentes, como é observado no exem-

### Objetivos Específicos

#### Construção do histograma

- Identificar, na construção de um gráfico que represente a distribuição de frequências, a necessidade de agrupar em intervalos de classe os valores observados de uma variável quantitativa contínua.

#### Sugestões e discussões

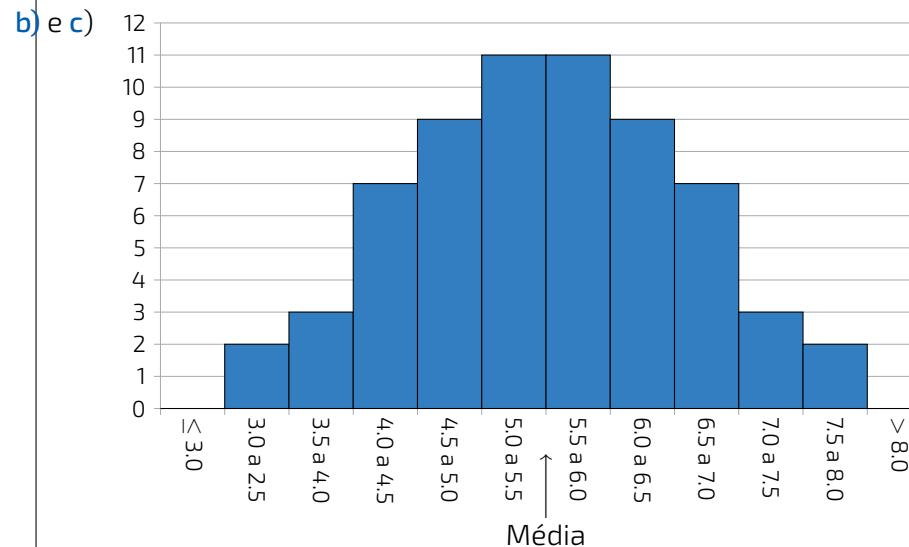
#### Construção do histograma

A construção do histograma será dirigida nessa atividade, mas recomenda-se fortemente o uso de recursos tecnológicos, como o GeoGebra, para esse tipo de construção.

Incluir link do arquivo desses dados em formato GeoGebra. Em alguns aplicativos, e é o caso do GeoGebra, é necessário substituir a vírgula como separador decimal do padrão brasileiro por ponto.

**Solução: Construção do histograma**

Intervalo de classe	Número de observações
[3,0; 3,5[	2
[3,5; 4,0[	3
[4,0; 4,5[	7
[4,5; 5,0[	9
[5,0; 5,5[	11
[5,5; 6,0[	11
[6,0; 6,5[	9
[6,5; 7,0[	7
[7,0; 7,5[	4
[7,5; 8,0[	2



d) A área do histograma construído é dada por

$$A = 0,5 \cdot (2 + 3 + 7 + 9 + 11 + 11 + 9 + 7 + 3 + 2) = 0,5 \cdot 64 = 32.$$

Por exemplo, o quociente da área do primeiro retângulo sobre a área total é

$$\frac{0,5 \cdot 2}{32} = \frac{13}{2} = \frac{1}{2} = 0,03125$$

que é a frequência relativa desse intervalo de classe. Da mesma forma, o quociente da área do terceiro retângulo sobre a área total é

$$\frac{0,5 \cdot 7}{32} = \frac{3,5}{32} = \frac{7}{64} = 0,109375$$

que é a frequência relativa desse intervalo de classe. Essa verificação ilustra a propriedade do histograma em representar a distribuição de frequências dos dados observados tal que cada retângulo do histograma representa a frequência do intervalo de classe correspondente no conjunto de dados como um todo.

Com o objetivo de identificar alguma estrutura no comportamento deste tipo de variável é necessário agrupar os valores em intervalos de classe, o que permite analisar a sua distribuição de frequências.

- a) Complete a tabela a seguir que utiliza de intervalos de amplitude 0,5 começando em 3,0. Observe que cada intervalo na tabela é fechado à esquerda e aberto à direita, isto quer dizer que, o limite inferior está incluso e o limite superior não está incluso.

Intervalo de classe	Número de observações
[3,0; 3,5[	
[3,5; 4,0[	
[4,0; 4,5[	
[4,5; 5,0[	
[5,0; 5,5[	
[5,5; 6,0[	
[6,0; 6,5[	
[6,5; 7,0[	
[7,0; 7,5[	
[7,5; 8,0[	

Para visualizar o comportamento desses dados, iremos construir um gráfico chamado histograma, composto por retângulos adjacentes cujas alturas representam a frequência de observações que ocorrem no intervalo correspondente. A base de cada retângulo corresponde aos limites do intervalo definido no agrupamento dos dados.

- b) Complete a figura a seguir com os demais retângulos do histograma.

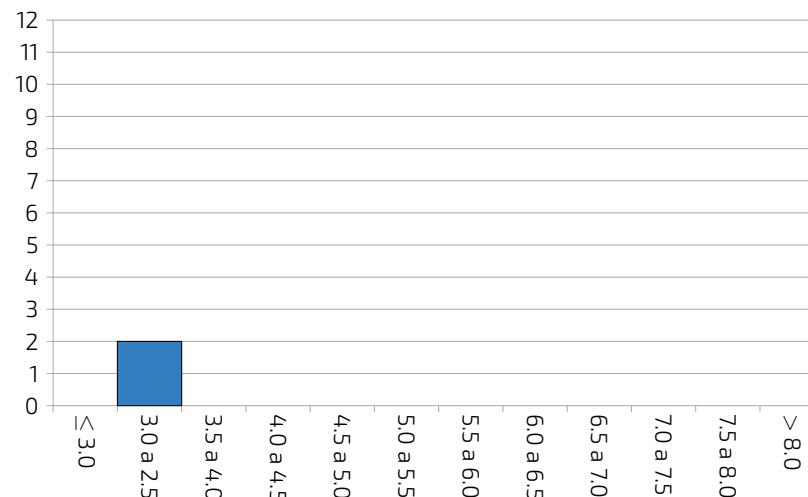


Figura 1.14: Histograma dos dados coletados pela grade de radiotelescópios

- c) Calcule a média dos dados da tabela e localize-a no gráfico, sabendo que a soma dos 64 registros de tempo é 351,95. O que você pode observar quanto à localização da média no histograma construído?

- d) Calcule a área correspondente ao histograma construído, somando as áreas dos 10 retângulos do histograma. Verifique que o quociente da área de cada retângulo e da área do histograma é igual à frequência relativa do intervalo de classe que ele representa.

### Medição da temperatura ao longo do tempo

#### Atividade 10

Você deve ter notado que a previsão do tempo é feita sempre a partir de dois números, isto ocorre porque a temperatura varia de forma contínua ao longo do dia e o que está sendo previsto são as temperaturas máxima e mínima. Por exemplo:  $28^\circ / 19^\circ$ , significa que a previsão da temperatura máxima durante o dia será aproximadamente de  $28^\circ\text{C}$  e a mínima  $19^\circ\text{C}$ .

Diversas variáveis meteorológicas (no sentido estatístico) são registradas nas estações meteorológicas: temperatura, precipitação (quantidade de chuva), umidade do ar, entre outras.

No Brasil, as estações estão a cargo do **Instituto Nacional de Meteorologia (INMET)** (<http://www.inmet.gov.br>) e as informações são armazenadas em bases de dados. Para poder tratar essas informações, frequentemente elas são resumidas por períodos de tempo de diferentes magnitudes: dias, semanas, meses ou anos.

Dados coletados ao longo do tempo (como a informação meteorológica) são conhecidos como séries de dados temporais ou, apenas, séries temporais, já que correspondem a variáveis que mudam continuamente ao longo do tempo e a informação só é útil se sabemos o momento em que foram realizadas as medições.

#### PARA REFLETIR

Forneça outros exemplos de séries temporais nas áreas de saúde, economia, finanças, educação etc.

A tabela a seguir fornece a média das temperaturas máximas para cada mês nos anos de 1991 a 2000 da cidade de Porto Alegre em graus centígrados (Fonte: [Banco de Dados Meteorológicos para Ensino e Pesquisa, BDMEP - INMET](#))

#### Objetivos Específicos

#### Medição da temperatura ao longo do tempo

- Identificar série temporal a partir de um conjunto de observações sobre uma variável quantitativa contínua variando no tempo.
- Elaborar gráficos de linha para ilustrar a evolução dos valores da variável ao longo do tempo.

#### Sugestões e discussões

#### Medição da temperatura ao longo do tempo

Para a construção do gráfico de linha será fornecida uma malha quadrículada para o preenchimento dos pontos, recomenda-se também uso de planilhas de cálculo para essa construção. Veja em [shorturl.at/pHJLU](http://shorturl.at/pHJLU) uma sugestão para realizar esta atividade.

Respostas possíveis na reflexão proposta são: índices de inflação, preços de diversos bens, índices da bolsa de valores, a população total em um território, a incidência de alguma enfermidade, a quantidade de vendas de um produto. É importante usar exemplos de dados que tenham aparecido recentemente na mídia ou que tenham relevância local.

Na discussão sobre sazonalidade, pedir aos alunos para trazer notícias de jornais ou revistas que contenham séries temporais. Mostrar que existem várias medições que são comparadas com as do ano anterior, por exemplo, inflação, crescimento do PIB, taxas de desemprego por trimestre, entre outras.

### Solução: Medição da temperatura ao longo do tempo

- a) Percebe-se temperaturas mais altas nos meses iniciais e finais do ano e, mais baixas, no meio do ano.
- b) Idem.

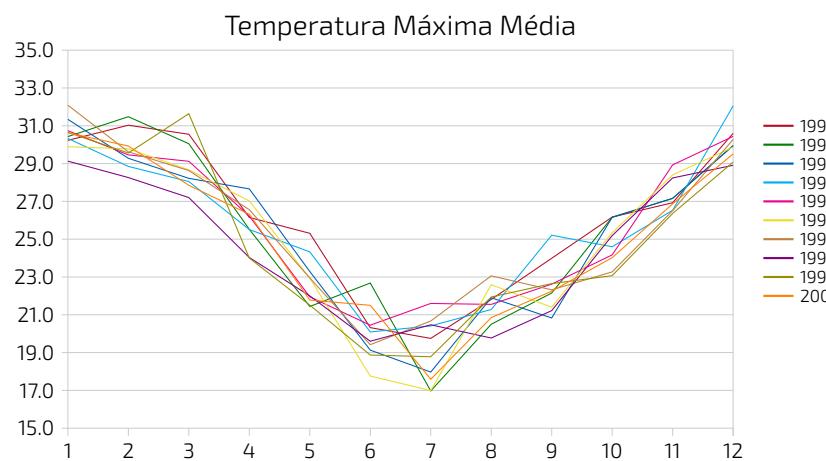


Figura 1.12: Gráficos de linhas com a temperatura máxima média mensal da cidade de Porto Alegre

- c) Idem ao item (b). Isso ocorre devido às estações do ano. No hemisfério sul temos temperaturas mais altas nos meses finais e iniciais do ano e temperaturas mais baixas no meio do ano.

Temperatura Máxima Média mensal nos anos 1991-2000 na cidade de Porto Alegre

Mês	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
1	30,23	30,43	31,34	30,33	30,74	29,89	32,09	29,13	30,65	30,63
2	31,03	31,48	29,28	28,85	29,46	29,78	29,62	28,26	29,56	29,93
3	30,55	30,05	28,22	28,05	29,12	28,67	28,63	27,20	31,64	27,85
4	26,15	25,52	27,66	25,51	26,22	27,03	26,56	24,03	24,00	26,32
5	25,31	21,44	23,29	24,33	21,95	22,94	22,95	22,00	21,51	21,78
6	20,32	22,68	19,13	20,09	20,45	17,76	19,42	19,60	18,87	21,50
7	19,75	16,91	17,97	20,41	21,60	16,99	20,67	20,47	18,78	17,59
8	21,81	20,50	21,90	21,28	21,55	22,59	23,06	19,77	21,94	20,85
9	23,99	22,14	20,83	25,21	22,62	21,40	22,32	21,22	22,65	22,25
10	26,17	26,16	26,40	24,60	24,17	25,34	23,27	25,19	23,07	24,02
11	26,93	27,16	28,07	26,53	28,93	28,40	26,51	28,24	26,36	26,87
12	30,60	29,95	29,73	32,05	30,44	29,87	30,28	28,91	29,08	29,51

- a) Escolha dois anos diferentes e localize os pontos da tabela na grade quadriculada usando o mês como abscissa ( $x$ ) e a temperatura como ordenada ( $y$ ). Utilize cores diferentes para a série de cada ano.
- b) Una os pontos correspondentes ao mesmo ano (mesma série) de meses consecutivos com um segmento e observe o resultado. Você percebe algum comportamento similar para a temperatura em anos diferentes?
- c) Compare seu gráfico com o de colegas que escolheram outros anos (ou acrescente séries de outros anos ao seu gráfico). O que você percebe com relação à temperatura nos meses iniciais, intermediários e finais do ano? A que se deve esse comportamento da temperatura?

#### VOCÊ SABIA?

Os gráficos construídos na atividade "Medição da temperatura ao longo do tempo" são chamados gráficos de linha. Esse tipo de gráfico é muito utilizado para variáveis quantitativas contínuas que dependem de uma outra variável quantitativa, neste caso o tempo. Quando a variável quantitativa é observada ao longo do tempo, o conjunto de dados resultante é chamado uma série temporal.

Como você já deve ter observado, a temperatura em Porto Alegre é mais baixa nos meses correspondentes ao inverno e mais alta na primavera e no verão, o que se repete cada ano. Este fenômeno, que se observa nos ciclos do gráfico, é chamado de sazonalidade. A origem deste conceito é exatamente o da sazonalidade que observamos na natureza com as estações ao longo do ano.

**Sazonalidade** Variações periódicas que se observam em séries temporais e que devem sua presença a um fenômeno implícito que incide de forma direta nas medições da variável observada.

Considere novamente os dados de temperatura da atividade anterior. Se representarmos todos os dados da tabela num único gráfico com a escala temporal das abscissas ao longo dos dez anos, obtemos o seguinte gráfico:

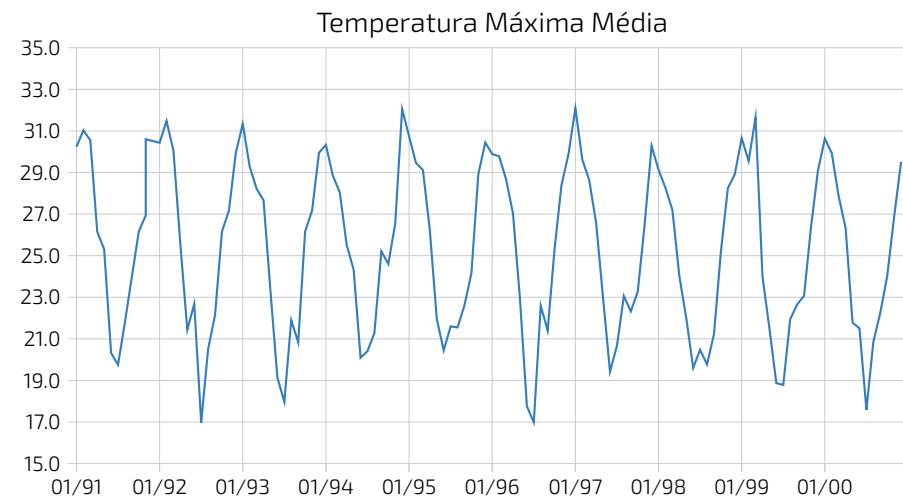


Figura 1.15: Efeito da sazonalidade no gráfico de linhas da temperatura máxima média

## ORGANIZANDO GRÁFICOS PARA VARIÁVEIS QUANTITATIVAS

Dois tipos de gráficos para representar variáveis quantitativas contínuas foram apresentados: o histograma e o gráfico de linha.

**Histograma** O histograma é uma representação gráfica da distribuição de frequências de uma variável quantitativa contínua agrupada em intervalos usando retângulos adjacentes. Cada retângulo no histograma corresponde a um intervalo considerado e a razão da área desse retângulo em relação à área total do histograma deve ser igual à frequência relativa de casos desse intervalo.

**Gráfico de linha** O gráfico de linha é uma representação útil quando os dados são uma série temporal, ou seja, os dados são coletados ao longo do tempo. Esse gráfico é construído marcando-se no plano Cartesiano os pontos  $(x, y)$  em que abscissa  $x$  representa o tempo e, a ordenada  $y$ , a variável quantitativa. Os pontos consecutivos são unidos por segmentos.

Quantos intervalos de classe considerar no agrupamento dos dados?

Quando existe a necessidade de agrupar os dados em intervalos, uma questão que se coloca é: quantos intervalos usar para que se possa reconhecer estruturas de frequências nesse conjunto? Não existe uma única resposta para essa questão. No entanto, devemos evitar tanto usar um número reduzido de intervalos, quanto usar um número grande de intervalos. Por exemplo, se usarmos um único intervalo, o histograma seria representado por um único retângulo que nada informaria sobre o comportamento dos dados, conforme o gráfico na [figura 1.16](#).

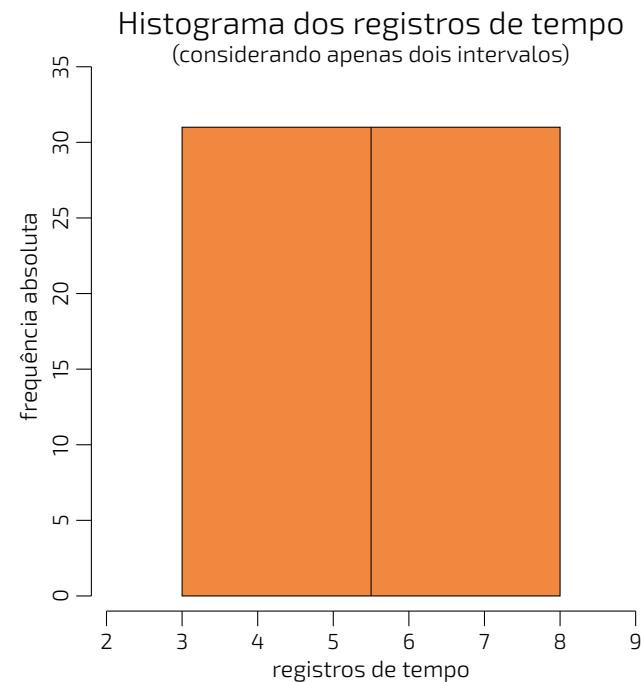


Figura 1.16: Histograma dos registros de tempo considerando apenas dois intervalos

Por outro lado, se o número de intervalos for igual ou superior ao número de observações, o histograma potencialmente teria apenas classes com uma única observação e o objetivo de

visualizar estruturas dos dados em análise se perderia. A [figura 1.17](#), construída a partir de 100 intervalos não revela a estrutura dos dados de registro de tempo, uma vez que cada classe contém no máximo duas observações.

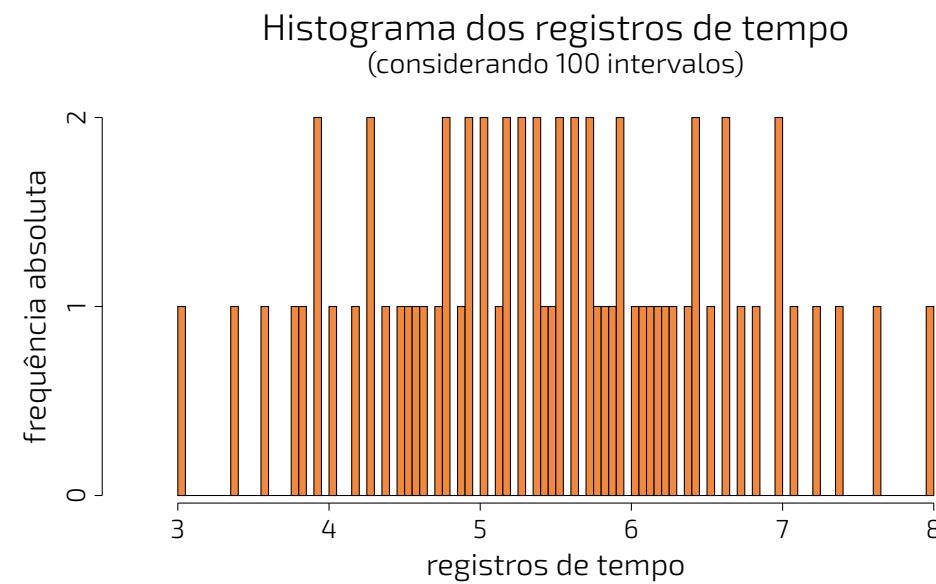


Figura 1.17: Histograma dos resgistros de tempo considerando cem intervalos

Embora não exista uma resposta única sobre quantos intervalos considerar, alguns autores sugerem usar o número inteiro mais próximo da raiz quadrada do número de observações, outros sugerem usar de 5 a 15 intervalos de amplitudes iguais. No GeoGebra, por exemplo, a função que constrói histogramas permite trabalhar com 3 a 20 intervalos. A [figura 1.18](#) apresenta um histograma construído com  $\sqrt{64} = 8$  intervalos.

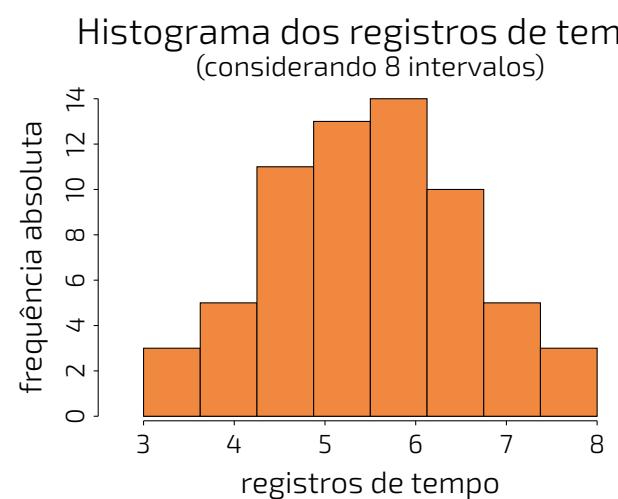


Figura 1.18: Histograma dos resgistros de tempo considerando oito intervalos

EXEMPLO 5 **Histograma cujos intervalos de classe têm o mesmo comprimento**

Suponha a seguinte distribuição de frequências de um conjunto de salários (em salários mínimos) de 50 funcionários de uma empresa.

Intervalo de classe	Frequência absoluta	frequência relativa
[ 1 ; 3 [	4	0,08
[ 3 ; 5 [	12	0,24
[ 5 ; 7 [	20	0,40
[ 7 ; 9 [	8	0,16
[ 9 ; 11 [	6	0,12

Observe que nessa tabela todos os intervalos têm amplitude 2. Veja um histograma construído para esses dados na [figura 1.19](#).

Histograma na escala da frequência absoluta  
intervalos de amplitudes iguais

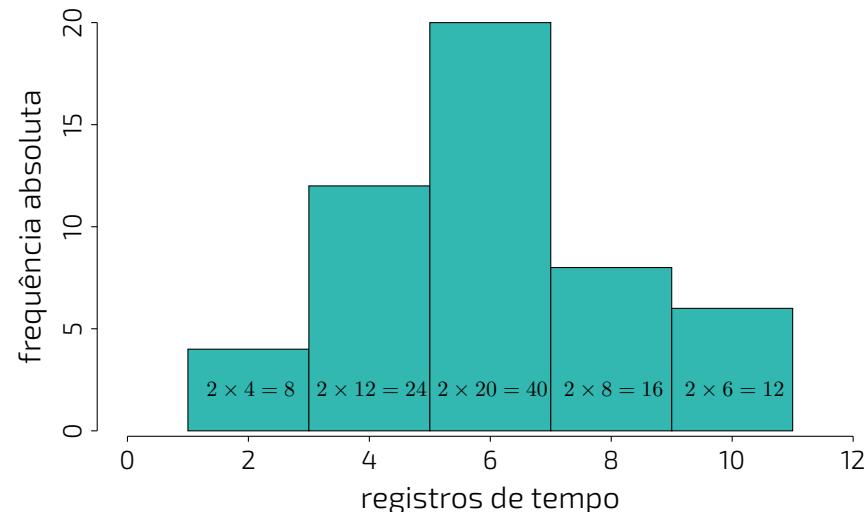


Figura 1.19: Histograma na escala da frequência absoluta

A área total desse histograma que foi construído na escala da frequência absoluta é dada pela soma das áreas dos 5 retângulos desse histograma que é igual a  $2 \cdot (4 + 12 + 20 + 8 + 6) = 2 \cdot 50$ . De fato, podemos dizer que a área total de um histograma construído na escala da frequência absoluta é dada pelo produto do comprimento dos intervalos, que deve ser o mesmo para todos, e o número total de observações.

Assim, a área relativa do primeiro retângulo é dada por  $\frac{2 \cdot 4}{2 \cdot 50} = \frac{4}{50} = 0,08$ . E, assim por diante, para cada um dos retângulos iremos verificar que a área relativa é igual a frequência relativa do intervalo que ele representa.

Se usássemos a escala da frequência relativa para a altura dos retângulos, a área total do histograma será dada pelo comprimento dos intervalos, comum a todos eles, que no caso desse exemplo será 2. Ainda assim, as áreas relativas para cada retângulo serão iguais às respectivas frequências relativas.

Até aqui consideramos intervalos de mesma amplitude e usamos, como a altura dos retângulos, a frequência absoluta ou a frequência relativa das observações no intervalo. Porém, quando os intervalos apresentam amplitudes desiguais, usar qualquer uma dessas duas frequências como a altura dos retângulos não será mais apropriado.

#### EXEMPLO 6 Histograma cujos intervalos de classe têm comprimento desiguais

Intervalo de classe	Frequência absoluta	Frequência relativa	Comprimento do intervalo
[1; 3[	4	0,08	2
[3; 5[	8	0,16	2
[5; 8[	18	0,36	3
[8; 12[	12	0,24	4
[12; 16[	8	0,16	4

Observe que os comprimentos dos intervalos de classe variam, sendo 2, 2, 3, 4 e 4, respectivamente.

Nesse caso devemos usar a densidade de frequência absoluta ou relativa obtida pela razão entre frequência e amplitude do intervalo.

$$\text{Densidade de frequência absoluta} = \frac{\text{frequência absoluta do intervalo}}{\text{amplitude do intervalo}}$$

$$\text{Densidade de frequência relativa} = \frac{\text{frequência relativa do intervalo}}{\text{amplitude do intervalo}}$$

A [figura 1.20](#) ilustra uma construção equivocada do histograma desses dados, usando a frequência absoluta.

Histograma incorreto na escala da frequência absoluta  
intervalos de amplitudes desiguais

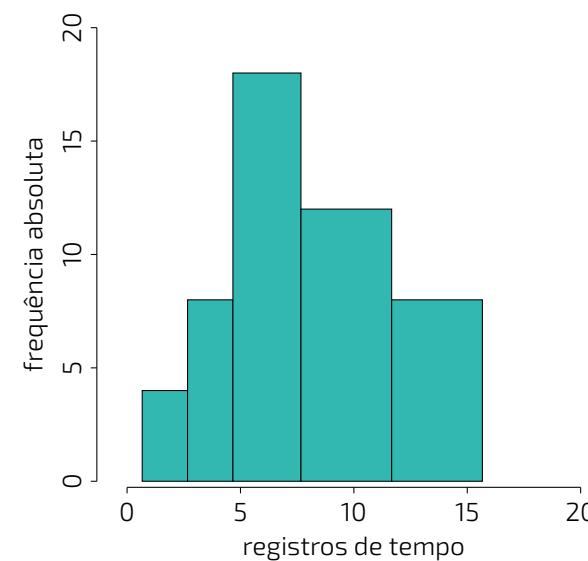


Figura 1.20: Histograma incorreto

Nesse histograma a área total é  $2 \times 4 + 2 \times 8 + 3 \times 18 + 4 \times 12 + 4 \times 8 = 158$ .

Observe que a razão da área do primeiro retângulo em relação à área total é dada por  $\frac{8}{158} \approx 0,051$ , porém a frequência relativa do primeiro intervalo é 0,08! A razão da área do último retângulo é  $\frac{4 \cdot 8}{158} \approx 0,20$ , porém a frequência relativa desse intervalo é 0,16! Ou seja, esse histograma não representa corretamente a distribuição de frequências desses dados. Na tabela a seguir, foram calculadas as densidades de frequência absoluta.

Intervalo de classe	Frequência absoluta	Amplitude do intervalo	Densidade freq. absoluta (freq. abs./comp. intervalo)
[1; 3[	4	2	2
[3; 5[	8	2	4
[5; 8[	18	3	6
[8; 12[	12	4	3
[12; 16[	8	4	2

Veja a seguir a construção do histograma na escala da densidade de frequência absoluta e observe que agora ele representa corretamente a distribuição de frequências.

Histograma na escala da densidade de frequência absoluta  
intervalos de amplitudes desiguais

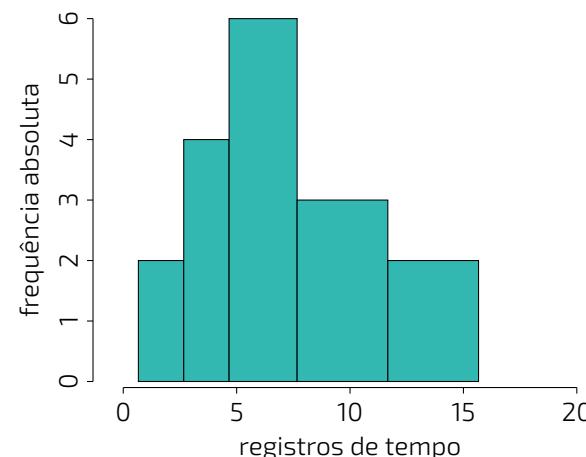


Figura 1.21: Histograma correto

Nesse histograma a área total é  $2 \times 2 + 2 \times 4 + 3 \times 6 + 4 \times 3 + 4 \times 2 = 50$ .

Comparando as [figuras 1.20](#) e [1.21](#), podemos perceber que a primeira distorce a estrutura da distribuição de frequências, atribuindo pesos maiores aos intervalos de maior amplitude e, menores, aos intervalos de menor amplitude. O uso da escala da densidade de frequência absoluta corrige essa distorção e garante a propriedade de que as áreas relativas de cada retângulo correspondam as suas frequências relativas.

As escalas de densidades de frequência, absoluta ou relativa, podem ser usadas sempre na construção do histograma. Já as escalas da frequência absoluta ou da frequência relativa apenas podem ser usadas quando todos os comprimentos dos intervalos de classe são iguais.

Quando usamos a escala da densidade de frequência absoluta, a área total do histograma será sempre igual ao número total de observações, pois a área de cada retângulo será igual à frequência absoluta correspondente e, a soma das frequências absolutas é sempre igual ao número total de observações.

Quando usamos a escala da densidade de frequência relativa, a área total do histograma será sempre igual a 1, pois a área de cada retângulo será igual à frequência relativa correspondente e, a soma das frequências relativas é sempre igual a 1.

Normalmente, na primeira construção dos intervalos de classe consideramos sempre intervalos de comprimentos iguais. Mas pode acontecer, nesse agrupamento, intervalos vazios ou intervalos com um número muito grande de observações. Quando essas situações ocorrem recomenda-se juntar dois intervalos consecutivos no primeiro caso ou subdividir o intervalo no segundo caso.

#### Gráfico de Barras versus Histograma

O gráfico de barras não é um histograma, apesar de suas representações serem parecidas. Os gráficos de barras são úteis para descrever a distribuição de frequências de uma variável qualitativa. Nesse gráfico só há um eixo com escala que corresponde aos valores das frequências das categorias (respostas) da variável. As barras podem ser tanto verticais como horizontais e são apresentadas de forma igualmente espaçada. Cada barra representa uma resposta da variável qualitativa e a altura da barra corresponde à frequência daquela resposta. Observe que o posicionamento das barras é livre, conforme os gráficos da [figura 1.22](#).

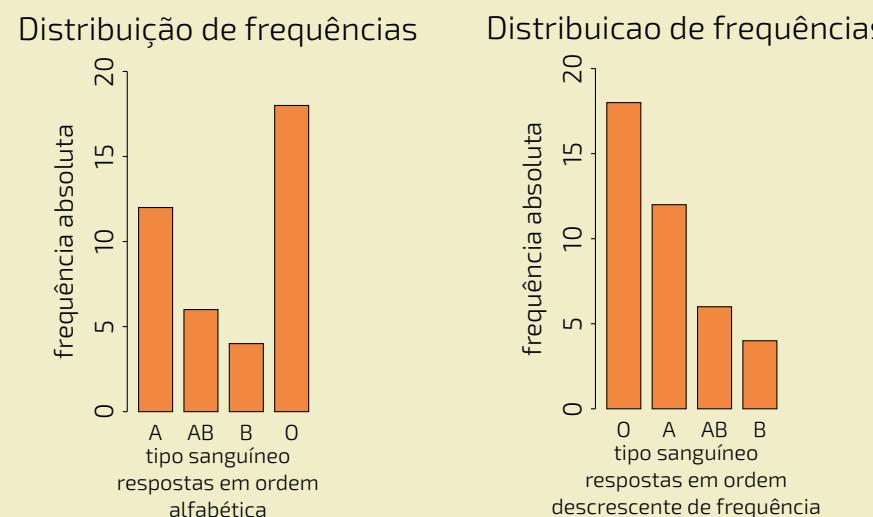


Figura 1.22: Gráfico de barras: duas formas de apresentação

O mais comum é dispor as respostas em ordem decrescente de frequência. Esse tipo de gráfico também pode ser usado para representar uma variável quantitativa discreta, sendo que nesse caso, as posições das barras correspondem aos valores assumidos pela

variável. Pela natureza discreta da variável, as barras não são adjacentes e, pela natureza quantitativa da variável, o posicionamento das barras não é livre.

Os histogramas são úteis para representar a distribuição de frequências de uma variável quantitativa contínua cujos valores foram agrupados em intervalos. No histograma, o eixo das abscissas (horizontal) representa a escala da variável contínua e, o eixo das ordenadas (vertical) representa a escala da frequência ou densidade de frequência que é definida como a razão entre a frequência e a amplitude do intervalo.

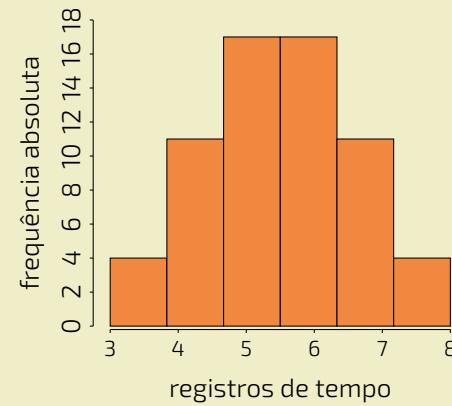


Figura 1.23: Histograma dos registros, considerando 5 intervalos

Não podemos variar livremente a posição dos intervalos nesse gráfico (figura 1.23). Ele revela uma estrutura importante desses dados, a saber, os registros de tempo ocorrem com maior frequência nos intervalos intermediários (de 4 a 6) e com frequência bem menor nos intervalos extremos (de 3 a 4 e de 7 a 8).

## PRATICANDO GRÁFICOS PARA VARIÁVEIS QUANTITATIVAS

### Construção de histogramas

#### Atividade 11

Refaça o histograma dos dados de registros de tempo variando o número de intervalos de classe. Compare a forma dos histogramas obtidos com a forma do histograma construído na atividade [Construção do Histograma](#).

### Histogramas com intervalos de amplitudes desiguais

#### Atividade 12

Suponha a seguinte distribuição de frequências de salários medidos em salários mínimos para 200 funcionários de uma empresa.

Intervalo de classe	Frequência Absoluta	Frequência Relativa
[ 2,0 ; 3,0 [	12	0,06
[ 3,0 ; 5,0 [	40	0,20
[ 5,0 ; 7,0 [	80	0,40
[ 7,0 ; 10,0 [	48	0,24
[ 10,0 ; 15,0 [	20	0,10

- Determine as amplitudes de cada intervalo considerado na tabela.
- Construa um histograma adequado para esses dados.

### Objetivos Específicos

#### Construção de histogramas

- Avaliar a forma do histograma a partir da variação do número de intervalos considerados

### Sugestões e discussões

#### Construção de histogramas

Essa atividade deve ser realizada com algum recurso tecnológico. Um exemplo de como realizá-la usando o Geogebra pode ser acessado em [shorturl.at/qDNQZ](http://shorturl.at/qDNQZ).

Arrastando o cursor na linha que representa classes (intervalos) é possível variar de três a 20 intervalos.

#### Nota 7

### Objetivos Específicos

#### Histogramas com intervalos de amplitudes desiguais

- Construir histogramas nos casos em que os intervalos apresentam amplitudes desiguais.
- Definir densidade de frequência absoluta e relativa.

### Sugestões e discussões

#### Histogramas com intervalos de amplitudes desiguais

Nessa atividade o histograma deve ser construído usando a escala de densidade de frequência (absoluta ou relativa).

#### Nota 8



## Descrição da parte final do capítulo

A duração prevista para trabalhar essa parte final do capítulo é de dois tempos de aula de 50 minutos.

Este bloco foi dividido em três seções: para saber mais, projeto aplicado e exercícios.

### a) Para saber mais

Nessa seção são apresentados alguns tipos de amostragem: aleatória simples, sistemática, por conglomerados e estratificada. Um exemplo, ilustrando os diferentes tipos foi incluído. O exercício de número 20, no final desse capítulo, refere-se ao conteúdo dessa seção.

### b) Projeto aplicado: Projeto a ser realizado de modo transversal paralelamente às aulas. O projeto envolverá

- a definição de um problema a ser investigado, de preferência incluindo outra disciplina,
- a definição da população,
- a construção de um questionário para a coleta de dados (amostra ou censo),
- a organização dos dados obtidos por meio de tabelas, gráficos e medidas resumo.
- a interpretação dos resultados obtidos e
- a confecção de relatório final.

Trabalhando o projeto de forma transversal, espera-se que na conclusão do mesmo, o capítulo "Medidas de Posição e Dispersão" já tenha sido trabalhado. As etapas sugeridas para o desenvolvimento do projeto incluem a formulação de uma questão de investigação estatística, o planejamento e a coleta dos dados, o tratamento dos dados e a interpretação dos resultados, com a elaboração de um relatório final do projeto.

### c) Exercícios



## PARA SABER + A SELEÇÃO DE AMOSTRAS

Quando queremos estender nossas observações provenientes de uma amostra para a população é importante ter cuidado na sua seleção, pois ela deve ser representativa da população. Embora não seja nosso objetivo aqui descrever métodos variados de seleção de amostras, cabe destacar que existem dois tipos principais de seleção: os probabilísticos e os não probabilísticos.

O primeiro tipo é fundamental para que seja possível avaliar a incerteza das conclusões devido à amostragem tais como margem erro e nível de confiança. Nesse tipo de seleção de amostra, conhecemos a probabilidade de seleção dos elementos da população na amostra. Entre os métodos probabilísticos mais comuns destacam-se

- a) amostragem aleatória simples: todas as amostras de igual tamanho têm probabilidades iguais de serem selecionadas.
- b) amostragem estratificada: a população é dividida em grupos de elementos homogêneos (similares nas características a serem investigadas) e os grupos são heterogêneos entre si. A amostra é composta por amostras aleatórias simples de cada grupo, em geral, proporcionalmente aos tamanhos dos grupos.
- c) amostragem por conglomerados: a população é subdividida em conglomerados (subpopulações). Uma amostra aleatória simples de conglomerados é obtida e, em seguida, todos os elementos dos conglomerados escolhidos são observados.
- d) amostragem sistemática: toda a população deve estar catalogada numa lista, por exemplo, lista dos alunos matriculados numa escola em ordem alfabética. Suponha que a lista contenha 1000 alunos e que se deseja obter uma amostra de tamanho 50. Para isso divide-se 1000 por 50 obtendo-se 20 blocos de 50 alunos. Sorteia-se ao acaso um número de 1 a 20, por exemplo, o número 9. Seleciona-se o aluno de número 9 e, depois, os próximos elementos são selecionados de 20 em 20 como uma Progressão Aritmética de razão 20 e primeiro termo 9.

Os casos mais comuns de métodos não probabilísticos são amostragem por conveniência e amostragem por julgamento. A amostragem por conveniência caracteriza-se por não ter um plano particular de amostragem. O objetivo nesse caso não seria generalizar conclusões e sim descrever as características principais do grupo de estudo. Nas amostras por julgamento, os elementos da amostra são escolhidos por um especialista no assunto sob investigação. A grande desvantagem dos métodos não probabilísticos é a impossibilidade de avaliar incertezas devido à amostragem.

### EXEMPLO 7 Horário de entrada

A direção de uma escola de Ensino Médio deseja realizar uma pesquisa para conhecer a opinião de seus 520 alunos sobre a antecipação em 30 minutos dos horários de seus turnos. Para tanto, devem decidir entre as seguintes estratégias de seleção de amostra.

- a) 40 alunos considerando os primeiros a chegar na escola na segunda-feira. *Temos, nesse caso, uma amostra de conveniência, pois a probabilidade de seleção dos alunos não é determinada no plano de amostragem: selecionar os 40 primeiros. Observe também que, esse esquema de seleção não parece razoável para essa pesquisa, pois*

é possível resultar numa resposta viesada, isto é, tendendo a favorecer à mudança de horário por considerar apenas os primeiros a chegar, não representando necessariamente a opinião da maioria dos 520 alunos da escola.

- b) 40 alunos escolhidos a partir do cadastro de 520 alunos matriculados da seguinte forma: como  $520/40 = 13$ , sorteia-se ao acaso um número entre 1 e 13, por exemplo 8; e, depois, seleciona-se do cadastro os alunos nas posições 8, 21, 34, 47, e, assim sucessivamente de 13 em 13, até o aluno de posição 515 no cadastro de alunos, totalizando 40 observações. Trata-se de uma amostra sistemática cuja probabilidade de seleção é conhecida, a saber,  $\frac{1}{13}$ . Nesse caso, espera-se que a amostra seja representativa do conjunto de estudantes do colégio.
- c) 40 alunos sendo 16 do primeiro ano, 14 do segundo ano e 10 do terceiro ano, escolhidos ao acaso, tendo em mente que na escola 40% dos alunos são de primeiro ano, 35% dos alunos são de segundo ano e 25% dos alunos são de terceiro ano. Trata-se de uma amostra estratificada proporcionalmente ao tamanho dos estratos que correspondem aos anos do Ensino Médio. As probabilidades de seleção de amostra são conhecidas. Nesse caso, também espera-se que a amostra seja representativa do conjunto de estudantes do colégio.
- d) 40 alunos de uma turma do segundo ano na qual está um filho do diretor. Trata-se de uma amostra de conveniência e que pode resultar num resultado duplamente viesado, tanto pelo fato de que a turma escolhida não representa necessariamente a maioria dos estudantes da escola, quanto pela presença do filho do diretor nessa turma que pode influenciar o resultado.

## PROJETO APLICADO

Faça uma investigação sobre algum tema de interesse. Essa atividade deve ser realizada em grupos. Após a aprovação do tema pelo professor, as seguintes etapas deverão ser realizadas:

- a) Elaborar um cronograma considerando um prazo de três meses para concluir o projeto.
- b) Definir a população e os elementos (unidades de observação) (pessoa, família, domicílio, cidade, escolas, turmas, estudantes etc.).
- c) Definir como a amostra será escolhida e quantos elementos serão considerados na amostra.
- d) Definir que variáveis serão observadas para cada elemento; em cada caso pode ser importante coletar informações de outras variáveis que podem estar relacionadas à questão de interesse para verificar se essas variáveis de alguma forma estão relacionadas.
- e) Construir um questionário para a coleta de informações.
- f) Coletar as informações.
- g) Construir uma planilha com os dados obtidos.
- h) Organizar os dados em tabelas e gráficos apropriados ao tipo de variável, resumindo a informação obtida.
- i) Elaborar um relatório de pesquisa, incluindo todas as etapas do projeto, resultados, análises e conclusões.

## Material suplementar

Sugestão de vídeos sobre o que é a Estatística, para que serve a Estatística e exemplos de aplicação da Estatística.

- O Prazer da Estatística — <https://www.youtube.com/watch?v=nB5l90W2eyo>
- O que é Estatística? - <https://www.youtube.com/watch?v=-Wm9cxiXUe0>
- Ação, Reação, Correlação - <http://m3.ime.unicamp.br/recursos/1043>

Sugestão de páginas para trabalhar com dados reais.

- Página do Programa de Desenvolvimento das Nações Unidas — [hdr.undp.org/en/data](http://hdr.undp.org/en/data)
- Organização Mundial de Saúde — [www.who.int/](http://www.who.int/)
- Instituto Brasileiro de Geografia e Estatística (IBGE) — [https://www.ibge.gov.br/](http://www.ibge.gov.br/)
- Estimativas de população dos municípios brasileiros — [https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/](http://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/)
- Atividades para o Ensino Médio — [https://vamoscontar.ibge.gov.br/atividades/ensino-medio.html](http://vamoscontar.ibge.gov.br/atividades/ensino-medio.html)
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) — <http://portal.inep.gov.br/inep-data>
- Censo Escolar INEP (último censo 2014) — [http://inepdata.inep.gov.br/analytics/saw.dll?Portal&PortalPath=%2Fshared%2FGeral%2F\\_portal%2FDissemina%C3%A7%C3%A3o%20dos%20Censos](http://inepdata.inep.gov.br/analytics/saw.dll?Portal&PortalPath=%2Fshared%2FGeral%2F_portal%2FDissemina%C3%A7%C3%A3o%20dos%20Censos)
- Frota de veículos por município do RJ — [http://www.detran.rj.gov.br/\\_estatisticas.veiculos/index.asp](http://www.detran.rj.gov.br/_estatisticas.veiculos/index.asp)
- Instituto de Pesquisa Econômica Aplicada (IPEA) — <http://www.ipea.gov.br/portal/>
- IPEA Data — <http://www.ipeadata.gov.br/Default.aspx>
- Ministério da Saúde - Datasus — <http://datasus.saude.gov.br/transferencia-download-de-arquivos/arquivos-de-dados>

## EXERCÍCIOS

1 Estabeleça se as seguintes conclusões a respeito de dados estatísticos provêm do uso da Estatística Descritiva (D) ou Estatística Inferencial (I), justificando as respostas.

- No estado do Rio de Janeiro, a média de gasto semanal de consumo de gasolina numa amostra de 500 proprietários de carros foi de R\$ 200,00. O governo do Rio de Janeiro afirma que a média semanal de gasto em gasolina no estado é R\$ 200,00.
- Uma amostra de 250 residentes de uma cidade indicou que 45 destes são funcionários públicos. Assim 18% desses 250 residentes trabalham para o governo.
- A média de idade de trabalhadores formais obtida de uma amostra de 380 habitantes de Nova Iguaçu foi de 34 anos.
- Numa pesquisa feita com 2000 habitantes da Grande São Paulo, 768 disseram fazer uso regular de sacola ecológica pessoal em suas compras de supermercado. A prefeitura conclui que mais de um terço dos habitantes da Grande São Paulo já aderiu à sacola ecológica.

2 Classifique as proposições abaixo como sendo de natureza Matemática (M) ou de natureza Estatística (E), justificando as respostas.

- Não existe número real  $x$  que satisfaça  $2x^2 + 3x + 2 = 0$ .
- Um dado que lançado 180 vezes não tenha gerado a face 6 é um dado viciado (não equilibrado).
- Se de 100 crianças vacinadas pela BCG contra a tuberculose apenas duas contraíram a doença, então a vacina tem eficácia de 98%.
- Todo número par maior que 2 pode ser representado pela soma de dois números primos.
- A precipitação de chuva amanhã será de aproximadamente 36 mm.
- Se um feixe de paralelas está cortado por duas transversais então os segmentos determinados sobre uma transversal são respectivamente proporcionais aos segmentos determinados na outra.

3 Suponha que pesquisadores desejam investigar como o peso pode afetar a pressão sanguínea. Classifique os tipos de variáveis em cada uma das situações a seguir.

- O peso e a pressão sanguínea dos indivíduos são registrados.
- Os indivíduos são classificados como abaixo do peso, normais ou acima do peso e suas pressões sanguíneas são registradas.
- Os indivíduos são classificados como tendo pressão alta, normal ou baixa e seus pesos são registrados.
- Os indivíduos são classificados como abaixo do peso, normais ou acima do peso e como tendo pressão alta, normal ou baixa.

4 Numa pesquisa realizada foi descoberto que 63% dos americanos adultos pesquisados não querem viver até os 100 anos. Em média, as pessoas pesquisadas disseram desejar viver até a idade de 91 anos.

### Solução: Exercícios

1

- Inferencial, pois a afirmação é feita para todo o Estado do Rio de Janeiro, não restringindo-se a uma descrição da amostra pesquisada.
- Descritiva, pois a afirmação refere-se apenas à amostra observada.
- Descritiva, pois a afirmação refere-se apenas à amostra pesquisada.
- Inferencial, pois a afirmação é feita para a população de São Paulo, baseando-se no resultado obtido na amostra investigada:  $\frac{768}{2000} \cdot 100 = 38,4\%$  das pessoas na amostra usam as sacolas ecológicas.

2

- Matemática, pois pode ser classificada em verdadeira ou falsa. Em particular, trata-se de uma afirmação verdadeira, pois de fato não existem raízes reais para essa equação.
- Estatística, pois embora seja esperado que ao lançar um dado honesto 180 vezes, ocorra a face 6 em média  $30\left(\frac{180}{6}\right)$  vezes, ainda assim não é possível afirmar se essa proposição essa falsa ou verdadeira. Por se tratar de uma proposição estatística, apenas podemos dizer que ela tem alto nível de confiança.
- Estatística, pois a verificação feita foi para uma amostra de 100 crianças e por se tratar de uma inferência estatística, não é possível dizer se ela é verdadeira ou falsa, apenas podemos dizer que ela tem alto nível de confiança.
- Matemática. Trata-se da conjectura de Goldbach, proposta pelo matemático prussiano Christian Goldbach e é um dos problemas mais antigos não resolvidos da matemática. Verificações por computador já confirmaram a conjectura de Goldbach para muitos números. No entanto, a efetiva demonstração matemática ainda não ocorreu. (Fonte: – [https://pt.wikipedia.org/wiki/Conjetura\\_de\\_Goldbach](https://pt.wikipedia.org/wiki/Conjetura_de_Goldbach) Ainda assim é uma proposição matemática: ou é verdadeira ou é falsa.
- Estatística, pois trata-se apenas de uma previsão e, portanto, podemos atribuir um nível de confiança para ela.
- Matemática, pois pode ser classificada em verdadeira ou falsa. De fato, trata-se do teorema de Tales e, portanto, é verdadeira.

### Solução: Exercícios

- 3**
- a) Peso e pressão são tratados como variáveis quantitativas contínuas.
  - b) O peso é tratado como variável qualitativa ordinal e a pressão é tratada como variável quantitativa contínua.
  - c) O peso é tratado como variável quantitativa contínua e a pressão como variável qualitativa ordinal.
  - d) Ambos são tratados como variáveis qualitativas ordinais.
- 4**
- a) Amostra, pois referem-se a um subconjunto da população de americanos adultos.
  - b) A pergunta poderia ser: "Você deseja viver até os 100 anos?", pois a resposta seria uma variável qualitativa: sim, não ou não sei.
  - c) A pergunta poderia ser "Até que idade você gostaria de viver?", pois a resposta seria um número.
- 5**
- a) Principal equipamento usado por aluno para acessar a internet. Variável qualitativa nominal.
  - b) Gráfico de barras.
  - c) De fato, vimos que o gráfico de setores é um gráfico adequado para representar as frequências de respostas de variáveis qualitativas, mas nesse caso, há frequências muito pequenas (1%, 2%, 5%, 6%) e essas pequenas diferenças levarão a setores pouco distinguíveis entre si.

- a) Devemos considerar os americanos adultos pesquisados como a amostra ou como a população? Por quê?
- b) Existe uma variável qualitativa de interesse nessa pesquisa; qual poderia ter sido a questão formulada na pesquisa para obter as informações dessa variável? Justifique a sua resposta.
- c) Existe uma variável quantitativa de interesse nessa pesquisa; qual poderia ter sido a questão formulada na pesquisa para obter as informações dessa variável? Justifique a sua resposta.

**5** "O telefone celular, antes tão mal visto no ambiente escolar, vai ocupando cada vez mais espaço na sala de aula: em 2016, 52% das escolas utilizavam o aparelho em atividades com os alunos. É o que aponta a pesquisa TIC Educação 2016, do Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (Cetic). Foram coletados dados de 1.106 escolas, em turmas de 5º e 9º ano do ensino fundamental e do 2º ano do ensino médio. Participaram das entrevistas 935 diretores, 922 coordenadores pedagógicos, 1.854 professores de diversas disciplinas e 11.069 estudantes. A pesquisa aconteceu entre agosto e dezembro de 2016." ([Leia a reportagem](#) publicada no G1.com.br).

O gráfico a seguir ilustra um dos resultados dessa pesquisa.

### Principais equipamentos usados por alunos para acessar a internet

Porcentagem sobre o total de estudantes participantes da pesquisa

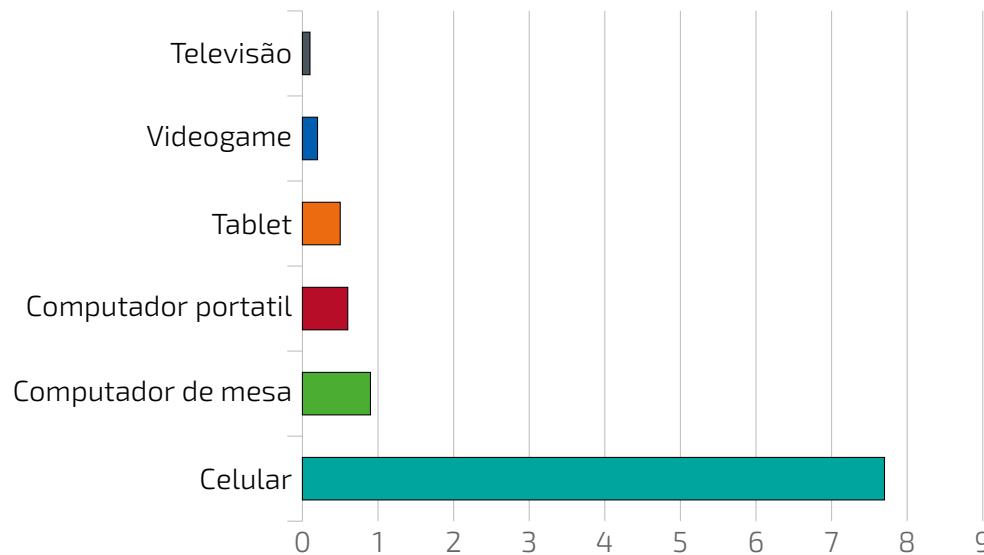
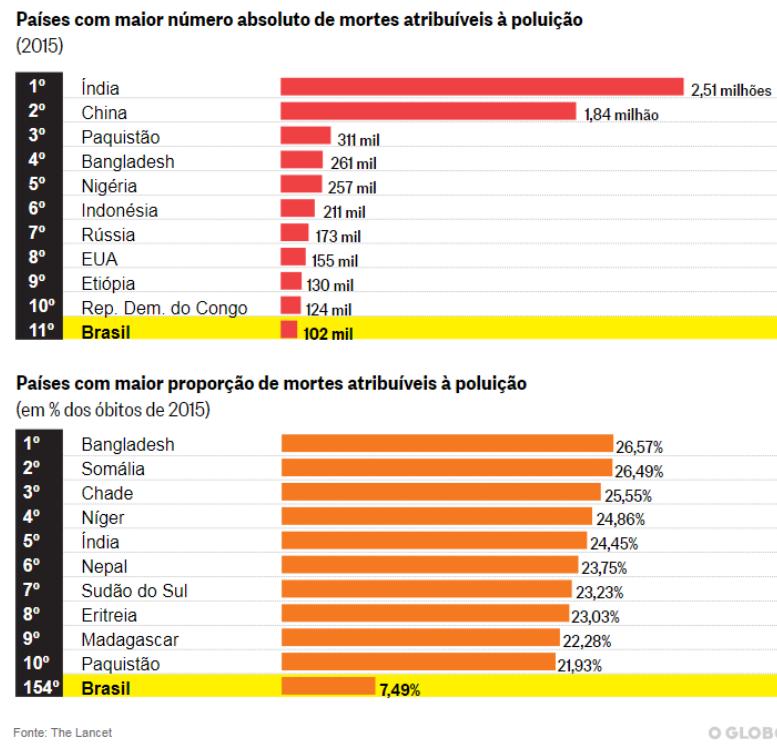


Figura 1.24: Fonte: CGI.br/NIC.br/Cetic.br/TIC Educação 2016

- a) Que variável foi analisada nesse gráfico e qual a sua classificação?
- b) Que tipo de gráfico foi usado para representar as respostas dessa variável?
- c) Você acha que o gráfico de setores seria adequado para representar essa distribuição de frequências? Por que?

- 6 "Poluição ambiental provoca uma em cada seis mortes no mundo": reportagem publicada pelo jornal O Globo em 20 de outubro de 2017. Analise os gráficos publicados. ([Leia a reportagem](#)).



- a) Que variável foi observada no primeiro gráfico? Que escala foi usada para o comprimento das barras nesse gráfico? Em que posição está o Brasil?
- b) Que variável foi observada no segundo gráfico? A que se deve a mudança radical de posição do Brasil nesse ranking?
- c) Por que é importante conhecer também o número absoluto de mortes atribuíveis à poluição e não olhar apenas para a proporção de mortes atribuíveis à poluição?
- d) Faça uma pesquisa para obter informações sobre as principais causas de óbito no Brasil.

- 7 (UFPR 2017-adaptado) O Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil (CERT.br) é responsável por tratar incidentes de segurança em computadores e redes conectadas à Internet no Brasil. A tabela abaixo apresenta o número de mensagens não solicitadas (spams) notificadas ao CERT.br no ano de 2015, por trimestre.

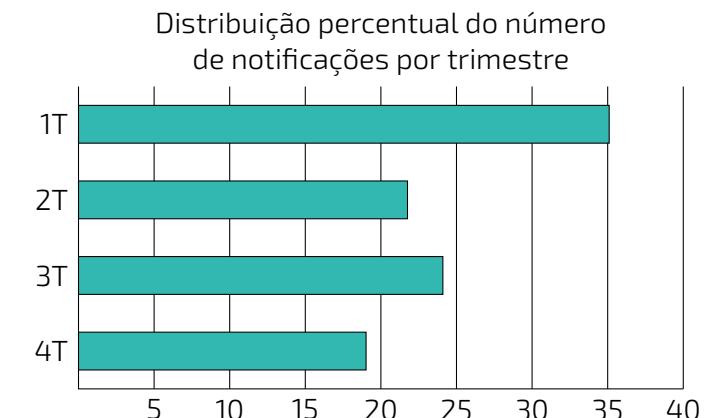
Trimestre	Notificações
4T	135.335
3T	171.523
2T	154.866
1T	249.743
total	711.467

Construa um gráfico para representar a distribuição do número de notificações por trimestre.

### Solução: Exercícios

- 6 a) Países com maior número absoluto de mortes atribuíveis à poluição em 2015, que é uma variável qualitativa nominal e foi organizada no gráfico em ordem decrescente de frequência. Frequência absoluta de casos. Décima primeira.
- b) A porcentagem de mortes atribuíveis em relação ao total de óbitos em 2015. De fato, cada óbito foi classificado em "atribuível à poluição" ou não (variável qualitativa) e em cada país calculou-se a porcentagem de óbitos atribuíveis à poluição. Trata-se de um gráfico de barras múltiplas, para comparar os diversos países em relação a essa porcentagem. A mudança radical de posição no Brasil se deve ao fato de que em relação ao total de óbitos, os atribuíveis à poluição correspondem a apenas 7,49%, não sendo nesse o caso mais comum. (Pesquise na internet sobre a distribuição de óbitos por causas no Brasil).
- c) O número absoluto é importante, por exemplo, para que seja possível fazer planejamento de alocação de recursos na saúde.

- 7 Existem diversas possibilidades. A figura apresentada aqui é um gráfico de barras em porcentagem com as barras na orientação horizontal.



**Solução: Exercícios****8**

- a) Falsa. As mulheres economicamente ativas eram cerca de 44 milhões e, os homens, cerca de 56 milhões, o que leva a concluir a população economicamente ativa de mulheres era cerca de 79% da população economicamente ativa de homens.
- b) Falsa. Para homens cresceu de cerca de 52 milhões para cerca de 58 milhões, o que dá um crescimento percentual relativo a 2003 de cerca de R. Para mulheres cresceu de cerca de 37,5 milhões para cerca de 47,5 milhões, o que dá um crescimento percentual relativo a 2003 de cerca de 27%.
- c) Falsa. Em 2005 eram cerca de 40 milhões e, em 2011, cerca de 45 milhões, o que dá um crescimento percentual relativo a 2005 de cerca de 12,5%.
- d) Verdadeira. Ver justificativa do item b).
- e) Falsa. Em 2007 eram cerca de 54 milhões e em 2015 cerca de 58 milhões, o que dá um crescimento percentual relativo a 2007 de cerca de 7%.

**8**

(UFRGS 2016 - adaptado) O gráfico a seguir representa a população economicamente ativa de homens e mulheres no Brasil de 2003 a 2015.

População economicamente ativa (em milhões)

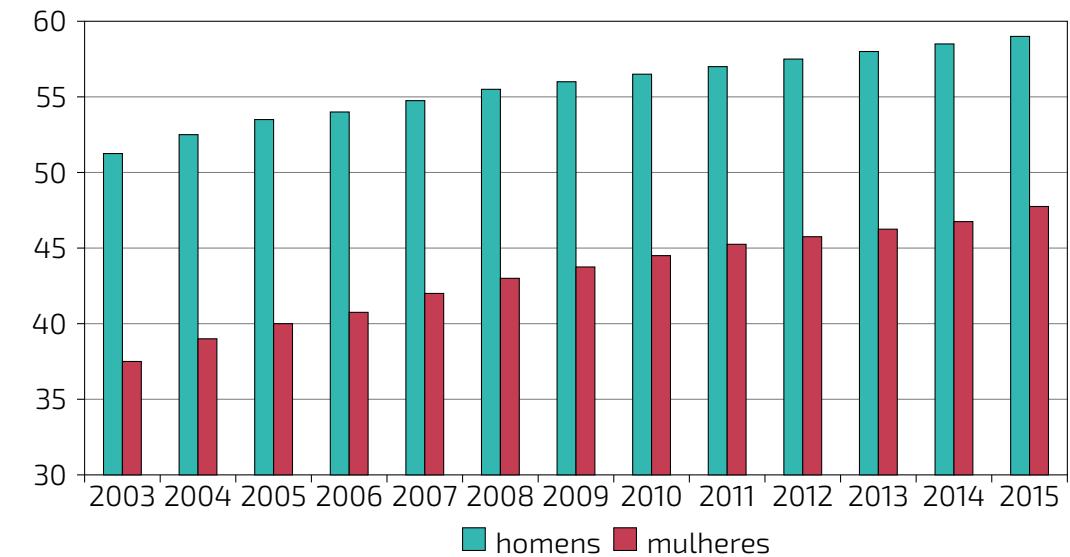
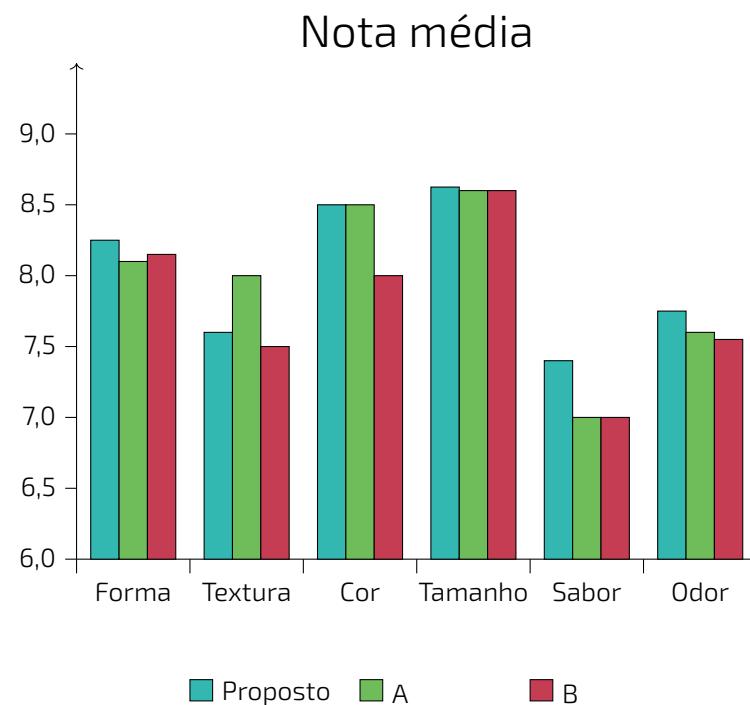


Figura 1.25: Fonte: Organização das Nações Unidas para Alimentação e agricultura

Classifique cada uma das afirmações a seguir em verdadeira ou falsa.

- a) No ano de 2009, a população economicamente ativa de mulheres era cerca de 50% da população economicamente ativa de homens.
- b) De 2003 a 2015, em termos percentuais, a população economicamente ativa de homens cresceu mais do que a de mulheres.
- c) Em relação a 2005, a população economicamente ativa de mulheres em 2011 cresceu cerca de 5%.
- d) De 2003 a 2015, em termos percentuais, a população economicamente ativa de mulheres cresceu mais do que a de homens.
- e) Em relação a 2007, a população economicamente ativa de homens em 2015 cresceu cerca de 3%.

- 9 (ENEM 2<sup>a</sup> aplicação 2016) A diretoria de uma empresa de alimentos resolve apresentar para seus acionistas uma proposta de novo produto. Nessa reunião, foram apresentadas as notas médias dadas por um grupo de consumidores que experimentaram o novo produto e dois produtos similares concorrentes (A e B).



A característica que dá a maior vantagem relativa ao produto proposto e que pode ser usada, pela diretoria, para incentivar a sua produção é a

- a) textura.
- b) cor.
- c) tamanho.
- d) sabor.
- e) odor.

**Solução: Exercícios**

- 9 Letra d). A maior vantagem relativa corresponde à maior diferença entre a nota do produto proposto e as notas dos produtos A e B de tal sorte que a nota do produto proposto seja maior do que as notas alcançadas por A e B. Desse modo, é fácil ver que a característica a ser escolhida é o sabor.

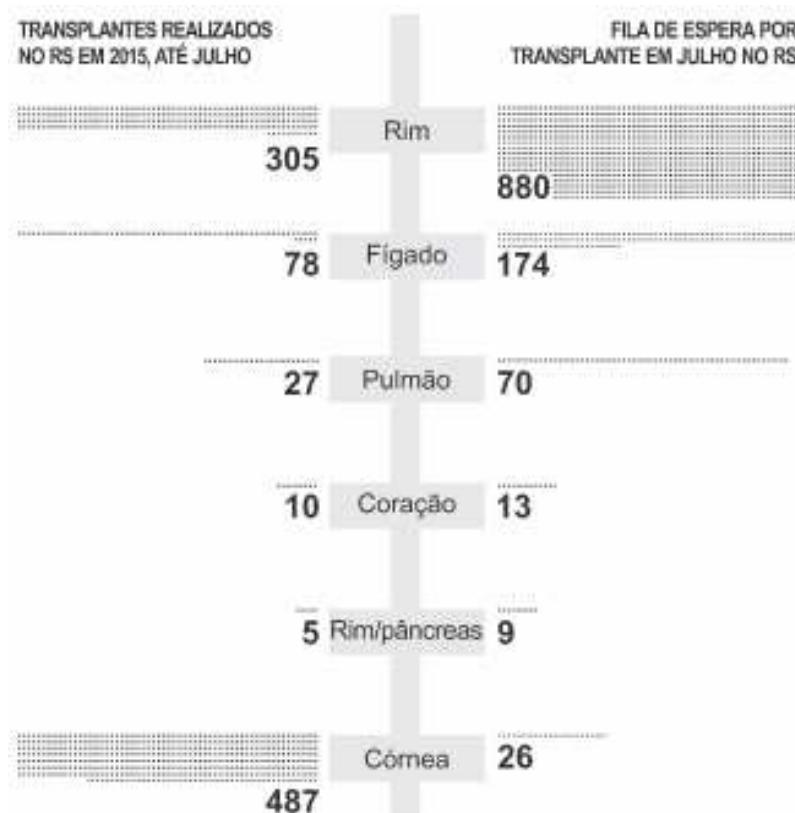
**Solução: Exercícios**

- 10 Considerando a tabela dos percentuais (valores relativos), a única afirmação correta é a da letra *a*.

Órgão	Transplantes	Fila de espera
Rim	33	75
Fígado	9	15
Pulmão	3	6
Coração	1	1
Rim/pâncreas	1	1
Córnea	53	2
Total	100	100

Tabela 1.1: Porcentagens dos números de transplantes até julho 2015 e das pessoas em fila de espera em julho de 2015 por órgão

- 10 (UFRGS 2016 - adaptado) Observe o gráfico a seguir.



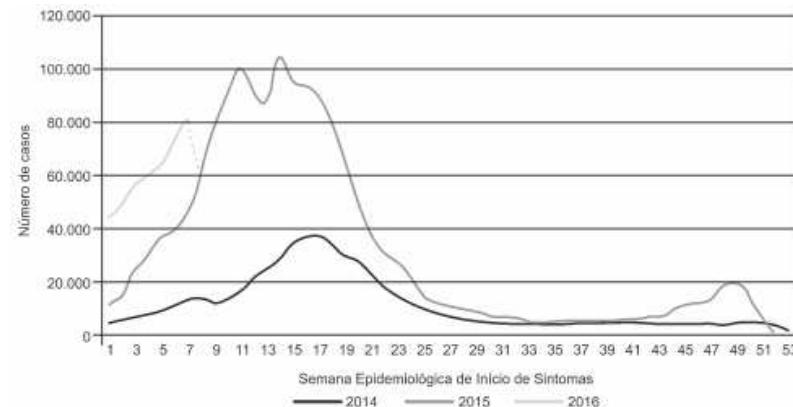
Fonte: Jornal Zero Hora

Nele está retratado o número de transplantes realizados no Rio Grande do Sul, até julho de 2015, e a quantidade de pessoas que aguardam na fila por um transplante no Estado, no mês de julho de 2015.

Com base no gráfico apresentado, classifique cada afirmação a seguir em verdadeira ou falsa.

- Mais da metade dos transplantes realizados até julho de 2015 foram transplantes de córnea.
- O percentual de pessoas que aguardavam transplante de pulmão em julho de 2015 correspondeu a 70% do total de pessoas na fila de espera por transplantes.
- O transplante de fígado é o que apresentou maior diferença percentual entre o número de transplantes realizados e o número de pessoas que aguardavam transplante.
- O número de transplantes de fígado realizados até julho de 2015 foi maior do que o número de transplantes de pulmão realizados no mesmo período.
- O transplante de córneas é o que tem a menor quantidade de pessoas aguardando transplante.

- 11 (UFPA 2016 - adaptado) O gráfico a seguir, retirado do Boletim Epidemiológico 16 de 2016 do Ministério da Saúde, registra os casos de dengue por semana, no Brasil, nos anos de 2014, 2015 e início de 2016.



Fonte: Sinan Online (atualizado em <sup>a</sup>13/07/2015; <sup>b</sup>04/01/2016; <sup>c</sup>07/03/2016). Dados sujeitos a alteração.

Figura – Casos prováveis de dengue, por semana epidemiológica de início de sintomas, Brasil, 2014<sup>a</sup>, 2015<sup>b</sup> e 2016<sup>c</sup>.

Com base no gráfico apresentado,

- o número de casos de dengue tem comportamento crescente próximo da vigésima segunda semana?
- os dados das 7 primeiras semanas de 2016 indicam uma diminuição do número de casos em relação a 2014 e 2015?
- no ano de 2015 houve mais de um milhão de casos?
- o maior número de casos ocorre em cada ano na décima quarta semana?
- em torno de que semana do ano 2016 é esperado o maior número de casos de dengue? Por que?

- 12 (ENEM 2016) O cultivo de uma flor rara só é viável se do mês do plantio para o mês subsequente o clima da região possuir as seguintes peculiaridades:

- a variação do nível de chuvas (pluviosidade), nesses meses, não for superior a 50 mm
- a temperatura mínima, nesses meses, for superior a 15°C;
- ocorrer, nesse período, um leve aumento não superior a 5 °C na temperatura máxima.

Um floricultor, pretendendo investir no plantio dessa flor em sua região, fez uma consulta a um meteorologista que lhe apresentou o gráfico com as condições previstas para os meses seguintes nessa região.

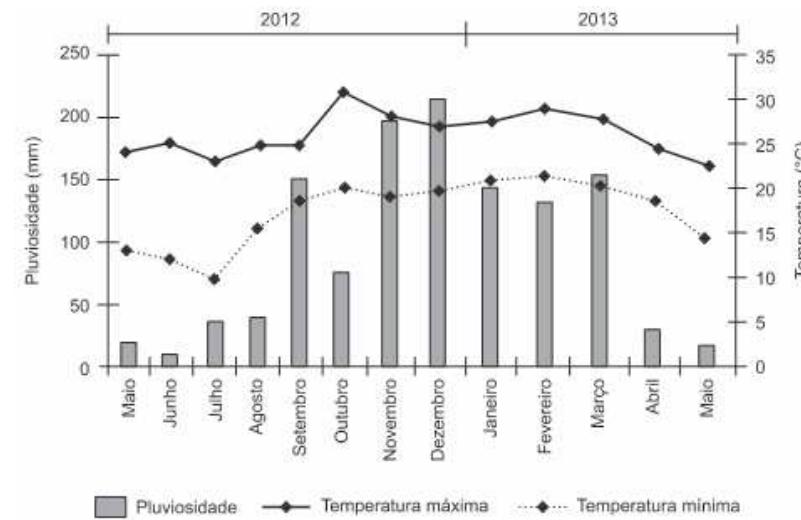
### Solução: Exercícios

- 11
- Não. Tanto em 2014 como em 2015 o comportamento é decrescente.
  - Não. O gráfico de 2016 está acima dos gráficos de 2014 e 2015 nas sete primeiras semanas.
  - Sim. Basta observar que entre as semanas 9 e 18 o número de casos foi maior do que ou igual a 80.000.
  - Não. Não há informações sobre o número de casos na décima quarta semana em 2016.
  - Entre a décima terceira e a décima sétima semana, pois nos anos anteriores, 2014 e 2015, foi entre essas semanas que ocorreu o maior número de casos.

### Solução: Exercícios

12 Letra a). O único mês que satisfaz todas as condições é janeiro. Com efeito,

- de fevereiro para março e de novembro para dezembro houve redução na temperatura máxima;
- a variação da pluviosidade de agosto para setembro e de dezembro para janeiro foi maior do que 50 mm.

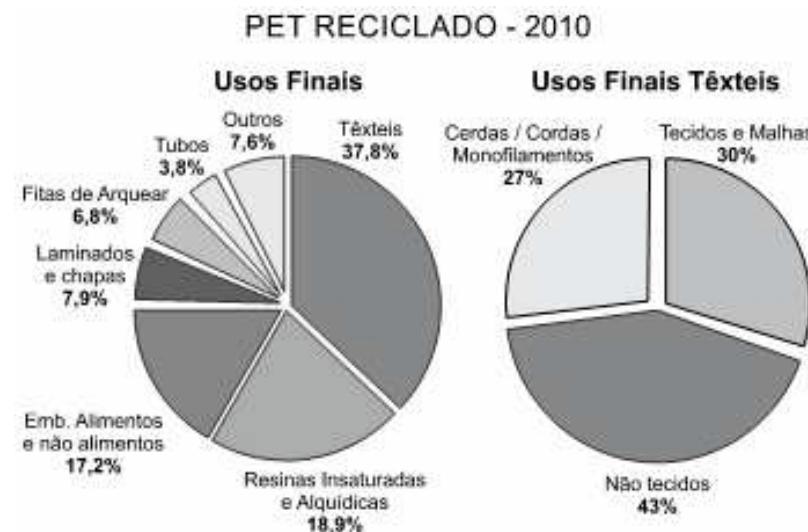


Com base nas informações do gráfico, o floricultor verificou que poderia plantar essa flor rara.

O mês escolhido para o plantio foi

- a) janeiro.
- b) fevereiro.
- c) agosto.
- d) novembro.
- e) dezembro.

13 (ENEM 2015) O polímero de PET (Politereftalato de Etileno) é um dos plásticos mais reciclados em todo o mundo devido à sua extensa gama de aplicações, entre elas, fibras têxteis, tapetes, embalagens, filmes e cordas. Os gráficos mostram o destino do PET reciclado no Brasil, sendo que, no ano de 2010, o total de PET reciclado foi de 282 kton (quilotoneladas).

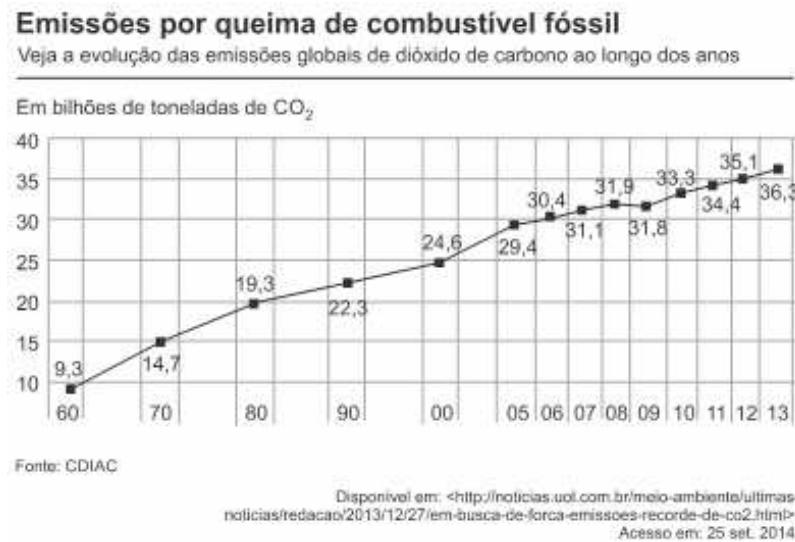


Disponível em: [www.abipet.org.br](http://www.abipet.org.br). Acesso em: 12 jul. 2012 (adaptado).

De acordo com os gráficos, a quantidade de embalagens PET recicladas destinadas a produção de tecidos e malhas, em kton é mais aproximada de

- a) 16,0
- b) 22,9
- c) 32,0
- d) 84,6
- e) 106,6

- 14 (UFRGS 2015 - adaptado) O gráfico abaixo apresenta a evolução da emissão de Dióxido de carbono ao longo dos anos.



Com base no gráfico apresentado, classifique cada afirmação a seguir em verdadeira ou falsa.

- a) Ao longo do período, a emissão de dióxido de carbono apresentou taxa de variação constante.
- b) Em relação aos anos 80, os anos 90 apresentaram emissão de dióxido de carbono 30% maior.
- c) O ano de 2009 apresentou menor valor de emissão de dióxido de carbono da primeira década do século XXI.
- d) De 2000 a 2013, houve crescimento percentual de 11,7% na emissão de dióxido de carbono.
- e) Em relação a 2000, o ano de 2013 apresentou emissão de dióxido de carbono aproximadamente 50% maior.

### Solução: Exercícios

- 13 Letra **c**). Sendo de 37,8% a porcentagem do total de PET reciclado para uso final têxtil, e de 30% dessa quantidade para tecidos e malhas, segue que a resposta é dada por  $\frac{37,8}{100} \cdot \frac{30}{100} \cdot 282 \approx 32,0\text{kton}$

- a) Falsa, houve períodos de crescimento e períodos de decaimento, portanto, com taxas de variação diferentes.
- b) Falsa, pois 22,3 – 19,3 não representam 30% de 19,3.
- c) Falsa, pois em 2005 e 2006 as emissões foram inferiores à emissão em 2008.
- d) Falsa, pois 36,3 – 24,6 = 11,7, aproximadamente 50%.
- e) Verdadeira, pois 36,3 – 24,6 = 11,7, aproximadamente 50% de 24,6.

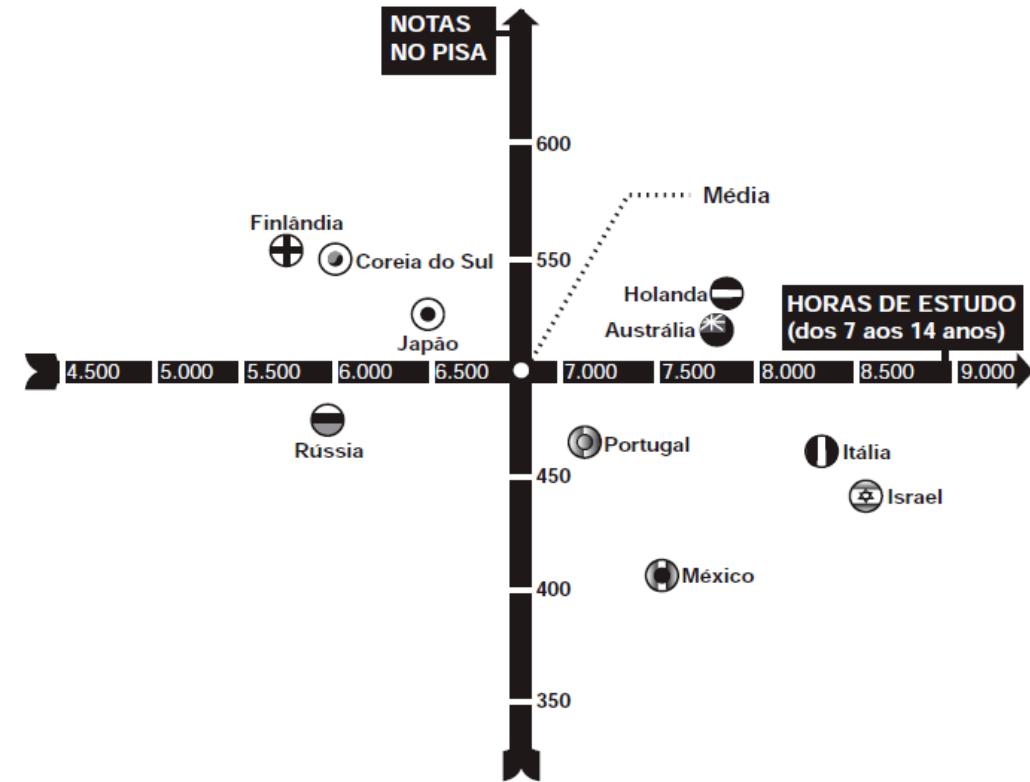
**Solução: Exercícios**

- 15 Letra **c**). Os países com notas abaixo da média de 24,6 são: Rússia, Portugal, México, Itália e Israel. Dentre esses países, o que apresenta maior quantidade de horas de estudo é Israel.

15

(ENEM 2013) Uma falsa relação

O cruzamento da quantidade de horas estudadas com o desempenho no Programa Internacional de Avaliação de Estudantes (Pisa) mostra que mais tempo na escola não é garantia de nota acima da média.

**NOTAS NO PISA E CARGA HORÁRIA (PAÍSES SELECIONADOS)\***

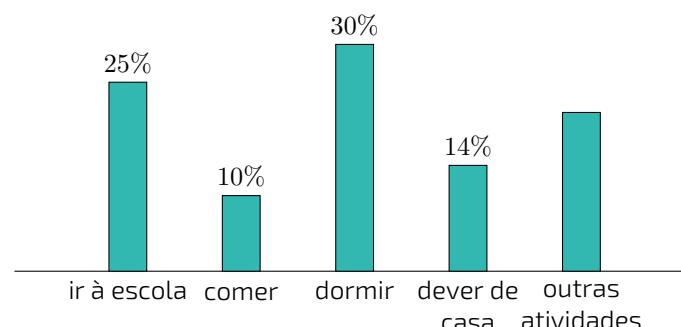
\* Considerando as médias de cada país no exame de matemática.

*Nova Escola, São Paulo, dez. 2010 (adaptado).*

Dos países com notas abaixo da média nesse exame, aquele que apresenta maior quantidade de horas de estudo é

- Finlândia.
- Holanda.
- Israel.
- México.
- Rússia.

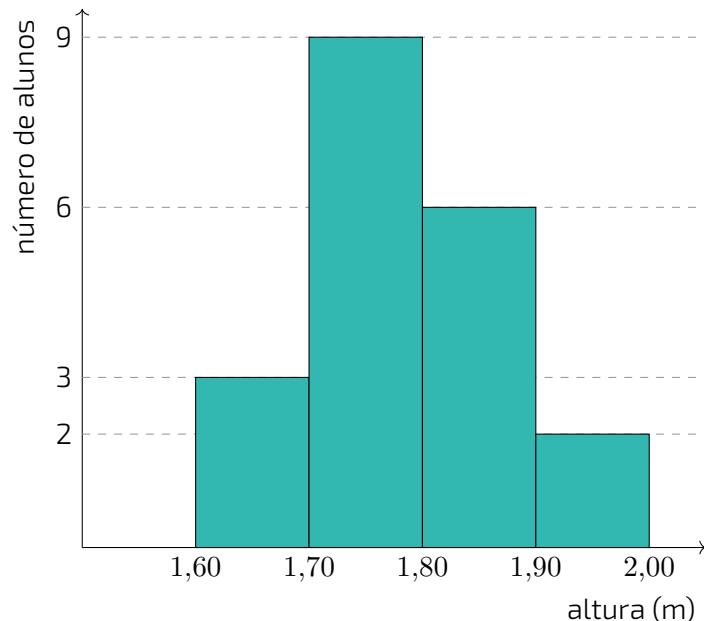
- 16 (UF-AM) O gráfico a seguir mostra quanto tempo um estudante gasta com suas atividades durante o dia.



A quantidade de horas gasta pelo estudante com outras atividades em um dia é de:

- a) 2,25h
- b) 3,02h
- c) 3,57h
- d) 5,04h
- e) 6,70h

- 17 (UERJ-adaptada) Após serem medidas as alturas dos alunos de uma turma, elaborou-se o seguinte histograma:



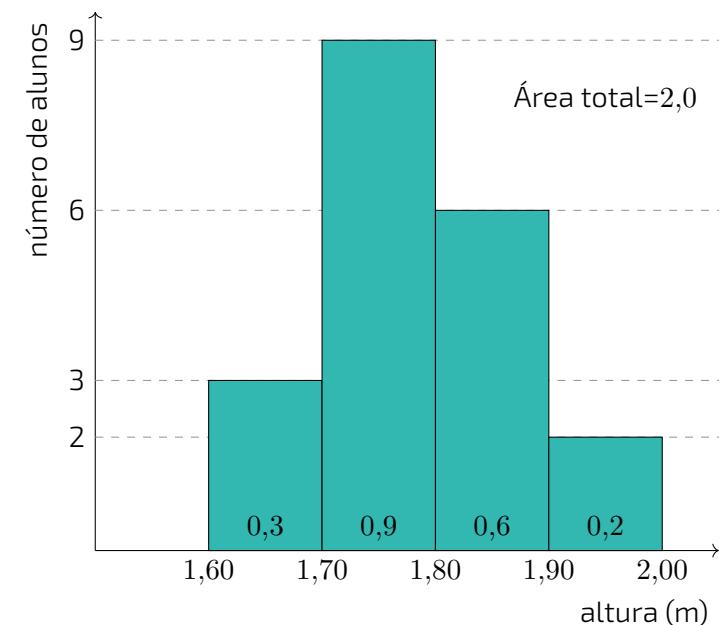
Em um histograma, se uma reta vertical de equação  $x = x_0$  divide o histograma em duas partes de mesma área, então o valor de  $x_0$  corresponde à mediana da distribuição representada no histograma. Calcule a mediana das alturas dos alunos com base no histograma apresentado.

### Solução: Exercícios

- 16 A porcentagem de horas do dia com outras atividades é dada por  $100 - (25 + 10 + 30 + 14) = 21$ . 2% de 24h é dado por  $\frac{21}{100} \cdot 24 = 5,04h$ . A opção correta é a letra d).

- 17 Observe que o histograma apresentado é composto por quatro retângulos cujas bases medem 0,1. Assim a área total do histograma é dada por  $0,1 \cdot (3 + 9 + 6 + 2) = 2,0$ , ou seja, soma das áreas dos 4 retângulos que compõem o histograma. Assim, a metade da área corresponde ao valor 1,0.

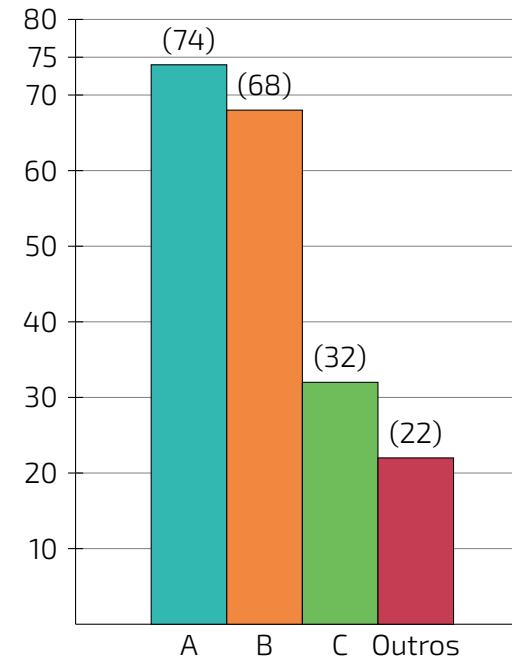
Considerando os dois primeiros retângulos, a área é  $0,1 \cdot 12 = 1,2$  que supera a metade da área total. Isso significa que a mediana será um valor que está entre 1,70 e 1,80. Considerando o primeiro retângulo, falta para completar 1 um sub-retângulo do segundo, com área igual a 0,7, ou seja,  $(x_0 - 1,7) \cdot 9 = 0,7$  tal que  $x_0 = 1,7 + 0,79 \approx 1,78m$ .



## Solução: Exercícios

- 18 De fato, pelo comprimento das barras, a barra correspondente ao tipo A tem comprimento maior do que o dobro dos comprimentos somados das barras correspondentes ao tipo C e outros. O gráfico não está correto, pois não respeita a escala de frequências: usa o mesmo tamanho para representar 10 unidades de 0 a 60, para representar apenas uma unidade entre 71 e 75.

Gráfico adequado:



- 18 Numa pesquisa sobre a preferência dos jovens por sucos, obteve-se, entre os tipos principais A, B e C, o seguinte resultado.

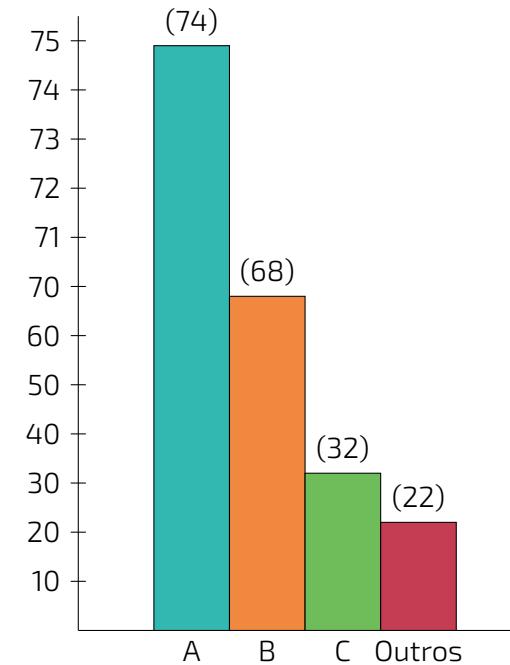
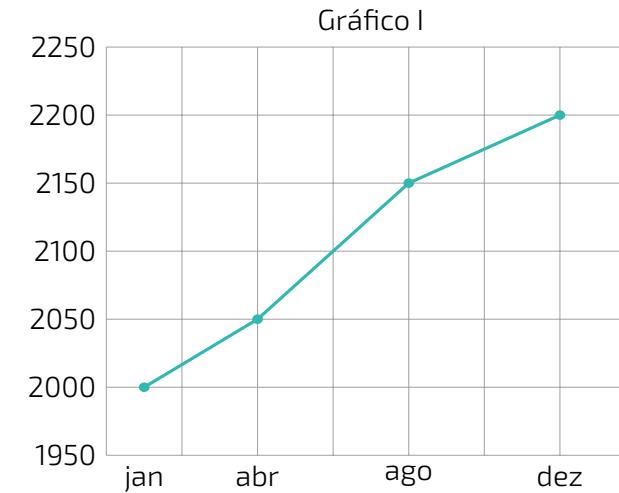
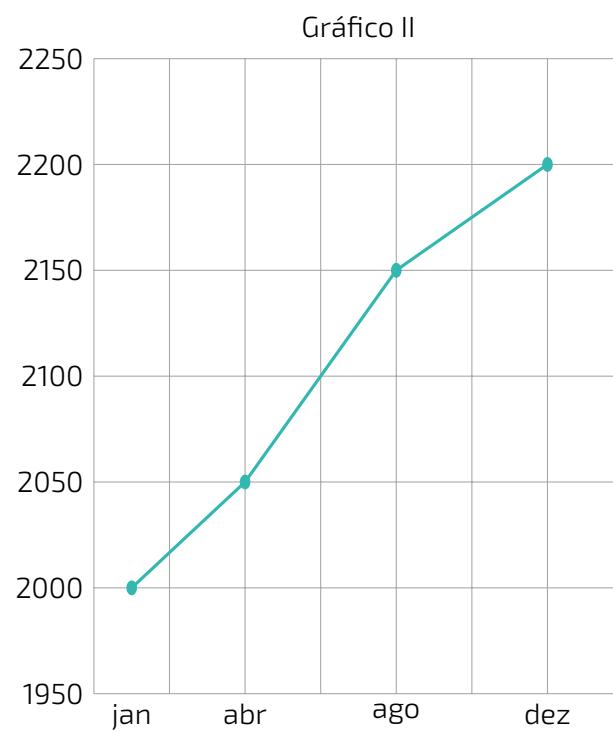


Figura 1.26: O número entre parênteses corresponde ao número de respostas para cada tipo de suco entre os jovens selecionados na pesquisa

- Olhando o gráfico é razoável dizer que a preferência pelo tipo A é maior que o dobro das preferências somadas pelo tipo C e outros tipos? Por que?
- Refaça o gráfico de barras.

- 19 (ENEM) Para convencer a população local da ineficiência da Companhia Telefônica Vilatel na expansão de oferta de linhas, um político publicou no jornal local o gráfico I, representado a seguir. A Companhia Vilatel respondeu dias depois publicando o gráfico II, com o qual pretende justificar um grande aumento na oferta de linhas. O fato é que, no período considerado, foram instaladas, efetivamente, 200 linhas telefônicas novas.





Analizando os gráficos, pode-se concluir que:

- o gráfico II apresenta um crescimento real maior do que o gráfico I.
- o gráfico I apresenta um crescimento real, sendo o gráfico II incorreto.
- o gráfico II apresenta o crescimento real, sendo o gráfico I incorreto.
- a aparente diferença de crescimento nos dois gráficos decorre da escolha de escalas diferentes.
- os dois gráficos são incomparáveis, pois usam escalas diferentes.

20 As fichas dos 800 funcionários de uma empresa estão catalogadas por ordem alfabética no setor de recursos humanos. Como você faria para obter uma amostra sistemática de tamanho 50 dos funcionários dessa empresa?

### Solução: Exercícios

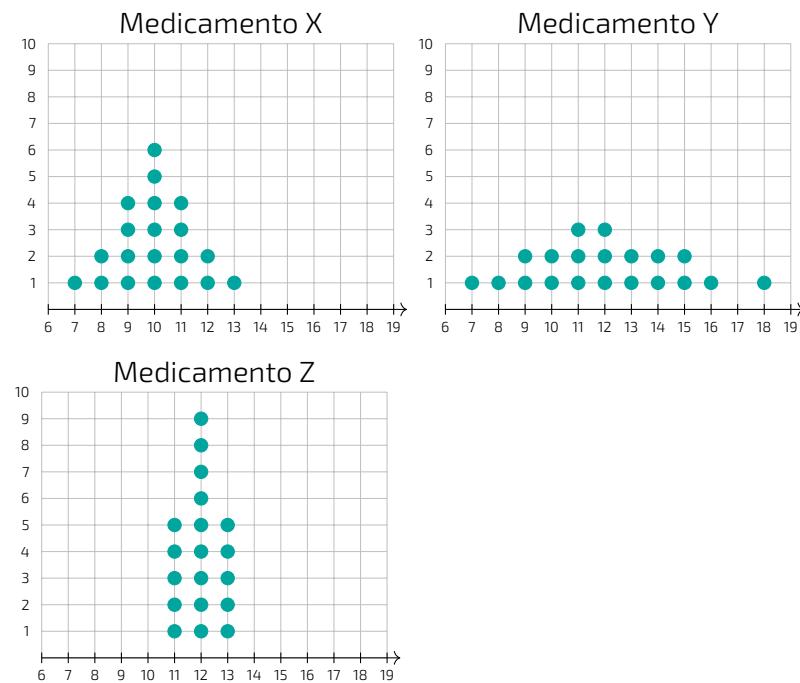
19 A opção correta é a letra **d**: a inclinação maior no gráfico II comparada ao gráfico I deve-se a escolha de escalas distintas. No gráfico I a amplitude do intervalo no eixo vertical para 50 unidades é um pouco menor do que a amplitude correspondente utilizada no gráfico II. As outras opções estão incorretas.

20 Considere a organização dos 800 funcionários em ordem alfabética: a cada funcionário corresponde uma posição de 1 a 800. Temos que  $\frac{800}{5} = 16$ . Sorteie ao acaso um número entre 1 e 16, por exemplo 5. Agora considere os 50 primeiros termos de uma Progressão Aritmética de primeiro termo 5 e razão 16: 5, 21, 37, ..., 789. Selecione então os 50 funcionários correspondentes a essas posições na listagem em ordem alfabética.

# Notas

## 1 Solução: Comparação de medicamentos

a)



b) Analisando os diagramas de pontos, percebe-se que o medicamento Y foi o que apresentou maior dispersão dos tempos de cura. Observe que é a distribuição que apresentou a maior amplitude.

c) De acordo com as somas informadas na tabela, as médias observadas de tempo de cura foram 10 minutos para o medicamento X; 12 minutos para o medicamento Y e 12 minutos para o medicamento Z.

d) Comparando os diagramas de pontos:

i) Observa-se que o medicamento X apresenta uma média amostral (10 min) inferior à do medicamento Y (12 min), porém existe uma interseção razoável, quando analisamos as duas distribuições empíricas dos tempos de cura para esses medicamentos. Isso potencialmente indicaria não existir uma diferença significativa entre os tempos médios de cura desses dois medicamentos. Uma forma de reforçar essa conclusão seria coletar mais dados para cada um dos medicamentos e observar se reproduzem o mesmo padrão observado na análise inicial.

ii) Quando analisamos as distribuições empíricas dos tempos de cura dos medicamentos X e Z, observamos que o medicamento X apresenta uma média amostral inferior à do medicamento Z. Neste caso, a interseção das duas distribuições é pequena. Além disso, todas as 20 medições do tempo de cura de Z são maiores do que a média de X. Nesta comparação, os dados se revelam mais favoráveis à escolha do medicamento X.

iii) Observa-se que ambos medicamentos apresentam a mesma média amostral, porém dispersões diferentes. Assim, esses dados favorecem o medicamento Z, que apresenta menor dispersão em torno do tempo médio de cura.

iv) Como já foi discutido, apenas os medicamentos X e Z apresentam uma diferença clara. No entanto, para uma conclusão mais confiável seria conveniente coletar mais informações.

2

## Sugestões e discussões

### Infográfico 2

Os itens f) e g) têm como objetivo estudar possíveis associações entre duas variáveis qualitativas, a saber, grau de instrução e prática de atividade física a) e rendimento per capita e prática de atividade física b). Observe que, embora rendimento seja uma variável quantitativa, quando ele é representado por intervalos de rendimento, se torna variável qualitativa. Novamente é importante destacar, nessa discussão, que o conjunto inteiro foi subdividido em subconjuntos ditados pelas categorias, grau de instrução ou faixas de rendimento, e que para cada subconjunto calculou-se a porcentagem de pessoas que praticam atividade física. Usar frequências absolutas não seria útil para comparar os diferentes grupos quando eles têm tamanhos diferentes.

3

### Prática de Atividade Física fora do horário escolar

Pesquisa na turma

\*Obrigatório

Gênero \*

Feminino

Masculino

Idade (em anos completos) \*

13 ou menos

14

15

16

17

18

19 ou mais

Que tipo de atividade física você pratica fora do horário escolar?  
Assinale apenas a principal.

Futebol

Caminhada

Vôlei

Natação

Lutas/Artes Marciais

Dança

Ginástica

Skate

Não pratica atividade física fora do horário escolar.

Outro: \_\_\_\_\_

Qual a frequência que você pratica atividade física?

até duas vezes por semana

três vezes por semana

mais de três vezes por semana

Qual é a duração média (em minutos) da atividade que você pratica?

30 minutos

mais de 30 até 60 minutos

mais de 60 até 90 minutos

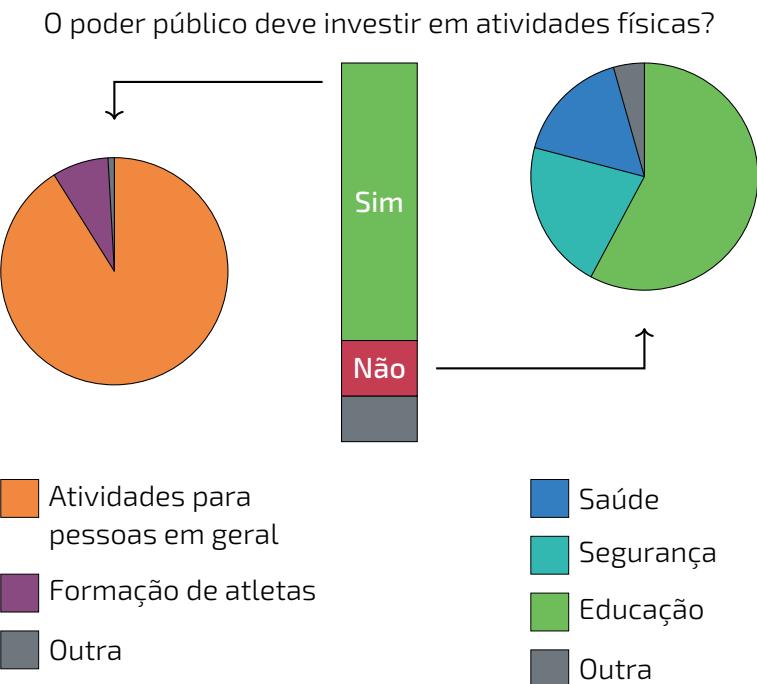
mais de 90 minutos

Figura 1.27: Sugestão de questionário

#### 4 Solução: Classificação de variáveis

- a) quantitativa.      e) qualitativa binária.      i) qualitativa binária.  
 b) quantitativa.      f) qualitativa.      j) quantitativa.  
 c) quantitativa.      g) quantitativa.      k) qualitativa binária.  
 d) quantitativa.      h) qualitativa.      l) qualitativa.

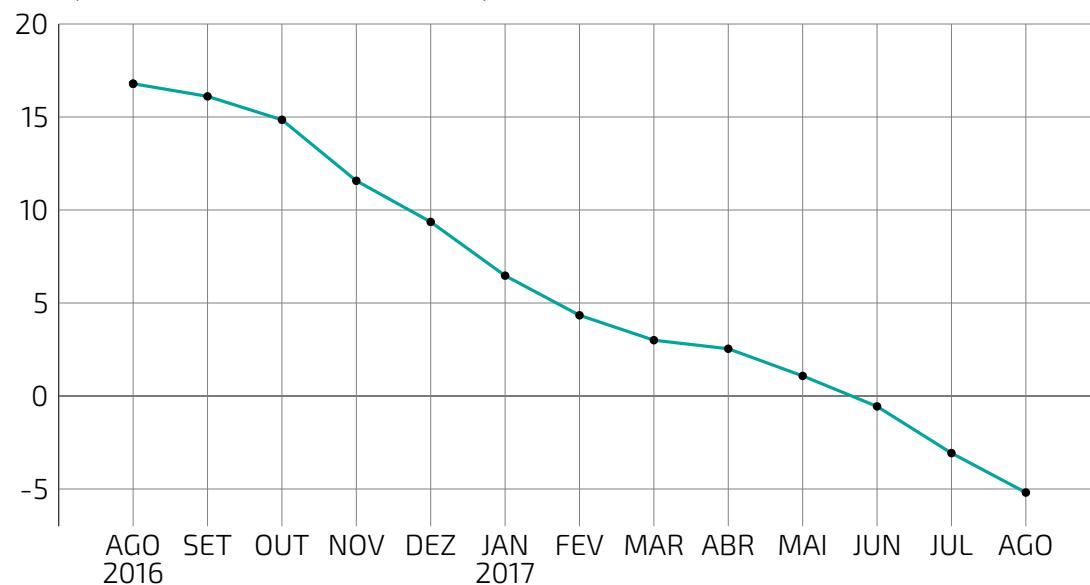
#### 5 Solução: Construção de gráficos para variáveis qualitativas



#### 6 Solução: Análise de gráfico

- a) É um gráfico que usa barras, mas nesse gráfico o comprimento das barras não é frequência.  
 b) Valor da inflação da alimentação acumulada nos últimos 12 meses. Esses valores são apresentados em função do período de tempo: agosto de 2016 até agosto de 2017.  
 c) O valor da inflação da alimentação acumulada nos últimos 12 meses é uma variável quantitativa, o período de tempo representado em meses/ano é uma variável qualitativa ordinal.  
 d) (gráfico) Como evoluiu a inflação da alimentação acumulada em 12 meses no período investigado, a saber, agosto de 2016 até agosto de 2017. Mais precisamente, percebe-se que a inflação da alimentação acumulada em 12 meses apresentou no período analisado uma forte tendência de queda.

INFLAÇÃO DA ALIMENTAÇÃO NO DOMICÍLIO  
(acumulado em 12 meses, em %)



#### 7 Solução: Construção de histogramas

Existem várias possibilidades e algumas delas estão apresentadas aqui. Na comparação é importante perceber que esses dados revelam uma estrutura simétrica, ocorrendo com frequências altas entre 4 e 6, e, ocorrendo com frequências bem menores nos intervalos extremos inferior e superior.

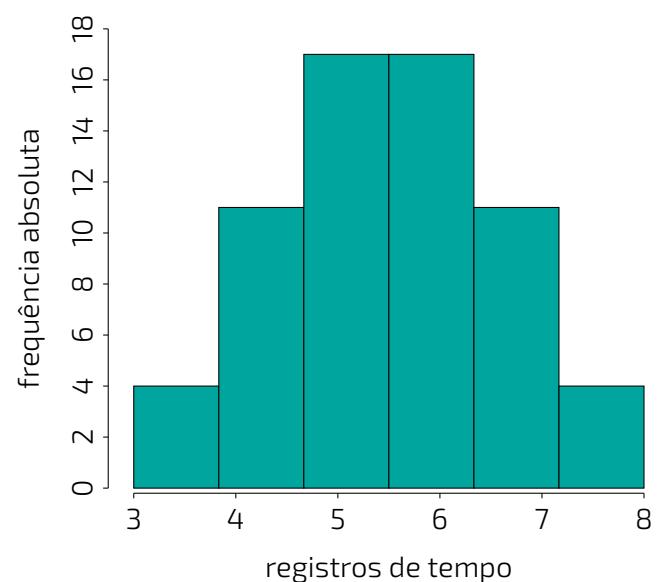


Figura 1.28: Histograma com 6 intervalos

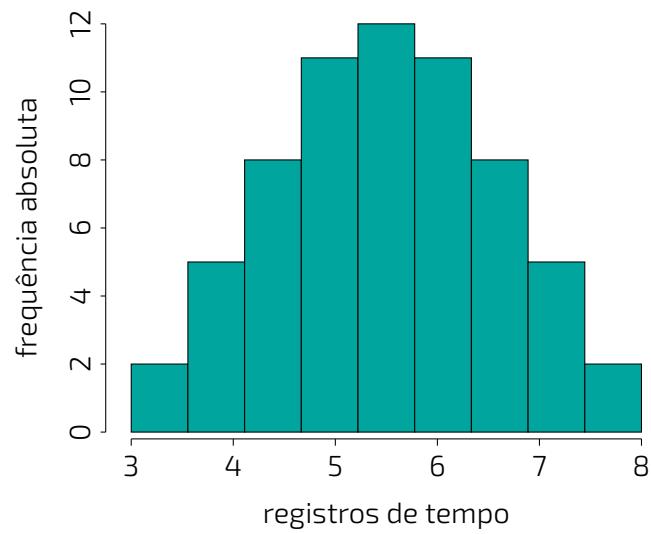


Figura 1.29: Histograma com 9 intervalos

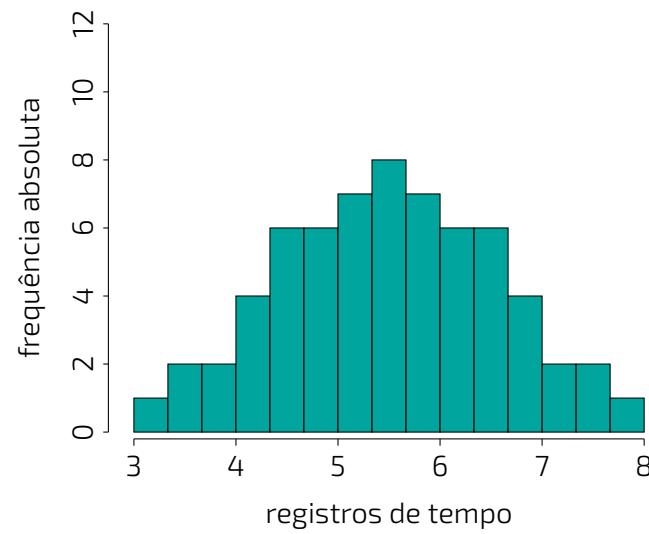


Figura 1.31: Histograma com 15 intervalos

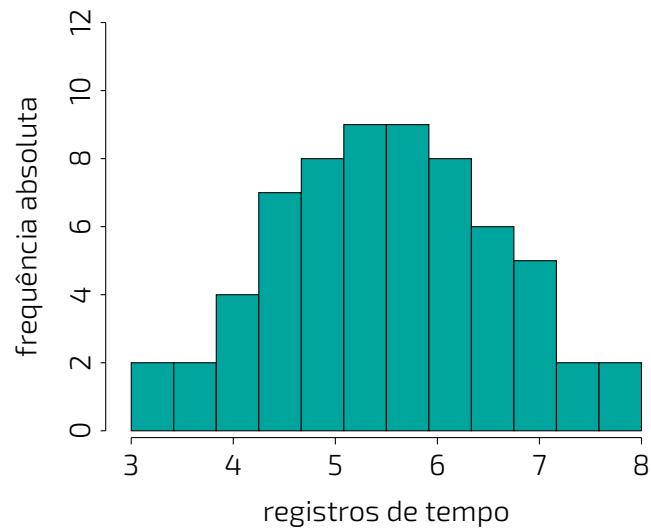


Figura 1.30: Histograma com 12 intervalos

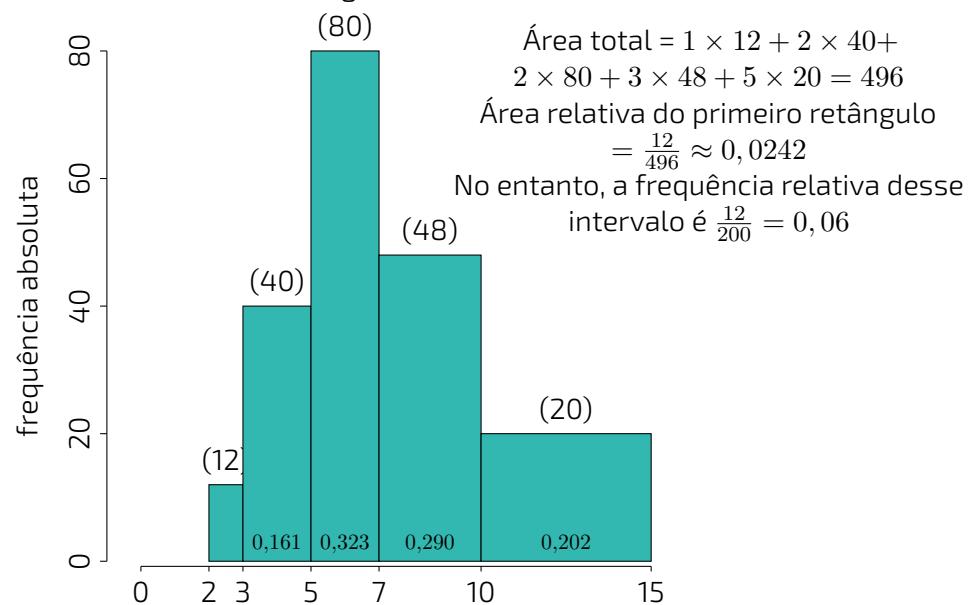
8 Solução: Histogramas com intervalos de amplitudes desiguais

a)

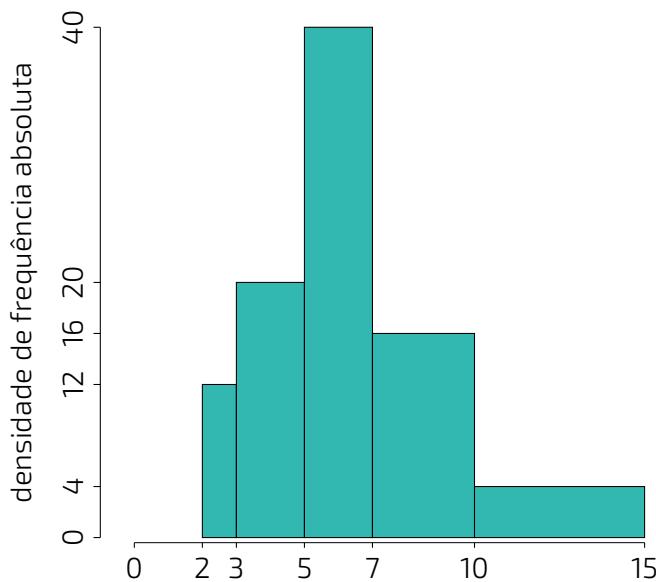
Intervalo de classe	Frequência absoluta	Amplitude	Dens. de freq. absoluta
[2,0; 3,0[	12	1	12
[3,0; 5,0[	40	2	20
[5,0; 7,0[	80	2	40
[7,0; 10,0[	48	3	16
[10,0; 15,0[	20	5	4

b)

Histograma errado



Histograma correto



## Referências Bibliográficas

- Batanero, C. e Borovcnik, M. (2016). *Statistics and Probability in High School*. Springer, Nova Iorque, Estados Unidos.
- Batanero, C., Burrill, G. e Reading, C. (2011). *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education*. A Joint ICMI/IASE Study: The 18th ICMI Study. Springer, Nova Iorque, Estados Unidos.
- Bussab, W. O. e Morettin, P. A. (2017). *Estatística Básica*. Saraiva, 9ª edição.
- Cordani, L. K. (2002). Estatística para todos. Oficina. Disponível em: <http://www.bienasbm.ufba.br/OF14.pdf>. Acesso em 24 de fevereiro de 2021.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M. e Scheaffer, R. (2007). Guidelines for assessment and instruction in statistics education (gaise) report: A pre-k-12 curriculum framework). *Amstat News*.
- IBGE (2017). Vamos contar.
- Louzada, F., Cordani, L., Bazan, J. e Barbosa, T. S. (2015). Reflexões a respeito dos conteúdos de probabilidade e estatística na escola no brasil - uma proposta. Relatório técnico, Associação Brasileira de Estatística, São Paulo, Brasil.
- Pfenning, N. (2011). *Elementary Statistics: Looking at the Big Picture*. Cengage Learning, Massachusetts, Estados Unidos, 1ª edição.
- Rossmann, A. J. e Chance, B. L. (1998). *Workshop Statistics: Discovery With Data*. Springer, Nova Iorque, Estados Unidos, 1ª edição.
- Veaux, R. D. D. e Velleman, P. F. (2008). Math is music; statistics is literature (or, why are there no six-year-old novelists?). *Amstat News*, páginas 54-57.