

UNIVERSIDADE FEDERAL DE ALFENAS  
INSTITUTO DE CIÊNCIAS EXATAS

---

# **Estatística Básica**

**Luiz Alberto Beijo**

---

ALFENAS  
Minas Gerais - Brasil  
23 de setembro de 2022

## DADOS DO MATERIAL: Notas de Aula de Estatística Básica

Este material tem por objetivo propiciar aos estudantes um roteiro para o curso de Estatística Básica. Ressalta-se que o texto aqui construído é apenas uma coleção das notas de aulas do professor e que havendo necessidade, as bibliografias referenciadas devem ser sempre consultadas.

Por se tratar de um texto em construção, erros podem ocorrer, por isso solicito, a gentileza, que sugestões e correções sejam encaminhadas para *luiz.beijo@unifal-mg.edu.br*.

BONS ESTUDOS!!!!

Elaboração:

*Luiz Alberto Beijo*

Licenciado em Ciências e Matemática pela PUC Minas. Mestre e Doutor em Estatística e Experimentação Agropecuária pela UFLA. Professor do Departamento de Estatística da Universidade Federal de Alfenas - UNIFAL-MG

## Sumário

<b>Lista de Tabelas</b>	<b>4</b>
<b>1 A ESTATÍSTICA NA CIÊNCIA</b>	<b>5</b>
1.1 A cientificidade do conhecimento . . . . .	5
1.2 Diferenças da atitude científica frente ao senso comum . . . . .	8
1.3 Conceito de método . . . . .	9
1.4 Reflexão . . . . .	9
1.5 Estatística . . . . .	10
1.5.1 Fases do trabalho estatístico . . . . .	12
1.6 Fluxo da Estatística Descritiva . . . . .	13
1.6.1 Coleta dos dados . . . . .	13
1.7 Conceitos básicos . . . . .	13
1.8 Uso do <i>Software R</i> . . . . .	14
1.9 Programa R . . . . .	15
<b>2 AMOSTRAGEM</b>	<b>16</b>
2.1 População . . . . .	16
2.1.1 Unidade Amostral . . . . .	16
2.1.2 População objetivo . . . . .	17
2.1.3 Característica . . . . .	17
2.2 Métodos de pesquisa descritiva (observacional) . . . . .	17
2.3 Amostragem . . . . .	18
2.4 Amostragem probabilística . . . . .	19
2.4.1 Amostragem simples ao acaso (ASA) ou aleatória simples . . . . .	19
2.4.2 Amostragem estratificada . . . . .	20
2.4.3 Amostragem Sistemática . . . . .	20
2.4.4 Amostragem por conglomerados . . . . .	21
2.5 Amostragem não probabilística . . . . .	22
2.6 Problemas de amostragem . . . . .	23
2.6.1 Amostras tendenciosas . . . . .	23
2.7 Exercícios . . . . .	24
<b>3 ESTATÍSTICA DESCRITIVA</b>	<b>27</b>
3.1 Conceitos e definições importantes . . . . .	27
3.2 Introdução . . . . .	28
3.3 Apresentação dos dados qualitativos . . . . .	29
3.4 Apresentação dos dados quantitativos discretos . . . . .	32
3.5 Organização e apresentação de dados quantitativos Contínuos . . . . .	34
3.5.1 Construção de uma tabela de distribuição de frequência (TDF) . . . . .	34
3.5.2 Representação gráfica de dados quantitativos contínuos . . . . .	36
3.5.3 Classificação das distribuições de frequências . . . . .	36
3.6 Exercícios . . . . .	38

<b>4</b>	<b>Medidas descritivas</b>	<b>39</b>
4.1	Medidas de posição . . . . .	39
4.1.1	Propriedades da média, mediana e moda . . . . .	40
4.1.2	EXEMPLO . . . . .	40
4.2	Medidas de variabilidade . . . . .	41
4.2.1	EXEMPLO . . . . .	43
4.3	Medidas de separatriz (quantil) . . . . .	44
4.4	Gráfico Boxplot ou Diagrama de caixas . . . . .	45
4.4.1	EXEMPLO . . . . .	46
4.5	Exercícios . . . . .	47

## Lista de Tabelas

1	Peso de 50 cobaias . . . . .	25
2	Classificação dos trabalhadores do ramo farmacêutico brasileiro, segundo a renda mensal, em 2020 . . . . .	29
3	Preferência pelo tipo de carne em uma amostra de 100 alunos da Escola "X". . . . .	29

## 1 A ESTATÍSTICA NA CIÊNCIA

- **Comunicado:** Segundo orientações via OFÍCIO CONJUNTO CIRCULAR Nº 1-2020-ProGrad-Reitoria-UNIFAL-MG.

"Este material é de uso exclusivo para estudo nesta disciplina e não pode ser compartilhado fora do ambiente desta turma virtual. Baixar os textos aqui para estudar em casa não viola a lei dos direitos autorais. Entretanto, do mesmo modo que você não pode pegar o livro emprestado na biblioteca e fazer cópia para distribuir para outras pessoas, aqui no ambiente virtual de aprendizagem é igual. Portanto, não compartilhem os textos, vídeos e/ou podcasts fora do ambiente da turma, para evitar problemas com a lei do direito autoral."

Antes de abordarmos os conceitos fundamentais de Estatística, é importante situarmos a *Estatística* dentro do contexto do **método científico**. Para isto precisamos nos fundamentar em alguns conceitos e definições que são importantes tanto para pesquisas científicas quanto para o processo de gestão.

### 1.1 A cientificidade do conhecimento

Segundo Oliveira (2005)<sup>i</sup>, o conhecimento está ligado visceralmente à qualidade de vida do ser humano, qualquer que seja a sua forma, quer seja ele do tipo místico-religioso, filosófico-científico ou estético-artístico. Conhecer está no centro de uma vida plena e abundante (sucesso). Observa-se que a nossa sociedade contemporânea é uma sociedade do conhecimento, tal é o avanço dos conhecimentos neste tempo.

Contrastamos conhecimento com saber, para definirmos uma diferença sutil entre estes dois conceitos.

- **Conhecimento** é a forma que o ser humano faz a representação (ou compreensão, entendimento, ou ainda uma interpretação) atribuindo significado à realidade. Esta realidade, composta por sistemas, fenômenos e/ou processos, é dissecada sempre segundo três aspectos desejáveis: *verdade, beleza e justiça*.
- **Conhecer** é, portanto, a consciência da natureza e da medida destes três aspectos no fenômeno em foco. O conhecimento permite ao homem se mover corretamente num mundo marcado por bases *verdadeiras, belas e justas*, as quais ele discerne. Conhecimento é, pois a coincidência, é o encontro, é a confirmação entre o que é a realidade e o que o homem entende que deva ser a sua representação do mundo e da vida. O conhecimento que nós temos é que dá sentido, ou significado, à realidade.

A finalidade do conhecimento é **a predição ou previsão** dos efeitos que serão observados a partir da consciência das causas, o que permite o domínio das circunstâncias e, finalmente, a capacidade de gestão da vida e do mundo.

*A gestão deve produzir em última instância o bem-estar e a felicidade coletiva e individual.*

- **Saber** é o conhecimento potencializado pela arte de conseguir aplicá-lo (arte é a capacidade de fazer acontecer).

---

<sup>i</sup> OLIVEIRA, M. S. **Orientações metodológicas para a construção de monografias**. Texto Acadêmico da Editora da UFLA: DEX-UFLA, 2005.

Enquanto o conhecimento pode existir apenas na mente de uma pessoa, *a sabedoria* necessariamente traduz-se em *palavras, gestos, ações, intervenções, com resultados concretos e observáveis*. O conhecedor é, portanto, aquele que percebe as coisas, conseguindo separar o que é belo, verdadeiro e justo do que não o é. O sábio porém, é aquele que **age** de modo belo, verdadeiro, e justo. O objetivo último **da educação é produzir sábios**.

Estes três modos de conhecer e de saber estão sempre presentes no ser humano, em qualquer situação. Porém, quando a **faculdade racional** está no comando das posturas e ações, destaca-se a forma de conhecimento chamada **científica**, ou, resumidamente, **Ciência**. A Ciência visa alcançar o **conhecimento provavelmente seguro e certo**, a fim de trazer o bem-estar e a felicidade nas experiências vividas.

Um importante resultado da conhecimento científico é a tecnologia.

- **Tecnologia** é a sabedoria científica, já que esta é a aplicação da ciência na construção artefatos (produtos, equipamento, métodos e modelos) que operam no mundo real externo ao ser humano (em alguns casos podemos denominar tecnologia como metodologia). Em outra definição, tecnologia é técnica com embasamento científico.

Assim como não podemos saber sem conhecer, não podemos gerar tecnologia sem ciência. Oliveira (2005) afirma que, para que um dado conhecimento seja considerado científico, é necessário que ele seja funcional, no sentido definido em seguida. A funcionalidade de um conhecimento consiste em:

1. **Organicidade (ou Sistemático)**: o conhecimento deve ser um sistema (ou uma organização) de proposições constituindo uma teoria. A Ciência é um conhecimento organizado e orgânico, incluindo nisto os conceitos de objetividade, consistência, coerência, relações de causa-e-efeito, e racionalidade.
2. **Instrumentalidade (ou Instrumental)**: a teoria científica deve servir como um instrumento para que o ser humano encontre e mantenha-se num estado de bem-estar e felicidade, resolvendo seus problemas. A Ciência é, então, um conhecimento instrumental, que visa aplicações "que deem certo".
3. **Intersubjetividade**: a teoria deve receber um julgamento com pronunciamento consensual e aprovação de uma comunidade de cientistas (pessoas que fazem ciência), isto é, de um corpo de pesquisadores e professores de determinada área da Ciência.

A Ciência é, portanto, um conhecimento compartilhável, transferível, ensinável, julgável e criticável. Como afirmado, a sistematização do conhecimento, preconizada no item 1, é o que denominamos como *teoria*.

- Teoria é a representação sistemática que o conhecimento científico faz do mundo e da vida. Das teorias é que derivamos os modelos e as hipóteses sobre o modo como funcionam o mundo e a vida.
- Modelos são representações organizadas de um dado fenômeno ou processo observado na realidade, com elementos constituintes justificados pela teoria utilizada, que faz pano-de-fundo ao modelo que está sendo proposto.
- Hipóteses são afirmações pontuais (não organizadas) sobre fenômenos ou processos reais, que providenciam uma resposta antecipada à um dado problema de pesquisa. Geralmente, precisam ser validadas.

Uma hipótese deve ser atrelada a uma teoria e a um modelo. Estes modelos e hipóteses definem a **gestão científica**, a qual é a utilização destes instrumentos conceituais derivados das teorias, para intervenção sobre os fenômenos, as organizações, os processos naturais e humanos, enfim, sobre todas as coisas que nos interessam.

As teorias, os modelos e as hipóteses recebem a homologação de uma comunidade legitimada para isto, reunidas em universidades e institutos de pesquisa, que se utilizam de defesas de documentos científicos perante bancas de pares, congressos científicos, simpósios, workshops, seminários, revistas científicas, entre outros, para proceder esta homologação. Tais eventos são os *eventos de comunicação e validação científica*, ou simplesmente **eventos científicos**.

A **pesquisa científica** é o modo como buscamos construir o conhecimento científico. Ela compreende *encontrar, selecionar, estruturar e resolver problemas* que interessam ao ser humano. Para apreender, representar e compreender problemas, o pesquisador lançará mão de hipóteses, modelos e teorias de modo crucial. Em outras palavras, antes de coletar dados, é necessário que nós tenhamos em mente um **referencial teórico**.

Conhecimento teórico é pré-requisito para se fazer uma pesquisa com qualidade, enquanto problemas são o ponto de partida para a pesquisa, percepção e a estruturação adequada destes problemas (que dependem essencialmente da postura filosófica e científica do pesquisador, ou grupo de pesquisa) são decisivos para o desenvolvimento da pesquisa.

O produto essencial de uma pesquisa científica é, portanto, um conhecimento que pretende ser seguro e certo (o conhecimento científico), apto para operar soluções para um problema humanamente relevante. A pesquisa científica deve ser conduzida segundo o método científico (veja Figura 1).

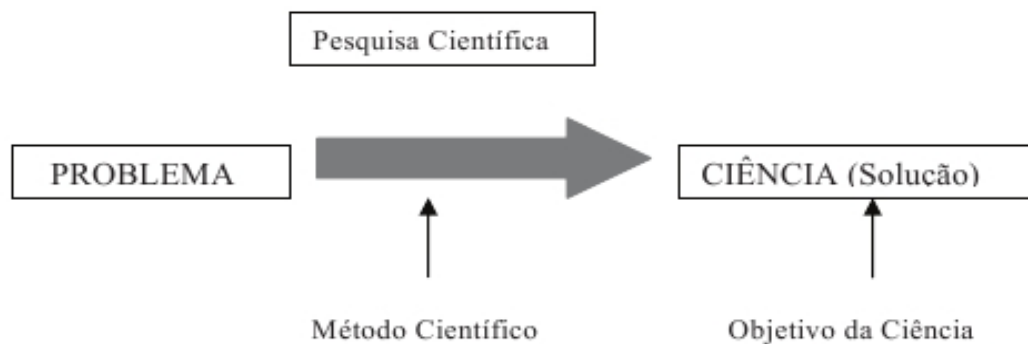


Figura 1: Pesquisa científica.

O **método científico** é o modo como nós devemos *entender logicamente* e como *devemos conduzir* todo o processo da pesquisa científica. Para entender por que precisamos de um método para guiar o pensamento e as práticas de pesquisa científica, vejamos como o conhecimento científico desenvolve-se.

As fases do ciclo do conhecimento científico, do ponto de vista da lógica da construção do conhecimento, podem ser explicitadas como:

1. Apropriação da realidade, registrando-se os fatos interpretados.
2. Sistematização dos fatos interpretados criando-se a teoria, os modelos e as hipóteses.
3. Verificação da concordância da teoria com os fatos.
4. Exposição destes fatos interpretados sistematizados e explicados pela teoria, disponibilizando-os para a conferência e absorção de outros.

Do ponto de vista dos artefatos redigidos, podemos dizer que a essência do método científico seria:

- i. Um problema verbalizado, que inicializa o processo da pesquisa.



ii. A partir daí é feita uma revisão de literatura, definindo um referencial teórico para a pesquisa; este referencial é adicionado a todo o ser do cientista, que parte então para o processo da pesquisa em si, estudando o mundo e a vida, num movimento de mão-dupla. Neste momento o pesquisador estabelece uma estratégia, *um projeto, um delineamento* para o processo da pesquisa (de acordo com tipo).

iii. Finalmente, o resultado da investigação é apresentado em uma obra científica (monografia, dissertação, tese, artigo, relatório, etc), que é escrita para mostrar a solução científica do problema.

**OBSERVAÇÃO:** Um profissional, com formação no Ensino Superior, mesmo que não vá realizar pesquisa científica, **com certeza**, ele utilizará conhecimentos advindos de pesquisas científicas. Então é muito importante que ele tenha conhecimentos básicos para compreender e criticar o resultado da investigação científica, de forma que possa utilizar ou aplicar aquele conhecimento ou tecnologia desenvolvida a partir dele.

Do ponto de vista das etapas de trabalho, apresenta esta essência como:

1. definição do problema;
2. delineamento ou projeto de pesquisa (ideias, possíveis soluções, elaboração das hipóteses);
3. obtenção dos dados (para certificação se as hipóteses são verdadeiras ou falsas);
4. análise dos dados (Descritiva, Inferencial, Modelagem);
5. conclusão e redação.

Define-se um **delineamento (ou projeto) de pesquisa** como a lógica que liga os dados, informações e fatos coletados (e as conclusões construídas) às questões iniciais, ao problema da pesquisa. Seria o método científico particularizado para uma pesquisa científica em especial. Todo estudo científico tem um projeto de pesquisa, explícito ou não.

## 1.2 Diferenças da atitude científica frente ao senso comum

Em uma formação para o "pensador em pesquisa" ou gestor, seja na graduação ou pós-graduação, é necessário que se diferencie o conhecimento científico do senso comum. Na filosofia, o *senso comum* (ou, conhecimento vulgar) é a primeira suposta compreensão do mundo resultante da herança fecunda de um grupo social e das experiências atuais que continuam sendo efetuadas. O senso comum descreve as crenças e proposições que aparecem como normal, sem depender de uma investigação detalhada para alcançar verdades mais profundas como as científicas. São características do senso comum:

1. opiniões diferentes;
2. ausência de critérios para avaliação;
3. generalizam opiniões;
4. baseiam-se, somente, na crença;
5. não questionam nem duvidam e não se admiram com nada no mundo.

Por outro lado, a atitude científica pode ser retratada como a atitude daquele que busca o conhecimento de *forma metódica e sistemática*, ou seja, baseado em argumentos racionais que por meio de experiências empíricas atestam a provável "verdade" sobre algo no mundo. São características da atitude científica:

1. separam elementos objetivos dos subjetivos;
2. elaboram teorias racionais;
3. demonstram as teorias com experimentos práticos;
4. não consideram os conhecimentos definitivos, mas abertos a mudanças;
5. são metódicos e rigorosos.

### 1.3 Conceito de método

A preocupação em descobrir e, portanto, explicar a natureza vem desde os primórdios da humanidade, quando as duas principais questões referiam-se às forças da natureza, a cuja mercê viviam os homens.

No século XVI se iniciou uma linha de pensamento que propunha encontrar um conhecimento embasado em maiores garantias, na procura do real. Não se buscam mais as causas absolutas ou a natureza íntima das coisas; ao contrário, procura-se compreender as relações entre elas, assim como a explicação dos acontecimentos, através da observação científica aliada ao raciocínio.

Com o passar do tempo, muitas modificações foram feitas nos métodos existentes, inclusive surgiram outros novos. No momento, o que nos interessa é o conceito moderno de método (independente do tipo). Para tal, consideramos como Bunge, *que o método científico é a teoria da investigação. Esta alcança seus objetivos, de forma científica, quando cumpre ou se propõe a cumprir suas etapas.*

Todas as ciências caracterizam-se pela utilização de métodos científicos; em contrapartida, nem todos os ramos de estudo que empregam estes métodos são ciências. Dessas afirmações podemos concluir que a utilização de métodos científicos não é da alçada exclusiva da ciência, mas sem o emprego de métodos científicos, não há ciência. Assim, o método é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo - conhecimentos válidos e verdadeiros (provavelmente), traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista.

Mais detalhes sobre os métodos científicos podem ser encontrados no livro Fundamentos de metodologia científica, de autoria de Marina de A. Marconi e Eva M. Lakatos. - 5. ed. - São Paulo : Atlas 2003.

### 1.4 Reflexão

- Porque o método é utilizado?

Para fornecer um suporte metodológico e representacional (sistemático) ao pensamento, permitindo o uso de ferramentas objetivas que permitam a superação das limitações individuais do pesquisador em suas análises e sínteses.

- O que é pesquisa?

É o processo através do qual as pessoas adquirem um novo conhecimento sobre si mesmas ou sobre o mundo em que vivem, com a finalidade de: Responder a um questionamento, resolver um problema, realizar a gestão de processos (órgãos, ou produtos), satisfazer uma necessidade.

O planejamento de uma pesquisa depende do tema do problema a ser estudado, da sua natureza e situação em que se encontra, área de atuação e nível de conhecimento do pesquisador. Isso significa que pode haver vários tipos de pesquisa em função dos objetivos a serem alcançados. Objetivos teóricos e/ou práticos.

- Tipos de pesquisa quanto a natureza. A Ciência apresenta uma diferenciação em relação a forma investigativa.

**Pesquisa básica:** consiste na aquisição do conhecimento sobre a natureza sem finalidades práticas ou imediatas.

**Pesquisa aplicada:** Consiste na utilização do conhecimento da pesquisa básica e da tecnologia para se obter aplicações práticas como produtos ou processos.

A pesquisa aplicada (tecnológica) tem como objetivo alcançar a inovação em um produto ou processo, frente a uma demanda ou necessidade preestabelecida.

Classificação da pesquisa:

#### 1. Quanto aos objetivos

- i. Pesquisa exploratória: é a coleta de dados e informações sobre um fenômeno (ou processo) de interesse sem grande teorização sobre o assunto, inspirando ou sugerindo uma hipótese explicativa;
- ii. Pesquisa Hipotético-Dedutiva de conhecimento: é a elaboração de uma ou mais hipóteses que relacione diversos fatos (variáveis), seguida da coleta de dados e da geração de informações que comprove ou não tal hipótese ou hipóteses.

#### 2. Quanto a intervenção

- i. Estudo observacional: o pesquisador anota os dados e extrai informações, mas faz o possível para não influenciar na ocorrência do fenômeno;
- ii. Estudo experimental: o pesquisador interfere e influencia na pesquisa, objetivando verificar os efeitos da intervenção.

#### 3. Classificação quanto ao tempo

- i. Estudo transversal: O pesquisador coleta os dados do experimento num único instante no tempo, obtendo um recorte momentâneo do fenômeno investigado;
- ii. Estudo longitudinal: o pesquisador coleta os dados do experimento em dois ou mais momentos, havendo um acompanhamento ao longo do tempo do fenômeno em estudo.

## 1.5 Estatística

**Estatística** é uma *Ciência* que fornece métodos para coleta, organização, descrição, análise e interpretação de dados (observacionais ou experimentais) e para a utilização dos mesmos na tomada de decisões.

**Bioestatística:** é a Estatística aplicada às ciências que estudam aspectos vitais (referentes à vida).

É importante destacar que a aplicação da Estatística não ocorre somente para a coleta e análise de dados na geração de conhecimento na Ciência, mas é largamente aplicada na Gestão (de empresas; órgãos; processos e produtos), como suporte na "tomada de Decisão".

A utilização de técnicas, destinadas à análise de situações complexas ou não, tem aumentado e faz parte do nosso cotidiano. O que tem levado a essa qualificação de nossas vidas no dia a dia?

Um fator importante é a popularização dos computadores. No passado, tratar uma grande massa de números era uma tarefa custosa e cansativa, que exigia horas de trabalho tedioso. Recentemente, no entanto, grandes quantidades de informações podem ser analisadas rapidamente com um computador pessoal e programas adequados.

Dessa forma o computador contribui, positivamente, na difusão e uso de métodos estatísticos. Por outro lado, o computador possibilita uma automação que pode levar um indivíduo sem preparo específico a utilizar técnicas inadequadas para resolver um dado problema.

**Assim, é necessária a compreensão dos conceitos básicos da Estatística, bem como as suposições necessárias para o seu uso de forma criteriosa.**

Podemos, de forma geral, dividir a Estatística em três áreas:

- Estatística Descritiva;
- Probabilidade;
- Inferência Estatística.

A **Estatística Descritiva** pode ser definida como um conjunto de técnicas destinadas a descrever e resumir dados, a fim de que possamos explorar as características de interesse. Em geral, utilizamos a Estatística Descritiva na etapa inicial da análise quando tomamos contato com os dados pela primeira vez.

Objetivando tirar conclusões de modo informal e direto, a maneira mais simples seria a observação dos valores colhidos. Entretanto ao depararmos com uma grande massa de dados percebemos, imediatamente, que a tarefa pode *não ser simples*. Para tentar retirar dos dados informações a respeito do fenômeno sob estudo, é preciso **aplicar algumas técnicas** que nos permitam **simplificar** a informação daquele particular conjunto de valores. A finalidade da Estatística Descritiva é tornar as *coisas mais fáceis de entender, de relatar e discutir*.

A **Probabilidade** pode ser pensada como a teoria matemática utilizada para estudar a **incerteza** oriunda de fenômenos que *envolvem o acaso* (**fenômenos aleatórios, ou experimentos aleatórios**). A decisão de um fabricante de empreender uma grande campanha de propaganda visando a aumentar sua participação no mercado, a decisão de parar de imunizar pessoas contra determinada doença, a decisão de arriscar-se a atravessar uma rua no meio do quarteirão, todas utilizam a probabilidade consciente ou inconscientemente.

**Inferência Estatística** é o estudo de técnicas que possibilitam a **extrapolação** (inferir), a um grande conjunto de dados (população), das informações e conclusões obtidas a partir de *subconjuntos* de valores, usualmente de dimensões muito menores (Amostras).

Estudos complexos que envolvem o tratamento estatístico dos dados, usualmente incluem as três áreas citadas acima, principalmente em trabalhos científicos e na gestão para tomada de decisão. Esta é a parte na qual estão incluídos os chamados **testes estatísticos**.

A ciência, de uma forma geral, objetiva a coleta de informações (fatos) na natureza e a formulação de modelos (ou hipóteses) que expliquem parte dos fenômenos ou permitam a sua previsão. Esse processo é baseado em critérios bem definidos pelo método científico, pelo qual, hipóteses formuladas são verificadas

posteriormente, com a coleta e interpretação de dados. Porém, esta metodologia é amplamente aplicada, além do Meio Científico.

As etapas da metodologia científica podem ser resumidas nos seguintes passos:

1. O problema a ser solucionado (Motivação);
2. Busca na teoria para fundamentar ideias, possíveis soluções;
3. Elaboração das hipóteses;
4. Verificação se as hipóteses são verdadeiras ou falsas;

*Nesta etapa, faz-se necessário a utilização de dados, que podem ser observacionais ou experimentais.*

5. Análise dos Resultados;
6. Comparação com as teorias existentes e Conclusões.

### 1.5.1 Fases do trabalho estatístico

Pela definição de Estatística apresentada anteriormente, podemos verificar a importância desta no método científico, chegando alguns pesquisadores a afirmarem que a Estatística é a "alma" do método científico.

Podemos descrever o trabalho estatístico em cinco etapas básicas:

1. Planejamento de como coletar os dados;
2. Coleta e crítica de dados;
3. Tratamento e apresentação dos dados;
4. Análise e interpretação dos resultados;
5. Conclusão.

O planejamento para a coleta dos dados está fortemente relacionado com os objetivos da pesquisa ou estudo, que, por sua vez, tem uma relação direta com o "Problema a ser solucionado". Portanto, qualquer "erro" em alguma das etapas do trabalho estatístico, levará a resultados equivocados e consequentemente a conclusões errôneas.

Para entendermos melhor as teorias envolvidas no trabalho estatístico, precisamos entender algumas definições e conceitos básicos.

## 1.6 Fluxo da Estatística Descritiva

A Estatística descritiva pode ser resumida de acordo com o diagrama apresentado na Figura 2:

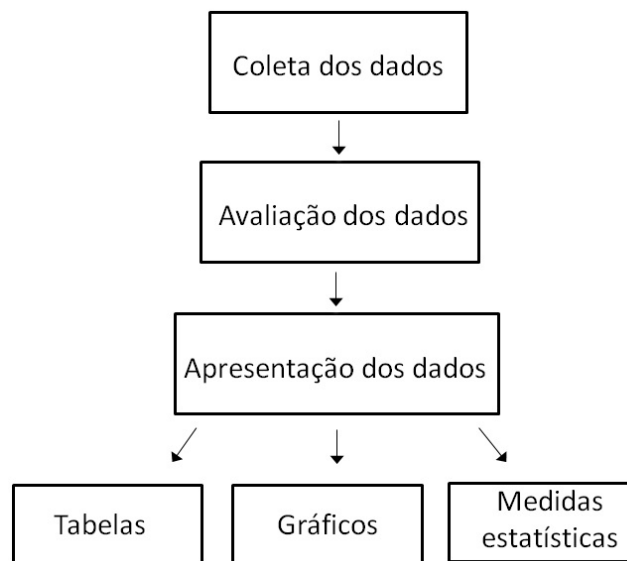


Figura 2: Diagrama resumo da Estatística Descritiva.

### 1.6.1 Coleta dos dados

Com a definição do problema/objetivo a ser estudado e o estabelecimento do planejamento da pesquisa (maneira pela qual os dados serão coletados; cronograma das atividades; custos envolvidos, etc.), o próximo passo é a coleta dos dados. A coleta dos dados consiste na busca ou compilação dos dados, variáveis ou componentes do fenômeno a ser estudado.

A coleta desses dados, pode ser realizada de modo direto ou indireto. Em que, a coleta direta ocorre quando os dados são obtidos diretamente da fonte originária, como no caso da avaliação da pressão arterial de um indivíduo ou índice de glicemia. A coleta é indireta quando é inferida a partir dos elementos conseguidos pela coleta direta ou utiliza-se prontuários.

Após a coleta dos dados faz-se necessário uma avaliação dos mesmos, na qual pode-se observar a presença de informações que possam ter sido coletadas/digitadas/anotadas erroneamente e, com isso, distorcer as análises e consequentemente os resultados.

Os dados não devem ser coletados de um modo qualquer, são necessárias técnicas adequadas para a obtenção dos mesmos. Tais técnicas são denominadas por técnicas de amostragem e são apresentadas no Capítulo 2.

Posterior a coleta e a avaliação dos dados faz-se necessário a sua apresentação, a qual pode ser realizada por meio de tabelas, gráficos ou das medidas estatísticas, descritas no Capítulo 3.

## 1.7 Conceitos básicos

1. Estatística: é uma Ciência que estuda e desenvolve métodos para planejar experimentos, coleta de dados, organizá-los, resumi-lo, analisá-los, interpretá-los e deles extrair conclusões. Podendo ser dividida em:

$$\text{Estatística} \Rightarrow \begin{cases} \text{Dedutiva (descritiva)} \\ \text{Indutiva (inferencial)} \end{cases}$$

- (a) A estatística dedutiva (descritiva): utilizada para descrever o conjunto de dados desde a elaboração da pesquisa até o cálculo de determinada medida.
  - (b) A estatística Indutiva (inferencial): relaciona-se com a incerteza, iniciando-se com o cálculo das Probabilidades e se desenvolve por todo a área da inferência.
2. **População:** conjunto de elementos (indivíduos ou objetos) que apresentam pelo menos uma característica em comum.
  3. **Amostra:** subconjunto finito e representativo da população.
  4. **Censo:** coleção de dados relativos a todos os elementos da população.
  5. **Variável aleatória:** pode ser definida como uma característica qualquer de interesse que associamos a uma população ou amostra a ser estudada. Estas podem ser classificadas como:
  6. Variáveis  $\Rightarrow \left\{ \begin{array}{l} \text{Qualitativa} \Rightarrow \left\{ \begin{array}{l} \text{Ordinal} \\ \text{Nomial} \end{array} \right. \\ \text{Quantitativas} \Rightarrow \left\{ \begin{array}{l} \text{Discreta} \\ \text{Contínua} \end{array} \right. \end{array} \right.$
  7. **Tipos de variáveis:** as variáveis podem ser:
    - a) **Variável qualitativa:** são aquelas variáveis que correspondem a atributos ou categorias, podem ser:
      - a.1) **Variável qualitativa nominal:** quando os atributos (ou categorias) **não são** possíveis de ordenação.  
**Exemplo:** Cor dos olhos dos alunos de uma turma.
      - a.2) **Variável qualitativa ordinal:** quando os atributos (ou categorias) **são** possíveis de ordenação.  
**Exemplo:** nível de conhecimento de inglês dos alunos de uma turma (baixo, médio, alto).
    - b) **Variável quantitativa:** são aquelas variáveis que correspondem a números resultantes de contagem ou medidas, podem ser:
      - b.1) **Variável quantitativa discreta:** são próprias de dados de contagem, isto é, estão definidas em um conjunto enumerável.  
**Exemplo:** número de pessoas de um determinado grupo que já sofreram algum tipo de fratura.
      - b.2) **Variável quantitativa contínua:** são aquelas em que as realizações resultam de uma medida (uma mensuração) que podem assumir qualquer valor real entre dois extremos. (em um intervalo real).  
**Exemplo:** peso e altura dos alunos de uma turma.
  8. **Dado ou observação:** é o valor (ou atributo) que descreve uma característica da amostra.
  9. **Parâmetro:** é a medida numérica que descreve uma característica da população.

## 1.8 Programa R

Nesse curso de Estatística Básica será utilizado o *Software R*

O R é uma linguagem orientada a objetos criada em 1996 por **Ross Ihaka** e **Robert Gentleman**, do Departamento de Estatística da Universidade de Auckland na Nova Zelândia, como uma alternativa ao software S-Plus, o qual exigia o pagamento de uma licença para seu uso.



Figura 3: \*

Fonte: <https://www.r-project.org/>

O programa R é de livre distribuição, foi inspirado no mesmo ambiente do S e permite a manipulação de dados, realização de cálculos e geração de gráficos. Seu uso tem aumentado nos últimos anos e usuários de todo o mundo têm contribuído com o desenvolvimento de novos recursos para o R.

Para instalar o programa R, siga os passos:

- (1) Entrar no site <https://www.r-project.org/>
- (2) Clicar no menu CRAN e escolher o espelho (mirror) Brasil;
- (3) Você verá que existem vários *mirrors* no Brasil. Escolha um, por exemplo o da Universidade Federal do Parana, no link <http://cran-r.c3sl.ufpr.br/>;
- (4) Se o seu sistema operacional for Windows, clicar em *Download R for Windows*, subdiretório **base** e baixar o programa em *Download R 4.2.1 for Windows*.

**OBS.:** A versão do programa pode ser diferente, mas isso não representa problema algum. A versão correspondente aos passos acima é da data de 23/06/2022.



## 2 AMOSTRAGEM

Como comentamos anteriormente, em uma pesquisa qualquer, deve-se pensar antes de tudo em elaborar um projeto que possa garantir sua viabilidade. O projeto faz a previsão e a provisão dos recursos necessários para atingir o objetivo proposto (solucionar um problema, investigar algo) e estabelece a ordem e a natureza das diversas tarefas a serem executadas dentro de um cronograma a ser observado.

Dentro dessas tarefas a delimitação da pesquisa, e posteriormente, a definição (identificação) da população e amostragem, com sua descrição e indicação dos critérios para sua constituição, são muito importantes para a uma conclusão (generalização) confiável dos resultados da pesquisa.

Após a definição do foco da pesquisa (assunto, tema principal), o pesquisador pode decidir ou pelo estudo de todo o universo da pesquisa (população) ou apenas sobre uma amostra. Neste caso, será aquele conjunto de informações que lhe possibilitará a escolha da amostra, que deve ser representativa ou significativa.

### 2.1 População

Antes de abordarmos população, é importante estar bem definido quais são as unidades ou sistemas na população.

#### 2.1.1 Unidade Amostral

Uma unidade amostral é um elemento ou um conjunto de entidades (elementos) relacionadas, que constituem um todo organizado globalmente e relacionado dinamicamente com o meio externo, e que realizam conjuntamente alguma função.

##### **Exemplos ilustrativos:**

- a) em uma pesquisa da incidência de um vírus em uma região, a unidade pode ser uma pessoa, uma família, um distrito ou um município;
- b) em uma pesquisa do controle da evasão escolar, a unidade pode ser um aluno, uma classe, uma escola, ou um município.

A definição da unidade depende do objetivo da pesquisa. Ela deve ser estabelecida na fase inicial da pesquisa, ou seja, na **formulação do problema a pesquisar**. Algumas vezes, a definição da unidade parece óbvia, entretanto, muito frequentemente, a definição ou escolha da unidade não é tão trivial.

A dificuldade da definição da unidade para uma pesquisa decorre do fato de existir na natureza uma hierarquia de sistemas, ou seja, sistemas dentro de sistemas, numa ordem decrescente de amplitude, tal que um determinado sistema é um subsistema em relação a nível hierárquico mais elevado e, por sua vez, contém subsistemas em nível mais baixo.

Muito frequentemente, a unidade é definida vagamente, através de uma sua característica global, ou de uma característica particular importante que é subentendido a identificar. Esse foi o critério adotado nas ilustrações anteriores e que será frequentemente utilizado nesse texto. Entretanto, deve ser claramente compreendido que a definição completa de uma unidade ou sistema deve abranger os seis aspectos listados na caracterização que segue sua definição.

1. função ou objetivos da unidade;
2. insumos - itens entram na unidade;
3. produtos - itens que saem da unidade;

4. componentes - propriedades internas que transformam insumos em produtos;
5. fluxo - movimento de itens entre os componentes da unidade;
6. fronteira ou limite - linha imaginária que demarca e que inclui todos os seus componentes e elementos.

### 2.1.2 População objetivo

Em uma pesquisa científica, a população objetivo ou, mais simplesmente, população, é a coleção bem definida das unidades (sistemas) de interesse para a qual é desejado estabelecer relações ou realizar inferência. O número de unidades é denominado **tamanho da população (N)**.

Uma população objetivo é definida pela especificação de suas unidades ou da caracterização das condições para que estas lhe integrem. A especificação da população, assim como de suas unidades, é determinada pelos objetivos da pesquisa e deve ser estabelecida na formulação do problema.

Populações existentes na natureza são populações finitas, ou seja, de tamanho expresso por um número natural  $N$ , muito frequentemente elevado e desconhecido. Ademais, tais populações têm constituição dinâmica, em decorrência da mutabilidade dos sistemas que lhe integram ao longo do tempo.

Em algumas pesquisas, a população objetivo é constituída por unidades existentes no momento da execução da pesquisa. Uma população nessas circunstâncias, cujas unidades podem ser identificadas, é uma *população real*. Muito frequentemente, entretanto, a população objetivo compreende unidades que não existem no momento da execução da pesquisa, mas que, supostamente, poderão existir no futuro. Nessas circunstâncias, cujas unidades não são identificáveis, mas apenas definidas pela caracterização das condições para que lhe integrem, é uma *população conceitual* (ou, população infinita). Em uma pesquisa de melhoramento da eficiência de um remédio, por exemplo, as unidades da população objetivo não são, somente, as pessoas doentes existentes na região de interesse no momento da execução da pesquisa, mas as pessoas doentes que existirão nessa região, no futuro.

### 2.1.3 Característica

A propriedade básica das populações de interesse na natureza é a heterogeneidade de suas unidades, o que caracteriza o que é comumente denominado de "*variabilidade natural*".

As unidades de uma população objetivo se caracterizam e se distinguem por um conjunto de particularidades, ou propriedades comuns. Cada uma dessas particularidades ou propriedades é uma característica ou atributo da população objetivo e de suas unidades. Cada característica pode manifestar-se nas unidades sob diferentes alternativas ou níveis. Assim, por exemplo, a idade é uma característica dos estudantes do ensino médio que constitui uma população objetivo e dos estudantes que são as unidades dessa população. Essa característica pode manifestar-se em cada um desses estudantes pode assumir qualquer valor numérico de certo intervalo de números reais.

## 2.2 Métodos de pesquisa descritiva (observacional)

Métodos de pesquisa descritiva ou observacional, também denominados levantamentos, são próprios para pesquisa descritiva, ou seja, pesquisa com propósito de prover a descrição do comportamento das unidades de uma população objetivo, através da identificação e descrição das características relevantes dessas unidades e das relações entre essas características. Podem ocorrer em dois níveis:

1. Nível elementar: a pesquisa descritiva detém-se apenas na mensuração de características e determinação de suas propriedades importantes, como médias e variâncias, ou distribuições de frequências.

2. Nível mais avançado: visa não apenas a descrição de características, mas, também, a identificação e descrição de relações de associação ou correlação de características (variáveis), ou seja, relações de características que variam conjuntamente.

A pesquisa descritiva requer:

- as definições claras dos objetivos da pesquisa e da população objetivo,
- a especificação das questões que visa responder,
- o planejamento cuidadoso com respeito às unidades que serão as fontes de informação e,
- aos procedimentos para a coleta de informação.

Em uma pesquisa descritiva as características das unidades (variáveis) se manifestam sem interferência do pesquisador. O pesquisador limita-se à mensuração e registro das informações segundo o plano pré-estabelecido, muito frequentemente por meio de um questionário ou instrumento de registro.

## 2.3 Amostragem

Uma pesquisa pode ser conduzida sobre todas as unidades da população objetivo ou sobre uma fração dessa população, ou seja, sobre **uma amostra** da população objetivo.

Se a pesquisa envolve a observação de todas as unidades da população objetivo, o método de pesquisa é denominado **censo**. Se é conduzida sobre uma amostra da população objetivo, o método de pesquisa é denominado levantamento **por amostragem**.

Como nem sempre há possibilidade de pesquisar todos os indivíduos população que se deseja estudar, devido ao tamanho da população, ou à escassez de recursos ou à premência do tempo. Nesse caso, utiliza-se o método da amostragem, que consiste em obter um juízo sobre o total (universo), mediante a compilação e exame de apenas uma parte, **a amostra**, selecionada por procedimentos científicos. O valor desse sistema vai depender da amostra:

- a) se ela for suficientemente representativa ou significativa;
- b) se contiver todos os traços característicos numa proporção relativa à população.

O censo somente é aplicável em situações em que a população objetivo é finita e suas unidades são identificáveis e disponíveis para coleta de informação. Mesmo nessas circunstâncias, por razões econômicas, éticas ou de outra origem, o levantamento por amostragem pode ser mais conveniente e apropriado. Por essa razão, o levantamento por amostragem é muito mais frequentemente utilizado.

A amostragem é preferível ao censo quando:

- **A população for infinita:** obviamente não seria possível examinar todos os itens da população o que tornaria então o censo impossível.
- **Aplicação de testes caráter destrutivo:** os itens examinados são destruídos no próprio ato do ensaio ou teste durante a pesquisa. Então o censo nos daria o panorama preciso de uma população que não existe mais.
- **O custo:** um censo pode apresentar um custo proibitivo.

- **O tempo:** uma amostra pode ser mais atualizada do que o censo. A necessidade de uma informação ser obtida mais rapidamente, um estudo de toda a população pode consumir demasiado tempo e perder utilidade. Além disso, se a população tende a modificar-se com o tempo, um censo poderá, na realidade, combinar várias populações.
- **A precisão:** a amostragem envolve menor número de elementos e, conseqüentemente, menor número de coletores de dados. Com grande número de agentes, há menor coordenação e controle, aumentando a chance de erros. A amostragem pode revelar maior uniformidade nos métodos de coleta de dados, e *maior comparabilidade, (menor variabilidade)* entre os dados, do que um censo.
- O tipo de informação pode depender da utilização de uma amostra ou de um censo.

Entretanto, há certas situações em que é mais vantajoso fazer um censo. Entre essas situações destacamos:

- a) A população pode ser tão pequena que o custo e o tempo de um censo sejam pouco maiores que para uma amostra.
- b) Se o tamanho da amostra é grande em relação ao da população, o esforço adicional requerido por um censo pode ser pequeno, além disso, o censo eliminará a variabilidade amostral.
- c) Se for exigida precisão completa, então o censo é o único método aceitável. Em face da variabilidade amostral, nunca podemos ter certeza de quais são os parâmetros verdadeiros da população. Um censo nos dará essa informação, embora erros na coleta dos dados e outros tipos de tendenciosidade possam afetar a precisão dos resultados.

O processo de amostragem determina o delineamento de amostragem, ou seja, a relação estrutural entre a amostra e a população objetivo e as chances de seleção das unidades da população objetivo. Há uma grande variedade de delineamentos de amostragem. A distinção mais importante entre esses delineamentos refere-se ao modo de seleção da amostra que pode ser objetivo e ter *base probabilista*, ou *subjetiva e não probabilista*.

## 2.4 Amostragem probabilística

Uma amostragem será probabilística se todos os elementos da população tiverem probabilidade conhecida, e diferente de zero, de pertencer à amostra. Desta forma, a amostragem probabilística implica um sorteio com regras bem determinadas, cuja realização só será possível se a população for finita e totalmente acessível.

Consideraremos aqui os seguintes planos de amostragem probabilística:

1. Amostragem Aleatória Simples.
2. Amostragem Proporcional Estratificada.
3. Amostragem Sistemática.
4. Amostragem por Conglomerado.

### 2.4.1 Amostragem simples ao acaso (ASA) ou aleatória simples

**Objetivo:** obter uma amostra representativa quando os elementos da população são *todos homogêneos* em relação às características que podem influenciar as variáveis que serão medidas ou identificadas.

**Procedimento:** enumeram-se todos os elementos da população  $(1, 2, \dots, N)$  e sorteiam-se " $n$ " elementos mediante um dispositivo aleatório (computador, calculadora, tabela de números aleatórios, entre outros).

**Exemplo:****População:** alunos da sala de aula.**Variável em estudo:** idade média da sala.**Amostra:** 10 alunos selecionados mediante um processo de Amostragem Simples ao Acaso.

Podemos avaliar esse parâmetro por meio de uma amostra simples ao acaso, devido a que todos os elementos da população são homogêneos.

**2.4.2 Amostragem estratificada**

**Objetivo:** melhorar a representatividade da amostra quando os elementos da população são heterogêneos, porém, podem ser agrupados em subpopulações (estratos) contendo elementos homogêneos.

**Procedimento:** a população é dividida em grupos ou estratos contendo elementos homogêneos e as amostras são retiradas separadamente de cada um desses grupos através de uma Amostragem Simples ao Acaso.

**Exemplo:** A diretora de uma determinada escola precisa de informação atualizada sobre determinadas variáveis dos seus 1803 alunos, sendo que por problemas de tempo e custo apenas é possível entrevistar 100 delas. As variáveis são, relação peso-altura, número de irmãos, entre outras. A população foi dividida em 5 estratos, segundo o período e a série.

Estrato	Período (série)	Número de propriedades ( $N_i$ )	$n_i$
1	Ensino básico (1,2)	873	
2	Ensino básico (3,4)	386	
3	Ensino fundamental (1,2)	246	
4	Ensino fundamental (3,4)	186	
5	Ensino médio (1,2,3)	112	
$N=1803$			$n=100$

**Pergunta:** qual deverá ser o tamanho da amostra retirado de cada estrato?

**Resposta:** a amostra de 100 alunos deverá ser retirada de maneira tal que os 5 estratos sejam considerados.

Existem 2 critérios para determinar o tamanho da amostra em cada estrato: Proporcional e Ótimo. **Critério**

**Proporcional:** o estrato " $i$ " fornece uma quantidade  $n_i$  de elementos proporcional ao tamanho  $N_i$  do estrato. Isto é:  $n_i = \frac{N_i}{N}n$ , em que  $N$  tamanho da população e  $n$  o tamanho da amostra. Este critério é adequado quando existe a mesma variabilidade dentro de estratos:  $\sigma_1 = \sigma_2 = \dots = \sigma_k$

**Critério Ótimo:**  $n_i = \frac{N_i S_i}{\sum_{i=1}^k N_i S_i} n$  quando se conhece a variabilidade ( $S_i$  desvio padrão) dentro dos " $k$ " estratos.

**2.4.3 Amostragem Sistemática**

**Objetivo:** aumentar a representatividade da amostra dando maior cobertura à população. É usada quando os elementos da população são *heterogêneos* e não podem ser agrupados em sub-populações homogêneas.

**Procedimento:** enumeram-se todos os elementos da população  $(1, 2, \dots, N)$ , calcula-se " $r$ " e sorteia-se um primeiro elemento para formar parte da amostra, este entre 1 e o " $r$ -ésimo" elemento. Os demais são retirados

em uma progressão aritmética, saltando " $r$ " elementos, até completar o total da amostra ( $n$ ). O valor " $r$ " é chamado passos de amostragem e é determinado por:

$$r = \frac{N}{n}$$

em que " $r$ " deve ser inteiro.

**Esquema:** população enumerada:  $1, 2, 3, \dots, i, \dots, N$ .

Esquema da Amostra Sistemática:

Primeiro elemento:	$i$
Segundo elemento:	$i + r$
Terceiro elemento:	$i + 2r$
Quarto elemento:	$i + 3r$
$\vdots$	$\vdots$
$n$ -ésimo elemento:	$i + (n-1)r$

**Exemplo:** Temos uma cidade com 1000 residências das quais serão amostradas 50 para verificar a qualidade e tratamento da água na residência.

$$r = \frac{N}{n} = \frac{1000}{50} = 20$$

- Enumeram-se todas as residências:  $1, 2, \dots, 1000$ .
- Sorteia-se a primeira residência, ou seja, um valor entre 1 e 20. Digamos que a primeira residência sorteada seja o número 10.
- As residências selecionadas serão:

Primeira residência:	10.
Segunda residência:	30.
Terceira residência:	50.
Quarta residência:	70.
$\vdots$	$\vdots$
Quinquagésima residência:	990.

#### 2.4.4 Amostragem por conglomerados

**Conglomerado:** subgrupo de elementos da população com a mesma heterogeneidade (mesmas características) que a população.

**Objetivo:** facilitar a coleta da informação. Cada conglomerado deve possuir a mesma heterogeneidade (mesmas características) que a população. Isto é, cada conglomerado deve representar bem toda a população.

**Procedimento:** consiste em subdividir a população em *conglomerados* de elementos que reproduzam bem as características da população. Sorteia-se um determinado número de conglomerados. Todos os elementos contidos nos conglomerados selecionados vão compor a amostra.

Esquema:

**Exemplo:** Em uma pesquisa sobre as características das farmácias da região Sul de Minas Gerais, para diminuir custos de transporte ao percorrer toda a região, o pesquisador pode concentrar suas entrevistas sobre um número limitado de municípios (ou sub-regiões) escolhidos ao acaso. Todas as farmácias que estão nas sub-regiões escolhidas podem ser considerados como uma amostra da população de agricultores do Sul de Minas Gerais. Teríamos então:

**População:** as " $N$ " farmácias do Sul de Minas Gerais.

**Conglomerados:** " $M$ " municípios do Sul de Minas Gerais.

**Amostra de conglomerados:** " $m$ " municípios.

**Amostra:** As " $n$ " farmácias contidas na amostra de " $m$ " municípios.

## 2.5 Amostragem não probabilística

Quando nem todos os elementos da população tiverem uma probabilidade diferente de zero de pertencerem à amostra, dizemos que a amostragem é não-probabilística. (ou seja, alguns elementos não terão chance fazer parte (compor) da amostra). Este processo de amostragem é *subjetivo* e seu regimento depende do *conhecimento que possui o pesquisador* a respeito da estrutura da população. É empregada, muitas vezes, por simplicidade ou pela impossibilidade de se obter amostragem probabilística.

Consideraremos aqui os seguintes planos de amostragem não probabilística:

### a) Amostragem a esmo.

É o caso em que o pesquisador procura ser "aleatório", sem, no entanto, utilizar um sorteio aleatório rigoroso.

**Exemplo:** Se tirarmos 100 comprimidos (analgésico) de uma caixa que contém 10000 comprimidos da mesma marca, de certo não faríamos uma amostragem aleatória simples, pois seria extremamente trabalhosa, mas faríamos retiradas a esmo.

Os resultados de uma amostragem a esmo são os mesmos de uma amostragem probabilística se a população é homogênea (elementos da população são *todos homogêneos* em relação às características que podem influenciar as variáveis) e se não existe a possibilidade do amostrador ser influenciado (mesmo que inconscientemente) por alguma característica dos elementos da população. Não seria o caso da amostra dos comprimidos, acima, se estes tivessem modelos e tamanhos diferentes, e isto afetasse a característica observada nos mesmos.

### b) Amostragem por julgamento.

Neste tipo de amostragem, a amostra é colhida na parte da população que é acessível. Então se faz uma distinção entre população-objeto (aquela que se tem em mente ao realizar o estudo) e a população-amostrada (a parte da população que é acessível). Se essas duas populações tiverem as mesmas características, este tipo de amostragem vai ser equivalente a uma amostragem probabilística.

Ocasionalmente, os itens amostrais, se apresentam convenientemente grupados. Uma pesquisa médica deve trabalhar com os pacientes disponíveis. Este grupo não pode ser considerado como uma amostra aleatória do público em geral e seria perigoso tentar tirar conclusões gerais com base em tal estudo. Entretanto, os resultados poderiam proporcionar uma base para a elaboração de um plano de amostragem aleatória para validar

os resultados básicos. Os perigos inerentes à pesquisa médica, bem como a outros tipos de pesquisa, frequentemente obrigam a limitar a pesquisa inicial a um pequeno grupo de voluntários. Outros exemplos similares seriam portadores de doenças fatais, cadáveres, animais, entre outros.

Finalmente, a amostragem por julgamento pode ser mais rápida e menos custosa porque não é preciso construir uma listagem dos itens da população. Tenha-se em mente que a amostragem por julgamento não permite a avaliação objetiva do erro amostral, de modo que é conveniente usar a amostragem probabilística sempre que possível.

Existem outros tipos de amostragem não probabilística que não serão abordadas aqui neste curso.

## 2.6 Problemas de amostragem

### 2.6.1 Amostras tendenciosas

Talvez você nunca faça um trabalho que exija amostragem. Mas muito provavelmente você lerá ou usará resultados de trabalhos cujos dados foram obtidos por amostragem. Então saiba que é importante entender o que é uma *amostra tendenciosa*.

Primeiro, as inferências devem ser feitas apenas para a população de onde a amostra foi retirada.

Não tem sentido, por exemplo, estudar os hábitos de higiene dos índios bolivianos e fazer inferência para a população da periferia da cidade de São Paulo. Também é preciso verificar se a amostra foi retirada da população usando um processo delineado segundo *critérios estatísticos*. A amostra deve ter o tamanho usual da área em que a pesquisa se enquadra.

As amostras *muito pequenas* podem ser excelentes estudos de casos, mas não permitem fazer inferência estatística. Mas também desconfie de amostras muito grandes. Será que o pesquisador tinha tempo e dinheiro para fazer um bom levantamento de tantos dados? E veja como foi feita a coleta de dados(questionário, registros etc).

São mais confiáveis as respostas obtidas através de entrevistas, desde que o entrevistador tenha sido treinado. Estude as perguntas. Elas eram claras? As respostas podem ser, por alguma razão, mentirosas? Leia o artigo e se pergunte:

1. Qual é a população?
2. Como a mostra foi selecionada?
3. Qual é o tamanho da amostra?
4. Como o questionário foi feito?
5. As perguntas eram claras?

Se estas perguntas não tiverem resposta satisfatória, a amostra pode ser tendenciosa.

**Exemplo:** Para estimar o tamanho dos morangos de uma caixa, não seria correto examinar o tamanho dos 10 morangos que estão na parte de cima, pois provavelmente, a amostra seria tendenciosa uma vez que os vendedores de morango arrumam as caixas de maneira a colocar as frutas maiores nas camadas mais superficiais.

- É preciso cuidado na forma de tomar a amostra porque os erros de amostragem fazem com que os resultados da amostragem sejam diferentes dos resultados do censo.



- Quando a pesquisa exige que os participantes respondam a um questionário, é preciso especial atenção à forma de obter as respostas.

Ética é extremamente importante.

- Um estatístico conhecido disse, certa vez, que é possível mentir usando estatística, mas que se mente mais, e melhor, sem estatísticas. É preciso entender que as amostras podem levar a conclusões erradas. Contudo, as opiniões pessoais, sem base em dados, levam em geral, a conclusões *muito mais erradas*.

## 2.7 Exercícios

### Lista: Exercícios de Amostragem

1. Para as situações descritas a seguir, indique a população e a amostra correspondente. Discuta a validade do processo de inferência estatística para cada um dos casos, ou discuta sobre a representatividade da amostra em cada caso.<sup>[1]</sup>
  - (a) Para avaliar a eficácia de uma campanha de vacinação no Estado de Minas Gerais, 200 mães de recém-nascidos, durante o primeiro semestre de um dado ano e em uma maternidade em Belo Horizonte, foram entrevistadas a respeito da última vez em que vacinaram seus filhos.
  - (b) Uma amostra de sangue foi retirada de um paciente com suspeita de anemia, para avaliar a taxa de ferro no sangue.
  - (c) Para verificar a audiência de um programa de televisão, 563 indivíduos foram entrevistados por telefone com relação ao canal em que estavam sintonizados.
  - (d) A fim de avaliar a intenção de voto para presidente dos brasileiros, 122 pessoas foram entrevistadas em Alfenas.
2. Para cada uma das pesquisas consideradas a seguir caracterize a unidade, a população e a amostra<sup>[2]</sup> :
  - (a) Levantamento da opinião dos produtores de café no Sul de Minas sobre as tecnologias geradas pela pesquisa nos últimos dez anos, através de questionário enviado a 100 produtores selecionados da lista de 500 produtores associados a uma cooperativa de produtores deste Estado.
  - (b) Levantamento da opinião sobre a qualidade do transporte urbano de uma cidade, através de consulta por telefone que utiliza um subconjunto de números extraídos da lista telefônica dessa cidade.
  - (c) Estudo da relação entre saúde da criança aos cinco anos e amplitude do tempo de aleitamento materno conduzida nos hospitais e postos de saúde do Município de Varginha
3. Os dados da Tabela 1, referem-se ao peso de cobaias (em Kg) disponíveis em um laboratório para experimentos, sendo cada observação pertence a um elemento, um coelho.
  - a) Considere que os dados acima correspondem a uma população com média populacional 4,03Kg.
  - b) Selecione uma amostra aleatória simples de tamanho 5. Determine a média amostral. Calcule o erro de estimação  $E = |\bar{x} - \mu|$ .
  - c) Selecione uma amostra sistemática de tamanho 5. Determine a média amostral. Qual é o erro de estimação?
  - d) Qual seria o melhor tipo de amostragem para este caso? Explique.

Tabela 1: Peso de 50 cobaias

Identificador	Peso	Identificador	Peso	Identificador	Peso	Identificador	Peso
1	2,85	16	3,91	31	4,16	46	4,53
2	2,85	17	3,91	32	4,19	47	4,54
3	2,89	18	3,92	33	4,20	48	4,59
4	3,56	19	3,92	34	4,21	49	4,87
5	3,62	20	3,95	35	4,22	50	4,90
6	3,65	21	4,00	36	4,22		
7	3,67	22	4,00	37	4,30		
8	3,68	23	4,00	38	4,30		
9	3,73	24	4,00	39	4,31		
10	3,76	25	4,01	40	4,35		
11	3,77	26	4,08	41	4,36		
12	3,77	27	4,08	42	4,50		
13	3,81	28	4,09	43	4,50		
14	3,85	29	4,09	44	4,51		
15	3,87	30	4,10	45	4,51		

4. Num hospital com 1650 pacientes hipertensos, deseja-se calcular o percentual médio de gordura desses pacientes, porém, sabe-se que o peso do paciente pode interferir na resposta desta variável, criando-se então estratos por peso.

Estrato	Classes de peso (kg)	Número de Pacientes	Tamanho da amostra por estrato
1	Até 50	30	
2	51 a 65	120	
3	66 a 80	230	
4	81 a 95	380	
5	Mais de 95	890	
Total		1650	

Escreva os passos para selecionar uma amostra de tamanho 50, aplicando o processo de amostragem que melhor se adapte à informação que você possui.

5. O laboratório de Análises Clínicas da UNIFAL-MG deseja realizar uma pesquisa com o objetivo de caracterizar seus pacientes atendidos. Como forma de obter estas informações, foram distribuídos questionários para 80 pacientes do cadastro geral, por meio do qual procurou-se avaliar o nível de escolaridade (alto, médio, baixo), altura (em cm), idade, religião e número de filhos.

- Qual é a população em estudo?
- Qual é a amostra?
- Quais foram as variáveis estudadas?
- Classifique as variáveis quanto à sua natureza.

## Referências

- [1] MAGALHÃES, M. N., LIMA, A. C. P. **Noções de Probabilidade e Estatística**. São Paulo: Edusp, 2010. [24](#)
- [2] SILVA, J. G. C. da. **Estatística Experimental: Planejamento de Experimentos**. UFP, 2007. [24](#)

### 3 ESTATÍSTICA DESCRITIVA

Para entendermos melhor a finalidade da Estatística Descritiva (Exploratória), vamos relembrar alguns conceitos e definições.

#### 3.1 Conceitos e definições importantes

1. **População:** é um conjunto de elementos com pelo menos uma característica comum. Essa(s) característica(s) comum(s) deve(m) delimitar inequivocamente quais elementos que pertencem ou não à população.

"A população pode ser: finita (quando pode ser enumerada) ou infinita (quando não pode ser enumerada)."

2. **Amostra:** é um subconjunto de uma população. É necessariamente finita, pois todos seus elementos serão examinados para efeito da realização do estudo estatístico desejado.

**Observação:** os elementos de uma *pesquisa experimental* são considerados como sendo uma amostra.

3. **Variável:** é a característica pela qual se deseja que a *população* seja descrita. Essa característica pode assumir diferentes valores de elemento para elemento.

4. **Dados ou observação:** é o valor que assume a variável para um elemento em particular.

Notação:

Tamanho da população: número de elementos da população  $\rightarrow "N"$ .

Tamanho da amostra: número de elementos da amostra  $\rightarrow "n"$ .

Variável:  $X, Y, Z$ , ou  $X_i, Y_i, Z_i$ , etc. (serão letras maiúsculas, geralmente as últimas do alfabeto).

5. **Tipos de variáveis:** as variáveis podem ser:

**a) Variável qualitativa:** são aquelas variáveis que correspondem a atributos ou categorias, podem ser:

**a.1) Variável qualitativa nominal:** quando os atributos (ou categorias) não são possíveis de ordenação.

**Exemplo:** sexo dos alunos de uma turma.

**a.2) Variável qualitativa ordinal:** quando os atributos (ou categorias) são possíveis de ordenação.

**Exemplo:** nível de conhecimento de inglês dos alunos de uma turma (baixo, médio, alto).

**b) Variável quantitativa:** são aquelas variáveis que correspondem a números resultantes de contagem ou medidas, podem ser:

**b.1) Variável quantitativa discreta:** são próprias de dados de contagem, isto é, estão definidas em um conjunto enumerável.

**Exemplo:** número de pessoas de um determinado grupo que já sofreram algum tipo de fratura.

**b.2) Variável quantitativa contínua:** são aquelas em que as realizações resultam de uma medida (uma mensuração) que podem assumir qualquer valor real entre dois extremos. (em um intervalo real).

**Exemplo:** peso e altura dos alunos de uma turma.

## 3.2 Introdução

**Definição:** Estatística descritiva é a parte da Estatística que preocupa-se com a organização, apresentação, e análise dos dados qualitativos, ou quantitativos, sejam eles amostrais ou experimentais.

A **apresentação de dados** é feita mediante alguma representação que pode ser tabular ou gráfica.

A **análise dos dados** é realizada com o cálculo de algumas medidas de interesse, por exemplo: média e desvio padrão.

A tabela é uma ferramenta que possibilita uma inspeção rigorosa dos dados, sendo bastante eficiente para mostrar o comportamento dos dados, facilitar a compreensão das informações bem como a interpretação do fenômeno ou processo estudado.

Uma tabela deve ser autoexplicativa, isto é, as informações nela apresentadas devem ser compreendidas pelo leitor sem que o mesmo tenha que ler ou consultar o texto que a antecede ou precede. Assim, uma tabela deve ser composta pelos seguintes itens, (Figura 4):

1. **Título:** parte que aparece sempre na parte superior da tabela e contem as informações que respondam às perguntas do fenômeno estudado, tais como: O que? Onde? Quando? (deve ser explicativo, com a identificação da população, da variável).

2. **Cabeçalho:** indica a natureza do conteúdo de cada coluna: o que está sendo representado em cada coluna da tabela.

3. **Corpo:** representado por colunas nas quais são registradas as informações. Nessas colunas são apresentadas as frequências absolutas ( $F_i$ ), frequências relativas ( $Fr$ ) e frequências relativas em porcentagens ( $Fp$ ). Sendo:  $F_i$  : indica quantas vezes (frequência) se repete uma categoria ou valor em um conjunto de dados.

$$Fr = \frac{F_i}{n}$$

$$Fp = Fr * 100\%$$

$n$  : tamanho da amostra ou a soma de todas as frequências absolutas.

4. **Caselas:** espaço destinado a um só número.

5. **Total:** deve ser destacado de algum modo.

6. **Fonte/Rodapé:** espaço na parte inferior da tabela onde pode-se inserir notas ou fonte dos dados.

Tabela nº: Título				}	Cabeçalho
Variável	$F_i$	$Fr$	$Fp$		
				}	Corpo
Total	n	1	100		
Fonte:					

Figura 4: Representação de uma tabela de distribuição de frequências.

Note que toda tabela deve ser delimitada por traços horizontais, não devem ser feitos traços verticais para delimitar a tabela e o cabeçalho é separado do corpo por um traço horizontal.

Quando a variável é qualitativa ou quantitativa discreta, a distribuição de frequências consiste em atributos ou categorias com as contagens numéricas referentes a cada conjunto, como pode ser observado no exemplo a seguir:

Exemplo: Considere uma amostra de 2560 trabalhadores do ramo farmacêutico e com suas respectivas classificações segundo a renda mensal.

Tabela 2: Classificação dos trabalhadores do ramo farmacêutico brasileiro, segundo a renda mensal, em 2020

Renda mensal	Fi
Baixa	1613
Média	701
Alta	246
Total	2560

Fonte:

### 3.3 Apresentação dos dados qualitativos

Para dados qualitativos a enumeração (contagem), é a forma mais simples de representá-los.

**Observação:** toda tabela e gráfico deve ter um título explicativo (identificação da população, da variável).

**Representação tabular** Geralmente após a apresentação de uma tabela ou gráfico, devem ser feitos co-

Tabela 3: Preferência pelo tipo de carne em uma amostra de 100 alunos da Escola "X".

Tipo de Carne	Número de indivíduos (Fi)
Branca	25
Vermelha	75
Total	100

mentários sobre os mesmos, sempre referenciando sua numeração.

**Exemplo:** Pode-se observar, na Tabela 3.3, que as crianças da Escola X têm uma maior preferência por carne vermelha.

#### Representação gráfica

O gráfico de barras é um gráfico no qual as barras retangulares tem comprimento proporcional aos valores que ele representa (frequência). As barras podem ser desenhadas verticalmente ou horizontalmente. O gráfico de barras vertical é chamado de gráfico de colunas.

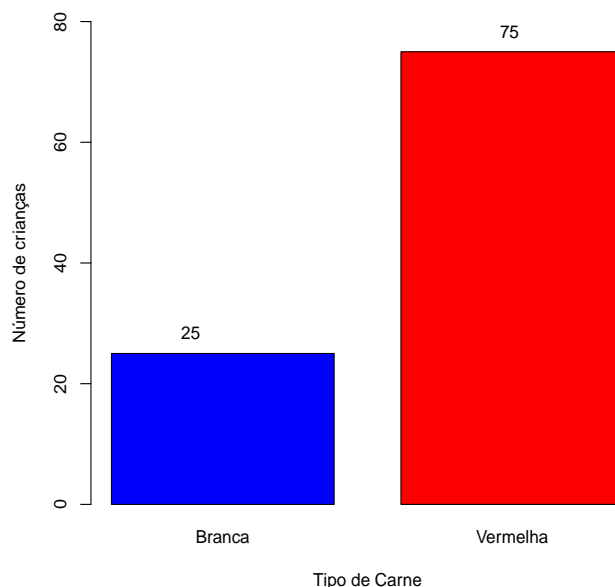


Figura 5: Gráfico de colunas da preferência pelo tipo de carne em uma amostra de 100 alunos da Escola X

O gráfico de setores, comumente chamado gráfico de pizza, é um diagrama circular em que os valores de cada categoria estatística representada são proporcionais à porcentagem da área ocupada, ou seja, cada "fatia" representa uma categoria de dados e o tamanho é proporcional à quantidade que representa.

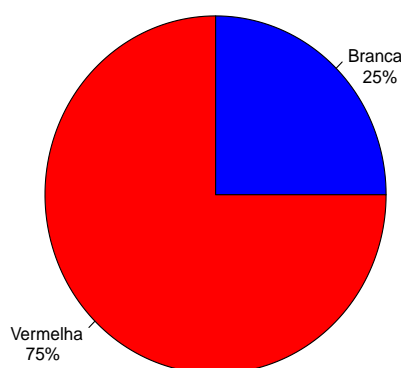


Figura 6: Gráfico de setores da preferência pelo tipo de carne em uma amostra de 100 alunos da Escola X

Os resultados devem ser apresentados apenas de uma forma, portanto, deve-se escolher entre o gráfico ou tabela. Cuidado com variáveis que tenham uma quantidade elevada de respostas (categorias).

Exemplo de gráfico pouco informativo

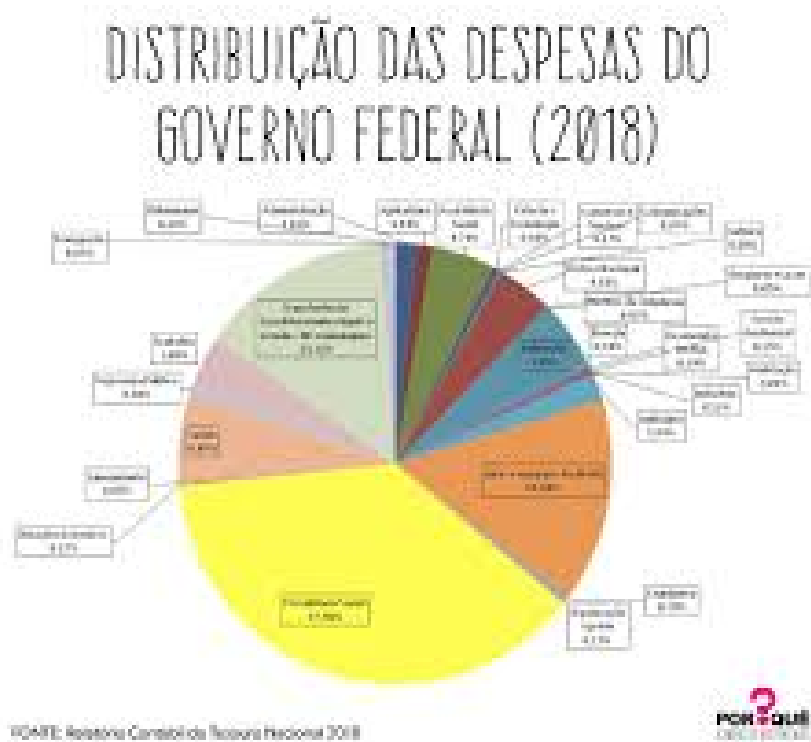


Figura 7: Gráfico de setores como exemplo de uma representação pouco informativa

Fonte: Tesouro Nacional 2018



### 3.4 Apresentação dos dados quantitativos discretos

Na maioria das situações em que temos dados referentes a variáveis quantitativas discretas (característico de contagens), é comum construir uma Tabela de frequência mais simples, na qual se representa a **frequência absoluta de cada valor assumido pela variável** ( $Fi$ ); as **frequências relativas** ( $Fr$ );  $Fr = Fi/n$ ; e as **frequências percentuais** ( $Fp$ );  $Fp = Fr \times 100$ .

**Exemplo:** Um pesquisador examinou, juntamente com sua equipe um lote de 40 caixas de um "produto alimentício", escolhido aleatoriamente num carregamento de 500 caixas (cada caixa possui 50 pacotes), anotando o número de pacotes contaminados por caixa. Os resultados foram os seguintes:

Caixa (ID)	1	2	3	4	5	6	7	8	9	10
pcts com fungo por caixa	0	1	0	0	2	3	1	0	0	2
Caixa (ID)	11	12	13	14	15	16	17	18	19	20
pcts com fungo por caixa	0	0	4	0	2	3	2	0	1	1
Caixa (ID)	21	22	23	24	25	26	27	28	29	30
pcts com fungo por caixa	3	1	0	4	2	1	2	0	0	3
Caixa (ID)	31	32	33	34	35	36	37	38	39	40
pcts com fungo por caixa	0	0	1	1	1	0	2	0	1	3

Efetuada-se a contagem, e organizando os resultados temos:

**Tabela 3.1:** Quantidade de pacotes contaminados por caixa em uma amostra de 40 caixas de um produto alimentício

$N^o$ de contaminados	$Fi$	$Fr$	$Fp(\%)$
0	16	0,400	40,0
1	10	0,250	25,0
2	7	0,175	17,5
3	5	0,125	12,5
4	2	0,050	5,0
TOTAL	40	1,00	100,0

**Observação:** quando a quantidade de resposta assumida pela variável discreta for muito elevada *deve-se criar classes* como para dados contínuos, que será visto à frente.

Perguntas interpretativas da Tabela 3.1.

1. Qual a porcentagem de caixas com mais de 1 pacote contaminado ?
2. Qual a porcentagem de caixas com no mínimo 3 pacotes contaminados ?
3. Qual a porcentagem de caixas com no máximo 1 pacote contaminado ?

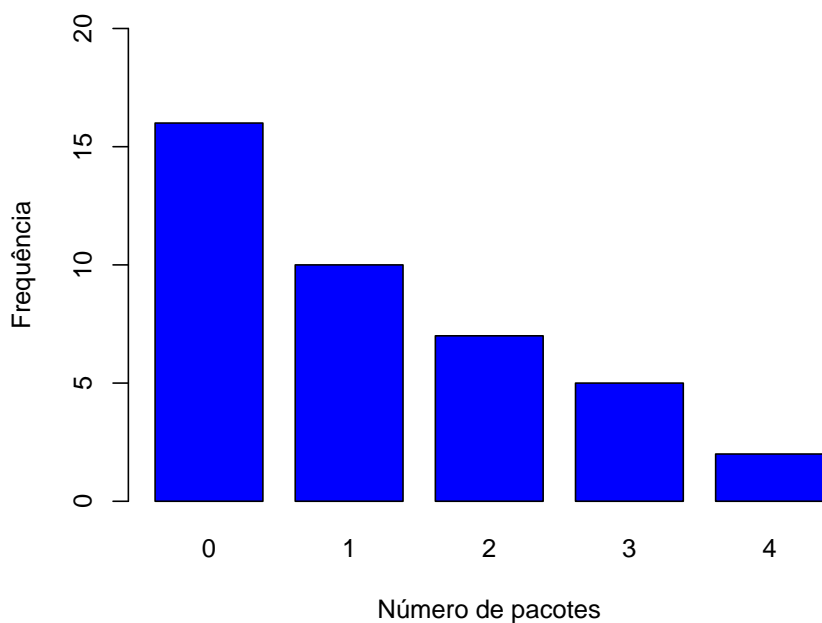


Figura 8: Gráfico de colunas da quantidade de pacotes contaminados por caixa obtida de uma amostra de 40 caixas de um produto alimentício

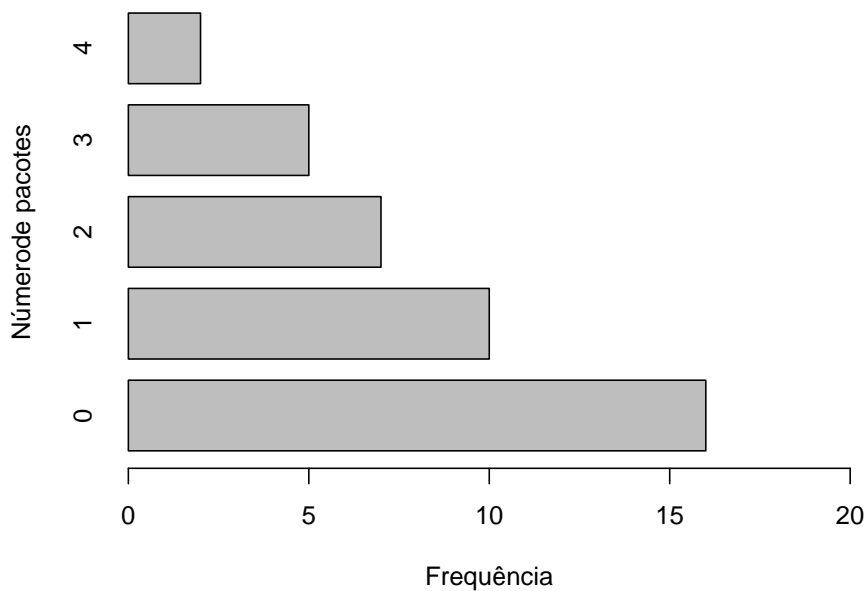


Figura 9: Gráfico de barras da quantidade de pacotes contaminados por caixa obtida de uma amostra de 40 caixas de um produto alimentício

Os resultados devem ser apresentados apenas de uma forma, portanto, deve-se escolher entre um gráfico ou tabela.

### 3.5 Organização e apresentação de dados quantitativos Contínuos

Como nos dados qualitativos, também é feita mediante uma representação. A forma mais simples e utilizada é a distribuição de frequências, definida a seguir.

**Distribuição de frequência:** é a distribuição dos dados em classes ou categorias. O número de elementos associados a cada classe representa a frequência de classe.

#### 3.5.1 Construção de uma tabela de distribuição de frequência (TDF)

**Exemplo:** Informações sobre os resultados de 100 exames de sangue, referente à taxa de glicose, em miligramas por decilitro (mg/dL), em uma amostra indivíduos do sexo feminino que trabalham na indústria EXEP.

86.6	88.6	99.4	90.4	90.8	100.3	92.8	82.4	85.9	87.3	97.3	92.2	92.4	90.7	86.7
100.7	93.0	78.2	94.2	87.2	83.6	88.7	83.8	85.6	86.2	79.9	95.0	90.9	83.2	97.5
92.6	88.2	95.4	95.3	94.9	94.1	93.3	89.6	88.2	87.7	85.8	88.8	82.4	103.0	97.2
83.3	87.6	87.2	94.7	89.5	91.5	89.8	89.7	98.2	88.6	99.1	80.7	93.5	90.7	91.3
92.3	87.0	88.0	83.9	83.6	91.8	92.7	90.3	95.5	102.3	87.1	76.1	96.0	85.7	85.9
96.2	88.3	82.7	91.1	89.2	90.0	92.3	87.8	93.9	88.7	92.0	96.6	92.6	88.0	96.9
96.0	93.3	91.4	86.2	98.2	86.4	103.1	99.2	88.6	83.8					

- Você consegue extrair alguma informação útil destes valores?
- Seria interessante apresentar desta forma?
- Você já imaginou se fossem analisados 5000 mulheres?

**Solução:** o ideal então seria sumarizar os dados por Construindo uma Tabelas e/ou um Gráfico.

Para de construir uma tabela, com critérios estatísticos, deve-se seguir os seguintes passos:

1. Determinar o número de classes ( $K$ ), para a qual podem-se os seguintes critérios:

- A familiaridade do pesquisador com os dados.
- Critério baseado no número de observações ( $n$ ): qual é o número ideal de classes? Oliveira, M. S (1994).

Número de observações ( $n$ )	Número de classes ( $K$ )
<b>Até 100</b>	$\sqrt{n}$ (se decimal, usar o inteiro superior)
<b>Acima de 100</b>	$5\log(n)$

2. Determinar amplitude da classe será determinada por um critério estatístico: Encontrar a amplitude total ( $A$ );

A amplitude total é a diferença entre a maior e a menor observação.

3. Encontrar a amplitude de cada classe ( $C$ );

A amplitude de cada classe a razão entre a amplitude total e o número de classes menos 1.

$$C = \frac{A}{K - 1}$$

4. Determinar as classes: sejam  $LI_i$  e  $LS_i$  os limites inferior e superior da classe " $i$ ", respectivamente.

(a)  $LI_1 = \text{menor Observação} - C/2$ .

(b)  $LI_1$  conhecido, logo  $LS_1 = LI_1 + C$ .

(c)  $LI_2 = LS_1$ .

(d)  $LS_2 = LI_2 + C$ .

Continuar os passos anteriores até completar " $K$ " classes.

5. Determinar o ponto médio da classe ( $X_i$ ):

$$X_i = (LI_1 + LS_1)/2;$$

6. Calcular os valores das frequências absolutas ( $F_i$ );

Para cada classe, o valor de  $F_i$  é o número de observações contidas na classe.

7. Calcular os valores das frequências relativas ( $Fr_i$ );  $Fr_i = F_i/n$

8. Calcular os valores das frequências percentuais ( $Fp_i$ );  $Fp_i = Fr_i \times 100$

9. Calcular os valores das frequências percentuais ACUMULADAS ( $FACp_i$ );  $FACp_i$ =soma das frequências percentuais até a classe " $i$ ".

**IMPORTANTE:** Para fins de cálculos de algumas medidas, todas as observações contidas num intervalo de classe serão considerados *iguais ao ponto médio da classe*. Essa é chamada a Hipótese Tabular Básica (HTB).

**Tabela 3.2:** Distribuição de frequência dos resultados de exames de sangue, referente à taxa de glicose (mg/dL), de indivíduos do sexo feminino que trabalham na indústria EXEP.

Classes	$X_i$	$F_i$	$Fr_i$	$Fp_i$	FACP
74,6 – 77,6	76,1	1	0,01	1,00	1,00
77,6 – 80,6	79,1	2	0,02	2,00	3,00
80,6 – 83,6	82,1	6	0,06	6,00	9,00
83,6 – 86,6	85,1	13	0,13	13,00	22,00
86,6 – 89,6	88,1	23	0,23	23,00	45,00
89,6 – 92,6	91,1	20	0,20	20,00	65,00
92,6 – 95,6	94,1	17	0,17	17,00	82,00
95,6 – 98,6	97,1	10	0,10	10,00	92,00
98,6 – 101,6	100,1	5	0,05	5,00	97,00
101,6 – 104,6	103,1	3	0,03	3,00	100,00
TOTAL		100	1,00	100,00%	-

Perguntas interpretativas da Tabela 3.2.

1. Qual o número de mulheres com taxa de glicose mínima de 83,6 mg/dL?
2. Qual a porcentagem de mulheres com taxa de glicose mínima de 98,6 mg/dL?
3. Qual a porcentagem de mulheres com taxa de glicose maiores que 80,6 mg/dL?
4. Qual a porcentagem de mulheres com taxa de glicose de no máximo que 83,6 mg/dL?

### 3.5.2 Representação gráfica de dados quantitativos contínuos

a) **Histograma:** gráfico representado por retângulos cujas bases são proporcionais às *amplitudes de classe* e as alturas proporcionais às *frequências das classes* (as frequências podem ser  $F_i$ ,  $Fr_i$ , ou  $Fp_i$ ).

b) **Polígono de frequência:** gráfico em que os pontos médios das classes, no topo dos retângulos do histograma, são unidos por linhas. O polígono começa e termina nos pontos médios das classes anterior a primeira e posterior à última respectivamente.

### 3.5.3 Classificação das distribuições de frequências

A forma do polígono de frequência permite classificar a distribuição de frequências em: simétrica, assimétrica à esquerda (negativa) e assimétrica à direita (positiva). Para isto, o polígono deve ser comparado com as seguintes curvas chamadas curvas de frequência.

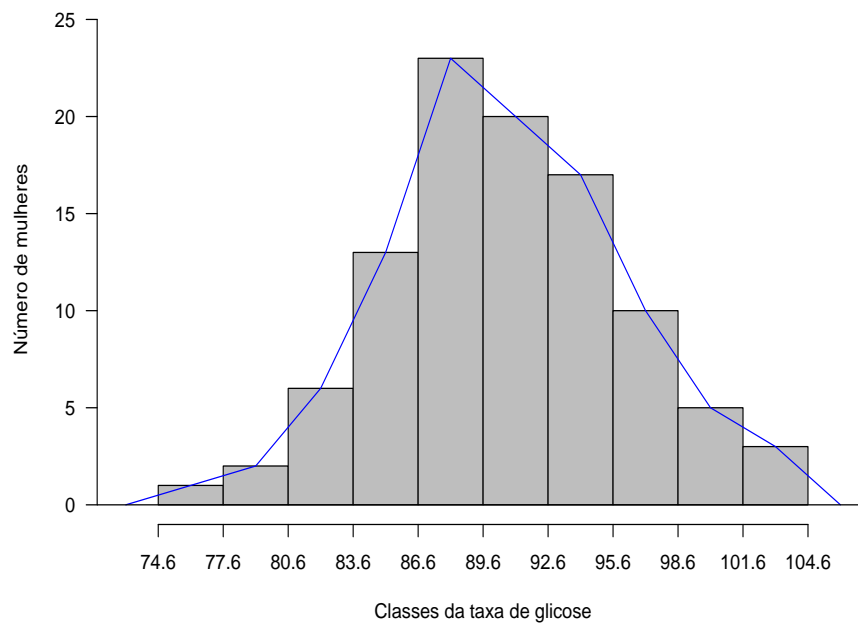


Figura 10: Histograma e polígono dos resultados de exames de sangue, referente à taxa de glicose (mg/dL), de indivíduos do sexo feminino que trabalham na indústria EXEP.

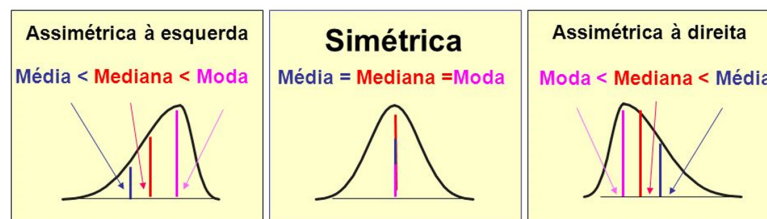


Figura 11: Exemplos de polígonos assimétrico à esquerda (negativa), simétrico e assimétrico à direita (positiva).

**Exemplo 2:** Peso (em kg) de 20 crianças, com 4 anos de idade, de um orfanato do sul de Minas.

14,2   16,2   20,6   21,4   23,4   15,8   18,4   20,6   22,0   26,6  
 14,4   18,4   20,6   21,8   25,2   16,2   19,2   20,8   22,2   27,0

Perguntas interpretativas da Tabela.

1. Qual o número de crianças com peso menor que 22,2kg?
2. Qual a porcentagem de crianças com peso de no mínimo 22,2kg?
3. Qual a porcentagem de crianças com peso de no mínimo 20,6kg?

### 3.6 Exercícios

- 1 Informações sobre os resultados de 36 exames de sangue, referente à fração de colesterol de muito baixa densidade (VLDL), em miligramas por decilitro (mg/dl), em indivíduos do sexo feminino.

5,25	9,57	7,59	13,85	11,06	17,26
4,56	9,55	7,39	13,60	10,76	16,61
6,66	9,80	8,46	15,99	12,79	22,00
5,73	9,65	7,44	14,69	11,59	18,75
4,00	9,54	6,86	13,23	10,53	16,22
6,26	9,77	8,12	14,71	12,00	19,40

- a) Construa a tabela de distribuição de frequência, o polígono e histograma.
- b) Qual o número de exames que apresentaram valores de VLDL de pelo menos 16,6 mg/dl?
- c) Qual o a percentagem de exames que apresentaram valores de VLDL menores que 13,0 mg/dl?
- d) Qual o a percentagem de exames que apresentaram valores de VLDL menores que 7,6 mg/dl?
- 2 Foi contado o número de crianças que cada uma de 50 famílias escolhidas aleatoriamente no Bairro Pinheirinhos em Alfenas, encontrando-se o seguinte resultado:

5	1	5	3	1	2	2	1	1	0
4	2	0	4	4	4	4	3	3	2
1	1	3	2	3	1	2	3	4	3
4	0	2	0	5	2	3	4	3	4
0	0	4	3	2	2	4	3	4	3

- a) Estes dados são de uma população ou de uma amostra?
- b) Classifique a variável em questão.
- c) Construa tabela de distribuição de frequência e o gráfico.
- 3 Seja os dados abaixo referentes à taxa de proteína (g/kg) na carne analisadas em 18 leitões de um experimento.

10,0	10,4	10,7	11,4	11,6	12,2
12,2	12,4	12,4	12,5	13,4	13,5
13,9	14,0	14,6	15,0	15,3	15,6

- a) Construa a tabela de distribuição de frequência, o polígono e histograma.
- b) Qual o número de animais que apresentaram taxa de proteína menor que 13,5g/kg?
- c) Que percentagem de animais apresentaram taxa de proteína menor que 14 g/kg?

## 4 Medidas descritivas

Uma vez que os dados são organizados e apresentados por meio de Tabelas e Gráficos, para se fazer uma melhor análise e interpretação dos mesmos faz-se necessário que as observações sejam bem representadas, sendo para isto utilizadas as chamadas *medidas descritivas*, sendo elas formadas por medidas de posição (média, mediana e moda), medidas de variabilidade ou dispersão (variância, desvio padrão e coeficiente de variação) e medidas separatrizes (quartis, decis e percentis)

### Notação

	Medida	Amostra	População
Medidas de Posição	Média	$\bar{X}$	$\mu$ (mi)
	Mediana	$M_d$	$\tilde{\mu}_d$
	Moda	$M_o$	$\tilde{\mu}_o$
Medidas de Variabilidade	Variância	$S^2$	$\sigma^2$
	Desvio padrão	$S$	$\sigma$ (sigma)
	Coeficiente de variação	$CV$	$\sigma_R$

### 4.1 Medidas de posição

Medidas de posição ou de tendência central são aquelas que procuram sintetizar as informações (os dados ou observações) em um único e informativo valor. Tais medidas têm tendência a posicionar-se no centro das distribuições. Estudaremos a média, moda e mediana.

- Média

É a medida mais usada por ser a mais comum e compreensível delas, bem como pela relativa simplicidade do seu cálculo, além de prestar-se bem ao tratamento algébrico.

A **média aritmética** ou simplesmente **média** de uma população de tamanho  $N$  é, por definição:

$$\mu = \frac{\sum_{i=1}^N x_i}{N}.$$

E, para uma amostra de tamanho  $n$  é, por definição:

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

1. Influenciada pelos valores extremos da distribuição; (Desvantagem)
2. Localiza-se, em geral, na classe de maior frequência;
3. Na sua determinação são considerados todos os dados da distribuição;
4. É única para um conjunto de dados.



- Mediana

Para um conjunto de dados ordenados (Rol), a mediana é o valor que é precedido e seguido pelo mesmo número de dados (observações). Isto é 50% dos dados são superiores à mediana e 50% são inferiores. Cálculo da mediana:

Quando o número de dados (n) for **ímpar**:

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

Ou seja, se "n" for ímpar, a mediana será o elemento da série ordenada que ocupa a posição  $((n + 1)/2)$ .

Quando o número de dados for **par**:

$$Md = \frac{x_{(\frac{n}{2})} + x_{(\frac{n+2}{2})}}{2}$$

Quando o número de elementos da série for par, nunca haverá coincidência da mediana com um dos elementos da série. A mediana será sempre a média aritmética dos dois elementos centrais da série, ou seja, a média dos elementos que estão nas posições  $(n/2)$  e  $(n + 2)/2$ .

- Moda

É o valor que ocorre com maior frequência no conjunto de dados, isto é, o valor mais comum. Quando o conjunto de dados apresentar dois valores com a mesma frequência máxima, cada um deles é uma moda, e o conjunto diz-se bimodal. Caso mais de dois valores ocorram com a mesma frequência máxima, o conjunto é multimodal. No entanto, quando nenhum valor é repetido, o conjunto de dados não tem moda, sendo dito ser amodal.

Quando os dados forem contínuos (sem repetições) para calcular a moda os dados devem estar em classes. O método de Czuber permite encontrar-se a moda em dados agrupados, sendo que, a moda estará contida na classe mais frequente ou, no histograma, a coluna mais alta.

#### 4.1.1 Propriedades da média, mediana e moda

##### Vantagens e desvantagens da média, mediana e moda

Embora a **média aritmética** seja amplamente utilizada como medida descritiva na ciência e para a tomada de decisão, deve-se ter cuidado com seu uso em situações em que a série apresenta "valores extremos".

	Média	Mediana	Moda
VANTAGENS	É tratável matematicamente	<b>Não é</b> influenciada por valores extremos	<b>Não é</b> influenciada por valores extremos
DESVANTAGENS	É influenciada por valores extremos	<b>Não é</b> tratável matematicamente	<b>Não é</b> tratável matematicamente

Portanto todas as vezes que não existirem "**valores extremos**"(outliers) no conjunto de dados deve-se usar a **média**.

#### 4.1.2 EXEMPLO

Seja uma amostra do peso, em mg, de certo comprimido utilizado no tratamento de controle de pressão arterial por pacientes de um PSF na cidade de Alfenas, em 2020.

ROL: 70 - 76 - 76 - 77 - 77 - 78 - 80 - 81 - 81 - 83 - 83 - 83 - 84 - 86 - 86 - 87 - 87 - 88 - 89 - 90 - 90 - 91 - 92 - 92 - 93 - 94 - 94 - 95 - 98 - 99

Calcule e interprete a média, a mediana e a moda. Indique qual a melhor medida de posição para representar o conjunto de dados.

**a) Média**

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2580}{30} = 86,0mg$$

O peso médio dos comprimidos é de 86 mg.

**b) Mediana**

Como temos  $n=30$ , tomaremos os elementos que estão na posição  $x_{15}$  e  $x_{16}$ .

$$Md = \frac{86 + 87}{2} = 86,5mg$$

50% dos pesos dos comprimidos estão abaixo (e acima) de 86,5mg.

**c) Moda**

$$Mo = 83,0mg$$

O peso com maior frequência é de 83,0 mg.

A Média é a melhor medida de posição para representar o conjunto de dados, pois não há valores extremos e ela é tratável matematicamente.

## 4.2 Medidas de variabilidade

São medidas que informam sobre a dispersão dos dados e são necessárias para, junto com a média, representar bem um conjunto de observações.

**a) Amplitude Total (A)** É a diferença entre o maior e menor dos valores da série. Ou seja:

$$A = X_{max} - X_{min}$$

A utilização da amplitude total como medida de dispersão é limitada, pois é uma medida que depende apenas dos valores extremos, não sendo afetada pela variabilidade interna dos valores da série.

**b) Variância e desvio padrão:** medem a variabilidade absoluta de um conjunto de observações.

São as duas medidas de dispersão mais usadas na ciência, sendo calculadas considerando-se todas as observações da série de dados. Elas são fundamentadas nos **desvios em relação à média**:

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

**Observação:** A variância é a soma de quadrados dos desvios em relação à média dividido por " $n-1$ "(amostral) ou " $N$ "(populacional).

**Variância amostral**

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \text{ ou } S^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n} \right]$$

em que:

- $n$  é o tamanho da amostra.
- $S = \sqrt{S^2}$  (desvio padrão amostral).

### Variância amostral para dados não agrupados

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{1}{N} \left[ \sum_{i=1}^N x_i^2 - \frac{\left( \sum_{i=1}^N x_i \right)^2}{N} \right]$$

**Observação 1:** a unidade da variância refere-se ao **quadrado da unidade dos dados**. O desvio padrão é expresso na mesma unidade dos dados originais. Para o exemplo: dados originais (g), desvio Padrão (g) e a variância ( $g^2$ ). Por isso, o desvio padrão possui uma interpretação mais realística dos dados.

**Observação 2:** a expressão  $\sum_{i=1}^n (x_i - \bar{x})^2$  é conhecida como soma de quadrados dos desvios em relação à média, porém, na prática aplicar esta fórmula é mais trabalhoso que utilizar a seguinte expressão:

$$\sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n},$$

em que  $\sum_{i=1}^n x_i^2$  e  $\sum_{i=1}^n x_i$  podem ser obtidos com certa facilidade em uma calculadora científica realizando os seguintes passos (para o modelo CASIO ou modelo semelhante):

1. Ativar o módulo estatístico: tecla "mode" → "2(SD)". No visor deve aparecer SD.
2. Com o módulo estatístico ativado digite cada dado e "guarde" na memória teclando "M+". No visor vai aparecer a contagem dos dados. Repita o processo do primeiro até o último dado.
3. Com os dados na memória podemos obter os somatórios:
 
$$\sum_{i=1}^n x_i^2: \text{tecla "shift"} \rightarrow "1 \text{ (S-SUM)}" \rightarrow "1" \rightarrow "=",$$

$$\sum_{i=1}^n x_i: \text{tecla "shift"} \rightarrow "1 \text{ (S-SUM)}" \rightarrow "2" \rightarrow "=",$$
 Para obter quantos dados estão na memória: tecla "shift" → "1 (S-SUM)" → "3" → "=".
4. Com os dados na memória podemos, também, obter de forma direta a média e o desvio padrão amostral:
 

média: tecla "shift" → "2 (S-VAR)" → "1" → "=".

desvio padrão amostral: tecla "shift" → "2 (S-VAR)" → "3" → "=".
5. PARA APAGAR A MEMÓRIA: Para sair deste módulo, e consequentemente "apagar" a memória, Tecla "mode" → "1(comp)".

**c) Coeficiente de variação:** também conhecido como desvio padrão relativo (DPR), é uma medida de variabilidade relativa. Refere-se à variabilidade dos dados em relação à média.

$$CV = \frac{S \times 100\%}{\bar{x}}$$

em que  $S$  é o desvio padrão amostral e  $\bar{x}$  é a média amostral. Ou seja, expressa percentualmente a variação dos dados em relação à média.

Sua **vantagem** é permitir a comparação de grandezas diferentes, que estão em unidades diferentes. Porém, possui algumas **restrições de uso**:

E, pode ser interpretado como:

Baixa variabilidade  $\Rightarrow CV < 10\%$ : dados homogêneos.

Média variabilidade  $\Rightarrow 10\% \leq CV < 20\%$

Alta variabilidade  $\Rightarrow 20\% \leq CV < 30\%$

Muito alta variabilidade  $\Rightarrow CV \geq 30\%$ : dados heterogêneos.

Observação: Essa interpretação é apresentada por Garcia (1989) e tem-se a ressalva de que esses valores podem ser diferentes para cada área. Deste modo, aconselha-se a realizar uma revisão de literatura.

Desvantagem:

I) Quando a média da variável em estudo tende a zero, o CV tende ao infinito (o que não tem sentido prático).

**d) Erro padrão da média:** refere-se à variabilidade dos dados em relação ao tamanho da amostra ( $n$ ).

$$EP = \frac{S}{\sqrt{n}}$$

em que  $S$  é o desvio padrão amostral e  $n$  é tamanho amostral.

**Observação:** como estaremos trabalhando com dados resultantes de experimentos, vamos sempre trabalhar com medidas amostrais.

### 4.2.1 EXEMPLO

Seja uma amostra do peso, em mg, de certo comprimido utilizado no tratamento de controle de pressão arterial por pacientes de um PSF na cidade de Alfenas, em 2020.

ROL: 70 - 76 - 76 - 77 - 77 - 78 - 80 - 81 - 81 - 83 - 83 - 83 - 84 - 86 - 86 - 87 - 87 - 88 - 89 - 90 - 90 - 91 - 92 - 92 - 93 - 94 - 94 - 95 - 98 - 99

Calcule a variância, o desvio padrão e o Coeficiente de variação (CV). Interprete o CV.

**a) Variância:** 
$$S^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n} \right]$$

$$S^2 = \frac{1}{30-1} \left[ 223394,0 - \frac{(2580)^2}{30} \right] S^2 = \frac{1514}{29} = 52,21mg^2$$

**b) Desvio padrão**

$$S = \sqrt{52,21mg^2} S = 7,23mg$$

**c) Coeficiente de variação** Lembrando que:  $\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2580}{30} = 86,0mg$

$$CV = \frac{7,23}{86} \times 100 = 8,40\%$$

Os pesos dos comprimidos apresentam uma variabilidade de 8,40% em relação à média, ou seja, apresentam baixa variabilidade.

### 4.3 Medidas de separatriz (quantil)

A mediana é uma medida separatriz, pois divide a série em duas partes iguais. Assim como há o interesse em dividir a série em duas partes iguais, pode haver interesse em dividi-la em 4, 10 ou 100 partes iguais. Surgiram então os Quartis, Decis e Percentis.

Dentre as medidas de separatriz, as mais utilizadas são os quartis. Os **Quartis** ( $Q_1$ ,  $Q_2$  e  $Q_3$ ), são três valores que dividem a série em quatro partes iguais. São as medidas de separatriz mais utilizadas.

1. Primeiro Quartil ( $Q_1$ ): valor situado de tal modo na série de dados que 25% das observações são menores que ele e 75% são maiores.
2. Segundo Quartil ( $Q_2$ ): valor situado de tal modo na série de dados que 50% das observações são menores que ele e 50% são maiores.
3. Terceiro Quartil ( $Q_3$ ): valor situado de tal modo na série de dados que 75% das observações são menores que ele e 25% são maiores.

O cálculo desses elementos segue o mesmo procedimento de mediana (que é igual ao  $Q_2$ ). Ordenam-se os dados de forma crescente (rol) e calcula-se a posição  $p^*$  em que está a medida desejada, em que  $p^*$  é dado por:

$$p^* = \frac{\text{Numero de dados}}{\text{Numero de partes}} \times \text{ordem}$$

$$Q_o = x_{p^*}, \text{ se } p^* \text{ for inteiro ou}$$

$$Q_o = \frac{(x_{p^*} + x_{p^*+1})}{2}, \text{ se } p^* \text{ for não for inteiro.}$$

Os Centis ou Percentis são 99 valores que dividem a série em cem partes iguais;

**Exemplo:** Considerando os dados do **exemplo 2.1**, calcular e interpretar os Quartis ( $Q_1$  e  $Q_3$ )

Para  $Q_1$ , como  $n=30$ , temos que  $p^* = \frac{30}{4} \times 1 = 7,5$ . Como não se tem a posição 7,5, fazemos a média dos elementos da posição 7 e 8.

$Q_1 = \frac{80+81}{2} = 80,5mg$ , isto indica que, 25% dos comprimidos tem pesos menores que 80,5 mg.

Para  $Q_3$ , como  $n=30$ , temos que  $p^* = \frac{30}{4} \times 3 = 22,5$ . Como não se tem a posição 22,5, fazemos a média dos elementos da posição 22 e 23.

$Q_3 = \frac{91+92}{2} = 91,5mg$ , isto indica que, 25% dos comprimidos tem pesos maiores 91,5 mg.

#### 4.4 Gráfico Boxplot ou Diagrama de caixas

O BOXPLOT é a forma gráfica de representar a distribuição dos dados por meio de cinco medidas estatísticas (MÍNIMO,  $Q_1$ , MEDIANA,  $Q_3$  e MÁXIMO), e permite obter um conjunto mais completo de informações sobre a forma da distribuição, conforme o exemplo:

O gráfico Boxplot nos fornece informações sobre a posição central, dispersão e assimetria da respectiva distribuição de frequências dos dados.

Eles também são usados para identificar valores discrepantes (**outliers**). O objetivo é analisar a distribuição dos dados. Assim, as conclusões que tiramos ao analisar um box plot são: centro dos dados (a média ou mediana), a amplitude dos dados (máximo - mínimo), a simetria ou assimetria do conjunto de dados e a presença de outliers.

Para identificar valores discrepantes (outliers), usa-se a média com o desvio padrão ou o **Desvio Interquartilico (DQI)** que é dado por: Dados a média ( $\bar{X}$ ) e o desvio padrão ( $S$ ) são possíveis *outliers* os valores da série de dados que são:

- menores que  $\bar{X} - 3 \times S$ , ou;
- maiores que  $\bar{X} + 3 \times S$ .

Dados o primeiro e o terceiro quartis são possíveis *outliers* os valores da série de dados que são:

- menores que  $Q_1 - 1.5 \times DQI$ , ou;
- maiores que  $Q_3 + 1.5 \times DQI$ .

Em que:

$$DQI = Q_3 - Q_1$$

## 4.4.1 EXEMPLO

**Exemplo:** Informações sobre os resultados de 100 exames de sangue, referente à taxa de glicose, em miligramas por decilitro (mg/dL), em uma amostra indivíduos do sexo feminino que trabalham na indústria EXEP.

86.6	88.6	99.4	90.4	90.8	100.3	92.8	82.4	85.9	87.3	97.3	92.2	92.4	90.7	86.7
100.7	93.0	78.2	94.2	87.2	83.6	88.7	83.8	85.6	86.2	79.9	95.0	90.9	83.2	97.5
92.6	88.2	95.4	95.3	94.9	94.1	93.3	89.6	88.2	87.7	85.8	88.8	82.4	103.0	97.2
83.3	87.6	87.2	94.7	89.5	91.5	89.8	89.7	98.2	88.6	99.1	80.7	93.5	90.7	91.3
92.3	87.0	88.0	83.9	83.6	91.8	92.7	90.3	95.5	102.3	87.1	76.1	96.0	85.7	85.9
96.2	88.3	82.7	91.1	89.2	90.0	92.3	87.8	93.9	88.7	92.0	96.6	92.6	88.0	96.9
96.0	93.3	91.4	86.2	98.2	86.4	103.1	99.2	88.6	83.8					

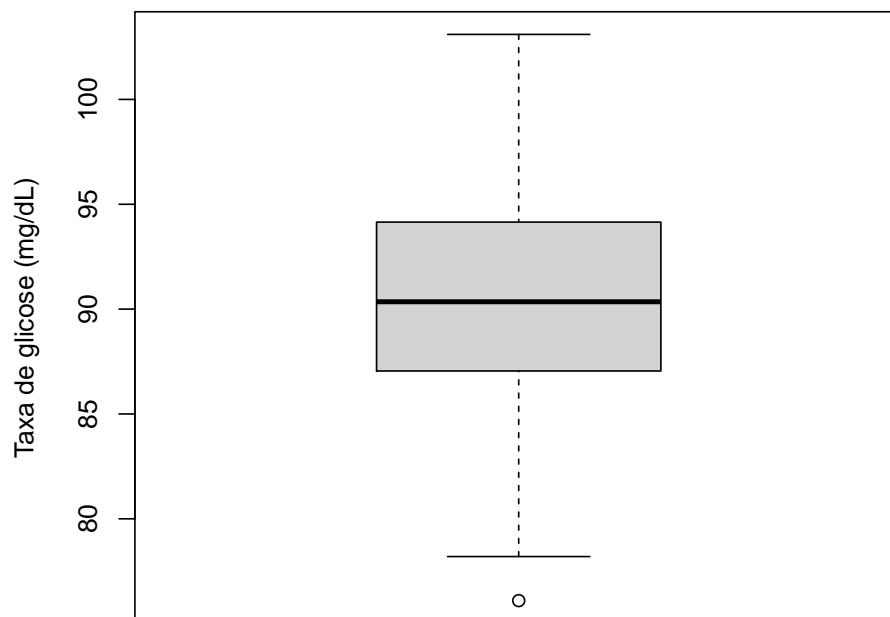


Figura 12: Boxplot dos resultados de exames de sangue, referente à taxa de glicose, em miligramas por decilitro (mg/dL), em uma amostra indivíduos do sexo feminino que trabalham na indústria EXEP..

**Interpretação:** Observa-se na Figura 12, que os dados de taxa de glicose se distribuem de forma simétrica e apresentam um *outlier* inferior.

## 4.5 Exercícios

- 1 Seja os dados abaixo referentes à taxa de proteína (g/kg) na carne de suínos analisadas em 18 amostras.

10,0 10,4 10,7 11,4 11,6 12,2  
 12,2 12,4 12,4 12,5 13,4 13,5  
 13,9 14,0 14,6 15,0 15,3 15,6

- Construa a tabela de distribuição de frequência, o polígono e histograma.
  - Qual o número de animais que apresentaram taxa de proteína menor que 13,5g/kg?
  - Que percentagem de animais apresentaram taxa de proteína menor que 14 g/kg?
  - Calcule e interprete a média, a mediana e a moda. Indique qual a melhor medida de posição para representar o conjunto de dados.
  - Calcule e interprete o quartil 1 ,quartil 3 e o coeficiente de variação.
  - Construa o box-plot e interprete.
- 2 Foi contado o número de crianças não-vacinadas em cada uma das 50 famílias escolhidas aleatoriamente no Bairro Pinheirinhos em Alfenas, encontrando-se o seguinte resultado:

5 1 5 3 1 2 2 1 1 0  
 4 2 0 4 4 4 4 3 3 2  
 1 1 3 2 3 1 2 3 4 3  
 4 0 2 0 5 2 3 4 3 4  
 0 0 4 3 2 2 4 3 4 3

- Construa o box-plot e interprete.
  - Calcule: a média, a moda, a mediana e o coeficiente de variabilidade e interprete.
- 3 Um laboratório que fabrica comprimidos analgésicos anuncia que seu remédio contra dor de cabeça tem a mesma média e uma variação menor na quantidade do princípio ativo em relação ao analgésico tradicional. Avaliaram se 8 amostras aleatórias de cada analgésico, obtendo-se os seguintes resultados (em mg).

Novo	6	5	8	5	6	7	6	6
Tradicional	8	6	5	7	4	9	5	6

Pode-se afirmar que o fabricante tem razão? Calcule as medidas de interesse e conclua.

- 4 Para estudar o rendimento de proteína de certo cultivar de aveia branca um pesquisador coletou 12 amostras, cujo rendimento de proteína foi medido. Os resultados obtidos, (em mg/g), foram os seguintes:

148 116 126 148 135 144 129 128 120 146 140 146

- Determine a média, a mediana, a moda, e o  $Q_1$  e  $Q_3$  do rendimento de proteína.
- Qual é a medida de posição mais apropriada para representar esse conjunto de dados?
- Você acha que a variabilidade dos dados é grande ou pequena? Por quê?



- 5 Examinou-se um lote de 100 caixas de um fármaco, escolhidas aleatoriamente num carregamento de 10000 caixas prontas para exportação, anotando o número de "vidros" com pH maior que o permitido (fora do padrão) por caixa (cada caixa contém 20 vidros). Os resultados foram os seguintes:

Número de vidros fora do padrão	0	1	2	3	4	5
Número de caixas (frequência)	40	28	12	10	6	4

- a) Qual é o número médio, mediano e modal, de vidros "Fora do Padrão" por caixa?
- b) Qual deverá ser o número total de vidros "fora do padrão" no carregamento?
- 6 Um medicamento utilizado em enfermidades decorrentes de contraturas musculares dolorosas contém em sua formulação a cafeína que é um estimulante do sistema nervoso central que atua sobre a musculatura estriada tornando-a menos susceptível à fadiga. Um pesquisador com o intuito de verificar a quantidade de cafeína (em mg) num "medicamento A" analisou uma amostra aleatória de 12 comprimidos obtendo os seguintes resultados:

3 4 5 4 3 6 3 4 2 4 4 3

- a) Calcule a média, mediana, moda, o terceiro quartil ( $Q_3$ ), o erro padrão e o coeficiente de variação.
- 7 Considere os conjuntos de dados da quantidade de um determinado exame realizado em dois laboratórios:

Laboratório A 380 360 370 1000 320

Laboratório B 720 750 740 730 760 720 770

Calcule (somente) a medida de posição mais apropriada para cada conjunto de dados. Comente o porque.

- 8 Considere os conjuntos de dados a seguir referentes aos preços (em reais) de um remédio em diferentes farmácias de cidades.

Cidade A	58	56	57	60	62	62	
Cidade B	72	75	74	73	74	72	77
Cidade C	48	49	45	47	43		

Qual é a cidade com maior variabilidade nos preços?

- 9 Para verificar o número de lesões musculares que os atletas de um clube já sofreram foram entrevistados aleatoriamente 50 atletas obtendo os seguintes resultados:

Número de lesões ( $x_i$ )	0	1	2	3	4	5
Frequência de atletas	14	20	9	4	2	1

- a) Calcule a média, mediana, moda e o coeficiente de variação do número de lesões musculares por atleta.