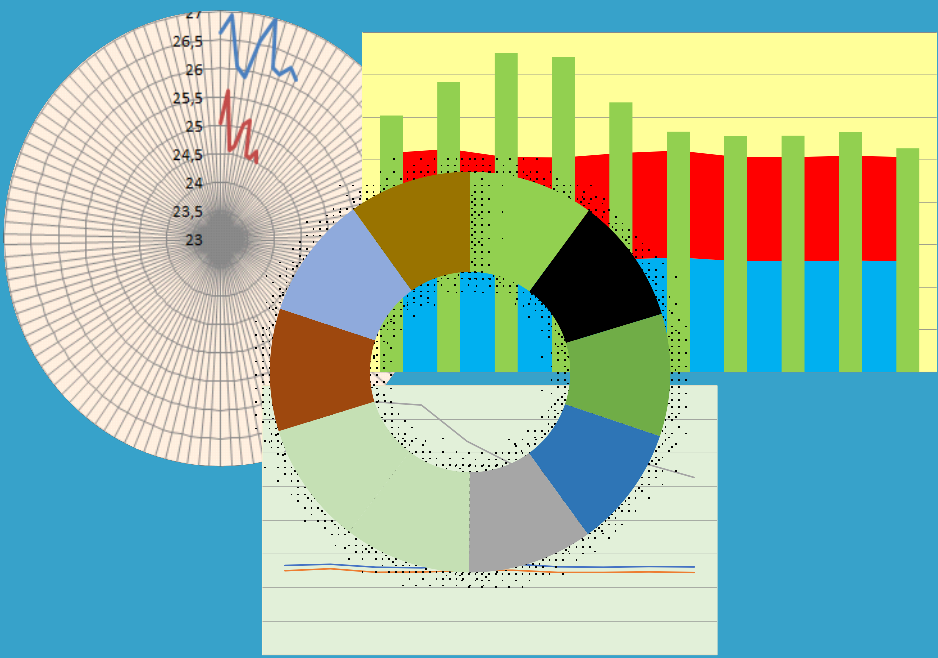
Terezinha de Jesus Rocha Vilanova Moreira
Marlei Rosa dos Santos
Antonia Leidiana Moreira
(Orgs.)

ESTATÍSTICA BÁSICA
para cursos de graduação



EdUESPI

Vol.1

Terezinha de Jesus Rocha Vilanova Moreira
Marlei Rosa dos Santos
Antonia Leidiana Moreira
(Orgs.)

ESTATÍSTICA BÁSICA
para cursos de graduação

Vol.1

EdUESPI

UniversidadeEstadual do Piauí
RuaJoão Cabral • n. 2231 • Bairro Pirajá • Teresina-PI

GOVERNO DO ESTADO DO PIAUÍ
UNIVERSIDADE ESTADUAL DO PIAUÍ - UESPI

José Wellington Barroso de Araújo Dias **Governador do Estado**
Maria Regina Sousa **Vice-governadora do Estado**
Evandro Alberto de Sousa **Reitor**

Marcelo de Sousa Neto **Editor**
Autoras **Revisão**
Editora e Gráfica - UESPI **E-book**

Ficha elaborada pelo Serviço de Catalogação da Biblioteca Central da UESPI

E79c Estatística básica / Terezinha de Jesus Rocha Vilanova Moreira,
Marlei Rosa dos Santos, Antonia Leidiana Moreira, organizadoras. –
Teresina: EdUESPI, 2021.
E-book.

ISBN: 978-65-88108-29-1

1. Matemática. 2. Estatística. 3. Probabilidade.
I. Moreira, Terezinha de Jesus Rocha Vilanova (Org.). II. Santos, Marlei Rosa dos (Org.). III. Moreira, Antonia Leidiana (Org.). IV. Título.

CDD: 519.5

Ficha Catalográfica elaborada pelo Serviço de Catalogação da Universidade Estadual do Piauí - UESPI
Nayla Kedma de Carvalho Santos (Bibliotecária) CRB 3ª Região/1188

Editora da Universidade Estadual do Piauí - EdUESPI
UESPI (*Campus Poeta Torquato Neto*)
Rua João Cabral, 2231 • Bairro Pirajá • Teresina-PI

Apresentação

Não há ramo da matemática, por mais abstrato que seja, que não
possa um dia vir a ser aplicado ao ramo do mundo real."
Lobachevsky

"Uma probabilidade razoável é a única certeza."
Samuel Howe

A experiência não permite nunca atingir a certeza absoluta.
Não devemos procurar obter mais que uma probabilidade."
Bertrand Russel

Ao pesquisarmos e elaborarmos esse material, procuramos levar em conta essas três afirmações citadas anteriormente, conservando o que julgamos ser fundamental: o desenvolvimento claro e compreensível de todos os conceitos básicos aliados ao rigor matemático aplicado à estatística, cujo objetivo é o de fazer com que o leitor compreenda as ideias básicas da Estatística e, quando necessário, saber aplicá-las na resolução de problemas do mundo real.

Observa-se que a utilização da Estatística é cada vez mais acentuada em diversas atividades profissionais, nos seus mais diversificados ramos, com maior ou menor intensidade, devido às múltiplas aplicações que o método estatístico proporciona àqueles que dele necessitam.

Este livro resume de forma clara e objetiva vários assuntos dentro da Estatística básica e contém exercícios para facilitar a compreensão do conteúdo proposto.

Uruçuí-PI, 10 de junho de 2021

Profa. Terezinha de Jesus Rocha Vilanova Moreira
Profa. Dra. Marlei Rosa dos Santos
Profa. Dra. Antonia Leidiana Moreira
(Organizadoras)

SUMÁRIO

CAPÍTULO I - A CIÊNCIA ESTATÍSTICA

1. PANORAMA HISTÓRICO

Embora a Estatística seja recente na área de pesquisa, já era observada de forma rudimentar e imprecisa na antiguidade, quando, então, os governantes coletavam e faziam registros de dados que consideravam importantes, tais como as informações sobre suas populações e suas riquezas, tendo como objetivo fins militares ou tributários (COSTA, 2011).

Dados têm sido coletados ao longo da história. Vejamos (COSTA, 2011):

I. Antiguidade – vários povos já registravam o número de habitantes, de nascimentos e óbitos. Já se faziam estimativas das riquezas sociais, distribuíam-se equitativamente terras aos povos, cobravam impostos e realizavam inquéritos quantitativos por processos que, hoje, chamaríamos de "estatísticas".

II. Idade Média – colhiam-se informações que, geralmente, eram tabuladas com finalidades tributárias ou bélicas.

III. Século XVI – surgem as primeiras análises sistemáticas, as primeiras tabelas, os números relativos e o cálculo de probabilidade. Periodicamente, nas civilizações antigas, eram feitos inquéritos sobre os quantitativos anuais de trigo e outros produtos e, com base nesses dados, eram estabelecidos os impostos. Essa prática também era utilizada na Idade Média. Até o início do século XVII, a Estatística servia apenas para assuntos de Estado e limitava-se a uma simples técnica de contagem, traduzindo numericamente fatos ou fenômenos observados. Era a fase da Estatística Descritiva.

IV. Século XVII – iniciou-se na Inglaterra uma nova fase de desenvolvimento da Estatística, voltada para a análise dos fenômenos observados. Era a fase da Estatística Analítica. John Graunt (1620-1674) foi quem publicou pela primeira vez, concretamente, um trabalho de estatística que se reportava à mortalidade dos habitantes de Londres. Esse trabalho esteve na base do aparecimento das primeiras tábuas de mortalidade e da elaboração de previsões sobre a duração de vida humana. Nasceu, assim, a Demografia.

V. Século XVIII – a origem do termo Estatística surgiu neste século. Quanto à origem da estatística, a data de seu aparecimento não parece ser encarada com unanimidade. Há quem diga que o seu autor foi Godofredo Achenwall (1719-1772), que usou pela primeira vez o termo estatística (*statistik*, do grego *statizein*). Outros também afirmam que tem origem na palavra estado, do latim *status*, pelo aproveitamento que dela tiravam os políticos e o Estado. Contudo, muito antes do termo Estatística, os romanos asseguravam o recenseamento dos cidadãos, e a Bíblia chega, inclusive, a testemunhar um desses recenseamentos.

VI Ao longo dos séculos XVIII e XIX – a Estatística desenvolveu-se muito, associada ao cálculo das probabilidades, que haviam se desenvolvido, e à realização de trabalhos de pesquisa científica nos domínios da Botânica, Biologia, Meteorologia, Astronomia etc. Mais tarde, a Estatística deixou de ser mera técnica de contagem de fenômenos para se transformar numa poderosa "alfaia[1]" científica a serviço dos diferentes ramos do saber. Surge então a fase da Estatística Aplicada.

É com essas características que a Estatística é hoje reconhecida, pois informações numéricas são necessárias para cidadãos e organizações de qualquer natureza e de qualquer parte do mundo globalizado. Portanto, é uma ciência moderna, imprescindível para entender aspectos e problemas em todas as áreas do conhecimento.

A Estatística é uma ciência meio, e não fim. Isso relaciona a Estatística com tabelas e gráficos (com dados experimentalmente obtidos: representados e analisados). Exemplos: Movimento das Bolsas de Valores, estatísticas da Loteria Esportiva, estatísticas da Saúde Pública etc.

2. MÉTODOS DE PESQUISA

Quando se faz uma pesquisa científica, o procedimento geral é formular hipóteses e testá-las. Inicialmente, essas hipóteses são formuladas em termos científicos dentro da área de estudo. Em seguida, devem ser expressas em termos estatísticos.

2.1 MÉTODO CIENTÍFICO - Conjunto de normas básicas que devem ser seguidas para a produção de conhecimentos que têm o rigor da ciência, ou seja, é um método[2] usado para a pesquisa e comprovação de um determinado conteúdo. Destacamos o método experimental e o método estatístico.

[1] Alfaia – objeto dedicado a um fim; adorno; etc. (https://www.dicionario.priberam.org/alfaia). Neste caso significa "referência".

[2] **MÉTODO:** é um conjunto de meios dispostos convenientemente, mais eficaz, para atingir determinada meta (CRESPO, 2002).

2.1.1 Método Experimental - Consiste em manter constante todas as causas (fatores), menos uma, e variá-la de modo que o pesquisador possa descobrir seus efeitos, caso existam. É o método preferido no estudo da Física, da Química, etc. (CRESPO, 2002).

2.1.2 Método Estatístico - Método que admite todas as causas presentes variando-as, dada a impossibilidade de manter as causas constantes, registrando essas variações e procurando determinar as influências que cabem a cada uma delas (COSTA, 2011). Exemplo: Quais as causas que definem o preço de uma mercadoria quando a sua oferta diminui?

2.1.2.1 Fases do Método Estatístico - As principais fases do método estatístico são: 1ª Fase - Definição do problema; 2ª Fase - Planejamento; 3ª Fase - Coleta de dados; 4ª Fase - Apuração dos dados; 5ª Fase - Apresentação dos dados e 6ª Fase - Análise e interpretação dos dados (CASTANHEIRA, 2003).

1ª Fase: Definição do problema - Saber exatamente aquilo que se pretende pesquisar.

2ª Fase: Planejamento - Planejar a realização das fases seguintes: determinando o objetivo da pesquisa, as características da amostra, os métodos que serão utilizados e o processamento de dados.

> **Nota:**
> Perguntas que deverão ser observadas:
> - Como levantar informações?
> - Que dados deverão ser obtidos?
> - Qual levantamento a ser utilizado: Censitário? Por amostragem?
> - Cronograma de atividades?
> - Custos envolvidos, etc.

3ª Fase: Coleta de dados - Fase operacional[3]. A coleta pode ser de dois tipos: coleta direta e coleta indireta.

I. Coleta direta - obtida diretamente da fonte. Levantamento de registros (nascimentos, óbitos, notas fiscais, impostos, entre outros) ou coletados diretamente através de

[3] É o registro sistemático de dados, com um objetivo determinado (TOLEDO; OVALLE, 2008).

inquéritos, questionários etc. Esses dados são chamados de dados[4] primários. Esta coleta pode ser classificada como:

***a) Coleta contínua**:* quando feita de forma contínua, como registros de nascimento, óbitos, casamentos, frequência de alunos às aulas, etc. (CRESPO, 2002).

b) Coleta periódica: quando feita em intervalos constantes de tempo, como censos (de 10 em 10 anos), avaliações mensais dos alunos, recenseamento demográfico, censo industrial, e outros (CRESPO, 2002).

c) Coleta ocasional: quando feita em determinada situação para atender a um objetivo, como pesquisa de mortalidade de um rebanho, pesquisa de um produto no mercado, registro de casos de dengue, etc.

- **Os dados podem ser:**

***a) Dados primários**:* quando são publicados pela própria pessoa ou organização que os haja recolhido (coleta direta) (TOLEDO; OVALLE, 2008).

Coleta de dados primários: o pesquisador não pergunta, **observa**.

Exemplo 1.1

Tabelas do censo demográfico do Instituto Brasileiro de Geografia e Estatística - IBGE (Tabela 1.1).

Tabela 1.1 - População residente, na capital e interior segundo as Grandes Regiões 1940/2000.

Grandes regiões	População residente					
	Total		Capital		Interior	
	1940	2000	1940	2000	1940	2000
Norte	1.632.917	12.900.704	368.528	3.895.400	1.264.389	9.005.304
Nordeste	14.434.080	47.741.711	1.270.729	10.162.351	13.163.351	37.579.365
Sudeste	18.278.837	72.412.411	3.346.991	18.822.986	14.931.846	53.589.425
Sul	5.735.305	25.107.616	459.659	3.290.220	5.275.646	21.817.396
Centro-Oeste	1.088.182	11.636.728	152.189	4.291.120	935.993	7.345.608

Fonte: IBGE, Censo Demográfico 1940/2000.

[4] Dado estatístico - é um dado numérico e é considerado a matéria-prima sobre a qual iremos aplicar os métodos estatísticos.

Exemplo 1.2

Pesquisa de observação para diagnosticar as necessidades de trânsito de uma cidade.

- **Levantamento** - É o método mais comum de se coletar dados. O instrumento pode ser um questionário estruturado ou um roteiro de itens em que o entrevistado disserta à vontade sobre cada item da pesquisa.

As vantagens e desvantagens das três principais formas de levantamento de dados são:

a) Entrevista pessoal – mais flexível e muito cara.

b) Telefone – mais barato, penetra em segmentos difíceis, mas é de fácil recusa.

c) Questionário (postal, WhatsApp e e-mail) – mais lento, média de retorno das respostas muito baixa, mas sem interferência do pesquisador.

II. Coleta indireta: é inferida de elementos conhecidos, através de uma coleta direta (dados primários), ou do conhecimento de características relacionadas ao fenômeno estudado.

a) Dados Secundários*:* quando são publicados por outra organização (dados primários) (TOLEDO; OVALLE, 2008).

Exemplo 1.3

Pesquisa sobre mortalidade infantil que é feita sobre a coleta direta de dados de nascimentos e óbitos.

Exemplo 1.4

Quando determinado jornal publica estatísticas referentes ao censo demográfico extraídas do IBGE.

Observação: É mais seguro trabalhar com fontes primárias. O uso da fonte secundária traz o grande risco de erros de transcrição.

4ª Fase: Apuração dos dados - Resumo dos dados através de sua contagem e agrupamento: condensação e a tabulação de dados; soma e processamento dos dados obtidos mediante critérios de classificação. Pode ser manual, mecânica ou eletrônica (TOLEDO; OVALLE, 2008).

5ª Fase: Apresentação dos dados - Há duas formas de apresentação que não se excluem mutuamente (TOLEDO; OVALLE, 2008).

- A apresentação tabular: é uma apresentação numérica dos dados em linhas e colunas distribuídas de modo ordenado, segundo regras práticas fixadas pelo Conselho Nacional de Estatística.
- A apresentação gráfica dos dados numéricos constitui uma apresentação geométrica permitindo uma visão rápida e clara do fenômeno[5].

6ª Fase: Análise e interpretação dos dados - A última fase do trabalho estatístico é a mais importante e delicada. Portanto, está ligada essencialmente ao cálculo de medidas e coeficientes, cuja finalidade principal é descrever o fenômeno (Estatística descritiva).

3. O QUE É ESTATÍSTICA?

A Estatística é uma parte da Matemática Aplicada que fornece métodos para a coleta, organização, apresentação, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões (CRESPO, 2002). Em outras palavras, estatística é a ciência que tem como base o estudo de uma população.

Esse estudo pode ser feito de duas maneiras:

a) Investigando todos os elementos da população.
b) Por amostragem, ou seja, selecionando alguns elementos da população.

Assim, a estatística divide-se em duas grandes áreas: Estatística Descritiva e Estatística Indutiva ou Inferencial.

Nota:
A coleta, a organização e a descrição dos dados estão a cargo da Estatística Descritiva, enquanto a análise e a interpretação desses dados ficam a cargo da Estatística Indutiva ou Inferencial (CRESPO, 2002).

[5] **FENÔMENO ESTATÍSTICO:** é qualquer evento que se pretenda analisar, cujo estudo seja passível da aplicação do método estatístico. Classificam-se em três grupos (TOLEDO; OVALLE, 2008):
a) Fenômenos de massa ou coletivo: são aqueles que não podem ser definidos por uma simples observação. A estatística dedica-se ao estudo desses fenômenos. *Ex: A natalidade em Uruçui, o preço médio da cerveja em Teresina, a mortalidade, a nupcialidade etc*
b) Fenômenos individuais: são aqueles que irão compor os fenômenos de massa. *Ex: Cada nascimento em Uruçuí, cada preço de cerveja em Teresina, cada indivíduo que morre, cada casamento que ocorre, etc.*
c) Fenômenos de multidão: quando as características observadas para a massa não se verificam para o particular, para o indivíduo isoladamente.

Figura 1.1 – Introdução à Estatística. Fonte: Contract Pharma Brasil, 2020. http://contractpharmabrasil.com.br/noticias_descricao.php?id=830

3.1 ESTATÍSTICA DESCRITIVA – É aquela que possui um conjunto de técnicas para planejar, organizar, coletar, resumir, classificar, apurar, descrever, comunicar e analisar os dados em tabelas, gráficos ou em outros recursos visuais, além do cálculo de estimativas de parâmetros representativos desses dados, interpretação de coeficientes e exposição que permitam descrever o fenômeno.

> **Nota:**
> Essa área apenas descreve e analisa um conjunto de dados, sem tirar conclusões.
> Exemplo: Os 150 trabalhadores de uma fábrica ganham em média R$ 1.200,00 por mês.

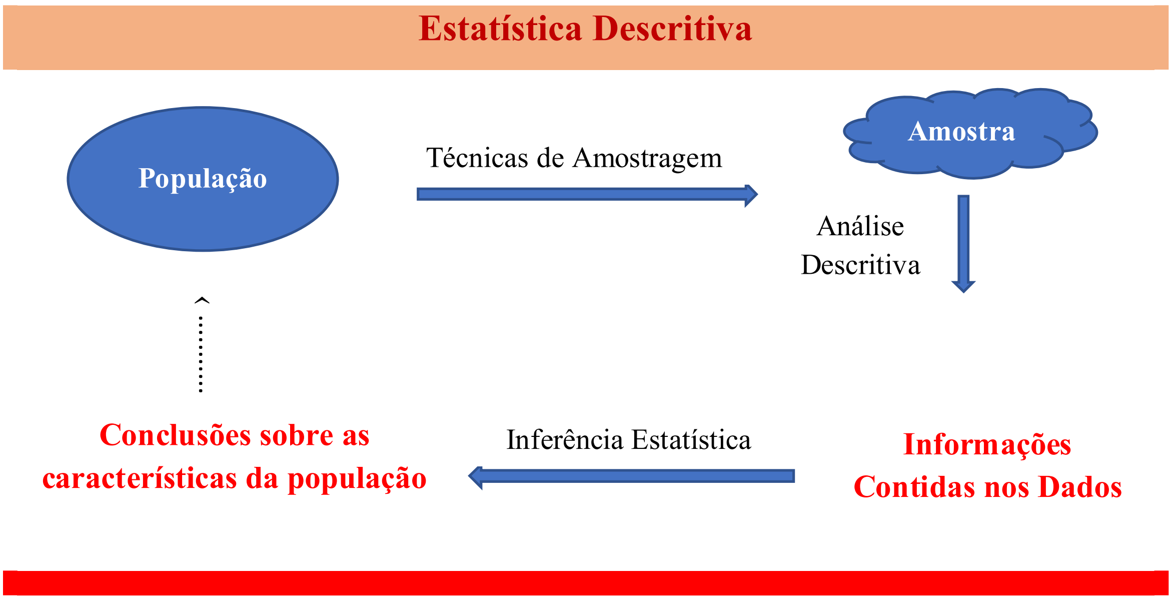


Figura 1.2 – Estatística descritiva. Fonte: PINTO (2019).

3.2 ESTATÍSTICA INDUTIVA OU INFERENCIAL – É o conjunto de técnicas que, partindo de uma amostra, estabelece hipóteses, tira as conclusões sobre a população de origem, formula previsões, fundamentando-se na teoria das probabilidades, e baseia-se na análise e na interpretação dos dados.

Para melhor entendimento dos propósitos da inferência estatística, faz-se necessário o entendimento de *população,* que é a totalidade de itens, objetos ou pessoas sob consideração, e a *amostra* é uma parte da população selecionada para análise (MARTINS, 2010).

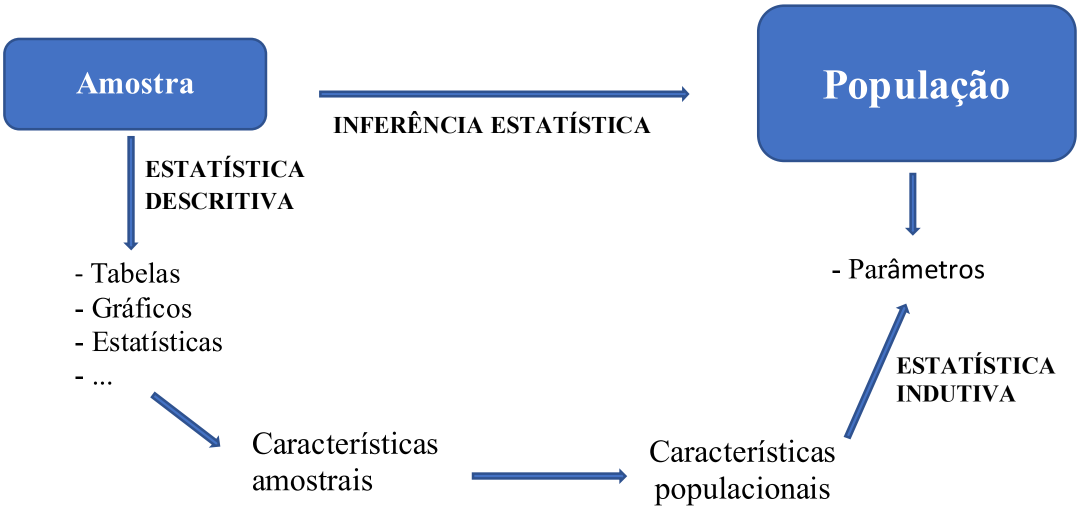


Figura 1.3 – Estatística inferencial. Fonte: Matemática estatística SPSS, 2014.

Notas:

- O processo de generalização do método indutivo está associado a uma margem de incerteza. Isso se deve ao fato de que a conclusão que se pretende obter para o conjunto de todos os indivíduos (população) analisados, quanto à determinadas características comuns, baseia-se em uma parcela (amostra) de observações.
- Na análise estatística de dados, pode-se obter os resultados de duas maneiras: através de um censo ou de uma amostragem.
- A partir do século XVII, matemáticos como Bernoulli, De Moivre e Gauss estabeleceram as bases da estatística inferencial, mas somente no século XX é que os métodos e as técnicas da estatística inferencial foram desenvolvidos por estatísticos como Pearson, Fisher, Gosset, entre outros (MARTINS, 2010).

Exemplo 1.5

Em um teste, a opinião revelou que 65% da população brasileira apoia um determinado candidato para presidente da república. Se este candidato for realmente inscrito para as eleições, é de se esperar que ele se eleja.

Exemplo 1.6

Pesquisa de mercado, pesquisa de opinião pública e praticamente todo experimento.

Exemplos 1.7

a) Proporção de crianças com a cobertura vacinal completa (estudo PI);

b) Número de latrocínios no Piauí, por região administrativa;

c) Percentual de intenção de votos para um candidato numa pesquisa eleitoral;

d) Medida do desempenho escolar de crianças expostas à violência doméstica do pai contra a mãe.

> **Nota:**
> Parâmetro - É uma característica fixa e desconhecida da população a qual se tem interesse em estudar (FARIAS, 2008). Os parâmetros representam quantidades numéricas que podem ser interpretadas pelo pesquisador, como por exemplo: média, proporção, variação, taxa de crescimento e outros.

4. POPULAÇÃO E AMOSTRA

4.1 POPULAÇÃO[6] - É o conjunto de informações ou conjunto de entes portadores de, pelo menos, uma característica comum (CRESPO, 2002). A população pode ser constituída por pessoas, animais, minerais, vegetais, etc.

- Os estudantes matriculados em um determinado curso de uma universidade constituem uma população, pois apresentam uma característica em comum.

Em um estudo estatístico, objetiva-se conhecer uma ou mais características dos elementos de uma população. É importante definir bem essas características de interesse para que sejam delimitados os elementos que pertencem à população e aos que não pertencem a esta.

A população pode ser constituída por pessoas, animais, minerais, vegetais etc.

[6] "População estatística ou universo estatístico" (CRESPO, 2002).

Exemplos 1.8

a) Deseja-se saber se nas indústrias situadas no Estado do Piauí, em 2018, existia algum tipo de controle ambiental.

   População ou universo – indústrias situadas no Estado do Piauí em 2018.

   Características – existência ou não de algum tipo de controle ambiental na indústria.

b) Deseja-se conhecer o consumo total de energia elétrica em Mwh nas residências da cidade de Uruçuí-PI, no ano de 2018.

   População ou universo – todas as residências que estavam ligadas à rede elétrica em Uruçuí-PI, no ano de 2018.

   Características – consumo anual de energia elétrica em Mwh.

4.1.1 Divisão da população

a) **População finita** – apresenta um número limitado de elementos. Dessa forma, é possível enumerar todos os elementos componentes.

Exemplo 1.9

Idade dos alunos do Curso de Agronomia da UESPI – Campus Uruçuí em 2018.

População – todos os alunos do Curso de Agronomia da UESPI – Campus Uruçuí em 2018.

b) População infinita – apresenta um número ilimitado de elementos, o que torna impossível a enumeração de todos os elementos componentes. Entretanto, tal definição existe apenas no campo teórico, porque, na prática, nunca encontraremos populações com infinitos elementos, mas somente populações com grande número de componentes. Nessas circunstâncias, tais populações são tratadas como se fossem infinitas.

Exemplo 1.10

Febre aviária.

População – aves.

4.1.2 Amostra - É uma parte ou um subconjunto finito de uma população (CRESPO, 2002).

Figura 1.4- População e amostra. Fonte: Google Sites, 2019.

- **Os dados de observação** registrados na amostra fornecem informações sobre a população. O processo pelo qual são tiradas conclusões sobre a população, com base nos resultados obtidos na amostra, refere-se à **inferência estatística** vista na definição.
- Para que as inferências sejam corretas, faz-se necessário que a amostra seja representativa da população, isto é, a amostra deve possuir as mesmas características básicas da população, no que diz respeito ao fenômeno a ser pesquisado
- As estatísticas obtidas na amostra são denominadas estimativas[7].

Exemplo 1.11

Para se verificar a concentração de poluentes no ar de algumas casas noturnas é feita a verificação por um aparelho que testa uma pequena quantidade de ar.

5. VARIÁVEIS[8] ESTATÍSTICAS - Representam o atributo ou característica que se pretende estudar em uma população ou amostra e são divididas em dois tipos: qualitativas e quantitativas.

5.1 VARIÁVEIS QUALITATIVAS - Quando seus valores são expressos por atributo[9]. Dividem-se em variáveis nominais e variáveis ordinais.

[7] Estimativa: é um **valor aproximado do parâmetro** e é calculado com o uso da amostra (RAMOS, 2016).
[8] Variável é o conjunto de resultados possíveis de um fenômeno (CRESO, 2002).
[9] **ATRIBUTO:** quando os dados estatísticos apresentam um caráter qualitativo, o levantamento e os estudos necessários ao tratamento desses dados são designados genericamente **de estatística de atributo** (TOLEDO; OVELLE, 2008).

5.1.1 Variáveis nominais – quando não existe ordenação nos atributos, envolve simplesmente o ato de nomear, rotular ou classificar um objeto, pessoa ou alguma característica, por meio de números ou outros símbolos.

- A variável pode assumir duas ou mais categorias (sem ordem hierárquica). O que se mede é colocado em outra categoria, indicando somente diferenças com respeito a uma ou mais características.
- Os números eventualmente utilizados têm função puramente de classificação e não podem ser operados aritmeticamente. Por exemplo, a religião é uma variável nominal.
- Quando variáveis nominais assumem duas categorias, são chamadas variáveis dicotômicas; e se assumirem três ou mais categorias, denominam-se variáveis categóricas (MARTINS, 2010).

Exemplos de variáveis com nível de mensuração nominal:

- Sexo (simbolizado por M e F, 1 e 2, etc.) filiação partidária, profissões, categorias funcionais, cor dos olhos, cor da pele, estado civil, cidade natal, marcas de carro, sexo, etc. (MARTINS, 2010).

5.1.2 Variáveis ordinais – quando os códigos numéricos podem agir como categorias ou ordenações. Elas envolvem variáveis que representam algum elemento de ordem. Uma classificação em anos pode ser um bom exemplo, assim como a faixa etária dos indivíduos.

Exemplos de variáveis com nível de mensuração ordinal (MARTINS, 2010):

- Status socioeconômico, grau de escolaridade, hierarquização de um conjunto de afirmações, atitudes de pessoas em relação a determinado fato, resultado de testes, etc.
- Grau de satisfação da população brasileira com relação ao trabalho de seu presidente (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmente satisfeito).
- Escolaridade (ensino fundamental, médio e superior), mês de observação (janeiro, fevereiro e março...), grau de satisfação (escala de 0 a 5).

Nota:
A resposta é expressa através de “palavras”.

5.2 VARIÁVEIS QUANTITATIVAS – Quando os dados são de caráter nitidamente quantitativo e o conjunto dos resultados possui uma estrutura numérica. Podem ser contínuas, discretas intervalar e razão.

5.2.1 Variáveis contínuas – são aquelas que podem assumir qualquer valor num certo intervalo (contínuo) da reta real. Essas variáveis, geralmente, provêm de medições.

Exemplo 1.12

A altura dos alunos é uma **variável contínua**, pois, teoricamente, um aluno poderá possuir altura igual a 1,70; 1,71; 1,711 e 1,712 m (medições, peso, estatura etc.).

5.2.2 Variáveis discretas ou descontínua – são aquelas que podem assumir apenas valores inteiros não negativos em pontos de uma reta. É possível enumerar todos os possíveis valores da variável.

Exemplo 1.13

Número de alunos de uma escola, número de mensagens em um e-mail, etc.

5.2.3 Variáveis de nível intervalar - a variável pode assumir várias categorias que mantém uma relação de ordem, além de intervalos iguais de medição. As distâncias são as mesmas em toda a escala. Existe um intervalo constante (unidade de medida). Por exemplo (MARTINS, 2010):

- Em uma prova de Estatística, com 30 questões de igual dificuldade, é aplicada a três garotas. Se Ana resolveu 10, Laura 20 e Breda 30, a distância entre Ana e Laura é igual à distância entre Laura e Breda.
- O zero (0) desse nível de mensuração é arbitrário, não é real.
- Um exemplo clássico em ciências naturais é a temperatura (em graus Centígrados ou Fahrenheit), em que o zero é arbitrário (não implicando que realmente haja zero temperatura).
- Trata-se de uma escala verdadeiramente quantitativa com possibilidade de aplicação a todas as estatísticas paramétricas comuns.
- São exemplos de variáveis com nível intervalar: peso, altura, volume, etc.

5.2.4 Variáveis de nível de razão – além de todas as características do nível intervalar, o zero é real, absoluto (não é arbitrário). Zero absoluto significa que há um ponto na escala onde não existe a propriedade (MARTINS, 2010).

Nota:
Para o uso de métodos e técnicas estatísticas, não se faz distinção entre os níveis intervalar e de razão (MARTINS, 2010). Para as variáveis quantitativas, a resposta é expressa em "valores numéricos" (TOLEDO; OVALLE, 2008).

5.3 DESCRIÇÃO GRÁFICA DAS VARIÁVEIS ESTATÍSTICAS - a organização, sumarização e descrição de dados podem ser feitas por meio da construção de gráficos e tabelas.

5.3.1 Descrição gráfica de variáveis qualitativas - o gráfico de barras horizontais ou verticais (Figura 1.5) e o gráfico em forma de pizza (Figura 1.6), são os gráficos mais comuns para a descrição de dados oriundos de variáveis qualitativas. Eles mostram as frequências de observações para cada nível, ou categoria, da variável que se deseja descrever (MARTINS, 2010).

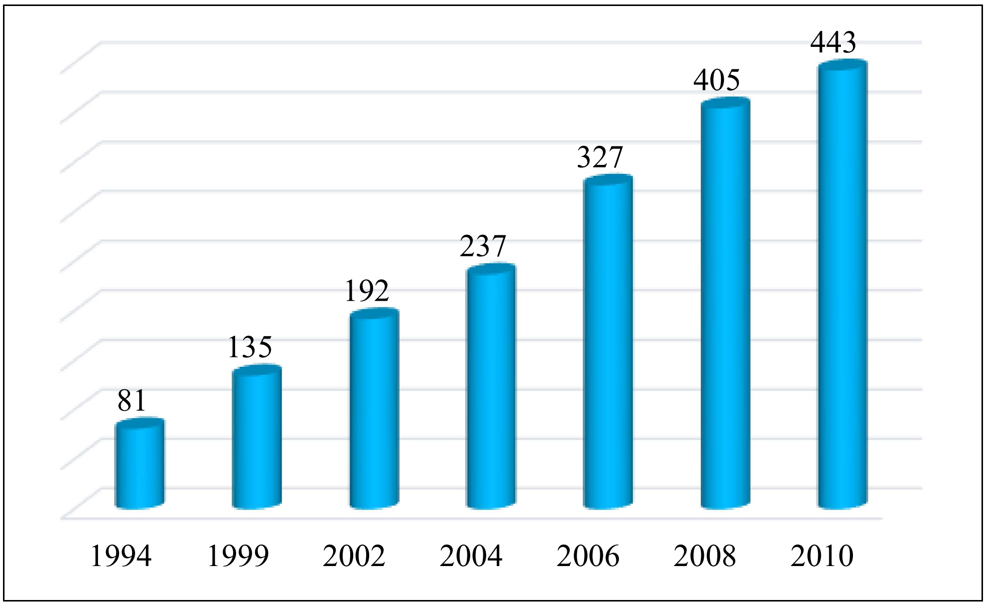


Figura 1.5 – A) Municípios com coleta seletiva de lixo sólido no Brasil em 2010. Fonte: www.cempre.org.br.Adaptado

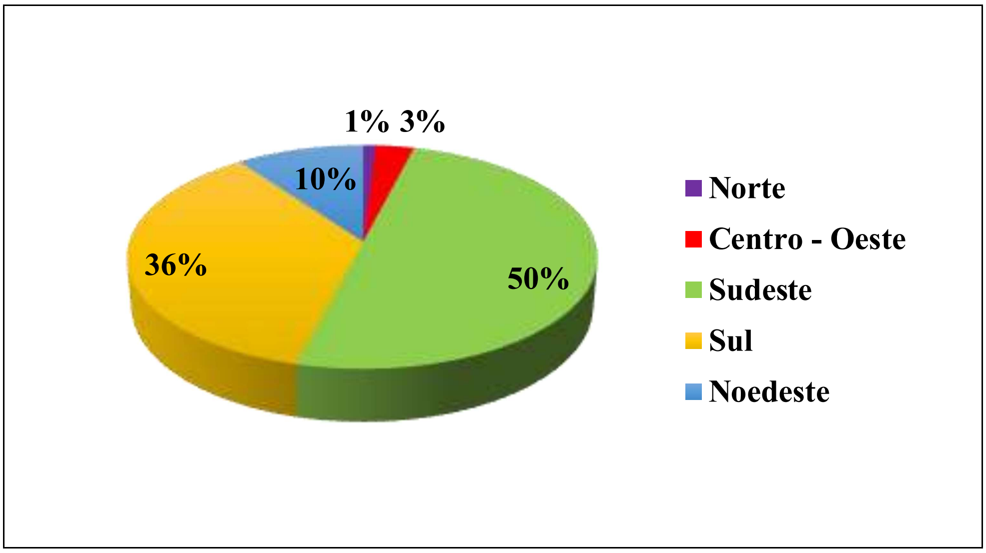


Figura 1.6 – A) Municípios com coleta seletiva de lixo sólido no Brasil e B) Percentual do total dos 443 municípios brasileiros, por região, que mantém coleta seletiva de lixo sólido em 2010. Fonte: www.cempre.org.br.Adaptado

5.3.2 Descrição gráfica de variáveis quantitativas - os histogramas são os gráficos mais adequados para a descrição de variáveis quantitativas. Basicamente, eles mostram as frequências de observações para cada valor ou conjunto de valores da variável que se deseja descrever. Com base em uma tabela de distribuição de frequências é construído o histograma (MARTINS, 2010).

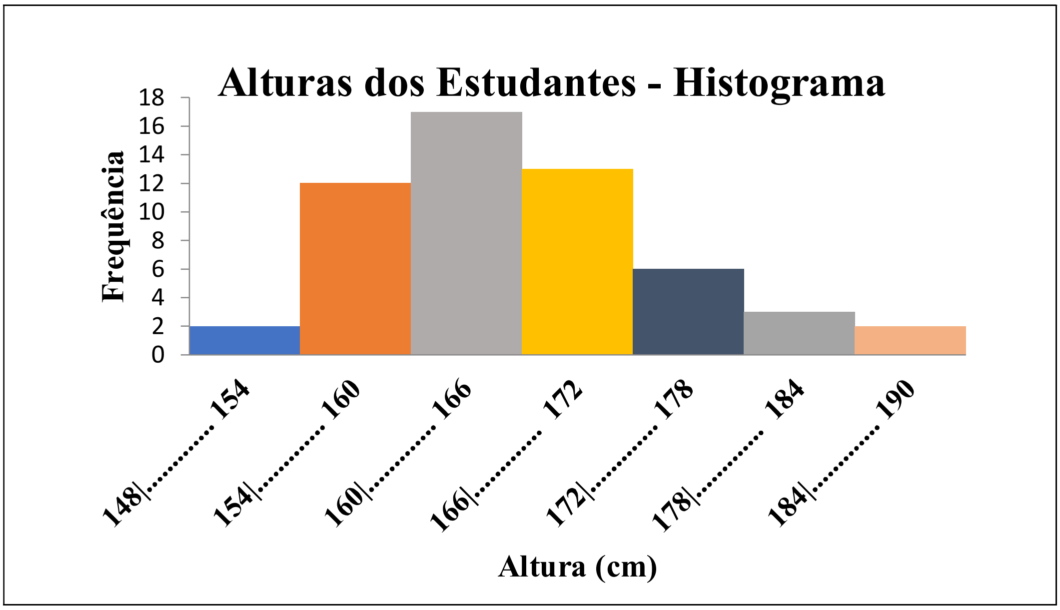


Figura 1.7 – Alturas dos estudantes – Histograma. Fonte: Oliveira (2019).

5.4 TABELAS ESTATÍSTICAS - É um quadro que resume um conjunto de observações (CRESPO, 2002).

- As tabelas são recursos utilizados pela estatística, com o objetivo de organizar e facilitar a visualização e comparação dos dados.
- As tabelas permitem uma visão geral dos valores assumidos pelas variáveis dentro de certos parâmetros.
- A representação tabular é uma apresentação numérica dos dados. Nesse tipo de representação, os dados são dispostos em linhas e colunas e distribuídos de modo ordenado, segundo algumas regras práticas ditadas pelo Conselho Nacional de Estatística e pelo Instituto Brasileiro de Geografia e Estatística (IBGE).
- A integração de valores que temos nas tabelas permite-nos, ainda, a utilização de representações gráficas, as quais, normalmente, são uma forma mais benéfica e elegante de demonstrar as características que estão sendo analisadas.
- Uma das vantagens das tabelas estatísticas é a de condensar, de forma consistente, as informações necessárias ao estudo desejado.

5.4.1 Elementos de uma tabela (CRESPO, 2002)

a) Corpo → Conjunto de linhas e colunas que contêm informações sobre a variável em estudo;

b) Título → Conjunto das informações (mais completas possíveis) que respondem às perguntas: O quê? Quando? Onde? localizado no topo da tabela;

c) Cabeçalho → Parte superior da tabela que especifica o conteúdo das colunas;

d) Coluna indicadora → Parte da tabela que especifica o conteúdo das linhas;

e) Casa ou célula → Espaço destinado a um só número;

f) Linhas → Retas imaginárias que facilitam a leitura, no sentido horizontal, de dados que se inscrevem nos seus cruzamentos com as colunas;

Há ainda a considerar os elementos complementares da tabela, que são a fonte[10], as notas e as chamadas, colocados de preferência, no seu rodapé.

Os lados direito e esquerdo de uma Tabela oficial devem ser abertos.

[10] **Fonte.** Indicação da entidade responsável pelo fornecimento dos dados ou pela sua Tabela.

Exemplo 1.14

Elementos de uma Tabela.

> **Nota:**
>
> De acordo com a Resolução 886 da Fundação IBGE, nas casas ou células, devemos colocar (CRESPO, 2002):
>
> - **Um traço horizontal** (—) quando o valor for zero, não só quanto à natureza das coisas, como quanto ao resultado do inquérito.
> - **Três pontos** (…) quando não temos os dados.
> - **Um ponto de interrogação** (?) quando temos dúvida quanto à exatidão de determinado valor.
> - **Zero** (0) quando o valor é muito pequeno para ser expresso pela unidade utilizada. Se os valores são expressos em numerais decimais, precisamos acrescentar à parte decimal um número correspondente de zeros (0,0; 0,00; 0,000; ...).

CAPÍTULO II - NOÇÕES DE AMOSTRAGEM

1. TIPOS DE AMOSTRAGEM - É a coleta das informações de parte da população chamada amostra mediante métodos adequados de seleção dessas unidades.

Amostragem é considerada uma técnica especial de escolher amostras, de forma a garantir o acaso[11] na escolha. Assim, cada elemento da população tem a mesma chance de ser escolhido, o que garante à amostra um caráter de representatividade da população.

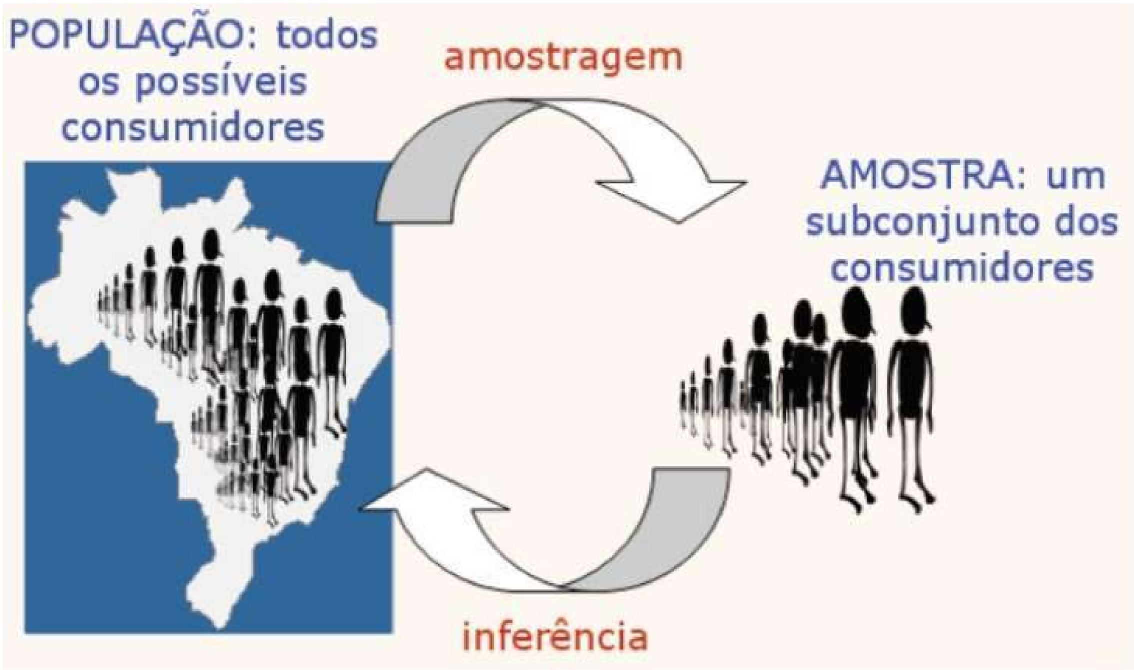


Figura 2.1 – Noções de amostragem. Fonte: VIEIRA; BESSEGATO (2013).

Existem técnicas adequadas para recolher amostras, de forma a garantir (tanto quanto possível) o sucesso da pesquisa e dos resultados. Definidos os objetivos e a população a ser estudada, deve-se pensar em como será constituída a amostra dos dados e quais as características ou variáveis a serem estudadas, ou seja, para cada tipo de amostragem existem técnicas de seleção da amostra. Distinguiremos dois tipos de amostragem: a probabilística e a não-probabilística.

[11] **Acaso** (do latim a casu, sem causa) é algo que surge ou acontece a esmo, sem motivo ou explicação aparente. Acaso = quando algo não depende de mim.

1.1 AMOSTRAGENS PROBABILÍSTICA - Se todos os elementos da população tem probabilidade conhecida, e diferente de zero, de ser incluídos na amostra, o que garante a representatividade da amostra em relação à população.

- Segundo essa definição, a amostragem probabilística implica um sorteio com regras bem determinadas, cuja realização só será possível se a população for finita e totalmente acessível.
- Assim, se N for o tamanho da população, a probabilidade de cada elemento ser selecionado será 1/N.
- Esse método garante cientificamente a aplicação das técnicas estatísticas de inferências. Somente com base em amostragens probabilísticas é que se podem realizar inferências ou induções sobre a população a partir do conhecimento da amostra.

Pode ser: aleatória, sistemática, estratificada e por conglomerado.

1.1.1 Amostragem casual ou aleatória simples - Este tipo de amostragem é equivalente a um sorteio lotérico.

- Sendo **N** o número de elementos da população, e **n** o número de elementos da amostra, cada elemento da população tem probabilidade **n/N** de pertencer a amostra. A essa relação **n/N,** denomina-se fração de amostragem.
- Na prática, a amostragem casual ou aleatória simples pode ser realizada numerando-se a população de 1 a N, sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer, k números dessa sequência, os quais corresponderão aos elementos pertencentes à amostra.
- Um instrumento útil para se realizar o sorteio acima descrito é a Tabela de Números Aleatórios (Tabela 2.1), construída de modo que os dez algarismos (0 a 9) são distribuídos ao acaso nas linhas e colunas. Ilustraremos sua utilização com um exemplo.

Tabela 2.1 - Tabela de números aleatórios.

5 7 7 2 0 0 3 9 8 4 8 4 4 1 7 9 6 7 7 1 4 0 2 1 1 3 9 7 5 6 4 9 8 6 5 4 0 8 9 3 2 9 6 8 7 4 5 4 8 3
2 8 8 0 5 3 5 1 5 9 0 9 9 3 9 8 8 7 5 8 7 0 2 7 7 1 7 7 1 7 0 6 3 2 0 2 7 8 6 2 1 6 7 4 6 9 6 5 1 7
9 2 5 9 1 8 5 2 8 7 3 0 4 8 8 6 9 7 4 8 3 5 2 5 1 8 8 8 7 4 0 3 6 2 9 8 3 8 5 8 6 5 8 6 4 2 4 1 0 3
9 0 3 8 1 2 9 1 7 4 3 0 1 9 7 5 8 9 0 7 5 0 6 4 1 5 5 9 7 1 8 8 1 3 7 4 9 5 3 0 5 2 7 8 3 0 1 1 7 5
8 0 9 1 1 6 9 4 6 7 5 8 6 0 8 2 0 6 6 6 9 0 4 7 5 6 1 8 4 6 4 5 1 1 1 2 3 5 3 2 4 5 5 0 4 1 1 3 4 3
2 2 0 1 7 0 3 1 3 2 9 6 9 1 9 2 7 5 4 0 1 6 5 4 2 9 7 2 7 4 9 9 0 0 9 5 9 7 6 1 0 0 9 8 2 4 3 0 0 7
5 6 2 4 1 0 0 4 3 0 2 0 4 6 2 9 9 0 5 3 5 3 1 1 0 5 8 4 4 1 2 1 6 4 7 9 1 9 7 6 2 9 5 1 6 2 6 0 6 6
7 9 4 4 9 2 6 2 0 2 9 6 8 6 6 4 3 0 0 0 9 4 5 6 6 9 3 0 2 0 5 9 8 7 8 7 3 5 4 4 2 2 5 0 9 7 7 8 1 9
5 3 9 9 6 6 4 5 0 8 8 9 7 8 5 0 7 7 5 3 3 7 2 5 7 7 4 1 2 7 6 2 3 8 0 2 2 3 5 7 6 2 0 1 4 1 6 0 3 5
1 8 9 2 8 7 3 5 8 8 5 5 0 5 2 1 3 6 5 1 3 9 2 8 5 0 1 4 6 6 8 5 7 9 3 0 1 9 7 9 7 2 6 6 6 4 3 1 4 5
5 3 0 8 5 8 9 6 6 3 0 5 6 1 2 5 7 0 2 2 5 0 4 1 2 8 9 6 6 2 6 6 4 3 6 3 0 6 6 3 0 1 3 2 7 9 8 5 2 2
0 3 5 8 8 0 2 9 2 8 7 6 8 9 5 1 1 8 2 4 8 8 8 9 4 6 4 7 4 8 5 9 1 9 2 9 8 7 0 3 1 0 3 3 9 9 6 7 1 2
2 7 0 7 8 1 8 8 6 5 6 9 4 9 9 8 0 0 2 8 0 4 7 0 5 1 3 0 0 1 4 7 1 8 9 7 3 3 2 1 8 5 8 2 4 5 4 3 2 4
0 5 2 1 0 8 5 9 0 1 0 6 2 2 2 4 9 8 9 1 8 1 1 7 5 5 4 4 6 6 1 6 0 7 7 3 0 7 6 6 1 0 1 2 3 1 7 8 5 8
4 0 3 6 1 3 2 7 8 4 3 0 8 2 3 3 3.6 3 9 6 9 4 2 0 5 5 8 6 4 6 1 1 2 3 3 8 9 2 7 8 9 5 2 6 6 7 1 9 3
5 4 6 0 2 5 2 8 8 5 8 8 2 0 0 0 1 0 5 9 6 1 0 5 3 6 6 1 3 3 7 2 0 1 0 1 1 9 0 1 6 1 1 0 5 1 2 0 9 1
7 1 5 1 6 3 4 0 7 6 7 1 1 1 7 3 7 3 5 2 3 7 3 1 6 0 4 5 8 8 9 2 7 3 4 3 7 1 2 8 0 4 9 8 0 9 0 2 4 8
6 1 0 2 0 1 8 1 7 3 9 2 6 0 6 6 7 3 5 8 5 3 3 4 4 2 6 8 2 6 3 8 3 4 0 3 2 7 4 4 9 6 0 4 4 6 6 5 9 3
8 2 5 5 9 3 1 3 4 6 3 0 9 5 2 6 5 5 0 6 9 6 1 7 6 5 9 1 7 2 3 9 7 9 9 6 1 2 4 9 5 2 8 0 6 3 2 6 9 9
8 9 9 8 5 4 1 4 2 1 7 4 1 3 5 7 6 8 1 9 8 6 2 8 6 0 8 9 4 7 3 3 1 5 2 6 2 8 7 7 4 5 3 8 4 8 0 8 0 8
0 0 9 9 8 4 8 4 1 4 6 7 9 5 1 3 7 7 5 8 9 0 1 4 5 0 7 9 4 2 7 3 6 3 3 1 0 6 6 0 4 3 4 0 1 2 5 5 0 4
6 2 4 1 5 0 7 8 2 0 4 8 0 5 8 8 4 3 5 2 9 8 0 3 1 9 9 3 9 2 0 3 0 4 9 7 2 5 8 4 9 5 9 5 0 3 6 3 3 1
9 4 2 7 9 0 6 9 2 4 6 8 0 9 9 2 1 1 8 6 0 7 6 3 8 3 1 9 3 2 9 9 5 1 1 5 5 5 7 1 0 9 2 7 0 2 6 7 0 0
4 4 8 9 2 9 2 8 8 4 3 6 2 8 2 5 1 5 8 2 8 7 7 4 1 8 9 7 2 5 7 6 1 0 6 3 2 6 7 6 0 2 2 6 7 4 5 3 2 8
9 7 3 0 7 6 9 5 3 3 2 1 1 0 5 5 2 6 9 5 6 6 6 5 5 2 0 4 9 9 3 6 5 8 4 8 0 3 0 8 9 3 6 3 5 8 1 7 9 6
3 9 1 6 5 8 0 4 4 4 8 0 1 5 5 9 5 9 8 3 9 0 9 5 5 4 6 6 8 1 8 4 3 9 6 0 8 5 3 8 8 8 6 6 3 3 3 5 6 9
6 0 7 8 1 1 0 3 2 6 6 7 5 0 3 4 0 9 6 1 3 1 3 0 2 0 7 6 9 3 6 6 3 0 8 3 5 1 0 9 3 3 8 3 6 4 7 6 0 5
0 3 1 9 2 3 4 7 6 2 8 9 5 7 7 7 9 1 3 3 8 8 4 7 6 0 5 9 3 7 5 4 3 9 4 8 7 7 6 7 4 9 8 5 3 8 4 3 9 1
4 1 2 8 5 2 6 7 5 6 2 5 3 9 5 9 9 6 6 5 5 1 3 6 9 0 3 2 2 2 3 9 3 3 0 5 2 2 9 9 0 3 3 9 9 7 9 6 9 9
7 7 5 4 9 8 5 0 3 9 2 5 3 7 4 2 5 2 9 7 1 0 0 3 5 6 0 4 9 2 8 1 6 6 8 6 7 0 0 1 4 8 8 9 5 5 8 2 1 0
2 8 6 3 4 1 6 1 9 1 6 4 2 4 8 3 8 1 3 7 3 4 4 8 8 3 2 7 9 6 3 8 7 1 6 9 7 3 0 6 7 7 5 0 2 5 6 4 6 0
7 4 2 4 4 8 8 5 4 0 1 2 3 3 5 9 6 7 5 0 1 4 9 8 1 4 2 6 4 2 7 9 7 9 1 3 5 2 8 9 6 9 7 8 8 0 4 4 7 1
0 0 2 4 0 3 3 7 9 6 4 6 6 8 7 5 0 5 3 2 4 2 1 6 6 3 3 3 2 8 9 7 2 6 3 6 4 7 2 7 7 3 6 5 3 8 3 4 4 6
0 0 2 4 0 3 3 7 9 6 4 6 6 8 7 5 0 5 3 2 4 2 1 6 6 3 3 3 2 8 9 7 2 6 3 6 4 7 2 7 7 3 6 5 3 8 3 4 4 6
0 5 4 1 4 7 6 9 6 9 4 5 3 6 1 6 7 1 1 8 9 5 5 1 9 7 2 2 0 4 1 3 2 3 9 6 5 8 6 0 0 3 6 9 4 8 7 9 8 3
6 2 6 9 8 4 9 7 9 7 4 7 2 3 6 6 5 1 5 6 1 3 0 8 6 9 1 1 5 2 7 5 5 9 2 6 8 6 8 1 8 0 4 3 0 0 9 8 9 2

NOTA: 0 - 10   00 – 100   000 – 1000   etc.

Fonte: www.passeidireto.com. 2019.

Exemplo 2.1

Vamos obter uma amostra, de 10%, representativa para a pesquisa da estatura de 90 alunos de uma escola:

Procedimentos:

a) Numeramos os alunos de 1 a 90.

b) Escrevemos os números dos alunos, de 1 a 90, em pedaços iguais de papel, colocando-os dentro de uma caixa (urna). Em seguida, agitamos sempre a caixa para misturar bem os pedaços de papel e retiramos um a um, nove números que formarão a amostra. Neste caso, 10% da população (CRESPO, 2002).

Nota:
Quando o número de elementos da amostra é muito grande, esse tipo de sorteio torna-se muito trabalhoso.

c) Sorteamos um algarismo qualquer da amostra, a partir do qual iremos considerar números de dois, três ou mais algarismos, conforme nossa necessidade. Os números, assim obtidos, irão indicar os elementos da amostra.
d) A leitura da tabela pode ser feita horizontalmente (da direita para esquerda ou vice-versa), verticalmente (de cima para baixo ou vice-versa), diagonalmente (no sentido ascendente ou descendente) ou formando o desenho de uma letra qualquer. A opção, deve ser feita antes de iniciar o processo.
e) Considerando a 18ª linha (para o nosso exemplo), tomamos os números de dois algarismos (tantos algarismos quantos formam o maior número da população), obtendo:
f) 02 01 81 73 92 60 66 73 58 53 34
g) Evidentemente, o numeral 92 será desprezado, pois não consta da população como será também abandonado um numeral que já tenha aparecido. Temos então:
h) 02 01 81 73 60 66 58 53
i) Medindo as alturas dos alunos correspondentes aos números sorteados, obteremos uma amostra das estaturas dos noventa alunos.

1.1.2 Amostragem proporcional ou estratificada - Quando a população se divide em subpopulações ou estratos.

Como é provável que a variável em estudo apresente, de estrato em estrato um comportamento heterogêneo e, dentro de cada estrato, um comportamento homogêneo, convém que o sorteio dos elementos da amostra leve em consideração tais estratos.

Dessa forma, ao ser empregado este tipo de amostragem, além de considerar a existência dos estratos, obtém os elementos da amostra proporcional ao número de elementos dos mesmos.

Exemplo 2.2

Supondo, no exemplo anterior, que, dos noventa alunos, 54 sejam meninos e 36 sejam meninas, vamos obter a amostra proporcional estratificada.

Procedimentos:

a) Teremos dois estratos (sexo masculino e sexo feminino) e queremos uma amostra de 10% da população. Logo, temos (Tabela 2.2):

Tabela 2.2- Amostragem proporcional ou estratificada.

Sexo	**População**	**10%**	**Amostra**
Meninos	54	(10 x 54)/100 = 5,4	5
Meninas	36	(10 x 36)/100 = 3,6	4
Total	90	(10 x 90)/100 = 9,0	9

b) Numeramos os alunos de 1 a 90, sendo que de 1 a 54 correspondem meninos e de 55 a 90, meninas. Tomando na tabela de Números Aleatórios, a primeira e segunda colunas da esquerda, de cima para baixo, obtemos os seguintes números:

57 28 92 90 80 22 56 79 53 18 53 03 27 05 40

Temos, então:

28 22 53 18 03 → para os meninos;

57 90 80 56 → para as meninas.

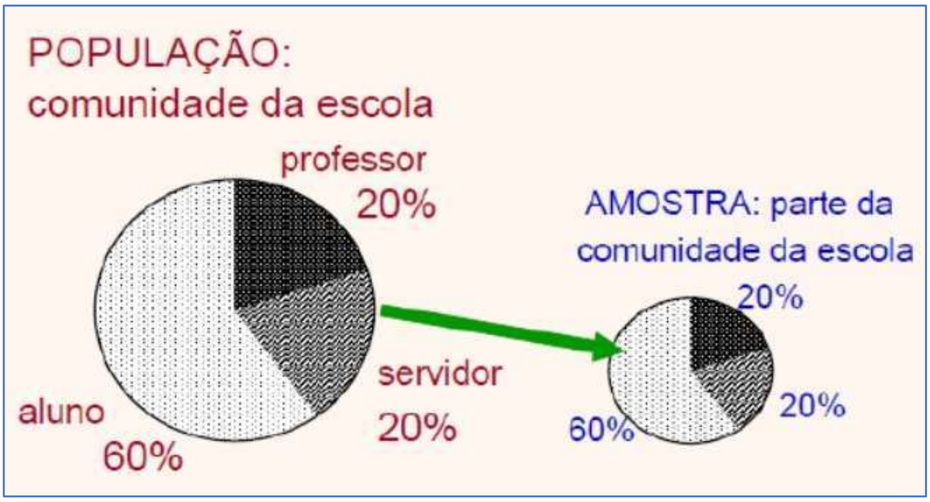


Figura 2.2 – Ilustração de uma amostragem estratificada proporcional.
Fonte: Wordpress.com. 2019.

1.1.3 Amostragem estratificada uniforme - Não utiliza o critério de proporcionalidade, pois se seleciona a mesma quantidade de elementos de cada estrato, devendo ser usada para comparar os estratos ou obter estimativas separadas para cada estrato.

Exemplo 2.3

Uma empresa de automação conta com 480 funcionários, dos quais 288 são do sexo feminino e os 192 restantes do sexo masculino. Considerando a variável "sexo" para estratificar essa população, vamos obter uma amostra estratificada uniforme de 50 funcionários.

Solução

Supondo que haja homogeneidade dentro de cada categoria, pode-se obter amostra estratificada uniforme de 50 funcionários com a seleção de 25 elementos de cada estrato.

Tabela 2.3 – Amostragem estratificada uniforme.

Estrato (por sexo)	População	Amostra estratificada uniforme
Feminino	288	$n_1 = 25$
Masculino	192	$n_2 = 25$
Total	480	50

1.1.4 Amostragem sistemática - Quando os elementos da população já se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, não há necessidade de construir o sistema de referência.

- São exemplos, os prontuários médicos de um hospital, os prédios de uma rua etc. Nestes casos, a seleção dos elementos que constituirão a amostra pode ser feita por um sistema imposto pelo pesquisador.
- No caso de uma linha de produção, podemos, a cada dez itens produzidos, retirar um para pertencer a uma amostra da produção diária. Neste caso, estaríamos fixando o tamanho da amostra em 10% da população.

Exemplo 2.4

Suponhamos uma rua com 900 prédios, dos quais desejamos obter uma amostra formada por 50 prédios. Podemos, neste caso, usar o seguinte procedimento: como 900/50 = 18, escolhemos por sorteio casual um número de 01 a 18 (inclusive), o qual indicaria o primeiro elemento sorteado para a amostra; os demais elementos seriam periodicamente considerados de 18 em 18. Assim, se o número sorteado fosse 4, tomaríamos, pelo lado direito da rua, o 4º prédio, o 22º, o 40º, etc., até voltarmos ao início da rua, pelo lado esquerdo.

Nota:
A principal vantagem da amostragem sistemática está na grande facilidade na determinação dos elementos da amostra. O perigo em adotá-la está na possibilidade de existirem ciclos de variação da variável de interesse, especialmente se o período desses ciclos coincidir com o período da retirada dos elementos da amostra. Por outro lado, se a ordem dos elementos na população não tiver qualquer relacionamento com a variável de interesse, então a amostragem sistemática terá efeitos equivalentes à casual simples, podendo ser utilizada sem restrições.

1.1.5 Amostragem por conglomerados (clusters) – É um procedimento de amostragem probabilística em que os elementos da população são selecionados aleatoriamente de forma natural por grupos (clusters). Os elementos da amostragem são selecionados a partir da população de maneira individual, um de cada vez.

- Usada quando a população pode ser dividida em grupos homogêneos selecionados aleatoriamente;
- A amostragem é feita a partir dos grupos e não dos indivíduos da população.
- Agrupamentos típicos são quarteirões, famílias, organizações, agências, edifícios, etc.

Exemplo 2.5

Suponha que desejamos selecionar uma amostra de chefes de família de uma cidade e não dispomos de uma relação (cadastro) de todas as residências.

Prática

- Podemos construir uma relação, numerando, em um mapa, cada quarteirão da cidade. A lista de todos os quarteirões poderá ser utilizada para a seleção de uma amostra aleatória simples de quarteirões, cada um deles representando um conglomerado (cluster) de residências.
- Após a amostragem dos conglomerados, são entrevistados todos os chefes de família dos quarteirões escolhidos.

- A amostragem por conglomerados é uma amostragem aleatória simples em que as unidades amostrais são os conglomerados.

1.2 AMOSTRAGENS NÃO-PROBABILÍSTICAS – São amostragens em que há uma escolha deliberada dos elementos da amostra. Não é possível generalizar os resultados da amostra para a população, pois amostras não probabilísticas não garantem a representatividade da população. Apresentamos a seguir alguns casos de amostragem não-probabilística.

1.2.1 Acidental ou conveniência - Amostra formada por aqueles elementos que vão aparecendo, ou seja, os entrevistados são acidentalmente escolhidos.

- Indicada para estudos exploratórios.
- Frequentemente utilizados em supermercados para testar produtos;
- Geralmente utilizada em pesquisas de opinião.

Exemplo 2.6

Pesquisas de opinião em praças públicas, ruas de grandes cidades.

1.2.2 Amostragem intencional - De acordo com determinado critério, é escolhido, intencionalmente, um grupo de elementos que irão compor a amostra. O entrevistador dirige-se, intencionalmente, a grupos de elementos dos quais deseja saber a opinião.

Exemplos 2.7

a) Quando de um estudo sobre automóveis, o pesquisador procura apenas oficinas.
b) Numa pesquisa sobre preferência por determinado cosmético, o pesquisador se dirige a um grande salão de beleza e entrevista as pessoas que ali se encontram.

1.2.3 Amostragem por quotas - Um dos métodos de amostragem mais comumente usados em levantamentos de mercado e em prévias eleitorais. Ele abrange três fases:

- **1ª Fase -** classificação da população em termos de propriedades que se sabe, ou presume, serem relevantes para a característica a ser estudada;
- **2ª Fase -** determinação da proporção (%) da população para cada característica (propriedade), relevante ao estudo;
- **3ª Fase** - fixação de quotas para cada observador, ou entrevistador, a quem cabe a responsabilidade de selecionar interlocutores, ou entrevistados, de modo que a amostra

total observada ou entrevistada contenha iguais proporções de cada característica que está sendo avaliada.

Exemplo 2.8

Admitamos que se deseja realizar uma pesquisa sobre intenções de votos em determinado município com 30.000 eleitores. Busca-se conhecer a intenção de votos dos homens e mulheres em função de seus níveis de escolaridade. O tamanho amostral foi dimensionado em 400 eleitores.

Procedimentos:

1) As variáveis **sexo e nível de escolaridade** são relevantes para o estudo.
2) Por meio dos registros dos cartórios eleitorais, ou com base em dados de outras pesquisas, são dimensionadas as porcentagens populacionais para cada um dos níveis das variáveis relevantes ao estudo.

Sejam as proporções:

Tabela 2.4 – Amostragem por quotas.

Em relação ao sexo	
Sexo	**%**
Masculino	60
Feminino	40

Em relação a escolaridade	
Escolaridade	**%**
Analfabeto	10
Ensino Fundamental	50
Ensino Médio	30
Superior	10

3) Dimensionamento das quotas.

- Como não dispomos das porcentagens de homens e mulheres para cada nível de escolaridade, vamos admitir que sejam (60% homens e 40% mulheres) para todos os níveis. Assim: serão escolhidos 240 homens (60% de 400) e 160 mulheres (40% de 400).
- Dos homens selecionados: 24 deverão ser eleitores analfabetos (10% de 240); 120 com ensino fundamental;72 com ensino médio; e 24 que tem o curso superior.
- Quanto às mulheres, serão escolhidas: 16 analfabetas; 80 com ensino fundamental; 48 com ensino médio; e 16 com curso superior.

Em síntese, o plano amostral será:

Tabela 2.5A – Dimensionamento das quotas.

Escolaridade	Homens	Mulheres
Analfabetos	24	16
Ensino Fundamental	120	80
Ensino Médio	72	48
Superior	24	16

Vamos admitir que temos quatro entrevistadores. Logo, a quota de cada um poderá ser:

Tabela 2.5B - Dimensionamento das quotas.

Escolaridade	Homens	Mulheres	Soma (Σ)
Analfabetos	6	4	10
Ensino Fundamental	30	20	50
Ensino Médio	18	12	30
Superior	6	4	10
	$\Sigma = 60$	**Σ=40**	**100**

Para que a amostra represente “todo o município”, os quatro entrevistadores poderiam cumprir suas quotas em cada uma das regiões: Norte, Sul, Leste e Oeste.

Nota:

O censo ou recenseamento demográfico é um estudo estatístico referente a uma população que possibilita o recolhimento de várias informações, tais como o número de homens, mulheres, crianças e idosos, onde e como vivem as pessoas. É o exame oficial, completo de toda população.

- Esse estudo é realizado, normalmente, de dez em dez anos, na maioria dos países.
- Quanto maior a amostra, mais precisa e confiáveis deverão ser as induções feitas sobre a população.
- Na prática, esta conclusão muitas vezes não acontece, pois o emprego de amostras com certo rigor técnico pode levar a resultados mais confiáveis ou até mesmo melhores do que os que seriam obtidos através de um censo.

CAPÍTULO III - SÉRIES ESTATÍSTICAS

1. SÉRIE ESTATÍSTICA - É toda e qualquer coleção de dados estatísticos referidos a uma mesma ordem de classificação: quantitativa.

- Em um sentido amplo, série é uma sucessão de números referidos a qualquer variável.
- Em sentido mais estreito, pode-se dizer que uma série estatística é uma sucessão de dados estatísticos referidos a caracteres qualitativos, ao passo que uma sucessão de dados estatísticos referidos a caracteres quantitativos configurará uma seriação.

Para diferenciar uma série estatística de outra, leva-se em conta, três caracteres presentes na tabela que as apresenta:

a) **A época** (fator temporal ou cronológico) a que se refere o fenômeno analisado.
b) **O local** (fator espacial ou geográfico) onde o fenômeno acontece.
c) **O fenômeno** (espécie do fato ou fator especificativo) que é descrito.

As séries estatísticas podem ser de quatro tipos, conforme varie um desses caracteres ou fatores. Costuma-se dividi-las em dois grupos: o das séries homógradas e o das heterógradas.

Nota:
Os quatro tipos de séries estatísticas diferenciam-se de acordo a variação de um desses três elementos: época, local, fato.

1.1 SÉRIE HOMÓGRADA - são aquelas em que a variável descrita apresenta variação discreta ou descontínua. Podem ser do tipo temporal, geográfica e a série específica.

1.1.1 Séries históricas, cronológicas, temporais ou marchas – identifica-se pelo caráter variável do fator cronológico. Assim, deve-se ter:

- Elemento Variável: época (fator cronológico)
- Elementos Fixos: local (fator geográfico); fenômeno (fator especificativo).

Exemplo 3.1

O diretor de marketing da empresa G.L.T. S.A, fabricante de componentes eletrônicos, deseja examinar a evolução de suas vendas em 2018, mês a mês. Para tanto, solicitou ao Departamento de Análise de Mercado a tabela da qual constam os valores de vendas no período desejado.

Tabela 3.1. – G.L.T. – Indústria de componentes eletrônicos. Vendas mercado interno 2018

Meses	Vendas (em milhares de R$)
Janeiro	2.300
Fevereiro	1.800
Março	2.200
Abril	2.210
Maio	2.360
Junho	2.600
Julho	2.690
Agosto	3.050
Setembro	3.500
Outubro	3.440
Novembro	3.100
Dezembro	2.760
TOTAL ANUAL	**32.010**

Fonte: Fictícia.

Nota:

- O único caráter variável é o tempo, representado pelos meses.
- Da mesma forma, os valores sucessivos apresentados pelo censo demográfico de uma cidade formarão a marcha da população dessa cidade.

1.1.2 Séries geográficas, espaciais, territoriais ou de localização – Apresenta como elemento ou caráter variável somente o fator geográfico.

- Elemento variável: local (fator geográfico).
- Elementos fixos: época (fator cronológico); fenômeno (fator especificativo).

Exemplo 3.2

Se o diretor de marketing da G.L.T. S.A deseja saber o comportamento das vendas dessa empresa efetuadas nos vários Estados do Brasil, durante o exercício fiscal de 2018, o fator diferenciador das vendas seria geográfico.

Tabela 3.2 - G.L.T. – Indústria de componentes eletrônicos. Vendas por Unidade da Federação – 2018

Unidades da Federação	Vendas (em milhares de R$)
Minas Gerais	4.000
Paraná	2.230
Rio Grande do Sul	6.470
Rio de Janeiro	8.300
São Paulo	10.090
Outros	920
TOTAL – BRASIL	**32.010**

Fonte: Fictícia.

1.1.3 Séries específicas ou categóricas - É a série cujos dados estão em correspondência com a espécie, ou seja, variam com o fenômeno.

a) Elemento variável: fenômeno (fator especificativo);

b) Elementos fixos: época (fator cronológico); local (fator geográfico).

Exemplo 3.3

Suponha que o diretor de marketing esteja agora interessado em conhecer o comportamento das vendas de cada um de seus produtos, os quais foram agrupados em três categorias ou linhas, dada a grande variedade de componentes fabricados pela empresa.

Tabela 3.3 - G.L.T. – Indústria de componentes eletrônicos. Vendas por Linha de Produto – 2018

Linha do Produto	Vendas (em milhares de R$)
Linha A	6.950
Linha B	9.310
Linha C	15.750
TODOS OS PRODUTOS	**32.010**

Fonte: Fictícia.

1.2 SÉRIE HETERÓGRADA – são aquelas na qual o fenômeno ou o fato apresenta gradações ou subdivisões. Embora fixo, o fenômeno varia em intensidade. Exemplo: A distribuição de frequência ou seriação.

1.2.1 Séries mistas, conjugadas ou tabela de dupla entrada - São apropriadas à apresentação de duas ou mais séries de maneira conjugada, havendo duas ordens de classificação: uma horizontal (linha) e outra vertical (coluna). O exemplo abaixo é de uma série geográfica-temporal.

Exemplo 3.4

Tabela 3.4 - ABC VEÍCULOS LTDA. Vendas no 1º bimestre de 2006

Filiais	Janeiro/2006	Fevereiro/2006
São Paulo	10.000	3.000
Rio de Janeiro	12.000	5.000
TOTAL	**22.000**	**8.000**

Fonte: Fictícia.

Nota:
Cada linha, encabeçada pela filial – São Paulo, Rio de Janeiro, constitui uma série geográfica. Por outro lado, cada coluna encabeçada pela época – janeiro 2006, fevereiro 2006, representa uma série temporal. O resultado é uma série geográfica-temporal.

1.3 DISTRIBUIÇÃO DE FREQUÊNCIAS – SERIAÇÃO – São séries heterógradas. Todos os elementos – época, local e fenômeno – são fixos, estando os dados agrupados de acordo com a intensidade ou variação quantitativa do fenômeno. Os dados estatísticos são dispostos ordenadamente em linhas (horizontal) e colunas (vertical).

Exemplo 3.5

Um exemplo de distribuição de frequência encontra-se na Tabela 3.5, que descreve as vendas da empresa G.L.T.S.A, agrupadas em classes de faturamento e a análise do número de meses em que se verificou os vários faturamentos incluídos nas diferentes classes.

Tabela 3.5 - G.L.T. – Indústria de componentes eletrônicos. Número de meses segundo o faturamento

Vendas (em milhares de reais)	Número de Meses
1.800 a 2.199	1
2.200 a 2.599	4
2.600 a 2.999	3
3.000 a 3.399	2
3.400 a 3.799	2
TOTAL DE MESES	**12**

Fonte: Fictício.

CAPÍTULO IV - DISTRIBUIÇÃO DE FREQUÊNCIA

1. DISTRIBUIÇÃO DE FREQUÊNCIA – É um tipo de tabela que condensa uma coleção de dados conforme as frequências (repetições de seus valores).

1.1 TABELA PRIMITIVA OU DADOS BRUTOS - são os dados coletados durante a pesquisa e que ainda não foram organizados, ordenados.

Exemplo 4.1

Suponhamos termos feito uma coleta de dados relativos às estaturas de 40 alunos, que compõem uma amostra dos alunos de um colégio A, resultando a seguintes tabela de valores.

Tabela 4.1 - ESTATURAS DE 40 ALUNOS DO COLÉGIO A

166	160	161	150	162	160	165	167	164	160
162	161	168	163	156	173	160	155	164	168
155	152	163	160	155	155	169	151	170	164
154	161	156	172	153	157	156	158	158	161

Fonte: CRESPO (2002).

1.2 ROL - é a ordenação dos valores obtidos (dados brutos) em **ordem crescente ou decrescente** de grandeza numérica ou qualitativa.

Exemplo: Ordenando os dados brutos da Tabela 4.1.

Tabela 4.2 - ESTATURAS DE 40 ALUNOS DO COLÉGIO A

150	154	155	157	160	161	162	164	166	169
151	155	156	158	160	161	162	164	167	170
152	155	156	158	160	161	163	164	168	172
153	155	156	160	160	161	163	165	168	173

Fonte: CRESPO (2002).

Essa ordenação dos dados proporciona algumas vantagens concretas com relação à sua forma original. Ela torna possível visualizar, de forma bem ampla, com relativa facilidade os seguintes dados:

- A menor estatura 150 cm e a maior estatura 173 cm;
- A amplitude de variação foi de 173 – 150 = 23 cm;
- A ordem que um valor particular da variável ocupa no conjunto;
- Observa-se que há uma concentração das estaturas em algum valor entre 160 cm e 165 cm, e que há poucos valores abaixo de 155 cm e acima de 170 cm.

a) O número de alunos que fica relacionado a determinado valor da variável denominamos frequência. Obtemos, assim, uma tabela que recebe o nome de distribuição de frequência.

b) Chama-se frequência de uma classe o número de valores da variável pertencentes à classe.

Nota:
A estatura 150 cm, tem 1 (uma) frequência. A estatura 160 cm, tem 5 (cinco) frequências, e assim por diante.

1.3 TABELA DE FREQUÊNCIAS – São representações nas quais os valores se apresentam em correspondência com suas repetições, evitando-se assim que eles apareçam mais de uma vez na tabela, como ocorre com o rol.

Exemplo 4.2

Distribuição dos dados da Tabela 4.2.

Tabela 4.3 - Estaturas de 40 alunos do Colégio A.

Estaturas (cm)	**Frequência**
150 ├ 154	4
154 ├ 158	9
158 ├ 162	11
162 ├ 166	8
166 ├ 170	5
170 ├ 174	3
	Σ = 40

Fonte: CRESPO (2002).

Notas:

- O símbolo ├ usado na Tabela 4.3. Exemplo 150├ 154 é um intervalo fechado à esquerda e aberto à direita, tal que $150 \leq x < 154$.

Observação:

" – ": não inclui os valores da direita nem os da esquerda.

" ┤ ": inclui o valor da direita, mas não o da esquerda.

" ├ ": inclui o valor da esquerda, mas não o da direita.

" ├┤ ": inclui tanto os valores da direita quanto os da esquerda.

- A tabela de frequências proporciona uma apresentação esteticamente mais vantajosa dos dados, facilitando ainda a verificação do comportamento do fenômeno.
- As tabelas de frequências podem representar tanto valores individuais como valores agrupados em classes.
- Ao agruparmos os valores da variável em classes, ganhamos em simplicidade, mas perdemos em pormenores.
- Na Tabela 4.2, podemos verificar que quatro alunos têm 1,61 cm de altura e que não existe nenhum aluno com 1,71 cm de altura.
- Na Tabela 4.3 não podemos ver se algum aluno tem a estatura 1,59 cm. Sabemos, com segurança, que onze alunos tem estatura compreendida entre 1,58 e 1,62 cm.
- A construção da Tabela 4.3 é para realçar o que há de essencial nos dados e, também, tornar possível o uso de técnicas analíticas para sua total descrição, porque a Estatística tem por finalidade específica analisar o conjunto de valores, desinteressando-se por casos isolados.

2. ELEMENTOS DE UMA DISTRIBUIÇÃO DE FREQUÊNCIA

Para construir uma tabela de frequências, faz-se necessário os termos próprios, bem como o procedimento técnico mais adequado. São eles: Frequência Simples Absoluta, Amplitude Total, Classe, Limites de Classes, Amplitude do intervalo de Classe, Ponto Médio de Classe, Amplitude Amostral.

2.1 FREQUÊNCIA SIMPLES ABSOLUTA (f_i) de uma classe ou de um valor individual é o número de observações correspondentes a essa classe ou a esse valor.

Exemplo 4.3

Na Tabela 4.3: $f_1 = 4$; $f_2 = 9$; $f_3 = 11$; $f_4 = 8$; $f_5 = 5$; $f_6 = 3$.

- A frequência simples é simbolizada por f_i (lê-se: f índice i ou frequência da classe i).
- A soma de todas as frequências é representada pelo símbolo do somatório Σ:

$$\sum_{i=1}^{k} f_i = n$$

Em que:

k é o extremo superior do intervalo de valores do índice i;

f_i é o número de observações de um valor;

n é o número total de valores observados.

- Para a distribuição em estudo (Tabela 4.3.), temos: $\Sigma fi = 40$

2.2 AMPLITUDE TOTAL DA DISTRIBUIÇÃO (AT): é a diferença entre o limite superior da última classe (limite superior máximo) e o limite inferior da primeira classe. (Limite inferior mínimo), ou seja, a diferença entre o maior e o menor valor observado da variável em estudo. AT = $L(max) - l(min)$.

Exemplo 4.4

Se na Tabela 4.3, a maior estatura tivesse sido 174 cm e a menor 150 cm, a amplitude total do conjunto de valores observados seria**:**

$$AT = 174 - 150 = 24 \rightarrow AT = 24 \text{ cm}$$

2.3 CLASSE DE FREQUÊNCIA, ou, simplesmente Classe (***i***): são os intervalos de variação da variável.

- As classes são representadas simbolicamente por ***i***, sendo $i = 1, 2, 3,.., k$ (onde k é o número total das classes da distribuição).
- Na Tabela 4.3, como a distribuição é formada de seis classes, podemos afirmar que $k = 6$.
- Uma determinada classe pode ser identificada por seus extremos ou pela ordem em que ela se encontra na Tabela 4.3.

Classe 150 ├ 154 ou primeira classe ($i = 1$)

Classe 162 ├ 166 ou quarta classe ($i = 4$)

2.3.1 NÚMERO DE CLASSES - INTERVALOS DE CLASSES

Para determinar o número de classes há diversos métodos. A regra de Sturges, um dos métodos, estabelece:

$$\boldsymbol{K = 1 + 3{,}322 \,.\, \log n}$$

Onde:

K = número de classes

n = número total de observações.

> **Nota:**
> A regra de Sturges, nos dá o número de classes em função do número de valores da variável.

Exemplo 4.5

a) Se o número de observações for 500:

$n = 500$

$K = 1 + 3{,}3 \log 500$ (log 500 = 2,69897)

$K = 1 + (3{,}3 \times 2{,}69897) \rightarrow K = 1 + 8{,}906601$

$K = 9{,}906601$ ou, arredondando $\boldsymbol{K = 10}$.

b) Se *n* = 50

$K = 1 + 3{,}3 \log 50$ (log 50 = 1,69897)

$K = 1 + (3{,}3 \times 1{,}69897) \rightarrow K = 1 + 5{,}606601$

$K = 6{,}606601$ ou, arredondando $K \approx 7$

> **Nota:**
> Existem outras fórmulas empíricas que pretendem determinar o número de classes que deve ter a distribuição, como por exemplo: $\boldsymbol{i \approx \sqrt{h}}$ (*h* = amplitude do intervalo de classe). Mesmo conhecendo estes critérios de determinação do número de classes, o analista deverá ter em mente que a escolha dependerá antes da natureza dos dados e da unidade de medida em que eles forem expressos do que de regras muitas vezes arbitrárias e pouco flexíveis.

2.4 LIMITES DE CLASSE (l_i e L_i): são seus valores extremos.

- O menor número é o limite inferior de classe (l_i) e o maior número, limite superior da classe (L_i);
- Na Tabela 4.3 a segunda classe, 154 ├ 158, por exemplo, tem como limite inferior: $l_2 = 154$ e limite superior: $L_2 = 158$

Para a construção de uma tabela de frequências é muito importante a escolha dos limites de classes, de forma que seus pontos médios coincidam, tanto quanto possível, com a concentração dos valores reais.

> **Notas:**
> 1. Os intervalos de classe devem ser escritos, de acordo com a Resolução 886/66 do IBGE, em termos desta quantidade até menos aquela, empregando, para isso, o símbolo ├ (inclusão de l_i e exclusão de L_i). Assim, o indivíduo com uma estatura de 158 cm está incluído na terceira classe ($i = 3$) e não na segunda classe.
> 2. A escolha dos limites dos intervalos deverão ser tais que forneçam, na medida do possível, para pontos médios, números que facilitem os cálculos – números naturais.

2.5 AMPLITUDE DO INTERVALO DE CLASSE (h_i): é a medida (comprimento) do intervalo que define a classe, sendo obtida através da diferença entre seus limites superior e inferior. Assim, $hi = Li - li$

Exemplo 4.6

Na distribuição da Tabela 4.3, tem-se a Amplitude do Intervalo de Classe entre:

$h_2 = L_2 - l_2 \quad \rightarrow \quad h_2 = 158 - 154 = 4 \rightarrow h_2 = 4$ cm

2.6 PONTO MÉDIO DE UMA CLASSE (x_i): é o ponto que divide o intervalo de classe em duas partes iguais. É o valor que a representa.

a) Para obter o ponto médio de uma classe, basta acrescentar ao seu limite inferior a metade da amplitude do intervalo de classe. Por exemplo, o ponto médio da primeira classe da distribuição da Tabela 4.3, será determinado como segue:

Primeira classe: 150 ├ 154

Amplitude do intervalo: $h_1 = L_1 - l_1 \rightarrow h_1 = 154 - 150 = 4$

Metade da amplitude = 2

Ponto médio: $Xi = 150 + 2 = 152$

b) Quando o limite superior de uma classe for igual ao inferior da seguinte (conforme aparece na Tabela 4.3) o ponto médio poderá ser calculado através da média aritmética dos limites do intervalo. Assim sendo, o ponto médio da segunda classe será:

Segunda Classe: 154 ├ 158

Ponto médio: $X_2 = (l_2 + L_2)/2 \rightarrow X_2 = (154 + 158)/2 = 156 \rightarrow$ **X_2 = 156 cm**

2.7 AMPLITUDE AMOSTRAL (AA): é a diferença entre o valor máximo e o valor mínimo da amostra:

$$AA = X(max) - X(min)$$

Na Tabela 4.3 temos:

Primeira classe 150 ├ 154, Última Classe 170 ├ 174

Intervalos fechados à esquerda e aberto à direita → Valor máximo = 173 e valor mínimo = 150

Então: $AA = X(max) - X(min)$.

$AA = 173 - 150 = 23 \rightarrow AA = 23$ cm

- Observe que a amplitude total da distribuição jamais coincide com a amplitude amostral. (*AT* sempre será maior que *AA*).

Nota:

$$\text{Amplitude de classe } (h) = \frac{\text{maior valor observado - menor valor observado}}{\text{número de classes}}$$

É a razão entre:

$$h = \frac{AA}{K}$$

Podemos, agora, dar à distribuição de frequência das estaturas dos quarenta alunos do Colégio A, a seguinte representação tabular técnica (Tabela 4.4)

Tabela 4.4 - Estatura de 40 alunos do Colégio A.

I	Estaturas (cm)	f_i
1	150 ├154	4
2	154 ├ 158	9
3	158 ├ 162	11
4	162 ├ 166	8
5	166 ├ 170	5
6	170 ├ 174	3
		$\Sigma f_i = 40$

Fonte: CRESPO (2002).

3. TIPOS DE FREQUÊNCIA - Em estatística podemos distinguir 4 tipos de frequências: frequência simples absoluta, frequência simples relativa, frequência absoluta acumulada e frequência relativa acumulada.

3.1 FREQUÊNCIA SIMPLES OU ABSOLUTA (f_i) – são os valores que realmente representam o número de dados de cada classe.

$$\sum_{i=1}^{k} f_i = n$$

Considerem-se os exemplos dados pelas Tabelas 4.5 e 4.6.

Tabela 4.5 – EMPRESA X. Relatório de inspeção. Número de defeitos por peça

Número de Defeitos (*xi*)	Número de peças (f_i)
0	5
1	10
2	18
3	12
4	5
	$\Sigma fi = 50$

Fonte: TOLEDO; OVALLE (2008).

Na Tabela 4.5, a frequência simples absoluta do valor zero é 5, indicando que esse número aparece 5 vezes no levantamento efetuado. Há, portanto, cinco peças com nenhum defeito.

- Essa distribuição da Tabela 4.5, trata de variável discreta de variação relativamente pequena, cada valor pode ser tomado como um intervalo de classe (intervalo degenerado) e, nesse caso, a distribuição é chamada de distribuição sem intervalos de classe.

Tabela 4.6 - Estatura de 40 alunos do colégio A

Estaturas (cm)	**Frequência**
150 ├ 154	4
154 ├ 158	9
158 ├ 162	11
162 ├ 166	**8**
166 ├ 170	5
170 ├ 174	3
	$\boldsymbol{\Sigma fi = 40}$

Fonte: CRESPO (2002).

Na Tabela 4.6, a frequência simples absoluta da quarta classe é 8. Há 8 alunos cujas estaturas se situaram no intervalo compreendido pelas estaturas de 162 cm, inclusive, a 166 cm, exclusive (e assim avalia-se todas as classes).

3.2 FREQUÊNCIA SIMPLES RELATIVA (fr_i) – são os valores das razões entre as frequências simples e a frequência total. Para calcular a frequência relativa, basta dividir a frequência absoluta da classe ou do valor individual pelo número total de observações.

Simbolicamente: $fr_i = f_i/\Sigma f_i$

- A frequência relativa da terceira classe, em nosso exemplo (Tabela 4.6.), é:

  $f_3 = 11;\quad \Sigma fi = 40$
  $fr_3 = f_3/\Sigma f_i$
  $fr_3 = 11/40 = 0{,}275$
  $fr_3 = 0{,}275$

- Desejando expressar o resultado em termos percentuais, multiplica-se o quociente obtido por 100.

  $fr_i = (f_i/\Sigma f_i)\ .\ 100$

- Observando essa última expressão, vê-se claramente que a frequência relativa é o resultado de uma regra de três simples:

  $\Sigma f_i \rightarrow 100$
  $f_i \rightarrow X\%$
  $X\% = (f_i/\Sigma f_i)\ .\ 100 = fri$

- A soma das frequências simples relativas de uma tabela de frequências é sempre igual a 1,00 ou 100%.
- A frequência simples relativa do valor 11 é igual a 0,275 ou 27,5%, indicando que há vinte e sete e meio por cento de alunos, dentre os quarenta, com a estatura de 158 cm, inclusive, a 162 cm, exclusive.

3.3 FREQUÊNCIA ACUMULADA (*Fi* ou *Fac*) – ("abaixo de" uma classe ou de um valor individual) - É a soma da frequência simples absoluta dessa classe ou desse valor com as frequências simples absolutas das classes ou dos valores anteriores.

$F_k = f_1 + f_2 + ... + f_k$ ou

$F_k = \Sigma f_1$ (1= 1, 2, ..., k)

Exemplo 4.7

A frequência acumulada correspondente à **terceira classe** do exemplo apresentado na Tabela 4.6.

$F_3 = \Sigma^3{}_i = f_i = f_1 + f_2 + f_3 \rightarrow (f_1 = 4; f_2 = 9; f_3 = 11)$

Assim: $F_3 = 4 + 9 + 11 = 24 \rightarrow F_3 = 24$

Existem 24 alunos com estatura inferior a 162 cm (limite superior do intervalo da terceira classe).

3.4 FREQUÊNCIA RELATIVA ACUMULADA (*FRi* ou *FRACi*): é a frequência acumulada da classe, dividida pela frequência total da distribuição.

Simbolicamente $\rightarrow Fr_i = F_i/\Sigma f_i$

Assim, para a terceira classe da Tabela 4.6, a frequência relativa acumulada é:

$F_3 = 24$ $\Sigma f_i = 40$

$Fr_3 = F_3/\Sigma f_i$

$Fr_3 = 24/40 = 0{,}600 \rightarrow Fr_3 = 0{,}600$

- Considerando a Tabela 4.4, podemos montar a seguinte tabela com as frequências estudadas.

Tabela 4.7 – Estatura de 40 alunos do colégio A.

I	Estatura (cm)	f_i	xi	Fri	F_i	Fr_i
1	150 ├ 154	4	152	0,100	4	0,100
2	154 ├ 158	9	156	0,225	13	0,325
3	158 ├ 162	11	160	0,275	24	0,600
4	162 ├ 166	8	164	0,200	32	0,800
5	166 ├ 170	5	168	0,125	37	0,925
6	170 ├ 174	3	172	0,075	40	1,000
		$\Sigma = 40$		**$\Sigma = 1{,}000$**		

Fonte: CRESPO (2002).

Exercícios resolvidos

O conhecimento dos vários tipos de frequência ajuda-nos a responder a muitas questões com relativa facilidade, como as seguintes:

a) Quantos alunos tem estatura entre 154 cm, inclusive, e 158 cm?

Esses são os valores da variável que formam a segunda classe. Como $f_2 = 9$, a resposta é: 9 alunos.

b) Qual a percentagem de alunos cujas estaturas são inferiores a 154 cm?

Esses valores são os que formam a primeira classe. Como $fr_1 = 0{,}100$, obtemos a resposta multiplicando a frequência relativa por 100.

Assim, 0,100 x 100 = 10

Logo, a percentagem de alunos é 10%.

c) Quantos alunos tem estatura abaixo de 166 cm?

É evidente que as estaturas consideradas são aquelas que formam as classes de ordem 1, 2, 3 e 4. Assim, o número de alunos é dado por:

$$F_4 = f_1 + f_2 + f_3 + f_4 \rightarrow F_4 = 4 + 9 + 11 + 8 = 32 \rightarrow F_4 = 32$$

Portanto, 32 alunos têm estatura abaixo de 166 cm

d) Quantos alunos tem estatura não-inferior a 158 cm? O número de alunos é dado por:

$$\Sigma_i = 3^6\, fi = f_3 + f_4 + f_5 + f_6 = 11 + 8 + 5 + 3 = 27 \quad \text{ou então:}$$

$$\Sigma_i = 3^6\, fi - F_2 = n - F_2 = 40 - 13 = 27$$

Portanto, 27 alunos têm estatura não-inferior a 158 cm.

4. REPRESENTAÇÃO GRÁFICA DE UMA DISTRIBUIÇÃO DE FREQUÊNCIA - Uma distribuição de frequência pode ser representada graficamente pelo histograma, pelo polígono de frequência e pelo polígono de frequência acumulada[12]

- Construímos qualquer um dos gráficos mencionados utilizando o primeiro quadrante do sistema de eixos coordenados cartesianos ortogonais. Na linha horizontal (eixo das abscissas) colocamos os valores da variável e na linha vertical (eixo das ordenadas), as frequências.

4.1 HISTOGRAMA - O histograma é formado por um conjunto de retângulos justapostos cujas bases se localizam sobre o eixo horizontal (eixo x), de tal modo que seus pontos médios coincidam com os pontos médios dos intervalos de classe.

- As larguras dos retângulos são iguais às amplitudes dos intervalos de classe.
- As alturas dos retângulos devem ser proporcionais às frequências das classes, sendo a amplitude dos intervalos igual. Isso nos permite tomar as alturas numericamente iguais às frequências.
- A distribuição da Tabela 4.4 corresponde ao seguinte histograma.

Figura 4.1 – Histograma da estatura dos alunos da escola A.B.C. Fonte: SILVA (2010).

A diferença entre o gráfico de coluna e o histograma é o distanciamento entre as colunas. No histograma, não há separação entre os retângulos adjacentes.

[12] Alguns autores preferem designá-lo por **ogiva de Galton**.

Notas:

- O histograma goza de uma propriedade da qual faremos considerável uso: a área de um histograma é proporcional à soma das frequências.
- No caso de usarmos as frequências relativas, obtemos um gráfico de área unitária.
- Quando queremos comparar duas distribuições, o ideal é fazê-lo pelo histograma de frequências relativas.

Examinando o histograma acima, responda:

A) Qual o intervalo de classe que tem maior frequência?

B) Qual a amplitude total da distribuição?

C) Qual o número total de alunos?

D) Qual a frequência do intervalo de classe 170 ├ 174.

4.2 POLÍGONO DE FREQUÊNCIA - é um gráfico de linha, sendo que as frequências marcadas sobre perpendiculares ao eixo horizontal, levantadas pelos pontos médios dos intervalos de classe.

- Para obtermos um polígono (linha fechada), devemos completar a figura, ligando os extremos da linha obtida aos pontos médios da classe anterior à primeira e da posterior à última, da distribuição.

À distribuição da Tabela 4.4 corresponde o seguinte polígono de frequência.

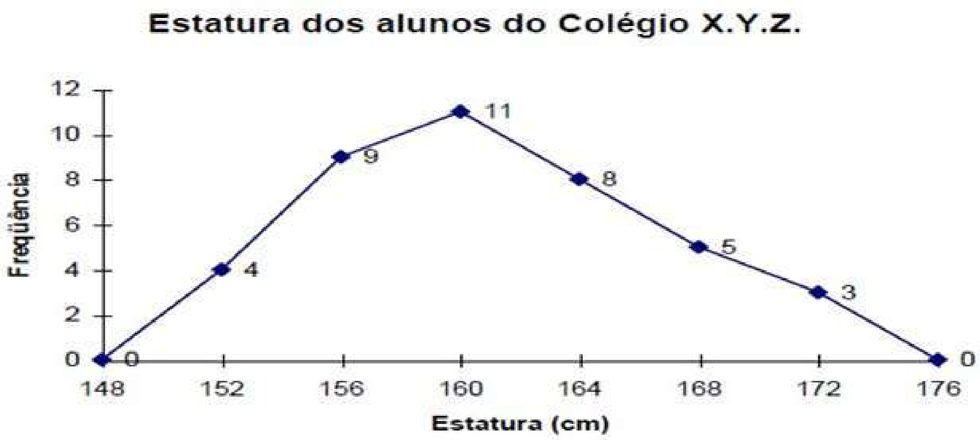


Figura 4.2 - Polígono de frequência da estatura dos alunos do colégio X.Y.Z. Fonte: SILVA (2010).

Procedimentos:

a) Pontos médios dos intervalos de classe: $x_I = (l_I + L_i)/2$

Primeira Classe: 150 ├ 154 → anterior à primeira classe: 146 ├ 150

Então: x_i = (146 + 150)/2= 148 (liga ao extremo);

Posterior à primeira classe: xi = (150 + 154)/2 = 152 (liga ao extremo); e calcula-se todas as classes.

Ultima classe: 170 ├ 174 → posterior à última classe: 174 ├ 178

Então: x_i = (174 + 178)/2 = 176 (liga ao extremo);

Nota:

- No caso de termos uma variável essencialmente positiva, cuja distribuição se inicie no valor zero, devemos considerar um intervalo anterior localizado no semi-eixo negativo. Porém, consideraremos apenas a parte positiva do segmento que liga o ponto médio desse intervalo com a frequência do intervalo 0 ├ .

4.3 POLÍGONO DE FREQUÊNCIA ACUMULADA - É traçado marcando-se as frequências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe.

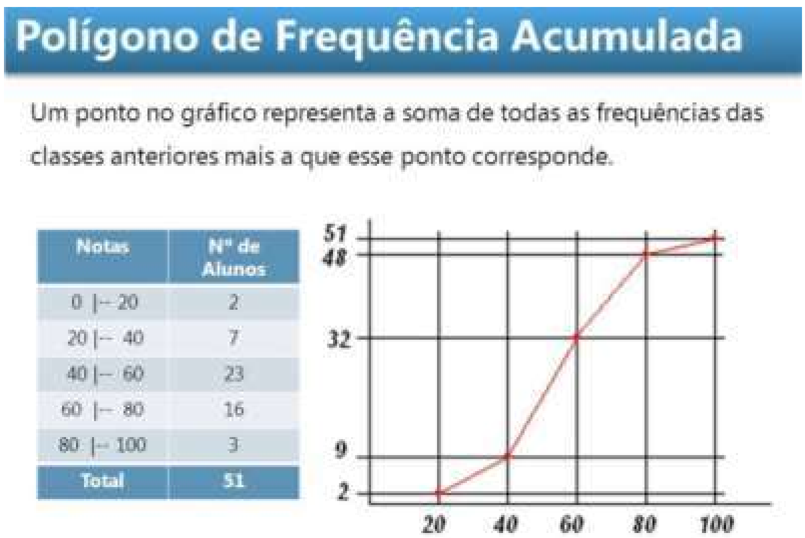


Figura 4.3 – Polígono de frequência acumulada. Fonte: ALESSANDRO (2012).

Uma distribuição de frequência sem intervalos de classe é representada graficamente por um diagrama onde cada valor da variável é representado por um segmento de reta vertical e de comprimento proporcional à respectiva frequência.

CAPÍTULO V - GRÁFICOS ESTATÍTICOS

1. GRÁFICOS ESTATÍSTICOS - São representações visuais dos dados estatísticos, cujo objetivo é o de produzir, no investigador ou no público, uma impressão mais rápida e viva do fenômeno em estudo.

A representação gráfica de um fenômeno deve obedecer a certos requisitos fundamentais para ser realmente útil:

a) **Simplicidade** – o gráfico deve ser destituído de detalhes de importância secundária, assim como de traços desnecessários que possam levar o observador a uma análise morosa ou com erros.
b) **Clareza** – o gráfico deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.
c) **Veracidade** – o gráfico deve expressar a verdade sobre o fenômeno em estudo.

1.1 CLASSIFICAÇÃO DOS GRÁFICOS SEGUNDO O OBJETIVO (USO)

- É possível distinguir, dois objetivos que justificariam o emprego de gráficos.
- Para apresentar visualmente dados numéricos, proporcionando maior facilidade e rapidez de compreensão dos mesmos, ou então;
- Para apresentar conclusões ou resultados de uma análise.

Há, portanto, dois tipos de gráficos conforme o objetivo: gráficos de informação e gráficos de análise.

1.1.1 Gráficos de informação - São gráficos destinados principalmente ao público em geral, objetivando proporcionar uma visualização rápida e clara da intensidade e dos valores relativos ao fenômeno observado.

- São gráficos tipicamente expositivos, dispensando comentários explicativos adicionais.
- Não se deve prescindir dos títulos, escritos em letra de forma;
- As legendas podem ser omitidas, desde que as informações desejadas estejam presentes, possibilitando a interpretação do gráfico.

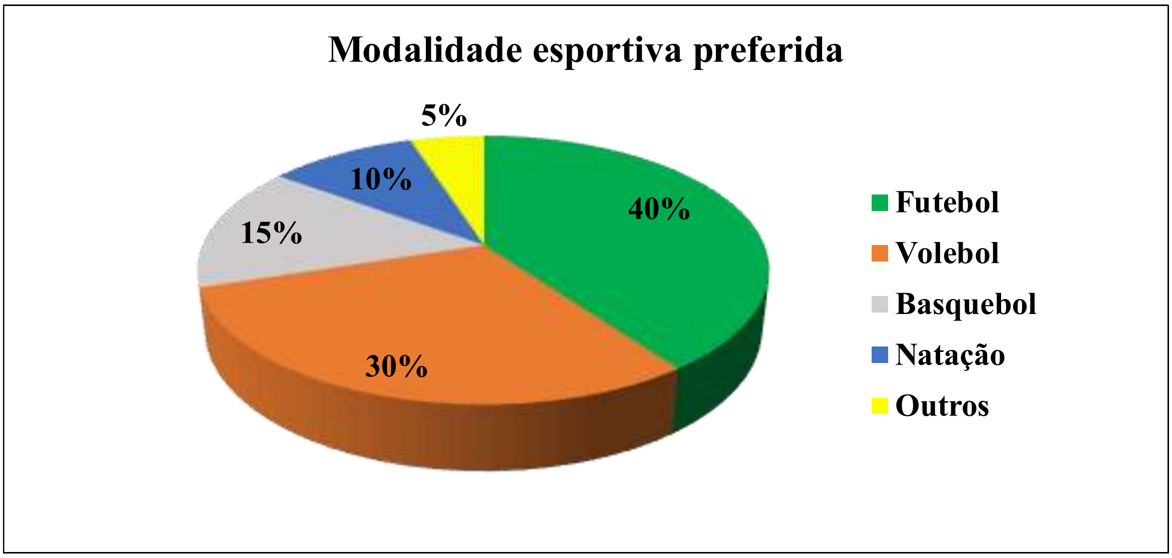


Figura 5.1 – Gráfico de informação. Fonte: Clubes de Matemática – Obmep (2019).

1.1.2 Gráficos de análise - Os gráficos de análise prestam-se melhor ao trabalho estatístico, fornecendo elementos úteis à fase de análise dos dados, sem deixar de ser, também, informativos.

- Os gráficos de análise frequentemente vêm acompanhados de uma tabela. Inclui-se, muitas vezes, um texto dissertativo, chamando a atenção para os pontos principais revelados pelo gráfico ou pela tabela.

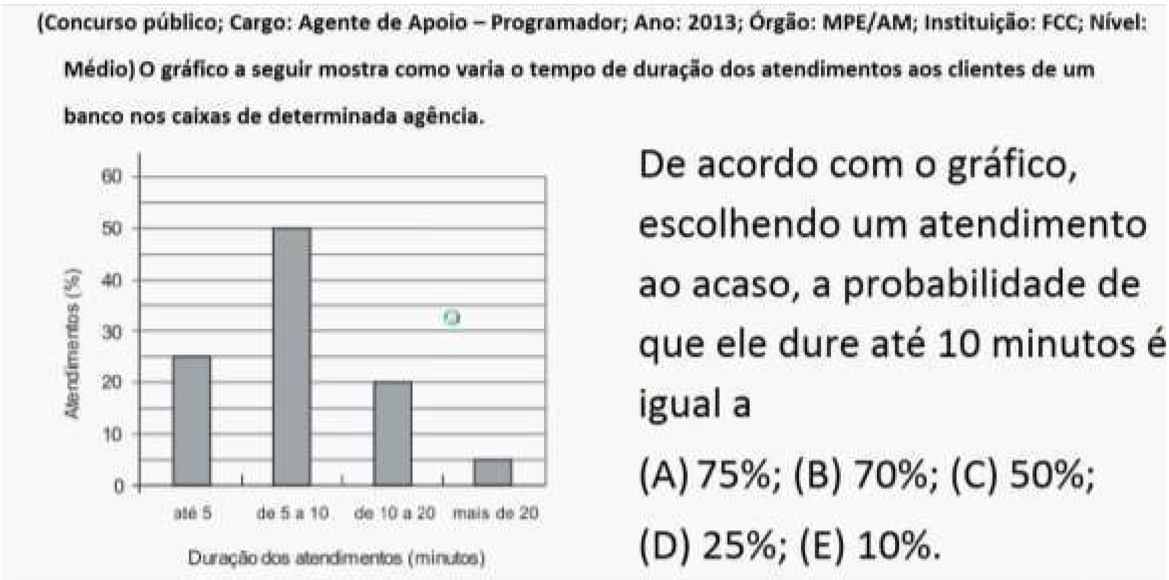


Figura 5.2 - Gráfico de análise. Fonte: CORREIA, 2016.

Nota:
Uso indevido de Gráficos podem trazer uma ideia falsa dos dados que estão sendo analisados, chegando mesmo a confundir o leitor. Trata-se, na realidade, de um problema de construção de escalas.

1.2 CLASSIFICAÇÃO DOS GRÁFICOS SEGUNDO A FORMA - Há quatro tipos de gráficos classificados quanto ao critério da forma: Diagramas, Cartogramas, Estereogramas e Pictogramas.

1.2.1 Diagramas - São gráficos geométricos dispostos em duas dimensões. São os mais usados na representação de séries estatísticas e se apresentam através de uma grande variedade de tipos.

- Para sua construção faz-se uso do sistema cartesiano.
- Eles podem ser: gráfico em linha ou em curva, gráfico em colunas ou em barras, gráfico em colunas ou em barras múltiplas, gráfico em setores.

A) Gráfico em linha ou em curva – Este tipo de gráfico se utiliza da linha poligonal para representar a série estatística temporal. Constitui uma aplicação do processo de representação das funções num sistema de coordenadas cartesianas.

Exemplo 5.1

Quantidade de acidentes de trânsito no Brasil na primeira semana de janeiro de 2014.

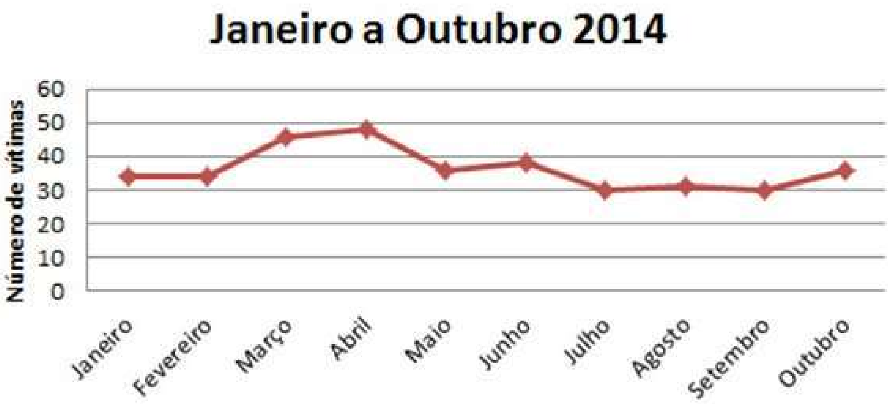


Figura 5.3 – Variação do número de acidentes de trânsito em Itaituba ao logo de dez meses. Fonte: GBM (2014).

Nota:

1. Podemos sombrear ou hachurar o gráfico.
2. Quando representamos, em um mesmo sistema de coordenadas, a variação de dois ou mais fenômenos (Figura 5.4), a parte interna da figura formada pelos gráficos desses fenômenos é denominada área de excesso.

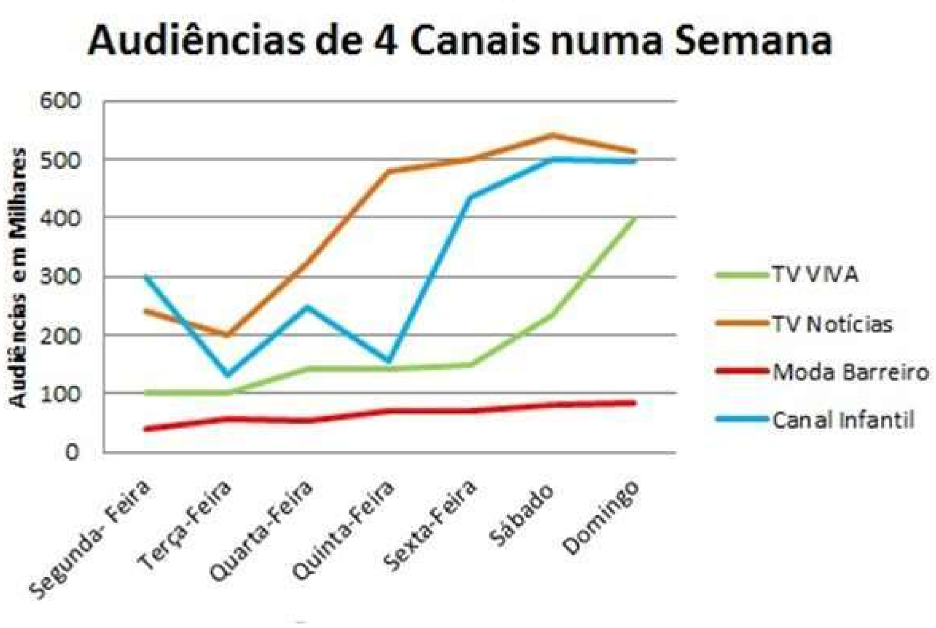


Figura 5.4 – Audiência de 4 canais de televisão numa semana. Fonte: MestAtisCticSa (2019).

B) Gráfico em colunas ou em barras – É a representação de uma série estatística através de retângulos dispostos na vertical (**colunas**) ou em retângulos dispostos na horizontal (**barras**).

I. Gráfico em colunas

- Os retângulos têm a mesma base e as alturas são proporcionais aos respectivos dados.
- Prestam-se, em especial, à representação, análise e interpretação de dados relacionados com séries de tempo, as colunas deverão estar dispostas em ordem cronológica
- Às vezes, coloca-se no topo ou no interior de cada coluna a valor correspondente à sua altura.
- Este tipo de gráfico pode aparecer de diversas maneiras, dentre elas: gráficos em colunas superpostas, gráficos de porcentagens etc.

Figura 5.5 - Número de casos de dengue no período de 2010 a 2015 em Iracemápolis. Fonte: Gazeta de Iracemápolis.

i) Gráfico em colunas superpostas - Corresponde ao gráfico de colunas compostas. Tanto um quanto o outro servem para representar comparativamente dois ou mais atributos.

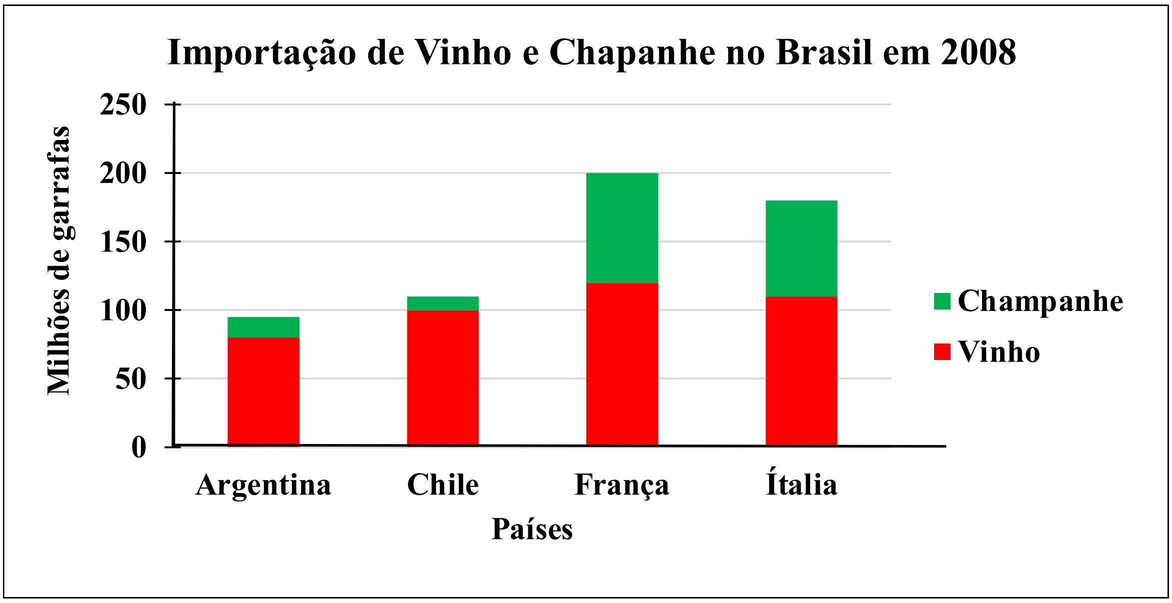


Figura 5.6 - Gráfica em colunas superpostas. Fonte: PINTO (2015).

ii) Gráficos de porcentagens - O mesmo fato representado pelo gráfico de colunas superpostas pode ser visto também em termos de participação percentual da importação no Brasil de vinho e champanhe de cada país.

- Cada coluna do gráfico corresponde a 100%, e suas componentes são obtidas por meio de uma regra de três simples. Para exemplificar: a importação do Chile, cujo valor é igual à altura da primeira coluna do gráfico (Figura 5.6), corresponde a 100%: então, a importação de vinho e champanhe do Chile, corresponde a **X** por cento. Esquematicamente.

Primeiro Passo: Calcular os 100%

110 → 100% (110 é o valor máximo aproximado)

100 → X

Então: X = (100 x 100)/110 = 91%

Segundo Passo: Obter a porcentagem do complemento por diferença:

(Observe as legendas)

Champanhe → 10 assim: 100 – 91 = **9%**

Vinho → 100 assim: 91 – 9 = **82%**

E, assim, se procede com os outros países até completar o gráfico.

iii) Gráficos em colunas remontadas - Permite que as comparações entre as grandezas dos atributos sejam feitas mais facilmente. Presta-se aos mesmos propósitos que o gráfico em colunas agrupadas.

Entrada de migrantes em três Estados do Brasil 1992 – 1994

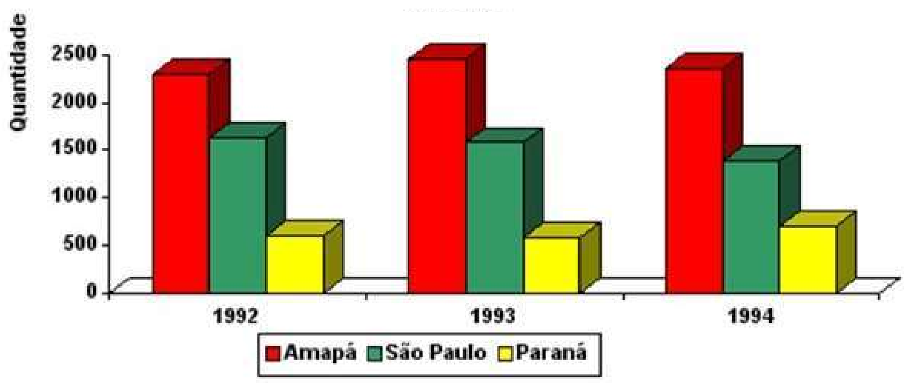


Figura 5.7 – Gráficos em colunas remontadas. Fonte: CV Magazine (2019).

II. Gráfico em barras:

- Os retângulos tem a mesma altura e os comprimentos são proporcionais aos respectivos dados: assegurando a proporcionalidade entre as áreas dos retângulos e os dados estatísticos.
- Quando **as legendas são extensas**, deve-se dar preferência ao **gráfico em barras horizontais** (séries geográficas e específicas). Porém, se preferirmos o **gráfico em colunas**, os dizeres (legendas) deverão ser dispostos de **baixo para cima**.
- **A ordem** a ser observada é a **cronológica**, se a **série for histórica**, e a **decrescente**, se for **geográfica ou categórica**.
- A distância entre as barras, por questões estéticas, não deverá ser menor que a metade nem maior que os dois terços da largura (ou da altura) dos retângulos.
- As barras só diferem em comprimento, e não em largura, a qual é arbitrária
- As barras devem ser desenhadas observando sua ordem de grandeza, **normalmente decrescente**.
- Um gráfico, construído para mostrar **grandezas absolutas**, deverá ter uma linha zero claramente definida e uma escala de quantidades ininterrupta, caso contrário a leitura e a interpretação do gráfico poderão ficar distorcidas.
- Existem outros tipos mais elaborados de gráficos em barras. Entre eles destacam-se três: **o gráfico de barras compostas, o gráfico de barras agrupadas e o gráfico de barras bidirecionais**

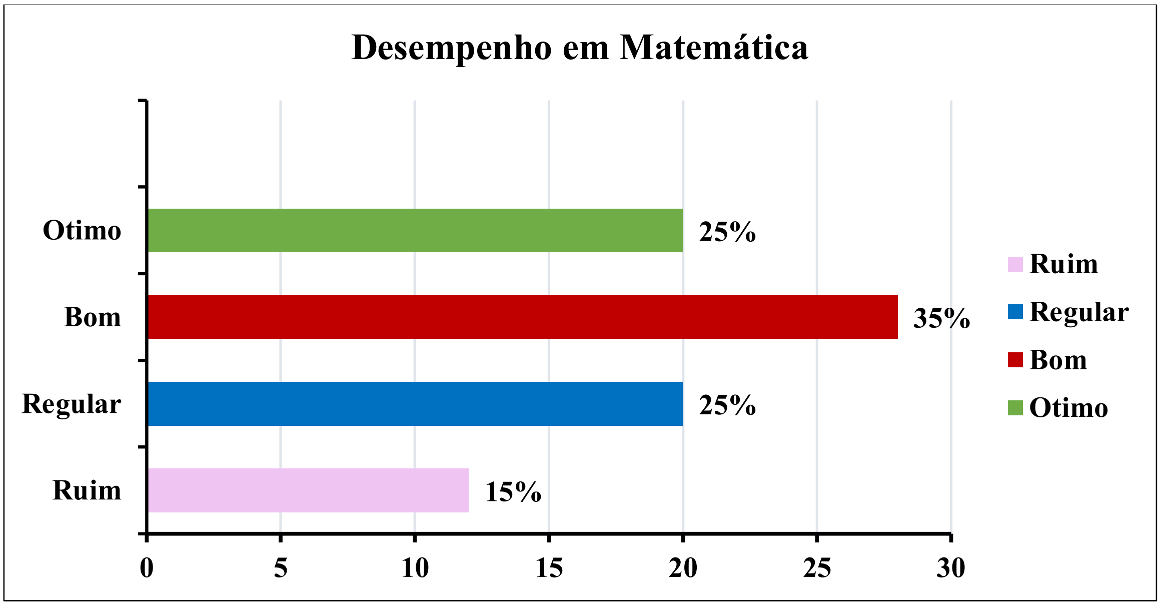


Figura 5.8 – Gráfico em barras horizontais. Fonte: SILVA (2019).

j) **Gráfico de barras compostas -** Diferente do gráfico de barras convencional apenas pelo fato de apresentar cada barra segmentada em partes componentes.

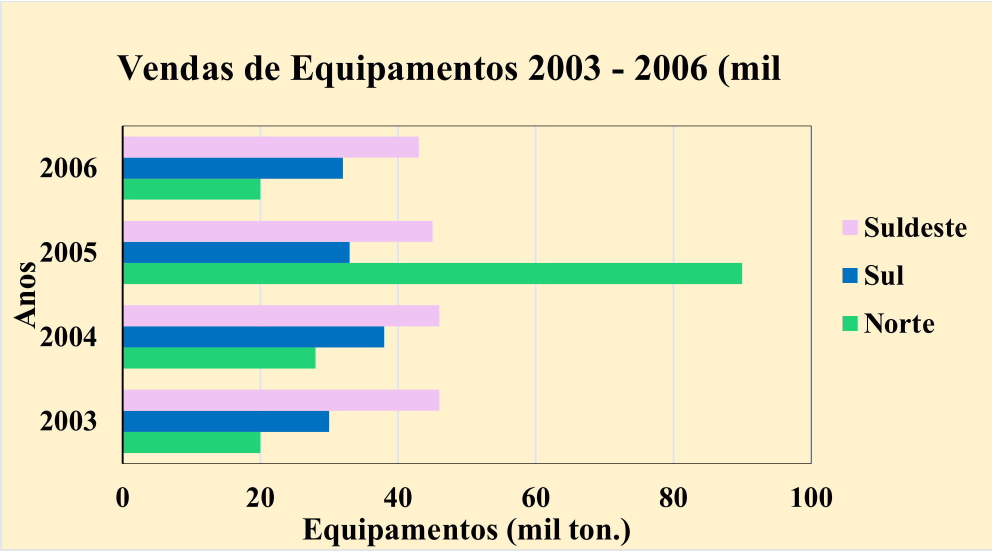


Figura 5.9 - Gráfico em barras compostas: Vendas de equipamentos.
Fonte: TROIA (2019).

- Através desse gráfico é possível comparar visualmente as vendas de equipamentos (de uma empresa) durante os anos de 2003 a 2006, por regiões (descritas na legenda).

ii) Gráfico de barras agrupadas - Presta-se aos mesmos propósitos que o gráfico de colunas remontadas.

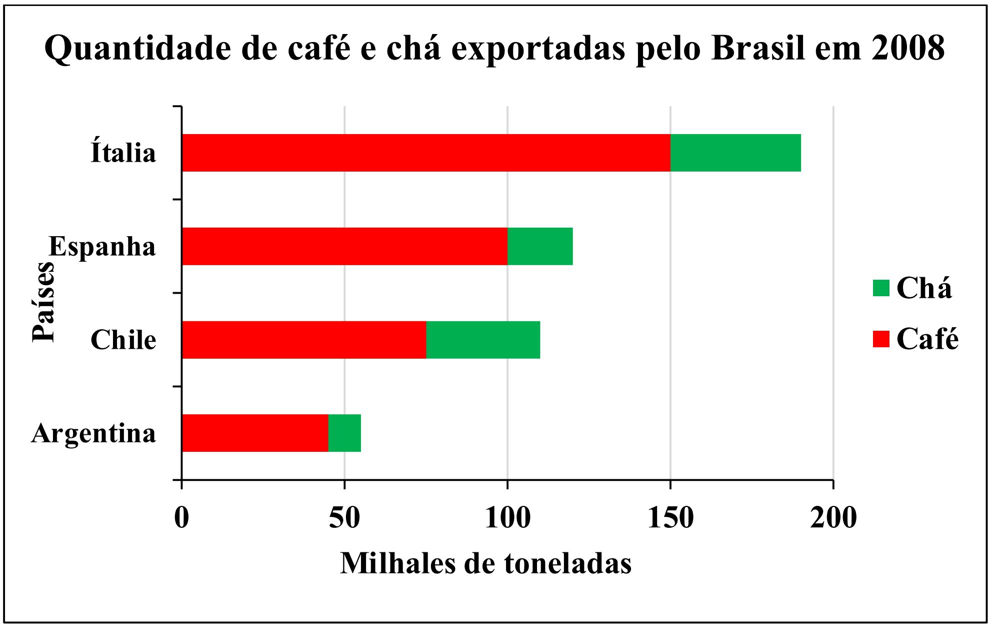


Figura 5.10 – Gráficos em barras agrupadas. Fonte: PINTO (2015).

iii) Gráfico de barras bidirecionais - Quando se deseja representar, graficamente, quantidades positivas e negativas, como, por exemplo, perdas e ganhos, mudanças de percentagens em períodos de tempo sucessivos e desvios positivos ou negativos em torno de um valor padrão ou coisas opostas, por exemplo, masculino e feminino.

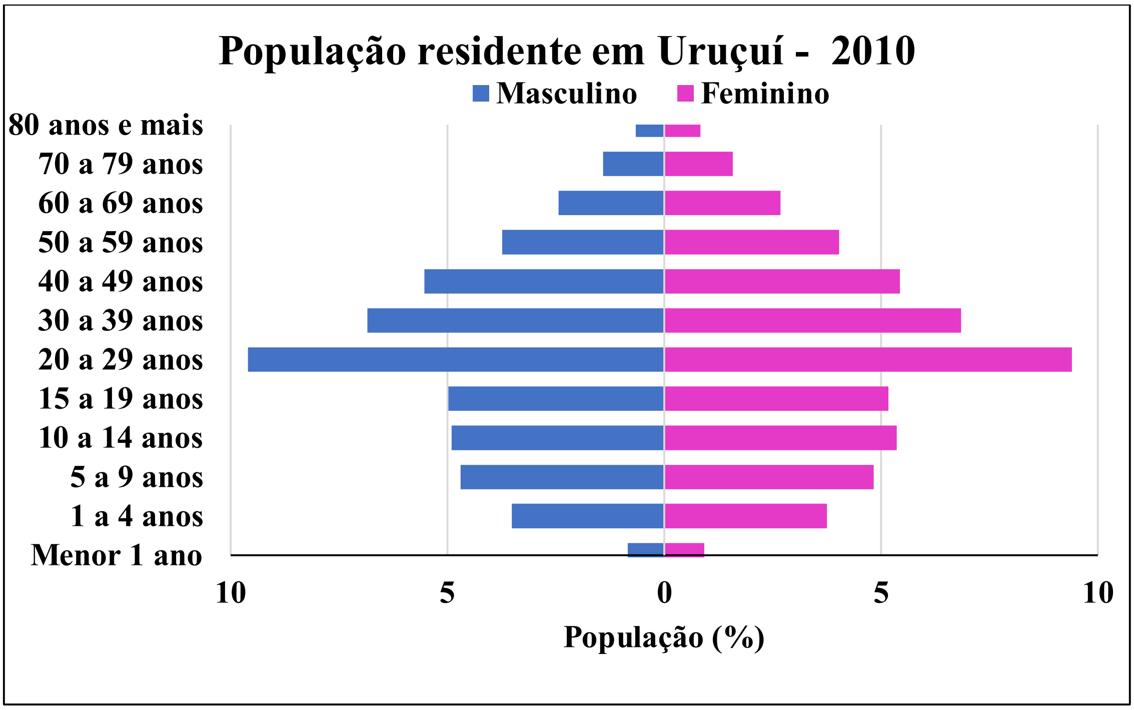


Figura 5.11 – Gráfico de barras bidirecionais. Fonte: IBGE Censo demografico/MS/SGEP/Datasus (2021).

C) Gráfico em colunas ou em barras múltiplas - Esse tipo é geralmente empregado quando queremos representar, simultaneamente, dois ou mais fenômenos estudados com o propósito de comparação.

Exemplos 5.2

I. Gráfico em colunas múltiplas

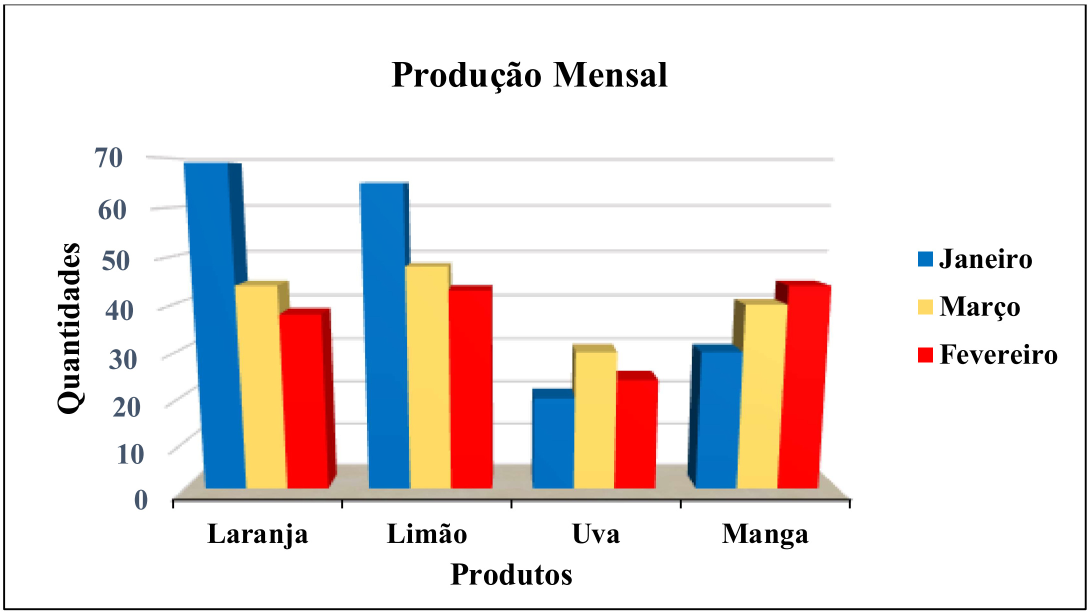


Figura 5.12 – Gráfico em colunas múltiplas. Fonte: FERREIRA (2013).

II Gráfico em barras múltiplas

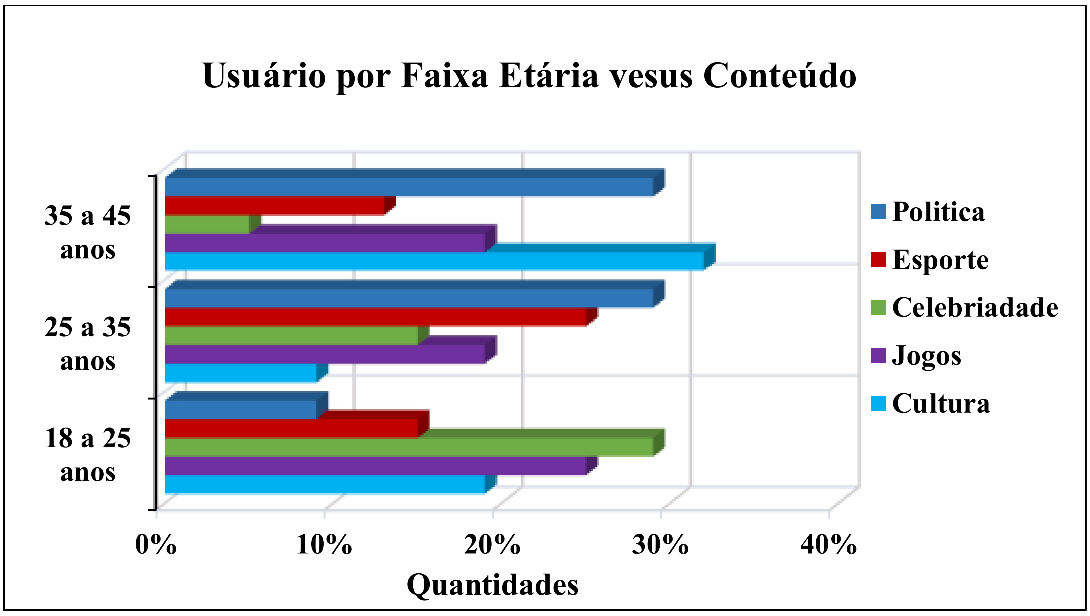


Figura 5.13 – Gráfico em barras múltiplas. Fonte: TECMUNDO (2009).

D) Gráfico em setores - Este gráfico é construído com base em um círculo, e é empregado sempre que desejamos ressaltar a participação do dado no total. Também é conhecido como gráfico circular ou cartograma em setores.

- São usados para representar valores absolutos ou porcentagens complementares.

- O total é representado pelo círculo que fica dividido em tantos setores quantas são as partes.
- Os setores são tais que suas áreas são respectivamente proporcionais aos dados da série.
- O gráfico em setores só deve ser empregado quando há, no máximo, sete dados.
- Obter cada setor por meio de uma regra de três simples e direta, lembrando que o total da série corresponde a 360º.
- Se a série já apresenta os dados percentuais, obter os respectivos valores em graus multiplicando o valor percentual por 3,6.
- São representadas por este tipo de gráfico, as séries geográficas, específicas e as categorias em nível nominal;
- As séries temporais geralmente não são representadas por este tipo de gráfico.

Nota:
A representação de porcentagens complementares pode ser feita através de diferentes tipos de gráficos:

a) Gráficos em setores.
b) Gráficos em barras compostas.
c) Gráficos em colunas (porcentagens complementares).
d) Gráficos em faixas complementares.

Exemplo 5.3: Gráfico em setores

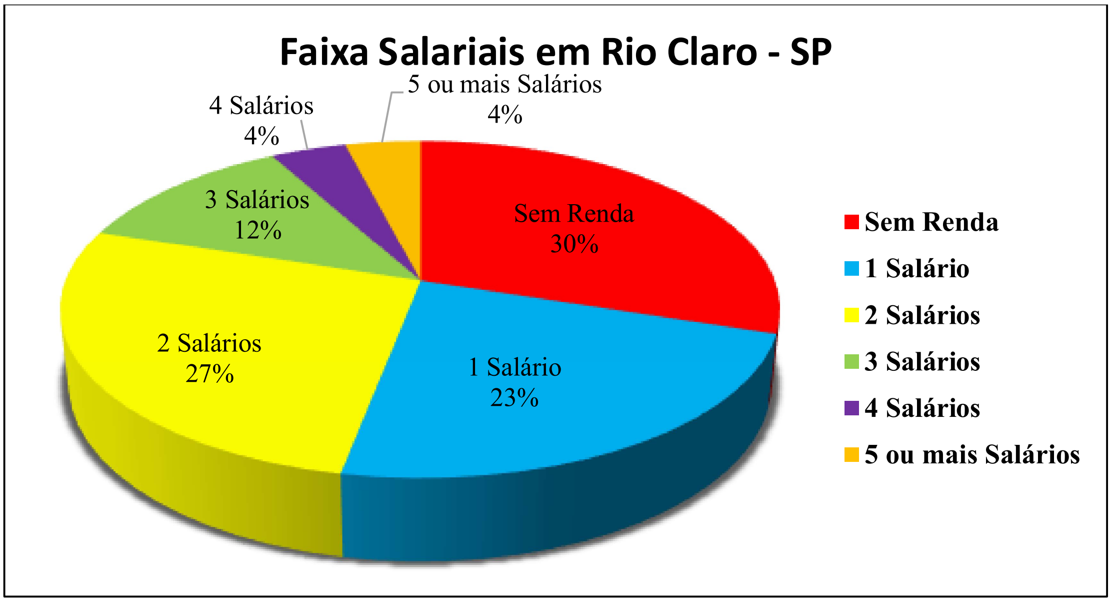


Figura 5.14 – Gráfico em setores. Fonte: PIZZOTTI (2011).

Represente a Tabela por meio de gráfico de setores:

Dada a série:

Tabela 5.1 - Rebanho de suíno do sudeste do Brasil em 1992

Estados	**Quantidade (mil cabeças)**
Minas Gerais	3.363,7
Espírito Santo	430,4
Rio de Janeiro	308,5
São Paulo	2.035,9
Total	6.138,5

Fonte: CRESPO (2002).

Procedimentos:

a) Para os cálculos aplica-se regra de três simples:

Minas Gerais → X_1 Espírito Santo →X_2 Rio de Janeiro → X_3 São Paulo→ X_4

6.138,5 → 360º Então: Xx_1 = (3.363,7 x 360) /6.138,5 → X_1 **= 197º**

3.363,7 → X_1

b) Use o mesmo procedimento para calcular os demais Estados. (Faça os cálculos)

Espírito santo →X_2 = 25,2 → X_2 **= 25º**

Rio de Janeiro → X_3 = 18,0 → X_3 **= 18º**

São Paulo →X_4 = 119,3 →X_4 **= 120°**

- Com estes dados (valores em graus) marcamos num círculo de raio arbitrário, com um transferidor, os arcos correspondentes, obtendo o gráfico.
- Coloque todos os dados obrigatórios: título, fonte, legenda.

Nota:
Gráfico polar – Faz uso do sistema de coordenadas polares. É o tipo ideal para representar séries temporais cíclicas - séries que apresentam em seu desenvolvimento determinada periodicidade, por exemplo, a variação da precipitação pluviométrica ao longo do ano ou a temperatura ao longo do dia, o consumo de energia durante o mês ou durante o ano, etc.

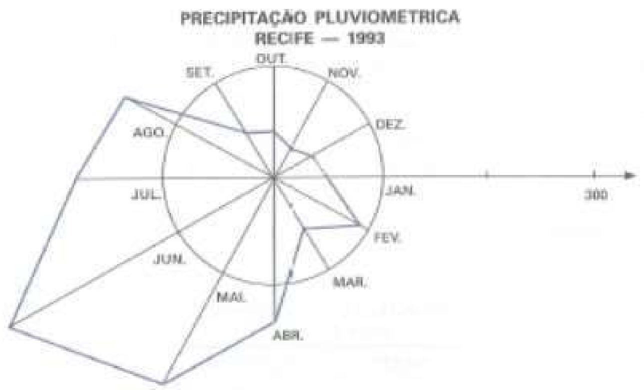


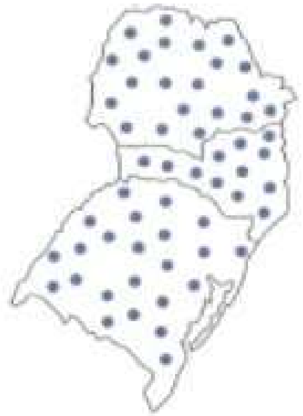
Figura 5.15 – Gráfico polar. Fonte: CRESPO (2002).

1.2.2 Cartogramas - É a representação sobre uma carta geográfica (mapas). O objetivo desse gráfico é o de figurar os dados estatísticos diretamente relacionados com áreas geográficas ou políticas.

Distinguimos duas aplicações:

a) Representar dados absolutos (população) – neste caso, lançamos mão, em geral, dos pontos, em número proporcional aos dados.

População projetada da região sul Brasil – 1994



400.000 habitantes

Figura 5.16 – Cartogramas. Fonte: CRESPO (2002).

> **Nota:**
> Quando os números absolutos a serem apresentados forem muito grandes, no lugar de pontos podemos empregar hachuras.

b) Representar dados relativos (densidade) - neste caso, lançamos mão, em geral, de hachuras ou cores.

Tipos de clima do Brasil

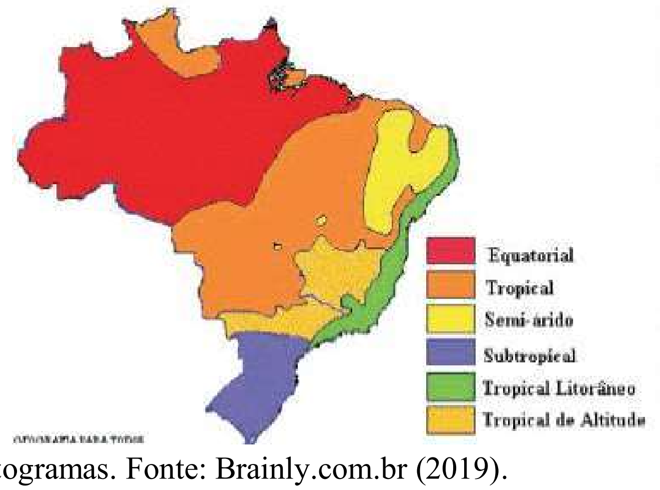


Figura 5.17 – Cartogramas. Fonte: Brainly.com.br (2019).

1.2.3 Estereogramas - São gráficos geométricos dispostos em **t**rês dimensões, pois representam volume. São usados nas representações gráficas das Tabelas de dupla entrada. Em alguns casos, esse tipo de gráfico fica difícil de ser interpretado dada a pequena precisão que oferecem.

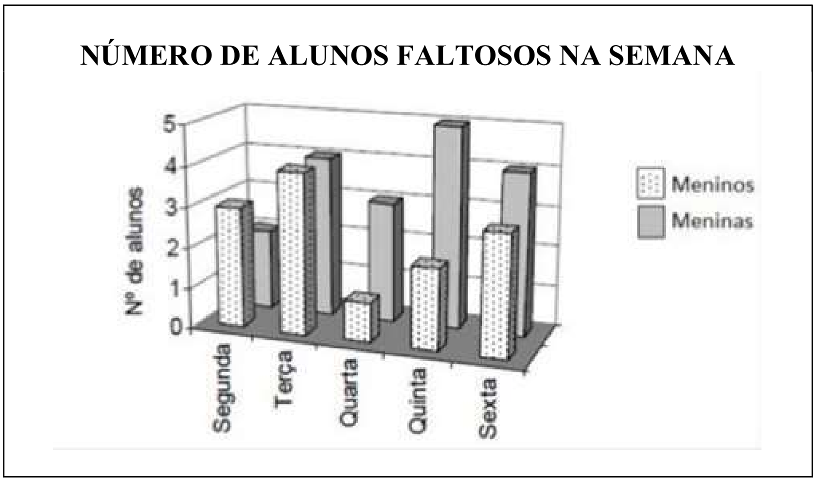


Figura 5.18 - Gráfico Estereogramas. Fonte: PINTO (2015).

1.2.4 Pictograma - Constitui um dos processos gráficos que melhor fala ao público pela sua forma, ao mesmo tempo, atraente e sugestiva. A representação gráfica consta de figuras.

- Os símbolos devem ser autoexplicativos.
- A desvantagem dos pictogramas é que mostram uma visão geral do fenômeno e não detalhes minuciosos.

Figura 5.19 – Pictograma. Fonte: matemática.pt. (2019).

CAPÍTULO VI - MEDIDAS DE POSIÇÃO

1. **ELEMENTOS TÍPICOS DA DISTRIBUIÇÃO** – Para ressaltar as tendências características de cada distribuição, isoladamente ou em confronto com outras, faz-se necessário introduzir conceitos que se expressem através de números para traduzir essas tendências. Esses conceitos são as:

- Medidas de Posição
- Medidas de Variabilidade ou Dispersão
- Medidas de Assimetria
- Medidas de Curtose

1.1 MEDIDAS DE POSIÇÃO – Estatísticas que representam uma série de dados que orientam quanto à posição da distribuição em relação ao eixo horizontal (eixo das abscissas). As medidas de posição mais importantes são as medidas de tendência central[13]. Destacamos:

- A média aritmética
- A mediana
- A moda

As outras medidas de posição são as **Separatrizes**:

- A própria mediana
- Os quartis
- Os decis
- Os percentis

1.1.1 Média aritmética ou média amostral ($\bar{x}$) - É o quociente da divisão da soma dos valores da variável pelo número deles:

$$\bar{x} = \frac{\sum x_i}{n}$$

[13] Medidas de tendência central: devido aos dados observados tenderem, em geral, a se agrupar em torno dos valores centrais (CRESPO,2002 p.79)

Onde:

$\bar{x}$ **:** símbolo da média aritmética (lê-se **x** barra**)**

Σx_i: soma dos valores de x (variável);

n: número de observações.

- **Cálculo da média**

Exemplo 6.1

Encontrar a média aritmética para o conjunto de observações: 5, 1, 6, 2,4.

Solução:

Temos cinco observações: n = 5, então:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5+1+6+2+4}{5} = \frac{18}{5} = 3{,}6$$

Exemplo 6.2

Sabendo-se que a venda diária de arroz tipo A, durante uma semana, foi de 10, 14, 13, 15, 16, 18 e 12 kilos, temos, para venda média diária na semana de:

$$\bar{x} = (10+14+13+15+16+18+12)/7 = 14 \text{ Kg}$$

Logo: $\bar{x} = 14$ Kg

i) Valores agrupados - Quando os valores $\mathbf{x_i}$ estão agrupados com suas respectivas frequências absolutas $\mathbf{f_i}$, a média aritmética ou média amostral é expressa por:

$$\bar{x} = (\Sigma\, x_1 f_i)/n \qquad (n = \Sigma f_i)$$

Exemplo 6.3

Como exemplo, vamos calcular a média da distribuição em classes de frequências dada na Tabela 6.1.

Tabela 6.1 – Estatura de 40 alunos do colégio A.

I	**Estaturas (cm)**	f_i	x_i	$x_i f_i$
1	150 ├ 154		152	608
2	154 ├ 158	9	156	1.404
3	158 ├ 162	11	160	1.760
4	162 ├ 166	8	164	1.312
5	166 ├ 170	5	168	840
6	170 ├ 174	3	172	516
		Σf_i **= 40**		$\Sigma x_i f_i$ **= 6.440**

Fonte: CRESPO (2002).

Solução:

Da tabela de distribuição de frequências, temos:

a) Ponto médio de uma classe (x_i): o ponto médio poderá ser calculado através da média aritmética dos limites do intervalo da classe: $_i = (l_i + L_i)/2$.

b) O ponto médio da primeira classe, 150 ├154:

$x_1 = (150 + 154)/2 = 152 \rightarrow x_1 = 152$ .... e assim calcula-se as demais classes.

$x_i f_i$, na primeira classe é: 152 * 4 = 608. E assim calcula-se as demais classes.

Então:

$\Sigma x_i f_i = 6.440$; $\Sigma f_i = 40$ $n = \Sigma f_i$

Logo: $\bar{x} = (\Sigma x_i f_i)/n \rightarrow \bar{x} = 6.440/40 \rightarrow \bar{x} = 161$ cm

Nota:

Emprego da média - a média é utilizada quando:

- Desejamos obter a medida de posição que possui a maior estabilidade:
- Houver necessidade de um tratamento algébrico anterior.

Questões propostas:

Questão 1

A média mínima para aprovação em determinada disciplina é 5,0. Se um estudante obtém as notas 7,5; 8,0; 6,0; 2,5; 2,0; 5,5; 4,0 nos trabalhos mensais da disciplina em questão, pergunta-se: ele foi ou não aprovado?

Questão 2

Calcular a média para cada uma das distribuições:

a)

x_i	3	4	7	8	12
f_i	2	5	8	4	3

b)

Aluguel em R$ 100	4 ├ 6	6 ├ 8	8 ├ 10	10 ├ 12	12 ├ 14	14 ├ 16	16 ├ 18
Quantidade de casas	18	25	32	40	30	18	12

1.1.2 A mediana (Md) - A mediana de um conjunto de valores, dispostos segundo uma ordem (crescente ou decrescente), é o valor que divide a amostra ou população, em duas partes iguais.

1.1.2.1 Mediana em dados não-agrupados - Dada uma série de valores como, por exemplo: **5, 2, 6, 13, 9, 15, 10**

De acordo com a definição de mediana, o primeiro passo a ser dado é o da ordenação (crescente ou decrescente) dos valores: 2, 5, 6, **9**, 10, 13, 15

O valor que divide a série acima em duas partes iguais é **9**;

Logo a **Md = 9**.

1.1.1.2 Cálculo da mediana - Variável discreta

Se a série dada tiver:

a) Se n ímpar (número ímpar de termos), a mediana será o elemento central de ordem:

[(n + 1)/2]

b) Se n for par, a mediana será a média entre os elementos centrais de ordem:

[n/2 e (n/2) +1].

Exemplo 6.4

Calcule a mediana da série **1, 3, 0, 0, 2, 4, 1, 2, 5**

Solução:

Primeiro passo: ordenar a série **0, 0, 1, 1, 2, 2, 3, 4, 5** **N = 9**

Segundo Passo: calcular a mediana: **Md = (n + 1)/2**

Md = (9 + 1)/2→ Md = 5

Logo, a mediana será o quinto elemento da série = 2

Exemplo 6.5

Calcule a mediana da série **1, 3, 0, 0, 2, 4, 1, 3, 5, 6**

Solução:

Primeiro passo: ordenar a série: 0, 0, 1, 1, 2, 3, 3, 4, 5, 6 n = 10

Segundo Passo: calcular a mediana **Md =** [(n/2) +(n/2)+ 1]

Md= [10/2 + (10/2)+1]

Terceiro Passo: identificar os termos; Md= (5 + 6) → **Md = (5º termo e 6º termo**)

Assim: **5º termo** corresponde a 2; e o **6º termo** corresponde a 3.

Quarto Passo: Encontrar a mediana através da média aritmética dos termos:

Md = (2 + 3)/2 = 2,5 → Md = 2,5

Notas:

- Quando o número de elementos da série estatística for ímpar haverá coincidência da mediana com um dos elementos da série;
- Quando o número de elementos da série estatística for par, nunca haverá coincidência da mediana com um dos elementos da série. A mediana será sempre a média aritmética dos 2 elementos centrais da série.
- Em uma série a mediana, a média e a moda não tem, necessariamente, o mesmo valor.
- A mediana depende da posição e não dos valores dos elementos na série ordenada. Essa é uma das diferenças marcantes entre mediana e média (que se deixa influenciar, e muito, pelos valores extremos).

Vejamos:

Em **5, 7, 10, 13, 15** a **média ($\bar{x}$) = 10** e **a mediana (Md) = 10**.

Exemplo 6.6

Consideremos a distribuição relativa a 34 famílias de quatro filhos, tomando para variável o número de filhos do sexo masculino. Calcular a mediana (Tabela 6.2).

Solução

Tabela 6.2 – Apresenta os dados do problema.

X_i	f_i	F_i	
0	2	**2**	
1	6	**8**	
2	10	**18**	**← classe Md**
3	12	**30**	
4	4	**34**	
	Σ = 34		

Procedimentos:

Tomemos a distribuição relativa à Tabela 6.2.

Xi = Nº de meninos (variável)

f_i = frequência; F_i = frequência acumulada

Primeiro passo: completar a tabela com a coluna correspondente à frequência acumulada (F_i)

Segundo passo: Encontrar a mediana (observando Σf_i)

$\Sigma f_i = 34 \rightarrow$ **n é par**.

- Mediana será a média entre os elementos da ordem [n/2 e (n/2) + 1]

Assim, 34/2 = 17º e 34/2 + 1 = 18º

Terceiro passo: identifica-se a classe mediana através da frequência acumulada.

$F_i = 18 \leftarrow$ classe mediana (contém os elementos de ordem 17º e 18º) que corresponde ao valor 2 da variável (x_i), sendo este o valor mediano.

Assim: 17º corresponde a 2

18º corresponde a 2, logo

Md = (2 + 2)/2 = 2 meninos

Exemplo 6.7

Calcular a mediana para as distribuições:

Tabela 6.3.A -

X_i	f_i	F_i	
1	1	1	
2	3	4	
3	5	9	←classe Md
4	2	11	
	$\Sigma = 11$		

Fonte: MARTINS (2010).

Procedimentos:

Tomemos a distribuição relativa à Tabela 6.3.A.

x_i = **ordem** (variável)

f_i = **frequência**

F_i = **frequência acumulada**

Primeiro passo: completar a tabela com a coluna correspondente à frequência acumulada (F_i)

Segundo passo: encontrar a mediana (observando Σf_i) → $\Sigma f_i = 11$ → n é ímpar

Mediana será o elemento de ordem (n + 1)/2,

Assim: (11+1)/2 = 6. → 6º (sexto elemento).

Terceiro passo: identifica-se a classe mediana através da frequência acumulada

$F_i = 9$ (contém o 6º elemento) ← **classe mediana** e $x_i = 3$, correspondente ao valor mediano. Assim: 6º corresponde a 3

Logo, **Md = 3**

1.1.1.3 Cálculo da mediana – Variável contínua

Primeiro Passo: calcula-se a ordem **n/2** → ($\Sigma f_i/2$).

A variável é contínua, independentemente se **n** é par ou ímpar.

Segundo Passo: pela F_i (frequência acumulada), identifica-se a classe que contém a mediana (classe Md)

Terceiro Passo: utiliza-se a fórmula:

$$\mathbf{Md} = l_{Md} + \frac{[(\Sigma f_i/2) - F_i\,(ant)] \cdot h}{f_{Md}}$$

Onde:

- l_{Md} = limite inferior da classe Md
- $\Sigma f_i/2$ = n: tamanho da amostra ou número de elementos
- F_i (ant) = frequência anterior à classe Md
- h= amplitude da classe Md
- f_{Md} = frequência da classe Md

Exemplo 6.8

Dada a distribuição amostral, calcular a mediana (utilizar a Tabela 6.4).

Tabela 6.4 - Estatura de 40 alunos do colégio A.

I	Estaturas (cm)	f_i	F_i	
1	150 ├154	4	4	
2	154 ├158	9	13	
3	158 ├162	11	24	←classe Md
4	162 ├166	8	32	
5	166 ├170	5	37	
6	170 ├174	3	40	
		Σ = 40		

Fonte: CRESPO (2002).

Procedimentos:

Primeiro Passo: Calcula-se: **$\Sigma f_i/2$ (**Como Σf_i = 40, temos: 40/2 = 20º**)**

$\Sigma f_i/2 = 20$

Segundo Passo: Identifica-se a classe **Md** pela **F_i.**

***Fi* = 24** → a classe Md é 3ª ordem (de ordem 3**)**

Classe Md = 158 ├162

Assim $\mathbf{l_{Md}}$ = 158; F_1 (ant) = 13; $\mathbf{f_{Md}}$ = 11 h = 4 **$\Sigma f_i/2 = 20$**

Terceiro Passo: Aplicar a fórmula

$$Md = l_{Md} + \frac{[(\Sigma f_i/2) - F_i\ (ant)]\ .\ h}{f_{Md}}$$

$$Md = 158 + \frac{[(40/2) - 13]\ .\ 4}{11}$$

$$Md = 158 + \frac{[(20 - 13)]\ .\ 4}{11}$$

Md = 158 + 28/11

Md = 158 + 2,54

Md = 160,54

Md = 160,5 cm

Empregamos a mediana quando:

- Desejamos obter o ponto que divide a distribuição em duas partes iguais.
- Há valores extremos que afetam de maneira acentuada a média aritmética.
- A variável em estudo é salário.

Exemplo 6.9

Dada a distribuição amostral, calcular a mediana:

Tabela 6.5 – Mulheres ativas.

Classes (x_i)	f_i	F_i
35 ├45	5	5
45 ├55	12	17
55 ├65	18	35
65 ├75	14	49
75 ├85	6	55
85 ├95	3	58
	Σf_i **= 58**	

1.1.3 A moda (Mo) - É o valor mais frequente da distribuição.

I. Para distribuições simples (sem agrupamento em classes), a identificação da Moda é facilitada pela simples observação do elemento que apresenta maior frequência. Assim:

a) A série de dados: 7, 8. 9, 10, 10, 10, 11, 12, 13, 15, tem moda igual a 10. (10 é o valor que mais se repete).

b) Há séries nas quais não exista valor modal, isto é, nas quais nenhum valor apareça mais vezes que outros.

Exemplo 6.10

A série de dados: 3, 5, 8, 10, 12,13 não apresenta moda. A série é amodal.

c) Em outros casos, pode haver dois ou mais valores de concentração. Dizemos, então, que a série tem dois ou mais valores modais.

Exemplo 6.11

Já a série de dados: 2, 3, 4 ,4, 4, 5, 6, 7, 7, 7, 8, 9 apresenta duas modas: 4 e 7.

A série é bimodal.

Para a distribuição:

x_i	243	245	248	251	307
f_i	7	17	**23**	20	8

a) Fixar o valor da variável de maior frequência, que é 23.

A moda será 248.

Indica-se **Mo = 248**

II Para dados agrupados em classes (variáveis contínuas) há diversas fórmulas para o cálculo da moda.

Destacaremos o cálculo da moda por meio da fórmula de Czuber.

Primeiro passo: Identifica-se a classe modal (classe com maior frequência)

Segundo passo: Aplica-se a fórmula

$$\boxed{\mathbf{Mo} = \mathbf{l_{Mo}} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot \mathbf{h}}$$

Onde:

$\mathbf{l_{Mo}}$ = é o limite inferior da classe modal

$\Delta_1 = \mathbf{f_{Mo} - f_{ant}}$ = diferença entre a frequência da classe modal e a frequência da classe imediatamente anterior.

$\Delta_2 = \mathbf{f_{Mo} - f_{post}}$ = diferença entre a frequência da classe modal e a frequência da classe imediatamente posterior.

h = amplitude da classe modal

$\mathbf{f_{Mo}}$ = frequência da classe modal

$\mathbf{f_{ant}}$ = frequência da classe imediatamente anterior

$\mathbf{f_{post}}$ = frequência da classe imediatamente posterior

Exemplo 6.12

Assim, para a distribuição (Tabela 6.6), temos a Moda:

Tabela 6.6 - Estatura de 40 alunos do colégio A

I	Estaturas (cm)	f_i
1	150 ├154	4
2	154 ├ 158	9
3	158 ├ 162	11 ←classe modal
4	162 ├ 166	8
5	166 ├ 170	5
6	170 ├ 174	3
		Σf_i = 40

Fonte: CRESPO (2002).

Procedimentos:

Primeiro Passo: Identifica-se a classe modal (classe com maior frequência): ***fi* = 11**

Classe Modal: 158 - 162

Segundo Passo: Aplica-se a fórmula**:**

$$\boxed{\mathbf{Mo = l_{Mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h}}$$

Onde:

$l_{Mo} = 158$

$\Delta_1 = f_{Mo} - f_{ant} = 11 - 9 = 2 \rightarrow \Delta_1 = 2$

$\Delta_2 = f_{Mo} - f_{post} = 11\text{-}8 = 3 \rightarrow \Delta_2 = 3$

$h = 162 - 158 = 2 \rightarrow h =$

Então:

$$\boxed{Mo = l_{Mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h}$$

Mo = 158 +{[2/(2+3)] x 4} = 158 + 1,6 = 159,6

Mo = 159,6 cm

Exemplo 6.13

Calcular a moda para a distribuição:

Salários (R$)	80 ├180	180 ├250	250 ├300	300 ├500
Nº de empregados	70	140	140	60

Fonte: MARTINS (2010).

Observe:

Que as amplitudes (h) das classes não são iguais. Nesse caso, é preciso calcular as densidades das classes: f_i/h, para identificar qual a classe modal (aquela com maior densidade). Assim:

Salários (R$)	f_i	f/h
(100) 80 ├180	70	70/100 = 0,7
(70) 180 ├250	140	140/70 = 2,0
(50) 250 ├300	140	140/50 = 2,8 ← classe modal
(200) 300 ├500	60	60/200 = 0,3
	$\Sigma f_i = 410$	

Procedimentos:

Primeiro passo: Identifica-se a classe modal (a de maior densidade). Será a 3ª classe: 250 ├300.

Segundo Passo: Aplica-se a fórmula:

$$\boxed{Mo = l_{Mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h}$$

onde:

$l_{Mo} = 250$

$\Delta_1 = f_{Mo} - f_{ant} = 2,8 - 2,0 = 0,8 \rightarrow \Delta_1 = 0,8$

$\Delta_2 = f_{Mo} - f_{post} = 2,8 - 0,3 = 2,5 \rightarrow \Delta_2 = 2,5$

$h = 300 - 250 = 50$

Então: $Mo = 250 + [0,8/(0,8 + 2,5)\ 50] = 250 + [0,8/3,3]\ 50 = 250 + 12,12 = 262,12$

$Mo = 262,12$

Portanto, o salário mais frequente para esse grupo de 410 empregados é **R$ 262,12**

Nota:

A moda é utilizada:

- Quando desejamos obter uma medida rápida e aproximada de posição;
- Quando a medida de posição deve ser o valor mais típico da distribuição.

Exercício Proposto

Complete o esquema para o cálculo da moda da distribuição de frequência:

Tabela 6.7 – Tabela de custos do produto A.

I	Custos (R$)	f_I
1	450 ├ 550	8
2	550 ├ 650	10
3	650 ├ 750	11
4	750 ├ 850	16
5	850 ├ 950	13
6	950 ├ 1.050	5
7	1.050 ├ 1.150	1
		Σ = 64

Fonte: Fictício.

A classe modal é a de ordem ...

Logo:

l_{Mo} = ?........... e L_{Mo}= ?

Temos, pois: Mo = ?

1.1.4 As separatrizes – não são medidas de tendência central, mas estão ligadas à mediana relativamente à sua característica de separar a série em duas partes que apresentam o mesmo número de valores. Essas medidas - os quartis, os decis e os percentis - são, juntamente com a mediana, conhecidas pelo nome genérico de separatrizes.

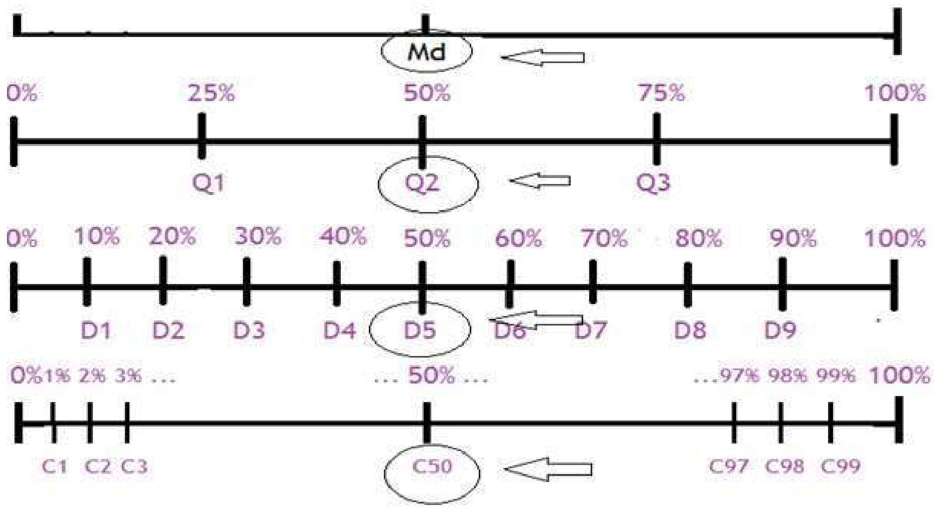


Figura 6.1 - Separatrizes. Fonte: OLIVEIRA (2016).

1.1.4.1 Os quartis - Denominamos quartis os valores de uma série que a dividem em quatro partes iguais (Figura 6.1).

Há, portanto, três quartis:

Primeiro Quartil: Q_1 = 1º quartil, deixa 25% dos elementos

Segundo Quartil: Q_2 = 2º quartil, coincide com a mediana, deixa 50% dos elementos

Terceiro Quartil: Q_3 = 3º quartil, deixa 75% dos elementos

- Eis as fórmulas para os cálculos de $\mathbf{Q_1}$ **e** $\mathbf{Q_3}$, para o caso de variáveis contínuas.
- Quando os dados são agrupados, para determinar os quartis usamos a mesma técnica do cálculo da mediana, substituir **n** por $\boldsymbol{\Sigma f_i/2}$ por $\boldsymbol{k\ \Sigma f_i/4}$, sendo k o número da ordem do quartil.

I. Determinação do 1º quartil:

Primeiro passo: Calcula-se a ordem $\boldsymbol{k\Sigma f_i/4}$ $\mathbf{(}\boldsymbol{k}\mathbf{ = 1)}$ e encontra-se $\boldsymbol{F_i}$

Segundo passo: Identifica-se Classe Q_1, através da F_i (contém o elemento identificador da classe Q_1), e identifica-se os termos: $\mathbf{l_{Q1}}$; F_i (ant); f_{Q1}; h

Terceiro passo: Aplica-se a fórmula;

$$\mathbf{Q_1 = l_{Q1} + \frac{[(\boldsymbol{k\Sigma f_i}/4) - \boldsymbol{F_i}\ (ant)]\ .\ h}{f_{Q1}}}$$

Onde:

$\mathbf{l_{Q1}}$ = limite inferior da classe Q_1
$\boldsymbol{k\Sigma f_i}$ **= n** → tamanho da amostra
$\boldsymbol{F_i}$ (ant)= frequência anterior à classe Q_1
$\mathbf{f_{Q1}}$ = frequência da classe Q_1
h = amplitude da classe Q_1

II. Determinação do 3º quartil

Primeiro passo: Calcula-se a ordem $3\boldsymbol{\Sigma f_i}/\mathbf{4}$ (k =3.) e encontra-se $\boldsymbol{F_i}$

Segundo passo: Identifica-se a classe $\mathbf{Q_3}$ **pela** $\mathbf{F_i}$ **(**contém o elemento identificador da classe Q_3).

Identifica-se os demais termos: $\mathbf{l_{Q3}}$; F_i (ant); f_{Q3}; h.

Terceiro passo: Aplica-se a fórmula:

$$Q_3 = \quad 1_{Q3} + \frac{[(3\Sigma fi/4) - F_i\,(ant)]\,.\,h}{f_{Q3}}$$

Em que:

l_{Q3} = limite inferior da classe Q_3
$k\Sigma f_i$ = **n** → tamanho da amostra
k **= 3**
F_i (ant)= frequência anterior à classe Q_3
f_{Q3} = frequência da classe Q_3
h = amplitude da classe Q_3

Exemplo 6.14

Dada a distribuição, determinar os quartis (**Q_1 e Q_3**) e a **mediana.**

Tabela 6.8 - Estatura de 40 alunos do colégio A.

Estaturas (cm)	f_i	F_i	
150 ├154	4	4	
154 ├ 158	9	13	←Classe Q_1 (contém o10º elemento)
158 ├ 162	11	24	**←Classe Md (contém o 20º elemento)**
162 ├ 166	8	32	←Classe Q_3 (contém o 30º elemento)
166 ├ 170	5	37	
170 ├ 174	3	40	
	$\Sigma f_I = 40$		

Fonte: CRESPO (2002).

Procedimentos:

- **Primeiro Quartil**

Primeiro passo: Calcula-se a ordem **$\Sigma f_i/4$ = n; (k = 1)**

$\Sigma f_i/4 = 40/4 = 10 \rightarrow F_I = 13$ (contém o 10º elemento)

Segundo passo: Identifica-se Classe Q_1 pela F_i =13 (contém o 10º elemento)

Classe Q_1 = 2ª classe: 154 ├ 158

Temos: F_i(ant) à Classe Q_1 = **4; → F_i (ant) = 4** **l_{Q1} =154** f_{Q1} = 9

h = 158 – 154 = 4 →h = 4 e $\Sigma f_i/4$= 10

Terceiro passo: Aplica-se a fórmula:

$$Q_1 = \quad l_{Q1} + \frac{[(k\Sigma f_i/4) - F_i\,(ant)]\,.\,h}{f_{Q1}}$$

$$Q_1 = 154 + \frac{(10 - 4) \cdot 4}{9}$$

$Q_1 = 154 + (24)/9$

$Q_1 = 154 + 2{,}66$

$\mathbf{Q_1 = 156{,}7}$ **cm**

- **Determinação do terceiro quartil**

Primeiro passo: Calcula-se a ordem $3\Sigma f_i/4$ ($k = 3$.)

$3\ \Sigma f_i/4 = (3 \times 40)/4 = 30$º elemento → $F_3 = 32$ (contém o 30º elemento)

Segundo passo: Identifica-se a classe Q_3 pela $F_3 = 32$(contém o 30º elemento)

Classe Q_3 é a 4ªclasse: 162 ├ 166

Temos: $h = 166 - 162 = 4$ $f_{Q3} = 8$ F_i (ant) à classe $Q_3 = 24$ → F_i (ant)=24

$l_{Q3} = 162$ $3\ \Sigma f_i/4 = 30$

Terceiro passo: Aplica-se a fórmula**:**

$$\mathbf{Q_3 = l_{Q3} + \frac{[(3\Sigma f_i/4) - F_i\ (ant)] \cdot h}{f_{Q3}}}$$

$$Q_3 = 162 + \frac{(30 - 24) \cdot 4}{8}$$

$Q_3 = 162 + 24/8$

$Q_3 = 162 + 3 = 165$

$Q_3 = 165$ cm

III. Determinação da mediana (= Q_2):

Procedimentos:

Primeiro passo: Calcula-se a ordem $\boldsymbol{\Sigma f_i/2}$ (A variável é contínua**,** independentemente se n é par ou ímpar**).**

Como: $\Sigma f_i = 40 \rightarrow 40/2 = 20$º (vigésimo elemento)

$\boldsymbol{F_1}$ **= 24 (**que contém o vigésimo elemento**)** ← **Classe mediana**

Segundo Passo: Identifica-se a classe **Md pela** $\boldsymbol{F_i}$.

$\boldsymbol{Fi}$ **= 24** → a classe Md é 3ª ordem: Classe Md = 158 ├ 162

Temos: $l_{Md} = 158$; $\boldsymbol{F_1}$ **(ant)** = 13; $\mathbf{f_{Md}} = 11$ $h = 4$ $\Sigma f_i = 20$

Terceiro Passo: Aplicar a fórmula:

$$\mathbf{Md} = \mathbf{l_{Md}} + \frac{\mathbf{[(\Sigma \mathit{f_i}/2) - \mathit{F_i}\ (ant)] \cdot h}}{\mathbf{f_{Md}}}$$

$$Md = 258 + \frac{(20 - 13) \cdot 4}{11}$$

Md= 158 + 28/11

Md = 158 + 2,54

Md = 160,5 cm

Então, $\mathbf{Q_2}$ **= 160,5 cm**

1.1.4.2 Os decis – são os valores que dividem a série em 10 partes iguais.

Indicamos os decis: $\mathbf{D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9}$.

$\mathbf{D_1}$ **= 1º decil**, deixa 10% dos elementos

$\mathbf{D_5}$ **= 5º decil**, coincide com a mediana, deixa 50% dos elementos

$\mathbf{D_9}$ **= 9º decil**, deixa 90% dos elementos.

Deste modo precisamos de **9 decis** para **dividirmos uma série em 10 partes iguais.**

I O cálculo para um decil (D_i) é dado por:

Primeiro passo: Calcula-se a ordem $\mathbf{k\Sigma f_i/10}$, em que **i= 1,2,3,4,5, 6,7,8** e **9**

k **= ordem**

Segundo passo: Identifica-se a classe D_i pela $\boldsymbol{F_i}$ **(**que contém o elemento da classe)

Terceiro passo: Aplica-se a fórmula:

$$\mathbf{Di} = \mathbf{1_{Di}} + \frac{\mathbf{[(\mathit{k}\Sigma \mathit{fi}/10) - \mathit{F_i}\ (ant)] \cdot h}}{\mathbf{f_{Di}}}$$

Onde:

$\mathbf{l_{Di}}$ = limite inferior da classe $D_i \rightarrow i = 1, 2, 3, 4, 5, 6,7,8$ e 9.

K = tamanho da amostra

$\boldsymbol{\Sigma fi}$ **=** soma das frequências anteriores à classe D_i

h = amplitude da classe D_i

$\mathbf{f_{Di}}$ **=** frequência da classe D_i

Exemplo 6.15

Determinar o **3º decil** da tabela anterior com classes (**Tabela 6.9**).

Tabela 6.9 - Estatura de 40 alunos do colégio A

Estaturas (cm)	f_i	F_i	
150 ├ 154	4	4	← **Classe** P_8 (contém o (3,2)º elemento)
154 ├ 158	9	13	← **Classe** D_3 (contém o 12º elemento)
158 ├ 162	11	24	
162 ├ 166	8	32	
166 ├ 170	5	37	
170 ├ 174	3	40	
	$\Sigma f_I = 40$		

Fonte: CRESPO (2002).

Solução:

Cálculo do 3º Decil: D_3

Procedimentos:

Temos: $k = 3$ $\Sigma fi = 40$

Primeiro passo: Calcula-se a ordem $k\Sigma f_i/10$

$k\Sigma fi/10 = (3 \text{ x } 40)/10 = 12º$ elemento→ $F_i = 13$ (contém o 12º elemento)

Segundo passo: $F_3 = 13$ **→ classe** D_3 é a 2ª classe: 154 ├ 158

Temos: $h = 158 - 154 = 4$ $f_{D3} = 9$ F_i (ant)= 4 $3\Sigma f_i = 12$ $l_{D3} = 154$

$$D_3 = l_{D3} + \frac{[(3\Sigma fi/10) - F_i \text{ (ant)}] \cdot h}{f_{Q3}}$$

$$D_3 = 154 + \frac{(12 - 4) \cdot 4}{9}$$

$D_3 = 154 + 32/9$

$D_3 = 154 + 3,55$

$D_3 = 157,55$

Logo: $D_3 = 157,55$ **cm**

1.1.4.3 Os percentis ou centis – são os noventa e nove valores que separam a série em 100 partes iguais (Figura 6.1).

Indicamos $P_1, P_2, ..., P_{32}, ..., P_{99}$ **ou** $C_1, C_2, ..., C_{32}, ..., C_{99}$

Observação:

- É evidente que: $\mathbf{P_{50} = Md}$, $\mathbf{P_{25} = Q_1}$ e $\mathbf{P_{75} = Q_3}$

I O cálculo de um percentil (P_I) é dado por:

Primeiro Passo: calcula-se a ordem $\mathbf{k\Sigma f_i/100}$, em que k é a ordem.

k = 1, 2, 3, ..., 98,99.

Segundo Passo: Pela $\mathbf{F_i}$ **(**frequência acumulada**)** identifica-se a classe P_i.

Terceiro Passo: Usa-se a fórmula:

$$P_i = 1_{pi} + \frac{[(k\Sigma fi/100) - F_i\,(ant)]\ .\ h}{f_{pi}}$$

Onde:

l_{Pi} = limite inferior da classe Pi i = 1, 2, 3,...99

k = tamanho da amostra

Σf_i = soma das frequências anteriores à classe P_i

h = amplitude da classe P_1

f_{Pi} = frequência da classe P_i

Exercício:

Considerando a distribuição da Tabela 6.9, determinar o **8º percentil**.

Solução:

Temos: $k = 8$ $\Sigma f_i = 40$

Primeiro Passo: calcula-se a ordem $k\Sigma f_i/100$

$k\Sigma f_i/100$ = (8 . 40)/100 = (3,2)º elemento

$F_1 = 4$ (contém o (3,2)º elemento→ F_i **(ant) = 0**

Segundo Passo: pela F_i identifica-se a classe P_i:

$F_1 = 4 \rightarrow P_8$ é a 1ª classe: 150 ├154

Temos: **h = 154 - 150 = 4** $\mathbf{f_{Pi} = 4}$ F_i **(ant) = 0** $k\Sigma f_i/100$ **=3,2**

Terceiro Passo: Usa-se a fórmula:

$$P_8 = 1_{p8} + \frac{[(k\Sigma fi/100) - F_i\,(ant)]\ .\ h}{f_{pi}}$$

$$P_8 = 150 + \frac{(32 - 0)\ .\ 4}{4}$$

$P_8 = 150 + 12{,}8/4$

$P_8 = 150 + 3{,}2$

$\mathbf{P_8 = 153{,}2\ cm}$

Exercício proposto

Dada a distribuição abaixo, determinar o 1º quartil, o 4º decil e o 72º percentil:

Classes	f_i	F_i
4 ├ 9	8	8
9 ├ 14	12	20
14 ├ 19	17	37
19 ├ 24	3	40
	$\Sigma f_i = 40$	

CAPÍTULO VII - MEDIDAS DE DISPERSÃO OU DE VARIABILIDADE

1. DISPERSÃO OU VARIABILIDADE – são medidas estatísticas utilizadas para avaliar o grau de variabilidade, ou dispersão, dos valores em torno da média. Servem para medir a representatividade da média.

Exemplo 7.1

Sejam as séries:

A: 20, 20 20

B: 15, 10, 20, 25, 30

a) Calculando a média aritmética de cada um desses conjuntos, obtemos:

$$\bar{x}_A = \Sigma x_i/n \rightarrow \bar{x}_A = 60/3 = 20 \rightarrow \bar{x}_A = 20$$

$$\bar{x}_B = \Sigma x_i/n \rightarrow \bar{x}_B = 100/5 = 20 \rightarrow {}_B = 20$$

Notas:

1) Os dois conjuntos apresentam a mesma média aritmética;
2) O conjunto A é mais homogêneo que o conjunto B, pois todos os valores são iguais a média (não apresenta dispersão em torno da média);
3) O conjunto B, apesar de ter média igual ao conjunto A, os valores da série B apresentam dispersão ou variabilidade em torno da média $\bar{x}_B$.
4) Para qualificar os valores de uma dada variável, ressaltando a maior ou menor dispersão ou variabilidade entre esses valores e sua medida de posição, a Estatística recorre às medidas de dispersão ou de variabilidade.

1.1 TIPOS DE MEDIDAS DE DISPERSÃO - Medidas de Dispersão Absoluta e Medidas de Dispersão Relativa.

1.1.1 Medidas de dispersão absoluta

I. Amplitude total (AT) – É a medida de dispersão dada pela diferença entre o maior e o menor valor da série.

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

Exemplo 7.2

Para as séries: A= {**10, 12, 20, 22, 25, 33, 38**}

B= {17, 18, 19, 20, 21, 22, 23} Calcular a amplitude total.

Solução:

Para o conjunto A, temos: **AT** = 38 – 10 = **28→ AT = 28**

Para o conjunto B, temos: **AT** = 23 – 17 = **6 → AT = 6**

a) A utilização da amplitude total como medida de dispersão é limitada, pois, sendo uma medida que depende apenas dos valores externos, não capta possíveis variações entre esses limites.
b) Se os dados vierem dispostos em uma tabela de frequências, com os valores agrupados em classes, há duas formas de se definir a amplitude total:

- **Primeiro Método**: **AT** = É a diferença entre o ponto médio da última classe e o ponto médio da primeira classe.
- **Segundo Método**: **AT** = É a diferença entre o limite superior da última classe e o limite inferior da primeira classe.

Exemplo 7.3

Calcular a amplitude total dos valores dispostos na Tabela 7.1.

Tabela 7.1 - Estatura de 40 alunos do colégio A

I	Classes	f_i
1	150 ├154	4
2	154 ├ 158	9
3	158 ├ 162	11
4	162 ├ 166	8
5	166 ├ 170	5
6	170 ├ 174	3
		$\Sigma f_1 = 40$

Fonte: CRESPO (2002).

Solução:

a) Pelo primeiro método:

Primeiro passo: Ponto médio da última classe: 170 ├ 174

$X_i = (l_i + L_i)/2 \rightarrow X_6 = (170 + 174)/2 = \mathbf{172}$

Segundo passo: Ponto médio da primeira classe: 150 ├154

$X_1 = (150 + 154)/2 = 152$

Terceiro passo: Diferença entre os pontos médios: $X_6 - X_1 = AT$

$AT = 172 – 152 = 20 \rightarrow \mathbf{AT = 20}$

Amplitude total = 20

Neste método os valores extremos são eliminados.

b) Pelo segundo método:

Última classe: 170 ├ 174 e primeira classe: 150 ├154

$AT = (L – l) \rightarrow AT= 174 – 150 = 24 \rightarrow AT = 24$

Amplitude total = 24

Nota: Embora a amplitude total seja a mais simples das medidas de dispersão, há uma forte restrição ao seu uso em virtude de sua grande instabilidade, uma vez que ela leva em conta apenas os valores extremos da série.

II Variância amostral (S^2) - A variância (S^2), de uma amostra de **n** medidas, é igual à soma dos quadrados dos desvios em relação à média, ou seja, a variância é expressa pelo quadrado da unidade de medida da variável que está sendo estudada.

[Σd_i^2, dividida por (n-1)], assim

$$S^2 = \frac{1}{(n-1)}\left[\Sigma x^2_i - \frac{(\Sigma x_i)^2}{n}\right]$$

ou

Para dados agrupados

$$S^2 = \frac{1}{(n-1)}\left[\Sigma x^2_i F_i - \frac{(\Sigma x_i F_i)^2}{n}\right]$$

- Quanto maior o valor de S^2, maior a dispersão dos dados amostrais.

Exemplo 7.4

Calcular a variância para as medidas amostrais de altura de mudas de alface: 3, 7, 2, 1, 8 cm.

Solução: Vamos determinar S^2 pela fórmula básica. Para tanto, é interessante a construção da seguinte tabela:

Tabela 7.2 - Altura de plântulas de cinco cultivares de alface, após 15 dias da semeadura.

Cultivar	**x_i (cm)**	**x_i^2**
1	3	9
2	7	49
3	2	4
4	1	1
5	8	64
	Σ = 21	**Σ =127**

Fonte: Fictício.

A variância amostral será:

$$\mathbf{S^2 = \frac{1}{(n-1)}\left[\Sigma x^2_i - \frac{(\Sigma x_i)^2}{n}\right]}$$

n (número de elementos da amostra) → $\mathbf{x_i = 5}$

$S^2 = 1/(5-1)\ [127 – (21)^2/5]$

$S^2 = ¼\ [127 – 441/5]$

$S^2 = ¼\ [127 - 88,2]$

$S^2 = ¼\ [38,8]$

$S^2 = 9,7$

Então, a variância amostral será: $S^2 = 9,7$

Exemplo 7.5

Calcular a variância da seguinte distribuição amostral:

Tabela 7.3 – Dados de distribuição amostral.

X_i	f_i	$X_i f_i$	$X_i^2 f_i$
5	2	10	50
7	3	21	147
8	5	40	320
9	4	36	324
11	2	22	242
	Σ = 16	**Σ = 129**	**Σ = 1.083**

Fonte: Fictício.

Procedimentos:

Temos:

$x_i^2 f_i = 5^2 * 2 = 50$; $x_2^2 f_2 = 7^2 * 3 = 147$ e assim até $x_5^2 f_5 = 11^2 * 2 = 242$

Lembrando que $\mathbf{\Sigma f_i = n \rightarrow n = 16}$

Aplicando a fórmula:

$$\mathbf{S^2 = \frac{1}{(n-1)}\left[\Sigma x^2_i f_i - \frac{(\Sigma x_i f_i)^2}{n}\right]}$$

$S^2 = 1/(16 - 1)\ [1.083 - (129)^2/16]$

$S^2 = 1/15\ [1.083 - (16.641)/16]$

$S^2 = 1/15\ [(17.328 - 16.641)/16]$

$S^2 = 1/15\ [(17.328 - 16.641)/16]$

$S^2 = 1/15(42{,}9375)$

$S^2 = 42{,}94/15$

$S^2 = 2{,}86$

Logo, a variância amostral é 2,86.

III. Desvio - padrão amostral (S): É uma medida definida como a raiz quadrada da variância, que será expresso na unidade de medida original. Assim. $\mathbf{S = \sqrt{S^2}}$.

a) O desvio padrão do exemplo 1 (variância), temos $S^2 = 9{,}7$, e dado por:

$S^2 = 9{,}7 \rightarrow S = \sqrt{9{,}7} = 3{,}1$;

Logo, o desvio padrão é 3,1

b) No **exemplo 7.5** (variância), temos $\mathbf{S^2 = 2{,}86}$, então o desvio-padrão será:

$S = \sqrt{2{,}86} = 1{,}69$;

Logo o desvio-padrão é 1,69

O desvio-padrão pode ser calculado por estes métodos mais práticos:

$$S = \sqrt{\frac{\Sigma xi^2}{n} - \left[\frac{\Sigma xi}{n}\right]^2} \qquad \text{Ou}$$

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

Exemplo 7.6

Calcular o desvio-padrão da produtividade de soja de sete cultivares (Sc/ha):

A = {40, 45, 48, 52, 54,62,70}

Solução

O modo mais prático é construir uma Tabela com duas colunas: uma para x_i e outra para x_i^2.

Assim:

Tabela 7.4 – Produtividade (Sc/ha) de sete cultivares de soja na fazenda Boa Esperança

Cultivares	**x_i**	**x_i^2**
1	40	1.600
2	45	2.025
3	48	2.304
4	52	2.704
5	54	2.916
6	62	3.844
7	70	4.900
	Σ = 371	**Σ = 20.293**

Fonte: Fictício

Procedimentos:

a) Como n = 7, temos:

$$S = \sqrt{\frac{\Sigma xi^2}{n} - \left[\frac{\Sigma xi}{n}\right]^2}$$

$$S = \sqrt{\frac{20.293}{7} - \left[\frac{371}{7}\right]^2}$$

$S = \sqrt{2,899 - (53)^2}$

$S = \sqrt{2.899 - 2.809}$

$S = \sqrt{90}$

$S = 9,486$

Logo: $S = 9,49 \rightarrow$ O desvio padrão é 9,5

Exemplo 7.7

Calcular o desvio-padrão da distribuição:

Tabela 7.5

x_i	f_i	$f_i\ x_i$	$f_i\ x_i^2$
0	2	0	0
1	6	6	6
2	12	24	48
3	7	21	63
4	3	12	48
	$\Sigma = 30$	$\Sigma = 63$	$\Sigma = 165$

Fonte: Fictício

Lembrando que **$\Sigma f_i = n \quad \rightarrow n = 30$**

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

$S = \sqrt{165/30 - (63/30)^2}$

$S = \sqrt{5,5 - 4,41}$

$S = \sqrt{1,09}$

$S = 1,044$

Logo, $S = 1,04 \rightarrow$ O desvio padrão é 1,04

Exemplo 7.8

A distribuição amostral representa a idade de 50 funcionários da empresa XPTO - determinar a variância, o desvio-padrão, e constatar as regras para interpretação do desvio-padrão.

Solução

Tabela 7.6 – Idade dos funcionários da empresa XPTO.

Intervalos das classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$
18 ├ 25	6	21,5	129	2.773,50
25 ├ 32	10	28,5	285	8.122,50
32 ├ 39	13	35,5	461,50	16.383,50
39 ├ 46	8	42,5	340	14.450,00
46 ├ 53	6	49,5	297	14.701,50
53 ├ 60	5	56,5	282,50	15.961,25
60 ├ 67	2	63,5	127	8.064,50
	Σ = 50		Σ = 1.922	Σ = 80.456,50

Fonte: MARTINS (2010).

Procedimentos:

Primeiro passo: Calcula-se o ponto médio das classes **(x_i):**

$$\mathbf{x_1 = (l_i + L_i)/2}$$

a) Para primeira classe: 18 ├ 25 →x_1 = (18 + 25)/2) = 21,5.... e assim calcula-se x2....x3...x4....x5....x6 e x7

b) Para última classe: x7 = 60 ├ 67 = (60 + 67)/2 = 63,5

c) $\Sigma f_i = n \rightarrow n = 50$

Segundo Passo: Calcula-se a Média amostral ($\bar{x}$)

Média amostral: $\bar{x} = (\Sigma x_i f_i)/n \rightarrow \bar{x} = 1.922/50 = 38,44$ anos

Média amostral = 38,44 anos

Terceiro Passo: Calcula-se a Variância Amostral (S^2)

$$\mathbf{S^2 = \frac{1}{(n-1)}\left[\Sigma x_i^2 f_i - \frac{(\Sigma\, x_i f_i)^2}{N}\right]}$$

$S^2 = 1/(50-1)\ [80.456,50 - (1.922)^2/50]$

$S^2 = 1/49\ [80.456,50 – 73.881,60]$

$S^2 = 6.574,90/49$

$S^2 = 134,18 \leftarrow$ Variância Amostral

Quarto Passo: Calcula-se o Desvio–padrão (S) → $\mathbf{S = \sqrt{S^2}}$

$S = \sqrt{134,18}$

$S = 11,58$

Logo, $S = 11,58$ anos ← Desvio padrão.

Exercício proposto

Calcular a variância e o desvio padrão da seguinte distribuição amostral:

Classes	2 ├4	4 ├6	6 ├8	8 ├10	10 ├12
f_i	3	5	8	6	3

Fonte: MARTINS (2010).

Solução:

a) É conveniente a construção da Tabela7.7:

Tabela 7.7 –

Classes	f_i	x_i	x_if_i	$x_i^2f_i$
2 ├4	3	3	9	81
4 ├6	5			
6 ├8	8			
8 ├10	6			
10 ├12	3			
	$\Sigma f_i= ?$		$\Sigma x_if_i= ?$	$\Sigma x_i^2f_i = ?$

Procedimentos:

Primeiro passo: Calcula-se o ponto médio das classes: $\mathbf{x_i} =$ ponto médio

$$\mathbf{X_1 = (l_i + L_i)/2}$$

a) Para primeira classe: 2 ├4 → $X_1 = (2+ 4)/2 = 3$→ $x_1 = 3$

b) Para segunda classe:

c) Para terceira classe:

d) Para quarta classe:

e) Para quinta classe**:**

Segundo Passo: completar a Tabela 7.7 com os dados.

Terceiro Passo: Calcula-se a Média amostral ($\bar{x}$).

Média amostral: $\bar{x} = (\mathbf{\Sigma xi\ f_i)/n}$

Quarto Passo: Calcula-se a Variância Amostral $(\mathbf{S^2})$

$$S^2 = \frac{1}{(n-1)}\left[\Sigma x_i^2 f_i - \frac{(\Sigma x_i f_i)^2}{n}\right]$$

Quinto Passo: Calcula-se o Desvio –padrão **(S)** → $S = \sqrt{S^2}$

1.1.2 MEDIDAS DE DISPERSÃO RELATIVA

I Coeficiente de variação[14] de pearson: CV_P – serve para medir a dispersão relativa de uma distribuição. É igual ao quociente entre o desvio –padrão e a média aritmética (TOLEDO, OVALLE, 2008).

$$\boxed{CV = \frac{S}{\bar{x}} \cdot 100}$$

Onde:

S = desvio padrão amostral;

$\bar{x}$ = média amostral

- Eis algumas regras empíricas para interpretações do coeficiente de variação**:**

  Se: CV < 15% → há baixa dispersão

  Se: 15% ≤ CV < 30% → há média dispersão

  Se: CV ≥ 30% → há elevada dispersão

Exemplo 7.9

Em uma empresa, o salário médio dos homens é de R$ 4.000, com desvio padrão de R$ 1.500, e o salário médio das mulheres é de R$ 3.000, com desvio padrão de R$ 1.200. A dispersão relativa dos salários é maior para os homens?

Solução:

Dos dados do problema temos:

Homens: $\bar{x}_H = 4.000$ $S_H = 1.500$

Mulheres: $\bar{x}_M = 3.000$ $S_M = 1.200$

Para os homens: $CV = (S_H / \bar{x}_H) \times 100 = (1.500/4.000) \times 100 = 37,5\%$

[14] Coeficiente de Variação ou coeficiente de variação relativa é uma porcentagem cujo cálculo resulta da comparação entre o desvio-padrão ou o desvio médio e a média ou a mediana. (TOLEDO, OVALLE, 2008.p.220)

Para as mulheres: CV = (S_M / $\bar{x}_M$) x 100 = (1.200/3.000) x 100 = 40%

Resposta:

Os salários das mulheres têm dispersão relativa maior do que os salários dos homens. As duas distribuições apresentam elevada dispersão (CV ≥ 30%).

Observação:

Para o Exemplo 7.8 (Tabela 7.6), onde:

- Média amostral: $\bar{x}$ = (Σ Xi F_i) n→ $\bar{x}$ = 1922/50 = 38,44 anos e
- Desvio –padrão: S = 11,58 anos, temos:
- CV_P = (11,58/38,44) x 100 = 0,30124 x 100 =30,12
- Daí, CV_P = 30,12% há elevada dispersão.

II. Coeficiente de variação de Thorndike: CV_T - é igual ao quociente entre o desvio padrão e a mediana.

$$CV = \frac{S}{Md} \cdot 100$$

Onde:

S = desvio padrão amostral

Md = mediana

Exemplo 7.10

Calcular o coeficiente de Thorndike com os dados do exemplo 7.7 (desvio-padrão) referente à distribuição (Tabela 7.5).

x_i	f_i	$f_i\ x_i$	$f_i\ x_i^2$
0	2	0	0
1	6	6	6
2	12	24	48
3	7	21	63
4	3	12	48
	Σ=30	Σ=63	Σ=165

Solução:

Os dados já são conhecidos:

S = 1,04

$\Sigma f_i/2 \rightarrow 30/2 = 15 \rightarrow Md = 2$

$CV_T = (1{,}04/2) \times 100 \rightarrow CV_T = 0{,}52 \times 100 = 52\%$

$CV_T = 52\%$.

III. Escore padronizado (Z_i): para uma medida xi. É dado por:

$$Z_i = \frac{X_i - \bar{x}}{S}$$

Onde:

$\bar{x}$ = média amostral;

S = desvio padrão amostral.

a) Um escore Z_i negativo indica que a observação x_i está à esquerda da média, enquanto um escore positivo indica que a observação está à direita da média.

Exemplo 7.11

Dadas as médias e os desvios padrões das avaliações de duas disciplinas:

Português	Matemática
$\bar{x}_P = 6{,}5$	$\bar{x}_M = 5{,}0$
$S_P = 1{,}2$	$S_M = 0{,}9$

Relativamente às disciplinas Português e Matemática, em qual delas obteve melhor performance um aluno com 7,5 em Português e 6,0 em Matemática?

Solução

Primeiro passo: Vamos determinar os escores padronizados para as notas obtidas:

Nota de Português: $Z_P = (7{,}5 - 6{,}5)/1{,}2 \rightarrow Z_P = 0{,}83$

Nota de Matemática: $Z_M = (6{,}0 - 5{,}0)/0{,}9 \rightarrow Z_M = 1{,}11$

A melhor performance relativa deu-se na disciplina Matemática, pois $Z_M > Z_P$. Observe que, em termos absolutos, o aluno conseguiu melhor nota em Português.

IV. Detectando outliers - dado descartável.

Nos trabalhos de coleta de dados, podem ocorrer observações que fogem das dimensões esperadas – os outliers. Para detectá-los, pode-se calcular o escore padronizado

(Z_i) e considerar outliers as observações cujos escores, em valor absoluto (módulo), sejam maiores que 3.

Exemplo 7.12

Os dados de uma pesquisa revelaram média 0,243 e desvio padrão 0,052 para determinada variável. Verificar se os dados 0,380 e 0,455 podem ser considerados observações da referida variável.

Solução:

$$Z_i = \frac{Xi - \bar{x}}{S}$$

a) $\bar{x}$ = 0,243 S= 0,052

b) Para x_i = 0,380 Z_i = (0,380 – 0,243)/0,052 = 2,63

c) Para x_i = 0,455 Z_i = (0,455 – 0,243)/0,052 = 4,08 (descartável)

Resposta:

O dado 0,380 pode ser considerado normal, por outro lado, 0,455 pode ser um outliers, (4,8 acima de 3), portanto descartável.

CAPÍTULO VIII - MEDIDAS DE ASSIMETRIA E CURTOSE

1. MEDIDAS DE ASSIMETRIA E CURTOSE - As medidas de assimetria e curtose, juntamente com as medidas de posição e de dispersão, completam o quadro das estatísticas descritivas que proporcionam a descrição e compreensão completas da distribuição de frequências estudadas.

1.1 MEDIDAS DE ASSIMETRIA - denomina-se assimetria (ou enviesamento) o grau de afastamento, de uma distribuição, da unidade de simetria.

- Em uma distribuição simétrica há igualdade de valores da média, mediana e moda.
- A assimetria, porém, torna-as diferentes, e essa diferença é tanto maior quanto maior a assimetria.

1.1.1 Principais medidas de assimetria - Método de comparação entre medidas de tendência central:

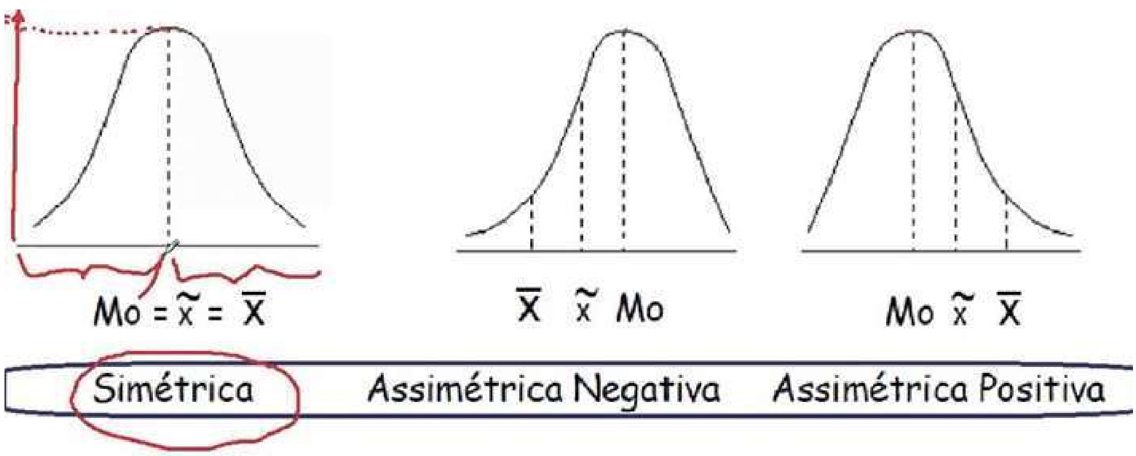


Figura 8.1 – Medidas de assimetrias. Fonte: CORREIA (2014).

- Assimétrica negativa ou assimétrica à esquerda: predominam valores inferiores à moda: $\overline{x}$ < Md < Mo (média, é menor que a mediana, e esta é menor que a moda) (Figura 8.1);
- Simétrica: as três medidas de tendência central mais usadas: moda, média aritmética e mediana, serem iguais: $\overline{x}$ = Md = Mo (Figura 8.1);

- Assimétrica positiva ou assimétrica à direita: predominam valores superiores à moda. Os valores concentram-se na extremidade inferior da escala (Figura 8.1).

$Mo < Md < \bar{x}$ (moda é menor que a mediana, e esta é menor que a média).

Estas medidas servem para avaliar o grau de assimetria de uma distribuição, ou seja, a intenção relativa com que uma curva de frequência se desvia da simetria. A comparação é simples:

$\bar{x}$ **> Mo** → Assimetria positiva

$\bar{x}$ **= Mo** → Simetria

$\bar{x}$ **< Mo** → Assimetria negativa

Exemplo 8.1

a) Verificar o tipo de assimetria das distribuições de frequências das Tabelas A, B e C:

Tabela A

Classes	**f_i**
10 ├ 20	5
20 ├ 30	10
30 ├ 40	15
40 ├ 50	20
50 ├ 60	5
	Σ = 55

Tabela B

Classes	**f_i**
10 ├ 20	5
20 ├ 30	10
30 ├ 40	15
40 ├ 50	10
50 ├ 60	5
	Σ = 45

Tabela C

Classes	**f_i**
10 ├ 20	5
20 ├ 30	20
30 ├ 40	15
40 ├ 50	10
50 ├ 60	5
	Σ = 55

Procedimentos:

Primeiro Passo: Calcula-se os pontos médios (x_i)

Primeira classe: 10 ├ 20 → ponto médio = x_i → $x_i = (10 + 20)/2 = 15$

Segunda classe: 20 ├ 30 → $x_i = (20 + 30)/2 = 25$

Terceira classe: 30 ├ 40 → $x_i = 35$

Quarta classe: 40 ├ 50 → $x_i = 45$

Quinta classe: 50 ├ 60 → $x_i = 55$

Segundo Passo: Calcular ($x_i f_i$, $x_i^2 f_i$ e F_i)

Terceiro Passo: Completar a Tabela A.1, com os dados encontrados

Tabela A.1 - x_i, $x_i f_i$, $x_i^2 f_i$ e F_i, referentes aos dados da Tabela A.

Classes	**f_i**	**Xi**	**$x_i f_i$**	**$x_i^2 f_i$**	**Fi**
10 ├ 20	5	15	75	1.125	5
20 ├ 30	10	25	250	6.250	15
30 ├ 40	15	35	525	18.375	30
40 ├ 50	**20**	45	900	40.500	50
50 ├ 60	5	55	275	15.125	55
	Σ = 55		Σ = 2.025	Σ = 81.375	

b) Proceder com as demais tabelas.

Tabela B.1 - x_i, $x_i f_i$, $x_i^2 f_i$ e F_i, referentes aos dados da Tabela B.

Classes	**f_i**	**x_i**	**$x_i f_i$**	**$x_i^2 f_i$**	**Fi**
10 ├ 20	5	15	75	1.125	5
20 ├ 30	10	25	250	6.250	15
30 ├ 40	**15**	35	525	18.375	30
40 ├ 50	10	45	45	20.250	40
50 ├ 60	5	55	275	15.1 5	45
	Σ = 45		**Σ = 1.575**	**Σ = 61.125**	

Tabela C.1 - x_i, $x_i f_i$, $x_i^2 f_i$ e F_i, referentes aos dados da Tabela C.

Classes	**f_i**	**x_i**	**$x_i f_i$**	**$x_i^2 f_i$**	**F_i**
10 ├ 20	5	15	75	1.125	5
20 ├ 30	**20**	25	500	12.500	25
30 ├ 40	15	35	525	18.375	40
40 ├ 50	10	45	450	20.250	50
50 ├ 60	5	55	275	15.125	55
	Σ = 55		**Σ = 1.825**	**Σ = 67.375**	

Quarto Passo: Cálculo das Médias

$\overline{x}$ **= (Σ x_i f_i)/n** → **n = Σf_i**

Da Tabela A.1 → $\overline{x}$ = 2.025/55 = 36,82

Da Tabela B.1. → $\overline{x}$ = 1.575/45 = 35

Da Tabela C.1. → $\overline{x}$ = 1.825/55 = 33,18

Quinto Passo: Cálculo das Modas (Método de Czber).

$$\boxed{Mo = l_{Mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h}$$

Da Tabela A.1 → Identifica-se a **classe modal** (classe com maior frequência):

fi = 20 → classe modal: 40 ├ 50 h = 50-40 = 10

Temos: l_{Mo} = 40

Δ_1 = f_{Mo} - f_{ant} → Δ_1 = 20 – 15 = 5 → Δ_1 = 5

Δ_2 = f_{Mo} - f_{post} → Δ_2 = 20 – 5 = 15 → Δ_2 = 15

Mo = 40 + [(5)/20]10→ M_o = 40 + 2,5 = 42,5

Da Tabela A.1 → Mo = 42,5

Da Tabela B.1.

f_i = 15 → classe modal: 30 ├ 40 h = 10

Temos: l_{Mo} = 30

Δ_1 = f_{Mo} - f_{ant} → Δ_1 = 15 – 10 = 5 → Δ_1 = 5

Δ_2 = f_{Mo} - f_{post} → Δ_2 = 15 – 10 = 5 → Δ_2 = 5

Mo = 30 + [(5)/10]10 → Mo = 30 + 5 = 3

Da Tabela B.1. Mo = 35

Da Tabela C.1.

fi = 20 → classe modal: 20 ├ 30 h = 10

Temos: l_{Mo} = 20

Δ_1 = f_{Mo} - f_{ant} → Δ_1 = 20 – 5 = 15 → Δ_1 = 15

Δ_2 = f_{Mo} - f_{post} → Δ_2 = 20 – 15 = 5 → Δ_2 = 5

Mo = 20 + [(15)/20]10 = 20+ 7,5 = 27,50

Da Tabela C.1: Mo = 27,50

Sexto Passo: Avaliar o grau de assimetria da distribuição ou distribuições**:**

Primeira Distribuição: $\bar{x}$ = 36,82 e Mo = 42,5

$\bar{x}$ **< Mo**→ Assimetria Negativa

Segunda Distribuição: $\bar{x}$ = 35 e Mo = 35

$\bar{x}$ **= Mo** → Simetria

Terceira Distribuição: $\bar{x}$ **= 33,18 e Mo = 27,50**

$\bar{x}$ >**Mo** → Assimetria Positiva

1.1.2 Coeficiente de assimetria – Uma medida usada para avaliar o grau de assimetria

I. 1º Coeficiente de assimetria de pearson[15] (AS):

$$AS = \frac{\bar{x} - M_o}{S}$$

II. 2º Coeficiente de assimetria de pearson (A_s):

$$A_S = \frac{3(\bar{x} - Med)}{S}$$

Se:

AS = 0, diz-se que a distribuição é simétrica;

AS > 0, diz-se que a distribuição é assimétrica positiva;

AS < 0, diz-se que a distribuição é assimétrica negativa.

Nota:

Coeficiente de Asssimetria de Pearson:

- Permite comparar duas ou mais distribuições diferentes e avaliar qual é a mais assimétrica.
- Quanto maior o Coeficiente de Assimetria de Pearson, mais assimétrica é a curva.

[15] Karl Pearson. (TOLEDO e OVALLE, 2008.p.253).

Exemplo 8.2

Considerando as distribuições Tabela A.1 dada anteriormente, calcular o primeiro e o segundo coeficientes de assimetria de Pearson.

Solução:

Os valores necessários para determinar as medidas que comporão o coeficiente de Pearson já se encontram calculados nas tabelas construídas na solução do exemplo 1.

Tabela A.1 - x_i, $x_i f_i$, $x_i^2 f_i$ e F_i, referentes aos dados da Tabela A.

Classes	f_i	Xi	$x_i f_i$	$x_i^2 f_i$	Fi
10 ├ 20	5	15	75	1.125	5
20 ├ 30	10	25	250	6.250	15
30 ├ 40	15	35	525	18.375	30←classe Md
40 ├ 50	**20**	45	900	40.500	50
50 ├ 60	5	55	275	15.125	55
	$\Sigma = 55$		$\Sigma = 2.025$	$\Sigma = 81.375$	

Procedimentos:

Para aplicar as fórmulas é necessário calcular média, moda, desvio-padrão, mediana.
Primeira distribuição: Tabela A.1

Primeiro Passo: Calcular

Média: $\bar{x} = (\Sigma x_i f_i)/\Sigma f_i \rightarrow 2.025/55 = 36,82 \rightarrow \bar{x} = 36,82$
Moda: $Mo = 40 + [(5)/20]10 \rightarrow Mo = 42,5$

Desvio–padrão: $n = \Sigma f_i \rightarrow n = 55$ (calcular pela regra da variância).

$$S^2 = \frac{1}{(n-1)}\left[\Sigma x^2_i f_i - \frac{(\Sigma x_i f_i)^2}{N}\right]$$

$S^2 = 1/54[(81.375) – (2.025)^2/55]$

$S^2 = [81.375–74.556,82]1/54$

$S^2 = 6.818,18/54$

$S^2 = 126,26$

$S = \sqrt{126,26}$

$S = 11,24$

Desvio padrão: $S = 11,24$

Segundo passo: Cálculo da Mediana:

$$\mathbf{Md = l_{Md} + \frac{[(\Sigma f_i/2) - F_i\ (ant)]\ .\ h}{f_{Md}}}$$

Temos: $\Sigma f_i/2 = 55/2 = 27{,}5 \rightarrow F_i = 30$ (contém 27,5º elemento).

A classe mediana é de 3ª ordem: 30 ├ 40 $l_{Md} = 30$ h=10 F_i (ant) = 15

$$Md = 30 + \frac{(27{,}5 - 15)\ .\ 10}{15}$$

Md = 30 + 8,33

Md = 38,33

Mediana: Md = 38,33

Terceiro Passo: 1º coeficiente de assimetria de pearson:

$$AS = \frac{\bar{x} - M_o}{S}$$

AS = (36,82 - 42,5)/11,24

AS = -5,68/11,24

AS = - 0,51

AS < 0 → distribuição assimétrica negativa

Quarto Passo: 2º coeficiente de assimetria de pearson:

$$As = \frac{3\ (\bar{x} - Med)}{S}$$

As = 3 (36,82 – 38,33)/11,24

As = 3(-1,51)/11,24

As = -0,40

AS < 0 → distribuição assimétrica negativa

$\bar{x}$ <	Md <	Mo	→	Assimetria Negativa
↓	↓	↓		
36,82	38,33	42,5		

Nota:
Os primeiro e segundo coeficientes de Pearson diferem em valor, mas não em sinal, como pode ser observado no exemplo acima.

Exercício proposto

Considerando as distribuições Tabela B.1 e da Tabela C.1, dadas anteriormente, calcular o primeiro e o segundo coeficientes de assimetria de Pearson.

Tabela B.1 - x_i, $x_i f_i$, $x_i^2 f_i$ e F_i, referentes aos dados da Tabela B.

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$	Fi
10 ├ 20	5	15	75	1.125	
20 ├ 30	10	25	250	6.250	15
30 ├ 40	**15**	35	525	18.375	30
40 ├ 50	10	45	450	20.250	40
50 ├ 60	5	55	275	15.125	45
	Σ = 45		Σ = 1.575	Σ = 61.125	

Tabela C.1 - x_i, $x_i f_i$, $x_i^2 f_i$ e F_i, referentes aos dados da Tabela C.

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$	F_i
10 ├ 20	5	15	75	1.125	5
20 ├ 30	**20**	25	500	12.500	25
30 ├ 40	15	35	525	18.375	40
40 ├ 50	10	45	450	20.250	50
50 ├ 60	5	55	275	15.125	55
	Σ = 55		Σ = 1.825	Σ = 67.375	

Solução:

Os valores necessários para determinar as medidas que comporão o coeficiente de Pearson já se encontram calculados nas Tabelas construídas na solução do exemplo 1.

Procedimentos:

Para aplicar as fórmulas é necessário calcular média, moda, desvio-padrão, mediana.

Segunda Distribuição: Tabela B.1

Primeiro Passo: Calcular

Média: $\bar{x} = (\Sigma\, x_i f_i)/\Sigma\, f_i \rightarrow$

Moda: ?.

Mo = ?.

Desvio–padrão: $n = \Sigma f_i \rightarrow n = 55$ (calcular pela regra da variância)

$$\mathbf{S^2 = \frac{1}{(n-1)}\left[\Sigma x^2_i f_i - \frac{(\Sigma x_i f_i)^2}{n}\right]}$$

$S = \sqrt{S^2}$

S = ?.

Mediana: $\mathbf{Md = 1_{Md} + \frac{[(\sum f_i/2) - Fi\ (ant)]\ .\ h}{f_{Md}}}$

Md = ?.

Segundo Passo: 1º coeficiente de assimetria de pearson:

$$AS = \frac{\bar{x} - M_o}{S}$$

AS = ?.

Terceiro Passo: 2º coeficiente de assimetria de pearson:

$$A_S = \frac{3\ (\bar{x} - Med)}{S}$$

AS = ?

Terceira Distribuição: Tabela C.1

Primeiro Passo: Calcular

Média: $\mathbf{\bar{x} = (\Sigma x_i f_i)/\Sigma f_i \rightarrow \bar{x} =}$

Moda: ?.

Mo = ?.

Desvio–padrão: $n = \Sigma f_i \rightarrow n = 55$ (calcular pela regra da variância)

$\mathbf{S^2 = 1/(n-1)\ [\Sigma x^2_i f_i - (\Sigma x_i f_i)^2/n]}$

$S = \sqrt{S^2}$

S = ?.

Mediana: $\mathbf{Md = 1_{Md} + \frac{[(\sum f_i/2) - Fi\ (ant)]\ .\ h}{f_{Md}}}$

Md = ?.

Segundo Passo: 1º coeficiente de assimetria de Pearson:

$$AS = \frac{\bar{x} - M_o}{S}$$

$AS = ?$.

Terceiro Passo: 2º coeficiente de assimetria de Pearson:

$$A_S = \frac{3\,(\bar{x} - Med)}{S}$$

$As = ?$.

1.1.3 Coeficiente quartil de assimetria (e_Q) – Medida aplicada quando não é possível empregar o desvio padrão como medida de dispersão. Em seu cálculo recorre aos três quartis.

O coeficiente quartil de assimetria é definido pela seguinte relação:

$$\boxed{e_Q = \frac{(Q_3 - Md) - (Md - Q_1)}{(Q3 - Md) + (Md - Q_1)}} \quad \text{ou} \quad \boxed{e_Q = \frac{Q3 - 2Md + Q1}{Q3 - Q1}}$$

A qual revela que a assimetria é uma quantidade tomada como o quociente entre a diferença entre os afastamentos dos quartis e sua soma.

- O coeficiente quartil de assimetria assume valores entre os limites + 1 e – 1, ou seja: **$-1 \leq e_Q \leq 1$**

Exemplo 8.3

Calcular o coeficiente quartil de assimetria para as três distribuições do Exemplo 1, Tabelas A.1; B.1; C.1.

a) Tabela A.1

Classes	f_i	xi	$x_i f_i$	$x_i^2 f_i$	Fi	
10 ├ 20	5	15	75	1.125	5	
20 ├ 30	10	25	250	6.250	15	←Classe Q_1(contém o 13,75º elem)
30 ├ 40	15	35	525	18.375	30	
40 ├ 50	**20**	45	900	40.500	50	←Classe Q_3(contém o 41,25º elem)
50 ├ 60	5	55	275	15.125	55	
	Σ = 55		**Σ = 2.025**	**Σ = 81.375**		

Procedimentos:

Calcular os quartis e a mediana

Primeira distribuição: Tabela A.1

Primeiro Passo:

Calcular $\mathbf{Q_1 = \quad l_{Q1} + \dfrac{[(k\ \Sigma fi/4) - F_i\ (ant)]\ .\ h}{f_{Q1}}}$

Temos:

$k\Sigma f_i/4 = 55/4 = 13{,}75$

$k\Sigma f_i/4 = 13{,}75$ $\quad (k = 1)$

$F_i = 15$ (contém o 13,75º elemento)

Classe $Q_1 = 20 \vdash 30$

$\mathbf{l_{Q1} = 20 \qquad F_i\ (ant) = 5 \quad h = 10 \quad f_{Q1} = 10}$

$Q_1 = 20 + [(13{,}75 - 5)10]/10$

$Q_1 = 20 + 8{,}75$

$\mathbf{Q_1 = 28{,}75}$

Segundo Passo:

Calcular $\mathbf{Q_3 = \quad l_{Q3} + \dfrac{[(3\Sigma fi/4) - F_i\ (ant)]\ .\ h}{f_{Q3}}}$

Temos: $3\Sigma f_i/4 = 3(55/4) = 41{,}25$

$3\Sigma f_i/4 = 41{,}25$

$F_i = 50$ (contém o 41,25º elemento)

Classe $Q_3 = 40 \vdash 50$

$l_{Q3} = 40 \qquad F_i\ (ant) = 30 \quad h = 10 \quad f_{Q3} = 20$

$Q_3 = 40 + [(41{,}25 - 30)10]/20$

$Q_3 = 40 + 5{,}63$

$\mathbf{Q_3 = 45{,}625}$

Terceiro Passo:

Calcular Md = ?.

(A mediana já foi calculada no exemplo 8.2 anterior página 114).

Md = 38,33

Quarto Passo: Aplicar a fórmula:

$$e_Q = \boxed{\frac{Q_3 - 2Md + Q_1}{Q_3 - Q_1}}$$

$$e_Q = \frac{45{,}625 - 2(38{,}33) + 28{,}75}{45{,}625 - 28{,}75}$$

$$e_Q = \frac{45{,}625 - 76{,}66 + 28{,}75}{16{,}875}$$

$$e_Q = \frac{-2{,}285}{16{,}875}$$

$e_Q = -0{,}135 \rightarrow$ Assimetria Negativa

b) Tabela B.1

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$	Fi
10 ├ 20	5	15	75	1.125	5
20 ├ 30	10	25	250	6.250	15
30 ├ 40	15	35	525	18.375	30
40 ├ 50	10	45	450	20.250	40
50 ├ 60	5	55	275	15.125	45
	Σ = 45		**Σ = 1.575**	**Σ = 61.125**	

Procedimentos:

Calcular os quartis e a mediana

Primeira distribuição: Tabela B.1

Primeiro Passo:

Calcular $\mathbf{Q_1 = 1_{Q1} + \dfrac{[(k\Sigma fi/4) - F_i\,(ant)]\,.\,h}{f_{Q1}}}$

Temos: $k\Sigma f_i /4 = ?$

$l_{Q1} = ?$ F_i (ant) = ? h = ? $f_{Q1} = ?$

$Q_1 = ?$

Segundo Passo:

Calcular $\mathbf{Q_3 = 1_{Q3} + \dfrac{[(3\Sigma fi/4) - F_i\,(ant)]\,.\,h}{f_{Q3}}}$

Temos: $3\Sigma f_i/4 = ?$

$l_{Q3} = ?$ F_i (ant) = ? h = ? $f_{Q3} = ?$

$Q_3 = ?$

Terceiro Passo: Calcular Md= ?
(A mediana já foi calculada no exemplo 3 anterior).

Quarto Passo: Aplicar a fórmula:

$$e_Q = \frac{Q_3 - 2Md + Q_1}{Q_3 - Q_1}$$

$e_Q = ?$

c) Tabela C.1

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$	F_i
10 ├ 20	5	15	75	1.125	5
20 ├ 30	20	25	500	12.500	25
30 ├ 40	15	35	525	18.375	40
40 ├ 50	10	45	450	20.250	50
50 ├ 60	5	55	275	15.125	55
	Σ = 55		**Σ = 1.825**	**Σ = 67.375**	

Procedimentos:

Calcular os quartis e a mediana

Primeira distribuição: Tabela C.1

Primeiro Passo: Calcular

$$\mathbf{Q_1 = \quad 1_{Q1} + \frac{[(\Sigma fi/4) - F_i\,(ant)]\,.\,h}{f_{Q1}}}$$

Temos:

$k\Sigma f_i/4 = ?$

$l_{Q1} = ?$ F_i (ant) = ? $h = ?$ $f_{Q1} = ?$

$Q_1 = ?$

Segundo Passo: Calcular

$$\mathbf{Q_3 = \quad 1_{Q3} + \frac{[(3\Sigma fi/4) - F_i\,(ant)]\,.\,h}{f_{Q3}}}$$

Temos:

$3\Sigma f_i/4 =$

$l_{Q3} = ?$ F_i (ant) = ? $h = ?$ $f_{Q3} = ?$ $Q_3 = ?$

Terceiro Passo: Calcular Md =

(A mediana já foi calculada no exemplo 3 anterior)

Md = ?

Quarto Passo: Aplicar a fórmula:

$$e_Q = \frac{Q_3 - 2Md + Q_1}{Q_3 - Q_1}$$

$e_Q =$?

1.1.4 Coeficiente de assimetria entre os percentis 10 e 90 (e_p)

Obedece ao mesmo princípio que o coeficiente quartil de assimetria, mas neste caso, temos um quociente entre a diferença entre os afastamentos dos percentis em relação à mediana (quinquagésimo centil) e sua soma:

$$\boxed{e_p = \frac{(P_{90} - Md) - (Md - P_{10})}{(P_{90} - Md) + (Md - P_{10})}} \quad \text{ou} \quad \boxed{e_p = \frac{P_{90} - 2Md + P_{10}}{P_{90} - P_{10}}}$$

Exemplo 8.4

Calcular o coeficiente de assimetria entre os percentis 10 e 90 das três distribuições do Exemplo 8.4, Tabelas A.1; B.1; C.1.

Solução:

As medianas das três distribuições, já calculadas, são iguais a:

Primeira distribuição: Md = 38,33

Segunda distribuição: Md = 35

Terceira distribuição: Md = 31,67

Cálculo do Décimo Percentil:

$$\mathbf{Pi = 1_{pi} + \frac{[(k\Sigma fi/100) - F_i\,(ant)] \,.\, h}{f_{pi}}}$$

(Os cálculos serão baseados nas Tabelas A.1; B.1; C.1.

Temos:

Tabela A. 1

Classes	f_i	xi	$x_i f_i$	$x_i^2 f_i$	Fi
10 ├ 20	5	15	75	1.125	5
20 ├ 30	10	25	250	6.250	15←Classe P_{10}
30 ├ 40	15	35	525	18.375	30
40 ├ 50	**20**	45	900	40.500	50 ←Classe P_{90}
50 ├ 60	5	55	275	15.125	55
	Σ = 55		**Σ = 2.025**	**Σ = 81.375**	

Procedimentos:

Primeira distribuição: Tabela A.1

Primeiro Passo: Calcular a Mediana: Md = 38,33

(já calculada no exemplo anterior).

Segundo Passo: Calcular P_{10}

Cálculo dos termos: $k\Sigma f_i/100 = 10\ (55)/100 = 5{,}5$

$F_i = 15$ [que contém o (5,5)º elemento] →Classe P_{10} é a 2ª classe: 20 ├ 30

$l_{P10} = 20$ $h = 10$ $F_i\ (ant) = 5$ $f_{P10}=10$

Aplicar a fórmula

$$\mathbf{P_{10} = \quad l_{P10} + \frac{[(10\Sigma fi/100) - F_i\ (ant)]\ .\ h}{f_{P10}}}$$

$P_{10}= 20 + [(5{,}5 – 5)10]/10$

$P_{10}= 20+ 0{,}5$

$P_{10}= 20{,}5$

Terceiro Passo: Calcular P_{90}

Temos:

$k\Sigma f_i/100 = (90 \times 55)/100= 49{,}5$

$F_1 =50$ [que contém (49,5)º elemento] → Classe P_{90} é a 4ª classe 40 ├ 50

l_{P90} = 40 **h = 10** F_i (ant)= 30 f_{P90}= 20

Aplicar a formula:

$$\mathbf{P_{90} = \quad 1_{P90} + \frac{[(90\Sigma fi/100) - F_i\ (ant)]\ .\ h}{f_{P90}}}$$

P_{90}= 40 + [(49,5 – 30)10]/ 20

P_{90} = 40 + 9,75

P_{90}= 49,75

Quarto Passo: Cálculo de e_p

$$e_p = \frac{P_{90} - 2Md + P_{10}}{P_{90} - P_{10}}$$

$$e_p = \frac{49{,}75 - 2(38{,}33) + 20{,}5}{49{,}75 - 20{,}5}$$

e_p = (70,25 – 76,66)/29,25

e_p = -6,41/29,25

e_p = - 0,219 → Assimetria Negativa.

Calcular a segunda e terceira distribuições (Tabelas B.1 e C.1) deste exemplo:

Tabela B.1.

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$	Fi
10 ├ 20	5	15	75	1.125	5
20 ├ 30	10	25	250	6.250	15
30 ├ 40	**15**	35	525	18.375	30
40 ├ 50	10	45	450	20.250	40
50 ├ 60	5	55	275	15.125	45
	Σ = 45		**Σ = 1.575**	**Σ = 61.125**	

Procedimentos:

Segunda distribuição: Tabela B.1.

Primeiro Passo: Calcular a Mediana: Md = ?

(Já calculada no exemplo anterior).

Segundo Passo: Calcular P_{10}

Aplicar a fórmula**:**

$$\mathbf{P_{10} = \quad 1_{P10} + \frac{[(10\Sigma fi/100) - F_i\ (ant)]\ .\ h}{f_{P10}}}$$

P_{10}= ?

Terceiro Passo: Calcular P_{90}

Aplicar à formula:

$$\mathbf{P_{90} = 1_{P90} + \frac{[(90\Sigma fi/100) - F_i\ (ant)]\ .\ h}{f_{P90}}}$$

$P_{90} = ?$

Quarto Passo: Cálculo de e_p

$$e_p = \frac{P_{90} - 2Md + P_{10}}{P_{90} - P_{10}}$$

$e_p = ?$

Tabela C.1.

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$	F_i
10 ├ 20	5	15	75	1.125	5
20 ├ 30	**20**	25	500	12.500	25
30 ├ 40	15	35	525	18.375	40
40 ├ 50	10	45	450	20.250	50
50 ├ 60	5	55	275	15.125	55
	Σ = 55		**Σ = 1.825**	**Σ = 67.375**	

Procedimentos:

Terceira distribuição: Tabela C.1.

Primeiro Passo: Calcular a Mediana: Md = ?

(Já calculada no exemplo anterior).

Segundo Passo: Calcular P_{10}

Aplicar a fórmula**:**

$$\mathbf{P_{10} = 1_{P10} + \frac{[(10\Sigma fi/100) - F_i\ (ant)]\ .\ h}{f_{P10}}}$$

$P_{10} = ?$

Terceiro Passo: Calcular P_{90}

Aplicar à formula:

$$\mathbf{P_{90} = 1_{P90} + \frac{[(90\Sigma fi/100) - F_i\ (ant)]\ .\ h}{f_{P90}}}$$

$\mathbf{P_{90}} = ?$

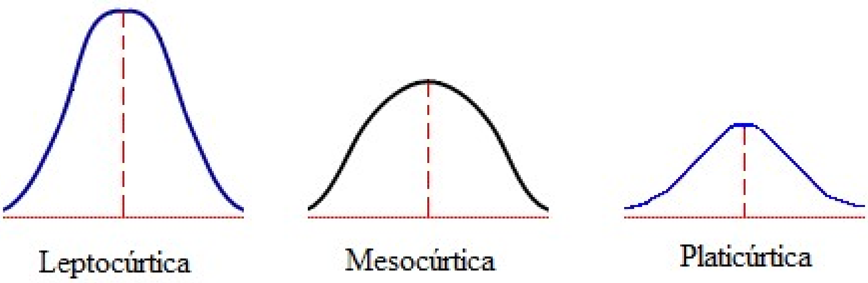
Quarto Passo: Cálculo de e_p

$$e_p = \frac{P_{90} - 2Md + P_{10}}{P_{90} - P_{10}}$$

$e_p = ?$

1.2 MEDIDAS DE CURTOSE – é o grau de achatamento de uma distribuição em relação a uma distribuição padrão denominada curva normal (curva correspondente a uma distribuição teórica de probabilidade).

- Quando a distribuição apresenta uma curva de frequência mais fechada que a normal (ou mais aguda em sua parte superior), recebe o nome de leptocúrtica
- Quando a distribuição apresenta uma curva de frequência mais aberta que a normal (ou mais achatada na sua parte superior), é chamada platicúrtica
- A curva normal, que é a nossa base referencial, recebe o nome de mesocúrtica.



8.2 – Medidas de curtose. Fonte: IBUJES (2019).

> **Nota:**
> O achatamento de cada uma dessas curvas refere-se de como os dados se apresentam em torno da média, podendo ser mais concentrados ou mais dispersos da média.

1.2.1 Coeficiente de curtose (C)– para avaliar o grau de curtose de uma curva ou distribuição de frequência:

$$\boxed{\mathbf{C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}}}$$

Onde:

C = Coeficiente de curtose

Q_1 = Primeiro quartil

Assim:

C = 0,263→ curva mesocúrtica

C < 0,263 →curva leptocúrtica

Q_3 = Terceiro quartil

P_{10} = Percentil 10

P_{90} = Percentil 90

C > 0,263→ curva platicúrtica

Observação:

Essa fórmula é conhecida como coeficiente percentílico de curtose.

Exemplo 8.5

Calcular o coeficiente percentílico de curtose das distribuições de frequências apresentadas no Exemplo 1 (de Assimetria) Tabela A.1.

Solução:

Tabela A. 1

Classes	f_i	Xi	$x_i f_i$	$x_i^2 f_i$	Fi	
10 ├ 20	5	15	75	1.125	5	
20 ├ 30	0	25	250	6.250	15	←classe Md
30 ├ 40	15	35	525	18.375	30	
40 ├ 50	**20**	45	900	40.500	50	←classe Md
50 ├ 60	5	55	275	15.12	55	
	Σ = 55		**Σ = 2.025**	**Σ = 81.375**		

Primeira Distribuição

Primeiro Passo: Calcular os quartis:

a) Primeiro Quartil: Q_1 :

$\Sigma f_i/4 \rightarrow 55/4 = 13{,}75 \rightarrow \Sigma f_i/4 = 13{,}75$ Fi = 15→ classe mediana

Q_1 = 2ª classe: 20 ├ 30 h = 30 – 20 = 10→ h = 10 F_i(ant)= 5

Aplicar a fórmula:

$$Q_1 = 1_{Q1} + \frac{[(\Sigma fi/4) - F_i\,(ant)]\,.\,h}{Fi}$$

$$Q_1 = 20 + \frac{(13{,}75) - 5)\,.\,h}{10}$$

Q_1 = 20+ 8,75

Q_1 = 28,75

b) Terceiro Quartil: Q_3

$3\ \Sigma\ f_i/4 \rightarrow 3(55)/4 = 41{,}25 \rightarrow 3\ \Sigma\ f_i/4 = 41{,}25$

$F_i = 50$ (classe mediana) $\quad Q_3 = 4^a$ classe: $40 \vdash 50$

Temos: $h^* = 50 - 40 = 10 \rightarrow h^* = 10 \quad f^* = 20 \quad F_i(ant) = 30$

$$\mathbf{Q_3 = 1^*_{Q1} + \frac{[(3\Sigma fi/4) - F\ (ant)]\ .\ h}{F^*}}$$

$$Q_3 = 40 + \frac{(41{,}25) - 30)\ .\ 10}{20}$$

$Q_3 = 40 + 5{,}625$

$Q_3 = 45{,}625$

$Q_3 = 45{,}63$

Segundo Passo: **Calcular os percentis**

a) Percentil: P_{10}

$k\Sigma f_i/100 \rightarrow K = 10$

$(10 \times 55)/100 = 5{,}5 \rightarrow k\Sigma\ f_i/100 = 5{,}5$

$F_i = 15 \rightarrow$ (classe mediana) $\rightarrow P_{10} = 2^a$ classe: $20 \vdash 30$

Temos: $F_1(ant) = 5 \quad h = 30 - 20 = 10 \quad f = 10 \quad k = 10$

Aplicar a fórmula:

$$\mathbf{P_{10} = 1_{P10} + \frac{[(10\Sigma fi/100) - F_i\ (ant)]\ .\ h}{f_{P10}}}$$

$$P_{10} = 20 + \frac{(5{,}5 - 5)\ .\ 10}{10}$$

$P_{10} = 20 + 0{,}5$

$P_{10} = 20{,}5$

b) Percentil: P_{90}

$k\Sigma f_{i\text{-}}/100 \rightarrow K = 90$

$(90 \times 55)/100 = 49{,}5 \rightarrow k\Sigma f_i/100 = 49{,}5$

$Fi = 50$ (classe mediana) $\rightarrow P_{90} = 4^a$ classe: $40 \vdash 50$

Temos: $F_i(ant) = 30 \quad h^* = 50 - 40 = 10 \quad f^* = 20 \quad K = 90$

Aplicar a fórmula:

$$P_{90} = 1_{P90} + \frac{[(90\Sigma fi/100) - F_i\,(ant)]\,.\,h}{f_{P90}}$$

$$P_{90} = 40 + \frac{(49,5 - 30)\,.\,10}{20}$$

$P_{90} = 40 + 9,75$

$P_{90} = 49,75$

Terceiro Passo: O cálculo do coeficiente é imediato:

Primeira Distribuição

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

$$C = \frac{45,625 - 28,75}{2(49,75 - 20,5)}$$

$C = 16,875/58,5$

C = 0,288

C > 0,263→ curva platicúrtica ou Distribuição Platicúrtica

Exemplo 8.6

Calcular o coeficiente percentílico de curtose das distribuições de frequências apresentadas no Exemplo 1 (de Assimetria): Tabela B.1 e Tabela C.1

- **Segunda distribuição**

Tabela B.1

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$	Fi
10 ├ 20	5	15	75	1.125	5
20 ├ 30	10	25	250	6.250	15
30 ├ 40	**15**	35	525	18.375	30
40 ├ 50	10	45	450	20.250	40
50 ├ 60	5	55	275	15.125	45
	Σ = 45		**Σ = 1.575**	**Σ = 61.125**	

- **Terceira Distribuição**

Tabela C.1

Classes	f_i	x_i	$x_i f_i$	$x_i^2 f_i$	F_i
10 ├ 20	5	15	75	1.125	5
20 ├ 30	**20**	25	500	12.500	25
30 ├ 40	15	35	525	18.375	40
40 ├ 50	10	45	450	20.250	50
50 ├ 60	5	55	275	15.125	55
	Σ = 55		**Σ = 1.825**	**Σ = 67.375**	

Procedimentos: (siga como modelo o Exemplo 8.5. Tabela A.1).

CAPÍTULO IX - DISTRIBUIÇÃO DE INTERVALOS DE CONFIANÇA

1. ESTIMATIVAS POR PONTO E INTERVALOS DE CONFIANÇA

O objetivo da Estatística é o de conhecer populações por meio das informações amostrais. Como as populações são caracterizadas por medidas numéricas descritivas, denominadas parâmetros, a estatística diz respeito à realização de inferências sobre esses parâmetros populacionais desconhecidos.

a) Parâmetros populacionais típicos - a média, o desvio-padrão e a proporção - de determinado evento populacional.
b) Os métodos para realizar inferências a respeito dos parâmetros pertencem a duas categorias:

- Estimação: determinação de estimativas dos parâmetros populacionais.
- Testes de Hipóteses: tomada de decisão relativa ao valor de um parâmetro populacional.

1.1 ESTIMATIVA POR PONTO – Quando, com base em dados amostrais, calculamos um valor da estimativa de um parâmetro populacional. Assim:

- O valor da média amostral ($\overline{x}$) é uma estimativa por ponto da média populacional (μ).
- O valor do desvio-padrão amostral (S) constitui uma estimativa do parâmetro (σ).

Exemplo 9.1

Uma amostra aleatória de 200 alunos de uma universidade de 20.000 estudantes revelou nota média amostral de 5,2. Logo, $\overline{x}$ = 5,2 é uma estimativa pontual da verdadeira nota média dos 20.000 alunos (MARTINS, 2010).

1.2 ESTIMATIVA POR INTERVALO – Uma estimativa por intervalo para um parâmetro populacional é um intervalo determinado por dois números, obtidos a partir dos elementos amostrais, que contenham o valor do parâmetro com dado nível de confiança ou probabilidade de (1 - α) %. Geralmente, (1 – α) % = 90%; 95%; 97,5%, ... (MARTINS, 2010).

Se o comprimento do intervalo é pequeno, temos um elevado grau de precisão da inferência realizada. As estimativas dessa natureza são denominadas intervalos de confiança.

Exemplo 9.2

Estimativas por Intervalo (MARTINS, 2010):

a) O intervalo [1,60 m; 1,64 m] contém a altura média dos moradores do município X, com nível de confiança de 95%;
b) Com 97,5% de confiança, o intervalo [8%; 10%] contém a proporção de analfabetos da cidade Y;
c) O intervalo [37 mm; 39 mm] contém o desvio padrão do comprimento de uma peça, com 90% de confiança.

Nota:

Quando se constrói um intervalo de confiança, é importante atentar para o risco do erro.

- Se o nível de confiança é de 95%, o risco do erro da inferência será de 5%. Assim: se construíssemos 100 intervalos, baseados em 100 amostras de tamanhos iguais, poderíamos esperar que 95 desses intervalos (95% deles) iriam conter o parâmetro populacional sob estudo, enquanto cinco intervalos (5% deles) não iriam conter o parâmetro.
- Uma configuração reforçará o conceito da estimativa por intervalo. Seja θ (lê-se teta) um parâmetro populacional. Vamos admitir a seleção de 10 amostras de mesmo tamanho e um nível de confiança de 90% (Figura 9.1) expõe os intervalos obtidos.

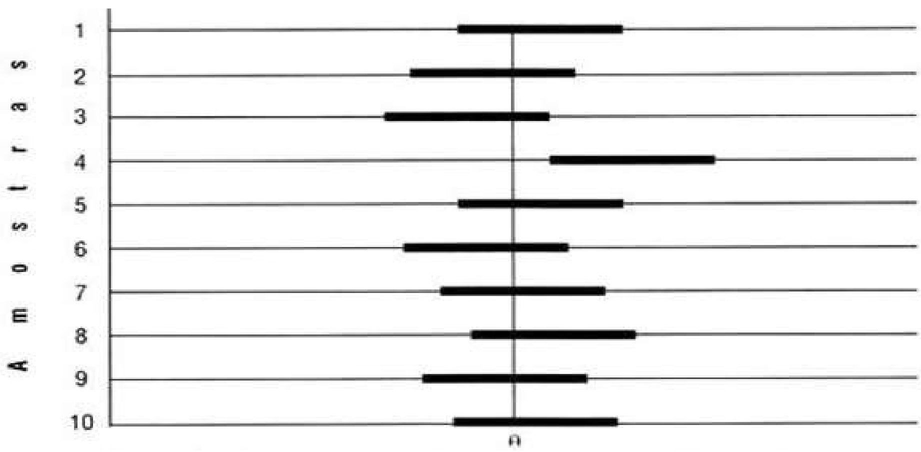


Figura 9.1 – Configuração: 10 intervalos de confiança para θ a partir de 10 amostras de mesmo tamanho e $(1 - \alpha)\% = 90\%$.Estimativa por intervalo. Fonte: MARTINS (2010).

- Os segmentos horizontais *representam os 10 intervalos*, e a reta vertical representa *a localização do parâmetro* θ. Nota-se que o parâmetro é fixo e que a localização do intervalo varia de amostra para amostra. Assim, podemos falar em termos da “probabilidade de o intervalo incluir θ”, e não em termos da “probabilidade de θ pertencer ao intervalo”, já que θ é fixo. O intervalo é aleatório. Na prática, somente um intervalo é construído por meio da amostra aleatória obtida.

2. INTERVALO DE CONFIANÇA PARA A MÉDIA POPULACIONAL QUANDO A VARIÂNCIA É CONHECIDA

I. O estimador da média populacional (μ) **é** $\overline{x}$,

II. O estimador do desvio padrão (**σ**)

III. O estimador da proporção (p)

Figura 9.2 - Paramentros versus Estimadores. Fonte: LOPES (2003).

- A distribuição de probabilidade de $\overline{x}$ é dada por: $\overline{x}$ d N (μ; σ^2/n) e

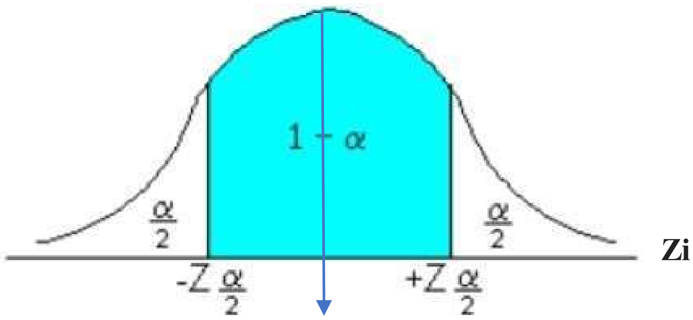
Para populações finitas:

$\bar{x}$ d N [μ; (σ^2/n) ((N – n)/(N - 1))]

Para populações infinitas, a variável normal padronizada de $\bar{x}$ será:

$$Z_i = \frac{(\bar{x} - \mu)}{\frac{\Sigma}{\sqrt{n}}}$$

Fixando um nível de confiança (1- α), temos:



0

Figura 9.3 – Intervalo de confiança para média populacional quando a variância é conhecida. Fonte: LOPES (2003).

Ou seja**:** $\mathbf{P\ (-Z_{\alpha/2} \leq Z_i \leq Z_{\alpha/2}) = 1 - \alpha}$

Substituindo o valor de $\mathbf{Z_i}$**:**

$$\mathbf{P\ -Z_{\alpha/2} \leq \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}} \leq Z_{\alpha/2}) = 1 - \alpha}$$

Resolvendo as duas inequações para **μ**, temos o intervalo de confiança para a média populacional (**μ**), quando a variância (**σ^2**) **é** conhecida:

$$\mathbf{P\ [\bar{x} - Z_{\alpha/2}(\sigma / \sqrt{n}) \leq \mu \leq \bar{x} + Z_{\alpha/2}(\sigma / \sqrt{n})] = 1 - \alpha}$$

Como pode ser verificado, a aplicação da fórmula é extremamente simples.

Primeiro Passo: Fixa-se o valor de (1 – α), ou (1 – α)100 = %

Segundo Passo: Observa-se na Tabela de distribuição normal padrão o valor das abscissas que deixam **a/2** em cada uma das caudas.

Terceiro Passo: Com os valores de $\overline{x}$ (média amostral), **σ** (o desvio padrão da população) e **n** (tamanho da amostra) constrói-se o intervalo.

Exemplo 9.3 (MARTINS, 2010)

A duração da vida de uma peça de equipamento é tal que **σ** = 5horas. Foram amostradas aleatoriamente 100 dessas peças, obtendo-se média de 500 horas. Desejamos construir um intervalo de confiança para a verdadeira duração média da peça com um **nível de 95% de confiança**.

Solução:

Temos: $\boldsymbol{\sigma = 5}$ $\quad$ **n= 100** $\quad$ $\boldsymbol{\overline{x} = 500}$ $(1 - \alpha)100 = 95\%$

O gráfico da distribuição normal padrão será:

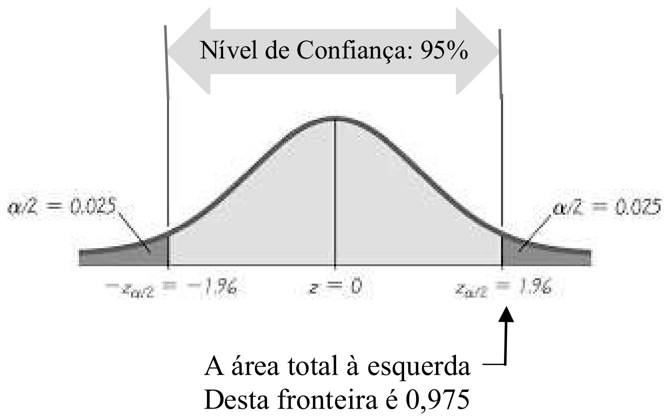


Figura 9.4 – Gráfico da distribuição normal padrão. Fonte: AMARAL (2012).

Para descobrir as abscissas $-Z_{a/2} = -1{,}96$ **e** $\boldsymbol{Z_{a/2} = 1{,}96}$ **(**buscamos na Tabela da distribuição normal padrão **(**Anexo C) com **0, 475**, já que a Tabela é de faixa central.

- Substituindo os dados na fórmula:

$$\mathbf{P\,[\,\overline{x} - Z_{a/2}\,(\,\sigma/\sqrt{n}) \le \mu \le \overline{x} + Z_{a/2}\,(\sigma/\sqrt{n})] = 1 - \alpha}$$

$$P\,[500 - 1{,}96\,(5/\sqrt{100}) \le \mu \le 500 + 1{,}96\,(5/\sqrt{100})] = 95\%$$

$$P\,[500 - 1{,}96\,(5/10) \le \mu \le 500 + 1{,}96\,(5/10)] = 95\%$$

$$P\,[500 - 0{,}98 \le \mu \le 500 + 0{,}98] = 95\%$$

$P\ (499{,}02 \leq \mu \leq 500{,}98) = 95\%$

O intervalo de confiança solicitado: **P (499,02 ≤ μ ≤ 500,98) = 95%**

- A interpretação desse resultado pode ser expressa por uma das duas maneiras a seguir:

a) "O intervalo [**499,02; 500,98**] contém a duração média da peça com 95% de confiança"; ou

b) "com 95% de confiança o intervalo **[499,02; 500,98]** contém a duração média da peça".

Isso significa que, se forem construídos intervalos dessa maneira, para um grande número de amostras, em 95% dos casos os intervalos incluiriam **μ.**

c) Para os casos de populações finitas, usa-se a seguinte fórmula:

$$P\left[\bar{x} - Z_{a/2}(\sigma/\sqrt{n})\ \sqrt{(N-n)/(N-1)} \leq \mu \leq \bar{x} + Z_{a/2}(\sigma/\sqrt{n})\ \sqrt{(N-n)/(N-1)}\right] = 1 - \alpha$$

Exemplo 9.4 (MARTINS, 2010)

Vamos admitir os mesmos dados do exemplo anterior, considerando como população a produção de 1.000 peças. Nesse caso o intervalo da média será:

$$\boldsymbol{\sigma = 5 \quad n = 100 \quad \bar{x} = 500} \quad (1-\alpha)100 = 95\% \quad N = 1.000$$

Aplicando a fórmula:

$P\left[\bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n})\ \sqrt{(N-n)/(N-1)} \leq \mu \leq \bar{x} + Z_{\alpha/2}(\sigma/\sqrt{n})\ \sqrt{(N-n)/(N-1)}\right] = 1 - \alpha$

$P\ [500 - 1{,}96\ (5/\sqrt{100})\ (\sqrt{1000-100/1000-1}) \leq \mu \leq 500 + 1{,}96\ (5/\sqrt{100})\ (\sqrt{1000-100)/1000-1})] = 95\%$

$P\ [500 - 1{,}96\ (5/10)\ (\sqrt{900/999}) \leq \mu \leq 500 + 1{,}96\ (5/10)\ (\sqrt{900/999})] = 95\%$

$P\ [500 - 1{,}96\ (0{,}5\ \sqrt{0{,}900}) \leq \mu \leq 500 + 1{,}96\ (0{,}5)\ (\sqrt{0{,}900}) = 95\%$

$P\ [500 - 1{,}96\ (0{,}5\ .\ 0{,}949) \leq \mu \leq 500 + 1{,}96\ (0{,}5\ .\ 0{,}949)] = 95\%$

$P\ [500 - 0{,}93 \leq \mu \leq 500 + 0{,}93] = 95\%$

P [499,07 ≤ μ ≤ + 500,93] = 95%

Logo, o intervalo **[499,07; 500,93]** contém a duração média das 1.000 peças com 95% de confiança.

3. INTERVALO DE CONFIANÇA (IC) PARA A MÉDIA POPULACIONAL QUANDO A VARIANCIA É DESCONHECIDA

Quando temos pequenas amostras e o valor do desvio padrão populacional é desconhecido, podemos construir intervalos de confiança para a média (fórmula a seguir) sendo necessário que a população de onde foi extraída a amostra aleatória tenha distribuição normal.

Como o σ (desvio padrão populacional) é desconhecido, é preciso substituí-lo por S (desvio padrão amostral) que, contrariamente ao σ (desvio padrão populacional), é uma variável aleatória Daí se ter o quociente entre duas variáveis aleatórias, $\bar{x}$ e S. O estimador:

$\dfrac{(\bar{x} - \mu)}{\frac{S}{\sqrt{n}}}$ tem distribuição **t de Student** (Anexo E) com (n-1) graus de liberdade.

Fixando um nível de confiança (1-α), temos:

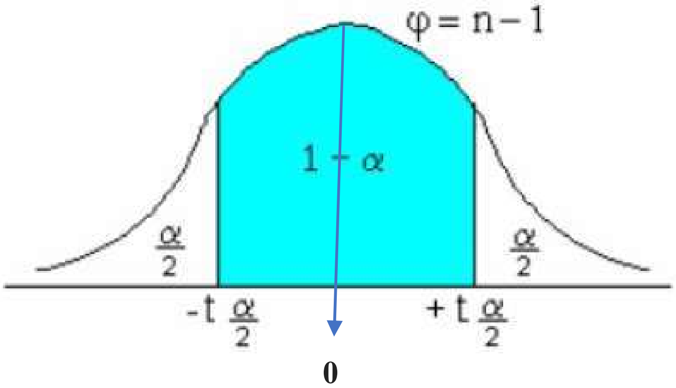


Figura 9.5 – Intervalo de confiança para média populacional quando a variância é desconhecida. Fonte: LOPES (2003).

Ou seja: $\mathbf{P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = 1 - \alpha}$

Substituindo o valor de t, e resolvendo as inequações para **μ**, obtemos o intervalo para a média, quando a variância (σ^2) é desconhecida:

$$\mathbf{P[\bar{x} - t\ \alpha/2\ (S/\sqrt{n}) \leq \mu \leq \bar{X} + t\ \alpha/2\ (S/\sqrt{n})] = 1 - \alpha}$$

Onde a variável **t** possui **(n – 1)** graus de liberdade.

Exemplo 9.5

A amostra **9, 8, 12, 7, 9, 6, 11, 6, 10, 9** foi extraída de uma população normal. Construa um intervalo de confiança para a média ao nível de 95%.

Solução: Calculando a média e o desvio padrão da amostra, obtemos

Primeiro Passo: Calculando a média da amostra: $\bar{x}$

$$\bar{x} = \frac{(9+8+12+7+9+6+11+6+10+9)}{10} = 8,7$$

Segundo Passo: Calculando o desvio padrão da amostra

Fórmula para o desvio padrão:

$$\boxed{S = \left[\frac{\sqrt{\Sigma(x^2_i)}}{N} - \frac{(\Sigma x_i)^2}{N} \right]}$$

Temos: = Σx_i = 9+8+12+7+9+6+11+6+10+9=87 n = 10

$\Sigma(x^2_i)$ = 81 + 64+ 144+ 49+ 81+ 36+121+36+100+81= 793

Aplicando a fórmula**:**

$$\boxed{S = \left[\frac{\sqrt{\Sigma(x^2_i)}}{n} - \frac{(\Sigma x_i)^2}{N} \right]}$$

$S = [\sqrt{(793/10 - (87/10)^2}\,]$

$S = \sqrt{79,3 - 75,7}$

$S = \sqrt{3,6}$

$S = 1,897 \approx 2 \rightarrow S = 2$

Assim: $\bar{x} = 8,7$ e $S = 2$

Como $((1 - \alpha)\% = 95\%$, e gl = 9, pois gl = φ = (n-1) →gl = 10 – 1 = 9→ gl = 9

Terceiro Passo: Consultar na Tabela de t de Student (Anexo E):

Temos: $-t_{\alpha/2} = -2,2622$ e $t_{\alpha/2} = 2,2622$

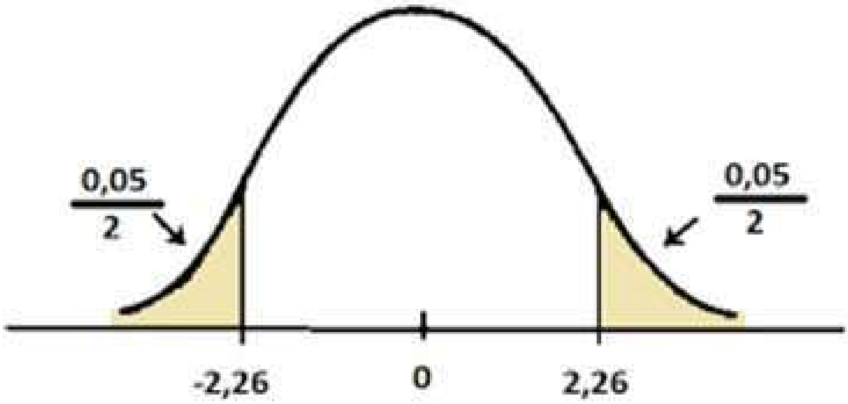


Figura 9.6 – Teste t de Student. Fonte: ALVES (2017).

Quarto passo: Substituir os valores na fórmula,

$$\mathbf{P\ [\ \overline{x}\ -\ t_{\alpha/2}\ (S/\sqrt{n}) \leq \mu \leq \overline{x}\ +\ t_{\alpha/2}\ (S/\sqrt{n})] = 1 - \alpha}$$

$$P\ [8{,}7 - 2{,}2622\ (2/\sqrt{10}) \leq \mu \leq 8{,}7 + 2{,}2622\ (2/\sqrt{10})] = 95\%$$

$$P\ [8{,}7 - 2{,}2622\ (2/3{,}16) \leq \mu \leq 8{,}7 + 2{,}2622\ (2/3{,}16)] = 95\%$$

$$P\ [8{,}7\ -2{,}2622(0{,}63) \leq \mu \leq 8{,}7 + 2{,}2622\ (0{,}63)] = 95\%$$

$$P\ [8{,}7 - 1{,}425 \leq \mu \leq 8{,}7 + 1{,}425] = 95\%$$

$$P\ [7{,}27 \leq \mu \leq 10,\ 13] = 95\%$$

O intervalo [7,27; 10,13] contém a média com 95% de confiança.

4. INTERVALO DE CONFIANÇA PARA A VARIÂNCIA

- O estimador de σ^2 é S^2.
- S^2 tem distribuição Qui-Quadrado (Anexo D), a exceção de constantes. Isto é:

  $[(n-1)\ S^2]/\sigma^2$ tem distribuição Qui-Quadrado com (n-1) graus de liberdade.

Ou seja:

$$\boldsymbol{\chi_{n}\ -\ 1^2\ d\ \ \frac{(n-1)\ S^2}{\sigma^2}}$$

Admitindo-se que a população de onde foi extraída a amostra tenha distribuição normal. Então o intervalo poderá ser:

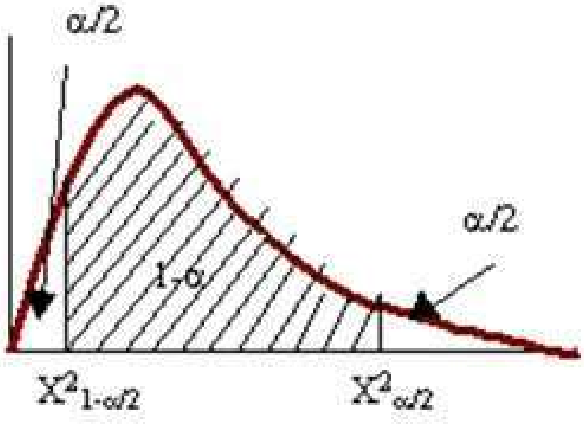


Figura 9.7 – Intervalo de confiança para a variância. Fonte: ITC (2019).

Observação:

$X^2_{1-\alpha/2} \rightarrow X^2$ inf e $X^2_{\alpha/2} \rightarrow X^2$ sup, na formula a seguir.

$$P\,(\chi^2_{inf} \le \chi^2_{n-1} \le \chi^2_{sup}) = 1\text{-}\alpha$$

Substituindo o valor de χ^2_{n-1} e resolvendo as duas inequações para o σ^2, obtemos o intervalo:

$$\boxed{P\left[\frac{(n-1)\,S^2}{\chi^2_{sup}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{inf}}\right] = 1 - \alpha}$$

Exemplo 9.6 (MARTINS, 2010)

Vamos admitir uma distribuição normal, n=10, $S^2 = 4$, e que desejamos construir um IC para a variância populacional ao nível de 90%.

Solução

Temos: n = 10, $S^2 = 4$, $\varphi = (n-1) = (10 - 1) = 9$ e $(1 - \alpha) = 90\%$

Consultando a Tabela da distribuição Qui-Quadrado (Anexo D), temos:

$\chi^2_{sup} = 16{,}9$ e $\chi^2_{inf} = 3{,}33$

Substituindo os valores na fórmula:

$$P\left[\frac{(n-1)\,S^2}{\chi^2_{sup}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{inf}}\right] = 1 - \alpha$$

$P\,[9(4)/16{,}9 \le \boldsymbol{\sigma^2} \le 9(4)/3{,}33] = 90\%$

$P\,[36/16{,}9 \le \boldsymbol{\sigma^2} \le 36/3{,}33] = 90\%$

$P\ [2{,}13 \leq \boldsymbol{\sigma^2} \leq 10{,}81] = 90\%$

Assim, podemos afirmar que o intervalo [2,13; 10,81] contém a variância populacional com 90% de confiança.

5. INTERVALO DE CONFIANÇA PARA O DESVIO PADRÃO

Admitindo que a distribuição de probabilidade populacional de onde se extraiu a amostra seja normal, um intervalo de confiança aproximado[16] para o desvio padrão ($\boldsymbol{\sigma}$) é dado pela raiz quadrada do IC da variância (σ^2).

$$\mathbf{P} \left[\frac{\sqrt{S^2}\ (n-1)}{\chi^2_{sup}} \leq \boldsymbol{\sigma} \leq \frac{\sqrt{S^2}\ (n-1)}{\chi^2_{inf}} \right] = \mathbf{1 - \alpha}$$

Onde a distribuição χ^2 é tomada com $\varphi = n-1$ graus de liberdade.

Exemplo 9.7

Com os dados do exemplo anterior, construa o IC para o desvio padrão populacional. Solução do exemplo anterior:

$2{,}13 \leq \boldsymbol{\sigma^2} \leq 10{,}81$; logo, um **IC** aproximado para $\boldsymbol{\sigma}$ será:

$\mathbf{P\ (\sqrt{2{,}13} \leq \sigma \leq \sqrt{10{,}81}) = 90\%}$ **ou,** $\mathbf{P\ (1{,}46 \leq \sigma \leq 3{,}29) = 90\%}$

O intervalo **[1,46; 3,29]** contém o desvio padrão populacional com 90% de confiança.

[16] Trata-se de uma aproximação, pois estaremos admitindo que a raiz quadrada de uma distribuição Qui-Quadrado também é Qui-Quadrado. (MARTINS, 2010, p.178)

CAPÍTULO X - TESTES DE HIPÓTESES

1. TESTES DE HIPÓTESES – método para fazer inferências sobre parâmetros populacionais. Em vez de calcular uma estimativa do parâmetro pontual ou do intervalo (exposto na Capítulo IX), iremos admitir um valor hipotético para um parâmetro populacional, e com base nas informações da amostra realizaremos o teste estatístico, para aceitar ou rejeitar o valor hipotético, ficando evidente que a decisão estará sujeita a erros.

2. PRINCIPAIS CONCEITOS

2.1 HIPÓTESE ESTATÍSTICA: é uma suposição quanto ao valor de um parâmetro populacional, ou quanto à natureza da distribuição de probabilidade de uma variável populacional.

São exemplos de hipóteses estatísticas:

a) A altura **média** da população brasileira é de 1,65 m, isto é: **H**: $\mu = 1{,}65$ **m.**

b) A **variância** populacional dos salários vale $\$500^2$, isto é: **H**: $\sigma^2 = 500^2$.

c) A proporção de piauienses com a doença X é de 40%, ou seja: **H**: $p = 0{,}40$

d) A distribuição de probabilidades dos pesos dos alunos de uma faculdade A é normal.

As hipóteses estatísticas são formuladas pelo pesquisador com base em suas conjecturas sobre o fenômeno, ou em função de informações teóricas.

2.2 TESTE DE HIPÓTESE – é uma regra de decisão para aceitar, ou rejeitar, uma hipótese estatística com base nos elementos amostrais. São dois os tipos de hipóteses:

1) H_0, chamada hipótese nula, a hipótese estatística a ser testada, expressa uma igualdade, tal como $\leq$ = ou $\geq$.

2) H_1, a hipótese alternativa. É geralmente o complemento da hipótese nula. É a afirmação que deve ser verdadeira se H_0 for falsa e contém uma afirmação de desigualdade, tal como < $\neq$ ou >.

Por exemplo, se o valor da afirmação for k e o parâmetro populacional for μ, as possíveis hipóteses a serem formuladas são:

I	II	III
I → H_0: $\mu = k$	II → H_0: $\mu = k$	III → H_0: $\mu = k$
H_1: $\mu \neq k$	H_1: $\mu > k$	H_1: $\mu < k$

IV → H_0: $\mathbf{\mu \leq k}$ V → H_0: $\mathbf{\mu \geq k}$

H_1: $\mathbf{\mu > k}$ H_1: $\mathbf{\mu < k}$

- A situação I corresponde a um teste bilateral ou bicaudal.
- As situações II, III, IV e V correspondem a testes unilaterais ou unicaudal.
- Em algumas situações, podemos formular as hipóteses da forma: I ou II, ambas corretas, ou, ainda, I ou III, também corretas
- Em situações críticas podemos obter conclusões diferentes.

Exemplos de Hipóteses para um Teste Estatístico:

a) H_0: $\mathbf{\mu = 1{,}65\ m}$

H_1: $\mathbf{\mu \neq 1{,}65\ m}$ ou $\mathbf{\mu > 1{,}65\ m}$ **ou** $\mathbf{\mu < 1{,}65\ m}$

b) H_0: $\mathbf{\sigma^2 = 500^2}$.

H_1: $\mathbf{\sigma^2 \neq 500^2}$. ou $\mathbf{\sigma^2 > 500^2}$ **ou** $\mathbf{\sigma^2 < 500^2}$

c) H_0: $\mathbf{p = 0{,}40}$

H_1: $\mathbf{p \neq 0{,}40}$ **ou** $\mathbf{p > 0{,}40}$ **ou** $\mathbf{p < 0{,}40}$

2.3 TIPOS DE ERROS E NÍVEL DE SIGNIFICÂNCIA- Há dois possíveis tipos de erros, quando realizamos um teste estatístico para aceitar ou rejeitar H_0.

a) Rejeitar $\mathbf{H_0}$**,** quando ela é verdadeira, e comete-se o Erro tipo I, e essa probabilidade de se cometer o Erro tipo I é designada $\boldsymbol{\alpha}$.

b) Aceitar $\mathbf{H_0}$, quando ela é falsa, comete-se o Erro tipo II, e a probabilidade de cometer o Erro tipo II é designada $\boldsymbol{\beta}$.

Quadro 10.1. Possíveis erros e acertos de uma decisão a partir de um teste de hipótese estão sintetizados no quadro a seguir:

		Realidade	
		H_0 verdadeira	**H_0 falsa**
Decisão	Aceitar H_0	Decisão correta (1-α)	Erro tipo II β
Decisão	Rejeitar H_0	Erro tipo I (α)	Decisão correta (1-β)

Notas:

1) O erro tipo I só pode ser cometido quando se rejeitar H_0, enquanto o erro tipo II, quando se aceitar H_0.
2) A redução simultânea dos erros poderá ser alcançada pelo *aumento do tamanho da amostra*, evidentemente, *com aumento dos custos.*
3) Para um mesmo tamanho de amostra, a probabilidade de incorrer em um *Erro tipo II* aumenta à medida que diminui a probabilidade do *Erro tipo I,* e vice-versa.

- O nível de significância de um teste é definido como sendo a probabilidade máxima permissível para cometer um Erro Tipo I. Ou seja, o nível de significância é igual ao valor α.
- A probabilidade de um Erro tipo II é o valor β. O valor $1-\beta$ é chamado de poder do teste. Ele representa a probabilidade de rejeitar a hipóteses nula quando a hipótese alternativa for verdadeira. O valor do poder é difícil (e às vezes impossível) de se encontrar na maioria dos casos.
- Se o valor $(1-\beta)$, tiver poder = 90% → $(1-\beta) = 0,9$. Assim, $\beta = 0,1$ que corresponde a 10%.

2.3.1 Configuração sobre o mecanismo dos erros

Exemplo 10.1

Desejamos testar $\mathbf{H_0: \mu = 20}$ contra $\mathbf{H_1: \mu > 20}$. Sabemos que a variância da população vale 16 e que foi retirada de uma amostra de 16 elementos. Ou seja:

$$\mathbf{H_0: \mu = 20 \quad \sigma^2 = 16 \quad n=16}$$
$$\mathbf{H_1: \mu > 20}$$

Como $\bar{x}$, estimulador de $\boldsymbol{\mu}$, tem distribuição normal (Capítulo IX) temos graficamente:

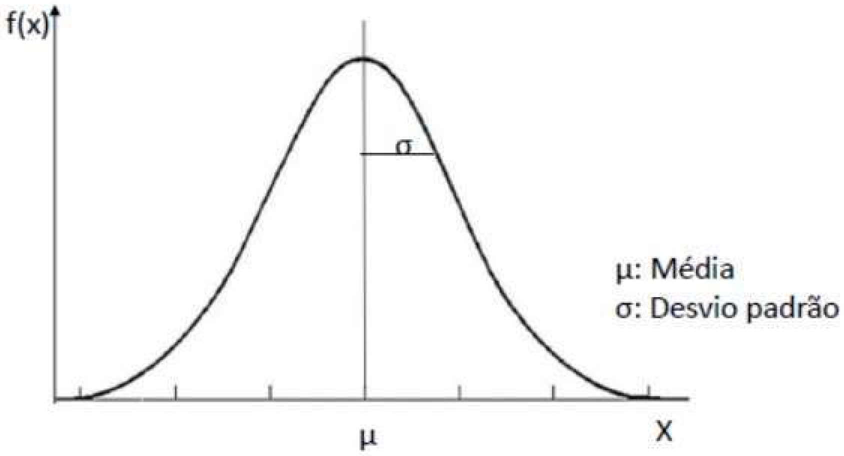


Figura 10.1 - Distribuição da média amostral para testar $\mathbf{H_0}$.

Distribuição da média amostral para testar $\mathbf{H_0}$: **μ = 20**.

Para valores de $\overline{x}$ (média amostral) próximos de 20 a hipótese $\mathbf{H_0}$ **(**agá-zero) poderá ser aceita. Como $\mathbf{H_1}$**: μ > 20**, devemos ter um limite crítico à direita **(** $\overline{x}_c$**)** para valores de $\overline{x}$ Assim:

- Região de rejeição para $\mathbf{H_0}$: **μ = 20 (**à direita**)**
- Região de aceitação para $\mathbf{H_0}$: **μ = 20 (**à esquerda**)**

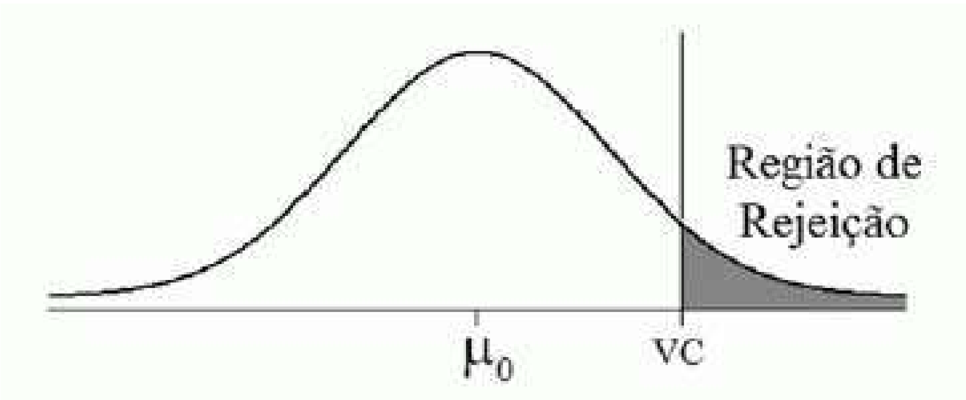


Figura 10.2 – Distribuição da média amostral para testar H_0. Fonte: FGV (2018).

Observar o gráfico:

- A área hachurada à direita, corresponde à probabilidade de **rejeitar** $\mathbf{H_0}$, quando $\mathbf{H_0}$: **μ = 20** é verdadeira. Essa área representa a probabilidade de cometer o **Erro tipo I**, ou seja, a área representa o risco **α**

Limite crítico: $\bar{x}_c$

- Para encontrar um **limite crítico** $\bar{x}_c$, atribui-se valores para **α.** Admitir que **α = 0,05 → α = 5%;** é preciso padronizar a distribuição normal das médias:

$$\boxed{\mathbf{Z^d = \frac{\bar{x}_c - \mu}{\sigma/\sqrt{n}} \approx N\,(0,1)}}$$

Ou seja:

$$\mathbf{Z = \frac{\bar{x}_c - \mu}{\sigma/\sqrt{n}}}$$

Temos: H_0: $\mu = 20$ $\sigma^2 = 16 \rightarrow \sigma = 4$ n=16

H_1: $\mu > 20$

Z (padronizado) = 1,64 (Se: 20/16 =1,25)

Então:

$$\mathbf{1{,}64 = \frac{\bar{x}_c - 20}{4/\sqrt{16}}}$$

$$1{,}64 = \frac{\bar{x}_c - 20}{4/4}$$

$$1{,}64 = \frac{\bar{x}_c - 20}{1}$$

$$1{,}64 = \bar{x}_c - 20$$

$$\bar{x}_c = 20 + 1{,}64$$

$$\bar{x}_c = 21{,}64$$

Logo, $\bar{x}_c = 21{,}64$

Assim, a regra de decisão para **H_0** será:

Rejeitar H_0, quando $\bar{x} > 21{,}64$

Aceitar H_0, quando $\bar{x} \leq 21{,}64$

2.3.2 Decisão de aceitar ou rejeitar h_0 - mecanismos dos erros:

Pelos gráficos observa-se, que há

a) Grande probabilidade de aceitar $\mathbf{H_0}$: μ **= 20- 95% -** e pouca probabilidade – apenas **5% -** de rejeitar $\mathbf{H_0}$: μ **= 20**.
b) Quando se aceita uma hipótese $\mathbf{H_0}$, quando $\mathbf{H_0}$ é falsa, comete-se um Erro tipo II.

Neste exemplo, essa probabilidade de Erro pode ser de 95% quando $\mathbf{H_0}$: μ **= 20** for falsa (fato preocupante).

c) Há apenas 5% de chances de rejeitar $\mathbf{H_0}$.
d) Quando se rejeita $\mathbf{H_0}$, a hipótese sendo verdadeira, comete-se um Erro tipo I.

Como a probabilidade neste caso é relativamente baixa - 5% -, a decisão de rejeitar H_0, é muito mais segura do que a decisão de aceitar H_0. Esta é a lógica do teste de significância.

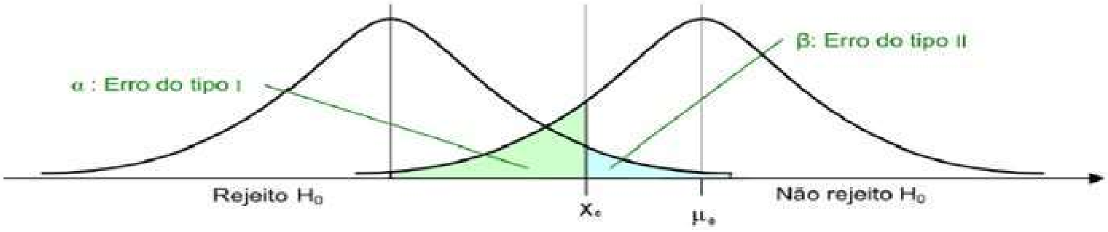
- Atribuem-se baixos valores para α, geralmente de 1% a 10%.
- Formula-se $\mathbf{H_0}$, com a pretensão de rejeitá-la, daí o nome de hipótese nula.
- Se o teste indicar a rejeição de H0, há um indicador mais seguro para a decisão.
- Caso o teste indique aceitação de $\mathbf{H_0}$, diz-se que, com o nível de significância **α,** não se pode rejeitar $\mathbf{H_0}$, e nestes casos a decisão não é tão segura quanto **a** decisão de rejeitar $\mathbf{H_0}$.

2.3.3 Determinar a probabilidade β de cometer o erro tipo II.

Para o cálculo de β (probabilidade de aceitar $\mathbf{H_0}$, quando $\mathbf{H_0}$, é falsa), é preciso admitir outros valores para $\mathbf{H_0}$.

a) No exemplo, $\mathbf{H_0}$: μ **= 20** seria falso, ou em realidade μ **> 20**. Isso corresponde a uma infinidade de possíveis valores para **μ,** por exemplo**: 20,2; 20,5; 21, 22**. Para cada um desses valores, podemos determinar o valor β condicionado à hipótese admitida.

Assim, para um valor qualquer, μ_1 > 20, temos a seguinte configuração para o Erro tipo II.



↓ ↓
20 21,64

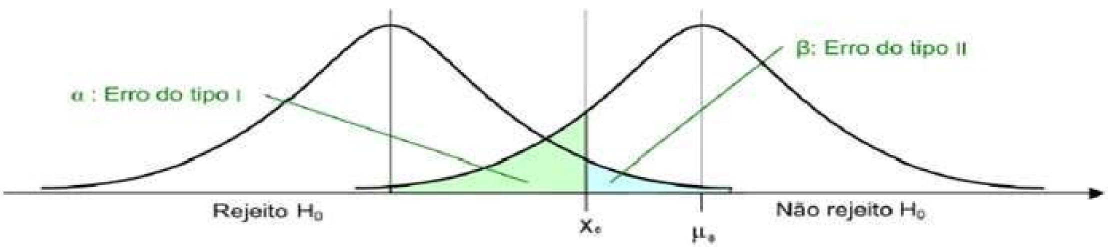
Figura 10.3 – Erros cometidos nos testes de hipóteses. Fonte: PORTAL ACTION (2018a).

A região hachurada é considerada, **P (β/μ = μ$_1$)** = probabilidade de aceitar **μ = 20,** quando **μ = μ$_1$**, é verdadeira.

- Se $\overline{X}$, por exemplo, for igual a 20,5, será aceita a hipótese falsa: **H$_0$**: **μ = 20,** quando na realidade a verdadeira hipótese é **H$_0$**: **μ = μ$_1$**
- Quando consideramos valores de **μ$_1$** próximos de 20, temos elevados índices para β.
- Imagine, olhando o gráfico, o deslocamento de **μ$_1$** para a esquerda. Quando **μ$_1$ = 21,64**, temos **P (β/μ = 21,64) = 50%** e esse valor irá crescendo, à medida que consideramos valores para **μ$_1$** menos que **21,64.**

Vamos calcular alguns valores de β.

a) Quanto valerá β, dado que **μ = 22?**



↓ ↓ ↓ ↓
20 21,64 22 $\overline{X}$

Figura 10.4 - Erros cometidos nos testes de hipóteses. Fonte: PORTAL ACTION (2018a).

Aplicar a fórmula:

$$\mathbf{Z} = \frac{\bar{x}_c - \mu}{\sigma/\sqrt{n}} \quad \text{Para } (\sigma = 4;\ n = 16)$$

$$P(\beta/\mu = 22) = P(\bar{x} < 21{,}64/\ \mu = 22) = P\left(Z < \frac{21{,}64 - 22}{4/\sqrt{16}} = -0{,}36\right) = 0{,}3594$$

b) Quanto valerá β, dado que **μ = 21,64**?

$$P(\beta/\mu = 21{,}64) = P(\bar{x} < 21{,}64/\mu = 21{,}64) = P\left(Z < \frac{21{,}64 - 21{,}64}{4/\sqrt{16}} = 0\right) = 0{,}50$$

Exemplo 10.2

Explicar os Erros tipo I e II que podem ocorrer quando um professor decide aprovar ou reprovar um aluno.

Resposta:

Quando um professor decide aprovar um aluno, poderá estar cometendo um Erro tipo II – aceitar $\mathbf{H_0}$, sendo $\mathbf{H_0}$ falsa: no caso, aprovar o aluno quando deveria reprová-lo. Por outro lado, quando um professor decide reprovar um aluno, poderá estar cometendo um Erro tipo I – rejeitar $\mathbf{H_0}$, sendo $\mathbf{H_0}$ verdadeira – no caso, reprovar o aluno quando deveria aprová-lo.

Exercício proposto

Explicar os Erros tipo I e II que podem ocorrer quando um gerente de recursos humanos decide contratar, ou não, determinado profissional.

Resposta:

2.4 TESTES DE SIGNIFICÂNCIA - Os testes de significância consideram apenas o erro **α.** O procedimento para realização dos testes de significância pode ser resumido nos seguintes passos:

Primeiro passo: enunciar as hipóteses $\mathbf{H_0}$ **e** $\mathbf{H_1}$**;**

$\mathbf{H_0}$ **(**hipótese nula**):** utilizar um dos símbolos**:** $\leq, =, \geq$

$\mathbf{H_1}$ **(**hipótese alternativa); utilizar um dos símbolos**:** $<, \neq, >$

Segundo passo: fixar o limite do erro e identificar a **variável do teste;**

Terceiro passo: faça um gráfico da distribuição amostral. De acordo com a hipótese alternativa (H_1), marque a região crítica (RC) do teste.

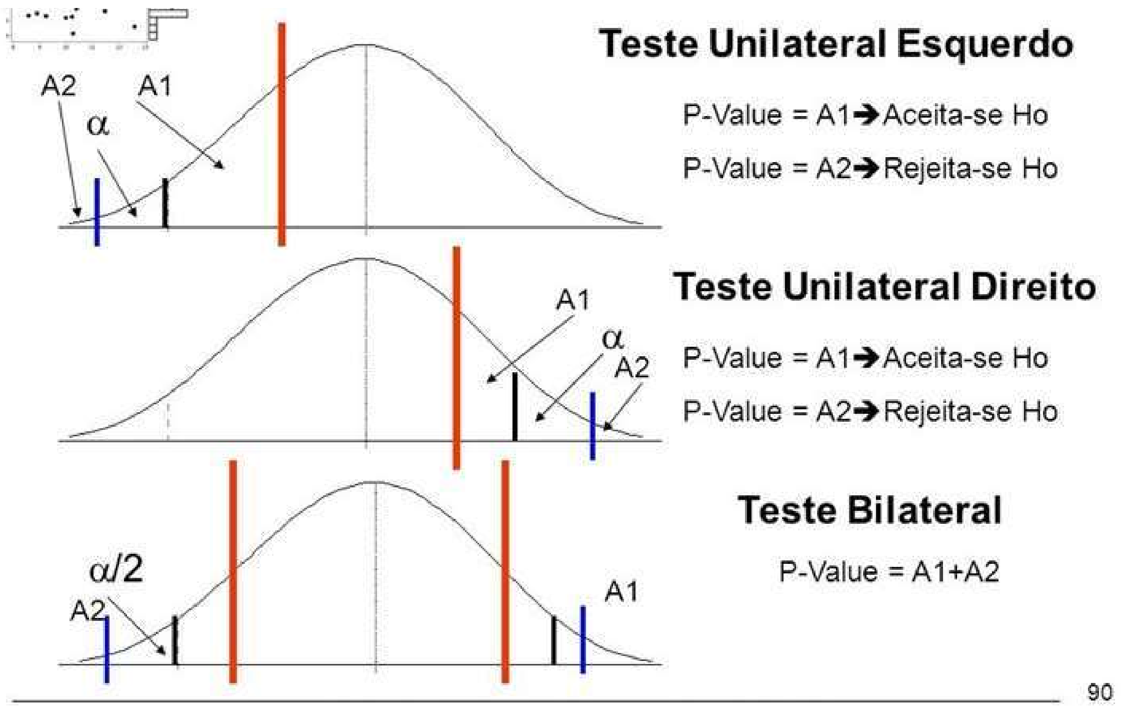


Figura 10.5 – Testes de significâncias. Fonte: BALESTRASSI (2019).

Notas:

1. Teste unilateral à direita: H_1 possui o símbolo >
2. Teste unilateral à esquerda: H_1 possui o símbolo <
3. Teste bilateral: H_1 possui o símbolo ≠

Quarto passo: com o auxílio das **Tabelas estatísticas (Normal**, t de Student, ...) considerando **α** (nível de significância) e a variável do teste, determinar a **RC (região crítica**) e a **RA (região de aceitação**) para H_0. Marque esse valor no gráfico.

Quinto passo: com os elementos amostrais, calcular o valor da variável do teste;

Sexto passo: concluir pela aceitação ou rejeição de H_0, comparando o valor obtido no passo (5º) com RA e RC.

- Se o valor observado pertence a RC, então rejeita H_0 e aceita H_1;
- Se o valor observado não pertence a RC, então aceite H_0 e rejeita H_1.

2.4.1 Teste de significância para médias populacionais - Quando o desvio padrão populacional **(σ),** for desconhecido, optamos pelo desvio padrão amostral **(S)**.

1) H_0: $\mu = \mu_0$

H_1: uma das alternativas

$\mu \neq \mu_0$ (a)

$\mu > \mu_0$ (b)

$\mu < \mu_0$ (c)

2) Fixar **α.** Admitindo-se variância **(σ^2)** desconhecida, a variável do teste será **t de Student,** com **φ = (n-1) graus de liberdade.**

3) Com auxílio da tabela **t,** determinam-se **RA e RC (**Anexo E)

4) Cálculo do valor da variável**:**

$$t_{cal} = \frac{\bar{x}_c - \mu_0}{S/\sqrt{n}}$$

Onde:

$\bar{x}$ = média amostral

μ_0 = valor da hipótese nula

S = desvio padrão amostral

n = tamanho da amostra

5) Conclusão:

a) Se $-t_{\alpha//2} \leq t_{cal} \leq t_{\alpha/2}$, não se pode rejeitar H_0.

Se $t_{cal} > t_{\alpha/2}$ ou $t_{cal} < -t_{\alpha/2}$, rejeita-se H_0.

b) Se $t_{cal} < t_\alpha$, não se pode rejeitar H_0.

Se $t_{cal} > t_\alpha$, rejeita-se H_0.

c) Se $t_{cal} > t_\alpha$, não se pode rejeitar H_0.

Se $t_{cal} < t_\alpha$, rejeita-se H_0.

Exemplo 10.3 (MARTINS, 2010)

Os registros dos últimos anos de um colégio atestam para os calouros admitidos uma nota média 115 (teste vocacional). Para testar a hipótese de que a média de uma nova turma é a mesma das turmas anteriores, retirou-se, ao acaso, uma amostra de 20 notas, obtendo-se média 118 e desvio padrão 20.

Admitir **α = 0,05**, para efetuar o teste.

Solução:

1) $\mathbf{H_0}$: $\mu = 115$

   $\mathbf{H_1}$: $\mu \neq 115$

2) Fixar: **α = 0,05**. Variável **t** como 19 graus de liberdade, pois $\boldsymbol{\varphi = n-1}$, e neste caso, $\boldsymbol{\varphi}$ $= 20 - 1 = 19$

3) Com auxílio da Tabela t, determinam-se RA e RC (Anexo E).

   gl=19 e área (α= 0,025) → 2,093

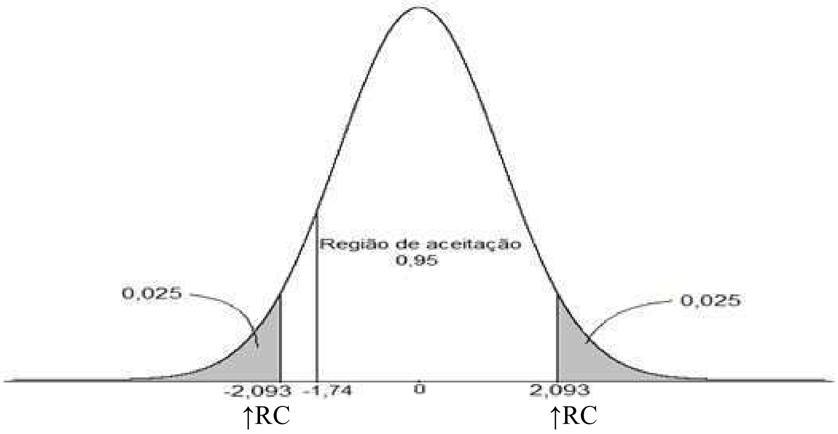


Figura 10.6 – Teste para média (Teste t). PORTAL ACTION (2018b).

5) Cálculo do valor da variável:

$$\mathbf{t_{cal}} = \frac{\bar{x}_c - \mu_0}{S/\sqrt{n}} \quad \text{(onde: } \bar{x} = 118 \quad \mu_0 = 115 \quad S = 20 \quad n = 20\text{)}$$

$$t_{cal} = \frac{118 - 115}{20/\sqrt{20}}$$

$$t_{cal} = \frac{3}{20/4,47}$$

$$t_{cal} = \frac{3}{4,47}$$

$$t_{cal} = 0,67$$

$t_{cal} = 0,67$ não pertence a RC, aceita H_0 e rejeita H_1.

6) Como $-2,093 \leq \mathbf{t_{cal}} \leq 2,093$, não se pode rejeitar $\mathbf{H_0}$: $\mu = 115$ com um nível de significância de 5%.

Exemplo 10.4

Uma amostra de 25 elementos, extraída de uma população normal, resultou média de 13,5 com desvio padrão de 4,4. Efetuar o teste ao nível de 0,05 para hipótese H_0: $\mu = 16$ contra $\mu \neq 16$.

Solução:

1) $\mathbf{H_0: \mu = 16}$
   $\mathbf{H_1: \mu \neq 16}$

2) Fixar $\boldsymbol{\alpha} = \mathbf{0,05}$. Variável **t** como 24 graus de liberdade, pois $\boldsymbol{\varphi} = \mathbf{n-1}$, e neste caso, $\varphi = 25 - 1 = 24$
3) Não solicita RA nem RC
4) Cálculo do valor da variável:

$$\mathbf{t_{cal}} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \quad (\text{onde: } \bar{x} = 13,5 \quad \mu_0 = 16 \quad S = 4,4 \quad n = 25)$$

$$t_{cal} = \frac{13,5 - 16}{4,4/\sqrt{25}}$$

$$t_{cal} = \frac{-2,5}{4,4/5}$$

$$t_{cal} = \frac{-2,5}{4,4} \cdot 5$$

$$\mathbf{t_{cal} = -2,84}$$

Como $\mathbf{t_{cal}} = -2,84$, rejeita-se H_0, concluindo-se, com risco de 5%, que a média é diferente de 16.

2.4.2 Teste para média com variância conhecida (σ^2): Quando o desvio padrão (σ) for conhecido a estatística do teste é:

$$\mathbf{Zt = \frac{(\bar{x} - \mu) . \sqrt{n}}{\Sigma}}$$

onde:

$\bar{x}$ = é a média amostral

μ = é a média populacional testada **(sob H_0)**

σ = é o desvio padrão populacional

n = é o tamanho da amostra

Exemplo 10.5

Funcionários de uma grande firma de contabilidade afirmam que a média dos salários dos contadores é menor que a de seu concorrente, que é de $45.000. Uma amostra aleatória de 30 contadores da firma mostrou que a média dos salários é de $43.500.e sabe-se, de estudos anteriores, que o desvio padrão dos salários é $5.200. Teste a afirmação dos funcionários ao nível de 5% de significância.

Solução:

H_0: μ ≥ 45.000

H_1: μ < 45.000

Temos: $\bar{x}$ = 43.500 σ = 5.200 n = 30 α = 5%

$$\mathbf{Zt = \frac{(\bar{x} - \mu) \sqrt{n}}{\Sigma}}$$

$$Zt = \frac{(43.500 - 45.000) \sqrt{30}}{5.200}$$

$$Zt = \frac{(-1.500) . 5,48}{5.200}$$

Zt = 1,58

↓5% ↓45%

Valores da distribuição normal:(1,64 + 1,65)/2 = 1,645

(Valores: 1,6 + 0,4495 = 1,64; 1,64 + 0,4505 = 1, 65)

Z não pertence a RC → aceita **H_0** e rejeita **H_1.**

Ao nível de 5%, não há evidencias de que o salário médio dos funcionários seja inferior ao dos concorrentes.

2.4.3 Teste de significância para variâncias populacionais

1) $\mathbf{H_0 = \sigma^2 = \sigma^2_0}$

H_1: uma das alternativas

$\sigma^2 \neq \sigma^2_0$ (a)

$\sigma^2 > \sigma^2_0$ (b)

$\sigma^2 < \sigma^2_0$ (c)

2) Fixar α. Escolher a variável **Qui-Quadrado** com $\varphi = (n-1)$ graus de liberdade.

3) Com auxílio da Tabela da distribuição **Qui – Quadrado,** determinam-se **RA e RC** (Anexo D**).**

4) Cálculo do valor da variável:

$\chi^2_{cal} = [(n-1)\,S^2]/\sigma^2_0$

Onde:

n = tamanho da amostra;

S^2 = variância amostral;

σ^2 = hipótese nula.

5) Conclusão:

a) Se $\chi^2_{inf} \leq \chi^2_{cal} \leq \chi^2_{sup}$, não se pode rejeitar $\mathbf{H_0}$

Se $\chi^2_{cal} > \chi^2_{sup}$ **ou** $\chi^2_{cal} < \chi^2_{inf}$, rejeita-se $\mathbf{H_0}$

b) Se $\chi^2_{cal} < \chi^2_{sup}$, não se pode rejeitar $\mathbf{H_0}$

Se $\chi^2_{cal} > \chi^2_{sup}$, rejeita-se $\mathbf{H_0}$

c) Se $\chi^2_{cal} > \chi^2_{inf}$, não se pode rejeitar $\mathbf{H_0}$

Se $\chi^2_{cal} < \chi^2_{inf}$, rejeita-se $\mathbf{H_0}$

Exemplo 10.6 (MARTINS, 2010)

Para testar a hipótese de que a variância de uma população com distribuição normal é 25, tirou-se uma amostra aleatória de 25 elementos obtendo-se $\mathbf{S^2 = 18{,}3}$. Admitindo α= 0,10, efetue o teste de significância unicaudal à esquerda.

Solução:

1) H_0: $\sigma^2 = 25$

H_1: $\sigma^2 < 25$

2) A = 0,10. Escolher uma variável Qui-Quadrado com 24 graus de liberdade, pois: $\varphi n = n-1 = 25 - 1 = 24$

3) Com auxílio da Tabela da distribuição Qui-Quadrado (Anexo E) determinam-se RC e RA.

4) $\boldsymbol{\chi^2_{cal} = [(n-1)\,S^2]/\sigma^2_0}$

Temos: $n = 25$; $S^2 = 18{,}03$; $\sigma^2 = 25$

$\chi^2_{cal} = [(25 - 1)18{,}3]/25 = 17{,}56$

5) Como $\boldsymbol{\chi^2_{cal} > 15{,}7}$ não se pode rejeitar $\boldsymbol{H_0: \sigma^2 = 25}$ ao nível de significância de 10%.

2.4.4 Teste de significância para proporções populacionais

1) $\boldsymbol{H_0: p = p_0}$

$\boldsymbol{H_1}$: uma das alternativas

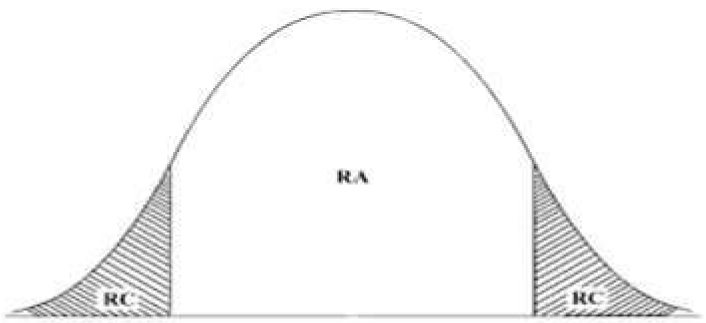
$\boldsymbol{p \neq p_0}$ (a)

$\boldsymbol{p > p_0}$ (b)

$\boldsymbol{p < p_0}$ (c)

2) Fixar α. Escolher a variável normal padrão.

3) Com auxílio da tabela da distribuição normal padrão (Anexo C), determinam-se RA e RC.



$\uparrow \boldsymbol{-Z_{\alpha/2}}$ **0** $\uparrow \boldsymbol{Z_{\alpha/2}}$ $\uparrow$ **Z**

Figura 10.7 – Gráfico do teste bilateral. Fonte: BELL0 (2011).

4) Cálculo do valor da variável:

$$Z_{cal} = \frac{f - p_0}{\sqrt{\dfrac{[p_0\,(1 - p_0)]}{N}}}$$

Onde:

f = frequência relativa do evento na amostra

p_0 = valor da hipótese nula

n = tamanho da amostra

5) Conclusões:

a) Se $-Z_{\alpha/2} \leq Z_{cal} \leq Z_{\alpha/2}$, não se pode rejeitar H_0

Se $Z_{cal} < Z_{\alpha/2}$ **ou** $Z_{cal} < -Z_{\alpha/2}$, rejeita-se H_0

b) **Se** $Z_{cal} < Z_{\alpha}$, não se pode rejeitar H_0

Se $Z_{cal} > Z_{\alpha}$, rejeita-se H_0

c) **Se** $Z_{cal} > -Z_{\alpha}$, não se pode rejeitar H_0

Se $Z_{cal} < -Z_{\alpha}$, rejeita-se H_0

Exemplo 10.7 (MARTINS, 2010)

As condições de mortalidade de uma região são tais que a proporção de nascidos que sobrevivem até 60 anos é de 60% (0,60). Testar essa hipótese, ao nível de 50%, se em 1.000 nascimentos amostrados aleatoriamente, verificaram-se 530 sobreviventes até 6 anos.

Solução:

1) $\mathbf{H_0: p = 0{,}60}$

   $\mathbf{H_1: p \neq 0{,}60}$

2) α**= 0,05**, **variável normal padrão**

3) Com auxílio da Tabela da distribuição normal padrão (Anexo C), determinam-se RA e RC.

   RC = 0,025

4) Cálculo do valor da variável**:**

$$Z_{cal} = \frac{f - p_0}{\dfrac{\sqrt{[p_0\,(1 - p_0)]}}{N}}$$

Temos:

f = frequência relativa do evento na amostra→ f = (530/1000 = 0,53)

p = 0,60 **n = 1.000**

$$Z_{cal} = \frac{0{,}53 - 0{,}60}{\sqrt{[0{,}60\,(1 - 0{,}60)]/1000}}$$

$$Z_{cal} = \frac{-0,07}{0,0155}$$

$$Zcal = -4,52$$

5) Como $\mathbf{Z_{cal} < -1,96}$, rejeita-se $\mathbf{H_0}$, concluindo-se, com nível de 5%, que a proporção de nascidos que sobrevivem até 60 anos é diferente de 60%.

CAPÍTULO XI - CORRELAÇÃO E REGRESSÃO

1. RELAÇÃO ENTRE VARIÁVEIS (X, Y)

Existe correlação entre duas ou mais variáveis quando as alterações sofridas por uma delas são acompanhadas por modificações nas outras, ou seja, no caso de duas variáveis X e Y, os aumentos ou diminuições em X correspondem a aumentos ou diminuições em Y. Assim:

- O consumo das famílias pode estar relacionado com sua renda;
- As vendas de uma empresa e os gastos promocionais podem relacionar-se, bem como a demanda de um determinado produto e seu preço.

a) Quando duas variáveis estão ligadas por uma relação estatística, dizemos que existe correlação entre elas.
b) Objeto de estudo da correlação: a verificação da existência e do grau de relação entre variáveis.
c) Caracterizada a relação, procura-se descrevê-la, sob a forma matemática, através de uma função.
d) A estimação dos parâmetros dessa função matemática é o objeto da regressão.

2. CORRELAÇÃO LINEAR SIMPLES

- O estudo da correlação tem por objetivo medir e avaliar o grau de relação existente entre duas variáveis aleatórias.
- A correlação linear procura medir as relações entre as variáveis X e Y através da disposição dos pontos (X, Y) em torno de uma reta.
- Sua interpretação dependerá do valor numérico e do sinal.
- A correlação linear simples pode ser: correlação linear positiva, correlação linear perfeita positiva, correlação negativa, correlação perfeita negativa, correlação nula e correlação espúria.
- O instrumento de medida da correlação linear é dado pelo Coeficiente de Correlação de Pearson.

2.1 COEFICIENTE DE CORRELAÇÃO LINEAR - É uma medida que avalia o quanto a “nuvem de pontos” no diagrama de dispersão aproxima-se de uma reta.

- O coeficiente deve indicar o grau de intensidade da correlação entre duas variáveis e, o sentido dessa correlação (positivo ou negativo).
- O coeficiente de Correlação linear é dado é pelo Coeficiente de Correlação do Produto de Momentos de Pearson, ou simplesmente Coeficiente de Pearson.

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\,\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

Onde:

n = número de observações.

- O campo do coeficiente **r** varia entre -1 ou +1, ou expresso em porcentagens, entre – 100% e + 100% ou seja, o valor de r pertence ao intervalo

  -1 ≤ r xy ≤ 1 =
- Quanto maior a qualidade do ajuste (ou associação linear), mais próximo de +1 ou -1 estará o valor do coeficiente r próximo de zero.
- Na prática, se r > 70% ou r < -70%, e n ≥ 30, diremos que há forte correlação linear entre as variáveis.

Os diagramas de dispersões ilustram as variações do coeficiente de correlação linear.

Correlação: “Coeficiente de Pearson”

O **coeficiente de correlaçãoo linear** mede o quão próximo estão os pontos de **uma reta**.

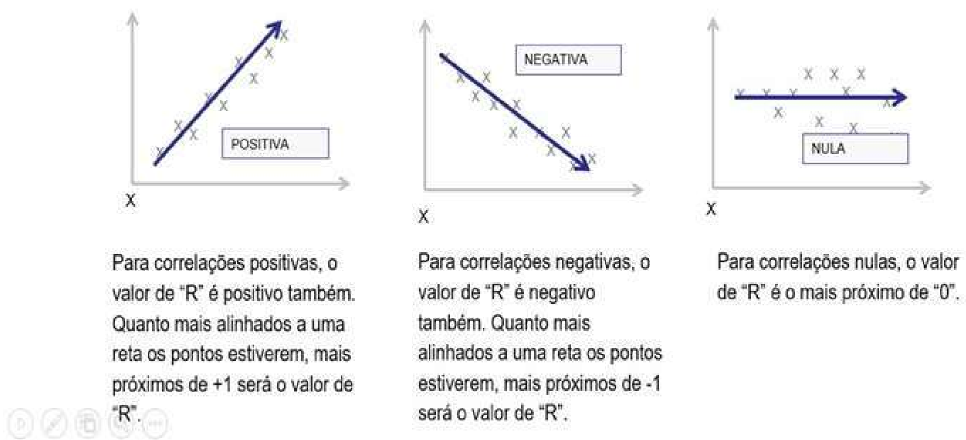


Figura 11.1 – Coeficiente de correlação linear de Pearson. Fonte: GRADUS (2018).

2.2 TIPOS DE CORRELAÇÃO LINEAR

2.2.1 Correlação linear positiva $0 < r_{xy} < 1$: se os pontos do diagrama têm como imagem uma reta ascendente. Valores crescentes de X estiverem associados a valores crescentes de Y, ou valores decrescentes de X estiverem associados a valores decrescentes de Y (Figura 11.2).

2.2.2 Correlação linear perfeita positiva: $r_{xy} = 1$ ou $r = 100\%$.

Os pontos (X, Y) estão perfeitamente alinhados.

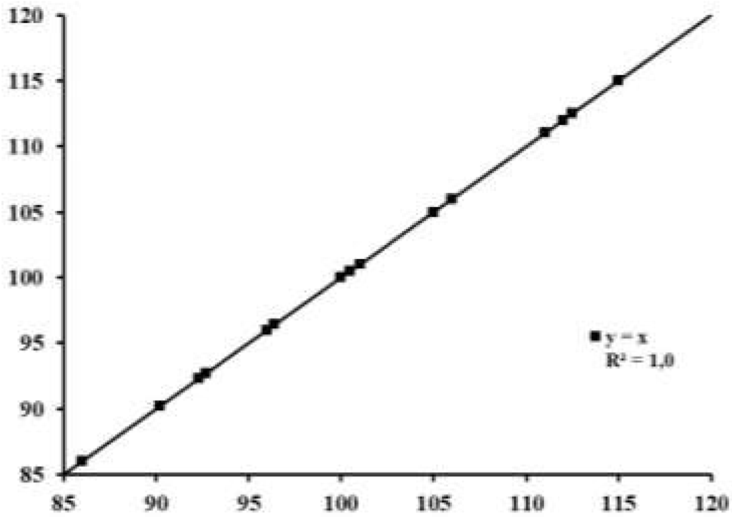


Figura 11.2 – Concordância perfeita entre as duas series de medidas. Fonte: MARTELLI FILHO et al. (2005).

2.2.3 Correlação negativa: $-1 < r_{xy} < 0$. Se os pontos tem uma reta descendente (valores crescentes de X estiverem associados a valores decrescentes de Y, ou valores decrescentes de X associados a valores crescentes de Y (Figura 11.3).

2.2.4 Correlação perfeita negativa: $r_{xy} = -1$ ou $r = -100\%$. Os pontos (X, Y) estão perfeitamente alinhados.

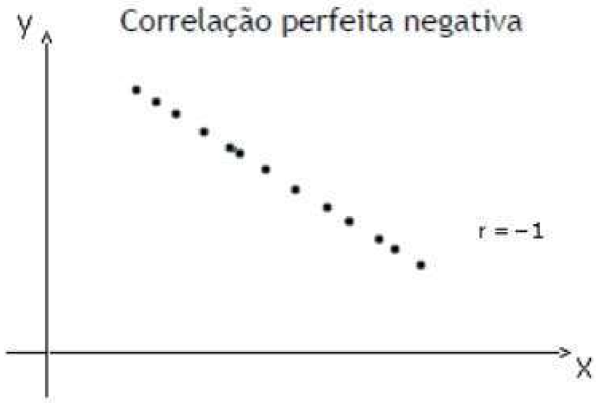


Figura 11.3 – Correlação perfeita negativa. Fonte: RODRIGUES (2012).

2.2.5 Correlação espúria: Quando o coeficiente apresentar um valor próximo de ± 1.

Notas:

1) Para que uma relação possa ser descrita por meio do coeficiente de correlação de Pearson é imprescindível que ela se aproxime de uma função linear.
2) Interpretação significativas sobre o comportamento simultâneo de variáveis analisadas:

- 0,00 a 0,19 → Uma correlação bem fraca
- 0,20 a 0,39 → Uma correlação fraca
- 0,40 a 0,69 → Uma correlação moderada
- 0,70 a 0,89 → Uma correlação forte
- 0,90 a 1,00 → Uma correlação muito forte

3) Para o cálculo do coeficiente de correlação é conveniente a construção de uma Tabela, onde, a partir dos valores X e Y, são determinadas todas as somas necessárias.

Exemplo 11.1

Calcular o coeficiente de correlação linear entre as variáveis X e Y, usando os dados da Tabela abaixo:

Y	10	8	6	10	12
X	2	4	6	8	10

Solução:

n = 5

Tabela 11.1

Y	X	X^2	Y^2	XY
10	2	4	100	20
8	4	16	64	32
6	6	36	36	36
10	8	64	10	80
12	10	100	144	120
Σ=46	**Σ= 30**	**Σ= 220**	**Σ= 444**	**Σ=288**

Fonte: TOLEDO; OVALLE (2008).

Aplicando a fórmula:

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$r = \frac{5\,(288) - (30)\,(46)}{\sqrt{[5\,(220) - (30)^2]}\,\sqrt{[5(444) - (46)^2]}}$$

$$r = \frac{1.440 - 1.380}{\sqrt{(1.100 - 900)}\,\sqrt{(2.220 - 2.116)}}$$

$r = 60/(14{,}14)\,(10{,}20)$

$r = 0{,}416$

O resultado mostra que a correlação linear entre X e Y é positiva (quando X cresce linearmente, Y também cresce linearmente), porém é moderada.

Exemplo 11.2

A Tabela seguinte mostra os resultados de uma pesquisa com 10 famílias de determinada região.

Tabela 11.2 – Resultado da pesquisa.

Famílias	Renda R$100	Poupança R$ 1.000	Números de Filhos	Média de anos de estudo da família
A	10	4	8	3
B	15	7	6	4
C	12	5	5	5
D	70	20	1	12
E	80	20	2	16
F	100	30	2	18
G	20	8	3	8
H	30	8	2	8
I	10	3	6	4
J	60	15	1	8

Fonte: TOLEDO; OVALLE (2008).

a) Calcular o coeficiente de correlação linear entre Rendas familiar e Poupança das dez famílias.

Solução:

Tabela 11.2A**.**

Renda (Y)	**Poupança (X)**	$\mathbf{X^2}$	$\mathbf{Y^2}$	**XY**
10	4	16	100	40
15	7	49	225	105
12	5	25	144	60
70	20	400	4.900	1.400
80	20	400	6.400	1.600
100	30	900	10.000	3.000
20	8	64	400	160
30	8	64	900	240
10	3	9	100	30
60	15	225	3.600	900
Σ = 407	**Σ = 120**	**Σ = 2.152**	**Σ = 26.769**	**Σ = 7.535**

Aplicando a fórmula:

Para n = 10

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$r = \frac{10\,(7.535) - (120)\,(407)}{\sqrt{[10\,(2.152) - (120)^2]}\,\sqrt{[10(26.769) - (407^2]}}$$

$$r = \frac{75.350 - 48.840}{\sqrt{(21.520 - 14.400)}\,\sqrt{(267.690 - 165.649)}}$$

$$r = \frac{26.510}{\sqrt{7.120}\,\sqrt{102.041}}$$

r = 26.510/(84,38) (319,43)

r = 26.510/26.953,5

r = 0,9835

Existe uma forte correlação linear entre renda e poupança familiar. O sinal do coeficiente mostra que as duas variáveis variam no mesmo sentido.

Exercício proposto

Calcular o coeficiente de correlação linear entre renda e números de filhos para as dez famílias.

Solução:

3. REGRESSÃO LINEAR SIMPLES

A análise de regressão tem por objetivo descrever, através de um modelo matemático, a relação existente entre duas variáveis, a partir de **n** observações dessas variáveis.

- A variável sobre a qual desejamos fazer uma estimativa recebe o nome de variável dependente (Y), também chamada de variável explicada e a outra X recebe o nome de variável independente, também chamada de variável explicativa.
- Em regressão considera-se apenas a variável Y como aleatória e a variável X como supostamente sem erro. Então a relação X e Y não é regida apenas por uma lei matemática, ou seja, para um dado valor de X, não necessariamente o mesmo Y.
- Com estas variáveis (X, Y) vamos procurar determinar o ajustamento de uma reta à relação entre essas variáveis.
- A função escolhida será aquela que for sugerida pelo conjunto dos pontos dispostos no diagrama (um conjunto de pontos sugerindo uma função linear)

3.1 MODELO DE REGRESSÃO LINEAR SIMPLES

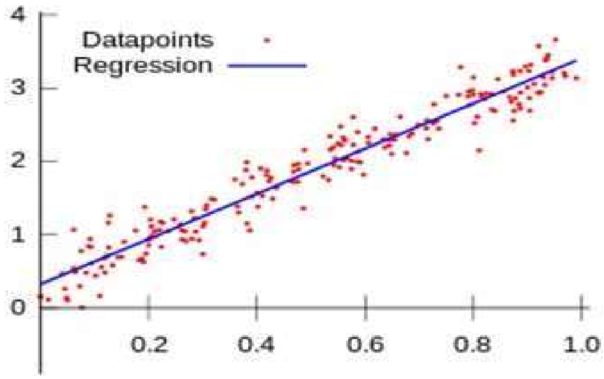


Figura 11.4 – Exemplo de regressão linear. Fonte: WIKIPÉDIA (2018).

Observando o diagrama de dispersão, podemos ter uma ideia do tipo de relação entre as duas variáveis.

- A natureza da relação pode tomar várias formas, desde uma simples relação linear até uma complicada função matemática.
- O modelo de regressão linear simples, pode ser representado como:

$$\mathbf{Y_i = \alpha + \beta X_i + \varepsilon_i}$$

Onde:

$\boldsymbol{\alpha}$ = intercepto da reta
$\boldsymbol{\beta}$ = inclinação da reta
$\boldsymbol{\varepsilon_i}$ = erro aleatório de Y para observação i.

Nota:
Assim, a inclinação β representa a mudança esperada de Y por unidade de X; isto é, representa a mudança de Y (tanto positiva quanto negativa) para uma particular unidade de X. Por outro lado, α representa a valor de Y quando X = 0, enquanto $\boldsymbol{\varepsilon_i}$ representa uma variável aleatória que descreve o erro de Y para cada observação i.

3.2 DETERMINAÇÃO DA EQUAÇÃO DE REGRESSÃO LINEAR SIMPLES

- Determina-se com base em uma amostra, a equação de regressão linear simples que melhor se ajuste aos dados amostrais. Encontrar os coeficientes da reta.
- A reta ajustada é representada por:

$$\mathbf{\hat{Y}_i = a + bX_i}$$

Onde:

$\hat{Y}_i$ = o valor da previsão de Y para uma observação;

X_i = o valor de X para a observação **i;**

a = o estimador de α;

b = o estimador de β.

- **a** e **b** são parâmetros: determinar os valores dos parâmetros, de modo que a reta se ajuste ao conjunto de pontos (estimar a e b de algum modo eficiente) Determinação dos parâmetros:

$$\mathbf{a = \bar{Y} - b\ \bar{X}}$$

$$\mathbf{b = \frac{n\,(\Sigma XY) - (\Sigma X)\,.\,(\Sigma Y)}{n\,(\Sigma X^2) - (\Sigma X)^2}}$$

Exemplo 11.3

Os dados abaixo referem-se ao volume de precipitação pluviométrica (mm) e ao volume de produção de leite tipo C (milhões de litros), em determinada região do país.

Tabela 11.3 – Análise de regressão linear.

Anos	Produção de leite tipo C (1.000.000 l)	Índice Pluviométrico (mm)
2009	26	23
2010	25	21
2011	31	28
2012	29	27
2013	27	23
2014	31	28
2015	32	27
2016	28	22
2017	30	26
2018	30	25

Fonte: TOLEDO; OVALLE (2008).

a) Ajustar os dados através de um modelo linear.

b) Admitindo-se em 2019, um índice pluviométrico de 24 mm, qual deverá ser o volume esperado de produção do leite tipo C?

Solução:

Primeiro Passo: Para efetuar os cálculos necessários ao ajustamento recorremos a uma Tabela auxiliar:

Tabela 11.3.1 - Resolução

Y	**X**	**X^2**	**XY**
26	23	529	598
25	21	441	525
31	28	784	868
29	27	729	783
27	23	529	621
31	28	784	868
32	27	729	864
28	22	484	616
30	26	676	780
30	25	625	750
Σ = 289	**Σ = 250**	**Σ = 6.310**	**Σ = 7.273**

Segundo Passo: Determinação do Valor do Parâmetro **b**.

n=10

$$b = \frac{n\,(\Sigma XY) - (\Sigma X)\,.\,(\Sigma Y)}{n\,(\Sigma X^2) - (\Sigma X)^2}$$

$$b = \frac{10\,(7.273) - (250)\,.\,(289)}{10\,(6.310) - (250)^2}$$

$$b = \frac{72.730 - 72.250}{63.100 - 62.500}$$

$$b = \frac{480}{600}$$

$$b = 0{,}8$$

Terceiro Passo: Determinação do Valor do Parâmetro a

$$\mathbf{a = \bar{Y} - b\,\bar{X}}$$

$$a = \Sigma Y/n - b\Sigma X/n$$

$$a = 289/10 - 0{,}8\,(250/10)$$

$$a = 28{,}9 - 20$$

$$a = 8{,}9$$

Quarto Passo: Equação da Reta Ajustada: $Y = a + bX = 8{,}9 + 0{,}8X$

a) Fazendo x = 24 mm, teremos: $Y = 8{,}9 + 0{,}8\ (24) = 28{,}1$

De acordo com o modelo, podemos esperar 28,1 milhões de litros produzidos para um índice pluviométrico de 24 mm.

Exercício proposto

Uma empresa está estudando como varia a demanda de certo produto em função do seu preço de venda. Para isso levantou as seguintes informações:

Tabela 11.4 – Demanda de produção em função do preço de um produto.

Meses	Unidades vendidas (Y)	Preço de venda (X) p/unidade
J	248	162,00
F	242	167,00
M	234	165,00
A	216	173,00
M	230	170,00
J	220	176,00
J	213	178,00
A	205	180,00
S	198	182,00
O	195	187,00

Fonte: TOLEDO; OVALLE (2008).

Com base nestes dados mostrar que a demanda o produto decresce linearmente com o acréscimo de preço.

REFERÊNCIAS BIBLIOGRAFICAS

ALESSANDRO, J. *Estatística: gráfico estatístico*. 2012. Disponível em: https://pt.slideshare.net/JoaoAlessandro/aula-05-grficos-estatsticos2. Acesso em: 20 fev. 2019.

ALVES, M. C. *Teste t de Student*. 2017. Disponível em: http://cmq.esalq.usp.br/wiki/lib/exe/fetch.php?media=publico:syllabvs:lcf5759a:teste_t.pdf. Acesso em: 20 nov. 2018.

AMARAL, E. F. L. *Estimativas e tamanhos amostrais*. 2012. Disponível em: http://www.ernestoamaral.com/docs/est-122/Aula04.pdf. Acesso em: 10 jan. 2019.

BALESTRASSI, P. P. Engenharia da qualidade I. Disponível em: https://slideplayer.com.br/slide/1847346/. Acesso em: 12 fev. 2019.

BELLO, P. *Teste de hipóteses*. 2011. Disponível em: https://www.docsity.com/pt/teste-hipoteses-3/4753288/. Acesso em: 01 mar. 2019. https://www.docsity.com/pt/teste-hipoteses-3/4753288/

Brainly.com.br. *Tipos de clima no Brasil*. Disponível em: https://brainly.com.br/tarefa/8771758. Acesso em: 10 fev. 2019.

CASTANHEIRA, N. P. Estatística aplicada a todos os níveis. 2 ed.rev. e atualizada. Curitiba: IBPEX, 2003. 315p.

Clubes de Matemática – Obmep. *Problema para ajudar na escola: esporte preferido*. Disponível em: http://clubes.obmep.org.br/blog/problema-para-ajudar-na-escola-esporte-preferido/. Acesso em: 18 jan. 2019.

Contract Pharma Brasil. Introdução à Estatística. Disponível em: http://contractpharmabrasil.com.br/noticias_descricao.php?id=830. Acesso em: 20 mar. 2020.

CORREIA, L. *Curso de estatística medidas de assimetria, positiva, negativa, simétrica normal e distribuição de frequência*. 2014. Disponível em: https://www.youtube.com/watch?v=fu5BVSd1rm4&list=RDCMUCMfhS4g0wOareGNW-BnVUxQ&index=2. Acesso em: 10 fev. 2019.

CORREIA, L. *Curso de estatística interpretação de gráficos e Tabelas análise gráfica dados representados colunas*. 2016. Disponível em: https://www.youtube.com/watch?v=xxA5T-mRWFU. Acesso em: 20 fev. 2019.

COSTA NETO, P. L. de O. *Estatística*. 3ª ed. São Paulo: Edgard Blucher Ltda, 2002.

COSTA, P. R. da. *Estatística*. 3 ed. Santa Maria: Universidade Federal de Santa Maria, Colégio Técnico Industrial de Santa Maria, Curso Técnico em Automação Industrial, 2011. 95p. Disponível em:

<https://www.ufsm.br/app/uploads/sites/413/2018/11/04_estatistica.pdf>. Acesso 19 de janeiro de 2021.

CRESPO, A. A. *Estatística Fácil*. 17ª ed. São Paulo: Saraiva, 2002. 224p.

CV Magazine. *Gráfico, colunas ou barras múltiplas* (CV Magazine). Disponível em: https://www.cvmagazine.info. Acesso em: 20 fev. 2019.

FARIAS, A. M. L. de. Inferência Estatística. 2008. Disponível em: <http://www.pt.wikepedia.org/wike/parametro_(estatistica)>. Acesso em 23 de junho de 2020.

FERREIRA, B. S. *Anotações da aula da disciplina de Epidemiologia geral*. UNIVILLE, 2013. Disponível em: https://aia1317.fandom.com/pt-br/wiki/An%C3%A1lise_Gr%C3%A1fica_-_Normas_e_Interpreta%C3%A7%C3%A3o. Acesso em: 18 dez. 2018.

FGV - *Fundação Getúlio Vargas. Estatística: Teste de hipóteses*. (2018). Disponível em: https://www.estudegratis.com.br/questoes-de-concurso/materia/estatistica/assunto/teste-de-hipotese. Acesso em: 19 fev. 2018.

IBGE – *Instituto Brasileiro de Geografia e Estatística. Tendências Demográficas: uma análise da população com base nos resultados dos censos demográficos 1940 e 2000*. IBGE: Rio de Janeiro, 2007. Disponível em: https: //www.biblioteca.ibge.gov.br/visualizacao/livros/liv34956.pdf. Acesso em: 18 dez. 2018.

IBGE - *Instituto Brasileiro de Geografia e Estatística. Censo Demográficas* população de Uruçuí-PI em 2010 pelo MS/SGEP/Datasus. Disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/cnv/poppi.def >. Acesso em 23 de janeiro de 2021.

GBM – Grupamento de Bombeiro Militar. *Dados estatísticos de acidentes de trânsito no município de Itaituba*. 7º GBM/3º seção, 2014. Disponível em: http://bombeirositaituba.blogspot.com/2014/11/dados-estatisticos-de-acidentes-de.html. Acesso em: 15 jan. 2019.

Google Sites. *População e Amostra*. Disponível em: https://sites.google.com/site/estatisticabasicacc/conteudo/estat/populacao. Acesso em: 15 jan. 2019.

GRADUS. *Regressão linear: o que é e qual a sua importância*? Disponível em: https://www.gradusct.com.br/regressao-linear/. Acesso em: 20 dez. 2018.

IBUJES, M. O. S. *Medidas de forma: asimetría y curtosis*. Disponível em: https://www.inf.ufsc.br/~vera.carmo/LIVROS/LIVROS/Luis%20Felipe%20Dias%20Lopes.pdf. Acesso em: 20 fev. 2019.

ITC – Instituto Tecnológico de Chihuahua. *Distribucion Ji-Cuadrada (X^2)*. Disponível em: http://www.itchihuahua.edu.mx/academic/industrial/estadistica1/cap03b.html. Acesso em: 03 fev. 2019.

LOPES, L. F. D. *Apostila Estatística*. 2003. Disponível em: https://www.monografias.com/trabajos87/medidas-forma-asimetria-curtosis/medidas-forma-asimetria-curtosis.shtml. Acesso em: 15 jan. 2019.

MARTELLI FILHO, J. A.; MALTAGLIATI, L. A.; TREVISAN, F.; GIL, C. T. L. A. *Novo método estatístico para análise da reprodutividade*. Revista Dental Press Ortodon Ortop Facial, v.10, n.5, p.122-129, 2005.

MARTINS, G. de A. *Estatística Geral e Aplicada*. 3ª ed. São Paulo: Atlas, 2010. 421p.

Matemática estatística SPSS. Formação-Explicações-Trabalhos - Apoio metodológico - Realização de análise estatística com SPSS. 2014. Disponível em: http://projetos-estatisticaspss.blogspot.com/2014/04/estatistica-inferencial.html. Acesso em: 20 mar. 2020.

Matematica.pt. *Tipos de gráficos estatísticos*. Disponível em: https://www.matematica.pt/util/resumos/tipos-graficos-estatisticos.php. Acesso em: 20 jan. 2019.

MestAtisCticSa. *Gráfico de linhas ou cronogramas*. Disponível em: https://estatisticasqn.webnode.pt/estatistica/graficos-de-linhas-ou-cronogramas-/. Acesso em: 25 jan. 2019.

MOORE, D. *A Estatística Básica e sua Prática*. 5. ed. Editora LTC, 2011.

MOROTTIN, P. A.; BUSSAB, W. de O. *Estatística Básica*. 6. ed. Editora Saraiva, 2010.

OLIVEIRA, A. *Representação gráfica de dados*. Disponível em: http://alexandreprofessor.blogspot.com/p/graficos.html. Acesso em: 18 jan. 2019.

OLIVEIRA, M. *Separatrizes com dados não agrupados*. 2016. Disponível em: https://manoeloliveiracontabil.blogspot.com/2016/10/separatrizes.html. Acesso em: 17 fev. 2019.

PINTO, A. G. *Estatística Descritiva*. Disponível em: https://pt.slideshare.net/AndersonGP/estatstica-descritiva-8427834. Acesso em: 15 jan. 2019.

PINTO, A. L. D. *Gráficos Estatísticos*. 2015. Disponível em: https://slidesshare.net/ThamirisFrade/aula-01-0915graficosestatisticos. Acesso em: Acesso em: 16 fev. 2019.

PIZZOTTI, J. *Tabelas e Gráficos Pesquisa Mercado*. 2011. Disponível em: http://jppesquisasemarketing.blogspot.com/2011/09/Tabelas-e-graficos-pesquisa-mercado.html. Acesso em: Acesso em: 16 jan. 2019.

PORTAL ACTION. *Erros cometidos nos testes de hipóteses*. Disponível em: http://www.portalaction.com.br/inferencia/511-erros-cometidos-nos-testes-de-hipoteses. Acesso em: Acesso em: 16 dez. 2018a.

PORTAL ACTION. T*este para média (Teste t)*. Disponível em: http://www.portalaction.com.br/inferencia/52-teste-para-media-teste-t. Acesso em: Acesso em: 16 dez. 2018b.

RAMOS, R. Estimativa e estimador: vamos extrapolar? 2016. Disponível em:<http://www.oestatistico.com.br?estimativa-e-estimador>. Acesso em 22 de maio de 2020.

RODRIGUES, S. C. A. *Modelo de regressão linear e suas aplicações*. 2012. Disponível em: https://ubibliorum.ubi.pt/bitstream/10400.6/1869/1/Tese%20Sandra%20Rodrigues.pdf . Acesso em: Acesso em: 12 jan. 2019. https://ubibliorum.ubi.pt/bitstream/10400.6/1869/1/Tese%20Sandra%20Rodrigues.pdf

SILVA, C. *Apostila Matemática e Estatística*. Disponível em: https://issuu.com/diadematematica/docs/apostila-mat-est-2010.2/39. Acesso em: Acesso em: 15 fev. 2019.

SILVA, M. N. P. da. *"Gráficos"*. Brasil Escola. Disponível em: https://brasilescola.uol.com.br/matematica/graficos.htm. Acesso em 03 de fev. de 2019.

TECMUNDO. *Saiba qual tipo de gráfico representa melhor os seus dados no Excel 2007*. Disponível em: https://www.tecmundo.com.br/excel/1745-saiba-qual-tipo-de-grafico-representa-melhor-os-seus-dados-no-excel-2007.htm. Acesso em 03 de fev. de 2019.

TOLEDO, G. L.; OVALLE, I. I. *Estatística Básica*. 2.ed. São Paulo: Atlas, 2008.

TROIA, L. *Resultado de análise de dados*. Disponível em: https://slideplayer.com.br/slide/85698/. Acesso em: 20 jan. 2019.

VIERIRA, M. T.; BESSEGATO, L. F. *Elementos de Estatística: noções de amostragem*. Juiz de Fora: UFJF, 2013. Disponível em: http://www.bessegato.com.br/UFJF/05_elem_amostragem1.pdf. Acesso em: 20 dez. 2018.

WIKIPÉDIA. *Regressão linear*. Disponível em: https://pt.wikipedia.org/wiki/Regress%C3%A3o_linear. Acesso em: 15 dez. 2018.

Wordpress.com. *Ilustração de uma Amostragem Estratificada Proporcional*. Disponível em: https://econpolrg.files.wordpress.com/2016/04/importancia-desenho-amostral.pdf. Acesso em: 20 jan. 2019.

www.cempre.org.br.Adaptado/imagem. *Municípios com coleta seletiva de lixo sólido no Brasil*. Acesso em: 16 jan. 2019.

www.passeidireto.com/arquivo/6064359/0-Tabela-de-numeros-aleatorios-1. *Tabela de Números Aleatórios*. Acesso em: 20 jan. 2019.

ANEXO

ANEXO A

Tabela de Números Aleatórios

57 72 00 39 84	84 41 79 67 71	40 21 13 97 56	49 86 54 08 93	29 68 74 54 83
28 80 53 51 59	09 93 98 87 58	70 27 71 77 17	06 32 02 78 62	16 74 69 65 17
92 59 18 52 87	30 48 86 97 48	35 25 18 88 74	03 62 98 38 58	65 86 42 41 03
90 38 12 91 74	30 19 75 89 07	50 64 15 59 71	88 13 74 95 30	52 78 30 11 75
80 91 16 94 67	58 60 82 06 66	90 47 56 18 46	45 11 12 35 32	45 50 41 13 43
22 01 70 31 32	96 91 92 75 40	16 54 29 72 74	99 00 95 97 61	00 98 24 30 07
56 24 10 04 30	20 46 29 90 53	53 11 05 84 41	21 64 79 19 76	29 51 62 60 66
79 44 92 62 02	96 86 64 30 00	94 56 69 30 20	59 87 87 35 44	22 50 97 78 19
53 99 66 45 08	89 78 50 77 53	37 25 77 41 27	62 38 02 23 57	62 01 41 60 35
18 92 87 35 88	56 05 21 36 51	39 28 50 14 66	85 79 30 19 79	72 66 64 31 45
53 08 58 96 63	05 61 25 70 22	50 41 28 96 62	66 43 63 06 63	01 32 79 85 22
03 58 80 29 28	76 89 51 18 24	88 89 46 47 48	59 19 29 87 03	10 33 99 67 12
27 07 81 88 65	69 49 98 00 28	04 70 51 30 01	47 18 97 33 21	85 82 45 43 24
05 21 08 59 01	06 22 24 98 91	81 17 55 44 66	16 07 73 07 66	10 12 31 78 58
40 36 13 27 84	30 82 33 36 39	69 42 05 58 64	61 12 33 89 27	89 52 66 71 93
54 60 25 28 85	88 20 00 10 59	61 05 36 61 33	72 01 01 19 01	61 10 51 20 91
71 51 63 40 76	71 11 73 73 52	37 31 60 45 88	92 73 43 71 28	04 98 09 02 48
61 02 01 81 73	92 60 66 73 58	53 34 42 68 26	38 34 03 27 44	96 04 46 65 93
82 55 93 13 46	30 95 26 55 06	96 17 65 91 72	39 79 96 12 49	52 80 63 26 99
89 98 54 14 21	74 13 57 68 19	86 28 60 89 47	33 15 26 28 77	45 38 48 08 08
00 99 84 84 14	67 95 13 77 58	90 14 50 79 42	73 63 31 06 60	43 40 12 55 04
62 41 50 78 20	48 05 88 43 52	98 03 19 93 92	03 04 97 25 84	95 95 03 63 31
94 27 90 69 24	68 09 92 11 86	07 63 83 19 32	99 51 15 55 71	09 27 02 67 00
44 89 29 28 84	36 28 25 15 82	87 74 18 97 25	76 10 63 26 76	02 26 74 53 28
97 30 76 95 33	21 10 54 26 95	66 65 52 04 99	36 58 48 03 08	93 63 58 17 96
39 16 58 04 44	80 15 59 59 83	90 95 54 66 81	84 39 60 85 38	88 66 33 35 69
60 78 11 03 26	67 50 34 09 61	31 30 20 76 93	66 30 83 51 09	33 83 64 76 05
03 19 23 47 62	89 57 77 91 33	88 47 60 59 37	54 39 48 77 67	49 85 38 43 91
41 28 52 67 56	25 39 59 96 65	51 36 90 32 22	39 33 05 22 99	03 39 97 96 99
77 54 98 50 39	25 37 42 52 97	10 03 56 04 92	81 66 86 70 01	48 89 55 82 10
28 63 41 61 91	64 24 83 81 37	34 48 83 27 96	38 71 69 73 06	77 50 25 64 60
74 24 48 85 40	12 33 59 67 50	14 98 14 26 42	79 79 13 52 89	69 78 80 44 71
00 24 03 37 96	46 68 75 05 32	42 16 63 33 28	97 26 36 47 27	73 65 38 34 46
05 41 47 69 69	45 36 16 71 18	95 51 97 22 04	13 23 96 58 60	03 69 48 79 83
62 69 84 97 97	47 23 66 51 56	13 08 69 11 52	73 59 26 86 81	80 43 00 98 92

Fonte: http://www.edisciplinas.usp.br/pluginfile.php/5710761/mod_resource/content/1/tabela-nemeros-aleatorios.pdf.

ANEXO B

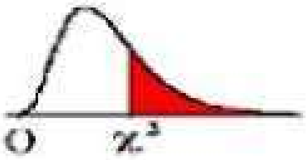
Tabela de Distribuição Normal Padrão (Z < z)

Z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0.6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7515	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Fonte: http://www.wike.icm.usp.br/imagens/f/f9/tabela_Normal.pdf

ANEXO C

Distribuição Qui-Quadrado



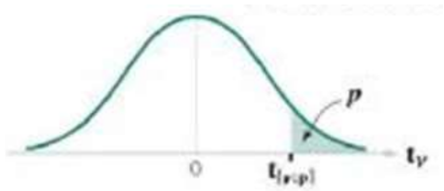
A tabela fornece os valores "c" tais que $P(x^2 > c) = p$
Onde "n" é o número de graus de liberdade e "p" é a probabilidade de sucesso.

Gl	0,995	0,990	0,975	0,950	0,900	0,500	0,100	0,050	0,025	0,010	0,005
1	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	5,348	10,605	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	12,340	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	17,338	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	18,338	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	19,337	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	20,337	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	21,337	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	22,337	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	23,337	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	24,337	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	25,336	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	26,336	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	27,336	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	28,336	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	29,336	40,256	43,773	46,979	50,892	53,672
35	17,192	18,509	20,569	22,465	24,797	34,336	46,059	49,802	53,203	57,342	60,275
40	20,707	22,164	24,433	26,509	29,051	39,335	51,805	55,758	59,342	63,691	66,766
45	24,311	25,901	28,366	30,612	33,350	44,335	57,505	61,656	65,410	69,957	73,166
50	27,991	29,707	32,357	34,764	37,689	49,335	63,167	67,505	71,420	76,154	79,490
55	31,735	33,570	36,398	38,958	42,060	54,335	68,796	73,311	77,380	82,292	85,749
60	35,534	37,485	40,482	43,188	46,459	59,335	74,397	79,082	83,298	88,379	91,952
70	43,275	45,442	48,758	51,739	55,329	69,334	85,527	90,531	95,023	100,425	104,215
80	51,172	53,540	57,153	60,391	64,278	79,334	96,578	101,879	106,629	112,329	116,321
90	59,196	61,754	65,647	69,126	73,291	89,334	107,565	113,145	118,136	124,116	128,299
100	67,328	70,065	74,222	77,929	82,358	99,334	118,498	124,342	129,561	135,807	140,169
110	75,550	78,458	82,867	86,792	91,471	109,334	129,385	135,480	140,917	147,414	151,948
120	83,852	86,923	91,573	95,705	100,624	119,334	140,233	146,567	152,211	158,950	163,648

Fonte: http://www.ime.unicamp.br/~cnaber/Tabela%20da%20Qui-quadrado.pdf.

ANEXO D

Tabela da distribuição T-Student



v: graus de liberdade

p: probabilidade acima

$P(t_v > t_{[v;p]}) = p$

	P													
p	**0,40**	**0,30**	**0,20**	**0,15**	**0,10**	**0,05**	**0,025**	**0,02**	**0,015**	**0,01**	**0,0075**	**0,005**	**0,0025**	**0,0005**
1	0,325	0,727	1,376	1,963	3,078	6,314	12,706	15,894	21,205	31,812	42,433	63,656	127,321	636,578
2	0,289	0,617	1,061	1,386	1,886	2,920	4,303	4,849	5,643	6,965	8,073	9,925	14,089	31,600
3	0,277	0,584	0,978	1,250	1,638	2,353	3,182	3,482	3,896	4,541	5,047	5,841	7,453	12,924
4	0,271	0,569	0,941	1,190	1,533	2,132	2,776	2,999	3,298	3,747	4,088	4,604	5,598	8,610
5	0,267	0,559	0,920	1,156	1,476	2,015	2,571	2,757	3,003	3,365	3,634	4,032	4,773	6,869
6	0,265	0,553	0,906	1,134	1,440	1,943	2,447	2,612	2,829	3,143	3,372	3,707	4,317	5,959
7	0,263	0,549	0,896	1,119	1,415	1,895	2,365	2,517	2,715	2,998	3,203	3,499	4,029	5,408
8	0,262	0,546	0,889	1,108	1,397	1,860	2,306	2,449	2,634	2,896	3,085	3,355	3,833	5,041
9	0,261	0,543	0,883	1,100	1,383	1,833	2,262	2,398	2,574	2,821	2,998	3,250	3,690	4,781
10	0,260	0,542	0,879	1,093	1,372	1,812	2,228	2,359	2,527	2,764	2,932	3,169	3,581	4,587
11	0,260	0,540	0,876	1,088	1,363	1,796	2,201	2,328	2,491	2,718	2,879	3,106	3,497	4,437
12	0,259	0,539	0,873	1,083	1,356	1,782	2,179	2,303	2,461	2,681	2,836	3,055	3,428	4,318
13	0,259	0,538	0,870	1,079	1,350	1,771	2,160	2,282	2,436	2,650	2,801	3,012	3,372	4,221
14	0,258	0,537	0,868	1,076	1,345	1,761	2,145	2,264	2,415	2,624	2,771	2,977	3,326	4,140
15	0,258	0,536	0,866	1,074	1,341	1,753	2,131	2,249	2,397	2,602	2,746	2,947	3,286	4,073
16	0,258	0,535	0,865	1,071	1,337	1,746	2,120	2,235	2,382	2,583	2,724	2,921	3,252	4,015
17	0,257	0,534	0,863	1,069	1,333	1,740	2,110	2,224	2,368	2,567	2,706	2,898	3,222	3,965
18	0,257	0,534	0,862	1,067	1,330	1,734	2,101	2,214	2,356	2,552	2,689	2,878	3,197	3,922
19	0,257	0,533	0,861	1,066	1,328	1,729	2,093	2,205	2,346	2,539	2,674	2,861	3,174	3,883
20	0,257	0,533	0,860	1,064	1,325	1,725	2,086	2,197	2,336	2,528	2,661	2,845	3,153	3,850
21	0,257	0,532	0,859	1,063	1,323	1,721	2,080	2,189	2,328	2,518	2,649	2,831	3,135	3,819
22	0,256	0,532	0,858	1,061	1,321	1,717	2,074	2,183	2,320	2,508	2,639	2,819	3,119	3,792
23	0,256	0,532	0,858	1,060	1,319	1,714	2,069	2,177	2,313	2,500	2,629	2,807	3,104	3,768
24	0,256	0,531	0,857	1,059	1,318	1,711	2,064	2,172	2,307	2,492	2,620	2,797	3,091	3,745
25	0,256	0,531	0,856	1,058	1,316	1,708	2,060	2,167	2,301	2,485	2,612	2,787	3,078	3,725
26	0,256	0,531	0,856	1,058	1,315	1,706	2,056	2,162	2,296	2,479	2,605	2,779	3,067	3,707
27	0,256	0,531	0,855	1,057	1,314	1,703	2,052	2,158	2,291	2,473	2,598	2,771	3,057	3,689
28	0,256	0,530	0,855	1,056	1,313	1,701	2,048	2,154	2,286	2,467	2,592	2,763	3,047	3,674
29	0,256	0,530	0,854	1,055	1,311	1,699	2,045	2,150	2,282	2,462	2,586	2,756	3,038	3,660
30	0,256	0,530	0,854	1,055	1,310	1,697	2,042	2,147	2,278	2,457	2,581	2,750	3,030	3,646
40	0,255	0,529	0,851	1,050	1,303	1,684	2,021	2,123	2,250	2,423	2,542	2,704	2,971	3,551
60	0,254	0,527	0,848	1,045	1,296	1,671	2,000	2,099	2,223	2,390	2,504	2,660	2,915	3,460
120	0,254	0,526	0,845	1,041	1,289	1,658	1,980	2,076	2,196	2,358	2,468	2,617	2,860	3,373
(∞)	0,253	0,524	0,842	1,036	1,282	1,645	1,960	2,054	2,170	2,326	2,432	2,576	2,807	3,290

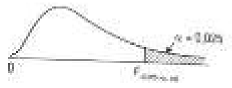
Fonte: http://www.sites.icme.usp.br/francisco/SMEo123/listas/Tabela-Dist_t.pdf

Exemplo: $t_{[10;0,05]} = 1,812$ **significa que $P(t_{10} > 1,812) = 0,05$**

ANEXO E – DISTRIBUIÇÃO F ($\alpha = 0,05$)

Denominador	Numerador φ_1																		
φ_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,46	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

ANEXO F – DISTRIBUIÇÃO F (α = 0,025)



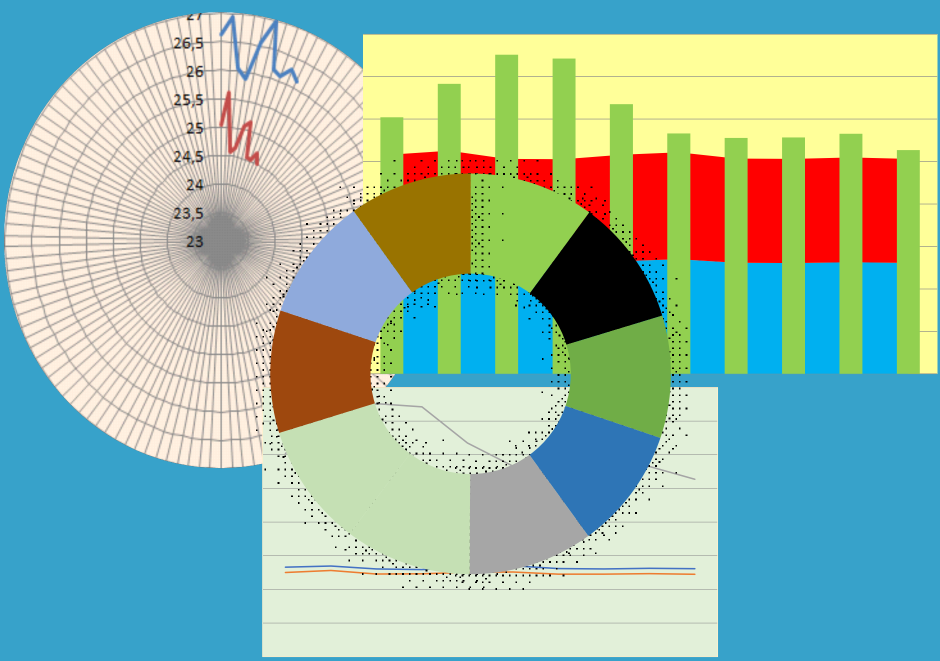
Denominador	Numerador φ_1																		
φ_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3	968,6	976,7	984,9	993,1	997,2	100,1	100,6	101,0	101,4	101,8
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,50
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95	13,90
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,46	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,77	2,67	2,58	2,50	2,44	2,38	2,32	2,26	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98	1,91
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95	1,88
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93	1,85
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91	1,83
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89	1,81
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58	1,48
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,43	1,31
	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,27	1,00

SOBRE AS AUTORAS

Teresinha de Jesus Rocha Vilanova Moreira - Graduada em Matemática pela Universidade Estadual do Piauí-UESPI, especialista em Políticas Educativas pela Faculdade Adelmar Rosado-FAR; Docência do Ensino Superior pela Universidade Estadual do Piauí-UESPI; e Controles na Administração Pública pela Faculdade de Ciências Humanas e Jurídicas de Teresina. Professora das disciplinas de Matemática, Estatística Básica, Administração Rural e Economia Rural no curso de Agronomia da UESPI, com atuação até 2020. Professor Formador na modalidade EAD, na disciplina de Matemática na Universidade Aberta do Brasil - UAB. Docente em cursos profissionalizantes de nível médio: Técnico em Agroindústria e Técnico em Agropecuária no Instituto Federal de Ciências e Tecnologias – IFPI, Campus de Uruçuí. Currículo Lattes: http://lattes.cnpq.br/7426121026369131. Email: teresinhasespi@hotmail.com

Marlei Rosa dos Santos - Engenheira Agrônoma pela Universidade Federal de Viçosa – UFV, com mestrado e Doutorado em Fitotecnia/Produção vegetal, pela Universidade Federal de Viçosa. Professora Associada II na Universidade Estadual do Piauí – UESPI, Campus Cerrado do Alto Parnaíba, Uruçuí-PI, atuando como professora, pesquisadora e extensionista. Responsável para ministrar a disciplina de Estatística Experimental para os alunos do curso de Bacharelado em Agronomia da UESPI, Campus Cerrado do Alto Parnaíba. Currículo Lattes: http://lattes.cnpq.br/9190456250674891. E-mail: marleirosa@urc.uespi.br

Antônia Leidiana Moreira – Graduada em Medicina Veterinária pela Universidade Federal do Piauí, atuou como professora do Curso de Agronomia da Universidade Estadual do Piauí, Campus Cerrado do Alto Parnaíba, Uruçuí-PI. Possui Mestrado e Doutorado em Ciência Animal, pela Universidade Federal do Piauí. Currículo Lattes: http://lattes.cnpq.br/1728120759402521. E-mail: leda.vet@hotmail.com



editora.uespi.br

Editora e Gráfica - UESPI