

Ricardo da Silva Souza

Dados Econômicos

 **Atena**
Editora
Ano 2022

Ricardo da Silva Souza

Dados Econômicos

 **Atena**
Editora
Ano 2022

Editora chefe

Profª Drª Antonella Carvalho de Oliveira

Editora executiva

Natalia Oliveira

Assistente editorial

Flávia Roberta Barão

Bibliotecária

Janaina Ramos

Projeto gráfico

Bruno Oliveira

Camila Alves de Cremo

Luiza Alves Batista

Imagem e Edição de arte

Luiza Alves Batista

2022 by Atena Editora

Copyright © Atena Editora

Copyright do texto © 2022 Os autores

Copyright da edição © 2022 Atena

Editora

Direitos para esta edição cedidos à

Atena Editora pelos autores.

Open access publication by Atena

Editora



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição *Creative Commons*. Atribuição-Não-Comercial-NãoDerivativos 4.0 Internacional (CC BY-NC-ND 4.0).

O conteúdo do texto e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva do autor, inclusive não representam necessariamente a posição oficial da Atena Editora. Permitido o *download* da obra e o compartilhamento desde que sejam atribuídos créditos ao autor, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Todos os manuscritos foram previamente submetidos à avaliação cega pelos pares, membros do Conselho Editorial desta Editora, tendo sido aprovados para a publicação com base em critérios de neutralidade e imparcialidade acadêmica.

A Atena Editora é comprometida em garantir a integridade editorial em todas as etapas do processo de publicação, evitando plágio, dados ou resultados fraudulentos e impedindo que interesses financeiros comprometam os padrões éticos da publicação. Situações suspeitas de má conduta científica serão investigadas sob o mais alto padrão de rigor acadêmico e ético.

Conselho Editorial**Ciências Humanas e Sociais Aplicadas**

Prof. Dr. Adilson Tadeu Basquerote Silva – Universidade para o Desenvolvimento do Alto Vale do Itajaí

Prof. Dr. Alexandre de Freitas Carneiro – Universidade Federal de Rondônia

Prof. Dr. Alexandre Jose Schumacher – Instituto Federal de Educação, Ciência e Tecnologia do Paraná

Prof. Dr. Américo Junior Nunes da Silva – Universidade do Estado da Bahia

Profª Drª Ana Maria Aguiar Frias – Universidade de Évora

Profª Drª Andréa Cristina Marques de Araújo – Universidade Fernando Pessoa

Prof. Dr. Antonio Carlos da Silva – Universidade Católica do Salvador

Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná

Prof. Dr. Antonio Gasparetto Júnior – Instituto Federal do Sudeste de Minas Gerais
Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília
Prof. Dr. Arnaldo Oliveira Souza Júnior – Universidade Federal do Piauí
Prof. Dr. Carlos Antonio de Souza Moraes – Universidade Federal Fluminense
Prof. Dr. Crisóstomo Lima do Nascimento – Universidade Federal Fluminense
Prof^ª Dr^ª Cristina Gaio – Universidade de Lisboa
Prof. Dr. Daniel Richard Sant’Ana – Universidade de Brasília
Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia
Prof^ª Dr^ª Dilma Antunes Silva – Universidade Federal de São Paulo
Prof. Dr. Edvaldo Antunes de Farias – Universidade Estácio de Sá
Prof. Dr. Elson Ferreira Costa – Universidade do Estado do Pará
Prof. Dr. Eloi Martins Senhora – Universidade Federal de Roraima
Prof. Dr. Gustavo Henrique Cepolini Ferreira – Universidade Estadual de Montes Claros
Prof. Dr. Humberto Costa – Universidade Federal do Paraná
Prof^ª Dr^ª Ivone Goulart Lopes – Istituto Internazionale delle Figlie di Maria Ausiliatrice
Prof. Dr. Jadilson Marinho da Silva – Secretaria de Educação de Pernambuco
Prof. Dr. Jadson Correia de Oliveira – Universidade Católica do Salvador
Prof. Dr. José Luis Montesillo-Cedillo – Universidad Autónoma del Estado de México
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Prof. Dr. Kárpio Márcio de Siqueira – Universidade do Estado da Bahia
Prof^ª Dr^ª Keyla Christina Almeida Portela – Instituto Federal do Paraná
Prof^ª Dr^ª Lina Maria Gonçalves – Universidade Federal do Tocantins
Prof^ª Dr^ª Lucicleia Barreto Queiroz – Universidade Federal do Acre
Prof. Dr. Luis Ricardo Fernandes da Costa – Universidade Estadual de Montes Claros
Prof. Dr. Lucio Marques Vieira Souza – Universidade do Estado de Minas Gerais
Prof^ª Dr^ª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof^ª Dr^ª Marianne Sousa Barbosa – Universidade Federal de Campina Grande
Prof. Dr. Marcelo Pereira da Silva – Pontifícia Universidade Católica de Campinas
Prof^ª Dr^ª Maria Luzia da Silva Santana – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Miguel Rodrigues Netto – Universidade do Estado de Mato Grosso
Prof. Dr. Pedro Henrique Máximo Pereira – Universidade Estadual de Goiás
Prof. Dr. Pablo Ricardo de Lima Falcão – Universidade de Pernambuco
Prof^ª Dr^ª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa
Prof^ª Dr^ª Rita de Cássia da Silva Oliveira – Universidade Estadual de Ponta Grossa
Prof. Dr. Rui Maia Diamantino – Universidade Salvador
Prof. Dr. Saulo Cerqueira de Aguiar Soares – Universidade Federal do Piauí
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará
Prof^ª Dr^ª Vanessa Bordin Viera – Universidade Federal de Campina Grande
Prof^ª Dr^ª Vanessa Ribeiro Simon Cavalcanti – Universidade Católica do Salvador
Prof. Dr. William Cleber Domingues Silva – Universidade Federal Rural do Rio de Janeiro
Prof. Dr. William Douglas Guilherme – Universidade Federal do Tocantins

Diagramação: Letícia Alves Vitral
Correção: Maiara Ferreira
Indexação: Amanda Kelly da Costa Veiga
Revisão: O autor
Autor: Ricardo da Silva Souza

Dados Internacionais de Catalogação na Publicação (CIP)	
S729	<p>Souza, Ricardo da Silva Dados econômicos / Ricardo da Silva Souza. – Ponta Grossa - PR: Atena, 2022.</p> <p>Formato: PDF Requisitos de sistema: Adobe Acrobat Reader Modo de acesso: World Wide Web Inclui bibliografia ISBN 978-65-258-0777-5 DOI: https://doi.org/10.22533/at.ed.775221212</p> <p>1. Economia. I. Souza, Ricardo da Silva. II. Título. CDD 330</p>
Elaborado por Bibliotecária Janaina Ramos – CRB-8/9166	

Atena Editora
 Ponta Grossa – Paraná – Brasil
 Telefone: +55 (42) 3323-5493
www.atenaeditora.com.br
contato@atenaeditora.com.br

DECLARAÇÃO DO AUTOR

O autor desta obra: 1. Atesta não possuir qualquer interesse comercial que constitua um conflito de interesses em relação ao conteúdo publicado; 2. Declara que participou ativamente da construção dos respectivos manuscritos, preferencialmente na: a) Concepção do estudo, e/ou aquisição de dados, e/ou análise e interpretação de dados; b) Elaboração do artigo ou revisão com vistas a tornar o material intelectualmente relevante; c) Aprovação final do manuscrito para submissão.; 3. Certifica que o texto publicado está completamente isento de dados e/ou resultados fraudulentos; 4. Confirma a citação e a referência correta de todos os dados e de interpretações de dados de outras pesquisas; 5. Reconhece ter informado todas as fontes de financiamento recebidas para a consecução da pesquisa; 6. Autoriza a edição da obra, que incluem os registros de ficha catalográfica, ISBN, DOI e demais indexadores, projeto visual e criação de capa, diagramação de miolo, assim como lançamento e divulgação da mesma conforme critérios da Atena Editora.

DECLARAÇÃO DA EDITORA

A Atena Editora declara, para os devidos fins de direito, que: 1. A presente publicação constitui apenas transferência temporária dos direitos autorais, direito sobre a publicação, inclusive não constitui responsabilidade solidária na criação dos manuscritos publicados, nos termos previstos na Lei sobre direitos autorais (Lei 9610/98), no art. 184 do Código Penal e no art. 927 do Código Civil; 2. Autoriza e incentiva os autores a assinarem contratos com repositórios institucionais, com fins exclusivos de divulgação da obra, desde que com o devido reconhecimento de autoria e edição e sem qualquer finalidade comercial; 3. Todos os e-book são *open access*, *desta forma* não os comercializa em seu site, sites parceiros, plataformas de *e-commerce*, ou qualquer outro meio virtual ou físico, portanto, está isenta de repasses de direitos autorais aos autores; 4. Todos os membros do conselho editorial são doutores e vinculados a instituições de ensino superior públicas, conforme recomendação da CAPES para obtenção do Qualis livro; 5. Não cede, comercializa ou autoriza a utilização dos nomes e e-mails dos autores, bem como nenhum outro dado dos mesmos, para qualquer finalidade que não o escopo da divulgação desta obra.

A finalidade deste livro é auxiliar aos alunos de Economia e de áreas correlatas a ter um primeiro contato com o tratamento da informação, como um pré-requisito a Estatística Descritiva. Os assuntos nele abordados são expostos de maneira elementar e, resumidamente, simples. Como articulador desta atividade, espero facilitar ou ao menos direcionar o trabalho do professor ao utilizar esta simples coletânea.

Espera-se que os leitores que utilizarão este conteúdo tenham ao menos um breve contato com a abordagem do conteúdo estruturante Tratamento da Informação no Ensino Básico (Ensino Fundamental e Médio), uma vez que o conhecimento prévio na interpretação de gráficos e tabelas potencializa o aprendizado deste conteúdo.

Além disso, para que haja um bom desenvolvimento do conteúdo da estatística descritiva é preferível que o aluno tenha experimentado os conceitos introdutórios da estatística: conceito de variável quantitativa contínua e discreta; e a noção básica de amostragem: amostragem aleatória ao caso, entretanto, este tipo de conhecimento torna-se não-obrigatório a medida que o mediador dessas informações trabalhe em concomitância com o conteúdo curricular.

Esta coletânea visa de maneira primal proporcionar ao aluno das Ciências Sociais Aplicadas (Administração, Atuária, Ciências Contábeis, Ciências Econômicas e Direito) um breve esclarecimento sobre como utilizar as informações quantitativas econômicas para descrever os fenômenos. Também, este livro pode ser útil para os profissionais das áreas para ser um norte para a realização de relatórios e pesquisas de cunho econômico. Também, para os colegas professores, este livro tem a utilidade de complementar o conteúdo de estatística nas disciplinas de Introdução a Estatística, Tratamento da Informação e Introdução a Econometria.

O processamento do texto em LaTeX foi revisado e realizado pela Dra. Camila Leão Cardozo, da Universidade Estadual de Londrina (UEL).

Londrina, agosto de 2022.

Ricardo da Silva Souza

SUMÁRIO

PREFÁCIO.....	2
----------------------	----------

CAPÍTULO 1

1.OS DADOS NAS CIÊNCIAS ECONÔMICAS.....	5
--	----------

1.1. ESTRUTURAÇÃO E TIPOS DE DADOS	5
--	---

1.2. FONTES DE DADOS	6
----------------------------	---

1.3. COLETA DE DADOS.....	7
---------------------------	---

1.4. TRATAMENTO DE DADOS	8
--------------------------------	---

1.5. EXERCÍCIOS REFERENTES AO CAPÍTULO 1	8
--	---

CAPÍTULO 2

O TRATAMENTO DA INFORMAÇÃO NAS CIÊNCIAS ECONÔMICAS: AS REPRESENTAÇÕES TABULAR E GRÁFICA.....	9
---	----------

2.1. VARIÁVEIS ESTATÍSTICAS.....	9
----------------------------------	---

2.2. SÉRIES ESTATÍSTICAS	11
--------------------------------	----

2.2.1. Série Cronológica.....	11
-------------------------------	----

2.2.2. Série Geográfica	11
-------------------------------	----

2.2.3. Série Específica.....	12
------------------------------	----

2.2.4. Série Mista	12
--------------------------	----

2.2.5. Distribuição de Frequências	13
--	----

2.3. REPRESENTAÇÃO TABULAR	13
----------------------------------	----

2.4. REPRESENTAÇÕES GRÁFICAS.....	14
-----------------------------------	----

2.4.1. Gráfico de Colunas	14
---------------------------------	----

2.4.2. Gráfico de Setores Circulares	15
--	----

2.4.3. Gráfico de Linhas	15
--------------------------------	----

2.5. EXERCÍCIOS DE REVISÃO	16
----------------------------------	----

CAPÍTULO 3**O TRATAMENTO DA INFORMAÇÃO NAS CIÊNCIAS ECONÔMICAS: MEDIDAS RESUMO 17**

3.1. MEDIDAS DE POSIÇÃO 17

3.1.1. Média 17

3.1.2. Mediana 18

3.1.3. Moda 18

3.2. MEDIDAS DE DISPERSÃO 19

3.2.1. Amplitude 19

3.2.2. Medidas de Dispersão: Variância e Desvio Padrão 19

3.3. EXERCÍCIOS DE REVISÃO 20

CAPÍTULO 4**OS DADOS E A ECONOMETRIA 21**

4.1. O MODELO DE REGRESSÃO LINEAR 21

4.1.1. Natureza da variável Y 224.1.2. Natureza das variáveis X ou regressores 234.1.3. Natureza do termo de erro estocástico, u 23

4.1.4. Natureza dos coeficientes de regressão 24

4.1.5. O significado de regressão linear 24

4.2. CLASSIFICAÇÃO DOS DADOS EM RELAÇÃO A SUA NATUREZA 24

4.2.1. Dados de séries temporais 24

4.2.2. Dados de corte transversal 25

4.2.3. Dados em painel ou longitudinais 25

4.3. FONTE E QUALIDADE DE DADOS 26

4.4. EXERCÍCIOS DE REVISÃO 26

CAPÍTULO 4**REFERÊNCIAS 28**

1. OS DADOS NAS CIÊNCIAS ECONÔMICAS

As informações obtidas nas pesquisas devem ser analisadas e interpretadas com o auxílio de métodos estatísticos, o que consiste na análise exploratória. Nas Ciências Econômicas, não é diferente, uma vez que se trata de uma ciência que utiliza informações de indivíduos, empresas e do governo.

Sabendo disso, o cientista econômico deverá ter em mente sobre quais informações coletar e analisar para obter uma pesquisa adequada ao seu objeto de estudo e para ganhar consistência nas possíveis hipóteses que a pesquisa pode ser embasada. A qualidade das informações obtidas também é importante, geralmente, fontes confiáveis de informações possibilitam uma pesquisa mais robusta. De modo análogo, uma fonte duvidosa permite uma pesquisa também duvidosa.

Para se realizar uma análise econômica, geralmente, é necessário buscar em fontes especializadas nas temáticas da Economia que são conhecidos como **fontes econômicas**. São localizações de domínio de rede (*internet*) fornecidos por grande parte instituições públicas, e outras, privadas, em que são disponibilizadas as informações. Tais informações, caracterizadas de maneira informal por **dados** e duas maneiras comuns dos economistas sobre utilizar as informações destas informações, se baseiam na **descrição estatística** e na **econometria**.

Nesta seção será apresentada, sequencialmente:

- Tipos de Dados
- Fontes de dados
- Coleta de dados
- Estruturação dos Dados

1.1. ESTRUTURAÇÃO E TIPOS DE DADOS

Existe uma forte corrente científica das Ciências Econômicas que estuda o comportamento das pessoas, das empresas e das nações por meio das informações disponíveis publicamente. Geralmente, estas informações estão construídas com um certo padrão que descreve uma característica tal como o salário de um indivíduo, os custos e receitas de uma empresa e a produção agregada de um país. Tais informações são chamados de dados econômicos.

Entretanto, dependendo da hospedagem do *site* em que estão alocados tais dados, a organização e a estruturação não é a ideal. Muitas vezes, as informações estão

concentradas de forma não regular, com os dados sem uma formatação adequada. Estes tipos de dados são conhecidos como **dados não-estruturados**. De outra forma, esses dados não estão representados por meio de tabelas ou organizados em linhas e colunas de forma uniforme.

Quando os dados estão organizados uniformemente entre linhas e colunas, em uma planilha ou condição semelhante, isto é, formatado em tabelas, este é conhecido como um conjunto de **dados estruturados**.

A estruturação dos dados dependerá do tipo de informações que o pesquisador tem em mãos (ou no computador). Quando o pesquisador possui informações quantitativas que não foram publicadas, principalmente em fontes oficiais, estes são conhecidos como **dados primários**. Geralmente, estas informações são oriundas de pesquisas de questionários que precisam de um refinamento ou por pesquisas em andamento, com dados parciais.

Quando os dados têm uma fonte (de preferência, fidedigna), com grande probabilidade de que estes dados estejam estruturados, com uma data ou período de publicação, são conhecidos como dados secundários. Nas Ciências Econômicas, estes tipos de dados são encontrados em maior frequência, pois há uma enorme fonte de dados disponíveis com facilidade.

1.2. FONTES DE DADOS

Especificamente, nas Ciências Sociais Aplicadas, exclusivamente em Economia, existe uma coleção de informações disponíveis nas bases científicas como *sites* institucionais como o IBGE e o IPEA em âmbito nacional e regionalmente no IPARDES e na Secretaria de Segurança Pública do Paraná (SSP-PR). Muitas vezes, estas informações podem estar organizadas de forma acessível, já estruturados com os dados em planilhas ou em texto de notas que softwares como o Excel e o Biostat e linguagens de programação como o R e Python. Outros bancos já possuem uma formatação específica e é necessário um tratamento ou uma transformação para uma extensão compatível ao software desejável, e como visto, é necessária uma atenção especial para qualificar as informações.

A seguir estão alguns exemplos de fontes de dados econômicos que disponibilizam informações quantitativas úteis para realizar as diversas análises econômicas em várias frentes:

Sigla	Instituição
ALICE- Web	Análise das Informações do Comércio Exterior
ANFAVEA	Associação Nacional de Veículos Automotores
BCB	Banco Central do Brasil
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
IBGE	Instituto Brasileiro de Geografia e Estatística
IPEA	Instituto de Pesquisa Econômica Aplicada
IPARDES	Instituto Paranaense de Pesquisa Desenvolvimento
MAPA	Ministério da Agricultura e Pecuária
RAIS	Relação Anual de Informações Sociais
SEADE- SP	Sistema Estadual de Análise de Dados
SSP-PR	Secretaria de Segurança Pública do Paraná
SUS	Sistema Único de Saúde

Tabela 1 - Algumas Fontes de dados Econômicos existentes no Brasil. Fonte: Elaboração Própria.

Há outras fontes de dados de natureza municipal, estadual e federal que não foram sobrecitadas que facilmente podem ser encontradas na internet. Além disso, as pesquisas em Ciências em Economia podem usufruir de dados originais, isto é, aqueles que não pertencem a uma fonte oficial de dados. Para tanto, faz-se necessário compreender como é realizada a coleta de dados.

1.3. COLETA DE DADOS

A partir do momento em que se objetiva a pesquisa quantitativa a fim de se realizar um estudo quantitativo em Ciências Econômicas, torna-se necessário compreender de qual forma esses dados serão coletados pelo pesquisador.

Se os dados são de natureza secundária, a coleta ocorre geralmente na navegação dos *sites* das fontes citadas na seção 1.2, por meio da rede mundial de computadores, a *internet*. A coleta geralmente ocorre fazendo o *download* de arquivos que estão associadas a tais fontes, a partir do direcionamento da própria página.

Nos dados primários, as duas formas mais usuais para coletar as informações são os questionários aplicados pelo pesquisador e a coletânea de informações de uma fonte que não foram revelados ao público. Para a segunda opção é necessário entrar em contato com o instituto de referência e solicitar a permissão legal para explorar os dados e transmitir as informações necessárias para a pesquisa.

Para o caso da obtenção de dados por meio de um questionário, o autor deverá elaborar um documento com um número limitado de questões, em que cada questão proporcionará uma resposta ou direcionamento para a pesquisa que está sendo realizada.

1.4. TRATAMENTO DE DADOS

Segundo uma das principais referências do ensino de estatística do ensino superior de Wilton Oliveira Bussab e Pedro Alberto Morettin (2012), a definição de **dados** são informações inerentes às variáveis que caracterizam os elementos que constituem a população ou amostra em estudo.

Quando se obtém os dados diretamente da pesquisa, sem nenhum tratamento, são classificados de **dados brutos** e quando o pesquisador realiza a organização dessas informações, ou seja, um tratamento, chamamos de **dados tratados**.

Para realizar um tratamento de dados de forma que as informações estejam razoavelmente apresentáveis ao público por meio de uma pesquisa é necessário ter em mente um processo organizacional, isto é, uma forma racional e funcional de proporcionar as informações em que as pessoas possam utilizar facilmente. Uma compilação adequada das informações permite também a interação com outras fontes de dados, uma vez que a natureza seja compatível. As ciências estatísticas proporcionam dois tipos de representações que descrevem as informações contidas nos dados de forma organizada. São conhecidas como representação tabular e gráfica, além de informações resumidas sobre os dados, que são as medidas resumo.

1.5. EXERCÍCIOS REFERENTES AO CAPÍTULO 1

1 - Por que os dados são importantes para as Ciências Econômicas?

2 - Qual é a diferença entre dados estruturados e não-estruturados?

3 - O que são dados primários? Dê um exemplo da obtenção de informações oriundas de dados primários.

4 - O que são dados secundários? Dê um exemplo da obtenção de informações oriundas de dados secundários.

5 - Cite pelo menos três instituições que possuem informações econômicas de natureza quantitativa.

6 - Como se realiza a coleta de dados de fonte primária? Dê exemplos.

7 - Como se realiza a coleta de dados de fonte secundária? Dê exemplos

8 - Diferencie dados brutos de dados tratados.

9 - Um questionário é uma ótima ferramenta para a coleta de dados primários. Uma excelente referência para elaborar questionários é livro da pesquisadora Sônia Vieira (2009). Elabore um esboço de uma pesquisa que utilize como ferramenta um questionário. Use como referência o livro citado para ajustar seu esboço.

O TRATAMENTO DA INFORMAÇÃO NAS CIÊNCIAS ECONÔMICAS: AS REPRESENTAÇÕES TABULAR E GRÁFICA

O tratamento da informação é a área da matemática que engloba as maneiras de se obter, organizar e analisar os dados para realizar conclusões e previsões sobre tais informações. Assim, é necessário entender ao menos a estrutura básica de uma coleção de informações em relação a sua natureza e a sua forma para obtê-las de tal forma que a sua organização seja realizada de forma eficiente para que ocorra uma análise também eficiente.

Com o avanço das Ciências Estatísticas, existem uma grande coleção de análises de dados que podem ser realizadas com um conjunto de informações. Basicamente, as análises na estatística são divididas em três grandes áreas: a estatística descritiva, a teoria das probabilidades e a inferência estatística. Este livro, de maneira sucinta, tem seu direcionamento a estatística descritiva, mas com condições de abordagem nas outras áreas.

Sendo assim, os dados são muito importantes para o tratamento da informação, principalmente para descrever as variáveis contidas em um certo conjunto de dados. Para este capítulo, a organização se associa a identificação dos dados em relação as suas características, mais conhecido por variáveis estatísticas. Após, entendido o conceito de variável é importante associar a natureza em que a variável está construída em relação aos dados (e que podem ser generalizados para quaisquer conjuntos) em relação as configurações das suas características, conhecidas como séries estatísticas. Posteriormente, as representações tabulares e gráficas de um conjunto de dados.

Este capítulo é muito importante para o avanço do uso dos dados econômicos por parte de um estudante ou pesquisador, uma vez que o tratamento das informações quantitativas permite qualificar uma pesquisa ou trabalho para além do escopo acadêmico ou profissional. Uma adequada apresentação dos dados por meio de tabelas e gráficos proporciona o acesso a informação para um número maior de pessoas que em condições normais não teriam condições de analisar os dados na sua forma original.

2.1. VARIÁVEIS ESTATÍSTICAS

Em matemática, variável é um símbolo que representa certa quantidade desconhecida, que passa por certas quantidades de grandeza. Em estatística, variável representa um resultado gerado por determinada característica ou fenômeno.

Por exemplo: um pesquisador deseja realizar uma análise do perfil de funcionários

de uma empresa de logística em relação a seis informações pessoais: Sexo, escolaridade, estado civil, idade, salário mensal e número de filhos.

Essas informações que determinam as características dos indivíduos são conhecidas como variáveis estatísticas. Para classificar uma variável estatística, utiliza-se uma letra maiúscula (A,B,C,...,X,Y,Z). De maneira bem simples, pode-se atribuir a variável sexo a letra S, escolaridade a letra E, estado civil que também começa com a letra E, podemos chamar de C, salário mensal a letra W, a idade por I e experiência, X.

De forma resumida, é possível listar as variáveis estatísticas em relação a sua classificação:

- C = Estado Civil
- E = Escolaridade
- I = Idade
- S = Sexo
- W = Salário
- X = Experiência

Deste conjunto de dados é possível classificá-las a partir das características da variável associada ao indivíduo. Geralmente, as variáveis que associam o indivíduo a um atributo ou qualidade são descritas como variáveis qualitativas.

Por outro lado, variáveis estatísticas classificadas como uma medida ou contagem são descritas como variáveis quantitativas. Para este conjunto de variáveis descritas pelo pesquisador temos que as variáveis estado civil, escolaridade e sexo são qualitativas, enquanto as variáveis escolaridade, salário e experiência são variáveis quantitativas.

Dentro de cada classificação de variável, também é possível fazer distinções em relação aos tipos: Na classificação das variáveis qualitativas, as variáveis estado civil e sexo não são possíveis realizar uma ordenação ou hierarquia em relação a um tipo em relação ao outro, apenas distinguir pelo seu nome. Este tipo de variável é conhecido como variável qualitativa nominal. A variável escolaridade é possível organizá-las em relação ao nível: Por exemplo, nível fundamental, médio e superior. A partir desta hierarquia é possível ordenar. Este tipo de variável é chamado de variável qualitativa ordinal.

No conjunto das variáveis quantitativas, considerando a idade e a experiência medida em anos completos, este é, uma medida de um número inteiro, especificamente, inteiro positivo. Para este tipo é possível classificar esta variável como variável quantitativa discreta. Se considerarmos a variável salário em reais, não considerando seus submúltiplos (os centavos) este também pode ser considerado uma variável quantitativa discreta.

Por outro lado, se considera a variável salário os seus submúltiplos e se considerar as partes de anos da idade do indivíduo, estes não serão números inteiros, mas sim, números racionais positivos. Para esta natureza de conjunto de dados, o tipo desta variável é chamado de variável quantitativa contínua.

Portanto é crucial para os pesquisadores classificar e identificar as variáveis estatísticas e os seus tipos para facilitar a realização das codificações necessárias e adequadas para tabular os dados e utilizá-los de forma eficiente por meio de uma planilha ou *software*.

2.2. SÉRIES ESTATÍSTICAS

Existem várias maneiras de representar um conjunto de dados por meio de tabelas, desde inserir a uma quantidade de dados sem nenhum tipo de tratamento até a sua delimitação e resumo para proporcionar as informações com uma maior densidade. Para tanto, a dependência, dos tipos de dados e da classificação que se queira dar.

Fixando uma das três modalidades que caracterizam um fato: tempo, local e fato observado, as séries se classificam em cinco tipos: Série Cronológica ou Temporal, Série Geográfica, Série Específica, Série Conjugada ou Mista e a Distribuição de Frequências. Segue os exemplos de cada tipo de série

2.2.1. Série Cronológica

A série cronológica varia o tempo e fixa o local e o fato. Geralmente em economia, as séries cronológicas têm como nome as séries temporais. Muitos dados macroeconômicos como a inflação, desemprego e o PIB de um país possuem a série cronológica como objeto de estudo.

Ano	Inflação (%)
2014	6,4
2015	10,6
2016	6,3
2017	2,9
2018	3,7
2019	4,3
2020	4,5

Tabela 2 - Inflação no Brasil entre 2014 a 2020, em percentual. Fonte: IBGE (2021).

2.2.2. Série Geográfica

Varia o local e fixa o tempo e o fato. Dentro das Ciências Econômicas existe uma coleção de fenômenos que analisam uma série geográfica tais como a produção de cada

município ou região, bem como o desempenho escolar e o nível de emprego. Existe uma grande área em franco desenvolvimento no campo da Economia conhecido como Economia Regional que utilizam massivamente este tipo de série.

Município	População
Curitiba	1.948.626
Londrina	575.377
Maringá	430.157
Ponta Grossa	355.356

Tabela 3 - População dos quatro maiores municípios do Paraná em 2021. Fonte: IBGE (2022).

2.2.3. Série Específica

Varia o fato e fixa o tempo e o local. Nas Ciências Econômicas, uma série específica é analisada com certa frequência em análises socioeconômicas em relação a um grupo de uma variável específica como sexo e cor e em relação as estruturas de mercado e a organização industrial..

Empresa	Percentual
Lacta	33
Nestlé	32
Neugbauer	21
Outros	14

Tabela 4 - Concentração de Mercado brasileiro de Chocolate em 2021. Fonte: ACB (2021).

2.2.4. Série Mista

Combinação de duas séries já citadas. A série mista entre uma série temporal e uma série geográfica ou específica é muito utilizada nas Ciências Econômicas em relação a análises de dados em painel em que leva em consideração os efeitos do tempo e do espaço.

Meses	Índice	
	Mensal	Anual
Agosto	0,65	5,29
Setembro	0,54	5,85
Outubro	0,61	6,50
Novembro	0,56	7,10
Dezembro	0,56	7,10

Tabela 5 - Índice percentual do rendimento da poupança no Brasil no ano de 2020. Fonte: Bacen (2020).

2.2.5. Distribuição de Frequências

A distribuição de frequências é um tipo de série que conta as determinadas características de uma amostra. As frequências podem ser definidas como frequência absoluta ou contagem de uma mesma categoria e frequência relativa que é a parte do todo. A distribuição de frequências por vezes pode se confundir com a série específica.

Sexo	Freq. Absoluta	Freq. Relativa
Feminino	12	60
Setembro	8	40
Total	20	100

Tabela 6 - Distribuição de frequências em relação ao sexo, de uma sala de aula da UENP, em 2022.
Fonte: Elaboração pelo autor.

2.3. REPRESENTAÇÃO TABULAR

A representação tabular é uma descrição dos dados em forma quantitativa, por meio das contagens de uma determinada característica em relação ao todo ou em relação a série na qual a variável estatística tem por natureza. Como visto na subseção das séries, se a contagem é representada por meio de uma tabela no qual estão contidas as informações absolutas, relacionadas a contagem dos elementos conhecidos como **frequências absolutas** e as informações em percentual, relativas, conhecidas como **frequências relativas**. Porém, há tabelas que contém informações das séries dos fenômenos analisados. Na seção de séries estatísticas, todas as séries citadas foram mostradas de forma sistemática por meio de tabelas.

As tabelas, pelas normas da ABNT, têm sua construção padronizada por linhas horizontais e não fechadas por linhas verticais. Entre as linhas horizontais superiores, está a descrição das variáveis ou o cabeçalho. Entre as linhas superior e inferior está o corpo e no rodapé, é citado a fonte.

A seguir, alguns exemplos de tabelas utilizadas para descrever dados das Ciências Econômicas

Ano	PIB per capita
2010	16050
2011	18815
2012	21367
2013	24667
2014	25618
2015	28339
2016	28944
2017	29749
2018	32330
2019	35291

Tabela 7 - PIB per capita em reais de Cornélio Procópio entre 2010 e 2019. Fonte: Elaborado a partir das informações do IBGE (2022).

Informações sobre a variável PIB de Cornélio Procópio:

- Também é uma variável quantitativa
- É uma variável contínua.
- São dados de corte seccional.

Para entender ainda mais sobre esta variável econômica em relação as suas informações, deve-se utilizar representações gráficas e medidas resumo (medidas de posição e dispersão).

2.4. REPRESENTAÇÕES GRÁFICAS

Há diversas formas de representar as informações sobre um conjunto de dados. Esta riqueza de representações permite ao pesquisador explorar as informações de forma variada. Entretanto, resumidamente, serão apresentados os três tipos de representações gráficas mais utilizados nas Ciências Econômicas: O gráfico de colunas, o de setor circular (popularmente conhecido como gráfico de pizza) e de linha.

2.4.1. Gráfico de Colunas

O gráfico de colunas é um dos tipos gráficos mais utilizados em estatísticas pela versatilidade, facilidade de construção e de interpretação. Ele é ideal para ser utilizada para descrever fenômenos cuja variável é qualitativa. Desta forma este gráfico é adequado para contagens, ou seja, as frequências absolutas. Porém, se especificado de forma correta uma variável quantitativa, este gráfico também pode ser utilizado, conforme observado na Figura 1.

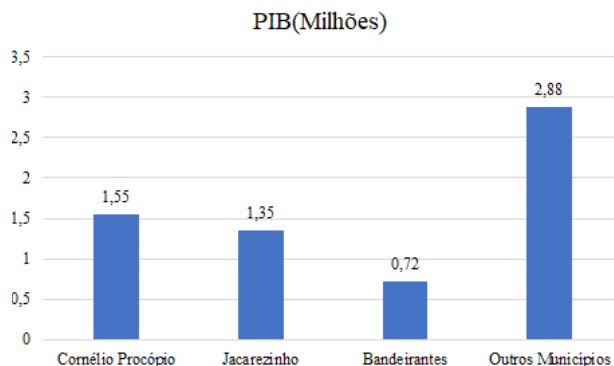


Figura 1- Representação Gráfica do PIB agregado do Norte Pioneiro, por município em 2019.

2.4.2. Gráfico de Setores Circulares

O gráfico de setor circular é uma representação versátil e que manualmente não é de fácil construção uma vez que se deve saber os graus dos respectivos setores. Mas as planilhas eletrônicas e os *softwares* estatísticos realizam a confecção deste tipo de gráfico com eficiência e exatidão. Esta representação gráfica é ideal para saber a parte referencial do todo. Ou seja é ideal para medidas relativas (frequência relativa), em porcentagem. Por exemplo, o Produto Interno Bruto da Região Norte Pioneiro pode ser representado em relação a sua participação relativa, conforme pode ser visto na Figura 2.

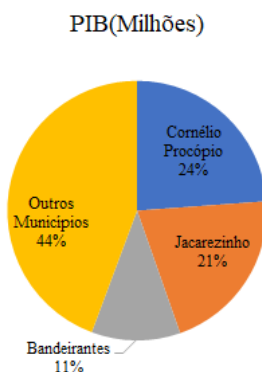


Figura 2 - Representação Gráfica da participação relativa do PIB na Região Norte Pioneiro em 2019.

2.4.3. Gráfico de Linhas

O gráfico de linhas é uma representação adequada e de fácil construção. Intuitivamente, tem-se no eixo horizontal a variável tempo como independente e a variável de referência dependente do tempo. Assim, esta representação gráfica é ideal para dados cujas séries são temporais ou cronológicas

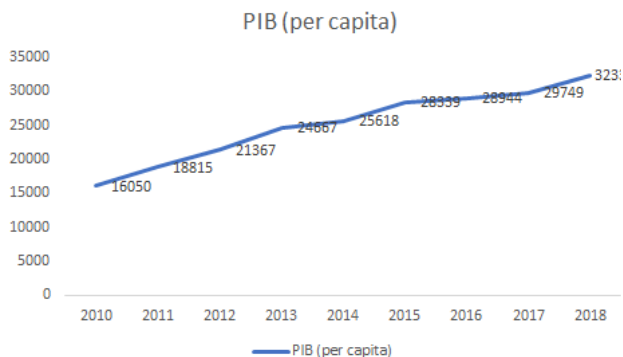


Figura 3 - Representação Gráfica do PIB per capita de Cornélio Procópio entre 2010 a 2018.

Para a ilustração gráfica, utilizou-se apenas informações quantitativas de uma cidade, de Cornélio Procópio. É possível realizar a análise para mais de um município dentro de um mesmo gráfico de linhas. Para isso, o pesquisador deverá tomar cuidado com a escala da medida de referência a ser empregada no eixo vertical. Por exemplo do PIB, um município que possui um PIB per capita muito baixo, apenas em centenas de reais, por exemplo, terá a visualização de suas linhas comprometida.

2.5. EXERCÍCIOS DE REVISÃO

1 - Para cada informação quantitativa econômica, determine a melhor representação gráfica.

- a) Desemprego no Brasil nos últimos 10 anos.
- b) Inflação dos cinco principais alimentos que fazem parte da cesta básica do Brasil, em 2020.
- c) População brasileira em relação ao número de pessoas economicamente ativas, por sexo, em 2014 e em 2015.

2 - A região Metropolitana de Londrina é um pólo consolidado no Paraná e no Brasil, principalmente, quando se trata em educação. Em relação a evasão escolar, a média da taxa percentual de evasão é menor que o Brasil, segundo o IBGE em 2019. Na mesma pesquisa, a desvio-padrão da RML relacionado a taxa durante toda a década de 2010 é menor que quando comparada ao Brasil. Sabendo destas informações, o que se pode concluir da taxa de evasão de Londrina em relação ao Brasil?

O TRATAMENTO DA INFORMAÇÃO NAS CIÊNCIAS ECONÔMICAS: MEDIDAS RESUMO

Vimos que a representação de dados por meios gráficos e tabelas fornece informações úteis. Mas, para complementar as informações existentes de uma variável é necessárias medidas numéricas que qualificam uma análise. Essas medidas são as medidas **resumo**. Nesta seção serão abordadas:

- Medidas de posição: Média, Mediana e Moda.
- Medidas de dispersão: Amplitude, Variância e Desvio-Padrão.

Para exemplificação, será utilizada as informações em uma área muito importante em economia: os dados sobre o mercado acionário. Serão utilizados os preços das ações da Companhia Paranaense de Energia Elétrica (COPEL) e o Serviço da Água e Esgoto do Paraná (SANEPAR), entre os dias 12 a 21 de janeiro de 2021, da base de dados da Bolsa de Valores de São Paulo, a IBOVESPA (Tabela 8).

Data	CPL6	SAPR4
12/01/21	67,07	4,65
13/01/21	65,40	4,70
14/01/21	67,05	4,70
15/01/21	66,01	4,73
18/01/21	65,26	4,67
19/01/21	64,99	4,64
20/01/21	65,13	4,51
21/01/21	66,28	4,38
22/01/21	65,31	4,30

Tabela 8 - Preço de fechamento da ação CPL6 (COPEL) e SAPR4(SANEPAR) entre 12/01/21 a 22/01/21. Fonte: Elaborado a partir dos dados da IBOVESPA.

3.1. MEDIDAS DE POSIÇÃO

São também chamadas de medidas de **tendência central** porque representam os fenômenos pelos valores médios, em torno dos quais tendem a concentrar seus valores. A análise de tais medidas se encontra na média aritmética simples, mediana e moda.

3.1.1. Média

Suponha um conjunto de dados que pode ser amostral ou populacional com de n elementos x_1, x_2, \dots, x_n . A **média aritmética simples** desses elementos é a soma das observações dividida pela quantidade de observações e é representado por:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^k x_i}{n}$$

Do exemplo Tabela 8, o número de observações sobre uma amostra de 9 dias (n=9), a média aritmética da ação CPLE6 será:

$$\bar{x} = \frac{67,07 + 65,40 + \dots + 66,28 + 65,31}{9} = \frac{592,5}{9} = 65,83$$

Ou seja, o preço médio de fechamento da ação da COPEL (CPLE6) é de R\$ 65,83. É sabido que a média é a medida de tendência central mais utilizada. De outra forma, este é o valor que quando comparado as outras observações, terá menor erro de previsão. Entretanto, será que este valor realmente é uma medida que está justamente no centro da distribuição, uma vez organizada de forma crescente ou decrescente?

3.1.2. Mediana

A mediana é outra medida de tendência central de uma variável. A mediana é o valor que fica no meio de uma sequência quando os dados são arranjados na ordem ascendente. Com um número ímpar de observações, a mediana é o valor que justamente separa em dois subconjuntos de mesma quantidade, de outra forma, o valor do meio. De maneira simples podemos encontrar a mediana a partir do conhecimento da cardinalidade do conjunto ou o número de elementos. Se um conjunto possui um número par de observações não tem valor único como mediana. Neste caso, seguimos a convenção de definir a mediana como sendo a média dos valores das duas observações do meio.

Para as ações da COPEL CPFL6, escrevendo-as em ordem crescente temos a seguinte configuração:

$$= \{64,99; 65,13; 65,26; 65,31; \mathbf{65,4}; 66,01; 66,28; 67,05; 67,07\}$$

A mediana para este conjunto de ações é R\$ 65,40 para estes conjuntos de dados. Agora, suponha um incremento de observação no dia 23 de janeiro de 2021 cujo valor é igual R\$ 66,19. Assim, o número de observações passa a ser dez preços de fechamento:

$$\{64,99; 65,13; 65,26; 65,31; \mathbf{65,40}; \mathbf{66,01}; 66,1; 66,28; 67,05; 67,07\}$$

3.1.3. Moda

A moda é o valor de dados que ocorre com maior frequência. Para ilustrar a identificação da moda, considere ainda o exemplo dos juros das ações da empresa SANEPAR (SAPR4), ordenado de forma ascendente:

Exemplo da ação SAPR4:

$$= \{4,30; 4,38; 4,51; 4,64; 4,65; 4,67; \mathbf{4,70}; \\ \mathbf{4,70}; 4,73\}$$

A moda para este conjunto de ações é 4,70.

3.2. MEDIDAS DE DISPERSÃO

Além das medidas de posição, frequentemente é desejável considerar as medidas de variabilidade ou de dispersão, utilizadas para avaliar o grau de variabilidade ou dispersão ou espalhamento, dos valores em torno da média. Nesta seção estudaremos:

- Amplitude
- Variância
- Desvio-Padrão

3.2.1. Amplitude

Esta é a medida de dispersão mais simples. É dada por:

$$\text{Amplitude} = \text{Valor Máximo} - \text{Valor Mínimo}$$

Para a ação CPFL6, o valor mínimo R\$64,99 e máximo R\$67,07. Logo:

$$\text{Amplitude} = 67,07 - 64,99 = 2,08$$

Embora a amplitude seja a medida da variabilidade mais fácil de calcular, raramente é usada como medida. A razão é que a amplitude é baseada em apenas duas observações e, portanto, é altamente influenciada pelos valores extremos.

3.2.2. Medidas de Dispersão: Variância e Desvio Padrão

A variância e o desvio-padrão são as medidas de dispersão normalmente mais aplicadas. São medidas que se relacionam. Já que a variância é o desvio-padrão ao quadrado. A variância considera a posição de cada observação em relação ao valor médio do conjunto de dados, e define-se como a média.

A variância é calculada pelas seguintes expressões:

$$s^2 = \frac{\sum_{i=1}^k x_i^2 - \frac{(\sum_{i=1}^k x_i)^2}{n}}{n-1} \quad \text{ou} \quad \sigma^2 = \frac{\sum_{i=1}^k x_i^2 - \frac{(\sum_{i=1}^k x_i)^2}{N}}{N}$$

O que difere os cálculos de s^2 e σ^2 ?

- s^2 é o estimador da variância amostral
- σ^2 é o estimador populacional

Para o desvio padrão, basta tomar a raiz quadrada da variância, isto é:

$$s = \sqrt{s^2} \text{ ou } \sigma = \sqrt{\sigma^2}$$

Para os dados das ações CPFL6 e SAPR4, temos:

$$s^2(\text{CPFL6}) = 0,66 \text{ e } s^n(\text{CPFL6}) = 0,81.$$

$$s^2(\text{SAPR4}) = 0,02 \text{ e } s^n(\text{CPFL6}) = 0,15.$$

A variância e o desvio-padrão da ação da COPEL é maior que a da SANEPAR. O que isso significa economicamente?

3.3. EXERCÍCIOS DE REVISÃO

- 1 - Calcular a média, a mediana e a moda da ação SAPR4 (SANEPAR).
- 2 - Calcular as medidas resumo das informações contidas na Tabela 2, da seção 2.2.
- 3 - Calcular as medidas resumo das informações contidas na Tabela 4, da seção 2.2.
- 4 - Calcular as medidas resumo das informações contidas na Tabela 6, da seção 2.3.

OS DADOS E A ECONOMETRIA

Para as Ciências Econômicas, os dados são importantes para realizar as pesquisas em diversos campos: Finanças, Microeconomia, Macroeconomia e o Desenvolvimento. É um consenso para os pesquisadores dessas áreas o uso de técnicas que utilizam simultaneamente a Teoria Econômica e os métodos estatísticos, principalmente que permitem associar a uma sistematização de um modelo econômico.

A técnica que permite esta união é conhecida como Econometria, que como o livro seminal de Damodar Gujarati define como a medida da economia. Outra definição comumente utilizada é a técnica de análise de dados para entender a relação entre as variáveis econômicas.

4.1. O MODELO DE REGRESSÃO LINEAR

A forma geral do modelo de regressão linear é descrito por:

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \dots + \alpha_n X_{ni} + u_i$$

A variável Y é conhecida como variável dependente, ou **regressando**, e as variáveis X são conhecidas como variáveis independentes, explanatórias, ou preditoras covariáveis ou **regressores** e u são conhecidos como termo de erro aleatório ou estocástico. O subscrito i denota a i -ésima observação.

De forma resumida, podemos escrever como $Y_i = \alpha X_n + u$. A equação é conhecida como modelo populacional ou verdadeiro. Ela é composta por dois componentes: (1) um componente determinístico α_x e um componente aleatório u .

O componente α pode ser interpretado como a **média condicional** de Y_i , isto é: $E[Y_i | X]$. Ou seja, Y condicionada aos valores de X .

Portanto, a equação $Y_i = \alpha X_n + u$ afirma que um valor individual de Y_i equivale ao valor médio da população ao qual ele ou ela é um integrante mais ou menos um termo aleatório.

Um exemplo muito adequado para exemplificar a relação entre a variável dependente e a variável independente é poderosamente adequada está no livro do famoso autor Damodar Gujarati (2019) que relaciona a relação econômica entre os gastos e a renda familiar: Se Y representar gastos familiares com alimentação e X representar a renda familiar, a equação $Y_i = \alpha X_n + u$ indica que os gastos médios com alimentação de uma família equivalem aos gastos médios com alimentação de todas as famílias com o mesmo nível de renda, mais ou menos um componente aleatório que pode variar de um indivíduo

ao outro e que pode depender de vários fatores".

A representação α_1 é conhecido como intercepto e o conjunto de representações $\alpha_2, \dots, \alpha_n$ são conhecidos como os coeficientes angulares. Juntos, são chamados de **coeficientes de regressão** ou **parâmetros de regressão**.

Na análise de regressão o **objetivo** é explicar o comportamento médio de Y em relação aos regressores, ou seja, como o Y médio reage às alterações nos valores das variáveis X .

OBS: É importante ter consciência de que a relação entre Y e $X_1 \dots X_n$, se houver, é a baseada na teoria relevante.

Cada coeficiente angular mede a taxa de variação parcial no valor médio de Y para uma unidade de variação no valor de um regressor. O número de regressores incluídos no modelo depende da natureza do problema e varia de problema para problema.

O termo de erro u é um termo genérico para se referir a todas as variáveis que não podem ser introduzidas no modelo por uma infinidade de razões. No entanto, presume-se que a influência média dessas variáveis sobre o regressando seja insignificante.

4.1.1. Natureza da variável Y

Em geral, presume-se que Y é uma variável aleatória que pode ser medida em quatro escalas diferentes: **escala de razão, escala de intervalo, escala ordinal e escala nominal**.

Escala de razão: A maioria das variáveis econômicas pertence a esta categoria. Uma variável de uma escala de razão apresenta três propriedades:

Razão entre variáveis: Y_2/Y_1

Distância entre variáveis: $Y_2 - Y_1$

Ordenação entre variáveis: $Y_2 > Y_1$

Exemplo: PIB da Região Sudeste do Brasil.

Estado	PIB (milhões de R\$) 2017
Espírito Santo	113.552
Minas Gerais	576.799
Rio de Janeiro	671.362
São Paulo	2.119.854

Escala de intervalo: As variáveis desta categoria não satisfazem a propriedade (1), isto é, por exemplo: Suponha os anos 2000 e 2007, a distância entre 2007 e 2000 é 7 anos, mas não faz sentido a razão 2007/2000.

Exemplo: Período medido em anos.

Período	Ano
1	2018
2	2019
3	2020
4	2021

Escala Ordinal: As variáveis dessa escala apenas satisfazem a propriedade (3). Por exemplo, sistemas de avaliação, como A, B, C são variáveis de escala ordinal.

Exemplo: Notas médias dos últimos quatro bimestres dos alunos do quarto ano de Economia.

Aluno	Ano
A	8
B	7
C	6
D	7

Escala Nominal: As variáveis dessa escala não obedecem a nenhuma das três propriedades. Por exemplo, sexo do indivíduo: masculino ou feminino. Tais variáveis são conhecidas como variáveis binárias ou categóricas.

Exemplo: Os últimos jogadores de futebol do planeta a serem escolhidos como Bola de Ouro.

Ano	Jogador
2018	Luka Modric (CRO)
2019	Lionel Messi (ARG)
2020	Robert Lewandowski (POL)
2021	Lionel Messi (ARG)

4.1.2. Natureza das variáveis X ou regressores

As variáveis independentes também podem ser medidas por qualquer uma das escalas já estudadas. No **modelo clássico de regressão linear** (MCRL), presume-se que os regressores sejam não aleatórios, no sentido que os valores são fixos em uma amostra repetida. Como resultado, a análise de regressão é condicional, isto é, depende dos valores dados dos regressores.

4.1.3. Natureza do termo de erro estocástico, u

O **termo de erro estocástico** é um termo genérico que inclui todas aquelas variáveis que não podem ser facilmente quantificadas. Pode representar variáveis que podem ou

não ser incluídas no modelo por falta de disponibilização dos dados, erros de medida ou pela aleatoriedade intrínseca do comportamento humano. Seja qual for a fonte do termo aleatório u , presume-se que o efeito médio do termo de erro sobre o regressando é no máximo "marginal".

4.1.4. Natureza dos coeficientes de regressão

No modelo clássico de regressão linear presume-se que os coeficientes de regressão sejam alguns números fixos e não aleatórios desconhecendo seus valores reais. É o objetivo de análise de regressão estimar seus valores com base nos dados da amostra.

4.1.5. O significado de regressão linear

O termo "linear" no modelo de regressão linear refere-se à **linearidade** dos coeficientes de regressão, o conjunto a_2, \dots, a_n , e não necessariamente, as linearidades das variáveis Y e X . Por exemplo, as variáveis Y e X podem ser logarítmicas ($\ln X_i$) ou recíprocas ($1/X_i$) ou elevadas a uma potência X_i^2 .

4.2. CLASSIFICAÇÃO DOS DADOS EM RELAÇÃO A SUA NATUREZA

Para realizar a análise de regressão, é necessário a utilização dos dados. Para inferir em relação as ciências econômicas, muitas informações são oriundas da Teoria Econômica. Os dados disponíveis para análise, em geral, podem ser classificados em três tipos:

- Dados de séries temporais
- Dados de corte transversal
- Dados empilhados (pooled) ou dados em painel.

4.2.1. Dados de séries temporais

Uma série temporal é um conjunto de observações de uma variável em períodos diferentes (dias, meses, anos). Exemplo: PIB de um país. Em período especificamente curto esse tipo de dados é chamado de dados de alta frequência. O exemplo mais comum são os dados dos mercados de ações e a produção de uma *commodity*, como a soja (Tabela 9) em um determinado período.

Ano	Produção
2016	110
2017	112
2018	117
2019	119

Tabela 9 – Produção de Soja, em milhões de toneladas, entre 2016 e 2019, no Brasil. Fonte: Elaboração do autor a partir de informações do MAPA.

4.2.2. Dados de corte transversal

Os dados de corte transversal são os dados de uma ou mais variáveis coletados no mesmo ponto do tempo. São alguns exemplos: o censo populacional realizado pelo IBGE, as pesquisas de opinião realizadas pelo IBOPE.

Do mesmo modo que os dados de séries temporais, os dados de corte transversal apresentam problemas específicos, especialmente o problema de heterogeneidade, como a análise da produção de soja por estado (Tabela 10).

Estado	Produção
Mato Grosso	35
Paraná	21
Rio Grande do Sul	13
Goiás	12

Tabela 10 – Produção de Soja, em milhões de toneladas, entre os estados do Brasil. Fonte: Elaboração do autor a partir das informações do MAPA.

4.2.3. Dados em painel ou longitudinais

Os dados em painel é a combinação dos dados de corte transversal e dados de séries temporais. Por exemplo, para estimar uma função produção, podemos ter dados de várias empresas em vários períodos, as observações em painel serão denotadas pelo duplo subscrito *it*. Para exemplificação, foi realizada a mescla das Tabelas 9 e 10 (Tabela 11):

Estado	MT	PR	RS	GO
Ano	Produção			
2017	32	19	17	11
2018	32	16	19	11
2019	35	21	13	12

Tabela 11 – Produção de Soja, em milhões de toneladas, entre os estados do Brasil, entre 2017 a 2019. Fonte: Elaboração do autor a partir das informações do MAPA.

4.3. FONTE E QUALIDADE DE DADOS

O sucesso de qualquer análise de regressão depende da disponibilidade de dados. Há dados coletados por um órgão do governo (IBGE, IPEADATA) ou órgãos internacionais como FMI e dados coletados por entidades privadas. Nos dias de hoje, o acesso à Internet facilitou a busca.

O fato de ser possível encontrar os dados em vários lugares não significa que os dados são bons. É preciso verificar com muito cuidado a qualidade do órgão que coletou estes dados, visto que haja erros de medida, omissão de dados. Damodar Gujarati no seu seminal livro de Econometria de 2019, em destaque diz que "Os pesquisadores devem ter em mente que a qualidade dos dados é importante de tal forma a pesquisa fica seriamente comprometida se a base de dados é duvidosa".

4.4. EXERCÍCIOS DE REVISÃO

1. Seja o fictício banco de dados de um conjunto de variáveis econômicas.

Y	X ₁	X ₂	X ₃	X ₄
20	Superior	1000,00	Feminino	2000
21	Médio	1230,18	Masculino	1999
25	Fundamental	814,20	Feminino	1995
28	Médio	760,00	Masculino	1992
31	Superior	1354,97	Feminino	1989

a) Identifique as variáveis

b) Classifique as Escalas de Variáveis

2. Seja o fictício banco de dados de um conjunto de variáveis econômicas. Classifique os bancos de dados a seguir em séries temporais, corte transversal e dados em painel.

a)

t	X ₁	X ₂
1	X ₁₁	X ₂₁
2	X ₁₂	X ₂₂
3	X ₁₃	X ₂₃

b)

t	X _t
1	X ₁
2	X ₂
3	X ₃

c)

i	X_i
1	X_1
2	X_2
3	X_3

REFERÊNCIAS

BUSSAB, W. de O; MORETTIN, P. Estatística Básica. 8ª ed. São Paulo: Saraiva, 2016.

BRUNI, A. L. Estatística Aplicada à Gestão Empresarial. 2º ed. São Paulo: Atlas, 2008.

CASTANHEIRA, N. P. Estatística Aplicada a Todos os Níveis. 5ª ed. São Paulo: IBPEX, 2010.

CRESPINO, A. A. Estatística Fácil. 19ª ed. São Paulo: Saraiva, 2009.

DOWNING, D.; CLARK, J. Estatística Aplicada. 2ª ed. São Paulo: Saraiva, 2006.

FONSECA, J. S. da, et al. Curso de Estatística. 6ª ed. São Paulo: Atlas, 2000.

GUJARATI, D. N. Econometria Básica. 5ª ed. Rio de Janeiro: Elsevier, 2011.

GUJARATI, D. N. Econometria: Princípios, Teoria e Aplicações Práticas. 1ª ed. São Paulo: Saraiva, 2019.

HOFFMANN, R. Estatística para Economistas. 4º ed. São Paulo: Pioneira, 2006.

LEVINE, D. M., et al. Estatística: Teoria e Aplicações Usando o Microsoft Excel. 3ª ed. São Paulo: LTC, 2005.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. Introduction to the Theory of Statistic. 3ª ed. New York: Mc GrawHill, 1973.

SARTORIS, A. N. Estatística e Introdução a Econometria. São Paulo: Saraiva, 2003.

SILVA, E. M. da, et al. Estatística: para os cursos de Economia, Administração e Ciências Contábeis. 3ª ed. v.1. São Paulo: Atlas, 2000.

SPIEGEL, M. R. Estatística. 3ª ed. São Paulo: Pearson, 2009.

VIEIRA, Sônia. Como elaborar questionários. São Paulo: Atlas, 2009.

Dados Econômicos



www.atenaeditora.com.br



contato@atenaeditora.com.br



[@atenaeditora](https://www.instagram.com/atenaeditora)



www.facebook.com/atenaeditora.com.br



Ano 2022

Dados Econômicos



www.atenaeditora.com.br



contato@atenaeditora.com.br



[@atenaeditora](https://www.instagram.com/atenaeditora)



www.facebook.com/atenaeditora.com.br



Ano 2022