

Marcelo Bragatto Dal Piaz

**Uso da linguagem python no auxílio à
aprendizagem de estatística descritiva na
educação básica**

Vitória

2023

Marcelo Bragatto Dal Piaz

Uso da linguagem python no auxílio à aprendizagem de estatística descritiva na educação básica

Dissertação de mestrado apresentada ao PROFMAT como parte dos requisitos exigidos para a obtenção do título de Mestre em Matemática

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL



PROFMAT

Orientador: Prof. Dr. Alancardek Pereira Araújo

Vitória

2023

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

D136u DAL PIAZ, MARCELO, 1987-
 Uso da linguagem python no auxílio à aprendizagem de estatística descritiva na educação básica / MARCELO DAL PIAZ. - 2023.
 60 f. : il.

 Orientador: Alancardek Araújo.
 Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade Federal do Espírito Santo, Centro de Ciências Exatas.

 1. Estatística. 2. Python (Linguagem de programação de computador). 3. Educação Básica. 4. Sequências Didáticas. I. Araújo, Alancardek. II. Universidade Federal do Espírito Santo. Centro de Ciências Exatas. III. Título.

CDU: 51



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Centro de Ciências Exatas

Programa de Pós-Graduação em Matemática em Rede Nacional – PROFMAT

“Uso da Linguagem Python no Auxílio à Aprendizagem de Estatística Descritiva na Educação Básica”

Marcelo Bragatto Dal Piaz

Defesa de Dissertação de Mestrado Profissional submetida ao Programa de Pós-Graduação em Matemática em Rede Nacional da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do título de Mestre em Matemática.

Aprovado em 28/09/2023 por:

Prof.(a) Dr.(a) Alancardek Pereira Araujo
Orientador(a) – UFES

Prof.(a) Dr.(a) Fábio Júlio Valentim
Orientador(a) – UFES

Prof. Dr.(a) Giselle Ribeiro de Azeredo Silva Strey
Membro Externo – SEDU-ES





Folha de Assinaturas Marcelo Bragatto dal Piaz

Data e Hora de Criação: 28/09/2023 às 12:26:29

Documentos que originaram esse envelope:

- Folha de Assinaturas Marcelo Bragatto dal Piaz.pdf (Arquivo PDF) - 1 página(s)



Hashs únicas referente à esse envelope de documentos

[SHA256]: 6901c819ea6a3ffff1021b0ecb973f61239d22844538dd87fba742f9ff09936d

[SHA512]: 230c042425e26308dce1bc182ccb36b9cc0807368995e6d70e1c9a9fb633655dca72482b31858510e6719fe2fb37b769dccb9f43312de972fbb53867775aec

Lista de assinaturas solicitadas e associadas à esse envelope



ASSINADO - Alancardek Pereira Araújo (alancardek.araujo@ufes.br)

Data/Hora: 28/09/2023 - 17:37:11, IP: 177.159.78.190, Geolocalização: [-20.289945, -40.298086]

[SHA256]: 5220625a85b8f0640ef11056571f56db00555d323a45f6fcbce617d376941e8e

Alancardek Pereira Araujo



ASSINADO - Fabio Julio da Silva Valentim (fabio.valentim@ufes.br)

Data/Hora: 28/09/2023 - 15:51:39, IP: 200.137.65.100, Geolocalização: [-20.277092, -40.301673]

[SHA256]: ca80b19eeac119468bcd55eb502dc5f44873b8429ae5c53f9b8b9d711b3e54e0



ASSINADO - Giselle Ribeiro de Azeredo Silva Strey (giselle.sarstrey@educador.edu.es.gov.br)

Data/Hora: 28/09/2023 - 15:20:10, IP: 179.109.143.57, Geolocalização: [-20.759502, -41.531571]

[SHA256]: 0cf6a3d4317e0a08277eda8b29feb68c7f1bb76da2e78ffd0fce73d3e88d08fd

Histórico de eventos registrados neste envelope

28/09/2023 17:37:11 - Envelope finalizado por alancardek.araujo@ufes.br, IP 177.159.78.190

28/09/2023 17:37:11 - Assinatura realizada por alancardek.araujo@ufes.br, IP 177.159.78.190

28/09/2023 17:36:20 - Envelope visualizado por alancardek.araujo@ufes.br, IP 177.159.78.190

28/09/2023 15:51:39 - Assinatura realizada por fabio.valentim@ufes.br, IP 200.137.65.100

28/09/2023 15:51:07 - Envelope visualizado por fabio.valentim@ufes.br, IP 200.137.65.100

28/09/2023 15:20:10 - Assinatura realizada por giselle.sarstrey@educador.edu.es.gov.br, IP 179.109.143.57

28/09/2023 15:19:38 - Envelope visualizado por giselle.sarstrey@educador.edu.es.gov.br, IP 179.109.143.57

28/09/2023 12:27:46 - Envelope registrado na Blockchain por ivan.barbosa@ufes.br, IP 200.137.65.108

28/09/2023 12:27:45 - Envelope encaminhado para assinaturas por ivan.barbosa@ufes.br, IP 200.137.65.108

28/09/2023 12:26:30 - Envelope criado por ivan.barbosa@ufes.br, IP 200.137.65.108

Dedico a minha família. Minha Esposa, Filha e meus Pais.

Agradecimentos

Em primeiro lugar, agradeço à Deus por ter proporcionado mais esta graça em minha vida.

Agradecimento especial ao meu amigo Ygor. Seu suporte foi fundamental para que tivesse êxito nas disciplinas do programa e também no exame de qualificação.

Agradeço ao meu orientador professor Alancardek Pereira Araujo por me orientar nesta dissertação sobre Estatística no Ensino Básico e o uso da linguagem Python. Suas colocações/sugestões trouxeram um ajuste fino fundamental para o trabalho.

Por fim, agradeço aos Professores do programa que tive a oportunidade de ter aula. Moacir, Florêncio, Domingos, Etereldes e Alancardek. Muito obrigado!

*“Você não tem que lutar para ser melhor que ninguém, você precisa lutar para ser melhor
a cada dia consigo mesmo.”
(Professor Olavo de Carvalho)*

Resumo

Nos dias atuais, é de fundamental importância que as pessoas tenham a capacidade de entender e compreender o grande fluxo de informações a que estão sendo expostas dia após dia. Diante disso, torna-se necessário o desenvolvimento da habilidade de leitura e interpretação de dados numéricos, tabelas e gráficos. No que se refere ao ensino da Estatística, uma aprendizagem significativa se torna crucial, visto que a sociedade vive cada vez mais interconectada e gera cada vez mais informações. Este trabalho tem como objetivo oferecer estratégias para abordar assuntos de Estatística Descritiva fazendo uso da linguagem de programação Python, procurando contribuir para uma formação crítica dos alunos e fornecer novas ferramentas didáticas aos professores da educação básica.

Palavras-chave: Estatística Descrita. Python. Educação Básica. Sequência Didáticas.

Abstract

In today's world, it is essential for people to have the ability to understand and comprehend the vast flow of information to which they are exposed on a daily basis. Therefore, the development of the skill of reading and interpreting numerical data, tables, and graphs is necessary. Regarding the teaching of Statistics, significant learning becomes crucial, given that society is becoming increasingly interconnected and generating more and more information. This work aims to offer strategies to approach Descriptive statistics topics using the Python programming language, seeking to contribute to a critical formation of students and provide new didactic tools to basic education teachers.

Keywords: Described Statistics. Python. Basic Education. Following Teaching.

Lista de ilustrações

Figura 1 – Divisão das áreas básicas da estatística	22
Figura 2 – Esquematização população x amostra.	23
Figura 3 – Esquematização dos tipos de variáveis.	24
Figura 4 – Tabela de frequência - Variável Qualitativa	29
Figura 5 – Tabela de frequência - Variável Quantitativa Contínua	30
Figura 6 – Gráfico de barras vertical e horizontal	31
Figura 7 – Gráfico de linhas	32
Figura 8 – Gráfico de pizza	32
Figura 9 – Gráfico de dispersão	33
Figura 10 – Histograma	34
Figura 11 – Box Plot	34
Figura 12 – Box Plot gerado a partir do Exemplo 3.4.3.	35
Figura 13 – Carregando as bibliotecas	40
Figura 14 – Carregando as respostas do questionário	40
Figura 15 – Selecionando uma amostra com 12 respostas	41
Figura 16 – Linhas de código - Medidas calculadas - Programadas manualmente	41
Figura 17 – Medidas calculadas - Programadas manualmente	41
Figura 18 – Linhas de código - Tabela de frequência - Variável qualitativa	42
Figura 19 – Tabela gerada - Tabela de frequência - Variável qualitativa	42
Figura 20 – Linhas de código - Tabela de frequência - Variável quantitativa	42
Figura 21 – Tabela gerada - Tabela de frequência - Variável quantitativa	43
Figura 22 – Linhas de código - Gráfico de barras	43
Figura 23 – Gráfico gerado - Gráfico de barras	43
Figura 24 – Linhas de código - Gráfico de linhas	44
Figura 25 – Gráfico gerado - Gráfico de linhas	44
Figura 26 – Linhas de código - Gráfico de pizza	44
Figura 27 – Gráfico gerado - Gráfico de pizza	45
Figura 28 – Linhas de código - Gráfico de dispersão	45
Figura 29 – Gráfico gerado - Gráfico de dispersão	45
Figura 30 – Linhas de código - Histograma	46
Figura 31 – Gráfico gerado - Histograma	46
Figura 32 – Linhas de código - Box plot	46
Figura 33 – Gráfico gerado - Box plot	47
Figura 34 – Página para download do ANACONDA.	52
Figura 35 – Instalação do ANACONDA.	52
Figura 36 – Tela após a execução do Jupyter Notebook.	53

Figura 37 – Tela para abrir novo arquivo .ipynb. 54

Lista de tabelas

Tabela 1 – Percentual alcançado por nível e etapa em Matemática no Saeb 2021. .	15
Tabela 2 – Habilidades esperadas de Prob. e Estatística dos concluintes da Ed. Básica.	18
Tabela 3 – Níveis de Letramento Estatístico	19
Tabela 4 – Trabalho Similares - Estatística Descritiva	20
Tabela 5 – Trabalho Similares - Python	21
Tabela 6 – Perguntas do questionário a serem respondidas pelo alunos da turma .	40
Tabela 7 – Escalas de Proficiências em Tratamento de Informações (Probabilidade e Estatística) consideradas no Saeb do 5º Ano – EF.	50
Tabela 8 – Escalas de Proficiências em Tratamento de Informações (Probabilidade e Estatística) consideradas no Saeb do 9º Ano – EF.	51
Tabela 9 – Escalas de Proficiências em Tratamento de Informações (Probabilidade e Estatística) consideradas no Saeb do 3ª/4ª Série - EMT + EMI. . . .	51
Tabela 10 – Dados gerados aleatoriamente.	56
Tabela 11 – Continuação - Dados gerados aleatoriamente.	57

Sumário

1	INTRODUÇÃO	14
2	REFERENCIAL TEÓRICO	17
2.1	Estatística na BNCC 2018	17
2.2	Letramento Estatístico (Literacia Estatística)	17
2.3	Trabalhos Similares	19
3	ESTATÍSTICA BÁSICA	22
3.1	Conceitos Básicos	23
3.2	Medidas de Posição	25
3.3	Medidas de Dispersão	27
3.4	Tabelas	29
3.5	Gráficos	30
4	POR QUE A LINGUAGEM DE PROGRAMAÇÃO PYTHON	36
5	ATIVIDADE PRÁTICA	39
5.1	Explorando os Dados da Pesquisa	39
6	CONSIDERAÇÕES FINAIS	48
	REFERÊNCIAS	49
	APÊNDICE A – ESCALAS DE PROFICIÊNCIAS EM TRATAMENTO DE INFORMAÇÕES (PROB. E ESTATÍSTICA) CONSIDERADAS NO SAEB	50
	APÊNDICE B – TUTORIAL DE INSTALAÇÃO E PRIMEIRO ACESSO DO PYTHON	52
	APÊNDICE C – CURSOS GRATUITOS PARA O APRENDIZADO DE PYTHON	55
	APÊNDICE D – RESPOSTAS GERADAS SIMULANDO UMA TURMA COM 36 ALUNOS	56
	APÊNDICE E – LINHAS DE CÓDIGOS UTILIZADAS EM FORMATO TEXTO	58

1 Introdução

A Estatística é ramo da Matemática Aplicada cujos princípios decorrem da teoria das probabilidades e que tem por objeto o estudo, bem como o agrupamento metódico, de séries de fatos ou de dados numéricos, conforme o Dicionário Online de Português (<https://www.dicio.com.br/estatistica/>). Ou seja, trata da coleta, da análise, da interpretação e da apresentação de massas de dados.

A Estatística é amplamente utilizada em várias áreas, incluindo ciência, negócios, saúde e governo, para tomar decisões informadas baseadas em evidências. Ela inclui métodos descritivos, que descrevem e resumem os dados, além de métodos inferenciais, que permitem inferir informações sobre uma população a partir de dados amostrais.

De acordo com a Wikipédia, atualmente, o volume de dados gerados por diversas ações é muito grande, assim como a capacidade de processamento desses dados. Com isso, surgiu o Business Intelligence (BI) e a ciência de dados, em que o primeiro busca analisar os dados de maneira descritiva, tentando responder "o que aconteceu?" e o segundo busca uma análise preditiva, tentando responder "o que vai acontecer?"

Um grande exemplo de aplicação da Estatística ligada a grandes massas de dados é a recomendação de vídeos pelo algoritmo do YouTube. O algoritmo considera o histórico de visualização do usuário, levando em conta os vídeos que ele assistiu anteriormente. Com base nesse histórico, o algoritmo tenta entender as preferências individuais e sugere vídeos semelhantes ou relacionados. Além disso, o algoritmo também usa técnicas de aprendizado de máquina (machine learning) para encontrar vídeos com características semelhantes aos que o usuário já assistiu, buscando oferecer recomendações relevantes com base em padrões e preferências semelhantes de outros usuários.

Diante disso, é importante que as pessoas tenham entendimento e compreensão das informações a que estão expostas diariamente, sendo necessário possuir habilidade em leitura e interpretação de dados numéricos, tabelas e gráficos.

De acordo com os resultados do Saeb 2021, que é um conjunto de avaliações externas em larga escala aplicadas a cada dois anos na rede pública e em uma amostra da rede privada, com o objetivo de realizar um diagnóstico da educação básica brasileira. Aproximadamente 5 milhões de estudantes dos 26 Estados e Distrito Federal das redes Federal, Estaduais, Municipais e Privada dos 5º e 9º anos do Ensino Fundamental e os concluintes do Ensino Médio (3ª e 4ª Séries) participaram do exame. Com base no desempenho, cada aluno pode ser classificado em até dez níveis.

Analisando os resultados da prova de matemática do Saeb 2021, observa-se que

38,80% dos alunos do 5º ano do Ensino Fundamental não atingiram o nível 4, que é o nível mais básico, no qual os alunos não conseguem sequer localizar um dado em tabelas de dupla entrada e reconhecer o maior valor em uma tabela cujos dados possuem até oito ordens. No 9º ano, constatou-se que 27,75% dos alunos não atingiram o nível 2, no qual os alunos não conseguem sequer interpretar dados apresentados em um gráfico de linha simples e associar dados apresentados em gráfico de colunas a uma tabela. Por fim, no Ensino Médio, 35,25% dos alunos não atingiram o nível 2, no qual os alunos não conseguem sequer associar um gráfico de setores a dados percentuais apresentados textualmente ou em uma tabela. A Tabela 1 apresenta o percentual de alunos em cada nível alcançado na avaliação, enquanto a escala de proficiência em Probabilidade e Estatística de cada etapa pode ser consultada no Apêndice A.

Tabela 1 – Percentual alcançado por nível e etapa em Matemática no Saeb 2021.

Nível	5º Ano - EF	9º Ano - EF	3ª/4ª Série - EMT + EMI
0	1,92%	14,69%	21,11%
1	6,50%	13,07%	14,14%
2	12,93%	16,63%	17,14%
3	17,46%	18,16%	17,43%
4	18,77%	17,51%	13,18%
5	16,69%	11,77%	8,64%
6	12,61%	5,30%	4,75%
7	8,02%	2,04%	2,31%
8	3,69%	0,83%	0,97%
9	1,38%	0,00%	0,33%
10	0,03%	-	0,00%

Fonte: Inep/Saeb 2021 - Resultados - Adaptado

Uma das principais determinantes da qualidade da educação é a formação dos professores. Professores bem preparados são essenciais para elevar o nível de conhecimento dos alunos. O professor de matemática licenciado começa a atuar a partir do 6º ano do Ensino Fundamental. Para que os alunos concluintes da Educação Básica alcancem todas as habilidades descritas na BNCC, é fundamental que, desde cedo, se construa uma base sólida de conhecimento para que os alunos possam atingir os objetivos requeridos.

O uso de tecnologias, como o Python, pode ser uma ferramenta valiosa para o ensino de estatística na educação básica. As ferramentas de análise de dados podem auxiliar os estudantes a coletar, limpar e analisar dados de forma mais eficiente. Além disso, bibliotecas de visualização de dados podem ajudar os estudantes a representar gráficos e visualizar dados de forma mais clara e intuitiva.

O Python possui uma grande comunidade e uma variedade de bibliotecas e ferramentas disponíveis, o que permite aos estudantes criar projetos interessantes e úteis, além de amplamente utilizado em muitas áreas, como ciência de dados, inteligência artificial e

desenvolvimento web, o que pode ajudar a preparar os estudantes para carreiras em áreas em alta demanda no mercado de trabalho.

É importante notar que o ensino de estatística na educação básica não se limita a ensinar conceitos matemáticos, mas também é fundamental ensinar habilidades críticas de pensamento, tais como avaliar fontes de dados, questionar hipóteses e tomar decisões informadas.

Além disso, o ensino de Estatística na educação básica pode ser usado para ajudar os estudantes a desenvolver habilidades que serão úteis para eles em suas vidas futuras, incluindo carreiras em ciência de dados, negócios, saúde e governo.

Este trabalho tem como objetivo oferecer estratégias para abordagem de assuntos de estatística descritiva, utilizando principalmente a linguagem de programação Python.

No Capítulo 2, será abordado sobre o referencial teórico. Será mostrado o que contém sobre Estatística na Base Nacional Curricular Comum da Educação Básica brasileira e o que alguns autores relatam sobre letramento estatístico. Por fim, serão citados alguns trabalhos similares a este.

No Capítulo 3, será apresentado o conteúdo teórico utilizado neste trabalho que pode ser utilizado por docentes como material teórico de apoio.

No Capítulo 4, será abordado um breve resumo sobre o uso de tecnologias no ensino, incluindo a linguagem de programação Python.

No Capítulo 5, serão apresentadas algumas estratégias para o ensino de estatística descritiva, mostrando a simulação de uma atividade prática.

2 Referencial Teórico

2.1 Estatística na BNCC 2018

O governo brasileiro promulgou a Lei nº 9.394 em 20 de dezembro de 1996, que estabelece as diretrizes e bases da Educação Básica Nacional. Em seguida, o Conselho Nacional de Educação (CNE) elaborou a Base Nacional Comum Curricular (BNCC), que foi promulgada em dezembro de 2017 por meio da resolução Conselho Nacional de Educação/Conselho Pleno (CNE/CP) nº 2. De caráter normativo, a BNCC estabelece os conhecimentos a serem abordados na Educação Infantil e Ensino Fundamental. Em dezembro de 2018, foi publicada uma nova resolução CNE/CP nº 4, que instituiu a BNCC também para o Ensino Médio.

A BNCC não é um currículo em si, mas um referencial que deve ser utilizado pelos sistemas de ensino, escolas e professores para elaboração de seus currículos e projetos pedagógicos.

A Estatística é uma das competências gerais que os estudantes devem desenvolver ao longo de sua formação escolar, de acordo com a BNCC. É esperado que eles sejam capazes de "utilizar dados e informações para elaborar argumentos, formular hipóteses, construir explicações e tomar decisões, com base em princípios científicos, utilizando conhecimentos matemáticos, científicos, históricos, geográficos, sociais, econômicos e culturais".

A BNCC reconhece a importância da Estatística como uma ferramenta fundamental para a análise e interpretação de dados e informações em diversas áreas do conhecimento. Além disso, ela prevê o desenvolvimento de habilidades estatísticas específicas em cada uma das áreas de conhecimento, tais como a coleta, organização e análise de dados em Matemática; a interpretação de gráficos e tabelas em Ciências da Natureza; e a compreensão de fenômenos sociais a partir de dados estatísticos em Ciências Humanas e Sociais.

De acordo com a BNCC, espera-se que os alunos concluintes da educação básica tenham domínio das habilidades descritas na Tabela 2.

2.2 Letramento Estatístico (Literacia Estatística)

O Letramento estatístico ou literacia estatística se refere à capacidade de compreender, interpretar e analisar dados e informações estatísticas. É a habilidade de usar a estatística para tomar decisões informadas, resolver problemas e comunicar resultados de forma clara e eficaz.

Tabela 2 – Habilidades esperadas de Prob. e Estatística dos concluintes da Ed. Básica.

Descrição
Analisar tabelas, gráficos e amostras de pesquisas estatísticas apresentada sem relatórios divulgados por diferentes meios de comunicação, identificando, quando for o caso, inadequações que possam induzir a erros de interpretação, como escalas e amostras não apropriadas.
Planejar e executar pesquisa amostral sobre questões relevantes, usando dados coletados diretamente ou em diferentes fontes, e comunicar os resultados por meio de relatório contendo gráficos e interpretação das medidas de tendência central e das medidas de dispersão (amplitude e desvio padrão), utilizando ou não recursos tecnológicos.
Resolver e elaborar problemas de contagem envolvendo agrupamentos ordenáveis ou não de elementos, por meio dos princípios multiplicativo e aditivo, recorrendo a estratégias diversas, como o diagrama de árvore.
Identificar e descrever o espaço amostral de eventos aleatórios, realizando contagem das possibilidades, para resolver e elaborar problemas que envolvem o cálculo da probabilidade.
Identificar situações da vida cotidiana nas quais seja necessário fazer escolhas levando-se em conta os riscos probabilísticos (usar este ou aquele método contraceptivo, optar por um tratamento médico em detrimento de outro etc.).
Resolver e elaborar problemas que envolvem o cálculo de probabilidade de eventos em experimentos aleatórios sucessivos.
Resolver e elaborar problemas, em diferentes contextos, que envolvem cálculo e interpretação das medidas de tendência central (média, moda, mediana) e das medidas de dispersão (amplitude, variância e desvio padrão).
Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.
Interpretar e comparar conjuntos de dados estatísticos por meio de diferentes diagramas e gráficos (histograma, de caixa (box-plot), de ramos e folhas, entre outros), reconhecendo os mais eficientes para sua análise.
Reconhecer a existência de diferentes tipos de espaços amostrais, discretos ou não, e de eventos, equiprováveis ou não, e investigar implicações no cálculo de probabilidades.

Fonte: BNCC 2018 (p. 546) - Adaptado.

O letramento estatístico envolve uma compreensão básica de conceitos estatísticos. Ele também envolve a capacidade de coletar e analisar dados, interpretar gráficos e tabelas, fazer inferências e tirar conclusões com base em dados.

Para (BARBOSA; VELASQUE; SILVA (2016)), é muito difícil promover o letramento estatístico se os professores não abordam o assunto em sala de aula, e quando abordado, é feito de maneira tradicional (quadro e giz) e unidimensional (professor => aluno). Para isso, sugere a utilização da tecnologia e de dinâmicas ativas de aprendizagem.

Watson e Callingham (2003) propuseram um modelo de níveis de letramento

estatístico, apresentado na Tabela 3.

Tabela 3 – Níveis de Letramento Estatístico

Nível	Descrição
6. Matemático Crítico	O aluno demonstra possuir habilidade matemática sofisticada para realizar muitas tarefas, desenvolver uma postura crítica, fazer interpretações e questionamentos.
5. Crítico	O aluno demonstra ser capaz de desenvolver uma postura crítica, fazer questionamentos em alguns contextos, usar a terminologia apropriada e interpretar quantitativamente.
4. Consistente Não Crítico	O aluno demonstra possuir habilidade estatística associada com a média, probabilidade simples, variação e interpretação gráfica.
3. Inconsistente	O aluno demonstra usar as ideias de estatística e conseguir obter algumas conclusões sem justificativas.
2. Informal	O aluno demonstra conseguir usar elementos simples da terminologia, faz cálculos básicos a partir de tabelas e gráficos.
1. Idiossincrático	O aluno demonstra uma habilidade matemática básica associada com a leitura e contagem (um a um) de valores em uma tabela, mas não consegue usar uma terminologia simples.

Fonte: Watson e Callingham (2003) – Adaptado.

Para Coutinho e Spina (2016), o professor deve identificar no livro didático a presença de elementos que o desenvolvimento do letramento estatístico. Caso o livro didático não contemple um conteúdo de maneira satisfatória do ponto de vista do letramento estatístico, o professor pode complementar o conteúdo.

Ter letramento estatístico é importante em muitos campos, incluindo negócios, ciência, saúde, finanças, governo e pesquisa acadêmica. Ele ajuda as pessoas a tomar decisões informadas com base em dados e a entender melhor o mundo ao seu redor.

2.3 Trabalhos Similares

Para o ensino de Estatística Descritiva, (SARAIVA; VICTER; SIQUEIRA (2017)) desenvolveram um software chamado SISTAT. O programa permite calcular diversas medidas e além de tabelas e gráficos.

Já Da Silva et al.(2018), construíram uma sequencia didática para o ensino da Estatística Descritiva utilizando o SISVAR, que é um software gratuito que tem o objetivo de fornecer ferramentas para análise estatística.

Em particular, consultando na base de 7023 dissertações do PROFMAT utilizando como critério de busca a palavra chave "Estatística Descritiva" foi observado um total de 10 (dez) trabalhos. Dentre eles, APOIO COMPUTACIONAL PARA O ENSINO DE ESTATÍSTICA DESCRITIVA fez uso do software R que é um programa semelhante ao Python. A Tabela 4 mostra os títulos encontrados. A consulta foi realizada em 05/10/2023.

Tabela 4 – Trabalho Similares - Estatística Descritiva

Aluno	Título da Dissertação
MARINEUSA APARECIDA RIBEIRO	CONTRIBUIÇÃO AO ESTUDO DO SOFTWARE R COMO FERRAMENTA DIDÁTICO- PEDAGÓGICA PARA O DESENVOLVIMENTO DE ESTATÍSTICA DESCRITIVA NO ENSINO MÉDIO
ALEXANDRE FONSECA SANTOS	APOIO COMPUTACIONAL PARA O ENSINO DE ESTATÍSTICA DESCRITIVA
ALAN KARDEK SANTOS FERREIRA	RELEVÂNCIA DA ESTATÍSTICA DESCRITIVA APLICADA A TÉCNICAS DE INVESTIGAÇÃO DE ACIDENTES DO TRABALHO: UM ESTUDO DE CASO NO CURSO TÉCNICO DE SEGURANÇA DO TRABALHO DO INSTITUTO FEDERAL DA BAHIA – CAMPUS DE VITÓRIA DA CONQUISTA
THIAGO BRAGA FERREIRA	UMA PROPOSTA DE ABORDAGEM DA ESTATÍSTICA DESCRITIVA NO ENSINO FUNDAMENTAL.
MARIO JOSE PEREIRA	ESTATÍSTICA DESCRITIVA: O USO DA ENGENHARIA DIDÁTICA NO ESTUDO DE MEDIDAS DE TENDÊNCIA CENTRAL
VIVIENNE DO VAL RODRIGUES	ENSINANDO ESTATISTICA DESCRITIVA: UM PROJETO PARA O ENSINO MEDIO
WESYLLIS DAS MERCES SALVADOR	ANÁLISE DO CONTEÚDO DE ESTATÍSTICA DESCRITIVA NO ENSINO MÉDIO
VIVIAM CIARINI DE SOUZA AMORIM	ESTATÍSTICA DESCRITIVA: BREVE HISTÓRICO, CONCEITOS E EXEMPLOS APLICÁVEIS NO ENSINO MÉDIO
EVANDRO DE MOURA RIOS	ESTATÍSTICA DESCRITIVA, PROBABILIDADE E ESTIMAÇÃO: NOÇÕES PARA O ENSINO BÁSICO
ALESSANDRO MORETTI	O ENSINO DA ESTATÍSTICA DESCRITIVA A PARTIR DA PROPOSTA DE RESOLUÇÃO DE PROBLEMAS

Fonte: <https://profmatt-sbm.org.br/dissertacoes/>

Adiante, utilizando como critério de busca a palavra chave "Python" foi observado um total de 10 (dez) trabalhos. Dentre eles, CLASSIFICAÇÃO DE ESTUDANTES ATRAVÉS DE ÁRVORES DE DECISÃO VIA PYTHON E RAPIDMINER fez uso do Python em uma aplicação que envolve Estatística . A Tabela 5 mostra os títulos encontrados. A consulta foi realizada em 05/10/2023.

Tabela 5 – Trabalho Similares - Python

Aluno	Título da Dissertação
CARLOS ANTONIO MARQUES DE SOUSA	UMA ABORDAGEM METODOLÓGICA PARA O CÁLCULO DE VOLUMES DE INTERSEÇÃO ENTRE SÓLIDOS UTILIZANDO A LINGUAGEM PYTHON
JOÃO EVAYR DE SOUZA	O USO DA LINGUAGEM DE PROGRAMAÇÃO PYTHON NA RESOLUÇÃO DE PROBLEMAS MATEMÁTICOS DO ENSINO MÉDIO
BRUNNO DE OLIVEIRA BORGES	A LINGUAGEM PYTHON COMO FERRAMENTA NO ENSINO BÁSICO
MARCOS CESAR CABRAL GALVAO	ENSINO E APRENDIZAGEM DA MATEMÁTICA NA EDUCAÇÃO BÁSICA UTILIZANDO TECNOLOGIAS E DESENVOLVENDO PENSAMENTO COMPUTACIONAL: ABORDAGEM COM SCRATCH, PORTUGOL, PYTHON E GEOGEBRA
LEOMIR AUGUSTO SEVERO GRAVE	O PENSAMENTO COMPUTACIONAL NA PRÁTICA: UMA EXPERIÊNCIA USANDO PYTHON EM AULAS DE MATEMÁTICA BÁSICA
GEOVANE MORATO PINTO	OBTENÇÃO DE CORPOS DE NÚMEROS E CORPOS FINITOS UTILIZANDO O PYTHON
GUSTAVO NOGUEIRA LEITE	PYTHON E FORTRAN COMO FERRAMENTA NA RESOLUÇÃO DE PROBLEMAS DE MATEMÁTICA E FÍSICA
ROGERIO DA SILVA CAVALCANTE	ARITMÉTICA COM PYTHON
FÁBIA VALÉRIA DE JESUS SILVA	EXPLORANDO AS ONDAS TRIGONOMÉTRICAS ATRAVÉS DO DESENVOLVIMENTO DE ANIMAÇÕES EM PYTHON, USANDO O RASPBERRY PI COMO TECNOLOGIA DE SUPORTE.
NATALIA GONÇALVES CAETANO	CLASSIFICAÇÃO DE ESTUDANTES ATRAVÉS DE ÁRVORES DE DECISÃO VIA PYTHON E RAPIDMINER

Fonte: <https://profmat-sbm.org.br/dissertacoes/>

3 Estatística básica

Este capítulo tem como objetivo de orientar os professores da Educação Básica a terem sugestões para a abordagem de alguns dos vários conceitos da Estatística.

Primeiramente, vale destacar a diferença entre dados e informação. Dados referem-se a fatos brutos e sem contexto, como números, palavras, imagens ou sons. Por si só, os dados não têm significado ou utilidade imediata.

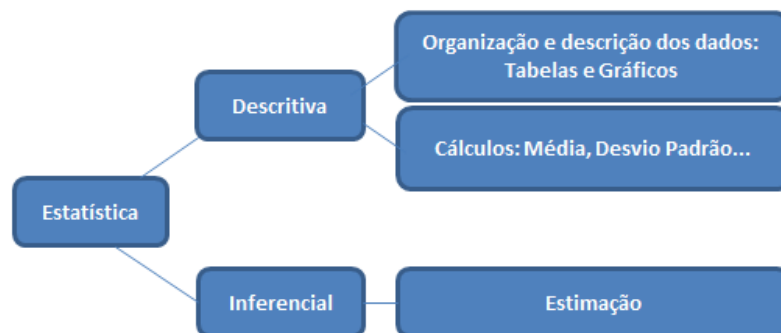
Já a informação é o significado ou o contexto que é extraído dos dados. Quando os dados são processados, organizados e interpretados, eles se tornam informações que podem ser usadas para tomar decisões, realizar análises ou obter percepções.

Portanto, dados e informação são conceitos complementares. Os dados fornecem a matéria-prima para a criação de informações úteis.

Posto isso, pode-se afirmar que a estatística é dividida em duas áreas principais:

- Estatística descritiva: é a área da estatística que se dedica à coleta, organização, resumo, apresentação e interpretação dos dados. Concentra-se em descrever e representar os dados através de medidas de tendência central, medidas de dispersão, tabelas e gráficos.
- Estatística inferencial: é a área da estatística que se dedica a fazer inferências sobre uma população com base em uma amostra. Faz uso de amostras para estimar parâmetros populacionais e testar hipóteses com diversas técnicas estatísticas. A estatística inferencial permite fazer generalizações sobre uma população com base em uma amostra e é amplamente utilizada em campos como pesquisa científica, negócios e saúde.

Figura 1 – Divisão das áreas básicas da estatística



Fonte: Elaboração própria

3.1 Conceitos Básicos

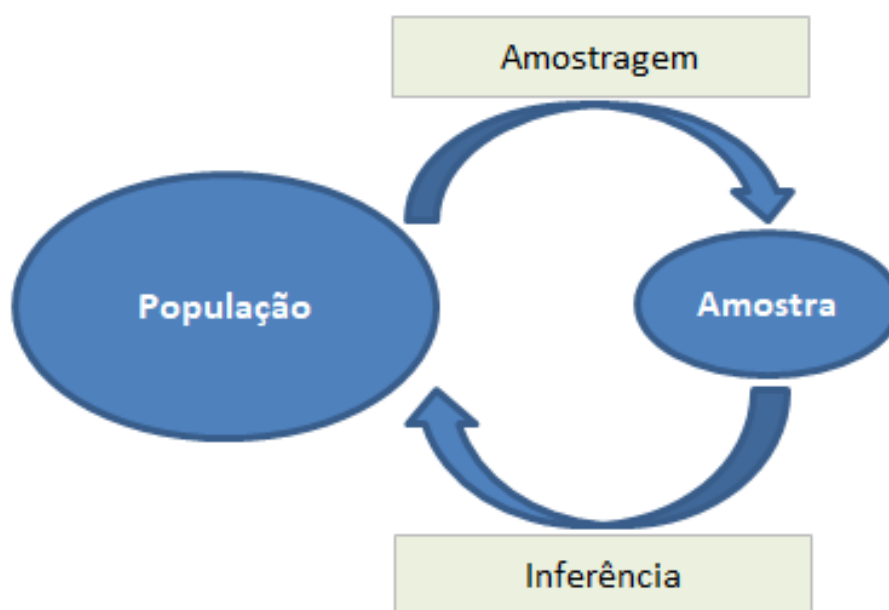
Nesta seção, serão discutidos os principais conceitos da estatística utilizados na educação básica.

Definição 3.1.1. *População* é o conjunto de todos os elementos ou resultados sob investigação. *Amostra* é qualquer subconjunto da população. (BUSSAB; MORETTIN, 2010)

A população é um conjunto de indivíduos de uma mesma espécie que vivem em um determinado espaço geográfico ou ambiente. Pode se referir também a um grupo de indivíduos de uma mesma espécie que podem interagir entre si, compartilhando características comuns e tendo acesso aos mesmos recursos. Em estatística, população é o conjunto completo de indivíduos, eventos ou objetos sobre os quais se deseja coletar informações.

A amostra é um subconjunto de indivíduos selecionados de uma população com o objetivo de representar a população como um todo. Em estatística, a amostra é utilizada para coletar informações sobre a população, com o objetivo de fazer inferências estatísticas e obter conclusões sobre a população completa. A amostra deve ser escolhida de forma aleatória e estratificada para garantir que ela seja representativa da população. Diferentes técnicas de amostragem podem ser utilizadas, como amostragem aleatória simples, amostragem estratificada e amostragem por conglomerados. (BOLFARINE; BUSSAB, 2005)

Figura 2 – Esquematização população x amostra.



Fonte: Elaboração própria

Exemplo 3.1.2. Suponha que a população de votantes do Espírito Santo seja de 3 milhões de pessoas. Deseja-se realizar uma pesquisa eleitoral para saber as intenções de voto dos candidatos ao governo do estado. Para isso, serão realizadas 1.200 entrevistas com o devido procedimento para seleção da amostra. Essa quantidade de entrevistas dará uma margem de erro de 2.95% (faixa de incerteza associada aos resultados da pesquisa), para mais ou para menos, a um nível de confiança de 95% (grau de certeza com que os resultados da pesquisa representam a opinião da população).

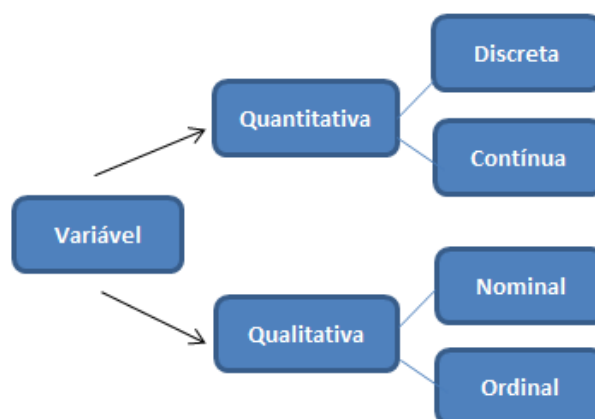
- População: 3 milhões de pessoas
- Amostra: 1.200 pessoas

Definição 3.1.3. *Variável* é uma característica ou medida que pode ser observada ou medida em uma amostra ou população.

Em estatística, as variáveis são utilizadas para descrever e analisar os dados. Elas podem ser classificadas de acordo com seu tipo de medida e são chamadas de *qualitativas* ou *quantitativas*.

- Variáveis quantitativas: são aquelas que podem ser medidas numericamente, como peso, idade, renda, etc. Elas podem ser discretas (com valores contáveis, como número de filhos) ou contínuas (com valores que podem ser medidos em uma escala contínua, como altura).
- Variáveis qualitativas: são aquelas que não podem ser medidas numericamente, como escolaridade e cor dos olhos. Elas podem ser nominais (com valores que não têm ordem, como cor dos olhos) ou ordinais (com valores que têm ordem, como escolaridade).

Figura 3 – Esquemática dos tipos de variáveis.



Fonte: Elaboração própria

Exemplo 3.1.4. Considere o Exemplo 3.1.2.

- Variável (qualitativa nominal): Intenção de voto dos candidatos ao governo do estado do Espírito Santo

□

3.2 Medidas de Posição

Definição 3.2.1. *Medidas de posição* são medidas de tendência central utilizadas para descrever e resumir a distribuição dos dados. Cada uma delas tem sua utilidade dependendo do tipo de dados e do objetivo da análise.

- Média (aritmética) (μ): é uma medida de tendência central que indica o valor central de uma série de dados.

Ela é calculada somando todos os valores da série e dividindo pelo quantidade de valores. A média é afetada por valores extremos (dados que estão muito distantes da média ou do centro de um conjunto de dados).

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

Exemplo 3.2.2. Considere as notas obtidas em uma prova de matemática de uma turma de dezenove alunos: 0, 1, 2, 2, 3, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7, 8, 8, 9, 10. Assim,

$$\mu = \frac{0 + 1 + 2 + \dots + 9 + 10}{19} = \frac{97}{19} = 5,11$$

□

- Mediana (Md): é uma medida de tendência central que indica o valor que divide uma série de dados em dois grupos de igual tamanho.

Ela é calculada ordenando os valores da série e encontrando o valor do meio. Se tivermos uma quantidade ímpar de termos, basta ordená-los e selecionar o termo central. Se tivermos uma quantidade par de termos, a mediana é calculada pela média aritmética dos termos centrais. A mediana não é afetada por valores extremos (outliers).

- Quantidade ímpar de termos: $\underbrace{1^\circ, 2^\circ, 3^\circ, 4^\circ}_{4 \text{ termos}} \quad \underbrace{5^\circ}_{\text{termo central}} \quad \underbrace{6^\circ, 7^\circ, 8^\circ, 9^\circ}_{4 \text{ termos}}$

- Quantidade par de termos: $\underbrace{1^\circ, 2^\circ, 3^\circ, 4^\circ}_{4 \text{ termos}} \quad \underbrace{5^\circ, 6^\circ}_{\text{termos centrais}} \quad \underbrace{7^\circ, 8^\circ, 9^\circ, 10^\circ}_{4 \text{ termos}}$

Exemplo 3.2.3. Considere o Exemplo 3.2.2.. Existem 19 notas e o termo central encontra-se na décima posição, ou seja, a nota 5. Logo, $Md = 5$.

$$\underbrace{0, 1, 2, 2, 3, 4, 4, 5, 5}_{9 \text{ termos}} \quad \underbrace{5}_{\text{termo central}} \quad \underbrace{5, 6, 6, 7, 7, 8, 8, 9, 10}_{9 \text{ termos}}$$

- **Moda (Mo):** é a medida de tendência central que indica o valor mais frequente em uma série de dados.

Ela é calculada encontrando o valor que ocorre com maior frequência. Uma série de dados pode ter várias modas (valores que ocorrem com a mesma frequência).

Exemplo 3.2.4. Considere novamente o Exemplo 3.2.2.. A nota que mais vezes se repete é a nota 5, tendo 4 aparições. Logo, $Mo = 5$. \square

Nota-se que as medidas (Média, Mediana e Moda) possuem valores bem próximos no Exemplo 3.2.2 e qualquer uma delas poderia ser usada como representativa do conjunto de dados. Geralmente, a média é a medida mais usada. Contudo, ela pode conduzir a erros de interpretação, visto que a média é bastante afetada por valores extremos. Nessas situações, a mediana se configura uma medida mais adequada.

Observe que, para calcular a moda de uma variável, precisa-se apenas da distribuição de frequências (contagem). Já para a mediana necessita-se minimamente ordenar as realizações da variável. Finalmente, a média só pode ser calculada para variáveis quantitativas.

Estas condições limitam bastante o cálculo para as variáveis qualitativas. Para variáveis nominais somente é permitido utilizar a moda. Para as variáveis ordinais, além da moda, pode-se usar também a mediana.

- **Quartil (Q):** é utilizado para dividir um conjunto de dados ordenado em partes iguais, especificamente em quatro partes. Cada uma dessas partes é chamada de quartil. O primeiro quartil (Q_1) é o valor que separa 25% dos dados abaixo e 75% acima dele. O segundo quartil (Q_2) é o mesmo que a mediana, que divide os dados em duas partes iguais, deixando 50% dos dados abaixo e 50% acima dele. O terceiro quartil (Q_3) é o valor que deixa 75% dos dados abaixo e 25% acima dele.

Para encontrar o segundo quartil (Q_2) basta calcular a mediana, visto que são a mesma coisa. Para calcular o primeiro quartil (Q_1), deve-se encontrar a mediana do conjunto de dados, em seguida considerar um novo conjunto do primeiro elemento até a mediana encontrada e calcular a mediana desse novo conjunto. Esse será o valor de Q_1 . Analogamente, vale para o terceiro quartil (Q_3).

- Quantidade ímpar de termos:

$$\underbrace{1^\circ, 2^\circ, 3^\circ, 4^\circ} \quad \underbrace{5^\circ} \quad \underbrace{6^\circ, 7^\circ, 8^\circ, 9^\circ}$$

. *mediana* .

$$\underbrace{1^\circ, 2^\circ} \quad \underbrace{3^\circ} \quad \underbrace{4^\circ, 5^\circ} \quad \vdots \quad \underbrace{5^\circ, 6^\circ} \quad \underbrace{7^\circ} \quad \underbrace{8^\circ, 9^\circ}$$

. Q1..... Q3..... .

- Quantidade par de termos:

$$\underbrace{1^\circ, 2^\circ, 3^\circ, 4^\circ} \quad \underbrace{5^\circ, 6^\circ} \quad \underbrace{7^\circ, 8^\circ, 9^\circ, 10^\circ}$$

. *mediana* .

$$\underbrace{1^\circ, 2^\circ} \quad \underbrace{3^\circ, 4^\circ} \quad \underbrace{5^\circ, 6^\circ} \quad \vdots \quad \underbrace{5^\circ, 6^\circ} \quad \underbrace{7^\circ, 8^\circ} \quad \underbrace{9^\circ, 10^\circ}$$

. Q1 : (3° + 4°)/2 . . Q3 : (7° + 8°)/2 .

Exemplo 3.2.5. Considere as notas obtidas em uma prova de português de uma turma de nove: 2, 4, 4, 5, 5, 6, 7, 8, 9. Tem-se que a mediana (Q_2) é 5, $Q_1 = 4$ e $Q_3 = 7$.

$$\underbrace{2, 4, 4, 5} \quad \underbrace{5} \quad \underbrace{6, 7, 8, 9}$$

. *mediana* .

$$\underbrace{2, 4} \quad \underbrace{4} \quad \underbrace{4, 5} \quad \vdots \quad \underbrace{5, 6} \quad \underbrace{7} \quad \underbrace{8, 9}$$

. Q1..... Q3..... .

□

3.3 Medidas de Dispersão

Definição 3.3.1. *Medidas de Dispersão* são medidas estatísticas utilizadas para descrever a variação ou o espalhamento dos dados em uma série ou amostra.

Elas são utilizadas para indicar como os dados estão distribuídos em relação à média, mediana ou moda.

O resumo de um conjunto de dados por uma única medida representativa de posição de tendência central esconde toda a informação sobre a variabilidade do conjunto de observações.

Exemplo 3.3.2. *Suponhamos que cinco grupos de alunos submeteram-se a um teste, obtendo-se as seguintes notas:*

- Grupo A: 3, 4, 5, 6, 7
- Grupo B: 1, 3, 5, 7, 9
- Grupo C: 5, 5, 5, 5, 5
- Grupo D: 3, 5, 5, 7
- Grupo E: 3, 5, 5, 6, 6

Tem-se que as médias dos grupos é a mesma (igual a 5). A identificação de cada uma destas séries de dados por sua média nada informa sobre as suas diferentes variabilidades. Nota-se, então, a conveniência de serem criadas medidas que sumerizem a variabilidade de um conjunto de observações e que se permita comparar conjuntos de dados distintos.

Um critério frequentemente usado para tal fim é aquele que mede a dispersão dos dados em torno de sua média. O princípio básico é analisar os desvios das observações em relação a média das observações.

- Desvio médio (dm): é uma medida estatística que representa a média dos desvios absolutos de cada valor em relação à média dos dados.

$$dm = \frac{|x_1 - \mu| + |x_2 - \mu| + \dots + |x_n - \mu|}{n} = \frac{\sum_{i=1}^n |x_i - \mu|}{n}$$

- Variância (σ^2): é uma medida de dispersão que indica a variação dos dados em relação à média.

Ela é calculada como a média dos quadrados dos desvios em relação à média.

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Voltando ao Exemplo 3.3.2, tem-se que

$$\begin{aligned} \text{Grupo A: } dm(A) &= 1,2 & \sigma^2(A) &= 2,0 \\ \text{Grupo D: } dm(D) &= 1,0 & \sigma^2(D) &= 2,0 \end{aligned}$$

Assim, pode-se dizer, então, que, segundo o desvio médio, o grupo D é mais homogêneo que A, enquanto ambos são igualmente homogêneos, segundo a variância.

As medidas de dispersão podem ser utilizadas para realizar comparações de homogeneidade e heterogeneidade dos conjuntos de dados quando comparáveis.

Sendo a variância uma medida de dimensão igual ao quadrado da dimensão dos dados (por exemplo, se os dados são expressos em cm, a variância será expressa em cm^2), pode causar problemas de interpretação.

- Desvio padrão ($\sigma > 0$): é uma medida de dispersão que indica a variação dos dados em relação à média. A vantagem dele é que sempre terá a mesma unidade de medida da média.

Ele é calculado como a raiz quadrada da variância.

$$\sigma = \sqrt{\sigma^2}$$

O desvio médio e o desvio padrão indicam em média qual será o "erro" cometido ao tentar substituir cada observação do conjunto de dados pela medida resumo (média).

□

3.4 Tabelas

Definição 3.4.1. *Tabelas de frequência* são ferramentas da estatística que permitem organizar e resumir dados quantitativos ou qualitativos.

Consiste em mostrar a frequência (número de vezes) em que cada valor ou categoria de uma variável de uma população (ou amostra). As frequências normalmente são exibidas com os números absolutos ou de maneira relativa (em percentual).

Para dados qualitativos, a tabela de frequência é construída listando cada categoria da variável e contando quantas vezes cada categoria ocorre no conjunto de dados.

Exemplo 3.4.2. Uma universidade pretende comprar novas carteiras escolares de braço. Visando estimar a quantidade de carteiras de braço direito e braço esquerdo, foi realizada uma pesquisa com 1000 alunos perguntando qual a mão o aluno utiliza para escrever. Destes, 860 responderam utilizar a mão direita, 100 a mão esquerda e 40 responderam utilizar ambas as mãos.

Assim, a tabela fica conforme a Figura 4.

Figura 4 – Tabela de frequência - Variável Qualitativa

Mão que utiliza para Escrever	Frequência Absoluta	Frequência Relativa
Direta	860	0,86 = 86%
Esquerda	100	0,10 = 10%
Ambidestro	40	0,04 = 4%
Total	1000	1,00 = 100%

Fonte: Elaboração própria

Para dados quantitativos, a tabela de frequência é construída dividindo a escala de medida da variável em intervalos (chamados de classes) e contando quantas vezes cada valor cai dentro de cada intervalo. Preferencialmente, cada classe deve possuir o mesmo tamanho (intervalos de mesma amplitude).

Exemplo 3.4.3. Em uma escola, foram medidas as alturas de 50 alunos com 12 anos de idade. As medidas, em metros, foram: 1,36, 1,38, 1,39, 1,41, 1,41, 1,42, 1,43, 1,43, 1,44, 1,44, 1,45, 1,46, 1,46, 1,47, 1,48, 1,49, 1,50, 1,50, 1,51, 1,51, 1,52, 1,52, 1,53, 1,53, 1,54, 1,54, 1,55, 1,55, 1,56, 1,56, 1,57, 1,57, 1,58, 1,58, 1,59, 1,59, 1,61, 1,61, 1,62, 1,63, 1,63, 1,64, 1,65, 1,66, 1,67, 1,68, 1,68, 1,69, 1,70, 1,76.

Assim, a tabela fica conforme a Figura 5.

Figura 5 – Tabela de frequência - Variável Quantitativa Contínua

Altura	Frequência Absoluta	Frequência Relativa
1,30 ----- 1,40	3	0,06 = 6%
1,40 ----- 1,50	13	0,26 = 26%
1,50 ----- 1,60	20	0,40 = 40%
1,60 ----- 1,70	12	0,24 = 24%
1,70 ----- 1,80	2	0,04 = 4%
Total	50	1,00 = 100%

Fonte: Elaboração própria

Note que foi utilizada a notação $a|----- b$, ou seja, indica que o intervalo contém o extremo a e não contém o extremo b . Analogamente, pode-se utilizar a notação $[a;b)$.

O leitor pode se perguntar como determinar a quantidade de intervalos de classes. Existem alguns métodos os quais podem ser utilizados.

Para (BUSSAB; MORETTIN, 2010), a escolha dos intervalos é arbitrária e a familiaridade do pesquisador com os dados é que indicará quantas e quais classes (intervalos) devem ser usadas. Normalmente, sugere-se o uso de 5 a 15 classes com a mesma amplitude.

Para (MILONE, 2004), em princípio, a forma e o número de classes são arbitrários. Porém, sugere como possíveis critérios de cálculo do número de intervalos de classes o critério da raiz ($k = \sqrt{N}$), o critério de Sturges ($k = 1 + 3,3 \log(N)$) e também critério de Milone ($k = -1 + 2 \ln(N)$), onde k é o número de classes e N é o número de elementos do conjunto de dados.

3.5 Gráficos

Definição 3.5.1. *Gráficos* são representações visuais dos dados que permitem comparar, analisar e interpretar informações de maneira mais clara e intuitiva.

Os gráficos são ferramentas valiosas para visualizar e comunicar dados, pois permitem identificar tendências, padrões e outliers (valores extremos) de forma rápida e eficaz.

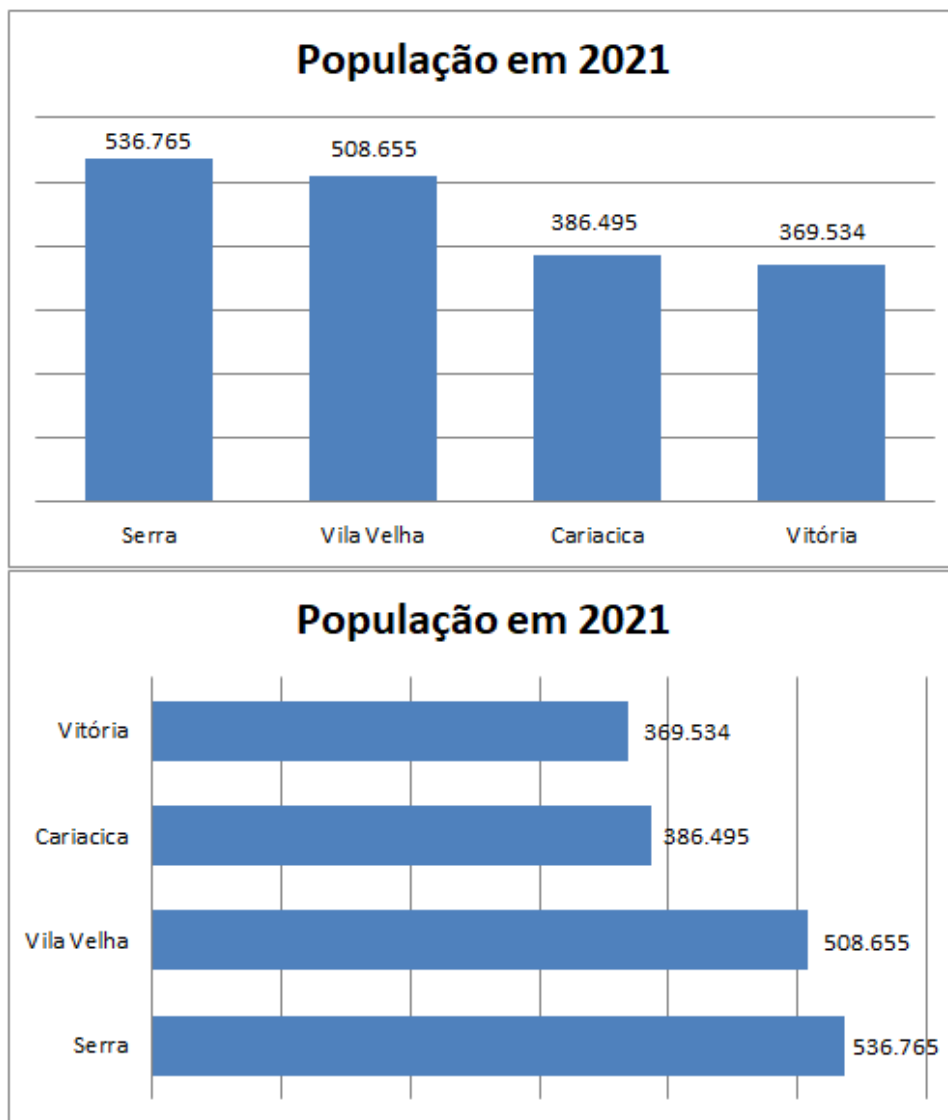
Existem vários tipos de gráficos, cada um adequado para representar diferentes tipos de dados e informações. Alguns dos tipos mais comuns de gráficos incluem:

- Gráfico de barras: utilizado para comparar a frequência ou ocorrência de diferentes categorias ou valores. Pode ser horizontal ou vertical.

Exemplo 3.5.2. Considere as populações de alguns municípios do Espírito Santo em 2021. Serra: 536.765 ; Vila Velha: 508.655 ; Cariacica: 386.495 ; Vitória: 369.534.

Assim, os gráficos de barras ficam conforme a Figura 6.

Figura 6 – Gráfico de barras vertical e horizontal



Fonte: Elaboração própria

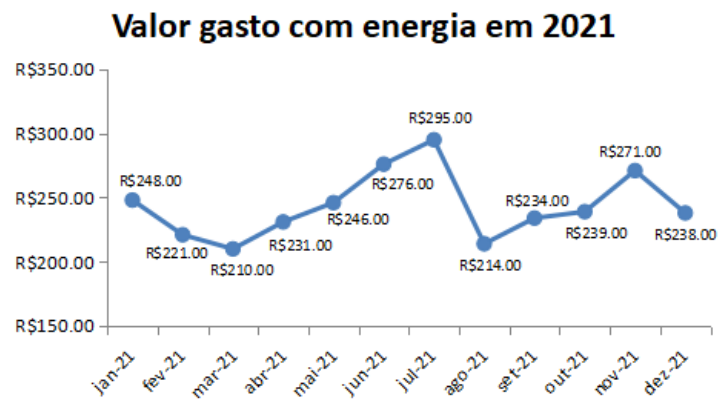
- Gráfico de linha: utilizado para mostrar a variação de uma variável ao longo do tempo.

Exemplo 3.5.3. Considere os gastos mensais com energia de uma residência em 2021.

Jan/21 - R\$ 248 ; Fev/21 - R\$ 221 ; Mar/21 - R\$ 210 ; Abr/21 - R\$ 231
 Mai/21 - R\$ 246 ; Jun/21 - R\$ 276 ; Jul/21 - R\$ 295 ; Ago/21 - R\$ 214
 Set/21 - R\$ 234 ; Out/21 - R\$ 239 ; Nov/21 - R\$ 271 ; Dez/21 - R\$ 238

Assim, o gráfico de linhas fica conforme a Figura 7.

Figura 7 – Gráfico de linhas



Fonte: Elaboração própria

- Gráfico de pizza: utilizado para mostrar a proporção de diferentes categorias ou valores em relação a um todo.

Exemplo 3.5.4. Considere os dados do Exemplo 3.4.2..

Assim, o gráfico de pizza fica conforme a Figura 8.

Figura 8 – Gráfico de pizza



Fonte: Elaboração própria

- Gráfico de dispersão: utilizado para mostrar a relação entre duas variáveis quantitativas.

Exemplo 3.5.5. Considere as informações de peso e altura de 10 pessoas num local de trabalho.

Pessoa 01 - Peso: 50kg | Altura: 1,60m

Pessoa 02 - Peso: 55kg | Altura: 1,58m

Pessoa 03 - Peso: 60kg | Altura: 1,70m

Pessoa 04 - Peso: 62kg | Altura: 1,68m

Pessoa 05 - Peso: 68kg | Altura: 1,72m

Pessoa 06 - Peso: 70kg | Altura: 1,75m

Pessoa 07 - Peso: 75kg | Altura: 1,85m

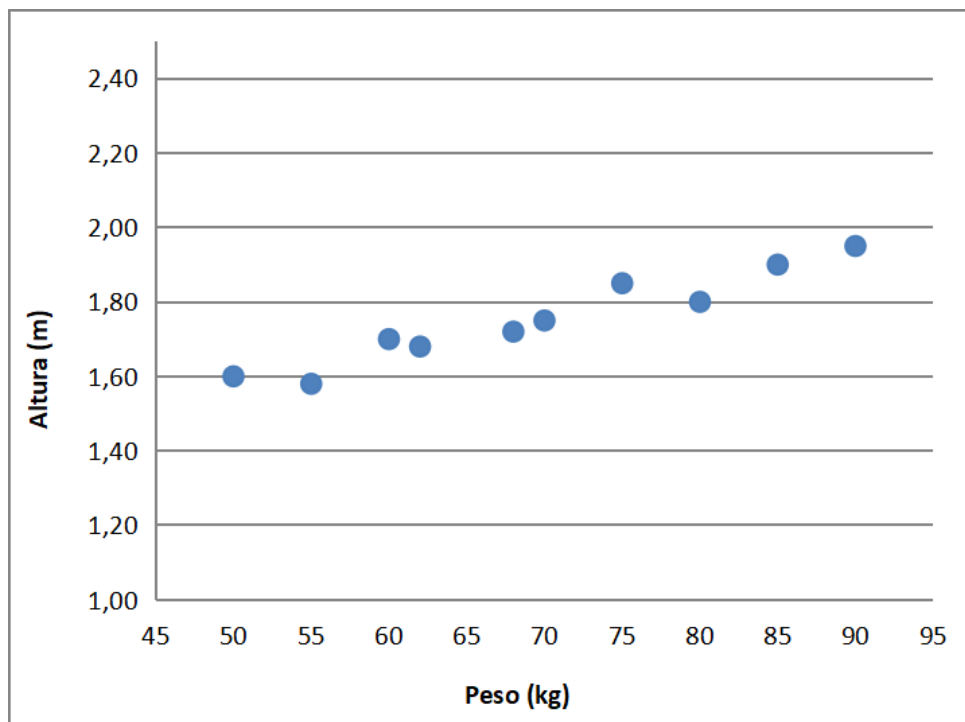
Pessoa 08 - Peso: 80kg | Altura: 1,80m

Pessoa 09 - Peso: 85kg | Altura: 1,90m

Pessoa 10 - Peso: 90kg | Altura: 1,95m

Assim, o gráfico de dispersão fica conforme a Figura 9.

Figura 9 – Gráfico de dispersão



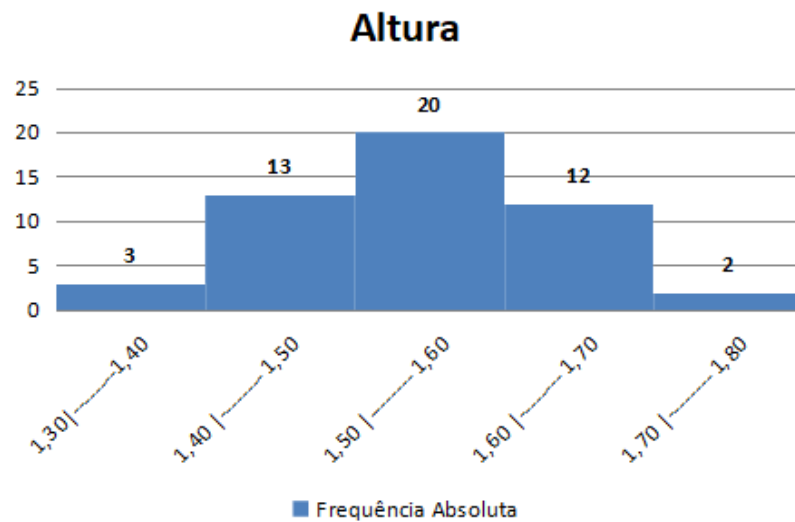
Fonte: Elaboração própria

- Histograma: utilizado para mostrar a distribuição de uma variável quantitativa.

Exemplo 3.5.6. Considere os dados do Exemplo 3.4.3..

Assim, o histograma fica conforme a Figura 10.

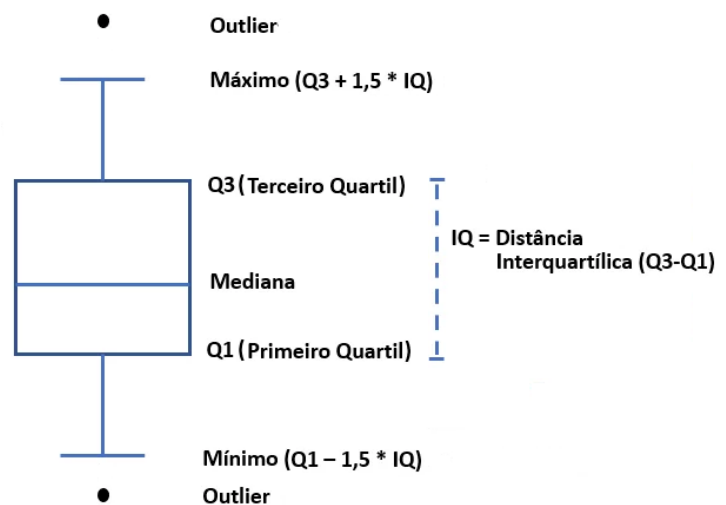
Figura 10 – Histograma



Fonte: Elaboração própria

- Box Plot: é uma representação visual das principais características do conjunto de dados, como a mediana, quartis e, principalmente, identificar possíveis valores extremos (outliers).

Figura 11 – Box Plot

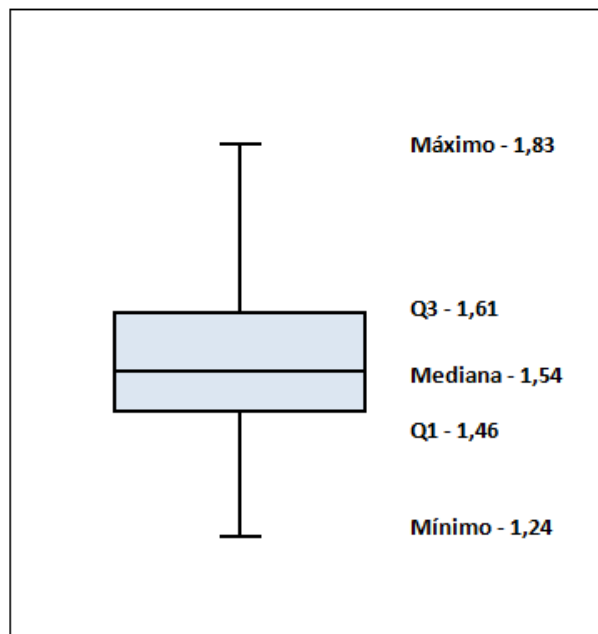


Fonte: Elaboração própria

Exemplo 3.5.7. Considere os dados do Exemplo 3.4.3..

Assim, o Box Plot fica conforme a Figura 12.

Figura 12 – Box Plot gerado a partir do Exemplo 3.4.3.



Fonte: Elaboração própria

□

4 Por que a linguagem de programação Python

O uso de tecnologias no ensino de matemática tem se tornado cada vez mais comum. O aprendizado interativo permite que os estudantes experimentem os conceitos matemáticos de forma interativa, o que pode ajudar a tornar o aprendizado mais robusto. As visualizações de conceitos com as tecnologias, por exemplo, com o uso de softwares de geometria dinâmica e simulações, permitem que os estudantes visualizem os conceitos matemáticos de forma mais clara e concreta, o que pode ajudar a esclarecer dúvidas e aprofundar o entendimento.

Nos dias atuais, o amplo uso de ferramentas educacionais como aplicativos, jogos e plataformas online, fornecem recursos para o ensino de matemática, ajudando a tornar o ensino mais atrativo. Além disso, uma preparação para o mundo real, visto que o uso de planilhas eletrônicas, programação e análise de dados preparam os estudantes para carreiras em áreas em diversas áreas.

Especificamente para o ensino de estatística, a utilização de recursos computacionais possui vantagens como:

- Análise de grandes conjuntos de dados: o uso de computadores permite aos estudantes analisar grandes conjuntos de dados, o que é impossível de fazer manualmente. Isso ajuda a desenvolver habilidades de análise e interpretação de dados.

- Visualização de dados: o uso de computadores permite aos estudantes criar gráficos e visualizações interativas de dados, o que pode ajudar a tornar os conceitos estatísticos mais fáceis de entender.

Diante disso, a utilização do Python que é uma linguagem de programação de alto nível e fácil de aprender (por possuir uma sintaxe mais simples do que as demais), torna ideal para introduzir conceitos de estatística aos estudantes no nível básico e como bônus, o aprendizado de programação. O aprendizado de programação pode oferecer benefícios como:

1. Pensamento lógico: a programação ajuda a desenvolver habilidades de pensamento lógico e resolução de problemas.

2. Criatividade: a programação permite que as pessoas criem e desenvolvam suas próprias ideias e projetos, o que pode ser altamente satisfatório e estimulante.

3. Autonomia: aprender a programar pode ajudar as pessoas a serem mais independentes e a automatizar tarefas cotidianas.

4. Melhor compreensão da tecnologia: aprender a programar pode ajudar as pessoas a entender como a tecnologia funciona e como ela pode ser usada de maneira eficaz.

Mas, por que o Python? Além do "sistema" Python ser gratuito, ele é multiplataforma, rodando em diversos sistemas operacionais e também extensível, com milhares de bibliotecas (conjuntos de módulos disponíveis com funcionalidades já disponíveis). E o principal, o Python possui fácil aprendizado, pois sua sintaxe é simples e foi criada para proporcionar agilidade, como consequência, dando maior produtividade ao usuário.

Existem vários pacotes (bibliotecas) estatísticos em Python que são amplamente utilizados para análise de dados e visualização. Alguns dos pacotes mais usados são:

- NumPy: é um pacote fundamental em Python para computação científica, que inclui funcionalidades para manipulação de arrays (matrizes) multidimensionais e cálculos matemáticos. Ele é amplamente utilizado para operações numéricas e estatísticas básicas.
- Pandas: é um pacote para análise de dados que fornece uma estrutura de dados flexível chamada DataFrame, que é semelhante a uma planilha do Excel. Ele é amplamente utilizado para importar, manipular, limpar e analisar dados.
- Matplotlib: é um pacote de visualização de dados em Python que permite criar gráficos e visualizações de dados de alta qualidade. Ele suporta vários tipos de gráficos, como linhas, barras, histogramas, dispersão, entre outros.
- Seaborn: é um pacote de visualização de dados em Python que fornece uma interface mais simplificada para criar gráficos estatísticos complexos. Ele é construído em cima do Matplotlib e oferece estilos de gráficos estatísticos pré-definidos.

Existem algumas outras opções de softwares disponíveis que também podem ser utilizados para o ensino de Estatística Descritiva. Pode-se destacar dois deles: R e o Calc/LibreOffice (planilhas eletrônicas similar ao Excel).

Ambos também são programas livres e possuem bastante material disponível para consulta na internet. A desvantagem de ambos é que as sintaxes das linguagens utilizadas são mais complexas, dificultando um pouco a parte do aprendizado da programação em si.

Consultando "software R" na base de dissertações do PROFMAT, existem 06 (seis) trabalhos os quais propuseram o uso do R (inclusive dois deles específicos para o ensino de estatística). Se consultado por "libreoffice calc", "planilha" ou "excel, existem 56 trabalhos os quais propuseram o uso de planilhas eletrônicas, sendo vários deles para o ensino de Estatística. A consulta foi realizada em 05/10/2023.

Portanto, além da melhor compreensão dos conceitos de estatística, o uso da ferramenta Python dará aos estudantes uma noção para a programação de computadores

com uma linguagem fácil e altamente utilizada nos dias atuais, também muito requisitada no mercado de trabalho.

5 Atividade Prática

A atividade prática proposta tem como objetivo servir de ideia para o docente, podendo ser utilizada e/ou adaptada de acordo com seus objetivos, buscando tornar o ensino mais dinâmico para os estudantes. A atividade sugerida envolve elementos do cotidiano, buscando promover a compreensão dos conceitos de estatística mencionados no Capítulo 2, utilizando Python como ferramenta para estimular um melhor entendimento por parte dos alunos de maneira prática.

Para a execução das atividades, o ideal seria utilizar o laboratório de informática da escola, com acesso à internet, e a exposição do professor utilizando o datashow. No entanto, sabe-se que nem todas as escolas do Brasil possuem computadores, e segundo o Censo Escolar 2022, apenas 43,1% das escolas que possuem Ensino Fundamental – Anos Finais e/ou Ensino Médio possuem laboratórios de informática. Além disso, de acordo com a PNAD Contínua - Acesso à Internet e à televisão e posse de telefone móvel celular para uso pessoal – 2021, apenas 32,7% dos domicílios possuem microcomputador.

Dessa forma, as limitações com a ausência de laboratório de informática e/ou internet podem ser contornadas de maneira parcial. Pode-se pedir aos alunos que preencham em casa ou no papel em sala de aula o questionário a partir do qual serão construídos os dados para a execução da atividade. O docente precisaria apenas de um computador (notebook) e um datashow para expor em sala de aula junto aos alunos. Neste caso, os alunos apenas assistirão à exposição do professor, ficando sem a prática no computador.

Para os docentes sem conhecimento das ferramentas utilizadas neste trabalho, é possível encontrar tutoriais e vídeos na internet que ensinam como manipulá-las. O tutorial de instalação do Python pode ser consultado no Apêndice B, e algumas sugestões de materiais gratuitos para o aprendizado da linguagem, tanto para os docentes quanto para os alunos que se interessarem em aprofundar na programação em Python, podem ser encontrados no Apêndice C.

5.1 Explorando os Dados da Pesquisa

Para iniciar o trabalho, o docente deve aplicar um questionário aos alunos de uma turma específica. Embora não seja exigido um número mínimo de respostas, é recomendável que haja pelo menos vinte respostas para permitir uma boa visualização dos dados. Recomenda-se a utilização do Google Forms, uma ferramenta gratuita e fácil de usar. O questionário deve incluir perguntas sobre sexo, peso, altura, tipo de transporte utilizado para chegar à escola e consumo de energia nos últimos seis meses em kWh.

Tabela 6 – Perguntas do questionário a serem respondidas pelo alunos da turma

Questionário
01. Qual é o seu sexo? <input type="checkbox"/> <i>Feminino</i> <input type="checkbox"/> <i>Masculino</i>
02. Qual é seu peso, em kg?
03. Qual é a sua altura, em metros?
04. Qual o tipo de transporte que utiliza para chegar a escola, na maioria das vezes? <input type="checkbox"/> <i>A pé</i> <input type="checkbox"/> <i>Ônibus</i> <input type="checkbox"/> <i>Motocicleta</i> <input type="checkbox"/> <i>Carro</i> <input type="checkbox"/> <i>Bicicleta</i> <input type="checkbox"/> <i>Transporte Escolar Particular</i> <input type="checkbox"/> <i>Outro</i>
05. Quanto foi o consumo de energia (em kW/h) da sua residência, há seis meses atrás?
06. Quanto foi o consumo de energia (em kW/h) da sua residência, há cinco meses atrás?
07. Quanto foi o consumo de energia (em kW/h) da sua residência, há quatro meses atrás?
08. Quanto foi o consumo de energia (em kW/h) da sua residência, no antepenúltimo mês?
09. Quanto foi o consumo de energia (em kW/h) da sua residência, no penúltimo mês?
10. Quanto foi o consumo de energia (em kW/h) da sua residência, no último mês?

Fonte: Elaboração própria

A partir deste momento, serão utilizadas bibliotecas do Python. Primeiramente, é necessário carregá-las para permitir o uso das diversas funções que serão utilizadas, conforme mostrado na Figura 13.

Figura 13 – Carregando as bibliotecas

```
# Importando Bibliotecas
import pandas as pd
import plotly.express as px
import statistics as sta
import seaborn as sns
```

Fonte: Elaboração própria

O Google Forms possui a opção de exportar as respostas organizadas em uma planilha em formato CSV. Dessa forma, é possível importar os dados coletados para o Python, utilizando a programação apresentada na Figura 14.

Figura 14 – Carregando as respostas do questionário

```
# Carrega os dados respondidos
dados = pd.read_csv('pesquisa1.csv', sep=';')
tabela = pd.DataFrame(dados)
display(tabela)
```

Fonte: Elaboração própria

Para ilustrar o processo, foram geradas aleatoriamente respostas ao questionário, simulando uma turma de 36 alunos. A tabela com os dados pode ser encontrada no Apêndice D. Seguindo os conceitos citados no capítulo 2, tem-se que:

- População: 36 alunos da turma.
- Amostra: 12 alunos da turma (por exemplo).

Figura 15 – Selecionando uma amostra com 12 respostas

```
# Seleciona uma Amostra de tamanho 12
print(dados.sample(n=12))
```

Fonte: Elaboração própria

- Variáveis:
 - > Qualitativas: Sexo e Tipo de transporte
 - > Quantitativas: Peso, Altura e Consumo de energia
- Medidas de Posição e Dispersão: Facilmente é possível calcular cada medida individualmente de maneira "manual". Como exemplo, será utilizada a variável Altura.

Figura 16 – Linhas de código - Medidas calculadas - Programadas manualmente

```
#Medidas de Posição e Dispersão
media = tabela["Altura"].mean()
mediana = tabela["Altura"].median()
moda = sta.mode(tabela["Altura"])
desvio_medio = tabela["Altura"].mad()
variancia = tabela["Altura"].var()
desvio_padrao = tabela["Altura"].std()
print("Média:", media)
print("Mediana:", mediana)
print("Moda:", moda)
print("Desvio Médio", desvio_medio)
print("Variância:", variancia)
print("Desvio Padrão:", desvio_padrao)
```

Fonte: Elaboração própria

Figura 17 – Medidas calculadas - Programadas manualmente

```
Média: 1.6438888888888887
Mediana: 1.66
Moda: 1.66
Desvio Médio 0.1031172839506173
Variância: 0.014527301587301584
Desvio Padrão: 0.12052925614680272
```

Fonte: Elaboração própria

- Tabelas de Frequências:
 - > Construindo a partir da variável Tipo de transporte.

Figura 18 – Linhas de código - Tabela de frequência - Variável qualitativa

```
#Tabela de Variável Qualitativa
frequencia = tabela['Transporte'].value_counts()
percentual = tabela['Transporte'].value_counts(normalize = True)*100
df = pd.DataFrame({"Frequência Absoluta": frequencia, "Frequência Relativa (%)":percentual})
df.rename_axis("Transporte", axis = "columns", inplace=True)
f.loc['Total'] = df.sum()
df
```

Fonte: Elaboração própria

Figura 19 – Tabela gerada - Tabela de frequência - Variável qualitativa

Transporte	Frequência Absoluta	Frequência Relativa (%)
Carro	16.0	44.444444
Motocicleta	7.0	19.444444
A pé	7.0	19.444444
Bicicleta	4.0	11.111111
Ônibus	2.0	5.555556
Total	36.0	100.000000

Fonte: Elaboração própria

-> Construindo a partir da variável Altura. Para determinar a quantidade de classes, será utilizado o critério de (BUSSAB; MORETTIN, 2010). Assim, serão consideradas cinco classes e cada uma com amplitude de 0.10 cm, iniciando por 1.40 cm.

Figura 20 – Linhas de código - Tabela de frequência - Variável quantitativa

```
# Tabela de Variável Contínua
int1 = tabela.query("1.40 <= Altura < 1.50")["Altura"]
cont_int1 = int1.value_counts().sum()
int2 = tabela.query("1.50 <= Altura < 1.60")["Altura"]
cont_int2 = int2.value_counts().sum()
int3 = tabela.query("1.60 <= Altura < 1.70")["Altura"]
cont_int3 = int3.value_counts().sum()
int4 = tabela.query("1.70 <= Altura < 1.80")["Altura"]
cont_int4 = int4.value_counts().sum()
int5 = tabela.query("1.80 <= Altura < 1.90")["Altura"]
cont_int5 = int5.value_counts().sum()
cont_total = cont_int1 + cont_int2 + cont_int3 + cont_int4 + cont_int5
mont_tab = {"Altura":["1.40 |– 1.50",
"1.50 |– 1.60",
"1.60 |– 1.70",
"1.70 |– 1.80",
"1.80 |– 1.90",
"Total"],
"Frequência Absoluta": [cont_int1, cont_int2, cont_int3, cont_int4,
cont_int5, cont_total],
"Frequência Relativa": [cont_int1*100 / cont_total, cont_int2*100 / cont_total,
cont_int3*100 / cont_total, cont_int4*100 / cont_total,
cont_int5*100 / cont_total, cont_total*100 / cont_total]}
tab_freq = pd.DataFrame(data = mont_tab)
tab_freq
```

Fonte: Elaboração própria

Figura 21 – Tabela gerada - Tabela de frequência - Variável quantitativa

	Altura	Frequência Absoluta	Frequência Relativa
0	1.40 -- 1.50	5	13.888889
1	1.50 -- 1.60	8	22.222222
2	1.60 -- 1.70	11	30.555556
3	1.70 -- 1.80	7	19.444444
4	1.80 -- 1.90	5	13.888889
5	Total	36	100.000000

Fonte: Elaboração própria

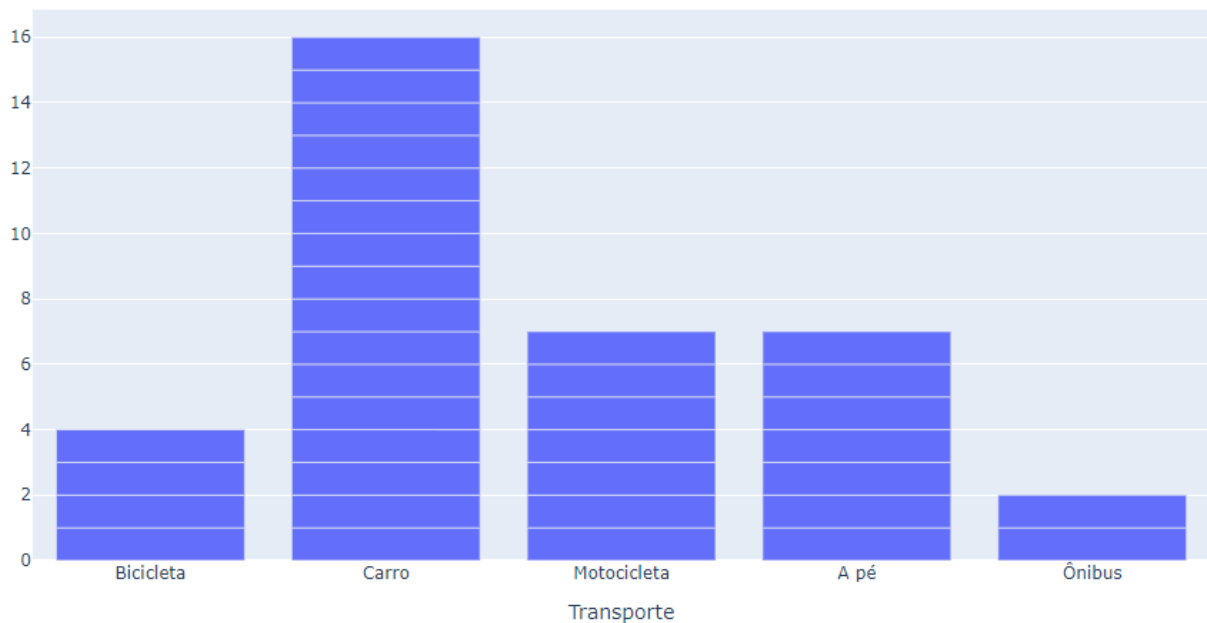
- Gráficos: Cada variável se encaixa de maneira mais harmônica em determinado tipo de gráfico. Assim, tem-se que:
 - > Gráfico de barras: Construindo a partir da variável Tipo de transporte.

Figura 22 – Linhas de código - Gráfico de barras

```
#Gráfico de Barras
gbarra = px.bar(tabela,x=tabela["Transporte"])
gbarra.show()
```

Fonte: Elaboração própria

Figura 23 – Gráfico gerado - Gráfico de barras



Fonte: Elaboração própria

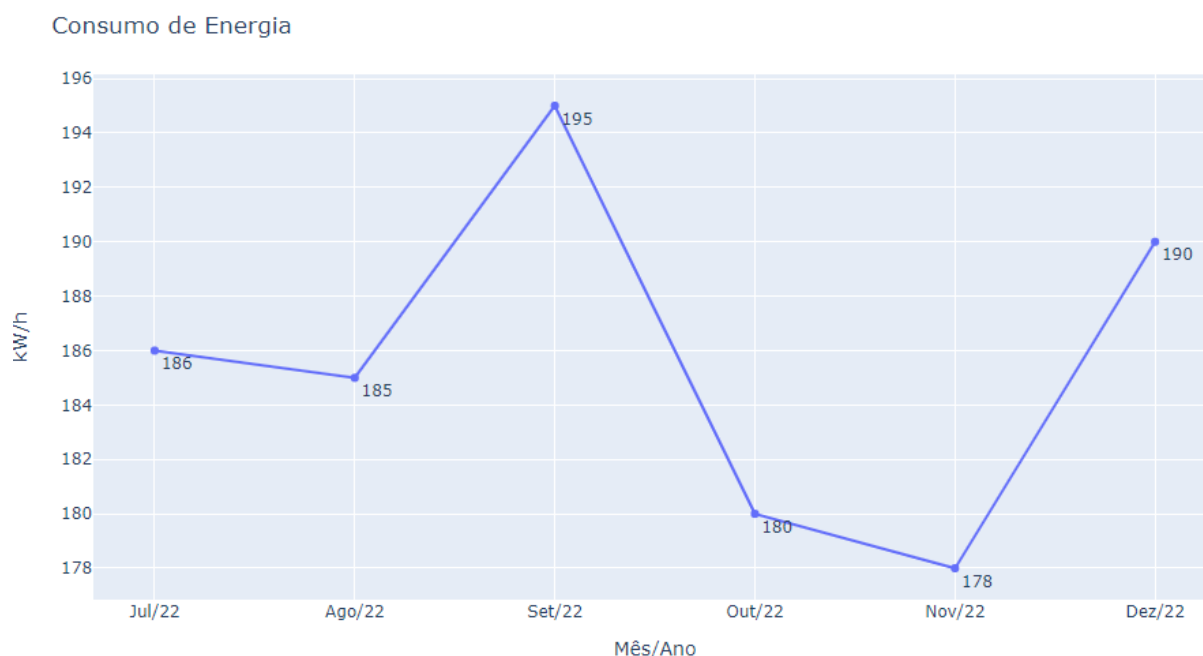
-> Gráfico de linhas: Construindo a partir dos dados da variável Consumo de energia da primeira resposta simulada.

Figura 24 – Linhas de código - Gráfico de linhas

```
#Gráfico de Linhas
glinha = px.line(tabela, x=["Jul/22","Ago/22","Set/22","Out/22","Nov/22","Dez/22"],
y=[186,185,195,180,178,190], text=["186","185","195","180","178","190"])
glinha.update_traces(textposition="bottom right")
glinha.update_layout(title="Consumo de Energia", xaxis_title="Mês/Ano",
yaxis_title="kW/h")
glinha.show()
```

Fonte: Elaboração própria

Figura 25 – Gráfico gerado - Gráfico de linhas



Fonte: Elaboração própria

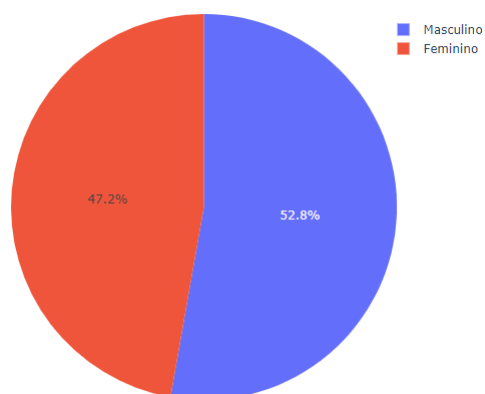
-> Gráfico de pizza: Construindo a partir da variável Sexo.

Figura 26 – Linhas de código - Gráfico de pizza

```
#Gráfico de Pizza
gpizza = px.pie(tabela, values=tabela["Sexo"].value_counts().values,
names = ["Masculino","Feminino"])
gpizza.show()
```

Fonte: Elaboração própria

Figura 27 – Gráfico gerado - Gráfico de pizza



Fonte: Elaboração própria

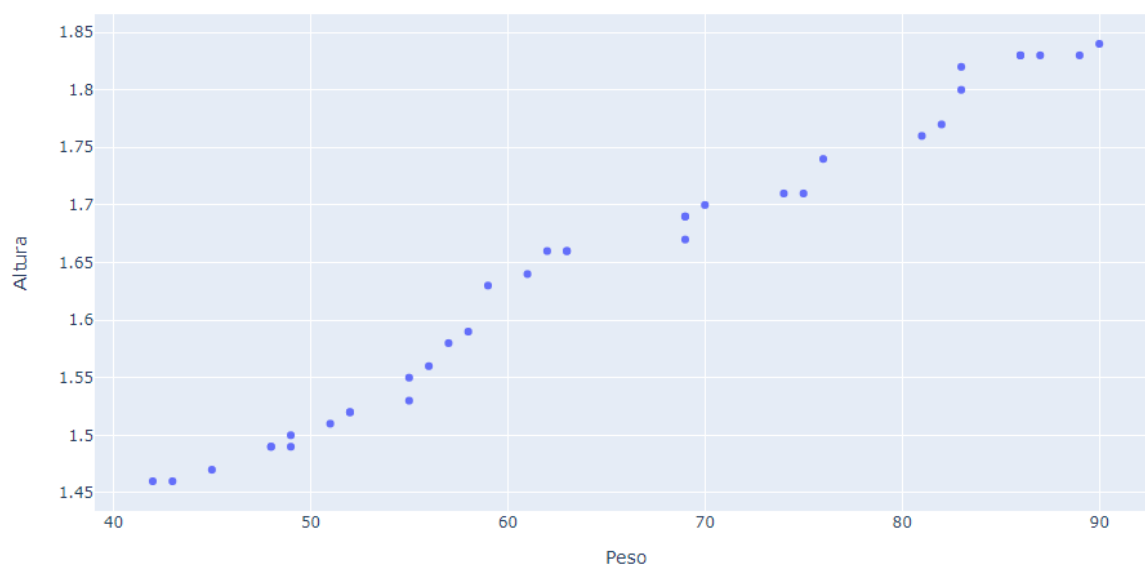
-> Gráfico de dispersão: Construindo a partir das variáveis Peso e Altura.

Figura 28 – Linhas de código - Gráfico de dispersão

```
#Gráfico de Dispersão
px.scatter(tabela, x="Peso", y="Altura")
```

Fonte: Elaboração própria

Figura 29 – Gráfico gerado - Gráfico de dispersão



Fonte: Elaboração própria

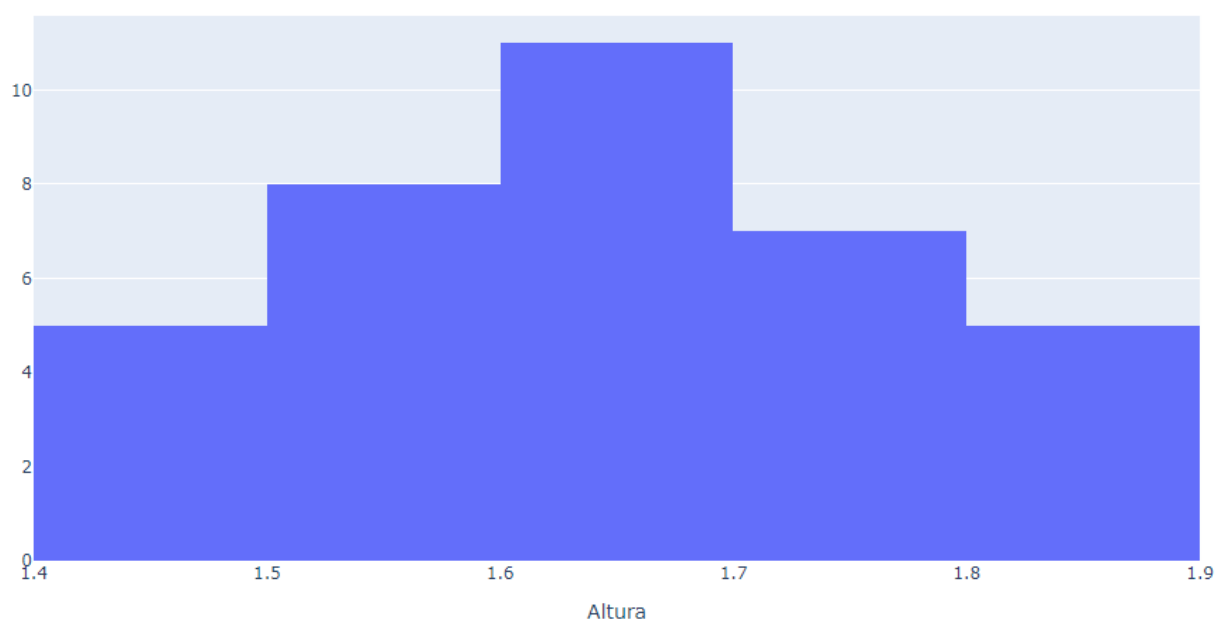
-> Histograma: Construindo a partir da variável Altura.

Figura 30 – Linhas de código - Histograma

```
#Histograma  
hist = px.histogram(tabela, x="Altura", nbins=5)  
hist.show()
```

Fonte: Elaboração própria

Figura 31 – Gráfico gerado - Histograma



Fonte: Elaboração própria

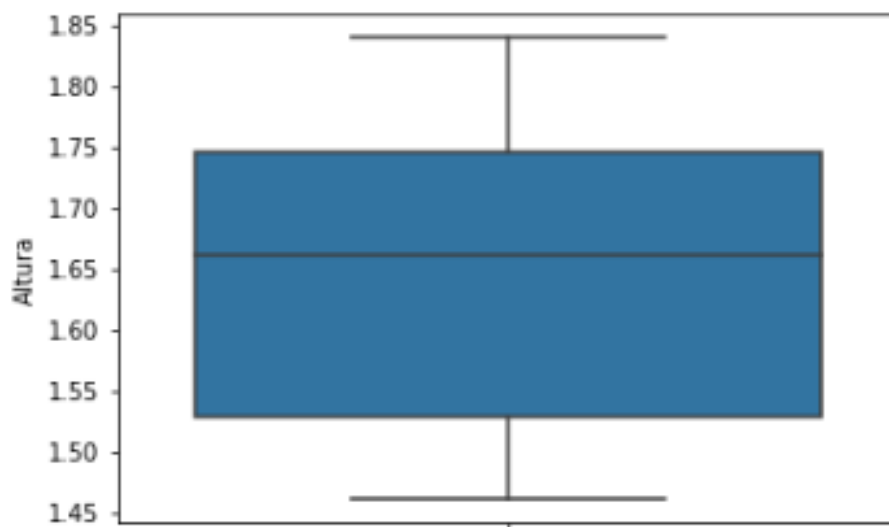
-> Box plot: Construindo a partir da variável Altura.

Figura 32 – Linhas de código - Box plot

```
#Box plot  
sns.boxplot(y=tabela["Altura"])
```

Fonte: Elaboração própria

Figura 33 – Gráfico gerado - Box plot



Fonte: Elaboração própria

6 Considerações Finais

Apesar das inúmeras situações presentes no cotidiano, o aprendizado da Estatística pelos estudantes da educação básica ainda deixa a desejar, de acordo com os resultados do Saeb 2021.

Vale ressaltar que o uso de tecnologias deve ser balanceado com outras abordagens de ensino, como as explicações orais e práticas escritas, visando garantir que os estudantes desenvolvam as habilidades de maneira completa.

A expectativa é que este trabalho possa contribuir para um melhor aprendizado dos alunos com relação aos temas básicos da Estatística, além de enriquecer os métodos de ensino de estatística na educação básica.

Uma sugestão para trabalhos futuros seria aplicar a programação em Python em uma atividade referente ao tema da probabilidade. Por exemplo, simular um número elevado de lançamentos de uma moeda para explicar o conceito de probabilidade frequentista.

Outra sugestão seria o docente separar suas turmas do mesmo nível em dois grupos: controle e experimental. No grupo experimental, o professor utilizaria a metodologia proposta neste trabalho, enquanto no grupo controle seria utilizado o método de ensino normalmente utilizado pelo próprio professor. Ao final, seria feita uma avaliação do desempenho das turmas e seriam utilizados testes estatísticos para verificar se houve diferença estatística entre o desempenho de aprendizado dos grupos.

Referências

- BARBOSA, Maria Tereza Serrano; VELASQUE, Luciane de Souza; SILVA, Alexandre Sousa. **O Letramento Estatístico na Formação dos Professores: Um Tutorial Metodológico**. VIDYA, v. 36, n. 2, p. 397-408, 2016.
- BOLFARINE, Heleno; BUSSAB, Wilton O. **Elementos de Amostragem**. Blucher, 2005.
- BRASIL. Ministério da Educação. **Base Nacional Comum Curricular**. Disponível em http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versa_oficial_site.pdf. Acesso em: 11 maio 2023.
- COUTINHO, Cileda de Queiroz Silva; SPINA, Gabriela. **A Estatística nos Livros Didáticos de Ensino Médio**. Statistics in Brazilian's High School Books. Ensino da Matemática em Debate. ISSN 2358-4122, v. 2, n. 2, 2016.
- DA SILVA, Adaias Milhomem et al. **Sequência Didática Aplicada ao Ensino da Estatística Descritiva**. Revista Destaques Acadêmicos, v. 13, n. 4, 2022.
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do Saeb 2021**. Disponível em <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/saeb>. Acesso em abr.2023.
- MILONE, Giuseppe. **Estatística: geral e aplicada**. Pioneira Thomson Learning, 2004.
- MORETTIN, Pedro A.; BUSSAB, Wilton O. **Estatística básica**. Saraiva Educação SA, 2010.
- Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) - Acesso à Internet e à televisão e posse de telefone móvel celular para uso pessoal - 2021**. Disponível em: <https://sidra.ibge.gov.br/tabela/7339#resultado>. Acesso em: 11 maio 2023.
- SARAIVA, J. C. C.; VICTER, E. F.; SIQUEIRA, A. S. **SISTAT: Ferramenta Computacional como Proposta para o Ensino da Estatística Descritiva**. Revista de Educação, Ciências e Matemática, 7(1), 2017.
- WATSON, Jane; CALLINGHAM, Rosemary. **statistical literacy: A complex hierarchical construct**. Statistics Education Research Journal, v. 2, n. 2, p. 3-46, 2003.
- WIKIPÉDIA. **Ciência de dados**. Disponível em: https://pt.wikipedia.org/wiki/Ciência_de_dados. Acesso em: 11 maio 2023.

APÊNDICE A – Escalas de Proficiências em Tratamento de Informações (Prob. e Estatística) consideradas no Saeb

Tabela 7 – Escalas de Proficiências em Tratamento de Informações (Probabilidade e Estatística) consideradas no Saeb do 5º Ano – EF.

Nível	Habilidades - 5º Ano - EF
0	Os estudantes localizados abaixo do desempenho de 125 requerem atenção especial, pois não demonstram habilidades muito elementares.
1	Os estudantes sequer sabem lidar com TRATAMENTO DE INFORMAÇÕES.
2	Localizar informações, relativas ao maior ou menor elemento, em tabelas ou gráficos.
3	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Reconhecer o maior valor em uma tabela de dupla entrada cujos dados possuem até duas ordens. • Reconhecer informações em um gráfico de colunas duplas.
4 e 5	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Reconhecer o maior valor em uma tabela cujos dados possuem até oito ordens. • Localizar um dado em tabelas de dupla entrada.
6	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Interpretar dados em uma tabela simples. • Comparar dados representados pelas alturas de colunas presentes em um gráfico.
7	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Interpretar dados em gráficos de setores.
8	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Interpretar dados em um gráfico de colunas duplas.
9 e 10	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Reconhecer o gráfico de linhas correspondente a uma sequência de valores ao longo do tempo (com valores positivos e negativos).

Fonte: Inep/Saeb - Adaptado

Tabela 8 – Escalas de Proficiências em Tratamento de Informações (Probabilidade e Estatística) consideradas no Saeb do 9º Ano – EF.

Nível	Habilidades - 9º Ano - EF
0	Os estudantes localizados abaixo do desempenho de 200 requerem atenção especial, pois não demonstram habilidades muito elementares.
1	Interpretar dados apresentados em tabela e gráfico de colunas.
2	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Interpretar dados apresentados em um gráfico de linha simples. • Associar dados apresentados em gráfico de colunas a uma tabela.
3	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Associar dados apresentados em tabela a gráfico de setores. • Analisar dados dispostos em uma tabela simples. • Analisar dados apresentados em um gráfico de linha com mais de uma grandeza representada.
4 e 5	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Analisar dados dispostos em uma tabela de dupla entrada.
6	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Resolver problemas que requerem a comparação de dois gráficos de colunas.
7, 8 e 9	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Determinar a média aritmética de um conjunto de valores. • Estimar quantidades em gráficos de setores. • Analisar dados dispostos em uma tabela de três ou mais entradas. • Interpretar dados fornecidos em gráficos envolvendo regiões do plano cartesiano. • Interpretar gráficos de linhas com duas sequências de valores.

Fonte: Inep/Saeb - Adaptado

Tabela 9 – Escalas de Proficiências em Tratamento de Informações (Probabilidade e Estatística) consideradas no Saeb do 3ª/4ª Série - EMT + EMI.

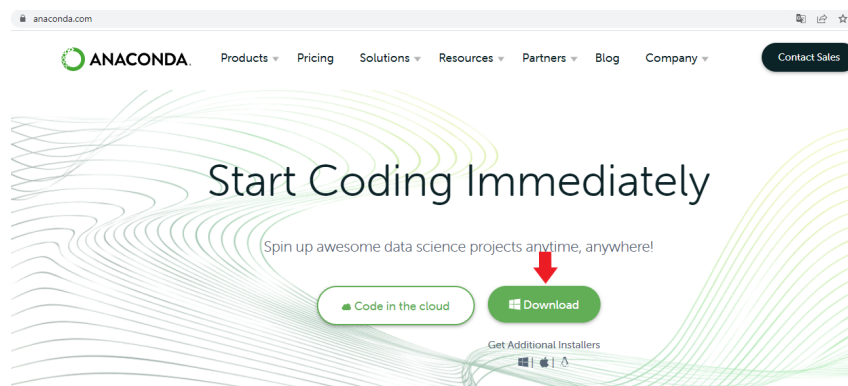
Nível	Habilidades - 3ª/4ª Série - EMT + EMI
0	Os estudantes localizados abaixo do desempenho de 225 requerem atenção especial, pois não demonstram habilidades muito elementares.
1	Associar uma tabela de até duas entradas a informações apresentadas textualmente ou em um gráfico de barras ou de linhas.
2 até 10	Além das habilidades anteriormente citadas, os estudantes provavelmente são capazes de: <ul style="list-style-type: none"> • Associar um gráfico de setores a dados percentuais apresentados textualmente ou em uma tabela.

Fonte: Inep/Saeb - Adaptado

APÊNDICE B – Tutorial de Instalação e Primeiro Acesso do Python

Primeiramente, deve se instalar o ANACONDA. Acessar <https://www.anaconda.com/> e solicitar o download do instalador (o padrão é para Windows). Existem também as opções para MAC e Linux.

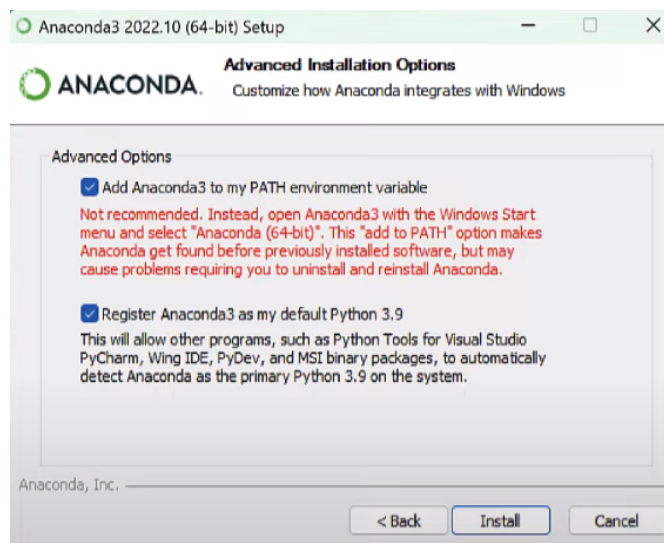
Figura 34 – Página para download do ANACONDA.



Fonte: anaconda.com

Em seguida, execute o arquivo .exe que foi baixado. Basta seguir a instalação normalmente. Ao chegar na tela a seguir contida na Figura 35, deve ficar marcadas as duas opções.

Figura 35 – Instalação do ANACONDA.

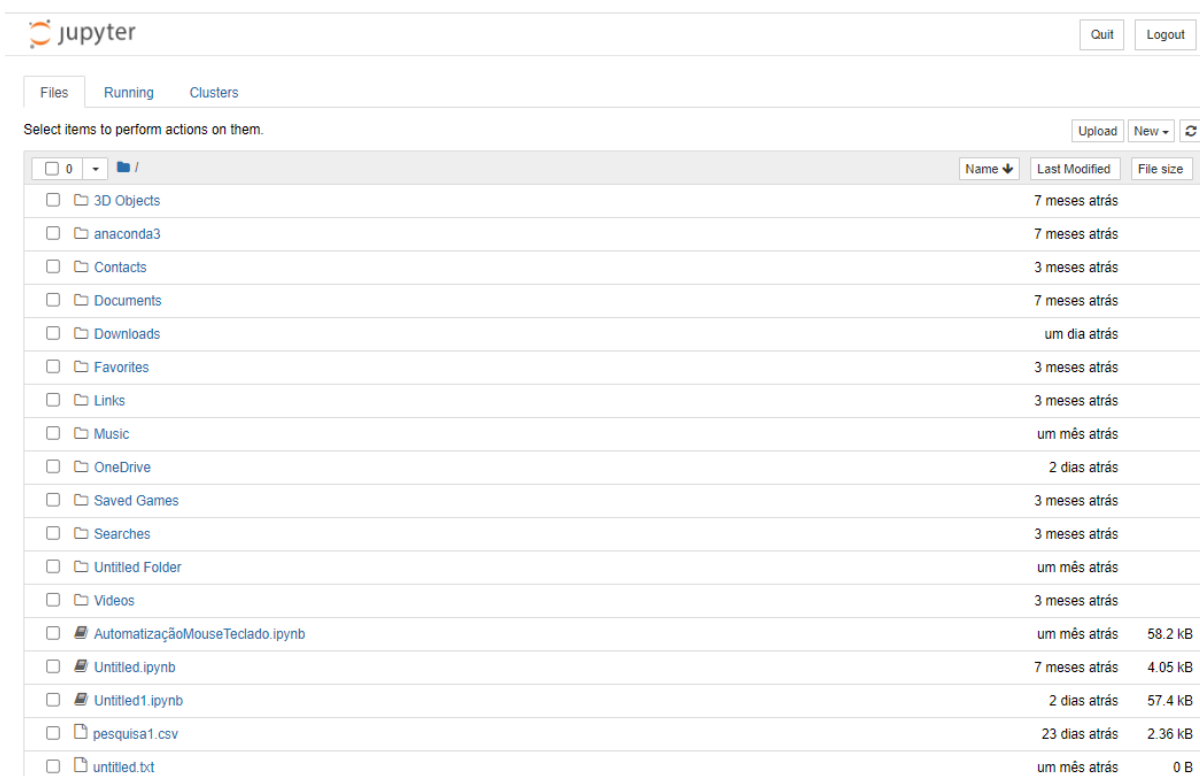


Fonte: Elaboração própria

Depois, basta seguir os passos normalmente e concluir a instalação.

Para abrir o programa, digite na barra de pesquisa do Windows "jupyter notebook". Ao executar o programa será aberta uma aba em seu navegador de internet como se fosse uma lista de arquivos e pastas, conforme a Figura 36.

Figura 36 – Tela após a execução do Jupyter Notebook.



Fonte: Elaboração própria

Para finalizar, basta agora criar um novo arquivo .ipynb clicando em "new" e em seguida em "Python 3 (ipykernel)" para que possam ser executados os códigos de programação em Python.

Figura 37 – Tela para abrir novo arquivo .ipynb.



Fonte: Elaboração própria

APÊNDICE C – Cursos Gratuitos para o Aprendizado de Python

No link abaixo, é possível consultar um minicurso de Python da hashtagtreinamentos.com. Um curso introdutório bastante didático.

<https://www.hashtagtreinamentos.com/curso-python>

No link abaixo, é possível acessar um curso básico de 72 horas de Fundamentos de Linguagem Python Para Análise de Dados e Data Science. Um curso introdutório robusto e bem completo. Inclusive, possui emissão de certificado.

<https://www.datascienceacademy.com.br/cursosgratuitos>

APÊNDICE D – Respostas geradas simulando uma turma com 36 alunos

Tabela 10 – Dados gerados aleatoriamente.

Sexo	Peso	Altura	Transporte	Ener.Jul/22
Masculino	42	1.46	Bicicleta	186.0
Feminino	43	1.46	Carro	250.0
Feminino	45	1.47	Motocicleta	269.0
Feminino	48	1.49	Carro	108.0
Masculino	48	1.49	Carro	127.0
Masculino	49	1.49	Carro	194.0
Feminino	49	1.50	Carro	267.0
Masculino	51	1.51	Motocicleta	269.0
Feminino	52	1.52	Bicicleta	117.0
Masculino	52	1.52	A pé	268.0
Masculino	55	1.53	Carro	102.0
Feminino	55	1.55	Carro	237.0
Feminino	56	1.56	Carro	267.0
Masculino	57	1.58	Ônibus	213.0
Masculino	58	1.59	Carro	159.0
Masculino	59	1.63	Carro	113.0
Masculino	61	1.64	Bicicleta	118.0
Feminino	62	1.66	Bicicleta	125.0
Masculino	63	1.66	Motocicleta	213.0
Feminino	63	1.66	Carro	230.0
Masculino	69	1.67	Motocicleta	107.0
Masculino	69	1.69	Carro	220.0
Masculino	69	1.69	Carro	261.0
Masculino	70	1.70	Carro	111.0
Feminino	74	1.71	Motocicleta	166.0
Feminino	75	1.71	Motocicleta	287.0
Feminino	76	1.74	Carro	188.0
Masculino	81	1.76	A pé	218.0
Masculino	82	1.77	A pé	244.0
Feminino	83	1.80	Ônibus	144.0
Masculino	83	1.82	A pé	294.0
Feminino	86	1.83	A pé	184.0
Masculino	86	1.83	A pé	196.0
Feminino	87	1.83	Carro	217.0
Feminino	89	1.83	A pé	285.0
Feminino	90	1.84	Motocicleta	262.0

Fonte: Elaboração própria

Tabela 11 – Continuação - Dados gerados aleatoriamente.

Ener.Ago/22	Ener.Set/22	Ener.Out/22	Ener.Nov/22	Ener.Dez/22
185.0	195.0	180.0	178.0	190.0
226.0	244.0	233.0	225.0	241.0
245.0	273.0	275.0	266.0	287.0
32.0	120.0	121.0	115.0	123.0
160.0	132.0	151.0	148.0	174.0
193.0	209.0	190.0	191.0	201.0
239.0	273.0	267.0	263.0	280.0
264.0	279.0	277.0	266.0	290.0
153.0	126.0	129.0	138.0	162.0
240.0	273.0	274.0	263.0	281.0
105.0	102.0	114.0	102.0	114.0
224.0	241.0	229.0	222.0	227.0
237.0	272.0	256.0	256.0	267.0
196.0	227.0	193.0	192.0	208.0
165.0	158.0	156.0	158.0	186.0
138.0	126.0	123.0	130.0	153.0
154.0	127.0	136.0	141.0	170.0
158.0	132.0	140.0	144.0	174.0
196.0	227.0	193.0	192.0	208.0
215.0	239.0	227.0	211.0	226.0
121.0	107.0	117.0	111.0	119.0
213.0	236.0	210.0	205.0	225.0
228.0	247.0	243.0	231.0	247.0
137.0	120.0	122.0	128.0	152.0
168.0	166.0	160.0	163.0	186.0
281.0	293.0	289.0	289.0	295.0
187.0	206.0	183.0	186.0	197.0
211.0	230.0	209.0	199.0	224.0
225.0	244.0	231.0	225.0	237.0
164.0	151.0	152.0	158.0	182.0
282.0	300.0	294.0	291.0	297.0
179.0	180.0	161.0	176.0	186.0
194.0	226.0	191.0	191.0	203.0
208.0	230.0	196.0	196.0	220.0
272.0	280.0	288.0	267.0	291.0
233.0	265.0	256.0	239.0	265.0

Fonte: Elaboração própria

APÊNDICE E – Linhas de códigos utilizadas em formato texto

```
# Importando Bibliotecas
import pandas as pd
import plotly.express as px
import statistics as sta
import seaborn as sns

# Carrega os dados respondidos
dados = pd.read_csv("pesquisa1.csv", sep=";")
tabela = pd.DataFrame(dados)
display(tabela)

# Seleciona uma Amostra de tamanho 12
print(dados.sample(n=12))

#Medidas de Posição e Dispersão
media = tabela["Altura"].mean()
mediana = tabela["Altura"].median()
moda = sta.mode(tabela["Altura"])
desvio_medio = tabela["Altura"].mad()
variância = tabela["Altura"].var()
desvio_padrao = tabela["Altura"].std()
print("Média:", media)
print("Mediana:", mediana)
print("Moda:", moda)
print("Desvio Médio:", desvio_medio)
print("Variância:", variância)
print("Desvio Padrão:", desvio_padrao)
```

```
#Tabela de Variável Discreta
frequencia = tabela["Transporte"].value_counts()
percentual = tabela["Transporte"].value_counts(normalize = True)*100
df = pd.DataFrame({"Frequência Absoluta": frequencia,
                  "Frequência Relativa(%)":percentual})
df.rename_axis("Transporte", axis = "columns", inplace=True)
df.loc["Total"] = df.sum()
df

# Tabela de Variável Contínua
int1 = tabela.query("1.40 <= Altura < 1.50")["Altura"]
cont_int1 = int1.value_counts().sum()
int2 = tabela.query("1.50 <= Altura < 1.60")["Altura"]
cont_int2 = int2.value_counts().sum()
int3 = tabela.query("1.60 <= Altura < 1.70")["Altura"]
cont_int3 = int3.value_counts().sum()
int4 = tabela.query("1.70 <= Altura < 1.80")["Altura"]
cont_int4 = int4.value_counts().sum()
int5 = tabela.query("1.80 <= Altura < 1.90")["Altura"]
cont_int5 = int5.value_counts().sum()
cont_total = cont_int1 + cont_int2 + cont_int3 + cont_int4 + cont_int5
mont_tab = {"Altura":["1.40 |— 1.50",
                    "1.50 |— 1.60",
                    "1.60 |— 1.70",
                    "1.70 |— 1.80",
                    "1.80 |— 1.90",
                    "Total"],
            "Frequência Absoluta": [cont_int1, cont_int2, cont_int3, cont_int4,
                                    cont_int5, cont_total],
            "Frequência Relativa": [cont_int1*100 / cont_total, cont_int2*100 / cont_total,
                                    cont_int3*100 / cont_total, cont_int4*100 / cont_total,
                                    cont_int5*100 / cont_total, cont_total*100 / cont_total]}
tab_freq = pd.DataFrame(data = mont_tab)
tab_freq

#Gráfico de Barras
gbarra = px.bar(tabela,x=tabela["Transporte"])
gbarra.show()
```

```
#Gráfico de Linhas
```

```
glinha = px.line(tabela, x=["Jul/22", "Ago/22", "Set/22", "Out/22", "Nov/22", "Dez/22"],  
y=[186,185,195,180,178,190], text=["186", "185", "195", "180", "178", "190"])  
glinha.update_traces(textposition="bottom right")  
glinha.update_layout(title="Consumo de Energia", xaxis_title="Mês/Ano",  
yaxis_title="kW/h")  
glinha.show()
```

```
#Gráfico de Pizza
```

```
gpizza = px.pie(tabela, values=tabela["Sexo"].value_counts().values,  
names = ["Masculino", "Feminino"])  
gpizza.show()
```

```
#Gráfico de Dispersão
```

```
px.scatter(tabela, x="Peso", y="Altura")
```

```
#Histograma
```

```
hist = px.histogram(tabela, x="Altura", nbins=5)  
hist.show()
```

```
#Box plot
```

```
sns.boxplot(y=tabela["Altura"])
```