



UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE ABAETETUBA
FACULDADE DE CIÊNCIAS EXATAS E TECNOLOGIA
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL

MICHEL CHARLES DA SILVA MOURA

**USO DO SOFTWARE BIOESTAT COMO MÉTODO EDUCATIVO
PARA O ENSINO E APRENDIZAGEM DA ESTATÍSTICA NA
EDUCAÇÃO BÁSICA**

ABAETETUBA-PA

2023



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARÁ
POS-GRADUACAO EM MATEMATICA EM REDE NACIONAL-ABAETETUBA

FOLHA DE APROVAÇÃO DE DISSERTAÇÃO Nº 2/2023 - PFMAB (11.09.37)

Nº do Protocolo: 23073.063854/2023-43

Abaetetuba-PA, 04 de setembro de 2023.

FOLHA DE APROVAÇÃO

MICHEL CHARLES DA SILVA MOURA

USO DO SOFTWARE BIOESTAT COMO MÉTODO EDUCATIVO PARA O ENSINO E
PARENDIZAGEM DA ESTATÍSTICA NA EDUCAÇÃO BÁSICA

Dissertação apresentada como requisito final para a obtenção do
título de Mestre em Matemática, pela Universidade Federal do
Pará

Aprovada em 09 de agosto de 2023.

Banca Examinadora:

(Assinado digitalmente em 04/09/2023 22:28)
ANTONIO MAIA DE JESUS CHAVES NETO
PROFESSOR DO MAGISTERIO SUPERIOR
ICEN (11.34)
Matrícula: ###529#2

(Assinado digitalmente em 04/09/2023 20:36)
JOSE FRANCISCO DA SILVA COSTA
PROFESSOR DO MAGISTERIO SUPERIOR
CABAE (11.09)
Matrícula: ###277#6

(Assinado digitalmente em 06/09/2023 17:05)
ROMULO CORREA LIMA
PROFESSOR DO MAGISTERIO SUPERIOR
CABAE (11.09)
Matrícula: ###046#3

(Assinado digitalmente em 11/09/2023 15:35)
SEBASTIAO MARTINS SIQUEIRA CORDEIRO
PROFESSOR DO MAGISTERIO SUPERIOR
CABAE (11.09)
Matrícula: ###496#3

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

M929u Moura, Michel Charles da Silva.
USO DO SOFTWARE BIOESTAT COMO MÉTODO
EDUCATIVO PARA O ENSINO E APRENDIZAGEM DA
ESTATÍSTICA NA EDUCAÇÃO BÁSICA / Michel Charles da
Silva Moura. — 2023.
110 f. : il. color.

Orientador(a): Prof. Dr. José Francisco da Silva Costa
Dissertação (Mestrado) - Universidade Federal do Pará,
Campus Universitário de Abaetetuba, Programa de Pós-Graduação
em Matemática em Rede Nacional, Abaetetuba, 2023.

1. Software BioEstat. I. Título.

CDD 510

Dedico este trabalho aos meus pais (*In memoriam*), à minha esposa e ao meu filho, por terem sido componente expressiva nas solenes realizações da minha vida.

AGRADECIMENTOS

Primeiramente, a Deus por ter concedido a vida que anima este corpo que vos escreve.

Em especial à minha irmã que sempre priorizou a educação dos que estão sob sua guarda, e que me oportunizou fazer essa caminhada que começamos na escola.

À minha esposa e ao meu filho que toleraram todas as adversidades inerentes a essa construção científica escrita.

À meu orientador nessa empreitada, professor Dr. José Francisco, que compartilhou sua vasta experiência em produções científicas e me dirigiu com a expertise de seu conhecimento físico-matemático para o surgimento deste trabalho.

À UFPA, campus de Abaetetuba, que se mantém credenciada a oferecer esse programa com os bons resultados de seus mestrados na Qualificação Nacional por meio do seu proficiente corpo docente - reponsável pela pós graduação desses alunos. Desta forma, dá oportunidade para que professores da Educação Básica dessa região possam realizar um mestrado *stricto sensu* tão bem conceituado nacionalmente.

Ao ex-coordenador, professor Rômulo Corrêa, que coordenava quando do meu ingresso no curso, e ao professor Manuel Costa que estava coordenador durante minha saída, obrigado por terem sido sempre muito solícitos comigo.

Aos meus colegas da turma PROFMAT 2021 – Abaetetuba, nos uníamos pra estudar sempre que algum complicado problema se apresentava para nos por à prova.

"O matemático, distraído naquele rio de símbolos, lidando tão somente com verdades formais, talvez ache resultados de incalculável importância na descrição que fazemos do universo."

(Karl Pearson)

RESUMO

Este trabalho tem como objetivo compreender o uso do *software* BioEstat como método educativo com ênfase no estudo da estatística para o ensino básico, além de instigar os alunos a fazerem pesquisa estatística usufruindo do processamento computacional dos dados. O estudo sugere ainda que a partir de conceitos estatísticos básicos aplicados em informações numéricas significativas, e com o auxílio de recursos tecnológicos, como planilhas eletrônicas e aplicativos estatísticos, se possa investigar tendências, padrões, pontos de inflexão ou associação entre variáveis, comportamentos imperceptíveis quando não são apreciados pormenorizadamente pela ciência. Utiliza-se como metodologia, a de caráter bibliográfica, empregada na coleta e pesquisa dos dados climáticos de longo prazo, coletados no INMET (Instituto Nacional de Meteorologia), como temperatura máxima e mínima, precipitação acumulada, umidade relativa do ar, nebulosidade, insolação total, intensidade vento e pressão atmosférica, para avaliação e aplicação de técnicas estatísticas no cálculo de médias, desvios padrão e outras medidas relevantes na análise de um conjunto de valores que se pretende interpretar. Com base nesses padrões, recorre-se ao método estatístico para empreender a análise exploratória dos dados, efetuar a geração de tabelas de frequências e ajustar modelos matemáticos às séries temporais. Conclui-se a pesquisa considerando que há uma notável eficiência do *software* BioEstat na educação devido ao seu acessível manual prático que contém exemplos de entrada de dados para obtenção de gráficos, medidas descritivas, realização de testes de regressão, dentre outras aplicações. E, ao mesmo tempo, serve de apoio didático para o ensino, especificamente para a 3ª série do ensino médio, onde o professor, intermediando o aprendizado dos conteúdos dessas habilidades, e com o uso da tecnologia disponível nesse contexto, pode motivar o interesse do aluno no tratamento da informação que demandam interpretação e análises de dados estatísticos.

Palavras-chave: BioEstat, estatística, ensino.

ABSTRACT

This work aims to understand the use of the BioEstat software as an educational method with an emphasis on the study of statistics for elementary education, as well as to encourage students to conduct statistical research using data processing. The study also suggests that starting from basic statistical concepts applied to meaningful numerical information, and with the aid of technological resources such as spreadsheets and statistical applications, one can investigate trends, patterns, inflection points, or associations between variables, behaviors that are imperceptible when not examined in detail by science. The methodology employed is bibliographical in nature, used in the collection and research of long-term climate data collected at INMET (National Institute of Meteorology), such as maximum and minimum temperature, accumulated precipitation, relative humidity, cloud cover, total insolation, wind intensity, and atmospheric pressure, for the evaluation and application of statistical techniques in calculating averages, standard deviations, and other relevant measures in the analysis of a set of values that are intended to be interpreted. Based on these patterns, statistical methods are used to undertake exploratory data analysis, generate frequency tables, and fit mathematical models to time series. The research concludes that there is a notable efficiency of the BioEstat software in education due to its accessible practical manual that contains examples of data input for obtaining graphs, descriptive measures, and performing regression tests, among other applications. At the same time, it serves as didactic support for teaching, specifically for the 3rd year of high school, where the teacher, mediating the learning of these skills, and with the use of technology available in this context, can motivate the student's interest in processing information that requires interpretation and analysis of statistical data.

Keywords: BioEstat. statistics. teaching.

LISTA DE QUADROS

Quadro 1: recursos e funcionalidades do BioEstat.	20
Quadro 2: etapas de uma pesquisa estatística.	25
Quadro 3: convenção na distribuição de frequência.	39
Quadro 4: elementos de uma distribuição de frequência.	39
Quadro 5: média aritmética dos três períodos para as variáveis analisadas no estudo.	66
Quadro 6: instruções das aplicações utilizadas na pesquisa.	67
Quadro 7: umidade relativa do ar horária mensal média em Belém-PA, para o período 1961-2020 (%).	93

LISTA DE FIGURAS

Figura 1: ilustração população-amostra.	27
Figura 2: classificação de uma variável.	30
Figura 3: cidades com mais raios no Brasil em janeiro de 2023.....	33
Figura 4: comparação das normais climatológicas (1991 a 2020) da temperatura mensal média, em Bagé-RS, Salvador-BA, Goiânia-GO, Belém-PA e Vitória-ES.	33
Figura 5: precipitação acumulada anual de alguns municípios do estado do Pará, no período 1991-2020 (mm).	34
Figura 6: histograma e curva de distribuição normal dos dados da radiação solar global medidos pelo INMET.	35
Figura 7: temperatura versus quantidade de vapor de um processo industrial.....	35
Figura 8: estrutura básica do gráfico box-plot.	36
Figura 9: histograma da intensidade do vento mensal média em Minas Gerais, no período 1961-1990.....	43
Figura 10: gráfico de uma distribuição simétrica.....	54
Figura 11: gráfico de uma distribuição simétrica e assimétrica.	55
Figura 12: medidas de curtose.....	57
Figura 13: análise da correlação entre duas variáveis a partir do gráfico de dispersão.	59
Figura 14: ajuste de curvas a pontos no plano cartesiano. (a) Representação no plano cartesiano de 7 pontos coletados em uma pesquisa. (b) Ajuste polinomial. (c) Ajuste por mínimos quadrados.....	60
Figura 15: janela do BioEstat.....	67
Figura 16: gráfico da nebulosidade mensal média em décimos em Belém-PA, no período 1961-2020.....	70
Figura 17: histograma da nebulosidade mensal média Belém-PA, para o período 1961-2020, gerado a partir da Tabela 6.	72
Figura 18: gráfico da pressão atmosférica mensal média ao nível do barômetro, em Belém-PA, no período 1961-2020 (hPa).....	74
Figura 19: histograma da pressão atmosférica mensal média em Belém-PA, para o período 1961-2020, obtida a partir da Tabela 9.....	75
Figura 20: gráfico da temperatura mínima mensal média em Belém-PA, no período 1961-2020 (°C).	76

Figura 21: histograma da temperatura mínima mensal média em Belém-PA, para o período 1961-2020, obtida a partir da Tabela 12.....	78
Figura 22: gráfico da temperatura máxima mensal média em Belém-PA, no período 1961-2020 (°C).....	79
Figura 23: histograma da temperatura máxima mensal média em Belém-PA, obtida a partir do Tabela 15.	82
Figura 24: gráfico da precipitação acumulada mensal média em Belém-PA, no período 1961-2020 (mm).	83
Figura 25: histograma da precipitação acumulada mensal média em Belém-PA, para o período 1961-2020, obtido a partir do Tabela 18.	85
Figura 26: temperatura da superfície do mar - 01-15/03/2023.....	86
Figura 27: gráfico da insolação total mensal média em Belém-PA, no período 1961-2020 (em horas).	87
Figura 28: histograma da insolação total mensal média em Belém-PA, para o período 1961-2020, obtido a partir da Tabela 21 (em horas).....	89
Figura 29: gráfico da intensidade do vento mensal média em Belém-PA, no período 1961-2020 (m/s).....	90
Figura 30: histograma da intensidade média mensal do vento em Belém-PA para o período 1961-2020, obtido a partir do Tabela 24.	92
Figura 31: gráfico box plot da umidade relativa do ar horária mensal média em Belém-PA, no período 1961-2020.....	94
Figura 32: dados de entrada para os testes de regressão.	95
Figura 33: gráfico de regressão logarítmica para as variáveis intensidade do vento versus umidade às 12h, em Belém-PA, para o período 1961-2020.....	96
Figura 34: gráfico de regressão linear para as variáveis temperatura máxima versus umidade às 12h, em Belém-PA, para o período 1961-2020.....	97

LISTA DE TABELAS

Tabela 1: índice de variação mensal e acumulado, adaptado do Índice Nacional de Preço ao Consumidor Amplo – IPCA – fevereiro de 2023.	31
Tabela 2: estatísticas do Registro Civil – 2021.	32
Tabela 3: distribuição de frequência da intensidade mensal média do vento em Minas Gerais, no período 1961-1990, obtida a partir do Exemplo 9.	42
Tabela 4: nebulosidade mensal média em décimos em Belém-PA, no período 1961-2020. .	69
Tabela 5: estatística descritiva da nebulosidade mensal média em décimos, em Belém-PA, no período 1961-2020.	70
Tabela 6: distribuição de frequência da nebulosidade mensal média para o período 1961-2020, obtida a partir do Quadro 5.	71
Tabela 7: pressão atmosférica mensal média ao nível do barômetro, em Belém-PA, no período 1961-2020 (hPa).	73
Tabela 8: estatística descritiva da pressão atmosférica mensal média em Belém-PA, no período 1961-2020 (hPa).	74
Tabela 9: distribuição de frequência da pressão atmosférica mensal média em Belém-PA, para o período 1961-2020, obtida a partir do Quadro 5.	75
Tabela 10: temperatura mínima mensal média em Belém-PA, no período 1961-2020 (°C). .	76
Tabela 11: estatística descritiva da temperatura mínima mensal média em Belém-PA, no período 1961-2020 (°C).	77
Tabela 12: distribuição de frequência da temperatura mínima mensal média em Belém-PA, para o período 1961-2020, obtida a partir do Quadro 5.	77
Tabela 13: temperatura máxima mensal média em Belém-PA, no período 1961-2020 (°C)..	79
Tabela 14: estatística descritiva da temperatura máxima mensal média em Belém-PA, no período 1961-2020 (°C).	80
Tabela 15: distribuição de frequência da temperatura máxima mensal média em Belém-PA para o período 1961-2020, obtida partir do quadro 5 (°C).	81
Tabela 16: precipitação acumulada mensal média em Belém-PA, no período 1961-2020 (mm).	83
Tabela 17: estatística descritiva da precipitação acumulada mensal média em Belém-PA, no período 1961-2020 (mm).	84

Tabela 18: distribuição de frequência da precipitação acumulada mensal média em Belém-PA, para o período 1961-2020, obtida a parti do quadro 5.	84
Tabela 19: insolação total mensal média em Belém-PA, no período 1961-2020 (em horas). 87	
Tabela 20: estatística descritiva da insolação total mensal média em Belém-PA, no período 1961-2020 (em horas).....	88
Tabela 21: distribuição de frequência da insolação total mensal média em Belém-PA, para o período 1961-2020, obtida a partir do Quadro 5.	88
Tabela 22: intensidade do vento mensal média em Belém-PA, no período 1961-2020 (m/s).90	
Tabela 23: estatística descritiva da intensidade do vento mensal média em Belém-PA, no período 1961-2020 (m/s).	91
Tabela 24: distribuição de frequência da intensidade do vento mensal média em Belém-PA, para o período 1961-2020, obtida a partir do Quadro 5.	92
Tabela 25: teste de regressão para as variáveis intensidade do vento versus umidade às 12h.	96
Tabela 26: teste de regressão para as variáveis temperatura máxima versus umidade às 12h.	97
Tabela 27: teste de regressão para as variáveis precipitação acumulada versus temperatura máxima em Belém-PA, no período 1961-2020.	98
Tabela 28: teste de regressão para as variáveis precipitação acumulada versus umidade às 18h.	99
Tabela 29: teste de regressão para as variáveis precipitação acumulada versus intensidade do vento.	100

SUMÁRIO

1 INTRODUÇÃO	14
1.1 OBJETIVO GERAL	15
1.2 OBJETIVOS ESPECIFICOS	15
2 HISTÓRIA, DEFINIÇÃO DE ESTATÍSTICA, ÁREAS DE APLICAÇÃO, O SOFTWARE BIOSTAT E TECNOLOGIA NA EDUCAÇÃO	16
2.1 ASPECTO HISTÓRICO	16
2.2 DEFINIÇÃO DE ESTATÍSTICA E ÁREAS DE APLICAÇÃO	18
2.3 O SOFTWARE BIOSTAT	19
2.4 TECNOLOGIA NA EDUCAÇÃO.....	21
3 MEDIDAS ESTATÍSTICAS, CORRELAÇÃO E REGRESSÃO	25
3.1 ESTATÍSTICA DESCRITIVA E CONCEITOS PRELIMINARES	25
3.2 ELEMENTOS DA ESTATÍSTICA	27
3.2.1 Amostra.....	27
3.2.2 Amplitude de uma amostra de dados numéricos.....	27
3.2.3 Variável estatística: quantitativa e qualitativas	28
3.3 TABELA ESTATÍSTICA	30
3.4 SÉRIES HISTÓRICAS	31
3.5 GRÁFICOS ESTATÍSTICOS	32
3.6 DIAGRAMA DE CULE E FOLHAS.....	36
3.7 TABELA PRIMITIVA.....	37
3.8 DISTRIBUIÇÃO DE FREQUÊNCIAS	38
3.8.1 Elementos de uma Distribuição de Frequência	39
3.8.2 Cálculo do número de classes.....	41
3.8.3 Representação Gráfica de uma Distribuição	42
3.9 MEDIDAS DE TENDÊNCIA CENTRAL.....	43
3.9.1 Média Aritmética.....	44
3.9.2 Mediana.....	44
3.9.3 Moda	46
3.9.4 Média geométrica.....	46
3.9.5 Média Harmônica.....	47
3.10 MEDIDAS SEPARATRIZES	48
3.10.2 Decis	50
3.10.3 Percentis	50
3.10.4 Desvio interquartilico D_i	50
3.11 MEDIDAS DE VARIABILIDADE OU DISPERSÃO	51
3.11.1. Variância	51

SUMÁRIO

3.11.2 Desvio padrão.....	53
3.11.3 Coeficiente de Variação.....	53
3.12 MEDIDAS DE ASSIMETRIA.....	54
3.12.1 Medidas de curtose.....	56
3.13 CORRELAÇÃO E REGRESSÃO.....	57
3.13.1 Coeficiente de determinação R^2	58
3.13.2 Coeficiente de correlação de Pearson	58
3.13.3 Reta De Regressão (ou reta dos mínimos quadrados ou reta de ajuste)	59
4 MÉTODO E RESULTADOS.....	64
4.1 METODOLOGIA DA PESQUISA.....	65
4.2 RESULTADOS E DISCUSSÃO DAS APLICAÇÕES	68
4.2.1. Nebulosidade.....	68
4.2.2 Pressão Atmosférica	72
4.2.3 Temperatura mínima	76
4.2.4 Temperatura Máxima	78
4.2.5 Precipitação Acumulada.....	82
4.2.6 Insolação	86
4.2.7 Intensidade do vento.....	89
4.2.8 Umidade relativa do ar	93
4.3 AJUSTAMENTO DE CURVAS.....	94
4.3.1 Intensidade do vento versus umidade às 12h.....	95
4.3.2 Temperatura máxima versus umidade às 12h.....	96
4.3.3 Precipitação acumulada versus temperatura máxima.....	97
4.3.4 Precipitação acumulada versus umidade às 18h	98
4.3.5 Precipitação acumulada versus intensidade do vento	100
4.4 SÍNTESE DOS RESULTADOS	101
5 CONCLUSÃO E RECOMENDAÇÃO.....	103
REFERÊNCIAS.....	105

1 INTRODUÇÃO

Uma alternativa de suporte metodológico às aulas de matemática são os programas de computador capazes de gerar gráficos de funções, as calculadoras científicas, *softwares* de geometria dinâmica, pacotes estatísticos, entre outros recursos tecnológicos disponíveis nesse segmento. Henz (2008), simplesmente define tecnologia como objetos, instrumentos, aparelhos eletrônicos, ou seja, todo os artifícios que possam beneficiar nossas vidas e em certas ocasiões faz-se inevitável.

Para Perius (2012), a internet e os computadores disponibilizam possibilidades que favorecem o progresso e a compreensão de conteúdos e métodos matemáticos. Ao mesmo tempo que para Campos e Campos (2001), o avanço de *softwares* educativos procura atender os atributos da educação que conduzem ao desenvolvimento integral do estudante que precisa aprender como se aprende, e a pensar para interceder melhor, modernizar e questionar, trabalhando assim com as funções cognitivas.

Os *softwares* foram concebidos originalmente com foco no processamento eficiente de grandes volumes de informações, além de possuírem características pedagógicas. Os desenvolvedores de tecnologia de computadores tinham como principal objetivo, aprimorar técnicas que permitissem minimizar os custos e despesas relacionados ao armazenamento e organização de dados, ao mesmo tempo em que reduziam significativamente o tempo necessário para o processamento (RIBEIRO; DAMÁSIO, 2006).

Para D'Ambrósio (1990), um dos maiores desafios da educação reside em cultivar a criatividade e a ética nos alunos, preparando-os a se tornarem protagonistas conscientes de sua história pessoal e coletiva. Isso implica em instruí-los a compreender e transformar o mundo ao seu redor.

O estudo tem duplo propósito, ser instrumento didático para os professores que se encontram com a responsabilidade de ensinar estatística aos jovens estudantes da etapa final da educação básica e produzir nestes, motivação e iniciativa de aprender estatística pesquisando dados notáveis com o uso da tecnologia disponível no tempo presente.

Portanto, indaga-se: de fato a utilização do recurso tecnológico no estudo de estatística, desperta o interesse do educando em aprender tal área da matemática pesquisando dados relevantes que podem apontar correlação de variáveis nos mais variados assuntos?

Parte-se da hipótese de que ensinar estatística somente conceituando temas, apresentando exemplos e resolvendo exercícios de aplicação, não são suficientes no desenvolvimento da assimilação e para mostrar a importância de se fazer pesquisas estatísticas, além de ser ineficaz para se obter dos estudantes o interesse necessário que os faça despertar para a iniciação científica.

Inicialmente, no primeiro capítulo, é feita uma apresentação histórica da evolução da estatística até ser considerada científica, com o desenvolvimento de métodos e modelos matemáticos. Informa-se a definição de estatística, citam-se áreas de aplicação e expõe-se um resumo do aplicativo juntamente com certos recursos disponíveis no *software* BioEstat. Expõe-se menções de auxílio tecnológico no ensino escolar em documentos oficiais que conduzem a educação básica brasileira.

No capítulo segundo, introduz-se a estatística descritiva, traz-se noções preparatórias, uso de tabelas, gráficos e medidas estatísticas, além da introdução aos modelos de regressão. No terceiro capítulo, declara-se a metodologia da pesquisa juntamente com os resultados e as observações das saídas de informações geradas pelo programa usado no estudo. No quinto capítulo, encontra-se a conclusão e a recomendação do trabalho.

1.1 OBJETIVO GERAL

Compreender o uso do *software* BioEstat como método educativo com ênfase no estudo da estatística para o ensino básico.

1.2 OBJETIVOS ESPECIFICOS

- Verificar o aspecto histórico da estatística descritiva, áreas de aplicação, o *software* BioEstat e a previsão do uso da tecnologia na educação;
- Mostrar a estatística descritiva, os conceitos preliminares, o uso de tabelas, gráficos de uma distribuição e medidas;
- Analisar padrões, tendências e possíveis associações entre pares de variáveis.

2 HISTÓRIA, DEFINIÇÃO DE ESTATÍSTICA, ÁREAS DE APLICAÇÃO, O SOFTWARE BIOESTAT E TECNOLOGIA NA EDUCAÇÃO

Neste capítulo será feita uma breve evolução histórica da estatística, mostrando o desenvolvimento desde a antiguidade até a atualidade. Informar-se-á, também, sobre algumas áreas de aplicação da estatística, além de uma breve apresentação do programa de computador utilizado na pesquisa e suas aplicações. Ao fim do mesmo, argumenta-se, com as evidências documentais oficiais, a necessidade e recomendação, na educação, da utilização de recursos tecnológicos pelo professor e aluno.

2.1 ASPECTO HISTÓRICO

A motivação de se apresentar preliminarmente o contexto da memória no presente capítulo faz-se necessária pois, compreender o processo histórico da estatística descritiva representa um tópico fundamental para se abordar durante o processo de aprendizagem no âmbito educacional. Isso porque, implementar o conhecimento sobre a história da estatística, torna o aprendizado mais interessante, além de motivar e aguçar a curiosidade do estudante.

O estudo do aspecto histórico não somente engrandece a etapa educacional dos alunos, mas também os certifica como estatísticos mais bem informados e preparados. Apreciar o progresso dessa área da matemática e aprender com as lições do pretérito, prepara o aprendiz para enfrentar os desafios estatísticos do presente e do futuro, colaborando para a ciência, a tomada de decisões e a progressão do conhecimento em um mundo em constante evolução.

Desde a antiguidade, diferentes povos faziam registros de contagens de informações como nascimentos, mortes, população local e estimativas de riquezas, cobrava-se impostos, fazia-se divisão justa de terras para o povo, produziam apurações quantitativas por processos que seria equivalente ao que a estatística, também, faz hoje (CRESPO, 2002).

Na idade média, o registro das informações eram em tabelas que tinham fins bélicos e tributários (COSTA, 2011). No entanto, nessa época, a análise quantitativa dos dados ainda era rudimentar e não tinha a abrangência e profundidade das práticas estatísticas modernas atuais.

Somente a partir do século XVI, segundo Crespo (2002), iniciam-se as primeiras análises em conjuntos de dados organizados e classificados, segundo um ou mais critérios, de fatos sociais: batizados, casamentos e funerais. No século XVII, em 1662, Jon Graunt publicou

o livro "Natural and Political Observations Made upon the Bills of Mortality". Graunt coletou dados sobre nascimentos, mortes e doenças em Londres e usou essas informações para identificar tendências e padrões.

Foi no século XVIII que foi dado, pelo economista e filósofo Godofredo Achenwall (1719 - 1772), o nome de estatística para a nova ciência (BARBOSA, 2012). Para os alemães, Achenwall é considerado o pai da estatística moderna. Ele é famoso pela sua obra "Princípios da Estatística". Com ele a estatística passa a ser não apenas catalogação de dados numéricos e se transforma em um estudo de como se chegar em conclusões sobre o todo (população) a partir de uma parte deste todo (amostra).

No século XIX, Segundo Rodrigues (2015), surgiu a estatística médica, com o estudo de William Farr surgem apontamentos de técnicas de correlação e ajuste de curvas. Karl Pearson conseguiu a equação que calcula o coeficiente de correlação e posteriormente desenvolveu a análise de regressão e teste de hipóteses qui quadrado. Ainda nesse período foram aplicados métodos estatísticos às questões da biologia relacionadas com a evolução e hereditariedade da teoria de Darwin.

No Brasil, em 1808, D. João VI, solicitou que fosse enviado a Portugal a elaboração de uma estatística detalhada sobre o Brasil, para entender melhor o país e seus habitantes (RODRIGUES, 2015). Foram organizadas informações sobre cultura, produção, consumo e exportação. Isso resultou na produção de documentos e tabelas que descreviam as condições econômicas, sociais e demográficas do Brasil na época.

No século XX, entre 1900 e 1915, iniciou-se a estatística moderna, sendo um período de transição entre a estatística clássica e a nova estatística. O estatístico irlandês William Sealy Gosset reconhecido por desenvolver o teste t de Student, que é usado para avaliar a diferença entre as médias de duas amostras. O teste t de Student é largamente usado na estatística inferencial. Gosset publicou seus trabalhos com o pseudônimo "Student" (RODRIGUES, 2015).

Ronald Aylmer Fisher, tido como um dos maiores estatísticos da história, reconhecido por desenvolver métodos estatísticos importantes, como o teste ANOVA, o teste de distribuição de Fisher, o teste de variação proporcional, e a teoria dos testes de hipótese. O período de 1925 a 1960 tornou-se o período de ouro da estatística (RODRIGUES, 2015).

Atualmente, século XXI, por causa dos crescentes e recorrentes avanços tecnológicos, existe um aumento no uso de recursos computacionais. Assim como, também, cresce a formalização da estatística, através de modelos matemáticos. Além disso, a quantidade e a variedade de dados gerados aumentaram rapidamente, o que resultou em uma maior necessidade de técnicas e ferramentas estatísticas avançadas para lidar com esses dados.

2.2 DEFINIÇÃO DE ESTATÍSTICA E ÁREAS DE APLICAÇÃO

Ensinar o significado de estatística é básico para que se forneça aos alunos uma compreensão simples e completa do que é essa disciplina e como ela se aplica em diversos cenários. Além disso, capacita-os a tomar decisões informadas, potencializar habilidades de pensamento crítico e cooperar de forma mais atuante e consciente na sociedade orientada por dados em que vivemos.

A palavra Estatística vem de *Status* (Estado, em Latim). Então ela significa estudo do Estado (COSTA, 2011). A estatística fornece instrumentos para aferir incertezas, examinar a relevância de efeitos e tomar decisões explicada com base em dados. Para Crespo (2002), estatística é uma parte da matemática aplicada que oferece técnicas para coleta, análise, interpretação, apresentação e organização de dados para servir-se dos mesmos na tomada de decisões.

Enquanto que para o Dicionário Escolar da Academia Brasileira de Letras (2008), estatística é o ramo da matemática que pesquisa, analisa e organiza dados numéricos. Ou ainda, que estatística é o resultado dessa pesquisa. Ferreira (2015) ensina que quando os dados e informações a serem tratados pertencem às ciências biológicas ou à saúde, usa-se o termo bioestatística.

Portanto, apesar das divergências encontradas, existe uma significação que é essencial e que deve pertencer a todas as definições de estatística, que é a análise de dados. O que se observa é que o significado depende do contexto. Se o contexto é lexical, a tradução é limitada e essencial, caso contrário, a intenção do sentido técnico tem um propósito bem claro que é fazer observações relevantes nas distribuições dos dados, ou ainda, tentar fazer previsões aproximadas a partir de uma amostra da população que se pesquisa.

Se informar sobre as áreas de aplicação da estatística no ensino vai além de simplesmente adquirir conceitos teóricos. Isso estimula a consciência sobre a importância da

estatística na vida cotidiana, abre novas perspectivas profissionais, consolida a compreensão dos conceitos em um contexto real e motiva uma aprendizagem mais ativa e comprometida. Ao se tornarem conhecedores em estatística, os alunos estão melhor preparados para enfrentar o futuro com confiança e contribuir positivamente em suas carreiras e na sociedade como um todo.

No cotidiano é comum o emprego da estatística em debates e notícias que tratam de economia, política, esportes, educação, saúde, alimentação, moradia, meteorologia, pesquisas de opinião, pesquisas de intenção de votos, meio ambiente, sociais, engenharia, entre outros (RODRIGUES, 2015). No comércio, por exemplo, moldar um produto para o consumidor requer conhecer o que o cliente anseia. Porém os consumidores têm escolhas distintas, tornando necessário efetuar pesquisas de observação, ou de sondagens para atender às necessidades de consumo do sujeito (RIBEIRO *et al*, 2016).

Na área médica, quando um médico generalista questiona a um paciente sobre seu antecedente familiar, ele o faz por causa do saber preexistente de padrões constatado pela estatística, tais como a tendência de acometimento de certas disfunções em uma mesma família (FIGUEIRÓ, 2014).

Os governos possuem estatísticos que coordenam censos de habitantes, bens manufaturados, agricultura, entre outros. São organizados e reunidos dados sobre consumo, fabricação, inventário, folha salarial, e outras informações das fábricas e estabelecimentos. Esses levantamentos advertem o administrador como sua empresa está se desenvolvendo, de seu crescimento em comparação com a concorrência existente no mercado, e disponibiliza-lhe condições de engendrar atos futuros. O estudo desses dados é considerável para um gerenciamento apropriado (CORREA, 2003).

2.3 O SOFTWARE BIOESTAT

O BioEstat segue uma abordagem que fortalece a compreensão da Estatística como uma disciplina científica, revelando modelos matemáticos e métodos de análise, tornando assim mais acessível a interpretação dos fenômenos que ocorrem de forma natural ou são provocados na esfera dos organismos vivos, principalmente pela intervenção humana. O pacote estatístico BioEstat tem como propósito oferecer aos estudantes de várias áreas do conhecimento uma

ferramenta altamente prática e de fácil utilização para a avaliação de informações resultantes de pesquisas (Ayres *et al*, 2007).

Esta elaboração, a do *software* BioEstat juntamente com o livro BIOESTAT (manual que vem acoplado com os arquivos de instalação do programa) foi realizado por três gerações dos Ayres: a primeira está representada por Manuel Ayres, ex-professor de Estatística da Universidade Federal do Pará; a segunda, por Manuel Ayres Jr., PhD em engenharia civil pela Universidade de Maryland (EUA); e a terceira, por Daniel Lima Ayres, engenheiro de computação pela Universidade de Illinois (EUA), que desenvolveu o aplicativo, juntamente com Alex de Assis Santos dos Santos, e a diagramação do manual, e, mais recentemente, Lucas Lima Ayres, estudante de medicina veterinária da Universidade Federal de Minas Gerais, o qual contribuiu no desenvolvimento do Capítulo 18 sobre Estatísticas Circulares (Ayres *et al*, 2007).

O Quadro 1, a seguir, exhibe os métodos estatísticos com as funções e aplicações disponíveis no *software* BioEstat.

Quadro 1: recursos e funcionalidades do BioEstat.

MÉTODOS ESTATÍSTICOS	FUNÇÕES E APLICAÇÕES
Estatísticas descritivas:	o software permite calcular medidas descritivas básicas, como média, mediana, desvio padrão, percentis, entre outros, para descrever e resumir os dados biológicos coletados.
Testes estatísticos:	o BioEstat oferece uma variedade de testes estatísticos paramétricos e não paramétricos como t de Student, ANOVA, teste de Mann-Whitney, teste de Kruskal-Wallis, qui quadrado, entre outros. Esses testes são utilizados para comparar grupos, verificar diferenças estatisticamente significativas e avaliar associações entre variáveis.
Análise de sobrevivência:	para estudos de sobrevida, o BioEstat oferece ferramentas para análise de dados censurados, como o método de Kaplan-Meier e modelos de regressão de Cox.
Gráficos e visualizações:	o software permite a criação de e visualizações personalizadas para representar os dados biológicos, como histogramas, gráficos de dispersão, boxplots, gráficos de barras, entre outros.
Amostragem e planejamento experimental:	o BioEstat auxilia na definição de tamanhos de amostra adequados e no planejamento experimental, considerando fatores como poder estatístico, níveis de significância e design experimental.

Fonte: elaborada pelo autor.

De acordo com Silva *et al.* (2014), o pacote estatístico do BioEstat tem como finalidade fornecer aos alunos, uma ferramenta de ampla funcionalidade e de simples manipulação na análise de dados produzidos por meio de pesquisa.

Podem surgir pequenas discrepâncias nos resultados ao compará-los com outros programas similares. Essas variações eventualmente identificadas podem estar ligadas à utilização de algoritmos diferentes (Ayres *et al.*, 2007).

2.4 TECNOLOGIA NA EDUCAÇÃO

A seguir serão expostas regulamentações, metas, propostas e menções ao uso da tecnologia na educação. Tais argumentos e coletâneas serão o apoio que justifica o propósito da utilização de recursos tecnológicos no processo de ensino e aprendizagem. Entenderemos que esse é um tema impreterível para o avanço cognitivo, para o desenvolvimento de competências e habilidades do sujeito que está sendo forjado na educação básica brasileira.

A Lei Federal nº 9394, de 20/12/96, Lei de Diretrizes e Bases da Educação Nacional (LDB), determina as normas e bases da educação no Brasil, ela é tida como a principal lei educacional federal e é encarregada de estabelecer a organização, as finalidades e os rumos da educação brasileira em todas as etapas e modalidades de ensino, da educação infantil até o nível superior. De acordo com a LDB (1996, apud ALMEIDA *et al.*, 2021, p148.), o artigo 32º da LDB expõe que o Ensino Fundamental tem como meta o desenvolvimento essencial do cidadão a partir do entendimento de diversos aspectos, entre eles a tecnologia.

Art. 32 - O Ensino Fundamental Obrigatório, [...] terá por objetivo a formação básica do cidadão, mediante:

II - a compreensão do ambiente natural e social, do sistema político, da tecnologia, das artes e dos valores em que se fundamenta a sociedade. (LDB, 1996, p 23.)

Em meados da década de noventa inicia-se uma relevante proposição no contexto educacional, com a finalidade de definir orientações para a formulação dos currículos escolares, pretendendo assegurar a qualidade da educação ofertada no Brasil, os Parâmetros Curriculares Nacionais foram concebidos pelo Ministério da Educação (MEC) em 1997. Os Parâmetros Curriculares Nacionais (PCNs) representam um referencial de qualidade para a educação no Ensino Fundamental brasileiro (BRASIL, 1997).

Desde os PCNs já se propunha a utilização de tecnologias educacionais para obtenção do aprendizado dos conteúdos curriculares inerentes à formação escolar. Segundo tal documento, não basta somente saber quais os códigos e comandos de entrada para alcançar respostas informatizadas. É preciso conhecer o processo que justifica as respostas dadas. Entender como manusear a inovação tecnológica acarreta o conhecimento da técnica para fazer uso dela e, basicamente, de habilidades referentes ao tratamento da informação (BRASIL, 1997). Tem-se aqui a alternativa pedagógica de se ensinar a partir do resultado.

Outro documento que versa sobre o programa educacional brasileiro são as Diretrizes Curriculares Nacionais, que constituem um conjunto estruturado de princípios teóricos e práticos que servem como base sólida na concepção dos currículos educacionais, enfocando questões pertinentes que promovem a formação de cidadãos críticos, criativos e preparados para enfrentar os desafios da atualidade.

As Diretrizes Curriculares Nacionais (DCN) apontam que a tecnologia tem a possibilidade de ser empregada como recurso para evolução das capacidades intelectivas, como a resolução de problemas, a tomada de decisão, a inventividade e a cooperação. As DCNs enfatizam que a aplicação de tecnologias pedagógicas é um relevante instrumento para o avanço de habilidades e competências do aprendiz. Dessa forma, os currículos escolares precisam antecipar sua integração de maneira organizada e responsável.

Tais diretrizes ressaltam ainda o valor da formação de professores para o manuseio dos artifícios tecnológicos educacionais no ensino. Advertem que o uso da tecnologia precisa estar em consonância com as metas educacionais do currículo juntamente com o perfil dos estudantes. Além do mais, com as tecnologias educacionais é possível favorecer a progressão da inclusão e a isonomia na educação.

Enquanto a escola permanece ancorada em metodologias tradicionais, concebendo o ensino e a aprendizagem como entidades distintas, os estudantes demandam abordagens diferentes, em que aprender, ensinar, pesquisar, investigar e avaliar se interligam de forma inseparável. Os alunos, dotados de características distintas, absorvem informações com rapidez, apreciam processos paralelos e desempenham múltiplas tarefas simultaneamente. Preferem elaborar gráficos antes mesmo de ler o texto em si. Ao mesmo tempo, os docentes acreditam acompanhar a era digital ao utilizarem computadores para digitar e imprimir textos e possuírem *e-mails*, porém, não percebem que os estudantes nasceram imersos nesse contexto digital. (BRASIL, 2013).

No que diz respeito às declarações nas Diretrizes próprias do Ensino Fundamental referentes às tecnologias nesta etapa, ressaltam que a tecnologia acompanhada com a ciência precisa passar pela elaboração dos assuntos da base nacional comum e a parte diversificada do currículo. Assim sendo, essas Diretrizes indicam que:

Art. 28 – a utilização qualificada das tecnologias e conteúdos das mídias como recurso aliado ao desenvolvimento do currículo contribui para o importante papel que tem a escola como ambiente de inclusão e de utilização crítica das tecnologias da informação e comunicação, requerendo o aporte dos sistemas de ensino no que se refere à:

I – provisão de recursos midiáticos atualizados e em número suficiente para o atendimento aos alunos;

II – adequada formação do professor e demais profissionais da escola. (BRASIL, 2013, p. 113).

Tais diretrizes são constantemente requalificadas para ajustar-se às transformações sociais e do mercado de trabalho. O Plano Nacional de Educação (PNE) é uma norma que define um conjunto de orientações, objetivos e planos para a política educacional brasileira a serem alcançadas em um período de dez anos. Para o decênio 2014-2024, atual PNE, tem-se o documento que mais faz referências e observações acerca do uso das tecnologias. Destacaremos aqui as estratégias acerca desse tema:

7.12 – Incentivar o desenvolvimento, selecionar, certificar e divulgar tecnologias educacionais para a educação infantil, o ensino fundamental e o ensino médio e incentivar práticas pedagógicas inovadoras que assegurem a melhoria do fluxo escolar e a aprendizagem, assegurada a diversidade de métodos e propostas pedagógicas, com preferência para softwares livres e recursos educacionais abertos, bem como o acompanhamento dos resultados nos sistemas de ensino em que forem aplicadas; [...]

10.6. estimular a diversificação curricular da educação de jovens e adultos, articulando a formação básica e a preparação para o mundo do trabalho e estabelecendo inter-relações entre teoria e prática, nos eixos da ciência, do trabalho, da tecnologia e da cultura e cidadania, de forma a organizar o tempo e o espaço pedagógicos adequados às características desses alunos e alunas; (BRASIL, 2014. p.63-70)

Em 2015, a primeira versão da proposta da BNCC foi divulgada pelo MEC, depois de um longo e complexo processo envolvendo debates, consultas públicas, análises e revisões. Em 2017 o Conselho Nacional de Educação (CNE) homologou a BNCC da Educação Infantil e do Ensino Fundamental, e em 2018 homologou a do Ensino Médio. A Base Nacional Comum Curricular (BNCC) é uma declaração regulamentada que institui as bases, habilidades e competências fundamentais para todos os estudantes da educação básica no Brasil. Criada pelo

Ministério da Educação (MEC) e legitimada pelo Conselho Nacional de Educação (CNE) em 2017. Exigindo das escolas públicas e privadas que observem o disposto em seus objetivos.

A BNCC estipula o que necessita ser ensinado, porém não se manifesta como se alcançará esse fim. Deixando clara autonomia para as escolas e professores definirem suas metodologias, seus materiais didáticos e planos educacionais mais apropriados à realidade de seus estudantes. É aqui nessas habilidades que a sugestão ao uso de softwares é mais específica e está direcionada para a etapa do ensino médio:

(EM13MAT202) 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.

(EM13MAT406) Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.

(EM13M AT407) 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.

(EM13MAT510) Investigar conjuntos de dados relativos ao comportamento de duas variáveis numéricas, usando ou não tecnologias da informação, e, quando apropriado, levar em conta a variação e utilizar uma reta para descrever a relação observada. (BRASIL, 2018, p.202-541)

Note que diante do exposto referente às citações documentais pedagógicas, não faltam prescrições normativas e declarações legais que orientem, proponham ou determinem utilizar, como material educativo no processo de ensino-aprendizagem, o artifício da tecnologia para auxiliar professores e estudantes.

Logo, para que esse ideal se concretize, investir na infraestrutura tecnológica das instituições educacionais é uma medida fundamental para impulsionar o progresso e a qualidade da educação. Ao priorizar recursos e esforços nesse sentido, abre-se caminho para um ambiente educacional mais moderno, dinâmico e alinhado com as demandas do século XXI.

Por essa razão, é inevitável no ensino e aprendizagem do tratamento da informação combinar as instruções teóricas que se adquirem no método tradicional, com a moderna disponibilidade de *softwares* programados para este fim. Os avanços tecnológicos revolucionaram a maneira como lidamos com dados e análises estatísticas. Atualmente, é indispensável a utilização de *softwares* especializados para simplificar e acelerar a coleta, organização, análise e interpretação de informações. Essas ferramentas computacionais

desempenham um papel crucial na eficiência do processo, permitindo que os dados sejam manipulados de maneira mais ágil e precisa.

3 MEDIDAS ESTATÍSTICAS, CORRELAÇÃO E REGRESSÃO

Das informações expostas no Capítulo anterior, parte-se para o referencial teórico, que dará subsídios à análise dos dados que serão apresentados na discussão dos resultados estatísticos gráficos e descritivos das normais climatológicas que o presente trabalho se propõe. Assim, é ponderoso compreender alguns conceitos estatísticos essenciais. Nesse contexto, apresentaremos tais conteúdos de modo descomplicado e direto.

3.1 ESTATÍSTICA DESCRITIVA E CONCEITOS PRELIMINARES

A estatística descritiva é constantemente utilizada como uma fase inicial na exploração de dados, porque permite interpretar o caráter dos dados, indicando tendências e modelos significativos. Além do mais, as informações adquiridas na fase descritiva são correntemente utilizadas como suporte para o estudo inferencial que procura determinar conclusões sobre uma população, apoiando-se em uma porção de dados.

De acordo com Reis (2009), a estatística descritiva consiste na coleta, exibição, estudo e exame de dados numéricos por meio da geração de ferramentas adequadas: tabelas, gráficos e indicadores numéricos.

Para realizar o planejamento de uma pesquisa estatística, segundo Paiva (2015), considera-se que, essencialmente, na estruturação de uma pesquisa estatística deve-se seguir as seguintes etapas:

Quadro 2: etapas de uma pesquisa estatística.

ETAPA	PROCESSO
Definição do problema:	Estabelecer o propósito, buscando circundar a causa que motivou a realização da pesquisa.
Planejamento da pesquisa:	Definir a população da pesquisa, isto é, quais elementos fornecem dados sobre o tema em questão.

Universo da pesquisa:	Definir se a pesquisa será feita com todos os elementos do conjunto relacionado à pesquisa (pesquisa censitária) ou somente com uma parte dele (pesquisa por amostragem).
Escolha da amostra:	Se a pesquisa for por amostragem, definir qual a parte da população a ser pesquisada, procurando uma amostra que possa representar a totalidade.
Questionário da pesquisa:	Formular as interrogações que serão feitas para obter informações sobre o assunto.
Tipo de coleta:	Definir como os dados serão pesquisados: por observação, questionário, entrevista ou outro método de coleta.
Levantamento	Coletar os dados e organizá-los.
Elaborar relatório:	Exibir os resultados no formato de tabela ou de gráfico.
Análise dos dados:	Estudar e explicar os dados e se necessário calcular medidas estatísticas.
Inferencial:	Fazer deduções sobre o tema pesquisado fundamentado na exploração dos dados.

Fonte: Paiva (2015).

O Universo estatístico ou população estatística, de acordo com Crespo (2002), é dado como sendo o conjunto de elementos que carrega, pelo menos, uma característica comum. Frequentemente, por incapacidade ou inviabilidade financeira, restringimos os estudos referentes à uma certa pesquisa a somente uma fração da população. A essa fração oriunda da população sob investigação que se denomina amostra.

Exemplo 1:

- a) Os estudantes constituem uma população, pois apresentam pelo menos uma característica comum: são os que estudam.
- b) Produtos entregues na região norte, constituem uma população, pois apresentam pelo menos uma característica comum: vão para a região norte.
- c) Homens que fazem musculação, constituem uma população, pois apresentam pelo menos uma característica comum: são os que se exercitam em academia.
- d) O departamento de controle de qualidade de uma fábrica de lâmpadas faz, diariamente, um teste por amostragem das unidades produzidas no dia. Nesse procedimento, o universo estatístico é o conjunto de todas as lâmpadas produzidas em um dia.

Em um estudo estatístico, temos que pesquisar uma ou mais características dos elementos de uma população. Para Crespo (2002), esta característica precisa estar distintamente definida. E isto ocorre quando, considerando um elemento qualquer, somos capazes de afirmar, sem incerteza, se esse elemento pertence ou não à população.

3.2 ELEMENTOS DA ESTATÍSTICA

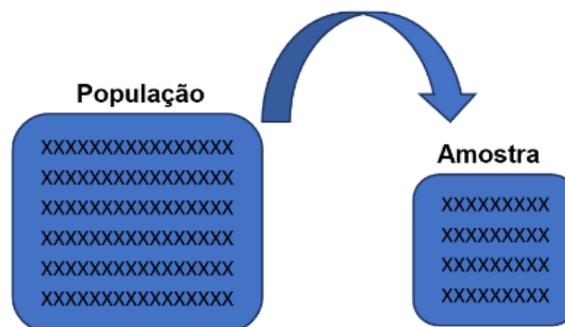
A estatística é uma ciência que se dedica ao estudo da coleta, organização, análise, interpretação e apresentação de dados. Para a execução dessas tarefas, a estatística se vale de diversos elementos essenciais ao seu funcionamento. A seguir, tem-se os elementos principais da estatística.

3.2.1 Amostra

Paiva (2015) explica que quando o universo estatístico é muito vasto ou quando for impossível colher dados de todos os elementos desse universo, seleciona-se um subconjunto dele, chamado amostra. Enquanto que Crespo (2002) é mais sucinto e define amostra como sendo um subconjunto finito de uma população.

Exemplo 2: Ao preparar uma sopa (universo estatístico), o cozinheiro prova uma colherada (amostra) para examinar a quantidade de sal. Não é necessário tomar toda a sopa para avaliar o tempero.

Figura 1: ilustração população-amostra.



Fonte: acervo do autor.

3.2.2 Amplitude de uma amostra de dados numéricos

Rodrigues (2017) define a amplitude como a diferença entre o maior e o menor valor observado. Ao mesmo tempo que Paiva (2015), prefere a formal definição de amplitude da

seguinte forma: sejam a e b , respectivamente, o menor e o maior elemento de um conjunto amostral de dados numéricos, denomina-se **amplitude da amostra** o número $b - a$.

Exemplo 3:

Em uma pesquisa foram coletados em vários supermercados os preços, em reais, da lata de determinada marca de refrigerante, obtendo-se a amostra:

3,28 3,35 3,26 3,30 3,20 3,38 3,28 3,29 3,30 3,25 3,26 3,32

Observa-se que o maior e o menor preço dessa amostra são, respectivamente, R\$ 3,38 e R\$ 3,20. Então, a **amplitude** dessa amostra é a diferença entre esses valores: R\$ 3,38 – R\$ 3,20 = R\$ 0,18. Isso significa que, dentre esses supermercados pesquisados, a maior variação de preço para esse refrigerante é 18 centavos.

3.2.3 Variável estatística: quantitativa e qualitativas

Segundo Costa (2011), as variáveis estatísticas representam as particularidades ou características que se planeja estudar em uma população ou amostra e são separadas em dois grupos: qualitativas e quantitativas. De acordo com Paiva (2015, p.36), “são atributos, numéricos ou não, pesquisados em cada elemento de uma amostra”.

A variável qualitativa é uma variável estatística que descreve uma característica ou qualidade de um objeto ou fenômeno, mas que não pode ser medida numericamente. Ou seja, é uma variável que se atribui uma qualidade não numérica como valor. Existem dois tipos de variáveis qualitativas: variável qualitativa nominal e variável qualitativa ordinal.

Entende-se como variáveis nominais quando não existe ordenação nas características e são classificadas em grupos diferentes. Os valores nominais são utilizados somente para separar um grupo de outro.

Exemplo 4:

Gênero (masculino, feminino); cor dos olhos (azul, verde, castanho, etc.); estado civil (solteiro, casado, divorciado, etc.); marcas.

As variáveis ordinais são assim classificadas quando possuem um conjunto de valores que podem ser colocados em uma ordem natural ou lógica. Esses valores são diferentes em quantidades e qualidades e possuem uma ordem predefinida.

Exemplo 5:

- a) Classificação de níveis de educação (ensino fundamental, ensino médio, ensino superior, etc.)
- b) Classificação de grau de satisfação do eleitorado brasileiro com relação à administração presidencial do país (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmente satisfeito).
- c) Escala de dor (sem dor, dor leve, dor moderada, dor forte, etc.)

Às variáveis quantitativas discretas corresponde a uma variável numérica que pode assumir somente valores inteiros dentro de um determinado intervalo. Ou seja, são dados obtidos por processo de contagem.

Exemplo 6:

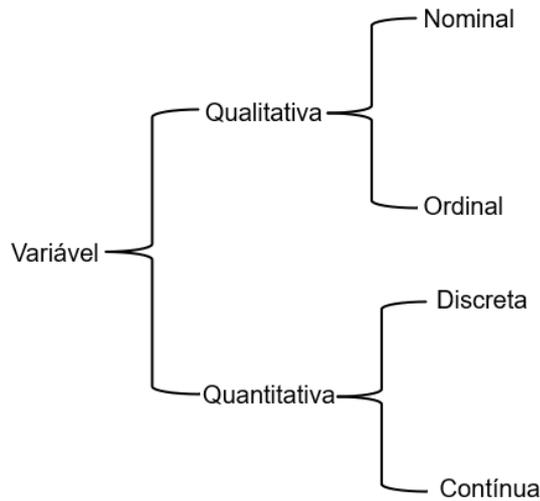
- a) O número de gols marcados em uma partida de futebol.
- b) O número de estudantes em uma classe.
- c) O número de cestas de 3 pontos em uma partida de basquete.

Às variáveis quantitativas contínuas são obtidos por processos de medição. Podem assumir qualquer valor do conjunto dos números reais.

Exemplo 7:

- a) Altura de uma pessoa.
- b) Peso.
- c) Temperatura.

A figura 2 esquematiza as classificações feitas sobre os tipos de variáveis.

Figura 2: classificação de uma variável.

Fonte: acervo do autor.

Em alguns casos, é possível atribuir valores numéricos aos atributos ou qualidades de uma variável qualitativa, e em seguida realizar a análise como se esta fosse quantitativa, contanto que essa estratégia seja suscetível à compreensão. A variável dicotômica é um caso onde essa quantificação da variável qualitativa se torna muito utilizável. Para a variável dicotômica, só ocorrem duas observações, frequentemente chamadas **sucesso** e **fracasso**. A variável “estado civil”, por exemplo, estaria nessa situação (MORETTIN; BUSSAB 2017).

3.3 TABELA ESTATÍSTICA

Uma tabela estatística é um método eficiente de expor dados estatísticos de maneira clara e sistematizada. Essa ferramenta é constituída por linhas e colunas, onde as linhas indicam as categorias ou variáveis de interesse que estão sendo observadas, enquanto as colunas significam as medidas estatísticas associadas a cada uma dessas variáveis. Essa organização facilita o entendimento das informações e possibilita que sejam naturalmente comparadas e examinadas.

Segundo Reis (2009), o êxito no uso de informações estatísticas necessita, em grande parte, da maneira como estas estão sendo transmitidas e podem ser usadas. A forma padronizada de apresentação e pormenorização das informações estatísticas tornam-se muito significativas principalmente porque as informações para quem se destina devem ser de entendimento simples e rápido. Neste sentido, estas metodologias foram elaboradas de maneira a simplificar operações de cálculos posteriores.

A Tabela 1, a seguir, mostra o índice de variação percentual de grupos de produtos e serviços em Belém – PA para o mês de fevereiro de 2023¹. Logo se pode notar, na coluna da variação mensal, que a maior variação positiva se deu com educação, enquanto que para o gênero vestuário houve uma suave deflação dos preços, na variação acumulada anual.

Tabela 1: índice de variação mensal e acumulado, adaptado do Índice Nacional de Preço ao Consumidor Amplo – IPCA – fevereiro de 2023.

Grupos de produtos e serviços	Variação mensal (%)	Variação acumulada no ano (%)
Alimentação e bebidas	0,51	1.63
Habitação	1,41	1.17
Artigos de residência	0,17	0.36
Vestuário	-0,38	-0.06
Transportes	0,03	0.36
Saúde e cuidados pessoais	1,92	0.77
Despesas pessoais	0,35	1.22
Educação	6,31	6.81
Comunicação	0,59	3.06

Fonte: IBGE.

3.4 SÉRIES HISTÓRICAS

As séries históricas dizem respeito a um conjunto de dados reunidos ao longo do tempo em períodos regulares. Esses dados normalmente se referem a eventos históricos, demografia, saúde, clima, economia ou outros assuntos que permitem serem observados ao longo do tempo. O estudo de séries históricas pode ser utilizado para reconhecer tendências, alterações sazonais, padrões e variações ao longo do tempo. Ademais, é um instrumento importante para antecipar tendências futuras.

As séries históricas são geralmente exibidas em gráficos, pois assim possibilitam uma visualização mais perceptível das modificações ao longo do tempo. Crespo (2002), define série estatística como sendo qualquer tabela que mostra o ordenamento de um conjunto de dados estatísticos em função do tempo, espaço ou do gênero.

¹ <https://sidra.ibge.gov.br/pesquisa/snipc/ipca/quadros/belem/fevereiro-2023> Acesso em 20/03/2023.

A série estatística que se apresenta a seguir (Tabela 2), explicita que nos anos em que se vivia o intensificado período pandêmico da COVID-19, houve expressiva diminuição de registros civis no Brasil; ainda nesse período, em 2020, o número de registros ignorados quase dobrou em comparação aos três anos anteriores observados. Outra observação que pode ser claramente feita é que historicamente, no Brasil, nascem ligeiramente mais homens que mulheres.

Tabela 2: estatísticas do Registro Civil – 2021.

Lugar do registro – Brasil					
Registros de nascidos vivos	2017	2018	2019	2020	2021
Homens	1 518 444	1 528 258	1 477 658	1 396 706	1 385 486
Mulheres	1 444 005	1 454 953	1 410 211	1 330 952	1 323 004
Ignorado	366	356	349	615	394
Total	2 962 815	2 983 567	2 888 218	2 728 273	2 708 884

Fonte: IBGE² (adaptado).

3.5 GRÁFICOS ESTATÍSTICOS

Usualmente, um exame detalhado de dados resume-se a aferir algumas medidas de posição e variabilidade, como por exemplo a média e a variância. Contrário a essa linha, uma corrente mais inovadora guiada por Tukey (1977), utiliza essencialmente métodos gráficos contrapondo sumários numéricos. Isto não quer dizer que sínteses não sejam necessárias, mas um estudo exploratório dos dados não tem que se restringir a aferir tais medidas (MORETTIN; BUSSAB, 2017).

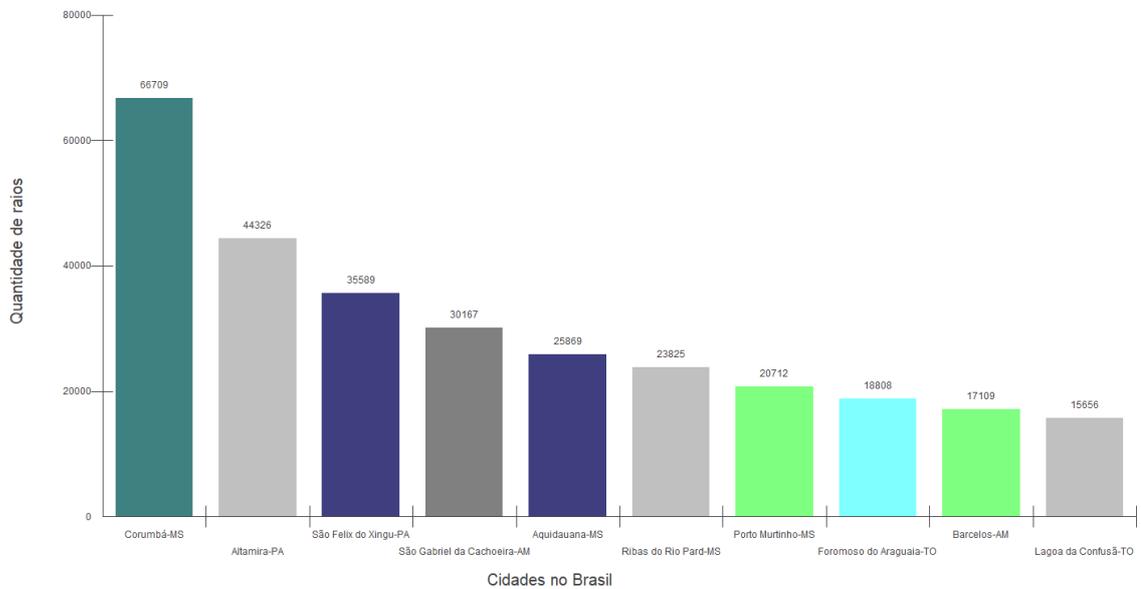
Costa (1992), de maneira sucinta, define que, representar graficamente significa realizar uma imagem que resuma de modo claro o comportamento de uma ou mais variáveis.

Alguns dos tipos mais comuns de gráficos estatísticos são:

- a) Gráfico de barras – utilizado para descrever dados discretos e exibir a comparação ou distribuição de valores. Neste tipo de gráfico, a altura ou comprimento das barras retangulares é proporcional à altura que ela representa.

² <https://sidra.ibge.gov.br/pesquisa/registro-civil/quadros/brasil/2021> Acesso em: 20/03/2023.

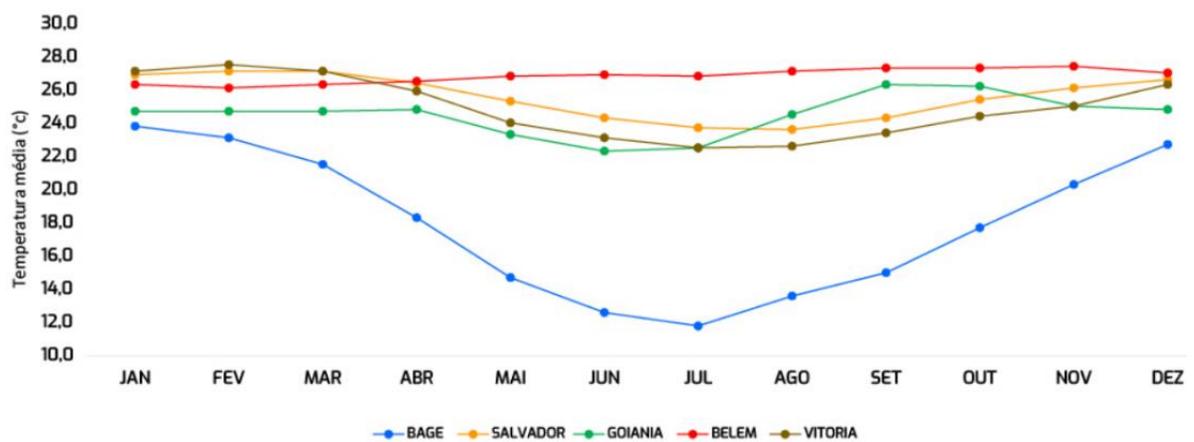
Figura 3: cidades com mais raios no Brasil em janeiro de 2023.



Fonte: Earth Networks³.

- b) Gráfico de linhas – usado para representar a relação entre duas variáveis contínuas. O gráfico de linhas mostra a evolução de uma variável ao longo de um determinado intervalo.

Figura 4: comparação das normais climatológicas (1991 a 2020) da temperatura mensal média, em Bagé-RS, Salvador-BA, Goiânia-GO, Belém-PA e Vitória-ES.

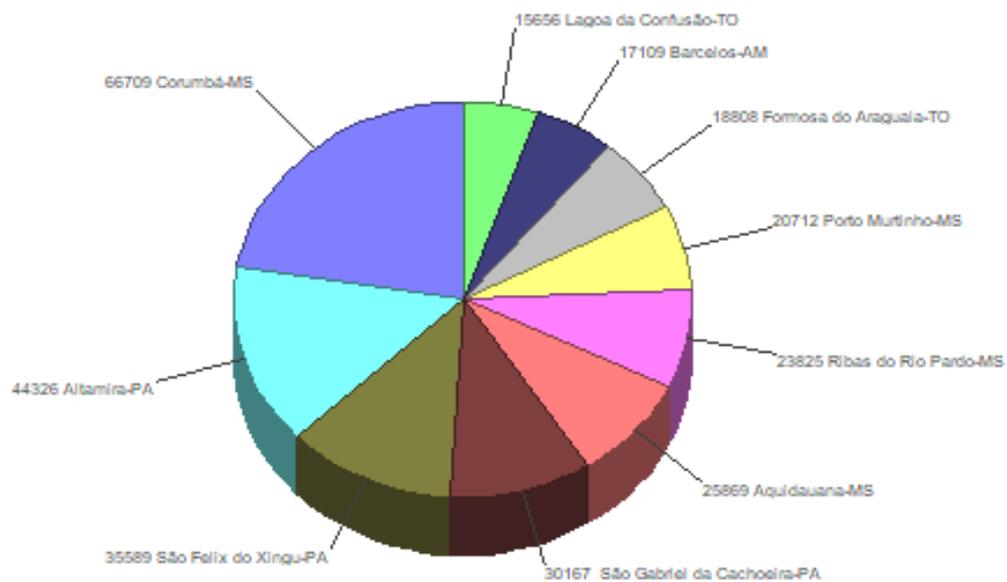


Fonte: Normais Climatológicas do Brasil (INMET, 2022).

³<https://tempoagora.uol.com.br/noticia/2023/02/04/janeiro-com-aumento-de-48-3-de-raios-no-estado-de-sao-paulo-9383>. Cidades com mais raios no Brasil em janeiro de 2023. Acesso em: 28 mar. 2023.

- c) Gráfico de setores – O gráfico de setores é uma representação visual adequada para dados nominais ou ordinais, especialmente quando o principal objetivo é comparar várias categorias em relação ao todo. Esse tipo de gráfico é construído dividindo-se um círculo em setores, onde cada setor representa proporcionalmente uma das categorias presentes na amostra (AYRES *et al*, 2007).

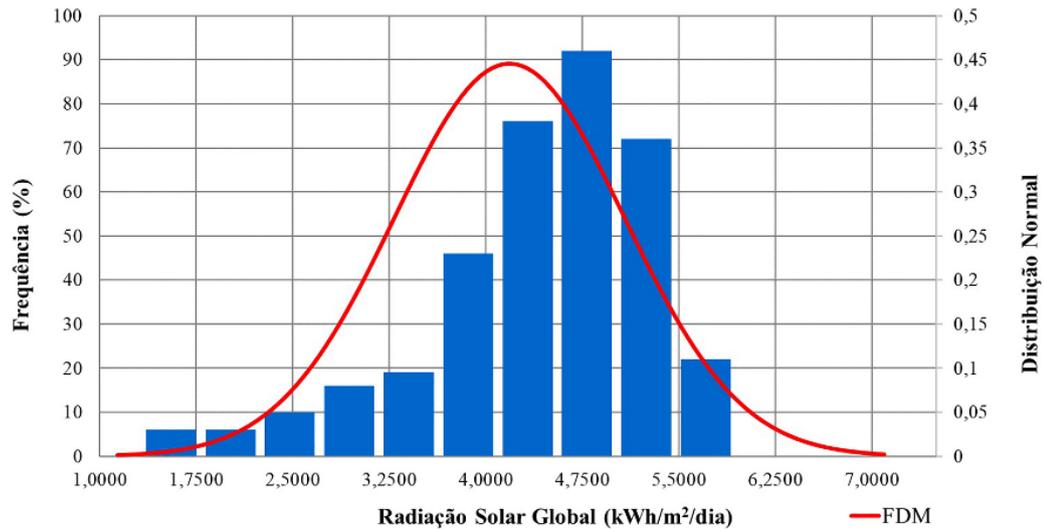
Figura 5: precipitação acumulada anual de alguns municípios do estado do Pará, no período 1991-2020 (mm).



Fonte: INMET.

- d) Histograma – Segundo Crespo (2002), o histograma é constituído por um conjunto de retângulos justapostos cujas bases estão sobre o eixo horizontal, de modo que seus pontos médios são coincidentes com os pontos médios dos intervalos de classe.

Figura 6: histograma e curva de distribuição normal dos dados da radiação solar global medidos pelo INMET.

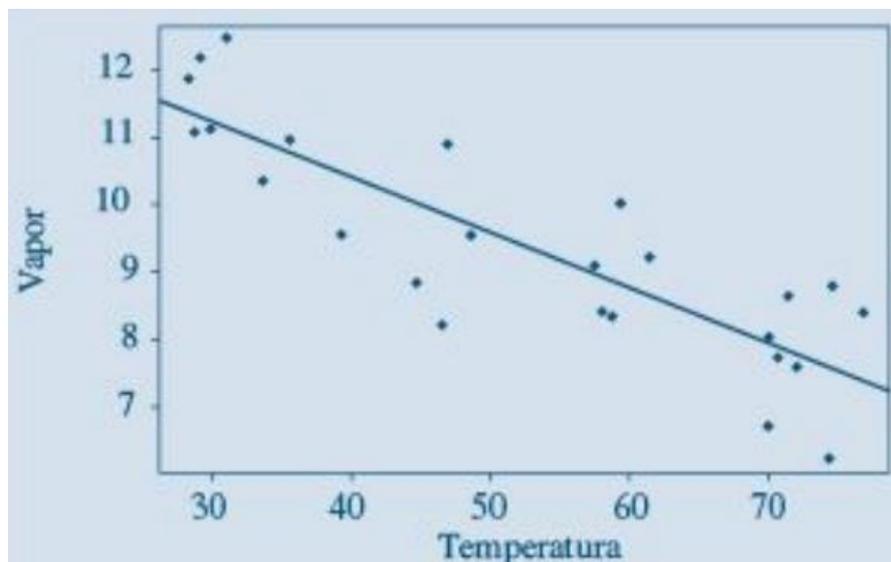


Fonte: Revista Brasileira de Meteorologia (2021).

- e) Gráfico de dispersão – Os diagramas de dispersão são representações de duas variáveis de natureza quantitativa que são coordenadas em um gráfico. Os gráficos de dispersão são adequados para verificar a associação ou não entre duas variáveis contínuas e indicar possíveis correlações entre elas. (ASSIS; SOUSA; DIAS, 2019).

No gráfico da Figura 8, tem-se a dispersão dos dados e uma reta ajustada obtidos em um processo industrial, onde X = temperatura média ($^{\circ}$ F) e Y = quantidade de vapor.

Figura 7: temperatura versus quantidade de vapor de um processo industrial.

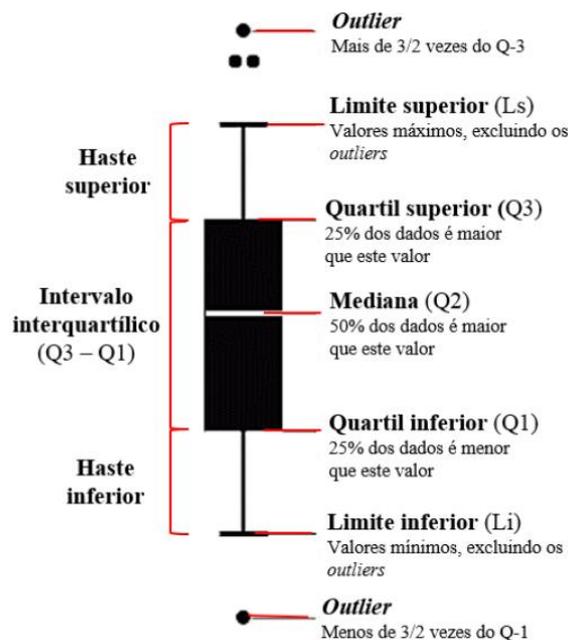


Fonte: Morettin e Bussab (2017).

- f) *Box plot* – é um método gráfico utilizado frequentemente em estatística para resumir e pesquisar dados quantitativos. O *box-plot* convencional mostra a mediana, os quartis, os valores pontuais mínimo e máximo, valores atípicos (*outliers*) e extremos.

A Figura 8 a seguir, traz os elementos, estrutura e significado do gráfico de caixa, *box plot*.

Figura 8: estrutura básica do gráfico *box-plot*.



Fonte: <https://www.robrac.org.br/seer/index.php/ROBRA/C/article/view/1132>. Acesso em: 11 jun. 2023.

3.6 DIAGRAMA DE CULE E FOLHAS

Segundo Reis (2009), este tipo de gráfico incorpora certas características do histograma enquanto preserva as informações presentes em cada dado individual. O Diagrama de caule e folhas contém um traço na vertical que separa o lado esquerdo, onde colocamos os algarismos das dezenas, do lado direito, onde colocamos os algarismos das unidades.

Exemplo 8: os dados a seguir, referem-se ao tempo gasto (em minutos) por 42 trabalhadores entre sua residência e o local de trabalho (REIS, 2009).

5	21	26	13	24	29	37
12	31	5	50	18	33	14
23	22	17	32	7	17	42
15	38	20	11	26	25	29
27	8	24	12	39	25	28
14	47	19	22	28	9	18

Organizando os dados do Exemplo 8 em um diagrama de caule e folhas, obtém-se a seguinte distribuição:

0		55789
1		122344577889
2		0122344556678899
3		123789
4		27
5		0

Note que se rodássemos o diagrama para a esquerda ele assemelhar-se-ia a um histograma, mas sem que perdesse a informação original. Por exemplo, ainda é possível saber que a maioria dos trabalhadores pesquisados gasta de 20 a 29 minutos no deslocamento entre sua casa e o trabalho.

3.7 TABELA PRIMITIVA

Segundo Crespo (2001), a tabela cujos dados de uma variável quantitativa não estão organizados numericamente (de maneira crescente ou decrescente) é chamada de tabela primitiva. Para ele, a não ordenação dos dados implica na dificuldade de analisar em torno de qual valor as medidas tendem a convergir, qual o maior e qual o menor valor observado, ou ainda, quantos estão abaixo ou acima de um valor tomado como referência. Define ainda que a tabela com uma certa ordenação recebe o nome de rol.

Exemplo 9: os dados do quadro a seguir, correspondem à intensidade média do vento em metros por segundo, de 1961-1990, para o mês de março no estado de Minas Gerais e que foram obtidos nas estações meteorológicas mantidas pelo INMET e que estão distribuídas por alguns municípios desse estado.

1.55	1.45	0.97	1.13	1.89	1.55	1.46
1.09	1.46	1.61	0.91	1.23	1.73	0.84
1.95	0.53	1.59	1.39	1.95	1.30	1.73
1.31	1.96	1.44	2.31	1.93	0.70	1.51
1.40	0.99	0.63	2.14	1.54	1.46	
1.40	1.57	1.90	1.18	1.85	1.18	
1.92	2.87	1.31	1.87	2.08	1.10	
0.74	1.60	0.68	1.09	1.89	1.67	

Organizando os dados do Exemplo 9 em ordem crescente, obtém-se o quadro a seguir:

0.53	0.99	1.30	1.46	1.59	1.89	2.08
0.63	1.09	1.31	1.46	1.60	1.89	2.14
0.68	1.09	1.31	1.46	1.61	1.90	2.31
0.70	1.10	1.39	1.51	1.67	1.92	2.87
0.74	1.13	1.40	1.54	1.73	1.93	
0.84	1.18	1.40	1.55	1.73	1.95	
0.91	1.18	1.44	1.55	1.85	1.95	
0.97	1.23	1.45	1.57	1.87	1.96	

Agora, é possível conhecer, sem muita dificuldade, qual a menor intensidade (0.53 m/s) e qual a maior (2.87 m/s); que a amplitude de variação foi de $2.87 - 0.53 = 2.34$ m/s; e, ainda, a ordem que um valor específico ocupa na distribuição dos dados. E fazendo uma análise mais apurada, vê-se que há uma concentração das intensidades dos ventos em algum valor entre 1.18 m/s e 2.14 m/s e, mais ainda, que há poucos valores acima de 1.87 m/s.

3.8 DISTRIBUIÇÃO DE FREQUÊNCIAS

De acordo com Spiegel (1978), se uma amostra ou população é ampla, fica complicado analisar as particularidades ou determinar medidas estatísticas tais como média, desvio padrão, etc. À vista disto, é adequado organizar os dados brutos (não ordenados). Para ilustrar, vamos considerar ainda a amostra do Quadro 4. Podemos usar esses dados arranjando-os em classes e afirmar o número de elementos que pertencem a cada classe; esse número é denominado

frequência da classe. A ordenação resultante é chamada distribuição de frequência ou tabela de frequência.

3.8.1 Elementos de uma Distribuição de Frequência

A organização de dados em intervalos ou classes, acompanhada da contagem da frequência de cada valor dentro desses intervalos, é um método utilizado na análise de dados. Essa abordagem permite uma forma mais simples e básica de resumir as informações, possibilitando uma visão mais aprofundada dos padrões e tendências presentes nos dados. No Quadro 4 estão dispostos os elementos essenciais de uma distribuição de frequência. Porém, faz-se necessário, antecipadamente, conhecer a simbologia para os intervalos (Quadro 3), que se convencionou pelo *software* utilizado.

Quadro 3: convenção na distribuição de frequência.

SÍMBOLO	SIGNIFICADO
—	Intervalo fechado à esquerda e aberto à direita (apenas o limite inferior pertence ao intervalo);
—	Intervalo fechado à direita e aberto à esquerda (apenas o limite superior pertence ao intervalo);
—	Intervalo fechado de ambos os lados: os dois limites pertencem ao intervalo;
—	Intervalo aberto em ambos os lados: os dois limites não pertencem ao intervalo.

Fonte: autor (2023).

Exemplo 10:

No intervalo de classe $5 \rightarrow 10$ tem-se dois limites, inferior e superior. Nesse intervalo o limite inferior é 5 e o limite superior é 10.

Quadro 4: elementos de uma distribuição de frequência.

ELEMENTO	CONCEITO
Limite Inferior Da Distribuição De Frequência (L_I):	é o valor a partir do qual são contadas as observações na distribuição de frequência.

Limite Superior Da Distribuição De Frequência (L_S):	é o valor até o qual são contadas as observações na distribuição de frequência.
Amplitude Total Da Distribuição De Frequência (A_T):	é a diferença entre o limite superior e o limite inferior da distribuição de frequência. $A_T = L_S - L_I$
Classes De Uma Distribuição De Frequência	são os subintervalos nos quais os dados são agrupados.
Limite Inferior De Classe (l_i):	é o valor a partir do qual são contadas as observações dentro da classe.
Limite Superior de Classe (l_s):	é o valo até o qual são contadas as observações dentro da classe.
Amplitude De Classe (a_t):	é a diferença entre o limite superior e o limite inferior dentro da classe, ou seja, é a diferença entre o maior e o menor valor observado dentro da classe. $a_t = l_s - l_i$
Amplitude Do Intervalo De Classe (h):	é o comprimento da classe. Para facilitação dos cálculos é conveniente arredondar o número correspondente à amplitude do intervalo de classe. $h = \frac{A_T}{K}$
Ponto Médio De Classe (x_i):	é o ponto que divide a classe em duas partes iguais e que é calculado através da média aritmética entre o limite inferior e o limite superior da classe. $x_i = \frac{l_s + l_i}{2}$
Frequência Simples Ou Frequência Absoluta Da Classe (f_i):	é o número de vezes que um determinado valor ocorre dentro da classe. A soma das frequências simples é igual ao número total dos dados.
Frequência Absoluta Acumulada De Classe (F_i):	é o acúmulo contínuo, desde a primeira classe até uma classe qualquer, das frequências absolutas das classes.

Frequência Relativa De Classe (f_r):	são os valores das razões entre as frequências simples e os números totais de dados. $f_r = \frac{f_i}{\sum f_i}$
Frequência Relativa Acumulada (F_r):	é o acúmulo contínuo, desde a primeira classe até uma classe qualquer, das frequências relativas das classes.

Fonte: adaptado – Costa (2011).

3.8.2 Cálculo do número de classes

Segundo Costa (2011), existem vários métodos para se calcular o número de classes em uma distribuição, depende da subjetividade do pesquisador, isto é, o número de classes K pode ser ajustado para se adequar ao conjunto de dados específico e à finalidade da análise. Porém existem algumas técnicas, como os exemplos citados a seguir.

- **Método Prático:** se $n < 25$, utilize $k = 5$. Se $n \geq 25$, utilize

$$k = \sqrt{n} \quad (1)$$

Onde,

K é o número de classes (ou intervalos);

n é o número de observações no conjunto de dados.

- **Fórmula de Sturges:**

$$K = 1 + 3,322 \log n \quad (2)$$

Essa fórmula é fundamentada no pressuposto de que o número ideal de classes é proporcional ao logaritmo do tamanho da amostra.

A distribuição de frequência a seguir, foi gerada pelo *software* estatístico BioEstat, a partir da inserção dos dados não ordenados do Exemplo 9. São produzidas quatro colunas: a primeira é a do intervalo de classes; a segunda corresponde ao ponto médio de cada classe; e a terceira e quarta colunas são, respectivamente, as frequências absolutas e frequências relativas.

Tabela 3: distribuição de frequência da intensidade mensal média do vento em Minas Gerais, no período 1961-1990, obtida a partir do Exemplo 9.

Classes	X_i	F_i	Percentual
0.53 — 0.87	0.70	6	11.54 %
0.87 — 1.20	1.03	9	17.31 %
1.20 — 1.54	1.37	13	25.00 %
1.54 — 1.87	1.71	12	23.08 %
1.87 — 2.21	2.04	10	19.23 %
2.21 — 2.54	2.38	1	1.92 %
2.54 — 2.88	2.71	1	1.92 %
TOTAL		52	100.00 %

Fonte: autor (2023).

A primeira classe consiste da intensidade média do vento de 0.53 a 0.87 metros por segundo, indicada por $0.53 \text{ † } 0.87$ (intervalo de classe). Como 6 valores acusam intensidades do vento pertencentes a esta classe, a frequência de classe correspondente é 6. No intervalo de classe $0.53 \text{ † } 0.87$, os números 0.53 e 0.87 são denominados limites de classe. A amplitude do intervalo, que geralmente é a mesma para todos os intervalos de classe, é a diferença entre os limites superior e limite inferior da classe. Nesse caso, $0.87 - 0.53 = 0.34$.

Pode-se tomar para representar uma classe o respectivo ponto médio x_i . No Quadro 6 acima, o ponto médio da classe $0.53 \text{ † } 0.87$ é 0.70.

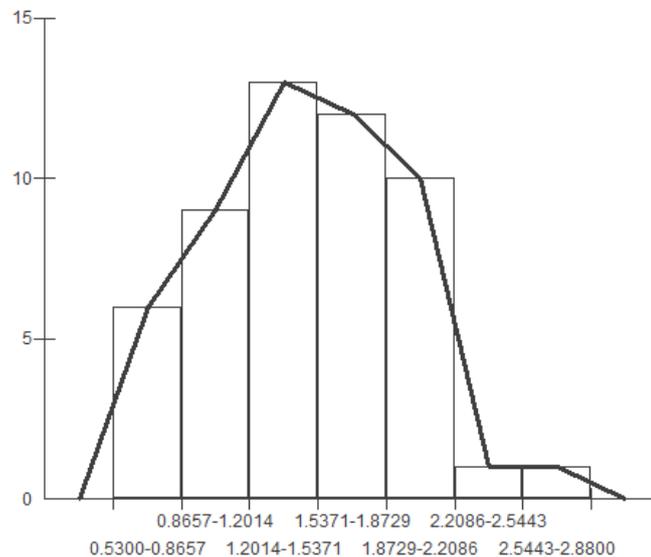
3.8.3 Representação Gráfica de uma Distribuição

Segundo Crespo (2002), uma distribuição de frequência pode ser representada graficamente pelo **histograma**, pelo **polígono de frequência** e pelo **polígono de frequência acumulada**. Qualquer um dos referidos gráficos é construído no 1º quadrante do sistema de eixos coordenados cartesianos ortogonais. No eixo das abscissas coloca-se os valores da variável; enquanto que no eixo das ordenadas coloca-se as frequências.

Apresentação visual para informações contínuas utilizando retângulos sobrepostos. Cada retângulo tem sua base representando o intervalo de uma classe específica e sua altura corresponde à frequência relativa (ASSIS, 2019). As larguras dos retângulos correspondem às amplitudes dos intervalos de classe; ao mesmo tempo que as alturas dos retângulos são proporcionais às frequências das classes e por tanto são numericamente iguais às frequências.

À distribuição do Exemplo 9 corresponde ao histograma contornado pelo polígono de frequência, representado, a seguir, na Figura 9.

Figura 9: histograma da intensidade do vento mensal média em Minas Gerais, no período 1961-1990.



Fonte: autor (2023).

Note que o polígono de frequência é um gráfico de linha que parte do ponto médio da classe que antecede à primeira e passa pelos pontos cujas coordenadas são o ponto médio de cada classe e sua respectiva frequência (base superior média dos retângulos justapostos). O polígono se fecha quando novamente intercepta o eixo horizontal no ponto médio da classe posterior à última classe da distribuição.

Comentários:

- O histograma tem a propriedade de ter sua área proporcional à soma das frequências.
- O gráfico terá área unitária quando forem consideradas as frequências relativas.
- Para comparar duas distribuições, o mais adequado é utilizar as frequências relativas.

3.9 MEDIDAS DE TENDÊNCIA CENTRAL

Azevedo (2016) define as medidas de tendência central como sendo os valores que traduzem a distribuição como um todo. De acordo com Reis (2009), eventualmente é útil

pormenorizar um conjunto de dados estatísticos por meio de um único valor; são as medidas da estatística descritiva que nos informam o valor central ou médio dos dados observados.

3.9.1 Média Aritmética

É a medida de tendência central mais usada e de mais simples análise. Seja referente a uma população (μ) ou de uma estatística amostral (\bar{X}), seu conceito e determinação são equivalentes.

A média aritmética é a soma de todos os valores observados dividida pelo número de observações.

$$\mu = \frac{X_1 + X_2 + \dots + X_n}{N} = \frac{\sum_{i=1}^N X_i}{N} \quad (3)$$

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n} \quad (4)$$

Onde n e N são respectivamente o número de elementos do conjunto amostral e o número de elementos do conjunto populacional.

Exemplo 11: Determinar a média aritmética simples dos valores 3, 7, 8, 10, 11.

$$\bar{X} = \frac{3 + 7 + 8 + 10 + 11}{5}$$

$$\bar{X} = 7,8$$

3.9.2 Mediana

A mediana (Md) é o valor central dos dados ordenados de modo crescente ou decrescente. A mediana é uma medida de posição definida como o valor da variável que ocupa a posição central da série de observações ou na distribuição de frequência. Ou seja, o número de observações inferiores à mediana é igual ao número de observações que são superiores à mediana.

Exemplo 12: No conjunto de dados {2; 4; 6; 8; 10; 12; 15; 17; 19; 21; 23}, tem-se que $Md = 12$, pois abaixo de 12 estão 5 valores (2; 4; 6; 8; 10) e acima de 12 estão também 5 valores (15; 17; 19; 21; 23).

Para dados discretos, se o conjunto de dados tiver um número ímpar de valores, a mediana é o valor central; se o conjunto de dados tiver um número par de valores, a mediana é o resultado da média aritmética dos dois valores centrais.

Exemplo 13: Para a série {3; 5; 7; 9; 11; 13}, tem-se que $Md = \frac{(7+9)}{2} = 8$.

Para o caso de uma variável contínua agrupada em classes, o cálculo da mediana pode ser assim resumido:

1º Como a variável é contínua, não é necessário diferenciar se o número de observações é par ou ímpar.

2º Analisando as frequências acumuladas, identifica-se a classe que contém a mediana e que será a classe mediana.

3º Calcula-se o valor da mediana por meio da Fórmula (5):

$$M_e = l_i + \frac{\left(\frac{n}{2} - F\right)}{f} \cdot L \quad (5)$$

Onde,

M_e é a mediana;

n é o tamanho da amostra;

l_i é o limite inferior da classe mediana;

F é a frequência acumulada da última classe anterior à classe mediana;

L é a amplitude da classe mediana (diferença entre o limite superior e inferior).

O uso dessa fórmula para o cálculo do valor mediano dentro da classe mediana admite que a distribuição seja simétrica, pois caso contrário a fórmula se torna inexata.

Segundo Ferreira (2015), a distribuição dos dados é assimétrica quando se alonga mais para um lado do que para o outro; e é simétrica se a metade esquerda do histograma se apresenta aproximadamente análoga à sua metade direita.

3.9.3 Moda

É o valor que mais se repete no conjunto dos dados, ou ainda, o valor mais frequente da distribuição. Por exemplo, a série de dados: 8, 9, 4, 11, 11, 11, 12, 13, 14, 16, tem moda igual a 11.

Porém, existem séries amodais, isto é, que não possuem um valor modal. É o caso da série: 4, 6, 9, 11, 13, 14, que não possui moda (amodal).

Pode ainda ocorrer casos em que a série apresente dois ou mais valores de acúmulo. Logo, diz-se que a série: 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 9, 10 tem como moda os valores 5 e 8 (bimodal).

Assim como a mediana, a moda também é uma medida de posição e sua determinação fica facilitada quando os dados estiverem ordenados. Reis (2009) ensina ainda que no caso de variáveis discretas, não há dificuldade para o cálculo da moda; enquanto que ao tratar de variáveis contínuas, exige-se fixar a classe modal, em seguida aplicar a fórmula:

$$M_o = l_{i(M_o)} + \frac{F_{(M_o+1)}}{F_{(M_o-1)} + F_{(M_o+1)}} a_{(M_o)} \quad (\text{Moda para frequências absolutas}) \quad (6)$$

$$M_o = l_{i(M_o)} + \frac{f_{(M_o+1)}}{f_{(M_o-1)} + f_{(M_o+1)}} a_{(M_o)} \quad (\text{Moda para frequências relativas}) \quad (7)$$

Onde,

$l_{i(M_o)}$ é o limite inferior da classe modal;

$F_{(M_o-1)}$, $f_{(M_o-1)}$ são respectivamente as frequências absoluta e relativa da classe anterior à classe modal;

$F_{(M_o+1)}$, $f_{(M_o+1)}$ são respectivamente as frequências absoluta e relativa subsequente à classe modal;

$a_{(M_o)}$ é a amplitude da classe modal.

3.9.4 Média geométrica

É particularmente útil para calcular a taxa percentual de crescimento médio, porém a média geométrica é sensível a valores extremos, e conseqüentemente, pode não ser uma medida de tendência central consistente em alguns casos.

Sejam $X_1, X_2, X_3, \dots, X_n$ valores de X , associados às frequências absolutas $F_1, F_2, F_3, \dots, F_n$, respectivamente. A média geométrica de X é definida por:

$$Mg = \sqrt[n]{X_1^{F_1} \cdot X_2^{F_2} \cdot X_3^{F_3} \cdot \dots \cdot X_n^{F_n}} \quad (8)$$

No caso em que $F_1 = F_2 = F_3 = \dots = F_n = 1$, tem -se:

$$Mg = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n} \quad (9)$$

Exemplo 14: Calcular a média geométrica dos valores 3, 6, 12, 24 e 48:

$$Mg = \sqrt[5]{3 \times 6 \times 12 \times 24 \times 48}$$

$$Mg = \sqrt[5]{248.832} \therefore Mg = 12$$

3.9.5 Média Harmônica

“Em situações onde proporcionalidade inversa esteja presente é aconselhável o cálculo da média harmônica. Por exemplo, quando estudamos fenômenos como a velocidade média, o custo médio de bens comprados com uma quantia fixa, etc., deverá ser usada a média harmônica” (REIS, 2009, p.75).

Sejam $X_1, X_2, X_3, \dots, X_n$ valores de X , associados às frequências absolutas $F_1, F_2, F_3, \dots, F_n$, respectivamente. A média harmônica de X é definida por:

$$Mh = \frac{n}{\frac{F_1}{X_1} + \frac{F_2}{X_2} + \frac{F_3}{X_3} + \dots + \frac{F_n}{X_n}} = \frac{n}{\sum_{i=1}^n \frac{F_i}{X_i}} \quad (10)$$

Se $F_1 = F_2 = F_3 = \dots = F_n = 1$, tem -se:

$$Mh = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + \dots + \frac{1}{X_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}} \quad (11)$$

Exemplo 15: Calcular a média harmônica para 2, 5, 8.

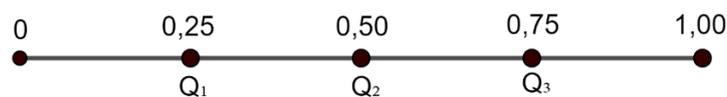
$$Mh = \frac{3}{\frac{1}{2} + \frac{1}{5} + \frac{1}{8}} = 3,64$$

3.10 MEDIDAS SEPARATRIZES

Conforme Reis (2009), são medidas de tendência não central que informam a posição dos valores das variáveis. Ferreira (2015), ensina que as medidas separatrizes ou de ordenamento são: quartis, decis e percentis.

3.10.1. Quartis

É uma delimitação para o percentil. Existem três quartis que vão dividir os dados ordenados em quatro partes iguais, assim:



O primeiro quartil (Q_1) separa os 25% inferiores dos 75% superiores, o segundo quartil (Q_2) é a mediana, ou seja, divide o conjunto de dados ao meio, e o terceiro quartil (Q_3) é o que separa os 75% valores inferiores dos 25% superiores. Como os quartis são medidas separatrizes, preliminarmente é necessário ordenar os dados da distribuição. O primeiro quartil (Q_1) é a observação que ocupa a posição $\frac{n}{4}$. O segundo quartil (Q_2) é a observação que ocupa a observação $\frac{2n}{4}$ e o terceiro quartil (Q_3) é a observação que ocupa a posição $\frac{3n}{4}$. Ao efetuar essas divisões, para encontrar os quartis, pode ocorrer do resultado ser um número inteiro ou um número fracionário. Então, adota-se a seguinte convenção:

- Se o resultado da divisão for um número fracionário, arredonda-se para cima e o valor do quartil será a observação encontrada nesta posição;
- Se o resultado da divisão for um número inteiro, o quartil será a média aritmética da observação que ocupar a posição encontrada com a observação que ocupar a posição imediatamente seguinte.

Exemplo 16: Abaixo estão listadas as medidas da pressão atmosférica ao nível da estação, em milibares (Mb), para o primeiro dia do ano de 2023 em Manaus, coletadas no *site* do INMET.

1005,0	1003,9	1006,6	1003,3
1005,8	1003,9	1007,2	1002,1
1005,8	1004,2	1007,4	1002,0
1005,6	1004,5	1006,9	1002,0
1005,8	1005,2	1006,0	1002,6
1004,6	1006,0	1004,7	1003,8

Para encontrar os quartis, é necessário ordenar esse conjunto de dados. Portanto:

1002,0	1003,9	1005,0	1006,0
1002,0	1003,9	1005,2	1006,0
1002,1	1004,2	1005,6	1006,6
1002,6	1004,5	1005,8	1006,9
1003,3	1004,6	1005,8	1007,2
1003,8	1004,7	1005,8	1007,4

- Posição do 1º quartil (Q_1): $Q_1 = \frac{n}{4} = \frac{24}{4} = 6$.

Como a divisão resultou em um número inteiro, o primeiro quartil é o resultado da média aritmética entre o valor que está na sexta posição e o valor que está na sétima posição.

$$Q_1 = \frac{1003,8 + 1003,9}{2} = 1003,85$$

Logo, 25% das observações são menores ou iguais a 1003,85 milibares e, 75% das observações são maiores ou iguais a 1003,85 milibares.

- Posição do 2º quartil (Q_2): $Q_2 = \frac{2n}{4} = \frac{2 \times 24}{4} = 12$

Como a divisão resultou em um número inteiro, o segundo quartil é o resultado da média aritmética entre o valor que está na décima segunda posição e o valor que está na décima terceira posição.

$$Q_2 = \frac{1004,7 + 1005,0}{2} = 1004,85$$

Tem-se que 50% das observações são menores ou iguais a 1004,85 milibares e 50% das observações são maiores ou iguais a 1004,85 milibares.

- Posição do 3º quartil (Q_3): $Q_3 = \frac{3n}{4} = \frac{3 \times 24}{4} = 18$

Como a divisão resultou em um número inteiro, o terceiro quartil é o resultado da média aritmética entre o valor que está na décima oitava posição e o valor que está na décima nona posição.

$$Q_3 = \frac{1005,8 + 1006,0}{2} = 1005,9$$

Neste conjunto de dados, 25% das observações são maiores ou iguais a 1005,9 milibares e 75% das observações são menores ou iguais a 1005,9 milibares.

3.10.2 Decis

Analogamente, segue-se o mesmo processo dos quartis para o cálculo dos decis, isto é, o decil D_k será a observação que ocupa a posição $\frac{(K \times n)}{10}$, onde $k = 1, 2, \dots, 9$ e n é número de dados do conjunto.

No exemplo anterior, o decil D_6 será a observação que ocupar a posição $\frac{(6 \times 24)}{10} = 14,4$. Como o resultado da divisão resultou em um valor fracionário (14,4), arredonda-se este pra cima, concluindo que na décima quinta posição estará o sexto decil, ou seja, $D_6 = 1005,6$. Portanto, 60% das observações são menores ou iguais a 1005,6 milibares.

3.10.3 Percentis

É a divisão (estratificação) de um conjunto de dados em cem partes. Refere-se sempre aos menores valores. Por exemplo, o décimo percentil refere-se aos 10% menores dados.

Para o cálculo do percentil, também, repete-se o método utilizado nos quartis e nos decis. O percentil P_k será a observação que ocupar a posição $\frac{(K \times n)}{100}$, onde $k = 1, 2, \dots, 99$ e n é o número de dados do conjunto.

Ainda no exemplo anterior, o percentil P_{87} será a observação que ocupar a posição $\frac{(87 \times 24)}{100} = 20,88$ no conjunto de dados ordenados. Como a divisão resultou em um valor fracionário, arredonda-se para 21. Assim, o percentil P_{87} é o valor que está na vigésima primeira posição.

$$P_{87} = 1006,6$$

Neste conjunto de dados, 13% das observações são maiores ou iguais a 1006,6 milibares.

3.10.4 Desvio interquartilico D_i

Também denominado de amplitude interquartilica, é uma medida estatística de dispersão. É o resultado da diferença entre o terceiro quartil (Q_3) e o primeiro quartil (Q_1):

$$D_i = Q_3 - Q_1 \quad (12)$$

Segundo Costa (2011), a amplitude interquartílica é menos afetada por valores extremos do que a amplitude total, sendo uma medida de amplo uso.

3.11 MEDIDAS DE VARIABILIDADE OU DISPERSÃO

As medidas de dispersão são empregadas para detalhar a variabilidade ou dispersão dos dados em um conjunto de observações. Essas medidas oferecem noções sobre o quanto estão dispersos ou próximos os dados estão em torno da média.

3.11.1. Variância

A variância serve pra mostrar o quão distante os dados estão da média. Se a variância é pequena, significa que os dados estão uniformemente distribuídos ou que todos estão muito próximos uns dos outros. E quanto maior for variância, maior será dispersão dos dados

Segundo Paiva (2015), é uma medida de dispersão que “indica o afastamento dos valores dos elementos de uma amostra em relação à média aritmética.” Simbolizada por σ^2 , denomina-se variância a média aritmética entre os quadrados dos desvios dos valores de uma amostra, ou seja:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} \quad (13)$$

Onde, \bar{x} é a média aritmética de uma amostra de números $x_1, x_2, x_3, \dots, x_n$.

Exemplo 17: Gustavo e Lucas tiveram a mesma média no vestibular conforme o boletim de notas a seguir:

Boletim de Gustavo		Boletim de Lucas	
Disciplina	Nota	Disciplina	Nota
Biologia	7,0	Biologia	7,0
História	7,5	História	6,5
Geografia	8,0	Geografia	8,0
Português	7,0	Português	6,5
Inglês	6,0	Inglês	7,5
Matemática	7,0	Matemática	7,5
Física	6,5	Física	6,0
Química	7,0	Química	7,0

Como eles disputavam a última vaga, foi adotada como critério de desempate, a variância do conjunto de notas em todas as disciplinas: o candidato com desempenho mais regular teve direito à vaga. Entende-se por desempenho mais regular aquele cujas notas apresentaram menor dispersão em relação à média aritmética (PAIVA, 2015).

Então, para saber quem ficou com a vaga, na situação hipotética do Exemplo 17, é necessário calcular a variância de ambos os candidatos.

Porém, primeiramente, calcula-se a média das notas, para em seguida determinar a variância.

- Cálculo da média aritmética de Gustavo (\bar{X}_G):

$$\bar{X}_G = \frac{7,0 + 7,5 + 8,0 + 7,0 + 6,0 + 7,0 + 6,5 + 7,0}{8} = \frac{56}{8} = 7,0$$

- Cálculo da variância das notas de Gustavo (σ_G^2):

$$\sigma_G^2 = \frac{(7,0 - 7,0)^2 + (7,5 - 7,0)^2 + (8,0 - 7,0)^2 + (7,0 - 7,0)^2 + (6,0 - 7,0)^2 + (7,0 - 7,0)^2 + (6,5 - 7,0)^2 + (7,0 - 7,0)^2}{8}$$

$$\sigma_G^2 = 0,3125$$

- Cálculo da média aritmética de Lucas (\bar{X}_L):

$$\bar{X}_L = \frac{7,0 + 6,5 + 8,0 + 6,5 + 7,5 + 7,5 + 6,0 + 7,0}{8} = \frac{56}{8} = 7,0$$

- Cálculo da variância das notas de Lucas (σ_L^2):

$$\sigma_L^2 = \frac{(7,0 - 7,0)^2 + (6,5 - 7,0)^2 + (8,0 - 7,0)^2 + (6,5 - 7,0)^2 + (7,5 - 7,0)^2 + (7,5 - 7,0)^2 + (6,0 - 7,0)^2 + (7,0 - 7,0)^2}{8}$$

$$\sigma_L^2 = 0,375$$

Portanto, depreende-se dos cálculos acima que o candidato Gustavo teve o desempenho mais regular, pois a dispersão de seu conjunto de notas foi menor, ficando assim, com a vaga.

3.11.2 Desvio padrão

O desvio padrão é a raiz quadrada positiva da variância (MORETTIN, 2017). É uma medida de variabilidade que traz o valor da variância para uma ordem de grandeza que está mais próximo dos valores do conjunto de dados em questão. De outro modo pode-se dizer que se fôssemos substituir cada valor da distribuição dos dados pelo valor da média aritmética, nos desviaríamos numericamente numa faixa de mais ou menos o valor obtido no desvio padrão.

Segundo Azevedo (2016), as principais propriedades do desvio padrão são:

1. Somando-se ou subtraindo-se a cada elemento de um conjunto de dados, o desvio padrão não se altera;
2. Dividindo-se ou multiplicando-se cada elemnto de um conjunto de dados por uma constante, o desvio padrão fica multiplicado ou dividido por esta constante.

3.11.3 Coeficiente de Variação

Para Fonseca (2011), o coeficiente de variação CV é uma medida relativa de dispersão adequada para uma comparação percentual da intensidade de acumulação dos dados ao redor da média entre séries distintas. É dada por:

$$CV = \frac{\sigma}{\bar{X}} \times 100 \quad (14)$$

ou

$$CV = \frac{S}{\bar{x}} \times 100 \quad (15)$$

Onde, σ = desvio-padrão populacional

\bar{X} = média populacional

S = desvio padrão amostral

\bar{x} = média amostral

Exemplo 18: Numa empresa, o salário médio dos homens é de R\$ 4.000,00, com desvio padrão de R\$ 1.500,00, e o das mulheres é de R\$ 3000,00, com desvio padrão de R\$ 1.200,00. Então:

$$\text{para os homens } CV = \frac{\sigma}{\bar{X}} \times 100 = \frac{1500}{4000} \times 100 = 37,5\% \text{ e}$$

para as mulheres $CV = \frac{\sigma}{\bar{X}} \times 100 = \frac{1200}{3000} \times 100 = 40\%$.

Portanto, conclui-se que os salários das mulheres indicam maior dispersão que o dos homens.

Ainda de acordo com o autor supracitado (FONSECA, 2011), a distribuição terá pequena dispersão quando o coeficiente der até 10%; terá média dispersão quando estiver acima de 10% até 20%; e a dispersão será alta quando ultrapassar 20%.

Determinados pesquisadores consideram:

Baixa dispersão: $CV \leq 15\%$

Média dispersão: $15\% < CV < 30\%$

Alta dispersão: $CV \geq 30\%$

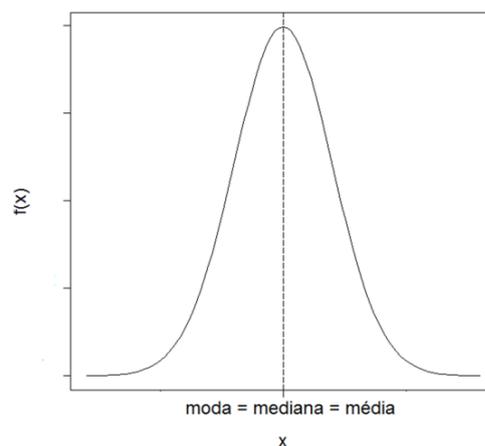
3.12 MEDIDAS DE ASSIMETRIA

De acordo com Fonseca e Martins (2011), a assimetria corresponde à intensidade do distanciamento de uma distribuição da unidade de simetria. Se uma distribuição é simétrica, então tem-se igualdade entre os valores da média, mediana e moda,

$$\bar{X} = M_d = M_o$$

A Figura 10 é um exemplo gráfico de uma distribuição simétrica.

Figura 10: gráfico de uma distribuição simétrica.



Tem-se em uma distribuição assimétrica positiva, ou assimétrica à direita, que o valor da moda é estritamente menor que o da mediana que por sua vez é estritamente menor que a média aritmética dos dados, ou seja,

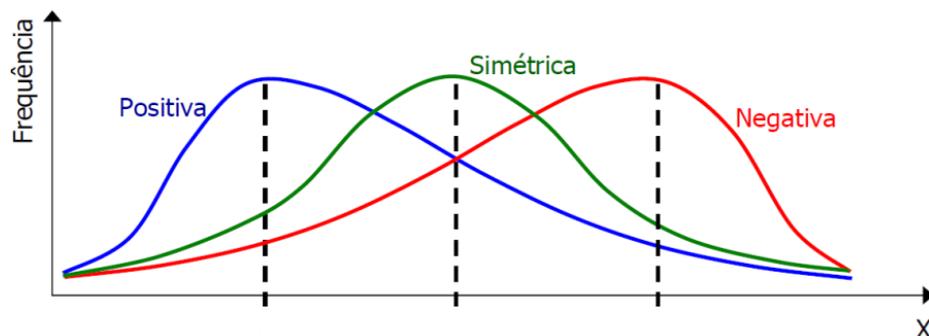
$$M_o < M_d < \bar{X}$$

Por outro lado, uma distribuição assimétrica negativa, ou assimétrica à esquerda, tem-se que a moda é estritamente maior que a mediana que é estritamente maior que a média aritmética dos dados, expressa pela desigualdade

$$\bar{X} < M_d < M_o$$

O próximo gráfico exibe uma sobreposição das três curvas conceituadas acima, de uma distribuição simétrica e assimétrica positiva e negativa. A curva simétrica (verde), é também conhecida como curva de Gauss, curva normal ou distribuição normal. As curvas em azul e vermelho são respectivamente de uma distribuição assimétrica à direita e à esquerda. Note que o ponto de máximo de cada uma das três curvas corresponde à moda de cada um dos três conjuntos de dados.

Figura 11: gráfico de uma distribuição simétrica e assimétrica.



Fonte: https://docs.ufpr.br/~rtkishi.dhs/TH024/TH024_10_Frequencia_cheias.pdf.
Acesso em: 03 jun. 2023.

Segundo Fonseca e Martins (2011), existem diversas fórmulas que calculam o coeficiente de assimetria, dentre elas são úteis:

1º Coeficiente De Pearson

$$AS = \frac{\bar{X} - M_o}{\sigma} \quad (16)$$

ou

$$AS = \frac{\bar{X} - M_o}{S} \quad (17)$$

Onde,

σ : é o desvio padrão populacional;

S : é o desvio padrão amostral;

\bar{X} : é a média aritmética dos dados;

M_o : é a moda dos dados.

2º Coeficiente de Pearson

$$AS = \frac{Q_1 + Q_3 - 2M_d}{Q_3 - Q_1} \quad (18)$$

Na condição de:

$AS = 0$ tem-se que a distribuição é simétrica;

$AS > 0$ tem-se que a distribuição é assimétrica positiva;

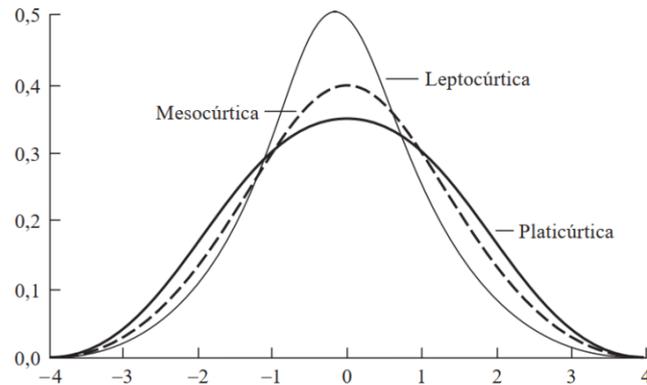
$AS < 0$ tem-se que a distribuição é assimétrica negativa.

3.12.1 Medidas de curtose

Curtose é o grau de achatamento da distribuição. Uma distribuição é dita mesocúrtica quando possui um achatamento moderado das caudas em relação à distribuição normal (simétrica); em outras palavras, pode-se dizer que não é uma curva achatada nem delgada.

A distribuição será leptocúrtica quando as caudas são mais estreitas e maiores em comparação com uma distribuição normal. Uma distribuição leptocúrtica mostra uma aglomeração maior de valores em torno da média, podendo assinalar uma maior instabilidade nos dados, implicando em uma maior probabilidade de ocorrer valores extremos.

Uma distribuição mais achatada em comparação com uma distribuição normal, é denominada de platicúrtica. A menor concentração de valores em torno da média pode ser um indicativo de menor variabilidade nos dados, com menor probabilidade de ocorrência de valores extremos.

Figura 12: medidas de curtose.

Fonte: <https://blog.proffernandamaciel.com.br/assimetria-e-curtose-dos-dados/>. Acesso em 06 jun. 2023.

Para medir o grau de curtose (coeficiente), utiliza-se a fórmula (16):

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} \quad (19)$$

Em que:

- Q_3 é o 3º quartil;
- Q_1 é o 1º quartil;
- Q_{90} é o 90º percentil;
- Q_{10} é o 10º percentil.

- Se $K = 0,263$, a distribuição de frequência é mesocúrtica;
- Se $K > 0,263$, a distribuição de frequência é palticúrtica;
- Se $K < 0,263$, a distribuição de frequência é leptocúrtica.

3.13 CORRELAÇÃO E REGRESSÃO

A correlação e a regressão são dois conceitos importantes na análise estatística de dados e estão estreitamente relacionados. Ambos são aplicados para apurar a relação entre duas ou mais variáveis e contribuem para a compreensão de padrões e tendências nos dados.

A correlação é uma ferramenta estatística que visa medir o grau de relação entre duas ou mais variáveis dentro de um conjunto de dados. Seu propósito é avaliar a intensidade e a direção dessa associação, indicando o quanto essas variáveis variam juntas.

A regressão é um método estatístico que possibilita modelar a associação entre uma variável dependente (variável que se pretende prever) e uma ou mais variáveis independentes (variável que se usa para prever). A finalidade da regressão é obter a equação que melhor representa essa associação.

3.13.1 Coeficiente de determinação R^2

Para Kennedy (2008), admite-se que o coeficiente de determinação seja uma medida que representa a proporção da variação da variável dependente (Y) que pode ser explicada pela variação da variável independente (X).

O coeficiente de determinação é obtido pela equação (20), de acordo com Heij *et al.* (2004, p. 83).

$$R^2 = \frac{(\sum(x_i - \bar{x})(y_i - \bar{y}))^2}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2} \quad (20)$$

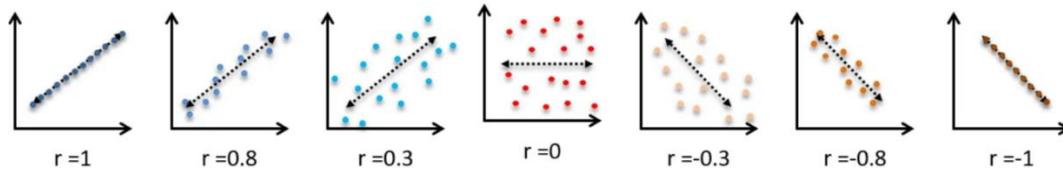
3.13.2 Coeficiente de correlação de Pearson

Segundo Ayres *et al* (2007), o coeficiente de correlação de Pearson é um teste estatístico, paramétrico, amplamente empregado em bioestatística que avalia a relação entre as variáveis X e Y , cujos valores são medidos em escala intervalar ou de razões. O coeficiente de Pearson, representado por "r", varia de -1 a +1, sendo que quanto mais próximos desses extremos os valores se encontrarem, maior será a força da associação entre as variáveis analisadas. Um valor de zero para esse coeficiente indica ausência de correlação.

O coeficiente de correlação de Pearson é dado pela equação (21), de acordo com Crespo (2002, p.151).

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}} \quad (21)$$

Figura 13: análise da correlação entre duas variáveis a partir do gráfico de dispersão.



Fonte: <https://medium.com/rela%C3%A7%C3%B5es-entre-dados/rela%C3%A7%C3%B5es-entre-dados-bc8329f02cea>. Acesso em: 17 jul. 2023.

- $r = 1$ significa perfeita correlação positiva.
- $r = 0,8$ significa alta correlação positiva.
- $r = 0,3$ significa baixa correlação positiva.
- $r = 0$ significa nenhuma correlação.
- $r = 0,3$ significa baixa correlação negativa.
- $r = -0,8$ significa alta correlação negativa.
- $r = -1$ significa perfeita correlação negativa.

3.13.3 Retas De Regressão (ou retas dos mínimos quadrados ou retas de ajuste)

De acordo com Correa (2003), a correlação linear é um tipo de relação entre duas variáveis em que o gráfico se aproxima de uma reta. Esse padrão de relação é visualmente representado no diagrama de dispersão, em um gráfico cartesiano. Para uma melhor avaliação da correlação entre as variáveis, é útil obter a equação dessa reta, conhecida como reta de regressão, e sua representação matemática é chamada de equação de regressão. O diagrama de dispersão é construído de acordo com os dados amostrais de n observações.

Segundo Chein (2019), a reta de regressão depende de cinco medidas estatísticas básicas: média de X ; desvio padrão de X ; média de Y ; desvio padrão de Y ; correlação de X e Y .

Conforme Reis (2009, p. 172), a reta de regressão é dada pela expressão:

$$Y = a + bX \quad (22)$$

Onde X é a variável independente e Y é a variável dependente.

E cujos parâmetros b e a são respectivamente coeficiente angular e intercepto, dados por:

$$a = \bar{Y} - b\bar{X} \quad (23)$$

$$b = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} \quad (24)$$

Sendo que:

\bar{Y} é o valor médio da variável y ;

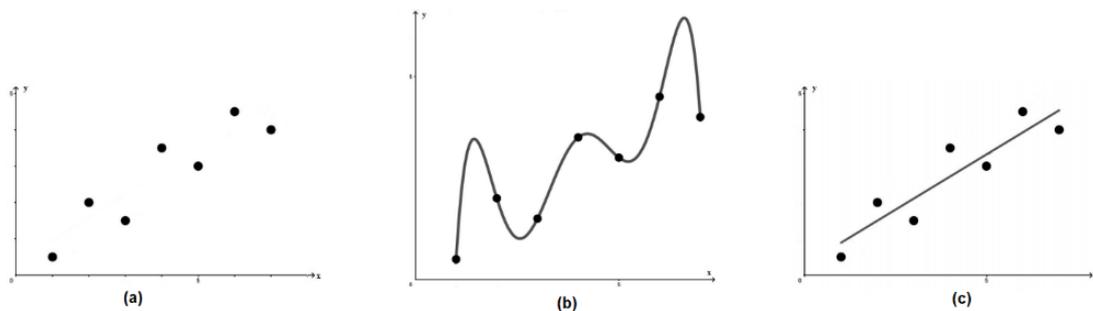
\bar{X} é o valor médio da variável x ;

n é o número de observações dos dados amostrais.

Vamos à explicação deste fato, Equações 23 e 24, pelo método dos mínimos quadrados, conforme Sousa Neto (2018):

A regressão linear por mínimos quadrados é uma técnica estatística utilizada para encontrar uma reta que melhor se ajuste a um conjunto de pares de dados. Essa reta é escolhida de forma que minimize a soma dos quadrados das diferenças entre os valores observados e os valores previstos pela reta.

Figura 14: ajuste de curvas a pontos no plano cartesiano. (a) Representação no plano cartesiano de 7 pontos coletados em uma pesquisa. (b) Ajuste polinomial. (c) Ajuste por mínimos quadrados.



Fonte: Sousa Neto (2018).

O método da interpolação polinomial Figura 14 (b), nem sempre é o mais adequado para obter um ajuste preciso de curva. Ao examinarmos a disposição dos pontos no plano cartesiano, torna-se evidente uma relação crescente entre as coordenadas x e y de cada ponto, por meio de um simples exame visual. Se optarmos por utilizar um polinômio interpolador de grau 6, Figura 14 (b), para associar os dados, obteremos uma curva que passará precisamente por todos os 7 pontos.

No entanto, devido à ampla variabilidade presente nos dados, a curva resultante oscilará, o que pode acarretar em valores intermediários previstos com uma significativa margem de

erro, pois esses valores preditos podem não seguir o mesmo padrão observado nos demais pontos coletados. Uma técnica para empregada para corrigir ou reduzir esse tipo de erro é traçar uma curva que siga a tendência geral dos dados, sem a necessidade de passar exatamente pelos pontos coletados, a Figura 14 (c), exibe que a utilização de uma reta ajustada a essa distribuição demonstra como esse modelo pode representar adequadamente a tendência geral dos dados, mesmo sem passar por nenhum dos pontos coletados.

A regressão por mínimos quadrados é um método matemático utilizado para encontrar a curva que melhor se ajusta aos dados coletados, minimizando a discrepância entre esses dados e os pontos da curva que representam a tendência geral dos mesmos.

3.13.4 Método dos mínimos quadrados

A finalidade deste método é determinar uma reta que melhor se ajuste aos pares de pontos: (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , ..., (x_n, y_n) . A reta é caracterizada pela equação:

$$y = a_0 + a_1x + e \quad (24)$$

Onde:

- a_0 coeficiente da interseção da reta de regressão com o eixo y ;
- a_1 coeficiente angular (inclinação) da reta;
- e erro (resíduo) entre a reta de regressão e os pontos dados.

Isolando o erro e na equação (24), tem-se que,

$$e = y - a_0 - a_1x. \quad (25)$$

Compreende-se claramente na equação 23, que o erro e caracteriza a discrepância entre o valor original y do dado e o valor aproximado, $a_0 + a_1x$, sugerido na equação (24).

Com o propósito de identificar uma linha reta que se adeque de forma mais precisa aos dados fornecidos, empreende-se o método dos mínimos quadrados para minimizar o erro inerente à equação (25) utilizada, isto é, o procedimento é minimizar a soma dos quadrados dos erros entre o valor medido y e o valor calculado pelo modelo linear $a_0 + a_1x$, conforme, a seguir, na equação (26). Este critério apresenta a vantagem de produzir uma única reta que melhor se ajusta a um conjunto específico de dados, evitando, dessa forma, a existência de múltiplas soluções para o mesmo problema.

$$S_r = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2 \quad (26)$$

Para determinar os coeficientes a_0 e a_1 que minimizem a equação (26), é necessário derivar a mesma em relação aos coeficientes e iguala-se à zero, pois, sabe-se do cálculo diferencial, que igualando-se a primeira derivada à zero, obtém-se seu valor de máximo ou de mínimo. No caso em questão tem-se o valor de mínimo, tendo em vista que o coeficiente angular a_1 é negativo.

A derivada (parcial) em relação a cada coeficiente está no sistema (27):

$$\begin{cases} \frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^n (y_i - a_0 - a_1 x_i) = 0, \\ \frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^n [(y_i - a_0 - a_1 x_i) x_i] = 0 \end{cases} \quad (27)$$

Os somatórios das duas equações do sistema (27) variam de $i = 1$ a n , se decide por simplificar os símbolos adotando a prática de separar as parcelas dentro dos somatórios em termos individuais, que resulta no sistema (28).

$$\begin{cases} \sum y_i - \sum a_0 - \sum a_1 x_i = 0, \\ \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2 = 0. \end{cases} \quad (28)$$

A expressão mencionada acima pode ser simplificada no contexto de um sistema de equações lineares com duas variáveis (a_0 e a_1), pois $\sum a_0 = n a_0$, resultando no sistema (29), cujas equações são chamadas de equações normais.

$$\begin{cases} n a_0 + (\sum x_i) a_1 = \sum y_i \\ (\sum x_i) a_0 + (\sum x_i^2) a_1 = \sum x_i y_i \end{cases} \quad (29)$$

A fim de garantir que o sistema apresente uma única solução, é imprescindível que a matriz formada pelos coeficientes das equações normais possua um determinante não nulo, isto significa:

$$\det \mathbf{A} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \neq 0,$$

e que ocorrerá, desde que $n \sum x_i^2 \neq (\sum x_i)^2$.

A resolução do sistema (29) para a variável a_1 e em seguida para a variável a_0 resulta que:

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (30)$$

$$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n} \quad (31)$$

Verifica-se que as equações (30) e (31) são respectivamente semelhantes as Equações (24) e (23) publicadas por Reis (2009), como queria-se demonstrar.

Exemplo 19: A partir dos pares ordenados (1, 70), (2, 50), (3, 40), (4, 30), (5, 20), (6, 10): calcular o coeficiente de correlação; determinar a reta ajustada; estimar o valor Y para $X = 0$. (CRESPO, 2002).

Inicialmente, para tornar o cálculo prático, faz-se necessário construir o quadro a seguir.

X	Y	X ²	Y ²	XY	
1	70	1	4900	70	
2	50	4	2500	100	
3	40	9	1600	120	
4	30	16	900	120	
5	20	25	400	100	
6	10	36	100	60	
SOMA	21	220	91	10400	570

Em seguida, faz-se as substituições na equação (18) para calcular o coeficiente de correlação de Pearson.

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}} = \frac{6 \times 570 - 21 \times 220}{\sqrt{[6 \times 91 - (21)^2][6 \times 10400 - (220)^2]}}$$

$$r \cong -0,99$$

Ou seja, tem-se um coeficiente de correlação próximo de -1 , o que indica uma forte associação negativa entre as variáveis x e y .

Para determinar a reta ajustada, faz-se uso da equação (21) para calcular o coeficiente angular, e em seguida usa-se a equação (20) para estimar o intercepto da equação linear de regressão.

$$b = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{6 \times 570 - 21 \times 220}{6 \times 91 - (21)^2} = \frac{-1200}{105} \cong -11,4$$

$$a = \bar{Y} - b\bar{X} = 36,7 - (-11,4 \cdot 3,5) = 76,6$$

Logo, a reta que melhor se ajusta aos pares de pontos em questão é:

$$Y = -11,4X + 76,6$$

Portanto, o valor estimado para Y , quando $X = 0$ é:

$$Y = 76,6$$

Diante de todo o básico conceitual da estatística deixado neste capítulo, reafirma-se a importância desse estudo em um mundo que tem e produz inúmeras informações e dados. Por isso é fundamental estudar e aplicar tais métodos, para auxiliar a entender e dar sentido a esses dados. Qualquer interpretação inferencial mais elaborada dos dados ou temas mais avançados, passou primordialmente pelos cálculos dessas medidas.

Através do uso de tabelas, gráficos e medidas de resumo como média, mediana, moda, desvio padrão, entre outras, a estatística descritiva torna os dados mais compreensíveis e contribui a esclarecer padrões e tendências, o que é fundamental para embasar decisões informadas. Além disso, possibilita a identificação de valores atípicos ou discrepantes nos dados e permite a comparação entre diferentes grupos ou categorias, destacando suas diferenças e semelhanças. Ao analisar padrões ao longo do tempo ou em diferentes grupos, é possível obter uma compreensão mais aprofundada do comportamento dos fenômenos estudados.

A estatística é empregada na avaliação de riscos em várias esferas, tais como seguros, investimentos e decisões estratégicas. Através da análise de dados históricos e do uso de modelos estatísticos, é viável estimar os riscos e implementar medidas apropriadas para suavizá-los. Em várias carreiras, o domínio de dados e habilidades em estatística é altamente apreciado. Ao adquirirem conhecimento em medidas estatísticas, os profissionais se tornam mais capacitados e aptos a desempenhar um papel significativo em suas respectivas áreas de atuação. Tendo em vista toda essa relevância, o professor nesse processo de ensino, deve articular uma metodologia que conduza o aluno pelos caminhos pedagógicos que os leve a conquistar o pleno entendimento dos gráficos, das formulações matemáticas e os indispensáveis saberes dessa matemática aplicada.

4 MÉTODO E RESULTADOS

De posse dos conceitos básicos em estatística, elencados no capítulo anterior, já se tem o necessário para se fazer interpretações dos relatórios descritivos gerados a partir dos dados de entrada, no *software*, que foram coletados no INMET⁴. Preliminarmente será apresentado os procedimentos e explicações referentes as informações que foram tratadas pelo BioEstat.

⁴ Instituto Nacional de Meteorologia (INMET). Estações de dados. Dados históricos. Disponível em: www.inmet.gov.br/portal. Acesso: 10 de dez. 2022.

4.1 METODOLOGIA DA PESQUISA

O procedimento foi dar tratamento estatístico às normais climatológicas disponíveis no portal do Instituto Nacional de Meteorologia. “As normais climatológicas são valores médios de variáveis meteorológicas calculados para um período relativamente longo e uniforme, compreendendo no mínimo três décadas consecutivas, e representa as características média do clima em um determinado local.” (INMET, 2022).

Foram organizados os dados aferidos na estação meteorológica de Belém do Pará (Norte do Brasil), cuja coordenada geográfica tem latitude -1.43 ($01^{\circ}26'S$) e longitude -48.43 ($48^{\circ}26'W$), situada a 10 metros acima do nível do mar. Com os recursos tecnológicos do programa BioEstat, obteve-se o fornecimento dos relatórios estatísticos-descritivos das variáveis e as explorações dos gráficos de linhas sobrepostas comparando as médias mensais das três séries históricas de cada variável; além da geração de tabelas de frequência, histograma e regressão. No *Microsoft Excel*⁵ foram organizadas as tabelas de apresentação dos dados coletados.

Segundo Miranda e Sousa (2013), a reunião, em um mesmo local, dos instrumentos meteorológicos que são utilizados para obter dados meteorológicos como o termômetro (medidor da temperatura do ar), barômetro (medidor da pressão atmosférica), higrômetro (medidor da umidade relativa do ar), e outros, é denominada estação meteorológica.

Para instalação do programa BioEstat, requer-se o sistema operacional Microsoft Windows XP ou subsequente, com um mínimo de 128 MB de memória RAM. No BioEstat, os valores com casas decimais devem ser grafados usando o ponto decimal. Isto é, não é aceito a vírgula na separação entre a parte inteira e decimal. Por exemplo, 1.5 ~~1,5~~ (Ayres *et al*, 2007).

As distribuições de frequência, e conseqüentemente os respectivos histogramas associados a elas, foram obtidos a partir da entrada do valor da média aritmética mensal das três séries históricas das variáveis estudadas, conforme exibido no Quadro 5:

⁵ O **Microsoft Excel** é um editor de planilhas produzido pela Microsoft para computadores que utilizam o sistema operacional Microsoft Windows, além de computadores Macintosh da Apple Inc. e dispositivos móveis como o Windows Phone, Android ou o iOS. Disponível em: https://pt.wikipedia.org/wiki/Microsoft_Excel Acesso em: 10 jul. 2023.

Quadro 5: média aritmética dos três períodos para as variáveis analisadas no estudo.

Mês	N. ⁶	Pres. ⁷	T.Mi. ⁸	T.Má ⁹ .	Prec. ¹⁰	I. ¹¹	V. ¹²	U.12 ¹³	U.18 ¹⁴	U.24 ¹⁵
JAN	0.7	1009.5	22.6	31.1	387.9	134.2	1.3	90.2	81.5	93.6
FEV	0.8	1009.7	22.7	30.8	416.6	103.4	1.4	91.6	83.3	95.0
MAR	0.8	1009.8	22.9	30.7	467.9	110.6	1.3	90.7	83.8	95.2
ABR	0.8	1009.8	23.1	31.1	414.4	131.1	1.3	90.1	82.7	95.4
MAI	0.7	1010.3	23.0	31.7	309.2	185.2	1.4	88.1	76.5	94.6
JUN	0.6	1011.2	22.7	32.0	182.1	228.3	1.6	86.2	70.9	92.7
JUL	0.6	1011.6	22.3	32.1	155.1	251.1	1.7	86.4	68.6	91.7
AGO	0.5	1011.3	22.4	32.5	130.0	264.9	1.8	85.0	68.7	90.4
SET	0.5	1010.7	22.3	32.6	131.0	240.0	1.9	83.4	70.4	89.0
OUT	0.6	1009.8	22.3	32.6	126.5	235.3	2.0	82.4	72.0	87.4
NOV	0.6	1008.9	22.6	32.7	132.3	200.5	1.9	83.1	73.0	87.0
DEZ	0.7	1009.1	22.6	32.2	251.6	169.8	1.6	86.5	77.2	90.2

Fonte: elaborado pelo autor.

Cada variável climatológica estudada teve suas informações inseridas clicando com o *mouse* na célula desejada no *grid* geral (planilha) e entrando com os valores digitando no teclado do computador. Na matriz de entrada de dados (*grid* geral), o conjunto de dados foi organizado em grupos de três colunas relativas aos três períodos de trinta anos observados, 1961 a 1990, 1981 a 2010 e de 1991 a 2020; e doze linhas representando os meses do ano, janeiro é 1, fevereiro é 2, e assim sucessivamente.

A Figura 15 é uma exibição da janela do BioEstat e alguns termos básicos do programa.

⁶ Nebulosidade em décimos.

⁷ Pressão atmosférica (hPa).

⁸ Temperatura mínima (°C).

⁹ Temperatura máxima (°C).

¹⁰ Precipitação acumulada (mm).

¹¹ Insolação total em horas.

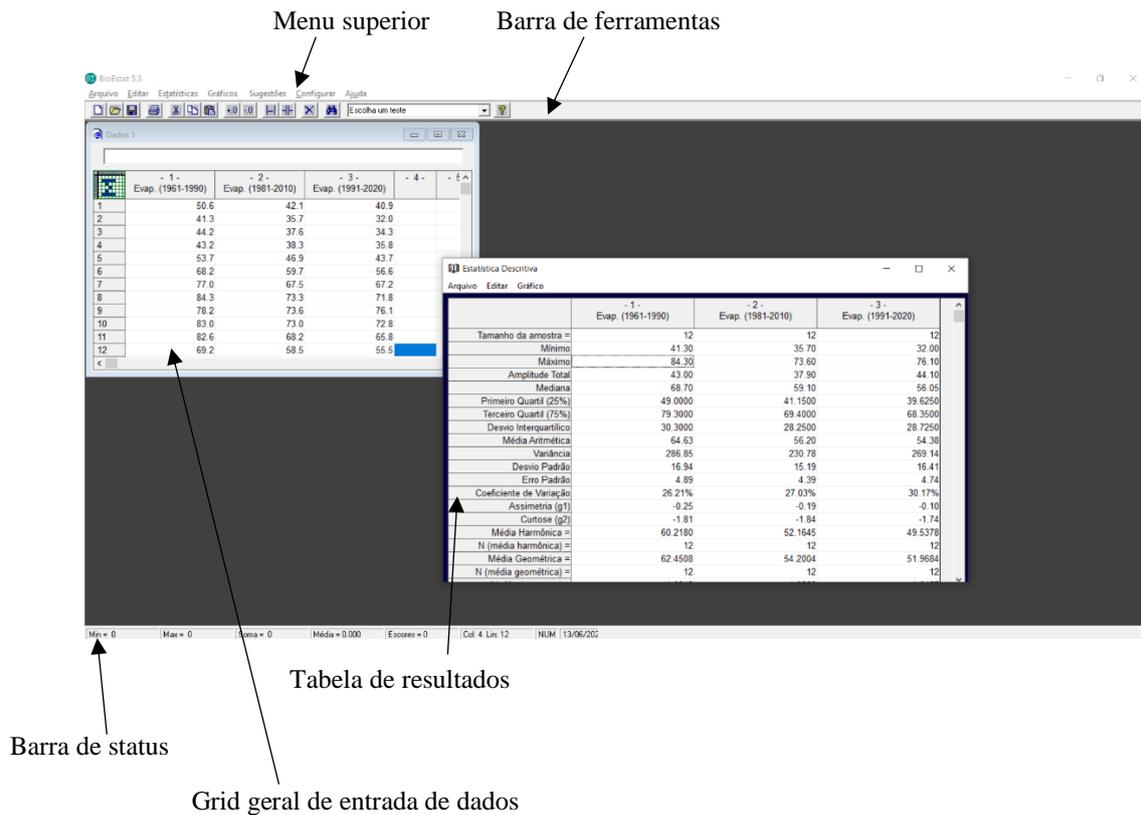
¹² Intensidade do vento (m/s).

¹³ Umidade relativa do ar horária (1200).

¹⁴ Umidade relativa do ar horária (1800).

¹⁵ Umidade relativa do ar horária (2400).

Figura 15: janela do BioEstat.



Fonte: acervo do autor.

O Quadro 6 mostra as instruções de cada aplicação utilizada na pesquisa.

Quadro 6: instruções das aplicações utilizadas na pesquisa.

APLICAÇÃO	INSTRUÇÕES
Gráfico de coluna	<ul style="list-style-type: none"> ➤ Clicar o mouse em Colunas e pressionar em Simples; ➤ Selecionar as duas colunas do <i>grid</i> e pressionar o <i>mouse</i> em Executar.
Gráfico de linha	<ul style="list-style-type: none"> ➤ Clicar em linha (curva) no <i>menu</i> superior, em Gráficos; ➤ Selecionar as colunas do <i>grid</i> e pressione em Executar.
Gráfico de setor	<ul style="list-style-type: none"> ➤ Clicar em Setor, no <i>menu</i> superior, em Gráficos; ➤ Na janela do gráfico selecione Configurações; ➤ Na aba Outras, selecione entre Absolutos ou Percentuais; ➤ e) Selecionar as colunas do <i>grid</i> e pressionar o <i>mouse</i> em Executar.
Gráfico <i>Box Plot</i>	<ul style="list-style-type: none"> ➤ No <i>menu</i> superior, clicar em Box Plot: mediana e quartis; ➤ Selecionar as colunas do <i>grid</i> e pressionar em Executar, para resultados.
Estatística Descritiva	<ul style="list-style-type: none"> ➤ Introduzir os dados no <i>grid</i> geral do BioEstat; ➤ No <i>menu</i> superior, clicar com o <i>mouse</i> Estatística;

	<ul style="list-style-type: none"> ➤ Clicar o <i>mouse</i> em Estatística Descritiva e, a seguir, em Dados Quantitativos; ➤ Selecionar as colunas do <i>grid</i> (1, 2, 3, ..., <i>n</i>); ➤ Pressionar em Executar o Teste, para resultados.
Tabela de Frequência	<ul style="list-style-type: none"> ➤ Pressione o <i>mouse</i> em Tabelas de Frequência, disponível no <i>menu</i> superior, em Gráficos; ➤ Selecionar a coluna do <i>grid</i> geral e clicar em Executar; ➤ Preencher o formulário Especificação das Classes; ➤ Pressionar o botão Confirmar para obter a Tabela de Frequência.
Histograma	<ul style="list-style-type: none"> ➤ Pressionar o <i>mouse</i> em Histograma disponível no <i>menu</i> superior; ➤ Selecionar a coluna do <i>grid</i> geral e clicar em Executar; ➤ Preencher o formulário Especificação das Classes; ➤ Pressionar o botão Confirmar para obter o histograma.
Regressão	<ul style="list-style-type: none"> ➤ Preencher o <i>grid</i> geral com os escores amostrais, representando: <ul style="list-style-type: none"> - Na primeira coluna os valores da variável X (independente); - Na segunda, os <i>escores</i> da variável Y (dependente); ➤ Clicar em Ajustamento de Curvas e selecionar as colunas do <i>grid</i> geral; ➤ Clicar em Executa Estatística, para obter os resultados.

Fonte: Ayres *et al* (2007).

4.2 RESULTADOS E DISCUSSÃO DAS APLICAÇÕES

A área de Belém, situada na junção do Rio Guamá com a Baía do Guajará, está aproximadamente a 120 km distante do Oceano Atlântico. Ela abrange uma extensão de 719 km², com uma altitude média de 12 m. Além de sua baixa altitude, uma das principais características fisiográficas de Belém é a presença de vários corpos d'água menores, conhecidos localmente como furos e igarapés. Esses elementos, juntamente com a localização em uma baixa latitude, contribuem para um clima quente e úmido na região. (BASTOS *et al*, 2002).

4.2.1. Nebulosidade

Denomina-se nebulosidade a fração da abóboda celeste que, em dado momento, encontra-se encoberto por nuvens. Efetivamente, é estimada visualmente, imaginando todas as nuvens juntas e atribuindo, aproximadamente, a fração do céu que essa união representa. A nebulosidade é expressa em décimos de céu encoberto. Nebulosidade de 0,5 significa que a metade da abóboda celeste está encoberta. O valor zero significa que nenhuma nuvem foi

observada no momento da medição. E 1 representa a totalidade da abóboda encoberta. (VAREJÃO-SILVA, 2006).

Segundo Pereira e Nunes (2013), a condição de céu encoberto implica em decaimento da intensidade do feixe de radiação recebida na estação comparada à recebida no topo da atmosfera. Essa atenuação é devida a alta umidade e a grande quantidade de nuvens. Pois a absorção pelos gases atmosféricos e o espalhamento pelas partículas suspensas na atmosfera fazem diminuir essa radiação ao atravessá-la.

A Tabela 4 refere-se a variável nebulosidade média mensal em décimos observadas na estação Belém-PA.

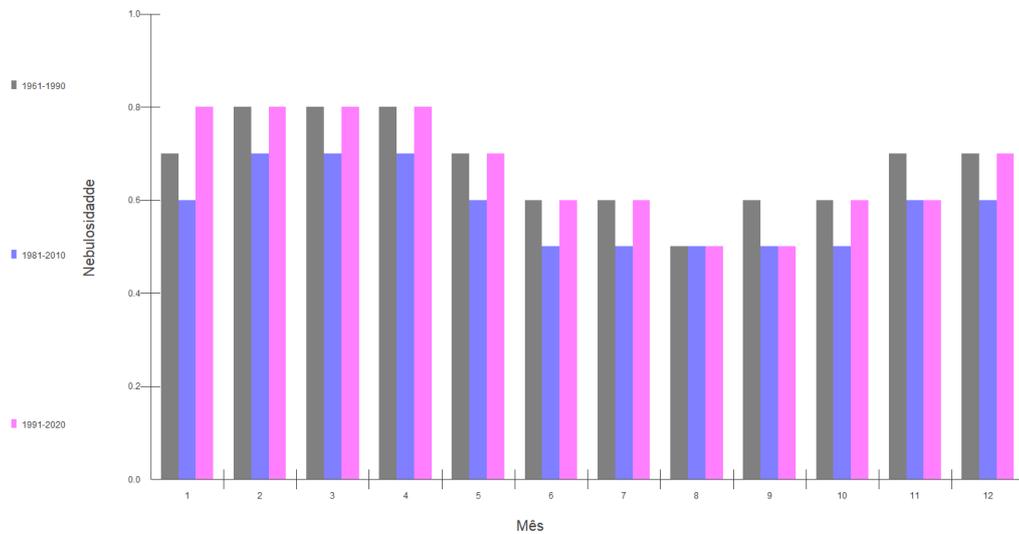
Tabela 4: nebulosidade mensal média em décimos em Belém-PA, no período 1961-2020.

MÊS	1961-1990	1981-2010	1991-2020
JAN	0.7	0.6	0.8
FEV	0.8	0.7	0.8
MAR	0.8	0.7	0.8
ABR	0.8	0.7	0.8
MAI	0.7	0.6	0.7
JUN	0.6	0.5	0.6
JUL	0.6	0.5	0.6
AGO	0.5	0.5	0.5
SET	0.6	0.5	0.5
OUT	0.6	0.5	0.6
NOV	0.7	0.6	0.6
DEZ	0.7	0.6	0.7

Fonte: INMET

Os dados da Tabela 4 estão exibidos no gráfico de colunas justapostas gerado pelo BioEstat a partir da entrada desses valores no programa (Figura15). Optou-se por esse modelo de gráfico, pois os dados desta variável apresentam uma característica discreta. Assim, o gráfico de barras mostrou-se mais apropriado do que o de linhas, pois neste tipo de apresentação há mais sinuosidade visual na distribuição para esses valores postos.

Figura 16: gráfico da nebulosidade mensal média em décimos em Belém-PA, no período 1961-2020.



Fonte: autor (2023).

Verifica-se no gráfico da Figura 16 que os únicos meses do último período estudado, onde a nebulosidade média não foi maior ou igual a nebulosidade média dos períodos anteriores, foram nos meses de setembro e novembro. Apontando que cada vez mais se tem o céu encoberto por nuvens neste local. Depreende-se também do mesmo gráfico que no mês de agosto ocorreu a menor média da nebulosidade para os três períodos.

A seguir, na Tabela 5, o relatório descritivo da nebulosidade mensal média para as três séries analisadas aponta que no último período (1991-2020), os dados se afastaram mais da média em comparação às duas séries dos dois períodos anteriores, fato explicado pelo maior valor do desvio padrão, no que implicou também em um maior coeficiente de variação, que ultrapassou os 17%, para este conjunto.

Tabela 5: estatística descritiva da nebulosidade mensal média em décimos, em Belém-PA, no período 1961-2020.

Período	1961-1990	1981-2010	1991-2020
Amostra (N)	12	12	12
Mínimo	0.5	0.5	0.5
Máximo	0.8	0.7	0.8
Amplitude Total	0.3	0.2	0.3
Mediana	0.7	0.6	0.65
Primeiro Quartil (25%)	0.6	0.5	0.6
Terceiro Quartil (75%)	0.725	0.625	0.8

Desvio Interquartilício	0.125	0.125	0.2
Média Aritmética	0.68	0.58	0.67
Variância	0.01	0.01	0.01
Desvio Padrão	0.1	0.08	0.12
Erro Padrão ¹⁶	0.03	0.02	0.03
Coefficiente de Variação	14.30%	14.31%	17.32%
Assimetria (g1)	-0.14	0.35	-0.06
Curtose (g2)	-0.77	-1.45	-1.47
Média Harmônica	0.6619	0.5727	0.6478
Média Geométrica	0.6685	0.578	0.6573
Variância (geom.) ¹⁷ ¹⁸	1.0094	1.0088	1.0137
Desvio Padrão (geom.) ¹⁹	1.1579	1.1522	1.1938

Fonte: autor (2023).

O valor mínimo observado nos três períodos foi de 0,5, isto é, verificou-se que no mínimo teve-se metade da abóboda celeste encoberta nos três conjuntos de dados. 0,5 também foi a média da nebulosida no período intermediário. Em relação ao coeficiente de assimetria, o período intermediário (1981-2010) é o único com com essa medida positiva, ou seja, com distribuição assimétrica à direita; também neste período identifica-se o menor valor do desvio padrão da nebulosidade.

Tabela 6: distribuição de frequência da nebulosidade mensal média para o período 1961-2020, obtida a partir do Quadro 5.

Classes	Xi	Fi	Percentual
0.50 — 0.58	0.54	2	16.67 %
0.58 — 0.66	0.62	4	33.33 %
0.66 — 0.74	0.70	3	25.00 %
0.74 — 0.82	0.78	3	25.00 %
0.82 — 0.90	0.86	0	0.00 %
TOTAL		12	100.00 %

Fonte: autor (2023).

¹⁶ “O erro padrão é uma medida da precisão da média amostral calculada. O erro padrão obtém-se dividindo o desvio padrão pela raiz quadrada do tamanho da amostra. Ou seja, $\frac{\sigma}{\sqrt{n}}$. Quando não se conhece o desvio padrão da população, usa-se o desvio padrão da amostra (s) ficando a fórmula: $\frac{s}{\sqrt{n}}$. Se de uma população, com média μ_e e desvio padrão σ , se retirarem muitas amostras todas do mesmo tamanho n, e para cada amostra se calcular a respectiva média, a distribuição de todas essas médias é normal com média μ_e desvio padrão $\frac{\sigma}{\sqrt{n}}$. Assim, o erro padrão não é mais do que o desvio padrão da distribuição das médias das amostras de uma população”. (ASSIS, 2019, p. 287).

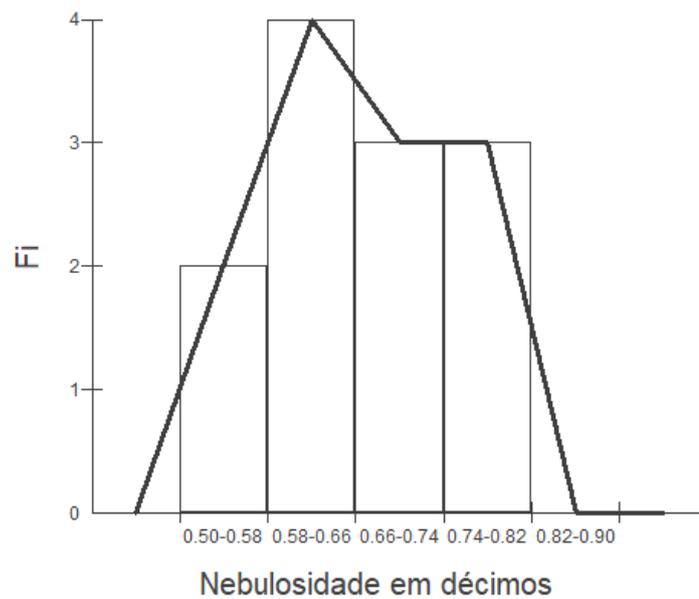
¹⁷ Nota: a variância geométrica é uma medida estatística empregada para medir a dispersão dos valores dentro de um conjunto de dados, porém levando em consideração o produto dos valores em vez de suas somas, como é feito na variância tradicional.

¹⁸ Nota: a variância geométrica é particularmente apropriada para conjuntos de dados que contenham valores positivos com interpretação multiplicativa, como taxas de crescimento ou retornos financeiros.

¹⁹ Nota: para obter o desvio padrão geométrico, basta calcular a raiz quadrada positiva da variância geométrica.

O quadro da distribuição de frequência explícita que a capital paraense tem, com maior frequência, sua abobáda celeste com mais de 60% encoberta por nuvens na localização da estação onde se afere essas medidas. Com base na distribuição acima, obtém-se o histograma da nebulosidade a seguir.

Figura 17: histograma da nebulosidade mensal média Belém-PA, para o período 1961-2020, gerado a partir da Tabela 6.



Fonte: autor (2023).

A assimetria fica evidenciada no gráfico, o que caracteriza uma distribuição não normal dos dados. Apresentou-se duas clases com a mesma frequência, acarretando que metade dos valores foi de no mínimo 70% do céu encoberto, observados na localização da estação que se afere esta medida.

4.2.2 Pressão Atmosférica

A pressão atmosférica é exercida pelos gases que formam a atmosfera, essa camada de gases que contorna o planeta Terra é confinada pela gravidade sobre a superfície. E como se vive na superfície, os que fazem dela sua habitação sofrem influencia de tal pressão. O ar pesa, e a pressão atmosférica equivale a pressão que os mergulhadores experimentam em maiores profundidades. O ar em regiões mais baixas se encontra mais acumulado, fazendo com que o

corpo humano se abasteça do oxigênio em menor tempo se comparado a localidades mais altas cujo acúmulo de oxigênio é menor. (CHAVES *et al*, 2020).

Além disso, a pressão atmosférica está relacionada com a chuva (precipitação): em locais com baixa pressão atmosférica, sendo assim, calor, o ar aquecido tende a subir e, ao atingir grandes altitudes, condensa-se, forma nuvens, e precipita-se. Segundo Miranda e Souza (2013), no sistema atmosférico de alta pressão (anticiclones) e baixa pressão atmosférica, o ar circula sempre das regiões de alta pressão para as de baixa pressão, buscando alcançar o equilíbrio.

Na próxima tabela (7), tem-se os dados da pressão atmosférica na unidade de mediada hectopascal (hPa) e determinada ao nível barômetro.

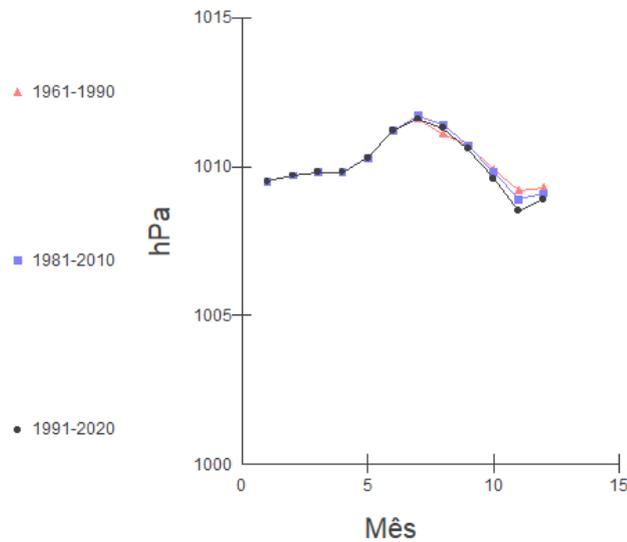
Inferese da Tabela 7 que, nas três séries, a distribuição dos dados referentes à pressão atmosférica, concentra os maiores valores nos meses de junho a agosto. Enquanto que os menores valores foram medidos no primeiro e último trimestre do ano de cada período estudado.

Tabela 7: pressão atmosférica mensal média ao nível do barômetro, em Belém-PA, no período 1961-2020 (hPa).

MÊS	1961-1990	1981-2010	1991-2020
JAN	1009.5	1009.5	1009.5
FEV	1009.7	1009.7	1009.7
MAR	1009.8	1009.8	1009.8
ABR	1009.8	1009.8	1009.8
MAI	1010.3	1010.3	1010.3
JUN	1011.2	1011.2	1011.2
JUL	1011.6	1011.7	1011.6
AGO	1011.1	1011.4	1011.3
SET	1010.7	1010.7	1010.6
OUT	1009.9	1009.8	1009.6
NOV	1009.2	1008.9	1008.5
DEZ	1009.3	1009.1	1008.9

Fonte: INMET

Figura 18: gráfico da pressão atmosférica mensal média ao nível do barômetro, em Belém-PA, no período 1961-2020 (hPa).



Fonte: autor (2023)

As variações de crescimento médio das três curvas mantiveram-se iguais nos três períodos, de janeiro a junho, sendo que de setembro a dezembro elas são divergentes no tempo, taxas de decréscimo e crescimento diferentes. A pressão atmosférica, de setembro a dezembro tem sido a menor dos últimos 60 anos, bem como nos meses de julho a agosto a pressão atmosférica foi menor conforme mostra o gráfico da Figura 18, das séries históricas. A pressão atmosférica mensal média mínima se deu no mês de novembro.

No próximo relatório descritivo, Tabela 8, tem-se um coeficiente de variação menor que 0,1% nas três séries consideradas na pesquisa, declarando pouca dispersão em torno da média, o que significa que tais dados são homogêneos. A assimetria da distribuição nos três conjuntos é positiva.

Tabela 8: estatística descritiva da pressão atmosférica mensal média em Belém-PA, no período 1961-2020 (hPa).

Período	1961-1990	1981-2010	1991-2020
Amostra (N)	12	12	12
Mínimo	1009.2	1008.9	1008.5
Máximo	1011.6	1011.7	1011.6
Amplitude Total	2.4	2.8	3.1
Mediana	1009.85	1009.8	1009.8
Primeiro Quartil (25%)	1009.65	1009.65	1009.575
Terceiro Quartil (75%)	1010.8	1010.825	1010.75
Desvio Interquartilico	1.15	1.175	1.175
Média Aritmética	1010.18	1010.16	1010.07

Variância	0.63	0.82	0.92
Desvio Padrão	0.8	0.91	0.96
Erro Padrão	0.23	0.26	0.28
Coefficiente de Variação	0.08%	0.09%	0.09%
Assimetria (g1)	0.6	0.46	0.16
curtose (g2)	-0.97	-0.93	-0.76
Média Harmônica	1010.1744	1010.1576	1010.0658
Média Geométrica	1010.1747	1010.158	1010.0662
Variância (geom.)	1	1	1
Desvio Padrão (geom.)	1.0008	1.0009	1.001

Fonte: autor (2023).

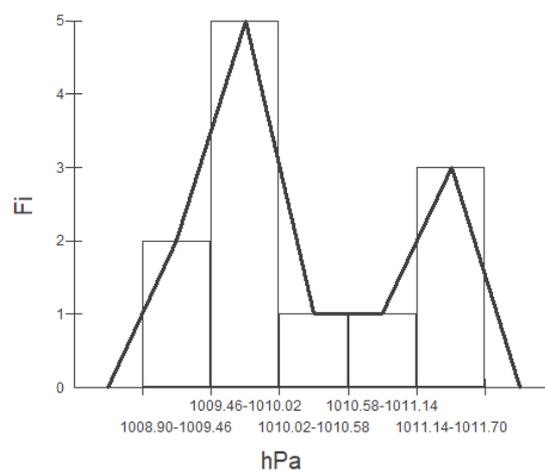
Na tabela 9, tem-se a informação de que a média da pressão atmosférica que ocorreu com maior frequência ficou no intervalo 1009,46 – 1010,02, para os três períodos analisados. E com ¼ das ocorrências ficaram os valores médios da pressão acima de 1011 hPa.

Tabela 9: distribuição de frequência da pressão atmosférica mensal média em Belém-PA, para o período 1961-2020, obtida a partir do Quadro 5.

Classes	Xi	Fi	Percentual
1008.90 — 1009.46	1009.18	2	16.67 %
1009.46 — 1010.02	1009.74	5	41.67 %
1010.02 — 1010.58	1010.30	1	8.33 %
1010.58 — 1011.14	1010.86	1	8.33 %
1011.14 — 1011.70	1011.42	3	25.00 %
TOTAL		12	100.00 %

Fonte: autor (2023).

Figura 19: histograma da pressão atmosférica mensal média em Belém-PA, para o período 1961-2020, obtida a partir da Tabela 9.



Fonte: autor (2023).

4.2.3 Temperatura mínima

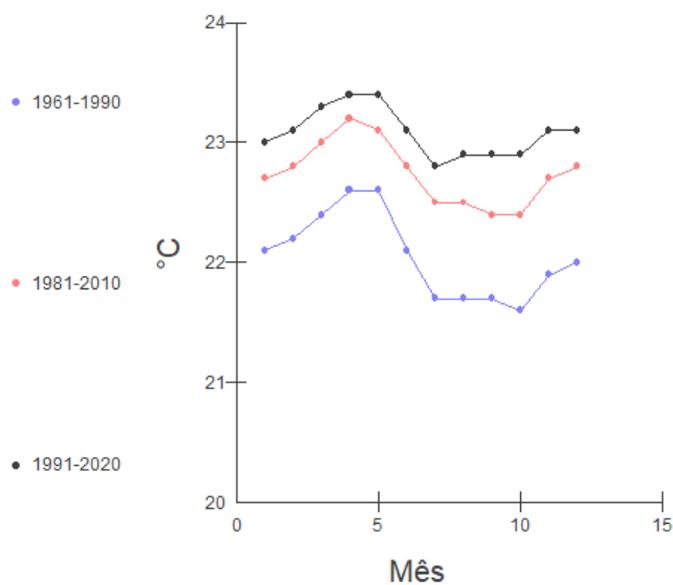
A variável temperatura mínima refere-se à menor temperatura registrada em um local específico durante um determinado período de tempo, geralmente ocorrendo geralmente ocorrendo quando as condições atmosféricas tendem a ser mais frias devido à ausência de radiação solar direta.

Tabela 10: temperatura mínima mensal média em Belém-PA, no período 1961-2020 (°C).

MÊS	1961-1990	1981-2010	1991-2020
JAN	22.1	22.70	23.0
FEV	22.2	22.80	23.1
MAR	22.4	23.00	23.3
ABR	22.6	23.20	23.4
MAI	22.6	23.10	23.4
JUN	22.1	22.80	23.1
JUL	21.7	22.50	22.8
AGO	21.7	22.50	22.9
SET	21.7	22.40	22.9
OUT	21.6	22.40	22.9
NOV	21.9	22.70	23.1
DEZ	22.0	22.80	23.1

Fonte: INMET

Figura 20: gráfico da temperatura mínima mensal média em Belém-PA, no período 1961-2020 (°C).



Fonte: autor (2023).

A evidência mostrada no gráfico da Figura 20 é de temperaturas mínimas cada vez maiores quando se compara as séries históricas no sentido crescente. O primeiro período não apresentou nenhuma mínima de 23 graus, todavia, o segundo e terceiro períodos já possuem esse valor em seus conjuntos. Outra pertinente observação é que há uma tendência de mínimas menores no segundo semestre das três distribuições.

Tabela 11: estatística descritiva da temperatura mínima mensal média em Belém-PA, no período 1961-2020 (°C).

Período	1961-1990	1981-2010	1991-2020
Amostra (N)	12	12	12
Mínimo	21.6	22.4	22.8
Máximo	22.6	23.2	23.4
Amplitude total	1	0.8	0.6
Mediana	22.05	22.75	23.1
Primeiro Quartil (25%)	21.7	22.5	22.9
Terceiro Quartil (75%)	22.25	22.85	23.15
Desvio Interquartílico	0.55	0.35	0.25
Média Aritmética	22.05	22.74	23.08
Variância	0.12	0.07	0.04
Desvio Padrão	0.35	0.26	0.2
Erro Padrão	0.1	0.08	0.06
Coefficiente de Variação	1.59%	1.16%	0.86%
Assimetria (g1)	0.4	0.31	0.44
Curtose (g2)	-1.05	-0.84	-0.81
Média Harmônica	22.0449	22.7389	23.0818
Média Geométrica	22.0475	22.7403	23.0825
Variância (geom.)	1.0001	1.0001	1
Desvio Padrão (geom.)	1.016	1.0117	1.0087

Fonte: autor (2023).

A pesar da primeira série de dados ter o menor conjunto de temperaturas mínimas, foi também a distribuição de dados que mais se afastou da média. Essa variabilidade no primeiro período é detectada não só pela maior variância, desvio padrão ou pelo coeficiente de variação, mas pode ser sentida no maior desvio interquartílico que aponta menos coesão dos valores nesse intervalo se comparado aos outros períodos.

Tabela 12: distribuição de frequência da temperatura mínima mensal média em Belém-PA, para o período 1961-2020, obtida a partir do Quadro 5.

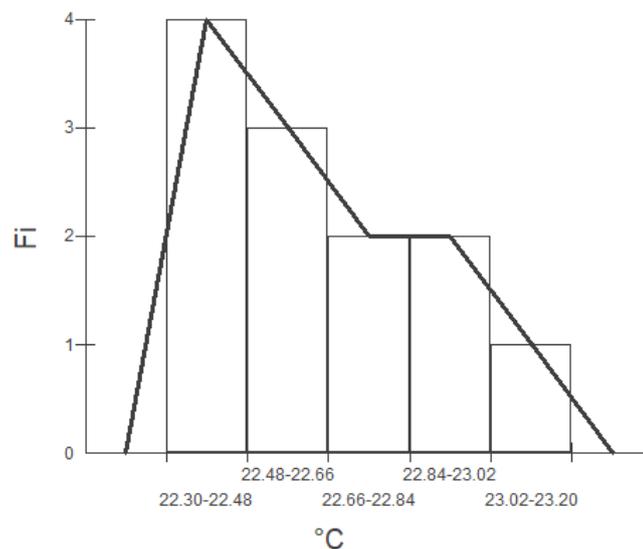
Classes	Xi	Fi	Percentual
22.30 — 22.48	22.39	4	33.33 %
22.48 — 22.66	22.57	3	25.00 %

22.66 — 22.84	22.75	2	16.67 %
22.84 — 23.02	22.93	2	16.67 %
23.02 — 23.20	23.11	1	8.33 %
TOTAL		12	100.00 %

Fonte: Autor (2023).

A maior temperatura mínima de 23 graus, foi a que se apresentou com menos frequência. Enquanto que a média mensal da temperatura mínima mais frequente foi de 22,39 °C, com um terço das ocorrências. No Histograma da Figura 21, a assimetria da distribuição é descrita pelo polígono de frequência e justificada pelos três coeficientes de assimetria positivos.

Figura 21: histograma da temperatura mínima mensal média em Belém-PA, para o período 1961-2020, obtida a partir da Tabela 12.



Fonte: autor (2023).

4.2.4 Temperatura Máxima

A temperatura do ar à superfície segue um ciclo diário, com oscilações entre um máximo (temperatura máxima do dia) e um mínimo (temperatura mínima do dia). Em condições normais, o valor máximo ocorre aproximadamente duas horas após o ápice do Sol, enquanto o valor mínimo acontece pouco antes do nascer do Sol. No entanto, a presença de fenômenos atmosféricos capazes de perturbar o padrão predominante pode alterar as expectativas em relação aos horários esperados para as temperaturas extremas, como por exemplo, a invasão de uma massa de ar frio (VAREJÃO-SILVA, 2006).

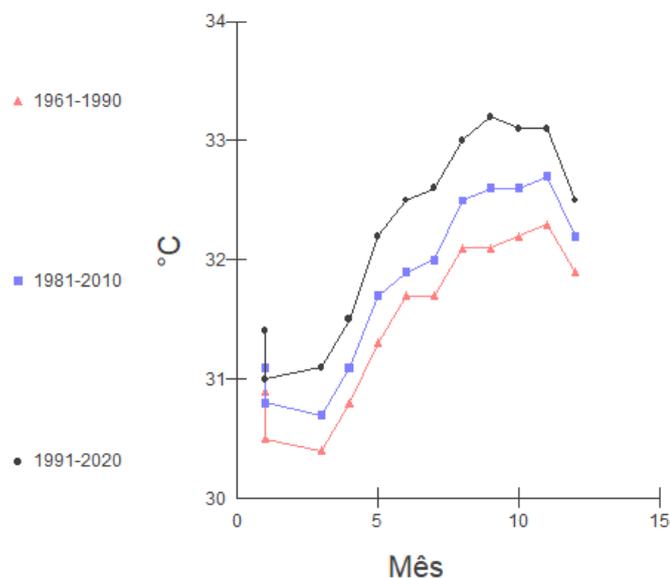
Tabela 13: temperatura máxima mensal média em Belém-PA, no período 1961-2020 (°C).

MÊS	1961-1990	1981-2010	1991-2020
JAN	30.9	31.1	31.4
FEV	30.5	30.8	31.0
MAR	30.4	30.7	31.1
ABR	30.8	31.1	31.5
MAI	31.3	31.7	32.2
JUN	31.7	31.9	32.5
JUL	31.7	32.0	32.6
AGO	32.1	32.5	33.0
SET	32.1	32.6	33.2
OUT	32.2	32.6	33.1
NOV	32.3	32.7	33.1
DEZ	31.9	32.2	32.5

Fonte: INMET

Note, na Tabela 13, que a temperatura máxima mensal média de 33°C só ocorre no último período. Igualmente neste conjunto, não há eventos de média mensal de 30°C no primeiro trimestre do ano, diferentemente do que se observa nos dois primeiros conjuntos de dados.

Figura 22: gráfico da temperatura máxima mensal média em Belém-PA, no período 1961-2020 (°C).



Fonte: autor (2023).

A observação aqui é que de fato estamos tendo temperaturas médias máximas cada vez mais altas com o passar das décadas. Trazendo à luz a reivindicação de ambientalistas que

defendem a existência de um aquecimento global causado pela ação humana no processo do efeito estufa.

Entretanto, MOLION²⁰, considera o efeito estufa benéfico para o planeta, pois sem ele, a temperatura da superfície da terra seria 18° C abaixo de zero, o que tornaria impossível gerar as condições para a vida que aqui se conhece.

Porém, há de se ponderar, segundo Silva (2009), que o aquecimento global é um fenômeno provocado por fatores internos e externos. Os fatores internos são complexos e estão relacionados a sistemas climáticos não lineares, pois as variáveis como a atividade solar, composição físico-química da atmosfera, o tectonismo e o vulcanismo são inconstantes e caóticos. Fatores externos são ações humanas e relacionados a emissões de gases que causam o efeito estufa através da queima de combustíveis fósseis, carvão e derivados de petróleo, indústrias, refinarias, motores, queimadas etc.

As previsões acerca do aquecimento gradual do planeta são realizadas por meio de modelos numéricos que incorporam diferentes hipóteses e aproximações, mas essas previsões ainda estão distantes de fornecer resultados precisos. Os resultados das simulações podem variar significativamente de acordo com as hipóteses ou aproximações utilizadas no modelo, o que leva a resultados que apenas indicam tendências. Esse cenário tem contribuído para intensificar a controvérsia científica em torno dos possíveis efeitos do aquecimento previsto. Contudo, todos os modelos concordam em uma questão: a tendência de aquecimento. (VAREJÃO-SILVA, 2006).

Tabela 14: estatística descritiva da temperatura máxima mensal média em Belém-PA, no período 1961-2020 (°C).

Período	1961-1990	1981-2010	1991-2020
Amostra (N)	12	12	12
Mínimo	30.4	30.7	31
Máximo	32.3	32.7	33.2
Amplitude Total	1.9	2	2.2
Mediana	31.7	31.95	32.5
Primeiro Quartil (25%)	30.875	31.1	31.475
Terceiro Quartil (75%)	32.1	32.525	33.025
Desvio Interquartilico	1.225	1.425	1.55
Média Aritmética	31.49	31.83	32.27
Variância	0.47	0.54	0.67
Desvio Padrão	0.69	0.74	0.82
Erro Padrão	0.2	0.21	0.24

²⁰ https://icat.ufal.br/laboratorio/clima/data/uploads/pdf/molion_desmist.pdf
 AQUECIMENTO GLOBAL. Acesso em: 02 jan. 2023.

Coefficiente de Variação	2.18%	2.32%	2.53%
Assimetria (g1)	-0.47	-0.35	-0.46
Curtose (g2)	-1.4	-1.48	-1.42
Média Harmônica	31.4778	31.8092	32.2476
Média Geométrica	31.4848	31.8171	32.2571
Variância (geom.)	1.0002	1.0002	1.0003
Desvio Padrão (geom.)	1.0222	1.0235	1.0258

Fonte: autor (2023).

A pesar da diferença de quase 1°C entre a maior média do terceiro (1991-2020) e a maior média do primeiro período (1961-1990), os coeficientes de variação, nos três conjuntos, não ultrapassaram os 3%, o que indica que os dados são homogêneos para esta variável no período aqui considerado. A maior média no último período também implicou em um maior desvio padrão para o mesmo conjunto. Dessa forma, aumentando a média da temperatura máxima, aumenta os desvios em torno da média, aumentando o erro na estimativa, o que significa clima mais instável e imprevisível.

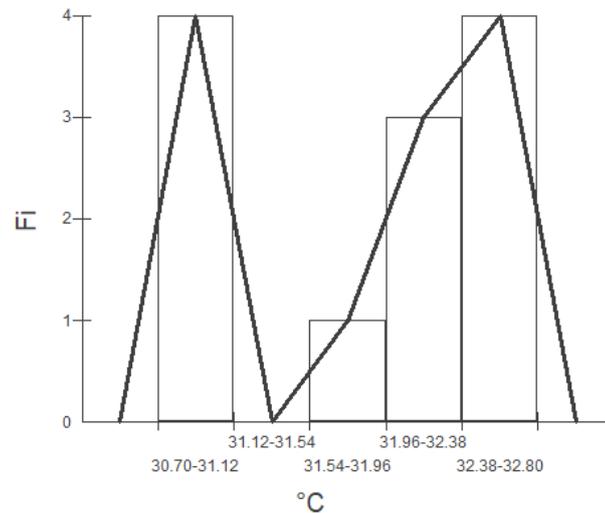
Tabela 15: distribuição de frequência da temperatura máxima mensal média em Belém-PA para o período 1961-2020, obtida partir do quadro 5 (°C).

Classes	Xi	Fi	Percentual
30.70 — 31.12	30.91	4	33.33 %
31.12 — 31.54	31.33	0	0.00 %
31.54 — 31.96	31.75	1	8.33 %
31.96 — 32.38	32.17	3	25.00 %
32.38 — 32.80	32.59	4	33.33 %
TOTAL		12	100.00 %

Fonte: autor (2023).

O histograma da Figura 23 exhibe, geometricamente, a observação de um conjunto de dados bimodal. Apresentando duas classes de temperatura média máxima ocorrendo com a mesma frequência. Em Belém, houve nesse período pesquisado, sempre temperaturas máximas média acima dos 30 graus celsius.

Figura 23: histograma da temperatura máxima mensal média em Belém-PA, obtida a partir do Tabela 15.



Fonte: autor (2023).

4.2.5 Precipitação Acumulada

Entre as várias variáveis climáticas, a precipitação pluviométrica assume um papel de destaque, uma vez que oferece suporte crucial para inúmeros setores da sociedade e é de vital importância para a preservação do equilíbrio do meio ambiente. A precipitação pluviométrica é reconhecida como a principal fonte de abastecimento hídrico para o nosso planeta e, portanto, desempenha um papel fundamental em diversos processos de tomada de decisão, seja de forma direta ou indireta, devido à sua dependência em relação ao espaço e ao tempo. (MIRANDA, 2013).

Conforme definido, a precipitação é o termo utilizado para descrever qualquer forma de água, seja ela líquida ou sólida, que cai das nuvens e alcança a superfície terrestre. No entanto, para os propósitos deste estudo em particular, se dará o enfoque apenas na precipitação na forma líquida, ou seja, a chuva, excluindo a consideração de neve, granizo e outras formas de precipitação sólida. (MOL, 2005).

Conforme dados fornecidos pelo INMET (Instituto Nacional de Meteorologia), a cidade de Belém, banhada pelo rio Guamá e pela baía do Guajará, é reconhecida como a capital brasileira com maior volume de chuvas. Seu índice pluviométrico anual atinge a marca de 2921,9 mm/ano (MORAES; FILHO, 2021).

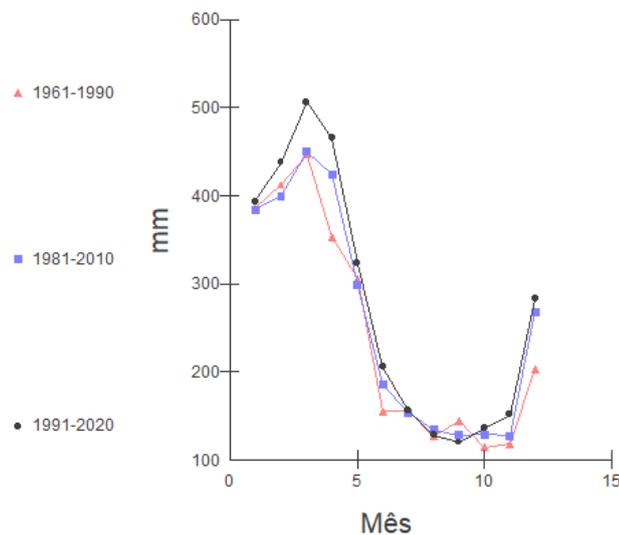
Tabela 16: precipitação acumulada mensal média em Belém-PA, no período 1961-2020 (mm).

MÊS	1961-1990	1981-2010	1991-2020
JAN	385.5	384.5	393.8
FEV	412.5	399.5	437.8
MAR	447.1	450.3	506.3
ABR	353.4	424.3	465.5
MAI	305.5	298.4	323.6
JUN	155.3	185.3	205.8
JUL	155.5	153.8	156.0
AGO	126.4	134.8	128.7
SET	144.8	128.2	120.1
OUT	114.6	129.2	135.8
NOV	118.2	127.4	151.4
DEZ	203.0	268.3	283.5

Fonte: INMET

Segundo os dados da Tabela 16, de janeiro para fevereiro, o último período foi o que apresentou o maior aumento percentual de precipitação média acumulada, com 11% de acréscimo. Convém ainda observar nesta tabela que no último período foi registrado, no mês de março, um acumulado de chuva que ultrapassou os 500 mm; em contrapartida, nos dois primeiros períodos, o maior acúmulo médio de precipitação ficou em torno de 450 mm, também no mês de março.

Figura 24: gráfico da precipitação acumulada mensal média em Belém-PA, no período 1961-2020 (mm).



Fonte: autor (2023).

O gráfico da Figura 24 também exhibe que os três conjuntos descreveram contornos parecidos, com os acúmulos máximos de chuva no fim do primeiro trimestre do ano e os volumes mínimos ficaram em meados do segundo semestre para as três distribuições.

Tabela 17: estatística descritiva da precipitação acumulada mensal média em Belém-PA, no período 1961-2020 (mm).

Período	1961-1990	1981-2010	1991-2020
Amostra (N)	12	12	12
Mínimo	114.6	127.4	120.1
Máximo	447.1	450.3	506.3
Amplitude total	332.5	322.9	386.2
Mediana	179.25	226.8	244.65
Primeiro Quartil (25%)	140.2	133.4	147.5
Terceiro Quartil (75%)	361.425	388.25	404.8
Desvio Interquartilico	221.225	254.85	257.3
Média Aritmética	243.48	257	275.69
Variância	16275.71	16745.52	21076.06
Desvio Padrão	127.58	129.4	145.18
Erro Padrão	36.83	37.36	41.91
Coefficiente de Variação	52.40%	50.35%	52.66%
Assimetria (g1)	0.5	0.37	0.41
Curtose (g2)	-1.64	-1.76	-1.58
Média Harmônica	190.0005	201.7946	210.9002
Média Geométrica	214.2005	227.3286	240.8129
Variância (geom.)	1.1292	1.1259	1.1406
Desvio Padrão (geom.)	1.6972	1.6861	1.734

Fonte: autor (2023).

Note que, com mais de 50%, o alto coeficiente de variação para os três períodos, demonstra que se tem dados não homogêneos, isto é, muitos valores do conjunto se afastaram da média, o que se traduz em meses de pouca chuva e meses mais chuvosos. O último período chama a atenção não só por conter a ocorrência do valor que ultrapassa os 500 milímetros de precipitação acumulada, mas também porque seu menor valor médio mensal não é maior que o mínimo do período intermediário. Outra consideração que se faz pertinente é que no mínimo, de 1991-2020, teve-se em média 5mm de chuva por dia em Belém-PA. O que pode explicar a ocorrência de pelo menos uma chuva fraca por dia na capital do Pará.

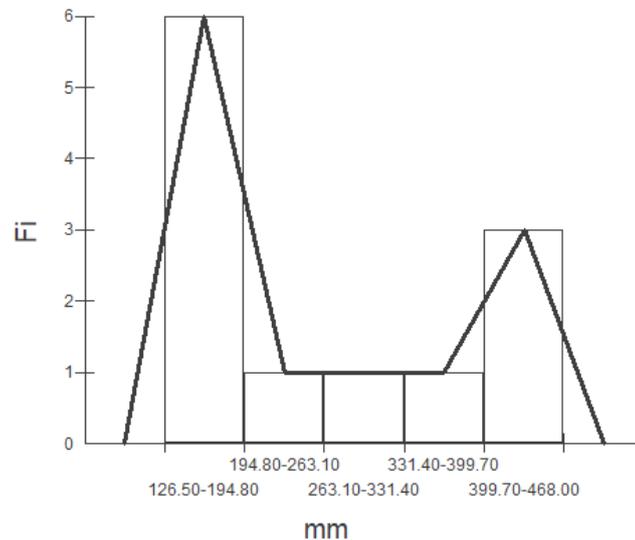
Tabela 18: distribuição de frequência da precipitação acumulada mensal média em Belém-PA, para o período 1961-2020, obtida a parti do quadro 5.

Classes	Xi	Fi	Percentual
126.50 — 194.80	160.65	6	50.00 %
194.80 — 263.10	228.95	1	8.33 %
263.10 — 331.40	297.25	1	8.33 %

331.40 — 399.70	365.55	1	8.33 %
399.70 — 468.00	433.85	3	25.00 %
TOTAL		12	100.00 %

Fonte: autor (2023).

Figura 25: histograma da precipitação acumulada mensal média em Belém-PA, para o período 1961-2020, obtido a partir do Tabela 18.



Fonte: autor (2023).

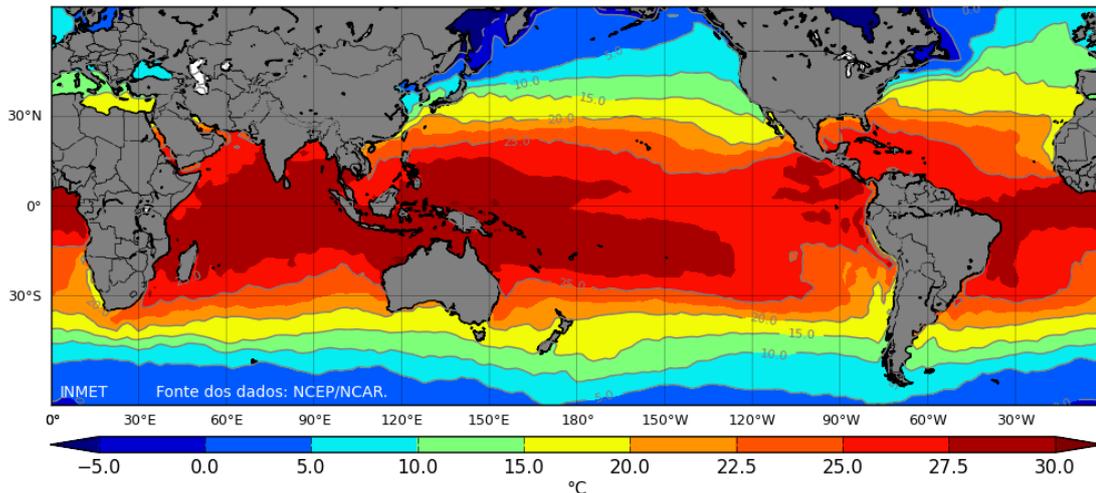
Verifica-se na figura (26) amostral seguinte, em um período de 15 dias do mês de março de 2023, a presença de águas oceânicas com a temperatura da superfície aos 30 °C banhando a costa norte e nordeste brasileiras e por conseguinte, chegando também até Belém e influenciando historicamente o regime de chuvas nesta época para este lugar, corroborando os resultados de Kemenys e Ferreira (2019) que concluiu que os desequilíbrios da temperatura da superfície do mar Atlântico aumentam a precipitação pluvial em algumas regiões do litoral leste da região nordeste, cuja interferência foi mais significativa que os desequilíbrios do Pacífico; ou seja, correlações significativas apontaram que o aquecimento superficial das águas do Atlântico Norte esteve associado ao aumento do volume de água de reservatórios do interior do Nordeste do Brasil.

Segundo o Boletim²¹ de Análise e Previsão Climática (2019): “Essa configuração de anomalias positivas na porção equatorial certamente favoreceu o desenvolvimento dos sistemas

²¹ https://www.semas.pa.gov.br/wp-content/uploads/2012/09/Boletim_RPCH_AnoXII_FEV_2019.pdf Acesso em: 08/01/2023.

produtores de chuva, o que contribuiu para as anomalias positivas de precipitação no litoral norte da América do Sul”.

Figura 26: temperatura da superfície do mar - 01-15/03/2023.



Fonte: <https://clima.inmet.gov.br/GraficosClimatologicos/DF/83377>. Acesso em: 20 abr. 2023.

4.2.6 Insolação

A insolação medida pelo INMET corresponde à quantidade de energia radiante proveniente do sol que é registrada em um local específico ao longo de um período determinado. Geralmente, essa medição é expressa em horas e representa a duração total da exposição solar durante esse intervalo de tempo. O instrumento destinado a medir a insolação é denominado de heliógrafo.

Varejão-Silva (2016) explica que é fundamental não confundir fotoperíodo com a insolação. O fotoperíodo refere-se ao número de horas em um dia em que o disco solar é visível para um observador localizado na superfície terrestre, em um local com horizonte desobstruído. A insolação refere-se ao período completo de tempo, desde o nascer até o ocaso, em que o disco solar permaneceu visível sem ser obscurecido por nuvens ou quaisquer fenômenos atmosféricos. A insolação é sempre menor ou, no máximo, igual ao fotoperíodo, que representa a duração máxima teoricamente possível de sol visível.

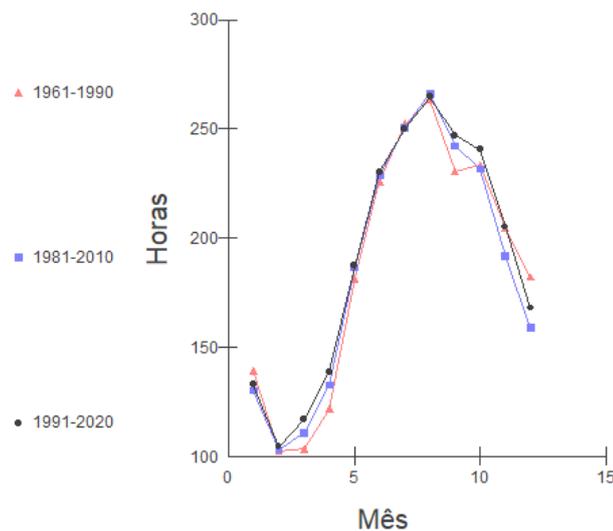
Verifica-se na Tabela 19 que a quantidade mensal média máxima de insolação total em horas, nos três períodos observados, se deu no mês de agosto; assim como também foi coincidente o valor mínimo nos três conjuntos, que ocorreu no mês de fevereiro.

Tabela 19: insolação total mensal média em Belém-PA, no período 1961-2020 (em horas).

MÊS	1961-1990	1981-2010	1991-2020
JAN	139.4	130.3	132.9
FEV	102.5	103.1	104.7
MAR	103.6	111.1	117.2
ABR	121.9	132.8	138.7
MAI	181.6	186.7	187.3
JUN	225.9	228.7	230.3
JUL	252.4	250.8	250.1
AGO	263.5	266.4	264.9
SET	230.5	242.4	247.1
OUT	233.4	231.8	240.6
NOV	204.6	191.7	205.1
DEZ	182.3	159.1	168.1

Fonte: INMET

Figura 27: gráfico da insolação total mensal média em Belém-PA, no período 1961-2020 (em horas).



Fonte: autor (2023).

O gráfico da Figura 27 revela suaves discrepâncias entre uma série e outra, o que indica que esta é uma variável que no mínimo não apresentou anomalias em um intervalo de tempo de 60 anos. Percebe-se, nos três conjuntos de dados, que a partir de fevereiro as médias entram em um intervalo crescente no total mensal de horas de insolação até alcançarem seus pontos de máximo no mês de agosto, e daí começam a decrescer.

Tabela 20: estatística descritiva da insolação total mensal média em Belém-PA, no período 1961-2020 (em horas).

Período	1961-1990	1981-2010	1991-2020
Amostra (N)	12	12	12
Mínimo	102.5	103.1	104.7
Máximo	263.5	266.4	264.9
Amplitude total	161	163.3	160.2
Mediana	193.45	189.2	196.2
Primeiro Quartil (25%)	135.025	132.175	137.25
terceiro quartil (75%)	231.225	234.45	242.225
desvio Interquartílico	96.2	102.275	104.975
média Aritmética	186.8	186.24	190.58
Variância	3331.63	3357.83	3258.7
Desvio Padrão	57.72	57.95	57.09
Erro Padrão	16.66	16.73	16.48
Coefficiente de Variação	30.90%	31.11%	29.95%
Assimetria (g1)	-0.32	-0.12	-0.22
Curtose (g2)	-1.41	-1.6	-1.6
Média Harmônica	167.898	168.2652	173.1599
Média Geométrica	177.6477	177.362	182.0471
Variância (geom.)	1.0524	1.0498	1.047
Desvio Padrão (geom.)	1.4091	1.3974	1.3841

Fonte: Autor (2023).

O alto coeficiente de variação nos três períodos mostra o quanto a nebulosidade tem influenciado para que os totais médios de horas de insolação se afastem da média, implicando em menor homogeneidade dos dados desta variável no período analisado.

Tabela 21: distribuição de frequência da insolação total mensal média em Belém-PA, para o período 1961-2020, obtida a partir do Quadro 5.

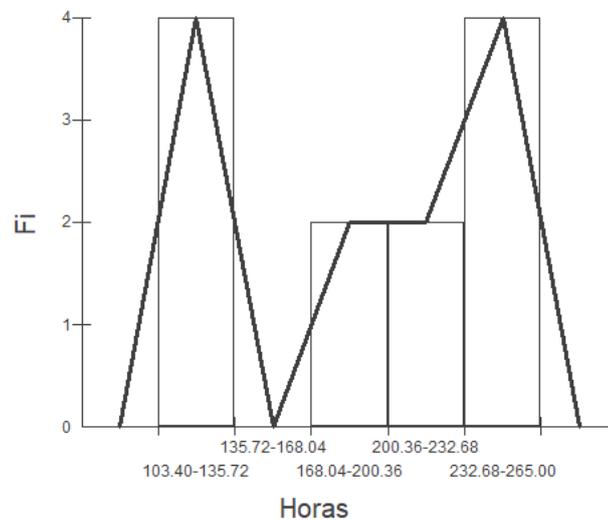
Classes	Xi	Fi	Percentual
103.40 — 135.72	119.56	4	33.33 %
135.72 — 168.04	151.88	0	0.00 %
168.04 — 200.36	184.20	2	16.67 %
200.36 — 232.68	216.52	2	16.67 %
232.68 — 265.00	248.84	4	33.33 %
TOTAL		12	100.00 %

Fonte: autor (2023).

A distribuição apresenta 1/3 terço dos casos com mais de 230 horas totais de média e outro terço dos casos com mais de 100 horas totais de insolação. O terço restante divide-se em duas classes ocorrendo com a mesma frequência, uma com mais de 200 horas e a outra com mais de 230 horas.

O gráfico da Figura 28 destaca uma distribuição bimodal, com duas classes de insolação ocorrendo com a mesma frequência.

Figura 28: histograma da insolação total mensal média em Belém-PA, para o período 1961-2020, obtido a partir da Tabela 21 (em horas).



Fonte: autor (2023).

4.2.7 Intensidade do vento

O vento próximo à superfície terrestre desempenha um papel significativo no clima e no tempo. Sua influência é fundamental para a dinâmica da vida no planeta, pois possibilita importantes interações, como a transferência de calor, dióxido de carbono (CO₂) e vapor d'água entre a atmosfera e a vegetação (OMETTO, 1981). A sua formação é resultado de uma combinação de fatores, incluindo diferenças na pressão atmosférica, variações de temperatura e instabilidade na atmosfera (LINACRE, 1997).

Considerando que o ar se desloca horizontalmente das regiões de alta pressão para as de baixa pressão, a velocidade do vento é influenciada pelo gradiente de pressão entre essas áreas. Quanto mais acentuado for o gradiente de pressão, maior será a velocidade do vento (MENDONÇA; DANI-OLIVEIRA, 2007).

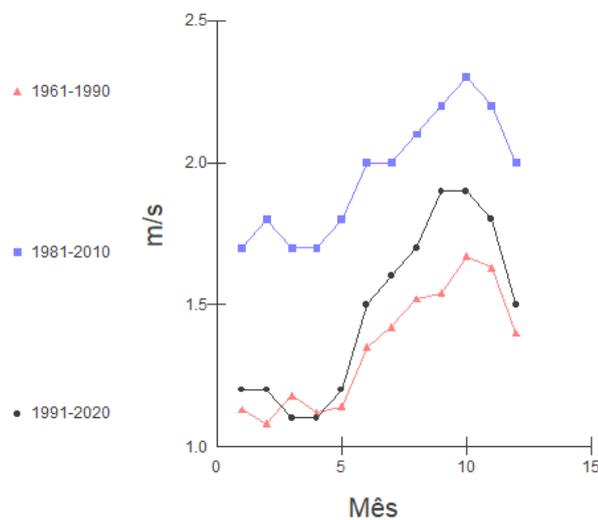
Tabela 22: intensidade do vento mensal média em Belém-PA, no período 1961-2020 (m/s).

MÊS	1961-1990	1981-2010	1991-2020
JAN	1.13	1.7	1.2
FEV	1.08	1.8	1.2
MAR	1.18	1.7	1.1
ABR	1.12	1.7	1.1
MAI	1.14	1.8	1.2
JUN	1.35	2.0	1.5
JUL	1.42	2.0	1.6
AGO	1.52	2.1	1.7
SET	1.54	2.2	1.9
OUT	1.67	2.3	1.9
NOV	1.63	2.2	1.8
DEZ	1.40	2.0	1.5

Fonte: INMET

Enfatiza-se na Figura 29 que o período intermediário (1981-2010) contém os maiores valores médios da intensidade do vento. Sugerindo que as anomalias climáticas em Belém e região, dos últimos anos, e que causaram danos, provavelmente não é o pior cenário por qual este município já passou neste contexto.

Figura 29: gráfico da intensidade do vento mensal média em Belém-PA, no período 1961-2020 (m/s).



Fonte: autor (2023).

Percebe-se que a partir do mês de maio, nos três períodos observados, as intensidades mensais médias do vento aumentaram até atingirem seu ponto máximo em outubro e a partir

daí decresceram. Essa tendência de escalada que se estende por cinco meses, até seu início de declínio, é justificada com os eventos atípicos noticiados neste período do ano pelo portal G1²² PA de notícias: na sexta-feira (30/09/2022) em Belém, ocorreu um vendaval com rajadas de vento que causaram a queda de árvores e o destelhamento de casas, possivelmente atingindo velocidades de até 80 km/h, de acordo com o INMET Pará.

Cinco dias antes deste fato que aconteceu no último dia de setembro de 2022, o mesmo portal²³ de notícia, também informa que no domingo (25/09/2022), em Belém, a velocidade dos ventos atingiu 80 km/h, conforme registrado pelo Instituto Nacional de Meteorologia (INMET). A intensa precipitação pluviométrica e os ventos resultaram em danos em diversos bairros da cidade.

Outra ocorrência meteorológica anômala, na noite de domingo (04/10/2021), foi divulgada neste *site*²⁴, anunciando que em Benevides, município da região metropolitana de Belém, um vendaval provocou o destelhamento de residências e a queda de árvores. Segundo o Instituto Nacional de Meteorologia (Inmet), a chuva que ocorreu no final da tarde foi acompanhada por ventos que alcançaram velocidades de até 60 km/h.

Tabela 23: estatística descritiva da intensidade do vento mensal média em Belém-PA, no período 1961-2020 (m/s).

Período	1961-1990	1981-2010	1991-2020
Amostra (N)	12	12	12
Mínimo	1.08	1.7	1.1
Máximo	1.67	2.3	1.9
Amplitude total	0.59	0.6	0.8
Mediana	1.375	2	1.5
Primeiro Quartil (25%)	1.1375	1.775	1.2
Terceiro Quartil (75%)	1.525	2.125	1.725
Desvio Interquartilico	0.3875	0.35	0.525
Média Aritmética	1.348	1.96	1.48
Variância	0.045	0.05	0.09
Desvio Padrão	0.213	0.22	0.31
Erro padrão	0.062	0.06	0.09
Coefficiente de Variação	15.82%	10.99%	20.87%
Assimetria (g1)	0.144	0.13	0.12
Curtose (g2)	-1.588	-1.42	-1.65
Média Harmônica	1.3176	1.9368	1.4163

²²<https://g1.globo.com/pa/para/noticia/2022/10/01/familias-contabilizam-prejuizos-apos-fortes-ventos-e-chuva-em-belem.ghtml> Acesso em: 09 jul. 2023.

²³<https://g1.globo.com/pa/para/noticia/2022/09/26/ventos-em-belem-chegam-a-80-kmh-aponta-inmet-meia-hora-para-destruir-tudo-diz-trabalhador.ghtml> Acesso em: 09 jul. 2023.

²⁴<https://g1.globo.com/pa/para/noticia/2020/10/05/vendaval-derruba-arvores-e-destelha-casas-em-benevides.ghtml> Acesso em: 09 jul. 2023.

Média Geométrica	1.3329	1.9475	1.4454
Variância (geom.)	1.011	1.0053	1.0195
Desvio Padrão (geom.)	1.172	1.1162	1.235

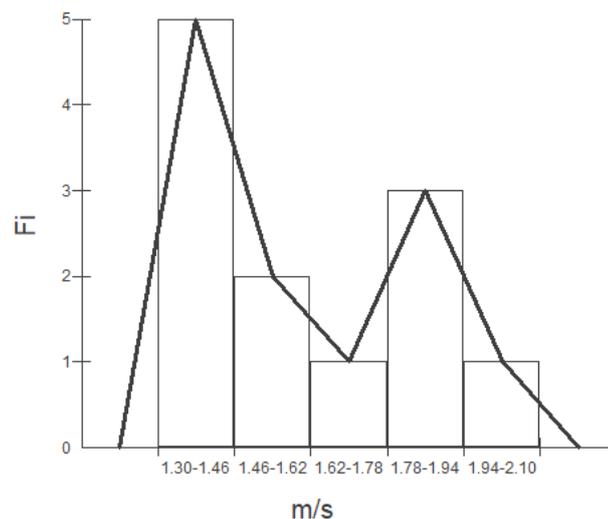
Fonte: autor (2023).

Tabela 24: distribuição de frequência da intensidade do vento mensal média em Belém-PA, para o período 1961-2020, obtida a partir do Quadro 5.

Classes	Xi	Fi	Percentual
1.30 — 1.46	1.38	5	41.67 %
1.46 — 1.62	1.54	2	16.67 %
1.62 — 1.78	1.70	1	8.33 %
1.78 — 1.94	1.86	3	25.00 %
1.94 — 2.10	2.02	1	8.33 %
TOTAL		12	100.00 %

Fonte: autor (2023).

Figura 30: histograma da intensidade média mensal do vento em Belém-PA para o período 1961-2020, obtido a partir do Tabela 24.



Fonte: autor (2023).

A classe da distribuição da intensidade do vento média mensal para o período estudado (1961-2020), com maior frequência, são as intensidades do vento compreendidas no intervalo que vai de 1,3m/s a 1,4 m/s, com quase metade das ocorrências. E 25% dos casos são valores contidos no intervalo de 1.7 m/s e 1.9 m/s. Notavelmente trata-se uma distribuição assimétrica.

4.2.8 Umidade relativa do ar

A umidade relativa do ar é a quantidade de moléculas de água que existe em um dado volume de ar. De acordo com Varejão-Silva (2006), a compreensão da quantidade de vapor de água é fundamental em diversas áreas da atividade humana. Por exemplo, é reconhecido que a umidade ambiente desempenha uma função que condiciona o desenvolvimento de diversos microrganismos patogênicos que afetam as plantas cultivadas. Além disso, a transpiração das plantas está diretamente relacionada ao nível de umidade do ar circundante. Por sua vez, a umidade atmosférica presente em determinado local é um dos parâmetros utilizados para avaliar o nível de conforto ambiental tanto para pessoas quanto para animais. Enfatiza-se ainda que garantir a umidade do ar dentro da faixa ideal é um aspecto constantemente monitorado ao armazenar uma ampla variedade de produtos.

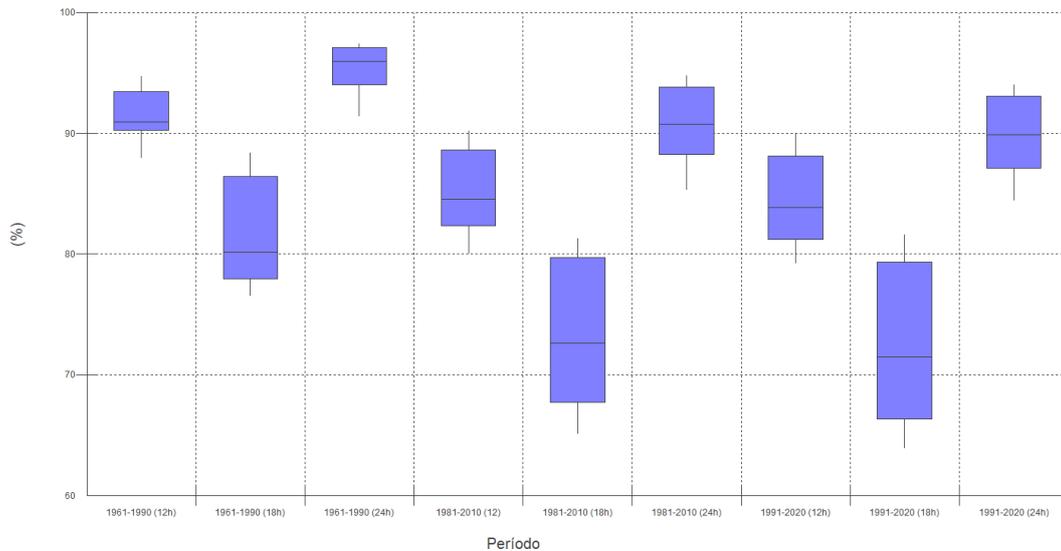
Quadro 7: umidade relativa do ar horária mensal média em Belém-PA, para o período 1961-2020 (%).

Mês	1200			1800			2400		
	1961-1990	1981-2010	1991-2020	1961-1990	1981-2010	1991-2020	1961-1990	1981-2010	1991-2020
JAN.	93.5	88.6	88.4	86	79.4	79.2	96.2	92.4	92.2
FEV.	94.7	90.2	90	87.9	81.0	80.9	97.3	94.2	93.6
MAR.	93.8	89.3	89	88.4	81.3	81.6	97.2	94.3	94.0
ABR.	93.4	88.8	88	87.7	80.7	79.7	97.4	94.8	94.0
MAL.	91.8	86.5	85.9	82.0	74.1	73.5	97.1	93.7	92.9
JUN.	90.8	84.4	83.5	78.3	67.8	66.5	96.2	91.4	90.6
JUL.	91.1	84.5	83.6	76.5	65.3	64.0	95.7	90.1	89.2
AGO.	90.6	82.8	81.6	77.1	65.1	63.9	94.7	88.8	87.6
SET.	89.2	81.0	80.0	77.7	67.6	65.9	93.9	87.4	85.7
OUT.	87.9	80.0	79.2	78.0	69.5	68.5	92.1	85.5	84.7
NOV.	88.4	81.0	79.9	78.4	71.1	69.5	91.4	85.3	84.4
DEZ.	90.8	84.6	84.1	81.9	75.1	74.7	94.1	88.5	87.9

Fonte: INMET.

Para se obter uma apresentação resumida e objetiva dos dados da umidade relativa do ar que se apresenta particionada em três horários, fez-se necessário utilizar o gráfico *box plot* que permite confrontar esses conjuntos.

Figura 31: gráfico *box plot* da umidade relativa do ar horária mensal média em Belém-PA, no período 1961-2020.



Fonte: Autor (2023).

Uma específica organização dos conjuntos de dados, referentes a umidade relativa do ar, favorece o aparecimento visual de distribuições periódicas no gráfico de caixa e que nos induz, inevitavelmente, a comparar tais grupos. Note que no período das 18h nas três séries, tem-se os menores valores da umidade relativa do ar medidos para o dia. Ainda vale enfatizar que nos últimos anos, 50% das observações, às 18h, estão próximas de 80% de umidade relativa.

4.3 AJUSTAMENTO DE CURVAS

O melhor ajuste de curvas é uma técnica utilizada em análise de dados e estatística para encontrar uma função matemática que melhor se ajuste aos pontos observados em um conjunto de dados. O objetivo é encontrar uma curva que represente o padrão geral dos dados de forma mais precisa possível, permitindo a previsão ou a interpolação de valores desconhecidos dentro do intervalo observado.

Para a inspeção da correlação entre pares de variáveis, partiu-se da entrada de dados no BioEstat, conforme mostra a Figura 32, o qual corresponde aos mesmos valores contidos no Quadro 5.

Figura 32: dados de entrada para os testes de regressão.

	- 1 - Nebulosidade	- 2 - Pressão	- 3 - Temp. Min.	- 4 - Temp. Máx.	- 5 - Precipitação	- 6 - Insolação	- 7 - Vento	- 8 - Umidade(1200)	- 9 - Umidade(1800)	- 10 - Umidade(2400)
1	0.7	1009.5	22.6	31.1	387.9	134	1.3	90.2	81.5	93.6
2	0.8	1009.7	22.7	30.8	416.6	103	1.4	91.6	83.3	95.0
3	0.8	1009.8	22.9	30.7	467.9	110	1.3	90.7	83.8	95.2
4	0.8	1009.8	23.1	31.1	414.4	131	1.3	90.1	82.7	95.4
5	0.7	1010.3	23.0	31.7	309.2	185	1.4	88.1	76.5	94.6
6	0.6	1011.2	22.7	32.0	182.1	228	1.6	86.2	70.9	92.7
7	0.6	1011.6	22.3	32.1	155.1	251	1.7	86.4	68.6	91.7
8	0.5	1011.3	22.4	32.5	130.0	264	1.8	85.0	68.7	90.4
9	0.5	1010.7	22.3	32.6	131.0	240	1.9	83.4	70.4	89.0
10	0.6	1009.8	22.3	32.6	126.5	235	2.0	82.4	72.0	87.4
11	0.6	1008.9	22.6	32.7	132.3	200	1.9	83.1	73.0	87.0
12	0.7	1009.1	22.6	32.2	251.6	169	1.6	86.5	77.2	90.2

Fonte: autor (2023).

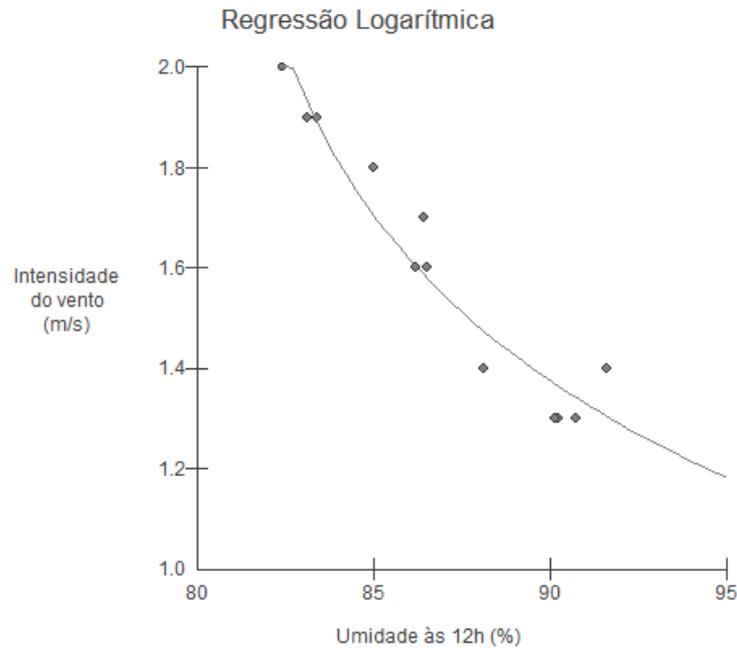
4.3.1 Intensidade do vento versus umidade às 12h

O gráfico da Figura 33 mostra uma tendência na intensidade do vento a partir das alterações da umidade às 12h. Observa-se uma diminuição na intensidade do vento conforme tem-se um aumento da umidade relativa do ar. Porém, esse decrescimento não é linear. Conforme a Tabela 25, o melhor ajuste de curva obtido pelo BioEstat indica a regressão logarítmica com o maior coeficiente de determinação R^2 .

O vento é um importante agente de transporte de umidade na atmosfera. Quando o vento sopra em uma determinada região, ele pode trazer ar úmido de outras áreas e, conseqüentemente, afetar a umidade local. Por exemplo, o vento que se origina em regiões marítimas tende a carregar mais umidade, enquanto o vento que se desloca sobre áreas continentais pode ser mais seco.

Além disso, a intensidade do vento também pode influenciar a taxa de evaporação da água na superfície terrestre. Ventos mais fortes tendem a aumentar a evaporação da água, o que pode levar a uma redução da umidade relativa do ar. Por outro lado, a umidade do ar também pode afetar a intensidade do vento. A variação na umidade pode resultar em mudanças nas forças de pressão atmosférica, o que, por sua vez, pode influenciar a velocidade do vento.

Figura 33: gráfico de regressão logarítmica para as variáveis intensidade do vento versus umidade às 12h, em Belém-PA, para o período 1961-2020.



Fonte: autor (2023).

Tabela 25: teste de regressão para as variáveis intensidade do vento versus umidade às 12h.

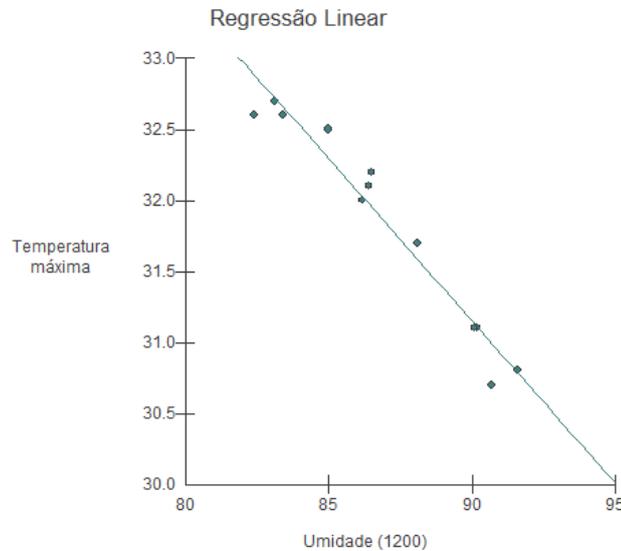
Regressão	Linear	Exponencial	Logarítmica	Geométrica
Tamanho da amostra	12	12	12	12
Intercepto (a)	8.441	113.4486	32.1987	312609179.4
Coef. regressão (b)	-0.0787	-0.0491	-6.853	-4.2783
Coef. determinação (R2)	92.37%	92.08%	92.76%	92.36%
Média (X)	86.975	86.975	86.9221	86.9221
Média (Y)	1.6	1.5808	1.6	1.5808
Var. independente	Coluna 8	Coluna 8	Coluna 8	Coluna 8
Var. dependente	Coluna 7	Coluna 7	Coluna 7	Coluna 7
Equação	$Y' = a + bX$	$Y' = a * e^{(bX)}$	$Y' = a + b * \ln(X)$	$Y' = a * X^b$

Fonte: autor (2023).

4.3.2 Temperatura máxima versus umidade às 12h

Novamente, o teste de regressão, agora para as variáveis umidade às 12h e temperatura máxima, apontou forte associação negativa linear entre essas variáveis. Ou seja, aumenta-se a umidade, diminui a temperatura máxima. O coeficiente de determinação R^2 (Tabela 26) para os quatro modelos de regressão foi de quase 95%, mas ainda assim o melhor ajuste foi a reta.

Figura 34: gráfico de regressão linear para as variáveis temperatura máxima versus umidade às 12h, em Belém-PA, para o período 1961-2020.



Fonte: Autor (2023).

Tabela 26: teste de regressão para as variáveis temperatura máxima versus umidade às 12h.

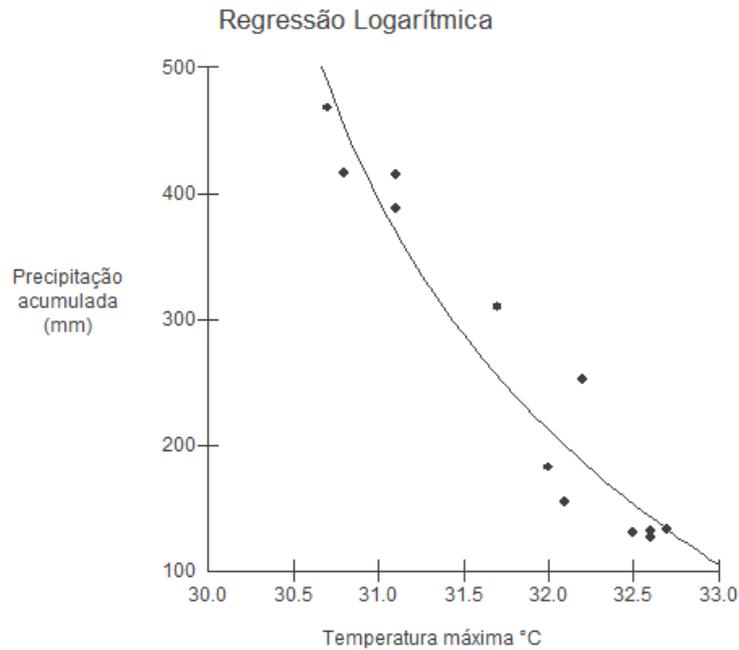
Regressão	Linear	Exponencial	Logarítmica	Geométrica
Tamanho da amostra	12	12	12	12
Intercepto (a)	51.67	59.4596	120.1265	513.902
Coef. regressão (b)	-0.228	-0.0072	-19.7726	-0.623
Coef. determinação (R ²)	94.93%	94.72%	94.46%	94.24%
Média (X)	86.975	86.975	86.9221	86.9221
Média (Y)	31.8417	31.8337	31.8417	31.8337
Var. independente	Coluna 8	Coluna 8	Coluna 8	Coluna 8
Var. dependente	Coluna 4	Coluna 4	Coluna 4	Coluna 4
Equação	$Y' = a + bX$	$Y' = a * e^{(bX)}$	$Y' = a + b * \ln(X)$	$Y' = a * X^b$

Fonte: autor (2023).

4.3.3 Precipitação acumulada versus temperatura máxima

No teste para essas duas variáveis, verificou-se, também, forte associação linear entre elas, porém o ajuste ótimo ficou com o modelo de regressão logarítmica, com um coeficiente de determinação de 94,06% contra 94,02% do modelo linear. Dessa forma, segundo o gráfico de regressão logarítmica (Figura 35), é possível notar que a medida que a média da temperatura máxima aumenta, tem-se um decréscimo na média de precipitação acumulada.

Figure 35: gráfico de regressão logarítmica para as variáveis precipitação acumulada versus temperatura máxima, em Belém-PA, no período 1961-2020.



Fonte: Autor (2023).

Tabela 27: teste de regressão para as variáveis precipitação acumulada versus temperatura máxima em Belém-PA, no período 1961-2020.

Regressão	Linear	Exponencial	Logarítmica	Geométrica
Tamanho da amostra	12	12	12	12
Intercepto (a)	5804.2375	6392.80 E+08	19369.6174	7692.37 E+31
Coef. regressão (b)	-174.1593	-0.6832	-5522.5427	-21.6439
Coef. determinação (R2)	94.02%	91.47%	94.06%	91.34%
Média (X)	31.8417	31.8417	31.8337	31.8337
Média (Y)	258.7167	227.892	258.7167	227.892
Var. independente	Coluna 4	Coluna 4	Coluna 4	Coluna 4
Var. dependente	Coluna 5	Coluna 5	Coluna 5	Coluna 5
Equação	$Y' = a + bX$	$Y' = a * e^{(bX)}$	$Y' = a + b * \ln(X)$	$Y' = a * X^b$

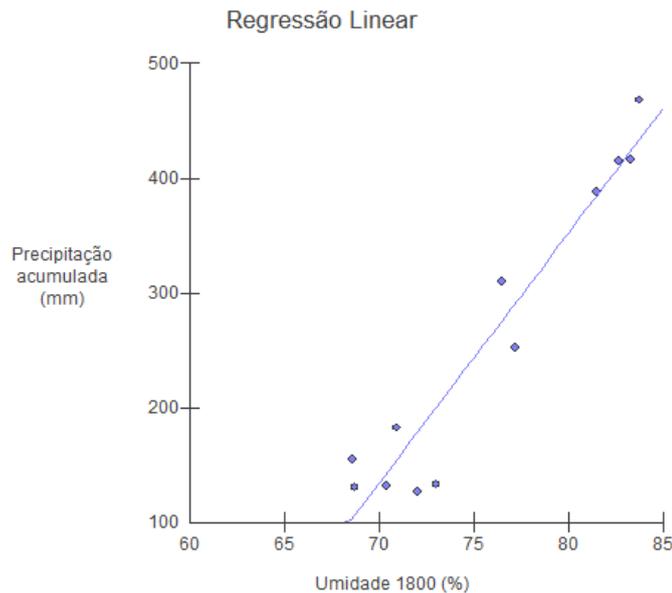
Fonte: autor (2023).

4.3.4 Precipitação acumulada versus umidade às 18h

Outra significativa associação linear, dessa vez positiva, foi certificada entre as variáveis, umidade relativa do ar às 18h e precipitação acumulada. O gráfico da Figura 36 de regressão linear aponta que a medida que aumenta a média de umidade às 18h, também aumenta a média da precipitação acumulada. Os coeficientes de determinação para os modelos de

regressão linear e de regressão logarítmica ultrapassaram os 90%; enquanto que os coeficientes de determinação para os modelos de regressão exponencial e geométrico ficaram um pouco abaixo de 90%. Contudo, o melhor ajuste foi o linear, com o coeficiente R^2 igual a 92,23%, contra 91,44% do coeficiente R^2 para o ajuste logarítmico.

Figure 36: gráfico de regressão linear para as variáveis precipitação acumulada versus umidades às 18h, em Belém-PA, no período 1961-2020.



Fonte: autor (2023).

Tabela 28: teste de regressão para as variáveis precipitação acumulada versus umidade às 18h.

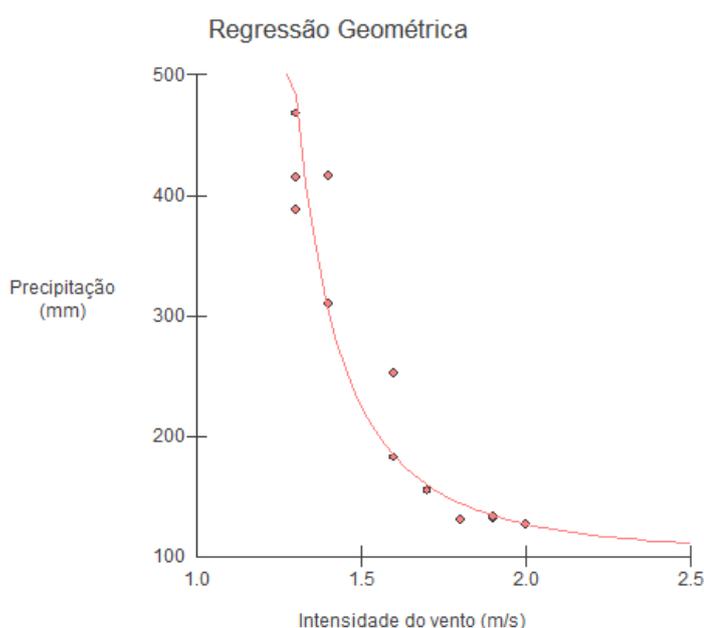
Regressão	Linear	Exponencial	Logarítmica	Geométrica
Tamanho da amostra	12	12	12	12
Intercepto (a)	-1388.2948	0.3611	-6869.1698	0
Coef. regressão (b)	21.7523	0.0852	1648.3513	6.4651
Coef. determinação (R2)	92.23%	89.36%	91.44%	88.93%
Média (X)	75.7167	75.7167	75.509	75.509
Média (Y)	258.7167	227.892	258.7167	227.892
Var. independente	Coluna 9	Coluna 9	Coluna 9	Coluna 9
Var. dependente	Coluna 5	Coluna 5	Coluna 5	Coluna 5
Equação	$Y' = a + bX$	$Y' = a * e^{(bX)}$	$Y' = a + b * \ln(X)$	$Y' = a * X^b$

Fonte: autor (2023).

4.3.5 Precipitação acumulada versus intensidade do vento

O modelo de regressão geométrica, exibido na Figura 37, tem um coeficiente de determinação R^2 altamente significativo, indicando que 93,56% da variabilidade dos dados (Tabela 30) é explicada por esse ajuste. A curva do modelo é o melhor, neste teste realizado pelo BioEstat, encontrado para adequar os dados da variável independente (precipitação acumulada) em relação à variável dependente (intensidade do vento).

Figure 37: gráfico de regressão geométrica para as variáveis precipitação acumulada versus intensidade do vento em Belém-PA, no período 1961-2020.



Fonte: autor (2023).

O modelo sugere ser capaz de capturar o padrão subjacente entre as duas variáveis. Isso aconselha que a intensidade do vento tem uma relação altamente associada com a precipitação acumulada e, portanto, as mudanças na intensidade do vento podem ser usadas para prever mudanças nos níveis de precipitação acumulada.

Tabela 29: teste de regressão para as variáveis precipitação acumulada versus intensidade do vento.

Regressão	Linear	Exponencial	Logarítmica	Geométrica
Tamanho da amostra	12	12	12	12
Intercepto (a)	1026.5437	5260.9927	614.4705	967.678
Coef. regressão (b)	-479.8919	-1.962	-776.8519	-3.1577

Coef. determinação (R2)	87.32%	92.28%	89.57%	93.56%
Média (X)	1.6	1.6	1.5808	1.5808
Média (Y)	258.7167	227.892	258.7167	227.892
Var. independente	Coluna 7	Coluna 7	Coluna 7	Coluna 7
Var. dependente	Coluna 5	Coluna 5	Coluna 5	Coluna 5
Equação	$Y' = a + bX$	$Y' = a * e^{(bX)}$	$Y' = a + b * \ln(X)$	$Y' = a * X^b$

Fonte: autor (2023).

Essa informação pode ter diversas aplicações práticas, especialmente em áreas relacionadas ao clima, meteorologia e estudos ambientais. A capacidade de estabelecer uma relação forte entre a precipitação acumulada e a intensidade do vento pode ser útil para previsão de tempestades, ventos fortes e eventos climáticos extremos, bem como para o planejamento e atenuação nos impactos ambientais.

4.4 SÍNTESE DOS RESULTADOS

As observações gráficas e de tabelas das normais climatológicas de Belém do Pará apontam para um período cíclico de maior precipitação acumulada nos meses de janeiro a abril. A partir de maio começa o decréscimo no acumulado médio de chuva, até atingir o mínimo no mês de outubro. E com o início do último trimestre do ano, o volume acumulado de chuva volta a ter uma tendência crescente.

Verificou-se que os meses com as maiores intensidades média de vento registradas, correspondem aos mesmos períodos de eventos climáticos atípicos nessa região, ainda que tais fatos não tenham ocorrido na mesma coordenada geográfica da estação meteorológica, seus medidores assinalaram velocidade acima da média, consoantes aos vendavais observados em Belém e noticiados em portais de notícias.

A média mensal da temperatura máxima mostrou-se maior a cada período estudado, porém, a análise é inconclusiva, pois não se investigou se essa escalada é cíclica, isto é, se a temperatura em algum período futuro retrocederá para novamente voltar a aumentar. Dessa forma, a variável temperatura máxima tem explicado a elevação da temperatura mínima mensal média, que tem se mostrado cada vez maior quando se compara os três períodos pesquisados.

Quanto a variável nebulosidade, as observações mostraram que o maior valor observado, para a média mensal de céu encoberto por nuvens, foi de 80%, e que se deu nos meses de fevereiro a abril, coincidindo com o período de maior acúmulo de precipitação para a

região, atestando o fato de que pode se ter céu encoberto e não se ter chuva, porém raramente se tem chuva com céu aberto.

A pressão atmosférica foi a única variável que menos se mostrou divergente nos três períodos em análise. Somente a partir do mês de outubro que os dados médios mensais começam a discrepar nas séries históricas.

A insolação total, nas três séries, teve um coeficiente de variação acima de 30%, indicando significativa dispersão dos dados em relação à média, porém, os pontos de máximo e de mínimo, nos três períodos, ocorreram nos mesmos meses. A umidade relativa do ar horária apresentou relativa regularidade nas três séries analisadas. Às 24h tem-se o ponto de máximo dessa variável, enquanto que às 18h, o ponto de mínimo.

Quanto as considerações sobre os modelos de regressão propostos, foram examinados cinco pares de variáveis que tiveram seus coeficientes de determinação R^2 acima de 90%, implicando em expressiva associação entre esses específicos pares apresentados: intensidade do vento e umidade às 12h; temperatura máxima e umidade às 12h; precipitação e temperatura máxima; precipitação e umidade às 18h e, precipitação e intensidade do vento.

Porém, a correlação não implica causalidade, ou seja, o fato de haver uma alta correlação não significa necessariamente que uma variável causa diretamente mudanças na outra. É importante ter cautela ao interpretar a correlação como uma relação causal, pois pode haver outros fatores ou variáveis ocultas influenciando ambos os fenômenos. No entanto vale recomendar que, ao entender a relação entre essas variáveis, os modelos de previsão meteorológica podem se tornar mais precisos e confiáveis.

5 CONCLUSÃO E RECOMENDAÇÃO

Antes que se comente os aspectos que se seguiram quanto a satisfação das expectativas postas inicialmente, referentes ao uso do programa BioEstat como recurso didático para o ensino e aprendizado de estatística, é necessário deixar patente que não se intencionou substituir a regência da docência pela máquina. É importante lembrar que a tecnologia é uma ferramenta que complementa, mas não substitui o papel fundamental dos professores na educação matemática.

A existência e exercício do profissional de ensino será sempre crucial quando se objetiva forjar estudantes críticos na recepção do aprender - entendedores da teoria que explica o funcionamento dos algoritmos e seus resultados técnicos - e capacitá-los a terem habilidade de modelar a natureza com a ciência que se inaugura na escola básica. Diante disso, é indubitável a função do professor no processo educacional como mediador do conhecimento que impulsiona os aprendizes a agir na produção de saberes que beneficiam a inteligência.

Tendo em vista que este estudo não é pioneiro na proposição de ensinar estatística com o apoio tecnológico, vale deixar a consideração da experiência que tiveram Silva *et al* (2014), e que publicaram:

“O software apresentou grande potencial de uso devido à facilidade de aquisição, por meio gratuito e virtual, além de apresentar fácil instalação e possuir um manual autoexplicativo. Estas características podem conferir ao BioEstat uma alta demanda de aquisição e utilização. [...] O emprego do software BioEstat em sala de aula torna-se um instrumento para o ensino de estatística, o qual os alunos além de aprenderem calcular manualmente poderão desenvolver os mesmos no programa. Este novo conhecimento tornará prático e ágil o trabalho com dados, proveniente de pesquisas. [...] O software apresentou grande potencial de uso no ensino de estatística, devido sua facilidade, aplicabilidade e adaptação aos sistemas computacional. [...] Suas características pedagógicas apresentaram conceito ótimo, podendo ser facilmente manuseado por um acadêmico sem a presença de um instrutor, apenas com a utilização de seu manual ou com a memorização de seus processos. [...] Em suas características de *interface* o programa obteve bom desempenho, apresentando uma linguagem simples, clareza nos procedimentos, facilidade de condução e indicando correção do conteúdo, mostrando-se satisfatórias, principalmente, por apresentar nacionalidade brasileira e conformidade com o idioma português. Isto faz com que os acadêmicos consigam desenvolver uma melhor compreensão dos procedimentos e resultados expostos pelo software. [...] O programa BioEstat apresenta uma área de trabalho bem ampla e didática, possuindo uma linguagem de fácil compreensão e com suas ferramentas dispostas de maneira simples, possibilitando sua aprendizagem mais rápida e eficiente. Apresenta seus resultados em editor do próprio programa de forma breve e clara, podendo o usuário imprimir ou salvar os mesmos. Desenvolve testes na área de estatística básica e experimental, sendo uma excelente ferramenta para pesquisadores nas mais diversas áreas do conhecimento, principalmente biológicas, agrárias e médicas. [...] Análise descritiva: a obtenção desta análise pelo software BioEstat se mostrou de grande eficácia uma vez que a mesma foi adquirida com um número mínimo de etapas e seus resultados apresentados de forma clara e didática”. (SILVA *et al*, 2014, p.377-380).

Assim, diante de toda essa básica análise estatística, sustentada pelo *software* empregado na pesquisa, a consumação está consoante com Ayres *et al* (2007), que afirma que o BioEstat gera maior interesse e abre oportunidade para que estudiosos de diversas áreas do conhecimento técnico-científico o utilizem como uma ferramenta de consulta para avaliar seus estudos experimentais e observacionais, contando com o apoio desse aplicativo.

Portanto, é decisivo que se dissemine essa prática, pois o uso estratégico da tecnologia permite ampliar os horizontes da educação, aproximando os estudantes de um mundo de conhecimento e informações atualizadas. Isso facilita o desenvolvimento de habilidades essenciais para o século XXI, como pensamento crítico, resolução de problemas e competências digitais, preparando-os para os desafios do futuro.

Assim, esta pesquisa deixa o impulso para que futuros trabalhos científicos dessa área do conhecimento, com ênfase na educação, explore não somente programas estatísticos, mas também outros *softwares* e ferramentas tecnológicas que auxiliem no processo de ensino-aprendizagem em diversos temas da matemática pura e aplicada.

As perspectivas vindouras do uso de *softwares* no ensino e aprendizagem da matemática são favoráveis, conforme a tecnologia continua a se desenvolver e integrar-se cada vez mais às salas de aula. A combinação entre matemática e informática tem a possibilidade de transformar a forma como os estudantes aprendem e professores ensinam essa disciplina fundamental.

REFERÊNCIAS

ALMEIDA, Fabiana dos Santos; LIMA, Daniela da costa Brito Pereira; RUAS, Kelly Cristina da Silva. **O uso das tecnologias digitais na educação básica**. Vitória – Espírito Santo: DECT, 2021.

ASSIS, Janilson de Pinheiro de; SOUSA, Roberto Pequeno de; DIAS, Carlos Tadeu dos Santos Dias. **Glossário de estatística**. Mossoró: EdUFERSA, 2019. 901p.

AYRES, Manuel; AYRES Jr., Manuel; AYRES, Daniel Lima; SANTOS, Alex de Assis dos Santos. **BioEstat: aplicações estatísticas nas áreas das ciências biológicas e médicas**. Belém; Sociedade Civil Mamirauá: MCT-CNPq, 2007. 364p.

AZEVEDO, Paulo Roberto Medeiros de; **Introdução à estatística** [recurso eletrônico]. – 3. ed. – Natal, RN: EDUFRN, 2016.

BASTOS, Therezinha Xavier; PACHECO, Nilza Araújo; NECHET, Dimitrie; SÁ, Tatiana Deane de Abreu. **Aspectos climáticos de Belém nos últimos cem anos**. Belém: Embrapa Amazônia Oriental, 2002. 31p.

BRASIL. Ministério da Educação. **Base Nacional Comum Curricular**. Brasília, 2018. 600p. Disponível em: http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_site.pdf. Acesso em: 17 jan. 2023. 600p.

BRASIL. Ministério da Educação. **Diretrizes Curriculares Nacionais Gerais da Educação Básica**. Brasília: MEC, SEB, DICELI, 2013. 562p.

BRASIL. Secretaria da Educação Fundamental. **Parâmetros Curriculares Nacionais: Matemática**. Brasília: MEC/SEF, 1998. 148 p.

CAMPOS, Gilda Helena Bernardino. **Qualidade de software: Teoria e Prática**. Campinas, Makron, 2001.

CHEIN, Flávia. **Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas**. Brasília: Enap, 2019. 76p.

COSTA, Fabrício Martins da; **Estatística**. Belém: UEPA – Universidade do Estado do Pará / Centro de Ciências Sociais e Educação, 2011. 76 p.

COSTA, Paulo Roberto da; **Estatística**. – 3. ed. – Santa Maria: Universidade Federal de Santa Catarina, Colégio técnico Industrial de Santa Maria, Curso Técnico em Automação industrial, 2011, 95 p.

COSTA, Sérgio Francisco; **Introdução Ilustrada à Estatística (com muito humor!)**. – 2. ed.– São Paulo: Harbra, 1992. 303p.

CRESPO, Antônio Arnot. **Estatística Fácil**. – 17. Ed. – São Paulo: Saraiva, 2002.

D'AMBROSIO, Ubiratan. **Etnomatemática: arte ou técnica de explicar e conhecer**. São Paulo: Ática, 1990.

DAMASIO, E.; RIBEIRO, C. E. N. Software livre para bibliotecas, sua importância e utilização: o caso GNUTECA. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 4, n. 2, p. 70–86, 2006. DOI: 10.20396/rdbci.v4i1.2036. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2036>. Acesso em: 9 jul. 2023.

Dicionário escolar da língua portuguesa / Academia Brasileira de Letras. – 2. ed. – São Paulo: Companhia Editora Nacional, 2008.

FELBERG, Ana Lúcia Cabral et al. **A INFLUÊNCIA DA PRESSÃO ATMOSFÉRICA NO CORPO HUMANO**. In: Anais da Semana Nacional de Ciência e Tecnologia do Instituto Federal de Rondônia - Campus Cacoal. Anais, Cacoal(RO) IFRO, 2020. Disponível em: <<https://www.even3.com.br/anais/snctcacoal2020/277674-A-INFLUENCIA-DA-PRESSAO-ATMOSFERICA-NO-CORPO-HUMANO>>. Acesso em: 25/06/2023 19:18.

FERREIRA, Valéria; **Estatística Básica**. – 17. Ed. – Rio de Janeiro: SESES, 2015.

FIGUEIRÓ, Ronaldo. **Manual prático de bioestatística computacional**. Centro Universitário de Volta Redonda - UniFOA – Campus Três Poços - 2014. 60 p.

FONSECA, Jairo Simon da. **Curso de estatística** / Jairo Simon da Fonseca, Gilberto de Andrade Martins. – 6. ed. – São Paulo: Atlas, 2011.

HEIJ, Christiaan; DE BOER, Paul; FRANSES, Philip Hans; KLOEK, Teun; VAN DIJK, Herman K. **Econometric Methods with Applications in Business and Economics**. Nova Iorque: Oxford University Press, 2004.

HENZ, Carla Cristina. **O uso das tecnologias no ensino-aprendizagem da matemática**. Universidade Regional Integrada do Alto Uruguai e das Missões – URI – Campus de Erechim – 2008.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Cidades, Censo**, 2010. Disponível em: <https://cidades.ibge.gov.br/>. Acesso: 29 de maio de 2023.

INSTITUTO NACIONAL DE METEOROLOGIA – INMET. **Normais climatológicas do Brasil 1991-2020**. Brasília – DF, 2022.

KEMENES, Alexandre; FERREIRA, Luanny Gabriele Cunha. **Influência das anomalias de temperatura da superfície do mar (Atlântico e Pacífico) sobre o acúmulo de água em reservatórios do interior do Nordeste do Brasil (1998-2015)** / Teresina : Embrapa Meio-Norte, 2019. 39p.

KENNEDY, Peter. **A Guide to Econometrics**. Massachusetts: Blackwell Publishing, 2008.
LINACRE, E.; GEERTS, B. **Climates & Weather Explained**. London and New York: Routledge, 1997.

MENDONÇA, F.; DANI-OLIVEIRA, I. M. **Climatologia: noções básicas e climas do Brasil**. São Paulo – SP: Oficina de textos, 2007.

MIRANDA, Ricardo Augusto Calheiros de; Souza, Lucio de. **Climatologia geográfica**. v. 2. – Rio de Janeiro: Fundação Cecierj, 2013. 240p.

MORAES, Dayse; FILHO, Manoel; **Contribuição Das Chuvas Do Período Da Tarde Em Belém E Possíveis Relações Com A Normal Climatológica**. Revista Brasileira de Climatologia, [S. l.], v. 23, 2021. DOI: 10.5380/abclima.v23i0.58364. Disponível em: <https://ojs.ufgd.edu.br/index.php/rbclima/article/view/14078>. Acesso em: 18 jul. 2023.

MORETTIN, Pedro Alberto; BUSSAB, Wilton de Oliveira. **Estatística Básica**. – 9.ed. – São Paulo: Saraiva, 2017.

OMETTO, J. C. **Bioclimatologia vegetal**. São Paulo: Ceres, 1981.

PEREIRA, Rodrigo da Silva; NUNES, André Becker. **Influência da nebulosidade na atenuação atmosférica na cidade de Pelotas-RS: Estudo de caso**. Revista ciência e natura, Santa Maria: Revista do Centro de Ciências Naturais e Exatas – Edição Esp. – UFSM, 2013.

PERIUS, Ana Amélia Butzen. **A tecnologia aliada ao ensino de matemática**. Universidade Federal do Rio Grande do Sul – Centro Interdisciplinar de Novas Tecnologias na Educação – Cerro Largo, 2012.

REIS, Elizabeth. **Estatística Descritiva**. – 7.ed. – Lisboa: Edições Sílabo, Lda, 2009.

MORAIS, Fabrício Higo Monturil. **Influência da Irradiação Solar na Análise de Viabilidade Econômica de Sistemas Fotovoltaicos**. Revista Brasileira De Meteorologia. Teresina, PI: v. 36, n. 4, 723– 734, 2021. Disponível em: <https://doi.org/10.1590/0102-7786360049>. Acesso em: 22 jul. 2023.

RIBEIRO, Matheus Henrique Dal Molin; PILAR, João Victor do; LIMA, José Donizete de. **Introdução ao estudo de Probabilidade e Estatística com auxílio do software R**. UTFPR – Universidade Tecnológica Federal do Paraná – Campus Pato Branco: Recursos Educacionais Aberto, agosto de 2016.

RODRIGUES, Aurélia Aparecida de Araújo. **ESTATÍSTICA: da educação básica ao ensino superior**. Uberlândia, MG: UFU, 2015.

SILVA, Robson Willians da Costa; PAULA, Beatriz Lima de. 2009. **Causa do aquecimento global: antropogênica versus natural**. Terræ Didática, 5(1):42-49. Disponível em: <http://www.ige.unicamp.br/terraedidatica/>. Acesso em: 02 jan. 2023.

SILVA, Antônio Carlos Silveiro da; VEIGA, Jessica Borges da; VEIGA, Bruna Borges da; CAIONI, Charles; OLIVEIRA, Ademilson Sampaio de. **Avaliação do software BioEstat para o ensino de estatística nos cursos de graduação**. Revista da Universidade Vale do Rio Verde, Três Corações, v.12, n. 2, p.375-385, ago./dez. 2014.

SOUSA NETO, Theófilo Machado de. **Ajuste de Curvas Usando Métodos Numéricos**. 2018. 134f. Dissertação (Mestrado - PROFMAT - Programa de Pós-graduação em Matemática em

Rede Nacional – Sociedade Brasileira de Matemática) – Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Goiânia, 2018.

VAREJÃO-SILVA, Mário Adelmo. **Meteorologia e Climatologia**. Brasília: INMET, Gráfica e Editora Pax, 2001. 532p.

VAREJÃO-SILVA, Mário Adelmo. **Meteorologia e climatologia**. Versão Digital 2, Recife, 2006. Disponível em: file:///C:/Users/miche/Downloads/METEOROLOGIA_E_CLIMATOLOGIA_VD2_Mar_2006.pdf. Acesso em: 09 jul. 2023.