Débora Souza Luz

Estatística no Ensino Médio: Um guia didático com teoria, aplicações e práticas

Vitória

2025

Débora Souza Luz

Estatística no Ensino Médio: Um guia didático com teoria, aplicações e práticas

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

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

PROFMAT

Orientador: Prof. Dr. Fábio Júlio da Silva Valentim

Vitória
2025

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

Luz, Débora Souza, 1993-
L979e Estatística no Ensino Médio: Um guia didático com teoria, aplicações e práticas / Débora Souza Luz. - 2025.
172 f. : il.

Orientador: Fábio Júlio da Silva Valentim.
Dissertação (Mestrado Profissional em Matemática em Rede Nacional) - Universidade Federal do Espírito Santo, Centro de Ciências Exatas.

1. Ensino de Estatística. 2. Novo Ensino Médio. 3. Itinerários Formativos. 4. Estatística. 5. Educação Financeira e Fiscal. 6. Prática Docente. I. Valentim, Fábio Júlio da Silva. II. Universidade Federal do Espírito Santo. Centro de Ciências Exatas. III. Título.

CDU: 51

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Centro de Ciências Exatas

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

“Estatística no Ensino Médio: Um guia didático com teoria, aplicações e práticas”

DÉBORA SOUZA LUZ

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

Aprovado em 28/07/2025 por:

____

Prof. Dr. Fábio Júlio da Silva Valentim
Presidente – Examinador Interno – UFES

____

Prof. Dr. Alancardek Pereira Araujo
Examinador Interno – UFES

____

Prof. Dr. Paulo Roberto Prezotti Filho
Examinador Externo – IFES

Campus Universitário Alaor de Queiroz Araújo – Av. Fernando Ferrari, 514, Goiabeiras, Vitória – ES │ 29075-910 │ (27) 4009-2474 │www.matematica.ufes.br/pt-br/pos-graduacao/PMPM – profmat.ufes@gmail.com

Folha de Assinaturas DEBORA SOUZA LUZ

Data e Hora de Criação: 24/07/2025 às 10:22:13

Documentos que originaram esse envelope:

- Folha de Assinaturas DEBORA SOUZA LUZ.pdf (Arquivo PDF) - 1 página(s)

Hashs únicas referente à esse envelope de documentos

[SHA256]: 444054ab890b3a2ab4c9cffbcb1e75a84c2485fbb88c65255a6ee9b5746e5773
[SHA512]: 9dac0cce40bf68a63d11342318f1b41f7f9a1707b4874e42e7a2d4b090144192acd9b571cd9f4b3fd14735ca19ce16fc968fb3356ab367a0772f8ca794047710

Lista de assinaturas solicitadas e associadas à esse envelope

ASSINADO - Alancardek Pereira Araújo (alancardek.araujo@ufes.br)
Data/Hora: 28/07/2025 - 17:37:19, IP: 172.86.185.6, Geolocalização: [-20.275577, -40.304163]
[SHA256]: 78fdc7fc8dc211cfccf73dbde7566619b2e664fe56792eb6c234d8a0f652ae8f
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

Alancardek Pereira Araujo

ASSINADO - Fábio Júlio da Silva Valentim (fabio.valentim@ufes.br)
Data/Hora: 28/07/2025 - 17:42:14, IP: 164.163.204.221
[SHA256]: 54cd8a5b9628008e90488a1b3918142ed8fdb8712e6dd3127f9885e1f40bc726
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

ASSINADO - Paulo Roberto Prezotti Filho (pprezotti@ifes.edu.br)
Data/Hora: 29/07/2025 - 06:29:44, IP: 187.64.156.5, Geolocalização: [-20.375907, -40.305695]
[SHA256]: 3d90b2ba299e6481cf2ca5c166e002ed828c6194e827c7a53c4defac51366c35
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

Histórico de eventos registrados neste envelope

29/07/2025 06:29:44 - Envelope finalizado por pprezotti@ifes.edu.br, IP 187.64.156.5
29/07/2025 06:29:44 - Assinatura realizada por pprezotti@ifes.edu.br, IP 187.64.156.5
29/07/2025 06:29:29 - Envelope visualizado por pprezotti@ifes.edu.br, IP 187.64.156.5
28/07/2025 17:42:14 - Assinatura realizada por fabio.valentim@ufes.br, IP 164.163.204.221
28/07/2025 17:42:02 - Envelope visualizado por fabio.valentim@ufes.br, IP 164.163.204.221
28/07/2025 17:37:19 - Assinatura realizada por alancardek.araujo@ufes.br, IP 172.86.185.6
28/07/2025 17:37:11 - Envelope visualizado por alancardek.araujo@ufes.br, IP 172.86.185.6
28/07/2025 08:00:25 - Envelope registrado na Blockchain por notificacao@astenassinatura.com.br
28/07/2025 08:00:24 - Envelope encaminhado para assinaturas por notificacao@astenassinatura.com.br
24/07/2025 10:22:14 - Envelope criado por ivan.barbosa@ufes.br, IP 200.137.65.106

A Deus, por ser meu sustento nos dias silenciosos
e minha força nas horas de cansaço.

Agradecimentos

A conclusão desta dissertação representa muito mais do que o encerramento de uma etapa acadêmica; é a síntese de uma jornada marcada por esforço, aprendizado, superações e, sobretudo, pelo apoio, incentivo e generosidade de inúmeras pessoas às quais sou profundamente grata.

Agradeço, primeiramente, a Deus, que me sustentou nos momentos mais difíceis e guiou cada passo do meu caminho. Sem Sua presença constante, nada disso teria sido possível.

Ao meu orientador, professor Fábio Júlio, expresso minha profunda gratidão pelas valiosas contribuições, pela escuta atenta e pela confiança depositada ao longo de todo o processo. Sua dedicação e comprometimento foram essenciais para a realização deste trabalho.

À Universidade Federal do Espírito Santo, pela oportunidade de formação e pelos ambientes de aprendizagem e reflexão que ampliaram minha visão acadêmica e profissional. Ao Programa de Pós-Graduação e a todos os professores que, direta ou indiretamente, fizeram parte da minha trajetória, deixo meu sincero reconhecimento.

À equipe de gestão da escola onde atuo, em especial ao diretor Helio Pettene, que sempre apoia e não mede esforços para encorajar seus profissionais a se capacitarem e buscarem seu aprimoramento contínuo. Sua liderança é inspiradora e fundamental para o desenvolvimento de todos nós.

Às minhas amigas Laís (Lala) e Gabriela, que estiveram ao meu lado durante toda essa caminhada — obrigada pela cumplicidade, pelas conversas esclarecedoras, pelo companheirismo e, claro, por suportarem minhas loucuras com tanto carinho e paciência. Minhas quintas-feiras jamais seriam as mesmas sem vocês. Aos pais da Lala, Marinete e Luís, e à sua família, que me acolheram como se fosse uma filha, desejo que Deus retribua em dobro todo esse acolhimento e afeto.

Aos colegas e amigos de profissão — especialmente aos professores de Matemática da escola onde atuo — agradeço pelas trocas diárias, pela convivência e pelas discussões enriquecedoras que inspiraram diretamente a proposta deste trabalho. Espero que este manual possa, de alguma forma, contribuir para a prática pedagógica de todos vocês.

À minha família, alicerce fundamental, agradeço pelo amor incondicional, pela força nas horas difíceis e por jamais deixarem de acreditar em mim. À minha mãe, que com suas orações e fé me sustentou espiritualmente a cada passo. Ao meu companheiro de vida, Luis Carlos, que me apoiou, me encorajou e nunca deixou que eu esquecesse a minha força e o quanto sou capaz. Ao meu pai, que partiu cedo demais e não pôde estar presente fisicamente nesta conquista, mas permanece vivo em meu coração e nas lembranças de cada vitória. E ao meu novo amor, que chegou no tempo perfeito de Deus — meu pequeno Arthur, que cresce em meu ventre e acompanha, mesmo em silêncio, os últimos passos desta jornada. A doce espera que me presenteou não só com um novo título acadêmico, mas com o eterno título de mamãe.

Aos amigos que celebraram comigo cada semestre vencido, que me levantaram nos momentos de cansaço e nunca se afastaram, mesmo diante da correria e das longas horas de estudo — meu muito obrigada por estarem sempre por perto.

Por fim, agradeço a todos que, de alguma forma, contribuíram para que esta dissertação se tornasse realidade. A cada um e cada uma, o meu mais sincero e afetuoso muito obrigada!

“Ensinar é um exercício de imortalidade.
De alguma forma continuamos a viver
naqueles cujos olhos
aprenderam a ver o mundo
pela magia da nossa palavra.
O professor, assim não morre
Jamais... "
(Rubem Alves)

Resumo

Esta dissertação dedica-se à elaboração de um manual de orientação didático-pedagógica para auxiliar professores no planejamento e desenvolvimento de aulas do componente curricular Estatística, inserido no itinerário formativo de Educação Financeira e Fiscal do Novo Ensino Médio, conforme as diretrizes da Base Nacional Comum Curricular (BNCC) e do currículo oficial do Estado do Espírito Santo. A pesquisa parte do reconhecimento das dificuldades enfrentadas pelos docentes ao lidar com novos componentes curriculares para os quais, muitas vezes, não receberam formação específica. O trabalho analisa os fundamentos do ensino de Estatística na BNCC e propõe sugestões práticas alinhadas à realidade escolar, visando tornar a aprendizagem mais contextualizada e significativa. A utilização de dados locais do Espírito Santo é incorporada em diversas atividades e exemplos desenvolvidos, buscando contextualizar a aprendizagem e promover maior engajamento dos estudantes. Como contribuição, o material desenvolvido oferece subsídios teóricos e práticos para fortalecer a atuação docente, promovendo competências como pensamento crítico e tomada de decisões fundamentadas.

Palavras-chave: Ensino de Estatística; Novo Ensino Médio; Itinerários Formativos; Estatística; Educação Financeira e Fiscal; Prática docente.

Abstract

This dissertation is dedicated to the development of a didactic-pedagogical guidance manual to support teachers in the planning and implementation of lessons on Statistics, a curricular component within the Financial and Fiscal Education track of the New High School model, in accordance with the guidelines of the Brazilian National Common Curricular Base (BNCC) and the official curriculum of the state of Espírito Santo. The research is based on the recognition of the challenges faced by teachers when working with new curricular components, often without having received specific training. The study examines the foundations of Statistics education as presented in the BNCC and offers practical suggestions aligned with school contexts, seeking to make learning more meaningful and contextualized. The use of local data from Espírito Santo is incorporated into several activities and examples developed throughout the work, aiming to enhance student engagement through contextual relevance. As its main contribution, the material provides both theoretical and practical support to strengthen teaching practices, promoting competencies such as critical thinking and informed decision-making.

Keywords: Statistics Education; New High School; Educational Tracks; Financial and Fiscal Education; Teaching Practice; Teaching Practice.

Lista de ilustrações

Lista de tabelas

Sumário

1 Introdução

"Sem dados, você é apenas mais uma pessoa com uma opinião."– *W. Edwards Deming*

A célebre frase de William Edwards Deming evidencia o papel essencial da informação no processo decisório. Em um contexto social marcado pela circulação massiva de conteúdos, a habilidade de interpretar dados com criticidade torna-se indispensável. Em um mundo onde todos têm algo a dizer, são os dados que oferecem solidez aos argumentos, direcionam escolhas e embasam posicionamentos de maneira responsável.

Contudo, os dados não se limitam a números ou estatísticas frias; eles traduzem comportamentos, revelam padrões e sinalizam possibilidades. Interpretá-los corretamente requer mais do que acesso à informação: requer análise criteriosa, contextualização e aplicação prática.

Nesse contexto, a formação estatística dos estudantes assume papel central desde os primeiros anos da Educação Básica. Para além da aplicação mecânica de fórmulas, o ensino de Estatística deve desenvolver habilidades voltadas à compreensão crítica da realidade. Conforme estabelecido pela Lei de Diretrizes e Bases da Educação Nacional (LDB, Lei nº 9.394/1996) e reafirmado pela Base Nacional Comum Curricular (BNCC), o estudo da Estatística no currículo escolar busca formar cidadãos capazes de interpretar, comunicar e tomar decisões com base em dados.

A BNCC contempla a Estatística na unidade temática "Probabilidade e Estatística", presente ao longo do Ensino Fundamental e Médio, enfatizando habilidades como coleta, organização, representação e análise de dados em diferentes contextos. Segundo o documento:

> "[...] habilidades para coletar, organizar, representar, interpretar e analisar dados em uma variedade de contextos, de maneira a fazer julgamentos bem fundamentados e tomar as decisões adequadas. Isso inclui raciocinar e utilizar conceitos, representações e índices estatísticos para descrever, explicar e predizer fenômenos." (BRASIL, 2017, p. 272)

Esse desenvolvimento ocorre de forma progressiva. Nos primeiros anos do Ensino Fundamental, os alunos realizam pesquisas de interesse próprio, representando dados em gráficos, tabelas e textos. Nos anos finais, são introduzidos conceitos como médias, medidas de dispersão e amostragem. No Ensino Médio, a Estatística passa a ser integrada ao componente curricular de Matemática, aproximando-se das vivências dos estudantes e dos recursos tecnológicos disponíveis.

Esse percurso revela que a Estatística não deve ser tratada como um conteúdo meramente técnico, mas como uma ferramenta essencial para a formação cidadã, capaz de ampliar a compreensão sobre fenômenos sociais, econômicos e ambientais, além de fomentar a tomada de decisões conscientes e fundamentadas. A integração de tecnologias digitais — como planilhas, bases de dados públicas (ex: IBGE) e recursos computacionais — reforça ainda mais seu papel na construção do pensamento crítico e reflexivo.

Com a reforma do Novo Ensino Médio, instituída pela Lei nº 13.415/2017 (Brasil, 2017), a estrutura curricular da educação básica foi reorganizada, com destaque para os itinerários formativos, que visam ao aprofundamento em áreas de interesse dos estudantes. No itinerário de *Educação Financeira e Fiscal*, incluído na área de Matemática e suas Tecnologias, a Estatística surge como componente autônomo, com o propósito de ampliar os conhecimentos abordados no currículo comum, por meio de uma abordagem aplicada, contextualizada e interdisciplinar.

Entretando, a implementação desses novos componentes revelou desafios significativos. Muitos docentes, ainda que experientes no ensino da Matemática tradicional, encontram dificuldades para planejar e ministrar aulas dessas novas disciplinas, incluindo *Estatística* com o enfoque proposto pelos itinerários formativos. Em diversos casos, esses profissionais não possuem formação específica na área de licenciatura em matemática, nem acesso a materiais adequados para adaptar conteúdos frequentemente presentes em livros de nível superior à realidade e às necessidades do Ensino Médio. Tais limitações impactam diretamente o processo de ensino-aprendizagem e enfraquecem o potencial da Estatística como ferramenta de leitura crítica da realidade.

Diante desse cenário, considera-se que a ausência de formação específica e de materiais didático-pedagógicos alinhados às diretrizes do Novo Ensino Médio representa um entrave significativo à atuação de professores de Matemática no ensino da *Estatística*. A hipótese que norteia esta pesquisa é que a elaboração de um manual orientador, fundamentado nas orientações curriculares do Estado do Espírito Santo e na BNCC, pode colaborar de forma efetiva com o planejamento pedagógico dos docentes, promovendo um ensino mais contextualizado, acessível e significativo.

Assim, esta dissertação tem como objetivo geral: **Elaborar um manual de orientação didático-pedagógica para auxiliar professores no planejamento e desenvolvimento de aulas do componente curricular Estatística, considerando o currículo oficial do Governo do Estado do Espírito Santo.**

Para atingir esse objetivo, propõem-se os seguintes objetivos específicos:

- Analisar as diretrizes da BNCC e do currículo estadual do Espírito Santo relacionadas ao ensino de Estatística no contexto dos itinerários formativos;

- Propor sugestões práticas de exemplos e atividades, voltadas para o ensino de Estatística de forma contextualizada e significativa;
- Promover a integração entre os conteúdos estatísticos, o uso de tecnologias digitais e os contextos cotidianos dos alunos, visando o desenvolvimento de competências previstas no currículo.

Ao oferecer aos docentes ferramentas pedagógicas alinhadas às diretrizes educacionais e à realidade escolar, pretende-se fortalecer o papel da Estatística na formação crítica dos estudantes, ampliando sua capacidade de compreender, interpretar e intervir no mundo. A relevância deste estudo está, portanto, tanto em seu aporte teórico quanto em sua aplicação prática, contribuindo para a qualificação do ensino e para o desenvolvimento de competências essenciais em uma sociedade orientada por dados.

Esta dissertação está organizada em cinco capítulos, além da introdução e da conclusão. O **Capítulo 2**, intitulado *O Contexto Atual do Ensino Médio: Estrutura Curricular e a Implementação dos Itinerários Formativos no Estado do Espírito Santo*, discute a reforma do Novo Ensino Médio, com foco na reestruturação curricular e na inserção dos Itinerários Formativos. São apresentados os fundamentos legais e pedagógicos que orientam essa mudança, com destaque para os quatro eixos estruturantes que norteiam a proposta. Em particular, é analisado o itinerário de *Educação Financeira e Fiscal*, pertencente à área de Matemática e suas Tecnologias, ressaltando-se o papel da Estatística na promoção da autonomia, da criticidade e da capacidade investigativa dos estudantes da rede estadual capixaba.

No **Capítulo 3**, *Fundamentos e Práticas em Estatística: Teoria e Exemplos para o Ensino Médio*, são abordados os principais objetos de conhecimentos estatísticos previstos nas orientações curriculares do referido itinerário. O capítulo propõe uma fundamentação teórica articulada a situações práticas do cotidiano dos alunos, reforçando a Estatística como ferramenta essencial para a interpretação da realidade e para o desenvolvimento de competências relacionadas à leitura, análise e comunicação de dados em contextos educacionais e sociais.

A base teórica deste capítulo apoia-se principalmente em autores clássicos e amplamente reconhecidos na área da Estatística, como Triola (2017), que oferece uma introdução clara e aplicada à disciplina; MORETTIN e BUSSAB (2010), referência consolidada em técnicas estatísticas; e Oliveira (2018), professor da Universidade Federal da Bahia, que elaborou material didático voltado à aplicação da Estatística nas ciências sociais, contribuindo com uma abordagem acessível e contextualizada para fins educacionais. Além desses, outros autores são citados pontualmente ao longo do capítulo para complementar discussões específicas, enriquecer o conteúdo e ampliar as possibilidades de aplicação pedagógica dos conceitos abordados.

O **Capítulo 4**, denominado *Sugestões de Práticas para o Ensino de Estatística*, apresenta um conjunto de propostas pedagógicas que buscam promover uma aprendizagem significativa e contextualizada. As atividades sugeridas contemplam conteúdos como medidas de tendência central, medidas de dispersão e noções introdutórias de probabilidade, associados a temas relevantes e atuais — como consumo consciente, desempenho escolar e tomada de decisão sob incerteza. Essas práticas favorecem o desenvolvimento do pensamento científico, da argumentação baseada em evidências e do engajamento dos estudantes em questões sociais e ambientais, em consonância com as competências gerais da Base Nacional Comum Curricular (BNCC).

2 O Contexto Atual do Ensino Médio: Estrutura Curricular e a Implementação dos Itinerários Formativos no Estado do Espírito Santo

O Ensino Médio, etapa final do ensino básico, desempenha um papel fundamental na trajetória educacional dos alunos, pois antecede escolhas cruciais em suas vidas, como a definição de uma carreira profissional, a decisão entre ingressar no ensino superior ou buscar oportunidades no mercado de trabalho, além de influenciar no desenvolvimento da autonomia, da identidade e da cidadania dos jovens.

Considerando esses fatores, torna-se essencial analisar e sugerir mudanças para aprimorar a experiência de aprendizado nessa fase, tornando o "aprender" mais significativo no projeto de vida dos estudantes. Nesse cenário, a preocupação com o desempenho insatisfatório e a alta taxa de abandono escolar no Ensino Médio evidenciou a necessidade de uma nova abordagem curricular, mais conectada às necessidades e aspirações dos jovens. Assim, em 2017, foi aprovada uma reforma no ensino médio com o objetivo de promover o protagonismo dos estudantes e incentivá-los a permanecer na escola. Essa reforma estabeleceu uma nova organização curricular, que combina aulas de formação geral básica, alinhadas à Base Nacional Comum Curricular (BNCC), com Itinerários Formativos[1].

De acordo com o documento *Novo Ensino Médio Capixaba - Itinerário Formativo* (SECRETARIA DE ESTADO DA EDUCAÇÃO DO ESPÍRITO SANTO, s.d.):

> Os Itinerários Formativos foram elaborados com base no documento "Referenciais Curriculares para a Elaboração dos Itinerários Formativos", Portaria 1.432 (BRASIL, 2018) que orienta os sistemas de ensino na construção dos itinerários formativos, visando atender as Diretrizes Curriculares Nacionais do Ensino Médio, publicadas na Resolução MEC/CNE/CEB nº 3, de 21 de novembro de 2018, e a Lei nº 13.415, de 16 de fevereiro de 2017.

Segundo os *Referenciais Curriculares para a Elaboração dos Itinerários Formativos* que consta na Portaria 1.432 (BRASIL, 2018), esse modelo de Ensino Médio busca preparar os alunos para enfrentar desafios pessoais, profissionais, sociais, culturais e ambientais,

[1] Itinerários Formativos é a parte flexível do currículo, se refere a uma porção do currículo do ensino médio que não é definida pelo conteúdo comum (que é obrigatório para todos os estudantes), mas sim pelas escolhas do aluno. O estudante tem a oportunidade de aprofundar e desenvolver habilidades e competências mais específicas para sua área escolhida.

considerando as rápidas transformações da sociedade contemporânea. As Diretrizes Curriculares Nacionais para o Ensino Médio (DCNEM), atualizadas em 2018, estabelecem que a nova organização curricular deve ser composta por:

- Formação Geral Básica: Conjunto de competências das áreas de Linguagens, Matemática, Ciências da Natureza e Ciências Humanas, com foco em aprofundar e consolidar os conhecimentos essenciais adquiridos no Ensino Fundamental. Essa parte do currículo tem carga horária máxima de 1.800 horas.
- Itinerários Formativos: Conjunto de unidades curriculares oferecidas pelas instituições e redes de ensino, que possibilita ao estudante desenvolver competências e habilidades em uma ou mais áreas do conhecimento ou na formação técnica e profissional, com carga horária mínima de 1.200 horas.

A estruturação dos Itinerários Formativos baseia-se em quatro eixos estruturantes: Investigação Científica, Processos Criativos, Mediação e Intervenção Sociocultural e Empreendedorismo, conforme representado na Figura 1.

Figura 1 – Quatro eixos fundamentais na construção dos Itinerários.

Fonte: (SEDU, S.db)

Esses eixos têm como finalidade promover uma formação integral dos estudantes — pessoal, profissional e cidadã — por meio de experiências de aprendizagem que os capacitem a gerar conhecimento, inovar, intervir criticamente na sociedade e desenvolver projetos com impacto no presente e no futuro.

Cada eixo apresenta um foco específico. A **Investigação Científica** visa o desenvolvimento de habilidades de pesquisa e análise crítica, incentivando o aluno a explorar e compreender o mundo por meio do método científico. Os **Processos Criativos** estão voltados ao estímulo da inovação, da expressão artística e do pensamento criativo, contribuindo para a formação de sujeitos capazes de propor soluções originais. A **Mediação e Intervenção Sociocultural** contempla práticas relacionas ao trabalho na comunidade, desenvolvendo a capacidade de atuar em contextos sociais e culturais e promovendo habilidades para mediar e intervir em diversas situações. Por fim, o **Empreendedorismo** busca desenvolver competências para criar e gerir negócios, além de capacitar o aluno para inovar em diferentes setores, preparando-o para o mercado de trabalho e para a liderança.

Nas escolas da Rede Estadual de Ensino do Espírito Santo, os Itinerários Formativos de Aprofundamento são ofertados conforme a demanda e as preferências manifestadas pelos estudantes. Esses itinerários são organizados com base nas Áreas de Conhecimento ou na integração entre elas, possibilitando uma formação mais flexível e alinhada aos interesses individuais. As modalidades disponíveis são:

- Linguagens e suas Tecnologias;
- Matemática e suas Tecnologias;
- Ciências da Natureza e suas Tecnologias;
- Ciências Humanas e Sociais Aplicadas;
- Formação Técnica e Profissional.

A oferta pode ocorrer de forma isolada, por área específica, ou de maneira integrada, envolvendo mais de uma área do conhecimento. As Figuras 2 e 3 ilustram, respectivamente, os itinerários organizados por área de conhecimento e os itinerários integrados entre diferentes áreas, evidenciando as possibilidades formativas disponibilizadas aos estudantes da rede estadual.

Dentre essas possibilidades, este trabalho tem como foco o Itinerário Formativo da área de Matemática e suas Tecnologias, intitulado “Educação Financeira e Fiscal”, conforme descrito no Catálogo dos Itinerários Formativos de Aprofundamento da SEDU (SEDU, S.db). A seguir, será apresentada a fundamentação e a proposta pedagógica que estruturam esse itinerário, destacando sua importância na formação crítica e empreendedora dos estudantes.

Figura 2 – Itinerários por área de conhecimento.

Fonte: (SEDU, S.da)

Figura 3 – Itinerários entre áreas de conhecimento.

Fonte: (SEDU, S.da)

2.1 Itinerário Formativo: Educação Financeira e Fiscal

> O aprofundamento em Matemática visa a contribuir de forma significativa na formação do estudante e, por consequência, de um cidadão com perfil empreendedor, competente, com consciência socioeconômica, investigativa e ética e com condições de desenvolver e realizar seus projetos individuais e coletivos, comprometido com o desenvolvimento local e regional (SEDU, S.db).

Nesse sentido, o aprofundamento em Matemática busca desempenhar um papel significativo na formação integral do adolescente, preparando-o para exercer a cidadania de forma consciente e responsável. Isso significa ensiná-lo a consumir e poupar de maneira ética, fornecendo conceitos e ferramentas para tomar decisões autônomas com base em uma mudança de atitude em seu planejamento financeiro a curto, médio e longo prazos. Além disso, a proposta visa criar oportunidades que permitam ao aluno refletir sobre sua realidade socioeconômica e, eventualmente, transformá-la.

Com o propósito de aprofundar os conhecimentos e habilidades desenvolvidos na Formação Geral Básica, esse itinerário propõe uma abordagem dinâmica para o ensino e a aprendizagem de conceitos relacionados à Matemática Financeira, Probabilidade, Estatística e Consumo Consciente. Esses conteúdos são organizados em módulos trimestrais, ofertados na 2ª e 3ª séries do Ensino Médio. Ainda que as duas séries compartilhem determinados objetos de conhecimento, a 3ª série possibilita uma abordagem mais aprofundada dos conteúdos (SEDU, 2023).

Para atingir os objetivos propostos, o itinerário integra componentes curriculares que fazem parte da parte flexível do currículo, permitindo uma organização dos objetos de conhecimento por séries e trimestres (as orientações curriculares podem ser observadas no Apêndice A). Dentre os componentes oferecidos, destacam-se: **Matemática Financeira**, **Estatística**, **Consumo Responsável e Educação Tributária**, **Educação Financeira** e **Projetos em Educação Financeira**, formando uma proposta interdisciplinar ao integrar a teoria e prática, além de estimular o protagonismo e a autonomia dos estudantes, oferecendo-lhes as ferramentas necessárias para uma tomada de decisão financeira consciente e responsável.

Nesse conjunto de componentes, a **Estatística** ocupa um papel relevante por possibilitar ao estudante o desenvolvimento de habilidades essenciais para a leitura e interpretação de dados socioeconômicos. Embora não constitua o foco central deste itinerário formativo, a Estatística é o componente curricular diretamente relacionado a esta pesquisa. Ao longo dos dois últimos anos do Ensino Médio, os estudantes têm acesso a uma sequência de objetos de conhecimento que vai desde fundamentos básicos até técnicas mais avançadas, como testes de hipóteses e distribuições probabilísticas, o que pode ser observado na Tabela 1. Essa progressão culmina na realização de projetos de pesquisa estatística, nos quais os alunos aplicam os conhecimentos adquiridos em situações reais e investigativas.

Além da aprendizagem conceitual, a disciplina de Estatística favorece o desenvolvimento de habilidades específicas estabelecidas na Base Nacional Comum Curricular para o Ensino Médio, como: **investigar e analisar situações-problema** (EMIFMAT01[2]), **formular e testar hipóteses** (EMIFMAT02), e **selecionar e comunicar informações com base em pesquisas** (EMIFMAT03). Tais competências fortalecem o pensamento crítico, a argumentação fundamentada e a capacidade de reconhecer a Matemática como ferramenta para interpretar fenômenos sociais, científicos e culturais.

[2] A sigla EMIFMAT refere-se às habilidades específicas de Matemática no âmbito do Itinerário Formativo do Ensino Médio. Cada código é composto pelas iniciais de Ensino Médio (EM), Itinerário Formativo (IF) e Matemática (MAT), seguido por um número que identifica a habilidade. Essas habilidades fazem parte da Base Nacional Comum Curricular (BNCC) e orientam o desenvolvimento de competências matemáticas voltadas à resolução de problemas, pensamento crítico, argumentação e aplicação em contextos reais.

Tabela 1 – Objetos de Conhecimento – Estatística

Série	**Trimestre**	**Objeto de Conhecimento**
2ª série	1º	Conceitos Básicos de Estatística; Estatística Descritiva/Inferencial/Probabilística; Variáveis quantitativas e qualitativas; Tipos de dados; Dados brutos/Rol; Variável quantitativa/qualitativa; Frequência: Simples, Relativa, Acumulada e Relativa acumulada; Introdução às medidas de tendência central; Introdução às medidas de dispersão.
	2º	Gráficos (conceitos e tipos); Técnicas de Amostragem: Tipos de amostragem; Erro amostral (conceitos e determinação); Valor esperado (conceitos e determinação).
	3º	Pesquisa Estatística: Teste de Hipóteses: Conceitos e aplicações; Pesquisas estatísticas: realização de projeto de pesquisa do aluno e aplicação dos conceitos
3ª série	1º	Medidas de Tendência Central: Média (conceitos, tipos e determinação); Moda (conceitos e determinação); Separatrizes (conceitos e determinação). Medidas de Dispersão: Desvio Padrão (conceito e determinação).
	2º	Intervalo de Confiança (conceitos e determinação); Distribuição Normal (conceitos e determinação).
	3º	Pesquisa Estatística: Teste de Hipóteses: Conceitos e aplicações; Pesquisas estatísticas: realização de projeto de pesquisa do aluno e aplicação dos conceitos.

Fonte: Adaptado de (SEDU, 2024)

A Estatística também se mostra relevante no contexto das **questões socioculturais e ambientais**, permitindo ao aluno propor estratégias de intervenção baseadas em dados concretos (EMIFMAT07, EMIFMAT08, EMIFMAT09). Além disso, as atividades desenvolvidas promovem a **criatividade e a inovação na resolução de problemas** (EMIFMAT04, EMIFMAT05, EMIFMAT06), incentivando a produção de novos conhecimentos matemáticos aplicáveis à realidade.

Dessa forma, o ensino de Estatística no itinerário de Educação Financeira e Fiscal contribui não apenas para a formação técnica dos estudantes, mas também para sua atuação consciente na sociedade, reforçando sua autonomia, protagonismo e capacidade de transformar sua realidade com base em decisões fundamentadas em dados.

3 Fundamentos e Exemplos em Estatística para o Ensino Médio

Este capítulo tem como objetivo explorar a teoria proposta na organização curricular para a disciplina de aprofundamento em Estatística, conforme apresentada no capítulo anterior. Busca-se contribuir para o ensino e o aprendizado dessa área, oferecendo uma fundamentação teórica sólida sobre os objetos de conhecimento destacados no documento curricular (Tabela 1), visando fortalecer a compreensão e a aplicação prática desses conceitos no contexto educacional.

A base teórica deste capítulo apoia-se principalmente em autores clássicos e amplamente reconhecidos na área da Estatística, como Triola (2017), que oferece uma introdução clara e aplicada à disciplina; MORETTIN e BUSSAB (2010), referência consolidada em técnicas estatísticas; e Oliveira (2018), professor da Universidade Federal da Bahia, que elaborou material didático voltado à aplicação da Estatística nas ciências sociais, contribuindo com uma abordagem acessível e contextualizada para fins educacionais. Além desses, outros autores são citados pontualmente ao longo do capítulo para complementar discussões específicas, enriquecer o conteúdo e ampliar as possibilidades de aplicação pedagógica dos conceitos abordados.

A ordem de apresentação dos conteúdos neste capítulo foi definida a partir da proposta curricular da disciplina, indicada anteriormente na Tabela 1, visando respeitar uma progressão coerente com a organização sugerida pelo currículo estadual. Embora reconheçamos que outras sequências sejam possíveis e até desejáveis dependendo do contexto pedagógico, optamos por manter a estrutura sugerida no documento oficial como forma de alinhar esta proposta à realidade curricular vigente e oferecer aos professores um ponto de partida claro e articulado com as diretrizes educacionais.

3.1 Definição de Estatística

Definição 3.1.1. *A Estatística é um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.*

O estudo da Estatística se faz necessário para compreender o mundo ao nosso redor e tomar decisões informadas. Em um cenário em que estamos constantemente cercados por dados, saber como coletá-los, organizá-los e interpretá-los de forma correta é fundamental. A Estatística nos permite analisar padrões, prever comportamentos e avaliar a efetividade de ações, o que é crucial em áreas como saúde, educação, negócios e até mesmo no nosso dia a dia. Além disso, com o avanço de tecnologias como o Big Data[1] e a Inteligência Artificial[2], ter conhecimentos estatísticos se tornou uma habilidade cada vez mais importante para navegar no vasto volume de informações que temos acesso.

A Estatística pode ser compreendida em duas áreas distintas:

- *Estatística Descritiva*, que tem como finalidade organizar, resumir e apresentar os dados de forma clara e compreensível. Para isso utiliza ferramentas como tabelas, gráficos e medidas numéricas, que ajudam a identificar padrões e tendências nos dados. O foco é fornecer uma visão geral das informações coletadas, facilitando a interpretação dos dados de maneira rápida e objetiva.
- *Estatística Inferencial*, que extrapola a descrição dos dados para realizar previsões ou inferências sobre uma população maior a partir de uma amostra. Por meio de métodos como testes de hipóteses, intervalos de confiança e modelos matemáticos, possibilita estimativas e generalizações mesmo quando não se dispõe de informações completas sobre o conjunto total de dados.

A Estatística Inferencial apoia-se fortemente na *Probabilidade*, ramo que estuda os fenômenos aleatórios e a incerteza, fornecendo as ferramentas necessárias para calcular probabilidades e compreender a distribuição dos eventos. Conceitos fundamentais, tais como variáveis aleatórias e distribuições de probabilidade, são essenciais para a análise, previsão do comportamento dos dados e fundamentação das inferências estatísticas.

Dessa forma, a Estatística e a Probabilidade são complementares: enquanto a Probabilidade oferece a base teórica, a Estatística Descritiva organiza e apresenta os dados, e a Estatística Inferencial permite a extração de conclusões e a tomada de decisões fundamentadas a partir da análise dos dados. Juntas, constituem uma estrutura essencial com ampla aplicação em áreas como ciência, saúde, negócios e educação.

3.2 Alguns conceitos importantes

Vamos estabelecer alguns conceitos preliminares:

- **Dado** - Fato ou número coletado, analisado e sintetizado para apresentação e interpretação.

[1] Big Data é o termo usado para descrever grandes volumes de dados complexos que não podem ser processados por métodos tradicionais. Ele envolve a coleta, armazenamento e análise de dados em grande escala para identificar padrões, tendências e insights valiosos, com aplicações em diversas áreas, como negócios, saúde e tecnologia.(ALECRIM, 2024)

[2] Inteligência Artificial consiste em sistemas projetados para fornecer respostas eficientes em formatos diversos (texto, imagem, áudio, vídeo, código), abordando uma ampla gama de assuntos, desde que as perguntas sejam claras e precisas.

- **População** ou **Universo estatístico** - Conjunto total de elementos ou indivíduos que possuem uma característica comum e que se deseja estudar. Pode ser uma população finita ou infinita.
- **Unidade estatística** ou **experimental** - Refere-se ao indivíduo ou objeto que está sendo observado ou analisado em um conjunto de dados.
- **Amostra** - Subconjunto de unidades experimentais da população.
- **Amostragem** - Procedimento para obtenção da amostra a partir de uma dada população.
- **Variável** - Característica ou propriedade de uma unidade experimental que pode assumir diferentes valores para cada elemento (ou indivíduo) de um grupo, seja ele uma população ou uma amostra. Por exemplo, a altura, o peso, a cor dos olhos e a idade são variáveis, pois seus valores podem variar — isto é, ser diferentes — entre as pessoas. Recebe o nome de "variável" justamente porque seu valor não é fixo, mas muda conforme a unidade analisada.

Do ponto de vista da teoria dos conjuntos, esses conceitos podem ser formalizados da seguinte forma: a população é representada por um conjunto Ω; cada unidade estatística corresponde a um elemento desse conjunto, denotado por $\omega \in \Omega$; a amostra é um subconjunto de Ω, ou seja, $A \subset \Omega$; e a variável pode ser interpretada como uma função cujo domínio é Ω e cujo contradomínio é um conjunto que contém todas as possíveis características dos elementos $\omega \in \Omega$. Frequentemente, essas características são representadas por números reais (ou vetores em $\mathbb{R}^n$), permitindo a aplicação de técnicas quantitativas na análise dos dados.

A seguir, apresenta-se um exemplo fictício que ilustra a aplicação desses conceitos em um cenário de investigação:

Exemplo 3.2.1. *Suponha que se deseje investigar os hábitos de consumo de frutas entre os habitantes de um município, com o objetivo de avaliar a viabilidade da abertura de um hortifrúti na região. Pretende-se compreender a frequência, a quantidade e os tipos de frutas consumidas pela população local, de modo a embasar decisões estratégicas para o novo empreendimento comercial.*

Considerando que a população do município seja de aproximadamente 100 mil habitantes, entrevistar todos os moradores seria uma tarefa custosa, demorada e, na maioria das vezes, inviável. Para contornar essa dificuldade, utiliza-se uma pesquisa amostral.

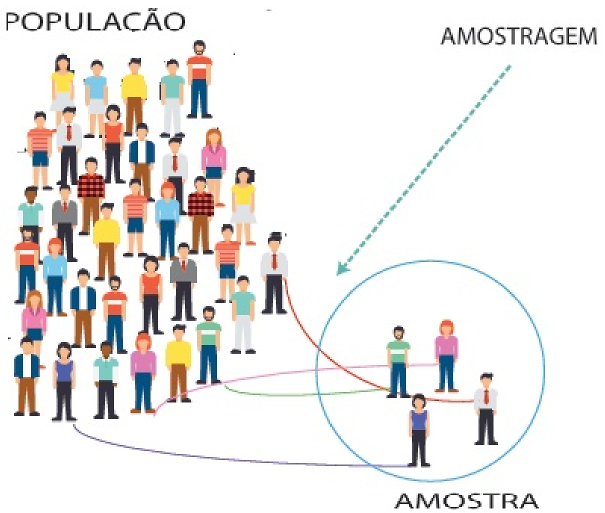
Nesse contexto, suponha-se que uma amostra aleatória de 500 moradores seja selecionada para responder a um questionário sobre seus hábitos alimentares relacionados ao consumo de frutas.

Delimitamos então:

- ***População ou universo estatístico:*** *Todos os habitantes do município (aproximadamente 100 mil pessoas).*
- ***Unidade estatística ou experimental:*** *Um morador do município.*
- ***Amostra:*** *Um grupo de 500 moradores sorteados aleatoriamente.*
- ***Dado:*** *O número de porções de frutas consumidas por dia por cada entrevistado.*
- ***Amostragem:*** *Processo de seleção aleatória simples, no qual todos os moradores têm a mesma probabilidade de serem escolhidos.*
- ***Variável:*** *Quantidade de porções de frutas consumidas diariamente por indivíduo.*

Vale observar, conforme ilustrado na Figura 4, que a amostra é um subconjunto da população. A forma como os 500 indivíduos foram selecionados caracteriza o processo de amostragem. Esse procedimento visa garantir que as informações obtidas por meio da amostra possam ser generalizadas para a população como um todo, com um grau aceitável de confiabilidade.

Figura 4 – Relação entre população, amostragem e amostra



Fonte: Produção da própria autora (2024), criado com a ferramenta disponível em https://www.visme.co/pt-br/.

A população é composta por todos os elementos que serão analisados, abrangendo todas as informações pertinentes ao estudo. Diante disso, por que optar pela amostra em vez de considerar a população completa? A razão para usar amostragem é bastante evidente: coletar dados de toda a população requer significativamente mais tempo e recursos. Além disso, em muitas situações, é impraticável obter informações de cada unidade da população. Portanto, a possibilidade de obter dados relevantes sobre a população a partir de uma amostra se torna uma alternativa bastante atraente.

3.3 Classificação das Variáveis

As variáveis são categorizadas com base na "origem"dos **dados**. Elas podem ser:

1. **Qualitativa** - também conhecidas como variáveis categóricas, são aquelas que descrevem características ou atributos e não podem ser medidas numericamente. Elas podem ser divididas em duas subcategorias:

   - **Ordinal** - as características ou atributos admitem alguma ordenação lógica/natural (ou de intensidade). Por exemplo Grau de escolaridade (fundamental, médio, superior), classificação de um filme, dentre outras;
   - **Nominal** - não possuem uma ordem específica. Por exemplo: cor dos olhos (azul, verde, castanho), tipos de frutas (maçã, banana, laranja), dentre outras.

   Mas também é possível converter a informação desse tipo de variável em um padrão numérico, entretanto, esse valor não tem um significado quantitativo. Por exemplo Ao investigar a frequência com que uma pessoa pratica atividades físicas (escala ordinal), podemos categorizar suas respostas em quatro níveis: nunca, raramente, frequentemente e sempre. Para facilitar a análise, podemos converter essas respostas em uma escala numérica, atribuindo 0 para nunca, 1 para raramente, 2 para frequentemente e 3 para sempre. Assim, conseguimos quantificar a frequência de atividades físicas de forma mais prática e comparativa.

2. **Quantitativa** - são aquelas que podem ser medidas numericamente e representam quantidades. Elas também se dividem em duas subcategorias:

   - **Discreta** - podem assumir valores inteiros e contáveis. Por exemplo: número de filhos, número de carros em um estacionamento (0, 1, 2, ...) ;
   - **Contínua** - podem assumir qualquer valor em um intervalo, incluindo frações e decimais. Por exemplo: altura de uma pessoa (1.70m, 1.75m), Temperatura de uma sala (19°C, 26ºC).

3.4 Estruturas de Organização e Exibição de Dados Coletados

Em uma pesquisa estatística as variáveis mensuradas em uma amostra podem ser dispostas em diferentes formatos, dependendo do tipo de análise e apresentação dos dados. Algumas formas comuns de dispor as observações de variáveis incluem:

1. **Dados Brutos** - são os dados coletados diretamente de uma pesquisa, experimento ou observação, sem qualquer organização, ordenação ou processamento. Isto é, eles estão na forma "original", como foram registrados, podendo conter valores repetidos e não seguem uma ordem específica.

   Exemplo 3.4.1. *Idade (anos) de dez alunos de uma turma específica de ensino médio.*

   16 15 15 17 19
   16 16 16 15 16

   Os dados brutos são fundamentais para garantir a transparência e a reprodutibilidade da pesquisa. Eles permitem que outros pesquisadores verifiquem os resultados, realizem novas análises ou utilizem os dados para outras finalidades.

2. **ROL ou Dados elaborados** - é o conjunto de dados brutos organizado em ordem crescente ou decrescente.

   Exemplo 3.4.2. *Idade (anos) de dez alunos de uma turma específica de ensino médio organizado em ordem crescente.*

   15 15 15 16 16
   16 16 16 17 19

   Essa organização facilita a identificação de valores importantes, como o menor e o maior dado coletado, além de auxiliar na análise de medidas estatísticas.

3. **Tabela de distribuição de frequência simples** - Os dados são agrupados em classes (i) ou categorias, e a frequência de cada grupo é registrada indicando:

   - Frequência Simples ou Absoluta (f_i): Indica quantas vezes cada valor ocorre.
   - Frequência Relativa (f_r): Representa a proporção ou percentual de cada valor em relação ao total. É especialmente útil ao comparar duas ou mais populações de tamanhos distintos.
   - Frequência Acumulada: Soma das frequências até um determinado ponto. Ela se divide em Frequência Absoluta Acumulada (f_{ia}) e Frequência Relativa Acumulada (f_{ra}).

Exemplo 3.4.3. *Consideremos a distribuição das idades dos dez alunos de uma turma do Ensino Médio, citada acima, conforme apresentada na Tabela 2. A tabela apresenta a frequência absoluta (f_i), relativa (f_r), acumulada absoluta (f_{ia}) e acumulada relativa (f_{ra}) de cada valor observado, possibilitando uma análise descritiva inicial dos dados obtidos.*

Tabela 2 – Distribuição de frequências simples das Idades dos alunos

i	**Idade(anos)**	f_i	f_r	f_{ia}	f_{ra}
1	15	3	$\frac{3}{10} = 0,3$ ou 30%	3	$\frac{3}{10} = 0,3$ ou 30%
2	16	5	$\frac{5}{10} = 0,5$ ou 50%	8	$\frac{8}{10} = 0,8$ ou 80%
3	17	1	$\frac{1}{10} = 0,1$ ou 10%	9	$\frac{9}{10} = 0,9$ ou 90%
4	18	0	$\frac{0}{10} = 0$ ou 0%	9	$\frac{9}{10} = 0,9$ ou 90%
5	19	1	$\frac{1}{10} = 0,1$ ou 10%	10	$\frac{10}{10} = 1$ ou 100%
		10	$\frac{10}{10} = 1$ ou 100%		

Fonte: Produção da própria autora (2024).

Exemplo 3.4.4. *Uma empresa composta por 500 funcionários deseja realizar um levantamento sobre dois aspectos de seu quadro de colaboradores: o nível de escolaridade e o número de filhos (Nº F.). Para isso, foi extraída uma amostra de 25 funcionários, cujas informações estão organizadas na Tabela 3, na qual apresenta os dados brutos obtidos na pesquisa.*

Analisar os dados diretamente na tabela de dados brutos pode ser confuso e pouco eficiente. Para facilitar a interpretação das informações coletadas e promover uma visualização mais clara das variáveis estudadas, os dados referentes à pesquisa foram reorganizados em tabelas de distribuição de frequência simples. Nessas tabelas, apresentam-se exclusivamente os valores de frequência absoluta das variáveis em questão, conforme ilustrado nas Tabelas 4 e 5.

A seleção das colunas e das informações apresentadas nas tabelas de distribuição de frequências pode ser ajustada conforme os objetivos específicos da análise estatística. Dependendo do nível de detalhamento desejado, o pesquisador pode incluir, além das frequências absolutas, outras medidas como as frequências relativas e as frequências acumuladas. Esse acréscimo possibilita uma compreensão mais aprofundada da distribuição dos dados, conforme ilustrado nas Tabelas 6 e 7.

Tabela 3 – Quantidade de filhos e escolaridade dos funcionários

Funcionário	Nº F	Escolaridade	Funcionário	Nº F	Escolaridade
1	2	Ensino Superior	14	1	Ensino Fundamental
2	1	Ensino Médio	15	4	Ensino Médio
3	0	Pós-graduação	16	2	Pós-graduação
4	3	Ensino Superior	17	1	Ensino Superior
5	2	Ensino Médio	18	0	Ensino Médio
6	1	Ensino Superior	19	3	Ensino Fundamental
7	0	Ensino Fundamental	20	2	Pós-graduação
8	4	Pós-graduação	21	1	Ensino Médio
9	2	Ensino Médio	22	2	Ensino Superior
10	1	Ensino Superior	23	0	Ensino Fundamental
11	3	Ensino Médio	24	4	Pós-graduação
12	0	Pós-graduação	25	1	Ensino Médio
13	2	Ensino Superior			

Fonte: Produção da própria autora (2024).

Tabela 4 – Nº de Filhos dos funcionários

Nº de Filhos	Frequência
0	5
1	6
2	7
3	4
4	3

Fonte: Produção da própria autora (2024).

Tabela 5 – Escolaridade dos funcionários

Escolaridade	Frequência
Ensino Fundamental	4
Ensino Médio	8
Ensino Superior	7
Pós-graduação	6

Fonte: Produção da própria autora (2024).

Tabela 6 – Nº de Filhos dos funcionários - 2

i	Nº de filhos	f_i	$f_r(\%)$	$f_{ra}(\%)$
1	0	5	20	20
2	1	6	24	44
3	2	7	28	72
4	3	4	16	88
5	4	3	12	100
		25	100	

Fonte: Produção da própria autora (2024).

Com base nas Tabelas 6 e 7, nota-se que o número de filhos entre os funcionários varia de 0 a 4, sendo que a maioria possui 1 ou 2 filhos, enquanto 20% não têm filhos. Em relação à escolaridade, há uma distribuição equilibrada entre os diferentes níveis educacionais, refletindo a diversidade do corpo de funcionários. A maior parte dos colaboradores possui Ensino Médio (32%), seguido por Ensino Superior (28%), Pós-graduação (24%) e Ensino Fundamental (16%), o que aponta para um nível educacional relativamente elevado na maioria dos casos. Além disso, aparentemente não há uma relação direta entre o número de filhos e o nível de escolaridade. Por

Tabela 7 – Escolaridade dos funcionários - 2

i	**Escolaridade**	f_i	$f_r(\%)$	$f_{ra}(\%)$
1	Ensino Fundamental	4	16	16
2	Ensino Médio	8	32	48
3	Ensino Superior	7	28	76
4	Pós-graduação	6	24	100
		25	100	

Fonte: Produção da própria autora (2024).

exemplo, funcionários com Ensino Fundamental podem ter de 0 a 3 filhos, enquanto aqueles com Pós-graduação apresentam uma variação de 0 a 4 filhos. Esses dados indicam que o nível de escolaridade não influencia diretamente o tamanho das famílias. Em geral, o perfil dos funcionários mostra famílias pequenas e uma boa formação acadêmica, o que sugere que a empresa pode criar programas de apoio que valorizem essas características, além de explorar o potencial para investir em formação continuada, considerando a quantidade expressiva de funcionários com Ensino Superior e Pós-graduação.

A tabela de distribuição de frequências é vantajosa em situações onde há necessidade de organizar e resumir dados para facilitar a análise e interpretação. Por exemplo: quando os dados coletados são em grande quantidade e se repetem; quando pretende-se comparar diferentes grupos ou categorias de forma clara; facilidade no cálculo de medidas estatísticas, dentre outras.

Observações: Quando os dados são quantitativos ou qualitativos ordinais, eles devem ser organizados nas classes (i) em ordem crescente ou decrescente. A distribuição de frequências oferece um resumo significativo dos dados, sem que haja perda de informação ao aplicar essa técnica. A frequência de ocorrência dos valores em cada classe torna-se clara. Esse tipo de distribuição de frequências é utilizado no caso discreto quando o número de classes é inferior ao total da amostra.

4. **Tabela de distribuição de frequência de dados agrupados** - Organiza os valores em intervalos de classe (i) , que representam faixas numéricas. Para cada intervalo, calcula-se a frequência de observações que pertencem a essa faixa. Essa forma de organização é especialmente útil para resumir grandes volumes de dados, destacando padrões e tendências de maneira clara e visual.

   Para construir uma tabela eficaz, é essencial determinar corretamente o número de classes e a amplitude de cada intervalo. Intervalos muito amplos podem obscurecer detalhes importantes, enquanto intervalos muito estreitos podem gerar confusão. Algumas regras práticas ajudam a definir a quantidade de classes:

   - Uma abordagem amplamente utilizada é a **regra de Sturges**:

$$k \approx 1 + 3,322 \log_{10} n, \tag{3.1}$$

onde k é o número de classes e n é o número total de observações.

Observação: Originalmente, a regra de Sturges utiliza o logaritmo na base 2:

$$k \approx 1 + \log_2 n.$$

Isso se baseia na ideia de que o número de classes deve crescer proporcionalmente ao número de vezes que os dados podem ser divididos em duas partes (divisões binárias).

Para utilizar o logaritmo na base 10, é aplicada a mudança de base do logaritmo:

$$\log_2 n = \frac{\log_{10} n}{\log_{10} 2}.$$

Sabendo que $\log_{10} 2 \approx 0,3010$, temos:

$$\log_2 n \approx \frac{\log_{10} n}{0,3010} \approx 3,322 \log_{10} n.$$

Substituindo na fórmula original:

$$k \approx 1 + \log_2 n \approx 1 + 3,322 \log_{10} n,$$

que corresponde à fórmula amplamente utilizada e expressa na equação (3.1).

Essa adaptação facilita os cálculos em contextos práticos, já que o logaritmo decimal é mais comumente empregado em ferramentas manuais e softwares básicos.

A regra de Sturges é eficaz para amostras de tamanho moderado, balanceando a quantidade de classes para evitar a perda de detalhes (com poucas classes) ou fragmentação excessiva (com muitas classes). Entretanto, para amostras muito grandes ou muito pequenas, outras regras podem ser mais adequadas.

- A quantidade de classes também pode ser estimada pela **raiz quadrada** do número total de observações:

$$k \approx \sqrt{n}, \tag{3.2}$$

onde k é o número de classes e n é o número total de observações. Essa técnica é prática e eficaz para a maioria das análises.

Além de determinar k, também é necessário calcular:

- A amplitude total (H) da amostra dado por:

$$H = Maior\ valor\ observado - Menor\ valor\ observado. \tag{3.3}$$

- A amplitude da classe (h), medida que define a largura de cada uma das classes da tabela de frequência é dada por:

$$h = \frac{H}{k}. \tag{3.4}$$

  Observação: O valor de h deve ser arredondado para cima (em unidade, décimos, centésimos, etc.) para garantir que o maior valor observado esteja incluído na última classe.

- Os limites de classes, correspondem aos valores extremos que definem o intervalo de cada classe em uma tabela de frequência. O limite mais baixo é denominado limite inferior, indicado por "l_i", enquanto o limite mais alto é conhecido como limite superior, indicado com letra Maiúscula "L_i" de cada classe. O limite inferior da primeira classe é igual ao menor valor observado. O limite superior da primeira classe é calculado como:

$$L_1 = l_1 + h.$$

  O limite inferior da classe seguinte será igual ao limite superior da classe anterior,

$$L_{i+1} = l_i + h \tag{3.5}$$

  e assim sucessivamente.

  Observação: Em cada classe utilizamos a notação: $l_i \mid - L_i$. O símbolo $\mid -$ indica que, ao contabilizar a frequência dos valores incluídos no intervalo da classe, serão considerados todos os valores que são maiores do que ou iguais a l_i e menores que L_i.

Exemplo 3.4.5. *Uma empresa com 500 funcionários deseja realizar um levantamento sobre a Idade e o Tempo de Serviço em anos completos dos mesmos. Uma amostra de 25 funcionários forneceu as informações aos quais estão apresentadas na Tabela 8. Essa tabela exibe os dados brutos coletados durante a pesquisa.*

Tabela 8 – Idade e tempo de serviço dos funcionários

Funcionário	Idade	Tempo de Serviço
1	28	2
2	34	5
3	45	10
4	31	3
5	39	7
6	50	15
7	29	1
8	36	6
9	42	12
10	38	8
11	27	2
12	44	14
13	33	4
14	41	11
15	30	3
16	37	9
17	26	1
18	48	16
19	35	5
20	40	12
21	32	4
22	46	13
23	25	0
24	43	11
25	49	17

Fonte: Produção da própria autora (2024).

Com essas informações, a empresa pretende analisar a relação entre a idade dos funcionários e o tempo que eles estão na empresa, buscando entender como esses fatores podem influenciar na produtividade e satisfação no trabalho.

Para realizar essa análise, iremos elaborar tabelas de dados agrupados para as variáveis idade e tempo de serviço, utilizando intervalos iguais para os colaboradores da empresa. Para isso, determinaremos:

a) A quantidade de classes utilizando a ***regra de Sturges*** *(fórmula 3.1), onde $n = 25$ representa o número total de observações:*

$$k \approx 1 + 3,322 \log_{10} 25$$

$$\approx 1{+}3,322 \cdot 1,398 \approx 5,504$$

$$Portanto,\ k = 6.$$

b) A amplitude total (3.3) das variáveis idade e tempo de serviço, onde os valores *mínimo e máximo observados para a idade são* 25 *e* 50*, respectivamente, e para o tempo de serviço são* 1 *e* 17*:*

Amplitude Total das Variáveis

Idade	*Tempo de Serviço*
$H = 50 - 25$	$H = 17 - 1$
$H = 25$	$H = 16$

c) a amplitude da classe (3.4) dada por:

Amplitude da classe das Variáveis

Idade	*Tempo de Serviço*
$h = \frac{25}{6}$	$h = \frac{16}{6}$
$h \approx 4,2$	$h \approx 2,7$
$h = 5$	$h = 3$

As tabelas 9 e 10 foram elaboradas com as classes correspondentes, especificando os limites inferiores e superiores de cada uma, para facilitar a análise e interpretação dos dados.

Tabela 9 – Distribuição de freq. da variável idade

i	**Idade**	f_i	$f_r(\%)$	$f_{ra}(\%)$
1	25 |- 30	5	20	20
2	30 |- 35	5	20	40
3	35 |- 40	5	20	60
4	40 |- 45	5	20	80
5	45 |- 50	4	16	96
6	50 |- 55	1	4	100
		25	100	

Fonte: Produção da própria autora (2024).

Observa-se que os funcionários estão distribuídos em faixas etárias regulares, com maior concentração entre 25 e 50 anos (96% dos funcionários). Este dado sugere que a força de trabalho é predominantemente jovem ou de meia-idade, o que pode ser vantajoso para a empresa em termos de energia e flexibilidade. Apenas 4% dos funcionários têm entre 50 e 55 anos, indicando uma presença limitada de profissionais mais experientes em idade.

Constatamos que a maioria dos funcionários possui até 6 anos de experiência na empresa, representando 44% do total acumulado. Essa informação pode indicar um

Tabela 10 – Distribuição de freq. da variável tempo de serviço

i	**Tempo de Serviço**	f_i	$f_r(\%)$	$f_{ra}(\%)$
1	0 |- 3	5	20	20
2	3 |- 6	6	24	44
3	6 |- 9	3	12	56
4	9 |- 12	4	16	72
5	12 |- 15	4	16	88
6	15 |- 18	3	12	100
		25	100	

Fonte: Produção da própria autora (2024).

recente aumento nas contratações ou uma rotatividade significativa. Os intervalos intermediários, que abrangem de 6 a 12 anos de serviço, correspondem a 28% dos funcionários, enquanto aqueles com mais de 12 anos de experiência também representam 28%, evidenciando uma base sólida de colaboradores com experiência acumulada na organização.

Ao cruzar os dados das tabelas, nota-se uma concentração de funcionários mais jovens com menos tempo de serviço, indicando uma possível renovação ou expansão da equipe. Por outro lado, a presença equilibrada de colaboradores experientes nas faixas etárias mais altas sugere uma retenção eficaz de talentos com maior vivência na empresa.

Esses resultados fornecem uma base sólida para que a empresa desenvolva estratégias alinhadas ao perfil de sua força de trabalho, como políticas de retenção, programas de treinamento e desenvolvimento, além de iniciativas específicas para atender às demandas de diferentes grupos etários e de experiência, visando aumentar a produtividade e a satisfação no trabalho.

Observação: É importante destacar algumas características fundamentais das tabelas de frequência. Diferente de outros tipos de tabelas, elas não apresentam delimitação externa com traços verticais. Em vez disso, sua estrutura é delimitada por traços horizontais que ajudam a organizar e separar as informações de forma clara.

Conforme estabelecido pelo Instituto Brasileiro de Geografia e Estatística - IBGE (1993), uma tabela deve seguir uma estrutura padrão que inclui os seguintes elementos:

- **Número -** Identifica a tabela de maneira única dentro do documento, facilitando a referência cruzada.
- **Título -** Descreve de forma breve e clara o conteúdo ou propósito da tabela, permitindo ao leitor entender imediatamente o tipo de informação apresentada.

- **Corpo -** É a parte principal da tabela, onde os dados estão organizados. O corpo é subdividido em:
  a) Cabeçalho - Localizado na parte superior do corpo, o cabeçalho indica o conteúdo de cada coluna, oferecendo um guia para interpretação dos dados.
  b) Coluna indicadora - Geralmente localizada à esquerda, esta coluna apresenta os itens ou categorias analisadas na tabela, como variáveis ou intervalos de classe.
  c) Linhas - As linhas horizontais contêm os dados correspondentes às categorias indicadas, seguindo o formato especificado no cabeçalho.
- **Fonte -** Apresenta a origem dos dados, garantindo a credibilidade da tabela e permitindo a verificação das informações quando necessário.

Esquematicamente:

Figura 5 – Elementos constitutivos de uma tabela

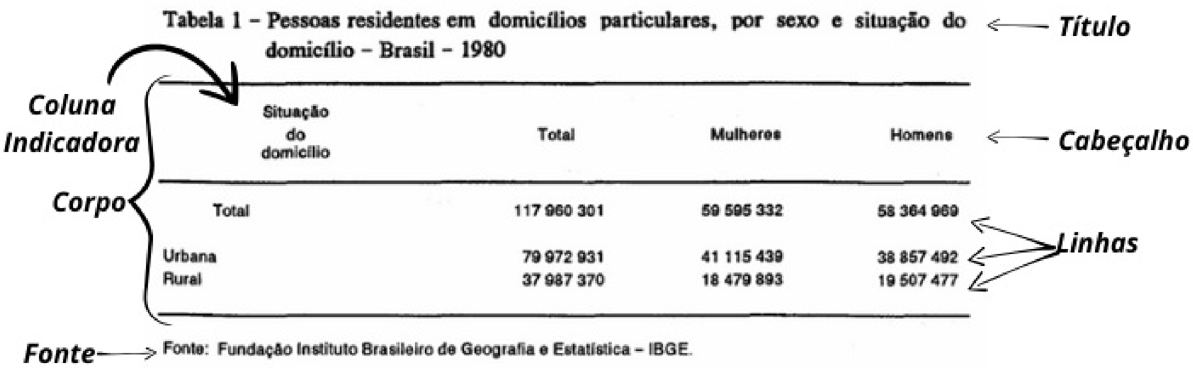


Tabela 1 – Pessoas residentes em domicílios particulares, por sexo e situação do domicílio – Brasil – 1980

Situação do domicílio	Total	Mulheres	Homens
Total	117 960 301	59 595 332	58 364 969
Urbana	79 972 931	41 115 439	38 857 492
Rural	37 987 370	18 479 893	19 507 477

Fonte: Fundação Instituto Brasileiro de Geografia e Estatística – IBGE.

Fonte: Adaptado de (Instituto Brasileiro de Geografia e Estatística - IBGE, 1993)

Seguindo essas diretrizes, é possível criar tabelas organizadas, claras e alinhadas aos padrões estatísticos reconhecidos, garantindo a legibilidade e a confiabilidade das informações apresentadas.

5. **Representação Gráfica**

A representação gráfica dos dados estatísticos de uma pesquisa desempenha um papel fundamental na comunicação dos resultados, pois permite uma visualização mais clara e objetiva das informações. Esse tipo de abordagem facilita a identificação de padrões, tendências e variações que podem ser difíceis de perceber apenas com tabelas ou textos descritivos. Além disso, a análise gráfica ajuda a simplificar a compreensão dos dados, tornando-os mais acessíveis para diferentes públicos, desde especialistas até leigos.

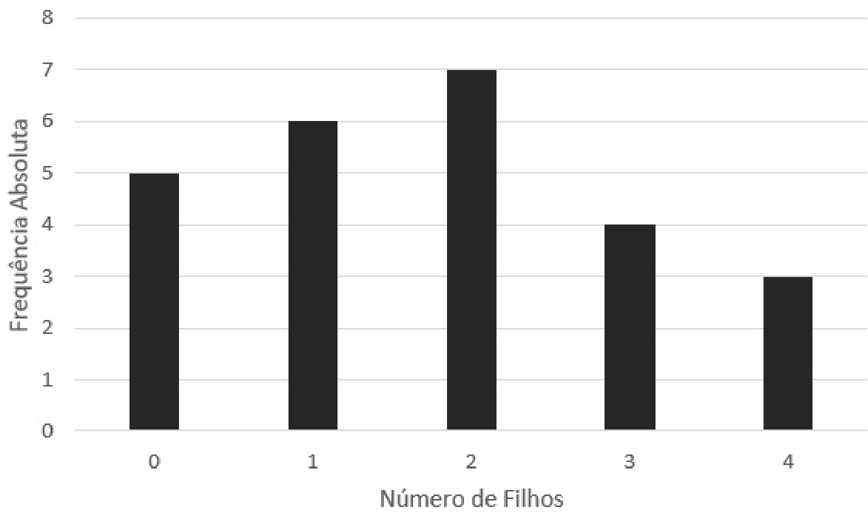
Existem diversos tipos de gráficos que podem ser escolhidos com base no tipo de variável e no objetivo da análise. É comum observar, em meios de comunicação, como mídias, jornais e livros, o uso frequente de alguns tipos específicos de gráficos, cada qual com suas particularidades e propósitos. Entre os mais utilizados, destacam-se:

- **Gráficos de Barras e de Colunas** – São utilizados para comparar valores entre diferentes categorias, representando visualmente a frequência ou ou o valor de cada uma. Para garantir a integridade da representação, é essencial manter proporções consistentes entre a largura das barras (ou colunas) e os espaçamentos, bem como preservar a escala dos eixos, evitando distorções.

  A distinção entre os dois tipos está na orientação: gráficos de colunas apresentam barras verticais, enquanto os de barras dispõem as categorias ao longo do eixo vertical, com as medidas no eixo horizontal. Quando as classes possuem a mesma amplitude, as colunas tendem a ter larguras uniformes, como ilustrado na Figura 6. Já os gráficos de barras, geralmente aplicados a variáveis qualitativas, seguem o mesmo princípio proporcional, conforme mostra a Figura 7.

  Um exemplo específico de gráfico de barras horizontais espelhadas por sexo e faixa etária é a pirâmide etária do município de Nova Venécia (ES), apresentada na Figura 8, amplamente utilizada em análises populacionais.

Figura 6 – Gráfico em Colunas correspondente a Tabela 6 do Ex. 3.4.4.

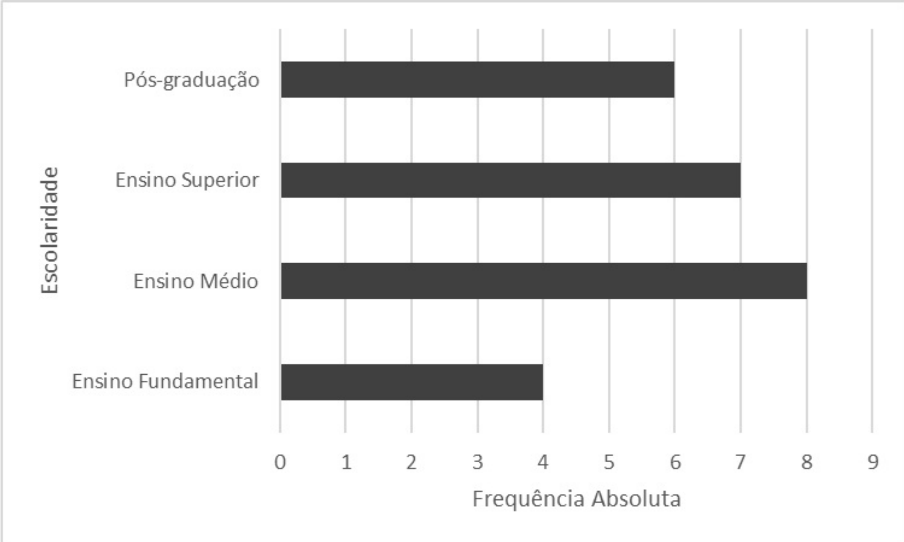


Fonte: Produção da própria autora (2024).

- **Gráfico de Setores (Gráfico de Pizza)** - é utilizado para representar as proporções de um todo, destacando a contribuição de cada classe em relação ao total, sendo frequentemente empregado para ilustrar dados em forma de porcentagens. Para que esse gráfico seja eficaz, é necessário que a variável compreenda um número reduzido de classes, de modo a evitar a sobrecarga visual e garantir a clareza na interpretação dos dados.

  Na elaboração do gráfico de setores, é fundamental garantir que haja uma proporção adequada entre o ângulo de cada setor e a frequência relativa de

Figura 7 – Gráfico de Barras correspondente a Tabela 7 do Ex. 3.4.4.

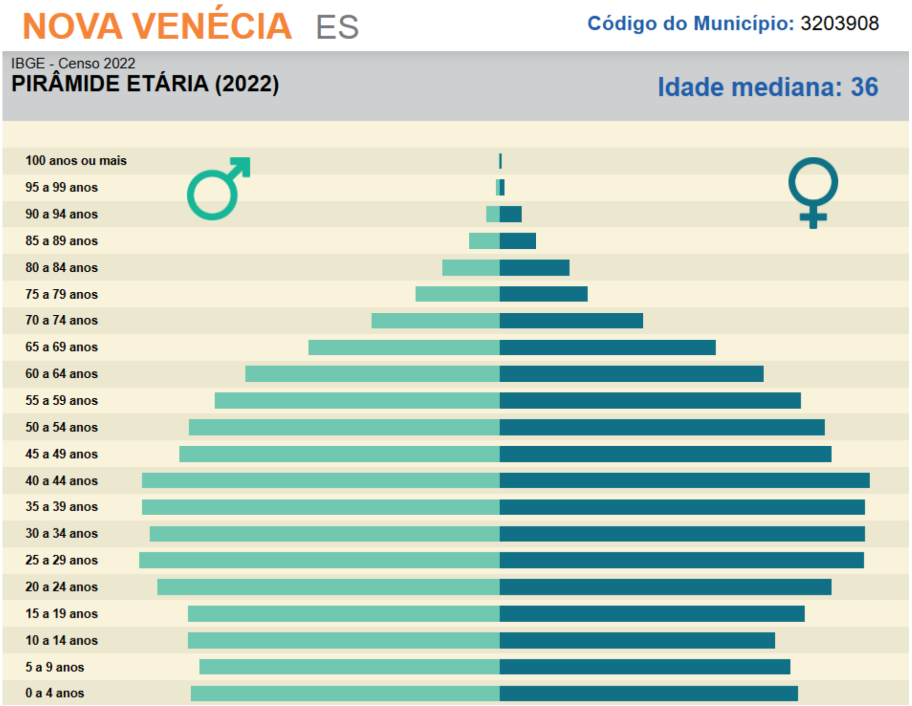


Fonte: Produção da própria autora (2024).

cada classe. Para determinar o ângulo correspondente a cada setor circular, podemos aplicar uma regra de três, como ilustrado abaixo:

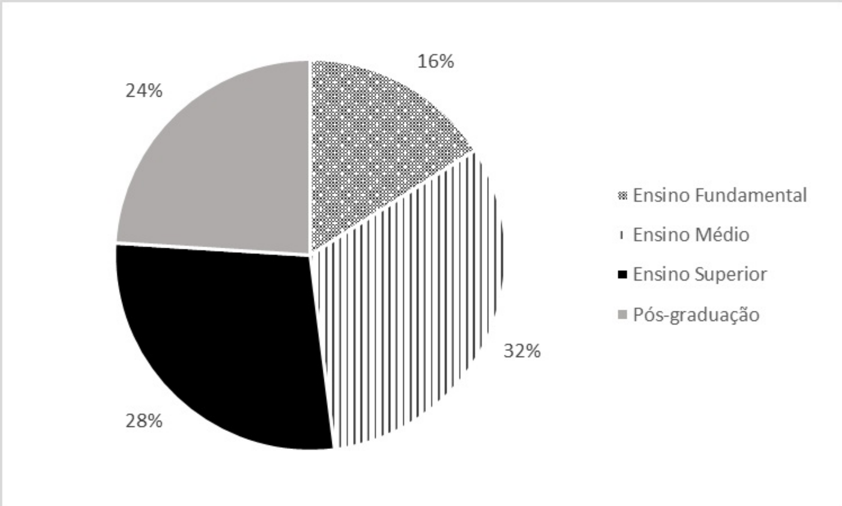
$$\begin{array}{ccc} 360^\circ & \rightarrow & 100\% \\ x^\circ & \rightarrow & f_r \end{array}$$

Figura 8 – Pirâmide etária (2022) do município de Nova Venécia (ES) — gráfico de barras horizontais espelhadas por sexo e faixa etária.



Fonte: Fonte: IBGE, Cidades. Panorama do Município de Nova Venécia (ES), disponível em <https://cidades.ibge.gov.br/panorama/3203906>, acesso em maio de 2025.

Figura 9 – Gráfico de Setores correspondente a Tabela 7 do Ex. 3.4.4.



Fonte: Produção da própria autora (2024).

Assim, temos a fórmula:

$$x^\circ = 3,60 \cdot f_r. \tag{3.6}$$

Com base nessa relação, os ângulos proporcionais para a construção do gráfico apresentado na Figura 9 são: Ensino Fundamental = 57,6°, Ensino Médio = 115,2°, Ensino Superior = 100,8° e Pós-graduação = 86,4°.

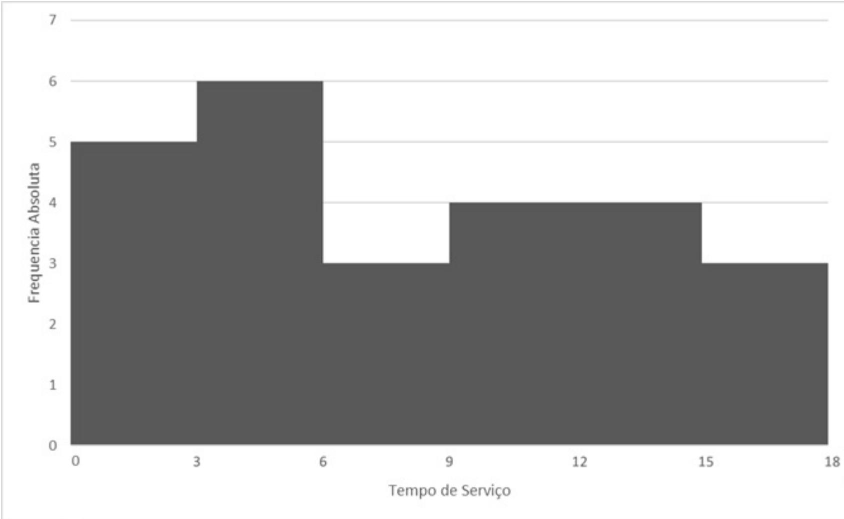
Observação: Os gráficos de barras, colunas e setores são usados principalmente para representar dados de variáveis qualitativas e quantitativas discretas. Eles são mais indicados quando os valores das variáveis não variam muito, facilitando a visualização e comparação das categorias ou grupos analisados.

- **Histograma** - Utilizado para dados quantitativos contínuos, exibindo a distribuição de frequências em intervalos de classes, o que possibilita visualizar como os dados se distribuem ao longo de um eixo. O gráfico da Figura 10 baseia-se nos dados da Tabela 9.

  É fundamental observar que, no histograma, não há distância entre as colunas das classes. Quando o limite inferior da primeira classe não está próximo de zero, utiliza-se um símbolo específico (frequentemente um pequeno zig-zag ou uma linha inclinada) no gráfico para indicar uma quebra de escala. Isso é feito no eixo das classes para sinalizar que o intervalo da classe está sendo comprimido, mantendo a clareza do gráfico sem distorcer visualmente a distribuição dos dados.

  Essa representação gráfica é de grande utilidade para analisar a variabilidade dos dados e as medidas de tendência central, além de contribuir para a identificação

Figura 10 – Histograma correspondente a Tabela 9 do Ex. 3.4.5



Fonte: Produção da própria autora (2024).

de padrões como normalidade, simetria ou assimetria na distribuição dos dados.

- **Gráfico Polígono de Frequência** - Uma variação do histograma, onde os pontos médios no topo das barras do histograma são conectados por uma linha contínua. Esse gráfico é útil para observar a tendência geral dos dados, oferecendo uma representação mais fluida e simplificada, o que facilita a interpretação de grandes volumes de dados e fornece uma visão clara da distribuição, sem as distrações visuais das barras.

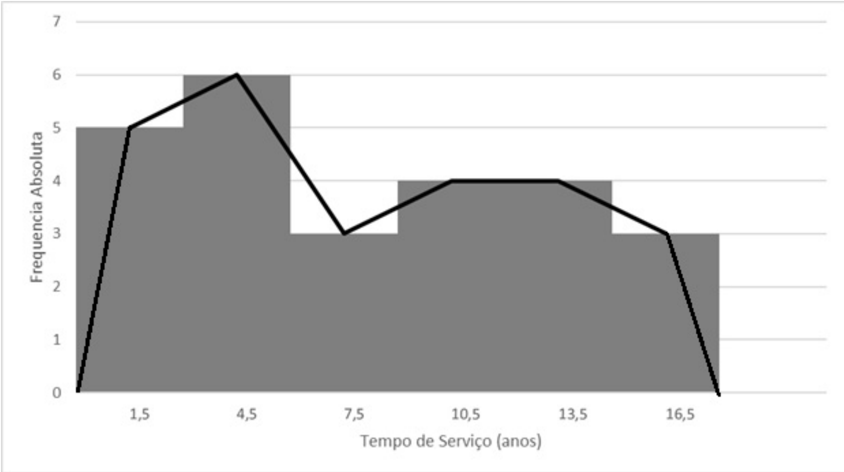
  Esse tipo de gráfico é mais adequado para dados contínuos ou intervalares, mas pode ser menos eficaz para dados discretos com poucas categorias. Além disso, o polígono de frequência pode não destacar variações locais nos dados, que seriam mais evidentes nas barras do histograma.

  Na sua construção, o gráfico polígono de frequência utiliza os mesmos dados do histograma, mas, em vez de representar as frequências por meio de barras, utiliza uma linha. O eixo horizontal (abscissas) representa as classes, com os pontos médios de cada classe sendo posicionados ao longo desse eixo, enquanto o eixo vertical (ordenadas) mostra as frequências correspondentes.

  Vale ressaltar que, como mostrado nos gráficos das Figuras 11 e 12, ambos se iniciam no limite inferior da primeira classe no eixo horizontal e terminam no limite superior da última classe, apresentando frequência nula em ambos os casos no eixo vertical.

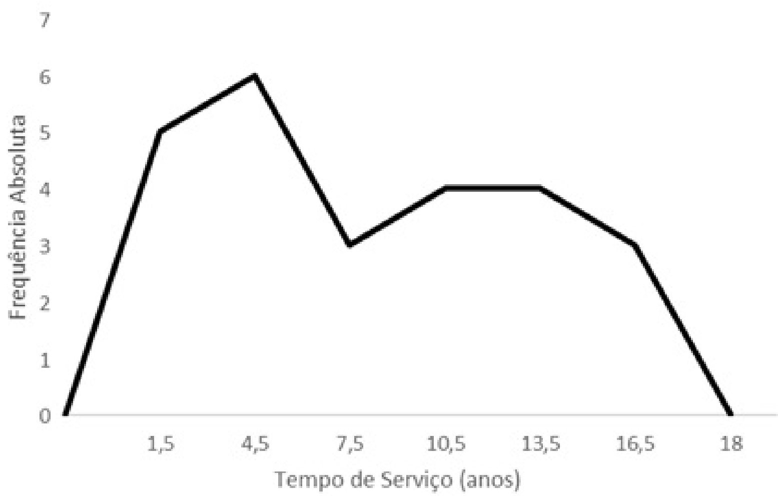
- **Pictograma** - Esse tipo de gráfico utiliza figuras ou ícones para representar a frequência de uma variável, sendo cada ícone correspondente a uma quantidade específica de unidades. Essa forma visual facilita a compreensão dos dados, tornando-os mais intuitivos e atraentes para o público. A Figura 13 exemplifica

Figura 11 – Polígono de Frequência e Histograma correspondente a Tabela 9 do Ex. 3.4.5



Fonte: Produção da própria autora (2024).

Figura 12 – Polígono de Frequência correspondente a Tabela 9 do Ex. 3.4.5



Fonte: Produção da própria autora (2024).

um pictograma elaborado a partir dos dados da Tabela 6 do Exemplo 3.4.4.

Além disso, pictogramas são amplamente empregados em contextos informativos e educativos, como ilustrado na Figura 14 que apresenta a distribuição da população por cor ou raça no município de Nova Venécia (ES), conforme dados do Censo 2022 do IBGE. Esses gráficos são especialmente eficazes para variáveis qualitativas (nominais ou ordinais) e variáveis quantitativas discretas, nas quais os dados podem ser agrupados em unidades representadas por ícones.

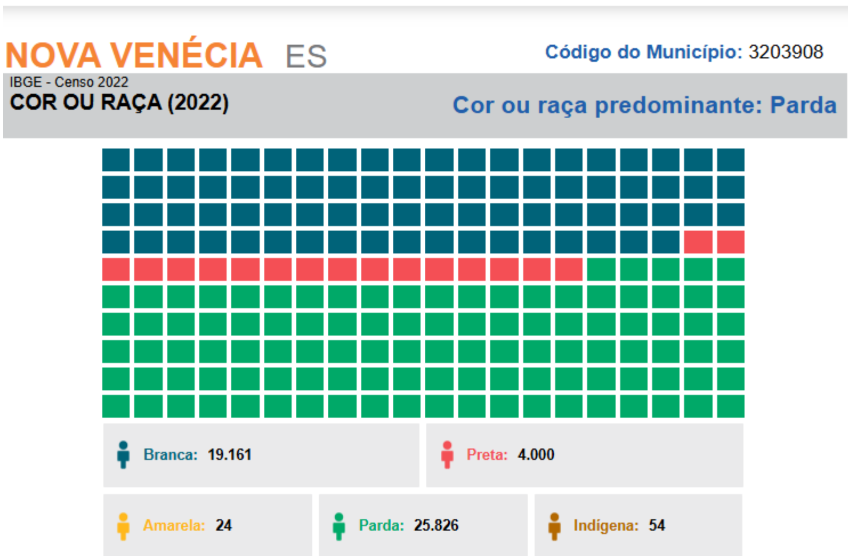
Por outro lado, pictogramas não são recomendados para variáveis contínuas, pois a representação proporcional por imagens torna-se pouco prática.

- **Gráfico de Dispersão** - Representa a relação entre duas variáveis quantitativas. Ele é composto por pontos no plano cartesiano, onde cada ponto reflete um par de valores correspondentes às variáveis analisadas. A disposição dos pontos

Figura 13 – Pictograma correspondente a Tabela 6 do Ex. 3.4.4

Fonte: Produção da própria autora (2024), criado com a ferramenta disponível em <https://www.visme.co/pt-br/>, acesso em novembro 2024.

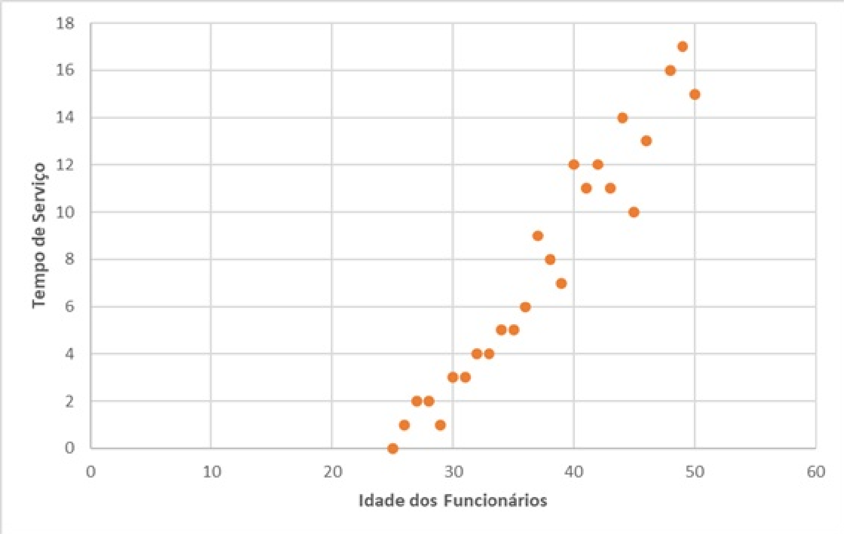
Figura 14 – Distribuição da população por cor ou raça no município de Nova Venécia (ES), conforme Censo 2022



Fonte: Fonte: IBGE, Cidades. Panorama do Município de Nova Venécia(ES), disponível em <https://cidades.ibge.gov.br/panorama/3203908>, acesso em maio de 2025.

conforme os valores das variáveis torna visível a relação entre elas, o que é útil para identificar padrões e correlações. Quando há uma correlação positiva, os pontos se alinham em uma linha ascendente, indicando que o aumento de uma variável está relacionado ao aumento da outra. Por outro lado, uma correlação negativa se manifesta em uma linha descendente, sugerindo que, à medida que uma variável cresce, a outra diminui. Assim, o gráfico de dispersão é útil para analisar variações, identificar valores extremos e entender as interações entre as variáveis, como exemplificado na Figura 15, que apresenta os dados da Tabela 8.

Figura 15 – Gráfico de dispersão correspondente a Tabela 8 do Ex. 3.4.5



Fonte: Produção da própria autora (2024).

Além disso, o gráfico de dispersão é eficaz na detecção de outliers ou valores atípicos, que aparecem como pontos fora do padrão. Essa característica é importante para identificar dados que fogem ao comportamento esperado, o que pode apontar para erros na coleta de dados ou a necessidade de considerar outras variáveis no modelo. Pela sua simplicidade e clareza, o gráfico de dispersão é amplamente utilizado como uma das primeiras ferramentas na análise exploratória de dados em estudos estatísticos e pesquisas científicas.

- **Gráfico Dot Plot** - Representa a distribuição de dados, especialmente quando se trata de dados discretos. Cada ponto no gráfico representa um valor individual do conjunto de dados, e pontos com o mesmo valor são sobrepostos, permitindo observar rapidamente a frequência de cada valor. A Figura 16 ilustra um Dot Plot baseado nos dados da Tabela 8.

Figura 16 – Gráfico Dot Plot correspondente a Tabela 8 do Ex. 3.4.5

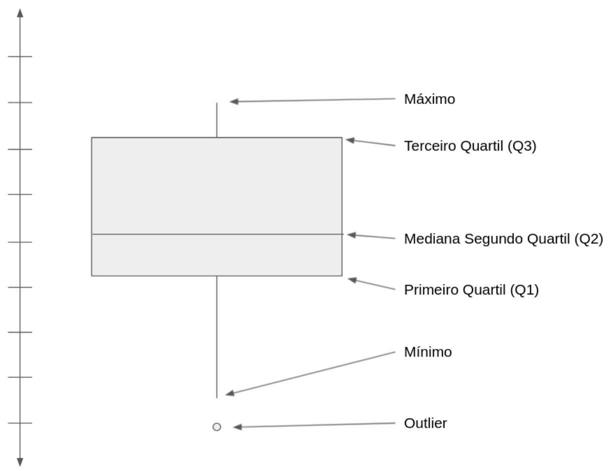
Fonte: Produção da própria autora (2024).

Esse tipo de gráfico é eficaz para identificar padrões, como a concentração de valores em determinadas faixas, além de destacar possíveis outliers. Ele é particularmente útil para pequenas quantidades de dados, proporcionando uma visão clara da distribuição e frequências de valores.

- **Gráfico Boxplot** - Também conhecido como “diagrama de caixas”, é um gráfico utilizado para mostrar a distribuição de uma variável numérica, ou até ordinal, facilitando a visualização da forma como os dados estão organizados. Ele é útil para comparar variáveis quantitativas com qualitativas, destacando padrões, dispersões e até valores extremos.

  No gráfico de Boxplot, são representados os “cinco números principais” que ajudam a resumir os dados: o mínimo, o primeiro quartil (Q1), a mediana (Q2), o terceiro quartil (Q3) e o máximo, conforme a Figura 17.

  Figura 17 – Estrutura do Boxplot

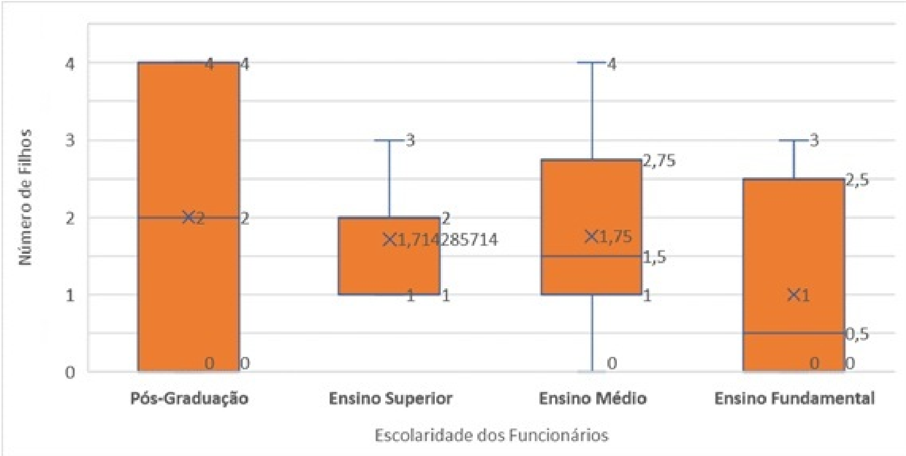


  Fonte: Produção da própria autora (2024).

  – O mínimo representa o menor valor do conjunto de dados dentro de um limite calculado, embora valores abaixo desse limite, conhecidos como outliers, possam ser identificados.
  – O primeiro quartil (Q1) indica que 25% dos dados estão abaixo dele, enquanto 75% dos dados estão acima.
  – A mediana (Q2), que aparece como uma linha no centro da caixa, divide os dados em duas metades iguais, com 50% dos valores abaixo e 50% acima.
  – O terceiro quartil (Q3), no topo da caixa, separa os 75% dos dados mais baixos dos 25% mais altos.
  – E o máximo, por sua vez, é o maior valor dentro de um limite calculado, mas, assim como o mínimo, também pode haver outliers acima dele.

  O conceito de quartis, bem como seus métodos de cálculo, será aprofundado na Seção 3.8.2, onde abordaremos as medidas separatrizes com mais detalhes. A Figura 18 apresenta o Gráfico Boxplot elaborado com base nos dados da Tabela 3, discutidos no Exemplo 3.4.4.

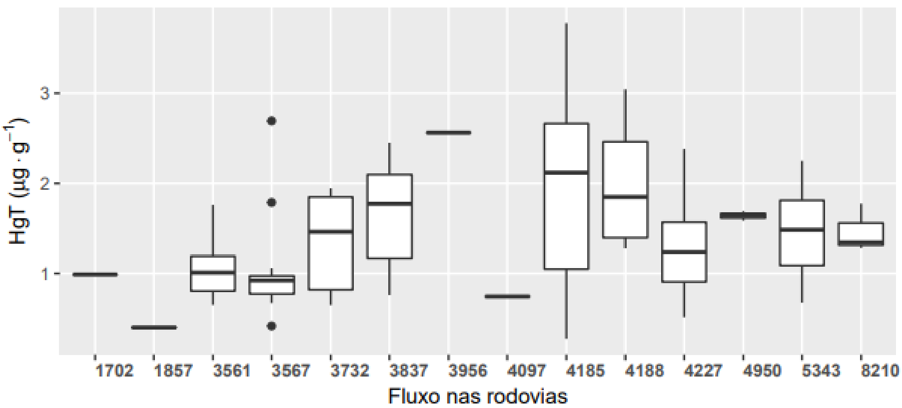
Figura 18 – Gráfico Boxplot correspondente a Tabela 3 do Ex. 3.4.4



Fonte: Produção da própria autora (2024).

Como exemplo prático da aplicação do gráfico boxplot em estudos científicos, apresentamos na Figura 19 um gráfico extraído da dissertação de Carvalho (2024)[3]. Esse boxplot representa as concentrações de HgT nos pelos de tamanduá-bandeira, categorizadas conforme o Volume Médio Diário Anual (VMDA) de automóveis nas rodovias analisadas. Valores considerados outliers — ou seja, superiores a 1,5 vezes o intervalo interquartil — são indicados como pontos isolados.

Figura 19 – Gráfico boxplot representando as concentrações de HgT em pelos de tamanduá-bandeira, categorizadas pelo Volume Médio Diário Anual (VMDA) de automóveis nas rodovias



Fonte: Adaptado de Carvalho (2018).

Esse exemplo demonstra a utilidade do boxplot para analisar e comparar a distribuição de variáveis ambientais em diferentes categorias, permitindo a identificação de padrões e valores atípicos. Embora o Boxplot seja uma

[3] Disponível em: <file:///C:/Users/dsluz/Downloads/GOCarvalho-Dissertacao-tamanduamercurio.pdf>

excelente ferramenta para análises detalhadas em estudos acadêmicos e em equipes de dados, ele não é o gráfico mais indicado para públicos leigos, pois a sua interpretação exige algum conhecimento prévio, o que o torna menos eficaz em contextos como divulgação científica ou matérias jornalísticas, onde a simplicidade e clareza são mais valorizadas.

Cada um desses gráficos possui características próprias que o tornam mais ou menos adequado dependendo do objetivo da análise e do público-alvo. A escolha correta do tipo de gráfico é essencial para garantir uma comunicação eficiente e a transmissão clara das informações.

3.5 Seleção de amostras: Técnicas de Amostragem e suas aplicações

Conforme abordado na Seção 3.2 deste estudo, a **amostragem** é definida como o procedimento utilizado para obter uma amostra a partir de uma população específica. Este processo deve ser realizado de maneira a assegurar uma adequada **representatividade** do conjunto populacional de interesse para a pesquisa. Em outras palavras, a amostra precisa refletir as características relevantes da população estudada. A qualidade e a precisão das conclusões tiradas de uma amostra dependem, em grande parte, da forma como ela é selecionada. Quando maior o conhecimento sobre a população, maior será a chance de a amostra refletir de maneira fiel suas propriedades. Por exemplo, a análise da quantidade de glóbulos brancos obtida de algumas gotas de sangue da ponta do dedo de um paciente pode fornecer uma boa estimativa sobre a quantidade desses glóbulos no corpo todo, pois sabemos que a distribuição dos glóbulos brancos é homogênea e qualquer amostra retirada seria representativa da população total de glóbulos brancos. No entanto, nem sempre essa escolha de amostra é tão simples. Em muitas situações, a seleção de uma amostra adequada exige atenção a detalhes específicos, já que a população pode não ser homogênea ou pode haver fatores que influenciem a representatividade.

Outro aspecto fundamental na seleção de amostras é garantir a imparcialidade no processo de escolha, de modo que todos os elementos da população tenham a mesma probabilidade de serem incluídos na amostra. Para que as conclusões extraídas sejam válidas, é essencial que o processo de amostragem não favoreça nem exclua grupos específicos da população. Para ilustrar esses conceitos, são apresentados a seguir alguns exemplos.

Exemplo 3.5.1. *Considere que se deseja avaliar o índice de aprovação de um novo medicamento em uma população de pacientes com uma doença específica.*

Pacientes de diversas faixas etárias e níveis de gravidade da doença devem ter a mesma probabilidade de serem selecionados para garantir a representatividade da amostra. Caso contrário, os resultados podem não refletir adequadamente o efeito do medicamento na população como um todo.

Exemplo 3.5.2. *Suponha que se queira investigar o tempo médio de deslocamento dos trabalhadores até o centro de uma grande cidade.*

Trabalhadores de diferentes bairros, com variações nas distâncias e nos meios de transporte utilizados, devem ter as mesmas chances de ser incluídos na amostra para assegurar que ela represente as diferentes condições de deslocamento da população.

Exemplo 3.5.3. *As pesquisas eleitorais são ferramentas essenciais para entender o comportamento do eleitorado durante o período das eleições. Elas fornecem uma estimativa de como os eleitores provavelmente votarão, ajudando a prever os resultados e a moldar estratégias de campanha. No entanto, os resultados dessas pesquisas podem variar em relação ao resultado final das eleições, o que destaca a importância de uma amostragem adequada.*

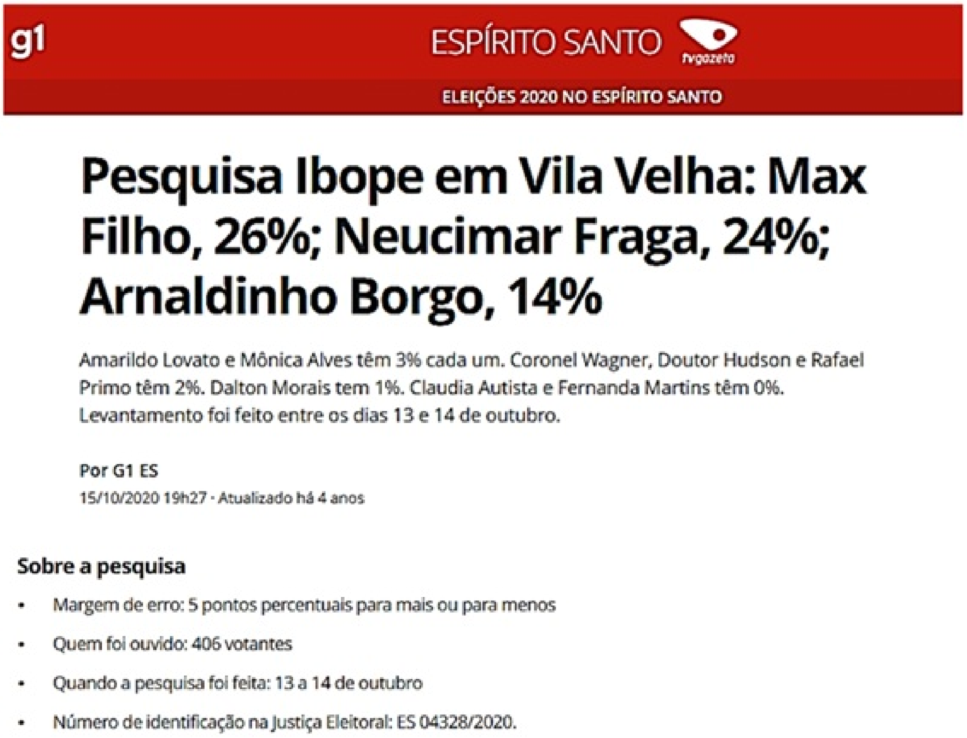
Uma amostragem mal feita ou inadequada pode levar a resultados distorcidos, afetando a confiabilidade da pesquisa. O exemplo das eleições municipais de 2020 em Vila Velha, no Espírito Santo, é um ótimo caso para analisar como a amostragem pode influenciar o resultado das previsões eleitorais. Conforme mostra a pesquisa Ibope realizada em outubro de 2020 (Figura 20), os dois principais candidatos eram dos partidos PSDB, com 26% das intenções de voto, e PSD, com 24%. O candidato do partido PODE estava significativamente atrás, com apenas 14%. Esses números sugeriam uma disputa acirrada entre os candidato dos partidos PSDB e PSD, com uma pequena vantagem para o candidato do PSDB.

No entanto, os resultados oficiais das urnas, após o término da apuração, revelaram uma reviravolta onde o candidato do partido PODE, o terceiro colocado na pesquisa, venceu as eleições (Figura 21). Essa mudança no resultado final em comparação com a pesquisa inicial destacou a complexidade da dinâmica eleitoral e a importância da precisão na amostragem.

Esse cenário evidencia como a amostragem e a metodologia da pesquisa podem impactar os resultados e gerar distorções. A pesquisa de intenção de voto, embora realizada com rigor, não conseguiu prever com precisão o comportamento do eleitorado, e o vencedor final ficou abaixo das intenções de voto que ele tinha na pesquisa. Isso reflete a importância de uma amostragem representativa, que deve considerar não apenas o número de entrevistados, mas também sua diversidade, distribuição geográfica e características demográficas.

Esse exemplo também ressalta que as pesquisas eleitorais devem ser interpretadas com cautela, pois o comportamento do eleitor pode mudar até o último momento. Pequenos erros de amostragem ou flutuações nas preferências dos eleitores podem levar a resultados distintos daqueles esperados. A vitória do candidato do partido PODE é um lembrete de como as dinâmicas eleitorais podem ser complexas e como a amostragem precisa ser

Figura 20 – Pesquisa Ibope Vila Velha-ES



g1 ESPÍRITO SANTO

ELEIÇÕES 2020 NO ESPÍRITO SANTO

Pesquisa Ibope em Vila Velha: Max Filho, 26%; Neucimar Fraga, 24%; Arnaldinho Borgo, 14%

Amarildo Lovato e Mônica Alves têm 3% cada um. Coronel Wagner, Doutor Hudson e Rafael Primo têm 2%. Dalton Morais tem 1%. Claudia Autista e Fernanda Martins têm 0%. Levantamento foi feito entre os dias 13 e 14 de outubro.

Por G1 ES

15/10/2020 19h27 · Atualizado há 4 anos

Sobre a pesquisa

- Margem de erro: 5 pontos percentuais para mais ou para menos
- Quem foi ouvido: 406 votantes
- Quando a pesquisa foi feita: 13 a 14 de outubro
- Número de identificação na Justiça Eleitoral: ES 04328/2020.

Fonte: G1. Disponível em: <https://g1.globo.com/es/espirito-santo/eleicoes/2020/noticia/2020/10/15/pesquisa-ibope-em-vila-velha-max-filho-26percent-neucimar-fraga-24percent-arnaldinho-borgo-14percent.ghtml>. Acesso em: 9 mar. 2025.

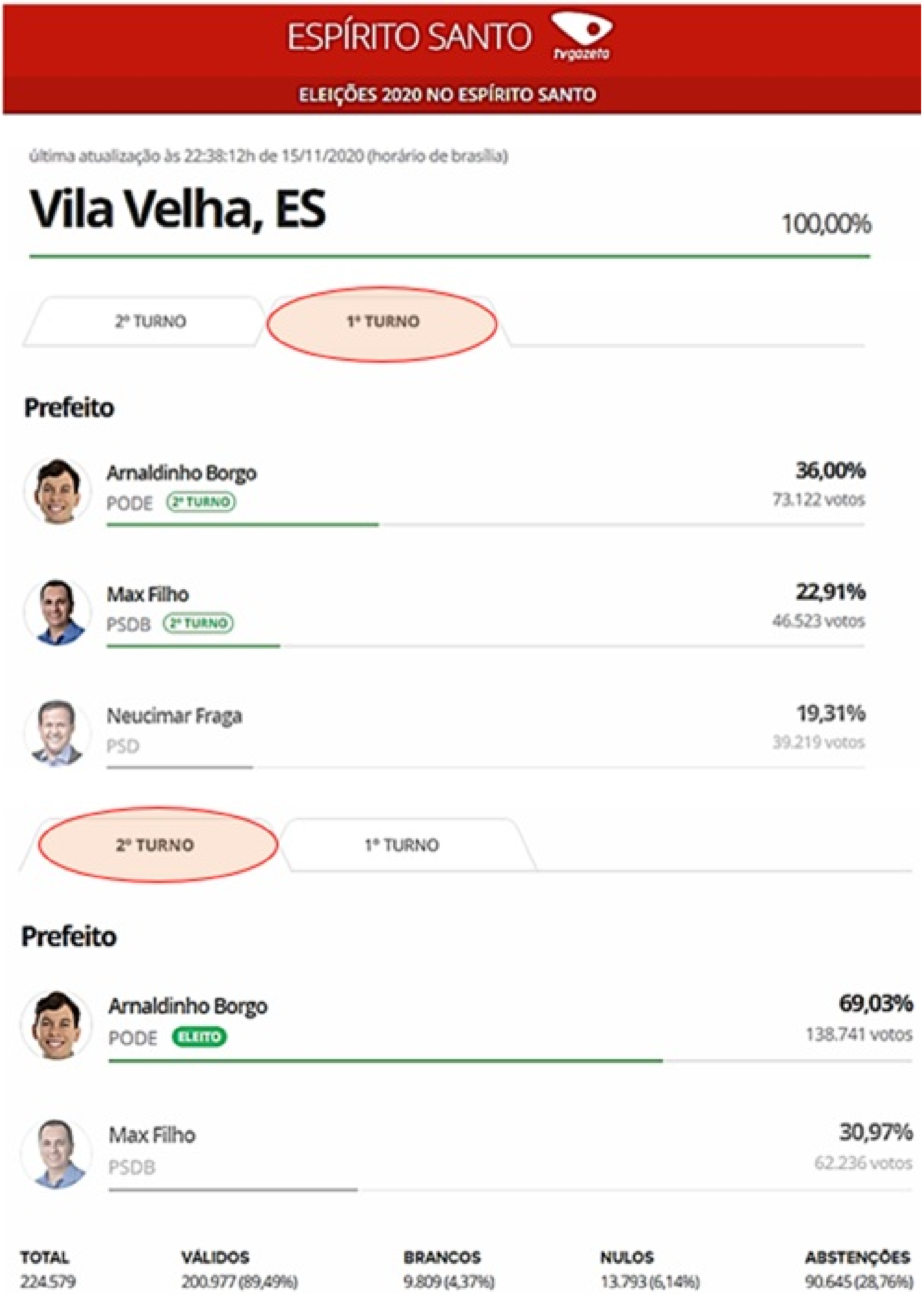
cuidadosamente planejada para refletir com precisão o eleitorado em sua totalidade.

Exemplo 3.5.4. *Suponha que a empresa responsável pelo transporte coletivo de Nova Venécia - ES queira melhorar o transporte público, mas, para isso, precisa entender melhor a demanda de ônibus na cidade. Ou seja, quantas pessoas utilizam o transporte coletivo e em quais horários os ônibus estão mais lotados.*

Para realizar essa análise, a empresa decide conduzir uma pesquisa. O instituto encarregado da coleta de dados opta por realizar a pesquisa apenas durante a manhã, entre 6h e 9h, pois é nesse período que as pessoas costumam ir ao trabalho ou à escola. Durante esse horário, os ônibus ficam superlotados, e muitas vezes os passageiros têm que esperar o próximo ônibus devido à falta de espaço. Com base nesses dados, a empresa conclui que, para resolver o problema da superlotação, seria suficiente aumentar a quantidade de ônibus apenas nesse horário da manhã.

No entanto, essa abordagem apresenta um erro importante. A pesquisa se concentrou unicamente no horário da manhã, sem considerar outros períodos do dia, como o meio-dia, a tarde ou os fins de semana. Após as 9h, os ônibus podem ficar quase vazios, pois a maioria das pessoas já chegou aos seus destinos. Nos fins de semana, a demanda pode ser ainda menor, uma vez que muitas pessoas não precisam utilizar o transporte público para ir ao trabalho.

Figura 21 – Resultado das apurações em Vila Velha-ES



Fonte: G1. Disponível em: <https://g1.globo.com/es/espirito-santo/eleicoes/2020/resultado-das-apuracoes/vila-velha.ghtml>. Acesso em: 9 mar. 2025.

O grande problema aqui é que a pesquisa não capturou a variação da demanda ao longo do dia, o que pode levar a uma visão distorcida da real necessidade de transporte coletivo na cidade. Se a pesquisa tivesse incluído dados de diferentes horários, como no meio da tarde e durante os fins de semana, a cidade teria uma imagem mais completa sobre o fluxo de passageiros. Com isso, as decisões seriam mais assertivas e eficazes.

Com um conjunto de dados mais abrangente, seria possível planejar melhor a *distribuição de ônibus: colocar mais veículos nos horários de pico, como pela manhã, e reduzir a quantidade nos horários de menor movimento, como à tarde ou nos fins de semana. Assim, a empresa poderia otimizar os recursos e garantir um transporte mais eficiente, sem desperdiçar dinheiro e oferecendo mais conforto aos passageiros.*

Portanto, a amostra precisa ser sempre representativa da realidade, e uma boa amostragem deve considerar todos os aspectos importantes para garantir a precisão e a confiabilidade dos resultados. Para alcançar esse objetivo, as técnicas de amostragem desempenham um papel fundamental, assegurando que a amostra seja representativa e imparcial em relação à população estudada, com o propósito de obter dados que possam ser **generalizados para o todo**. Esses métodos de amostragem se dividem em dois grandes grupos: **probabilística** e **não-probabilística**. No caso da amostragem **Probabilística**, é possível realizar inferências sobre a população com base na amostra, pois todos os elementos da população possuem uma probabilidade conhecida e diferente de zero de serem selecionados. Por outro lado, na amostragem **Não-Probabilística**, a seleção dos indivíduos é feita com base na conveniência ou facilidade de acesso, como ao escolher apenas aqueles que estão próximos ou são facilmente acessíveis. Embora essa abordagem seja mais prática e econômica, ela compromete a imparcialidade e a representatividade da amostra, impossibilitando a generalização dos resultados para a população como um todo.

Considerando que o método probabilístico é o mais adequado para garantir a imparcialidade e a generalização dos resultados, exploraremos a seguir as diversas técnicas que este método abrange.

1. **Amostragem Aleatória Simples**

   Na Amostragem Aleatória Simples, uma amostra de tamanho n é selecionada aleatoriamente a partir dos N elementos da população amostral. O procedimento de sorteio ocorre da seguinte maneira:

   a) Inicialmente rotula-se todos os elementos da população;
   b) Posteriormente, um indivíduo é selecionado aleatoriamente entre os N elementos disponíveis.
   c) Em seguida, o segundo indivíduo é sorteado aleatoriamente entre os $(N-1)$ elementos restantes.
   d) Esse processo é repetido sucessivamente até que todos os n indivíduos sejam selecionados.

   Vale ressaltar que esse procedimento ocorre sem reposição, ou seja, cada indivíduo pode ser sorteado apenas uma vez, garantindo que não haja duplicação de elementos na amostra. Nessa técnica a probabilidade de qualquer indivíduo, ou elemento, da população fazer parte da amostra é igual a $\dfrac{n}{N}$.

Algumas possíveis formas de realizar o sorteio aleatório incluem:

- Uso de algoritmos computacionais para geração de números aleatórios;
- Utilização de dados;
- Sorteio de bolinhas numeradas.

Exemplo 3.5.5. *Considere uma população de 10 pés de plantas da mesma espécie (Espécie Z), com a mesma idade e cultivadas sob mesmo solo e as mesmas condições climáticas.*

O objetivo é realizar medições, como a espessura das folhas e o diâmetro dos galhos, em uma amostra de 5 pés de planta $(n = 5)$.

Rotulamos os 10 pés de planta com números de 1 a 10 e, utilizando a Amostragem Aleatória Simples sem reposição, sorteamos 5 números aleatórios entre 1 e 10. Suponha que os números sorteados sejam 2, 4, 6, 8 e 9. Assim, a amostra será composta por esses pés, e suas medições serão realizadas.

Essa abordagem garante que a amostra seja representativa da população, permitindo uma análise precisa do desenvolvimento médio da plantação, com base em características como a espessura das folhas, o diâmetro dos galhos e outros parâmetros relevantes.

2. **Amostragem Estratificada**

Essa técnica de amostragem é indicada para aplicação quando a população da pesquisa é heterogênea, ou seja, quando as características dos indivíduos variam significativamente. Nesse caso, é recomendável subdividir a população em K *estratos homogêneos*, de forma que cada estrato represente uma subpopulação com características mais semelhantes entre seus elementos, com isso a técnica busca garantir que cada estrato seja adequadamente representado na amostra, melhorando a precisão dos resultados. Após a divisão dos estratos ou subgrupos, aplica-se a técnica da amostragem aleatória simples, isso significa que a amostra é extraída de maneira aleatória dentro de cada estrato, proporcionando uma amostra mais representativa e, consequentemente, mais precisa do que a amostragem aleatória simples com o mesmo tamanho de amostra.

Apesar de ser uma técnica eficaz, a amostragem estratificada tende a ser mais cara, pois exige uma segmentação da população e o processamento de dados mais detalhados.

Vejamos os dois componente principais da estrutura da amostragem estratificada:

- Tamanho da população (N): Refere-se ao total de indivíduos na população, que é a soma dos tamanhos de todos os estratos. Ou seja, o tamanho da população é dado por $N = N_1 + N_2 + \cdots + N_k$, onde $N_1, N_2, \ldots, N_k$ representam o número de indivíduos em cada estrato;

- Tamanho da amostra (n): Representa o número total de indivíduos selecionados na amostra, sendo a soma das amostras extraídas de cada estrato. Assim, o tamanho da amostra é expresso como $n = n_1 + n_2 + \cdots + n_k$, onde $n_1, n_2, \ldots, n_k$ são os tamanhos das amostras extraídas de cada estrato individualmente.

A alocação da amostra nos estratos pode ser feita de várias maneiras, dependendo das informações disponíveis e dos objetivos do estudo. As principais formas de alocação são:

- Alocação por igual: Quando se presume que todos os estratos possuem tamanhos semelhantes $(N_1 \approx N_2 \approx \cdots \approx N_k)$, cada estrato recebe a mesma quantidade de amostra, isto é,

$$n_1 = n_2 = \cdots = n_k = \frac{n}{k}.$$

Exemplo 3.5.6. *Se a amostra total for $n = 42$ e o número de estratos $k = 6$, cada estrato terá 7 elementos selecionados.*

- Alocação proporcional ao tamanho do estrato: Neste caso, a amostra de cada estrato é proporcional ao seu tamanho em relação à população total, isto é,

$$\frac{n_1}{n} = \frac{N_1}{N}, \frac{n_2}{n} = \frac{N_2}{N}, \ldots, \frac{n_k}{n} = \frac{N_k}{N}.$$

Dessa forma, a quantidade de amostras de cada estrato é dado por:

$$n_i = \frac{n \cdot N_i}{N} \text{ , para todo } i \in \{1, 2, 3, \ldots, k\}.$$

Exemplo 3.5.7. *Para uma população com $N = 120$ e uma amostra de tamanho $n = 24$ a ser selecionada divida em três estratos $N_1 = 20, N_2 = 40$ e $N_3 = 60$ a alocação proporcional das amostras coletadas dos três estratos será de:*

$$n_1 = \frac{24 \cdot 20}{120} = 4 \text{ , } n_2 = \frac{24 \cdot 40}{120} = 8 \text{ e } n_3 = \frac{24 \cdot 60}{120} = 12.$$

Desse modo deve ser selecionado 4 indivíduos do primeiro estrato, 8 do segundo estrato e 12 do terceiro.

- Alocação ótima: Busca otimizar a alocação de amostras em estratos com base em uma função específica, considerando o tamanho dos estratos, a variabilidade (heterogeneidade) interna de cada estrato e o custo de amostragem. O objetivo dessa técnica é maximizar a precisão das estimativas ou minimizar o custo total da amostragem, dependendo do critério escolhido.

Exemplo 3.5.8. *Considere uma pesquisa em uma fábrica com três setores (estratos): produção, vendas e administração. Cada setor tem características distintas, como variabilidade no desempenho e custo de amostragem.*

- *Estrato 1 - Produção: alta variabilidade, mas custo de amostragem baixo.*
- *Estrato 2 - Vendas: média variabilidade, custo de amostragem médio.*
- *Estrato 3 - Administração: baixa variabilidade, mas custo de amostragem alto.*

Se o objetivo for minimizar a variabilidade nas estimativas, a alocação ótima pode distribuir mais amostras para o estrato de maior variabilidade, ou seja, o estrato de produção, para garantir estimativas mais precisas.

Por outro lado, se o objetivo for minimizar o custo da pesquisa, a alocação deverá ser maior para os estratos de baixo custo (produção), mesmo que isso reduza a precisão das estimativas nos setores de alta variabilidade.

Exemplo 3.5.9. *Considere a seguinte situação de uma pesquisa em uma fábrica com 1000 funcionários dividos em três estratos : produção ($N_1 = 600$), vendas ($N_2 = 300$) e administração ($N_3 = 100$). Vamos agora assumir que temos um orçamento de $R\$10.000$ para a pesquisa. Assim:*

- *Estrato 1 - Produção: com variabilidade $\sigma_1 = 30$ e o custo de amostragem $C_1 = 100$ por amostra.*
- *Estrato 2 - Vendas: com variabilidade $\sigma_2 = 20$ e o custo de amostragem $C_2 = 150$ por amostra.*
- *Estrato 3 - Administração: com variabilidade $\sigma_3 = 10$ e o custo de amostragem $C_3 = 200$ por amostra.*

Vamos calcular as amostras utilizando a fórmula de Neyman, que busca minimizar o erro quadrático médio de estimativas para cada estrato.

$$n_i = \frac{N_i \cdot \sigma_i}{C_i} \cdot \frac{1}{\sum_{i=1}^{k} \frac{N_i \cdot \sigma_i}{C_i}}, \quad \textit{para todo } i \in \{1, 2, 3, \ldots, k\}. \tag{3.7}$$

Primeiramente calcularemos para cada estrato $\dfrac{N_i \cdot \sigma_i}{C_i}$*:*

- *Estrato 1:*

$$\frac{600 \cdot 30}{100} = 180.$$

- *Estrato 2:*

$$\frac{300 \cdot 20}{150} = 40.$$

– *Estrato 3:*

$$\frac{100 \cdot 10}{200} = 5.$$

Agora, somaremos os três resultados:

$$\sum_{i=1}^{3} \frac{N_i \cdot \sigma_i}{C_i} = 180 + 40 + 5 = 225.$$

E o terceiro passo é calcular a alocação de amostras para cada estrato:

– *Estrato 1:*

$$\frac{600 \cdot 30}{100} \cdot \frac{1}{225} = 80 \ \textit{amostras}.$$

– *Estrato 2:*

$$\frac{300 \cdot 20}{150} \cdot \frac{1}{225} \approx 18 \ \textit{amostras}.$$

– *Estrato 3:*

$$\frac{100 \cdot 10}{200} \cdot \frac{1}{225} \approx 2 \ \textit{amostras}.$$

Deste modo, teremos no total $n = 100$ amostras. O estrado de Produção recebe mais amostras devido à sua maior variabilidade, apesar de ter um custo mais baixo, o que ajuda a reduzir o erro na estimativa. Já o estrato de Administração, embora tenha menor variabilidade, tem um custo de amostragem mais alto, o que justifica a alocação de menos amostras. Com essa distribuição de amostras, conseguimos equilibrar a variabilidade e o custo de amostragem.

Portanto, esse tipo de locação de amostras determina a quantidade de amostras a ser retirada de cada estrato com base: na variabilidade dentro de cada estrato, que exige mais amostras para estimativas precisas, e no custo de amostragem, que indica que estratos com maior custo devem receber menos amostras para controlar os custos totais. O objetivo é equilibrar precisão e custo na pesquisa.

Em conclusão, a amostragem estratificada é uma técnica valiosa para populações heterogêneas, pois garante maior representatividade e precisão nas estimativas. Embora implique custos mais elevados devido à segmentação da população, seus benefícios superam a amostragem simples, especialmente quando são aplicados métodos de alocação adequados, como igual, proporcional ou ótima. Apesar do custo adicional, a precisão e a qualidade das estimativas obtidas justificam o investimento, tornando essa abordagem altamente recomendada em diversos contextos de pesquisa.

3. **Amostragem Sistemática**

   A amostragem sistemática é uma técnica usada quando a informação a ser coletada está facilmente acessível, facilitando o processo e garantindo uma distribuição adequada dos elementos na população. Nesse método, as unidades amostrais são selecionadas de forma regular e estruturada a partir de um sistema de referência bem definido. Esse processo é especialmente eficaz quando se dispõe de uma lista ou banco de dados, como:

   - Fichas de cadastro de assinantes (revistas, provedores de acesso à internet, serviço telefônico, etc.);
   - Cadastro de funcionários de uma empresa;
   - Peças em uma linha de produção;
   - Mudas em um canteiro de plantas, entre outros.

   A amostra é extraída de maneira que cada elemento da população tenha uma chance constante e conhecida de ser selecionado, o que favorece a representatividade da amostra.

   O processo de amostragem sistemática pode ser descrito nas etapas a seguir:

   a) Determinação do intervalo de seleção: O primeiro passo é calcular o intervalo de seleção, que indica a distância entre cada unidade selecionada da população. Esse intervalo é dado pela fórmula:
   $$S = \frac{N}{n}.$$
   Onde:
   S é o intervalo de seleção;
   N é o tamanho total da população;
   n é o tamanho da amostra;

   b) Seleção da primeira unidade amostral aleatoriamente: Uma vez definido o intervalo S, o próximo passo é sortear aleatoriamente o primeiro item a ser selecionado. Este item será escolhido dentro dos primeiros S elementos da lista.

   c) Seleção dos demais elementos de forma sistemática: A partir da primeira unidade amostral escolhida, os próximos itens são selecionados de forma sistemática, respeitando o intervalo S, ou seja, após o primeiro item, o próximo será o de posição S na lista, o seguinte será a posição $2S$, e assim por diante, até completar o número de elementos desejados na amostra.

Exemplo 3.5.10. *Suponha que uma empresa tem uma lista de 120 produtos que foram produzidos ao longo de um mês e deseja realizar uma inspeção de qualidade em uma amostra de 10 produtos. Para isso, a amostragem sistemática pode ser aplicada.*

O procedimento seria o seguinte:

a) *Determinar o intervalo de seleção:*

$$S = \frac{120}{10} = 12.$$

Ou seja, a cada 12 produtos, um novo item será selecionado para a amostra.

b) *Escolher aleatoriamente o primeiro produto dentro dos 12 primeiros da lista. Suponhamos que o 4º produto da lista seja o primeiro a ser selecionado.*

c) *A seleção dos demais produtos se dá a partir do 4º produto, continuando de forma sistemática, respeitando o intervalo de 12 produtos. Os produtos selecionados serão:*
O segundo elemento será o $4 + 12 = 16$*;*
O terceiro elemento será o $16 + 12 = 28$*;*
O quarto elemento será o $28 + 12 = 40$*;*
O quinto elemento será o $40 + 12 = 52$*;*
O sexto elemento será o $52 + 12 = 64$*;*
O sétimo elemento será o $64 + 12 = 76$*;*
O oitavo elemento será o $76 + 12 = 88$*;*
O nono elemento será o $88 + 12 = 100$*;*
O décimo elemento será o $100 + 12 = 112$*.*
Portanto, os 10 produtos selecionados para inspeção seriam os $4°$, $16°$, $28°$, $40°$, $52°$, $64°$, $76°$, $88°$, $100°$ *e* $112°$ *produtos da lista.*

Embora eficaz, a amostragem sistemática exige que a população esteja representada por uma lista completa e ordenada, o que nem sempre é possível. Além disso, caso haja um padrão ou repetição cíclica na lista de elementos (ou seja, uma sequência previsível de características que se repetem regularmente), o método pode resultar em uma amostra não representativa. Isso ocorre porque o intervalo de seleção pode coincidir com o padrão cíclico [4], distorcendo a amostra.

4. **Amostragem por Conglomerados**

Consiste em uma técnica em que a população é segmentada em grupos, denominados conglomerados. Cada conglomerado deve ser representativo da população, ou seja, ele deve refletir as características e a diversidade da população como um todo. A divisão pode ser feita com base em diferentes critérios, como localização geográfica,

[4] Um exemplo de padrão cíclico pode ser a distribuição de funcionários em turnos de trabalho:
Sequência: Manhã, Tarde, Noite, Manhã, Tarde, Noite, ...
Neste caso, os turnos de trabalho se repetem a cada 3 funcionários.
Se você escolher um intervalo de 3 para a amostragem, selecionando a cada 3º funcionário, você pode acabar selecionando funcionários sempre no mesmo turno de trabalho.

unidades de tempo ou qualquer outra característica que agrupe unidades de maneira significativa.

Após a formação dos conglomerados, aplica-se a técnica de amostragem aleatória simples para selecionar aleatoriamente os conglomerados que farão parte da amostra. O número de conglomerados a serem selecionados pode ser definido com base no tamanho total da amostra desejada, considerando as características da população e os objetivos do estudo.

Uma vez escolhidos os conglomerados, todos os elementos pertencentes a eles são incluídos na amostra. Este procedimento é o mais simples e amplamente utilizado quando a amostragem é realizada em um *único estágio*.

Entretanto, a amostragem por conglomerados pode ser aplicada em *dois ou mais estágios*, configurando um processo mais complexo. Essa técnica em dois estágios, após a seleção aleatória dos grupos, realiza-se um novo sorteio dentro de cada conglomerado para escolher os elementos que efetivamente farão parte da amostra.

Exemplo 3.5.11. *Suponha que se deseja realizar uma pesquisa sobre a qualidade de serviços de saúde em diferentes bairros de uma cidade.*

É possível observar que, devido à divisão da cidade em bairros, cada bairro pode ser considerado um conglomerado. Dessa forma, a pesquisa pode ser realizada por meio da aplicação da técnica de amostragem por conglomerados em dois estágios:

- *1º estágio - Seleção aleatória dos conglomerados : Realiza-se um sorteio aleatório de alguns bairros da cidade para integrar a amostra.*
- *2º estágio - Seleção de unidades de saúde dentro dos conglomerados: Dentro dos bairros selecionados, realiza-se um novo sorteio aleatório para selecionar as unidades de saúde que serão avaliadas. Em seguida, todos os pacientes ou serviços dessas unidades são entrevistados para avaliar a qualidade dos serviços prestados.*

A divisão em dois estágios, sendo o primeiro focado na seleção dos bairros e o segundo na seleção das unidades de saúde, possibilita uma amostragem mais controlada e direcionada, o que contribui para um aumento na precisão da coleta de dados. Esse método é útil quando a população é grande e a amostragem de cada indivíduo seria inviável ou muito cara.

5. **Amostragem Combinada ou Multietápica**

A amostragem combinada ou multietápica é uma técnica que combina diferentes métodos de amostragem em várias etapas sucessivas, com o objetivo de otimizar a coleta de dados em populações grandes e dispersas. Esse processo envolve a aplicação de diferentes critérios de amostragem em cada estágio, podendo incluir métodos como a amostragem aleatória simples, por conglomerados ou sistemática, dependendo das características da população e dos objetivos da pesquisa.

Uma das principais vantagens dessa abordagem é a possibilidade de utilizar métodos mais simples e econômicos nas etapas iniciais, reservando técnicas mais complexas ou precisas para estágios posteriores. Isso contribui para a redução dos custos e do tempo necessário para a coleta de dados, sem comprometer a representatividade e a precisão da amostra. Dessa forma, a amostragem multietápica é especialmente útil em pesquisas que exigem flexibilidade, permitindo a obtenção de uma amostra representativa de maneira eficaz, mesmo em grandes populações.

Exemplo 3.5.12. *Suponha que se deseje realizar uma pesquisa sobre o acesso e a qualidade dos serviços educacionais em uma cidade com grande extensão territorial e múltiplos bairros.*

Devido à dispersão geográfica da população e ao elevado custo de uma amostragem de todos os indivíduos, pode-se optar pela técnica de amostragem combinada.

O processo pode ser estruturado da seguinte forma:

- *Seleção dos bairros da cidade: inicialmente, a cidade é subdividida em bairros, que são tratados como conglomerados. A partir dessa divisão, realiza-se um sorteio aleatório para escolher os bairros que integrarão a amostra;*
- *Seleção de escolas dentro dos bairros selecionados: após a definição dos bairros, realiza-se um novo sorteio em cada um deles para determinar as escolas que serão avaliadas. Esse sorteio pode seguir critérios específicos, como a escolha entre escolas públicas ou privadas;*
- *Seleção de alunos dentro das escolas: posteriormente, dentro das escolas escolhidas, é realizada uma seleção aleatória ou sistemática de alunos que serão entrevistados, de modo a avaliar a qualidade dos serviços educacionais, como infraestrutura, formação dos professores e recursos pedagógicos disponíveis.*

Essa abordagem multietápica permite uma amostragem mais eficiente, reduzindo os custos e o tempo necessários, ao mesmo tempo em que assegura a representatividade e a precisão dos dados coletados, refletindo adequadamente as características da população.

3.6 Erro Amostral e Erros Não Amostrais: Compreendendo as fontes de imprecisão na Pesquisa

O erro amostral é a diferença entre o valor estimado de um parâmetro populacional[5], obtido a partir de uma amostra, e o verdadeiro valor desse parâmetro na população total. Esse erro ocorre devido à variação existente entre as unidades amostrais, pois representam apenas uma parte da população. Desse modo os resultados obtidos a partir da amostra podem diferir do que se encontraria se fosse possível medir toda a população.

Exemplo 3.6.1. *Suponha que queiramos saber a altura média de todos os estudantes de uma escola, mas não podemos medir a altura de todos. Para isso, selecionamos uma amostra de 10 estudantes para medir as alturas. Se a média das alturas dessa amostra for, por exemplo, 1,60 metros, e a média de altura de todos os estudantes da escola for, na realidade, 1,62 metros, a diferença entre 1,60 e 1,62 metros é o erro amostral.*

Na prática, não existe amostra totalmente isenta de erro, pois toda amostra é, por natureza, uma representação parcial da população. No entanto, é possível minimizar o erro amostral por meio de algumas precauções. Por exemplo, uma amostra maior tende a reduzir a variabilidade das estimativas. Além disso, utilizar métodos de amostragem adequados, como a amostragem aleatória, estratificada ou sistemática, pode aumentar as chances de a amostra refletir com maior precisão as características da população.

A determinação adequada do tamanho da amostra está diretamente relacionada à redução do erro amostral, por que permite que uma amostra seja suficientemente representativa da população, com uma margem de erro controlada. Essa margem pode ser para mais ou para menos, e seu valor depende do nível de precisão desejado para a pesquisa. A probabilidade de que a estimativa do parâmetro populacional esteja dentro da margem de erro deve ser alta, geralmente 95%. Ao calcular o tamanho da amostra levando em conta o *erro tolerável* e a probabilidade de cometer esse erro, o pesquisador aumenta as chances de que as estimativas do parâmetro populacional sejam precisas. Mas como efetuar esse cálculo?

Na prática, pode-se escolher um tamanho inicial para a amostra n_0 em função do erro relativo tal que:

$$n_0 = \left(\frac{1}{\text{erro relativo}}\right)^2 . \tag{3.8}$$

[5] Parâmetro populacional é o valor numérico fixo que descreve uma característica de toda a população, como média, proporção, variância, desvio, dentre outros. Por exemplo, a média de idade de todos os moradores de uma cidade é um parâmetro populacional.

Ou ainda, para calcular n_0 podemos considerar uma distribuição de probabilidades[6] para a estimativa, normalmente a *Gaussiana*, ou *normal*, usando a seguinte fórmula:

$$n_0 = \frac{z^2 \cdot \sigma^2}{E^2}\,. \tag{3.9}$$

Onde:

- z é o valor crítico da distribuição normal para o nível de confiança da pesquisa;
- σ é o desvio padrão da população;
- E é o erro relativo desejado.

Após, calcula-se o tamanho da amostra (n) apropriada para a pesquisa por meio de uma correção para a população finita (se você conhece o tamanho total da população (N). usando a fórmula:

$$n = \frac{n_0 \cdot N}{N + n_0 - 1}\,. \tag{3.10}$$

Exemplo 3.6.2. *Suponha que uma empresa deseja realizar uma pesquisa sobre a satisfação de seus 1.200 clientes. A empresa quer um erro relativo de 4%.*

Desse modo, calculamos o tamanho inicial para a amostra n_0 *em função do erro relativo* $0,04$*:*

$$n_0 = \left(\frac{1}{0,04}\right)^2\,.$$

$$n_0 = 625$$

Agora, aplicando a correção de população finita, já que a população não é muito grande ($N = 1200$*), temos:*

$$n = \frac{625 \cdot 1200}{1200 + 625 - 1}$$

$$n \approx 411\ \ .$$

Portanto, o tamanho da amostra corrigido seria de 411 unidades amostrais.

6 Vale ressaltar que de acordo com as orientações curriculares na tabela 1, esse conceito será apenas estudado na 3ª série do Ensino Médio. Desse modo, para efeito de cálculo de n_0 utilizaremos a fórmula anterior.

É importante destacar que, quando a população é muito grande, a correção se torna insignificante, e o tamanho da amostra n_0 pode ser usado diretamente como o tamanho da amostra final, pois na fórmula 3.10, quando N é suficientemente grande $\frac{n_0}{N} \approx 0$, assim:

$$n = \frac{n_0 \cdot N}{N + n_0 - 1} = \frac{n_0}{1 + \frac{n_0}{N} - \frac{1}{N}} \approx n_0\,.$$

Além do erro amostral, há também os *erros não amostrais*, que são causados por falhas no processo de coleta de dados, mas não estão diretamente relacionados à amostragem em si. Esses erros podem surgir de diversas fontes e impactar significativamente a qualidade da pesquisa. Dentre os principais erros não amostrais, podemos destacar:

- População acessível diferente da população alvo: ocorre quando a amostra não representa adequadamente a população a ser estudada, impedindo que a amostra capture a diversidade da população-alvo.
- Falta de resposta: esse erro acontece quando unidades amostrais não respondem à pesquisa ou não fornecem dados completos, gerando um viés de não resposta, o que compromete a representatividade da amostra.
- Erros de Mensuração: são causados por falhas no processo de coleta ou registro de dados, como o uso de instrumentos inadequados ou perguntas mal formuladas, afetando a precisão das estimativas e distorcendo os resultados.

A redução do erro amostral e a prevenção dos erros não amostrais são de suma importância para assegurar que os resultados da pesquisa reflitam com precisão as características da população estudada. A minimização dessas fontes de erro contribui diretamente para a confiabilidade e representatividade dos dados, garantindo que as conclusões tiradas a partir da amostra sejam mais próximas da realidade da população como um todo.

3.7 Valor esperado: Conceito e aplicações

Para compreender o valor esperado de um experimento aleatório, é necessário primeiramente entender os conceitos fundamentais de **probabilidade** e as **distribuições de probabilidade**, que são essenciais para a análise de variáveis aleatórias.

Conforme destacado por MORETTIN e BUSSAB (2010), a **probabilidade** é a teoria matemática voltada ao estudo da incerteza presente em fenômenos de caráter aleatório. Essa teoria é essencial para a realização de previsões e estimativas em diversas áreas, como clima, mercado financeiro e processos eleitorais. Além de facilitar a análise de dados experimentais, a probabilidade fundamenta a inferência estatística e o teste de hipóteses. Além disso, é essencial em algoritmos de aprendizado de máquina, permitindo fazer previsões e decisões baseadas em padrões encontrados em grandes volumes de dados.

Os conceitos fundamentais da teoria da probabilidade, incluindo eventos, espaço amostral, probabilidade de eventos e variáveis aleatórias, fornecem a base para compreender e calcular a chance de ocorrência de diferentes resultados em situações incertas. Esses conceitos são detalhados no trabalho de Souza (SOUZA, 2024, seção 2.3, p. 36). Na estatística, essas ideias se refletem na coleta e análise de dados, onde a probabilidade auxilia na modelagem das incertezas e na inferência sobre populações a partir de amostras. Assim, enquanto a probabilidade trata do estudo teórico da incerteza, a estatística aplica esses princípios para interpretar dados reais e tomar decisões informadas.

A **distribuição de probabilidade**, por sua vez, descreve a chance de ocorrência de cada um dos resultados possíveis em um experimento aleatório. As distribuições de probabilidade podem ser classificadas em **discretas** e **contínuas**, dependendo da natureza dos resultados possíveis.

- **Distribuições Discretas** - As distribuições discretas são aquelas que podem assumir um **número finito** ou contável de resultados. Nos concentremos aqui no caso finito. Situação na qual a probabilidade de ocorrência de um evento A é dada pela razão entre o número de resultados favoráveis e o número total de resultados possíveis, conforme a fórmula:

$$P(A) = \frac{\textit{número de resultados favoráveis}}{\textit{número total de resultados possíveis}}. \tag{3.11}$$

  Essa definição é válida apenas para distribuições discretas finitas, onde os resultados podem ser contados e são finitos. Um exemplo disso é o lançamento de uma moeda, onde os resultados possíveis são "cara" ou "coroa", e cada um tem a mesma probabilidade de ocorrer, ou seja, $0,5$ para "cara" e $0,5$ para "coroa". Para obter esse valor basta utilizar a fórmula acima, por exemplo, para o evento E — ocorrer cara no lançamento de uma moeda — temos 1 resultado favorável entre 2 resultados possíveis:

$$P(E) = \frac{1}{2} = 0,5.$$

  Outro exemplo clássico é o lançamento de um dado, onde os resultados possíveis são $1, 2, 3, 4, 5$, e 6, e cada um tem a mesma probabilidade de $\frac{1}{6}$. Para uma distribuição discreta, a soma das probabilidades de todos os resultados possíveis deve ser igual a 1.

- **Distribuições Contínuas** - As distribuições contínuas se aplicam a variáveis aleatórias que podem assumir qualquer **valor dentro de um intervalo contínuo**. Um exemplo dessa distribuição é a temperatura de uma cidade, que pode variar de maneira contínua ao longo do dia. A probabilidade de uma temperatura exata ocorrer em um momento específico é zero, sendo descrita por uma função de densidade de probabilidade, como a distribuição normal (a qual será abordada na Seção 3.9).

O **valor esperado** de uma variável aleatória é um conceito fundamental na teoria das probabilidades e fornece uma medida do "valor médio" esperado de um experimento aleatório. O valor esperado é calculado como a média ponderada[7] dos resultados possíveis, onde os pesos são dados pelas probabilidades associadas a esses resultados. Para distribuições discretas, o valor esperado $E(X)$ é dado por:

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(x_i)\,. \tag{3.12}$$

Onde,

- x_i são os valores possíveis;
- $P(x_i)$ são as probabilidades associadas a esses valores.

Para distribuições contínuas, o valor esperado é obtido por meio da integral da função densidade de probabilidade $f(x)$ multiplicada pela variável x. Esse conceito é frequentemente abordado no ensino superior, sobretudo em disciplinas de Estatística e Probabilidade. Diante disso, este estudo concentrar-se-á nas distribuições discretas finitas.

Exemplo 3.7.1. *Ao lançar um dado não viciado de seis faces, o* ***valor esperado*** *é calculado utilizando todos os resultados possíveis 1, 2, 3, 4, 5 e 6, levando em conta que cada face tem a mesma probabilidade de ocorrer* $\left(\frac{1}{6}\right)$*. Nesse caso, o valor esperado seria* $3,5$*, pois utilizando a fórmula 3.12 temos:*

$$E(X) = \left(1 \cdot \frac{1}{6}\right) + \left(2 \cdot \frac{1}{6}\right) + \left(3 \cdot \frac{1}{6}\right) + \left(4 \cdot \frac{1}{6}\right) + \left(5 \cdot \frac{1}{6}\right) + \left(6 \cdot \frac{1}{6}\right)$$

$$E(X) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3,5\,.$$

Esse valor reflete o comportamento médio esperado do experimento após muitas repetições. Vale destacar que, embora $3,5$ *represente o valor esperado, ele não é um dos possíveis resultados em um único lançamento do dado.*

Exemplo 3.7.2. ***Análise Estatística: Tempo Diário de Estudo de Estudantes do Ensino Médio*** *- Uma turma do* 3^a *série do ensino médio realizou uma pesquisa para saber quanto tempo, em média, os estudantes dedicam diariamente aos estudos fora da sala de aula. Participaram 50 alunos, que registraram seu tempo diário de estudo durante*

Tabela 11 – Distribuição do tempo diário de estudo dos estudantes

i	**Intervalo de Tempo (h)**	**Ponto Médio ($\bar{x}_i$)**[8]	f_i	fr_i
1	0 |- 1	0.5	8	0.16
2	1 |- 2	1.5	15	0.30
3	2 |- 3	2.5	18	0.36
4	3 |- 4	3.5	6	0.12
5	4 |- 5	4.5	3	0.06
Total			**50**	**1.00**

Fonte: Produção da própria autora (2025).

uma semana. Os dados foram organizados em intervalos de tempo, formando a seguinte tabela de frequência de dados agrupados (Tabela 11).

Utilizando a fórmula 3.12 de valor esperado temos:

$$E(X) = (0{,}5 \cdot 0{,}16) + (1{,}5 \cdot 0{,}30) + (2{,}5 \cdot 0{,}36) + (3{,}5 \cdot 0{,}12) + (4{,}5 \cdot 0{,}06)$$

$$E(X) = 0{,}08 + 0{,}45 + 0{,}90 + 0{,}42 + 0{,}27 = 2{,}12$$

*Deste modo, o tempo médio esperado de estudo diário entre os alunos pesquisados é de aproximadamente **2,12 horas**. Isso significa que, em média, um estudante dessa turma dedica cerca de 2 horas e 7 minutos por dia aos estudos fora da sala de aula.*

A faixa de 2 a 3 horas concentra a maior frequência relativa (36%), o que influencia diretamente no valor esperado. Esse resultado mostra como a frequência relativa serve de ponderador no cálculo da média, tornando o valor esperado uma boa estimativa do comportamento geral da turma[9].

Exemplo 3.7.3. *Suponha que um investidor esteja avaliando a possibilidade de aplicar $R\$100,00$ em uma oportunidade de investimento de alto risco. Essa aplicação apresenta dois possíveis resultados:*

- *Com probabilidade de $0,3$ (ou 30%), o investidor obtém um retorno de $R\$300,00$.*
- *Com probabilidade de $0,7$ (ou 70%), o investidor perde todo o valor aplicado, ou seja, $R\$100,00$.*

7 Na próxima seção deste estudo, abordaremos o cálculo de médias, incluindo as diferentes formas de média, como a aritmética e a ponderada, e suas aplicações em diversos contextos estatísticos.

9 Ao abordar este trecho, é importante que o professor estabeleça conexões explícitas entre os objetos de conhecimento desta seção (frequência, valor esperado) e os da próxima seção, em que o conceito de média será estudado com mais profundidade. Essa articulação entre conteúdos contribui para dar mais significado à aprendizagem dos alunos, mostrando a coerência interna da matemática e sua aplicabilidade em contextos reais.

Seja X a variável aleatória que representa o retorno líquido do investimento. Os valores possíveis de R são:

$$X = \begin{cases} \text{300,} & \text{com probabilidade } \ 0{,}3 \\ \text{-100,} & \text{com probabilidade } 0{,}7 \end{cases}$$

Assim, o valor esperado (ou esperança matemática) do retorno é calculado pela fórmula 3.12. Substituindo os valores:

$$E(X) = (300 \cdot 0,3) + (-100 \cdot 0,7)$$
$$E(X) = 90 - 70 = 20\,.$$

Portanto, o valor esperado do retorno do investimento é de $R\$20,00$. Isso indica que, em média, espera-se um ganho de $R\$20,00$ por investimento, caso a operação fosse repetida muitas vezes sob as mesmas condições.

Exemplo 3.7.4. *Uma fábrica está decidindo se deve produzir um lote adicional de um produto. O custo de produção é de $R\$2.000,00$ e a empresa estima dois cenários possíveis de demanda para esse lote:*

- *Com probabilidade de $0,4$, a demanda será alta e o lucro obtido será de $R\$5.000,00$;*
- *Com probabilidade de $0,6$, a demanda será baixa e a empresa conseguirá vender apenas parte do lote, obtendo um lucro de $R\$1.000,00$.*

Seja L a variável aleatória que representa o ***lucro líquido*** *do lote. Subtraindo o custo de produção dos lucros brutos, temos:*

$$L = \begin{cases} \text{5000 - 2000 = 3000,} & \text{com probabilidade } \ 0{,}4 \\ \text{1000 - 2000 = - 1000,} & \text{com probabilidade } 0{,}6 \end{cases}$$

Calculando o valor esperado do lucro com a fórmula 3.12:

$$E(X) = (3000 \cdot 0,4) + (-1000 \cdot 0,6)$$
$$E(X) = 1200 - 600 = 600.$$

Deste modo, o lucro esperado com a produção do lote é de $R\$600,00$. Isso indica que, em média, a decisão de produzir esse lote resulta em ganho financeiro, o que pode justificar a produção do ponto de vista estatístico.

Além de ser útil para entender a média de uma distribuição de probabilidades, o valor esperado tem aplicações práticas em diversos campos, especialmente em situações de **tomada de decisões** sob incerteza. Em finanças, por exemplo, ele é utilizado para calcular o retorno médio de um investimento, levando em conta as diferentes probabilidades associadas a cenários econômicos variados. O valor esperado também é uma ferramenta fundamental em pesquisas empresariais, como nas análises de produção ou satisfação do cliente. Em uma pesquisa de satisfação, o valor esperado pode ser utilizado para calcular a média ponderada das respostas dos clientes, considerando as probabilidades de diferentes níveis de satisfação. Da mesma forma, em pesquisas de produção, ele pode ser usado para prever o desempenho médio de uma linha de produção, considerando diferentes condições operacionais e variabilidades.

Dessa forma, o valor esperado não apenas ajuda a entender o comportamento central de uma variável aleatória, mas também desempenha um papel importante na formulação de estratégias e decisões informadas em contextos variados, como investimentos financeiros e a melhoria de processos empresariais.

3.8 Medidas Estatísticas

3.8.1 Medidas de Tendência Central ou Medidas de Posição

Ao analisar um conjunto de dados, é comum recorrer ao uso de tabelas de frequência ou gráficos, que oferecem uma visão mais detalhada e informativa sobre o comportamento de uma variável do que a simples tabela de dados brutos, como visto na Seção 3.4 deste estudo. Esses métodos de resumo permitem identificar padrões e distribuições de forma mais clara e eficaz. A apresentação dos dados por meio dessas ferramentas é uma fonte valiosa de informações, muitas vezes mais eficiente do que a própria tabela original de dados (MORETTIN; BUSSAB, 2010). No entanto, em determinados casos, se faz necessário reduzir ainda mais essas informações para apresentar um valor único ou alguns valores que representem de maneira concisa toda a série de dados (CHAVANTE; PRESTES, 2016), ou seja, de forma sintetizada, para expressar características específicas da população em estudo.

Essa redução drástica dos dados pode ser feita por meio das **medidas de tendência central**, que são ferramentas estatísticas essenciais para descrever os valores mais representativos de uma distribuição. Elas nos ajudam a entender o comportamento global dos dados, fornecendo uma indicação de onde a maioria dos valores está concentrada. Entre as principais medidas de tendência central, temos a **média**, a **mediana** e a **moda**, que têm diferentes formas de representar o "centro" dos dados, dependendo da distribuição e das características da série.

Cada uma dessas medidas oferece uma perspectiva única sobre os dados, vejamos:

3.8.1.1 **Moda** (M_o)

É o valor que ocorre com maior frequência em um conjunto de dados. Ela é especialmente útil para dados qualitativos ou quando o objetivo é identificar a tendência mais prevalente entre os dados coletados.

Exemplo 3.8.1. *No exemplo 3.4.2, que se refere às idades (em anos) de dez alunos de uma turma de ensino médio, a idade que mais se repete é 16 anos, sendo esta a moda do conjunto de dados. Ou seja, a maioria dos alunos tem 16 anos, sendo essa a idade mais comum entre os membros da amostra.*

Exemplo 3.8.2. *Utilizando os dados apresentados na tabela de frequência simples 4 ou na 6 do exemplo 3.4.4, podemos identificar a moda ao observar o maior valor na coluna f_i. Neste caso, o valor máximo é $f_3 = 7$, o que indica que a terceira classe correspondente ao número de filhos igual a 2 é a moda. Ou seja, entre os dados coletados sobre o número de filhos, a maior parte dos indivíduos da amostra tem 2 filhos. Isso revela a tendência mais comum dentro dessa variável.*

Entretanto, para o cálculo da moda em uma tabela de distribuição de frequências com dados agrupados, a abordagem difere daquela utilizada em dados não agrupados. Nesse caso, utiliza-se uma fórmula específica que considera a classe modal, ou seja, a classe com a maior frequência absoluta. O método mais comum para calcular a moda em dados agrupados é por meio de **interpolação linear**, baseada em uma construção geométrica fundamentada na semelhança de triângulos no histograma de frequências. A ideia consiste em assumir que, na vizinhança da classe modal, a distribuição das frequências apresenta um comportamento aproximadamente linear. Considera-se que o ponto de máximo da curva de densidade (a moda) localiza-se em algum ponto dentro da classe modal, e sua posição relativa pode ser estimada geometricamente, observando o "topo" dos retângulos que representam as frequências das classes anterior, modal e posterior. A interpolação parte do pressuposto de que dois triângulos — formados pelas diferenças de frequências entre as classes adjacentes e a classe modal — são semelhantes. A partir dessa semelhança, Czuber derivou a seguinte expressão:

$$M_o = l_i + \left(\frac{f_1 - f_o}{2f_1 - f_0 - f_2} \right) \cdot h\,. \tag{3.13}$$

Onde,

l_i = Limite inferior da classe modal

f_1 = Frequência absoluta da classe modal

f_0 = Frequência absoluta da classe anterior à classe modal

f_2 = Frequência absoluta da classe posterior à classe modal

h = Amplitude da classe (diferença entre o limite superior e o limite inferior da classe)

Essa fórmula fornece uma estimativa mais precisa da moda em distribuições agrupadas, especialmente quando as classes são de mesma amplitude. Essa abordagem, conforme discutido por Zat (2022), oferece aos alunos uma oportunidade de aplicar conceitos geométricos — como proporcionalidade e semelhança de triângulos — aliados a ferramentas algébricas, promovendo uma compreensão mais rica e interdisciplinar da Estatística. No entanto, vale destacar que, em distribuições de frequência com dados agrupados, a presença de múltiplas classes com a mesma frequência máxima caracteriza um conjunto de dados multimodal, e a aplicação direta da fórmula pode apresentar limitações.

Um caso particular ocorre quando há três classes modais consecutivas, ou seja, três classes com frequências iguais e superiores às demais. Nessa situação, ao se aplicar a fórmula de interpolação da moda para uma dessas classes, observa-se que:

$$f_1 = f_0 = f_2 = f \ .$$

Substituindo esses valores na Equação 3.13, tem-se:

$$M_o = l_i + \left(\frac{f - f}{2f - f - f}\right) \cdot h = l_i + \left(\frac{0}{0}\right) \cdot h \ .$$

Essa substituição conduz a uma indeterminação do tipo $\frac{0}{0}$, o que impossibilita o uso da fórmula tradicional para o cálculo da moda nesse contexto. Tal resultado evidencia que, quando há múltiplas classes modais consecutivas com a mesma frequência, a interpolação não fornece um valor definido para a moda. Nesses casos, considera-se que a distribuição é multimodal e, portanto, a moda não pode ser representada por um único valor.

Apesar disso, a moda ainda pode ser estimada por meio de métodos alternativos. Uma abordagem simples consiste em considerar o ponto médio da classe modal como uma estimativa da moda, especialmente quando não há assimetrias relevantes na distribuição dos dados.

Exemplo 3.8.3. *Vamos calcular a moda do tempo de serviço com base nos dados coletados de 500 funcionários de uma empresa, conforme o exemplo 3.4.5, utilizando a tabela de dados agrupados 10, por meio da fórmula 3.13.*

Primeiramente, identificamos a classe modal, localizando o maior valor na coluna f_i*. Neste caso, o valor máximo é* $f_2 = 6$*, o que indica que a segunda classe é a classe modal. Em seguida, coletamos os seguintes dados:*

$$l_i = 3 \ ;$$

$$f_1 = 6 \ ;$$
$$f_0 = 5 \ ;$$
$$f_2 = 3 \ ;$$
$$h = L_i - l_i = 6 - 3 = 3 \ .$$

Agora, aplicamos a fórmula para calcular a moda:

$$M_o = 3 + \left(\frac{6-5}{2 \cdot 6 - 5 - 3}\right) \cdot 3$$
$$M_o = 3 + \left(\frac{1}{4}\right) \cdot 3$$
$$M_o = 3 + 0,75$$
$$M_o = 3,75 \ .$$

Portanto, a moda do conjunto de dados é $3,75$ *anos. Isso significa que o tempo de serviço mais comum entre os funcionários dessa empresa é de aproximadamente* 3 *anos e* 9 *meses.*

Observação:

- É importante destacar que, ao calcular a moda para dados agrupados, quando a primeira classe é a classe modal, deve-se considerar $f_0 = 0$ (frequência da classe anterior). Da mesma forma, quando a última classe é a classe modal, utiliza-se $f_2 = 0$ (frequência da classe posterior), uma vez que essas classes não possuem uma classe anterior ou posterior, respectivamente;
- Adicionalmente, os cálculos da moda por meio de gráficos seguem a mesma lógica. Quando utilizamos um histograma, o cálculo da moda é realizado por meio da fórmula de interpolação linear, da mesma forma que é feito com os dados em tabelas de distribuição de frequências agrupadas.

Vale destacar que, um conjuntos de dados que apresentam dois valores com a mesma frequência máxima são denominados *bimodais*. Quando há mais de dois valores com frequência máxima igual, a distribuição é classificada como *multimodal*. Por outro lado, se nenhum valor se repete, a distribuição é considerada *amodal*, ou seja, não apresenta moda.

3.8.1.2 **Média** $\bar{X}$

Há diversos tipos de médias, e a escolha de qual utilizar depende da natureza dos dados e do objetivo da análise. Dentre as principais médias, destacam-se a média aritmética, a média aritmética ponderada, a média geométrica e a média harmônica. A seguir, será apresentada uma análise detalhada de cada uma dessas médias, com explicações sobre suas fórmulas, aplicações e características.

1. **Média Aritmética**

   A média aritmética é, sem dúvida, a medida de tendência central mais conhecida e utilizada. Ela é calculada somando todos os valores de um conjunto de dados dividindo o resultado pelo número total de observações, isto é, a média aritmética $(\bar{x})$ de uma lista de n números x_1, x_2, x_3, ..., x_n é definida por:

   $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{3.14}$$

   O subscrito i indica a posição do valor na sequência, sendo $i = \{1, 2, \ldots, n\}$. Logo, x_1 representa o primeiro valor observado, x_2 representa o segundo valor e assim por diante; x_i é o i-ésimo valor no conjunto de n valores.

   A letra grega maiúscula Σ (sigma) é usada para indicar uma soma. A indicação $\sum_{i=1}^{n} x_i$ significa o somatório dos valores de x_i, para i variando de 1 até n, isto é,

   $$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} .$$

Exemplo 3.8.4. *Na 36° rodada do Campeonato Brasileiro de Futebol 2024, foi registrado um novo recorde de gols,conforme ilustrado na Figura 22.*

Figura 22 – Registro do Recorde de Gols no Brasileirão 2024

Fonte: Ge. Disponível em: <https://ge.globo.com/espiao-estatistico/noticia/2024/12/01/rodada-36-estabelece-um-novo-recorde-de-gols-no-brasileirao-2024.ghtml>. Acesso em: 23 mar. 2025.

A seguir, são apresentados os dados referentes à quantidade de gols marcados em cada uma das 10 partidas disputadas nesta rodada. A tabela 12 exibe o número de gols de cada partida.

Tabela 12 – Número de Gols na 36° Rodada do Brasileirão 2024

Partida	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Número de Gols	5	4	3	6	4	3	5	2	2	2

Com base nos dados apresentados na tabela, podemos calcular a média de gols na 36° rodada do Campeonato Brasileiro 2024. O número de gols em cada uma das 10 partidas foi o seguinte:

$$5, 4, 3, 6, 4, 3, 5, 2, 2, 2 \quad .$$

Para calcular a média $\bar{x}$, utilizamos a fórmula da média aritmética (Equação 3.14). Substituindo os valores dos gols e a quantidade $n = 10$:

$$\bar{x} = \frac{5 + 4 + 3 + 6 + 4 + 3 + 5 + 2 + 2 + 2}{10}$$

$$\bar{x} = \frac{36}{10} = 3,6 \quad .$$

Portanto, a média de gols na 36° rodada do Campeonato Brasileiro 2024 foi de "3,6 gols por partida".

A média de gols em um campeonato serve como um indicador da competitividade do torneio, refletindo a capacidade ofensiva e defensiva das equipes. Ela ajuda a analisar o desempenho das equipes, planejar estratégias, identificar tendências de jogo e pode até ser utilizada como critério de desempate. Além disso, a média de gols contribui para o entretenimento dos torcedores e influencia o marketing e os patrocinadores do evento.

Outra observação importante é que a média obtida não coincidiu com nenhum dos números de gols por partida, pois a média não é necessariamente igual a um dos valores da variável. Isso ocorre porque a média é um valor central que representa o comportamento global dos dados, mas não reflete valores específicos que possam ocorrer em partidas individuais.

Em resumo, a média de gols oferece uma visão geral importante sobre a dinâmica da competição e o estilo de jogo das equipes.

Exemplo 3.8.5. *Suponha que queiramos calcular a média aritmética das idades (em anos) dos dez alunos listados no exemplo 3.4.2:*

15	15	15	16	16
16	16	16	17	19

Aplicando a fórmula da média aritmética (Equação 3.14), substituímos os valores das idades e a quantidade de alunos $n = 10$:

$$\bar{x} = \frac{15 + 15 + 15 + 16 + 16 + 16 + 16 + 16 + 17 + 19}{10}$$

$$\bar{x} = \frac{161}{10} = 16,1 \quad .$$

Portanto, a média das idades dos dez alunos é 16,1 anos. Isso significa que, em média, os alunos têm essa idade, o que pode indicar que a maioria dos alunos está na faixa etária entre 15 e 19 anos, com uma leve concentração em torno de 16 anos.

2. **Média Aritmética Ponderada**

A média aritmética é calculada somando todas as observações e dividindo o total pelo número de elementos. No entanto, em situações onde algumas observações têm maior relevância ou frequência, utiliza-se a **média aritmética ponderada**. Esse tipo de média é aplicado quando as **observações se repetem** ou possuem **pesos diferentes**. Normalmente, os dados que exigem esse cálculo são apresentados em tabelas de frequência, podendo ser agrupados ou não.

A fórmula da média aritmética ponderada é dada por:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}. \tag{3.15}$$

onde:

- $\bar{x}$ é a média ponderada,
- x_i representa os valores das observações,
- w_i são os pesos (ou repetições) atribuídos a cada valor observado x_i, sendo $w_i \geq 0$ para todo i e pelo menos um $w_i > 0$,
- n é o número total de observações.

Ou ainda,

$$\bar{x} = \frac{x_1 \cdot w_1 + x_2 \cdot w_2 + x_3 \cdot w_3 + \cdots + x_n \cdot w_n}{w_1 + w_2 + w_3 + \cdots + w_n}.$$

Exemplo 3.8.6. *Suponha que uma pessoa assista a diferentes tipos de programas de TV ao longo da semana. O tempo gasto assistindo a cada tipo de programa varia, e esses programas têm diferentes níveis de importância, que podem ser representados como "pesos". Vamos calcular a média ponderada do tempo total de visualização, considerando a frequência com que ela assiste a cada tipo de programa também como um fator relevante.*

A tabela a seguir mostra o tempo (em horas) gasto em cada tipo de programa e a frequência com que cada tipo de programa é assistido por semana:

Agora, vamos calcular a média ponderada do tempo de visualização de TV. O cálculo da média ponderada é dado pela Equação 3.15. Substituindo os valores na fórmula:

Tabela 13 – Tempo gasto assistindo diferentes tipos de programas de TV por semana

Tipo de Programa	**Tempo (h)**	**Frequência (n° dias)**	**Peso (Freq.)**
Novelas	2	4	4
Noticiários	1	5	5
Filmes	3	2	2
Documentários	1.5	3	3

Fonte: Produção da própria autora (2025).

$$\bar{x} = \frac{(2 \cdot 4) + (1 \cdot 5) + (3 \cdot 2) + (1.5 \cdot 3)}{4 + 5 + 2 + 3}$$

$$\bar{x} = \frac{8 + 5 + 6 + 4.5}{14}$$

$$\bar{x} = \frac{23.5}{14} = 1.68 \ horas \ \ .$$

Portanto, a média ponderada do tempo de visualização de TV por semana é de "1,68 horas" por dia, considerando a frequência com que os diferentes tipos de programas são assistidos. A média ponderada permite que programas com maior frequência de visualização tenham maior influência no cálculo final, refletindo de forma mais precisa os hábitos de consumo de mídia dessa pessoa.

Exemplo 3.8.7. *No exemplo apresentado em 3.8.5, é possível calcular a média das idades também por meio da média ponderada, já que alguns valores se repetem, como o 15, que ocorre 3 vezes, e o 16, que ocorre 5 vezes. Dessa forma, o cálculo da média ponderada é realizado da seguinte maneira:*

$$\bar{x} = \frac{15 \cdot 3 + 16 \cdot 5 + 17 \cdot 1 + 19 \cdot 1}{3 + 5 + 1 + 1}$$

$$\bar{x} = \frac{161}{10} = 16,1 \ \ .$$

Observe que os pesos, que representam a importância de cada valor, podem ser interpretados como a frequência de ocorrência de cada valor. Essa abordagem é frequentemente utilizada no cálculo de médias em contextos envolvendo tabelas de frequência. A seguir, apresentamos alguns exemplos para ilustrar essa aplicação.

Exemplo 3.8.8. *Suponha que queiramos calcular a média de notas, em uma turma de 10 alunos, aos quais foram organizadas em uma tabela de frequência simples conforma a tabela 14:*

Vamos calcular a média utilizando a fórmula da média aritmética ponderada (Equação 3.15). Substituindo x_i pelas notas dos estudantes e w_i pelas frequências absolutas de cada classe, obtemos:

Tabela 14 – Notas obtidas por dez estudantes

Nota	f_i
4	2
5	3
6	1
7	2
8	2
Total de alunos	10

Fonte: Produção da própria autora 2025

$$\bar{x} = \frac{(4 \cdot 2) + (5 \cdot 3) + (6 \cdot 1) + (7 \cdot 2) + (8 \cdot 2)}{2 + 3 + 1 + 2 + 2}$$

$$\bar{x} = \frac{59}{10}$$

$$\bar{x} = 5,9 \quad .$$

Portanto, a média ponderada das notas da turma é "5,9" pontos.

No exemplo acima, a média ponderada foi calculada utilizando a frequência de cada nota como peso, o que reflete a quantidade de alunos em cada faixa de notas. A seguir, vamos explorar a aplicação dessa metodologia em tabelas de dados agrupados[10].

Exemplo 3.8.9. *Calcular a média do tempo de serviço (anos) de uma amostra de 25 funcionários, com base nos dados fornecidos na tabela 10 do exemplo 3.4.5.*

Primeiramente, é necessário calcular o ponto médio de cada intervalo de classe, utilizando a média aritmética dos limites inferior e superior de cada classe, conforme a fórmula:

$$\bar{x} = \frac{l_i + L_i}{2} \quad .$$

Onde l_i e L_i são, respectivamente, os limites inferior e superior de cada intervalo de classe.

Esse valor será utilizado como base para os cálculos subsequentes. Em seguida, uma coluna adicional é inserida para simplificar o processo de cálculo, conforme mostrado na tabela 15.

[10] Curiosidade: Existem diversos procedimentos para calcular a média em tabelas de frequência com dados agrupados. Neste trabalho, será adotado o método mais comum: a média ponderada, em que cada classe é representada por um valor típico (geralmente o ponto médio), e sua importância é definida pela frequência correspondente. Em distribuições contínuas, pode-se empregar o **método da média de classes contínuas**, que considera os limites dos intervalos. Já em distribuições acumuladas, utiliza-se a **média ponderada com base nas frequências acumuladas**. Independentemente do método, o propósito é sempre obter uma medida representativa da tendência central dos dados.

Tabela 15 – Distribuição de freq. da variável tempo de serviço (alterada)

i	**Tempo de Serviço**	Ponto médio (PM)	f_i	$f_r(\%)$	$f_{ra}(\%)$
1	0 |- 3	1,5	5	20	20
2	3 |- 6	4,5	6	24	44
3	6 |- 9	7,5	3	12	56
4	9 |- 12	10,5	4	16	72
5	12 |- 15	13,5	4	16	88
6	15 |- 18	16,5	3	12	100
			25	100	

Fonte: Produção da própria autora (2024).

Com base na Equação 3.15, calculamos a média ponderada do tempo de serviço da amostra composta por 25 funcionários.

$$\bar{x} = \frac{(1,5 \cdot 5) + (4,5 \cdot 6) + (7,5 \cdot 3) + (10,5 \cdot 4) + (13,5 \cdot 4) + (16,5 \cdot 3)}{5 + 6 + 3 + 4 + 4 + 3}$$
$$\bar{x} = \frac{202,5}{25}$$
$$\bar{x} = 8,1 \quad .$$

Assim, a média de tempo de serviço dos funcionários é de 8,1 anos, o que indica que, em média, os funcionários possuem essa quantidade de anos de experiência na empresa.

3. **Média Geométrica**

A média geométrica é uma medida de tendência central usada para calcular a média de conjuntos de dados cujos valores possuem uma relação multiplicativa ou exponencial. Ela é especialmente útil quando se lida com taxas de crescimento ou variações relativas, como no cálculo de índices financeiros, crescimento populacional ou produtividade.

A fórmula da média geométrica para um conjunto de n números não negativos $x_1, x_2, \ldots, x_n$ é dada por:

$$\bar{x}_g = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} . \tag{3.16}$$

onde:

- $\bar{x}_g$ é a média geométrica,
- x_i representa os valores das observações,
- n é o número total de observações.

O nome desse símbolo, $\prod$, vem da palavra "produc" em inglês, que significa "produto". $\prod_{i=1}^{n} x_i$ significa o produtório dos valores x_i, para i variando de 1 até n.

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot \ldots \cdot x_n \,.$$

Assim, a média geométrica pode ser entendida como a raiz n-ésima do produto dos valores, isto é,

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}.$$

Exemplo 3.8.10. *Considere que um atleta participou de 4 corridas ao longo de um ano, obtendo os seguintes tempos (em minutos) em cada uma delas:*

- *Corrida 1: 12,5 minutos;*
- *Corrida 2: 11,8 minutos;*
- *Corrida 3: 12,0 minutos;*
- *Corrida 4: 11,7 minutos.*

O objetivo é calcular a média geométrica desses tempos, a fim de obter uma medida representativa do desempenho do atleta ao longo das corridas.

Para calcular a média geométrica dos tempos, devemos multiplicar os tempos e, em seguida, tirar a raiz n-enésima, onde n é o número de corridas. Nesse caso, temos 4 tempos. Assim:

$$\bar{x}_g = \sqrt[4]{12,5 \cdot 11,8 \cdot 12 \cdot 11,7}$$

$$\bar{x}_g = \sqrt[4]{20,709} \approx 12 \quad .$$

Portanto, o tempo médio do atleta é de 12 minutos, representando a medida mais precisa do desempenho, levando em consideração o efeito composto de cada corrida. A média geométrica é ideal para analisar grandezas que envolvem multiplicação ou variação exponencial, como no caso do desempenho atlético em competições.

A média geométrica é amplamente utilizada em diversas áreas, como:

- Cálculo de Taxas de Crescimento: Ao analisar o crescimento de uma variável ao longo do tempo, como no caso do crescimento populacional ou da valorização de um ativo financeiro, a média geométrica oferece uma maneira precisa de calcular a taxa média de crescimento.

- Índices Econômicos: No setor financeiro, a média geométrica é empregada para calcular a rentabilidade média de investimentos ao longo de um período, levando em conta o efeito composto das taxas de crescimento.
- Análise de Dados em Escala Logarítmica: Em áreas como biologia, química e física, a média geométrica é usada para representar dados que seguem uma escala logarítmica, como a intensidade de radiação ou a concentração de substâncias em reações químicas.

Exemplo 3.8.11. *Suponha que um investidor tenha aplicado seu dinheiro em três diferentes investimentos financeiros ao longo de três anos consecutivos. Ele deseja calcular a rentabilidade média anual desses ativos, mas como os retornos de cada ano são reinvestidos, ou seja, os rendimentos são multiplicativos, a média geométrica* ($\bar{x}_g$) *é a medida adequada para essa situação.*

Os retornos anuais desses ativos foram os seguintes:

- *No primeiro ano, o retorno foi de 10% (ou* $100\% + 10\% = 110\% = 1,10$ *em termos multiplicativos);*
- *No segundo ano, o retorno foi de -5% (ou* $100\% - 5\% = 95\% = 0,95$ *em termos multiplicativos).*
- *No terceiro ano, o retorno foi de 8% (ou* $100\% + 8\% = 108\% = 1,08$ *em termos multiplicativos).*

Para calcular a rentabilidade média anual dos três anos, utilizamos a fórmula da média geométrica (Equação 3.16) . Substituímos x_i *pelos valores dos retornos anuais e* n *pelo número total de anos. Dessa forma,*

$$\bar{x}_g = \sqrt[3]{1,10 \cdot 0,95 \cdot 1,08}$$
$$\bar{x}_g = \sqrt[3]{1,1286}$$
$$\bar{x}_g \approx 1,0412 \quad .$$

Portanto, a rentabilidade média anual, em termos multiplicativos, é aproximadamente $1,0412$*.*

Convertendo para porcentagem:

$$\bar{x}_g = 1,0412 - 1 = 0,0412 \ ou \ 4,12\% \quad .$$

A rentabilidade média anual desse investidor foi de aproximadamente $4,12\%$ *ao ano. Ou seja, apesar de um retorno negativo no segundo ano, o investidor obteve um crescimento médio de* $4,12\%$ *por ano ao longo do período de três anos.*

Exemplo 3.8.12. *O Instituto Brasileiro de Geografia e Estatística (IBGE) calcula a **Taxa Média Geométrica de Crescimento Anual** (TGCA) da população brasileira entre os censos demográficos. Essa taxa reflete o ritmo anual de crescimento populacional durante o período entre dois censos sucessivos, considerando o efeito composto do crescimento populacional. Em 2023, o IBGE divulgou os resultados do último censo da população de Nova Venécia, no Espírito Santo (ver Figura 23):*

Figura 23 – Censo de 2022 da cidade de Nova Venécia/ES

Fonte: IBGE. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/es/nova-venecia.html>. Acesso em: 23 mar. 2025.

No censo anterior, realizado em 2010, a população era de 46.031 habitantes. Vale ressaltar que o intervalo entre os censos foi de doze anos, considerando que a pandemia de COVID-19, ocorrida em 2020, provocou o adiamento da coleta de dados. Com base nos dois últimos censos, a TGCA pode ser calculada utilizando a fórmula da média geométrica.

Primeiramente calculamos a Taxa de Crescimento Total (TCT) entre os 12 anos:

$$TCT = \frac{49.065}{46.031} \approx 1,0659 \ .$$

A partir desse valor, calcula-se a Taxa de Crescimento Anual (TGA) utilizando a média geométrica:

$$TGCA = \bar{x}_g - 1 = \sqrt[12]{1,0659} - 1 \approx 1,00533 - 1 = 0,00533 \ ou \ 0,533\% \ .$$

Assim, a TGCA da cidade de Nova Venécia/ES é de aproximadamente 0,533% ao ano, indicando que a população cresceu, em média, 0,533% ao ano durante o período de 2010 a 2022. A aplicação da média geométrica, por meio da TGCA, permite uma análise detalhada do crescimento populacional ao longo do tempo, essencial para o desenvolvimento de estratégias eficazes de planejamento e gestão pública.

Para justificar o uso da média geométrica, observe que:

$$\bar{x}_g = \sqrt[12]{\frac{49.065}{46.031}} \approx \sqrt[12]{1,0659} \ .$$

Denotando $P_0 = 46.031$ *como a população em 2010, e* $P_{12} = 49.065$ *como a população em 2022, temos as taxas de crescimento anuais sucessivas dadas por:*

$$\frac{P_1}{P_0}, \frac{P_2}{P_1}, \ldots, \frac{P_{12}}{P_{11}} .$$

Logo, a média geométrica dessas razões é:

$$\bar{x}_g = \sqrt[12]{\frac{P_1}{P_0} \cdot \frac{P_2}{P_1} \cdot \ldots \cdot \frac{P_{11}}{P_{10}} \cdot \frac{P_{12}}{P_{11}}} = \sqrt[12]{\frac{P_{12}}{P_0}} = \sqrt[12]{\frac{49.065}{46.031}} = \sqrt[12]{1,0659} .$$

Essa abordagem ilustra uma vantagem essencial da média geométrica: quando lidamos com várias taxas de variação sucessivas expressas como razões (como no caso de crescimento populacional composto), basta conhecer o valor inicial e o final para calcular a taxa média de crescimento ao longo do período.

4. **Média Harmônica**

A média harmônica é uma medida estatística utilizada quando se trabalha com conjuntos de dados que envolvem taxas, razões ou valores inversamente proporcionais. Ao contrário da média aritmética, que calcula a soma dos valores e divide pelo número de elementos, a média harmônica leva em consideração os inversos dos valores, o que a torna particularmente útil em situações onde os dados estão relacionados a fenômenos que envolvem razões ou fluxos.

A fórmula da média harmônica para um conjunto de n números $x_1, x_2, \ldots, x_n$ é dada por:

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{3.17}$$

ou

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} \ .$$

Onde:

- $\bar{x}_h$ é a média harmônica,
- x_i representa os valores das observações,
- n é o número total de observações.

Essa medida é comumente aplicada em áreas como economia, física, finanças e engenharia, especialmente quando se deseja calcular a média de grandezas como velocidades, eficiências ou taxas de retorno. Devido à sua natureza, a média harmônica tende a ser mais sensível aos valores menores do que as outras médias, o que a torna uma ferramenta valiosa para situações em que se pretende enfatizar os valores extremos ou aqueles que têm maior impacto no resultado final.

Exemplo 3.8.13. *Suponha que um carro percorra duas distâncias com velocidades diferentes:*

- *Primeira distância: 120 km a 50 km/h.*
- *Segunda distância: 120 km a 70 km/h.*

A velocidade média harmônica para o percurso total pode ser calculada utilizando a fórmula da média harmônica, pois as distâncias são iguais.

Utilizando a fórmula da média harmônica, substituindo $n = 2$ *e os dados observados* x_i *por 50 e 70 temos:*

$$\bar{v}_h = \frac{2}{\frac{1}{50} + \frac{1}{70}} = \frac{2}{\frac{7}{350} + \frac{5}{350}} = \frac{2}{\frac{12}{350}} = \frac{700}{12} \approx 58,33\, km/h \ .$$

Assim a média harmônica fornece uma velocidade média de "58,33 km/h", considerando que as distâncias percorridas são iguais e que o tempo para cada trecho depende inversamente das velocidades. A média harmônica é apropriada neste caso, pois lida com taxas, como a velocidade, e leva em consideração a maior "desvantagem" das velocidades mais baixas, resultando em uma velocidade média menor do que a média aritmética simples.

A origem da média harmônica nesse contexto pode ser compreendida pela fórmula da velocidade média geral. Sabendo que a velocidade média V_m é definida pela razão entre a distância total e o tempo total, ou seja,

$$V_m = \frac{d}{t} \tag{3.18}$$

considere um veículo que percorre duas distâncias d_1 e d_2 , com velocidades médias V_1 e V_2 , respectivamente. O tempo total será:

$$t = \tfrac{d_1}{V_1} + \tfrac{d_2}{V_2}$$

e a velocidade média total será:

$$V_m = \frac{d_1 + d_2}{\frac{d_1}{V_1} + \frac{d_2}{V_2}} = \frac{V_1 V_2 (d_1 + d_2)}{V_1 d_2 + V_2 d_1}\,. \tag{3.19}$$

A Equação 3.19 representa uma **forma ponderada da média harmônica**. Em particular, quando as distâncias são iguais ($d_1 = d_2 = d$), essa fórmula se reduz à média harmônica simples das velocidades.

$$V_m = \frac{d_1 + d_2}{\frac{d_1}{V_1} + \frac{d_2}{V_2}} = \frac{d + d}{\frac{d}{V_1} + \frac{d}{V_2}} = \frac{2d}{\frac{d}{V_1} + \frac{d}{V_2}} = \frac{2}{\frac{1}{V_1} + \frac{1}{V_2}}\,.$$

Assim, a média harmônica surge naturalmente da definição física de velocidade média quando as distâncias são fixas e os tempos são variáveis em função das velocidades.

Exemplo 3.8.14. *Considere que um trabalhador execute três tarefas distintas em uma fábrica, com tempos e quantidades produzidas específicas:*

- *Tarefa 1: 30 minutos para produzir 12 unidades.*
- *Tarefa 2: 40 minutos para produzir 18 unidades.*
- *Tarefa 3: 50 minutos para produzir 25 unidades.*

Queremos descobrir eficiência média do trabalhador, levando em conta a quantidade de unidades produzidas em relação ao tempo gasto em cada tarefa. A eficiência é determinada pela razão entre o número de unidades produzidas e o tempo utilizado. Quando os tempos de execução de diferentes tarefas variam, a média harmônica é a medida mais apropriada para calcular a eficiência média, isso porque ela reflete corretamente o inverso dos tempos, ajustando o impacto das variações nas tarefas.

Primeiramente antes de aplicar a média harmônica nesse contexto, temos que calcular a eficiência de cada tarefa fazendo a razão entre as unidades produzidas em relação ao tempo.

$$\text{Eficiência}_1 = \frac{12}{30} = 0,4 \text{ unidades por minuto.}$$
$$\text{Eficiência}_2 = \frac{18}{40} = 0,45 \text{ unidades por minuto.}$$
$$\text{Eficiência}_2 = \frac{25}{50} = 0,5 \text{ unidades por minuto.}$$

A partir desses valores, calcula-se a média harmônica da eficiência:

$$\bar{x}_h = \frac{3}{\frac{1}{0,4} + \frac{1}{0,45} + \frac{1}{0,5}}$$

$$\bar{x}_h = \frac{3}{2,5 + 2,2222 + 2}$$

$$\bar{x}_h = \frac{3}{6,7222} \approx 0,446 \text{ unidades por minuto}\,.$$

Portanto, a eficiência média do trabalhador é de aproximadamente "0,446 unidades por minuto". Nessa média consideramos a variabilidade dos tempos de produção, proporcionando uma medida mais precisa da eficiência global do trabalhador. Diferente da média aritmética, que seria mais sensível a valores extremos e não refletiria adequadamente o impacto do tempo em cada tarefa.

Exemplo 3.8.15. *Considere que um motorista viajou por dois tipos de trajeto: no primeiro, ele dirigiu 100 km na cidade, onde o carro consumiu 12 litros de combustível. No segundo trajeto, ele dirigiu 200 km na estrada, onde o consumo foi de 14 litros. Queremos determinar qual foi o consumo médio de combustível (em litros por quilômetro) para a viagem inteira. Como temos uma relação inversa entre a distância percorrida e o consumo de combustível a média mais indicada é a harmônica.*

Primeiramente vamos calcular o consumo de combustível em cada trajeto da seguinte maneira:

- *1º trajeto: Consumo de combustível*

$$c_1 = \tfrac{12}{100} = 0,12 \text{ litros por km;}$$

- *2º trajeto: Consumo de combustível*

$$c_2 = \tfrac{14}{200} = 0,07 \text{ litros por km.}$$

Utilizando os valores acima na fórmula da média harmônica, temos:

$$\bar{x}_h = \frac{2}{\frac{1}{0,12} + \frac{1}{0,07}}$$

$$\bar{x}_h = \frac{2}{8,33... + 14,2857}$$

$$\bar{x}_h \approx 0,0884 \text{ litros por km.}$$

Portanto o consumo médio de combustível para a viagem inteira foi de aproximadamente "0,088 litros por quilômetro". Isso nos dá uma ideia de quão eficiente foi o carro considerando os dois tipos de trajeto.

Em resumo, ao considerar os quatro tipos de médias, temos a seguinte tabela:

Tabela 16 – Resumo dos Tipos de Médias

Tipo de Média	**Fórmula**	**Contexto de Uso**	**Características**
Aritmética	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	Usada quando os dados são **iguais ou têm o mesmo peso**. Aplica-se a **quantidades constantes ou uniformes**.	Caracteriza-se por ser a média mais simples, sendo adequada para dados que possuem a mesma importância.
Ponderada	$\bar{x}_p = \frac{\sum w_i x_i}{\sum w_i}$	Usada quando os dados têm **pesos diferentes**. Considera que algumas observações têm mais importância que outras.	Considera a "importância relativa" das observações, quando há diferentes pesos associados aos dados.
Geométrica	$\bar{x}_g = (\prod_{i=1}^{n} x_i)^{1/n}$	Usada para dados que envolvem **taxas de crescimento ou variação percentual**. Adequada para valores com uma **escala multiplicativa**.	Utilizada para calcular médias de "crescimento exponencial" ou "acréscimos percentuais".
Harmônica	$\bar{x}_h = \frac{n}{\sum \frac{1}{x_i}}$	Usada quando o dado é uma **taxa ou razão**. Ideal para **média de taxas inversas**, como em cálculos de velocidade, eficiência e consumo.	Adequada para dados relacionados a "taxas inversas", como tempo ou velocidade, onde valores menores têm maior impacto.

Fonte: Produção da própria autora (2025).

Relações entre as Médias

As médias aritmética ($\bar{x}$), geométrica ($\bar{x_g}$) e harmônica ($\bar{x_h}$) apresentam uma relação de desigualdade fundamental, válida para qualquer conjunto de números reais positivos $x_1, x_2, \ldots, x_n$. Essa relação pode ser expressa como:

$$\bar{x_h} \leq \bar{x_g} \leq \bar{x} \,.$$

Ou seja, a média harmônica é sempre menor ou igual à média geométrica, que por sua vez é menor ou igual à média aritmética.

Essa desigualdade é consequência das propriedades matemáticas das médias e está associada à chamada *desigualdade entre as médias*, também relacionada à desigualdade de Jensen. A intuição por trás dessa relação é que:

- A média aritmética dá mais peso aos valores altos;
- A média geométrica pondera de maneira mais equilibrada os valores;
- A média harmônica dá mais peso aos valores baixos.

Como resultado, a média harmônica tende a ser mais sensível a valores pequenos, enquanto a aritmética é mais influenciada por valores extremos elevados.

Exemplo 3.8.16. *Velocidade Média em um Percurso - Considere um carro que percorre dois trechos iguais de 60 km cada, porém em velocidades diferentes:*

- ***Trecho 1:*** *60 km a uma velocidade constante de 60 km/h;*
- ***Trecho 2:*** *60 km a uma velocidade constante de 120 km/h.*

Queremos analisar as médias aritmética, geométrica e harmônica dessas duas velocidades.

$$\bar{x} = \frac{60+120}{2} = \frac{180}{2} = 90 \ km/h\,.$$

$$\bar{x_g} = \sqrt{60 \cdot 120} = \sqrt{7200} \approx 84,85 \ km/h\,.$$

$$\bar{x_h} = \frac{2}{\frac{1}{60}+\frac{1}{120}} = \frac{2}{\frac{3}{120}} = \frac{2 \cdot 120}{3} = 80 \ km/h\,.$$

Comparando os resultados das Médias

$$80 \leq 84,85 \leq 90$$
$$i.e$$
$$\bar{x_h} \leq \bar{x_g} \leq \bar{x}$$

Neste cenário:

- *A **média harmônica** reflete a **velocidade média real** quando as distâncias são iguais.*

- *A* ***média aritmética*** *considera apenas os valores e desconsidera o tempo gasto em cada trecho, superestimando a média real.*
- *A* ***média geométrica*** *oferece um valor intermediário.*

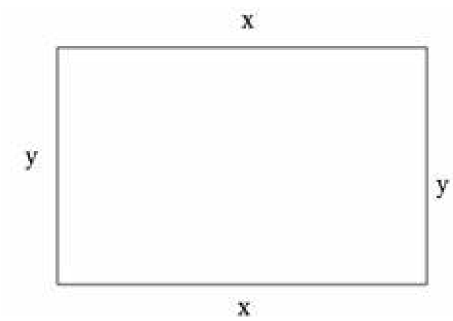
Dessa forma, a desigualdade entre as médias não apenas evidencia propriedades matemáticas fundamentais, como também oferece ferramentas analíticas valiosas para a resolução de problemas em diversos contextos. Para uma compreensão mais aprofundada sobre o tema, recomenda-se a leitura do trabalho de Silva (2019), no qual a autora apresenta, entre as páginas 60 e 66, a demonstração formal dessa desigualdade. Além disso, o material dedica um capítulo exclusivo às aplicações das desigualdades das médias, abordando problemas algébricos, geométricos, de máximos e mínimos de funções, bem como situações contextualizadas relacionadas à otimização de grandezas. Essas contribuições reforçam o potencial didático e prático do tema, tanto para o ensino quanto para a modelagem matemática em diferentes níveis de complexidade.

A desigualdade entre a média aritmética e a média geométrica pode ser aplicada em diversos contextos, inclusive na geometria. A seguir, apresenta-se um exemplo que ilustra essa aplicação, relacionando a área e o perímetro de um retângulo.

Exemplo 3.8.17. *Qual a relação entre a área e o perímetro de um retângulo?*

Solução: *Tome que A seja a área e P o perímetro de um retângulo em que os lados medem x e y*

Figura 24 – Retângulo em que os lados medem x e y.



Fonte: Disponível em: <https://brainly.com.br/tarefa/18661100>
. Acesso em: 22 de maio. 2025.

$$A = x \cdot y$$
$$P = 2(x + y) \quad .$$

Usando a desigualdade das médias, em que $\bar{x} \geq \bar{x}_g$:

$$\frac{x+y}{2} \geq \sqrt{xy}$$

$$4\cdot\frac{x+y}{2} \geq 4\cdot\sqrt{xy}$$

$$2(x+y) \geq 4\sqrt{xy}$$

$$2^2(x+y)^2 \geq 4^2\sqrt{xy}^2$$

$$4(x+y)^2 \geq 16xy$$

$$\boxed{P^2 \geq 16A\ .}$$

Esse resultado evidencia uma relação significativa entre as grandezas geométricas do retângulo: embora tanto o perímetro quanto a área estejam diretamente associados às medidas dos lados, a aplicação da desigualdade das médias permite estabelecer que o perímetro ao quadrado é sempre maior ou igual a 16 vezes a área. Essa relação é válida para qualquer retângulo e atinge a igualdade apenas no caso particular em que os lados são iguais, ou seja, quando o retângulo é um quadrado. Dessa forma, a utilização da média aritmética e da média geométrica nesse contexto destaca como a comparação entre médias fornece limites precisos para grandezas envolvidas em figuras geométricas, reforçando um princípio central da otimização: entre todos os retângulos com o mesmo perímetro, o quadrado é o que apresenta a maior área.

Desigualdade das Médias (Aritmética $\geq$ Geométrica)

Proposição 3.8.18. *Se $x_1, x_2, \cdots, x_n$ são números reais positivos, com $n \geq 2$, então*

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdots x_n}\,. \tag{3.20}$$

Caso Particular: Provemos que (3.20) vale para $n = 2$. Temos a sequência de equivalências:

$$(x_1 - x_2)^2 \geq 0 \iff$$

$${x_1}^2 - 2x_1x_2 + {x_2}^2 \geq 0 \iff$$

$${x_1}^2 + {x_2}^2 \geq 2x_1x_2 \iff$$

$$\frac{{x_1}^2 + {x_2}^2}{2} \geq x_1x_2 \iff$$

$$\frac{{x_1}^2 + {x_2}^2}{2} + \frac{2x_1x_2}{2} \geq x_1x_2 + x_1x_2 \iff$$

$$\frac{(x_1 + x_2)^2}{2} \geq 2x_1x_2 \iff$$

$$\frac{(x_1 + x_2)^2}{4} \geq x_1x_2 \iff$$

$$\frac{x_1 + x_2}{2} \geq \sqrt{x_1 x_2} \ .$$

Assim, a desigualdade entre a média aritmética e a média geométrica é válida para dois valores.

Caracterização do Caso de Igualdade

Ida: Suponha que a média aritmética seja igual à média geométrica, ou seja, se

$$\bar{x}_g = \bar{x}, \text{ então } x_1 = x_2 = \cdots = x_n \, .$$

Volta: Analogamente, se $x_1 = x_2 = \cdots = x_n$, então $\bar{x}_g = \bar{x}$.

Pela desigualdade das médias aritmética e geométrica, temos que

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

$$\frac{x_1 + x_1 + \cdots + x_1}{n} \geq \sqrt[n]{x_1 \cdot x_1 \cdots x_1}$$

$$\frac{n \cdot x_1}{n} \geq \sqrt[n]{x_1^n}$$

$$x_1 \geq x_1$$

Logo, $x_1 = x_1$.

3.8.1.3 **Mediana** ($\mathbf{M_d}$)

A mediana é uma medida de tendência central (também uma das chamadas *separatrizes*, as quais vamos estudar e analisar na Seção 3.8.2 deste estudo) que representa o valor que divide um conjunto de dados em duas partes iguais. Em outras palavras, é o valor central de uma distribuição de dados.

Ao contrário da média, que pode ser influenciada por valores extremos (outliers), a mediana é mais robusta e confiável em situações onde os dados apresentam grande variação (distribuições assimétricas) ou valores discrepantes. Nesse caso, quando a amostra contém valores muito maiores ou menores do que os demais, a mediana se torna a medida de posição mais adequada para descrever a tendência central dos dados.

Para o estudo da mediana, consideraremos as diferentes maneiras de organizar os dados:

- **ROL de dados brutos e Tabela de distribuição de frequência simples**

  Primeiramente, os n dados devem ser organizados em ordem crescente ou decrescente.

  Para determinar a posição da mediana, basta calcular: $\frac{n+1}{2}$.
  Vale ressaltar que, quando n é ímpar, a posição da mediana corresponderá a um único termo, ou seja, a mediana será o termo central da distribuição e, nesse caso, fará parte do conjunto de valores observado. Por outro lado, quando n é par, a

mediana estará localizada entre dois termos centrais da distribuição, e será calculada como a média aritmética desses dois termos, podendo não pertencer diretamente ao conjunto de valores observados.

Exemplo 3.8.19. *No Exemplo 3.4.2, temos as idades (em anos) de dez alunos de uma turma de ensino médio:*

15 15 15 16 16 16 16 16 17 19

Como $n = 10$ *é par, a posição da mediana será dada por* $\dfrac{10+1}{2} = 5,5^{\circ}$.
Isso significa que a mediana está entre o 5° *e o* 6° *termos, que são ambos iguais a 16. A média aritmética de dois termos iguais é o próprio termo (*$\bar{X} = \dfrac{16+16}{2} = 16$*), portanto, a mediana desse conjunto de dados é 16, ou seja,* $M_d = 16$. *O que significa que, em uma turma de dez alunos, metade dos alunos tem idades abaixo de 16 anos, e a outra metade tem idades acima de 16 anos. Ou seja, 50% dos alunos têm 16 anos ou menos, e 50% têm 16 anos ou mais.*

Exemplo 3.8.20. *Caso os dados apresentados no Exemplo 3.8.19 estivessem organizados por meio de uma tabela de distribuição de frequência simples, a análise da mediana poderia ser facilitada com o uso da coluna de frequência absoluta acumulada (*f_{ia}*).*

A Tabela 17 apresenta a distribuição de frequências da variável idade, permitindo a identificação mais direta da posição central dos dados.

Tabela 17 – Distribuição de frequências da variável idade

i	**Idade(anos)**	**Freq. Abs** (f_i)	**Freq. Rel.** (f_{ia})
1	15	3	3
2	16	5	8
3	17	1	9
4	18	0	9
5	19	1	10
	Total	10	

Fonte: Produção da própria autora (2024).

Prosseguimos da mesma forma: primeiro, calculamos a posição da mediana, dada por $\dfrac{10+1}{2} = 5,5^{\circ}$, *como no Exemplo 3.8.19.*
Isso indica que a mediana está entre o 5° *e o* 6° *termos. Ao observar a coluna de frequência absoluta acumulada (*f_{ia}*), podemos perceber que do* 4° *ao* 8° *termo, todos pertencem à segunda classe, correspondente à idade de 16 anos. Isso nos leva a concluir que a mediana (*M_d*) é 16, ou seja,* $M_d = 16$.

Exemplo 3.8.21. *No Exemplo 3.4.4, foi coletado o número de filhos de uma amostra de 25 funcionários de uma empresa, conforme tabulado na Tabela 4. Vamos adicionar a coluna f_{ia} para facilitar a identificação da classe da mediana por meio de sua posição.*

Tabela 18 – Nº de Filhos dos funcionários (alterada)

Número de Filhos	f_i	f_{ia}
0	5	5
1	6	11
2	7	18
3	4	22
4	3	25
Total	25	

Fonte: Produção da própria autora (2024).

Prosseguimos da mesma forma: primeiro, calculamos a posição da mediana, dada por $\dfrac{25+1}{2} = 13°$.
Isso indica que a mediana está na décima terceira posição dos dados. Observando a coluna de frequência absoluta acumulada (f_{ia}), podemos perceber que os dados que ocupam as posições do $12°$ *ao* $18°$ *pertencem à terceira classe, correspondente ao número de 2 filhos. Portanto, concluímos que a mediana (M_d) é 2, ou seja,* $M_d = 2$*. O indica que, entre os 25 funcionários da empresa, metade deles tem 2 filhos ou menos, e a outra metade tem 2 filhos ou mais. Em outras palavras, 50% dos funcionários têm até 2 filhos, e 50% têm mais de 2 filhos.*

- **Tabela de frequência de dados agrupados**

  No caso de dados agrupados em tabelas de frequência, o cálculo da mediana envolve a identificação da classe mediana, que é a classe que contém o valor central. Para calcular a mediana, os seguintes passos podem ser seguidos:

  1. Determinar as frequências acumuladas de cada classe;
  2. Calcular a posição da mediana usando a fórmula $\dfrac{n+1}{2}$, onde n é o total de dados, ou seja, o somatório das frequências absolutas de todas as classes;
  3. Identificar a classe correspondente à frequência acumulada que é imediatamente superior à posição calculada. Esta classe é chamada de classe mediana. Após isso, aplicar a fórmula:

$$M_d = l_i + \left(\frac{\frac{n}{2} - f_{(i-1)a}}{f_i}\right) \cdot h\,. \tag{3.21}$$

  Onde,

l_i = Limite inferior da classe mediana.
n = Total de observações.
$f_{(i-1)a}$ = Frequência acumulada antes da classe mediana.
f_i = Frequência da classe mediana.
h = Amplitude da classe.

Exemplo 3.8.22. *No Exemplo 3.4.5, consideramos a coleta das idades de 25 funcionários de uma empresa. Com base nesses dados, foi elaborada uma tabela de frequência com classes, apresentada na Tabela 9, a qual servirá de base para o cálculo da mediana das idades.*

Conforme os procedimentos descritos anteriormente, iniciaremos pela obtenção das frequências absolutas acumuladas (f_{ia}) para cada classe, já dispostas na Tabela 19, o que facilita a identificação da classe mediana e a aplicação da fórmula correspondente (Equação 3.21).

Tabela 19 – Distribuição de freq. da variável idade (alterada)

i	**Idade**	f_i	f_{ia}
1	25 |- 30	5	5
2	30 |- 35	5	10
3	35 |- 40	5	15
4	40 |- 45	5	20
5	45 |- 50	4	24
6	50 |- 55	1	25
	Total	25	

Fonte: Produção da própria autora (2024).

Em seguida, calculamos a posição da mediana utilizando a fórmula:

$$\frac{n+1}{2} = \frac{25+1}{2} = 13.$$

Isso significa que a classe mediana é aquela que contém o dado na 13ª posição da amostra. Observando as frequências acumuladas, vemos que a classe mediana corresponde à terceira classe, pois os dados nas posições do 11º ao 15º estão contidos nela.

Agora, coletamos os seguintes dados para aplicar na fórmula da mediana (Equação 3.21):

$$l_3 = 35;$$
$$n = 25;$$

$$f_{2a} = 10;$$
$$f_3 = 5;$$
$$h = 5$$

Substituindo esses valores na fórmula, temos:

$$M_d = 35 + \left(\frac{\frac{25}{2} - 10}{5}\right) \cdot 5$$
$$M_d = 35 + \left(\frac{12,5 - 10}{5}\right) \cdot 5$$
$$M_d = 35 + \left(\frac{2,5}{5}\right) \cdot 5$$
$$M_d = 37,5 \quad .$$

Portanto, a mediana das idades dos funcionários é 37,5 anos. Esse valor indica que, em uma amostra de 25 funcionários, metade dos funcionários tem idades abaixo de 37 anos e 6 meses e a outra metade tem idades acima desse valor.

A mediana não leva em consideração todos os valores do conjunto de dados, como acontece com a média. Por isso, em distribuições simétricas, a média pode ser mais informativa, pois usa todas as observações. No entanto, em distribuições assimétricas ou com dados extremos, a mediana se torna uma medida mais robusta e confiável, já que não é tão influenciada por outliers.

Além disso, como discutido anteriormente, o cálculo da mediana por meio de gráficos segue os mesmos princípios aplicados às tabelas de frequência. Em gráficos como o histograma, a mediana pode ser determinada visualmente ou com interpolação, localizando a classe que contém a mediana com base na área acumulada das frequências. Esse processo gráfico é análogo ao método utilizado com as tabelas, mas oferece uma representação visual das informações e uma forma alternativa de calcular a mediana de maneira mais prática.

3.8.2 Medidas Separatrizes

As medidas separatrizes são ferramentas essenciais na estatística para analisar e descrever a distribuição dos dados de um conjunto. Elas têm a função de dividir o conjunto de dados em segmentos ou partes, o que facilita a compreensão da dispersão, simetria e concentração dos valores. As medidas separatrizes mais comuns são aquelas que dividem o rol (dados ordenados) ao meio (**mediana**), em quatro partes (**quartis**), em dez partes (**decil**) ou em cem partes (**percentis ou centis**). É importante destacar que, para sua aplicação correta, os dados precisam estar organizados em ordem crescente.

Na seção anterior, estudamos a mediana, e agora vamos abordar as demais medidas separatrizes:

- **Quartis**

  Os quartis são medidas separatrizes que dividem um conjunto de dados em quatro partes iguais. Eles são compostos por três valores:

  1. **Primeiro Quartil** (Q_1) - valor situado de tal modo na série de dados que 25% das observações são menores que ele e 75% são maiores;
  2. **Segundo Quartil** (Q_2) - valor situado de tal modo na série de dados que 50% das observações são menores que ele e 50% são maiores (o que corresponde ao valor da mediana);
  3. **Terceiro Quartil** (Q_3) - valor situado de tal modo na série de dados que 75% das observações são menores que ele e 25% são maiores

Figura 25 – Representação Gráfica dos Quartis em um conjunto de dados

Fonte: Produção da própria autora (2025).

A diferença entre o terceiro e o primeiro quartil, conhecida como intervalo interquartil (**IIQ**), é uma medida relevante de dispersão, pois indica a faixa onde 50% dos dados estão agrupados. O IIQ é menos influenciado por outliers em relação ao desvio padrão (que será abordado na Seção 3.8.3 desse estudo), tornando-o uma medida útil para distribuições assimétricas.

Exemplo 3.8.23. *Considere que uma empresa de transporte de mercadorias deseja analisar o tempo de entrega de pacotes em uma determinada região. Eles têm os dados de tempo de entrega de 200 pacotes, medidos em horas, e querem entender melhor a distribuição dos tempos para identificar possíveis melhorias no processo.*

Primeiro, os tempos de entrega são ordenados de forma crescente. Em seguida, a empresa calcula os quartis:

- *Suponha que Q_1 seja 2 horas. Isso significa que 25% dos pacotes foram entregues em menos de 2 horas;*
- *Digamos que Q_2 seja 4,5 horas. Isso indica que 50% dos pacotes foram entregues em até 4,5 horas e 50% foram entregues após esse tempo;*

– *Se Q_3 for 7 horas, isso significa que 75% dos pacotes foram entregues em até 7 horas, e apenas 25% levaram mais tempo do que isso.*

Após os resultados dos quartis, a empresa consegue ter uma visão clara de como os tempos de entrega estão distribuídos. O intervalo entre o primeiro e o terceiro quartil (IIQ) será de 5 horas, mostrando que a maioria dos pacotes (50%) foram entregues em um intervalo de 2 a 7 horas.

Essa análise ajuda a empresa a entender onde a maior parte das entregas se concentra e identificar pontos de melhoria. Por exemplo, se uma porcentagem significativa dos pacotes está sendo entregue muito rapidamente (abaixo de 2 horas), pode ser uma oportunidade para melhorar a logística e focar em otimizar as entregas mais lentas. Além disso, o intervalo interquartil pode ajudar a empresa a perceber se há outliers, como pacotes entregues com muito mais demora, que precisam de uma análise mais aprofundada.

Exemplo 3.8.24. *Suponha que temos os tempos de entrega (em horas) de 10 pacotes em uma pequena empresa. Os tempos de entrega são os seguintes:*

$$6, 8, 7, 5, 9, 6, 4, 8, 7, 5 \ .$$

Para calcular os quartis, o primeiro passo é organizar os dados em ordem crescente. Isso pode ser feito por meio de um rol ou de tabelas de frequência. Como temos poucos dados, vamos organizá-los em um rol:

$$4, 5, 5, 6, 6, 7, 7, 8, 8, 9 \ .$$

O próximo passo é calcular a Q_2 (mediana). Como temos 10 números, a mediana será a média dos 5º e 6º valores, que são 6 e 7, respectivamente:

$$Q_2 = \frac{6+7}{2} = 6,5 \ .$$

Agora, para determinar o Q_1, basta calcular a mediana da primeira metade dos dados (valores à esquerda da mediana). A primeira metade dos dados é: $4, 5, 5, 6, 6$. *A mediana dessa metade é o 3º valor, que é 5.*

Logo, $Q_1 = 5$.

Analogamente, para determinar o Q_3, calcula-se a mediana da segunda metade dos dados (valores à direita da mediana). A segunda metade dos dados é: $7, 7, 8, 8, 9$. *A mediana dessa metade é o 3º valor, que é 8.*

Portanto, $Q_3 = 8$.

*Por fim, o **intervalo interquartil (IIQ)** diferença entre o terceiro e o primeiro quartil:*

$$IIQ = Q_3 - Q_1 = 8 - 5 = 3 \ .$$

O $IIQ = 3$ indica que 50% dos pacotes foram entregues dentro de um intervalo de 3 horas, entre 5 e 8 horas.

Esse exemplo ilustra como os quartis ajudam a analisar a distribuição dos tempos de entrega, mostrando a concentração dos dados e a dispersão.

Para calcular os quartis (Q_1, Q_2, Q_3) com dados apresentados em tabela de frequência com dados agrupados, basta seguir ose seguintes passos:

1. Localização do Quartil

   A localização do q-ésimo quartil, que é dado por L_q, é calculada pela fórmula:

   $$L_q = \frac{q \cdot n}{4} \ . \tag{3.22}$$

   Onde:

   q é o número do quartil, variando de 1 a 3 (por exemplo, 25% para Q_1, 50% para Q_2, etc.);

   n é o número total de observações (a soma das frequências absolutas).

2. Identificar o Intervalo do Quartil

   Com a localização L_q, identifique o intervalo de classe que contém esse valor. Em seguida, calcule o valor do quartil utilizando a fórmula de interpolação.

3. Fórmula de Interpolação

   A interpolação para o cálculo do q-ésimo quartil no intervalo de classe é dada por:

   $$Q_q = l_i + \left(\frac{L_q - f_{(i-1)a}}{f_i} \right) \cdot h \ . \tag{3.23}$$

   Onde:

   Q_q é o valor do q-ésimo quartil;

   l_i é o limite inferior da classe que contém o quartil;

   L_q é a localização do quartil;

   $f_{(i-1)a}$ é a frequência acumulada antes da classe que contém o quartil;

   f_i é a frequência da classe que contém o quartil;

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

Exemplo 3.8.25. *Considere a tabela de frequências agrupadas apresentada na Tabela 20, que mostra a distribuição das idades de um grupo de 30 pessoas.*

Agora, vamos calcular os quartis Q_1, Q_2 e Q_3.

Tabela 20 – Distribuição de frequência das idades de um grupo de pessoas

i	**Idade**	f_i	f_a
1	10 |- 19	4	4
2	19 |- 28	6	10
3	28 |- 37	5	15
4	37 |- 46	8	23
5	46 |- 55	7	30
	Total	30	

Fonte: Produção da própria autora (2025).

1. *Cálculo do Primeiro Quartil.*

   Primeiramente calcula-se a posição de Q_1 dada por:

   $$L_1 = \frac{1 \cdot 30}{4} = 7,5 \ \ .$$

   A posição $7,5$ está dentro da segunda classe ($19| - 28$). Para calcular o valor de Q_1, usamos a fórmula de interpolação.

   Sendo:

   $l_2 = 19$*(limite inferior do intervalo);*
   $L_1 = 7,5$ *(posição de Q_1);*
   $f_{(2-1)a} = 4$ *(frequência acumulada antes do intervalo $19| - 28$);*
   $f_2 = 6$ *(frequência da classe $19| - 28$);*
   $h = 9$ *(amplitude da classe).*

   Substituindo na fórmula de interpolação:

   $$Q_1 = 19 + \left(\frac{7,5 - 4}{6}\right) \cdot 9 = 19 + \left(\frac{3,5}{6}\right) \cdot 9 = 19 + 5,25 = 24,25\,.$$

   Portanto, $Q_1 = 24,25$.

2. *Cálculo do Segundo Quartil (mediana).*

   Analogamente, primeiro calculamos a posição de Q_2 dada por:

   $$L_2 = \frac{2 \cdot 30}{4} = 15 \ \ .$$

   A posição 15 está dentro da terceira classe ($28| - 37$). Para calcular o valor de Q_2, usamos a fórmula de interpolação.

   Sendo:

   $l_3 = 28$*(limite inferior do intervalo);*
   $L_2 = 15$ *(posição de Q_2);*

$f_{(3-1)a} = 10$ *(frequência acumulada antes do intervalo* $28| - 37$*);*
$f_3 = 5$ *(frequência da classe* $28| - 37$*);*
$h = 9$ *(amplitude da classe).*

Substituindo na fórmula de interpolação:

$$Q_2 = 28 + \left(\frac{15-10}{5}\right) \cdot 9 = 28 + \left(\frac{5}{5}\right) \cdot 9 = 28 + 9 = 37\,.$$

Portanto, $Q_2 = 37$*.*

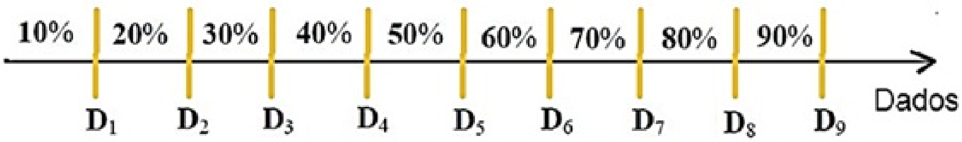
3. *Cálculo do Terceiro Quartil.*

   Seguindo os passos anteriores para o cálculo de Q_3*, que se encontra na quarta classe (*$37| - 46$*) obtemos que que* $Q_3 \approx 45,44$*.*

- **Decis**

  Os decis dividem um conjunto de dados ordenado em 10 partes iguais. Eles são valores que marcam as divisões do conjunto de dados, de forma que cada décimo (ou decil) representa 10% dos dados. O primeiro decil (D_1) é o valor abaixo do qual estão 10% dos dados. O segundo decil (D_2) marca o valor abaixo do qual estão 20% dos dados, e assim por diante, até o nono decil (D_9), que é o valor abaixo do qual 90% dos dados se encontram.

  Figura 26 – Representação Gráfica dos decis em um conjunto de dados



  Fonte: Produção da própria autora (2025).

  O uso de decis pode ser particularmente útil quando é necessário uma análise mais detalhada dos dados, oferecendo uma visão mais precisa da distribuição em comparação aos quartis. Ao dividir os dados em 10 partes, os decis permitem observar variações sutis entre as diferentes porções do conjunto, o que pode ser relevante, por exemplo, para identificar padrões ou outliers que não seriam perceptíveis apenas com quartis.

  O cálculo dos decis segue um procedimento semelhante ao dos quartis. Inicialmente, os dados devem ser ordenados. Em seguida, determina-se a localização L_d, que pode ser calculada pela fórmula:

$$L_d = \frac{d \cdot n}{10}\,. \tag{3.24}$$

Onde:

d é o número do decil, variando de 1 a 9 (por exemplo, 10% para Q_1, 20% para Q_2, etc.);

n é o número total de observações (a soma das frequências absolutas).

Após determinar a localização L_d, identificamos o decil, que pode estar posicionado entre dois números. Nesse caso, o valor do decil é calculado utilizando a média aritmética dos dois números. Caso o decil ocupe uma posição exata, o valor do decil será simplesmente o número correspondente a essa posição.

Este método é aplicável quando os dados estão apresentados em forma de rol ou em tabelas de frequência simples. Quando os dados são apresentados em tabelas de frequência com dados agrupados, a localização do decil é utilizada para identificar a classe em que o decil se encontra. Em seguida, a mesma fórmula de interpolação (Equação 3.23) pode ser utilizada, substituindo Q_q por D_d para representar o decil.

Exemplo 3.8.26. *Considere que uma empresa de médio porte está analisando os salários de seus funcionários para entender melhor a distribuição salarial. A empresa tem 50 funcionários e coletou os dados sobre os salários (em reais) que variam entre 1.000 e 10.000 reais.*

A empresa quer saber qual o valor de salário abaixo do qual 30% dos funcionários ganham menos, ou seja, o terceiro decil (D_3).

Primeiramente, precisamos organizar os salários dos 50 funcionários em ordem crescente, em rol ou por meio de tabela de frequência. Suponha que, após organizar os dados, temos a lista de salários. Não vou listar todos os 50 salários aqui, mas imagine que eles variam de 1.000 a 10.000 reais.

Para calcular a localização do D_3, usamos a Equação 3.24:

$$L_3 = \frac{3 \cdot 50}{10} = 15 \quad .$$

Isso significa que o terceiro decil está na 15^a posição dos dados ordenados.

Agora, devemos localizar o valor na 15^a posição da lista de salários. Vamos supor que, após organizar os dados, o salário na 15^a posição seja de 3.500 reais.

*Portanto, o D_3 é **3.500 reais**.*

Observações:

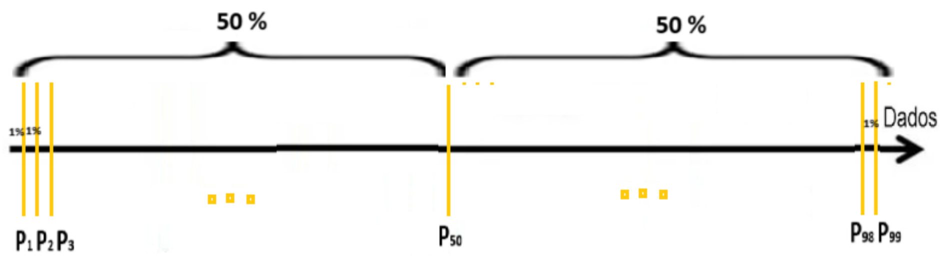
- *D_3: O valor de **3.500 reais** significa que **30% dos funcionários da empresa ganham menos do que 3.500 reais** por mês;*

- *Com base nesse resultado, a empresa pode analisar se a distribuição salarial está adequada e se há necessidade de ajustar as faixas salariais. Por exemplo:*
  - *Se 30% dos funcionários estão abaixo de 3.500 reais, a empresa pode verificar se isso está de acordo com a política salarial da organização ou se há um desnível significativo que precisa ser corrigido.*
  - *A empresa pode também decidir oferecer treinamentos para melhorar a qualificação dos funcionários que ganham menos ou revisar suas políticas de remuneração para garantir uma distribuição mais equitativa.*

- **Percentis ou Centis**

Os percentis são medidas separatrizes que dividem um conjunto de dados ordenados em 100 partes iguais. O p-ésimo percentil é o valor abaixo do qual p% dos dados estão localizados. Por exemplo, o 25° percentil (P_{25}) corresponde ao valor abaixo do qual 25% dos dados estão localizados, o 50° percentil (P_{50}) é a mediana (ou 2° quartil (Q_2) e está abaixo do qual 50% dos dados estão localizados.

Figura 27 – Representação Gráfica dos Percentis em um Conjunto de Dados



Fonte: Produção da própria autora (2025).

Os percentis fornecem uma visão mais detalhada da distribuição dos dados em comparação com outras medidas, como os quartis e os decis. Eles são particularmente úteis em análises onde se deseja uma segmentação mais refinada dos dados.

Analogamente, o cálculo dos Centis segue um procedimento semelhante ao dos quartis. Inicialmente, os dados devem ser ordenados. Em seguida, determina-se a localização L_p, que pode ser calculada pela fórmula:

$$L_p = \frac{p \cdot n}{100}\,. \tag{3.25}$$

Onde:

p é o número do percenti, variando de 1 a 99 (por exemplo, 1% para P_1, 20% para P_2, etc.);

n é o número total de observações (a soma das frequências absolutas).

Após determinar a localização L_p, identificamos o percentil, que pode estar posicionado entre dois números. Nesse caso, o valor do percentil é calculado utilizando a média aritmética dos dois números. Caso o centil ocupe uma posição exata, o valor do centil será simplesmente o número correspondente a essa posição.

Este método é aplicável quando os dados estão apresentados em forma de rol ou em tabelas de frequência simples. Quando os dados são apresentados em tabelas de frequência com dados agrupados, a localização do percentil é utilizada para identificar a classe em que o centil se encontra. Em seguida, a mesma fórmula de interpolação (Equação 3.23) pode ser utilizada, substituindo Q_q por P_p para representar o Percentil.

Exemplo 3.8.27. *Considere que uma comissão organizadora de um concurso público deseja analisar a distribuição das pontuações obtidas pelos candidatos em uma prova objetiva. As notas, que variam de 0 a 100, foram agrupadas e organizadas na tabela de frequência apresentada na Tabela 21.*

Tabela 21 – Distribuição das Notas dos Candidatos

i	**Notas**	f_i	f_a
1	0 |– 20	5	5
2	20 |– 40	8	13
3	40 |– 60	12	25
4	60 |– 80	15	40
5	80 |– 100	10	50
	Total	50	

Fonte: Produção da própria autora (2025).

O interesse da comissão recai sobre a identificação da nota que delimita os 85% dos candidatos com os desempenhos mais baixos, isto é, o $85°$ *percentil* (P_{85})*. Essa informação é fundamental para uma compreensão mais aprofundada do rendimento geral dos participantes, auxiliando na definição de critérios de corte ou na formulação de estratégias de classificação.*

Para o cálculo do P_{85}*, inicialmente identifica-se a classe em que esse percentil se encontra, utilizando-se a fórmula da posição percentual (Equação 3.25). Considerando-se* $p = 85$ *e* $n = 50$*, tem-se:*

$$L_{85} = \frac{85 \cdot 50}{100} = 42,5° \quad .$$

Ou seja, o P_{85} *está na posição* $42,5°$ *dos dados acumulados. Observando a tabela, verifica-se que essa posição corresponde à última classe:*$80| - 100$*. Aplica-se, então, a fórmula do percentil com interpolação:*

$$P_p = l_i + \left(\frac{L_p - f_{a(i-1)}}{f_i}\right) \cdot h$$
$$P_{85} = 80 + \left(\frac{42,5 - 40}{10}\right) \cdot 20$$
$$P_{85} = 80 + \left(\frac{2,5}{10}\right) \cdot 20$$
$$P_{85} = 85 \quad .$$

Assim, conclui-se que 85% dos candidatos obtiveram nota igual ou inferior a 85. Em outras palavras, apenas 15% dos participantes alcançaram pontuação superior a esse valor.

Com base nessa informação, a banca examinadora pode estabelecer a nota mínima exigida para a convocação à segunda fase do concurso, além de avaliar se a prova apresentou um grau de dificuldade compatível com os objetivos do certame. Adicionalmente, o percentil calculado pode servir como parâmetro para a seleção automática de candidatos em programas de reconhecimento por mérito, caso se deseje destacar os 15% com melhor desempenho.

As medidas separatrizes ajudam a representar um conjunto de dados de forma que se possa compreender tanto o comportamento central quanto a dispersão dos valores, sendo de extrema importância para diversas áreas da pesquisa e tomada de decisão.

3.8.3 Medidas de Dispersão

As medidas de dispersão são fundamentais para entender a variabilidade dos dados em torno de uma medida central, como a média. Enquanto a média fornece um valor representativo, ela não revela a diversidade ou concentração dos dados ao seu redor. Em muitos casos, conhecer apenas a média de um conjunto de dados não é suficiente para compreendê-lo por completo.

Exemplo 3.8.28. *Considere os seguintes grupos de alunos, cujas médias aritméticas de notas são todas iguais a 5 ($\bar{x} = 5$) :*

- *Grupo A: {3, 3, 4, 5, 10}*
- *Grupo B: {1, 2, 6, 7, 9}*
- *Grupo C: {5, 5, 5, 5, 5}*
- *Grupo D: {2, 5, 6, 7}*
- *Grupo E: {3, 4, 4, 5, 9}*

Embora todos tenham a mesma média, as distribuições das notas são bem diferentes. O grupo C tem notas idênticas, ou seja, sem variabilidade, enquanto o grupo B apresenta uma ampla dispersão entre 1 e 10.

A partir do exemplo apresentado, observa-se que, embora a média aritmética seja uma medida útil para representar um conjunto de dados, ela não fornece informações sobre a variabilidade entre os valores. Nesse contexto, as medidas de dispersão tornam-se fundamentais, pois permitem avaliar o grau de dispersão dos dados em torno da média, contribuindo para uma compreensão mais aprofundada da distribuição dos valores. Tais medidas são particularmente relevantes em análises comparativas entre diferentes conjuntos de dados, uma vez que dois conjuntos podem apresentar a mesma média, mas comportamentos distintos quanto à dispersão dos seus elementos.

Em áreas como educação, as medidas de dispersão podem indicar que, apesar de uma média satisfatória, há uma grande variação no desempenho dos alunos, sugerindo a necessidade de intervenções. No contexto financeiro, essas medidas ajudam a entender a volatilidade dos preços ou a estabilidade de um investimento.

Portanto, as medidas de dispersão são essenciais para qualificar a variabilidade dos dados, oferecendo uma visão mais completa e precisa do comportamento da variável em análise. Elas vão além da média, proporcionando insights valiosos para a tomada de decisões em diversas áreas.

A **variância** e o **desvio padrão** são duas das principais **medidas de dispersão**. Elas nos ajudam a entender o quanto os dados se afastam da média e ambas possuem versões específicas para populações e amostras.

- A **variância** (σ^2) mede o grau de dispersão dos dados em relação à média. A *variância populacional* é calculada quando se tem acesso a todos os dados da população. Em contrapartida, a *variância amostral* é uma estimativa da variância populacional, obtida a partir de uma amostra.

  Assim, a variância amostral é calculada pela média dos quadrados das diferenças entre cada valor do conjunto de dados e a média amostral. Para um conjunto de dados $X = \{x_1, x_2, \ldots, x_n\}$, a fórmula da variância amostral é dada por:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \,. \tag{3.26}$$

  Onde:

  σ^2 é a variância;

  x_i são os dados observados com $1 \leq i \leq n$;

  n total de dados observados;

$\bar{x}$ é média do conjunto de dados.

- O **desvio padrão** (σ) é a raiz quadrada da variância e indica, em média, o quanto os valores do conjunto se afastam da média. O desvio *padrão populacional* é calculado a partir de todos os dados da população, enquanto o desvio *padrão amostral* é uma estimativa obtida a partir de uma amostra.

  A principal vantagem do desvio padrão é estar expresso na mesma unidade dos dados coletados, o que facilita a interpretação. A fórmula do desvio padrão é dada por:

$$\sigma = \sqrt{\sigma^2}\,. \tag{3.27}$$

Neste trabalho, sempre que mencionarmos variância e desvio padrão, estaremos nos referindo às suas versões amostrais, uma vez que a análise é realizada a partir de amostras representativas das populações de interesse.

Exemplo 3.8.29. *Considere novamente o **Grupo A**, cujas notas são: {3, 3, 4, 5, 10}, com média aritmética $\bar{x} = 5$.*

*Vamos calcular a **variância** e o **desvio padrão** desse grupo.*

Primeiro, determinamos o desvio de cada valor em relação à média (diferença), elevamos ao quadrado, somamos os resultados e dividimos pelo número total de elementos ($n = 5$). Esses cálculos estão organizados na Tabela 22.

Tabela 22 – Cálculo para a Variância das Notas dos Candidatos do Grupo A

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
3	3 - 5 = -2	4
3	3 - 5 = -2	4
4	4 - 5 = -1	1
5	5 - 5 = 0	0
10	10 - 5 = 5	25

Fonte: Produção da própria autora (2025).

Somando os quadrados dos desvios: $4 + 4 + 1 + 0 + 25 = 34$.

Variância:

$$\sigma^2 = \frac{34}{5} = 6,8 \quad .$$

Desvio padrão:

$$\sigma = \sqrt{6,8} \approx 2,61 \quad .$$

Portanto, a variância do Grupo A é 6,8, e o desvio padrão é aproximadamente 2,61. Isso indica que, em média, as notas desse grupo variam cerca de 2,61 pontos em relação à média.

Exemplo 3.8.30. *Consideremos agora o* ***Grupo C****, com as notas: {5, 5, 5, 5, 5}, cuja média aritmética é* $\bar{x} = 5$.

Como todos os valores são iguais à média, não há nenhuma variação entre as notas. Vamos verificar isso por meio do cálculo da variância e do desvio padrão:

Soma dos quadrados dos desvios: $0 + 0 + 0 + 0 + 0 = 0$.

Variância:

$$\sigma^2 = \frac{0}{5} = 0 \ .$$

Desvio padrão:

$$\sigma = \sqrt{0} = 0 \ .$$

Como era esperado, tanto a variância quanto o desvio padrão são iguais a zero, indicando que não há nenhuma dispersão entre os dados. Todos os alunos do Grupo C obtiveram exatamente a mesma nota.

Exemplo 3.8.31. *A variação nos preços de combustíveis é um aspecto relevante para consumidores e órgãos reguladores, especialmente em contextos locais, nos quais fatores como concorrência, logística e políticas fiscais podem gerar diferenças significativas nos valores praticados. Nesse sentido, o Procon-ES realizou um levantamento dos preços do óleo diesel S-10 em postos da Grande Vitória, com o objetivo de monitorar as oscilações de mercado e fornecer dados transparentes à população.*

A Tabela 23 apresenta os preços do óleo diesel S-10 coletados entre os dias 4 e 15 de junho de 2018, em 12 estabelecimentos localizados no município de Vitória.

Com base nesses dados, procederemos ao cálculo das medidas de dispersão dos preços.

Primeiramente, calculamos a média aritmética dos preços que é dada por:

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

$$\bar{x} = \frac{3,699 + 3,459 + 3,339 + 3,496 + \cdots + 3,350 + 3,439 + 3,590 + 3,659}{12}$$

$$\bar{x} = \frac{40,038}{12} \approx 3,3365 \ .$$

Tabela 23 – Preços do Óleo Diesel S-10 em Postos de Combustíveis de Vitória (junho/2018)

Estabelecimento	**Preço (R$)**
Posto Iate	R$ 3,699
Posto Atlântico	R$ 3,459
Posto Oceano	R$ 3,339
Posto Trilha	R$ 3,496
Posto Camburi do Gás	R$ 3,339
Posto Monza	R$ 3,339
Posto Aerovix	R$ 3,65
Posto Eucalipto	R$ 3,339
Posto Leitão	R$ 3,35
Posto Ouro Negro	R$ 3,439
Posto Brasil	R$ 3,59
Posto Moscoso	R$ 3,659

Fonte: PROCON-ES adaptado. Disponível em: <https://procon.es.gov.br/Media/procon/Documentos/Relat%C3%B3rio%20Fiscaliza%C3%A7%C3%A3o%20-%20Levantamento%20de%20pre%C3%A7os%20do%20%C3%B3leo%20diesel%20na%20Grande%20Vit%C3%B3ria%20-%202018.pdf>. Acesso em: 13 abril 2025.

Logo, a média dos preços é R$ 3,3365.

Agora, para calcular a variância, antes determinamos os desvios ao quadrado de cada preço em relação à média. Os resultados dos cálculos dos desvios quadráticos são:

$$(3,699 - 3,3365)^2 = 0,1314$$
$$(3,459 - 3,3365)^2 = 0,015$$
$$(3,339 - 3,3365)^2 = 0,00000625$$
$$(3,496 - 3,3365)^2 = 0,02544$$
$$(3,339 - 3,3365)^2 = 0,00000625$$
$$(3,339 - 3,3365)^2 = 0,00000625$$
$$(3,650 - 3,3365)^2 = 0,09828$$
$$(3,339 - 3,3365)^2 = 0,00000625$$
$$(3,350 - 3,3365)^2 = 0,00018225$$
$$(3,439 - 3,3365)^2 = 0,010506$$
$$(3,590 - 3,3365)^2 = 0,06426$$
$$(3,659 - 3,3365)^2 = 0,104001 \quad .$$

Soma dos desvios ao quadrado:

$$\sum(x_i - \bar{x})^2 \approx 0,44909 \quad .$$

Assim, substituindo os valores na Equação 3.26 da Variância:

$$\sigma^2 = \frac{0,44909}{11} \approx 0,04083 \ .$$

O desvio padrão:

$$\sigma = \sqrt{0,04083} \approx 0,2021 \ .$$

A análise da variação nos preços do óleo diesel S-10 nos postos de combustíveis de Vitória revela uma dispersão moderada entre os valores praticados. A média aritmética dos preços foi de R$ 3,34, e o desvio padrão calculado foi de R$ 0,20, o que indica que, embora haja flutuações nos preços entre os estabelecimentos, as diferenças não são excessivas.

Este comportamento pode sugerir que o mercado local de combustíveis está relativamente equilibrado, com uma concorrência moderada entre os postos. As oscilações nos preços podem ser influenciadas por fatores como custos logísticos, políticas fiscais e estratégias de precificação dos postos, mas de forma geral, as variações observadas não apontam para uma grande heterogeneidade no mercado.

Além disso, a presença de um desvio padrão de aproximadamente R$ 0,20 indica que a maioria dos preços está concentrada em torno da média, o que pode ser interpretado como uma margem de variação que não impacta significativamente o consumidor, reforçando a ideia de um mercado competitivo. Tais resultados são úteis para órgãos reguladores, como o Procon-ES, que monitoram a transparência e a justiça nas práticas de precificação no setor de combustíveis.

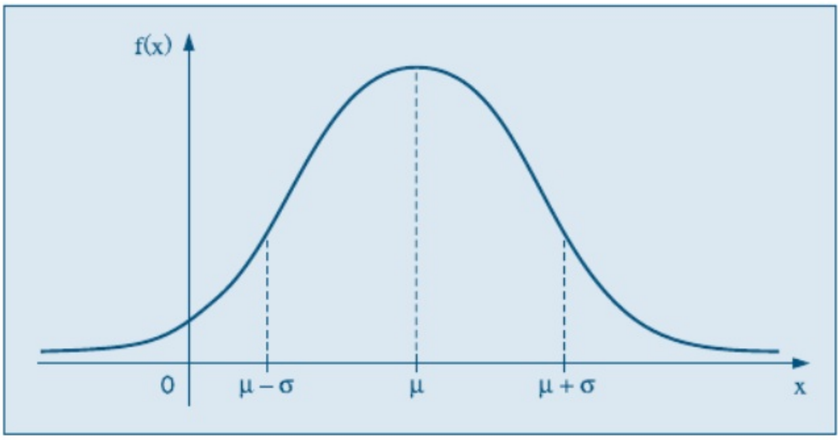
Esse estudo contribui para uma compreensão mais detalhada do mercado local de combustíveis e pode servir como base para futuras análises sobre os preços dos combustíveis em contextos semelhantes.

3.9 Distribuição Normal

A distribuição normal, também conhecida como **curva de Gauss**, é um dos modelos estatísticos mais importantes e amplamente utilizados para representar fenômenos naturais e sociais. Ela descreve como os dados se distribuem ao redor de uma média, sendo mais comuns os valores próximos a essa média e menos frequentes os valores extremos.

Essa distribuição possui uma forma característica de **sino simétrico**, concentrando a maior parte dos dados no centro e apresentando uma dispersão equilibrada para os dois lados. Diversos fenômenos do cotidiano seguem essa tendência, como a altura de pessoas, peso corporal, tempos de reação e notas escolares (TRIOLA, 2017).

Figura 28 – Representação gráfica de uma função densidade de distribuição normal com média μ e desvio padrão σ



Fonte: MORETTIN e BUSSAB (2010), p. 176.

A forma exata da curva é determinada por sua **função densidade de probabilidade**, dada pela equação:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty\,. \tag{3.28}$$

Onde:

- x é o valor observado (variável aleatória contínua);
- μ representa a média da distribuição;
- σ é o desvio padrão, que indica a dispersão dos dados;
- e é a base do logaritmo natural ($e \approx 2{,}718$);
- π é a constante pi ($\pi \approx 3{,}1416$).

Embora a fórmula possa parecer complexa, sua função principal é definir a "altura" da curva para cada valor x. Quanto mais próximo um valor estiver da média, maior será sua densidade de probabilidade — ou seja, sua ocorrência é mais provável.

Neste trabalho, apresentaremos uma abordagem introdutória da distribuição normal, com foco na compreensão conceitual e em aplicações práticas, evitando o uso de cálculos avançados, como integrais. Essa escolha está alinhada aos objetivos do itinerário formativo *Educação Financeira e Fiscal*, voltado ao Ensino Médio.

3.9.1 Propriedades da Distribuição Normal

A distribuição normal apresenta características fundamentais que a tornam uma ferramenta estatística poderosa. Entre as principais, destacam-se:

1. **Simetria**
   A curva normal é simétrica em torno da média, ou seja, os dados estão igualmente distribuídos dos dois lados do centro.

   Exemplo 3.9.1. *Se a média das idades de uma turma for de 15 anos, isso indica que há aproximadamente a mesma quantidade de estudantes com idades inferiores e superiores a esse valor.*

2. **Coincidência das medidas de tendência central**
   Em uma distribuição normal, a média, a mediana e a moda coincidem. Esses valores representam o ponto central da distribuição.

   Exemplo 3.9.2. *Em uma turma de 50 alunos cuja média de notas foi 7, a maioria dos alunos terá desempenho próximo a esse valor, com poucos casos extremos de notas muito altas ou muito baixas.*

3. **Desvio padrão (σ)**
   Essa medida quantifica a dispersão dos dados em torno da média. Distribuições com menor desvio padrão são mais concentradas; com maior desvio, mais espalhadas.

   Exemplo 3.9.3. *No Exemplo 3.9.2, se o desvio padrão das notas for 1, a maioria ficará entre 6 e 8. Se for 3, as notas estarão distribuídas entre 4 e 10.*

3.9.2 Regra Empírica e Escore-z

Um dos conceitos mais úteis associados à distribuição normal é a chamada **Regra Empírica** (ou regra dos 68-95-99,7), que nos ajuda a entender como os dados se distribuem em torno da média.

Essa regra afirma que, em uma distribuição normal:

- Cerca de 68% dos valores estão dentro de ± 1 desvio padrão da média ($\mu \pm \sigma$);
- Cerca de 95% dos valores estão dentro de ± 2 desvios padrão da média ($\mu \pm 2\sigma$);
- Cerca de 99,7% dos valores estão dentro de ± 3 desvios padrão da média ($\mu \pm 3\sigma$).

Exemplo 3.9.4. *Se a média das notas de uma prova é 6 e o desvio padrão é 2, então:*

- 68% *dos alunos tiraram notas entre 4 e 8 (1 desvio padrão da média);*
- 95% *dos alunos tiraram notas entre 2 e 10 (2 desvios padrões da média);*
- $99,7\%$ *dos alunos tiraram notas entre 0 e 12 (3 desvios padrões da média).*

Essa propriedade é especialmente importante para identificar valores *atípicos* ou *outliers*, que estão fora do esperado para uma dada distribuição, auxiliando em análises de controle de qualidade, finanças, saúde, entre outras áreas.

Para facilitar essa análise, utilizamos o conceito do **escore-z**, que transforma qualquer valor x observado em uma medida padronizada da distância, em termos de desvios padrão, em relação à média:

$$z = \frac{x - \mu}{\sigma} . \tag{3.29}$$

O escore-z indica quantos desvios padrão um valor está acima (se $z > 0$) ou abaixo (se $z < 0$) da média. Valores de z próximos a zero são comuns, enquanto valores muito grandes (positivos ou negativos) indicam dados incomuns.

Exemplo 3.9.5. *Considere uma turma de estudantes com média das notas igual a 70 e desvio padrão 10. Suponha que um aluno obteve a nota de 85 pontos. Vamos calcular o* ***escore*** *z para essa nota.*

Substituindo os valores indicados na fórmula de escore-z temos:

$$z = \frac{85 - 70}{10} = \frac{15}{10} = 1{,}5 \ .$$

Isso significa que o aluno está 1,5 desvios padrão acima da média da turma. Portanto, seu desempenho foi significativamente superior à média esperada.

Exemplo 3.9.6. *De acordo com a Organização Mundial da Saúde (OMS), em crianças de 5 anos de idade:*

- *A altura média global para meninos é de aproximadamente 110 cm;*
- *O desvio padrão (σ) é de cerca de 4,8 cm.*

Fonte: SOCIEDADE BRASILEIRA DE PEDIATRIA. Gráficos de Crescimento. Disponível em: <https://www.sbp.com.br/departamentos/endocrinologia/graficos-de-crescimento/>. Acesso em: 1 maio 2025.

Suponha que um menino de 5 anos, tenha sido medido em uma unidade de saúde e apresentou uma altura de 101 cm.

Usamos a fórmula do escore z:

$$z = \frac{101 - 110}{4,8}$$
$$z \approx -1,88 \ .$$

O escore z de $-1,88$ *indica que João está 1,88 desvios padrão abaixo da média, o que pode ser um sinal de risco de baixa estatura para a idade, segundo os critérios da OMS. Valores abaixo de -2 são considerados indicativos de baixa estatura, um problema associado à desnutrição crônica. A equipe de saúde pode usar essa informação para investigar as causas possíveis como alimentação inadequada, problemas de saúde ou condições socioeconômicas e intervir precocemente para melhorar o crescimento do menino.*

Para interpretar os valores do escore-z e estimar as probabilidades associadas a eventos dentro de uma distribuição normal, utilizamos a **Tabela da Distribuição Normal Padrão**, também conhecida como **Tabela** z[11]. Essa tabela fornece a área sob a curva normal padrão à esquerda de um determinado valor de z, representando a probabilidade acumulada até esse ponto.

A utilização da Tabela z é um recurso fundamental na prática estatística, pois permite que qualquer valor observado em uma distribuição normal seja transformado em um escore padronizado. A partir disso, torna-se possível calcular sua posição relativa dentro da distribuição e determinar a probabilidade de ocorrência de valores maiores ou menores que ele.

Exemplo 3.9.7. *Considere que em uma pesquisa sobre as idades dos funcionários de uma empresa, a média de idade seja de 40 anos, com desvio padrão de 5 anos. Queremos saber a probabilidade de um funcionário ter menos de 35 anos. Para isso, vamos calcular o* ***escore*** z *correspondente à idade de 35 anos e consultar a Tabela* z *para determinar essa probabilidade.*

Substituindo os valores fornecidos na fórmula de escore z *temos:*

$$z = \frac{35 - 40}{5} = \frac{-5}{5} = -1 \quad .$$

Com o escore z *igual a -1, consultamos a Tabela* z *para encontrar a área sob a curva normal padrão à esquerda desse valor. De acordo com a Tabela* z*, o valor correspondente a* $z = -1$ *é aproximadamente* 0,1587*, o que indica que a probabilidade de um funcionário ter menos de 35 anos é de cerca de* 15,87%.

A Figura 29 apresenta a Tabela z *utilizada para encontrar essa área, destacando a linha referente ao valor de* $z = -1$.

Exemplo 3.9.8. *Considere que um laboratório realiza exames de colesterol total em uma população adulta. Os dados mostram que o nível médio de colesterol é de 200 mg/dL, com desvio padrão de 25 mg/dL. A equipe de saúde quer saber qual a porcentagem da população apresenta colesterol acima de 242 mg/dL.*

[11] A tabela pode ser acessada em: NABER, C. Tabela normal padrão. Disponível em: <https://www.ime.unicamp.br/~cnaber/tabela_normal.pdf>. Acesso em: 1 maio 2025.

Figura 29 – Encontrando o valor da área correspondente ao valor de $z = -1$

z	0,0	0,01	0,02	0,03	0,04	0,05
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885

Fonte: Tabela z adaptada. Disponível em: <https://www.ime.unicamp.br/~cnaber/tabela_normal.pdf>. Acesso em: 1 maio 2025.

Primeiro, calculamos o escore z correspondente a 242 mg/dL:

$$z = \frac{242 - 200}{25} = \frac{42}{25} = 1,68 \quad .$$

*Com o escore $z = 1,6$, consultamos a **Tabela** z para encontrar a área à esquerda deste valor.*

A Figura 30 apresenta a tabela z com o valor correspondente a $z = 1{,}6$, que indica a área acumulada até esse ponto em uma distribuição normal padrão.

Figura 30 – Encontrando o valor da área correspondente ao valor de $z = 1,6$

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633

Fonte: Tabela Z adaptada. Disponível em: <https://www.ime.unicamp.br/~cnaber/tabela_normal.pdf>. Acesso em: 1 maio 2025.

*Segundo a tabela, a área acumulada até $z = 1{,}6$ é aproximadamente **0,9535**, ou seja, 95,35% da população apresenta níveis de colesterol iguais ou inferiores a 242 mg/dL.*

Portanto, a porcentagem de pessoas com colesterol acima de 242 mg/dL é:

$$100\% - 95,35\% = 4,65\% \ .$$

*A análise mostra que cerca de **4,65%** da população apresenta colesterol em níveis potencialmente preocupantes, o que pode embasar ações preventivas de saúde pública.*

A padronização por meio do escore-z e o uso da Tabela z têm ampla aplicação prática em diversas áreas do conhecimento. Na educação, possibilita a comparação de desempenho entre alunos ou turmas distintas; na saúde, auxilia na avaliação de crescimento e no monitoramento de condições clínicas; na psicologia, é usada para interpretar escores em testes padronizados; e na economia, contribui para a identificação de anomalias em séries temporais.

Além disso, o escore-z é uma ferramenta muito útil para analisar dados e tirar conclusões a partir deles. Ele permite comparar valores de diferentes situações e identificar se estão dentro do esperado ou se são excepcionais. Por ser simples de calcular e interpretar, o escore-z é amplamente utilizado tanto em estudos escolares quanto em pesquisas científicas e aplicações profissionais em diversas áreas, como saúde, economia e educação.

3.9.3 Como Verificar se os Dados de uma pesquisa seguem uma Distribuição Normal?

Existem várias maneiras de verificar se os dados seguem uma distribuição normal. Mas iremos considerar uma da qual já abordamos nesse estudo na Seção 3.4 na representação gráfica utilizando o **histograma**. Se os dados seguem uma distribuição normal, o histograma terá a forma de um **sino**, com a maior parte dos dados concentrada em torno da média. Esse comportamento é representado pelo escore z, que ajuda a identificar como os dados se distribuem em torno da média e qual é a probabilidade de ocorrência de determinados valores dentro dessa distribuição.

Exemplo 3.9.9. *Se você coletar as idades de 100 pessoas e construir um histograma, e os dados seguirem uma distribuição normal, o gráfico terá a forma de um sino, com as idades mais comuns concentradas ao redor de uma faixa central, como 30 anos. Esse comportamento é refletido por escores z próximos de 0, indicando que a maior parte dos dados está próxima da média.*

Exemplo 3.9.10. *Considere que sejam lançadas 30 moedas ao mesmo tempo e será registrado o número de moedas que saem com a face **cara** voltada para cima. Este procedimento será repetido 50 vezes, resultando em 50 valores a serem observados. Cada valor representa o número de caras obtidas em um lançamento coletivo de 30 moedas.*

A chance de uma moeda cair com a face **cara** *é de* 50% *(isto é,* $p = 0{,}5$ *). Logo, a variável aleatória* ***número de caras*** *segue uma* ***distribuição binomial*** *com* $n = 30$ *tentativas e probabilidade* $p = 0{,}5$.

Com base na teoria, espera-se que a distribuição dos dados se aproxime de uma ***distribuição normal*** *com:*

- *Média esperada:*

  $\mu = n \cdot p = 30 \cdot 0{,}5 = 15$.

- *Desvio padrão esperado:*

  $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{30 \cdot 0{,}5 \cdot 0{,}5} \approx 2{,}74$.

Para realizar o experimento de forma prática e automatizada, utilizaremos o ***Microsoft Excel****. A seguir, descrevemos os passos para realizar a simulação:*

1. *Gerar os Dados*

   Em uma planilha do Excel:

   - *Crie 30 colunas representando cada moeda.*
   - *Em cada célula use a fórmula:* `=SE(ALEATÓRIO()<0,5;1;0)` *para simular uma moeda (1 = cara, 0 = coroa).*

     A Figura 31 ilustra um exemplo da geração de dados aleatórios no Excel utilizando a fórmula mencionada.

Figura 31 – Gerando os Dados aleatórios

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	[illegible]
2	0	0	0	1	0	0	1	0	1	1	1	1	0	1	0	[illegible]
3	1	1	1	1	0	0	0	1	1	1	0	0	1	0	1	[illegible]
4	0	0	0	1	1	1	0	1	1	0	1	1	1	1	1	[illegible]
5	0	0	0	1	1	1	1	0	0	1	1	0	0	0	1	[illegible]
6	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1	[illegible]
7	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	[illegible]
8	1	0	0	0	0	0	0	1	0	1	1	0	0	1	1	[illegible]
9	0	0	1	1	1	0	0	1	0	0	0	1	0	1	0	[illegible]
10	1	1	0	0	0	1	0	1	1	1	0	1	0	1	1	[illegible]
11	0	0	1	1	1	0	0	0	1	0	1	0	0	1	0	[illegible]
12	1	0	1	0	1	1	0	0	0	1	0	1	0	0	0	[illegible]
13	0	1	0	0	1	1	1	0	0	0	0	1	0	1	0	[illegible]
14	0	0	1	1	0	0	1	1	1	1	0	1	0	1	0	[illegible]
15	1	0	0	0	1	1	0	0	0	0	0	1	1	0	0	[illegible]
16	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1	[illegible]
17	1	1	0	0	1	1	1	0	1	0	0	0	1	0	0	[illegible]
18	1	1	1	0	1	1	0	1	1	0	1	0	0	1	1	[illegible]
19	1	0	0	1	1	1	0	0	1	0	0	1	0	0	1	[illegible]
20	0	0	0	1	0	0	0	1	1	1	1	0	0	1	1	[illegible]
21	0	1	1	0	0	0	0	0	1	0	0	0	0	1	1	[illegible]
22	0	0	1	0	1	1	0	1	0	0	1	1	0	0	0	[illegible]

Fonte: Produção da própria autora (2025)

   - *Na 31ª coluna, some os valores de cada linha:* `=SOMA(A1:AD1)`.
   - *Copie essa linha para baixo até completar 50 linhas (simulando 50 lançamentos).*

2. *Cálculo da Média e Desvio Padrão*

   Utilize as funções do Excel:

   - *Média:* `=MÉDIA(AE1:AE50)`
   - *Desvio padrão:* `=DESVPAD.P(AE1:AE50)`

   A Figura 32 apresenta a aplicação das fórmulas no Excel para o cálculo da média e do desvio padrão da amostra de dados simulados.

Figura 32 – Cálculo da Média e Desvio Padrão

Média	14,74
Desvio Padrão	2,133635395

Fonte: Produção da própria autora (2025)

3. ***Construir o Histograma***

   - *Crie intervalos de classes (exemplo: 10|- 11, 12|- 13, 14|- 15, etc.).*
   - *Conte quantos valores se encontram em cada intervalo e insira como gráfico de colunas.*

   A Figura 33 mostra a tabela com os intervalos de classes utilizados para a construção do histograma. Em seguida, a Figura 34 apresenta o gráfico correspondente, representando a distribuição de frequências dos dados simulados.

Figura 33 – Tabela com intervalos de classes

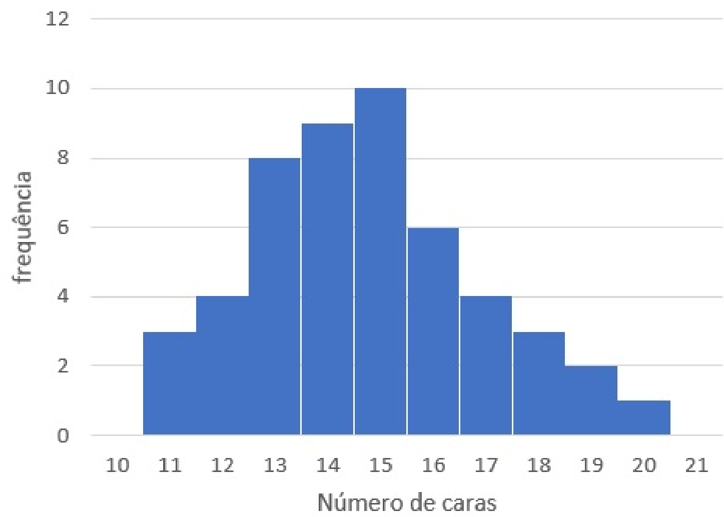
Número de Caras	Frequência
10 |-11	0
11 |-12	3
12 |-13	4
13 |-14	8
14 |-15	9
15 |-16	10
16 |-17	6
17 |-18	4
18 |-19	3
19 |-20	2
20 |-21	1
21 |-12	0

Fonte: Produção da própria autora (2025)

4. *Adicionar a Curva Normal*

   Para sobrepor a curva normal ao histograma:

Figura 34 – Gráfico



Fonte: Produção da própria autora (2025)

- *Calcule os valores da função densidade normal dada pela Equação 3.28; Em Excel, utilize:* `=DIST.NORM(x; média; desvio_padrão; FALSO)` *para cada valor* x *(ponto médio de cada classe). A Tabela com os valores obtidos da função densidade está representada na Figura 35.*

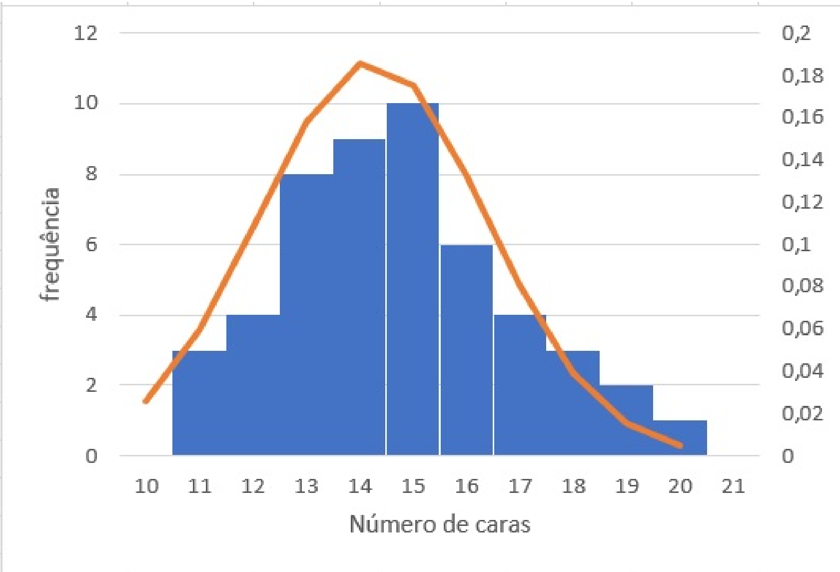
Figura 35 – Tabela com os valores da função densidade normal

Número de Caras	Frequência	Função densidade normal
10 |-11	0	0,025957758
11 |-12	3	0,059028195
12 |-13	4	0,107758815
13 |-14	8	0,157923622
14 |-15	9	0,18579857
15 |-16	10	0,175484443
16 |-17	6	0,133056429
17 |-18	4	0,080990501
18 |-19	3	0,039576127
19 |-20	2	0,015525067
20 |-21	1	0,004889165

Fonte: Produção da própria autora (2025)

- *Em seguida, esses valores devem ser inseridos no gráfico como uma nova série de dados, representada por uma linha sobreposta ao histograma. Esse procedimento permite visualizar como os dados observados se aproximam da distribuição* *normal teórica. O resultado final dessa sobreposição pode ser visto na Figura 36.*

Figura 36 – Tabela com os valores da função densidade normal



Fonte: Produção da própria autora (2025)

Após a realização do experimento, é possível observar que a maioria dos resultados ficou concentrada em torno de 15 caras, valor que corresponde à média esperada para o lançamento de 30 moedas. Com base nos dados simulados e no histograma construído, percebe-se que a distribuição do número de caras tende a assumir uma forma simétrica e centralizada, muito próxima de uma curva em sino.

A sobreposição da curva normal ao histograma evidencia que fenômenos aleatórios simples, como o lançamento de moedas, podem apresentar comportamento semelhante ao de uma distribuição normal. Essa observação ilustra, de maneira concreta, o conceito estatístico conhecido como o **Teorema Central do Limite (TCL)**.

Embora esse teorema seja tradicionalmente estudado no ensino superior, ele pode ser compreendido de forma intuitiva: quando repetimos muitas vezes experimentos aleatórios independentes — como lançar moedas ou calcular médias de diferentes amostras —, os resultados obtidos (sejam somas ou médias) tendem a se distribuir de forma aproximadamente normal, mesmo que as variáveis individuais não sigam uma distribuição normal inicialmente. Em outras palavras, quanto maior o número de repetições, mais os resultados se aproximam da forma de sino típica da distribuição normal.

De modo mais técnico, segundo James (2007), o Teorema Central do Limite afirma que, sob certas condições, a seguinte padronização das somas parciais de variáveis aleatórias independentes:

$$Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} \tag{3.30}$$

converge, em distribuição, para uma variável aleatória com distribuição normal padrão.

Nessa equação, $S_n = X_1 + X_2 + \cdots + X_n$ representa a soma de variáveis aleatórias independentes com variância finita, $\mathbb{E}[S_n]$ é a média esperada dessa soma e $\mathrm{Var}(S_n)$ sua variância.

De forma mais acessível, isso significa que, ao somarmos muitas variáveis aleatórias independentes (como resultados de experimentos repetidos), e aplicarmos esse processo de ajuste — subtraindo a média esperada e dividindo pelo desvio padrão —, os valores resultantes (Z_n) passam a seguir uma distribuição muito próxima da curva normal padrão, isto é, com média zero e desvio padrão igual a 1.

Esse processo é bastante semelhante ao cálculo do escore-z, apresentado na Equação 3.29, que também envolve a subtração da média e a divisão pelo desvio padrão. Essa semelhança reforça a relação entre os conceitos de **normalização** (como no Teorema Central do Limite) e **padronização** (como no escore-z), ambos essenciais para a análise estatística de dados.

A distribuição normal é amplamente utilizada porque muitos fenômenos naturais e sociais seguem esse padrão. Algumas das aplicações incluem:

- Análise de *resultados de provas* e exames, onde a maioria dos estudantes obtém notas próximas à média.
- Estudos em *ciências naturais*, como a distribuição das velocidades de partículas ou a altura de plantas.
- Previsões e análise de *investimentos financeiros*, onde a variação dos preços de ações é frequentemente modelada usando a distribuição normal.

A seguir, apresentamos dois exemplos que ilustram a aplicabilidade da distribuição normal em diferentes cenários práticos:

Exemplo 3.9.11. *Suponha que uma fábrica de bebidas realize o controle de qualidade no peso das garrafas produzidas. Sabe-se que o peso das garrafas vazias segue uma **distribuição normal** com média* 500 *gramas e desvio padrão* 10 *gramas.*

Para garantir que todas as garrafas estejam dentro do padrão aceitável, a empresa estabelece que garrafas com peso inferior a 480g ou superior a 520g devem ser descartadas.

Vamos determinar qual a porcentagem de garrafas que estão dentro do intervalo aceitável?

Primeiramente transformaremos os limites de peso das garrafal em ***escores*** *z, utilizando a fórmula:*

$$z = \frac{x - \bar{x}}{\sigma}$$

$$z_1 = \frac{480 - 500}{10} = -2 \qquad z_2 = \frac{520 - 500}{10} = 2 \ .$$

Usando a ***Regra Empírica (68-95-99,7)****, sabemos que aproximadamente* ***95%*** *dos dados em uma distribuição normal estão entre dois desvios padrão da média, ou seja, entre $z = -2$ e $z = 2$.*

Podemos concluir então que cerca de ***95%*** *das garrafas produzidas estão dentro do intervalo de peso aceitável, e* ***5%*** *serão descartadas por estarem fora do padrão (2,5% abaixo de 480g e 2,5% acima de 520g).*

Esse exemplo mostra como a distribuição normal e o escore z são ferramentas eficazes para prever a variabilidade e garantir a qualidade em processos industriais.

Exemplo 3.9.12. *Considere uma escola onde os resultados de uma prova de matemática aplicada a 200 alunos seguem uma distribuição normal com média* 70 *pontos e desvio padrão* 10 *pontos. A coordenação pedagógica deseja premiar os alunos que estão entre os melhores 2,5% da turma. Para isso, é necessário determinar qual deve ser a nota mínima x para receber essa premiação.*

Sabemos que os $2,5\%$ superiores de uma distribuição normal correspondem a escores z ***maiores que aproximadamente 1,96****, conforme indicado pela tabela da distribuição normal padrão (ver Figura 37). Isso significa que o valor de $z = 1,96$ deixa cerca de $97,5\%$* ***da área acumulada à sua esquerda*** *(pois $100\% - 2,5\% = 97,5\%$).*

A partir disso, podemos substituir os valores na fórmula do escore z para encontrar a nota correspondente:

$$1,96 = \frac{x - 70}{10}$$

$$x = 1{,}96 \cdot 10 + 70 = 89{,}6 \ .$$

Portanto para estar entre os 2,5% melhores da turma, um aluno deve tirar pelo menos ***89,6 pontos*** *na prova. A distribuição normal, nesse caso, permite determinar limites de desempenho de forma estatisticamente fundamentada.*

Para um aprofundamento conceitual e prático sobre a distribuição normal, recomenda-se a leitura de Cabariti (2025), que apresenta de forma acessível uma análise detalhada da

Figura 37 – Encontrando o valor de z correspondente ao valor da área de $2,5\%$ superiores de uma distribuição normal (0,975)

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: Tabela z adaptada. Disponível em: <https://www.ime.unicamp.br/~cnaber/tabela_normal.pdf>. Acesso em: 1 maio 2025.

função densidade da distribuição normal, bem como a interpretação de seus principais parâmetros — média e desvio padrão. O material também destaca as propriedades fundamentais dessa distribuição, como a simetria em torno da média e o fato de que a área sob a curva é igual a 1, ilustrando visualmente a determinação de regiões de probabilidade. Além disso, são propostas atividades e exercícios que contribuem para a consolidação do entendimento da distribuição normal como um modelo probabilístico amplamente utilizado na estatística e em diversas áreas das ciências.

3.10 Intervalo de Confiança

O intervalo de confiança é uma ferramenta estatística que permite estimar, com determinado grau de certeza, o valor de um parâmetro populacional com base em dados amostrais. Segundo Triola (2017), o intervalo de confiança consiste em uma faixa de valores calculada a partir dos dados amostrais que, com determinado nível de confiança (por exemplo, 95%), tem grande probabilidade de conter o valor verdadeiro do parâmetro populacional, como a média ou a proporção.

Matematicamente, o intervalo é construído com base na média amostral ($\bar{x}$), no desvio padrão da amostra (σ), no tamanho da amostra (n) e em um fator de correção (z), derivado da distribuição normal padrão. A fórmula básica para o intervalo de confiança da média é dada por:

$$IC = \bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}. \tag{3.31}$$

Onde:

- $\bar{x}$ é a média amostral;
- σ é o desvio padrão populacional;
- n é o tamanho da amostra.

Esse intervalo oferece uma maneira prática de expressar a precisão de uma estimativa amostral. Por exemplo, um intervalo de confiança de 95% nos diz que, se repetirmos o processo de amostragem inúmeras vezes, em 95% das amostras o valor estimado do parâmetro estará dentro do intervalo calculado.

O intervalo de confiança tem uma aplicação fundamental em várias áreas do conhecimento e das ciências aplicadas. Em contextos industriais, como o exemplo a seguir, ele é utilizado para controlar a qualidade de processos e garantir que as medidas estejam dentro dos limites aceitáveis. Além disso, é frequentemente usado em pesquisas de mercado, estudos de saúde pública, e na avaliação de investimentos financeiros, pois fornece uma maneira robusta de estimar variáveis desconhecidas.

Exemplo 3.10.1. *Considere que uma fábrica declara que os parafusos que produz possuem, em média,* $100g$*. Para verificar se o processo produtivo continua atendendo a esse padrão, foi coletada uma amostra aleatória de* 64 *parafusos, e a média amostral obtida foi de* $98g$*. Sabemos, a partir de dados históricos, que o desvio padrão populacional do peso dos parafusos é de* $8g$*.*

Neste caso, deseja-se calcular o ***intervalo de confiança de*** 95% *para a média populacional do peso dos parafusos. Esse intervalo de confiança proporcionará uma estimativa, com um grau conhecido de confiabilidade, sobre o valor verdadeiro da média da população, permitindo verificar se existem indícios de que o processo produtivo tenha se desviado do valor de referência de* $100g$*.*

Para calcular esse intervalo, utilizamos a seguinte fórmula:

$$IC = \bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

Os valores de $\bar{x}$*,* σ *e* n *já estão claramente especificados no problema. O próximo passo é determinar o valor do escore-z, que pode ser obtido consultando a tabela da distribuição normal padrão. Para um nível de confiança de* 95%*, buscamos dois valores de* z *(um negativo e um positivo) que satisfaçam as condições a seguir:*

- *A área sob a curva normal entre esses dois valores seja de* 95%*;*
- *Os* 5% *restantes da área sejam distribuídos igualmente nas duas caudas da curva, ou seja,* $2,5\%$ *à esquerda e* $2,5\%$ *à direita.*

Esse valor de z corresponde à área acumulada à esquerda do valor de z. Queremos encontrar o valor de z tal que a área à esquerda dele seja $0,975$, pois:

$$100\% - 2,5\% = 97,5\% \quad .$$

Observando a Figura 38 obtemos $z = 1,96$ (pois $P(z < 1,96) \approx 0,975$).

Isso significa que 95% *da área sob a curva normal está entre $z = -1,96$ e $z = +1,96$.*

Figura 38 – Encontrando o valor de z correspondente ao valor da área de $2,5\%$ superiores em uma distribuição normal (0,975)

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: Tabela Z adaptada. Disponível em: <https://www.ime.unicamp.br/~cnaber/tabela_normal.pdf>. Acesso em: 1 maio 2025.

Agora, substituindo os valores na fórmula do intervalo de confiança (Equação 3.31), temos:

$$IC = 98 \pm 1,96 \cdot \frac{8}{\sqrt{64}} = 98 \pm 1,96$$
$$IC = [96{,}04 \ ; \ 99{,}96] \quad .$$

Assim, com 95% *de confiança, podemos afirmar que a média real do peso dos parafusos produzidos está entre **96,04g** e **99,96g**. Como esse intervalo **não inclui** o* *valor de referência de **100g**, há evidências de que o processo produtivo pode ter sofrido alterações.*

Ou seja, se a média populacional ainda fosse de fato $100g$*, seria esperado que ela estivesse dentro do intervalo estimado com base na amostra. O fato de estar fora desse intervalo sugere, com um grau razoável de confiança, que a média real é inferior a* $100g$*. Isso não é uma prova absoluta, mas é um indício estatisticamente significativo de que a fábrica deve investigar o processo produtivo para verificar se houve algum desvio, falha ou mudança nas condições de produção.*

A interpretação correta do intervalo de confiança é fundamental para sua aplicação. Quando calculamos um intervalo de confiança de 95%, não estamos dizendo que há 95% de chance de que o valor verdadeiro da média esteja dentro do intervalo. Em vez disso, estamos afirmando que, se repetíssemos o processo de amostragem muitas vezes, 95% dos intervalos construídos conteriam o valor verdadeiro da média populacional.

Além disso, o intervalo de confiança não é uma garantia absoluta, mas sim uma estimativa com base nos dados da amostra. Essa estimativa está sujeita a erros amostrais e deve ser utilizada com cautela, especialmente quando as suposições para o uso da fórmula (como a normalidade dos dados e o conhecimento do desvio padrão populacional) não são atendidas.

3.11 Teste de Hipóteses

O teste de hipóteses é um procedimento estatístico utilizado para avaliar se os dados amostrais fornecem evidências suficientes para rejeitar uma hipótese previamente formulada. Essa técnica permite estimar a probabilidade de que a diferença observada nos dados tenha ocorrido apenas por acaso, considerando que a hipótese nula (H_0) seja verdadeira (ASSIS; SOUSA; LINHARES, 2020).

A formulação de hipóteses é uma etapa essencial do processo científico. As hipóteses devem ser baseadas em conhecimento prévio, estudos anteriores e observações iniciais, e precisam ser claras, objetivas e passíveis de verificação por meio de dados coletados de forma sistemática. Para garantir a imparcialidade, a definição da hipótese a ser testada deve ocorrer antes da análise dos dados.

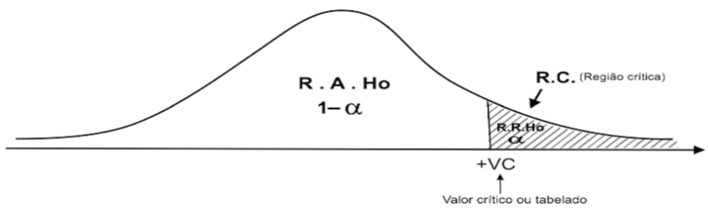
Em um teste de hipóteses, são formuladas duas proposições opostas:

- **Hipótese nula** (H_0): representa o estado de neutralidade, ausência de efeito ou diferença.
- **Hipótese alternativa** (H_1): representa uma possível mudança, efeito ou diferença a ser investigada.

A decisão de rejeitar ou não H_0 é feita com base em uma estatística calculada a partir da amostra, chamada **estatística do teste**. Essa estatística é comparada a valores críticos definidos por um **nível de significância** (α), que representa a probabilidade máxima de rejeitar H_0 quando ela for verdadeira. Comumente, utiliza-se $\alpha = 0{,}05$, o que indica uma margem de erro de 5%.

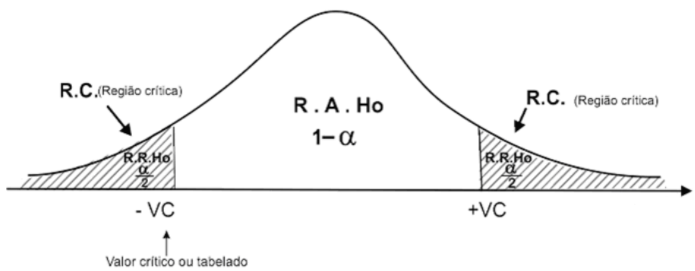
No **teste bilateral**, α é dividido igualmente entre as duas extremidades (caudas) da distribuição, sendo adequado quando se deseja detectar desvios tanto para mais quanto para menos (ver Figura 40). No **teste unilateral**, todo o valor de α é alocado em uma única cauda — à direita (se espera aumento) ou à esquerda (se espera diminuição), como ilustrado na Figura 39.

Figura 39 – Curva do teste de hipótese unilateral à direita, mostrando a região crítica ou de rejeição de H_0



Fonte: Assis, Sousa e Linhares (2020), p. 42

Figura 40 – Curva do teste de hipótese bilateral, mostrando as regiões críticas ou de rejeição de H_0



Fonte: Assis, Sousa e Linhares (2020), p. 44

A estatística de teste utilizada para esse estudo será para dados com distribuição normal conhecida, o **teste** z, dado por:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\,. \tag{3.32}$$

onde:

- $\bar{x}$ é a média amostral;
- μ_0 é o valor da média sob a hipótese nula (H_0);

- σ é o desvio padrão populacional;
- n é o tamanho da amostra.

Esse valor de Z é então comparado com os valores críticos da distribuição normal padrão para tomar a decisão estatística.

Exemplo 3.11.1. *Suponha que se deseje testar se a média de idade de uma população é igual a 50 anos. Para isso, foi coletada uma amostra aleatória de $n = 36$ elementos, com média amostral $\bar{x} = 52$ e desvio padrão populacional $\sigma = 6$. Adote um nível de significância de $\alpha = 0{,}05$.*

Etapas do teste:

- *Hipóteses:*
  - H_0: $\mu = 50$
  - H_1: $\mu \neq 50$
- *Estatística do teste (Equação 3.32):*

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{52 - 50}{6/\sqrt{36}} = \frac{2}{1} = 2$$

- *Valor crítico (teste bilateral com $\alpha = 0{,}05$): $Z_{crítico} = \pm 1{,}96$*

  (obtido na Tabela Z disponível em: <https://www.ime.unicamp.br/~cnaber/tabela_normal.pdf>)

- ***Decisão**: Como $|z| = 2 > 1,96$, rejeita-se a hipótese nula H_0, pois o valor da estatística do teste encontra-se na região crítica, localizada nas caudas da distribuição normal no teste bilateral.*

Desse modo, existe evidência estatística, ao nível de 5% de significância, de que a média populacional é diferente de 50 anos.

4 Sugestões de Práticas

Este capítulo apresenta um conjunto de propostas de práticas pedagógicas voltadas ao ensino da Estatística, com o intuito de favorecer uma aprendizagem significativa e contextualizada. As atividades sugeridas contemplam conteúdos fundamentais, como medidas de tendência central, medidas de dispersão e noções introdutórias de probabilidade, articulando-os a situações-problema envolvendo contextos reais. O objetivo principal é integrar o conhecimento estatístico a dimensões sociais, ambientais, econômicas e culturais, em consonância com as competências gerais estabelecidas pela Base Nacional Comum Curricular (BNCC), tais como o desenvolvimento do pensamento científico, a capacidade de argumentação com base em evidências, a empatia e o exercício da responsabilidade social.

Os temas abordados nas práticas pedagógicas contemplam questões contemporâneas, como o consumo consciente de recursos hídricos, a análise de variações no desempenho escolar e a tomada de decisões sob incerteza, promovendo, assim, o engajamento dos estudantes em atividades que estimulam a curiosidade, o raciocínio lógico, e a criticidade. Ademais, as propostas visam fornecer ao professor instrumentos didáticos que o auxiliem na organização de aulas mais dinâmicas, contextualizadas e alinhadas aos desafios atuais do ensino e da formação cidadã.

4.1 Prática 01 – Introdução à Estatística por Meio da Investigação de Expectativas

A proposta desta prática é iniciar a disciplina de Estatística por meio de uma abordagem investigativa, utilizando um formulário como instrumento de coleta de dados para compreender as expectativas e percepções dos alunos em relação à matemática/estatística e sua aplicação no cotidiano.

A proposta visa, principalmente, despertar o interesse dos discentes pela Estatística, evidenciando seu potencial para a análise de informações reais e significativas para o grupo. Além disso, essa atividade inicial permite ao docente conhecer melhor o perfil de sua turma, favorecendo a elaboração de aulas mais contextualizadas e alinhadas às necessidades e interesses dos estudantes.

Objetivos

- Investigar a percepção dos alunos sobre a matemática e sua relação com a tomada de decisões;
- Apresentar a estatística como uma ferramenta útil para organizar e interpretar dados;
- Estimular a participação ativa dos alunos desde o início da disciplina.

Descrição da Atividade

A atividade tem início com o preenchimento, por parte dos estudantes, de um formulário composto por questões fechadas e abertas, abordando aspectos como perfil pessoal, sentimentos em relação à matemática/estatística e conteúdos de maior interesse. O instrumento utilizado para essa coleta de dados está disponível no Apêndice B, como sugestão utilizar utilizando a ferramenta *Google Forms*, a qual permite a coleta e organização de dados de forma prática e acessível. Essa plataforma é especialmente útil para aplicações em ambientes educacionais, possibilitando a análise imediata das respostas fornecidas pelos estudantes.

Os dados obtidos podem ser organizados em tabelas e representações gráficas simples, proporcionando uma introdução aos conceitos básicos da Estatística Descritiva. Além disso, o docente poderá propor reflexões acerca de situações cotidianas nas quais a análise de dados se mostra relevante para a tomada de decisões, ampliando o repertório dos estudantes quanto às aplicações práticas da Estatística.

Após a aplicação do formulário, o professor pode seguir algumas etapas para garantir que a atividade seja efetiva e que os dados coletados contribuam para o aprendizado dos alunos e para o aprimoramento das aulas. Aqui estão algumas sugestões de ações que o professor pode realizar:

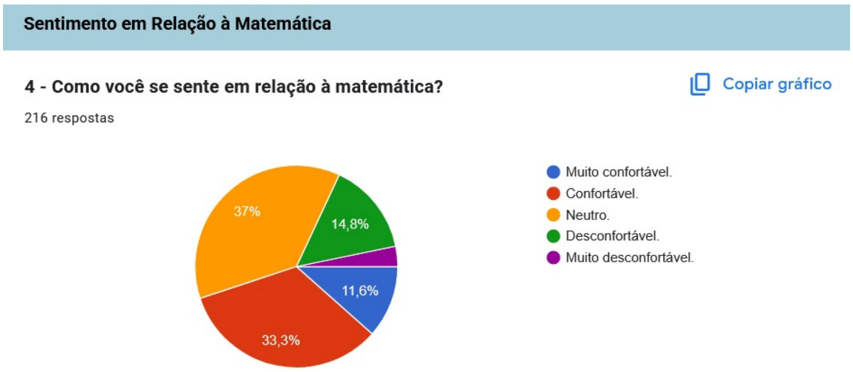
1. **Análise dos Dados:** O professor pode utilizar as ferramentas do Google Forms ou exportar os dados para planilhas, gerando gráficos e tabelas para identificar padrões e tendências.

   Por exemplo, os dados coletados da pergunta nº 4 são apresentados na Figura 41, os da pergunta nº 6 na Figura 42, e os da pergunta nº 12 na Figura 43.

2. **Discussão em Sala de Aula:** Utilizar os resultados do formulário para promover uma discussão sobre as percepções dos alunos em relação à matemática e sua aplicação no cotidiano, estimulando a reflexão sobre a importância da estatística nas decisões, mais informadas.

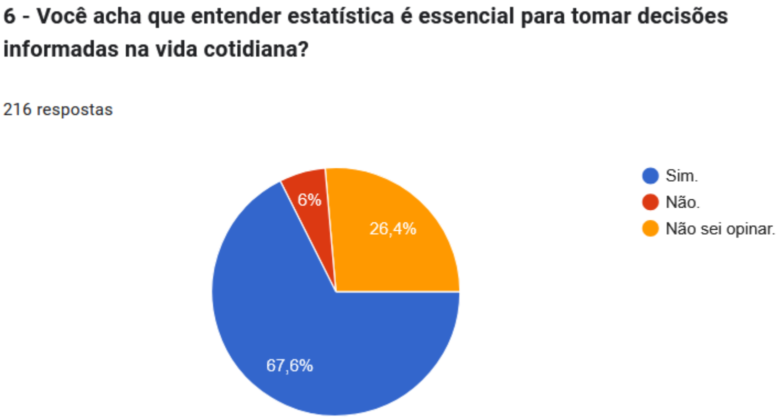
   Por exemplo, nos gráficos de setores apresentados acima relativos aos dados da pergunta 6 e 12, observa-se que mais da metade dos respondentes acredita que é

Figura 41 – Dados coletados da pergunta nº 4



Fonte: Produção da própria autora (2024)

Figura 42 – Dados coletados da pergunta nº 6



Fonte: Produção da própria autora (2024)

essencial tomar decisões informadas na vida cotidiana. No entanto, verifica-se que o número de estudantes que prefere tomar decisões baseadas pela intuição é superior. Isso sugere que, embora os alunos reconheçam a importância de tomar decisões informadas, muitos ainda não aplicam esse princípio na prática, o que pode abrir

Figura 43 – Dados coletados da pergunta nº 12

Fonte: Produção da própria autora (2024) espaço para reflexões e discussões sobre como incorporar dados em suas escolhas diárias.

3. **Introdução aos conceitos preliminares de Estatística:** O professor pode iniciar a introdução de conceitos fundamentais de Estatística, como dados, população, amostra e variáveis.

4. **Atividades Práticas:** Os alunos podem ser solicitados a organizar e apresentar os dados coletados, utilizando tabelas de frequência ou outros métodos de visualização de dados.

5. **Planejamento de Aulas Futuras:** Planejar novas aulas considerando os interesses e desafios revelados pelo formulário, buscando tornar o conteúdo mais relevante e acessível.

4.2 Prática 02 – Análise Estatística de Preços e Consumo Consciente (Interdisciplinar)

"Quanto custa se alimentar? Uma investigação sobre variação de preços e consumo responsável"

Descrição Geral

Esta prática propõe uma atividade interdisciplinar que integra os componentes curriculares de Estatística, Matemática Financeira, Consumo Responsável e Educação Tributária, tendo como foco a observação e análise sistemática da variação de preços dos principais itens da cesta básica ao longo de um período de dois meses, promovendo uma compreensão mais ampla e crítica da dinâmica de preços no mercado local, dos hábitos de consumo da população e da importância da educação financeira e tributária na vida cotidiana.

O desenvolvimento dessa proposta está alinhado a uma abordagem de pesquisa do tipo **longitudinal**, uma vez que envolve o acompanhamento sistemático e repetido de um mesmo grupo de estudantes ao longo de um período definido — neste caso, oito semanas — com coletas periódicas de dados sobre a variação de preços. Uma pesquisa longitudinal caracteriza-se pela observação do mesmo grupo de participantes por certo período de tempo, com coletas repetidas de informações baseadas nos mesmos métodos e variáveis (MORLING, 2015). Esse tipo de abordagem permite não apenas descrever situações pontuais, mas analisar **tendências**, **padrões** e **transformações** ao longo do tempo, conferindo maior profundidade à análise.

Como apontam Magnusson e Cairns (2009) e UNICEF (2015), os estudos longitudinais ocupam um lugar privilegiado quando se busca compreender processos de mudança e desenvolvimento, sendo cada vez mais utilizados também em pesquisas educacionais. No contexto escolar, práticas dessa natureza permitem que os estudantes não apenas apliquem conceitos estatísticos e financeiros, mas também desenvolvam um olhar crítico sobre os fenômenos que se transformam com o tempo — como os preços de alimentos, os hábitos de consumo e os impactos sociais da economia.

A atividade pode ser realizada em grupos, e os estudantes deverão acompanhar, semanalmente, os preços de determinados produtos em supermercados previamente definidos, assegurando a padronização quanto à marca e à quantidade dos itens pesquisados.

Ao final do período estipulado, os dados obtidos serão organizados e analisados com o auxílio de ferramentas estatísticas, permitindo a construção de gráficos, tabelas e indicadores que possibilitem comparações, identificação de tendências de preços, cálculo de médias, variações percentuais e avaliação dos períodos mais vantajosos para a realização de compras. Essa abordagem contribuirá para o desenvolvimento de práticas de consumo mais conscientes e sustentáveis, além de favorecer a compreensão da relação entre finanças pessoais, economia e cidadania.

Após a análise dos dados, cada grupo será convidado a apresentar um seminário à turma, relatando suas experiências durante o processo de coleta, organização e interpretação das informações, bem como apresentando os principais resultados obtidos. Em seguida, o professor poderá conduzir uma discussão coletiva para que os alunos identifiquem possíveis padrões, discrepâncias ou semelhanças entre os diferentes estabelecimentos pesquisados, aprofundando a reflexão sobre os fatores que influenciam as variações de preços e os impactos dessas diferenças no consumo das famílias.

Objetivos

- Desenvolver habilidades de coleta, organização e análise de dados estatísticos;
- Compreender conceitos de variação, média, mediana e desvio padrão em contextos reais;
- Aplicar conhecimentos de matemática financeira para interpretar aumentos percentuais e identificar tendências de mercado;
- Estimular práticas de consumo consciente e sustentável;
- Refletir sobre o papel da educação financeira e tributária na formação cidadã;
- Promover o trabalho colaborativo e a pesquisa de campo como metodologias ativas de aprendizagem.

Etapas para aplicação da prática

Etapa 1 – Formação dos Grupos e Planejamento da Coleta de Dados

- A turma será dividida em grupos (preferencialmente de 4 a 5 alunos).
- Cada grupo ficará responsável por monitorar os preços em um supermercado específico, escolhido previamente para evitar duplicidade.
- Os estudantes deverão registrar, semanalmente, os preços de pelo menos 10 itens da cesta básica (por exemplo: arroz, feijão, leite, óleo, açúcar, café, farinha, macarrão, ovos, pão).
- Para garantir a comparabilidade dos dados, os alunos deverão registrar sempre a mesma marca e mesma quantidade para cada item.

Etapa 2 – Coleta de Dados (Duração: 8 semanas)

- Os preços serão anotados semanalmente em uma planilha (digital ou impressa), contendo as seguintes colunas: Produto, Marca, Quantidade, Preço (R$), Data, Nome do Supermercado.
- Sugere-se o uso de ferramentas como Google Planilhas para facilitar o armazenamento e o compartilhamento das informações.

Etapa 3 – Organização e Análise Estatística dos Dados

Ao final das oito semanas, os grupos deverão:

- Calcular a média de preço semanal de cada produto;
- Determinar a variação de preços (absoluta e percentual);
- Identificar o produto com maior e menor variação de preço;
- Apresentar gráficos (colunas, linhas, pizza) comparando os dados ao longo do tempo e entre supermercados;
- Verificar se há padrões de aumento ou redução de preços em determinados períodos (ex: fim de mês, promoções, sazonalidade).

Etapa 4 – Análise Financeira e Reflexão Crítica

- Utilizar conceitos de matemática financeira para calcular a média de aumento percentual dos produtos por supermercado.

- Simular uma compra mensal e comparar os valores entre os diferentes estabelecimentos.
- Discutir estratégias de economia no consumo alimentar: melhor período para compras, comparação de preços, marcas alternativas, promoções, planejamento financeiro.
- Introduzir no debate questões como carga tributária sobre produtos alimentícios, desigualdade no acesso à alimentação de qualidade e consumo sustentável.

Etapa 5 – Produção e Apresentação dos Resultados

- Os dados e análises serão consolidados em um relatório final, contendo: introdução, metodologia, análise estatística, discussão financeira, reflexões sobre o consumo consciente e considerações finais.
- Os grupos também farão uma apresentação oral ou em formato de pôster para socializar os resultados com a turma e a comunidade escolar.

Considerações Finais

A prática proposta possibilita uma avaliação ampla e formativa, considerando aspectos como a participação ativa dos estudantes em todas as etapas do processo, a organização e precisão na coleta e análise dos dados, a aplicação correta dos conceitos estatísticos e financeiros, além da clareza e consistência na apresentação das informações e reflexões durante os seminários. Ao longo da atividade, os alunos desenvolvem competências essenciais, como o raciocínio lógico-matemático, a leitura crítica da realidade socioeconômica, o planejamento e a tomada de decisões, bem como habilidades de comunicação oral e escrita. Além disso, são estimulados a adotar uma postura cidadã, consciente e ética diante do consumo e da gestão de recursos. A relevância da prática reside justamente em sua capacidade de articular teoria e prática, aproximando os conteúdos escolares do cotidiano dos estudantes e promovendo uma aprendizagem significativa, crítica e voltada para a formação integral.

4.3 Prática 03 – De Olho no Orçamento: A Média como Ferramenta

Descrição Geral

Nesta prática, os estudantes serão convidados a manter os olhos atentos no próprio orçamento. A partir de uma situação inspirada na rotina de um jovem aprendiz, vamos explorar como a média aritmética pode se tornar uma ferramenta poderosa para planejar gastos, evitar surpresas financeiras e desenvolver o hábito do consumo consciente. Mais do que números, esta atividade traz reflexões importantes sobre o equilíbrio entre desejos, necessidades e responsabilidade com o dinheiro.

O foco principal da prática é a aplicação das medidas de tendência central, com ênfase na média aritmética, a partir de uma situação próxima da realidade dos estudantes do Ensino Médio: o gerenciamento financeiro de um jovem aprendiz. Considerando que muitos alunos já vivenciam experiências profissionais por meio do programa Jovem Aprendiz, propõe-se uma análise dos gastos diários com base em uma simulação de despesas ocorridas nos primeiros 12 dias do mês.

A situação-problema busca responder à seguinte questão: *Será que o estudante poderá manter o mesmo ritmo de gastos diários ao longo do mês, ou precisará reduzir seus gastos para não ficar sem dinheiro?* Essa problemática serve como ponto de partida para discussões e reflexões, incentivando o pensamento crítico e a aplicação prática dos conceitos estatísticos.

Inicialmente, os alunos deverão analisar os dados fornecidos de forma intuitiva, discutir em grupo possíveis estratégias de controle de gastos e anotar suas sugestões. Posteriormente, será realizado o estudo teórico sobre as medidas de tendência central, especialmente a média aritmética. Após a consolidação dos conceitos, os alunos retornarão ao problema gerador para aplicar os conhecimentos adquiridos na resolução da situação proposta.

Objetivos

- Compreender e aplicar o conceito de média aritmética em contextos reais;
- Estimular o raciocínio lógico e o pensamento crítico por meio da análise de dados financeiros;
- Promover a reflexão sobre o consumo responsável e a importância do planejamento financeiro;

Etapas para aplicação da prática

Etapa 1 – Apresentação da Situação-Problema

- Apresentar aos estudantes uma tabela com os gastos diários simulados de um jovem aprendiz nos 12 primeiros dias do mês, como por exemplo a ver Tabela 24.
- Explicar que o objetivo será analisar se ele conseguirá manter esse padrão de gastos ao longo dos 30 dias, considerando um salário mensal fixo (Por exemplo: $R\$723,00$).

Tabela 24 – Quadro de gastos iniciais do Jovem Aprendiz nos 12 primeiros dias do mês

Dia do Mês	Tipo de Gasto	Gasto Diário (R$)
1 (Seg)	Transporte + Lanche + Doce	21,10
2 (Ter)	Transporte + Lanche	17,90
3 (Qua)	Transporte + Lanche + Fast food	22,70
4 (Qui)	Transporte + Lanche	17,80
5 (Sex)	Transporte + Lanche + Extra	31,90
6 (Sáb)	Transporte + Lanche + Rolê	42,70
7 (Dom)	Lanche + Rolê	36,80
8 (Seg)	Transporte + Lanche + App pago	23,40
9 (Ter)	Transporte + Lanche	17,40
10 (Qua)	Transporte + Lanche + Bebida energética	21,90
11 (Qui)	Transporte + Lanche	17,95
12 (Sex)	Transporte + Lanche + Extra	32,40

Fonte: Produção da própria autora (2025)

- Levantar, em grupo, hipóteses, estratégias e possíveis soluções para o problema, antes do estudo formal da teoria.

Etapa 2 – Discussão Inicial e Levantamento de Estratégias

- Os alunos, organizados em grupos, devem debater entre si sobre as possíveis soluções e registrar suas ideias.
- As sugestões serão compartilhadas com a turma para fomentar o diálogo e o pensamento coletivo.

Etapa 3 – Estudo Teórico das Medidas de Tendência Central

- Introduzir os conceitos de média aritmética, mediana e moda, com foco na média.
- Realizar exemplos simples e contextualizados para fixação dos conteúdos.
- Explorar situações reais e cotidianas para mostrar a aplicabilidade do conteúdo.

Etapa 4 – Retorno ao Problema e Aplicação dos Conceitos

- Com base nos conceitos estudados, os grupos devem calcular a média dos gastos diários simulados.
- Comparar a média encontrada com o orçamento mensal (salário do jovem aprendiz).
- Analisar se é possível manter o padrão de gastos ou se será necessário ajustar.

- Calcular, se necessário, a nova média ideal de gastos diários para que o salário dure até o final do mês.

Etapa 5 – Apresentação das Soluções e Discussão Final

- Cada grupo apresentará sua análise, justificando os cálculos e as decisões tomadas.
- O professor poderá conduzir uma discussão coletiva sobre a importância do controle financeiro, uso da média como ferramenta de planejamento e os riscos de gastos excessivos.

Considerações Finais

Esta prática permite aos estudantes vivenciar a aplicação concreta dos conceitos de estatística, especialmente a média aritmética, ao mesmo tempo em que promove uma reflexão crítica sobre o consumo consciente e o planejamento financeiro pessoal. A proposta articula teoria e prática, aproxima os conteúdos escolares da realidade dos jovens e contribui para a formação de cidadãos mais conscientes e preparados para lidar com os desafios cotidianos da vida adulta. Além de desenvolver competências matemáticas, a atividade estimula o trabalho em equipe, a argumentação e a tomada de decisões com base em dados reais e relevantes.

4.4 Prática 04 – Calculando Soluções: A Estatística no Enfrentamento da Estiagem

Figura 44 – Reportagem sobre a estiagem no Espírito Santo e seus impactos na qualidade da água

Estiagem no Espírito Santo: água das torneiras fica salgada por causa do avanço do mar sobre rios

A vazão do Rio Cricaré baixou tanto que, ao invés de o rio desaguar na praia, é o mar que está entrando no rio. Agora tem tanto sal na água que a companhia de abastecimento parou de fazer a captação.

Por Jornal Nacional

23/09/2024 21h02 · Atualizado há 7 meses

Fonte: G1. Disponível em: <https://g1.globo.com/jornal-nacional/noticia/2024/09/23/estiagem-no-espirito-santo-agua-das-torneiras-fica-salgada-por-causa-do-avanco-do-mar-sobre-rios.ghtml>. Acesso em: 13 maio de 2025.

Descrição Geral

A presente prática como objetivo principal promover a compreensão e a aplicação das medidas de tendência central a partir de uma problemática socioambiental contemporânea de grande relevância: a **estiagem no estado do Espírito Santo** e o consequente **avanço da água salgada sobre os rios**, o que tem comprometido a potabilidade da água distribuída às populações ribeirinhas e urbanas. A proposta visa integrar o ensino da Estatística com a Educação Ambiental, fomentando nos estudantes o desenvolvimento do pensamento crítico e da consciência cidadã.

A situação-problema norteadora propõe o seguinte questionamento: *Como uma família pode adequar seu consumo diário de água diante da redução da oferta e da qualidade da água distribuída?*

Para tanto, os alunos serão convidados a analisar uma base de dados simulada, que representa o consumo diário de água de uma família de quatro pessoas ao longo de dez dias. Por meio do cálculo das medidas de tendência central, os estudantes deverão avaliar a sustentabilidade do padrão atual de consumo, comparar os resultados com o limite imposto pela companhia de abastecimento, e propor alternativas viáveis de redução ou reaproveitamento da água.

Como alternativa à simulação proposta, também poderá ser realizado o levantamento de dados reais, por meio da análise das contas de água das próprias famílias dos estudantes, o que tornaria a atividade ainda mais significativa e contextualizada.

A prática favorece uma abordagem interdisciplinar, envolvendo os componentes curriculares de Estatística, Consumo Responsável e Educação Tributária, Biologia, Química e Geografia. Dessa forma, promove a integração entre conteúdos estatísticos e questões socioambientais contemporâneas, ampliando a compreensão dos alunos sobre os desafios relacionados à gestão sustentável dos recursos hídricos.

Objetivos

- Compreender e aplicar os conceitos de média aritmética, mediana e moda no tratamento e análise de dados;
- Interpretar dados estatísticos referentes ao consumo doméstico de água e relacioná-los a situações reais de escassez hídrica;
- Estimular o pensamento crítico e a tomada de decisões com base em evidências matemáticas;
- Promover a interdisciplinaridade e a reflexão sobre práticas sustentáveis de uso dos recursos naturais.

Etapas para Aplicação da Prática

Etapa 1 – Contextualização da Situação-Problema

- Iniciar a atividade com a leitura da manchete publicada no portal G1: *"Estiagem no Espírito Santo: água das torneiras fica salgada por causa do avanço do mar sobre rios"* (G1 - Jornal Nacional, 23/09/2024).
- Apresentar aos alunos os impactos ambientais e sociais da estiagem prolongada e da intrusão salina, destacando a necessidade de repensar os padrões de consumo de água.
- Apresentar a tabela abaixo, que simula o consumo diário de água de uma família de quatro pessoas durante dez dias consecutivos:

Tabela 25 – Consumo Diário de Água por uma Família de 4 Pessoas (em Litros)

Dia	**Consumo (litros)**
1 (Seg)	540
2 (Ter)	510
3 (Qua)	525
4 (Qui)	580
5 (Sex)	490
6 (Sáb)	610
7 (Dom)	620
8 (Seg)	505
9 (Ter)	495
10 (Qua)	550

Fonte: Produção da própria autora (2025)

- Propor o seguinte questionamento: *Considerando que a companhia de abastecimento estabeleceu um limite de 15.000 litros por mês por residência durante a estiagem, essa família está dentro do limite permitido? Qual seria o consumo ideal diário para respeitar esse teto?*

Etapa 2 – Discussão Inicial e Levantamento de Hipóteses

- Organizar os estudantes em grupos e propor uma discussão inicial com base nos seguintes pontos:
  - Qual o perfil de consumo apresentado na tabela?
  - O padrão atual é sustentável dentro da nova realidade hídrica?
  - Quais mudanças de hábito poderiam contribuir para a redução do consumo?

- Os grupos devem registrar suas hipóteses, argumentos e sugestões preliminares antes do estudo formal dos conceitos estatísticos.

Etapa 3 – Aplicação Prática dos Conceitos das Medidas de Tendência Central

- Os estudantes deverão:
    1. Calcular a média aritmética do consumo diário apresentado na tabela;
    2. Determinar a mediana e a moda dos dados;
    3. Estimar o consumo mensal com base na média diária obtida (multiplicando por 30);
    4. Comparar esse valor com o limite imposto de 15.000 litros mensais;
    5. Caso o limite seja ultrapassado, calcular a nova média ideal para manter o consumo dentro do valor permitido;
    6. Avaliar se a mediana e/ou a moda poderiam representar melhor o comportamento do consumo familiar em determinados contextos.
- Com base nas análises, os grupos devem elaborar propostas de adaptação, como: reutilização da água da máquina de lavar, captação de água da chuva, banhos mais curtos, entre outros.

Etapa 4 – Apresentação e Socialização dos Resultados

- Cada grupo apresentará à turma os resultados dos cálculos realizados, bem como suas reflexões e propostas para um consumo consciente de água.
- O professor poderá fomentar uma discussão crítica a respeito do papel da Matemática como ferramenta de análise e transformação social.
- Promover uma reflexão final sobre o impacto das ações individuais e coletivas no enfrentamento das crises ambientais.

Considerações Finais

A prática desenvolvida propicia um ambiente de aprendizagem significativo, em que os estudantes utilizam as medidas de tendência central não apenas como ferramentas matemáticas abstratas, mas como instrumentos concretos para interpretar e intervir na realidade. Ao integrar o ensino da Estatística com um problema socioambiental urgente — a escassez de água —, a atividade estimula o pensamento crítico, o raciocínio lógico, a argumentação com base em dados e o engajamento cidadão.

O uso conjunto da média, da mediana e da moda permite uma leitura mais abrangente do conjunto de dados, possibilitando que os alunos compreendam diferentes aspectos do consumo doméstico de água e reconheçam a importância de tomar decisões sustentáveis. Além disso, a proposta contribui para o desenvolvimento de competências gerais da BNCC, como o pensamento científico, o protagonismo social e a responsabilidade com o meio ambiente.

4.5 Prática 05 – Aposta Inteligente: Como a Probabilidade Informa Nossas Escolhas

"Qual é a soma mais provável ao lançar dois dados?"

Descrição Geral

Esta prática tem como objetivo motivar os estudantes a compreender o conceito de **probabilidade** por meio de uma situação-problema envolvendo o lançamento de dois dados. A atividade propõe que os alunos reflitam sobre quais somas têm maior ou menor chance de ocorrer e como o conhecimento matemático pode ser usado para tomar decisões fundamentadas.

Inicialmente, o professor apresenta uma pergunta desafiadora: *"Se eu lançar dois dados, qual será a soma mais provável que pode aparecer?"* Ou ainda: *"Se você tivesse que adivinhar a soma dos dados antes de eu lançá-los, qual soma você escolheria para ter mais chances de acertar?"*

A partir dessa indagação, os estudantes são convidados a formular hipóteses e discutir estratégias. Em seguida, são incentivados a representar todas as possibilidades do lançamento de dois dados, construindo a tabela de combinações possíveis e suas respectivas somas (total de 36 possibilidades). A visualização desse quadro permite observar quais somas são mais frequentes e favorece a compreensão da distribuição de probabilidades.

Com base nisso, os conceitos de *evento, espaço amostral, probabilidade e frequência relativa* são introduzidos ou aprofundados. A prática culmina com atividades práticas e simulações que conectam os conteúdos à tomada de decisões em situações de incerteza, como jogos, sorteios, previsões e cenários reais.

Objetivos

- Compreender o conceito de espaço amostral e eventos;
- Calcular probabilidades em experimentos aleatórios simples;

- Comparar diferentes estratégias com base em evidências matemáticas;
- Utilizar representações (tabelas, gráficos e diagramas) para organizar dados e analisar padrões;
- Desenvolver o pensamento lógico e habilidades de argumentação;
- Refletir sobre a aplicabilidade da probabilidade em decisões cotidianas.

Etapas para aplicação da prática

Etapa 1 – Apresentação do Desafio

- O professor propõe oralmente: *"Se você tivesse que adivinhar a soma dos dois dados que vou lançar, qual soma você escolheria para ter mais chances de acertar?"*
- Os alunos registram individualmente ou em duplas suas hipóteses, justificando suas escolhas com base na intuição.
- O professor lança os dados algumas vezes e registra os resultados como exemplo inicial.

Etapa 2 – Discussão e Levantamento de Estratégias

- Em grupos, os alunos discutem possíveis estratégias para descobrir qual soma tem maior chance de acontecer.
- Alguns podem propor simulações de jogadas (lançar dados várias vezes e registrar as somas).
- Outros podem começar a tentar listar combinações possíveis.

Etapa 3 – Construção do Espaço Amostral

- O professor, com apoio dos alunos, constrói o quadro de todas as 36 combinações possíveis ao lançar dois dados (de 1 a 6).
- Cada combinação (por exemplo, (1,1), (1,2), ..., (6,6)) é registrada e associada à sua soma.
- Em seguida, constrói-se uma tabela de frequências das somas, mostrando quantas vezes cada valor pode ocorrer.

Etapa 4 – Cálculo de Probabilidades

- Os alunos calculam a probabilidade de ocorrência de cada soma:
  - Soma 2: $\frac{1}{36}$
  - Soma 3: $\frac{2}{36}$
  - Soma 4: $\frac{3}{36}$
  - Soma 5: $\frac{4}{36}$
  - Soma 6: $\frac{5}{36}$
  - Soma 7: $\frac{6}{36}$ (maior probabilidade)
  - Soma 8: $\frac{5}{36}$
  - Soma 9: $\frac{4}{36}$
  - Soma 10: $\frac{3}{36}$
  - Soma 11: $\frac{2}{36}$
  - Soma 12: $\frac{1}{36}$
- Comparar os resultados com as hipóteses iniciais dos alunos.

Integração com a Estatística

Embora o foco inicial da prática seja a probabilidade, ela oferece uma rica oportunidade para trabalhar conceitos estatísticos de forma integrada. Ao lançar os dados várias vezes e registrar as somas obtidas, os estudantes geram um conjunto de dados reais, que podem ser organizados em tabelas de frequência e representados por gráficos.

Com esses dados, é possível calcular e analisar:

- A frequência absoluta e frequência relativa de cada soma;
- A moda, ou seja, a soma que mais aparece nas simulações;
- A discrepância entre a frequência observada e a probabilidade, introduzindo a ideia de variabilidade;

Dessa forma, os alunos não apenas entendem a teoria da probabilidade, mas também aplicam ferramentas estatísticas para validar hipóteses, interpretar dados observados e refletir sobre o comportamento dos fenômenos aleatórios no mundo real.

Etapa 5 – Aplicação em Situações de Decisão

- Discutir como a probabilidade pode ser usada em outras decisões: previsões meteorológicas, seguros, jogos de azar, etc.
- Refletir: É sempre vantajoso escolher a opção mais provável? Por quê?

Curiosidade: E se fossem três dados?

Após concluir a análise dos dois dados, o professor pode propor uma nova questão para instigar a curiosidade dos alunos:

*"E se fossem lançados **três dados**, qual seria a soma mais provável?"*

Neste caso, o número de combinações possíveis passa a ser $6 \cdot 6 \cdot 6 = 216$, e a soma mais provável ao lançar três dados é:

$$\boxed{11}$$

A distribuição é simétrica, e o valor mais provável de soma ao lançar n dados de 6 faces pode ser estimado pela fórmula:

$$\text{Soma mais provável} = 3.5 \cdot n \quad .$$

Como 3.5 é a média aritmética de um único dado (valores de 1 a 6), essa fórmula nos dá a soma que está no centro da distribuição. Para três dados, temos:

$$3.5 \cdot 3 = 10.5 \quad .$$

Portanto, as somas mais prováveis serão aquelas ao redor de 10 e 11, sendo o 11 o valor mais comum na prática. A fórmula é útil para estimar a região central das somas possíveis ao lançar múltiplos dados.

Considerações Finais

Esta prática promove o desenvolvimento de habilidades fundamentais para a educação matemática, como a análise crítica, a resolução de problemas e a tomada de decisões baseadas em dados. Ao articular experimentação, argumentação e visualização, os estudantes percebem a relevância da probabilidade em contextos reais e desenvolvem maior confiança para aplicá-la em diferentes situações. Além disso, a atividade favorece a aprendizagem colaborativa, a escuta ativa e o respeito a diferentes estratégias, consolidando a matemática como uma ferramenta para a leitura do mundo.

4.6 Prática 06 – Quando a Média Engana: Entendendo a Dispersão

Descrição Geral

A prática tem como objetivo o desenvolvimento da compreensão dos alunos sobre as medidas de dispersão, complementando as análises feitas com as medidas de tendência central, especialmente no contexto de dados que apresentam grande variação. O problema central abordado é: Como as diferentes variações de desempenho em grupos de alunos, com médias semelhantes, podem influenciar decisões educacionais?

A partir de dados simulados sobre o desempenho de dois grupos de alunos em uma prova, a atividade visa aplicar conceitos estatísticos (como amplitude, desvio padrão e variância) para avaliar a **consistência do desempenho** de cada grupo, além de discutir como essas medidas podem influenciar práticas pedagógicas e decisões na gestão educacional.

A situação-problema proposta aos alunos é: *Como você avaliaria o desempenho de dois grupos de alunos, ambos com a mesma média, mas com comportamentos diferentes em relação à variação das notas? Como isso pode influenciar as estratégias de ensino?*

Objetivos

- Compreender e aplicar as medidas de dispersão (amplitude, desvio padrão e variância) no tratamento e análise de dados.
- Relacionar dados sobre desempenho acadêmico com o conceito de variabilidade e discutir a importância dessa informação.
- Estimular o pensamento crítico sobre a utilidade das medidas de tendência central e dispersão na análise de dados reais.

Etapas para Aplicação da Prática

Etapa 1 – Contextualização da Situação-Problema

- **Apresentação do Problema:** Divida a turma em grupos pequenos e apresente a seguinte questão para discussão: *Imagine que você é um coordenador de um curso preparatório para vestibular. Você recebeu os resultados das notas de 10 alunos que fizeram uma prova, e, a partir da média, você precisa avaliar o desempenho dos grupos. A média das notas é de 5,0. Agora, o que você faria se soubesse que um dos grupos tem notas bem homogêneas, mas o outro grupo tem uma grande variação nas suas notas? Qual grupo, em sua opinião, teve um desempenho mais consistente?*

- **Distribuição da Tabela de Notas:** Apresente aos alunos a tabela com as notas de dois grupos de alunos:

Tabela 26 – Notas dos Grupos de Alunos

Grupo	**Notas**
Grupo 1	5, 5, 5, 5, 5 ,5, 5, 5, 5, 5
Grupo 2	10, 10, 10, 10, 10, 0, 0, 0, 0, 0

Fonte: Produção da própria autora (2025)

- Pergunte aos alunos: *O que as médias dos grupos indicam sobre seu desempenho?* Pergunte também: *Como você avaliaria o desempenho de cada grupo levando em consideração apenas a média de 5,0 para ambos os grupos?*

Etapa 2 – Discussão Inicial e Levantamento de Hipóteses

- **Discussão em Grupo:** Organize os estudantes em grupos para discutirem a seguinte questão:
  - O que as medidas de tendência central (média, mediana e moda) indicam sobre o desempenho de cada grupo?
  - Como as diferentes distribuições das notas podem influenciar a avaliação do desempenho dos alunos?
  - O que pode ser perdido ou distorcido ao olhar apenas para a média?
- **Registro de Hipóteses:** Solicite que cada grupo registre suas hipóteses sobre a relação entre média e variabilidade nas notas, refletindo sobre os pontos discutidos.

Etapa 3 – Introdução às Medidas de Dispersão

- **Explicação sobre Medidas de Dispersão:** Explique aos alunos que, para uma análise mais completa, é necessário avaliar como as notas se distribuem em torno da média. As medidas de dispersão ajudam a entender a variação dos dados.
- **Cálculos das Medidas de Dispersão:** Oriente os alunos no cálculo das medidas de dispersão para os dois grupos:
  - **Amplitude:** A diferença entre o maior e o menor valor no conjunto de dados.
  - **Desvio Padrão e Variância:** Indicadores que mostram o quanto os dados se afastam da média.

Etapa 4 – Discussão dos Resultados das Medidas de Dispersão

- **Apresentação dos Cálculos:** Após os cálculos, apresente os resultados para os alunos.
- **Discussão Crítica:** Pergunte aos alunos:
  - O que as medidas de dispersão nos mostram sobre os dois grupos que a média não revela?
  - Como você usaria essas informações em uma análise real, por exemplo, para identificar se os alunos de um grupo precisam de mais atenção ou se todos estão performando de forma semelhante?

  Reflita sobre como a variabilidade das notas pode influenciar a análise do desempenho dos alunos.

Etapa 5 – Contextualização no Mundo Real

- **Aplicação Prática em Outro Contexto:** Para ilustrar a importância das medidas de dispersão, forneça outro exemplo real:
  - Imagine que você está analisando os custos mensais de consumo de energia de duas casas diferentes, ambas com uma média de consumo de 200 kWh por mês. A casa A tem consumos muito próximos da média, enquanto a casa B tem picos de consumo extremamente altos em alguns meses (por exemplo, um mês com 800 kWh e outro com 100 kWh).
- **Discussão sobre Consumo de Energia:** Pergunte aos alunos:
  - Como você avaliaria a situação das duas casas apenas com a média de consumo?
  - A média sozinha é suficiente para entender a realidade do consumo em ambas as casas? Como a dispersão ajuda a complementar a análise?

Etapa 6 – Conclusão e Reflexão Final

- **Reflexão sobre a Aplicação das Medidas de Dispersão:** Finalize a atividade com uma reflexão crítica:
  - Como as medidas de dispersão ajudam a tomar decisões mais informadas em diferentes contextos, como no ambiente educacional e nas finanças pessoais?
  - Por que a combinação das medidas de tendência central e dispersão é importante para uma análise mais precisa dos dados?

Considerações Finais

Esta prática proporciona aos alunos uma oportunidade de entender como as **medidas de dispersão** complementam as **medidas de tendência central** na análise de dados, ajudando a revelar padrões e variações que a média, sozinha, não seria capaz de demonstrar. Além disso, ao relacionar esses conceitos com exemplos do cotidiano, como desempenho acadêmico e consumo de energia, a atividade possibilita uma aplicação prática da Matemática em contextos reais e significativos.

O uso conjunto das medidas de dispersão e das medidas de tendência central permite uma análise mais precisa e profunda dos dados, contribuindo para a formação de um pensamento crítico e para a tomada de decisões informadas. A prática também promove a interdisciplinaridade, envolvendo aspectos de Estatística e do cotidiano, preparando os alunos para a análise e interpretação de dados em diversas áreas.

4.7 Prática 07 – Quanto Você Confia no que Sabe? Julgamentos sob Incerteza

Descrição Geral

A presente prática tem como objetivo central fomentar a capacidade dos estudantes de expressar, quantificar e avaliar suas incertezas diante de informações incompletas, utilizando conceitos estatísticos de probabilidade subjetiva [1] e regra de escore quadrático. A atividade parte de um conjunto de afirmações de conhecimento geral, sobre as quais os alunos devem estimar a veracidade atribuindo uma probabilidade. Posteriormente, esse julgamento é avaliado por meio de uma função de penalidade estatística.

A situação-problema proposta é: *Como avaliar se uma pessoa toma boas decisões quando está diante de informações incertas? Como representar numericamente sua confiança?*

Essa prática foi adaptada da sugestão proposta por Rifo (2024) em sua obra *Probabilidade e Estatística: Aspectos de tomada de decisões e incerteza para o Ensino Fundamental e Médio* , que apresenta abordagens inovadoras para integrar o raciocínio probabilístico ao processo educacional com base em situações reais e de incerteza.

A atividade promove o raciocínio estatístico, a autorreflexão e o pensamento crítico, demonstrando que boas decisões não dependem apenas de estar certo, mas da forma como se lida com a incerteza de maneira quantificável.

[1] A abordagem subjetivista de probabilidade pode ser aplicada em situações de incerteza, nas quais é necessário tomar uma decisão. Nesses casos, embora não seja possível determinar com precisão todas as implicações da escolha, o conceito de probabilidade permite expressar quantitativamente a incerteza envolvida (MOREIRA, 2015).

Objetivos

- Estimular o pensamento probabilístico em situações de incerteza;
- Compreender e aplicar a regra de escore quadrático como medida da qualidade de julgamentos;
- Analisar comparativamente diferentes estratégias de decisão com base em informação parcial;
- Desenvolver habilidades de argumentação e reflexão sobre os próprios conhecimentos;

Etapas para Aplicação da Prática

Etapa 1 – Introdução à Situação-Problema

- Apresentar aos alunos o seguinte contexto: *Muitas decisões na vida são tomadas com base em julgamentos subjetivos. Como saber se confiamos demais ou de menos nas nossas opiniões?*
- Explicar que o objetivo da atividade é estimar a veracidade de uma série de afirmações usando probabilidades subjetivas — e depois avaliar a qualidade desses julgamentos com base em uma regra estatística de penalidade.
- Introduzir o conceito de probabilidade subjetiva: uma forma de expressar numericamente o grau de confiança de que uma informação é verdadeira.
- Apresentar a regra de escore quadrático, que penaliza decisões imprecisas com base na confiança declarada:
  - Se o evento for **verdadeiro**: $\text{perda}(p) = 100 \times (1 - p)^2$;
  - Se o evento for **falso**: $\text{perda}(p) = 100 \times p^2$.

  Conforme Rifo (2024), ao considerarmos um evento que de fato ocorre, é possível observar diferentes cenários de previsão. Quando a probabilidade atribuída é $p = 1$, isso significa que a previsão estava totalmente correta, sem qualquer penalização. Já uma previsão com $p = 0.9$ indica uma grande confiança de que o evento seria verdadeiro, resultando em um pequeno escore de $(1-0.9)^2 = 0.01$, ou, se multiplicado por 100, o valor 1. Com uma previsão de $p = 0.7$, o escore sobe para 9. Para a previsão $p = 0.5$, que reflete a crença de que o evento tem a mesma probabilidade de ser verdadeiro ou falso, o escore é 25. Por fim, quando a previsão é $p = 0$, ou seja, a expectativa de que o evento seja falso, o escore é o maior possível, igual a 100. Esses valores são sintetizados na Figura 45, que ilustra as perdas associadas à regra quadrática para diferentes probabilidades atribuídas, tanto no caso de o evento ser verdadeiro quanto falso.

Figura 45 – Valor da perda, de acordo com a regra quadrática, para diversas probabilidades atribuídas, p, se o evento for verdadeiro ou se o evento for falso

	probabilidade atribuída *p*										
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
evento verdadeiro	100	81	64	49	36	25	16	9	4	1	0
evento falso	0	1	4	9	16	25	36	49	64	81	100

Fonte: (RIFO, 2024), p.7

- Ressaltar que quanto **menor o score**, melhor a qualidade do julgamento.

Etapa 2 – Atividade de Julgamento com Probabilidades Subjetivas

- Dividir a turma em dois grupos: **Grupo A** e **Grupo B**.
- Entregar aos dois grupos o questionário abaixo, mas apenas o Grupo A responde neste momento.
- Os alunos do **Grupo A** atribuirão, individualmente, uma probabilidade (com uma casa decimal) à veracidade de cada afirmação, sem consultar fontes externas.
- **Questionário – Veracidade das Afirmações (atribua p entre 0 e 1):**
  1. Em torno de 15% da população mundial é muçulmana.
  2. O perigeu é o ponto mais afastado da Terra na órbita de um satélite artificial.
  3. O primeiro romance de Jorge Amado foi *O país do carnaval* .
  4. O número 23711 é primo.
  5. A ilha de Nova Bretanha está localizada no arquipélago de Bismarck.
  6. A temperatura média da superfície de Vênus é em torno de 450°C .
  7. O diprotodonte foi extinto há aproximadamente 25 mil anos.
  8. Santiago do Chile está mais ao sul que Campinas.
  9. *Domingo no parque* é uma música de Gilberto Gil .
  10. O mito grego de Eurídice explica as estações do ano.
- Após a coleta das respostas do Grupo A, o professor jucalcula a **média das probabilidades** (valor de p) atribuídas para cada questão.

Etapa 3 – Segunda Rodada com Informação Adicional (Grupo B)

- O **Grupo B** terá acesso às médias de p do Grupo A para cada afirmação.
- Com base nessa informação, o Grupo B responde o mesmo questionário, estimando sua própria probabilidade p para cada questão.
- Essa etapa visa observar se a **influência de informações alheias** melhora a tomada de decisão.

Etapa 4 – Cálculo do Score e Comparação de Resultados

- Divulgar o gabarito com as respostas corretas das afirmações.

  Resposta para as 10 afirmações: F, F, V, F, V, V, F, V, V, F.
- Cada estudante deve calcular seu escore individual com base nas fórmulas:
  - Se a afirmação for verdadeira: $100 \times (1-p)^2$;
  - Se a afirmação for falsa: $100 \times p^2$.
- Somar os scores de todas as 10 afirmações para obter o **score total individual**.
- Comparar os resultados entre os grupos:
  - Qual grupo teve menor escore médio?
  - Houve melhora no julgamento com base em informações do grupo anterior?

Etapa 5 – Discussão e Reflexão

- Conduzir uma discussão com base em perguntas orientadoras:
  - O que foi mais difícil: saber ou decidir a confiança sobre o que se sabia?
  - Qual foi o impacto de errar com muita certeza (p alto em afirmação falsa)?
  - Como isso se aplica a decisões cotidianas com base em “achismos”?
- Solicitar uma breve reflexão escrita sobre:
  - O que aprendi sobre expressar incerteza?
  - Em que situações da vida essa prática pode ajudar?
  - O que faria diferente da próxima vez que tivesse que tomar uma decisão incerta?

Considerações Finais

Esta prática permite que os estudantes vivenciem o raciocínio estatístico de maneira aplicada, refletindo sobre suas próprias crenças, confiabilidade do conhecimento e impactos de julgamentos mal fundamentados. Ao adotar uma abordagem quantitativa da incerteza, a atividade promove uma compreensão mais refinada da tomada de decisão.

Baseada na proposta de Rifo (2024) a atividade conecta a Estatística com a realidade dos alunos, estimula o autoconhecimento e promove o desenvolvimento de competências da BNCC, como o pensamento crítico, a argumentação e a responsabilidade com as próprias decisões.

4.8 Prática 08 – Teste de Hipóteses Aplicado: Investigando Suspeitas com Estatística

Descrição Geral

Esta prática tem como objetivo introduzir os estudantes ao conceito de **teste de hipóteses** por meio de uma situação-problema realista: verificar, a partir de dados amostrais, se pacotes de galinhas comercializadas como “pé duro” realmente possuem o peso médio de 1kg (1000g), como indicado pelo fornecedor.

A proposta parte considerando uma suspeita fictícia levantada por um supermercado, que observou variações nos pesos dos pacotes. Para investigar se essas variações são aceitáveis ou se há indícios de fraude, os estudantes são desafiados a utilizar ferramentas estatísticas de inferência, especificamente o teste de hipóteses.

A prática visa tornar os conceitos de *hipótese nula, hipótese alternativa, valor crítico, estatística de teste e rejeição da hipótese* mais significativos, ao mesmo tempo em que evidencia a importância da estatística na tomada de decisões baseadas em evidências.

Objetivos

- Compreender o papel do teste de hipóteses em situações reais;
- Formular e interpretar hipóteses nula e alternativa;
- Calcular e interpretar a estatística de teste (Z);
- Tomar decisões fundamentadas em dados e critérios estatísticos;
- Refletir sobre o papel da estatística na detecção de erros e fraudes;
- Aplicar conceitos de inferência estatística em contextos cotidianos.

Etapas para Aplicação da Prática

Etapa 1 – Apresentação do Desafio

- O professor apresenta a situação: *Um supermercado suspeita que os pacotes de galinhas 'pé duro', que deveriam pesar 1kg, podem estar com peso inferior. Como podemos investigar essa suspeita de forma científica?*
- Os estudantes registram suas hipóteses iniciais, com base na intuição.
- Pergunta orientadora: *Você acha que os pacotes estão, em média, com menos de 1000g? Por quê?*

Etapa 2 – Coleta e Organização dos Dados

- O professor apresenta uma amostra de 64 pacotes, cujos pesos (em gramas) estão registrados a seguir:

Tabela 27 – Pesos amostrados dos pacotes de galinha (em gramas)

985	1012	995	1003	998	1010	992	1001
979	1005	988	1020	1008	983	1002	1011
992	1007	985	1010	996	1004	1006	987
981	1002	993	1004	990	998	1000	990
989	994	1001	1010	1015	1006	984	988
992	1003	980	1001	1009	989	1005	995
988	1018	997	985	1007	989	1002	1000
986	990	996	1000	1006	981	1012	1003

Fonte: Produção da própria autora (2025)

- Os alunos analisam a tabela e discutem se visualmente conseguem identificar algum padrão de subpeso.

Etapa 3 – Formulação do Teste de Hipóteses

- O professor orienta os alunos a formular as hipóteses estatísticas:
  - Hipótese nula (H_0): $\mu = 1000$.
  - Hipótese alternativa (H_1): $\mu \neq 1000$.
- Explica-se que se trata de um **teste bilateral**, pois queremos saber se o peso médio é diferente (para mais ou para menos) de 1000g.

Etapa 4 – Cálculo da Estatística de Teste

- O professor introduz os conceitos fundamentais de **teste de hipótese**, explicando sua importância na verificação de afirmações sobre parâmetros populacionais com base em dados amostrais.

- Em seguida, o professor solicita que os alunos calculem a **média amostral** a partir da tabela com os 64 valores (pesos dos pacotes), e informa o **desvio padrão populacional** fornecido pela empresa.

  - Média amostral: $\bar{x} = 997{,}78$.
  - Desvio padrão populacional: $\sigma = 4$.
  - Tamanho da amostra: $n = 64$.
  - Nível de significância: $\alpha = 0{,}05$.

- Os alunos devem então aplicar o **teste Z**, utilizando a Equação 3.32:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- Substituindo os valores fornecidos: $Z = \dfrac{997{,}78 - 1000}{4/\sqrt{64}} = \dfrac{-2{,}22}{0{,}5} = -4{,}44$

- Para um teste **bilateral** com $\alpha = 0{,}05$, os valores críticos são:

$$Z = \pm 1{,}96$$

  Fonte da Tabela z:
  https://www.ime.unicamp.br/ cnaber/tabela_normal.pdf.
  Acesso em: 1 maio 2025.

- Como $Z = -4{,}44$ está fora da região de não rejeição (isto é, $Z < -1{,}96$), os alunos devem concluir que:

  Há evidências suficientes para rejeitar a hipótese nula.

- Conclui-se, portanto, que a média dos pacotes difere significativamente de 1000g, o que pode indicar erro no empacotamento ou inconsistência no processo.

Etapa 5 – Interpretação e Tomada de Decisão

- Os alunos interpretam o resultado: *Há evidências estatísticas de que o peso médio dos pacotes não é igual a 1000g.*
- Refletem: *Isso significa que o fornecedor cometeu fraude? Ou pode haver outra explicação?*
- O professor destaca o papel do erro tipo I: rejeitar H_0 quando ela é verdadeira, e o risco envolvido.

Atividades de Consolidação

1. Se a média fosse 999,5g, o teste ainda rejeitaria H_0?
2. Como a decisão seria afetada se o desvio padrão fosse 8g?
3. O que mudaria se fosse um teste unilateral?
4. Por que não é possível afirmar com 100% de certeza que houve fraude?

Considerações Finais

Esta prática promove o desenvolvimento do raciocínio estatístico e da argumentação lógica, ao explorar uma situação real de forma investigativa. Os estudantes compreendem que, mesmo com incertezas, é possível usar dados e métodos científicos para tomar decisões fundamentadas.

A atividade favorece a aprendizagem colaborativa, a escuta ativa e o uso da matemática como uma ferramenta para interpretar o mundo e resolver problemas concretos.

5 Conclusão

Este trabalho teve como objetivo central a elaboração de um manual de orientação didático-pedagógica destinado a apoiar professores no planejamento e desenvolvimento de aulas do componente curricular Estatística, inserido no itinerário formativo de Educação Financeira e Fiscal, em consonância com as diretrizes curriculares do Estado do Espírito Santo e da Base Nacional Comum Curricular (BNCC). Para isso, foram cumpridos os objetivos específicos estabelecidos, que contemplaram a análise detalhada das diretrizes da BNCC e do currículo estadual do Espírito Santo referentes ao ensino de Estatística no contexto dos itinerários formativos, conforme apresentado no Capítulo 2. Ademais, foram propostas sugestões práticas, exemplos ilustrativos e atividades pedagógicas voltadas para um ensino contextualizado e significativo da Estatística, explorados nos Capítulos 3 e 4. Ressalta-se, ainda, a promoção da integração entre os conteúdos estatísticos, o uso de tecnologias digitais — exemplificado no exemplo 3.9.10 — e a articulação com os contextos cotidianos dos estudantes, com o propósito de fomentar o desenvolvimento das competências previstas no currículo.

A partir da análise das mudanças decorrentes da implementação do Novo Ensino Médio, em especial da criação dos itinerários formativos, constatou-se a urgência de repensar práticas pedagógicas diante dos novos desafios impostos aos docentes. Muitos professores passaram a atuar em componentes curriculares para os quais não receberam formação específica ou aprofundada durante sua trajetória acadêmica, o que evidencia lacunas tanto na formação inicial quanto na disponibilidade de materiais pedagógicos compatíveis com as exigências da nova proposta curricular.

O estudo discutiu os objetos de conhecimento da Estatística previstos na BNCC e propôs sugestões práticas para o ensino desse componente. Ainda que a aplicação de dados reais e regionais do Espírito Santo não tenha sido amplamente desenvolvida ao longo desta pesquisa, reconhece-se a relevância dessa abordagem. Exemplos contextualizados com dados locais — como estatísticas socioeconômicas, ambientais e educacionais dos municípios capixabas — são estratégias valiosas para tornar a aprendizagem mais significativa e promover maior engajamento dos estudantes, uma vez que aproximam o conteúdo escolar de suas vivências cotidianas.

O manual proposto visa suprir uma lacuna identificada por professores que enfrentam dificuldades para planejar aulas de Estatística de forma integrada, significativa e aplicada. A proposta busca fomentar o desenvolvimento de competências como o pensamento crítico, a argumentação com base em evidências e a capacidade de tomar decisões fundamentadas, essenciais para a formação de cidadãos atuantes e conscientes em um mundo orientado por dados.

Entre as contribuições deste trabalho, destaca-se a oferta de um referencial didático alinhado ao currículo oficial, que pode auxiliar os professores em sua prática docente ao sugerir atividades e caminhos possíveis para o ensino de Estatística de forma contextualizada. Apesar dos limites da pesquisa — como a não aplicação prática do manual em sala de aula e a a ausência de um processo sistemático de testagem e retorno por parte dos professores —, o trabalho abre caminhos para estudos futuros.

Como desdobramento, recomenda-se a aplicação piloto do manual em escolas públicas da rede estadual, acompanhada da realização de oficinas de formação continuada voltadas aos professores que atuam nos componentes curriculares introduzidos com o Novo Ensino Médio. Sugere-se, ainda, a produção de materiais complementares que integrem dados reais de municípios capixabas, com o objetivo de aproximar o ensino da Estatística do cotidiano dos estudantes e potencializar sua aprendizagem por meio de contextos próximos à realidade dos alunos.

Conclui-se que esta dissertação representa uma contribuição relevante para a qualificação do ensino de Estatística no Ensino Médio, oferecendo um ponto de partida para a construção de práticas pedagógicas mais contextualizadas, críticas e efetivas. Ao reconhecer a importância de articular teoria, prática e realidade local, o trabalho reafirma o papel da Estatística na formação de sujeitos capazes de compreender, interpretar e intervir em um mundo cada vez mais mediado por dados.

Referências

ALECRIM, E. *O que é Big Data, para que serve e como funciona*. 2024. <https://www.infowester.com/big-data.php>. Acesso em: 23 dez. 2024. Citado na página 27.

ASSIS, J. P. de; SOUSA, R. P. de; LINHARES, P. C. F. *Testes de Hipóteses Estatísticas*. Mossoró, RN: EdUFERSA, 2020. Acesso em: 16 maio 2025. Disponível em: <https://livraria.ufersa.edu.br/wp-content/uploads/sites/165/2020/08/testes-de-hipoteses-estatisticas-edufersa.pdf>. Citado 2 vezes nas páginas 125 e 126.

BRASIL. *Base Nacional Comum Curricular*. 2017. Acesso em: 04 maio 2025. Disponível em: <https://basenacionalcomum.mec.gov.br/images/BNCC_20dez_site.pdf>. Citado na página 16.

Brasil. *Lei nº 13.415, de 16 de fevereiro de 2017*. 2017. Acesso em: 24 jul. 2024. Disponível em: <https://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Lei/L13415.htm>. Citado na página 17.

BRASIL. *PORTARIA Nº 1.432, DE 28 DE DEZEMBRO DE 2018*. 2018. Accessed: 17 jun. 2024. Disponível em: <https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/70268199.> Citado na página 20.

CABARITI, E. *Distribuição Normal*. 2025. Acesso em: 8 de abril 2025. Disponível em: <https://www.ime.usp.br/~salles/fatec/estatistica/aula8.pdf>. Citado na página 121.

CARVALHO, G. O. de. *Ocorrência de Mercúrio em pelos de Tamanduá-Bandeira (Myrmecophaga tridactyla): monitoramento nas rodovias do Mato Grosso do Sul e na Estação Ecológica de Santa Bárbara, São Paulo, Brasil*. Dissertação (Dissertação de Mestrado) — Universidade Federal do Rio de Janeiro, Instituto de Biofísica Carlos Chagas Filho, Rio de Janeiro, 2018. Orientador: João Paulo Machado Torres. Co-orientador: Rodrigo Ornellas Meire. 93 f.: il.; 31 cm. Disponível em: <file:///C:/Users/dsluz/Downloads/GOCarvalho-Dissertacao-tamanduamercurio.pdf>. Citado na página 49.

CHAVANTE, E.; PRESTES, D. *Quadrante Matemática, 3º ano: ensino médio*. 1ª. ed. São Paulo: Edições SM, 2016. Citado na página 70.

Instituto Brasileiro de Geografia e Estatística - IBGE. *Normas de apresentação tabular*. 1993. <https://biblioteca.ibge.gov.br/visualizacao/livros/liv23907.pdf>. Acesso em: 06 jan. 2025. Citado 2 vezes nas páginas 39 e 40.

JAMES, B. C. A. *Probabilidade: Um Curso na Teoria das Probabilidades*. 2. ed. Rio de Janeiro: IMPA – Instituto Nacional de Matemática Pura e Aplicada, 2007. Citado na página 119.

MAGNUSSON, D.; CAIRNS, R. B. Developmental science: Toward a unified framework. In: CAIRNS, R. B.; ELDER, G. H. J.; COSTELLO, J. E. (Ed.). *Developmental science*. Cambridge, UK: Cambridge University Press, 2009. p. 7–30. Citado na página 132.

MOREIRA, A. d. P. M. *Aplicações da teoria da decisão e probabilidade subjetiva em sala de aula do ensino médio*. Campinas, SP: [s.n.], 2015. Disponível em: <https://sca.profmat-sbm.org.br/profmat_tcc.php?id1=1973&id2=72861>. Citado na página 148.

MORETTIN, P. A.; BUSSAB, W. O. *Estatística Básica*. 6ª. ed. São Paulo: Saraiva, 2010. Citado 5 vezes nas páginas 18, 26, 65, 70 e 110.

MORLING, B. *Research Methods in Psychology: Evaluating a World of Information*. [S.l.]: W. W. Norton & Company, 2015. Citado na página 131.

OLIVEIRA, J. S. C. de. *Estatística aplicada às ciências sociais aplicadas II*. Salvador: Universidade Federal da Bahia, Faculdade de Ciências Contábeis; Superintendência de Educação a Distância, 2018. 112 p. Acesso em: 15 nov. 2024. ISBN 978-85-8292-162-3. Disponível em: <http://repositorio.ufba.br/ri/handle/ri/28125>. Citado 2 vezes nas páginas 18 e 26.

RIFO, L. *Probabilidade e estatística: aspectos de tomada de decisões e incerteza para o ensino fundamental e médio*. Edição especial. Rio de Janeiro: Sociedade Brasileira de Matemática, 2024. Citado 4 vezes nas páginas 148, 149, 150 e 152.

SEDU, S. D. E. D. E. D. E. S. *Catálogo Novo Ensino Médio*. S.d. Accessed: 17 jun. 2024. Disponível em: <https://novoensinomedio.sedu.es.gov.br/Media/NovoEnsinoMedio/Arquivos/CAT%C3%81LOGO%20NOVO%20ENSINO%20MEDIO.pdf.> Citado na página 23.

SEDU, S. D. E. D. E. D. E. S. *Itinerário Formativo*. S.d. Accessed: 17 jun. 2024. Disponível em: <https://novoensinomedio.sedu.es.gov.br/itinerario-formativo.> Citado 3 vezes nas páginas 21, 22 e 23.

SEDU, S. da Educação do Estado do E. S. *Currículo do Ensino Médio: Aprofundamento da Área de Matemática (Alterado em 15-09-23)*. 2023. Acesso em: 3 dez. 2024. Disponível em: <https://curriculo.sedu.es.gov.br/curriculo/wp-content/uploads/2023/09/Curriculo-EM_Aprofundamento-da-area_-Matematica_-Alterado_15-09-23.pdf>. Citado na página 24.

SEDU, S. de Estado da Educação do E. S. *Currículo - Orientações Curriculares*. 2024. <https://curriculo.sedu.es.gov.br/curriculo/orientacoescurriculares/>. Acesso em: 4 dez. 2024. Citado 5 vezes nas páginas 25, 161, 162, 166 e 167.

SILVA, M. F. T. da. *Médias, desigualdade das médias e a resolução de problemas*. Dissertação (Mestrado) — Universidade Federal do Espírito Santo, 2019. Acesso em: 22 maio 2025. Disponível em: <https://repositorio.ufes.br/handle/10/16848>. Citado na página 89.

SOUZA, R. P. de. *Os impactos do Programa de Mestrado Profissional em Matemática - PROFMAT/UFES na prática profissional de seus egressos*. Dissertação (Dissertação de Mestrado) — Universidade Federal do Espírito Santo, 2024. Programa de Mestrado Profissional em Matemática - PROFMAT/UFES. Citado na página 66.

TRIOLA, M. F. *Introdução à Estatística*. 12. ed. São Paulo: Pearson, 2017. Citado 4 vezes nas páginas 18, 26, 109 e 122.

UNICEF. *Strength in Numbers: How Longitudinal Research Can Support Child Development*. [S.l.]: UNICEF Office of Research – Innocenti, 2015. Citado na página 132.

ZAT, A. D. *Moda Estatística: Relações Conceituais*. 2022. <https://editora.pucrs.br/anais/erematsul/minicursos/modaestatistica.pdf>. Disponível em: PUCRS - Anais do EREMATSUL, Minicursos. Citado na página 72.

A Orientações Curriculares da Parte Flexível

Tabela 28 – Objetos de Conhecimento – Consumo Responsável e Educação Tributária

Série	Trimestre	Objeto de Conhecimento
2ª série	1º	Economia Doméstica: Dinheiro (valor, conceito, origem e evolução); As dimensões da saúde e sua relação com o dinheiro (saúde intelectual, saúde financeira, ...); Identificando o perfil do consumidor (compulsivo, consciente, ...).
	2º	Estilos de vida sustentáveis: Conceitos de consumo; Impactos socioambientais do consumo; Responsabilidades do cidadão consumidor; Proposta de mudança de padrão de consumo; Certificação e Rotulagem ambientais; Maquiagem Verde (Greenwashing); Introdução: Universo da Água (Usos da água; Dicas práticas para economia de água; Técnicas de reaproveitamento e armazenamento da água); Introdução: Energia é tudo (Fontes de energia; Dicas práticas para economia de energia).
	3º	Educação Tributária: O que é educação fiscal; Impostos, tributos e taxas; O controle do cidadão dos gastos públicos; Os principais impostos federais, estaduais e municipais; A isenção de impostos no brasil; Documentos fiscais; Sonegação fiscal.
3ª série	1º	Economia Doméstica: Etapas da mudança em busca de uma relação mais eficaz e vantajosa para a pessoa; Estágios da mudança; Pré-contemplação; A contemplação; A preparação; A ação; A manutenção; O planejamento financeiro; Elaborando o planejamento financeiro e definindo metas; Identificando os sonhos e se permitindo sonhar; Como economizar; Alfabetização e organização financeira; Cuidados com as armadilhas; Direitos do consumidor; Organizando a vida financeira; Cadastro negativo e cadastro positivo; A agenda financeira.
	2º	Estilos de vida sustentáveis: Resíduos sólidos no Brasil; Resíduos perigosos; Impactos dos resíduos e o que o consumo tem a ver com isso; Repensar, reduzir, recusar, reutilizar e reciclar (5Rs); Impactos da construção; Ecovilas, Permacultura e Bioconstrução; Sustentabilidade na sua construção; Gestão dos resíduos de construção; Cuidados ao adquirir um imóvel; Materiais de limpeza; Transporte Sustentável; Meios de transporte não motorizados.
	3º	Educação Tributária: Orçamentos Públicos; Porque tudo no Brasil é tão caro; Estado e tributação: do estado liberal à contemporaneidade brasileira; Tributação e fundamentos do estado republicano: cidadania e educação fiscal ancoradas no dever fundamental de pagar tributos; Tributação, igualdade e coesão social na américa latina; Tributação e direitos fundamentais: dos direitos sociais aos limites do poder de tributar; Tributo: definição e tipologia.

Fonte: Adaptado de (SEDU, 2024)

Tabela 29 – Objetos de Conhecimento – Matemática Financeira

Série	**Trimestre**	**Objeto de Conhecimento**
2ª série	1º	Situações cotidianas e o uso da Matemática Financeira; Capitalização Simples: conceito de juro, capital e taxa de juros; Capitalização Composta: montante e valor atual para pagamento único; Equivalência de taxas; Desconto simples.
	2º	Sistema de Amortização: Sistema Francês (Tabela Price); Sistema de Amortização Constante (SAC); Sistema de Amortização Misto (SAM); Investimentos de renda fixa e renda variável.
	3º	Operações Financeiras Realizadas no Mercado: Inflação e correção monetária, indexador; Aplicações financeiras com renda fixa; Desenvolvimento do projeto do estudante através da criação e/ou utilização de recursos digitais de aplicação dos conceitos.
3ª série	1º	Situações cotidianas e uso da Matemática Financeira: série de pagamentos; noção de fluxo de caixa; pagamentos vencidos e antecipados; equivalência de capitais e planos de pagamentos.
	2º	Método de Avaliação de Fluxo de Caixa (Valor presente líquido; Taxa interna de retorno;); Classificação de Taxas de Juros (Conceito e classificação das taxas de juros; Taxas equivalentes e proporcionais; Juros pagos antecipadamente); Taxa Média e Prazo Médio (média e prazo médio para operações de desconto simples; Taxa média e prazo médio para operações com juro simples; Taxa média e prazo médio para operações com juro composto.).
	3º	Operações Financeiras Realizadas no Mercado: Inflação e correção monetária, indexador; Aplicações financeiras com renda fixa; Desenvolvimento do projeto do estudante através da criação e/ou utilização de recursos digitais de aplicação dos conceitos.

Fonte: Adaptado de (SEDU, 2024)

Tabela 30 – Objetos de Conhecimento – Educação Financeira

Série	Trimestre	Objeto de Conhecimento
3ª série	1º	Vida Familiar e Cotidiana: Registrar despesas regularmente; Saber como se gasta o próprio dinheiro mensalmente; Estimar o valor das próprias despesas; Listar as despesas familiares; Classificar as despesas familiares em "fixas", "variáveis" e "eventuais ou extraordinárias"; Elaborar um orçamento mensal organizando as despesas de acordo com a classificação atribuída; Comparar orçamentos; Pesquisar taxas de juros e o CET de empréstimo pessoal para assalariado; Decidir entre tomar um empréstimo e utilizar dinheiro da poupança; Comparar o CET de empréstimos de diferentes instituições financeiras; Compreender que há comportamentos que nos levam a gastar mais dinheiro do que o previsto na hora de ir às compras; Distinguir os comportamentos positivos dos negativos na hora de ir às compras; Tomar decisões de compra diante de certos imprevistos; Categorizar despesas pessoais e familiares; Avaliar a importância das despesas no contexto familiar próprio; Identificar categorias cujas despesas podem ser reduzidas; Calcular o peso relativo das categorias de despesa; Elaborar planejamento de redução de despesas em 5%; Levantar situações em que o seguro pode fazer diferença; Compreender vocabulário específico de seguros; Identificar alternativas de prevenção coerentes com o próprio contexto familiar; Classificar as receitas da família em fixas e variáveis; Elaborar tabela com as receitas da família ao longo de vários meses; Analisar como a própria família gasta ou poupa o dinheiro extra de rendas sazonais. Vida Social: Identificar desperdícios nas próprias despesas; Evitar desperdícios; Dimensionar despesas utilizando estimativas; Orçar eventos sociais; Planejar eventos sociais ambientalmente responsáveis; Identificar armadilhas ao fazer estimativas; Explicar conceitos financeiros para outras pessoas; Relacionar conceitos de taxa de juros e de risco e retorno com situações cotidianas; Identificar os elementos de uma fatura de cartão de crédito; Identificar os comportamentos financeiros que provocam endividamento no cartão; Utilizar o cartão de crédito de forma consciente e responsável; Elaborar hipóteses e conclusões sobre a vida de pessoas a partir de seus dados financeiros; Levantar despesas envolvidas em um acampamento; Prever verba para imprevistos; Elaborar planejamento financeiro para acampar; Elaborar planejamento de festa junina nos moldes de um plano de negócio; Identificar armadilhas financeiras em anúncios de parcelamento; Analisar opções para sair de situação de endividamento no cartão de crédito. Bens Pessoais: Calcular a diferença entre valores à vista e a prazo; Buscar informações específicas no Código de Defesa do Consumidor; Tomar decisões financeiras considerando o custo de oportunidade; Equilibrar desejos e necessidades na escolha de um produto; Comparar preços; Calcular poupança necessária para comprar computador à vista; Calcular a diferença entre taxa de juros e taxa de empréstimo; Calcular o rendimento de uma poupança; Distinguir poupança de financiamento; Tomar decisão de poupança ou financiamento de acordo com as necessidades e possibilidades; Identificar os elementos dos textos publicitários voltados para despertar desejo de consumo; Identificar

		o conflito entre desejo e necessidade na situação de consumo; Utilizar o conhecimento sobre as principais armadilhas ligadas ao consumo, para se proteger do impulso irrefletido de consumir; Escolher o tipo de aparelho celular de acordo com as próprias necessidades; Escolher o plano de telefonia celular que atende às próprias necessidades; Compreender a fatura do celular; Identificar casos de práticas abusivas e de violação de direitos do consumidor; Redigir os possíveis encaminhamentos para um problema de consumo, inclusive descrevendo os direitos básicos do consumidor violados; Utilizar a taxa de câmbio para converter moedas estrangeiras em moeda nacional; Decifrar como é feita a cobrança em reais de compra realizada com cartão de crédito em outra moeda.
	2º	Trabalho: Identificar o tipo de trabalho que mais atrai; Analisar o trabalho que mais atrai em comparação com o seu projeto de vida; Elaborar o próprio currículo; Identificar habilidades compatíveis com cargos anunciados; Mostrar qualidades profissionais próprias de forma adequada em uma entrevista de emprego simulada; Alinhar o emprego desejado ao estilo de vida pretendido; Distinguir renda bruta e líquida; Explicar quais descontos são aplicados à renda bruta e como são calculados; Utilizar o conceito de desemprego estrutural em uma história de sua autoria; Levantar possibilidades de superação de uma situação de desemprego, considerando a tensão entre desejos e necessidades; Elaborar orçamento com base em dados e estimativas; Elaborar planejamento financeiro simulado visando a um resultado superavitário; Fazer provisões; Incluir situações futuras no planejamento financeiro atual; Elaborar um esboço de plano de aposentadoria, alinhando metas de longo prazo com meios para alcançá-las; Reconhecer o Seguro DPVAT e como ficar atento às práticas dos aproveitadores (Golpe do DPVAT); Elaborar mensagem e produto para campanha informativa junto à comunidade escolar sobre seguros; Utilizar vocabulário de seguros aplicado à campanha informativa. Empreendedorismo: Diferenciar “empreendedorismo por necessidade” de “empreendedorismo por oportunidade”; Relacionar características próprias com oportunidades de negócio; Identificar problemas da comunidade que possam gerar novos negócios; Utilizar uma versão simples de tempestade mental para gerar boas ideias de negócio; Distinguir “conhecimentos”, “habilidades”, “atitudes” e “competências” no contexto de empreendedorismo; Avaliar se possui os conhecimentos, habilidades e atitudes necessários para abrir um negócio; Avaliar se é preciso adquirir ou aprimorar as competências necessárias para abrir um determinado negócio; Identificar características comuns aos empreendedores; Diferenciar empreendedorismo de intraempreendedorismo; Verificar se tem perfil empreendedor; Identificar o público-alvo de um negócio; Criar marca e slogan para um produto ou serviço; Montar um plano de marketing; Realizar pesquisa de mercado para um produto ou serviço; Diferenciar tipos de recursos necessários para a abertura e operação de um negócio; Cotar custos básicos de abertura e operação de um negócio; Definir competências, conhecimentos, habilidades e atitudes do pessoal necessário para trabalhar em um negócio;

		Fazer projeções de vendas e lucro de um negócio; Dimensionar o lucro de um negócio; Descontar os tributos devidos no caso de produtos ou de serviços; Distinguir filantropia e responsabilidade socioambiental; Fazer um plano de responsabilidade socioambiental de um negócio; Juntar em um Plano de Negócios todos os dados colhidos nas etapas anteriores de trabalho nesse tema de empreendedorismo. Grandes Projetos: Harmonizar desejos e necessidades na escolha da casa própria; Buscar informações de preço e financiamento da casa própria; Decidir o valor da casa, da entrada e das prestações em função do orçamento familiar; Planejar-se financeiramente para pagar a entrada e as prestações da casa própria; Orçar uma festa; Adequar uma festa ideal à própria realidade financeira; Fazer provisões; Cortar despesas de acordo com as prioridades estabelecidas; Tomar decisões de investimento em uma situação de mercado simulada; Tomar uma decisão inicial de investimento considerando a situação atual própria e da família; Identificar direitos ou deveres que não estão sendo atendidos em determinadas situações; Elaborar argumentos para debater questões de direitos e deveres dos investidores; Reconhecer esquemas irregulares de captação de recursos, conhecidos como Pirâmides Financeiras; Conhecer os prós e contras de um consórcio; Tomar decisões de consumo e poupança em situação simulada de consórcio; Tomar decisões de lance e de manter ou desistir do consórcio em situação simulada; Decidir entre duas opções de aplicação, os considerando as taxas de juros; Explicar como se evita endividamentos; Levantar opções para sair de um endividamento de R$ 1.000,00; Estimar despesas fixas e variáveis para estudar em outra cidade; Calcular a receita mensal necessária para estudar em outra cidade; Elaborar um planejamento financeiro para estudar em outra cidade.
	3º	Bens Públicos: Buscar informações sobre bens e serviços públicos; Calcular, de forma simulada, quanto o poder público gasta para manter a escola em que o aluno estuda; Identificar as áreas de atuação da escola em relação a determinadas categorias; Indicar a categoria de atuação da escola que demanda prioridade de aplicação de recursos no orçamento escolar; Identificar os fatores responsáveis pelo alto custo ambiental do livro escolar; Calcular o consumo aproximado de papel na própria escola; Realizar várias ações para economizar papel; Elaborar e engajar-se em campanha para se economizar papel; Distinguir diversos tipos de tributos e suas funções; Relacionar nota fiscal com tributos e com oferta de bens e serviços públicos; Propagar o comportamento cidadão de exigência de nota fiscal de bens e serviços consumidos; Identificar semelhanças e diferenças entre orçamento público e orçamento familiar; Relacionar o dever de os cidadãos pagarem os tributos devidos com o dever de o governo oferecer serviços públicos; Levantar necessidades da comunidade em relação aos serviços públicos para servir de base para um ensaio de elaboração de orçamento público; Relacionar os efeitos da corrupção com a restrição de verbas públicas para provimento de serviços públicos; Saber buscar informação sobre as contas públicas; Tomar iniciativas de combate à corrupção; Distinguir diversos tipos de tributos e suas funções; Relacionar nota fiscal com tributos e com oferta de bens e serviços públicos; Propagar o comportamento cidadão de exigência de nota fiscal de bens e serviços consumidos. Economia do País: Elaborar esboço de projeto de atividade

		cultural ou esportiva; Alinhar objetivos de um projeto a objetivos da Lei Rouanet; Compreender trechos de leis com vocabulário de Educação Financeira; Identificar problemas que a inflação pode gerar quando a receita não acompanha o aumento dos preços; Realizar ajustes em um orçamento familiar em função da inflação; Explicar a outro jovem o que é inflação; Matemática na explicação de fenômenos de natureza científica, social, profissional, cultural, de processos tecnológicos, identificando os diversos pontos de vista e posicionando-se mediante argumentação, com o cuidado de citar as fontes dos recursos utilizados na pesquisa e buscando apresentar conclusões com o uso de diferentes mídias. Explicar o que são e como funcionam os órgãos supervisores do sistema financeiro; Estabelecer relações entre crescimento econômico e crescimento da renda individual; Planejar ações simuladas de organização financeira para famílias de baixa renda; Criar situações fictícias que ilustrem a lei da oferta e da demanda; Estimar receitas e despesas de uma pessoa aposentada; Elaborar um planejamento financeiro simulado para uma pessoa aposentada; Levantar preços para estimar o valor total das necessidades básicas de uma pessoa; Relacionar o valor das necessidades básicas de uma pessoa com o valor do salário mínimo. Economia do Mundo: Contextualizar a função do dinheiro na nossa sociedade; Conservar o dinheiro; Identificar diversas ciladas envolvidas em negociações; Compreender que os blocos econômicos se organizam por meio de acordos negociados; Compreender, por meio de vivência lúdica, algumas problemáticas referentes aos blocos econômicos; Aprender a pensar estratégias simuladas de resolução de questões mundiais, a partir de fatos que se apresentam na economia; Identificar os produtos importados utilizados no cotidiano; Localizar os países dos quais o Brasil importa os produtos utilizados no cotidiano; Buscar dados sobre as exportações nacionais e internacionais; Refletir sobre o perfil e a atuação dos representantes de um país em uma comunidade internacional; Propor apoio de uma instituição financeira internacional para um projeto local; Comparar o índice de IDH e o PIB per capita de diversos países; Relacionar o desenvolvimento econômico de um país aos impactos ambientais que causa; Buscar referenciais para a futura decisão profissional; Estabelecer relações entre a crise econômica de um país e situações da vida pessoal; Criar textos que divulguem impactos de uma crise econômica no cotidiano.

Fonte: Adaptado de (SEDU, 2024)

Tabela 31 – Objetos de Conhecimento – Projetos em Educação Financeira e Fiscal

Série	**Trimestre**	**Objeto de Conhecimento**
3ª série	1º	Raciocínio Lógico: Conjuntos numéricos fundamentais; Intervalos numéricos; Operações com conjuntos; Partição de um conjunto; Número de elementos da união de dois conjuntos; Símbolos utilizados na Lógica Matemática; O Modificador Negação; Operações lógicas; Resumo da Teoria; Relações binárias; Regra de três composta; Cuidado com a regra de três; O conjunto vazio; Fatorial; Princípio fundamental da contagem – PFC.
	2º	Análise Combinatória: Permutações simples; Permutações com elementos repetidos; Arranjos simples. Combinações simples.
	3º	Probabilidades: Conceito elementar de Probabilidade; Propriedades; Probabilidade condicional; Distribuição binomial de probabilidades; Utilização e criação de recursos digitais de aplicação dos conceitos estudados durante o aprofundamento.

Fonte: Adaptado de (SEDU, 2024)

B Formulário: Como a Matemática Influencia Suas Decisões?

Olá!
Estamos realizando uma pesquisa com estudantes do Ensino Médio para entender melhor suas experiências e opiniões sobre a matemática e a tomada de decisão. As respostas que você fornecer serão muito valiosas para melhorar nosso conteúdo pedagógico e ajudar a tornar as aulas mais relevantes e envolventes.

Este formulário é totalmente anônimo e suas respostas serão utilizadas apenas para fins educacionais. Não se preocupe; não há respostas certas ou erradas. O importante é que você compartilhe sua perspectiva honesta.

Agradecemos muito sua participação!

Informações Gerais

1. Qual é seu sexo?
   ☐ Feminino
   ☐ Masculino
   ☐ Prefiro não declarar

2. Qual a sua idade (anos): ____________________

3. Qual é a sua série/ano escolar?
   ☐ 1ª série do Ensino Médio
   ☐ 2ª série do Ensino Médio
   ☐ 3ª série do Ensino Médio

Sentimentos em relação à Matemática

4. Como você se sente em relação à matemática?
   ☐ Muito confortável
   ☐ Confortável
   ☐ Neutro(a)

☐ Desconfortável

☐ Muito desconfortável

5. Qual conteúdo de matemática você mais gosta?

   ☐ Geometria

   ☐ Álgebra

   ☐ Estatística

   ☐ Cálculos de aritmética

   ☐ Outros: ____________________

Estatística e Tomada de Decisão

6. Você acredita que entender estatística é importante para tomar decisões informadas?

   ☐ Sim

   ☐ Não

   ☐ Não sei opinar

7. Ao tomar decisões, você confia mais em:

   ☐ Dados estatísticos

   ☐ Relatos pessoais

   ☐ Ambos igualmente

   ☐ Nenhum dos dois

8. Você já utilizou dados estatísticos em decisões relacionadas a: (Marque todas que se aplicam)

   ☐ Saúde

   ☐ Finanças

   ☐ Estudos

   ☐ Jogos

   ☐ Resolução de problemas

   ☐ Outros: ____________________

9. Você já se arrependeu de uma decisão por não considerar dados estatísticos?

   ☐ Sim

   ☐ Não

Resolução de Problemas com Matemática

10. Você já resolveu algum problema pessoal ou escolar utilizando matemática?
   ☐ Sim
   ☐ Não

   Se sim, que tipo de cálculo ou modelo utilizou?

   ____

Processo de Tomada de Decisão

11. Você costuma analisar prós e contras antes de tomar uma decisão?
   ☐ Sempre
   ☐ Às vezes
   ☐ Raramente
   ☐ Nunca
12. Qual método você prefere para tomar decisões?
   ☐ Intuição
   ☐ Análise de dados
   ☐ Opinião de amigos/familiares
   ☐ Outros: __________________
13. Você costuma refletir sobre os resultados das suas decisões?
   ☐ Sim
   ☐ Não

Influências e Reflexões

14. Quais fatores mais influenciam suas decisões? (Marque todos que se aplicam)
   ☐ Emoções
   ☐ Influências externas (mídia, amigos, etc.)
   ☐ Informações e dados
   ☐ Outros: __________________
15. Você acredita que a matemática pode ajudar a combater a desinformação?
   ☐ Sim
   ☐ Não
   ☐ Não sei opinar

16. Você já tomou uma decisão importante sem ter todas as informações necessárias?
    ☐ Sim
    ☐ Não

Considerações Finais

17. Você gostaria de aprender mais sobre como usar a matemática para tomar decisões?
    ☐ Sim
    ☐ Não

18. Qual tema você gostaria de explorar mais em relação à matemática e tomada de decisão?
    ☐ Finanças pessoais
    ☐ Planejamento de carreira
    ☐ Resolução de conflitos
    ☐ Resolução de problemas
    ☐ Jogos e estratégias
    ☐ Outros: ____________________