



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I – CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE MATEMÁTICA
CURSO DE GRADUAÇÃO EM LICENCIATURA EM MATEMÁTICA**

DANILO RODRIGUES DOS SANTOS

**ALGUMAS REFLEXÕES SOBRE OS DADOS EDUCACIONAIS RELATIVOS AS
QUESTÕES DE ESTATÍSTICA NO ENEM 2023**

**CAMPINA GRANDE-PB
2024**

DANILO RODRIGUES DOS SANTOS

ALGUMAS REFLEXÕES SOBRE OS DADOS EDUCACIONAIS RELATIVOS AS
QUESTÕES DE ESTATÍSTICA NO ENEM 2023

Trabalho de Conclusão de Curso apresentado à
Coordenação do Curso de Matemática da
Universidade Estadual da Paraíba, como
requisito parcial à obtenção do título de
Licenciado em Matemática.

Orientador: Prof. Me. Josevandro Barros Nascimento

CAMPINA GRANDE-PB
2024

É expressamente proibida a comercialização deste documento, tanto em versão impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que, na reprodução, figure a identificação do autor, título, instituição e ano do trabalho.

S237a Santos, Danilo Rodrigues dos.

Algumas reflexões sobre os dados educacionais relativos
as questões de estatísticas no ENEM 2023 [manuscrito] /
Danilo Rodrigues dos Santos. - 2024.

58 f. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em
Matemática) - Universidade Estadual da Paraíba, Centro de
Ciências e Tecnologia, 2024.

"Orientação : Prof. Me. Josevandro Barros Nascimento,
Departamento de Matemática - CCT".

1. Mineração de dados educacionais. 2. Disciplina
estatística - ENEM 2023. 3. Análise estatística descritiva. I.
Título

21. ed. CDD 371.26

DANILO RODRIGUES DOS SANTOS

ALGUMAS REFLEXÕES SOBRE OS DADOS EDUCACIONAIS RELATIVOS AS
QUESTÕES DE ESTATÍSTICAS NO ENEM 2023

Trabalho de Conclusão de Curso
apresentado à Coordenação do Curso
de Matemática da Universidade
Estadual da Paraíba, como requisito
parcial à obtenção do título de
Licenciado em Matemática

Aprovada em: 22/11/2024.

BANCA EXAMINADORA

Documento assinado eletronicamente por:

- **Adailson Ribeiro da Silva** (***.475.294-**), em **26/11/2024 18:05:15** com chave **22319cb0ac3a11ef9bb306adb0a3afce**.
- **Emanuela Régia de Sousa Coelho** (***.622.214-**), em **26/11/2024 17:04:13** com chave **9bb48010ac3111efa9a41a7cc27eb1f9**.
- **Josevandro Barros Nascimento** (***.063.584-**), em **26/11/2024 16:52:29** com chave **f8385b24ac2f11efacf606adb0a3afce**.

Documento emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse https://suap.uepb.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Folha de Aprovação do Projeto Final

Data da Emissão: 31/03/2025

Código de Autenticação: 4904b7



Dedico este trabalho a Deus, ele que sempre foi minha força para viver e seguir em frente. A Ele devo todas as minhas realizações e conquistas, assim como a todos que se fizeram presentes em minha formação.

AGRADECIMENTOS

Gostaria de expressar minha imensa gratidão a Deus que me permitiu traçar este caminho, principalmente em meio às dificuldades durante esta graduação, dificuldades estas que vão além do físico e mental, mas que fazem parte do processo de realização deste trabalho de conclusão de curso.

Agradeço imensamente aos meus pais e irmãos que estiveram sempre ao meu lado, me incentivando desde o início da minha vida, com ênfase na vida acadêmica, de modo que eu nunca perdesse o rumo dos meus objetivos. Pois sempre confiaram que a educação abre as portas para um futuro melhor.

Também sou grato profundamente a minha namorada pelo amor incondicional e compreensão neste trajeto, onde sempre me colocava para cima nos momentos mais difíceis, assim como seus familiares.

Quero agradecer aos professores da banca, Prof. Me. Adailson Ribeiro da Silva Universidade Estadual da Paraíba (UEPB) e Prof.^a Dr.^a. Emanuela Régia de Sousa Coelho Universidade Estadual da Paraíba (UEPB), pelas contribuições relevantes para a finalização deste trabalho. Assim como ao Prof. Me. Josevandro Barros Nascimento, ao qual agradeço a sua orientação e apresentação ao mundo da análise de dados, onde pude me aventurar novamente na programação, além de cada palavra de apoio e sua imensa dedicação.

Também sou grato aos meus colegas e amigos, pois seu apoio e incentivo foram essenciais para me manter resiliente, agradeço a cada palavra de apoio e risadas, em especial nas pessoas de Elielson, Suzany, Geise e Maria Vitória.

Finalmente, agradeço a todos que puderam fazer parte deste capítulo da minha história, mesmo que não mencionadas neste espaço, todos têm uma importância única em que nada seria igual em sua ausência. Agradeço do fundo do meu coração.

“A matemática do tempo é simples. Você tem menos do que pensa e precisa mais do que acha”.

(Kevin Ashton)

RESUMO

O Exame Nacional do Ensino Médio (ENEM) é um processo seletivo ao qual dá acesso ao ensino superior o conteúdo de estatística presente na prova de matemática deste exame é fundamental na formação do indivíduo tendo em vista que esta ciência aproxima elementos abstratos com a realidade. A Mineração de Dados Educacionais (MDE) permite conhecer fatores que melhoram o projeto educacional, bem como prever o desempenho dos alunos e os elementos que influenciam o aprendizado. Neste sentido, esta pesquisa tem por objetivo apresentar a abordagem das questões de estatística do ENEM do ano de 2023, utilizando bases de dados fornecidas pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) e aplicando técnicas de mineração de dados com o propósito de melhor explicar as dificuldades dos alunos no contexto das questões do ensino de estatística. A pesquisa adota uma abordagem quantitativa, à luz da análise Estatística Descritiva e Inferencial. Após a análise dos resultados, destacam-se as dificuldades e fatores que as originam em relação às respostas dos alunos nas questões de estatística, proporcionando uma maior compreensão e reflexão sobre a prática pedagógica do professor de matemática em sala de aula.

Palavras-Chave: mineração de dados educacionais; estatística do ENEM; análise estatística descritiva.

ABSTRACT

The National High School Exam (ENEM) is a selection process that grants access to higher education. The statistics content present in the mathematics test of this exam is fundamental in the formation of the individual, considering that this science brings abstract elements closer to reality. Educational Data Mining (EDM) allows us to know factors that improve the educational project, as well as predict student performance and the elements that influence learning. In this sense, this research aims to present the approach to the statistics questions of the ENEM of the year 2023, using databases provided by INEP (National Institute of Studies and Educational Research Anísio Teixeira) and applying data mining techniques with the purpose of better explaining the difficulties of students in the context of the questions of teaching statistics. The research adopts a quantitative approach, in the light of Descriptive and Inferential Statistical analysis. After analyzing the results, the difficulties and factors that cause them in relation to the students' responses to the statistics questions are highlighted, providing a greater understanding and reflection on the pedagogical practice of the mathematics teacher in the classroom.

Keywords: educational data mining; ENEM statistics; descriptive statistical analysis.

SUMÁRIO

1	INTRODUÇÃO.....	10
1.1	JUSTIFICATIVA.....	11
1.2	QUESTÕES DE PESQUISA.....	11
1.3	OBJETIVOS.....	11
1.3.1	Objetivo Geral.....	11
1.3.2	Objetivos Específicos.....	11
2	REVISÃO BIBLIOGRÁFICA.....	13
2.1	MINERAÇÃO DE DADOS EDUCACIONAIS (MDE)	13
2.2	ÁREAS RELACIONADAS A EDM.....	14
2.3	ATIVIDADES DE MINERAÇÃO DE DADOS EM CONTEXTOS EDUCACIONAIS.....	15
2.4	APRENDIZAGEM NO ENSINO DE ESTATÍSTICA NO ENSINO MÉDIO NA EDUCAÇÃO BÁSICA.....	16
2.5	FATORES IMPORTANTES NA PREVISÃO DO DESEMPENHO DO ALUNO.....	17
2.6	LINGUAGEM R.....	18
3	METODOLOGIA PROPOSTA.....	19
3.1	CLASSIFICAÇÃO DA PESQUISA.....	19
3.2	PROCEDIMENTOS METODOLÓGICOS.....	19
3.2.1	Coleta de dados.....	20
3.2.2	Extração de recursos e limpeza de dados (Pré-Processamento e Transformação).....	21
3.2.3	Processamento analítico e algoritmos.....	21
4	RESULTADOS E DISCUSSÕES	22
4.1	ANÁLISE DAS QUESTÕES DO ENEM-2023.....	22
4.2	VISUALIZAÇÃO GRÁFICA DAS QUESTÕES DE MATEMÁTICA E SUAS TECNOLOGIAS, ESPECIFICAMENTE AS QUESTÕES QUE ENVOLVEM OS CONCEITOS DO ENSINO DE ESTATÍSTICA.....	31
4.2.1	Relação de escolhas dos participantes das questões 139, 159, 166, 173.....	31
4.2.2	Relação de escolhas dos participantes das questões 145, 162, 140, 172.....	33
4.2.3	Relação de escolhas dos participantes das questões 154, 171, 157, 159.....	35

4.2.4	Relação de escolhas dos participantes das questões 162, 176, 151, 154.....	37
4.2.5	Relação de escolhas dos participantes das questões 165, 167, 137, 151.....	39
4.2.6	Relação de escolhas dos participantes das questões 167, 163, 178, 147.....	40
4.2.7	Relação de escolhas dos participantes das questões 176, 139, 156, 143.....	42
4.2.8	Relação de escolhas dos participantes das questões 177, 169, 164, 138.....	44
5	CONSIDERAÇÕES FINAIS.....	46
	REFERÊNCIAS.....	47
	APÊNDICE A - GRÁFICOS REFERENTES A FIGURA 10.....	51
	APÊNDICE B - GRÁFICOS REFERENTES A FIGURA 11.....	52
	APÊNDICE C - GRÁFICOS REFERENTES A FIGURA 12.....	53
	APÊNDICE D - GRÁFICOS REFERENTES A FIGURA 13.....	54
	APÊNDICE E - GRÁFICOS REFERENTES A FIGURA 14.....	55
	APÊNDICE F - GRÁFICOS REFERENTES A FIGURA 15.....	56
	APÊNDICE G - GRÁFICOS REFERENTES A FIGURA 16.....	57
	APÊNDICE H - GRÁFICOS REFERENTES A FIGURA 17.....	58

1 INTRODUÇÃO

O Exame Nacional do Ensino Médio (ENEM) surgiu em 1998 e ao longo dos anos sofreu algumas modificações, seja no estilo da prova, seja na distribuição dos conteúdos abordados. A partir disto, o olhar sobre as questões de estatística exibem aspectos visíveis no ensino e aprendizagem em Matemática, de forma que tais conhecimentos são necessários para observar e analisar características do nosso mundo real. Ressalta-se a importância de uma formação inicial e continuada de professores de matemática, utilizando de ferramentas que facilitem o processo de ensino e acompanhamento do desenvolvimento de seus alunos.

O uso da mineração de dados educacionais permite ao professor analisar padrões e organizar os dados obtidos em categorias, prevendo o desempenho dos alunos e de fatores que influenciam o aprendizado, partindo das características dos alunos, proporcionando-lhes desenvolver soluções que melhorem o processo de avaliação contínua. Além disso, o intenso avanço tecnológico e digital exige que o professor faça parte deste processo, sendo necessário estar sempre atualizado com relação às novas tecnologias, as quais proporcionam habilidades e inovações em sua prática pedagógica.

Assim, o objetivo principal deste trabalho é analisar as questões de estatística abordados no ENEM-2023, observando quais conteúdos são mais abordados, dificuldades dos estudantes e explorar como a mineração de dados educacionais contribui para a avaliação contínua dos estudantes.

Para atingir os objetivos propostos nesta pesquisa, utilizamos uma metodologia quantitativa para investigar o desempenho dos estudantes nas questões analisadas e observar os conteúdos com maior índice de erros, destacando as possíveis causas da falta de êxito na resolução das questões.

Apesar dos benefícios da mineração de dados educacionais, no Brasil a introdução desse método na formação de professores apresenta diversos desafios. Questões relacionadas à falta de incentivo, infraestrutura tecnológica, formação docente adequada, falta de incentivo ao aprendizado de uma linguagem de programação, são alguns dos obstáculos enfrentados. Dessa maneira, demonstramos que a mineração de dados educacionais é um instrumento poderoso para capacitar educadores e ampliar seu repertório de habilidades e competências, o qual favorece consequentemente o desenvolvimento do ensino e aprendizagem.

1.1 JUSTIFICATIVA

As pesquisas sobre o uso da Mineração de Dados Educacionais (MDE) justificam-se pelos avanços no combate a problemas como evasão e reprovação escolar, além de diversos trabalhos desenvolvidos voltados para a tomada de decisão (Nascimento; Junior; Fagundes, 2018). O propósito da aplicação da MDE é identificar informações que contribuam para aprimorar a proposta educacional, melhorar as condições de infraestrutura escolar, otimizar o processo de ensino-aprendizagem e prever o desempenho dos alunos (Baker; Isotani; Carvalho, 2011).

A evolução e importância do ensino de estatística nos currículos escolares são oriundos da inclusão de seus conteúdos no componente de matemática, no entanto, muitos professores receberam conhecimentos insuficientes sobre esta ciência ao longo de sua formação, impedindo-os de trabalhar eficientemente os conteúdos de estatística (Quedi; Darroz, 2018).

1.2 QUESTÕES DE PESQUISA

Com base no que foi mencionado nas considerações, quais são as principais questões do ensino de estatística que os alunos das escolas públicas têm mais dificuldades na prova do ENEM do ano de 2023?

Assim apresentamos os objetivos, proposto nesta pesquisa.

1.3 OBJETIVOS

Apresento nesta seção os objetivos gerais e específicos da pesquisa.

1.3.1 Objetivo Geral

Analisar os padrões das questões do ensino de estatística ocorrido na avaliação de larga escala do Exame Nacional do Ensino Médio (ENEM) no ano de 2023.

1.3.2 Objetivos Específicos

- Analisar as questões de estatística do ENEM 2023 para identificar padrões de dificuldade e compreensão dos alunos;
- Investigar a distribuição das escolhas dos participantes nas questões do ENEM 2023;

- Identificando padrões de respostas corretas e incorretas para compreender as dificuldades dos alunos em estatística.
- Desenvolver visualizações gráficas das questões de Matemática e suas Tecnologias do ENEM 2023.

2 REVISÃO BIBLIOGRÁFICA

Nesta seção é apresentado a revisão de literatura referente a investigação desta pesquisa.

2.1 MINERAÇÃO DE DADOS EDUCACIONAIS (MDE)

A Mineração de Dados Educacionais (do inglês, “*Educational Data Mining*”, ou EDM), se faz necessária a partir da ampla difusão do uso de sistemas informatizados nas escolas e universidades, estes espaços geram uma enorme quantidade de dados que podem ser armazenados em bases de dados (Rodrigues; Gomes; Adeodato, 2017). A EDM surgiu pela primeira vez em uma série de workshops a iniciada em 2005, após estudos de Corbett e Anderson; já em 2006 foi publicado o artigo teórico sobre o potencial dos métodos em EDM de Zaiane e o primeiro livro de EDM por Romero e Ventura, os workshops se tornaram uma conferência anual, em 2008, e em seguida a Sociedade Internacional de Mineração de Dados Educacionais (Baker; Isotani; Carvalho, 2011).

A EDM pode ser vista não apenas como uma comunidade científica, mas também como uma área de investigação científica. No que tange à comunidade de pesquisa, a EDM pode ser considerada semelhante à análise de aprendizagem e como área de investigação científica, a EDM se preocupa com a análise de dados educacionais de grande escala, com interesse em métodos automatizados (Baker; Martin; Rossi, 2016). Com o crescimento do volume de dados, abre-se um leque de possibilidades para a comunidade científica, na qual muitos pesquisadores da área de Informática na Educação despertaram interesse nas técnicas de mineração de dados.

O processo de EDM consiste em transformar dados brutos de sistemas educacionais em informações úteis que influenciam a prática e a pesquisa educacional, além disso, este processo é semelhante ao processo de Data Mining, que consiste em pré-processamento, análise e pós-processamento. A EDM também possui algumas características como descoberta de informação automatizada, abordagem reducionista, origem no software educacional, na modelagem dos alunos, entre outros, adaptação automática, além de técnicas e métodos, afirmam (Moissa; Gasparini; Kemczinski, 2015).

A Mineração de Dados Educacionais ou EDM é definida de diversas maneiras. Segundo Romero e Ventura (2010), EDM é definida como “a aplicação de técnicas de mineração de dados (DM) para tipos específicos de conjuntos de dados provenientes de ambientes educacionais para abordar questões educacionais importantes”.

Segundo a Sociedade Internacional de Mineração de Dados Educacionais, a EDM pode ser definida da seguinte forma:

É uma disciplina emergente, preocupada com o desenvolvimento de métodos para explorar dados únicos e cada vez mais em larga escala, provenientes de contextos educacionais e usa esses métodos para entender melhor os alunos e as configurações em que aprendem (educationaldatamining.org, [s. d.]).

A EDM demanda de transformações das técnicas já existentes e o amadurecimento de novas técnicas. Romero e Ventura (2010) indicam sobre a diversidade da natureza dos dados analisados em EDM, os quais são mais diversos que a natureza observada nos dados utilizados em abordagens discretas de mineração de dados, de forma que esta diversidade auxilia a majoração da educação.

A integração entre pesquisadores da área da educação e da ciência de dados na Mineração de Dados Educacionais (MDE) é de extrema importância por vários motivos, um destes sendo a complementaridade de conhecimentos, que permite uma abordagem mais abrangente e eficaz na análise dos dados, aos quais os pesquisadores da área da educação possuem propriedade em metodologias de ensino e aprendizagem, enquanto os cientistas de dados têm habilidades em análise estatística, modelagem de dados e desenvolvimento de algoritmos. Melhoria na interpretação dos resultados sendo mais precisa e contextualizada. Desenvolvimento de soluções personalizadas, adaptadas às necessidades específicas do ambiente educacional. Além do avanço da pesquisa interdisciplinar, possibilitando a criação de novas técnicas, abordagens e metodologias na área de Mineração de Dados Educacionais (de Souza; de Souza, 2021).

Ademais é perceptível a evolução de alguns aspectos da Mineração de Dados Educacionais ao longo do tempo, como a adoção de grandes bases de dados na educação, criação de cursos *e-learning* a exemplo do MOOCs. Avanço das tecnologias computacionais, sendo este fundamental para a aplicação eficaz de técnicas de EDM permitindo analisar grandes volumes de dados. Melhorias nas técnicas e ferramentas de Mineração de Dados já existentes, além da aplicação e teste dessas técnicas (de Souza; de Souza, 2021).

2.2 ÁREAS RELACIONADAS A EDM

No que tange a área relacionada a Mineração de Dados Educacionais, esta por sua vez é uma área interdisciplinar, ou seja, está entrelaçada com outras áreas do conhecimento sendo estas como recuperação de informação, sistemas de recomendação, dentre outras. Romero e Ventura (2013) afirmam que EDM contempla a combinação de três áreas como Ciência da

Computação, Educação e Estatística. Dessa intersecção surge a relação com a Aprendizagem de Máquina e Aprendizagem Analítica (LA), oriundas da área de educação baseada em computador (*E-learning*).

A Aprendizagem Analítica (LA) é a área mais relacionada com EDM, devido a suas técnicas mais utilizadas como aplicação de inferência estatística, visualização de dados, análise de redes sociais (SNA), análise de sentimento, análise de influência e análise de discurso. Romero e Ventura (2010) apontam sobre os sistemas educacionais, enfatizando que estes possuem características distintas e que se fazem necessárias o desenvolvimento de novas abordagens de mineração de dados, além disso, um trabalho na área de EDM pode ser classificado em mais de uma abordagem educacional.

Os avanços observados nos modelos de aprendizagem de máquina (AM) permitiram obter uma alta taxa de precisão, no entanto, os modelos ficaram cada vez mais sofisticados e de difícil interpretação. Para algumas áreas como saúde e educação compreender e interpretar sistemas inteligentes é essencial (Neto; Vasconcelos; Zanchettin, 2021).

Ademais em EDM existem diversas linhas de pesquisa, sendo boa parte derivadas da área de mineração de dados. As principais subáreas de pesquisa em EDM são predição (*Prediction*), agrupamento (*Clustering*), mineração de relações (*Relationship Mining*), destilação de dados para facilitar decisões humanas (*Distillation of Data for Human Judgment*), descobertas com modelos (*Discovery with Models*) (Baker; Isotani; Carvalho, 2011).

2.3 ATIVIDADES DE MINERAÇÃO DE DADOS EM CONTEXTOS EDUCACIONAIS

Na área de EDM são utilizadas tarefas convencionais de mineração de dados, a exemplo de classificação, regressão, agrupamento e análise de associação. Peña-Ayala (2014) elaborou um estudo sobre a frequência das atividades mais comuns em EDM, dentre elas a classificação e agrupamento foram as mais citadas em trabalhos da área.

Além do mais, existem alguns processos importantes que permitem uma aplicação mais eficaz na Mineração de Dados Educacionais e para obter benefícios significativos no contexto educacional.

A **coleta de dados** corresponde à etapa inicial onde os dados educacionais são reunidos de diversas fontes majoritariamente oriundas do meio digital, como sistemas de gestão acadêmica, plataformas de ensino online, questionários, entre outros. Em seguida o **pré-processamento de dados**, onde os dados educacionais passam por um processo de limpeza, transformação e integração garantindo sua qualidade e consistência para análise. **Análise**

exploratória de dados, nesta fase são utilizadas técnicas estatísticas a fim de explorar padrões, tendências e relações nos dados educacionais. **Aplicação de técnicas de Mineração de Dados**, consiste na utilização de algoritmos de *Machine Learning*, Redes Neurais, entre outros para obter informações e executar previsões a partir dos dados educacionais. **Avaliação e interpretação dos resultados**, nesta etapa os dados são avaliados e interpretados considerando sua qualidade, adequação de técnicas utilizadas e o contexto educacional a fim de gerar insights que contribuam na tomada de decisão no ambiente educacional. Por fim, a implementação de ações e feedback, que se utilizam dos resultados obtidos da análise de dados educacionais no uso de estratégias para melhoria do ensino, aprendizagem e desempenho dos alunos, destacam (de Souza; de Souza, 2021).

2.4 APRENDIZAGEM NO ENSINO DE ESTATÍSTICA NO ENSINO MÉDIO NA EDUCAÇÃO BÁSICA

A importância da estatística na sociedade atual é uma realidade presente que abrange diversos aspectos, aos quais a produção e divulgação de dados não podem ser transmitidos baseados em intuições. O pensamento estatístico envolve a análise das informações considerando a variação característica dos dados e como essa variação pode auxiliar na tomada de decisões (Frei; Rosa; Biazi, 2023).

Frei, Rosa e Biazi (2023) afirmam ser comum que alguns indivíduos que recebem a informação em suma apenas a absorvem e culminam por inutilizá-la, seja por não possuírem ferramentas e noções básicas de estatística para que possam usufruir da informação de maneira útil em seu cotidiano.

No ensino e a aprendizagem da estatística na educação básica, essencialmente no ensino médio, é preciso levar em consideração os obstáculos que o diferem da Matemática, pois o pensamento estatístico possui um caráter que quebra com o paradigma lógico e determinista oriundo da matemática, além de que os problemas estatísticos costumam ser abertos, podendo existir mais de um método de solução correta (Lima *et al.*, 2022).

A compreensão eficaz dos dados se dá por meio do letramento estatístico, o qual refere-se à habilidade de compreender textos e as possíveis implicações das informações estatísticas no contexto, incluindo a familiaridade com sua terminologia, linguagem e conceitos dentro de um cenário social, além do desenvolvimento de atitudes investigativas e críticas (Lima; Giordano, 2021).

Sob este viés, Lima *et al* disserta acerca de letramento estatístico:

O letramento estatístico é formado por cinco componentes cognitivos a ressaltar: o próprio letramento, que envolve leitura de textos, gráficos, tabelas; conhecimentos estatísticos; conhecimentos matemáticos; conhecimentos do contexto; capacidade de elaboração de questões críticas (Lima *et al.*, 2022).

O letramento estatístico possui duas dimensões: a individual e a social, a dimensão individual abrange a capacidade de estabelecer conexões entre ideias e entre informações textuais e extratextuais. Já a dimensão social inclui as interações entre os participantes, as exigências dos contextos sociais e as representações e valores inerentes aos atos de leitura e escrita (Lima; Giordano, 2021).

Dessa forma, com o pensamento de tornar o estudante um indivíduo ativo na edificação do saber a BNCC (Brasil, 2018) é contrária às práticas tradicionais de ensino, a fim de formar cidadãos críticos e participante na vida social, pois a desinformação acaba sendo banalizada devido ao baixo nível de escolarização (Lima *et al.*, 2022).

2.5 FATORES IMPORTANTES NA PREVISÃO DO DESEMPENHO DO ALUNO

A educação gera uma imensa quantidade de dados onde podemos dizer que cada estudante é um pequeno universo, com isso se abre a possibilidade de utilizar ferramentas digitais e tecnologias digitais para melhorar o desempenho dos estudantes, incluindo inteligência artificial (IA), algoritmos de aprendizagem de máquina (AM). Assim, é possível prever comportamentos e características em relação ao desempenho dos estudantes o qual é chamado de previsão do desempenho acadêmico (PDA), sendo necessário para o acompanhamento do progresso dos alunos e adaptar o ensino a necessidades individuais, além de reconhecer eventuais dificuldades de aprendizagem o PDA fornece informações precisas para a escola, de forma que esta possa assumir práticas e abordagens de ensino cada vez mais eficazes (Benevento *et al.*, 2023).

O PDA é habitualmente aplicado nas instituições de ensino, avaliada por notas e frequências, bem como trabalhos extras a fim de proporcionar melhorias no processo educativo. É possível utilizar técnicas de mineração de dados educacionais, usando análises estatísticas e de AM no aprendizado de alunos, efetivando a oferta de educação (Benevento *et al.*, 2023).

2.6 LINGUAGEM R

Antes da criação da linguagem R, segundo Parga e Faria (2021) eram utilizadas duas linguagens que forneceram uma forte influência no desenvolvimento da linguagem R, sendo estas as linguagens S e Scheme, de forma que a linguagem R se assemelha à própria sintaxe da linguagem S. A linguagem R nasceu na década de 90, com o projeto de pesquisa de Ross Ihaka e Robert Gentleman, ambos estatísticos e pesquisadores associados ao departamento de estatística da Universidade de Auckland (Parga; Faria, 2021). Ademais, a linguagem R é uma linguagem interpretada, ou seja, ao enviar um comando escrito para o console do programa, ele irá avaliar os comandos enviados antes de executá-lo. O console corresponde a ferramenta mais importante do R, ao ser nele que se localiza o interpretador que avalia e executa os comandos realizados. A análise de dados por meio de uma linguagem de programação se torna uma maneira bastante eficiente de realizar esta tarefa, ao adquirir um aspecto interativo no R, de forma que se assemelha a uma conversa com o console (Parga; Faria, 2021).

Sobre algumas características da linguagem R, a Fundação R discorre:

R fornece uma ampla variedade de técnicas estatísticas (modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento, ...) e técnicas, e é altamente extensível. A linguagem S é muitas vezes o veículo de escolha para a pesquisa em metodologia estatística, e R fornece uma rota Open Source para a participação nessa atividade (R: O que é R?, [s. d.]).

Ainda, o R é dividido em duas partes, o sistema básico que consiste em toda composição de pacotes básicos da linguagem, sendo estes a base do R. E todos os pacotes externos ao sistema básico, desenvolvidos pela comunidade, o que é possível já que o R é gratuito e *open source*¹, o que permite que vários usuários desenvolvam novas funcionalidades e expandindo efetivamente o universo da linguagem R. Por fim, temos os Ambientes de Desenvolvimento Integrado (IDE), sendo o mais utilizado o *RStudio*. Ao qual oferece um ambiente com atalhos e ferramentas que facilitam o trabalho com o R (Parga; Faria, 2021).

¹ Open source corresponde a “código aberto”, utilizado em ferramentas que fazem uso de alguma linguagem de programação.

3 METODOLOGIA PROPOSTA

Nesta seção, apresenta-se o percurso metodológico seguido durante a pesquisa, para o alcance dos objetivos propostos e para trazer os resultados relacionados à análise de dados.

3.1 CLASSIFICAÇÃO DA PESQUISA

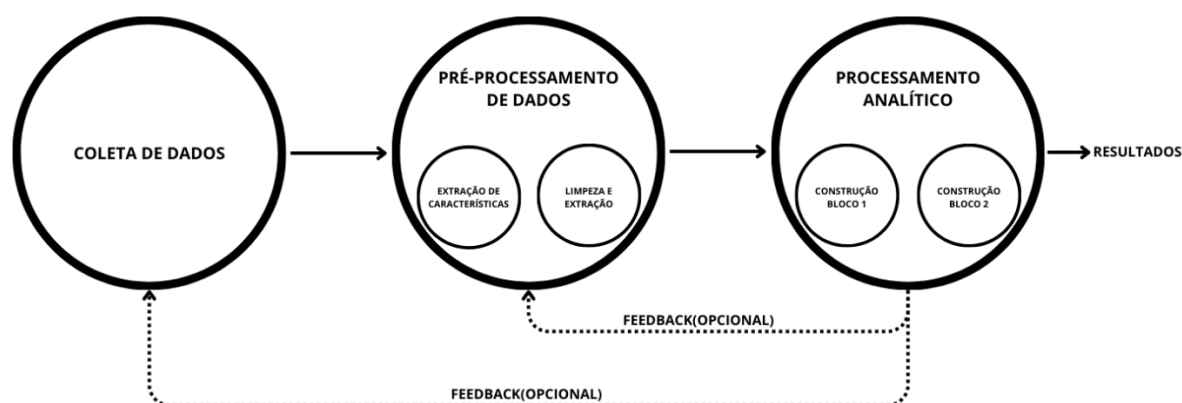
Uma investigação acadêmica científica, existem várias maneiras e formas de ser classificados, onde podemos citar como: natureza, abordagem e por seus objetivos propostos (Fontelles et al., 2009).

Neste sentido, conforme (Borgman, 2015; National Science Board, 2005), classifica-se, pesquisa quanto a sua natureza em uma pesquisa observacional, em que os dados científicos são gerados de origem observacional, como, por exemplo, computacional, experimental e de registro. Quanto à abordagem metodológica da pesquisa enquadra-se “A pesquisa quantitativa pretende e permite a determinação de indicadores e tendências presentes na realidade, ou seja, dados representativos e objetivos, opondo-se à ciência aristotélica, com a desconfiança sistemática das evidências e experiência imediata.” (MUSSI *et al.*, 2019). As pesquisas quantitativas, que se relacionam com as pesquisas de valores numéricos.

3.2 PROCEDIMENTOS METODOLÓGICOS

No que se refere ao procedimento metodológico desta pesquisa tem como base o processo da aplicação da MDE, que é similar ao da MD. Tomando como base a proposta de Aggarwal (2015) que propõe 3 etapas para o fluxo do trabalho de uma análise de Mineração de Dados conforme a figura 1.

Figura 1 – Processo de Data Mining proposto por Aggarwal (2015)



Fonte: Adaptado de Aggarwal (2024).

Conforme, Aggarwal (2015) às etapas para a utilização da MD são compostas de 3 etapas: 1) Coleta de dados; 2) Extração de recursos e limpeza de dados (Pré-Processamento e Transformação) –para tornar os dados adequados para processamento; 3) Processamento analítico e algoritmos. A sequência das etapas aplicadas nesta pesquisa é apresentar subseções a seguir.

3.2.1 Coleta de dados

Para Aggarwal (p.3 2015) classifica a “coleta de dados pode exigir o uso de *hardware* especializado, como uma rede de sensores, trabalho manual, como a coleta de pesquisas de usuários, ou ferramentas de *software*, como um motor de rastreamento de documentos da *web* para coletar documentos”, assim a coleta de dados nesta pesquisa se deu na base do Plano de Dados Abertos (PDA) é a principais bases de dados do INEP com ênfase aos dados do ENEM -2023 de maneira que dados podem ser classificados como abertos quando têm acesso, utilização, modificação e distribuição dissoluta para qualquer finalidade, com a condição de que o autor seja mencionado e que os dados sejam compartilhados sob a mesma licença. A disponibilização de dados tem tomado relevância como uma ferramenta que usa livremente dados de domínios distintos a fim de agregar valor para diferentes áreas, no entanto, a exploração de dados abertos ainda se depara com barreiras políticas organizacionais, legais, técnicas e financeiras (Silva, 2022). Após a etapa da coleta dos dados são feitas pré-processamento dos dados.

3.2.2 Extração de recursos e limpeza de dados (Pré-Processamento e Transformação)

Em posse da base de dados do ENEM-2023, ao serem dados brutos é necessário fazer a limpeza adequada, pois esses dados são feitos o *download* e não estão no formato adequado para o processamento. Neste sentido, é utilizando o *software* R como sendo, “R é um *software* estatístico que oferece um ambiente para análise interativa de dados, e que conta com uma poderosa linguagem de programação.” (Parga e Faria, 2021, p.11). Fornecendo recursos para o processamento analítico e algoritmos.

3.2.3 Processamento analítico e algoritmos

Quanto à etapa do processamento analítico e algoritmo, “A parte final do processo de mineração é projetar métodos analíticos eficazes a partir dos dados processados.” Aggarwal (p.3 2015), existem várias técnicas que podem ser utilizadas para aplicações e análises dos dados, podendo ser citada como: a Estatística Descritiva e Inferencial, algoritmos de Aprendizagem de Máquina (*Machine Learning*) e as redes neurais profundas associadas à Aprendizagem Profunda (*Deep Learning*). Diante desta investigação, esse estudo parte utilizar por meio da estatística descritiva para termos os resultados e discussões eficazes quanto aos objetivos propostos.

4 RESULTADOS E DISCUSSÕES

Nesta sessão faremos uma análise das questões de estatística abordadas na prova do 2º dia do Enem 2023, especificamente as questões que envolvem os conteúdos de estatística, além de discutir sobre a abordagem das questões na prova, bem como conceitos abordados nas questões.

4.1 ANÁLISE DAS QUESTÕES DO ENEM-2023

O ENEM, foi criado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) em 1998, com a perspectiva de avaliar o rendimento da Educação Básica brasileira (Brasil, 2016). Neste sentido, atualmente o ENEM é aplicado em dois domingos consecutivos desde o ano de 2017, sendo composto por quatro blocos de questões. No primeiro domingo, são abordadas as áreas de Ciências Humanas e suas Tecnologias (História, Geografia, Filosofia e Sociologia) e Linguagens, Códigos e suas Tecnologias (Língua Portuguesa, Literatura, Língua Estrangeira, Artes, Educação Física e Tecnologias da Informação e Comunicação). Nesse dia, os estudantes responderam a 45 questões de cada área, além de uma redação e cinco questões de inglês ou espanhol, conforme a opção de cada candidato.

Já no segundo domingo de aplicação do exame, é constituído pelas seguintes áreas: Ciências da Natureza e suas Tecnologias (Química, Física e Biologia) e Matemática e suas Tecnologias, também com 45 questões para cada área (Silva; Morino; Sato, 2014). Aqui, ao analisarmos a prova de Matemática e suas Tecnologias aplicado em 2023, especificamente no que se refere as 8 questões de estatística que foram encontradas.

No quadro 1, a seguir, apresentamos a distribuição dos conhecimentos de Estatística das 8 questões por cor da prova.

Quadro 1 – Distribuição dos Conhecimentos de Estatística no ENEM 2023

Ano	Cor do Caderno	Nº da Questão no caderno	Conteúdo abordado
2023	Amarela	139	Medidas de Tendência Central
2023	Amarela	145	Média, Moda e Mediana
2023	Amarela	154	Médias
2023	Amarela	162	Leitura de Gráficos e Tabelas

2023	Amarela	165	Médias
2023	Amarela	167	Leitura de Gráficos e Tabelas
2023	Amarela	176	Leitura de Gráficos e Tabelas
2023	Amarela	177	Leitura de Gráficos e Tabelas
2023	Azul	137	Médias
2023	Azul	140	Média, Moda e Mediana
2023	Azul	151	Leitura de Gráficos e Tabelas
2023	Azul	157	Médias
2023	Azul	156	Leitura de Gráficos e Tabelas
2023	Azul	164	Leitura de Gráficos e Tabelas
2023	Azul	166	Medidas de Tendência Central
2023	Azul	178	Leitura de Gráficos e Tabelas
2023	Rosa	138	Leitura de Gráficos e Tabelas
2023	Rosa	143	Leitura de Gráficos e Tabelas
2023	Rosa	147	Leitura de Gráficos e Tabelas
2023	Rosa	151	Médias
2023	Rosa	154	Medidas de Tendência Central
2023	Rosa	159	Média, Moda e Mediana
2023	Rosa	172	Médias
2023	Rosa	173	Leitura de Gráficos e Tabelas
2023	Cinza	139	Leitura de Gráficos e Tabelas
2023	Cinza	159	Medidas de Tendência Central

2023	Cinza	162	Média, Moda e Mediana
2023	Cinza	163	Leitura de Gráficos e Tabelas
2023	Cinza	167	Médias
2023	Cinza	169	Leitura de Gráficos e Tabelas
2023	Cinza	171	Médias
2023	Cinza	176	Leitura de Gráficos e Tabelas

Fonte: Dados da Pesquisa do autor (2024)

O quadro 1, trata da relação de distribuição dos conteúdos por questão em cada caderno de provas e suas respectivas cores, sendo estas as questões analisadas neste trabalho.

Apresentamos na figura 2 a seguir, a questão 139 do ano 2023 do caderno amarelo e sua resolução.

Figura 2 – Questão 139 prova do Enem 2023

QUESTÃO 139

Uma pessoa pratica quatro atividades físicas — caminhar, correr, andar de bicicleta e jogar futebol — como parte de seu programa de emagrecimento. Essas atividades são praticadas semanalmente de acordo com o quadro, que apresenta o número de horas diárias por atividade.

Dias da semana	Caminhar	Correr	Andar de bicicleta	Jogar futebol
Segunda-feira	1,0	0,5	0,0	2,0
Terça-feira	0,5	1,0	0,5	1,0
Quarta-feira	0,0	1,5	1,0	0,5
Quinta-feira	0,0	2,0	0,0	0,0
Sexta-feira	0,0	0,5	0,0	2,5

Ela deseja comemorar seu aniversário e escolhe o dia da semana em que o gasto calórico com as atividades físicas praticadas for o maior. Para tanto, considera que os valores dos gastos calóricos das atividades por hora (cal/h) são os seguintes:

Atividade física	Caminhar	Correr	Andar de bicicleta	Jogar futebol
Gasto calórico (cal/h)	248	764	356	492

O dia da semana em que será comemorado o aniversário é

A segunda-feira.
B terça-feira.
C quarta-feira.
D quinta-feira.
E sexta-feira.

Fonte: Inep (2023)

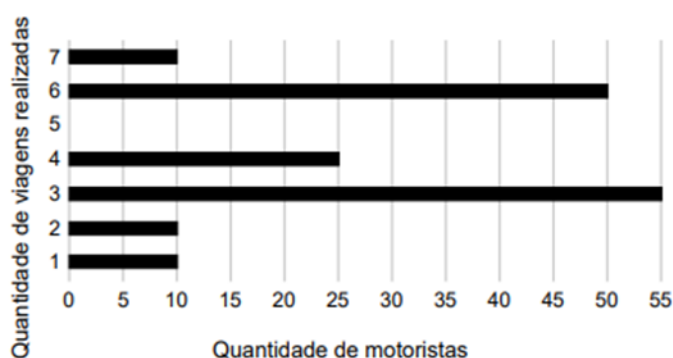
A questão 139 aborda conceitos de análise gráfica e tabelas, exigindo a leitura e interpretação dos dados apresentados. O enunciado descreve uma situação em que uma pessoa pratica quatro atividades físicas visando emagrecer, distribuídas em horas diárias ao longo da semana. Com o desejo de comemorar seu aniversário, o indivíduo escolhe o dia em que seu gasto calórico foi maior. Para isso, é necessário analisar as informações nas tabelas, que

mostram a quantidade de horas diárias dedicadas a cada atividade e seus respectivos gastos calóricos. Finalmente, para identificar o dia com o maior gasto calórico, basta somar os gastos calóricos diários; o dia em que esse gasto foi mais elevado é a quarta-feira. Já na figura 3, a questão 145 apresenta conceitos de média, moda e mediana a partir dos dados expostos.

Figura 3 – Questão 145 prova do Enem 2023

QUESTÃO 145

Uma empresa de transporte faz regularmente um levantamento do número de viagens realizadas durante o dia por todos os 160 motoristas cadastrados em seu aplicativo. Em um certo dia, foi gerado um relatório, por meio de um gráfico de barras, no qual se relacionaram a quantidade de motoristas com a quantidade de viagens realizadas até aquele instante do dia.



Comparando os valores da média, da mediana e da moda da distribuição das quantidades de viagens realizadas pelos motoristas cadastrados nessa empresa, obtém-se

- ☐ A mediana = média < moda.
- ☐ B mediana = moda < média.
- ☐ C mediana < média < moda.
- ☐ D moda < média < mediana.
- ☐ E moda < mediana < média.

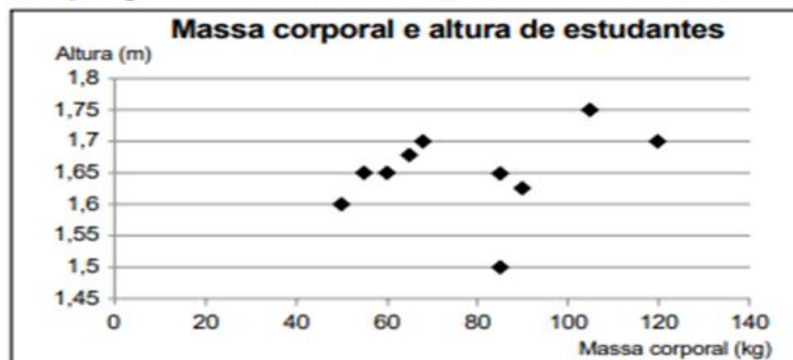
Fonte: Inep (2023)

A questão trata de um levantamento do número de viagens realizadas pelos motoristas de uma empresa de transporte. A moda, que corresponde ao número de viagens de maior frequência, é 3, conforme observado no gráfico. A mediana, o qual é o valor central, é obtida pela média entre os valores das posições 80 e 81, já que foram realizadas 160 viagens; nesse caso, a mediana é igual a 4. A média é calculada pela razão entre a quantidade total de viagens e a quantidade de motoristas, sendo seu valor igual a 4,15. Por fim, ao comparar os valores da média, da mediana e da moda, observa-se que a distribuição das quantidades de viagens realizadas pelos motoristas cadastrados na empresa é: moda < mediana < média. A questão 154, figura 4, trata do conceito de média, apresentando uma situação em que o professor propôs uma atividade para verificar o percentual de estudantes com massa corporal abaixo da média e altura acima da média em um grupo de alunos.

Figura 4 – Questão 154 prova do Enem 2023

QUESTÃO 154

Um professor, para promover a aprendizagem dos estudantes em estatística, propôs uma atividade. O objetivo era verificar o percentual de estudantes com massa corporal abaixo da média e altura acima da média de um grupo de estudantes. Para isso, usando uma balança e uma fita métrica, avaliou uma amostra de dez estudantes, anotando as medidas observadas. O gráfico apresenta a massa corporal, em quilograma, e a altura, em metro, obtidas na atividade.



Após a coleta dos dados, os estudantes calcularam a média dos valores obtidos, referentes à massa corporal e à altura, obtendo, respectivamente, 80 kg e 1,65 m.

Qual é o percentual de estudantes dessa amostra com massa corporal abaixo da média e altura acima da média?

- A** 10
- B** 20
- C** 30
- D** 50
- E** 70

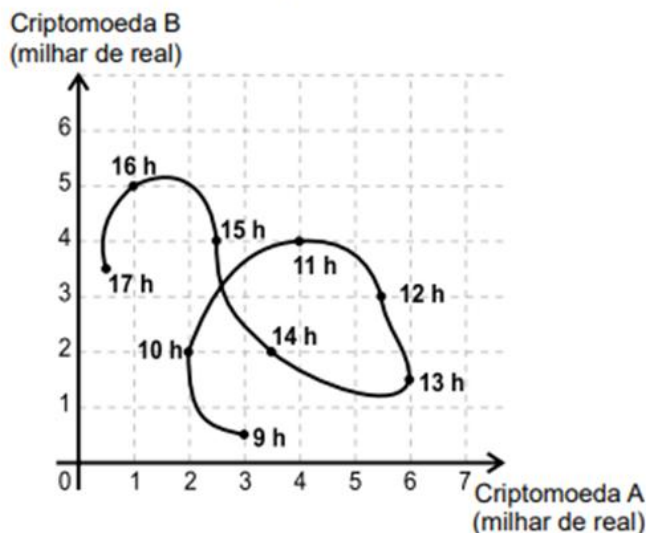
Fonte: Inep (2023)

O gráfico da questão apresenta esses dados para um grupo de 10 alunos, cada um com uma massa corporal e uma altura específica. Analisando o gráfico com base nas médias apresentadas, é possível dividi-lo em quadrantes para facilitar a visualização e calcular o percentual exigido nas condições dadas. Dessa forma, observa-se que dois estudantes se enquadram nessas condições, correspondendo a 20% do total de estudantes analisados na amostra. A questão 162 na figura 5, envolve uma análise gráfica na qual um investidor aplicou em duas criptomoedas, e o gráfico exibe a evolução de cada uma delas.

Figura 5 – Questão 162 prova do Enem 2023

QUESTÃO 162

Um investidor iniciante observou o gráfico que apresenta a evolução dos valores de duas criptomoedas A e B em relação ao tempo.



Durante horas consecutivas, esses valores foram observados em nove instantes, representados por horas exatas.

Em quantos desses instantes a criptomoeda A estava mais valorizada do que a criptomoeda B?

- ☐ A 3
- ☐ B 4
- ☐ C 6
- ☐ D 7
- ☐ E 9

Fonte: Inep (2023)

A partir do gráfico, é possível observar que há diferentes momentos em que a moeda A está mais valorizada que a moeda B, e vice-versa, representados em um plano cartesiano. Ao traçar uma diagonal que passa pela origem do plano cartesiano, é possível perceber que a moeda A está mais valorizada que a moeda B em quatro ocasiões, nos instantes 9h; 12h; 13h; 14h. A questão 165, figura 6 retrata a distribuição dos funcionários de acordo com dois setores, em que 75 funcionários estão no setor de produção e 25 funcionários estão no setor de administração.

Figura 6 – Questão 165 prova do Enem 2023

QUESTÃO 165

Os 100 funcionários de uma empresa estão distribuídos em dois setores: Produção e Administração. Os funcionários de um mesmo setor recebem salários com valores iguais. O quadro apresenta a quantidade de funcionários por setor e seus respectivos salários.

Setor	Quantidade de funcionários	Salário (em real)
Produção	75	2 000,00
Administração	25	7 000,00

A média dos salários dos 100 funcionários dessa empresa, em real, é

- A** 2 000,00.
- B** 2 500,00.
- C** 3 250,00.
- D** 4 500,00.
- E** 9 000,00.

Fonte: Inep (2023)

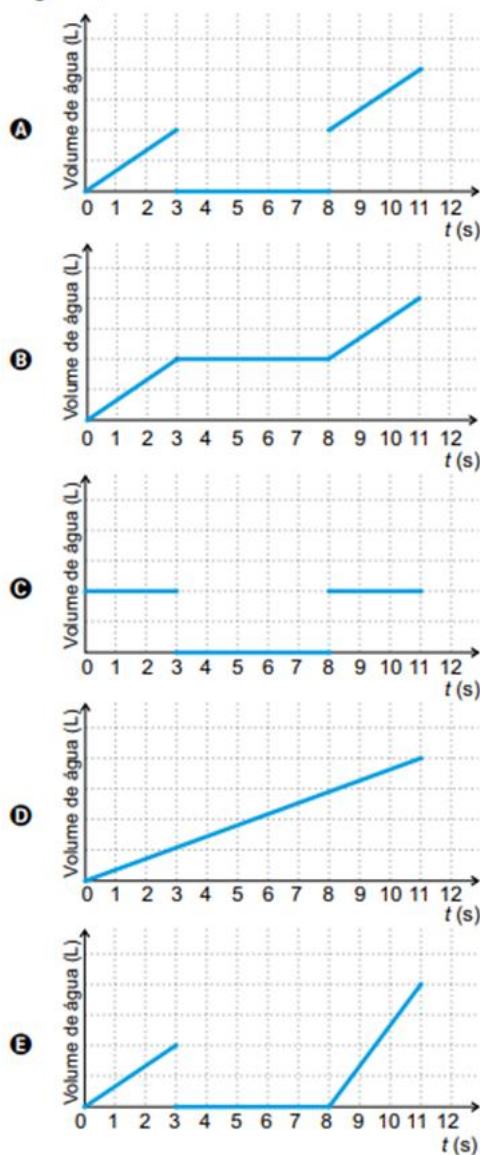
É possível obter o resultado da questão a partir da utilização do cálculo da média ponderada, a partir da razão entre o valor total em reais do salário de todos os funcionários pela quantidade de funcionários, sendo esta igual a R\$ 3250,00 reais. Na questão 167, são abordados conceitos de análise de gráficos, razão e proporção, e estatística.

Figura 7 – Questão 167 prova do Enem 2023

QUESTÃO 167

Estudantes trabalhando com robótica criaram uma “torneira inteligente” que automatiza sua abertura e seu fechamento durante a limpeza das mãos. A tecnologia funciona da seguinte forma: ao se colocarem as mãos sob a torneira, ela libera água durante 3 segundos para que a pessoa possa molhá-las. Em seguida, interrompe o fornecimento de água por 5 segundos, enquanto a pessoa ensaboa suas mãos, e finaliza o ciclo liberando água para o enxágue por mais 3 segundos. Considere o tempo (t), em segundo, contado a partir do instante em que se inicia o ciclo. A vazão de água nessa torneira é constante.

Um esboço de gráfico que descreve o volume de água acumulado, em litro, liberado por essa torneira durante um ciclo de lavagem das mãos, em função do tempo (t), em segundo, é



Fonte: Inep (2023)

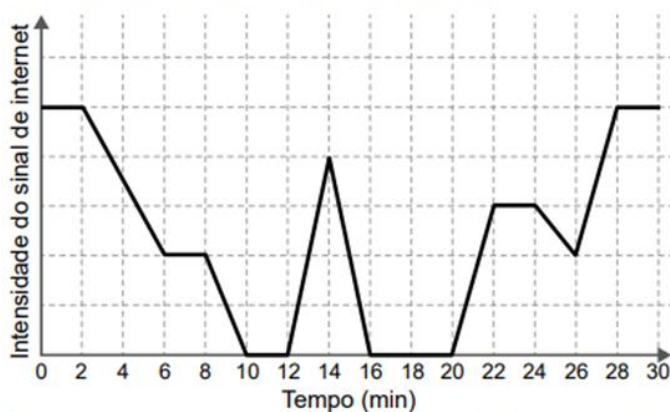
A questão trata da vazão de uma torneira inteligente, onde, em determinados momentos, a torneira está aberta, em outros, o fornecimento de água é interrompido, e depois é reaberta.

Quando a água está fluindo, o gráfico mostra um crescimento; quando a vazão é interrompida, o gráfico se comporta como uma constante. O gráfico que representa exatamente essa análise é o da alternativa B. A questão 176 envolve a leitura e interpretação de um gráfico que retrata a situação de uma pessoa usando um aplicativo de celular para monitorar a variação da intensidade do sinal de internet enquanto caminha.

Figura 8 – Questão 176 prova do Enem 2023

QUESTÃO 176

Uma pessoa caminha por 30 minutos e utiliza um aplicativo instalado em seu celular para monitorar a variação da intensidade do sinal de internet recebido pelo aparelho durante o deslocamento. Chegando ao seu destino, o aplicativo forneceu este gráfico:



Por quantos minutos, durante essa caminhada, o celular dessa pessoa ficou sem receber sinal de internet?

- A** 6
- B** 8
- C** 10
- D** 14
- E** 24

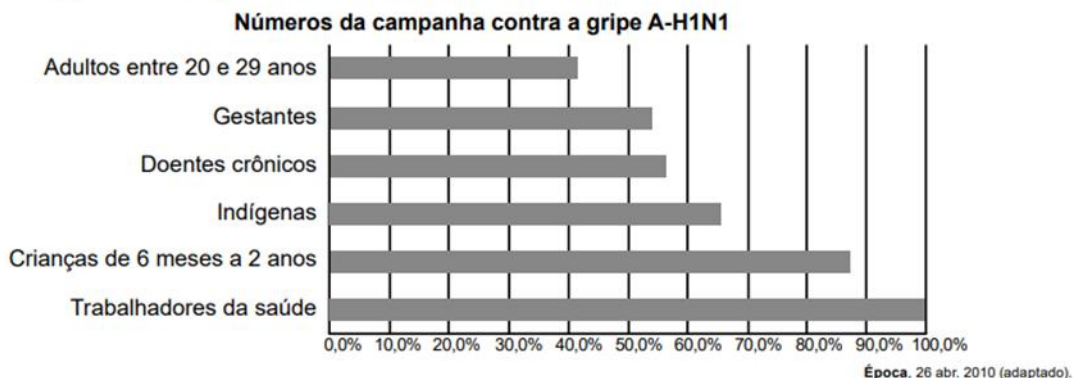
Fonte: Inep (2023)

A questão pede que se analise no gráfico quanto tempo, em minutos, o celular ficou sem receber sinal de internet durante a caminhada. Conforme os dados apresentados, o eixo x representa o tempo gasto na atividade em minutos, e o eixo y corresponde à intensidade do sinal de internet. Para atingir o objetivo, basta observar os pontos onde o sinal de internet é igual a zero e, em seguida, verificar quanto tempo se passou até que houvesse uma nova variação do sinal nesses pontos. Assim, o total de tempo em que o celular ficou sem receber sinal de internet foi de 6 minutos. A questão 177 apresenta um gráfico de colunas sobre a imunização contra a gripe A-H1N1, mostrando o percentual de pessoas imunizadas em diferentes categorias.

Figura 9 – Questão 177 prova do Enem 2023

QUESTÃO 177

O gráfico expõe alguns números da gripe A-H1N1. Entre as categorias que estão em processo de imunização, uma já está completamente imunizada, a dos trabalhadores da saúde.



De acordo com o gráfico, entre as demais categorias, a que está mais exposta ao vírus da gripe A-H1N1 é a categoria de

- A** indígenas.
- B** gestantes.
- C** doentes crônicos.
- D** adultos entre 20 e 29 anos.
- E** crianças de 6 meses a 2 anos.

Fonte: Inep (2023)

O ponto central da questão é identificar, a partir da análise do gráfico, qual dos grupos mencionados está mais exposto ao vírus da gripe A-H1N1. Ao observar o gráfico, percebe-se que esse grupo corresponde aos adultos entre 20 e 29 anos, ao apresentar o menor percentual de imunização em comparação aos demais grupos.

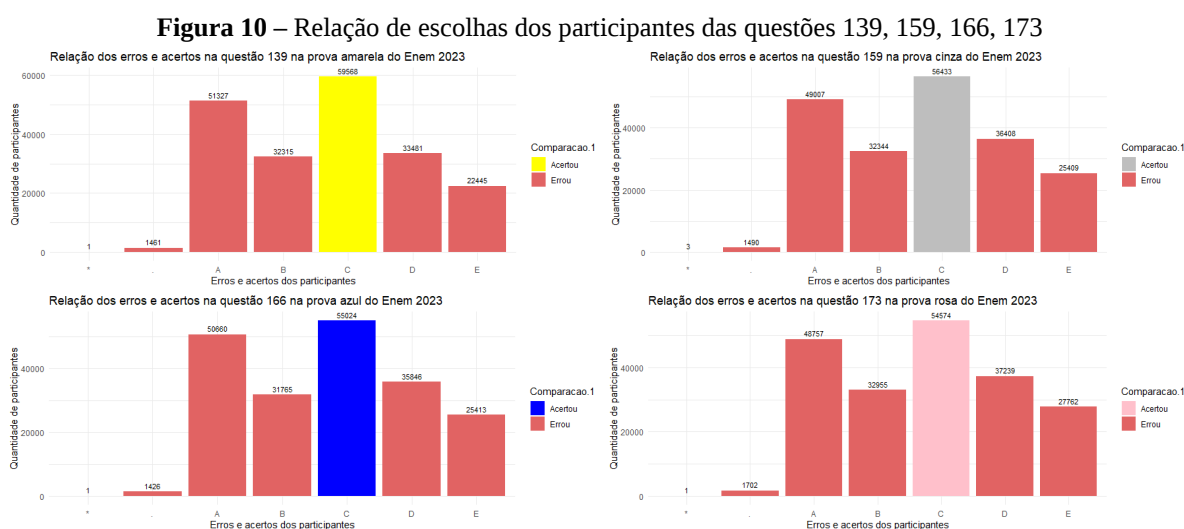
4.2 VISUALIZAÇÃO GRÁFICA DAS QUESTÕES DE MATEMÁTICA E SUAS TECNOLOGIAS, ESPECIFICAMENTE AS QUESTÕES QUE ENVOLVEM OS CONCEITOS DO ENSINO DE ESTATÍSTICA

Após o destaque das questões foram desenvolvidos gráficos para exemplificar as escolhas dos participantes, bem como apresentar os resultados da análise estatística das escolhas das alternativas.

4.2.1 Relação de escolhas dos participantes das questões 139, 159, 166, 173

Com base na questão 139 do caderno amarelo, observamos que as medidas de tendência central vêm sendo abordadas em questões que geralmente usam situações cotidianas, a fim de extrair alguma informação a partir do contexto apresentado. As Orientações Curriculares para o Ensino Médio, enfatizam que, no ensino médio, é importante proporcionar aos alunos a

compreensão tanto intuitiva quanto formal das principais ideias matemáticas presentes em representações estatísticas, procedimentos ou conceitos (Brasil, 2009). A partir da análise da questão, a fim de exemplificar o desempenho dos alunos apresentamos a figura 10 a seguir ² a relação de escolhas nas alternativas das questões 139, 159, 166, 173, dos cadernos amarelo, cinza, azul e rosa respectivamente. Os gráficos em questão trazem as alternativas escolhidas pelos participantes, distribuídas nas alternativas de A à E, além do símbolo “*” (asterisco) que corresponde a uma dupla marcação no gabarito e “.” (ponto) para questão em branco.



Fonte: Dados da Pesquisa (2024)

No caderno amarelo, apenas um participante marcou * ou seja, fez uma dupla marcação. Já 1461 participantes não marcaram nenhuma alternativa, correspondendo a 0,7% das escolhas na questão. Na alternativa A, 51327 respostas que equivalem a 25,6%. Na alternativa B, 32315 escolhas ou 16,1%. Na alternativa C, 59568 ou 29,7%. 33481 dos participantes marcaram a alternativa D, o que representa 16,7%. Por fim, 22445 participantes escolheram a alternativa E, ou seja, 11,2% do total de escolhas nesta questão. A partir da análise do gráfico correspondente ao caderno amarelo, na questão 139, cerca de 70,3% dos participantes erraram a questão, enquanto 29,7% acertaram.

No caderno Cinza, três estudantes marcaram *, ou seja, fizeram uma dupla marcação. 1490 participantes não marcaram nenhuma alternativa, sendo 0,7% das respostas na questão. Na alternativa A, 49007 escolhas, o que equivale a 24,4% das respostas. Na alternativa B, 32344

² Figura 10 encontra-se em tamanho maior no apêndice deste trabalho.

respostas, ou 16,1%. Na alternativa C, 56433 escolhas, o que corresponde a 28,1% das respostas na questão. Alternativa D, 36408 ou 18,1%. E por fim, alternativa E com 25409 representando 12,6% das respostas. Com base na análise do gráfico proporcional ao caderno cinza, na questão 159, cerca de 28,1% dos participantes acertaram a questão, enquanto 71,9% não assinalaram a alternativa correta.

No caderno azul, apenas um candidato assinalou *, sendo esta uma dupla marcação. 1426 participantes não marcaram nenhuma alternativa, condizente a 0,7% das respostas. Na alternativa A, 50660 escolhas, ou seja, 25,3% das respostas. Alternativa B 31765, sendo igual a 15,9% das respostas. Alternativa C, 55024 participantes, correspondendo a 27,5%. Alternativa D, 35846 participantes ou 17,9% das respostas. E na alternativa E, 25413 respostas, o que equivale a 12,7%. Segundo a análise do gráfico, no caderno azul, na questão 166, cerca de 72,5% dos participantes erraram a questão, enquanto 27,5% acertaram.

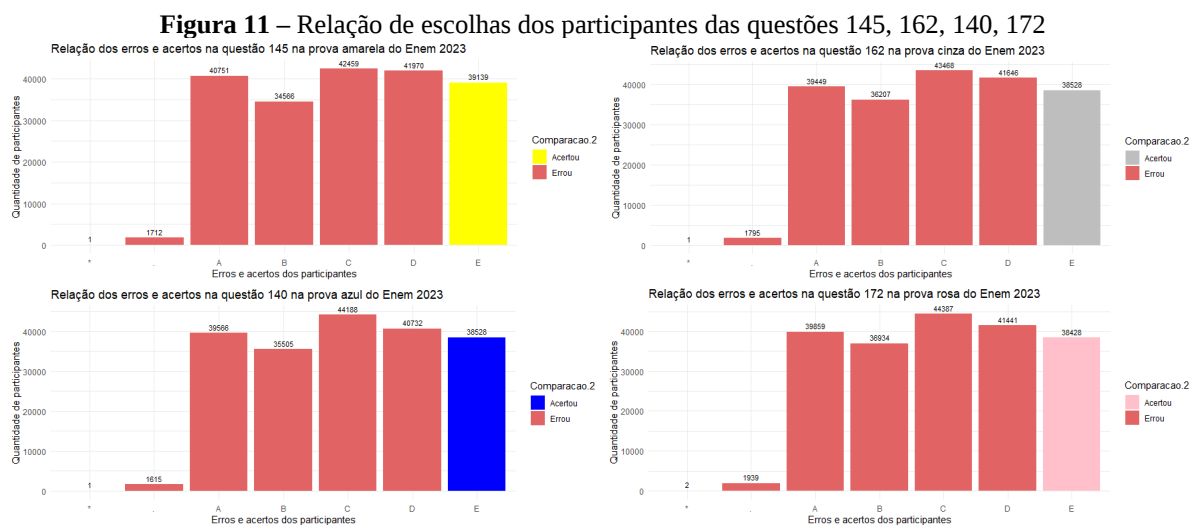
No caderno rosa, um participante realizou uma dupla marcação. 1702 candidatos não marcaram nenhuma alternativa. 48757 assinalaram a alternativa A, o que corresponde a 24% das respostas. 32955 escolheram a alternativa B, ou seja, 16,2%. 54574 responderam à alternativa C, sendo assim, 26,9%. Alternativa D 37239 participantes, proporcional a 18,3%. Alternativa E, 27762 respostas que equivale a 13,7%. A partir da análise do gráfico do caderno rosa, na questão 173, cerca de 73,1% dos participantes erraram a questão, enquanto 26,9% acertaram a alternativa correta.

De acordo com (Giordano, 2020) esta dificuldade em saber quando e como se utilizar as principais medidas de tendência central pode ocorrer com os estudantes, em alguns casos, simplesmente por não saberem calculá-las ou a aplicação mecânica do algoritmo sem compreensão do seu significado.

4.2.2 Relação de escolhas dos participantes das questões 145, 162, 140, 172

Seguindo ainda o caderno amarelo do ENEM 2023, com base na questão 145 do caderno amarelo, os conceitos de moda, média e mediana são tratados como conceitos estatísticos básicos, no qual assumem um papel importante de ferramentas de análise de dados mais simples. Sendo assim, levantamos a hipótese da falta de preparo impede em muitos casos os professores trabalharem tais conteúdos com seus alunos no Ensino Médio, originando um mau aprendizado em relação a estes conceitos imprescindíveis aos estudantes (Quedi; Darroz, 2018). A partir da análise da questão, a fim de exemplificar o desempenho dos alunos apresentamos

na figura 16 a seguir, a relação de escolhas nas alternativas das questões 145, 162, 140, 172 dos cadernos amarelo, cinza, azul e rosa respectivamente.



Fonte: Dados da Pesquisa (2024)

No caderno amarelo, as respostas dos participantes estão distribuídas da seguinte forma: apenas uma dupla marcação, 1712 respostas em branco, correspondendo a 0,8%. Na alternativa A, 40751 respostas, ou 20,3%. Alternativa B, 34566 equivalentes a 17,2% das escolhas na questão. Alternativa C 42459, representando 21,2% das respostas dos participantes. Alternativa D 41970, proporcional a 20,9% das escolhas na questão. Alternativa E 39139 participantes, ou seja, 19,5% das respostas. Segundo a análise do gráfico, no caderno amarelo, na questão 145, cerca de 80,5% dos participantes erraram a questão, enquanto 19,5% acertaram.

No caderno cinza, as respostas analisadas estão exibidas da seguinte forma: apenas uma dupla marcação. 1795 participantes não marcaram nenhuma alternativa, equivalendo a 0,8% das respostas. 39449 participantes assinalaram a alternativa A, o que corresponde a 19,6%. 36207 marcaram a alternativa B, ou seja, 18%. Alternativa C, 43468 dos participantes, sendo proporcional a 21,6%. Alternativa D 41646 escolhas, sendo igual a 20,7%. E por fim, na alternativa E, 38528 participantes assinalaram a alternativa, sendo 19,2% das respostas na questão. Conforme a análise do gráfico, no caderno cinza, na questão 162, cerca de 80,8% dos participantes erraram a questão, enquanto 19,2% acertaram.

No caderno azul, apenas um participante realizou uma dupla marcação. 1615 não marcaram nenhuma alternativa, correspondendo a 0,8% das respostas. 39566 assinalaram a alternativa A, sendo equivalente a 19,8% das respostas. 35505 a alternativa B, proporcional a 17,7% das respostas. Na alternativa C 44188 participantes, ou seja, 22,1%. Alternativa D 40732 escolhas sendo um total de 20,4% das respostas. Alternativa E 38528 das respostas, o que é

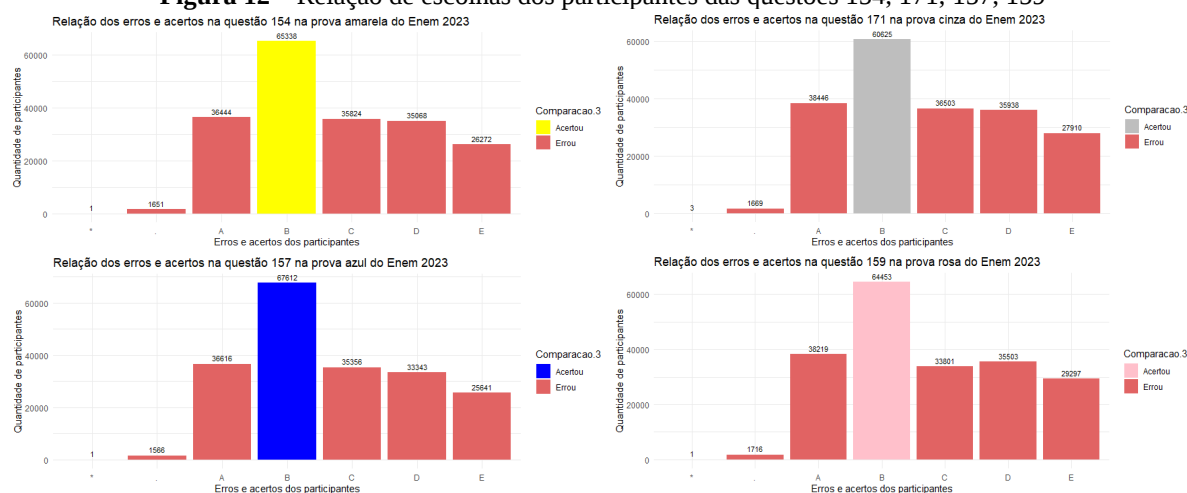
igual a 19,3% das escolhas dos participantes nesta questão. A partir da análise do gráfico, no caderno azul, na questão 140, cerca de 80,7% dos participantes erraram a questão, enquanto 19,3% acertaram.

No caderno rosa, apenas 2 participantes executaram uma dupla marcação. 1939 participantes não marcaram nenhuma alternativa, correspondendo a 0,9% das respostas. 39859 assinalaram a alternativa A, sendo 19,6% das escolhas. 36934 na alternativa B, 18,2%. Na alternativa C 44387 dos participantes, equivalendo a 21,9%. Alternativa D 41441, sendo proporcional a 20,4% das respostas. E por fim na alternativa E 38428, o que coincide com 18,9% das escolhas dos participantes nesta questão. Conforme a análise do gráfico, no caderno rosa, na questão 172, cerca de 81,1% dos participantes erraram a questão, enquanto 18,9% acertaram.

Segundo (Quedi; Darroz, 2018) a dificuldade apresentada pelos participantes neste tipo de questão desenrola-se por muitas vezes pela ausência da abordagem de alguns conceitos como os de média, moda e mediana, sendo estas dificuldades preocupantes, pois estes conhecimentos auxiliam na interpretação da realidade atual.

4.2.3 Relação de escolhas dos participantes das questões 154, 171, 157, 159

Seguindo ainda o caderno amarelo do ENEM 2023, com base na questão 154 do caderno amarelo observamos que o conceito de média também é abordado em conjunto com a análise de gráficos, utilizando informações no meio de um contexto cotidiano a fim de obter um dado específico. Dessa forma, as habilidades referentes ao conceito de média bem como as de medidas de tendência central em alguns casos, não são desenvolvidas de maneira satisfatória sob a hipótese dos estudantes apresentarem dificuldades na interpretação dos problemas, principalmente pela falta de leitura de acordo com (Veras *et al.*, 2021). A partir da análise da questão, com o propósito de exemplificar o desempenho dos alunos, apresentamos na figura 22 a seguir, a relação de escolhas nas alternativas das questões 154, 171, 157, 159 dos cadernos amarelo, cinza, azul e rosa respectivamente.

Figura 12 – Relação de escolhas dos participantes das questões 154, 171, 157, 159

Fonte: Dados da Pesquisa (2024)

No caderno amarelo, as escolhas dos participantes estão distribuídas da seguinte forma: apenas uma dupla marcação. 1651 participantes não marcaram nenhuma alternativa, o que representa 0,8% das respostas. Na alternativa A, 36444 escolhas ou 18,2%. Alternativa B, 65338 o que equivale a 32,6%. Alternativa C 35824 correspondendo a 17,9% das respostas. Alternativa D, 35068 participantes sendo proporcional à 17,5%. E por fim alternativa E, 26272 escolhas que coincidem com 13,1% das respostas. A partir da análise do gráfico, no caderno amarelo questão 154, cerca de 67,4% dos participantes erraram a alternativa correta da questão, enquanto 32,6% acertaram.

No caderno cinza, apenas três dos participantes realizaram uma dupla marcação. 1669 não assinalaram nenhuma alternativa, sendo igual a 0,8% das respostas. 38446 participantes escolheram a alternativa A, o que corresponde a 19,1%. Na alternativa B 60625, equivalendo a 30,1% das respostas. Alternativa C 36503 ou seja, 18,2%. Alternativa D 35938 das escolhas representando 17,9%. E finalmente a alternativa E 27910, o que é proporcional a 13,9% das respostas. Com base na análise do gráfico, no caderno cinza questão 171, cerca de 69,9% dos participantes erraram a questão, enquanto 30,1% acertaram.

No caderno azul, apenas uma dupla marcação, enquanto 1566 participantes não assinalaram nenhuma alternativa, correspondendo a 0,7% das respostas. Na alternativa A 36616 escolhas, o que equivale a 18,3%. Na alternativa B 67612 sendo igual a 33,8%. Alternativa C 35356, ou seja, 17,7%. Alternativa D 33343 respostas, o que representa 16,7%. E na alternativa E 25641 participantes, proporcional a 12,8% das respostas. De acordo com a análise do gráfico, no caderno azul, questão 157, cerca de 66,2% dos participantes erraram a questão, enquanto 33,8% acertaram.

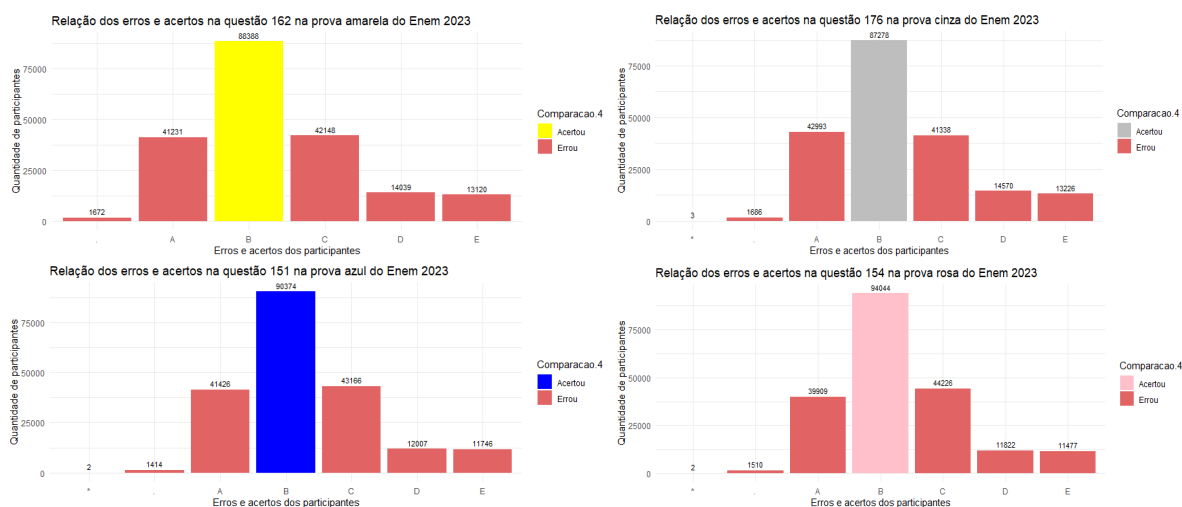
No caderno rosa, apenas uma dupla marcação. 1716 participantes não marcaram nenhuma alternativa, correspondendo a 0,8% das respostas na questão. 38219 assinalaram a alternativa A, sendo este equivalente a 18,8%. Alternativa B 64453 participantes, ou seja, 31,8% das respostas. Alternativa C 33801, representando 16,7%. Alternativa D 35503 escolhas, sendo igual a 17,5%. E na alternativa E 29297, o que é proporcional a 14,4% das respostas. Conforme a análise do gráfico, no caderno rosa, questão 159, cerca de 68,2% erraram a questão, enquanto 31,8% acertaram.

Sob a ótica de (Hollas; Bernardi, 2019) a dificuldade dos estudantes nesta questão pode suceder-se principalmente pela tarefa de interpretar o enunciado da questão, dessa forma se faz necessário a potencialização no uso das competências da Educação Estatística Crítica.

4.2.4 Relação de escolhas dos participantes das questões 162, 176, 151, 154

Continuando no caderno amarelo do ENEM 2023, com base na questão 162, a qual é proporcional às questões 176, 151, 154 dos cadernos cinza, azul e rosa respectivamente. Observamos que este tipo de questão exige ao participante a análise de gráficos e tabelas. Sob este viés se faz necessário o desenvolvimento do pensamento estatístico crítico (Hollas; Bernardi, 2019). A partir da análise da questão, com o propósito de exemplificar o desempenho dos alunos, apresentamos na figura 28 a seguir, a relação de escolhas nas alternativas das questões 162, 176, 151, 154 dos cadernos amarelo, cinza, azul e rosa respectivamente.

Figura 13 – Relação de escolhas dos participantes das questões 162, 176, 151, 154



Fonte: Dados da Pesquisa (2024)

No caderno amarelo, nenhum dos participantes realizou uma dupla marcação. 1672 participantes não assinalaram nenhuma alternativa, sendo 0,8% das respostas. Na alternativa A 41231 escolhas, o que corresponde a 20,6%. Na alternativa B 88388, equivalente a 44,1%. Na alternativa C 42148, ou seja, 21% das respostas. Na alternativa D 14039, o que é proporcional a 7%. E na alternativa E 13120, igual a 6,5% das respostas na questão. Conforme a análise do gráfico, na questão 162, cerca de 55,9% dos participantes assinalaram a alternativa incorreta na questão, enquanto 44,1% acertaram.

No caderno cinza, apenas três duplas marcações. 1686 participantes não marcaram nenhuma alternativa, correspondendo a 0,8% das respostas. Alternativa A 42993 equivalente a 21,4%. Na alternativa B 87278, 43,4% das escolhas. Na alternativa C 41338, ou seja, 20,6%. Alternativa D 14570, representando 7,2% das respostas. E por fim, alternativa E com 13226 escolhas, portanto, 6,5%. Segundo a análise da questão a partir do gráfico, na questão 176, cerca de 56,6% participantes erraram a questão, enquanto 43,4% acertaram.

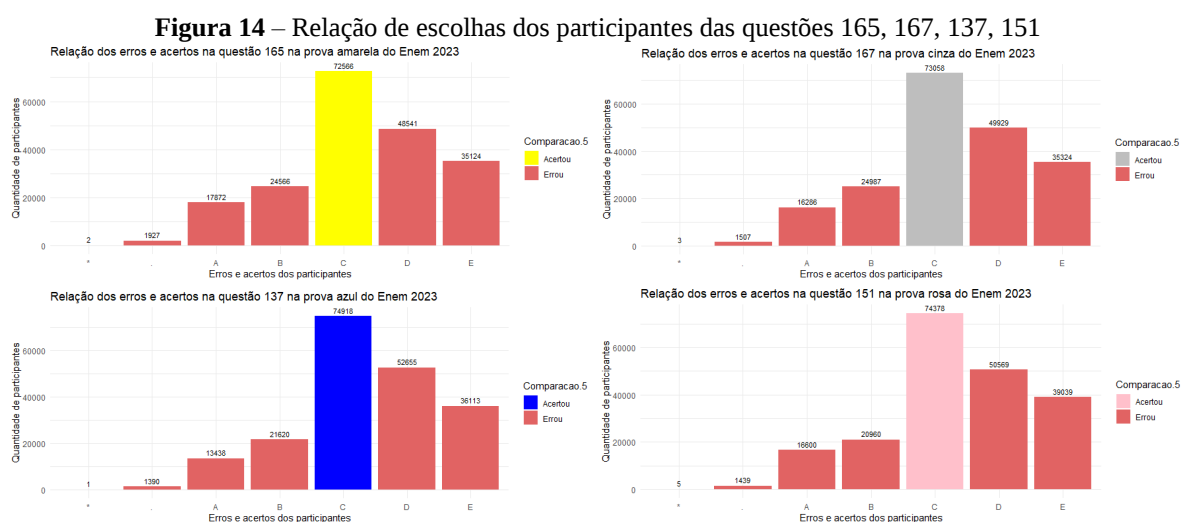
No caderno azul apenas duas duplas marcações. 1414 participantes não assinalaram nenhuma das alternativas, correspondendo a 0,7% das respostas. 41426 escolhas na alternativa A, ou 20,7%. 90374 escolhas na alternativa B, o que equivale a 45,2% das respostas. 43166 na alternativa C, ou seja, 21,6%. 12007 escolhas na alternativa D, sendo igual a 6% das respostas. E por fim na alternativa E 11746 participantes o que é proporcional a 5,8% das escolhas na questão. Segundo a análise do gráfico, na questão 151, cerca de 54,8% dos participantes erraram a questão, enquanto 45,2% acertaram.

No caderno rosa, apenas duas duplas marcações. 1510 participantes não assinalaram nenhuma das alternativas, correspondendo a 0,7% das respostas. Na alternativa A 39909, ou seja, 19,7%. Na alternativa B 94044 escolhas, o que representa 46,3%. Alternativa C 44226, equivalente a 21,8%. Alternativa D 11822, sendo igual a 5,8% das respostas. E finalmente a alternativa E com 11477 escolhas, ou 5,6%. Conforme a análise do gráfico, na questão 154, cerca de 53,7% dos participantes erraram a questão, enquanto 46,3% acertaram.

Conforme (Lima; Giordano, 2021) a dificuldade apresentada por alguns estudantes nestas questões pode ser oriunda da falta de um bom letramento estatístico, onde os conhecimentos são adquiridos em cada fase do aprendizado. O ato de ultrapassar algumas destas fases ou mau aproveitamento leva a dificuldade na interpretação de gráficos e tabelas devido ao baixo desenvolvimento investigativo e crítico.

4.2.5 Relação de escolhas dos participantes das questões 165, 167, 137, 151

Prosseguindo no caderno amarelo do ENEM 2023, com base na questão 165, a qual é proporcional às questões 167, 137, 151 dos cadernos cinza, azul e rosa respectivamente. Percebe-se que as questões abordam novamente o conceito de médias. Em relação aos problemas sobre medidas de tendência central, principalmente a ideia de média, deve-se valorizar problemas que tratem mais com o significado destas medidas (Lopes; Corral; Resende, 2012). A partir da análise da questão, com o propósito de exemplificar o desempenho dos alunos, apresentamos na figura 34 a seguir, a relação de escolhas nas alternativas das questões 165, 167, 137, 151 dos cadernos amarelo, cinza, azul e rosa respectivamente.



Fonte: Dados da Pesquisa (2024)

No caderno amarelo, apenas duas duplas marcações. 1927 participantes não marcaram nenhuma das alternativas, correspondendo a 0,9% das respostas. Na alternativa A 17872 escolhas, o que resulta em 8,9%. Alternativa B 24566, representando 12,2% das escolhas. Alternativa C 72566, sendo igual a 36,2%. Alternativa D 48541, sendo equivalente a 24,2% das respostas. E na alternativa E 35124, representando 17,5% das respostas dos participantes. Segundo a análise do gráfico, na questão 165, cerca de 63,8% dos participantes erraram a questão, enquanto 36,2% acertaram.

No caderno cinza, apenas três participantes realizaram uma dupla marcação. 1507 não assinalaram nenhuma das alternativas, correspondendo a 0,7% das respostas. Na alternativa A 16286, ou seja, 8,1%. Na alternativa B 24987, o que equivale a 12,4% das escolhas. Na alternativa C 73058, sendo igual a 36,3%. Na alternativa D 49929, o que representa 24,8%. E

na alternativa E 35324, é proporcional à 17,6% das respostas. Segundo a análise do gráfico, na questão 167, cerca de 63,7% dos participantes erraram a questão, enquanto 36,3% acertaram.

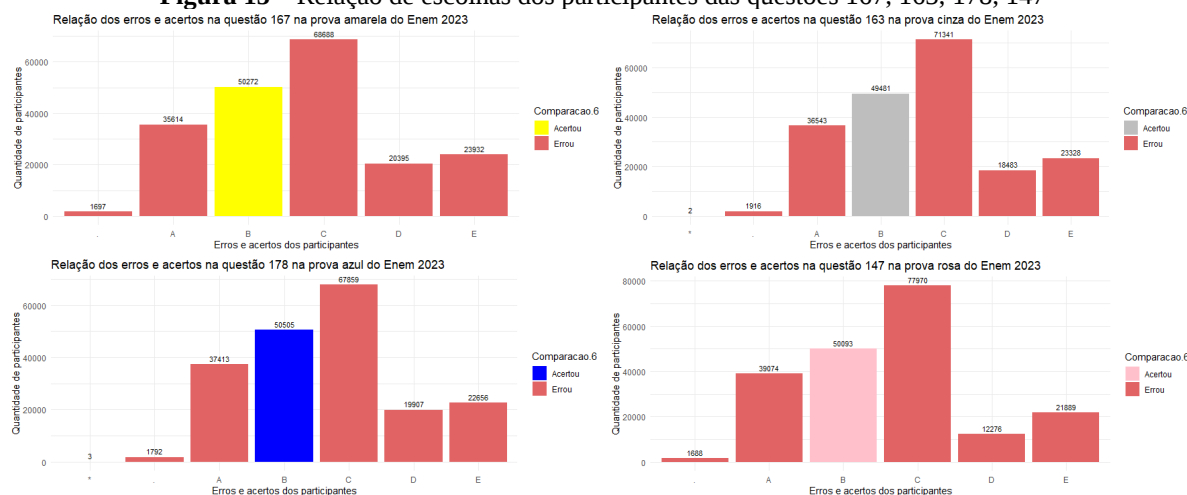
No caderno azul, apenas uma dupla marcação. 1390 participantes não marcaram nenhuma das alternativas, o que corresponde a 0,6% das respostas na questão. Na alternativa A 13438 escolhas o que equivale a 6,7% das respostas. Na alternativa B 21620 participantes, ou seja, 10,8% das respostas. Na alternativa C 74918, sendo igual a 37,4%. Na alternativa D 52655 escolhas, sendo proporcional a 26,3%. E na alternativa E 36113, o que representa 18% das respostas. A partir da análise do gráfico, na questão 137, cerca de 62,6% dos participantes erraram a questão, enquanto 37,4% acertaram.

No caderno rosa, houve cinco duplas marcações. 1439 participantes não assinalaram nenhuma das alternativas, representando 0,7% das respostas na questão. 16600 escolhas na alternativa A, sendo 8,1% das respostas na questão. 20960 na alternativa B, ou seja, 10,3%. Na alternativa C 74378, o que corresponde a 36,6%. Na alternativa D 50569, proporcional a 24,9% das respostas. E na alternativa E 39039, igual a 19,2% das escolhas dos participantes. Segundo a análise do gráfico, na questão 151, cerca de 63,4% dos participantes erraram a questão, enquanto 36,6% acertaram.

Segundo (Lopes; Corral; Resende, 2012) a principal dificuldade encontrada pelos participantes neste tipo de questão pode estar associada à falta de resolução de situações-problema, aos quais os conhecimentos estatísticos devem estar inseridos no cotidiano dos estudantes.

4.2.6 Relação de escolhas dos participantes das questões 167, 163, 178, 147

Seguindo no caderno amarelo do ENEM 2023, com base na questão 167, a qual é proporcional às questões 163, 178, 147 dos cadernos cinza, azul e rosa respectivamente. Observamos que a questão se trata da análise de gráficos e tabelas. Sobre os problemas que envolvem analisar tabelas e gráficos, os estudantes necessitam de um processo investigativo no qual envolvem variáveis categóricas para sua realização (Lima; Giordano, 2021). A partir da análise da questão, com o propósito de exemplificar o desempenho dos alunos, apresentamos na figura 40 a seguir, a relação de escolhas nas alternativas das questões 167, 163, 178, 147 dos cadernos amarelo, cinza, azul e rosa respectivamente.

Figura 15 – Relação de escolhas dos participantes das questões 167, 163, 178, 147

Fonte: Dados da Pesquisa (2024)

No caderno amarelo, nenhum dos participantes realizou uma dupla marcação. 1697 participantes não assinalaram nenhuma das alternativas, correspondendo a 0,8% das respostas. Na alternativa A 35614 escolhas, o que representa 17,8% das respostas na questão. Alternativa B 50272, sendo igual a 25,1%. Na alternativa C 68688, proporcional a 34,2%. Alternativa D 20395 escolhas, ou seja, 10,2%. Alternativa E 23932, 11,9% das respostas. De acordo com a análise do gráfico na questão 167, cerca de 74,9% dos participantes erraram a questão, enquanto 25,1% acertaram.

No caderno cinza, apenas duas duplas marcações. 1916 participantes não marcaram nenhuma alternativa, o que representa 0,9% das respostas. Na alternativa A 36543 escolhas, correspondendo a 18,2%. Alternativa B 49481 participantes, o equivalente a 24,6%. Alternativa C 71341 escolhas, ou seja, 35,5%. Alternativa D 18483 escolhas dos participantes, o que é igual a 9,1% das respostas. E na alternativa E 23328, 11,6%. Conforme a análise do gráfico da questão 163, cerca de 75,4% dos participantes erraram a questão, enquanto 24,6% acertaram.

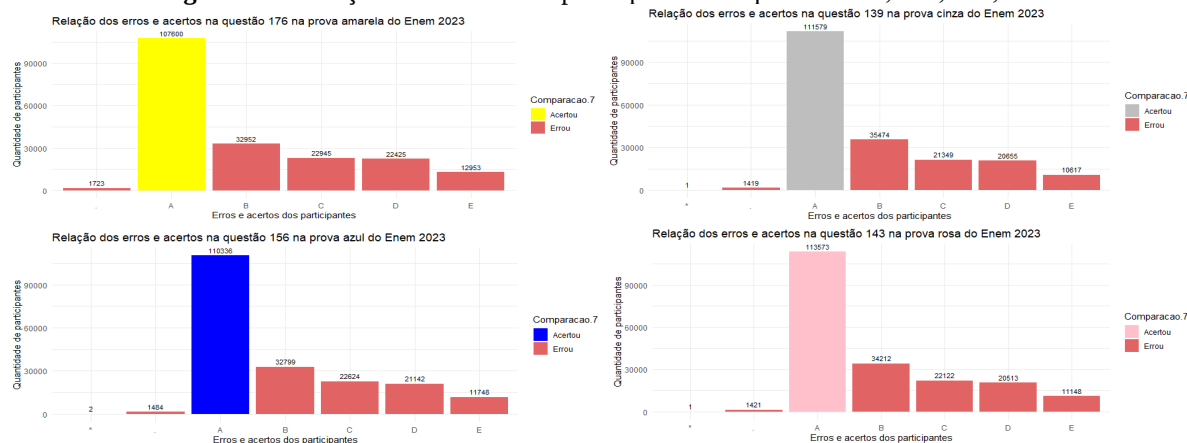
No caderno azul, apenas três participantes realizaram duplas marcações. 1792 participantes não assinalaram nenhuma alternativa, o que corresponde a 0,8% das respostas. 37413 participantes escolheram a alternativa A, o que representa 18,7% das respostas. 50505 escolhas na alternativa B, ou seja, 25,2% das respostas. 67859 participantes escolheram a alternativa C, o que equivale a 33,9% das respostas. 19907 escolhas na alternativa D, proporcional a 9,9% das respostas. E por fim na alternativa E, 22656 escolhas sendo igual a 11,3%. Em consoante com a análise do gráfico, na questão 178, cerca de 74,8% dos participantes erraram a questão, enquanto 25,2% acertaram.

No caderno rosa, não houve dupla marcação. 1688 participantes não assinalaram nenhuma das alternativas, sendo 0,8% do total das respostas na questão. 39074 participantes escolheram a alternativa A, o que corresponde a 19,2%. 50093 participantes escolheram a alternativa B, ou seja, 24,7%. 77970 participantes escolheram a alternativa C, o que representa 38,4%. 12276 participantes na alternativa D, sendo igual a 6%. E na alternativa E 21889 escolhas, portanto, 10,8%. A partir da análise do gráfico, na questão 147, cerca de 75,3% dos participantes erraram a questão, enquanto 24,7% dos participantes acertaram.

Conforme (Lima; Giordano, 2021) a principal eventualidade de dificuldade encontrada pelos participantes neste tipo de questão diz respeito ao conceito de espaço amostral introduzido por meio dos gráficos, o estudante é desafiado a estimar e compreender aspectos da realidade seguindo um processo investigativo, sendo uma habilidade desenvolvida inicialmente no 3º ano, mostrando que boa parte dos estudantes sofrem uma defasagem no que tange a esta habilidade.

4.2.7 Relação de escolhas dos participantes das questões 176, 139, 156, 143

Seguindo no caderno amarelo do ENEM 2023, com base na questão 176, a qual é equivalente às questões 139, 156, 143 dos cadernos cinza, azul e rosa respectivamente. A questão trata-se de mais uma análise de gráficos e tabelas. Neste tipo de questão voltamos nosso olhar novamente para educação estatística crítica, de forma que se faz uso da literacia estatística (Hollas; Bernardi, 2019). A partir da análise da questão, com o propósito de apresentar o desempenho dos alunos, apresentamos na figura 46 a seguir, a relação de escolhas nas alternativas das questões 176, 139, 156, 143 dos cadernos amarelo, cinza, azul e rosa respectivamente.

Figura 16 – Relação de escolhas dos participantes das questões 176, 139, 156, 143

Fonte: Dados da Pesquisa (2024)

No caderno amarelo, nenhum dos participantes realizou uma dupla marcação. 1723 não assinalaram nenhuma das alternativas o que corresponde a 0,8% das respostas. Na alternativa A 107600 escolhas, o que representa 53,6%. Na alternativa B 32952 participantes, o que equivale a 16,4%. 22945 participantes escolheram a alternativa C, sendo 11,4%. 22425 escolheram a alternativa D, ou seja, 11,2%. E na alternativa E 12953, sendo igual a 6,4%. De acordo com a análise do gráfico, na questão 176, cerca de 46,4% dos participantes erraram a questão, enquanto 53,6% dos participantes acertaram.

No caderno cinza, apenas um participante realizou uma dupla marcação. 1419 participantes não assinalaram nenhuma alternativa, representando 0,7% das respostas. 111579 participantes escolheram a alternativa A, correspondendo a 55,5%. 35474 escolheram a alternativa B, ou seja, 17,6%. 21349 escolheram a alternativa C, o que equivale a 10,6%. Na alternativa D 20655 participantes, sendo proporcional a 10,3%. E na alternativa E 10617, sendo 5,2% das respostas. Segundo a análise do gráfico, na questão 139, cerca de 44,5% dos participantes erraram a questão, enquanto 55,5% acertaram.

No caderno azul, apenas dois participantes realizaram uma dupla marcação. 1484 participantes não assinalaram nenhuma alternativa, o que representa 0,7% das respostas. 110336 participantes escolheram a alternativa A, o que corresponde a 55,1%. 32799 escolhas na alternativa B, o que equivale a 16,4% das respostas. Na alternativa C 22624 escolhas, sendo 11,3% das respostas. Na alternativa D 21142, ou seja, 10,6%. E na alternativa E 11748 escolhas dos participantes, sendo igual a 5,8% das respostas. Conforme a análise do gráfico, na questão 156, cerca de 44,9% dos participantes erraram a questão, enquanto 55,1% acertaram.

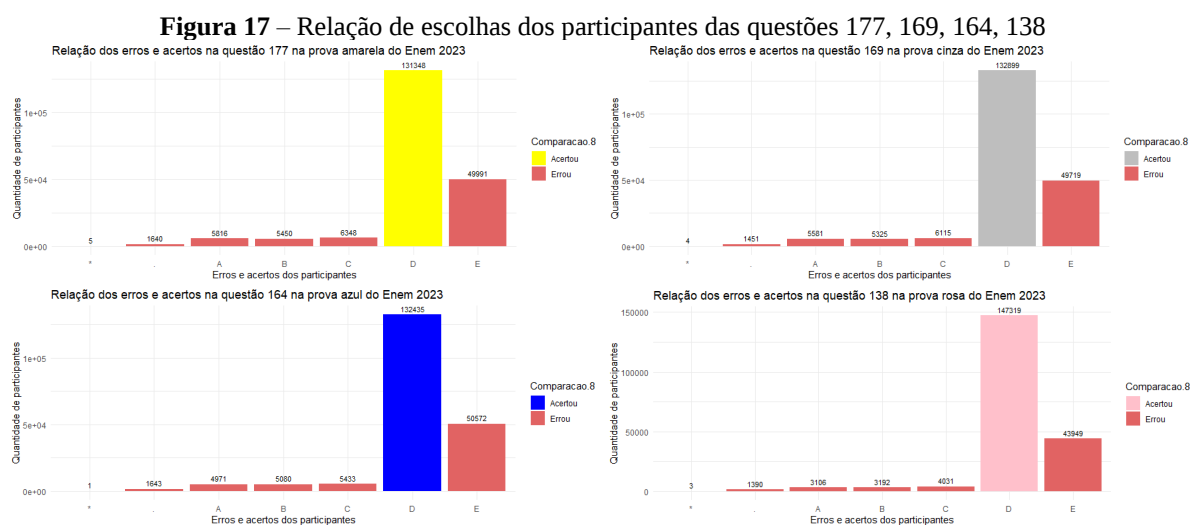
No caderno rosa, apenas um participante realizou uma dupla marcação. 1421 participantes não assinalaram nenhuma das alternativas, sendo proporcional a 0,7% das

respostas na questão. 113573 participantes escolheram a alternativa A, ou seja, 56% das respostas. Na alternativa B 34212 escolhas, o que equivale a 16,9%. Na alternativa C 22122 escolhas, ou seja, 10,9%. Alternativa D 20513 participantes, o que representa 10,1%. E na alternativa E 11148 participantes, portanto, 5,4% das respostas. A partir da análise do gráfico, na questão 143, cerca de 44% dos participantes erraram a questão, enquanto 56% acertaram.

De acordo com (Skovsmose, 2015) a principal hipótese que justifica a dificuldade encontrada pelos estudantes na questão se dá na leitura das informações do gráfico, a noção de contagem está além de operar com números, sendo assim a matemacia é descrita como um modo de ler o mundo, utilizando destes artifícios e ferramentas.

4.2.8 Relação de escolhas dos participantes das questões 177, 169, 164, 138

Ainda no caderno amarelo do ENEM 2023, com base na questão 177, em que equivalem às questões 169, 164, 138 dos cadernos cinza, azul e rosa respectivamente. A questão trata-se mais uma vez de uma análise de gráficos e tabelas. Neste tipo de questão é imprescindível que o estudante trate a representação oriunda dos gráficos e tabelas como uma aproximação da realidade, bem como a passagem de um tipo de representação para outro, a fim de construir novos conhecimentos (Bivar; Selva, 2011). A partir da análise da questão, com o propósito de apresentar o desempenho dos alunos, apresentamos na figura 52 a seguir, a relação de escolhas nas alternativas das questões 177, 169, 164, 138 dos cadernos amarelo, cinza, azul e rosa respectivamente.



Fonte: Dados da Pesquisa (2024)

No caderno amarelo, apenas 5 participantes realizaram uma dupla marcação. 1640 não assinalaram nenhuma alternativa, o que corresponde a 0,8% das respostas. Na alternativa A 5816 escolhas, ou seja, 2,9% das respostas. Na alternativa B 5450, representando 2,7%. Na alternativa C 6348 escolhas, equivalente a 3,1%. Na alternativa D 131348 escolhas dos participantes, sendo igual a 65,5% das respostas. E na alternativa E 49991, proporcional a 24,9% das respostas na questão. De acordo com a análise do gráfico, na questão 177, cerca de 34,5% dos participantes erraram a questão, enquanto 65,5% acertaram.

No caderno cinza, apenas 4 participantes realizaram uma dupla marcação. 1451 participantes não marcaram nenhuma alternativa, ou seja, 0,7% das respostas. 5581 participantes escolheram a alternativa A, representando 2,7%. 5325 participantes escolheram a alternativa B, o que equivale a 2,6% das respostas. 6115 escolhas na alternativa C, sendo igual a 3%. 132899 participantes escolheram a alternativa D, sendo proporcional a 66,1% das respostas. E na alternativa E 49719 escolhas, portanto, 24,7%. A partir da análise do gráfico, na questão 169, cerca de 33,9% dos participantes erraram a questão, enquanto 66,1% acertaram.

No caderno azul, apenas um participante efetuou uma dupla marcação. 1643 não assinalaram nenhuma alternativa, o que corresponde a 0,8% das respostas. 4971 participantes escolheram a alternativa A, o que representa 2,4%. 5080 participantes escolheram a alternativa B, sendo igual a 2,5% das respostas. Na alternativa C 5433 escolhas, ou seja, 2,7%. Na alternativa D 132435 escolhas, o que equivale a 66,2% das respostas. E na alternativa E 50572 escolhas dos participantes, sendo proporcional a 25,3%. Conforme a análise do gráfico, na questão 164, cerca de 33,8% dos participantes erraram a questão, enquanto 66,2% acertaram.

No caderno rosa, apenas 3 participantes realizaram uma dupla marcação. 1390 não assinalaram nenhuma das alternativas, correspondendo a 0,6% das respostas. 3106 participantes escolheram a alternativa A, o que representa 1,5% das respostas. Na alternativa B 3192 escolhas, sendo 1,5% das respostas. Na alternativa C 4031 participantes, o que é equivalente a 1,9% das respostas. Alternativa D 147319 escolhas, sendo igual a 72,6%. E na alternativa E 43949 escolhas, proporcional a 21,7% das respostas. Conforme a análise do gráfico, na questão 138, cerca de 27,4% dos participantes erraram a questão, enquanto 72,6% acertaram.

Segundo (Frei; Rosa; Biazi, 2023) a principal dificuldade encontrada pelos participantes nesta questão pode ter sucedido pela falta de habilidade ao ler e interpretar dados oriundos de gráficos e tabelas, seja por própria defasagem do aluno ou carência de uma formação continuada para o professor no que diz respeito ao ensino de estatística.

5 CONSIDERAÇÕES FINAIS

Ao longo deste trabalho, exploramos o papel da mineração de dados educacionais aberto, a fim de identificar padrões e analisar as questões de estatística na prova de Matemática do ENEM 2023. Por meio de uma extensa análise gráfica das escolhas dos participantes, foi possível constatar o potencial da mineração de dados para reconhecer padrões e categorias de um dado que esteja sendo estudado e analisado.

Além disso, percebemos que a mineração de dados educacionais permite ao professor e futuro professor, analisar os dados dos seus alunos de forma que acompanhe os níveis de aprendizagem tanto da turma quanto individualmente, fazendo uso desta ferramenta a fim de melhorar sua prática pedagógica e obter resultados satisfatórios.

No entanto, é necessário ressaltar que a eficácia do uso da mineração de dados educacionais está relacionada tanto à disponibilidade de tais recursos, como à capacitação docente e o contexto em que será inserida. É importante que as políticas educacionais, as instituições de ensino e os professores busquem a flexibilidade de utilizar estas ferramentas, além de reconhecer sua relevância tecnológica.

Outro fator que é válido ressaltar diz respeito às questões de estatística, a partir da análise observamos que em suma os conteúdos mais abordados referem-se a análise de gráficos e tabelas, bem como medidas de tendência central. Estes conteúdos são considerados introdutórios para os métodos estatísticos, no entanto, os estudantes apresentaram menos acertos nestes tipos de questão.

Estas dificuldades ocorrem por diversos fatores, sendo infortúnios individuais ou coletivos, devido à falta de preparo de alguns professores em relação ao ensino de estatística e até mesmo insuficiência no aprendizado de conceitos estatísticos por parte dos estudantes.

Por fim, o presente trabalho destaca a necessidade de ampliar os estudos na integração da mineração de dados educacionais na sala de aula, de forma que possamos garantir um fortalecimento do ensino de matemática e das ferramentas de avaliação contínua, bem como capacitar os professores ao seu uso de forma que atenda às necessidades individuais de seus alunos.

REFERÊNCIAS

- AGGARWAL, Charu C. **Data Mining: The Textbook**. 1. ed. New York, USA: Springer, 2015. v. 1Ebook. Disponível em: <https://doi.org/10.1007/978-3-319-14142-8>.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de Dados Educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, [s. l.], v. 19, n. 02, p. 03, 2011.
- BAKER, R. S.; MARTIN, T.; ROSSI, L. M. Educational Data Mining and Learning Analytics. In: THE WILEY HANDBOOK OF COGNITION AND ASSESSMENT. [S. l.]: John Wiley & Sons, Ltd, 2016. p. 379–396. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118956588.ch16>. Acesso em: 9 abr. 2024.
- BENEVENTO, M. *et al.* PREVER E MELHORAR O DESEMPENHO DOS ALUNOS COM O USO COMBINADO DE APRENDIZAGEM DE MÁQUINA E GPT. **Regae: Revista de Gestão e Avaliação Educacional**, [s. l.], v. 12, n. 21, 2023. Disponível em: http://educa.fcc.org.br/scielo.php?script=sci_abstract&pid=S2318-13382023000100222&lng=pt&nrm=iso&tlng=pt. Acesso em: 5 jul. 2024.
- BIVAR, D. B.; SELVA, A. C. Analisando atividades envolvendo gráficos e tabelas nos livros didáticos de matemática (CO). In: XIII CONFERÊNCIA INTERAMERICANA DE EDUCAÇÃO MATEMÁTICA, 2011. **Anais [...]**. [S. l.: s. n.], 2011. Disponível em: https://xiii.ciaem-redumate.org/index.php/xiii_ciaem/xiii_ciaem/paper/view/1332. Acesso em: 8 out. 2024.
- BORGMAN, C. L. **Big data, little data, no data: scholarship in the networked world**. London: The MIT Press, 2015.
- BRASIL. Base Nacional Comum Curricular. Brasília: Ministério da Educação, 2018.
- BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Sobre o Enem - Inep. Brasília, 2016b. Disponível em: <http://portal.inep.gov.br/web/enem/sobre-oenem>. Acesso em: 18 dez. 2016.
- COSTA, E. *et al.* Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. **Jornada de Atualização em Informática na Educação**, [s. l.], v. 1, n. 1, p. 1–29, 2012.
- DE SOUZA, V. F.; DE SOUZA, M. F. Os avanços da mineração de dados educacionais: definições, processo e evolução Advances in educational data mining: definitions, process and evolution. **Brazilian Journal of Development**, [s. l.], v. 7, n. 8, p. 80798–80816, 2021.
- EDUCACIONALDATAMINING.ORG. In: [s. d.]. Disponível em: <https://educationaldatamining.org/>. Acesso em: 9 abr. 2024.

EYNG, A. M. Análise de agrupamento pelos métodos hierárquico aglomerativo e particional fuzzy utilizados para educational data mining em dados de educação a distância. [s. l.], 2021. Disponível em: <http://repositorio.unesc.net/handle/1/8174>. Acesso em: 24 maio 2024.

FREI, F.; ROSA, J. S.; BIAZI, Â. H. Professores de Matemática estão preparados para o ensino de Estatística e Probabilidade?. **Revista Internacional de Pesquisa em Educação Matemática**, [s. l.], v. 13, n. 2, p. 1–17, 2023.

FONTELLES, Mauro José et al. Metodologia da pesquisa científica: diretrizes para a elaboração de um protocolo de pesquisa. Rev. para. med, 2009.

GIORDANO, C. **Concepções sobre Estatística: um estudo com alunos do ensino médio**. 2020. [s. l.], 2020.

HOLLAS, J.; BERNARDI, L. T. M. dos S. O Exame Nacional do Ensino Médio (Enem) e as competências para uma Educação Estatística Crítica. **Ensaio: Avaliação e Políticas Públicas em Educação**, [s. l.], v. 28, p. 110–134, 2019.

LIMA, S. de O. *et al.* Ensino de Estatística, Probabilidade e Combinatória na Educação Básica: : os novos desafios da BNCC. **Revista Baiana de Educação Matemática**, [s. l.], v. 3, n. 01, p. e202209–e202209, 2022.

LIMA, S.; GIORDANO, C. Letramento estatístico: um olhar sobre a BNCC. In: [S. l.: s. n.], 2021. p. 473–494.

LOPES, J. M.; CORRAL, R. S.; RESENDE, J. S. O ESTUDO DA MÉDIA, DA MEDIANA E DA MODA ATRAVÉS DE UM JOGO E DA RESOLUÇÃO DE PROBLEMAS. *Revista Eletrônica de Educação*, [s. l.], v. 6, n. 2, p. 250–270, 2012.

MENDES, R. de Á.; SILVA, L. A. Mineração de Dados Educacionais Aplicada à Análise Preditiva em Fóruns no Moodle. **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**, [s. l.], v. 5, n. 1, p. 940, 2016.

MOISSA, B.; GASPARINI, I.; KEMCZINSKI, A. Educational Data Mining versus Learning Analytics: estamos reinventando a roda? Um mapeamento sistemático. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, [s. l.], v. 26, n. 1, p. 1167, 2015.

MUSSI, R. F. de F.; MUSSI, L. M. P. T.; ASSUNÇÃO, E. T. C.; NUNES, C. P. Pesquisa Quantitativa e/ou Qualitativa: distanciamentos, aproximações e possibilidades. **Revista Sustinere**, v. 7, n. 2, p. 414–430, 2019. <https://doi.org/10.12957/sustinere.2019.41193>.

NATIONAL SCIENCE BOARD (NSB). Long-lived digital data collections: enabling research and education in the 21st century. 2005. Disponível em: <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em: 28 jul. 2024.

NASCIMENTO, R. L. S. do; JUNIOR, G. G. da C.; FAGUNDES, R. A. de A. Mineração de Dados Educacionais: Um Estudo Sobre Indicadores da Educação em Bases de Dados do INEP. **Revista Novas Tecnologias na Educação**, [s. l.], v. 16, n. 1, 2018. Disponível em: <https://seer.ufrgs.br/index.php/renote/article/view/85989>. Acesso em: 14 out. 2024.

NETO, M. V. G.; VASCONCELOS, G. C.; ZANCHETTIN, C. Mineração de Dados Aplicada à Predição do Desempenho de Escolas e Técnicas de Interpretabilidade dos Modelos. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 2021. **Simpósio Brasileiro de Informática na Educação (SBIE)**. [S. l.]: SBC, 2021. p. 773–782. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/18105>. Acesso em: 5 jul. 2024.

PARGA, J. P. F. A.; FARIA, P. Introdução à Linguagem R: seus fundamentos e sua prática. [s. l.], 2021. Disponível em: https://www.academia.edu/44520051/Introdu%C3%A7%C3%A3o_%C3%A0_Linguagem_R_seus_fundamentos_e_sua_pr%C3%A1tica. Acesso em: 3 jul. 2024.

PEÑA-AYALA, A. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, v. 4, n. 41, p. 1432-1462, 2014.

QUEDI, R. P.; DARROZ, L. M. Conceitos básicos de estatística: as lacunas conceituais de acadêmicos ingressantes no curso superior. **Revista Thema**, [s. l.], v. 15, n. 4, p. 1256–1268, 2018.

R: O QUE É R? [S. l.], [s. d.]. Disponível em: <https://www.r-project.org/about.html>. Acesso em: 3 jul. 2024.

RODRIGUES, R.; GOMES, A. S.; ADEODATO, P. Uma abordagem de Mineração de Dados Educacionais para previsão de desempenho a partir de padrões comportamentais de Autorregulação da Aprendizagem. In: VI CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2017, Recife, Pernambuco, Brasil. **Anais [...]**. Recife, Pernambuco, Brasil: [s. n.], 2017. p. 13. Disponível em: <http://www.br-ie.org/pub/index.php/wcbie/article/view/7355>. Acesso em: 4 abr. 2024.

ROMERO, C; VENTURA, C. Data mining in education. *WIREs Data Mining Knowl Discov*, February 2013. 12-27.

ROMERO, C.; VENTURA, C. Educational data mining: a review of the state of the art. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev*, 2010. 601-618.

SILVA, Leandro A.; MORINO, Anderson Hideki; SATO, Thiago Massahiro Conti. Prática de mineração de dados no Exame Nacional do Ensino Médio. In: WORKSHOPS DO CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 3., n. 1, p. 651-660, 2014, Dourados, MS. **Anais [...]**. Dourados, MS: Sociedade Brasileira de Computação, 2014. Disponível em: <http://www.br-ie.org/pub/index.php/wcbie/article/view/3289>. Acesso em: 22 abr. 2024.

SILVA, L. F. da. **Dados Abertos Governamentais conectados em Big Data: framework conceitual**. [S. l.]: Editora Dialética, 2022.

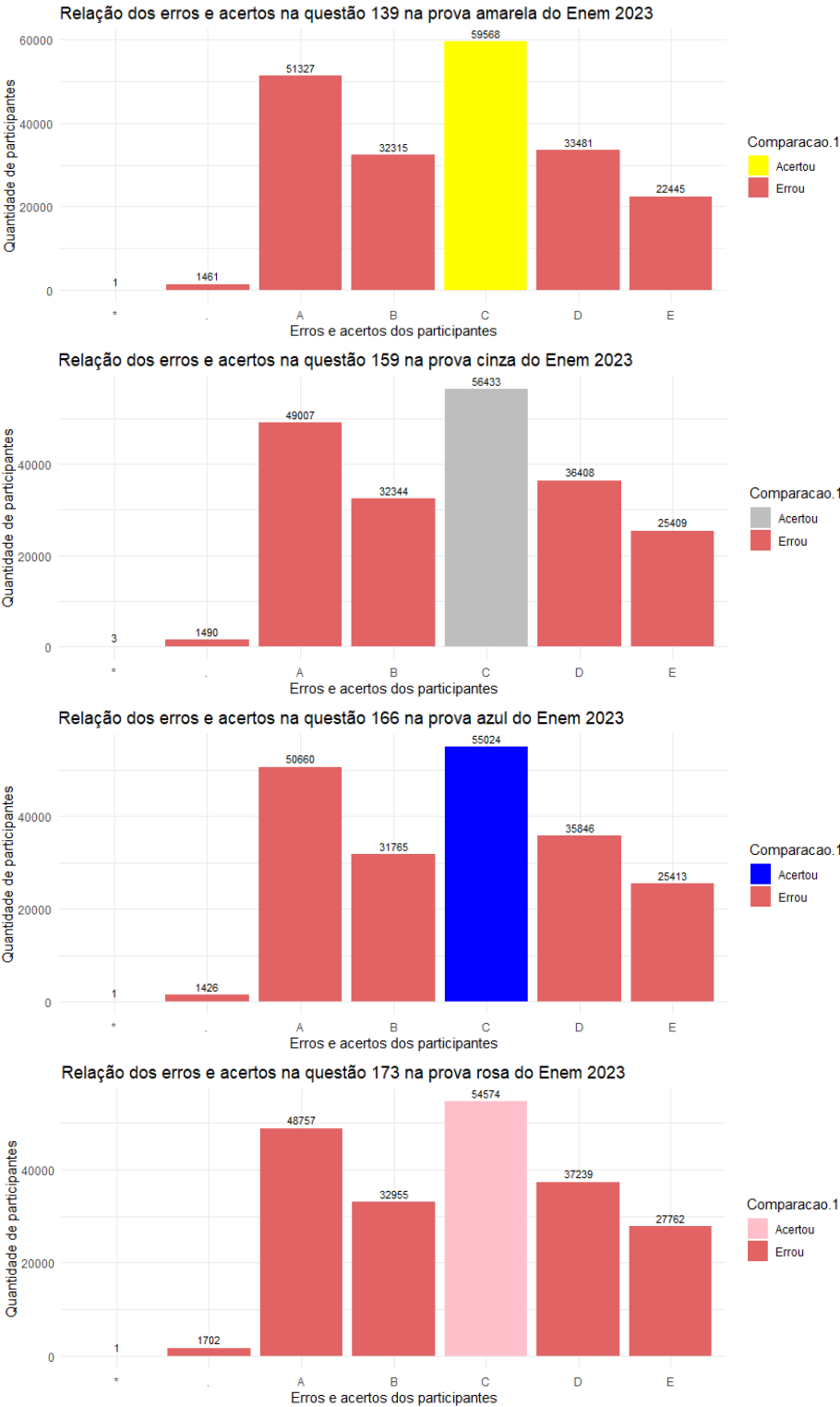
SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 2021. **Simpósio Brasileiro de Informática na Educação (SBIE)**. [S. l.]: SBC, 2021. p. 773–782. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/18105>. Acesso em: 5 jul. 2024.

SILVA, P. *et al.* Modelos de Regressão Aplicados a Predição do Desempenho Escolar de Estudantes do Ensino Fundamental. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, [s. l.], v. 30, n. 1, p. 1621, 2019.

SKOVSMOSE, O. Um convite à educação matemática crítica. Campinas: Papirus, 2015.

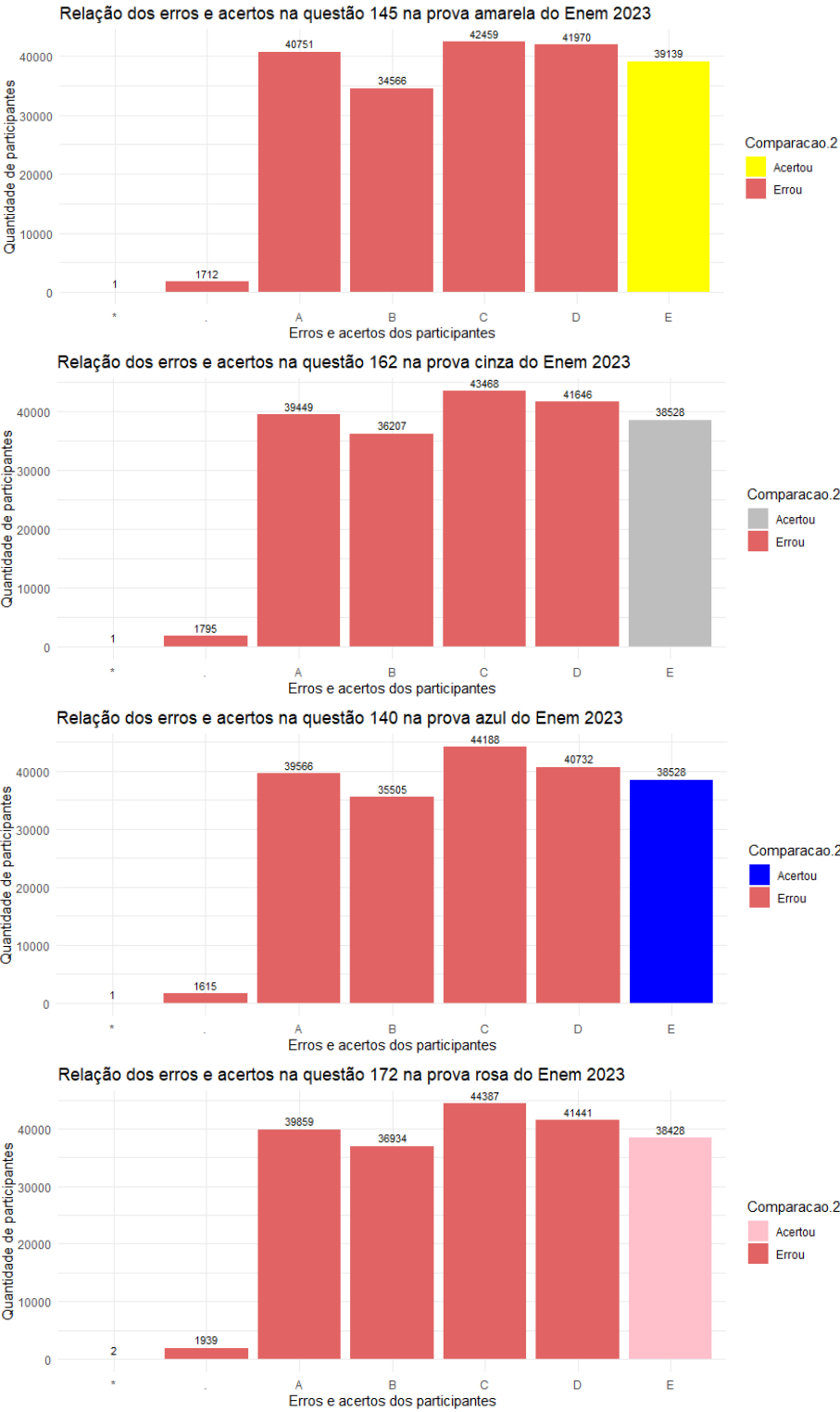
VERAS, W. S. *et al.* Exame Nacional do Ensino Médio (ENEM): Conteúdo matemático mais abordado e suas dificuldades. **Research, Society and Development**, [s. l.], v. 10, n. 6, p. e51610616041–e51610616041, 2021.

APÊNDICE A - GRÁFICOS REFERENTES A FIGURA 10



Fonte: Dados da Pesquisa (2024)

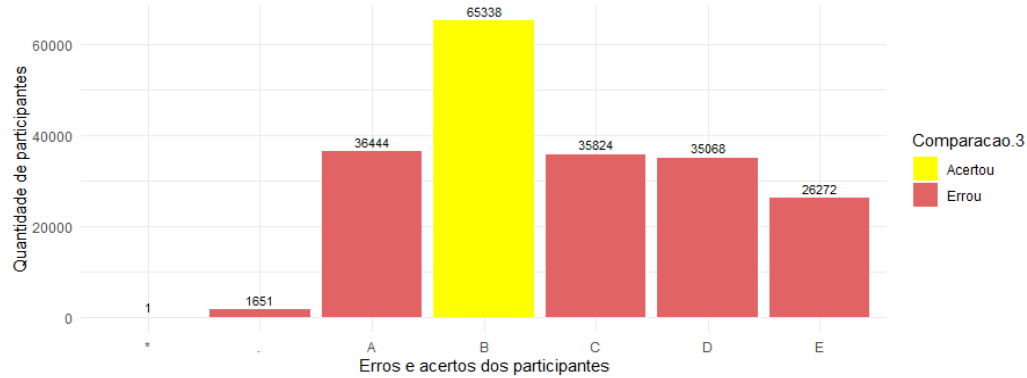
APÊNDICE B - GRÁFICOS REFERENTES A FIGURA 11



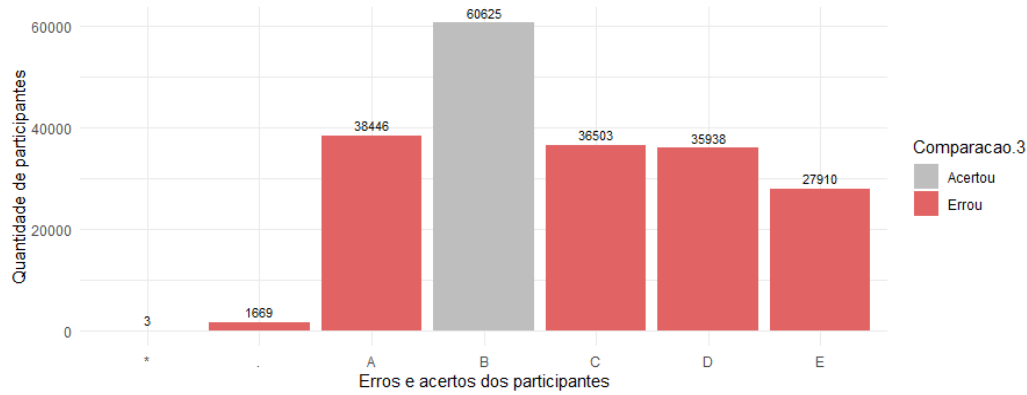
Fonte: Dados da Pesquisa (2024)

APÊNDICE C - GRÁFICOS REFERENTES A FIGURA 12

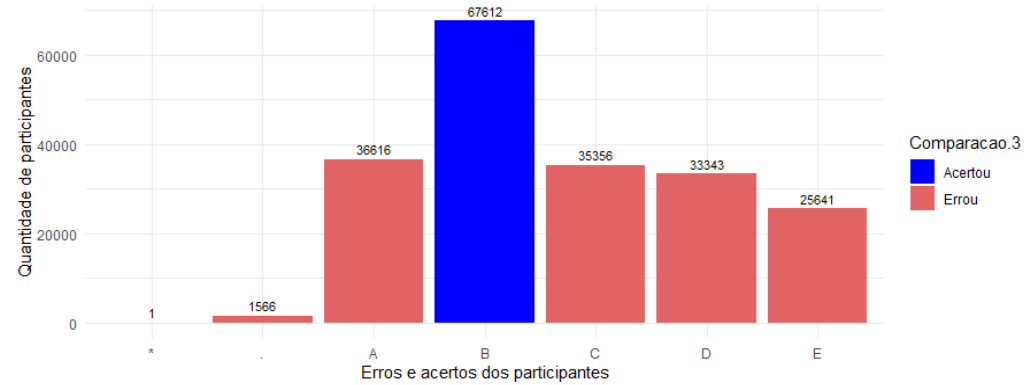
Relação dos erros e acertos na questão 154 na prova amarela do Enem 2023



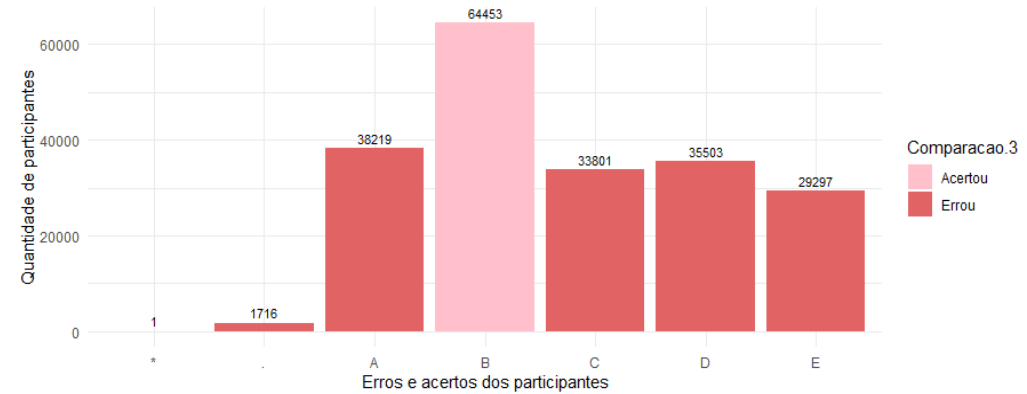
Relação dos erros e acertos na questão 171 na prova cinza do Enem 2023



Relação dos erros e acertos na questão 157 na prova azul do Enem 2023

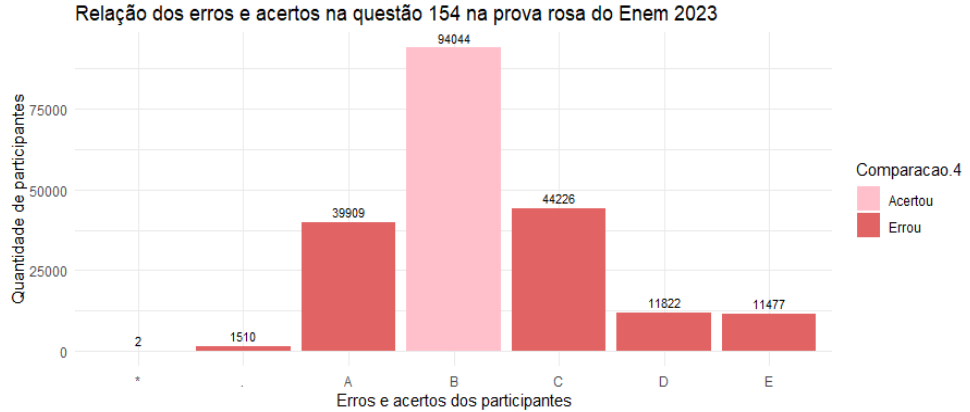
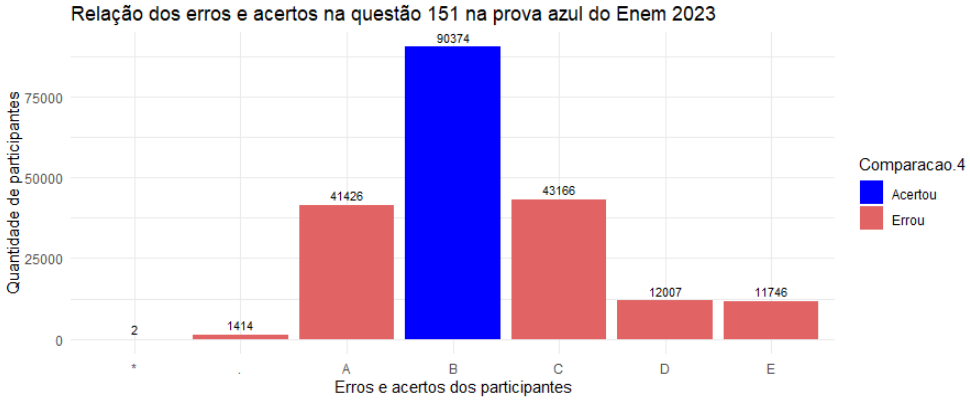
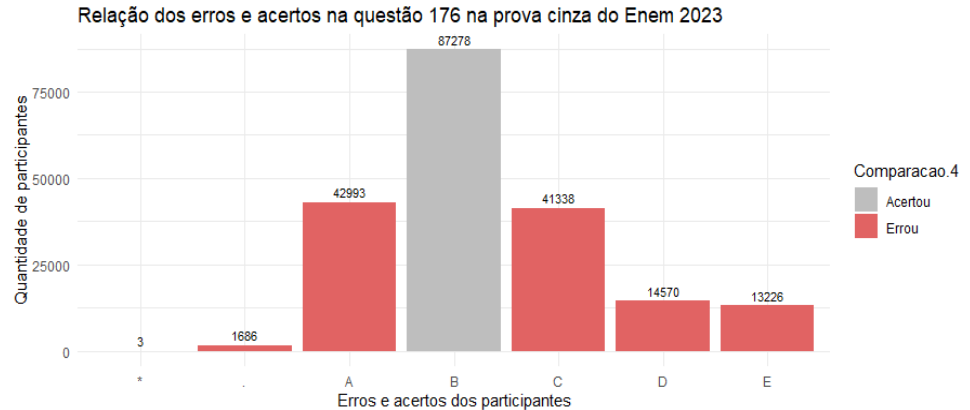
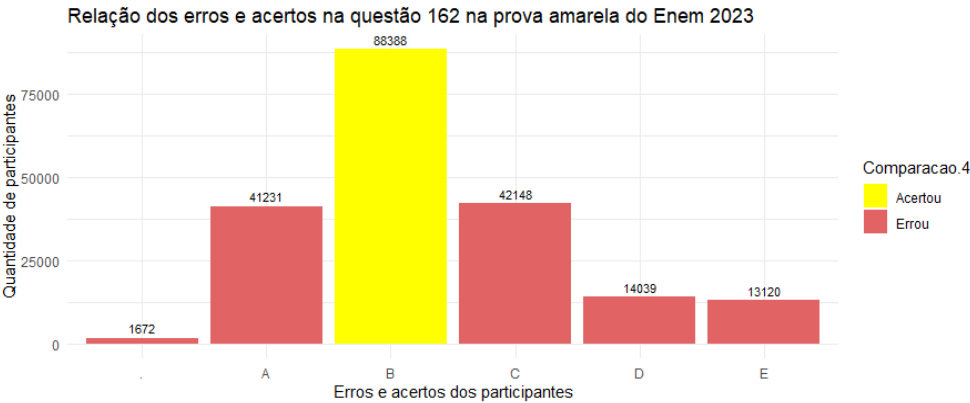


Relação dos erros e acertos na questão 159 na prova rosa do Enem 2023



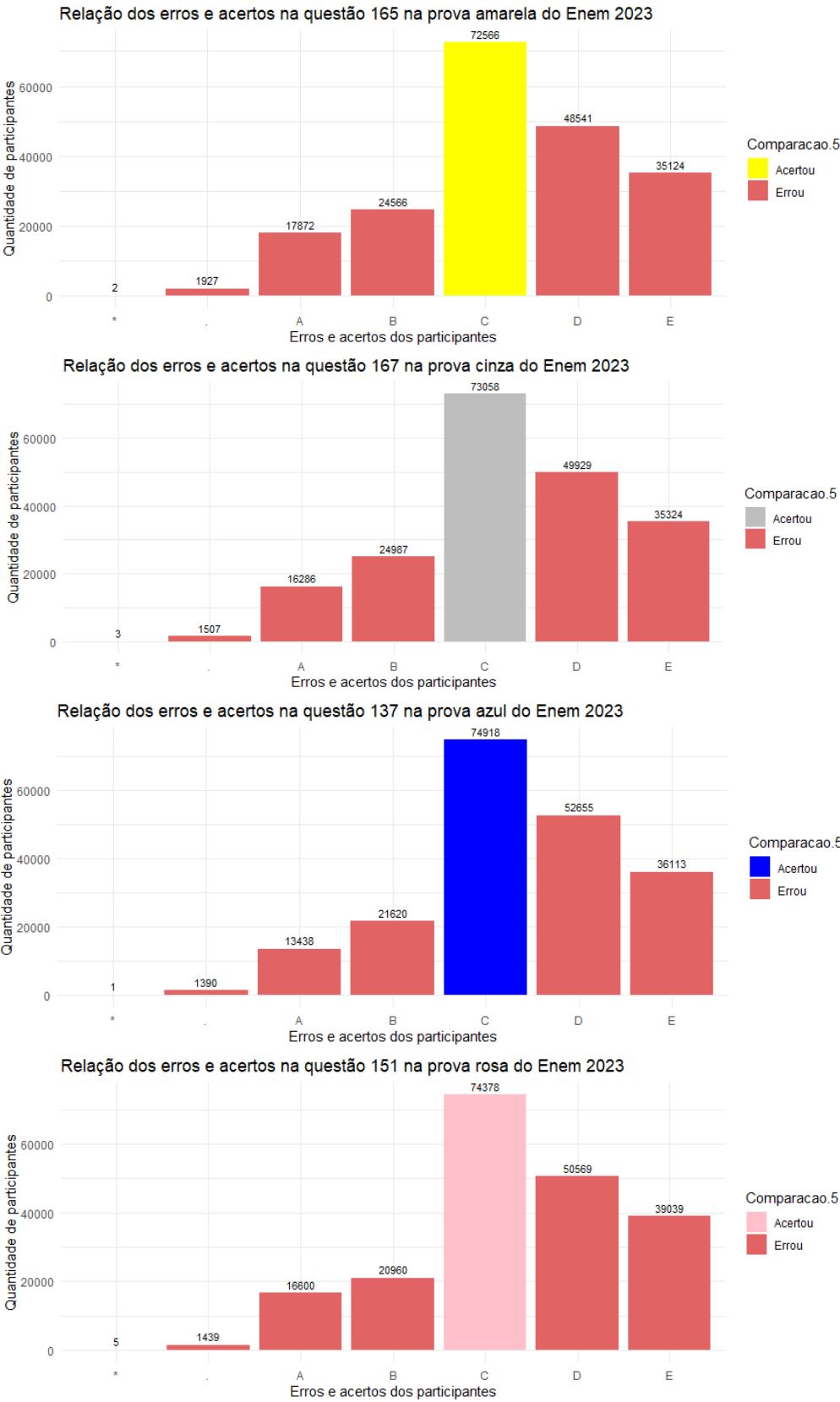
Fonte: Dados da Pesquisa (2024)

APÊNDICE D - GRÁFICOS REFERENTES A FIGURA 13



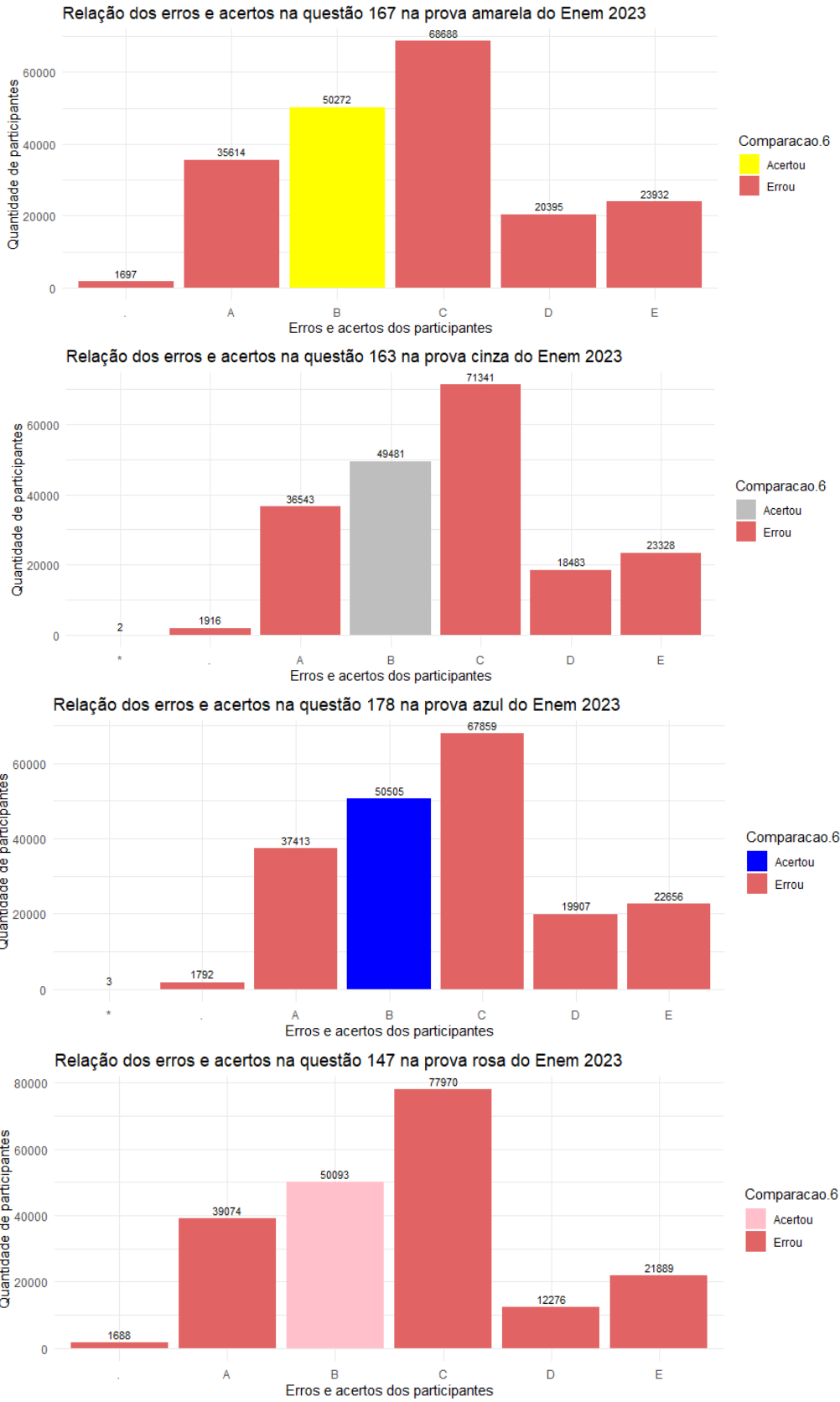
Fonte: Dados da Pesquisa (2024)

APÊNDICE E - GRÁFICOS REFERENTES A FIGURA 14



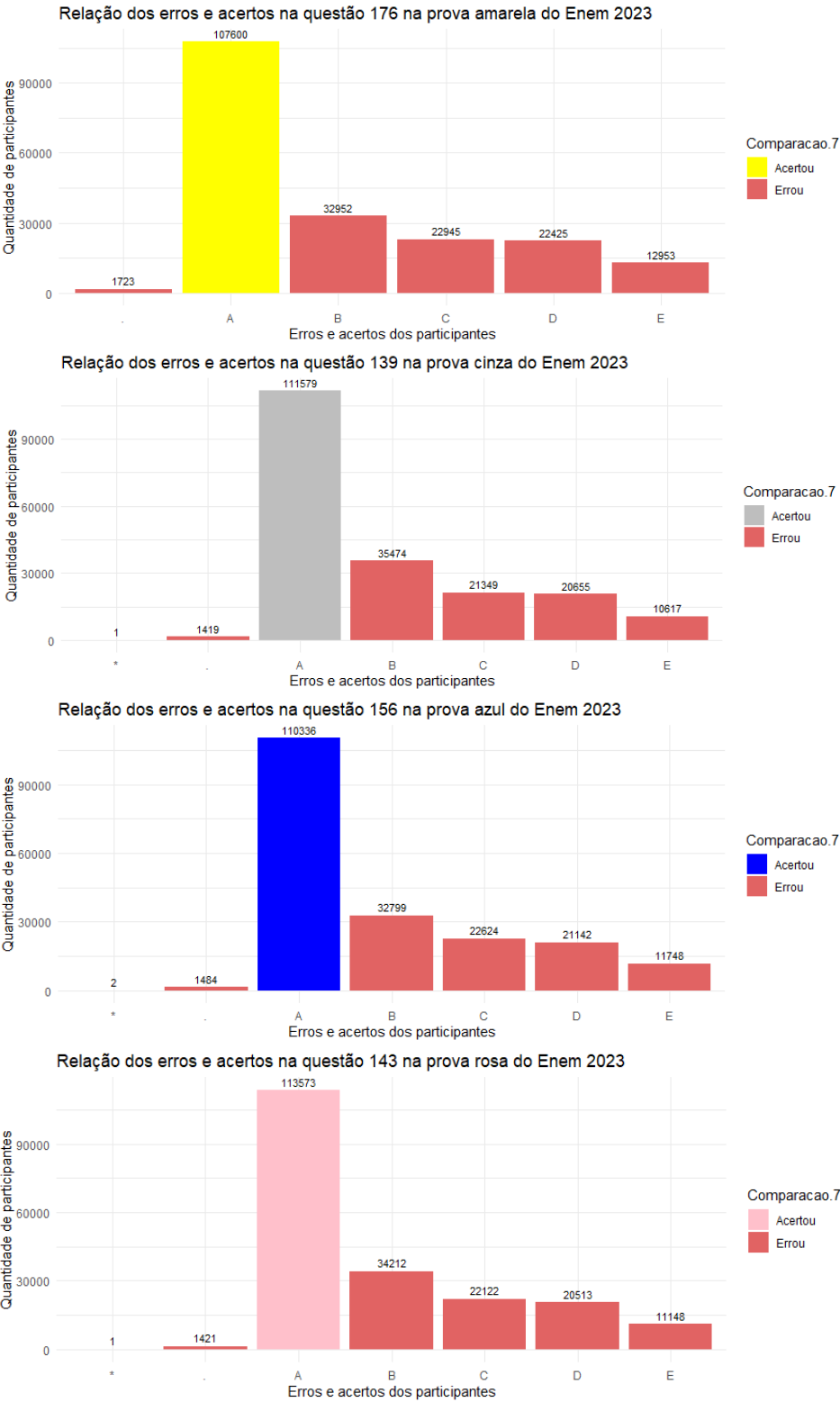
Fonte: Dados da Pesquisa (2024)

APÊNDICE F - GRÁFICOS REFERENTES A FIGURA 15



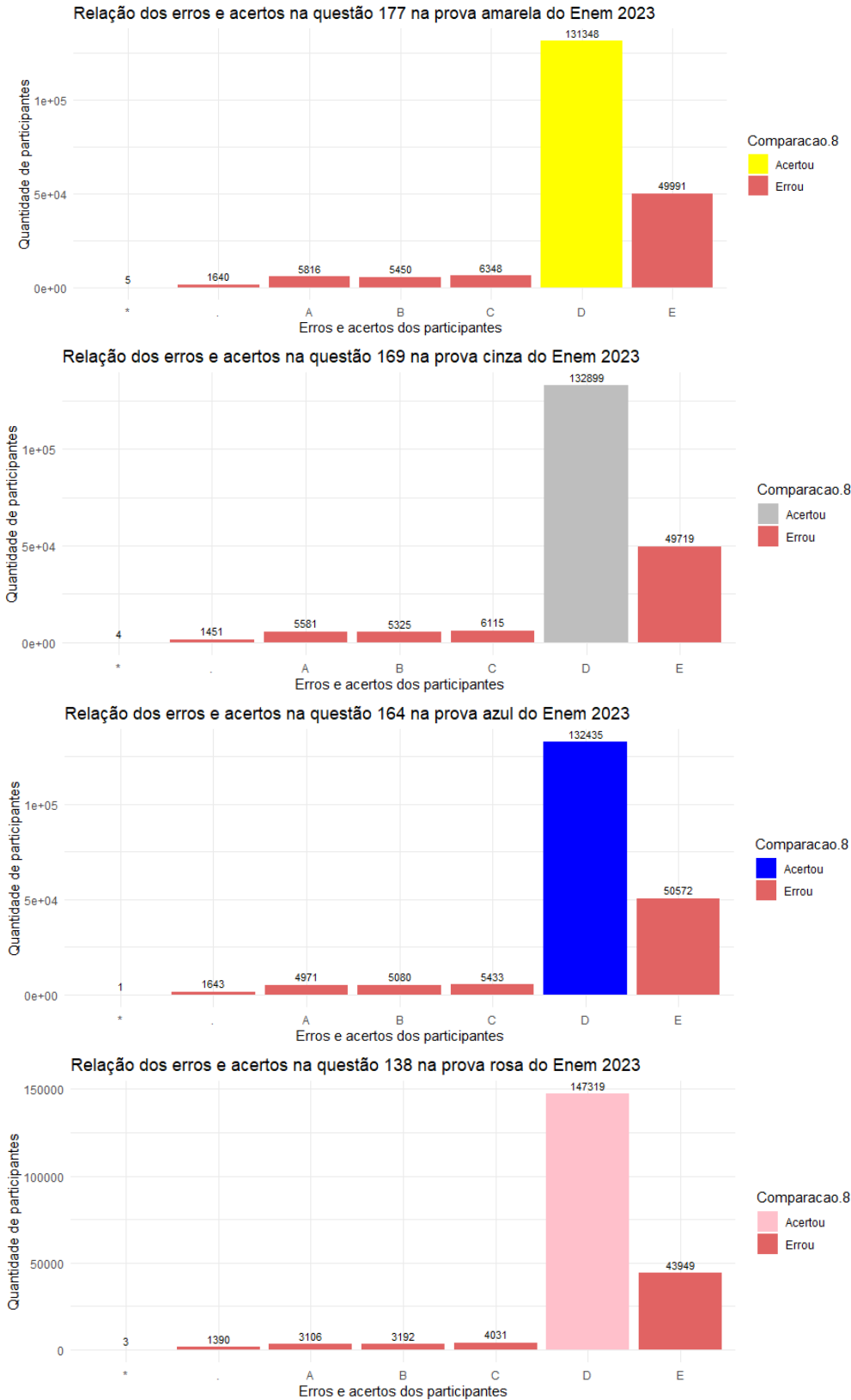
Fonte: Dados da Pesquisa (2024)

APÊNDICE G - GRÁFICOS REFERENTES A FIGURA 16



Fonte: Dados da Pesquisa (2024)

APÊNDICE H - GRÁFICOS REFERENTES A FIGURA 17



Fonte: Dados da Pesquisa (2024)