



Fundação Universidade Federal do ABC
Pró reitoria de pesquisa
Av. dos Estados, 5001, Santa Terezinha, Santo André/SP, CEP 09210-580
Bloco L, 3ºAndar, Fone (11) 3356-7617
iniciacao@ufabc.edu.br

Relatório do Primeiro Hackthon de Dados pela Universidade Federal do ABC (UFABC)

Nomes do alunos: Arthur Ortega (ra: 11202321805); Rodrigo Polino (ra: 11202320594); Kaique Bezerra (ra: 11202320362); João Marcelino (ra: 11202321602).

Título do projeto: Análise comparativa: Escola pública X Escola Privada no ENEM

São Paulo

12/06/2025



Resumo

Este projeto de dados relacionado à estatística desempenha papel central na análise de grandes volumes de dados, permitindo validar hipóteses e apoiar decisões baseadas em evidências. Este trabalho tem como objetivo avaliar, com base em dados do ENEM 2021/2022/2023, possíveis diferenças de desempenho entre estudantes de escolas públicas e privadas, com ênfase na área de Matemática. Foram aplicados métodos estatísticos, como o teste t de Student, para comparar médias entre grupos, conferindo rigor ao estudo. Os resultados sugerem desigualdades persistentes, com estudantes de escolas privadas geralmente obtendo melhores resultados. A pesquisa evidencia como a análise estatística é fundamental para a compreensão dos padrões educacionais e o panorama geral da maneira com que os dados se configuram para fornecer bases para o desenvolvimento de novas teorias e maneiras de abordar problemáticas.

Palavras-chave: análise de dados; estatística desempenho escolar; análise de dados ENEM; teste t de Student.

Abstract

This data project, grounded in statistical methods, plays a central role in analyzing large volumes of information, enabling hypothesis validation and supporting evidence-based decision making. The aim of this study is to evaluate, using ENEM 2021–2023 data, potential performance differences in Mathematics between students from public and private schools. Statistical methods, notably the Student's t-test, were applied to compare group means, thereby conferring rigor to the analysis. The results suggest persistent inequalities, with private school students generally achieving higher scores. This research demonstrates that statistical analysis is essential for understanding educational patterns and the overall data landscape, providing a basis for the development of new theories and approaches to problem solving.

Keywords: data analysis; educational performance statistics; ENEM data analysis; Student's t-test

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 5 |
| 2 | Objetivo Geral | 6 |
| 3 | Fundamentação teórica | 6 |
| 3.1 | Uma análise Estatística | 9 |
| 4 | Metodologia | 10 |
| 4.1 | Ferramentas e Métodos | 10 |
| 4.2 | Etapas da pesquisa | 10 |
| 5 | Resultados e Discussão | 11 |
| 5.1 | Resultados para 2023 | 15 |
| 5.2 | Resultados para 2022 | 15 |
| 5.3 | Resultados para 2021 | 17 |
| 5.4 | Análise geral | 18 |
| 5.5 | Possível Impacto da Pandemia | 18 |
| 6 | Conclusões | 19 |

1 Introdução

Para a primeira etapa desse projeto, é necessário pensar e entender um pouco mais sobre o que significa a ciência de dados, para o que ela serve e quais são as possibilidades de mercado que oferece:

A Ciência de Dados é uma área interdisciplinar que combina técnicas de estatística, computação e conhecimento de domínio para extrair conhecimento e gerar insights a partir de grandes volumes de dados. Em um mundo cada vez mais orientado por informações, a ciência de dados torna-se essencial para a tomada de decisões inteligentes baseadas em evidências, tanto no setor público quanto no privado.

O avanço da tecnologia e o crescimento exponencial da produção de dados — oriundos de redes sociais, sistemas empresariais, sensores, aplicativos, entre outros — impulsionaram a demanda por profissionais capazes de lidar com essas informações de forma estruturada e estratégica. A ciência de dados permite transformar esse grande volume de dados brutos em informações valiosas, por meio de análises estatísticas, modelagem preditiva, aprendizado de máquina e visualizações interpretáveis.

Entre suas aplicações práticas, destacam-se: previsões de comportamento de consumidores, detecção de fraudes, otimização de processos logísticos, personalização de serviços online, monitoramento de saúde pública, análises educacionais, entre muitas outras. Em contextos governamentais e educacionais, como no presente projeto, a ciência de dados pode ser aplicada para avaliar políticas públicas, identificar desigualdades e propor melhorias baseadas em evidências concretas.

A ciência de dados é uma área multidisciplinar que oferece diversas possibilidades profissionais, com aplicações práticas em setores públicos e privados.

Um **cientista de dados** envolve a construção de modelos estatísticos e algoritmos com o objetivo de gerar previsões, identificar padrões e propor soluções baseadas em evidências. Profissionais dessa área utilizam estatística, programação e conhecimento de negócio para extrair valor de grandes volume de dados.

Um **analista de dados** envolve a coleta, organização e interpretação de dados com o objetivo de responder perguntas específicas de negócio ou pesquisa. É uma área essencial para transformar dados brutos em informações úteis para a tomada de decisão. Exemplos de atuação: avaliar o desempenho de campanhas publicitárias, analisar notas escolares para identificar dificuldades de aprendizagem, verificar padrões de produtividade.

Já os **engenheiros de dados** são responsáveis por construir e manter a infraestrutura necessária para coleta, armazenamento e processamento eficiente de grandes volumes de dados. Exemplos de atuação: criar pipelines que unificam dados de diferentes fontes, estruturar bases de dados para análises rápidas, assegurar a qualidade e integridade dos

dados ao longo do tempo. Entre muitas outras possibilidades profissionais.

2 Objetivo Geral

Avaliar, com base nos dados do ENEM 2023, se há evidências estatísticas de que estudantes oriundos de escolas privadas obtêm, em média, notas mais altas nas provas — com foco principal na prova de Matemática — do que aqueles oriundos de escolas públicas. Além disso, inferir se isso pode estar associado a uma maior chance de ingresso em universidades federais.

Além dos objetivos específicos de pesquisa, este projeto também tem os objetivos característicos de um projeto de Iniciação Científica padrão:

- Obter resultados iniciais, e na medida inéditos, para os problemas estudados.
- Incentivar o gosto do aluno pela pesquisa e a criatividade para solução de problemas.
- Envolver o aluno em problemas de pesquisa na área de Ciência de Dados.
- Desenvolver a maturidade algorítmica (base) do aluno.
- Ampliar o conhecimento nas questões relacionadas ao desenvolvimento de pesquisa em dados, em geral, tais como escrita de artigos e relatórios, preparação e apresentação.
- Desenvolver no aluno a prática de busca, leitura e compreensão de artigos de dados através do estudo de artigos selecionados ad-hoc.
- Incentivar o aluno a trabalho em equipe com membros do grupo de pesquisa.

3 Fundamentação teórica

Para tornar este projeto mais didático e comprehensível, adota-se uma abordagem fictícia: assume-se o papel de analistas de dados contratados pela Presidência da República. A missão, dentro desse cenário, foi realizar um estudo sobre os dados do ENEM 2021/2022/2023. O objetivo dessa análise é identificar padrões de desempenho entre os estudantes, principalmente em Matemática, com foco na comparação entre alunos de escolas públicas e privadas.

O Exame Nacional do Ensino Médio (ENEM) é uma das principais portas de entrada para o ensino superior no Brasil. Dentre os diversos fatores que podem influenciar o

desempenho dos candidatos, destaca-se o tipo de escola frequentada no ensino médio, geralmente classificada como pública ou privada. Este trabalho tem como objetivo investigar, sob uma perspectiva estatística, se estudantes de escolas privadas apresentam desempenho significativamente superior, especialmente em Matemática, em comparação com estudantes de escolas públicas.

Ao trabalhar com estatística, sempre podem ser elaboradas hipóteses a serem comprovadas pelo estudo em questão, neste caso, uma análise de dados referentes à base do ENEM. Para a comparação entre os dois grupos (pública vs. privada), formulam-se as seguintes hipóteses:

H_0 (**hipótese nula**): A média das notas dos estudantes de escolas públicas é igual à média das notas dos estudantes de escolas privadas.

$$H_0 : \mu_{\text{pública}} = \mu_{\text{privada}}$$

H_1 (**hipótese alternativa**): A média das notas dos estudantes de escolas privadas é maior que a dos estudantes de escolas públicas.

$$H_1 : \mu_{\text{pública}} > \mu_{\text{privada}}$$

Hipótese Específica para Matemática:

H_0 : Não há diferença significativa nas médias de nota em Matemática entre os dois grupos.

H_1 : A média da nota em Matemática dos alunos de escola privada é maior que a de escola pública.

Hipótese Específica para Linguagens:

H_0 : Não há diferença significativa nas médias de nota em Linguagens entre os dois grupos.

H_1 : A média da nota em Linguagens dos alunos de escola privada é maior que a de escola pública.

Na estatística, quando quer-se testar uma ideia ou uma suposição sobre uma população a partir de uma amostra, precisa-se de um modelo lógico que permita avaliar se os dados realmente sustentam essa ideia. Esse modelo é construído a partir de duas hipóteses complementares.

1. Hipótese Nula(H_0):

É a suposição de que não existe diferença ou efeito significativo entre os grupos analisados. Ela funciona como um "ponto de partida" neutro — ou seja, a ideia que aceitamos até que tenhamos evidência suficiente para rejeitá-la.

2. Hipótese Alternativa(H_1):

É a suposição que propõe uma diferença real ou efeito estatisticamente significativo. Ela é o que tentamos demonstrar com os dados, mas só será aceita se houver evidência suficiente contra a hipótese nula.

Variância e Desvio-Padrão

A variância amostral s^2 quantifica a dispersão dos valores de uma variável em torno de sua média e é definida por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

onde x_i representa cada observação, \bar{x} é a média amostral e n o número de observações. No contexto dos dados do ENEM, a variância reflete o grau de heterogeneidade das notas dos estudantes dentro de cada grupo (público vs. privado).

O desvio-padrão s é a raiz quadrada da variância,

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

e fornece uma medida de dispersão na mesma unidade das observações (pontos de nota). Em nossa análise, o desvio-padrão das médias gerais permite avaliar como as notas se afastam, em média, do desempenho esperado, servindo de base para a comparação da consistência dos resultados entre os grupos.

Teste t de Student

O teste t de Student, introduzido por Gosset (1908) sob o pseudônimo “Student”, é um procedimento estatístico para comparar médias de duas populações quando as variâncias são desconhecidas e amostras de tamanho finito. Dados dois grupos independentes com tamanhos amostrais n_1 e n_2 , médias \bar{X}_1 e \bar{X}_2 e variâncias amostrais s_1^2 e s_2^2 , a estatística de teste é definida por

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

em que

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

e os graus de liberdade associados são

$$\nu = n_1 + n_2 - 2.$$

Este teste baseia-se em três pressupostos principais:

- **Independência das amostras:** as observações em um grupo não influenciam as do outro.
- **Normalidade:** os valores em cada grupo seguem, aproximadamente, distribuição normal (para amostras grandes, aplica-se o Teorema do Limite Central).
- **Homoscedasticidade:** as variâncias populacionais dos dois grupos são iguais. Caso esta hipótese não se sustente, utiliza-se a correção de Welch, que ajusta os graus de liberdade.

3.1 Uma análise Estatística

O fato de gerarmos hipóteses, torna o estudo mais qualificado, quando entende-se que seguindo esse padrão as decisões baseadas em achismos são evitadas. é necessário quantificar a incerteza de forma objetiva, através do valor-p e dos testes estatísticos. Há a garantia de que conclusões só sejam tiradas quando a evidência dos dados for realmente forte e assim há validade científica ao estudo, permitindo replicação e análise crítica por outros.

O projeto tem por objetivo descobrir se alunos de escolas privadas têm desempenho superior no ENEM em relação aos de escolas públicas.

A mera comparação das médias sugere um desempenho superior dos alunos de escolas privadas; entretanto, esse procedimento não é suficiente para uma conclusão estatisticamente válida.

Por isso, define-se:

- Hipótese Nula(H_0) :

Não há diferença significativa entre as notas de alunos de escolas públicas e privadas.

- Hipótese Alternativa(H_1):

Alunos de escolas privadas têm, em média, notas maiores do que as escolas públicas.

A decisão estatística fundamenta-se no valor-*p* resultante da comparação do valor calculado de *t* com a distribuição *t* de Student de ν graus de liberdade. Um valor-*p* abaixo do nível de significância escolhido (por exemplo, $\alpha = 0,05$) indica evidência

estatística suficiente para rejeitar a hipótese nula

$$H_0: \mu_1 = \mu_2.$$

4 Metodologia

4.1 Ferramentas e Métodos

- Coleta de dados: Os dados apresentados neste trabalho foram obtidos a partir de etapas rigorosamente definidas, de modo que o código desenvolvido foi elaborado de maneira coesa e coerente, atendendo aos critérios metodológicos estabelecidos para a pesquisa. Para assegurar a assertividade e a confiabilidade do método adotado, optou-se pela obtenção dos dados por planilhas fornecidas pelos proponentes do projeto, ou bases de dados disponíveis em instituições governamentais de caráter federal e da Universidade Federal do ABC e seus órgãos competentes.(PROGRAD, n.d.) (INEP, 2025)

- **Ferramentas Utilizadas**

Para apoiar a análise e assegurar a reproduzibilidade, utilizou-se o **Jupyter Notebook** como ambiente interativo de desenvolvimento e registro dos experimentos. A manipulação e o tratamento dos dados foram realizados com a biblioteca **pandas** em **Python** (versão 3.12), fundamentadas nos princípios estatísticos de amostragem, estimação pontual e teste de hipóteses. Para a geração de gráficos e a representação das distribuições, adotou-se o módulo **matplotlib.pyplot**, complementado pela biblioteca **seaborn**, cuja malha de grid aprimora significativamente a legibilidade e facilita a interpretação dos resultados. A elaboração e a formatação do manuscrito científico foram conduzidas no **Overleaf**, utilizando a linguagem **LaTeX**, garantindo padronização tipográfica, gestão automatizada de referências bibliográficas e alta qualidade de layout, todos de fácil reproduzibilidade, dado na documentação de código:

Acesso ao arquivo no Google Drive

4.2 Etapas da pesquisa

1. **Tratar os dados:** Obtenção das tabelas dos indivíduos que participaram do ENEM, fornecidas pelo INEP,e utilização dos códigos em python no ambiente Jupyter, com o auxílio das biblioteca pandas, a fim de eliminar as linhas que

continham células vazias, e filtrar os dados pelas colunas que seriam utilizadas na análise estatística, como todas as notas de cada aluno por eixo dos dias da prova, e o tipo da escola.

2. Definir Quantas pessoas têm na amostra, a fim de obter quantas delas fizeram escola pública ou privada, para assim tirar a média de nota de todas as pessoas envolvidas;
3. Tirar a média das notas das pessoas de escola pública e privada, e calcular as variâncias de cada ano;
 - Com as variâncias amostrais s_1^2 e s_2^2 calculadas, define-se

$$k = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)}.$$

- * Se $k < 2$, assume-se homocedasticidade e aplica-se o *teste t de Student* para variâncias iguais, podendo considerar essa uma distribuição normal, em que a análise pode ser feita com acurácia:

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \quad t = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad \nu = n_1 + n_2 - 2.$$

- * Se $k \geq 2$, adota-se o *teste de Welch*:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad \nu \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}.$$

Em ambos os casos, o *valor-p* é obtido pela cauda dupla:

$$p = 2 \left[1 - F_\nu(|t|) \right],$$

onde F_ν é a CDF da distribuição t de Student com ν graus de liberdade.

4. Observar se há diferença significativa e se caso houver, refuta-se a hipótese nula e avaliamos a hipótese alternativa;
5. Nesse caso, a hipótese alternativa significa que escolas privadas possuem notas melhores que as públicas.

5 Resultados e Discussão

Ao longo dos anos, os resultados do exame têm revelado importantes desigualdades no desempenho dos estudantes, principalmente nas áreas de Matemática e Lingua-

| NU_INSCRIÇÃO | TP_COR_RACA | NU_ANO | TP_ESCOLA | NO_MUNICIPIO_PROVA | NU_NOTA_MT |
|--------------|--------------|--------|-----------|--------------------|---------------------------------|
| 3 | 210060214087 | parda | 2023 | Escola pública | Fortaleza 466.7 |
| 4 | 210059980948 | parda | 2023 | Escola pública | Quixadá 338.3 |
| 11 | 210060801595 | parda | 2023 | Escola pública | Ubiratã 691.9 |
| 16 | 210058061534 | branca | 2023 | Escola pública | Belém 437.0 |
| 41 | 210059855121 | branca | 2023 | Escola pública | São Paulo 628.1 |
| ... | ... | ... | ... | ... | ... |
| 3933936 | 210061965954 | branca | 2023 | Escola pública | Santo Antônio da Patrulha 494.5 |
| 3933940 | 210061965958 | branca | 2023 | Escola pública | São Leopoldo 666.2 |
| 3933944 | 210061965962 | branca | 2023 | Privada | Capão da Canoa 444.9 |
| 3933948 | 210061959674 | branca | 2023 | Escola pública | São Leopoldo 530.5 |
| 3933953 | 210061932304 | branca | 2023 | Privada | Bragança Paulista 569.0 |

1039069 rows × 10 columns

Figura 1: Exemplo de tabela de referência utilizada(2023)

gens.

Segundo os dados do ENEM 2023, a média da prova de Linguagens, Códigos e suas Tecnologias foi a mais baixa entre todas as áreas avaliadas, evidenciando sérias dificuldades dos estudantes em interpretação de texto, compreensão de gêneros discursivos e domínio da norma padrão da língua portuguesa. Embora Matemática tenha apresentado uma média maior, ela também figura entre as disciplinas com maior número de acertos baixos, demonstrando uma defasagem no raciocínio lógico e na resolução de problemas básicos.

Esses resultados refletem desafios estruturais no sistema educacional brasileiro, como a falta de base sólida nas etapas anteriores de ensino, desigualdade de acesso a materiais de qualidade, e a distância entre o conteúdo trabalhado em sala de aula e as competências exigidas na prova. Além disso, tais defasagens têm impacto direto nas chances de ingresso dos estudantes em instituições públicas de ensino superior, como a Universidade Federal do ABC (UFABC), cujos cursos mais concorridos exigem desempenho elevado justamente nessas duas áreas.

Interpretação do critério de homocedasticidade No presente estudo, obteve-se

$$k = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)} < 2,$$

o que indica que as variâncias amostrais dos grupos público e privado são suficientemente próximas para admitirmos homocedasticidade.

Dessa forma, as principais implicações são:

- **Uso do teste t de Student clássico:** podemos aplicar o teste para variâncias iguais, aproveitando a estimativa combinada de variância

$$S_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2},$$

garantindo maior eficiência estatística.

- **Graus de liberdade plenos:** o número total de graus de liberdade é dado por $\nu = n_1 + n_2 - 2$, maximizando o poder do teste.
- **Interpretação do valor-p:** ao comparar a estatística

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

com a distribuição t de Student de ν graus de liberdade, o valor- p resultante reflete com precisão a probabilidade de obtenção de diferença tão extrema sob a hipótese nula de médias iguais.

Em síntese, $k < 2$ legitima o emprego do teste t de Student clássico, proporcionando estimativas mais estáveis da variância e maior sensibilidade para detectar diferenças reais entre as médias dos grupos analisados.

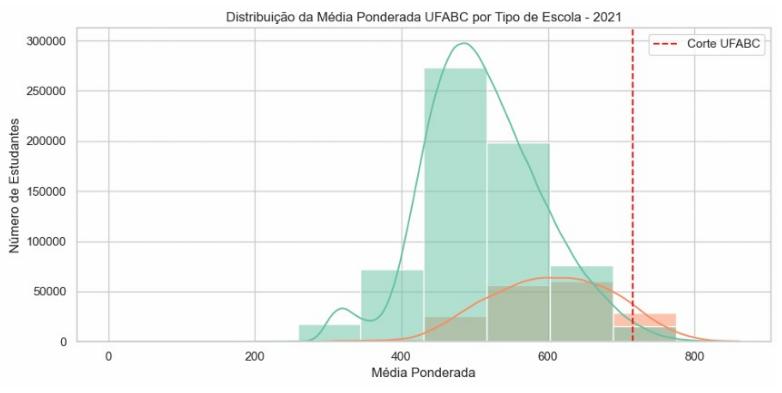


Figura 2: Distribuição média ponderada por tipo de escola(2021)

Para continuar nosso estudo, vamos construir uma estimativa hipotética realista, com base nos padrões de notas do ENEM e das notas de corte do SiSU para faculdades federais — como a UFABC, UFRJ, UFPE, etc.

Estimativa Hipotética: Notas mínimas para ter boa chance

supõe-se um cenário realista, considerando um curso competitivo como o Bacharelado em Ciência e Tecnologia da UFABC ou Engenharia em outras federais:

Chega-se a um valor de nota em matemática de 720 a 800 que a pessoa deve tirar para obter boas chances de entrar em um curso o Bacharelado em Ciência e

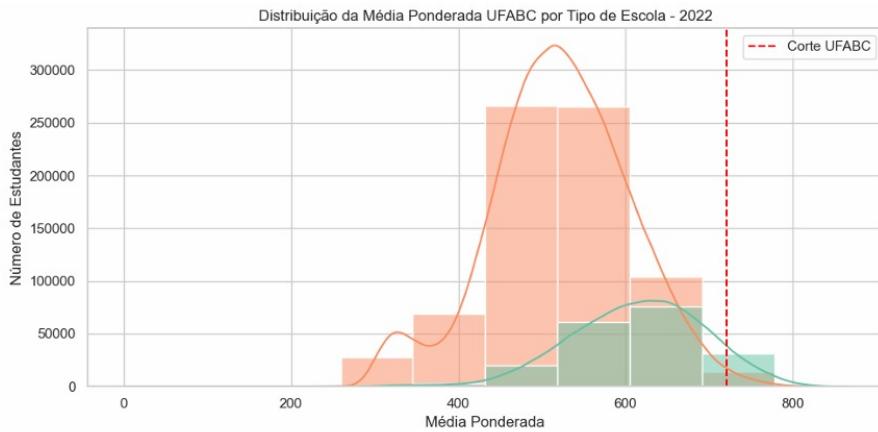


Figura 3: Distribuição média ponderada por tipo de escola(2022)

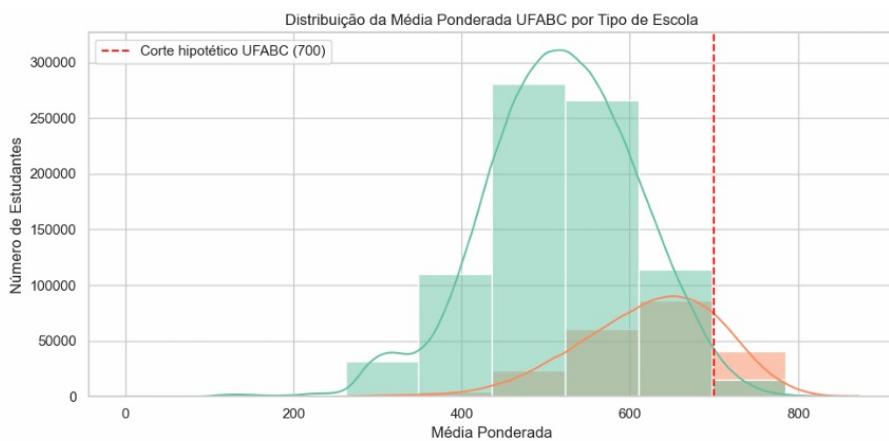


Figura 4: Distribuição média ponderada por tipo de escola(2023)

Tecnologia da UFABC ou Engenharia em outras federais.

Já para as notas em Linguagens foi estipulada uma nota entre 650 a 720.

Para essa etapa, procedeu-se ao tratamento dos dados, incluindo limpeza e remoção de registros incompletos e outliers. Em seguida, foi quantificado o tamanho da amostra total e classificado cada participante segundo o tipo de escola de origem (pública ou privada). Para cada grupo, determinou-se o número de estudantes que alcançaram a nota de corte estipulada, tanto no segmento privado quanto no público. Por fim, compararam-se essas quantidades com o total de estudantes de escolas privadas, de modo a avaliar a proporção de aprovados e a representatividade relativa de cada subgrupo no desempenho geral da amostra, levando em consideração os eixos de matemática e linguagens.

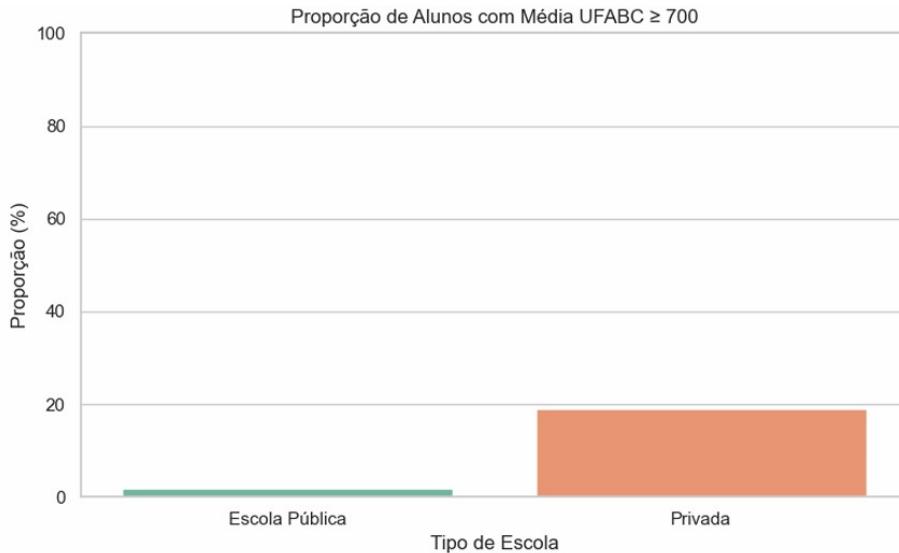


Figura 5: Proporção de alunos com média UFABC maior ou igual a 700

5.1 Resultados para 2023

No ano de 2023, a amostra incluiu $N = 1.039.069$ estudantes, dos quais 820.141 (78,9%) pertencem a escolas públicas e 218.928 (21,1%) a escolas privadas. As médias gerais foram

$$\bar{X}_{\text{públicos}} = 515,97. \quad \bar{X}_{\text{privados}} = 616,17.$$

com variâncias

$$s_{\text{públicos}}^2 = 7.650,14. \quad s_{\text{privados}}^2 = 6.752,81.$$

Considerando alunos com nota de Matemática superior a 720 pontos, verificou-se que apenas 3,2% (26.602) dos estudantes de escolas públicas e 24,0% (53.952) dos de escolas privadas atingiram esse patamar. Para notas de Linguagens acima de 650 pontos, observaram-se 0,7% (6.499) dos estudantes de escolas públicas e 6,4% (14.035) dos de escolas privadas.

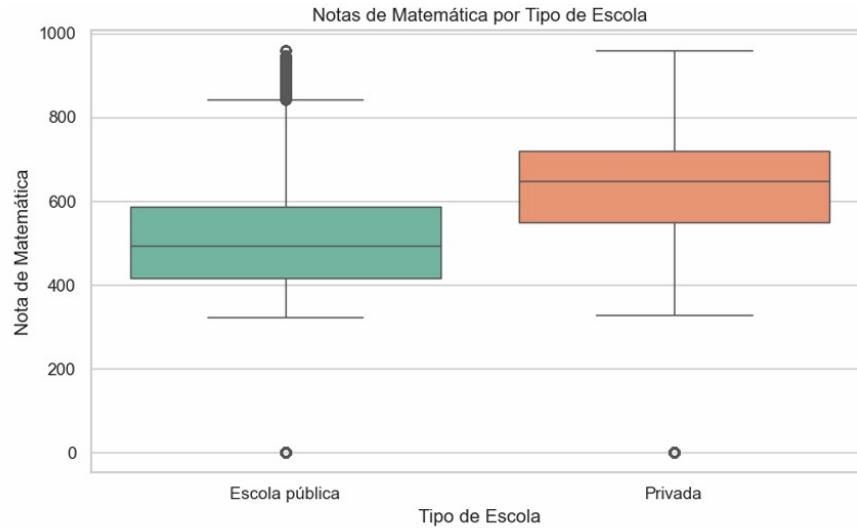
Considerando a média dos três anos abordados, identifica-se que a média para alcançar uma aprovação na UFABC é de 717,70, e que em 2023 identifica-se que o total de pessoas que teriam nota suficiente para ingressar na UFABC, de escolas privadas, seria de 28.151(12,0%), já para escolas públicas 7.709(0,9%).

5.2 Resultados para 2022

No ano de 2022, a análise abrangeu uma amostra de $N = 940.354$ estudantes, dos quais 746.292 (79,3%) eram oriundos de escolas públicas e 194.062 (21,7%) de

Tabela 1: Resumo didático dos resultados de 2023

| Indicador | Escola Pública | Escola Privada |
|---------------------------|-----------------|-----------------|
| Número de estudantes | 820 141 (78,9%) | 218 928 (21,1%) |
| Média geral (\bar{X}) | 515,97 | 616,17 |
| Variância (s^2) | 7.650,14 | 6.752,81 |
| Média $\geq 717,70$ | 7 709 (0,9%) | 28 151 (12,0%) |
| Matemática > 720 | 26 602 (3,2%) | 53 952 (24,0%) |
| Linguagens > 650 | 6 499 (0,7%) | 14 035 (6,4%) |

**Figura 6:** Nota de matemática das escolas

instituições privadas. As médias gerais encontradas foram

$$\bar{X}_{\text{públicas}} = 519,27. \quad \bar{X}_{\text{privadas}} = 608,47.$$

com variâncias estimadas em

$$s_{\text{públicas}}^2 = 6\,610,06. \quad s_{\text{privadas}}^2 = 5\,938,25.$$

Quando se consideram somente os estudantes com nota de Matemática acima de 720 pontos, observou-se que 2,9% ($n = 22.385$) dos participantes de escolas públicas e 18,0% ($n = 36.635$) dos de escolas privadas atingiram esse patamar. De forma semelhante, para notas de Linguagens superiores a 650 pontos, registraram-se 0,7% ($n = 5.467$) de estudantes públicos e 5,4% ($n = 10.628$) de privados.

Em 2022 identifica-se que o total de pessoas que teriam média suficiente para ingressar na UFABC, de escolas privadas, seria de 18.390(9,4%), já para escolas públicas 6.378(0,8%).

Tabela 2: Resumo didático dos resultados de 2022

| Indicador | Escola Pública | Escola Privada |
|---------------------------|-----------------------|-----------------------|
| Número de estudantes | 746 292 (79,3%) | 194 062 (21,7%) |
| Média geral (\bar{X}) | 519,27 | 608,47 |
| Variância (s^2) | 6 610,06 | 5 938,25 |
| Média $\geq 717,70$ | 6 378 (0,8%) | 18 390 (9,4%) |
| Matemática > 720 | 22 385 (2,9%) | 36 635 (18,0%) |
| Linguagens > 650 | 5 467 (0,7%) | 10 628 (5,4%) |

5.3 Resultados para 2021

No ano de 2021, a análise abrangeu uma amostra de $N = 829.640$ estudantes, dos quais 653.335 (78,7%) eram oriundos de escolas públicas e 176.305 (22,3%) de instituições privadas. As médias gerais encontradas foram

$$\bar{X}_{\text{públicas}} = 509,37. \quad \bar{X}_{\text{privadas}} = 598,92.$$

com variâncias estimadas em

$$s^2_{\text{públicas}} = 6 358,43. \quad s^2_{\text{privadas}} = 6 857,25.$$

Quando se consideram somente os estudantes com nota de Matemática acima de 720 pontos, observou-se que 2,5% ($n = 16.874$) dos participantes de escolas públicas e 16,4% ($n = 29.884$) dos de escolas privadas atingiram esse patamar. De forma semelhante, para notas de Linguagens superiores a 650 pontos, registraram-se 0,7% ($n = 4694$) de estudantes públicos e 5,4% ($n = 9403$) de privados.

Em 2021 identifica-se que o total de pessoas que teriam média suficiente para ingressar na UFABC, de escolas privadas, seria de 16.859(9,5%), já para escolas públicas 6.680(1%).

Tabela 3: Resumo didático dos resultados de 2021

| Indicador | Escola Pública | Escola Privada |
|---------------------------|-----------------------|-----------------------|
| Número de estudantes | 653 335 (78,7%) | 176 305 (22,3%) |
| Média geral (\bar{X}) | 509,37 | 598,92 |
| Variância (s^2) | 6 358,43 | 6 857,25 |
| Média $\geq 717,70$ | 6 680 (1,0%) | 16 859 (9,5%) |
| Matemática > 720 | 16 874 (2,5%) | 29 884 (16,4%) |
| Linguagens > 650 | 4 694 (0,7%) | 9 403 (5,4%) |

5.4 Análise geral

Baseado numa análise em que os três anos são considerados 2.219.768 alunos de escola pública realizaram a prova, e somente 20.768 deles alcançaram uma média que os possibilitem a aprovação na Universidade Federal do ABC, o que representa 0,9% e 589.295 foram alunos de escola privada e 63.400 deles atingiram a média suficiente, o que representa 10,0% da amostra. Analogamente, somente, 65.861 estudantes de escola pública atingiram uma nota de matemática superior a 720, representando 2,9%, e 120.471, dos quais estudaram em escola privada, atingiram uma nota superior à essa, representando 20,0% da nota.

Para linguagens somente 16.660 de escola pública atingiram uma nota superior à 650, representando 0,75%, e 34.066 de escola privada atingiram uma nota superior a essa, representando 5,7%

Tabela 4: Desempenho agregado de estudantes públicos e privados (2021–2023)

| Indicador | Escola Pública | Escola Privada |
|-------------------------------|----------------|----------------|
| Total de estudantes | 2219768 | 589295 |
| Aprovados UFABC (média geral) | 20768 (0,9%) | 63400 (10,0%) |
| Matemática > 720 pontos | 65861 (2,9%) | 120471 (20,0%) |
| Linguagens > 650 pontos | 16660 (0,75%) | 34066 (5,7%) |

5.5 Possível Impacto da Pandemia

Os resultados de 2021, ainda fortemente influenciados pelas restrições ao ensino presencial durante a pandemia de COVID-19, apresentaram médias gerais em Matemática inferiores às observadas em 2022 e 2023 (públicas: 509,37; privadas: 598,92). A variância mais elevada em 2021 sugere maior heterogeneidade de desempenho, possivelmente decorrente de condições desiguais de acesso a recursos de ensino remoto. Em 2022, com a retomada parcial das atividades presenciais, verificou-se aumento das médias gerais (públicas: 519,27; privadas: 608,47) e leve redução da variabilidade, indicando recuperação acadêmica inicial. Contudo, o alargamento da diferença média em 2023 (públicas: 515,97; privadas: 616,17) pode refletir efeitos residuais da pandemia e consolidação de desigualdades estruturais. Tais padrões reforçam a hipótese de que o ensino remoto e as interrupções prolongadas no calendário escolar impactaram negativamente o aprendizado, sobretudo em contextos de menor suporte tecnológico e pedagógico.

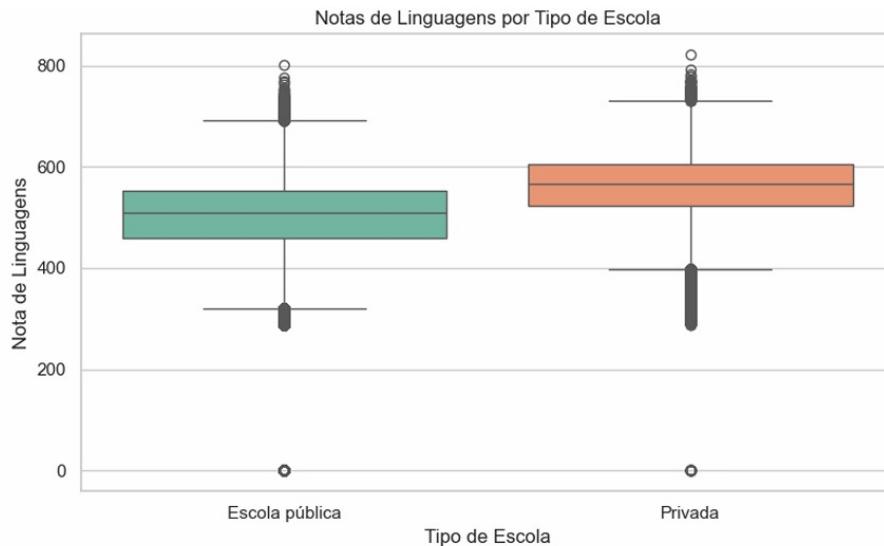


Figura 7: Nota de linguagens das escolas

6 Conclusões

A análise dos microdados do ENEM 2021–2023 revelou diferenças estatisticamente significativas no desempenho em Matemática entre estudantes de escolas públicas e privadas, confirmadas pelo teste t de Student ($\alpha = 0,05$) sob a hipótese de homocedasticidade ($k < 2$). Alunos de instituições privadas apresentaram médias consistentemente superiores, maior proporção acima dos cortes competitivos (média $\geq 717,70$; Matemática > 720 ; Linguagens > 650) e variabilidade ligeiramente menor em comparação aos pares de escolas públicas.

O ano de 2021, marcado pela pandemia de COVID-19, registrou maior dispersão (s^2 elevado) e médias reduzidas, refletindo os impactos do ensino remoto desigual. Em 2022 observou-se recuperação parcial das médias e redução da variabilidade, enquanto em 2023 consolidou-se a vantagem dos estudantes de escolas privadas, sugerindo efeitos residuais das interrupções escolares e desigualdades estruturais pré-existentes.

Esses achados ressaltam a importância de políticas públicas voltadas à equiparação de oportunidades, por meio de investimentos em infraestrutura tecnológica, capacitação docente e programas de reforço para alunos de escolas públicas.

Perspectivas Futuras

- **Controle multivariado:** incorporar variáveis socioeconômicas, regionais e de infraestrutura escolar em modelos de regressão para identificar fatores explicativos adicionais.

- **Análise longitudinal:** estender o estudo a coortes posteriores do ENEM para avaliar tendências de longo prazo e o impacto de intervenções educacionais.
- **Avaliação de programas:** investigar a eficácia de políticas e programas de apoio implementados no ensino médio nos últimos anos.
- **Modelagem preditiva:** desenvolver algoritmos de aprendizado de máquina para prever desempenho e direcionar ações de suporte individualizado.

Dessa forma, este trabalho não só mapeia desigualdades educacionais como também estabelece uma base científica para o desenvolvimento de novas teorias e estratégias de intervenção que promovam maior equidade e qualidade no ensino médio brasileiro.

Referências Bibliográficas

INEP. 2025. *Microdados enem.*

PROGRAD. n.d.. *Histórico de ingresso na ufabc.*