

**UNIVERSIDADE DE SÃO PAULO**  
**Instituto de Ciências Matemáticas e de Computação**

**Estudo Observacional do Desempenho de Estudantes  
de Escolas Públicas no ENEM de 2023**

Relatório do trabalho prático da disciplina SME0807 - Técnicas de  
Amostragem cursada no 1º semestre de 2024

**Ada Maris Pereira Mário** *12725432*

**Emanuel Victor da Silva Favorato** *12558151*

**Gabriela Scaranello Teixeira de Barros** *9284334*

**Karen Neves Perciliano** *12559357*

**Tayane Leandro Guerrero** *11838203*

**Thalita Sousa Pereira** *12559187*

**SÃO CARLOS**

**2024**

# 1 Introdução

O Exame Nacional do Ensino Médio (ENEM) é uma das avaliações mais significativas no cenário educacional do Brasil. Além de avaliar o desempenho do estudante ao fim da escolaridade básica, o ENEM é uma das portas de entrada para as universidades públicas e privadas do país [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2024]. Desse modo, compreender os fatores que influenciam o desempenho dos alunos no exame é crucial para a formulação de políticas educacionais mais eficazes e justas [Dutra et al., 2023].

Neste projeto, temos como objetivo geral analisar as médias ensino básico público e o quanto influencia nas notas gerais, e específicos nas notas parciais dos alunos no ENEM. Para tanto, foram formuladas as seguintes hipóteses centrais e hipóteses secundárias:

- **Hipótese Central:**

$$\begin{cases} H_0 : & \text{A nota média geral das escolas públicas é no máximo 520.} \\ H_1 : & \text{A nota média geral das escolas públicas é maior que 520.} \end{cases}$$

- **Hipóteses Secundárias:**

- **Redação:**

$$\begin{cases} H_0 : & \text{A nota média das escolas públicas na redação é no máximo 650.} \\ H_1 : & \text{A nota média das escolas públicas na redação é maior que 650.} \end{cases}$$

- **Linguagens:**

$$\begin{cases} H_0 : & \text{A nota média das escolas públicas em linguagens é no máximo 510.} \\ H_1 : & \text{A nota média das escolas públicas em linguagens é maior que 510.} \end{cases}$$

- **Matemática:**

$$\begin{cases} H_0 : & \text{A nota média das escolas públicas em matemática é no máximo 500.} \\ H_1 : & \text{A nota média das escolas públicas em matemática é maior que 500.} \end{cases}$$

Os resultados desta pesquisa podem fornecer ideias para a formulação de políticas educacionais, ajudando a identificar áreas de ensinos que necessitam de maior atenção e investimento.

## 2 Revisão Bibliográfica

Em um estudo de estatística aplicada na área de Educação, pesquisadores da Comissão Permanente de Concursos (COMPERVE) do Rio Grande do Norte e do Instituto Federal do Rio Grande do Norte (IFRN) utilizaram dados do ENEM de 2014 para descrever procedimentos computacionais e aspectos teóricos sobre seleção de amostra, baseados no desempenho dos alunos [de Melo Pereira et al., 2019]. A população estudada foram de escolas privadas do estado do Rio Grande do Norte, e a amostra foi obtida por meio de amostragem estratificada e agrupamento.

O estudo teve como objetivo generalizar os resultados para toda a população, utilizando amostragem probabilística. A população da amostra foi composta por 3.118 candidatos de 45 escolas privadas de Natal que ingressaram na graduação da UFRN via SISU em 2015. A amostragem probabilística permitiu a obtenção da amostra de forma aleatória, conforme descrito por [Marôco, 2018]. A amostragem estratificada foi empregada no primeiro estágio, onde a população foi dividida em subgrupos homogêneos, com seleção final dos elementos pertencentes a todos os estratos [de Melo Pereira et al., 2019].

Jackson Costa, em sua pesquisa, analisou disparidades entre as regiões Nordeste e Sudeste quanto ao impacto de variáveis socioeconômicas no desempenho dos candidatos nas provas do ENEM [Costa, 2024]. O estudo incluiu modelagem estatística dos resultados de 2019 e 2022, investigando as diferenças antes e após a pandemia nas duas regiões. Os modelos estatísticos aplicados foram Mínimos Quadrados Ordinários (MQO) e Propensity Score Matching (PSM).

As variáveis principais para o desempenho no ENEM incluem escolaridade dos pais, faixa de renda familiar, idade, raça, sexo, dependência administrativa das escolas, renda familiar, alta escolaridade da mãe e acesso à internet [Costa, 2024]. A análise revelou uma forte discrepância nos resultados entre Nordeste e Sudeste, intensificada pela pandemia.

## 3 Metodologia

Para esta pesquisa, foi utilizado o conjunto de microdados do ENEM 2023 [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2024], com 179885

entradas e 76 atributos, dentre eles, dados de inscrição, gabarito, localidade, dentre outros. Como o interesse é na análise do desempenho de estudantes de escolas públicas, foram eliminadas as colunas com informações sobre código diversos, vetores de respostas, notas de cada competência da redação, e algumas perguntas do questionário socioeconômico que não serão exploradas neste trabalho.

Este estudo é de natureza explicativa, buscando explicar as médias das notas gerais e específicas (redação, linguagens e matemática) dos alunos da rede pública que participaram do ENEM em 2023 (unidades elementares). O objetivo é entender os fatores que influenciam essas notas.

A amostra foi selecionada utilizando a técnica de amostragem aleatória simples sem reposição. Esse método foi escolhido por sua simplicidade e eficiência na obtenção de uma amostra representativa da população de interesse, garantindo que todos os alunos da rede pública que realizaram o ENEM em 2023 tivessem a mesma probabilidade de serem selecionados.

Inicialmente, foi retirada uma amostra piloto com  $n = 10$  para estimar a variabilidade das notas nas disciplinas de interesse (redação, linguagens e matemática). Isso foi feito para calcular o tamanho necessário da amostra final, considerando um erro máximo de 50 pontos e 95% de confiança, encontrando-se o tamanho ideal de  $n = 2125$ .

As variáveis principais de interesse são as médias das notas nas seguintes categorias:

- **Média geral:** Representada pela variável `nota_media`.
- **Redação:** Representada pela variável `NU_NOTA_REDACAO`.
- **Linguagens:** Representada pela variável `NU_NOTA_LINGUAGENS`.
- **Matemática:** Representada pela variável `NU_NOTA_MATEMATICA`.

Essas variáveis foram escolhidas por serem indicadores do desempenho dos alunos no ENEM, refletindo habilidades específicas e compreensão em áreas fundamentais do currículo escolar.

A seguir, para cada variável, foram encontrados os valores necessários (1) (2) para o cálculo das estatísticas de teste (3).

$$\bar{x} = \frac{\sum_{i \in s} X_i}{n} = \frac{t(\mathbf{s})}{n} = \hat{\mu} \quad (1)$$

$$\text{var}(\bar{x}) = (1 - f) \left( \frac{s^2}{n} \right) \quad (2)$$

$$Z = \frac{\bar{x} - \mu_0}{\sqrt{\text{var}(\bar{x})}} = \frac{\bar{x} - \mu_0}{\sqrt{(1 - f) \frac{s^2}{n}}} \sim N(0, 1). \quad (3)$$

Vale ressaltar que a amostragem e demais cálculos foram feitos com suporte de ferramentas computacionais da linguagem de programação *Python*, utilizando-se de bibliotecas e módulos numéricos e de ciência de dados como *pandas*, *numpy* e *scipy.stats*.

Por fim, baseado nesses valores e ao nível de significância de  $\alpha = 5\%$ , as hipóteses foram avaliadas. Os resultados encontrados serão discutidos na próxima seção.

## 4 Resultados

Feita a coleta da amostra, realizou-se uma pequena análise exploratória dos dados amostrais e populacionais. A seguir, apresentam-se os histogramas das notas para cada categoria.

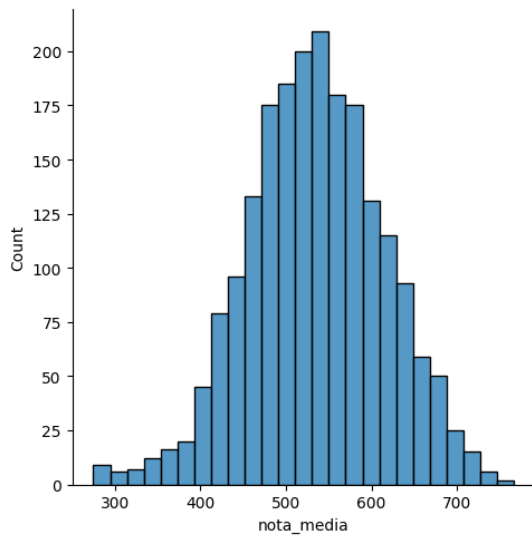


Figura 1: Histograma das notas gerais

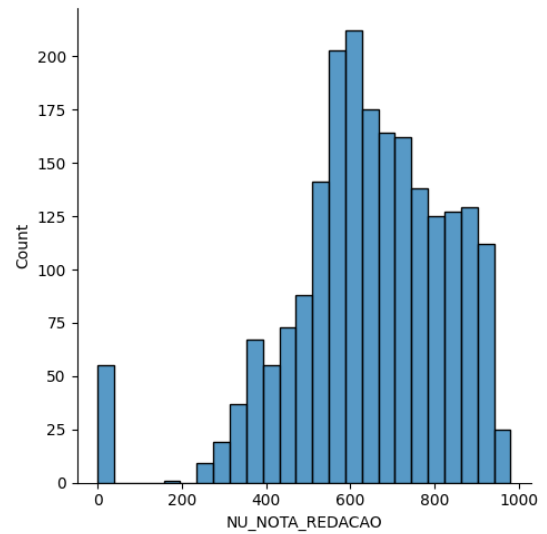


Figura 2: Histograma das notas em redação

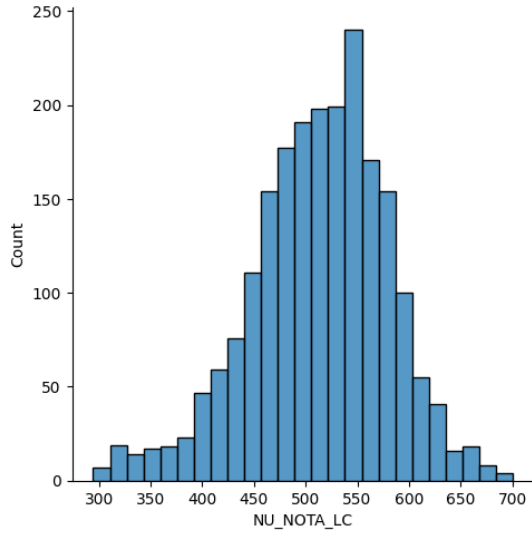


Figura 3: Histograma das notas em linguagens

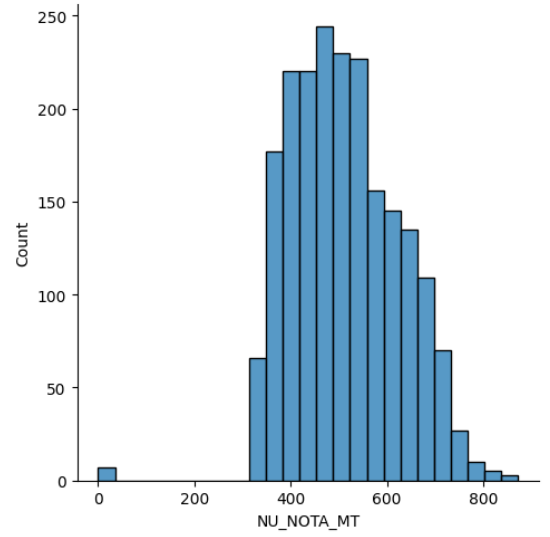


Figura 4: Histograma das notas em matemática

Analisando as distribuições empíricas dos dados pelos histogramas e realizando os cálculos necessários, encontraram-se, para a população, as médias de  $\mu_{\text{geral}} = 534.42$ ,  $\mu_{\text{redação}} = 661.74$ ,  $\mu_{\text{linguagens}} = 512.39$ ,  $\mu_{\text{matemática}} = 511.83$ . Da mesma forma, encontraram-se médias amostrais de  $\bar{x}_{\text{geral}} = 534.42$ ,  $\bar{x}_{\text{redação}} = 661.74$ ,  $\bar{x}_{\text{linguagens}} = 512.39$ ,  $\bar{x}_{\text{matemática}} = 511.83$ .

Desse modo, considerando as modas e médias obtiveram-se as estatísticas de teste (3) para as hipóteses anteriormente mencionadas e suas respectivas regiões críticas para o nível  $\alpha = 5\%$  de significância, encontrando-se  $RC_{\text{geral}} = \{\bar{x} > 522.64\}$ ,  $RC_{\text{redação}} = \{\bar{x} > 655.54\}$ ,  $RC_{\text{linguagens}} = \{\bar{x} > 512.22\}$ ,  $RC_{\text{matemática}} = \{\bar{x} > 503.71\}$ . Logo, rejeitou-se  $H_0$  para todos os casos.

Portanto, ao nível de  $\alpha = 5\%$  de significância, os dados da amostra aleatória simples sem reposição evidenciam que as notas médias gerais, em redação, em linguagens e em matemática dos alunos da rede pública no ENEM são, respectivamente, maiores que 520, 650, 510, 500, superando nossas expectativas.

## 5 Considerações Finais

Neste estudo, analisou-se o impacto do ensino básico público no desempenho dos alunos no Exame Nacional do Ensino Médio (ENEM) de 2023. Foram formuladas hipóteses centrais e secundárias focadas nas notas médias gerais e específicas (redação, linguagens e matemática) dos estudantes de escolas públicas. Através da utilização de técnicas de amostragem aleatória simples e análise estatística, obtiveram-se resultados significativos que contribuem para a compreensão da influência do ensino público no desempenho acadêmico.

Os resultados evidenciaram que as notas médias gerais e específicas dos alunos da rede pública superaram os valores esperados nas hipóteses nulas para todas as categorias analisadas. As estatísticas de teste, ao nível de significância de 5%, indicaram que as notas médias são significativamente maiores que 520 para a média geral, 650 para a redação, 510 para linguagens e 500 para matemática. Esse desempenho superior ao esperado das escolas públicas no ENEM é muito importante por várias razões, tais como: ser um indicador de qualidade da escola pública e servir também como motivação para a formulação de políticas educacionais que buscam melhorar a qualidade do ensino público e reduzir as disparidades de desempenho entre diferentes tipos de escolas.

Adicionalmente, a análise pode-se abrir portas para analisar outros fatores que influenciam no desempenho dos alunos no ENEM, podendo guiar iniciativas específicas para áreas que necessitam de maior atenção e investimentos, como o aprimoramento de recursos didáticos e treinamento de professores. O estudo contribui para o debate sobre a equidade educacional no Brasil e reforça a necessidade de intervenções direcionadas para elevar o padrão de ensino nas escolas públicas.

## Referências

- Jackson Antônio Costa. Análise regional das diferenças nas notas do enem entre as escolas públicas e privadas no pós-pandemia. B.S. thesis, Brasil, 2024.
- Fernando Antonio de Melo Pereira, Raquel Alves Basílio, Maria Jessiane Alexandre da Silva, and Roberto Junior J Oliveira. Plano de amostragem: um estudo de caso utilizando dados do enem. *Brazilian Applied Science Review*, 3(2):1036–1046, 2019.
- Janderson Ferreira Dutra, João Batista Firmino Júnior, and Damires Yluska de Souza Fernandes. Fatores que podem interferir no desempenho de estudantes no enem: uma revisão sistemática da literatura. *Revista Brasileira de Informática na Educação*, 31(1): 323–351, 2023.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. ENEM, 2024. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>. Acesso em: 18 jun. 2024.
- João Marôco. *Análise Estatística com o SPSS Statistics.: 7ª edição*. ReportNumber, Lda, 2018.