

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**

**DEPARTAMENTO DE ECONOMIA**

**BACHARELADO EM CIÊNCIAS ECONÔMICAS**

**JACKSON ANTÔNIO COSTA**

**Análise regional das diferenças nas notas do ENEM entre as escolas  
públicas e privadas no pós-pandemia**

Recife - PE

2023

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO**  
**DEPARTAMENTO DE ECONOMIA**  
**BACHARELADO EM CIÊNCIAS ECONÔMICAS**

**JACKSON ANTÔNIO COSTA**

**Análise regional das diferenças nas notas do ENEM entre as escolas  
públicas e privadas no pós-pandemia**

Trabalho de Conclusão de Curso apresentado pelo aluno **Jackson Antônio Costa** ao Curso de Ciências Econômicas da Universidade Federal Rural de Pernambuco - UFRPE, como pré-requisito parcial para obtenção do grau de Bacharel em Ciências Econômicas, sob a orientação da professora **Dra. Poema Isis Andrade de Souza**.

Recife - PE

2023

Dados Internacionais de Catalogação na Publicação  
Universidade Federal Rural de Pernambuco  
Sistema Integrado de Bibliotecas  
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

---

C837a Costa, Jackson Antônio  
Análise regional das diferenças nas notas do ENEM entre as escolas públicas e privadas no pós-pandemia / Jackson  
Antônio Costa. - 2024.  
62 f. : il.

Orientadora: Poema Isis Andrade de Souza.  
Inclui referências e apêndice(s).

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco, Bacharelado em  
Ciências Econômicas, Recife, 2024.

1. ENEM. 2. PSM. 3. MQO. 4. Sudeste. 5. Nordeste. I. Souza, Poema Isis Andrade de, orient. II. Título

CDD 330

---

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO**  
**DEPARTAMENTO DE ECONOMIA**  
**BACHARELADO EM CIÊNCIAS ECONÔMICAS**

Monografia apresentada como requisito necessário para a obtenção do título de Bacharel  
em Ciências Econômicas.

**Análise regional das diferenças nas notas do ENEM entre as escolas  
públicas e privadas no pós-pandemia**

**JACKSON ANTÔNIO COSTA**

BANCA EXAMINADORA

Trabalho de Conclusão de Curso aprovado com nota \_\_\_\_ apresentado em \_\_\_\_/\_\_\_\_/\_\_\_\_

—

---

Orientadora: Prof<sup>ª</sup> Dra. Poema Isis Andrade de Souza

---

1º Examinador: Prof. Dr. André de Souza Melo

---

2º Examinador: Prof<sup>ª</sup> Dra. Gisleia Benini Duarte

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO**  
**DEPARTAMENTO DE ECONOMIA**  
**BACHARELADO EM CIÊNCIAS ECONÔMICAS**

**AGRADECIMENTOS**

Agradeço, em primeiro lugar, a Deus por me permitir concluir mais uma etapa importante em minha vida. Aos meus pais, agradeço pela formação moral e ética que me proporcionaram, além de todo o apoio emocional e financeiro que tornou minha graduação uma realidade. À minha noiva, expresso minha gratidão por todo o apoio emocional nos momentos tão fáceis quanto difíceis, os quais foram fundamentais para a minha saúde mental ao longo dos quatro anos de estudo. Quero estender meus agradecimentos a todos os professores do Bacharelado em Ciências Econômicas da UFRPE, que compartilharam seus conhecimentos comigo. À minha orientadora, Professora Poema, sou muito grato por seu suporte e pelos *insights* valiosos, que foram de grande importância para a conclusão deste trabalho. Aos meus colegas de estudo, com os quais tive a oportunidade de aprender e adquirir diferentes perspectivas, agradeço pela colaboração. À coordenação do curso de Economia, meu reconhecimento por sua dedicação. Por fim, agradeço a todos aqueles que, de uma forma ou de outra, contribuíram para a realização deste objetivo.

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO****PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO****DEPARTAMENTO DE ECONOMIA****BACHARELADO EM CIÊNCIAS ECONÔMICAS****RESUMO**

O ENEM (Exame Nacional do Ensino Médio), criado em 1998, é o principal meio de acesso ao ensino superior no Brasil, proporcionando acesso a programas como o SISU (Sistema de Seleção Unificado), PROUNI (Programa Universidade para Todos) e FIES (Fundo de Financiamento Estudantil). Portanto, este estudo visa compreender os fatores que influenciam as notas do ENEM dos candidatos, analisando duas bases de dados (2019 e 2022) para examinar as variações nas notas no período antes e depois da pandemia da COVID-19 e, conseqüentemente, as disparidades entre as regiões Sudeste e Nordeste do Brasil. No primeiro momento, foi estimado o modelo de Mínimos Quadrados Ordinários (MQO) com a proposição de analisar a correlação das notas obtidas pelo candidato com a escola de origem, utilizando variáveis socioeconômicas como variáveis de controle. É importante considerar o viés de origem da escola e do perfil socioeconômico dos estudantes, visto que alunos de escolas privadas tendem a ter melhores notas. Para mitigar esse viés, utilizou-se o método de *Propensity Score Matching* (PSM), que é uma ferramenta valiosa quando o intuito ao controle de vieses de seleção. Os resultados encontrados revelam uma discrepância entre os anos de 2019 e 2022, demonstrando uma intensificação das distorções regionais na educação, com melhores rendimentos de nota para os alunos das escolas privadas no pós-pandemia da COVID-19 na região Sudeste.

**Palavras-chave:** ENEM; PSM; MQO; Sudeste; Nordeste; COVID-19; variáveis socioeconômicas.

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO**  
**DEPARTAMENTO DE ECONOMIA**  
**BACHARELADO EM CIÊNCIAS ECONÔMICAS**

**ABSTRACT**

The ENEM (National High School Exam), created in 1998, is the main means of access to higher education in Brazil, providing access to programs such as SISU (Unified Selection System), PROUNI (University for All Program) and FIES (Student Financing Fund). Therefore, this study aims to understand the factors that influence the ENEM scores of candidates, analyzing two databases (2019 and 2022) to examine the variations in scores in the period before and after the COVID-19 pandemic and, consequently, the disparities between the Southeast and Northeast regions of Brazil. At first, the Ordinary Least Squares (OLS) model was estimated with the proposition of analyzing the correlation of the scores obtained by the candidate with the school of origin, using socioeconomic variables as control variables. It is important to consider the bias of the school's origin and the socioeconomic profile of the students, as students from private schools tend to have better grades. To mitigate this bias, the Propensity Score Matching (PSM) method was used, which is a valuable tool when the intention is to control selection biases. The results found reveal a discrepancy between the years 2019 and 2022, demonstrating an intensification of regional distortions in education, with better grade yields for students from private schools in the post-COVID-19 pandemic in the Southeast region.

**Keywords:** ENEM; PSM; MQO; Southeast; Northeast; COVID-19; socioeconomic variables.

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**

**PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO**

**DEPARTAMENTO DE ECONOMIA**

**BACHARELADO EM CIÊNCIAS ECONÔMICAS**

**LISTA DE GRÁFICOS**

Figura 01 - Histórico de inscritos ENEM e Fuvest	10
Figura 02 - Nota média do ENEM 2019 por unidade da federação	11
Figura 03 - Nota média do ENEM 2022 por unidade da federação	12
Figura 04 - Notas médias no ENEM por tipo de escola	24
Figura 05 - Gráfico QQ (quantile-quantile) Microdados ENEM 2019	54
Figura 06 - Gráfico QQ (quantile-quantile) Microdados ENEM 2022	54



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO****PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO****DEPARTAMENTO DE ECONOMIA****BACHARELADO EM CIÊNCIAS ECONÔMICAS****LISTA DE TABELAS**

Tabela 01 - Disparidade de notas entre escolas privadas em 2019	14
Tabela 02 - Disparidade de notas entre escolas públicas em 2019	14
Tabela 03 - Disparidade de notas entre escolas privadas em 2022	15
Tabela 04 - Disparidade de notas entre escolas públicas em 2022	15
Tabela 05 – Análise descritiva das variáveis socioeconômicas do Enem	22
Tabela 06 - Descrição das variáveis utilizadas nas estimações	26
Tabela 07 – Regressões modelo MQO 2019	28
Tabela 08 – Regressões modelo MQO 2022	29
Tabela 09 – Regressões modelo PSM 2019	32
Tabela 10 – Regressões modelo PSM 2022	33
Tabela 11 - Teste de Kolmogorov-Smirnov	42
Tabela 12 - Teste de Anderson-Darling	42
Tabela 13 - Inflação de variância	42
Tabela 14 - Teste de Durbin-Watson	43
Tabela 15 - Nordeste 2019	44
Tabela 16 - Sudeste 2019	44
Tabela 17 - Nordeste 2022	44
Tabela 18 - Sudeste 2022	45
Tabela 19 - Nordeste 2019	45
Tabela 20 - Sudeste 2019	45
Tabela 21 - Nordeste 2022	46
Tabela 22 - Sudeste 2022	46
Tabela 23 – Regressão do modelo MQO - Nordeste 2019	46
Tabela 24 – Regressão do modelo MQO - Sudeste 2019	47
Tabela 25 – Regressão do modelo MQO - Nordeste 2022	48
Tabela 26 - Regressão Modelo MQO - Sudeste 2022	48
Tabela 27 - Resultados Nordeste Modelo PSM 2019	49

Tabela 28 – Resultados do Sudeste Modelo PSM 2019	50
Tabela 29 - Coeficientes Nordeste Modelo PSM 2022	50
Tabela 30 - Coeficientes Sudeste Modelo PSM 2022	51
Tabela 31 - Modelo Logit para o ano de 2019	52
Tabela 32 - Modelo Logit para o ano de 2022	52
Tabela 33 - Regressão logística para o ano de 2019	53
Tabela 34 - Regressão logística para o ano de 2022	54

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**

**PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO**

**DEPARTAMENTO DE ECONOMIA**

**BACHARELADO EM CIÊNCIAS ECONÔMICAS**

**LISTA DE SIGLAS E ABREVIATURAS**

ENEM - Exame Nacional do Ensino Médio

PSM - *Propensity Score Matching*

MQO - Mínimos Quadrados Ordinários

LC - Linguagens, Códigos e suas Tecnologias

MT - Matemática e suas Tecnologias

CN - Ciências da Natureza e suas Tecnologias

CH - Ciências Humanas e suas Tecnologias

RD - Redação

PIB - Produto Interno Bruto

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

COVID - *Coronavirus Disease*

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**

**PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO**

**DEPARTAMENTO DE ECONOMIA**

**BACHARELADO EM CIÊNCIAS ECONÔMICAS**

**SUMÁRIO**

<b>1. INTRODUÇÃO</b>	<b>2</b>
<b>2. REVISÃO LITERÁRIA</b>	<b>5</b>
2.1. Relação entre os fatores socioeconômicos e o desempenho escolar	5
2.2. Efeito da pandemia da COVID-19 sobre o desempenho escolar	8
<b>3. O ENEM COMO FERRAMENTA DE ESTIMAÇÃO DA DESIGUALDADE REGIONAL</b>	<b>10</b>
<b>4. METODOLOGIA</b>	<b>16</b>
4.1. Mínimos Quadrados Ordinários	17
4.2. <i>Propensity Score Matching</i> (PSM)	18
4.3. Estratégia empírica	21
4.4. Dados	25
<b>5. RESULTADOS</b>	<b>27</b>
5.1. Mínimos Quadrados Ordinários (MQO)	27
5.2. <i>Propensity Score Matching</i> (PSM)	30
<b>6. CONSIDERAÇÕES FINAIS</b>	<b>33</b>
<b>7. REFERÊNCIAS</b>	<b>37</b>
<b>8. APÊNDICE</b>	<b>41</b>
8.1. Testes de normalidade dos resíduos	41
8.2. Multicolineariedade	42
8.3. Autocorrelação dos resíduos	43
8.4. Indicadores de ajustes dos Mínimos quadrados ordinários	44
8.5. Indicadores de ajustes do <i>Propensity Score Matching</i> (PSM)	45
8.6. Regressões completas Mínimos Quadrados Ordinários	46
8.7. Regressões completas <i>Propensity Score Matching</i>	49
8.8. Regressões <i>Logit</i> e função logística	52

## 1. INTRODUÇÃO

A educação desempenha um papel crucial no desenvolvimento econômico de um país, sendo um fator determinante para o crescimento sustentável e a melhoria das condições de vida da população. Diversos estudos e análises demonstram a estreita relação entre o nível educacional de uma nação e seu desempenho econômico. Segundo o economista Theodore W. Schultz (1961), em seu livro "*Investment in Human Capital*", o investimento em capital humano, representado pela educação e formação da força de trabalho, é essencial para aumentar a produtividade e a eficiência econômica. Schultz (1961) argumenta que os indivíduos que recebem uma educação sólida tornam-se mais capacitados para contribuir para a economia, gerando inovação, empreendedorismo e promovendo o avanço tecnológico.

Além disso, a educação não apenas influencia a produtividade individual, mas também contribui para a redução da desigualdade social e para o desenvolvimento de uma sociedade mais justa. A formação educacional proporciona oportunidades de acesso, permitindo que os cidadãos participem ativamente da economia, tenham acesso a empregos mais qualificados e contribuam para o crescimento econômico. Dentro desse contexto, Claudia Goldin e Lawrence F. Katz (2008), em seu trabalho "*The Race Between Education and Technology*", destacam a importância de investir na educação como uma resposta necessária ao avanço tecnológico. Eles argumentam que sociedades que priorizam a educação estão mais preparadas para enfrentar os desafios impostos pela evolução tecnológica, garantindo uma mão de obra mais qualificada e adaptável.

Nessa perspectiva, é de fundamental importância a utilização de ferramentas avaliativas para compreender o panorama da educação e identificar seus principais entraves. No contexto brasileiro, o Exame Nacional do Ensino Médio (ENEM) destaca-se como um dos mais relevantes e abrangentes instrumentos de avaliação educacional nas últimas décadas. Desde sua criação, em 1998, o ENEM passou por diversas mudanças e assumiu um papel cada vez mais importante no cenário educacional brasileiro. Funcionando não apenas como um processo de seleção para ingresso no ensino superior, mas também como uma ferramenta capaz de diagnosticar habilidades e competências dos estudantes ao longo dos anos.

O ENEM, ao longo de suas edições, passou por diversas mudanças metodológicas, por meio das quais procurou adequá-lo melhor às demandas do ensino e às necessidades dos alunos. Nesse sentido, vale destacar a mudança do exame como instrumento de seleção para ingresso

no ensino superior desde 2009, quando foi criado o Sistema de Seleção Unificada (SISU), que utiliza a nota do ENEM como único critério para ingresso em diversas instituições de ensino superior no Brasil. Essa mudança foi um marco significativo na história do ENEM, pois rompeu com o tradicional vestibular e abriu novas formas de acesso ao ensino superior, promovendo maior inclusão e democratização do aprendizado. De acordo com o INEP (2023), no ano de 2023, o SISU ofertou 226.399 mil vagas entre cursos de licenciatura, bacharelado e tecnólogos por 128 instituições públicas. No entanto, o exame tornou-se ainda mais importante, pois além de critério para admissão no SISU, foi também adotado como etapa obrigatória na candidatura ao financiamento estudantil por meio do Programa Universidade para Todos (ProUni) e do Fundo de Amparo ao Estudante (FIES).

Dessa forma, o Exame Nacional do Ensino Médio (ENEM) assumiu uma importância crucial tanto para os estudantes que desejam ingressar no ensino superior quanto para as instituições de ensino, sejam elas públicas ou privadas. Isso se dá pela sua capacidade de fornecer um modelo de avaliação padronizado, permitindo que os estudantes se preparem de maneira mais clara e precisa em relação às suas habilidades e características.

No entanto, é válido ressaltar que, conforme argumentado por Da Silva (2014), é de suma importância criar uma diversidade de alternativas de avaliação, todas elas independentes e externas, embora não necessariamente idênticas. Esse enfoque visa evitar a uniformização do sistema educacional e, ao mesmo tempo, promover a diversificação de abordagens na avaliação do conhecimento e das habilidades dos estudantes.

Além de o exame ser utilizado como critério de ingresso em faculdades, ele se destaca como uma valiosa ferramenta de diagnóstico na educação devido à sua abrangência nacional, avaliação de distintas áreas de competência e à sua periodicidade anual. Isso proporciona uma visão panorâmica da qualidade do ensino nas diferentes regiões do país. Nesse sentido, o ENEM ajuda a identificar as fragilidades e potencialidades da educação brasileira, o que permite um direcionamento mais efetivo das políticas públicas para a melhoria da educação básica.

Apesar do avanço e do reconhecimento de sua importância, o ENEM também enfrenta críticas significativas. Entre essas críticas, destaca-se a preocupação com a pesada carga de conteúdo que os alunos devem dominar para obter um desempenho satisfatório no exame. O extenso rol de disciplinas exigido no ENEM torna-se um desafio considerável, especialmente para os estudantes com acesso limitado a recursos educacionais de qualidade.

Além disso, surgem questionamentos acerca do TRI (Teoria de Resposta ao Item). O TRI é um conjunto de modelos matemáticos que considera o item como unidade básica de análise e procura representar a probabilidade de um examinando dar uma certa resposta a um item como função dos parâmetros do item e dos traços latentes do indivíduo (COUTO; PRIMI, 2011). Isso levanta a possibilidade de se alcançar pontuações elevadas mesmo sem um conhecimento aprofundado das disciplinas, o que pode causar distorções dos resultados.

Dessa forma, torna-se essencial conduzir uma análise dos dados abertos da prova, os quais oferecem informações detalhadas sobre os participantes do exame, abrangendo aspectos socioeconômicos, escolares e de desempenho. O presente estudo visa compreender as disparidades entre as regiões Nordeste e Sudeste quanto ao impacto das variáveis socioeconômicas no desempenho dos candidatos nas cinco provas do ENEM (Linguagens, Ciências Humanas, Ciências da Natureza, Matemática e Redação). Este estudo é crucial para compreender os impactos desiguais no desempenho educacional entre diferentes estratos socioeconômicos, especialmente considerando as disparidades econômicas historicamente presentes nas duas regiões. A região Sudeste, caracterizada por sua forte presença econômica e centros urbanos consolidados, apresenta uma realidade educacional distinta do Nordeste, que enfrenta desafios socioeconômicos mais pronunciados. A pandemia acentuou as desigualdades preexistentes, destacando a vulnerabilidade de estudantes de escolas públicas, predominantemente localizadas em áreas economicamente desfavorecidas. Para esta análise, será empregada a modelagem estatística nos resultados de 2019 e 2022, com o objetivo de investigar como se manifestam essas diferenças antes e após a pandemia nas duas regiões em estudo.

Para realizar esta análise, serão empregados dois modelos estatísticos distintos. O Mínimo Quadrado Ordinário (MQO), que tem como objetivo calcular a correlação entre as variáveis socioeconômicas e a nota obtida pelo candidato em cada prova. Além disso, será aplicado o *Propensity Score Matching* (PSM), um modelo voltado para o controle de viés onde neste estudo será empregado para mitigar o viés relacionado à escola de origem (pública e privada). A vantagem do PSM frente ao MQO reside na capacidade de eliminar possíveis vieses de seleção da escola de origem, controlando o grupo de tratamento e de controle por um conjunto de variáveis socioeconômicas similares. Por fim, será apresentado no apêndice o modelo *Logit* juntamente com a função sigmoide que será utilizado com o intuito de calcular a probabilidade de um candidato ser proveniente de escola privada a partir de suas características socioeconômicas.

Esta monografia está estruturada em oito capítulos, a saber: Introdução, Referencial Teórico - no qual serão apresentados estudos com diversas abordagens realizados sobre o Exame Nacional do Ensino Médio (ENEM), o ENEM como ferramenta de estimação da desigualdade regional - que buscará demonstrar como o ENEM pode fornecer informações valiosas para expor os níveis de desigualdades entre escolas públicas e privadas de diferentes regiões, Metodologia - neste capítulo serão apresentadas as características e uma descrição minuciosa dos métodos estatísticos utilizados, Resultados - onde serão apresentados os resultados das regressões do MQO e PSM, Considerações Finais - que buscará sintetizar os resultados da pesquisa e destacar a relevância e as implicações do estudo, Referências - onde serão apresentadas as referências utilizadas em todos os estudos, e Apêndice - que apresentará os testes estatísticos realizados.

## **2. REVISÃO LITERÁRIA**

### 2.1. Relação entre os fatores socioeconômicos e o desempenho escolar

A desigualdade nas notas obtidas no Exame Nacional do Ensino Médio (ENEM) é um tema de extrema relevância no contexto educacional brasileiro. Essa disparidade, que tem sido objeto de inúmeras discussões acadêmicas e políticas, reflete não apenas as diferenças individuais de desempenho dos estudantes, mas também uma série de fatores socioeconômicos e educacionais que moldam o acesso à educação de qualidade e, conseqüentemente, o resultado obtido nas provas do ENEM. Por esse motivo é de extrema importância explorar as principais teorias e pesquisas que abordam as causas e as conseqüências das desigualdades nas notas do ENEM, buscando compreender os mecanismos subjacentes a esse fenômeno.

De acordo com Feijó e França (2021), a diferença de desempenho entre alunos de escolas públicas e privadas é explicada predominantemente pelo que os autores nomeiam de efeito composição, que se refere às dotações dos estudantes, ou seja, às suas características individuais e de seus familiares. Esse efeito é influenciado também pelo *background* da turma em que o estudante está inserido características socioeconômicas dos colegas como a educação dos pais, a educação da mãe e a renda familiar mensal dos estudantes aliado ao *background* familiar dos estudantes. Além disso, os autores mencionam o “efeito estrutural”, que se refere à forma como os alunos utilizam em seu benefício as características dos docentes e da escola, conseguindo



transformá-los em maiores pontuações. Esse efeito é mais importante nos quantis inferiores da distribuição de notas, com exceção da prova de redação (Feijó e França, 2021).

Em estudo realizado por Santos e Mariano (2018) ficou evidenciado que a participação dos pais em reuniões escolares tem um impacto positivo no desempenho acadêmico dos alunos, especialmente em relação à proficiência em português e matemática. Este estudo também encontrou um diferencial de desempenho do Sistema de Avaliação da Educação Básica (Saeb) de 2013 entre escolas públicas e privadas no estado do Ceará, com os alunos das escolas privadas apresentando um desempenho superior aos alunos das escolas públicas, as escolas privadas apresentam um diferencial de desempenho significativo em relação às escolas públicas. Em língua portuguesa, as escolas privadas tiveram uma vantagem de 28,08, 35,68 e 40,72 pontos para o 5º, 9º e 3º ano, respectivamente. Já em matemática, as diferenças são de 25,01, 42,77 e 49,95 pontos a mais para os mesmos anos escolares. Esses números destacam a discrepância no desempenho entre as duas redes de ensino.

O artigo "Igualdade de Oportunidades: Analisando o Papel das Circunstâncias no Desempenho do ENEM", de autoria de Figueirêdo, Nogueira e Santana (2014), aborda a questão da igualdade de oportunidades no contexto educacional, com foco no desempenho dos estudantes no Exame Nacional do Ensino Médio (ENEM) e do Sistema de Avaliação da Educação Básica (SAEB) de 2010. O estudo utiliza estratégias empíricas para investigar se a origem familiar tem um papel relevante no desempenho dos alunos que concluíram o Ensino Médio no Brasil e no Nordeste, e como as circunstâncias sociais influenciam o resultado econômico. Os resultados deste estudo mostraram que o esforço empregado é um fator determinante para um bom desempenho e que o efeito indireto da educação da mãe como o ambiente familiar, o apoio educacional oferecido em casa e a valorização da educação é surpreendentemente maior do que os efeitos diretos que se trata do nível de escolaridade da mãe. Além disso, o estudo destaca que a raça e o gênero também influenciam o desempenho educacional, com estudantes não negros e do sexo feminino apresentando vantagem em relação aos demais.

No estudo "*Equality of Opportunity: A Progress Report*", o autor Roemer (1998) aprofundou a discussão sobre igualdade de oportunidades, ressaltando a importância de assegurar acesso equitativo às condições para alcançar metas, independentemente das circunstâncias iniciais. Ele destaca que a igualdade de oportunidades vai além da eliminação de barreiras legais, requerendo a criação de um ambiente que capacite todos os indivíduos a

perseguirem seus objetivos de maneira justa. Roemer (1998) argumenta que alcançar verdadeira igualdade demanda não apenas ausência de discriminação explícita, mas também a implementação de políticas que abordem disparidades sistêmicas.

Roemer (1998) sublinha a necessidade de reconhecer e combater desigualdades estruturais decorrentes de condições socioeconômicas desfavoráveis, propondo abordagens para transformar a distribuição de recursos e oportunidades. Suas análises provavelmente exploraram estratégias políticas e econômicas visando erradicar, não apenas mitigar, disparidades persistentes, garantindo a genuína igualdade de oportunidades em todas as esferas da sociedade.

No que diz respeito ao desempenho no ENEM, o estudo de Albernaz *et al.* (2002) é relevante. Os autores utilizaram dados do Sistema de Avaliação da Educação Básica (SAEB) para estimar a função de produção educacional no ensino fundamental. Eles concluíram que cerca de 80% da variância do desempenho médio entre as escolas pode ser atribuída a fatores relacionados às circunstâncias dos estudantes. Outro aspecto abordado no artigo foi a influência do background familiar na obtenção de um bom desempenho no ENEM. Ferreira *et al.* (2008); e Ferreira e Gignoux (2011) destacam a importância da educação dos pais como uma das circunstâncias mais relevantes que afetam os resultados educacionais. Essa heterogeneidade regional também é observada, com as regiões Norte e Nordeste apresentando maiores desafios em termos de oportunidades educacionais (Gamboa e Waltenberg, 2012).

Portanto, é importante ressaltar que a igualdade de oportunidades é um tema complexo e multidimensional, envolvendo não apenas fatores individuais, mas também questões sociais, econômicas e políticas. Dessa forma, a análise do desempenho no ENEM e a identificação das disparidades de oportunidades entre os grupos são fundamentais para promover políticas educacionais mais justas e igualitárias.

## 2.2. Efeito da pandemia da COVID-19 sobre o desempenho escolar

A pandemia do COVID-19, que assolou o mundo no final de 2019, porém trouxe maiores reverberações no ano de, com restrições de circulação e *lockdowns*, atingindo todos os setores da sociedade incluindo as instituições educacionais que tiveram que se adaptar a uma nova realidade com o distanciamento social, trouxe efeitos adversos no desempenho educacional dos estudantes em todos os níveis educacionais. Essa crise sanitária gerou desafios inesperados para

os sistemas de ensino, obrigando educadores, estudantes e famílias a se adaptarem rapidamente a novas realidades e modalidades de aprendizagem, onde o ensino passou a ser virtual e exigiu diversas adaptações físicas, tecnológicas e sociais. Diante desse contexto, tornou-se fundamental uma análise criteriosa de aspectos cruciais, como desigualdade no acesso à educação, isolamento social e saúde mental, perda de aprendizado e os desafios enfrentados pelos educadores.

Segundo o estudo *'The potential effects of the COVID-19 pandemic on learning'* realizado por Luis Monroy-Gómez-Franco, Roberto Vélez-Grajales e Luis F. López-Calva (2022), a pandemia do COVID-19 teve o potencial de gerar impactos significativos no ensino, com repercussões de longo prazo. A transição abrupta do ensino presencial para o remoto resultou em perdas de aprendizado, afetando não apenas o desempenho em testes, mas também o tempo dedicado pelos estudantes aos estudos. Estudos anteriores indicam que a redução do tempo de estudo durante a pandemia foi substancial, com uma diminuição de até 18 horas por semana, evidenciando um declínio preocupante no engajamento acadêmico. Além disso, os autores destacam que mesmo sob a premissa de que o modelo de instrução remota implementado foi um substituto perfeito para as aulas presenciais, a perda média de aprendizado a curto prazo foi maior do que zero. Essa constatação sugere que mesmo choques temporários podem ser amplificados por um sistema educacional que não se ajusta aos perfis de aprendizado dos alunos, o que pode acarretar em custos de aprendizado associados à pandemia.

No âmbito nacional, no estudo desenvolvido por Weber Neto *et al.* (2022), foi realizada uma análise comparativa entre os dados do Exame Nacional do Ensino Médio (ENEM) nos anos de 2019 e 2020, com o intuito de investigar o impacto da pandemia da COVID-19. Através da mineração de dados educacionais do ENEM, os pesquisadores identificaram mudanças significativas no desempenho dos estudantes, especialmente nas notas médias gerais. Os resultados revelaram uma alteração nas médias das notas entre os dois anos, indicando uma possível influência da pandemia nos resultados obtidos. Além disso, a disparidade de notas entre candidatos de diferentes níveis socioeconômicos também foi analisada. Em geral, houve um aumento leve nas notas médias dos candidatos, com exceção da classe de renda “nenhuma renda”, que teve uma redução de 6,8 pontos, evidenciando variações que sugerem um impacto diferenciado da pandemia nos estudantes de acordo com sua condição financeira. Esses achados ressaltam a importância de compreender os efeitos da pandemia no contexto educacional e suas implicações para a equidade no acesso à educação.

Já no estudo realizado por Willian H. de Brito e Fabiano P. Pedroso (2022) buscou-se compreender como as características socioeconômicas individuais e contextuais influenciaram o desempenho dos estudantes paranaenses que realizaram o Enem nos anos de 2020 e 2021. Os resultados apontam que os efeitos negativos da pandemia foram mais intensos sobre os alunos com renda familiar inferior a três salários-mínimos, matriculados em escolas públicas e autodeclarados pretos, pardos ou indígenas. Esses grupos foram os mais prejudicados, refletindo desigualdades socioeconômicas e étnico-raciais no acesso à educação durante a crise sanitária. Além disso, a variável acesso à internet mostrou-se relevante, com os estudantes que possuíam esse recurso apresentando impacto positivo em seus desempenhos. No entanto, quando analisado alunos de escolas públicas versus particulares ocorreu uma queda na disparidade durante a pandemia, encarada pelos autores como uma queda artificial.

Em estudo desenvolvido por Weber Neto *et al.* (2022), foi empregada uma análise exploratória dos dados do ENEM nos estados do Ceará, Maranhão e Piauí durante a pandemia da Covid-19. Os resultados mostram uma melhora no desempenho geral dos estados, um aumento no número de inscritos que não compareceram à prova e um crescimento da desigualdade entre os participantes provenientes de instituições públicas e privadas. O estudo também destaca a importância da infraestrutura financeira para o sucesso dos alunos na modalidade de ensino a distância.

Diante do exposto, à medida que o mundo se recupera da pandemia da covid-19, é fundamental realizar estudos sobre os possíveis efeitos da COVID-19 no desempenho educacional, para possibilitar implementação de estratégias voltadas ao combate das desigualdades de acesso à educação, apoiar a saúde mental dos estudantes e desenvolver métodos de ensino híbridos que aproveitem as lições aprendidas durante a pandemia. A educação é um pilar fundamental para o futuro, e é crucial garantir que nenhum aluno fique para trás devido aos impactos da COVID-19 por esse motivo se faz necessário analisar quais os efeitos da pandemia no desempenho de estudantes com acesso a ensino privado versus público.

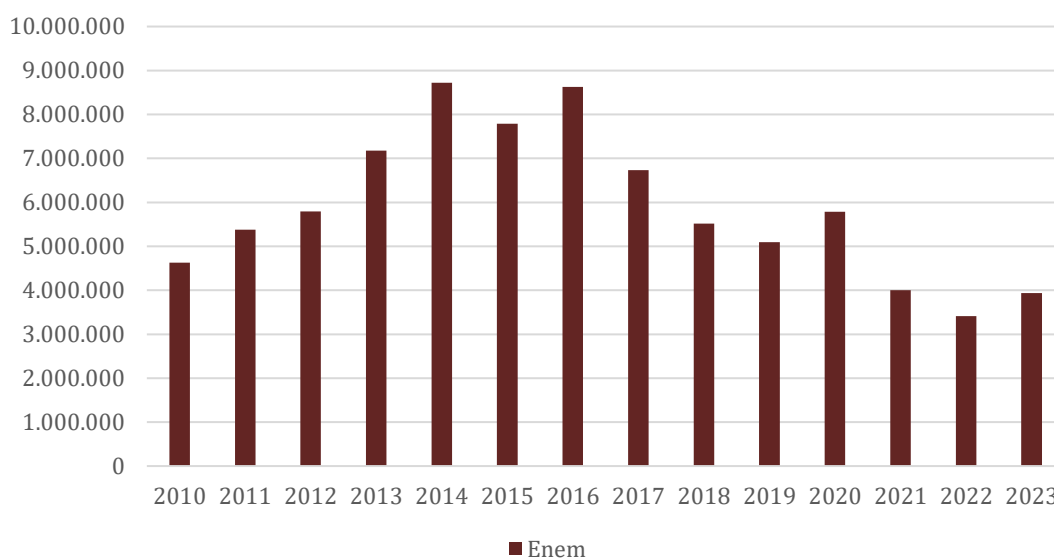
### **3. O ENEM COMO FERRAMENTA DE ESTIMAÇÃO DA DESIGUALDADE REGIONAL**

O Exame Nacional do Ensino Médio (ENEM) destaca-se como uma ferramenta crucial para a compreensão e análise da desigualdade regional no contexto educacional brasileiro. Sua importância vai além da simples avaliação individual dos estudantes, tornando-se um indicador

robusto das disparidades presentes nas diversas regiões do país. Essa robustez decorre de várias características intrínsecas ao exame, sendo as principais a sua frequência anual, que possibilita a captura contínua de mudanças e tendências no desempenho dos estudantes ao longo do tempo e em diferentes regiões. Essa característica é crucial para monitorar o impacto de políticas públicas, reformas educacionais e outras iniciativas. Outro aspecto relevante que torna o ENEM extremamente eficaz como ferramenta de análise educacional é a amplitude de conteúdos abordados na prova. Essa variedade proporciona uma avaliação abrangente do conhecimento dos estudantes, não se limitando apenas a língua portuguesa e matemática, como ocorre em exames como a Prova Brasil e o SAEPE. O ENEM inclui também outras ciências, avaliando assim a qualidade e a diversidade dos currículos escolares.

A terceira vantagem reside em sua abrangência nacional, sendo o maior vestibular do Brasil, com milhões de inscritos, como demonstrado na Figura 01. Sua amplitude reflete a intenção de proporcionar oportunidades equitativas a estudantes de diferentes estados e contextos socioeconômicos.

**Figura 01 - Histórico de inscritos ENEM**



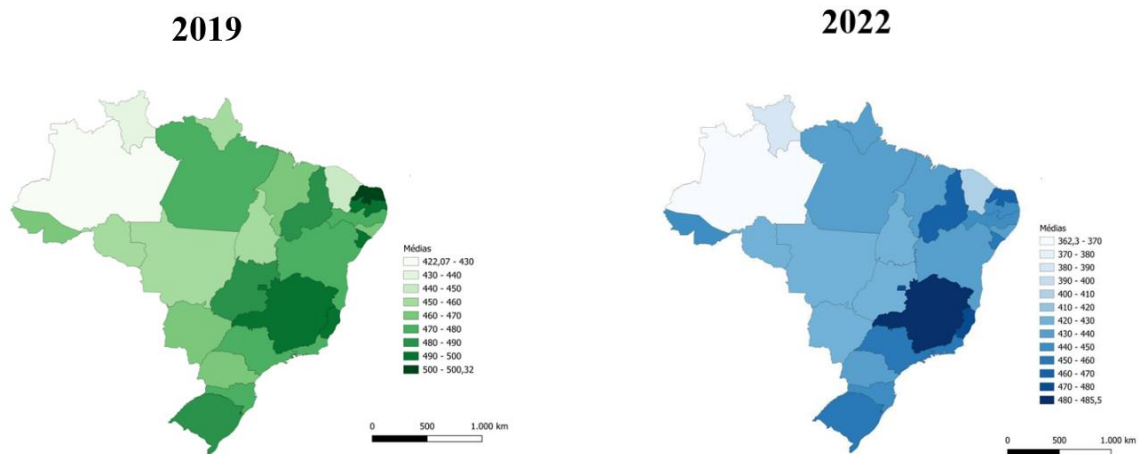
Fonte: Elaboração própria a partir de dados do INEP.

Devido à sua abrangência, o ENEM permite realizar uma excelente fotografia anual da qualidade da situação educacional de diferentes perfis de estudantes em todas as regiões do

país, podendo contribuir para uma análise regional mais aprimorada da situação educacional. A Figura 02, a seguir, contém informações que correspondem às médias aritméticas de 2019 e 2022 por estado, onde é possível ter uma visão panorâmica de como as notas mudam drasticamente de um estado brasileiro para outro. Quando analisada a média por região, observamos que o Sudeste apresentou a melhor média em 2019, atingindo 484,39 pontos, enquanto o Nordeste registrou uma média ligeiramente inferior, marcando 477,48 pontos. Essa diferença de 6,91 pontos entre as duas regiões é um aspecto notável e que merece uma análise mais aprofundada, visto que reflete não apenas desigualdades regionais, mas também potenciais disparidades nas condições socioeconômicas e estruturais das escolas e dos estudantes. No mesmo período, o Centro-Oeste se posicionou como a terceira região com a média de 474,95 pontos, enquanto o Sul alcançou a quarta maior média, com 473,73 pontos. A região Norte, por sua vez, apresentou a pior média nacional, registrando 452,66 pontos. Essa variação de quase 22 pontos entre a melhor e a pior média nacional ressalta a magnitude das discrepâncias educacionais no Brasil. Essas diferenças entre as médias do Sudeste e demais regiões não se limita a uma mera questão estatística, mas aponta para desafios substanciais no sistema educacional brasileiro. Essa disparidade pode estar relacionada a diversos fatores, como investimentos em infraestrutura escolar, qualidade do corpo docente, acesso a recursos educacionais e condições socioeconômicas dos estudantes.

No ano de 2019 o Amazonas foi o estado que teve a menor média, com apenas 422,07 pontos, enquanto o estado com melhor desempenho médio foi o Rio Grande do Norte, com 500,32 pontos, seguido de Sergipe, com 497,18 pontos. Ambos ficaram acima da média de estados com maiores PIB per capita, como São Paulo, com 471,40, Rio de Janeiro, 477,61, e Minas Gerais, o estado com a terceira melhor média em 2019, com 495,87.

**Figura 02 - Nota média por estado do ENEM 2019 e 2022**



Fonte: Elaboração própria a partir dos Microdados ENEM 2019 e 2022.

Na Figura 02, também é apresentado as médias de nota do ENEM por Unidade da Federação em 2022. Essas médias são especialmente relevantes, pois se trata do primeiro ENEM após o retorno de 100% das aulas, podendo fornecer uma análise importante quando comparadas à Figura 02, pois possibilita visualizar o impacto da pandemia na média geral de cada estado e região. Quando analisada por região, todas regrediram em termos de médias. A região Sudeste teve a melhor média, 469,49, uma redução de 14,89 pontos em relação a 2019. A segunda melhor média foi do Sul, com 445,81 de média, uma redução de 27,92 pontos. O Nordeste tornou-se a terceira região com melhor desempenho, com 443,10, uma diferença de 34,38 pontos em relação a 2019. O Centro-Oeste teve uma média de 438,13, uma queda de 36,81, e a região com pior desempenho foi o Norte, com média de 413,82 pontos, tendo sido a maior discrepância em relação a 2019, com uma queda de 38,84 pontos na média. Essas quedas nas notas de todas as regiões revelam que o impacto da pandemia afetou todos, porém de forma distinta; regiões em situação de pior status socioeconômico tiveram efeitos mais severos do que regiões com melhor estrutura econômica.

Em relação aos estados individualmente o Amazonas manteve-se como o estado com pior desempenho, com uma nota média de 362,33, tendo tido uma queda de 102,5 pontos em relação ao ano de 2019. Já o Rio Grande do Norte, que era o estado com melhor desempenho em 2019, tornou-se a quarta melhor nota, com 465,87, uma queda de 34,45 pontos. Sergipe saiu da segunda melhor nota para a sexta melhor nota, com uma média estadual de 458,39, uma perda de 38,79 pontos. Os dois estados com melhor desempenho em 2022 foram Minas Gerais, com 485,46, e Espírito Santo, com 478,97. Embora sejam as melhores médias, ambos tiveram quedas em comparação a 2019, respectivamente de 10,41 e 13,68 pontos.

Para melhor compreensão das desigualdades regionais enfrentadas na educação, é necessário desagregar os dados e analisar escola pública versus escola pública e o mesmo ser feito com instituições privadas. Essa abordagem será apresentada entre as Tabelas 01 e 04, onde se encontram os valores das diferenças das médias entre escolas de diferentes regiões porém de mesma natureza (pública ou privada) e o p-valor do teste t, usado para determinar se há uma diferença estatisticamente significativa entre as médias de dois grupos em que na hipótese nula ( $p > 0,05$ ), não há diferença significativa entre as médias dos grupos, enquanto na hipótese alternativa ( $p < 0,05$ ), existe diferença significativa entre as médias. Para melhor visualização será apresentado com realce verde aqueles valores que apresenta um  $p < 0,001$ , em amarelo aqueles com  $p < 0,01$ , em vermelho aqueles com  $p < 0,05$  e aqueles em brancos se trata de valores em que não há diferença significativa entre as médias.

Na Tabela 01, são apresentadas as disparidades entre instituições privadas no ano de 2019, juntamente com os indicativos de significância do teste t. A maior diferença entre médias é observada na comparação entre a Região Nordeste e a Sul, com um valor de 10,92 pontos e uma diferença estatisticamente significativa entre essas médias. Por outro lado, a menor disparidade é encontrada na média das escolas nordestinas em relação às sudestinas, com apenas 0,70 pontos. Essa diferença indica que, apesar das discrepâncias regionais em termos de renda, as instituições privadas, quando comparadas, apresentam resultados semelhantes. Essa suposição de que a região tem pouca influência quando se trata de instituições privadas vai além da região Nordeste; o Norte, por exemplo, tem 1,98 pontos a mais que o Sul, uma região consideravelmente mais rica.

**Tabela 01 - Disparidade de notas entre escolas privadas em 2019**

<b>Média</b> <sub>coluna - Média</sub> <b>linha</b>	Média <sub>Centro-Oeste</sub>	Média <sub>Nordeste</sub>	Média <sub>Norte</sub>	Média <sub>Sudeste</sub>
Média <sub>Nordeste</sub>	-5,274643*			
Média <sub>Norte</sub>	3,665837	8,940479***		
Média <sub>Sudeste</sub>	-4,571392	0,703251*	-8,237228***	
Média <sub>Sul</sub>	5,648436	10,92308***	1,982599*	10,21983***

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Teste-t: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ .

Ao realizar a análise das instituições públicas, os valores foram bastante discrepantes entre todas as regiões, especialmente quando comparamos as duas regiões mais ricas (Sul e Sudeste) com as demais. Isso evidencia que as distorções regionais são mais pronunciadas em escolas



públicas do que em escolas privadas, que são mais equidistantes quando comparadas às públicas. A maior discrepância, apresentada no ano de 2019, foi entre as regiões Norte e Centro-Oeste, com uma diferença de 17,12 pontos a favor da região Centro-Oeste. A menor discrepância ocorreu entre Centro-Oeste e Sul, com uma vantagem de 0,85 pontos para o Centro-Oeste. Além disso, podem ser observadas diversas discrepâncias, como entre Nordeste e Sudeste, com uma diferença de 6,62 pontos entre as regiões, ou entre Norte e Sul, com uma vantagem de 16,27 pontos para o Sul em comparação com o Norte.

**Tabela 02 - Disparidade de notas entre escolas públicas em 2019**

<b>Média</b> <sub>coluna - Média</sub> <sub>linha</sub>	Média <sub>Centro-Oeste</sub>	Média <sub>Nordeste</sub>	Média <sub>Norte</sub>	Média <sub>Sudeste</sub>
Média <sub>Nordeste</sub>	9,524306***			
Média <sub>Norte</sub>	17,12074***	7,596437***		
Média <sub>Sudeste</sub>	2,904322**	-6,619984***	-14,21642***	
Média <sub>Sul</sub>	0,8521391	-8,672167***	-16,2686***	-2,052183*

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Teste-t: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

As Tabelas 03 e 04 são especialmente importantes, pois fornecem os valores das diferenças de notas entre instituições pós-pandemia, possibilitando a análise de como a pandemia do COVID-19 pode ter aprofundado ainda mais as distorções e quais instituições foram mais impactadas. A Tabela 03 apresenta as diferenças entre escolas privadas no ano de 2022. Com exceção das comparações Centro-Oeste menos Nordeste e Centro-Oeste menos Sudeste, que registraram uma redução na diferença entre as notas, as demais comparações apresentaram um agravamento nas diferenças de notas. Isso evidencia que mesmo as instituições privadas foram afetadas pelo impacto da pandemia do COVID-19, com um aumento nas distorções regionais.

**Tabela 03 - Disparidade de notas entre escolas privadas em 2022**

<b>Média</b> <sub>coluna - Média</sub> <sub>linha</sub>	Média <sub>Centro-Oeste</sub>	Média <sub>Nordeste</sub>	Média <sub>Norte</sub>	Média <sub>Sudeste</sub>
Média <sub>Nordeste</sub>	-2,068755			
Média <sub>Norte</sub>	13,70622***	15,77498***		
Média <sub>Sudeste</sub>	-1,135465	0,9332905	-14,84169***	
Média <sub>Sul</sub>	9,309575**	-11,37833***	-4,396648	10,44504***

Fonte: Elaboração própria a partir dos Microdados ENEM 2022.

Teste-t: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

A Tabela 04 apresenta os valores das discrepâncias entre instituições públicas, e o que observamos é uma aceleração significativa das distorções entre as escolas públicas, com exceção da comparação entre Centro-Oeste e Nordeste. Destaca-se, especialmente, as regiões do Sudeste e Sul, que apresentaram resultados significativamente superior em comparação com regiões mais pobres, como o Norte. Ao comparar o Norte com o Sudeste e o Sul, observa-se uma diferença de 34,98 e 26,44 pontos, respectivamente, representando as maiores discrepâncias apresentadas no ano de 2022.

**Tabela 04 - Disparidade de notas entre escolas públicas em 2022**

<b>Média</b> <sub>coluna - Média</sub> <sub>linha</sub>	<b>Média</b> <sub>Centro-Oeste</sub>	<b>Média</b> <sub>Nordeste</sub>	<b>Média</b> <sub>Norte</sub>	<b>Média</b> <sub>Sudeste</sub>
<b>Média</b> <sub>Nordeste</sub>	0,6078489			
<b>Média</b> <sub>Norte</sub>	14,06644***	13,45859***		
<b>Média</b> <sub>Sudeste</sub>	-20,91731***	-21,52516***	-34,98345***	
<b>Média</b> <sub>Sul</sub>	-12,37274***	-12,98059***	-26,43918***	8,544567***

Fonte: Elaboração própria a partir dos Microdados ENEM 2022.

Teste-t: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

Os dados apresentados nas Tabelas de 01 a 04 destacam a importância de utilizar o ENEM como ferramenta de análise das desigualdades regionais, permitindo uma visualização clara das distorções educacionais entre as diversas regiões do Brasil. Além disso, fica evidente que, embora existam distorções também entre escolas privadas, estas são consideravelmente menores em comparação com as escolas públicas. Esse cenário foi agravado pelo impacto da COVID-19, que intensificou as disparidades entre as escolas públicas de regiões mais ricas, como Sul e Sudeste, em comparação com as demais regiões.

Quando analisamos as discrepâncias entre as regiões Nordeste e Sudeste, objeto de investigação neste estudo, observamos na Tabela 02 uma diferença significativa entre as médias das duas regiões em relação às escolas públicas. Contudo, o mesmo não é evidenciado no ano de 2019 para as escolas privadas, onde existe uma diferença menor que um ponto na média, conforme apresentado na Tabela 01. Entretanto, tal diferença não apresenta significância estatística no Teste-t. Ao examinarmos as diferenças de médias em 2022, percebemos a manutenção da não significância e uma diferença inferior a 1 ponto entre as médias das regiões Sudeste e Nordeste, conforme destacado na Tabela 03. No entanto, nota-se uma piora na discrepância em relação às escolas públicas na Tabela 04, onde há uma diferença de 21,52 pontos. Esses resultados derivam das características distintas das duas regiões. A região

Nordeste é a mais carente em termos de renda per capita no país, enquanto o Sudeste possui um nível de renda per capita acima da média nacional. Além disso, o Nordeste apresenta um nível de escolaridade inferior ao do Sudeste.

#### 4. METODOLOGIA

O objetivo deste estudo é realizar uma análise regional do impacto da escola de origem do candidato na nota obtida no ENEM utilizando como controle variáveis socioeconômicas nos anos de 2019 e 2022, ou seja, antes e depois da pandemia do COVID-19. Pretende-se investigar como essas variáveis afetam de maneira distinta as regiões Nordeste e Sudeste.

Para a condução dessa análise, serão empregadas duas modelagens estatísticas, juntamente com um conjunto de estatísticas descritivas. Será utilizado o modelo de Mínimos Quadrados Ordinários (MQO), buscando estimar a correlação entre variáveis independentes (características socioeconômicas) e dependentes (desempenho por prova). Além disso, também será empregado o método *Propensity Score Matching* (PSM) que tem por objetivo obter coeficientes de correlação mais verossímeis em comparação aos estimados pelos MQO, considerando que seu uso neste trabalho visa controlar o viés da escola de origem (privada ou pública). Por fim, no apêndice consta regressões do modelo *Logit* e da função sigmoide que foram utilizados em conjunto para obter a probabilidade de um indivíduo ser proveniente de uma escola privada com base em suas características socioeconômicas.

##### 4.1. Mínimos Quadrados Ordinários

O Modelo de Mínimos Quadrados Ordinários (MQO) representa uma técnica fundamental na estatística e na econometria, usada para estimar parâmetros desconhecidos em um modelo linear. Segundo Wooldridge (2009, p. 37), "os Mínimos Quadrados Ordinários são uma técnica amplamente utilizada para estimar parâmetros em modelos de regressão linear, devido à sua simplicidade e eficiência". A ideia central do MQO é minimizar a soma dos quadrados dos resíduos, isto é, a diferença entre os valores observados e os valores preditos pelo modelo. Conforme Greene (2012, p. 74) destaca, "a minimização dos quadrados dos resíduos é a base matemática dos Mínimos Quadrados Ordinários e torna essa técnica altamente eficaz na busca dos melhores parâmetros do modelo linear".

No entanto, o modelo enfrenta diversas problemáticas que podem causar viés nos resultados, devido a motivos como a multicolinearidade, ela se manifesta quando duas ou mais variáveis independentes no modelo de regressão estão altamente correlacionadas entre si. De acordo com Hair *et al.* (2019), a multicolinearidade pode tornar difícil distinguir o impacto individual de cada variável independente no resultado, uma vez que seus coeficientes se tornam instáveis e pouco confiáveis. Outro desafio é a heteroscedasticidade, que ocorre quando a dispersão dos erros não é uniforme em todos os níveis das variáveis independentes, afetando a precisão das estimativas dos parâmetros do modelo e a validade das inferências estatísticas.

Além disso, a endogeneidade é um terceiro problema que pode surgir na modelagem MQO, quando uma variável independente está correlacionada com os resíduos do modelo, levando a estimativas viesadas dos coeficientes. Como salientado por Greene (2018), a endogeneidade pode levar a estimativas viesadas dos coeficientes de regressão e à perda da eficiência dos estimadores MQO. Por fim, devemos mencionar a possibilidade da presença de outliers, que são valores atípicos nos dados e podem distorcer as estimativas dos coeficientes, de acordo com McCulloch (2008), outliers podem levar a estimativas dos parâmetros do MQO que não são mais representativas da tendência subjacente dos dados. Isso ocorre porque o MQO atribui pesos iguais a todas as observações, tratando-as como igualmente importantes. No entanto, os outliers, por sua natureza, geralmente têm um impacto desproporcional nas estimativas, puxando os parâmetros do modelo na direção deles.

A equação que representa o modelo MQO é a seguinte:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (03)$$

Nesta equação:

- Y: representa a variável dependente que estamos tentando prever.
- $B_0$ : é o intercepto, o valor de Y quando todas as variáveis independentes (X) são iguais a zero.
- $B_1, \beta_2, \dots, \beta_n$ : são os coeficientes de regressão para as variáveis independentes  $X_1, X_2, \dots, X_n$ , que representam o efeito dessas variáveis sobre Y.
- $\varepsilon$ : é o termo de erro, que captura a variação não explicada pelo modelo.

Devido aos problemas enfrentados pelo MQO, são necessários testes estatísticos para estimativas precisas, a fim de assumir a linearidade da amostra. Esses testes estão apresentados

no capítulo apêndice. Ao realizar o teste de normalidade dos resíduos, foi encontrada uma distribuição não normal, conforme apresentado na Figura 04, que exibe o QQ (quantile-quantile) para o ano de 2019, e na Figura 05, que mostra o QQ (quantile-quantile) para o ano de 2022. Para uma melhor estimação, além da representação gráfica, foram realizados dois testes de normalidade: o Kolmogorov-Smirnov (Tabela 11) e Anderson-Darling (Tabela 12). Ambos indicaram a presença de uma distribuição não normal dos resíduos. Para solucionar tal problemática, foi realizado um ajuste por meio da transformação logarítmica. Esse método de ajuste foi empregado, pois a transformação logarítmica (log) é frequentemente aplicada quando se lida com dados que exibem uma relação exponencial ou multiplicativa. Além disso, é útil quando os dados têm uma distribuição assimétrica positiva.

Com relação aos demais testes, como o Durbin-Watson (Tabela 14), que buscou estimar se existia autocorrelação entre os resíduos, e a Inflação de Variância (Tabela 13), que buscou averiguar se existia ou não multicolinearidade, a amostra não apresentou a necessidade de transformações, pois não viola os pressupostos da linearidade, que são pré-requisitos para uma estimação desse tipo.

#### 4.2. *Propensity Score Matching* (PSM)

É de fundamental importância reconhecer os desafios enfrentados pela modelagem MQO fazendo necessário adotar abordagens apropriadas para mitigar as problemáticas enfrentadas ao utilizar o modelo MQO por isso será utilizado neste estudo o modelo PSM, ou *Propensity Score Matching*, que é uma técnica estatística frequentemente utilizada em pesquisas sociais e na econometria para equilibrar grupos de tratamento e controle em estudos observacionais. Ela visa mitigar possíveis vieses de seleção e permitir comparações mais confiáveis entre grupos.

No estudo feito por Rosenbaum e Rubin (1983), os autores destacaram a importância do PSM ao afirmar que "a comparação de tratamento com controle pode ser realizada dentro de subgrupos bem definidos que se assemelham no score de propensão." Essa abordagem possibilita a redução do viés de seleção e a melhoria da validade interna dos estudos observacionais.

Outro estudo importante, conduzido por Austin (2011), enfatiza a utilidade do PSM em ajustar para covariáveis de confusão não observadas, afirmando que "o PSM pode ser uma ferramenta valiosa para controlar vieses não observados em estudos observacionais." Isso é

particularmente relevante quando informações detalhadas sobre todas as variáveis de confusão não estão disponíveis.

Essa técnica envolve a criação de um escore de propensão (*propensity score*), que é uma estimativa da probabilidade de um indivíduo ou unidade ter recebido um tratamento ou exposição com base em suas características observáveis (covariáveis)

Rosenbaum e Rubin (1983), definem esta técnica (*Propensity Score Matching*) como sendo a probabilidade condicional de exposição levando em conta um grupo de variáveis, determinando as que recebem o tratamento, considerando previamente as características relacionadas ao tratamento onde  $T = \{0,1\}$  representa os indivíduos tratados e não tratados, onde, para  $T = 1$  o indivíduo está no grupo dos tratados onde neste estudo correspondera a alunos oriundos de escola privada, se  $T = 0$  o indivíduo está no grupo dos não tratados que será assumido por estudantes de escolas públicas nas regressões do PSM. Em seguida, os indivíduos são emparelhados ou combinados de forma que os grupos de tratamento e controle tenham escores de propensão semelhantes, tornando-os comparáveis em termos dessas características observáveis. Isso permite a avaliação do efeito causal do tratamento, controlando potenciais vieses de seleção em estudos observacionais. Portanto, o PSM é uma ferramenta útil para estimar o efeito causal quando não é possível realizar um experimento randomizado.

O *propensity score matching* pode ser representado pela equação 04:

$$\ln\left(\frac{P(Y=1)}{P(Y=0)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (04)$$

Onde:

- $P(Y=1)$ : é a probabilidade de pertencer ao grupo de tratamento.
- $P(Y=0)$ : é a probabilidade de pertencer ao grupo de controle.
- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ : são coeficientes a serem estimados.
- $X_1, X_2, \dots, X_n$ : são as variáveis observadas.

O escore de propensão (PS) para cada indivíduo é calculado usando a equação de regressão logística:

$$PS = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (05)$$

O escore de propensão é uma estimativa da probabilidade de pertencer ao grupo de tratamento, que, neste caso, significa ser de instituição privada ou não. Após o cálculo dos escores de propensão, será utilizado o método de emparelhamento de *nearest neighbor*, no qual cada unidade no grupo de tratamento é emparelhada com uma unidade no grupo de controle que possui o escore de propensão mais próximo. Em outras palavras, a unidade tratada é associada à unidade de controle que possui o escore de propensão mais semelhante. Com os grupos emparelhados, as comparações entre o grupo de tratamento e controle são realizadas com base nessas unidades emparelhadas, auxiliando no controle das diferenças nas características observáveis entre os grupos.

A metodologia utilizada no *propensity score matching* permite outras abordagens além do *nearest*, tais como:

- *Kernel Matching*: Método que suaviza a estimativa do escore de propensão utilizando uma função *kernel*. Este método pondera as observações no grupo de controle com base na proximidade no espaço de características.
- *Stratification*: Envolve a divisão dos dados em estratos (grupos) com base nos escores de propensão. As comparações são, então, realizadas dentro de cada estrato.
- *Exact Matching*: Neste método, as unidades do grupo de tratamento são emparelhadas exatamente com as unidades correspondentes no grupo de controle, considerando características idênticas. Encontrar correspondências exatas pode ser desafiador, especialmente em conjuntos de dados extensos.

A escolha do método “*nearest neighbor*” em detrimento de outros métodos reside na sua flexibilidade, pois pode ser aplicado tanto em conjuntos de dados grandes quanto pequenos. Este método não requer a especificação de uma função de *kernel* ou estratificação. Em comparação com métodos paramétricos, o “*nearest neighbor*” é menos sensível a certas suposições sobre a distribuição dos dados, tornando-o útil quando tais suposições podem ser questionáveis. Além disso, destaca-se pela sua capacidade de ajuste dinâmico às características específicas dos dados, adaptando-se melhor a padrões complexos e não lineares.

#### 4.3. Estratégia empírica

De acordo com Souza *et al.* (2021, p. 07), os principais determinantes para o desempenho no ENEM incluem a escolaridade dos pais, a faixa de renda familiar, a idade, a raça, o sexo, a

dependência administrativa das escolas, a renda da família, a alta escolaridade da mãe e o acesso à internet. Na Tabela 05 é apresentado dados descritivos sobre os microdados do ENEM em 2019 e 2022.

Pode-se destacar alguns indicadores socioeconômicos que fornecem *insights* empíricos sobre as desigualdades e principais características dos estudantes de ensino público em comparação com os estudantes de ensino privado. Inicialmente, foi observada a diferença nas notas dos estudantes no ano de 2019, que foi de 71,8 pontos. Houve uma queda inexplicável nessa diferença, que se reduziu para 48,82 em 2022. Isso indica que, mesmo após três anos, a disparidade nas notas entre os estudantes de escolas públicas e privadas ainda persiste, com os estudantes de escolas privadas obtendo notas mais altas. Essa diferença pode ser explicada por variáveis socioeconômicas, como o acesso à internet, em ambos os anos, observamos uma maior porcentagem de estudantes do ensino superior com acesso à internet. Em 2019, a porcentagem era de 78,55% para os estudantes de escolas públicas e 96,77% para os alunos de escolas particulares.

Em relação a raça temos em 2019, 61,75% dos negros estavam matriculados em escolas públicas, enquanto em 2022 esse número caiu para 60,09%. Já a porcentagem de negros em escolas privadas aumentou de 35,08% em 2019 para 28,91% em 2022. Por outro lado, a porcentagem de não-negros em escolas públicas passou de 38,25% em 2019 para 39,91% em 2022, enquanto a presença em escolas privadas aumentou de 64,92% para 71,09% no mesmo período. Esses dados sugerem uma tendência oposta à observada entre os negros, com um aumento da presença de não-negros em escolas privadas e uma leve variação nas escolas públicas.

Além disso, deve-se destacar a influência do nível de educação do pai e da mãe na nota dos estudantes. Em 2019, 47,12% dos pais de estudantes de escolas públicas possuíam ensino médio ou superior, enquanto 53,67% das mães possuíam o mesmo nível de educação. No entanto, os dados revelam que, no mesmo ano, 82,34% dos pais de estudantes de escolas privadas tinham ensino médio ou superior, e 88,81% das mães tinham o mesmo nível de educação.

Em relação à posse de computadores, observamos uma marcante discrepância nos dados. Em 2019, a posse de computadores era de 47,77% para escolas públicas e 84,59% para escolas privadas. No entanto, em 2022, a posse de computadores diminuiu para 39,99% nas escolas públicas, enquanto nas escolas privadas aumentou para 86,89%. Esses números evidenciam



uma significativa disparidade financeira entre os alunos de escolas públicas e privadas. A esmagadora maioria dos alunos provenientes de instituições privadas tem acesso a computadores, enquanto nas escolas públicas esse percentual não chega a 50%.

Outro fator significativo na disparidade das médias é a renda familiar dos estudantes. Em 2019, apenas 18,63% dos estudantes de escolas públicas tinham uma renda familiar igual ou superior a dois salários mínimos, enquanto em 2022, esse valor aumentou para 22,46%. Em contraste, em 2019, 69,94% dos estudantes de escolas privadas tinham renda familiar igual ou superior a dois salários mínimos, e em 2022, esse valor aumentou para 79,64%. Isso evidencia que a disparidade entre os estudantes está diretamente relacionada a fatores socioeconômicos, sendo a renda o fator principal.

**Tabela 05 – Análise descritiva das variáveis socioeconômicas do Enem**

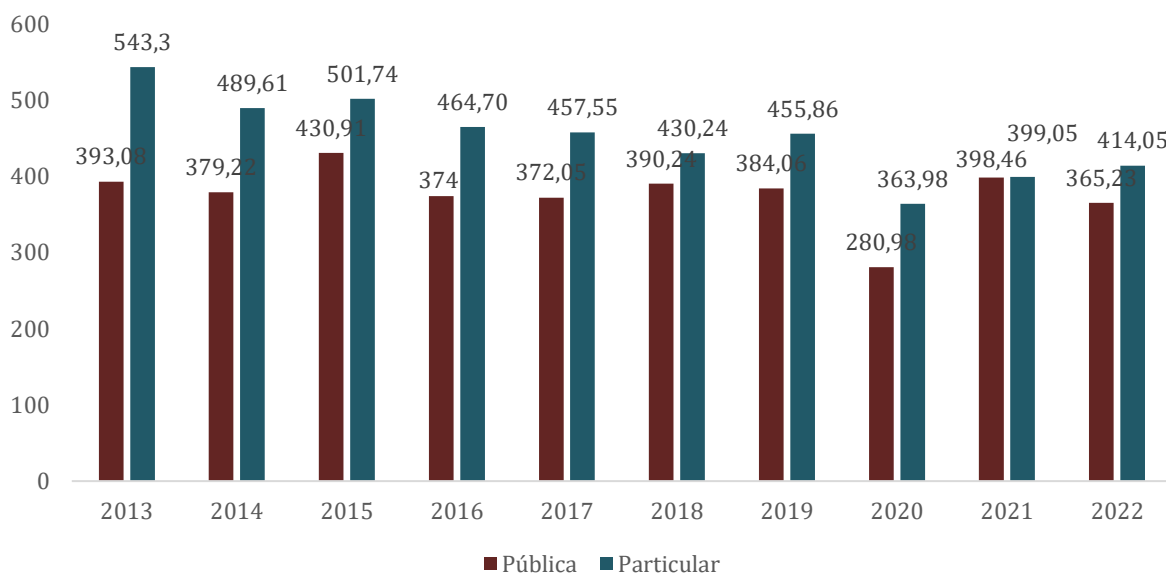
Variável	2019		2022	
	Pública	Privada	Pública	Privada
<b>Escola</b>	<b>Pública</b>	<b>Privada</b>	<b>Pública</b>	<b>Privada</b>
<b>Notas médias</b>	384,06	455,86	365,23	414,05
<b>Mulheres</b>	66,25%	62,65%	66,58%	63,09%
<b>Homens</b>	33,75%	37,35%	33,42%	36,91%
<b>Negros (pardos + pretos)</b>	61,75%	35,08%	60,09%	28,91%
<b>Não-Negros</b>	38,25%	64,92%	39,91%	71,09%
<b>Possui acesso a internet</b>	78,55%	96,77%	88,84%	98,92%
<b>Possui celular</b>	97,33%	98,67%	96,98%	98,35%
<b>Possui computador</b>	47,77%	84,59%	39,99%	86,89%
<b>Concluiu o ensino médio</b>	66%	65%	54,27%	54,28%
<b>Escola de origem</b>	87,60%	12,40%	89,26%	10,74%
<b>Casado(a)</b>	1,39%	0,27%	0,59%	0,24%

<b>Pai concluiu o ensino médio</b>	47,12%	82,34%	51,10%	51,85%
<b>Mãe Concluiu o ensino médio</b>	53,67%	88,81%	58,67%	92,14%
<b>Renda familiar igual ou superior a 2 salários mínimos</b>	18,63%	69,94%	22,46%	79,64%
<b>Cargo do pai exige ensino médio ou superior</b>	29,76%	74,22%	32,33%	76,26%
<b>Cargo da mãe exige ensino médio ou superior</b>	27,70%	72,60%	31,04%	75,70%

Fonte: Elaboração própria a partir dos Microdados ENEM 2019 e 2022.

A disparidade nas notas entre estudantes de escolas públicas e privadas resulta de diversas características observáveis e não observáveis, como discutido anteriormente. As notas são apenas o reflexo de um conjunto de características intrínsecas do candidato. A Tabela 05 apresentada demonstra que os estudantes provenientes de escolas públicas apresentam indicadores mais baixos em comparação com os estudantes de escolas privadas. Isso explica o fato de que, como mostrado na Figura 03, houve uma disparidade constante nas médias de desempenho entre os anos de 2013 a 2022.

**Figura 03 - Notas médias no ENEM por tipo de escola**



Fonte: Elaboração própria, a partir dos Microdados ENEM.

O *background* social dos estudantes pode ser utilizado como uma ferramenta importante para entender os motivos para essas diferenças encontradas nas notas dos estudantes, para

Hanushek e Luque (2003), "o ambiente socioeconômico é um determinante fundamental do desempenho escolar." Em pesquisa realizada por eles aponta que crianças de famílias com menor renda e níveis educacionais mais baixos têm maior probabilidade de enfrentar dificuldades acadêmicas. Essa disparidade é frequentemente atribuída à falta de acesso a materiais de aprendizado, suporte educacional adequado e oportunidades extracurriculares enriquecedoras.

Essa disparidade nas médias do ENEM reflete além de características socioeconômicas também desafios subjacentes presentes no sistema educacional público brasileiro, desafios significativos que impactam a qualidade da educação oferecida a milhões de estudantes em todo o país. Entre os problemas mais prementes estão a falta de infraestrutura adequada, a defasagem no currículo, a desigualdade de acesso e a falta de investimentos suficientes. A infraestrutura precária das escolas públicas é um problema persistente. Muitas instituições carecem de estruturas físicas adequadas, salas de aula superlotadas e falta de recursos básicos, como bibliotecas, laboratórios e material didático de qualidade.

A defasagem escolar é outro desafio significativo. No Brasil é uma questão persistente que afeta negativamente o desempenho escolar de muitos estudantes. Essa defasagem, que ocorre quando os alunos não alcançam o nível de aprendizado esperado para sua série ou idade, tem implicações profundas para a qualidade da educação no país. Segundo dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), em 2021, 54,2% dos alunos matriculados no ensino médio tinham idades superiores à faixa etária adequada para essa etapa de ensino.

Além disso, a falta de investimentos adequados é um problema crônico. O financiamento insuficiente resulta na falta de capacitação adequada para os professores, remuneração inadequada e falta de recursos pedagógicos. Esses fatores desencorajam muitos profissionais talentosos de seguir a carreira docente e prejudicam a qualidade do ensino.

Para enfrentar esses problemas, é essencial que o sistema educacional público brasileiro receba investimentos significativos e que políticas eficazes sejam implementadas para melhorar a infraestrutura, ocorram atualizações no currículo, promover a igualdade de acesso e apoio ao corpo docente.

#### 4.4. Dados

Nesta seção, serão apresentados os dados e as variáveis socioeconômicas que compõem este trabalho. A amostra utilizada refere-se aos microdados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), nos quais foram utilizados os dados dos exames realizados em 2019 e 2022. Os microdados do exame são obtidos por meio de um questionário não obrigatório, o qual o candidato preenche no momento da inscrição. Por esse motivo, no ano de 2019, 66,3% das observações foram desconsideradas, pois os candidatos não responderam ao questionário, o que representou 3.378.189 dos 5.095.308 inscritos. Já no ano de 2022, esse valor caiu para 59,2%, representando 2.018.532 pessoas que não responderam ao questionário dos 3.409.682 candidatos confirmados. Dessa forma, foram considerados para análise, nos anos de 2019 e 2022, respectivamente, 1.717.119 e 1.391.150 candidatos.

Aqueles que optam por responder ao questionário são submetidos a um total de 42 perguntas, as quais o INEP divide em dados dos participantes, compreendendo 10 perguntas; dados da escola do participante, compreendendo 7 perguntas; e um questionário socioeconômico com 25 perguntas, todas de múltipla escolha. Além das 42 perguntas realizadas diretamente ao estudante, são adicionadas aos microdados 37 variáveis que incluem dados das provas objetivas, dados da redação e dados do local da aplicação da prova. Por fim, são disponibilizados na seção de dados do candidato o número da inscrição do candidato e o ano de realização do exame. Para a análise da dissertação, serão utilizadas 14 variáveis descritas na Tabela 06, somadas às notas obtidas nas provas de linguagens, códigos e suas tecnologias; ciências humanas e suas tecnologias; ciências da natureza e suas tecnologias; matemática e suas tecnologias; e a redação além da média aritmética simples que foi construída a partir das cinco provas realizadas pelo candidato.

**Tabela 06 - Descrição das variáveis utilizadas nas estimações**

<b>Natureza da variável</b>	<b>Variável</b>	<b>Descrição</b>
Dependente	Escola	Variável <i>dummy</i> para identificar se o indivíduo é oriundo de escola pública = 0 ou particular = 1.

Independentes	Sexo	Variável <i>dummy</i> para representar o sexo do candidato, masculino = 0 e feminino = 1.
	Cor/Etnia	Variável <i>dummy</i> para identificar se o indivíduo é preto(a) ou pardo(a), se sim = 0, caso contrário = 1.
	Ensino	Variável <i>dummy</i> para identificar se o indivíduo realizou o ensino médio regular, se sim = 0, caso contrário = 1.
	Estado Civil	Variável <i>dummy</i> para identificar se o indivíduo é casado(a), se sim = 1, caso contrário = 0.
	Pai ensino	Variável <i>dummy</i> para indicar o nível de escolaridade do pai, se possui nível médio completo ou superior = 1, se possui qualquer formação até nível fundamental completo = 0.
	Mãe ensino	Variável <i>dummy</i> para indicar o nível de escolaridade da mãe, se possui nível médio completo ou superior = 1, se possui qualquer formação até nível fundamental completo = 0.
	Ocupação do pai	Variável <i>dummy</i> para indicar o nível de ensino do cargo do pai, se o cargo exige nível médio completo ou superior = 1, se possui qualquer formação até nível fundamental completo = 0.
	Ocupação da mãe	Variável <i>dummy</i> para indicar o nível de ensino do cargo da mãe, se o cargo exige nível médio completo ou superior = 1, se possui qualquer formação até nível fundamental completo = 0.
	Residentes na casa	Variável <i>dummy</i> para indicar se o indivíduo reside sozinho = 1, senão = 0.
	Renda domiciliar	Variável <i>dummy</i> para indicar se a renda domiciliar é de 2 salários mínimos ou maior = 1, senão = 0.
	Celular	Variável <i>dummy</i> para indicar se na residência possui 1 ou mais celulares = 1, senão = 0.

---

Computador *Dummy* para indicar se na residência possui 1 ou mais computadores = 1, senão = 0.

Acesso à internet *Dummy* para identificar se possui acesso a internet, se sim = 1, senão = 0.

---

Fonte: Elaboração própria a partir dos microdados ENEM 2019 e 2022

## 5. RESULTADOS

Nesta seção serão apresentados os resultados obtidos pelos métodos estatísticos discutidos no capítulo Metodologia. Os resultados serão apresentados na ordem Mínimos Quadrados Ordinários (MQO) e *Propensity Score Matching* (PSM).

### 5.1. Mínimos Quadrados Ordinários (MQO)

O Modelo de Mínimos Quadrados Ordinários (MQO) é uma técnica amplamente utilizada na estatística e na econometria para estimar os parâmetros de um modelo estatístico, quando se deseja encontrar a "melhor" linha de ajuste em um conjunto de dados. Essa técnica é particularmente importante em análises de regressão, onde procuramos entender e modelar as relações entre variáveis independentes e dependentes. No caso deste estudo a variável dependente serão as notas obtidas por prova segmentada pelas regiões Sudeste e Nordeste já a variáveis de interesse correspondem a escola de origem (privada ou pública) e as variáveis de controle são sexo, estado civil, etnia, tipo de instituição que concluiu o ensino médio, nível de educação do pai e da mãe, ocupação do pai, ocupação da mãe, quantidade de pessoas que moram na sua residência, possui celular, possui computador e possui acesso à internet. A estimação é representada através da equação a seguir:

$$\begin{aligned} \text{Nota} = & \beta_0 + \beta_1\text{Escola} + \beta_2\text{Sexo} + \beta_3\text{EstadoCivil} + \beta_4\text{Etnia} + \beta_5\text{Ensino} + \beta_6\text{PaiEnsino} + \\ & \beta_7\text{MãeEnsino} + \beta_8\text{OcupaçãoPai} + \beta_9\text{OcupaçãoMãe} + \beta_{10}\text{Residentes} + \beta_{11}\text{Renda} + \beta_{12}\text{Celular} + \\ & \beta_{13}\text{Computador} + \beta_{14}\text{Internet} + \varepsilon \quad (08) \end{aligned}$$

As regressões completas, juntamente com os coeficientes das variáveis de controle, estão disponíveis no capítulo do apêndice. No item 8.4, são apresentados os indicadores de ajuste das regressões, que incluem informações como  $R^2$ ,  $R^2$  ajustado, erro padrão residual e estatística F. Já no item 8.6, estão listados os valores dos coeficientes de todas as variáveis utilizadas na modelagem, conforme descrito pela Equação 08.

Na Tabela 07, temos os resultados da correlação entre escola de origem e a nota obtida pelo candidato por prova em 2019 na região Nordeste e Sudeste, onde podemos visualizar, a priori, uma ligação positiva entre a nota e o candidato ser oriundo de uma instituição privada para todas as provas nas duas regiões. Isso demonstra, como sugerido pela literatura, que candidatos de instituição privada possuem vantagens em relação àqueles vindos de uma escola pública, devido ao arcabouço escolar superior. Quando analisamos as evidências, observamos um coeficiente ligeiramente superior na região Sudeste em linguagens, sendo apenas 0,3968 maior do que na região Nordeste. Isso não indica uma diferença significativa. No entanto, ao examinar as notas em matemática, observa-se uma disparidade notável no coeficiente entre o Sudeste e o Nordeste, com uma diferença de 9,4284. Isso sugere uma correlação mais forte entre escolas privadas e notas de matemática na região Sudeste em comparação com o Nordeste. O mesmo padrão é observado para ciências da natureza, apresentando uma diferença de 3,5710. Em relação à prova de ciências humanas, a diferença é insignificante, sendo apenas 0,4271. Por fim, o maior coeficiente de correlação entre a nota obtida e a escola privada é encontrado na redação. A região do Nordeste possui um coeficiente superior, com uma diferença de 5,0.

**Tabela 07 – Regressões modelo MQO 2019**

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>Escola (Nordeste)</b>	44,13548*** (0.29)	78,1458*** (0.60)	62,29337*** (0.39)	56,1235*** (0.37)	114,6396*** (0.83)
<b>Escola (Sudeste)</b>	44,53230*** (0.21)	87,5742*** (0.45)	65,86443*** (0.31)	56,55063*** (0.80)	109,6391*** (0.63)

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Total de observações: 99.583 (Nordeste) e 137.558 (Sudeste).

P-valor: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ .

(Erro padrão)

Os coeficientes da Tabela 07 evidenciam uma forte correlação para o ano de 2019 entre a escola privada e a nota obtida em todas as provas nas duas regiões estudadas. No entanto, é importante destacar que em algumas provas há uma diferença significativa no tamanho do coeficiente quando comparadas entre as regiões, como ocorre na prova de matemática e redação. Com relação ao ano de 2022, os coeficientes representados na Tabela 12 indicam um aumento em todas as provas quando comparados ao ano de 2019. Isso sugere um possível aumento nas disparidades entre escolas públicas e privadas, tanto na região Sudeste quanto na Região Nordeste, devido ao aumento do coeficiente de correlação entre escolas privadas e as provas. Ao analisar a prova de linguagens, observa-se uma discrepância modesta de 2,23, com uma correlação mais forte para o Nordeste. Isso pode indicar uma piora no desempenho dos estudantes de escolas públicas, afetados de forma mais severa pela pandemia. Em relação à

prova de matemática, a disparidade continua sendo pequena, apenas 1,19. No entanto, houve um aumento significativo em comparação com 2019, com o coeficiente do Nordeste subindo 21,24 e o do Sudeste 13,01.

Ao observar o coeficiente de correlação entre a escola de origem e a prova de ciências da natureza, o Sudeste apresenta o maior valor, gerando uma discrepância de 3,56 em relação ao Nordeste. Comparando com os coeficientes da Tabela 07, nota-se um aumento percentual de 21,66% no coeficiente do Nordeste, uma diferença numérica de 13,48. O Sudeste teve um aumento de 13,48 pontos no coeficiente de correlação, representando um aumento de 20,47%. Na área de ciências humanas, o Nordeste possui um coeficiente superior, com uma diferença de 3,228 em relação ao Sudeste. Em termos percentuais, o Nordeste apresentou um crescimento de 31,72%, com uma diferença numérica de 20,961 entre 2019 e 2022. No Sudeste, houve um aumento percentual de 30,60%, representando um acréscimo de 17,306 em valor numérico. Ao analisar a correlação da redação, observa-se os maiores coeficientes, com uma diferença entre as regiões de 21,114. Uma análise segmentada revela que o coeficiente do Nordeste teve um aumento percentual de 32,14%, enquanto o Sudeste registrou um aumento de 18,91%.

**Tabela 08 – Regressões modelo MQO 2022**

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>Escola (Nordeste)</b>	65,52037*** (0.40)	99,3906*** (0.64)	75,7793*** (0.41)	77,0846*** (0.41)	151,4895*** (0.90)
<b>Escola (Sudeste)</b>	63,2865*** (0.28)	100,5867*** (0.47)	79,3421*** (0.33)	73,8565*** (0.29)	130,3755*** (0.63)

Fonte: Elaboração própria a partir dos Microdados ENEM 2022.

Total de observações: 142.622 (Nordeste) e 126.423 (Sudeste).

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

Em síntese, é possível observar, por meio dos resultados apresentados nas Tabelas 07 e 08, uma forte correlação entre escolas privadas e um desempenho satisfatório nas provas do ENEM. Há um aumento significativo no ano de 2022 em comparação com 2019. Esses resultados destacam as distorções existentes na educação brasileira, as quais se tornam ainda mais acentuadas entre regiões com disparidades econômicas, como é o caso do Nordeste e Sudeste. Essa disparidade foi intensificada no pós-pandemia, como evidenciado pelo aumento em todas as provas da correlação entre escolas privadas e as notas. Esse cenário aponta para um incremento na discrepância entre as instituições no período pós-pandêmico.



## 5.2. *Propensity Score Matching* (PSM)

Os modelos de mínimos quadrados ordinários (MQO) enfrentam desafios significativos em muitos cenários já discutidos no capítulo metodologia. Um dos principais problemas é o viés sistemático na seleção de escolas, o MQO assume que as escolhas das escolas são aleatórias e independentes. No entanto, em muitos casos, a escolha da escola pode depender de fatores não observados, como a situação socioeconômica dos pais. Isso pode introduzir viés nas estimativas, pois diferentes escolas podem atrair alunos com características distintas. Além disso, existe a problemática da heterogeneidade não observada entre as escolas pois o MQO assume que todas as escolas têm efeito semelhante sobre as notas dos alunos. Entretanto, escolas públicas e privadas podem diferir em diversos aspectos, como infraestrutura, qualidade do corpo docente e recursos disponíveis. Ignorar tais diferenças pode levar a uma estimativa distorcida do verdadeiro efeito da escola na pontuação do Enem.

Por outro lado, o uso do *propensity score matching* (PSM) oferece vantagens significativas, para tratar possível viés enfrentado pela amostra. Se trata de uma abordagem estatística que permite no momento da modelagem controlar variáveis latentes não observáveis, o que pode ajudar a reduzir o viés na análise. Portanto, o PSM é especialmente útil quando se lida com viés de mensuração, uma vez que permitem a modelagem das relações subjacentes entre as variáveis observáveis e latentes, o que pode melhorar a precisão das estimativas. Isso os torna uma escolha valiosa quando o viés é uma preocupação em sua análise de dados. Por esse motivo foi realizado neste estudo o PSM com intuito de controlar o viés escola de origem (pública ou privada) do estudante onde buscou analisar de forma isolada como os fatores socioeconômicos podem afetar de forma direta na nota do estudante sem a influência dos diversos fatores que distinguem em termos de qualidade de ensino aqueles estudantes provenientes de instituições públicas ou privadas.

Após o processo de pareamento, observou-se uma redução na quantidade de amostras. Para o ano de 2019, na região Nordeste, foram inicialmente registradas 99.583 observações no grupo Controle e 13.130 no grupo Tratado. Após o pareamento, o total foi de 26.260 observações, distribuídas igualmente entre os dois grupos. No ano de 2022, a região Nordeste apresentou 142.622 observações no grupo Controle e 11.684 no grupo Tratado. Após o pareamento, o total manteve-se em 23.368, com 11.684 observações em cada grupo.

Na região Sudeste, em 2019, houve inicialmente 137.558 observações antes do pareamento, reduzindo-se para 40.892 após o processo, com 20.446 observações em cada grupo. Já em 2022, a região Sudeste registrou 126.423 observações antes do pareamento, resultando em 41.558 observações após o processo, com 20.799 em cada grupo. A diminuição no número de observações está associada ao método de pareamento mais próximo (*nearest neighbor*), que requer um número igual de observações para os grupos tratado e não tratado a fim de realizar o pareamento de forma eficaz.

As regressões completas, juntamente com os coeficientes das variáveis de controle, estão disponíveis no capítulo do apêndice. No item 8.5, são apresentados os indicadores de ajuste das regressões, que incluem informações como  $R^2$ ,  $R^2$  ajustado, erro padrão residual e estatística F. Já no item 8.7, estão listados os valores dos coeficientes de todas as variáveis utilizadas na modelagem.

Na Tabela 09, apresentamos os resultados do PSM para as regiões Nordeste e Sudeste no ano de 2019. Para fins de comparação, também estão dispostos os resultados das regressões MQO. Ao analisarmos os coeficientes do Nordeste em 2019, observamos, em linguagens, uma diminuição singela de apenas 3,86%. Em relação à prova de matemática, constatamos um aumento também pequeno, de apenas 1,37%, no valor do coeficiente. No que diz respeito à prova de ciências da natureza, há uma diferença de 0,52% no valor do coeficiente de correlação entre o método do PSM e o MQO. Na prova de ciências humanas, a diferença entre o MQO e o PSM é de 0,11%. Por fim, destacamos a diferença na correlação da prova de redação, onde foi apresentada a maior discrepância, com o PSM apresentando um valor 6,04% maior que o MQO. corrija: Em relação ao sudeste ocorreu uma diminuição de 10,8% em linguagens quando se comparado ao MQO, para matemática ocorre uma mudança de 1,79%, em ciências da natureza a diferença entre o PSM e o MQO é de 1,84%, já em ciências humanas temos uma diferença nos valores dos coeficientes do PSM e MQO de 5,66% e por fim em redação existe uma diferença de 2,94%.

Os valores dos coeficientes no Propensity Score Matching diferem dos Mínimos Quadrados Ordinários devido às abordagens metodológicas distintas utilizadas em cada técnica. Enquanto os Mínimos Quadrados Ordinários buscam minimizar a soma dos quadrados dos resíduos, o *Propensity Score Matching* visa equilibrar as características observáveis entre os grupos tratado e de controle. Assim, essas diferenças nos coeficientes refletem a natureza específica de ajuste

para lidar com o viés de seleção inerente a análises baseadas em observações, ao contrário da abordagem mais ampla dos Mínimos Quadrados Ordinários.

**Tabela 09 – Regressões modelo PSM 2019**

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>PSM</b>	42,493***	79,221***	62,618***	56,185***	121,559***
<b>(Nordeste)</b>	(1.28)	(1.82)	(1.45)	(1.39)	(2.42)
<b>MQO</b>	44,13548***	78,1458***	62,29337***	56,1235***	114,6396***
<b>(Nordeste)</b>	(0.29)	(0.60)	(0.39)	(0.37)	(0.83)
<b>PSM (Sudeste)</b>	41,0908***	86,034***	64,668***	53,524***	112,858***
	(1.21)	(1.73)	(1.40)	(1.29)	(2.00)
<b>MQO</b>	44,53230***	87,5742***	65,86443***	56,55063***	109,6391***
<b>(Sudeste)</b>	(0.21)	(0.45)	(0.31)	(0.80)	(0.63)

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Total de observações: MQO (99.583 Nordeste e 137.558 Sudeste) e PSM (26.260 Nordeste e 40.892 Sudeste).

P-valor: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ .

(Erro padrão)

Ao realizar uma análise comparativa entre o PSM do Nordeste e do Sudeste em 2019, os coeficientes apresentados na Tabela 09 indicam uma relação positiva entre o candidato do ENEM ser oriundo de uma instituição privada e um bom desempenho em todas as provas prestadas. Destacam-se como principais diferenças os coeficientes na prova de linguagem, onde a discrepância é mínima, com apenas 1,40. As duas maiores diferenças estão na prova de redação, com 13,36, seguida de ciências humanas, com 2,66.

Ao analisar a Tabela 10, observamos os resultados do PSM para 2022 com relação às regiões do Sudeste e Nordeste. Esses resultados nos permitem identificar os principais ajustes em relação ao MQO do mesmo ano. No Nordeste, as principais disparidades entre o MQO e o PSM são observadas nos coeficientes das provas de ciências da natureza e ciências humanas, com diferenças percentuais de 9,59% e 7,22%, respectivamente. No entanto, a menor diferença foi encontrada na prova de redação, com apenas 3,07%. Em relação à região Sudeste, a menor diferença foi observada na prova de redação, com uma diferença percentual de 2,48%. As duas provas com a maior diferença de correlação entre os modelos foram ciências humanas e linguagem, com diferenças de 9,58% e 9,56%, respectivamente. Ao comparar as diferenças entre as regiões, a principal discrepância entre os coeficientes do PSM é encontrada na redação, onde a diferença é de 13,36 pontos. Por outro lado, a menor diferença entre os coeficientes é observada em matemática, com apenas 0,76 pontos de diferença.

**Tabela 10 – Regressões modelo PSM 2022**

Método	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
--------	-----------	------------	----------------------	------------------	---------

<b>PSM</b>	60,4379***	94,743***	69,1477***	71,8889***	146,966***
<b>(Nordeste)</b>	(1.76)	(2.32)	(1.88)	(1.82)	(3.00)
<b>MQO</b>	65,52037***	99,3906***	75,7793***	77,0846***	151,4895***
<b>(Nordeste)</b>	(0.40)	(0.64)	(0.41)	(0.41)	(0.90)
<b>PSM (Sudeste)</b>	57,767***	93,976***	72,438***	67,3979***	133,601***
	(1.49)	(1.94)	(1.59)	(1.53)	(2.29)
<b>MQO</b>	63,2865***	100,5867***	79,3421***	73,8565***	130,3755***
<b>(Sudeste)</b>	(0.28)	(0.47)	(0.33)	(0.29)	(0.63)

Fonte: Elaboração própria a partir dos Microdados ENEM 2022.

Total de observações: MQO (142.622 Nordeste e 126.423 Sudeste) e PSM (23.368 Nordeste e 41.558 Sudeste).

P-valor: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ .

(Erro padrão)

## 6. CONSIDERAÇÕES FINAIS

O intuito do trabalho, assim como discutido no capítulo de introdução, é investigar as possíveis influências das variáveis socioeconômicas antes e depois da pandemia do COVID-19 nas notas dos estudantes do Nordeste e Sudestes, utilizando como ferramenta de estimação modelos econométricos como forma de estimar tais influências. O que podemos visualizar através dos resultados é uma forte discrepância nos resultados entre o Nordeste e o Sudeste, que se intensificou ainda mais com a pandemia. Ao observar a Figura 02, é notável a presença de médias mais elevadas na região Sudeste em comparação com a região Nordeste. No entanto, ao analisar o mapa para o ano de 2022, observa-se uma piora nas médias gerais em todos os estados. O impacto da pandemia foi mais severo na região Nordeste, com estados como o Rio Grande do Norte, que possuía a melhor média nacional, sofrendo uma queda extrema na média. Em contraste, estados da região Sudeste tiveram perdas menos severas.

Ao analisar as Tabelas 01 a 04, fica evidente que as maiores distorções estão entre as escolas públicas. Nas Tabelas 01 e 03, as diferenças de médias em escolas particulares são mínimas e não significativas entre as regiões Nordeste e Sudeste. No entanto, isso não se aplica ao comparar as diferenças em escolas públicas por região, como demonstrado na Tabela 02, onde há uma diferença significativa de 6,62 pontos a mais na região Sudeste em comparação com o Nordeste, com um nível de significância elevado no Teste-t. Essa discrepância se intensificou com o impacto da pandemia. A nota média das escolas públicas do Sudeste agora é 21,52 pontos mais alta do que a média das escolas públicas do Nordeste. Esses dados evidenciam que já existia uma disparidade entre as regiões (Nordeste e Sudeste), a qual foi acentuada pela pandemia do COVID-19. Embora tenha afetado todos os estados, o impacto foi mais severo no Nordeste em virtude de fatores socioeconômicos. A região Sudeste, com uma renda per capita

mais elevada, foi impactada de maneira menos intensa. Essa disparidade também se reflete nas instituições mais vulneráveis economicamente, como as escolas públicas.

Quando analisado os resultados das modelagens estatísticas podemos ver a forte influência do *background* socioeconômico do participante como explicação para as desigualdades de notas. Quando analisado os resultados do PSM podemos destacar a influência de diversas variáveis de controle como sexo em que, pode-se destacar como resultado a relação negativa entre nota nas ciências da natureza e matemática e o gênero feminino, porém uma correlação positiva entre o gênero feminino e as notas nas demais provas, apontando que existe uma desigualdade de gênero quanto as ciências exatas esses resultados estão presentes tanto na região Sudeste quanto Nordeste, onde ocorreu uma redução do tamanho dos coeficientes de 2019 em relação a 2022. Diversos pesquisadores têm observado essa discrepância, demonstrando que, embora meninas e meninos possam começar com desempenhos semelhantes nessas áreas, as diferenças se tornam mais pronunciadas à medida que avançam nos estudos, sendo o ENEM uma ferramenta interessante para tal análise tendo em vista que avalia estudantes que estão no final do ensino médio ou já terminaram. Autores como Carol S. Dweck (2006), em seu livro "*Mindset: The New Psychology of Success*", apontam que fatores sociais e culturais desempenham um papel significativo nessa disparidade. Ela destaca como estereótipos de gênero e expectativas sociais podem influenciar o desempenho das meninas em matérias como matemática e física. Além disso, estudos como o de Hyde, Fennema e Lamon (1990), publicado no periódico "*Psychological Bulletin*", examinaram diferenças de gênero em habilidades matemáticas e encontraram que, embora as diferenças sejam pequenas, elas são amplificadas ao longo do tempo devido a variáveis sociais e ambientais.

Outro resultado que pode ser destacado se trata das variáveis de nível de escolaridade dos pais em que através dos coeficientes estimados pelo PSM é visto que estudantes com pais com maior nível educacional tendem a ter uma nota melhor com destaque para os coeficientes de escolaridade da mãe que possui maiores influências nas notas que do pai, indo ao encontro do previsto na literatura. Segundo Richard E. Nisbett (2009) no livro "*Intelligence and How to Get It: Why Schools and Cultures Count*", há uma considerável evidência de que o nível educacional dos pais está positivamente relacionado com o sucesso acadêmico dos filhos. Nisbett (2009) argumenta que o ambiente educacional proporcionado pelos pais, incluindo conversas intelectualmente estimulantes, leitura regular e apoio na resolução de problemas, contribui significativamente para o desenvolvimento cognitivo das crianças.

Outros resultados obtidos incluem a forte correlação entre renda e as notas dos estudantes principalmente na região nordeste onde os coeficientes estimados pelo PSM são maiores em todas as provas no Nordeste que no Sudeste onde estudantes com uma família de maior renda tendem a ter um melhor desempenho, evitando o abandono dos estudos para trabalhar ou realizar trabalho juntamente aos estudos, acesso a uma estrutura socioeconômica mais flexível e diversificada e por último a melhor qualidade de ensino da região sudeste em que mesmo estudantes menos abastados tem a possibilidade de ter melhor desempenho se comparado a alunos com a mesma situação financeira moradores do Nordeste. Portanto, pode ser visto de forma evidente, tanto através da estimação dos Mínimos quadrados ordinários quanto pelo *Propensity score matching*, que as variáveis independentes socioeconômicas são fortemente influentes na nota dos estudantes.

Por conseguinte, fica evidente que estudantes de escola privada e alunos da região Sudeste quando comparados com alunos de escola pública e alunos da região Nordeste possuem, além de uma estrutura socioeconômica favorecida uma qualidade de ensino que tem importância fundamental no seu desempenho nas cinco provas do exame. Essa desigualdade de acesso a recursos educacionais é abordada por Amartya Sen em seu livro "Desenvolvimento como Liberdade", no qual ele argumenta que a educação não deve ser vista apenas como uma mercadoria, mas como uma ferramenta para ampliar as liberdades individuais. No contexto brasileiro, a disparidade entre escolas públicas e privadas limita o acesso igualitário a essas liberdades, perpetuando um ciclo de desigualdade que se reflete em diversas esferas sociais, e o Enem é apenas um reflexo dessas desigualdades, evidenciado pelas estimativas obtidas neste estudo. Onde a pandemia de COVID-19 foi um fator preponderante no aumento da desigualdade, em que estudantes com melhores condições socioeconômicas conseguiram obter um desempenho ainda mais distante daqueles menos favorecidos. Isso evidencia a necessidade ainda maior da continuidade e promoção de políticas públicas que busquem mitigar as desigualdades educacionais enfrentadas por estudantes desfavorecidos.

## 7. REFERÊNCIAS

Agresti, A. (2013). *Categorical Data Analysis*. Wiley.

Albernaz, A. L., ... (2002). Igualdade de oportunidades e desempenho educacional: uma análise do ensino fundamental no Brasil. *Revista Brasileira de Economia*, 56(3), 373-392.

Almeida, Rafael Toledo Costa de. *Análise de fatores associados ao desempenho dos participantes do ENEM 2018 utilizando o sparklyr*. 2021. 59 f. TCC (Graduação em Estatística) Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador, BA, 2021.

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399-424.

Bourdieu, P. (1996). *A Reprodução: Elementos para uma Teoria do Sistema de Ensino*. Rio de Janeiro: Francisco Alves.

Bourguignon, F., Ferreira, F. H. G., & Menendez, M. (2007). Inequality of opportunity in Brazil. *Review of Income and Wealth*, 53:585–618.

BRITO, W. H. de; PEDROSO, F. P. Impactos de variáveis socioeconômicas no desempenho no Enem no primeiro biênio da pandemia de Covid-19. *Revista de Gestão e Avaliação Educacional*, [S. l.], p. e84069, 2023. DOI: 10.5902/2318133884069. Disponível em: <https://periodicos.ufsm.br/regae/article/view/84069>. Acesso em: 11 fev. 2024.

COUTO, Gleiber; PRIMI, Ricardo. Teoria de resposta ao item (TRI): Conceitos elementares dos modelos para itens dicotômicos. *Boletim de Psicologia*, São Paulo, v. 61, n. 134, 2011.

DA SILVA XAVIER BARROS, A. Vestibular e Enem: um debate contemporâneo. v. 22, p. 1057–1090, 2014.

Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. Random House.

Elias de Souza, A. .; De Sá Santos, L. M.; De Menezes Larruscaim, I.; Da Nobrega Besarria, C. Determinantes do Desempenho no Enem na Região Nordeste: Uma Análise de Dados em Painel do Período de 2015 a 2019. *Revista Brasileira de Estudos Regionais e Urbanos*, [S. l.], v. 15, n. 4, p. 690–711, 2022. DOI: 10.54766/rberu.v15i4.915. Disponível em: <https://revistaaber.emnuvens.com.br/rberu/article/view/915>. Acesso em: 22 out. 2023.

Feijó, Janaína Rodrigues; França, João Mário Santos de. Diferencial de desempenho entre jovens das escolas públicas e privadas. *Estudos Econômicos (São Paulo)*, São Paulo, v. 51, n. 2, p. 373-408, abr.-jun. 2021. DOI: <http://dx.doi.org/10.1590/0101-41615126jffj>.

Ferreira, F. H. G., & Gignoux, J. (2011). The measurement of educational inequality. *Review of Income and Wealth*, 57(4), 622-657.

Ferreira, F. H. G., ... (2008). Towards an understanding of socially-inherited inequalities in educational achievement: Evidence from Latin America and the OECD. Background paper, World Bank, Washington DC.

FIGUEIRÊDO, E.; NOGUEIRAY, L.; SANTANAZ, F. L. Igualdade de Oportunidades: Analisando o Papel das Circunstâncias no Desempenho do ENEM. *Revista Brasileira de Economia*, v. 68, n. 3, p. 373-392, 2014.

Gamboa, L. F., & Waltenberg, F. D. (2012). Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006-2009. *Economics of Education Review*, 31(5), 694-708.

Goldin, C., & Katz, L. F. (2008). *The Race Between Education and Technology*. Harvard University Press.

Greene, W. H. (2012). *Econometric Analysis*. Pearson Education.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and Equity in Schools around the World. *Economics of Education Review*, 22(5), 481-502.

Histórico de inscritos no Exame Nacional do Ensino Médio (ENEM). INEP. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/noticias/enem/historico-de-inscritos-no-exame-nacional-do-ensino-medio>. Acesso em: 05 ago. 2023.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107(2), 139-155.



McCulloch, C. E. (2008). Variance and bias in the error rates of the OLS estimator. *Journal of Econometrics*, 146(1), 151-155.

MONROY-GÓMEZ-FRANCO, L.; VÉLEZ-GRAJALES, R.; LÓPEZ-CALVA, L. F. The potential effects of the COVID-19 pandemic on learnings. *International journal of educational development*, v. 91, n. 102581, p. 102581, 2022.

NISBETT, Richard E. *Intelligence and How to Get It: Why Schools and Cultures Count*. W. W. Norton & Company, 2009.

Roemer, J. E. (1998). *Equality of opportunity*. Harvard University Press, Cambridge.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Santos, Mariano. Família, escola e desempenho dos alunos. *Economia Ensaios*, Uberlândia, v. 33, n. 1, p. 239-270, jul./dez. 2018.

Schultz, T. W. (1961). Investment in Human Capital: The Role of Education and of Research. *The Journal of Political Economy*, 70(5), 9–49.

Sen, A. (2000). *Desenvolvimento como Liberdade*. São Paulo: Companhia das Letras.

WEBER NETO, N. .; C. SOARES, R. .; REIS COUTINHO, L. .; SOARES TELES, A. . A Pandemia da COVID-19 impactou o ENEM? Uma Análise Comparativa de Dados dos Anos de 2019 e 2020. *Revista Novas Tecnologias na Educação*, Porto Alegre, v. 20, n. 1, p. 223–232, 2022. DOI: 10.22456/1679-1916.126655. Disponível em: <https://seer.ufrgs.br/index.php/renote/article/view/126655>. Acesso em: 11 fev. 2024.

Weber Neto, Nelson; Soares, Raimundo C.; Coutinho, Luciano R.; Teles, Ariel S.. Análise Exploratória de Dados para Identificar o Impacto da Pandemia da Covid-19 no Enem dos estados do Ceará, Maranhão e Piauí. In: *Escola Regional de Computação do Ceará, Maranhão e Piauí (ERCEMAPI)*, 10. , 2022, São Luís. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 31-40. DOI: <https://doi.org/10.5753/ercemapi.2022.225842>.

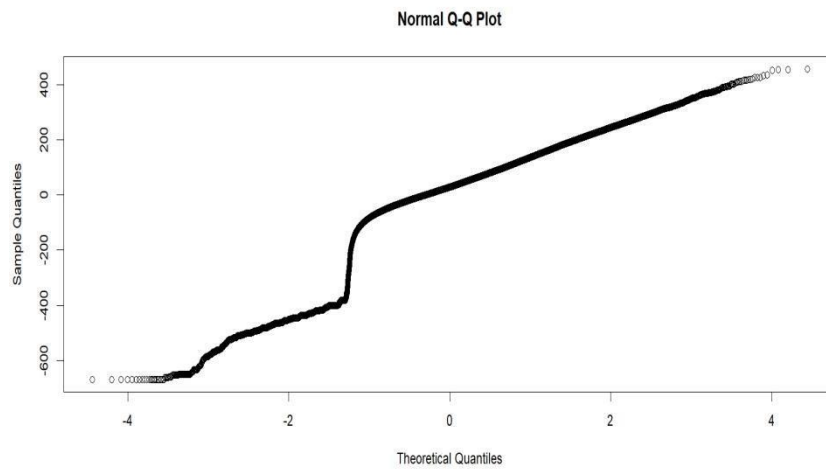
Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning.

## 8. APÊNDICE

### 8.1. Testes de normalidade dos resíduos

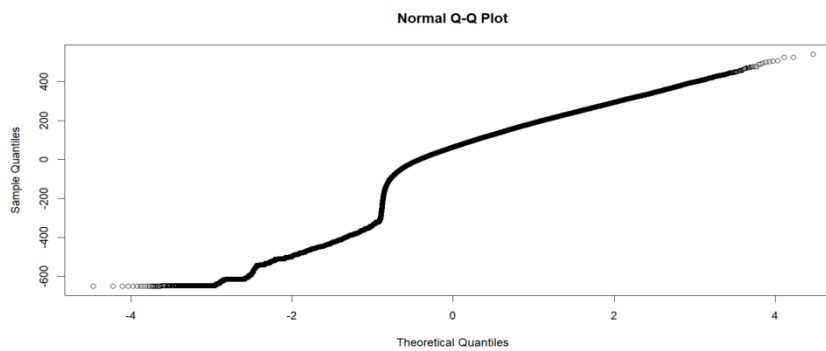
- QQnorm

**Figura 04 - Gráfico QQ (quantile-quantile) Microdados ENEM 2019**



Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

**Figura 05 - Gráfico QQ (quantile-quantile) Microdados ENEM 2022**



Fonte: Elaboração própria a partir dos Microdados ENEM 2022.

- Kolmogorov-Smirnov

**Tabela 11 - Teste de Kolmogorov-Smirnov**

<b>Prova</b>	<b>Nordeste 2019</b>	<b>Sudeste 2019</b>	<b>Nordeste 2022</b>	<b>Sudeste 2022</b>
<b>LC</b>	0,67903***	0,72895***	0,68657***	0,71528***
<b>MT</b>	0,60656***	0,63739***	0,65999***	0,6549***
<b>CN</b>	0,65512***	0,68362***	0,72804***	0,70592***
<b>CH</b>	0,62803***	0,67039***	0,69755***	0,71585***
<b>RD</b>	0,59629***	0,6045***	0,61492***	0,63199***

Fonte: Elaboração própria a partir dos Microdados do ENEM.

- Anderson-Darling

**Tabela 12 - Teste de Anderson-Darling**

<b>Prova</b>	<b>Nordeste 2019</b>	<b>Sudeste 2019</b>	<b>Nordeste 2022</b>	<b>Sudeste 2022</b>
<b>LC</b>	11410***	16896***	12355***	12036***
<b>MT</b>	5868,7***	7130,1***	8185,7***	6302,9***
<b>CN</b>	8930,5***	10969***	14235***	10169***
<b>CH</b>	7964,5***	11707***	13247***	12107***
<b>RD</b>	2852,9***	4053,9***	4562,2***	4716,1***

Fonte: Elaboração própria a partir dos Microdados do ENEM.

## 8.2. Multicolinariedade

- Inflação de variância (VIF)

**Tabela 13 - Inflação de variância**

<b>Variáveis</b>	<b>Nordeste 2019</b>	<b>Sudeste 2019</b>	<b>Nordeste 2022</b>	<b>Sudeste 2022</b>
<b>Escola</b>	1,42831	1,31056	1,344496	1,311749
<b>Sexo</b>	1,007847	1,017595	1,006513	1,017719
<b>Estado civil</b>	1,003889	1,006593	1,001149	1,001507
<b>Cor/raça</b>	1,048249	1,083088	1,040513	1,088322
<b>Ensino</b>	1,000482	1,000671	1,000979	1,001877

<b>Escolaridade do pai</b>	1,364701	1,311091	1,375352	1,271992
<b>Escolaridade da mãe</b>	1,362692	1,31397	1,376943	1,277479
<b>Ocupação do pai</b>	1,419178	1,337849	1,361356	1,285816
<b>Ocupação da mãe</b>	1,430425	1,377342	1,393442	1,3323
<b>Residentes na casa</b>	1,00137	1,002816	1,001756	1,002401
<b>Renda domiciliar</b>	1,552269	1,409983	1,490583	1,416719
<b>Celular</b>	1,033143	1,007385	1,032925	1,007412
<b>Computador</b>	1,397205	1,252095	1,318921	1,270733
<b>Internet</b>	1,247506	1,144004	1,106267	1,059986

Fonte: Elaboração própria a partir dos Microdados do ENEM.

### 8.3. Autocorrelação dos resíduos

- Durbin-Watson

**Tabela 14 - Teste de Durbin-Watson**

<b>Prova</b>	<b>Estimativa</b>	<b>Nordeste 2019</b>	<b>Sudeste 2019</b>	<b>Nordeste 2022</b>	<b>Sudeste 2022</b>
	<b>Autocorrelação</b>	-0,00252569	0,00384760	0,001559289	0,00489647
<b>LC</b>	<b>D-W Statistic</b>	2,005039	1,992304	1,996881	1,990178
	<i>p-value</i>	0,404	0,142	0,524	0,068
	<b>Autocorrelação</b>	0,001341193	0,00517497	0,005973988	-0,000450237
<b>MT</b>	<b>D-W Statistic</b>	1,997278	1,989646	1,988013	2,000882
	<i>p-value</i>	0,654	0,068	0,012	0,904
	<b>Autocorrelação</b>	0,003249254	0,00419135	0,003802656	0,000238555
<b>CN</b>	<b>D-W Statistic</b>	1,993486	1,991616	1,992355	1,999502
	<i>p-value</i>	0,296	0,11	0,136	0,97
	<b>Autocorrelação</b>	-0,00194446	0,00643368	0,002172653	0,005566397
<b>CH</b>	<b>D-W Statistic</b>	2,003872	1,98713	1,995654	1,988837
	<i>p-value</i>	0,514	0,012	0,388	0,038

	<b>Autocorrelação</b>	0,001037116	0,00689301	0,01096634	0,008410482
<b>RD</b>	<b>D-W Statistic</b>	1,997902	1,986206	1,978059	1,983149
	<i>p-value</i>	0,692	0,006	0	0,002

Fonte: Elaboração própria a partir dos Microdados do ENEM.

#### 8.4. Indicadores de ajustes dos Mínimos quadrados ordinários

**Tabela 15 - Nordeste 2019**

<b>Indicadores de ajuste</b>	<b>LC</b>	<b>MT</b>	<b>CN</b>	<b>CH</b>	<b>RD</b>
Erro padrão residual	21,16	39,56	27,22	26,02	56,47
R <sup>2</sup>	0,8148	0,844	0,8379	0,829	0,8099
R <sup>2</sup> ajustado	0,8147	0,844	0,8378	0,829	0,8099
Estatística F	2,307x10 <sup>4</sup>	2,539x10 <sup>4</sup>	2,618x10 <sup>4</sup>	2,351x10 <sup>4</sup>	1,931x10 <sup>4</sup>

Fonte: Elaboração própria a partir dos Microdados do ENEM.

**Tabela 16 - Sudeste 2019**

<b>Indicadores de ajuste</b>	<b>LC</b>	<b>MT</b>	<b>CN</b>	<b>CH</b>	<b>RD</b>
Erro padrão residual	20,32	40,04	28,7	25,56	54,64
R <sup>2</sup>	0,8019	0,8597	0,8334	0,8294	0,8204
R <sup>2</sup> ajustado	0,8018	0,8596	0,8334	0,8294	0,8204
Estatística F	2,766x10 <sup>4</sup>	3,651x10 <sup>4</sup>	3,203x10 <sup>4</sup>	3,058x10 <sup>4</sup>	2,567x10 <sup>4</sup>

Fonte: Elaboração própria a partir dos Microdados do ENEM.

**Tabela 17 - Nordeste 2022**

<b>Indicadores de ajuste</b>	<b>LC</b>	<b>MT</b>	<b>CN</b>	<b>CH</b>	<b>RD</b>
Erro padrão residual	28,98	43,7	28,76	29,18	63,8
R <sup>2</sup>	0,8004	0,8158	0,7923	0,7981	0,7863
R <sup>2</sup> ajustado	0,8004	0,8158	0,7922	0,7981	0,7862
Estatística F	2,882x10 <sup>4</sup>	3,061x10 <sup>4</sup>	2,891x10 <sup>4</sup>	2,88x10 <sup>4</sup>	2,359x10 <sup>4</sup>

Fonte: Elaboração própria a partir dos Microdados do ENEM.

**Tabela 18 - Sudeste 2022**

<b>Indicadores de ajuste</b>	<b>LC</b>	<b>MT</b>	<b>CN</b>	<b>CH</b>	<b>RD</b>
Erro padrão residual	26,57	42,81	30,56	27,39	56,97
R <sup>2</sup>	0,786	0,8399	0,804	0,801	0,8079
R <sup>2</sup> ajustado	0,7859	0,8398	0,804	0,8009	0,8078
Estatística F	2,249x10 <sup>4</sup>	2,935x10 <sup>4</sup>	2,476x10 <sup>4</sup>	2,465x10 <sup>4</sup>	2,261x10 <sup>4</sup>

Fonte: Elaboração própria a partir dos Microdados do ENEM.

### 8.5. Indicadores de ajustes do Propensity Score Matching (PSM)

**Tabela 19 - Nordeste 2019**

<b>Indicadores de ajuste</b>	<b>LC</b>	<b>MT</b>	<b>CN</b>	<b>CH</b>	<b>RD</b>
Erro padrão residual	102,01	145,5	115,5	111,3	193,5
R <sup>2</sup>	0,1026	0,1846	0,1484	0,1359	0,1674
R <sup>2</sup> ajustado	0,1021	0,1842	0,1479	0,1354	0,167
Estatística F	214,3	424,4	326,6	294,8	377

Fonte: Elaboração própria a partir dos Microdados do ENEM.

**Tabela 20 - Sudeste 2019**

<b>Indicadores de ajuste</b>	<b>LC</b>	<b>MT</b>	<b>CN</b>	<b>CH</b>	<b>RD</b>
Erro padrão residual	121,5	174	141,3	129,5	201,5
R <sup>2</sup>	0,6577	0,3471	0,104	0,9159	0,1294
R <sup>2</sup> ajustado	0,6545	0,3471	0,1037	0,9128	0,1291
Estatística F	205,6	454,6	338,9	294,4	434,2

Fonte: Elaboração própria a partir dos Microdados do ENEM.

**Tabela 21 - Nordeste 2022**

<b>Indicadores de ajuste</b>	<b>LC</b>	<b>MT</b>	<b>CN</b>	<b>CH</b>	<b>RD</b>
Erro padrão residual	133,7	176,5	142,5	138,4	228,1

R <sup>2</sup>	0,09563	0,1372	0,1003	0,107	0,146
R <sup>2</sup> ajustado	0,09508	0,1367	0,09976	0,1065	0,1455
Estatística F	176,4	265,3	185,9	199,9	285,1

Fonte: Elaboração própria a partir dos Microdados do ENEM.

**Tabela 22 - Sudeste 2022**

Indicadores de ajuste	LC	MT	CN	CH	RD
Erro padrão residual	150,9	196,5	161,1	154,9	232,1
R <sup>2</sup>	0,05831	0,09659	0,07556	0,06873	0,1076
R <sup>2</sup> ajustado	0,058	0,09628	0,07525	0,06842	0,1073
Estatística F	183,8	317,2	242,5	219	357,9

Fonte: Elaboração própria a partir dos Microdados do ENEM.

## 8.6. Regressões completas Mínimos Quadrados Ordinários

**Tabela 23 – Regressão do modelo MQO - Nordeste 2019**

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>Escola</b>	44,13548*** (0.29)	78,1458*** (0.60)	62,29337*** (0.39)	56,1235*** (0.37)	114,6396*** (0.83)
<b>Sexo</b>	7,92548*** (0.16)	-26,6063*** (0.31)	-8,70814*** (0.20)	-0,7856*** (0.20)	28,2337*** (0.45)
<b>Estado civil</b>	-76,2875*** (0.70)	-105,423*** (1.30)	-72,60939*** (0.89)	-81,2364*** (0.83)	-151,150*** (2.02)
<b>Cor/raça</b>	10,13053*** (0.18)	16,7933*** (0.36)	10,86809*** (0.24)	10,9389*** (0.23)	20,3229*** (0.51)
<b>Ensino</b>	-4,09830** (1.31)	-30,6736*** (2.55)	-12,41325*** (1.70)	-9,4730*** (1.65)	-39,0105*** (3.75)
<b>Escolaridade do pai</b>	5,64451*** (0.18)	5,8768*** (0.35)	4,52298*** (0.23)	5,3894*** (0.23)	5,0843*** (0.51)
<b>Escolaridade da mãe</b>	21,30154*** (0.17)	29,0775*** (0.35)	23,1870*** (0.23)	21,5779*** (0.22)	44,3969*** (0.51)
<b>Ocupação do pai</b>	0,40896* (0.20)	2,5310*** (0.40)	0,66131* (0.26)	1,0993*** (0.25)	-2,5846*** (0.57)
<b>Ocupação da mãe</b>	4,57491*** (0.20)	7,3928*** (0.41)	6,05062*** (0.27)	5,2832*** (0.26)	8,3471*** (0.58)
<b>Residentes na casa</b>	-44,9860*** (1.22)	-78,2267*** (2.39)	-55,87489*** (1.63)	-51,2186*** (1.52)	-89,7825*** (3.62)
<b>Renda domiciliar</b>	26,08322*** (0.30)	57,0145*** (0.61)	35,56046*** (0.40)	32,7926*** (0.38)	54,4932*** (0.85)
<b>Celular</b>	21,39427*** (0.39)	30,2886*** (0.74)	21,80851*** (0.50)	23,6064*** (0.50)	40,3646*** (1.13)
<b>Computador</b>	31,06479*** (0.19)	47,1637*** (0.38)	36,25933*** (0.25)	33,9055*** (0.24)	67,0156*** (0.54)

<b>Acesso à internet</b>	15,55642*** (0.18)	18,3981*** (0.36)	14,84179*** (0.24)	16,1643*** (0.23)	32,8564*** (0.53)
<b>R<sup>2</sup></b>	0,8148	0,844	0,8379	0,829	0,8099
<b>R<sup>2</sup> ajustado</b>	0,8147	0,844	0,8378	0,829	0,8099

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Total de observações: 99.583

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

**Tabela 24 – Regressão do modelo MQO - Sudeste 2019**

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>Escola</b>	44,53230*** (0.21)	87,5742*** (0.45)	65,86443*** (0.31)	56,55063*** (0.80)	109,6391*** (0.63)
<b>Sexo</b>	-1,80581*** (0.14)	-41,0291*** (0.29)	-19,74546*** (0.20)	-11,72039*** (0.18)	15,4714*** (0.41)
<b>Estado civil</b>	-64,2451*** (0.63)	-100,094*** (1.31)	-66,1180*** (0.93)	-73,75318*** (0.80)	-132,225*** (1.85)
<b>Cor/raça</b>	8,87194*** (0.13)	18,3999*** (0.28)	11,77160*** (0.19)	10,58914*** (0.17)	17,4636*** (0.39)
<b>Ensino</b>	-26,6740*** (1.20)	-60,3117*** (2.49)	-38,48605*** (1.73)	-36,99648*** (1.57)	-78,3137*** (3.53)
<b>Escolaridade do pai</b>	7,83081*** (0.14)	9,5543*** (0.31)	7,96340*** (0.21)	7,96503*** (0.19)	7,3663*** (0.43)
<b>Escolaridade da mãe</b>	19,74795*** (0.15)	31,8480*** (0.31)	23,40493*** (0.22)	20,85541*** (0.19)	34,2451*** (0.44)
<b>Ocupação do pai</b>	0,13960 (0.15)	3,3744*** (0.32)	1,09874*** (0.22)	1,66471*** (0.19)	2,5068*** (0.45)
<b>Ocupação da mãe</b>	5,02304*** (0.15)	11,9601*** (0.33)	8,02708*** (0.23)	5,62979*** (0.20)	10,1248*** (0.46)
<b>Residentes na casa</b>	-50,4150*** (1.10)	-76,6636*** (2.27)	-51,51812*** (1.62)	-56,06897*** (1.38)	-89,0867*** (3.20)
<b>Renda domiciliar</b>	27,86844*** (0.16)	54,7150*** (0.35)	36,86048*** (0.24)	33,28094*** (0.21)	61,2815*** (0.49)
<b>Celular</b>	17,37929*** (0.50)	24,4265*** (1.05)	18,42745*** (0.73)	20,58196*** (0.65)	24,2552*** (1.48)
<b>Computador</b>	29,86241*** (0.15)	51,6925*** (0.31)	40,67419*** (0.21)	35,23870*** (0.19)	63,0353*** (0.43)
<b>Acesso à internet</b>	8,19975*** (0.22)	10,8919*** (0.45)	7,52042*** (0.31)	8,64050*** (0.28)	24,1697*** (0.64)
<b>R<sup>2</sup></b>	0,8019	0,8597	0,8334	0,8294	0,8204
<b>R<sup>2</sup> ajustado</b>	0,8018	0,8596	0,8334	0,8294	0,8204

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Total de observações: 137.558

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

**Tabela 25 – Regressão do modelo MQO - Nordeste 2022**

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>Escola</b>	65,52037*** (0.40)	99,3906*** (0.64)	75,7793*** (0.41)	77,0846*** (0.41)	151,4895*** (0.90)



<b>Sexo</b>	17,31436*** (0.18)	-11,6554*** (0.28)	3,2171*** (0.18)	10,8699*** (0.18)	59,3999*** (0.43)
<b>Estado civil</b>	-82,4220*** (1.22)	-97,2104*** (1.93)	-65,1482*** (1.27)	-74,7085*** (1.22)	-150,647*** (2.95)
<b>Cor/raça</b>	17,28118*** (0.20)	22,7161*** (0.31)	15,6634*** (0.19)	16,3284*** (0.20)	32,1909*** (0.46)
<b>Ensino</b>	-36,0079*** (1.06)	-44,7012*** (1.61)	-31,9750*** (1.04)	-31,8510*** (1.06)	-62,9798*** (2.53)
<b>Escolaridade do pai</b>	4,69353*** (0.20)	2,9934*** (0.32)	1,0953*** (0.20)	3,6360*** (0.21)	0,8952 (0.48)
<b>Escolaridade da mãe</b>	27,50078*** (0.20)	35,6637*** (0.32)	26,7879*** (0.20)	26,0338*** (0.21)	54,1224*** (0.48)
<b>Ocupação do pai</b>	-5,26671*** (0.22)	-8,5273*** (0.35)	-8,5618*** (0.22)	-6,4505*** (0.22)	-14,4254*** (0.52)
<b>Ocupação da mãe</b>	-0,62692** (0.23)	0,5964 (0.36)	0,3736 (0.22)	-0,6798** (0.23)	-3,2457*** (0.53)
<b>Residentes na casa</b>	-57,6323*** (1.56)	-73,6105*** (2.43)	-53,5983*** (1.58)	-56,4744*** (1.57)	-105,185*** (3.73)
<b>Renda domiciliar</b>	34,84600*** (0.33)	57,4491*** (0.53)	37,0269*** (0.33)	36,9480*** (0.33)	67,9387*** (0.75)
<b>Celular</b>	22,63713*** (0.50)	30,5200*** (0.76)	19,3876*** (0.49)	21,1487*** (0.50)	35,4860*** (1.18)
<b>Computador</b>	45,51868*** (0.23)	61,0040*** (0.36)	49,7933*** (0.22)	47,0002*** (0.23)	82,5381*** (0.53)
<b>Acesso à internet</b>	25,37644*** (0.26)	27,9795*** (0.39)	19,0493*** (0.25)	22,2900*** (0.10)	44,6339*** (0.61)
<b>R<sup>2</sup></b>	0,8004	0,8158	0,7923	0,7981	0,7863
<b>R<sup>2</sup> ajustado</b>	0,8004	0,8158	0,7922	0,7981	0,7862

Fonte: Elaboração própria a partir dos Microdados ENEM 2022.

Total de observações: 142.622

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

**Tabela 26 - Regressão Modelo MQO - Sudeste 2022**

<b>Variáveis</b>	<b>Linguagem</b>	<b>Matemática</b>	<b>Ciências da natureza</b>	<b>Ciências humanas</b>	<b>Redação</b>
<b>Escola</b>	63,2865*** (0.28)	100,5867*** (0.47)	79,3421*** (0.33)	73,8565*** (0.29)	130,3755*** (0.63)
<b>Sexo</b>	1,4994*** (0.19)	-34,5119*** (0.32)	-17,7194*** (0.22)	-7,8883*** (0.20)	27,3536*** (0.44)
<b>Estado civil</b>	-60,303*** (1.60)	-71,6189*** (2.70)	-52,6027*** (1.90)	-56,0791*** (1.65)	-96,1201*** (3.66)
<b>Cor/raça</b>	13,0716*** (0.18)	23,6791*** (0.31)	14,5303*** (0.21)	12,9255*** (0.19)	23,4740*** (0.42)
<b>Ensino</b>	-39,392*** (1.17)	-50,5500*** (1.95)	-31,0118*** (1.37)	-40,3430*** (1.19)	-88,4011*** (2.69)
<b>Escolaridade do pai</b>	6,4383*** (0.20)	7,8066*** (0.34)	5,8411*** (0.23)	5,1239*** (0.21)	3,5352*** (0.47)
<b>Escolaridade da mãe</b>	20,8204*** (0.21)	30,6928*** (0.36)	21,9390*** (0.25)	19,8370*** (0.22)	33,5605*** (0.49)
<b>Ocupação do pai</b>	-1,8681*** (0.20)	-0,2675 (0.34)	-2,0441*** (0.23)	-1,4752*** (0.21)	-4,0301*** (0.46)
<b>Ocupação da mãe</b>	5,5771*** (0.20)	9,9644*** (0.35)	8,8356*** (0.24)	6,7754*** (0.21)	8,1257*** (0.47)
<b>Residentes na casa</b>	-40,050*** (1.75)	-59,0126*** (3.04)	-38,5596*** (2.14)	-41,2696*** (1.82)	-67,0664*** (3.98)
<b>Renda domiciliar</b>	30,8842*** (0.21)	52,0228*** (0.36)	34,3138*** (2.14)	32,3465*** (0.21)	58,2841*** (0.48)

<b>Celular</b>	36,6450*** (0.64)	49,1080*** (1.07)	31,8305*** (0.75)	32,8513*** (0.65)	54,9649*** (1.45)
<b>Computador</b>	36,6431*** (0.20)	55,2804*** (0.34)	43,0172*** (0.23)	38,4396*** (0.21)	72,6827*** (0.46)
<b>Acesso à internet</b>	13,1478*** (0.45)	13,8109*** (0.75)	7,9205*** (0.53)	10,8428*** (0.46)	28,5329*** (1.04)
<b>R<sup>2</sup></b>	0,786	0,8399	0,804	0,801	0,8079
<b>R<sup>2</sup> ajustado</b>	0,7859	0,8398	0,804	0,8009	0,8078

Fonte: Elaboração própria a partir dos Microdados ENEM 2022.

Total de observações: 126.423

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

## 8.7. Regressões completas *Propensity Score Matching*

**Tabela 27 - Resultados Nordeste Modelo PSM 2019**

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>Escola</b>	42,493*** (1.28)	79,221*** (1.82)	62,618*** (1.45)	56,185*** (1.39)	121,559*** (2.42)
<b>Sexo</b>	5,225*** (1.30)	-34,049*** (1.85)	-12,022*** (1.47)	-3,605* (1.41)	34,520*** (2.46)
<b>Estado civil</b>	-36,001* (15.24)	-33,712 (21.73)	-26,352 (17.25)	-31,047 (16.62)	-81,915** (28.89)
<b>Cor/raça</b>	8,077*** (1.28)	17,249*** (1.83)	9,893*** (1.45)	9,810*** (1.40)	14,696*** (2.43)
<b>Ensino</b>	-25,339* (10.84)	-56,157*** (15.46)	-36,317** (12.27)	-31,424** (11.82)	-77,808*** (20.55)
<b>Escolaridade do pai</b>	5,926*** (1.60)	11,006*** (2.28)	8,247*** (1.81)	6,203*** (1.74)	9,451** (3.03)
<b>Escolaridade da mãe</b>	16,915*** (1.95)	25,324*** (2.78)	18,399*** (2.21)	19,065*** (2.13)	31,865*** (3.70)
<b>Ocupação do pai</b>	4,558** (1.47)	11,832*** (2.10)	5,999*** (1.67)	8,060*** (1.61)	9,300*** (2.79)
<b>Ocupação da mãe</b>	5,806*** (1.47)	13,829*** (2.10)	9,022*** (1.67)	7,438*** (1.60)	14,108*** (2.79)
<b>Residentes na casa</b>	-45,497** (15.24)	-24,991 (21.72)	-26,547 (17.24)	-35,335* (16.61)	-51,243 (28.88)
<b>Renda domiciliar</b>	25,442*** (1.38)	55,763*** (1.97)	35,029*** (1.56)	32,058*** (1.50)	58,633*** (2.62)
<b>Celular</b>	10,522* (4.95)	11,302 (7.05)	8,548 (5.60)	10,286 (5.39)	17,703 (9.37)
<b>Computador</b>	24,467*** (1.62)	41,288*** (2.30)	30,166*** (1.83)	29,584*** (1.76)	53,273*** (3.06)
<b>Acesso à internet</b>	10,011** (3.09)	8,913* (4.40)	10,932** (3.50)	10,381** (3.37)	20,655*** (5.86)
<b>R<sup>2</sup></b>	0,1021	0,1842	0,1479	0,1354	0,167
<b>R<sup>2</sup> ajustado</b>	214,3	424,4	326,6	294,8	377

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Total de observações: 26.260

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

Tabela 28 – Resultados do Sudeste Modelo PSM 2019

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>Escola</b>	41,0908*** (1.21)	86,034*** (1.73)	64,668*** (1.40)	53,524*** (1.29)	112,858*** (2.00)
<b>Sexo</b>	0,2892 (1.25)	-42,596*** (1.79)	-16,656*** (1.46)	-9,605*** (1.34)	24,985*** (2.08)
<b>Estado civil</b>	- 38,7913*** (10.65)	-70,258*** (15.25)	-53,539*** (12.39)	-42,756*** (11.36)	-59,674*** (17.67)
<b>Cor/raça</b>	8,6717*** (1.38)	18,602*** (1.98)	9,915*** (1.61)	10,365*** (1.48)	15,096*** (2.30)
<b>Ensino</b>	-17,1785 (13.86)	-85,802*** (19.85)	-62,557*** (16.13)	-28,309 (14.78)	- 102,758*** (23.00)
<b>Escolaridade do pai</b>	10,1922*** (1.75)	14,556*** (2.50)	10,136*** (2.03)	12,147*** (1.86)	15,577*** (2.90)
<b>Escolaridade da mãe</b>	22,0379*** (2.11)	38,270*** (3.02)	27,424*** (2.46)	24,992*** (2.25)	36,157*** (3.50)
<b>Ocupação do pai</b>	4,1965** (1.50)	11,382*** (2.15)	6,911*** (1.75)	7,894*** (1.60)	15,719*** (2.49)
<b>Ocupação da mãe</b>	5,3317*** (1.49)	15,636*** (2.13)	11,592*** (1.73)	6,621*** (1.58)	14,775*** (2.46)
<b>Residentes na casa</b>	-24,5664 (14.24)	-20,831 (20.39)	-20,282 (16.56)	-24,85 (15.18)	-9,402 (23.62)
<b>Renda domiciliar</b>	27,9374*** (1.48)	56,987*** (2.12)	38,736*** (1.72)	34,299*** (1.58)	61,425*** (2.46)
<b>Celular</b>	16,6586** (5.81)	18,958* (8.32)	14,071* (6.76)	16,701** (6.19)	15,776 (9.64)
<b>Computador</b>	26,5859*** (2.02)	44,614*** (2.90)	37,654*** (2.35)	33,120*** (2.16)	56,273*** (3.36)
<b>Acesso à internet</b>	14,2780** (14.28)	10,701 (6.44)	11,775* (5.23)	12,733** (4.79)	28,795*** (7.46)
<b>R<sup>2</sup></b>	0,6545	0,3471	0,1037	0,9128	0,1291
<b>R<sup>2</sup> ajustado</b>	205,6	454,6	338,9	294,4	434,2

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Total de observações: 40.892

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

Tabela 29 - Coeficientes Nordeste Modelo PSM 2022

Variáveis	Linguagem	Matemática	Ciências da natureza	Ciências humanas	Redação
<b>Escola</b>	60,4379*** (1.76)	94,743*** (2.32)	69,1477*** (1.88)	71,8889*** (1.82)	146,966*** (3.00)
<b>Sexo</b>	11,6761*** (1.81)	-24,054*** (2.39)	-8,7153*** (1.93)	1,4964 (1.87)	53,48*** (3.09)
<b>Estado civil</b>	-57,5215** (19.75)	-67,860** (26.07)	-46,1870* (21.05)	-56,1196** (20.44)	-74,019* (33.70)
<b>Cor/raça</b>	14,7681*** (1.77)	23,437*** (2.33)	14,3595*** (1.88)	14,7545*** (1.82)	19,850*** (3.01)
<b>Ensino</b>	-60,7991** (21.73)	-62,820* (28.67)	-56,6959* (23.15)	-58,8248** (22.48)	-93,635* (37.07)
<b>Escolaridade do pai</b>	6,6404** (2.31)	10,508*** (3.05)	4,8951* (2.47)	5,8478* (2.40)	4,370 (3.95)

<b>Escolaridade da mãe</b>	24,2876*** (2.99)	36,286*** (3.94)	22,5127*** (3.18)	22,7387*** (3.09)	47,385*** (5.10)
<b>Ocupação do pai</b>	-0,7835 (2.03)	5,026 (2.67)	-0,9291 (2.16)	-0,4712 (2.10)	-3,438 (3.46)
<b>Ocupação da mãe</b>	2,6339 (2.08)	7,871** (2.74)	4,0963 (2.19)	3,5573 (2.13)	4,339 (3.51)
<b>Residentes na casa</b>	-54,6522** (18.24)	-29,153 (24.07)	-12,7859 (19.43)	-56,3116** (18.87)	-56,913 (31.11)
<b>Renda domiciliar</b>	29,2940*** (2.08)	50,390*** (2.74)	30,8231*** (2.21)	30,3351*** (2.15)	59,961*** (3.545)
<b>Celular</b>	26,0809*** (6.47)	32,989*** (8.54)	29,6771*** (6.89)	26,3876*** (6.69)	39,836*** (11.04)
<b>Computador</b>	34,2096*** (2.30)	46,105*** (3.03)	34,9208*** (2.45)	34,7340*** (2.38)	58,173*** (3.92)
<b>Acesso à internet</b>	15,1233* (6.62)	17,686* (8.74)	17,5759* (7.05)	8,7872 (6.85)	16,399 (11.29)
<b>R<sup>2</sup></b>	0,09563	0,1372	0,1003	0,107	0,146
<b>R<sup>2</sup> ajustado</b>	0,09508	0,1367	0,09976	0,1065	0,1455

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

Total de observações: 23.368

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

**Tabela 30 - Coeficientes Sudeste Modelo PSM 2022**

<b>Variáveis</b>	<b>Linguagem</b>	<b>Matemática</b>	<b>Ciências da natureza</b>	<b>Ciências humanas</b>	<b>Redação</b>
<b>Escola</b>	57,767*** (1.49)	93,976*** (1.94)	72,438*** (1.59)	67,3979*** (1.53)	133,601*** (2.29)
<b>Sexo</b>	5,737*** (1.55)	-35,259*** (2.02)	-17,199*** (1.66)	-5,8399*** (1.59)	40,888*** (2.39)
<b>Estado civil</b>	-32,197* (15.13)	-16,664 (19.70)	-25,594 (16.15)	-32,2489* (15.53)	-17,69 (23.27)
<b>Cor/raça</b>	9,990*** (1.76)	23,942*** (2.29)	14,335*** (1.88)	10,2673*** (1.81)	17,949*** (2.71)
<b>Ensino</b>	-59,914*** (17.22)	-99,093*** (22.42)	-67,421*** (18.39)	-59,4248*** (17.68)	-98,781*** (26.49)
<b>Escolaridade do pai</b>	12,078*** (2.26)	17,778*** (2.95)	14,023*** (1.88)	11,0329*** (2.32)	13,802*** (3.48)
<b>Escolaridade da mãe</b>	21,314*** (2.91)	31,418*** (3.79)	22,811*** (3.10)	20,5875*** (2.98)	27,805*** (4.47)
<b>Ocupação do pai</b>	3,356 (1.82)	8,860*** (2.37)	4,984* (1.94)	4,4877* (1.87)	10,389*** (2.80)
<b>Ocupação da mãe</b>	5,749** (1.82)	12,414*** (2.38)	7,519*** (1.95)	6,8093*** (1.87)	10,919*** (2.81)
<b>Residentes na casa</b>	-42,247* (17.56)	-40,973 (22.87)	-31,936 (18.75)	-34,8283 (18.03)	-56,353* (27.02)
<b>Renda domiciliar</b>	28,838*** (2.23)	46,211*** (2.90)	27,316*** (2.38)	29,0968*** (2.28)	53,343*** (3.42)
<b>Celular</b>	37,903*** (6.23)	55,639*** (8.12)	42,331*** (6.66)	35,3086*** (6.40)	49,055*** (9.59)
<b>Computador</b>	32,565*** (2.64)	45,766*** (3.44)	36,122*** (2.82)	34,7762*** (2.71)	56,387*** (4.06)
<b>Acesso à internet</b>	-2,179 (9.55)	6,118 (12.44)	-3,238 (10.20)	-0,8739 (9.81)	25,97 (14.70)
<b>R<sup>2</sup></b>	0,05831	0,09659	0,07556	0,06873	0,1076
<b>R<sup>2</sup> ajustado</b>	0,058	0,09628	0,07525	0,06842	0,1073

Fonte: Elaboração própria a partir dos Microdados ENEM 2022.

Total de Observações: 41.558

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

8.8. Regressões *Logit* e função logísticaTabela 31 - Modelo *Logit* para o ano de 2019

Variáveis	Nordeste	Sudeste	Brasil
<b>Sexo</b>	0,23837*** (0.02)	0,03064 (0.02)	0,048595*** (0.01)
<b>Estado civil</b>	-1,31713*** (0.21)	-0,56079*** (0.13)	-0,917135*** (0.10)
<b>Cor/Raça</b>	0,57690*** (0.02)	0,73943*** (0.02)	0,43651*** (0.01)
<b>Ensino</b>	-0,10159 (0.20)	-0,23605 (0.20)	-0,019518 (0.12)
<b>Escolaridade do pai</b>	0,42267*** (0.01)	0,51372*** (0.02)	0,482434*** (0.01)
<b>Escolaridade da mãe</b>	0,73541*** (0.03)	0,69213*** (0.03)	0,734424*** (0.02)
<b>Ocupação do pai</b>	0,68970*** (0.03)	0,94389*** (0.02)	0,882267*** (0.01)
<b>Ocupação da mãe</b>	0,55440*** (0.03)	0,75131*** (0.02)	0,730222*** (0.01)
<b>Residentes na casa</b>	-0,21292 (0.24)	-0,02294 (0.20)	-0,007275 (0.11)
<b>Renda domiciliar</b>	1,55902*** (0.03)	1,29817*** (0.02)	1,182327*** (0.01)
<b>Celular</b>	0,12977 (0.08)	-0,06406 (0.08)	-0,042631 (0.05)
<b>Computador</b>	0,89132*** (0.03)	0,61981*** (0.03)	0,601075*** (0.02)
<b>Acesso à internet</b>	0,83402*** (0.05)	0,55897*** (0.06)	0,660684*** (0.03)
<b>Total de observações</b>	112.713	137.558	352.836

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

P-valor: \* p &lt; 0,05, \*\* p &lt; 0,01, \*\*\* p &lt; 0,001.

(Erro padrão)

Tabela 32 - Modelo *Logit* para o ano de 2022

Variáveis	Nordeste	Sudeste	Brasil
<b>Sexo</b>	0,27803*** (0.02)	-0,04393* (0.02)	0,01871 (0.01)
<b>Estado civil</b>	-0,37322 (0.21)	0,08197 (0.18)	-0,28378* (0.11)
<b>Cor/Raça</b>	0,59354*** (0.02)	0,77457*** (0.02)	0,57524*** (0.01)
<b>Ensino</b>	-1,14740*** (0.25)	-0,94663*** (0.19)	-0,92583*** (0.12)
<b>Escolaridade do pai</b>	0,43559*** (0.03)	0,55162*** (0.03)	0,54746*** (0.02)
<b>Escolaridade da mãe</b>	0,79057*** (0.03)	0,70147*** (0.03)	0,75633** (0.02)
<b>Ocupação do pai</b>	0,64997*** (0.03)	0,94989*** (0.02)	0,89319*** (0.01)

<b>Ocupação da mãe</b>	0,50833*** (0.03)	0,69494*** (0.02)	0,67431*** (0.01)
<b>Residentes na casa</b>	-0,03151 (0.22)	0.16309 (0.20)	0,05709 (0.11)
<b>Renda domiciliar</b>	1,73509*** (0.02)	1,42299*** (0.02)	1,40303*** (0.01)
<b>Celular</b>	-0,21831** (0.08)	-0,17675* (0.07)	-0,09643* (0.04)
<b>Computador</b>	1,20970*** (0.03)	1,02031*** (0.03)	1,0046*** (0.02)
<b>Acesso à internet</b>	0,67777*** (0.07)	0,38069*** (0.10)	0,53941*** (0.05)
<b>Total de observações</b>	154.306	126.423	407.444

Fonte: Elaboração própria a partir dos Microdados Enem 2022

P-valor: \* p < 0,05, \*\* p < 0,01, \*\*\* p < 0,001.

(Erro padrão)

**Tabela 33 - Regressão logística para o ano de 2019**

<b>Variáveis</b>	<b>Nordeste</b>	<b>Sudeste</b>	<b>Brasil</b>
<b>Sexo</b>	0,5593	0,5077	0,5121
<b>Estado civil</b>	0,2113	0,3634	0,2855
<b>Cor/Raça</b>	0,6404	0,6769	0,6074
<b>Ensino</b>	0,4746	0,4413	0,4951
<b>Escolaridade do pai</b>	0,6041	0,6257	0,6183
<b>Escolaridade da mãe</b>	0,6760	0,6664	0,6758
<b>Ocupação do pai</b>	0,6659	0,7199	0,7073
<b>Ocupação da mãe</b>	0,6352	0,6795	0,6748
<b>Residentes na casa</b>	0,4470	0,4943	0,4982
<b>Renda domiciliar</b>	0,8262	0,7855	0,7654
<b>Celular</b>	0,5324	0,4840	0,4893
<b>Computador</b>	0,7092	0,6502	0,6459
<b>Acesso à internet</b>	0,6972	0,6362	0,6594

Fonte: Elaboração própria a partir dos Microdados ENEM 2019.

**Tabela 34 - Regressão logística para o ano de 2022**

<b>Variáveis</b>	<b>Nordeste</b>	<b>Sudeste</b>	<b>Brasil</b>
<b>Sexo</b>	0,5691	0,4890	0,5048
<b>Estado civil</b>	0,4078	0,5205	0,4295
<b>Cor/Raça</b>	0,6442	0,6845	0,6400
<b>Ensino</b>	0,2410	0,2796	0,2838
<b>Escolaridade do pai</b>	0,6072	0,6345	0,6335
<b>Escolaridade da mãe</b>	0,6880	0,6685	0,6806
<b>Ocupação do pai</b>	0,6570	0,7211	0,7095
<b>Ocupação da mãe</b>	0,6244	0,6671	0,6625
<b>Residentes na casa</b>	0,4921	0,5407	0,5143

<b>Renda domiciliar</b>	0,8501	0,8058	0,8027
<b>Celular</b>	0,4456	0,4559	0,4759
<b>Computador</b>	0,7702	0,7350	0,7320
<b>Acesso a internet</b>	0,6632	0,5940	0,6318

---

Fonte: Elaboração própria a partir dos Microdados ENEM 2022.