

idp

idn

MESTRADO PROFISSIONAL EM ECONOMIA

**DESEMPENHO NO ENEM NO DISTRITO FEDERAL (2017–
2023):** UMA ANÁLISE COM GAMLSS-SHASHO2 E TESTES NÃO
PARAMÉTRICOS

SÂMEA LACERDA LIMA

Brasília-DF, 2025

SÂMEA LACERDA LIMA

**DESEMPENHO NO ENEM NO DISTRITO FEDERAL (2017–
2023):
UMA ANÁLISE COM GAMLSS-SHASHO2 E TESTES NÃO
PARAMÉTRICOS**

Dissertação apresentada ao Programa de Pós Graduação em Economia, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como requisito parcial para obtenção do grau de Mestre.

Orientador

Professor Doutor Roberta Moreira Wichmann

Brasília-DF 2025

SÂMEA LACERDA LIMA

DESEMPENHO NO ENEM NO DISTRITO FEDERAL (2017–2023):

UMA ANÁLISE COM GAMLSS-SHASHO2 E TESTES NÃO PARAMÉTRICOS

Dissertação apresentada ao Programa de Pós Graduação em Economia, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como requisito parcial para obtenção do grau de Mestre.

Aprovado em 30 / 06 / 2025

Banca Examinadora

Prof. Dr. Roberta Moreira Wichmann - Orientador

Prof. Dr. Carlos Eduardo Gasparini

Prof. Dr. Tiago Almeida de Oliveira

Código de catalogação na publicação – CIP

L732d Lima, Sâmea Lacerda
Desempenho no ENEM no Distrito Federal (2017-2023) uma
análise com GAMLSS-SHASH02 e testes não paramétricos /
Sâmea Lacerda Lima. — Brasília: Instituto Brasileiro Ensino,
Desenvolvimento e Pesquisa, 2025.
78 f. :

Orientador: Profa. Dra. Roberta Moreira Wichmann

Dissertação (Mestrado Profissional em Economia) —
Instituto Brasileiro Ensino, Desenvolvimento e Pesquisa – IDP,
2025.

1. Rendimento escolar. 2. Estatística educacional. 3.
Igualdade na educação. 4. Sistema de educação - Brasil. I.Título

CDD 330

Elaborada pela Biblioteca Ministro Moreira Alves

RESUMO

Este estudo tem como objetivo investigar os efeitos das variáveis socioeducacionais no desempenho dos estudantes do Distrito Federal no Exame Nacional do Ensino Médio (ENEM), com base na análise dos microdados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) no período de 2017 a 2023. As variáveis selecionadas para análise incluem a escolaridade dos pais, sexo, raça, a renda familiar e o tipo de escola, como fatores que influenciam o desempenho acadêmico dos estudantes.

Diferentemente de abordagens tradicionais que concentram a análise apenas na média dos resultados, este trabalho adota uma metodologia inovadora por meio da aplicação do modelo GAMLSS (Generalized Additive Models for Location, Scale and Shape) com a distribuição SHASHo2 (Sinh-Arcsinh), que possibilita modelar simultaneamente os parâmetros de localização (média), escala (dispersão), assimetria e curtose da distribuição das notas do ENEM. Tal abordagem oferece uma modelagem mais flexível e robusta, capaz de capturar características complexas e heterogêneas dos dados educacionais que modelos convencionais não conseguem representar adequadamente.

Além disso, foram conduzidos testes estatísticos não paramétricos complementares como o teste de Mann-Whitney, para comparar distribuições entre dois grupos; o teste de Kruskal-Wallis, para analisar diferenças entre múltiplos grupos; e o teste de Dunn, para realizar comparações múltiplas post hoc. Esses testes foram fundamentais para validar a significância estatística das diferenças observadas nos grupos de estudantes, garantindo a consistência e a confiabilidade dos resultados.

Os resultados indicam que as variáveis socioeconômicas e escolares não influenciam apenas o valor médio das notas, mas também aspectos mais sutis da distribuição do desempenho, como a variabilidade, a assimetria e o peso das caudas da distribuição. Observou-se que estudantes oriundos de famílias com maior renda e escolaridade paterna apresentam desempenho significativamente mais elevado e menos disperso, refletindo maior estabilidade nos resultados acadêmicos. Por outro lado, o tipo de escola demonstrou

influenciar de forma multifacetada os diferentes parâmetros da distribuição das notas, evidenciando desigualdades estruturais presentes no sistema educacional do Distrito Federal.

Este trabalho contribui para o aprofundamento do conhecimento acerca dos determinantes do desempenho escolar, ao introduzir uma abordagem estatística avançada e ao fornecer evidências empíricas que subsidiem políticas públicas voltadas à redução das desigualdades educacionais. A modelagem detalhada dos diferentes parâmetros da distribuição das notas possibilita uma compreensão mais abrangente dos fatores que afetam o desempenho dos estudantes, indicando caminhos para intervenções mais precisas e eficazes no contexto educacional.

Palavras chave: Desempenho escolar, ENEM, variáveis socioeducacionais, GAMLSS, testes não paramétricos, desigualdade educacional, Distrito Federal.

ABSTRACT

This study aims to investigate the effects of socio-educational variables on the performance of students in the Federal District in the National High School Exam (ENEM), based on the analysis of microdata made available by the National Institute of Studies and Educational Research Anísio Teixeira (INEP) from 2017 to 2023. The variables selected for analysis include parents' education, gender, race, family income and type of school, as factors that influence students' academic performance.

Unlike traditional approaches that focus the analysis only on the average of the results, this work adopts an innovative methodology through the application of the GAMLSS (Generalized Additive Models for Location, Scale and Shape) model with the SHASHo2 (Sinh-Arcsinh) distribution, which makes it possible to simultaneously model the parameters of location (average), scale (dispersion), asymmetry and kurtosis of the distribution of ENEM scores. This approach offers a more flexible and robust modeling, capable of capturing complex and heterogeneous characteristics of educational data that conventional models cannot adequately represent.

In addition, complementary nonparametric statistical tests were conducted, such as the Mann-Whitney test to compare distributions between two groups; the Kruskal-Wallis test to analyze differences between multiple groups; and the Dunn test to perform multiple post hoc comparisons. These tests were essential to validate the statistical significance of the differences observed in the student groups, ensuring the consistency and reliability of the results.

The results indicate that socioeconomic and educational variables influence not only the average value of grades, but also more subtle aspects of the distribution of performance, such as variability, asymmetry and the weight of the tails of the distribution. It was observed that students from families with higher income and paternal education had significantly higher and less dispersed performance, reflecting greater stability in academic results. On the other hand, the type of school was shown to influence in a multifaceted way the different parameters of the distribution of grades, evidencing structural inequalities present in the educational system of the Federal District.

This work contributes to the deepening of knowledge about the determinants of school performance, by introducing an advanced statistical approach and by providing empirical evidence to support public policies aimed at reducing educational inequalities. The detailed modeling of the different parameters of the distribution of grades allows a more comprehensive understanding of the factors that affect student performance, indicating paths for more precise and effective interventions in the educational context.

Keywords: School performance, ENEM, socio-educational variables, GAMLSS, non-parametric tests, educational inequality, Federal District.

LISTA DE ABREVIATURAS E SIGLAS

ENEM	Exame Nacional do Ensino Médio
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IPEA	Instituto de Pesquisas Aplicadas
SISU	Sistema de Seleção Unificada
PROUNI	Programa de Universidade para Todos
FIES	Fundo de Financiamento Estudantil
GAMLSS	GENERALIZED ADDITIVE MODELS FOR LOCATION, SCALE AND SHAPE

LISTA DE ILUSTRAÇÕES

Figura 1 parâmetro de dispersão (σ)	65
Figura 2 parâmetro da assimetria (v)	66
Figura 3 parâmetro de curtose (τ)	66
Figura 4 Diagnóstico dos resíduos quantílicos do modelo GAMLSS com distribuição SHASHo2	67
Figura 5 Efeitos marginais das variáveis explicativas sobre a assimetria (v) da distribuição das notas	69
Histograma 1 Histograma da distribuição da variável nota_media	39
Histograma 2 Distribuição da nota média por gênero	41
Histograma 3 Distribuição da nota média por raça	43
Histograma 4 Distribuição da nota média por tipo de escola	46
Histograma 5 Distribuição da nota média segundo a escolaridade do pai	49
Histograma 6 Distribuição da nota média segundo a escolaridade da mãe	52
Histograma 7 Distribuição da nota média por faixa de renda familiar	55
Histograma 8 Distribuição da nota média ao longo dos anos (2017-2023)	56

LISTA DE TABELAS

Tabela 1	Variáveis explicativas consideradas na análise	31
Tabela 2	Distribuição percentual das variáveis categóricas da amostra	36
Tabela 3	Estatísticas descritivas para a variável nota média (N = 107.824)	38
Tabela 4	Análise descritiva das notas médias por gênero	40
Tabela 5	Resultado do teste de Mann-Whitney para comparação das notas médias por gênero	41
Tabela 6	Tamanho de efeito (r) e intervalo de confiança para o teste de Mann-Whitney	42
Tabela 7	Estatísticas descritivas das notas médias por raça	42
Tabela 8	Resultado do teste de Kruskal-Wallis para notas médias por raça/cor	43
Tabela 9	Comparações múltiplas (teste de Dunn) entre grupos de raça/cor	44
Tabela 10	Estatísticas descritivas das notas médias por tipo de escola	45
Tabela 11	Resultado do teste de Kruskal-Wallis para notas médias por tipo de escola	47
Tabela 12	Comparações múltiplas (teste de Dunn) entre tipos de escola	47
Tabela 13	Estatísticas descritivas das notas médias por escolaridade do pai	48
Tabela 14	Resultado do teste de Kruskal-Wallis para notas por escolaridade do pai	49
Tabela 15	Comparações múltiplas (teste de Dunn) por escolaridade do pai	50

Tabela 16	Estatísticas descritivas das notas médias por escolaridade da mãe	51
Tabela 17	Resultado do teste de Kruskal-Wallis para notas médias por escolaridade	51
Tabela 18	Comparações múltiplas (teste de Dunn) entre grupos de escolaridade da mãe	52
Tabela 19	Estatísticas descritivas das notas médias por renda familiar	53
Tabela 20	Resultado do teste de Kruskal-Wallis para notas médias por renda familiar	54
Tabela 21	Comparações múltiplas (teste de Dunn) entre grupos de renda familiar	54
Tabela 22	Valores do Fator de Inflação da Variância Generalizado (GVIF ajustado)	57
Tabela 23	Estimativas dos coeficientes para o parâmetro μ (média) do modelo GAMLSS com distribuição SHASHo2	59
Tabela 24	Estimativas dos coeficientes para o parâmetro σ (sigma) do modelo GAMLSS com distribuição SHASHo2	61
Tabela 25	Estimativas dos coeficientes para o parâmetro ν (nu) do modelo GAMLSS com distribuição SHASHo2	62
Tabela 26	Estimativas dos coeficientes para o parâmetro τ (tau) do modelo GAMLSS com distribuição SHASHo2	63
Tabela 27	Resumo estatístico dos resíduos quantílicos do modelo GAMLSS	67

SUMÁRIO

1. INTRODUÇÃO	15
----------------------------	-----------

2. REFERENCIAL TEÓRICO	20
-------------------------------------	-----------

3. METODOLOGIA	25
-----------------------------	-----------

3.1 MODELO GAMLSS	25
-------------------------	-----------

3.1.1 MODELAGEM COM GAMLSS	27
----------------------------------	-----------

3.1.2 ESCOLHA DA DISTRIBUIÇÃO SHASHO2	28
---	-----------

3.2 BASE DE DADOS	30
-------------------------	-----------

3.3 ESTRATÉGIA DE ANÁLISE ESTATÍSTICA	31
---	-----------

3.3.1 TESTES ESTATÍSTICOS NÃO PARAMÉTRICOS	31
--	-----------

3.3.2 CRITÉRIOS DE COMPARAÇÃO E AJUSTE	33
--	-----------

4. RESULTADOS	36
----------------------------	-----------

4.1 ANÁLISE DESCRITIVA DA BASE DE DADOS	36
---	-----------

4.2 ANÁLISE DAS VARIÁVEIS E TESTES NÃO PARAMÉTRICOS	39
---	-----------

4.3 MODELAGEM ESTATÍSTICA COM GAMLSS	57
--	-----------

4.3.1 DIAGNÓSTICO DE MULTICOLINEARIDADE	57
---	-----------

4.3.2 AJUSTE DO MODELO GAMLSS	58
-------------------------------------	-----------

4.3.3 EFEITO DAS VARIÁVEIS NA MÉDIA DAS NOTAS (M)	64
---	-----------

4.3.4 VARIABILIDADE, ASSIMETRIA E CURTOSE DAS NOTAS	64
---	-----------

5. CONCLUSÃO	71
---------------------------	-----------

REFERÊNCIAS	74
--------------------------	-----------



1

INTRODUÇÃO

O Exame Nacional do Ensino Médio (ENEM) se consolidou como um dos principais mecanismos de acesso ao ensino superior no país, influenciando diretamente a trajetória acadêmica e profissional de milhões de estudantes, além de ser um dos principais instrumentos de avaliação educacional do Brasil (IPEA, 2015).

No contexto do Distrito Federal, essa avaliação assume uma importância também significativa, dado o cenário educacional peculiar da região, que abriga uma diversidade socioeconômica e cultural, devido à grande concentração do funcionalismo público. Assim, o ENEM não apenas serve como um critério de seleção para instituições de ensino superior, mas também é uma ferramenta de mobilidade social, oferecendo oportunidades para estudantes de diferentes origens. Diversos estudos realizados pelo Instituto de Pesquisa Econômica Aplicada - IPEA analisam a contribuição do ENEM e do Sistema de Seleção Unificada - SISU para a mobilidade social no Brasil. Um exemplo é o estudo de 2015, "Mobilidade Social e Acesso ao Ensino Superior no Brasil", que aborda o impacto de políticas como o ENEM, o Programa de Universidade para Todos - PROUNI e o Fundo de Financiamento Estudantil - FIES na inclusão de estudantes de classes sociais mais baixas no ensino superior.

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) desempenha um papel fundamental na estrutura educacional brasileira, atuando como a principal entidade responsável pela avaliação e pelo monitoramento da educação no país¹. As avaliações realizadas pelo INEP têm um impacto significativo na educação nacional, fornecendo uma base empírica para o diagnóstico da qualidade do ensino, permitindo que educadores e gestores identifiquem áreas de sucesso e aspectos que necessitam de intervenção.

¹ Essa atribuição se justifica pelo fato de o INEP ser uma autarquia federal vinculada ao Ministério da Educação, legalmente incumbida de coordenar os processos avaliativos e estatísticos da educação nacional, conforme estabelecem a Lei de Diretrizes e Bases da Educação Nacional (Lei nº 9.394/1996) e a Lei nº 10.861/2004, que instituiu o Sistema Nacional de Avaliação da Educação Superior (SINAES) (BRASIL, 1996; BRASIL, 2004).

Os resultados dessas avaliações são utilizados para promover a transparência e a responsabilização nas escolas, estimulando a melhoria contínua da qualidade educacional². A disseminação de dados estatísticos e análises detalhadas sobre o desempenho estudantil e as condições das escolas é crucial para a formulação de políticas públicas que buscam equidade e inclusão no acesso à educação de qualidade.

O Distrito Federal - DF tem uma das menores taxas de analfabetismo do Brasil, o que indica um avanço significativo no acesso à educação básica. De acordo com o IBGE (2020), a taxa de analfabetismo no DF era de 5%, abaixo da média nacional, que é de 6,6%. Embora o analfabetismo tenha diminuído substancialmente, ainda existem áreas mais afastadas da capital, especialmente no Entorno, que enfrentam maiores desafios educacionais (CENSO 2020). No Índice de Desenvolvimento da Educação Básica (IDEB) o DF se destaca positivamente, principalmente no ensino fundamental. No IDEB 2019, o DF obteve uma média de 6,2 no ensino fundamental, superando a média nacional de 5,9. Isso reflete os avanços em qualidade e infraestrutura nas escolas. No entanto, no ensino médio, o desempenho do DF foi de 4,2, ainda superior à média nacional de 3,9, mas demonstrando disparidade de qualidade entre diferentes regiões da capital e que os desafios estruturais nesse nível de ensino permanecem (INEP, 2019).

Compreender o desempenho no ENEM vai além da análise de médias. Trata-se de identificar os fatores sociais e institucionais que moldam os resultados dos estudantes, revelando desigualdades e permitindo o direcionamento de políticas públicas mais eficazes. Entre esses fatores, destacam-se a renda familiar, a escolaridade dos pais e o tipo de escola frequentada. Estudantes com maior poder aquisitivo tendem a ter acesso a recursos pedagógicos, acompanhamento educacional e ambiente doméstico mais favorável ao aprendizado (Bourdieu, Passeron, 1975). De modo similar, o nível educacional dos pais — sobretudo da mãe — representa uma importante dimensão do capital cultural que influencia o desempenho escolar.

² BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Transparência e Prestação de Contas. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/transparencia-e-prestacao-de-contas>. Acesso em: 7 jul. 2025.

Assim, o objetivo deste estudo é analisar como variáveis socioeconômicas e escolares influenciam o desempenho dos estudantes do Distrito Federal no ENEM, no período de 2017 a 2023. Busca-se identificar padrões de desempenho entre os grupos sociais, verificar tendências ao longo do tempo e mapear desigualdades educacionais associadas à renda, escolaridade dos pais e tipo de escola (pública ou privada). A investigação fundamenta-se na hipótese, amplamente sustentada pela literatura³, de que essas variáveis exercem influência significativa nos resultados obtidos pelos alunos. Para isso, recorre-se a uma abordagem estatística inovadora, que combina testes não paramétricos e modelagem por meio dos Modelos Aditivos para Local, Dispersão e Forma (GAMLSS), com distribuição SHASHo2. Essa escolha metodológica permite analisar múltiplas dimensões da distribuição de desempenho — como variabilidade, assimetria e curtose — respeitando as características dos dados do ENEM e ampliando a compreensão das desigualdades educacionais.

A pergunta que norteia esta pesquisa é: Como as variáveis socioeconômicas — como renda familiar, escolaridade dos pais e tipo de escola — influenciam o desempenho dos estudantes do Distrito Federal no ENEM, entre os anos de 2017 a 2023?

Ao examinar as desigualdades de desempenho com base em dados longitudinais e por meio de ferramentas estatísticas robustas, esta pesquisa pretende contribuir para o diagnóstico da educação básica na capital federal e fornecer subsídios empíricos para o desenvolvimento de estratégias que promovam maior equidade no acesso ao ensino superior.

Pois compreender as barreiras enfrentadas por estudantes em contextos diversos é essencial para a construção de políticas públicas que promovam uma educação mais justa, inclusiva e transformadora.

Dessa forma, ao lançar mão de métodos estatísticos mais sensíveis às especificidades dos dados educacionais, esta pesquisa busca oferecer uma leitura aprofundada das desigualdades de desempenho no ENEM entre os estudantes do Distrito Federal. Espera-se que os achados contribuam tanto para o debate acadêmico quanto

³ BOURDIEU, Pierre; PASSERON, Jean-Claude. *A reprodução: elementos para uma teoria do sistema de ensino*. Rio de Janeiro: Francisco Alves, 1975.



para a formulação de políticas públicas voltadas à promoção de uma educação mais equitativa.





?

2

REFERENCIAL TEÓRICO

A relação entre o desempenho no Exame Nacional do Ensino Médio (ENEM) e as variáveis socioeconômicas têm sido objeto de diversos estudos, que buscam entender como fatores sociais e econômicos influenciam a *performance* dos estudantes. A literatura aponta para a complexidade dessa relação, envolvendo múltiplas dimensões que afetam o acesso e a qualidade da educação no Brasil.

O desempenho dos estudantes no Exame Nacional do Ensino Médio (ENEM) é um fenômeno multifacetado que pode ser profundamente influenciado por uma série de variáveis socioeconômicas. Para compreender essa complexidade, diversas teorias podem ser mobilizadas, cada uma oferecendo uma perspectiva única sobre como fatores sociais e econômicos moldam as oportunidades educacionais e, conseqüentemente, os resultados acadêmicos.

De acordo com Oliveira (2016) o ENEM é fundamental no processo de reestruturação educacional, pois, ao avaliar competências e não apenas conhecimentos teóricos, fornece informações essenciais para aprimorar a prática docente, orientar os alunos em seu desempenho e apoiar a formulação de políticas públicas para a educação.

Um dos principais fatores socioeconômicos que afetam o desempenho no ENEM é a renda familiar. Isso se deve ao acesso a recursos educacionais, como livros, internet e cursos preparatórios, que são mais frequentes em famílias com melhores condições financeiras. A pesquisa de Silva et al. (2017) confirma essa correlação, apontando que a desigualdade de renda está diretamente ligada às disparidades no desempenho acadêmico.

A Teoria do Capital Cultural de Pierre Bourdieu (1960) sugere que o conhecimento e as habilidades adquiridos no ambiente familiar influenciam diretamente o desempenho escolar. Bourdieu argumenta que as diferenças no capital cultural entre as classes sociais contribuem para a reprodução das desigualdades no sistema educacional,



evidenciando como a escolaridade dos pais e o ambiente familiar impactam as expectativas e o engajamento dos alunos.

Complementando essa visão, a Teoria da Mobilidade Social, proposta por Erikson e Goldthorpe (1992) analisa como as condições socioeconômicas de uma geração podem restringir ou facilitar as oportunidades educacionais da próxima. Essa teoria é essencial para entender as barreiras que estudantes de contextos menos favorecidos enfrentam ao longo de sua trajetória escolar.

Alves e Ribeiro (2017) indicam que alunos cujos pais possuem níveis mais altos de escolaridade tendem a ter um desempenho superior no ENEM. Essa relação pode ser explicada pela transmissão de capital cultural e social, que prepara os estudantes para enfrentar os desafios educacionais.

Diversos estudos sobre o desempenho acadêmico nas escolas brasileiras mostram que alunos de escolas públicas de ensino fundamental e médio têm desempenho médio significativamente inferior ao de alunos de escolas privadas ou federais (Albanez, Ferreira e Franco, 2002; Alves, 2007; Barbosa, 2009). Essas diferenças refletem a grande estratificação do sistema educacional brasileiro, que, embora siga um currículo unificado, reproduz desigualdades. Segundo a teoria da "desigualdade efetivamente mantida" (Lucas, 2001), a estratificação educacional é um dos principais meios de reprodução das desigualdades sociais e educacionais. No Brasil, essa situação é agravada pelo fato de as melhores universidades serem públicas e gratuitas, enquanto as escolas privadas de ensino fundamental e médio apresentam melhor qualidade. Esse sistema, em seu desenho, favorece a desigualdade. Assim, é essencial estudar as desigualdades de oportunidades entre estudantes de escolas públicas e privadas.

Além das variáveis socioeconômicas tradicionais, a literatura também destaca a interseccionalidade das desigualdades de gênero e raça. De acordo com o relatório da organização Todos pela Educação (2021) a combinação de desigualdades raciais e de classe social contribui para um cenário em que o acesso e a qualidade da educação são desiguais⁴.

⁴ TODOS PELA EDUCAÇÃO. *A educação no Brasil: uma perspectiva internacional*. São Paulo: Todos pela Educação, 2021. Disponível em: https://todospelaeducacao.org.br/wordpress/wp-content/uploads/2021/06/A-Educacao-no-Brasil_uma-perspectiva-internacional.pdf. Acesso em: 7 jan. 2025.

As análises sobre desempenho no ENEM e variáveis socioeconômicas não são apenas descritivas, mas também oferecem implicações importantes para a formulação de políticas públicas. Entender as relações entre esses fatores pode ajudar na criação de intervenções que visem a equidade educacional, promovendo recursos e suporte direcionados para estudantes em situações de vulnerabilidade.

Esses referenciais teóricos oferecem uma base sólida para a análise do desempenho no ENEM em relação às variáveis socioeconômicas. Ao integrar esses conceitos, a pesquisa pode proporcionar uma compreensão mais profunda das dinâmicas que influenciam os resultados dos alunos e contribuir para o desenvolvimento de políticas educacionais mais eficazes e equitativas.

O impacto da pandemia de COVID-19, em particular, pode ter alterado significativamente o desempenho dos alunos em 2020, devido a fatores como a transição para o ensino remoto, a inacessibilidade à educação de qualidade durante o período de distanciamento social e as dificuldades socioeconômicas exacerbadas por essa crise.

No campo metodológico, diversos trabalhos sobre o desempenho escolar utilizam abordagens estatísticas convencionais, como análises descritivas, testes de hipótese paramétricos e modelos lineares (LM) ou modelos lineares generalizados (GLM). Embora úteis, essas abordagens assumem pressupostos restritivos, como normalidade dos resíduos e homocedasticidade, que nem sempre são atendidos por dados educacionais de larga escala. Além disso, tais modelos concentram-se apenas na média da variável resposta, desconsiderando outros aspectos relevantes da distribuição, como a variabilidade, a assimetria e a curtose, que também refletem desigualdades estruturais.

Diante dessas limitações, esta pesquisa propõe uma abordagem estatística mais flexível e robusta. Inicialmente, utilizam-se testes não paramétricos, como Kruskal-Wallis e Mann-Whitney, apropriados para a comparação entre grupos socioeconômicos quando os pressupostos paramétricos são violados. Em seguida, adota-se a modelagem por meio dos Modelos Aditivos para Local, Dispersão e Forma (GAMLSS), que permitem incorporar múltiplas dimensões da distribuição da variável resposta. Esse tipo de modelo se mostra especialmente

adequado para tratar dados assimétricos e com caudas pesadas, características frequentemente observadas nas notas do ENEM.

Dentro do escopo de distribuições oferecidas pelo GAMLSS, a distribuição SHASHo2 (Sinh-Arcsinh, parametrização 2), proposta por Jones e Pewsey (2009), foi selecionada por apresentar o melhor desempenho segundo os critérios de informação (AIC e BIC) e qualidade dos resíduos. Essa distribuição estende a normal clássica, permitindo ajustes flexíveis para assimetria e curtose, sendo, portanto, particularmente indicada para o tipo de dado analisado neste estudo.

A adoção dessa metodologia, ainda pouco explorada nos estudos sobre desempenho educacional no Brasil, representa uma inovação na forma de investigar as desigualdades educacionais. Ao permitir modelar diferentes aspectos da distribuição das notas — e não apenas suas médias — o GAMLSS com distribuição SHASHo2 oferece uma visão mais abrangente e sensível às nuances do fenômeno, respeitando a complexidade empírica dos dados e fornecendo subsídios mais robustos para a formulação de políticas públicas educacionais.



3

3

METODOLOGIA

Devido à natureza dos dados e à violação de pressupostos como normalidade e homocedasticidade, optou-se por uma abordagem metodológica mais flexível e robusta, combinando testes não paramétricos e modelagem GAMLSS. Na etapa inicial, foram aplicados testes estatísticos não paramétricos — como o teste de Kruskal-Wallis para comparação entre três ou mais grupos e o teste de Dunn para comparações múltiplas post hoc — a fim de identificar diferenças significativas entre os grupos definidos por variáveis socioeconômicas e escolares. Esses testes foram escolhidos por sua adequação a dados que não seguem distribuição normal e que apresentam variância desigual entre os grupos.

Em uma segunda etapa, foi utilizada a modelagem estatística por meio do modelo GAMLSS (Generalized Additive Models for Location, Scale and Shape), ajustado com a distribuição SHASHo2 (Sinh-Arcsinh). Essa escolha metodológica representa um avanço em relação aos modelos tradicionais, pois permite modelar simultaneamente não apenas a média (μ), mas também a dispersão (σ), a assimetria (ν) e a curtose (τ) da variável resposta — no caso, a nota média dos estudantes no ENEM. A distribuição SHASHo2, por sua vez, é especialmente apropriada para dados com caudas pesadas e assimetria, características frequentemente observadas em avaliações em larga escala.

Essa abordagem integrada possibilitou uma análise mais abrangente e sensível das desigualdades educacionais, capturando nuances que vão além da média dos desempenhos e revelando padrões de variação associados ao perfil socioeconômico dos estudantes. A seguir, são apresentados os dados utilizados, os critérios de seleção e limpeza da base, bem como os procedimentos estatísticos detalhados empregados na análise.

3.1 MODELO GAMLSS

A escolha do modelo estatístico nesta pesquisa foi orientada por três fatores principais: (i) a natureza contínua e assimétrica da variável resposta (nota média no ENEM), (ii) a presença de heterocedasticidade

e caudas pesadas nos dados e (iii) a necessidade de modelar múltiplos aspectos da distribuição condicional da variável resposta, não se restringindo à média.

Inicialmente, considerou-se a utilização do Modelo Linear (LM), baseado na suposição de normalidade dos resíduos e homocedasticidade. No entanto, a análise exploratória revelou violações sistemáticas desses pressupostos, especialmente assimetria negativa e aumento da variabilidade das notas ao longo de subgrupos socioeconômicos. A presença de outliers e a estrutura heterogênea dos dados — oriundos de diferentes estratos sociais e escolares — comprometem a validade das inferências sob o paradigma clássico de mínimos quadrados.

Como alternativa, testou-se o uso do Modelo Linear Generalizado (GLM), cuja flexibilidade permite especificar distribuições não normais para a variável resposta, bem como utilizar funções de ligação adequadas à escala dos dados. Todavia, os GLMs permanecem restritos à modelagem do parâmetro de localização (média), assumindo que os demais momentos da distribuição (dispersão, assimetria, curtose) permanecem constantes e exógenos ao modelo. Essa limitação é particularmente problemática em contextos educacionais, nos quais a dispersão do desempenho pode variar substancialmente entre grupos (por exemplo, estudantes de escolas públicas e privadas), e a distribuição das notas pode apresentar assimetrias e curtose elevadas, associadas à desigualdade de oportunidades.

Diante dessas limitações, optou-se pela adoção do modelo GAMLSS (Generalized Additive Models for Location, Scale and Shape), desenvolvido por Rigby e Stasinopoulos (2005). O GAMLSS estende a estrutura dos GLMs ao permitir que múltiplos parâmetros da distribuição da variável resposta — incluindo não apenas a média (μ), mas também a dispersão (σ), a assimetria (ν) e a curtose (τ) — sejam modelados como funções de covariáveis. Essa abordagem oferece maior flexibilidade para capturar variações estruturais na forma da distribuição condicional, sendo especialmente útil em contextos com heterogeneidade populacional e distribuições não simétricas.

Para esta aplicação, selecionou-se a distribuição SHASHo2 (Sinh-Arcsinh), cuja parametrização permite controlar diretamente a assimetria e a curtose da distribuição, por meio de transformações hiperbólicas que ajustam a cauda e o formato da densidade. Essa

distribuição pertence à classe das transformações de Johnson e é indicada para dados contínuos que apresentam caudas pesadas e assimetrias moderadas — características identificadas nos dados do ENEM do Distrito Federal.

A flexibilidade do GAMLSS com distribuição SHASHo2 permite uma modelagem mais realista da variável resposta, adequando-se às propriedades empíricas dos dados e possibilitando inferências mais robustas sobre o impacto das variáveis socioeconômicas no desempenho dos estudantes.

3.1.1 MODELAGEM COM GAMLSS

A modelagem por meio do Generalized Additive Models for Location, Scale and Shape (GAMLSS) é uma abordagem estatística altamente flexível que permite modelar diferentes parâmetros da distribuição da variável resposta — no caso, a nota obtida no ENEM — como função das variáveis explicativas.

Por meio dessa abordagem é possível avaliar não apenas o efeito médio das variáveis sobre o desempenho, mas também como elas influenciam a variabilidade e a forma da distribuição das notas, o que oferece uma visão mais abrangente sobre as desigualdades educacionais.

Portanto, a combinação entre os testes não paramétricos e o modelo GAMLSS permite uma análise estatística robusta neste caso, capturando particularidades importantes no desempenho dos estudantes ao longo do tempo, considerando múltiplas dimensões das desigualdades socioeconômicas.

A formulação geral de um modelo GAMLSS é dada por:

$$g_1(\mu) = \eta_1 = X_1\beta_1 + \sum s_{1j}(X_{1j})$$

$$g_2(\sigma) = \eta_2 = X_2\beta_2 + \sum s_{2j}(X_{2j})$$

$$g_3(v) = \eta_3 = X_3\beta_3 + \sum s_{3j}(X_{3j})$$

$$g_4(\tau) = \eta_4 = X_4\beta_4 + \sum s_{4j}(X_{4j})$$

Em que:

- a) μ : parâmetro de localização (média ou mediana da distribuição)
- b) σ : parâmetro de escala (dispersão)
- c) v : parâmetro de assimetria
- d) τ : parâmetro de curtose (cauda)
- e) $g_k(\cdot)$: função de ligação específica para o parâmetro k
- f) η_k : preditor linear (ou aditivo) de cada parâmetro
- g) $X_k\beta_k$ é a parte linear do modelo, com X_k sendo a matriz de covariáveis e β_k os coeficientes correspondentes;
- h) $\sum s_{kj}(x_{kj})$ são os termos suavizados (como splines) que permitem modelar efeitos não lineares das variáveis explicativas contínuas.

Cada um dos parâmetros pode ser modelado como uma função linear ou aditiva de covariáveis distintas.

3.1.2 ESCOLHA DA DISTRIBUIÇÃO SHASHO2

No escopo das distribuições oferecidas pelo GAMLSS, foram avaliadas diversas opções capazes de acomodar assimetrias e caudas pesadas, características observadas nos dados referentes às notas médias dos estudantes. Após comparações baseadas nos critérios de informação (Akaike Information Criterion – AIC, e Bayesian Information Criterion – BIC) e na análise dos resíduos quantílicos normalizados, a distribuição SHASHo2 (Sinh-Arcsinh, parametrização 2) demonstrou o melhor desempenho.

A distribuição SHASHo2 é uma extensão da normal padrão proposta por Jones e Pewsey (2009), posteriormente implementada no pacote `gamlss` por Stasinopoulos et al. (2017). Trata-se de uma distribuição contínua e altamente flexível, indicada para modelar variáveis que apresentam comportamento não normal, como assimetrias moderadas e caudas espessas. Essa flexibilidade se mostrou adequada para o contexto educacional analisado, no qual os pressupostos de normalidade foram claramente violados.

Sua densidade de probabilidade é dada por:

$$f(y; \mu, \sigma, \nu, \tau) = \frac{1}{\sigma} \cdot \phi \left(\sinh \left(\tau \cdot \sinh^{-1} \left(\frac{y - \mu}{\sigma} \right) - \nu \right) \right) \cdot \frac{\tau \cdot \cosh \left(\tau \cdot \sinh^{-1} \left(\frac{y - \mu}{\sigma} \right) - \nu \right)}{\sqrt{1 + \left(\frac{y - \mu}{\sigma} \right)^2}}$$

Em que:

- a) $\phi(\cdot)$: função densidade da normal padrão
- b) μ : parâmetro de localização
- c) $\sigma > 0$: parâmetro de escala
- d) $\nu \in \mathbb{R}$: controla a assimetria
- e) $\tau > 0$: controla a espessura das caudas

Componentes:

- a) $\phi(z)$:
É a função densidade da distribuição normal padrão (média 0, desvio padrão 1).
- b) $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$
- a) $\sinh^{-1}(x)$:
É o arcseno hiperbólico. É uma função inversa da seno hiperbólica.
- b) $\cosh(x)$:
É o cosseno hiperbólico.

A SHASHo2 generaliza a normal: quando $\nu=0$ e $\tau=1$, a distribuição se reduz à normal padrão, simétrica e mesocúrtica. Valores de $\nu \neq 0$ introduzem assimetria, enquanto $\tau \neq 1$ altera a curtose, possibilitando caudas mais leves ou mais pesadas.

Os modelos foram ajustados no ambiente estatístico R, por meio do pacote `gamlss`, especificando a SHASHo2 como distribuição da variável resposta. Foram adotadas funções de ligação identidade para o parâmetro de localização (μ), e logarítmica para os parâmetros de escala (σ) e curtose (τ), a fim de garantir positividade e interpretabilidade.

A avaliação da qualidade do ajuste foi conduzida por meio de ferramentas gráficas e numéricas. Destacam-se os resíduos quantílicos normalizados, cuja distribuição deve se aproximar da normal padrão, e os gráficos Q-Q, que permitem verificação visual da normalidade. Além

disso, utilizou-se o worm plot (Van Buuren & Fredriks, 2001), que fornece uma análise local dos desvios da normalidade, bem como gráficos dos resíduos em função dos preditores, a fim de identificar possíveis padrões sistemáticos.

Complementarmente, foram analisadas estatísticas descritivas dos resíduos — incluindo média próxima de zero, variância próxima de um, assimetria próxima de zero e curtose próxima de três — e a correlação de Filliben, cujos valores acima de 0,998 indicam forte aderência à normalidade. Também foram consideradas métricas globais de ajuste, como AIC, BIC e desvio global (global deviance), que permitiram comparar diferentes distribuições candidatas.

Foram testadas, ainda, distribuições alternativas, como GA (Gama), BCCG (Box-Cox Cole and Green) e BCPE (Box-Cox Power Exponential). No entanto, a SHASHo2 foi selecionada como modelo final com base em sua superioridade nos critérios de informação e na adequada distribuição dos resíduos.

3.2 BASE DE DADOS

Os dados foram coletados do site do INEP no link <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>. Esses microdados são compostos por informações fornecidas pelos candidatos no ato da inscrição. E desses microdados foram selecionadas as variáveis para este estudo.

A variável dependente utilizada foi a nota média do ENEM, calculada como a média aritmética das cinco áreas avaliadas: Linguagens e Códigos, Matemática, Ciências da Natureza, Ciências Humanas e Redação. Esta média foi adotada como uma medida sintética de desempenho global do estudante.

As variáveis explicativas incluíram aspectos socioeconômicos, escolares e individuais, conforme descrito abaixo:

Tabela 1 – Variáveis explicativas consideradas na análise

Variável	Categorias
Escolaridade do pai	Baixa (sem instrução ou ensino fundamental) Média (ensino médio), Alta (graduação ou mais)
Escolaridade da mãe	Baixa (sem instrução ou ensino fundamental) Média (ensino médio), Alta (graduação ou mais)
Renda familiar	Baixa, Média, Alta (faixas salariais categorizadas)
Tipo de escola	Pública Privada
Raça/cor	Branca, Preta, Parda, Amarela, Indígena
Gênero	Masculino, Feminino
Ano de participação	2017 a 2023

Fonte: Elaboração própria

As variáveis categóricas foram tratadas como fatores (factor), respeitando a ordem das categorias quando necessário. A variável ano foi mantida como numérica para possibilitar sua modelagem como tendência linear no tempo.

O processo de limpeza da base de dados envolveu a exclusão de participantes que não concluíram todas as provas ou deixaram a redação em branco, registros com informações faltantes nas variáveis de interesse; e casos com inconsistências (como notas iguais a zero em todas as áreas).

Após esse tratamento, obteve-se uma base final composta por aproximadamente 107.824 estudantes distribuídos ao longo dos sete anos analisados.

3.3 ESTRATÉGIA DE ANÁLISE ESTATÍSTICA

3.3.1 TESTES ESTATÍSTICOS NÃO PARAMÉTRICOS

Como parte dos procedimentos metodológicos, foram empregados testes estatísticos não paramétricos para avaliar possíveis diferenças significativas no desempenho médio dos estudantes em

relação às variáveis categóricas analisadas. A utilização de testes não paramétricos justificou-se pela violação do pressuposto de normalidade da variável resposta, verificada por meio de inspeções gráficas, como histogramas e boxplots, e confirmada pelo teste de Shapiro-Wilk. Dentre os testes utilizados, destaca-se o teste de Mann-Whitney, adequado para comparações entre dois grupos independentes. Esse teste tem por objetivo verificar se as distribuições das duas populações são estatisticamente iguais, assumindo como hipótese nula que ambas pertencem à mesma população, ou seja, não há diferença entre elas.

A hipótese alternativa, por sua vez, assume que as distribuições são diferentes. A estatística U é calculada a partir da soma dos postos das observações nos dois grupos, sendo que, em amostras grandes, essa estatística pode ser padronizada e aproximada por uma distribuição normal.

Além da significância estatística, foi considerada a análise do tamanho do efeito, uma medida complementar que expressa a magnitude da diferença entre os grupos comparados. Enquanto os testes de hipóteses indicam apenas se a diferença é estatisticamente significativa, o tamanho do efeito permite avaliar o quão relevante é essa diferença em termos práticos. No caso do teste de Mann-Whitney, o tamanho do efeito pode ser obtido pela razão entre o valor z padronizado e a raiz quadrada do número total de observações, o que resulta em um coeficiente interpretado de forma análoga ao coeficiente de correlação.

Para variáveis com mais de dois grupos independentes, como faixa de renda familiar, raça/cor e escolaridade dos pais, foi empregado o teste de Kruskal-Wallis, considerado uma extensão do teste de Mann-Whitney. Esse teste permite comparar simultaneamente três ou mais grupos, sem pressupor normalidade ou homogeneidade de variâncias, sendo indicado para dados ordinais ou intervalares com distribuição assimétrica. A hipótese nula estabelece que as distribuições das populações são iguais, enquanto a hipótese alternativa sugere que pelo menos uma das distribuições difere das demais.

Quando o teste de Kruskal-Wallis aponta a rejeição da hipótese nula, é necessário realizar uma análise de comparações múltiplas para identificar quais grupos diferem entre si. Para isso, foi adotado o teste de Dunn, indicado como procedimento post-hoc para comparações

pareadas entre os grupos. As hipóteses envolvem, nesse caso, a igualdade ou diferença entre pares de grupos. Com o intuito de controlar o erro do tipo I associado à realização de múltiplos testes, aplicou-se a correção de Bonferroni, que ajusta os valores de p para manter o rigor estatístico das conclusões inferenciais.

Esses procedimentos forneceram evidências importantes sobre desigualdades no desempenho associadas às condições socioeconômicas e escolares dos estudantes.

3.3.2 CRITÉRIOS DE COMPARAÇÃO E AJUSTE

A avaliação do ajuste dos modelos GAMLSS aplicados ao desempenho dos estudantes no ENEM foi conduzida com base em critérios estatísticos e gráficos. Os resíduos quantílicos normalizados foram examinados quanto à aderência à distribuição normal padrão, com o auxílio de gráficos Q-Q, worm plots (Van Buuren & Fredriks, 2001) e diagramas de resíduos em função dos preditores. Também foram analisadas estatísticas descritivas dos resíduos — como média, variância, assimetria e curtose — e a correlação de Filliben, considerando valores acima de 0,998 como indicativos de um ajuste satisfatório. Para a comparação entre diferentes especificações do modelo, foram utilizados os critérios AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) e o desvio global. Testaram-se distribuições alternativas, como GA, BCCG e BCPE, sendo selecionado o modelo que apresentou o menor AIC e o melhor comportamento dos resíduos, assegurando maior adequação estatística à natureza assimétrica e heterocedástica dos dados educacionais analisados.

Essas ferramentas permitiram avaliar o ajuste dos modelos e verificar a adequação dos resíduos em termos de simetria, curtose e comportamento das caudas.

Além disso, a multicolinearidade entre os preditores foi verificada por meio do índice de inflação da variância (VIF), confirmando a ausência de colinearidade severa (todos os GVIFs ajustados estavam abaixo de 1.1).

Toda a análise foi realizada no ambiente estatístico R, utilizando os seguintes pacotes: `gamlss` (RIGBY; STASINOPULOS, 2023), `car` (FOX; WEISBERG, 2019), `FSA` (OGG; MUNN, 2023), `dunn.test` (DINNO, 2017),

ggplot2 (WICKHAM, 2016), dplyr, tidyr (WICKHAM et al., 2023) e data.table (DOWLE; SRINIVASAN, 2021).

O código-fonte e as etapas de processamento estão documentados para fins de reprodutibilidade.



4

4

RESULTADOS

4.1 ANÁLISE DESCRITIVA DA BASE DE DADOS

A base de dados utilizada na análise contém um total de 107.824 registros, abrangendo informações sobre as notas médias do ENEM e características socioeconômicas dos estudantes do Distrito Federal, no período de 2017 a 2023. As variáveis disponíveis são: nota média, escolaridade do pai, escolaridade da mãe, renda familiar, tipo de escola, raça, gênero e ano da prova.

Abaixo segue tabela com uma descrição inicial das variáveis qualitativas, destacando suas frequências relativas, conforme demonstrado na Tabela 2.

Tabela 2 – Distribuição percentual das variáveis categóricas da amostra		
Variável	Categoria	Frequência (%)
Sexo	Feminino	58,8
	Masculino	41,2
Cor/Raça	Parda	48,3
	Branca	32,3
	Preta	14,3
	Amarela	2,6
	Não declarado	2
	Indígena	0,4
Tipo de escola	Pública	81,2
	Privada	15
	Exterior	3,3
	Não respondeu	0,4
Escolaridade do pai	Médio	56,8

	Baixo	26
	Alto	17,1
Escolaridade da mãe	Médio	67,5
	Baixo	19,1
	Alto	13,3
Renda familiar	Baixa	42,9
	Alta	34,5
	Média	22,6

Fonte: Dados da pesquisa. Elaboração própria.

Em relação ao perfil sociodemográfico dos participantes, observa-se que a maioria dos estudantes se identifica com o sexo feminino, representando 58,8% da amostra, enquanto 41,2% se identificam como do sexo masculino. Quanto à cor ou raça, a maior parte dos respondentes se declarou parda (48,3%), seguida por branca (32,3%) e preta (14,3%). Participantes que se identificaram como amarelos (2,6%) e indígenas (0,4%) compuseram uma proporção menor, assim como aqueles que não declararam essa informação (2,0%).

No que se refere ao tipo de escola em que cursaram o ensino médio, observa-se predominância da rede pública, com 81,2% dos estudantes, enquanto 15,0% frequentaram escolas privadas. Além disso, 3,3% estudaram no exterior e 0,4% não informaram o tipo de instituição. Em relação à escolaridade dos pais, nota-se maior concentração na faixa de escolaridade média, tanto para os pais (56,8%) quanto para as mães (67,5%). A escolaridade baixa foi observada em 26,0% dos pais e 19,1% das mães, ao passo que a escolaridade alta foi menos frequente, sendo relatada por 17,1% dos estudantes em relação ao pai e por 13,3% em relação à mãe.

Por fim, no que diz respeito à renda familiar, a maior parte dos participantes pertence à faixa de renda baixa (42,9%), seguida pela faixa de renda alta (34,5%) e, por fim, pela faixa de renda média (22,6%).

Esses resultados evidenciam um perfil de maioria feminina, parda, oriunda da rede pública de ensino, com pais e mães majoritariamente com escolaridade média e com predomínio de baixa renda familiar. Tal panorama socioeconômico e demográfico reflete

desigualdades estruturais que potencialmente influenciam o desempenho educacional dos estudantes, sendo essencial para a compreensão dos resultados posteriores da análise inferencial e modelagem estatística.

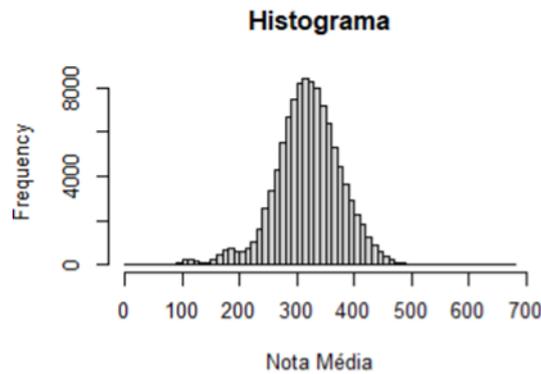
A análise descritiva das notas médias, demonstrada na tabela 3, revelou valores entre 0,01 e 675,26 pontos, com média de 318,67 e desvio padrão de 57,66. A mediana (319,81) é próxima da média, sugerindo uma distribuição relativamente simétrica, embora o teste de Anderson-Darling indique violação da normalidade ($A = 305,24$; $p < 0,001$).

Tabela 3 – Estatísticas descritivas para a variável nota média (N = 107.824)	
Estatística	Valor
Mínimo	0,01
1º Quartil	285,7
Mediana	319,8
Média	318,7
3º Quartil	355
Máximo	675,3
Desvio Padrão (sd)	57,66
Assimetria (skewness)	-0,46
Curtose (kurtosis)	1,49

Fonte: Dados da pesquisa. Elaboração própria.

A distribuição dos dados, conforme evidenciado no histograma abaixo, apresenta assimetria de -0,46 e curtose de 1,49, justificando o uso da distribuição SHASHo2 no modelo GAMLSS, dada sua capacidade de ajustar assimetria e curtose simultaneamente.

Histograma 1 – Histograma da distribuição da variável nota_media



Fonte: Dados da pesquisa. Elaboração própria.

4.2 ANÁLISE DAS VARIÁVEIS E TESTES NÃO PARAMÉTRICOS

Nesta seção, apresentam-se as estatísticas descritivas das notas médias obtidas no Exame Nacional do Ensino Médio (ENEM), organizadas por grupos definidos a partir de variáveis sociodemográficas, tais como tipo de escola, escolaridade dos pais, renda familiar, gênero e raça/cor. Para cada uma dessas categorias, exibem-se medidas de tendência central, dispersão e forma, com o objetivo de proporcionar uma compreensão inicial dos padrões de desempenho observados na amostra.

Para verificar a existência de diferenças estatisticamente significativas no desempenho médio entre os grupos, foram aplicados testes estatísticos não paramétricos adequados às características dos dados. As análises envolveram o teste de Mann-Whitney para comparação entre dois grupos independentes e o teste de Kruskal-Wallis para múltiplos grupos, seguido, quando necessário, do teste de Dunn como procedimento post-hoc, com correção de Bonferroni para múltiplas comparações.

A apresentação dos resultados é complementada por representações gráficas, por meio de histogramas que ilustram a distribuição das notas em cada grupo, facilitando a visualização de assimetrias, dispersões e padrões de concentração. Nos histogramas, o eixo horizontal representa os intervalos das notas médias, enquanto o eixo vertical indica a frequência absoluta de estudantes em cada faixa de desempenho. Dessa forma, os gráficos permitem observações mais

amplas sobre possíveis desigualdades de desempenho entre os grupos sociodemográficos investigados.

Gênero

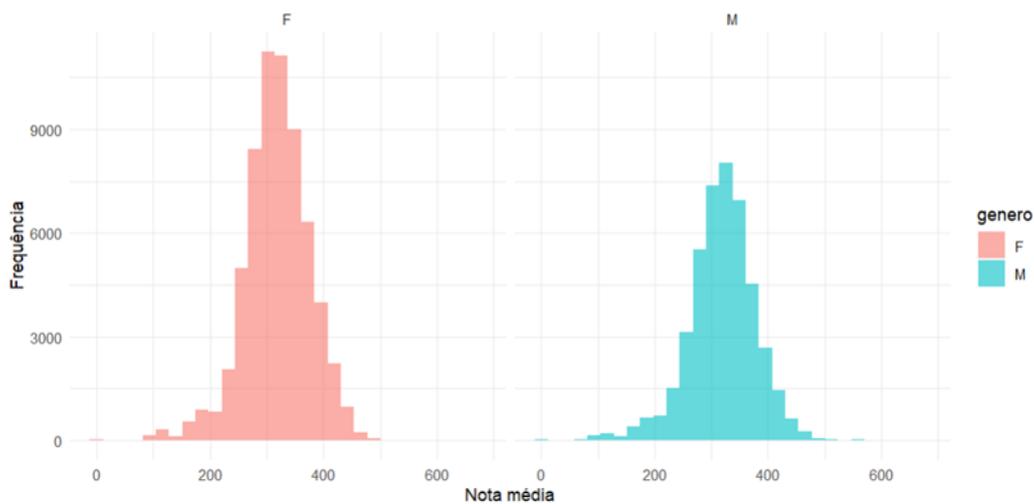
A análise das notas médias por gênero constitui a primeira etapa da investigação das possíveis desigualdades educacionais entre os grupos sociodemográficos. Para isso, apresentam-se, a seguir, as estatísticas descritivas dos participantes de acordo com o gênero informado, bem como histogramas com a distribuição das notas e os resultados dos testes estatísticos aplicados. O objetivo é verificar se há diferenças estatisticamente significativas entre os desempenhos médios de homens e mulheres e qual a magnitude dessas diferenças.

A Tabela 4 apresenta as estatísticas descritivas das notas médias segundo o gênero dos participantes. Observa-se que os estudantes do gênero masculino ($n = 44.400$) obtiveram média de 319,03 (DP = 58,48), enquanto as estudantes do gênero feminino ($n = 63.424$) apresentaram média de 318,42 (DP = 57,08). A mediana foi de 321,34 para homens e de 318,70 para mulheres. As medidas de dispersão (desvio padrão, MAD, amplitude) e de forma (assimetria e curtose) foram similares entre os grupos, sugerindo distribuições com formatos próximos.

Tabela 4 – Estatísticas descritivas das notas médias por gênero										
Gênero	n	Média	DP	Mediana	MAD	Mínimo	Máximo	Assimetria	Curtose	Erro padrão
Masculino	44.400	319	58,5	321,34	51,1	0,01	658,61	-0,47	1,65	0,28
Feminino	63.424	318,4	57,1	318,7	51,5	0,01	675,26	-0,44	1,36	0,23

Fonte: Elaboração própria.

O histograma 2 mostra que estudantes do gênero feminino tendem a concentrar-se em faixas ligeiramente superiores em relação aos do gênero masculino.

Histograma 2 – Distribuição da nota média por gênero.

Fonte: Dados da pesquisa. Elaboração própria.

Para investigar se essa diferença entre os grupos seria estatisticamente significativa, empregou-se o teste de Mann-Whitney (Tabela 5), por tratar-se de uma alternativa não paramétrica ao teste t para amostras independentes, adequada à natureza da variável dependente. Os resultados indicaram uma diferença estatisticamente significativa entre as distribuições das notas médias entre os gêneros ($W = 1.426.546.419$; $p = 0,0002292$).

Tabela 5 – Resultado do teste de Mann-Whitney para comparação das notas médias por gênero

Estatística	Valor
Teste	Wilcoxon rank sum (Mann-Whitney)
Estatística W	1.426.546.419
Valor p	0,0002292
Hipótese alternativa	Diferença de local \neq 0

Fonte: Elaboração própria.

Entretanto, a magnitude dessa diferença foi avaliada por meio do coeficiente r de efeito para o teste de Mann-Whitney (Tabela 6), o qual apresentou valor de 0,0112 (IC95%: 0,0053–0,0174). De acordo com os critérios estabelecidos por Cohen (1988), trata-se de um efeito de magnitude muito pequena. Assim, embora estatisticamente significativa, a diferença apresenta baixa relevância prática.

Tabela 6 – Tamanho de efeito (r) e intervalo de confiança para o teste de Mann-Whitney

Coeficiente r	IC 95% Inferior	IC 95% Superior
0,0112	0,0053	0,0174

Fonte: Elaboração própria.

Raça

A variável raça foi considerada nesta análise em razão de sua relevância nas discussões sobre desigualdades educacionais no Brasil. As estatísticas descritivas das notas médias dos participantes, por grupo racial autodeclarado, estão dispostas na Tabela 7. Observa-se que estudantes que se autodeclararam brancos apresentaram o maior desempenho médio (330,95), enquanto o menor desempenho foi registrado entre participantes indígenas (296,36). O grupo que não declarou raça/cor obteve média intermediária (323,30). As demais medidas de dispersão e forma revelam variabilidade relativamente semelhante entre os grupos, embora os dados sugiram maior assimetria e curtose em grupos historicamente vulnerabilizados, como o indígena e o preto.

Tabela 7 – Estatísticas descritivas das notas médias por raça

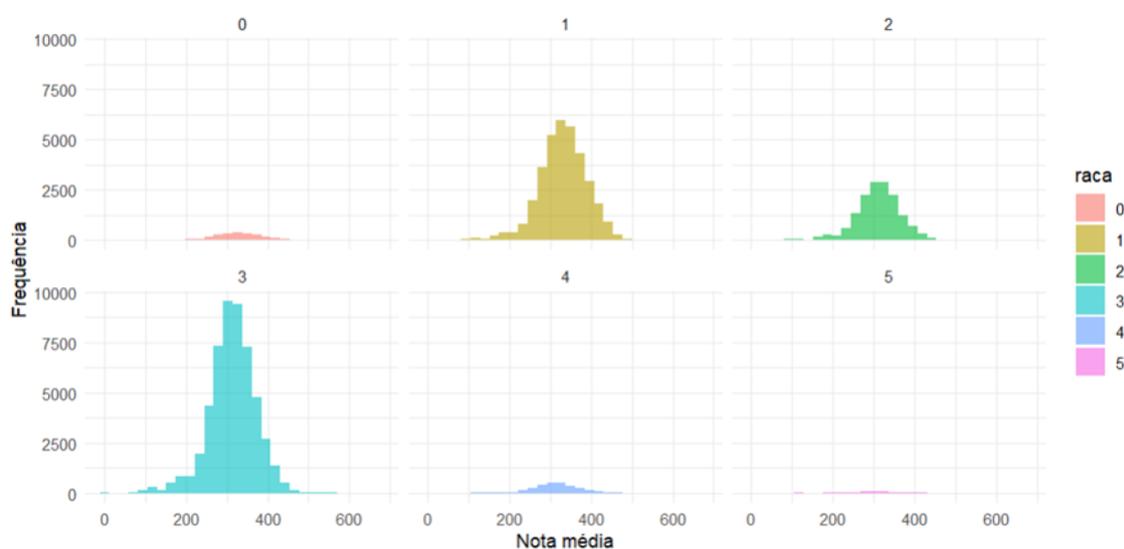
Raça/Cor	n	Média	DP	Mediana	MAD	Mínimo	Máximo	Assimetria	Curtose	Erro padrão
Branca	34.823	330,95	58,2	332,24	53,7	0,01	636,96	-0,47	1,27	0,31
Preta	15.449	310,7	55,9	313,05	47,8	0,01	675,26	-0,54	2,24	0,45
Parda	52.096	313,2	56,5	314,81	49,7	0,01	642,66	-0,5	1,6	0,25
Amarela	2.770	311,9	55,3	312,22	48	91,56	588,06	-0,3	1,43	1,05
Indígena	477	296,36	58,9	298,8	49,7	40,12	453,28	-0,62	1,61	2,69

Não declarado	2.209	323,3	59,7	324,08	55,2	49,12	586,26	-0,39	1,17	1,27
---------------	-------	-------	------	--------	------	-------	--------	-------	------	------

Fonte: Elaboração própria.

Essas tendências são reforçadas pela visualização apresentada abaixo, que exibe a distribuição das notas médias por raça/cor. O histograma mostra que o grupo branco apresenta maior concentração de estudantes em faixas de desempenho elevadas, ao passo que grupos como o indígena e o pardo demonstram distribuições mais dispersas e centradas em valores inferiores.

Histograma 3 – Distribuição da nota média dos estudantes por raça.



Legenda: Raça – 0 não declarado, 1 branca, 2 preta, 3 parda, 4 amarela, 5 indígena;

Fonte: Dados da pesquisa. Elaboração própria.

A fim de verificar se as diferenças observadas entre os grupos são estatisticamente significativas, aplicou-se o teste de Kruskal-Wallis, adequado para comparar múltiplos grupos com dados não paramétricos. Os resultados indicaram diferença estatisticamente significativa nas distribuições das notas médias entre os grupos raciais ($\chi^2 = 2662,2$; $gl = 5$; $p < 0,0001$), conforme demonstrado na Tabela 9.

Tabela 8 – Resultado do teste de Kruskal-Wallis para notas médias por raça/cor

Estatística	Valor
Teste	Kruskal-Wallis

Qui-quadrado (χ^2)	2662,2
Craus de liberdade	5
Valor p	< 0,0001

Fonte: Elaboração própria.

Com o intuito de identificar quais pares de grupos apresentaram diferenças significativas, realizou-se o teste post hoc de Dunn com correção de Bonferroni (Tabela 9). Os resultados mostram que a maioria das comparações entre pares de grupos revelou diferenças estatisticamente significativas. Exceções foram observadas nas comparações entre os grupos preto e amarelo (p-ajustado = 1,00), bem como entre os grupos pardo e amarelo (p-ajustado = 0,63), sugerindo proximidade nos padrões de desempenho entre esses perfis.

Tabela 9 – Comparações múltiplas (teste de Dunn) entre grupos de raça/cor					
Grupo 1	Grupo 2	Estatística z	p bruta	p-ajustada (Bonferroni)	Significância
Branca	Preta	-37,6	< 0,0001	< 0,0001	***
Branca	Parda	-46,1	< 0,0001	< 0,0001	***
Branca	Indígena	-13,4	< 0,0001	< 0,0001	***
Branca	Amarela	-18,2	< 0,0001	< 0,0001	***
Branca	Não declarado	6,33	< 0,0001	< 0,0001	***
Preta	Parda	4,83	< 0,0001	2,03e-5	***
Preta	Amarela	0,22	0,824	1	n.s.
Parda	Amarela	-2,03	0,0419	0,628	n.s.
Preta	Indígena	-5,45	< 0,0001	7,54e-7	***
Parda	Indígena	-6,47	< 0,0001	1,46e-9	***
Amarela	Indígena	-5,2	< 0,0001	2,93e-6	***

Fonte: Elaboração própria com base no teste `dunn_test()` com correção de Bonferroni.

Esses achados evidenciam disparidades estatisticamente significativas nos desempenhos médios entre os grupos raciais, com vantagem sistemática do grupo branco e menor desempenho entre

grupos historicamente marginalizados. Tais diferenças, embora reforçadas numericamente, devem ser interpretadas à luz dos processos estruturais que afetam as oportunidades educacionais no contexto brasileiro.

Tipo de Escola

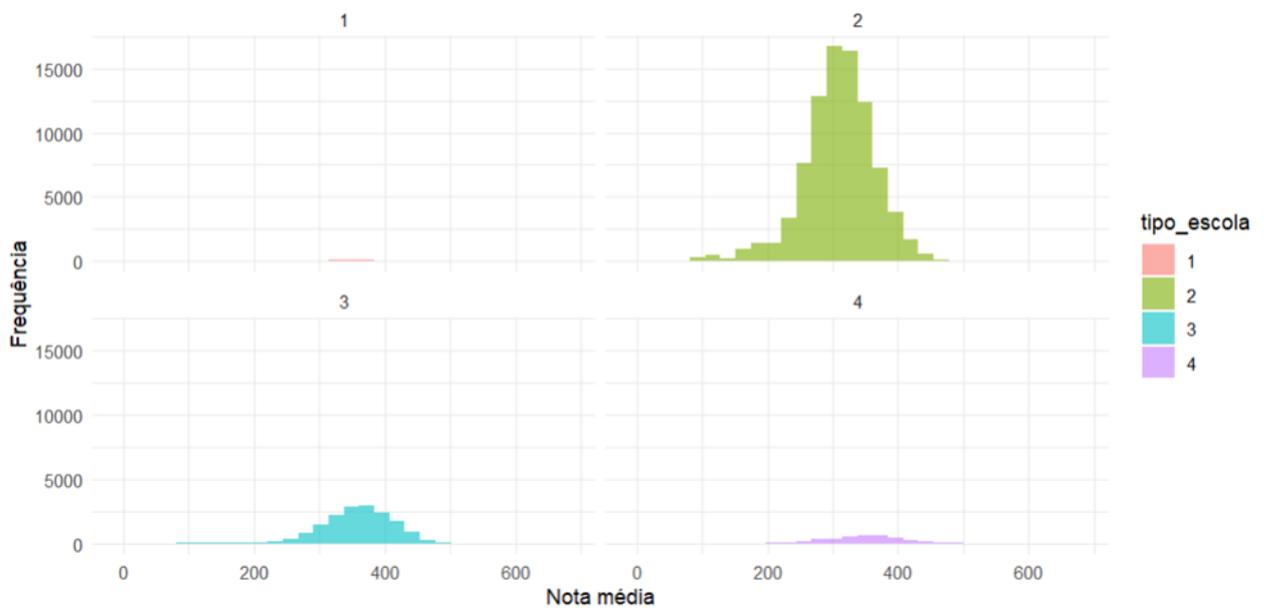
O tipo de escola frequentada pelos estudantes é uma variável relevante para a compreensão das desigualdades educacionais no Brasil. Nesta subseção, são apresentadas as estatísticas descritivas das notas médias segundo o tipo de escola informado pelos participantes: pública, privada, exterior e não declarado.

A Tabela 10 mostra que estudantes de escolas privadas obtiveram a maior média (359,28), seguidos pelos que estudaram no exterior (350,12). Participantes de escolas públicas apresentaram a menor média (309,75). O grupo que não declarou o tipo de escola teve desempenho intermediário (343,05). As medidas de dispersão foram semelhantes entre os grupos, indicando relativa homogeneidade na variabilidade interna.

Tabela 10 – Estatísticas descritivas das notas médias por tipo de escola										
Tipo de escola	n	Média	DP	Mediana	MAD	Mínimo	Máximo	Assimetria	Curto e	Erro padrão
Pública	87.587	309,75	55	312,25	47,5	0,01	675,26	-0,58	1,87	0,19
Privada	16.213	359,28	52,7	361,71	51,1	0,01	636,96	-0,56	1,6	0,41
Exterior	3.590	350,12	53	351,65	52,7	0,01	552,91	-0,27	0,96	0,88
Não respondeu	434	343,05	51	342,17	49,3	125,12	472,62	-0,13	0,49	2,45

Fonte: Elaboração própria.

Histograma 4 – Distribuição da nota média por tipo de escola (pública e privada).



Legenda: Tipo de Escola – 1 não respondeu, 2 pública, 3 privada, 4 exterior.

Fonte: Dados da pesquisa. Elaboração própria.

O histograma 4 mostra que a distribuição das notas médias do ENEM, segmentada por tipo de escola, evidencia disparidades relevantes no desempenho dos estudantes. A maior parte dos participantes está vinculada a escolas públicas, cuja distribuição é aproximadamente simétrica, com concentração das notas entre 300 e 500 pontos, e pico em torno de 400 pontos. Este padrão indica desempenho intermediário, com variação moderada entre os estudantes.

Em contraste, estudantes provenientes de escolas particulares apresentam uma distribuição mais concentrada e deslocada para a direita, com maior frequência de notas acima de 400 pontos, refletindo desempenho superior e menor dispersão. Já os estudantes oriundos de escolas no exterior apresentam distribuição mais achatada, com frequência significativamente menor e maior variabilidade nos resultados, o que pode refletir a heterogeneidade desse grupo.

O grupo de estudantes que não respondeu ao tipo de escola é numericamente irrelevante, com frequência praticamente nula.

Os dados reforçam a influência do tipo de instituição de ensino sobre o desempenho no ENEM, revelando um padrão de vantagem para os estudantes de escolas particulares e desvantagem relativa para os de escolas públicas. Tais resultados refletem desigualdades

estruturais no acesso a recursos educacionais, qualidade do ensino e suporte pedagógico.

A análise inferencial foi realizada por meio do teste de Kruskal-Wallis, apropriado para comparação de múltiplos grupos em dados não paramétricos. O teste indicou diferenças estatisticamente significativas entre os grupos ($\chi^2 = 11.872$; gl = 3; $p < 0,0001$), conforme indicado na Tabela 11.

Tabela 11 – Resultado do teste de Kruskal-Wallis para notas médias por tipo de escola	
Estatística	Valor
Teste	Kruskal-Wallis
Qui-quadrado (χ^2)	11.872
Graus de liberdade	3
Valor p	< 0,0001

Fonte: Elaboração própria.

Posteriormente, aplicou-se o teste post hoc de Dunn com correção de Bonferroni para identificar os pares de grupos com diferenças significativas. Os resultados, apresentados na Tabela 12, mostram que praticamente todas as comparações entre pares de grupos apresentaram diferenças estatisticamente significativas, com exceção da comparação entre o grupo “Não respondeu” e “Exterior”, cujo valor ajustado de p foi de 0,102.

Tabela 12 – Comparações múltiplas (teste de Dunn) entre tipos de escola					
Grupo 1	Grupo 2	Estatística z	p bruta	p-ajustada (Bonferroni)	Significância
Não respondeu	Pública	-12,2	< 0,0001	< 0,0001	***
Não respondeu	Privada	6,08	1,22e-9	7,29e-9	***
Não respondeu	Exterior	2,39	0,017	0,102	n.s.

Pública	Privada	103	< 0,0001	< 0,0001	***
Pública	Exterior	41,6	< 0,0001	< 0,0001	***
Privada	Exterior	-9,45	< 0,0001	2,01e-20	***

Fonte: Elaboração própria com base no teste `dunn_test()` com correção de Bonferroni.

Esses resultados evidenciam disparidades significativas no desempenho médio conforme o tipo de escola frequentada. O melhor desempenho entre estudantes de escolas privadas e do exterior pode refletir desigualdades de acesso a recursos educacionais, infraestrutura e preparação para o exame.

Escolaridade do Pai

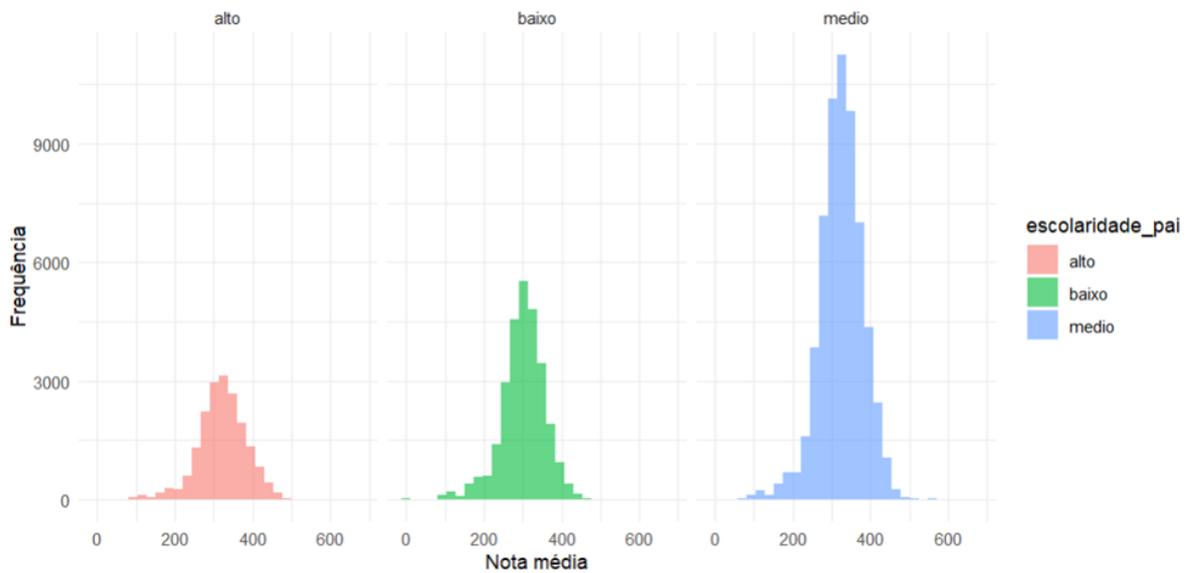
A escolaridade paterna é considerada um importante indicador de capital cultural e de contexto familiar, frequentemente associado ao desempenho educacional dos filhos. Nesta subseção, são apresentadas as estatísticas descritivas das notas médias no ENEM segundo três níveis de escolaridade do pai: alto, médio e baixo.

Como mostra a Tabela 13, estudantes cujos pais possuem escolaridade média (ensino médio completo/superior incompleto) obtiveram a maior média de notas (325,13), seguidos por aqueles do grupo com escolaridade alta (323,02). O grupo de menor desempenho foi o dos estudantes cujo pai apresenta baixa escolaridade (301,72). As distribuições são levemente assimétricas à esquerda, com valores semelhantes de desvio padrão e curtose.

Tabela 13 – Estatísticas descritivas das notas médias por escolaridade do pai										
Escolaridade do Pai	n	Média	DP	Mediana	MAD	Mínimo	Máximo	Assimetria	Curtose	Erro padrão
Baixo	28.082	301,7	56	304,53	48	0,01	636,96	-0,61	1,83	0,33
Médio	61.257	325,1	55,7	326,12	50,6	0,01	675,26	-0,43	1,5	0,23
Alto	18.485	323	61,2	323,58	55,1	0,01	623,46	-0,43	1,22	0,45

Fonte: Elaboração própria

Histograma 5 – Distribuição da nota média segundo a escolaridade do pai.



Fonte: Dados da pesquisa. Elaboração própria.

O histograma acima apresenta as notas médias por nível de escolaridade paterna. Observa-se que estudantes com pais de escolaridade média ou alta tendem a se concentrar em faixas superiores de desempenho. Já o grupo associado à baixa escolaridade apresenta maiores frequências em faixas inferiores da distribuição.

O teste de Kruskal-Wallis foi aplicado para verificar diferenças estatisticamente significativas entre os grupos (Tabela 14). O resultado evidenciou a rejeição da hipótese de igualdade entre as distribuições ($\chi^2 = 3.533,6$; $gl = 2$; $p < 0,0001$).

Tabela 14 – Resultado do teste de Kruskal-Wallis para notas por escolaridade do pai	
Estatística	Valor
Teste	Kruskal-Wallis
Qui-quadrado (χ^2)	3.533,60
Graus de liberdade	2
Valor p	< 0,0001

Fonte: Elaboração própria com base na função `kruskal.test()` do R.

O teste post hoc de Dunn com correção de Bonferroni (Tabela 15) indicou diferenças significativas entre todos os pares de grupos. As

maiores discrepâncias foram observadas nas comparações entre os grupos de escolaridade baixa e os demais, sugerindo um impacto substancial da escolaridade paterna sobre o desempenho médio.

Tabela 15 – Comparações múltiplas (teste de Dunn) por escolaridade do pai					
Grupo 1	Grupo 2	Estatística z	p bruta	p-ajustada (Bonferroni)	Significância
Alto	Médio	4,88	1,07E-06	0,00000322	***
Alto	Baixo	-40,1	< 0,0001	< 0,0001	***
Médio	Baixo	-58,4	< 0,0001	< 0,0001	***

Fonte: Elaboração própria com base no teste `dunn_test()` com correção de Bonferroni.

Os resultados reforçam a influência da escolaridade parental no desempenho educacional dos estudantes, sugerindo que níveis mais altos de instrução por parte do pai estão associados a melhores resultados no ENEM. A associação pode refletir fatores como apoio acadêmico, expectativa familiar e acesso a recursos.

Escolaridade da Mãe

A escolaridade materna é amplamente reconhecida como uma das variáveis sociodemográficas mais fortemente associadas ao desempenho escolar dos filhos, em função de fatores como incentivo educacional, engajamento familiar e disponibilidade de capital cultural. De acordo com Bourdieu e Passeron (1977), o nível educacional da mãe constitui um dos principais indicadores do capital cultural transmitido no ambiente familiar, sendo decisivo especialmente nas etapas iniciais da trajetória escolar. Ainda que o efeito da escolaridade materna tenda a diminuir ao longo das transições educacionais mais avançadas, sua influência permanece significativa inclusive no acesso e permanência no ensino superior. Esta subseção apresenta as estatísticas descritivas das notas médias segundo três faixas de escolaridade da mãe: baixa, média e alta.

A Tabela 16 mostra que quanto maior a escolaridade materna, maior a média de desempenho no ENEM. Estudantes pertencentes ao grupo de escolaridade alta apresentaram média de 336,25 pontos, seguidos pelo grupo médio (321,43). O grupo com mães de escolaridade

baixa obteve o menor desempenho médio (296,70). As distribuições indicam assimetrias negativas em todos os grupos, com maior concentração de participantes em faixas inferiores de desempenho no grupo de baixa escolaridade.

Tabela 16 – Estatísticas descritivas das notas médias por escolaridade da mãe

Escolaridade da Mãe	n	Média	DP	Mediana	MAD	Mínimo	Máximo	Assimetria	Curtose	Erro padrão
Baixo	20.625	296,7	56	300,13	47,5	0,01	588,06	-0,64	1,63	0,39
Médio	72.821	321,4	55,9	322,55	50	0,01	658,61	-0,46	1,64	0,21
Alto	14.378	336,3	60	338,09	56,4	0,01	675,26	-0,47	1,22	0,5

Fonte: Elaboração própria

As diferenças entre os grupos foram submetidas ao teste de Kruskal-Wallis, que revelou diferença estatisticamente significativa nas distribuições das notas médias conforme a escolaridade materna ($\chi^2 = 4.752,3$; gl = 2; $p < 0,0001$), conforme a Tabela 17.

Tabela 17 – Resultado do teste de Kruskal-Wallis para notas médias por escolaridade da mãe

Estatística	Valor
Teste	Kruskal-Wallis
Qui-quadrado (χ^2)	4.752,30
Graus de liberdade	2
Valor p	< 0,0001

Fonte: Elaboração própria com base na função `kruskal.test()` do R.

O teste post hoc de Dunn com correção de Bonferroni (Tabela 18) apontou diferenças estatisticamente significativas entre todos os pares de grupos, com destaque para a grande disparidade entre o grupo de escolaridade alta e o de baixa.

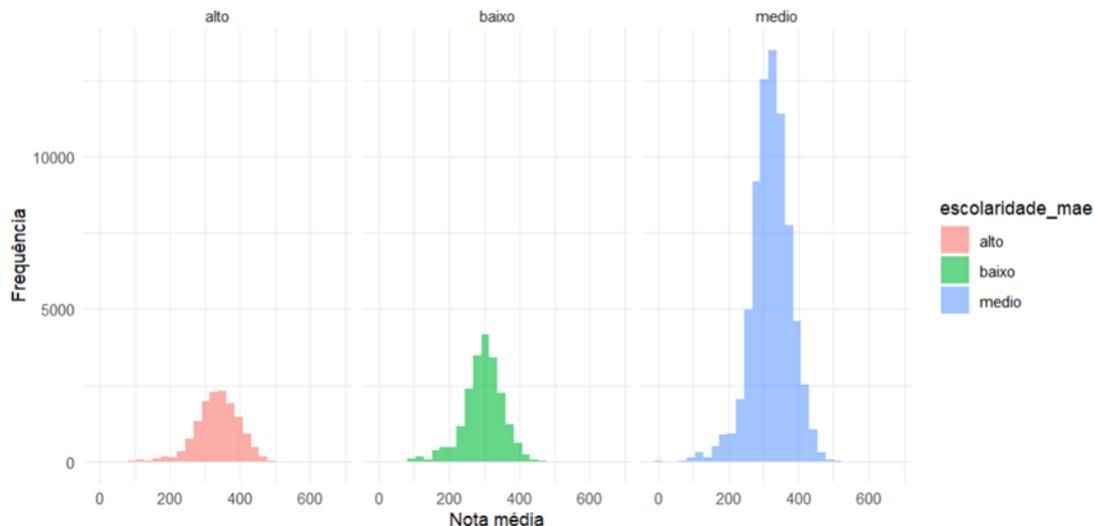
Tabela 18 – Comparações múltiplas (teste de Dunn) entre grupos de escolaridade da mãe

Grupo 1	Grupo 2	Estatística z	p bruta	p-ajustada (Bonferroni)	Significância
Alto	Médio	-29,1	< 0,0001	< 0,0001	***
Alto	Baixo	-64,9	< 0,0001	< 0,0001	***
Médio	Baixo	-55,8	< 0,0001	< 0,0001	***

Fonte: Elaboração própria com base no teste `dunn_test()` com correção de Bonferroni.

Os resultados demonstram uma associação significativa entre o nível de escolaridade da mãe e o desempenho dos estudantes no ENEM, apontando um gradiente claro: quanto maior a instrução materna, maior tende a ser a nota média dos filhos. Essa relação corrobora achados da literatura (Bourdieu e Passeron, 1977) sobre o papel da figura materna na construção de trajetórias escolares mais bem-sucedidas.

Histograma 6 – Distribuição da nota média segundo a escolaridade da mãe.



Fonte: Dados da pesquisa. Elaboração própria.

A distribuição das notas médias por escolaridade materna é apresentada no histograma 6. Observa-se nitidamente que o grupo de alta escolaridade concentra maior proporção de participantes com notas elevadas, enquanto o grupo de baixa escolaridade exibe maior frequência em faixas inferiores da distribuição. O grupo intermediário

apresenta distribuição entre os dois extremos, com média e dispersão mais próximas do grupo superior.

Renda Familiar

A renda familiar constitui uma das variáveis mais fortemente associadas ao desempenho escolar, refletindo desigualdades estruturais de acesso a recursos educacionais, culturais e tecnológicos. Nesta subseção, são apresentadas as análises descritivas e inferenciais das notas médias no ENEM segundo três faixas de renda: baixa, média e alta.

A Tabela 19 mostra uma relação direta entre o nível de renda familiar e o desempenho no exame. Estudantes pertencentes ao grupo de renda alta apresentaram média de 342,02 pontos, seguidos pelo grupo de renda média (316,90), enquanto os participantes de famílias com menor renda obtiveram a menor média (300,83). Os valores de desvio padrão e de assimetria indicam consistência nas distribuições, com leve assimetria negativa mais acentuada no grupo de renda baixa.

Tabela 19 – Estatísticas descritivas das notas médias por renda familiar										
Renda Familiar	n	Média	DP	Mediana	MAD	Mínimo	Máximo	Assimetria	Curtose	Erro padrão
Baixa	46.239	300,8	55,2	303,71	46,8	0,01	630,01	-0,62	1,88	0,26
Média	24.386	316,9	53,8	318,4	46,7	0,01	675,26	-0,54	2,11	0,34
Alta	37.199	342	54,8	343,2	51,6	0,01	642,66	-0,43	1,28	0,28

Fonte: Elaboração própria

A aplicação do teste de Kruskal-Wallis confirmou a existência de diferenças estatisticamente significativas entre os grupos ($\chi^2 = 11.406$; $gl = 2$; $p < 0,0001$), conforme apresentado na Tabela 20.

Tabela 20 – Resultado do teste de Kruskal-Wallis para notas médias por renda familiar

Estatística	Valor
Teste	Kruskal-Wallis
Qui-quadrado (χ^2)	11.406
Graus de liberdade	2
Valor p	< 0,0001

Fonte: Elaboração própria com base na função `kruskal.test()` do R.

Em seguida, o teste post hoc de Dunn com correção de Bonferroni (Tabela 21) revelou diferenças estatisticamente significativas entre todos os pares de grupos, com destaque para a disparidade entre os grupos de renda alta e baixa (estatística $z = -107$).

Tabela 21 – Comparações múltiplas (teste de Dunn) entre grupos de renda familiar

Grupo 1	Grupo 2	Estatística z	p bruta	p-ajustada (Bonferroni)	Significância
Alta	Média	-55	< 0,0001	< 0,0001	***
Alta	Baixa	-107	< 0,0001	< 0,0001	***
Média	Baixa	-36,6	< 0,0001	< 0,0001	***

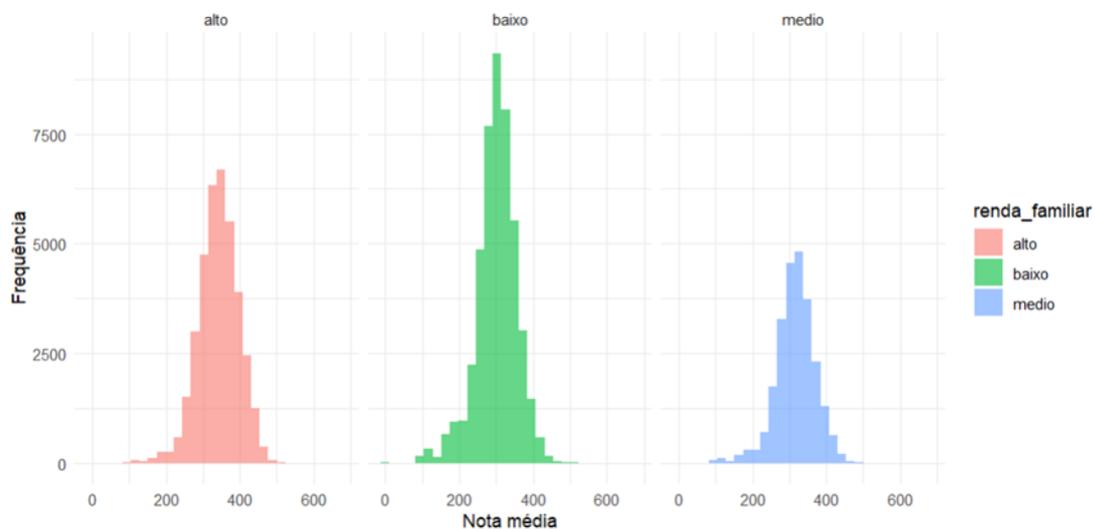
Fonte: Elaboração própria com base no teste `dunn_test()` com correção de Bonferroni.

Esses resultados reforçam a existência de um gradiente de desempenho diretamente relacionado à renda familiar. Estudantes de maior renda tendem a obter notas mais altas, o que pode refletir o acesso ampliado a recursos educacionais, melhor infraestrutura escolar, maior tempo de dedicação aos estudos e apoio extracurricular.

O histograma 7 mostra que a distribuição das notas médias do ENEM, estratificada por nível de renda familiar (alta, média e baixa), evidencia um padrão consistente de associação entre condição socioeconômica e desempenho acadêmico. Observa-se que os estudantes pertencentes ao grupo de renda alta apresentam

distribuição aproximadamente simétrica, com concentração das notas em faixas mais elevadas (entre 400 e 550 pontos), refletindo maior centralidade e menor dispersão. No grupo de renda média, a distribuição também é relativamente simétrica, porém deslocada para valores ligeiramente inferiores (300 a 450 pontos), sugerindo desempenho intermediário. Já os estudantes de renda baixa, que compõem a maior parte da amostra, apresentam notas médias concentradas entre 300 e 450 pontos, com maior dispersão e assimetria leve à esquerda.

Histograma 7 – Distribuição da nota média por faixa de renda familiar.



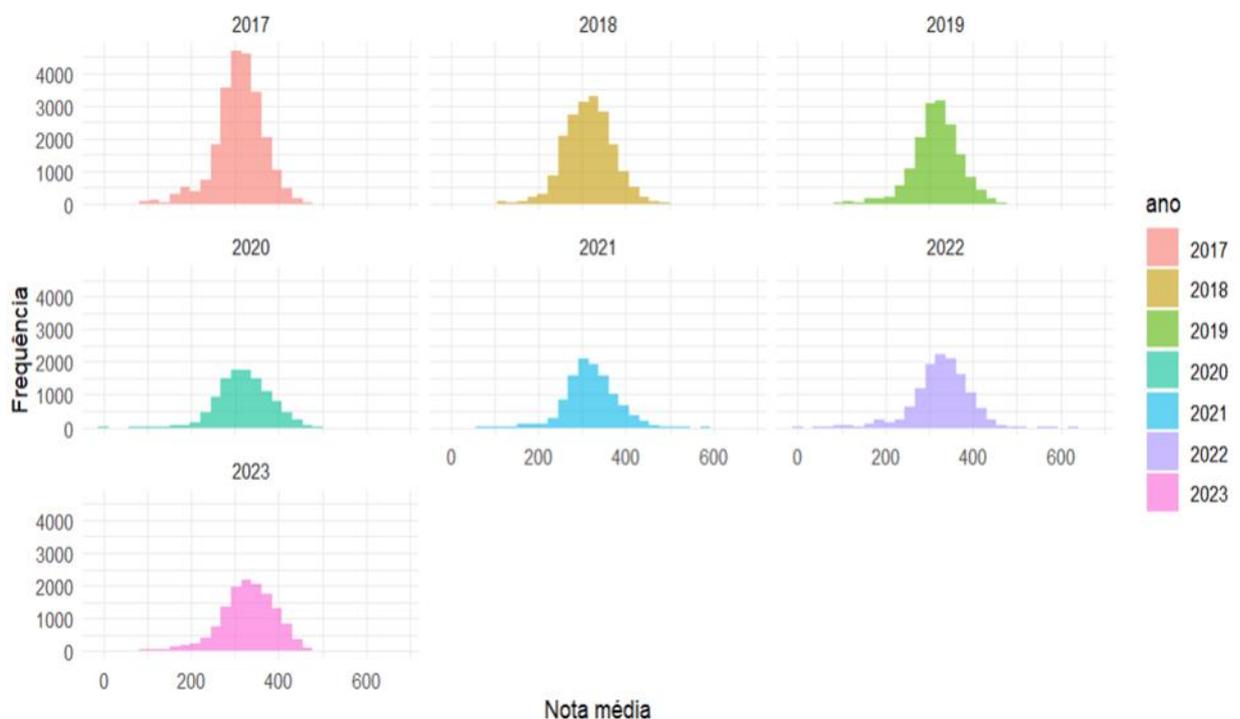
Fonte: Dados da pesquisa. Elaboração própria.

Esses resultados reforçam a hipótese de que a renda familiar exerce influência significativa sobre o desempenho dos estudantes no ENEM, revelando uma tendência de melhora no rendimento escolar conforme aumentam os recursos socioeconômicos disponíveis. A predominância de notas mais baixas entre os estudantes de baixa renda sugere a presença de desigualdades estruturais que impactam negativamente o acesso a condições adequadas de ensino e aprendizagem.

As análises descritivas e inferenciais conduzidas nesta seção revelam a presença de desigualdades significativas no desempenho médio dos participantes do ENEM, associadas a diferentes variáveis sociodemográficas. A média de notas variou sistematicamente de acordo com o tipo de escola, a raça, o gênero, a renda familiar e a escolaridade dos pais, sendo as maiores discrepâncias observadas nas variáveis de natureza estrutural, como renda e escolaridade materna. Os testes não paramétricos confirmaram que essas diferenças são

estatisticamente significativas, embora a magnitude do efeito tenha variado entre os fatores. As representações gráficas reforçaram visualmente os padrões identificados nas tabelas e facilitaram a percepção da desigualdade de distribuição entre os grupos. Esses achados fornecem evidências preliminares sobre o papel das condições sociais e educacionais no desempenho dos estudantes e fundamentam as análises inferenciais que serão conduzidas nas seções seguintes.

Histograma 8 – Distribuição da nota média ao longo dos anos (2017-2023).



Fonte: Dados da pesquisa. Elaboração própria.

No histograma 8 a análise descritiva das notas médias do ENEM entre os anos de 2017 e 2023, por meio de histogramas, permite observar variações importantes no desempenho dos estudantes ao longo do tempo. Entre 2017 e 2019, as distribuições apresentam formato aproximadamente simétrico, com maior concentração das notas na faixa de 350 a 500 pontos, sugerindo desempenho relativamente estável no período pré-pandêmico. O ano de 2017, em particular, exibe menor dispersão, indicando maior homogeneidade entre os participantes.

A partir de 2020, observam-se alterações na forma das distribuições: há redução na frequência total de estudantes e maior dispersão das notas, com caudas mais longas e distribuições levemente

assimétricas. Essas mudanças coincidem com o período da pandemia de COVID-19, o que sugere impactos estruturais sobre o desempenho escolar, provavelmente associados à suspensão das aulas presenciais, desigual acesso ao ensino remoto e intensificação das desigualdades educacionais.

Em 2022 e 2023, nota-se uma tendência de recuperação parcial, embora a variabilidade permaneça superior à observada nos anos anteriores à pandemia. As notas médias continuam concentradas entre 350 e 450 pontos, mas com maior heterogeneidade entre os participantes.

4.3 MODELAGEM ESTATÍSTICA COM GAMLSS

4.3.1 DIAGNÓSTICO DE MULTICOLINEARIDADE

Antes de iniciar a modelagem, foi realizada a verificação da multicolinearidade entre as variáveis explicativas selecionadas por meio do cálculo do Fator de Inflação da Variância (Variance Inflation Factor – VIF), utilizando um modelo de regressão linear múltipla com a nota do Exame Nacional do Ensino Médio (ENEM) como variável dependente.

Os resultados, apresentados na Tabela 19, demonstram que todos os valores ajustados de $GVIF^{1/(2 \cdot df)}$ encontram-se abaixo de 1,1, indicando níveis baixos de multicolinearidade. De acordo com Hair et al. (2009), valores de VIF abaixo de 5 não configuram problemas graves de colinearidade. Assim, conclui-se que não há necessidade de exclusão ou transformação das variáveis, podendo todas serem mantidas na modelagem estatística subsequente com o uso do modelo GAMLSS.

Tabela 22 – Valores do Fator de Inflação da Variância Generalizado (GVIF ajustado)			
Variável	GVIF	GL (Df)	$GVIF^{1/(2 \cdot df)}$
Escolaridade do pai	1,29	2	1,066
Escolaridade da mãe	1,36	2	1,08
Renda familiar	1,3	2	1,068

Tipo de escola	1,16	1	1,077
Raça/cor	1,04	1	1,019
Gênero	1,01	1	1,006
Ano	1,04	1	1,017

Fonte: Elaboração própria

4.3.2 AJUSTE DO MODELO GAMLSS

Dando continuidade à análise estatística, foi ajustado um modelo da classe GAMLSS (Generalized Additive Models for Location, Scale and Shape) para investigar a influência das variáveis socioeconômicas sobre o desempenho médio dos estudantes no Exame Nacional do Ensino Médio (ENEM) no Distrito Federal, entre 2017 e 2023. O modelo permite a modelagem simultânea de múltiplos parâmetros da distribuição da variável resposta — localização (μ), escala (σ), assimetria (ν) e curtose (τ) — proporcionando maior flexibilidade na análise dos dados.

A variável dependente desta pesquisa é a nota final do ENEM, cuja distribuição se mostrou assimétrica ao longo dos anos observados. Por essa razão, foi conduzida uma avaliação exploratória de diferentes distribuições candidatas, e a distribuição Sinh-Arcsinh tipo 2 (SHASHo2) foi selecionada por sua capacidade de acomodar assimetrias e curtose excessiva nas notas médias. Essa escolha foi fundamentada tanto na inspeção gráfica da variável resposta quanto em critérios estatísticos de ajuste de modelos preliminares.

A estimação dos parâmetros foi realizada pelo algoritmo RS (Rigby & Stasinopoulos), com convergência obtida após 21 iterações, resultando em uma Deviance Global final de 1.152.621, indicando um ajuste satisfatório do modelo aos dados.

As variáveis explicativas incluídas no modelo representam dimensões relevantes do perfil dos estudantes, como gênero, raça/cor, tipo de escola, escolaridade dos pais e renda familiar.

O modelo ajustado considerou os seguintes parâmetros da distribuição SHASHo2: a média (μ), modelada em função da escolaridade dos pais, renda familiar, tipo de escola, raça, gênero e ano; a dispersão (σ), modelada em função do gênero, tipo de escola e ano (este último com termo suavizado por B-splines); a assimetria (ν),

modelada apenas em função do tipo de escola; e a curtose (τ), modelada com base na escolaridade paterna.

Tabela 23 – Estimativas dos coeficientes para o parâmetro μ (média) do modelo GAMLSS com distribuição SHASHo2					
Variável	Coefficiente	Erro Padrão	t- valor	p-valor	Significância
(Intercepto)	355,29	3,86	92,08	<0,001	***
Escolaridad e pai (baixo)	-3,41	0,49	-6,93	<0,001	***
Escolaridad e pai (médio)	2,94	0,42	6,98	<0,001	***
Escolaridad e mãe (baixa)	-8,74	0,6	-14,67	<0,001	***
Escolaridad e mãe (média)	0,04	0,48	0,08	0,933	
Renda baixa	-23,86	0,39	-61,57	<0,001	***
Renda média	-13,61	0,42	-32,35	<0,001	***
Escola publica	-22,7	3,69	-6,15	<0,001	***
Escola privada	10,96	3,74	2,93	0,003	**
Escola exterior	1,13	3,96	0,29	0,776	
Raça branca	1,08	1,09	1	0,319	
Raça preta	-5,88	1,12	-5,24	<0,001	***
Raça parda	-6,02	1,08	-5,6	<0,001	***
Raça amarela	-7,72	1,39	-5,57	<0,001	***
Raça indígena	-15,24	2,45	-6,21	<0,001	***
Masculino	-1,72	0,3	-5,67	<0,001	***

Ano	2,38	0,07			
-----	------	------	--	--	--

Significância: " $p < 0,001$; " $p < 0,01$; " $p < 0,05$; '!' $p < 0,1$; '' não significativo.

Fonte: Elaboração própria

O intercepto (355,29) representa o valor esperado do parâmetro μ quando todas as variáveis explicativas categóricas estão no nível de referência e as contínuas iguais a zero.

Os coeficientes das variáveis categóricas indicam o efeito incremental em relação à categoria base. Por exemplo, as variáveis relacionadas à escolaridade dos pais indicam que a escolaridade baixa do pai e da mãe estão negativamente associadas ao parâmetro μ , com estimativas de -3,41 e -8,74 respectivamente, e altamente significativas ($p < 0.001$). A escolaridade média do pai tem efeito positivo (+2.94), enquanto a escolaridade média da mãe não apresenta efeito significativo ($p = 0.93$).

As variáveis de renda familiar demonstram grande impacto negativo sobre o parâmetro μ para as categorias baixa (-23,86) e média (-13,61), ambas com significância estatística elevada ($p < 0.001$), indicando uma forte associação entre menor renda e menor média da nota.

O tipo de escola também afeta significativamente μ : o tipo_escola pública possui coeficiente negativo (-22,7), indicando uma diminuição da média em relação ao grupo base, enquanto tipo_escola privada apresenta um coeficiente positivo (+10,96). O tipo_escola exterior não apresenta efeito significativo.

As categorias raciais (preta a indígena) apresentam coeficientes negativos significativos, sugerindo que esses grupos têm média da nota menor em comparação à categoria de referência (branca).

O coeficiente para o sexo masculino ($M = -1,72$) indica uma diminuição média significativa em relação ao sexo feminino.

Finalmente, o coeficiente da variável ano (+2,38) indica um aumento médio significativo da nota a cada ano, sugerindo tendência temporal positiva no desempenho.

Todos os coeficientes significativos possuem valores-p inferiores a 0.001, reforçando a robustez dos efeitos estimados. Os valores dos

erros padrão são relativamente baixos, conferindo confiabilidade às estimativas.

Tabela 24 – Estimativas dos coeficientes para o parâmetro σ (sigma) do modelo GAMLSS com distribuição SHASHo2					
Variável	Estimativa	Erro Padrão	Valor-t	p-valor	Significância
(Intercepto)	3,6458	0,04	90,3	< 2e-16	***
Escola pública	0,0217	0,04	0,54	0,5906	
Escola privada	0,0037	0,041	0,09	0,9284	
Escola exterior	0,0735	0,043	1,73	0,0835	.
Sexo Masculino	0,054	0,005	10,35	< 2e-16	***
pb(ano)	0,0208	0,001	16,7	< 2e-16	***

Significância: " p < 0,001; " p < 0,01; " p < 0,05; ' p < 0,1; ' ' não significativo.

Fonte: Elaboração própria

No modelo ajustado para o parâmetro σ (sigma), responsável por modelar a dispersão da variável resposta na distribuição SHASHo2, foi estimado com o uso da função de ligação logarítmica (log). O intercepto do modelo apresentou valor de 3,6458 (erro padrão = 0,0404; p < 0,001), representando o logaritmo da variabilidade esperada das notas médias do ENEM para o grupo de referência.

Dentre as covariáveis analisadas, o gênero masculino mostrou-se positivamente associado à dispersão das notas (coeficiente = 0,0539; p < 0,001), sugerindo maior heterogeneidade no desempenho dos estudantes desse grupo em comparação ao feminino. A variável ano, modelada de forma não paramétrica por meio da função pb(ano), também apresentou efeito estatisticamente significativo (p < 0,001), indicando que a variabilidade das notas aumentou ao longo do período analisado (2017–2023).

Por sua vez, as categorias do tipo de escola (pública, privada e exterior) não apresentaram efeitos estatisticamente significativos sobre a dispersão da nota média, com coeficientes próximos de zero e

elevados valores de p. A única exceção foi a categoria de escola no exterior, que apresentou significância marginal (coeficiente = 0,0735; $p \approx 0,084$), sugerindo uma possível tendência de maior variabilidade nesse grupo.

Considerando o uso da função de ligação logarítmica, os coeficientes devem ser interpretados de forma exponencial, de modo que valores positivos implicam aumentos percentuais na dispersão da variável resposta. Esses resultados evidenciam que fatores como gênero e ano de aplicação do exame possuem maior influência sobre a heterogeneidade do desempenho dos estudantes do que o tipo de escola frequentada.

Tabela 25 – Estimativas dos coeficientes para o parâmetro ν (nu) do modelo GAMLSS com distribuição SHASHo2

Variável	Estimativa	Erro Padrão	Valor-t	p-valor	Significância
(Intercepto)	-0,0671	0,0512	-1,309		
Escola publica	-0,0755	0,0514	-1,47		
Escola privada	-0,0673	0,052	-1,295		
Escola exterior	0,0265	0,0545	0,486		

Significância: " $p < 0,001$; " $p < 0,01$; " $p < 0,05$; ' ' $p < 0,1$; ' ' não significativo.

Fonte: Elaboração própria

No que se refere ao parâmetro ν (nu), responsável por capturar a assimetria da distribuição da nota média, este foi modelado em função da variável tipo de escola. Os resultados indicam que não há diferenças estatisticamente significativas entre as categorias dessa variável, sugerindo que a assimetria permanece relativamente constante independentemente do tipo de escola frequentado.

O intercepto foi estimado em -0,0671 (erro padrão = 0,0512; $p = 0,190$), representando o valor médio de ν para a categoria de referência. Os coeficientes associados às demais categorias — tipo_escola2 (pública), tipo_escola privada e tipo_escola exterior — variaram entre -0,075 e +0,027, todos com valores de p superiores a 0,10, o que reforça a ausência de associação estatisticamente significativa.

Assim, os resultados sugerem que, nas condições do modelo ajustado, o tipo de escola não exerce influência sistemática sobre a assimetria da distribuição das notas médias dos estudantes.

Tabela 26 – Estimativas dos coeficientes para o parâmetro τ (tau) do modelo GAMLSS com distribuição SHASHo2				
Variável	Estimativa	Erro Padrão	Valor t	Valor-p
(Intercept)	-0.30354	0.00749	-40.502	< 2e-16 ***
escolaridade_pai_baixo	-0.01267	0.00859	-1.475	0.140
escolaridade_pai_medio	0.04532	0.00764	5.932	2.99e-09 ***

Significância: " p < 0,001; " p < 0,01; " p < 0,05; ' ! p < 0,1; ' ' não significativo.

Fonte: Elaboração própria

O parâmetro τ (tau), associado à curtose da distribuição da variável resposta, reflete o grau de concentração dos valores em torno da média. A modelagem desse parâmetro foi realizada com o uso da função de ligação logarítmica. O intercepto apresentou efeito altamente significativo ($p < 0,001$), representando o valor de referência para a curtose no grupo-base.

Dentre as covariáveis analisadas, apenas a categoria “escolaridade_pai_medio” apresentou efeito positivo e estatisticamente significativo ($p < 0,001$), sugerindo que estudantes cujos pais possuem escolaridade de nível médio tendem a apresentar maior concentração das notas em torno da média — ou seja, uma distribuição com curtose mais elevada. Por outro lado, a categoria “escolaridade_pai_baixo” não apresentou significância estatística ($p = 0,140$), indicando ausência de efeito sistemático sobre a curtose.

Esses resultados sugerem que a escolaridade dos pais exerce influência sobre o formato da distribuição das notas médias, sendo o nível médio de escolaridade paterna um fator associado à maior concentração dos escores próximos à média.

4.3.3 EFEITO DAS VARIÁVEIS NA MÉDIA DAS NOTAS (M)

Os resultados indicam que a escolaridade dos pais, a renda familiar, o tipo de escola, a raça, o gênero e o ano têm impactos estatisticamente significativos na média das notas dos estudantes.

No que diz respeito à escolaridade paterna, observou-se efeitos diferenciados: alunos cujos pais possuem baixa escolaridade obtiveram, em média, 3,41 pontos a menos do que aqueles com escolaridade paterna alta, enquanto estudantes com escolaridade paterna média apresentaram um aumento médio de 2,94 pontos ($p < 0,001$). Já a escolaridade materna baixa teve um impacto ainda mais expressivo, reduzindo a nota média em 8,74 pontos ($p < 0,001$), ao passo que a escolaridade materna média não apresentou efeito significativo sobre o desempenho dos estudantes.

A renda familiar revelou-se fortemente associada ao desempenho acadêmico, com estudantes provenientes de famílias de baixa renda apresentando notas 23,86 pontos inferiores e aqueles de renda média, 13,61 pontos menores, em comparação aos estudantes de renda alta ($p < 0,001$). O tipo de escola também se mostrou um fator determinante: os alunos da rede pública tiveram uma redução média de 22,70 pontos em suas notas ($p < 0,001$), enquanto os da rede privada registraram um aumento médio de 10,96 pontos ($p = 0,003$). A categoria referente a escolas no exterior, entretanto, não evidenciou efeito estatisticamente significativo.

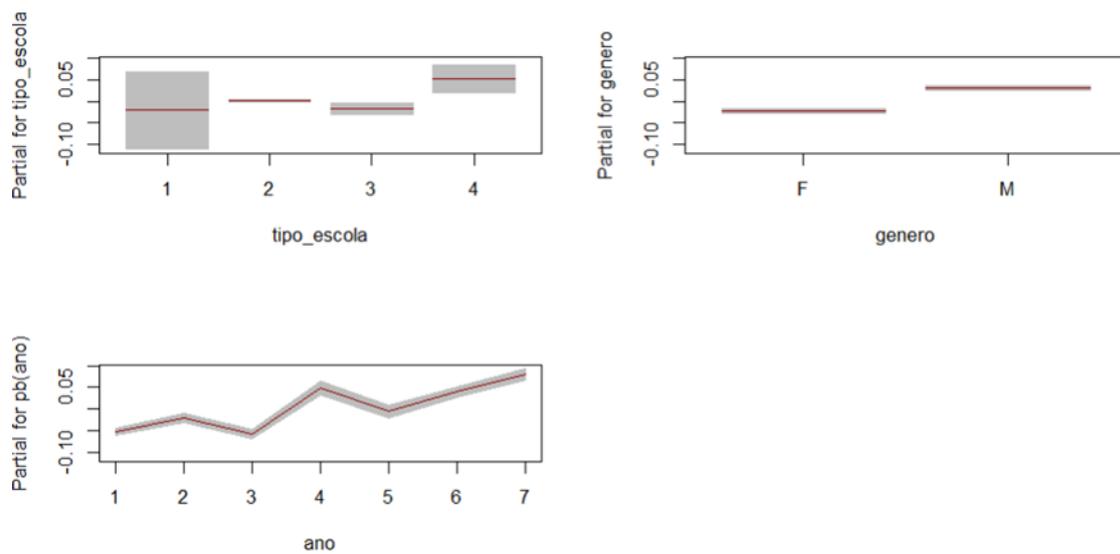
Quanto à raça, os coeficientes indicaram reduções significativas nas médias das notas quando comparadas ao grupo de referência, com variações que vão de -5,88 pontos para estudantes autodeclarados pretos até -15,24 pontos para os indígenas, todos com $p < 0,001$. Além disso, o gênero masculino foi associado a uma redução média de 1,72 pontos em relação ao feminino ($p < 0,001$). Por fim, a variável ano apresentou coeficiente positivo (2,38; $p < 0,001$), evidenciando uma tendência de aumento contínuo nas notas médias dos estudantes no período de 2017 a 2023.

4.3.4 VARIABILIDADE, ASSIMETRIA E CURTOSE DAS NOTAS

O parâmetro de dispersão (σ) foi modelado com base nas variáveis tipo de escola, gênero e ano. Observou-se que estudantes do

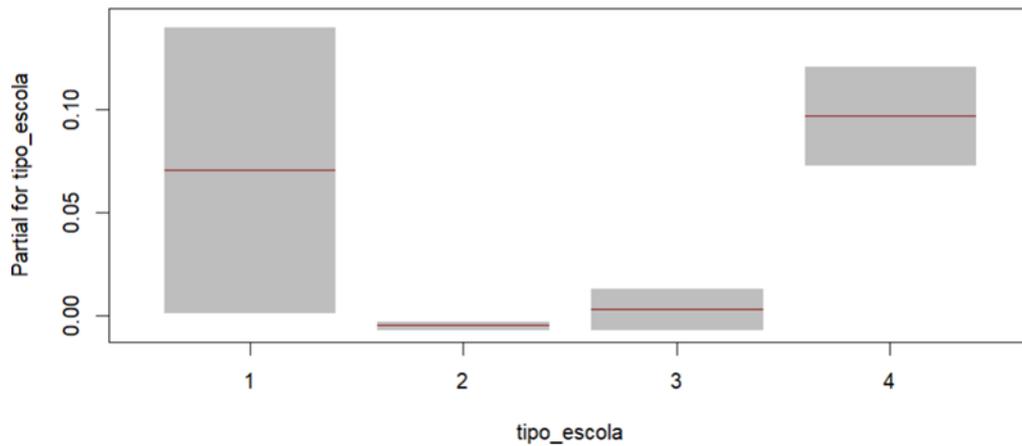
sexo masculino apresentaram maior variabilidade nos resultados, indicando maior heterogeneidade nas notas nesse grupo. A variável ano, modelada por meio de função B-spline, revelou uma tendência de aumento progressivo na dispersão ao longo do tempo, evidenciando maior variação das notas nos anos mais recentes. Por outro lado, o tipo de escola não apresentou efeito estatisticamente significativo sobre esse parâmetro.

Figura 1 – parâmetro de dispersão (σ)



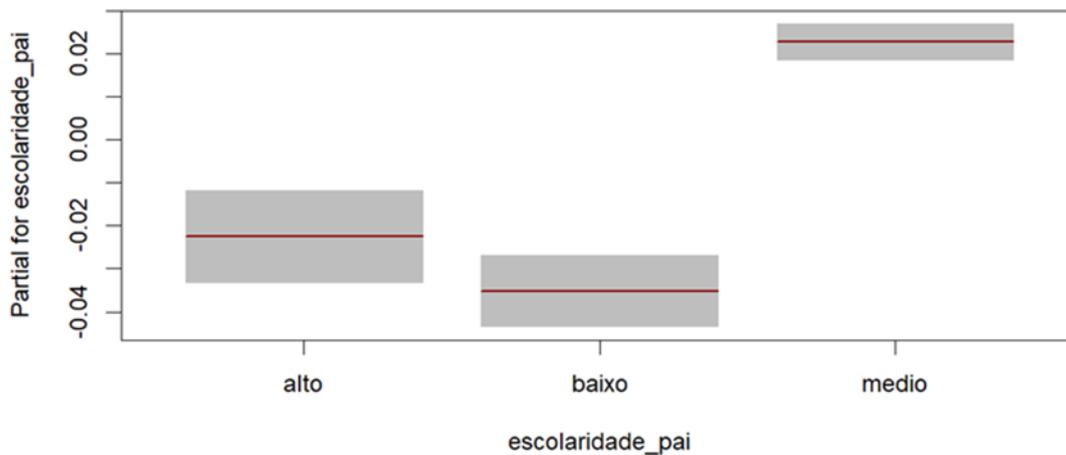
Fonte: Elaboração própria

No que se refere à assimetria (ν), que foi modelada exclusivamente em função do tipo de escola, os coeficientes estimados não atingiram significância estatística. Isso sugere que a simetria da distribuição das notas manteve-se relativamente constante entre os diferentes tipos de escola, sem distorções relevantes nas caudas.

Figura 2 – parâmetro da assimetria (v)

Fonte: Elaboração própria

Por fim, o parâmetro de curtose (τ) foi significativamente influenciado pela escolaridade paterna. Especificamente, estudantes cujos pais possuem escolaridade média apresentaram maior curtose (coeficiente positivo), indicando uma distribuição mais concentrada em torno da média, isto é, mais pontiaguda. Esse padrão pode refletir menor variabilidade interna de desempenho dentro desse grupo.

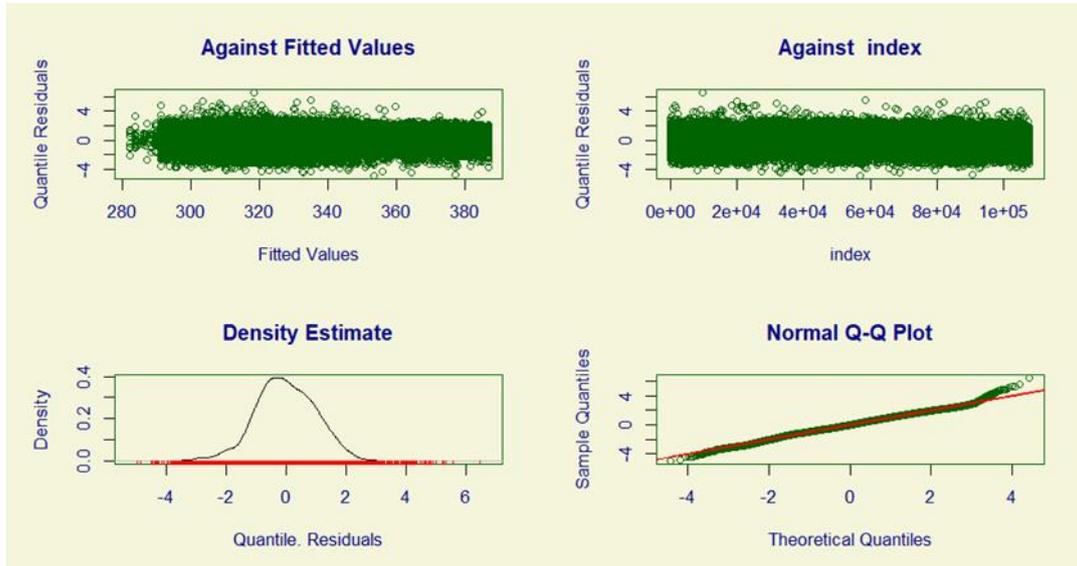
Figura 3 – parâmetro de curtose (τ)

Fonte: Elaboração própria

A qualidade do ajuste do modelo foi avaliada por meio da análise dos resíduos quantílicos. Os gráficos de diagnóstico indicam que os resíduos apresentam comportamento aproximadamente aleatório em relação aos valores ajustados e à ordem dos dados, o que sugere a ausência de padrões sistemáticos. A distribuição dos resíduos aproxima-se da normalidade, com leve assimetria e caudas discretamente mais longas, o que é coerente com o uso da distribuição

SHASHo2, capaz de modelar assimetrias e curtose excessiva. Tais resultados indicam que o modelo é adequado para representar a estrutura dos dados.

Figura 4 – Diagnóstico dos resíduos quantílicos do modelo GAMLSS com distribuição SHASHo2



Fonte: Elaboração própria

A Tabela 27 sintetiza os principais indicadores numéricos de adequação dos resíduos quantílicos do modelo ajustado. Os valores observados estão próximos aos esperados sob a hipótese de resíduos com distribuição normal padrão — reforçando, de forma complementar à análise gráfica, a qualidade do ajuste obtido por meio do GAMLSS.

Tabela 27 – Resumo estatístico dos resíduos quantílicos do modelo GAMLSS	
Estatística	Valor
Média dos resíduos	0,0052
Variância dos resíduos	0,9956
Coefficiente de assimetria	-0,0631
Coefficiente de curtose	3,2004
Correlação de Filliben (Q-Q normal plot)	0,9988

Fonte: Elaboração própria com base nos resíduos do modelo_flex2.

A fim de visualizar os efeitos individuais de cada preditor sobre os parâmetros do modelo GAMLSS, foram gerados gráficos marginais para os componentes da distribuição SHASHo2 estimada: média (μ), dispersão (σ), assimetria (ν) e curtose (τ). Esses gráficos ilustram como cada variável contribui para explicar a forma da distribuição das notas médias, considerando os demais efeitos mantidos constantes.

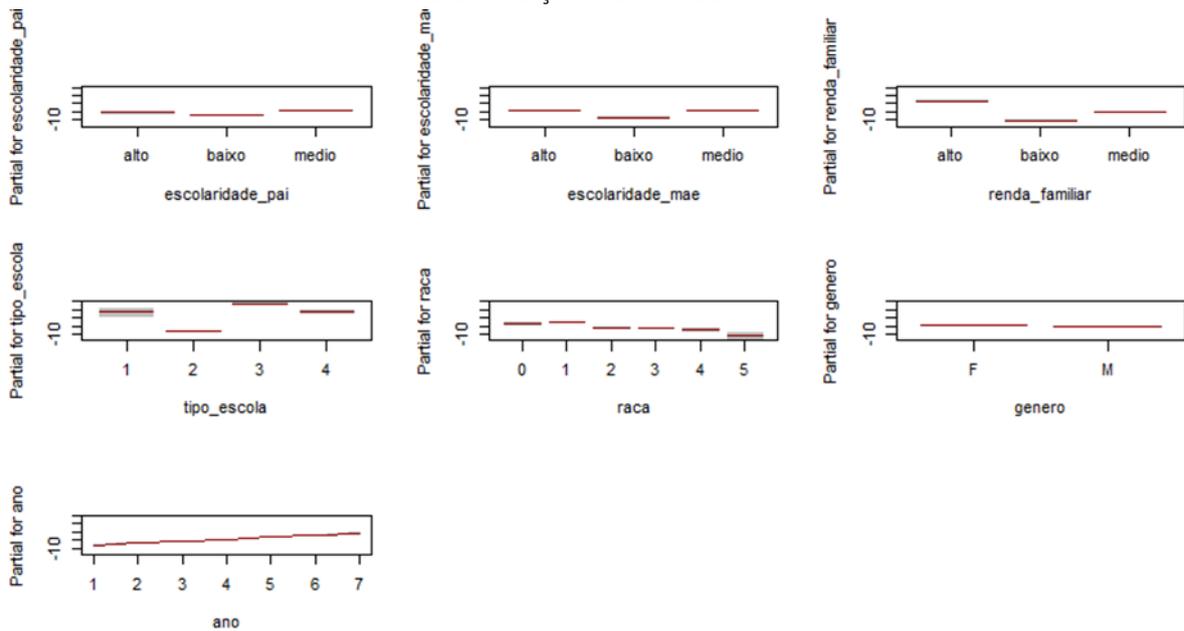
No contexto do modelo GAMLSS, os efeitos marginais correspondem à variação esperada nos parâmetros da distribuição da variável resposta (média, dispersão, assimetria e curtose) em função de pequenas mudanças nas variáveis explicativas, mantendo-se as demais constantes. Esses efeitos fornecem uma interpretação mais intuitiva do impacto das covariáveis sobre diferentes aspectos da distribuição das notas, além do simples efeito médio.

Diferentemente dos modelos tradicionais que focam apenas na média da variável resposta, o GAMLSS permite modelar simultaneamente múltiplos parâmetros da distribuição. Assim, os efeitos marginais revelam como, por exemplo, o gênero ou o tipo de escola influenciam não só o desempenho médio, mas também a variabilidade (σ), a simetria (ν) e a concentração dos valores extremos (τ) das notas.

Os gráficos marginais ilustram esses efeitos ao apresentar as mudanças estimadas nos parâmetros da distribuição conforme variáveis específicas variam, possibilitando uma análise visual e detalhada das diferenças entre grupos e ao longo do tempo. Dessa forma, a utilização dos efeitos marginais torna-se essencial para compreender as múltiplas dimensões do desempenho escolar e as desigualdades educacionais refletidas nos dados.

Assim para visualizar os efeitos marginais das variáveis explicativas sobre os parâmetros da distribuição da nota média dos estudantes, foram gerados gráficos dos efeitos ajustados pelo modelo GAMLSS com a distribuição SHASHo2. Esses gráficos estão apresentados abaixo:

Figura 5 – Efeitos marginais das variáveis explicativas sobre a assimetria (v) da distribuição das notas



Fonte: Elaboração própria

No caso do parâmetro σ (dispersão), observou-se que o gênero masculino e o avanço dos anos contribuem para o aumento da variabilidade das notas. Isso indica que, além das diferenças médias de desempenho, há maior heterogeneidade interna em certos grupos, o que pode refletir desigualdade de oportunidades educacionais.

O parâmetro v (assimetria), embora não tenha apresentado significância estatística elevada, sugere diferenças sutis na simetria das distribuições de notas entre os tipos de escola. Essa característica pode estar relacionada à presença de subgrupos extremos dentro de determinados contextos escolares.

Por fim, o parâmetro τ (curtose) apresentou associação significativa com a escolaridade paterna. Estudantes com pais de escolaridade média apresentaram distribuições mais concentradas em torno da média, com menor dispersão nos extremos, refletindo padrões de desempenho mais homogêneos.



5

5

CONCLUSÃO

Este estudo teve como objetivo investigar os efeitos das variáveis socioeconômicas e demográficas sobre o desempenho médio dos estudantes do Distrito Federal no ENEM, no período de 2017 a 2023, por meio da modelagem estatística GAMLSS com distribuição SHASHo2. Essa abordagem demonstrou-se adequada e eficaz para capturar não apenas os impactos sobre a média das notas, mas também sobre sua dispersão, assimetria e curtose.

Os resultados principais evidenciam que fatores como renda familiar, escolaridade dos pais, tipo de escola, raça e gênero exercem influência significativa no desempenho dos estudantes. Destaca-se que alunos oriundos de contextos menos favorecidos apresentam médias inferiores e maior variabilidade nas notas, refletindo desigualdades estruturais ainda presentes no sistema educacional brasileiro.

Embora a evolução positiva da média das notas ao longo dos anos indique avanços, o aumento da dispersão revela que tais progressos não se distribuem de forma homogênea entre os diferentes grupos sociais. Esses achados reforçam a urgência de políticas públicas voltadas à redução das desigualdades educacionais, com ênfase na valorização da escola pública, no apoio às famílias de baixa renda e no combate às disparidades raciais e de gênero.

Do ponto de vista metodológico, o modelo GAMLSS mostrou-se uma ferramenta robusta para análises educacionais complexas, proporcionando maior precisão e detalhamento em comparação a abordagens tradicionais. Espera-se que este trabalho contribua não apenas para o entendimento das desigualdades educacionais no Distrito Federal, mas também para a disseminação de metodologias estatísticas mais avançadas em estudos correlatos.

Adicionalmente, reconhece-se que este estudo não abordou aspectos territoriais das desigualdades, como as variações de desempenho entre escolas localizadas nas diferentes regiões administrativas do Distrito Federal (cidades-satélites), e outras questões acerca de instalações, formação docente e outras questões que podem influenciar nos resultados. Futuras investigações poderão

ampliar essa análise, assim como incorporar outras dimensões institucionais e contextuais não consideradas no presente modelo.

Os resultados aqui apresentados oferecem subsídios valiosos para a formulação e o aprimoramento de políticas públicas educacionais no Distrito Federal. Ao identificar grupos sociais mais vulneráveis e as dimensões específicas em que as desigualdades se manifestam, espera-se que gestores e formuladores de políticas possam direcionar esforços e recursos de forma mais estratégica — promovendo ações voltadas à equidade e inclusão, com potencial para impactar positivamente a qualidade e a justiça educacional no Distrito Federal.



REFERÊNCIAS

REFERÊNCIAS

REFERÊNCIAS

ALBANEZ, Alicia; FERREIRA, Francisco e FRANCO, Creso. (2002), "A Escola Importa? Determinantes da Eficiência e Equidade no ensino Fundamental Brasileiro". Pesquisa e Planejamento Econômico, vol. 23, pp. 453-476.

ALMEIDA, L. C. As desigualdades e o trabalho das escolas: problematizando a relação entre desempenho e localização socioespacial. Revista Brasileira de Educação, v. 22, n. 69, p. 361–384, abr. 2017.

ALVES, Fátima. (2007), Qualidade na Educação Fundamental Pública nas Capitais Brasileiras: Tendências, Contextos e Desafios. Rio de Janeiro, Departamento de Educação, Pontifícia Universidade Católica do Rio de Janeiro, Tese de doutorado.

ALVES, João; RIBEIRO, Maria. A relação entre a escolaridade dos pais e o desempenho no ENEM. Revista Brasileira de Educação, v. 22, n. 3, p. 123-145, 2017.

BARBOSA, Maria Lígia de O. (2009), Desigualdade e Desempenho Belo Horizonte, Argymentvm.

BOURDIEU, Pierre; PASSERON, Jean-Claude. A reprodução: elementos para uma teoria do sistema de ensino. Rio de Janeiro: Francisco Alves, 1975.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Desempenho no ENEM: análise de fatores socioeconômicos e educacionais [online]. Brasília, 2021. Disponível em: <https://www.inep.gov.br>. Acesso em: 31 out. 2024.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Microdados do ENEM. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>. Acesso em: 14 out. 2024.

CASTRO, G.; TIEZZI, S. A reforma do ensino médio e a implantação do Enem no Brasil. Desafios, v. 65, n. 11, pp. 46-115. 2004.

DINNO, Alexis. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. R package version 1.3.5, 2017. Disponível em: <https://CRAN.R-project.org/package=dunn.test>. Acesso em: 23 jun. 2025.

DOWLE, Matt; SRINIVASAN, Arun. data.table: Extension of 'data.frame'. R package version 1.14.8, 2021. Disponível em: <https://CRAN.R-project.org/package=data.table>. Acesso em: 23 jun. 2025.

ERIKSON, Robert; GOLDTHORPE, John H. The Constant Flux: A Study of Class Mobility in Industrial Societies. Oxford: Clarendon Press, 1992.

ERNICA, Maurício; BATISTA, Antônio Augusto Gomes. A escola, a metrópole e a vizinhança vulnerável. Cadernos de Pesquisa, São Luiz, v. 42, p. 640-666, 2012.

FEIJÓ, J. R.; FRANÇA, J. M. S. D.; PINHO NETO, V. R. D. Desempenho dos estudantes ao final do ensino médio: Mensurando a influência direta e indireta da educação dos pais. Revista Brasileira de Economia, v. 76, n. 1, p. 30-56, jan. 2022.

FOX, John; WEISBERG, Sanford. An R Companion to Applied Regression. 3. ed. Thousand Oaks: Sage, 2019.

FRANCO, Ana Maria Paiva; MENEZES FILHO, Naércio. Uma análise de rankings de escolas brasileiras com dados do SAEB. Estudos Econômicos, São Paulo, v. 42, n. 2, p. 263-283, 2012.

JONES, M. C., & Pewsey, A. (2009). Sinh-arcsinh distributions. Biometrika, 96(4), 761-780.

LIBÂNEO, J. C. (2012). O dualismo perverso da escola pública brasileira: escola do conhecimento para os ricos, escola do acolhimento social para os pobres. Educação e Pesquisa, 38(1), 13-28.

LIMA, C. C. V.; BRIGHENTI, C. R. G. Desempenho de estudantes de Minas Gerais no Exame Nacional do Ensino Médio considerando variáveis socioeconômicas. Educação e Pesquisa, v. 49, p. e253303, 2023.

LIMA, Cayo César Viana de; BRIGHENTI, Carla Regina Guimarães. Desempenho de estudantes de Minas Gerais no Exame Nacional do Ensino Médio considerando variáveis socioeconômicas. Educação e Pesquisa, São Paulo, v. 49, e253303, 2023

LIMA, P. J. (2015). Crítica sociológica do Exame Nacional do Ensino Médio: uma análise bourdiana. Encontro Nacional de Pesquisa em Educação em Ciências, 10, 1-8.

LOUZANO, P. (2013). Fracasso escolar: evolução das oportunidades educacionais de estudantes de diferentes grupos raciais. Cadernos Cenpec, 3(1), 111-133.

LUCAS, S. (2001), "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects". *American Journal of Sociology*, vol. 106, nº 6, pp. 1642-1690.

MELLO, R. D. N., MEDEIROS, H. A. V., PAIVA, F. S., & SIMÕES, J. L. (2014). O impacto do Enem nas políticas de democratização do acesso ao ensino superior brasileiro. *Comunicações*, 21(3), 109-123.

MUZZETI, L. R. (1999). Consenso ou conflito: contribuições das teorias sociológicas de Emile Durkheim e de Pierre Bourdieu. *Boletim do Departamento de Didática*, 15, 43-62.

OGG, Andrew; MUNN, Martin B. FSA: Simple Fisheries Stock Assessment Methods. R package version 0.9.5, 2023. Disponível em: <https://CRAN.R-project.org/package=FSA>. Acesso em: 23 jun. 2025.

OLIVEIRA, Thiago Soares de. O ENEM: breves considerações sobre importância avaliativa e reforma educacional. *Educação Por Escrito*, Porto Alegre, v. 7, n. 2, p. 278-288, jul.-dez. 2016

RIGBY, R. A., & Stasinopoulos, D. M. (2005). Generalized Additive Models for Location, Scale and Shape (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507–554.

RIGBY, Robert A.; STASINOPULOS, Mikis. gamlss: Generalised Additive Models for Location Scale and Shape. R package version 5.4-11, 2023. Disponível em: <https://CRAN.R-project.org/package=gamlss>. Acesso em: 23 jun. 2025.

SILVA, Anna Camila Lima e; MOTA, Renata de Oliveira; LIMA, João Cardim Ferreira; QUEIROZ, Fernanda Cristina Barbosa Pereira; NORONHA, Samara Leite. A influência da escolaridade dos pais e da renda familiar no desempenho dos candidatos do ENEM. In: ANEGEP. *Anais da ANEGEP 2017*. Joinville/SC 2017.

SOARES, José Francisco. (2004), "Qualidade e Equidade na Educação Básica Brasileira: A Evidência do SAEB-2001". *Archivos Analíticos de Políticas Educativas*, nº 12, pp. 112-167.

STASINOPOULOS, D. M., Rigby, R. A., & Heller, G. Z. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*. CRC Press.

VAN BUUREN, S., & Fredriks, A. M. (2001). Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, 20(8), 1259–1277.

VIGGIANO, Esdras; MATTOS, Cristiano. O desempenho de estudantes no ENEM 2010 em diferentes regiões brasileiras. Revista Brasileira de Estudos Pedagógicos, Brasília, DF, v. 94, n. 237, p. 417-438, 2013

WEIS, C. (2014) Os direitos humanos contemporâneos. São Paulo: Malheiros.

WICKHAM, Hadley et al. dplyr: A Grammar of Data Manipulation; tidyr: Tidy Messy Data. R packages version 1.1.4 e 1.3.0, 2023. Disponível em: <https://CRAN.R-project.org/package=dplyr> e <https://CRAN.R-project.org/package=tidyr>. Acesso em: 23 fev. 2025.

WICKHAM, Hadley. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag, 2016.



idn

Bo
pro
cit
ref
Ness
são e

idp

A ESCOLHA QUE
TRANSFORMA
O SEU CONHECIMENTO