



Universidade Federal
de Ouro Preto

**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Departamento de Computação e Sistemas**

**Associação entre Perfil
Socioeconômico e as Redes de Ensino
Médio no Desempenho do ENEM**

Carlos Vitor Ferreira de Moura

**TRABALHO DE
CONCLUSÃO DE CURSO**

ORIENTAÇÃO:

Felipo Bacani

COORIENTAÇÃO:

Helen de Cássia Sousa da Costa Lima

Julho, 2026

João Monlevade–MG

Carlos Vitor Ferreira de Moura

**Associação entre Perfil Socioeconômico e as
Redes de Ensino Médio no Desempenho do
ENEM**

Orientador: Felipe Bacani

Coorientadora: Helen de Cássia Sousa da Costa Lima

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Universidade Federal de Ouro Preto

João Monlevade

Julho de 2026



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIÊNCIAS EXATAS E APLICADAS
DEPARTAMENTO DE CIÊNCIAS EXATAS E APLICADAS



FOLHA DE APROVAÇÃO

Nome do discente: Carlos Vitor Ferreira de Moura

Nome do trabalho: Associação entre Perfil Socioeconômico e as Redes de Ensino Médio no Desempenho do ENEM

Monografia apresentada ao Curso de Sistemas de Informação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Sistemas de Informação

Aprovada em 28 de julho de 2026

Membros da banca

Prof. Dr. Felipe Bacani - Orientador (Universidade Federal de Ouro Preto)
Profa. Dra. Helen de Cassia Sousa da Costa Lima - Coorientadora (Universidade Federal de Ouro Preto)
Prof. Dr. Fernando Bernardes de Oliveira - (Universidade Federal de Ouro Preto)
Profa. Dra. Erica de Matos Reis Ferreira Ferreira

Felipo Bacani, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 03/08/2026



Documento assinado eletronicamente por **Felipo Bacani, PROFESSOR DE MAGISTERIO SUPERIOR**, em 03/08/2026, às 19:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1151673** e o código CRC **E9E39046**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.008688/2026-19

SEI nº 1151673

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163
Telefone: (31)3808-0816 - www.ufop.br

Dedico este trabalho à minha família, minha base e minha fortaleza. Obrigado por todo o amor, apoio e compreensão ao longo desta jornada. Vocês são a minha inspiração e a razão de todas as minhas conquistas. Esta vitória também é de vocês.

Agradecimentos

Primeiramente, agradeço a Deus por me conceder força, sabedoria e perseverança para concluir esta importante etapa da minha vida. À minha amada família, meu eterno agradecimento por todo o amor, apoio incondicional e por sempre acreditarem em mim, mesmo nos momentos mais desafiadores.

Um agradecimento mais que especial aos meus queridos amigos Jhonatan e Eutran, que foram meus companheiros não apenas de apartamento, mas de vida durante a graduação. Juntos, enfrentamos os desafios de montar nosso apartamento, de equilibrar as contas e as responsabilidades. Mas, acima de tudo, construímos uma amizade sólida, compartilhando momentos únicos, risadas, lágrimas e aprendizados que levarei para sempre em meu coração.

Ao meu dedicado orientador, Felipo, minha sincera gratidão por todo o conhecimento compartilhado, pela paciência e por me guiar nessa jornada acadêmica.

“A morte não extingue a luz, apenas apaga a lamparina porque chegou a aurora.”

— Chico Xavier

Resumo

Este trabalho investiga a relação entre variáveis socioeconômicas e o tipo de escola dos participantes do ENEM entre 2019 e 2023, a partir dos microdados públicos disponibilizados pelo INEP. Após o pré-processamento das cinco edições do exame, com a remoção de ausentes, treineiros e registros incompletos, foi realizada uma análise exploratória comparando o desempenho e o perfil dos candidatos das redes pública e privada. Em seguida, foram ajustados dois modelos de regressão logística binária, sempre com a rede privada como categoria de referência: um comparando a rede estadual, com cerca de 2,77 milhões de registros, e outro comparando a rede federal, com cerca de 1,08 milhão. As variáveis analisadas foram renda familiar, escolaridade do pai e da mãe, cor/raça, sexo, acesso à internet e quantidade de computadores no domicílio. Todos os coeficientes foram estatisticamente significativos, com p-valor abaixo de 0,01, e todas as razões de chances ficaram abaixo de 1, confirmando a associação entre condições socioeconômicas mais favoráveis e a rede privada. O acesso à internet foi a variável que mais separou a rede estadual da privada ($OR \approx 0,46$), enquanto a cor/raça foi a que mais separou a federal da privada ($OR \approx 0,63$). Já a quantidade de computadores diferenciou bem a rede estadual da privada ($OR \approx 0,67$), mas praticamente não diferenciou a federal ($OR \approx 0,97$), indicando que o perfil socioeconômico do aluno da rede federal se aproxima do perfil do aluno da rede privada, em linha com o efeito de filtro social das instituições públicas seletivas apontado na literatura.

Palavras-chave: ENEM. desigualdade educacional. regressão logística. razão de chances. microdados.

Abstract

This work investigates the relationship between socioeconomic variables and the school type of ENEM participants from 2019 to 2023, using the public microdata provided by INEP. After preprocessing the five editions of the exam, removing absentees, non-graduating participants and incomplete records, an exploratory analysis was carried out comparing the performance and the profile of candidates from public and private schools. Then, two binary logistic regression models were fitted, always with the private network as the reference category: one comparing the state network, with about 2.77 million records, and another comparing the federal network, with about 1.08 million. The analyzed variables were family income, father's and mother's education, race, sex, internet access and the number of computers at home. All coefficients were statistically significant, with p-value below 0.01, and all odds ratios were below 1, confirming the association between more favorable socioeconomic conditions and the private network. Internet access was the variable that most separated the state network from the private one ($OR \approx 0.46$), while race was the one that most separated the federal network from the private one ($OR \approx 0.63$). The number of computers clearly differentiated the state network from the private one ($OR \approx 0.67$), but barely differentiated the federal network ($OR \approx 0.97$), indicating that the socioeconomic profile of federal school students is close to that of private school students, in line with the social filter effect of selective public institutions reported in the literature.

Key-words: ENEM. educational inequality. logistic regression. odds ratio. microdata.

Lista de ilustrações

Figura 1 – Média geral dos participantes por ano	27
Figura 2 – Média por matéria, por ano, separada por rede de ensino	28
Figura 3 – Treineiros em relação ao total de inscritos por ano	29
Figura 4 – Percentual de desistentes e ausentes por ano	29
Figura 5 – Distribuição de candidatos por dependência administrativa	30
Figura 6 – Distribuição da nota por classe de renda, pública <i>versus</i> privada	31
Figura 7 – Nota média por classe de renda, pública <i>versus</i> privada	31
Figura 8 – Distribuição da escolaridade dos pais, pública <i>versus</i> privada	32
Figura 9 – Distribuição acumulada da escolaridade dos pais, pública <i>versus</i> privada	33
Figura 10 – Distribuição das notas entre brancos e não-brancos, pública <i>versus</i> privada	33
Figura 11 – Percentual de alunos com acesso à internet em casa por ano, pública <i>versus</i> privada	34
Figura 12 – Acesso a recursos digitais, pública <i>versus</i> privada	35
Figura 13 – Distribuição da nota por dependência administrativa	35
Figura 14 – Média geral por dependência administrativa por ano	36
Figura 15 – Média por matéria nas três dependências administrativas	36
Figura 16 – Nota média por dependência administrativa e classe de renda	37
Figura 17 – Matriz de correlação com filtro $ \rho > 0,3$, pública <i>versus</i> privada	38
Figura 18 – Matriz de correlação com filtro $ \rho > 0,8$, pública <i>versus</i> privada	38
Figura 19 – Razão de chances por variável socioeconômica nas comparações Estadual <i>versus</i> Privada e Federal <i>versus</i> Privada	40

Lista de tabelas

Tabela 1	–	CrITÉrios de filtragem aplicados no pré-processamento	21
Tabela 2	–	VariÁveis selecionadas dos microdados e seus valores possÍveis	22
Tabela 3	–	Escala de renda familiar mensal do questionário socioeconômico	23
Tabela 4	–	Agrupamento das faixas de renda em classes	24
Tabela 5	–	VariÁveis utilizadas nos modelos de regressão logística	26
Tabela 6	–	Razão de chances e intervalos de confiança de 95% nos dois modelos . .	39

Lista de abreviaturas e siglas

ENEM	Exame Nacional do Ensino Médio
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
OR	<i>Odds Ratio</i> (Razão de Chances)
SISU	Sistema de Seleção Unificada
COMPUTADOR_NUM	Quantidade de computadores no domicílio
COR_RACA_BIN	Cor/raça do candidato (variável binária)
DEP_ADM_BIN	Rede de ensino do candidato (variável dependente binária)
ESC_MAE_NUM	Escolaridade da mãe ou mulher responsável
ESC_PAI_NUM	Escolaridade do pai ou homem responsável
INTERNET_BIN	Acesso à internet no domicílio (variável binária)
MEDIA_GERAL	Média aritmética das notas das cinco provas do ENEM
RENDA_NUM	Renda mensal familiar declarada pelo candidato
SEXO_BIN	Sexo do candidato (variável binária)
TP_COR_RACA	Cor/raça autodeclarada nos microdados do ENEM
TP_DEP_ADM	Dependência administrativa da escola nos microdados do ENEM
TP_ESCOLA	Tipo de escola do Ensino Médio cursada pelo candidato
TP_SEXO	Sexo do candidato nos microdados do ENEM

Lista de símbolos

α	Grau de significância estatística
β	Coefficiente da regressão logística
e	Base do logaritmo natural
H_0	Hipótese nula
OR	Odds Ratio (Razão de Chances)
p	Probabilidade da categoria de interesse
X_i	Variável preditora do modelo
Y	Variável dependente do modelo
ρ	Coefficiente de correlação de Pearson

Sumário

1	INTRODUÇÃO	13
1.1	O problema de pesquisa	13
1.2	Objetivos	14
1.3	Contribuições	14
1.4	Organização do trabalho	15
2	REVISÃO BIBLIOGRÁFICA	17
2.1	O ENEM como instrumento de acesso ao ensino superior	17
2.2	Desigualdade educacional e fatores socioeconômicos	17
2.3	Diferenciação institucional e o filtro seletivo na rede pública	18
2.4	Trabalhos relacionados e diferenciais desta abordagem	19
3	METODOLOGIA	20
3.1	Base de dados	20
3.2	Pré-processamento	21
3.2.1	Adaptação das variáveis entre edições	22
3.2.2	Codificação das variáveis	23
3.3	Regressão Logística e Razão de Chances	24
4	RESULTADOS	27
4.1	Análise exploratória	27
4.2	Análise de correlação entre as variáveis	37
4.3	Regressão Logística (<i>odds ratio</i>) entre as redes de ensino	39
4.3.1	Variáveis preditoras de renda familiar	40
4.3.2	Variáveis preditoras de escolaridade dos pais e responsáveis	40
4.3.3	Variáveis preditoras de acesso à internet e computadores	41
4.3.4	Demais análises: variáveis preditoras de raça e sexo	41
5	CONCLUSÃO	43
	REFERÊNCIAS	45

1 Introdução

O Exame Nacional do Ensino Médio (ENEM) é, desde sua criação em 1998, uma das principais ferramentas de avaliação da educação básica no Brasil e a principal porta de entrada para o ensino superior público no país. Ao longo dos anos, o exame acumulou um volume expressivo de dados sobre os candidatos, incluindo não apenas suas notas, mas também informações detalhadas sobre sua origem socioeconômica, trajetória escolar e condições de vida. Esse conjunto de informações torna os microdados do ENEM uma fonte privilegiada para investigar como fatores estruturais se relacionam com o desempenho educacional.

A desigualdade no desempenho escolar no Brasil é um fenômeno amplamente documentado. Estudantes de famílias com maior renda, com pais mais escolarizados e oriundos de escolas privadas consistentemente apresentam resultados superiores nas avaliações nacionais. [Senkevics, Carvalhaes e Ribeiro \(2022\)](#) demonstram, a partir de um painel de egressos do ensino médio de 2012, que a nota média no ENEM dos candidatos pertencentes ao quinto de renda mais rico supera em 112 pontos a dos pertencentes ao quinto mais pobre, e que 92,3% dos jovens do estrato mais abastado ingressam no ensino superior, contra apenas 48,1% dos mais pobres. No nível das provas, [Carvalho \(2022\)](#) confirmam esse padrão ao analisar os microdados do ENEM de 2017 a 2020: variáveis como renda familiar mensal e escolaridade dos pais figuram entre os fatores com maior influência sobre as notas de Matemática, independentemente do contexto pandêmico. Compreender a magnitude dessas diferenças e identificar quais variáveis mais se associam a elas é relevante tanto para o diagnóstico da qualidade educacional quanto para a formulação de políticas públicas mais efetivas.

Diante desse contexto, este trabalho analisa os microdados do ENEM referentes ao período de 2019 a 2023, com o objetivo de investigar a relação entre variáveis socioeconômicas e o desempenho dos candidatos. Em especial, busca-se compreender o papel da renda familiar, do tipo de escola frequentada e da escolaridade dos responsáveis como fatores associados ao desempenho ao longo da série histórica analisada.

1.1 O problema de pesquisa

Apesar da expansão do acesso ao ENEM e ao ensino superior nas últimas décadas, as desigualdades no desempenho entre diferentes grupos socioeconômicos permanecem expressivas. Candidatos de escolas públicas, de famílias de menor renda e cujos responsáveis têm menor escolaridade tendem a obter notas significativamente mais baixas do que seus pares em condições mais favoráveis. [Motta \(2025\)](#) quantificam essa disparidade nos

microdados do ENEM 2023: a mediana das notas de Matemática alcança 647 pontos entre alunos de escolas privadas, contra 493 pontos entre os de escolas públicas. O autor ajusta um modelo de regressão logística binária e reporta os resultados por meio da razão de chances (*odds ratio* – OR), confirmando que a associação entre o tipo de escola e o desempenho é estatisticamente significativa. Essa é a mesma técnica adotada neste trabalho, detalhada na Seção 3.3, com a diferença de que aqui as variáveis preditoras são socioeconômicas, e não as notas das provas. Senkevics, Carvalhaes e Ribeiro (2022) acrescentam que essa desigualdade opera por dois mecanismos articulados: (1) o desempenho escolar, fortemente condicionado pela origem social ao longo da trajetória educacional, e (2) o efeito direto da renda sobre as chances de ingresso, independente da nota obtida. Compreender a magnitude dessas diferenças e quais variáveis mais se associam a elas é relevante tanto para o diagnóstico da qualidade educacional quanto para a formulação de políticas públicas mais efetivas.

1.2 Objetivos

O objetivo geral deste trabalho é analisar a relação entre variáveis socioeconômicas e o desempenho no ENEM entre 2019 e 2023, investigando como fatores como renda familiar, escolaridade dos responsáveis, raça, sexo e acesso a recursos tecnológicos se associam ao tipo de rede escolar frequentada pelos candidatos.

Este trabalho possui os seguintes objetivos específicos:

- Realizar uma análise exploratória do desempenho dos candidatos por grupo socioeconômico, raça, tipo de escola e ano de realização do exame, por meio de visualizações geradas a partir dos microdados do INEP.
- Quantificar a associação entre variáveis socioeconômicas e o tipo de dependência administrativa da escola, comparando Estadual e Federal à rede Privada, por meio do cálculo das razões de chances (*odds ratio*), utilizando regressão logística binária.
- Identificar quais variáveis apresentam maior poder de diferenciação entre as redes públicas e a rede privada, com ênfase em renda familiar, escolaridade dos responsáveis, raça e acesso à internet.

1.3 Contribuições

A literatura brasileira sobre desigualdade educacional no ENEM é extensa, mas costuma tratar a rede pública como um bloco único e concentrar-se em uma única edição do exame. Diante desse cenário, este trabalho apresenta as seguintes contribuições:

- **Separação entre as redes estadual e federal.** Em vez de adotar a dicotomia usual entre escola pública e privada, o trabalho ajusta dois modelos independentes, comparando cada rede pública à rede privada. Essa decisão é sustentada teoricamente pelo efeito de filtro social das instituições seletivas descrito por [Feinmann e Rocha \(2024\)](#) e é confirmada empiricamente pelos resultados: variáveis como escolaridade da mãe e quantidade de computadores em casa separam nitidamente o aluno da rede estadual do aluno da rede privada, mas quase não separam o da rede federal ([seção 4.3](#)).
- **Janela temporal de cinco edições.** A análise cobre o período de 2019 a 2023, o que permite verificar a estabilidade dos padrões ao longo do tempo, incluindo os anos afetados pela pandemia. Os trabalhos correlatos mais próximos analisam uma única edição: [Motta \(2025\)](#) trabalha com o ENEM 2023 e [Campos, Moraes e Peres \(2025\)](#) com o ENEM 2022.
- **Quantificação conjunta das variáveis socioeconômicas.** O modelo estima o efeito de cada variável controlando pelas demais, e os resultados são reportados por razão de chances com intervalos de confiança de 95%. Isso permite ordenar as variáveis por poder de diferenciação entre as redes, algo que a análise de correlação, apresentada na [seção 4.2](#), não permitiu.
- **Registro do recorte racial.** A revisão realizada não identificou estudos que tratem diretamente do recorte racial no ENEM sob essa perspectiva. Os resultados mostram que a desigualdade racial persiste *dentro* de cada rede, e não apenas entre elas, o que fica registrado como ponto de partida para investigações futuras.

1.4 Organização do trabalho

Este trabalho está organizado em cinco capítulos. Neste primeiro capítulo foram apresentados o contexto do problema, os objetivos da pesquisa, as contribuições e a motivação que levou à escolha do tema. O [Capítulo 2](#) traz a revisão bibliográfica, que aborda o ENEM como a principal forma de acesso ao ensino superior para os estudantes, a relação entre a desigualdade educacional enfrentada por eles e os fatores socioeconômicos, os fundamentos da regressão logística e da razão de chances, além dos trabalhos relacionados que serviram de base para esta pesquisa. O [Capítulo 3](#) descreve a metodologia adotada, incluindo a base de dados utilizada, as etapas de pré-processamento dos dados extraídos e a especificação do modelo de regressão logística, com a definição da variável dependente e das variáveis preditoras. O [Capítulo 4](#) apresenta os resultados obtidos, iniciando pela análise exploratória dos dados, que retoma pontos discutidos na revisão bibliográfica, e seguindo para os resultados do modelo de regressão logística, com as razões de chances (OR) estimadas nas comparações entre as redes de ensino Federal x Privada e Estadual x Privada.

Por fim, o [Capítulo 5](#) retoma os objetivos do trabalho, discute as principais conclusões, aponta algumas limitações que a pesquisa sofreu e deixa sugestões para trabalhos futuros sobre esse tema amplo.

2 Revisão Bibliográfica

Este capítulo apresenta a fundamentação teórica em que o trabalho se referencia. Os conceitos abordados instruem a compreensão da análise realizada, bem como os trabalhos correlatos ao tema.

2.1 O ENEM como instrumento de acesso ao ensino superior

O Exame Nacional do Ensino Médio (ENEM) teve sua criação em 1998 com o objetivo principal de avaliar o desempenho dos estudantes ao final do ciclo de ensino ([Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2019](#)). Com o passar dos anos, o exame passou a desempenhar um papel crucial no acesso ao ensino superior, especialmente após a criação do Sistema de Seleção Unificada (SISU), que utiliza as notas do ENEM como critério de seleção para vagas em instituições públicas federais ([MOTTA, 2025](#)).

A prova se resume em cinco áreas do conhecimento: Linguagens, Códigos e suas Tecnologias; Matemática e suas Tecnologias; Ciências Humanas e suas Tecnologias; Ciências da Natureza e suas Tecnologias; e Redação, dentre as quais são distribuídas questões que o aluno ao final do ciclo de ensino médio deve ser capaz de responder. Além dessas questões, o participante responde a um questionário socioeconômico cujas respostas abordam as condições sociais em que está inserido, como renda familiar, escolaridade dos responsáveis, entre outros fatores. Esses dados são disponibilizados anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) ([Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2023](#)). Dado o número expressivo de participantes anuais, esses microdados tornaram-se uma fonte relevante para cruzar o desempenho dos estudantes com variáveis socioeconômicas detalhadas ([SENKEVICS; CARVALHAES; RIBEIRO, 2022](#)).

2.2 Desigualdade educacional e fatores socioeconômicos

A relação entre as condições socioeconômicas do participante e o desempenho escolar é um tema que vem ganhando força na literatura educacional brasileira ([CARVALHO, 2022](#)). Fatores como renda familiar, escolaridade dos pais e acesso a recursos tecnológicos durante o ciclo de educação estão associados ao rendimento dos estudantes em avaliações de larga escala ([SENKEVICS; CARVALHAES; RIBEIRO, 2022](#)).

No contexto do ENEM, estudos apontam que alunos que tiveram sua formação

em escolas privadas apresentam desempenho superior ao dos alunos de escolas públicas, e essa diferença está, em sua maior parte, associada ao perfil socioeconômico em que o aluno está inserido, e não apenas ao tipo de instituição frequentada (CAMPOS; MORAES; PERES, 2025; MOTTA, 2025). Essa constatação reforça a importância de se tratar a análise das variáveis socioeconômicas e o desempenho entre as redes de ensino de modo que seja possível separar o efeito da escola do efeito das condições sociais do aluno (CAMPOS; MORAES; PERES, 2025).

A escolaridade dos pais, em particular, é apontada como um dos preditores mais relevantes no desempenho escolar do aluno, uma vez que o capital cultural acumulado pela família ao longo dos anos influencia diretamente o ambiente de aprendizagem do participante do ENEM (LOURENÇO, 2016). De forma semelhante, o acesso à internet e a computadores em casa representa uma vantagem concreta no processo de preparação para exames como o ENEM (CARVALHO, 2022).

2.3 Diferenciação institucional e o filtro seletivo na rede pública

Quando o desempenho educacional no Brasil é analisado, agrupar todas as redes consideradas de “ensino público” pode ocultar questões sociais profundas, pois elas podem ser geridas pelo governo federal ou estadual. A literatura aponta que a rede de ensino médio é gerida em sua maioria pelos governos estaduais, cujas matrículas costumam ocorrer de acordo com o zoneamento geográfico, ou seja, encaminhando o aluno para a escola estadual mais próxima da sua residência (FEINMANN; ROCHA, 2024). Essa característica faz com que a rede estadual de ensino absorva grande parte dos alunos e reflita as desigualdades referentes a onde o aluno reside.

Por outro lado, instituições públicas do governo federal que fornecem o ensino médio e até mesmo a opção integrada ao ensino técnico operam sob um sistema de admissão baseado em notas, estabelecendo, assim, uma lógica excludente. Feinmann e Rocha (2024) demonstram, no contexto do ensino superior brasileiro, que realizar processos seletivos estritamente baseados em desempenho acaba favorecendo os alunos de origens socioeconômicas mais privilegiadas: nas instituições públicas de elite, apenas 2,5% dos alunos vêm dos 20% mais pobres, um nível de segregação de renda equivalente ao encontrado nas instituições privadas de elite.

Esse mesmo cenário de triagem social, fortemente influenciado pelo capital cultural e pela herança familiar do estudante, ocorre antes mesmo de o estudante chegar ao ensino superior (LOURENÇO, 2016). No Brasil, as redes federais de ensino médio, como os Institutos Federais, Centros Federais e demais instituições do mesmo perfil, comumente adotam processos seletivos baseados em desempenho como forma de ingresso. Diante disso, a metodologia deste trabalho optou por separar a rede estadual da rede federal

nas comparações feitas com a rede privada de ensino. Essa divisão justifica-se para evitar o efeito de filtro social, em que as escolas federais poderiam mascarar o desempenho e a realidade do aluno da rede estadual de ensino, evidenciando assim os reais fatores de desigualdade no sistema público. Essa comparação entre as três redes é realizada na [seção 4.1](#) e na [seção 4.3](#).

2.4 Trabalhos relacionados e diferenciais desta abordagem

Os trabalhos discutidos nas seções anteriores abordam a desigualdade educacional a partir de bases, métodos e variáveis-resposta distintos. [Senkevics, Carvalhaes e Ribeiro \(2022\)](#) e [Feinmann e Rocha \(2024\)](#) acompanham concluintes do ensino médio ao longo de vários anos e tomam o ingresso no ensino superior como desfecho. [Campos, Moraes e Peres \(2025\)](#) e [Carvalho \(2022\)](#) trabalham com edições específicas do ENEM e analisam a nota, o primeiro por meio de análise descritiva e o segundo por regressão linear múltipla. [Lourengo \(2016\)](#) compara o perfil dos aprovados na UnB antes e depois da adoção do SiSU. Por fim, [Motta \(2025\)](#) é o mais próximo desta pesquisa, ao aplicar regressão logística binária sobre os microdados do ENEM 2023 para comparar escola pública e privada.

Diante desse panorama, três diferenciais podem ser destacados. O primeiro é a **variável resposta**: enquanto [Campos, Moraes e Peres \(2025\)](#) e [Carvalho \(2022\)](#) tomam a nota como desfecho, e [Senkevics, Carvalhaes e Ribeiro \(2022\)](#) e [Feinmann e Rocha \(2024\)](#) tomam o ingresso no ensino superior, este trabalho toma o *tipo de escola* como desfecho. A escolha desloca a análise para um estágio anterior da trajetória educacional, em que a desigualdade já se manifesta antes da realização da prova.

O segundo é o **grão da comparação**. [Motta \(2025\)](#) adota a mesma técnica empregada aqui, mas trata a rede pública como categoria única. Este trabalho ajusta dois modelos independentes, separando a rede estadual da federal, o que permite mostrar que as duas apresentam perfis socioeconômicos bastante distintos sob o mesmo rótulo.

O terceiro é a **amplitude temporal**. Os trabalhos que utilizam microdados do ENEM concentram-se em uma ou poucas edições, ao passo que esta pesquisa cobre cinco edições consecutivas, o que permite avaliar a estabilidade das associações encontradas ao longo do período, inclusive nos anos afetados pela pandemia.

3 Metodologia

Os dados utilizados neste trabalho são os microdados do ENEM disponibilizados publicamente pelo INEP, abrangendo as edições de 2019 a 2023. Cada edição contém registros individuais de todos os candidatos inscritos, incluindo as notas obtidas nas cinco áreas do exame: Ciências da Natureza, Ciências Humanas, Linguagens e Códigos, Matemática e Redação, além das respostas ao questionário socioeconômico, que traz informações sobre escolaridade e ocupação dos responsáveis, renda familiar, tipo de escola cursada no ensino médio e acesso a bens e serviços no domicílio, entre outras.

Por se tratar de arquivos com volume entre três e seis milhões de registros por edição, o processamento foi realizado em ambiente de nuvem, com os dados armazenados no Google Drive e acessados diretamente pelo Google Colab durante a execução dos *scripts*.

Os passos para execução deste trabalho são assim definidos:

- Revisão da literatura sobre desigualdade educacional e fatores socioeconômicos associados ao desempenho no ENEM.
- Pré-processamento dos microdados do INEP, com filtragem de registros inválidos e organização das variáveis de interesse.
- Análise exploratória dos dados, com geração de visualizações para identificação de padrões de desempenho por grupo socioeconômico, raça, tipo de escola e período analisado.
- Cálculo das razões de chances (*odds ratio*) para quantificar a associação entre as variáveis socioeconômicas e o tipo de rede escolar frequentada pelos candidatos.
- Análise e discussão dos resultados obtidos.

3.1 Base de dados

Os microdados do ENEM são disponibilizados anualmente pelo INEP e constituem uma das principais fontes de informação sobre o desempenho educacional no Brasil. Para este trabalho, foram utilizadas as edições de 2019 a 2023, período que permite observar variações no desempenho ao longo do tempo e identificar tendências na relação entre fatores socioeconômicos e resultado no exame.

3.2 Pré-processamento

Antes das análises, os dados passaram por uma etapa de preparação com o objetivo de garantir a qualidade e a consistência dos registros utilizados. Essa etapa foi inteiramente implementada em Python, com uso das bibliotecas *pandas* e *numpy*. A execução do código desta etapa gera arquivos de formato *csv* que são reutilizados nas etapas seguintes, de forma que o pré-processamento só precisa ser executado uma única vez.

Dado o tamanho dos arquivos, não foi possível carregá-los integralmente na memória de uma só vez. Para contornar isso, a leitura foi feita em partes: o arquivo de cada ano foi processado em blocos de 400 mil linhas, aplicando os filtros de limpeza em cada bloco antes de seguir para o próximo. Em cada bloco lido, os critérios de filtragem sintetizados na [Tabela 1](#) foram aplicados em sequência.

Tabela 1 – Critérios de filtragem aplicados no pré-processamento

Critério	Registros removidos	Justificativa
Ausentes	Candidatos que não compareceram a todos os dias do exame	As notas desses candidatos estão faltantes ou incompletas, o que inviabiliza a análise do desempenho
Treineiros	Candidatos que não concluem o ensino médio no ano vigente ($IN_TREINEIRO = 1$)	O trabalho analisa o vínculo do candidato com a rede em que cursa o ensino médio, o que não se aplica a esse grupo
Nota zero na Redação	Candidatos com $NU_NOTA_REDACAO = 0$	A pontuação costuma indicar redação em branco, fuga ao tema ou texto anulado, situações que não refletem desempenho baixo no sentido usual
Escopo de rede	Candidatos fora das redes pública e privada do Brasil ($TP_ESCOLA =$ não respondeu ou exterior)	Categorias fora do escopo desta pesquisa; a variável é detalhada na Seção 3.2.1

Fonte: elaborado pelo autor.

Das 76 variáveis presentes nos microdados, foram selecionadas aquelas relevantes aos objetivos do trabalho, apresentadas na [Tabela 2](#) com os valores que cada uma pode assumir. No grupo de desempenho, foram utilizadas as notas de cada área do exame e calculada a média das cinco provas (Ciências da Natureza, Ciências Humanas, Linguagens e Códigos, Matemática e Redação). No grupo socioeconômico, foram consideradas a dependência administrativa da escola, cor/raça, sexo, escolaridade dos responsáveis, renda familiar e situação de conclusão do ensino médio. Para facilitar a leitura dos gráficos e tabelas, os códigos numéricos das variáveis categóricas foram substituídos por seus respectivos rótulos descritivos, conforme o dicionário de variáveis do INEP ([Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2019](#); [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2020](#); [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2021](#); [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio](#)

Teixeira, 2022; Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2023).

Tabela 2 – Variáveis selecionadas dos microdados e seus valores possíveis

Variável	Descrição	Valores possíveis
<i>Desempenho</i>		
NU_NOTA_*	Nota de cada prova	Numérica, de 0 a 1000
MEDIA_GERAL	Média das cinco provas	Numérica, calculada pelo autor
<i>Socioeconômicas</i>		
TP_DEP_ADM	Dependência administrativa	Federal; Estadual; Municipal; Privada
TP_ESCOLA	Tipo de escola	Não respondeu; Pública; Privada; Exterior (suprimida a partir de 2021)
TP_ST_CONCLUSAO	Situação de conclusão	Já concluiu; concluirá no ano corrente; concluirá após o ano corrente; não concluiu e não está cursando
IN_TREINEIRO	Candidato treineiro	0 = não; 1 = sim
TP_COR_RACA	Cor/raça autodeclarada	Não declarado; Branca; Preta; Parda; Amarela; Indígena
TP_SEXO	Sexo	Masculino; Feminino
Q001 e Q002	Escolaridade dos responsáveis	De “nunca estudou” a “pós-graduação”, mais “não sei”
Q006	Renda familiar mensal	17 faixas ordinais, de A a Q (Tabela 3)
Q024	Computadores em casa	Nenhum; 1; 2; 3; 4 ou mais
Q025	Acesso à internet	Não; Sim

Fonte: elaborado pelo autor.

Vale observar que a categoria *Municipal* da variável TP_DEP_ADM não é utilizada nas comparações deste trabalho, uma vez que a oferta municipal de ensino médio é residual no país. As análises que separam as três redes consideram apenas Federal, Estadual e Privada.

3.2.1 Adaptação das variáveis entre edições

Uma adaptação necessária durante o pré-processamento foi garantir a comparabilidade da variável de renda familiar (‘Q006’) entre as cinco edições analisadas. O INEP reajusta anualmente os limites monetários das faixas de renda declaradas no questionário socioeconômico, acompanhando variações no salário mínimo vigente. Embora a estrutura da variável se mantenha estável, com 17 categorias ordinais identificadas pelas letras A a Q, da menor para a maior renda, os valores em reais correspondentes a cada faixa diferem entre os anos. Por exemplo, a categoria B correspondia a “Até R\$ 998,00” em 2019 e passou a “Até R\$ 1.320,00” em 2023 (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2019; Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira,

2023). A [Tabela 3](#) apresenta a escala completa da questão, com os valores utilizados na edição de 2023.

Tabela 3 – Escala de renda familiar mensal do questionário socioeconômico (‘Q006’).

Código	Renda mensal familiar
A	Nenhuma renda
B	Até R\$ 1.045,00
C	De R\$ 1.045,01 até R\$ 1.567,50
D	De R\$ 1.567,51 até R\$ 2.090,00
E	De R\$ 2.090,01 até R\$ 2.612,50
F	De R\$ 2.612,51 até R\$ 3.135,00
G	De R\$ 3.135,01 até R\$ 4.180,00
H	De R\$ 4.180,01 até R\$ 5.225,00
I	De R\$ 5.225,01 até R\$ 6.270,00
J	De R\$ 6.270,01 até R\$ 7.315,00
K	De R\$ 7.315,01 até R\$ 8.360,00
L	De R\$ 8.360,01 até R\$ 9.405,00
M	De R\$ 9.405,01 até R\$ 10.450,00
N	De R\$ 10.450,01 até R\$ 12.540,00
O	De R\$ 12.540,01 até R\$ 15.675,00
P	De R\$ 15.675,01 até R\$ 20.900,00
Q	Acima de R\$ 20.900,00

Fonte: [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira \(2023\)](#)

Para contornar esse problema sem perder a ordenação relativa das faixas, a variável foi convertida para uma escala ordinal numérica de 0 a 16, em que o valor 0 representa ausência de renda e o valor 16 representa a faixa de maior renda declarada. Esse procedimento preserva a hierarquia entre os grupos e permite que candidatos de diferentes edições sejam comparados de forma consistente, sem depender dos limites monetários absolutos de cada ano.

Verificou-se também que a variável TP_ESCOLA apresentava a categoria 4 = Exterior nas edições de 2019 e 2020, sendo esta suprimida a partir de 2021 ([Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2019](#); [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2020](#); [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2021](#)). Como o escopo deste trabalho se restringe a candidatos de escolas públicas ou privadas do Brasil, registros com essa classificação foram excluídos em todas as edições em que apareceram.

3.2.2 Codificação das variáveis

Além da conversão da renda para a escala numérica, as demais variáveis do questionário também passaram por conversões antes de entrar nas análises.

Para as análises exploratórias, as 17 faixas de renda foram agrupadas em quatro classes, de acordo com a [Tabela 4](#). Esse agrupamento foi feito porque comparar as 17 faixas de uma vez só deixaria os gráficos muito carregados, dificultando a leitura. Já no

modelo de regressão logística foi utilizada a escala numérica completa (RENDA_NUM, de 0 a 16), apresentada anteriormente.

Tabela 4 – Agrupamento das faixas de renda do questionário ‘Q006’ em classes.

Código Q006	Classe de renda
A, B, C	Baixa
D, E, F	Média-Bx
G, H, I, J	Média-Alt
K, L, M, N, O, P, Q	Alta

Fonte: elaborado pelo autor.

A escolaridade dos responsáveis é perguntada nas questões ‘Q001’, para o pai ou homem responsável, e ‘Q002’, para a mãe ou mulher responsável. As respostas vão desde nunca ter estudado até possuir pós-graduação, e foram convertidas para uma escala de 0 a 6, seguindo a ordem dos níveis de ensino (ESC_PAI_NUM e ESC_MAE_NUM).

Na questão de cor/raça, o participante se autodeclara entre as opções Branca, Preta, Parda, Amarela e Indígena, podendo também não declarar. Para o modelo, essas categorias foram unificadas em dois grupos: branco e não-branco (COR_RACA_BIN: 1 = branco, 0 = não-branco). Essa unificação simplifica a interpretação da razão de chances, que passa a comparar diretamente os dois grupos, e concentra a análise no contraste entre o grupo branco e os demais, que é o ponto de interesse deste trabalho.

As demais variáveis exigiram conversões mais simples. O sexo do participante já é registrado de forma binária nos microdados e foi apenas recodificado (SEXO_BIN: 0 = masculino, 1 = feminino). A questão ‘Q025’ pergunta se existe acesso à internet na residência do aluno (INTERNET_BIN: 0 = não tem, 1 = tem). Por fim, a questão ‘Q024’ pergunta a quantidade de computadores em casa, com respostas que vão de nenhum até quatro ou mais, convertidas para a escala de 0 a 4 (COMPUTADOR_NUM).

O resumo de todas as variáveis que entraram nos modelos, com suas origens e codificações, é apresentado na [Tabela 5](#), na [seção 3.3](#).

3.3 Regressão Logística e Razão de Chances

Foram realizados dois comparativos independentes, sempre com a rede Privada como referência: Estadual versus Privada e Federal versus Privada, seguindo abordagem similar à adotada por [Motta \(2025\)](#). Para cada comparativo, define-se uma rede pública como categoria de interesse (valor 1) e a rede Privada como categoria de referência (valor 0).

Para quantificar a associação entre as variáveis socioeconômicas e o tipo de rede escolar frequentada pelos candidatos, este trabalho emprega a regressão logística binária e o cálculo das razões de chances (*odds ratio* – OR). Essa abordagem é adequada quando

a variável dependente assume apenas dois valores possíveis (HOSMER; LEMESHOW; STURDIVANT, 2013; AGRESTI, 2002).

Em um dos casos do problema deste trabalho, explorado na seção 4.3, consideram-se as seguintes probabilidades envolvendo a variável dependente Y :

$$\begin{aligned} p &= \text{Prob}(Y = 1) = \text{Prob}(\text{rede} = \text{Estadual}) \\ 1 - p &= \text{Prob}(Y = 0) = \text{Prob}(\text{rede} = \text{Privada}), \end{aligned}$$

em que “rede” representa a dependência administrativa da escola do candidato, de forma que

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log\left(\frac{\text{Prob}(\text{rede} = \text{Estadual})}{\text{Prob}(\text{rede} = \text{Privada})}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n. \quad (3.1)$$

Encontrar os coeficientes β_i do “logit” de p , em que logit é o inverso da função logística, é o que se chama de regressão logística (MOTTA, 2025). Cada coeficiente β_i representa o efeito da variável X_i sobre o $\log$ das chances de pertencer à categoria de interesse, e a hipótese nula associada a cada β_i é de que a variável preditora X_i não tem efeito sobre a variável dependente, ou seja, $H_0 : \beta_i = 0$. O OR é obtido pela exponenciação desse coeficiente (AGRESTI, 2002):

$$OR = e^{\beta_i} \quad (3.2)$$

Um OR acima de 1 indica que o aumento naquela variável está associado a maior chance de pertencer à rede comparada em relação à Privada; abaixo de 1, indica que o aumento naquela variável está associado a menor chance de pertencer à rede comparada em relação à Privada, que é a referência; igual a 1, aponta ausência de associação (HOSMER; LEMESHOW; STURDIVANT, 2013; FIELD, 2018). Essa característica torna o OR uma medida útil para comparar grupos com perfis socioeconômicos distintos, como os analisados por Senkevics, Carvalhaes e Ribeiro (2022) ao investigar as desigualdades de acesso ao ensino superior no Brasil.

Para ilustrar essa leitura, Motta (2025) estima, em um modelo com a mesma estrutura adotada aqui, um OR de 0,995 para a nota de Ciências da Natureza: cada ponto a mais nessa prova reduz em cerca de 0,5% as chances de o aluno ser de escola pública em relação à privada. Neste trabalho a interpretação é a mesma, trocando as notas pelas variáveis socioeconômicas do questionário.

Para variáveis ordinais, como renda familiar (RENDA_NUM, de 0 a 16) e escolaridade dos pais (ESC_PAI_NUM e ESC_MAE_NUM, de 0 a 6), o OR representa a variação nas chances a cada incremento de uma unidade na escala (HOSMER; LEMESHOW; STURDIVANT, 2013). Para comparar níveis mais distantes, eleva-se o OR ao número de degraus de diferença

entre eles. Para variáveis binárias, como raça (COR_RACA_BIN: 0 = não-branco, 1 = branco), sexo (SEXO_BIN: 0 = masculino, 1 = feminino) e acesso à internet (INTERNET_BIN: 0 = não tem, 1 = tem), o OR compara diretamente os dois grupos (AGRESTI, 2002; HOSMER; LEMESHOW; STURDIVANT, 2013).

As variáveis preditoras incluídas foram renda familiar, escolaridade do pai ou homem responsável, escolaridade da mãe ou mulher responsável, raça, sexo, acesso à internet e quantidade de computadores em casa. A Tabela 5 resume a variável dependente e as variáveis preditoras utilizadas nos modelos, com a questão de origem no questionário do ENEM e a codificação adotada. Na variável dependente, o valor 1 corresponde à rede pública de cada comparativo, Estadual ou Federal, conforme o modelo. Todos os modelos foram ajustados com a biblioteca *statsmodels* em Python, e os resultados são apresentados com o p-valor de cada coeficiente.

Com o objetivo de permitir a verificação e a reprodução dos procedimentos descritos neste capítulo, o código utilizado em todas as etapas da análise está disponível publicamente em <<https://github.com/carviit/tcc-enem-socioeconomico>>.

Tabela 5 – Variáveis utilizadas nos modelos de regressão logística.

Variável	Origem (ENEM)	Descrição	Codificação
<i>Variável dependente</i>			
DEP_ADM_BIN	TP_DEP_ADM	Rede de ensino	1 = Estadual ou Federal 0 = Privada
<i>Variáveis preditoras</i>			
RENDA_NUM	Q006	Renda familiar mensal	Ordinal, 0 a 16
ESC_PAI_NUM	Q001	Escolaridade do pai ou homem responsável	Ordinal, 0 a 6
ESC_MAE_NUM	Q002	Escolaridade da mãe ou mulher responsável	Ordinal, 0 a 6
COR_RACA_BIN	TP_COR_RACA	Raça/cor autodeclarada	0 = não-branco 1 = branco
SEXO_BIN	TP_SEXO	Sexo	0 = masculino 1 = feminino
INTERNET_BIN	Q025	Acesso à internet no domicílio	0 = não tem 1 = tem
COMPUTADOR_NUM	Q024	Quantidade de computadores em casa	Ordinal, 0 a 4

Fonte: elaborado pelo autor.

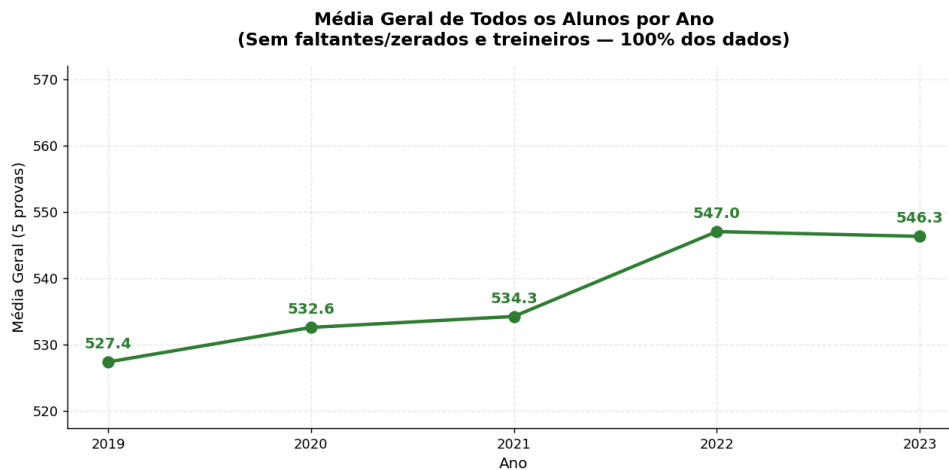
4 Resultados

Este capítulo apresenta os resultados obtidos na pesquisa. A primeira parte traz a análise exploratória dos dados, retomando pontos discutidos no [Capítulo 2](#), seguida da análise de correlação entre as variáveis. A última parte traz os resultados do modelo de regressão logística, com as razões de chances estimadas nas comparações entre as redes de ensino.

4.1 Análise exploratória

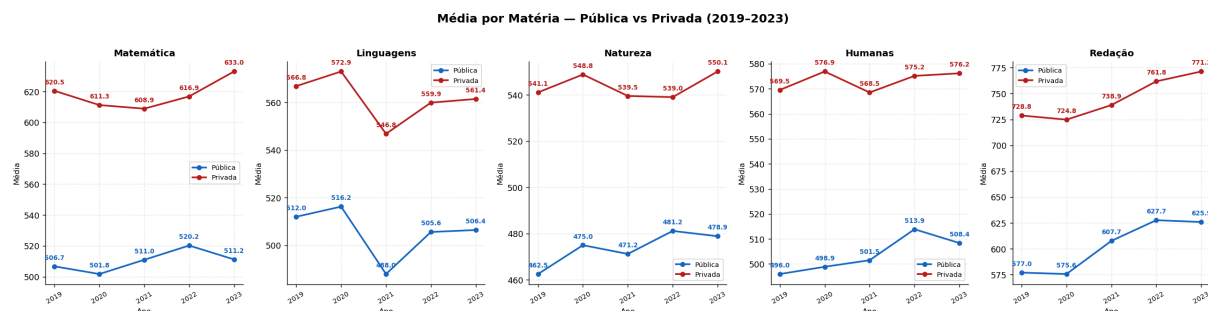
Inicialmente foi realizada uma análise exploratória com os dados de 2019 a 2023, excluindo os alunos que não realizaram a prova em algum dos dois dias e os alunos treineiros, que são alunos que não terminarão o ensino médio no ano vigente. A [Figura 1](#) apresenta a média geral dos participantes em cada ano, já a [Figura 2](#) apresenta a média por matéria em cada ano, separando a rede pública de ensino da rede privada.

Figura 1 – Média geral dos participantes por ano



Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Figura 2 – Média por matéria, por ano, separada por rede de ensino

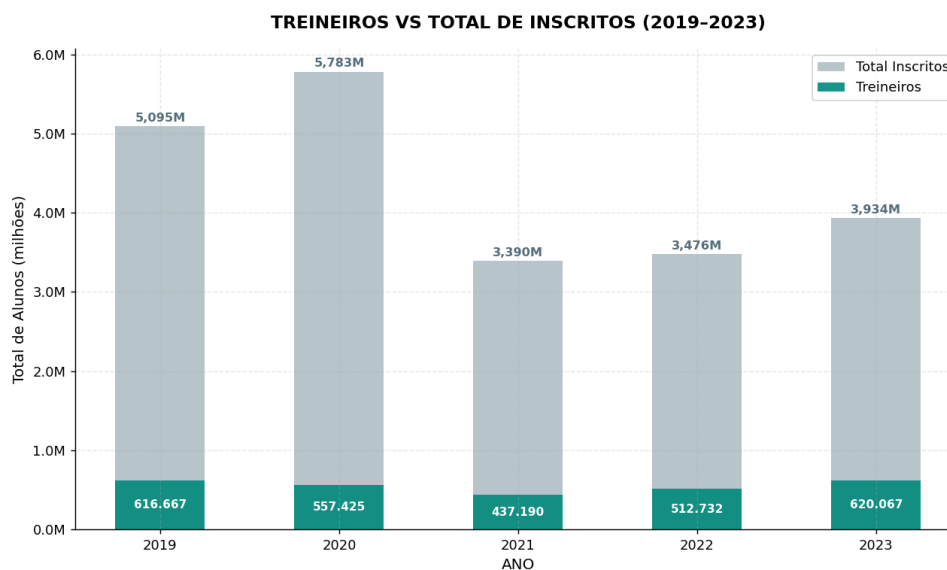


Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

A diferença de desempenho entre as redes, discutida no [Capítulo 2](#) a partir dos trabalhos de [Campos, Moraes e Peres \(2025\)](#) e [Motta \(2025\)](#), aparece em todas as matérias e em todos os anos do período. Em Matemática, porém, a distância entre as redes cresceu de 113,8 pontos em 2019 para 121,8 pontos em 2023, enquanto nas demais matérias ela ficou estável, como em Linguagens, ou diminuiu, como em Natureza, Humanas e Redação. Este trabalho não aprofunda as causas desse movimento, que fica como sugestão para análises futuras.

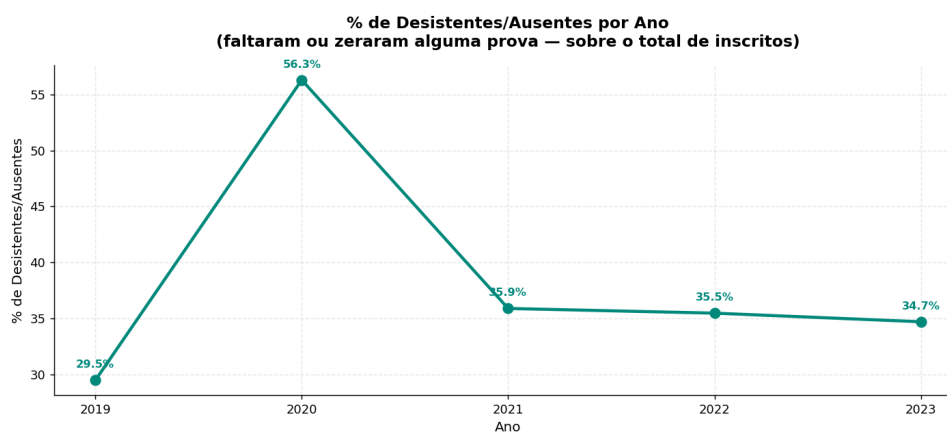
Quanto ao tamanho dos recortes aplicados, os treineiros representam algo entre 10% e 16% dos inscritos em cada edição ([Figura 3](#)), e a parcela de participantes que faltou ou zerou alguma prova ficou acima de um terço na maior parte do período, com pico de 56,3% em 2020 ([Figura 4](#)).

Figura 3 – Treineiros em relação ao total de inscritos por ano



Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

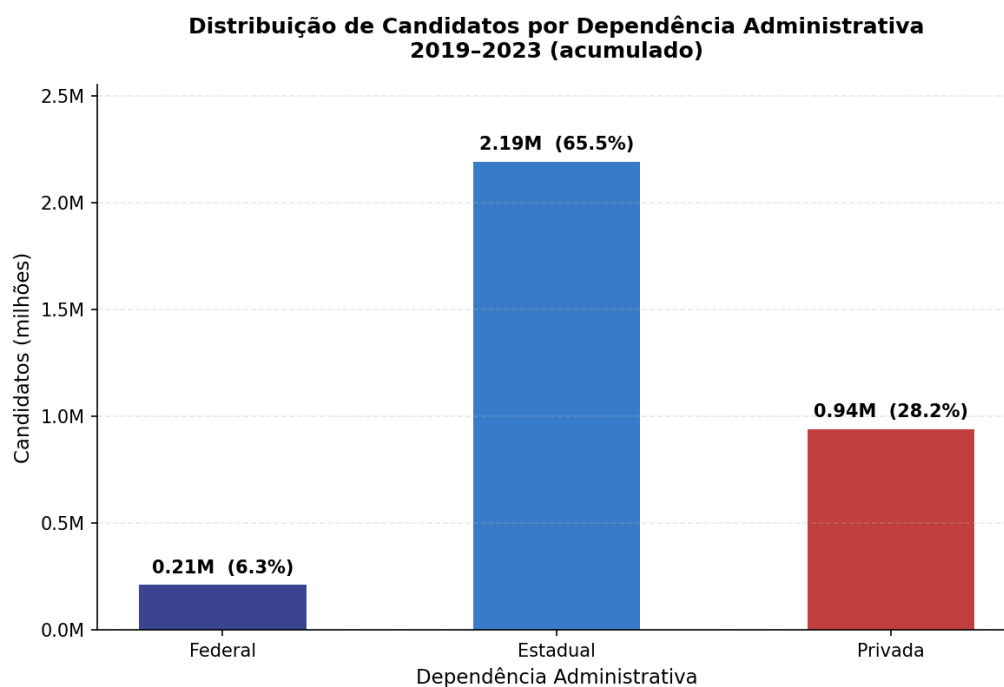
Figura 4 – Percentual de desistentes e ausentes por ano



Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

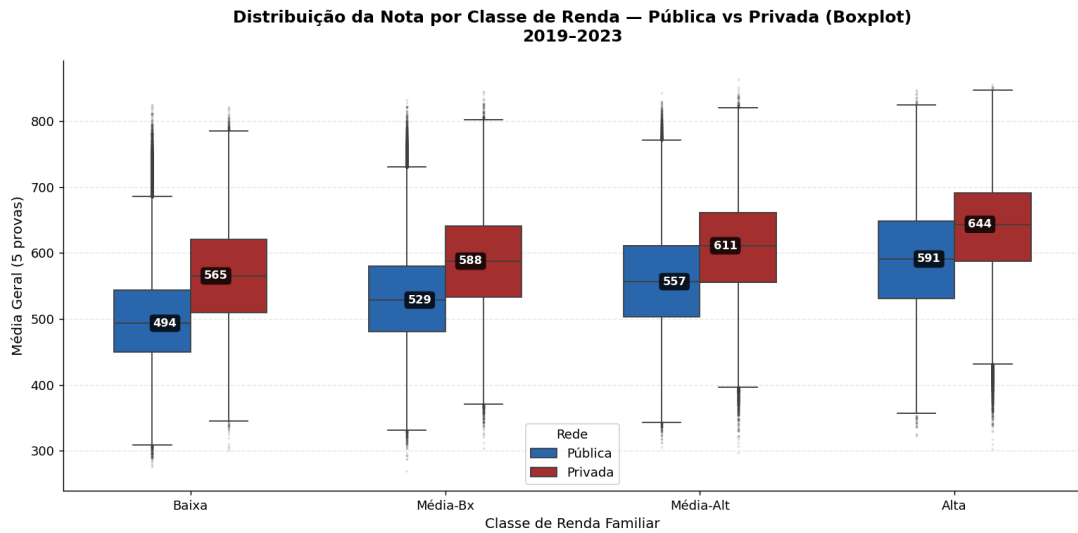
Depois dos filtros, a base acumulada do período ficou com cerca de 3,34 milhões de participantes com escola identificada, dos quais 65,5% vêm da rede estadual, 28,2% da privada e apenas 6,3% da federal (Figura 5). Essa concentração na rede estadual vai ao encontro do que [Feinmann e Rocha \(2024\)](#) apontam sobre ela ser a principal responsável pelas matrículas do ensino médio no país.

Figura 5 – Distribuição de candidatos por dependência administrativa



Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

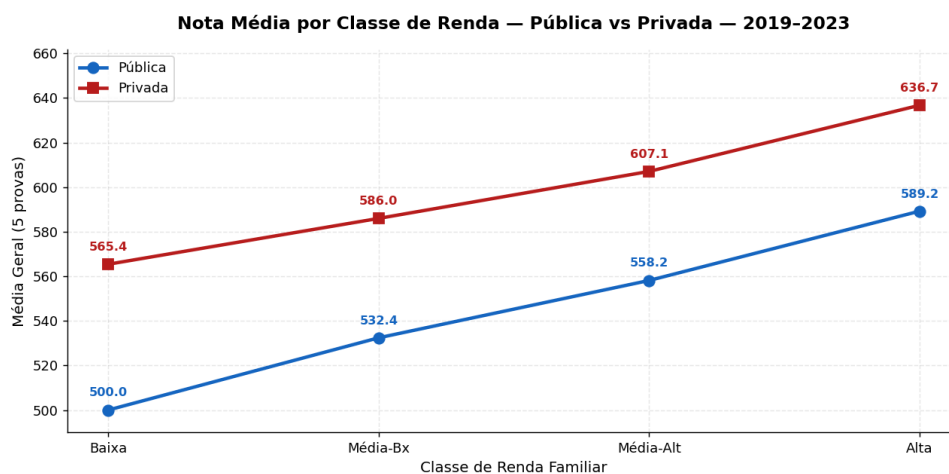
Em seguida, as notas foram analisadas em relação à renda familiar declarada pelos alunos, utilizando as classes de renda definidas na [Tabela 4](#). A [Figura 6](#) apresenta a distribuição de notas dos alunos separada por classe de renda e por rede de ensino, já a [Figura 7](#) apresenta a média das notas entre esses alunos. Essa análise retoma o que foi visto no [Capítulo 2](#), em que a renda familiar aparece como um dos fatores mais associados ao rendimento em avaliações de larga escala ([SENKEVICS; CARVALHAES; RIBEIRO, 2022](#); [CARVALHO, 2022](#)).

Figura 6 – Distribuição da nota por classe de renda, pública *versus* privada

Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Na análise da [Figura 6](#), podemos observar uma quantidade elevada de outliers na rede pública de ensino, mostrando que, mesmo entre os alunos de renda mais baixa, existe uma parcela que atinge notas bem acima da mediana do seu grupo, aproximando-se das notas mais altas registradas na rede privada.

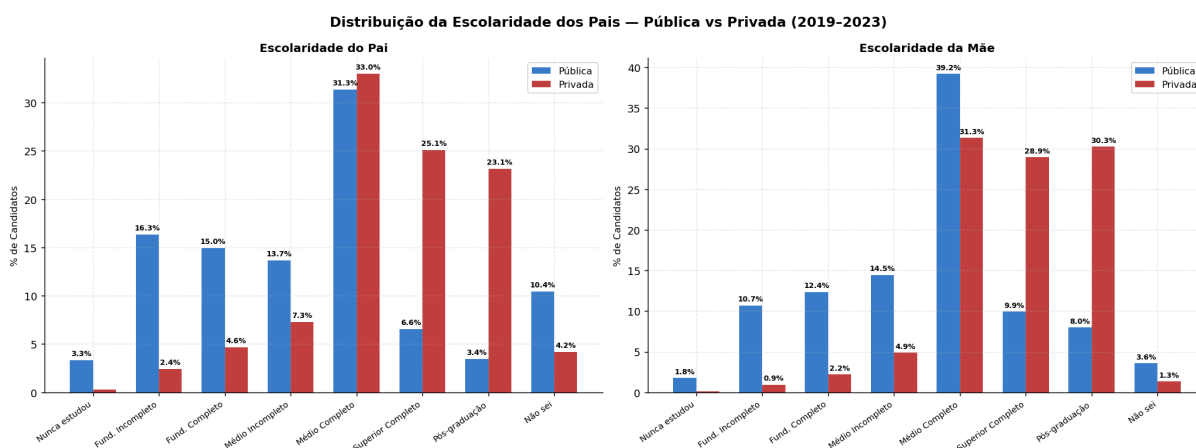
A [Figura 7](#) apresenta a nota média entre esses alunos, reforçando o padrão de crescimento da nota conforme a classe de renda aumenta, tanto na rede pública quanto na rede privada.

Figura 7 – Nota média por classe de renda, pública *versus* privada

Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

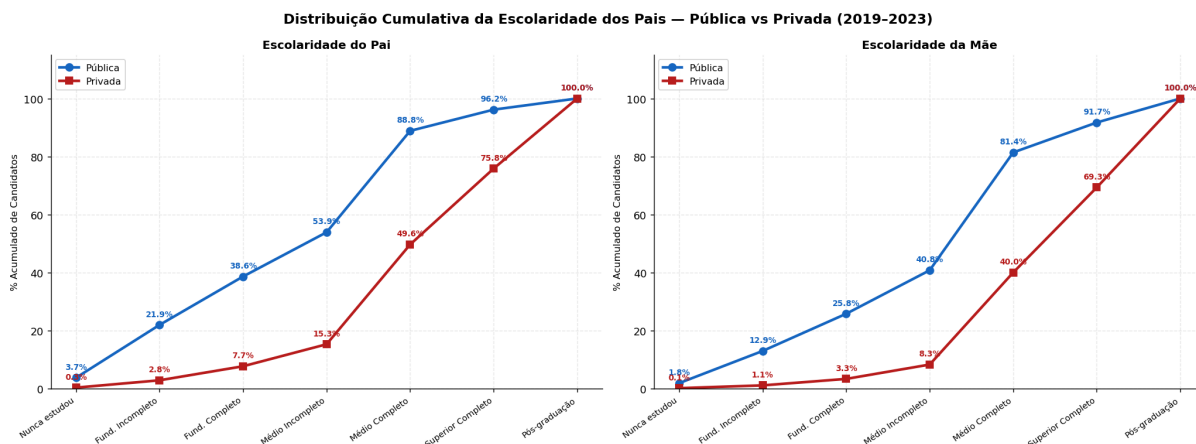
Outro fator analisado foi a escolaridade dos pais. Como discutido no [Capítulo 2](#), a escolaridade dos responsáveis é apontada como um dos preditores mais relevantes do desempenho escolar, já que o capital cultural acumulado pela família influencia diretamente o ambiente de aprendizagem do aluno (LOURENÇO, 2016). Nos dados, a diferença de perfil entre as redes é grande: na privada, quase metade dos pais e mais da metade das mães possuem ensino superior completo ou pós-graduação, enquanto na pública esses grupos ficam em torno de 10% e 18%, respectivamente ([Figura 8](#)). A categoria “Não sei” também é bem mais frequente entre os alunos da rede pública, principalmente em relação ao pai.

Figura 8 – Distribuição da escolaridade dos pais, pública *versus* privada



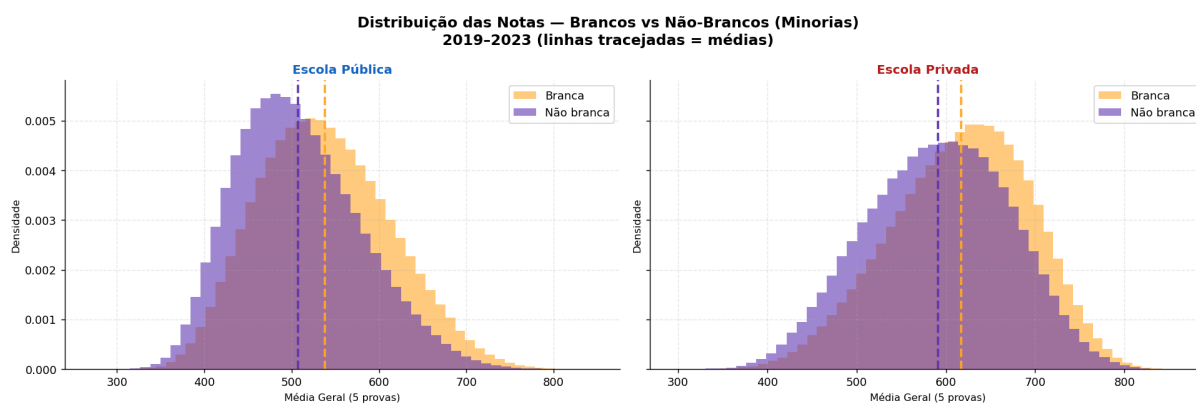
Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Acumulando esses percentuais, o contraste fica ainda mais visível: cerca de 54% dos pais dos alunos da rede pública não chegaram a concluir o ensino médio, contra aproximadamente 15% na rede privada. No caso das mães, os percentuais são respectivamente de 41% e 8% ([Figura 9](#)).

Figura 9 – Distribuição acumulada da escolaridade dos pais, pública *versus* privada

Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

O recorte por raça e cor segue a mesma direção. Agrupando os candidatos em brancos e não-brancos, mesma codificação utilizada no modelo de regressão (subseção 3.2.2), as distribuições de notas se sobrepõem em boa parte, mas a curva dos alunos não-brancos aparece deslocada para a esquerda nas duas redes, com diferença de cerca de 30 pontos nas médias da rede pública e de 25 pontos na privada (Figura 10). Ou seja, foi observado que a desigualdade racial não aparece apenas na comparação entre redes, ela se mantém dentro de cada uma. A revisão bibliográfica deste trabalho não trouxe estudos que tratem diretamente desse recorte no ENEM, então essas observações ficam registradas como ponto de partida para investigações futuras.

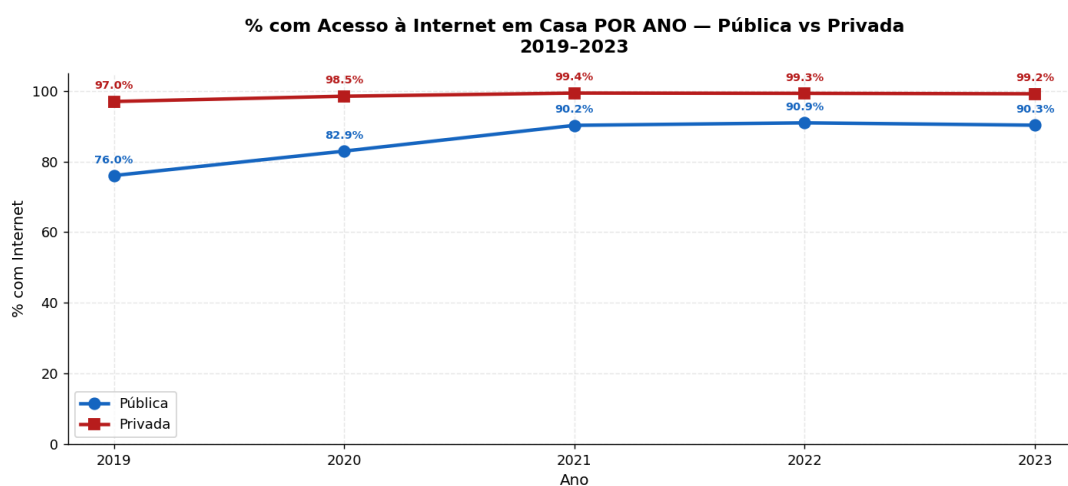
Figura 10 – Distribuição das notas entre brancos e não-brancos, pública *versus* privada

Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Sobre o acesso a recursos digitais, o Capítulo 2 destacou que a internet e o

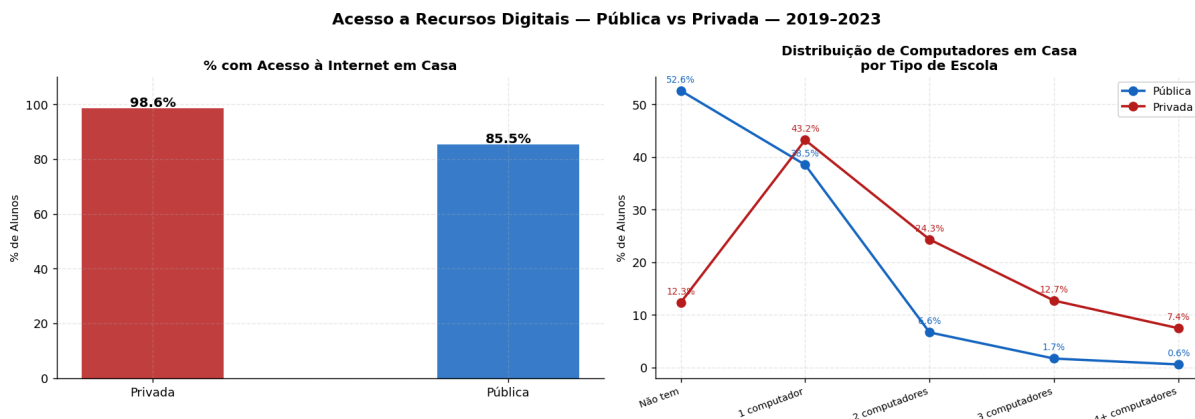
computador em casa representam uma vantagem concreta na preparação para exames como o ENEM (CARVALHO, 2022). Nos dados, o acesso à internet é praticamente universal na rede privada durante todo o período, acima de 97%, enquanto na rede pública ele sai de 76,0% em 2019 e chega a 90,3% em 2023, reduzindo a distância entre as redes de 21 para cerca de 9 pontos percentuais, mas sem eliminá-la (Figura 11).

Figura 11 – Percentual de alunos com acesso à internet em casa por ano, pública *versus* privada



Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

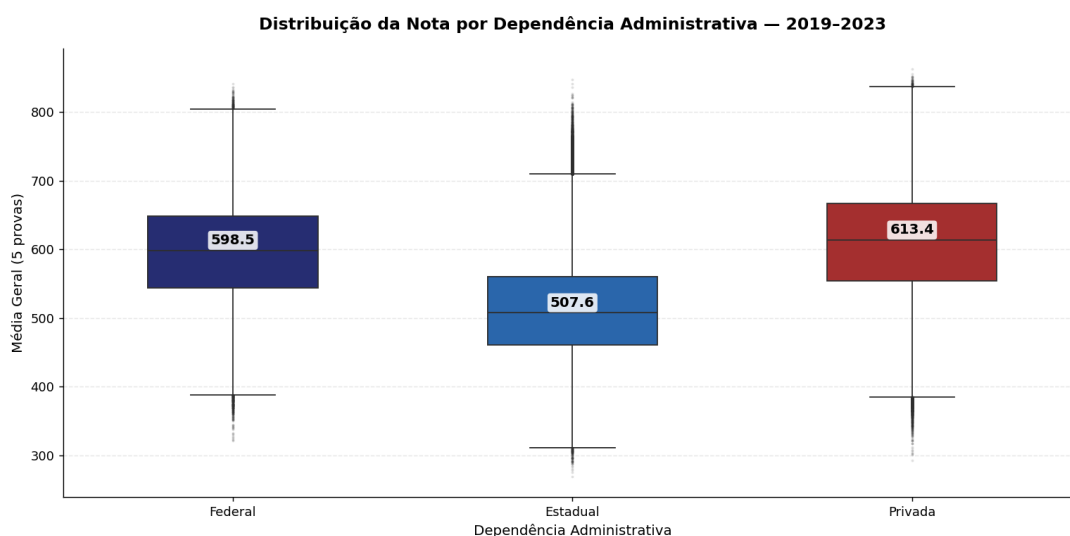
Quando o assunto é acesso a computador, a diferença é bem maior: mais da metade dos alunos da rede pública (52,6%) declara não ter nenhum em casa, contra 12,3% na rede privada. Na outra ponta, cerca de 44% dos alunos da privada possuem dois ou mais computadores, situação de menos de 9% dos alunos da pública (Figura 12). Acumulando essas faixas, cerca de 91% dos alunos da rede pública têm no máximo um computador em casa, enquanto na rede privada esse grupo fica em torno de 56%.

Figura 12 – Acesso a recursos digitais, pública *versus* privada

Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

As análises seguintes separam a rede pública em estadual e federal, divisão que foi justificada no [Capítulo 2](#): as escolas federais adotam processos seletivos baseados em desempenho, o que tende a favorecer alunos de origens mais privilegiadas e aproximar essas instituições do perfil das escolas privadas ([FEINMANN; ROCHA, 2024](#)). A distribuição das notas nas três redes dá o primeiro sinal desse efeito: a mediana da rede federal (598,5) fica muito mais perto da privada (613,4) do que da estadual (507,6), como podemos observar na [Figura 13](#).

Figura 13 – Distribuição da nota por dependência administrativa

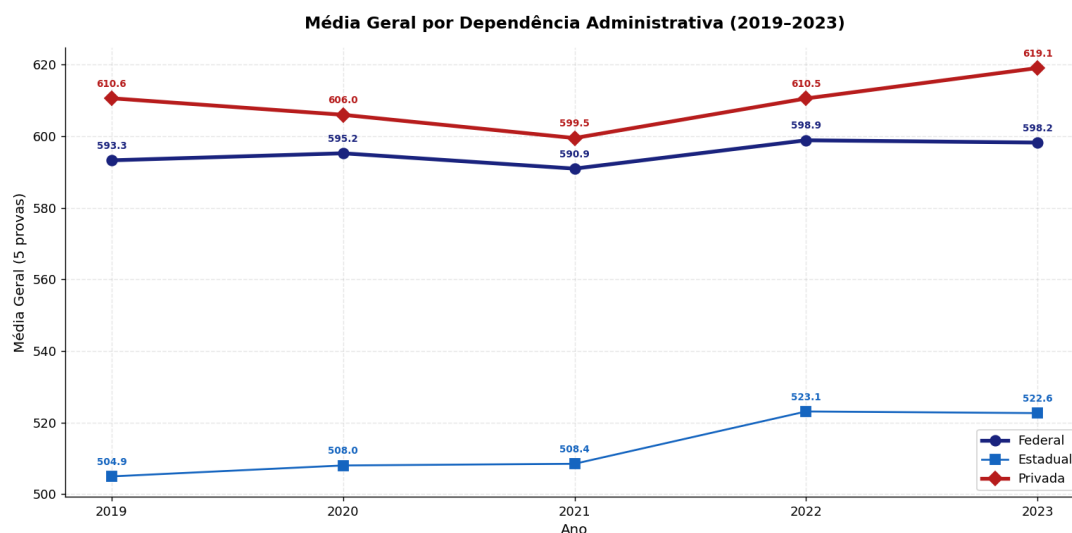


Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Esse comportamento se repete ano a ano: a curva da rede federal acompanha a da

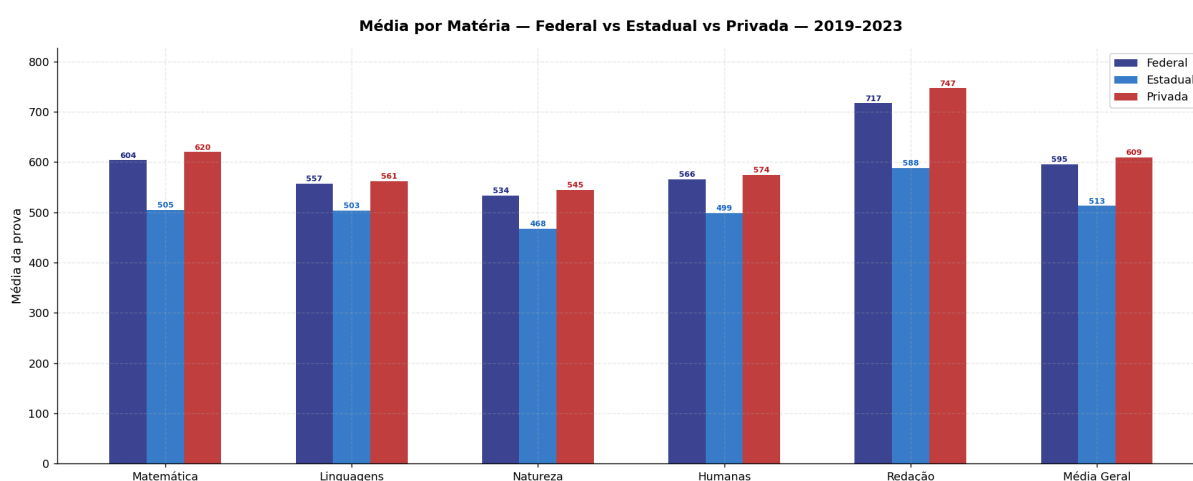
privada durante todo o período, com distância entre 10 e 20 pontos, enquanto a estadual permanece cerca de 90 pontos abaixo das outras duas (Figura 14). O mesmo padrão aparece em todas as matérias (Figura 15).

Figura 14 – Média geral por dependência administrativa por ano



Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Figura 15 – Média por matéria nas três dependências administrativas

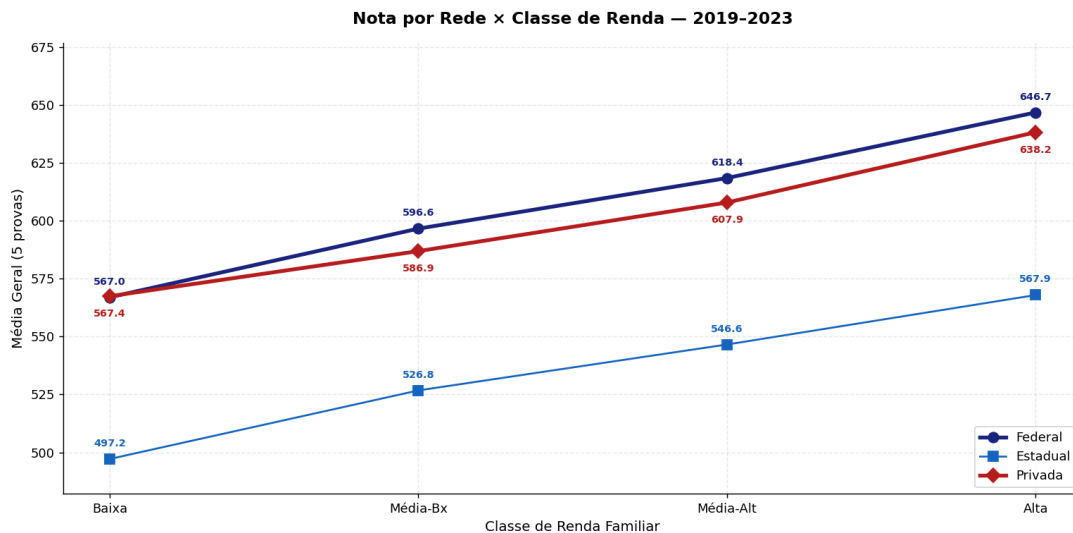


Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

O cruzamento entre a rede e a classe de renda complementa essa comparação. Na classe baixa, os alunos da rede federal praticamente empatam com os da privada (567,0 contra 567,4) e, nas demais classes, a federal passa a apresentar média superior à da própria rede privada (Figura 16). Esse comportamento reforça o argumento de Feinmann

e Rocha (2024) de que as instituições públicas seletivas atendem um público diferente do restante da rede pública, e que agrupar todas as escolas públicas em um grupo único poderia mascarar a realidade do aluno da rede estadual.

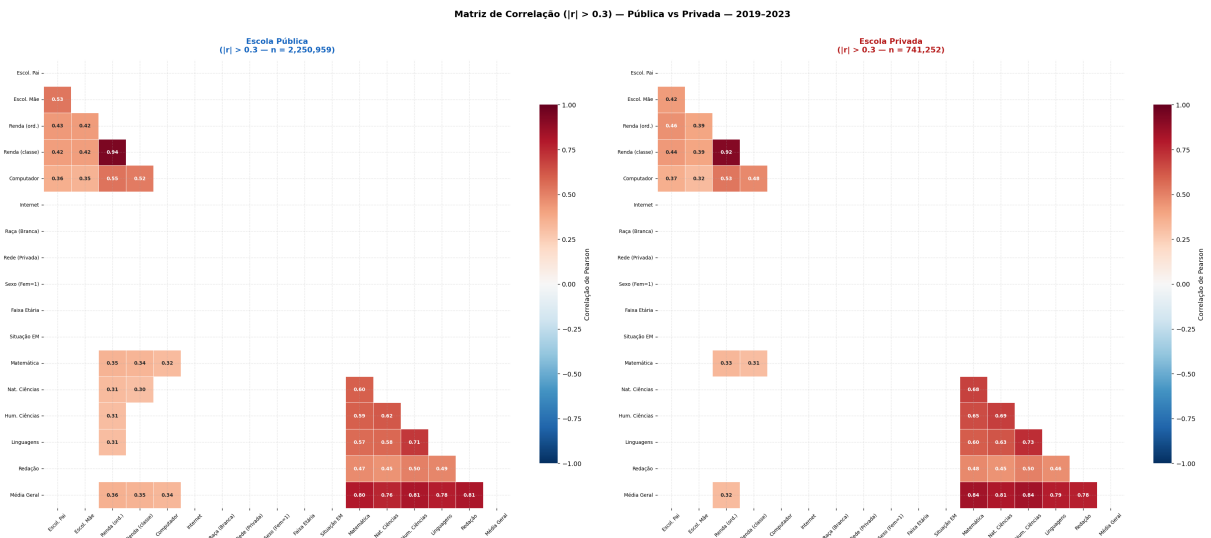
Figura 16 – Nota média por dependência administrativa e classe de renda



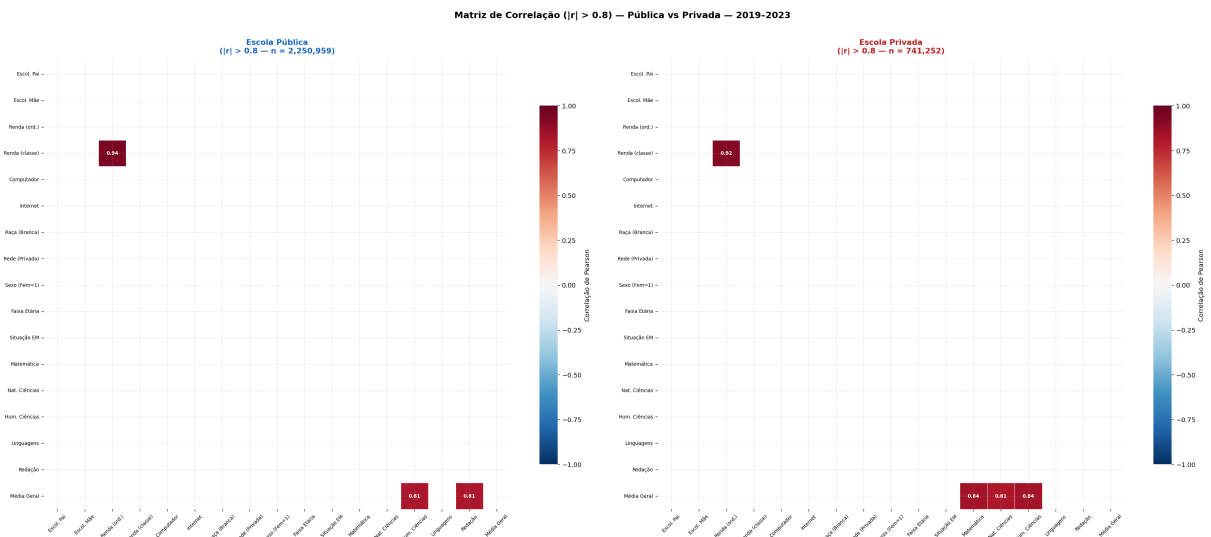
Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

4.2 Análise de correlação entre as variáveis

Antes de partir para o modelo, calculamos o coeficiente de correlação de Pearson, ρ , entre as variáveis socioeconômicas e as notas, na expectativa de encontrar associações diretas mais fortes. Essa análise, porém, não trouxe grandes resultados. Mesmo com um filtro exploratório de $|\rho| > 0,3$, a maior correlação entre uma variável socioeconômica e a nota foi a da renda com a média geral, de apenas 0,36 na rede pública (Figura 17). Quando aplicamos o filtro mais rígido, de $|\rho| > 0,8$, sobram apenas as correlações entre as notas das provas e a média geral e entre as duas codificações de renda, que carregam a mesma informação (Figura 18).

Figura 17 – Matriz de correlação com filtro $|\rho| > 0,3$, pública *versus* privada

Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Figura 18 – Matriz de correlação com filtro $|\rho| > 0,8$, pública *versus* privada

Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Ainda assim, a análise deixou uma informação útil. As associações moderadas aparecem entre as próprias variáveis socioeconômicas, como renda e quantidade de computadores (0,55) e a escolaridade do pai com a da mãe (0,53), mostrando que esses fatores andam juntos na vida do aluno. Como nenhuma variável isolada explica a nota e os fatores se sobrepõem, esse resultado acaba reforçando a escolha metodológica do trabalho: a

regressão logística permite estimar o efeito de cada variável controlando as demais, como descrito no [Capítulo 3](#).

4.3 Regressão Logística (*odds ratio*) entre as redes de ensino

Depois da análise exploratória e da análise de correlação, partimos para a regressão logística binária. Foram ajustados dois modelos, sempre com a rede pública comparada como categoria de interesse ($Y = 1$) e a rede privada como categoria de referência ($Y = 0$): um comparando a rede estadual com a privada, com cerca de 2,77 milhões de registros, e outro comparando a rede federal com a privada, com cerca de 1,08 milhão. As variáveis preditoras e suas codificações estão descritas na [Tabela 5](#), no [Capítulo 3](#).

Todos os coeficientes estimados foram estatisticamente significativos, com p-valor abaixo de 0,01. Ou seja, considerando grau de significância de $\alpha = 0,01$, cada uma das variáveis preditoras permitiu recusar, com significância estatística, a hipótese nula de que não exerce nenhuma influência sobre a variável dependente “tipo de escola” ([seção 3.3](#)). A [Tabela 6](#) apresenta os valores estimados de razão de chances (OR) de cada variável nos dois modelos, com os respectivos intervalos de confiança de 95%, e a [Figura 19](#) exibe esses mesmos valores em forma gráfica.

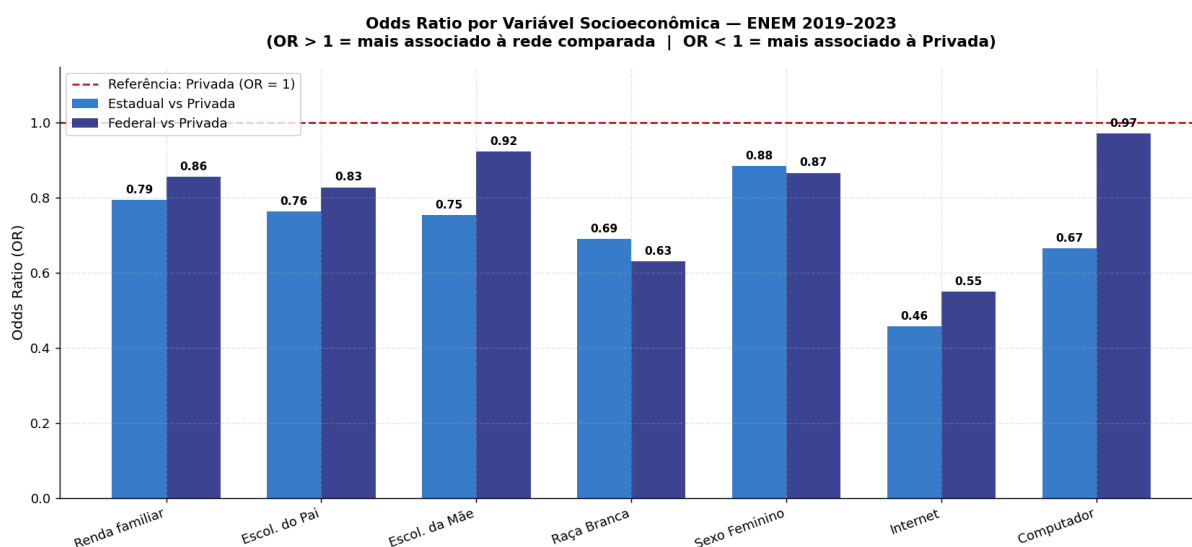
Tabela 6 – Razão de chances e intervalos de confiança de 95% nos dois modelos

Variável	Estadual <i>vs</i> Privada		Federal <i>vs</i> Privada	
	OR	IC 95%	OR	IC 95%
Intercepto	245,2231	[239,6432; 250,9330]	4,5119	[4,3700; 4,6585]
Renda familiar	0,7940	[0,7930; 0,7949]	0,8566	[0,8550; 0,8582]
Escolaridade do pai	0,7631	[0,7610; 0,7654]	0,8283	[0,8246; 0,8320]
Escolaridade da mãe	0,7543	[0,7520; 0,7566]	0,9239	[0,9194; 0,9284]
Raça branca	0,6900	[0,6854; 0,6947]	0,6311	[0,6242; 0,6380]
Sexo feminino	0,8849	[0,8791; 0,8909]	0,8669	[0,8577; 0,8761]
Acesso à internet	0,4567	[0,4476; 0,4660]	0,5491	[0,5342; 0,5645]
Qtd. de computadores	0,6650	[0,6621; 0,6680]	0,9720	[0,9654; 0,9786]

Fonte: elaborado pelo autor. Todos os coeficientes com p-valor $< 0,01$.

Nenhum dos intervalos de confiança inclui o valor 1, o que reforça a significância estatística das associações estimadas. A amplitude reduzida dos intervalos decorre do tamanho das amostras utilizadas.

Figura 19 – Razão de chances por variável socioeconômica nas comparações Estadual *versus* Privada e Federal *versus* Privada



Fonte: elaborado pelo autor a partir dos microdados do ENEM (INEP).

Todos os OR ficaram abaixo de 1, ou seja, o aumento de qualquer uma das variáveis reduz a chance de o aluno pertencer à rede pública comparada, sempre em relação à privada. Uma forma direta de ler esses valores é em percentual, como explicado na [seção 3.3](#): um OR de 0,75, por exemplo, indica que aquela característica reduz em 25% as chances de o aluno estar na rede pública comparada, em relação a um aluno de mesmo perfil nas demais variáveis.

4.3.1 Variáveis preditoras de renda familiar

Começando pela renda familiar, os OR foram de 0,79 na comparação da rede estadual e de 0,86 na da federal, ou seja, cada nível a mais na escala de renda reduz em 21% as chances de o aluno estar na rede estadual e em 14% as de estar na federal. Como a escala tem 17 níveis, o efeito acumulado entre uma família das primeiras faixas e uma das últimas é grande. Na prática, quanto maior a renda, mais o perfil do aluno se aproxima da rede privada, na linha do peso que a renda tem nas desigualdades educacionais discutidas no [Capítulo 2](#) (SENKEVICS; CARVALHAES; RIBEIRO, 2022).

4.3.2 Variáveis preditoras de escolaridade dos pais e responsáveis

A escolaridade da mãe se divide entre os dois modelos. Na comparação com a rede estadual, o OR de 0,75 é o maior efeito entre as variáveis de contexto familiar: cada nível a mais na formação da mãe reduz em 25% as chances de o aluno estar na estadual.

Vale lembrar que essa escala tem 7 níveis, de 0 a 6, indo de quem nunca estudou até a pós-graduação ([subseção 3.2.2](#)). Na federal, porém, o OR sobe para 0,92, uma redução de apenas 8% por nível. Isso mostra que a formação da mãe separa bem o aluno da estadual do aluno da privada, mas quase não separa o da federal. Esses resultados vão na direção do capital cultural familiar apontado por [Lourenço \(2016\)](#) como um dos preditores mais relevantes do desempenho escolar.

Já a escolaridade do pai segue o caminho da renda, com OR de 0,76 e 0,83: cada nível a mais na formação do pai reduz em 24% as chances de o aluno estar na rede estadual e em 17% as de estar na federal, em relação à privada.

4.3.3 Variáveis preditoras de acesso à internet e computadores

O acesso à internet foi a variável que mais separou a rede estadual da privada, com o maior efeito individual dos dois modelos: ter internet em casa reduz em 54% as chances de o aluno estar na rede estadual (OR de 0,46) e em 45% as de estar na federal (OR de 0,55). Na prática, o aluno sem internet em casa é o retrato mais nítido da rede estadual, o que conversa com a vantagem concreta do acesso a recursos digitais apontada por [Carvalho \(2022\)](#) e com o que vimos na [Figura 12](#).

A quantidade de computadores em casa se comportou de forma diferente nos dois modelos: cada computador a mais reduz em 33% as chances de o aluno estar na rede estadual (OR de 0,67), mas em apenas 3% as de estar na federal, com OR de 0,97, muito próximo de 1. Em outras palavras, nesse aspecto o aluno da rede federal se parece com o da privada, e não com o da estadual. Esse achado se soma aos anteriores para reforçar o efeito de filtro social das escolas federais discutido por [Feinmann e Rocha \(2024\)](#): por trás do rótulo de escola pública, a rede federal atende um público com perfil socioeconômico próximo do da rede privada.

4.3.4 Demais análises: variáveis preditoras de raça e sexo

Na variável de raça, os candidatos foram agrupados em brancos e não-brancos, como descrito na [subseção 3.2.2](#). O aluno autodeclarado branco tem 31% menos chances de estar na rede estadual (OR de 0,69) e 37% menos de estar na federal (OR de 0,63) do que um aluno não-branco de mesmo perfil. Foi a variável de maior separação entre a federal e a privada: mesmo comparando alunos iguais em renda, escolaridade dos pais e recursos em casa, a rede privada concentra mais alunos brancos.

O sexo teve o efeito mais fraco dos dois modelos, com OR de 0,88 e 0,87: ser do sexo feminino reduz em cerca de 12% as chances de o aluno estar na rede estadual e em 13% as de estar na federal, sempre em comparação com um aluno do sexo masculino de mesmo perfil socioeconômico. É uma diferença pequena perto das demais variáveis, mas

ainda assim estatisticamente significativa.

No conjunto, os resultados dos dois modelos confirmam a associação entre as condições socioeconômicas do aluno e o tipo de escola frequentada, discutida ao longo do Capítulo 2 (CAMPOS; MORAES; PERES, 2025; MOTTA, 2025; SENKEVICS; CARVALHAES; RIBEIRO, 2022).

5 Conclusão

Este trabalho analisou a relação entre variáveis socioeconômicas e o tipo de escola frequentada pelos participantes do ENEM entre 2019 e 2023, utilizando os microdados públicos do INEP. A análise partiu de uma etapa exploratória (seção 4.1), que comparou o desempenho e o perfil dos candidatos das redes pública e privada, e terminou na regressão logística binária, que estimou a razão de chances de cada variável nas comparações da rede estadual e da rede federal com a privada.

Os objetivos definidos no Capítulo 1 foram atendidos. A análise exploratória confirmou, nos dados recentes do exame, os padrões discutidos no Capítulo 2: a nota cresce junto com a renda familiar e com a escolaridade dos pais nas duas redes, e o perfil dos alunos da rede privada é bem diferente do perfil da rede pública, seja na formação dos responsáveis, seja no acesso à internet e computador em casa. Percebemos também que a desigualdade não aparece só entre as redes, mas dentro delas, como no recorte por raça, em que os alunos não-brancos ficam cerca de 30 pontos abaixo na rede pública e 25 na privada.

Na parte do modelo de regressão logística (seção 4.3), todas as variáveis apresentaram OR abaixo de 1 e significância estatística, confirmando a associação entre as condições socioeconômicas e o tipo de escola. O acesso à internet foi o fator que mais separou a rede estadual da privada, enquanto a raça foi o que mais separou a federal da privada. O resultado que melhor resume o trabalho, porém, é o da quantidade de computadores em casa: a variável diferencia bem o aluno da rede estadual do aluno da privada, mas praticamente não diferencia o da federal. Somado ao desempenho das escolas federais, que empata ou até supera o da rede privada quando controlamos a classe de renda, esse achado sustenta a decisão metodológica de separar as redes estadual e federal e confirma o efeito de filtro social das instituições públicas seletivas apontado na literatura.

O trabalho tem limitações que precisam ser registradas. As variáveis do questionário socioeconômico são autodeclaradas, o que traz imprecisões. A regressão logística binária foi o único modelo estatístico utilizado, e os resultados indicam associação entre as variáveis, não relação de causa e efeito. Além disso, a comparação entre edições do ENEM exigiu adaptações nas variáveis, como no caso da renda, que muda de valores entre os anos (subseção 3.2.1).

Ainda que os resultados indiquem associação e não causalidade, eles sustentam algumas implicações práticas em termos de política pública. A primeira diz respeito ao **acesso digital**. O acesso à internet foi o preditor de maior magnitude entre a rede estadual e a privada (OR de 0,46), e a análise exploratória mostrou que 52,6% dos alunos

da rede pública declaram não ter nenhum computador em casa, contra 12,3% na privada. Programas de inclusão digital voltados especificamente ao aluno da rede estadual, e não à rede pública tomada como bloco único, atuariam sobre o fator que mais diferencia esse público.

A segunda diz respeito ao **desenho dos processos seletivos da rede federal**. Os resultados mostram que a escolaridade da mãe reduz em 25% as chances de o aluno estar na rede estadual, mas em apenas 8% as de estar na federal, e que a quantidade de computadores praticamente não diferencia o aluno da federal do aluno da rede privada. Isso sugere que a simples expansão da rede federal de ensino médio não amplia por si só o acesso dos estudantes em maior desvantagem socioeconômica, uma vez que o ingresso por desempenho tende a selecionar um público já socioeconomicamente favorecido. Políticas de reserva de vagas ou de bonificação por origem escolar nos processos seletivos dessas instituições atuariam diretamente sobre esse filtro.

A terceira diz respeito ao **monitoramento e à produção de indicadores**. Os achados mostram que a dicotomia entre escola pública e privada, adotada com frequência em avaliações e diagnósticos educacionais, agrega realidades bastante distintas: a rede federal apresenta perfil socioeconômico e desempenho próximos aos da rede privada, enquanto a estadual permanece cerca de 90 pontos abaixo de ambas. Indicadores que desagreguem a rede pública por dependência administrativa evitariam que a média nacional mascare a realidade do aluno da rede estadual, que concentra 65,5% dos participantes do período analisado. Na mesma direção, o fato de a desigualdade racial persistir *dentro* de cada rede indica que ações voltadas apenas à redução das diferenças *entre* redes não seriam suficientes.

Para trabalhos futuros, fica a sugestão de explorar melhor o recorte racial no ENEM, pouco presente na literatura que revisamos, e de aprofundar as questões sociais levantadas ao longo da análise, buscando compreender melhor os mecanismos por trás das associações que encontramos.

Referências

- AGRESTI, A. *Categorical Data Analysis*. 2. ed. Hoboken: John Wiley & Sons, 2002. Citado 2 vezes nas páginas 25 e 26.
- CAMPOS, L. B. A.; MORAES, C. P. de; PERES, R. T. Considerações acerca dos impactos da renda familiar e da escolaridade materna no desempenho em Matemática no ENEM 2022. *Revista Educação e Políticas em Debate*, v. 14, n. 2, p. 1–15, 2025. Citado 5 vezes nas páginas 15, 18, 19, 28 e 42.
- CARVALHO, G. A. *O impacto da pandemia nas notas do ENEM na área de Matemática e suas tecnologias*. Dissertação (Monografia de Graduação) — Universidade Federal de Ouro Preto, Ouro Preto, 2022. Citado 7 vezes nas páginas 13, 17, 18, 19, 30, 34 e 41.
- FEINMANN, J.; ROCHA, R. H. Social mobility and higher education in Brazil. Working paper, UC Berkeley. 2024. Citado 7 vezes nas páginas 15, 18, 19, 29, 35, 37 e 41.
- FIELD, A. *Discovering Statistics Using IBM SPSS Statistics*. 5. ed. London: SAGE Publications, 2018. Citado na página 25.
- HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied Logistic Regression*. 3. ed. Hoboken: Wiley, 2013. Citado 2 vezes nas páginas 25 e 26.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Dicionário de Microdados do ENEM 2019*. Brasília, 2019. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>>. Citado 4 vezes nas páginas 17, 21, 22 e 23.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Dicionário de Microdados do ENEM 2020*. Brasília, 2020. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>>. Citado 3 vezes nas páginas 21, 22 e 23.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Dicionário de Microdados do ENEM 2021*. Brasília, 2021. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>>. Citado 3 vezes nas páginas 21, 22 e 23.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Dicionário de Microdados do ENEM 2022*. Brasília, 2022. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>>. Citado 2 vezes nas páginas 21 e 22.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Dicionário de Microdados do ENEM 2023*. Brasília, 2023. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>>. Citado 4 vezes nas páginas 17, 21, 22 e 23.
- LOURENÇO, V. M. *Limites e possibilidades do Enem no processo de democratização do acesso à educação superior brasileira*. Dissertação (Dissertação de Mestrado) — Universidade de Brasília, Brasília, 2016. Citado 4 vezes nas páginas 18, 19, 32 e 41.

MOTTA, S. I. B. da. *Impacto do tipo de escola nas notas do ENEM: análise estatística via regressão logística*. Dissertação (Monografia de Graduação) — Universidade Federal de Ouro Preto, Ouro Preto, 2025. Citado 9 vezes nas páginas 13, 15, 17, 18, 19, 24, 25, 28 e 42.

SENKEVICS, A. S.; CARVALHAES, F.; RIBEIRO, C. A. C. Mérito ou berço? Origem social e desempenho no acesso ao ensino superior. *Cadernos de Pesquisa*, v. 52, p. e09528, 2022. Citado 8 vezes nas páginas 13, 14, 17, 19, 25, 30, 40 e 42.