



AVALIAÇÃO EDUCACIONAL NO ENEM: UM ESTUDO SOBRE QUALIDADE E EQUIDADE

Caroline Ponce de Moraes

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Carlos Eduardo Pedreira

Rio de Janeiro
Novembro de 2021

**AVALIAÇÃO EDUCACIONAL NO ENEM: UM ESTUDO SOBRE QUALIDADE
E EQUIDADE**

Caroline Ponce de Moraes

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA
DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS
PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE
SISTEMAS E COMPUTAÇÃO.

Orientador: Prof. Carlos Eduardo Pedreira, Ph.D.

Aprovada por: Prof. Henrique Luiz Cukierman, D.Sc.

Prof. Eduardo Fonseca Mendes, Ph.D.

Prof.^a Maria Tereza Serrano Barbosa, D.Sc.

Prof. Rodrigo Tosta Peres, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

NOVEMBRO DE 2021

Moraes, Caroline Ponce de.

Avaliação Educacional no ENEM um estudo sobre
qualidade e equidade / Caroline Ponce de Moraes. – Rio de
Janeiro: UFRJ/COPPE, 2021.

XII, 100 p.: il.; 29,7 cm.

Orientadores: Carlos Eduardo Pedreira

Tese (doutorado) – UFRJ/ COPPE/ Programa de
Engenharia de Sistemas e Computação, 2021.

Referências Bibliográficas: p. 73-89.

1. ENEM. 2. Modelos Multiníveis. 3. Equidade. 4.
Qualidade. I. Pedreira, Carlos Eduardo. II. Universidade
Federal do Rio de Janeiro, COPPE, Programa de Engenharia
de Sistemas e Computação. III. Título.

DEDICATÓRIA

Aos meus pais.
Ao meu marido.
Ao meu irmão.
Aos meus sobrinhos.
Família e amigos.
Com amor, admiração e gratidão.

AGRADECIMENTOS

Por muitas vezes achei que esse dia pudesse não chegar. Não foi um caminho fácil, como todos que já fizeram um doutorado sabem. Foram mais de cinco anos de intensa dedicação e doação aos estudos, pesquisas e revisões que culminaram nesta tese. O que me fez chegar até aqui foi a persistência, o objetivo, a dedicação, a minha família, que o tempo todo acreditou em mim, os meus amigos e, não menos importantes, os meus alunos.

Começo agradecendo ao professor e amigo Rodrigo, que ao longo dessa caminhada depositou muita confiança em mim, foi sempre paciente, otimista e soube me conduzir até aqui.

Um agradecimento especial ao meu marido, Leonardo, que incansavelmente me apoiou, me incentivou e que me lembrava que ia dar certo.

Aos meus pais, por me mostrarem a importância da educação e por me colocarem nesse caminho desde muita nova. Se hoje estou aqui, o mérito é de vocês.

Ao meu irmão, com quem compartilho não só os nossos pais, mas também nossa profissão, que me ouviu, me ajudou e vibrou em todas as conquistas.

Em memória, ao meu avô, que mesmo com pouco estudo foi uma das pessoas que mais me ensinou. Ensinou sobre coisas que realmente são importantes e que devemos dar valor na vida. Onde você estiver, o meu muito obrigada.

Aos meus amigos e familiares, que sabem o quanto foi difícil todo esse caminho até aqui e que mantiveram suas torcidas para que esse dia chegasse.

Por fim, mas não menos importante, agradeço ao professor Carlos Eduardo Pedreira por todas as trocas; aos membros da banca por aceitarem o convite e por todas as contribuições sugeridas; ao coordenador Daniel por sua perene disposição em responder minhas dúvidas, assim como ao Guty e ao Ricardo, ao departamento e ao PESC como um todo.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

AVALIAÇÃO EDUCACIONAL NO ENEM: UM ESTUDO SOBRE QUALIDADE E EQUIDADE

Caroline Ponce de Moraes

Novembro/2021

Orientador: Carlos Eduardo Pedreira

Programa: Engenharia de Sistemas e Computação

Os objetivos centrais desta tese são analisar a importância escolar no desempenho dos alunos e apresentar um debate sobre equidade, propondo um estudo sobre o ENEM. Este exame é relevante, pois é uma das formas de mobilidade social em um país tão desigual como o Brasil, além de ser pouco explorado na literatura de avaliação educacional. As principais contribuições são: o cálculo do impacto escolar no ENEM para escolas públicas e privadas; a comparação entre os perfis de alunos de desempenho alto e baixo e a análise sobre equidade, tendo em vista as variáveis de sexo e raça. Utilizam-se modelos multiníveis para o cálculo da importância escolar com base em indicadores educacionais. O resultado é de uma contribuição escolar, denominado efeito escola, de 9% para escolas públicas e 13% para privadas. Usam-se, também, modelos de regressão quantílica para a comparação entre alunos de desempenho alto e baixo, motivando um estudo sobre desigualdade e equidade, que resultou em uma análise do impacto intramunicipal para as variáveis sexo e raça feita através de modelos multiníveis com coeficientes aleatórios. Os resultados indicam que, em muitos casos, obter qualidade implica em queda na equidade, o que mostra a necessidade de se repensar questões envolvendo políticas públicas educacionais. Por fim, vide questões atuais relacionadas a pandemia, uma reflexão sobre possíveis consequências do Coronavírus na educação é feita, discutindo o quanto a ausência de aulas presenciais poderá interferir no desempenho de alunos de baixa renda de escolas públicas em todo o Brasil no ENEM.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements
for the degree of Doctor of Science (D.Sc.)

EDUCATIONAL ASSESSMENT AT ENEM: A STUDY ON QUALITY AND EQUITY

Caroline Ponce de Moraes

November/2021

Advisor: Carlos Eduardo Pedreira

Department: Systems Engineering and Computer Science

The main objectives of this thesis are to analyze the importance of school in student performance and present a debate on equity, proposing a study on the National High School Exam, the ENEM. This exam is particularly relevant, because it is one of the forms of social mobility in an unequal country as Brazil. Besides, it is less explored in the educational assessment literature. The main contributions are: the calculation of school importance in ENEM for public and private schools; the comparison between the profiles of high- and low-performing students; analysis on equity, considering the variables of sex and race. Multilevel models were used to calculate school importance based on educational indicators. The result is a school contribution, called the school effect, of 9% for public schools and 13% for private ones. Quantile regression models were used to compare high- and low-students, thus motivating a study on inequality and equity, which resulted in an analysis of the intramunicipal impact for the gender and race variables, carried out using multilevel models with random coefficients. The results indicate that, in many cases, obtaining quality implies a decrease in equity, which indicates the need to rethink issues involving educational public policies. Finally, considering actual issues about the pandemic, a reflection on possible consequences of the Coronavirus in education is done, discussing how the absence of in-person classes may interfere the ENEM's performance of low-income students in public schools throughout Brazil.

SUMÁRIO

LISTA DE FIGURAS.....	IX
LISTA DE TABELAS.....	X
LISTA DE ABREVIATURAS	XII
1. INTRODUÇÃO	1
2. CONTEXTUALIZAÇÃO.....	5
2.1 PANORAMA DA AVALIAÇÃO EDUCACIONAL NO BRASIL	5
2.1.1 MODELAGEM ESTATÍSTICA APLICADA A EDUCAÇÃO.....	5
2.1.2 LITERATURA BRASILEIRA DE ESTATÍSTICA EM EDUCAÇÃO: UMA CONTEXTUALIZAÇÃO	6
2.1.3 VARIÁVEIS DE SAÍDA.....	7
2.1.4 VARIÁVEIS DE ENTRADA	9
2.2 EFICÁCIA ESCOLAR: O EFEITO ESCOLA	13
2.3 EFICÁCIA ESCOLAR: DESIGUALDADE DE DESEMPENHO E EQUIDADE	16
3. MATERIAIS E MÉTODOS	19
3.1 BASES DE DADOS	19
3.1.1 BASE DE DADOS: EFEITO ESCOLA	19
3.1.2 BASE DE DADOS: DESIGUALDADE DE DESEMPENHO	23
3.1.3 BASE DE DADOS: EQUIDADE	24
3.2 MODELOS ESTATÍSTICOS.....	24
3.2.1 MODELO CONCEITUAL: EFEITO ESCOLA.....	24
3.2.2 MODELO ESTATÍSTICO: EFEITO ESCOLA	25
3.2.3 MODELO ESTATÍSTICO: DESIGUALDADE DE DESEMPENHO.....	27
3.2.4 MODELOS ESTATÍSTICOS: EQUIDADE	28
4. RESULTADOS E DISCUSSÕES	32
4.1 RESULTADOS E DISCUSSÕES: EFEITO ESCOLA	32
4.2 RESULTADOS E DISCUSSÕES: DESIGUALDADE DE DESEMPENHO	40
4.2.1 RESULTADOS DOS MODELOS	41
4.2.2 ANÁLISES SOBRE VARIÁVEIS FAMILIARES E ESCOLARES	45
4.3 RESULTADOS E DISCUSSÕES: EQUIDADE	51
4.3.1 ANÁLISE EXPLORATÓRIA	52
4.3.2 RESULTADOS DOS MODELOS ESTATÍSTICOS E DISCUSSÕES	55
5. VARIÁVEIS FAMILIARES EM TEMPOS DE PANDEMIA: UMA REFLEXÃO A PARTIR DE DADOS PRÉ-PANDÊMICOS DO ENEM	62
5.1 CONSIDERAÇÕES INICIAIS.....	62
5.2 DESAFIOS PARA A EDUCAÇÃO EM TEMPOS DE PANDEMIA	64
5.3 MATERIAIS E ANÁLISE DESCRIPTIVA	65
5.4 ENEM EM TEMPOS DE PANDEMIA	71
6 CONCLUSÃO.....	72
7 REFERÊNCIAS BIBLIOGRÁFICAS.....	75
APÊNDICE A – BASE DE DADOS E DADOS FALTANTES	90
A.1 BASE DE DADOS	90
A.2 ANÁLISE DE DADOS FALTANTES	90
APÊNDICE B – CÓDIGOS “R”	91
APÊNDICE C – TABELAS	94
APÊNDICE D – ANÁLISE RESIDUAL	97

LISTA DE FIGURAS

Figura 3.1 – Diagrama das variáveis associadas a cada nível do modelo.....	25
Figura 4.1– Comparação entre variáveis percentuais para escolas públicas e privadas	33
Figura 4.2 – Complexidade de gestão e nível socioeconômico médio dos candidatos ao ENEM 2017	34
Figura 4.3 – Histogramas das notas em matemática dos candidatos ao ENEM 2017.	35
Figura 4.4 – Renda familiar dos candidatos ao ENEM 2017 por dependência administrativa	36
Figura 4.5 – Densidades estimadas de notas matemáticas de estudantes de escolas públicas e privadas	46
Figura 4.6 – Distribuição da educação materna de alunos de escolas públicas e privadas	47
Figura 4.7 – Mapeamento de cores do número de alunos por renda e desempenho por quantil para escolas públicas	49
Figura 4.8 – Mapeamento de cores do número de alunos por renda e desempenho por quantil para escolas privadas.....	49
Figura 4.9 – Equidade x Qualidade para os 30 melhores e 30 piores municípios em relação à raça no ENEM 2017.....	58
Figura 4.10 – Equidade x Qualidade para os 30 melhores e 30 piores municípios em relação a sexo no ENEM 2017.	59
Figura 5.1 – Proporções dos níveis de escolaridade materna nas cidades entre 50 e 500 mil habitantes que tiveram alunos prestando o ENEM 2017, por Região do Brasil	67
Figura 5.2 – Proporções de bibliotecas e salas de leitura nas cidades entre 50 e 500 mil habitantes que tiveram alunos prestando o ENEM 2017, por Região do Brasil.....	70

LISTA DE TABELAS

Tabela 3.1 – Descrição das variáveis dependentes e independentes relacionadas às características individuais e familiares	21
Tabela 3.2 – Descrição das variáveis independentes relacionadas ao efeito escola... <td>22</td>	22
Tabela 3.3 – Descrição das variáveis independentes relacionadas ao efeito escola – infraestrutura	23
Tabela 3.4 – Variáveis independentes e coeficientes aleatórios por modelo, sendo escolaridade: escolaridade materna; renda: renda familiar; NSE: nível socioeconômico médio escolar	29
Tabela 4.1 – Distribuição das notas em matemática no ENEM 2017 por dependência administrativa	35
Tabela 4.2 – Resultado da modelagem multinível para escolas públicas e privadas .	37
Tabela 4.3 – Resultados do modelo de regressão quantílica e multinível para matemática em escolas públicas	42
Tabela 4.4 – Resultados do modelo de regressão quantílica e multinível para matemática em escolas privadas	43
Tabela 4.5 – Valores dos quantis para a proficiência matemática do aluno em escolas públicas e privadas	46
Tabela 4.6 – Distribuições das proficiências em matemática no ENEM 2017 por raça. <td>52</td>	52
Tabela 4.7 – Distribuições das proficiências em matemática no ENEM 2017 por sexo.	53
Tabela 4.8 – Desigualdades em desempenho segmentadas por sexo, raça e escolaridade materna	53
Tabela 4.9 – Desigualdades em desempenho segmentadas por sexo, raça e renda familiar.....	54
Tabela 4.10 – Resultados dos modelos 1 a 5 para efeitos fixos e aleatórios.	55
Tabela 5.1 – Distribuição de acesso, em casa, de computadores e internet por Região em, porcentagem (%).	65
Tabela 5.2 – Distribuição dos estudantes e escolas por Região.	66
Tabela 5.3 – Distribuição da renda familiar por Região nos quartis 25, 50, 75 e na média.	68
Tabela 5.4 – Médias das proficiências em matemática no ENEM 2017 por Região e escolaridade da mãe.....	69

Tabela 5.5 – Médias das proficiências em matemática no ENEM 2017 por Região considerando renda familiar inferior (<) ou superior (>) a três salários mínimos. 70

LISTA DE ABREVIATURAS

ABREVIATURAS

- CEC - Commission of the European Communities
- ENADE - Exame Nacional de Desempenho dos Estudantes
- ENEM - Exame Nacional do Ensino Médio
- FUNDEF - Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério
- IBGE - Instituto Brasileiro de Geografia e Estatística
- INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
- ICC - Intraclass correlation coefficient
- IDEB - Índice de Desenvolvimento da Educação Básica
- PIRLS - Progress in International Reading Literacy Study
- PISA - Programa Internacional de Avaliação de Alunos
- ProUni - Programa Universidade para Todos
- SAEB - Sistema de Avaliação da Educação Básica
- SEDES - Secretaria de Desenvolvimento Social
- SEDUCE-GO - Secretaria de Estado da Educação de Goiás
- SISU - Sistema de Seleção Unificada
- TIMSS - Trends in International Mathematics and Science Study

1. INTRODUÇÃO

Os estudos na área de avaliação educacional foram iniciados no século XX e têm sido frequentes na literatura internacional a partir da década de 60, onde contribuições (COLEMAN *et al.*, 1966, BOURDIEU e PASSERON, 1975) indicam pouca relação entre a escola e o desempenho, considerando as condições socioeconômicas dos alunos como fator relevante para o sucesso estudantil.

A pergunta sobre qual o real impacto escolar no desempenho tem sido estudada e debatida desde então (SOARES, 2004) a partir dos processos e fatores escolares, além das variáveis individuais e socioeconômicas dos alunos.

Nesta tese, propõe-se um estudo sobre o Exame Nacional do Ensino Médio, o ENEM. As duas principais motivações para a escolha deste exame são: a importância do ENEM no Brasil em termos de mobilidade social e sua pouca utilização na literatura de avaliação educacional. A base de dados é referente ao exame de 2017 e os objetivos globais são calcular a contribuição das escolas públicas e privadas nas notas dos alunos da Região Sudeste, analisar os alunos de desempenho alto e baixo, com enfoque nas desigualdades sociais e apresentar um estudo sobre equidade para raça e sexo, considerando alunos de escolas públicas de todo o Brasil. Para finalizar a tese, uma discussão sobre as possíveis implicações da pandemia do Coronavírus será realizada. Mais especificamente, os objetivos podem ser descritos como segue:

- (i) Apresenta-se uma contextualização de estatística aplicada a educação, com um breve referencial sobre variáveis de entrada e saída que compõem modelos de avaliação educacional, além de conceitos sobre eficácia escolar, efeito escola, efeito família e equidade.
- (ii) Calcula-se a contribuição escolar no ENEM 2017, denominada efeito escola, tendo por base os indicadores educacionais disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) como fatores escolares. O banco de dados utilizado é composto por alunos de escolas públicas e privadas que vivem em cidades entre 50 e 500 mil habitantes da Região Sudeste do Brasil. Este recorte foi feito em virtude da Região Sudeste apresentar o maior percentual de candidatos ao ENEM e para evitar comparações entre municípios discrepantes em termos populacionais, já que cidades entre 50 e 500 mil

habitantes podem ser consideradas como cidades médias (SPOSITO, 2001, WENDEL, 2010). É utilizada uma modelagem multinível (aluno, escola) apropriada para dados hierárquicos. O algoritmo de seleção de variáveis Lasso (HASTIE *et al.*, 2001, HASTIE *et al.*, 2015, MASCI *et al.*, 2016) é usado como comparação e conclui-se que as variáveis selecionadas pelo Lasso são muito similares às consideradas significativas pelo modelo multinível. Como principal resultado, tem-se que o efeito escola para escolas públicas é de 9% e o efeito escola para escolas privadas é de 13%. É importante mencionar que, ao se utilizar os indicadores, chega-se a um efeito escola para o ENEM compatível com a literatura quando se controla o modelo pelas variáveis socioeconômicas (TRAVITZKI e BOTO, 2013), o que mostra que as variáveis do INEP podem ser consideradas boas *proxies* para qualidade docente, informação sobre os pares e gestão escolar. Deve-se considerar, também, que pequenas proporções atribuídas às escolas podem ser muito importantes para o desempenho individual de um aluno (REYNOLDS *et al.*, 2011);

- (iii) Após o cálculo da contribuição escolar, o objetivo é identificar características de alunos de desempenho alto, médio ou baixo, além de desigualdade entre fatores escolares e familiares de estudantes de escolas públicas e privadas no ENEM 2017. Para isso, uma modelagem de regressão quantílica é proposta a fim de analisar as diferenças entre os fatores associados à proficiência. Além das variáveis individuais e socioeconômicas e dos indicadores utilizados anteriormente, são adicionadas ao banco de dados variáveis de infraestrutura escolar. Uma análise descritiva sobre desigualdade destes fatores também é apresentada e se observa que, como já seria esperado, os alunos de escolas públicas estão em uma situação de prejuízo, considerando fatores socioeconômicos e escolares em comparação aos alunos de escolas privadas;
- (iv) A análise de desigualdade proposta motiva um debate sobre equidade, que é apresentado a seguir. Para medir a equidade, utilizam-se dados de alunos de todo o Brasil, que prestaram o ENEM 2017 e cujas escolas pertencem a municípios brasileiros entre 50 e 500 mil habitantes. A ideia é analisar as variáveis individuais: raça e sexo. Os resultados indicam desigualdade de desempenho entre os grupos sociais. O estudo proposto é intramunicipal, o que possibilita identificar regiões que se destacam por

qualidade ou equidade. Cinco modelos multiníveis são ajustados (alunos, escolas e municípios), com estudo de coeficientes aleatórios para variáveis individuais. Verifica-se que, para raça e sexo, há uma discrepância grande entre os municípios na comparação qualidade *versus* equidade. Isso significa que, para estas variáveis, obter qualidade implica em queda na equidade, o que indica a necessidade de se repensar questões envolvendo políticas públicas educacionais.

- (v) Uma reflexão sobre possíveis consequências da pandemia do Coronavírus na educação é feita, a partir da discussão de o quanto a ausência de aulas presenciais pode interferir no desempenho de alunos de baixa renda de escolas públicas em todo o Brasil no ENEM.

Conforme mencionado anteriormente, o ENEM é pouco explorado na literatura de avaliação educacional. O motivo está associado ao viés induzido pelas diferenças entre a amostra da população pré-universitária, que é aquela que se candidata ao exame, e uma amostra aleatória de estudantes do ensino médio. É razoável esperar que haja diferenças importantes de preparação entre estes grupos, uma vez que os que escolhem se candidatar provavelmente estão mais bem preparados que os demais por acreditar que podem concorrer a uma vaga de ensino superior. Pode-se considerar que os resultados encontrados para o ENEM não podem ser generalizados para toda a população de alunos do ensino médio, uma vez que apenas um subconjunto desta população tenha se candidatado ao exame. Por outro lado, isso ocasionou uma lacuna na literatura em termos de avaliações da população pré-universitária. Esta tese possui o objetivo de avançar no estudo deste grupo. Apesar dos resultados obtidos não serem generalizáveis para toda a população de alunos do ensino médio, conjectura-se que as diferenças encontradas entre alunos de escolas públicas e privadas e entre os desempenhos de diferentes grupos sociais, se fossem considerados todos os alunos ao final do ensino médio, seriam superiores àquelas aqui encontradas.

As contribuições aqui apresentadas estão relacionadas às análises quantitativas sobre a importância escolar e às diferenças de desempenho entre grupos sociais. Está fora do escopo deste trabalho criar propostas para elaborações de políticas públicas. Nosso objetivo global consiste em prover informações para que, oportunamente, políticas educacionais e sociais sejam criadas.

O capítulo 2 contextualiza a tese, apresentando um panorama geral sobre o cenário da avaliação educacional no Brasil, além dos conceitos de eficácia escolar, qualidade e equidade. As descrições dos dados utilizados bem como a metodologia se encontram

no capítulo 3. No capítulo 4, encontram-se os resultados e discussões. Uma breve reflexão sobre o cenário do ENEM em tempos de pandemia pode ser encontrada no capítulo 5. O capítulo 6 conclui a tese.

2. CONTEXTUALIZAÇÃO

Neste capítulo, apresenta-se uma contextualização do trabalho, iniciando com o panorama do cenário de avaliação educacional no Brasil. Propõe-se uma breve análise crítica, acompanhada de discussões sobre modelagem de dados educacionais e variáveis típicas de entrada e saída. Posteriormente, o conceito de eficácia escolar é discutido, com foco tanto na qualidade escolar quanto na desigualdade de desempenho e equidade.

2.1 PANORAMA DA AVALIAÇÃO EDUCACIONAL NO BRASIL

A relação técnica entre a saída de um processo e as entradas dos fatores associados a ela pode ser definida, em economia, como função de produção. Através desta função, pode-se avaliar a produtividade de um processo (MONL, 1989, HANUSHEK, 2015). Em termos educacionais, a função de produção em educação tipicamente relaciona entradas como, por exemplo, recursos escolares, qualidade docente e atributos familiares a uma saída que, usualmente, é uma medida de sucesso ou produtividade do aluno (HANUSHEK, 2007), ou, genericamente, da escola (BOWLES, 1970). Identificar os fatores que implicam em melhor desempenho do aluno é uma tarefa importante, uma vez que possibilita à criação de políticas públicas bem sucedidas. Entretanto, esta busca muitas vezes é deixada de lado em discussões sobre políticas educacionais e o objetivo passa a ser aumentar a qualidade sem a identificação apropriada de quais fatores implicam, de fato, nesta melhora (HANUSHEK e ETTEMA, 2017). Além de acarretar em investimento público ou privado em fatores inconsistentes para melhorar o aprendizado, isso também causa prejuízo aos estudantes, uma vez que gerações são afetadas por políticas de ensino ineficientes.

2.1.1 MODELAGEM ESTATÍSTICA APLICADA A EDUCAÇÃO

Do ponto de vista econômico, uma função de produção de educação (FPE) pode ser definida de formas relativamente diferentes na literatura, com pequenas variações conceituais e de notação (ESPINOSA, 2017). Em sua forma mais simples, a definição de FPE aqui apresentada é inspirada na abordagem de BOWLES (1970), da seguinte forma:

$$Y = f(X, F, P) \quad (2.1)$$

onde Y é uma medida de desempenho dos alunos; X é um vetor de medidas da escola, professores e pares, F é um vetor com características familiares e P é um vetor com características individuais do aluno. É importante citar que a estimativa empírica desta função, segundo BOWLES (1970), apresenta a saída individual para cada aluno. Exemplos de medidas do vetor X podem ser o tamanho da turma, o grau de formação do docente ou o desempenho dos colegas de classe; do vetor F, a escolaridade dos pais e, do vetor P, as habilidades do aluno ou características pessoais, como raça ou gênero. A partir dos estudos de Coleman, este tipo de formulação econômica do problema possibilita uma série de contribuições na literatura de estatística aplicada à educação, com objetivo de mapear a relação entre os insumos escolares e suas influências no desempenho dos alunos.

À medida que os estudos avançavam, o caminho metodológico se solidificava através de modelagens multiníveis, que são apropriadas a dados educacionais, e possibilitam a identificação da contribuição escolar na proficiência. Este tipo de abordagem permite, através dos níveis do modelo, agrupar os alunos em escolas. O número de níveis depende do objetivo do estudo e pode ter enfoque em turmas ou escolas, ou ainda em análises intramunicipais ou estaduais. As contribuições escolar e familiar no desempenho são definidas como efeito escola e efeito família, respectivamente. Do ponto de vista metodológico, esses efeitos são obtidos a partir da contribuição de cada nível do modelo na variação das notas.

Neste capítulo, apresenta-se uma introdução à avaliação educacional no Brasil, onde as questões fundamentais são colocadas à luz das literaturas nacional e internacional, contribuindo para uma visão crítica, embora breve, das contribuições obtidas nas últimas décadas no Brasil. O objetivo deste capítulo não é prover um estado da arte ou uma busca exaustiva da literatura nacional no assunto, mas estabelecer o cenário em que esta tese está inserida, além de contextualizar as contribuições aqui propostas, analisando criticamente a área através de trabalhos considerados relevantes relacionados ao nosso país.

2.1.2 LITERATURA BRASILEIRA DE ESTATÍSTICA EM EDUCAÇÃO: UMA CONTEXTUALIZAÇÃO

Dentro do contexto de modelagens estatísticas aplicadas à educação, as variáveis de saída ou dependentes mais comuns são os testes realizados por estudantes, embora outras análises possam ser levadas em consideração e serão comentadas oportunamente. Quanto às medidas de desempenho, são elas o Sistema de Avaliação

de Educação Básica (SAEB)/Prova Brasil (SAEB, 2021) e o Exame Nacional de Ensino Médio, o ENEM, uma das formas de acesso à universidade no Brasil (INEP, 2021a) ou ainda as metas do IDEB (INEP, 2021b). A principal justificativa é que se tratam dos únicos exames nacionais, até a presente data, que permitem comparar desempenho para os ensinos fundamental e médio, respectivamente (CASTRO, 2009). Diversas variáveis de entrada ou independentes podem ser possíveis para um modelo estatístico, dependendo da análise em questão. É muito comum a utilização de variáveis do censo escolar (INEP, 2021c) do questionário do ENEM (INEP, 2021a) e dos indicadores educacionais INEP (2021d), disponíveis no site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Estas informações em conjunto com variáveis individuais e socioeconômicas permitem medir o efeito escola e o efeito família.

Há concordância entre as medidas usuais de proficiência utilizadas no Brasil em comparação com a literatura internacional. Segundo WERFHORS e MIJS (2010), normalmente são usados dados referentes ao PISA, TIMSS ou PIRLS. O que se observa é que avaliações em larga escala tendem a ser escolhidas para a comparação de desempenho de alunos.

Dependendo do objetivo do estudo, pode haver dificuldades em se obter alguns requisitos metodológicos. Por exemplo, no ENEM o aluno realiza um único exame, e embora não seja o cenário ideal, o conhecimento prévio pode ser substituído por variáveis contextuais. O fato é que, apesar das dificuldades, há um interesse social em se medir o efeito escola em exames como o ENEM, algo feito em TRAVITZKI e BOTO (2013), identificando que o desempenho do aluno é explicado por fatores que estão fora do contexto escolar.

2.1.3 VARIÁVEIS DE SAÍDA

Na maior parte das vezes, as variáveis de saída estão associadas a medidas de proficiência relacionadas aos exames citados na seção anterior, embora isso possa variar de acordo com o objetivo do trabalho. Por outro lado, há contribuições onde se analisa a probabilidade de determinado fato acontecer, algo modelado através de regressões logísticas.

Se a variável de saída for uma medida de proficiência, há alguns estudos que analisam o desempenho em Matemática e Languages (GREMAUD *et al.*, 2007, MARIONI *et al.*, 2014, OLIVEIRA *et al.*, 2017, TAVARES e PONCZEK, 2018, ROCHA e FUNCHAL, 2019, BARTHOLO *et al.*, 2020) e outros apenas em Matemática, (ALBERNAZ *et al.*, 2002, SOARES e COLLARES, 2006, FRANCO *et al.*, 2007,

MENEZES-FILHO, 2007, FRANCO e MENEZES-FILHO, 2008, OLIVEIRA e CARVALHO, 2018). Poucas vezes a análise é focada apenas em Languages (SOARES, 2003), e isso ocorre independente do exame escolhido (Prova Brasil/SAEB, ENEM ou outro). Como a habilidade em matemática é usualmente atribuída à escola, existem muitos estudos que utilizam este exame para medir eficácia escolar. Não há, portanto, a obrigatoriedade de utilizar todos os testes relacionados ao exame escolhido.

Desempenhos médios ou outras funções das proficiências podem ser usados. Em AMÉRICO e LACRUZ (2017), calcula-se a média de Português e Matemática para cada aluno e a variável de saída é a média da escola, considerando a nota média dos alunos como esse indicador. O indicador do IDEB, uma função determinada a partir do produto das médias das proficiências padronizadas por um indicador de rendimento, é usado em ALVES e SOARES (2013). Em relação ao ENEM, há estudos que levam em consideração a nota média (CURI e MENEZES-FILHO, 2013) ou outras notas, além de Matemática e Português (OLIVEIRA *et al.*, 2017). Ainda considerando acesso à universidade, SOARES *et al.* (2001) apresentam uma análise de três anos consecutivos para o vestibular da UFMG.

Considerando outras medidas de saída, RIANI e RIOS-NETO (2008) estimam a probabilidade de o aluno frequentar a escola na idade correta, LEON e MENEZES-FILHO (2002) analisaram as taxas de rendimento (aprovação, reprovação ou evasão), SOARES e SÁTYRO (2008) estudaram a distorção idade-série e OLIVEIRA e SILVA (2018) analisaram a nota do aluno no ENADE (Exame Nacional de Desempenho dos Estudantes) no ano de 2012, encontrando uma relação forte entre nível socioeconômico e desempenho. Modelagens com saídas binárias também são implementadas e regressões logísticas são usadas em MATOS e RODRIGUES (2016), onde a variável resposta era se a escola cumpriu a meta do IDEB 2013 e em CURI e MENEZES-FILHO (2010), com a análise dos fatores determinantes da escolha entre as redes pública ou privada. Em TERRA *et al.* (2012), também foi apresentada uma modelagem para determinar a escolha entre escolas públicas e privadas, mas usando como base o desempenho em Matemática e Português. Uma análise por gênero foi apresentada em CARVALHO e KASSOUF (2009) tendo as variáveis dependentes relacionadas a gastos familiares, onde os resultados empíricos encontraram um cenário favorável à escolarização feminina, no ano de 2009.

2.1.4 VARIÁVEIS DE ENTRADA

Em geral, as análises estatísticas em educação estão relacionadas à medição da eficácia escolar. O objetivo principal varia de acordo com o estudo, mas boa parte se relaciona com medir a qualidade da escola ou a equidade em algum contexto. Para se medir o efeito escola, associado à qualidade escolar, o objetivo é caracterizar a contribuição da escola no desempenho do aluno e, do ponto de vista metodológico, este efeito é calculado através do erro do modelo multinível no nível escola. As variáveis de entrada normalmente utilizadas estão relacionadas à qualidade docente, gestão escolar e pares, esta última normalmente medida através do nível socioeconômico escolar médio. É comum que variáveis individuais e socioeconômicas sejam incorporadas ao modelo tanto em estudos longitudinais quanto transversais. Do ponto de vista de equidade, geralmente se observa se há desigualdade de desempenho a partir de uma ou mais variáveis, por exemplo: entre alunos por grupos raciais, entre meninos e meninas ou por grupos de origens socioeconômicas distintas. Metodologicamente, medidas de equidade podem ser obtidas através da utilização de coeficientes aleatórios associados ao nível de interesse. TRAVITZKI *et al.* (2016) analisou equidade no ENEM no nível intraescolar. Finalmente, as variáveis individuais e familiares possibilitam o cálculo do efeito família. Normalmente, o efeito família é calculado conjuntamente com o efeito escola, sendo caracterizado pela proporção do erro do modelo no nível do aluno. Assim, dificilmente um estudo foca em calcular o efeito família isoladamente. O que se observa nas contribuições de avaliação educacional é uma variação em relação ao enfoque do trabalho e ao uso destas variáveis: a motivação pode ser mais relacionada à equidade ou a quanto estas variáveis estão relacionadas ao desempenho (que se trata do efeito família, propriamente dito).

As diferenças socioeconômicas impactam no desempenho dos alunos, refletindo as desigualdades que existem entre os grupos (RODRIGUES *et al.*, 2011). Em ALVES *et al.* (2016), pode-se observar diferenças sistemáticas entre as proficiências de alunos segmentados por raça, sexo ou nível socioeconômico, tanto para o 5º quanto para o 9º ano.

Os resultados em análises equitativas mostram que, embora haja muitas possibilidades de novos estudos, o retrato sobre as discrepâncias entre os desempenhos de alunos de diferentes classes sociais indica que existe um abismo entre o aprendizado de diferentes classes. Um debate mais acentuado sobre políticas públicas necessárias para atenuar esse hiato está fora do escopo da tese, mas acredita-se que o foco na educação básica, com facilidade de acesso e qualidade escolar seria uma abordagem apropriada com o objetivo de diminuir a iniquidade em nosso país

(FERRÃO *et al.*, 2018). Sob o ponto de vista internacional, sabe-se que a tentativa de manter sistemas educativos equitativos é um problema mundial. Em RIEGLE-CRUMB e GRODSKY (2010), pode-se encontrar uma boa discussão sobre equidade nos EUA, considerando turmas avançadas de Matemática.

De maneira mais prática, ao considerar o estudo da eficácia escolar focado na qualidade e mensuração do efeito escola, pode-se considerar que esta tarefa está relacionada à medição da importância da qualidade docente, da gestão, da infraestrutura escolar ou dos pares (FRANCO *et al.*, 2007, MENEZES-FILHO, 2007, SOARES e SÁTYRO 2008, MARIONI *et al.*, 2014, AMÉRICO e LACRUZ, 2017, OLIVEIRA e CARVALHO, 2018). Em FRANCO *et al.* (2007), uma investigação nacional sobre quais características escolares estão associadas à melhora do desempenho médio das escolas em Matemática, na 4^a série do Ensino Fundamental, medido pelo SAEB 2001, é apresentada. Aqui foram utilizadas muitas variáveis relacionadas a efeito escola, com três exceções: nível socioeconômico individual do aluno, gênero e repetência. Foi implementado um modelo hierárquico e as variáveis nível socioeconômico médio da escola e dever de casa tiveram uma importância muito grande na explicação da nota em Matemática. Em MENEZES-FILHO (2007), é utilizada uma modelagem transversal com dados nacionais do SAEB de 2003 para medir a proficiência dos alunos da 4^a e 8^a séries do ensino fundamental e do terceiro ano do ensino médio em Matemática. Foram implementados modelos de regressão múltipla separadamente para escolas públicas e privadas. Uma proposta interessante em SOARES (2003) foi analisar o ambiente da sala a partir de um questionário sobre comportamento do professor e da turma, interesse do docente em ajudar os alunos, entre outros itens. Este tipo de abordagem não é usual e isso ocorre, principalmente, pela dificuldade em mensurar essa informação em larga escala. No entanto, a preocupação com variáveis docentes que vão além de quantidade de horas aula e grau de instrução já foi considerada em 1970, onde se media a capacidade verbal do docente para estimar a capacidade verbal do aluno (BOWLES, 1970).

Outras contribuições têm algumas conclusões com ênfase no efeito família, onde relações com as escolaridades dos pais ou outras características familiares que podem contribuir para a proficiência, como a renda familiar, são buscadas (SOARES e COLLARES, 2006, FRANCO e MENEZES-FILHO, 2008, RIANI e RIOS-NETO, 2008, OLIVEIRA e SILVA, 2018). Pode-se considerar, também, características pessoais, como raça ou gênero, mas normalmente são análises agregadas ao efeito família. Em geral, a conclusão quase que majoritária, é que existem fatores importantes tanto em efeito escola, quanto em efeito família na proficiência dos alunos, muitas vezes com evidência do impacto familiar e social maior que o escolar. Dados do SAEB 2001 de

alunos da 8^a série de escolas públicas brasileiras são utilizados em SOARES e COLLARES (2006), que usam a regressão hierárquica para explicar a proficiência em Matemática e modelos estruturais com algumas variáveis endógenas como saída do modelo. Os fatores exógenos foram recursos econômicos da família, características do aluno e localização da escola, e se ressalta a importância de dividir o efeito família em fatores econômicos e culturais. Uma discussão mais ampla sobre o SAEB é proposta em FRANCO e MENEZES-FILHO (2008). Um modelo em painel é usado para estimar proficiência em Matemática na 4^a série com análises, também, em escolas públicas e privadas, separadamente. Os autores ressaltam a contribuição usando variáveis não usadas anteriormente em âmbito nacional para SAEB, identificando a importância das variáveis socioeconômicas.

A segmentação aqui realizada diz respeito apenas a algumas das conclusões centrais de cada pesquisa. Conforme já citado anteriormente, em muitas abordagens se encontram como fatores relevantes o nível socioeconômico familiar e a escolaridade dos pais e, em alguns casos, o efeito escola consegue ser medido com mais impacto, como, por exemplo, em ALBERNAZ *et al.* (2002), onde o nível socioeconômico é importante, mas, uma vez controlado, se comprova a importância escolar. Um indicador do efeito escola foi criado pelo INEP em 2005, ano da primeira Prova Brasil (GREMAUD *et al.*, 2007). A ideia era medir a influência escolar a partir de variáveis familiares e sociais fixas. Em ALVES e SOARES (2013), também se ressalta a importância no nível socioeconômico, além de um impacto positivo na infraestrutura escolar e negativo na complexidade, indicando que escolas com gestões menos complexas podem ter melhor desempenho. Uma diferença entre anos iniciais e finais do ensino fundamental foi encontrada em MATOS e RODRIGUES (2016), onde se verifica se a escola alcançou a meta do IDEB 2013. Para os primeiros anos escolares, a infraestrutura apresenta o maior impacto e, para os anos finais, o nível socioeconômico e a infraestrutura aumentam a probabilidade de a escola atingir a meta. Uma abordagem em painel que encontra forte relação entre infraestrutura escolar e distorção idade-série é feita em SOARES e SÁTYRO (2008). Embora a maioria das contribuições envolvendo modelos estatísticos em educação seja para mensurar o impacto do efeito escola e efeito família na proficiência, existem outras análises, como por exemplo, em ROCHA e FUNCHAL (2019), CURI e MENEZES-FILHO (2013) e OLIVEIRA *et al.* (2017), onde a avaliação de custos ou o impacto das mensalidades no desempenho são avaliadas. Em ROCHA e FUNCHAL (2019), uma abordagem de dados em painel para avaliar o desempenho médio da escola no ENEM entre 2012 e 2015 é usado. O objetivo é entender a relação custo *versus* desempenho. As variáveis de efeito escola e efeito família são usadas, além de variáveis de custos de cada escola. A amostra é composta por 32 escolas

estaduais de ensino médio do Espírito Santo. São implementadas regressões com efeito fixo e regressões quantílicas, ambas para notas em matemática e linguagem e códigos e encontra-se correlações entre o desempenho no ENEM e variáveis de custos, mas há um baixo poder explicativo dos custos no desempenho, indicando que não necessariamente a alocação de mais recursos em escolas públicas garante um desempenho melhor no ENEM. Já em CURI e MENEZES-FILHO (2013), encontra-se uma avaliação sobre a importância do valor da mensalidade para escolas no estado de São Paulo na nota do ENEM de 2006. Também são consideradas avaliações sobre efeitos de escola, família e pares no desempenho de alunos de escolas públicas e privadas. A média das provas objetivas no ENEM é usada como medida de saída. Em relação à mensalidade, chega-se à conclusão de que está positivamente relacionada ao desempenho. Duas amostras são utilizadas: uma com as mil melhores escolas do estado e a outra com as demais. Além das notas em matemática e linguagens e códigos, OLIVEIRA *et al.* (2017) também usa notas médias em ciências da natureza, ciências humanas e redação. Microdados do INEP e informações obtidas a partir de uma Secretaria de Estado da Educação de Goiás (SEDUCE-GO) são usados neste trabalho. A base de dados é composta por 356 escolas públicas de Goiás com pelo menos 50% de alunos participantes do ENEM 2014. A conclusão é que maior educação formal e assiduidade para professores não promovem aumento na proficiência em comparação ao custo. Escolas de tempo integral e com maior média de hora aula também não se mostram eficientes. Fora do contexto do estudo sobre custos e mensalidades, existem ainda outras aplicações. Os fatores determinantes na decisão de se matricular um filho em uma escola privada, segundo CURI e MENEZES-FILHO (2010), são a educação da mãe, a renda familiar, o custo da educação, a oferta relativa de escolas públicas ou privadas (ambas negativamente) e a Região de moradia. Uma avaliação sobre ampliação da jornada escolar é investigada em OLIVEIRA e TERRA (2016). Em CARVALHO e FIRPO (2014), uma comparação entre o desempenho escolar para escolas seriadas e por ciclo é feita, não encontrando diferença significativa no desempenho. Em TAVARES e PONCZEK (2018) é medido o efeito do aumento do salário dos professores na proficiência do aluno, usando notas de matemática e português do Sistema de Avaliação de Rendimento do Estado de São Paulo e foi possível encontrar evidências de que o aumento salarial melhora o rendimento dos alunos. Pode-se argumentar que o aumento não é suficiente para a realocação do tempo do docente ou que, como não é relacionado ao desempenho dos alunos, não há incentivo para que os docentes se aprimorem mais. Por outro lado, em MENEZES-FILHO e PAZELLO (2007), inicialmente não se observa evidência entre o salário do professor e a proficiência nas escolas públicas, mas encontra impacto positivo nas

escolas privadas. Posteriormente, é feita uma análise a partir do Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério (FUNDEF), que indica que este fundo aumentou o salário dos professores de escolas públicas e que este aumento estava relacionado com melhora na proficiência.

A utilização de indicadores educacionais em avaliação educacional é muito comum no Brasil (VITELLI *et al.*, 2018, LACRUZ *et al.*, 2019, MENEZES *et al.*, 2020). Sob o ponto de vista de estatística aplicada à educação, os indicadores educacionais podem ser considerados *proxies* para caracterizar as informações escolares, como gestão, qualidade docente e informações sobre os pares. Diversas contribuições usam estes indicadores para se medir efeito escola, como, por exemplo, (OLIVEIRA e TERRA, 2016, AMÉRICO e LACRUZ, 2017, OLIVEIRA *et al.*, 2017, OLIVEIRA e SILVA, 2018). É importante ter em mente que os indicadores não refletem a qualidade escolar, mas, sim, características particulares da escola. Além disso, políticas públicas podem se basear em indicadores, porém desenvolver políticas focadas unicamente neles pode refletir em um sistema educacional de qualidade, mas iníquo.

2.2 EFICÁCIA ESCOLAR: O EFEITO ESCOLA

Os estudos sobre avaliação educacional iniciaram a partir do começo do século XX e, segundo SILVA e GOMES (2018), desde então, diferentes concepções foram se constituindo e se modificando. No início do século, considerava-se que a avaliação educacional consistia em mensurar o conhecimento com medidas objetivas e, atualmente, as concepções existentes englobam uma visão mais ampla que envolve negociações entre os atores que buscam, a partir de métodos qualitativos e quantitativos, avaliar as mudanças nos desempenhos dos alunos. Pode-se considerar que os estudos relacionados à área de avaliação educacional tiveram um grande ponto de inflexão na segunda metade da década de 1960, onde COLEMAN (1966) identificou que as características socioeconômicas do aluno são mais associadas ao seu desempenho escolar do que as variáveis relacionadas à escola.

Estes resultados contradiziam o senso comum existente nas primeiras décadas do século XX (REYNOLDS *et al.*, 2014), onde se colocava a escola como um ambiente capaz de proporcionar aprendizado e influenciar na vida dos alunos. A partir dessa constatação, muitos estudos procuraram identificar os efeitos do que se denominou clima escolar nos desempenhos acadêmicos, relacionando-os com o contexto. BERKOWITZ *et al.* (2016) faz uma extensa síntese desses estudos e verifica a existência de uma grande variabilidade, tanto na definição e forma de mensuração do que viria a ser o clima escolar, quanto nos caminhos metodológicos percorridos nestes

estudos. No Brasil, um estudo recente realizado por MELO e MORAES (2019) demonstra que o clima escolar, avaliado através de questionários aplicados aos alunos, professores e gestores, é um fator protetivo em escolas que obtiveram um desempenho no Exame Nacional do Ensino Médio (ENEM) muito acima do esperado, em contextos socioeconômicos desfavoráveis. Outro conjunto de estudos da área de avaliação educacional é realizado para medir de forma mais objetiva, não apenas o clima escolar, mas o que se denomina de eficácia escolar, conceito que abrange fatores relacionados tanto ao clima escolar, quanto ao processo de gestão (PASSONE, 2019). Diversos estudos vêm sendo realizados a fim de identificar quais os fatores mais relacionados ao desempenho dos estudantes e possibilitar a mensuração da eficácia escolar.

Um retrato do estado da arte sobre a pesquisa em eficácia escolar é apresentado em REYNOLDS *et al.* (2014), onde aspectos históricos, problemas na área e avanços teóricos e metodológicos são discutidos. Além disso, é considerado que a pesquisa internacional em eficácia pode ser dividida em quatro fases: a primeira se iniciou após a publicação dos resultados de Coleman, onde um incômodo foi gerado e estudos foram iniciados a fim de se medir a importância escolar; a segunda fase, a partir do meio dos anos 80, é associada às abordagens multiníveis que são apropriadas a dados educacionais e possibilitaram uma melhor mensuração do efeito gerado pelas escolas no desempenho dos alunos; a terceira fase, no início dos anos 90, se caracteriza pelo interesse dos processos escolares. Finalmente, a partir dos anos 2000, a quarta fase se dá até hoje e é constituída pela internacionalização dos estudos, pelas fusões das abordagens e por avanços teóricos e metodológicos.

Entre as reflexões trazidas pelos estudos nestas décadas, passou a haver um consenso em relação ao conceito trazido por MORTIMORE (1998), de que uma escola é eficaz se os seus alunos melhoram o desempenho ao longo do tempo, independente de fatores socioeconômicos. Para se medir esta eficácia, a metodologia que passou a ser utilizada foi a de modelagem multinível, que permite identificar a contribuição das variáveis escolares no desempenho, distinguindo-as de variáveis individuais e socioeconômicas associadas aos alunos.

Esta metodologia tornou possível as mensurações separadas do efeito escola, que estima a influência da escola no desempenho do aluno, e do efeito família, que mede quanto os fatores individuais e socioeconômicos afetam este desempenho. No Brasil, os estudos sobre eficácia escolar tiveram início nos anos 90, momento em que a literatura internacional já estava na terceira fase. Apenas a partir deste período é que dados educacionais foram disponibilizados no país (FERRÃO *et al.*, 2018). Ao longo destas três décadas há resultados importantes referentes ao cálculo da eficácia escolar, tanto no que diz respeito ao efeito escola quanto à equidade no Brasil.

Em SOARES e ALVES (2013) fatores escolares relacionados ao desempenho dos alunos são discutidos, assim como a melhora da eficácia na escola pública brasileira. Os resultados apresentados neste estudo confirmam que escolas e municípios contribuem para o desempenho dos alunos, e identificam que há diferenças na proficiência ao levar em consideração as variáveis individuais e socioeconômicas.

Conforme debatido em KEEVES *et al.* (2005), os resultados relacionados ao efeito escola variam na literatura, com algumas contribuições atribuindo uma maior importância escolar em comparação a outras. Esta variação reflete diferenças nas escolhas, tanto em relação à modelagem quanto das variáveis escolares e socioeconômicas. Mensurar a gestão escolar ou a qualidade docente costuma representar um desafio no cálculo do efeito escola, uma vez que alguns atributos são difíceis de serem medidos. Indicadores educacionais costumam ser utilizados como *proxies* destas variáveis.

A utilização de indicadores em estudos sobre avaliação educacional vem sendo frequente no Brasil (VITELLI *et al.*, 2018, LACRUZ *et al.*, 2019, MENEZES *et al.*, 2020). A maioria desses indicadores são calculados a partir dos dados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O INEP “...é uma autarquia federal vinculada ao Ministério da Educação (MEC), responsável pelas avaliações e exames, pelas estatísticas e indicadores, e pela gestão do conhecimento e estudos educacionais” (INEP, 2021a) sendo, portanto, um importante aliado para a pesquisa educacional do país. Os indicadores educacionais podem ser usados como *proxies* para mensurar, por exemplo, gestão escolar, qualidade docente e influência dos pares.

Em contribuições relacionadas à eficácia escolar, os indicadores e outras medidas escolares, individuais e socioeconômicas são as variáveis independentes do modelo. A variável de saída, em geral, é uma medida de desempenho do aluno. Uma possibilidade de medida de desempenho pouco explorada na literatura é o Exame Nacional do Ensino Médio (ENEM), que foi criado com objetivo de avaliar o desempenho do estudante ao término da escolaridade básica e é utilizado como principal acesso ao ensino superior público no Brasil. Por não possibilitar análises longitudinais para medir o desempenho de um aluno e de haver um viés de seleção relacionado ao fato dos alunos escolherem se farão ou não o exame, há uma escassez sobre o tema na literatura. Por outro lado, o ENEM tem um caráter de mobilidade social que aumenta a sua importância para a sociedade (CARVALHES e RIBEIRO, 2019). Boa parte dos jovens brasileiros busca, no acesso ao ensino superior através do ENEM, uma alternativa para mudança de vida. Tendo em vista a importância deste exame, estudos recentes foram feitos para mensurar o efeito escola no ENEM (TRAVITZKI *et al.*, 2014, TRAVITZKI, 2017).

2.3 EFICÁCIA ESCOLAR: DESIGUALDADE DE DESEMPENHO E EQUIDADE

O debate sobre avaliação educacional e os impactos causados por variáveis familiares, individuais e escolares no desempenho dos alunos vem sendo conduzido por diferentes comunidades científicas nas últimas décadas. A inclusão das variáveis socioeconômicas neste cenário leva à reflexão sobre a comparação entre alunos de alto e baixo desempenho.

É fato que a origem social do aluno é relevante para a sua proficiência e o impacto causado por ela é denominado efeito família, conforme citado anteriormente. Desta forma, a avaliação educacional deve levar em consideração a desigualdade social, que não deve ser contemplada isoladamente através do capital econômico, que está relacionado à renda de cada família, mas também levando em consideração os capitais social e cultural (ALVES *et al.*, 2007; BONAMINO *et al.*, 2010). O capital social está relacionado às redes de relações sociais e à quantidade e à qualidade de recursos desta rede. Já o capital cultural está relacionado ao contexto familiar, como o grau de domínio da língua culta, as informações sobre o mundo escolar e as heranças culturais familiares, que podem facilitar o aprendizado escolar. Além disso, este capital também está presente na forma de bens culturais e no grau de investimento na carreira escolar, que depende do retorno a ser obtido no mercado de trabalho (BONAMINO *et al.*, 2010). Alunos provenientes de população de baixa renda, com pouca infraestrutura domiciliar e pais com baixa escolaridade se encontram em um cenário com muitos obstáculos a serem superados em comparação a alunos com mais recursos sob o ponto de vista econômico, social e cultural.

Conforme mencionado anteriormente, o acesso ao ensino superior é uma forma de mobilidade social, onde indivíduos buscam ascender socialmente através de melhores empregos e oportunidades (CARVALHAES e RIBEIRO, 2019). Sob o ponto de vista de políticas públicas educacionais, é fundamental que ações sejam adotadas a fim de minimizar situações de desvantagem, especialmente para estudantes oriundos de escolas públicas que possam não ter tido acesso a recursos e oportunidades em condições de igualdade a alunos de escolas privadas. No Brasil, foram criados alguns programas para tentar amenizar o problema da desigualdade em relação ao acesso ao ensino superior, porém, ainda há muita discussão sobre a sua efetividade (MOTTA e LOPES, 2012, ROSSETTO e GONÇALVES, 2015, WAINER e MELGUIZO, 2017). De acordo com SALES *et al.*, (2017), uma população com escolaridade mais alta, tende a ter um maior desenvolvimento social e, como consequência, sua intervenção social se torna mais ampla e efetiva.

As análises de qualidade e equidade em um sistema educacional são importantes e dizem respeito, de uma forma geral, à avaliação de quanto os ambientes escolares são produtivos e justos. A definição de equidade educacional tem variado na literatura e uma segmentação deste assunto pode ser encontrada em SIMIELLI (2015), que consolida o debate sobre equidade em três linhas diferentes: o foco no acesso aos recursos e processos; o foco nos resultados obtidos pelos alunos e a abordagem pluralista, onde recursos, processos e resultados são avaliados a partir de uma visão mais global. Esta linha mais plural está em concordância com a opinião da comissão europeia, que caracteriza a equidade como uma medida de como os indivíduos podem tirar vantagem da educação em termos de igualdade de oportunidades, acesso, tratamento e resultados (CEC, 2006).

HIPPE *et al.* (2016), afirma que a qualidade dos professores é fundamental para o sucesso nos resultados relacionados a equidade e qualidade e que as políticas públicas devem levar em consideração os diferentes contextos da população, como, por exemplo, socioeconômico, político, geográfico, entre outros. Em geral, considera-se que políticas públicas relacionadas à equidade devem ser elaboradas com foco na primeira infância. Apesar da ênfase no desenvolvimento de políticas públicas nos anos iniciais de ensino, estes autores consideram que a diferença de desempenho entre alunos de grupos diferentes pode aumentar na medida em que os anos de educação se acumulam.

Um estudo longitudinal feito com dados de estudantes norte americanos relacionou o agrupamento escolar por habilidades à reprodução de desigualdades sociais (BUTTARO e CATSAMBI, 2019). Os resultados demonstraram que as lacunas de aproveitamento entre os alunos agrupados e desagrupados aumentaram a cada ano de estudo. Embora essa segmentação de alunos por habilidade não seja usual no sistema educacional brasileiro, este estudo corrobora a opinião de que diferenças entre grupos podem aumentar no decorrer dos anos, o que desperta o interesse de como seriam os resultados em termos de equidade e qualidade para um público de alunos nos anos finais da educação básica, faixa estudada neste capítulo.

Em um país com tanta desigualdade social como o Brasil, é fundamental o estudo sobre equidade no âmbito educacional. A ideia de equidade utilizada neste trabalho está relacionada a um ambiente onde fatores individuais e socioeconômicos não interfiram no aprendizado do aluno. Dessa forma, este trabalho está situado na abordagem pluralista, avaliando o contexto socioeconômico frente aos resultados obtidos pelos alunos. Em ambientes iníquos, o desempenho pode ser mais bem explicado a partir de variáveis individuais e familiares, em comparação a variáveis escolares. Conforme

mencionado em SIMIELLI (2017), “Para que haja equidade, os recursos e processos devem ter impacto sobre o desempenho do aluno”.

A educação básica no Brasil engloba educação infantil, ensino fundamental e ensino médio. Normalmente, o debate sobre equidade se dá no contexto do ensino fundamental e as desigualdades já são conhecidas, tanto em relação às escolas públicas e privadas (BARBOSA e FERNANDES, 2001), quanto em relação aos municípios (ALVES *et al.*, 2016). No entanto, considerando a importância que o acesso ao ensino superior no Brasil pode ter na mobilidade social (CARVALHAES e RIBEIRO, 2019), e sabendo que as notas obtidas no ENEM são utilizadas para conseguir uma vaga em uma instituição pública a partir do Sistema de Seleção Unificada (SISU) ou ainda uma bolsa de estudos pelo Programa Universidade para Todos (ProUni), entre outros fins, entende-se a relevância do ENEM, tanto para os jovens brasileiros, quanto para os responsáveis pelas políticas voltadas para o ensino médio.

Apesar disso, há uma escassez de estudos recentes que mensurem a iniquidade devido à raça e sexo entre a população que está terminando o ensino médio ou com possibilidade de se candidatar ao ensino superior. Entre estudos anteriores à reformulação realizada no ENEM em 2009, encontra-se o de DIAZ (2010) que investiga os resultados do ENEM de 2005 e encontra uma desigualdade entre os resultados obtidos pelos indivíduos que realizaram o ENEM na idade correta em comparação aos mais velhos. Outro estudo, mais recente, é o de TRAVITZKI *et al.* (2016), que analisa desigualdades sociais para a população pré-universitária em diferentes estados brasileiros a partir do ENEM, e encontra que o progresso educacional segue sensível a fatores de natureza sócio espacial e racial.

A discussão a respeito da equidade no âmbito da educação que é afetada pelas desigualdades sociais não pode estar separada da necessidade de que o sistema público de educação desenhe políticas públicas que promovam a qualidade escolar com vistas a compensar as iniquidades. Esta qualidade tem sido mensurada através de avaliações sobre condições de acesso, trajetória escolar e aprendizado, conforme citado em ALVES *et al.* (2016).

3. MATERIAIS E MÉTODOS

Neste capítulo são descritas as bases de dados utilizadas neste trabalho e as modelagens propostas. Como há três propostas centrais, a saber, cálculo do efeito a escola, desigualdade de desempenho entre alunos de alto e baixo rendimento e equidade, apresentam-se as descrições para cada etapa da pesquisa.

3.1 BASES DE DADOS

A seguir, serão descritas as bases de dados utilizadas para cada etapa do trabalho.

3.1.1 BASE DE DADOS: EFEITO ESCOLA

Trata-se de um estudo transversal, com a população alvo constituída pelos alunos das escolas da Região Sudeste, residentes em municípios com população entre 50 e 500 mil habitantes (IBGE, 2021), que realizaram o ENEM em 2017. O recorte regional é definido pelo fato da Região Sudeste atingir a maior porcentagem (36%) de inscritos (MEC, 2021) e a exclusão dos municípios menores e maiores visa evitar a comparação entre escolas de municípios muito discrepantes em termos populacionais. Conforme mencionado anteriormente, cidades entre 50 e 500 mil habitantes podem ser consideradas como cidades médias (SPOSITO, 2001, WENDEL, 2010).

O desempenho dos estudantes é medido pelas notas em matemática (matemática e suas tecnologias), representadas aqui como variável dependente (Tabela 3.1). A escolha pelo exame de Matemática é por conta do sucesso neste desempenho ser usualmente uma medida atribuída à escola (FRANCO *et al.*, 2005, RODRIGUES *et al.*, 2011, OLIVEIRA e CARVALHO, 2018). Em relação às demais variáveis apresentadas nesta Tabela 3.1, a renda familiar, o número de pessoas na casa e a escolaridade da mãe são as variáveis socioeconômicas escolhidas que, adicionalmente com idade, sexo e raça serão vistas como variáveis associadas ao efeito família. A variável renda familiar, medida no questionário em 17 intervalos, foi utilizada como contínua a partir do ponto médio de cada classe. Para a primeira categoria, o valor atribuído foi zero e, para a última, o valor de R\$ 20.145,50, considerando que foi mantida a amplitude intraclass igual às demais.

Já na Tabela 3.2, encontram-se os indicadores educacionais associados à escola que serão usados como variáveis independentes na modelagem para o cálculo do efeito escola. O indicador de complexidade está relacionado à gestão escolar; as horas-aula

diárias e alunos por turma refletem a questão organizacional da escola, além das taxas de distorção idade-série e de aprovação e abandono. Há variáveis que são proxies para a qualidade docente, como adequação, esforço e regularidade, embora não estejam relacionadas diretamente à didática ou a qualidade da aula. O nível socioeconômico médio é um valor médio relacionado aos alunos da escola e fornece uma medida do padrão de vida das famílias.

Para a construção do banco de dados, são selecionadas as informações dos alunos e de suas famílias (Tabela 3.1), obtidas no questionário do ENEM (INEP, 2021c), e as informações das escolas (Tabela 3.2), retiradas dos indicadores educacionais (INEP, 2021b), produzidos a partir da base de dados do Censo Escolar do ano de 2017, disponíveis no portal do INEP.

Informações provenientes destas duas bases são unificadas em um único banco de dados através do código da escola associado a cada aluno. Dessa forma, a base de dados final do estudo, contemplou as informações de 218.745 alunos, matriculados em 171.733 de escolas públicas e 47.012 de escolas privadas com 22 variáveis independentes associadas a características individuais, familiares e escolares dos candidatos e uma variável dependente, a nota de Matemática. Uma análise sobre os dados faltantes desta base de dados se encontra no Apêndice A.

Tabela 3.1 – Descrição das variáveis dependentes e independentes relacionadas às características individuais e familiares

Variáveis	Descrição
Dependente Notas no ENEM 2017	Nota dos alunos em Matemática na prova do ENEM de 2017 podendo variar de 0 a 1000
Variáveis associadas ao Efeito Família:	
Independentes Escolaridade da mãe: Nunca estudou, Parou antes do 5º ano, Ensino Fundamental , Ensino Médio, Ensino Superior, Pós Graduação, Não sei	Sendo, 1 se sim e 0 caso contrário
Renda familiar	Ponto médio de cada classe
Número de pessoas	Número de pessoas que residem no seu domicílio
Idade	Idade do aluno
Sexo	Sendo, 1 feminino e 0 masculino.
Raça	Sendo, 1 branco e 0 caso contrário.

Fonte: Própria autoria, 2021.

Tabela 3.2 – Descrição das variáveis independentes relacionadas ao efeito escola

Variáveis associadas ao Efeito Escola:	
Independentes	
Alunos por turma	Número médio que corresponde à divisão do número de matrículas pelo número de turmas da escola.
Horas-aula diária	Número médio de horas-aula diária.
Taxa de distorção idade-série	Taxas de distorção idade-série.
Taxa de aprovação	Taxa de aprovação dos discentes.
Taxa de abandono	Taxa de abandono dos discentes.
Indicador de regularidade do corpo docente	Observação da permanência dos professores nas escolas.
Indicador de esforço docente	Percentual de docentes por nível de esforço. As características consideradas para mensurar o nível de esforço são: número de escolas, turnos, alunos e etapas que o professor leciona. As escolas são classificadas do nível 1 (até 25 alunos, um único turno, escola e etapa) ao nível 6 (mais de 400 alunos, nos três turnos, em duas /três escolas e etapas). Utiliza-se como variável explicativa a soma dos percentuais dos níveis 4, 5 e 6.
Indicador de complexidade de gestão da escola	Características consideradas para mensurar o ICG: porte das escolas; número de turnos, etapas e modalidades oferecidas. O nível 1 representa a menor complexidade de gestão escolar, enquanto o nível 6, a maior.
Nível socioeconômico dos discentes	Média aritmética simples da medida nível socioeconômico dos alunos, no intervalo definido com base em bens domésticos, renda, contratação de serviços e nível de escolaridade da família, a fim de possibilitar a visão geral do padrão de vida dos alunos. Para melhor descrever o nível socioeconômico das escolas, foram criados, a partir da análise de cluster por método hierárquico, seis grupos, de forma que o Grupo 1 congrega as escolas com Inse mais baixo e o Grupo 6, com mais alto.
Adequação de formação docente	Vai do grupo 1 (licenciatura ou bacharelado com complementação pedagógica na disciplina que leciona) ao grupo 5 (não possui curso superior). Considera-se como variável explicativa o percentual de disciplinas, em cada etapa, ministradas por professores do nível 1.

Fonte: Própria autoria, 2021.

3.1.2 BASE DE DADOS: DESIGUALDADE DE DESEMPENHO

A base de dados é a mesma da seção 3.1.1, ou seja, com um total 218.745 alunos, sendo 171.733 de escolas públicas e 47.012 de escolas privadas. No entanto, no presente capítulo variáveis de infraestrutura da escola (INEP, 2021c) são levadas em consideração.

Sendo assim, há 35 variáveis independentes e uma variável dependente que são associadas a diferentes modelos, descritos na próxima subseção. As variáveis são relacionadas a informações individuais e socioeconômicas dos alunos, além de informações escolares, relacionadas a gestão, corpo docente, pares e infraestrutura escolar.

Os estados pertencentes à Região Sudeste do Brasil são Rio de Janeiro, São Paulo, Minas Gerais e Espírito Santo. Em torno de 53% dos alunos analisados são de São Paulo, seguidos por 22% de Minas Gerais, 16% do Rio de Janeiro e 9% do Espírito Santo. Considerando a dependência administrativa, a distribuição por estado é similar à distribuição geral. Tem-se que 56% dos alunos de escolas privadas são de São Paulo, 18% de Minas Gerais, 17% do Rio de Janeiro e 9% do Espírito Santo. Já em relação às escolas públicas, 52% dos alunos são de São Paulo, 24% de Minas Gerais, 15% do Rio de Janeiro e 9% do Espírito Santo.

As variáveis individuais, socioeconômicas, bem como os indicadores do INEP, estão descritos nas Tabelas 3.1 e 3.2 da seção anterior. Na Tabela 3.3, podem-se encontrar as variáveis de infraestrutura utilizadas neste capítulo.

Tabela 3.3 – Descrição das variáveis independentes relacionadas ao efeito escola – infraestrutura

Efeito Escola - Infraestrutura	
Independentes	
Localização	Sendo, 1 para escola urbana e 0 para rural
Sala Diretoria, Sala Professor, Laboratório de Informática, Quadra de esportes, Cozinha, Biblioteca,	Sendo, 1 se tem e 0 caso contrário
Sala de Leitura, Refeitório, Computador, Internet, Banda Larga, Alimentação	

Fonte: Própria autoria, 2021.

3.1.3 BASE DE DADOS: EQUIDADE

A base de dados é composta por alunos de escolas públicas que pertencem a municípios com população entre 50 e 500 mil habitantes de todo o Brasil que realizaram o Exame Nacional do Ensino Médio (ENEM), no ano de 2017. A escolha de contemplar estudantes do país como um todo ocorreu pela motivação de apresentar uma análise intramunicipal para equidade. Para isso, foram utilizadas variáveis individuais, familiares e, diferentemente das análises anteriores, uma única variável medida ao nível da escola. Como atributos individuais, tem-se a raça e o sexo dos alunos. O nível socioeconômico familiar foi medido através da escolaridade da mãe e da renda familiar. No total, a base de dados contemplou informações sobre 375.670 alunos.

A escolaridade materna assume, no estudo, o valor 1 para mães que tenham escolaridade a partir do ensino médio e 0 caso contrário. Já a raça teve os alunos brancos por base, em relação às demais raças.

3.2 MODELOS ESTATÍSTICOS

Nesta seção apresentam-se as modelagens utilizadas. Para mensuração do efeito escola, são propostos modelos multiníveis (aluno, escola). Para análise de desempenho de alunos de alto e baixo rendimento, são utilizadas regressões quantílicas e modelos multiníveis. Finalmente, para a análise intramunicipal de equidade, modelos multiníveis de três níveis (aluno, escola, município) são usados.

3.2.1 MODELO CONCEITUAL: EFEITO ESCOLA

O modelo aqui proposto pretende decompor o desempenho em Matemática dos estudantes das escolas públicas e privadas para avaliar o quanto ele pode ser atribuído à escola. A Figura 3.1 ilustra a modelagem proposta: no primeiro nível do modelo encontram-se as variáveis associadas ao efeito família, que são medidas para cada estudante; no segundo nível, têm-se as variáveis para o cálculo do efeito escola, que são compostas pelos indicadores educacionais descritos anteriormente.

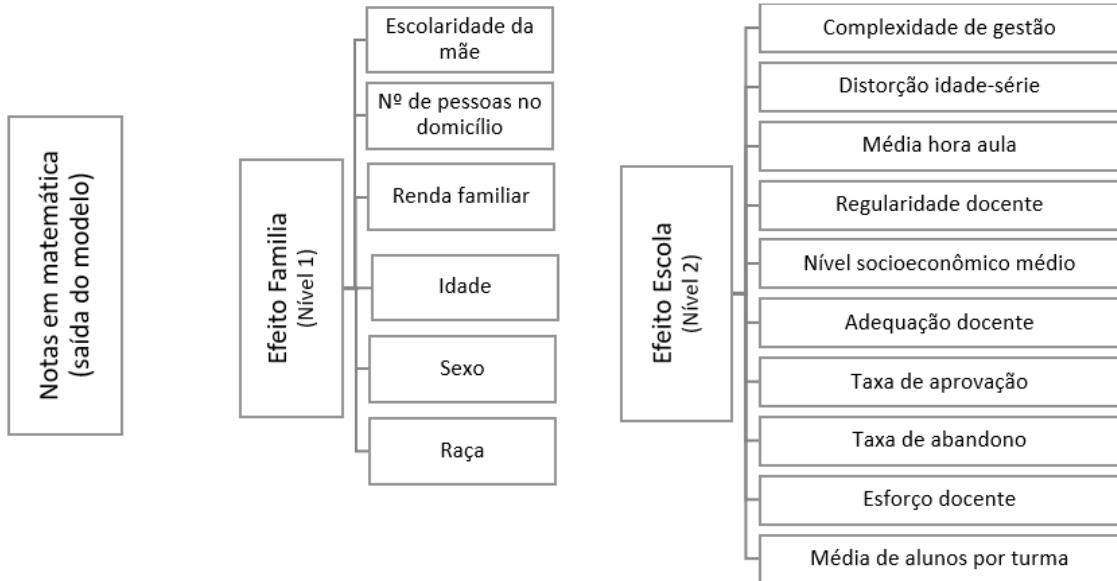


Figura 3.1 – Diagrama das variáveis associadas a cada nível do modelo

Fonte: Própria autoria, 2021.

3.2.2 MODELO ESTATÍSTICO: EFEITO ESCOLA

Para se estudar o desempenho dos alunos em Matemática no ENEM, a partir de variáveis individuais, socioeconômicas e escolares, o modelo multinível é o indicado. Uma vez que esta modelagem contempla a hierarquia presente no conjunto de dados, onde alunos estão agrupados em escolas.

O modelo de regressão multinível proposto é o de intercepto aleatório com dois níveis, aluno e escola. Nas etapas da modelagem são ajustados modelos diferentes para alunos de escolas públicas e privadas. O modelo formal está descrito na equação (3.1), de acordo com HOX (2010).

$$Y_{i,j} = \gamma_{00} + \sum_{p=1}^{12} \gamma_{p0} X_{pij} + \sum_{q=1}^{10} \gamma_{0q} Z_{qj} + u_{0j} + \varepsilon_{ij} \quad (3.1)$$

onde:

i é um índice para os estudantes;

$Y_{i,j}$ é a proficiência em matemática para o aluno i da escola j;

γ_{00} é o intercepto da regressão;

X_{pij} , p = 1, ..., 12 são as 12 variáveis exploratórias no nível do aluno, sendo 7 delas dummies associadas a escolaridade da mãe;

Z_{qj} , q = 1, ..., 10 são as 10 variáveis exploratórias no nível da escola;

u_{0j} são os erros no nível da escola;

ε_{ij} são os erros no nível do aluno.

A opção pelo ajuste de dois modelos, sendo um deles para os alunos das escolas públicas e o outro para os das escolas privadas, foi porque tanto a carreira docente quanto as condições de gestão estão muito mais associadas a políticas públicas educacionais nas escolas públicas em comparação as escolas privadas, já que as escolas privadas possuem maior independência no sentido de obter melhora nos indicadores.

A metodologia utilizada para a estimativa dos parâmetros em modelos multiníveis é através do método da máxima verossimilhança (HOX, 2010), implementada, no presente estudo, pelo pacote lme4 do software estatístico R (BATES, 2015).

Para a avaliação do efeito escola, uma medida proporcional pode ser calculada através do índice de correlação intraclasse ICC (HOX, 2010) conforme equação (3.2):

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{\varepsilon}^2} \quad (3.2)$$

onde σ_{u0}^2 é a variância dos resíduos do nível escola, e σ_{ε}^2 é a variância dos resíduos do nível aluno.

Quanto maior a variabilidade das notas em decorrência do efeito escola, maior a diferença entre as escolas. Isto significa que há instituições que contribuem mais para o desempenho dos alunos em comparação a outras, aumentando a relevância escolar no processo. O cálculo do efeito escola, tanto em estudos longitudinais quanto em transversais, produzem resultados variados e isso é dependente da identificação do modelo, ou seja, das variáveis escolhidas (KEEVES, HUNGI e AFRASSAT, 2005).

Embora a modelagem multinível com controle de variáveis individuais e socioeconômicas seja a mais adequada neste contexto, apresenta-se, para efeito de comparação, um procedimento de seleção de variáveis. O algoritmo utilizado é o Lasso (HASTIE *et al.*, 2001, HASTIE *et al.*, 2015, MASCI *et al.*, 2016) e o objetivo é verificar se as variáveis consideradas significativas pelo modelo da equação (3.1) coincidem com as variáveis selecionadas pelo Lasso.

3.2.3 MODELO ESTATÍSTICO: DESIGUALDADE DE DESEMPENHO

Para alcançar um dos objetivos centrais, que é a avaliação de alunos de escolas públicas e privadas com desempenhos alto e baixo no ENEM 2017, é utilizada a modelagem de regressão quantílica (KOENKER e BASSETT, 1978, KOENKER e HALLOK, 2001). Esta abordagem estima modelos por quantil, o que permite uma inferência a partir das características das estimativas para diferentes desempenhos dos alunos. Define-se desempenho alto como acima do quantil 75 e baixo como abaixo do quantil 25. A análise do desempenho em torno da média será feita pela regressão quantílica no quantil 50, assim como através da estimativa de um modelo multinível. A modelagem multinível, permite uma comparação e avaliação da modelagem quantílica no quantil 50.

No modelo de regressão quantílica, a ideia central é que, quando a nota do estudante corresponde exatamente ao quantil τ , isto significa que o seu desempenho é maior que o de uma proporção τ de alunos e pior que uma proporção $(1 - \tau)$ (KOENKER e HALLOK, 2001). O modelo linear quantílico para o τ^{th} quantil é dado por:

$$Y_i | X_{pi}, Z_q, \tau = \beta_0(\tau) + \sum_{p=1}^{12} \beta_p(\tau) X_{pi} + \sum_{q=1}^{23} \delta_q(\tau) Z_q + e_i(\tau) \quad (3.3)$$

onde:

i é um índice para os estudantes;

$Y_i | X_{pi}, Z_q, \tau$ é a proficiência em matemática para o estudante i , dadas as variáveis individuais e escolares X_{pi} e Z_q e o quantil τ ;

$\beta_0(\tau)$ é o intercepto da regressão para cada quantil τ ;

X_{pi} , $p = 1, \dots, 12$ são as 12 variáveis exploratórias individuais e familiares;

Z_q , $q = 1, \dots, 23$ são as 23 variáveis exploratórias escolares;

$e_i(\tau)$ são os erros para cada quantil τ ;

Computacionalmente, as ferramentas utilizadas para este modelo podem ser descritas em KOENKER, (2010). A partir do modelo presente na equação (3.3) dois modelos empíricos são estimados, um para alunos de escolas públicas e outro para os alunos das escolas privadas.

Ao variar o valor do quantil τ , pode-se analisar o impacto das variáveis individuais, familiares e escolares, a partir da modelagem quantílica, no desempenho de estudantes de baixo e alto desempenho.

Dados escolares possuem natureza hierárquica (HOX, 2010). A fim de levar isto em consideração, apresentam-se dois modelos multiníveis de dois níveis com intercepto aleatório, propostos com o objetivo de prever a proficiência em Matemática para escolas públicas e privadas. Os alunos estão no primeiro nível, enquanto que as escolas se encontram no segundo nível. Foram estabelecidos dois modelos diferentes, um para escolas públicas e outro para privadas. O modelo a seguir inclui novas variáveis ao efeito escola em comparação com o modelo da equação 3.1.

$$Y_{ij} = \gamma_{00} + \sum_{p=1}^{12} \gamma_{p0} X_{pij} + \sum_{q=1}^{23} \gamma_{0q} Z_{qj} + u_{0j} + \varepsilon_{ij} \quad (3.4)$$

onde:

i é um índice para os estudantes;

Y_{ij} é a proficiência em matemática para o aluno i da escola j ;

γ_{00} é o intercepto da regressão;

X_{pij} , $p = 1, \dots, 12$ são as 12 variáveis exploratórias no nível do aluno, sendo 7 delas *dummies* associadas a escolaridade da mãe;

Z_{qj} , $q = 1, \dots, 23$ são as 23 variáveis exploratórias no nível da escola;

u_{0j} são os erros no nível da escola;

ε_{ij} são os erros no nível do aluno.

Metodologicamente, a diferença da equação (3.4) para a equação (3.1) é a adição das variáveis de infraestrutura escolar. Computacionalmente, as ferramentas utilizadas para este modelo podem ser descritas em BATES *et al.*, 1996. Enquanto o modelo da equação (3.3) permite avaliar os coeficientes estimados para alunos de desempenho baixo e alto, o modelo da equação (3.4) permite avaliar o comportamento médio dos estudantes, assim como uma comparação com a modelagem quantílica no quantil 50. A discussão aqui apresentada terá ênfase na desigualdade entre os alunos de escolas públicas e privadas, que será constatada a partir da diferença entre as notas obtidas, nas estimativas dos modelos e nas análises descritivas.

3.2.4 MODELOS ESTATÍSTICOS: EQUIDADE

A equidade dos municípios é avaliada a partir de modelos multiníveis onde a variável dependente será a nota de Matemática do aluno no ENEM 2017, as variáveis individuais e socioeconômicas do aluno estão em um primeiro nível, da escola em um segundo

nível e do município em um terceiro nível. As avaliações se dão a partir de cinco modelos com diferentes variáveis independentes, representadas pelos coeficientes aleatórios, permitindo avaliar se o município é mais ou menos equitativo que a média.

É avaliado se a equidade diminui ou aumenta na medida em que as variáveis socioeconômicas são adicionadas ao modelo, estratégia comum na literatura e algo parecido ao feito em FERRÃO *et al.* (2018) para conhecimento prévio do aluno. Em termos de modelos, a ideia é similar às avaliações propostas em ALVES *et al.* (2016), TRAVITZKI *et al.* (2016) e TRAVITZKI (2017). Da mesma forma que em ALVES *et al.* (2016), não é adicionada nenhuma variável no nível do município, uma vez que o objetivo é avaliar apenas os coeficientes aleatórios das variáveis onde a equidade é estudada. As variáveis utilizadas são similares as discutidas em TRAVITZKI *et al.* (2016) e TRAVITZKI (2017), inclusive com o nível socioeconômico escolar médio como única variável no nível escola. Os modelos propostos são apresentados a seguir e descritos de forma resumida na Tabela 3.4.

- (i) Modelo 1 (M1): análise de efeitos fixos, através de modelagem multinível, com o primeiro nível representando apenas as variáveis individuais, o segundo nível representando as escolas com o nível socioeconômico médio escolar e o terceiro nível representando os municípios aos quais as escolas pertencem;
- (ii) Modelos 2 e 3 (M2 e M3): cálculo da equidade a partir de dois novos modelos, com análise dos coeficientes aleatórios para raça e sexo, respectivamente;
- (iii) Modelos 4 e 5 (M4 e M5): inserção de variáveis socioeconômicas nos modelos 2 e 3 a fim de observar se as equidades perdem ou ganham magnitude (tanto em relação a efeitos fixos quanto a efeitos aleatórios);

Tabela 3.4 – Variáveis independentes e coeficientes aleatórios por modelo, sendo escolaridade: escolaridade materna; renda: renda familiar; NSE: nível socioeconômico médio escolar.

Modelos	Variáveis Independentes		Coeficiente Aleatório (município)
	Nível 1	Nível 2	
M1	Raça, Sexo	NSE	-
M2	Raça, Sexo	NSE	Raça
M3	Raça, Sexo	NSE	Sexo
M4	Raça, Sexo, Escolaridade, Renda	NSE	Raça
M5	Raça, Sexo, Escolaridade, Renda	NSE	Sexo

Fonte: Própria autoria, 2021.

As análises são apresentadas tanto em relação aos efeitos fixos quanto aos efeitos aleatórios. A verificação de equidade para cada variável se dá a partir da análise dos coeficientes aleatórios, onde se pode verificar o quanto um município está acima ou abaixo da média em relação à raça e sexo. A descrição formal do modelo M5 está apresentada a seguir:

$$Y_{ijk} = \beta_0 + B_1^T X_{1ijk} + \beta_{2k} x_{2ijk} + \beta_3 x_{3ijk} + \beta_4 x_{4ijk} + \beta_5 x_{50jk} + e_{ijk} + r_{0jk} + u_{00k}$$

$$\beta_2 = \gamma_{200k} + u_{200k} \quad (3.5)$$

onde:

Y_{ijk} é o desempenho do aluno i, da escola j, que pertence ao município k;

B_1 é o vetor de efeitos fixos referentes a raça;

X_{1ijk} é o vetor que corresponde a raça do aluno i, da escola j, que pertence ao município k;

β_{2k} é o coeficiente aleatório do sexo, que varia entre os municípios;

x_{2ijk} é o sexo do aluno i, da escola j, que pertence ao município k;

γ_{200k} é o efeito fixo do sexo;

u_{200k} é o efeito aleatório do sexo;

β_3 é o efeito fixo referente a escolaridade materna;

x_{3ijk} é a escolaridade materna do aluno i, da escola j, que pertence ao município k;

β_4 é o efeito fixo da renda familiar;

x_{4ijk} é a renda familiar do aluno i, da escola j, que pertence ao município k;

β_5 é o efeito fixo do nível socioeconômico escolar médio;

x_{50jk} é o nível socioeconômico escolar médio da escola j, que pertence ao município k;

e_{ijk} é o efeito aleatório do nível dos alunos;

r_{0jk} é o efeito aleatório do nível das escolas;

u_{00k} é o efeito aleatório do nível dos municípios.

O método utilizado para o ajuste dos modelos é implementado, no presente estudo, pelo pacote *lme4* do software estatístico R (BATES *et al.*, 2015) e a avaliação da

significância dos coeficientes aleatórios é realizada conforme o descrito em KUZNETSOVA *et al.* (2017).

No próximo capítulo, são apresentados os resultados referentes às estimativas dos efeitos fixos e aleatórios dos cinco modelos multiníveis definidos na seção anterior, bem como a discussão sobre equidade referente aos municípios brasileiros.

4. RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados os resultados referentes ao cálculo do efeito escola para o ENEM, levando em consideração os indicadores educacionais como variáveis escolares; da análise de desempenho entre alunos de alto e baixo rendimento e do estudo sobre equidade intramunicipal.

4.1 RESULTADOS E DISCUSSÕES: EFEITO ESCOLA

Na primeira etapa, apresenta-se a distribuição dos indicadores atribuídos às escolas públicas e privadas, permitindo, com isso, traçar um perfil das escolas aqui analisadas. Na Figura 4.1, dois desses indicadores são relacionados aos docentes, que são adequação e esforço, e dizem respeito, respectivamente, ao percentual de disciplinas ministradas por professores com qualificação e percentual de professores com excesso de carga de trabalho. Percebe-se que os perfis das escolas públicas e privadas são semelhantes, com 25% das escolas com mais de 75% de docentes com licenciatura ou complementação pedagógica. Quanto ao esforço docente, existe um percentual alto das escolas públicas com professores com uma carga elevada de trabalho, sendo *outliers* as escolas cuja porcentagem de docentes nesta condição seja menor do que 40%. Já no *boxplot* relacionado às escolas privadas, este percentual possui uma maior variabilidade, o que indica mais escolas com um percentual menor de docentes com carga elevada.

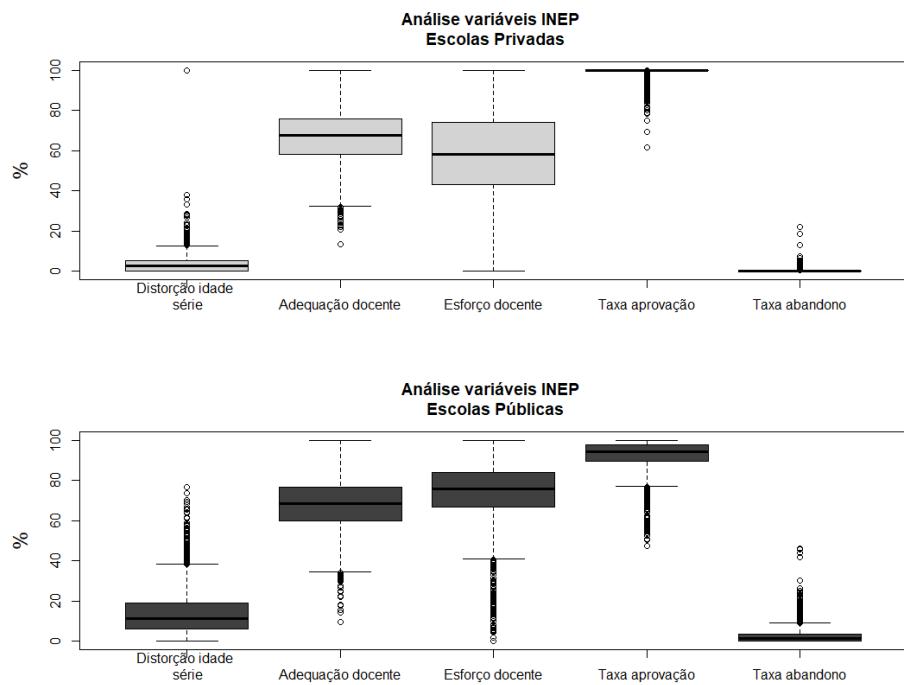


Figura 4.1– Comparação entre variáveis percentuais para escolas públicas e privadas
Fonte: Própria autoria, 2021.

Os outros três indicadores desta Figura 4.1 são relacionados à escola: porcentagem de alunos com distorção idade-série e as taxas de aprovação e abandono. Em relação ao indicador que mede a distorção idade série, verifica-se que ele apresenta maior variabilidade nas escolas públicas, onde sua mediana equivale a uma taxa de distorção próxima a 11%, enquanto esse número cai para menos de 3% nas escolas privadas. Isto pode estar relacionado a uma taxa de aprovação ligeiramente menor em comparação às escolas privadas, aumentando a distorção de idade. As taxas de abandono são similares em ambas as dependências administrativas, com uma variabilidade ligeiramente maior nas escolas públicas.

Outros dois indicadores são apresentados na Figura 4.2. Um deles mede a complexidade de gestão que está associada ao porte da escola, ou seja, ao número de turnos, turmas e modalidades ofertadas (do nível 1, menor complexidade de gestão, ao nível 6, maior complexidade de gestão) e o nível socioeconômico das escolas que leva em consideração bens domésticos dos alunos, renda, nível de educação familiar, entre outras coisas, conforme visto na Tabela 3.2 (do nível 1, escolas com menor nível socioeconômico ao nível 6 com maior nível socioeconômico). Percebe-se que, enquanto apenas 7% das escolas privadas são classificadas com níveis 5 e 6 de complexidade, 37% das escolas públicas estão nestes níveis. Além disso, a maioria das escolas nas duas dependências administrativas possuem complexidade de gestão no nível 4.

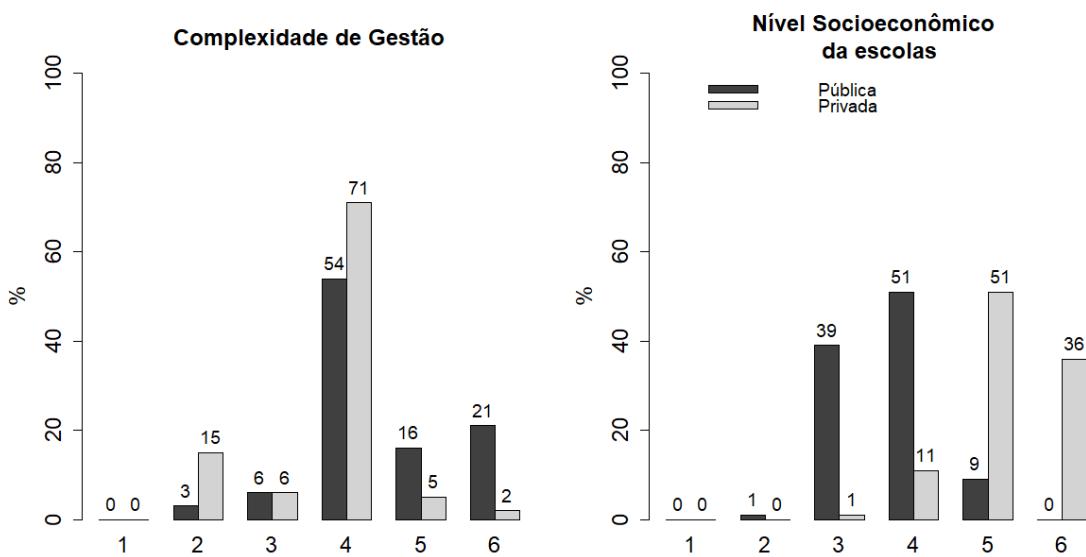


Figura 4.2 – Complexidade de gestão e nível socioeconômico médio dos candidatos ao ENEM 2017

Fonte: Própria autoria, 2021.

Em relação ao nível socioeconômico médio, ainda na Figura 4.2, verifica-se, como seria de se esperar, que 87% das escolas privadas estão nos níveis 5 e 6, enquanto 90% das escolas públicas possui nível socioeconômico entre os níveis 3 ou 4. Assim, as escolas públicas lidam com mais complexidade na gestão e com alunos com níveis socioeconômico menores.

O objetivo dos indicadores educacionais é medir, sob diferentes prismas, a qualidade de um determinado atributo ou ação escolar. As Figuras 4.1 e 4.2 indicam que há condições desfavoráveis para os alunos de escolas públicas, tanto ao considerar o indicador que reflete a origem social dos alunos, quanto ao tomar o indicador relacionado a uma maior taxa de alunos com defasagem idade-série, além de uma maior carga de trabalho dos docentes. Todas estas desvantagens podem influenciar nas diferenças entre os desempenhos, constatadas na Figura 4.3, através da distribuição das notas em Matemática dos alunos de escolas públicas e privadas no ENEM 2017.

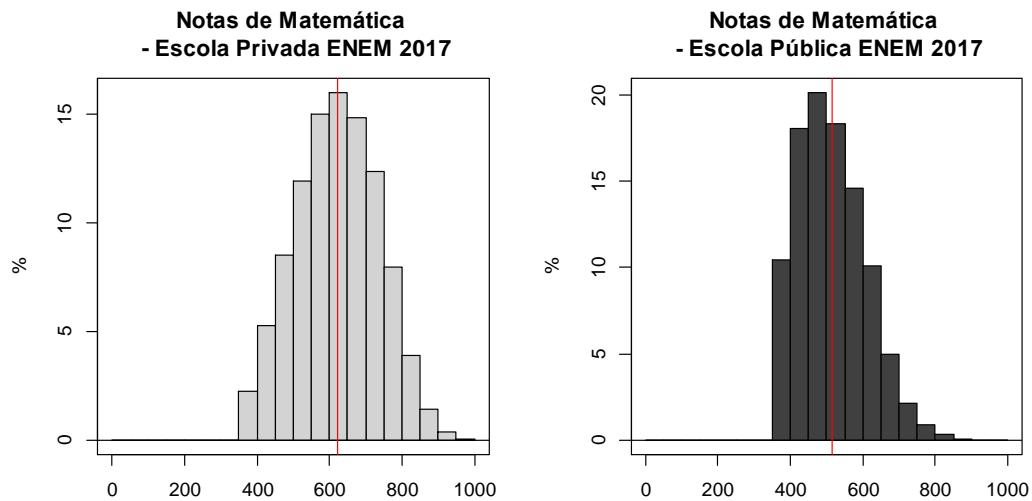


Figura 4.3 – Histogramas das notas em matemática dos candidatos ao ENEM 2017

Fonte: Própria autoria, 2021.

De acordo com o histograma das notas, percebe-se que, enquanto a média das notas dos alunos das escolas privadas está um pouco acima de 600 (linha vermelha), essa média se reduz consideravelmente em comparação aos alunos das escolas públicas. Visualmente, enquanto as notas parecem estar mais distribuídas em torno da média no primeiro histograma, no segundo, os dados apresentam uma assimetria positiva, ou seja, obtendo maior quantidade de alunos concentrados com notas menores. A seguir, na Tabela 4.1, encontram-se o primeiro e o terceiro quartis, a mediana e a média das notas em Matemática para as duas dependências administrativas.

Tabela 4.1 – Distribuição das notas em matemática no ENEM 2017 por dependência administrativa

Notas ENEM - Matemática	1º Quartil	Mediana	Média	3º Quartil
Escolas Públicas	441	503,7	513,9	575,9
Escolas Privadas	538,5	622,3	621	704

Fonte: Própria autoria, 2021.

Enquanto apenas 25% dos alunos de escolas públicas tiram notas superiores a 575,9, para esse mesmo percentual, os alunos das escolas privadas atingem notas superiores a 704. Considerando a nota de 704, a porcentagem equivalente de alunos cujas notas são superiores a essa nas escolas públicas é de menos de 4%. Analisando os valores da mediana, 50% dos alunos das escolas públicas tiram até 503,7 já a nota

mediana das escolas privadas é de 622,3, número superior ao terceiro quartil das notas das escolas públicas, evidenciando assim, a discrepância entre as dependências administrativas.

Em relação às diferenças de renda familiar dos alunos por dependência administrativa das escolas, pode-se verificar, na Figura 4.4, o quanto a renda dos estudantes de escolas públicas que fizeram ENEM é muito inferior em relação aos das escolas privadas. Enquanto a renda familiar média nas escolas privadas é de R\$ 6.246,00, nas escolas públicas é de R\$ 2.047,00. Apenas 10% das famílias dos alunos de escolas públicas possuem renda acima de R\$ 4.216,50. Já nas escolas privadas, 10% das famílias possuem renda superior ao valor de R\$ 12.649,50.

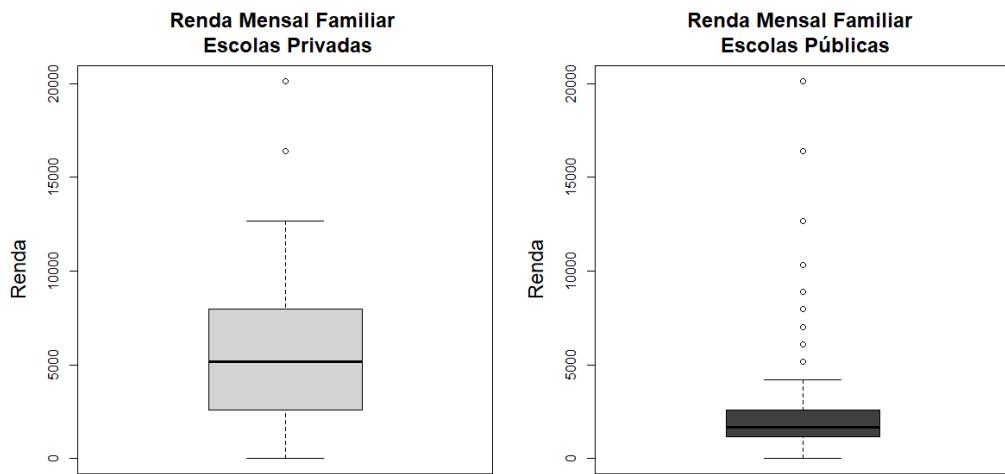


Figura 4.4 – Renda familiar dos candidatos ao ENEM 2017 por dependência administrativa

Fonte: Própria autoria, 2021.

Esta análise confirma os diversos estudos recentes na área que ressaltam as diferenças entre as escolas públicas e privadas, tanto no que concerne aos indicadores, quanto em relação à origem social dos alunos. As estimativas dos coeficientes das variáveis atribuídas ao efeito família e efeito escola estão apresentadas na Tabela 4.2 e são realizadas através do ajuste de um modelo multinível para cada uma das dependências administrativas. Nesta tabela, só estão incluídas as estimativas dos efeitos que foram considerados significativos ao nível de 5% e, ao final, o ICC relativo ao efeito escola. Os códigos em R referentes a estes resultados se encontram no Apêndice B. No Apêndice C, pode-se encontrar a Tabela C1, que inclui intervalos de confiança e p-valores.

Tabela 4.2 – Resultado da modelagem multinível para escolas públicas e privadas

ENEM - Matemática		
Variáveis	Escolas Privadas	Escolas Públicas
Constante	769,92	502,16
Efeito Família		
Nunca estudou		-7,61
Parou antes do 5ºano		-6,41
Ensino Fundamental		
Ensino Médio	7,87	6,42
Ensino Superior	15,95	13,19
Completou a pós graduação	17,09	12,90
Não sei		-9,92
Número de pessoas doméstico		-1,99
Renda Familiar	0,002	0,005
Idade	-16,94	-2,48
Sexo	-41,92	-29,98
Raça	6,03	9,04
Efeito Escola - Inep		
Complexidade de Gestão		3,63
Distorção Idade-Série	-1,37	-0,27
Média hora aula	6,34	4,09
Regularidade Docente	14,61	9,96
Nível socioeconômico Médio	22,71	12,28
Adequação Docente	0,42	0,12
Taxa de Aprovação	-0,84	-0,25
Taxa de Abandono		-0,50
Média de Alunos por turma	0,44	
Esforço Docente	-0,31	-0,39
ICC	9%	13%

Fonte: Própria autoria, 2021.

Antes de analisar os efeitos de cada indicador na nota de Matemática dos estudantes que fizeram o ENEM em 2017, ressalta-se que, segundo a medida ICC, que avalia a porcentagem de contribuição das variáveis que compõem o efeito escola nesta nota, verifica-se que, para as escolas públicas, esta proporção é de 9% e, para as escolas privadas, de 13%. Embora esses percentuais sejam baixos, eles estão dentro da faixa esperada de acordo com a literatura quando há controle de nível socioeconômico dos alunos nos modelos. Resultados similares e uma discussão importante sobre o papel

da escola são apresentados em TRAVITZKI e BOTO (2013), embora com modelagens e variáveis diferentes, mas com resultados de 6% ou 7% de explicação de variância pela escola com controle econômico-social.

Os resultados do ICC mostram que os indicadores educacionais contribuem mais para o desempenho dos alunos de escolas privadas do que para os alunos de escolas públicas, na Região Sudeste do Brasil, com a população pertencente a faixa estudada. Dito de outra forma, este resultado significa que 87% da variação das notas de Matemática nas escolas privadas se dá fora do contexto escolar, assim como 91% nas escolas públicas.

Em relação à comparação com o Lasso, o que se observa é um equilíbrio entre as variáveis selecionadas por este algoritmo e as consideradas significativas pelo modelo multinível. Para escolas públicas, a única discordância é em taxa de aprovação, descartada pelo Lasso e considerada significativa pelo modelo multinível. Houve concordância em todas as outras variáveis. Para escolas privadas, há concordância total nas variáveis escolares e apenas três discordâncias nas variáveis individuais e socioeconômicas.

Ao analisar as estimativas separadas de cada um dos indicadores que medem o efeito escola, destacam-se o nível socioeconômico médio e o indicador de regularidade docente. Estes efeitos chegam a aumentar a nota em 22,7 pontos para cada nível socioeconômico e em 14,6 para cada nível de regularidade docente nas escolas privadas e quase 10 pontos em cada um desses níveis nas escolas públicas. O indicador que mede a complexidade de gestão (número de alunos, turmas e turnos) só teve efeito significativo para os estudantes de escolas públicas. Este efeito, que é de 3,6 pontos a mais para cada nível de complexidade, deve traduzir o fato de que escolas públicas maiores e com mais turnos estejam de alguma forma mais alinhadas à preparação dos seus estudantes para o ENEM. As variáveis que mediram a distorção idade série, esforço docente e taxa de aprovação causam impactos negativos nas notas. Em relação aos docentes, por exemplo, uma escola onde 100% dos seus professores têm esforço considerado alto, há perda de, em média, 31 pontos em escolas privadas e 39 pontos em públicas quando comparados às escolas que não tenham nenhum docente nessa situação. Em contraponto ao esforço docente, se 100% dos professores são considerados adequados, isto é, tiverem licenciatura ou complementação pedagógica na área que lecionam, há um aumento médio de 42 pontos em escolas privadas e 12 pontos nas públicas. Média hora aula e regularidade docente se relacionam linearmente com coeficientes positivos para as notas. É razoável esperar que o impacto da média de horas aula nas notas pode ser positivo, em um exame como o ENEM. Já o comportamento de regularidade docente é esperado por ser uma variável interessante

do ponto de vista educacional, que mede a continuidade dos professores nas escolas. Isso significa que um professor com mais regularidade pode estar mais inserido na cultura escolar e possuir maior conhecimento do perfil dos alunos.

Boa parte das variáveis se mostrou significativa para os modelos. Sendo assim, com exceção do número médio de alunos por turma, todos os outros indicadores utilizados são significativos no modelo das escolas públicas. No caso do modelo para as escolas privadas, os únicos indicadores não significantes são os de complexidade de gestão e taxa de abandono.

A análise de cada uma das estimativas dos efeitos das variáveis que medem o efeito família mostra que, à medida que cresce a escolaridade materna do estudante, aumenta seu efeito na nota de Matemática. Por exemplo, o estudante cuja mãe tem o ensino médio, tem nota 6,4 pontos maior se pertence a uma escola pública e 7,9 maior se estuda em uma escola privada, esses efeitos sendo ainda maiores no caso de a escolaridade ser nível superior ou pós-graduação. O número de pessoas no domicílio não é significativo para os estudantes de escolas privadas, mas diminui em 2 pontos para cada pessoa a mais no domicílio de estudantes de escolas públicas, provavelmente por refletir a falta de condições apropriadas para o estudo. Já quanto ao efeito da renda familiar, verifica-se que cada R\$ 1.000,00 a mais de renda, tem o efeito de aumentar 2 e 5 pontos respectivamente, nas notas dos estudantes das escolas públicas e privadas. Apesar do impacto aparentemente pequeno, ao considerar a amplitude da variável, que chega a R\$ 20.000,00, as notas podem variar em mais de 100 pontos. As variáveis renda familiar e escolaridade materna indicam a importância dos capitais econômico e social no desempenho dos alunos.

Em relação às variáveis individuais e socioeconômicas, verifica-se que os estudantes que fazem o ENEM com mais idade e que estudaram em escolas privadas, chegam a ter quase 17 pontos em desvantagem para cada ano. Este efeito negativo diminui para cerca de 2,5 entre os estudantes de escolas públicas. Em relação ao sexo, os de sexo masculino têm uma vantagem de 42 pontos nas escolas privadas e de 30 pontos, nas públicas e os brancos têm vantagens de 6 e 9 pontos nessas duas dependências.

Melhorias e adaptações nas escolas privadas podem ser feitas de acordo com o planejamento dos gestores e com as condições econômicas da escola. Em relação aos docentes, as contratações e estímulos à carreira, como melhorias nas condições de trabalho dentro e fora da escola, treinamento e salários, são controlados individualmente pelos gestores e pela comunidade escolar, ao contrário do cenário público, onde é necessário que haja políticas educacionais apropriadas. A conclusão é de que é mais factível administrar a melhora dos indicadores para as escolas privadas, o que acaba levando ao maior impacto observado pelos modelos de regressão.

Uma dificuldade usual em estudos desta natureza é identificar um maior efeito escola em modelos que também mensuram o efeito família. A discussão pode ser encaminhada sob diversos aspectos, mas consideramos dois, particularmente relevantes: um sob o ponto de vista educacional e, o outro, relacionado à análise estatística. Educacionalmente, o real papel da escola não é, ou não deve ser, preparar para exames em larga escala, mas sim contribuir para a formação do cidadão em um sentido social mais amplo. Estatisticamente, há de se levar em consideração que a mensuração da contribuição escolar é, usualmente, feita por *proxies* que não necessariamente medem o que se deseja. É muito difícil mapear qualidade docente, além de diversos outros atributos relacionados à influência que a escola proporciona aos alunos. Como essas características são difíceis de serem medidas, especialmente em exames de larga escala, há a possibilidade do efeito escola não ser corretamente estabelecido por fatores não observados pelo modelo.

Finalmente, é importante frisar que o efeito escola, neste caso, não pode ser tomado como uma medida de qualidade escolar, mesmo porque os próprios indicadores, isoladamente, não oferecem informação sobre a qualidade da escola, mas, sim, sobre a variável que está sendo medida. A interpretação correta é que a relação linear entre os indicadores e o desempenho em Matemática em uma modelagem multinível, com intercepto aleatório, indica que 13% da variação das notas é explicada pela escola para alunos da iniciativa privada e 9% da variação é explicada pela escola para alunos da rede pública.

No Apêndice D, questões referentes à análise de resíduos podem ser encontradas. Autores como BOX (1976) defendem que, na prática, “todos os modelos estão errados”. Já McCULLOCH e NEUHAUS (2011) e HOX (2010) ressaltam a robustez da estimativa de máxima verossimilhança, mesmo quando os pressupostos não são atendidos.

4.2 RESULTADOS E DISCUSSÕES: DESIGUALDADE DE DESEMPENHO

Nesta seção, são apresentados os resultados relativos aos modelos de regressão quantílica e multinível; uma análise sobre escolaridade materna, renda familiar, raça, regularidade, esforço docente e nível socioeconômico escolar médio em comparação ao desempenho dos estudantes em Matemática no ENEM 2017 e um breve debate sobre a desigualdade existente entre alunos de escolas públicas e privadas na Região Sudeste do Brasil.

4.2.1 RESULTADOS DOS MODELOS

Os resultados dos modelos empíricos estimados a partir das equações (3.3) e (3.4) se encontram nas Tabelas 4.3 e 4.4. O não preenchimento de um valor em uma célula das tabelas significa que a variável não é considerada significativa pelo modelo ao nível de 5%. Para as regressões quantíticas, tanto para alunos de escolas públicas quanto de escolas privadas, são usados os quantis $\tau = 10, 25, 50, 75$ e 90 . Os códigos em R referentes aos resultados das escolas privadas se encontram no Apêndice B.

Tabela 4.3 – Resultados do modelo de regressão quantílica e multinível para matemática em escolas públicas

ENEM - Matemática	Quantil					Modelagem Multinível
	10%	25%	50%	75%	90%	
Variáveis						
Constante	452,82	528,56	601,50	670,17	735,12	582,71
Efeito Família						
Nunca estudou	-2,68	-4,11	-7,08	-12,81	-16,54	-7,50
Parou antes do 5ºano	-1,97	-5,34	-7,10	-10,12	-8,75	-6,47
Ensino Fundamental						
Ensino Médio	3,69	5,68	8,48	8,14	7,46	6,50
Ensino Superior	9,07	13,47	17,31	19,26	18,48	13,11
Completou a pós graduação	8,69	17,96	18,60	21,29	21,09	12,71
Não sei	-4,86	-11,08	-12,34	-15,27	-13,36	-9,84
Número de pessoas domicílio	-1,24	-1,94	-2,70	-3,16	-2,98	-1,98
Renda Familiar	0,004	0,006	0,007	0,008	0,009	0,005
Idade	-1,06	-1,66	-2,35	-2,57	-2,70	-2,48
Sexo	-10,57	-21,46	-32,46	-40,14	-42,39	-29,97
Raça	4,95	7,97	11,21	11,69	10,77	9,28
Efeito Escola - Infraestrutura						
Localização	-5,75	-10,03	-9,64	-10,26	-9,26	-8,16
Sala Diretoria						
Sala Professor	5,56	4,05	7,56	12,76		
Laboratório de Informática		3,41	6,22	6,53	7,71	4,41
Quadra de Esportes	-5,52	-8,32	-9,54	-13,13	-13,67	-9,11
Cozinha	-22,31	-29,99	-28,82	-21,83	-23,55	-26,75
Biblioteca	10,72	18,39	26,82	31,95	34,18	22,12
Sala de Leitura	-4,75	-5,77	-4,58	-1,89		-3,86
Refeitório	-2,30	-2,59	-2,90			-2,67
Computador						
Internet						
Banda Larga						
Alimentação	-47,74	-75,02	-88,47	-97,61	-91,31	-68,88
Efeito Escola - Inep						
Complexidade de Gestão						
Distorção Idade-Série	-0,33	-0,54	-0,63	-0,73	-0,69	-0,52
Média hora aula	2,78	3,89	4,51	4,51	4,53	4,71
Regularidade Docente	4,38	6,50	7,06	6,49	6,61	7,44
Nível socioeconômico Médio	8,81	13,88	17,24	18,71	18,69	14,68
Adequação Docente						
Taxa de Aprovação	-0,12	-0,18	-0,17	-0,14	-0,14	
Taxa de Abandono	-0,21	-0,34	-0,45	-0,39	-0,36	
Média de Alunos por turma				-0,15	-0,27	
Esforço Docente	-0,04	-0,10	-0,21	-0,32	-0,40	-0,16

Fonte: Própria autoria, 2021.

Tabela 4.4 – Resultados do modelo de regressão quantílica e multinível para matemática em escolas privadas

ENEM - Matemática	Quantil					Modelagem Multinível
	Escolas Privadas	10%	25%	50%	75%	
Variáveis						
Constante	576,41	696,30	776,63	781,96	791,29	707,31
Efeito Família						
Nunca estudou	-19,13					
Parou antes do 5ºano						
Ensino Fundamental	-14,63	-12,91				
Ensino Médio			9,10	10,89		7,77
Ensino Superior		14,44	19,60	23,31	16,44	15,91
Completou a pós graduação	10,44	20,01	25,94	27,77	20,23	17,03
Não sei						
Número de pessoas domicílio						
Renda Familiar	0,0020	0,0028	0,0030	0,0027	0,0025	0,0020
Idade	-16,28	-18,32	-15,70	-13,05	-10,06	-17,00
Sexo	-27,68	-38,99	-44,47	-45,89	-45,40	-41,97
Raça	5,10	4,61	3,65			6,17
Efeito Escola - Infraestrutura						
Localização						
Sala Diretoria						
Sala Professor		16,53	16,52	12,96	17,99	
Laboratório de Informática	-7,11	-8,56	-8,65	-9,49	-8,48	-8,28
Quadra de Esportes						
Cozinha		4,77		4,27		6,68
Biblioteca	16,34	18,04	18,93	19,52	19,04	15,11
Sala de Leitura	8,36	10,40	12,21	10,98	9,79	6,24
Refeitório						
Computador			19,27	33,12	43,41	
Internet					-37,54	
Banda Larga					17,40	
Alimentação		-4,85	-7,76	-12,94	-10,10	-10,49
Efeito Escola - Inep						
Complexidade de Gestão	-3,23	-4,20	-2,61	-1,63		
Distorção Idade-Série	-1,56	-2,12	-2,42	-2,36	-2,50	-1,55
Média hora aula	6,27	6,88	4,33	3,48	2,76	6,27
Regularidade Docente	14,08	15,00	15,42	15,40	13,62	14,44
Nível socioeconômico Médio	18,28	21,10	24,47	26,61	25,76	22,21
Adequação Docente	0,29	0,41	0,38	0,40	0,26	0,43
Taxa de Aprovação		-1,15	-1,70	-1,74	-1,50	
Taxa de Abandono						
Média de Alunos por turma	0,25	0,37	0,40	0,39	0,39	0,38
Esforço Docente	-0,12	-0,18	-0,25	-0,27	-0,27	-0,29

Fonte: Própria autoria, 2021.

De uma forma geral, os resultados são consistentes com a literatura em educação. Escolaridade materna aumenta as notas sistematicamente a partir do ensino médio, mas possui impacto negativo para mães com até 9 anos de educação. No caso de escolas privadas, poucas variáveis são significativas para mães com escolaridade

abaixo do ensino médio. A mãe possuir pós-graduação impacta em 21 pontos nas notas de alunos de desempenho alto e 18 pontos em desempenho baixo nas escolas públicas (Tabela 4.3). No extremo, se avaliamos as diferenças entre os quantis 90 e 10 para esta variável, o impacto causado pela mãe possuir pós-graduação chega a ser em torno de duas vezes maior. A renda familiar aumenta as notas entre 2 e 9 pontos a cada R\$ 1.000,00 a mais na renda da família. Esta variável, que possui contribuição estável nas notas ao longo dos quantis das escolas privadas, apresenta comportamento diferente para as escolas públicas: os alunos de desempenho mais alto (quantil 90) aumentam 9 pontos a cada R\$ 1.000,00, enquanto os de desempenho mais baixo (quantil 10) aumentam apenas 4 pontos a cada R\$ 1.000,00. Este resultado indica que o aumento da renda tem maior impacto para os alunos de maior desempenho. Em outras palavras, deve-se saber usar as circunstâncias favoráveis, como uma renda maior, para incentivar o aprendizado. Quanto à raça, alunos de raça branca tiram notas maiores, o que pode estar ligado à questão social e será debatido na sequência. Esta variável possui um perfil completamente diferente entre as escolas públicas e privadas. Para as públicas, como pode-se observar na Tabela 4.3, há um crescimento até o quantil 50 (entre 5 e 11 pontos) e depois os valores se estabilizam. O valor médio no modelo multinível, que é em torno de 9 pontos, mostra a assimetria dos coeficientes ao longo dos quantis. Nas escolas privadas (Tabela 4.4), a variável não é significativa acima do quantil 50, o que corrobora a hipótese de estar associada a uma questão social, já que para alunos de alto desempenho, a raça não é uma variável significativa nas escolas privadas. Em relação a sexo, meninos tiram entre 10 e 45 pontos a mais em Matemática em comparação com as meninas, e a diferença fica maior conforme o desempenho aumenta.

Considerando as variáveis de infraestrutura, a biblioteca é considerada importante tanto no contexto público quanto no privado, enquanto outras variáveis possuem um padrão diferente. Por exemplo, laboratório de informática tende a ser importante para escolas públicas e sala de leitura e computador para escolas privadas. A combinação entre acervo disponível e espaço para estudo parece funcionar melhor em escolas privadas, uma vez que é provável que estas disponham de mais recursos financeiros e estruturais para prover ambientes confortáveis e que estimulem os alunos a estudarem na escola fora dos momentos de aula. A importância da biblioteca pode, também, remeter a questões relacionadas a variáveis não observáveis, uma vez que escolas com bibliotecas podem ser mais organizadas e melhor estruturadas, características difíceis de serem medidas em modelagens quantitativas.

Quanto aos indicadores públicos do INEP, média hora aula, regularidade e adequação docente e nível socioeconômico médio impactam positivamente nas notas,

enquanto que esforço docente, distorção idade-série e complexidade de gestão escolar tendem a produzir queda nas notas. O docente ser regular na escola já garante entre 4 e 7 pontos nas escolas públicas e entre 13 e 15 pontos nas privadas multiplicados pelo valor de regularidade. Por outro lado, se 100% dos professores da escola possuem alto esforço docente, a nota decresce, em média, 29 pontos para alunos de escolas privadas e 16 para os alunos das públicas. Vale lembrar que esforço docente está ligado ao número de turnos, turmas e escolas que o professor leciona. Um esforço maior significa uma demanda maior de tempo. Algumas discussões específicas que tangem a desigualdade social são abordadas na próxima subseção. A análise por quantis permite medir o impacto das variáveis para diferentes desempenhos de alunos. Ao analisar a variável de esforço docente nas escolas públicas, o impacto é 10 vezes pior para notas no quantil 90 do que para notas no quantil 10. O mesmo acontece nas escolas privadas, só que em uma proporção entre 2 e 3 vezes. Ou seja, uma porcentagem de professores com esforços altos influencia negativamente todos os quantis, em especial os mais altos. Considerando o quantil 50 e as modelagens multiníveis para escolas públicas e privadas nas Tabelas 4.3 e 4.4, pode-se concluir que as variáveis significativas coincidem em grande parte das vezes e observa-se uma tendência das estimativas da regressão quantílica serem um pouco superiores em valor absoluto aos da modelagem multinível. Mas de uma forma geral, as duas estimativas apresentam padrões de resultados similares.

4.2.2 ANÁLISES SOBRE VARIÁVEIS FAMILIARES E ESCOLARES

Para analisar as diferenças entre os perfis sociais de alunos de escolas públicas e privadas, são levados em consideração os resultados dos modelos apresentados nas Tabelas 4.3 e 4.4, assim como algumas análises descritivas que são apresentadas nesta discussão. Algumas variáveis recebem destaque, como renda familiar, escolaridade materna e raça, nos contextos individual e familiar, e esforço e regularidade docente e nível socioeconômico médio, no contexto escolar.

Para verificar as diferenças entre escolas públicas e privadas, a primeira questão a se levar em consideração é a variável dependente aqui utilizada, isto é, a proficiência em Matemática. Na Figura 4.5, podem-se observar densidades estimadas para as duas dependências administrativas e, na Tabela 4.5, os valores estimados para os quantis, indicando uma tendência de melhor desempenho para alunos de escolas privadas. Como comparação pela definição de desempenho aqui apresentada na seção de Materiais e Métodos o desempenho considerado alto nas escolas públicas é atingido com 576 pontos e, nas privadas, com 704 pontos. Já o desempenho baixo, se atinge

com 441 pontos nas públicas e 539 nas privadas. O quantil 90 nas escolas públicas, cujo valor é em torno de 640 pontos, é próximo ao quantil 50 nas escolas privadas, que é em torno de 620 pontos. Outra possibilidade de mensurar estas diferenças está no cálculo de riscos relativos, que se trata da razão entre a proporção de escolas privadas e a proporção de escolas públicas. O risco de um aluno de escola privada tirar mais de 600 no exame de Matemática do ENEM é 3 vezes o de um aluno de escola pública. Já para notas menores que 400, há 4 vezes mais risco de um aluno de escola pública em comparação a um aluno de escola privada.

Tabela 4.5 – Valores dos quantis para a proficiência matemática do aluno em escolas públicas e privadas

Notas ENEM - Matemática	Quantil				
	10%	25%	50%	75%	90%
Escolas Públcas	398,7	441,0	503,7	575,9	640,1
Escolas Privadas	466,1	538,5	622,3	704,3	769,5

Fonte: Própria autoria, 2021.

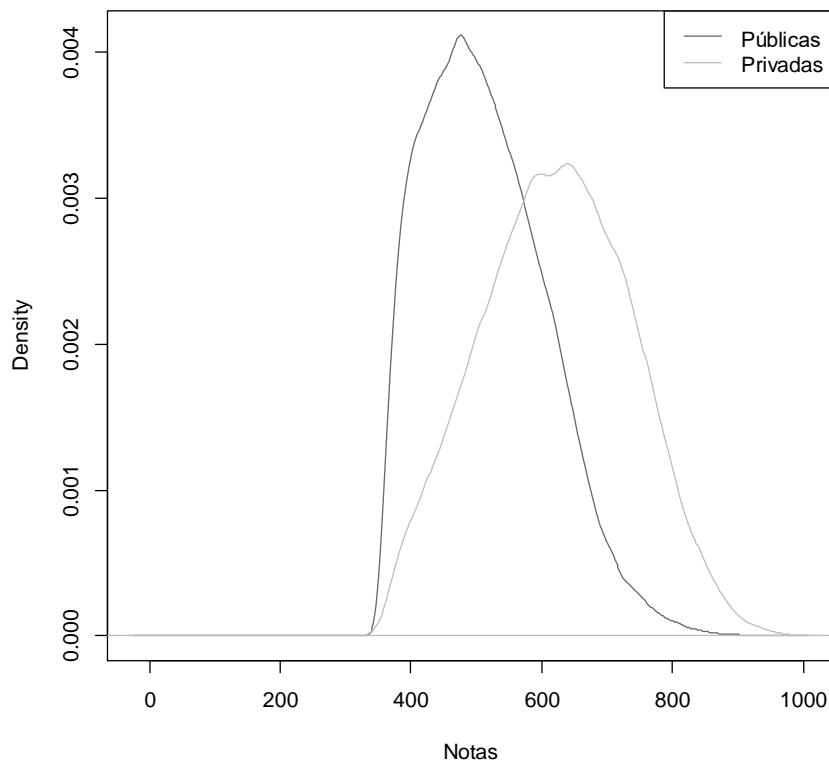


Figura 4.5 – Densidades estimadas de notas matemáticas de estudantes de escolas públicas e privadas

Fonte: Própria autoria, 2021.

Em relação às variáveis independentes, a escolaridade materna é uma variável tipicamente relacionada à proficiência na literatura em educação. Os resultados dos modelos presentes nas Tabelas 4.3 e 4.4 indicam que a mãe possuir ensino superior ou pós-graduação aumenta as notas, até mesmo em quantis baixos, enquanto mães com escolaridade menor ou igual a 9 anos de estudo causa impacto negativo nas notas. A distribuição da escolaridade materna pode ser encontrada na Figura 4.6, para alunos de escolas públicas e privadas. As 8 categorias de escolaridade materna são consideradas para estas análises.

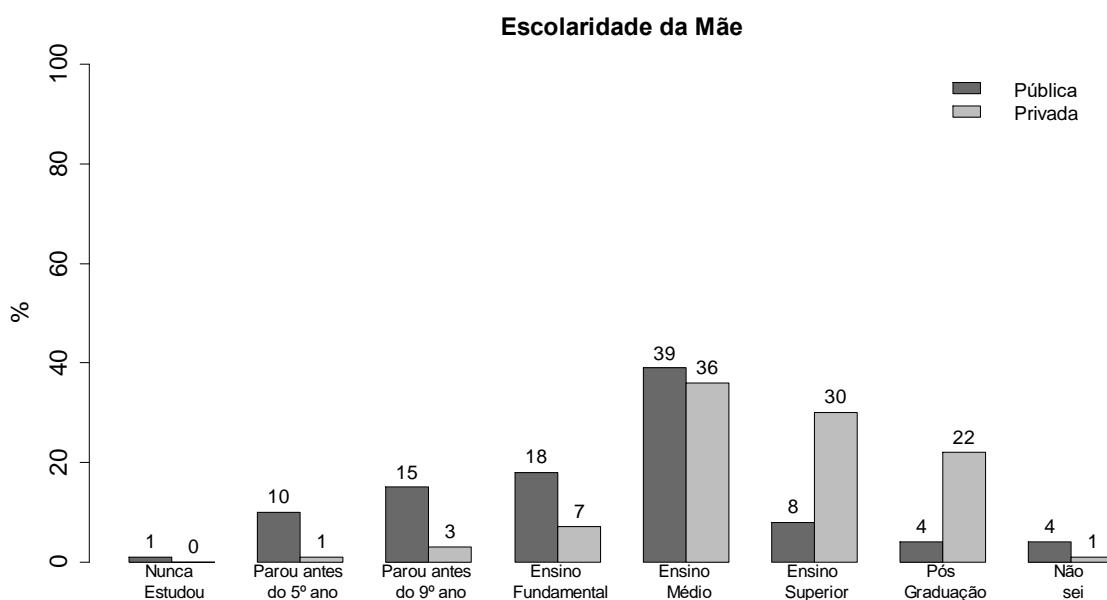


Figura 4.6 – Distribuição da educação materna de alunos de escolas públicas e privadas

Fonte: Própria autoria, 2021.

Apenas para o ensino médio há um equilíbrio entre as duas amostras. Para ensino superior e pós-graduação há uma supremacia em termos proporcionais de alunos de escolas privadas e, para mães com escolaridade abaixo do ensino médio, a maioria proporcional é de alunos de escolas públicas. Levando-se em consideração a informação presente na Figura 4.6 e nas Tabelas 4.3 e 4.4, percebe-se o impacto favorável que a escolaridade da mãe produz nas notas, principalmente de alunos de escolas privadas.

Para a renda familiar, variável originalmente intervalar, há 17 categorias que, neste banco de dados, são representadas pelo ponto médio de cada intervalo. Para as análises descritivas, as categorias são agrupadas de 3 em 3, correspondendo a: menor

que R\$ 1.171,50; entre R\$ 1.171,50 e menor que R\$ 2.576,50; entre R\$ 2.576,50 e menor que R\$ 5.133,50; entre R\$ 5.133,50 e menor que R\$ 7.964,50; entre R\$ 7.964,50 e menor que R\$ 12.649,50; maior que R\$ 12.649,50. O último grupo é composto pelas duas últimas categorias. No Brasil, considera-se que uma família é de baixa renda se possui renda familiar inferior a 3 salários mínimos (SEDES, 2018). Em 2017, o salário mínimo era de R\$ 937,00, o que significa que os dois primeiros grupos são integralmente de famílias de baixa renda e parte do terceiro. Ainda hoje, com o salário mínimo atual de R\$ 1.100,00, a situação seria exatamente a mesma.

Dois mapas de cores são apresentados nas Figuras 4.7 e 4.8, comparando renda familiar com os desempenhos dos alunos de escolas públicas e privadas, respectivamente. No eixo x se encontram os grupos de renda e, no eixo y, os quantis. Em cada retângulo, tem-se a quantidade de alunos que teve seu desempenho associado aquele quantil e possui a renda familiar correspondente. Quanto mais escuro o retângulo, maior a quantidade de alunos pertencentes a ele. É esperado que as regiões mais escuras nos mapas de cores sejam as pertencentes aos intervalos dos quartis 25 a 75, pois há 50% dos dados nestas regiões contra 25% acima e 25% abaixo.

Na Figura 4.7, que se trata da análise de renda familiar x desempenho para escolas públicas, pode-se observar que a maior parte dos alunos pertence a famílias com renda inferior a R\$ 5.133,50. Além disso, como esperado, para rendas mais baixas há mais alunos com notas nos quantis mais baixos e maiores rendas estão relacionadas a maiores notas, mas em menor quantidade. Na Figura 4.8, que se trata da análise renda familiar x desempenho para escolas privadas, é importante levar em consideração que a escala de cores é diferente, uma vez que há muito menos alunos de escolas privadas do que de escolas públicas. O perfil destes alunos é diferente, concentrando a renda entre pouco mais de 2 até 7 salários mínimos. Há um padrão linear neste segundo gráfico, o que indica que maiores rendas estão associadas a maiores notas.

Para efeito de comparação entre o perfil de escolas públicas e privadas, deve-se levar em consideração a Tabela 4.5, que mostra que os valores dos quantis são diferentes entre elas. O valor da nota correspondente a um determinado quantil em uma escola privada é sempre maior que a nota associada ao mesmo quantil em uma escola pública.

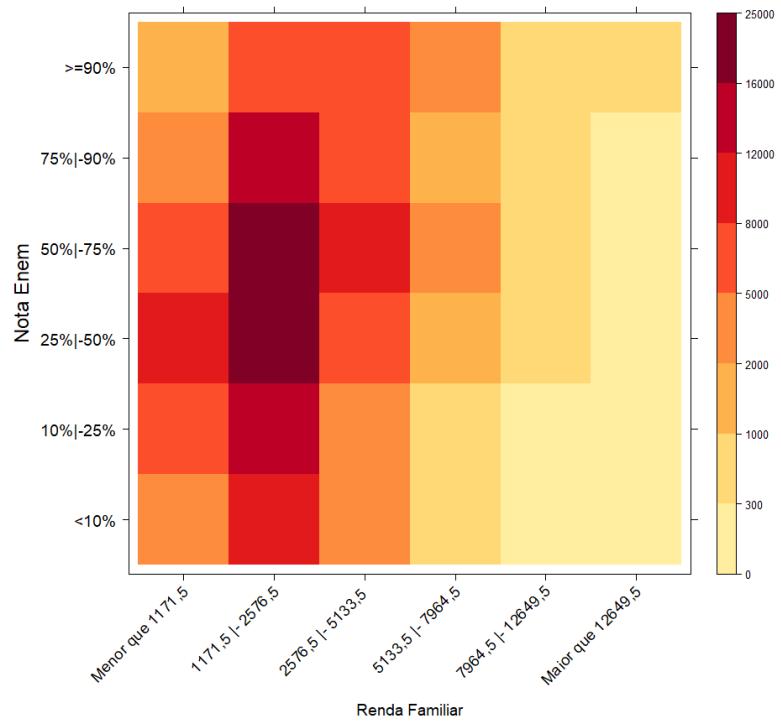


Figura 4.7 – Mapeamento de cores do número de alunos por renda e desempenho por quantil para escolas públicas

Fonte: Própria autoria, 2021.

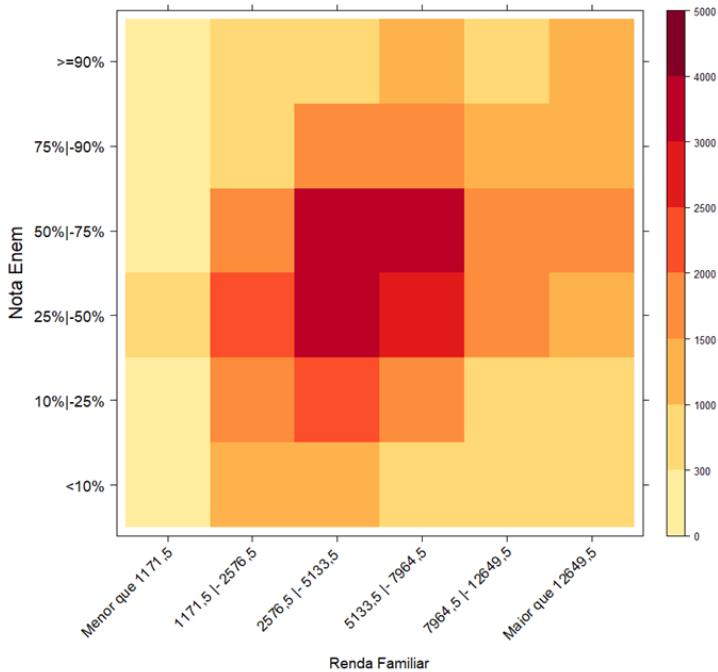


Figura 4.8 – Mapeamento de cores do número de alunos por renda e desempenho por quantil para escolas privadas

Fonte: Própria autoria, 2021.

Outra variável importante dentro do contexto individual e familiar quando se estuda os fatores que influenciam a proficiência a fim de contribuir para políticas públicas e gestão escolar é a raça. Nesta seção dividimos entre brancos e não brancos. Os resultados dos modelos indicam uma diferença grande entre escolas públicas e privadas. Na média, alunos brancos tiram 9 pontos a mais em escolas públicas e 6 pontos a mais em escolas privadas. A regressão quantílica apontou um decrescimento dos coeficientes de raça em escolas privadas até o quantil 50 e, a partir dele, a raça não foi mais significativa. Já para escolas públicas, a raça foi significativa em todos os quantis, com um aumento progressivo até o quantil 50 e estabilização nos quantis superiores. Este comportamento pode apontar para uma associação com desigualdade social uma vez que, para as notas intermediárias e altas, a variável não é significativa para escolas privadas e possui a mesma contribuição para escolas públicas. Além disso, os alunos brancos de escolas privadas possuem famílias com renda maior ou igual que os alunos brancos de escolas públicas, assim como alunos não brancos de escolas privadas possuem renda maior ou igual a alunos não brancos de escolas públicas ($p < 0,01$ em todos os casos). Para os testes de hipótese entre duas amostras utiliza-se o teste de Wilcoxon. Um debate mais profundo sobre raça será apresentado nos resultados referentes a equidade.

Para as variáveis escolares, os docentes das escolas privadas têm índices de regularidade maiores que os docentes de escolas públicas ($p < 0,01$). Trata-se de um resultado surpreendente porque a carreira de magistério de ensino público possui estabilidade, o que significa que o docente concursado tem o direito de permanecer na instituição até a aposentadoria. Nas escolas particulares, mesmo que possuam política de carreira docente, não há estabilidade.

Para o esforço docente observa-se que professores de escolas públicas possuem um esforço maior comparado aos de escolas privadas ($p < 0,01$). Isso ocorre pela diferença salarial, o que implica em lecionar em mais escolas, mais turnos e para mais alunos.

Já o nível socioeconômico médio, variável também importante para a proficiência na literatura, possui um impacto médio de 14 pontos em escolas públicas e 22 pontos para escolas privadas. Essa variável avalia a contribuição dos pares nas notas e os alunos de escolas privadas possuem nível socioeconômico médio maiores que os alunos de escolas públicas ($p < 0,01$).

Essas três análises sobre variáveis escolares mostram que os docentes das escolas privadas são mais regulares e atuam com um esforço menor comparado aos docentes de escolas públicas. Já o nível socioeconômico médio dos pares nas escolas privadas

também é superior se comparado aos das escolas públicas. Consolidando os resultados individuais e familiares aos escolares, se observa que o cenário é muito mais favorável aos alunos de escolas privadas tanto sob o ponto de vista econômico social quanto sob a ótica escolar, com o desempenho entre bom e ótimo de um aluno de uma escola pública podendo ser comparado a um resultado entre mediano e bom de um aluno de escola privada.

Analizando conjuntamente os resultados podem-se comparar os coeficientes das regressões associados aos mapas de cores e traçar o perfil da base de dados estudada. É razoável esperar que baixa renda e baixa escolaridade estejam relacionadas negativamente nas notas e o contrário acontecer para as notas altas. O problema é que há grandes diferenças entre as distribuições de renda familiar e escolaridade materna para os alunos de escolas públicas e privadas e a dificuldade dos alunos de escolas públicas em terem acesso ao ensino superior pode indicar mais um componente para que este círculo vicioso seja mantido, onde famílias de baixa renda não conseguem, na grande maioria das vezes, acesso à educação de qualidade e não encontram uma saída para aumentar a sua renda, através de mobilidade social.

Sob o ponto de vista escolar, é razoável conjecturar se a infraestrutura da escola e a qualidade docente podem amortizar o impacto negativo de uma realidade familiar que seja difícil de colaborar para o desempenho do aluno, como debatido na introdução com o conceito de escola efetiva. Entretanto, o que se observa é um cenário similar à comparação familiar entre alunos de escolas públicas e privadas. Escolas com professores menos regulares e com mais esforço docente acabarão tendo mais dificuldade em reverter este quadro próximo ao exame de acesso a universidade, especialmente após o aluno ter uma vida escolar inteira inserido neste contexto. Outras discussões sobre equidade e o âmbito escolar serão apresentadas na próxima seção.

4.3 RESULTADOS E DISCUSSÕES: EQUIDADE

Os resultados referentes a equidade são divididos em uma análise exploratória, onde a desigualdade de desempenho entre diferentes grupos sociais é verificada e, posteriormente, os coeficientes são estimados e uma análise dos modelos descritos na seção de Materiais e Métodos é realizada. Esses resultados contemplam alunos de todo o Brasil e não apenas da Região Sudeste.

4.3.1 ANÁLISE EXPLORATÓRIA

Nesta seção é apresentada uma análise da distribuição das notas em Matemática no ENEM nas categorias das variáveis individuais e socioeconômicas (sexo, raça, nível de escolaridade e renda familiar) dos alunos. Para essa análise, as variáveis socioeconômicas são categorizadas da seguinte maneira: escolaridade materna, sendo ensino médio em diante e abaixo de ensino médio e renda, dividida em alta, média e baixa.

De acordo com a Tabela 4.6, verifica-se que as notas dos alunos brancos e amarelos são superiores em relação aos alunos pardos, pretos e indígenas, o que está de acordo com o encontrado por TRAVITZKI *et al.* (2016). Se a comparação for entre as medianas, tem-se 510,2 para brancos contra 479,3 de alunos pretos, uma diferença em torno de 30 pontos. Isto indica que a análise de desempenho por raça para a população pré-universitária está de acordo com a literatura para educação básica no ensino fundamental: o desempenho de alunos brancos e amarelos costuma ser melhor em comparação ao desempenho de alunos pretos e pardos, mas, em geral, os alunos pardos possuem notas melhores que os pretos (ALVES *et al.*, 2016).

Tabela 4.6 – Distribuições das proficiências em matemática no ENEM 2017 por raça.

Notas em Matemática	Raça					
	Não Declarado	Branca	Preta	Parda	Amarela	Indígena
1º Quartil	430,10	445,00	424,10	426,20	429,90	422,60
Mediana	487,60	510,20	479,30	482,70	489,30	473,10
Média	501,30	519,60	490,80	495,20	501,70	483,30
3º Quartil	560,40	583,90	546,60	551,50	559,80	534,50

Fonte: Própria autoria, 2021.

A partir das medidas apresentadas na Tabela 4.7, observa-se que os alunos do sexo masculino tiram notas mais altas em Matemática do que o feminino, o que também reforça resultados anteriores na literatura em educação. Em relação à mediana, a diferença fica em torno de 34 pontos a mais para o sexo masculino. No primeiro quartil a diferença fica em torno de 20 pontos e, no terceiro quartil, alcança mais de 46 pontos. Em comparação com a análise anterior, as diferenças entre brancos e pretos são de 20, 30 e 37 pontos nos percentis 25, 50 e 75 favoráveis aos brancos. Já para sexo, os valores são em torno de 20, 34 e 46 nos mesmos percentis em favor do sexo masculino,

o que sugere que, para grupo de alunos com desempenho maiores, a diferença da variável sexo é maior do que em relação à raça.

Tabela 4.7 – Distribuições das proficiências em matemática no ENEM 2017 por sexo.

Notas em Matemática	Sexo	
	Feminino	Masculino
1º Quartil	424,80	445,30
Mediana	479,50	513,80
Média	490,20	523,20
3º Quartil	543,90	590,80

Fonte: Própria autoria, 2021.

Adicionando a escolaridade materna aos agrupamentos anteriores, tem-se os resultados da Tabela 4.8.

Tabela 4.8 – Desigualdades em desempenho segmentadas por sexo, raça e escolaridade materna.

Escolaridade Materna		Meninas		Meninos	
		Brancas	Pretas	Brancos	Pretos
Abaixo Ensino Médio	1º Quartil	425,20	412,10	448,60	426,10
	Mediana	479,30	458,70	516,50	481,50
	Média	488,84	469,59	523,44	493,50
Pelo menos Ensino Médio	3º Quartil	542,10	517,00	590,20	551,10
	1º Quartil	448,40	427,90	473,90	444,90
	Mediana	510,20	483,20	551,70	509,90
Ensino Médio	Média	517,83	492,70	554,59	518,99
	3º Quartil	578,10	548,00	626,80	584,90

Fonte: Própria autoria, 2021.

Verifica-se que alunos do sexo masculino e de raça branca possuem os melhores resultados ($p < 0,01$); há um equilíbrio nas notas entre sexo masculino e raça preta em comparação a sexo feminino e raça branca; alunas do sexo feminino e raça preta possuem o pior desempenho. Para ilustrar a desigualdade, pode-se observar que apenas 25% das alunas do sexo feminino e de raça preta, cujas mães possuem escolaridade abaixo do ensino médio (quarta linha da tabela) tiram notas superiores a 517, o que corresponde aproximadamente a 25 pontos a menos que alunas do sexo feminino e raça branca (542,10), 34 a menos que o sexo masculino e raça preta (551,10) e 73 pontos a menos que sexo masculino e raça branca (590), todos no terceiro quartil.

Isso mostra uma realidade muito difícil para estas alunas, apesar de estarem entre as 25% melhores notas considerando as três variáveis em análise.

O mesmo ocorre quando se considera a escolaridade materna do ensino médio em diante onde novamente se destacam os alunos do sexo masculino e de raça branca, chegando a diferenças mínimas em torno de 25 pontos para o primeiro quartil e quase 42 pontos no terceiro, em comparação aos demais. A próxima e última análise dessa seção será segmentando sexo e raça por renda familiar, o que pode ser observado na Tabela 4.9.

Tabela 4.9 – Desigualdades em desempenho segmentadas por sexo, raça e renda familiar.

Renda	Meninas		Meninos	
	Brancas	Pretas	Brancos	Pretos
Renda Baixa (<3.748,00)	1º Quartil	432,30	417,60	456,10
	Mediana	489,30	467,00	527,30
	Média	497,87	477,60	532,58
Renda Média (3.748,00 - 7.496,00)	3º Quartil	553,60	527,10	601,80
	1º Quartil	469,50	441,10	493,00
	Mediana	536,25	507,80	572,80
Renda Alta (> 7.496,00)	Média	541,68	515,07	573,98
	3º Quartil	607,02	579,15	648,40
	1º Quartil	494,80	456,82	526,80
	Mediana	579,40	532,60	619,00
	Média	581,13	538,79	618,25
	3º Quartil	655,80	602,35	703,90
				632,85

Fonte: Própria autoria, 2021.

O padrão de desempenho é similar ao encontrado na análise anterior. Alunas do sexo feminino e de raça preta possuem o pior desempenho, independente da renda. Alunos do sexo masculino e raça branca tiram as notas mais altas ($p < 0,01$) e há equilíbrio entre os desempenhos dos alunos do sexo masculino de raça preta com o sexo feminino de raça branca. Para alunos de renda alta, os brancos tiram notas mais altas que os pretos, independente do sexo. As diferenças médias entre alunos de renda alta e de renda baixa são: 83,26 (581,13 – 497,87) entre sexo feminino de raça branca; 61,19 entre sexo feminino de raça preta; 85,67 entre sexo masculino de raça branca; 60,21 entre sexo masculino de raça preta. As diferenças são maiores comparando alunos de renda alta e baixa em comparação a escolaridade materna. Observe que as duas variáveis são importantes e estes resultados indicam relação entre elas e a proficiência.

Os estudos apresentados anteriormente permitem mostrar a variação das notas dos alunos em Matemática no ENEM entre as categorias de uma ou mais variáveis individuais. Os resultados refletiram as desigualdades de gênero, raça, renda e escolaridade materna. Por conta disso, a modelagem estatística a ser realizada leva em consideração essas variáveis.

4.3.2 RESULTADOS DOS MODELOS ESTATÍSTICOS E DISCUSSÕES

Os resultados para os efeitos fixos e aleatórios dos cinco modelos que foram ajustados se encontram na Tabela 4.10. As células em branco indicam que a variável não foi contemplada pelo modelo em questão. Todas as variáveis de efeitos fixos, assim como todas as componentes de efeitos aleatórios, são consideradas significativas em todos os modelos em nível de 5%. Os códigos em R referentes a estes resultados encontram-se no Apêndice B. No Apêndice C, pode-se encontrar a Tabela C2, que inclui os intervalos de confiança, os p-valores e os desvios padrão para os componentes da variância.

Tabela 4.10 – Resultados dos modelos 1 a 5 para efeitos fixos e aleatórios.

ENEM - Matemática					
	Escolas PÚblicas	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Variáveis					
Efeito Fixo					
Constante	468,02	469,54	471,43	463,31	465,11
Preta	-17,96	-17,92	-17,9	-14,90	-14,86
Parda	-9,40	-9,32	-9,36	-7,13	-7,15
Amarela	-5,12	-5,02	-5,06	-2,74	-2,77
Indígena	-20,89	-20,81	-20,85	-16,99	-17,02
Não declarado	-7,80	-7,74	-7,79	-5,39	-5,43
Meninas	-31,21	-31,2	-31,25	-28,84	-28,9
Nse	17,07	16,56	15,93	12,94	12,31
Ensino Médio ou mais				13,34	13,35
Renda Familiar				5,82	5,81
Componentes da variância					
Efeito Aleatório					
Variância do Nível 1 (aluno)	6516,6	6514,38	6509,41	6392,01	6387,39
Variância do Nível 2 (escola)	843,9	844,38	846,25	688,17	689,5
Variância do Nível 3 (município)	137,6	141,36	184,15	142,41	184,95
Preta/Município		22,68		19,76	
Meninas/Município			31,51		28,65
Ensino Médio ou mais/Município					
Renda/Município					

Nota: Preta/Município - fonte de variação do coeficiente de raça Preta entre os municípios.

Fonte: Própria autoria, 2021.

Considerando os modelos de 1 a 3, em termos de efeitos fixos, são analisados a seguir os impactos da raça e do sexo do aluno em sua proficiência no ENEM, controlando a ação escolar através do nível socioeconômico escolar médio (NSE). Pode-se observar que todas as raças tiram notas inferiores aos alunos brancos, com o pior cenário relacionado aos alunos indígenas, com quase 21 pontos a menos. Nas comparações entre brancos, pardos e pretos, observa-se que alunos pardos têm desempenho inferior a alunos brancos, e alunos pretos possuem desempenho inferior aos pardos, com uma diferença em torno de 18 pontos entre brancos e pretos. Para o sexo feminino, as alunas tiram notas em torno de 31 pontos a menos que os alunos do sexo masculino. O nível socioeconômico (NSE) impacta positivamente, com um aumento entre 16 e 17 pontos em cada nível.

Os modelos 4 e 5, comparativamente aos anteriores, incluem as variáveis socioeconômicas dos alunos, representadas por escolaridade materna (ensino médio ou mais) e renda familiar. Observa-se que os impactos da raça, do sexo e do nível socioeconômico médio escolar permanecem significativos, suas relações com a proficiência seguem o mesmo perfil de importância, entretanto, há uma pequena perda de magnitude. Alunos de raça preta tiram, em média, 15 pontos a menos que os de raça branca, para o sexo feminino o resultado fica em torno de 29 pontos a menos comparado aos do sexo masculino e, finalmente, o nível socioeconômico impactou entre 12 e 14 pontos a cada pontuação do NSE. Com o acréscimo da escolaridade materna e da renda familiar, tem-se um impacto entre 13 a 14 pontos para alunos que têm mães com escolaridade igual ou superior ao ensino médio e em torno de 6 pontos a cada R\$ 1.000,00 para a variável renda familiar. Deve-se levar em consideração que a amplitude da renda é em torno de R\$ 20.000,00, o que pode resultar em torno de 120 pontos de diferença na prova.

Quanto aos efeitos aleatórios, há uma redução do valor do intercepto da escola ao controlar os modelos pelas variáveis socioeconômicas (modelos 1 a 3 contra 4 a 5), o que indica modelos com médias menores atribuídas às escolas a partir deste controle. Também há redução nas variações dos coeficientes aleatórios, com a componente relacionada à cor preta por município reduzida de 22,68 para 19,76 (modelos 2 e 4) e a relacionada a sexo por município reduzindo de 31,51 para 28,65 (modelos 3 e 5), após o controle. Cabe ressaltar que, entre os modelos 4 e 5, o coeficiente de correlação intraclasse, ICC (HOX, 2010), que se trata da proporção da variância explicada em cada nível do modelo e pode ser usado como medidas para o efeito escola e o efeito dos municípios, fica entre 9% e 10% para o efeito escola e entre 2% e 3% para o efeito dos municípios, mostrando que a maior parte do impacto das notas se dá por fatores

externos ao controle de escolas e municípios, levando-se em consideração as variáveis estudadas.

Sabe-se da complexidade de medir quais são todas as possíveis variáveis que impactam os resultados escolares. Muitas delas vão além das escolas, como por exemplo, as variáveis individuais de cada aluno, o que contribui para que o impacto da escola seja menor do que o esperado. Mesmo assim, sua importância é ressaltada em SIMIELLI (2015): “a escola, e seus recursos e processos internos, ainda são centrais porque são possíveis de serem modificados a partir das políticas públicas”.

Para uma melhor visualização dos resultados encontrados a partir dos modelos, são apresentadas duas figuras relacionadas à qualidade *versus* equidade para os modelos 4 e 5. O objetivo dessa análise é avaliar o quanto equidade e qualidade caminham juntas ou se o aumento da equidade está relacionado à diminuição da qualidade. Em cada uma das figuras, o eixo x corresponde aos valores da medida de equidade obtidas pelos municípios, através das estimativas dos efeitos aleatórios dos modelos 4 e 5. Por exemplo, um município que possua seu valor no eixo x igual a zero, tem seu coeficiente aleatório exatamente igual ao efeito fixo estimado pelo modelo. Já um município que possua seu valor no eixo x igual a 2, tem seu coeficiente aleatório igual à soma do efeito fixo com 2. Para o eixo y, considera-se a média dos alunos do município em Matemática no ENEM.

A análise do gráfico é feita da seguinte maneira: para valores do eixo x à direita de zero, encontram-se os municípios mais equitativos em relação à variável em discussão. Para a medida de qualidade, a comparação é feita a partir da nota média de todos os alunos da base de dados no exame de Matemática do ENEM 2017, e está representada por uma reta horizontal em cada figura. Municípios que estão acima desta reta possuem qualidade acima da média. Em cada figura, mostram-se os 30 melhores e os 30 piores municípios em relação à equidade em questão. Foi realizada, em cada figura, uma segmentação por Região, a fim de que seja informado se há alguma discrepância entre as regiões para os municípios mais ou menos equitativos. No Apêndice C, pode-se encontrar a Tabela C3 com o nome e os respectivos efeitos para cada município.

Na Figura 4.9, tem-se a análise de qualidade *versus* equidade para os municípios em relação aos alunos de raça preta, referentes ao modelo 4. Quanto maior o valor do eixo x, mais equidade em relação a variável há para o município, já que, ao somar com o efeito médio de -14,90 (Tabela 4.10), o resultado ficaria mais próximo a zero, o que representa maior equidade. A informação mais impactante é que, enquanto a maioria dos municípios com menos equidade para a raça preta estão acima da média, o oposto ocorre com os municípios mais equitativos. Dentre os trinta municípios mais equitativos, se destacam Jundiaí e Goianésia, destacados por setas na Figura 4.9, que possuem as

melhores medidas de qualidade deste grupo. Em compensação, há alguns municípios com média abaixo de 480 pontos. Um tema já debatido na literatura é a dificuldade em se alcançar ao mesmo tempo qualidade e equidade, algo que já ocorre desde anos iniciais da educação básica e já esperávamos identificar neste recorte entre alunos pré-universitários. Há, também, maioria de municípios da Região Sudeste e nenhum representante das regiões norte e centro-oeste entre os mais iníquos em relação à raça.

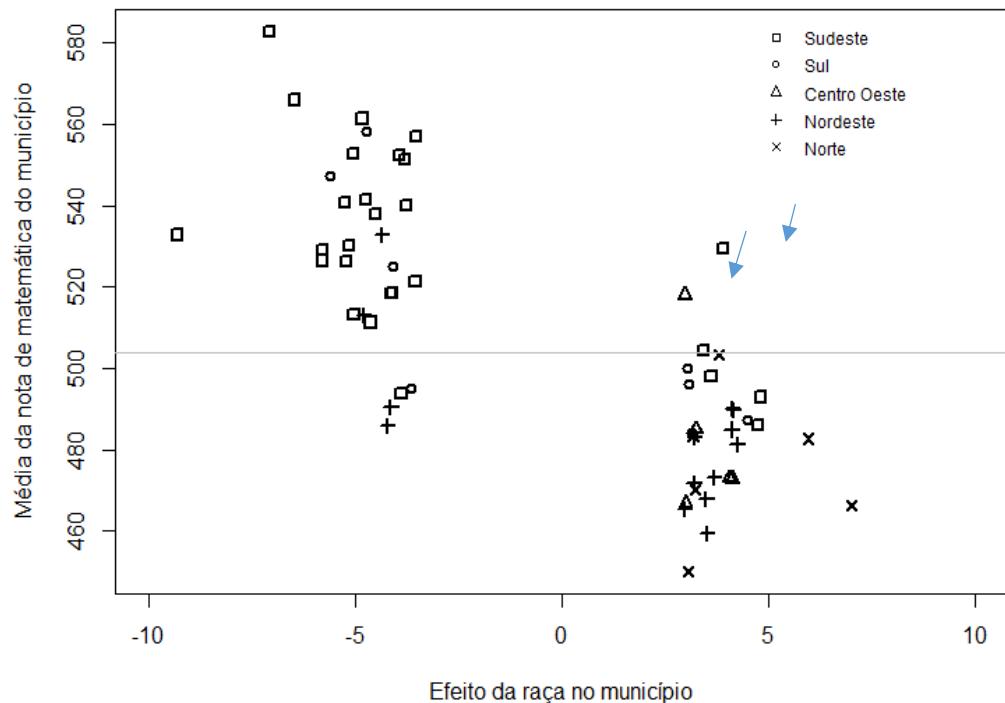


Figura 4.9 – Equidade x Qualidade para os 30 melhores e 30 piores municípios em relação à raça no ENEM 2017.

Fonte: Própria autoria, 2021.

Comportamento similar pode ser encontrado na Figura 4.10, na avaliação qualidade versus equidade para a variável sexo, referente ao modelo 5. Da mesma forma que na análise anterior, maiores valores no eixo x significam maior equidade. A maioria dos municípios mais iníquos possui desempenho acima da média, enquanto apenas duas cidades, dentre as mais equitativas, possuem média de notas superior à média geral, Jaguariúna e Franca, destacadas por setas na Figura 4.10. Novamente, muitos municípios alcançam equidade à custa de uma piora na qualidade. No grupo de menor equidade, há equilíbrio entre as Regiões Sudeste e Nordeste e algumas cidades da Região Sul. É importante considerar que, mesmo entre cidades mais equitativas, ainda há diferenças consideráveis para alunos do sexo masculino, que tiram em torno de 18

pontos a mais, comparados as do sexo feminino. Estas cidades estão em torno de 10 pontos acima do efeito médio da variável sexo feminino, representada pelo valor -28,9, como pode-se observar na Tabela 4.10. Já nas cidades mais iníquas, a diferença chega a ser em torno de 40 pontos. Isso acontece, pois o município mais iníquo pela figura abaixo está em torno de -12 pontos e é acrescido ao efeito médio de -28,90.

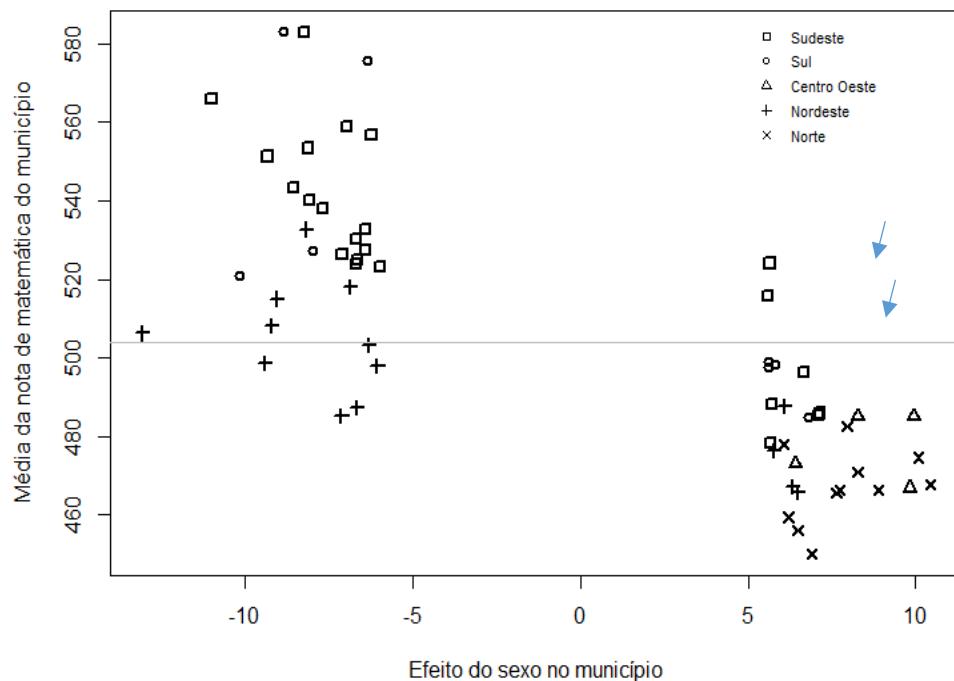


Figura 4.10 – Equidade x Qualidade para os 30 melhores e 30 piores municípios em relação a sexo no ENEM 2017.

Fonte: Própria autoria, 2021.

Observando as duas figuras em questão, destaca-se a concentração das melhores notas (acima da linha horizontal cinza) para municípios mais iníquos independente da variável em questão, sexo ou raça. A fim de tentar compreender melhor o resultado, uma análise levando em consideração todos os 595 municípios da base dados foi realizada para cada um dos efeitos das figuras anteriores. Considere uma divisão dos municípios em quatro grupos: acima ou abaixo da média geral (reta horizontal nos gráficos) e os mais ou menos equitativos (maior ou menor que o valor zero, no eixo x). Ao elaborar políticas públicas educacionais, o objetivo é que o município possua equidade positiva e esteja acima da média, uma vez que este grupo combina qualidade e equidade. A seguir, os perfis desses quatro grupos, em termos das variáveis aqui

utilizadas, são apresentados com o objetivo de contribuir para a discussão socioeconômica no âmbito de equidade.

Inicia-se a discussão sobre o perfil do grupo de interesse: o que combina qualidade e equidade. Trata-se do menor grupo dentre os quatro, com 88 municípios para raça e 87 para sexo. Seus municípios possuem, sistematicamente, maior proporção de alunos brancos, menor proporção de alunos negros, maiores renda e nível socioeconômico escolar e maior proporção de mães cuja escolaridade esteja acima do ensino médio. Perfil muito parecido com este grupo é o composto por municípios com qualidade acima da média, porém, mais iníquos no que diz respeito a sexo e raça. Estes municípios seguem o mesmo padrão dos anteriores, e, as principais diferenças estão em proporções um pouco menores em comparação ao primeiro grupo e com uma maior amplitude das variáveis analisadas. A amplitude das variáveis indica que, em todos os grupos, existem municípios com diferentes características. Isso significa que, quem combina qualidade e equidade, possui populações mais homogêneas, ou seja, com menores amplitudes entre as variáveis, seguindo o mesmo perfil considerando raça, renda e escolaridade materna.

Considerando os municípios abaixo da média geral do ENEM, observam-se proporções bem menores de alunos brancos, um pequeno aumento da proporção de alunos pretos, renda e nível socioeconômico escolar menores que os grupos anteriores. Para o grupo abaixo da média, mas equitativo, a escolaridade materna é ligeiramente menor que os grupos de maior qualidade. Já o grupo que combina baixas qualidade e equidade, possui menores percentuais de mães cuja escolaridade é acima do ensino médio.

As análises apresentadas podem ajudar nas propostas de políticas públicas. A associação entre maior renda, maior proporção de alunos brancos e baixa proporção de alunos negros, com resultados de melhor qualidade e equidade, mostra uma realidade incompatível com boa parte dos municípios brasileiros. É possível que a realidade das famílias dos alunos pretos nos municípios que se destacam seja muito diferente da maioria das famílias pretas no Brasil.

É razoável supor que questões relacionadas à renda familiar tenha relevância na discussão qualidade *versus* equidade. A literatura em avaliação educacional reconhece o capital econômico familiar, aqui representado pela variável renda, como importante e a sua influência para os demais capitais associados ao ambiente familiar e socioeconômico do aluno (BONAMINO *et al.*, 2010). Em termos de políticas públicas, uma possibilidade seria tentar suprir a ausência desses capitais através de iniciativas governamentais e projetos para melhorar o sistema público educacional.

Apesar do escopo deste trabalho não incluir a elaboração de políticas públicas educacionais, mas sim fornecer subsídios para tal, projetos como o de uma escola integral (GONÇALVES, 2006), onde o aluno estude, se alimente, pratique esportes, tenha acesso à internet e biblioteca ao longo do dia, possa ser interessante. Nesse caso, a escola ocuparia um papel e ofereceria infraestrutura, fora do horário de aula, que as famílias não poderiam oferecer aos seus filhos. Melhorar a infraestrutura das escolas, desde materiais pedagógicos apropriados, como computador e internet, até alimentação para todos os alunos, requer uma política com investimento.

É necessário também oferecer vagas para toda a população de baixa renda. A qualificação do corpo docente, com salários adequados e projeto de carreira, também é fundamental para contar com professores que contribuam para a vida escolar do aluno. Oferecer educação pública dessa natureza, e em todas as escolas, poderia amenizar as diferenças por renda e fatores socioeconômicos que observamos.

Finalmente, ainda que esse tipo de projeto fosse implementado, seria necessário um período de transição, para as gerações que iniciaram a vida escolar em diferentes realidades. Políticas de cotas deveriam ser mantidas, na tentativa de atenuar as desigualdades sociais, até que não fossem mais necessárias.

Do ponto de vista da discrepância entre as notas de Matemática de alunos e alunas no ENEM, políticas de incentivo às meninas nas escolas podem ser realizadas. Eventos de olimpíadas de Matemática, divulgação científica, entre outros, podem ser efetivos e incentivar a participação de meninas na Matemática, a fim de diminuir essa diferença que, infelizmente, vem ocorrendo e é debatida na literatura.

Os resultados aqui apresentados em relação à equidade reforçam o debate já citado neste trabalho sobre a dificuldade em conciliar qualidade e equidade em educação. Ao escolher o ENEM, é como se fossem potencializados todos os problemas acumulados em relação ao aprendizado por toda a vida escolar do aluno, avaliando em um único exame, ao final do ensino médio, as diferenças equitativas. Trata-se de uma consequência de toda a desigualdade e iniquidade presente desde a educação básica dos estudantes e que acaba sendo refletida na população pré-universitária.

Pode-se observar também que, ainda com o exame aplicado apenas ao subconjunto da população que decidiu prestar o ENEM, os resultados aqui obtidos são exatamente os esperados e, infelizmente, há problemas equitativos que variam entre os municípios brasileiros considerando raça e sexo.

5. VARIÁVEIS FAMILIARES EM TEMPOS DE PANDEMIA: UMA REFLEXÃO A PARTIR DE DADOS PRÉ-PANDÊMICOS DO ENEM

Um dos objetivos de políticas públicas educacionais é o desenvolvimento de um sistema escolar que combine qualidade e equidade, sendo justo em igualdade de oportunidades para diferentes grupos sociais e alcançando excelência de desempenho. Atualmente, o impacto causado pela pandemia da Covid-19 na educação afetou boa parte dos estudantes em todo o mundo. O objetivo deste capítulo é identificar, a partir de análises descritivas, as relações entre algumas variáveis tipicamente utilizadas em avaliação educacional e o desempenho no exame em matemática do ENEM. Fatores familiares, como escolaridade materna e renda, e de infraestrutura escolar serão segmentados por Região. As análises apresentadas indicam relação entre estes fatores familiares e o desempenho, o que abre a discussão sobre o quanto a ausência de aulas presenciais poderá interferir no desempenho de alunos de baixa renda de escolas públicas em todo o Brasil. Os resultados mostram a desigualdade social no país e os desafios que estão por vir para que não haja um abismo ainda maior do que o atual no que tange ao acesso à educação superior gratuita no Brasil. Os estudos aqui apresentados podem servir como base para gestores na discussão sobre políticas públicas educacionais.

5.1 CONSIDERAÇÕES INICIAIS

A pandemia causada pelo novo Coronavírus (WU, 2020), cuja doença (Covid-19) mudou a realidade de bilhões de pessoas ao redor do planeta, impôs restrições que impactaram a vida de todos. O impacto educacional foi enorme uma vez que, em virtude de medidas de prevenção, as instituições educacionais de diversos países ficaram fechadas (UNESCO, 2020a). No mês de maio de 2020 havia, segundo a UNESCO (UNESCO, 2020a), 1.198.530.172 estudantes de todos os níveis (pré-primário, primário, secundário e terciário) afetados em todo o mundo, o que representava 68,5% do total de alunos no mundo. Modelagens para medir o impacto de transmissão em um contexto similar ao atual com o fechamento de escolas foram propostas em KAWANO e KAKEHASHI (2015) e LUCA *et al.* (2018). O fechamento de escolas em epidemias e pandemias é, usualmente, uma medida considerada eficaz para a diminuição do contágio (KAWANO e KAKEHASHI, 2015), mas trata-se de um assunto controverso na comunidade científica e sua efetividade vem sendo questionada na literatura (LUCA *et al.*, 2018).

Recentemente, uma revisão sistemática foi apresentada já no contexto da Covid-19 (VINNER *et al.*, 2020), debatendo algumas evidências encontradas e sugerindo que combinações de medidas de distanciamento social devem ser consideradas. Embora haja uma óbvia importância desse debate, em uma situação como a atual, a tendência é que as medidas de isolamento sejam adotadas pelos governantes com o objetivo de mitigar o máximo possível os efeitos da pandemia.

Com uma grande quantidade de alunos de todo o mundo sem aulas presenciais, esforços vêm sendo feitos por educadores, professores, gestores e familiares em busca de alternativas para seguir remotamente o processo educacional. Em países com alto desenvolvimento econômico, pode-se esperar que uma boa parcela de sua população possua as condições financeiras e estruturais de oferecer aos seus filhos as possibilidades para o aprendizado em casa. Entretanto, este cenário não é compatível com a realidade de famílias de várias partes do mundo. Apenas para citar um único fator: o acesso digital, segundo a UNESCO (2020b), o relatório The State of Broadband de 2019, mostra que há apenas 54,8% dos domicílios no mundo conectados à internet. Em termos de cobertura de rede móvel, houve uma melhora de apenas 22% na cobertura 4G nos últimos 5 anos (até 2019) nos países de baixa renda em comparação a um aumento de 66% nos países de renda baixa a média. Se a adaptação à nova realidade da educação mundial será online ou parcialmente online e se a democratização da educação já é algo distante das populações de baixa renda (CAREGNATO e OLIVEN, 2017), torna-se necessário um debate sobre como tornar a educação a mais democrática possível.

O despertar de novas necessidades educacionais que surge em um ambiente pós-pandêmico parece ser incompatível com as famílias que convivem com altas taxas de desigualdades sociais, altos índices de pobreza, vivendo em condições de miséria, sem saneamento básico e acesso à educação para seus filhos. A discussão em termos educacionais de democratização da educação, com internet de qualidade, softwares especializados, inteligência artificial e sistemas de recomendações está relacionada com a origem social do indivíduo. O objetivo central deste capítulo é, a partir de análises descritivas, compreender melhor algumas variáveis escolares e familiares em um dos exames que permitem o acesso a universidade pública no Brasil, levando em consideração o exame em Matemática dos alunos. A partir destes resultados, discutimos o quanto a ausência de aulas presenciais poderá influenciar no desempenho de alunos de baixa renda.

A utilização do ENEM como ferramenta de estudo para aplicação de políticas públicas já foi previamente estudada em ALMEIDA *et al.* (2017). Além disso, por possibilitar ao aluno o acesso ao ensino superior, muitas famílias escolhem as escolas

que matricularão seus filhos levando em consideração a preparação que esta escola pode proporcionar a criança para o ENEM (SAUER e SARAIVA, 2019). Os dados utilizados são de alunos de escolas públicas brasileiras, divididos em cidades de todas as regiões do país: Norte, Nordeste, Centro-Oeste, Sul e Sudeste com população entre 50 e 500 mil habitantes.

5.2 DESAFIOS PARA A EDUCAÇÃO EM TEMPOS DE PANDEMIA

O desafio atual da educação mundial é superar as barreiras do isolamento social e prover uma educação de qualidade: (i) técnica, ensinando os conteúdos necessários; (ii) pedagógica, a ponto de utilizar as ferramentas tecnológicas a fim de criar um ambiente que proporcione a criança e o adolescente o interesse pelo saber, e (iii) humana, sendo capaz de lidar com todos os atores deste processo (alunos, professores e educadores em geral, pais e gestores) em uma situação de estresse contínuo para todos.

É interessante considerar que, do ponto de vista educacional, um debate recorrente nos últimos anos é sobre a escola como a conhecemos. A mudança ocorrida no mundo nas últimas décadas, com a evolução tecnológica, já tinha alterado a maneira como nos relacionamos, como compramos e como agimos diariamente. A escola, entretanto, continua, de forma geral, muito parecida com a de décadas atrás.

Curiosamente, neste momento há a necessidade dos alunos aprenderem fora de sala de aula. Educadores têm recorrido à utilização de softwares que permitem aula à distância e a maior interatividade possível com seus alunos. Este processo, que não sabemos quanto tempo durará, provavelmente acarretará em mudanças definitivas em muitas instituições. Toda a discussão que ocorre há anos, especificamente na comunidade de educação, passa a fazer parte do cotidiano do mundo inteiro. Há uma grande necessidade de um novo paradigma de ensino-aprendizagem que dê mais protagonismo aos alunos e a utilização de recursos digitais contribui para a construção deste novo paradigma. Com o passar do tempo, pode haver uma democratização do ensino através da internet, uma vez que há possibilidade de trabalhos colaborativos e, possivelmente, os casos de sucesso não ficarão restritos a uma só comunidade, mas serão compartilhados mundialmente.

Embora o contexto indique que as restrições impostas pela pandemia contribuem para que um novo cenário seja construído (o que pode ser animador dentro de algum tempo) todo esse discurso contrasta com a informação de que pouco mais de 45% da população mundial não possui acesso à internet no domicílio. No caso do Brasil, a informação sobre o percentual da população que possui computador e internet no

domicílio, segmentado por Região, pode ser encontrado na Tabela 5.1, cuja fonte é CETIC (2017).

Tabela 5.1 – Distribuição de acesso, em casa, de computadores e internet por Região em, porcentagem (%).

Regiões	Apenas Computador	Apenas Internet	Nem Computador, nem Internet	Computador e Internet
Sudeste	5	18	26	51
Nordeste	5	21	46	28
Sul	5	16	35	44
Norte	6	26	46	22
Centro-Oeste	4	20	29	48

Fonte: CGI.br/NIC.br (CETIC, 2017)

A Região que possui maior porcentagem de domicílios com computador e internet é a Sudeste, com apenas 51%. Por outro lado, as Regiões Norte e Nordeste possuem apenas 22% e 28%, respectivamente. Domicílios com apenas internet provavelmente possuem acesso a partir de um celular, o que ainda assim dificultaria o processo de aprendizagem. O número de domicílios sem computador nem internet é preocupante; quase metade dos domicílios das regiões Norte e Nordeste não possuem ambos. O melhor cenário é na Região Sudeste, onde um a cada quatro domicílios não possui acesso a computador e internet. Este levantamento pode proporcionar um temor considerando o futuro da educação no Brasil ou em países que possuam um perfil similar, considerando que o isolamento social pode seguir até que toda a população seja vacinada.

5.3 MATERIAIS E ANÁLISE DESCRIPTIVA

O banco de dados é composto por estudantes de escolas pertencentes às cinco regiões brasileiras: Norte, Nordeste, Centro-Oeste, Sul e Sudeste que pleitearam acesso à educação superior no Brasil através do Exame Nacional do Ensino Médio (ENEM), em 2017. A base é similar à da análise de equidade, contemplando um total de 375.670 estudantes de 5.897 escolas.

Dentre as variáveis usualmente consideradas pela literatura em avaliação educacional, utiliza-se, neste trabalho, escolaridade da mãe e renda familiar, como fatores familiares (BARROS *et al.*, 2006, REIS e RAMOS, 2011, BASSETTO, 2019). Para fatores escolares, consideramos a escola possuir biblioteca e sala de leitura, a fim

de ressaltar a importância da infraestrutura escolar para os alunos (SOARES, 2005, FRANCO e BONAMINO, 2005). Quanto ao desempenho escolar, as notas em Matemática no ENEM foram utilizadas (FRANCO *et al.*, 2005, RODRIGUES *et al.*, 2011, OLIVEIRA e CARVALHO, 2018). Em relação às variáveis familiares, estas também fizeram parte do estudo de BASSETTO (2019) para medir desempenho em Matemática de alunos do 3º ano do ensino médio. O critério para se utilizar a escolaridade da mãe é que esta variável está relacionada à chance de indivíduos entre 15 e 24 anos estarem estudando (KUBOTA, 2019), faixa etária que inclui a população pré-universitária. Segundo LONGO e VIEIRA (2017), a associação entre a escolaridade da mãe e a condição de vida dos filhos também pode se refletir no nível de escolaridade que eles alcançam. Já a renda familiar, exemplo de capital econômico da família, é uma variável frequentemente relacionada ao desempenho (BASSETTO, 2019). A biblioteca e a sala de leitura são ambientes onde, provavelmente, a sua utilização pelos alunos está restrita em tempos de pandemia. A distribuição por Região pode ser encontrada na Tabela 5.2.

Tabela 5.2 – Distribuição dos estudantes e escolas por Região.

Regiões	Estudantes	Escolas
Sudeste	171.733 (46%)	2.978 (51%)
Nordeste	92.927 (25%)	1.028 (17%)
Sul	51.259 (14%)	1.094 (19%)
Norte	37.985 (10%)	425 (7%)
Centro-Oeste	21.766 (6%)	372 (6%)
Total	375670	5897

Fonte: Própria autoria, 2021.

Pode-se observar, na Tabela 5.2, que a Região Sudeste possui quase metade dos alunos e pouco mais da metade das escolas da base de dados. Seguido pela Região Nordeste quanto a número de estudantes. Vale relembrar que a Região Sudeste, de acordo com a Tabela 5.1, também foi a que apresentou maior porcentagem de domicílios com computador e internet. Quanto às regiões Norte e Centro-Oeste, há uma proporção pequena de alunos e escolas em comparação às demais regiões, considerando a base de dados aqui utilizada.

Na Figura 5.1, podem-se observar as distribuições das escolaridades maternas por Região do Brasil. Em torno de 50% das mães (46% no Centro-Oeste, 54% no Nordeste, 43% no Norte, 44% no Sudeste e 48% no Sul do país) possuem, no máximo, até o ensino fundamental. Quando o candidato ao ENEM preenche o questionário e seleciona à opção de resposta “Não Sei”, é provável que a escolaridade real das mães esteja

dentro desta faixa de escolaridade mais baixa, o que aumentaria mais um pouco a proporção citada anteriormente. Se considerarmos que, em um período de isolamento social, os candidatos terão que se preparar para o ENEM de casa, boa parte dos alunos em todas as regiões do Brasil dificilmente contaria com a ajuda materna sob o ponto de vista acadêmico. A maior parte em todas as regiões é de mães com ensino secundário e, apenas uma pequena parte, com ensino superior e pós-graduação.

A escolaridade dos pais é um fator que influencia a melhora na proficiência dos alunos, tema este muito debatido na literatura (SOARES e COLLARES, 2006, OLIVEIRA e SILVA, 2018). Em tempos de isolamento social, é possível que o impacto desta influência aumente na proficiência dos filhos. Pelos dados apresentados na Figura 5.1, pode-se esperar que a disparidade existente prejudicará em torno de 50% dos candidatos ao ENEM.

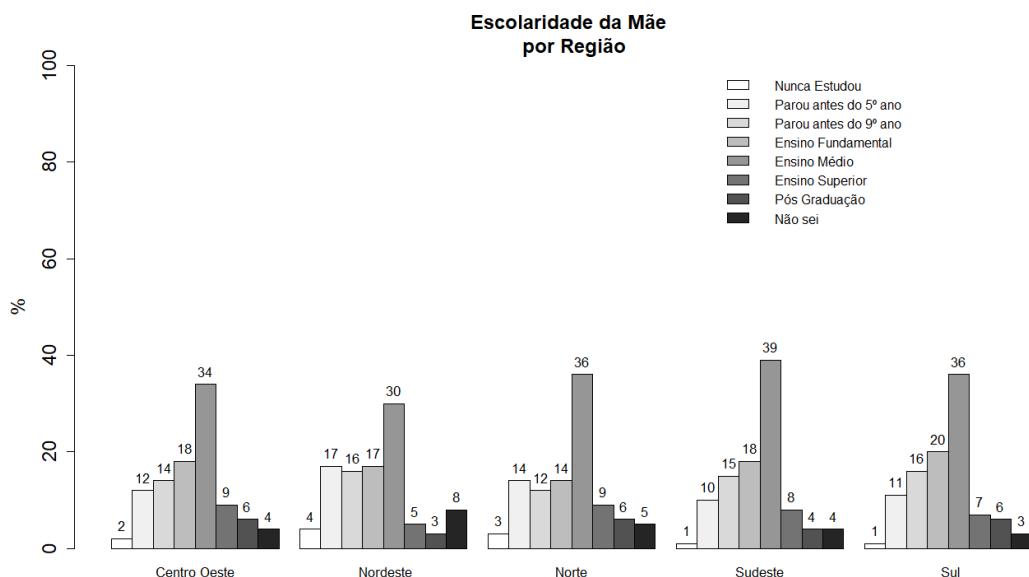


Figura 5.1 – Proporções dos níveis de escolaridade materna nas cidades entre 50 e 500 mil habitantes que tiveram alunos prestando o ENEM 2017, por Região do Brasil

Fonte: Própria autoria, 2021.

Na Tabela 5.3, pode-se observar a distribuição da renda familiar por Região nos quartis 25, 50, 75 e na média.

Mesmo considerando que se trata de um subconjunto da população, cujos filhos se candidataram ao ENEM 2017, observar a renda familiar indica as dificuldades que a população brasileira enfrenta. Metade das famílias da Região Nordeste vive com menos de R\$500,00 por mês e apenas 25% vivem com mais de R\$1.171,50. Mesmo na Região Sul, que possui a maior média (R\$ 2.382,00), 50% das famílias possuem renda igual ou

menor a R\$ 1.640,00. A partir dos dados da Tabela 5.3, pode-se concluir que é difícil para as famílias dos alunos de escolas públicas brasileiras se adequarem às condições de estudo a distância, onde são necessários dispositivos tecnológicos e internet de boa qualidade. Em 2017, o salário mínimo era de R\$ 937,00. Em 2021, com o salário mínimo de R\$ 1.100,00, 50% das famílias em todas as regiões possuem renda mensal inferior a dois salários mínimos. Neste trabalho, considera-se que famílias brasileiras de baixa renda são aquelas com rendimento familiar de até 3 salários mínimos (SEDES, 2018). Dessa forma, mais de 75% das famílias das regiões Sudeste, Centro-Oeste, Norte e Nordeste são de baixa renda, assim como mais de 50% das famílias da Região Sul.

Tabela 5.3 – Distribuição da renda familiar por Região nos quartis 25, 50, 75 e na média.

Regiões	25%	50%	Média	75%
Centro-Oeste	1172	1172	1814	2108
Nordeste	468,5	468,5	978,4	1171,5
Norte	468,5	1171,5	1353,5	1639,5
Sudeste	1172	1640	2047	2576
Sul	1172	1640	2382	3280

Fonte: Própria autoria, 2021.

Para uma análise a partir da proficiência em Matemática dos alunos no ENEM, pode-se observar, na Tabela 5.4, as médias em Matemática dos alunos por Região e escolaridade da mãe. Na primeira linha, encontram-se os alunos que responderam que não sabiam qual era o nível de escolaridade materno. É razoável esperar que, nesses casos, a escolaridade materna não seja alta. Isso pode ser corroborado pelas médias das notas dos alunos que responderam esta opção em todas as regiões, uma vez que se assemelham as médias dos alunos cujas mães possuem escolaridades mais baixas. A partir da segunda linha, se fixarmos a coluna, a tabela possui um padrão em que as médias aumentam conforme o aumento da escolaridade.

Tabela 5.4 – Médias das proficiências em matemática no ENEM 2017 por Região e escolaridade da mãe.

Nota por Escolaridade	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Não sei	474,64	470,39	465,25	488,81	494,22
Nunca Estudou	463,99	468,5	451,68	470,97	470,91
Parou antes do 5º ano	476,62	472,14	460,84	488,13	492,04
Parou antes do 9º ano	482,43	483,3	466,64	501,83	502,53
Ensino Fundamental	491,31	487,15	471,93	506,23	505,83
Ensino Médio	501,25	505,05	486,02	520,51	523,51
Ensino Superior	523,63	527,36	504,47	547,97	548,71
Pós Graduação	528,86	535,59	518,99	561,51	562,52

Fonte: Própria autoria, 2021.

A conclusão é que a escolaridade da mãe, que é uma variável associada ao capital cultural familiar, aparenta estar relacionada positivamente ao desempenho dos alunos. Na Região Sudeste, por exemplo, a média dos alunos cujas mães possuem pós-graduação é quase 20% maior em comparação a média dos alunos cujas mães nunca estudaram. As médias nas regiões Sul e Sudeste também são sistematicamente superiores às das outras regiões.

Na Tabela 5.5, pode-se verificar as médias de proficiências de alunos cujas famílias possuem renda maior ou menor que três salários mínimos, segmentadas por Região. Alunos de famílias com maior renda possuem maior média, mas aqui há um resultado diferente do encontrado na Tabela 5.4: Nordeste e Sudeste possuem as maiores médias. A importância de variáveis socioeconômicas no desempenho dos alunos é algo conhecido na literatura (BARROS *et al.*, 2006, BASSETTO, 2019) e, o que mostramos nessas segmentações, é que todo o conhecimento que possuímos em relação à educação básica se reflete no ENEM. Como em 2020 a educação foi a distância e os alunos não dispuseram da infraestrutura escolar para auxiliar seus estudos, é razoável esperar que o desempenho de alunos de baixa renda, cujas mães possuam baixa escolaridade, seja fortemente prejudicado no próximo ENEM.

Tabela 5.5 – Médias das proficiências em matemática no ENEM 2017 por Região considerando renda familiar inferior (<) ou superior (>) a três salários mínimos.

Regiões	<	>
Centro-Oeste	489,13	532,36
Nordeste	487,8	551,57
Norte	474,34	524,78
Sudeste	504,34	551,85
Sul	505,08	546,95

Fonte: Própria autoria, 2021.

Na Figura 5.2, tem-se as proporções para biblioteca e sala de leitura. Apenas nas escolas das cidades da Região Sudeste entre 50 e 500 mil habitantes onde houve candidatos ao ENEM há equilíbrio na proporção de escolas públicas com bibliotecas. Nas outras regiões, as porcentagens das escolas que não possuem biblioteca variam entre 3% e 20%. Ainda assim, há uma quantidade considerável de alunos que sequer possuem uma biblioteca na escola em todas as regiões. Quanto à sala de leitura, há menos escolas públicas que possuem este importante ambiente que, em conjunto com as bibliotecas, permitem estudo coletivo entre os alunos e a possibilidade de estudar na escola fora do horário da aula.

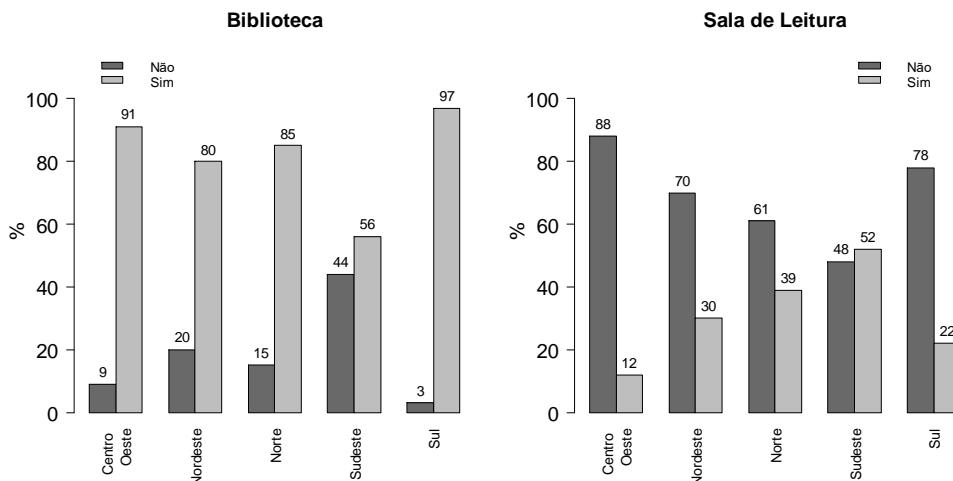


Figura 5.2 – Proporções de bibliotecas e salas de leitura nas cidades entre 50 e 500 mil habitantes que tiveram alunos prestando o ENEM 2017, por Região do Brasil.

Fonte: Própria autoria, 2021.

5.4 ENEM EM TEMPOS DE PANDEMIA

Consolidando as análises vistas anteriormente, o cenário para a educação brasileira, tendo em vista as necessidades impostas pela pandemia gerada pelo Coronavírus, tende a ser dramático e aumentar ainda mais a desigualdade social existente no país.

As análises descritivas apresentadas anteriormente indicam que a escolaridade materna de quase 50% das mães dos alunos do banco de dados aqui utilizado é até o ensino fundamental. Observamos que, independente da Região, há um aumento sistemático nas médias das notas dos alunos à medida em que a escolaridade aumenta. Os resultados indicam uma tendência de notas maiores para alunos das regiões Sul e Sudeste em comparação com as outras regiões. Considerando a renda familiar, alunos provenientes de famílias de renda superior a três salários mínimos possuem desempenho muito superior ao de alunos de renda inferior a três salários mínimos em todas as regiões. Alunos das regiões Nordeste, Sul e Sudeste possuem as maiores notas considerando a renda. A maioria das escolas possui biblioteca, mas boa parte delas não possui sala de leitura. Ainda assim, a não utilização dos acervos das bibliotecas e do seu espaço de estudo em termos de pandemia pode ser um fator importante para alunos de baixa renda que não possuem estrutura familiar para o seu estudo regular.

Os resultados já mostram uma disparidade entre os alunos. O agravamento proporcionado pela pandemia pode tornar essa discrepância ainda maior, causando um verdadeiro abismo social.

Os caminhos para conciliar as necessidades educacionais discutidas com as dificuldades sociais apontadas nesta seção incluem: (i) políticas de inclusão digital, com acesso a internet e recursos digitais, como computadores ou tablets para os alunos; (ii) apoio à carreira docente, qualificando os professores para ensino online; (iii) convênios com empresas e órgãos públicos com expertise em desenvolvimento de conteúdo digital. Além disso, os gestores devem desenvolver políticas que estimulem o acesso e a permanência escolar das crianças desde os primeiros anos de estudo até o final do ensino médio, aumentando a equidade entre os alunos e provendo ambientes escolares de excelência a fim de manter qualidade e equidade, características fundamentais para a eficácia escolar e que nem sempre são alcançadas, muitas vezes em virtude de políticas inadequadas.

6 CONCLUSÃO

Nesta tese propõe-se uma série de estudos a partir de variáveis individuais, familiares e escolares de alunos que pleitearam uma vaga no ensino superior no Brasil através do ENEM, no ano de 2017. Apresentam-se análises relacionadas à eficácia escolar no ENEM, agregando novas contribuições à literatura, principalmente através da mensuração do efeito escola a partir dos indicadores educacionais do INEP e uma análise sobre equidade e qualidade neste exame.

As principais contribuições desta tese podem ser resumidas da seguinte forma:

Uma breve análise crítica da literatura nacional de avaliação educacional foi feita a luz de contribuições internacionais consideradas importantes. O objetivo é situar a pesquisa realizada, discutindo algumas contribuições que vêm sendo obtidas ao longo de pouco mais de duas décadas de estudo no Brasil. É importante citar que não se propõe apresentar o estado da arte ou uma pesquisa exaustiva sobre avaliação educacional no Brasil.

Calculam-se o efeito escola no ENEM para alunos de escolas públicas e privadas que pertencem a cidades que possuem entre 50 e 500 mil habitantes na Região Sudeste do Brasil, usando por base o exame de Matemática. A principal contribuição relacionada ao efeito escola é mensurá-lo levando em consideração apenas os indicadores educacionais do INEP, o que permitiu analisar a representatividade destes indicadores no exame. O cálculo do ICC mostra que o efeito escola é de 9% para as escolas públicas e 13% para as escolas privadas. Estes valores são um pouco maiores que os encontrados para o ENEM anteriormente na literatura, o que significa que as *proxies* para qualidade docente, gestão escolar e nível socioeconômico médio das escolas apresentadas pelo INEP mapeiam essas informações de forma adequada considerando as variáveis normalmente usadas em avaliação educacional. A modelagem utilizada foi multinível (alunos, escolas), apropriada para dados educacionais. Como medir o efeito escola em estudos transversais é um tema discutido na literatura pela ausência de dados longitudinais e dificuldade em mensurar o conhecimento prévio do aluno, comparam-se as variáveis significativas do modelo com uma seleção de variáveis obtida pelo algoritmo Lasso. A conclusão é que houve uma similaridade grande entre as variáveis selecionadas pelo Lasso com as consideradas significativas pelo modelo multinível. Esta contribuição caracteriza os indicadores do INEP como boas *proxies* para as variáveis escolares ao utilizar o ENEM como medida de desempenho. É importante citar que não se consideraram, nesta tese, os indicadores educacionais como medidas de qualidade escolar, mas, sim, variáveis que mapeiam características individuais da

escola. A qualidade escolar é uma variável difícil de ser mensurada, que não está relacionada unicamente com indicadores ou medidas de desempenho, conforme discutido ao longo deste texto.

Posteriormente, dados relacionados à infraestrutura escolar são adicionados ao banco de dados e o objetivo é identificar as desigualdades de desempenho entre alunos de escolas públicas e privadas a partir da desigualdade social que existe entre estes dois grupos. Para isso, um modelo de regressão quantílica possibilita uma inferência sobre os alunos de desempenho alto, médio e baixo. Como a regressão quantílica não é multinível é feita uma comparação do quantil 50 com um modelo de dois níveis, metodologicamente equivalente ao proposto anteriormente, e os resultados são similares. Identificam-se diferenças entre os desempenhos dos alunos de escolas públicas e privadas, além de desigualdade entre as variáveis familiares e escolares. Do ponto de vista familiar, escolaridade materna e renda familiar possuem diferenças entre suas distribuições dependendo da dependência administrativa. Quanto às variáveis escolares, professores de escolas públicas possuem menor regularidade docente e maior esforço docente, o que faz com que, ao invés de conseguir amortizar as dificuldades familiares, a escola aumente a desigualdade entre alunos de escolas públicas e privadas.

Em seguida, um debate sobre equidade e qualidade no ENEM 2017 é apresentado. Foram utilizados dados apenas de alunos de escolas públicas, mas alunos de escolas pertencentes a municípios entre 50 e 500 mil habitantes de todo o Brasil foram considerados. Agrega-se a literatura ao apresentar uma análise intramunicipal para o ENEM, considerando equidade em relação à raça (especialmente na comparação entre alunos brancos e pretos) e sexo. Foram utilizados cinco modelos multiníveis (alunos, escolas, municípios), quatro deles com coeficientes aleatórios, a fim de medir a equidade. Os resultados indicam que os municípios mais iníquos para raça e sexo costumam ter qualidade acima da média, enquanto os mais equitativos geralmente estão abaixo da média. Este resultado está em linha com uma discussão relacionada a políticas públicas, onde desenvolver políticas educacionais na tentativa de melhorar os indicadores pode ser uma estratégia equivocada, já que pode promover a diminuição de equidade. Outra possibilidade seria desenvolver políticas com foco especial nos anos iniciais, dando condições de acesso e escolaridade de qualidade a alunos de diferentes grupos sociais. Com este processo se mantendo ao longo dos anos de escolaridade, a tendência é de diminuição de iniquidade entre os grupos. Se a qualidade for mantida, é possível que haja um sistema educacional eficaz, com qualidade e equidade simultaneamente. O desenvolvimento e análise de políticas públicas de educação estão

fora do escopo desta tese, apenas apresentamos subsídios para que estas decisões sejam tomadas.

Finalmente, apresenta-se uma reflexão sobre o momento atual em que vivemos, com a pandemia do Coronavírus e a dificuldade da população pré-universitária. O debate está inserido no contexto de aulas remotas e população de baixa renda e sem acesso a internet, o que pode levar a um hiato ainda maior entre classes sociais no Brasil.

Estas contribuições permitem avançar no conhecimento sobre o ENEM, que é um exame importante do ponto de vista social, uma vez muitos jovens de baixa renda buscam mobilidade social através do acesso ao ensino superior. Com o efeito escola em 9% e 13% para as escolas públicas e privadas, concluímos que boa parte da variação das notas diz respeito a fatores extraescolares. Apesar disso, não se deve considerar que não há importância escolar no desempenho dos alunos. A diferença entre as escolas públicas e privadas provavelmente é referente a uma melhor organização escolar do ponto de vista de gestão e docente, mas também está associada à questão socioeconômica, pois, em geral, escolas privadas possuem mais recursos do que as públicas. A iniquidade intramunicipal presente no ENEM fornece um retrato da dificuldade social em se conciliar equidade e qualidade do ponto de vista de variáveis individuais dos alunos e abre possibilidades para outros estudos que demonstrem a necessidade de adequação e aperfeiçoamento de nosso sistema educacional, em busca de prover para os nossos jovens ambientes equitativos e de qualidade, a fim de prepará-los para uma vida adulta digna e justa.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ALBERNAZ, A., FERREIRA, F., FRANCO, C., 2002, “Qualidade e equidade no ensino fundamental brasileiro”, *Pesquisa e Planejamento Econômico*, v.32, n. 3. Disponível: <http://repositorio.ipea.gov.br/handle/11058/4340>

ALMEIDA, D. C., PIERRE, V. H. L., COSTA, E. S., et al., 2017, “Enem: uma análise do efeito das políticas públicas educacionais em Alagoas e a comparação do resultado do ranking com os demais estados do Nordeste”, *Revista Brasileira de Política e Administração da Educação*, v. 33, n. 2, pp. 399 - 419. Disponível: <http://doi.org/10.21573/vol33n22017.70879>

ALVES, F., ORTIGÃO, I., FRANCO, C, 2007, “Origem social e risco de repetência escolar: interação raça-capital econômico”, *Cadernos de Pesquisa*, v.37, n.130, pp.161-180. Disponível: doi.org/10.1590/S0100-15742007000100008

ALVES, M. T. G., SOARES, J. F., XAVIER, F. P., 2016, “Desigualdades Educacionais no Ensino Fundamental de 2005 a 2013: Hiato entre Grupos Sociais”, *Revista Brasileira de Sociologia*, v. 4, n. 7, pp. 49-81. Disponível: <http://dx.doi.org/10.20336/rbs.150>

ALVES, M. T., SOARES, J. F., 2013, “Contexto escolar e indicadores educacionais: condições desiguais para a efetivação de uma política de avaliação educacional,” *Educação e Pesquisa*, v. 39, n. 1, pp. 177-194. Disponível: <https://doi.org/10.1590/S1517-97022013000100012>

AMÉRICO, B., LACRUZ, A., 2017, “Contexto e Desempenho Escolar: Análise das Notas na Prova Brasil das Escolas Capixabas por Meio de Regressão Linear Múltipla,” *Revista de Administração Pública*, v. 51, n.5, pp. 854-878. Disponível: <https://doi.org/10.1590/0034-7612160483>

ARRAES, R., MARIANO, F., 2019, “Decomposição quantílica incondicional dos diferenciais de desempenho entre alunos de escolas privadas e públicas profissionalizantes”, *Pesquisa e Planejamento econômico*, v. 49, n. 3. Disponível: <http://repositorio.ipea.gov.br/handle/11058/9770>

BARBETTA, P., ANDRADE, D., TAVARES, H., “Estudo de fatores associados através de regressão quantílica hierárquica”, *Estudo em Avaliação Educacional*, v.29, n.71, 2018. Disponível: <https://doi.org/10.18222/eae.v29i71.4973>

BARBOSA, M. E., FERNANDES, C., 2001, “A escola brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em Matemática dos alunos da 4a série.” Em C. Franco (org.) *Avaliação, Ciclos e Promoção na educação*, pp. 155-172.

BARTHOLO, T. L., KOSLINSKI, M. C., DE ANDRADE, F. M., et al., 2020, “Segregação Escolar e Desigualdades Educacionais no Início da Escolarização no Brasil”, *REICE. Revista Iberoamericana Sobre Calidad, Eficacia Y Cambio En Educación*, v. 18, n. 4. Disponível: <http://www.rinace.net/reice/numeros/arts/vol18num4/art3.htm>

BARROS, R.; FOGUEL, M.; ULYSSEA, G., 2006, “Desigualdade de Renda no Brasil: Uma Análise da Queda Recente. IPEA, Rio de Janeiro, 2006”, Disponível em: <<http://repositorio.ipea.gov.br/handle/11058/3249>>

BASSETTO, C. F., 2019, “Background familiar e desempenho escolar: uma abordagem com variáveis binárias a partir dos resultados do Saresp”, *Revista Brasileira de Estudos de População*, v.36 Disponível: <https://doi.org/10.1590/S0034-71402011000200004>

BATES, D., MACHLER, M., BOLKER, B., et al., 2015, “Fitting Linear Mixed-Effects Models Using lme4”, *Journal of Statistical Software*, v. 67, n.1, pp.1–48. Disponível: doi.org/10.18637/jss.v067.i01.

BERKOWITZ R., MOORE H., ASTOR R. A., et al., 2017, “A Research Synthesis of the Associations Between Socioeconomic Background, Inequality, School Climate, and Academic Achievement”, *Review of Educational Research*, v.87, n.2, pp.425-469. Disponível:<https://journals.sagepub.com/doi/abs/10.3102/0034654316669821?journalCode=rera>

BONAMINO, A., ALVES F., FRANCO C., et al., 2010, “Os Efeitos das Diferentes Formas de Capital no Desempenho Escolar: Um Estudo à Luz de Bourdieu e de Coleman”, *Revista Brasileira de Educação*, v. 15, n. 45, pp. 487-499.
Disponível: doi.org/10.1590/S1413-24782010000300007

BOURDIEU, P., PASSERON, J., 1975, *A reprodução: elementos para uma teoria do sistema de ensino*. Rio de Janeiro: Editora Francisco Alves.

BOWLES, S., 1970, "Towards an Educational Production Function", In: Education Income, and Human Capital, Studies in Income and Wealth, v. 35, pp. 11-70, National Bureau of Economic Research.

BOX, G., 1976, "Journal of the American Statistical Association", v. 71, n. 356, pp. 791-799. Disponível em: links.jstor.org/sici?&sici=0162-1459%28197612%2971%3A356%3C791%3ASAS%3E2.0.CO%3B2-W

BUTTARO, A. Jr., CATSAMBISS., 2019, "Ability Grouping in the Early Grades: Long-Term Consequences for Educational Equity in the United States", Teachers College Record, v.121, n.2, pp. 1-50. Disponível: <http://doi.org/10.1086/595667>

CAREGNATO, C. E., OLIVEN, A. C., 2017, "Educação superior e políticas de ação afirmativa no Rio Grande do Sul: desigualdades e equidade", *Educar em Revista*, n. 64, pp.171-187. Diponível: <http://doi.org/10.1509/0104-4060.47764>

CARVALHAES, F., RIBEIRO, C., 2019, "Estratificação horizontal da educação superior no Brasil: Desigualdades de classe, gênero e raça em um contexto de expansão educacional", *Tempo soc. [online]*, v.31, n.1, pp.195-233. Disponível: <http://dx.doi.org/10.11606/0103-2070.ts.2019.135035>.

CARVALHO, S., KASSOUF, A. L., 2009, "As despesas familiares com educação no Brasil e a composição de gênero do grupo de irmãos," *Economia Aplicada*, v. 13, n. 3., pp. 353-375. Disponível: <https://doi.org/10.1590/S1413-80502009000300001>

CARVALHO, S., FIRPO, S., 2014. "O regime de ciclos de aprendizagem e a heterogeneidade de seus efeitos sobre a proficiência dos alunos". *Economia Aplicada*, v.18, n.2, pp.199-214. Disponível: <https://doi.org/10.1590/1413-8050/ea374>

CASTRO, M. H., 2009, "A Consolidação da Política de Avaliação da Educação Básica no Brasil", *Meta Avaliação*, v.1, n.3. Disponível: <http://dx.doi.org/10.22347/2175-2753v1i3.51>

CETIC - Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação, 2017. Disponível em: <<https://www.cetic.br/pt/tics/domicilios/2017/domicilios/A4B/>>. Acesso em 04 junho 2021, 11:30.

Commission of the European Communities, 2006, "Efficiency and equity in European education and training systems", *Communication from the European Commission to the Council and to the European Parliament*. Brussels: COM 481.

COLEMAN, J., HOPKINS, J., CAMPBELL, E., et al., 1966, *Equality of Educational Opportunity*, Washington, DC: U.S. Government Printing Office.

CURI, A., MENEZES-FILHO, N., 2010, "Determinantes dos gastos com educação no Brasil," *Pesquisa e Planejamento Econômico*, v. 40, n. 1, pp. 1-8. Disponível: <http://repositorio.ipea.gov.br/handle/11058/5102>

CURI, A., MENEZES-FILHO, N., 2013, "Mensalidade escolar, background familiar e os resultados do Exame Nacional do Ensino Médio," *Pesquisa e Planejamento Econômico*, v. 43, n. 2, pp - 223-254. Disponível: <http://repositorio.ipea.gov.br/handle/11058/4858>

DIAZ, M. D., 2010, "Desigualdade de oportunidades no ensino médio: ENEM". Revista Economia & Tecnologia, v.6, n.3, pp.121-128. Disponível: <https://doi.org/10.5380/ret.v6i3.26960>

DUTRA, R., COELHO, A. C., DUTRA, G., 2019, "Indicadores Educacionais e Proficiência no ENEM: um estudo nos Institutos Federais do Brasil", *Meta: Avaliação*. v.11, n.31, 2019. Disponível: <http://dx.doi.org/10.22347/2175-2753v11i31.1781>

ESPINOSA, A.N.G., 2017, *Estimating the education production function for cognitive and non-cognitive development of children in Vietnam through structural equation modelling using Young Lives data base*. MSc in Quantitative Research Methods Institute of Education, University College London.

ERNICA, M., RODRIGUES, E., 2020, “Desigualdades educacionais em metrópoles: território, nível socioeconômico, raça e gênero”. *Educação e Sociedade*, v.41. Disponível: <https://doi.org/10.1590/ES.228514>

FARIA, E., GUIMARÃES, R., 2015, “Excelência com equidade: fatores escolares para o sucesso educacional em circunstâncias desfavoráveis”, *Estudos em Avaliação Educacional*, v.26, n.61, pp.192-215. Disponível: <https://doi.org/10.18222/eae266103115>

FERRÃO, M. E., FERNANDES, C., 2003, “O efeito-escola e a mudança - dá para mudar? Evidências da investigação brasileira”. *REICE - Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, v.1, n.1, pp.1-13. Disponível: <https://revistas.uam.es/reice/article/view/5343>

FERRÃO, M. E., BARROS, G., BOF, A., OLIVEIRA, A., 2018, “Estudo Longitudinal sobre Eficácia Educacional no Brasil: Comparação entre Resultados Contextualizados e Valor Acresentado”. *DADOS-Revista de Ciências Sociais*, v. 61, n.4, pp. 265-300. Disponível: <https://doi.org/10.1590/001152582018160>

FRANCO, C., BONAMINO, A., 2005, “A pesquisa sobre características de escolas eficazes no Brasil: breve revisão dos principais achados e alguns problemas em aberto”, *Revista Educação On-line*, n. 1, 2005. Disponível em: < <https://www.maxwell.vrac.puc-rio.br/7378/7378.PDF> >

FRANCO, C., ORTIGAO, I., ALBERNAZ A., et al., 2007, “Qualidade e equidade em educação: reconsiderando o significado de fatores intra-escolares”, *Ensaio: Avaliação de Políticas Públicas Educacionais*, v.15, n.55, pp. 277-298G. Disponível: doi.org/10.1590/S0104-40362007000200007

FRANCO, A. M., MENEZES-FILHO, N., 2008, “Os Determinantes da Qualidade da Educação no Brasil,” Tese de D.Sc, Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, SP, Brasil.

GONÇALVES, S. A., 2006. “Reflexões sobre educação integral e escola de tempo integral”, *Revista Cadernos cenpec.* v.1, n.2. Disponível: <http://cadernos.cenpec.org.br/cadernos/index.php/cadernos/article/view/136>

GREMAUD, A., FELICIO, F., BIONDI, R., 2007, “Indicador de Efeito Escola: uma metodologia para a identificação dos sucessos escolares a partir dos dados da Prova Brasil,” *Textos para Discussão*, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Ministério da Educação (MEC).

GRINER, A., 2012, *O argumento de inclusão enquanto política de acesso à universidade pública*. Dissertação Mestrado em Políticas e Gestão Públicas; Gestão Organizacional, Universidade Federal do Rio Grande do Norte, Natal.

HASSAN, I., 2014, *Hierarchical quantile regression*. Thesis Master of Science in Biostatistics, University of Alberta, Canada, 2014. Disponível em: <https://era.library.ualberta.ca/files/m326m3245/Hassan_Imran_201408_MSc.pdf> Disponível: em 05 de junho 2021.

HANUSHEK, E., A., 2007, *Education Production Functions*, Stanford: Hoover Institution, Stanford University.

HANUSHEK, E, A., 2015, “Time in Education: Introduction”. *The Economic Journal*, v.125, n.588, pp. F394–F396. Disponível: <https://doi.org/10.1111/ecoj.12266>

HANUSHEK, E. A., ETTEMA, E., 2017, “Defining productivity in education: Issues and illustrations”, *The American Economist* , v.65, n.2, pp.165-183.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., 2001, *The Elements of Statistical Learning*. New York, Springer Series in Statistics.

HASTIE, T., TIBSHIRANI, R, WAINWRIGHT, M., 2015, *Statistical Learning with Sparsity: The Lasso and Generalizations*. New York, Chapman & Hall/CRC.

HIPPE, R., ARAÚJO, L., DINIS DA COSTA, P. 2016, “Equity in Education in Europe”, Retrieved from Luxembourg (Luxembourg): Publications Office of the European Union. Disponível: <https://doi.org/10.2791/255948>

HOX, J., 2010, *Multilevel Analysis: Techniques and Applications*. New York: Routledge.

IBGE - Instituto Brasileiro de Geografia e Estatística. Estimativas populacionais para os municípios. 2021. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados>. Acesso em 04 junho 2021, 11:30.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *ENEM*. 2021a. Disponível em: <<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>>. Acesso em 04 junho 2021, 11:30.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *IDEB*. 2021b. Disponível em: <<http://ideb.inep.gov.br/>>. Acesso em 04 junho 2021, 11:30.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Censo Escolar. 2021c. Disponível em: <<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-escolar>>. Acesso em 04 junho 2021, 11:30.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Indicadores Educacionais. 2021d. Disponível em: <<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais>>. Acesso em 04 junho 2021, 11:30.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. 2021e. Disponível em: <<http://portal.inep.gov.br/conheca-o-inep>>. Acesso em 04 junho 2021, 11:30.

KAWANO, S., KAKEHASHI, M., 2015, "Substantial impact of school closure on the transmission dynamics during the pandemic flu H1N1-2009 in Oita, Japan", *PLoS One*, v.10, n. 12, 2015. Disponível: <http://doi.org/10.1371/journal.pone.0144839>

KEEVES, J., HUNGI, N., AFRASSA T., 2005, "Measuring value added effects across schools: Should schools be compared in performance?", *Studies in Educational Evaluation*, v. 31, pp. 247-266. Disponível: <https://doi.org/10.1016/j.stueduc.2005.05.012>

KOENKER, R., 2010, *Quantile Regression in R: A Vignette*. Cambridge. University Press 2010. Disponível: <https://doi.org/10.1017/CBO9780511754098.011>

KOENKER, R, BASSETT, G., 1978, *Regression Quantiles*. The Econometric Society, v.46, n.1, pp. 33–50, 1978. Disponível: <https://doi.org/10.2307/1913643>

KOENKER, R., HALLOK, K., 2001, “Quantile Regression”, *Journal of Economic Perspectives*, v.15, n.4, pp. 143–156. Disponível: <https://doi.org/10.1257/jep.15.4.143>

KUBOTA, L. C., 2019, O Peso do passado no futuro do trabalho: a transmissão intergeracional de letramento, Nota Técnica n. 54, Diretoria de Estudos e Políticas Setoriais de Inovação e Infraestrutura. Disponível em<<http://repositorio.ipea.gov.br/handle/11058/9527>>

KUZNETSOVA, A., BROCKHOFF, P., CHRISTENSEN, R., 2017, “lmerTest Package: Tests in Linear Mixed Effects Models”, *Journal of Statistical Software*, v. 82, n.13, pp.1–26. Disponível: <https://doi.org/10.18637/jss.v082.i13>

LACRUZ, J., AMERICO, L., CARNIEL, F., 2019, “Indicadores de qualidade na educação: análise discriminante dos desempenhos na Prova Brasil”, *Revista Brasileira de Educação [online]*. v.24. Disponível: <https://doi.org/10.1590/S1413-24782019240002>

LAROS, J., MACIANO, J., 2008, “Análise multinível aplicada aos dados do NELS 88”, *Estudos em Avaliação Educacional*, v.19, n. 40, pp. 263-78, 2008.

LEON, F., MENEZES-FILHO, N., 2002, “Reprovação, avanço e evasão escolar no Brasil,” *Pesquisa e Planejamento Econômico – IPEA*, v. 32.

LONGO, F. V., VIEIRA, J. M., 2017, “Educação de mãe para filho: fatores associados à mobilidade educacional no Brasil”, *Educação e Sociedade*, v.38, n.141, pp.1051-1071. Disponível: <https://doi.org/10.1590/es0101-73302017162420>

LUCA, G., KERCKHOVE, K., COLETTI, P., et al., 2018, “The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for

Belgium”, BMC Infect. Dis., v.18, n.1, pp. 18-29, 2018. Disponível:
doi.org/10.1186/s12879-017-2934-3

MARIONI, L., FREGUGLIA, R., MENEZES-FILHO, N., 2014, “Teacher quality and student achievement: Evidence from Brazilian longitudinal data”, In: *The Annual Meeting of The Latin American and Caribbean Economic Association*. Disponível: <https://repositorio.usp.br/item/002730893>

MASCI, C., LEVA, F., AGASISTI, T., et al., 2016, “Does class matter more than school? Evidence from a multilevel statistical analysis on Italian junior secondary school students”. *Socio-Economic Planning Sciences*, Elsevier, v. 54, pp 47-57. Disponível: <https://doi.org/10.1016/j.seps.2016.03.001>

MAAS, C., HOX, J., 2004, “D The influence of violations of assumptions on multilevel parameter estimates and their standard errors”. Computational Statistics & Data Analysis, v. 46, n.3, pp. 427-440. Disponível: <https://doi.org/10.1016/j.csda.2003.08.006>

McCULLOCH, C., NEUHAUS, J., 2011, “Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter,” Statistical Science, v. 26, n. 3, pp. 388-402. Disponível: <https://doi.org/10.1214/11-STS361>

MATOS, D., RODRIGUES, E., 2016, “Indicadores educacionais e contexto escolar: uma análise das metas do IDEB,” *Estudos em Avaliação Educacional*, v. 27, n. 66, pp. 662-688. Disponível: <http://www.repositorio.ufop.br/jspui/handle/123456789/9393>

MEC – Ministério da Educação, 2021, “*Exame Nacional do Ensino Médio 2017 tem 7,6 milhões de inscrições com 80,7% confirmadas.*”. Disponível em: <<http://portal.mec.gov.br/publicacoes-para-professores/30000-uncategorised/49491-exame-nacional-do-ensino-medio-2017-tem-7-6-milhoes-de-inscricoes-com-80-7-confirmadas>>. Acesso em 04 junho 2021, 11:30.

MELO, S., MORAIS, A., 2019, “Clima escolar como fator protetivo ao desempenho em condições socioeconômicas desfavoráveis”, *Cadernos de Pesquisa*, São Paulo, v. 49, n. 172, pp. 10-34. Disponível: <https://doi.org/10.1590/198053145305>

MENEZES-FILHO, N., 2007, "Os determinantes do desempenho escolar no Brasil". *Sumário Executivo*, Instituto Futuro Brasil, IBMEC São Paulo e Faculdade de Economia e Administração da Universidade de São Paulo. Disponível: <https://repositorio.usp.br/item/001624821>

MENEZES-FILHO, N., PAZELLO, E., 2007, "Do teachers' wages matter for proficiency? Evidence from a funding reform in Brazil," *Economics of Education Review*, Elsevier, v. 26, n.6, pp. 660-672. Disponível: <https://doi.org/10.1016/j.econedurev.2007.08.003>

MENEZES, D., MORAES, J., BATISTA, B., 2020, "Efeito escola na educação básica: observações a partir da perspectiva Bourdiesiana". *Revista Temas em Educação*, v. 29, n.1. Disponível: <https://doi.org/10.22478/ufpb.2359-7003.2020v29n1.47642>

MONL, D. H., 1989, "The Education Production Function: It's Evolving Role in Policy Analysis", *Educational Evaluation and Policy Analysis*, v.11, pp.31-45.

MORTIMORE, P., 1998, *The Road to Improvement: Reflections on School Effectiveness*. Lisse, Swets & Zeitlinger Publishers.

MOTTA, I. D., LOPES, H. M., 2012. "O sistema de cotas sociais para ingresso na universidade pública", *Revista do Instituto do Direito Brasileiro*, n. 11, pp. 6823-6857. Disponível: https://www.cidp.pt/revistas/ridb/2012/11/2012_11_6823_6857.pdf

NASCIMENTO C. A. de O., GUIMARÃES, S.,2020, "O Enem nas pesquisas acadêmicas: saberes e práticas docentes em foco", *Ensino Em Re-Vista*, 27(Especial), pp.1357-1379. Disponível: <https://doi.org/10.14393/ER-v27nEa2020-8>

PASSONE, E., 2019, "Gestão escolar e democracia: o que nos ensinam os estudos de Eficácia Escolar", *Laplace Em Revista*, v.5, n.2, pp.142-156. Disponível: <https://dialnet.unirioja.es/servlet/articulo?codigo=7053005>

OLIVEIRA, G, LIMA, A., JÚNIOR, S., et al., 2017, "Avaliação de eficiência das escolas públicas de ensino médio em Goiás: Uma análise de dois estágios," *Economia Aplicada*, v. 21, n. 2, pp. 163-181. Disponível: <https://doi.org/10.11606/1413-8050/ea118947>

OLIVEIRA, A. C., CARVALHO, C., 2018, “Gestão escolar, liderança do diretor e resultados educacionais no Brasil”, *Revista Brasileira de Educação*, v.23. Disponível: <https://doi.org/10.1590/s1413-24782018230015>

OLIVEIRA, L. F. B., TERRA, R., 2016 “Impacto do Programa Mais Educação em indicadores educacionais”. In: *Working Paper - International Policy Centre for Inclusive Growth*.

OLIVEIRA, A., SILVA, I., 2018, “ Indicadores educacionais no Ensino Superior Brasileiro: possíveis articulações entre desempenho e características do alunado,” *Avaliação: Revista da Avaliação da Educação Superior*, v.23, n.1, pp.157-177. Disponível: <https://doi.org/10.1590/S1414-40772018000100009>

REIS, M. C., RAMOS, L., “Escolaridade dos pais, desempenho no mercado de trabalho e desigualdade de rendimentos”, *Revista Brasileira de Economia*, v.65, n.2, 2011. Diponível: <https://doi.org/10.1590/S0034-71402011000200004>

REYNOLDS, D., SAMMONS, P., De FRAINE, B., et al., 2011, “Educational effectiveness research (EER): a state of the art review”. In: *Paper presented at the 24th Annual Meeting of the International Congressfor School Effectiveness and Improvement*, Limassol, Cyprus, pp. 4–7, Janeiro 2011.

REYNOLDS, D., SAMMONS, P., De FRAINE, B., et al. 2014, “Educational effectiveness research (EER): a state of the art review”, *An International Journal of Research, Policy and Practice*, v. 25, pp.197-230. Disponível: <https://doi.org/10.1080/09243453.2014.885450>

RIANI, J., RIOS-NETO, E., 2008, “Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros?”, *Revista Brasileira de Estudos de População*, v. 25, n. 2, pp. 251-259. Disponível: <http://www.scielo.br/pdf/rbepop/v25n2/v25n2a04.pdf>

RIEGLE-CRUMB C., GRODSKY E., 2010, “Racial-ethnic differences at the intersection of math course-taking and achievement”, *Sociology of Education*, v. 83, n.3, pp- 248–270. Disponível: <https://doi.org/10.1177/0038040710375689>

ROCHA, A., FUNCHAL, B., 2019, "Mais recursos, melhores resultados? As relações entre custos escolares diretos e desempenho no Ensino Médio," *Revista de Administração Pública*, v. 53, n. 2, pp. 291-309. Disponível: <https://doi.org/10.1590/0034-761220170175>

RODRIGUES, C. G., RIOS-NETO, E. L. G., PINTO, C. C. de X, 2011, "Diferenças intertemporais na média e distribuição do desempenho escolar no Brasil: o papel do nível socioeconômico, 1997 a 2005", *Revista Brasileira de Estudos de População*, v. 28, n. 1, pp. 5-36. Disponível: <https://doi.org/10.1590/S0102-30982011000100002>

ROSSETTO, C., GONÇALVES, F, 2015, "Equidade na educação superior no Brasil: Uma análise multinomial das políticas públicas de acesso", *DADOS – Revista de Ciências Sociais*, v. 58, n. 3, pp. 791-824. Disponível: doi.org/10.1590/00115258201559

SAEB - Sistema de Avaliação da Educação Básica Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>>. Acesso em 04 junho 2021, 11:30.

SALES, P. E. N., HEIJMANS, R. D., SILVA, C. E. G., 2017, "Análise multinível da transição estudantil do curso técnico para o ensino superior", *Estudos em Avaliação Educacional*, v. 28, n. 69. Disponível: <https://doi.org/10.18222/eae.v0ix.4095>

SAUER, M., SARAIVA, K. "Uma escola diferente do mundo lá fora", *Revista Brasileira de Política e Administração da Educação*, v.35, n.3, 2019. Disponível: <http://doi.org/10.21573/vol35n32019.95346>

SEDES – Secretaria de Desenvolvimento Social. Disponível em: <<http://www.sedes.df.gov.br/cadastro-unico/#:~:text=S%C3%A3o%20consideradas%20fam%C3%ADlias%20de%20baixa,R%24%202.811%2C00>>. Acesso em 04 junho 2021, 11:30.

SILVA, A., GOMES, A., 2018, "Avaliação educacional: concepções e embates teóricos", *Estudos em Avaliação Educacional*, v.29, n.71, 2018. Disponível: <https://doi.org/10.18222/eae.v29i71.5048>

SIMIELLI, L. E. R., 2015. Equidade Educacional no Brasil: Análise das oportunidades educacionais em 2001 e 2011, Tese de Doutorado, Escola de Administração de Empresas de São Paulo.

SIMIELLI, L. E., 2017, “Equidade e Oportunidades Educacionais: O Acesso a Professores no Brasil”, *Arquivos Analíticos de Políticas Educativas*, v. 25, n. 46.

SOARES, J. F., ALVES, M. T., OLIVEIRA, R., 2001, “O efeito de 248 escolas de nível médio no vestibular da UFMG nos anos de 1998, 1999 e 2000,” *Estudos em Avaliação Educacional*, n. 24, pp. 69-118. Disponível: <https://doi.org/10.18222/eae02420012201>

SOARES, T., 2003, “Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no Simave 2002,” *Estudos em Avaliação Educacional*, v. 28, pp. 103-123. Disponível: <https://doi.org/10.18222/eae02820032172>

SOARES, J. F., 2004, “O efeito da escola no desempenho cognitivo de seus alunos”. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, v. 2, n.2, pp. 83-104. Disponível: <https://revistas.uam.es/reice/article/view/5550>

SOARES, J. F., 2005, “O efeito da escola no desempenho cognitivo de seus alunos”. In: MELLO E SOUZA, Alberto (Org.). *Dimensões da avaliação educacional*, Petrópolis: Vozes, pp. 174-204, 2005.

SOARES, J. F., ALVES, M. T., 2003, “Desigualdades raciais no sistema brasileiro de educação básica. Educação e Pesquisa”, v.29, n.1, pp.147-165. Disponível: <http://doi.org/10.1590/S1517-97022003000100011>

SOARES, J. F., ALVES, M. T., 2013, “Efeitos de escolas e municípios na qualidade do ensino fundamental”, *Cadernos de Pesquisa*, v. 43, n. 149, pp. 492-517. Disponível: <https://doi.org/10.1590/S0100-15742013000200007>

SOARES, J. F., COLLARES, A. C., 2006, “Recursos familiares e o desempenho cognitivo dos alunos do ensino básico brasileiro,” *Dados*, v.49, no. 3, pp. 1-23. Disponível: <https://doi.org/10.1590/S0011-52582006000300007>

SOARES, S., SÁTYRO, N., 2008, "O impacto da infraestrutura escolar na taxa de distorção idade-série das escolas brasileiras de ensino fundamental - 1998 a 2005," *Texto para discussão do IPEA*, n. 1338.

TAVARES, P., PONCZEK, V., 2018, "Teacher pay and student performance: Evidence from Brazil," *Brazilian Review of Econometrics*, v. 38, n.2, pp. 179-219. Disponível: <https://doi.org/10.12660/bre.v38n22018.73437>

TRAVITZKI, R., BOTO, C., 2013, *ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar*, Tese de D.Sc., Universidade de São Paulo, São Paulo, SP, Brasil. Disponível: <https://www.teses.usp.br/teses/disponiveis/48/48134/tde-28062013-162014/pt-br.php>

TRAVITZKI, R., CALERO, J., BOTO, C., 2014, "What does the National High School Exam (ENEM) tell Brazilian society?", *CEPAL Review*, v.113, n.9, pp.157–174. Disponível: <https://doi.org/10.18356/5db107a1-en>

TRAVITZKI, R., FERRÃO, M. E., COUTO, A., 2016, "Desigualdades educacionais e socioeconômicas na população brasileira pré-universitária: Uma visão a partir da análise de dados do ENEM". *Arquivos Analíticos de Políticas Educativas*, v. 24, n. 74., pp. 1-32. Disponível: <http://doi.org/10.14507/epaa.24.2199>

TRAVITZKI, R., 2017, "Qualidade com Equidade Escolar: Obstáculos e Desafios na Educação Brasileira", *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, v. 15, n. 4, pp. 27-49. Disponível: <https://doi.org/10.15366/reice2017.15.4.002>

TERRA, R., ZOGHBI, A. C., FELÍCIO, F., 2012. "Produtividade Relativa dos Setores Público e Privado em Educação: impactos sobre a escolha da escola pela família". *Economia Aplicada*, v. 16, n. 4, pp. 579-611. Disponível: <https://doi.org/10.1590/S1413-80502012000400003>

UNESCO - United Nations Educational, Scientific and Cultural Organization. *COVID 19 Educational Disruption and Response*, 2020a. Disponível em: <<https://en.unesco.org/covid19/educationresponse>>. Acesso em 04 junho 2021, 11:30.

UNESCO - United Nations Educational, Scientific and Cultural Organization. *New report on global broadband access underscores urgent need to reach the half of the world still unconnected*, 2020b. Disponível em: <<https://en.unesco.org/news/new-report-global-broadband-access-underscores-urgent-need-reach-half-world-still-unconnected>>. Acesso em 04 junho 2021, 11:30.

VITELLI, R., FRITSCH, R., CORSETTI, B, 2018, "Indicadores educacionais na avaliação da educação básica e possíveis impactos em escolas de Ensino Médio no município de Porto Alegre, Rio Grande do Sul", *Revista Brasileira de Educação. [online]*, v.23. Disponível: <https://doi.org/10.1590/S1413-24782018230065>

VINER, R. M., RUSSEL, S. J., CROKER, H., 2020, "School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review", *Lancet Child Adolesc Health*, v.4, n.5, pp. 397-404, 2020. Disponível: [http://doi.org/10.1016/S2352-4642\(20\)30095-X](http://doi.org/10.1016/S2352-4642(20)30095-X)

WAINER, J., MELGUIZO, T., 2018, "Políticas de inclusão no ensino superior: avaliação do desempenho dos alunos baseado no Enade de 2012 a 2014", *Educação e Pesquisa*, n.44. Disponível: <http://doi.org/10.1590/s1517-9702201612162807>

WERFHORS, H., MIJS, J., 2010, "Achievement Inequality and the Institutional Structure of Educational Systems: A comparative Perspective". *Annual Review of Sociology*, v. 36, pp. 407-428.

WU, F., Zhao, S., Yu, B., et al., 2020, "A new coronavirus associated with human respiratory disease in China". *Nature*, n. 579, pp. 265–269. Disponível: <http://doi.org/10.1038/s41586-020-2008-3>

APÊNDICE A – BASE DE DADOS E DADOS FALTANTES

A.1 BASE DE DADOS

A base de dados do trabalho contempla os microdados do ENEM (INEP, 2021a). Foram usadas, também, as variáveis do censo escolar (INEP, 2021c), assim como os indicadores educacionais (INEP, 2021d), todos disponíveis no site do INEP. Utilizam-se municípios com mais de 50 e menos de 500 mil habitantes e essas estimativas podem ser obtidas a partir do site do IBGE (IBGE, 2021). Uma análise de dados faltantes é apresentada a seguir. Consideram-se no trabalho, os alunos que tinham informações para todas as variáveis analisadas.

A.2 ANÁLISE DE DADOS FALTANTES

A análise de dados faltantes foi feita a partir do banco de dados obtido com a restrição de domínio, ou seja, para municípios da Região Sudeste entre 50 e 500 mil habitantes. Sete variáveis do INEP possuíam dados faltantes, além das notas do ENEM. As observações com dados faltantes, representadas na tabela por NA, do inglês *not available*, foram retiradas do banco de dados. Para as variáveis do INEP, há menos de 2% de dados faltantes, como pode-se observar na Tabela A1.

Tabela A1: Dados faltantes

Variáveis com NA Indicadores Educacionais	% faltantes
Taxa distorção idade - série	1,36%
Média hora-aula	1,36%
Adequação docente	1,36%
Taxa de aprovação	1,36%
Número médio de alunos por turma	1,36%
Complexidade de Gestão	1,36%
Taxa de abandono	1,36%

Fonte: Própria autoria, 2021.

APÊNDICE B – CÓDIGOS “R”

A seguir, encontram-se os códigos escritos em R para os modelos referentes aos resultados das Tabelas 4.2, 4.3 e 4.10. Em relação às tabelas 4.2 e 4.3, foram colocados apenas os códigos referentes às escolas privadas, sendo análogos aos referentes às escolas públicas

Modelo referente a Tabela 4.2 para escolas privadas

```
# Início do script  
# Análises para as notas de matemática  
require(lme4)  
require(sjPlot)  
  
# bancomodelobr é planilha consolidada com todas as informações (ENEM, Censo  
#Escolar e Indicadores Educacionais)  
  
#selecionando a Região sudeste  
bancomodelosudeste=bancomodelobr[c(bancomodelobr$SG_UF_RESIDENCIA=="RJ"  
|bancomodelobr$SG_UF_RESIDENCIA=="SP"|  
bancomodelobr$SG_UF_RESIDENCIA=="MG"|bancomodelobr$SG_UF_RESIDENCIA  
=="ES"),]  
  
# Sudeste escolas Privadas  
bancomodelosudestepriv=bancomodelosudeste[bancomodelosudeste$TP_DEPENDE  
NCIA==1,]  
  
##### Modelo multinível  
  
attach(bancomodelosudestepriv)  
modelo1sudesteprivada=lmer(NU_NOTA_MT~Nunca.estudou + Parou.antes.do.5º.ano  
+ Ensino.Fundamental + Ensino.Médio + Ensino.Superior +  
Completou.a.pós.graduação +Não.sei + Q005 + rendafamiliar + NU_IDADE +
```

```

TP_SEXO + TP_COR_RACA +ICG + TDI_M03 + HAD_M03 + MIRD +
INSE_CLASSIFICACAO + AFD_MED1 + tap_M03 + tab_M03 + ATU_M03 +
IED + (1|CO_ENTIDADE))

```

```

# icc
icc(modelo1sudesteprivada)

```

```

#visualizando em formato de tabela
tab_model(modelo1sudesteprivada,digits = 3)

```

Modelo referente a Tabela 4.4 para escolas privadas

```

# Início do script
# Análises para as notas de matemática
require(quantreg)
require(sjPlot)

```

```

#selecionando a Região sudeste
bancomodelosudeste=bancomodelobr[c(bancomodelobr$SG_UF_RESIDENCIA=="RJ"
|bancomodelobr$SG_UF_RESIDENCIA=="SP"|
bancomodelobr$SG_UF_RESIDENCIA=="MG"|bancomodelobr$SG_UF_RESIDENCIA
=="ES"),]

```

```

# Sudeste escolas Privadas
bancomodelosudestepriv=bancomodelosudeste[bancomodelosudeste$TP_DEPENDE
NCIA==1,]

```

```

##### Modelo de regressão quantílica

```

```

##### Análise para alguns taus

```

```

taus=c(0.1,0.25,0.50,0.75,0.90)

modelosudesteprivrqtabela=rq(NU_NOTA_MT ~ TP_LOCALIZACAO +
IN_SALA_DIRETORIA +IN_SALA_PROFESSOR
+IN_LABORATORIO_INFORMATICA + IN_QUADRA_ESPORTES + IN_COZINHA
+ IN_BIBLIOTECA +IN_SALA_LEITURA + IN_REFEITORIO + IN_COMPUTADOR +
IN_INTERNET +IN_BANDA_LARGA + IN_ALIMENTACAO + Nunca.estudou +
Parou.antes.do.5º.ano +Ensino.Fundamental +

```

Ensino.Médio + Ensino.Superior + Completou.a.pós.graduação +Não.sei + Q005 + rendafamiliar + NU_IDADE + TP_SEXO + TP_COR_RACA +ICG + TDI_M03 + HAD_M03 + MIRD + INSE_CLASSIFICACAO + AFD_MED1 + tap_M03 + tab_M03 + ATU_M03 + IED, data=bancomodelosudestepriv, tau=taus)

- (1) A ideia para as escolas públicas seria a mesma. Só que seleciona-se a dependência sendo “TP_DEPENDENCIA==0”.

Modelo referente a Tabela 4.10

Início do script

Análises para as notas de matemática

require(lme4)

require(sjPlot)

Brasil escolas Públicas

bancomodelobrpub=bancomodelobr[bancomodelobr\$TP_DEPENDENCIA==0,]

Modelo Multinível de três níveis sendo a variável sexo como coeficiente aleatório ao nível do município

Equivalente ao modelo 5

modelo5 <- lmer(NU_NOTA_MT ~ Preta + Parda + Amarela + Indígena + Não.declarado + TP_SEXO + INSE_CLASSIFICACAO + + ensinomediooumais + rendafamiliardiv+ (1|CO_ENTIDADE) + (1 + TP_SEXO|CO_MUNICIPIO))

summ(modelo5)

APÊNDICE C – TABELAS

As tabelas C1, C2 e C3 a seguir contemplam, além das estimativas dos coeficientes, os intervalos de confiança e os p-valores exatos.

Tabela C1 (Tabela 4.2 do texto)

Variáveis	ENEM - Matemática					
	Escolas Privadas			Escolas Públicas		
	Estimativas	IC	P-valor	Estimativas	IC	P-valor
Constante	769,92	(684,675; 855,170)	<0,001	502,16	(476,874; 527,450)	<0,001
Efeito Família						
Nunca estudou	2,88	(-22,376; 28,130)	0,82	-7,61	(-11,381; -3,847)	<0,001
Parou antes do 5ºano	-7,41	(-16,879; 2,061)	0,13	-6,41	(-7,989; -4,831)	<0,001
Ensino Fundamental	-4,93	(-10,885; 1,032)	0,11	0,46	(-0,894; 1,803)	0,51
Ensino Médio	7,87	(2,593; 12,982)	0,003	6,42	(5,222; 7,611)	<0,001
Ensino Superior	15,95	(10,632; 21,258)	<0,001	13,19	(11,424; 14,956)	<0,001
Completo a pós graduação	17,09	(11,646; 22,533)	<0,001	12,90	(10,683; 15,122)	<0,001
Não sei	-5,69	(-15,139; 3,767)	0,24	-9,92	(-12,180; -7,664)	<0,001
Número de pessoas domicílio	0,072	(-0,779; 0,923)	0,87	-1,99	(-2,281; -1,695)	<0,001
Renda Familiar	0,002	(0,001; 0,002)	<0,001	0,005	(0,005; 0,005)	<0,001
Idade	-16,94	(-18,27; -15,6)	<0,001	-2,48	(-2,624; -2,334)	<0,001
Sexo	-41,92	(-43,740; -40,117)	<0,001	-29,98	(-30,783; -29,187)	<0,001
Raça	6,03	(3,915; 8,145)	<0,001	9,04	(8,215; 9,875)	<0,001
Efeito Escola - Inep						
Complexidade de Gestão	-1,32	(-4,026; 1,390)	0,34	3,63	(2,326; 4,923)	<0,001
Distorção Idade-Série	-1,37	(-1,858; -0,884)	<0,001	-0,27	(-0,388; -0,157)	<0,001
Média hora aula	6,34	(3,274; 9,401)	<0,001	4,09	(2,919; 5,259)	<0,001
Regularidade Docente	14,61	(9,729; 19,492)	<0,001	9,96	(7,859; 12,069)	<0,001
Nível socioeconómico Médio	22,71	(19,148; 26,270)	<0,001	12,28	(10,409; 14,154)	<0,001
Adequação Docente	0,42	(0,250; 0,597)	<0,001	0,12	(0,033; 0,199)	0,006
Taxa de Aprovação	-0,84	(-1,633; -0,054)	0,036	-0,25	(-0,460; -0,047)	0,016
Taxa de Abandono	-0,92	(-3,371; -1,29)	0,461	-0,50	(-0,883; -0,121)	<0,001
Média de Alunos por turma	0,44	(0,225; 0,659)	<0,001	-0,08	(-0,264; 0,105)	0,40
Esforço Docente	-0,31	(-0,434; -0,192)	<0,001	-0,39	(-0,460; -0,320)	<0,001
ICC		9%			13%	

Fonte: Própria autoria, 2021.

Tabela C2 (Tabela 4.10 do texto)

ENEM - Matemática		Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Variáveis						
Efeito Fixo						
Constante		468,02 IC (463,90; 472,14)	469,54 IC (463,90; 472,14)	471,43 IC (467,35; 475,50)	463,31 IC (459,45; 467,17)	465,11 IC (461,27; 468,95)
	p-valor	0,001	0,001	0,001	0,001	0,001
Preta		-17,96 IC (-18,87; -17,06)	-17,92 IC (-18,94; -16,89)	-17,9 IC (-18,80; -17,00)	-14,90 IC (-15,91; -13,89)	-14,86 IC (-15,76; -13,97)
	p-valor	0,001	0,001	0,001	0,001	0,001
Parda		-9,40 IC (-10,04; -8,77)	-9,32 IC (-9,96; -8,68)	-9,36 IC (-9,99; -8,724)	-7,13 IC (-7,76; -6,49)	-7,15 IC (-7,79; -6,52)
	p-valor	0,001	0,001	0,001	0,001	0,001
Amarela		-5,12 IC (-6,94; -3,29)	-5,02 IC (-6,85; -3,20)	-5,06 IC (-6,88; -3,24)	-2,74 IC (-4,55; -0,93)	-2,77 IC (-4,58; -0,96)
	p-valor	0,001	0,001	0,001	0,003	0,003
Indígena		-20,89 IC (-24,30; -17,49)	-20,81 IC (-24,21; -17,41)	-20,85 IC (-24,26; -17,45)	-16,99 IC (-20,36; -13,62)	-17,02 IC (-20,39; -13,65)
	p-valor	0,001	0,001	0,001	0,001	0,001
Não declarado		-7,80 IC (-9,81; -5,79)	-7,74 IC (-9,75; -5,73)	-7,79 IC (-9,80; -5,78)	-5,39 IC (-7,38; -3,40)	-5,43 IC (-7,42; -3,44)
	p-valor	0,001	0,001	0,001	0,001	0,001
Feminino		-31,21 IC (-31,74; -30,68)	-31,2 IC (-31,73; -30,67)	-31,25 IC (-31,99; -30,51)	-28,84 IC (-29,37; -28,32)	-28,9 IC (-29,62; -28,18)
	p-valor	0,001	0,001	0,001	0,001	0,001
Nse		17,07 IC (15,90; 18,25)	16,56 IC (15,39; 17,73)	15,93 IC (14,78; 17,07)	12,94 IC (11,82; 14,02)	12,31 IC (11,23; 13,39)
	p-valor	0,001	0,001	0,001	0,001	0,001
Ensino Médio ou mais					13,34 (12,79; 13,89)	13,35 (12,80; 13,90)
	IC				0,001	0,001
Renda Familiar					5,82 (5,65; 5,99)	5,81 (5,65; 5,99)
	IC				0,001	0,001
	p-valor					
Componentes da variância						
Efeito Aleatório						
Variância do Nível 1 - aluno		6516,6 (80,72)	6514,38 (80,71)	6509,41 (80,68)	6392,01 (79,95)	6387,39 (79,92)
Variância do Nível 2 - escola		843,9 (29,04)	844,38 (29,06)	846,25 (29,09)	688,17 (26,23)	689,5 (26,25)
Variância do Nível 3 - município		137,6 (11,73)	141,36 (11,89)	184,15 (13,57)	142,41 (11,93)	184,95 (13,59)
Preta/Município			22,68 (4,76)		19,76 (4,45)	
Feminino/Município				31,51 (5,61)		28,65 (5,35)
Ensino Médio ou mais/Município						
Renda/Município						

Nota 1: IC é o intervalo de confiança.

Nota 2: Preta/Município - fonte de variação do coeficiente de raça Preta entre os municípios.

Nota 3: O valor em parênteses apresentado na componente da variância é o desvio padrão.

Fonte: Própria autoria, 2021.

Os municípios apresentados nas Figuras 4.9 e 4.10 são listados na Tabela C3 com os efeitos associados e a região a qual pertencem.

Tabela C3: 30 melhores e 30 piores municípios

Nome município	Efeito raça	Região	Nome município	Efeito sexo	Região
Rio Branco	5,99	Norte	Jaru	6,09	Norte
Bragança	3,23	Norte	Rio Branco	7,96	Norte
Castanhal	3,19	Norte	Coari	6,21	Norte
Itupiranga	3,07	Norte	Humaitá	6,49	Norte
Santana	7,03	Norte	Itacoatiara	10,44	Norte
Palmas	3,81	Norte	Manacapuru	8,28	Norte
Caxias	3,21	Nordeste	Abaetetuba	7,64	Norte
Imperatriz	4,17	Nordeste	Capanema	8,91	Norte
Itapajé**	-4,20	Nordeste	Itupiranga	6,91	Norte
Pacajus	-4,13	Nordeste	Macapá	10,08	Norte
Cabedelo	3,21	Nordeste	Santana	7,74	Norte
Cajazeiras	-4,79	Nordeste	Paço do Lumiar	6,34	Nordeste
Camaragibe	4,14	Nordeste	Juazeiro do Norte	-6,30	Nordeste
Igarassu	3,69	Nordeste	Maranguape	-7,12	Nordeste
Olinda	4,26	Nordeste	Caicó	-6,86	Nordeste
Paudalho	2,98	Nordeste	Parnamirim	-9,05	Nordeste
Timbaúba	-4,34	Nordeste	Santa Rita	6,49	Nordeste
São Cristóvão	4,12	Nordeste	Timbaúba	-8,16	Nordeste
Candeias	3,51	Nordeste	Arapiraca	-6,66	Nordeste
Casa Nova	3,53	Nordeste	Lagarto	-13,06	Nordeste
Barbacena	-6,47	Sudeste	Barreiras	-9,20	Nordeste
Campo Belo	-5,25	Sudeste	Ilhéus	6,08	Nordeste
Caratinga	-4,13	Sudeste	Itabuna	5,77	Nordeste
Cataguases	-3,54	Sudeste	Teixeira de Freitas	-6,05	Nordeste
Congonhas	-5,05	Sudeste	Vitória da Conquista	-9,39	Nordeste
Divinópolis	-3,80	Sudeste	Barbacena	-10,99	Sudeste
Itaúna	-4,50	Sudeste	Divinópolis	-9,32	Sudeste
Leopoldina	-3,51	Sudeste	Itaúna	-7,69	Sudeste
Manhuaçu	-5,20	Sudeste	Leopoldina	-6,23	Sudeste
Muriaé	-5,14	Sudeste	Muriaé	-6,69	Sudeste
Patos de Minas	-3,75	Sudeste	São Sebastião do Paraíso	-8,55	Sudeste
Timóteo	-4,82	Sudeste	Viçosa	-8,22	Sudeste
Viçosa	-7,07	Sudeste	Vitória	-6,40	Sudeste
Cachoeiro de Itapemirim	-4,11	Sudeste	Nova Friburgo	-7,10	Sudeste
Colatina	-4,74	Sudeste	Queimados	5,68	Sudeste
Linhares	-5,03	Sudeste	Saquarema	-6,67	Sudeste
São Mateus	-4,62	Sudeste	Americana	-8,06	Sudeste
Vitória	-9,32	Sudeste	Carapicuíba	7,11	Sudeste
Niterói	-5,78	Sudeste	Cubatão	-6,40	Sudeste
Nova Friburgo	-5,78	Sudeste	Embu das Artes	6,67	Sudeste
Arujá	-3,88	Sudeste	Franca	5,59	Sudeste
Barueri	-3,94	Sudeste	Itaquaquecetuba	7,15	Sudeste
Franco da Rocha	3,43	Sudeste	Jaguariúna	5,66	Sudeste
Itaquaquecetuba	4,77	Sudeste	Jandira	5,70	Sudeste
Jundiaí	3,92	Sudeste	Limeira	-8,09	Sudeste
Mauá	3,63	Sudeste	Matão	-6,97	Sudeste
Santana de Parnaíba	4,83	Sudeste	Piracicaba	-6,63	Sudeste
Araucária	3,14	Sul	Santa Bárbara d'Oeste	-5,97	Sudeste
Cianorte	3,08	Sul	Ibiporã	5,63	Sul
Sarandi	4,50	Sul	Tubarão	5,63	Sul
Criciúma	3,07	Sul	Bento Gonçalves	-6,36	Sul
Florianópolis	-4,72	Sul	Cachoeirinha	-7,99	Sul
Alvorada	-3,64	Sul	Novo Hamburgo	-8,84	Sul
Pelotas	-4,07	Sul	Santo Ângelo	-10,17	Sul
Santa Maria	-5,61	Sul	São Borja	5,83	Sul
Corumbá	3,00	Centro Oeste	Uruguaiana	6,81	Sul
Cáceres	4,14	Centro Oeste	Corumbá	9,84	Centro Oeste
Várzea Grande	4,06	Centro Oeste	Várzea Grande	6,42	Centro Oeste
Goianésia	2,98	Centro Oeste	Senador Canedo	8,27	Centro Oeste
Trindade	3,26	Centro Oeste	Trindade	9,94	Centro Oeste

Fonte: Própria autoria, 2021.

APÊNDICE D – ANÁLISE RESIDUAL

Uma das hipóteses em que à estimação dos parâmetros dos modelos utilizados se apoia é a distribuição normal dos resíduos.

O histograma dos resíduos apresentado na Figura D.1 mostra o seu comportamento geral. Pode observar que a normal superposta é, visualmente, uma aproximação razoável para a distribuição empírica dos resíduos.

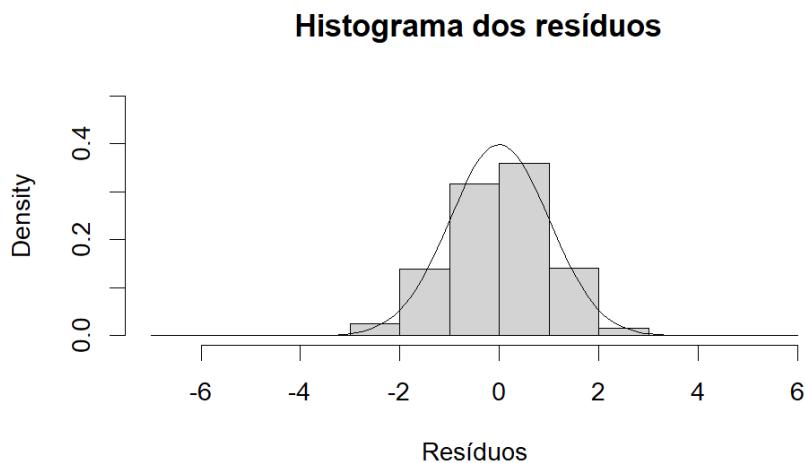


Figura D.1: Histograma para o resíduo da equação 3.2

A estatística de teste usual para se testar esta hipótese é a correlação entre os resíduos e os percentis de uma distribuição normal padrão. Ou seja, o teste é baseado nos gráficos denominados de QQ-Plot, Figura D.2.

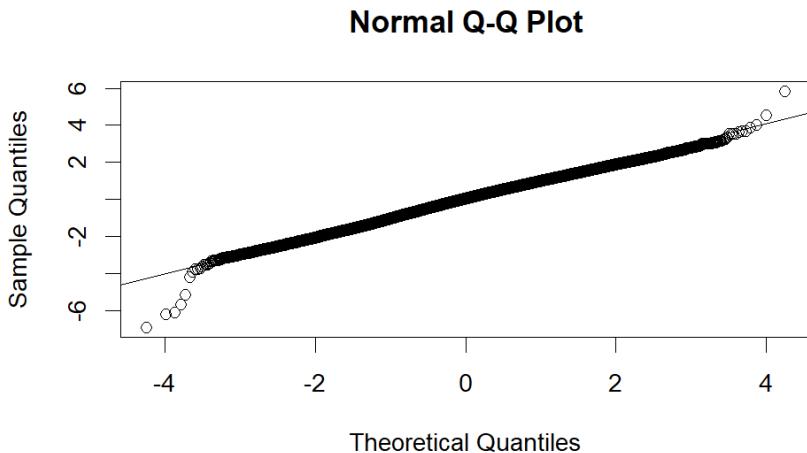


Figura D.2: Gráfico QQ-plot

A correlação entre as duas medidas é de 0,9985. Apesar deste alto valor, pela quantidade de pontos disponíveis, a hipótese é rejeitada. Ou seja, pelo gráfico pode-se verificar que a cauda da distribuição dos resíduos é mais pesada do que a da distribuição normal. Também é possível perceber, nas figuras acima, que o modelo se ajusta razoavelmente para grande parte dos dados e que os resíduos possuem distribuição simétrica.

Aqui é importante observar que, na análise apresentada no capítulo 3, o número de observações é muito grande. Isso tem como consequência que, mesmo pequenas diferenças, são detectadas.

Um exercício de simulação foi feito em que amostras de tamanhos variados foram retiradas e a normalidade dos resíduos, testada. A partir do momento em que se selecionam mais de 1000 observações, já se torna difícil de encontrar a normalidade para esses conjuntos. O que nos leva que a rejeição do teste está associada mais à precisão implícita das grandes amostras do que a um comportamento anômalo. Cenário similar acontece para todas as análises dos resíduos dos modelos propostos ao longo do trabalho. Sendo que, no modelo da equação 3.5, os resíduos são considerados normais ao nível do município, o que é razoável esperar já que há um grupo de aproximadamente 600 municípios, bem menor comparado aos demais.

Ou seja, como dito por BOX (1976) “todos modelos são errados, alguns são úteis”. No nosso caso, é a lupa do grande número de observações que indica um problema. Nada sugere que as estimativas são viciadas. Isso pode levantar discussões importantes sobre identificação de modelos, parcimônia e análise de pressupostos. As bases de dados aqui utilizadas contemplam um grande conjunto de observações, o que dificulta as verificações de normalidade. Existe uma discussão mais recente na literatura

que ilustra a robustez da estimativa de máxima verossimilhança em muitos aspectos para os efeitos aleatórios, mesmo quando pressupostos de normalidade não são atendidos (McCULLOCH e NEUHAUS, 2011). Como existe uma discordância sobre a importância dos pressupostos, McCULLOCH e NEUHAUS (2011) citam em seu trabalho que “uma conclusão importante é o grande grau de robustez da máxima verossimilhança para uma grande variedade de situações comumente encontradas” (tradução própria).

Além disso, nesta tese os modelos multiníveis foram estimados a partir de um grande conjunto de observações, o que de acordo com HOX (2010) favorece a robustez das estimativas, além de uma numerosa quantidade de grupos, tanto em relação a escolas quanto de municípios. Uma citação ao longo do livro é feita por esse autor: “Uma vantagem do método de estimativa de máxima verossimilhança é que é geralmente robusto e produz estimativas que são assintoticamente eficientes e consistentes. Com grandes amostras, as estimativas de máxima verossimilhança são geralmente robustas contra violações moderadas dos pressupostos, tais como ter erros não normais” (HOX, 2010, pp.40, tradução própria).

Segundo MAAS e HOX (2004), “erros residuais com distribuições não normais no segundo nível de um modelo de regressão multinível, aparentam ter pouco ou nenhum efeito nas estimativas dos efeitos fixos” (tradução própria). Ainda segundo os autores, os erros não normais possuem efeito na parte aleatória do segundo nível do modelo, com estimativas da variância não viesadas, porém erros padrão nem sempre precisos. Já no nível mais baixo do modelo, os erros padrão da estimativa de máxima verossimilhança são precisos.

Por fim como comparação encontra-se, na Figura D.2, o gráfico do QQ-Plot usando a distribuição T-Student que apresenta caudas mais pesadas para alguns graus de liberdade diferentes. A correlação para as diferentes distribuições de T-Student apresentadas abaixo, variam de 0,9977 até 0,9984.

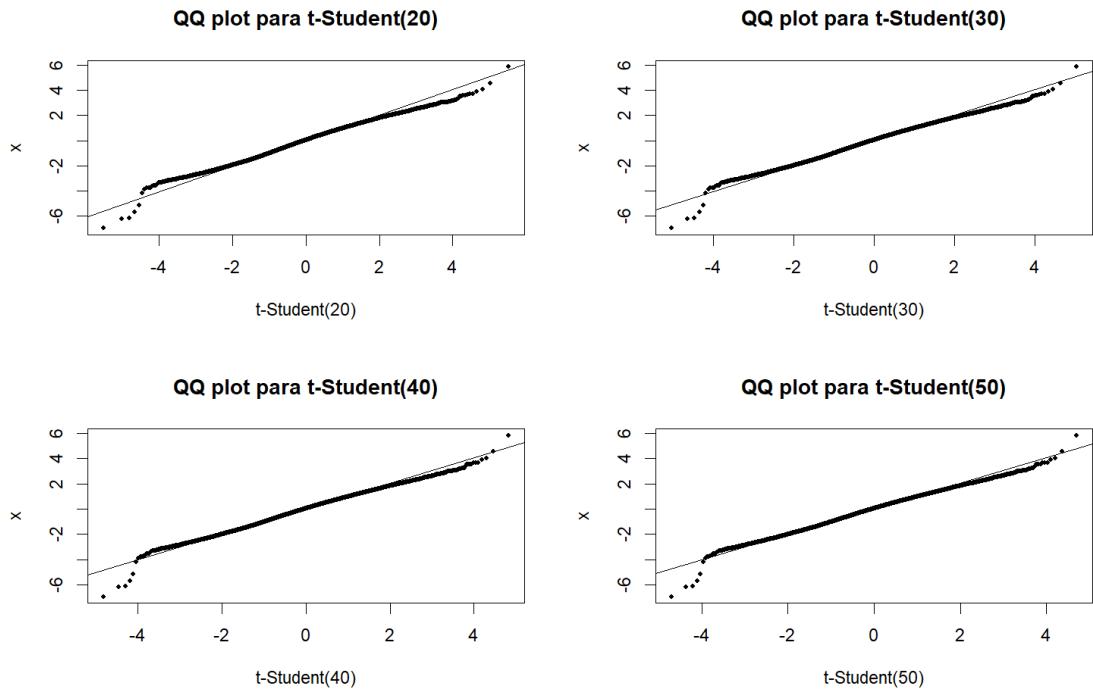


Figura D.2: QQ-Plot para distribuição T-Student variando o grau de liberdade.

Dessa forma, os resíduos não indicam dificuldades substanciais para o modelo. Portanto, os resultados apresentados na tese suportam a realidade e estão condizentes com evidências já encontradas na literatura.