



Universidade de Brasília
Departamento de Estatística

**Fatores associados ao desempenho em Matemática dos alunos do 9º ano do
Ensino Fundamental no Centro-Oeste: Uma abordagem multinível**

Tailine Juliana dos Santos Nonato

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2025**

Tailine Juliana dos Santos Nonato

**Fatores associados ao desempenho em Matemática dos alunos do 9º ano do
Ensino Fundamental no Centro-Oeste: Uma abordagem multinível**

Orientador(a): Prof. Dr. Luís Gustavo do Amaral Vinha

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2025**

Aos meus primeiros professores: meus pais Eunice e Ivo.

Agradecimentos

Assim como dedico a eles esta monografia, também os agradeço por permitirem e garantirem que ela acontecesse. Agradeço os meus pais, Eunice e Ivo, por terem me oferecido apoio incondicional e terem sido grandes inspirações para mim. A missão deles sempre foi, por meio da educação, transformar a vida das pessoas e, seguindo seus passos, também tornarei esta a minha missão.

Agradeço as minhas avós, Raimunda e Iolanda, por terem me oferecido abraços calorosos e olhares preocupados do início da minha graduação até o fim de suas vidas. Agradeço as minhas tias Judith e Ideides por oferecerem refúgios de confiança e apoio.

Agradeço a Mariana por ter me oferecido um amor e uma parceria genuína que me conduziu à luz e me motivou a seguir. Agradeço a Juliana por ter sido uma grande amiga e braço direito nessa jornada. E, aos Reis Perdidos, agradeço pelo espaço de pertencimento e torço para podermos nutri-lo por muito tempo.

Agradeço a Universidade de Brasília por oferecer uma graduação repleta de oportunidades de crescimento acadêmico, profissional e pessoal. Cito entre essas oportunidades a ESTAT, as *M²ICEs*, o DataUnb/Cecampe Centro-Oeste e o programa de mobilidade internacional. Agradeço também ao SEDAP/INEP que me recebeu de forma muito aberta e garantiu a segurança da informação, ao passo que também incentivou o uso dos dados em prol da sociedade.

Agradeço a Universidade de Macau que me recebeu como intercambista e bolsista e proporcionou experiências inesquecíveis ao meu período final de graduação.

Por fim, agradeço aos docentes e a equipe administrativa do Departamento de Estatística que sempre demonstraram muita paciência e perseveraram no objetivo de criar grandes Estatísticos. Em especial, agradeço ao professor Doutor Luis Gustavo do Amaral Vinha, que ofereceu seu tempo, competência e materiais para que fosse possível chegar nos resultados presentes neste estudo.

Resumo

O presente estudo tem como objetivo identificar características de alunos e escolas associadas ao desempenho em Matemática dos estudantes do 9º ano do ensino fundamental na região Centro-Oeste. A análise considerou 99.422 alunos distribuídos em 1.955 escolas. Para a identificação dos fatores associados, foi ajustado um modelo de regressão linear multinível com dois níveis hierárquicos: alunos no nível 1 e escolas no nível 2. Os dados utilizados foram provenientes da edição de 2019 do Sistema Nacional de Avaliação da Educação Básica (SAEB), complementados por variáveis da edição de 2015 do SAEB e do Programa Dinheiro Direto nas Escolas (PDDE) no período de 2015 a 2018. No nível do aluno, foram identificados efeitos significativos do grupo racial, da escolaridade da mãe, do histórico de reprovações e do nível socioeconômico. No nível da escola, observaram-se efeitos significativos da área da escola (capital ou interior), do nível socioeconômico médio e do desempenho médio da escola no SAEB de 2015. Por outro lado, o repasse total do PDDE não apresentou efeitos significativos no desempenho dos alunos.

Palavras-chaves: Modelagem multinível, Avaliação educacional, Desempenho escolar, SAEB, PDDE.

Abstract

The following study aims to identify student and school characteristics associated with mathematics performance among 9th-grade students in the Central-West region of Brazil. The analysis considered 99,422 students distributed across 1,955 schools. To identify the associated factors, a multilevel linear regression model was fitted with two hierarchical levels: students at level 1 and schools at level 2. The data used were obtained from the 2019 edition of SAEB, supplemented by variables from the 2015 edition of SAEB and the PDDE dataset covering the period from 2015 to 2018. At the student level, significant effects were identified for racial group, mother's education, grade repetition history, and socioeconomic status. At the school level, significant effects were observed for school location (capital or countryside), average socioeconomic status, and the school's average performance in the 2015 SAEB. Conversely, the total funds received from PDDE did not show significant effects on student performance.

Keywords: Multilevel modeling, Education assessment, Academic performance, SAEB, PDDE.

Lista de Tabelas

1	Descrições das variáveis utilizadas no estudo	19
2	Medidas resumo da proficiência em Matemática	21
3	Frequências da escolaridade da mãe (%)	24
4	Frequências de reprovações (%)	24
5	Frequências das características das escolas (%)	25
6	Medidas resumo do valor total de repasses do PDDE (R\$)	27
7	Estimativas do modelo multinível nulo	28
8	Estimativas do modelo multinível com variáveis do Nível 1	29
9	Estimativas do modelo multinível com variáveis dos níveis 1 e 2	30
10	Estimativas do modelo multinível com variáveis dos níveis 1 e 2 com efeitos aleatórios	31
11	Estimativas do modelo multinível com variáveis dos níveis 1 e 2 com efeitos aleatórios e interações	32
12	Estimativas do modelo multinível com adição de variáveis do PDDE e do SAEB 2015	33

Lista de Figuras

1	Representação gráfica de uma estrutura hierárquica com dois níveis	10
2	Representação gráfica da linha do tempo do estudo	17
3	Distribuição da proficiência em Matemática	22
4	Gráfico de quantis da proficiência em Matemática	22
5	Proficiência em Matemática por raça/cor	23
6	Proficiência em Matemática por Nível Socioeconômico (NSE) do aluno . .	23
7	Proficiência em Matemática por escolaridade da mãe	24
8	Proficiência em Matemática por reprovação	25
9	Proficiência em Matemática por NSE da escola	26
10	Proficiência em Matemática por área da escola	26
11	Proficiência em Matemática por localização da escola	27
12	Proficiência em Matemática por valor total do PDDE	28
13	Distribuição dos resíduos no Nível 1	34
14	Distribuição dos resíduos no Nível 2: Intercepto	34
15	Distribuição dos resíduos no Nível 2: NSE do aluno	35
16	Distribuição dos resíduos no Nível 2: Raça/cor	35
17	Distribuição dos resíduos no Nível 2: Reprovação	35

Lista de Abreviações e Siglas

EMV Estimação de Máxima Verossimilhança.

FNDE Fundo Nacional de Desenvolvimento da Educação.

ICC Coeficiente de Correlação Intraclasse.

INEP Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

LDBEN Lei de Diretrizes e Bases da Educação Nacional.

MMV Método de Máxima Verossimilhança.

NSE Nível Socioeconômico.

PDDE Programa Dinheiro Direto nas Escolas.

REML Estimação de Máxima Verossimilhança Restrita.

SAEB Sistema Nacional de Avaliação da Educação Básica.

SEDAP Serviço de Acesso à Dados Protegidos.

Sumário

1 Introdução	8
2 Referencial Teórico	10
2.1 Análise multinível	10
2.2 Definição do modelo	10
2.2.1 Suposições do modelo	12
2.2.2 Estimação dos parâmetros	13
2.2.3 Passo a passo para formulação do modelo	13
2.2.4 Variância explicada pelo modelo	15
2.2.5 Análise dos resíduos	15
3 Metodologia	17
3.1 Conjunto de dados	17
3.1.1 SAEB	17
3.1.2 PDDE	18
3.1.3 Período e variáveis	19
3.2 Análise dos dados	20
4 Resultados	21
4.1 Análise Descritiva	21
4.1.1 Características dos alunos - Nível 1	21
4.1.2 Características das escolas - Nível 2	25
4.2 Análise Multinível	28
4.2.1 Etapa 1: SAEB 2019	28
4.2.2 Etapa 2: SAEB 2019, SAEB 2015 e PDDE 2015-2018	32
4.2.3 Análise dos resíduos	34
5 Conclusão	36
Referências	38

1 Introdução

A avaliação de políticas públicas é uma ferramenta fundamental para a tomada de decisões e para a melhoria dos serviços prestados à população. No contexto educacional, as políticas públicas visam garantir o cumprimento do direito de acesso à educação, conforme estabelecido no artigo 6º da Constituição Federal do Brasil (BRASIL, 1988) e regulamentado pela Lei de Diretrizes e Bases da Educação Nacional (LDBEN) (BRASIL, 1996).

Um dos alvos das políticas públicas educacionais é a qualidade do ensino e tem-se o desempenho escolar como um dos principais indicadores dessa qualidade (NETO; ROSENBERG, 1995). No Brasil, o desempenho escolar na Educação Básica é aferido por meio do SAEB, estudo conduzido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Por meio desse estudo, provas de Língua Portuguesa e Matemática são aplicadas a cada dois anos para estudantes do 5º e 9º ano do ensino fundamental e do 3º ano do ensino médio em escolas públicas e privadas. Além do desempenho escolar, os resultados do SAEB disponibilizam o acesso às características individuais dos alunos e características das escolas.

No contexto de avaliação e financiamento educacional, o PDDE também se destaca como um importante instrumento de apoio às escolas públicas brasileiras. Administrado pelo Fundo Nacional de Desenvolvimento da Educação (FNDE), o PDDE realiza repasses anuais de recursos financeiros às escolas, visando fortalecer sua autonomia na gestão de recursos e contribuir para a melhoria das infraestruturas educacionais. O valor do repasse é calculado com base no número de alunos, localização da escola e sua rede de ensino, abrangendo escolas públicas e escolas privadas de educação especial.

Analizar os determinantes da qualidade do ensino envolve não apenas a avaliação do desempenho escolar, é fundamental compreender os fatores associados a esse desempenho, sejam eles com efeitos positivos ou negativos. Nesse contexto, a modelagem multinível tem sido amplamente empregada na literatura para identificar os fatores relacionados ao desempenho dos alunos da educação básica com base nos dados do SAEB. Por exemplo, no estudo de Andrade e Laros (2007), foi constatado que, para cada aumento de um desvio-padrão no nível socioeconômico da escola, observa-se um incremento médio de 11,62 pontos no desempenho escolar dos alunos na escala de proficiência do SAEB, que varia de 0 a 500. De forma complementar, Laros, Marciano e Andrade (2012) verificaram que a desigualdade regional no desempenho médio das escolas em Língua Portuguesa está fortemente associada ao nível socioeconômico dos alunos e das escolas. Ademais, Vinha,

Karino e Laros (2016) ajustaram um modelo para explicar o desempenho em Matemática que incluiu variáveis de controle, 13 variáveis no nível do aluno, 9 no nível da escola, além de efeitos randômicos e interações, explicando aproximadamente 50% da variância no nível da escola e mais de 11% da variância no nível do aluno.

Contudo, a utilização de modelos multiníveis que integrem informações de diferentes edições do SAEB e que considerem dados externos, como os do PDDE, é menos comum. Diante do apresentado, este estudo, que está dividido em 5 capítulos, tem por objetivo identificar características dos alunos e das escolas associadas ao desempenho dos alunos em Matemática no 9º ano do ensino fundamental na região Centro-Oeste. Essa identificação será realizada por meio da modelagem multinível utilizando os dados da edição de 2019 do SAEB, além de variáveis da edição de 2015 do SAEB e do PDDE de 2015 a 2018. Especificamente, aplicar a modelagem multinível utilizando como variável preditora o desempenho médio das escolas em 2015 para avaliar a associação do dinheiro investido nas escolas por meio do PDDE com o desempenho escolar. Portanto, este estudo não se configura como uma avaliação de impacto de política pública, mas oferece uma análise aprofundada e detalhada dos dados e das políticas educacionais em questão.

No Capítulo 2 é apresentado o referencial teórico descrevendo a abordagem multinível e um modelo multinível, as suposições atreladas a este tipo de modelo, estimativa dos parâmetros, variância explicada pelo modelo e análise dos resíduos. No Capítulo 3, é descrita a metodologia do estudo, especificando o conjunto de dados e a análise destes dados. No Capítulo 4, são descritos os resultados, desde uma análise descritiva até a análise multinível realizada em duas etapas. Por fim, no Capítulo 5, são enunciadas as considerações finais do estudo.

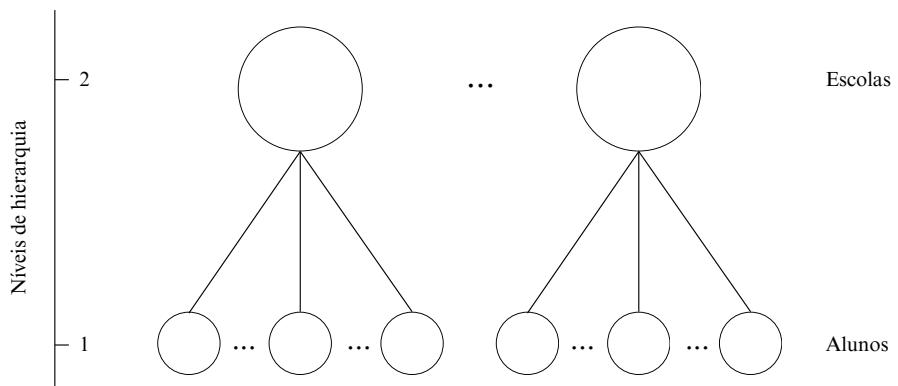
2 Referencial Teórico

2.1 Análise multinível

Frequentemente é necessário investigar as relações entre indivíduos e os grupos aos quais eles fazem parte. Quando o fenômeno estudado envolve características dos indivíduos e também características dos grupos, esse tipo de investigação pode ser denominada pesquisa multinível. Esse contexto surge a partir do estudo de populações em estrutura hierárquica, ou seja, uma amostra desse tipo de população deve ser coletada em etapas (HOX; MOERBEEK; SCHOOT, 2010).

Aplicações podem ser encontradas em diversas esferas de conhecimento e, para exemplificar a estrutura multinível, descreve-se comumente a estrutura hierárquica das redes de ensino, onde alunos estão agrupados em escolas. Inicialmente selecionam-se as escolas e, a partir delas, selecionam-se os alunos, gerando camadas de informação, chamadas de níveis. Geralmente, o nível mais baixo (1) é definido pelos indivíduos, no exemplo, os alunos, e o ponto inicial de coleta das informações é dito o nível mais alto ou grupo, no exemplo, as escolas, como ilustra a Figura 1.

Figura 1: Representação gráfica de uma estrutura hierárquica com dois níveis



2.2 Definição do modelo

A regressão linear é um dos métodos estatísticos mais utilizados para modelar a relação entre uma variável resposta e uma ou mais variáveis explicativas. No caso mais simples, com somente uma variável preditora, o modelo pode ser descrito por

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

onde Y_i representa a variável resposta do indivíduo i , X_i a variável preditora, β_0 e β_1 os coeficientes a serem estimados e e_i o termo de erro aleatório (CHEIN, 2019). Esse modelo assume que as observações são independentes e que a variabilidade nos dados pode ser explicada exclusivamente pelos preditores incluídos no modelo. No entanto, em estruturas hierárquicas, as observações individuais geralmente não são independentes. Nesse tipo de estrutura é necessário recorrer às técnicas de análise multinível, como a modelagem de regressão linear multinível. Neste caso, considera-se a existência de efeitos fixos, que modelam características populacionais, e efeitos aleatórios, que modelam características individuais, tanto no nível mais baixo quanto em níveis mais altos.

No contexto escolar, um modelo multinível com 2 níveis pode ser descrito considerando J escolas, n_j indivíduos em cada escola ($j = 1, 2, \dots, J$) e uma variável explicativa para cada nível. Por exemplo, no nível mais baixo (alunos), considera-se a variável resposta Y como o desempenho dos alunos em Matemática e uma variável explicativa X como o NSE do aluno, enquanto no nível mais alto (escolas), considera-se uma variável explicativa Z que representa a localização da escola. Segundo Hox (1998), com $i = 1, 2, \dots, n_j$, o modelo pode ser definido nos dois níveis. No Nível 1, tem-se

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}, \quad (2.2.1)$$

em que

- Y_{ij} : variável aleatória referente ao desempenho em Matemática do i -ésimo aluno da j -ésima escola;
- X_{ij} : NSE do i -ésimo aluno da j -ésima escola;
- β_{0j} : variável aleatória referente ao intercepto da j -ésima escola;
- β_{1j} : variável aleatória referente ao efeito do NSE do aluno na j -ésima escola;
- e_{ij} : erro aleatório associado ao i -ésimo aluno da j -ésima escola.

E no Nível 2, tem-se

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}Z_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_j + u_{1j}, \end{aligned} \quad (2.2.2)$$

onde

- Z_j : localização da j -ésima escola;
- γ_{00} : intercepto geral;
- γ_{10} : efeito médio do NSE do aluno;
- γ_{01} : efeito associado à localização da escola no intercepto;
- γ_{11} : efeito de interação do NSE do aluno e a localização da escola;
- u_{0j} e u_{1j} : erros aleatórios associados à β_{0j} (intercepto) e β_{1j} (efeito do NSE do aluno), respectivamente.

Substituindo a equação (2.2.2) na equação (2.2.1), a equação final do modelo com uma variável explicativa no nível do aluno e uma variável explicativa no nível da escola fica expressa por

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + u_{1j}X_{ij} + u_{0j} + e_{ij}. \quad (2.2.3)$$

Em um modelo de regressão linear multinível, assume-se que os modelos de diferentes escolas podem ter diferentes interceptos e diferentes coeficientes de inclinação, enquanto em modelo de regressão linear tradicional o intercepto e os efeitos das variáveis explicativas seriam considerados iguais para todas as escolas (OSIO, 2013).

2.2.1 Suposições do modelo

Analogamente ao modelo de regressão linear tradicional, na modelagem multinível também existem pressupostos para viabilizar inferências sob o modelo e, consequentemente, garantir que o respectivo modelo ajustado possa ser utilizado para inferências populacionais. As suposições do modelo tradicional também se aplicam ao modelo multinível. Assume-se que a relação entre as variáveis explicativas e a variável resposta pode ser modelada como uma combinação linear dos preditores, ou seja, linearidade, e que os preditores não devem ser altamente correlacionados entre si, ou seja, não colinearidade. Espera-se também que os erros aleatórios do nível aluno tenham média zero, ou seja, $E(e_{ij}) = 0$, variância constante, neste estudo dada por $Var(e_{ij}) = \sigma_e^2$, e não sejam correlacionados entre si, ou seja, $Cov(e_{ij}, e_{kj}) = 0$ para $i \neq k$.

No modelo multinível, por ser considerado mais de um nível, é necessário adicionar suposições específicas a este tipo de modelagem. Segundo Hox, Moerbeek e Schoot (2010), os efeitos aleatórios do nível da escola devem ter média zero, ou seja, $E(u_j) = 0$, variância

constante, neste estudo dada por $Var(u_j) = \sigma_u^2$, e são não correlacionados dos erros do nível do aluno, ou seja, $Cov(u_j, e_{ij}) = 0$. Por fim, espera-se que os erros aleatórios, tanto do nível do aluno quanto do nível da escola, atendam a normalidade, ou seja, $e_{ij} \sim N(0, \sigma_e^2)$ e $u_j \sim N(0, \sigma_u^2)$.

2.2.2 Estimação dos parâmetros

Os parâmetros dos modelos podem ser estimados utilizando o Método de Máxima Verossimilhança (MMV). Geralmente, esse método é robusto, ou seja, tolera pequenas violações das suposições, e produz estimações assintoticamente eficientes e consistentes (HOX; MOERBEEK; SCHOOT, 2010). No contexto da regressão multinível, pode-se aplicar a Estimação de Máxima Verossimilhança (EMV) Completa, onde os coeficientes de regressão e os componentes da variância são incluídos na função de verossimilhança, e a Estimação de Máxima Verossimilhança Restrita (REML), onde somente os componentes da variância são incluídos na função e os estimadores são calculados em uma segunda etapa (HOX; MOERBEEK; SCHOOT, 2010). No presente estudo, o modelo foi ajustado utilizando o Método de Máxima Verossimilhança Restrita (REML). Nesse procedimento, os valores são obtidos a partir da integração da densidade marginal da variável resposta com respeito aos efeitos fixos (LAIRD; WARE, 1982).

2.2.3 Passo a passo para formulação do modelo

Dado que nos modelos multiníveis existem níveis distintos e cada nível deve ser considerado, o modelo pode ser construído com uma abordagem diferente da feita em modelos tradicionais. Hox, Moerbeek e Schoot (2010) sugerem um processo em 5 passos:

- Passo 1

Análise de um modelo nulo, ou seja, sem variáveis explicativas, tendo somente o intercepto como coeficiente, que representa o desempenho médio dos alunos. O modelo nulo é dado por

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}.$$

Neste passo calcula-se o Coeficiente de Correlação Intraclass (ICC), valor que fornece indicação suficiente para justificar uma abordagem multinível. Este coeficiente apresenta valores no intervalo contínuo $(0, 1)$, onde, quanto mais próximo de zero, maior

a homogeneidade entre as escolas (FERRAO, 2004), pois a estimativa de σ_u^2 será muito próxima de zero. Um valor acima de 0,1 indica que a análise multinível deve ser considerada (LEE, 2008). O ICC é dado por

$$ICC = \frac{\sigma_u^2}{\sigma_e^2 + \sigma_u^2},$$

onde σ_u^2 é variância dos erros aleatórios associados às variáveis explicativas do Nível 2 (escola) e σ_e^2 é a variância dos erros no Nível 1 (aluno).

- Passo 2

Análise de um modelo com adição de variáveis explicativas com efeitos fixos do primeiro nível (aluno). Assim, as variâncias correspondentes aos coeficientes de inclinação das variáveis explicativas são fixadas em zero, ou seja, a inclinação das variáveis explicativas são consideradas constantes entre as escolas. Considerando somente uma variável preditora, tem-se

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + u_{0j} + e_{ij}.$$

- Passo 3

Análise de um modelo com adição de variáveis explicativas do segundo nível (escola). Assim como nos passos anteriores, os interceptos variam conforme as escolas, e, neste passo, adicionam-se variáveis de efeito fixo do nível escola que auxiliam na explicação das variações dos interceptos. Considerando a adição de uma variável preditora, o modelo é dado por

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + u_{0j} + e_{ij}.$$

- Passo 4

Análise de um modelo com adição de efeitos aleatórios do primeiro nível (aluno), verificando quais variáveis explicativas têm efeitos diferentes para diferentes escolas. Neste passo, tem-se

$$Y_{ij} = \gamma_{00}X_{pij} + \gamma_{0q}Z_{qj} + u_{pj}X_{pij} + u_{0j} + e_{ij}.$$

- Passo 5

Análise de um modelo com adição de interações entre variáveis de diferentes níveis, chegando a um modelo final que pode ser dado por

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{1j}X_{1ij} + u_{0j} + e_{ij}.$$

2.2.4 Variância explicada pelo modelo

Para verificar o quanto da variabilidade da resposta é explicada por um modelo linear, calcula-se comumente o coeficiente de determinação R^2 . A estatística indica a qualidade de ajuste do modelo, no entanto, a abordagem utilizada em um modelo de regressão tradicional não se aplica exatamente da mesma forma no modelo de regressão multinível. Segundo Hox, Moerbeek e Schoot (2010), grande parte da variância não pode ser explicada e, pela presença de efeitos aleatórios, é inerente a maior complexidade do modelo. Raudenbush e Bryk (2002) propõem o cálculo de um coeficiente R^2 para cada um dos níveis do modelo. Para proporção da variância explicada no primeiro nível, calcula-se

$$R_1^2 = \left(\frac{\sigma_{e|b}^2 - \sigma_{e|m}^2}{\sigma_{e|b}^2} \right),$$

em que $\sigma_{e|b}^2$ é a variância dos resíduos do Nível 1 no modelo nulo e $\sigma_{e|m}^2$ é variância dos resíduos do Nível 1 no modelo de interesse para comparação. Para a proporção da variância explicada no segundo nível calcula-se

$$R_2^2 = \left(\frac{\sigma_{u|b}^2 - \sigma_{u|m}^2}{\sigma_{u|b}^2} \right),$$

em $\sigma_{u|b}^2$ é a variância dos resíduos do Nível 2 no modelo nulo e $\sigma_{u|m}^2$ é variância dos resíduos do Nível 2 no modelo de interesse para comparação. A interpretação dos dois coeficientes pode ser feita similarmente à feita na regressão tradicional.

2.2.5 Análise dos resíduos

Considerando que os erros aleatórios não são observáveis, cria-se uma quantidade, o resíduo, que depende somente da amostra e que tenta imitar o comportamento do erro. Segundo Nunes (2010), “os resíduos consistem, na prática, na diferença entre a resposta observada e o respectivo valor ajustado pelo modelo dentro de cada grupo”.

Os resíduos do Nível 1 são estimativas do componente e_{ij} e os resíduos do Nível 2 são estimativas dos componentes u_{0j} e u_{1j} . Assim, avalia-se as características que os resíduos apresentam sob o modelo ajustado para verificar se as suposições do modelo relacionadas a normalidade, independência e homocedasticidade (Seção 2.2.1) parecem válidas. Neste estudo, essa análise ocorre a partir de gráficos de dispersão dos valores observados em relação aos valores estimados para verificar homocedasticidade e independência e gráficos quantil-quantil para verificar normalidade.

3 Metodologia

3.1 Conjunto de dados

Os dados utilizados são provenientes de duas fontes públicas, o INEP e o FNDE. A principal base de dados é a edição de 2019 do SAEB (INEP/MEC, 2020) e a partir desta, busca-se agregar informações da edição de 2015 do SAEB e repasses do PDDE em 2015, 2016, 2017 e 2018. Para viabilizar a integração, são definidos como focos de estudo os estudantes da rede pública do 9º ano do ensino fundamental na região Centro-Oeste. A Figura 2, que apresenta uma linha do tempo do estudo, demonstra como foram considerados os dados do SAEB e do PDDE segundo ordem cronológica.

Figura 2: Representação gráfica da linha do tempo do estudo



Fonte: Ilustrações por Thirasak e Ainul Muttaqin (Noun Project)

Os dados do SAEB 2019 abrangiam 143.223 alunos da região Centro-Oeste e 2.536 escolas, das quais 1.998 também foram avaliadas em 2015. Os dados do PDDE de 2015 a 2018 incluíam informações de 7.371 escolas, sendo que 1.857 delas estavam entre as avaliadas em ambas as edições do SAEB. Após a filtragem para considerar somente as escolas avaliadas tanto em 2015 quanto em 2019, que possuíam informações válidas (não nulas) sobre o desempenho em Matemática, a amostra final deste estudo compreende 99.422 alunos distribuídos em 1.955 escolas. Para as escolas não presentes nos dados do PDDE, o valor total dos repasses foi considerado zero.

3.1.1 SAEB

Os dados do SAEB são produzidos a partir dos questionários aplicados a cada dois anos. As informações estão disponíveis ao público com acesso aos microdados, no entanto, existe um processo de desidentificação que torna os códigos das escolas fictícios, não

sendo possível integrar informações utilizando o código de escola ou mesmo acompanhar a evolução das escolas ao longo das edições do SAEB. Portanto, os dados utilizados no presente trabalho decorrem de pesquisa realizada com as bases de dados protegidos do (SAEB 2015, base da escola, e SAEB 2019, base da escola e base do aluno), acessadas mediante ingresso autorizado na Sala Segura do Serviço de Acesso à Dados Protegidos (SEDAP) do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Os resultados, análises e interpretações apresentados são de responsabilidade única do autor, não representando a visão oficial do INEP, nem se constituindo em estatística oficial. O acesso à Sala Protegida do INEP permite a manipulação dos dados sigilosos, neste caso, identificações das escolas, desde que se mantenha a proteção e não exista nenhuma forma de identificar as escolas na divulgação dos resultados. A concessão da autorização também impõe que o estudo seja de relevância para sociedade e que, ao final, seja formalizado em relatório para o INEP.

Entre os diversos instrumentos do SAEB, os utilizados neste estudo são:

- Teste de Matemática

Segundo o DAEB/INEP/MEC (2021a), o teste de matemática visa a “avaliação das capacidades como observação, estabelecimento de relações, comunicação (diferentes linguagens), argumentação e validação de processos, estimulando formas de raciocínio como intuição, indução, dedução e estimativa”. Os escores são obtidos pela Teoria de Resposta ao Item (TRI) e variam de 0 a 500 pontos, com média de 250 e desvio-padrão de 50 (DAEB/INEP/MEC, 2020).

- Questionário do aluno

Segundo o DAEB/INEP/MEC (2021a), o questionário do aluno visa fornecer “informações socioeconômicas das famílias dos alunos, aspectos do perfil do estudante, sua trajetória escolar, transporte utilizado para ir à escola e utilização do tempo fora da escola”.

3.1.2 PDDE

Os dados do PDDE (BRASIL, 2021) são consolidados majoritariamente a partir de cadastros manuais por parte de cada gestor ou secretaria de escola do Brasil utilizando o sistema online PDDEWeb do FNDE e registros gerados pelo banco responsável pelo pagamento dos recursos (Banco do Brasil). Os dados são públicos e podem ser acessados no também sistema online Consulta Escola do FNDE, com a devida identificação do

código único da escola (Código INEP). Para este estudo, a variável de maior relevância é o montante do repasse financeiro destinado a cada escola. Este valor é calculado com base em um valor fixo anual de R\$1.000,00, que varia conforme o tipo de instituição e a presença de Unidade Executora (UEx), e um valor per capita anual de R\$20,00 por aluno, ajustado por multiplicadores relacionados à localização (urbana ou rural), existência de UEx, natureza da instituição (pública ou privada de educação especial) e atendimento a alunos da educação especial ou de polos presenciais da Universidade Aberta do Brasil.

3.1.3 Período e variáveis

O período de avaliação foi definido em quatro anos para possibilitar uma análise mais clara e robusta de eventuais mudanças, em comparação com a observação em ciclos bienais do SAEB. A seleção do intervalo de 2015 a 2019 foi estrategicamente realizada para excluir os efeitos disruptivos causados pela pandemia de COVID-19 no sistema educacional.

Com o apoio das informações do Dicionário de Variáveis do SAEB 2019 (INEP/MEC, 2020), a Tabela 1 descreve as variáveis do estudo.

Tabela 1: Descrições das variáveis utilizadas no estudo

Variável	Descrição	Fonte
Aluno (Nível 1)		
PROFICIENCIA_MT_SAEB	Proficiência do aluno em Matemática	SAEB
TX_RESP_Q002	Qual é a sua cor ou raça?	SAEB
TX_RESP_Q004	Qual é a maior escolaridade da sua mãe?	SAEB
TX_RESP_Q015	Reprovação escolar	SAEB
VL_INSE_ALUNO	NSE do Aluno	SAEB
Escola (Nível 2)		
ID_AREA	Área (Capital/Interior)	SAEB
ID_ESCOLA	Código da Escola	PDDE e SAEB*
ID_LOCALIZACAO	Localização (Urbana/Rural)	SAEB
NIVEL_SOCIO_ECONOMICO	NSE da Escola	SAEB
MEDIA_9EF_MT	Média em Matemática 9º ano	SAEB
VALOR TOTAL	Valor total repassado a escola (ano, parcela)	PDDE

*Variável de integração entre as fontes

As variáveis explicativas testadas para integrar o modelo foram selecionadas com base na literatura. Para proporcionar uma melhor interpretação dos parâmetros do modelo, as variáveis descritas na Tabela 1 foram codificadas de forma que no modelo

considera-se a raça/cor (0=outros, 1=brancos e amarelos), a escolaridade da mãe (0 a 4), as reprovações (0 a 2), a área da escola (0=interior, 1=capital), a localização da escola (0=rural, 1=urbana) e o NSE da escola (1 a 7) de forma numérica. Seguindo as recomendações de Hox, Moerbeek e Schoot (2010), todas as variáveis explicativas foram centralizadas em suas médias visando facilitar a interpretação de possível interações no modelo.

No presente estudo, os escores relacionados ao NSE dos alunos foram calculadas pelo INEP. É comum na literatura que pesquisadores proponham técnicas para estimar esse indicador devido à ausência dessa informação nos microdados públicos do SAEB. Neste estudo, o acesso ao SEDAP permitiu a utilização das estimativas oficiais do INEP, disponibilizadas, até então, somente na Sala Protegida. De acordo com DAEB/INEP/MEC (2021b), os escores do Indicador de NSE dos estudantes foram estimados por meio do método a posteriori (EAP) e, posteriormente, submetidos a uma transformação linear, garantindo que o Indicador de NSE agregado para o Brasil apresentasse média 5 e desvio-padrão 1 e pudesse ser dividido em oito níveis, sendo o nível 1 o mais desfavorecido economicamente e o nível 8 o mais favorecido.

3.2 Análise dos dados

O SAEB tem seus dados dispostos em estrutura hierárquica, visto que, conforme o Relatório de Amostragem do SAEB (INEP/MEC, 2020), são coletadas informações das escolas e a partir delas são coletadas informações sobre os alunos. Diante disso, este estudo terá o rendimento em Matemática como variável resposta e as características da escola e dos alunos como variáveis explicativas. Em uma primeira etapa, todas as variáveis são provenientes do SAEB 2019 e o modelo foi construído a partir dos passos propostos por Hox, Moerbeek e Schoot (2010), descritos na Seção 2.2.3. Em uma segunda etapa, adiciona-se o rendimento em Matemática médio das escolas em 2015 como variável preditora como um controle para avaliar uma possível associação do rendimento em Matemática de 2019 com os repasses do PDDE de 2015 a 2018.

Para todos os processos desta monografia, as linguagens utilizadas foram o R (Versão 4.4.1) (R Core Team, 2021) com auxílio dos pacotes *lme4* (BATES et al., 2014) e *performance* (LUDECKE et al., 2021) e o Python (Versão 3.12.7) com auxílio da biblioteca *pandas*.

4 Resultados

4.1 Análise Descritiva

Tendo a proficiência dos alunos em Matemática como variável de interesse do estudo, assume-se que sua distribuição pode se apresentar de diferentes formas consoante o comportamento, o ambiente e as oportunidades dos alunos. Com base nas respostas do questionário do SAEB, é possível avaliar essas diferenças de acordo com escolaridade dos pais, raça/cor, entre outros.

4.1.1 Características dos alunos - Nível 1

Inicialmente, analisa-se a proficiência em Matemática. A Tabela 2 apresenta algumas medidas resumo para facilitar a interpretação da proficiência e, a partir dela, observa-se uma média e uma mediana próximas. A Figura 3 apresenta o histograma da proficiência e sugere uma distribuição aproximadamente simétrica. A Figura 4 apresenta o gráfico quantil-quantil da proficiência, no qual é possível observar que os desvios nas caudas a distâncias levemente de uma distribuição Normal.

Tabela 2: Medidas resumo da proficiência em Matemática

Estatística	Proficiência em Matemática
Média	265,9
Desvio Padrão	46,5
Mínimo	126,3
Primeiro Quartil	234,8
Mediana	267,8
Terceiro Quartil	298,0
Máximo	397,5

Figura 3: Distribuição da proficiência em Matemática

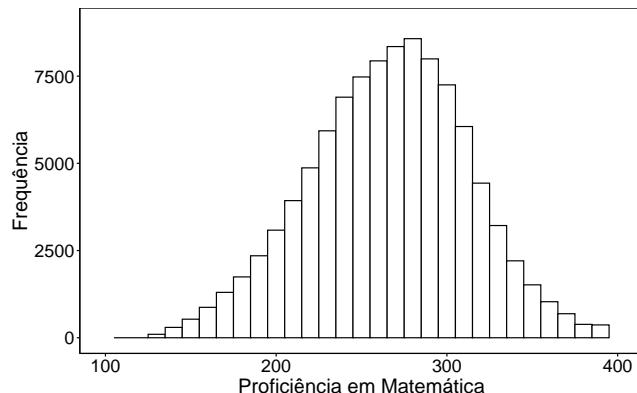
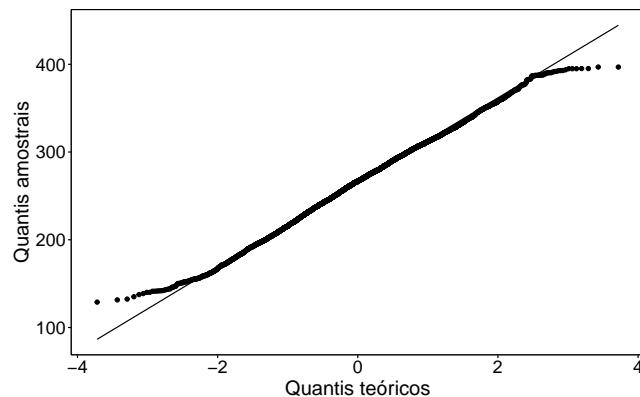
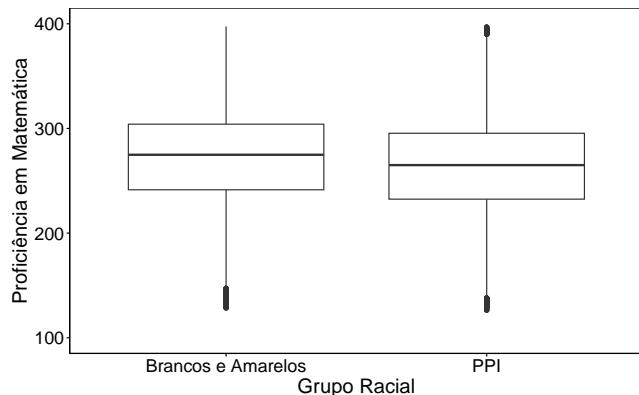


Figura 4: Gráfico de quantis da proficiência em Matemática



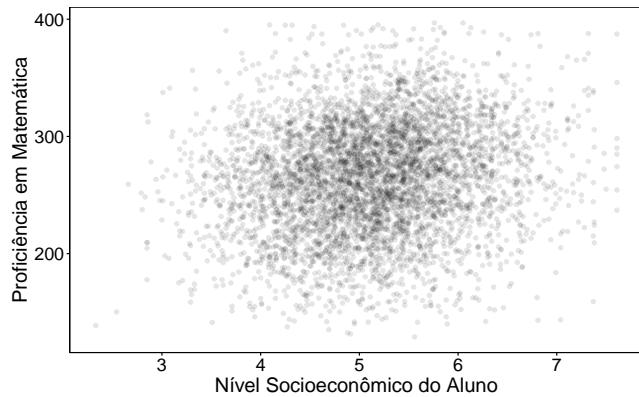
Para analisar a relação entre as proficiências e as características sociodemográficas, inicia-se pela raça/cor. Seguindo a abordagem adotada em outros estudos da literatura, como o trabalho de Vinha, Karino e Laros (2016), optou-se por categorizar a autodeclaração de raça/cor em dois grupos, de forma que cerca de 29% dos alunos são do grupo de brancos e amarelos e 71% do outro grupo, incluindo alunos pretos, pardos e indígenas. Pela Figura 5, que apresenta os boxplots da proficiência por raça/cor, é possível observar pequena diferença entre as medianas, de forma que, há uma sugestão de que o grupo de brancos e amarelos tendem a ter notas maiores que o outro grupo.

Figura 5: Proficiência em Matemática por raça/cor



Ao analisar a proficiência em Matemática tendo em vista o NSE do aluno, observa-se na Figura 6, que apresenta o gráfico de dispersão da proficiência por NSE do aluno, que não existe relação clara entre as duas variáveis. Observa-se pontos mais concentrados no centro com algumas dispersões, não indicando tendências crescentes.

Figura 6: Proficiência em Matemática por NSE do aluno

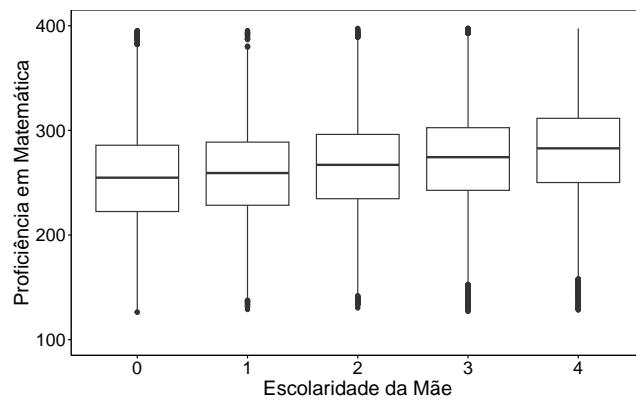


A escolaridade dos pais é um fator relevante para a compreensão do ambiente em que o estudante está inserido. Considerando a escolaridade da mãe, a Tabela 3, onde estão descritas as frequências relativas da escolaridade da mãe, mostra que uma parcela significativa dos alunos declara que suas mães não concluíram o ensino médio ou não souberam informar esse dado. Embora neste estudo, essa informação não esteja desagregada, pela informação ser baseada na autodeclaração dos estudantes, considera-se a possibilidade de inconsistências em relação à realidade. Ainda assim, a Figura 7, que apresenta os boxplots da proficiência por escolaridade da mãe, indica uma suave tendência de aumento na proficiência dos alunos à medida que a escolaridade da mãe se eleva.

Tabela 3: Frequências da escolaridade da mãe (%)

Escolaridade da mãe	Total
	n=99.422
0: Não completou o 5º ano do Ensino Fundamental/Não sabe	29,80
1: Ensino Fundamental, até o 5º ano	8,49
2: Ensino Fundamental completo	13,32
3: Ensino Médio completo	30,60
4: Ensino Superior completo (faculdade ou graduação)	17,79

Figura 7: Proficiência em Matemática por escolaridade da mãe

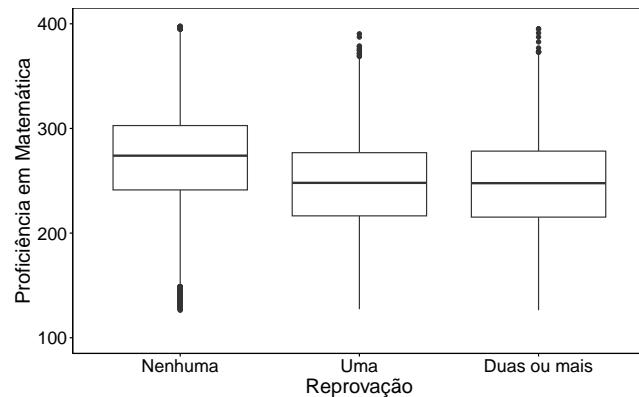


A autodeclaração do aluno sobre reprovações é uma variável importante para estudo, pois pela Figura 8, que apresenta os boxplots da proficiência por quantidade de reprovações, nota-se uma potente diferença entre a mediana da proficiência dos que nunca reprovaram e as medianas dos que reprovaram uma ou duas vezes. Destaca-se que, segundo a Tabela 4, onde observa-se as frequências relativas de reprovações, cerca de 8% reprovou mais de uma vez.

Tabela 4: Frequências de reprovações (%)

Reprovação	Total
	n=99.422
Não	77,17
1 vez	14,85
2 ou mais	7,98

Figura 8: Proficiência em Matemática por reaprovação



4.1.2 Características das escolas - Nível 2

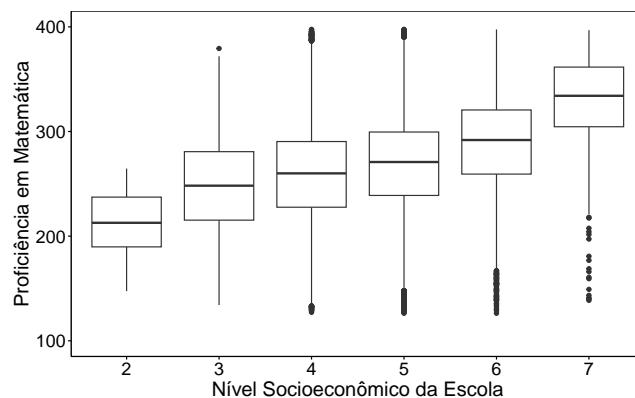
Características institucionais, como o nível socioeconômico médio e a localização da escola (urbana ou rural), desempenham um papel crucial, pois podem refletir desigualdades estruturais e impactar diretamente os resultados educacionais. Adicionalmente, a Tabela 5 apresenta as frequências relativas das características das escolas e, nesta, observa-se que quase 90% das escolas são de Nível IV e V quanto ao NSE médio e que 97% das escolas estão em localização urbana.

Tabela 5: Frequências das características das escolas (%)

Característica	Total n=1.955
Nível Socioeconômico	
1: Nível I	0
2: Nível II	0,02
3: Nível III	1,60
4: Nível IV	42,74
5: Nível V	46,78
6: Nível VI	8,25
7: Nível VII	0,60
Área	
Capital	32,27
Interior	67,73
Localização	
Urbana	97,03
Rural	2,97

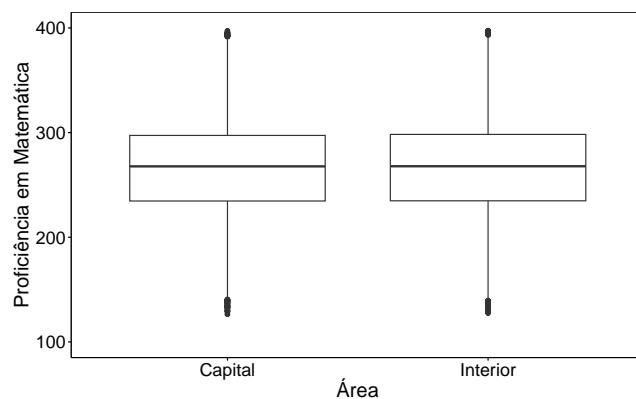
Na Figura 9, onde são apresentados os boxplots da proficiência por NSE médio (da escola), é possível observar grandes diferenças nas proficiências conforme o NSE da escola, sugerindo que quanto maior o nível, maior o desempenho. Ainda, na Figura 9, observa-se a presença de *outliers* na maioria dos níveis.

Figura 9: Proficiência em Matemática por NSE da escola



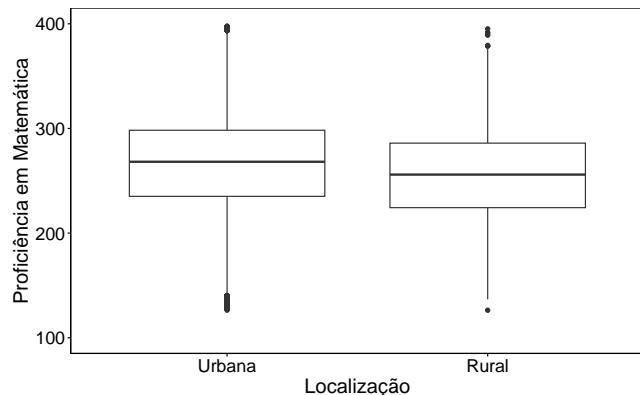
Quanto a área da escola, capital ou interior, a Figura 10, que apresenta os boxplots da proficiência por área da escola, sugere que as medianas das proficiências em Matemática são similares.

Figura 10: Proficiência em Matemática por área da escola



Na Figura 11, as medianas da proficiência sugerem diferenças conforme a localização da escola, urbana ou rural, de forma que alunos de escolas urbanas apresentam notas mais altas que os alunos de escolas rurais. No entanto, observa-se na Tabela 5, vista anteriormente, que cerca de 3% das escolas estão em localização rural.

Figura 11: Proficiência em Matemática por localização da escola



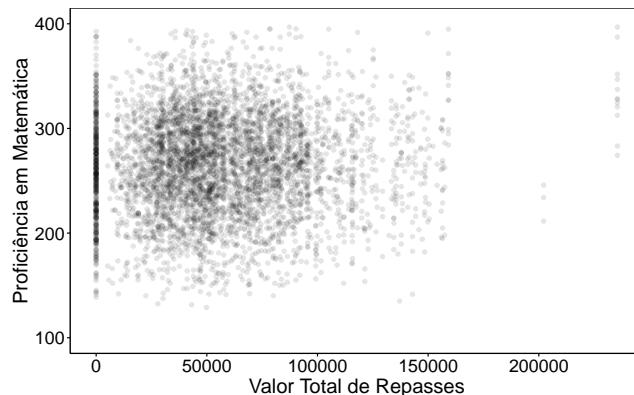
Para a análise das informações relacionadas ao PDDE, foi considerada a soma dos repasses financeiros recebidos por cada escola no período de 2015 a 2018. Além disso, foram incluídas na amostra escolas que não receberam recursos em todos os anos, atribuindo-se o valor zero. Conforme a Tabela 6, em que são apresentadas as medidas resumo do valor total, a soma dos repasses foi de em média R\$55.278,00, chegando a atingir um valor máximo de R\$235.500,00 de repasses nesses quatro anos avaliados.

Tabela 6: Medidas resumo do valor total de repasses do PDDE (R\$)

Estatística	Valor Total PDDE
Média	55.278
Desvio Padrão	35.164
Mínimo	0
Primeiro Quartil	32.060
Mediana	51.160
Terceiro Quartil	75.280
Máximo	235.500

Ao verificar a proficiência em Matemática por valor total, observa-se na Figura 12, onde é apresentado o gráfico de dispersão da proficiência por valor total, que não há indicações de tendências crescentes. Para ter ainda mais informações sobre essa relação, foi calculada também a Correlação de Pearson entre o valor total do PDDE e a proficiência e obtém-se 0,05, valor considerado muito baixo, indicando uma correlação fraca entre essas variáveis.

Figura 12: Proficiência em Matemática por valor total do PDDE



4.2 Análise Multinível

Tendo como base as informações obtidas na análise descritiva (Seção 4.1) e o interesse inicial do estudo, ajusta-se o modelo multinível.

4.2.1 Etapa 1: SAEB 2019

- Passo 1: Modelo nulo

Conforme a Tabela 7, onde é apresentado o modelo nulo, pode-se verificar que, pelo modelo, a proficiência média dos alunos é de 263,21. Observa-se também que o ICC foi de 0,121, logo, 12% da variância do desempenho dos alunos em Matemática está associada ao nível da escola, sendo razoável a aplicação da modelagem multinível.

Tabela 7: Estimativas do modelo multinível nulo

Efeito fixo	Efeito	P-valor
Intercepto	263,215	<0,001
Efeito aleatório		
Variância no Nível 1	1.881,33	
Variância no Nível 2	260,17	
ICC	0,121	

- Passo 2: Modelo com adição de variáveis do Nível 1 (aluno)

Do Nível 1, são inseridas as variáveis relacionadas à raça/cor, a escolaridade da mãe, a quantidade de reprovações e o nível socioeconômico do aluno. Nota-se na Tabela

8, em que é apresentado o modelo com adição de variáveis explicativas de efeito fixo do Nível 1, que a cada nível de reprovação (nenhuma, uma, duas ou mais), tem-se um efeito negativo de cerca de -13,5 pontos na proficiência em Matemática. Observando a escolaridade da mãe, observa-se que com o aumento de um nível de escolaridade (de 0 a 4), existe um aumento médio de aproximadamente 4 pontos na proficiência. Em raça/cor, observa-se que em média, os alunos do grupo de brancos e amarelos apresentam cerca de 4,1 pontos a mais que os demais. Destaca-se que este modelo explica 6,47% da variância do Nível 1 (dentro dos grupos) e 20,86% da variância do Nível 2 (diferença entre grupos).

Tabela 8: Estimativas do modelo multinível com variáveis do Nível 1

Efeito fixo	Efeito	P-valor
Intercepto	263,781	<0,001
Variáveis do Nível 1		
Raça/cor	4,124	<0,001
Escolaridade da mãe	4,081	<0,001
Reprovação	-13,496	<0,001
NSE do aluno	2,649	<0,001
Efeito aleatório		
Variância no Nível 1	1.759,66	
Variância no Nível 2	205,91	
Variância Explicada		
R^2_1 (Nível 1)	6,47%	
R^2_2 (Nível 2)	20,86%	

- Passo 3: Modelo com adição de variáveis do Nível 2 (escola)

Com adição da área, da localização e do NSE da escola, é possível perceber na Tabela 9, onde é apresentado o modelo com adição de variáveis explicativas de efeito fixo do Nível 2, que a cada aumento no nível socieconômico da escola (1 a 7), espera-se um aumento médio de cerca de 11,1 pontos da proficiência média da escola. Quanto a área da escola, o modelo sugere que em média, as escolas em capitais têm proficiência média 6,4 pontos menor que escolas no interior. Quanto a localização, espera-se que escolas urbanas tenham em média cerca de 3,3 pontos a mais do que as escolas em áreas rurais. Observa-se que a variância explicada no Nível 1 permanece a mesma do modelo anterior (6,47%), o que é esperado dado que foram adicionadas variáveis do Nível 2, e a variância

explicada no Nível 2 aumenta para 39,56%, indicando que a inclusão de variáveis de Nível 2 melhora substancialmente o ajuste do modelo.

Tabela 9: Estimativas do modelo multinível com variáveis dos níveis 1 e 2

Efeito fixo	Efeito	P-valor
Intercepto	265,152	<0,001
Variáveis do Nível 1		
Raça/cor	4,033	<0,001
Escolaridade da mãe	4,029	<0,001
Reprovação	-13,505	<0,001
NSE do aluno	2,325	<0,001
Variáveis do Nível 2		
NSE da escola	11,178	<0,001
Área da escola	-6,375	<0,001
Localização da escola	3,307	0,018
Efeito aleatório		
Variância no Nível 1	1.759,71	
Variância no Nível 2	157,24	
Variância Explicada		
R^2_1 (Nível 1)	6,47%	
R^2_2 (Nível 2)	39,56%	

- Passo 4: Modelo com adição de efeitos aleatórios do Nível 1

Ao adicionar efeitos aleatórios, nota-se a partir das variâncias na Tabela 10, que apresenta o modelo com adição de efeitos aleatórios do Nível 1, que o nível socioeconômico do aluno, a quantidade de reprovações, a escolaridade da mãe e a raça/cor tem efeitos diferenciados sobre o desempenho em Matemática dependendo da escola que o aluno frequenta. Nota-se também que a quantidade de reprovações especificamente apresenta grande efeito aleatório, com desvio padrão de 4,45 ($\sqrt{19,84}$). Com isso, assume-se que as reprovações podem ter grande efeito no desempenho em algumas escolas, enquanto em outras pode ter menor efeito.

Tabela 10: Estimativas do modelo multinível com variáveis dos níveis 1 e 2 com efeitos aleatórios

Efeito fixo	Efeito	P-valor
Intercepto	265,058	<0,001
Variáveis do Nível 1		
Raça/cor	4,078	<0,001
Escolaridade da mãe	4,045	<0,001
Reprovação	-13,610	<0,001
NSE do aluno	2,321	<0,001
Variáveis do Nível 2		
NSE da escola	10,869	<0,001
Área da escola	-6,314	<0,001
Localização da escola	3,045	0,027
Efeito aleatório		
Variância no Nível 1	1.749,48	
Variância no Nível 2	154,63	
Variância NSE do aluno	0,09	
Variância Reprovação	19,84	
Variância Escolaridade da mãe	0,13	
Variância Raça/cor	6,05	

- Passo 5: Modelo com adição de interações

Ao adicionar interações entre os diferentes níveis, houve significância nas interações entre nível socieconômico da escola e quantidade de reprovações e entre nível socieconômico da escola e escolaridade da mãe. Com isso, escolaridade da mãe deixou de ser considerado um efeito aleatório. Nota-se na Tabela 11, que apresenta o modelo com adição de interações entre variáveis explicativas do Nível 1 e do Nível 2, que com o aumento do nível socioeconômico da escola, há uma redução de cerca de 1,5 no efeito das reprovações no desempenho do aluno e uma redução de cerca de 0,4 no efeito da escolaridade da mãe no desempenho do aluno. Foram consideradas somente as interações entre o NSE da escola e as variáveis do nível do aluno, interações entre outras variáveis não foram testadas.

Tabela 11: Estimativas do modelo multinível com variáveis dos níveis 1 e 2 com efeitos aleatórios e interações

Efeito fixo	Efeito	P-valor
Intercepto	265,179	<0,001
Variáveis do Nível 1		
Raça/cor	4,065	<0,001
Escolaridade da mãe	4,027	<0,001
Reprovação	-13,667	<0,001
NSE do aluno	2,305	<0,001
Variáveis do Nível 2		
NSE da escola	11,104	<0,001
Área da escola	-6,212	<0,001
Localização da escola	2,934	0,033
Efeito de interação		
Reprovação x NSE da Escola	-1,551	<0,001
Escolaridade da mãe x NSE da Escola	-0,411	0,003
Efeito aleatório		
Variância no Nível 1	1.753,81	
Variância no Nível 2	155,02	
Variância NSE do aluno	0,02	
Variância Reprovação	20,58	
Variância Raça/cor	5,66	

4.2.2 Etapa 2: SAEB 2019, SAEB 2015 e PDDE 2015-2018

Nesta etapa ajusta-se o modelo utilizando como variável preditora o desempenho médio das escolas em 2015 para avaliar a associação do dinheiro investido nas escolas por meio do PDDE de 2015 a 2018 com o desempenho dos alunos.

- Modelo com adição de novas variáveis do Nível 2 (escola)

A Tabela 12, que mostra o modelo com adição de variáveis do Nível 2 referentes ao SAEB 2015 e ao PDDE, evidencia que com as técnicas aplicadas neste estudo, o valor total de repasses do PDDE não apresenta efeito significativo no desempenho dos alunos. Nota-se também que ao adicionar as novas variáveis, a localização da escola

perdeu significância, sendo retirada do modelo. Esse resultado confirma a ausência de correlação entre a proficiência em Matemática e o valor total do PDDE observada na análise descritiva. Por outro lado, com a entrada da média da escola em Matemática em 2015, observa-se uma diminuição de 40% na variância no Nível 2 em relação ao modelo anterior (Tabela 11). O efeito de 0,564 da média em 2015 é considerado elevado, uma vez que a proficiência em Matemática em 2019 (variável resposta) segue a mesma escala. Dessa forma, um incremento de 1 ponto na média de 2015 está associado a um aumento esperado de 0,564 pontos na proficiência em 2019.

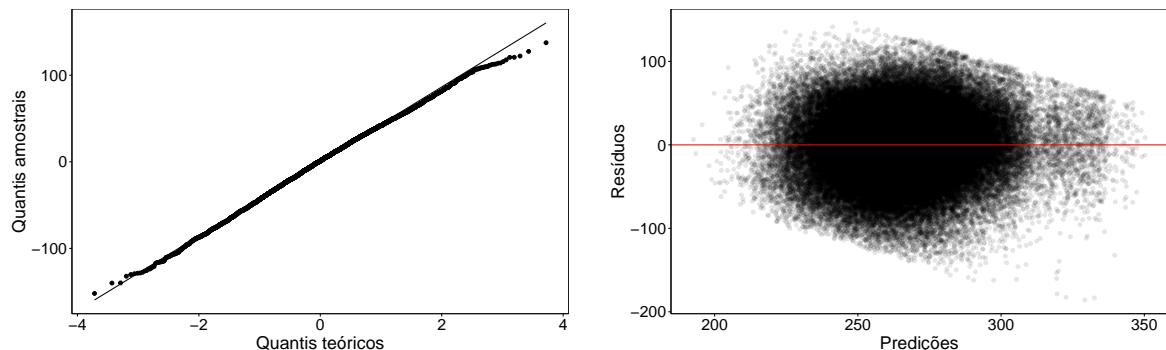
Tabela 12: Estimativas do modelo multinível com adição de variáveis do PDDE e do SAEB 2015

Efeito fixo	Efeito	P-valor
Intercepto	265,440	<0,001
Variáveis do Nível 1		
Raça/cor	3,955	<0,001
Escolaridade da mãe	4,017	<0,001
Reprovação	-13,770	<0,001
NSE do aluno	2,277	<0,001
Variáveis do Nível 2		
NSE da escola	4,341	<0,001
Área da escola	-4,674	<0,001
Valor total PDDE	0,000	0,432
Média SAEB em 2015	0,564	<0,001
Efeito de interação		
Reprovação x NSE da Escola	-1,370	<0,001
Escolaridade da mãe x NSE da Escola	-0,468	0,001
Efeito aleatório		
Variância no Nível 1	1.751,70	
Variância no Nível 2	91,64	
Variância NSE do aluno	0,65	
Variância Reprovação	7,31	
Variância Raça/cor	3,42	

4.2.3 Análise dos resíduos

Para verificar se as conclusões feitas sob o modelo ajustado na Seção 4.2.2 cumprem com as suposições da modelagem multinível, analisa-se os resíduos do Nível 1 e do Nível 2. Percebe-se no gráfico quantil-quantil da Figura 13 que os resíduos do Nível 1 (aluno) apresentam pequenos desvios nas caudas, onde há sugestão de violação de normalidade. Na análise do gráfico de dispersão dos resíduos em relação aos valores preditos é necessário atentar-se ao comportamento dos pontos no canto superior direito e no canto inferior esquerdo, que sugerem uma não aleatoriedade dos dados, evidenciando uma possível violação da suposição de homocedasticidade.

Figura 13: Distribuição dos resíduos no Nível 1



Quanto aos resíduos no nível da escola, não há sugestões de violação da homocedasticidade para os diferentes efeitos aleatórios. No entanto, observa-se principalmente no NSE do aluno na Figura 15 violação da suposição de normalidade.

Figura 14: Distribuição dos resíduos no Nível 2: Intercepto

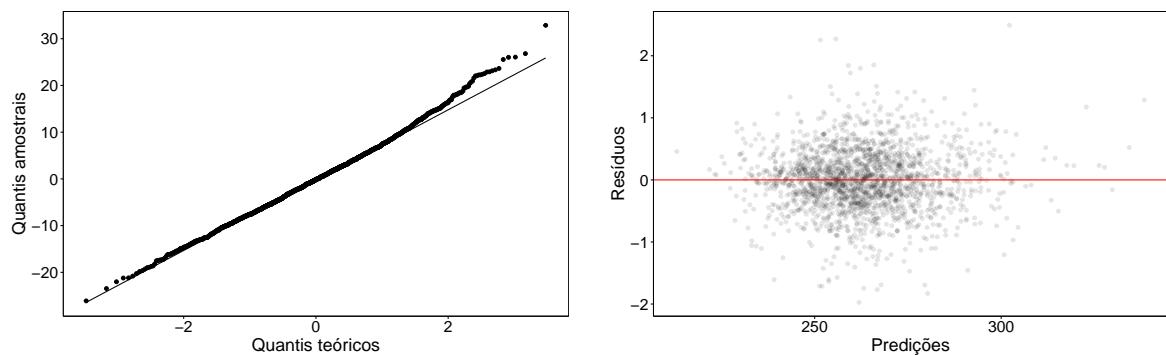


Figura 15: Distribuição dos resíduos no Nível 2: NSE do aluno

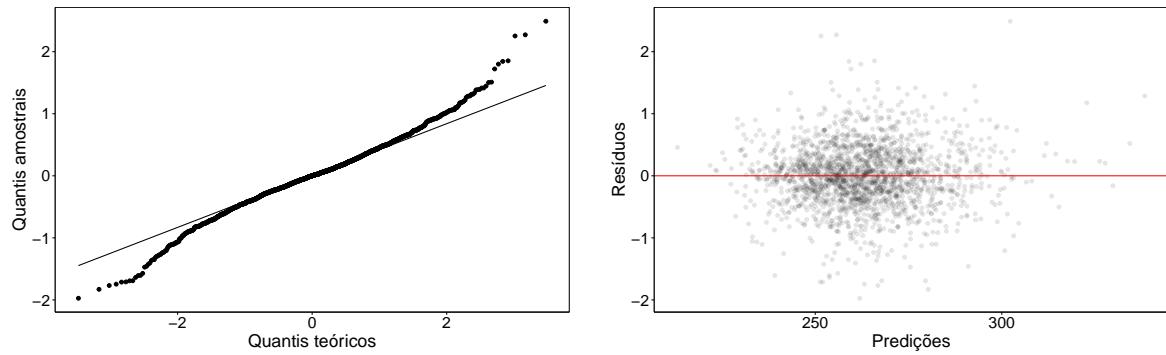


Figura 16: Distribuição dos resíduos no Nível 2: Raça/cor

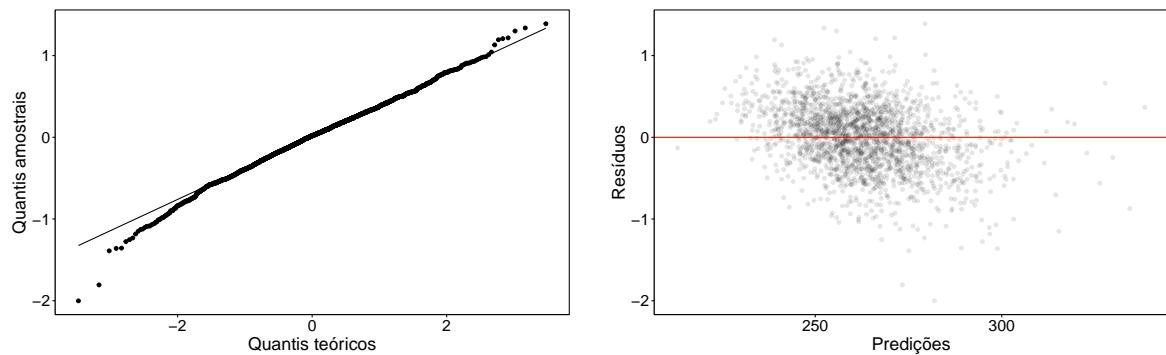
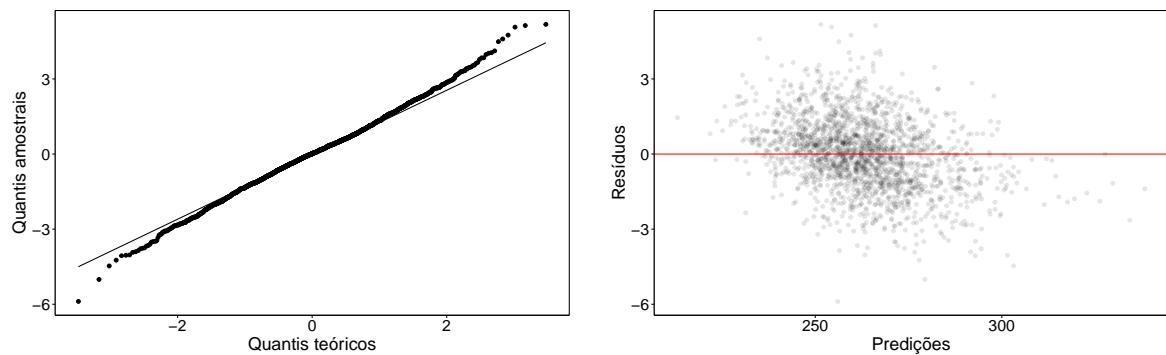


Figura 17: Distribuição dos resíduos no Nível 2: Reprovação



5 Conclusão

O presente estudo teve como objetivo identificar características dos alunos e das escolas associadas ao desempenho dos alunos em Matemática no 9º ano do ensino fundamental na região Centro-Oeste. A identificação foi realizada por meio da modelagem multinível utilizando os dados da edição de 2019 do SAEB, além de variáveis da edição de 2015 do SAEB e do PDDE de 2015 a 2019.

No modelo ajustado, as variáveis do nível do aluno que apresentaram efeito foram relacionadas à quantidade de reprovações, ao grupo racial, à escolaridade da mãe e ao nível socioeconômico do aluno. No nível da escola, os efeitos significativos são relacionados ao nível socioeconômico médio, área da escola (capital ou interior) e desempenho médio em Matemática da escola no SAEB de 2015. Destaca-se que com o aumento do nível socioeconômico da escola, há uma redução de cerca de 1,3 no efeito das reprovações no desempenho do aluno e uma redução de cerca de 0,5 no efeito da escolaridade da mãe no desempenho do aluno. Ademais, o nível socioeconômico do aluno, a quantidade de reprovações e o grupo racial tem efeitos diferenciados sobre o desempenho em Matemática dependendo da escola que o aluno frequenta. Por fim, não foi identificada diferença significativa no valor total de repasses do PDDE na explicação do desempenho dos alunos.

Uma das limitações deste estudo refere-se às possíveis violações das suposições de homocedasticidade e normalidade no nível do aluno e de normalidade no nível da escola, as quais poderiam ser tratadas por meio de técnicas estatísticas não consideradas neste estudo. No entanto, segundo a DAEB/INEP/MEC (2021a), os dados do SAEB são de natureza quase censitária, o que implica que, caso toda a população fosse considerada, os coeficientes estimados seriam semelhantes. Assim, espera-se que, dada a escala da amostra, eventuais desvios dessas suposições não comprometem significativamente as estimativas pontuais.

Para estudos futuros, pode-se considerar o uso de pacotes como o *misty* (YANAGIDA, 2024), que implementam métodos mais robustos de centralização de variáveis explicativas para modelos multinível, possibilitando a obtenção de resultados potencialmente distintos. Com relação ao dinheiro recebido pelas escolas, o SAEB inclui uma questão respondida pelos diretores das escolas sobre a suficiência dos recursos financeiros, que pode ser explorada em pesquisas futuras em conjunto com as informações do PDDE, permitindo uma análise mais aprofundada dos efeitos do financiamento direto na educação.

O estudo cumpriu com os objetivos de estudar o uso de modelos multinível no contexto das avaliações educacionais, integrar os dados do SAEB 2019 com os dados do SAEB 2015 e do PDDE de 2015 a 2018 e aplicar a modelagem multinível utilizando como variável preditora o desempenho médio das escolas em 2015 para avaliar a associação do dinheiro investido nas escolas por meio do PDDE com o desempenho escolar. Além disso, abriu portas para novos estudos no mesmo contexto, com melhores técnicas e variáveis e com o acesso ao SEDAP.

Referências

ANDRADE, J. M. d.; LAROS, J. A. Fatores associados ao desempenho escolar: estudo multinível com dados do saeb/2001. *Psicologia: Teoria e Pesquisa*, Instituto de Psicologia, Universidade de Brasília, v. 23, n. 1, p. 33–41, Jan 2007. ISSN 0102-3772. Disponível em: <<https://doi.org/10.1590/S0102-37722007000100005>>.

BATES, D. et al. Fitting linear mixed-effects models using the lme4 package in r. *J Stat Softw*, v. 67, p. 1–48, 2014.

BRASIL. *Constituição da República Federativa do Brasil promulgada em 5 de outubro de 1988*. [s.n.], 1988. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constitui%C3%A7ao.htm>.

BRASIL. *Lei de Diretrizes e Base da Educação Nacional n.º 9.394/96*. [s.n.], 1996. Disponível em: <<https://legislacao.presidencia.gov.br/atos/?tipo=LEI&numero=9394&ano=1996&ato=3f5o3Y61UMJpWT25a>>.

BRASIL. *Resolução N.º 15, de 16 de setembro de 2021*. [s.n.], 2021. Disponível em: <https://www.gov.br/fnde/pt-br/acesso-a-informacao/acoes-e-programas/programas/pdde/media-pdde/funcionamento/copy_of-RESOLUOPDDEN15DE16DESETEMBRODE2021.pdf>.

CHEIN, F. Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas. Escola Nacional de Administração Pública (Enap), 2019.

DAEB/INEP/MEC. Escalas de proficiência do saeb. 2020. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_examenes_da_educacao_basica/escalas_de_proficiencia_do_saeb.pdf>.

DAEB/INEP/MEC. Avaliação de Ciências Humanas e Ciências da Natureza no 9º ano do Ensino Fundamental. *Relatório de Resultados do SAEB 2019*, gov.br, v. 3, 2021. Disponível em: <https://download.inep.gov.br/educacao_basica/saeb/2019/resultados/relatorio_de_resultados_do_saeb_2019_volume_1.pdf>.

DAEB/INEP/MEC. Indicador de nível socioeconômico do saeb/2019. 2021. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e-indicadores/indicador_nivel_socioeconomico_saeb_2019_nota_tecnica.pdf>.

FERRAO, M. E. Introdução aos modelos de regressão multinível em educação. *Cadernos de Pesquisa*, v. 34, n. 121, p. 248, abr. 2004. Disponível em: <<https://publicacoes.fcc.org.br/cp/article/view/499>>.

HOX, J. Multilevel modeling: When and why. Nova Iorque, 1998. Disponível em: <<https://www.joophox.net/publist/whenwhy.pdf>>.

HOX, J.; MOERBEEK, M.; SCHOOT, R. van de. *Multilevel Analysis: Techniques and Applications, Second Edition*. Taylor & Francis, 2010. (Quantitative Methodology Series). ISBN 9781136975349. Disponível em: <<https://books.google.com.br/books?id=jLPHBQAAQBAJ>>.

- INEP/MEC. Microdados do saeb 2019. 2020. Disponível em: <<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/saeb>>.
- LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. *Biometrics*, [Wiley, International Biometric Society], v. 38, n. 4, p. 963–974, 1982. ISSN 0006341X, 15410420. Disponível em: <<http://www.jstor.org/stable/2529876>>.
- LAROS, J. A.; MARCIANO, J. L.; ANDRADE, J. M. d. Fatores associados ao desempenho escolar em Português: um estudo multinível por regiões. *Ensaio: Avaliação e Políticas Públicas em Educação*, scielo, v. 20, n. 4, p. 623–646, Dec 2012. ISSN 0104-4036. Disponível em: <http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S0104-40362012000400002&nrm=iso>.
- LEE, V. Utilização de modelos lineares hierárquicos lineares para estudar contextos sociais: o caso dos efeitos da escola. *Pesquisa em eficácia escolar: origem e trajetórias*, p. 273–296, 2008.
- LUDECKE, D. et al. performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, v. 6, n. 60, p. 3139, 2021.
- NETO, J.; ROSENBERG, L. Indicadores de qualidade do ensino e seu papel no sistema nacional de avaliação. *Em Aberto*, v. 15, n. 66, 1995. Disponível em: <<https://emaberto.inep.gov.br/ojs3/index.php/emaberto/article/view/2331>>.
- NUNES, A. M. R. *O modelo linear misto multinível na análise do efeito do desbaste de pinheiros na recuperação ecológica de uma pedreira calcária*. Dissertação (Mestrado) — Universidade de Lisboa, 2010.
- OSIO, M. *Análise de modelos de regressão multiníveis simétricos*. Dissertação (Mestrado) — Universidade de São Paulo, 2013. Disponível em: <https://www.teses.usp.br/teses/disponiveis/55/55134/tde-05072013-161440/publico/Dissertacao_Marina2.pdf>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>.
- RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical Linear Models: Applications and Data Analysis Methods*. [S.l.]: Sage, 2002.
- VINHA, L. G. d. A.; KARINO, C. A.; LAROS, J. A. Factors associated with mathematics performance in brazilian basic education. *Psico-USF*, SciELO Brasil, v. 21, p. 87–100, 2016.
- YANAGIDA, T. *misty: Miscellaneous Functions 'T. Yanagida'*. [S.l.], 2024. R package version 0.6.8. Disponível em: <<https://CRAN.R-project.org/package=misty>>.