

## EFEITOS DO CONTEXTO FAMILIAR E ESCOLAR NO DESEMPENHO EM MATEMÁTICA: UMA ABORDAGEM BAYESIANA COM DADOS DO SARESP

Camila Fernanda Bassetto Sampaio<sup>1</sup>

### Resumo

O presente estudo tem como principal objetivo a aplicação de modelos lineares generalizados hierárquicos para identificar os fatores que atuam sobre o desempenho escolar do aluno. Para alcançar tal objetivo, foram utilizados dados do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP), referentes à edição de 2019. Além dos microdados associados ao desempenho obtido pelos alunos da 3ª série do Ensino Médio pertencentes à rede pública de ensino, considerou-se também as respostas dos pais dos alunos dadas ao questionário socioeconômico aplicado por esta avaliação de larga escala. O Índice Socioeconômico (INSE) das escolas estaduais de São Paulo, que participaram da avaliação em 2019, foram inseridos na amostra elaborada na presente pesquisa. Nos modelos utilizados, busca-se medir a probabilidade de o estudante obter uma classificação satisfatória na proficiência em matemática, considerando variáveis específicas dos alunos e de contexto escolar e familiar. A estimação dos parâmetros dar-se-á via modelos hierárquicos estruturados em dois níveis, por meio do software RStudio. A identificação e a compreensão dos fatores extra e intra-escolares que atuam sobre o desempenho escolar do aluno é importante para subsidiar os agentes da Secretaria Educacional da Educação nas definições voltadas às políticas educacionais direcionadas às melhorias na Educação Básica brasileira. **Palavras-chave:** Modelos lineares generalizados, proficiência em matemática, Ensino Médio.

### Abstract

The objective of the present study is to apply hierarchical generalized linear models to identify the factors that influence students' academic performance. To this end, data from the 2019 edition of the São Paulo State School Performance Assessment System (SARESP) were employed. In addition to the microdata concerning the performance of 3rd-year high school students enrolled in the public education system, responses provided by their parents to the socioeconomic questionnaire administered as part of this large-scale assessment were also considered. The Socioeconomic Index (INSE) of São Paulo state schools that participated in the 2019 assessment was incorporated into the sample used in this research. The models aim to estimate the probability that a student achieves a satisfactory mathematics proficiency level, taking into account individual-level variables as well as contextual factors related to the school and family environment. Parameter estimation is

---

<sup>1</sup>Universidade Estadual Paulista "Júlio de Mesquita Filho- Faculdade de Ciências e Letras de Araraquara - UNESP/FCLAR

conducted through two-level hierarchical models using the RStudio software environment. Identifying and understanding both intra- and extra-school factors that affect student performance is important for informing decision-making by educational authorities and guiding the development of public policies aimed at improving the quality of Basic Education in Brazil. **Keywords:** Generalized linear models, mathematics proficiency, High school.

## Introdução

A partir da década de 1960, o interesse pela investigação dos fatores que interferem no desempenho escolar tem crescido, gradativamente, no âmbito nacional e internacional. Parte superior do formulário

No Brasil, as pesquisas sobre o efeito e a qualidade da escola são bem recentes, com os primeiros resultados publicados na década de 1990, quando o Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (INEP), do Ministério da Educação (MEC), passou a disponibilizar dados educacionais. Os estudos intensificaram-se a partir da consolidação do Sistema de Avaliação da Educação Básica (SAEB), cujos resultados despertaram atenção por evidenciarem um quadro de desigualdades educacionais no país, impulsionando a realização dos primeiros estudos que analisaram o desempenho escolar (Fletcher, 1998; Barbosa e Fernandes, 2001; Ferrão et al., 2001).

A avaliação externa é entendida como aquela que pode aquilatar o desempenho dos estudantes e, conseqüentemente, a qualidade social da educação, utilizando seus dados para balizar políticas públicas educacionais voltadas para a melhoria do ensino. Para tanto, se vale de escalas de proficiência nas quais seus intervalos mostram a aquisição de competências no componente curricular avaliado (Borges et al., 2022). Dessa forma, as competências se referem a uma matriz de referência que não pode ser confundida com propostas curriculares, estratégias de ensino ou diretrizes pedagógicas.

Intensificada no início da década de 1990, a avaliação dos sistemas educacionais tornou-se um ponto relevante nas propostas de políticas públicas ligadas à educação, estando relacionada com a necessidade de um melhor gerenciamento dos recursos disponíveis, direcionando, de forma adequada, os investimentos na escola pública, e com a necessidade de diagnosticar o ensino público objetivando efetivas melhorias. Nos últimos anos, é perceptível o avanço obtido na área educacional considerando os recursos metodológicos disponíveis e os atuais critérios de avaliação, tais como o Sistema Nacional de Avaliação da Educação Básica (SAEB), o Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (SARESP), o Exame Nacional do Ensino Médio (ENEM), entre outros. Tais sistemas têm revelado um quadro crítico para a educação, o que fortalece a necessidade de investigar os fatores que influenciam o desempenho dos alunos com o objetivo de esclarecer o processo de produção das desigualdades educacionais observadas nos resultados das avaliações.

Nas avaliações educacionais, vários fatores associados às variáveis internas e externas à escola, podem influenciar o desempenho escolar (Jesus e Laros, 2004; Riani e Rios-Neto, 2008; Laros et al., 2010). Neste sentido, buscando contribuir com a literatura no que se refere à investigação de fatores que impactam o desempenho escolar do indivíduo, o objetivo proposto na presente pesquisa consiste na utilização de modelos lineares generalizados hierárquicos com dados resultantes do SARESP, correspondentes às notas e classificações obtidas em Matemática e às respostas dadas ao questionário socioeconômico respondido pelos pais dos alunos participantes da avaliação de 2019. O índice de nível socioeconômico (INSE) das escolas estaduais do estado de São Paulo também foi considerado. A estimação dos modelos propostos foi obtida por simulação usando o *software* RStudio.

Os resultados obtidos mostram que as variáveis associadas à reprovação exerceram um efeito negativo sobre a proficiência em Matemática, demonstrando que alunos com histórico de defasagem escolar, devido a reprovações, apresentam coeficientes negativos de regressão em relação a essas variáveis. Por outro lado, a escolaridade materna foi identificada como um fator que influencia positivamente o desempenho escolar dos alunos.

## Uso de modelos hierárquicos na análise do desempenho escolar

Indivíduos tendem a viver dentro de estruturas organizacionais, tais como família, escolas, igrejas, cidades, estados e países. No ambiente educacional, estudantes são organizados em classes ou turmas, as quais, por sua vez, são definidas dentro de escolas. Esta distribuição sugere, para análise destes dados, a utilização de um procedimento estabelecido em dois estágios, sendo o primeiro composto por uma amostra de escolas e o segundo por amostras de alunos dentro de cada escola. A existência de dados hierarquicamente estruturados não é acidental e, na investigação dos fatores que atuam sobre o desempenho escolar do aluno, tal estrutura não pode ser ignorada (Goldstein, 2001). Uma metodologia estatística utilizada para analisar, simultaneamente, as relações a nível de aluno, consiste nos modelos lineares hierárquicos (Raudenbush e Bryk, 2002).

Os modelos de regressão hierárquica tem sido implementados no Brasil, desde o final dos anos 90, para analisar os dados das avaliações educacionais em larga escala, como mostram os estudos de Fletcher (1998), Ferrão et al. (2001), Alves e Soares (2008), Soares (2005), Andrade e Laros (2007), entre outros. Considerado um dos pioneiros na utilização de modelos lineares hierárquicos no Brasil, a pesquisa educacional realizada por Fletcher (1998) verificou que a proporção de alunos brancos e do sexo masculino na escola causam efeitos positivos sobre o desempenho em matemática. Nos estudos de Soares e Mendonça (2003), Riani e Rios-Neto (2008) e de Brooke et al. (2014), os resultados indicaram que os alunos masculinos e negros têm desempenho inferior. A identificação dos fatores que afetam o desempenho dos estudantes foi objeto de estudo também de Laros et al. (2010), os quais propuseram um modelo multinível com dados do SAEB de 2001 e

consideraram o modelo empírico apresentado por Jesus e Laros (2004), com dois níveis, isto é, aluno e escola. De acordo com resultados, a variável que mais afetou o desempenho escolar dos alunos foi o atraso escolar, seguida de o aluno gosta de estudar matemática. A quantidade de faltas e o índice de repetência afetaram negativamente o desempenho acadêmico. Diferentemente dos demais estudos citados, Laros et al. (2010) sugerem intervenções que a família e a escola podem realizar para aumentar o desempenho do aluno. No que se refere à família, os autores destacam o incentivo ao estudo desde a infância por meio de jogos e atividades lúdicas e evitar que os filhos ingressem e permaneçam na escola com distorção idade-série. Investigações na área de educação escolar mostram que vários fatores podem influenciar a trajetória escolar dos indivíduos e que a educação gira em torno de dois eixos, os quais devem ser monitorados de forma a captar os efeitos de cada um sobre a proficiência e sobre a qualidade do ensino.

Moreira et al. (2017) utilizaram dados do Sistema de Avaliação do Rendimento Escolar do Rio Grande do Sul (SAERS) do ano de 2007 para investigar os determinantes do desempenho escolar no estado do Rio Grande do Sul. Os autores concluíram que características socioeconômicas dos alunos, representadas pelos itens que os alunos possuem e não possuem em casa, são importantes determinantes do desempenho escolar, enquanto as características das escolas, representadas por características dos professores, diretores e infraestrutura da escola, pouco contribuem na explicação do rendimento escolar.

O estudo de Palermo et al. (2014) considerou a proficiência em Matemática como variável dependente e utilizou dados da Prova Brasil de 2007 referentes aos alunos do 5º ano do Ensino Fundamental no Rio de Janeiro. Seguindo a metodologia proposta por Barbosa e Fernandes (2001), os autores consideraram a modelagem hierárquica com dois e três níveis. Para o modelo com dois níveis, foram testados alunos e turmas e também alunos e escolas, e para o modelo com três níveis foram considerados alunos, turmas e escolas, nesta ordem. Em relação ao nível dos alunos, os autores identificaram uma relação diretamente proporcional entre o nível socioeconômico, representado pela soma ponderada de bens de consumo disponíveis nos domicílios e da escolaridade máxima dos pais, sobre o desempenho em Matemática, sugerindo que alunos com nível socioeconômico acima da média da turma teria a proficiência aumentada em um décimo.

Os estudos realizados no Brasil sobre os determinantes educacionais enfatizam o papel da família na educação dos filhos, mostrando que os fatores familiares, principalmente o nível de instrução dos pais, possuem grande influência sobre o desempenho educacional. Pesquisas que incluem, além dos fatores familiares, aqueles relacionados à escola e à sociedade em geral mostram que, apesar do importante impacto das características, os demais fatores citados também provocam efeito significativo na proficiência do aluno. Concomitantemente, a partir de estudos citados anteriormente, é evidente o impacto do nível socioeconômico no desempenho escolar dos alunos e, por isso, no estudo de qualquer realidade educacional deve-se, necessariamente, considerar a condição socioeconômica do

aluno. Características sociodemográficas como sexo e defasagem existente entre a idade do aluno e da série em que se encontra, sintetizam experiências que impactam seu aprendizado.

As análises desenvolvidas mostram-se relevantes, pois, além de auxiliar na formulação de políticas públicas educacionais mais eficazes, possibilitam investigar até que ponto os fatores escolares minimizam a influência da família no resultado educacional do indivíduo.

## Material e Método

É pauta frequente, na maioria das escolas, a realização de avaliações externas do desempenho educacional dos alunos. O estado de São Paulo, além de participar das avaliações nacionais, promove o Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo – SARESP, cuja finalidade é fornecer informações consistentes sobre a situação da escolaridade na rede pública estadual, capazes de orientar os gestores no monitoramento das políticas voltadas à melhoria da qualidade do ensino.

A partir das expectativas de aprendizagem quanto ao conteúdo, competências e habilidades, estabelecidas para cada ano/série da disciplina no Currículo do Estado de São Paulo, os pontos da escala do SARESP são agrupados em quatro níveis: Abaixo do Básico (AB), Básico (BA), Adequado (AD) e Avançado (AV). O Quadro 1 contém informações referentes aos níveis de proficiência, intervalos de pontuação e descrição, utilizados pelo SARESP. A primeira coluna mostra os níveis de proficiência, os quais, combinados com seus intervalos de pontuação, são classificados em Insuficiente (AB), Suficiente (BA e AD) e Avançado (AV). A segunda coluna contém a descrição de cada nível de proficiência.

**Quadro 1** – Pontuação e descrição dos níveis de proficiência em Matemática

| Nível de proficiência                                  | Pontuação e descrição                                                                                                                            |
|--------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Abaixo do Básico (AB)</b>                           | $< 275$ : Domínio insuficiente dos conteúdos, competências e habilidades desejáveis para o ano/série em que se encontra.                         |
| <b>Básico (BA)</b>                                     | $275 \text{ a } < 350$ : Domínio mínimo dos conteúdos, competências e habilidades desejáveis para o ano/série em que se encontra.                |
| <b>Adequado (AD)</b>                                   | $350 \text{ a } < 400$ : Domínio pleno dos conteúdos, competências e habilidades desejáveis para o ano/série em que se encontra.                 |
| <b>Avançado (AV)</b>                                   | $\geq 400$ : Conhecimento e domínio acima do requerido dos conteúdos, competências e habilidades desejáveis para o ano/série em que se encontra. |
| FONTE: Sumário Executivo SARESP 2019 (São Paulo, 2019) |                                                                                                                                                  |

A Secretaria da Educação do Estado de São Paulo (SEDUC/SP) realizou, em 2019, a 22ª edição do SARESP, caracterizada como uma avaliação em larga escala da Educação

Básica, aplicada desde 1996. Participam desta avaliação escolas municipais, estaduais e particulares, e as Escolas Técnicas Estaduais (ETE), administradas pelo Centro Educacional Tecnológico Paula Souza e vinculadas à Secretaria Estadual de Desenvolvimento do estado de São Paulo. A realização do SARESP envolve alunos, pais de alunos, escolas, diretores, coordenadores, professores, aplicadores e fiscais, e utiliza-se de dois instrumentos de avaliação, sendo o primeiro caracterizado pela aplicação de provas, e o segundo correspondendo à aplicação de questionários direcionados aos pais e alunos, diretores, professores coordenadores e demais professores (Brasil, 2019).

Os questionários respondidos pelos pais e por alunos, permitem coletar informações sobre o contexto social, econômico, cultural e familiar dos alunos, sobre a trajetória escolar, hábitos de estudo e percepções e expectativas a respeito dos professores e da gestão escolar, e sobre o perfil e aspectos da gestão escolar e da prática pedagógica. Dada a grande quantidade de variáveis disponibilizadas pelos questionários aplicados, aos alunos e pais dos alunos, o método Lasso (Izbicky, 2018), foi utilizado com o pacote glmnet e da função cv.glmnet, a fim de reduzir a dimensão dos dados.

O método selecionou as variáveis associadas às questões apresentadas no Quadro 2.

**Quadro 2** – Questões selecionadas via método Lasso

| Variáveis                                                                           | Alternativas             |
|-------------------------------------------------------------------------------------|--------------------------|
| <b>Sexo</b>                                                                         | (A) Feminino             |
|                                                                                     | (B) Masculino            |
| <b>Período</b>                                                                      | (A) Manhã                |
|                                                                                     | (B) Tarde                |
|                                                                                     | (C) Noite                |
| <b>10A.</b> Você já foi reprovado?                                                  | (A) Não                  |
|                                                                                     | (B) Sim, 1 vez           |
|                                                                                     | (C) Sim, 2 vezes ou mais |
| Seus pais ou responsáveis: <b>Q13A.</b> Ajudam na lição ou conferem se fez a lição? | (A) Sim, muito           |
|                                                                                     | (B) Sim, um pouco        |
|                                                                                     | (C) Não                  |

| Variáveis                                                 | Alternativas                                               |
|-----------------------------------------------------------|------------------------------------------------------------|
| <b>Q22A.</b> Gosta de estudar Matemática?                 | (A) Sim                                                    |
|                                                           | (B) Não                                                    |
| <b>Q23A.</b> Faz a lição de Matemática?                   | (A) Sempre ou quase sempre                                 |
|                                                           | (B) De vez em quando                                       |
|                                                           | (C) Nunca                                                  |
|                                                           | (D) Nunca tenho lição de casa                              |
| <b>Q24A.</b> Gosta de estudar Língua Portuguesa?          | (A) Sim                                                    |
|                                                           | (B) Não                                                    |
| Qual sua forma preferida de estudar:                      |                                                            |
| <b>Q27A.</b> Prestando atenção nas aulas?                 | (A) Não é o que gosto/faço                                 |
| <b>Q28A.</b> Lendo anotações em casa? <b>Q31A.</b>        | (B) Gosto só um pouco                                      |
| Fazendo pesquisas na Internet.                            | (C) Gosto bastante                                         |
| <b>Q19P.</b> Até que série/nível de ensino a mãe estudou? | (A) Nunca estudou ou não completou a 4ª série/5º ano       |
|                                                           | (B) Completou a 4ª série/5º ano, mas não a 8ª série/9º ano |
|                                                           | (C) Completou a 8ª série/9º ano, mas não o Ensino Médio    |
|                                                           | (D) Completou o EM, mas não o ES                           |
|                                                           | (E) Completou o ES                                         |
|                                                           | (F) Completou a Pós-graduação                              |
|                                                           | (G) Não sei                                                |

| Variáveis                                                           | Alternativas                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Q23P.</b> Qual é a renda familiar no seu domicílio?              | <p>(A) Até 1 SM (R\$ 998,00)</p> <p>(B) De 1 a 2 SM (R\$ 998,01 a R\$ 1.996,00)</p> <p>(C) De 2 a 3 SM (R\$ 1.996,01 a R\$ 2.994,00)</p> <p>(D) De 3 a 5 SM (R\$ 2.994,01 a R\$ 4.990,00)</p> <p>(E) De 5 a 8 SM (R\$ 4.990,01 a R\$ 7.984,00)</p> <p>(F) De 8 a 15 SM (R\$ 7.984,01 a R\$ 14.970,00)</p> <p>(G) Mais de 15 SM (mais de R\$ 14.970,01)</p> <p>(H) Não sei/não quero responder</p> |
| Na sua casa você tem: <b>Q27P.</b> Livros (romances, poesias etc.). | <p>(A) Sim</p> <p>(B) Não</p>                                                                                                                                                                                                                                                                                                                                                                     |
| FONTE: Questionário socioeconômico aplicado pelo SARESP             |                                                                                                                                                                                                                                                                                                                                                                                                   |

Uma vez que a presente pesquisa considera uma abordagem com modelos lineares generalizados (McCullagh, 2019), faz-se necessário categorizar as variáveis. Para as questões apresentadas no Quadro 2, utilizou-se a notação dada no Quadro 3.

**Quadro 3** – Codificação das variáveis

| Questão       | Notação   | Codificação                            | Questão | Notação    | Codificação                                             |
|---------------|-----------|----------------------------------------|---------|------------|---------------------------------------------------------|
| Classificação | $Y_{ij}$  | 0, se Insuficiente<br>1, se Suficiente | Q27A    | $x8_{ij}$  | 0, se A<br>1, se B ou C                                 |
| Sexo          | $x1_{ij}$ | 0, se feminino<br>1, se masculino      | Q28A    | $x9_{ij}$  | 0, se A<br>1, se B ou C                                 |
| Período       | $x2_{ij}$ | 1, se manhã<br>0, se tarde ou noite    | Q31A    | $x10_{ij}$ | 0, se A<br>1, se B ou C                                 |
| Q10A          | $x3_{ij}$ | 1, se A<br>0, se B ou C                | Q20P    | $x11_{ij}$ | 0, se A ou B<br>1, se C ou D<br>2, se E ou F            |
| Q13A          | $x4_{ij}$ | 0, se C<br>1, se A ou B                | Q23P    | $x12_{ij}$ | 0, se A ou B<br>1, se C ou D<br>2, se E ou F<br>3, se G |
| Q22A          | $x5_{ij}$ | 0, se B<br>1, se A                     | Q27P    | $x13_{ij}$ | 0, se B<br>1, se A                                      |

| Questão                       | Notação   | Codificação                                                                        | Questão | Notação   | Codificação        |
|-------------------------------|-----------|------------------------------------------------------------------------------------|---------|-----------|--------------------|
| Q23A                          | $x6_{ij}$ | 0, se C ou D<br>1, se A ou B                                                       | Q24A    | $x7_{ij}$ | 0, se B<br>1, se A |
| INSE                          |           | 0, se [3,98; 4,99)<br>1, se [4,99; 5,1)<br>2, se [5,1; 5,22)<br>3, se [5,22; 5,92] |         |           |                    |
| FONTE: Elaboração dos autores |           |                                                                                    |         |           |                    |

Para a Classificação, a codificação denominada Insuficiente contempla os níveis de proficiência Abaixo do Básico (AB) e Básico (BA), e a Suficiente abarca os níveis Adequado (AD) e Avançado (AV). O valor do INSE varia entre 3,98 e 5,92. Foi codificado como 0 se pertencer ao intervalo [3,98; 4,99). A combinação destas variáveis permitiu elaborar diferentes modelos logísticos com o objetivo de estimar a probabilidade de o aluno da 3ª série do EM classificar-se em Suficiente na proficiência em Matemática. Usando a codificação das variáveis selecionadas, o modelo logístico resultante é dado como na equação (1).

$$P(Y_{ij} = 1 | X = x_{ij}) = \frac{e^{\left(\beta_0 + \sum_{k=1}^{31} \beta_k x_{kij}\right)}}{1 + e^{\left(\beta_0 + \sum_{k=1}^{31} \beta_k x_{kij}\right)}} \quad (1)$$

Utiliza-se o logit, dado na expressão (2) para facilitar a compreensão dos resultados.

$$\text{logit}(\pi_{ij}) = \ln \left[ \frac{\pi_{ij}}{1 - \pi_{ij}} \right] = \beta_0 + \sum_{k=1}^{31} \beta_k x_{kij} \quad (2)$$

Para os modelos utilizados, o valor da *deviance* foi considerado como critério de seleção, buscando-se, assim, identificar aquele com maior ajuste aos dados e, simultaneamente, maior poder preditivo. O modelo selecionado por este critério será empregado na estimação dos parâmetros associados às variáveis preditoras que o compõe.

O modelo selecionado é dado como na expressão (3).

$$P(Y_{ij} = 1 | X = x_{ij}) = \frac{e^{\left(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij} + \beta_7 x_{7ij} + \beta_8 x_{8ij} + \beta_9 x_{9ij} + \beta_{10} x_{10ij} + \beta_{11} x_{11ij} + \beta_{12} x_{12ij} + \beta_{13} x_{13ij} + \beta_{14} x_{14ij} + \beta_{15} x_{15ij} + \beta_{16} x_{16ij} + \beta_{17} x_{17ij} + \beta_{18} x_{18ij} + \beta_{19} x_{19ij} + \beta_{20} x_{20ij} + \beta_{21} x_{21ij} + \beta_{22} x_{22ij} + \beta_{23} x_{23ij} + \beta_{24} x_{24ij} + \beta_{25} x_{25ij} + \beta_{26} x_{26ij} + \beta_{27} x_{27ij} + \beta_{28} x_{28ij} + \beta_{29} x_{29ij} + \beta_{30} x_{30ij} + \beta_{31} x_{31ij}\right)}}{1 + e^{\left(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij} + \beta_7 x_{7ij} + \beta_8 x_{8ij} + \beta_9 x_{9ij} + \beta_{10} x_{10ij} + \beta_{11} x_{11ij} + \beta_{12} x_{12ij} + \beta_{13} x_{13ij} + \beta_{14} x_{14ij} + \beta_{15} x_{15ij} + \beta_{16} x_{16ij} + \beta_{17} x_{17ij} + \beta_{18} x_{18ij} + \beta_{19} x_{19ij} + \beta_{20} x_{20ij} + \beta_{21} x_{21ij} + \beta_{22} x_{22ij} + \beta_{23} x_{23ij} + \beta_{24} x_{24ij} + \beta_{25} x_{25ij} + \beta_{26} x_{26ij} + \beta_{27} x_{27ij} + \beta_{28} x_{28ij} + \beta_{29} x_{29ij} + \beta_{30} x_{30ij} + \beta_{31} x_{31ij}\right)}} \quad (3)$$

Considerando o modelo linear generalizado hierárquico, com o nível 1 representado por alunos e o nível 2 pelas escolas, estimativas foram obtidas a fim de conhecer a probabilidade de o aluno alcançar a classificação Suficiente na proficiência em Matemática.

## Resultados e Discussão

O modelo hierárquico dado na expressão (3) foi considerado para todas as regiões do estado de São Paulo, isto é, Capital, Centro, Litoral, Nordeste, Noroeste, Norte, Sudeste e Sudoeste, estabelecidas conforme a Diretoria de Ensino, permitindo fazer uma análise regional. Na Tabela 1 são apresentadas as estimativas junto com os erros padrão (Est.) para cada variável e também as razões de chances (Exp.).

**Tabela 1** – Estimativas regionais associadas ao modelo selecionado.

| Variável                      | Est.              | Exp.  | Variável | Est.              | Exp.  | Variável | Est.              | Exp.  |
|-------------------------------|-------------------|-------|----------|-------------------|-------|----------|-------------------|-------|
| <b>Capital – INSE = 5,188</b> |                   |       |          |                   |       |          |                   |       |
| Intercepto                    | -1,332<br>(0,059) | 0,264 | Q22A1    | 0,952<br>(0,024)  | 2,591 | Q23P2    | 0,459<br>(0,054)  | 1,583 |
| SEXO1                         | 0,298<br>(0,027)  | 1,348 | Q24A1    | -0,416<br>(0,034) | 0,660 | Q23P3    | 0,022<br>(0,215)  | 1,019 |
| PERIODO1                      | 0,454<br>(0,031)  | 1,574 | Q20P1    | 0,218<br>(0,029)  | 1,245 | INSE1    | 0,173<br>(0,046)  | 1,188 |
| Q10A1                         | 0,812<br>(0,036)  | 2,255 | Q20P2    | 0,157<br>(0,042)  | 1,171 | INSE2    | 0,305<br>(0,048)  | 1,356 |
| Q13A1                         | -0,393<br>(0,026) | 0,676 | Q23P1    | 0,301<br>(0,028)  | 1,352 | INSE3    | 0,625<br>(0,059)  | 1,869 |
| <b>Centro – INSE = 5,317</b>  |                   |       |          |                   |       |          |                   |       |
| Intercepto                    | -1,280<br>(0,171) | 0,277 | Q22A1    | 1,110<br>(0,073)  | 3,025 | Q23P2    | 0,810<br>(0,178)  | 2,240 |
| Intercepto                    | -1,332<br>(0,059) | 0,264 | Q22A1    | 0,952<br>(0,024)  | 2,591 | Q23P2    | 0,459<br>(0,054)  | 1,583 |
| SEXO1                         | 0,298<br>(0,027)  | 1,348 | Q24A1    | -0,416<br>(0,034) | 0,660 | Q23P3    | 0,022<br>(0,215)  | 1,019 |
| PERIODO1                      | 0,454<br>(0,031)  | 1,574 | Q20P1    | 0,218<br>(0,029)  | 1,245 | INSE1    | 0,173<br>(0,046)  | 1,188 |
| Q10A1                         | 0,812<br>(0,036)  | 2,255 | Q20P2    | 0,157<br>(0,042)  | 1,171 | INSE2    | 0,305<br>(0,048)  | 1,356 |
| Q13A1                         | -0,393<br>(0,026) | 0,676 | Q23P1    | 0,301<br>(0,028)  | 1,352 | INSE3    | 0,625<br>(0,059)  | 1,869 |
| <b>Litoral — INSE = 5,075</b> |                   |       |          |                   |       |          |                   |       |
| Intercepto                    | -1,192<br>(0,184) | 0,305 | Q22A1    | 1,041<br>(0,091)  | 2,830 | Q23P2    | 0,606<br>(0,207)  | 1,820 |
| SEXO1                         | 0,360<br>(0,091)  | 1,432 | Q24A1    | -0,397<br>(0,108) | 0,672 | Q23P3    | -0,380<br>(1,049) | 0,674 |

| Variável                       | Est.              | Exp.  | Variável | Est.              | Exp.  | Variável | Est.              | Exp.  |
|--------------------------------|-------------------|-------|----------|-------------------|-------|----------|-------------------|-------|
| PERIODO1                       | 0,353<br>(0,097)  | 1,419 | Q20P1    | 0,258<br>(0,105)  | 1,293 | INSE1    | 0,315<br>(0,147)  | 1,364 |
| Q10A1                          | 0,740<br>(0,114)  | 2,096 | Q20P2    | 0,043<br>(0,145)  | 1,044 | INSE2    | 0,670<br>(0,164)  | 1,953 |
| Q13A1                          | -0,340<br>(0,090) | 0,714 | Q23P1    | 0,267<br>(0,092)  | 1,303 | INSE3    | 1,355<br>(0,397)  | 3,838 |
| <b>Nordeste — INSE = 5,262</b> |                   |       |          |                   |       |          |                   |       |
| Intercepto                     | -0,897<br>(0,158) | 0,408 | Q22A1    | 1,000<br>(0,069)  | 2,721 | Q23P2    | 0,432<br>(0,151)  | 1,539 |
| SEXO1                          | 0,176<br>(0,067)  | 1,192 | Q24A1    | -0,384<br>(0,084) | 0,679 | Q23P3    | 2,128<br>(1,136)  | 7,255 |
| PERIODO1                       | 0,339<br>(0,083)  | 1,403 | Q20P1    | 0,273<br>(0,077)  | 1,311 | INSE1    | -0,103<br>(0,128) | 0,902 |
| Q10A1                          | 0,974<br>(0,091)  | 2,644 | Q20P2    | 0,253<br>(0,113)  | 1,290 | INSE2    | 0,289<br>(0,126)  | 1,335 |
| Q13A1                          | -0,468<br>(0,067) | 0,626 | Q23P1    | 0,239<br>(0,070)  | 1,267 | INSE3    | 0,544<br>((0,134) | 1,719 |
| <b>Noroeste — INSE = 5,299</b> |                   |       |          |                   |       |          |                   |       |
| Intercepto                     | -1,112<br>(0,209) | 0,329 | Q22A1    | 1,173<br>(0,082)  | 3,229 | Q23P2    | 0,850<br>(0,179)  | 2,332 |
| SEXO1                          | 0,373<br>(0,082)  | 1,452 | Q24A1    | -0,048<br>(0,105) | 0,952 | Q23P3    | 0,117<br>(0,778)  | 1,067 |
| PERIODO1                       | 0,325<br>(0,095)  | 1,383 | Q20P1    | 0,220<br>(0,090)  | 1,243 | INSE1    | 0,097<br>(0,167)  | 1,102 |
| Q10A1                          | 0,959<br>(0,116)  | 2,611 | Q20P2    | 0,189<br>(0,126)  | 1,210 | INSE2    | 0,095<br>(0,154)  | 1,098 |
| Q13A1                          | -0,501<br>(0,081) | 0,606 | Q23P1    | 0,439<br>(0,085)  | 1,547 | INSE3    | 0,224<br>(0,170)  | 1,245 |
| <b>Norte — INSE = 5,259</b>    |                   |       |          |                   |       |          |                   |       |
| Intercepto                     | -0,845<br>(0,168) | 0,429 | Q22A1    | 1,012<br>(0,067)  | 2,749 | Q23P2    | 0,321<br>(0,141)  | 1,382 |
| SEXO1                          | 0,246<br>(0,069)  | 1,280 | Q24A1    | -0,298<br>(0,089) | 0,743 | Q23P3    | -0,923<br>(0,694) | 0,394 |
| PERIODO1                       | 0,234<br>(0,090)  | 1,263 | Q20P1    | 0,282<br>(0,074)  | 1,324 | INSE1    | 0,091<br>(0,124)  | 1,094 |
| Q10A1                          | 0,722<br>(0,089)  | 2,058 | Q20P2    | 0,243<br>(0,115)  | 1,273 | INSE2    | 0,139<br>(0,124)  | 1,149 |
| Q13A1                          | -0,497<br>(0,066) | 0,608 | Q23P1    | 0,418<br>(0,069)  | 1,521 | INSE3    | 0,406<br>(0,135)  | 1,500 |
| <b>Sudeste — INSE = 5,230</b>  |                   |       |          |                   |       |          |                   |       |

| Variável                                                 | Est.              | Exp.  | Variável | Est.              | Exp.  | Variável | Est.              | Exp.  |
|----------------------------------------------------------|-------------------|-------|----------|-------------------|-------|----------|-------------------|-------|
| Intercepto                                               | -1,095<br>(0,154) | 0,335 | Q22A1    | 1,066<br>(0,063)  | 2,905 | Q23P2    | 0,331<br>(0,146)  | 1,389 |
| SEXO1                                                    | 0,201<br>(0,067)  | 1,224 | Q24A1    | -0,242<br>(0,082) | 0,785 | Q23P3    | -0,134<br>(0,569) | 0,879 |
| PERÍODO1                                                 | 0,477<br>(0,075)  | 1,612 | Q20P1    | 0,333<br>(0,072)  | 1,395 | INSE1    | -0,175<br>(0,117) | 0,838 |
| Q10A1                                                    | 0,941<br>(0,097)  | 2,565 | Q20P2    | 0,197<br>(0,106)  | 1,216 | INSE2    | 0,043<br>(0,130)  | 1,040 |
| Q13A1                                                    | -0,370<br>(0,067) | 0,690 | Q23P1    | 0,218<br>(0,071)  | 1,244 | INSE3    | 0,113<br>(0,119)  | 1,120 |
| <b>Sudoeste – INSE = 5,278</b>                           |                   |       |          |                   |       |          |                   |       |
| Intercepto                                               | -1,280<br>(0,171) | 0,277 | Q22A1    | 1,110<br>(0,073)  | 3,025 | Q23P2    | 0,810<br>(0,178)  | 2,240 |
| SEXO1                                                    | 0,198<br>(0,075)  | 1,219 | Q24A1    | -0,208<br>(0,091) | 0,809 | Q23P3    | -0,465<br>(0,859) | 0,622 |
| PERÍODO1                                                 | 0,437<br>(0,085)  | 1,547 | Q20P1    | 0,286<br>(0,078)  | 1,330 | INSE1    | 0,049<br>(0,130)  | 1,051 |
| Q10A1                                                    | 0,863<br>(0,105)  | 2,368 | Q20P2    | 0,179<br>(0,121)  | 1,197 | INSE2    | 0,160<br>(0,128)  | 1,173 |
| Q13A1                                                    | -0,341<br>(0,074) | 0,710 | Q23P1    | 0,375<br>(0,073)  | 1,455 | INSE3    | 0,293<br>(0,136)  | 1,339 |
| FONTE: Elaboração própria a partir dos valores estimados |                   |       |          |                   |       |          |                   |       |

De acordo com os valores mostrados na Tabela 1, é possível verificar que, em todas as regiões, alunos do sexo masculino tem maior probabilidade de se classificar em suficiente na proficiência em Matemática quando comparados com os alunos do sexo feminino, pois a estimativa da variável SEXO1 foi maior que 1. Isso vale para os alunos matriculados no período matutino, quando comparados com aqueles que estudam no período vespertino ou noturno, o que pode ser comprovado pelas estimativas maiores que 1 para a variável PERÍODO1. A probabilidade em questão também é maior para os alunos que nunca foram reprovados e que gostam de estudar matemática, quando comparados com aqueles que já reprovaram uma ou mais vezes e que não gostam de estudar matemática. Em todas as regiões, para os alunos que nunca reprovaram, a probabilidade de se classificar em Suficiente na proficiência em Matemática é duas vezes maior que a dos alunos que reprovaram pelo menos uma vez na disciplina, de acordo com os valores exponenciais (Exp.) obtidos.

A Tabela 1 mostra que, para aqueles que gostam de estudar Matemática, a chance de obter tal classificação é quase três vezes maior, quando comparada à probabilidade dos demais alunos. Tal afirmação deve-se ao fato de as estimativas para as variáveis Q10A1

e Q22A1 terem sido próximas de dois e três, respectivamente. Por outro lado, quando comparado com o aluno que não recebe ajuda dos pais na lição de casa, aquele que recebe tem probabilidade menor de conseguir a classificação Suficiente, haja vista o valor da estimativa menor que um para a variável Q10A1. Uma explicação para tal resultado pode ser o fato de o aluno que precisa de ajuda dos pais apresentar maior dificuldade, comparado àquele que é capaz de lidar sozinho com as lições escolares. O mesmo ocorreu para os alunos que gostam de estudar Língua Portuguesa. Quando a preferência é por esta disciplina, conseqüentemente, não gostam de estudar Matemática.

Considerando a escolaridade materna e a renda familiar, pode-se afirmar que, para os alunos matriculados nas escolas situadas nas regiões do estado de São Paulo, a probabilidade de se classificar em Suficiente na proficiência em Matemática, segundo dados do SARESP de 2019, é maior para aqueles cujas mães têm o EM e o ES completos, comparados aos alunos filhos de mães com que não concluíram nem o EF. Tal probabilidade também aumenta à medida em que a renda familiar aumenta, porém até o limite de 15 SM.

De acordo com a Tabela 1, as estimativas obtidas para as variáveis Q23P1, Q23P2 e Q23P3, sugerem que a probabilidade de alcançar a classificação Suficiente na proficiência em Matemática do aluno com renda familiar entre 2 SM e 5 SM é de, aproximadamente, 1,5 vezes a probabilidade daquele cuja renda familiar é inferior a 2 SM. Para os alunos com renda familiar entre 5 SM e 15 SM, tal probabilidade é 1,6 vezes a probabilidade daquele com a renda mais baixa. A renda familiar superior a 15 SM não se mostrou significativa para explicar a referida probabilidade. Em relação à variável da escola, INSE, nota-se que a probabilidade aumenta à medida em que o nível socioeconômico da escola aumenta, isto é, escolas com melhores condições e recursos financeiros contribuem positivamente para o desempenho dos alunos.

## Conclusão

Embora esteja em vias de desenvolvimento no Brasil, a avaliação educacional em larga escala tem se mostrado importante nas últimas décadas, constituindo-se como ferramenta para auxiliar a identificação de fatores atuantes sobre o desempenho escolar do aluno. Nos últimos 25 anos, aproximadamente, os sistemas de avaliação tiveram, como objetivo prioritário, encontrar mecanismos que oferecessem melhorias, de maneira eficaz e eficiente, na qualidade do ensino ofertado na sociedade. Por sua vez, a avaliação educacional tem a finalidade de fornecer elementos que permitam diagnosticar a situação do sistema educacional de determinado local e, concomitantemente, subsidiar políticas e diretrizes adequadas no contexto municipal, estadual e nacional, visando a contínua pela melhoria na qualidade da educação.

Nesta investigação, foram considerados dados do SARESP, referentes à edição de 2019,

especificamente as notas obtidas em Matemática pelos alunos da 3<sup>a</sup> série do Ensino Médio pertencentes à rede pública de ensino. As respostas dos alunos e pais dos alunos participantes, dadas aos questionários aplicados pelo SARESP constituíram dados significativos para a análise proposta. Nos modelos hierárquicos generalizados, a variável resposta caracterizou-se como categórica e representou a classificação do aluno em Insuficiente ou Suficiente, de acordo com o nível de proficiência em Matemática alcançado.

Os resultados obtidos nesta pesquisa corroboraram com as conclusões de estudos realizados na literatura, os quais enfatizam que os estudos que consideram modelos hierárquicos para dados educacionais, realizados no Brasil, mostram que existe uma significativa variação entre as escolas, mas que a variabilidade observada entre os alunos é sempre maior. Os valores estimados usando o modelo linear generalizado hierárquico proposto neste estudo permitiram observar que as variáveis associadas à reprovação apresentaram efeito negativo sobre a proficiência em Matemática enquanto a escolaridade da mãe e a condição socioeconômica contribuíram positivamente para um resultado satisfatório.

As efetivas contribuições de estudos como este apresentado ao longo do artigo buscam obter conhecimento para subsidiar soluções para problemas que são constantes preocupações de todos aqueles que estão envolvidos com o sistema educacional. A compreensão dos fatores extra e intra-escolares que atuam sobre a probabilidade de o aluno alcançar um resultado adequado no SARESP, influenciando diretamente o nível de classificação em que ele se encontra, é importante para subsidiar os agentes da Secretaria Educacional da Educação nas definições voltadas às políticas educacionais direcionadas às melhorias na Educação Básica brasileira.

Assim, a partir dos resultados alcançados, espera-se ter contribuído com a literatura que aborda a identificação de fatores que impactam o desempenho educacional do aluno e ressaltar a importância de formulação e implementação de políticas públicas voltadas à educação que visem, simultaneamente, melhorar a qualidade do ensino ofertado à sociedade e diminuir o impacto das características socioeconômicas sobre o desempenho do aluno.

## Referências

ALVES, M. T. G.; SOARES, J. F. O efeito das escolas no aprendizado dos alunos: um estudo com dados longitudinais no ensino fundamental. *Educação e Pesquisa*, São Paulo, v. 34, n. 3, p. 527–544, 2008.

ANDRADE, J. M. e; LAROS, J. A. Fatores associados ao desempenho escolar: estudo multinível com dados do SAEB/2001. *Psicologia: teoria e pesquisa*, Brasília, v. 23, p. 33–41, 2007.

BARBOSA, M. E. F.; FERNANDES, C. A escola brasileira faz diferença? Uma investigação dos efeitos da escola na proficiência em matemática dos alunos da 4<sup>a</sup> série. In:

FRANCO, C. (Org.). *Promoção, ciclos e avaliação educacional*. Porto Alegre: ArtMed, 2001. p. 155–172.

BORGES, A. P. N. R.; CASTRO E ALMEIDA, S. P. N.; LIMA, K. Avaliação Educacional: o Saeb, seus pressupostos, finalidades e repercussões. *Revista Internacional de Pesquisa em Educação Matemática*, [s.l.], v. 12, n. 4, p. 1–19, 2022.

BRASIL. Lei n.º 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. *Diário Oficial da União: seção 1*, Brasília, DF, 23 dez. 1996.

BROOKE, N. et al. Modelagem do crescimento da aprendizagem nos anos iniciais com dados longitudinais da pesquisa GERES. *Educação e Pesquisa*, São Paulo, v. 40, n. 1, p. 77–94, 2014.

FERRÃO, M. E. et al. O SAEB – Sistema Nacional de Avaliação da Educação Básica: objetivos, características e contribuições na investigação da escola eficaz. *Revista Brasileira de Estudos de População*, São Paulo, v. 18, n. 102, p. 111–130, 2001.

FLETCHER, P. À procura do ensino eficaz. Rio de Janeiro: Ministério da Educação e Cultura, Departamento da Avaliação da Educação Básica, 1998. (mimeo).

GOLDSTEIN, H. Modelos de realidade: novas abordagens para a compreensão de processos educacionais. In: FRANCO, C. (Org.). *Promoção, ciclos e avaliação educacional*. Porto Alegre: ArtMed, 2001. p. 84–99.

IZBICKI, R.; DOS SANTOS, T. M. *Machine Learning sob a ótica estatística*. São Carlos: UFSCar/Insper, 2018.

JESUS, G. R. de; LAROS, J. A. Eficácia escolar: regressão multinível com dados de avaliação em larga escala. *Avaliação Psicológica*, São Paulo, v. 3, n. 2, p. 93–106, 2004.

LAROS, J. A.; MARCIANO, J. L.; ANDRADE, J. M. de. Fatores que afetam o desempenho na prova de matemática do SAEB: um estudo multinível. *Avaliação Psicológica*, São Paulo, v. 9, n. 2, p. 173–186, 2010.

MCCULLAGH, P. *Generalized linear models*. Nova Iorque: Routledge, 2019.

MOREIRA, K. da S. G.; ANDRADE, P. de A.; BEGOLIN, I. P. Determinantes da proficiência em matemática no Rio Grande do Sul: uma análise a partir de modelos hierárquicos. *Ensaio FEE*, Porto Alegre, v. 38, n. 1, p. 7–34, 2017.

PALERMO, G. A.; SILVA, D. B. do N.; NOVELLINO, M. S. F. Fatores associados ao desempenho escolar: uma análise da proficiência em matemática dos alunos do 5º ano do ensino fundamental da rede municipal do Rio de Janeiro. *Revista Brasileira de Estudos de População*, São Paulo, v. 31, n. 2, p. 367–394, 2014.

RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical linear models: applications and data analysis methods*. 2. ed. Thousand Oaks, CA: Sage Publications, 2002.

RIANI, J. de L. R.; RIOS-NETO, E. L. G. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros? *Revista Brasileira de Estudos de População*, São Paulo, v. 25, n. 2, p. 251–269, 2008.

RSTUDIO. Disponível em: <https://cran.rstudio.com/>. Citation(“pacman”,

“dplyr”, “memisc”, “mlmRev”, “lme4”, “rstanarm”, “bayesplot”, “readxl”, “glmnet”, “brms”).

SÃO PAULO (Estado). SECRETARIA DA EDUCAÇÃO. *Sumário Executivo SA-RESP 2019*. São Paulo: Secretaria da Educação do Estado de São Paulo, 2019.

SOARES, J. F.; CÉSAR, C. C.; MAMBRINI, J. Determinantes de desempenho dos alunos do ensino básico brasileiro: evidências do SAEB de 1997. In: FRANCO, C. (Org.). *Promoção, ciclos e avaliação educacional*. Porto Alegre: ArtMed, 2001. p. 121–153.

SOARES, T. M. Modelo de três níveis hierárquicos para a proficiência dos alunos de 4<sup>a</sup> série de língua portuguesa do SIMAVE/PROEB-2002. *Revista Brasileira de Educação*, Rio de Janeiro, n. 29, p. 73–87, 2005.

SOARES, T. M.; MENDONÇA, M. C. M. Construção de um modelo de regressão hierárquico para os dados do SIMAVE-2000. *Pesquisa Operacional*, São Paulo, v. 23, n. 3, p. 421–441, 2003.

## Anexo

```
pacman::p_load(dplyr, memisc, mlmRev, lme4, rstanarm, bayesplot, readxl, glmnet, brms)
```

```
# Split the data into training and test set
set.seed(162)
training.samples <- dados$CLASSIFIC %>%
createDataPartition(p = 0.7, list = FALSE)
train.data <- dados[training.samples, ]
test.data <- dados[-training.samples, ]
x <- model.matrix(CLASSIFIC~., train.data)[,-1]
y <- train.data$CLASSIFIC
glmnet(x, y, family = "binomial", alpha = 1, lambda = NULL)
cv.lasso <- cv.glmnet(x, y, alpha = 1, family = "binomial")
plot(cv.lasso)
coef(cv.lasso, cv.lasso$lambda.min)
coef(cv.lasso, cv.lasso$lambda.1se)
model <- glmnet(x, y, alpha = 1, family = "binomial", lambda = cv.lasso$lambda.min)
x.test <- model.matrix(diabetes ~., test.data)[,-1]
probabilities <- model %>% predict(newx = x.test)
predicted.classes <- ifelse(probabilities > 0.5, "pos", "neg")
# Lasso regression
set.seed(123)
cv.lasso <- cv.glmnet(x, y, alpha = 1, family = "binomial")
coef(cv.lasso, cv.lasso$lambda.min)
coef(cv.lasso, cv.lasso$lambda.1se)
# Variables coding
options(mc.cores = parallel::detectCores())
Banco_2019_Centro <- read_excel("Banco_2019_Centro.xlsx")
Banco_2019_Centro$classific_mat <- recode(Banco_2019_Centro$classific_mat,
"0"<- "Insuficiente", "1"<- "Suficiente")
Banco_2019_Centro$SEXO <- recode(Banco_2019_Centro$SEXO, "1"<- "M", "2"<-
"F")
Banco_2019_Centro$SEXO <- factor(Banco_2019_Centro$SEXO)
Banco_2019_Centro$PERIODO <- recode(Banco_2019_Centro$PERIODO,
"1"<- "MANHÃ", "2"<- "TARDE", "3"<- "NOITE")
Banco_2019_Centro$inicio_estudos <- recode(Banco_2019_Centro$inicio_estudos,
"0"<- "A", "1"<- "B")
```

```
Banco_2019_Centro$inicio_estudos <- factor(Banco_2019_Centro$inicio_estudos)
Banco_2019_Centro$ja_reprovou <- recode(Banco_2019_Centro$ja_reprovou,

  "0"<- "A", "1"<- c("B","C"))

Banco_2019_Centro$ja_reprovou <- factor(Banco_2019_Centro$ja_reprovou)
Banco_2019_Centro$pais_incentivam <- recode(Banco_2019_Centro$pais_incentivam,
  "1"<- c("A","B"), "0"<- "C")
Banco_2019_Centro$pais_incentivam <- factor(Banco_2019_Centro$pais_incentivam)
Banco_2019_Centro$gosta_matematica <- recode(Banco_2019_Centro$gosta_matematica,
  "1"<- "A", "0"<- "B")
Banco_2019_Centro$gosta_matematica <- factor(Banco_2019_Centro$gosta_matematica)
Banco_2019_Centro$faz_licao <- recode(Banco_2019_Centro$faz_licao,
  "1"<- c("A","B"), "0"<- c("C","D"))
Banco_2019_Centro$faz_licao <- factor(Banco_2019_Centro$faz_licao)
Banco_2019_Centro$esc_pai <- recode(Banco_2019_Centro$esc_pai, "0"<-
c("A","B"),
  "1"<- "C", "2"<- "D", "3"<- "E", "4"<- "F")
Banco_2019_Centro$esc_pai <- factor(Banco_2019_Centro$esc_pai)
Banco_2019_Centro$esc_mae <- recode(Banco_2019_Centro$esc_mae, "0"<-
c("A","B"),
  "1"<- "C", "2"<- "D", "3"<- "E", "4"<- "F")
Banco_2019_Centro$trabalho_mae <- recode(Banco_2019_Centro$trabalho_mae,
  "1"<- c("A","B","C"), "2"<- "D", "3"<- "E")
Banco_2019_Centro$trabalho_mae <- factor(Banco_2019_Centro$trabalho_mae)
Banco_2019_Centro$renda_familiar <- recode(Banco_2019_Centro$renda_familiar,
  "1"<- c("A","B"), "2"<- c("C","D"), "3"<- "E", "4"<- c("F","G"))
Banco_2019_Centro$renda_familiar <- factor(Banco_2019_Centro$renda_familiar)
# Model
prior2 <- c( prior(normal(0,10), class = Intercept), prior(normal(0,10), class = b),
prior(cauchy(0,10), class = sd))
MLH1_B_Centro <- brm( classific_mat| trials(nivel_profic_mat) ~ SEXO + PE-
RIODO + ja_reprovou + pais_incentivam + gosta_matematica + faz_licao + esc_pai
+ esc_mae + renda_familiar + INSE + (1|CD_UE), family = binomial(link = "logit"),
data = Banco_2019_Centro, prior = prior2, warmup = 2000, iter = 4000, chains = 3,
control = list(adapt_delta = 0.95))
summary(MLH1_B_Centro, waic = TRUE)
exp(fixef(MLH1_B_Centro))
# MCMC diagnostic: Rhat: potential scale reduction statistic
rhats <- rhat(M3_Centro_HB)
```

```
structure(rhats[1:22], my_attribute = "This is a vector")  
color_scheme_set("brightblue")  
mcmc_rhat(rhats)
```