

A relação entre nível socioeconômico e proficiência em matemática de estudantes pernambucanos do 9º ano através da Mineração de Dados Educacionais

The relationship between socioeconomic level and mathematics proficiency of 9th grade students from Pernambuco through Educational Data Mining

Jairo José Ribeiro Toscano de Brito Júnior¹
Gabriel Cândido da Silva²
Rodrigo Lins Rodrigues³
Americo Nobre Gonçalves Ferreira Amorim⁴

Resumo

O desempenho escolar dos estudantes da educação básica, mensurado pelo Ideb a partir dos índices de aprovação das instituições escolares e de proficiência no Saeb (Prova Brasil), possui relação com aspectos sociais, como: renda familiar, trabalho doméstico, nível de escolaridade dos pais, dentre outros. Nesse sentido, esse trabalho tem como propósito demonstrar o quanto os aspectos relativos aos Índices de Nível Socioeconômico (Inse) afetam a proficiência em matemática dos estudantes dos Anos Finais das escolas públicas do Estado de Pernambuco. Para tal, aplicou-se técnicas de mineração de dados sobre os microdados coletados no banco de dados do INEP. Os resultados demonstram uma diferença estatisticamente significativa entre proficiências de matemática dos estudantes de Inses diferentes e destaca os tópicos socioeconômicos mais relevantes relacionados ao desempenho no Saeb.

Palavras chave: Mineração de dados; dados abertos; proficiência em matemática.

Abstract

The Basic education student's performance, measured by Ideb based on the approval rates of school institutions and proficiency in the Saeb (Brazilian National Exams), is related to social aspects, such as: family income, domestic work, parents' education level, among others. In this sense, this work is focused on demonstrating how aspects related to the Socioeconomic Level Index (Inse) affects the Final Years of public schools student's proficiency in mathematics in the State of Pernambuco. To this end, educational data mining techniques were applied to the microdata collected on the INEP's database. The results

¹ Universidade Federal Rural de Pernambuco | jairo.toscano@ufrpe.br

² Universidade Federal Rural de Pernambuco | gabcandidods@gmail.com

³ Universidade Federal Rural de Pernambuco | rodrigo.linsrodrigues@ufrpe.br

⁴ Johns Hopkins University | ang6011@nyu.edu

demonstrate a statistically significant difference between the student's math proficiencies belonging to different Inses and highlight the most relevant socioeconomic topics related to performance on the Saeb.

Keywords: Data mining; open data; math proficiency.

Introdução

O desempenho escolar dos estudantes da educação básica, mensurado pelo Ideb a partir dos índices de aprovação das instituições escolares e de proficiência no Saeb (Prova Brasil), possui relação com aspectos sociais, como: renda familiar, trabalho doméstico, nível de escolaridade dos pais, dentre outros. De acordo com as pesquisas, tais fatores externos influenciam diretamente no rendimento escolar do estudante (Ferreira et al., 2016; Proença et al., 2020; Gomes et al., 2021), e analisá-los são essenciais para que se possa mensurar as competências e habilidades a serem desenvolvidas pelos estudantes. (Alves et al., 2007; Luna e Carvalho, 2019; Paula, 2021).

Araújo et al. 2020 destaca que, "(...) melhoria na condição socioeconômica e na renda da população afeta positivamente as notas dos alunos, corroborando a importância das políticas públicas de redistribuição de renda que diminuem as desigualdades existentes" (ARAÚJO et al., 2020, p. 3.521). Em concordância, Silva et al. 2019 investigou a aplicabilidade de modelos de regressão para a previsão de desempenho escolar dos estudantes de escolas públicas de Pernambuco usando dados do Saeb dos anos de 2015 e 2017, e seus resultados mostram que fatores como a quantidade de pessoas que moram numa residência, o incentivo (ou não) dos pais aos estudos e a região onde o aluno reside influenciam no desempenho escolar.

Em 2021, o estado de Pernambuco apresentou o terceiro melhor IDEB do país, ficando atrás apenas do Paraná e Goiás, essa informação foi disponibilizada pelo INEP em 2022. Diante disso, estabeleceu-se como objetivo deste estudo, investigar a relação entre o índice de nível socioeconômico (Inse) e a proficiência em matemática dos estudantes do 9º ano de escolas públicas do estado de Pernambuco, fornecidos pelos microdados do Saeb das edições de 2011 a 2019. A fim de alcançar tal objetivo foram elaboradas as seguintes questões de pesquisa: Existe diferença estatisticamente significativa entre as médias de proficiência em matemática dos estudantes com níveis socioeconômicos diferentes? Quais fatores socioeconômicos são mais determinantes para a diferença no nível de proficiência?

Por fim, este trabalho está estruturado da seguinte forma: Introdução, Metodologia, Resultados e discussões, e Considerações finais.

Metodologia

Contexto da pesquisa

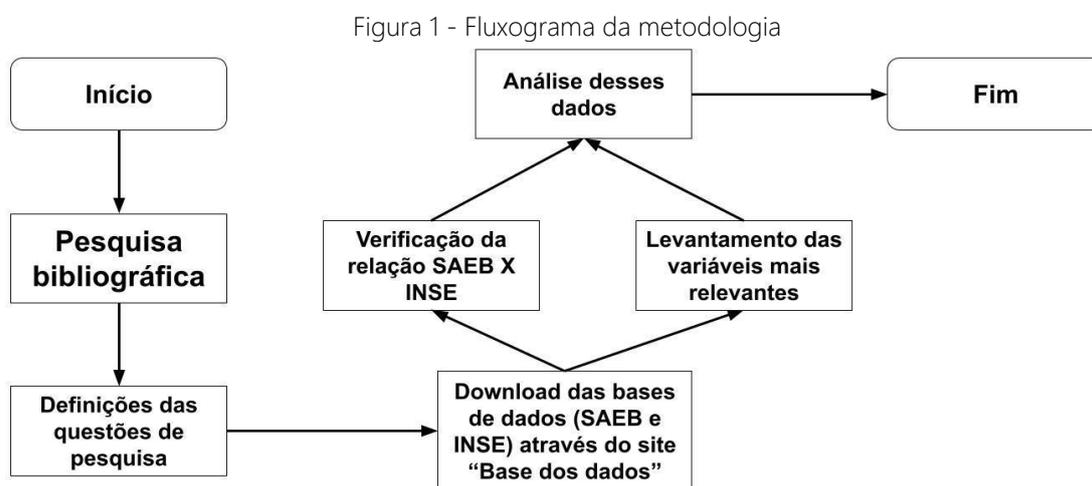
O estado de Pernambuco em 2020, de acordo com a sinopse do Censo Escolar do site do IBGE (Acessado pelo link: <https://cidades.ibge.gov.br/brasil/pe/recife/pesquisa/13/5908>, em 07/12/2021), possui 13.962 escolas distribuídas em 185 municípios. Além disso, dessas escolas, 6.105 são de educação infantil, 6.742 de ensino fundamental e 1.115 de ensino médio. E os alunos dessas escolas têm o seu desempenho avaliado a partir da mensuração

de suas proficiências nas disciplinas de matemática e língua portuguesa, que ocorre em avaliações externas, como Saeb (Américo e Lacruz, 2017; Araújo et al., 2020).

Neste trabalho foi utilizada a base de dados aberta do Saeb (2011-2019) referente aos alunos do 9º ano do ensino fundamental, pois nessa base temos informações relacionadas aos fatores socioeconômicos desses alunos, bem como seus respectivos desempenhos em matemática. Os dados referentes ao Saeb 2011-2019 foram retirados do INEP através do site “Base dos dados” (Acessado pelo link: <https://basedosdados.org/>, em 22/10/2021). Esses dados estavam dentro de um arquivo nomeado “TS_ALUNO_9EF” que tinha um total de 2.388.931 linhas e 130 colunas contendo informações sobre os estudantes, suas proficiências em matemática e suas respostas ao questionário socioeconômico. Além dessa base, utilizou-se os dados do Inse dos anos de 2011-2019. A partir desses dados abertos, além da verificação da diferença estatisticamente significativa entre as médias de matemática do Saeb dos estudantes de diferentes níveis socioeconômicos, buscaremos compreender o quanto os fatores socioeconômicos, levantados por este questionário de 49 questões, estão relacionados ao desempenho dos estudantes na prova de matemática, entendendo também quais destas questões levantadas demonstram maior relação com essa nota.

Análise dos dados

O método de análise aplicado neste estudo ocorreu dentro de duas etapas principais, que podem ser descritas como: Verificação da existência de diferença entre a proficiência de matemática de estudantes de diferentes Inse; Levantamento de aspectos sociais mais relevantes segundo os microdados do Saeb. Abaixo segue o fluxograma completo da metodologia utilizada no artigo contendo todas as etapas referentes ao desenvolvimento do artigo:



Fonte: Dados da pesquisa

Como demonstrado no fluxograma, deu-se início ao nosso processo de estudo realizando pesquisas bibliográficas relacionadas a estudos que buscaram e realizaram análises a partir dos microdados do Saeb, para identificar de que maneira este trabalho poderia contribuir para a literatura. Feito isso, definiu-se as questões de pesquisa que norteariam as nossas análises em prol de solucionar tais questões.

Com o delineamento dos objetivos da pesquisa bem definidos, prosseguiu-se para o download das bases a serem utilizadas no processo de análise, do qual dividiu-se em dois setores: 1) Verificação da existência de diferença estatisticamente significativa entre a proficiência em matemática de estudante com níveis socioeconômicos (Inse) diferentes ao longo dos anos; 2) Utilização da base de dados "TS_ALUNO_9EF" para levantamento das variáveis mais relevantes. Por fim, realizamos o cruzamento final de informações e a análise dos resultados obtidos para gerar as discussões pertinentes ao tema.

Verificação da existência de diferença entre a proficiência de matemática de estudantes de diferentes INSE

Nesta etapa pretendeu-se verificar, da perspectiva da estatística, a existência de diferença entre a proficiência em matemática dos estudantes pernambucanos de escola pública de cada nível socioeconômico. Para tal, utilizou-se recursos da estatística descritiva e inferencial a partir do software RStudio, versão 4.1.0, e das funcionalidades dos pacotes tidyverse, ggpubr e rstatix. Além disso, usou-se o teste de Kruskal-Wallis para verificar se existe diferença significativa entre os dados observados, pois este "é o teste apropriado para comparar as duas ou mais distribuições de uma variável, pelo menos ordinal, observada em duas ou mais amostras independentes" (MAROCO, 2018, p. 316).

Este processo ocorreu utilizando as bases de dados referente ao Saeb aplicado de 2011 a 2019 e ao Inse estabelecido nos anos de 2011, 2015 e 2019. Embora já exista a informação acerca do nível socioeconômico do aluno nos microdados do Saeb, optou-se por recorrer diretamente aos dados do Inse por dois motivos: Os dados do Inse oferecem uma classificação de nível socioeconômico justificada e fundamentada em nota técnica oficial publicada; Os microdados do Saeb referente aos anos de 2011 e 2013 não apresentavam a classificação dos níveis socioeconômicos dos estudantes da mesma forma que o Inse. Dessa forma, as variáveis usadas nessa etapa do processo foram: código de identificação da escola (id_escola); a média da nota de matemática do Saeb (nota_saeb_matematica); rede (rede); e o indicador de nível socioeconômico (inse).

A verificação da existência de diferença entre a proficiência de matemática de estudantes de diferentes Inses deu-se início a partir da seleção das variáveis das duas bases, filtrando, nos microdados do Saeb, por escolas públicas estaduais e municipais (rede = 2 e rede = 3, respectivamente) de Pernambuco, unindo aos dados referentes aos níveis socioeconômicos do Inse e retirando os valores NA (não avaliáveis). Em seguida utilizou-se o teste de Kolmogorov-Smirnov para identificar se os dados seguiam uma distribuição normal, referentes às médias de matemática do Saeb, pois essa é uma condição necessária para a utilização do teste de Kruskal-Wallis. Por fim, aplicou-se o teste de Kruskal-Wallis com post hoc do teste de Dunn. O nível de significância adotado foi de 95%.

Levantamento de aspectos sociais relevantes segundo os microdados do Saeb

Esta etapa avaliou quais características socioeconômicas são capazes de explicar a proficiência em matemática obtida pelos estudantes na prova do Saeb. Para a execução desta etapa foi realizado o método Stepwise Logistic Regression, que consiste de uma seleção automatizada do conjunto de variáveis independentes que formam a melhor

construção possível do modelo de predição para uma variável dependente (Kassambara, 2017).

Para a construção do conjunto de dados, foram utilizados os microdados do Sistema de Avaliação da Educação Básica (Saeb) do ano de 2019, considerando apenas a tabela chamada "TS_ALUNO_9EF", que corresponde ao levantamento das informações referentes aos alunos que cursam o 9º ano do Ensino Fundamental. No processo de construção e tratamento dos dados, foram considerados apenas os estudantes pertencentes às escolas estaduais e municipais do estado de Pernambuco, e que não possuíam valores NA nos dados correspondentes à Proficiência em Matemática e no Questionário Socioeconômico. Dito isto, a amostra final contou com 13.411 estudantes da rede estadual e 25.547 estudantes da rede municipal.

A execução do algoritmo foi realizada duas vezes, uma vez para a rede estadual e outra para a municipal. Neste processo, um total de 49 perguntas que compõe o questionário socioeconômico foram utilizadas como as variáveis independentes, já a variável dependente foi construída a partir de um processo de tornar binário os valores de proficiência em matemática, sendo esta realizada a partir da média geral do estado de Pernambuco, considerando uma média diferente para redes estaduais e municipais, ou seja, a variável alvo se trata de uma variável binária que indica se os estudantes estão acima ou abaixo da média geral da rede de ensino da qual pertence.

Ao fim desta etapa, utilizamos o algoritmo Odds Ratio, também chamado de Razão de Possibilidades, que nos indica um valor estatístico que considera a probabilidade de que o evento ocorra dividido pela propriedade de que não ocorra (Guedes, Lemos e Lacruz, 2021). Com isto, realizamos um levantamento que indica quais são as 5 características socioeconômicas mais determinantes para classificar o estudante como acima ou abaixo da média para as redes estaduais e municipais.

Resultados e Discussões

Durante a etapa de tratamento dos dados foi observado a ausência de alguns dados das bases utilizadas.

Quadro 1 - Quantidade de escolas contempladas nesta etapa

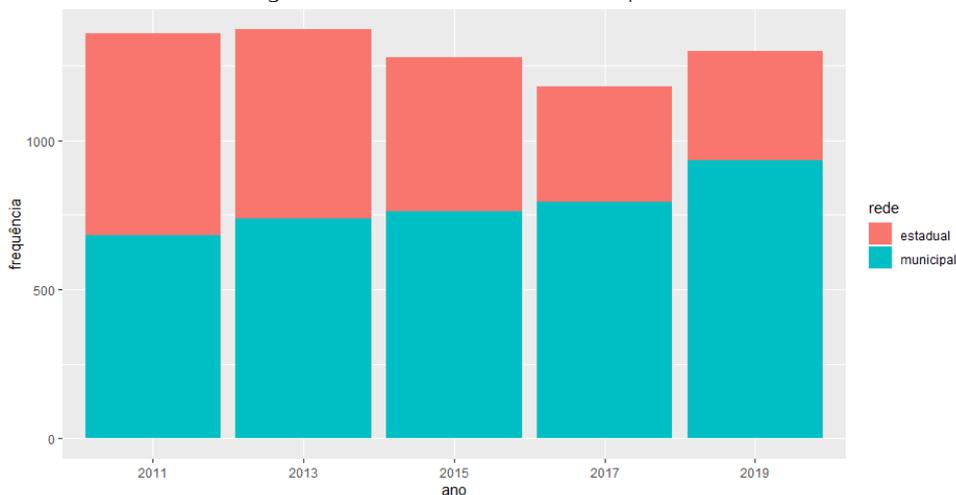
	2011	2013	2015	2017	2019
Quantidade de escolas na base do Saeb	1414	2538	2466	3155	3138
Quantidade de escolas com a MEDIA_9EF_MAT e INSE - CLASSIFICAÇÃO	1360	1371	1279	1180	1300
Porcentagem de escolas contempladas neste estudo	96.18	54.02	51.87	37.40	41.43

Fonte: Dados da pesquisa

Percebe-se, no Quadro 1, que apenas no ano de 2011 havia um retrato social significativo, 96,18% das escolas participaram e realizaram o preenchimento dos questionários do Saeb para o 9º Ano. É possível observar, que nos demais anos, a quantidade de escolas que tinham a sua nota de matemática do SAEB e o seu INSE registrados era menor que 60%. Esse quantitativo aparenta ser insuficiente para se alcançar um dos objetivos propostos por essas bases, que seria a produção de indicadores

representativos para auxiliar na implementação, monitoramento e avaliação de políticas públicas (Brasil, 2019).

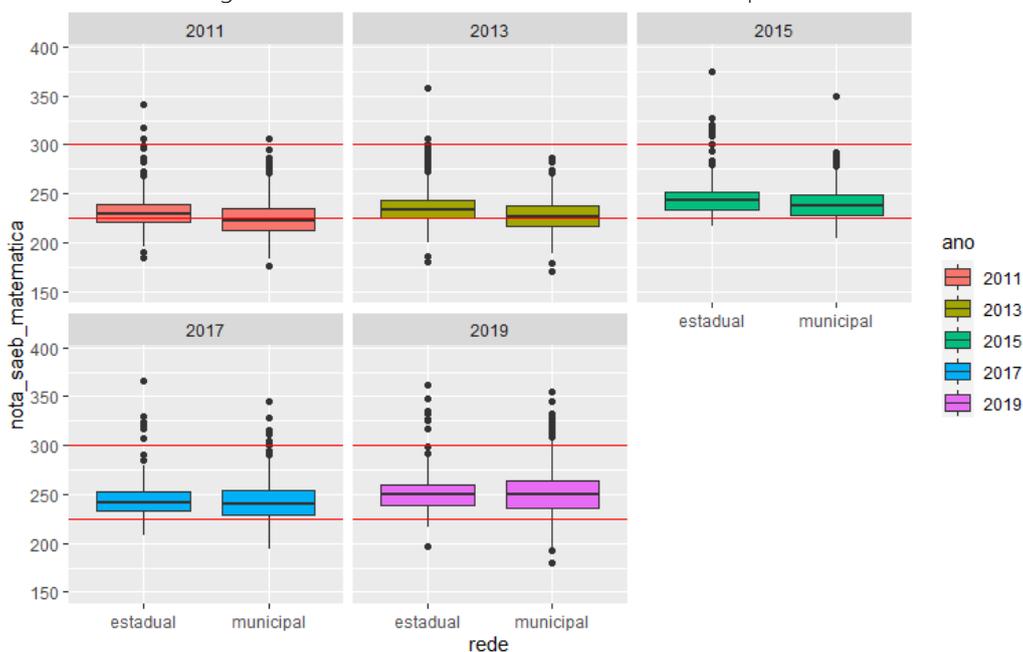
Figura 2 - Quantidade de escolas por rede



Fonte: Dados da pesquisa

A partir da Figura 2, pode-se visualizar a distribuição das escolas abordadas por sua dependência administrativa. Observa-se que, no ano de 2011, os dados representavam escolas estaduais e municipais em proporções semelhantes, porém, nota-se que, com o passar dos anos, essa proporção tornou a mudar com a maior parte da representação sendo definida pelas escolas municipais.

Figura 3 - Proficiência em matemática das redes por ano

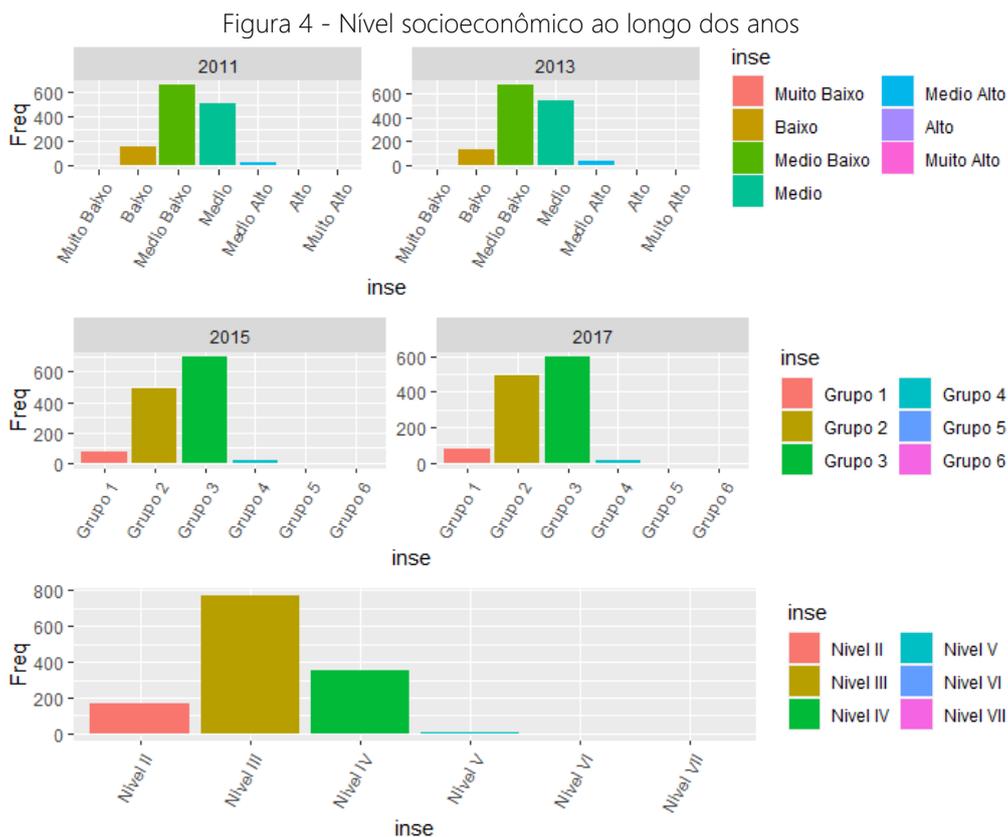


Fonte: Dados da pesquisa

Corroborando com o problema abordado por este estudo, visualizou-se na Figura 3 que, mesmo com o crescimento da quantidade de estudantes presente no nível básico de proficiência - cuja nota de matemática varia entre 225 e 300 (entre as linhas vermelhas),

cerca de 7,54% das escolas apresentaram proficiência no nível insuficiente. Pode-se destacar, a partir da Figura 3, que em 2011 as escolas municipais apresentaram uma mediana de 222.71 de proficiência (IQR: 212.16, 235.21), ou seja, em torno de 50% das escolas municipais apresentaram proficiência insuficiente segundo a escala de proficiência do Saeb.

É observável na Figura 3 que a mediana de proficiência das escolas estaduais (2011 - 229.05 (IQR: 220.54, 239.41), 2013 - 233.57 (IQR: 224.73, 243.47), 2015 - 242.70 (IQR: 233.84, 251.85), 2017 - 241.73 (IQR: 232.97, 252.41), 2019 - 250.18 (IQR: 239.26, 259.20)) aparecem maiores que as das municipais (2011 - 222.71 (IQR: 212.16, 235.21), 2013 - 227.00 (IQR: 216.77, 237.87), 2015 - 237.91 (IQR: 228.41, 248.18), 2017 - 240.26 (IQR: 229.37, 253.44), 2019 - 249.85 (IQR: 235.56, 264.03)) ao longo dos anos. As médias das escolas estaduais apresentaram, de 2011 a 2019, os seguintes desvios padrões (16.03, 16.39, 16.05, 18.41, 19.29), enquanto que as escolas municipais apresentaram, no mesmo período, os seguintes desvios padrões em suas médias (17.88, 16.83, 16.09, 21.83, 22.81). Note que o desvio padrão das médias é pequeno, por isso é possível observar a presença de alguns valores discrepantes (outliers). Por fim, percebe-se um discreto aumento da proficiência ao longo dos anos em ambas as redes.



Fonte: Dados da pesquisa

A Figura 4 apresenta um panorama do nível socioeconômico das escolas ao longo dos anos a partir da relação da variável frequência (Freq) e índice de nível socioeconômico (inse). No ano de 2019 não constava nenhum dado referente ao Nível I do Inse, por isso o mesmo não aparece no gráfico. Pode-se destacar dois pontos a partir do gráfico acima: O terceiro grupo socioeconômico predomina como o nível que representa grande parte das

escolas; E a frequência do quarto grupo apresenta uma queda significativa após 2013, assim como o grupo do “Nível II” em 2019.

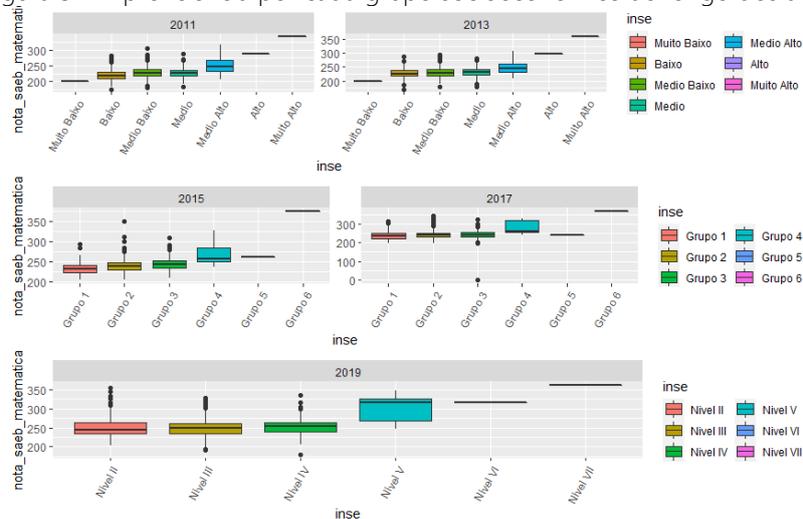
Embora existam algumas mudanças na caracterização do nível socioeconômico entre as edições do ano de 2013, 2015 e 2019 do Inse, percebe-se que o terceiro grupo socioeconômico permanece como o predominante ao longo dos anos. Os aspectos do terceiro grupo em comum entre as edições do Inse, segundo as notas técnicas oficiais do Inse, são: os responsáveis terem o ensino médio completo (a edição de 2013 considera a conclusão do ensino fundamental); terem em casa uma geladeira, um banheiro, uma televisão e uma máquina de lavar. O quarto grupo das edições de 2011 e 2013 apresentam uma queda brusca a posteriori pois as edições seguintes passaram a adotar no terceiro grupo o item “acesso à internet”, e assim vários estudantes do quarto grupo migraram para o terceiro grupo. Já a queda do “Grupo 2” de 2017 para o quantitativo do segundo grupo de 2019 (“Nível II”) ocorreu devido às mudanças no Inse 2019, que no critério escolaridade dos pais passou a englobar os estudantes que tinham pais com ensino fundamental completo a partir do “Nível III”, diferente do Inse 2017 que contemplava esse critério no “Grupo 2”.

A subseção a seguir aborda a aplicação do teste de Kruskal-Wallis com o objetivo de verificar que existe uma diferença estatisticamente significativa entre as médias de matemática de diferentes grupos socioeconômicos.

Diferença estatisticamente significativa entre as proficiências dos níveis socioeconômicos

Para verificar se a diferença entre as médias da proficiência de matemática dos diferentes grupos socioeconômicos é estatisticamente significativa optou-se por usar o teste de Kruskal-Wallis. Por se tratar de um teste não-paramétrico, o teste de Kruskal-Wallis pode ser utilizado se os dados dependentes não seguem uma distribuição normal. Visando verificar se esse pressuposto é satisfeito pelos dados, aplicou-se o teste de Kolmogorov-Smirnov e obteve-se um $p\text{-valor} < 0.001$ para a variável dependente *nota_saeb_matematica* em cada ano de aplicação do Saeb, o que indica que não esses dados não seguem uma distribuição normal. Sendo assim, seguiu-se para o teste de Kruskal-Wallis.

Figura 5 - A proficiência por cada grupo socioeconômico ao longo dos anos

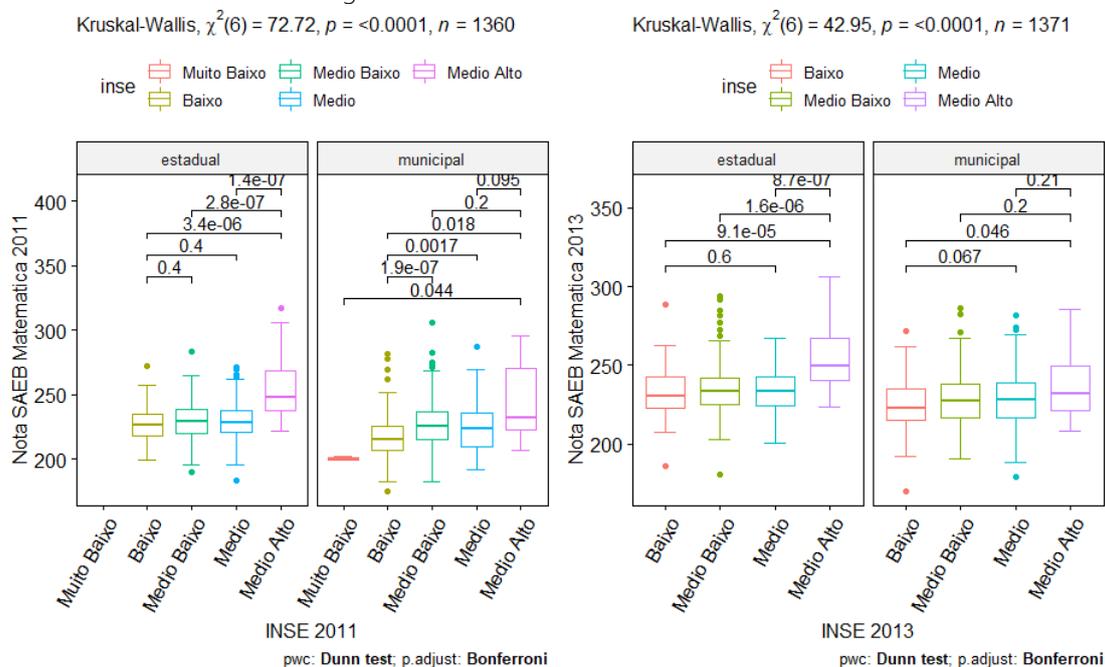


Fonte: Dados da pesquisa

Olhando a Figura 5, pode-se perceber a existência de uma relação diretamente proporcional entre a proficiência e o nível socioeconômico. Ao aplicar o teste de Kruskal-Wallis, comparando as médias de proficiência por grupo socioeconômico em cada ano de aplicação do Saeb, ou seja, em 2011, 2013, 2015, 2017 e 2019, obteve-se p-valores menores que 0,0001. Pelos resultados, percebe-se que ao menos dois grupos sociais apresentam médias diferentes. Por fim, aplicou-se o teste de Dunn para a comparação dois a dois, para identificar os grupos que apresentaram diferença estatisticamente significativa entre si.

A aplicação do teste de Dunn foi feita em cada ano e exposta por rede para melhor identificação da diferença em questão. Os níveis socioeconômicos que não aparecem nos eixos x dos gráficos a seguir, não demonstraram diferença significativa com nenhum outro nível. Acima dos gráficos aparecem uma expressão matemática que informa o valor do qui-quadrado em função do grau de liberdade ($\chi^2(df)$), o p-valor do teste de Kruskal-Wallis (p) e o número de observações (n).

Figura 6 - Resultado dos anos 2011 e 2013



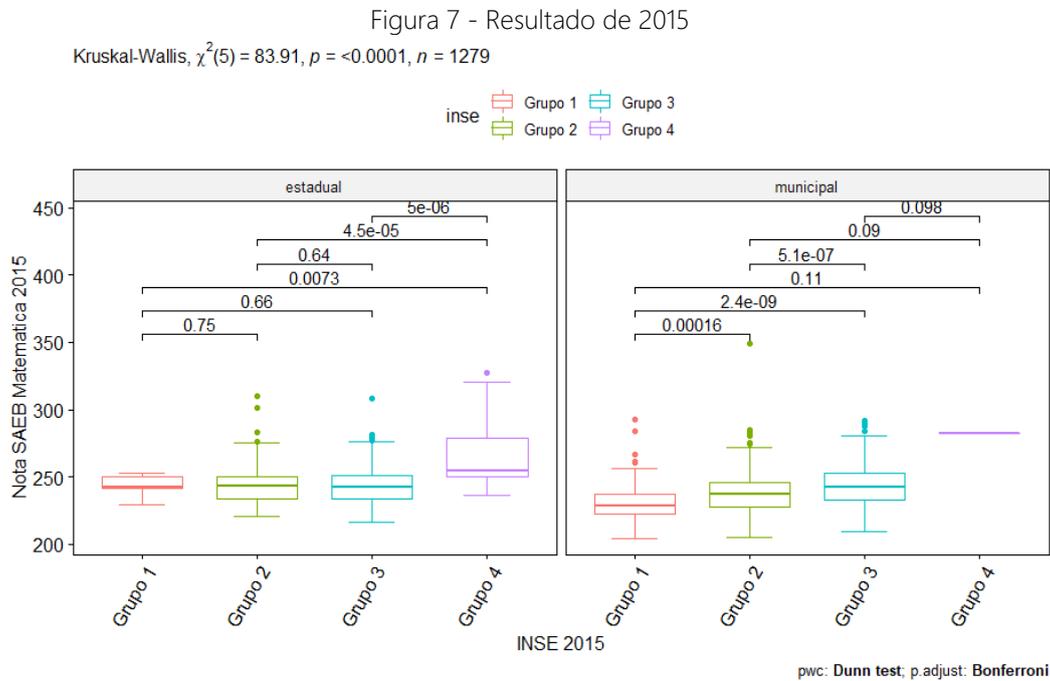
Fonte: Dados da pesquisa

Pode-se observar o resultado dos anos de 2011 e 2013 na Figura 6. Percebe-se que, no ano de 2011, a rede estadual apresenta três diferenças significativas entre o nível socioeconômico “Médio Alto” e os níveis “Médio”, “Médio Baixo” e “Baixo”, com um p-valor < 0.0001. Pode-se notar que a maior parte dos estudantes do Inse “Médio Alto” apresentam proficiência maior que os demais níveis.

A rede municipal, em 2011, apresentou diferença significativa entre “Médio Alto” e os níveis “Baixo” e “Muito Baixo”, com um p-valor de 0.018 e 0.044, respectivamente. Ainda tratando da rede municipal, o teste também identificou diferença significativa entre os níveis “Médio” e “Baixo”, e “Médio Baixo” e “Baixo”, com p-valor < 0.01. É possível perceber, nessas diferenças citadas, que o nível social maior sempre apresenta médias de matemática superiores ao nível no social inferior ao qual mostrou diferença significativa.

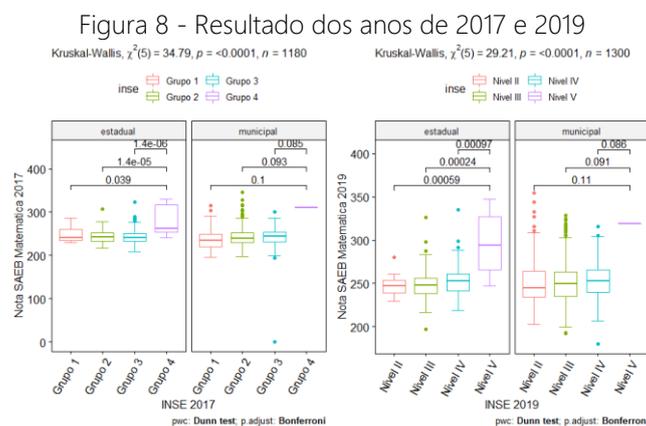
Já no ano de 2013, da perspectiva da rede municipal, apresentou diferença significativa apenas entre os níveis “Médio Alto” e “Baixo” com um p-valor de 0.046. Já na rede estadual

percebe-se que há uma diferença estatisticamente significativa entre o nível “Médio Alto” e os níveis “Médio”, “Médio Baixo” e “Baixo” com um p-valor < 0.001. Pode-se ver uma relação semelhante a observada em 2011 quanto as médias de matemáticas dos diferentes grupos socioeconômicos.



Fonte: Dados da pesquisa

O ano de 2015 apresentou um tamanho de efeito moderado, diferente dos demais anos que possuíam um tamanho de efeito pequeno. Dentre os 5 anos analisados, este ano foi o que apresentou a relação mais forte entre proficiência em matemática e nível socioeconômico. Na Figura 7, em rede estadual, pode-se ver que há uma diferença significativa entre o nível “Grupo 4” e os níveis “Grupo 3” e “Grupo 2”, com um p-valor menores que 0.0001. Já na municipal, mostra uma diferença estatisticamente significativa entre o nível “Grupo 3” e os níveis “Grupo 2” e “Grupo 1”, com um p-valor < 0.000001, além da diferença entre os níveis “Grupo 2” e “Grupo 1”, com um p-valor de 0.00016. Pode-se notar também que a proficiência em matemática dos estudantes é diretamente proporcional ao seu nível socioeconômico.



Fonte: Dados da pesquisa

Nos anos de 2017 e 2019 (Figura 8) é possível notar uma quantidade menor de diferença estatisticamente significativa entre os níveis socioeconômicos com relação aos anos anteriores. Em 2017, apenas a rede estadual apresentou alguma diferença significativa, e essa diferença se deu entre o nível "Grupo 4" e os níveis "Grupo 3", "Grupo 2" e "Grupo 1, com um p -valor < 0.05 . Enquanto que, em 2019, novamente, apenas a estadual apresentou alguma diferença, que se deu através da relação entre os níveis "Nível V" e os níveis "Nível IV", "Nível III" e "Nível II", com p -valor menor que 0.001. Pode-se perceber que os grupos socioeconômicos maiores apresentam médias de matemática superiores aos grupos de nível socioeconômico inferior. O fato da rede municipal não mostrar diferença significativa pode implicar, mais provavelmente, que o questionário socioeconômico não foi preenchido corretamente, ou que o nível socioeconômico não está influenciando mais na nota de matemática do Saeb aplicada à rede municipal.

A partir desses resultados, pode-se concordar com Alves et al. 2007, Ferreira et al. 2016, Silva et al. 2019, Araújo et al. 2020, Proença et al. 2020, Gomes et al. 2021 e Paula, 2021 quanto a influência que o Inse tem sobre a proficiência dos estudantes, embora o resultado da rede municipal de 2019 não tenha se comportando como o esperado.

Descoberta das características socioeconômicas mais relevantes dos estudantes

Como descrito na seção 2.2.2, foi realizado um processo de levantamento das características socioeconômicas que possuem maior significância na proficiência em matemática, considerando os estudantes pertencentes ao 9º ano do Ensino Fundamental das redes estaduais e municipais do estado de Pernambuco. As 5 características socioeconômicas mais relevantes estão elencadas no Quadro 2.

Quadro 2 - Características socioeconômicas mais relevantes.

Questão	Respostas	Grupo Abaixo da Média	Grupo Acima da Média
TX_RESP_Q017A: Fora da escola em dias de aula, quanto tempo você usa para: - Lazer (TV, internet, jogar bola, música etc.).	Não uso meu tempo para isso. Menos de 1 hora. Entre 1 e 2 horas. Mais de 2 horas.	07,95% 15,87% 23,31% 52,87%	02,45% 07,54% 23,36% 66,65%
TX_RESP_Q015: Você já foi reprovado?	Não. Sim, uma vez. Sim, duas vezes ou mais.	67,92% 23,81% 08,27%	84,56% 11,57% 03,87%
TX_RESP_Q010A: Na sua casa tem: - Tv a cabo (Ex.:Net©, Sky© etc.).	Não. Sim.	46,26% 53,74%	57,42% 42,58%
TX_RESP_Q010I: Na sua casa tem: - Freezer (independente ou segunda porta da geladeira).	Não. Sim.	39,83% 60,17%	29,59% 70,41%
TX_RESP_Q016: Alguma vez você	Nunca. Sim, uma vez.	94,77% 04,48%	95,65% 03,58%

abandonou a escola deixando de frequentá-la até o final do ano escolar?	Sim, duas vezes ou mais.	00,75%	00,77%
-------------------------------------------------------------------------	--------------------------	--------	--------

Fonte: Dados da pesquisa

Na primeira coluna do quadro acima, estão as questões em ordem decrescente de relevância, na segunda coluna estão presentes as alternativas de respostas e, na terceira e quarta coluna, a distribuição em porcentagem das respostas dadas, considerando uma divisão entre alunos pertencentes ao grupo abaixo da média e o grupo acima da média. É possível perceber que existe destaque nas questões relacionadas a reprovação e abandono escolar, presença de tv a cabo e freezer em casa, e a mais determinante, o tempo que o estudante possui de lazer fora da escola em dias de aula.

Os dados da questão "TX_RESP_Q017A: Fora da escola em dias de aula, quanto tempo você usa para: - Lazer (TV, internet, jogar bola, música etc.)" está diretamente relacionado ao tempo que a criança e/ou adolescente usufrui para o seu lazer e o tempo que ele se dedica a outras afazeres. Alguns desses afazeres são os trabalhos domésticos e o próprio trabalho para ajudar na renda familiar. Paula (2021) indica que há muitas situações assim, e quando isso acontece, ela afirma que quando os estudantes exercem trabalhos domésticos, seu desempenho é fragilizado, e aqueles que exercem trabalho não-doméstico são ainda mais penalizados.

Em relação à questão "TX_RESP_Q015: Você já foi reprovado?", tem-se que 32% dos estudantes considerados abaixo da média já foram reprovados ao menos uma vez, o que é o dobro, considerando os estudantes que estão acima da média, 15,5%. Essas constatações vão de encontro com a questão "TX_RESP_Q016: Alguma vez você abandonou a escola deixando de frequentá-la até o final do ano escolar?" apresentando percentuais de aproximadamente 5% para estudantes que estão abaixo da média e em torno de 4% para estudantes com proficiência acima da média.

Nesse sentido, Viana et al. (2019) apontam em seus estudos que há "maior índice de reprovação para os alunos do Ensino Fundamental e maior índice de abandono para os alunos do Ensino Médio" (VIANA et al., 2019, p. 55). Essa constatação dos pesquisadores também é explicitada em Carvalho et al. (2020). Os mesmos enfatizam que é comum na rede pública essas altas taxas de reprovação e abandono, e que a rede estadual de ensino perde mais alunos também influenciados pela busca por trabalho e formas de contribuir na renda familiar.

As questões "TX_RESP_Q010A: Na sua casa tem: - TV a cabo (Ex.: Net©, Sky© etc.)" e "TX_RESP_Q010I: Na sua casa tem: - Freezer (independente ou segunda porta da geladeira)", segundo Alves et al. (2014), estão relacionadas ao poder de consumo e à renda da família. Entretanto, constatou-se em nesta análise que 53,74% dos estudantes com notas de proficiência abaixo da média afirmaram possuírem TV a cabo em casa. Assim como 60,17% desses estudantes afirmaram possuírem Freezer. Essas informações já foram analisadas por Paula (2021), na qual identificou que menos tempo dedicado à TV, ao videogame e/ou computador significa maior proficiência.

De acordo com Viana et al. (2019) a renda familiar afeta diretamente no futuro de seus descendentes, uma vez que "famílias com menor renda não conseguem investir economicamente na vida escolar dos filhos" (VIANA et al., 2019, p. 55). Essa é uma constante na literatura, Gomes et al. (2020) apontaram em sua pesquisa que famílias com

renda de até 2 salários mínimos apresentam proficiência muito abaixo da perspectiva, e quanto menor a renda familiar, maiores são as chances do fracasso escolar. Araújo et al. (2020) constata que a desigualdade de renda é o que mais afeta o desempenho escolar dos estudantes, visto que as dificuldades enfrentadas pelas famílias acabam sendo transferidas, seja psicológica ou emocionalmente, para o estudante. Esses estudos apresentam um panorama real evidenciados nos índices e indicadores educacionais. Implementar políticas públicas que tornem essa realidade mais igualitária se faz necessária e urgente.

Considerações finais

Esta pesquisa objetivou investigar a relação entre o Inse e a proficiência em matemática dos estudantes do 9º ano de escolas públicas de Pernambuco fornecidos pelos microdados do Saeb das edições de 2011 a 2019. Para isso, fez-se uso dos microdados do Saeb, produzidos pelo Inep dentre os anos de 2011 até 2019, empregando estatística descritiva e inferencial por intermédio do software RStudio, versão 4.1.0, e dos pacotes tidyverse, ggpubr e rstatix.

A literatura pesquisada já indicava que fatores socioeconômicos afetam os estudantes em seu desempenho nas atividades escolares, seja por questões relacionadas à renda familiar, e/ou por questões relativas ao lugar onde mora. Os dados utilizados também indicaram para esse sentido ao mostrarem que grupos socioeconômicos superiores apresentam índices de proficiência em matemática maiores que os grupos socioeconômicos inferiores. As informações produzidas nesta pesquisa, apontam também que as escolas dos estados apresentaram melhor essa relação com o passar dos anos quando comparadas com as da rede municipal.

Ao analisar os indicadores socioeconômicos, notou-se cinco fatores mais relevantes dentre as quarenta e nove questões disponíveis no questionário 2019 desenvolvido pelo Inep. Essas questões estão relacionadas ao tempo de lazer do estudante, ao poder de consumo de sua família e a relação de abandono e reprovação escolar. Segundo a literatura, essas questões sendo avaliadas negativamente, as tornam fatores preponderantes à evasão, ao baixo desempenho junto às atividades escolares.

Em síntese, ponderou-se que muito mais poderia ser extraído dos dados apresentados, porém percebe-se que nem todas as escolas do país foram contempladas com as avaliações externas, e conseqüentemente, com o questionário socioeconômico. E esse fato dificulta a precisão da avaliação destas provas, uma vez que, leva-se em conta esses aspectos socioeconômicos para tal.

Agradecimentos

Agradecemos a fundação CAPES e a FACEPE pelo financiamento, tornando possível o desenvolvimento deste trabalho, assim como à Universidade Federal Rural de Pernambuco, ao programa PPGE/UFPE – Programa de Pós-Graduação em Ensino das Ciências, e ao grupo de pesquisa EVANTE que auxiliaram o processo de construção do mesmo.

Referências

- ALVES, T., PASSADOR, C. S., & NORONHA, A. B. **A relação entre desempenho escolar, condições de oferta de ensino e origem socioeconômica em escolas da rede pública de ensino.** In XXIII Simpósio Brasileiro, V congresso Luso-Brasileiro e I Colóquio Ibero-Americano de Política e Administração da Educação, Porto Alegre, 2007. Disponível em: https://www.researchgate.net/profile/Thiago-Alves-9/publication/228593289_A_relacao_entre_desempenho_escolar_condicoes_de_oferta_de_ensino_e_origem_socio-economica_em_escolas_da_rede_publica_de_ensino/links/00b4953712d2a86e0c000000/A-relacao-entre-desempenho-escolar-condicoes-de-oferta-de-ensino-e-origem-socio-economica-em-escolas-da-rede-publica-de-ensino.pdf. Acessado em: 08 out. 2021.
- ALVES, T., GOUVÊA, M. A., & VIANA, A. B. N. **Proposta de um indicador socioeconômico para os alunos das escolas públicas dos municípios brasileiros.** Dados, vol. 57, p. 129-166, 2014. Disponível em: <https://www.scielo.br/j/dados/a/SJx5twFq7Y6VjdJJhsqfcYS/?lang=pt>. Acessado em: 15 out. 2021.
- AMÉRICO, B. L., & LACRUZ, A. J. **Contexto e desempenho escolar: análise das notas na Prova Brasil das escolas capixabas por meio de regressão linear múltipla.** Revista de Administração Pública, vol. 51, p. 854-878, 2017. Disponível em: <https://www.scielo.br/j/dados/a/SJx5twFq7Y6VjdJJhsqfcYS/?lang=pt>. Acessado em: 15 out. 2021.
- ARAÚJO, J. M., MARTIN, D. G., FERREIRA, M. A. M., & FARIA, E. D. **Fatores determinantes do desempenho educacional no Sudeste Brasileiro.** Revista Eletrônica Gestão & Sociedade, vol. 14, no. 38, p. 3507-3525, 2020. Disponível em: https://www.researchgate.net/profile/Juliana-Araujo-13/publication/340738558_Fatores_determinantes_do_desempenho_educacional_no_sudeste_brasileiro/links/5e9b079d299bf13079a3f95e/Fatores-determinantes-do-desempenho-educacional-no-sudeste-brasileiro.pdf. Acessado em: 08 out. 2021.
- BRASIL. Ministério da Educação (MEC). **Sistema de Avaliação da Educação Básica: Documentos de referência versão preliminar.** Brasília, 2019. Disponível em: https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_basica/saeb_documentos_referencia-versao_preliminar.pdf. Acessado em: 15 out. 2021.
- CARVALHO, I. D., SANTOS, J. D., & CHRISPINO, Á. **Sucesso e fracasso no ensino fundamental: uma relação entre reprovação, abandono e proficiência.** Revista Meta: Avaliação, Rio de Janeiro, vol. 12, no. 34, p. 136-161, 2020. Disponível em: <https://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/2248/pdf>. Acessado em: 08 out. 2021.
- FERREIRA, R. C., DA SILVA GUILHERME, O. D., & DE OLIVEIRA, A. F. **Indicadores socioeconômicos e desempenho educacional no estado do Tocantins-Br: uma realidade desafiadora apresentada a partir das microrregiões tocantinenses.** Entorno Geográfico, 2016. Disponível em: <https://bibliotecadigital.univalle.edu.co/handle/10893/11530>. Acessado em: 21 out. 2021.
- GOMES, C. M. A., FLEITH, D. D. S., MARINHO-ARAÚJO, C. M., & RABELO, M. L. **Preditores do Desempenho em Matemática de Estudantes do Ensino Médio.** Psicologia: Teoria e Pesquisa,

vol. 36, 2021. Disponível em:

<https://www.scielo.br/j/ptp/a/nmFpbxGtkNVM9x96ZSdLLnr/abstract/?lang=pt>. Acessado em: 15 out. 2021.

GUEDES, T., LEMOS, K. R., & LACRUZ, A. J. **Modelo de regressão logística aplicado na classificação do desempenho de alunos do 5º ano do ensino fundamental de escolas estaduais do Espírito Santo no Saeb**. Navus: Revista de Gestão e Tecnologia, vol. 11, p. 8-19, 2021. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=7774471>. Acessado em: 08 out. 2021.

KASSAMBARA, A. Machine learning essentials: Practical guide in R. Sthda, 2018.

LUNA, L. C.; CARVALHO, J. I. F. **“Oi, Quem está olhando minhas estatísticas?”—Uma discussão do desempenho de estudantes da Educação Básica sobre média aritmética**. Amazônia: Revista de educação em ciências e matemáticas, v. 15, n. 33, p. 151-166, 2019. Disponível em: <https://periodicos.ufpa.br/index.php/revistaamazonia/article/view/6116>. Acesso em: 21 dez. 2022.

MAROCO, J. **Análise Estatística com o SPSS Statistics: 7ª edição**. ReportNumber, Lda, 2018.

PAULA, J. S. **Elementos associados à proficiência em Matemática: um estudo aplicado as escolas públicas em Alagoas**. Diversitas Journal, vol. 6, no. 1, 1114-1141, 2021. Disponível em: https://diversitasjournal.com.br/diversitas_journal/article/view/1487. Acessado em: 15 out. 2021.

PROENÇA, M. C.; MAIA-AFONSO, E. J.; TRAVASSOS, W. B. & CASTILHO, G. R. **Resolução de Problemas de Matemática: análise das dificuldades de alunos do 9.º ano do ensino fundamental**. Amazônia: Revista de educação em ciências e matemáticas, v. 16, n. 36, p. 224-243, 2020. Disponível em: <https://periodicos.ufpa.br/index.php/revistaamazonia/article/view/8639>. Acesso em: 21 dez. 2022.

SILVA, P., DO NASCIMENTO, R. L. S., LIMA, M., FAGUNDES, R., & DE SOUZA, F. D. F. **Modelos de regressão aplicados a predição do desempenho escolar de estudantes do ensino fundamental**. In Brazilian Symposium on Computers in Education, vol. 30, no. 1, p. 1621, 2019. Disponível em: <http://ojs.sector3.com.br/index.php/sbie/article/view/8895>. Acessado em: 08 out. 2021.

VIANA, K. C., DO NASCIMENTO, T. S., MEIRELES, K. D., HOMRICH, G. H., & SEIBERT, C. S. **O impacto das condições sociais na formação dos alunos do Colégio Estadual Doutor Pedro Ludovico Teixeira em Porto Nacional, Tocantins, Brasil**. Revista Interface (Porto Nacional), vol. 18, no. 18, 47-56, 2019. Disponível em: <https://sistemas.uft.edu.br/periodicos/index.php/interface/article/view/8311>. Acessado em: 30 out. 2021.