

Variáveis com impacto no desempenho educacional das Diretorias de Ensino paulistas no IDESP 2022

Lucas Pereira Sperandio 1

Gerente de Políticas Públicas 1
Associação Parceiros da Educação
São Paulo, SP, Brasil
lucas@parceirosedu.org.br

Larissa Brêtas Moura 2

Mestra em Ciências
Universidade de São Paulo 2
Piracicaba, São Paulo, Brasil idade,
larissa.bretas@usp.br

Resumo

As diretorias regionais de ensino são fundamentais para a implementação de políticas educacionais em São Paulo. Este estudo analisou variáveis que influenciam o desempenho dos estudantes de escolas públicas dos Anos Finais e do Ensino Médio em nível regional. A revisão da literatura destacou que condicionantes iniciais de aprendizagem, dimensões socioeconômicas e raciais, e o modelo pedagógico são determinantes do desempenho educacional. Utilizando o método “stepwise” e a técnica de “clustering”, as diretorias foram agrupadas em grupos homogêneos internamente e heterogêneos entre si, permitindo a aplicação do modelo “diferenças em diferenças” para identificar boas práticas. Os resultados apontaram que o perfil racial apresenta a maior correlação com o rendimento estudantil, seguido pelo ensino integral e pelo nível de aprendizagem inicial. O estudo possibilitou a criação de “clusters” para medir o avanço entre 2021 e 2022, identificando “outliers” de desempenho entre as regionais.

Palavras-chave: Educação; indicadores educacionais; desigualdade racial; “Machine Learning”.

1 Introdução

A importância das instâncias regionais no sistema educacional público brasileiro é amplamente reconhecida, pois elas atuam como elo entre o nível central e as escolas, garantindo a implementação eficaz das políticas públicas (Flório e Montoro, 2017). No estado de São Paulo, a Secretaria da Educação (SEDUC-SP) e as 91 Diretorias de Ensino (DEs) representam, respectivamente, os níveis central e regional. Entre as atribuições das DEs, previstas no Art. 72 do Decreto N° 64.187/2019 (São Paulo, 2019), destaca-se o monitoramento dos indicadores de desempenho escolar, especialmente o Índice de Desenvolvimento da Educação do Estado de São Paulo (IDESP), que avalia a qualidade educacional combinando proficiência em Língua Portuguesa e Matemática com o fluxo escolar (SEDUC-SP, 2022a).

A análise de indicadores educacionais é essencial para o planejamento e a tomada de decisões fundamentadas. Segundo Robert Dahl (2009), a decisão em políticas públicas é o processo de escolha eficaz entre alternativas, e o uso de métodos de análise como o procedimento “stepwise” (Fávero e Belfiore, 2017) e técnicas de “machine learning” são fundamentais para aprimorar esse processo.

Com base nesses princípios, o presente estudo tem dois objetivos principais: (i) identificar as variáveis explicativas que influenciam o desempenho dos alunos dos Anos Finais (AF) do Ensino Fundamental e do Ensino Médio (EM) das escolas estaduais paulistas no IDESP de 2022, utilizando o método “stepwise”; e (ii) aplicar técnicas não supervisionadas de “machine learning” para agrupar as Diretorias de Ensino em grupos homogêneos internamente e heterogêneos entre si, criando grupos de controle para o modelo “diferenças em diferenças”. Esses grupos visam contribuir para a identificação, disseminação e premiação de boas práticas, definir metas realistas e permitir comparações mais justas entre as regionais, considerando suas especificidades.

2 Metodologia

Este estudo adotou uma abordagem mista, combinando métodos qualitativos e quantitativos. Segundo Crozatti (2019), o desempenho escolar é um fenômeno multifatorial, exigindo a consideração de múltiplas variáveis. O modelo foi construído com base em evidências, a partir de análise documental (Caulley, 1981; Ludke e André, 1986), e selecionou variáveis explicativas relevantes.

Na análise quantitativa, aplicou-se o método “stepwise” (Fávero e Belfiore, 2017) para selecionar as variáveis mais significativas, mitigando problemas de multicolinearidade.

Em seguida, utilizou-se a técnica de “clustering” para agrupar as Diretorias de Ensino (DEs) em grupos homogêneos internamente e heterogêneos entre si, com o método “K-means” e o critério de “Elbow”.

Por fim, o modelo “diferenças em diferenças” foi utilizado para avaliar o impacto das políticas ao longo do tempo, comparando os resultados das DEs com seus respectivos clusters. O percurso metodológico segue as oito etapas apresentadas na figura a seguir:

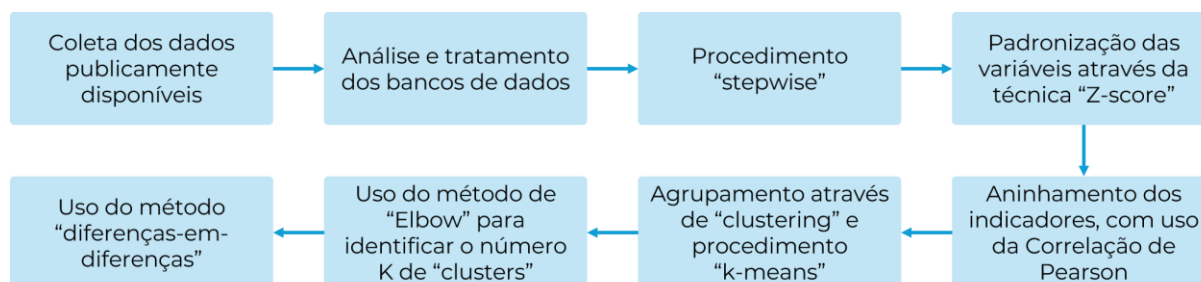
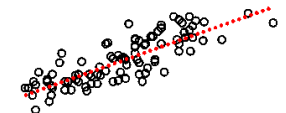
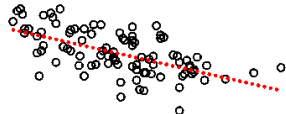
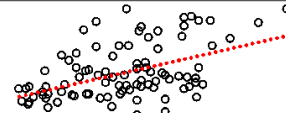


Figura 1. Etapas do percurso metodológico da pesquisa
Fonte: elaboração própria

Os dados foram obtidos de bases públicas estaduais e federais, incluindo acesso via Lei de Acesso à Informação (Brasil, 2011), e as análises foram realizadas no software R (R Core Team, 2023).

3 Resultados

Os resultados deste estudo foram organizados em duas etapas principais, correspondendo aos Anos Finais do Ensino Fundamental (AF) e ao Ensino Médio (EM). As variáveis explicativas selecionadas para o modelo foram analisadas, com destaque para aquelas mantidas após o procedimento “stepwise”, eliminando as que apresentavam multicolinearidade. Nos AF, as variáveis IDESP_2021 (Nota obtida pela DE no Índice de Desenvolvimento da Educação do Estado de São Paulo, em 2021), NB_2022 (% de alunos não-brancos, por DE, em 2022) e PEI_2022 (% de escolas aderidas ao Programa de Ensino Integral, no nível das DEs, em 2022) foram mantidas, sendo que o percentual de alunos não-brancos apresentou a maior correlação negativa com o desempenho escolar. No EM, as mesmas variáveis foram mantidas, mas o PEI_2022 (Programa de Ensino Integral) demonstrou menor significância estatística. A Figura 2 apresenta os valores de correlação, distribuições das variáveis e respectivas significâncias da variável resposta com as variáveis explicativas, para o ciclo dos Anos Finais:

IDESP_2022 [AF]	0,82	-0,61	0,48
	IDESP_2021 [AF]		
		NB_2022 [AF]	
			PEI_2022 [AF]

Legenda: $p < 0.001$; $p < 0.01$; $p < 0.05$

Figura 2. Matriz de correlação do IDESP 2022 e as variáveis explicativas, nos AF
Fonte: resultados originais da pesquisa

Definidas as variáveis explicativas e mensurado o grau de correlação entre os componentes, a partir do coeficiente de correlação de Pearson (r), foi calculado o Indicador Finalístico das Diretorias de Ensino (I_{DE}), padronizado por meio do “Z-Score” e ponderado pela correlação de cada variável. Esse indicador permitiu agrupar as Diretorias de Ensino (DEs) em clusters homogêneos internamente e heterogêneos entre si, utilizando a técnica de “clustering”, aplicada separadamente para os AF e o EM.

O modelo “diferenças em diferenças” foi então utilizado para medir o desempenho das DEs ao longo do tempo, comparando-as com a média de seus clusters e identificando aquelas com avanços mais expressivos. As DEs com os melhores resultados destacaram-se por seu progresso no período analisado, com Itapeva e Taquaritinga se sobressaindo em ambos os ciclos de ensino, sendo potenciais referências para boas práticas.

4 Conclusões e Considerações Finais

As análises e resultados constataram que as variáveis relacionadas às (i) condições socioeconômicas, (ii) à pobreza e (iii) ao perfil racial dos estudantes capturam fenômenos multicolineares. No entanto, diferente da ampla literatura acadêmica, no nível das Diretorias de Ensino, há uma conexão inequívoca - e ainda mais relevante do que no nível das escolas - entre a raça dos alunos e seu desempenho no IDESP, identificando o percentual de alunos não-brancos, como a variável de maior correlação com o indicador. Essa constatação não apenas reforça a persistência do racismo estrutural na sociedade brasileira, mas também

evidencia a urgência de políticas públicas educacionais e sociais que confrontem essa desigualdade racial.

Para além disso, a pesquisa ainda propôs a utilização da técnica de “clustering” para agrupar as Diretorias de Ensino do estado de São Paulo, permitindo análises mais justas e contextualizadas, segundo a realidade das diferentes regiões. Ao identificar padrões e similaridades entre as Diretorias, este estudo abre espaço para uma compreensão mais profunda das necessidades e desafios específicos enfrentados por cada grupo. No entanto, é crucial ressaltar que essas comparações devem ser realizadas com sensibilidade e considerando as particularidades de cada contexto. A simples avaliação de desempenho com base em indicadores isolados pode levar a conclusões injustas e simplistas, ignorando as complexidades e os fatores socioeconômicos que influenciam o cenário educacional.

É necessário, portanto, adotar abordagens analíticas que levem em conta a diversidade de realidades existentes nas Diretorias de Ensino, garantindo que as políticas e intervenções sejam adequadas e eficazes em cada contexto específico. Por fim, identificou-se, a partir do método “diferenças em diferenças”, que algumas regionais registraram desempenhos significativamente maiores do que os seus próprios “clusters”, destacando que as regionais: Itapeva e Taquaritinga, figuraram no “ranking” dos cinco melhores desempenhos, tanto nos Anos Finais, quanto no Ensino Médio, o que deixa um gancho para possíveis futuros estudos da área da pedagogia e da gestão escolar, em busca da identificação de práticas exitosas para publicizá-las, apoiando, com dados e evidências, a formulação, implementação e avaliação de políticas públicas educacionais, no estado de São Paulo.

5 Referências

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal. Diário Oficial da União: seção 1, Brasília, DF, 18 nov. 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm. Acesso em: 5 ago. 2023.

CAULLEY, D. N. Document analysis in program evaluation. *Evaluation and Program Planning*, v. 6, n. 1, p. 19-29, 1983. Disponível em: [https://doi.org/10.1016/0149-7189\(83\)90041-1](https://doi.org/10.1016/0149-7189(83)90041-1). Acesso em: 12 ago. 2023.

CROZATTI, J. Variáveis de influência no IDEB do ensino fundamental dos municípios paulistas em 2017. In: CONGRESSO BRASILEIRO DE CUSTOS, 26., 2019, Curitiba. Anais [...]. Curitiba: Associação Brasileira de Custos, 2019. p. 1-15. Disponível em: <https://anaiscbc.emnuvens.com.br/anais/article/view/4649>. Acesso em: 5 ago. 2023.

DAHL, R. A. A democracia e seus críticos. São Paulo: Martins Fontes, 2009.

FÁVERO, L. P.; BELFIORE, P. Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata. Rio de Janeiro: Elsevier, 2017.

GOMES, E. B. Q.; TARANTIN JUNIOR, W. Uma análise multinível dos efeitos das unidades federativas brasileiras na renda disponível per capita por domicílio. 2022. Monografia (MBA em Data Science e Analytics) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, SP, 2022. Disponível em: <http://biblioteca.pecege.org.br/>. Acesso em: 12 ago. 2023.

LÜDKE, M.; ANDRÉ, M. E. D. A. Pesquisa em educação: abordagens qualitativas. São Paulo: EPU, 1986.

R CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2023. Disponível em: <https://www.r-project.org/>. Acesso em: 15 jul. 2023.

SÃO PAULO (Estado). Decreto nº 64.187, de 17 de abril de 2019. Reorganiza a Secretaria da Educação e dá providências correlatas. Diário Oficial do Estado de São Paulo, São Paulo, SP, 18 abr. 2019. Disponível em: <https://www.al.sp.gov.br/repositorio/legislacao/decreto/2019/decreto-64187-17.04.2019.html>. Acesso em: 15 jul. 2023.

SECRETARIA DA EDUCAÇÃO DO ESTADO DE SÃO PAULO (SEDUC-SP). Índice de Desenvolvimento da Educação do Estado de São Paulo (IDESP) por escola. São Paulo, 2022a. Disponível em: <https://dados.educacao.sp.gov.br/dataset/%C3%ADndice-de-desenvolvimento-da-educa%C3%A7%C3%A3o-do-estado-de-s%C3%A3o-paulo-idesp-por-escola>. Acesso em: 22 jul. 2023.

SECCHI, L. Análise de políticas públicas: diagnóstico de problemas, recomendação de soluções. São Paulo: Cengage Learning, 2016.

TARANTIN JUNIOR, W.; FÁVERO, L. P. Diferença-em-diferenças: a ciência de dados aplicada à avaliação de impactos sociais e políticas públicas. IT Forum, 2020. Disponível em: <https://itforum.com.br/colunas/diferenca-em-diferencas-a-ciencia-de-dados-aplicada-a-avaliacao-de-impactos-sociais-e-politicas-publicas/>. Acesso em: 5 ago. 2023.