

Um estudo da desigualdade em redes municipais de ensino do estado de São Paulo

Pedro Sartori Dias dos Reis¹

Cátedra Sérgio Henrique Ferreira do Instituto de Estudos Avançados da USP de Ribeirão Preto

Ribeirão Preto, São Paulo, Brasil

pedro.sartori@usp.br

Rafael Naime Ruggiero²

Departamento de Psicobiologia da Universidade Federal de São Paulo (UNIFESP) e Cátedra Sérgio Henrique Ferreira do Instituto de Estudos Avançados da USP de Ribeirão Preto

São Paulo, São Paulo, Brasil

rnuggiero@unifesp.br

João Henrique Rafael Junior³

Cátedra Sérgio Henrique Ferreira do Instituto de Estudos Avançados da USP de Ribeirão Preto

Ribeirão Preto, São Paulo, Brasil

jhenrique@usp.br

Mozart Neves Ramos⁴

Cátedra Sérgio Henrique Ferreira do Instituto de Estudos Avançados da USP de Ribeirão Preto

Ribeirão Preto, São Paulo, Brasil

mozartnramos@gmail.com

Resumo

Este estudo analisa desigualdades educacionais em 68 redes municipais do estado de São Paulo, com população entre 100 mil e 600 mil habitantes, nos Anos Iniciais do Ensino Fundamental. Foram consideradas 1.347 escolas, a partir de cinco indicadores: proficiência em língua portuguesa e matemática no SAEB, nota padronizada, fluxo escolar e taxa de distorção idade-série. Aplicou-se uma análise de componentes principais (PCA), que permitiu classificar as escolas em quatro níveis de desempenho. Aquelas no nível 4 apresentaram os melhores resultados (75% a 100%) e as de nível 1, os mais baixos (0% a 25%). Para investigar os fatores associados aos extremos, utilizou-se regressão logística com dados contextuais do INEP (Prova Brasil). Os resultados mostraram que o nível socioeconômico dos alunos, o nível de formação dos professores e a infraestrutura escolar estão significativamente associados ao desempenho, especialmente nas escolas dos extremos. Conclui-se que persistem desigualdades significativas entre e dentro das redes, e a análise combinada de dados contextuais e técnicas estatísticas revela não só essas disparidades, mas também caminhos para políticas mais eficazes e equitativas.

Palavras-chave: Desigualdade educacional, Análise de componentes principais, Regressão logística, Estatística multivariada.

1 Introdução

O novo Fundeb (BRASIL, 2020) reforça a necessidade das redes públicas de ensino em melhorar a aprendizagem adequada dos estudantes com base nos resultados do Saeb, profissionalizar a gestão escolar e reduzir as desigualdades educacionais. Esses condicionantes ainda representam um grande desafio para as redes públicas de ensino, mesmo em um estado com elevados recursos, como São Paulo. Esse cenário pode ser observado, por exemplo, na própria capital paulista, como mostramos no IX Conbratri (REIS et. al., 2024) – uma rede com grandes desigualdades e desafios imensos no campo da aprendizagem escolar. O não cumprimento desses condicionantes pode representar perdas de recursos, como recentemente divulgado pelo Fundo Nacional de Desenvolvimento da Educação (FNDE), do Ministério da Educação (MEC), que publicou a planilha de municípios que deixarão de receber a Complementação de Valor Aluno Resultado (Vaar) (BRASIL, 2024).

Portanto, conhecer como está o quadro da aprendizagem e da desigualdade escolar torna-se imperativo para o financiamento de estados e municípios. Um dos caminhos pode ser por meio da análise em municípios de porte populacional entre 100 mil e 600 mil habitantes, já que exercem influência decisiva em torno daqueles de pequeno porte que gravitam em torno deles, como apontam (STAM, 2013).

Diante desse contexto, este trabalho tem como objetivo analisar a desigualdade escolar em 1.347 escolas de 68 redes municipais de ensino, de 100.000 a 600.000 habitantes, no estado de São Paulo para os Anos Iniciais do Ensino Fundamental. Para esta finalidade levamos em conta os resultados de língua portuguesa e matemática no exame do SAEB (BRASIL, 2005), a nota padronizada e o fluxo escolar, empregados no cálculo do IDEB (BRASIL, 2007, FERNANDES, 2007), e a taxa de distorção idade-série. A matriz de 1.347 escolas por 5 indicadores foi submetida a uma análise estatística multivariada de componentes principais (MARDIA, 1979). Nesta análise, as escolas foram agrupadas em quatro níveis de desempenho escolar. As de nível 4, com melhor desempenho, apresentam valores da primeira componente principal acima do terceiro quartil (entre os 75% e 100%). Já as de nível 1, com pior desempenho, estão abaixo do primeiro quartil da mesma componente (entre 0% e 25%). Para melhor compreender a diferença entre as escolas de nível 1 e 4 foi empregada a técnica de análise de regressão logística (FERNANDES et. al., 2020) usando um conjunto de indicadores fornecidos pelo INEP na análise do Prova Brasil.

2 Metodologia

A matriz de 1.347 x 5 foi submetida a uma análise estatística multivariada de componentes principais que descreve uma rotação dos eixos originais na busca de novos eixos (componentes principais) nos quais a variância dos dados projetados é maximizada. Matematicamente, essa rotação é realizada pela diagonalização da matriz de covariância dos dados, definida por $(X_{cent}^t \cdot X_{cent}) / (n - 1)$, onde X_{cent} é a matriz de dados X centrada, obtida ao subtrair de cada elemento de X a média de sua respectiva coluna, X_{cent}^t é sua transposta e n é o número de linhas (observações) de X . Como resultado, os elementos dos autovetores representam os cossenos diretores, ou seja, a contribuição (ou peso) dos eixos originais (neste trabalho, os cinco indicadores educacionais) para as componentes principais. Os autovalores, por sua vez, representam a variância em cada componente principal. Assim, por exemplo, o primeiro autovalor da matriz de covariância é igual à quantidade de variância estatística explicada pelo primeiro autovetor. Este autovetor, que define a primeira componente principal (CP1), aponta na direção da variância máxima estatística dos dados investigados. A segunda componente

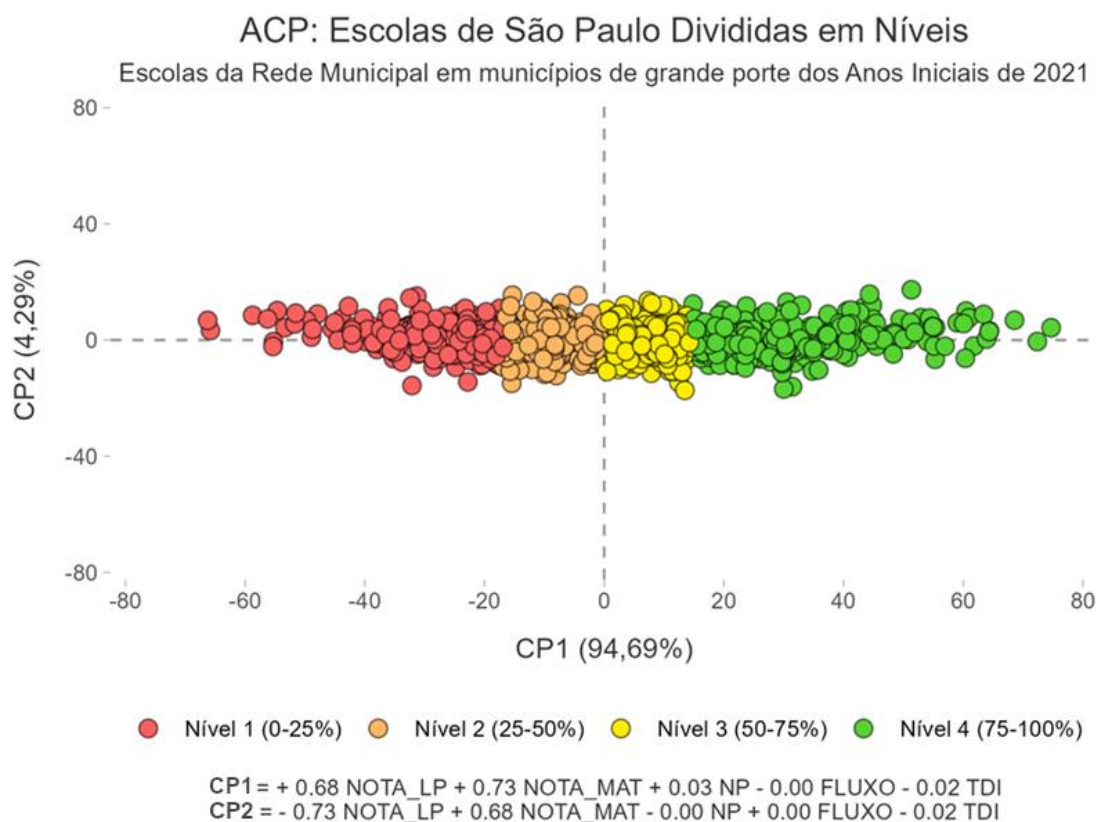
principal (CP2) é perpendicular à CP1 e explica uma quantidade máxima da variância residual dos dados que não foi explicada pela primeira componente. Se as duas primeiras componentes principais explicarem uma quantidade significativa da variância total, um gráfico com a projeção dos pontos nessas CP's pode revelar padrões que não são de simples inspeção na matriz de dados originais.

A partir dessa ACP foi possível dividir as escolas em quatro níveis de desempenho escolar. As de nível 4, de maior desempenho, com valor da primeira componente principal acima do terceiro quartil (entre 75% e 100%), juntamente com as de nível 1, abaixo do primeiro quartil (entre 0% e 25%) foram submetidas a uma análise de regressão logística, que nos permite modelar a relação entre uma variável dependente categórica (geralmente binária) e uma ou mais variáveis independentes, que podem ser contínuas ou categóricas. Ela estima a probabilidade de ocorrência de um evento, com base nas variáveis independentes. Isso nos permitiu melhor compreender a diferença entre as escolas de nível 1 e 4 usando um conjunto de indicadores fornecidos pelo INEP na análise do Prova Brasil.

3 Resultados

Na Figura 1 mostramos o gráfico de escores das duas componentes principais CP1 e CP2 relativo à matriz de 1.347 escolas por 5 indicadores para os Anos Iniciais do Ensino Fundamental de 2021, que explicam juntas, quase 99% de toda a variância original.

Figura 1



A primeira componente principal (CP1) foi utilizada como *proxy* do desempenho escolar porque ela concentra quase toda a variância da matriz original (94,69%) e apresenta coeficientes (cossenos diretores) elevados para as notas de língua portuguesa

e matemática (0,63 e 0,73, respectivamente), o que indica uma forte contribuição dessas variáveis na composição da componente. Como CP1 é uma combinação linear das variáveis originais, esses coeficientes indicam quanto o valor da componente varia em função de cada variável: um acréscimo de um ponto na nota de Matemática, por exemplo, eleva o escore da escola em CP1 em 0,73 unidades, mantendo todas as demais variáveis constantes, e desloca-a para a direita no gráfico. Já variações no fluxo escolar, na nota padronizada ou na taxa de distorção idade-série resultam em alterações muito pequenas nesse eixo.

Já a CP2 apresenta coeficientes de $-0,73$ para língua portuguesa e $+0,68$ para matemática, o que indica que, para um mesmo valor de CP1, as escolas com escores mais altos em CP2 tendem a ter melhor desempenho relativo em matemática do que em língua portuguesa. Embora essa interpretação possa ser útil para análises diferenciais por disciplina, a CP2 não representa o desempenho escolar global de forma direta e, além disso, explica apenas uma fração muito pequena da variância total (4,29%).

Dessa forma, utilizamos apenas a CP1 como base para classificar as escolas em quatro níveis de desempenho: as escolas à direita do gráfico, destacadas em verde, correspondem ao nível 4 (percentil entre 75% e 100% de CP1), enquanto as escolas à esquerda, em vermelho, representam o nível 1 (percentil entre 0% e 25%). Para melhor compreender os fatores que podem explicar a diferença entre elas, realizamos uma análise de regressão logística usando dados do questionário do Prova Brasil e do Censo Escolar, entre eles, a taxa de distorção idade-série, a reprovação escolar, a porcentagem de alunos brancos e pretos, o nível da complexidade de gestão, o número médio de alunos por turma, o nível socioeconômico (NSE), o número de horas-aula por dia e o abandono escolar; além desses, foi também incluído o PIB per capita do município. De acordo com os coeficientes da regressão logística, as escolas de nível 4 apresentam altas taxas de NSE, de % de alunos brancos e de PIB per capita; enquanto aquelas de menor desempenho (nível 1) mostram altas taxas de abandono escolar e de % de alunos pretos.

Fazendo esse procedimento, ao longo dos anos de 2007 a 2021, foi possível identificar os municípios com maior percentual de escolas nos níveis 4 e 1, cujos resultados são mostrados nas Tabelas 1 e 2, respectivamente.

Tabela 1: Os três principais municípios por ano avaliado, entre 2007 e 2021, com maior percentual de escolas no nível 4

2007	2009	2011	2013	2015	2017	2019	2021
Sertãozinho (92,3%)	Marília (100%)	S.C. do Sul (93,7%)	Birigui (100%)	S.C. do Sul (84,2%)	S.C. do Sul (100%)	Indaiatuba (85,2%)	S. Carlos (100%)
Indaiatuba (83,3%)	Indaiatuba (89,5%)	Sertãozinho (92,9%)	S.C. do Sul (100%)	Sertãozinho (80,0%)	Indaiatuba (82,1%)	Birigui (84,6%)	Indaiatuba (80,8%)
S.B. d'Oeste (81,2%)	Sertãozinho (85,7%)	M. Guapu (80,0%)	Sertãozinho (80,0%)	Catanduba (78,6%)	Sertãozinho (80,0%)	Sertãozinho (73,3%)	S.C. do Sul (70,0%)

Tabela 2: Os três principais municípios por ano avaliado, entre 2007 e 2021, com maior percentual de escolas no nível 1

2007	2009	2011	2013	2015	2017	2019	2021
------	------	------	------	------	------	------	------

Mauá (100%)	Carapicuíba (100%)	Carapicuíba (100%)	Jandira (100%)	Mauá (100%)	Mauá (100%)	Mauá (100%)	Itaquaquecetuba (84,0%)
Itaquaquecetuba (93,1%)	Mauá (100%)	Jandira (100%)	Mauá (100%)	F. Rocha (92,3%)	F. Rocha (92,9%)	Itaquaquecetuba (92,1%)	Itapevi (82,4%)
F. Vasconcelos (87,5%)	E. Artes (94,4%)	Itaquaquecetuba (88,9%)	Itaquaquecetuba (83,3%)	Jandira (90,0%)	Itapevi (80,0%)	F. Rocha (87,5%)	F. Rocha (80,0%)

De acordo com a Tabela 1, o município de Sertãozinho aparece com maior frequência nos oito anos avaliados com o percentual de escolas no nível 4 – seis vezes, seguido pelos municípios de Indaiatuba e São Caetano do Sul – cinco vezes. Por outro lado, de acordo com a Tabela 2, o município de Mauá aparece com maior frequência nos oito anos avaliados com o percentual de escolas no nível 1 – seis vezes, seguido pelo município de Itaquaquecetuba – cinco vezes e Franco da Rocha – quatro vezes. Estes dois últimos deixaram de receber a complementação do Vaar do Fundeb por não terem apresentado melhoria em nenhum dos indicadores de atendimento e da aprendizagem com redução das desigualdades, conforme metodologia constante da Portaria MEC nº 975, de 13 de dezembro de 2022 (BRASIL, 2022). Além disso, os municípios de Mauá e de Itaquaquecetuba ocuparam as duas últimas posições no ranking de uma análise longitudinal do Ideb em 49 municípios de médio porte do estado de São Paulo. Vale salientar que o município de Franco da Rocha não participou dessa análise (RAMOS, 2023).

4 Conclusões e Considerações Finais

O estudo combinado das técnicas estatísticas de componentes principais e de regressão logística se mostrou eficaz para analisar as desigualdades, em termos de desempenho escolar, de 68 redes municipais de ensino, de 100 mil a 600 mil habitantes, do estado de São Paulo para os Anos Iniciais do Ensino Fundamental. Foi possível identificar os municípios com maior percentual de escolas com nível 4 – de maior desempenho, ao longo do período de 2007 a 2021, assim como aqueles com maior percentual de escolas no nível mais baixo em desempenho – nível 1. Encontra-se em processo, em nosso Laboratório de Ciência de dados Educacionais, a extensão desse estudo para o ano de 2023.

5 Referências

BRASIL. Ministério da Educação. Portaria n. 931, de 21 de março de 2005. Institui o Sistema de Avaliação da Educação Básica – Saeb. Diário Oficial da União: seção 1, Brasília, DF, n. 55, p. 16-17, 22 mar. 2005.

BRASIL. Ministério da Educação. Decreto n. 6.094, de 24 de abril de 2007. Diário Oficial da União: seção 1, Brasília, DF, n. 79, p. 5, 24 abr. 2007. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/d6094.htm.

BRASIL, Ministério da Educação, Lei nº 14.113, de 25 de dezembro de 2020, que trata da regulamentação do Fundo de Manutenção e Desenvolvimento da Educação Básica e

de Valorização dos Profissionais da Educação (Fundeb), art. 212-A da Constituição Federal, (2020).

BRASIL, Ministério da Educação, Portaria nº 975, de 13 de dezembro de 2022, que aprova a metodologia de cálculo dos indicadores de atendimento e de melhoria de aprendizagem, com redução de desigualdades, consoante o art. 5º, inciso III, o art. 14, caput e §§ 2º e 3º, e o art. 15, inciso III, da Lei nº 14.113, de 25 de dezembro de 2020, que regulamenta o Fundeb, para o exercício financeiro de 2023.
<https://www.gov.br/fnde/pt-br/aceso-a-informacao/acoes-e-programas/financiamento/fundeb/legislacao/2022/portaria-no-975-de-13-de-dezembro-de-2022/view>

BRASIL, Ministério da Educação, Redes de ensino inabilitadas à complementação VAAR 2024 pelo não cumprimento de condicionalidades de melhoria de gestão previstas no art. 14, § 1º, incisos I a V, da Lei nº 14.113/2020, ou por não terem alcançado evolução em indicadores de atendimento e de melhoria da aprendizagem com redução das desigualdades previstos no art. 14, § 2º da Lei nº 14.113/2020. Disponível em:
<https://www.gov.br/fnde/pt-br/aceso-a-informacao/acoes-e-programas/financiamento/fundeb/2024/Redesinabilitadaspormotivo.pdf>

FERNANDES, R.; Índice de Desenvolvimento da Educação Básica (Ideb). Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília, DF, 2007. Disponível em:
https://download.inep.gov.br/educacao_basica/portal_ideb/o_que_sao_as_metas/Artigo_projecoes.pdf.

FERNANDES A.A.T.; FILHO, D.B.F.; ROCHA, E.C.; NASCIMENTO, W.S.; “Read this paper if you want to learn logistic regression”, *Rev. Sociol. Polit.*, vol. 28 n. 74, p. 1-19 (2020). Disponível em: <https://doi.org/10.1590/1678-987320287406en>

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M.; *Multivariate Analysis*, Academic Press, London, 1979. Disponível em:
<https://statisticalsupportandresearch.files.wordpress.com/2017/06/k-v-mardia-j-t-kent-j-m-bibby-multivariate-analysis-probability-and-mathematical-statistics-academic-press-inc-1979.pdf>.

RAMOS, M.N.; SILVA, L.; SILVA, J.B.P.; FILHO, A.J.C.; “Uma análise longitudinal do Ideb em municípios de médio porte do estado de São Paulo”, *Ensaio: aval. pol. públ. Educ.*, Rio de Janeiro, v.31, n.118, p. 1-22, jan./mar. 2023. Disponível em: <https://doi.org/10.1590/S0104-40362023003103818>

REIS, P.S.D.; JÚNIOR, J.H.R.; RUGGIERO, R.N.; RAMOS, M.N.; “Uso de modelagens estatísticas e de ciências dos dados para reduzir as desigualdades escolares da rede pública municipal do Município de São Paulo”, IX Congresso Brasileiro de Teoria de Resposta ao Item (TRI) e Métodos Quantitativos em Avaliação – IX CONBRATRI, Rio de Janeiro (RJ), 2024.

STAMM, C.; STADUTO, J. A. R.; LIMA, J. F.; WADI, Y. M.; “A população urbana e a difusão das cidades de porte médio no Brasil”, *Interações*, vol. 14, no. 2, p. 251-265, 2013. Disponível em: <https://www.scielo.br/j/inter/a/9bhn6dCRJkt8qpV9t6c49JK/?lang=pt>.