



10.24065/re.v15i1.2981

ARTIGO

ESTATÍSTICA MULTIVARIADA APLICADA NO DESEMPENHO ESCOLAR DE PONTA GROSSA-PR

MULTIVARIATE STATISTICS APPLIED TO SCHOOL PERFORMANCE IN PONTA GROSSA-PR

ESTADÍSTICA MULTIVARIADA APLICADA AL RENDIMIENTO ESCOLAR EN PONTA GROSSA-PR

Marcos Teixeira Alves¹ <https://orcid.org/0000-0001-7595-8940>
Scheila Valechenski Biehl² <https://orcid.org/0000-0003-2323-868X>
Deyse Márcia Pacheco Gebert³ <https://orcid.org/0000-0001-7561-7775>

¹ Universidade Estadual de Ponta Grossa; Departamento de Matemática e Estatística - Ponta Grossa, Paraná, Brasil, CEP 84030-900, e-mail: mtalves@uepg.br

² Universidade Estadual de Ponta Grossa; Departamento de Matemática e Estatística - Ponta Grossa, Paraná, Brasil, CEP 84030-900, e-mail: svbiehl@uepg.br

³ Universidade Estadual de Ponta Grossa; Departamento de Matemática e Estatística - Ponta Grossa, Paraná, Brasil, CEP 84030-900, e-mail: dmpgebert@uepg.br

RESUMO

Este trabalho tem como objetivo investigar padrões de desempenho das escolas participantes do Saeb 2019 relativas ao 9º ano do Ensino Fundamental da cidade de Ponta Grossa-PR. A pesquisa quantitativa utilizou da estatística multivariada, em particular, da análise de componentes principais e de agrupamento, baseando-se nas variáveis de interesse: percentuais de alunos com aprendizado adequado em Língua Portuguesa e Matemática no Saeb 2019, a taxa de aprovação e a proporção de alunos sem defasagem escolar no referido ano. Constatou-se que 84,77% da variabilidade dos dados é explicada pelas duas primeiras componentes principais, com peso significativo das variáveis relacionadas com o aprendizado escolar. Com o uso desses componentes e via uma translação de eixos, foi construído um indicador de desempenho das escolas participantes da avaliação.

Palavras-chave: análise de componentes principais; análise de agrupamento; desempenho escolar; aprendizado adequado.

ABSTRACT

This study aims to investigate performance patterns of schools participating in Saeb 2019 related to the 9th grade of elementary school in the city of Ponta Grossa, Paraná. The quantitative research used multivariate statistics, in particular principal component analysis and clustering, based on the variables of interest: percentages of students with adequate learning in Portuguese Language and Mathematics in Saeb 2019, the pass rate, and the proportion of students without academic gaps in that year. It was found that 84.77% of the data variability is explained by the first two principal components, with significant weight given to variables related to school learning. Using these



components and via axis translation, a performance indicator was constructed for the schools participating in this assessment.

Keywords: principal component analysis; cluster analysis; school performance; adequate learning.

RESUMEN

Este trabajo tiene como objetivo investigar los patrones de desempeño de las escuelas participantes en el SAEB 2019 relacionadas con el 9º grado de la Escuela Primaria en la ciudad de Ponta Grossa-PR. La investigación cuantitativa utilizó estadística multivariada, en particular análisis de componentes principales y conglomerados, basada en las variables de interés: porcentajes de estudiantes con aprendizaje adecuado en Lengua Portuguesa y Matemáticas en el SAEB 2019, la tasa de aprobación y la proporción de estudiantes sin discrepancia escolar en ese año. Se encontró que el 84,77% de la variabilidad de los datos se explica por los dos primeros componentes principales, con un peso significativo de las variables relacionadas con el aprendizaje escolar. Con el uso de estos componentes y a través de una traslación de ejes, se construyó un indicador de desempeño de las escuelas participantes en esta evaluación.

Palabras clave: análisis de componentes principales; análisis de conglomerados; rendimiento escolar; aprendizaje adecuado.

INTRODUÇÃO

No que diz respeito às avaliações da Educação Básica, o Brasil tem investido na mensuração da qualidade desse nível de ensino a partir de 1990 com a criação do Sistema de Avaliação da Educação Básica (Saeb). Os objetivos centrais focam-se em obter um diagnóstico da educação básica brasileira e de fatores que podem interferir no desempenho dos alunos (Santos; Ferreira; Simões, 2019). Desde 2019, o sistema contempla todas as etapas do ensino básico, da educação infantil ao ensino médio.

Os dados obtidos através da aplicação de testes e questionários socioeconômicos são um dos requisitos para a construção do Índice de Desenvolvimento da Educação Básica (Ideb). Esse indicador oficial da qualidade do Ensino Básico é o resultado da multiplicação da nota média obtida pelos alunos na Prova Brasil referente às disciplinas de Língua Portuguesa e Matemática com o fluxo escolar (taxa de aprovação).

Embora em sua essência esse índice busque servir de subsídios para diagnóstico, reflexão e planejamento do trabalho pedagógico da escola, bem como para a formulação de ações e políticas públicas, alguns teóricos argumentam a fragilidade e as limitações de seu uso. Pinto (2008) destaca que quanto mais insumos a escola pode prover, como recursos humanos qualificados e infraestrutura, maior será a qualidade de ensino, ocasionando resultados melhores nas avaliações de larga escala. O autor explica que

O atual sistema chamado de avaliação [...], mas que na verdade não passa de um conjunto de testes padronizados de português e matemática, é claramente insuficiente para aferir a qualidade da educação oferecida pelos sistemas de ensino, mesmo considerando os limites dos objetivos postos pela legislação (Pinto, 2008, p. 59).

Os testes de Língua Portuguesa e Matemática são aplicados nos sistemas públicos e particulares de educação, a cada dois anos, contemplando estudantes do 5º e do 9º anos do ensino fundamental, e alunos do 3º ano do ensino médio. As competências exigidas nessas avaliações consistem na capacidade de resolução de problemas para a Matemática, e a leitura e interpretação de textos para a Língua Portuguesa.

Os resultados dessas avaliações demonstram um grande desafio para a educação básica brasileira: além do desempenho estar aquém do esperado pelo INEP, há um visível decréscimo com o passar dos anos. Segundo Ortigão (2008, p. 78), considerando as últimas avaliações para os anos finais do Ensino Fundamental, “após oito anos de escolarização, não construíram competências básicas necessárias para o cotidiano e para prosseguirem nos estudos”.

Para se ter uma ideia do desafio mencionado anteriormente, considerou-se como público-alvo os alunos dos anos finais do Ensino Fundamental II, a saber do 9º ano. A proficiência em Língua Portuguesa e em Matemática é considerada adequada quando o estudante atinge um desempenho acima de 275 e de 300 pontos, respectivamente. De acordo com o Portal QEdu (2025), no estado do Paraná, 43% dos estudantes apresentaram aprendizado adequado em Língua Portuguesa, enquanto 25% aprendizado adequado em Matemática no Saeb de 2019. Em nível nacional, 36% possuíram aprendizado adequado em Língua Portuguesa e apenas 18% em Matemática.

Ponta Grossa, situada a 103 km da capital do Paraná, Curitiba, é a quarta cidade mais populosa do estado, contando atualmente com 358 371 habitantes (IBGE, 2022). De acordo com o Portal QEdu (2025), o município conta com indicador de fluxo, em 2019, de 0,99, o que indica que a cada 100 alunos, um não foi aprovado. Com relação às taxas de aprendizado adequado, o portal apresenta os seguintes valores: 41% para Língua Portuguesa e 24% para Matemática, pouco abaixo dos valores do estado. Destaca-se que as escolas responsáveis pelos anos finais do Ensino Fundamental são estaduais.

Neste trabalho, o objetivo central consiste na aplicação de duas técnicas de estatística multivariada, a saber: análise de componentes principais e de agrupamento, visando investigar padrões de desempenho das escolas participantes do Saeb 2019 relativas ao 9º ano do Ensino Fundamental da cidade de Ponta Grossa. Para tanto, considerou-se as variáveis: percentual de alunos com aprendizado adequado em Língua Portuguesa e em Matemática, a taxa de aprovação e a taxa de distorção série, que mede a proporção de alunos com mais de dois anos de atraso escolar.

A Lei de Diretrizes e Base da Educação 9.394/1996 recomenda que a criança deve ingressar aos 6 anos no 1º ano do ensino fundamental e concluir a etapa aos 14. Na faixa etária dos 15 aos 17 anos, o jovem deve estar matriculado no ensino médio. Desse modo, o valor da distorção é calculado em anos e representa a defasagem entre a idade do aluno e a idade recomendada para a série que ele está cursando. O aluno é considerado em situação de distorção ou defasagem idade-série quando a diferença entre a idade do aluno e a idade prevista para a série é de dois anos ou mais. Para fins de interpretação, introduzimos a variável NDis que compreende o complementar da taxa de distorção idade-série descrita acima, ou seja, o percentual de alunos obedecendo as recomendações dispostas na LDB 9.394/1996.

Os objetivos específicos consistem em analisar os pesos das variáveis descritas no parágrafo anterior na obtenção das componentes principais, bem como realizar a análise de agrupamento para verificar a similaridade entre as escolas participantes. Estudo semelhante a este já foi realizado com sucesso em outros trabalhos, como Ramos et al. (2020) e Ramos et al. (2021). Como desfecho da pesquisa, os autores construíram um indicador de desempenho das 36 escolas participantes do Saeb 2019 com base nos escores obtidos com a aplicação desta metodologia exploratória.

METODOLOGIA

Nesta pesquisa, foram utilizados os dados de 36 escolas do município de Ponta Grossa-PR participantes do Saeb 2019 relativos ao 9º ano do Ensino Fundamental. Para cada escola foi coletado o percentual de alunos com aprendizado adequado em Língua Portuguesa (Port) e em Matemática (Mat), a taxa de aprovação (FI) e a taxa de alunos sem distorção idade-série (NDis). As informações das 36 escolas foram alocadas de acordo com a nota do Ideb e estão apresentadas no Quadro 1.

Quadro 1 - Escolas de Ponta Grossa-PR participantes do Saeb 2019

Posição Ideb	Escola	Port	Mat	FI	NDis
1	Medalha Milagrosa	65	55	95	95,5
2	Prof. Júlio Teodorico	63	47	89	89,2
3	De Vila Velha	52	24	98	91,8
4	Profª. Margarete Mazur	44	13	98	88,9
5	Francisco P. Machado	37	13	96	84,7
6	Alberto R. Valente	39	32	94	90,2
7	Pe. Carlos Zelesny	35	20	96	89,8
8	Espírito Santo	49	20	97	90,1

9	Jesus Divino Operário	47	28	92	92,5
10	Nossa Sra. das Graças	41	25	92	89,7
11	Prof. Edison Pietrobelli	47	17	92	90,1
12	Profa. Elzira C. de Sá	40	23	93	92,5
13	Prof. João R. von Borell	43	27	89	87,1
14	Prof. Meneleu A. Torres	50	21	92	86,1
15	Pe Arnaldo Jansen	50	31	89	83,8
16	Prof. Colares	40	17	95	77,5
17	Monteiro Lobato	48	18	92	90,9
18	Bento Mossuranga	41	21	90	88,2
19	Pres. Kennedy	39	30	90	89,3
20	Dorah Daitschman	55	14	84	77,3
21	Dr. Epaminondas Ribas	53	22	81	86,4
22	Prof. Cesar Martinez	48	20	83	88
23	Dr Munhoz da Rocha	46	24	84	88,1
24	Prof. Eugenio Malanski	42	25	85	83,6
25	Prof. Iolando Taques	34	11	89	84
26	Prof. Amálio Pinheiro	38	25	84	79,4
27	Ana Boratto	28	14	91	89,5
28	Polivalente	28	21	84	79,6
29	Profa. Sirley Jagas	28	10	88	85,8
30	José Elias da Rocha	41	21	81	77,6
31	Nossa Sra. da Glória	26	20	82	79,6
32	Antonio Sampaio	24	19	87	77,8
33	Pe. Pedro Grzelczaki	27	25	79	83,2
34	31 de Março	35	21	84	86,2
35	Prof. José do Amaral	20	8	81	80,8
36	Prof. Becker e Silva	26	20	70	72,7

Fonte: Portal QEdU (2025).

As variáveis Port e Mat indicam os percentuais de alunos com aprendizado adequado em Língua Portuguesa e Matemática no Saeb 2019, respectivamente. O fluxo (Fl) apresenta a taxa de aprovação em 2019. Finalmente, a variável NDis é a taxa complementar da distorção idade-série e mede a proporção de alunos sem defasagem entre sua idade e a idade recomendada para a série que ele está cursando. Os dados relativos às três primeiras variáveis foram extraídos da plataforma QEdU (2025), enquanto os dados da variável NDis foram obtidos no Censo Escolar do Inep/MEC.

Uma das técnicas de estatística multivariada utilizada neste trabalho consiste na análise de componentes principais (ACP). Nesta técnica, segundo Mingoti (2005), transforma-se o conjunto de variáveis originais em um outro conjunto de variáveis de mesma dimensão, chamadas componentes principais, que possuem propriedades importantes: são escritas como combinação linear das variáveis originais, não correlacionadas e estimadas com o propósito de reter o máximo de informação em termos da variação total contida nos dados. Além destas propriedades, a ACP é utilizada para geração de índices e agrupamentos entre as observações conforme percebe-se adiante neste estudo.

Denotando por X a matriz dos dados contidos no Quadro 1, em que as linhas representam as escolas e as colunas as variáveis de interesse: Port, Mat, Fl e NDis, as componentes principais são obtidas a partir dos autovetores associados a matriz de covariâncias de $X^T X$ com X^T a matriz transposta de X . Já os autovalores estão associados com a variância explicada por cada componente principal.

A primeira componente principal CP_1 tem variância máxima em relação às demais, isto é, os valores desta componente têm variância maior do que os de quaisquer outras componentes. Na ACP, cada nova componente obtida, explica cada vez menos a variabilidade total. Dessa forma, a segunda componente principal CP_2 explica uma quantidade da variância total que não foi explicada pela CP_1.

Para Hair et al. (2019), geometricamente, a ACP consiste em realizar uma rotação no sistema de coordenadas das variáveis originais, de forma que o novo sistema capte a máxima variabilidade existente entre as observações.

Considerando que as duas primeiras componentes principais explicam uma quantidade significativa da variância total, pode-se construir um gráfico com os eixos dados pela direção destas componentes, sendo que a projeção dos pontos nestes eixos pode mostrar padrões que não seriam captados imediatamente observando a matriz dos dados originais. Para a análise das componentes principais discutidas na próxima seção, utilizou-se os dados centrados na média.

Outra abordagem multivariada empregada nesta pesquisa foi a análise de agrupamento. Essa técnica, segundo Ferreira (2009), tem como objetivo formar grupos que apresentam alta similaridade entre si e ao mesmo tempo apresentem a máxima distinção com os demais grupos formados. Essa similaridade é baseada nas distâncias calculadas entre as escolas, considerando as quatro variáveis em estudo: Port, Mat, Fl e NDis.

Inicialmente, foram calculadas as distâncias euclidianas entre todas as escolas, duas a duas, gerando uma matriz de distâncias. A partir disso, procedeu-se com o método hierárquico de agrupamento chamado método da ligação completa, ou “vizinho mais distante”. Nesse método, inicialmente verifica-se a menor distância entre duas escolas, essas escolas são unidas em um grupo. Em seguida a matriz de distâncias é atualizada com esse novo grupo, mantendo sempre a maior distância do novo grupo com cada uma das demais escolas. As distâncias entre escolas que não foram afetadas pelo novo grupo são mantidas. A matriz de distâncias estando atualizada, procede-se novamente com a verificação da menor distância entre escolas (ou grupo formado anteriormente) e faz-se a união. Isso é repetido até que todas as escolas estejam em

um único grupo. A visualização dessas uniões é feita por meio de um dendrograma ilustrado na próxima seção.

A validação do agrupamento foi realizada por meio da correlação cofenética, que nada mais é do que a correlação entre as distâncias originais dos dados com a as distâncias verificadas no dendrograma: quanto mais alta a correlação, mais fiel à estrutura real dos dados é o agrupamento obtido. As análises estatísticas foram feitas utilizando-se o software R (R Core Team, 2019).

RESULTADOS E DISCUSSÕES

Uma análise estatística descritiva é apresentada no Quadro 2. Iniciando pelo aprendizado adequado em Língua Portuguesa, o percentual médio, entre todas as escolas em estudo, foi de 40,8%, com um desvio padrão de 10,68%, sendo esta a variável que apresentou a maior variabilidade. O menor percentual de aprendizado adequado em Língua Portuguesa é da escola Prof. José do Amaral com 20%, já a escola Medalha Milagrosa apresentou o maior aprendizado desta disciplina com 65%.

Quadro 2 - Estatística Descritiva

Variáveis	Média	Desvio Padrão	mín.	Quartil 1	Quartil 2 (mediana)	Quartil 3	máx.
<i>Por</i>	40,8	10,68	20	34,75	41	48	65
<i>Mat</i>	22,27	9,12	8	17,75	21	25	55
<i>Fl</i>	88,5	6,24	70	84	89	92,25	98
<i>NDis</i>	85,7	5,36	72,7	82,6	86,75	89,72	95,5

Fonte: Dados da pesquisa (2025).

No que diz respeito ao aprendizado adequado em Matemática, observou-se uma média significativamente menor, de 22,27%. Considerando o Quartil 2 (mediana), percebe-se que metade das escolas teve porcentagens de aprendizado inferiores a 21%. Notadamente, o menor valor registrado para o aprendizado adequado em Língua Portuguesa foi de 20%, o que indica que, em Matemática, metade das escolas apresentou desempenhos ainda mais baixos do que este mínimo. As escolas que apresentaram maior e menor percentual de aprendizado adequado em Matemática com 55% e 8% foram novamente as escolas Medalha Milagrosa e Prof. José do Amaral, respectivamente.

Em relação à taxa de aprovação, a média foi de 88,5%, com desvio padrão de 6,24%. Pode-se observar que 50% das escolas apresentaram percentuais de aprovação maiores que

89,9%. A escola Prof. Becker e Silva apresentou o menor percentual, 70%, e as escolas com o maior percentual foram Profa. Margarete Mazur e a de Vila Velha, as duas com taxa de aprovação de 98%. Vale ressaltar que essas escolas não foram as mesmas que apresentaram os maiores percentuais de aprendizado adequado em Matemática e em Língua Portuguesa.

Sobre a proporção de alunos sem defasagem de idade-série, a média das escolas em estudo foi de 85,7%, com desvio padrão de 5,36%. A escola com menor percentual de alunos sem defasagem foi a escola Prof. Becker e Silva, com 72,7% e a escola Medalha Milagrosa apresentou o maior percentual, 95,5%.

O próximo interesse consistiu na análise de componentes principais. As correlações entre as variáveis estudadas são apresentadas no Quadro 3.

Quadro 3 - Correlações entre as variáveis: *Por*, *Mat*, *Fl* e *NDis*

Variáveis	Port	Mat	Fl	NDis
Port	1			
Mat	0,585	1		
Fl	0,392	0,122	1	
Ndis	0,521	0,404	0,672	1

Fonte: Dados da pesquisa (2025).

De modo geral, observou-se que as variáveis estão correlacionadas moderadamente. O maior valor pode ser visto para a correlação entre as variáveis: *NDis* e *Fl*. Isso é esperado, uma vez que maiores taxas de aprovação resultam em menor distorção idade-série.

A análise de correlação justifica o uso da análise de componentes principais que passamos a descrever. As duas primeiras componentes principais obtidas: CP_1 e CP_2 explicam 84,77% da variabilidade total dos dados. Suas equações são apresentadas abaixo:

$$CP_1 = 0,759 \cdot Port + 0,554 \cdot Mat + 0,219 \cdot Fl + 0,261 \cdot NDis,$$

$$CP_2 = 0,205 \cdot Port - 0,678 \cdot Mat + 0,636 \cdot Fl + 0,306 \cdot NDis.$$

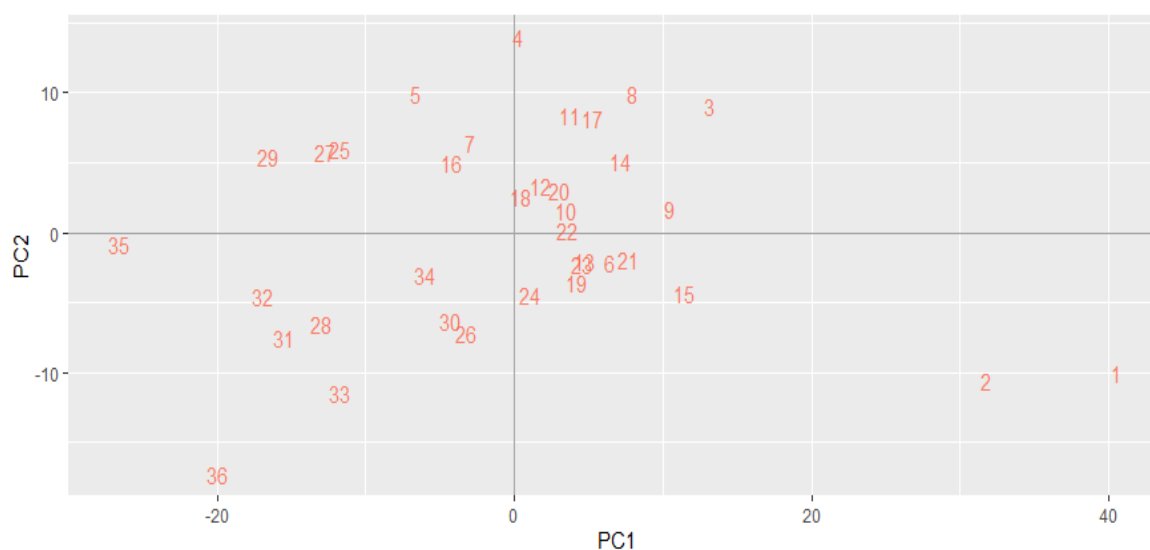
Observou-se que a CP_1 retém 65,61% da informação original, e com maiores cargas para as variáveis que indicam os percentuais de alunos com aprendizado adequado nas disciplinas: Língua Portuguesa e Matemática no Saeb de 2019. Pode-se dizer que esta componente fornece um índice de desempenho geral das escolas, de tal modo que taxas maiores para as variáveis em estudo, resultam valores maiores para a CP_1 .

A CP_2 é responsável pela retenção de 19,16% da informação dos dados. Na sua equação são notórios os maiores pesos para as variáveis: *Mat* e *Fl*, com sinais opostos. Baseados nessa equação, pode-se dizer que esta componente fornece um indicativo do desempenho na disciplina de matemática no Saeb 2019, ou seja, valores maiores para esta componente é um indicativo de aprendizado em matemática baixo.

A Figura 1 apresenta o gráfico contendo os *escores* das duas componentes principais referentes às trinta e seis escolas cuja numeração segue a disposição do Quadro 1. Nota-se que as escolas 1, 2 e 3, respectivamente: Medalha Milagrosa, Prof. Júlio Teodorico e De Vila Velha possuem os maiores valores para a CP_1 . Coincidentemente, são as escolas com os maiores valores do Ideb. No sentido contrário, as escolas Prof. José do Amaral (35) e Prof. Becker e Silva (36) apresentaram os menores valores para este componente, contrastando com seus desempenhos no Ideb.

Figura 1 - Gráfico dos escores das escolas

Fonte: Dados da pesquisa (2025).

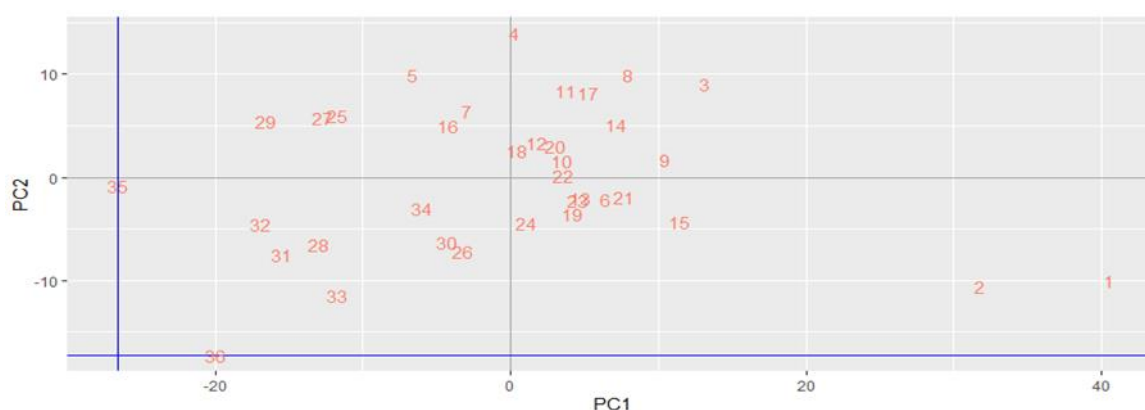


Ainda na Figura 1, as posições das observações em termos do eixo vertical mostram que as escolas Profa Margarete Mazur (4) e Francisco P. Machado (5) obtiveram os maiores valores para a CP_2 , demonstrando um baixo desempenho relacionado com a disciplina de Matemática, como é possível observar no Quadro 1.

Com o intuito de construir um índice de desempenho escolar com base nas duas primeiras componentes principais, utilizou-se a metodologia descrita em (Ramos *et al.*, 2021).

A construção está alicerçada na translação dos eixos coordenados da Figura 1 de tal modo que a nova origem coincida com $O' = (-26.63504; -17.24465)$, em que a primeira coordenada corresponde ao menor escore com relação à CP_1 e a segunda coordenada ao mínimo dos escores referentes à CP_2 . A Figura 2 ilustra a inserção dos novos eixos:

Figura 2 - Obtenção dos novos eixos para definição do desempenho escolar



Fonte: Dados da pesquisa (2025).

Observa-se que a escola com pior escore referente à CP_1 é a Prof. José do Amaral (35), enquanto a escola com pior desempenho para a CP_2 é a Prof. Becker e Silva (36). Uma vez instituídos os novos eixos, nota-se que as escolas mais afastadas de O' apresentam maiores escores para as duas componentes. Dessa forma, utilizamos a distância Euclidiana destes pontos em relação a nova origem O' como indicadora do desempenho das escolas com base nas variáveis consideradas nesse estudo. Essa métrica será designada por D_i para a i -ésima escola. Por exemplo, as escolas: Medalha Milagrosa (1) e Prof. Júlio Teodorico (2) são tais que $D_1 = 67,5$ e $D_2 = 58,74$, obtendo os melhores desempenhos desta análise. Por outro lado, Prof. Becker e Silva tem desempenho dado por $D_{36} = 6,67$, sendo o pior dentre todas as escolas participantes. Os valores obtidos para D_i são apresentados no Quadro 4:

Quadro 4 - Ordenamento das escolas via métrica D_i

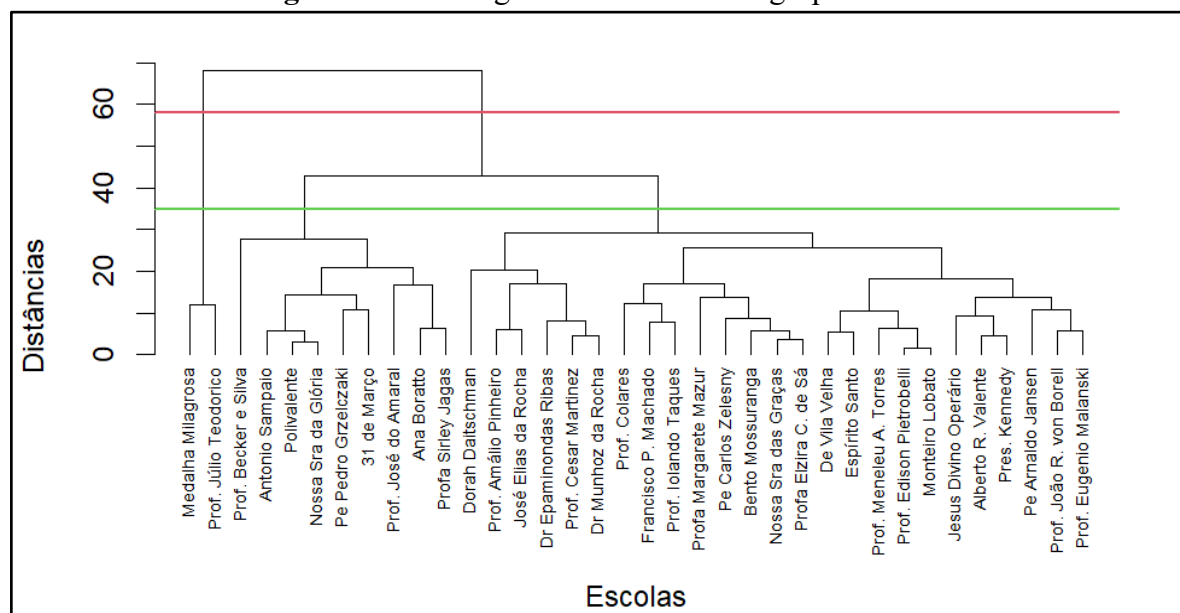
Posição (Ideb)	Escola	Port	Mat	Fl	NDis	D_i	Posição (D_i)
1	Medalha Milagrosa	65	55	95	95,5	67,50	1
2	Prof. Júlio Teodorico	63	47	89	89,2	58,74	2
3	De Vila Velha	52	24	98	91,8	47,67	3
4	Profa. Margarete Mazur	44	13	98	88,9	41,17	6
5	Francisco P. Machado	37	13	96	84,7	33,77	19
6	Alberto R. Valente	39	32	94	90,2	36,35	12

7	Pe Carlos Zelesny	35	20	96	89,8	33,45	22
8	Espírito Santo	49	20	97	90,1	44,04	4
9	Jesus Divino Operário	47	28	92	92,5	41,62	5
10	Nossa Sra. das Graças	41	25	92	89,7	35,54	14
11	Prof. Edison Pietrobelli	47	17	92	90,1	39,74	10
12	Profa. Elzira C. de Sá	40	23	93	92,5	35,14	15
13	Prof. João R. von Borell	43	27	89	87,1	34,90	17
14	Prof. Meneleu A. Torres	50	21	92	86,1	40,50	8
15	Pe. Arnaldo Jansen	50	31	89	83,8	40,21	9
16	Prof. Colares	40	17	95	77,5	31,58	23
17	Monteiro Lobato	48	18	92	90,9	40,77	7
18	Bento Mossuranga	41	21	90	88,2	33,57	21
19	Pres. Kennedy	39	30	90	89,3	33,72	20
20	Dorah Daitschman	55	14	84	77,3	35,94	13
21	Dr. Epaminondas Ribas	53	22	81	86,4	37,57	11
22	Prof. Cesar Martinez	48	20	83	88	34,91	16
23	Dr. Munhoz da Rocha	46	24	84	88,1	34,60	18
24	Prof. Eugenio Malanski	42	25	85	83,6	30,53	24
25	Prof. Iolando Taques	34	11	89	84	27,64	25
26	Prof. Amálio Pinheiro	38	25	84	79,4	25,44	27
27	Ana Boratto	28	14	91	89,5	26,84	26
28	Polivalente	28	21	84	79,6	17,35	31
29	Profa. Sirley Jagas	28	10	88	85,8	24,77	30
30	José Elias da Rocha	41	21	81	77,6	24,83	29
31	Nossa Sra da Glória	26	20	82	79,6	14,78	35
32	Antonio Sampaio	24	19	87	77,8	15,92	34
33	Pe Pedro Grzelczaki	27	25	79	83,2	15,99	33
34	31 de Março	35	21	84	86,2	25,08	28
35	Prof. José do Amaral	20	8	81	80,8	16,38	32
36	Prof. Becker e Silva	26	20	70	72,7	6,67	36

Fonte: Dados da pesquisa (2025).

As três primeiras escolas em relação ao Ideb continuam com as mesmas posições em relação ao indicador D_i , o mesmo para a escola Prof. Becker e Silva (36) que apresentou o pior índice. Em geral, percebe-se que D_i ordena as escolas considerando as maiores taxas de aprendizado nas disciplinas básicas.

Na etapa da estatística multivariada de análise de agrupamento, foi construído o dendrograma que está apresentado na Figura 3. Esse agrupamento foi validado com uma correlação cofenética de 80,2%, ou seja, o agrupamento possui uma fidelidade com a estrutura real dos dados de 80,2%.

Figura 3 - Dendrograma da análise de agrupamento

Fonte: Dados da pesquisa (2025).

Observando este gráfico, é possível verificar a possível formação de 2 ou de 3 grupos de escolas. A formação de 2 grupos está indicada com a linha vermelha do gráfico e a formação de 3 grupos pela linha verde. Considerando a formação de apenas dois grupos, pode-se identificar um grupo formado pelas escolas Medalha Milagrosa e Prof. Júlio Teodorico, que são as duas maiores notas do Ideb, e um segundo grupo composto pelas demais escolas.

Agora considerando a formação de três grupos, teríamos novamente um grupo formado pelas duas escolas com as maiores notas do Ideb (Medalha Milagrosa e Prof. Júlio Teodorico), um segundo grupo formado por 9 escolas (Ver Quadro 5) e um terceiro grupo composto pelas escolas restantes.

Quadro 5 - Composição dos 3 grupos da análise de agrupamento

Grupo 1	Grupo 2	Grupo 3
1 Medalha Milagrosa 2 Prof. Júlio Teodorico	27 Ana Boratto 28 Polivalente 29 Profa. Sirley Jagas 31 Nossa Sra. da Glória 32 Antonio Sampaio 33 Pe. Pedro Grzelczaki 34 31 de Março 35 Prof. José do Amaral 36 Prof. Becker e Silva	Demais escolas

Fonte: Dados da pesquisa (2025).

Com essa formação é possível identificar que o agrupamento preservou a classificação do Ideb, sendo o grupo 1 formado pelas duas escolas com maiores notas, o grupo 2 formado pelas 9 escolas com as menores notas (exceto a escola José Elias da Rocha) e o grupo 3 com as escolas que conseguiram notas intermediárias no Ideb.

CONSIDERAÇÕES FINAIS

O estudo revelou a importância das taxas de aprendizado adequado das disciplinas de Língua Portuguesa e Matemática no desempenho das escolas participantes do Saeb 2019. No caso da CP_2 , a influência do aprendizado em Matemática é intensificada, e este fato é considerado no indicador D_i construído pelos autores. Para exemplificar, vê-se que a escola Espírito Santo na 8ª posição quanto ao Ideb passa para a 4ª posição com o uso do indicador proposto, justamente devido às taxas destacáveis de aprendizado nas disciplinas básicas.

O indicador aqui proposto, assim como outros existentes, revela o grande desafio para a educação brasileira: a melhoria da qualidade da aprendizagem nas disciplinas de Língua Portuguesa e Matemática.

Exigir taxas de aprovação alta e de distorção idade-série baixa não são suficientes para garantir aprendizagem de fato. Devem-se investir em políticas públicas que promovam a qualidade do ensino, como a formação de professores da Educação Básica, o planejamento e a gestão escolar eficientes, entre outras.

REFERÊNCIAS

BRASIL. Instituto Brasileiro de Geografia e Estatística (IBGE). **Censo Brasileiro de 2022**. Rio de Janeiro: IBGE, 2022. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/pr/ponta-grossa.html>. Acesso em: 14 set. 2025.

BRASIL. Lei n.º 9394/96. **Lei de Diretrizes e Bases da Educação Nacional**. Estabelece as diretrizes e bases da educação nacional. Brasília: 1996. Disponível em: https://www.planalto.gov.br/ccivil_03/Leis/L9394.htm. Acesso em: 14 set. 2025.

FERREIRA, D. F. **Estatística multivariada**. Lavras: Editora UFLA, 2008.

HAIR JR., J. F.; WILLIAM, B.; BABIN, B.; ANDERSON, R. E. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

ORTIGÃO, M. I. R. Avaliação e Políticas Públicas: possibilidades e desafios para a Educação Matemática. **Bolema**, São Paulo, v. 21, n. 29, p. 71-98, 2008. Disponível em: <https://www.periodicos.rc.biblioteca.unesp.br/index.php/bolema/article/view/1719>. Acesso em: 14 set. 2025.

PINTO, J. M. R. O custo de uma educação de qualidade. In: CORREA, B. C.; GARCIA, T. O. (Orgs.). **Política educacionais e organização do trabalho na escola**. São Paulo: Xamã, 2008.

QEdU. **Use dados. Transforme a educação**. 2020. Disponível em: <https://novo.qedu.org.br/>. Acesso em: 06 set. 2025.

RAMOS, M. N.; COSTA FILHO, A. Uma análise do aprendizado escolar nas escolas municipais de Ribeirão Preto. **Nova Escola**, São Paulo, maio 2020. Disponível em: <https://nova-escola-producao.s3.amazonaws.com/myxvP9uy9xmdsgRSEsJUgjY7qRTMTFtvYrxzcphzzFW59uFVDdRJDkVK6GEr/mozart-ramos-uma-analise-do-aprendizado-escolar-nas-escolas-municipais-de-ribeirao-preto.pdf>. Acesso em: 06 mar. 2022.

RAMOS, M. N. et al. Uma análise estatística multivariada do desempenho das escolas municipais de Ribeirão Preto. Ensaio: Avaliação e Políticas Públicas em Educação, [S.l.], v. 29, n. 113, p. 857-873, oct. 2021. ISSN 1809-4465. Disponível em: <https://revistas.cesgranrio.org.br/index.php/ensaio/article/view/3286> Acesso em: 06 set. 2025.

R CORE TEAM R. **A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2025. Disponível em: <https://www.R-project.org/> Acesso em: 06 set. 2025.

SANTOS, M. A. B. dos; FERREIRA, H. S.; SIMÕES, L. L. F. Formação de professor e profissionalismo: Reflexões acerca da avaliação externa. **Educ. Form.**, [S. l.], v. 4, n. 11, p. 161–178, 2019. DOI: 10.25053/redufor.v4i11.896. Disponível em: <https://revistas.uece.br/index.php/redufor/article/view/896>. Acesso em: 15 set. 2025.

VIEIRA, G. M.; VIEIRA, V. M. de O. Avaliação educacional: ressignificação e reconstrução da prática. **Práticas Educativas, Memórias e Oralidades - Rev. Pemo**, [S. l.], v. 6, p. e12529, 2024. DOI: 10.47149/pemo.v6.e12529. Disponível em: <https://revistas.uece.br/index.php/revpemo/article/view/12529>. Acesso em: 15 set. 2025.

Histórico Editorial

Submetido: 20 de outubro de 2025.

Publicado: 03 de dezembro de 2025.

Minicurrículo

Marcos Teixeira Alves

Doutor em Matemática Aplicada pela Universidade Federal do Paraná (UFPR). Licenciado em Matemática pela Universidade Federal de Santa Catarina (UFSC). Professor da Universidade Estadual de Ponta Grossa (UEPG).

Contribuição de autoria: Desenvolvimento do trabalho, análises estatísticas e correções do texto.

Lattes: <http://lattes.cnpq.br/5135226249313330>.

Scheila Valechenski Biehl

Doutora em Engenharia Elétrica pela Universidade de São Paulo (USP). Mestre em Matemática Computacional pelo Instituto de Ciências Matemáticas e Computacional (ICMC). Professora Associada Universidade Estadual de Ponta Grossa (UEPG). Contribuição de autoria: Desenvolvimento do trabalho e correções do texto.

Lattes: <http://lattes.cnpq.br/2491113026629291>.

Deyse Márcia Pacheco Gebert

Doutora em Estatística e Experimentação Agropecuária pela Universidade Federal de Lavras (UFLA). Graduada em Agronomia pela Universidade do Estado de Santa Catarina (UDESC). Professora da Universidade Estadual de Ponta Grossa (UEPG).

Contribuição de autoria: Desenvolvimento do trabalho, análises estatísticas e correções do texto.

Lattes: <http://lattes.cnpq.br/7731588187059070>.

COMO REFERENCIAR - ABNT

ALVES, Marcos Teixeira; BIEHL, Scheila Valechenski; GEBERT, Deyse Márcia Pacheco. Estatística multivariada aplicada no desempenho escolar de Ponta Grossa-PR. **Revista Exitus**, Santarém/PA, v. 15, e025059, n. 1, 2025. <https://doi.org/10.24065/re.v15i1.2981>

COMO REFERENCIAR - APA

Alves, M. T., Biehl, S. V. & Gebert, D. M P. (2025). Estatística multivariada aplicada no desempenho escolar de Ponta Grossa-PR. *Revista Exitus*, 15, e025059. <https://doi.org/10.24065/re.v15i1.2981>

Licença de Uso

Licenciado sob a Licença Creative Commons Atribuição 4.0 Internacional (CC BY 4.0). Esta licença permite compartilhar, copiar, redistribuir o manuscrito em qualquer meio ou formato. Além disso, permite adaptar, remixar, transformar e construir sobre o material, desde que seja atribuído o devido crédito de autoria e publicação inicial nesta revista.