

## DIAGNÓSTICO DA QUALIDADE DA EDUCAÇÃO BÁSICA BRASILEIRA NAS REDES MUNICIPAIS

## DIAGNOSIS OF THE QUALITY OF BRAZILIAN BASIC EDUCATION IN MUNICIPAL NETWORKS

**Pedro Henrique Alves Barros**

Mestre em Matemática  
Universidade do Estado do Rio de Janeiro (UERJ)  
E-mail: pedro.barros@pos.ime.uerj.br

**Regina Serrão Lanzillotti**

Doutora em Engenharia de Transportes  
Coordenação dos Programas de Pós-Graduação e Pesquisa de Engenharia (COPPE - UFRJ)  
E-mail: reginalanzillotti@ime.uerj.br

**RESUMO:** A motivação pontua a educação básica brasileira nas redes municipais, Saeb (2021). Objetiva-se verificar a proficiência dos valores médios em Matemática, Língua Portuguesa e Taxa de Alfabetização, entre estados nas escolas municipais e privadas, para capital e interior. Calculou-se a média, moda e mediana para avaliar simetria, curtose e coeficiente de variação, tendo o histograma para visualização gráfica. As covariâncias para pares de disciplinas foram obtidas em relação às autarquias e regionalidades tanto na capital quanto interior. O teste de igualdade de covariância aferiu a similaridade no desempenho das disciplinas. O gráfico de dispersão entre Distâncias Euclidiana e Mahalanobis para a Taxa de Alfabetização por estado, mostrou sensibilidade que as regiões Sudeste e Centro Oeste.

**Palavras-chave:** Teste de Covariância. Distância de Mahalanobis. Distância Euclidiana. Estatística. Desempenho.

**ABSTRACT:** Motivation punctuates Brazilian basic education in municipal networks, Saeb (2021). The objective is to verify the proficiency of the average values in Mathematics, Portuguese Language and Literacy Rate, between states in municipal and private schools, for the capital and interior. The mean, mode and median were calculated to evaluate symmetry, kurtosis and coefficient of variation, with the histogram for graphic visualization. The covariances for pairs of disciplines were obtained in relation to municipalities and regionalities both in the capital and in the interior. The equality of covariance test measured the similarity in the performance of the subjects. The scatterplot between Euclidean and Mahalanobis Distances for the Literacy Rate by state showed sensitivity than the Southeast and Central West regions.

**Keywords:** Covariance Test. Distance from Mahalanobis. Euclidean Distance. Statistic. Performance.

## 1 INTRODUÇÃO

A educação básica é a primeira etapa na formação do indivíduo. Esta etapa possui divisões, cada qual com seus objetivos e metodologias, que estão sempre sendo avaliados e reavaliados-para acompanhar as mudanças vivenciadas pelas gerações estudantis. A educação possui um papel central no projeto de nação, pois “[...] sozinha não transforma a sociedade, sem ela tampouco a sociedade muda” (Freire, 2000), porém também atua constituindo “[...] relações sociais mais amplas, contribuindo, contraditoriamente, desse modo, para a transformação e a manutenção dessas relações.” O ensino básico brasileiro precisa ser constantemente analisado, para que haja oferta de um ensino de qualidade. Interpretar os diferentes níveis de aprendizagem no Ensino Fundamental em países de dimensões continentais, território brasileiro, torna-se desafiante. Estima-se que sejam 178 mil escolas que agregam mais de 47 milhões de matrículas, o que se configura em tarefa complexa, porém necessária, pois os resultados propiciam direcionar os investimentos para regiões de maior demanda. Uma das alternativas de diagnosticar a educação básica brasileira tem sido as avaliações externas em larga escala. O modelo avaliativo em questão, surgiu no Brasil na década de 1990, objetivando aferir a qualidade do ensino, sendo elaborada e realizada pelo Ministério da Educação. Esse é um dos meios utilizados como ferramenta para mapear os aspectos positivos e negativos, indicando quais as possíveis dificuldades encontradas, principalmente, na alfabetização, medindo assim os resultados obtidos por grupos de estudantes, fornecendo insumos para o constante monitoramento, planejamento, orientação e intervenção no processo educativo. Luckesi cita que:

Avaliação pode ser caracterizada como uma forma de ajuizamento da qualidade do objeto avaliado, fator que implica uma tomada de posição a respeito do mesmo, para aceitá-lo ou para transformá-lo. A avaliação é um julgamento de valor sobre manifestações relevantes da realidade, tendo em vista uma tomada de decisão. (LUCKESI, 2000, p. 33)

Foram surgindo a partir daí outras avaliações externas como a Avaliação Nacional de Alfabetização (ANA), o Exame Nacional do Ensino Médio (Enem) e o Exame Nacional de Desempenho de Estudantes (Enade).

Uma das primeiras avaliações externas na educação foi o Sistema de Avaliação da Educação Básica (Saeb), que tinha o objetivo de aferir a proficiência estudantil nas séries iniciais em Língua Portuguesa e Matemática. O exame constitui-se em instrumento de valor na tomada de decisões que visem alcançar melhor qualidade de ensino, através desses resultados. o Instituto Nacional de Pesquisas e Estudos (Inep) cita que:

O Saeb permite que as escolas e as redes municipais e estaduais de ensino avaliem a qualidade da educação oferecida aos estudantes. O resultado da avaliação é um indicativo da qualidade do ensino brasileiro e oferece subsídios para a elaboração, o monitoramento e o aprimoramento de políticas educacionais com base em evidências. (INEP/MEC, 2019)

Desde 1990, o Saeb vem passando por alterações metodológicas, pois o educar vem sofrendo transformações pedagógicas e administrativas, sendo a mais recente a implementação da Base Nacional

Comum Curricular (BNCC), um dos principais documentos que versa sobre a educação básica brasileira. A avaliação censitária do Saeb mapeia a educação em sete dimensões: atendimento escolar, ensino e aprendizagem, investimento, profissionais da educação, gestão, equidade, direitos humanos. Em 2007, o Inep determinou o Índice de Educação Básica que associa as notas dos estudantes no Saeb, com as taxas de aprovação, reprovação e abandono, oriundas do Censo Escolar.

Em constantes melhorias, principalmente a partir de 2019, ocorre uma reestruturação do exame para se adequar a BNCC, pois foram inseridas as matrizes de Língua Portuguesa e Matemática no 2º ano do Ensino Fundamental, enquanto Ciências Humanas e Ciências da Natureza para o 9º ano e deste segmento escolar, as quais a partir de 2023, estas disciplinas foram incluídas a partir do 5º ano. O exame de 2021 foi considerado um estudo piloto, exame com publicação mais recente de resultados, disponível pelo Inep, que discriminam as taxas de alfabetização, Linguagem e Matemática a nível das escolas públicas municipais, estaduais, federais e privadas, também pela agregação estadual e nacional. As notas de Matemática e Língua Portuguesa e das demais disciplinas são determinadas pelas escalas de proficiência pré-estipuladas para exame, portanto há uma associação em relação a escala de desempenho em consonância com as habilidades e competências presentes na BNCC, consideradas adquiridas ou desenvolvidas. A taxa de alfabetização corresponde ao índice relativo da proporção de estudantes que obtiveram um nível satisfatório de alfabetização de acordo com a Avaliação Nacional da Alfabetização (ANA), tal exame é realizado anualmente para aferir o nível de alfabetização e letramento em Língua Portuguesa e alfabetização em Matemática das crianças regularmente matriculadas no 3º ano do ensino fundamental.

Cabe ainda acrescentar que segundo Creso Franco:

[...] estudos que testam apenas uma vez os alunos apresentam importantes limitações para a explicação do efeito das escolas na aprendizagem. A razão disso reside no fato de que a medida de proficiência é uma medida do aprendizado dos alunos ao longo de muitos anos. Portanto, essa medida não pode ser explicada em função do passado recente dos alunos (variáveis da escola que o aluno frequentou no ano do SAEB). É por essa razão que os estudos longitudinais são considerados o "padrão ouro" dos estudos de avaliação de escolas e de sistemas educacionais. Nos estudos longitudinais, a medida de proficiência prévia é usada para filtrar o efeito do aprendizado prévio do aluno. Deste modo, a segunda proficiência mensurada (após a filtragem do aprendizado prévio) representa o aprendizado ao longo do ano. Por certo, esse aprendizado ao longo do ano pode ser explicado pelas características da escola frequentada. É essa a principal limitação do SAEB para a explicação dos fatores que influenciam o aprendizado. (Creso Franco, 2001)

Ainda assim, acredita-se que a análise global de informações pode ser reveladora referente ao desempenho estudantil, conforme Alavarase indica:

Conhecer e utilizar os resultados das avaliações externas nas salas de aula e cotejá-los com as avaliações internas significa compreendê-los não como um fim em si mesmo, mas sim como possibilidade de associá-los às transformações necessárias no sentido de fortalecer a qualidade da escola pública democrática, que é aquela que se organiza para garantir a aprendizagem de todos e todas (ALAVARASE, 2013, p. 75).

Em tempos de pós-pandemia, uma análise do cenário revelado em 2021, o mais recente, pode ser demonstrativo e indicativo para futuras políticas de combate ao analfabetismo, pois, pois devido a

necessidade das escolas precisarem manterem-se fechadas por conta das questões sanitárias, verificou-se que muitos estudantes, ainda que com esforços das respectivas autarquias de ensino, ficaram lesados no processo de ensino e aprendizagem naquele período, carregando uma lacuna que irá perdurar nos anos escolares seguintes. Esta proposta visa mapear a divergência das realidades do ensino público e do particular em termos de desempenho estudantil no Ensino Fundamental a nível estadual, utilizando como banco de dados os resultados fornecidos pelo Inep do Saeb 2021, ano considerado o mais letal da Covid-19 no Brasil. Naquele momento crítico, em que o país terminou o ano com 619.109 mortes por conta da pandemia, as medidas sanitárias restritivas determinavam um ensino, inicialmente remoto, e, posteriormente, híbrido. O artigo 80 da LDB assegura:

[...] o Poder Público incentivará o desenvolvimento e a veiculação de programas de ensino a distância, em todos os níveis e modalidades de ensino, e de educação continuada”, cujo § 1º define que “a educação a distância, organizada com abertura e regime especiais, será oferecida por instituições especificamente credenciadas pela União (Brasil, 1996).

A Educação à Distância (EaD), regulamentada atualmente pelo Decreto nº 9.057/2017, é aplicável tanto a cursos de Educação Superior, quanto na Educação Básica. Porém, em 2021, durante a pandemia, a sociedade mostrou-se pouco confiante neste método de ensino para a educação básica, principalmente por conta da imensa desigualdade social existente no país, onde parcela considerável da população sequer tinha acesso a utensílios básicos de higiene pessoal. Cunha e Ferrari citam:

[...] é importante destacar que a pandemia apenas acentuou vulnerabilidades que já estavam presentes no tecido social brasileiro. Pelo menos desde 2014, embora com mais ênfase a partir de 2016, passamos não apenas pelo esgotamento do ciclo econômico expansivo (2004 a 2014) como, principalmente, pela recessão econômica (2014 a 2016) e pela estagnação econômica (2016 até os dias atuais), já definida por alguns autores como “estagflação” (Cunha; Ferrari, 2021).

Em razão deste cenário, pode-se perceber que a pandemia exerceu influência desproporcional nas variadas camadas sociais, exercendo maior impacto, sobretudo, nas classes menos favorecidas, e mais dependentes dos serviços públicos, sobretudo na oferta da educação pública. Este estudo, é, portanto, também uma maneira de verificar a eficácia no cumprimento do artigo 206 da Constituição Federal de 1988, que garante a Educação dever do Estado e da família, e direito de todos.

## 1.1 OBJETIVOS

O objetivo geral deste trabalho é diagnosticar o nível de proficiência dos estudantes na estrutura geográfica tanto intra quanto inter a nível estadual no tocante as disciplinas de Matemática, Língua Portuguesa e Alfabetização nas instituições públicas e privadas através dos resultados do Saeb 2021, o mais recente divulgado pelo Inep no momento da finalização deste trabalho.

A especificidade concentra-se no parágrafo único do artigo 211 da Constituição Federal, que dita que a Prefeitura dos municípios atuará prioritariamente no Ensino Fundamental e na Educação Infantil,

enquanto as escolas de Ensino Médio sob a gestão dos Estados da federação. Esta taxionomia do ensino público foi confrontada tanto para as escolas privadas da capital, quanto as do interior. ao das escolas privadas tanto da capital quanto do interior. Desta forma, será possível mensurar as diferenças nas aprendizagens geograficamente e de acordo com a autarquia, que reflete o grau de dissimilaridade na aquisição de conhecimento adquirido para oportunizar conscientemente os destinos dos investimentos do Ministério da Educação.

## 2 REFERENCIAL TEÓRICO

Inicialmente foi feita uma revisão da literatura desenvolvida com artigos publicados no período de 2017 a 2024 nas bases eletrônicas: *Scientific Electronic Library Online* - Scielo e Google Acadêmico, empregando os descritores: Saeb. Acrescentou-se uma pesquisa na base de dados do Inep acerca do Saeb e suas propostas, abordagens, histórico de alterações e resultados publicados. Foi realizada também uma consulta a avaliações anteriores de Matemática e Língua Portuguesa para verificação do modelo de avaliação aplicado, além de uma busca por avaliações de alfabetização e critérios de alfabetização estipulados pela ANA.

## 3 METODOLOGIA

Os resultados mais recentes do Saeb realizado em 2021 constam no site do Inep. No momento, os dados de 2023 ainda não foram divulgados, serão em fase preliminar. Este Sistema de Informações Gerenciais - SIG subsidiou o estudo estatístico descritivo e inferencial para os quantitativos referentes aos anos da pandemia mundial, momento crítico em que muitos estudantes estavam afastados das escolas, apesar do ensino remoto ter sido ofertado em muitas redes municipais, mas poucos estudantes tiveram acesso ao uso de tecnologia assistiva da tecnologia e educação.

No site do Inep constam os valores médios da proficiência referentes ao ensino privado da capital, mas observou-se a ausência para estado do Ceará. Houve uma tentativa de contato com o Inep, porém sem sucesso. Portanto, o estado foi desconsiderado neste estudo. A disponibilidade dos desempenhos médios em Língua Portuguesa, Matemática e Alfabetização foram verificadas, porém os órgãos competentes não incluíram aferição da variabilidade em torno deste promédio.

Inicialmente foram criados os histogramas das médias disponibilizadas que permitiram uma visualização do perfil dos alunos em relação ao aprendizado. As medidas de posição média, mediana, moda foram obtidas para avaliar a ocorrência ou não da heterogeneidade dos dados divulgados, sendo que em uma distribuição pode ser considerada simétrica, se estas medidas forem iguais (Correa, 2002), mas a distribuição assimétrica é positiva, se o valor da média é maior que a moda, enquanto na negativa, a moda é maior que a média. A medida de dispersão, coeficiente de variação (CV), pode ser obtido pela razão entre o desvio padrão e a média aritmética, sendo o desvio padrão, raiz quadrada da variância.

$$CV = \frac{S}{\bar{x}}, \quad (1)$$

onde S = desvio padrão

$\bar{x}$  = média aritmética.

Considera-se baixa dispersão quando os valores de CV forem próximos de zero, pois o aumento desta medida, infere na expressividade da variação em torno do valor médio. O coeficiente percentílico de curtose indica o grau de achatamento do polígono de frequência do histograma, aferido pela relação entre a distância média do distanciamento dos quartis 3 e 1, em relação a distância dos Percentis extremos, percentil 90 e percentil 10. (Naghetini; Pinto, 2007). A equação está expressa a seguir:

$$K = \frac{(Q_3 - Q_1)/2}{(P_{90} - P_{10})} \text{ onde} \quad (2)$$

$Q_3$ , 3ª quartil,  $Q_1$ , 1ª quartil,  $P_{90}$ , percentil 90,  $P_{10}$ , percentil 10.

A Distribuição de Probabilidade Normal apresenta coeficiente percentílico de curtose correspondente a 0,263, quando se classifica distribuição mesocúrtica. Se o valor for maior, designa-se como platicúrtica, senão, leptocúrtica.

A análise de agrupamento é uma técnica multivariada que tem por objetivo proporcionar uma ou várias partições na massa de dados, em grupos, por algum critério de classificação, de tal forma que exista homogeneidade dentro e heterogeneidade entre grupos (Sneath & Sokal, 1973; Mardia et al., 1997) O teste de covariância foi realizado para verificar a igualdade nas covariâncias dos desempenhos em Matemática, Língua Portuguesa e Taxa de Alfabetização no confronto do ensino ofertado na capital e no interior, tanto a nível privado e quanto público, a nível municipal de cada estado da federação. A metodologia testa a hipótese nula  $H_0$  da igualdade das covariâncias.

$$H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_k, \quad (3)$$

onde  $\Sigma_1, \Sigma_2, \dots, \Sigma_k$  representam as  $k$  matrizes de covariâncias de um sistema populacional  $p$ -dimensional multinormal. Seja  $S_i$  uma estimativa não tendenciosa amostral de  $\Sigma_i$ , baseada em  $n_i$  graus de liberdade, onde  $n_i = N_i - 1$  para o caso de uma amostra aleatória com  $N_i$  vetores observados da  $i$  – ésima população. Se  $H_0$  for verdadeira, tem-se:

$$S = \frac{1}{\sum n_i} \sum_{i=1}^k n_i S_i, \quad (4)$$

onde S é a estimativa agrupada das matrizes de covariâncias. A estatística teste:

$$M = \sum n_i \ln|S| - \sum_{i=1}^k n_i \ln S_i \quad (5)$$

Box (1949) mostrou que se o fator de escala

$$C^{-1} = 1 - \frac{2p^2+3p-1}{6(p+1)(k-1)} \left( \sum_{i=1}^k \frac{1}{n_i} - \frac{1}{\sum n_i} \right) \quad (6)$$

for introduzido, a quantidade  $MC^{-1}$  apresenta uma distribuição que pode ser aproximada da Chi-quadrado ( $\chi$ ) com  $\alpha = 1/2 (k - 1)p(p + 1)$  graus de liberdade, sendo que na medida em que se aumenta o valor de  $n_i$ , tamanhos amostrais dos grupos, se todos os  $n_i$  forem iguais a  $n$ , tem-se:

$$C^{-1} = 1 - \frac{(2p^2+3p-1)(k+1)}{6(p+1)kn} \quad (7)$$

Se,  $MC^{-1} < \chi_{\alpha(\text{tabelado})}$  aceita-se a veracidade de  $H_0$ .

Em grupamento multivariado referente a cada Unidade da Federação foram realizadas medidas de similaridade, com o objetivo de diagnosticar a intensidade das discrepâncias interestaduais no desempenho estudantil dentro das regiões Norte, Sul, Nordeste, Sudeste e Centro-Oeste. Nos agrupamentos é possível fazer uma análise exploratória de similaridade através do cálculo de distâncias. Com base nessa medida os indivíduos similares são agrupados e os demais são colocados em grupos separados (Aaker et al., 2001). Diferentes escalas de medidas de grandeza podem distorcer a realidade, sendo que a forma mais adequada para verificação é o uso de medidas padronizadas em unidades do desvio padrão, a qual é explicitada a seguir:

$$z_i = \frac{x_i - \bar{x}}{s}, \quad (8)$$

onde  $z_i$  é o valor observado de  $x_i$  na escala de razão,  $\bar{x}$  é a média aritmética,  $s$  o desvio padrão amostral.

A maioria dos métodos de análise de grupos, requer a análise das relações entre eles. Segundo Barroso & Artes (2003) há dois tipos de parença: as medidas de similaridade (diretamente proporcionais à semelhança entre objetos) e as medidas de dissimilaridade (inversamente proporcional a semelhança entre objetos). Pode-se definir similaridade como “a medida de correspondência, ou semelhança, entre os objetos a serem agrupados.” (CORRAR; PAULO e DIAS FILHO, 2007, p.333) Tais medidas são facilmente interrelacionáveis e transformáveis entre si (Bussab et al., 1990). Normalmente, a medida de similaridade ou dissimilaridade é expressa como uma função distância ou métrica (Doni, 2004). Segundo Khattree e Naik (2000) e Cruz e Carneiro (2006), dentre as medidas de dissimilaridades, a distância Euclidiana e a distância de Mahalanobis estão em destaque, devido a sua maior utilização.

A formulação da Distância Euclidiana ( $D_E$ ) é baseada puramente em pontos espaciais, e desconsidera a variabilidade entre os dados, é demonstrada pela aplicação do Teorema de Pitágoras em relação a distância, pois o espaço euclidiano torna-se um espaço métrico.

$$D_E(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (9)$$

A adoção da Distância Generalizada de Mahalanobis ( $D_m$ ), distância introduzida pelo matemático indiano Prasanta Chandra Mahalanobis em 1936, baseia-se nas variáveis expressas em unidades do desvio padrão ( $z$ ) e a inversa da matriz de covariância  $S^{-1}$  para identificação de distintos padrões, tendo a formulação estimada segundo (Rao, 1952) por:

$$D_m(z) = \sqrt{z^T \cdot S^{-1} \cdot z}, \quad (10)$$

onde  $S^{-1}$  é a inversa da matriz de covariância.

A distância de Mahalanobis considera, portanto, a variabilidade dentro de cada unidade amostral, e não somente a tendência central, sendo uma medida mais aceitável para análise grupos (Riboldi, 1986).

#### 4 RESULTADOS E DISCUSSÕES

A análise do desempenho estudantil para o Saeb 2021 iniciou-se com a obtenção da média e mediana das taxas de alfabetização, Língua Portuguesa e Matemática nas instituições de ensino públicas e privadas da capital e interior, com exceção do Ceará, discriminando desvio padrão e coeficiente de variação. Os resultados obtidos para a Taxa de Alfabetização, Língua Portuguesa e Matemática, tabela 1, indicaram que os melhores resultados se encontram na capital para o ensino privado, tendo médias entre 76,63, 772,34 e 772,60 com respectivos coeficientes de variação 0,26, 0,01 e 0,02. Os valores médios menos expressivos foram encontrados na capital nas escolas municipais para as disciplinas em análise, 28,03 para Taxa de Alfabetização, 704,04 para Língua Portuguesa e 724,78 para Matemática. Os maiores coeficientes de variação para as três disciplinas estão registrados nas escolas públicas municipais do interior, o que vem diminuir a fidedignidade dos valores médios.

Tabela 1: Média, mediana e coeficiente de variação para as Taxas de Alfabetização, Língua Portuguesa e Matemática para as escolas municipais na capital e interior dos estados brasileiros, Saeb 2021. (continua)

Discriminação	Escola Municipal		Escola Privada	
	Capital	Interior	Capital	Interior
Taxa de alfabetização				
Média	28,03	28,84	76,63	74,16
Mediana	27,56	25,36	79,34	74,86
Desvio Padrão	10,10	13,38	19,73	10,04
Coeficiente de Variação	0,36	0,46	0,26	0,14
Língua Portuguesa				
Média	704,04	705,37	772,34	766,56
Discriminação	Escola Municipal		Escola Privada	
	Capital	Interior	Capital	Interior
Língua Portuguesa				
Mediana	705,75	702,98	770,46	765,53
Desvio Padrão	16,46	20,11	15,12	12,25
Coeficiente de Variação	0,02	0,03	0,01	0,02
Matemática				
Média	724,78	726,23	772,60	764,92
Mediana	722,41	724,29	777,08	761,76
Desvio Padrão	18,92	19,95	19,17	14,20
Coeficiente de Variação	0,03	0,03	0,02	0,02

Fonte: O autor.

Uma opção para avaliar a dispersão em torno dos valores médios simultaneamente corresponde ao cálculo da covariância combinada de dimensionalidade quatro, construída segundo as covariâncias obtidas para a taxonomia referente a capital e interior tanto a nível público municipal e privado, para as disciplinas Matemática, Língua Portuguesa e Taxa de Alfabetização.

Segundo (Morrison, 1971), foi obtido o valor do fator de escala ( $C^{-1}$ ) 0,92, enquanto para a estatística teste (M), o valor 26,15. O produto dessas estatísticas resultou em 24,15, que comparado com o valor 31,14 associado ao valor Chi-quadrado para 20 graus de liberdade ao nível de significância 5%, configura a igualdade das covariâncias entre os desempenhos de Matemática, Taxa de Alfabetização e Língua Portuguesa. Esse resultado representa que as notas, independentes das disciplinas, variam de forma similar independente da região e autarquia. Ou seja, a análise estatística das Taxas de Alfabetização pode representar tão bem o cenário da aprendizagem, nas demais disciplinas. Este achado, pode ser questionado por que as taxas de alfabetização, estão aquém da avaliação de Matemática e Língua Portuguesa, porém cabe ressaltar que as taxas de alfabetização impõem que os estudantes que alcancem o mínimo esperado de alfabetização segundo a ANA. Porém isso não informa sobre a veracidade observada das discrepâncias avaliativas observadas na tabela 1.

Torna-se necessário, obter as covariâncias segundo os pares de disciplinas, tanto para público quanto privado e, também, a nível regional. Similarmente, ao procedimento anterior, foi obtido o valor do fator

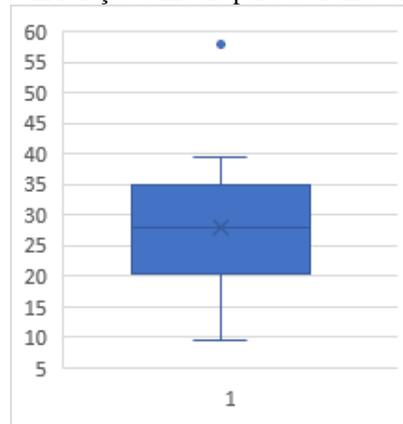
de escala ( $C^{-1}$ ) que se igualou 0,95, enquanto a Estatística Teste “M”, o valor 37,98. O produto dessas estatísticas resultou em 35,92, que comparado com o valor 28,87 associado ao valor Chi-quadrado para 18 graus de liberdade ao nível de significância 5%, configura a desigualdade das covariâncias entre os desempenhos de Matemática, Taxa de Alfabetização e Língua Portuguesa. Este resultado, mostra que dependendo da região e autarquia, o desempenho escolar torna-se diferenciado. Interpreta-se que o primeiro teste de Covariância Combinada não informou a fidedignidade, e sim, talvez, a perpetuação do cenário, pois se há deficiência em pelo menos uma das disciplinas, pode vir a atingir as demais. O desempenho dos estudantes pós alfabetização não evolui de maneira igualitária, inclusive a aferição adequada da taxa de alfabetização é precursora da proficiência esperada dos estudantes.

O coeficiente percentílico de curtose é obtido segundo a razão entre a distância média dos quartis em relação a distância dos extremos, percentil 90 e percentil 10. Se esse valor tender a zero, indica que a variabilidade em torno do valor mediano é pequena, senão quanto maior o coeficiente, maior heterogeneidade se observa nos valores observados. As Taxas de Alfabetização das escolas municipais da capital e do interior apresentaram valores deste coeficiente de 0,309 e 0,289, respectivamente, indicando distribuições platicúrticas, ou seja, não há uma variabilidade expressiva em torno da mediana que apresentaram valores 27,56 e 25,36, respectivamente, não distanciando dos valores médios para capital avaliado em 28,03, mas para o interior, a média está acima deste valor, 28,84, o que configura uma tendência a assimetria. Neste caso, a capital mostrou coeficiente de assimetria 0,78, descrevendo uma distribuição assimétrica à direita, o que implica que a maior parte dos alunos está abaixo da média.

Em contrapartida, a distribuição da Taxa de Alfabetização nas escolas privadas da capital e do interior, apresentaram curtoses percentílicas de 0,247 e 0,207, respectivamente, não se distanciando da curtose da distribuição normal, da ordem de 0,263. Em relação à assimetria, os resultados para o interior mostraram valores próximos dos promédios, o que pode ser inferido que esses resultados aderem a distribuição normal. No caso da capital, o valor foi de -0,23, configurando uma distribuição assimétrica à esquerda, o que permite mostrar que a maioria dos alunos está acima da média.

A análise a nível estadual, permitiu considerar o estado do Piauí como discrepante em relação à média para as Taxas de Alfabetização das escolas municipais da capital (figura 1), sendo que este estado se destacou a nível nacional, praticamente o dobro da média. Apesar da surpresa, torna-se fato relevante, o que leva ao questionamento da técnica de alfabetização.

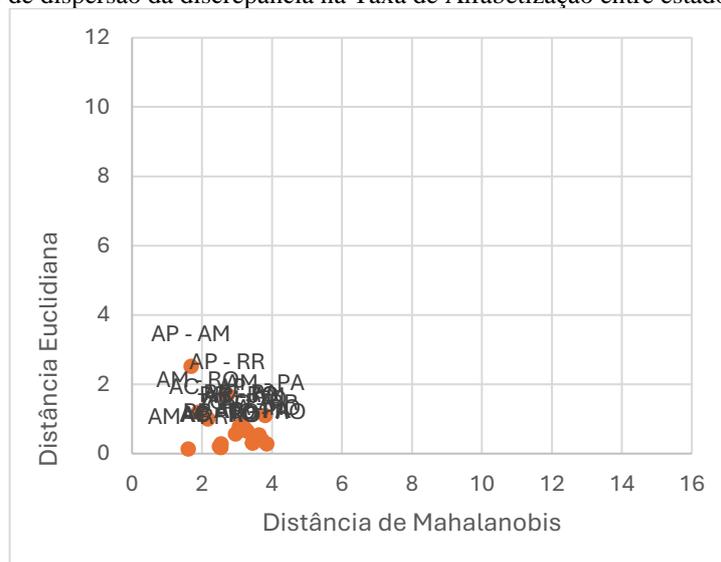
Figura 1: Gráfico boxplot em relação à média para as Taxas de Alfabetização da Capital.



Fonte: O autor.

Um dos objetivos é avaliar as discrepâncias dos resultados da Alfabetização a nível estadual, agrupando os estados a nível regional, Norte, Nordeste, Sul, Sudeste e Centro-Oeste, optando-se pela Distância Euclidiana e Distância de Mahalanobis que foram plotadas no gráfico de dispersão para cada região. Cabe ressaltar que a Distância Euclidiana é obtida em relação ao vizinho mais próximo, enquanto a distância de Mahalanobis, considera a variabilidade segundo a matriz de covariância. Na figura 2, está representado o clusteamento para a região Norte.

Figura 2: Gráfico de dispersão da discrepância na Taxa de Alfabetização entre estados da Região Norte.

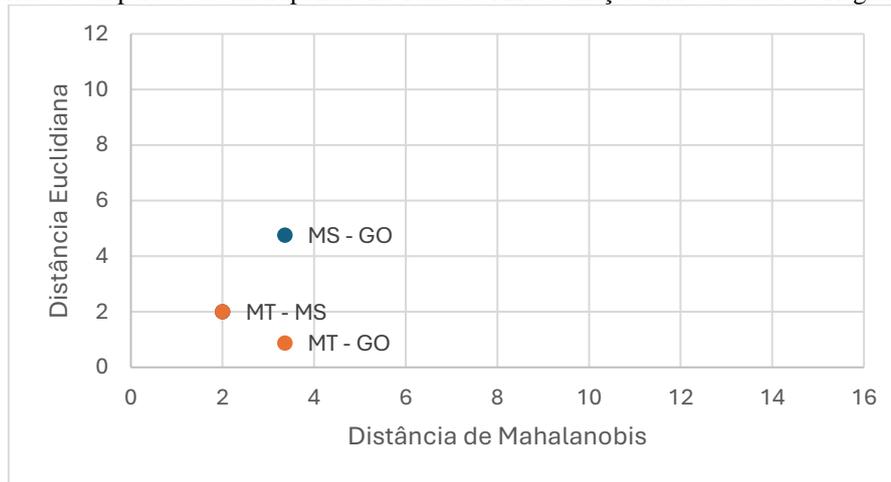


Fonte: O autor.

Os estados do Amazonas e Pará devem estar separados dos demais agrupamentos, sendo Amazonas com tendência para se agregar aos estados da parte inferior esquerda da figura, e o Pará com tendência para se agregar aos estados na parte inferior direita do gráfico, tendo menores Distâncias Euclidianas. Do observado, esses são os que contiveram maior discrepância nos pares ordenados em função das Distâncias Euclidiana e de Mahalanobis, sendo que a de Mahalanobis contempla a influência da variabilidade em torno da média. As figuras 3, 4 e 5 mostram os gráficos de dispersão para os estados das regiões Sudeste, Sul e Centro-Oeste, respectivamente.



Figura 5: Gráfico de dispersão da discrepância na Taxa de Alfabetização entre estados da Região Centro-Oeste.

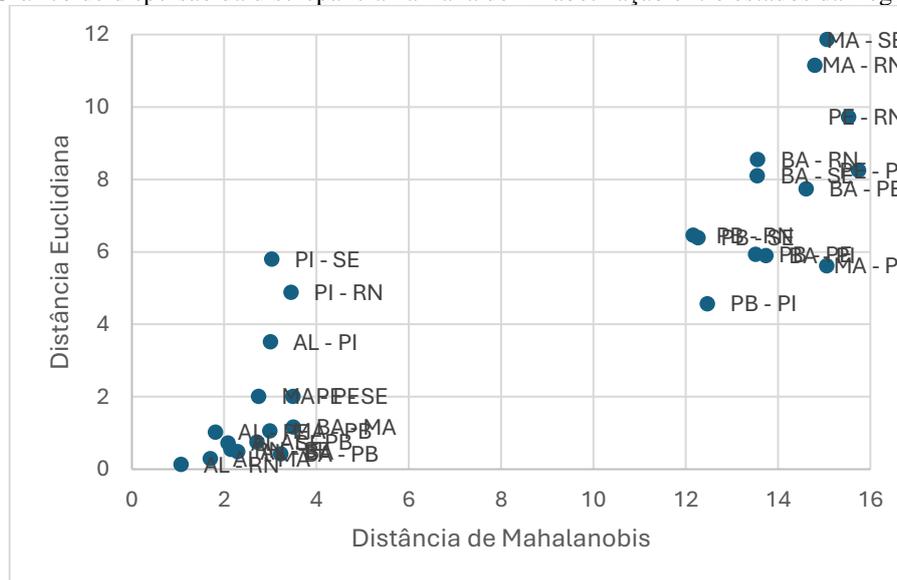


Fonte: O autor.

Pode-se observar que na região Sul, houve pouca expressividade nas diferenças dos valores limites obtidos para a Distância Euclidiana e Distância de Mahalanobis, o que contribuiu para o agrupamento do Rio Grande do Sul e Santa Catarina.

A distância de Mahalanobis para a região Centro Oeste, sinalizou um destaque de Goiás em relação a Mato Grosso, quando se analisa segundo a distância euclidiana evidencia a discriminação de Mato Grosso. Por fim, o gráfico da figura 6, ilustra a região Nordeste.

Figura 6: Gráfico de dispersão da discrepância na Taxa de Alfabetização entre estados da Região Nordeste.



Fonte: O autor.

Na região Nordeste, observou-se claramente dois agrupamentos, o da direita inferior sob forte influência do Piauí e Alagoas devido a frequência das duas distâncias aferidas por pares ordenados. O agrupamento a esquerda com maior influência de Manaus e Bahia, bem mais sensível aos efeitos de variabilidade em torno da média.

## 5 CONCLUSÃO

As taxas de alfabetização nas escolas municipais da capital, escolas municipais do interior, e nas escolas privadas da capital e do interior, são representativas do nível de proficiência estudantil mensurado pelo exame nacional em 2021. Este ano foi um ano crítico, as escolas ainda se adaptavam às mudanças impostas pelo ensino híbrido. Ainda hoje, esses efeitos são sentidos pelos profissionais da educação em todo o país que contribuiu para o mínimo necessário que alavanque o aprendizado das disciplinas analisadas. Portanto, é coerente adotar a taxa de alfabetização como uma forma de diagnosticar cenários futuros do déficit de aprendizagem escolar, permitindo debates entre os educadores sobre maneiras de contornar tais lacunas e capacitar o estudante para o exercício futuro de sua cidadania. O tratamento estatístico realizado permite perceber a igualdade na variação dos aproveitamentos em relação à média nas matérias de Matemática, Língua Portuguesa e Taxa de Alfabetização, segundo teste de covariância tratados tanto nas escolas municipais e privadas, para a capital e interior. Cabe ressaltar, que os valores médios 28,03 das escolas municipais da capital e 76,63 das escolas privadas da capital, representaram os menores e maiores índices de Taxa de Alfabetização, respectivamente. A avaliação da curtose e da simetria, permitiu aferir que há uma diferença entre os resultados das escolas municipais e privadas, pois nas segundas os resultados dos alunos são superiores que o valor médio, enquanto nas municipais, há uma tendência, dos valores estarem próximos dos promédios. Percebeu-se que há uma semelhança entre a avaliação de Matemática e Língua Portuguesa, mas ao confrontar com a taxa de Alfabetização, a maioria está abaixo de 50%, mostrando que, talvez, a forma de construir o instrumento de avaliação em sala de aula, se distancia dos conceitos utilizadas na formação desta taxa. Acresce que na época da pandemia o processo de alfabetização foi interrompido. Fica evidenciado a necessidade das aulas suplementares, da Língua Portuguesa no sentido de suprir a deficiência de alfabetização precária.

O teste de Covariância ao ser aplicado de acordo com o local e o tipo de estudo a ser ofertado, comprovou que há uma diferença na variabilidade em torno da média dos resultados no desempenho nas disciplinas de Português e Matemática e nas Taxas de Alfabetização, comprovando que, no cenário de pandemia, a oferta de ensino não foi a mesma, uma vez que se percebeu que o aproveitamento se diferenciava em relação a autarquia e a regionalidade. Esta análise estatística descritiva, indicou que no cenário de pandemia, o ensino privado, ainda que em um momento de crise, conseguiu prover um melhor ensino, apesar da adoção do ensino híbrido.

Uma vez que o interesse constituiu avaliar o grau de similaridade no ensino, e que o teste de Covariância indicou a representatividade da Taxas de Alfabetização, foi realizado o cálculo das distâncias entre os estados da federação. As análises das distâncias Euclidianas e distâncias de Mahalanobis entre os estados no interior das regiões Sul, Sudeste, Norte, Nordeste e Centro-Oeste, mostrou que na região Sudeste há uma forte influência da variabilidade em torno da média no cálculo das distâncias, sendo esta portanto a que contém o maior grau de dissimilaridade no nível de proficiência medido no Saeb 2021. A região

Centro Oeste foi a única onde os efeitos de variabilidade em torno da média, reduziram distâncias entre os estados em termos de Taxa de Alfabetização.

Desta forma, fica evidente a necessidade do uso de métodos quali quantitativos na mensuração do comprometimento do processo de ensino e aprendizagem, a nível regional, estadual, e municipal, uma vez que subsidia os investimentos a serem aplicados pelas autarquias nas respectivas regionalidades.

## REFERÊNCIAS

- AAKER, D. et al. Pesquisa de marketing. São Paulo: Atlas, 2001
- ALAVARSE, O.M. A organização do ensino fundamental em ciclos: algumas questões. Revista Brasileira de Educação, Rio de Janeiro, v. 14, n. 40, p. 35-50, jan./abr. 2009.
- BARROSO, L.P.; ARTES, R. Análise multivariada. Lavras: UFLA, 2003.
- BONAMINO, A.; FRANCO, C.. Avaliação e política educacional: o processo de institucionalização do Saeb. Cadernos de Pesquisa, São Paulo, n. 108, p. 101-132, nov. 1999.
- BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Censo da Educação Básica 2019: resumo Técnico. Brasília, 2020.
- BRASIL. Lei de Diretrizes e Bases da Educação Nacional, LDB. 9394/1996.
- BUSSAB, W.O.; MIAZAK, E.S.; ANDRADE, D.F. Introdução à Análise de Agrupamentos. 9º Simpósio Brasileiro de Probabilidade e Estatística. São Paulo: IME – USP, 1990.
- CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. (Eds.). . Análise multivariada: para os cursos de administração, ciências contábeis e economia. São Paulo: Atlas, 2007. p. 73–130.
- CORREA, S.M.B.B., Probabilidade e estatística, 2ª ed., Belo Horizonte, PUC Minas Virtual, 2003.
- CRUZ, C. D.; CARNEIRO, P. C. S. Modelos Biométricos aplicados ao melhoramento genético. Viçosa: UFV, 2006.
- FRANCO, C., FERNANDES, C, SOARES, J.F., BELTRÃO, K., BARBOSA, M. E, ALVES, M. T. G. O referencial teórico na construção dos questionários contextuais do SAEB 2001. Estudos em Avaliação Educacional. São Paulo: n.28, p.39-71, 2003.
- FREIRE, P. Pedagogia da indignação: cartas pedagógicas e outros escritos. São Paulo: UNESP, 2000.
- KHATTREE, R.; NAIK, D.N. Multivariate data reduction and discrimination with SAS Software. New York: BBU Press and John Wiley Sons Inc., 2000.
- LUCKESI, J.C. Avaliação da Aprendizagem Escolar: estudos e proposições. 10 ed. São Paulo: Cortez, 2000.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. Cluster analysis. Academic Press, 1997. p.360–393.
- MORRISON, D.F. – Multivariate Statistical Methods - McGraw Hill, N.Y., 1971.
- NAGHETTINI, M.; PINTO, E. J. A. Hidrologia Estatística. Belo Horizonte: CPRM, 2007.
- RAO, C. R. An advanced statistical method in biometric research. New York: Ed. John Wiley e Sons, 1952.
- RIBOLDI, J. Análise de agrupamento "Cluster Analysis" e suas aplicações. Piracicaba: ESALQ, 1986.
- SNEATH, P. H. A.; SOKAL, R. R. Numeric taxonomy: the principles and practice of numerical classification. San Francisco: W. H. Freeman, 1973. p. 15-20.