

## Combinando Diferentes Formas de Cálculo de Indicadores do PNE e Agrupamento para o Apoio à Decisão

Ebony Marques<sup>1</sup>, Gabriel Alves<sup>1</sup>, Andréza Alencar<sup>1</sup>,  
Rafael Mello<sup>1,2</sup>, Ibsen Bittencourt<sup>3</sup>

<sup>1</sup> Universidade Federal Rural de Pernambuco (UFRPE)

<sup>2</sup> CESAR School

<sup>3</sup> Universidade Federal de Alagoas (UFAL)

{ebony.marquesr, gabriel.alves, andreza.leite, rafael.mello}@ufrpe.br,

ibsen@feac.ufal.br

**Abstract.** *The National Education Plan (PNE) establishes goals, guidelines and strategies for developing education in Brazil. The current calculation method for the goals' indicators is uniform across municipalities and states. This study proposes alternative approaches to calculate the PNE indicators at the state level, investigating their potential biases. The alternative calculation methods associated with a clustering analysis allow one to identify states with more opportunities to improve their indicators and highlight similar municipalities with significant deviation from expected results. This paper shows that the current calculation method yields biases for several states. Besides, we conducted a cluster analysis for the Rio Grande do Sul (RS) state, detecting outliers and clusters within municipalities that present results different from expectations.*

**Resumo.** *O Plano Nacional de Educação (PNE) estabelece metas, diretrizes e estratégias para o desenvolvimento da educação no Brasil. A forma de cálculo para os indicadores das metas é padronizada entre municípios e estados. Este estudo propõe abordagens alternativas para calcular os indicadores do PNE a nível estadual, investigando seus vieses potenciais. Os métodos alternativos de cálculo e a análise de grupos permitem que se identifiquem estados com mais oportunidades de melhorias para seus indicadores e ressaltam municípios cujos resultados diferem significativamente do esperado. Este trabalho observou que a forma de cálculo atual apresenta vieses para vários estados. Além disso, a análise de grupos foi realizada para o Estado do Rio Grande do Sul, detectando outliers e grupos de municípios com resultados diferentes da expectativa.*

### 1. Introdução

O Plano Nacional de Educação (PNE) é um instrumento de planejamento estratégico que trata da educação no Brasil. O atual PNE possui vigência de 2014 a 2024 e, ao fim desse período, deve ser criado um novo plano, conforme disposição legal [Brasil 2014]. Compreendendo todas as etapas e modalidades de ensino, o PNE estabelece metas, diretrizes e estratégias com o objetivo de orientar a elaboração de políticas públicas educacionais e promover a melhoria da qualidade da educação no país. O plano traz à tona temas relacionados ao acesso à educação, à qualidade do ensino, à gestão educacional, à formação de professores, à valorização de profissionais e ao investimento na área, entre outros.

O PNE estabelece um conjunto de 20 metas, sendo cada uma delas acompanhada de um ou mais indicadores. Esses indicadores desempenham um papel crucial ao permitir que os gestores da educação monitorem o progresso das metas em níveis que vão desde os municípios e estados até as regiões e o país como um todo. A Lei Nº 13.005 de 2014 [Brasil 2014] determina que o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) seja responsável por definir a metodologia de cálculo dos indicadores e apresentar relatórios periódicos com os resultados correspondentes [INEP 2015]. No contexto dos estados, a fórmula definida para calcular os indicadores é a mesma que é aplicada aos municípios [Brasil 2014].

Esta pesquisa visa analisar a forma atual do cálculo dos indicadores para os estados e propor abordagens alternativas. A motivação deste trabalho se baseia na hipótese de que o cálculo agregado padrão dos indicadores para os estados possui forte influência de municípios populosos. Portanto, estados que tenham municípios populosos com bons resultados podem apresentar bons resultados, ainda que populações inteiras de municípios menores estejam desatendidas. Além disso, buscando auxiliar o gestor estadual a identificar municípios em situação crítica quanto ao atingimento de um indicador do PNE, este trabalho propõe uma estratégia baseada na criação de grupos a partir dos municípios dos estados. A fim de responder às questões de pesquisa listadas abaixo, o método proposto será aplicado sobre o Indicador 1A da Meta 1 do PNE, que determina que todas as crianças de 4 a 5 anos precisam estar matriculadas em escolas ou creches.

Questão de Pesquisa 1 (QP1): *A forma padrão de cálculo de indicadores do PNE para os estados pode influenciar a avaliação de seu desempenho?*

Questão de Pesquisa 2 (QP2): *É possível identificar, no âmbito de um estado, municípios com desempenhos discrepantes quanto a um indicador do PNE por meio de métodos de agrupamento?*

## **2. Trabalhos Relacionados**

Tendo em vista os objetivos do PNE, existem diversos trabalhos que tratam do monitoramento do cumprimento do plano por municípios e estados. Quanto à fiscalização realizada por órgãos públicos, que é legalmente prevista, trabalhos como os de [Lauris et al. 2018] e [Souza et al. 2021] abordam a atuação de tribunais de contas estaduais no monitoramento do PNE. O estudo de [Lauris et al. 2018] apresenta o *software* TC Educa, que permite o monitoramento com base em parâmetros nacionais e uniformes. A pesquisa de [Souza et al. 2021] analisa a ação do Tribunal de Contas do Paraná na fiscalização da Meta 1, destacando a ampliação do escopo de atuação do tribunal para além da execução orçamentária. Esses estudos ressaltam a importância dos órgãos no acompanhamento e garantia da implementação adequada das políticas educacionais previstas pelo PNE.

O estudo de [Simões 2018] investiga os aspectos metodológicos do cumprimento da Meta 1, incluindo indicadores, composição do atendimento e desigualdades de acesso. O trabalho de [Coutinho and Alves 2019] trata do atendimento da Meta 1 por municípios do Paraná, destacando a desigualdade entre os municípios. Por sua vez, o estudo de [Ximenes and Grinkraut 2014] examina o conteúdo da Meta 1 e resalta a necessidade de indicadores transparentes e efetivação do direito de acesso à educação infantil. Esses estudos fornecem *insights* relevantes sobre o alcance da Meta 1 e apontam para a importância de políticas educacionais que visem reduzir as disparidades no acesso à educação infantil.

A pesquisa de [Caseiro and Azevedo 2018] aborda a construção dos indicadores da Meta 12, destacando as decisões teóricas e metodológicas consideradas pelo INEP. O trabalho de [Oliveira 2019] compara os procedimentos metodológicos e os resultados dos indicadores da Meta 17, analisando a substituição da PNAD pela PNAD Contínua. Ambos os estudos destacam a importância de compreender as escolhas metodológicas e as consequências na interpretação dos dados e resultados obtidos a partir dos indicadores de monitoramento.

Técnicas de agrupamento são empregadas em contextos diversos, com diferentes objetivos. Tais técnicas podem ser utilizadas, por exemplo, para agrupar municípios com características semelhantes, buscando fornecer suporte à decisão sobre a destinação de recursos públicos em áreas como saúde [Lima et al. 2009] e educação [Spanol et al. 2022], ou para identificar disparidades entre os municípios de um estado [Costa et al. 2012]. Em especial, o trabalho de [Spanol et al. 2022] propõe a execução de agrupamentos de municípios de todo o Brasil visando à identificação de entes com desempenhos semelhantes quanto a indicadores do PNE.

Além disso, há trabalhos que propõem o uso de agrupamento para a detecção de *outliers*. O trabalho de [Pamula et al. 2011] propõe um método baseado em agrupamento para detectar *outliers*, utilizando o algoritmo de K-Means para dividir o conjunto de dados em grupos e calcular uma pontuação de *outlier* baseada em distâncias. O estudo de [Chawla and Gionis 2013] apresenta uma abordagem unificada para agrupamento e descoberta de *outliers*, formalizando o problema como uma generalização do K-Means. Já o trabalho de [Jiang and An 2008] apresenta um método de detecção de *outliers* baseado em agrupamento utilizando um algoritmo de passagem única e determinação de grupos *outliers* com base no fator de *outlier*. Os resultados experimentais demonstram a eficácia e praticidade dos métodos propostos nos três trabalhos, fornecendo abordagens eficientes e escaláveis para detecção de *outliers* em diferentes tipos de dados.

Os trabalhos citados que trataram do PNE consideraram abordagens distintas para o cálculo de indicadores, incluindo o monitoramento da implementação e o cumprimento das metas. Já os trabalhos de detecção de *outliers* exploraram técnicas de agrupamento, análise de dados e comparação de resultados. O presente trabalho difere-se dos demais ao propor distintas formas de cálculo agregado dos indicadores do PNE. Com isso, é possível observar e comparar o desempenho dos estados, regiões e país sob diferentes pontos de vista. Além disso, quanto ao uso de agrupamento, este trabalho aplica essa técnica tanto para a detecção de *outliers* quanto para a comparação de municípios com características semelhantes a nível estadual. Assim, o método proposto neste estudo pode auxiliar gestores públicos a otimizar o uso dos recursos para educação, priorizando municípios que gerem maior impacto nos indicadores do PNE.

### 3. Método para Análise de Indicadores do PNE em Estados

Considerando as duas questões de pesquisa apresentadas, o método utilizado para o desenvolvimento deste trabalho é constituído por duas etapas: o **cálculo agregado do indicador dos estados brasileiros** e a **análise de grupos dos municípios de um estado**. As subseções seguintes apresentam as etapas detalhadamente.

### 3.1. Cálculo Agregado do Indicador para os Estados Brasileiros

A primeira etapa do método proposto possui o objetivo de responder à QP1, observando os impactos que a forma padrão de cálculo do indicador agregado para os estados causa sobre a análise de desempenho. O cálculo padrão do indicador do estado é obtido com a mesma fórmula empregada no cálculo padrão do indicador do município, que trata da razão das populações atendida e estimada do ente observado [Brasil 2014], agora considerando as populações do estado. Além do cálculo padrão, este estudo propõe a análise do indicador estadual a partir da média e da mediana dos indicadores dos municípios do estado. Os dados utilizados para a execução deste trabalho são oriundos de bases de dados públicas disponibilizadas pelo IBGE e pelo INEP.

O diagrama de caixa (*boxplot*) pode ser utilizado a fim de visualizar a diferença entre o cálculo padrão e o cálculo com base na mediana, considerando os indicadores do PNE aferidos para todos os municípios do estado. Um marcador extra deve ser incluído a fim de representar o valor do cálculo padrão para o estado. Quanto maior o Intervalo Inter-Quartil (IQR), maior a variação do indicador para os municípios associados àquele estado. Por conseguinte, quanto menor o IQR, maior a equidade entre os municípios do estado, com relação ao indicador [Bruce and Bruce 2019].

O *boxplot* permite identificar a existência de *outliers* [Bruce and Bruce 2019]. Os *outliers* positivos e negativos representam, respectivamente, municípios que possuem um valor do indicador muito superior e muito inferior com relação aos demais municípios do estado. Caso um município considerado *outlier* seja muito populoso, o impacto sobre o cálculo padrão será expressivo. Se um estado possui a maior parte de sua população concentrada em um único município e esse é um *outlier* positivo para o indicador, a forma de cálculo padrão pode apontar para o atingimento da meta por parte do estado, mesmo que todos os demais municípios estejam bem abaixo do objetivo definido para aquele indicador. De forma geral, quanto maior a distância do marcador do cálculo padrão para a mediana, maior o impacto de municípios populosos sobre o cálculo padrão um estado.

Além da análise com base na mediana, a análise equivalente com base na média dos indicadores para os municípios do estado provém uma outra visão. Enquanto a mediana e IQR são medidas menos sensíveis a grandes variações e *outliers*, a média e o desvio padrão são mais sensíveis a essas variações [Bruce and Bruce 2019]. Portanto, para observar a variabilidade dos dados dos municípios de um estado, a análise da média em conjunto com o desvio padrão fornece uma perspectiva mais apropriada. Para a visualização desses dados, convém utilizar um gráfico com um marcador e uma barra de erro para representar a média e o desvio padrão do valor dos indicadores para cada município do estado. As mesmas análises realizadas para a mediana também são válidas nesse caso. Vale salientar que, como a média é mais sensível a extremos, as diferenças entre os valores da média e do cálculo padrão tendem a ser menores do que no caso da mediana. Ademais, quanto maior a barra de erro, que indica o desvio padrão, maior a diferença entre os resultados dos municípios com relação ao indicador analisado. Portanto, enquanto a análise com base na mediana e IQR é mais adequada para observar *outliers*, a análise proposta mostra de melhor forma a variabilidade dos valores do indicador em um estado.

Por fim, deve-se aferir o erro médio ao comparar os valores do cálculo padrão e os valores com base na mediana e na média dos indicadores. Para tanto, o cálculo do *Mean Squared Error* (MSE) (Erro Quadrático Médio) e do *Mean Absolute Error* (MAE)

(Erro Absoluto Médio) permite observar essa diferença [Han et al. 2022]. Ambas medidas mostram a diferença entre o valor “real” (cálculo padrão) e o valor no “modelo” (cálculo proposto pela mediana ou média dos indicadores municipais). Contudo, a primeira amplifica os erros, uma vez que eleva os valores ao quadrado, enquanto a segunda afere apenas o valor absoluto da diferença. Assim, quanto maior o valor do MSE e do MAE para um indicador, maior o impacto de municípios muito populosos do Brasil sobre os resultados dos indicadores dos estados. Conseqüentemente, haverá indícios de que alguns municípios, em especial os menos populosos, podem estar sendo relegados com relação àquele indicador em análise, ou mesmo que municípios muito populosos precisam de mais investimentos. A Seção 4.1 apresenta um estudo detalhado para esta etapa do método, aplicando-a no cálculo do Indicador 1A do PNE para estados brasileiros.

### 3.2. Análise de Grupos para os Municípios do Estado

A segunda parte do método proposto tem o objetivo de responder à QP2. Para tanto, procura-se criar grupos de municípios no âmbito de um estado, observando o indicador analisado e algumas variáveis relacionadas aos municípios. A criação de grupos permite identificar não apenas *outliers*, mas também municípios que, apesar de terem características semelhantes, têm o resultado acima ou aquém do esperado para um indicador.

Neste estudo, usa-se o K-Means, que tem visa agrupar os pontos do conjunto de dados em  $k$  grupos de modo que a soma dos quadrados das distâncias dos pontos aos centros dos grupos obtidos seja minimizada. O algoritmo, que recebe  $k$  como entrada, define aleatoriamente  $k$  pontos da amostra como centroides de cada grupo e, com base na distância euclidiana, inclui os demais elementos da amostra nos grupos, calculando-se o erro das distâncias. O processo é executado de forma iterativa até que o erro permaneça estável, dentro de limiares pré-estabelecidos [Han et al. 2022].

Para a criação dos grupos, deve-se utilizar variáveis socioeconômicas e educacionais dos municípios e o indicador do PNE que se deseja analisar. Assim, este trabalho emprega as dimensões de educação e renda do Índice de Desenvolvimento Humano Municipal (IDHM), além da população e do desempenho do indicador do PNE dos municípios. Dessa forma, a escolha das variáveis se dá no sentido de que se espera que municípios que possuem níveis equivalentes de renda, educação e população também deveriam possuir resultados equivalentes para os indicadores do PNE.

Ao criar grupos utilizando variáveis que possuem uma grande diferença nas escalas, deve-se normalizar ou padronizar essas variáveis. Com isso, evita-se que o algoritmo de agrupamento priorize indevidamente variáveis com escalas maiores [Han et al. 2022]. Como este trabalho lida com variáveis com escalas muito discrepantes, especialmente a população, realiza-se a padronização das variáveis por meio do *Z-Score*, calculado como  $Z = (x - \mu) / \sigma$ , em que  $x$ ,  $\mu$  e  $\sigma$  são, respectivamente, o valor, a média e o desvio padrão da variável padronizada.

Considerando o K-Means e visando definir a quantidade de grupos, devem ser realizados testes com diferentes valores para  $k$ , com  $2 \leq k \leq 10$ . Para escolher o valor mais adequado de  $k$ , calculam-se os coeficientes de inércia, normalmente observados no Método do Cotovelo, e de silhueta. Com base nos coeficientes, escolhe-se o  $k$  ideal para o problema analisado [Han et al. 2022]. Vale salientar que, como muitos municípios têm valores próximos para as variáveis analisadas, alguns dos grupos podem apresentar

sobreposição, impactando diretamente os valores dos referidos coeficientes. Note que, em uma situação hipotética e ideal, em que todos os municípios tivessem excelentes níveis de educação, renda e o indicador máximo, os grupos iriam diferir apenas quanto à população. Contudo, é exatamente devido à discrepância nesses indicadores que a técnica de agrupamento se mostra eficaz para identificar resultados diferentes do esperado.

Por fim, a apresentação dos resultados dos grupos auxilia o processo de responder à QP2. Quanto maior a quantidade de municípios e de variáveis utilizadas para a criação dos grupos, maior o desafio para exibir esses resultados. Assim, criam-se dois tipos de gráficos para a exibição dos grupos: um gráfico de dispersão para cada par de variáveis usadas no agrupamento, em que cada município é representado por um marcador associado ao seu grupo, e um gráfico de dispersão com barra de erro para cada variável usada no agrupamento, em que o eixo  $x$  é uma variável utilizada no agrupamento, o eixo  $y$  é o grupo e a barra de erro indica o desvio padrão da variável naquele grupo.

O primeiro gráfico permite observar a relação entre as variáveis para cada grupo e é especialmente útil para visualizar o tamanho dos grupos e elementos com pouca sobreposição. A estratégia de criar gráficos para cada par de variáveis também permite ter uma visão multidimensional dos grupos. Contudo, especialmente quando há muita sobreposição dos grupos, essa estratégia se mostra ineficaz. Dessa maneira, um gráfico de dispersão com barra de erro permite visualizar a média e dispersão dos grupos para cada variável utilizada. Esses dois tipos de gráficos combinados se mostram eficazes para apresentar os dados independentemente da quantidade de municípios dos estados e das variáveis empregadas para o agrupamento.

Ao criar os grupos, é importante observar que, quanto maior a variabilidade de um grupo com relação a uma variável, menor a importância da variável para a criação daquele grupo. Caso essa dispersão ocorra apenas para o indicador do PNE, trata-se de um indício de que o grupo tenha capturado municípios que, apesar de terem características muito parecidas, possuem desempenhos muito discrepantes com relação ao indicador. Com isso, seria importante que os gestores observassem boas práticas de outros municípios, especialmente daqueles que estão no mesmo grupo e que tiveram resultados melhores para o indicador. No caso de grupos com poucos elementos, há bons indícios de que esses representam grupos de municípios *outliers*. Nesse caso, convém observar — especialmente com os gráficos de dispersão com barra de erro — quais as variáveis geraram as maiores discrepâncias. Caso tenha sido o indicador do PNE e esse tenha um alto valor, o município pode ser considerado uma referência para boas práticas que levem a uma melhoria com relação ao indicador. Contudo, caso o valor do indicador seja baixo, há indícios de problemas, como, por exemplo, na coleta de dados para o município. A Seção 4.2 apresenta um estudo detalhado para esta etapa do método, observando o Estado do Rio Grande do Sul (RS) e seus municípios e considerando o Indicador 1A do PNE.

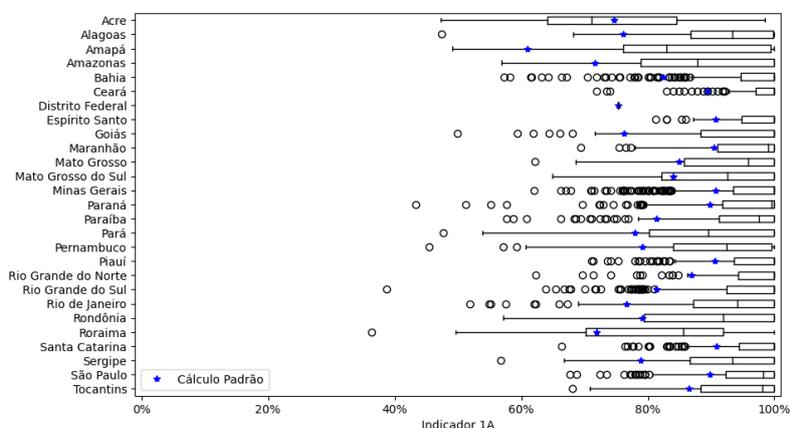
#### **4. Análise do Indicador 1A do PNE**

O Indicador 1A do PNE procura aferir o *percentual da população de 4 e 5 anos que frequenta a escola*. Esse indicador será usado como base para responder às perguntas de pesquisa deste trabalho, seguindo o método apresentado na Seção 3. Assim, os resultados do Indicador 1A para os estados brasileiros são discutidos mediante diferentes formas de agregação do indicador, na Seção 4.1. Em seguida, a Seção 4.2 apresenta uma análise de

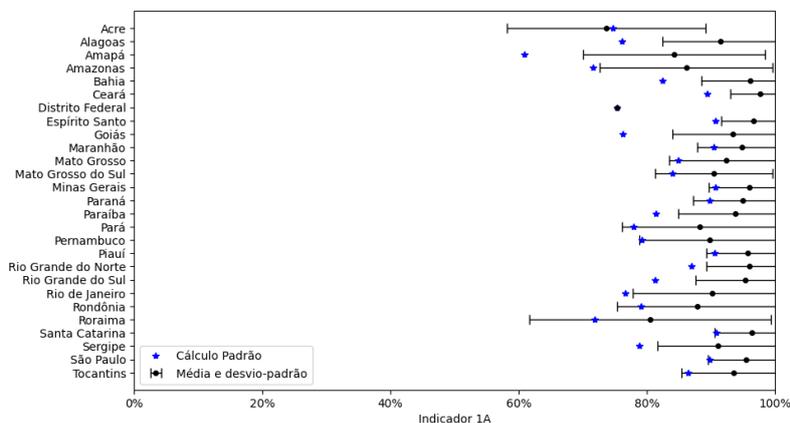
grupos para os municípios do Estado do Rio Grande do Sul (RS).

#### 4.1. Cálculo Agregado do Indicador 1A para os Estados Brasileiros

A partir do método apresentado na Seção 3.1, esta seção procura responder à QP1, utilizando como estudo de caso o Indicador 1A, para todos os estados do Brasil. Com o cálculo da média e o desvio padrão do Indicador 1A de todos os municípios brasileiros agrupados por estado, a Figura 1b apresenta um gráfico de dispersão com barra de erro. Esse gráfico também expõe um marcador com o valor do Indicador 1A para cada estado, aferido com a forma de cálculo padrão [INEP 2015]. Por sua vez, a Figura 1a exibe um diagrama de caixa (*boxplot*) para os valores do Indicador 1A dos municípios brasileiros, agrupados por estado, e o valor do Indicador 1A de cada estado com o cálculo padrão.



(a) Diagrama de caixa com a mediana e o cálculo padrão.



(b) Gráfico de dispersão com barra de erro com a média e o cálculo padrão.

**Figura 1. Análise da diferença entre o cálculo padrão, mediana e média do Indicador 1A para os estados brasileiros.**

A partir do *boxplot*, é possível observar que muitos estados possuem *outliers* negativos para o Indicador 1A. Esses *outliers* são municípios que têm valores muito inferiores aos demais municípios do estado para o Indicador 1A. Nesses casos, o valor do cálculo padrão tende a ser menor que a mediana (ou mesmo a média) dos demais municípios do estado, como observado para estados como o Rio Grande do Sul e Bahia. No caso da Bahia,

o cálculo padrão é ainda menor que o valor inferior da barra de erro, como observado na Figura 1a, mesmo com o seu já alto desvio padrão.

No caso do Ceará, apesar de o cálculo padrão ficar abaixo do desvio padrão com o cálculo da média, esse estado apresenta uma das menores variabilidades, quando comparado com os demais estados. Contudo, ao analisar o diagrama de caixas, é possível observar a ocorrência de alguns *outliers*. Pode-se inferir que esse estado possui uma boa uniformidade entre os seus municípios, mas que o cálculo padrão é diretamente impactado pelos *outliers*, especialmente aqueles que representam municípios mais populosos. Assim, a fim de melhorar os resultados do Indicador 1A para esse estado, seus esforços poderiam ser direcionados exatamente a esses municípios.

O erro entre os valores do cálculo com base na média e o cálculo padrão para os estados é de 111,78 para o MSE e de 9,29 para o MAE. Já quando se analisa a diferença entre a mediana e o cálculo padrão, o MSE é de 181,78, enquanto o MAE é de 12,48. Como esperado, a diferença para mediana é maior, uma vez que essa possui menos influência de *outliers*, que estavam presentes em vários estados (Figura 1a). Ademais, esses valores indicam que há uma diferença significativa entre o cálculo padrão e o cálculo com base nas médias ou mediana dos indicadores dos municípios. Sendo assim, como discutido na Seção 3.1, esses cálculos com base média ou na mediana dão uma visão completamente diferente sobre os resultados dos estados para o Indicador 1A.

#### **4.2. Análise dos Grupos de Municípios do Rio Grande do Sul para o Indicador 1A**

Esta seção procura responder à QP2, fundamentando-se na análise do Indicador 1A para os municípios do Rio Grande do Sul (RS). Como mostrado na seção anterior, o estado tem muitos municípios que atingiram a meta de 100% do Indicador 1A (ao menos metade deles), além de possuir diferenças significativas entre as métricas observadas. Analisar os desempenhos dos municípios por meio do agrupamento proposto pode permitir uma melhor compreensão acerca do desempenho do estado.

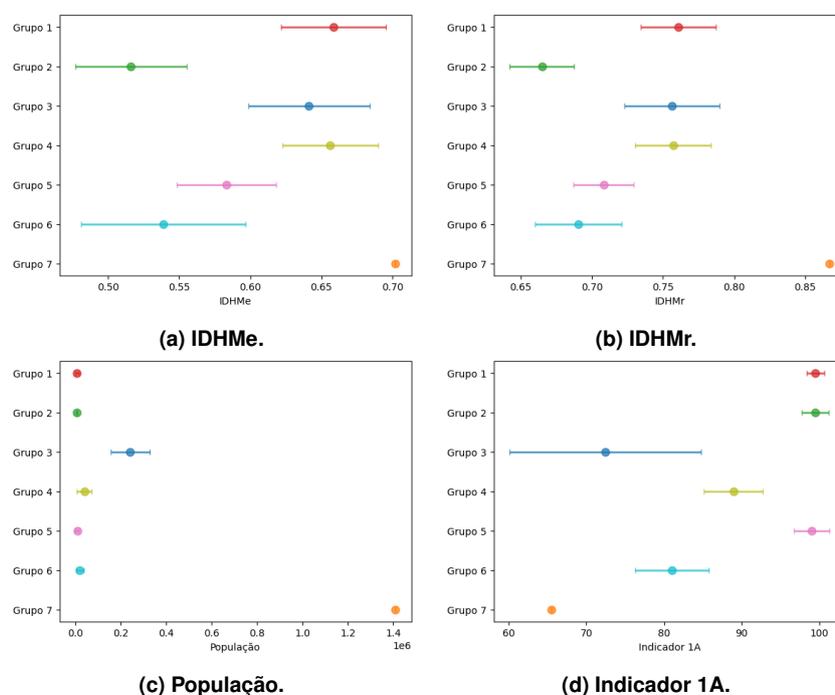
De acordo com o IBGE, o Rio Grande do Sul possui atualmente 497 municípios, mas um deles, Pinto Bandeira, foi criado após o Censo Demográfico de 2010. Com isso, até a divulgação dos resultados de um novo Censo, não há dados disponíveis sobre ele. Dessa forma, este estudo trata de agrupar 496 municípios do estado. A Tabela 1 exibe os coeficientes de inércia e de silhueta, utilizados para a avaliação da coesão e separação dos grupos, variando o  $k$  de 2 a 10 grupos com o algoritmo K-Means. A coluna  $\Delta$  *Inércia* apresenta a variação do coeficiente de inércia com relação à quantidade de grupos anterior, sendo utilizado o próprio coeficiente de inércia para o  $k = 2$ . Assim, quanto maior o valor absoluto dessa coluna, maior será a variação resultante no gráfico do cotovelo, indicando uma melhor definição entre os grupos gerados.

A Tabela 1 mostra que, para 7 grupos, tem-se um bom resultado tanto no coeficiente da silhueta quanto na diferença da inércia. A Figura 3 mostra cada município do Rio Grande do Sul em gráficos de dispersão da relação entre as variáveis, com o grupo sendo representado pela forma do marcador do município. Já a Figura 2 apresenta a média e o desvio padrão dos grupos para cada uma das variáveis. Os grupos de 1 a 7 possuem respectivamente 158, 80, 13, 76, 126, 42 e 1 município(s). Percebe-se que o *Grupo 7* representa um *outlier* na variável populacional (Figura 2c) e é composto apenas pelo Município de Porto Alegre. Apesar de o grupo possuir valores altos para o IDHMe e IDHMr,

**Tabela 1. Coeficiente de inércia, Variação da inércia e Coeficiente de silhueta para o agrupamento de 2 a 10 grupos.**

K	Inércia	$\Delta$ Inércia	Coef. Silhueta
2	1376.0031	1376.0031	0.3535
3	1025.0120	-350.9911	0.3596
4	743.6564	-281.3556	0.3856
5	620.8042	-122.8522	0.3960
6	523.4703	-97.3339	0.3054
<b>7</b>	<b>438.6117</b>	<b>-84.8586</b>	<b>0.5381</b>
8	408.6376	-29.9741	0.4466
9	379.1304	-29.5072	0.4245
10	350.5414	-28.5890	0.3864

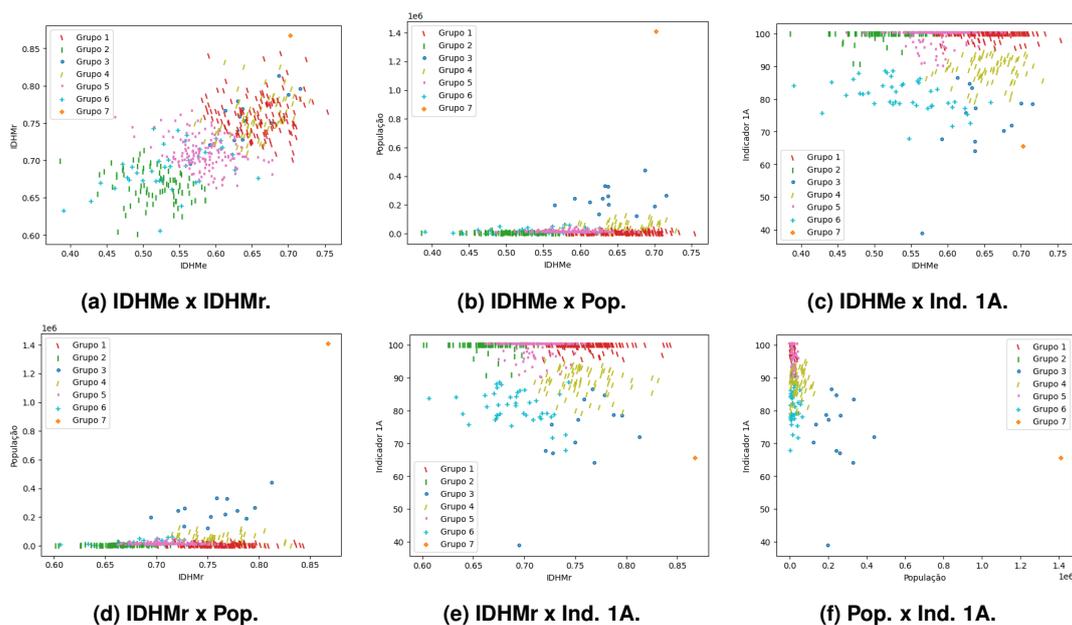
o seu Indicador 1A está entre os piores do estado. Logo, o cálculo padrão do Indicador 1A para o Rio Grande do Sul sofre uma forte influência negativa desse município. O fato de o IDHMe do município ser alto aponta para indícios de falhas pontuais, seja no atendimento do indicador, seja no cômputo e informação do dado no Censo.



**Figura 2. Média e desvio padrão dos grupos para as variáveis utilizadas no agrupamento de municípios do Rio Grande do Sul (RS).**

Possuindo apenas 13 municípios, o *Grupo 3* se mostra mais suscetível a variações e também pode indicar valores um pouco divergentes dos demais. De fato, ao observar a Figura 2c, à exceção do Município de Porto Alegre, os municípios pertencentes a esse grupo são os mais populosos do estado. Assim, esses municípios também possuem maior impacto sobre o cálculo padrão do Indicador 1A para o estado. Uma vez que esses municípios possuem uma maior variabilidade no Indicador 1A (vide Figura 2d), os municípios com os piores valores para o indicador seriam bons candidatos a terem ações direcionadas buscando a melhoria do Indicador 1A.

É possível observar na Figura 2d que os municípios pertencentes aos grupos 1, 2 e 5 possuem os maiores valores para o Indicador 1A, com baixo desvio padrão. Isso



**Figura 3. Relação entre as variáveis utilizadas no agrupamento de municípios do Rio Grande do Sul (RS).**

indica que os municípios desses grupos possuem um comportamento similar quanto a esse indicador. Contudo, a Figura 2a mostra que o IDHMe dos municípios do *Grupo 2* está entre os piores do estado. Dessa forma, apesar de terem bons resultados para esse indicador específico, é válido ter uma maior atenção voltada aos demais indicadores do PNE para os municípios do grupo em questão, ou mesmo observar se estão ocorrendo falhas ao calcular esse indicador, que elevariam indevidamente seus resultados.

Por fim, o *Grupo 4* tem características de IDHMe, IDHMr e população similares às dos municípios do *Grupo 1*, com valores inferiores para o Indicador 1A. Acredita-se que ações e boas práticas adotadas por municípios pertencentes ao *Grupo 1* possam ser adaptadas para melhorar os resultados do indicador para os municípios do *Grupo 4*.

## 5. Considerações Finais

A análise das formas de cálculo agregado dos indicadores do PNE para os estados apresentada neste trabalho mostrou as limitações e vieses associados ao cálculo padrão dos indicadores, definido conforme previsão legal. Nota-se que, quanto mais populoso um município, maior a sua influência sobre o resultado do estado. Dessa forma, estados podem atingir o objetivo de um indicador deixando muitos municípios total ou parcialmente desassistidos. Por outro lado, se um único município muito populoso tem um baixo desempenho para um indicador, os resultados do estado são impactados de forma excessiva.

A análise de grupos mostrou que os municípios de um estado podem ser agrupados tanto a fim de identificar *outliers* quanto para que haja um compartilhamento de ações e boas práticas entre municípios semelhantes. Ademais, os grupos com maior dispersão e com muitos elementos podem ser priorizados para uniformizar e melhorar o desempenho do estado como um todo. Além da análise de outros indicadores do PNE e estados, trabalhos futuros podem incluir outras variáveis associadas à alocação de recursos financeiros nos municípios para encontrar discrepâncias nos investimentos de verbas públicas.

## Referências

- Brasil (2014). Lei Nº 13.005, de 25 de junho de 2014. *Diário Oficial da República Federativa do Brasil*.
- Bruce, A. and Bruce, P. (2019). *Estatística Prática para Cientistas de Dados*. Alta Books.
- Caseiro, L. and Azevedo, A. (2018). A construção dos indicadores de monitoramento da Meta 12 do PNE. *Cadernos de Estudos e Pesquisas em Políticas Educacionais*, 1.
- Chawla, S. and Gionis, A. (2013). k-means–: A unified approach to clustering and outlier detection. In *SIAM international conference on data mining*, pages 189–197. SIAM.
- Costa, C. C. d. M., Ferreira, M. A. M., Braga, M. J., and Abrantes, L. A. (2012). Disparidades inter-regionais e características dos municípios do estado de minas gerais. *Desenvolvimento em Questão*, 10(20):52–88.
- Coutinho, Â. S. and Alves, T. (2019). Desigualdade de acesso à Educação Infantil: uma análise da meta 1 do PNE na região metropolitana de Maringá. *Educar em Revista*, 1.
- Han, J., Pei, J., and Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.
- INEP (2015). Plano nacional de educação pne 2014-2024: linha de base.
- Jiang, S. and An, Q. (2008). Clustering-based outlier detection method. In *International conf. on fuzzy systems and knowledge discovery*, volume 2, pages 429–433. IEEE.
- Lauris, R. P., de Oliveira, P. P., and Grosser, V. P. (2018). Tc educa: rumo a parâmetros de fiscalização das metas do plano nacional de educação pelos tribunais de contas. *Revista do Tribunal de Contas do Estado de Minas Gerais*, 36.
- Lima, L. D. d., de, A., and Tavares, C. L. (2009). Condições de financiamento em saúde nos grandes municípios do Brasil. *Cadernos de Saúde Pública*.
- Oliveira, M. L. d. (2019). Estudo comparativo entre os procedimentos metodológicos e os resultados dos indicadores de monitoramento da Meta 17 do Plano Nacional de Educação. *Cadernos de Estudos e Pesquisas em Políticas Educacionais*, 2.
- Pamula, R., Deka, J. K., and Nandi, S. (2011). An outlier detection method based on clustering. In *2011 second international conference on emerging applications of information technology*, pages 253–256. IEEE.
- Simões, A. A. (2018). Os indicadores da Meta 1 do PNE. *Cadernos de Estudos e Pesquisas em Políticas Educacionais*, 1.
- Souza, D. N. M. d., Silva, S. C. d., Costa, G. S. A., Rodrigues, R. S., and Júnior, M. A. d. A. E. (2021). O tribunal de contas e o cumprimento da meta 1 do pne pelos municípios alagoanos. *Essentia-Revista de Cultura, Ciência e Tecnologia da UVA*, 22(2):34–42.
- Spanol, M., Oliveira, E., Alves, G., Bittencourt, I., Falcão, T. P., and Mello, R. F. (2022). Uso de agrupamento para avaliação de desempenho educacional e apoio à gestão em Áreas de investimento. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, pages 944–955, Porto Alegre, RS, Brasil. SBC.
- Ximenes, S. and Grinkraut, A. (2014). Acesso à educação infantil no novo pne: parâmetros de planejamento, efetivação e exigibilidade do direito. *Cadernos Cenpec — Nova série*, 4(1).