TIPOLOGIA SOCIOECONÔMICA DOS MUNICÍPIOS DO RIO DE JANEIRO POR ÍNDICE COMPOSTO E AGRUPAMENTOS MULTIVARIADO.

Renato de Queiroz Machado[1], Evelyn Nunes Goulart da Silva Pereira[2,3], Soraya de Oliveira Bandeira[4], Yuri Duarte Porto[5], Igor Campos de Almeida Lima[3,6]

[1]Instituto Federal do Rio de Janeiro, Nilópolis, Brasil (nhoquim@gmail.com)
[2]Fundação Oswaldo Cruz, Rio de Janeiro, Brasil
[3]Universidade Federal Rural do Rio de Janeiro, Três Rios, Brasil
[4]Instituto Alberto Luiz Coimbra de Pós-graduação de Pesquisa e Engenharia (COPPE) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil
[5]Universidade Federal de Mato Grosso, Cuiabá, Brasil
[6]Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil

Resumo: Este estudo classifica os 92 municípios do Rio de Janeiro por IDH-M, PIB per capita, remuneração média e Gini. Após transformações, padronização e índice composto, aplicaram-se PCA, Ward.D2, k-means, PAM e modelos de mistura gaussianos. Os resultados indicam dois perfis principais, com maior renda e IDH-M associados também a maior desigualdade média.

Palavras-chave: Indicadores socioeconômicos; Índice composto; PCA; Ward.D2; Rio de Janeiro.

INTRODUÇÃO

O estado do Rio de Janeiro reúne municípios com trajetórias econômicas, demográficas e sociais muito diferentes. Na mesma unidade federativa convivem áreas metropolitanas densas, polos industriais, cidades turísticas, municípios ligados à cadeia do petróleo e localidades com menor dinamismo produtivo. Por isso, a leitura apoiada em um único indicador é limitada: crescimento econômico não se traduz automaticamente em bem-estar social, melhores condições de trabalho ou menor desigualdade. Essa ressalva é recorrente na literatura sobre desenvolvimento humano, na qual indicadores sintéticos são úteis, mas precisam ser interpretados com cautela porque resumem fenômenos multidimensionais (Sagar; Najam, 1998; Postali, 2009).

Nesse cenário, classificar os municípios fluminenses por indicadores socioeconômicos ajuda a reconhecer padrões de semelhança e diferença entre localidades. A análise pode apoiar diagnósticos territoriais, planejamento público e definição de prioridades, desde que seus limites sejam explicitados. A combinação de IDH-M, PIB per capita, remuneração média por ocupado (RMO) e coeficiente de Gini reúne dimensões complementares: desenvolvimento humano, capacidade produtiva, renda do trabalho formal e desigualdade. A presença do Gini é decisiva, pois médias de renda ou produção podem esconder concentração interna de renda, aspecto central na mensuração da desigualdade (Atkinson, 1970; Ceriani; Verme, 2012).

O uso de índice composto e de métodos multivariados é adequado porque organiza indicadores heterogêneos em uma estrutura comparável entre municípios. Ainda assim, essa estratégia não produz uma verdade única sobre desenvolvimento municipal. A literatura sobre indicadores compostos mostra que normalização, orientação das variáveis, ponderação e forma de agregação influenciam a classificação final. Assim, cada decisão metodológica precisa ser declarada para que o índice seja lido como síntese analítica, e não como medida absoluta e imune a escolhas do pesquisador (Saisana; Saltelli; Tarantola, 2005; Mazziotta; Pareto, 2013; Greco et al., 2019).

Diante disso, este trabalho teve como objetivo classificar e agrupar os 92 municípios do estado do Rio de Janeiro com base em indicadores socioeconômicos. A estratégia combinou análise de componentes principais, construção de um índice socioeconômico municipal e métodos de agrupamento, com ênfase em Ward.D2, k-means, PAM/k-medoids e modelos de mistura gaussianos. O conjunto dessas técnicas permite examinar a estrutura dos dados por diferentes ângulos: redução de dimensionalidade, formação de grupos homogêneos, comparação entre partições e avaliação da

consistência dos agrupamentos (Ward, 1963; Jolliffe; Cadima, 2016; Rousseeuw, 1987).

MATERIAL E MÉTODOS

A base analisada incluiu os 92 municípios do estado do Rio de Janeiro e quatro indicadores: IDH-M, PIB per capita, remuneração média por ocupado e coeficiente de Gini. Segundo o Programa das Nações Unidas para o Desenvolvimento, o Instituto de Pesquisa Econômica Aplicada e a Fundação João Pinheiro (PNUD; IPEA; FJP, [s.d.]), o Atlas do Desenvolvimento Humano no Brasil disponibiliza indicadores relacionados ao desenvolvimento humano municipal. Antes da modelagem, verificaram-se valores faltantes e coerência dos tipos numéricos. Como as variáveis têm unidades e escalas distintas, o tratamento dos dados buscou preservar o significado substantivo dos indicadores e, ao mesmo tempo, permitir sua utilização em técnicas baseadas em distância. A seleção das variáveis seguiu a lógica de reunir desenvolvimento humano, produção econômica, renda do trabalho formal e desigualdade, em diálogo com a literatura sobre desenvolvimento humano e mensuração de desigualdade (Sagar; Najam, 1998; Atkinson, 1970).

As variáveis PIB per capita e remuneração média por ocupado foram transformadas por $\log(1+x)$, conforme a Equação (1), para reduzir a influência da assimetria à direita, comum em bases socioeconômicas. Nessa expressão, x_{ir} representa o valor original do município *i* no indicador *r*, e x'_{ir} representa o valor transformado. O uso de transformações para melhorar o comportamento distribucional e a estabilidade analítica é compatível com a tradição inaugurada por Box e Cox (1964).

$$x'_{ir} = \log(1 + x_{ir}) \tag{1}$$

Como valores maiores do coeficiente de Gini indicam maior desigualdade, a Equação (2) inverte esse indicador, fazendo com que valores maiores passem a representar situação relativamente mais favorável.

$$G_i^* = 1 - G_i \tag{2}$$

Em seguida, foi aplicada a normalização min-max: o valor de cada município é reescalonado a partir do mínimo e do máximo observados na respectiva variável, resultando em valores entre 0 e 1 (a Equação 3).

$$x_{ir}^* = \frac{x_{ir} - \min(x_r)}{\max(x_r) - \min(x_r)} \tag{3}$$

Por fim, a Equação (4) define o índice socioeconômico municipal como a média aritmética de quatro termos já normalizados: IDH-M, PIB per capita transformado, RMO transformada e Gini invertido. A opção por pesos iguais favorece transparência, mas permanece uma escolha normativa, pois polaridade, normalização, ponderação e agregação afetam o resultado final (Mazziotta; Pareto, 2013; Saisana; Saltelli; Tarantola, 2005).

$$I_i = \frac{1}{4}\left(\mathrm{IDH}_i^* + \mathrm{PIB}_i^* + \mathrm{RMO}_i^* + \mathrm{GiniInv}_i^*\right) \tag{4}$$

Para a PCA e para os agrupamentos, utilizou-se a matriz padronizada composta por IDH-M, PIB per capita transformado, remuneração média transformada e Gini invertido. A padronização, dada pela Equação (5), calcula z_{ir} a partir do valor x_{ir}, da média da variável *r* e do desvio-padrão s_r.

$$z_{ir} = \frac{x_{ir} - \bar{x}_r}{s_r} \tag{5}$$

Em seguida, a dissimilaridade entre dois municípios *i* e *j* foi medida pela distância euclidiana da Equação (6), em que *p* é o número de variáveis e z_{ir} e z_jr z_{jr} são os valores padronizados desses municípios na variável *r*. A PCA foi usada com finalidade exploratória e interpretativa, como técnica de redução de dimensionalidade, e não como modelo causal (Jolliffe; Cadima, 2016).

$$d_{ij} = \sqrt{\sum_{r=1}^{p}\left(z_{ir} - z_{jr}\right)^2} \tag{6}$$

O agrupamento hierárquico foi realizado pelo método Ward.D2, cujo critério procura minimizar o aumento da soma de quadrados intragrupo a cada fusão. Na Equação (7), A e B são grupos candidatos à fusão, |A| e |B| são suas cardinalidades, e os termos vetoriais com barras representam os centroides dos grupos no espaço padronizado.

$$\Delta(A,B) = \frac{|A||B|}{|A|+|B|}\left\|\bar{\mathbf{z}}_A - \bar{\mathbf{z}}_B\right\|^2 \tag{7}$$

A escolha por Ward.D2 também considerou a discussão de Murtagh e Legendre (2011), que mostram diferenças relevantes entre implementações associadas a Ward.D e Ward.D2, especialmente quanto ao uso de dissimilaridades ao quadrado e às alturas do dendrograma. Foram avaliadas soluções de k = 2 a k = 8. A Equação (8) define a silhouette do município *i*, em que *a(i)* é a dissimilaridade média dentro do próprio grupo e *b(i)* é a menor dissimilaridade média em relação aos demais grupos.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{8}$$

A Equação (9) apresenta o gap statistic, em que W_k mede a dispersão intragrupo para *k* grupos e E_n^* indica a esperança sob amostras de referência. A rotina em linguagem R usou B = 200 simulações, suficiente para avaliação exploratória reprodutível; em estudos

futuros, B pode ser ampliado para 500 ou 1000 para maior estabilidade (Ward, 1963; Murtagh; Legendre, 2011; Rousseeuw, 1987; Tibshirani; Walther; Hastie, 2001).

$$Gap(k) = E_n^*\{\log(W_k)\} - \log(W_k) \qquad (9)$$

Além de Ward.D2, foram ajustados k-means, PAM/k-medoids e modelos de mistura gaussianos. O k-means minimiza o critério J da Equação (10), em que K é o número de grupos, C_k é o k-ésimo grupo, $\boldsymbol{z_i}$ é o vetor padronizado do município i e $\boldsymbol{\mu_k}$ é o centroide do grupo k.

$$J = \min_{C_1,\cdots,C_K} \sum_{k=1}^{K} \sum_{i \in C_k} \|\mathbf{z}_i - \boldsymbol{\mu}_k\|^2 \qquad (10)$$

No ajuste final do k-means, usaram-se nstart = 100 e iter.max = 100 para reduzir a dependência de inicializações aleatórias; no gap statistic, adotou-se nstart = 50. O PAM foi usado como alternativa baseada em medoides, isto é, municípios reais representativos dos grupos. O modelo de mistura gaussiana foi ajustado para G = 2 a G = 8 e selecionado por BIC.

$$BIC_M = -2l_M(\hat{\theta}_M) - m_M \ln(n) \qquad (10)$$

Onde BIC_M é o critério bayesiano de informação do modelo M, $l_M(\hat{\theta}_M)$ é a log-verossimilhança maximizada, m_M é o número de parâmetros livres estimados e n é o número de municípios analisados. Adotou-se o modelo com maior valor de BIC.

As partições foram comparadas pelo Adjusted Rand Index, com a finalidade de medir concordância entre métodos distintos de agrupamento (Lloyd, 1982; Schubert; Rousseeuw, 2021; Fraley; Raftery, 2002; Hubert; Arabie, 1985).

Por fim, foram gerados mapas temáticos para representar o índice socioeconômico municipal e os grupos obtidos pelo método Ward.D2. A união entre base tabular e malha municipal foi feita pela normalização dos nomes dos municípios. Como o agrupamento não impôs restrição de contiguidade espacial, municípios do mesmo grupo podem aparecer territorialmente separados. Caso o objetivo futuro seja produzir regiões socioeconômicas contíguas, será necessário empregar métodos de regionalização espacialmente restritos, como aqueles baseados em grafos de vizinhança e conectividade territorial (Assunção et al., 2006).

RESULTADOS E DISCUSSÃO

O índice socioeconômico municipal variou de 0.2146 a 0.6925, com média de 0.4299 e mediana de 0.4147. A Figura 1 apresenta a distribuição espacial desse índice e mostra que os municípios fluminenses não formam um gradiente territorial simples e contínuo. Valores relativamente mais altos aparecem em áreas metropolitanas, litorâneas, industriais ou associadas a maior dinamismo produtivo. Essa leitura reforça a utilidade de índices compostos como instrumentos de síntese, mas também evidencia que tais índices devem ser interpretados como aproximações dependentes das variáveis, transformações e pesos adotados (Mazziotta; Pareto, 2013; Greco et al., 2019).

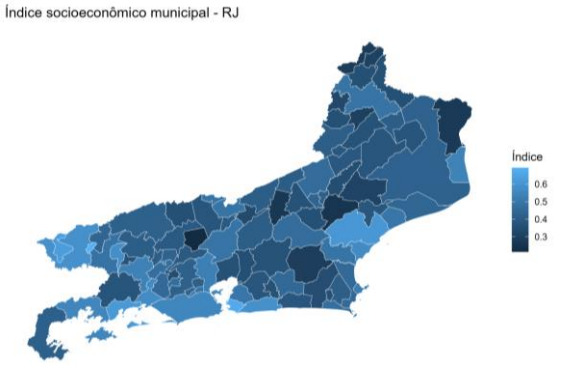

Figura 1. Distribuição espacial do índice socioeconômico municipal no estado do Rio de Janeiro.

A Figura 2 apresenta o biplot da PCA. A primeira componente principal explicou 63.9% da variância total, enquanto a segunda explicou 20.0%; juntas, as duas componentes explicaram 83.9%. Assim, boa parte da estrutura multivariada dos quatro indicadores pode ser examinada em um plano bidimensional. No biplot, IDH-M, remuneração média e PIB per capita transformado apontaram predominantemente para o lado negativo da PC1, enquanto Gini invertido apontou em direção oposta. Como o sinal das componentes principais é arbitrário, a interpretação substantiva é que a PC1 separou municípios com maior renda, maior IDH-M e maior dinamismo econômico daqueles com menor desempenho nesses indicadores, mas também evidenciou que alguns municípios de maior renda apresentam desigualdade mais elevada (Jolliffe; Cadima, 2016; Atkinson, 1970).

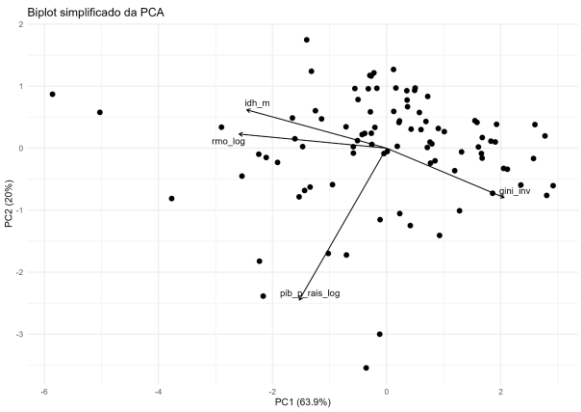

Figura 2. Biplot para os indicadores socioeconômicos padronizados.

A Figura 3 sintetiza a validação interna dos agrupamentos ao mostrar a silhouette média por número de grupos. Para Ward.D2, a maior silhouette média ocorreu em k = 2, com valor igual a 0.3719; o k-means também apresentou melhor resultado em k = 2, com silhouette média de 0.3657. O PAM, por sua vez, teve melhor silhouette em k = 4, com valor de 0.2957, sugerindo uma partição mais fina. Já o gap statistic indicou k = 1 para k-means e PAM, o que sugere que a estrutura de grupos não é fortíssima quando comparada a uma distribuição de referência. Assim, a adoção de k = 2 deve ser entendida como escolha parcimoniosa e interpretável, sustentada sobretudo pela silhouette de Ward.D2 e pela visualização do dendrograma, e não como prova de que existem apenas dois tipos socioeconômicos naturais no estado (Rousseeuw, 1987; Tibshirani; Walther; Hastie, 2001).

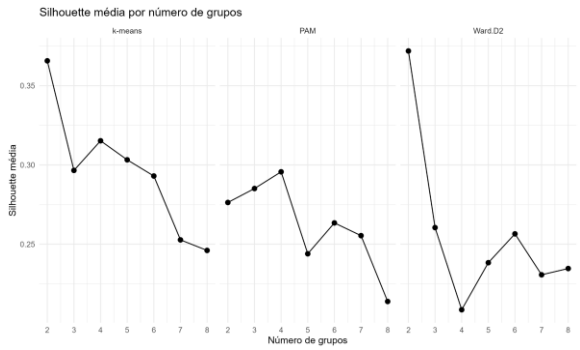

Figura 3. Silhouette média por número de grupos para k-means, PAM e Ward.D2.

No agrupamento hierárquico Ward.D2 com k = 2, o Grupo 1 reuniu 78 municípios e apresentou médias de IDH-M = 0.7007, PIB per capita = 17.16, remuneração média por ocupado = 972.45 e Gini = 0.4836. O Grupo 2 reuniu 14 municípios - Niterói, Itatiaia, Mangaratiba, Macaé, Maricá, Resende, Volta Redonda, Itaguaí, Rio das Ostras, Rio de Janeiro, Angra dos Reis, Piraí, Armação dos Búzios e Três Rios - e apresentou médias de IDH-M = 0.7548, PIB per capita = 38.99, remuneração média por ocupado = 1514.14 e Gini = 0.5236. Portanto, o Grupo 2 não deve ser descrito simplesmente como "melhor": ele reúne municípios com maior renda, maior IDH-M e maior dinamismo produtivo, mas também maior desigualdade média (Atkinson, 1970; Sagar; Najam, 1998).

Tabela 1. Perfil médio dos grupos obtidos pelo método Ward.D2.

Grupo	G_1	G_2
n	78	14
IDH-M médio	0,7007	0,7548
PIB pc médio	17,1617	38,9929
RMO média	972,45	1514,14
Gini médio	0,4836	0,5236
Índice médio	0,4036	0,5764

A Tabela 1 evidencia que os dois grupos não diferem apenas em tamanho, mas também no perfil socioeconômico médio. O Grupo 1 concentrou a maior parte dos municípios analisados, com 78 localidades, e apresentou valores médios inferiores de IDH-M, PIB per capita, remuneração média por ocupado e índice socioeconômico. O Grupo 2, embora menor, com 14 municípios, reuniu localidades com maior desempenho médio nesses indicadores: o IDH-M passou de 0,7007 no Grupo 1 para 0,7548 no Grupo 2; o PIB per capita médio aumentou de 17,1617 para 38,9929; a remuneração média por ocupado passou de 972,45 para 1514,14; e o índice socioeconômico médio elevou-se de 0,4036 para 0,5764.

Entretanto, a esta tabela também mostra que o Grupo 2 não deve ser interpretado simplesmente como o conjunto dos municípios "melhores" em sentido amplo. O coeficiente de Gini médio foi maior nesse grupo, passando de 0,4836 no Grupo 1 para 0,5236 no Grupo 2. Como valores mais altos de Gini indicam maior desigualdade, esse resultado sugere que os municípios com maior dinamismo econômico e melhores médias de desenvolvimento humano e renda também apresentaram, em média, maior concentração de renda. Assim, a tabela reforça a principal interpretação substantiva do estudo: a classificação dos municípios fluminenses exige leitura multidimensional, pois maior desempenho econômico não implica, necessariamente, menor desigualdade social.

A Figura 4 complementa essa leitura ao apresentar o perfil médio padronizado dos grupos. Nela, o Grupo 2 aparece acima da média geral em IDH-M, PIB per capita transformado e remuneração média transformada, mas abaixo da média no Gini invertido, confirmando a combinação entre maior dinamismo socioeconômico e maior desigualdade relativa. O Grupo 1, por sua vez, mostra padrão oposto: menor desempenho médio nos indicadores de renda e desenvolvimento humano, mas posição relativamente mais favorável no indicador invertido de desigualdade (Ceriani; Verme, 2012; Saisana; Saltelli; Tarantola, 2005).

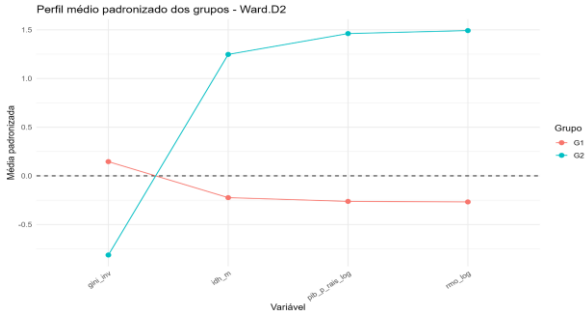

Figura 4. Perfil médio padronizado dos grupos obtidos pelo método Ward.D2.

A Figura 5 mapeia a solução Ward.D2. O Grupo 2 aparece em pontos separados do estado, evidenciando que a classificação não impõe contiguidade: a semelhança foi calculada no espaço dos indicadores, não pela vizinhança geográfica. Para converter essa tipologia em regionalização, seria necessário usar métodos espacialmente restritos, baseados em grafos de vizinhança e conectividade territorial (Assunção et al., 2006).

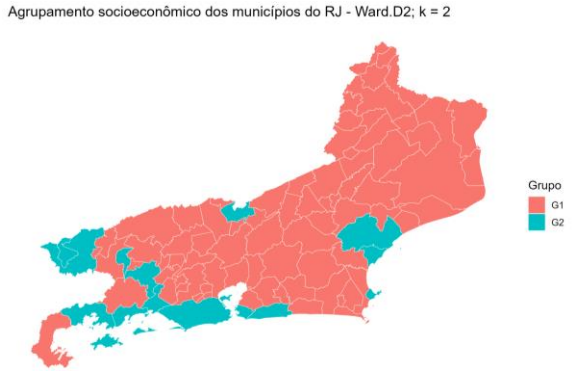

Figura 5. Distribuição espacial dos grupos socioeconômicos obtidos pelo método Ward.D2.

Considerando as partições discutidas nas Figuras 3 a 5, a comparação entre os métodos indicou concordância parcial. O Adjusted Rand Index foi 0.4682 entre Ward.D2 e k-means, 0.4981 entre Ward.D2 e modelos de mistura gaussianos, e 0.5615 entre k-means e modelos de mistura gaussianos. Esses valores sugerem concordância moderada, mas não identidade entre as soluções. Já a comparação entre Ward.D2 e PAM resultou em ARI negativo, indicando discordância expressiva quando o PAM foi forçado a operar com k = 2. Esse comportamento é compatível com a validação interna, pois o PAM havia sugerido melhor desempenho com k = 4. Portanto, para este conjunto de dados, Ward.D2, k-means e modelos de mistura gaussianos convergem para uma separação geral em dois grandes perfis, enquanto o PAM parece captar uma estrutura mais segmentada (Hubert; Arabie, 1985; Lloyd, 1982; Schubert; Rousseeuw, 2021; Fraley; Raftery, 2002).

Os resultados têm limites claros. A análise é transversal e usa apenas quatro indicadores; portanto, não permite inferir causalidade, tendência temporal ou impacto de políticas públicas. Além disso, o índice socioeconômico depende de escolhas metodológicas explícitas, como transformação logarítmica, inversão do Gini, normalização min-max e pesos iguais. Por fim, os agrupamentos são exploratórios e sensíveis ao método escolhido, como demonstrado pelas diferenças entre Ward.D2, k-means, PAM e modelos de misturas gaussianos. Essas limitações não invalidam a análise, mas delimitam seu alcance: o estudo oferece uma tipologia inicial e transparente dos municípios, útil para diagnóstico, e não uma classificação definitiva ou normativa do desenvolvimento municipal (Saisana; Saltelli; Tarantola, 2005; Greco et al., 2019; Rousseeuw, 1987).

CONCLUSÃO

A análise classificou os 92 municípios do estado do Rio de Janeiro a partir de um conjunto reduzido, mas substantivamente relevante, de indicadores socioeconômicos. A PCA mostrou que duas componentes sintetizaram grande parte da variabilidade observada, facilitando a visualização dos municípios e a identificação de perfis contrastantes.

O agrupamento Ward.D2 indicou uma solução parcimoniosa com dois grandes grupos. O primeiro reuniu a maior parte dos municípios e apresentou menor dinamismo médio em IDH-M, PIB per capita e remuneração. O segundo reuniu 14 municípios com maior desempenho médio nesses indicadores, mas também maior desigualdade média medida pelo coeficiente de Gini.

A classificação dos municípios fluminenses não pode ser reduzida a uma oposição simples entre “mais desenvolvidos” e “menos desenvolvidos”. O grupo com maior renda e maior IDH-M também apresentou maior desigualdade média, o que reforça a necessidade de interpretar, em conjunto, produção econômica, renda do trabalho, desenvolvimento humano e distribuição de renda.

Como encaminhamento, recomenda-se ampliar a base com indicadores de educação, saúde, saneamento, pobreza, emprego formal, finanças públicas e dinâmica demográfica. Também seria pertinente testar agrupamentos espacialmente restritos, caso o objetivo passe a ser construir regiões socioeconômicas contíguas. Por fim, sugere-se realizar análise de sensibilidade do índice, variando pesos, métodos de normalização e número de grupos, para avaliar a robustez da classificação obtida.

AGRADECIMENTOS

Y.D.P. agradece à Secretaria de Estado de Desenvolvimento Econômico de Mato Grosso

(SEDEC-MT), Cuiabá, Brasil, pela bolsa de pós-doutorado concedida no âmbito do projeto SEDEC-PRO-2022/02316 (Projeto Número: 3.008.004). S.O.B. agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro concedido por meio de bolsa, processo n. 130848/2026-3.

REFERÊNCIAS

ASSUNÇÃO, R. M.; NEVES, M. C.; CÂMARA, G.; FREITAS, C. C. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. International Journal of Geographical Information Science, v. 20, n. 7, p. 797-811, 2006.

ATKINSON, A. B. On the measurement of inequality. Journal of Economic Theory, v. 2, n. 3, p. 244-263, 1970.

BOX, G. E. P.; COX, D. R. An analysis of transformations. Journal of the Royal Statistical Society: Series B, v. 26, n. 2, p. 211-252, 1964.

CERIANI, L.; VERME, P. The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. The Journal of Economic Inequality, v. 10, p. 421-443, 2012.

FRALEY, C.; RAFTERY, A. E. Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association, v. 97, n. 458, p. 611-631, 2002.

GRECO, S.; ISHIZAKA, A.; TASIOU, M.; TORRISI, G. On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. Social Indicators Research, v. 141, p. 61-94, 2019.

HUBERT, L.; ARABIE, P. Comparing partitions. Journal of Classification, v. 2, p. 193-218, 1985.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A, v. 374, n. 2065, 2016.

LLOYD, S. Least squares quantization in PCM. IEEE Transactions on Information Theory, v. 28, n. 2, p. 129-137, 1982.

MAZZIOTTA, M.; PARETO, A. Methods for constructing composite indices: one for all or all for one? Rivista Italiana di Economia Demografia e Statistica, v. 67, n. 2, p. 67-80, 2013.

MURTAGH, F.; LEGENDRE, P. Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. arXiv:1111.6285v2 [stat.ML], 2011.

POSTALI, F. A. S. Petroleum royalties and regional development in Brazil: the economic growth of recipient towns. Resources Policy, v. 34, n. 4, p. 205-213, 2009.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO; INSTITUTO DE PESQUISA ECONÔMICA APLICADA; FUNDAÇÃO JOÃO PINHEIRO. Atlas do Desenvolvimento Humano no Brasil. [S. l.]: PNUD; Ipea; FJP, [s.d.]. Disponível em: <COLOCAR O ENDEREÇO DELE>. Acesso em: 23 março 2026.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interprSetation and validation of cluster analysis. Journal of Computational and Applied Mathematics, v. 20, p. 53-65, 1987.

SAGAR, A. D.; NAJAM, A. The human development index: a critical review. Ecological Economics, v. 25, n. 3, p. 249-264, 1998.

SAISANA, M.; SALTELLI, A.; TARANTOLA, S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society: Series A, v. 168, n. 2, p. 307-323, 2005.

SCHUBERT, E.; ROUSSEEUW, P. J. Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. Information Systems, v. 101, 101804, 2021.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B, v. 63, n. 2, p. 411-423, 2001.

WARD JR., J. H. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, v. 58, n. 301, p. 236-244, 1963.