Indicadores Sociais e Óbitos por Faixa Etária no Brasil (2022–2023): Uma Análise de Dados do IBGE com Técnicas de Mineração de Dados

Pedro Igor Gomes de Morais
Departamento de Informática
Universidade Federal do Espírito Santo
Vitória, Brasil
pedroigorgm@gmail.com

Matheus Saick de Martin
Departamento de Informática
Universidade Federal do Espírito Santo
Vitória, Brasil
matheus.martin@edu.ufes.br

Renzo Henrique Guzzo Leão
Departamento de Informática
Universidade Federal do Espírito Santo
Vitória, Brasil
renzolealguzzo@gmail.com

***Abstract*—Este trabalho tem como objetivo investigar padrões de mortalidade por faixa etária e causa específica nas unidades federativas brasileiras, com base em dados socioeconômicos do IBGE referentes aos anos de 2022 e 2023. Foram aplicadas técnicas de mineração de dados, como agrupamento, PCA, normalização populacional, detecção de outliers e regressão linear simples. Os resultados mostram que não foram encontrados padrões homogêneos de mortalidade entre estados de uma mesma macro-região, tampouco relações lineares consistentes entre proporção etária e taxas de óbito, quando consideradas causas de óbitos selecionadas. A análise revelou *outliers* nas unidades federativas com indicadores sociais destoantes, associados a fatores estruturais. O estudo reforça a necessidade de políticas públicas sensíveis às desigualdades regionais bem como a necessidade de análises internas de para cada unidade federativa.**

***Index Terms*—mortalidade, desigualdade social, mineração de dados, agrupamento, análise exploratória, IBGE, indicadores socioeconômicos, heurística**

I. Introdução

A pandemia da COVID-19 evidenciou a defasagem das políticas públicas, especialmente aquelas voltadas à saúde, em relação à realidade de indicadores socioeconômicos do país [1]. Tal defasagem contribuiu para a rápida disseminação da doença, tornando o Brasil um dos países mais afetados pela pandemia [2]. Esse cenário é claramente discutido em [3], que demonstra como as vulnerabilidades sociais preexistentes foram determinantes para o agravamento dos impactos da doença.

Esses estudos reforçam a relevância de compreender como desigualdades estruturais, refletidas em indicadores sociais, influenciam diretamente desfechos de saúde. Embora a pandemia tenha evidenciado essas desigualdades de forma aguda, tais disparidades não são exclusivas desse evento [1] e historicamente afetam diversos aspectos da saúde e da mortalidade no Brasil. Por esse motivo, é fundamental dar continuidade a essa linha de investigação, buscando ampliar o entendimento das relações entre condições socioeconômicas e os desfechos de saúde em diferentes contextos e períodos.

Nesse sentido, este trabalho se propõe a aplicação de métodos de mineração de dados bem estabelecidos na literatura, sobre indicadores sociais e mortalidades por diferentes causas selecionadas e/ou faixas etárias nas unidades federativas brasileiras, com base em dados tabulares oficiais do Instituto Brasileiro de Geografia e Estatística (IBGE)[1]. O objetivo é identificar padrões nos números de óbitos regionais, bem como os principais indicadores sociais associados, oferecendo subsídios para a compreensão das desigualdades socioeconômicas entre as regiões do país. Portanto, por meio de uma análise exploratória, espera-se que os resultados contribuam para reforçar a importância de políticas públicas com especificidades regionais e sociais que possuem maior probabilidade de mitigarem as desigualdades que impactam diretamente a quantidade de óbitos no Brasil.

II. Perguntas de Pesquisa e Metodologia

Para fins de direcionamento do trabalho, foram definidas as seguintes perguntas de pesquisa (PP):

- **PP1:** Unidades federativas de uma mesma grande região possuem padrões semelhantes de óbitos?
- **PP2:** Existe uma tendência linear entre a proporção da população de uma unidade federativa em uma determinada faixa etária e o número de óbitos por uma causa específica nessa mesma faixa etária, ajustada pela população correspondente? Essa pergunta de pesquisa pode ser descrita como a seguinte equação:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

$$x_i = \frac{P_{i,a}}{P_i} \tag{2}$$

$$y_i = \frac{D^C_{i,a}}{P_{i,a}} \tag{3}$$

Onde:

- i: índice da unidade federativa;
- a: faixa etária considerada;

[1] https://www.ibge.gov.br/

- C: causa específica de óbito;
- $P_{i,a}$: população da unidade federativa i na faixa etária a;
- P_i: população total da unidade federativa i;
- $D_{i,a}^{C}$: número de óbitos por causa C na faixa etária a na unidade federativa i.
- β_0, β_1: coeficientes da regressão linear;
- ε_i: termo de erro aleatório.

Ao considerar as perguntas de pesquisa, esse estudo foi organizado nas seguintes etapas principais: revisão de trabalhos relacionados, apresentação do procedimento de coleta e critérios de seleção dos dados, escolhas das técnicas de mineração de dados, e, por fim, uma breve discussão dos resultados, conclusão e apresentação de possibilidades de trabalhos futuros.

III. Trabalhos relacionados

É notável que estudos existentes se propõem a investigar hipóteses sobre relações que permeiam indicadores socioeconômicos e óbitos de causas diversas, como a COVID-19, com resultados que corroboram as hipóteses das desigualdades sociais serem associadas à mortalidade de diferentes origens e à expectativa de vida da população. Avançando além das análises tradicionais de mortalidade, pesquisas recentes utilizam o indicador de Expectativa de Vida Saudável (EVS) objetivando retratar que desigualdades se acentuam ao se considerar a qualidade de vida, além de disparidades significativas entre as Unidades Federativas e estratos sociais [4], [5].

Comumente, em campos de investigação metodológicos que visam analisar relações entre desigualdade social e mortalidade, indicadores como índice de desenvolvimento humano (IDH), renda per capita e escolaridade têm sido utilizados de forma exaustiva, ainda que os efeitos desses fatores variem mediante a causa específica e o recorte geográfico considerado [6], [7]. Além disso, abordagens estatísticas específicas para observar localizações geográficas próximas, i.e., cidades e unidades federativas, são empregadas para corroborar conexões entre variáveis socioeconômicas relacionadas aos padrões espaciais de mortalidade [7], [8].

Todavia, apesar desses trabalhos consolidarem a importância de indicadores socioeconômicos nas disparidades entre óbitos em unidades federativas, carece-se ainda de indicadores que expliquem a complexidade da desigualdade em diferentes micro e macro regiões do Brasil. Diante disso, análises exploratórias sobre dados socioeconômicos brasileiros, como as que são propostas neste trabalho, permanecem indispensáveis no aprofundamento da compreensão dessas relações e no auxílio à criação de políticas públicas mais eficazes nas unidades federativas do Brasil.

IV. Coleta e seleção dos dados

Os dados obtidos para esse trabalho foram coletados a partir da base de dados do IBGE, especificamente, todas as 147 tabelas da base *Síntese de Indicadores Sociais* [9] e dados do censo demográfico de 2022 [10]. As informações contidas nessas bases abrangem indicadores relacionados à estrutura econômica, mercado de trabalho, padrão de vida, distribuição de rendimentos, condições de moradia, educação, condições de saúde da população do Brasil, e o quantitativo de pessoas vivas por faixa etária das unidades federativas e das cidades. A seleção dos dados foi realizada em quatro etapas, com o objetivo de refinar e estruturar os dados para os experimentos.

- **Etapa 1 — Seleção inicial de tabelas**: Das 147 tabelas iniciais, foram excluídas aquelas que não continham separação por unidades federativas ou que apresentavam uma quantidade majoritária de dados ausentes (acima de 50% de dados faltantes por coluna).Com a aplicação desses critérios, somente 91 das tabelas iniciais permaneceram.
- **Etapa 2 — Compatibilidade de faixas etárias**: Foram eliminadas tabelas que não contemplavam pelo menos três faixas etárias distintas ou que apresentavam dados redundantes em relação a outras tabelas mantidas. Ao final, manteram-se 67 tabelas.
- **Etapa 3 — Recorte temporal**: Para preservar a consistência temporal, foram excluídas as tabelas que não continham dados de 2022 e 2023, restando 66 tabelas.
- **Etapa 4 — Integração**: Por fim, as tabelas selecionadas foram integradas por unidade federativa. Desta forma, obteve-se uma base de dados consolidada de atributos por estado para os anos de 2022 e 2023 com os indicadores socioeconômicos a serem analisados.

V. Aplicação de técnicas de mineração de dados

Para os anos de 2022 e 2023, a primeira etapa da análise exploratória foi conduzida sobre os dados da *Tabela 5.2 – Número de óbitos por causas selecionadas e grupos de idade, segundo Grandes Regiões, Unidades da Federação e Municípios das Capitais – Brasil*, considerando apenas as unidades da federação. Primeiro, foi aplicado um ajuste de escala pela população relativa à unidade federativa e faixa-etária considerada. É importante salientar que para o ano de 2023, a população precisou ser estimada através do uso da taxa de crescimento geométrica do Brasil do ano de 2022, visto que o IBGE não disponibilizou os dados populacionais por unidade federativa para o ano de 2023 tampouco a taxa de crescimento geométrica por unidade federativa, e nem mesmo a estimativa populacional de 2023 por unidade federativa, embora a estimativa para o ano de 2024 estivesse disponível. Segue , abaixo, a estimativa por faixa etária por unidade federativa em 2023.

Seja $P_{i,e,2022}$ a população da **unidade federativa** i na **faixa etária** e, no ano de 2022. Seja $\widehat{P}_{i,e,2023}$ a estimativa correspondente para 2023. Supondo uma taxa de crescimento geométrico constante $r = 0.0052$, e considerando $t = 1$ ano, a população pode ser estimada da seguinte forma:

$$\widehat{P}_{i,e,2023} - P_{i,e,2022} = P_{i,e,2022} \cdot (1+r)^1$$

$$\widehat{P}_{i,2023} - P_{i,2022} = (\sum_{e} \widehat{P}_{i,e,2023}) - P_{i,2022} = \sum_{e} P_{i,e,2022} \cdot (1+r)$$

$$\widehat{P}_{i,2023} - P_{i,2022} = (1+r) \cdot \sum_{e} P_{i,e,2022} = (1+r) \cdot P_{i,2022}$$

$$\widehat{P}_{\text{Brasil},2023} - P_{\text{Brasil},2022} = \sum_{i} \sum_{e} \widehat{P}_{i,e,2023} = \sum_{i} \sum_{e} P_{i,e,2022} \cdot (1+r)$$

$$\widehat{P}_{\text{Brasil},2023} = (1+r) \cdot \sum_{i} \sum_{e} P_{i,e,2022} = (1+r) \cdot P_{\text{Brasil},2022}$$

No entanto, essa aproximação, apesar de válida para o total do Brasil, pode gerar distorções na estimativa da população por faixa etária e por unidade federativa, uma vez que:

- A taxa de crescimento r foi considerada igual para todas as faixas etárias e unidades federativas;
- Na prática, faixas etárias e regiões não necessariamente crescem no mesmo ritmo.

Em seguida, foram aplicados algoritmos de agrupamento com o objetivo de identificar padrões de macro-regiões [11]:

- Agrupamento aglomerativo utilizando os critérios *ward* [12], completo e média [13].
- *Affinity Propagation* [14], configurado com até 200 iterações, fator de amortecimento (*damping*) igual a 0.5, 15 iterações para convergência e distância euclidiana.
- Para o algoritmo K-Means [15], foi feita a análise do *silhouette score* [16] para diferentes valores de número de grupos, foi definido $k = 4$ considerando a análise.

Para permitir a visualização dos agrupamentos obtidos, foi empregada a Análise de Componentes Principais (PCA), com a seleção de duas componentes principais. Com o intuito de facilitar a interpretação da separação entre os grupos e identificar variáveis representativas, adotou-se a seguinte heurística: para cada componente, foram selecionadas as duas variáveis referentes aos dois maiores coeficientes em valor absoluto da combinação linear da componente em questão. Dessa forma, obteve-se um conjunto de quatro variáveis mais influentes.

Para investigar a presença de possíveis *outliers*, construiu-se uma visualização em formato de mapa de calor com base nessas variáveis. Considerando que todos os valores estão definidos no intervalo $[0, 1]$, a detecção de *outliers* seguiu uma metodologia robusta, baseada na regra do intervalo interquartil (*Interquartile Range* – IQR), amplamente utilizada na literatura [17], com os seguintes limiares:

$$\text{Limiar superior} = Q_3 + 1{,}5 \times \text{IQR}$$

$$\text{Limiar inferior} = Q_1 - 1.5 \times \text{IQR}$$

Em que Q_1 e Q_3 representam, respectivamente, o primeiro e o terceiro quartis da distribuição da variável, e $\text{IQR} = Q_3 - Q_1$. Como as variáveis estão restritas ao intervalo [0, 1], eventuais valores negativos no limiar inferior foram desconsiderados.

Também foram construídos modelos de regressão linear simples para cada uma das causa de mortalidade por intervalo de idade, seguindo a definição apresentada pela **PP2**. Em um cenário sem diferenças significativas entre unidades federativas, espera-se que o modelo linear seja razoável. Vale ressaltar que, para todas as regressões, os respectivos *outliers* de cada causa por faixa etária identificados, através dos limiares definidos, não foram removidos para construção dos respectivos modelos.

Para complementar a análise dos possíveis motivos dos fenômenos de óbito por faixa etária, foi feito de forma análoga ao método de seleção das principais características da tabela de óbitos, a realização da extração de variáveis por meio do PCA aplicada à tabela unificada, composta pelas demais 65 tabelas (com sua devida escala ajustada), selecionando um número de componentes capazes de capturar mais de 90% da variância explicada. Para cada componente, identificou-se a variável original (isto é, a coluna da tabela unificada) referente ao coeficiente de maior contribuição em valor absoluto para a combinação linear da componente. Em seguida, para cada uma dessas variáveis selecionadas, foi gerado um gráfico de barras com ordenação não decrescente para tentar identificar características dos *outliers* resultantes.

VI. Discussão dos Resultados

Para o ano de 2022, os agrupamentos não conseguem capturar diferenças no padrão de óbitos entre unidades federativas de uma mesma grande região, assim como pode ser visto na figura 1.

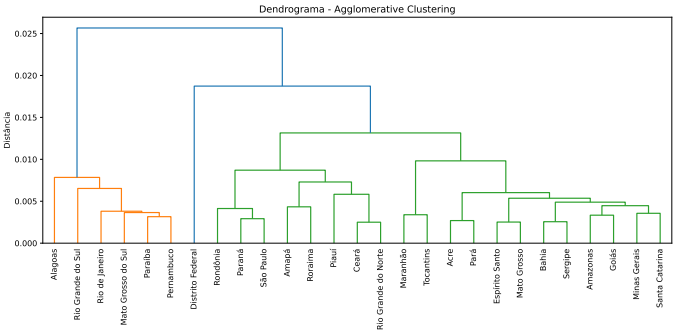

Fig. 1. Agrupamento aglomerativo (ward) - 2022

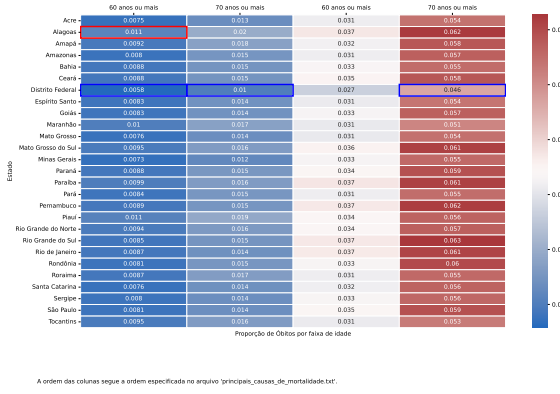

Fig. 2. Mapa de calor - 2022

A partir do mapa de calor da figura 2, é possível observar os *outliers* positivos e negativos, destacados em retângulos de cor azul e vermelha, respectivamente, podemos destacar

o estado de Alagoas para *doenças hipertensivas, isquemicas do coração, insuficiência cardíaca e cerebrovasculares* no grupo de idade de 60 anos ou mais, bem como o Distrito Federal para *doenças hipertensivas, isquemicas do coração, insuficiência cardíaca e cerebrovasculares* no grupo de idade de 60 anos ou mais, e, também, para o número de óbitos total da população de 60 anos ou mais e 70 anos ou mais. Se considerarmos as características selecionadas pela heurística sobre as 65 tabelas de indicadores sociais, podemos destacar que o estado de Alagoas possui um número levemente superior de número mensal médio de leitos de internação pelo SUS a cada dez mil habitantes, vide figura 4. De forma análoga, na figura 3 a capital do Brasil apresenta uma maior concentração de renda de pessoas sem benefícios de programas sociais governamentais.

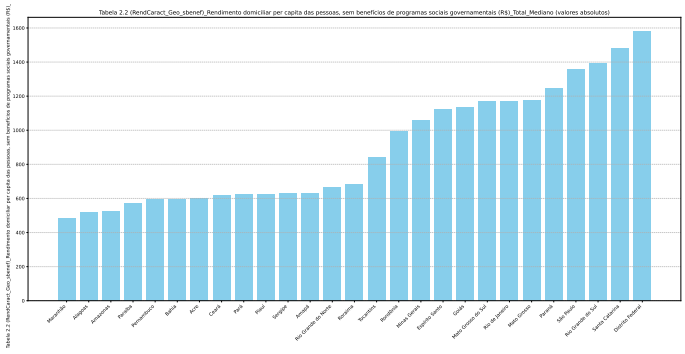

Fig. 3. Rendimento Domiciliar per capita (pessoas sem benefícios sociais governamentais) - 2022

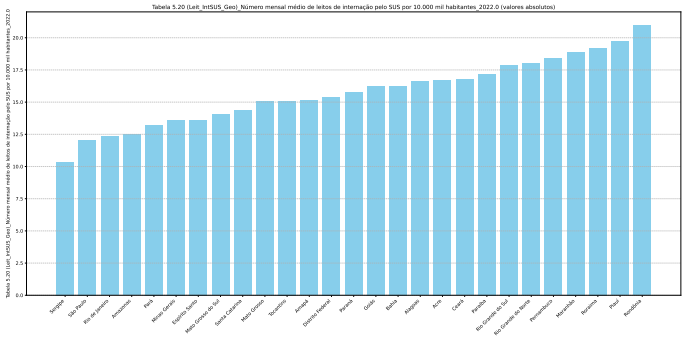

Fig. 4. Número mensal médio de leitos de internação pelo SUS a cada 10 mil habitantes - 2022

Para os 4 modelos de regressão referentes ao ano de 2022, nenhum deles traz evidência da existência de uma tendência linear, visto que todos os R^2 ficaram iguais ou inferiores a 0.270. [2]

Para o ano de 2023 os agrupamentos também não conseguiram capturar evidências de diferenças no padrão de óbitos entre unidades federativas de uma mesma grande região, assim como pode ser visto na figura 5. O número de grupos foi definido como 5, tendo em vista a análise da silhueta.

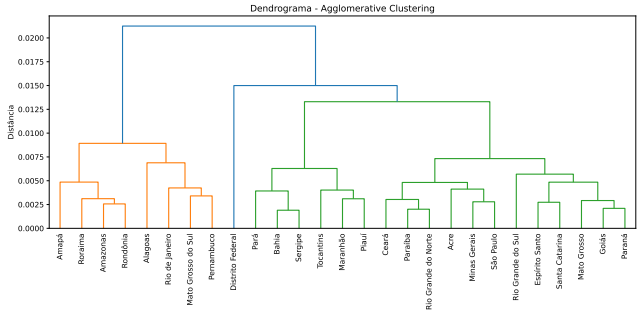

Fig. 5. Agrupamento aglomerativo (ward) – 2023

Para os 4 modelos de regressão referentes ao ano de 2023, nenhum dos modelos apresentou tendência linear (i.e. todos os R^2 abaixo de 0.175).

Em contrapartida, no mapa de calor da figura 6, é possível observar que o Distrito Federal, Maranhão, Rio Grande do Sul e Rio de Janeiro são destacados como outliers. O estado do Rio de Janeiro, destaca-se pelo percentual de pessoas quase idosas que exercem atividade remunerada, vide figura 7. Já o Rio Grande do Sul, destaca-se por um maior número mensal médio de leitos de internação pelo SUS a cada 10 mil habitantes como pode ser observado na figura9, assim como um número inferior de percentual de pessoas com 25 anos ou mais com apenas o ensino médio completo visto na figura 8, em contrapartida, o Rio de Janeiro destaca-se por um maior percentual de pessoas apenas com ensino médio completo nessa mesma faixa-etária como também pode ser observado na figura 8.

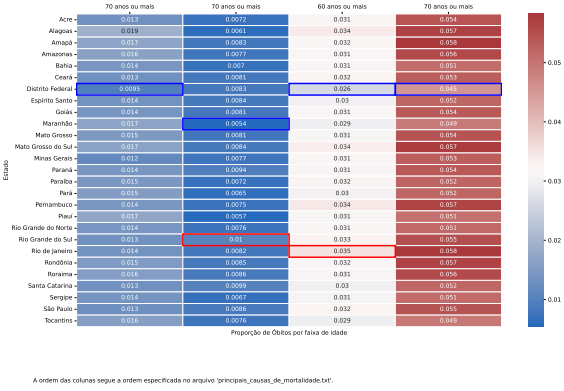

Fig. 6. Mapa de calor - 2023

A partir do destaque dos *outliers*, pode-se formular hipóteses para análises futuras à associação destes fatos com a influência do movimento anti-vacina no Brasil [18], visto que existem esforços constantes do poder público e no ambiente de pesquisa, para campanhas e desenvolvimento de técnicas [19] de combate a esses movimentos.

[2]Os gráficos foram omitidos por falta de espaço, as figuras estão disponíveis no repositório do projeto https://github.com/intel-comp-saude-ufes/2025-1-P1-IBGE-obitos-em-unidades-federativas

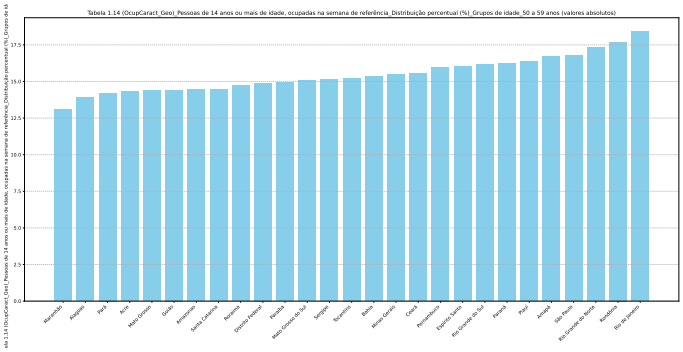

Fig. 7. Distribuição percentual (50-59 anos) - 2023

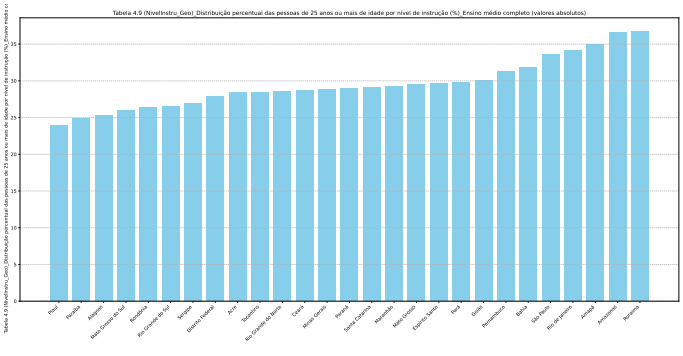

Fig. 8. Percentual de pessoas de 25 anos ou mais com apenas o ensino médio completo - 2023

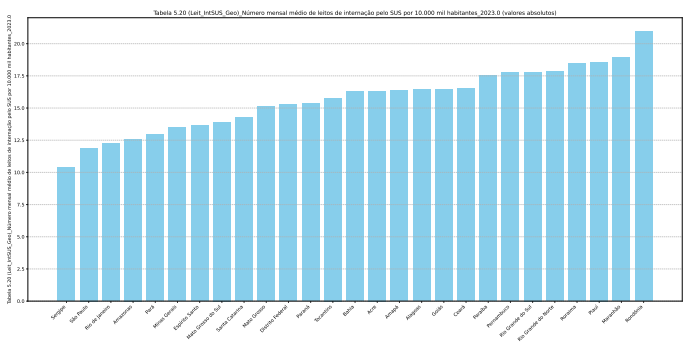

Fig. 9. Número mensal médio de leitos de internação pelo SUS a cada 10 mil habitantes - 2023

VII. Conclusão

Este trabalho propôs uma abordagem exploratória para análise de padrões de mortalidade por faixa etária e causa específica nas unidades federativas brasileiras, com base em indicadores socioeconômicos do IBGE referentes a 2022 e 2023. A metodologia integrou técnicas de agrupamento, redução de dimensionalidade, normalização populacional, detecção de outliers e regressão linear, configurando uma estratégia válida para a identificação de desigualdades regionais em saúde.

Os resultados das estratégias de agrupamento, demonstram que com a abordagem utilizada, os algoritmos selecionados não foram capazes de identificar padrões de óbitos entre unidades federativas dentro de uma mesma macro-região. Ademais, observou-se que a proporção da população em uma faixa etária não implica, necessariamente, numa relação linear com as taxas de óbito por causas específicas.

Apesar dos resultados apresentados neste estudo, a fim de aprofundar a caracterização das vulnerabilidades regionais e subsidiar políticas públicas mais precisas e equitativas, destacam-se as seguintes limitações deste trabalho e consequentes possibilidades de trabalhos futuros: a necessidade de estimativas populacionais para o ano de 2023, a curta janela de evolução temporal dos fenômenos analisados, bem como a restrição a somente uma base de dados de indicadores sociais (i.e. somente uso da base do IBGE), e por fim a ausência de análises por unidade federativa, visto a possibilidade de diferenças internas.

References

[1] P. H. Souza, "A history of inequality: Top incomes in brazil, 1926–2015," *Research in Social Stratification and Mobility*, vol. 57, pp. 35–45, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0276562417302068

[2] J. L. Kephart, X. Delclòs-Alió, D. A. Rodríguez, O. L. Sarmiento, T. Barrientos-Gutiérrez, M. Ramirez-Zea, D. A. Quistberg, U. Bilal, and A. V. D. Roux, "The effect of population mobility on covid-19 incidence in 314 latin american cities: a longitudinal ecological study with mobile phone location data," *The Lancet Digital Health*, vol. 3, pp. e716–e722, 11 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/34456179/

[3] J. F. d. Sousa Filho, U. M. Silva, L. L. Lima, A. S. S. Paiva, G. F. Santos, R. F. S. Andrade, N. Gouveia, I. H. Silveira, A. A. de Lima Friche, M. L. Barreto, and W. T. Caiaffa, "Association of urban inequality and income segregation with covid-19 mortality in brazil," *PLOS ONE*, vol. 17, no. 11, pp. 1–12, 11 2022. [Online]. Available: https://doi.org/10.1371/journal.pone.0277441

[4] C. L. Szwarcwald, D. E. R. Montilla, A. P. Marques, G. N. Damacena, W. d. S. d. Almeida, and D. C. Malta, "Inequalities in healthy life expectancy by federated states," *Revista de Saúde Pública*, vol. 51, no. Suppl 1, p. 7s, 2017.

[5] C. L. Szwarcwald, W. d. S. d. Almeida, P. R. B. d. Souza Júnior, J. M. Rodrigues, and D. E. Romero, "Socio-spatial inequalities in healthy life expectancy in the elderly, brazil, 2013 and 2019," *Cadernos de Saúde Pública*, vol. 38, no. Sup 1, p. e00124421, 2022.

[6] A. C. d. O. Costa, D. d. O. Ramos, and R. Paes de Sousa, "Indicators of social inequalities associated with cancer mortality in brazilian adults: scoping review," *Ciência & Saúde Coletiva*, vol. 29, no. 8, p. e19602022, 2024.

[7] V. Dias Marques, M. Massago, M. T. da Silva, I. Roskowski, D. A. N. de Lima, L. dos Santos, E. Louro, S. Tomás Gonçalves, R. B. Pedroso, A. M. Obale, S. M. Pelloso, J. R. N. Vissoci, C. A. Staton, O. K. Nihei, M. D. de Barros Carvalho, A. d. C. Dutra, and L. de Andrade, "Exploring regional disparities in lung cancer mortality in a brazilian state: A cross-sectional ecological study," *PLOS ONE*, vol. 18, no. 6, p. e0287371, 06 2023.

[8] A. d. C. Dutra, L. L. Silva, R. B. Pedroso, Y. P. Tchuisseu, M. T. da Silva, M. Bergamini, J. F. H. C. Scheidt, P. H. Iora, R. d. L. Franco, C. A. Staton, J. R. N. Vissoci, O. K. Nihei, and L. de Andrade, "The impact of socioeconomic factors, coverage and access to health on heart ischemic disease mortality in a brazilian southern state: A geospatial analysis," *Global Heart*, vol. 16, no. 1, p. 5, 01 2021.

[9] Instituto Brasileiro de Geografia e Estatística (IBGE), "Síntese de indicadores sociais," https://www.ibge.gov.br/estatisticas/sociais/saude/9221-sintese-de-indicadores-sociais.html, 2023, accessed: Jul. 18, 2025.

[10] ——. (2024) Tabela 9514 - população residente, por sexo, idade e forma de declaração da idade. Acesso em: 20 jul. 2025. [Online]. Available: https://sidra.ibge.gov.br/tabela/9514

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Duchesnay,

"Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[12] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[13] R. R. Sokal and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas Scientific Bulletin*, vol. 38, pp. 1409–1438, 1958.

[14] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1136800

[15] X. Jin and J. Han, *K-Means Clustering*. Boston, MA: Springer US, 2010, pp. 563–564. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_425

[16] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[17] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*. Wiley, 2014.

[18] F. Malini, F. Sodré, A. Cavalini, G. Herkenhoff, and F. Goveia, *Five Patterns of Vaccine Misinformation on Telegram*, 01 2025, pp. 181–196.

[19] A. Cavalini, T. Donadia, F. Malini, and G. Comarela, "Detecting misinformation on telegram anti-vaccine communities," in *Anais do XXXIX Simpósio Brasileiro de Bancos de Dados*. Porto Alegre, RS, Brasil: SBC, 2024, pp. 729–735. [Online]. Available: https://sol.sbc.org.br/index.php/sbbd/article/view/30739