

Análise de Dados Sobre Suicídio no Brasil: Identificação de Padrões

Érika Regina de Souza
Dep. de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
erikars@ufmg.br

Resumo— Este trabalho apresenta uma análise de dados sobre suicídio e tentativas de suicídio no Brasil, utilizando bases públicas do DATASUS e indicadores socioeconômicos do IBGE. A pesquisa examinou tendências temporais, diferenças entre faixas etárias e regiões, fatores associados à recorrência e perfis de risco identificados por meio de técnicas de clusterização. Os resultados mostram crescimento contínuo das taxas de óbitos e tentativas nas últimas décadas, especialmente entre adultos e em regiões como Sul e Centro-Oeste. A recorrência também aumentou de forma significativa, indicando maior vulnerabilidade entre pessoas que já tiveram uma primeira tentativa. Fatores socioeconômicos, educacionais e de saúde mental mostraram associações variadas, destacando a complexidade do fenômeno. Os achados reforçam a importância de políticas públicas direcionadas e de monitoramento contínuo para prevenção do suicídio no país.

Palavras-chave — suicídio; tentativas de suicídio; DATASUS; IBGE; saúde mental; saúde pública; clusterização.

I. INTRODUÇÃO

O suicídio é reconhecido pela Organização Mundial da Saúde como um grave problema de saúde pública, responsável por mais de 720 mil mortes por ano no mundo [1]. No Brasil, tanto os óbitos quanto as tentativas de suicídio apresentam crescimento contínuo, o que reforça a necessidade de estudos que permitam compreender melhor seus padrões, fatores associados e grupos mais vulneráveis.

O avanço das tecnologias de informação e a disponibilização de bases públicas de saúde possibilitam análises mais amplas sobre o tema. Pesquisas recentes utilizam desde dados clínicos e sensores móveis [2] até modelos causais construídos com especialistas [3]. No contexto brasileiro, a base BrSuicides [4] tem se destacado por organizar, de forma acessível, dados do Sistema de Informações sobre Mortalidade (SIM) relacionados a óbitos por suicídio.

Além dos fatores individuais, o suicídio está ligado a aspectos sociais, econômicos, demográficos e culturais. Estudos mostram que escolaridade, renda, desigualdade regional e prevalência de transtornos mentais influenciam o comportamento suicida [1], [5]. Outras pesquisas analisam padrões de risco por meio de técnicas de clusterização [6], [7], enquanto revisões sistemáticas reforçam a importância da análise temporal [8].

Nesse contexto, este trabalho tem como objetivo investigar padrões de suicídio e tentativas no Brasil, integrando dados do DATASUS e indicadores socioeconômicos do IBGE. A pesquisa inclui análises temporais em janelas de 10 anos, estudo de perfis sociodemográficos, associação com fatores socioeconômicos e modelagem de recorrência por regressão logística, além da identificação de grupos de risco por técnicas de clusterização. Os resultados buscam contribuir para uma

compreensão mais ampla do fenômeno e apoiar estratégias de prevenção em saúde pública.

II. REFERENCIAL TEÓRICO

O suicídio é reconhecido pela Organização Mundial da Saúde como um grave problema de saúde pública [1]. Nesse contexto, diversos estudos vêm sendo desenvolvidos para compreender os fatores que influenciam sua ocorrência e buscar formas de prevenção.

Pesquisas recentes exploram o uso de tecnologias avançadas e dados de sensores em smartphones para prever riscos de comportamento suicida [2]. Outras abordagens utilizam modelos causais construídos com especialistas para mapear relações entre fatores sociais e emocionais associados ao suicídio [3]. No contexto brasileiro, o conjunto de dados BrSuicides se destaca como uma ferramenta relevante para a análise de casos diários, permitindo observar padrões demográficos, temporais e geográficos a partir de registros oficiais do DATASUS [4].

Os estudos citados utilizam diferentes tipos de dados e metodologias. Trabalhos baseados em dados ecológicos e sensores móveis permitem acompanhamento individual em tempo real [2], enquanto modelos causais derivados de entrevistas com profissionais de saúde mental identificam ciclos de retroalimentação entre variáveis emocionais e sociais [3]. Já o BrSuicides fornece um conjunto de dados estruturado com variáveis como sexo, idade, escolaridade, local da ocorrência e data do óbito, permitindo análises populacionais robustas [4].

A análise desses dados permite a identificação de diversos padrões. Nos trabalhos com dados quantitativos, é possível observar tendências sazonais, variações entre regiões e diferenças significativas entre faixas etárias e gêneros. Os estudos com foco qualitativo, como [3], identificam ciclos de retroalimentação entre fatores emocionais e sociais, como isolamento, estigma e histórico de trauma. Tais padrões ajudam a compreender tanto os fatores de risco objetivos quanto as dinâmicas sociais e psicológicas que envolvem o comportamento suicida.

O uso do conjunto de dados BrSuicides [4] como base desta pesquisa se justifica por sua abrangência, qualidade e origem confiável. Ele foi construído com base nos registros oficiais do Sistema de Informações sobre Mortalidade (SIM), disponibilizados pelo DATASUS, o que garante a confiabilidade e a padronização dos dados.

O BrSuicides contém dados com abrangência geográfica de todo o território brasileiro e é mantido atualizado pelos autores do estudo. Assim, ele permite a análise de variáveis demográficas, temporais e geográficas relacionadas aos casos de suicídio, sendo especialmente útil para estudos populacionais. Além disso, seu formato acessível e bem documentado facilita a manipulação e cruzamento com

outros conjuntos de dados, como outras bases do DATASUS ou do IBGE.

Outro aspecto relevante na literatura é a recorrência de tentativas de suicídio, reconhecida como um dos fatores de risco mais importantes para o óbito futuro [1]. Estudos demonstram que pessoas com tentativas prévias apresentam risco significativamente maior de novas ocorrências, o que torna fundamental caracterizar esse grupo em análises epidemiológicas. Além dos fatores individuais, o suicídio também está relacionado a determinantes sociais da saúde.

Pesquisas brasileiras apontam que variáveis como renda, escolaridade e desigualdade regional estão associadas às taxas de suicídio [5], [9]. A integração de bases do IBGE permite analisar esse contexto social e estrutural, enriquecendo a interpretação dos dados do DATASUS.

No campo metodológico, técnicas de agrupamento têm sido utilizadas para identificar perfis distintos de risco em populações específicas. Alguns estudos classificam indivíduos com ideação suicida em grupos com características clínicas e sociodemográficas semelhantes [6], enquanto outros identificam clusters espaço-temporais de suicídio, evidenciando desigualdades regionais e padrões de concentração dos casos [7]. Essas abordagens revelam combinações de fatores que não seriam perceptíveis em análises isoladas.

Por fim, a análise temporal de longo prazo, estruturada em janelas de 10 anos, tem sido amplamente utilizada em estudos epidemiológicos para identificar mudanças nos perfis de risco ao longo das décadas. Revisões sistemáticas destacam o uso desse tipo de abordagem para observar variações geracionais e avaliar o impacto de transformações sociais ou políticas públicas sobre os casos de suicídio [8].

III. METODOLOGIA

A primeira etapa do trabalho utilizou uma abordagem exploratória, com base na análise de dados de quatro bases públicas do sistema de saúde brasileiro: o Sistema de Informações sobre Mortalidade (SIM), o Sistema de Informação de Agravos de Notificação (SINAN), o Sistema de Informações Hospitalares (SIH) e o Sistema de Informações Ambulatoriais (SIA).

Nesta segunda etapa da pesquisa, foram utilizados principalmente os dados do Sistema de Informações sobre Mortalidade (SIM), por meio da base BrSuicides, do Sistema de Informação de Agravos de Notificação (SINAN), coletados do DATASUS, e dos dados do IBGE, coletados do sistema SIDRA.

A. Coleta de dados do DATASUS

A coleta dos dados foi feita diretamente no servidor do DATASUS por acesso via cliente FTP (File Transfer Protocol). Em seguida, foi utilizado o programa TabWin¹ para importar os arquivos baixados em formato “.dbc” e convertê-los para o formato “.dbf”. Ainda utilizando o TabWin, os arquivos no formato final foram filtrados e limpos para serem utilizados de forma adequada nesta pesquisa.

¹ TabWin:

<http://siab.datasus.gov.br/DATASUS/index.php?area=0608>

Os dados analisados compreendem o período de 2009 a 2023, por ser um período com dados contidos em todas as bases analisadas. Para garantir que os dados analisados estivessem de fato relacionados a tentativas ou óbitos por suicídio, foi utilizado como critério de filtragem o intervalo de códigos CID-10 X60 a X84, que corresponde às lesões autoprovocadas intencionalmente.

Especificamente no caso dos dados do SINAN, foi utilizado um filtro para considerar somente os registros que tiveram preenchidos corretamente o campo correspondente a lesão provocada pela própria vítima e o campo de número de agentes envolvidos definido como 1.

B. Coleta de dados do IBGE

A coleta de dados sociodemográficos foi feita via web pelo Sistema IBGE de Recuperação Automática (SIDRA), que armazena e disponibiliza dados de pesquisas realizadas pelo Instituto. Alguns dados são disponibilizados a nível municipal e outros a nível estadual e regional. Além disso, maioria dos dados disponibilizados são de pesquisas referentes aos anos de 2013 e 2019.

No Acervo, foram listadas as tabelas de dados agregados e foram selecionadas oito delas para serem analisadas nesta pesquisa, conforme a Tabela I. Em todos os casos, buscou-se utilizar as versões mais recentes de cada um dos agregados, ignorando-se as séries encerradas.

TABELA I. AGREGADOS IBGE ANALISADOS

Tabela de dados agregados	
<i>Id</i>	<i>Descrição</i>
4099	Taxas de desocupação e de subutilização da força de trabalho
4694	Pessoas de 18 anos ou mais de idade que referem diagnóstico de depressão por profissional de saúde mental
5885	Municípios com ou sem estrutura organizacional nas áreas de educação, cultura, esporte e saúde
5938	PIB a preços correntes
6579	População residente estimada por município
7903	Pessoas de 18 anos ou mais de idade que referem diagnóstico de outra doença mental (esquizofrenia, transtorno bipolar, psicose ou TOC) por profissional de saúde mental
10139	Taxa de escolarização das pessoas de 6 anos ou mais de idade
10061	Pessoas de 18 anos ou mais de idade, por nível de instrução

C. Análise dos dados

A análise dos dados do SINAN e IBGE foi realizada com foco na caracterização dos óbitos e das tentativas de suicídio no Brasil, considerando variáveis sociodemográficas, dados da Pesquisa Nacional de Saúde (PNS) e contextuais, como sexo, faixa etária, escolaridade, método utilizado, localização e recorrência.

Após a padronização e limpeza dos registros, foram produzidas estatísticas descritivas, gráficos comparativos e análises temporais para identificar padrões consistentes entre os diferentes grupos populacionais.

A análise de recorrência das tentativas de suicídio foi realizada a partir das variáveis disponibilizadas no SINAN, especificamente o campo que identifica se a pessoa já havia realizado uma ou mais tentativas anteriores.

Inicialmente, os registros foram filtrados para remover valores ignorados ou inconsistentes, garantindo que apenas notificações com informação válida sobre recorrência fossem utilizadas. Em seguida, os casos foram classificados em três categorias: tentativas recorrentes, não recorrentes e ignoradas.

A partir dessa classificação, foram geradas distribuições por sexo, faixa etária, método utilizado e região, permitindo identificar os grupos mais afetados pela repetição de tentativas. Também foram produzidas análises temporais, avaliando a evolução da recorrência ao longo dos anos, além de comparações entre perfis sociodemográficos. Para complementar essa etapa, aplicou-se um modelo de regressão logística com o objetivo de investigar fatores associados à recorrência, considerando variáveis como idade, sexo, escolaridade, raça/cor e presença de transtornos mentais.

Todas as etapas foram conduzidas em Python, com uso de bibliotecas como pandas, matplotlib e statsmodels.

Para investigar quais fatores estão associados à probabilidade de recorrência das tentativas de suicídio, foi ajustado um modelo de regressão logística binária utilizando os registros válidos do SINAN. A variável dependente do modelo foi a recorrência (0 = não recorrente, 1 = recorrente). As variáveis explicativas foram selecionadas com base na literatura e na disponibilidade dos dados, incluindo idade (em anos), sexo, escolaridade, raça/cor, situação conjugal, presença de transtorno mental, comorbidades e região de residência.

As variáveis categóricas foram transformadas em indicadores (dummies) para permitir sua incorporação no modelo. Após a construção da matriz de dados, o modelo foi estimado com a biblioteca statsmodels, utilizando o método de máxima verossimilhança.

Foram analisados os coeficientes, os odds ratios e seus respectivos intervalos de confiança, permitindo avaliar a força e o sentido da associação entre cada variável e o risco de recorrência. Essa abordagem tentou identificar quais características individuais aumentam ou reduzem a chance de nova tentativa, complementando a análise descritiva e

oferecendo uma visão estatística mais robusta dos fatores de risco envolvidos.

D. Clusterização

Nesta etapa do trabalho, foram aplicadas técnicas de aprendizado não supervisionado com o objetivo de identificar padrões de perfis de pessoas relacionadas às tentativas de suicídio registradas no SINAN.

Os métodos escolhidos foram K-Means e DBSCAN, que permitem explorar agrupamentos naturais nos dados sem utilizar rótulos pré-definidos.

Para o pré-processamento dos dados, inicialmente foram selecionadas variáveis relevantes para a construção dos perfis individuais: idade, sexo, raça/cor, escolaridade, situação conjugal, estado de residência, método utilizado na tentativa e recorrência (0 para primeira tentativa, 1 para caso recorrente). As variáveis categóricas foram transformadas em variáveis binárias (one-hot encoding) por meio da função `get_dummies`, enquanto idade e recorrência foram mantidas em formato numérico.

Em seguida, todos os atributos foram padronizados com o algoritmo `StandardScaler`, que transforma cada variável para média zero e desvio padrão igual a um. Essa etapa é essencial para evitar que variáveis com maior escala influenciem de forma desproporcional os algoritmos de clusterização, especialmente o K-Means, que se baseia em distâncias euclidianas.

Devido ao grande volume da base (mais de 900 mil registros), foi aplicada uma amostragem aleatória para permitir a execução eficiente dos algoritmos, preservando representatividade. Para visualização dos clusters, utilizou-se a técnica de Análise de Componentes Principais (PCA), reduzindo o espaço multidimensional para dois componentes principais, sem alterar os resultados dos agrupamentos.

IV. RESULTADOS

Nesta seção, são apresentados os resultados obtidos na segunda etapa da pesquisa, que dá continuidade às análises exploratórias realizadas e aprofunda a compreensão dos fatores associados aos casos de suicídio e tentativas no Brasil.

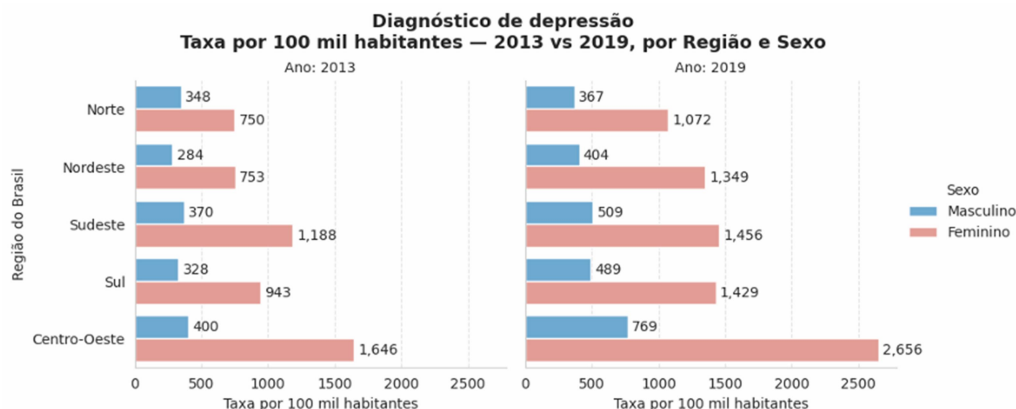


Fig. 1. Presença de diagnósticos de depressão (PNS)

A. Integração com dados do IBGE

Conforme mostra a Fig. 1., a comparação das taxas de diagnóstico de depressão entre 2013 e 2019 revela um crescimento expressivo em todas as regiões do Brasil, afetando ambos os sexos, mas com maior impacto sobre mulheres.

Em 2013, já se observava um padrão consistente: as taxas femininas eram superiores às masculinas em todas as regiões, especialmente no Centro-Oeste, onde o valor feminino ultrapassava 1.600 casos por 100 mil habitantes. Em 2019, esse padrão se acentuou, com aumentos significativos que elevaram as taxas femininas para níveis ainda mais altos, novamente com destaque para o Centro-Oeste, que ultrapassou 2.600 casos por 100 mil habitantes.

O gráfico de crescimento percentual, apresentado na Fig. 2, deixa ainda mais evidente essa tendência. Entre 2013 e 2019, todas as regiões apresentaram aumento nas taxas, porém a magnitude do crescimento variou de forma marcante. O maior crescimento masculino ocorreu no Centro-Oeste (+92%), seguido pelo Sul (+49%) e Nordeste (+42,5%), indicando uma intensificação do diagnóstico principalmente nessas áreas. Já entre as mulheres, o destaque foi a região Nordeste (+79,2%), seguida de Centro-Oeste (+61,4%) e Sul (+51,6%). Mesmo regiões com valores absolutos menores, como o Norte, apresentaram aumentos relevantes, especialmente no sexo feminino (+43%).

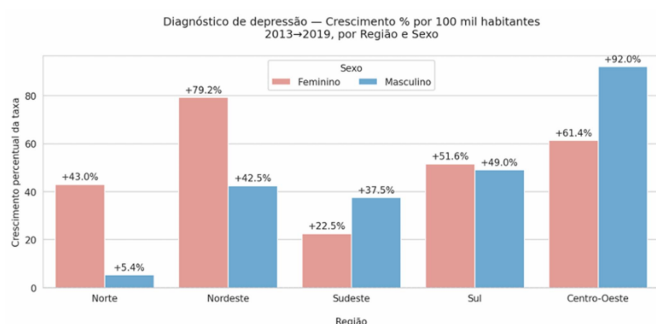


Fig. 2. Crescimento percentual de diagnóstico de depressão entre 2013 e 2019, por região.

A comparação entre os anos de 2013 e 2019 revela um aumento expressivo dos diagnósticos de depressão em todas as regiões, ao mesmo tempo em que as taxas de tentativas de suicídio nas capitais também cresceram, embora em proporções muito menores, como mostra a Fig. 3.

Como os gráficos utilizam escala logarítmica, fica

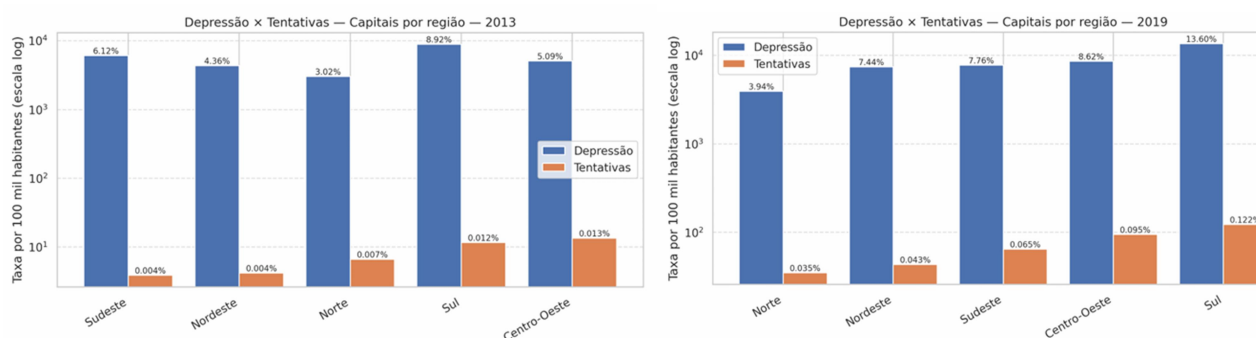


Fig. 3. Comparativo entre taxas de depressão e tentativas de suicídio.

evidente a diferença de magnitude entre os dois fenômenos: enquanto a depressão aparece em valores de milhares de diagnósticos por 100 mil habitantes, as tentativas registradas representam apenas uma fração percentual dessa taxa. É importante também ressaltar que foram considerados também somente os dados das capitais de cada estado das regiões, presentes na Pesquisa Nacional de Saúde (PNS).

Em 2013, as maiores taxas de depressão entre capitais se concentravam nas regiões Sul (8,92%), Sudeste (6,12%) e Centro-Oeste (5,09%), enquanto Norte e Nordeste apresentavam valores mais baixos. As tentativas, embora muito menos frequentes, seguiam um padrão semelhante: regiões com mais diagnósticos também apresentavam taxas relativamente maiores de tentativas, especialmente o Sul (0,012%) e Centro-Oeste (0,013%). Esse alinhamento inicial sugere uma possível relação regional entre maior prevalência de sofrimento mental diagnosticado e maior probabilidade de tentativa registrada.

Em 2019, esse contraste regional se acentuou. As taxas de depressão aumentaram substancialmente em todas as regiões, chegando a 13,60% no Sul e 8,62% no Centro-Oeste, e as tentativas também cresceram, principalmente no Sul (0,122%), que passou a apresentar o maior valor nacional. Além disso, observa-se um crescimento significativo das tentativas no Centro-Oeste e Sudeste, reforçando a presença de padrões regionais consistentes entre os dois indicadores.

Em síntese, os gráficos evidenciam que, mesmo com ordens de magnitude muito distintas, as regiões que apresentam maiores taxas de depressão também concentram maiores índices de tentativa de suicídio, tanto em 2013 quanto em 2019. O aumento observado em todas as capitais reforça a importância do monitoramento contínuo de indicadores de saúde mental, especialmente em regiões que já apresentavam maior risco inicialmente.

O gráfico da Fig. 4. mostra uma comparação entre a prevalência de transtornos mentais graves (esquizofrenia, transtorno bipolar, psicose ou TOC) diagnosticados por profissional de saúde mental, entre pessoas com 18 anos ou mais (dados da PNS) e as taxas de tentativas de suicídio registradas nas regiões brasileiras.

Em todas as regiões, observa-se uma diferença expressiva entre os dois indicadores: enquanto o diagnóstico de transtornos mentais aparece em valores relativamente elevados, variando entre 1,83% (Norte) e 6,04% (Sul), as tentativas de suicídio ocorrem em uma ordem de magnitude muito menor, com taxas variando de 0,026% a 0,108%. Essa diferença é destacada pela escala logarítmica do gráfico.

Regionalmente, a Região Sul se destaca com a maior prevalência de transtornos mentais graves (6,04%), seguida do Sudeste (5,84%) e Centro-Oeste (5,08%). Essas regiões

também apresentam as maiores taxas de tentativa, com destaque para o Sul (0,108%) e Centro-Oeste (0,071%). Já as regiões Norte e Nordeste exibem tanto menor prevalência desses transtornos quanto menores taxas de tentativas, sugerindo um padrão consistente entre sofrimento mental e risco registrado de suicídio.

Embora não seja possível inferir causalidade apenas a partir desses dados, o alinhamento entre maior prevalência de transtornos mentais graves e maior taxa de tentativas reforça um padrão amplamente reconhecido na literatura: pessoas com transtornos como bipolaridade, psicose ou esquizofrenia tendem a apresentar risco significativamente maior de comportamento suicida. Além disso, diferenças regionais podem refletir fatores como disponibilidade de serviços de saúde mental, acesso ao diagnóstico especializado e níveis de desigualdade socioeconômica.

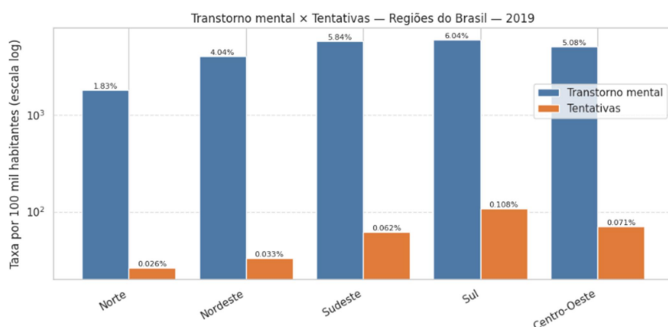


Fig. 4. Diagnóstico de outros transtornos mentais por região (esquizofrenia, transtorno bipolar, psicose ou TOC).

Em resumo, os resultados indicam que regiões com maior carga de transtornos mentais graves também tendem a apresentar maior incidência de tentativas de suicídio, com destaque para o Sul e Centro-Oeste. Esse padrão reforça a necessidade de políticas regionais específicas, especialmente nas áreas onde tanto a prevalência quanto o risco são mais elevados.

Outra variável analisada foi o PIB per capita a preços correntes dos municípios brasileiros, considerando os dados disponibilizados do ano de 2021, como mostra a Fig. 5.

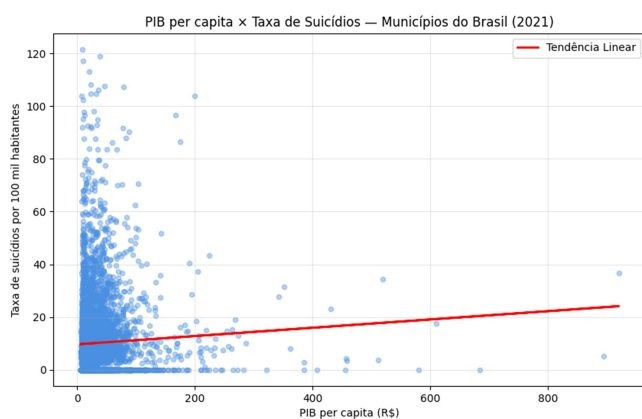


Fig. 5. Relação entre tentativas de suicídio e PIB per capita a preços correntes dos municípios brasileiros, em milhares de reais.

A grande maioria dos municípios brasileiros tem PIB per capita abaixo de R\$ 50 mil, o que explica a grande concentração de pontos à esquerda do gráfico. Apesar dessa concentração, observa-se uma grande variabilidade nas taxas de suicídio dentro desse grupo, com municípios de PIB baixo

apresentando tanto taxas muito baixas quanto muito altas (acima de 100 por 100 mil habitantes).

Nos valores mais altos de PIB per capita (acima de R\$ 100 mil e especialmente acima de R\$ 200 mil), o número de municípios é muito pequeno. Nesses municípios, a taxa de suicídio tende a ser baixa, embora existam alguns pontos isolados com valores mais altos. Esse comportamento reforça que, mesmo em municípios com maior renda, não existe um padrão consistente de redução ou aumento da taxa de suicídio.

A linha de tendência, em vermelho no gráfico, apresenta coeficiente positivo pequeno, o que indica uma relação positiva fraca entre o PIB per capita dos municípios e a taxa de suicídios. Embora a linha tenha uma leve inclinação ascendente, isso não representa um padrão consistente, mas sim o efeito de alguns poucos municípios com PIB muito elevado, que acabam puxando a regressão para cima.

A grande massa de municípios, concentrada em valores baixos de PIB per capita, mostra taxas de suicídio muito variadas, o que evidencia ausência de correlação clara entre as duas variáveis.

Em relação aos dados agregados de municípios com e sem estrutura organizacional nas áreas de educação, cultura, esporte e saúde, o gráfico da Fig. 6. compara, em cada região do Brasil, a taxa média de tentativas de suicídio entre dois grupos: todos os municípios da região e apenas os municípios classificados como sem estrutura organizacional nas áreas citadas anteriormente.

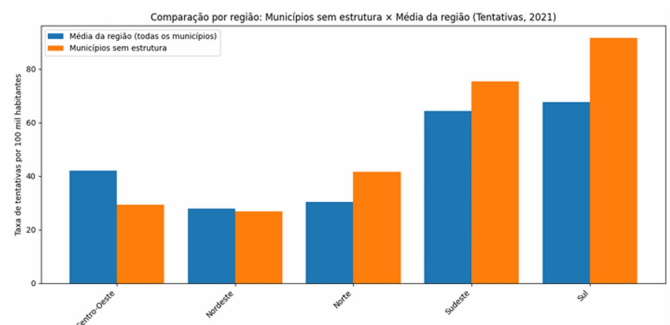


Fig. 6. Presença de estrutura organizacional nas áreas de educação, cultura, esporte e saúde.

Os resultados mostram que o impacto da ausência de estrutura varia entre as regiões, mas segue um padrão geral: em regiões onde as taxas de tentativas já são mais elevadas, como Sul e Sudeste, os municípios sem estrutura apresentam valores ainda maiores do que a média regional.

No Sul, por exemplo, a taxa média passa de aproximadamente 70 para mais de 90 tentativas por 100 mil habitantes nos municípios sem estrutura; e no Sudeste, o salto vai de cerca de 65 para aproximadamente 75. Essa diferença sugere que a falta de organização administrativa mínima nessas áreas pode estar associada a maior vulnerabilidade e maior risco de tentativa de suicídio.

Em regiões com taxas historicamente mais baixas, como Nordeste, os municípios sem estrutura apresentam taxas iguais ou menores que a média regional. Isso indica que, nessas regiões, a presença ou ausência de estrutura administrativa não se relaciona da mesma forma com o risco de tentativas, podendo refletir diferenças socioeconômicas,

distribuição populacional, menor capacidade de registro ou outros fatores estruturais regionais.

A relação entre escolarização e tentativas de suicídio varia conforme a idade, mas tende a ser fraca em todas as faixas etárias, como mostra a Fig. 7.

Nas idades escolares obrigatórias entre 6 e 14 anos, praticamente não há associação entre os indicadores. Entre 15 e 17 anos, observa-se uma correlação ligeiramente negativa, indicando que a evasão escolar pode estar ligada a maior vulnerabilidade e risco de tentativas, embora o efeito seja muito pequeno.

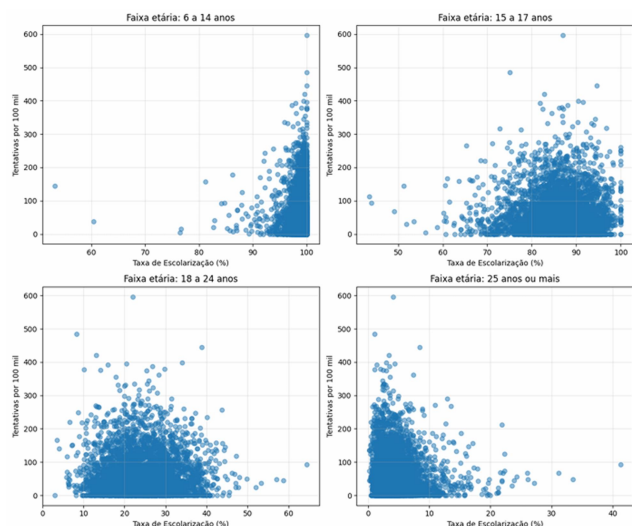


Fig. 7. Relação entre escolarização e tentativas de suicídio, com dados de 2021.

A associação se torna mais evidente na faixa etária entre 18 e 24 anos, fase em que a escolarização passa a variar bastante entre municípios. Ainda assim, a correlação permanece fraca, sugerindo apenas que estar fora da escola ou faculdade nessa etapa pode ampliar riscos em alguns contextos. Após os 25 anos, a escolarização deixa de funcionar como indicador relevante, pois grande parte da população já não está mais inserida no sistema educacional.

O gráfico da Fig. 8. mostra a distribuição da população por nível de instrução nas regiões brasileiras e evidencia diferenças relevantes.

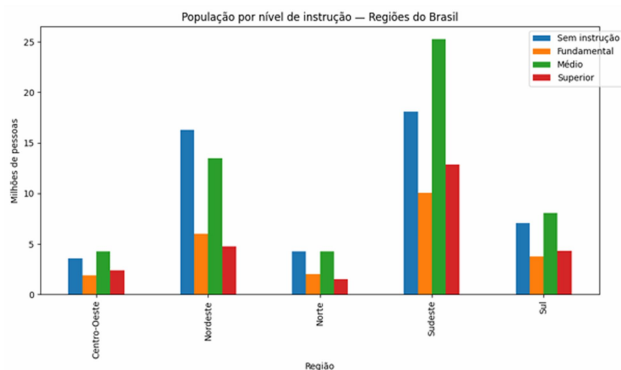


Fig. 8. População por nível de instrução, com dados de 2022.

O Sudeste concentra o maior número de pessoas com ensino médio e superior, refletindo maior escolarização

média, enquanto o Nordeste e o Norte apresentam proporções mais altas de pessoas sem instrução ou com apenas o fundamental.

Essas desigualdades educacionais regionais se refletem na comparação com as tentativas de suicídio, como mostra o gráfico da Fig. 9., que mostra a relação da proporção de baixa escolaridade (sem instrução) com a taxa regional de tentativas de suicídio.

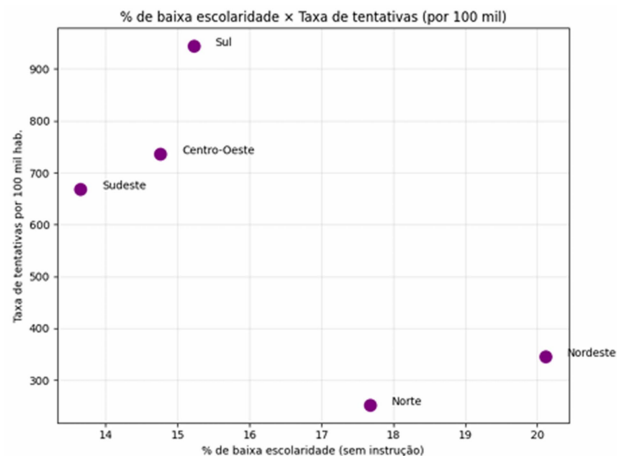


Fig. 9. Relação entre nível de instrução e tentativas de suicídio, com dados de 2022.

O padrão apresentado é contraintuitivo, com regiões com maior escolarização média, como Sul, Sudeste e Centro-Oeste, apresentam as maiores taxas de tentativas, enquanto regiões com maior proporção de pessoas sem instrução, como o Nordeste, exibem taxas menores.

O Norte aparece com a menor taxa entre todas as regiões, apesar de ter escolaridade média baixa. Esse comportamento indica que, no nível agregado regional, o nível de instrução não funciona como um preditor direto das tentativas. Ao contrário, a escolaridade mais alta se associa a taxas maiores, sugerindo a influência de fatores culturais, demográficos e socioeconômicos específicos de cada região.

Foi analisada também a influência da evolução da taxa de desocupação por região (2012–2025), como mostra a Fig. 10.

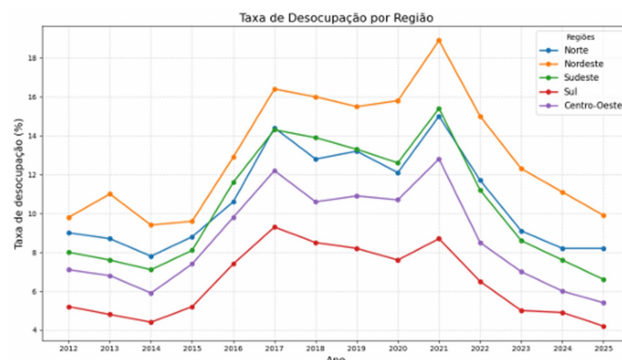


Fig. 10. Evolução da taxa de desocupação por região (2012–2025).

O gráfico mostra que todas as regiões brasileiras passaram por um ciclo semelhante de aumento e posterior queda da desocupação ao longo do período analisado. Entre 2012 e 2014, as taxas se mantêm relativamente estáveis e em patamares baixos, com o Sul e o Centro-Oeste apresentando consistentemente os menores valores, enquanto o Nordeste

mantém as maiores taxas. A partir de 2015, observa-se um aumento expressivo em todas as regiões.

O ano de 2020 marca uma nova aceleração do desemprego, provavelmente associada aos impactos sociais e econômicos da pandemia de COVID-19. O Nordeste volta a apresentar o maior valor, aproximando-se de 19%, seguido por Norte e Sudeste. A partir de 2021, inicia-se uma trajetória de queda contínua nas taxas de desocupação, com todas as regiões retornando a níveis próximos aos pré-crise.

Foram calculadas as correlações por região entre a taxa de desocupação e as taxas de suicídio, sendo que a correlação global ficou aproximadamente 0.17, o que indica que indica que essa relação não é linear nem uniforme em todo o país.

Porém, quando analisada por região, a correlação se torna moderada: Nordeste (0,53), Sul (0,47), Sudeste (0,41) e Norte (0,43). Isso significa que, em várias regiões, períodos de maior desemprego tendem a estar associados a aumentos nas tentativas de suicídio, ainda que não de forma determinante. O Centro-Oeste apresenta correlação mais baixa (0,24), sugerindo que outros fatores regionais podem ser mais relevantes para explicar seu padrão de crescimento.

A associação não é forte o suficiente para indicar causalidade, mas reforça que vulnerabilidades socioeconômicas, como desemprego, podem intensificar riscos em determinados contextos regionais.

B. Análise de Recorrência e fatores de risco

Para entender quais fatores aumentam ou diminuem a chance de uma pessoa já ter tentado suicídio antes, foi feita uma regressão logística usando a variável “recorrência” como resposta. As variáveis utilizadas foram: idade, sexo, escolaridade, raça, estado civil, transtorno mental, transtorno de comportamento e região.

As categorias foram transformadas em variáveis numéricas (dummies), e os registros com informações faltando foram retirados. O modelo identificou como cada fator está associado à probabilidade de recorrência, permitindo comparar o impacto relativo de cada um deles.

Os resultados do modelo de regressão logística evidenciam que diversos fatores individuais e contextuais influenciam a probabilidade de recorrência das tentativas de suicídio. A idade apresentou associação positiva, indicando que, a cada aumento de um ano, há um pequeno crescimento na chance de repetição da tentativa.

Assim como observado por estudos epidemiológicos [1] e reforçado por pesquisas brasileiras com dados do DATASUS, o sexo feminino mostrou maior probabilidade de recorrência, o que se alinha ao padrão amplamente documentado de que mulheres realizam mais tentativas ao longo da vida, enquanto homens apresentam maior letalidade.

Em relação à escolaridade, observaram-se coeficientes positivos para categorias intermediárias, sugerindo que indivíduos com ensino fundamental ou médio incompleto apresentam maior risco de reincidência do que aqueles com escolaridade mais alta.

A variável transtorno mental também se mostrou significativamente associada, com aumento expressivo nas chances de recorrência entre pessoas diagnosticadas com algum transtorno psiquiátrico.

Quando consideradas as regiões, verificou-se que residentes do Nordeste e do Sudeste apresentaram maior probabilidade de recorrência em comparação ao Centro-Oeste, que foi utilizado como categoria de referência.

Os resultados do modelo de regressão logística estão de acordo com o que a literatura aponta sobre a recorrência das tentativas de suicídio. Assim como em estudos nacionais e internacionais, o modelo mostrou maior probabilidade de repetição entre mulheres, pessoas com transtornos mentais e indivíduos em situação de vulnerabilidade social.

Esses achados são compatíveis com a literatura científica, que reconhece transtornos mentais como fatores de risco relevantes para recorrência [1], [5], e também com estudos que ressaltam a influência do contexto social na manutenção do risco [3].

As diferenças regionais identificadas pelo modelo também refletem padrões já observados em estudos nacionais [4], [9], mostrando que o comportamento suicida não é distribuído de forma uniforme no país. Assim, o modelo reforça conclusões presentes na literatura e confirma que a recorrência é influenciada por múltiplos fatores individuais e sociais.

C. Análise Temporal

Foram realizadas análises temporais a partir das bases SIM e SINAN, considerando janelas de 10 anos e observando faixas etárias, regiões do país e a evolução da recorrência entre as tentativas de suicídio. O objetivo é mostrar como o fenômeno tem mudado ao longo do tempo e quais grupos e locais apresentam maior crescimento.

As análises mostram que a taxa de óbitos por suicídio vem aumentando de forma contínua no Brasil ao longo das últimas décadas. Entre 1996 e 2003, os valores eram mais baixos e relativamente estáveis. Já entre 2004 e 2013, esse crescimento se torna mais evidente, e na janela mais recente, de 2014 a 2023, o aumento é ainda mais forte, como mostra a Fig. 11.



Fig. 11. Evolução da taxa de óbitos por suicídio (2014–2023).

A partir de 2020, observa-se um salto mais acentuado, compatível com os efeitos da pandemia de COVID-19, como isolamento social, piora da saúde mental e redução na oferta de serviços.

Quando se observa a evolução por faixa etária, na Fig. 12., percebemos que o crescimento ocorre em praticamente todos os grupos, mas não de maneira uniforme. As faixas 25–34, 35–44 e 45–54 anos são as que mais chamam atenção, pois concentram tanto o maior número de óbitos quanto o maior crescimento ao longo dos anos. Esses grupos

representam adultos em idade produtiva e, muitas vezes, sob maior pressão econômica, social e emocional.

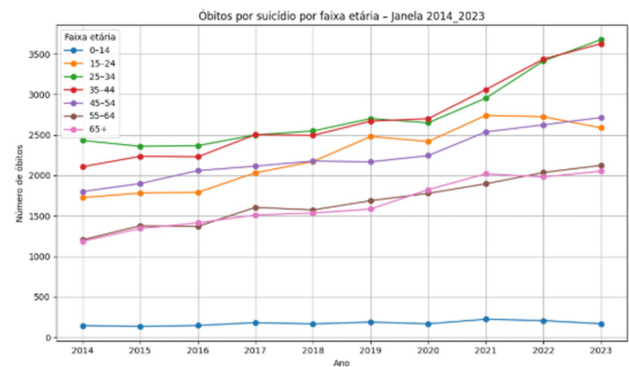


Fig. 12. Evolução da taxa de óbitos por suicídio por faixa etária (2014–2023).

As faixas 55–64 e 65+ também apresentam aumento constante, sugerindo um envelhecimento do perfil das vítimas.

A faixa 0–14 anos permanece estável e em valores muito abaixo das demais, o que mostra que, apesar do crescimento geral, o impacto sobre crianças permanece pequeno ou subnotificado.

No conjunto, os resultados indicam que o suicídio afeta principalmente adultos e pessoas de meia-idade, e que esses grupos têm se tornado cada vez mais vulneráveis com o tempo.

As análises regionais revelam que todas as regiões apresentam crescimento nas taxas de suicídio, mas em intensidades bem diferentes. A Região Sul se destaca como a mais afetada em todas as janelas temporais, com taxas sempre acima das demais. Em seguida aparecem Centro-Oeste e Sudeste, também apresentando aumento consistente.

Já Norte e Nordeste possuem as menores taxas, embora também estejam em crescimento, como mostra a Fig. 13.

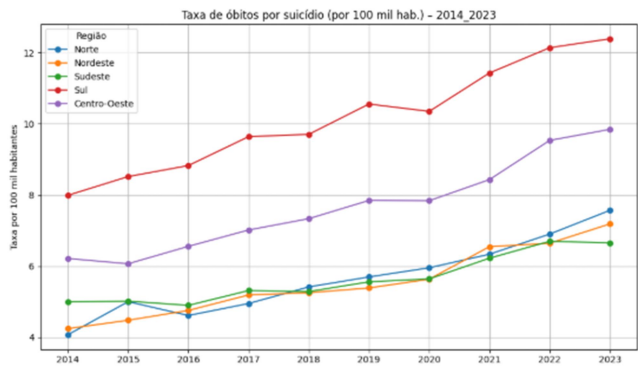


Fig. 13. Evolução da taxa de óbitos por suicídio por região do Brasil (2014–2023).

Outro resultado importante diz respeito à recorrência, isto é, a proporção de pessoas que voltam a tentar suicídio depois da primeira ocorrência. Os dados mostram que essa taxa tem aumentado de maneira contínua em todas as janelas analisadas. Entre 2009 e 2018, o crescimento foi mais lento,

com algumas oscilações nos primeiros anos. Porém, entre 2014 e 2023, a subida é muito mais rápida, especialmente a partir de 2018. Em 2023, a taxa ultrapassa 55%, indicando que mais da metade das pessoas que tentam suicídio está retornando com novas tentativas.

D. Análise de Clusters

A aplicação das técnicas de clusterização teve como objetivo identificar perfis de pessoas envolvidas em tentativas de suicídio, considerando variáveis individuais (idade, sexo, raça/cor, escolaridade), contextuais (UF, situação conjugal) e características do evento (método e recorrência).

A análise se concentrou principalmente no algoritmo K-Means, que apresentou clusters interpretáveis e consistentes, enquanto o DBSCAN foi utilizado de forma exploratória para avaliar a existência de regiões de alta densidade no espaço multidimensional dos dados.

A aplicação do algoritmo DBSCAN aos dados de tentativas de suicídio evidenciou a presença de um grande agrupamento denso, acompanhado por registros classificados como ruído. Esse resultado indica que as observações se organizam de forma contínua no espaço dos dados, sem a existência de múltiplas regiões densas naturalmente bem delimitadas, sugerindo a ausência de divisões naturais claras entre os perfis analisados.

Após amostragem aleatória e padronização dos atributos, o K-Means convergiu para quatro agrupamentos distintos, como mostra a Tabela II. Os resultados foram visualizados em duas dimensões por meio de Análise de Componentes Principais (PCA), permitindo observar a separação relativa entre os clusters e sua distribuição geral na amostra. Os quatro perfis identificados representam grupos populacionais relevantes para ações de saúde pública.

TABELA II. CLUSTERS IDENTIFICADOS

CLUSTERS		
Id	Perfil resumido	Risco
1	Jovens adultos, maioria mulheres, envenenamento	Moderado
2	Jovens adultos, envenenamento acentuado	Moderado
3	Homens mais velhos, métodos agressivos, maior recorrência	Alto
4	Jovens com automutilação, recorrência intermediária	Alto

A Fig. 14 apresenta a projeção das tentativas de suicídio nos dois primeiros componentes principais, com os agrupamentos obtidos pelos algoritmos K-Means e DBSCAN. No painel referente ao K-Means, observam-se quatro grupos parcialmente sobrepostos, indicando que os perfis não são totalmente separados e apresentam transições graduais. Já o resultado do DBSCAN evidencia um grande agrupamento contínuo, acompanhado de registros dispersos, sugerindo a ausência de divisões naturais claras entre os perfis analisados.

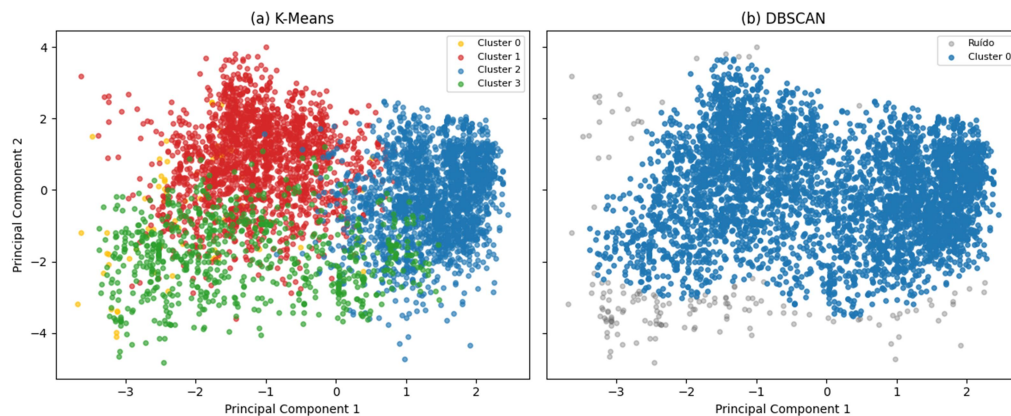


Fig. 14. Comparação entre os agrupamentos obtidos por K-Means e DBSCAN na projeção dos dados em PCA.

Para fins de apresentação, os clusters foram renumerados de 1 a 4 na Tabela II, enquanto nos gráficos são apresentados conforme a numeração original do algoritmo, de 0 a 3.

O primeiro cluster reúne adultos jovens, majoritariamente mulheres, cujas tentativas ocorrem principalmente por envenenamento. A recorrência é moderada, sugerindo um padrão de comportamento associado a crises emocionais, mas sem predominância de métodos altamente letais.

O segundo cluster apresenta um perfil semelhante, também composto por jovens adultos e predominantemente femininos, mas com ainda maior concentração de tentativas por ingestão de substâncias tóxicas. Em ambos os casos, observa-se que os métodos utilizados possuem menor letalidade imediata, embora indiquem sofrimento persistente.

O terceiro cluster apresenta um perfil mais crítico. Com idade média mais elevada e maior proporção de homens, esse grupo concentra as maiores taxas de recorrência e um uso expressivo de métodos mais letais, como enforcamento. Esses resultados indicam que esse cluster representa um grupo de maior gravidade, cujo risco para evolução desfavorável é maior e, portanto, demanda atenção especial em estratégias de prevenção e acompanhamento contínuo.

O quarto cluster é composto majoritariamente por jovens que utilizam objetos cortantes, refletindo um padrão associado à automutilação. A recorrência também é elevada, reforçando a necessidade de intervenção precoce e cuidados em saúde mental.

A análise exploratória com o algoritmo DBSCAN complementou esses achados. Embora o método seja capaz de identificar agrupamentos com base na densidade dos dados, neste estudo ele indicou a presença de um grande agrupamento principal, além de registros classificados como ruído, não evidenciando a formação de múltiplos grupos bem definidos.

Esse resultado sugere que a estrutura dos dados é mais contínua do que segmentada e que não há fronteiras naturais bem definidas entre perfis, o que reforça a adequação do K-Means como método principal para caracterização dos grupos. Assim, os resultados indicam que as tentativas de suicídio no Brasil apresentam heterogeneidade significativa, mas ainda assim permitem a identificação de perfis que podem subsidiar ações de prevenção direcionadas.

Após a análise dos clusters associados às tentativas de suicídio, aplicou-se a técnica de clusterização aos dados de

suicídio consumado, com o objetivo de comparar a estrutura dos perfis identificados. Para essa etapa, foi utilizado o algoritmo K-Means com três clusters, considerando as variáveis sociodemográficas e o método utilizado, sendo a PCA empregada apenas para visualização.

A clusterização resultou em dois grandes agrupamentos com características muito semelhantes e um terceiro cluster residual, composto por poucos registros, como mostra a Fig. 15.

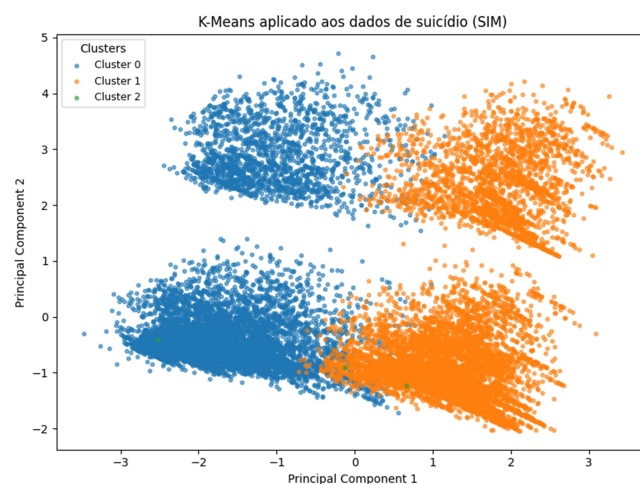


Fig. 15. Agrupamentos obtidos por K-Means para óbitos por suicídio.

Os dois clusters principais concentram majoritariamente indivíduos do sexo masculino, com idades médias mais elevadas e uso predominante de métodos altamente letais, como enforcamento e arma de fogo. As diferenças observadas entre esses grupos são sutis, refletindo principalmente variações etárias, o que indica elevada homogeneidade dos perfis associados aos óbitos por suicídio.

O terceiro cluster, composto por um número reduzido de registros, não apresentou características consistentes que permitissem sua interpretação como um perfil distinto, sendo considerado um conjunto de casos atípicos decorrente do processo de clusterização.

Diferentemente do observado nas tentativas de suicídio, cuja clusterização revelou maior diversidade de perfis, os óbitos por suicídio apresentaram estrutura significativamente mais homogênea. Esse contraste sugere que, embora as tentativas envolvam trajetórias variadas, o suicídio consumado tende a concentrar-se em um conjunto mais

restrito de características, especialmente relacionadas ao sexo masculino, maior idade e uso de métodos mais letais.

V. CONCLUSÃO

Este trabalho investigou padrões de suicídio e tentativas no Brasil utilizando dados integrados do SIM, SINAN e IBGE. As análises mostraram que tanto os óbitos quanto as tentativas vêm aumentando ao longo das últimas décadas, com maior concentração entre adultos e nas regiões Sul, Centro-Oeste e Sudeste. A recorrência também cresce de forma consistente, indicando maior vulnerabilidade entre pessoas que já passaram por uma primeira tentativa. Variáveis socioeconômicas e educacionais apresentaram associações distintas conforme a região, reforçando a complexidade do fenômeno.

Apesar dos resultados consistentes, esta pesquisa apresenta limitações importantes. A principal delas é a subnotificação, especialmente nas tentativas de suicídio, que dependem de registro adequado pelas unidades de saúde e podem variar entre regiões e ao longo do tempo. Além disso, o SINAN ampliou sua cobertura e qualidade de preenchimento nos últimos anos, o que significa que parte do aumento observado pode refletir melhoria na notificação, e não necessariamente crescimento real no número de casos. Outras limitações incluem possíveis erros de preenchimento no sistema pelos profissionais de saúde, ausência de algumas variáveis em determinados anos e impossibilidade de inferir causalidade a partir dos dados utilizados.

Ainda assim, os padrões identificados ajudam a compreender melhor os grupos mais afetados e reforçam a necessidade de políticas públicas direcionadas, com atenção especial às regiões e perfis populacionais que apresentam maior risco. Os óbitos concentram-se principalmente nas regiões Sul e Centro-Oeste e afetam, em maior proporção, homens adultos, especialmente entre 30 e 59 anos, com menor escolaridade e maior utilização de métodos de alta letalidade. Por outro lado, as tentativas de suicídio ocorrem com maior frequência nas regiões Sudeste, Sul e Centro-Oeste e estão mais associadas a mulheres, adolescentes e adultos jovens, caracterizando maior recorrência e o uso predominante de métodos menos letais, como a intoxicação.

A identificação desses perfis evidencia a importância de estratégias de prevenção que considerem as diferenças regionais e populacionais. Estudos futuros podem aprofundar relações causais, incorporar variáveis adicionais e explorar modelos preditivos para apoiar estratégias de prevenção mais eficazes.

REFERÊNCIAS

- [1] Organização Mundial da Saúde, "Suicídio – Ficha informativa", OMS, mar. 2025. [Online]. Disponível: <https://www.who.int/news-room/fact-folhas/detalhes/suicidio>
- [2] E. M. Kleiman et al., "O uso de tecnologia avançada e métodos estatísticos para prever e prevenir o suicídio," *Molecular Psychiatry*, vol. 28, pp. 1821–1834, 2023. [Online]. Disponível: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10426769>
- [3] A. M. Giabbanelli et al., "Caminhos para o suicídio ou conjuntos de ciclos viciosos? Entendendo a complexidade do suicídio por meio do mapeamento causal," *Social Network Analysis and Mining*, vol. 12, n° 97, pp. 1, 17, out. 2022. [Online]. Disponível: <https://link.springer.com/article/10.1007/s13278-022-00886-9>
- [4] V. W. Klann and E. P. M. Sousa, "BrSuicides: Conjunto de dados para análise de casos diários de suicídio no Brasil," in *Anais do Dataset Showcase Workshop (DSW)*, Florianópolis, SC, Brasil, 2024, pp. 12–22. DOI: 10.5753/dsw.2024.244020
- [5] Ministério da Saúde (Brasil), *Prevenção do suicídio: manual dirigido a profissionais das equipes de saúde mental*, Brasília: Ministério da Saúde, 2006. [Online]. Available: https://bvsms.saude.gov.br/bvs/publicacoes/manual_editoracao.pdf
- [6] E. Namgung, H. Kang, and B. Kim, "Clusters of outpatients with suicidal ideation: a latent class analysis of clinical and sociodemographic characteristics," *BMC Psychiatry*, vol. 24, no. 94, 2024. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/38394926>
- [7] S. Choi et al., "Identification of spatiotemporal suicide clusters among adolescents in Korea using a density-based clustering algorithm," *Child and Adolescent Psychiatry and Mental Health*, vol. 18, no. 19, 2024. [Online]. Available: <https://capmh.biomedcentral.com/articles/10.1186/s13034-024-00745-9>
- [8] R. Benson et al., "A Systematic Review of Quantitative Methods Used to Identify or Evaluate Suicide and Self-Harm Clusters," *International Journal of Environmental Research and Public Health*, vol. 19, no. 9, p. 5313, 2022. [Online]. Available: <https://www.mdpi.com/1660-4601/19/9/5313>
- [9] V. W. Klann and E. P. M. Sousa, "Análise de similaridades e padrões estatísticos em séries temporais de casos de suicídio no Brasil," in *Anais do Workshop de Teses e Dissertações (WTDBD) - SBBD*, Florianópolis, SC, Brasil, 2024, pp. 165–171. DOI: 10.5753/sbbd_estendido.2024.243810