

Análise de similaridades e padrões estatísticos em séries temporais de casos de suicídio no Brasil

Victor William Klann¹, Elaine Parros Machado de Sousa¹

¹Universidade De São Paulo
Instituto De Ciências Matemáticas e de Computação
Av. Trab. São Carlense, 400 - Centro, São Carlos - SP, 13566-590

victor.klann@usp.br, parros@icmc.usp.br

Level: Master's in Computer Science and Computational Mathematics

Admission: 03/2023 **Qualifying Exam:** 28/05/2024 **Defense:** 03/2025

Completed Activities: Completion of mandatory credits, qualifying exam and presentation, foreign language proficiency exam, data collection, modeling and preprocessing, data exploratory analysis

Future Activities: Similarity search, clustering, temporal pattern mining, multi-modal data integration

Publications: DSW 2024

***Abstract.** Suicide is a serious global public health problem and is among the leading causes of death in the world. In Brazil, suicide rates have seen alarming increases and continue to rise. Some public policies have been implemented to prevent the rise in suicide cases. Despite this, studies in the field indicate that such measures are not enough to cause a reduction in cases. The collection and analysis of suicide data are important for creating and reevaluating monitoring and prevention processes for suicide attempts. The goal of this master's project is to contribute to the recognition of temporal patterns of suicide cases in Brazil from suicide deaths data published by DATASUS, using a time series-based approach. The main steps of this work are the collection of data and modeling of time series, provision of a pre-processed database containing daily suicide cases data from Brazil, exploratory analysis of the time series, similarity search between time series and application of clustering; mining of time series subsequence patterns, integration of multimodal data, and finally, the evaluation of the results obtained. It is expected to identify inherent temporal patterns in the time series to assist in understanding the behavior of suicide cases in the states and regions of Brazil.*

1. Introdução

Ao longo dos anos, o suicídio tem se tornado um problema sério de saúde pública global. Estima-se que em 2019, 703.000 pessoas cometeram suicídio, e está entre as principais causas de morte mundialmente (WHO 2021). O suicídio é definido como um ato de lesão ou intoxicação autodirigida intencionalmente (WHO 1992). Esforços em inúmeros países para combate e prevenção do suicídio permitiram que os casos de suicídio caíssem desde os anos 2000. A tendência de casos de suicídio no Brasil, contudo, mostrou-se crescente

durante a mesma época (Ilic and Ilic 2022). Durante o período de 2002 a 2012, todas as regiões reportaram aumentos: no norte, 77,7%; no nordeste, 51,7%; no centro-oeste, 16,3%; no sul, 15,2% e no sudeste, 35,7%. No entanto, o aumento não foi proporcional geograficamente, visto que, no sudeste, o Rio de Janeiro quase não apresentou crescimento, em contraste com Minas Gerais, cujos casos cresceram em 58,3% (Junior 2015).

Segundo dados da Organização Mundial de Saúde (OMS), divulgados em 2014, o Brasil, apesar de apresentar altos números absolutos de ocorrências de suicídio, eram consideradas baixas taxas de suicídio quando comparadas com outros países. Porém, algumas regiões do Brasil, como o extremo norte e extremo sul, se equiparavam a países com as maiores taxas de óbito por suicídio. A então Secretaria de Vigilância em Saúde (SVS)¹ reportou que a taxa de suicídio no país teve um aumento alarmante de 43% comparando o número anual de mortes entre 2010 e 2019, e que a tendência das taxas de suicídio observadas na década anterior das cinco regiões do país continuaram ascendentes, com as regiões sul e centro-oeste possuindo as maiores taxas de suicídio do país (SVS 2021).

Algumas medidas e políticas públicas foram aplicadas como tentativa de prevenir o aumento de incidências de suicídio. A Campanha do Setembro Amarelo (CSA)² é uma medida criada pela Associação Brasileira de Psiquiatria (ABP) em conjunto com o Conselho Federal de Medicina (CFM). A campanha tem como objetivo conscientizar a população sobre as circunstâncias que causam comportamentos suicidas e prestar esclarecimentos quanto aos tratamentos recomendados para esse tipo de transtorno mental.

A coleta e análise desses dados são importantes para que se possam criar e reavaliar processos de monitoramento e prevenção das tentativas de suicídio (WHO 2021; SVS 2021). Nota-se também que profissionais da saúde defendem que é de extrema importância a análise dos casos bem sucedidos de prevenção para avaliar políticas existentes e guiar implementações futuras (Machado et al. 2014).

2. Motivação e justificativa

Desde o início do século, diversos trabalhos revisaram os dados de óbitos causados por suicídio no Brasil, porém a maioria desses trabalhos realizam apenas uma análise exploratória rasa dos dados, geralmente sem considerar a evolução temporal dos mesmos, e com pouca ou nenhuma utilização de técnicas de análise estatística.

O estudo de (de Oliveira et al. 2020) teve como objetivo analisar a prevalência das notificações de suicídio no Brasil antes e após o lançamento da Campanha Setembro Amarelo pelo Ministério da Saúde (MS). Os autores utilizaram os dados contidos no Sistema de Informações sobre Mortalidade (SIM) e no Sistema de Informação de Agravos de Notificação (SINAN) disponibilizados pelo Departamento de Informática do Sistema Único de Saúde (DATASUS)³.

Os resultados da pesquisa mostraram que as taxas de incidência relacionadas ao suicídio apresentaram dados crescentes após a aplicação do programa. Os autores concluíram que, apesar da campanha ter sido iniciada em 2015 com o objetivo de prevenir o

¹<https://svs.aids.gov.br/daent/cgiae/sim/apresentacao/> Acesso em: 3 mar. 2024.

²<https://www.abp.org.br/setembro-amarelo>

³<https://datasus.saude.gov.br/> Acesso em: 3 mar. 2024

suicídio, as taxas de incidência apresentaram dados crescentes após a aplicação da mesma, levando a questionamentos sobre a sua efetividade. Também sugeriram acompanhar a campanha por um intervalo maior de tempo para avaliar seus impactos futuros.

(Cruz et al. 2023) buscaram avaliar a eficácia da Campanha de Setembro Amarelo comparando com dados antes e após a sua implementação em escala nacional, a partir de 2015. Foi observado um aumento nas taxas anuais de mortes por suicídio entre os anos analisados de 2011 a 2019, concluindo que a implementação da campanha não resultou na redução de casos de suicídio.

(da Mata et al. 2020) realizaram uma análise sociodemográfica das vítimas de suicídio no Brasil entre 2006 e 2015, utilizando os dados do SIM e DATASUS. A análise foi realizada nos dados absolutos de suicídio, mostrando um aumento acumulado de 37.707 casos de suicídio e um aumento de 37.7% no número de ocorrências entre 2006 e 2015 na região sudeste, 22.685 casos e 23,1% no sul, 22.617 casos e 23% no nordeste, 8.596 casos e 8.75% no centro-oeste, e 6.591 casos com aumento de 6.71% de casos de suicídio no norte. Os mesmos concluíram que a mortalidade por suicídio continuou a crescer no país, destacando variações regionais. Eles destacaram que o Brasil ainda carece de programas governamentais que trabalhem efetivamente na prevenção do suicídio.

2.1. Solução proposta

Nesse contexto, este projeto de mestrado visa contribuir para o reconhecimento de padrões temporais a partir de dados de casos de suicídio no Brasil, realizando uma análise exploratória dos dados relacionados a casos de suicídio no Brasil e identificar padrões que possibilitem tanto uma avaliação mais ampla do comportamento temporal desses dados no país como um todo, quanto uma avaliação mais focada em regiões e estados brasileiros, ao longo do tempo. Pretende-se utilizar técnicas de análise e mineração de séries temporais que sejam robustas à natureza não-determinística e irregular dos dados.

2.2. Objetivos específicos

1. Organizar e disponibilizar uma base de dados pré-processada contendo os dados de casos diários de suicídio de todas as unidades federativas do Brasil;
2. Obter e analisar propriedades estatísticas e de comportamentos como a tendência, sazonalidade e distribuição de dados residuais das séries temporais por meio de técnicas de pré-processamento e decomposição de séries.
3. Identificar possíveis relações entre os dados representados nas séries temporais, como similaridade no comportamento dos dados de estados geograficamente próximos ou distantes, a partir da aplicação de *clustering* de séries temporais.
4. Verificar a eficácia das campanhas do setembro amarelo, aplicando métodos de busca de similaridade entre subsequências para a mineração de padrões dentro das séries temporais. Realizar também a extração de intervalos de início e fim das campanhas para definição do tamanho das subsequências, através da utilização de técnicas de multimodalidade.

3. Metodologia

O desenvolvimento dessa pesquisa envolve as seguintes etapas principais: coleta dos dados e modelagem das séries temporais; pré-processamento dos dados; análise exploratória e de componentes das séries temporais; busca de similaridade entre séries temporais

e aplicação de *clustering* nas séries similares; mineração de padrões de subsequências de séries temporais e integração de dados multimodais; e avaliação dos resultados obtidos. Até o momento, foram realizadas as etapas de coleta, modelagem e pré-processamento dos dados, e o início da análise exploratória e de componentes das séries temporais de cada região e unidade federal do país.

O DATASUS disponibiliza os dados de óbitos de todas as unidades federativas do Brasil, e um dicionário de atributos extensivo ⁴. Os dados são divididos em duas classificações, baseadas na 9ª Revisão de Classificação Internacional de Doenças (CID-9), até 1995, e a 10ª Revisão (CID-10), após 1996. Como as duas classificações são diferentes (WHO 1992), o DATASUS não disponibiliza uma lista integrada⁵.

O código CID-10 é composto por uma letra e dois dígitos, representando a causa principal da morbidade ou mortalidade. É possível incluir até 4 novos dígitos decimais para incluir detalhes como subcategoria, severidade, local de ocorrência e outros subtipos, sendo determinado pelo início do código (WHO 1992). Os códigos de suicídio começam com a letra X e os dois primeiros dígitos devem pertencer ao intervalo fechado de [60, 84] (WHO 1992). O terceiro dígito é uma subclassificação realizada pelo DATASUS para indicar o local de ocorrência do suicídio ⁶. Um exemplo é o código X650, definido como o óbito de auto-intoxicação voluntária por álcool, cometido na própria residência da vítima.

Os dados utilizados são de casos de óbito no Brasil ocorridos durante o período de 1996 e o último ano consolidado, que atualmente é 2022. A granularidade dos dados é de natureza diária, baseado na declaração de óbito individual do falecido registrada em um Cartório de Registro Civil. Os registros dos cartórios são coletados pelas Secretarias de Saúde estaduais e agregadas pelo SIM. A consolidação da base de dados nacional é realizada somente após a agregação dos dados de todas as unidades federais (SVSA 2024). Os dados foram obtidos pelo do servidor do DATASUS através do uso da biblioteca de Python pySUS (Coelho et al. 2021). Para a manipulação dos dados, foi utilizada a linguagem de programação Python e as bibliotecas Pandas e NumPy.

A base de dados gerada resulta em 268.028 registros e 103 atributos. Após realizar a modelagem da base de dados e remoção de valores inválidos, restaram 266.236 registros e 17 atributos. Para criar as séries temporais para o Brasil, para cada UF e regiões, foram aplicadas agregações de dados, de forma diária, semanal, mensal e anual, gerando um número de observações de, respectivamente, 9.862, 1.409, 324 e 27 observações. A base de séries temporais resultante possui 33 séries temporais com até 52.912 observações diárias, para cada tipo de agregação, totalizando 132 séries temporais.

4. Resultados preliminares

Para iniciar a análise exploratória das séries temporais de casos de suicídio no Brasil, foram visualizados os comportamentos das séries temporais com diferentes graus de agregação, conforme mostra a figura 1. Verifica-se uma dificuldade na interpretabilidade da

⁴<https://pcdas.icict.fiocruz.br/conjunto-de-dados/sistema-de-informacoes-de-mortalidade-sim/dicionario-de-variaveis/> Acesso em: 03 mar. 2024

⁵<http://tabnet.saude.es.gov.br/cgi/tabnet/sim/sim96/obtdescr.htm> Acesso em: 3 mar. 2024

⁶<http://www2.datasus.gov.br/cid10/V2008/cid10.htm> Acesso em: 3 mar. 2024

agregação de observações diárias, e como a agregação dos dados de forma mensal melhora a visualização.

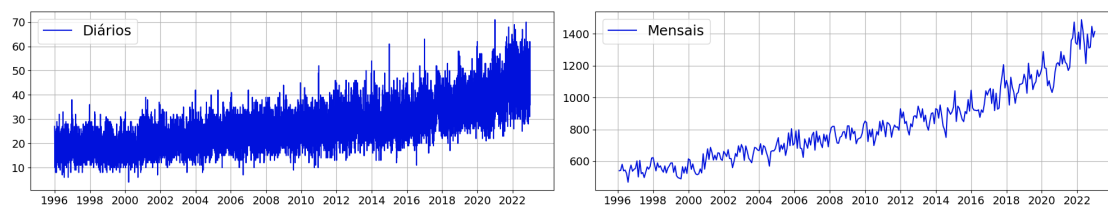


Figura 1. Casos diários e mensais de suicídio do Brasil entre 1996 e 2022

4.1. Decomposição das séries temporais

Para realizar a decomposição da série temporal, foi utilizado o algoritmo STL, presente na figura 2. Essa abordagem permite separar componentes sazonais, tendências e ruídos com alta precisão, sendo flexível e eficaz para capturar padrões não lineares em dados complexos (Cleveland et al. 1990).

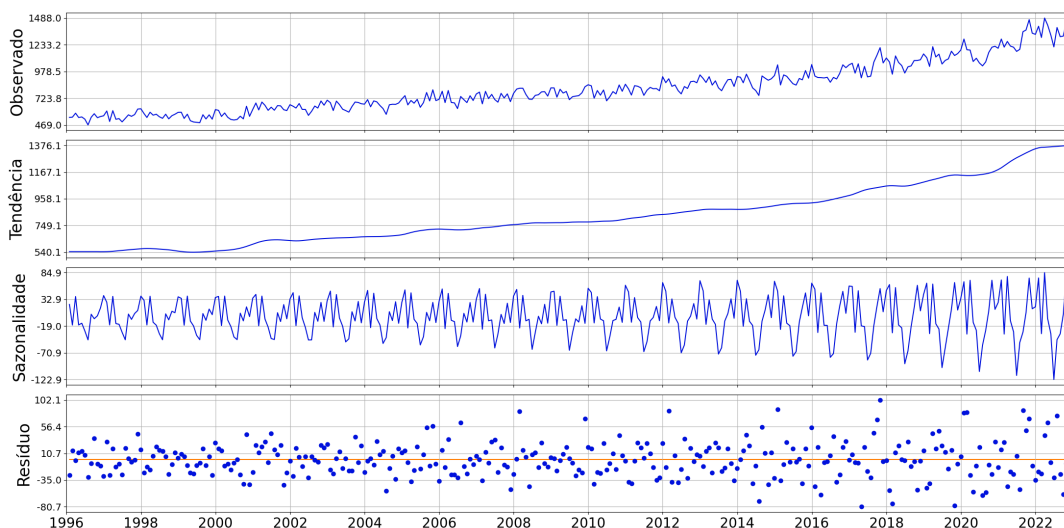


Figura 2. Decomposição da série temporal do Brasil: Casos mensais de suicídio entre 1996 e 2022

É possível visualizar que a tendência se manteve crescente durante todo o período coletado, e que a sazonalidade dos dados inicialmente apresentava um modelo aditivo de composição, mas que nos anos mais recentes têm apresentado um modelo híbrido ou multiplicativo. Esse fenômeno não parece ser exclusivo da sazonalidade, já que a amplitude no volume das observações consideradas ruído também aumenta.

4.2. Similaridade entre séries temporais

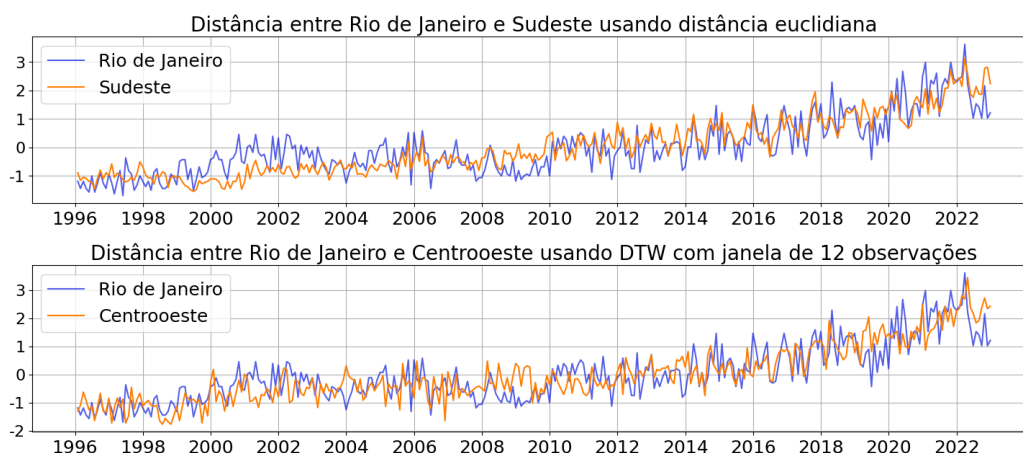


Figura 3. Comparação entre diferentes medidas de distância para o cálculo de similaridade

Para iniciar a medida de similaridade entre as séries temporais, foram utilizadas uma medida *lock-step* com distância euclidiana para medir a distância ponto a ponto, e uma medida elástica, DTW, para avaliar se o tipo de medida de distância geraria um resultado diferente.

Após calculadas as similaridades, foi avaliado se o estado em questão apresentava similaridade entre si e a região a qual pertence e depois entre os estados de sua mesma região. O exemplo de maior contraste, mostrado na figura 3, foi o Rio de Janeiro, que, para distância euclidiana, apresentou maior semelhança com a sua região, Sudeste, e com os estados da Bahia, Goiás e Minas gerais. Usando DTW, assemelhou-se com a região centro-oeste e os estados do Maranhão, Santa Catarina e Goiás.

Referências

- [Cleveland et al. 1990] Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition. *J. Off. Stat*, 6(1):3–73.
- [Coelho et al. 2021] Coelho, F. C., Baron, B. C., de Castro Fonseca, G. M., Reck, P., and Palumbo, D. (2021). Alertadengue/pysus: Vaccine.
- [Cruz et al. 2023] Cruz, W. G. N., Jesuino, T. A., Moreno, H. F., Santos, L. G., and de Almeida, A. G. (2023). Impact analysis of the brazilian suicide prevention campaign. *Trends in psychiatry and psychotherapy*, (AheadOfPrint):0–0.
- [da Mata et al. 2020] da Mata, K. C. R., Daltro, M. R., and Ponde, M. P. (2020). Perfil epidemiológico de mortalidade por suicídio no brasil entre 2006 e 2015. *Revista Psicologia, Diversidade e Saúde*, 9(1):74–87.
- [de Oliveira et al. 2020] de Oliveira, M. E. C., de Lima Gomes, K. A., Nóbrega, W. F. S., Gusmão, E. C. R., dos Santos, R. D., and Franklin, R. G. (2020). Série temporal do suicídio no brasil: o que mudou após o setembro amarelo? *Revista Eletrônica Acervo Saúde*, 48(48):e3191–e3191.
- [Ilic and Ilic 2022] Ilic, M. and Ilic, I. (2022). Worldwide suicide mortality trends (2000–2019): A joinpoint regression analysis. *World journal of psychiatry*, 12(8):1044.

- [Junior 2015] Junior, A. (2015). O comportamento suicida no brasil e no mundo the suicidal behavior in brazil and in the world. *Revista Brasileira de Psicologia*, 02:15–28.
- [Machado et al. 2014] Machado, M. F. S., Leite, C. K. S., and Bando, D. H. (2014). Políticas públicas de prevenção do suicídio no brasil: uma revisão sistemática. *RG&PP*, 4(2):348.
- [SVS 2021] SVS (2021). Boletim epidemiológico. secretaria de vigilância em saúde. Technical report, SVS.
- [SVSA 2024] SVSA (2024). Sistema de informação sobre mortalidade (sim).
- [WHO 1992] WHO (1992). The icd-10 classification of mental and behavioural disorders : clinical descriptions and diagnostic guidelines. World Health Organization. Disponível em: <https://apps.who.int/iris/handle/10665/37958>. Acesso em: 10 mai. 2023.
- [WHO 2021] WHO (2021). Suicide worldwide in 2019: global health estimates. Geneva: World Health Organization. License: CC BY-NC-SA 3.0 IGO, ISBN 978-92-4-002664-3.