INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SÃO PAULO

Danilo Bento Lessa

Felipe Domingues Bonfim

Lucas Nogueira De Souza

SP3071715

SP3071227

SP3072703

**Projeto de Estatística e Probabilidade -   
 Análise exploratória de dados**

SÃO PAULO

2023

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SÃO PAULO

Danilo Bento Lessa

Felipe Domingues Bonfim

Lucas Nogueira De Souza

SP3071715

SP3071227

SP3072703

**Projeto de Estatística e Probabilidade -   
 Análise exploratória de dados**

Atividade apresentada ao curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo.

Orientadora Prof.ª Dra. Josceli M. Tenorio

**SÃO PAULO**

**2023**

Sumário

[1 Introdução 3](#_Toc137938323)

[1.1 Descrição do problema 3](#_Toc137938324)

[1.2 Proposta da análise 4](#_Toc137938325)

[1.3 Descrição da base de dados 5](#_Toc137938326)

[2 Metodologia 5](#_Toc137938327)

[3 Análises 6](#_Toc137938328)

[4 Conclusão 6](#_Toc137938329)

[5 Referência bibliográfica 8](#_Toc137938330)

# Introdução

Este projeto tem como objetivo realizar uma análise exploratória dos dados de homicídio no Brasil, comparando o período pré-pandemia de COVID-19 (2018-2019) com o período durante a pandemia (2020-2021). Os dados de mortalidade utilizados foram extraídos das bases de dados do Sistema de Informação sobre Mortalidade(SIM) disponibilizadas pelo Ministério da Saúde através do portal openDataSUS.

O Sistema de Informação sobre Mortalidade (SIM) foi desenvolvido pelo Ministério da Saúde em 1975 com o objetivo de unificar mais de quarenta modelos de Declaração de Óbito utilizados anteriormente. Essa unificação permitiu a coleta de dados sobre mortalidade de forma padronizada em todo o país. O SIM desempenha um papel fundamental na obtenção de informações precisas e consistentes sobre as causas de morte no Brasil, auxiliando na formulação de políticas públicas de saúde e no monitoramento epidemiológico.

A análise foi conduzida com o objetivo de verificar se houve uma diminuição estatisticamente significativa nos casos de homicídios durante a pandemia (2020-2021) em comparação com os anos anteriores (2018-2019). Além disso, buscou-se correlacionar os resultados obtidos com fontes externas de informação confiáveis, como notícias baseadas em dados e bases complementares, a fim de verificar a consistência das conclusões obtidas. Essa abordagem permite uma análise mais precisa e embasada sobre o impacto da pandemia nos índices de homicídio no Brasil.

## Descrição do problema

Durante os anos de 2019 e 2020, o Brasil enfrentou um período de transição complexo em relação à segurança pública, com um impacto significativo da pandemia de COVID-19 na criminalidade e nos índices de homicídios. Essa crise sanitária teve efeitos diversos nas diferentes regiões do país, resultando em variações na quantidade de homicídios e em suas taxas.

Em 2019, o Brasil apresentava uma taxa de homicídios de aproximadamente 27,8 por 100.000 habitantes, de acordo com dados do Atlas da Violência 2021, produzido pelo Instituto de Pesquisa Econômica Aplicada (IPEA) em parceria com o Fórum Brasileiro de Segurança Pública. Com estes números, o Brasil ocupava a posição de número 10 no ranking mundial e a segunda colocação na América do Sul no ranking de países com maior taxa de homicídio em 2019, segundo o Relatório Global de Homicídios 2019 do Escritório das Nações Unidas sobre Drogas e Crime (UNODC).

Gráfico, Gráfico de barras

Descrição gerada automaticamenteEntretanto, apesar dos altos números apresentados, dados coletados pelo projeto Monitor da Violência, realizado pelo Núcleo de Estudos da Violência da USP, o G1 e o Fórum Brasileiro de Segurança Pública, indica que a taxa de homicídios no Brasil vem decrescendo de forma consistente desde 2017, como aponta a Figura 1 abaixo:

Figura 1 – Homicídios no Brasil (Fonte: Monitor da Violência, 2021)

## Proposta da análise

Este projeto tem como objetivo realizar uma análise exploratória dos dados coletados em relação à taxa de homicídio no Brasil, em cada uma das cinco regiões: Norte, Nordeste, Centro-Oeste, Sudeste e Sul. A proposta é confirmar se houve uma redução nessa taxa, realizando a verificação através da análise das bases de dados públicas do DataSUS do período entre 2018 e 2021 utilizando índices de significância e testes de hipótese para inferir informações.

## Descrição da base de dados

Os arquivos utilizados como base de dados para este projeto foram disponibilizados pelo Ministério da Saúde através do openDataSUS. As bases de dados utilizadas são denominadas Mortalidade Geral 2018, Mortalidade Geral 2019, Mortalidade Geral 2020 e Mortalidade 2021. Cada arquivo contém informações sobre os óbitos registrados nos respectivos anos. O arquivo Mortalidade Geral 2018 possui aproximadamente 438MB de dados, o Mortalidade Geral 2019 possui aproximadamente 449MB, o Mortalidade Geral 2020 possui aproximadamente 519MB e o Mortalidade 2021 possui 617MB de dados.

Os dados brutos estão estruturados em tabelas com 1.048.575 registros e 87 colunas cada. Essas tabelas fornecem informações detalhadas sobre cada óbito registrado, incluindo características como tipo de morte, contagem de homicídios, contagem total de mortes no Brasil, data do óbito, nome da região e idade da pessoa.

Para o propósito desta análise, os dados foram tratados e separados em grupos de amostras contendo 200.000 registros por ano. Essas amostras foram agrupadas em duas categorias: "Pré-pandemia de COVID-19" e "Durante a pandemia de COVID-19". O grupo "Pré-pandemia" consiste nas amostras coletadas nos anos de 2018 e 2019, enquanto as amostras coletadas nos dois anos subsequentes foram alocadas no grupo que representa o período durante a pandemia.

As variáveis selecionadas para a análise incluem o tipo de morte, a contagem de homicídios, a contagem total de mortes no Brasil, a data do óbito, o nome da região(Norte, Nordeste, Centro-oeste, Sudeste e Sul) e a idade da pessoa.

# Metodologia

Inicialmente, os dados foram agrupados em duas amostras distintas divididos em pré-pandemia e durante a pandemia, período considerado pré-pandêmico engloba os anos 2018 e 2019, já o período durante a pandemia considerou os dados dos anos 2020 e 2021. Essa divisão tem o objetivo de analisar possíveis efeitos causados pela pandemia de COVID-19 na taxa de homicídio.

Utilizando a linguagem R, foram criadas funções para transformar e adequar os dados às necessidades da análise. As informações relacionadas à região geográfica, mês de óbito e outras variáveis relevantes foram tratadas e organizadas para facilitar a visualização e compreensão dos resultados.

Foram geradas tabelas agregando o número total de mortes e homicídios em cada região, permitindo uma visão geral dos dados. Além disso, gráficos foram criados para ilustrar os homicídios totais ao longo do período analisado, evidenciando possíveis tendências ou variações significativas.

A análise também inclui a geração de intervalos de confiança para cada região, o que possibilita avaliar a precisão das estimativas. Além disso, foram realizados testes de hipótese para verificar se há diferenças estatisticamente significantes na taxa de homicídio entre as regiões e no Brasil como um todo.

Todo o processo de análise, incluindo tabelas e gráficos gerados, é demonstrado com detalhes na seção 3 deste documento.

# Análises

# Conclusão

Com base na análise realizada, foi possível confirmar uma redução estatisticamente significativa na taxa de homicídios em todas as cinco regiões do Brasil. Esses resultados respaldam a premissa inicial de que houve uma diminuição dos índices de violência letal no país. No entanto, é importante ressaltar que, com base nas informações analisadas, não é possível atribuir exclusivamente o impacto da pandemia de COVID-19 como o único responsável por essas reduções.

Diversos fatores devem ser considerados para uma compreensão mais abrangente dessas mudanças. Entre eles, destacam-se as políticas de segurança pública implementadas pelas autoridades em cada região, bem como a influência do crime organizado e do tráfico de drogas. Além disso, fatores socioeconômicos, como níveis de desigualdade, pobreza e acesso a oportunidades, também desempenham um papel importante na incidência de homicídios.

Embora a pandemia possa ter contribuído indiretamente para a redução da taxa de homicídios, através de medidas de isolamento social e restrições à circulação, não é possível afirmar categoricamente que esse foi o único fator determinante. A análise estatística indica uma tendência geral de queda nos índices de violência letal, mas é fundamental considerar a interação complexa entre esses diversos elementos e seu impacto na segurança pública.

# Referência bibliográfica

United Nations Office on Drugs and Crime (UNODC). **Global Study on Homicide**. [Online]. Disponível em: <https://www.unodc.org/documents/data-and-analysis/gsh/Booklet1.pdf>. Acesso em: 17 jun. 2023.

Instituto de Pesquisa Econômica Aplicada (IPEA) e Fórum Brasileiro de Segurança Pública. (2021). **Atlas da Violência 2021**. [Online]. Disponível em: <https://www.ipea.gov.br/atlasviolencia/arquivos/artigos/5141-atlasdaviolencia2021completo.pdf>. Acesso em: 17 jun. 2023.

Monitor da Violência. **As mortes violentas mês a mês no país** [Online]. Disponível em: <http://especiais.g1.globo.com/monitor-da-violencia/2018/mortes-violentas-no-brasil/#/dados-mensais-2022>. Acesso em: 17 de junho de 2023.