



# CRIME, VIOLÊNCIA E COVID-19: UMA ANÁLISE DE SÉRIES TEMPORAIS DAS REGIÕES SUL E SUDESTE DO BRASIL

Palavras-Chave: CRIMINALIDADE, VIOLÊNCIA, COVID-19

Autores(as):

RAPHAEL JORGE BULASCOSCHI, IE – UNICAMP

Prof. Dr. MARCELO JUSTUS DOS SANTOS (orientador), IE - UNICAMP

---

## INTRODUÇÃO:

A violência no Brasil é um problema histórico que permeia o debate público há muito tempo. De acordo com o Global Peace Index, o país ocupa o 128º lugar entre 163 países no ranking de paz e segurança. Homicídios e crimes violentos são peremptórios problemas sócio-políticos que animam diversas discussões no debate público; grupos e facções políticas se dividem em propor diversas soluções para o problema, mas o debate parece ser imobilizado ao longo das décadas.

Esse imobilismo pode ser associado a uma dificuldade em obter, tratar e interpretar dados sólidos referentes a estratégias de segurança pública no país. Assim, sem uma base estatística sólida que forneça evidências referentes a experiências de segurança pública, o debate fica limitado a uma questão de pontos de vista e opiniões infundadas.

Portanto, o presente trabalho busca explorar os dados de segurança pública no país (com o foco nos estados das regiões Sul e Sudeste) para fazer uma análise, erigida sobre bases de dados sólidas, sobre a trajetória de diversos indicadores de violência, levando também em conta os impactos da pandemia de COVID-19 e como a política de distanciamento social impacta essa trajetória nas séries temporais coletadas e exploradas. A partir dessa análise, desenvolver também modelos preditivos para as variáveis de interesse exploradas.

## METODOLOGIA:

A porção analítica deste trabalho se concentrará na análise das séries temporais e construção de modelos, a partir dessas séries temporais, preditivos que buscarão observar como tenderão a se comportar as variáveis analisadas. Para a análise de séries temporais, costuma-se fazer uso de modelos autorregressivos integrados de médias móveis (ARIMA); além disso, buscando explorar também o alto componente de variação sazonal das variáveis envolvidas nos dados de segurança pública, pode-se também construir modelos sazonais autorregressivos integrados de média móvel (SARIMA).

Para outra porção analítica do trabalho, buscando avaliar o impacto da pandemia de COVID-19 nas séries temporais, buscar-se-á aplicar análises de intervenção por meio da modelagem Box-Jenkins.

No entanto, para todos os processos envolvidos na porção analítica do trabalho, necessita-se que disponhamos de uma quantidade significativa (principalmente por se tratar de uma análise de séries temporais) de dados. Portanto, imbuí-se que uma parte significativa no tempo investido no desenvolvimento desse projeto fosse focado na construção e padronização de bases de dados que dispusessem de todos os dados essenciais para que fossem construídas as análises e modelos da melhor forma possível. Optou-se por uma coleta dos dados a partir de bases construídas diretamente pelo órgão de Segurança Pública de cada estado, pois julgou-se que, nesse caso, poderíamos estar em contato com dados mais confiáveis; além de serem também muitas vezes dados com um alcance temporal maior, o que nos interessaria mais nesse caso específico de se fazer uma análise de séries temporais. Sendo assim, uma parte substancial do projeto é focado justamente na coleta dos dados, sendo essa parte, embora demorada (por ser demasiadamente manual, em muitos casos), de importância substancial do trabalho. No entanto, por mais que a etapa de coleta, organização e padronização dos dados possa parecer mais intermediária no processo de concepção desse projeto, alguns resultados podem ser derivados da experiência derivada dessa etapa; resultados esses que serão discutidos mais à frente.

## RESULTADOS E DISCUSSÃO:

Tendo coletado e organizado os dados, buscou-se montar os modelos preditivos das variáveis selecionadas. Por um motivo de fatura de dados, focou-se em explorar os dados relacionados a furtos ou também, para alguns estados que não contavam com essa classificação em uma escala temporal satisfatória, relacionados a furtos de veículos, uma vez que se considerou essa variável uma *proxy* dos furtos para os estados que não contam com essa classificação.

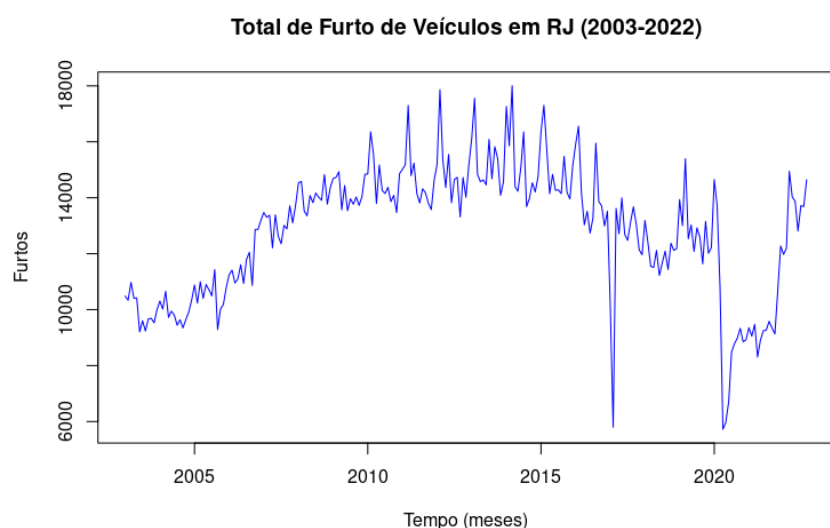


Gráfico 1 - Representação da série temporal de registros de furtos no estado do Rio de Janeiro (2003-2022). Elaboração própria. Com dados do ISP-RJ.

Observando o gráfico 1, podemos perceber dois fatores que são de suma importância para os modelos desenvolvidos ao longo do projeto. O primeiro deles é que a simples visualização gráfica dos dados nos permite notar um padrão sazonal em para os dados de furtos no estado do Rio de Janeiro. Esse tipo de comportamento poderá ser notado também em alguns outros estados explorados no projeto, embora o

componente sazonal não seja significativo para todos os estados e regiões explorados. Assim, para alguns conjuntos de dados serão preferíveis modelos ARIMA, enquanto para outros, que contam com o componente sazonal de maneira mais substantiva, usaremos modelos SARIMA. O segundo fator importante é o impacto da pandemia de COVID-19, bem como das decorrentes políticas de distanciamento social, na variável de interesse. Percebe-se, a partir de março de 2020 uma queda substancial da linha do gráfico. No entanto, o registro de furtos de veículo já volta a crescer com força a partir de 2021. Esse comportamento também poderá ser observado para outros estados explorados.

No que concerne o desenvolvimento dos modelos preditivos, buscou-se desenvolvê-los passando por uma série de etapas. Em um primeiro momento, executamos o upload dos arquivos de dados no programa Rstudio. Feito isso, dividimos a série temporal em duas séries separadamente, uma série de teste e uma de treino. A série de teste possui as 12 últimas observações da série, enquanto a série de treino possui n-12 observações, sendo n o número de observações totais da série, ou seja, a série de treino compreende todas as observações da série excetuadas aquelas contidas na série teste. Feito isso, buscaremos determinar o modelo a partir da função “*auto.arima*” do pacote analítico “Forecast” no Rstudio.

Feita essa divisão, partimos para uma avaliação da previsão desse modelo. Nessa parte, analisaremos os resíduos e os acertos de previsão. Verificando os resíduos, verificamos a autocorrelação dos resíduos a partir do teste de Ljung-Box. Os resultados são resumidos em um gráfico (Gráfico 2), que resumem em um único tanto o gráfico dos resíduos, a função de autocorrelação (FAC, ou ACF, na sigla em inglês), que demonstra se obtemos defasagens (correlações significativas), bem como a distribuição dos resíduos, que nos permite observar a normalidade da distribuição dos resíduos.

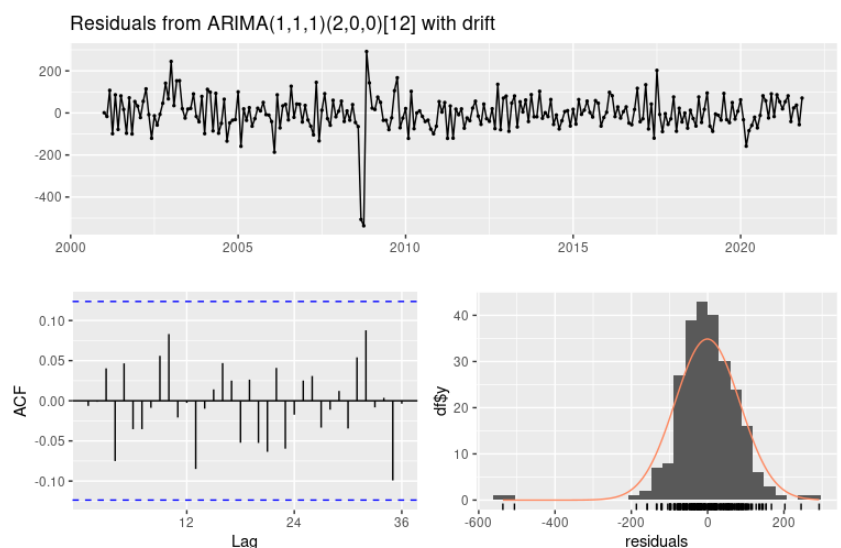


Gráfico 2 - Representação gráfica da análise dos resíduos do modelo SARIMA construído para a região de Araçatuba (SP). Elaboração própria.

Todos esses gráficos dispostos são resultados de uma análise a partir do modelo SARIMA já estimado. Nesse caso, para a região de Araçatuba, teremos um modelo SARIMA(1, 1, 1)(2, 0, 0). Ou seja, um modelo com um coeficiente autorregressivo, um coeficiente de integração e um coeficiente de médias móveis, para a parte não sazonal do modelo, e um coeficiente autorregressivo com sazonalidade e nenhum coeficiente de integração ou de médias móveis com sazonalidade.

Feita essa etapa, prosseguimos em sentido de testar a capacidade do modelo de prever as observações da amostra de teste a partir do quadrado médio do erro (MSE). Dado o modelo de treino, testamos uma previsão de 12 observações para frente. Feito isso, compararemos a acurácia da previsão em comparação com a amostra teste. Em seguida, compararemos o desempenho com alguma previsão de linha base que assume que o valor observado mais recente é o verdadeiro representante do futuro. Com isso, faremos uma comparação entre os valores do MSE encontrados pelo modelo “benchmark” e o modelo estimado. Com um valor do MSE menor para o modelo estimado, temos que o modelo fornece uma melhor precisão, pois possui valores melhores para os indicadores.

Assim, feita a análise dos acertos de previsão, bem como conhecendo o modelo, partimos para o momento que usaremos o modelo para construir previsão de fato. Nesse trabalho nos limitamos a fazer uma previsão para 12 períodos (meses) no futuro. Usando a função “*auto.arima*”, deixamos que o modelo ajuste automaticamente os parâmetros do modelo para a série temporal como um todo. Feito isso, obtemos um valor de previsão, bem como um intervalo de confiança para a variação, conforme exposto no Gráfico 3.

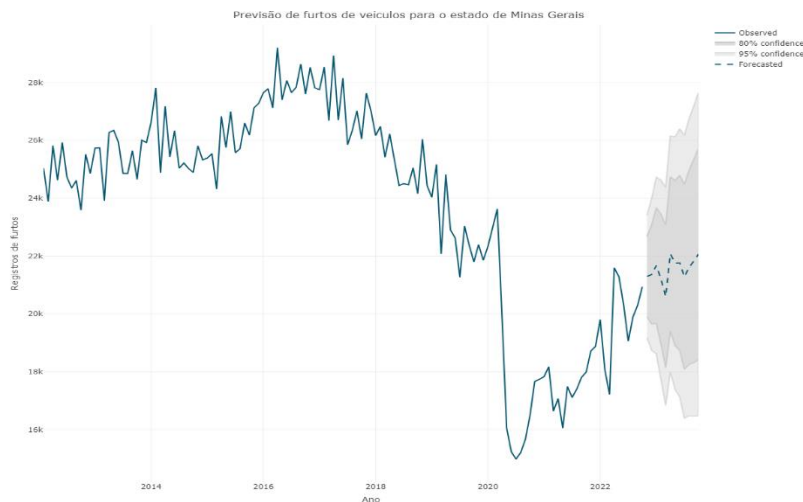


Gráfico 3 - Série temporal para o furto de veículos no estado de Minas Gerais, com a inclusão da previsão para 12 meses seguintes, bem como intervalos de variação estimados para 80% e 95% de nível de confiança. Elaboração própria, com dados da SSP-MG

## CONCLUSÕES:

Por se tratar de um projeto que se desenvolve a partir da análise de uma vasta gama de estados diferentes, cada um com as suas particularidades no que tange o estado da segurança pública, as conclusões generalizadas são um pouco limitadas. Não se pode, por exemplo, definir uma única previsão de tendência das séries de tempo para todos os tempos, dadas as suas particularidades. Podemos, ainda assim, desenvolveremos algumas reflexões acerca do processo envolvido na elaboração deste trabalho.

Compreendidos na primeira parte do projeto, conclusões importantes relacionadas a qualidade dos dados são relevantes. Primeiramente, torna-se visível que não contamos, no Brasil, com uma base de dados unificada de alto alcance temporal com dados mensais para todos os estados brasileiros. Por mais que haja um esforço nesse sentido, a partir da plataforma SINESP do Ministério da Justiça, o seu uso ainda é muito limitado, sendo usado nesse trabalho apenas em casos extremos de completa escassez de dados. Em segundo lugar, outro obstáculo na primeira parte do projeto foi a diferenciação metodológica entre os estados com relação a definição de certos crimes, o que fez com que

buscássemos tratar os dados a partir de um processo de padronização com atenção aos dicionários metodológicos de cada estado. Assim, a conclusão derivada da primeira parte do projeto é relacionada à uma dificuldade de acesso a dados públicos que seriam importantes para o desenvolvimento dessa e de outras pesquisas na área.

Quanto à parte analítica, conforme já citado, as conclusões variam de estado para estado. No entanto conclusões gerais que podem ser tiradas são relacionadas ao fato de crise sanitária relacionada ao alastramento da COVID-19, decorrente a partir do mês de março de 2020, ter efeitos significativos sobre as variáveis analisadas para praticamente todos os estados e regiões do Sul e Sudeste do Brasil. Além disso, um outro padrão que se segue para a maioria das séries temporais tratadas é a de um aumento dos registros de crimes a partir da metade do ano de 2020, com um crescimento ainda maior a partir de meados de 2021. Esse movimento pode estar fortemente relacionado à flexibilização de medidas de distanciamento social envolvidas nas políticas de controle da pandemia. Assim, configura-se uma tendência de alta para as séries temporais de maneira geral, podendo esse efeito ser derivado de um “carrego estatístico” com relação a uma queda abrupta nos meses subsequentes a março de 2020, mas não necessariamente significa necessariamente que a criminalidade está crescendo substantivamente, pode significar apenas que as variáveis estão convergindo para um nível de acordo com aquele observado para o período pré-pandemia.

---

## BIBLIOGRAFIA

Enders, W. **Applied Econometric Time Series** 3ª Edición. Wiley, 2010.

Institute for Economics and Peace. (2021). **Global Peace Index 2021**. Disponível em: <http://visionofhumanity.org/reports>

PENSSAN. **Inquérito Nacional sobre Insegurança Alimentar no Contexto da Pandemia no Brasil**. Disponível em: [http://olheparaafome.com.br/VIGISAN\\_Inseguranca\\_alimentar.pdf](http://olheparaafome.com.br/VIGISAN_Inseguranca_alimentar.pdf)

LIMA, Renato Sérgio de; BUENO, Samira; ALCADIPANI, Rafael. **Evolução das mortes violentas intencionais no Brasil**. in: Anuário Brasileiro de Segurança Pública 2021 - São Paulo: FBSP, 2021. Disponível em: <https://forumseguranca.org.br/anuario-brasileiro-seguranca-publica/>

CERQUEIRA, Daniel et al. **Atlas da Violência 2021** - São Paulo: FBSP, 2021. Disponível em: <https://forumseguranca.org.br/atlas-da-violencia/>

Razbash S, Wang E, Yasmeeen F (2023). **forecast: Forecasting functions for time series and linear models**. R package version 8.21, <<https://pkg.robjhyndman.com/forecast/>>.

BUENO, Rodrigo De Losso da Silveira. **Econometria de séries temporais - 2ª edição**. Cengage (2012)