



CRIME, VIOLÊNCIA E COVID-19: UMA ANÁLISE DE SÉRIES TEMPORAIS DAS REGIÕES NORTE E NORDESTE DO BRASIL

Palavras-Chave: Crime, Homicídios, COVID-19

Autores:

GIOVANNI DOS SANTOS VARGETTE, IE – UNICAMP

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

INTRODUÇÃO:

Segundo levantamento realizado pelo Monitor da Violência, parceria entre o G1, o Núcleo de Estudos da Violência da USP (NEV) e o Fórum Brasileiro de Segurança Pública, no ano de 2021 ocorreram 41.069 casos de mortes violentas no Brasil, com isso o país apresenta uma taxa de 19,3 casos a cada 100 mil habitantes. Apesar de ser um número elevado, o mesmo apresenta uma redução de 7% se comparado com o ano anterior. Porém, tal redução é causada principalmente por Unidades Federativas (UFs) das regiões Sul e Sudeste. Nos Estados que compõem estas duas regiões apenas o Rio de Janeiro e o Espírito Santo estão acima da média nacional, com 19,4 e 27,3 casos, respectivamente. Enquanto isso, Estados das regiões Norte e Nordeste despontam como os mais violentos do país, sendo Amazonas (36,8), Ceará (35,7) e Amapá (35,6) os líderes do ranking.

Se analisarmos o período de isolamento social durante a pandemia de Covid-19, tivemos um aumento dos casos de mortes violentas de 5% no país como um todo, mas alavancado principalmente pela região Nordeste, que contou com um aumento de 20% (Monitor da Violência, 2021).

De acordo com estudos recentes do Instituto Datafolha, e a reportagem da Agência Câmara de Notícias em 2021, houve uma redução nos casos de violência contra mulheres nas ruas, enquanto a violência doméstica aumentou. Em 2017, 39% das mulheres relataram violência nas ruas, mas esse número caiu para 19% em 2021. Porém, os casos de violência dentro das casas das vítimas aumentaram, como dito pela pesquisadora Samira Nunes.

A partir desse contexto, a presente pesquisa parte da hipótese de que, além de uma possível influência temporal, com sazonalidades e outros aspectos que infiram correlação no número de ocorrências de crimes, - e aqui trataremos mais especificamente da série temporal de homicídios - a crise sanitária, decorrente da pandemia de COVID-19, iniciada em Março de 2020, gerou algum tipo de impacto sobre a série temporal analisada, bem como sobre a sua tendência de crescimento, decréscimo ou estagnação.

METODOLOGIA:

Este projeto de Iniciação Científica foi dividido em duas etapas, sendo a primeira, a construção do banco de dados (BD), composto com índices de criminalidade, violência e produtividade policial, para os Estados que compõem as regiões Norte e Nordeste do Brasil. A segunda etapa foi dedicada para a realização das análises de impacto sobre as séries temporais de homicídio de cada Estado. Ambos os processos serão descritos abaixo.

Na construção dos BD's demos preferência para os dados divulgados diretamente pelos Estados como uma forma de reduzir a possibilidade de erros na divulgação dessas informações, porém nem sempre foi possível acessá-los. A busca foi realizada nos sites dos órgãos responsáveis pela segurança pública, como o nome dos órgãos variam em alguns Estados, optou-se por padronizá-los como Secretaria de Segurança Pública (SSP) referente a cada Estado, assim como é utilizado no Estado de São Paulo. Algumas UFs não possuíam divulgação em seu site, como foi o caso do Acre. Para contornar esse e outros problemas de divulgação de dados com que nos deparamos, utilizaram-se as informações do Sistema Nacional de Informações de Segurança Pública (SINESP).

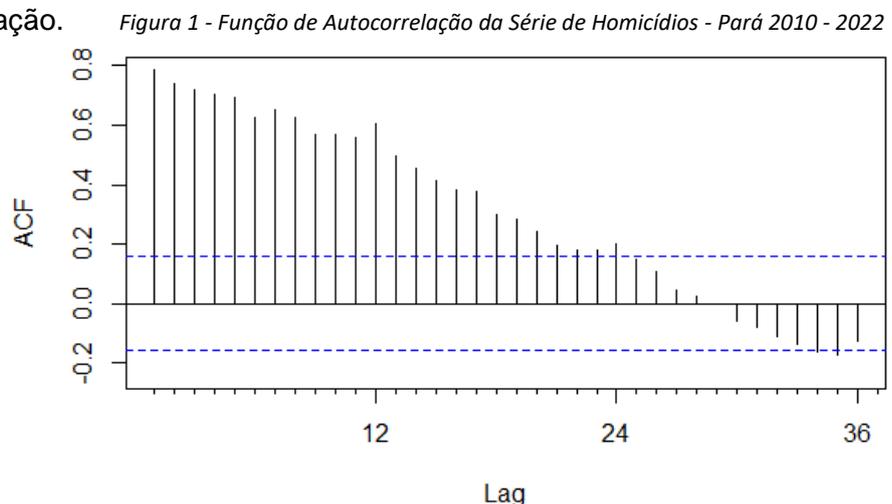
Com os bancos de dados estruturados optou-se por focar em uma série temporal para a realização das análises de impacto, a série escolhida foi a de homicídios. Os Estados do Piauí, Ceará e Paraíba não possuem divulgação específica para homicídios, nesses casos utilizou-se os dados de Mortes Violentas e Intencionais (MVI) para o Piauí, e os dados de Crimes Violentos Letais e Intencionais (CVLI) para Ceará e Paraíba.

Sendo assim, a metodologia aplicada sobre os dados coletados foi o Método Box-Jenkins, que consiste na divisão em etapas para a criação de um modelo matemático do estilo ARIMA ou SARIMA – Auto Regressivo Integrado de Médias Móveis e Auto Regressivo Integrado Sazonal de Médias Móveis – que seja capaz de simular o processo estocástico formador dos dados, possibilitando assim uma análise preditiva, baseando-se em rigor estatístico de nível de confiança de 95%. Importante ressaltar que todos os gráficos e análises construídos, bem como os modelos, foram feitos através da linguagem R e do *Software* RStudio, e que serão disponibilizadas as linhas de código construídas no processo.

A etapa inicial consiste na Especificação do modelo, ou seja, observar se os dados coletados possuem algum tipo de autocorrelação.

Tal análise é feita graficamente, através da análise da Função de Autocorrelação (FAC) da série, assim como no exemplo ao lado:

Ao analisarmos a figura 1, temos que a série de homicídios no Pará de 2010 até 2022, possui correlação estatisticamente significativa entre ela mesma até o



lag 24, que representa os meses anteriores, o que nos fornece algumas suspeitas, como por exemplo a existência de sazonalidade nesta série, dada por exemplo pelo aumento da correlação nos *lags* 12 e 24. A autocorrelação é uma característica essencial para modelos preditivos criados a partir de séries temporais, pois permitem que se tenha uma projeção do futuro baseada no passado.

A etapa seguinte corresponde a estimação do modelo, para isso utilizou-se a função `auto.arima` presente no pacote `forecast` do `software` RStudio como um meio de agilizar tanto a especificação quanto a estimação do modelo. Importante dizer que a função `auto.arima` nos serviu mais como um meio de nortear a especificação e a estimação do modelo, trazendo uma sugestão que comumente necessitou de alterações para se chegar no modelo buscado.

O terceiro passo é o diagnóstico sobre os resíduos provenientes do modelo construído, aqui se busca resíduos do tipo ruído branco gaussiano. Com isso, queremos dizer que buscamos modelos que nos gerem resíduos, ou seja os erros contidos naquele modelo, que não possuam autocorrelação, já que isso indicaria que estamos omitindo fatores importantes e que estes estão contidos dentro dos resíduos. Os resíduos também precisam ser homocedásticos, isto é, possuir variâncias constante. Caso isso não ocorra estaremos lidando com heterocedasticidade, e para isso deve-se aplicar um tipo de modelagem própria para contornar o problema. A última característica necessária para os resíduos é que os mesmos possuam distribuição normal. O modelo que apresentar essas três características é considerado bom, permitindo o avanço para a próxima etapa, a de previsão.

Na etapa da previsão aplicamos efetivamente o modelo, neste projeto de pesquisa nossa aplicação se deu através de uma análise de intervenção. Ou seja, estabelecemos o mês de março de 2020 como o início da pandemia de COVID-19. Sendo assim, estipulou-se um modelo pré-covid, com dados até março de 2020, após isso criou-se uma variável binária, que representa a ação da crise sanitária e do isolamento social. Através dessa ferramenta, com rigor estatístico, foi possível avaliar o impacto da pandemia de COVID-19 sobre as ocorrências de homicídios – CVLI e MVI para localidades específicas – dos Estados das Regiões Norte e Nordeste do Brasil.

Portanto, os modelos construídos seguem o seguinte formato:

Figura 2 - Notação Padrão de Modelo SARIMA

$$\text{SARIMA } (\underbrace{p, d, q}_{\text{Coeficientes não sazonais}}) (\underbrace{P, D, Q}_{\text{Coeficientes sazonais}})$$

Em que p é referente aos coeficientes autorregressivos, d ao coeficiente de integração e q ao coeficiente de médias móveis, todos não sazonais. Já P , é relativo aos coeficientes autorregressivos com sazonalidade, bem como D é o coeficiente de integração e Q o de médias móveis, ambos com sazonalidade.

RESULTADOS E DISCUSSÃO:

Dada a limitação de páginas do presente resumo, optou-se pela seleção de um Estado como exemplificação dos resultados obtidos. Sendo assim, apresentaremos os resultados referentes à série temporal de homicídios no Estado do Acre, utilizando dados provenientes do SINESP, com divulgação a partir de janeiro de 2015 até junho de 2022. Importante ressaltar também que nesta pesquisa seguiu-se a notação convencional para as significâncias estatísticas. Logo, (***) corresponde a significância estatística de 1%, (**) significância estatística de 5%, (*) significância estatisticamente de 10%, e NS para dados não estatisticamente significantes.

O modelo gerado pela função *auto.arima* para o Estado do Acre no período pré-pandemia, foi um ARIMA (1,0,0), ou seja, apenas um modelo autorregressivo de ordem 1. Após a implementação da variável *dummy* (variável binária), controle dos *outliers* encontrados, e a aplicação desse modelo sobre a série completa, obteve-se os resultados demonstrados na tabela ao lado:

Variável	Estimativa
Intercepto	3,1080 (0,0837)***
Ar1	0,4985 (0,0922)***
Xreg	-0,3479 (0,1442)**

Tabela 1 – Coeficientes do Modelo do Estado do Acre

Ao analisar a Tabela 1 temos que todos os coeficientes estimado são estatisticamente significantes. Podemos dar maior enfoque maior para o xreg, que diz respeito à variável *dummy*, portanto a mesma se mostra estatisticamente significativa, ou seja, temos evidência estatística de efeito da pandemia de COVID-19 sobre a série de homicídios do Estado do Acre, nesse caso de uma forma redutora, dado o sinal negativo do coeficiente. Importante ressaltar que o valor estimado do coeficiente atua marginalmente sobre a variável explicada do modelo, nesse caso as ocorrências de homicídio.

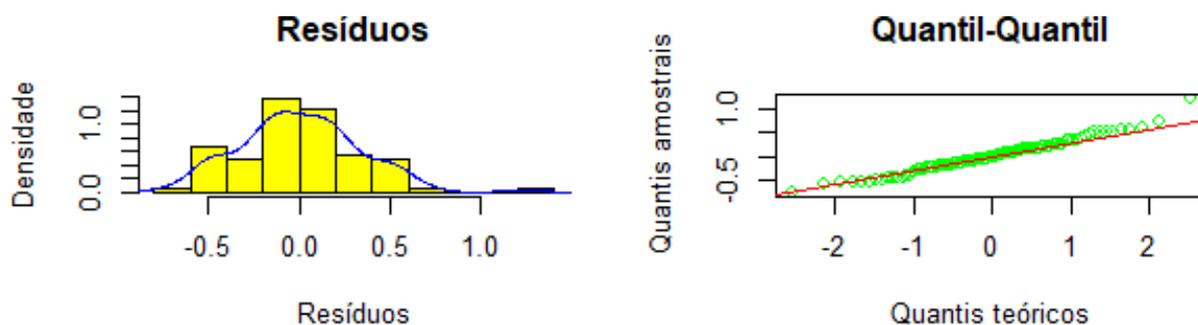
Os resíduos gerados pelo modelo também foram analisados, para autocorrelação utilizou-se o teste de Ljung-Box, para heterocedasticidade o teste de heterocedasticidade condicional autorregressiva (ARCH), já para inferir sobre a distribuição dos resíduos, o teste de Shapiro-Wilk.

Teste	P-valor
Ljung-Box	0,5705 ^{NS}
ARCH	0,789 ^{NS}
Shapiro-Wilk	0,1763 ^{NS}

Tabela 2 – Resultado dos Testes Realizados nos Resíduos do Modelo do Estado do Acre

Dados os resultados apresentados na Tabela 2, a hipótese nula de todos os testes não é recusada, logo o modelo apresenta resíduos do tipo Ruído Branco Gaussiano. Já a partir da análise da Figura 3, temos um reforço visual do comportamento dos resíduos da série, demonstrando graficamente o comportamento Ruído Branco Gaussiano.

Figura 3 – Análise Gráfica dos Resíduos do Modelo do Estado do Acre



CONCLUSÕES:

A partir da análise apresentada, temos um caso em que encontra-se evidências estatísticas no impacto da covid sobre a série temporal de homicídios, porém o comportamento de relação inversa entre o impacto pandêmico e os casos de homicídio não vão se repetir para todos Estados, nem mesmo serão todos os Estados que terão evidências de alguma relação entre os dois acontecimentos.

Portanto, podemos concluir a partir dessa breve exposição e focando nos resultados obtidos na investigação realizada no Estado do Acre, que a hipótese nula proposta por essa pesquisa foi recusada, logo a pandemia de COVID-19 apresentou fundamentos estatísticos de influência sobre os casos de homicídios no Estado. Nesse caso, esse impacto se deu negativamente, ou seja, segundo os dados apresentados e o modelo construído, o impacto da crise sanitária e do isolamento social causado por ela, atua sobre os registros de homicídios reduzindo-os em torno de 34%.

Vale a ressalva de que os resultados obtidos se deram a partir da metodologia aplicada, e que outras aplicações e métodos de análise de intervenção são de extrema importância, tanto para a validação do que foi proposto nesta pesquisa, quanto para a proposição de outras hipóteses.

BIBLIOGRAFIA

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

Violência contra as mulheres nas ruas cai durante a pandemia, mas aumenta dentro de casa. BITTAR, Paula, 2021. Disponível em: <https://www.camara.leg.br/noticias/797543-violencia-contra-as-mulheres-nas-ruas-cai-durante-a-pandemia-mas-aumenta-dentro-de-casa/>. Acesso em: 13/05/2022

Número de assassinatos cai 7% no Brasil em 2021 e é o menor da série histórica. G1, 2022.

Disponível em: <https://g1.globo.com/monitor-da-violencia/noticia/2022/02/21/numero-de-assassinatos-cai-7percent-no-brasil-em-2021-e-e-o-menor-da-serie-historica.ghtml>. Acesso em: 13/05/2022

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016.

WICKHAM, H., et al. (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, <<https://doi.org/10.21105/joss.01686>>.

WICKHAM H., et al. (2023). **dplyr: A Grammar of Data Manipulation**. R package version 1.1.1, <<https://CRAN.R-project.org/package=dplyr>>.

PEREIRA, R., GONCALVES, C. (2022). **geobr: Download Official Spatial Data Sets of Brazil**. R package version 1.7.0, <<https://CRAN.R-project.org/package=geobr>>.

WICKHAM, H., BRYAN, J. (2023). **readxl: Read Excel Files**. R package version 1.4.2, <<https://CRAN.R-project.org/package=readxl>>.

PFAFF, B. (2008) **Analysis of Integrated and Cointegrated Time Series with R**. Second Edition. Springer, New York. ISBN 0-387-27960-1

GRAVES, S. (2019). **FinTS: Companion to Tsay (2005) Analysis of Financial Time Series**. R package version 0.4-6, <<https://CRAN.R-project.org/package=FinTS>>.

HYNDMAN, R., et al., (2023). **forecast: Forecasting functions for time series and linear models**. R package version 8.21, <<https://pkg.robjhyndman.com/forecast/>>.