

CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ
ISSN on-line 2317-4536 / ISSN impresso 1413-9022 - v.51, p.38-53, 2021
DOI: 10.12957/cadest.2021.65627

HOMICÍDIO DOLOSO NO RIO DE JANEIRO: PREVISÕES USANDO MODELOS DE ESTADO E SUAVIZAÇÃO EXPONENCIAL, ARIMA E REDES NEURAIS AUTORREGRESSIVAS

Emerson Gonçalves Araújo (UERJ)
emerson.g.1322@gmail.com

Marcello Montillo Provenza (UERJ)
mprovenza@gmail.com

José Fabiano da Serra Costa (UERJ)
fabiano@ime.uerj.br

Resumo

O presente artigo tem como objetivo a análise dos números mensais de homicídios do estado do Rio de Janeiro ao longo do tempo, e de alguns modelos estatísticos de predição, para um melhor entendimento do contexto da violência letal. A base de dados provém de informações divulgadas pelo Instituto de Segurança Pública (ISP), durante o período entre janeiro de 2000 e dezembro de 2020. Com o uso do software R, foi possível gerar gráficos comparativos, analisar as estatísticas dos dados e fazer testes de normalidade, estacionariedade, tendência e sazonalidade para buscar entender melhor o comportamento desse grave delito no estado. Após essa primeira etapa, a série temporal foi ajustada para os modelos Autorregressivo Integrado de Média Móveis (ARIMA), de Estado e Suavização Exponencial (ETS) e Redes Neurais Autoregressivas (AR-NN). Para avaliação dos modelos, a base foi dividida em dois períodos: a base de treino correspondeu ao período entre 2000 e 2017 e a base de teste entre 2018 e 2019. Como 2020 foi um ano atípico devido a pandemia, optou-se por excluir esse ano das previsões. Observou-se através das métricas de previsão (MAE, RMSE e MAPE) que o modelo AR-NN(15,8) apresentara comportamento mais próximo dos valores observados.

Palavras-chave: Violência letal; Séries temporais; Modelos estatísticos; Previsão.

1. Introdução

A violência urbana é um tema recorrente na sociedade brasileira por razões bastante justificadas. A violência atinge a população como um todo, tanto urbana quanto rural e, em níveis significativos. Dentre as formas de violência que afetam a população de forma impactante, destaca-se o homicídio. Diversos estudos são feitos tendo como base esse assunto. Grande parte dos trabalhos sobre homicídios usam a abordagem quantitativa, avaliando essas ocorrências através de variáveis demográficas como sexo, idade, etc. (SOUZA *et al.*, 2007).

A elaboração de previsões para dados de violência não é uma tarefa fácil. A maior dificuldade em prever essas informações ocorre principalmente porque antecipar um movimento humano não é um processo simples. Alguns autores correlacionam a violência com problemas existentes na sociedade, como desigualdade social, impunidade, corrupção, crime organizado, etc. (SOARES, 2008; TAVARES *et al.*, 2016). Entretanto, modelos estatísticos têm sido estudados e utilizados em busca de um diagnóstico promissor.

A representatividade mais grave de violência é, com certeza, o homicídio. Sobre a população em geral e sobre as mídias, este crime tem grande impacto negativo, causando insegurança, medo, aumentando os níveis de ansiedade, comprometendo a cidadania e os direitos humanos. As pesquisas realizadas ao longo das últimas décadas destacam as intensidade e apreensão da área da saúde acerca dos homicídios (OLIVEIRA *et al.*, 2020).

Nos últimos dez anos, as taxas dos homicídios têm apresentado redução na região Sudeste do Brasil, especialmente em São Paulo e Rio de Janeiro. Essa diminuição também é notada em algumas variáveis demográficas como gênero, e nas regiões metropolitanas desses estados (SOARES FILHO *et al.*, 2020).

Diante deste cenário, é importante a atuação de pesquisadores no estudo do comportamento deste delito. A previsão de dados em segurança pública contribui para orientação de políticas públicas e o trato dos agentes criminais na direção da tomada de decisões. Sendo assim, o objetivo deste trabalho é elaborar estimativas dos indicadores relacionados aos homicídios no estado do Rio de Janeiro, utilizando técnicas quantitativas.

2. Estatística e Segurança Pública

Fundamental para trabalhar com estatística é poder utilizar dados confiáveis. Em segurança pública e violência letal, tal princípio também se aplica. Afinal, os resultados somente poderão ser considerados adequados se o banco de dados utilizado refletir de forma bastante objetiva a situação existente.

Sobre o tema, Provenza *et al.* (2017) realizam um estudo estatístico das bases do Instituto de Segurança Pública¹ (ISP) e do Ministério da Saúde através de testes não paramétricos, com objetivo de buscar entendimento do cenário da violência no município do Rio de Janeiro e comparar os dados publicados entre as bases da segurança e da saúde.

De toda forma, desde que utilizada uma base de dados confiável, é bem plausível realizar análises estatísticas interessantes sobre segurança pública, violência letal e demais crimes, o que vem sendo demonstrados em alguns trabalhos recentes no Brasil.

Felix (2013) utiliza os dados da Polícia Militar do Estado de São Paulo, no período de janeiro de 2009 a agosto de 2012, para realizar uma análise de regressão avaliando estatisticamente o potencial probabilístico de mortes evitadas com a apreensão de armas de fogo pela polícia, numa evidente abordagem sobre a questão do desarmamento no estado e na cidade de São Paulo.

Nery *et al.* (2012) realizam um estudo avaliando 13.205 setores censitários e os 96 distritos censitários na cidade de São Paulo. Através de registros dos homicídios dolosos registrados no município entre 2000 e 2008 são apresentadas as taxas brutas de homicídios dolosos por 100.000 habitantes para cada setor censitário, assim como as taxas bayesianas globais e locais. Para identificar a existência de padrões de distribuição espacial dos homicídios, são utilizadas as técnicas BoxMap e o índice de Moran.

Nunes *et al.* (2020) apresentam um modelo de previsão ARIMA (*Autoregressive Integrated Moving Average*) para realizar um estudo do homicídio doloso, no período de 2012 a 2016, no estado da Bahia, utilizando os dados registrados e disponibilizados pela Secretaria de Segurança Pública do Estado da Bahia.

Cardoso *et al.* (2016) utilizam os dados do Instituto de Segurança Pública do estado do Rio de Janeiro, para análise da evolução temporal e tendência das taxas de homicídio doloso, de autos de resistência e de pessoas desaparecidas, através de modelo de regressão *Joinpoint*, considerando como variável dependente as referidas taxas e como variável independente o ano calendário.

3. Metodologia

O banco de dados utilizado neste estudo provém dos registros de ocorrência da Polícia Civil do estado do Rio de Janeiro. As informações são consolidadas e divulgadas pelo Instituto de Segurança Pública (ISP). O homicídio doloso foi observado no período entre janeiro de 2000 e dezembro de 2020. Foi utilizado o *software* R (R CORE TEAM, 2019) para elaboração dos gráficos e previsões.

A base de treino dos dados correspondeu ao período compreendido entre janeiro de 2000 e dezembro de 2017 e a base de teste entre janeiro de 2018 e dezembro de 2019. Considerando que o ano de 2020 foi atípico devido a pandemia, este trabalho optou por não considerar os dados de 2020.

Os modelos estatísticos empregados para as previsões foram o Autorregressivo Integrado de Média Móveis (ARIMA), de Estado e Suavização Exponencial (ETS) e Redes Neurais Autorregressivas (AR-NN).

3.1. Testes de Normalidade

Os testes de distribuição teórica são utilizados para verificar se probabilidade associada a um conjunto de dados pode ser aproximada pela distribuição Gaussiana. Neste trabalho foram usados os testes Shapiro-Wilk (W), Kolmogorov-Smirnoff (D_n), Anderson-Darling (A), Cramer-von Misses (CVM), Lilliefors (D), Pearson (P) e Shapiro-Francia (W') expostos nas equações (1 a 7, respectivamente). Na hipótese nula (H_0), os dados seguem uma distribuição normal. Na hipótese alternativa (H_1), os dados não seguem uma distribuição normal (LUCAMBIO, 2008).

$$W = \frac{(\sum_{i=1}^n a_i y_{(i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

$$D_n = \sup_x |F_n(x) - F(x)| \quad (2)$$

$$A = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \ln(p_{(i)}) + \ln(1 - p_{(n-i+1)}) \quad (3)$$

$$CVM = \frac{1}{12n} + \sum_{i=1}^n \left(p_{(i)} - \frac{2i-1}{2n} \right) \quad (4)$$

$$D = \max \left(\max_{i=1, \dots, n} \left(\frac{i}{n} - p_{(i)} \right), \max_{i=1, \dots, n} \left(p_{(i)} - \frac{i-1}{n} \right) \right) \quad (5)$$

$$P = \sum \frac{(C_i - E_i)^2}{E_i} \quad (6)$$

$$W' = \frac{(\sum_{j=1}^n a_j y_{(j)i})^2}{\sum_{j=1}^n (y_{(j)i} - \bar{y})^2} \quad (7)$$

3.2. Testes de Estacionariedade

Uma série temporal é estacionária quando suas características estatísticas (média, variância, autocorrelação, etc.) são constantes ao longo do tempo. Ou seja, trata-se de uma série que se desenvolve aleatoriamente no tempo, em torno de uma média constante, refletindo alguma forma de equilíbrio estatístico estável. Na prática, a maioria das séries apresentam algum tipo de não estacionariedade (MADDALA; LAHIRI, 2009).

Na hipótese nula (H_0), a série temporal é estacionária. Na hipótese alternativa (H_1), a série temporal não é estacionária. Os testes usados para verificar a estacionariedade foram Dickey-Fuller Aumentado (ADF) e Phillips-Perron (PP) expostos nas equações (8 e 9, respectivamente) (LEYBOURNE; NEWBOLD, 1999).

$$ADF = \delta + bt + \lambda Y_{t-1} + \sum_{s=1}^p \beta_s \Delta Y_{t-1} + \varepsilon_t \quad (8)$$

$$PP = \delta + bt + \lambda Y_{t-1} + \varepsilon_t \quad (9)$$

3.3. Testes de Tendência

O movimento sistemático de aumento ou redução, linear ou não linear, de uma série temporal ao longo do tempo é chamado de tendência. Na hipótese nula (H_0), a série temporal não apresenta tendência. Na hipótese alternativa (H_1), a série temporal apresenta tendência. Os testes usados para verificar a tendência foram Wald-Wolfowitz (Z), Mann-Kendall (S) e o coeficiente de correlação de Spearman (ρ) expostos nas equações (10, 11 e 12, respectivamente) (SIEGEL; CASTELLAN JR, 2006; STEINKE, 2020).

$$Z = \frac{R - \left(\frac{2n_1n_2}{n_1+n_2} + 1\right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}} \quad (10)$$

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sign}(x_j - x_k) \quad (11)$$

$$\rho = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2 - 1)} \quad (12)$$

3.4. Testes de Sazonalidade

Sazonalidade é a presença de variações que ocorrem em intervalos regulares específicos de tempo. Pode ser semanal, mensal, trimestral, etc. Existem dois tipos de sazonalidade (determinística e estocástica). A sazonalidade determinista ocorre quando se pressupõe um padrão sazonal regular e estável no tempo. Desta forma, pode-se prever o comportamento sazonal perfeitamente a partir de dados anteriores. A sazonalidade estocástica ocorre quando a componente sazonal da série varia com o tempo. Este procedimento pode ser utilizado quando se tem um padrão sazonal constante.

Na hipótese nula (H_0), a série temporal não apresenta sazonalidade. Na hipótese alternativa (H_1), a série temporal apresenta sazonalidade. Os testes usados para verificar a sazonalidade foram Kruskal-Wallis (T_1) e Friedman (H), expostos nas equações (13 e 14, respectivamente) (SIEGEL; CASTELLAN JR, 2006).

$$T_1 = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_j^2}{n_j} - 3(N+1) \quad (13)$$

$$H = \frac{12}{nk(k+1)} \sum_{j=1}^k R_j^2 - 3n(k+1) \quad (14)$$

3.5. Modelos de Estado e Suavização Exponencial (ETS)

O nome das técnicas que em geral as classes de métodos de previsão usam para o tratamento das causas de flutuações em séries de tempo é o alisamento (ou suavização). Os modelos ETS são uma classe especial dos modelos de alisamento exponencial. É usada

uma metodologia para identificar três caracteres utilizando a terminologia de framework. O primeiro caractere revela o erro (A, M ou Z), o segundo demonstra a tendência (N, A, M ou Z) e o terceiro a sazonalidade (N, A, M ou Z). Deste modo, tem-se que: N = nenhum; A = aditivo; M = multiplicativo; e Z = seleção automática (HYNDMAN; ATHANASOPOULOS, 2018). A equação das observações é (15):

$$Y_t = z_t \alpha_t + \varepsilon_t \quad (15)$$

E a equação de estado é (16):

$$\alpha_t = T_t \alpha_{t-1} + R_t n_t \quad (16)$$

onde α_t é o vetor de estados; ε_t são os ruídos não correlacionados; n_t é o vetor de ruídos serialmente não correlacionados; z_t , T_t e R_t são matrizes do sistema.

3.6. Modelos ARIMA

A metodologia Box-Jenkins consiste em ajustar os modelos ARIMA a um conjunto de dados. Tal método consiste em propor e adequar séries temporais de dados lineares estacionários (ou não-estacionários). A estratégia para a construção do modelo é baseada em um ciclo iterativo, no qual a escolha da estrutura é baseada nas próprias observações. Os estágios do ciclo iterativo são especificação, identificação, estimação e diagnóstico (BOX *et al.*, 2015, MORETTIN; TOLÓI, 2018).

Os modelos ARIMA partem do pressuposto que os valores de uma série temporal armazenam uma relação de dependência onde cada valor atual pode ser explicado pelo anterior. O objetivo é determinar os três componentes que configuram qualquer estrutura: p (parâmetros autorregressivos), d (processos de diferenciação) e q (parâmetros de médias móveis), formando assim os modelos ARIMA(p,d,q) (BOX *et al.*, 2015, MORETTIN; TOLÓI, 2018).

A Função de Autocorrelação (FAC) e a Função de Autocorrelação Parcial (FACP) são usadas para estimar os parâmetros da modelagem Box-Jenkins. A FAC é uma medida da correlação que fornece todas as autocorrelações da variável X_t (divididas por k unidades de tempo) nos tempos anteriores ($X_{t-1}, X_{t-2}, \dots, X_{t-k}$). A FACP é a correlação parcial entre X_t e X_{t-k} , ou seja, é a correlação entre X_t e X_{t-k} após terem sido descontadas as interferências passadas de $X_{t-1}, X_{t-2}, \dots, X_{t-k-1}$ (BUENO, 2011; BOX *et*

al., 2015).

O modelo ARIMA(p,d,q) é apropriado para a previsão de séries temporais cujo processo estocástico não é estacionário. Logo, a série original passará por diferenciações para torná-la estacionária. Estes são os processos de diferenciação (ou ordem de integração - d) (BUENO, 2011; GUJARATI; PORTER, 2011; BOX *et al.*, 2015).

Após identificada a ordem de integração do modelo ARIMA, precisa-se estimar os parâmetros autorregressivo (p) e média móvel (q). Sendo assim, pode-se utilizar três métodos de estimação: momentos, mínimos quadrados e máxima verossimilhança (GUJARATI; PORTER, 2011; BOX *et al.*, 2015). O método geral para estimação de modelos de Box-Jenkins (e usado neste trabalho) é o da máxima verossimilhança (17).

$$W_t = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \text{ com } W_t^d = \Delta^d Z_t \quad (17)$$

3.7. Modelos de Redes Neurais Autorregressivos

A técnica de Redes Neurais Autorregressivos (AR-NN) agrega o modelo estatístico autorregressivo (AR) com o aprendizado de máquina das redes neurais artificiais. As redes neurais artificiais habitualmente contêm uma parte linear e outra parte não linear. Enquanto nas redes neurais a variável resposta é função de um vetor de variáveis explicativas, o modelo redes neurais autorregressivas contempla os valores defasados da variável resposta como variáveis explicativas (18) (HAYKIN, 2009).

$$Z_t = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i} + \sum_{j=1}^h \Psi \left(\gamma_{0j} + \sum_{i=1}^p \gamma_{ij} Z_{t-i} \right) \beta_j + \varepsilon_t \quad (18)$$

3.8. Métricas de Previsão

Quando modelos de projeções são efetuados, é interessante observar os erros associados para determinar se os resultados se encontram dentro de limites toleráveis. Há medidas conhecidas de previsões para examinar a qualidade dos modelos propostos, das quais podem abranger valores absolutos ou percentuais. Neste trabalho, foram usadas a Raiz do Erro Quadrado Médio (RMSE), o Erro Médio Absoluto (MAE) e o Erro Percentual Médio Absoluto (MAPE), conforme (19, 20 e 21), respectivamente (OLIVEIRA *et al.*, 2010).

$$RMSE = \sqrt{\frac{\sum (Z_t - \hat{Z}_t)^2}{n}} \quad (19)$$

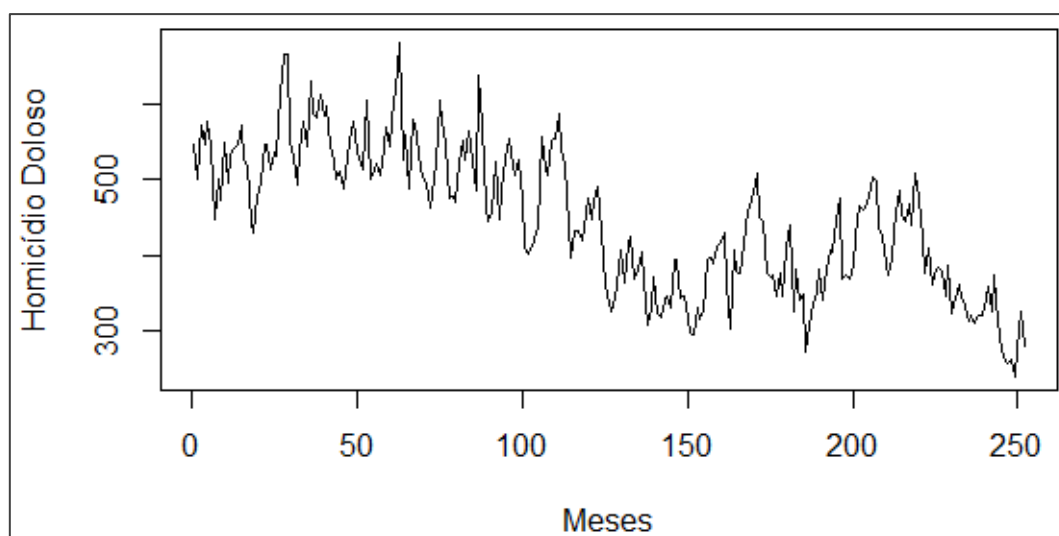
$$MAE = \frac{\sum |Z_t - \hat{Z}_t|}{n} \quad (20)$$

$$MAPE = \frac{\sum |Z_t - \hat{Z}_t|}{n} * 100 \quad (21)$$

4. Resultados

A Figura 1 apresenta a série histórica dos homicídios dolosos no estado do Rio de Janeiro no período compreendido entre janeiro de 2000 e dezembro de 2020. Em uma primeira análise visual, é possível verificar a existência de tendência de queda ao longo do tempo. Também são notadas algumas oscilações ocorrendo durante determinados períodos. É provável que se tenha sazonalidade na série.

Figura 1 - Série histórica dos homicídios dolosos no Rio de Janeiro entre 2000 e 2020



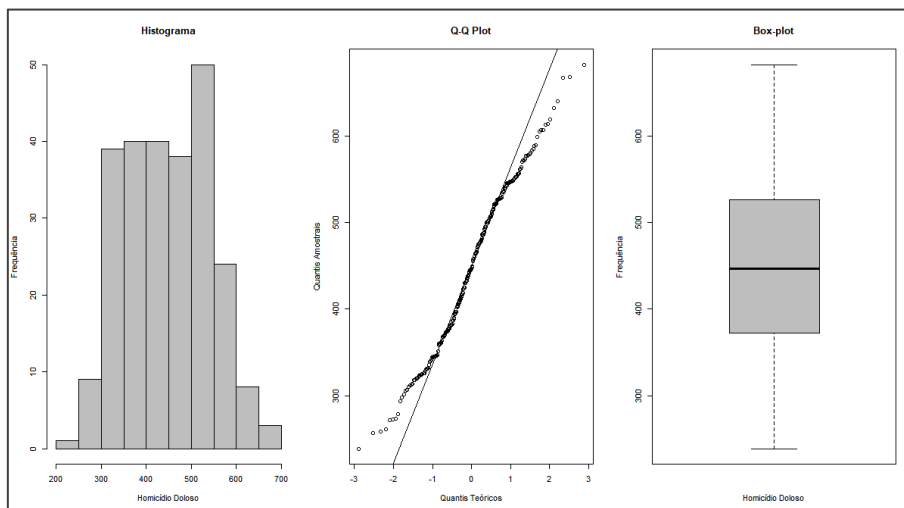
Fonte: ISP (2021)

A Tabela 1 mostra as estatísticas referentes aos dados da série histórica. Ao longo do período de 252 meses, a média de homicídios foi de 447,9 e o desvio padrão de 94,2. O valor mínimo foi de 239 vítimas, em setembro de 2020, e o valor máximo 682 vítimas, em março de 2005.

Tabela 1 - Estatísticas dos homicídios

Estatística	Valor
Tamanho amostral (n)	252
Média	447,9
Desvio padrão	94,2
Valor mínimo	239
1º quartil (Q ₁)	372,5
Mediana	447
3º quartil (Q ₃)	526
Valor máximo	682

A Figura 2 demonstra o histograma, o *Q-Q plot* e o gráfico de caixa (*box-plot*) da série. A análise visual do histograma mostra que, a princípio, os dados não se distribuem normalmente. Os valores estão concentrados entre 300 e 550. Essa hipótese é reforçada pelo *Q-Q plot*, onde a linha traçada parece não estar alinhada aos pontos. O *box-plot* revela que não há valores discrepantes (*outliers*).

Figura 2 - Histograma, *Q-Q plot* e *box-plot* dos homicídios

Para confirmar a hipótese de não normalidade das observações, a Tabela 2 apresenta os testes realizados sobre distribuição dos dados. Considerando um nível de significância de 5%, foi rejeitada a hipótese nula. O p-valor de todos os testes mostraram resultados abaixo de 0,05.

Tabela 2 - Testes sobre distribuição dos dados

Teste	Estatística de teste	p-valor
Shapiro-Wilk	0,981	0,002
Kolmogorov-Smirnov	1,000	< 0,001
Anderson-Darling	1,666	< 0,001
Cramer-Von mises	0,286	< 0,001
Lilliefors	0,068	0,006
Pearson	28,627	0,027
Shapiro-Francia	0,983	0,006

A partir da análise exploratória dos dados, alguns diagnósticos específicos para séries temporais devem ser elaborados. A Tabela 3 exhibe os testes de estacionariedade executados. Considerando um nível de significância de 5%, foi rejeitada a hipótese nula. A série de homicídios não é estacionária. O p-valor dos testes indicam resultados abaixo de 0,05.

Tabela 3 - Testes de estacionariedade

Teste	Estatística de teste	p-valor
Dickey-Fuller	-5,9	0,01
Phillips-Perron	-60,06	0,01

Rejeitada a hipótese de estacionariedade, a Tabela 4 apresenta os testes de tendência. Considerando um nível de significância de 5%, foi rejeitada a hipótese nula. Os testes feitos revelaram que existe tendência relativa a série de homicídios. O p-valor dos testes ficaram abaixo de 0,05. Pela análise visual gráfica, confirma-se a inclinação de queda nos valores.

Tabela 4 - Testes de tendência

Teste	Estatística de teste	p-valor
Wald	-12,687	< 0,001
Spearman	-0,746	< 0,001
Kendall	-0,537	< 0,001

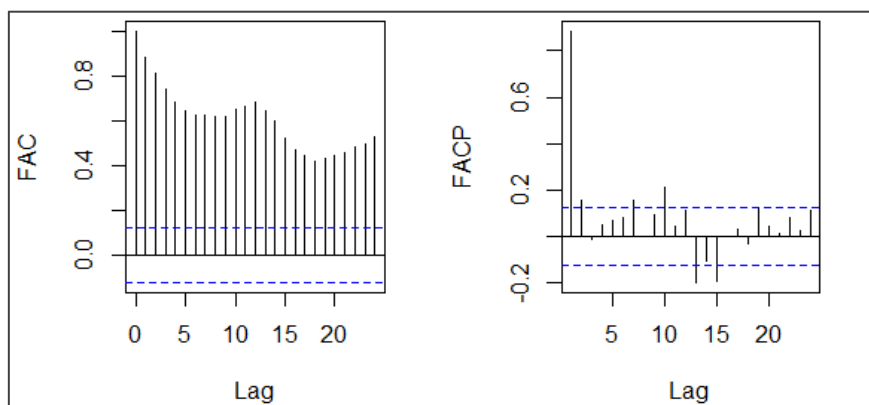
Para comprovar as oscilações na série de homicídios, foram realizados testes de sazonalidade demonstrados através da Tabela 5. Considerando um nível de significância de 5%, foi rejeitada a hipótese nula. O p-valor dos testes estão abaixo de 0,05. Sendo assim, existe sazonalidade nos valores em determinados períodos específicos.

Tabela 5 - Testes de sazonalidade

Teste	Estatística de teste	p-valor
Kruskal-Wallis	26,06	0,006
Friedman	110,74	< 0,001

O próximo passo, antes da elaboração das previsões, é avaliar a FAC e a FACP da série temporal. Na Figura 3 pode-se ver que a FAC não apresenta decaimento exponencial (ou esse decaimento é muito lento). A FACP parece ser truncada no lag 2, porém, outros lags extrapolam os limites de confiança. Desse modo, não se pode definir um modelo mais simples para a série de homicídios. A análise desses dois gráficos revela que é necessário realizar diferenciações para torná-la estacionária e, posteriormente, treinar a modelagem para que sejam executadas as previsões.

Figura 3 - Função de Autocorrelação e Função de Autocorrelação Parcial



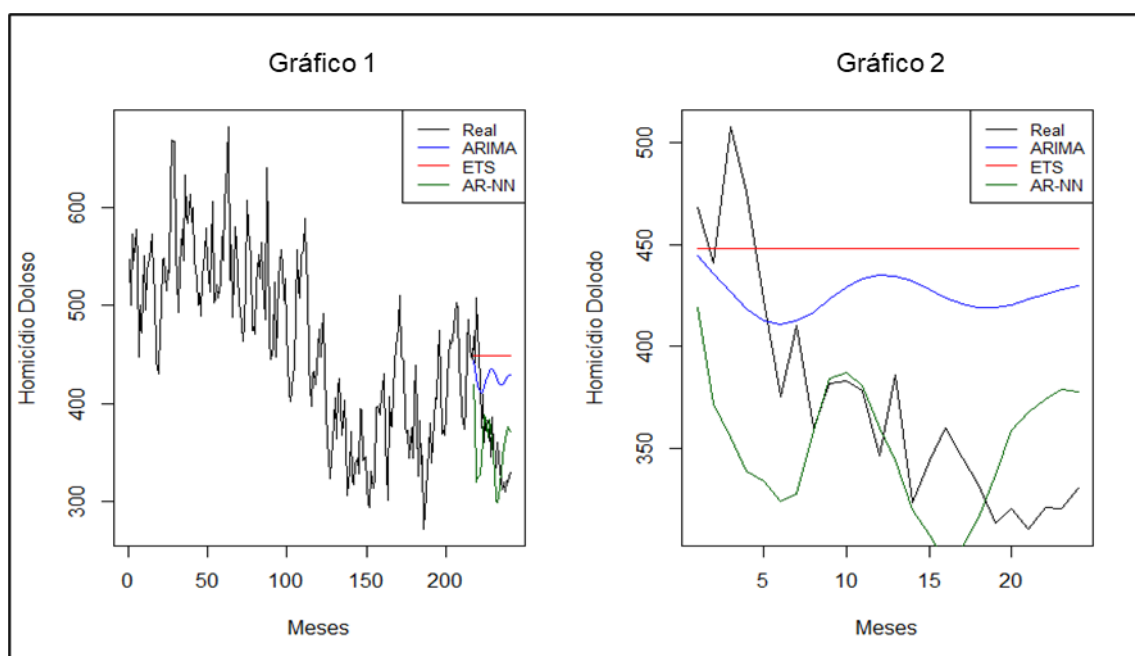
Os modelos ajustados para as previsões ficaram próximos. Através do pacote *forecast* do R, foram usadas as funções “auto.arima” para a modelagem ARIMA, “ets” para o modelo ETS e “nnetar” para as Redes Neurais Autorregressivas. Os resultados foram: ARIMA(3,1,2), ETS(A,N,N) e AR-NN(15,8). O ETS(A,N,N) configura uma suavização exponencial simples com erros aditivos.

Considerando a pandemia, é necessário notar que o ano de 2020 foi atípico, devido inclusive aos diversos tipos de restrições ocorridas diante da expansão do vírus da COVID-19, incluindo fechamento de bares, shoppings, restaurantes, escolas, universidades, etc. No Brasil, alguns municípios estabeleceram o *lockdown*, aonde somente serviços essenciais estavam disponíveis a população, como supermercados e farmácias. Desse modo, optou-se por excluir o ano de 2020, visto que esses dados podem

ter sofrido interferência devido a esse novo momento vivido pela população.

A Figura 4 apresenta dois gráficos que compara os dados reais com as previsões para 2018 e 2019. O Gráfico 1 demonstra todo o período da série de homicídios entre janeiro de 2000 e dezembro de 2019, juntamente com as bases de treino, teste e as previsões para os últimos 24 meses. O Gráfico 2 revela os dados comparados na base de teste e as previsões (período compreendido entre janeiro de 2018 e dezembro de 2019). Graficamente, o modelo AR-NN(15,8) parece apresentar comportamento mais próximo dos valores observados.

Figura 4 - Previsão dos homicídios entre janeiro de 2018 e dezembro de 2019



Para avaliar esses três modelos, foram utilizadas três métricas de previsão. A Tabela 6 mostra os resultados encontrados. O AR-NN(15,8) obteve melhores resultados para as previsões 2018-2019.

Tabela 6 - Métricas de previsão para 2018-2019

Modelos	MAE	RMSE	MAPE
ETS(A,N,N)	84,0	92,8	24,3
ARIMA(3,1,2)	66,8	75,1	19,2
AR-NN(15,8)	48,3	63,9	12,3

5. Considerações finais

O objetivo deste trabalho foi analisar o contexto dos homicídios dolosos ao longo do tempo e realizar previsões com o uso de três modelos: Box-Jenkins, ETS e Redes Neurais Autorregressivas. Através de testes estatísticos, foi visto que a série não dispõe de normalidade nos dados, não possui valores discrepantes, não é estacionária, e, que há tendência de queda e sazonalidade.

A análise temporal foi elaborada separando a base em dois períodos: a base de treino compreendeu o período entre janeiro de 2000 e dezembro de 2017; e a base de teste para os modelos foi utilizada no período compreendido entre janeiro de 2018 e dezembro de 2019. Dado que houve o início de uma pandemia em 2020, este ano não foi considerado para as previsões. O modelo AR-NN(15,8) estabeleceu os melhores resultados nas três métricas (MAE, RMSE e MAPE).

Como sugestão para estudos futuros no desenvolvimento das previsões, podem ser usados os modelos de regressão (GUJARATI; PORTER, 2011) e as técnicas de aprendizado de máquina (*machine learning*) (HASTIE *et al.*, 2009; IZBICKI; DOS SANTOS, 2020).

Referências

- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time Series Analysis: Forecasting and Control**. 5th Edition, John Wiley & Sons, 2015.
- BUENO, R. L da S. **Econometria de Séries Temporais**. 2ª Edição, Cengage Learning, 2011.
- CARDOSO, F. L. M. G.; CECCHETTO, F. R.; CORRÊA, J. S.; SOUZA, T. O. D. Homicídios no Rio de Janeiro, Brasil: uma análise da violência letal. **Ciência & Saúde Coletiva**, 21(4), 1277-1288, 2016.
- FELIX, S. A. Armas versus Vidas: Análise de Regressão sobre o Impacto da Apreensão de Armas nos Homicídios. **Revista do Laboratório de Estudos da Violência da UNESP/Marília**, edição 11, 2013.
- GUJARATI, D. N.; PORTER, D. C. **Econometria Básica-5**. Amgh Editora, 2011.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer Science & Business Media, 2009.
- HAYKIN, S. S. **Neural Networks and Learning Machines**. Simon Haykin, 2009.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. OTexts, 2018.
- IZBICKI, R.; DOS SANTOS, T. M. **Aprendizado de Máquina: uma Abordagem Estatística**. Rafael Izbicki, 2020.
- LEYBOURNE, J.; NEWBOLD, P. The behaviour of Dickey–Fuller and Phillips–Perron tests under the alternative hypothesis. **The Econometrics Journal**, v. 2, n. 1, p. 92-106, 1999.

LUCAMBIO, F. Diferentes testes para verificar normalidade de uma amostra aleatória. **UFPR**, PR, Brasil. 2008. Disponível em: https://docs.ufpr.br/~lucambio/CE225/2S2009/Normal_test.pdf. Acesso em 16 fev. 2021.

MADDALA, G. S; LAHIRI, K. **Introduction to Econometrics**. 4^a. Ed, John Wiley & Sons Ltd, UK, 2009.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise De Séries Temporais: Modelos Lineares Univariados**. Editora Blucher, 2018.

NERY M. B.; PERES M. F. T.; CARDIA, N.; VICENTIN, D.; ADORNO, S. Regimes espaciais: dinâmica dos homicídios dolosos na cidade de São Paulo entre 2000 e 2008. **Rev Panam Salud Publica**. 32(6):405–12, 2012.

NUNES, S. I. O. S.; SILVA FILHO, A. M.; DA SILVA, C. A. L. Modelagem temporal dos homicídios dolosos registrados na Bahia no período 2012 a 2016. Uma abordagem com o modelo ARIMA. **In: A Produção do Conhecimento Interdisciplinar nas Ciências Ambientais**, DOI 10.22533/at.ed.18420100214, Atena Editora, 2020.

OLIVEIRA, A. C. S.; Souza, A. A.; Lacerda, W. S.; Gonçalves, L. R. Aplicação de redes neurais artificiais na previsão da produção de álcool. **Ciência e Agrotecnologia**, v. 34, n. 2, p. 279-284, 2010.

OLIVEIRA, A. L. S.; LUNA, C. F.; SILVA, M. G. P. Homicídios do Brasil na última década: uma revisão integrativa. **Ciência & Saúde Coletiva**, v. 25, p. 1925-1934, 2020.

PROVENZA, M. M.; SERRA COSTA, J. F.; MIRANDA, I. P.; SILLI, M. T.; MARQUES, K. S. Homicídio doloso na cidade do Rio de Janeiro: Uma comparação entre as bases da segurança e da saúde. **DILEMAS: Revista de Estudos de Conflito e Controle Social**, Vol.10, no 2, pp. 297-321, 2017.

SIEGEL, S.; CASTELLAN JR. N. J. **Estatística Não-Paramétrica para Ciências do Comportamento**. Artmed Editora, 2006.

SOARES, G. A. D. **Não Matarás: Desenvolvimento, Desigualdade E Homicídios**. Editora FGV, 2008.

SOARES FILHO, A. M.; DUARTE, E. C.; MERCHAN-HAMANN, E. Tendência e distribuição da taxa de mortalidade por homicídios segundo porte populacional dos municípios do Brasil, 2000 e 2015. **Ciência & Saúde Coletiva**, v. 25, p. 1147-1156, 2020.

SOARES FILHO, A. M.; SOUZA, M. F. M.; GAZAL-CARVALHO, C.; MALTA, D. C.; ALENCAR, A. P.; SILVA, M. M. A.; MORAES NETO, O. L. Análise da mortalidade por homicídios no Brasil. **Epidemiologia e Serviços de Saúde**, v. 16, n. 1, p. 7-18, 2007.

STEINKE, V. A.; MELO, L. A. M. P.; MELO, M. L.; DA FRANCA, R. R.; LUCENA, R. L.; STEINKE, E. T. Trend analysis of air temperature in the federal district of Brazil: 1980–2010. **Climate**, v. 8, n. 8, p. 89, 2020.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <https://www.R-project.org>. 2019.

TAVARES, R.; CATALAN V. D. B.; ROMANO, P. M. M.; MELO, E. M. Homicídios e vulnerabilidade social. **Ciência & Saúde Coletiva**, v. 21, p. 923-934, 2016.

MURDER IN RIO DE JANEIRO: FORECASTS USING EXPONENTIAL SMOOTHING, ARIMA, AND AUTOREGRESSIVE NEURAL NETWORKS MODELS

Abstract

This article aims to analyze the monthly numbers of homicides in the state of Rio de Janeiro and some statistical prediction models for a better understanding of the context of lethal violence. The database comes from the Instituto de Segurança Pública (ISP) during the period between January 2000 and December 2020. Using the R software it was possible to generate comparative graphs, analyze data statistics, perform tests of normality, stationarity, trend, and seasonality to have a better understand the behavior of this crime in the state. After this first step, the time series was adjusted to the Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing (ETS), and Autoregressive Neural Networks (AR-NN) models. To evaluate the models, the base was divided into two periods: the training base (2000 to 2017), and the test base (2018 to 2019). As 2020 was an atypical year due to the pandemic, it was decided to exclude this forecast year. It was observed through the forecast metrics (MAE, RMSE, and MAPE) that the AR-NN(15.8) model presented a behavior closer to the observed values.

Key-words: *Lethal violence; Time series; Statistical models; Forecast.*