

**Leonardo Luís Barcellos Raposo**

**Análise e previsão dos homicídios dolosos de  
jovens no estado do Rio de Janeiro**

Niterói - RJ, Brasil

04 de maio de 2021

**Leonardo Luís Barcellos Raposo**

**Análise e previsão dos homicídios  
dolosos de jovens no estado do Rio  
de Janeiro**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Moisés Lima de Menezes

Co-Orientador(a): Núbia Karla de Oliveira Almeida

Niterói - RJ, Brasil

04 de maio de 2021

**Leonardo Luís Barcellos Raposo**

**Análise e previsão dos homicídios dolosos de  
jovens no estado do Rio de Janeiro**

Monografia de Projeto Final de Graduação sob o título “*Análise e previsão dos homicídios dolosos de jovens no estado do Rio de Janeiro*”, defendida por Leonardo Luís Barcellos Raposo e aprovada em 04 de maio de 2021, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof. Dr. Moisés Lima de Menezes**  
Departamento de Estatística – UFF

---

**Profa. Dra. Núbia Karla de Oliveira Almeida**  
Departamento de Estatística – UFF

---

**Prof. Dra. Ludmilla Da Silva Viana Jacobson**  
Departamento de Estatística – UFF

---

**Prof. Dra. Maria Cristina Bessa Moreira**  
Departamento de Estatística – UFF

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

R219a Raposo, Leonardo Luís Barcellos  
Análise e previsão dos homicídios dolosos de jovens no  
estado do Rio de Janeiro / Leonardo Luís Barcellos Raposo ;  
Moisés Lima de Menezes, orientador ; Núbia Karla de Oliveira  
Almeida, coorientador. Niterói, 2021.  
57 f. : il.

Trabalho de Conclusão de Curso (Graduação em  
Estatística)-Universidade Federal Fluminense, Instituto de  
Matemática e Estatística, Niterói, 2021.

1. Violência. 2. Homicídios Dolosos. 3. Rio de Janeiro. 4.  
Séries Temporais. 5. Produção intelectual. I. Menezes,  
Moisés Lima de, orientador. II. Almeida, Núbia Karla de  
Oliveira, coorientador. III. Universidade Federal Fluminense.  
Instituto de Matemática e Estatística. IV. Título.

CDD -

# Resumo

A violência em todo mundo é algo amplamente destacado pela mídia como um grande problema para toda a sociedade de forma geral, gerando consequências irreversíveis a todos de forma direta e indireta. A população mais jovem é comumente afetada por diversas formas de violência, principalmente em relação aos homicídios dolosos, quando há intenção de cometer tal ato. Visando monitorar o desenvolvimento deste tipo de violência, utilizou-se a análise de séries temporais. Este trabalho tem como objetivo identificar o perfil da vítima, bem como do agressor e a possível relação entre ambos, visando também destacar períodos do dia e locais do estado do Rio de Janeiro onde seriam mais frequentes as ocorrências e realizar a previsão de homicídios dolosos para o ano de 2020 através dos modelos de Holt-Winters e Box & Jenkins em busca de um modelo que ajuste bem os dados. Estatísticas de aderência foram utilizadas para comparação de modelos e com a análise feita, foi visto que os Modelos de Box & Jenkins obtiveram os melhores resultados quando comparadas as estatísticas de aderência com os Modelos de Holt-Winters.

Palavras-chave: Violência. Homicídios Dolosos. Rio de Janeiro. Séries Temporais. Holt-Winters e Box & Jenkins.

# Dedicatória

Foi pensando nas pessoas que executei este projeto, por isso dedico este trabalho a todos aqueles a quem esta pesquisa possa ajudar de alguma forma.

# Agradecimentos

Em primeiro lugar, à Deus, que fez com que meus objetivos fossem alcançados durante todos os meus anos de estudos.

Aos amigos Bernardo, Isabelly, Larissa, Natan e Rodrigo, por todo o apoio e pela ajuda durante toda graduação, que muito contribuíram para a realização deste trabalho.

Aos meus pais, família e amigos, que me incentivaram nos momentos difíceis e compreenderam a minha ausência enquanto eu me dedicava à realização deste trabalho.

Aos professores Moisés e Núbia, por terem sido meus orientadores e terem desempenhado tal função com dedicação e amizade.

Agradeço também os professores que me acompanharam mais de perto durante a graduação.

Agradeço à Camilla por ter estado ao meu lado em todos os momentos dessa jornada e ter me dado todo apoio e amor.

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 12
1.1	Violência Urbana . . . . .	p. 12
1.2	Revisão Bibliográfica . . . . .	p. 16
1.3	Objetivos . . . . .	p. 18
1.4	Organização do Trabalho . . . . .	p. 18
<b>2</b>	<b>Materiais e Métodos</b>	p. 19
2.1	Base de dados . . . . .	p. 19
2.2	Séries Temporais . . . . .	p. 20
2.2.1	Autocorrelação . . . . .	p. 21
2.2.2	Função de Autocorrelação . . . . .	p. 22
2.2.3	Função de Autocorrelação Parcial . . . . .	p. 22
2.2.4	Correlogramas . . . . .	p. 23
2.3	Operadores . . . . .	p. 24
2.3.1	Operador de defasagem . . . . .	p. 24
2.3.2	Operador de diferenças . . . . .	p. 24
2.4	Teste de Significância Individual . . . . .	p. 24
2.5	Testes de avaliação do comportamento da série temporal . . . . .	p. 25
2.5.1	Teste de Raiz Unitária . . . . .	p. 25



2.5.2	Testes de Normalidade . . . . .	p. 27
2.5.2.1	Teste de Shapiro-Wilk . . . . .	p. 27
2.5.2.2	Teste de Jarque-Bera . . . . .	p. 28
2.5.2.3	Teste Lilliefors . . . . .	p. 28
2.6	Modelos de Previsão . . . . .	p. 29
2.6.1	Modelos de Holt-Winters . . . . .	p. 29
2.6.2	Modelos de Box & Jenkins . . . . .	p. 30
2.6.2.1	Processos Autorregressivos . . . . .	p. 31
2.6.2.2	Processos de Médias Móveis . . . . .	p. 31
2.6.2.3	Modelos Autorregressivos e de Médias Móveis . . . . .	p. 32
2.6.2.4	Modelos ARIMA . . . . .	p. 32
2.6.2.5	Modelos SARIMA . . . . .	p. 33
2.7	Análise de Resíduos . . . . .	p. 34
2.8	Estatísticas de Aderência . . . . .	p. 35
2.8.1	MAD ( <i>Mean Absolute Deviation</i> ) ou Desvio Médio Absoluto . . . . .	p. 35
2.8.2	MSE ( <i>Mean Square Error</i> ) ou Erro Quadrático Médio . . . . .	p. 36
2.8.3	RMSE ( <i>Root Mean Square Error</i> ) ou Raiz Quadrada do Erro Quadrado Médio . . . . .	p. 36
2.8.4	MAPE ( <i>Mean Absolute Percentage Error</i> ) ou Erro Médio Percentual Absoluto . . . . .	p. 36
2.8.5	R <sup>2</sup> - máximo . . . . .	p. 36
2.8.6	AIC . . . . .	p. 37
2.8.7	BIC . . . . .	p. 37
2.8.8	HQC . . . . .	p. 38
2.9	Metodologia . . . . .	p. 38
<b>3</b>	<b>Análise dos Resultados</b>	p. 40

3.1	Características dos homicídios . . . . .	p. 40
3.2	Modelos de Holt-Winters . . . . .	p. 43
3.3	Modelos de Box & Jenkins . . . . .	p. 45
3.4	Escolha do modelo final . . . . .	p. 51
<b>4</b>	<b>Conclusão</b>	p. 53
	<b>Referências</b>	p. 55

# Lista de Figuras

1	Exemplo de Correlograma da FAC . . . . .	p. 23
2	Exemplo de Correlograma da FAC dos resíduos não correlacionados. . .	p. 35
3	Homicídios dolosos por mês no estado do Rio de Janeiro . . . . .	p. 40
4	Homicídios dolosos por dia da semana no estado do Rio de Janeiro . . .	p. 42
5	Série original e previsão modelada via Holt-Winters . . . . .	p. 44
6	Correlograma da FAC dos Resíduos . . . . .	p. 44
7	FAC e FACP do log dos homicídios . . . . .	p. 46
8	FAC e FACP com uma diferença sazonal do log dos homicídios . . . . .	p. 47
9	Série original e previsão modelada via Box & Jenkins . . . . .	p. 49
10	Correlograma da FAC dos Resíduos . . . . .	p. 50
11	Normalidade dos Resíduos . . . . .	p. 50
12	Série dos dados de 2018 e 2019 com previsão para 2020 . . . . .	p. 52

# Lista de Tabelas

1	Equações do Modelo de Holt Winters . . . . .	p. 29
2	Equações do Modelo de Holt Winters com <i>Damped Trend</i> . . . . .	p. 30
3	Comportamento teórico da FAC e FACP para alguns modelos . . . . .	p. 32
4	Estatísticas Descritivas . . . . .	p. 41
5	Distribuição de frequências absolutas e relativas ao total de óbitos por homicídio de características referentes às ocorrências, à vítima e ao agressor	p. 41
6	Estatísticas de Aderência para os Modelos de Holt-Winters . . . . .	p. 43
7	Constantes de Amortecimento . . . . .	p. 43
8	Componentes Sazonais . . . . .	p. 43
9	Testes de Normalidade . . . . .	p. 45
10	Testes de Normalidade com transformação logarítmica . . . . .	p. 45
11	Significância dos Parâmetros dos Modelos com $P = 2$ . . . . .	p. 48
12	Estatísticas de Aderência para os Modelos de Box & Jenkins . . . . .	p. 48
13	Parâmetros do Modelo de Box & Jenkins . . . . .	p. 49
14	Testes de Normalidade para os Resíduos . . . . .	p. 51
15	Estatísticas de Aderência para os Modelos selecionados . . . . .	p. 51
16	Previsões Mês a Mês . . . . .	p. 51

# 1 Introdução

## 1.1 Violência Urbana

Um problema social que está constantemente em evidência no dia a dia das pessoas, principalmente nos grandes centros urbanos, é a violência. É recorrente nos meios de comunicação de massa, notícias diárias de ações violentas que amedrontam a população, causando uma sensação constante de insegurança. De acordo com a Organização Mundial da Saúde (OMS), não existe um único fator que poderia explicar o motivo de algumas violências ocorrerem em determinadas regiões ou porque determinados indivíduos se comportam de maneira violenta. Um conjunto relacional de diversos fatores socioculturais resultam na prática da violência. Dessa maneira, as discussões acerca da saúde e da segurança pública são fundamentais para sua prevenção na sociedade. [1].

A OMS define o ato violento como o uso intencional do poder ou da força, de modo que possa causar lesões físicas e/ou psicológicas. Apesar da consequência ocasionada pela atitude violenta, essa definição refere-se especificamente a intencionalidade que resultou nessa prática [2]. Dessa maneira, acidentes domésticos ou grande parte das lesões provocadas por acidentes de trânsito são considerados incidentes não intencionais, e não se encaixam nessa definição [1].

Há uma dificuldade muito grande quando discutimos a intencionalidade dentro da descrição de violência. A intenção de usar a força pode não significar obrigatoriamente o desejo de causar um dano que resulte em uma lesão, pois a violência se diferencia de episódios não intencionais que podem resultar em estragos físicos ou mentais. Nesse sentido, é possível ocorrer uma dissemelhança entre o que se pretende e a consequência obtida. Uma pessoa pode cometer um ato violento e não interpretar sua atitude como tal. Nas relações familiares, por exemplo, uma mãe pode agredir seu filho na intenção de educá-lo, mas causar nessa criança, ainda que sem intenção, traumas e consequências psíquicas. Segundo Walters & Parke (1964) [3], a violência é determinada culturalmente. Em algumas culturas agredir sua esposa é aceitável socialmente, de forma que nem a

vítima nem o agressor considere esse ato como violento. No entanto, mesmo que esses comportamentos não sejam reconhecidos como uma forma de violência, eles trazem danos à saúde das pessoas que foram violentadas. [1].

Em 2017, no mundo, homicídios mataram mais que conflitos armados, 463.821 mortes contra 89 mil mortes, segundo relatório do Escritório das Nações Unidas sobre Drogas e Crime (UNODC), uma das agências especializadas no enfrentamento ao crime organizado transnacional da Organização das Nações Unidas (ONU) [4]. Este número de homicídios representa 6,1 mortes a cada 100.000 habitantes [5]. Quanto à população jovem, segundo o Fundo das Nações Unidas para a Infância (UNICEF), em 2015, a cada sete minutos, em algum lugar do mundo, uma criança ou adolescente foi morto pela violência. Somente em 2015, mais de 82 mil jovens de 10 a 19 anos perderam suas vidas a partir de homicídios ou de alguma forma de conflito armado. Desses óbitos, 24.5 mil foram registrados na América Latina e no Caribe [6].

Alguns estudos investigaram possíveis fatores associados à violência contra jovens e identificaram como prováveis causas as questões demográficas e o tráfico de drogas na América Latina, embora afirmem que esse não seja um fenômeno exclusivamente latino-americano. Citam por exemplo, os Estados Unidos, onde os jovens são duas vezes mais vítimas de crime do que os adultos acima de 25 anos de idade [7].

Entre os países latino-americanos, o Brasil aparece com a segunda maior taxa de homicídios a cada cem mil habitantes, ficando atrás apenas da Venezuela. Com 30,5 mortes a cada 100.000 pessoas, segundo o relatório do UNODC, a taxa brasileira é cinco vezes maior que a média global. Cabe ainda ressaltar que, no ano de 2017, o Brasil registrou o maior número bruto de homicídios em todo o mundo [5]. Ainda sobre esta problemática, o Instituto de Pesquisa Econômica Aplicada (IPEA), em seu mapeamento das mortes violentas com dados de 2007 a 2017, indica uma taxa de homicídios ainda maior, 31,6 mortes a cada 100.000 habitantes, que se trata do maior nível histórico de letalidade violenta intencional no país, sendo mais que o dobro da taxa do Iraque, por exemplo [8]. Esta diferença entre as taxas das pesquisas pode ser justificada pois a definição legal de homicídio com intenção de matar difere entre os países, no Brasil, de acordo com o Código Penal, homicídio é o ato de matar alguém [9] enquanto que o UNODC em seu Estudo Global Sobre Homicídio define:

Dentro da grande gama de tipos de morte violenta, o elemento primordial do homicídio intencional é sua completa ligação com o perpetrador direto, o que consequentemente exclui mortes causadas por guerras e conflitos, mortes auto infligidas

(suicídio), mortes devido a intervenções legais ou motivos justos (como autodefesa, por exemplo) e mortes quando houve negligência do perpetrador mas este não tinha a intenção de tirar uma vida (homicídio não intencional) [10].

No Brasil, de acordo com o mapeamento do IPEA as taxas de homicídios nos anos de 2007 a 2017 variaram de forma distinta, tendo aumento no Acre, no Rio Grande do Norte e na Bahia. Em contrapartida, foi verificada uma diminuição destas taxas, neste período, no Distrito Federal, no Espírito Santo e no Paraná. Em relação aos números totais de óbitos, no ano de 2017, o estado da Bahia registrou a maior quantidade de homicídios no país, com 7.487, seguido do Rio de Janeiro e Ceará, com 6.416 e 5.433, respectivamente. Observou-se também que 75,5% das vítimas eram negros (pretos ou pardos), resultando em uma taxa de 43,1 homicídios a cada 100.000 negros, enquanto que a taxa em não negros (brancos, amarelos e indígenas) foi de 16,0, indicando que para cada indivíduo não negro que morreu vítima de homicídio, aproximadamente 2,7 negros foram mortos por este motivo. Os estados com maiores taxas de homicídios de negros encontram-se na região Nordeste, sendo eles: Rio Grande do Norte (87,0), Ceará (75,6), Pernambuco (73,2), Sergipe (68,8) e Alagoas (67,9). Ao mesmo tempo que os que possuem a menor taxa são: São Paulo (12,6), Paraná (19,0) e Piauí (21,5), todas essas taxas são para cada 100.000 habitantes negros. Também foi constatado que no Brasil a violência letal remete sobretudo à população jovem (15 a 29 anos), visto que em 2017 35.783 jovens foram assassinados. Esse número representa uma taxa de 69,9 homicídios a cada 100.000 habitantes jovens, sendo que no Rio de Janeiro esta taxa eleva-se para 92,6 [8].

Em relação à faixa etária jovem, há de se fazer algumas ressalvas. O Ministério da Saúde segue a convenção elaborada pela OMS que define o intervalo entre 10 e 19 anos, como adolescência, e o situado entre 15 e 24 anos como juventude [11]. O Estatuto da Juventude (Lei nº 12.852, de 5 de agosto de 2013) define, no entanto, juventudes diversas de acordo com a faixa etária. Dos 15 a 17 anos são os chamados adolescentes-jovens; dos 18 a 24 anos de jovens-jovens e entre os 25 a 29 anos são nomeados como jovens-adultos [12].

Ainda tendo como base informações do IPEA (2007 a 2017), o comportamento das mortes considerando a faixa etária e o sexo indicou que, no ano de 2017 no Brasil, 51,8% dos óbitos entre 15 a 19 anos e 49,4% do total de óbitos entre 20 a 24 anos foram causados por homicídio. Este cenário também é preocupante se observado o sexo. Entre as mulheres o total de óbitos por homicídio corresponde a 17,4% e 15,5%, respectivamente, nas faixas etárias de 15 a 19 anos e 20 a 24 anos, enquanto que entre os homens este total representa

59,1% e 55,7%, respectivamente, nessas mesmas faixas etárias [8].

No que se refere ao grau de instrução e faixa etária, a Pesquisa Nacional por Amostra de Domicílios (PNAD) do primeiro trimestre de 2019 registrou que 35% dos brasileiros em idade de trabalhar, acima de 14 anos, não concluíram o ensino fundamental. Os dados ainda apontam que 48% dessas pessoas não concluíram sequer o ensino médio [13]. Essa baixa escolaridade dos jovens afeta diretamente as chances dos mesmos conseguirem um emprego. A população jovem, cresce significativamente sem expectativa de um futuro promissor, com diminuição de oportunidades de emprego, com aumento da pobreza, com o envolvimento com o tráfico e abuso de álcool, o que pode refletir no aumento da violência [14]. Este cenário, que inclui a falha do sistema educacional, agregado à falta de emprego para esta importante parcela de pessoas da sociedade, promove desigualdades sociais, sendo um dos principais determinantes da elevada mortalidade relacionada à violência, cujos perpetradores e vítimas são principalmente os mesmos jovens [15].

O combate à violência em geral tem sido cada vez mais evidente. Em 2013 o IPEA sugeriu ações ao poder público com o objetivo de encarar o problema com resultados quase instantâneos, sendo eles: maior efetivo policial, mais prisões e menor evasão escolar. Não são as únicas maneiras de lidar com o problema, mas apresentam, segundo o IPEA, um plano para ajudar a reter o quadro de mortes em relação aos homicídios. O Instituto calcula que aumentar em 10% o efetivo policial (polícia militar e polícia civil) reduz a taxa de homicídios no ano seguinte entre 0,8% e 3,4%. Também foi estimado que aumentar em 10% o número de presos reduz a taxa de homicídios do ano seguinte em, aproximadamente, 0,5% e por fim um aumentar em 1% as taxas de evasão escolar pode contribuir com o aumento da taxa de homicídios em 0,1% [16].

No estado do Rio de Janeiro, de acordo com o Instituto de Segurança Pública (ISP), há um indicador estratégico de criminalidade chamado letalidade violenta, este indicador é composto por quatro títulos, sendo eles: homicídio doloso, morte por intervenção de agente do estado, latrocínio (roubo seguido de morte) e lesão corporal seguida de morte. Segundo o ISP, em 2017 no estado do Rio de Janeiro, foram registrados 6.749 casos de letalidade violenta, ou seja, 40 mortes a cada 100.000 habitantes e no título de homicídios dolosos foram contabilizados 5.322 casos, assim sendo 31,8 mortes a cada 100.000 habitantes. Dentre as localidades do estado citam-se as seguintes taxas de mortalidade por homicídio doloso em 2017: Capital, 22,7; Baixada Fluminense, 49,6; Grande Niterói, 32,9; e Interior do Rio de Janeiro, 29,8. Este comportamento é preocupante, pois a OMS considera epidêmicas taxas de homicídio superiores a 10 [17].



Análises descritivas, testes não paramétricos e análise de séries temporais são algumas ferramentas estatísticas que podem ajudar nas análises destes dados de violência.

## 1.2 Revisão Bibliográfica

Nesta seção são descritos de forma sucinta alguns dos principais trabalhos encontrados na literatura abordando o tema de interesse.

Neves *et. al* (2015) [18] investigaram a mortalidade de jovens brasileiros entre 15 e 29 anos, ocorridas no período de 2000 a 2012. Foram feitas análises descritivas e estudo de séries temporais com dados do Sistema de Informações sobre Mortalidade (SIM) e utilizou-se a regressão Prais-Winsten para análise de tendências nas taxas de mortalidade. Registraram-se 958.224 óbitos de jovens neste período, 79,6% do sexo masculino, as taxas de mortalidade foram de 1,6 em 2000 e de 1,5 em 2012 a cada 1000 jovens, constatando taxas de incremento anual com tendências estacionárias (-0,34%; IC95%: -1,05 ; 0,37). Constatou-se, em 2012, tendência crescente nessas taxas entre homens nas regiões Nordeste (3,08%; IC95%: 2,56 ; 3,61) e Sul (0,88%; IC95%: 0,09 ; 1,66). Nos homens, a frequência de óbitos foi maior entre aqueles com 20 a 24 anos de idade (39,0%), de cor da pele parda (52,1%) e escolaridade menor que oito anos de estudo (69,9%). Entre as mulheres, a frequência de óbitos foi maior naquelas com 25 a 29 anos de idade (39,4%), de cor da pele branca (47,2%) e escolaridade inferior a oito anos de estudo (57,8%). Para ambos os sexos, os óbitos foram mais frequentes aos sábados e domingos, com distribuição relativamente homogênea entre os meses do ano.

Provenza (2015) [19] investigou homicídios dolosos no período de 2001 até 2011. Com micro dados, que tem como fonte o ISP, registraram-se 65.760 homicídios dolosos. Foram analisados o espaço-tempo e o perfil das vítimas para no ano 2011. Em um segundo momento o estudo da série histórica teve a utilização de quatro modelos de alisamento na previsão de séries temporais: alisamento exponencial simples (AES), alisamento exponencial linear de brown (AELB), alisamento exponencial de holt (AEH) e alisamento exponencial sazonal de Holt-Winters (AESHW). Constatou-se que há maior incidência de homicídios entre homens negros/pardos entre 20 a 29 anos e com um predomínio de uso de armas de fogo. As regiões do estado mais afetadas foram a Baixada Fluminense e a Capital, com maior ocorrência em vias públicas e à noite. Para avaliar os modelos de séries temporais, foram usadas duas medidas de erro de previsão, o erro quadrado médio (EQM) e o erro percentual médio absoluto (MAPE). Segundo o EQM, a melhor

alternativa para estimação da série foi o método AES, enquanto que, segundo o MAPE revelou que o melhor resultado obtido foi o AESHW.

Melo *et. al* (2017) [20] analisaram a associação entre o risco de morte por agressões em jovens do sexo masculino e características sociodemográficas dos municípios brasileiros, tendo como unidades de análise municípios com mais de 20.000 habitantes no período de 2010 até 2014. Neste intervalo, foram registrados 127.137 óbitos por agressão a jovens de 15 a 29 anos. Modelos de regressão binomial negativa brutos e ajustado foram construídos para estimar razões de taxas de mortalidade (RTM) e seus respectivos intervalos de 95% de confiança. RTM mais elevadas no modelo ajustado foram observadas nos municípios mais urbanizados (1,95; IC95%: 1,70 ; 2,23), em categorias intermediárias de desigualdade de renda (1,10; IC95%: 1,01-1,20) e proporção de pobreza (1,69; IC95%: 1,51 ; 1,89), com menor proporção de jovens frequentando o Ensino Médio (2,05; IC95% 1,83 ; 2,30), maior proporção de jovens de 18-24 anos desocupados (1,27; IC95%: 1,16 ; 1,40) e maior número de mulheres em relação ao de homens (1,28; IC95% 1,05 ; 1,58). A mortalidade de homens jovens causada por agressão é alta, especialmente nas grandes cidades urbanizadas, com maior proporção de jovens buscando emprego e fora do ensino médio. Ele destaca a importância das políticas sociais para lidar com a violência entre jovens.

Provenza *et. al* (2018) [21] investigaram o homicídio doloso no estado no período de 2001 a 2016. Assim com o trabalho realizado no Rio de Janeiro já mencionado [19], foram apresentados o perfil das vítimas e a análise e previsão da série temporal através dos modelos: alisamento exponencial simples, alisamento exponencial linear de Brown, alisamento exponencial de Holt e alisamento exponencial sazonal de Holt-Winters. A eficácia dos modelos foi testada por duas medidas: o erro quadrático médio e o erro percentual médio absoluto. Para a elaboração da modelagem da série temporal dos homicídios dolosos no estado do foi utilizado o teste não paramétrico de Kurskal-Wallis. Foi observada uma gradual tendência de redução no número de ocorrências de 2001 até 2016, porém os números ainda estão muito altos comparados com outras regiões similares. Em 2016, a arma de fogo foi o instrumento mais utilizado e a Baixada Fluminense foi a região mais afetada. A maior parte das ocorrências ocorre à noite e nos fins de semana e os jovens negros são os mais afetados.

Motivado pelos resultados dos estudos aqui mencionados e a relevância do tema, surgiu o interesse de verificar o comportamento de mortes violentas em jovens residentes no estado do Rio de Janeiro. Compreender a dimensão da violência de maneira mais ampla e profunda por seu impacto na mortalidade é uma maneira de esclarecer a magnitude desse

problema e um passo necessário para a criação de estratégias de políticas públicas para prevenir e abordar suas causas e consequências.

## 1.3 Objetivos

Este trabalho tem como objetivo geral investigar o comportamento dos homicídios dolosos contra jovens (15-24 anos) ocorridos no estado do Rio de Janeiro no período de 2006 a 2019, sendo os objetivos específicos:

- (i) identificar regiões do estado, localidades e turnos do dia onde os homicídios predominaram;
- (ii) averiguar descritivamente o comportamento temporal da série mensal dos homicídios, identificando aspectos como tendência e sazonalidade;
- (iii) descrever o comportamento dos homicídios no período investigado;
- (iv) prever o número de casos de homicídios para os meses de janeiro a dezembro de 2020.

## 1.4 Organização do Trabalho

Este trabalho está organizado em 3 capítulos. No capítulo 2 são apresentados os materiais e métodos utilizados no estudo, no capítulo 3 encontram-se as análises dos resultados e, por fim, no capítulo 4 encontra-se a conclusão.

## 2 Materiais e Métodos

### 2.1 Base de dados

A base de dados para este trabalho foi obtida através do Instituto de Segurança Pública (ISP) [22]. Foram consideradas ocorrências diárias de mortes por violência no estado do Rio de Janeiro, de janeiro de 2006 a dezembro de 2019, restritas a vítimas adolescentes (de 15 a 24 anos) e os registros foram aglutinados de forma mensal. O Quadro 1 apresenta as variáveis referentes às ocorrências de mortes por violência, bem como informações da vítima, que serão consideradas no estudo, com suas categorias originais e as categorias que serão utilizadas neste TCC.

Quadro 1: Descrição das variáveis do banco de dados

Variáveis	Valores e/ou categorias	
	Originais	TCC
Ano	2006 a 2019	2006 a 2019
Mês	Janeiro a Dezembro	Janeiro a Dezembro
Título da ocorrência	Homicídio doloso	Homicídio doloso
Conteúdo	Vítimas e Ambos	Vítimas e Ambos
Data	Período entre 01/01/2006 e 31/12/2019	Período entre 01/01/2006 e 31/12/2019
Hora	Entre 00:00 e 23:59	0h † 6h, 6h † 12h, 12h † 18h e 18h † 0h
Local	Chácara, Condomínio, Conjunto habitacional, Favela, Fazenda, Residência, Sítio e Vila. Via pública. Aeroporto, Autódromo, Banco, Bar/Restaurante, Boate, Caixa eletrônico, Campo, Casa de massagem, Casa de show, Centro de convenções, Clube, Delegacia, Escritório, Estabelecimento público, Estabelecimentos diversos; Estação de barcas, metrô, ônibus e trem; Estádio, Floresta, Garagem, Hotel, Hipódromo, Hospital, Instituição religiosa, Interior de coletivos, Lagoa, Linha férrea, Mar, Morro, Motel, Parque, Parque de diversões, Posto, Praia, Rio, Semáforo, Shopping, Supermercado e Teatro. Ignorado, Outros e Sem informação	Espaço residencial, Via pública, Estabelecimentos e Ignorado/Outros
Município	Municípios existentes no RJ, segundo IBGE	Capital, Grande Niterói, Baixada Fluminense e Interior do Estado
Bairro	Bairros existentes no RJ, segundo IBGE	Bairros dos municípios das regiões Capital, Grande Niterói, Baixada Fluminense e Interior do Estado

Quadro 1 (Continuação)

Variáveis	Valores e/ou categorias	
	Originais	TCC
Sexo	Masculino e Feminino	Masculino e Feminino
Raça/Cor	Albino, Amarelo, Branco, Índio, Negro e Pardo	Branco, Preto, Pardo e Outros
Relação da vítima com o agressor	Filho, Irmão(ã), Mãe e Pai. Cônjuge, Cunhado(a), Enteado(a), Ex-cônjuge, Ex-Cunhado(a), Ex-Noivo, Genro, Neto(a), Nora, Padrasto, Primo(a), Sobrinho(a), Sogro(a) e Tio(a). Aluno(a), Amante, Amigo(a), Companheiro(a), Comparsa, Ex-companheiro(a), Ex-comparsa, Ex-namorado, Namorado(a) e Vizinho(a). Cliente, Colega de trabalho, Empregado(a) e Patrão(ao). Ignorado, Outra e Sem informação	Familiar de 1º grau, Familiar de 2º grau, Conhecido, Trabalho e Nenhuma
Idade	15 a 24	15 a 19 e 20 a 24

## 2.2 Séries Temporais

**Definição 2.1 (Processo Estocástico)** Um processo estocástico é coleção de variáveis aleatórias ordenadas no tempo e definidas em um conjunto de pontos  $T$ , que pode ser contínuo ou discreto onde  $Z = \{Z(t), t \in T\}$ , de forma que para cada  $t \in T$ ,  $Z(t)$  é uma variável aleatória [23].

**Definição 2.2 (Processo Estacionário)** Se o comportamento do processo tem as características de não mudar com o tempo, ou seja, o processo se desenvolve aleatoriamente com o tempo em torno do valor médio, então o processo é denominado de processo estacionário.

Há três tipos de estacionariedade:

- **Estrita:** a estrutura probabilística não se altera com o tempo;
- **Ampla:** os momentos,  $E[Z_t^{(i)}], i = 1, 2, \dots$ , não variam com o tempo;
- **Segunda ordem:** os momentos até segunda ordem não variam com o tempo (média e variância não variam) e autocovariância depende apenas da diferença entre os tempos.

**Definição 2.3 (Ruído Branco)** Um caso especial de processo estocástico é o ruído branco ( $a_t$ ). Ele é independente e identicamente distribuído (i.i.d.) com média 0 e variância constante e finita sendo denotado como  $a_t \sim RB(0, \sigma^2)$ .

**Definição 2.4 (Série Temporal)** *Uma série temporal é um conjunto de observações  $z_t, t \in T$  ordenadas no tempo e igualmente espaçadas. Neste tipo de dado as observações vizinhas são dependentes e há interesse em modelar esta dependência. É também definida como sendo uma das realizações do processo estocástico [23].*

As séries temporais podem ser classificadas em contínuas e discretas. As séries ditas discretas são observações em instantes pontuais, sendo  $T = \{t_1, \dots, t_n\}$  o conjunto dos tempos discretos até  $n$ , a série temporal será denotada por  $\{Z_t : t \in T\}$ .

Séries temporais podem ser decompostas em 4 elementos, são eles:

- **Tendência:** indica sentido de crescimento ou decrescimento;
- **Ciclo:** comportamento repetido ao longo do tempo;
- **Sazonalidade:** comportamento que se repete com igual intervalo de tempo;
- **Ruído ou erro aleatório:** tudo aquilo que não é explicado pelas outras componentes da série e são elementos que não podem ser medidos mas interferem no comportamento da série.

As séries estacionárias são diferentes das séries não estacionárias. Nas séries não estacionárias, há tendências para cima e para baixo, então a média e a variância são funções do tempo.

### 2.2.1 Autocorrelação

Durante a análise de séries temporais, há a necessidade de identificar elementos que possibilitem a obtenção do modelo adequado para descrever o comportamento e as características da série. Estes elementos são identificados a partir do coeficiente de autocorrelação  $\{r_k\}$  que, diferente do coeficiente de correlação usual, realiza correlação em uma única variável comparando em diferentes instantes de tempo [24].

Na prática, mais simples do que calcular diretamente o coeficiente de autocorrelação é obter inicialmente o coeficiente de autocovariância  $\{c_k\}$  para em seguida obter o coeficiente de autocorrelação. O coeficiente de autocovariância é definido por:

$$c_k = \sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t-k} - \bar{z})/n \quad (2.1)$$

Os coeficientes de autocorrelação são então obtidos como:

$$r_k = c_k/c_0 \quad (2.2)$$

onde  $k$  é a defasagem.

Note ainda que substituindo  $k$  por 0 em  $c_k$  tem-se:

$$c_0 = \sum_{t=1}^n (z_t - \bar{z})^2/n \quad (2.3)$$

Logo  $c_0$  é igual a variância.

### 2.2.2 Função de Autocorrelação

A Função de Autocorrelação (FAC) é uma ferramenta importante para acessar as propriedades de uma série temporal. Considerando uma série  $Z_t$ , o coeficiente de correlação entre  $Z_t$  e  $Z_{t-k}$  é chamado de autocorrelação de  $k$ -ésima ordem e é denotado por:

$$\rho_k = \frac{Cov(Z_t, Z_{t-k})}{Var(Z_t)} = \frac{\gamma_k}{\gamma_0} \quad (2.4)$$

**Proposição 2.1** *A FAC  $\rho_k$  satisfaz as seguintes propriedades:*

- (i)  $\rho_0 = 1$ ;
- (ii) *A correlação entre  $Z_t$  e  $Z_{t+k}$  é a mesma que entre  $Z_t$  e  $Z_{t-k}$ , ou seja  $\rho_k = \rho_{-k}$ ;*
- (iii)  $-1 \leq \rho_k \leq 1$

### 2.2.3 Função de Autocorrelação Parcial

A idéia de autocorrelação pode ser estendida. Se for medida a correlação entre duas observações seriais,  $Z_t$  e  $Z_{t+k}$ , eliminando a dependência dos termos intermediários,  $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$ , tem-se o que se denomina Função de Autocorrelação Parcial (FACP), representada por:

$$\phi_{kk} = Corr(Z_t, Z_{t-k} | Z_{t-1}, \dots, Z_{t-(k+1)}) \quad (2.5)$$

### 2.2.4 Correlogramas

O correlograma é um tipo de gráfico utilizado para entender se a série temporal possui características sazonais ou algum tipo de tendência e consiste em traçar as autocorrelações em diversas  $k$  defasagens [24].

A Figura 1 apresenta um exemplo de um correlograma da função de autocorrelação e da função de autocorrelação parcial, respectivamente:

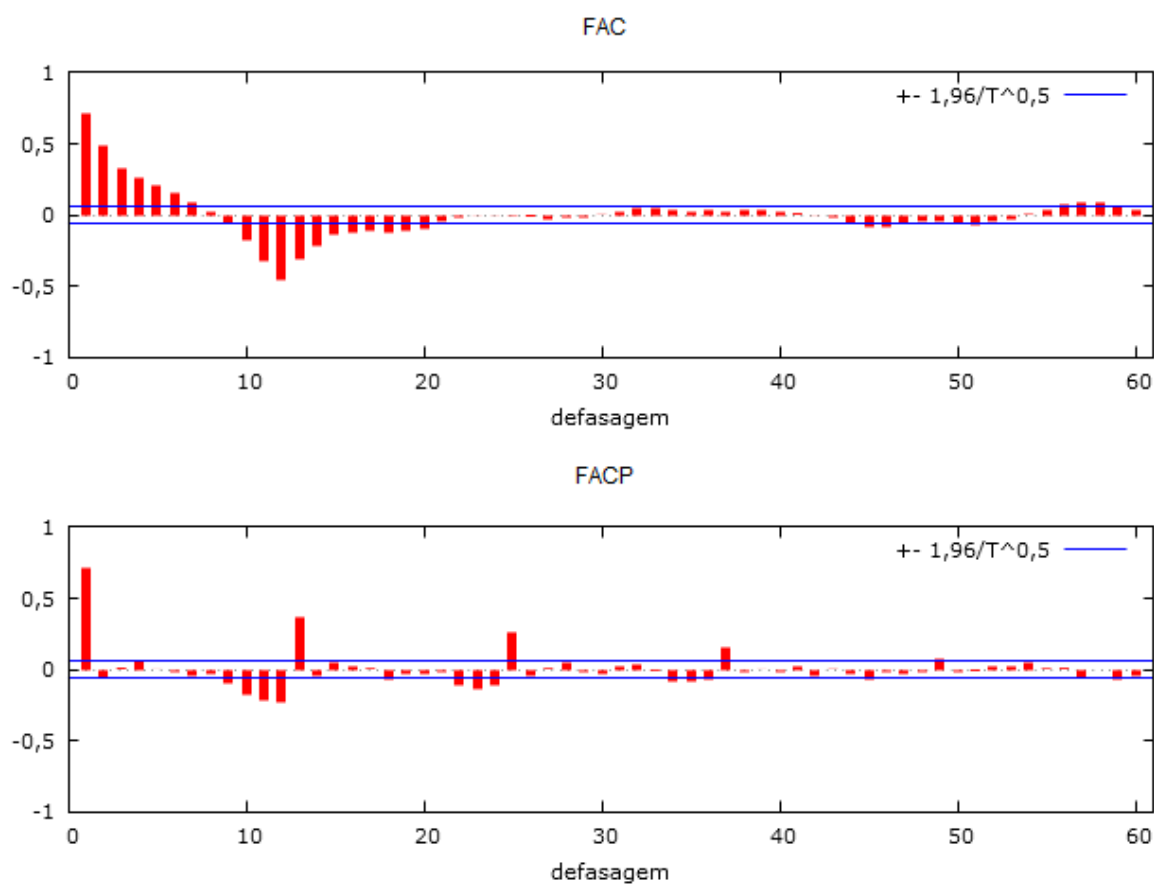


Figura 1: Exemplo de Correlograma da FAC

Fonte: elaboração própria.

Note que a diferença entre a FAC e a FACP mostra que na imagem de cima há uma autocorrelação significativa até a defasagem 20, enquanto que na imagem de baixo apenas 4 defasagens da série temporal realmente são importantes para modelar seu valor esperado em  $t$ .



## 2.3 Operadores

### 2.3.1 Operador de defasagem

Um operador de defasagem  $B$  é o operador que retorna o número de períodos associados a uma observação precedente.

$$B^k Z_t = Z_{t-k} \quad (2.6)$$

### 2.3.2 Operador de diferenças

Ao aplicar um operador de diferenças ( $\Delta$ ) a uma série temporal  $Z_t$ , isso resultará em uma nova série com  $t - 1$  observações. A forma geral deste operador é:

$$\Delta = 1 - B \quad (2.7)$$

$$\Delta Z_t = (1 - B)Z_t = Z_t - Z_{t-1} \quad (2.8)$$

Para se obter a estacionariedade em uma série não estacionária, uma possibilidade é realizar sucessivas diferenças até que a série se torne estacionária.

## 2.4 Teste de Significância Individual

Teste de hipóteses de significância individual dos parâmetros estimados servem para determinar se cada parâmetro é importante ou significativo para o modelo em questão.

As hipóteses para testar a significância são dadas por:

$$\begin{cases} H_0 : \beta_k = 0 \\ H_1 : \beta_k \neq 0 \end{cases} ; k = 0, 1, \dots \quad (2.9)$$

E a estatística de teste como sendo:

$$t_{\text{obs}} = \frac{\hat{\beta}_{\mathbf{k}}}{\sqrt{\text{V\hat{A}R}(\hat{\beta}_{\mathbf{k}})}} \quad (2.10)$$

Se  $H_0$  não é rejeitada, então pode-se retirar o respectivo parâmetro do modelo já que este não o influencia de forma significativa.

## 2.5 Testes de avaliação do comportamento da série temporal

Antes de começar a modelagem da série temporal, serão necessários testes estatísticos para analisar o comportamento antes da modelagem das mesmas. Os pressupostos para a modelagem ARIMA são de estacionariedade e normalidade dos dados, para isso, serão feitos testes de raiz unitária e normalidade [25].

### 2.5.1 Teste de Raiz Unitária

Um dos pressupostos para a modelagem ARIMA é o de estacionariedade, como visto na Definição 2.2. Antes de tudo, uma série temporal estacionária de segunda ordem, deve manter ao longo do tempo suas características estatísticas, que são:

- $\mu_t$  constante
- $\sigma_t^2$  constante

Para a realização do teste de raiz unitária é necessário definir antes os seguintes tópicos:

**Definição 2.5 (Polinômio Autorregressivo)** O polinômio autorregressivo de ordem  $p$  é dado por:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.11)$$

**Definição 2.6 (Polinômio Médias Móveis)** O polinômio médias móveis de ordem  $q$  é dado por:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.12)$$

Seja  $\Psi(B) = \frac{\theta(B)}{\phi(B)}$  e dado um modelo linear geral  $\phi(B)Z_t = \theta(B)a_t$ , pode-se dizer que  $Z_t$  é estacionário se as raízes de  $\Psi(B) = 0$  estão dentro de um círculo unitário, caso

este processo tenha uma raiz unitária e se 1 é a raiz da equação, é dito que tal processo é não-estacionário.

Pode-se detectar estacionariedade através do teste de raiz unitária, verifica se o polinômio autorregressivo ou médias móveis utilizado na análise possui raiz sobre o círculo unitário.

Será utilizado o teste de *Dickey-Fuller Aumentado* que é conhecido como teste *ADF* (*Augmented Dickey-Fuller*) e estuda a seguinte regressão [26]:

$$\Delta \mathbf{y}_t = \beta_1 + \beta_2 \mathbf{t} + \delta \mathbf{y}_{t-1} + \sum_{i=1}^m \mathbf{a}_i \Delta \mathbf{y}_{t-1} + \varepsilon_t \quad (2.13)$$

onde:

- $\beta_1$ : é o intercepto;
- $\beta_2$ : é o coeficiente de tendência;
- $\delta$ : é o coeficiente de presença de raiz unitária;
- $m$ : é o número de defasagens tomadas na série;
- $\varepsilon_t$ : é o erro.

Hipóteses do teste:

$$\begin{cases} H_0 : \text{existe raiz unitária (não é estacionário)} \\ H_1 : \text{não existe raiz unitária (é estacionário)} \end{cases} \quad (2.14)$$

E a estatística do teste como sendo:

$$\mathbf{T} = \frac{\hat{\delta}}{\text{se}(\hat{\delta})} \quad (2.15)$$

onde:

- $\hat{\delta}$ : é o estimador de  $\delta$ , já definido anteriormente;
- $\text{se}(\hat{\delta})$ : é o estimador para o desvio padrão do erro de  $\delta$ .

Dickey e Fuller tabelaram os valores críticos da estatística  $\mathbf{T}$  através da simulação de Monte Carlo e tais valores variam nos casos de presença somente de intercepto, presença somente de tendência e presença de ambos.

## 2.5.2 Testes de Normalidade

Outra pressuposição para a modelagem de Box & Jenkins é a de que os dados sigam distribuição normal [25].

**Definição 2.7 (Processo Normal)** *Um processo seria normal se a cada instante de tempo, sua distribuição é normal e, conseqüentemente, uma série seria normal se seus dados seguissem uma distribuição normal de probabilidade.*

Para a verificação de Normalidade da série, serão considerados os testes de normalidade. Neste TCC seram considerados os teste de Shapiro-Wilk, Jarque-Beta e Lilliefors.

As hipóteses nula e alternativa para os testes citados são:

Hipóteses do teste:

$$\begin{cases} H_0 : \text{os dados seguem distribuição normal, } N(\mu, \sigma^2) \\ H_1 : \text{os dados não seguem distribuição normal} \end{cases} \quad (2.16)$$

### 2.5.2.1 Teste de Shapiro-Wilk

Proposto por Shapiro & Wilk (1965) [27] e utiliza a seguinte estatística de teste abaixo.

$$\mathbf{W} = \frac{\mathbf{b}^2}{\sum_{i=1}^n (\mathbf{x}_{(i)} - \bar{\mathbf{x}})^2} \quad (2.17)$$

onde:

- $\mathbf{x}_{(i)}$ : são os valores da amostra ordenados;
- $\mathbf{b}$ : é uma constante determinada da seguinte forma:

$$b = \begin{cases} \sum_{i=1}^{\frac{n}{2}} a_{n-i+1} \times (x_{(n-i+1)} - x_{(i)}), & \text{se } n \text{ é par} \\ \sum_{i=1}^{\frac{n+1}{2}} a_{n-i+1} \times (x_{(n-i+1)} - x_{(i)}), & \text{se } n \text{ é ímpar} \end{cases} \quad (2.18)$$

Em que  $a_{n-i+1}$  são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho  $n$  de uma distribuição Normal e esses valores podem ser encontrados de forma tabelada [28].

### 2.5.2.2 Teste de Jarque-Bera

Proposto por Bera & Jarque (1980) [29] e utiliza a seguinte estatística de teste abaixo.

$$\mathbf{JB} = \mathbf{n} \left( \frac{\alpha_3^2}{6} + \frac{(\alpha_4 - 3)^2}{24} \right), \quad (2.19)$$

tal que:

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}.$$

$$\alpha_3 = \frac{\sum_{i=1}^n (y_i - \bar{y})^3}{ns^3},$$

$$\alpha_4 = \frac{\sum_{i=1}^n (y_i - \bar{y})^4}{ns^4},$$

### 2.5.2.3 Teste Lilliefors

O teste Lilliefors é uma modificação do teste de Kolmogorov-Smirnov e utiliza a seguinte estatística de teste abaixo [30].

$$\mathbf{D} = \max \{ \mathbf{D}^+, \mathbf{D}^- \}, \quad (2.20)$$

onde  $D^+ = \max_{i=1, \dots, n} \left\{ \frac{i}{n} - p_{(i)} \right\}$ ,  $D^- = \max_{i=1, \dots, n} \left\{ p_{(i)} - \frac{i-1}{n} \right\}$  e  $p_{(i)} = \Phi([x_{(i)} - \bar{x}]/s)$

de forma que:

- $\bar{x}$ : média amostral;
- $s$ : desvio padrão amostral;
- $\Phi$ : função de distribuição acumulada normal padrão

## 2.6 Modelos de Previsão

### 2.6.1 Modelos de Holt-Winters

Wheelwright *et. al* (1998) [31] afirmam que o modelo de Holt Winters é entendido como um método de suavização exponencial incluindo uma equação adicional utilizada para ajustar a previsão de sazonalidade. Em 1957, Holt expandiu o modelo de suavização exponencial simples, para lidar com dados que apresentavam tendência linear. Em 1960, Winters estendeu o modelo de Holt, incluindo uma nova equação para lidar com a previsão do comportamento da componente sazonal dos dados, gerando assim o método de Holt Winters.

De acordo com Morettin e Tolo (2004) [23] há dois tipos de procedimentos cuja aplicação depende das características da série: aditivo e multiplicativo. O modelo aditivo é usado quando a amplitude da variação sazonal mantém-se constante e o modelo multiplicativo é utilizado quando a amplitude da variação sazonal aumenta com o tempo. Esses procedimentos são baseados em três equações com constantes de suavização diferentes e que são associadas a cada uma das componentes do padrão da série: nível, tendência e sazonalidade. Como é mostrado na Tabela 1 onde na equação de previsão do modelo aditivo, a componente sazonal é adicionada ao modelo e no modelo multiplicativo, a componente sazonal é multiplicada ao modelo.

Tabela 1: Equações do Modelo de Holt Winters	
Modelo Aditivo	
Nível	$L_t = \alpha(Z_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1})$
Tendência	$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$
Sazonalidade	$S_t = \gamma(Z_t - L_t) + (1 - \gamma)S_{t-s}$
Previsão	$Z_{t+h} = (L_t + T_t h) + S_{t-s+h}$
Modelo Multiplicativo	
Nível	$L_t = \alpha \left( \frac{Z_t}{S_{t-s}} \right) + (1 - \alpha)(L_{t-1} + T_{t-1})$
Tendência	$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$
Sazonalidade	$S_t = \gamma \left( \frac{Z_t}{L_t} \right) + (1 - \gamma)S_{t-s}$
Previsão	$Z_{t+h} = (L_t + T_t h)S_{t-s+h}$

onde:

- **S**: comprimento da sazonalidade;

- $L_t$ : nível da série;
- $T_t$ : tendência da série;
- $S_t$ : componente sazonal;
- $Z_{t+h}$ : previsão  $h$  passos-à-frente;
- $Z_t$ : valor observado;
- $\alpha, \beta, \gamma$ : constantes de amortecimento de nível, tendência e sazonalidade, respectivamente.

As previsões geradas a partir do método linear de Holt exibem uma tendência constante indefinidamente no futuro. Há evidências que indiquem que esses métodos tendem a superestimar, especialmente para horizontes de previsão mais longos. Motivados por isso, Gardner & McKenzie (1985) [32] introduziram um parâmetro que “amortece” a tendência para uma linha plana em algum momento no futuro, chamado de *Damped Trend*. A Tabela 2 mostra as equações para este método.

Tabela 2: Equações do Modelo de Holt Winters com *Damped Trend*

Modelo Aditivo com <i>Damped Trend</i>	
Nível	$L_t = \alpha(Z_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + \phi T_{t-1})$
Tendência	$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)\phi T_{t-1}$
Sazonalidade	$S_t = \gamma(Z_t - L_t) + (1 - \gamma)S_{t-s}$
Previsão	$Z_{t+h} = (L_t + T_t h [\sum_{i=1}^h \phi^i]) + S_{t-s+h}$
Modelo Multiplicativo com <i>Damped Trend</i>	
Nível	$L_t = \alpha \left( \frac{Z_t}{S_{t-s}} \right) + (1 - \alpha)(L_{t-1} + \phi T_{t-1})$
Tendência	$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)\phi T_{t-1}$
Sazonalidade	$S_t = \gamma \left( \frac{Z_t}{L_t} \right) + (1 - \gamma)S_{t-s}$
Previsão	$Z_{t+h} = (L_t + T_t h [\sum_{i=1}^h \phi^i]) S_{t-s+h}$

onde:

- $\phi$ : parâmetro *Damped Trend*, tal que  $\phi \in (0, 1)$ .

### 2.6.2 Modelos de Box & Jenkins

A metodologia de Box & Jenkins para a previsão se baseia no ajuste de modelos tentativos denominados *ARIMA* (ou *SARIMA*, caso a série seja influenciada por fatores

sazonais) a séries temporais de valores observados de forma que a diferença entre os valores gerados pelos modelos e os valores observados resulte em séries de resíduos de comportamento aleatório em torno de zero.

As estruturas de Box & Jenkins são identificadas através do comportamento teórico das funções de autocorrelação e da função de autocorrelação parcial. Os critérios de estacionariedade e normalidade devem ser satisfeitos. Caso o processo não seja estacionário, realiza-se sucessivas diferenças até torná-lo estacionário.

Existem vários critérios para identificação de um modelo, por isso, é possível identificar modelos diferentes dependendo do critério que foi escolhido para identificação.

### 2.6.2.1 Processos Autorregressivos

Suponha que  $\mathbf{a}_t$  seja um processo com média zero e variância  $\sigma_t^2$ . Um processo é chamado de autorregressivo de ordem  $p$ , ou AR( $p$ ), se:

$$\mathbf{Z}_t = \phi_1 \mathbf{Z}_{t-1} + \dots + \phi_p \mathbf{Z}_{t-p} + \mathbf{a}_t \quad (2.21)$$

onde:

- $\phi_1, \dots, \phi_p$ : são parâmetros que assumem valores no conjunto dos números reais;
- $\mathbf{a}_t$ : é o ruído branco.

### 2.6.2.2 Processos de Médias Móveis

Suponha que  $\mathbf{a}_t$  seja um processo com média zero e variância  $\sigma_t^2$ . Um processo é chamado de médias móveis de ordem  $q$ , ou MA( $q$ ), se:

$$\mathbf{Z}_t = \mathbf{a}_t - \theta_1 \mathbf{a}_{t-1} - \dots - \theta_q \mathbf{a}_{t-q} \quad (2.22)$$

onde:

- $\theta_1, \dots, \theta_q$ : são parâmetros reais;
- $\mathbf{a}_t$ : é o ruído branco.



### 2.6.2.3 Modelos Autorregressivos e de Médias Móveis

Modelos autorregressivos e de médias móveis é a junção dos modelos AR e MA. Denota-se por ARMA(p,q) um processo autorregressivo e de médias móveis de ordem (p,q) e pode ser representado por:

$$\mathbf{Z}_t = \phi_1 \mathbf{Z}_{t-1} + \dots + \phi_p \mathbf{Z}_{t-p} + \mathbf{a}_t - \theta_1 \mathbf{a}_{t-1} - \dots - \theta_q \mathbf{a}_{t-q} \quad (2.23)$$

onde:

- $\phi_1, \dots, \phi_p$ : são parâmetros da estrutura autorregressiva;
- $\theta_1, \dots, \theta_q$ : são parâmetros da estrutura médias móveis;
- $\mathbf{a}_t$  : é o ruído branco.

Os modelos *AR*, *MA* e *ARMA* são apropriados para descrever séries temporais estacionárias, isto é, séries que se desenvolvem no tempo ao redor de uma média constante. Para a escolha de  $p$  e  $q$ , a Tabela 3 apresenta alguns comportamentos esperados da FAC e FACP de alguns modelos da classe ARMA mais comuns [33].

Tabela 3: Comportamento teórico da FAC e FACP para alguns modelos

Modelo	FAC	FACP
MA(1)	1 pico na defasagem 1 ( $\rho_1 \neq 0$ )	Decaimento exponencial
AR(1)	Decaimento exponencial	1 pico na defasagem 1 ( $\phi_{11} \neq 0$ )
MA(2)	1 pico na defasagem 1 e 1 pico na defasagem 2 ( $\rho_1 \neq 0$ e $\rho_2 \neq 0$ )	Mistura de exponenciais ou ondas senóides amortecidas
AR(2)	Mistura de exponenciais ou ondas senóides amortecidas	1 pico na defasagem 1 e um pico na defasagem 2 ( $\phi_{11} \neq 0$ e $\phi_{22} \neq 0$ )
ARMA(1,1)	Decaimento exponencial	Decaimento exponencial

### 2.6.2.4 Modelos ARIMA

São modelos para descrever séries temporais não estacionárias. Para tornar a série estacionária é necessário fazer sucessivas ( $d$ ) diferenças na série até obter a estacionariedade de segunda ordem.

A equação do modelo ARIMA é representada da seguinte forma:

$$\text{ARIMA}(p, d, q) : \phi(\mathbf{B}) \Delta^d \mathbf{Z}_t = \theta(\mathbf{B}) \mathbf{a}_t \quad (2.24)$$

onde:

- $\phi(\mathbf{B})$ : é o polinômio auto regressivo (AR) de ordem  $\mathbf{p}$ ;
- $\theta(\mathbf{B})$ : é o polinômio de médias móveis (MA) de ordem  $\mathbf{q}$ ;
- $\mathbf{d}$ : número de diferenças necessárias para tornar a série estacionária;
- $\Delta^{\mathbf{d}}$ : é o operador de diferença de ordem  $\mathbf{d}$ ;
- $\mathbf{a}_t$  : é o ruído branco.

### 2.6.2.5 Modelos SARIMA

A equação do modelo ARIMA Sazonal (SARIMA) é representada da seguinte forma:

$$\text{SARIMA}(\mathbf{p}, \mathbf{d}, \mathbf{q}) \times (\mathbf{P}, \mathbf{D}, \mathbf{Q})_S : \phi(\mathbf{B})\Phi(\mathbf{B})\Delta^{\mathbf{d}}\Delta_S^{\mathbf{D}}\mathbf{Z}_t = \theta(\mathbf{B})\Theta(\mathbf{B})\mathbf{a}_t \quad (2.25)$$

onde:

- $\phi(\mathbf{B})$ : é o polinômio auto regressivo (AR) de ordem  $\mathbf{p}$ ;
- $\Phi(\mathbf{B})$ : o polinômio AR-sazonal de ordem  $\mathbf{P}$ ;
- $\theta(\mathbf{B})$ : é o polinômio de médias móveis (MA) de ordem  $\mathbf{q}$ ;
- $\Theta(\mathbf{B})$ : é o polinômio MA-sazonal de ordem  $\mathbf{Q}$ ;
- $\mathbf{d}$ : número de diferenças necessárias para tornar a série estacionária;
- $\mathbf{D}$ : número de diferenças sazonais necessárias para tornar a série estacionária;
- $\Delta^{\mathbf{d}}$ : é o operador de diferença de ordem  $\mathbf{d}$ ;
- $\Delta_S^{\mathbf{D}}$ : é o operador de diferença sazonal de ordem  $\mathbf{D}$ ;
- $\mathbf{a}_t$  : é o ruído branco.

## 2.7 Análise de Resíduos

Considerando um modelo linear do tipo:

$$\mathbf{Z}_t = \mu(\mathbf{t}) + \varepsilon_t, \quad \varepsilon_t \sim \mathbf{N}(\mathbf{0}, \sigma_{\varepsilon_t}^2) \quad (2.26)$$

sendo  $\varepsilon_t$  o ruído branco.

A equação do modelo ajustado é:

$$\hat{\mathbf{Z}}_t = \hat{\mu}(\mathbf{t}), \quad \text{ignorando } \hat{\varepsilon}_t \quad (2.27)$$

Espera-se que  $\hat{\varepsilon}_t$ , possua as características do ruído branco, que seria, possuir média zero, variância constante e os erros serem independentes e também seguir uma distribuição normal [23].

Para testar normalidade, serão usados os testes já mencionados em 2.5.2 e para a análise dos resíduos será feita a análise do correlograma dos resíduos, os mesmos devem ser não correlacionados, então basta que todas as defasagens apresentem autocorrelação não significativa, dentro do seguinte intervalo:

$$\rho_k \sim N(0, 1), \text{ logo : } \left( 0 \pm 1,96 \cdot \frac{1}{\sqrt{n}} \right) \quad (2.28)$$

Sendo  $n$  o número de observações da série.

A Figura 2 apresenta um exemplo de um correlograma dos resíduos. As linhas pontilhadas representam o intervalo citado em (2.28).

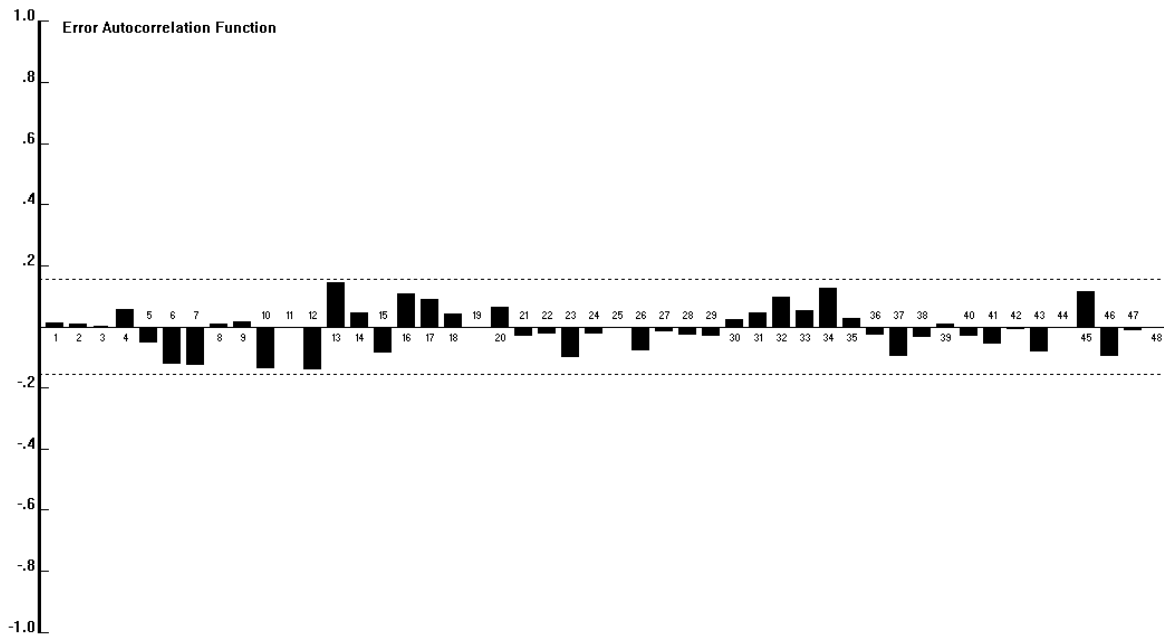


Figura 2: Exemplo de Correlograma da FAC dos resíduos não correlacionados.

Fonte: elaboração própria.

## 2.8 Estatísticas de Aderência

As estatísticas de aderência são usadas para avaliar o poder preditivo dos modelos. Espera-se que o modelo escolhido minimize as estatísticas de aderência. Serão usadas neste trabalho, as estatísticas de aderência a seguir:

### 2.8.1 MAD (*Mean Absolute Deviation*) ou Desvio Médio Absoluto

Esta estatística representa a diferença absoluta média entre os valores ajustados e os valores reais, considerando assim a média dos desvios absolutos.

$$\text{MAD} = \frac{\sum_{t=1}^n |Z_t - \hat{Z}_t|}{n} \quad (2.29)$$

onde:

- $Z_t$ : é o valor observado;
- $\hat{Z}_t$ : é o valor ajustado;

- $n$ : é o tamanho da série.

### 2.8.2 MSE (*Mean Square Error*) ou Erro Quadrático Médio

O MSE avalia a diferença entre o valor real e o valor ajustado de maneira quadrática, desta forma serão obtidos valores positivos. Quanto menor for o MSE, melhor será a qualidade do ajuste.

$$\text{MSE} = \frac{\sum_{t=1}^n (\mathbf{Z}_t - \hat{\mathbf{Z}}_t)^2}{n} \quad (2.30)$$

### 2.8.3 RMSE (*Root Mean Square Error*) ou Raiz Quadrada do Erro Quadrado Médio

É uma medida de magnitude que, assim como o MSE, possui apenas valores positivos e melhor será a qualidade do ajuste quando RMSE será próximo a zero.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^n (\mathbf{Z}_t - \hat{\mathbf{Z}}_t)^2}{n}} \quad (2.31)$$

### 2.8.4 MAPE (*Mean Absolute Percentage Error*) ou Erro Médio Percentual Absoluto

Apresenta o quanto se erra, em média, a nível de agregação de cálculo, sem compensar erros negativos com erros positivos. Seus valores variam de 0 a 100%, quanto menor o resultado melhor.

$$\text{MAPE} = \sum_{t=1}^n \left| \frac{\mathbf{Z}_t - \hat{\mathbf{Z}}_t}{\mathbf{Z}_t} \right| \times \frac{100\%}{n} \quad (2.32)$$

### 2.8.5 R<sup>2</sup> - máximo

O  $R^2$  representa o poder de explicação do modelo para a variação da demanda. Quanto maior o valor desta estatística, melhor.

$$R^2 = 1 - \frac{\sum_{t=1}^n (\mathbf{Z}_t - \hat{\mathbf{Z}}_t)^2}{\sum_{t=1}^n (\mathbf{Z}_t - \bar{\mathbf{Z}})^2} \quad (2.33)$$

onde:

- $\bar{Z}_t$ : é o valor médio.

### 2.8.6 AIC

O critério de informação de Akaike (AIC), proposto por Akaike (1974) [34], mostrou que o viés é dado assintoticamente por  $p$ , em que  $p$  é o número de parâmetros a serem estimados no modelo, e definiu seu critério de informação como:

$$\mathbf{AIC} = -2\log(\mathbf{L}_p) + 2[(\mathbf{p} + 1) + 1] \quad (2.34)$$

onde:

- $\mathbf{L}_p$ : é a função de máxima verossimilhança do modelo;
- $\mathbf{p}$ : é o número de variáveis explicativas consideradas no modelo;

### 2.8.7 BIC

Proposto por Schwarz (1978) [35], o critério de informação bayesiano (BIC), penaliza modelos com muitas variáveis, sendo assim, se utiliza o Princípio da Parcimônia.

$$\mathbf{BIC} = -2\log(\mathbf{L}_p) + [(\mathbf{p} + 1) + 1] \log(\mathbf{n}) \quad (2.35)$$

onde:

- $\mathbf{L}_p$ : é a função de máxima verossimilhança do modelo;
- $\mathbf{p}$ : é o número de variáveis explicativas consideradas no modelo;
- $\mathbf{n}$ : é o tamanho da série.

### 2.8.8 HQC

O critério de informação de Hannan-Quinn (HQC) é um critério para a seleção de um modelo. É uma alternativa ao AIC e ao BIC.

$$\text{HQC} = n \log \left( \frac{\text{RSS}}{n} \right) + 2k \log(\log(n)) \quad (2.36)$$

onde:

- **k**: é o número de parâmetros do modelo;
- **RSS**: é a soma residual de quadrados que resulta de uma regressão linear ou de outro modelo estatístico;
- **n**: é o tamanho da série.

Os critérios *AIC*, *BIC* e *HQC* penalizam modelos com grande número de parâmetros e quanto menor o valor da estatística, melhor.

## 2.9 Metodologia

Visando alcançar os objetivos estabelecidos neste projeto, inicialmente serão realizadas análises descritiva dos casos para que se possa averiguar o perfil dos jovens e das ocorrências. Em segundo momento, a série original será modelada via métodos de amortecimento exponencial e modelos de Box & Jenkins.

Posteriormente aos ajustes dos modelos, serão calculadas estatísticas de aderência para avaliar a capacidade preditiva dos modelos. O melhor modelo a ser selecionado será proveniente de qual minimizar as estatísticas de aderência e será considerado o mais apropriado para gerar previsões desta série.

O nível de significância ( $\alpha$ ) adotado em todos os teste será de 0,05 e os *softwares* que serão utilizados nas análises e etapas são:

- **Excel, Power BI e R**: para manipular os dados e realizar a análise descritiva;
- **FPW**: para o ajuste dos modelos de Holt Winters e Box & Jenkins, cálculo das estatísticas de aderência e análise dos resíduos;

- **Gretl:** para testes de raiz unitária, testes de normalidade, análise dos correlogramas, modelagens Box & Jenkins e análise dos resíduos



## 3 Análise dos Resultados

### 3.1 Características dos homicídios

Os registros avaliados se referem a 16.905 ocorrências mensais de homicídios dolosos entre jovens de 15 a 24 anos no estado do Rio de Janeiro, coletadas no período de janeiro de 2006 a dezembro de 2019 e aglutinados de forma mensal. Os dados são do Instituto de Segurança Pública. A Figura 3 retrata o comportamento dessa série, indicando picos de ocorrência no início dos anos e após o mês de julho em alguns anos, o que pode indicar que em períodos de festas de rua e férias escolares há mais crimes.

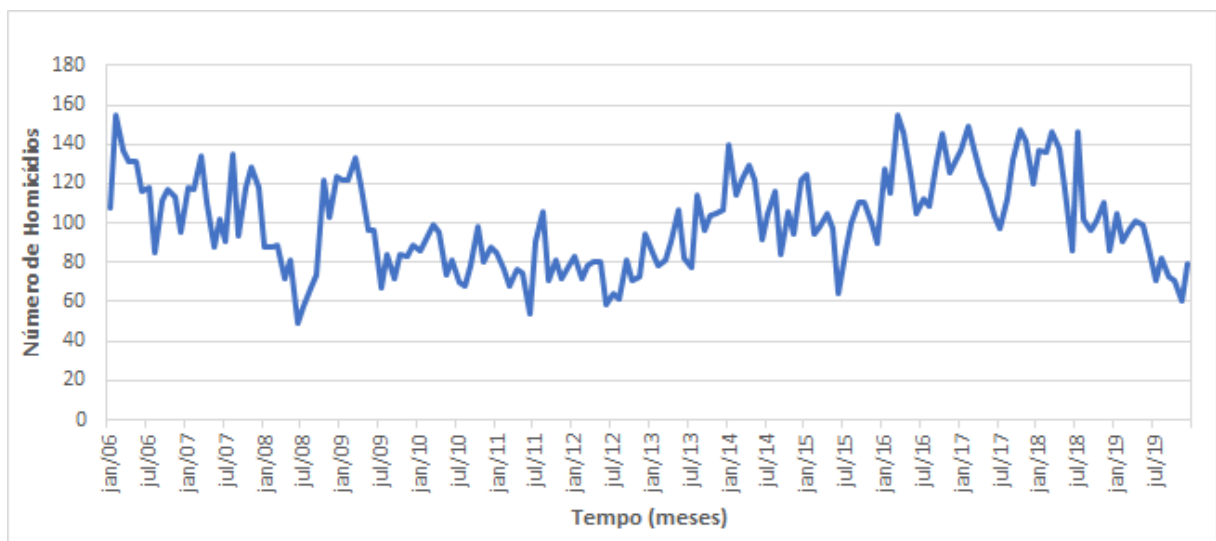


Figura 3: Homicídios dolosos por mês no estado do Rio de Janeiro

A Tabela 4 apresenta algumas estatísticas descritivas sobre o número de ocorrências das violências, identifica-se que ocorrem no estado do Rio de Janeiro em média 100,62 homicídios dolosos por mês. Em Junho de 2008 foi registrado a quantidade mínima de ocorrências (49) ao longo de todo período investigado, e em fevereiro de 2006 e em março de 2016 o número máximo de casos (155).

Tabela 4: Estatísticas Descritivas

<b>Estatísticas</b>	<b>Valor</b>
Média	100,62
Mediana	98,00
Mínimo	49,00
Máximo	155,00
Desvio Padrão	23,77
Percentil de 5%	65,35
Percentil de 95%	143,65
Coefficiente de Variação (CV) <sup>1</sup>	0,24

Finalmente a Tabela 5 apresenta os perfis das ocorrências, das vítimas e dos agressores. Nota-se que a identidade do agressor é na maioria (67,65%) desconhecida ou que não há nenhuma relação (29,73%) com a vítima do homicídio e levando-se em conta grande parte dos crimes ocorrerem em vias públicas (70,34%), pode-se levantar suspeitas de que os crimes se dão de maneira mais violenta e por motivos fúteis ou por impulso, sem um motivo ou planejamento do ato.

Tabela 5: Distribuição de frequências absolutas e relativas ao total de óbitos por homicídio de características referentes às ocorrências, à vítima e ao agressor

<b>Características</b>	<b>Categorias</b>	<b>Frequência Absoluta</b>	<b>Frequência Relativa</b>
Ocorrências			
<b>Hora</b>	0h † 6h	4.230	25,02 %
	6h † 12h	3.402	20,12 %
	12h † 18h	3.300	19,52 %
	18h † 0h	5.973	35,34 %
<b>Local</b>	Espaço residencial	2.021	11,95 %
	Estabelecimentos	1.492	8,83 %
	Via pública	11.891	70,34 %
	Ignorado/Outros	1.501	8,88 %
<b>Município</b>	Capital	4.765	28,19 %
	Grande Niterói	1.950	11,53 %
	Baixada Fluminense	5.352	31,66 %
	Interior do Estado	4.838	28,62 %
Vítimas			
<b>Sexo</b>	Feminino	1.003	5,94 %
	Masculino	15.734	93,07 %
	Sem informação	168	0,99 %
<b>Raça/Cor</b>	Branco	3.564	21,08 %
	Preto	4.287	25,36 %
	Pardo	8.356	49,43 %
	Outros/Ignorado	698	4,13 %

<sup>1</sup>Medida Adimensional

Tabela 5 (Continuação)

Características	Categorias	Frequência Absoluta	Frequência Relativa
<b>Idade</b>	15 a 19	7.075	41,85 %
	20 a 24	9.830	58,15 %
Agressor			
<b>Relação</b>	Familiar 1º grau	52	0,31 %
	Familiar 2º grau	71	0,42 %
	Conhecido	313	1,85 %
	Trabalho	11	0,06 %
	Nenhuma	5.026	29,73 %
	Desconhecido	11.432	67,63 %

Na Tabela 5, percebe-se também um padrão que se assemelha ao que já foi relatado na literatura quanto ao sexo e a raça das vítimas [18], em sua maioria, as vítimas são homens e pessoas pretas e pardas. Também constata-se que os crimes ocorrem com mais frequência, de forma absoluta, à noite e de madrugada. A região com maior quantidade de ocorrências, em termos absolutos, foi a Baixada Fluminense, o que não significa dizer que é o local onde ocorre a maior taxa de homicídios no estado, para isto cabe uma análise com os dados da população residente de cada região.

Também foi investigado a partir dos dados obtidos, os dias da semana em que mais houve ocorrências em todo o estado. A Figura 4 exibe estes resultados. Verifica-se, de forma absoluta, que de segunda-feira a quinta-feira os índices se comportam de forma parecida, porém ao chegar perto do fim de semana, começando por sexta-feira, sobe de forma considerável no sábado e domingo, onde as ocorrências atingem seu pico e permanecem semelhantes nestes dias da semana.

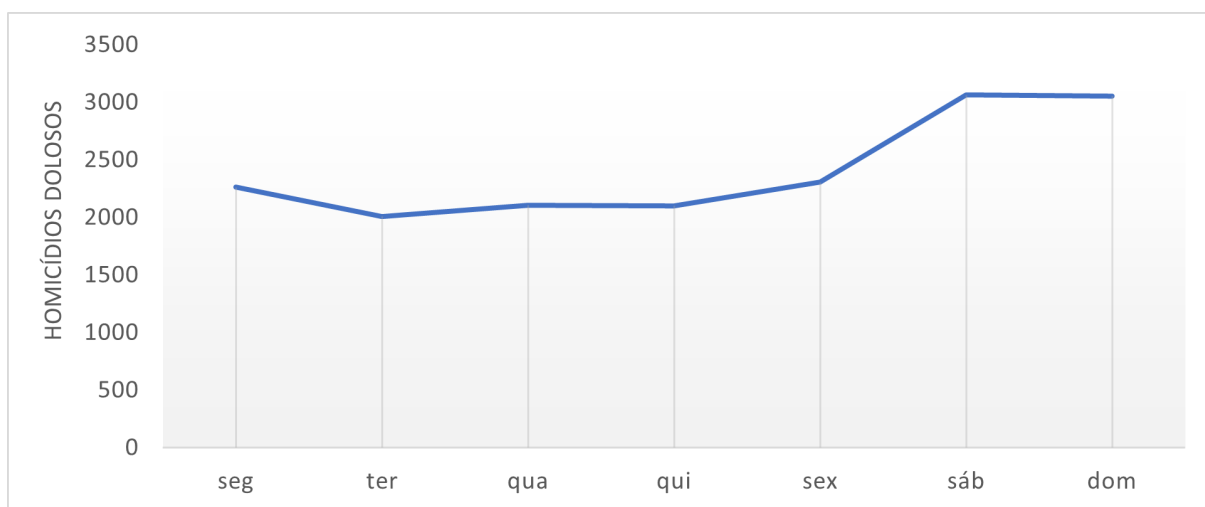


Figura 4: Homicídios dolosos por dia da semana no estado do Rio de Janeiro

## 3.2 Modelos de Holt-Winters

Primeiramente para a modelagem da série foram ajustados modelos de Holt-Winters. Foram considerados modelos com ou sem tendência linear e com *damped trend*, estes com ou sem sazonalidade, que pode ser aditiva ou multiplicativa. Na Tabela 6 são apresentadas as estatísticas de aderência de cada modelo de Holt-Winters.

Tabela 6: Estatísticas de Aderência para os Modelos de Holt-Winters

Modelos	$R^2$	BIC	MAPE	RMSE	MAD
Constante e sem sazonalidade	0,5084	16,87	0,1318	16,62	12,73
Constante e com sazonalidade aditiva	0,5817	15,80	0,1214	15,33	11,67
Constante e com sazonalidade multiplicativa	<b>0,5868</b>	<b>15,71</b>	<b>0,1192</b>	<b>15,24</b>	<b>11,46</b>
Linear e sem sazonalidade	0,5084	17,17	0,1318	16,62	12,73
Linear e com sazonalidade aditiva	0,5817	16,05	0,1214	15,33	11,67
Linear e com sazonalidade multiplicativa	0,5864	15,96	<b>0,1192</b>	<b>15,24</b>	11,47
<i>Damped</i> e sem sazonalidade	0,5084	17,40	0,1318	16,62	12,73
<i>Damped</i> e com sazonalidade aditiva	0,5817	16,29	0,1214	15,33	11,67
<i>Damped</i> e com sazonalidade multiplicativa	0,5865	16,20	<b>0,1192</b>	<b>15,24</b>	<b>11,46</b>

De acordo com a Tabela 6. o modelo que obteve o melhor ajuste foi o que apresentou tendência linear constante e sazonalidade multiplicativa, já que possui o maior  $R^2$  e nas demais estatísticas o menor valor.

Na Tabela 7 estão as constantes de amortecimento para este modelo e na Tabela 8 suas componentes sazonais.

Tabela 7: Constantes de Amortecimento

Nível ( $\alpha$ )	0,43322
Sazonalidade ( $\gamma$ )	0,12932

Tabela 8: Componentes Sazonais

Janeiro	1,80105	Julho	0,91834
Fevereiro	1,02573	Agosto	0,97289
Março	1,08894	Setembro	0,94562
Abril	1,07433	Outubro	1,04053
Mai	1,01596	Novembro	0,98528
Junho	0,85545	Dezembro	1,02402

Após escolhido o modelo de Holt-Winters mais adequado para modelar a série, uma previsão será feita. A Figura 5 apresenta a série original sobreposta à série ajustada, bem como da previsão que este modelo realiza para o ano de 2020, representadas nas cores preto e vermelho, respectivamente.

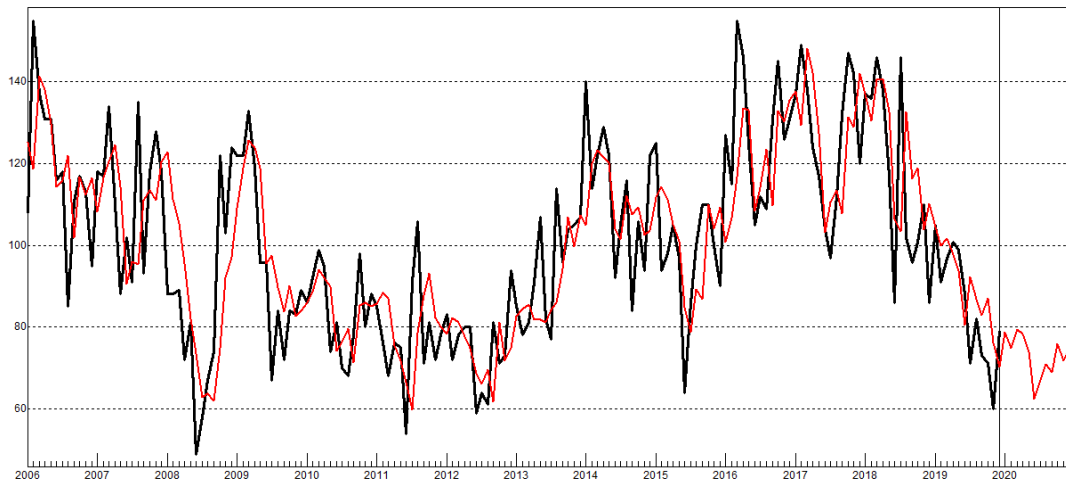


Figura 5: Série original e previsão modelada via Holt-Winters

Com a finalidade de se fazer a análise de resíduos para ver a adequação do modelo, o correlograma dos resíduos foi feito. Como pode ser observado na Figura 6, nota-se que os resíduos não apresentam forte correlação, indicando um comportamento de um ruído branco com média zero e variância constante e também o teste de normalidade apresentou um p-valor  $> 0,05$ , não rejeitando a hipótese de normalidade, indicando que o modelo tem um bom ajuste.

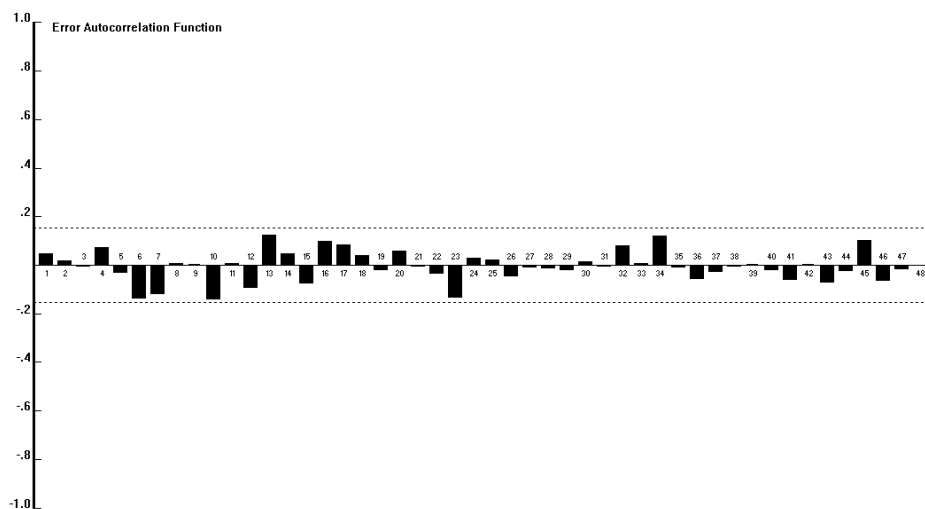


Figura 6: Correlograma da FAC dos Resíduos

### 3.3 Modelos de Box & Jenkins

Primeiramente foram feitos testes de normalidade, sendo eles o teste de Shapiro-Wilk, teste de Lilliefors e o teste de Jarque-Bera. Estes têm o intuito de verificar a normalidade da série original e suas hipóteses estão apresentadas em 2.16.

Feitos os testes, a Tabela 9 apresenta os respectivos p-valores para cada um dos testes mencionados.

Tabela 9: Testes de Normalidade

Testes	P-Valor
Teste de Shapiro-Wilk	0,02
Teste Lilliefors	0,08
Teste de Jarque-Bera	0,07

Sob um nível de significância de 0,05, tem-se que apenas o teste de Shapiro-Wilk teve a hipótese nula rejeitada, dando indícios de não normalidade nos dados.

Posteriormente, a fim de obter que todos os testes não rejeitem a hipótese nula de normalidade, foi realizada, devido a natureza dos dados, uma transformação logarítmica nos dados para normalizar a série. A Tabela 10 apresenta estes resultados.

Tabela 10: Testes de Normalidade com transformação logarítmica

Testes	P-Valor
Teste de Shapiro-Wilk	0,07
Teste Lilliefors	0,19
Teste de Jarque-Bera	0,17

Agora pode-se observar que em todos os testes feitos a hipótese nula não foi rejeitada, portanto assume-se normalidade e as análises adiante serão feitas com base na série com o log dos dados originais.

Em seguida, foi realizado o teste de raiz unitária de Dickey-Fuller Aumentado (teste ADF), para saber se a série é estacionária ou não. Suas hipóteses podem ser observadas em 2.14.

Após realizado o teste, foi obtido um p-valor = 0,01, o que resulta em rejeitar a hipótese nula. Isso significa dizer que a série é estacionária e portanto não será necessário realizar diferenças simples a fim de tornar a série estacionária. Assim, até o momento se tem o seguinte modelo:

$$ARIMA(p, d = 0, q)$$

A análise das funções de autocorrelação e autocorrelação parcial auxilia na identificação de possíveis candidatos à modelos para ajustar os dados, sendo possível definir valores de  $p$  e  $q$  viáveis para que se verifique qual modelo terá o melhor ajuste. A Figura 7 apresenta os correlogramas.

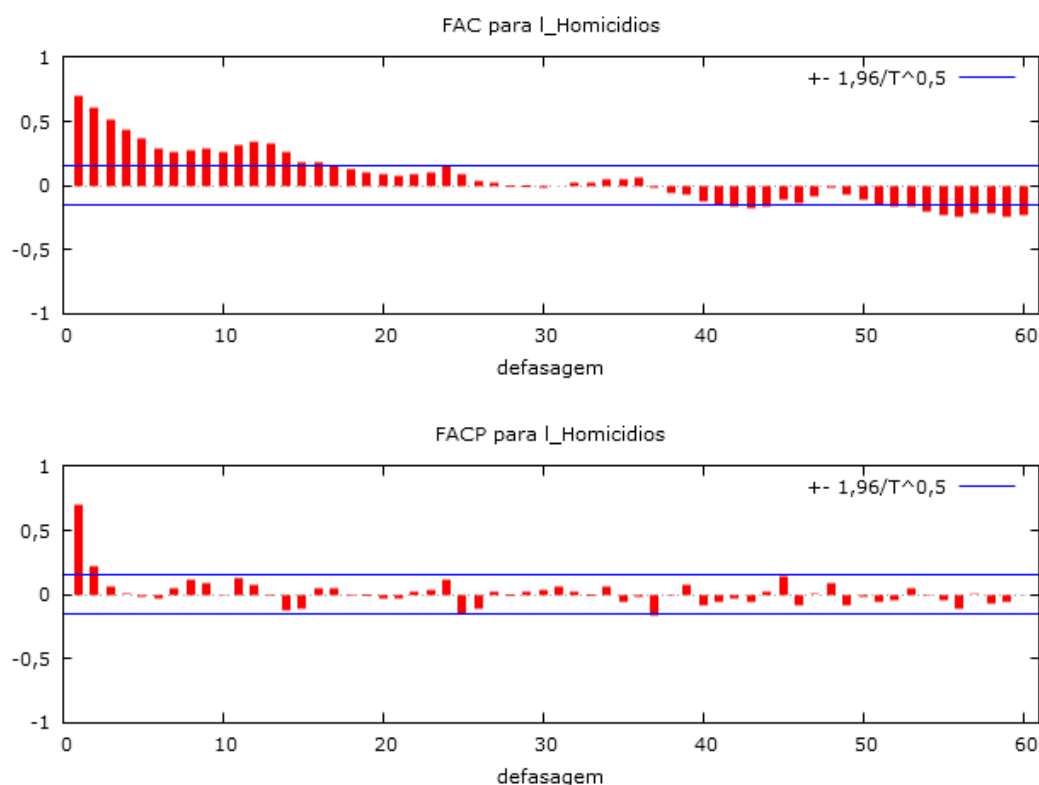


Figura 7: FAC e FACP do log dos homicídios

É possível identificar no gráfico da FAC que há uma interferência sazonal, pois nos lags 12, 24, 48 e 60 há significância. No lag 36 não há significância, porém a partir dele nota-se um leve crescimento e após isso um novo decaimentos das barras. Isso indica que será necessário aplicar uma diferença sazonal nos dados para identificar o comportamento neste novo correlograma e portanto a utilização dos modelos da classe *SARIMA* serão necessários.

Na parte simples do modelo, pode-se observar na FAC um decaimento exponencial a partir do primeiro lag e na FACP há dois lags significativos, sendo o lag 2 quase sendo não significativo, o que indica um comportamento  $AR(p)$ . Portanto surge daí dois candidatos de modelos na parte simples, sendo eles:

- $\text{SARIMA}(1, 0, 0) \times (P, D, Q)$ ;
- $\text{SARIMA}(2, 0, 0) \times (P, D, Q)$ .

Após realizada a diferença sazonal, pode-se observar o novo correlograma e identificar os valores da parte sazonal de  $P$  e  $Q$ , como é apresentado na Figura 8.

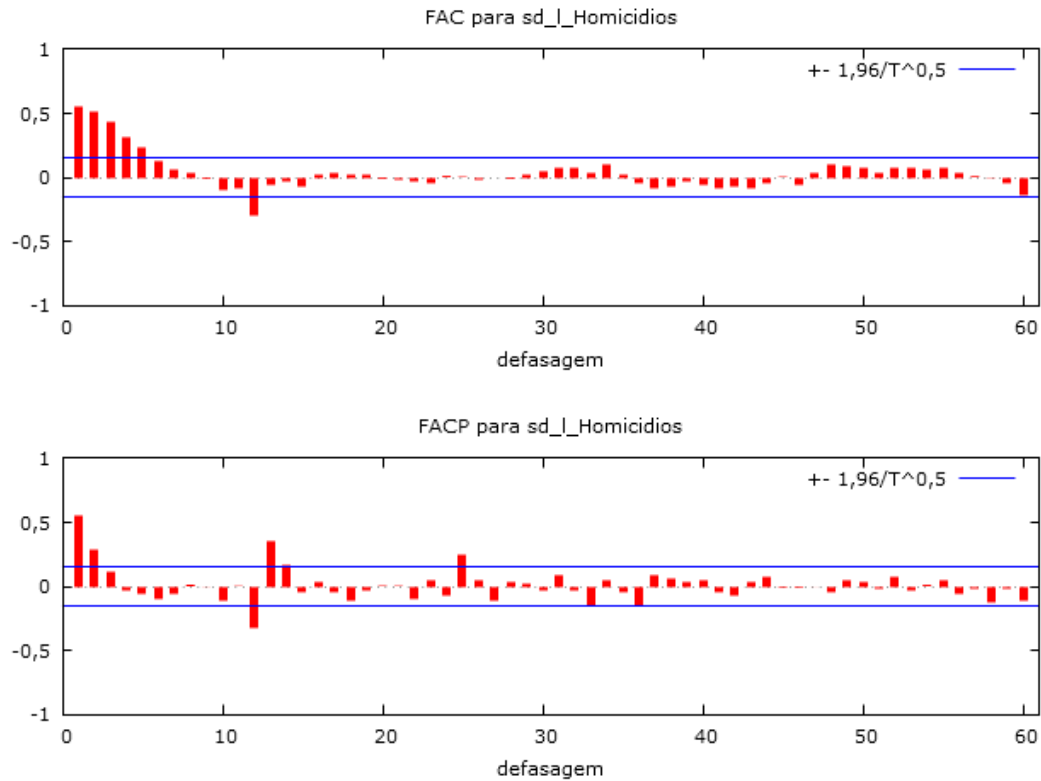


Figura 8: FAC e FACP com uma diferença sazonal do log dos homicídios

Pode ser notado na Figura 8 que na FAC no lag 12 há um pico e na FACP no mesmo lag 12 há um decaimento, o que indica um comportamento de um MA(1) sazonal. Também observa-se que na FACP há dois picos nos lags 12 e 24, que de forma concomitante com o pico no lag 12 da FAC sugere um ARMA(2,1) sazonal. Portanto com a adição destes dois candidatos, tem-se os seguintes modelos que serão testados:

- $\text{SARIMA}(1, 0, 0) \times (0, 1, 1)$ ;
- $\text{SARIMA}(2, 0, 0) \times (0, 1, 1)$ ;
- $\text{SARIMA}(1, 0, 0) \times (2, 1, 1)$ ;
- $\text{SARIMA}(2, 0, 0) \times (2, 1, 1)$ .



Ao ajustar os modelos propostos para obter as estatísticas de aderência, foi registrado que nos modelos  $\text{SARIMA}(1, 0, 0) \times (2, 1, 1)$  e  $\text{SARIMA}(2, 0, 0) \times (2, 1, 1)$  os parâmetros sazonais estimados do  $AR(P)$ , com  $P = 2$ , apresentou uma não significância dado a presença dos outros, como pode ser observado na Tabela 11 abaixo. Sendo assim, decidiu-se retirar um destes parâmetros indicados em vermelho do modelo para que se possa testar um modelo com  $P = 1$  e compará-lo aos demais, tornando-se necessário adicionar mais dois candidatos à modelos com para ajustar a série.

Tabela 11: Significância dos Parâmetros dos Modelos com  $P = 2$ 

Parâmetros	P-Valor
<i>SARIMA</i> (1, 0, 0) <i>X</i> (2, 1, 1)	
$\phi_1$	< 0,0001
$\Phi_1$	0,0160
$\Phi_2$	0,7649
$\Theta_1$	< 0,0001
<i>SARIMA</i> (2, 0, 0) <i>X</i> (2, 1, 1)	
$\phi_1$	< 0,0001
$\phi_2$	< 0,0001
$\Phi_1$	0,0369
$\Phi_2$	0,8651
$\Theta_1$	< 0,0001

Os seguintes modelos serão ajustados e comparados para a escolha do melhor dentre eles.

- $\text{SARIMA}(1, 0, 0) \times (0, 1, 1)$ ;
- $\text{SARIMA}(1, 0, 0) \times (1, 1, 1)$ ;
- $\text{SARIMA}(1, 0, 0) \times (2, 1, 1)$ ;
- $\text{SARIMA}(2, 0, 0) \times (0, 1, 1)$ ;
- $\text{SARIMA}(2, 0, 0) \times (1, 1, 1)$ ;
- $\text{SARIMA}(2, 0, 0) \times (2, 1, 1)$ .

Após obtidos esses candidatos foram encontradas estatísticas de aderência, que são apresentadas na Tabela 12.

Tabela 12: Estatísticas de Aderência para os Modelos de Box &amp; Jenkins

Modelos	$R^2$	BIC	MAPE	RMSE	MAD
<i>SARIMA</i> (1, 0, 0) $\times$ (0, 1, 1)	0,6002	15,39	0,1135	14,88	10,93
<i>SARIMA</i> (1, 0, 0) $\times$ (1, 1, 1)	0,5968	15,69	0,1171	14,95	11,17
<i>SARIMA</i> (1, 0, 0) $\times$ (2, 1, 1)	0,5982	15,90	0,1173	14,93	11,17
<i>SARIMA</i> (2, 0, 0) $\times$ (0, 1, 1)	0,6304	15,02	0,1117	14,24	10,63
<i>SARIMA</i> (2, 0, 0) $\times$ (1, 1, 1)	0,6326	15,21	0,1145	14,24	10,90
<i>SARIMA</i> (2, 0, 0) $\times$ (2, 1, 1)	0,6374	15,34	0,1135	14,22	10,84

Pelos resultados obtidos na Tabela 12, os modelos  $\text{SARIMA}(2, 0, 0) \times (0, 1, 1)$  e  $\text{SARIMA}(2, 0, 0) \times (2, 1, 1)$  apresentam as melhores estatísticas de aderência, porém o primeiro possui o menor BIC e também menos parâmetros. Portanto o modelo que melhor se ajusta aos dados é o  $\text{SARIMA}(2, 0, 0) \times (0, 1, 1)$ . A Figura 9 mostra a série original sobreposta à série ajustada pelo modelo escolhido de Box & Jenkins, representados nas cores preto e vermelho, respectivamente, assim como a sua previsão para o ano de 2020.

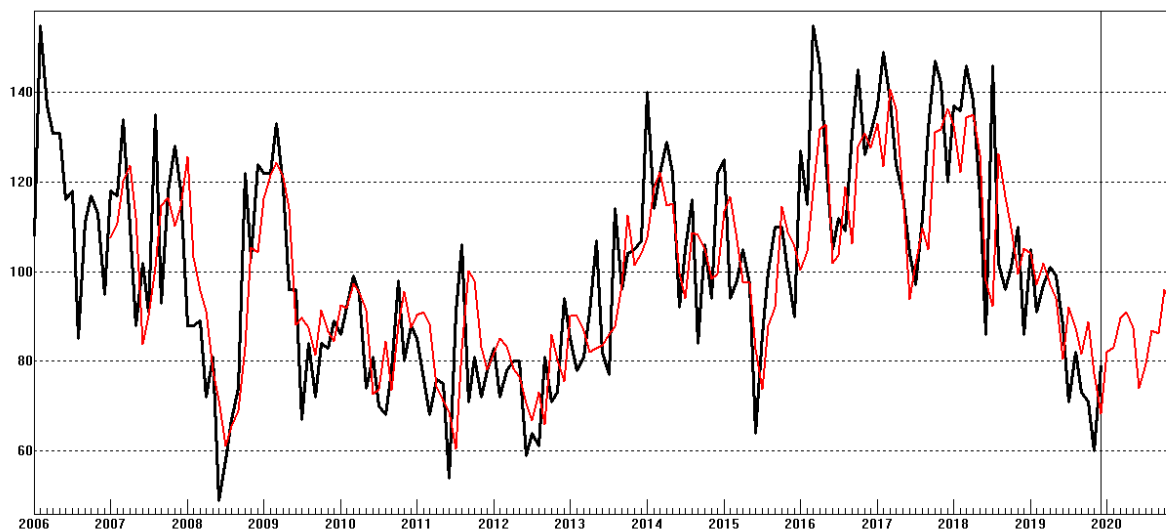


Figura 9: Série original e previsão modelada via Box & Jenkins

A Tabela 13 apresenta as estimativas dos parâmetros do modelo ajustado.

Tabela 13: Parâmetros do Modelo de Box & Jenkins

Parâmetros	Estimativa
$\phi_1$	0,4935
$\phi_2$	0,3227
$\Theta_1$	0,9098

Com o objetivo de se fazer a análise dos resíduos para ver a adequação do modelo, o correlograma dos resíduos foi feito. Como pode ser observado na Figura 10, nota-se que os resíduos não apresentam forte correlação, apresentando um comportamento de um ruído branco com média zero e variância constante, indicando que o modelo tem um bom ajuste.

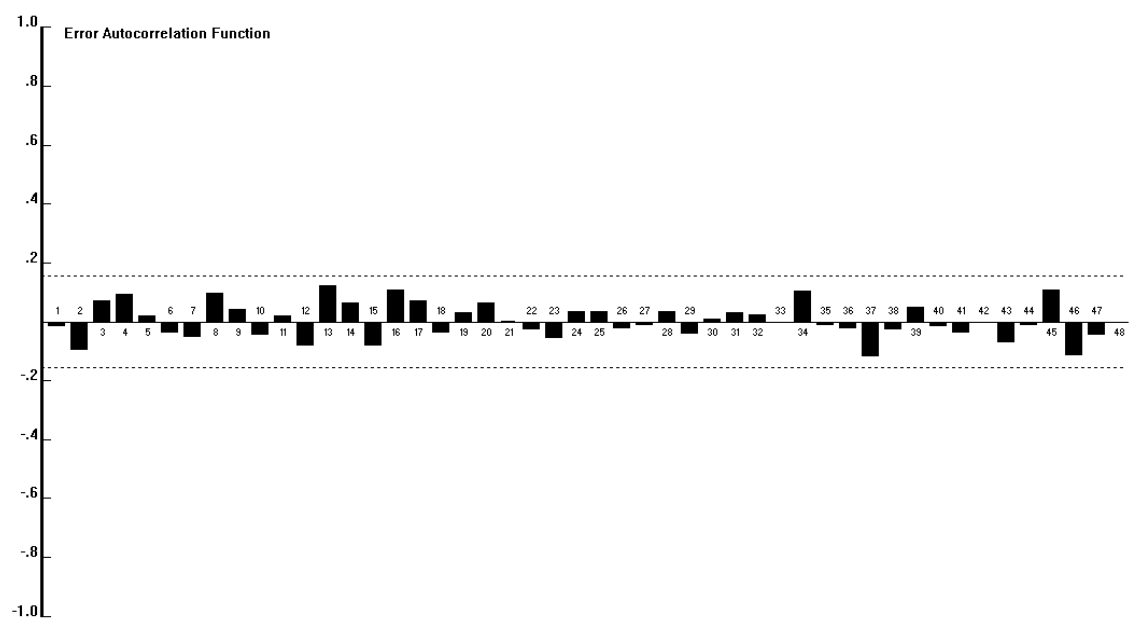


Figura 10: Correlograma da FAC dos Resíduos

E por fim, a normalidade destes resíduos foi testada, como pode ser observado na Figura 11 e na Tabela 14.

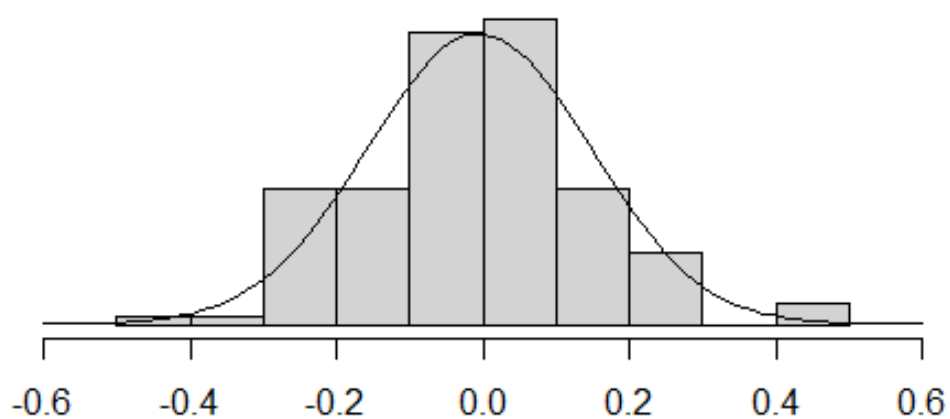


Figura 11: Normalidade dos Resíduos

Tabela 14: Testes de Normalidade para os Resíduos

Testes	P-Valor
Teste de Shapiro-Wilk	0,139
Teste Lilliefors	0,410
Teste de Jarque-Bera	0,485

Pode ser observado então que os resíduos seguem uma distribuição normal, portanto o modelo tem, de fato, um bom ajuste.

### 3.4 Escolha do modelo final

Após escolhidos os modelos nas Seções 3.2 e 3.3, restaram dois candidatos de cada método de modelagem. O modelo de Holt-Winters selecionado foi o modelo sem tendência linear e com sazonalidade multiplicativa, enquanto que o de Box & Jenkins escolhido foi o **SARIMA(2, 0, 0) × (0, 1, 1)**.

Foram comparadas suas respectivas estatísticas de aderência para a escolha do melhor dentre eles, como pode ser observado na Tabela 15 abaixo.

Tabela 15: Estatísticas de Aderência para os Modelos selecionados

Modelos	$R^2$	BIC	MAPE	RMSE	MAD
Constante e com sazonalidade multiplicativa	0,5868	15,71	0,1192	15,24	11,46
<b>SARIMA(2, 0, 0) × (0, 1, 1)</b>	<b>0,6304</b>	<b>15,02</b>	<b>0,1117</b>	<b>14,24</b>	<b>10,63</b>

Portanto, o modelo escolhido para modelar a série de dados e realizar sua devida previsão será o modelo de Box & Jenkins, **SARIMA(2, 0, 0) × (0, 1, 1)**. Seu gráfico de previsão de um ano, junto com dados dos dois últimos anos, representados nas cores vermelho e preto respectivamente, pode ser visto com mais clareza na Figura 12 e os valores de homicídios previstos mês a mês para o ano de 2020 podem ser observados na Tabela 16.

Tabela 16: Previsões Mês a Mês

Mês/Ano	Previsão	Mês/Ano	Previsão
Jan/2020	83	Jul/2020	80
Fev/2020	84	Ago/2020	87
Mar/2020	90	Set/2020	87
Abr/2020	92	Out/2020	97
Mai/2020	88	Nov/2020	93
Jun/2020	74	Dez/2020	95

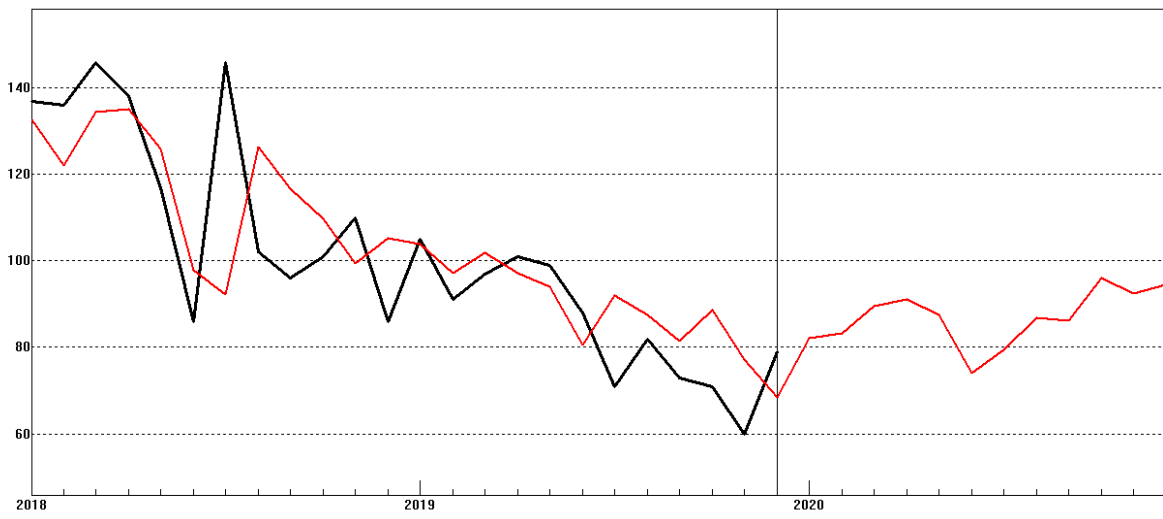


Figura 12: Série dos dados de 2018 e 2019 com previsão para 2020

## 4 Conclusão

Os homicídios dolosos não ocorrem de forma uniforme durante os dias da semanas, meses e anos, as ocorrências, como vistas, se comportam de maneira sazonal e em determinadas épocas as mortes têm um índice mais elevado que em outras, principalmente em temporadas de férias e quando há mais pessoas circulando. Para isso, são necessárias análises mais apuradas do comportamento dos homicídios no estado do Rio de Janeiro, com a finalidade de que possam ser tomadas medidas como forma de prevenção e monitoramento da vigilância por parte das autoridades responsáveis.

As técnicas estatísticas se tornam importantes neste processo, para que com os modelos de previsões de séries temporais possa se ter um panorama do que pode vir a ocorrer em um determinado período de tempo à frente. Realizar esta previsão foi um dos objetivos deste projeto, além de identificar descritivamente os dados, de forma que se possa ter clareza sobre o que acontece no estado quando se fala de homicídios.

Os modelos de Holt-Winters e Box & Jenkins foram utilizados na série original e diversas técnicas foram testadas para que os modelos pudessem ser aplicados. Após comparadas as estatísticas de aderência ( $R^2$ ), ( $BIC$ ), ( $MAPE$ ), ( $RMSE$ ) e ( $MAD$ ), os resultados demonstraram que a série, via modelos de Box & Jenkins, obteve a melhor modelagem comparado aos modelos de Holt-Winters.

Observou-se também que o estado do Rio de Janeiro registrou maiores números de ocorrências de homicídios no período da noite e da madrugada, principalmente aos fins de semana e em via pública nas cidades de maior índice urbano. Quanto às vítimas, pessoas do sexo masculino e da cor preta e parda sofreram mais com este tipo de violência e na maioria dos casos esta vítima não tinha nenhum grau de parentesco ou proximidade com seu agressor.

Este conglomerado de informações demonstra que este tipo de violência impacta na vida das pessoas de uma sociedade como um todo, então este estudo contribuiu para um melhor entendimento do comportamento dos homicídios no estado por meio de previsões e

descrições. Futuramente poderá servir como apoio para outros estudos e também utilizado para dar auxílio na criação de políticas públicas e para colaborar em uma melhora da segurança pública.

## Referências

- 1 KRUG, E. G. et al. The world report on violence and health. *The lancet*, Elsevier, v. 360, n. 9339, p. 1083–1088, 2002.
- 2 ORGANIZATION, W. H. et al. Global consultation on violence and health. *Violence: a public health priority*. Geneva: WHO, 1996.
- 3 WALTERS, R. H.; PARKE, R. D. Social motivation, dependency, and susceptibility to social influence. In: *Advances in experimental social psychology*. [S.l.]: Elsevier, 1964. v. 1, p. 231–276.
- 4 G1. *Homicídios mataram mais que conflitos armados em 2017, diz ONU*. 2019. Disponível em: <https://g1.globo.com/mundo/noticia/2019/07/08/homicidios-mataram-mais-que-conflitos-armados-em-2017-diz-onu.ghtml>. Acesso em: 04 mai. 2021.
- 5 DRUGS, U. N. O. on; CRIME. *Global study on homicide 2019: trends, contexts, data*. [S.l.]: UNODC, 2019.
- 6 UNICEF et al. A familiar face: Violence in the lives of children and adolescents. *New York, NY: UNICEF*, 2017.
- 7 ABRAMOVAY, M. et al. Juventude, violência e vulnerabilidade social na América Latina: desafios para políticas públicas. UNESCO Brasília/Interamerican Development Bank, 2002.
- 8 CERQUEIRA, D. et al. Atlas da violência 2019. Instituto de Pesquisa Econômica Aplicada (Ipea), 2019.
- 9 PENAL, C. Decreto-lei nº 2.848, de 7 de dezembro de 1940. *JusBrasil*, 1940.
- 10 DRUGS, U. N. O. on; CRIME. *Global study on homicide 2013: trends, contexts, data*. [S.l.]: UNODC, 2013.
- 11 BRASIL. *Diretrizes nacionais para a atenção integral à saúde de adolescentes e jovens na promoção, proteção e recuperação da saúde*. [S.l.]: Ministério da Saúde Brasília, 2010.
- 12 BRASIL. Lei nº 12.852, de 5 de agosto de 2013. institui o Estatuto da Juventude e dispõe sobre os direitos dos jovens, os princípios e diretrizes das políticas públicas de juventude eo sistema nacional de juventude-sinajuve. *Diário Oficial da União*, 2013.
- 13 ESTATÍSTICA, I. B. de. *Pesquisa nacional por amostra de domicílios*. [S.l.: s.n.], 2019.



- 14 SANTOS, F. dos; SILVA, R. S.; NERY, F. S. Comparativo das taxas de homicídio da população jovem masculina, Brasil e Sergipe, entre 1996 a 2014. In: *Congresso Internacional de Enfermagem*. [S.l.: s.n.], 2017. v. 1, n. 1.
- 15 REICHENHEIM, M. E. et al. Violence and injuries in Brazil: the effect, progress made, and challenges ahead. *The Lancet*, Elsevier, v. 377, n. 9781, p. 1962–1975, 2011.
- 16 SACHSIDA, A.; MENDONÇA, M. J. C. de. *Evolução e determinantes da taxa de homicídios no Brasil*. [S.l.], 2013.
- 17 ESTATÍSTICA, I. B. de. *Séries Históricas Anuais de Taxa de Letalidade Violenta*. [S.l.: s.n.], 2020.
- 18 NEVES, A. C. M. d.; GARCIA, L. P. Mortalidade de jovens brasileiros: perfil e tendências no período 2000-2012. *Epidemiologia e Serviços de Saúde*, SciELO Public Health, v. 24, p. 595–606, 2015.
- 19 PROVENZA, M. M. Análise e previsão de séries temporais do homicídio doloso no Rio de Janeiro. *Cadernos de Estudos Sociais e Políticos*, v. 4, n. 7, p. 63–83, 2015.
- 20 MELO, A. C. M.; SILVA, G. D. M. d.; GARCIA, L. P. Mortalidade de homens jovens por agressões no Brasil, 2010-2014: estudo ecológico. *Cadernos de Saúde Pública*, SciELO Public Health, v. 33, p. e00168316, 2017.
- 21 PROVENZA, M. M.; COSTA, J. F. da S.; SILVA, L. de C. Análise de dados e previsão de séries temporais do homicídio doloso no estado do Rio de Janeiro entre 2001 e 2016. *Produção em Foco*, v. 8, n. 2, 2018.
- 22 ISP. *Letalidade Violenta*. 2020. Disponível em: <http://www.isp.rj.gov.br/>. Acesso em: 20 mai. 2020.
- 23 TOLOI, C. M. de C.; MORETTIN, P. A. *Análise de Séries Temporais*. [S.l.]: Edgard Blucher, 2004.
- 24 EHLERS, R. S. Análise de séries temporais. *Universidade Federal do Paraná*, 2007.
- 25 SOUZA, R. C.; CAMARGO, M. E. Análise e previsão de séries temporais. *Gráfica e Editora Regional*. Rio de Janeiro, 2004.
- 26 PORTAL ACTION. *Teste de Dickey-Fuller Aumentado*. s.d. Disponível em: <http://www.portalaction.com.br/inferencia/64-teste-de-shapiro-wilk>. Acesso em: 01 mai. 2021.
- 27 SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, JSTOR, v. 52, n. 3/4, p. 591–611, 1965.
- 28 PORTAL ACTION. *Teste de Shapiro-Wilk*. s.d. Disponível em: <http://www.portalaction.com.br/inferencia/64-teste-de-shapiro-wilk>. Acesso em: 01 mai. 2021.
- 29 JARQUE, C. M.; BERA, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, Elsevier, v. 6, n. 3, p. 255–259, 1980.

- 30 DALLAL, G. E.; WILKINSON, L. An analytic approximation to the distribution of Lilliefors's test statistic for normality. *The American Statistician*, Taylor & Francis, v. 40, n. 4, p. 294–296, 1986.
- 31 WHEELWRIGHT, S.; MAKRIDAKIS, S.; HYNDMAN, R. J. *Forecasting: methods and applications*. [S.l.]: John Wiley & Sons, 1998.
- 32 GARDNER JR, E. S.; MCKENZIE, E. Forecasting trends in time series. *Management science*, INFORMS, v. 31, n. 10, p. 1237–1246, 1985.
- 33 FRANCO, G. C. Apostila sobre modelagem box and jenkins.
- 34 AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974.
- 35 SCHWARZ, G. et al. Estimating the dimension of a model. *The annals of statistics*, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.