



UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA
BACHARELADO EM ESTATÍSTICA



Análise Temporal das Taxas de Suicídios no Estado de Minas Gerais

Rodnei Alves Marques

Ouro Preto-MG
Novembro / 2020

Rodnei Alves Marques

Análise Temporal das Taxas de Suicídios no Estado de Minas Gerais

Monografia de Graduação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do grau de bacharel em Estatística.

Orientador

Dr. Spencer Barbosa da Silva

UNIVERSIDADE FEDERAL DE OURO PRETO – UFOP
DEPARTAMENTO DE ESTATÍSTICA – DEEST

Ouro Preto - MG

Novembro / 2020



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE ESTATÍSTICA

**FOLHA DE APROVAÇÃO****Rodnei Alves Marques****Análise Temporal das Taxas de Suicídios no Estado de Minas Gerais**

Membros da banca

Spencer Barbosa da Silva - Doutor em Estatística - UFOP
Dr. Tiago Martins Pereira - Doutor em Estatística - UFOP
Dr. Rívert Paulo Braga Oliveira - Doutor em Estatística - UFOP

Versão final

Aprovado em 25 de novembro de 2020

De acordo

Professor Orientador Spencer Barbosa da Silva



Documento assinado eletronicamente por **Rivert Paulo Braga Oliveira, PROFESSOR DE MAGISTERIO SUPERIOR**, em 25/11/2020, às 16:43, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Spencer Barbosa da Silva, CHEFE DO DEPARTAMENTO DE ESTATÍSTICA**, em 25/11/2020, às 16:43, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Tiago Martins Pereira, COORDENADOR(A) DO CURSO DE ESTATÍSTICA**, em 25/11/2020, às 16:44, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0105628** e o código CRC **9C10005B**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.008289/2020-54

SEI nº 0105628

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: - www.ufop.br

Resumo

O autoextermínio tem aumentado em Minas Gerais em todas as faixas etárias. Tal fato, além da perda de vidas humanas, demanda recursos públicos que poderiam ser aplicados em outras áreas da saúde. Diante deste tema, esse trabalho tem como objetivo modelar e fazer previsões para as taxas específicas de suicídio por sexo no estado de Minas Gerais. Estas análises, juntamente com estudos clínicos, podem compor mais uma alternativa para os órgãos de saúde, auxiliando-os no preparo de ações que possam diminuir de forma efetiva o número de suicídios. As séries foram ajustadas com a metodologia de Box & Jenkins com utilização dos softwares R e GRETLL. Foi observada uma tendência crescente histórica nas séries ao longo dos anos para as taxas específicas de suicídios masculinos e femininos e conclui-se que este crescimento ainda prevalece para os próximos anos caso não haja nenhuma intervenção mais efetiva.

Palavras-chave: Séries Temporais, Taxa de Suicídios, Modelos ARIMA, Previsões.

Abstract

Self-extermination has increased in Minas Gerais in all age groups. This fact, in addition to the loss of human lives, demands public resources that could be applied in other areas of health. Given this theme, this work aims to model and make predictions for specific suicide rates by gender in the state of Minas Gerais, so these analyses are another alternative for health agencies to prepare, together with clinical studies, actions that can effectively reduce the number of suicides. The series were adjusted using the Box & Jenkins methodology using the R and GRETLM softwares. A historical growing trend has been observed in the series over the years for specific rates of male and female suicides and we concluded this growth still prevails for the next years if there is no more effective intervention.

Keywords: Time Series, Suicide Rate, ARIMA Models, Forecasts.

Lista de figuras

1	Exemplo de uma série temporal $Z(t)$	p. 13
2	Tendências de um série temporal $Z(t)$	p. 13
3	Decomposição de uma série no domínio da frequência	p. 14
4	Gráfico da precipitação de chuva em Porto Velho - RO	p. 15
5	Processo estocástico interpretado como uma família de variáveis, Moret- tin e Tolo (2006).	p. 21
6	Modelo ARIMA(1,1,1).	p. 23
7	Funções fac e $facp$ de um modelo ARIMA(1,0,2).	p. 24
8	Séries de suicídios por sexo no estado de Minas Gerais (1992-2019) . . .	p. 29
9	Gráfico Amplitude x Média em a) Série Masculina e b) Série feminina .	p. 30
10	Série diferenciada da taxa específica de suicídios em Minas Gerais . . .	p. 31
11	Funções de autocorrelação e autocorrelação parcial das séries de Suicídios	p. 32
12	Funções de autocorrelação e autocorrelação parcial das séries diferenciadas	p. 33
13	Série de suicídios (efetivo x ajustado)	p. 35
14	Funções de autocorrelação e autocorrelação parcial dos resíduos	p. 36
15	Q-Q plot das séries de suicídios masculino e feminino em Minas Gerais	p. 37
16	Gráficos com previsões para as séries de suicídios por sexo em Minas Gerais	p. 39
17	Diagnósticos dos modelos ajustados	p. 39

Lista de tabelas

1	Teste Dickey Fuller Aumentado (Série suicídios em MG)	p. 31
2	Teste Dickey Fuller Aumentado (Série diferenciada)	p. 31
3	Teste Kruskal-Wallis (Séries de suicídios em MG)	p. 32
4	Critério Akaike para os modelos estimados	p. 34
5	Parâmetros dos modelos ajustados	p. 34
6	Teste Ljung-Box (Resíduos)	p. 36
7	Previsões das séries sem as últimas observações	p. 38
8	Previsões futuras das séries	p. 38

Lista de Siglas

ADF Teste Dickey Fuller Aumentado

AIC Critério de Akaike

ARIMA Autorregressivo Integrados de Médias Móveis

FAC Função de Autocorrelação

FACP Função de Autocorrelação Parcial

GRETLL Gnu Regression, Econometrics and Time-series Library

IBGE Instituto Brasileiro de Geografia e Estatística

IPEA Instituto de Pesquisas Econômicas Aplicadas

OMS Organização Mundial da Saúde

TES Taxa Específica de Suicídios

Sumário

1	Introdução	p. 9
2	Revisão de Literatura	p. 12
2.1	Séries temporais	p. 12
2.2	Metodologias estatísticas de previsão	p. 15
2.2.1	Estacionalidade	p. 16
2.2.2	Função de autocorrelação	p. 19
2.2.3	Função de autocorrelação parcial	p. 20
2.3	Metodologia Box & Jenkins	p. 21
2.3.1	Etapas da modelagem	p. 23
3	Materiais e Métodos	p. 27
4	Resultados e Discussões	p. 29
4.1	Identificação dos modelos	p. 29
4.2	Estimação dos modelos	p. 33
4.3	Verificação dos modelos	p. 35
4.4	Previsões	p. 37
5	Considerações Finais	p. 40
	Referências	p. 41

1 Introdução

As ocorrências de suicídios têm se tornado cada vez mais frequentes no Brasil e no mundo, com crescimento progressivo nos últimos anos. De acordo com a Organização Mundial da Saúde (OMS, 2019), para cada adulto que morreu por suicídio, pode ter havido mais vinte outros tentando suicidar-se. O ato suicídio é considerado a terceira maior causa de óbitos por fatores externos, perdendo apenas para homicídios e acidentes de trânsito, nesta ordem. Enquadram-se como suicídios atos que causam a própria morte de forma intencional.

De acordo com Filho e Zerbini (2016), os índices de suicídios aumentaram em todas as faixas etárias, mas tentativas de suicídios são maiores entre as mulheres e jovens, sendo a intoxicação por medicamentos o método mais utilizado. No entanto, homens brasileiros provocam autoextermínio efetivo em maior intensidade, ou seja, a população masculina utiliza métodos de maior letalidade. Segundo os autores, os métodos mais utilizados pelos homens são enforcamentos e arma de fogo.

Resultados semelhantes foram encontrados por Ribeiro et al. (2018). Os autores analisaram 169 tentativas de suicídios na cidade de Uberaba em Minas Gerais e concluíram que, nas tentativas de suicídio por pessoa do sexo feminino, destacaram-se o uso de medicamentos e envenenamento e, para homens, os principais meios de auto agressão foram enforcamento, seguido de autointoxicação. Para as análises da série temporal os autores utilizaram análises descritivas e regressão polinomial.

No trabalho realizado por Gonçalves, Gonçalves e Júnior (2011), os autores abordam a dependência espacial do número de casos, ou seja, a constatação do "efeito contágio" espacial a partir de amostras de microrregiões geográficas brasileiras. Eles defendem a criação de um plano nacional de prevenção ao suicídio, que objetive sensibilizar a sociedade e mostrar que o suicídio não está limitado ao indivíduo, mas sim relacionado à sociedade como um todo.

Do ponto de vista econômico, o suicídio e suas tentativas representam enorme custo

para a sociedade. Primeiro, porque demanda recursos públicos que poderiam ser aplicados em outras áreas da saúde, e segundo, porque resulta na perda do capital humano. Em uma pesquisa realizada pelo Instituto de Pesquisas Econômicas Aplicadas (IPEA, 2020), o valor médio das internações registradas por tentativas de suicídios entre os anos de 1998 e 2004 foi de R\$ 507,00 (quinhentos e sete reais), com média de internação de quatro dias.

Analogamente, ainda considerando o ponto de vista econômico, Branas et al. (2015) utilizaram os modelos ARIMA para analisar as séries temporais de suicídios mensais na Grécia entre os anos de 1983 e 2012. Os autores alertam sobre o aumento da taxa de suicídios no mesmo período em que medidas de austeridade econômica são aplicadas à população. As análises mostraram que as medidas de austeridade econômica aplicadas em janeiro de 2002, outubro de 2008, junho de 2011 e abril de 2012 provocaram aumentos significativos nas ocorrências de suicídios entre a população.

Silveira, Santos e Ferreira (2012), alertam que a prevalência de autoextermínio entre os homens condiciona um valor total de custo 1,66 vezes maior que o valor gasto para as internações de mulheres. Além disso, a média de permanência e valor médio de internações do sexo masculino são maiores, o que denota um quadro de maior gravidade. Porém as internações do sexo feminino são mais dispendiosas, requerendo mais gastos com equipamentos e um maior valor agregado por dia.

Diante deste tema, o objetivo deste trabalho é modelar e fazer previsões para as taxas específicas de suicídios no estado de Minas Gerais, de modo que as análises sejam mais uma alternativa às secretarias de saúde, conselhos regionais de psicologia e também organizações não governamentais prepararem, juntamente com estudos clínicos, no planejamento de ações que possam reduzir de forma efetiva as taxas de suicídios no estado. Os dados das taxas de suicídios por sexo foram obtidas no banco de dados do Instituto de Pesquisas Econômicas Aplicadas (IPEA) e do Instituto Brasileiro de Geografia e Estatística (IBGE).

Para ajuste dos modelos foi utilizada a metodologia de Box & Jenkins, que consiste nos modelos autorregressivos integrados de médias móveis, às séries das taxas de suicídios em Minas Gerais. Essa metodologia parte do princípio de que os modelos temporais podem ser desenvolvidos em ciclos iterativos, no qual a estrutura do modelo é baseada nos próprios dados. Foram escolhidos os modelos que apresentaram melhores ajustes e feitas previsões para os anos de 2020, 2021 e 2022.

As séries foram analisadas no *software* R e no *software* GRETL. O trabalho está

organizado da seguinte forma: Inicialmente é apresentada uma revisão de literatura com citações de trabalhos que desenvolveram previsões com modelos ARIMA; posteriormente são abordados os materiais e métodos com apresentação dos dados e a metodologia adotada; os resultados e discussões com apresentação dos melhores modelos, consolidados com testes estatísticos: Por fim, os resultados e discussões do presente estudo.

2 Revisão de Literatura

2.1 Séries temporais

Uma série temporal pode ser definida como um conjunto de observações discretas ou contínuas ao longo do tempo, que apresentam dependência de um instante para outro. Estas observações podem ser decompostas nos componentes de tendência, ciclo e sazonalidade. As análises podem ser feitas no domínio temporal com modelos paramétricos ou no domínio de frequências com modelos não paramétricos.

Em todos os casos, modelos probabilísticos ou modelos estocásticos são construídos, no domínio do tempo ou de frequências. Estes modelos devem ser simples de forma que o número de parâmetros deva ser o menor possível, (MORETTIN; TOLOI, 2006).

O conceito de fazer previsões é antigo e pode ser encontrado nas mais diversas áreas do conhecimento, como por exemplo: Engenharia, Economia, Saúde, Ambientais , etc. Com previsões futuras é possível tomar decisões prévias e fazer correções em uma sequência de eventos. Araujo et al. (2009), comentam que o pressuposto básico da análise de séries temporais é a de que fatores que influenciam o comportamento da série no passado continuarão a fazê-lo no futuro.

Com a análise de uma série temporal podemos investigar o mecanismo gerador da série, fazer previsões, descrever o comportamento e procurar periodicidades pertinentes. Então, a notação utilizada aqui para denotar uma série temporal de tamanho n , será:

$$Z(t_1), Z(t_2), Z(t_3), \dots, Z(t_n) , \quad (2.1)$$

em que Z_j é a componente da série no tempo j . Na Figura 1, é apresentado um gráfico que representa uma série temporal $Z(t)$.

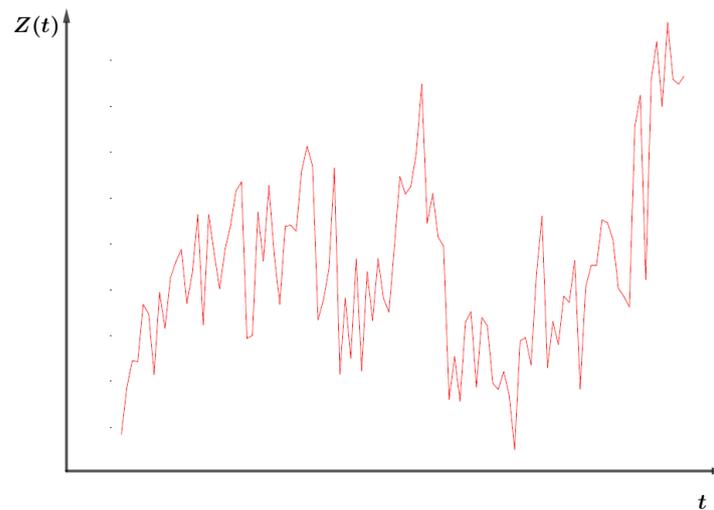


Figura 1: Exemplo de uma série temporal $Z(t)$

De acordo com Morettin e Tolo (2006), uma das suposições mais frequentes que se faz a respeito de uma série temporal é sobre sua estacionalidade, ou seja, se o seu comportamento, ao longo do tempo, se desenvolve aleatoriamente ao redor de uma média. A maneira mais tradicional de analisar uma série é através da sua decomposição nas componentes de tendência, ciclo e sazonalidade.

A tendência de uma série indica se ela cresce, decresce ou permanece estável e qual a velocidade destas mudanças. De acordo com Campos (2008), tendência em uma série temporal é a mudança gradual observada por meio da variação dos seus valores ao longo do tempo, e que se mantém ao se moverem as componentes de ciclos, sazonalidade e fatores aleatórios. A tendência pode ser visualizada pela linha em torno da qual os dados flutuam e os casos mais trabalhados são com tendência constante, linear ou quadrática, como mostrado na Figura 2.

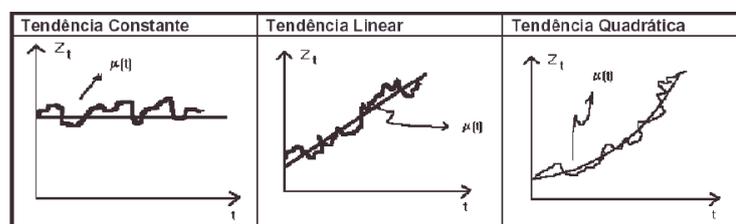


Figura 2: Tendências de um série temporal $Z(t)$

Os ciclos são caracterizados pelas oscilações nas séries, subida e descida de forma suave ou instantânea, ao longo da componente de tendência. Como um exemplo, pode-se pensar em ciclos relacionados à atividade econômica ou ciclos meteorológicos.

De acordo com Fonseca, Martins e Toledo (1985), a componente cíclica é entendida como parte da série temporal que apresenta um movimento em torno da tendência ao longo do tempo. O ciclo é o movimento de aparência quase periódica com fases de picos e fases de vale, alternadamente, que pode ser analisado através do seu espectro. A análise espectral é uma decomposição da série temporal em componentes cíclicas, e isso pode ser feito com utilização das transformadas de Fourier ou *wavelets*. Na Figura 3, é mostrado uma série temporal decomposta em suas componentes cíclicas, com utilização da transformada de Fourier (FT).

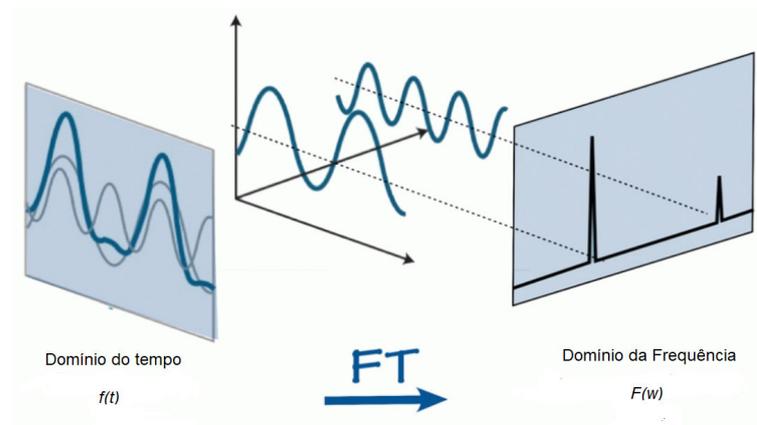


Figura 3: Decomposição de uma série no domínio da frequência

A sazonalidade em uma série temporal, muito confundida com ciclos, são as oscilações de subida e descida que ocorrem, frequentemente, em um determinado período de tempo. A principal diferença entre as componentes sazonal e cíclica é que a primeira possui movimentação de fácil previsão, que ocorrem em intervalos regulares, já os movimentos cíclicos tendem a serem irregulares.

As sazonalidades referem-se a movimentos de padrões regulares ou repetidos de curto prazo em torno de uma linha de tendência, em que os dados da série temporal parecem obedecer a períodos como dias, semanas, meses, etc. Morettin e Tolo (2006), classificam como movimentos sistemáticos causados por fenômenos não econômicos e aparecem quando as observações são intra-anuais. Na Figura 4, é mostrado como exemplo de sazonalidade, o gráfico da precipitação mensal de chuva na cidade de Porto Velho entre julho de 2007 até agosto de 2015.

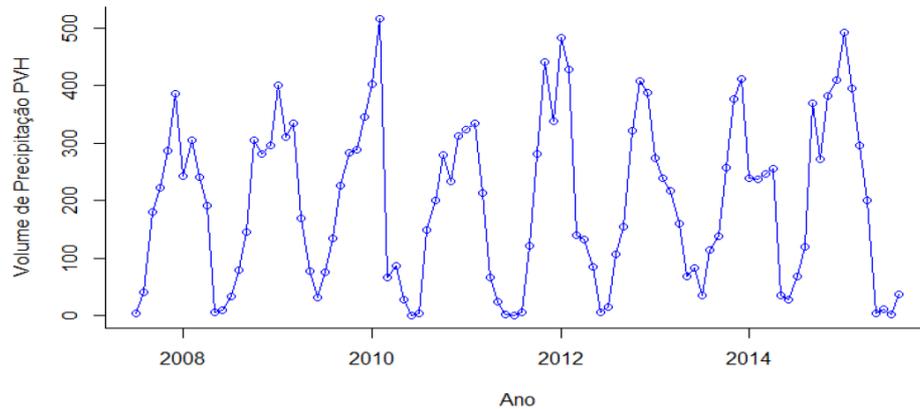


Figura 4: Gráfico da precipitação de chuva em Porto Velho - RO

Ao estudar uma série temporal estamos interessados em analisar e modelar a série, de forma a descrever e verificar suas características mais relevantes e possíveis relações com outras séries. Além disso, fazer previsões a partir de valores históricos procurando estimar previsões de curto prazo. O número de instantes à frente para o qual é feita a previsão é chamado de horizonte de previsão.

2.2 Metodologias estatísticas de previsão

Os sistemas de previsão podem ser simples e imediatos com pouca análise de dados ou complexos e substanciais, que necessitam de um trabalho considerável de interpretação. Dentre os procedimentos de previsão podemos destacar:

- **Modelos Univariados** - se baseia em um única série histórica ao longo do tempo. Como por exemplo, os modelos de Box et al. (1994) que consiste em uma classe de modelos auto-regressivos integrados e de médias móveis, conhecidos como modelos ARIMA.
- **Modelos de Função de Transparência** - nesse tipo de modelo a série também é explicada por outras séries temporais não correlacionadas entre si.
- **Modelos Multivariados** - são modelados simultaneamente duas ou mais séries sem qualquer exigência em relação a direção da causalidade entre elas.

Antes de introduzir os modelos de Box et al. (1994), vamos apresentar definições de estacionalidade e função de autocorrelação de um série temporal.

2.2.1 Estacionalidade

Uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente em torno de uma média constante, que reflete alguma forma de equilíbrio estável, (MORETTIN; TOLOI, 2006).

Uma série temporal pode ser estacionária em um período longo ou curto de tempo, se mudar de nível ou inclinação ela é não estacionária. Pode-se dizer ainda que é uma série que não mostra tendência, somente sazonalidade. Consequentemente uma série não estacionária apresenta as duas componentes, tendência e sazonalidade, ou uma delas, (SÁFADI, 2004).

De acordo com Barbosa, Iquiapaza e Bressan (2008), uma série estacionária é aquela cujo valor não é afetado permanentemente pelos valores de erros contidos nas observações passadas, um exemplo é o processo de ruído branco:

$$y_t = \epsilon_t , \quad (2.2)$$

em que t representa o tempo, y representa a variável dependente e ϵ a variável aleatória não correlacionada com média zero e variância constante. Também pode ter estacionalidade em torno de uma tendência:

$$y_t = T_t + \epsilon_t , \quad (2.3)$$

em que T_t caracteriza a tendência ao longo do tempo t .

De acordo com Morettin e Tolo (2006), existem a estacionalidade fraca, uma série ampla de segunda ordem e a forte, uma série mais estrita.

Definição 1. *Um processo estocástico $Z = Z(t), t \in T$, diz-se estritamente estacionário se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, ou seja,*

$$F(z_1, \dots, z_n, t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n, t_1, \dots, t_n) , \quad (2.4)$$

para quaisquer t_1, \dots, t_n, τ de T

Isso determina que todas as distribuições unidimensionais são invariantes no tempo, com isso, tanto a média como a variância são constantes.

Definição 2. *Um processo estocástico $Z = z(t), t \in T$, diz-se fracamente estacionário ou*

estacionário de segunda ordem, se e somente se,

i) $E\{Z(t)\} = \mu(t) = \mu$, constante, para todo $t \in T$;

ii) $E\{Z^2(t)\} < \infty$, para todo $t \in T$;

iii) $\psi(t_1, t_2) = Cov\{Z(t_1), Z(t_2)\}$ é uma função de $|t_1 - t_2|$

Existem diversos testes para identificação de tendência e sazonalidade. Os testes usados na presente proposta para identificar essas componentes serão os testes de Dickey-Fuller aumentado (ADF) e Kruskal-Wallis respectivamente.

• Teste Dickey Fuller Aumentado

A ideia inicial para análise de uma série temporal, ou seja, o primeiro passo é construir o gráfico da série para perceber características importantes como tendência e sazonalidade. Ademais, é necessário utilizar testes de hipóteses estatísticos para verificar a existência da componente. O teste de Dickey Fuller Aumentado (ADF), relaciona a presença de tendência com a existência de uma raiz dentro de um círculo unitário no polinômio autorregressivo.

De acordo com Dickey e Fuller (1979), o teste é usado para verificar se a série é estacionária, através do estudo sobre as raízes dos operadores retardados dentro de um círculo de raio um. A raiz unitária em modelos autorregressivos de médias móveis (ARMA) é um problema quando quando apresentam pelo menos uma raiz sobre o círculo unitário no polinômio autorregressivo.

Uma série temporal é estacionária se as características de sua média, variância ou covariância não variam ao longo do tempo, ou seja, ela é estacionária se desenvolve no tempo dentro de uma média constante, (GUJARATI; PORTER, 2011). O teste da raiz unitária tem a finalidade de detectar se a série é ou não estacionária.

O teste apresenta as seguintes hipóteses:

$$\begin{cases} H_0 : \text{Existe pelo menos uma raiz dentro do círculo unitário (não estacionária)} ; \\ H_1 : \text{Não existe raiz dentro do círculo unitário (estacionária)}. \end{cases}$$

A estatística ADF, usada no teste, é um número negativo, e quanto mais negativo, mais indicativo o teste se torna de rejeitar a hipótese nula, ele requer o estudo sobre a regressão:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-i} + \epsilon_t, \quad (2.5)$$

em que β_1 é o intercepto, também denominado como *drift* da série; β_2 é o coeficiente de tendência; δ é o coeficiente de presença da raiz unitária e m o número de defasagens tomadas na série. É feita uma regressão de Δy_t em $y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t+p-1}$ e calculada a estatística de teste dada por:

$$T = \frac{\hat{\delta}}{s(\hat{\delta})}, \quad (2.6)$$

em que $\hat{\delta}$ é o estimador de δ e $s(\hat{\delta})$ é o estimador para o desvio padrão do erro de δ . O resultado do teste é negativo, e quanto menor esse número, mais indicativo para rejeitar a hipótese nula de que existe raiz unitária na série.

Após verificar que uma série não é estacionária, existem técnicas para torná-la estacionária. Uma delas consiste em diferenciar os dados da série Z_t até que se torne estacionária, com isso temos uma nova série $Y_t = Z_t - Z_{t-1}$.

• Teste Kruskal Wallis

É um teste não paramétrico para detecção de sazonalidade que, de acordo com Morrettin e Tolo (2006), antes de ser usado é conveniente eliminar a tendência, se ela estiver presente. Assim, o teste será aplicado aos resíduos, salientando que por se tratar de um teste não paramétrico, ele não assume distribuição normal dos resíduos. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{Não existe sazonalidade;} \\ H_1 : \text{Existe sazonalidade.} \end{cases}$$

São selecionadas k amostras de tamanho n_j da série. Então as observações serão:

$$Y_{ij}, j = 1, \dots, k, i = 1, \dots, n_j, N = \sum_{j=1}^k n_j.$$

As observações Y_{ij} , são substituídas pelos seus postos R_{ij} , ordenando todas as N observações. A soma de todos os postos associados a j -ésima amostra,

$$R.j = \sum_{i=1}^{n_j} R_{ij}, \quad j = 1, \dots, k.$$

A hipótese H_0 de não existência de sazonalidade é rejeitada se a estatística,

$$T_1 = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1), \quad (2.7)$$

for maior ou igual ao valor crítico T_{1c} , sendo que T_{1c} é tal que $P_H(T_1 \geq T_{1c}) = \alpha$, α é o nível de significância do teste.

Uma das suposições do teste diz que as variáveis dentro de cada amostra devem ser independentes e que as amostras são independentes entre si, o que não ocorre em séries temporais. Essa é uma das críticas da aplicação do teste.

Portanto, devemos ser cauteloso ao aplicar estes testes, devido as suposições envolvidas para sua aplicação e a possibilidade das mesmas não serem válidas.

2.2.2 Função de autocorrelação

Um processo autorregressivo de ordem p pode ser definido por:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t. \quad (2.8)$$

A função de autocovariância pode ser obtida multiplicando-se ambos os lados da equação (2.6) por \tilde{Z}_{t-j} e calculando a esperança. Deste modo, tem-se:

$$E(\tilde{Z}_t \tilde{Z}_{t-j}) = \phi_1 E(\tilde{Z}_{t-1} \tilde{Z}_{t-j}) + \phi_2 E(\tilde{Z}_{t-2} \tilde{Z}_{t-j}) + \dots + \phi_p E(\tilde{Z}_{t-p} \tilde{Z}_{t-j}) + E(\tilde{Z}_t \tilde{Z}_{t-j}), \quad (2.9)$$

como $E(\tilde{Z}_t \tilde{Z}_{t-j}) = \gamma_j$, $E(\tilde{Z}_t \tilde{Z}_{t-j}) = 0$ e $j > 0$, obtém-se a função de autocorrelação (*fac*) pela expressão:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p}. \quad (2.10)$$

Dividindo ambos os lados por $\gamma_0 = Var(Z_t)$, obtém-se a função de autocorrelação:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}. \quad (2.11)$$

Considerando $j = 1, 2, \dots, p$ na equação (2.9), sabendo que $\rho_0 = 1$, obtemos as equações de Yule-Walker:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1}$$

$$\begin{aligned}
\rho_2 &= \phi_1\rho_1 + \phi_2 + \dots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p.
\end{aligned} \tag{2.12}$$

Os coeficientes $\phi_2, \phi_3, \dots, \phi_p$ do modelo $AR(p)$ podem ser estimados utilizando as equações (2.10) e substituindo as *fac* ρ_j por suas estimativas.

2.2.3 Função de autocorrelação parcial

De acordo com Box et al. (1994), a função que ajuda a identificar a ordem do processo autoregressivo que se adapta a série temporal é a função de autocorrelação parcial (*facp*).

Seja ϕ_{kj} o j -ésimo coeficiente do modelo $AR(k)$ e o ϕ_{kk} o último coeficiente. Dado as equações de Yule-Walker:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, 3, \dots, k. \tag{2.13}$$

Elas podem ser escritas na forma de um sistema matricial, como:

$$\begin{bmatrix}
1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\
\rho_1 & 1 & \rho_2 & \dots & \rho_{k-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1
\end{bmatrix}
\begin{bmatrix}
\phi_{k1} \\
\phi_{k2} \\
\vdots \\
\phi_{kk}
\end{bmatrix}
=
\begin{bmatrix}
\rho_1 \\
\rho_2 \\
\vdots \\
\rho_k
\end{bmatrix} \tag{2.14}$$

A solução geral para as equações para $k = 1, 2, 3, \dots$ pode ser dada por:

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|}, \tag{2.15}$$

em que ϕ_{kk} é a função de autocorrelação parcial, P_k é a matriz de autocorrelação e P_k^* é a matriz de autocorrelação com a última coluna substituída pelo vetor de autocorrelação.

Um processo auto regressivo de ordem p , tem uma função *facp* $\phi_{kk} \neq 0$, para $k \leq p$ e $\phi_{kk} = 0$ para $k > p$. Então no processo $AR(p)$, tem-se:

$$Var(\hat{\phi}_{jj}) \simeq \frac{1}{N}, \quad j > p \tag{2.16}$$

Sob a hipótese de que o processo é $AR(p)$ e para N suficientemente grande, a função

f_{acp} possui uma distribuição aproximadamente normal com média zero e variância de acordo com (2.14), (MORETTIN; TOLOI, 2006).

Com isso, ϕ_{kk} será significativamente diferente de zero se:

$$|\hat{\phi}_{kk}| > \frac{2}{\sqrt{N}}, \quad j > p \quad (2.17)$$

2.3 Metodologia Box & Jenkins

Uma série temporal é um conjunto de observações feitas e armazenadas ao longo do tempo que possuem tendência serial. São observações que podem ser contínuas no tempo ou discretas ou aquelas que são feitas em um intervalo de tempo específico. Os modelos utilizados para descrever séries temporais são processos controlados por leis probabilísticas chamados processo estocástico, definido abaixo: (MORETTIN; TOLOI, 2006).

Definição 3. *Seja T um conjunto arbitrário. Um processo estocástico é uma família $\{Z = Z(t), t \in T\}$, tal que, para cada $t \in T$, $Z(t)$ é uma variável aleatória.*

Na Figura 5, percebe-se que para cada $t \in T$, tem-se uma variável aleatória $Z(t, w)$, com uma distribuição de probabilidade. Existe a possibilidade que se tenha funções densidades de probabilidades distintas em tempos distintos. No entanto, a situação mais comum é aquela em que as funções de densidades $Z(t, w)$ são as mesmas, para todo $t \in T$.

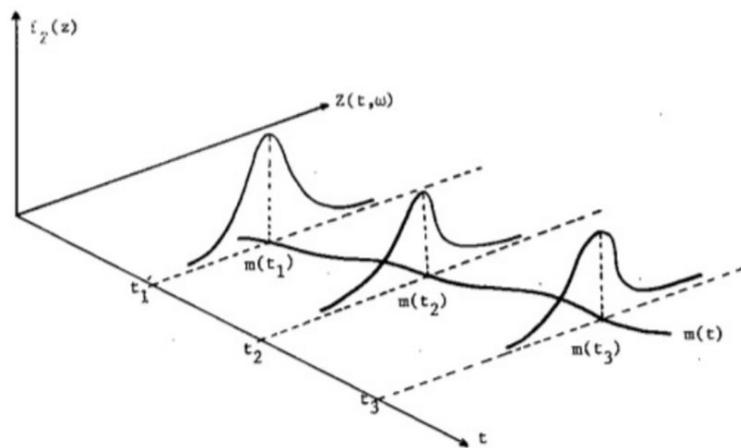


Figura 5: Processo estocástico interpretado como uma família de variáveis, Morettin e Tolo (2006).

O modelo ARIMA, também conhecido como metodologia Box & Jenkins, é composto por um processo autorregressivo (AR) que utiliza termos defasados da série temporal

como variável explicativa. Após aplicação de testes, se houver indicação do uso de variável defasado uma vez, o modelo é chamado de autorregressivo de primeira ordem, defasado duas vezes, de segunda ordem, e assim por diante, (GUJARATI; PORTER, 2011).

O modelo tem como princípio que a série temporal seja gerada por um processo estocástico cuja natureza pode ser representada através de um modelo. A notação empregada para designação do modelo é $ARIMA(p, d, q)$, em que p representa o número de parâmetros autorregressivos, d o número de diferenciações para que a série se torne estacionária e q o número de parâmetros de médias móveis, ou seja, uma combinação linear dos termos que ocorrem atualmente e em momentos do passado.

O modelo ARMA é um caso particular do modelo ARIMA no caso em que $d = 0$. Neste caso a série é estacionária que se subdivide em dois modelos, AR (autorregressivo) e MA (médias móveis). Os modelos estacionários, partem de que Z_t seja um processo com média zero e variância σ^2 , (MORETTIN; TOLOI, 2006).

Em situações em que a série a ser analisada não apresente estacionalidade, aplicam-se diferenças até que a nova variável seja estacionária; se a série indicar estacionalidade após aplicação da primeira diferença, então a série é integrada de ordem $d = 1$. Em caso negativo, aplica-se uma segunda diferença à nova variável e a série passa a ser de ordem $d = 2$, e assim por diante.

Definição 4. *Um processo Y_t , quando estacionário, é conhecido como auto-regressivo de ordem p , se:*

$$Y_t = \mu + \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t, \quad (2.18)$$

em que $\epsilon_t \sim N(0, \sigma_{\epsilon_t}^2)$.

A estimativa da variável Y_t depende de uma combinação linear de p termos da série observada, incluindo o termo aleatório ϵ_t de ruído branco (erros de estimação não correlacionados com distribuição normal, média zero e variância constante). Os coeficientes ϕ_i são os parâmetros que ponderam os valores de Y_t do instante imediatamente anterior $t - 1$ até o mais distante $t - p$, sendo obtidos através de técnicas de minimização do erro.

O modelo $MA(q)$ assume que a série é gerada através de uma combinação linear de q sinais de ruídos ϵ_{t-1} , aleatórios e independentes entre si, definido como:

Definição 5. *Um processo Y_t , quando estacionário, é conhecido como médias móveis de ordem q , se:*

$$Y_t = - \sum_{i=0}^q \theta_i \epsilon_{t-i} \quad (2.19)$$

em que $\epsilon_t \sim N(0, \sigma_{\epsilon_t}^2)$.

A combinação dos modelos $AR(p)$ e $MA(q)$, quando o processo é estacionário, dá origem ao modelo $ARMA(p, q)$, no qual

$$Y_t = \mu + \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{i=0}^q \theta_i \epsilon_{t-i}, \quad (2.20)$$

em que $\epsilon_t \sim N(0, \sigma_{\epsilon_t}^2)$. Com combinações do modelo AR com MA, chamado modelo $ARMA(p, q)$, obtém-se representações mais relevantes com um número menor de parâmetros. Na proposta para modelos não estacionários, com os modelos $ARIMA(p, d, q)$, temos uma série com valores finitos para realizar as diferenças d , para tornar a série estacionária, representado por $W_t = \nabla^d Y_t$. A Figura 5 mostra um organograma do modelo $ARIMA(1,1,1)$.

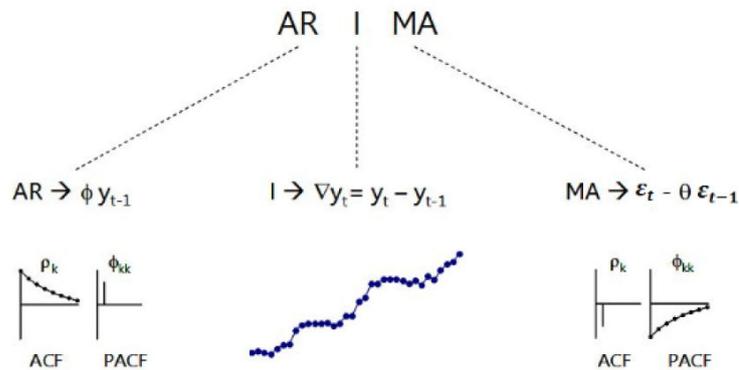


Figura 6: Modelo $ARIMA(1,1,1)$.

2.3.1 Etapas da modelagem

Na aplicação dos modelos de Box & Jenkins, com objetivo de fazer previsões em séries temporais, a primeira fase consiste em identificar o processo aleatório gerador dos dados para, em seguida, estimar os parâmetros que o caracterizam e verificar se as hipóteses do modelo são cumpridas. Caso contrário, uma nova fase de identificação deve ser considerada até que a verificação das hipóteses sejam finalmente satisfatória, permitindo assim, fazer previsões.

i) Identificação do modelo

Para a identificação do modelo ARIMA, são empregadas técnicas que possam identificar a estrutura do modelo de forma a permitir a identificação dos valores de p, d e q , que caracterizam o processo estocástico. Primeiramente, são feitas diferenciações na série original, quantas forem necessárias, para que a série resultante seja estacionária. O número de diferenciações d é aquele para que a função de autocorrelação amostral da série transformada decresça rapidamente para zero.

Para apontar o processo $AR(p)$, é feita análise da função de autocorrelação parcial ($facp$) e, para o processo $MA(q)$, é feita uma análise da função de autocorrelação (fac), para determinação dos parâmetros p e q , (MORETTIN; TOLOI, 2006). Vale salientar que, deve-se usar um modelo com melhor ajuste e menor número de parâmetros, princípio da parcimônia. Em outras palavras, dados dois modelos com ajustes iguais, é escolhido aquele com menor número de parâmetros. Na Figura 7 é apresentado um exemplo das funções fac e $facp$ de um modelo estacionário ARIMA(1,0,2).

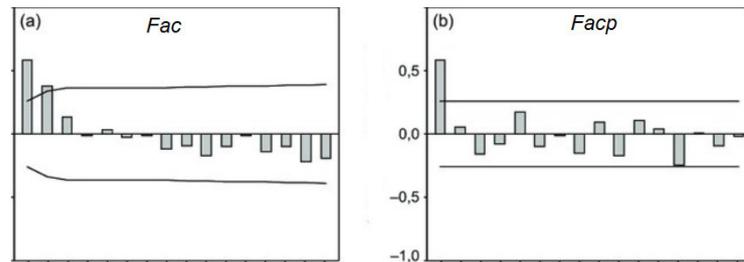


Figura 7: Funções fac e $facp$ de um modelo ARIMA(1,0,2).

ii) Estimação dos parâmetros

Esta etapa é executada por meio de *software* específico para análise de séries temporais. São diversos métodos para estimação dos parâmetros com resultados semelhantes, embora exista diferença quanto à eficiência computacional. De modo mais geral, usa-se algum procedimento iterativo de estimação de mínimos quadrados não linear ou utilizam-se de métodos de máxima verossimilhança, (MORETTIN; TOLOI, 2006).

iii) Verificação

Para que o uso do modelo seja eficaz na predição da variável dependente em estudo, o pesquisador necessita examinar previamente os pressupostos da regressão, bem como identificar as consequências da sua violação. Entre os pressupostos estão: normalidade, independência e média dos resíduos.

De acordo com Graybill (1994), ao fazer uma análise gráfica das distribuições dos resíduos, verifica-se a existência de homocedasticidade, ou seja, a hipótese de variância constante, que é aceita quando não há nenhuma tendência dos resíduos em relação ao valor estimado. A heterogeneidade das variâncias, ou violação da homogeneidade das variâncias, pode ser reduzida por intermédio da transformação de variáveis que não possuem distribuição normal.

Ao ajustar um modelo ARIMA, o objetivo é capturar toda estrutura de dependência presente na série. Então, os resíduos do modelo não devem apresentar nenhum tipo de dependência no tempo. Em um modelo ARIMA ajustado adequadamente, os erros de previsão (resíduos) devem ser uma sequência de ruídos brancos, ou seja, não devem existir auto-correlações e auto-correlações parciais significativas. Os testes estatísticos nessa etapa envolvem verificações da condição de estacionalidade, normalidade dos resíduos, análise de significância dos parâmetros estimados e critérios de seleção de modelo. Seguem os testes utilizados neste trabalho.

- **Teste de Box & Pierce**

Este é um teste para autocorrelações dos resíduos, foi sugerido por Box e Pierce (1970). Ele utiliza as k primeiras autocorrelações dos resíduos \hat{r}_k para verificar se os resíduos são um ruído branco, ou seja, independentes e identicamente distribuídos, com média zero e variância constante. Uma variação desse teste foi sugerida por Ljung e Box (1978) em que estatística do teste é dada por:

$$Q(k) = N(N - 2) \sum_{j=1}^k \frac{\hat{r}_j^2}{(N - j)}, \quad (2.21)$$

Se $Q > \chi_{k-p-q}^2$, rejeita-se as hipóteses de ruído branco. O valor k é o número de "lags", p a parte autorregressiva e q a parte das médias móveis.

- **Critério de Akaike**

O critério de Akaike (AIC) demonstra capacidade de ajuste do modelo estimado. O AIC avalia se o ajuste do modelo através do método máxima verossimilhança, determinando o logaritmos de razões comparando entre os modelos e penalizando os modelos pelos parâmetros utilizados, (GUJARATI; PORTER, 2011).

$$AIC = T \ln \hat{\sigma}_\epsilon^2 + \frac{T(1 + \frac{p}{T})}{1 - \frac{(p+2)}{T}}, \quad (2.22)$$

em que p representa a quantidade de parâmetros do modelo.

- **Teste Shapiro-Wilk**

O teste Shapiro-wilk, proposto em 1965, é baseado na estatística de teste W dada por:

$$W = \frac{b^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (2.23)$$

em que x_i são os valores da amostra ordenados, \bar{x} a média e b é determinada da seguinte forma:

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{(n-i+1)}(x_{(n-i+1)} - x_i) & \text{para } n \text{ ímpar;} \\ \sum_{i=1}^{(n+1)/2} a_{(n-i+1)}(x_{(n-i+1)} - x_i) & \text{para } n \text{ par,} \end{cases}$$

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, esses valores são tabelados.

- iv*) **Previsões**

Um dos principais objetivos da análise de séries temporais é prever um valor Z_{t+h} , $h \geq 1$, supondo que temos as observações $\dots, Z_{t-2}, Z_{t-1}, Z_t$, até o instante t , que é chamado origem das previsões. A notação mais utilizada é denotada como $\hat{Z}_t(h)$, ou seja, previsão de origem t e horizonte h .

Essas previsões também podem ser obtidas por meio de intervalos de confiança para os Z_{t+1}, Z_{t+2} , etc. No entanto, a variância aumenta com h , logo as amplitudes desses intervalos aumentam à medida que se afasta da origem t , caracterizando o aumento da incerteza das previsões à medida que h cresce.

Neste trabalho será utilizada uma técnica para avaliar as previsões, que consiste na retirada de valores reais da série, fazer o ajuste com ausência destes valores para fazer as previsões e compará-las com os valores reais. Verificando assim se os valores reais estão dentro do intervalo de confiança gerados pela previsão.

3 Materiais e Métodos

Nesta seção são descritas as séries analisadas, bem como os procedimentos para análises.

Os dados dos números de suicídios por sexo no estado de Minas Gerais, entre os anos de 1992 e 2019, foram obtidos no banco de dados do Instituto de Pesquisas Econômicas Aplicadas (IPEA, 2020), vinculado ao Ministério da Economia. Os trabalhos do IPEA estão disponíveis gratuitamente por meio de publicações, seminários e um canal de TV. Os dados relativos à população por sexo do estado de Minas Gerais, também entre os anos 1992 e 2019, foram obtidos por meio de estimativas feitas pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2010). O IBGE constitui o principal provedor de dados e informações do país, que atende às necessidades dos mais diversos segmentos da sociedade civil.

Para padronizar os dados por sexo foi utilizada a taxa específica de suicídios (TES) por dez mil habitantes, de acordo com a equação (3.1):

$${}_nTES_x = \frac{{}_nS_x}{{}_nP_x} \cdot 10.000, \quad (3.1)$$

em que ${}_nS_x$ é o número de suicídios por sexo e ${}_nP_x$ a população por sexo no estado de Minas Gerais.

Inicialmente foi considerada uma análise descritiva das séries com construção dos gráficos para observar indícios de tendência e sazonalidade. Verificou-se a necessidade de transformação dos dados, por meio do gráfico amplitude versus média.

Foram aplicados os testes de Dickey Fuller Aumentado (ADF) e Kruskal-Wallis para tendência e sazonalidade, em seguida analisadas as funções de autocorrelação (fac) e autocorrelação parcial ($facp$) na série livre de tendência para estimar a ordem do modelo. Após o ajuste dos modelos adequados, foram analisados as funções fac e $facp$ dos resíduos e aplicado o teste de Ljung-Box para verificar se os resíduos são ruído branco.

A princípio foram retiradas as taxas específicas de suicídios dos anos 2017, 2018 e

2019 e feitas previsões com os melhores modelos ajustados, escolhido de acordo com o critério de Akaike (AIC), para esses anos. Com objetivo de fazer comparações das taxas específicas reais com os valores previstos pelo modelo melhor ajustado. Após essa análise preliminar, foi incluído novamente às séries os valores retirados, ajustado novamente os modelos e feitas previsões para os anos 2020, 2021 e 2022. Por fim, um diagnóstico gráfico do modelo ajustado.

As análises foram feitas com utilização do *software* R (*Core Team*, 2018), com emprego dos pacotes de ferramentas *forecast* - Hyndman e Khandakar (2008), *urca* - Pfaff (2008), *FitAR* - McLeod e Zhang (2008), *runstats* - Karas e Urbanek (2019), *randtests* - Caeiro e Mateus (2014) e *lmtest* - Zeileis e Hothorn (2002). Também foi utilizado o *software* GRETL, Andrade (2013).

4 Resultados e Discussões

Nesta seção serão apresentados os modelos de ajustes obtidos com aplicação da metodologia de Box et al. (1994) com objetivo de fazer previsões para as taxas específicas de suicídios por sexo no estados de Minas Gerais, a fim de que ações sociais possam ser direcionadas a essa temática.

4.1 Identificação dos modelos

Nessa proposta foram trabalhados os dados das taxas específica de suicídios na população masculina e feminina no estado de Minas gerais entre os anos de 1992 e 2019, obtidos no banco de dados do Instituto de Pesquisas Econômicas Aplicadas (IPEA), vinculado ao Ministério da Economia do Governo Federal. Na Figura 8, são mostradas as séries anuais das taxas específicas de suicídios para a população feminina (em vermelho) e para população masculina (em azul) no estado de Minas Gerais no período citado.

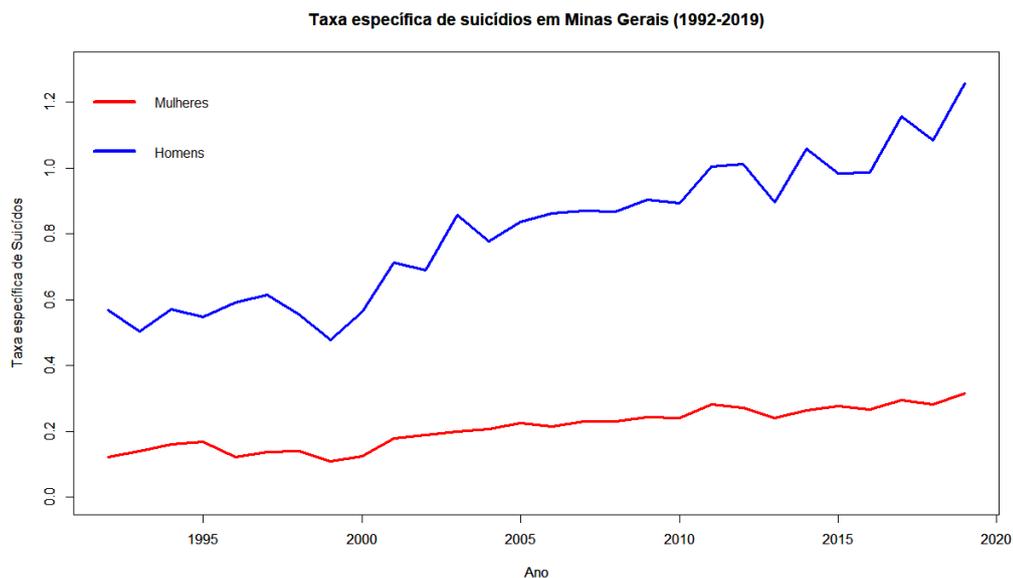


Figura 8: Séries de suicídios por sexo no estado de Minas Gerais (1992-2019)

A análise da série em estudo foi iniciada a partir de uma inspeção visual do seu gráfico mostrado na Figura 8, que, a princípio, mostram indícios da presença de tendência ao longo do tempo. Percebe-se também a informação dada por Filho e Zerbini (2016), de que, apesar das tentativas de suicídios serem mais significativas entre as mulheres, os homens provocam autoextermínio de forma mais efetiva.

Para verificar a necessidade de transformação dos dados nas séries, foram feitos os gráficos amplitude versus média mostrado na Figura 9. Constatou-se a não necessidade de transformação dos dados, já que as médias não é diretamente proporcional as amplitudes. Essa informação pode ser observada com análise do gráfico na Figura 9, em que as séries foram divididas em subgrupos de cinco observações consecutivas e para cada grupo são calculadas as médias e as amplitudes, com objetivo de verificar visualmente a necessidade ou não da transformação logarítmica na série. Nota-se que a amplitude não é diretamente proporcional à média.

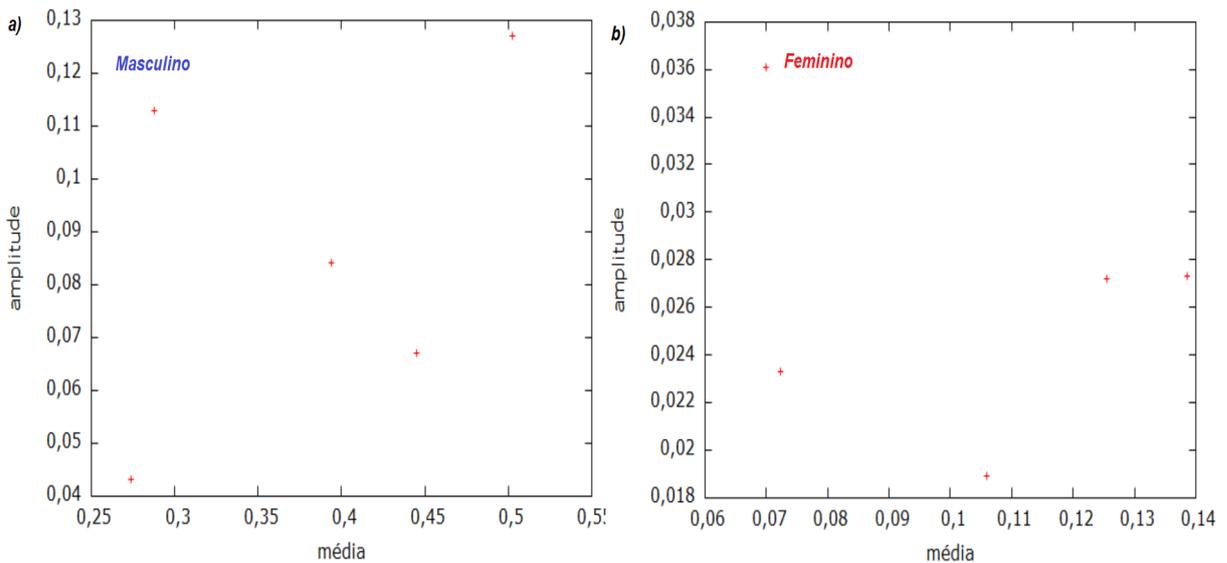


Figura 9: Gráfico Amplitude x Média em a) Série Masculina e b) Série feminina

No estudo da análise de séries temporais é necessário que a série seja estacionária, pois séries não estacionárias podem fornecer resultados preditivos viesados. Logo, as séries foram submetidas ao teste de raiz unitária Dickey Fuller aumentado (ADF) para verificar a existência ou não de estacionalidade em nível. Na Tabela 1, são mostrados os resultados dos testes comprovando que as séries não são estacionárias, ou seja, o número de suicídios por 10.000 habitantes tem aumentado, ao longo dos anos, para as duas populações (feminino e masculino).

Tabela 1: Teste Dickey Fuller Aumentado (Série suicídios em MG)

Variável	p-valor	Estacionária
N° de suicídios F	0,2042	Não
N° de suicídios M	0,0763	Não

Para transformar uma série não estacionária em estacionária, aplica-se a ela diferenças até que a série apresente estacionalidade. O número de diferenças necessárias para transformá-la em estacionária será a componente de integração (d) do modelo ARIMA(p, d, q). Após a aplicação da primeira diferença, percebe-se pelo gráfico na Figura 10, que as séries diferenciadas apresentam estacionalidade.

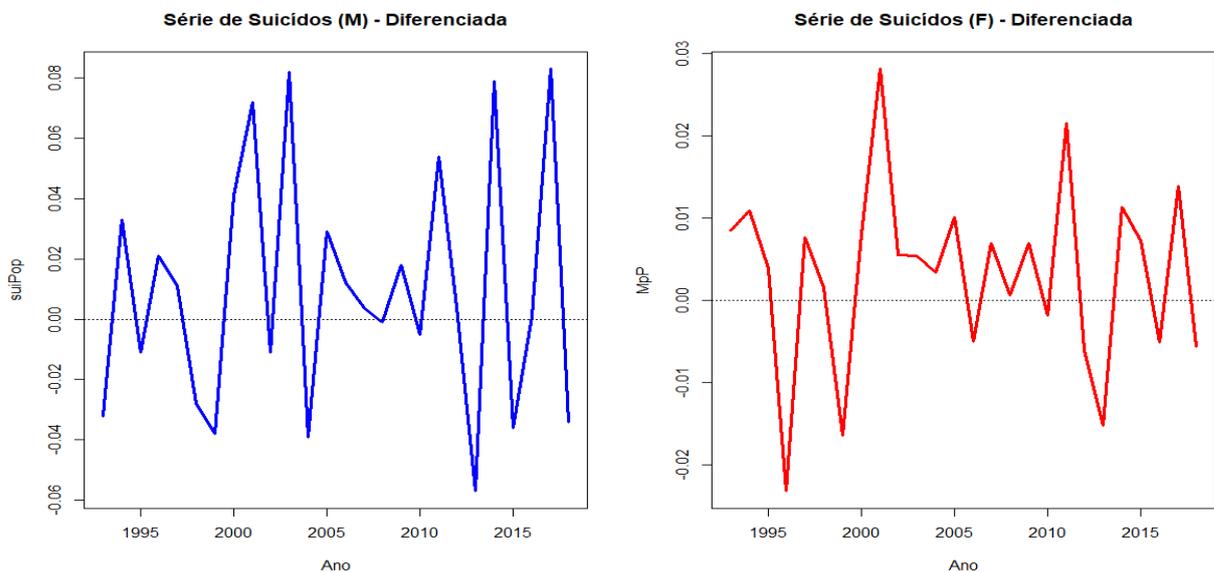


Figura 10: Série diferenciada da taxa específica de suicídios em Minas Gerais

Para comprovar que após a primeira diferença as séries tornam-se estacionárias, foi aplicado novamente o teste ADF às séries diferenciadas. Na Tabela 2, são mostrados os resultados dos testes indicando que as séries diferenciadas são estacionárias, portanto as séries são integradas de ordem $d = 1$.

Tabela 2: Teste Dickey Fuller Aumentado (Série diferenciada)

Variável	p-valor	Estacionária
N° de suicídios F	0,0044	Sim
N° de suicídios M	0,0010	Sim

A próxima verificação realizada foi a presença ou não de sazonalidade, para isso foi aplicado o teste Kruskal Wallis, mostrado na Tabela 3, ao nível de 5% de significância, concluímos que temos evidências estatísticas de que as séries não possuem sazonalidade.

Tabela 3: Teste Kruskal-Wallis (Séries de suicídios em MG)

Variável	p-valor	Sazonalidade
N° de suicídios F	0,4631	Não
N° de suicídios M	0,4672	Não

As observações feitas nas funções de autocorrelação e autocorrelação parcial ajudam na identificação da escolha das componentes do modelo ARIMA. De acordo com Morettin e Tolo (2006), em uma série estacionária, a função *fac* ajuda a identificar o parâmetro do modelo $MA(q)$, enquanto a função *facp* ajuda na identificação do parâmetro do modelo $AR(p)$. Na Figura 11, são mostradas as funções autocorrelação e autocorrelação parcial das séries de suicídios masculinos e femininos em Minas gerais.

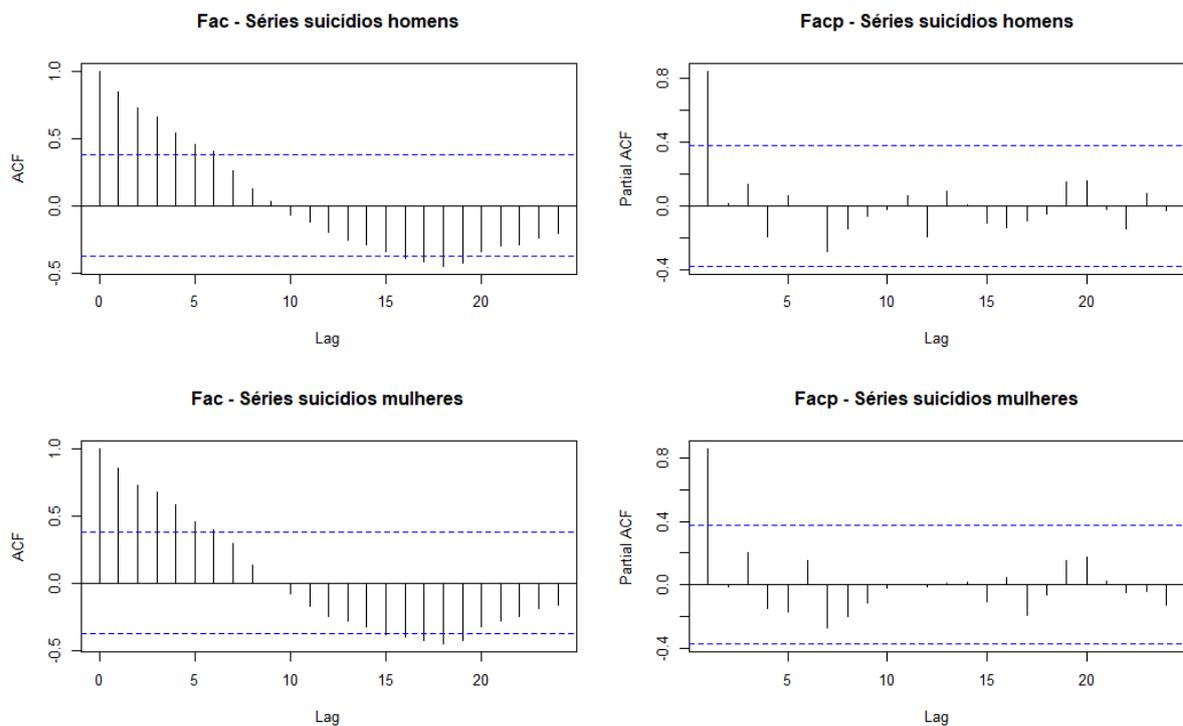


Figura 11: Funções de autocorrelação e autocorrelação parcial das séries de Suicídios

As funções de autocorrelação, Figura 11, mostram a correlação existente no tempo com a mesma série defasada em relação ao tempo inicial, ou seja, existe uma dependência temporal na taxa específica de suicídios ao longo do tempo. As funções de autocorrelação sugerem um modelo $ARMA(p, q)$ pois apresentam extensão infinita, a qual decai de acordo com exponenciais e/ou senoides amortecidas após o "lag" $p - q$, Morettin e Tolo (2006). Ainda de acordo com os autores, uma função de autocorrelação parcial de um processo $ARMA(p, q)$ se comporta de maneira similar a *fac*.

4.2 Estimação dos modelos

Afim de ajustar um modelo ARIMA, é preciso determinar as ordens p , d e q que fornecem o melhor ajuste possível do modelo. Com apoio do *software* R e o *software* GRETL, foram efetuados testes, supondo as ordens do modelo, afim de identificar os que apresentassem maior grau de significância.

Para verificar a ordem do modelo a ser ajustado, foram analisados os gráficos das funções de autocorrelação e autocorrelação parcial das séries diferenciadas, que indicaram os primeiros modelos a serem experimentados. Na Figura 12, são mostradas as funções de autocorrelação e autocorrelação parcial das séries diferenciadas.

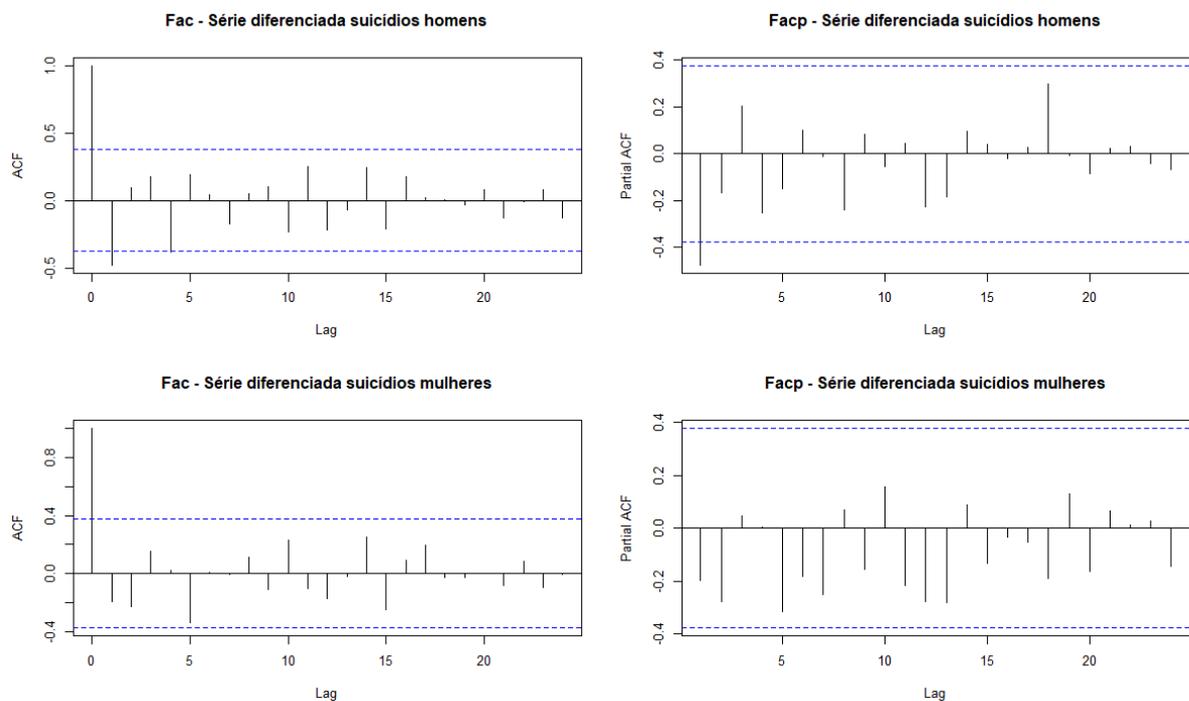


Figura 12: Funções de autocorrelação e autocorrelação parcial das séries diferenciadas

Com as análises dos correlogramas na Figura 12, foram testados os modelos ARIMA (1,1,2), ARIMA (1,1,1) e ARIMA (1,1,5) para a série masculina e ARIMA (0,1,0), ARIMA (0,1,1) e ARIMA (1,1,1) para a série feminina. Os critérios AIC para cada modelo são apresentados na Tabela 4.

Tabela 4: Critério Akaike para os modelos estimados

Sexo	Modelos	AIC
Masculino	ARIMA(1,1,5)	-59,58
	ARIMA(1,1,2)	-55,18
	ARIMA(1,1,1)	-57,63
Feminino	ARIMA(0,1,1)	-126,45
	ARIMA(0,1,0)	-125,20
	ARIMA(1,1,1)	-128,35

Como já havia sido testado anteriormente, com aplicação do teste ADF, sabe-se que as séries são estacionárias em primeira diferença. Portanto, todos os modelos foram rodados com $d = 1$.

Analisando a Tabela 4, observa-se a partir do critério de AIC, que os melhores resultados foram o modelo ARIMA(1,1,5) para a série masculina e ARIMA (1,1,1) para a série feminina. Na Tabela 5, são mostrados os coeficientes, o erro padrão e o p-valor para cada modelo escolhido.

Tabela 5: Parâmetros dos modelos ajustados

Modelo	Coefficientes	Erro padrão	p-valor
Masculino	$\phi_1 = 0,2519$	0,1351	0,00021
ARIMA(1,1,5)	$const = 0,0255$	$9,2 \cdot 10^{-4}$	$6,22 \cdot 10^{-16}$
	$\theta_1 = -1,2398$	0,1359	$7,61 \cdot 10^{-20}$
	$\theta_5 = 0,2534$	0,1287	0,049
Feminino	$\phi_1 = 0,5017$	0,1716	0,0043
ARIMA(1,1,1)	$const = 0,007$	$7,8 \cdot 10^{-4}$	$3,01 \cdot 10^{-19}$
	$\theta_1 = -1,000$	0,1016	$7,63 \cdot 10^{-23}$

Os modelos para previsão podem ser escritos da seguinte forma:

$$\hat{Z}_t = 0,0255 + 0,2519Z_{t-1} + 0,2534a_{t-5} - 1,2398a_{t-1} + a_t, \quad (4.1)$$

para o modelo ARIMA(1,1,5) da taxa específica de suicídios da população masculina, e

$$\hat{Z}_t = 0,007 + 0,5017Z_{t-1} - 1,000a_{t-1} + a_t, \quad (4.2)$$

para o modelo ARIMA(1,1,1) da taxa específica de suicídios da população feminina.

Na Figura 13, são mostrados os gráficos que fazem a representação do poder preditivo de cada série, ao fazer uma comparação das taxas específicas de suicídios com os valores previstos pelos modelos ajustados. As previsões efetuadas através do modelo ARIMA (1,1,5) para a série masculina e ARIMA (1,1,1) para a série feminina, se aproximam

muito da taxa específica de suicídios real, gerando uma conclusão da eficiência do modelo para gerar previsões de suicídios no estado de Minas Gerais.

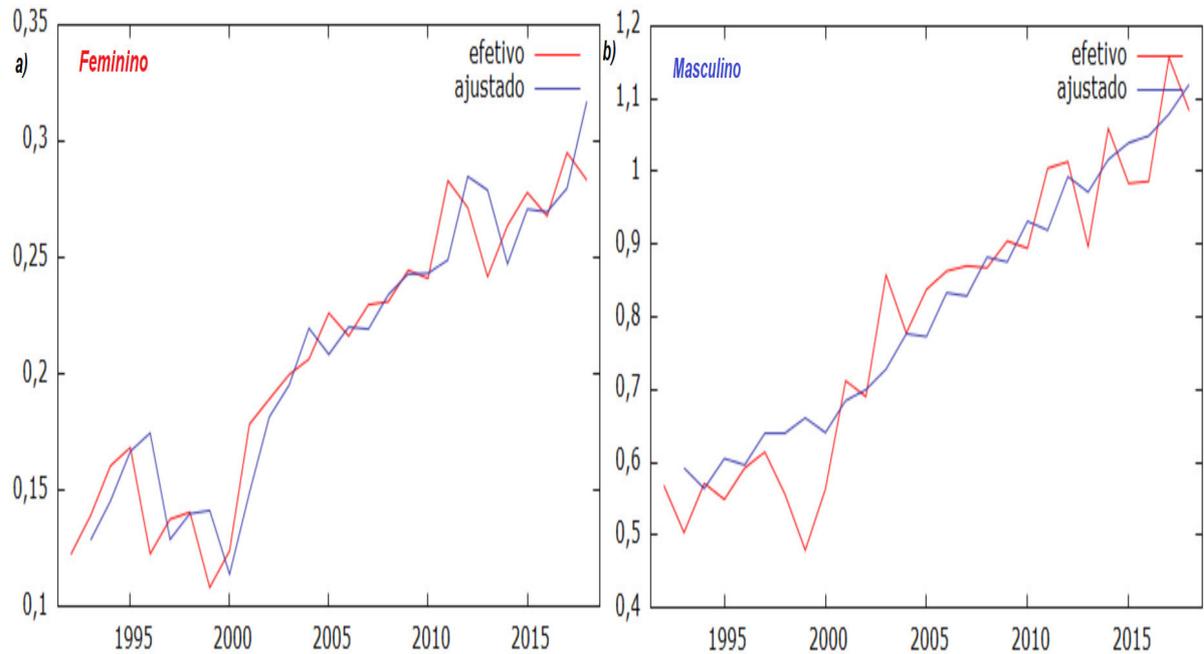


Figura 13: Série de suicídios (efetivo x ajustado)

4.3 Verificação dos modelos

O próximo passo da análise de série temporal é verificar se o modelo ajusta-se de forma adequada aos dados, com um nível de significância pré estabelecido. Na Figura 14, são mostrados os correlogramas dos resíduos para as séries de suicídios masculino e feminino. Observa-se que os erros não são correlacionados, indicando um bom ajuste do modelo aos dados.

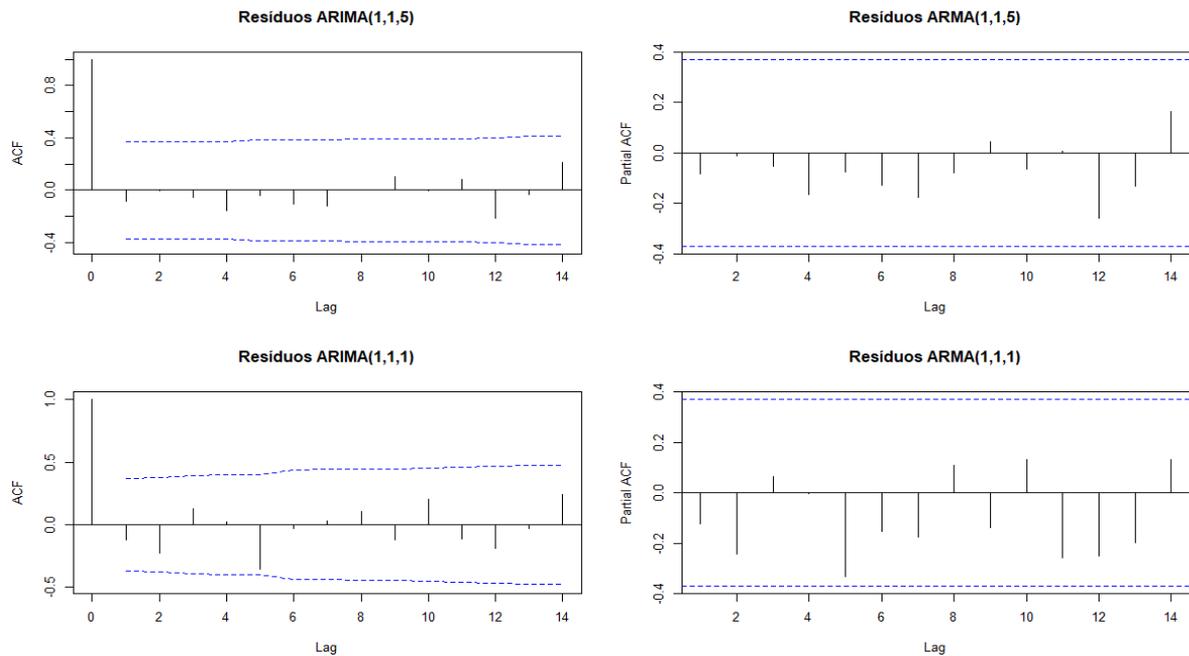


Figura 14: Funções de autocorrelação e autocorrelação parcial dos resíduos

Considerando que o principal interesse no ajuste são valores futuros, então o modelo mais adequado é aquele que apresenta as melhores previsões. Na Tabela 6, são apresentados os resultados do teste Ljung-Box em que mostram que as séries ajustadas não apresentam correlação nos resíduos, comprovando o que já foi observado com as funções de autocorrelação, Figura 14.

Tabela 6: Teste Ljung-Box (Resíduos)

Variável	p-valor	Correlação
N° de suicídios F	0,3770	Não
N° de suicídios M	0,6153	Não

Outra forma de verificar a qualidade de ajuste do modelo é fazer uma análise do gráfico de normalidade dos resíduos. Na Figura 15, podemos verificar que os valores estimados estão dispostos em uma linha diagonal indicando uma linearidade satisfatória, a hipóteses de normalidade foram aceitas com base no teste Shapiro-Wilk, em que foram obtidos p-valores $p = 0,1785$ e $p = 0,4263$ para as séries de suicídios masculinos e femininos respectivamente.

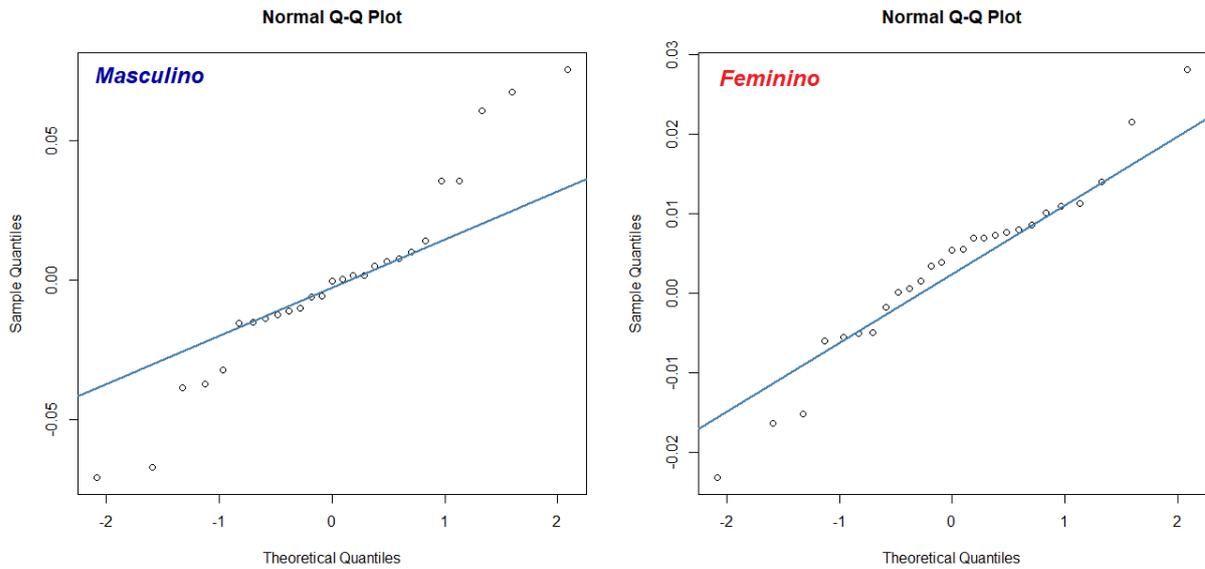


Figura 15: Q-Q plot das séries de suicídios masculino e feminino em Minas Gerais

Considerando, no geral, as estatísticas apresentadas pelo ajuste dos modelos aos dados observados de suicídios em Minas Gerais, verifica-se que esses são eficientes para a realização das previsões.

4.4 Previsões

A busca por valores previstos, para uma meta previamente estabelecida, é uma das principais finalidades da utilização de modelos de séries temporais. Para o trabalho proposto, as previsões podem ser utilizadas para nortear políticas públicas na área da saúde mental da população em Minas Gerais.

As séries em estudo apresentam valores das taxas específicas de suicídios entre os anos de 1992 e 2019. Para verificar se os modelos apresentam previsões dentro de um intervalo de confiança, foram retiradas as três últimas observações (2017, 2018 e 2019), ajustados os modelos propostos e feitas as previsões para essas observações retiradas. Na Tabela 7, são apresentadas as previsões para as duas séries por intervalo, com 95% de confiança e os respectivos valores reais.

Tabela 7: Previsões das séries sem as últimas observações

Séries	Ano	Intervalo	Valor real (TES)
Prev (M)	2017	0,9406 : 1,1919	1,1566
	2018	0,8347 : 1,1326	1,0836
	2019	0,8140 : 1,1952	1,2566
Prev (F)	2017	0,2245 : 0,3123	0,2950
	2018	0,2084 : 0,3284	0,2831
	2019	0,1958 : 0,3410	0,3156

Como podemos perceber, a maior parte dos valores das taxas específicas de suicídios reais estão dentro do intervalo de confiança previsto pelos modelos, isso sugere o bom ajuste dos modelos às séries analisadas. Após essa verificação foram feitas previsões para os anos de 2020, 2021 e 2022 com os modelos propostos e com a série completa, ou seja, incluindo os taxas reais de 2017, 2018 e 2019. Na Tabela 8, são apresentados os valores previstos para cada taxa específica e também por intervalo, ao nível de 5% de significância.

Tabela 8: Previsões futuras das séries

Séries	Ano	Intervalo	Previsão (TES)
Prev (M)	2020	1,1536 : 1,4106	1,2821
	2021	1,1415 : 1,4466	1,2940
	2022	1,1383 : 1,5328	1,3382
Prev (F)	2020	0,2741 : 0,3553	0,3148
	2021	0,2731 : 0,3699	0,3215
	2022	0,2731 : 0,3833	0,3282

De acordo com a Tabela 8, nota-se que as previsões para a taxa de suicídios da população masculina e feminina apontam crescimento médio anual de 2,1% e 1,3% respectivamente, com crescimento mais acentuado de 2021 para 2022 na população masculina. Portanto, considerando o aumento de ocorrências de suicídios para as duas populações, aspecto também observado por Ribeiro et al. (2018) na cidade de Uberaba em Minas Gerais, salienta-se a importância de políticas públicas serem direcionadas com mais atenção para essa temática, de forma a diminuir sofrimentos com a perda de vidas humanas. Na Figura 16 são mostradas graficamente essas conclusões.

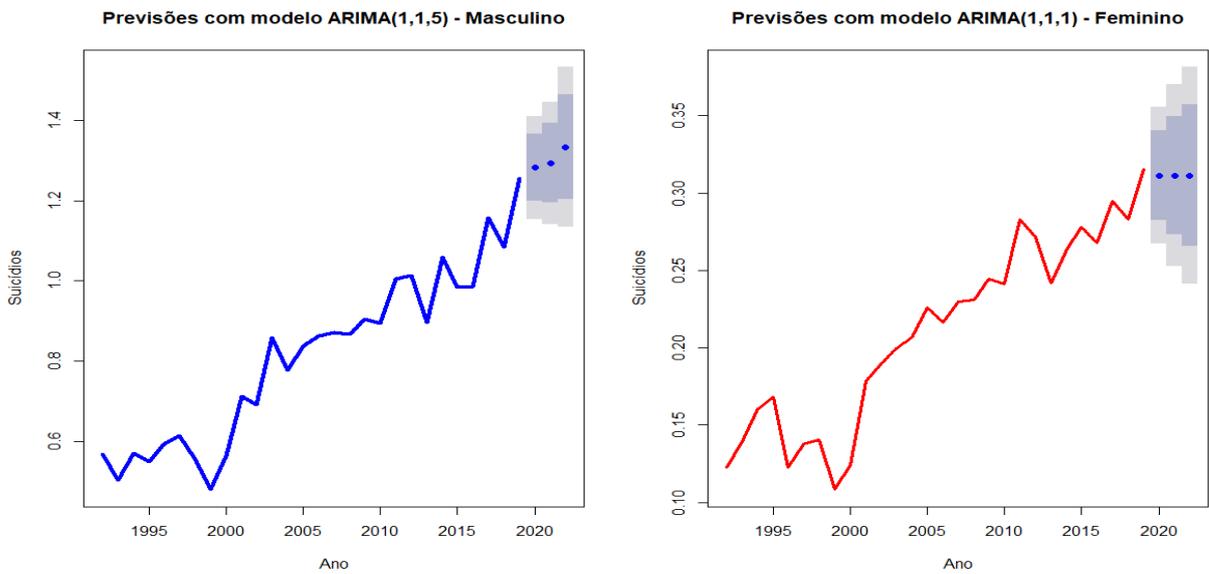


Figura 16: Gráficos com previsões para as séries de suicídios por sexo em Minas Gerais

A função *tsdiag* do *software* R proporciona um diagnóstico gráfico dos modelos de ajuste. Portanto, antes de encerrar as análises dos dados, na Figura 17, são apresentados os gráficos do resumo diagnóstico dos modelos ajustados.

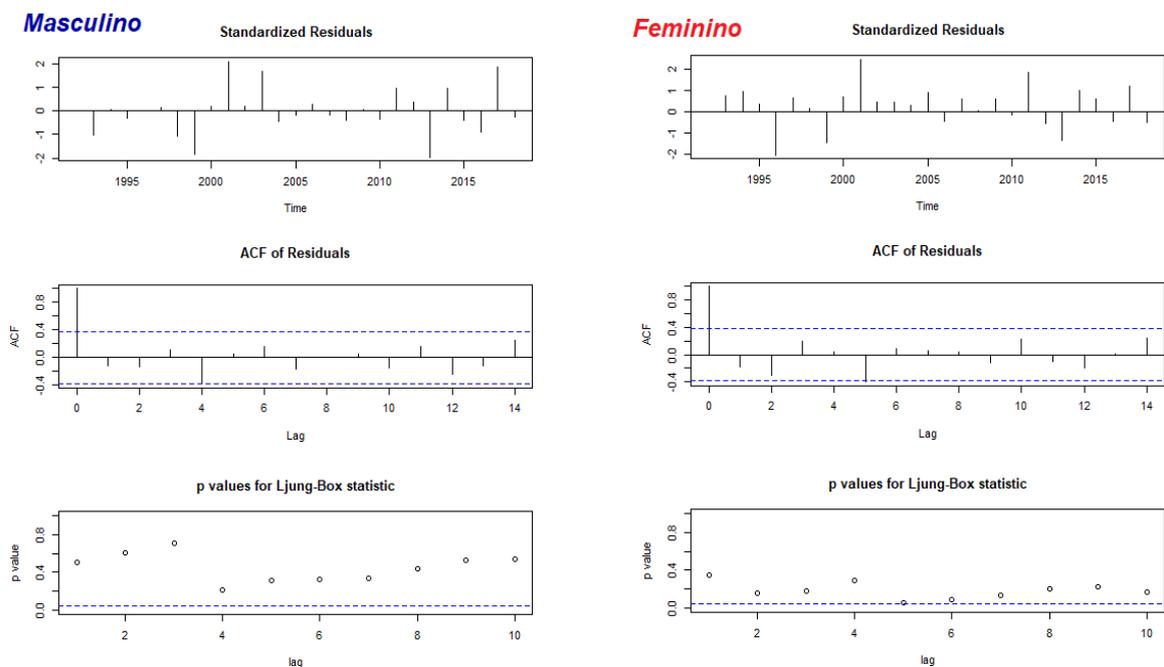


Figura 17: Diagnósticos dos modelos ajustados

5 Considerações Finais

Esse estudo aplicou a metodologia Box et al. (1994) para ajustar modelos ARIMA às séries da taxa específicas de suicídios de homens e mulheres no estado de Minas Gerais entre os anos de 1992 e 2019, e assim, obter previsões para os anos 2020, 2021 e 2022. Foi verificado, de maneira geral, que os modelos de Box & Jenkins ajustaram-se bem aos dados a ponto de fazer previsões seguras com nível de significância de 5%. Com isso, políticas públicas podem ser direcionadas com base nessas análises para salvar vidas. Foi observada uma tendência histórica de crescimento para as duas séries de suicídios e conclui-se que permanece a tendência de crescimento nas taxas específicas de suicídios nas duas populações (masculina e feminina) para os anos 2020, 2021 e 2022. Pretende-se para trabalhos futuros, fazer uma análise múltipla de séries que possam estar correlacionadas com as taxas específicas de suicídios, para com isso, compreender possíveis causas do aumento ao longo do tempo.

Referências

- ANDRADE, C. H. C. Manual de introdução ao pacote econométrico gretl. 2013. Disponível em: <<https://s3.amazonaws.com/academia.edu.documents/39579087/Gretl.pdf>>. Acesso em: 05 març. 2020.
- ARAUJO, M. F. C. et al. Precipitação pluviométrica mensal no estado do rio de janeiro: sazonalidade e tendência. *Bioscience Journal*, Universidade Federal de Uberlândia, v. 25, n. 4, p. 90–100, 2009.
- BARBOSA, F. V.; IQUIAPAZA, R. A.; BRESSAN, A. A. Testes de não estacionariedade em séries financeiras com dados em painel: Uma síntese aplicada. In: *V Encontro Brasileiro de Finanças*. São Paulo: [s.n.], 2008.
- BOX, G. E. et al. *Time series analysis: forecasting and control*. 3. ed. New Jersey: Prentice Hall, 1994.
- BOX, G. E.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, Taylor & Francis Group, v. 65, n. 332, p. 1509–1526, 1970.
- BRANAS, C. C. et al. The impact of economic austerity and prosperity events on suicide in greece: a 30-year interrupted time-series analysis. *BMJ open*, British Medical Journal Publishing Group, v. 5, n. 1, 2015.
- CAEIRO, F.; MATEUS, A. *randtests: Testing randomness in R*. [S.l.], 2014. R package version 1.0. Disponível em: <<https://CRAN.R-project.org/package=randtests>>. Acesso em: 10 març.2020.
- CAMPOS, R. J. Previsão de séries temporais com aplicações a séries de consumo de energia elétrica. 2008. Disponível em: <<<https://repositorio.ufmg.br/bitstream/1843/BUOS-8CTETD/1/290m.pdf>>. Acesso em Març 24, 2020.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979.
- FILHO, M. C.; ZERBINI, T. Epidemiologia do suicídio no brasil entre os anos de 2000 e 2010. *Saúde, Ética & Justiça*, v. 21, n. 2, p. 45–51, 2016.
- FONSECA, J. S. d.; MARTINS, G. d. A.; TOLEDO, G. L. Estatística aplicada. *Atlas*, 1985.
- GONÇALVES, L. R.; GONÇALVES, E.; JÚNIOR, L. B. d. O. Determinantes espaciais e socioeconômicos do suicídio no brasil: uma abordagem regional. *Nova Economia*, SciELO Brasil, v. 21, n. 2, p. 281–316, 2011.

- GRAYBILL, F. A. *Regression analysis: concepts and applications*. [S.l.]: Duxbury Resource Center, 1994.
- GUJARATI, D. N.; PORTER, D. C. *Econometria Básica-5*. [S.l.]: Amgh Editora, 2011.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, v. 26, n. 3, p. 1–22, 2008. Disponível em : <<http://www.jstatsoft.org/article/view/v027i03>>. Acesso em: 10 març. 2020.
- IBGE. Instituto brasileiro de geografia e estatística. Censo Demográficos, 2010. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html>>. Acesso em: 05 abr. 2020.
- IPEA. Instituto de pesquisas econômicas e aplicadas. Tema: Saúde, 2020. Disponível em: <<http://www.ipeadata.gov.br/Default.aspx>>. Acesso em: 10 març. 2020.
- KARAS, M.; URBANEK, J. *runstats: Fast Computation of Running Statistics for Time Series*. [S.l.], 2019. R package version 1.1.0. Disponível em: <<https://CRAN.R-project.org/package=runstats>>. Acesso em: 10 març. 2020.
- LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. *Biometrika*, Oxford University Press, v. 65, n. 2, p. 297–303, 1978.
- MCLEOD, A. I.; ZHANG, Y. Improved subset autoregression: With R package. *Journal of Statistical Software*, v. 28, n. 2, 2008. Disponível em: <<http://www.jstatsoft.org/v28/i02/>>. Acesso em: 10 març. 2020.
- MORETTIN, P. A.; TOLOI, C. M. d. C. *Análise de Séries Temporais*. 2. ed. São Paulo: Edgard Blucher, 2006.
- OMS. Organização mundial da saúde. preventing suicide: A resource series. Genebra, 2019. Disponível em: <<https://www.who.int/publications-detail/preventing-suicide-a-resource-series>>. Acesso em: 05 abr. 2020.
- PFAFF, B. *Analysis of Integrated and Cointegrated Time Series with R*. Second. New York: Springer, 2008. ISBN 0-387-27960-1. Disponível em: <<http://www.pfaffikus.de>>. Acesso em: 05 març. 2020.
- RIBEIRO, N. M. et al. Análise da tendência temporal do suicídio e de sistemas de informações em saúde em relação às tentativas de suicídio. *Texto & Contexto-Enfermagem*, SciELO Brasil, v. 27, n. 2, 2018.
- SÁFADI, T. Uso de séries temporais na análise de vazão de água na represa de furnas. *Ciência e Agrotecnologia*, Scielo Brasil, v. 28, n. 1, p. 142–148, 2004.
- SILVEIRA, R. E.; SANTOS, S. A.; FERREIRA, L. A. Impacto da morbi-mortalidade e gastos com suicídio no brasil de 1998 a 2007. *Revista de Pesquisa Cuidado é Fundamental Online*, Rio de Janeiro, v. 4, n. 4, p. 3033–3042, 2012.
- ZEILEIS, A.; HOTHORN, T. Diagnostic checking in regression relationships. *R News*, v. 2, n. 3, p. 7–10, 2002. Disponível em : <<https://CRAN.R-project.org/doc/Rnews/>>. Acesso em: 15 març. 2020.