

Natan Moraes Vaz

**Impacto da COVID-19 no número de mortes no Brasil em 2020**

Niterói - RJ, Brasil

2021

Natan Moraes Vaz

**Impacto da COVID-19 no número de mortes no Brasil em 2020**

Trabalho de Conclusão de Curso apresentado para obtenção do grau de Bacharel em Estatística pela Universidade Federal Fluminense.

Orientadora: Prof. Dra. Márcia Marques de Carvalho

Niterói - RJ, Brasil

2021

Natan Moraes Vaz

**Impacto da COVID-19 no número de mortes no  
Brasil em 2020**

Monografia de Projeto Final de Graduação sob o título "Impacto da COVID-19 no número de mortes no Brasil em 2020", defendida por Natan Moraes Vaz e aprovada em 17 de setembro de 2021, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

Prof. Dra. Márcia Marques de Carvalho  
Departamento de Estatística - UFF

---

Prof. Dra. Ludmila da Silva Viana Jacobson  
Departamento de Estatística - UFF

---

Prof. Dr. Moisés Lima de Menezes  
Departamento de Estatística - UFF

Niterói - RJ, Brasil

2021

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

V393i Vaz, Natan Moraes  
Impacto da COVID-19 no número de mortes no Brasil em 2020 /  
Natan Moraes Vaz ; Márcia Marques de Carvalho, orientadora.  
Niterói, 2021.  
59 f.

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

1. COVID-19. 2. Brasil. 3. Séries temporais. 4. Box &  
Jenkins. 5. Produção intelectual. I. Carvalho, Márcia  
Marques de, orientadora. II. Universidade Federal Fluminense.  
Instituto de Matemática e Estatística. III. Título.

CDD -

## RESUMO

Em 2020 a COVID-19 chegou ao Brasil e, por conta dos danos causados à saúde e da facilidade de sua transmissão, logo se tornou destaque nas mídias e nas redes sociais. Desde então, a busca por uma melhor compreensão da doença tornou-se prioridade, visando a adoção de medidas que pudessem diminuir o contágio e o número de óbitos. Buscando uma melhor compreensão sobre o efeito desta doença no número de mortes no país, este trabalho propôs o uso de análise de séries temporais como ferramenta para verificar o comportamento da evolução do número de óbitos no Brasil, mês a mês, nos últimos 10 anos. Utilizando-se o Modelo de Regressão Linear e o Modelo de Box & Jenkins, foram realizadas previsões para 2020, sobre a quantidade de óbitos esperados no país caso não houvesse a pandemia de COVID-19, e então comparou-se os valores obtidos com os dados reais fornecidos pela organização de saúde. Sobre os resultados obtidos, e em relação aos dados utilizados, tem-se que o Modelo Box & Jenkins apresentou uma melhor capacidade de previsão do que o Modelo de Regressão Linear. Verificou-se que o aumento do número de mortes previstos para o Brasil em 2020 foi de 13,83%, o que pode representar o impacto da pandemia no país.

Palavras-chave: COVID-19, coronavírus, Brasil, séries temporais, Modelo de Regressão Linear e Box & Jenkins

## **DEDICATÓRIA**

Pensando nas pessoas que gostariam de um esclarecimento sobre o impacto do coronavírus no número de óbitos no Brasil em 2020, este trabalho visa entender o quanto a COVID-19 pode ter alterado o comportamento da mortalidade no país. Espero que esta pesquisa possa ajudar de alguma forma.

## **AGRADECIMENTOS**

Agradeço em primeiro lugar a Deus que vem me ajudando, me sustentando e me guardando em todos os desafios da vida, inclusive nos meus anos de estudo.

Agradeço a minha família e meus familiares que de forma direta ou indireta me ajudaram ao longo da minha caminhada acadêmica.

Agradeço aos amigos que fiz ao longo da Faculdade, que me apoiaram e me ajudaram em vários momentos difíceis da vida acadêmica (não citarei nomes, pois posso me esquecer de alguém e não ser justo com todos que me ajudaram).

Agradeço a todos os professores que me transmitiram os seus conhecimentos e experiências.

## SUMÁRIO

1. INTRODUÇÃO	13
1.1. Motivação	13
1.2. Objetivos Geral e Específicos	14
2. MATERIAIS E MÉTODOS	15
2.1. Base de dados	15
DATASUS	15
openDataSUS	15
CORONAVÍRUS//BRASIL Portal Coronavírus	15
Utilização dos dados	16
2.2. Processos estocásticos	16
Processos estacionários	17
Ruido branco	17
2.3. Séries temporais	18
Componentes de uma série temporal	18
Função de autocorrelação	19
Função de autocorrelação parcial	19
Correlograma	20
Teste da Raiz Unitária	20
2.4. Modelos Econométricos	21
Modelo de Regressão Linear	22
Modelo ARIMA	25
Modelo SARIMA	26
Metodologia Box & Jenkins	27
Identificação de um ARIMA	27
Estimação de um ARIMA	28



Análise de resíduos _____	28
2.5. Comparação entre os modelos _____	29
2.6. Comparação entre os órbitos observados e previstos para 2020 _____	34
2.7. Metodologia _____	34
3. RESULTADOS _____	35
3.1. Estatísticas descritivas _____	35
3.2. Características da Série Temporal _____	39
3.3. Modelo de Regressão Linear _____	40
3.4. Modelos Box & Jenkins _____	45
3.5. Escolha do modelo para fazer a previsão _____	53
3.6. Impacto da COVID-19 _____	55
4. CONCLUSÃO _____	57
Referências _____	58

## LISTA DE QUADROS

**Quadro 1:** Comportamento teórico da FAC e FACP para alguns modelos \_\_\_\_\_ 28

## LISTA DE FIGURAS

<b>Figura 1:</b> Exemplo de correlograma da FAC e FACP	20
<b>Figura 2:</b> Exemplo de correlograma da FAC dos resíduos.	29
<b>Figura 3:</b> Série temporal do número de óbitos por COVID-19 em 2020.	38
<b>Figura 4:</b> Série temporal do número de óbitos no Brasil de 2010 a 2019.	39
<b>Figura 5:</b> Série temporal com previsão dos anos de 2018 e 2019 MQO.	42
<b>Figura 6:</b> Gráfico de dispersão da FAC dos resíduos para o MQO.	44
<b>Figura 7:</b> Histograma comparado à normal dos resíduos MQO.	44
<b>Figura 8:</b> Primeira diferença da série temporal do número total mensal de óbitos no Brasil no período de janeiro/2010 a dezembro/2017.	46
<b>Figura 9:</b> FAC e FACP da 1ª diferença dos óbitos no Brasil no período de janeiro/2010 a dezembro/2017.	47
<b>Figura 10:</b> FAC e FACP da 1ª diferença sazonal do número de óbitos no Brasil no período de janeiro/2010 a dezembro/2017.	48
<b>Figura 11:</b> Série temporal com previsão dos anos de 2018 e 2019 SARIMA(0,1,1)(1,1,4).	50
<b>Figura 12:</b> Gráfico de dispersão da FAC dos resíduos para o SARIMA.	52
<b>Figura 13:</b> Histograma comparado à normal dos resíduos SARIMA.	52
<b>Figura 14:</b> Previsão para 2020 caso não houvesse a pandemia.	54

## LISTA DE TABELAS

<b>Tabela 1:</b> Dados DATASUS número de óbitos ano a ano de 2010 a 2019. _____	35
<b>Tabela 2:</b> Aumento percentual médio para cada mês. _____	36
<b>Tabela 3:</b> Aumento percentual de óbitos a cada ano. _____	36
<b>Tabela 4:</b> N° de óbitos preliminares para 2020 emitidos pelo openDataSUS _____	37
<b>Tabela 5:</b> N° de óbitos por COVID-19 em 2020 emitidos pelo Painel Coronavírus _____	38
<b>Tabela 6:</b> Estatística descritiva 2010 a 2019. _____	40
<b>Tabela 7:</b> Estatísticas de aderência para os modelos MQO. _____	41
<b>Tabela 8:</b> Parâmetros da tendência ( $\alpha$ ) e da sazonalidade ( $\beta$ ). _____	41
<b>Tabela 9:</b> Valores reais x valores previstos 2018 e 2019 MQO. _____	43
<b>Tabela 10:</b> Teste de normalidade dos resíduos MQO. _____	45
<b>Tabela 11:</b> Teste de normalidade dos dados da série temporal do número total mensal de óbitos no Brasil no período de janeiro/2010 a dezembro/2017. _____	47
<b>Tabela 12:</b> Estatísticas de aderência para os modelos SARIMA. _____	49
<b>Tabela 13:</b> Parâmetros estimados da modelagem SARIMA(0,1,1)(1,1,4). _____	49
<b>Tabela 14:</b> Valores reais x valores previstos 2018 e 2019 SARIMA(0,1,1)(1,1,4). _____	51
<b>Tabela 15:</b> Teste de Normalidade dos resíduos SARIMA. _____	53
<b>Tabela 16:</b> Estatísticas de aderência para o modelo selecionado. _____	53
<b>Tabela 17:</b> N° de óbitos esperados para 2020 caso não houvesse pandemia. _____	55
<b>Tabela 18:</b> Efetividade do modelo através do erro de previsão mês a mês para o Brasil em 2020. _____	55
<b>Tabela 19:</b> Impacto da COVID-19 no aumento do número de óbitos previstos para o Brasil em 2020. _____	56

# 1. INTRODUÇÃO

## 1.1. Motivação

O novo coronavírus (SARS-CoV-2), popularmente conhecido por COVID-19 (sigla que denomina a doença causada pelo vírus), vem causando óbitos desde seu surgimento, em dezembro de 2019, na cidade de Wuhan, capital da província de Hubei, na China [1,2,3]. Em pouco tempo, o vírus se espalhou e o país entrou em estado de epidemia. Porém o vírus não permaneceu em sua cidade de origem. Logo a doença se espalhou e começou a ocasionar mortes por todos os continentes. Então, no dia 11 de março de 2020 foi instaurado pela Organização Mundial de Saúde (OMS) o status de pandemia [4].

Como o vírus é de fácil disseminação, vários casos da doença começaram a aparecer no Brasil e muitas mortes surgiram por intermédio do seu contágio. É fato que a COVID-19 é uma doença contagiosa e letal, por isso houve a necessidade da criação de medidas para evitar o seu contágio e assim diminuir o número de óbitos, não só no Brasil, mas em todo o mundo. Então, diversas instruções começaram a ser passadas pela OMS aos países, como o uso do álcool gel (para higienização das mãos e superfícies) e o uso da máscara (para evitar que pessoas contaminadas propaguem o vírus, ao tossir ou espirrar, através da liberação de gotículas no ar) [5].

Existem nações que sofrem mais do que outras por conta de influências naturais (clima, temperatura, localização geográfica, umidade do ar etc.) e culturais (proximidade social, higienização, uso excessivo de tabaco etc.), o que dificulta o trabalho da OMS ao tomar medidas generalizadas, que funcionem para todos os países.

Desta forma, este trabalho visa estudar o comportamento do número de óbitos no Brasil em 2020 caso não houvesse a pandemia e comparar com os valores reais, para melhor entendimento do impacto da pandemia no país.

## **1.2. Objetivos Geral e Específicos**

O objetivo geral é avaliar o impacto da pandemia no número de óbitos total no Brasil em 2020, através dos dados disponibilizados pelos órgãos responsáveis por pesquisas sobre mortalidade no país.

Entre os objetivos específicos, encontram-se:

- Estimar o número de óbitos no ano de 2020 caso não houvesse a pandemia mês a mês;
- Comparar os valores esperados com os valores reais;

## **2. MATERIAIS E MÉTODOS**

### **2.1. Base de dados**

#### **DATASUS**

O Departamento de Informática do Sistema Único de Saúde (DATASUS) foi criado em 1991 com a finalidade de prover suporte de informática e sistemas de informação ao Sistema Único de Saúde (SUS), de forma que informações mais precisas pudessem ser utilizadas na realização de planejamentos, operação e controle.

Na atualidade o DATASUS é um provedor de software para as secretarias estaduais e municipais de saúde do Brasil, com a principal finalidade de prover modernização por meio da tecnologia da informação para apoiar o SUS. [6]

Neste trabalho serão utilizados os dados de 2010 a 2019 das estatísticas vitais; mortalidade pela CID-10 e mortalidade geral, para todo o Brasil. Os filtros para extração do banco de dados foram os óbitos registrados por local de residência, de mês nos anos últimos 10 anos.

#### **openDataSUS**

Também foram utilizadas as informações contidas no openDataSUS, onde estão disponibilizados dados do DATASUS de forma completa e prévia, para subsidiar estudos objetivos da situação sanitária e para ajudar em tomadas de decisões na elaboração de ações de saúde [7]. No presente trabalho, foram utilizados os valores preliminares do número de mortes no ano de 2020 como comparação ao valor esperado resultante das análises estatísticas.

#### **CORONAVÍRUS//BRASIL Portal Coronavírus**

Painel desenvolvido para veicular oficialmente a situação epidemiológica da COVID-19 no Brasil. [20]

Os valores relacionados a quantidade de óbitos no Brasil no ano de 2020 por coronavírus foram extraídos deste portal.

### **Utilização dos dados**

Como o foco do trabalho é estimar o número de mortes no Brasil em 2020 caso não houvesse a pandemia, necessita-se de dados precisos e confiáveis, por isso a base de dados do site oficial do SUS é visada. Entretanto, o site do DATASUS não disponibiliza, até o momento, os dados referentes ao ano de estudo, o que torna necessária a utilização de outro banco de dados, o do openDataSUS.

Como os valores referentes ao ano de 2020 são dados preliminares, poderá ocorrer alguma alteração no número de casos fatais registrados, a base teve como data de extração o mês de agosto de 2021, e última atualização dos dados em abril de 2021.

Assim, para fazer a previsão do número de mortes caso não houvesse a pandemia, foram utilizados os últimos 10 anos do DATASUS, de forma a embasar numa série histórica sólida o estudo do comportamento dos casos de óbitos no Brasil e fazer uma previsão para o ano de 2020. Após isso, tais dados serão comparados com aqueles disponibilizados no openDataSUS.

Para fazer as previsões e análises do trabalho, os principais modelos utilizados serão o de Regressão Linear e o modelo de Box & Jenkins.

### **2.2. Processos estocásticos**

Um processo estocástico é uma sequência de variáveis aleatórias (variáveis quantitativas que dependem de fatores aleatórios) ordenadas no tempo e definidas por um conjunto de pontos  $T$ , ou seja,  $T$  sendo um conjunto aleatório  $Z = \{Z(t), t \in T\}$  tal que para cada  $t \in T$ ,  $Z(t)$  é uma variável aleatória. [10]

O processo estocástico pode ser definido como estacionário e não estacionário, para este trabalho destaca-se o processo estacionário.



## Processos estacionários

Um processo estocástico pode-se definir como estacionário, quando todas as características do comportamento da base de dados são mantidas ao longo do tempo, isto é, os dados se desenvolvem aleatoriamente ao longo do tempo em torno da média, fazendo com que a escolha de uma data de origem não seja importante.

. O processo estacionário diferencia-se do processo não estacionário, onde possui momentos em que o processo cresce ou decresce sem limites, portanto, a média e a variância são funções do tempo.

Existem 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 primeiro e segundo momentos não variam com o tempo e a auto covariância depende apenas da diferença entre os tempos.
  - $E[Z(t)] = \mu_{(t)} = \mu$ , constante para todo  $t \in T$ ;
  - $E[Z_t^2] < \infty$ , para todo  $t \in T$ ;
  - $\gamma(t_1, t_2) = Cov[Z(t_1), Z(t_2)]$  é uma função de  $|t_1 - t_2|$

Neste trabalho o processo estacionário utilizado foi o de segunda ordem.

## Ruído branco

Caracteriza-se como ruído branco  $\{a_t\}$ ,  $t = 1, 2, \dots$  uma sequência onde a cada instante  $t$ , o processo tenha média 0, variância constante  $\sigma_a^2$  e não apresenta correlação serial.

- $E[a_t] = 0$ , para  $t = 1, 2, \dots$ ;
- $Var(a_t) = \sigma_a^2$ , para  $t = 1, 2, \dots$ ;
- $Cov(a_t, a_s) = 0$  para  $s \neq t$ ,  $t = 1, 2, \dots$ ;

## 2.3. Séries temporais

Uma série temporal é um conjunto de observações ordenadas no tempo e igualmente espaçadas. Neste tipo de dado, as observações vizinhas geralmente são dependentes e há interesse em modelar esta dependência. A série temporal é definida como sendo uma das realizações do processo estocástico, que é uma coleção de variáveis aleatórias ordenadas no tempo e definidas em um conjunto de pontos  $T$ . [10]

A análise de séries temporais visa em um primeiro momento inferir sobre as características básicas de como o processo estocástico das observações da série é gerado. [11]

(MORETTIN, 1981) define os objetivos da análise de séries temporais como:

- *Descrição*: Determinar as propriedades de uma série como, por exemplo, o padrão de tendência, a existência de alterações estruturais e sazonalidades etc. Antes de se tentar modelar e prever uma determinada série de tempo, é desejável ter uma ideia preliminar dos dados, para obter descrever algumas das suas principais propriedades. Isto será uma ajuda muito importante na realização do processo de modelação.
- *Explicação/Modelação*: Encontrar um modelo estatístico adequado que permita explicar o comportamento da série no período observado.
- *Previsão*: Estimar valores futuros de uma Série Temporal, com base em valores passados.
- *Controle de processos*: Boas previsões permitem ao estatístico, tomar medidas de forma a controlar um determinado processo. [8]

Análise de séries temporais é utilizada nas mais variadas áreas de aplicação: finanças, seguros, meteorologia, energia, estudos demografia etc.

### Componentes de uma série temporal

Considerando-se as observações  $Z_t$ ,  $t = 1, 2, \dots, n$  de uma série temporal, em que um modelo de decomposição aditivo consiste em escrever  $Z_t$  como uma função de três componentes não observáveis, isto é,

$$Z_t = T_t + S_t + a_t, \quad (2.1)$$

onde  $T_t$  e  $S_t$  representam a tendência e sazonalidade, respectivamente, enquanto  $a_t$  é um componente aleatório, de média 0, variância constante  $\sigma_a^2$  e  $Cov(a_t, a_s) = 0$  para  $s \neq t$ .

A tendência  $T_t$  pode ser definida como o padrão de crescimento, ou decréscimo dos dados em um período do tempo, enquanto a sazonalidade  $S_t$  pode ser definida como a repetição do comportamento em períodos idênticos no tempo.

Existem algumas formas de estimar a tendência de uma série, como: ajustar uma função tempo, como um polinômio, uma exponencial, ou outra função suavizada de  $t$ ; suavizar os valores da série ao redor de um ponto; médias móveis. No presente trabalho, o método que será utilizado para estimar a tendência e a sazonalidade da série será o modelo de regressão. [12]

### Função de autocorrelação

A Função de autocorrelação (FAC) é uma ferramenta muito utilizada para verificar se a série é estacionária. Considerando uma série  $Z_t$ , o coeficiente de correlação entre  $Z_t$  e  $Z_{t-k}$  é chamado de autocorrelação de  $k$ -ésima ordem dado por:

$$\rho_k = \frac{Cov(Z_t, Z_{t-k})}{Var(Z_t)} \quad (2.2)$$

onde:

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

### Função de autocorrelação parcial

A Função de Autocorrelação Parcial (FACP) é a medida da correlação entre duas observações seriais,  $Z_t$  e  $Z_{t+k}$ , eliminando a dependência dos termos intermediários,  $Z_{t+1}$ ,  $Z_{t+2}$ , ...,  $Z_{t+k-1}$ , dada por:

$$\varphi_{kk} = Corr(Z_t, Z_{t-k} | Z_{t-1}, \dots, Z_{t-(k+1)}) \quad (2.3)$$

## Correlograma

O gráfico que contém os  $k$  primeiros coeficientes de autocorrelação como função de  $k$  é chamado de correlograma. [13]

A Figura 1 apresenta um exemplo de FAC e FACP.

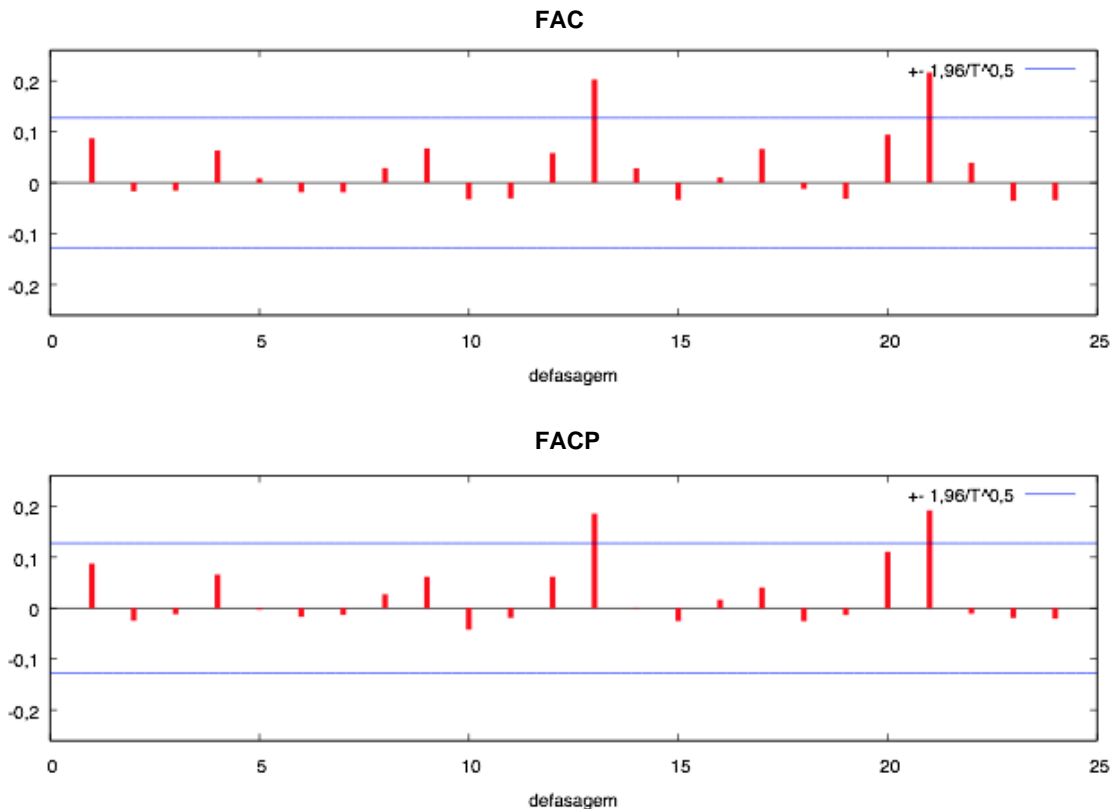


Figura 1: Exemplo de correlograma da FAC e FACP

## Teste da Raiz Unitária

Para que uma série temporal seja estacionária, suas características devem se manter ao longo do tempo. E estas características são:  $\mu_t$  e  $\sigma_t^2$  devem ser constantes.

Detecta-se estacionariedade de duas maneiras: através do teste da raiz unitária e do gráfico de correlograma.

- Teste da raiz unitária: verifica se o polinômio autorregressivo ou médias móveis utilizadas na análise possuem raiz sobre o círculo unitário.
- Teste com base no correlograma: é um teste visual, que consiste em verificar se o correlograma apresenta autocorrelações amostrais que diminuem gradualmente,

implicando em uma longa dependência, indicando, assim, não estacionariedade na série temporal.

O teste de *Dickey-Fuller Aumentado (ADF – Augmented Dickey-Fuller)* é utilizado para a verificação da raiz unitária. Com a seguinte expressão:

$$\Delta y_t = \beta_1 + \beta_2 + \delta y_{t-1} + \sum_{i=1}^m a_i \Delta y_{t-1} + \varepsilon_t \quad (2.4)$$

onde:

- $\Delta y_t$ : é a primeira diferença de  $Y_t$
- $\beta_1$ : é o intercepto;
- $\beta_2$ : é o coeficiente de tendência;
- $\delta$ : é o coeficiente de presença de raiz unitária e;
- $m$ : é o número de defasagens tomadas na série.

Teste de hipóteses:

- $H_0$ : “Existe raiz unitária” (não é estacionária)
- $H_1$ : “Não existe raiz unitária” (é estacionária)

Com estatística do teste:

$$T = \frac{\hat{\delta}}{se(\hat{\delta})} \quad (2.5)$$

onde:

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

## 2.4. Modelos Econométricos

Neste trabalho, dois modelos econométricos serão utilizados para verificar e prever o número de óbitos caso não houvesse a pandemia do coronavírus no Brasil, sendo eles: o modelo de regressão linear e o modelo Box & Jenkins.

## Modelo de Regressão Linear

O modelo de regressão linear múltiplo foi usado neste trabalho é definido da seguinte forma:

$$z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \alpha_1 D_{1,t} + \alpha_2 D_{2,t} + \dots + \alpha_{11} D_{11,t} + \varepsilon_t \quad (2.6)$$

Que também pode ser escrito desta forma:

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1^2 & 1 & \dots & 0 \\ 1 & 2 & 2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & n & n^2 & 0 & \dots & 0 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \alpha_{11} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.7)$$

Em que:

$$\underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}_{nx1}, X = \begin{bmatrix} 1 & 1 & 1^2 & 1 & \dots & 0 \\ 1 & 2 & 2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & n & n^2 & 0 & \dots & 0 \end{bmatrix}_{nxp}, \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \alpha_{11} \end{bmatrix}_{px1} \text{ e } \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{nx1}$$

Assim, pode-se escrever o modelo da forma matricial:

$$\underline{z} = X\underline{\beta} + \underline{\varepsilon} \quad (2.8)$$

onde:

- $z_t$ : é o valor da variável resposta para a t-ésima observação, onde  $t = 1, 2, \dots, n$ ;
- $n$ : é o tamanho da amostra
- $X$ : é uma matriz de pesos, que define quais parâmetros  $\beta_i, \alpha_j$ , para  $i = 0, 1, 2$  e  $j = 1, 2, \dots, 11$  serão utilizados
- $\beta_0, \beta_1, \beta_2, \alpha_1, \alpha_2, \dots, \alpha_{11}$ : são parâmetros desconhecidos;
- $p$ : é a quantidade de parâmetros desconhecidos,  $p$  neste trabalho é igual a 14;
- $D_{1,t}, D_{2,t}, \dots, D_{11,t}$ : são variáveis *dummies*, variável binária (0 ou 1);
- $\varepsilon_t$ : é o t-ésimo termo de erro aleatório com média 0 e variância  $\sigma^2$ .

Como a média dos erros aleatórios é igual a 0, tem-se que:

$$E(Z_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \alpha_1 D_{1,t} + \alpha_2 D_{2,t} + \dots + \alpha_{11} D_{11,t} \quad (2.9)$$

Logo:

$$\hat{z}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\alpha}_1 D_{1,t} + \hat{\alpha}_2 D_{2,t} + \dots + \hat{\alpha}_{11} D_{11,t} \quad (2.10)$$

Com isso, pode-se definir o resíduo da seguinte forma:

$$\hat{\varepsilon}_t = z_t - \hat{z}_t \quad (2.11)$$

Caso haja a interferência de sazonalidade, as *dummies* se fazem necessárias para explicar o comportamento do modelo, define-se como:

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{j=1}^{11} \alpha_j D_{j,t} \quad (2.12)$$

Onde  $D_{j,t}$  é a  $j$ -ésima variável *dummy* e  $\alpha_j$  é o parâmetro de impacto do  $j$ -ésimo mês.

Como existe influência dos meses de cada ano, a *dummy* influencia os dados da seguinte forma: suponha que o mês de janeiro seja importante, então  $D_{1,t}$  será igual a 1 e o  $\alpha_1$  terá um valor estimado para este mês, enquanto todos os outros meses não terão influência, uma vez que cada  $\alpha_j$  será multiplicado por  $D_j$ , que vale 0 para  $j = 2, \dots, 11$ .

### Mínimos Quadrados Ordinários (MQO)

O objetivo do MQO é encontrar valores de  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  e  $\alpha_j$ ,  $j = 1, 2, \dots, 11$  que minimizem o erro, e para que isso aconteça, é preciso encontrar todas as diferenças quadráticas do erro. Defino  $Q$  como o somatório dos erros quadráticos:

$$Q = \sum_{t=1}^n (\hat{\varepsilon}_t)^2 \quad (2.13)$$

É preciso escrever Q na forma matricial, para se encontrar mais facilmente os parâmetros que minimizam os erros, e deve-se usar a equação (2.8):

$$\begin{aligned}
 Q &= \sum_{t=1}^n (\hat{\varepsilon}_t)^2 = \underline{\varepsilon}^T \underline{\varepsilon} \\
 &= (\underline{z} - X\underline{\beta})^T (\underline{z} - X\underline{\beta}) \\
 &= (\underline{z}^T - X^T \underline{\beta}^T) (\underline{z} - X\underline{\beta}) \\
 &= \underline{z}^T \underline{z} - \underline{z}^T X \underline{\beta} - \underline{\beta}^T X^T \underline{z} + \underline{\beta}^T X^T X \underline{\beta} \\
 &= \underline{z}^T \underline{z} - 2\underline{\beta}^T X \underline{z} + \underline{\beta}^T X^T X \underline{\beta}
 \end{aligned} \tag{2.14}$$

E assim tem-se que:

$$Q(\underline{\beta}) = \underline{z}^T \underline{z} - 2\underline{\beta}^T X \underline{z} + \underline{\beta}^T X^T X \underline{\beta} \tag{2.15}$$

Para minimizar os erros deve-se derivar a função (2.15) para a matriz dos parâmetros desconhecidos  $\underline{\beta}$ , após isso, igualar 0 para assim poder estimar os parâmetros que minimizam os erros.

$$\frac{dQ}{d\underline{\beta}} = -2X^T \underline{z} + 2X^T X \underline{\beta} = 0 \tag{2.16}$$

Com isso pode-se estimar o melhor modelo de regressão para os dados, dado que:

$$\underline{\hat{\beta}} = (X^T X)^{-1} X^T \underline{z} \tag{2.17}$$

Após isso, é esperado que os resíduos do modelo ( $\hat{\varepsilon}_t$ ) estejam distribuídos entorno de 0 e sigam uma distribuição normal, definindo assim um modelo bem ajustado.



## Modelo ARIMA

O modelo ARIMA é um dos mais utilizados para fazer previsões, e neste trabalho ele também será usado.

O modelo ARIMA (p,d,q) tem como parâmetros o p que é o número de termos autorregressivos, o d que é o número de diferenças e o q que é o número de termos da média móvel.

Um modelo ARIMA (p,0,q) é especificado da seguinte forma, onde  $Z_t$  é uma série estacionária:

$$Z_t = c + \varphi_1 Z_{t-1} + \dots + \varphi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.18)$$

onde  $\varphi_1, \dots, \varphi_p$  são parâmetros da estrutura autorregressiva,  $\theta_1, \dots, \theta_q$  são parâmetros da estrutura médias móveis e  $a_t$  ruído branco e segue uma distribuição normal.

A equação (2.18) mostra que a variável dependente  $Z_t$  aparece, independentemente do número de defasagens, p vezes, e o termo  $a_t$  aparece defasado q vezes, assim tem-se que o p e o q indicam o número de defasagens de  $Z_t$  e  $a_t$ , respectivamente.

Caso a série não seja estacionária é necessário fazer d diferenças para torná-la estacionária, então o modelo ficará da seguinte forma: ARIMA (p,d,q) em termos de diferença  $Z_t$  ( $\Delta Z_t$ ), e sua representação é dada por:

$$\Delta^d Z_t = c + \varphi_1 \Delta^d Z_{t-1} + \dots + \varphi_p \Delta^d Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.19)$$

Onde  $\Delta^d$  é o operador de diferença, calculado por:  $Z_t - Z_{t-1}$ , para  $t = 2, \dots, n$ , onde n é a quantidade total da amostra.

$a_t$  é o ruído branco independente da  $f(t)$  em que a média é igual a 0, a variância é igual a  $\sigma^2$  e a  $cov(a_t, a_s)$  é igual a 0 para  $t \neq s$ . Pode-se seguir as seguintes hipóteses para  $a_t$ :

- $H_0$ : “Existe raiz unitária” (não é estacionária)
- $H_1$ : “Não existe raiz unitária” (é estacionária)

## Modelo SARIMA

O modelo ARIMA sazonal, conhecido comumente como SARIMA, é composto por uma parte não sazonal (p,d,q) e uma parte sazonal (P,D,Q).

Sua representação é dada por SARIMA(p,d,q)(P,D,Q) e sua fórmula é dada da seguinte forma:

$$\Delta_S^D \Delta^d Z_t = c + \varphi_1 \Delta^d Z_{t-1} + \dots + \varphi_p \Delta^d Z_{t-p} + \Phi_1 \Delta_S^D Z_{t-1} + \dots + \Phi_P \Delta_S^D Z_{t-P} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} - \Theta_1 a_{t-1} - \dots - \Theta_Q a_{t-Q} \quad (2.20)$$

onde:

- p é o número de termos autorregressivos da parte não sazonal;
- q é o número de termos de médias móveis da parte não sazonal;
- $\Delta^d$  é o operador de diferença da parte não sazonal de ordem d, calculado por:  $Z_t - Z_{t-1}$ , para  $t = 2, \dots, n$ , onde n é a quantidade total da amostra;
- P é o número de termos autorregressivos da parte sazonal;
- Q é o número de termos de médias móveis da parte sazonal;
- $\Delta_S^D$  é o operador de diferença da parte sazonal de ordem D, calculado por:  $Z_t - Z_{t-S}$ , para  $t = S + 1, \dots, n$  e  $S = 12$ , onde n é a quantidade total da amostra;
- $\varphi_1, \dots, \varphi_p$  são parâmetros da estrutura autorregressiva da parte não sazonal;
- $\theta_1, \dots, \theta_q$  são parâmetros da estrutura média móveis da parte não sazonal;
- $\Phi_1, \dots, \Phi_P$  são parâmetros da estrutura autorregressiva da parte sazonal;
- $\Theta_1, \dots, \Theta_Q$  são parâmetros da estrutura média móveis da parte sazonal;
- $a_t$  é o ruído branco

Para este trabalho, todas as referências feitas ao Modelo ARIMA englobam também o Modelo SARIMA, pois para obtenção dos termos P e Q os mesmos passos são seguidos, porém realizados com a série de diferença sazonal, explicada acima em  $\Delta_S^D$ .

## Metodologia Box & Jenkins

A metodologia de Box & Jenkins para a previsão se baseia no ajuste de modelos ARIMA e ARIMA sazonal (SARIMA), para que se estabeleça um mecanismo no qual valores futuros possam ser previstos, dados valores presentes e passados. As técnicas deste processo de previsão estão fundamentadas na exploração da correlação temporal que pode existir entre os valores da série.

O nome desta metodologia advém dos estatísticos George Box e Gwilym Jenkins, que a criaram em 1976, e que se tornou o modelo mais utilizado para fazer previsões a curto prazo. O modelo possui três principais etapas, vindo de uma série temporal estacionária:

- Identificação: Descobre-se os valores apropriados de  $p$ ,  $d$ , e  $q$  através da análise do correlograma e do correlograma parcial;
- Checagem de diagnóstico: Deve-se estimar e testar os parâmetros dos termos autorregressivos e da média móvel incluídos no modelo. Um teste possível é ver se os resíduos estimados são ruídos brancos.
- Previsão: Prevê valores futuros da série considerando o erro padrão da previsão, e assim é possível construir intervalos de confiança para a previsão.

### Identificação de um ARIMA

Esta etapa é considerada a mais crítica ao se utilizar o modelo ARIMA, pois a identificação dos parâmetros  $p$  e  $q$  são feitas com base nas autocorrelações e autocorrelações parciais, e este processo é bastante subjetivo.

O Quadro 1 apresenta alguns comportamentos esperados da FAC e FACP de alguns modelos da classe ARMA mais comuns, para a identificação. [15]

**Quadro 1:** 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 ( $\varphi_{11} \neq 0$ )
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 senoides amortecidas
AR(2)	Mistura de exponenciais ou ondas senoides amortecidas	1 pico na defasagem 1 e um pico na defasagem 2 ( $\varphi_{11} \neq 0$ e $\varphi_{22} \neq 0$ )
ARMA(1,1)	Decaimento exponencial	Decaimento exponencial

## Estimação de um ARIMA

Após identificar o modelo para a série temporal, o passo seguinte é estimar os parâmetros, o que é feito através de métodos iterativos, como o método dos momentos, estimadores de mínimos quadrados, ou o estimador de máxima verossimilhança.

## Análise de resíduos

É esperado que o resíduo ( $a_t$ ) seja um ruído branco, ou seja, que tenha média 0 e variância constante  $\sigma_t^2$  e as observações sejam independentes seguindo uma distribuição normal. [16]

Para testar a normalidade, será utilizado o teste de Shapiro-Wilk e para análise dos resíduos será usado o gráfico de correlograma dos resíduos. Espera-se que eles sejam não correlacionados, o que significa que as defasagens não devem apresentar autocorrelação significativa dentro do seguinte intervalo:

$$\rho_k \sim N(0,1), \text{ Logo: } (0 \pm 1,96 \cdot \frac{1}{\sqrt{n}}) \quad (2.21)$$

Sendo  $n$ , o número de observações da série.

Com testes de hipótese:

- $H_0: \rho_k = 0$
- $H_1: \rho_k \neq 0$

A Figura 2 apresenta um exemplo de um correlograma dos resíduos. As linhas vermelhas representam o intervalo citado acima.

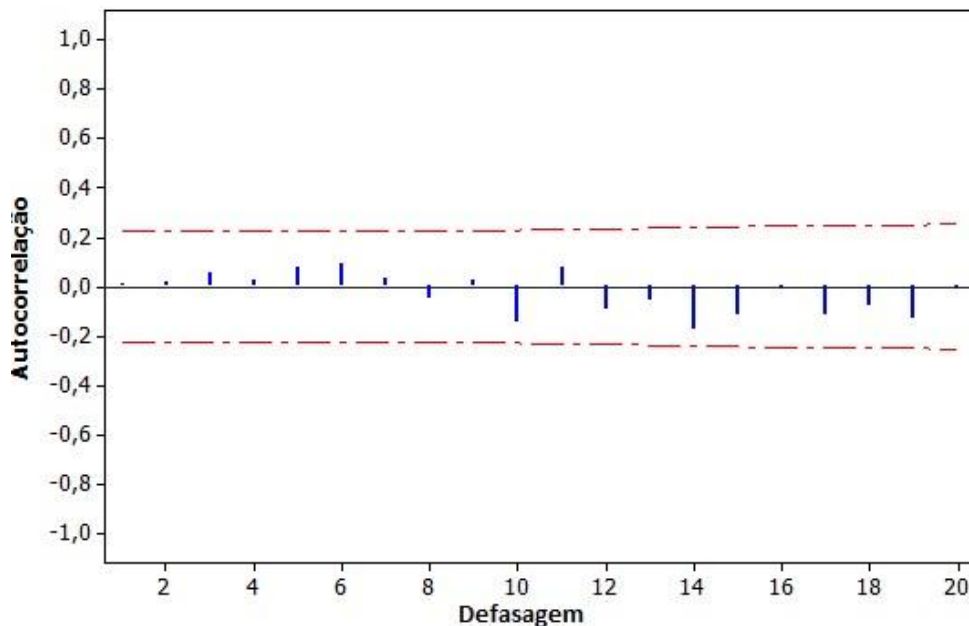


Figura 2: Exemplo de correlograma da FAC dos resíduos.

## 2.5. Comparação entre os modelos

Para avaliar o quão boa é a previsão, algumas estatísticas de aderência se fazem necessárias.

Espera-se que o modelo escolhido apresente o melhor resultado em tais estatísticas, para modelagem mais assertiva para o número de mortes no Brasil em 2020 caso não houvesse a pandemia.

Os modelos foram estimados para os anos de 2010 a 2017, para verificar se sua capacidade de previsão foi adequada, foi feita a previsão dos modelos para os anos de 2018 e 2019 e comparado com os dados reais. O modelo que apresentou melhor ajuste para a série, de acordo com as estatísticas de aderência, foi o modelo utilizado para prever o ano de 2020.

A seguir as estatísticas de aderência que serão utilizadas, onde  $t = 1, 2, \dots, 24$  se referem aos anos de 2018 e 2019,  $Z_t$  os dados observados pelo DATASUS para estes dois anos e  $\hat{Z}_t$  os dados previstos pelos modelos MQO e SARIMA.

### **Desvio Médio Absoluto (Mean Absolute Deviation - MAD)**

O desvio médio absoluto representa a diferença absoluta entre os valores ajustados e observados, considerando assim a média dos desvios absolutos.

$$MAD = \frac{\sum_{t=1}^n |z_t - \hat{z}_t|}{n} \quad (2.22)$$

onde:

- $Z_t$ : é o valor observado;
- $\hat{Z}_t$ : é o valor ajustado;
- $n$ : é o tamanho da série.

### **Erro Quadrático Médio (Mean Square Error - MSE)**

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

Sua expressão é dada por:

$$MSE = \frac{\sum_{t=1}^n (z_t - \hat{z}_t)^2}{n} \quad (2.23)$$

onde:

- $Z_t$ : é o valor observado;
- $\hat{Z}_t$ : é o valor ajustado;
- $n$ : é o tamanho da série.

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

A 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 está próximo de zero.

Sua expressão é dada por:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (z_t - \hat{z}_t)^2}{n}} \quad (2.24)$$

onde:

- $Z_t$ : é o valor observado;
- $\hat{Z}_t$ : é o valor ajustado;
- $n$ : é o tamanho da série.

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

O MAPE representa 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.

Sua expressão é dada por:

$$MAPE = \sum_{t=1}^n \left| \frac{z_t - \hat{z}_t}{z_t} \right| \cdot \frac{100\%}{n} \quad (2.25)$$

onde:

- $Z_t$ : é o valor observado;
- $\hat{Z}_t$ : é o valor ajustado;
- $n$ : é o tamanho da série.

### **R<sup>2</sup> ajustado**

O R<sup>2</sup> representa o poder de explicação do modelo para a variável resposta. Quanto mais próximo de 1 for o R<sup>2</sup>, melhor será para explicar o modelo.

Sua expressão é dada por:

$$R^2 = 1 - \frac{\sum_{t=1}^n (z_t - \hat{z}_t)^2}{\sum_{t=1}^n (z_t - \bar{z}_t)^2} \quad (2.26)$$

onde:

- $Z_t$ : é o valor observado;
- $\hat{Z}_t$ : é o valor ajustado;
- $n$ : é o tamanho da série.

O  $R^2$  é uma estatística de aderência, que não leva em consideração a quantidade de variáveis explicativas do modelo, e apresenta valores mais altos para modelos com várias variáveis, e por este motivo será utilizado o  $R^2$  ajustado, que também melhor explicará o modelo se o valor for próximo de 1, porém penaliza a quantidade de variáveis, sendo uma escolha mais assertiva, pois leva em conta todas as variáveis explicativas do modelo e é definido por:

$$R_{ajus}^2 = 1 - \frac{(n-1)}{(n-1)-k}(1 - R^2) \quad (2.27)$$

onde:

- $n$ : é o tamanho da série e;
- $k$ : é a quantidade de variáveis explicativas do modelo.

### **Critério de Informação de Akaike (AIC)**

O critério de informação de Akaike (AIC), proposto por Akaike (1974) [17], 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:

$$AIC = -2 \ln(Lp) + 2[(p + 1) + 1] \quad (2.28)$$

onde:

- $L_p$ : é a função de máxima verossimilhança do modelo;
- $p$ : é o número de variáveis explicativas consideradas no modelo.



## **Critério de Informação Bayesiano (BIC)**

O critério de informação bayesiano (BIC), proposto por Schwarz (1978) [18], penaliza os modelos com muitas variáveis, sendo assim, se utiliza o Princípio da Parcimônia.

Sua expressão é dada por:

$$BIC = -2 \ln(Lp) + [(p + 1) + 1] \ln(n) \quad (2.29)$$

onde:

- $L_p$ : é a função de máxima verossimilhança do modelo;
- $p$ : é o número de variáveis explicativas consideradas no modelo.
- $n$ : é o tamanho da série

## **Critério de Informação de Hannan-Quinn (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.

Sua expressão é dada por:

$$HQC = n \ln\left(\frac{RSS}{n}\right) + 2k \ln(\ln(n)) \quad (2.30)$$

onde:

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

## 2.6. Comparação entre os óbitos observados e previstos para 2020

Após utilizar os métodos de previsão apresentados nas seções anteriores e verificar, através dos testes de aderência, qual modelo se adequa melhor e tem uma previsão mais correta com relação aos óbitos nos anos de 2018 e 2019, foi feita uma projeção para verificar qual seria o comportamento do número de mortes no Brasil em 2020 sem a influência do coronavírus. A próxima etapa foi verificar a quantidade de mortes previstas e comparar com a quantidade de óbitos informada no site OpenDataSUS, observando se os valores se encontram próximos, para assim concluir o sobre impacto da COVID-19 no Brasil.

Espera-se desta forma atingir o objetivo do trabalho.

## 2.7. Metodologia

Inicialmente serão realizadas análises descritivas do número de óbitos dos últimos 10 anos no Brasil. Em seguida a série será modelada para encontrar o melhor modelo ajustado e prever o ano de 2020 caso a pandemia não se fizesse presente no país no ano de estudo, os modelos de regressão linear e Box & Jenkins serão utilizados para tal.

Posteriormente serão calculadas estatísticas de aderência aos modelos para avaliar a capacidade preditiva de cada um. O melhor modelo ajustado será o que encontrar estatísticas de aderência mais satisfatórias para a série, e assim realizar as previsões.

O nível de significância adotado em todo o trabalho será de 0,05 e os *softwares* que serão utilizados para estas etapas serão:

- **Excel do office 365 e R 4.0.3:** para manipular os dados e realização das análises descritivas;
- **Gretl 2021b:** para teste de estacionariedade e normalidade, modelagem, estatísticas de aderência e previsão dos modelos de regressão linear e Box & Jenkins.

### 3. RESULTADOS

Este capítulo é responsável por apresentar os dados de forma descritiva, os resultados das análises dos modelos, o melhor modelo ajustado e comparar os valores de óbitos esperados no Brasil em 2020, caso não houvesse a pandemia, com os valores reais.

#### 3.1. Estatísticas descritivas

Neste tópico, serão apresentados os dados coletados para as análises mencionadas nos capítulos anteriores, e algumas breves estatísticas que podem dar algum direcionamento para o que irá se encontrar no desenvolvimento dos modelos.

Primeiramente, serão discriminados os valores dos bancos de dados, e, a fim de realizar uma exploração inicial, a variação percentual do número de óbitos ocorridos no período analisado frente ao período anterior, que poderá ser uma comparação entre meses de anos distintos, ou entre os próprios totais anuais.

Seguem na Tabela 1 os dados disponibilizados pelo DATASUS com a quantidade de mortes ano a ano, de 2010 a 2019.

**Tabela 1:** Dados DATASUS número de óbitos ano a ano de 2010 a 2019.

<b>ANO</b>	<b>Nº DE ÓBITOS</b>
<b>2010</b>	1.136.947
<b>2011</b>	1.170.498
<b>2012</b>	1.181.166
<b>2013</b>	1.210.474
<b>2014</b>	1.227.039
<b>2015</b>	1.264.175
<b>2016</b>	1.309.774
<b>2017</b>	1.312.663
<b>2018</b>	1.316.719
<b>2019</b>	1.349.801

Fonte DATASUS: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10br.defmortalidade>

Percebe-se na Tabela 1 uma tendência crescente dos dados, visto que o ano seguinte é sempre maior do que o ano anterior, e que o ano de 2010 apresentou menor valor (1.136.947) e 2019 apresentou a maior quantidade de óbitos (1.349.801).

Na Tabela 2 destaca-se a média dos últimos 9 anos, visto que 2010 não foi comparado com o ano anterior (que não faz parte dos anos a serem analisados), do aumento percentual para cada mês do ano.

**Tabela 2:** Aumento percentual médio para cada mês.

<b>MÊS</b>	<b>VARIAÇÃO (%)</b>
JANEIRO	1,92%
FEVEREIRO	1,26%
MARÇO	1,93%
ABRIL	2,34%
MAIO	2,44%
JUNHO	2,30%
JULHO	2,65%
AGOSTO	1,85%
SETEMBRO	1,85%
OUTUBRO	1,78%
NOVEMBRO	1,74%
DEZEMBRO	1,54%

Fonte DATASUS: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10br.defmortalidade>

Na Tabela 2 calculou-se a média da variação percentual de cada mês com relação ao mesmo mês do ano anterior dos anos de 2011 a 2019. Ela mostra que a média da variação percentual no mês de fevereiro possui o valor de 1,26%, sendo o menor crescimento no número de mortes ano após ano, já o mês de julho possui o maior crescimento médio ao longo dos anos (2,65%).

A Tabela 3 mostra a evolução percentual no aumento do número de mortes a cada ano no Brasil.

**Tabela 3:** Aumento percentual de óbitos a cada ano.

<b>ANO</b>	<b>VARIAÇÃO (%)</b>
2010	-
2011	2,95%
2012	0,91%
2013	2,48%
2014	1,37%
2015	3,03%
2016	3,61%
2017	0,22%
2018	0,31%
2019	2,51%
<b>MÉDIA</b>	<b>1,93%</b>

Fonte DATASUS: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10br.defmortalidade>

Na Tabela 3 calculou-se a variação percentual de cara ano com relação ao ano anterior dos anos de 2011 a 2019. Pode-se perceber que o ano de 2016 apresentou o maior aumento percentual no número de mortes, sendo a quantidade de óbitos esse ano 3,61% superior àquela auferida em 2015. Já o ano de 2017 teve o menor aumento percentual em relação ao ano anterior (0,22%).

A Tabela 4 contém os dados extraídos no banco de dados do openDataSUS, onde encontram-se os valores preliminares (podem sofrer alteração até a divulgação dos dados oficiais) dos números de óbitos totais, coletados para o ano de 2020.

**Tabela 4:** Nº de óbitos preliminares para 2020 emitidos pelo openDataSUS

<b>MÊS</b>	<b>Nº DE ÓBITOS 2020 (PRÉVIO)</b>
JANEIRO	113.292
FEVEREIRO	106.229
MARÇO	119.998
ABRIL	125.726
MAIO	156.544
JUNHO	142.337
JULHO	147.109
AGOSTO	142.756
SETEMBRO	127.773
OUTUBRO	128.250
NOVEMBRO	118.723
DEZEMBRO	131.351
<b>TOTAL</b>	<b>1.560.088</b>

Fonte openDataSUS: <https://opendatasus.saude.gov.br/dataset/sistema-de-informacao-sobre->

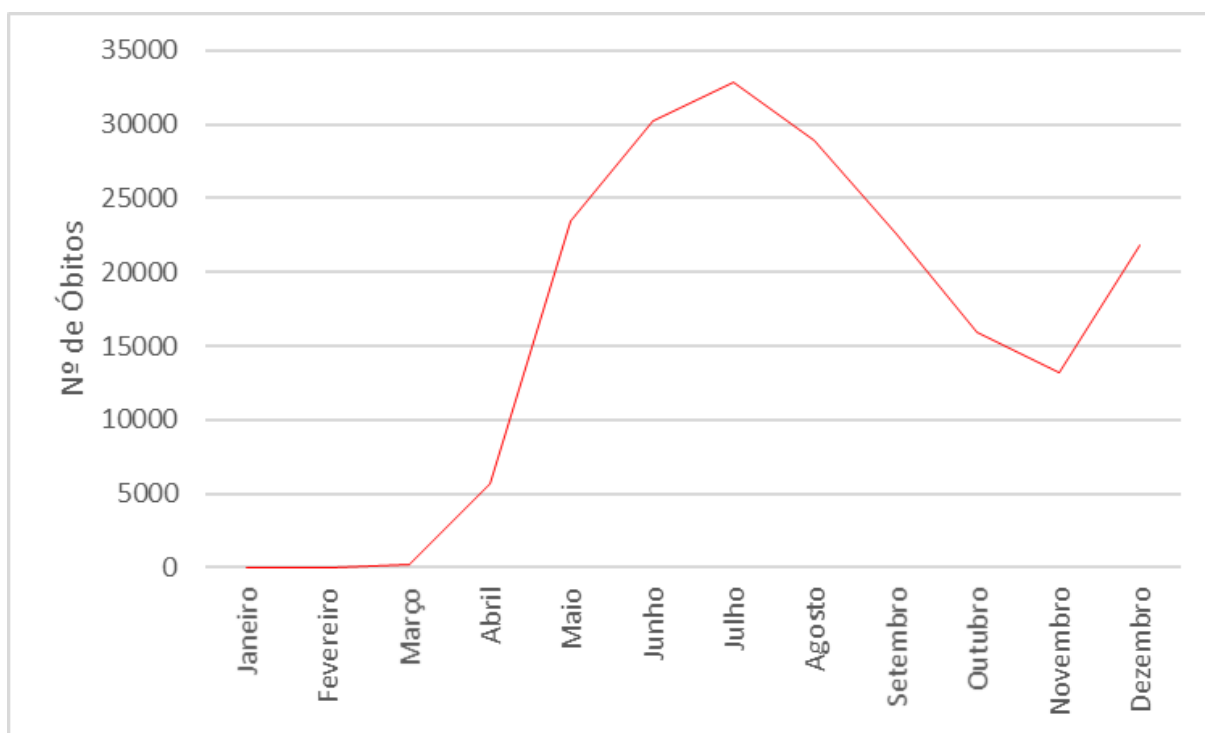
Em relação aos dados preliminares para 2020, apresentados na Tabela 4, percebe-se que o mês de maio teve o valor mais elevado de óbitos do ano (156.544), e o menor valor registrado foi o de fevereiro (106.229). Já a quantidade total de óbitos, como informada pelo Open Data, foi de 1.560.088.

A Tabela 5 retrata o número de mortes por COVID-19 no ano de 2020 pelo Painel Coronavírus e a Figura 3 a distribuição das mortes pela doença ao longo do tempo.

**Tabela 5:** Nº de óbitos por COVID-19 em 2020 emitidos pelo Painel Coronavírus

MÊS	Nº DE ÓBITOS POR COVID-19 EM 2020
JANEIRO	-
FEVEREIRO	-
MARÇO	201
ABRIL	5.700
MAIO	23.413
JUNHO	30.280
JULHO	32.881
AGOSTO	28.906
SETEMBRO	22.571
OUTUBRO	15.932
NOVEMBRO	13.236
DEZEMBRO	21.829
<b>TOTAL</b>	<b>194.949</b>

Fonte Painel Coronavírus: <https://covid.saude.gov.br>



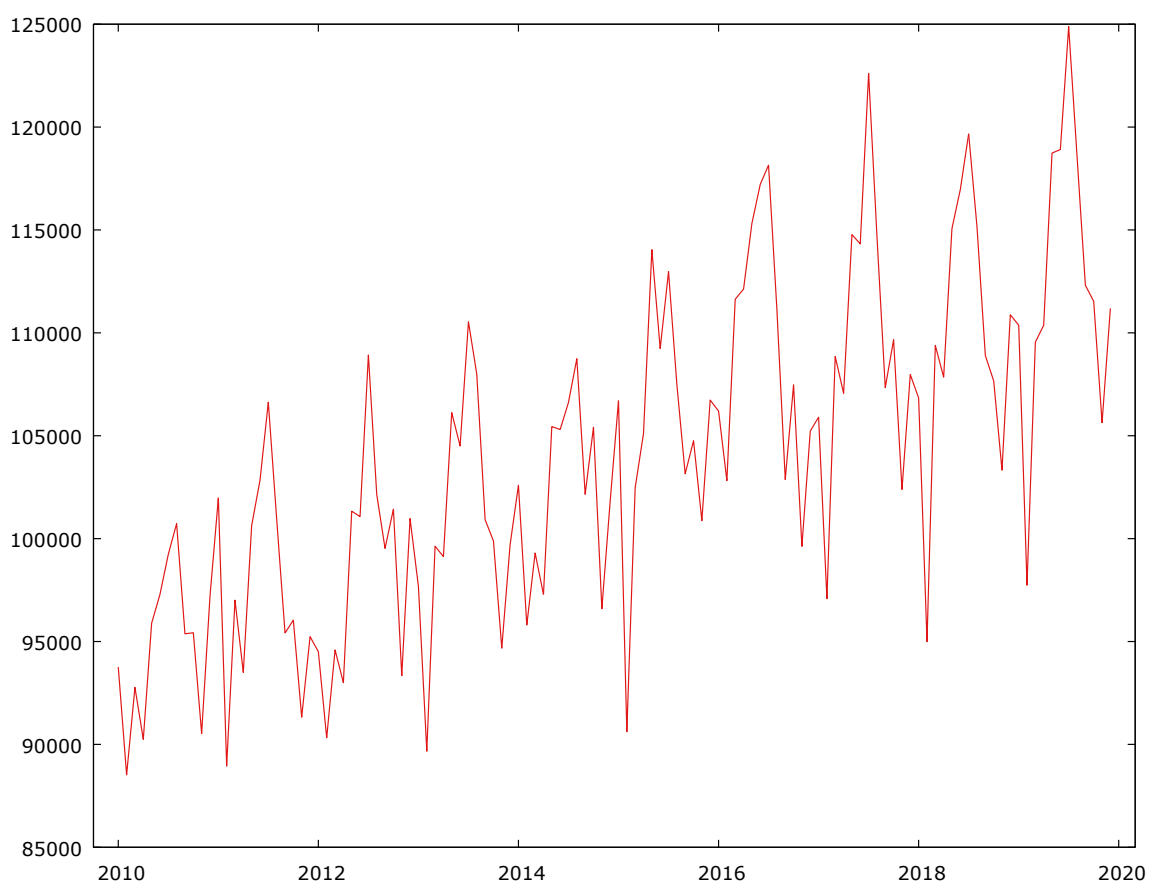
**Figura 3:** Série temporal do número de óbitos por COVID-19 em 2020.

A Tabela 5 e Figura 3, mostram a quantidade de óbitos causados por coronavírus no Brasil em 2020 e é possível identificar que o mês de julho teve a maior prevalência de mortes por conta do contágio do COVID-19, 32.881. Os meses de janeiro e fevereiro encontram-se zerados, pois a doença chegou ao Brasil em março de 2020.

### 3.2. Características da Série Temporal

Os dados encontram-se regularmente distribuídos (mês a mês, durante os últimos 10 anos), caracterizando-se uma série temporal. Portanto, estudar o comportamento desta série é importante para obtenção de informações que poderão direcionar os modelos e análises que este trabalho visa usar.

A Figura 4 mostra o número de órbitos no Brasil, de 2010 a 2019, ao longo dos meses.



**Figura 4:** Série temporal do número de órbitos no Brasil de 2010 a 2019.

A Figura 4 mostra que o comportamento na quantidade de órbitos no país tem uma tendência crescente, visto que o número de órbitos aumenta com o passar dos anos. Outra característica visualmente aparente é a da sazonalidade, pois a cada ano o comportamento do gráfico se repete, isto é, a cada ciclo de ano, os dados possuem um comportamento típico.

Com relação aos valores, tem-se em média cerca de 103.990 óbitos mensais, em fevereiro de 2010 ocorreu o menor número de mortes em todos os anos de análise (88.516) e em julho de 2019 ocorreu o maior número de mortes (124.885), o que corrobora a tendência crescente da série. A Tabela 6 mostra mais resultados descritivos.

**Tabela 6:** Estatística descritiva 2010 a 2019.

<b>ESTATÍSTICAS DESCRITIVAS, USANDO AS OBSERVAÇÕES 2010:01 - 2019:12</b>	
<b>MÉDIA</b>	103.990
<b>MEDIANA</b>	103.240
<b>MÍNIMO</b>	88.516
<b>MÁXIMO</b>	124.885
<b>DESVIO PADRÃO</b>	8.016,20
<b>COEFICIENTE DE VARIAÇÃO (CV)</b>	0,08
<b>PERCENTIL DE 5%</b>	90.529
<b>PERCENTIL DE 95%</b>	118.550

Para se fazer as análises deste trabalho, foram utilizados os dados de 2010 a 2017 desta série temporal. Para identificação do modelo melhor ajustado que preverá os valores de 2018 e 2019 de forma mais aproximada. Após a verificação e a escolha do modelo, este é utilizado para prever os óbitos esperados, sem o impacto do coronavírus, no ano de 2020.

### **3.3. Modelo de Regressão Linear**

Para a modelagem da série temporal foram ajustados diversos modelos de regressão linear, utilizando-se os Mínimos Quadrados Ordinários, conhecido como MQO.

Considerou-se diversos modelos, com tendência, sem tendência, com ou sem a constante, com a influência da sazonalidade (com o uso das *dummies*) e sem a influência da sazonalidade. Na Tabela 8 são apresentados os resultados das estatísticas de aderência de cada MQO.



**Tabela 7:** Estatísticas de aderência para os modelos MQO.

PARÂMETROS MQO	MAPE	RMSE	R <sup>2</sup> AJUSTADO	AIC	BIC	HQC
SEM CONSTANTE E T	38,868%	215,95	-	2.338,45	2.341,01	2.339,48
CONSTANTE E T	4,084%	72,22	0,4765	1.919,86	1.924,99	1.921,94
CONSTANTE E T <sup>2</sup>	4,227%	73,08	0,4511	1.924,41	1.929,54	1.926,48
SEM CONSTANTE E T <sup>2</sup>	57,136%	253,93	-	2.400,67	2.403,23	2.401,70
CONSTANTE, T E T <sup>2</sup>	4,085%	72,22	0,4711	1.921,83	1.929,52	1.924,94
SEM CONSTANTE, T E <i>DUMMIES</i>	<b>1,832%</b>	<b>48,77</b>	<b>0,8767</b>	<b>1.791,13</b>	<b>1.824,47</b>	<b>1.804,60</b>
SEM CONSTANTE, T <sup>2</sup> E <i>DUMMIES</i>	2,028%	51,18	0,8504	1.809,65	1.842,99	1.823,13
SEM CONSTANTE, T, T <sup>2</sup> E <i>DUMMIES</i>	1,833%	48,73	0,8756	1.792,79	1.828,70	1.807,31
SEM CONSTANTE E <i>DUMMIES</i>	4,519%	74,31	0,3436	1.950,78	1.981,55	1.963,22
CONSTANTE, T E <i>DUMMIES</i>	<b>1,832%</b>	<b>48,77</b>	<b>0,8767</b>	<b>1.791,13</b>	<b>1.824,47</b>	<b>1.804,60</b>
CONSTANTE, T <sup>2</sup> E <i>DUMMIES</i>	2,028%	51,18	0,8504	1.809,65	1.842,99	1.823,13
CONSTANTE, T, T <sup>2</sup> E <i>DUMMIES</i>	1,833%	48,73	0,8756	1.792,79	1.828,70	1.807,31
CONSTANTE E <i>DUMMIES</i>	4,519%	74,31	0,3436	1.950,78	1.981,55	1.963,22

A Tabela 7 mostra que, de acordo com as estatísticas de aderência, dois modelos apresentaram os mesmos resultados e o melhor ajuste. Os modelos foram: sem a constante, com a tendência linear e com a sazonalidade; com a constante, a tendência e a sazonalidade. Por este motivo, levando em consideração a igualdade dos modelos, os resultados que serão apresentados a partir de agora levam em consideração o melhor modelo como o que possui constante, tendência linear e sazonalidade, por conter mais variáveis explicativas.

Na Tabela 8 apresentam-se os parâmetros das tendências e das sazonalidades.

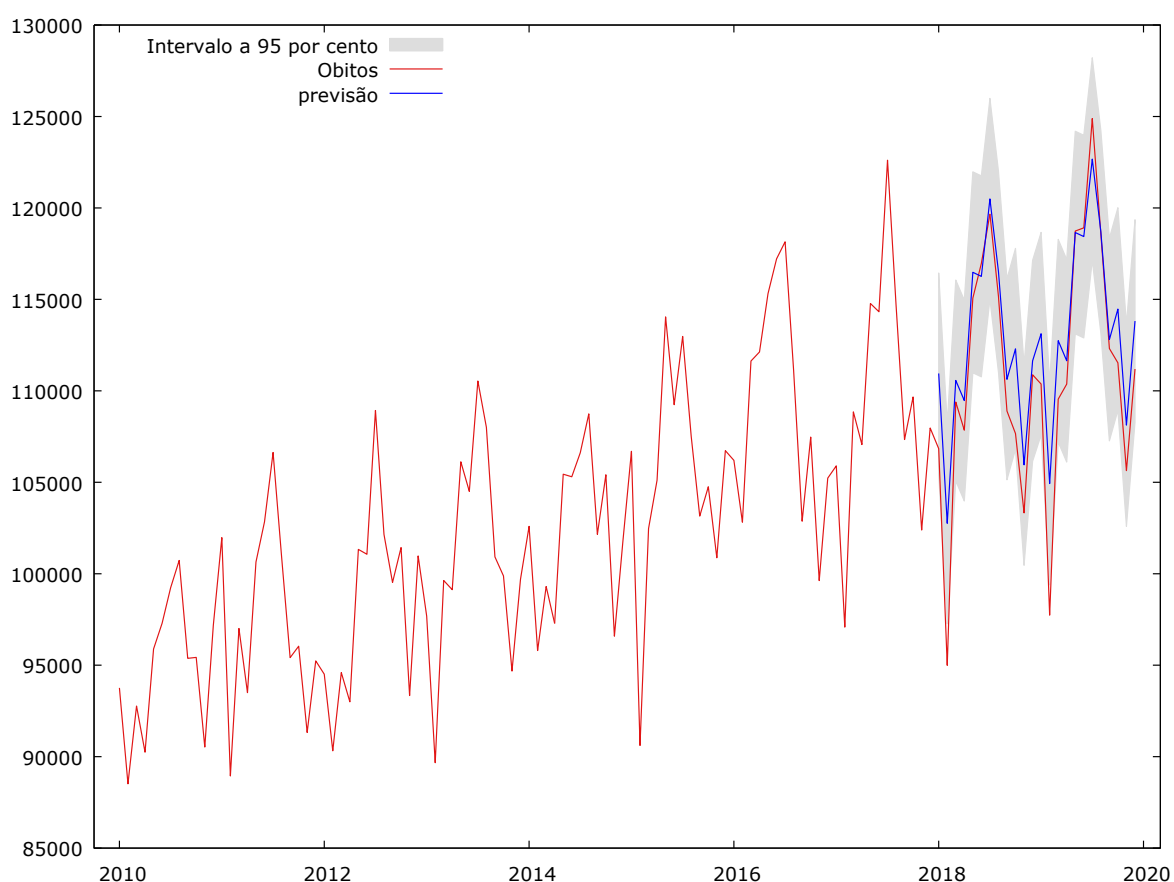
**Tabela 8:** Parâmetros da tendência ( $\alpha$ ) e da sazonalidade ( $\beta$ ).

PARÂMETROS	COEFICIENTES	P-VALOR
$\beta_0$	92.069,60	6,06E-84
$\beta_1$	181,22	4,32E-32
$\alpha_1$	<b>1.302,37</b>	<b>0,31320</b>
$\alpha_2$	-7.072,72	3,89E-07
$\alpha_3$	<b>558,07</b>	<b>0,66450</b>
$\alpha_4$	<b>-723,40</b>	<b>0,57390</b>
$\alpha_5$	6.108,13	7,84E-06
$\alpha_6$	5.705,17	2,60E-05
$\alpha_7$	9.760,95	3,62E-11
$\alpha_8$	5.606,61	3,43E-05
$\alpha_9$	<b>-467,73</b>	<b>0,71560</b>
$\alpha_{10}$	<b>1.017,93</b>	<b>0,42850</b>
$\alpha_{11}$	-5.506,16	4,56E-05

Não há  $\alpha_{12}$  devido a colinearidade exata, onde o mês de dezembro foi utilizado como categoria de referência para as *dummies*.

Na Tabela 8 destacam-se os coeficientes dos parâmetros e os p-valores. Como pode ser visto, a constante, a tendência linear e as *dummies* 2, 5, 6, 7, 8 e 11 são significativas para o modelo. Apesar das *dummies* 1, 3, 4, 9 e 10 apresentarem valores não significativo, não podem ser retiradas do modelo, pois são importantes para a parte sazonal da série.

A Figura 5 apresenta o comportamento da previsão dentro da amostra para os anos 2018 e 2019 a fim de verificar a capacidade preditiva do modelo.



**Figura 5:** Série temporal com previsão dos anos de 2018 e 2019 MQO.

O Figura 5 mostra que todos os pontos da modelagem se encontram dentro do intervalo de confiança de 95% e os valores encontram-se próximos dos valores reais.

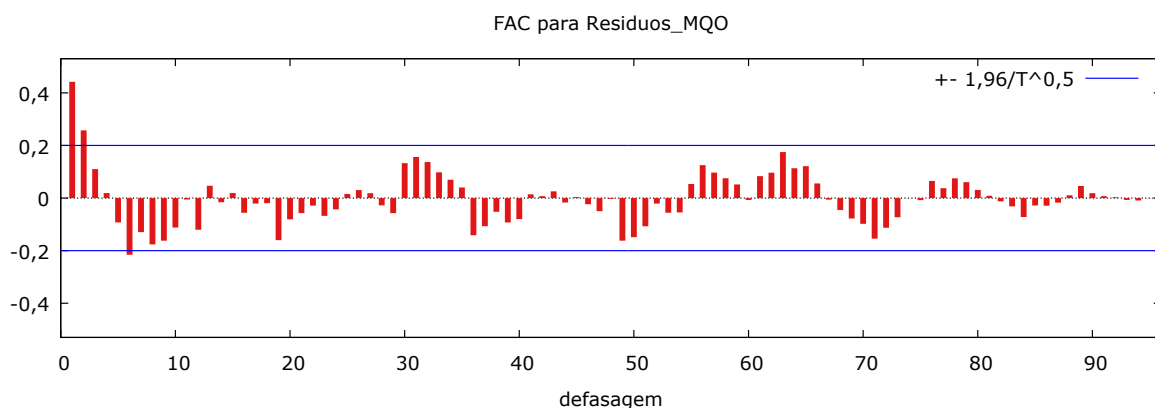
A Tabela 9 contém os valores reais e valores previstos para 2018 e 2019, além da diferença e a diferença percentual entre os dois valores, e nos dois casos o valor previsto foi usado como referência, além de fornecer o total de óbitos dos anos analisados e a variação percentual dos totais.

**Tabela 9:** Valores reais x valores previstos 2018 e 2019 MQO.

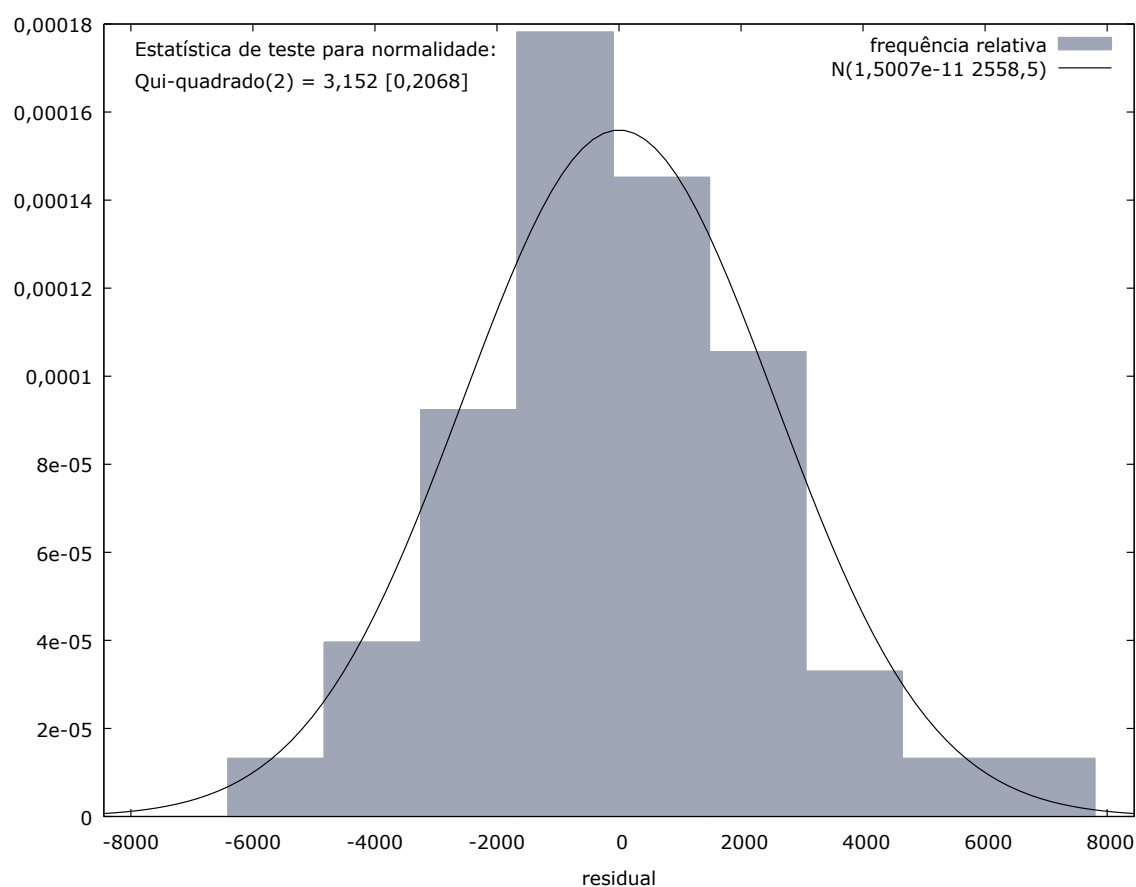
MÊS/ANO	VALOR REAL	VALOR PREVISTO	VARIAÇÃO	VARIAÇÃO %
JAN/18	106.838	110.950	4.112	3,85%
FEV/18	94.994	102.756	7.762	8,17%
MAR/18	109.390	110.568	1.178	1,08%
ABR/18	107.849	109.468	1.619	1,50%
MAI/18	115.068	116.481	1.413	1,23%
JUN/18	116.967	116.259	-708	-0,61%
JUL/18	119.675	120.496	821	0,69%
AGO/18	115.169	116.523	1.354	1,18%
SET/18	108.897	110.630	1.733	1,59%
OUT/18	107.666	112.296	4.630	4,30%
NOV/18	103.328	105.954	2.626	2,54%
DEZ/18	110.878	111.641	763	0,69%
JAN/19	110.373	113.124	2.751	2,49%
FEV/19	97.737	104.931	7.194	7,36%
MAR/19	109.549	112.743	3.194	2,92%
ABR/19	110.370	111.642	1.272	1,15%
MAI/19	118.736	118.655	-81	-0,07%
JUN/19	118.911	118.433	-478	-0,40%
JUL/19	124.885	122.670	-2.215	-1,77%
AGO/19	118.568	118.697	129	0,11%
SET/19	112.312	112.804	492	0,44%
OUT/19	111.532	114.471	2.939	2,64%
NOV/19	105.635	108.128	2.493	2,36%
DEZ/19	111.193	113.815	2.622	2,36%
<b>TOTAL</b>	<b>2.666.520</b>	<b>2.714.135</b>	<b>47.615</b>	<b>1,79%</b>

A previsão realizada na Tabela 9 tem como menor variação absoluta maio de 2019, onde a variação no número de óbitos foi de 81 mortes e a maior variação absoluta prevista foi em fevereiro de 2018, onde a variação foi de 7.762 mortes. O total da variação no número de mortes dos dois anos foi de 47.615 e variação percentual de 1,79%.

As Figura 6 e 7 e a Tabela 10 apresentam o comportamento dos resíduos.



**Figura 6:** Gráfico de dispersão da FAC dos resíduos para o MQO.



**Figura 7:** Histograma comparado à normal dos resíduos MQO.

Como pode ser observado na Figura 6, os resíduos se distribuem aleatoriamente em torno do 0 com uma variância  $\sigma^2$  constante, mesmo que as duas primeiras lags apresentem alta significância, a Figura 7 apresenta a confirmação da normalidade dos resíduos.

A Tabela 10 apresenta o teste de normalidade para os resíduos do modelo apresentados graficamente pela Figura 6.

**Tabela 10:** Teste de normalidade dos resíduos MQO.

TESTE	P-VALOR
SHAPIRO-WILK	0,2729

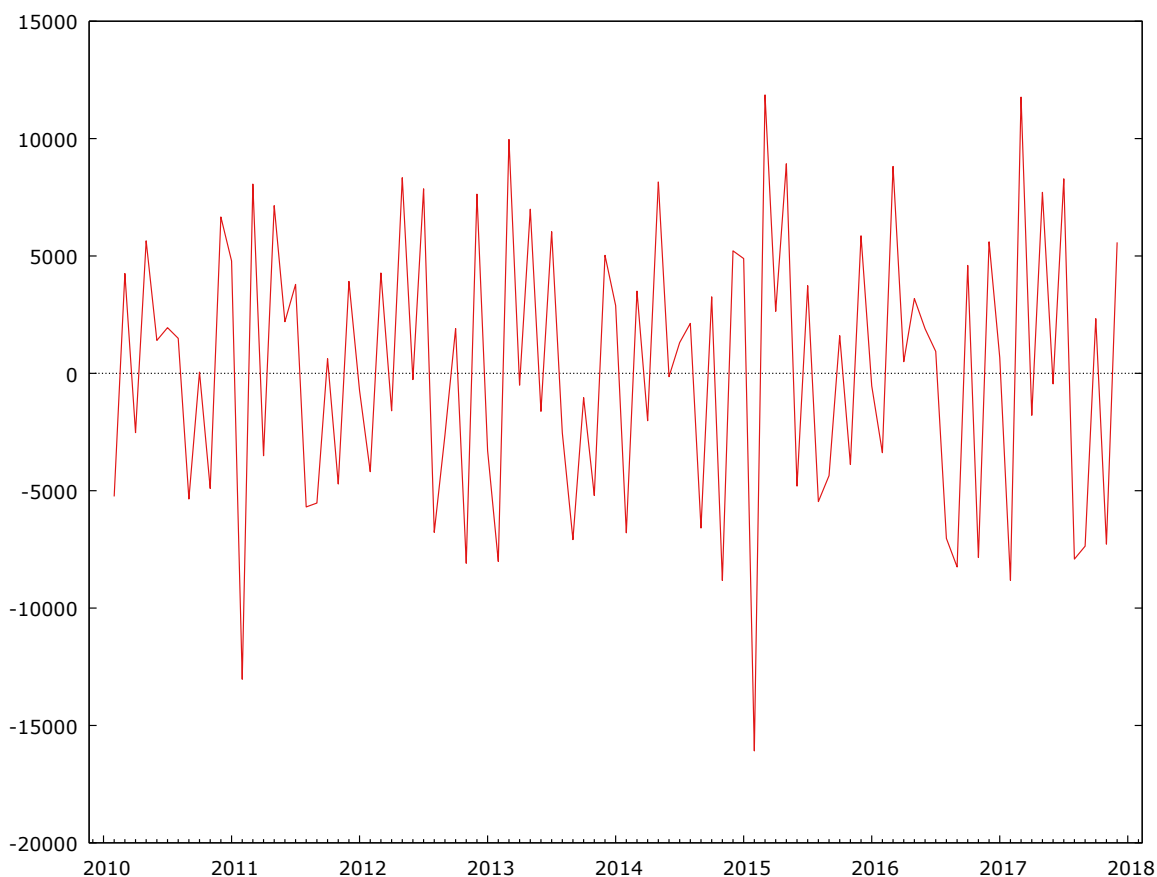
Como o p-valor foi alto (acima do nível de significância de 5%) pode-se dizer que o modelo tem um bom ajuste.

### **3.4. Modelos Box & Jenkins**

Para ajustar o modelo Box & Jenkins, a série tem que satisfazer dois critérios: o de estacionariedade e o de normalidade.

Como visto na Figura 4 a série original parece não apresentar um comportamento estacionário, por esse motivo o teste de Dickey-Fuller aumentado foi feito para verificar esse comportamento, e com o p-valor de 0,8857 não se rejeita a hipótese de que há raiz unitária, por isso a série não é estacionaria.

E para que a série se torne estacionária é preciso fazer a primeira diferença, que consta em Modelo ARIMA na Seção 2.4. A Figura 8 apresenta o gráfico da série após a realização da primeira diferença.



**Figura 8:** Primeira diferença da série temporal do número total mensal de órbitas no Brasil no período de janeiro/2010 a dezembro/2017.

Como pode ser visto na Figura 8, a série se desenvolve de forma regular ao longo do tempo em torno de uma média, característica normalmente encontrada em uma série estacionária. Para verificar se a série é estacionária o teste de Dickey-Fuller aumentado foi realizado. E o resultado do teste mostrou que o p-valor ( $4,305e-006$ ) foi menor que o grau de significância fixado no início do trabalho (0,05), assim rejeitando a hipótese nula de que há raiz unitária, representando então estacionariedade na série.

Além de verificar a estacionariedade, também foi avaliada a normalidade da série através do teste: Shapiro-Wilk. Neste teste, p-valor menor que o nível de significância indicam a rejeição da hipótese de normalidade. A Tabela 11 apresenta o resultado do teste realizado.

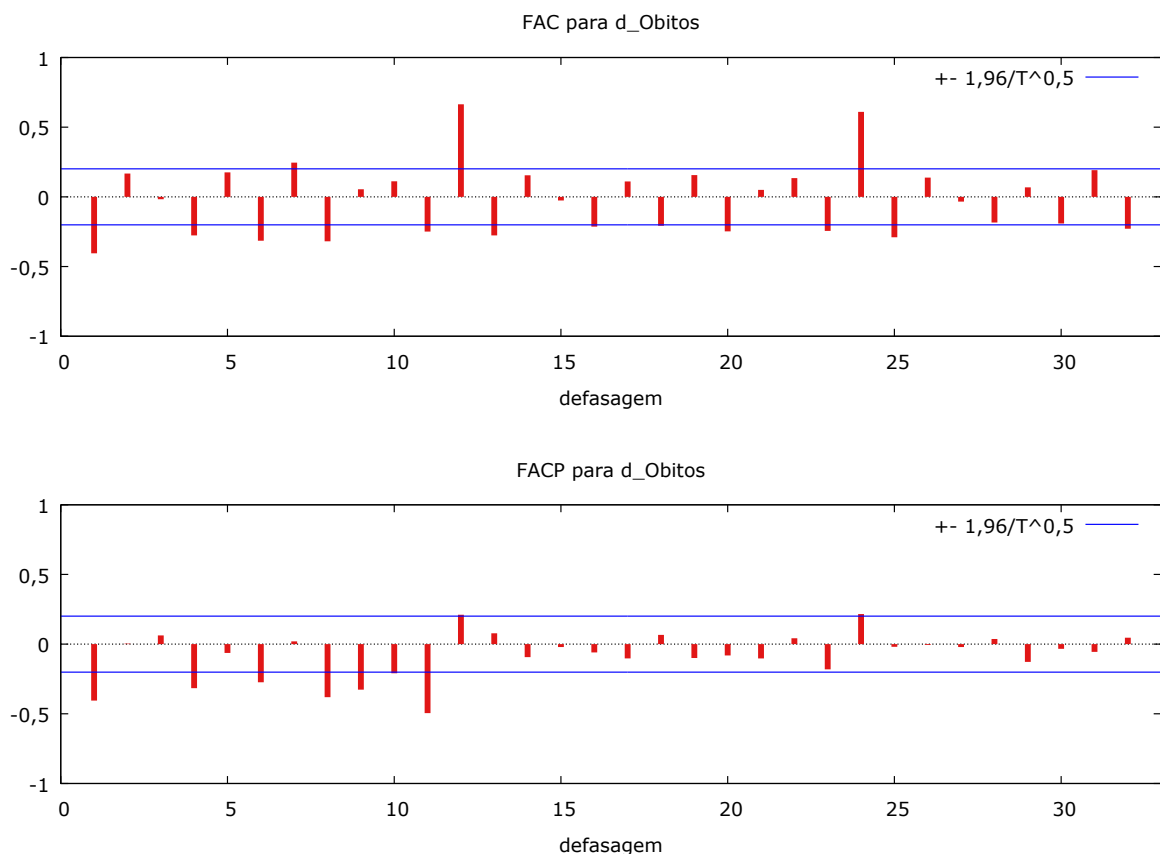
**Tabela 11:** Teste de normalidade dos dados da série temporal do número total mensal de óbitos no Brasil no período de janeiro/2010 a dezembro/2017.

TESTE	P-VALOR
SHAPIRO-WILK	0,2044

Como o p-valor do teste aplicado é maior que 0,05, então a hipótese de normalidade não é rejeitada. Tem-se, portanto, que o critério de normalidade é atendido pela série.

A análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) auxilia na escolha dos possíveis candidatos a termos autorregressivos e termos de médias móveis p e q, respectivamente, para ajuste do modelo.

A Figura 9 apresenta o correlograma com a FAC e FACP da série de primeira diferença dos óbitos, pois como visto anteriormente, a série se torna estacionária após realização de tal ação.

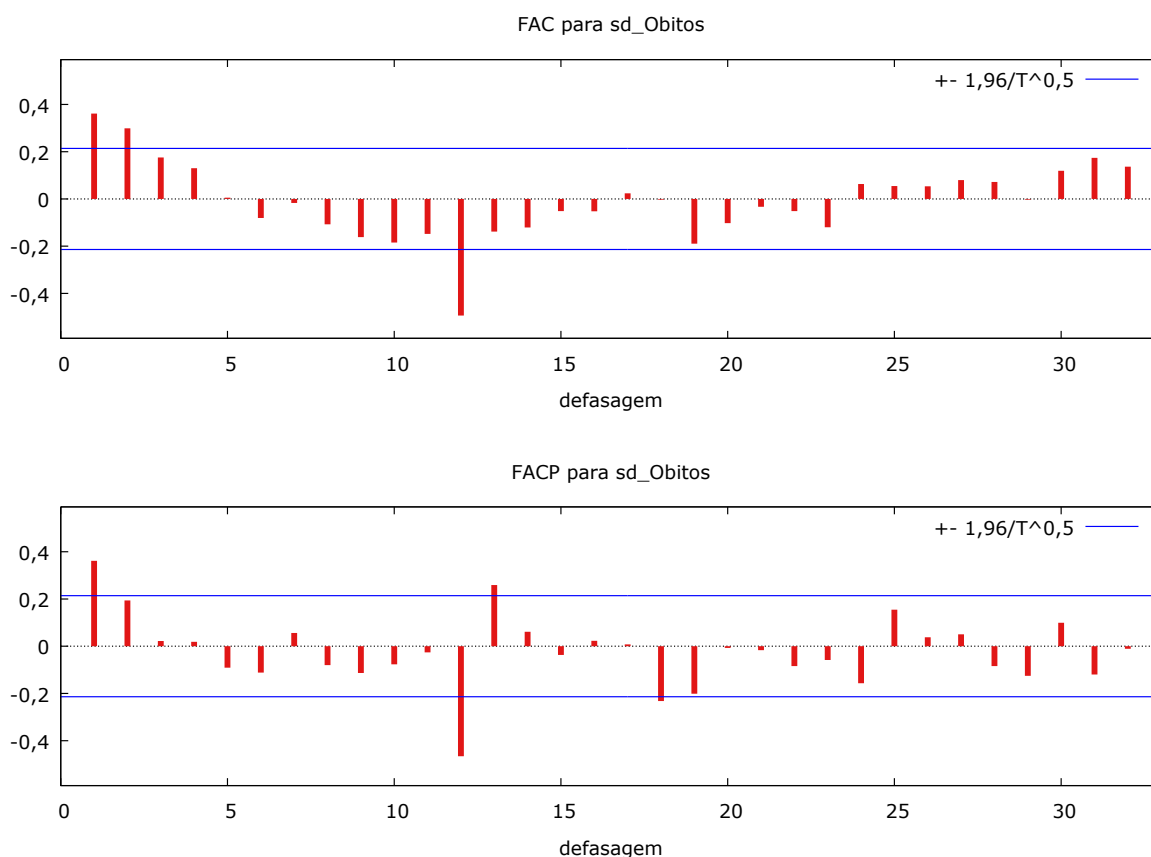


**Figura 9:** FAC e FACP da 1ª diferença dos óbitos no Brasil no período de janeiro/2010 a dezembro/2017.

É possível notar uma interferência sazonal na Figura 9, pelas significâncias dos valores da FAC nos lags 12 e 24.

Isso indica que será necessário realizar uma diferença sazonal na série.

A Figura 10 apresenta a FAC e FACP do número de óbitos com uma diferença sazonal.



**Figura 10:** FAC e FACP da 1ª diferença sazonal do número de óbitos no Brasil no período de janeiro/2010 a dezembro/2017.

Uma vez que a série possui sazonalidade, pode ser vantajoso utilizar o modelo SARIMA. Assim, observando os gráficos representados nas Figuras 9 e 10, pode-se partir dos termos:  $p = 0$ ;  $q = 1$ ;  $P = 2$  e  $Q = 4$ , pois nessas defasagens os valores ficaram fora do intervalo de confiança apresentado na FAC e na FACP. Como visto anteriormente, foi preciso diferenciar a série uma vez para 1ª diferença ( $d$ ) e uma vez de forma sazonal ( $D$ ).

De acordo com os termos encontrados, o SARIMA(0,1,1)(2,1,4) foi o modelo inicialmente utilizado para, após a realização de diversos testes, encontrar o melhor modelo ajustado Box & Jenkins.



A Tabela 12 mostra os resultados das estatísticas de aderência de vários testes realizados, partindo dos termos extraídos das Figuras 9 e 10.

**Tabela 12:** Estatísticas de aderência para os modelos SARIMA.

PARÂMETROS SARIMA	MAPE	RMSE	R <sup>2</sup> AJUSTADO	AIC	BIC	HQC
SARIMA(0,1,0)(1,1,4)	1,214%	39,81	0,9479	1.568,32	1.582,84	1.574,15
SARIMA(0,1,1)(1,1,4)	<b>1,183%</b>	39,33	<b>0,9492</b>	1.561,82	<b>1.578,76</b>	1.568,63
SARIMA(0,1,1)(1,1,3)	1,555%	40,21	0,9139	1.564,68	1.579,19	1.570,51
SARIMA(0,1,1)(2,1,4)	1,245%	40,16	0,9444	1.563,62	1.582,97	1.571,39
SARIMA(0,1,1)(2,1,3)	1,336%	41,72	0,9354	1.565,64	1.582,57	1.572,44
SARIMA(0,1,2)(2,1,3)	1,364%	42,02	0,9325	1.566,84	1.586,19	1.574,61
SARIMA(0,1,2)(1,1,4)	1,239%	40,16	0,9438	1.563,02	1.582,37	1.570,79
SARIMA(1,1,2)(1,1,4)	1,303%	41,98	0,9317	<b>1.558,88</b>	1.580,65	<b>1.567,63</b>
SARIMA(2,1,2)(1,1,4)	1,192%	<b>39,20</b>	0,9474	1.560,33	1.584,52	1.570,05

A Tabela 12 mostra 9 testes realizados com diversas variações dos termos p, q, P e Q, a fim de encontrar o melhor ajuste para a série.

Nota-se que em nenhum dos casos houve unanimidade em relação às estatísticas de aderência adotadas, isto é, não houve modelagem que apresentou melhores estatísticas em todos os testes realizados. Desta forma, será escolhido o modelo em que houve a maior quantidade de aderências satisfatórias: SARIMA(0,1,1)(1,1,4).

Importante considerar que o RMSE, AIC e HQC estão bem próximos dos melhores valores encontrados na Tabela 13.

A Tabela 13 mostra a relação dos parâmetros obtidos através do modelo escolhido.

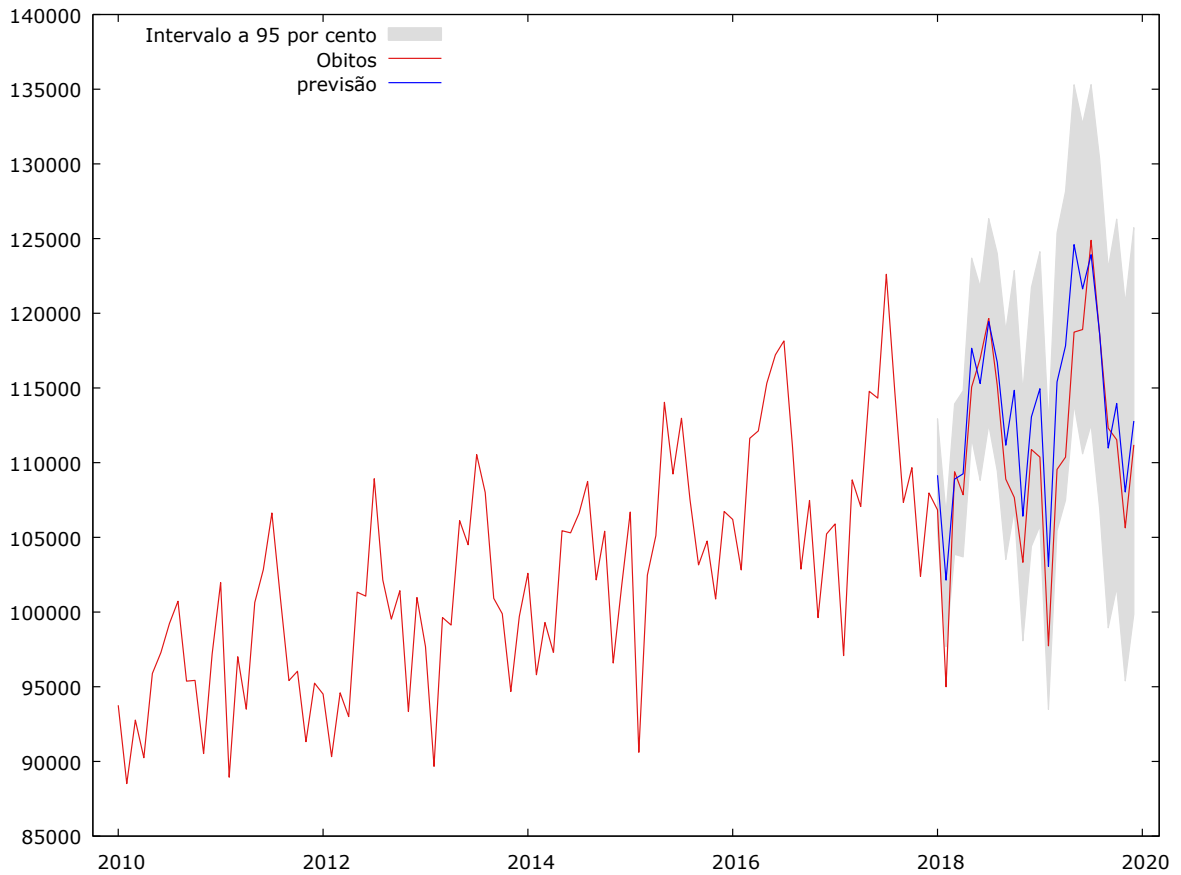
**Tabela 13:** Parâmetros estimados da modelagem SARIMA(0,1,1)(1,1,4).

PARÂMETROS	COEFICIENTES	P-VALOR
$\Phi_1$	-0,9617	0,00000
$\vartheta_1$	-0,3770	3,57e-07
$\theta_1$	<b>0,1309</b>	<b>0,22060</b>
$\theta_2$	-1,0535	3,96e-18
$\theta_3$	0,3241	0,00230
$\theta_4$	-0,8829	1,16e-11

A Tabela 13 mostra os parâmetros e p-valores dos coeficientes que resultaram no melhor ajuste, onde todos os parâmetros foram significativos (ao nível de 5%) para

o modelo, exceto o  $\Theta_1$ , porém ao testar o modelo sem este parâmetro há uma piora no modelo, como visto na Tabela 13.

A Figura 11 apresenta a previsão dos anos de 2018 e 2019 utilizando o SARIMA(0,1,1)(1,1,4).



**Figura 11:** Série temporal com previsão dos anos de 2018 e 2019 SARIMA(0,1,1)(1,1,4).

A Figura 11 representa em azul os valores previstos pelo modelo para os anos de 2018 e 2019, e o fato de não ultrapassarem os limites da área sombreada significa que toda a previsão está dentro do intervalo de confiança.

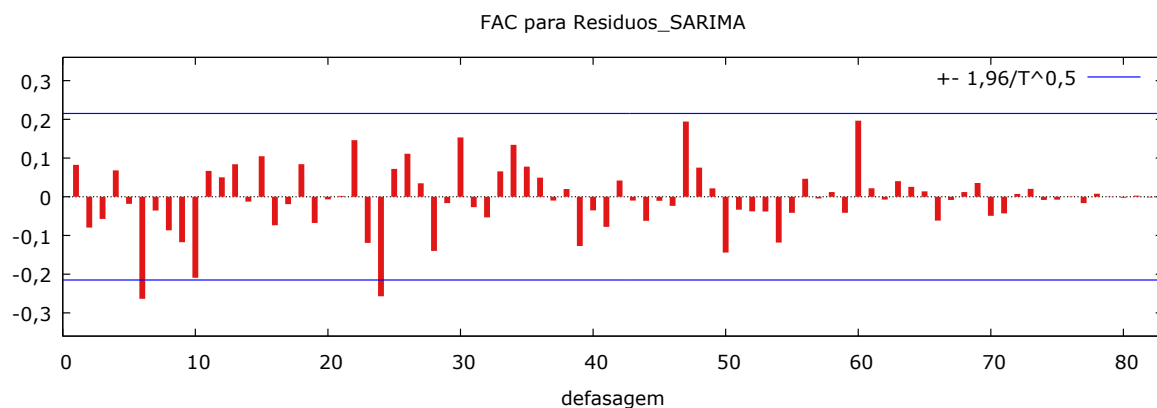
A Tabela 14 destaca os valores previstos e reais dos anos de 2018 e 2019, de acordo com o modelo ajustado SARIMA.

**Tabela 14:** Valores reais x valores previstos 2018 e 2019 SARIMA(0,1,1)(1,1,4).

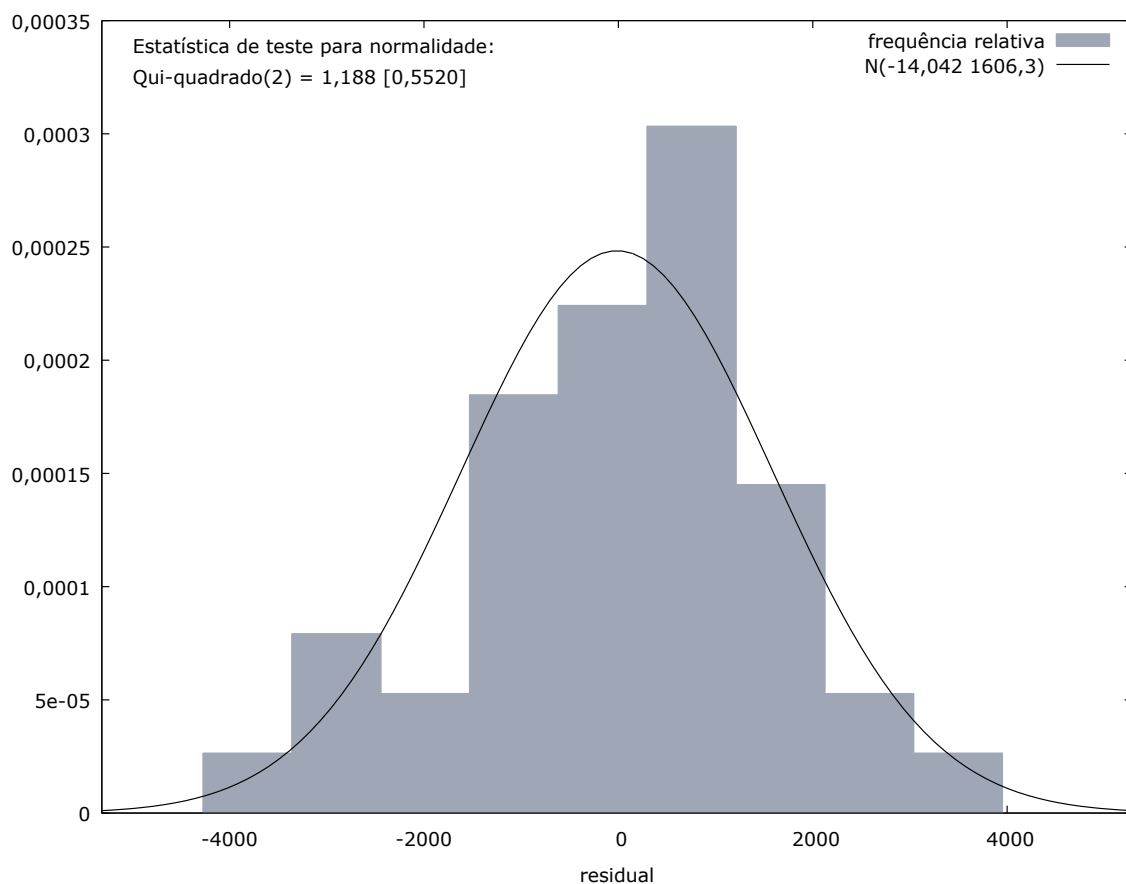
<b>MÊS/ANO</b>	<b>VALOR REAL</b>	<b>VALOR PREVISTO</b>	<b>VARIAÇÃO</b>	<b>VARIAÇÃO (%)</b>
JAN/18	106.838	109.158	2.320	2,17%
FEV/18	94.994	102.138	7.144	7,52%
MAR/18	109.390	108.900	-490	-0,45%
ABR/18	107.849	109.252	1.403	1,30%
MAI/18	115.068	117.660	2.592	2,25%
JUN/18	116.967	115.294	-1.673	-1,43%
JUL/18	119.675	119.442	-233	-0,19%
AGO/18	115.169	116.721	1.552	1,35%
SET/18	108.897	111.171	2.274	2,09%
OUT/18	107.666	114.844	7.178	6,67%
NOV/18	103.328	106.419	3.091	2,99%
DEZ/18	110.878	113.069	2.191	1,98%
JAN/19	110.373	114.950	4.577	4,15%
FEV/19	97.737	103.054	5.317	5,44%
MAR/19	109.549	115.406	5.857	5,35%
ABR/19	110.370	117.820	7.450	6,75%
MAI/19	118.736	124.598	5.862	4,94%
JUN/19	118.911	121.631	2.720	2,29%
JUL/19	124.885	123.923	-962	-0,77%
AGO/19	118.568	118.611	43	0,04%
SET/19	112.312	110.978	-1.334	-1,19%
OUT/19	111.532	113.962	2.430	2,18%
NOV/19	105.635	108.035	2.400	2,27%
DEZ/19	111.193	112.791	1.598	1,44%
<b>TOTAL</b>	<b>2.666.520</b>	<b>2.729.829</b>	<b>63.309</b>	<b>2,37%</b>

Dentre as previsões tabeladas acima, a que apresentou menor variação absoluta em relação aos dados reais foi a do mês de agosto de 2019 (43 mortes), e a maior variação absoluta se deu no mês de abril de 2019 (7.450 mortes). Já em relação ao total de óbitos ocorridos em 2018 e 2019, tem-se que a previsão foi de 63.309 óbitos a mais que os dados reais, uma variação de 2,37%.

As Figuras 12 e 13 e a Tabela 15 apresentam o comportamento dos resíduos.



**Figura 12:** Gráfico de dispersão da FAC dos resíduos para o SARIMA.



**Figura 13:** Histograma comparado à normal dos resíduos SARIMA.

A Tabela 15 apresenta o teste de normalidade para os resíduos do modelo apresentados graficamente pela Figura 12.

**Tabela 15:** Teste de Normalidade dos resíduos SARIMA.

TESTE	P-VALOR
SHAPIRO-WILK	0,4244

Como pode ser observado na Figura 12, os resíduos se distribuem aleatoriamente em torno do 0 com uma variância  $\sigma^2$  constante e na Figura 13, e Tabela 15 apresentam a confirmação da normalidade, o que indica que o modelo tem um bom ajuste.

### 3.5. Escolha do modelo para fazer a previsão

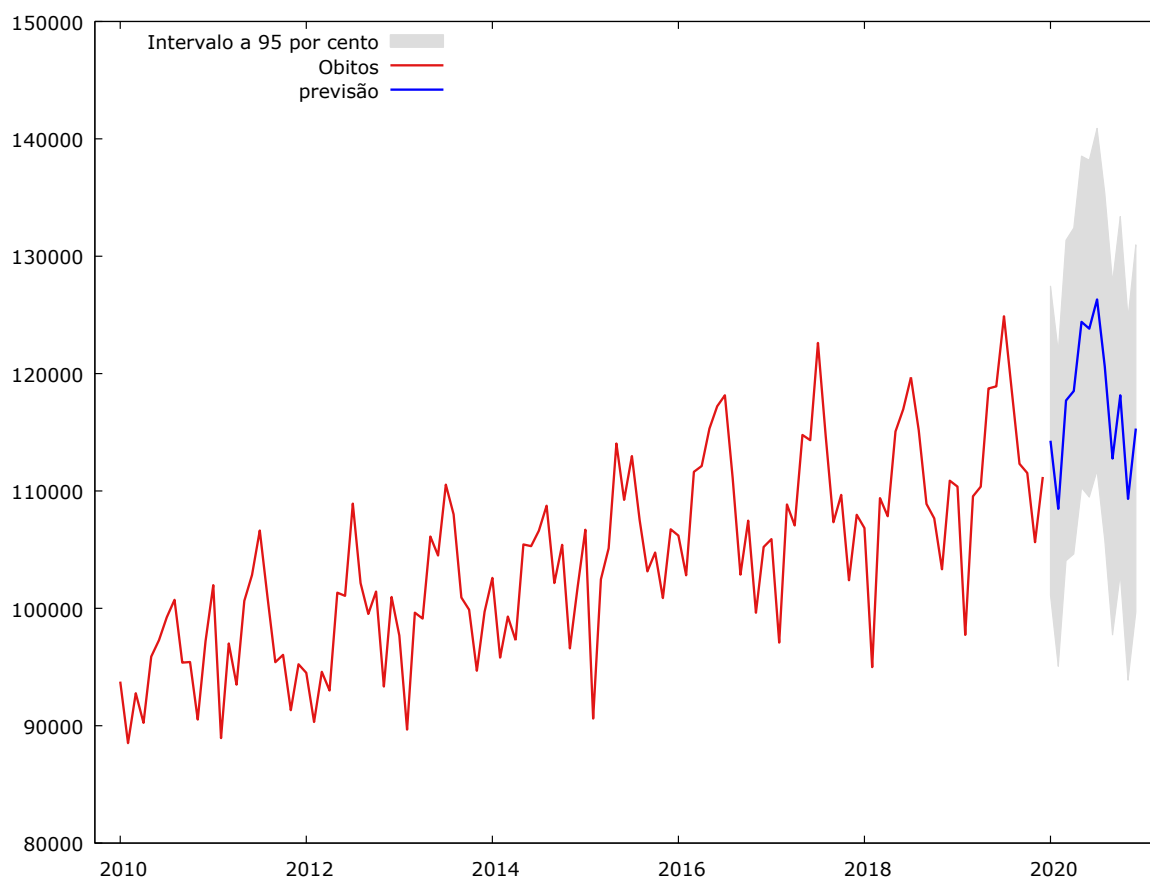
Dentre os modelos apresentados nas Seções 3.3 e 3.4, faz-se necessário escolher aquele que melhor se ajustou à série. Para isso, foram comparadas as estatísticas de aderência do modelo de regressão linear e do Box & Jenkins, presentes na Tabela 16.

**Tabela 16:** Estatísticas de aderência para o modelo selecionado.

MODELOS	MAPE	RMSE	R <sup>2</sup> AJUSTADO	AIC	BIC	HQC
CONSTANTE, T E DUMMIES	1,83%	48,77	0,8767	1.791,13	1.824,47	1.804,60
SARIMA(0,1,1)(1,1,4)	1,18%	39,33	0,9492	1.561,82	1.578,76	1.568,63

De acordo com a Tabela 16, o modelo que mais se adequa à série, por apresentar as melhores estatísticas de aderência, é o modelo de Box & Jenkins com o SARIMA (0,1,1)(1,1,4). Este será, portanto, o modelo selecionado.

A Figura 14 mostra as quantidades de óbitos esperadas para o ano de 2020, previstas pelo modelo selecionado, caso não houvesse a pandemia.



**Figura 14:** Previsão para 2020 caso não houvesse a pandemia.

A Figura 14 representa em azul os valores previstos pelo modelo para o ano 2020, e o fato de não ultrapassarem os limites da área sombreada significa que toda a previsão está dentro do intervalo de confiança.

A Tabela 17 contém a quantidade de óbitos estimada pontualmente pelo modelo e seu intervalo de confiança a 95% selecionado para cada mês do ano de 2020.

**Tabela 17:** N° de óbitos esperados para 2020 caso não houvesse pandemia.

MÊS	ESTIMATIVA PONTUAL	ERRO PADRÃO	ESTIMATIVA INTERVALAR A 95%
JANEIRO	114.280	6.718	(101.112 - 127.447)
FEVEREIRO	108.475	6.843	(95.063 - 121.888)
MARÇO	117.706	6.966	(104.053 - 131.359)
ABRIL	118.519	7.086	(104.629 - 132.408)
MAIO	124.407	7.205	(110.285 - 138.528)
JUNHO	123.827	7.322	(109.477 - 138.177)
JULHO	126.319	7.436	(111.744 - 140.894)
AGOSTO	120.594	7.549	(105.797 - 135.391)
SETEMBRO	112.763	7.661	(97.748 - 127.778)
OUTUBRO	118.151	7.771	(102.921 - 133.381)
NOVEMBRO	109.324	7.879	(93.881 - 124.766)
DEZEMBRO	115.317	7.986	(99.666 - 130.969)
<b>TOTAL</b>	<b>1.409.682</b>	-	-

A Tabela 17 mostra que a soma dos valores mensais na estimação pontual é de 1.409.682 óbitos no ano de 2020 caso não houvesse a pandemia do coronavírus. É possível verificar também que todos os valores da estimação pontual se encontram dentro do intervalo de confiança a um nível de confiança de 95%.

### 3.6. Impacto da COVID-19

A Tabela 18 destaca a assertividade do modelo através do erro previsto.

**Tabela 18:** Efetividade do modelo através do erro de previsão mês a mês para o Brasil em 2020.

MESES DE 2020	ÓBITOS PRELIMINARES (OPENDATUSUS)	ÓBITOS POR COVID-19 (PORTAL CORONAVÍRUS)	PRÉVIA DO NÚMERO DE ÓBITOS SEM A COVID (ÓBITOS PRELIMINARES - ÓBITOS POR COVID-19)	ÓBITOS PREVISTOS (VALORES PREVISTOS PELO SARIMA)	ERRO DE PREVISÃO (PRÉVIA DOS ÓBITOS SE COVID - ÓBITOS PREVISTOS)
JANEIRO	113.292	0	113.292	114.280	-988
FEVEREIRO	106.229	0	106.229	108.475	-2.246
MARÇO	119.998	201	119.797	117.706	2.091
ABRIL	125.726	5.700	120.026	118.519	1.507
MAIO	156.544	23.413	133.131	124.407	8.724
JUNHO	142.337	30.280	112.057	123.827	-11.770
JULHO	147.109	32.881	114.228	126.319	-12.091
AGOSTO	142.756	28.906	113.850	120.594	-6.744
SETEMBRO	127.773	22.571	105.202	112.763	-7.561
OUTUBRO	128.250	15.932	112.318	118.151	-5.833
NOVEMBRO	118.723	13.236	105.487	109.324	-3.837
DEZEMBRO	131.351	21.829	109.522	115.317	-5.795

A Tabela 18 mostra que todos os valores prévios mês a mês para o ano de 2020 encontram-se dentro do intervalo de confiança do modelo estimado com o melhor ajuste, como pode ser visto na Tabela 17. Com isso os erros previstos não ultrapassam os limites do intervalo de confiança, e isso evidencia mais uma vez que o modelo pode ter acertado ao prever o número de óbitos para 2020 no Brasil caso não houvesse a pandemia.

A Tabela 19 revela o impacto que a pandemia da COVID-19 causou no aumento do número de óbitos mês a mês no Brasil em 2020, dado os valores mensais previstos pelo modelo para o ano em questão.

**Tabela 19:** Impacto da COVID-19 no aumento do número de óbitos previstos para o Brasil em 2020.

<b>MESES DE 2020</b>	<b>ÓBITOS PREVISTOS (VALORES PREVISTOS PELO SARIMA)</b>	<b>ÓBITOS POR COVID-19 (PORTAL CORONAVÍRUS)</b>	<b>IMPACTO DA COVID-19 NO AUMENTO DO NÚMERO DE ÓBITOS PREVISTOS MÊS A MÊS EM 2020</b>
<b>JANEIRO</b>	114.280	0	0,00%
<b>FEVEREIRO</b>	108.475	0	0,00%
<b>MARÇO</b>	117.706	201	0,17%
<b>ABRIL</b>	118.519	5.700	4,81%
<b>MAIO</b>	124.407	23.413	18,82%
<b>JUNHO</b>	123.827	30.280	24,45%
<b>JULHO</b>	126.319	32.881	26,03%
<b>AGOSTO</b>	120.594	28.906	23,97%
<b>SETEMBRO</b>	112.763	22.571	20,02%
<b>OUTUBRO</b>	118.151	15.932	13,48%
<b>NOVEMBRO</b>	109.324	13.236	12,11%
<b>DEZEMBRO</b>	115.317	21.829	18,93%
<b>TOTAL</b>	1.409.682	194.949	13,83%

A Tabela 19 mostra que a COVID-19 teve um impacto de 13,83% no aumento do número de óbitos previstos para o Brasil em 2020.



#### 4. CONCLUSÃO

Em 2020 a COVID-19 chegou ao Brasil e de acordo com o portal CORONAVÍRUS//BRASIL, fez cerca de 194.949 (cento e noventa e quatro mil, novecentos e quarenta e nove) vítimas até o mês de dezembro do mesmo ano. Por este motivo o presente trabalho visou prever o número de óbitos no Brasil em 2020 caso não houvesse a pandemia, para melhor entendimento do seu impacto no aumento das mortes no país.

Com os dados fornecidos pelo DATASUS, dois modelos estatísticos foram estimados para prever os valores mensais de óbitos para o Brasil em 2020: o modelo de regressão linear, com a modelagem dos Mínimos Quadrados Ordinários (MQO) e o modelo de Box & Jenkins, com a modelagem do SARIMA. Após estimados os modelos, foram feitas algumas estatísticas de aderência para a escolha do modelo melhor ajustado para a série, dentre os dois modelos apresentados: o  $R^2$  ajustado; a MAPE; a RMSE; o AIC; o BIC e o HQC foram essenciais para a escolha da modelagem mais assertiva. E o que melhor se ajustou aos dados, de acordo com as estatísticas de aderência foi o modelo de Box & Jenkins.

Com isso, verificou-se que o número de óbitos esperados para o Brasil em 2020 caso não houvesse a pandemia foi de 1.409.682. Após a previsão realizou-se a análise dos erros de previsão, e mais uma vez o modelo se mostrou adequado, dado que todos os valores encontrados para esses erros estavam dentro do intervalo de confiança a um nível de confiabilidade de 95%. Por fim, efetuou-se a verificação do impacto da COVID-19 no número total de óbitos esperados para o Brasil em 2020, onde observou-se um acréscimo de 13,83% nos óbitos totais para o ano de estudo, este aumento pode ser explicado em parte pela pandemia.

Para mais esclarecimentos sobre o impacto da pandemia no Brasil em 2020, estudar o número de mortes por faixa etária pode ser relevante.

## REFERÊNCIAS

- 1 COSTA, Isabela Bispo Santos da Silva, BITTAR Cristina Salvadori, RIZK Stephanie Itala, *et al.* **O Coração e a COVID-19: O que o Cardiologista Precisa Saber**. Arq. Bras. Cardiol. vol.114, no.5, São Paulo, maio 2020.
- 2 CLERKIN KJ, FRIED JA, RAIKHELKAR J, Sayer G, GRIFFIN JM, MASOUMI A, *et al.* **Coronavirus disease 2019 (COVID-19) and cardiovascular disease**. Circulation. 2020 Mar 21.
- 3 GANATRA S HS, HAMMOND SP, NOHRIA A. **The novel coronavirus disease (COVID-19) threat for patients with cardiovascular disease and cancer**. JACC CardioOncology. 2020 Mar.
- 4 GONÇALVES, Ferreira Netto, R., DO NASCIMENTO CORRÊA, J. W. **EPIDEMIOLOGIA DO SURTO DE DOENÇA POR CORONAVÍRUS (COVID-19)**. DESAFIOS - Revista Interdisciplinar Da Universidade Federal Do Tocantins, 7(Especial-3), 18-25. 2020.
- 5 **Orientações da OMS para prevenção da COVID-19**. Disponível em: <https://sbpt.org.br/portal/covid-19-oms/>. Acesso em 07/04/2021
- 6 **DATASUS**. Disponível em <https://datasus.saude.gov.br/sobre-o-datasus/>. Acesso em 10/04/2021
- 7 **openDataSUS** <https://opendatasus.saude.gov.br/about>. Acesso em 28/08/2021
- 8 MORETTIN, Pedro A., TOLOI, Clélia M. C. **Análise de Séries Temporais**, São Paulo: Edgard Blucher. 1981.
- 9 MORETTIN, Pedro A., TOLOI, Clélia M. C. **Análise de Séries Temporais**, 2ª ed., São Paulo: Edgard Blucher. 2004.

- 10 MORETTIN, Pedro A., TOLOI, Clélia M. C. **Análise de Séries Temporais**, 2ª ed., São Paulo: Edgard Blucher. 2006.
- 11 GRANGER, C. W. J., NEWBOLD, Paul. **Forecasting Economic Time Series**. Journal of the American Statistical Association, V.83, N. 402, p 574, 1988.
- 12 RPUBS. **Correlação, Autocorrelação e Autocorrelação Parcial**. Disponível em < [https://rpubs.com/hudsonchavs/fac\\_facp](https://rpubs.com/hudsonchavs/fac_facp) >. Acesso em 16/04/2021
- 13 SHAPIRO, S. S.; WILK, M. B. **An analysis of variance test for normality**. Biometrika, JSTOR, v. 52, n. 3/4, p. 591–611, 1965.
- 14 FISCHER, S. **Séries Univariantes de Tempo-Metodologia de box & jenkins**. Fundação de Economia e Estatística, p. 450, 1982.
- 15 FRANCO, G. C. **Apostila sobre modelagem box and jenkins**. Depto. Estatística - UFMG
- 16 CAMELO, H. d. N. *et al.* **Modelagem box-jenkins aplicada a previsão de velocidade do vento em regiões do nordeste brasileiro para fins de geração eólica**. Ciência e Natura, 2018.
- 17 AKAIKE, H. **A new look at the statistical model identification**. *IEEE transactions on automatic control*, IEEE, v. 19, n. 6, p. 716{723, 1974.
- 18 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.
- 19 EHLERS, R. S. **Análise de séries temporais**. Universidade Federal do Paraná, 2007.
- 20 **CORONAVÍRUS//BRASIL, Painel Coronavírus** <<https://covid.saude.gov.br/>>. Acesso em 25/08/2021.