UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO ACADÊMICO DO AGRESTE
NÚCLEO DE TECNOLOGIA
CURSO DE ENGENHARIA DE PRODUÇÃO

ITALO IVELDO LUIZ DOS SANTOS PEREIRA

UMA ANÁLISE DE SÉRIES TEMPORAIS DO NÚMERO DE SUICÍDIO NO BRASIL ENTRE O PERÍODO DE 1996 ATÉ 2020.

Caruaru
2022

ITALO IVELDO LUIZ DOS SANTOS PEREIRA

UMA ANÁLISE DE SÉRIES TEMPORAIS DO NÚMERO DE SUICÍDIO NO BRASIL ENTRE O PERÍODO DE 1996 ATÉ 2020.

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Produção do Centro Acadêmico do Agreste – CAA, da Universidade Federal de Pernambuco, como requisito parcial para aprovação na disciplina Projeto Final de Curso.

Área de concentração: Series temporais

Orientador: Profº. Dr. Thyago Celso Cavalcante Nepomuceno

Caruaru

2022

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Pereira, Italo Iveldo Luiz dos Santos.
Uma análise de séries temporais do número de suicídios no Brasil entre o periodo de 1996 até 2020 / Italo Iveldo Luiz dos Santos Pereira. - Caruaru, 2022.
89 : il., tab.

Orientador(a): Thyago Celso Cavalcante Nepomuceno
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro Acadêmico do Agreste, Engenharia de Produção - Bacharelado, 2022.
Inclui referências, apêndices.

1. Séries temporais. 2. ARIMA. 3. Suavização Exponêncial. 4. Suicídios. I. Nepomuceno, Thyago Celso Cavalcante. (Orientação). II. Título.

620 CDD (22.ed.)

ITALO IVELDO LUIZ DOS SANTOS PEREIRA

UMA ANÁLISE DE SÉRIES TEMPORAIS DO NÚMERO DE SUICÍDIO NO BRASIL ENTRE O PERÍODO DE 1996 ATÉ 2020

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Produção do Centro Acadêmico do Agreste da Universidade Federal de Pernambuco - UFPE, na modalidade de monografia, como requisito parcial para a obtenção do grau de Bacharel em Engenharia de Produção.

Aprovado em: 23/05/2022

BANCA EXAMINADORA

_____

Prof. Dr. Thyago Celso Cavalcante Nepomuceno (Orientador)

Universidade Federal de Pernambuco

_____

Prof. Dr. Lúcio Camara e Silva (Examinador Interno)

Universidade Federal de Pernambuco

_____

Dr. Henrique Pinto dos Santos Zaidan (Examinador Externo)

Universidade Federal Rural de Pernambuco

RESUMO

Com o aumento de fatores de risco para comportamentos suicidas como ansiedade e depressão devido a pandemia da COVID-19. Este trabalho apresenta uma análise descritiva e temporal sobre o número de suicídios no Brasil durante os períodos de 1996 a 2020 e avalia quais foram os efeitos da pandemia sobre este tema a partir da construção de modelos ARIMA e de suavização exponencial. Tal questionamento coloca em evidência também qual a eficiência de ações para o combate e prevenção de suicídios como a campanha Setembro Amarelo. Conclui-se que não existem evidências de que tal doença tenha aumentado ou diminuído os números de mortes autoprovocadas no Brasil, tal qual, não existem evidências da eficácia da Campanha Setembro amarelo.

Palavras-chave: séries temporais; suicídios; ARIMA; suavização exponencial

ABSTRACT

With the increase in risk factors for suicidal behaviors such as anxiety and depression due to the COVID-19 pandemic. This paper presents a descriptive and temporal analysis of the number of suicides in Brazil during the periods from 1996 to 2020 and evaluates the effects of the pandemic on this theme through the construction of ARIMA models and exponential smoothing. Such questioning also highlights the efficiency of actions to combat and prevent suicides, such as the Yellow September campaign. We conclude that there is no evidence that this disease has increased or decreased the number of self-caused deaths in Brazil, just as there is no evidence of the effectiveness of the Yellow September campaign.

Keywords: time series; suicides; ARIMA; exponential smoothing

SUMÁRIO

1	**INTRODUÇÃO**	**9**
1.1	JUSTIFICATIVA	11
1.2	OBJETIVO GERAL	12
1.3	OBJETIVO ESPECÍFICOS	12
2	**REFERENCIAL TEÓRICO**	**13**
2.1	SERIES TEMPORAIS	13
2.2	MODELO DE SERIES TEMPORAIS	15
2.2.1	**Médias Móveis**	16
2.2.2	**Suavização Exponencial Simples**	17
2.2.3	**Suavização Exponencial Linear de Holt (SEH)**	18
2.2.4	**Suavização exponencial de Holt-Winters**	19
2.2.5	**Modelos ARIMA**	20
2.2.5.1	Modelos autorregressivos (AR) e de média móveis (MA)	21
2.2.5.2	Modelos ARIMA e SARIMA	22
2.2.5.3	Cenários Temporais	22
2.3	APLICAÇÕES DE SÉRIES TEMPORAIS EM PESQUISAS SOBRE SUICÍDIOS	23
2.3.1	**Aplicações de séries temporais em pesquisas de suicídio no Brasil**	24
2.3.2	**Aplicações de séries temporais em pesquisas analisando o impacto do COVID no número de suicídios**	25
3	**METODOLOGIA**	**27**
4	**DESENVOLVIMENTO**	**30**
4.1	ANÁLISE DESCRITIVA GERAL	30
4.2	ANÁLISE DESCRITIVA POR CATEGORIAS	31
4.3	ANÁLISE DE SÉRIES TEMPORAIS	37
4.4	DEFINIÇÂO DO MELHOR MODELO PARA PREVISÃO DE SUICÍDIOS NO BRASIL	41
4.4.1	**Escolha de um modelo pelo método de suavização exponencial**	41

4.4.1.1 Análise dos resíduos..... 42
4.4.1.2 Previsão com dados de teste..... 43
4.4.2 Escolha de um modelo utilizando o método ARIMA..... 45
4.4.2.1 Função de autocorrelação e autocorrelação parcial..... 47
4.4.2.2 Modelos estimados..... 48
4.4.2.3 Análise dos resíduos..... 49
4.4.2.4 Previsão com dados de teste..... 51
4.4.3 Definindo o melhor modelo, ARIMA vs ETS..... 52
4.5 IMPACTOS DO SETEMBRO AMARELO NOS NÚMEROS DE SUICÍDIOS..... 53
4.6 IMPACTOS DO COVID-19 NOS ANOS DE 2020..... 57
5 CONCLUSÃO..... 63
REFERÊNCIAS..... 66
APÊNDICE A – CÓDIGO FONTE..... 72

1. INTRODUÇÃO

Segundo a OMS (OMS,2018) mais de 800 mil pessoas no mundo morrem de suicídio todos os anos. Além disso, no Brasil (Ministério da Saúde, 2021) o suicídio foi a quarta maior causa de morte entre adultos jovens do sexo masculino de 20 a 39 anos. Entre os períodos de 2007 a 2016 foram registrados 106 mil óbitos, com uma taxa de 5,3/100 mil habitantes.

Ainda segundo a OMS, o Brasil é o campeão mundial em casos de transtornos de ansiedade e ocupa a quinta colocação no ranking mundial em casos de depressão, sendo essas duas doenças fortes candidatas a evoluir e gerar casos de comportamento suicida. (2017)

Além das vidas perdidas o suicídio também tem impacto direto e indireto nas vidas de outras pessoas que se relacionam de alguma forma com esse indivíduo através de núcleos como família e trabalho por exemplo, estima-se que a cada suicídio realizado, pelo menos seis pessoas sofrem algum impacto negativo. (BOTEGA, 2014; MINAYO; MENEGHEL; CAVALCANTE, 2012).

Temas relacionados a saúde mental merecem ainda mais atenção atualmente, tendo em vista que desde janeiro de 2020 o mundo vive uma pandemia. Pesquisas apontam que o coronavírus está potencializando sentimentos de angústia, ansiedade e depressão, principalmente em profissionais de saúde. (OMS, 2022).

O COVID-19 ampliou diversos fatores de risco ligados ao suicídio como isolamento em decorrência do distanciamento social, desemprego, crises econômicas e diversos traumas. Pesquisas realizados no Chile, Brasil, Peru e Canadá apontam que após 18 meses de pandemia 50% das pessoas que participaram relataram que sua saúde mental havia piorado. (OMS, 2021)

Nessa perspectiva, diante dos impactos negativos que os suicídios causam para a sociedade, o presente trabalho visa elencar quais as aplicações de séries

temporais para análise de óbitos por suicídios no Brasil entre os períodos de 1996 até 2020.

Portanto, indaga-se: Quais os impactos da pandemia da COVID-19 na distribuição dos suicídios no Brasil? Então, o objetivo geral da presente pesquisa é aplicar séries temporais para análise de suicídios no Brasil, entre os períodos de 1996 até 2020.

Para tanto, foram definidos os seguintes objetivos específicos: Identificar as principais características dos indivíduos que cometeram suicídios; identificar padrões temporais a partir dos componentes de tendência, sazonalidade e ciclo; Desenvolver modelos de previsão a partir do método de suavização exponencial e ARIMA; selecionar o melhor modelo de previsão a partir das métricas AIC e RMSE; esclarecer se o COVID-19 trouxe impactos nos números de óbitos por suicídios no Brasil no ano de 2020; Investigar a eficácia da campanha Setembro Amarelo no combate dos suicídios; Sugerir o melhor mês para a campanha.

Partisse da hipótese de que existem diversas contribuições de séries temporais para o tema estudado, tendo em vista que os dados históricos disponíveis permitem avaliar tendências, sazonalidades e determinar se houve um aumento no número de óbitos por suicídio devido a pandemia.

Assim, realiza-se uma pesquisa de finalidade básica estratégica, com objetivo descritivo, sob o método hipotético-dedutivo, por meio de uma abordagem quantitativa a respeito do tema de séries temporais e suicídios.

No quarto capitulo, subtópico 4.1 realiza-se uma análise descritiva com objetivo de determinar a taxa de suicídios anual e mensal no Brasil.

No quarto capitulo, subtópico 4.2, realiza-se uma análise exploratória com foco em entender quais os perfis dos indivíduos que cometem suicídio, bem como sua idade, sexo, estado civil e as principais causas de óbito.

No quarto capitulo, subtópico 4.3, realiza-se uma análise temporal da série histórica disponibilizada pelo SUS, indicando seus componentes de tendência e

sazonalidade. Bem como esclarecer quais meses possuem os maiores e menores número de óbitos.

No quarto capitulo, subtópico 4.4, utiliza-se a métrica do erro quadrático médio e o critério de Akaike para encontrar o melhor modelo entre o método ARIMA e de suavização exponencial com objetivo de prever os números de óbitos por suicídio.

No quarto capitulo, subtópicos 4.5 e 4.6 aplica-se o modelo de série temporal encontrado com a finalidade de avaliar se a pandemia do COVID-19 elevou o número de suicídios no Brasil em 2020 e quais os impactos da campanha Setembro amarelo para combate de ações suicidas.

Ao final, conclui-se que os objetivos são atendidos e a hipótese levantada é aceita, indicando que que não houve aumento no número de suicídios em 2020, mesmo com a ocorrência da pandemia.

1.1 JUSTIFICATIVA

Com a pandemia do COVID-19 tonou-se mais frequente temas relacionados a saúde mental e abriu-se diversos discursões sobre seus impactos no número de suicídio no Brasil e no mundo. Somado a isso o país vive uma crise econômica e política que traz diversas incertezas gerando gatilhos como ansiedade e depressão que podem aumentar as ocorrências de lesões autoprovocadas voluntariamente com morte.

O estudo dos dados históricos de suicídios do Brasil por meio de séries temporais permite mensurar os impactos de eventos, como crises e pandemias, além de encontrar padrões de comportamento suicida, tendências, sazonalidades e a criação de modelos de previsão que podem auxiliar no processo de tomada de decisão de órgãos governamentais de saúde e a criação de novas políticas públicas. Desta forma, se justifica o desenvolvimento de um estudo de séries temporais dos dados históricos de suicídios do Brasil.

1.2 OBJETIVO GERAL

- Avaliar a evolução e o impacto da Pandemia da COVID-19 na distribuição dos suicídios no Brasil considerando a série história entre os anos de 1996 a 2020.

1.3 OBJETIVOS ESPECIFICOS

- Identificar as principais características dos indivíduos que cometeram suicídios no Brasil.
- Identificar padrões temporais a partir dos componentes de tendência, sazonalidade e ciclo.
- Desenvolver modelos de previsão a partir do método de suavização exponencial e ARIMA que possa prever o número de mortes por suicídio no Brasil.
- Selecionar o melhor modelo de previsão a partir das métricas AIC e RMSE.
- Esclarecer se o COVID-19 trouxe impactos nos números de óbitos por suicídios no ano de 2020.
- Investigar a eficácia da campanha Setembro amarelo no combate dos suicídios no Brasil.
- Sugerir o melhor mês para a campanha Setembro Amarelo.

2 REFERENCIAL TEÓRICO

2.1 SERIES TEMPORAIS

Segundo Gujarati (2008, p45), uma série temporal é um conjunto de observações de valores que uma variável assume em diferentes momentos do tempo. Esses dados podem ser coletados a intervalos regulares, como diariamente (preços das ações, relatórios meteorológicos), semanalmente (informações sobre oferta de moeda), mensalmente (taxa de desemprego, índice de preços ao consumidor [IPC]), trimestralmente (PIB), anualmente (orçamento do governo), quinquenalmente, isto é, a cada cinco anos (censo industrial dos Estados Unidos), ou decenalmente (censo demográfico).

As séries temporais são utilizadas com diversos propósitos, Morettin e Toloi (2006, p24) destacam os seguintes objetivos:

a. Investigar o mecanismo gerador da série temporal;
b. Fazer previsões de valores futuros da série;
c. Descrever apenas o comportamento da série através de gráficos e verificando a existência de tendência, ciclos e variações sazonais;
d. Procurar periodicidades relevantes nos dados.

Em relação as suas características, uma série pode ser definida como a soma de três componentes não-observáveis: Tendência, sazonalidade e erro. Para uma série {Zt, t=1,...,N} pode-se apresenta-la como:(Moretti e Toloi,2006,p.46)

$$Zt = Tt + St + at \quad (1)$$

Onde,

T = Tendência

S = Sazonalidade

a = Erro

Para HYNDMAN, esses componentes podem ser definidos da seguinte forma (2021):

- Tendência: “Existe uma tendência quando há um aumento ou diminuição de longo prazo nos dados”
- Sazonalidade: “Ocorre quando uma série temporal é afetada por fatores sazonais, como a época do ano ou o dia da semana. A sazonalidade é sempre de frequência fixa e conhecida.”
- Erro: Componente aleatório que resta ao retirar a Sazonalidade e a tendência.

Segundo Bueno (2008, p7) uma série temporal pode ser classificada como estacionária ou não estacionária. A estacionariedade implica um comportamento aleatório ao longo do tempo sob as seguintes condições:

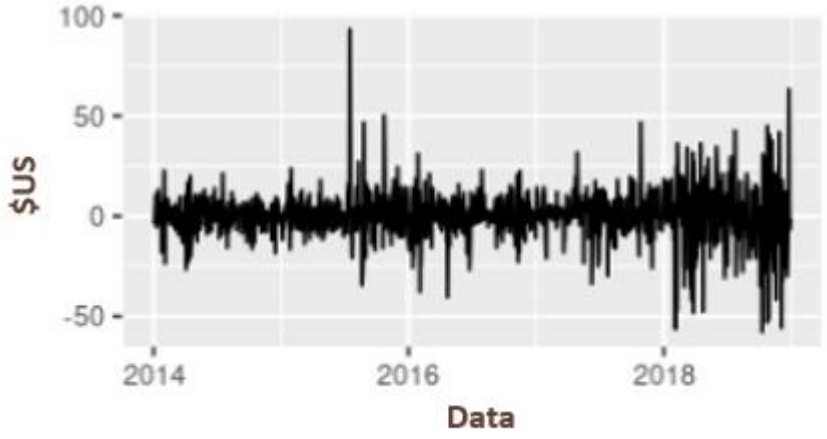
$$Média: E(Y_t) = \mu \quad (2)$$

$$Variância: var(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \quad (3)$$

$$Covariância: (Y_k) = E[(Y_t - \mu)(Y_{t+k} - \mu)] \quad (4)$$

A figura 1 apresenta a variação diária no preço de fechamento das ações do google, a mesma é um exemplo de uma série estacionária pois todas as condições apresentadas anteriormente de média constante e variância igual a zero foram satisfeitas.

Figura 1 – Preços diários das ações do Google entre 2014 a 2018



Fonte: HYNDMAN (2021)

Morettin e Toloi (2006, p.5) destaca que

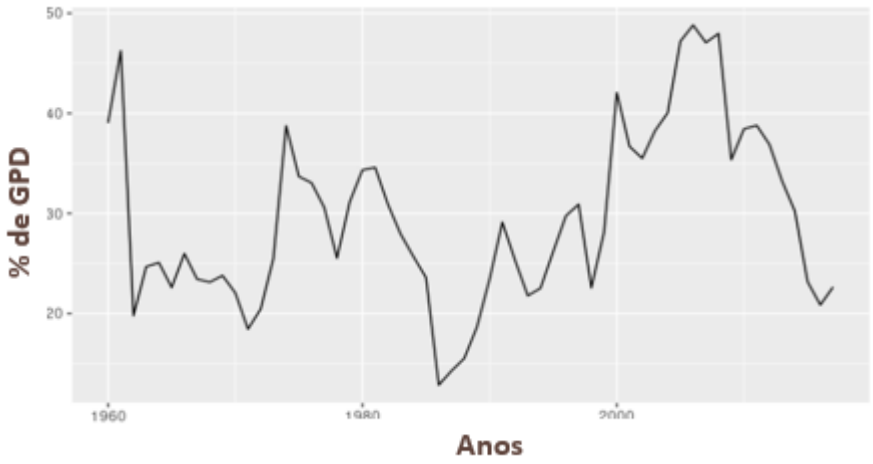
> Uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio.

Para muitos modelos de séries temporais é necessário que o mesmo seja estacionário ou que os dados sejam transformados utilizando processos de diferenciação, um exemplo de modelo muito utilizado sob as condições informadas acima são os da classe ARIMA.

Uma série temporal pode ser classificada como não estacionária quando suas propriedades estatísticas dependem da janela de tempo em que a mesma é observada. (HYNDMAN, 2021)

A figura 2 apresenta dados sobre as exportações de bens e serviços da Argélia durante o período de 1960 e 2017, a mesma é uma série não estacionária pois sua média não é constante ao longo do tempo.

Figura 1 - Exportações de Bens e Serviços da Argélia



Fonte: HYNDMAN (2008)

2.2 MODELOS DE SERIES TEMPORAIS

Os modelos de séries temporais têm como objetivo diferenciar padrões de comportamento de qualquer ruído que possa estar presente nas observações passadas, com isso, utilizar os padrões encontrados nos dados históricos para prever valores futuros. (MORETTIN e TOLOI, 2006)

Segundo Pellegrini e Fogliatto (2001) as técnicas de previsão variam consideravelmente, onde cada uma delas foi desenvolvido com um propósito, grau de precisão e custo de utilização distintos. A escolha do melhor modelo dependerá de como as previsões futuras podem ser explicadas através de seus componentes (sazonalidade, tendência, entre outros), da agregação temporal dos dados e o intervalo de previsões escolhida.

Para Morettin (2006) os modelos de séries temporais podem ser divididos em dois grandes grupos:

- Modelos paramétricos: Nestes modelos o número de parâmetros é finito e a análise é feita no domínio do tempo. Como exemplo de métodos podemos citar: regressão linear, ARMA, ARIMA, modelos não-lineares;
- Modelos não-paramétricos: Envolvem um número infinito de parâmetros, muito utilizado para aprendizado de máquina. Alguns dos principais modelos são a função de auto covariância e sua transformada de Fourier.

2.2.1 Médias Móveis

O método de médias móveis é um modelo que utiliza da média aritmética das observações anteriores da série temporal com objetivo de prever seu próximo valor. (GONÇALVES, 2018)

Recebe o nome de média móvel, porque as observações mais antigas vão saindo da média aritmética à medida que novos valores são adicionados a série. Para uma série Z(t) com r observações anteriores o valor Z(t+1) é definido pela seguinte equação (MORETTIN e TOLOI 2006):

$$Z_{t+1} = \frac{Z_t + Z_{t-1} + Z_{t-2} + \cdots + Z_{t-r+1}}{r} \quad (5)$$

A vantagem de se utilizar este método é que por ser uma técnica extremamente simples é necessário um baixo esforço computacional e necessitam de poucos dados históricos. (MAKRIDAKIS, WHEELWRIGHT, & HYNDMAN, 1998).

Para Lemos, seu uso é recomendado para previsões de curto prazo e tem como desvantagem não lidar bem com séries que possuam a presença de tendência e sazonalidade. (2006).

2.2.2 Suavização Exponencial Simples

É considerado o método mais simples entre as suavizações exponenciais, este modelo calcula os valores futuros da série utilizando uma média ponderada onde os pesos diminuem exponencialmente à medida que as observações ficam mais antigas. (HYNDMAN, 2008)

Ainda segundo Hyndman, este método é muito utilizado para dados sem tendência clara ou sazonalidade. Para calcular o valor futuro de uma série, o pode ser definido pela equação 6: (2008)

$$\hat{y}_{\frac{t+1}{T}} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ldots, \quad (6)$$

Onde,

- α: constante de suavização (0 ≤ α≤ 1);
- $\hat{y}_{t+1}$: previsão para o período (t+1);
- y_t: valor da série relativa ao período t;

A principal vantagem da suavização exponencial simples é sua grande flexibilidade por permitir que a constante de suavização seja modificada sempre que o padrão dos dados da série mudar. Além disso, possui uma boa aderência para processos preditivos onde se tem grandes quantidades de dados. (LEMOS, 2006).

Como principal desvantagem, destaca-se a dificuldade em encontrar um valor ótimo para a constante de suavização. (GONÇALVEZ, 2008; MORETTIN; TOLOI,2006)

2.2.3 Suavização Exponencial Linear de Holt (SEH)

Este método permite a previsão de dados com tendência, mas sem sazonalidade. Além da constante de suavização exponencial que é utilizada no método anterior foi feito um acréscimo de uma outra constante, chamada de β, com objetivo de modelar a tendência da série temporal. (KOENING, 2014).

Para sua utilização, além da equação de previsão são necessárias duas equações de suavização, uma para o nível e outra para a tendência que tem o objetivo respectivamente de calcular onde a série está em um determinado período e o quanto a mesma vai crescer.

Holt (1957) propões as seguintes Equações para determinar respectivamente os valores da previsão, e dos coeficientes de nível e tendência da série.

$$Equação\ de\ previsão: \hat{y}_{\frac{t+h}{t}} = l_t + hb_t \quad (7)$$

$$Equação\ de\ nível: l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (8)$$

$$Equação\ de\ tendência: b_t = \beta * (l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (9)$$

Onde,

- α: constante de suavização (0 ≤ α≤ 1);
- y_t: valor da série relativa ao período t;
- l_t: nível da série no tempo t
- b_t: tendência da série no período t
- β: constante de amortização de tendência (0 ≤ β ≤ 1)

Para Morettin e Toloi (2006), o método compartilha da mesma dificuldade encontrada no modelo de suavização exponencial simples, determinar os valores ótimos para as constantes α (coeficiente de amortização de nível) e β (constante de amortização de tendência)

2.2.4 Suavização exponencial de Holt-Winters

Com objetivo de capturar não só a tendência, mas também a sazonalidade, foi desenvolvido um método que utiliza três equações de suavização (uma para o nível l_t, uma para a tendência b_t, e um para o componente sazonal s_t. (HYNDMAN, 2008).

Segundo Gonçalves, o método de suavização exponencial de Holt-Winters se divide em dois modelos que dependerá do padrão sazonal da série estudada (2008):

1) Método multiplicativo: É utilizado quando a variação sazonal aumenta ou diminui à medida que o período analisado se altera. Para Winters, as equações que definem esse método são:

$$\hat{y}_{\frac{t+h}{t}} = (l_t + hb_t)_{S_{t+h-m(k+1)}} \quad (10)$$

$$l_t = \alpha \frac{y_t}{S_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \quad (11)$$

$$S_t = \gamma \frac{y_t}{(l_{t-1} + b_{t-1})} + (1-\gamma)S_{t-m} \quad (12)$$

Onde:

- α,β e y são constantes de suavização
- S é a duração do período sazonal
- $\hat{y}_{(t+h)/t}$ é o valor previsto para o período t+h

2) Método aditivo: É utilizado quando a variação sazonal continua padrão ao longo do tempo na série estudada. É definido pelas seguintes equações:

$$\hat{y}_{\frac{t+h}{t}} = (l_t + hb_t)_{S_{t+h-m(k+1)}} \quad (13)$$

$$l_t = \alpha(y_t - S_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \quad (14)$$

$$b_t = \beta * (l_t - l_{t-1}) + (1-\beta)b_{t-1} \quad (15)$$

$$S_t = \gamma(y_t - l_{t-1} + b_{t-1}) + (1-\gamma)S_{t-m} \quad (16)$$

Onde:

- α,β,y e são constantes de suavização
- S é a duração do período sazonal
- $\hat{y}_{t+h/t}$ é o valor previsto para o período t+h

2.2.5 Modelos ARIMA

Os modelos da família ARIMA foram desenvolvidos por Box e Jenkins e tem o intuito de utilizar modelos autorregressivos e de médias moveis para ajustar e prever séries temporais que podem ser estacionárias (ARIMA) ou com presença de sazonalidade (SARIMA). (Box e Jenkins, 1994).

Segundo Bueno, a metodologia desenvolvida por Box e Jenkins para modelar uma série temporal possuem os seguintes passos: (2012)

1. Confirmar se a série estudada é considerada estacionária, caso não seja, deve-se diferencia-la antes de passar para a próxima etapa.
2. Identificar a classe geral do Modelo (AR, MA, ARMA, ARIMA ou SARIMA).
3. Estimar o modelo, para isso deverá ser analisado as funções de autocorrelação (FAC) e autocorrelação parcial (FACP). As duas funções ajudaram a encontrar as ordens p e q do modelo.
4. Diagnosticar se os resíduos gerados satisfazem a hipótese de ser um ruído branco. Se a hipótese não for rejeitada, pode-se seguir para a próxima etapa, para isso as funções de autocorrelação e autocorrelação parcial dos resíduos

estimados não devem conter memória. Caso a hipótese seja rejeitada, será necessário voltar para a primeira etapa.

5. Prever os valores futuros da série estudada e definir seus níveis de confiança. Para encontrar o desempenho das previsões feitas, pode-se utilizar as métricas do erro quadrático médio (MSPE) e do erro absoluto médio (MAPE).

2.2.5.1 Modelos autorregressivos (AR) e de média móveis (MA)

Em um modelo autorregressivo as previsões são feitas a partir de uma combinação linear dos valores passados da série estudada. Assim, para um modelo AR de ordem p pode ser escrito como (HYNDMAN,2008):

$$Y_y = c + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + \varepsilon_t \quad (17)$$

Onde,

- c : constante da equação;
- θ_q : paramento autorregressivo;
- ε_t : ruído branco.

Ainda segundo Hyndman para um modelo de médias móveis (MA) as previsões são feitas a partir de uma média móvel ponderada dos últimos erros da previsão em relação aos valores reais da série estudada. Um modelo pode ser descrito pela equação 18: (2008)

$$Y_t = c + \varepsilon_t + \phi_t \varepsilon_{t-1} + \ldots + \phi_q \varepsilon_{t-q} \quad (18)$$

Onde,

- c : constante da equação;
- $\emptyset_t$: paramento de médias móveis;
- ε_t : ruído branco.

2.2.5.2 Modelos ARIMA e SARIMA

O modelo ARIMA é a combinação dos modelos de média móvel, diferenciação e autorregressão. Para a modelagem de uma série temporal pode-se utilizar tanto os três componentes quanto um subconjunto destes. Seu modelo pode ser descrito pela equação 19: (HYNDMAN; 2008, FAVA; 2000).

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (19)$$

Onde,

- p : Ordem da parte autorregressiva;
- d : grau de primeira diferenciação envolvido;
- q = Ordem da parte da média móvel;
- y'_t : Série temporal diferenciada.

O modelo SARIMA é a adição de termos sazonais no modelo ARIMA que foi apresentado anteriormente. Para Hyndman os componentes sazonais do modelo são semelhantes a não sazonal, mas envolvem retrocessos do período sazonal e pode ser definido pela equação 20, onde os termos sazonais são multiplicados pelos termos não sazonais:

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4) y_t = (1 + \theta_1 B)(1 + \Theta_1 B^4)\varepsilon_t. \quad (20)$$

2.2.5.3 Cenários Temporais

Os modelos da família ARIMA possuem um importante papel para analisar os impactos de políticas públicas no Brasil, destaca-se a pesquisa feita por Nepomuceno et al. (2017) que avaliou a eficácia da lei estadual brasileira 13.748 de abril de 2019, que proibia a venda de bebidas alcoólicas nos estádios de futebol de Pernambuco. Tal pesquisa concluiu que a venda de bebidas exerce pouca influência sobre o

comportamento violento dos torcedores a partir da análise de 1363 ocorrências em estádios durante os períodos de fevereiro de 2009 a dezembro de 2015.

Os resultados deste trabalho foram utilizados como suporte para extinguir em janeiro de 2016 esta lei, sendo este um exemplo prático de como o estudo de séries temporais podem ser aplicados no auxílio de políticas públicas.

2.3 APLICAÇÕES DE SÉRIES TEMPORAIS EM PESQUISAS SOBRE SUICÍDIOS

O estudo de séries temporais no contexto de doenças e saúde pública é bem difundido, principalmente para análise de doenças crônicas e previsão da sua mortalidade. Pode-se destacar por exemplo, o estudo de (RAZZAQUE; BISHAI, 2018) que buscou avaliar os efeitos dos extremos de temperatura na mortalidade infantil em Bangladesh utilizando séries temporais, ou ainda, o estudo feito por (OLUKANMI; NELWAMONDO; NWULU, 2021) que utilizou o modelo ARIMA para modelar dados de pesquisa no google com objetivo de prever doenças semelhante a gripe na África do sul.

Quando o tema é suicídios, existem diversas aplicações para a utilização de séries temporais, vale ressaltar a pesquisa feita por (YOONHEE K. HO K.; DONG-SIK K., 2010) que identificou uma associação entre a temperatura ambiental diária e a mortalidade por suicídios na Coréia do Sul utilizando dados de 2001 a 2005 coletados do escritório nacional de estatística da Coréia (KNSO) e dados de 64 estações meteorológicas do país, para isso foi utilizado o método GAM para modelar a série temporal. O estudo identificou uma correlação positiva onde a cada 1 ºC de aumento médio diário na temperatura ocorreu um aumento de 1.4% no número de suicídios.

Outro uso bem difundido são os impactos de crises econômicas no número de óbitos por suicídio, destaca-se a pesquisa desenvolvida por (BRANAS et al.,2015) que avaliou utilizando o método ARIMA o impacto dos eventos de austeridade econômica e prosperidade no número de suicídios na Grécia nos períodos de 1983 a 2012. O estudo identificou um aumento de 35% no número de óbitos por suicídio a partir de junho de 2011, assim que novas medidas mais rigorosas foram sancionadas para conter a crise econômica que assolava o país.

Modelos de séries temporais podem ser utilizados não só para previsões da taxa de mortalidade autoprovocada, mas para encontrar padrões nos dados históricos.

O trabalhou feito por (ROSTAMI et al., 2019) identificou a partir de um modelo de suavização exponencial um padrão na sazonalidade no número de suicídios no Oeste do Irã, onde foi identificado picos de mortalidade nos meses de agosto entre os períodos de 2006 a 2013.

2.3.1 Aplicações de séries temporais em pesquisas de suicídio no Brasil.

No Brasil encontram-se pesquisas sobre o tema tanto para descrever aspectos sociodemográficos como para avaliar tendências de aumento ou diminuição dos números de mortalidade.

A pesquisa feita por (ARRUDA et al., 2021) analisou a partir de séries temporais (método de regressão de Prais-Winsten) a mortalidade por suicídio em adultos jovens brasileiros entre os anos de 1997 e 2019. O trabalho identificou uma tendência de crescimento tanto para o sexo masculino (VPA: 1; IC95%) quanto para o sexo feminino (VPA: 1,45, IC95%).

Vale destacar também, o trabalho feito por (ROSA et al., 2017) que utilizou regressões polinomiais para avaliar os impactos das políticas públicas de saúde mental no estado do Pará que veio reestruturando desde o ano de 2001 as redes de atendimento à saúde mental como os CAPS e APS. Para realização do estudo foram utilizados dados de mortalidade entre 1996 a 2012 e identificaram uma diminuição de 0,16 ao ano na taxa de suicídio do estado.

Já (OLIVEIRA et al., 2020) questionou os impactos no número de suicídios após o lançamento da campanha Setembro Amarelo (CSA) feita pelo ministério da saúde. O trabalho identificou um aumento na taxa de suicídios de 66% entre os períodos de 2013 a 2017, onde o programa CSA foi iniciado no ano de 2015, a partir de um estudo descritivo de séries temporais.

Utilizando o método de suavização exponencial, destaca-se o trabalho de (SILVA A.V et al.,2021) que fez uma análise epidemiológica acerca do suicídio na região metropolitana de Maceió, ao longo dos últimos 20 anos (2000 até 2019). Como resultados, chegou-se que do total de suicídios, 77% eram do gênero masculino e que os dados coletados do sistema de informação de mortalidade (SIM) apresentavam

uma tendência de crescimento médio da mortalidade de 2,2 por 100.000 habitantes ao ano.

Uma pesquisa descritiva de séries temporais desenvolvida por (GALVÃO et al. 2017) em Pernambuco (1997 a 2010) identificou uma taxa de 71,5 suicídios a cada 100 mil habitantes, onde desses, houve um predomínio de homens, adultos (25 a 39 anos), pardos, solteiros e com baixa escolaridade.

2.3.2 Aplicações de séries temporais em pesquisas analisando o impacto do COVID no número de suicídios.

Algumas pesquisas tentam avaliar o impacto do COVID-19 no número de suicídios, vale destacar o trabalho desenvolvido pela Secretaria de planejamento (2020) do estado da Bahia que constatou uma diminuição da taxa de mortalidade por suicídio no Brasil teve uma redução de 5.7% em comparação ao ano de 2019. Foi utilizado uma abordagem descritiva de séries temporais comparando as médias e variações ano a ano, utilizando como insumo dados da superintendência e proteção e Vigilância nos períodos de 2010 a 2020.

Pesquisa com objetivo semelhante foi feito por (PIRKIS et al., 2021) que avaliou o impacto da pandemia no número de mortes autoprovocadas em 21 países entre 1º de setembro de 2020 e 1º de novembro de 2020 a partir de dados coletados em diversas fontes como ministério da saúde, agências policias entre outras. Como resultado, para os países de renda alta e média, os números de suicídio permaneceram praticamente inalterados ou diminuíram, no caso do Brasil, durante a fase inicial de contágio não houve nenhuma evidência de aumento ou diminuição por conta da pandemia.

Diversos países tentaram encontrar alguma correlação entre o número de suicídios e a pandemia de COVID-19, um trabalho desenvolvido em Queensland (Austrália) (LESKE et al., 2021) avaliou os impactos da COVID em mortes autoprovocadas através de um modelo de serie temporal interrompida em um comparativo entre o período anterior a pandemia (2015 – 2019) e durante a mesma (2020), chegou-se à conclusão de que não existem evidências de aumento nas mortes durante o período estudado.

Um estudo desenvolvido na Coreia (KIM, 2021) a partir de um comparativo dos primeiros 8 meses de pandemia (2020) em relação ao ano anterior foi constatado uma diminuição na taxa de suicídios de 6,9%.

No Canadá, um estudo desenvolvido em várias províncias do país (ISNAR; OREMUS, 2022) identificou que as diferenças entre as contagens de mortes por automutilação observadas e previstas não foram estatisticamente significativa, foi desenvolvido um modelo ARIMA utilizando dados de 218 a 2020.

3 METODOLOGIA

Em relação ao desenho do estudo, este trabalho se trata de uma pesquisa básica estratégica, com objetivo descritivo, por meio de uma abordagem quantitativa a respeito do tema de séries temporais e suicídios.

Em relação aos dados do número de suicídios no Brasil, foram coletados através do sistema de informação sobre mortalidade (SIM), por meio eletrônico, utilizando o portal do SUS (DATASUS) que possuem as informações oficiais do Brasil sobre o tema estudado. Considerou-se como óbito por suicídio causas externas de mortalidade que utilizam os códigos X60 a X84, classificação adotada internacionalmente segundo o CID-10.

Para este estudo foram utilizadas as seguintes bases de dados e para os seguintes objetivos:

- Suicídios por tempo, ano e mês, por estado: Foi utilizado como fonte de dados para criação dos modelos de séries temporais.
- Suicídios por cor e Faixa Etária, suicídios por causa, suicídios por estado civil, Suicídios por faixa etária, suicídios por local, suicídios por sexo: Foram utilizados para entender o perfil desses indivíduos através de uma análise exploratória.
- População estimada do Brasil por ano, entre 1996 a 2020, e por Estado: Foi utilizada para encontrar a taxa de suicídios para algumas análises exploratórias.
- População estimada do Brasil por mês, entre 2010 a 2020: Foi utilizada para encontrar a taxa de suicídios mensal.

Para as bases de dados informadas acima foram utilizadas as seguintes variáveis: Sexo (Masculino, Feminino, Ignorado); Cor (Branca, Pretos e pardos, Ignorado, Indígena e Amarela); Faixa Etária (10 a 14, 15 a 19, 20 a 29, 30 a 39, 40 a 49, 50 a 59, 60 a 69, 70 a 79 e 80); Causa (X60 – X69: autointoxicação e X70-x84:Lesões autoprovocada intencionalmente); Estado Civil (Solteiro, Casado, Ignorado, Separado judicialmente, Viúvo e Outro); Local (Domicílio, Hospital, Via Pública, Outro estabelecimento de saúde, ignorado e Outros).

Em relação ao tratamento dos dados foi criado uma rotina no software R para tratar as bases utilizadas, como:

- Colocar todos os estados em uma única coluna caso fosse necessário;
- Retirar textos que continham descrição dos dados coletados;
- Retirar linhas de subtotais que apareciam ao longo das bases de dados;
- Formatar apropriadamente coluna de mês e ano para o formato utilizado no Brasil e que o software R conseguisse identificar.

Para o cálculo da taxa de mortalidade, foram utilizados dados do Instituto Brasileiro de Geografia e Estatísticas (IBGE) sobre projeções populacionais. (DATASUS). Foi utilizado o método descrito na equação 21:

$$\frac{N°\ de\ óbitos\ pela\ causa\ específica, em\ determinado\ local\ e\ periodo}{População\ total\ do\ mesmo\ local\ e\ período\ x100.000} \quad (21)$$

Os dados de suicídios foram analisados utilizando a estatística descritiva e o método de séries temporais proposto por Box & Jenkins para testes de significâncias, análises gráficas, modelagem e previsões

Como modelos de previsão de séries temporais foi aplicado o método de suavização exponencial, onde utilizando a função ETS do pacote fpp3, que utiliza o critério AIC para selecionar o melhor modelo, chegou em um modelo de tendência aditiva, erro multiplicativo e sazonalidade aditiva com as seguintes equações:

$$y_t = (l_{t-1} + \emptyset b_{t-1} + s_{t-m})(1 + e_t) \quad (22)$$

$$l_t = (l_{t-1} + \emptyset b_{t-1} + a)(l_{t-1} + \emptyset b_{t-1} + s_{t-m})(e_t) \quad (23)$$

$$b_t = \emptyset b_{t-1} + \beta(l_{t-1} + \emptyset b_{t-1} + s_{t-m})e_t \quad (24)$$

$$s_t = s_{t-m} + \gamma(l_{t-1} + \emptyset b_{t-1} + s_{t-m})e_t \quad (25)$$

Também foi analisado os ajustes para um modelo ARIMA. Para essa fase do trabalho foi avaliado alguns modelos e chegou-se nas métricas explanadas na tabela 1:

Tabela 1 - Modelos analisados e seus respectivos AIC's

MODELO	AIC
ARIMA (1,1,1)	1.251
ARIMA (4,1,1)	1251
ARIMA (11,1,1)	1238
ARIMA (4,1,1) (1,0,1) [12]	1197
ARIMA (4,1,1) (1,0,2) [6]	1215
ARIMA (0,1,3) (0,0,2) [12]	1222

Fonte: O Autor (2022)

Para todos os modelos analisados, foram repartidos os dados sobre suicídios em dados de treino (1996 a 2017), dados de teste (2018 a 2019).

Foi utilizado a métrica RMSE para analisar a acuracidade entre o melhor modelo ARIMA e o modelo de suavização exponencial com objetivo de definir o modelo que será empregado para analisar os dados de suicídio de 2020.

Com objetivo de avaliar a eficácia da campanha Setembro Amarelo foi utilizado o modelo selecionado pela métrica RMSE e reajustado com dados de treino de 1996 a 2014. Logo em seguida, foi comparado com os dados de 2015 a 2020 as previsões geradas pelo modelo no mesmo período.

Para tabulação foi utilizado o programa Microsoft Office Excel 2019 e para análise descritiva, construção de gráfico, análise de tendência e criação dos modelos foi utilizado o software Rstudio versão 4.0.3.

Foram utilizados os seguintes pacotes no Rstudio para análise e tratamento dos dados: tidyr, dplyr, ggplot2, readxl, stringr, forecast, fpp3, TSA.

4 DESENVOLVIMENTO

Esta seção está reservada a análise descritiva dos dados relativos aos suicídios no Brasil, considerando a série história de 1996 a 2020. Ademais o Subtópico 4.1 apresentará uma análise descritiva da taxa de suicídios no Brasil durante os períodos de 2010 a 2020. No subtópico 4.2 será apresentado uma análise descritiva por categorias como sexo, raça, faixa etária, causa de suicídios e estado. No subtópico 4.3 será identificado quais os componentes de tendência e sazonalidade a partir da decomposição da série temporal de óbitos por suicídio. No subtópico 4.4 será definido o melhor modelo de previsão entre os métodos de suavização exponencial e ARIMA utilizando como critério de escolha as métricas AIC e RMSE. Nos subtópicos 4.5 e 4.6 será utilizado o melhor modelo definido no tópico anterior para avaliar os impactos da Campanha de prevenção aos suicídios (setembro Amarelo) e da pandemia da COVID-19 nos números de óbitos.

4.1 ANÁLISE DESCRITIVA GERAL

A taxa de óbitos entre os anos de 2010 e 2020 apresentam uma média de 23 suicídios por ano a cada 100 mil Hab., a diferença entre a média e a mediana mostram que os dados possuem uma distribuição uniforme sem muitos outliers tendo seu maior valor, 26 suicídios no ano de 2020 e seu menor valor, 20 suicídios no ano de 2011.

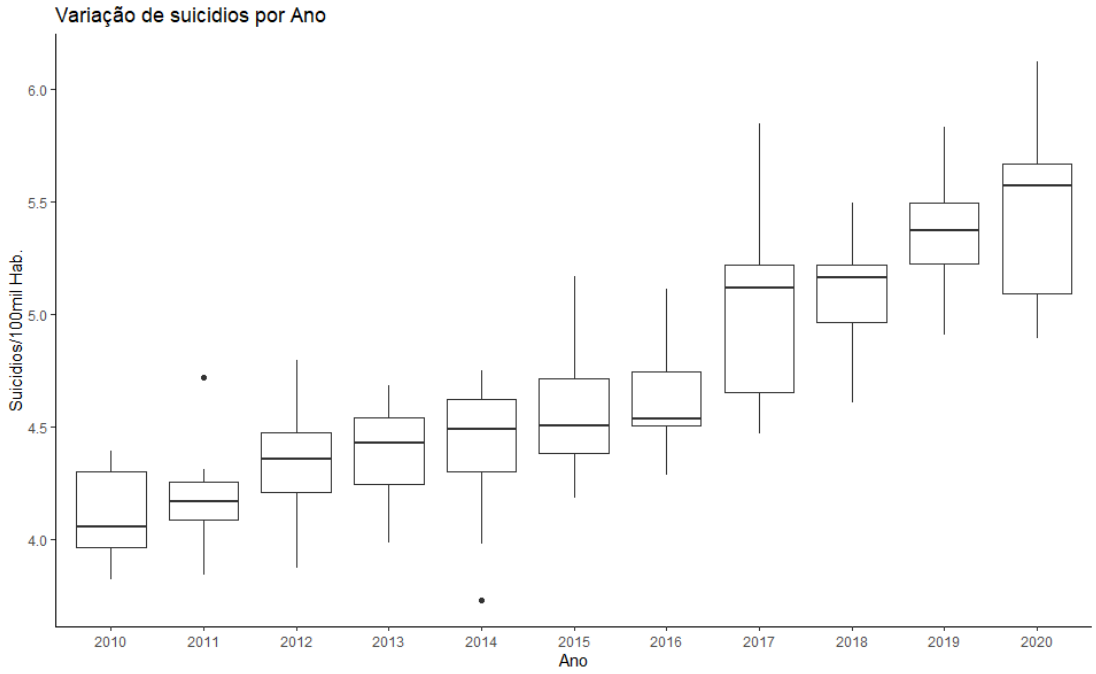
Tabela 2. Estatística descritiva

Min	1st Qu.	Mediana	Média	3rd Qu.	Max.
20	21	22	23	24	26

Fonte: O Autor (2022)

A figura 3 apresenta a variação anual do número de suicídios entre 2010 a 2020, embora tenha-se encontrado apenas 2 outliers a diferença não é significativa devido a escala ser baixa. O gráfico expõe uma tendência crescente na taxa de suicídios e uma distribuição não uniforme ao longo dos anos, na maior parte dos casos abaixo da mediana, podendo ser um indicio de padrões sazonais.

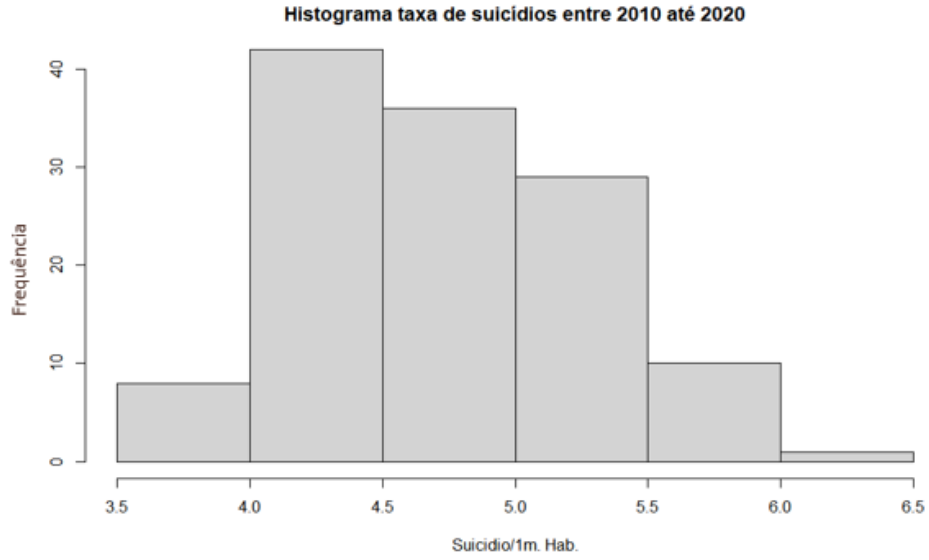
Figura 2 - Variação de Suicídios anualmente



Fonte: O autor (2022)

O histograma na figura 4 apresenta a distribuição do número de suicídios mensais para cada 1 milhão de Hab., o mesmo expõe que no Brasil morre-se entre 4 a 6 pessoas por mês.

Figura 3 - Histograma do número de Suicídios no Brasil

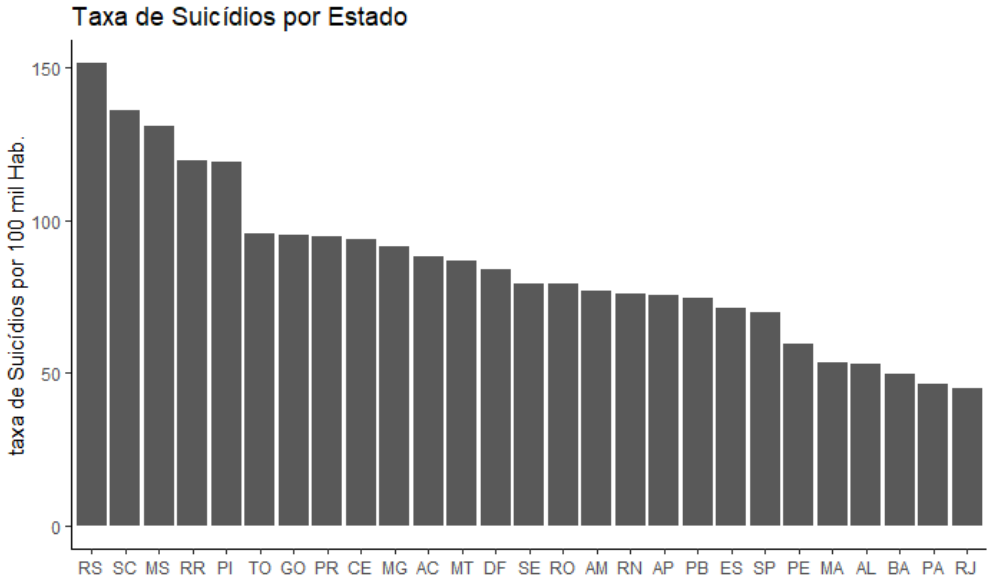


Fonte: O autor (2022)

4.2 ANÁLISE DESCRITIVA POR CATEGORIAS

A figura 5 mostra a taxa de suicídios entre os períodos de 2010 até 2020 para os estados brasileiros.

Figura 4 - Suicídios por Estado

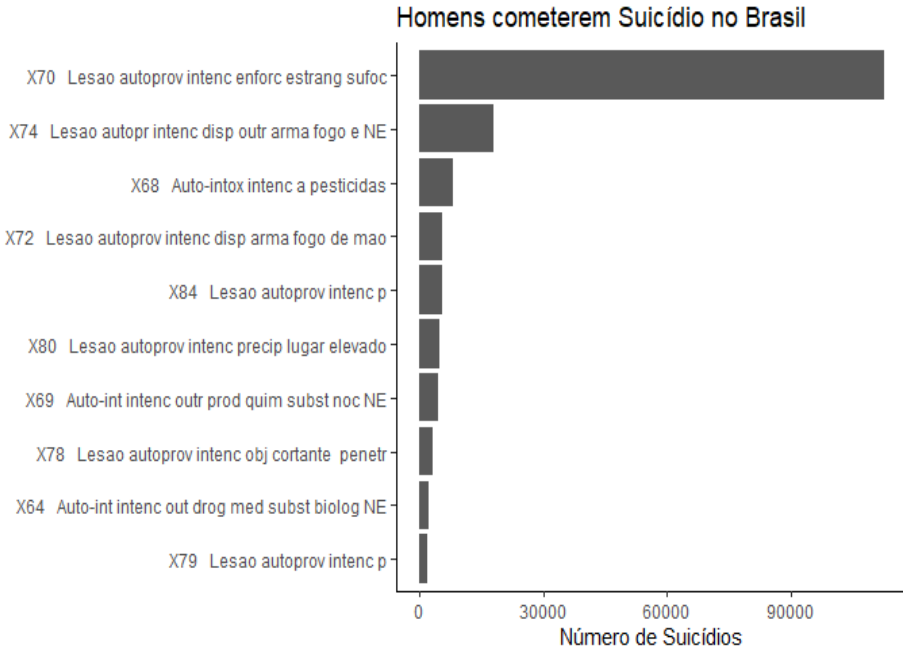


Fonte: O autor (2022)

O estado que possui a maior taxa de suicídios proporcional a sua população é Rio Grande do Sul. Para um melhor entendimento é importante investigar quais os possíveis motivos.

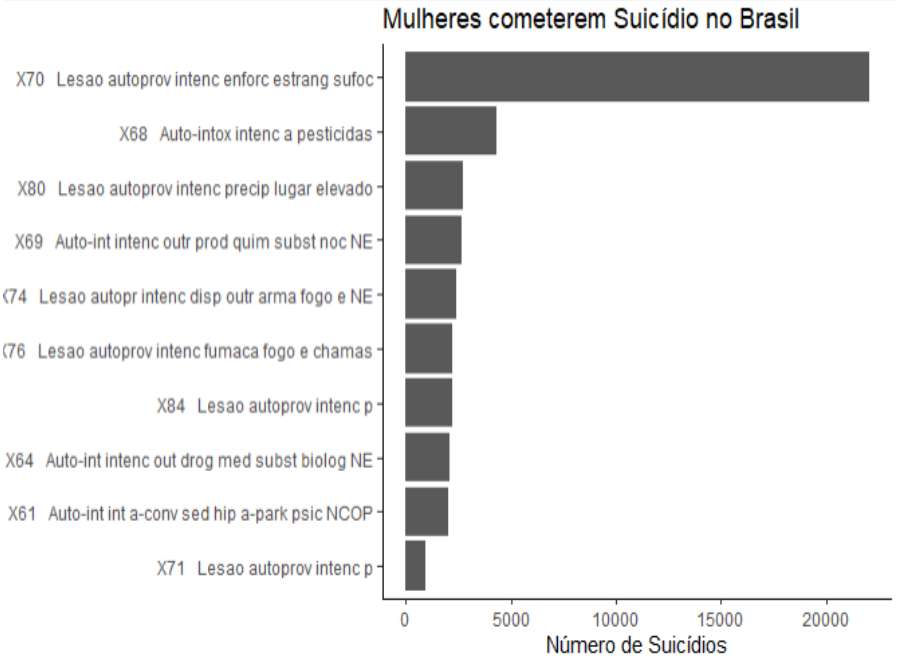
Analisando agora as principais formas de suicídios utilizadas, é possível notar que essas escolhas podem depender do perfil do indivíduo como idade, sexo, localização, escolaridade e etc. Para uma visão apenas de sexo, temos as seguintes escolhas:

Figura 5 - Número de Suicídios por Causa - Sexo Masculino



Fonte: O autor (2022)

Figura 6 - Número de Suicídios por Causa - Sexo Feminino



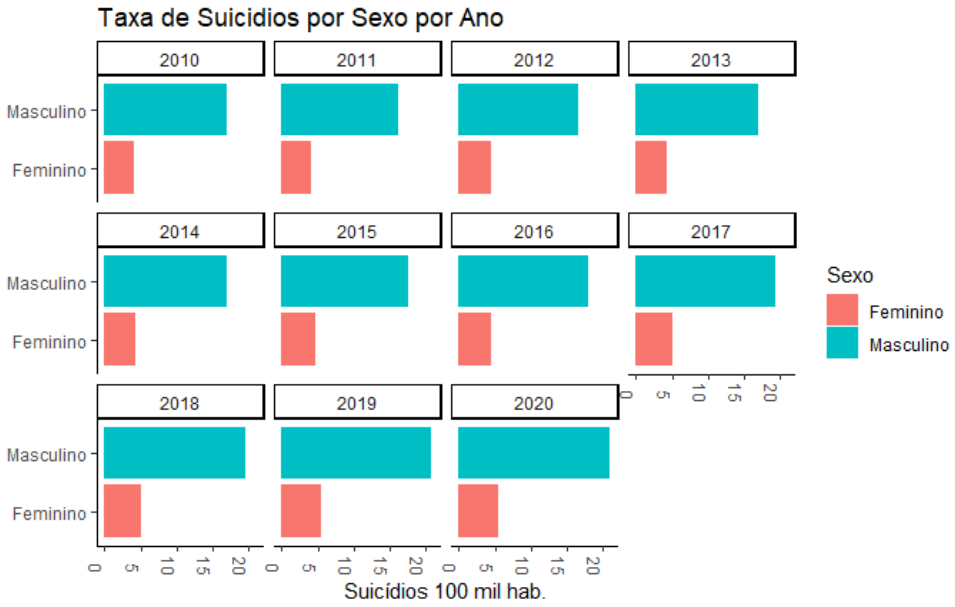
Fonte: O autor (2022)

A principal forma utilizada tanto pelo sexo masculino quanto pelo sexo feminino é o enforcamento, porém após isso o ranking possui variações, por exemplo,

para homens a arma de fogo ocupa o segundo lugar enquanto para as mulheres essa colocação foi para o uso de venenos. Vale ressaltar que esses dados só levam em consideração os óbitos, para um entendimento mais amplo seria interessante analisar informações das tentativas de suicídio que não se concretizaram.

Em relação ao sexo, pessoas do sexo masculino se suicidam mais que pessoas do sexo feminino, para todos os anos analisados. Porém segundo o SINAN, entre 2011 a 2018, 67% das notificações de tentativas de suicídios eram indivíduos do sexo feminino (2019, p.4). Essa taxa de sucesso pode estar ligada aos métodos utilizados, como apresentado anteriormente, pessoas do sexo masculino optam por formas mais violentas e letais como a arma de fogo. A figura 8 apresenta a taxa de suicídios por sexo nos anos de 2010 a 2020, nota-se que em todos os anos pessoas do sexo masculino tiveram um número maior de mortes autoprovocadas.

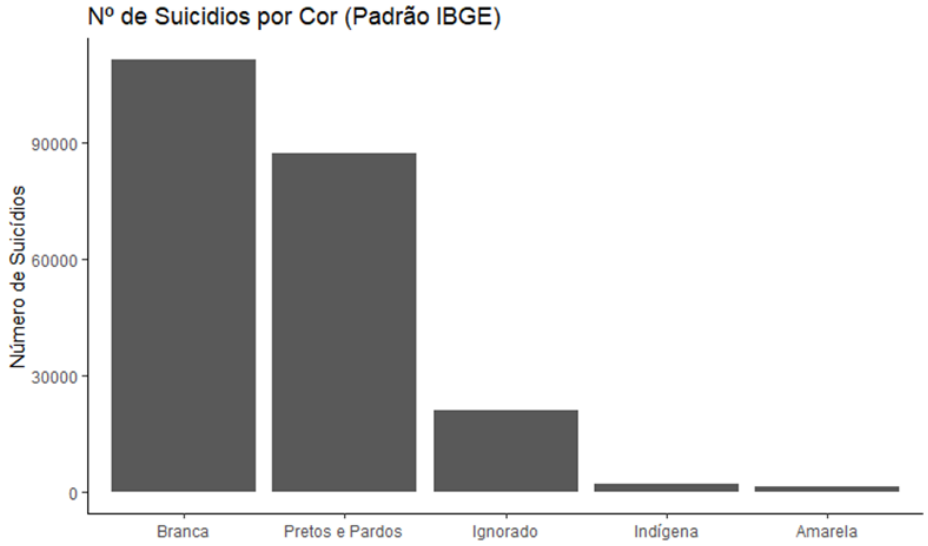
Figura 7 - Suicídios por Sexo e Ano



Fonte: O autor (2022)

Para entender o perfil epistemológico é importante também perceber como esse problema se comporta por raça, a figura 9 mostra que durante o período de 1996 até 2020 pessoas brancas possuem o maior número de óbitos.

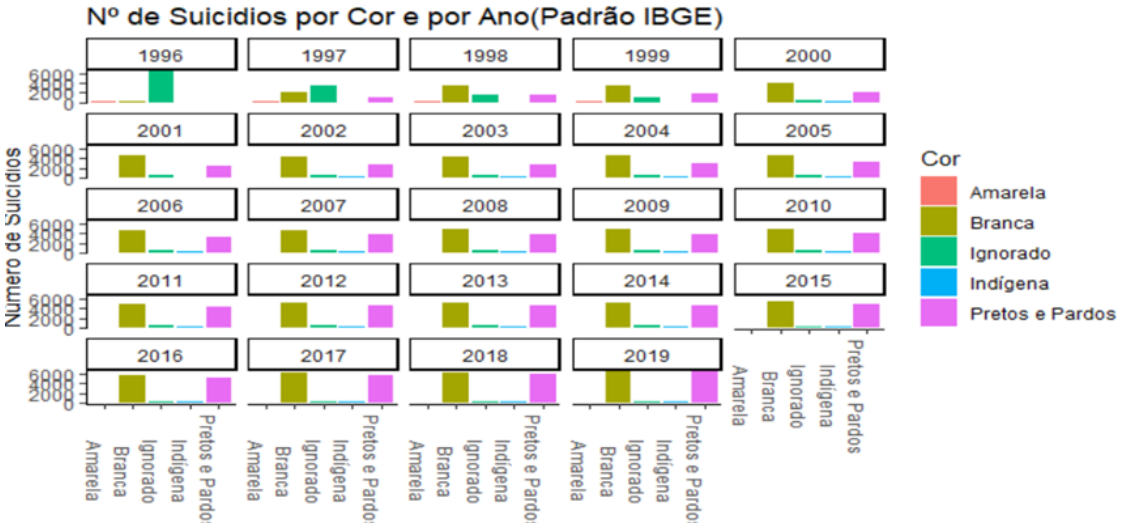
Figura 8 - Suicídios por Cor



Fonte: O autor (2022)

A figura 10 mostra que para todos os anos observados indivíduos brancos se suicidam mais.

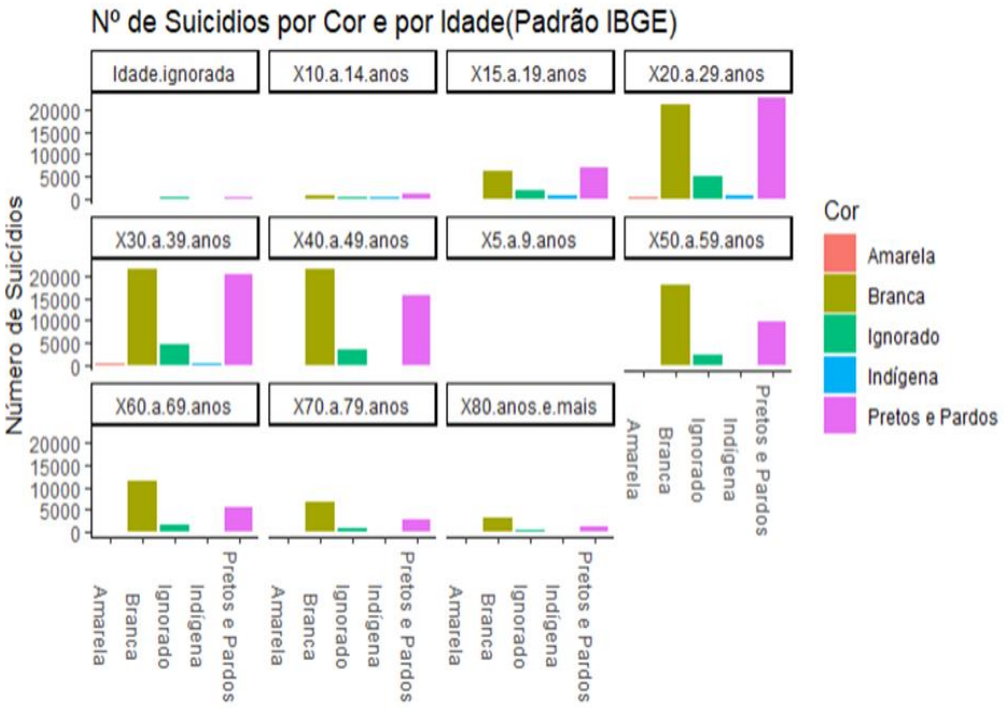
Figura 9 - Suicídios por Cor e Ano



Fonte: O autor (2022)

Avaliando também o número de suicídios por faixa etária, pessoas negras e pardas se suicidam mais que as outras raças quando a idade das vítimas está entre os 15 aos 29 anos. Para pesquisas futuras seria interessante isolar esses dados que fugiram do padrão observado e analisar as possíveis causas para esse comportamento.

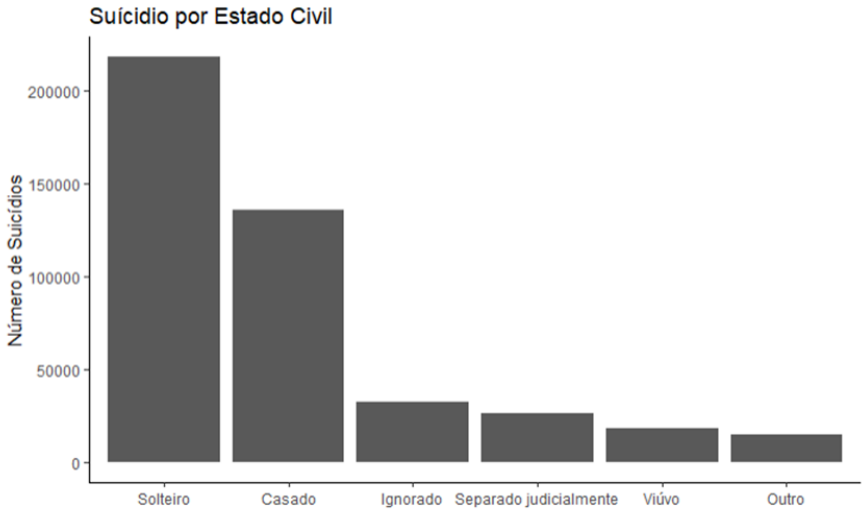
Figura 10 - Suicídios por Cor e Idade



Fonte: O autor (2022)

Para finalizar a análise exploratória, a figura 12 mostra que em relação ao estado civil os indivíduos solteiros possuem os maiores números de óbitos durante o período observado.

Figura 11 - Suicídios por Estado Civil

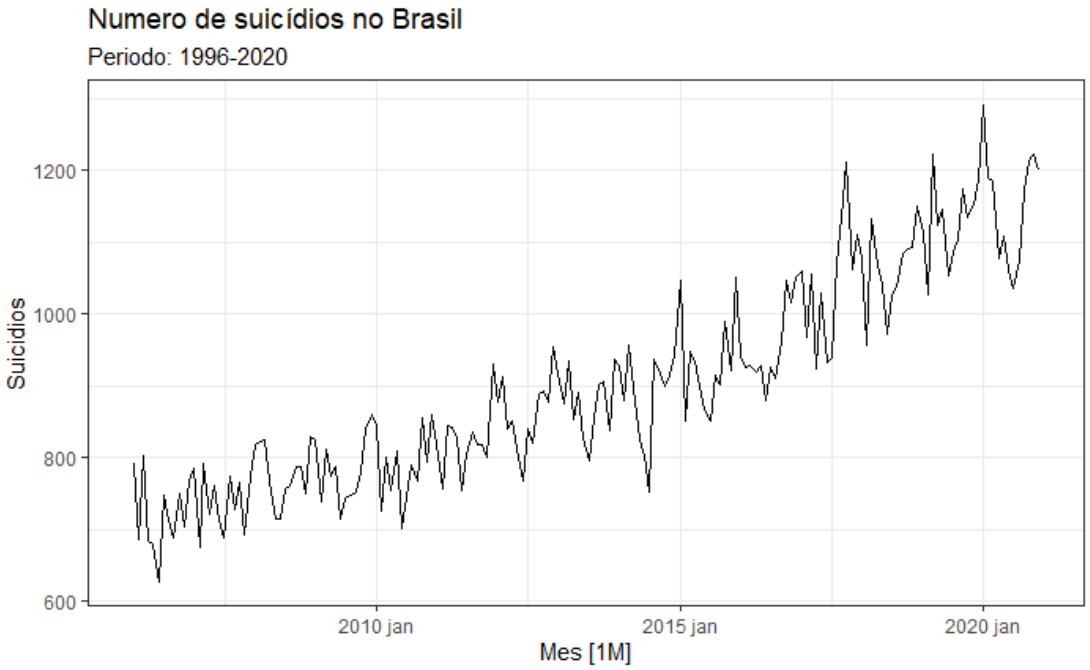


Fonte: O autor (2022)

4.3 ANÁLISE DE SÉRIES TEMPORAIS

Esse tópico do trabalho terá como objetivo analisar os dados sob a ótica de séries temporais. A figura 13 apresenta o comportamento do número de suicídios no Brasil ao longo do tempo, nota-se uma tendência de crescimento ao longo dos anos e uma sazonalidade a princípio não evidente.

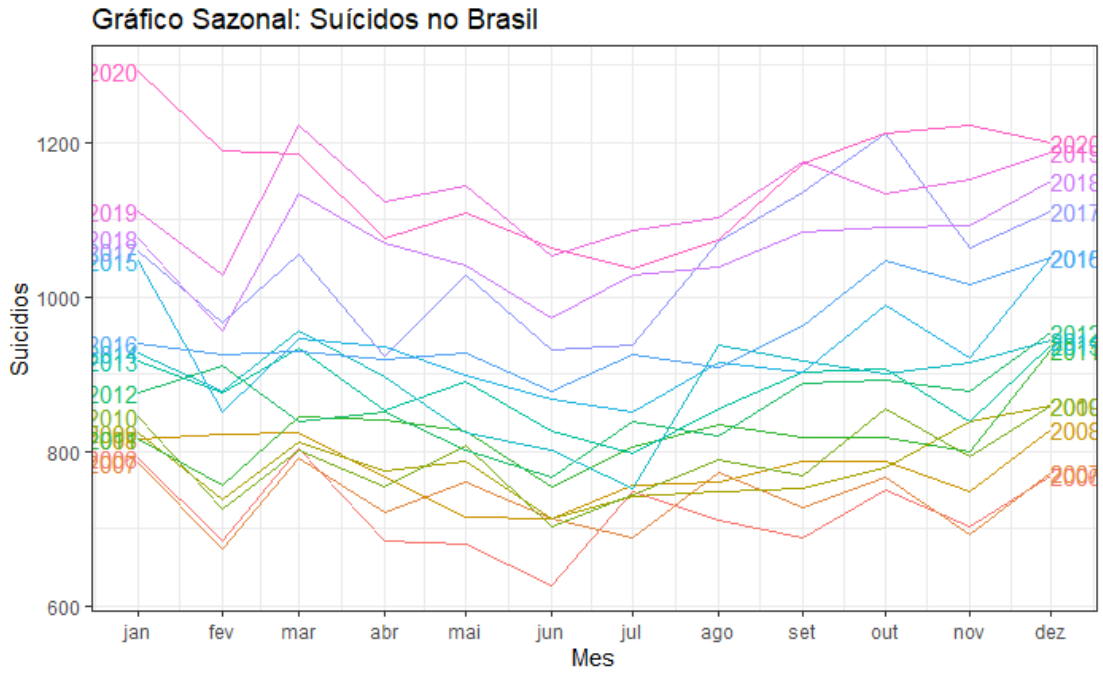
Figura 13 - Número de Suicídios no Brasil



Fonte: O autor (2021)

Foi avaliado o comportamento da série por mês utilizando um gráfico sazonal para encontrar possíveis padrões nos dados.

Figura 12 - Variação Sazonal de Suicídios no Brasil

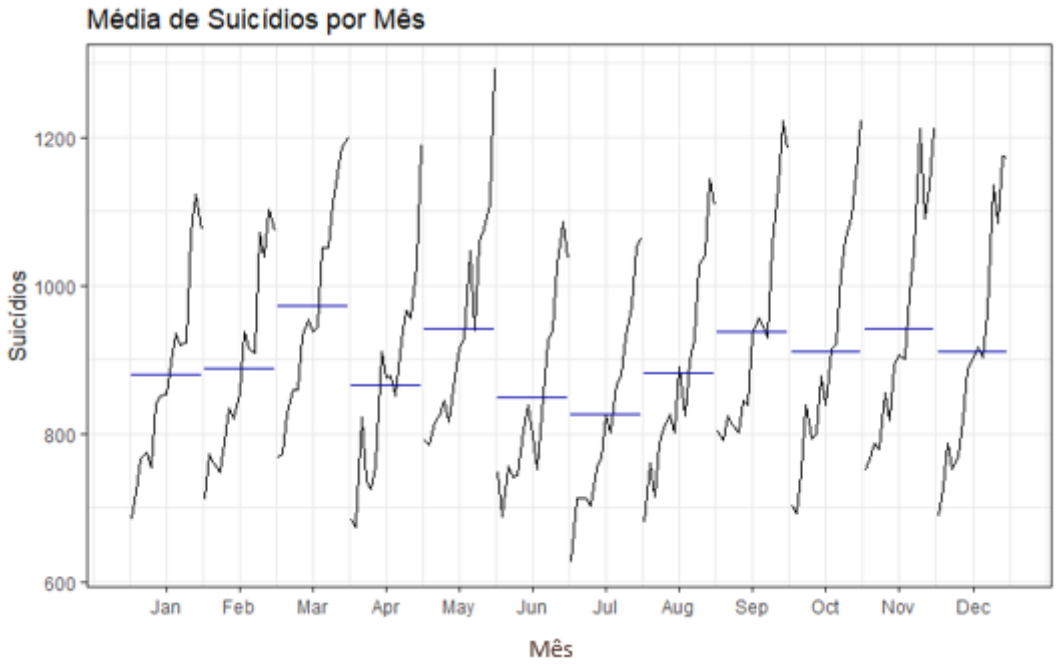


Fonte: O autor (2022)

Analisando os dados Sazonais, de certo há um crescimento das series todos os anos, embora não se pode afirmar que isso é algo puramente devido ao número de suicídios, pela melhora do processo de coleta dessas informações ou ao crescimento da população brasileira. Além disso, é possível identificar padrões sazonais ao longo dos meses com picos em março, agosto e outubro.

Observando a média de suicídios ao longo dos meses a partir na figura 15, março é o período onde merece um cuidado especial dos setores de saúde pública. Já o mês de junho e julho possuem uma média e variação menor se comparado a outros meses.

Figura 13 - Média de Suicídios por Mês

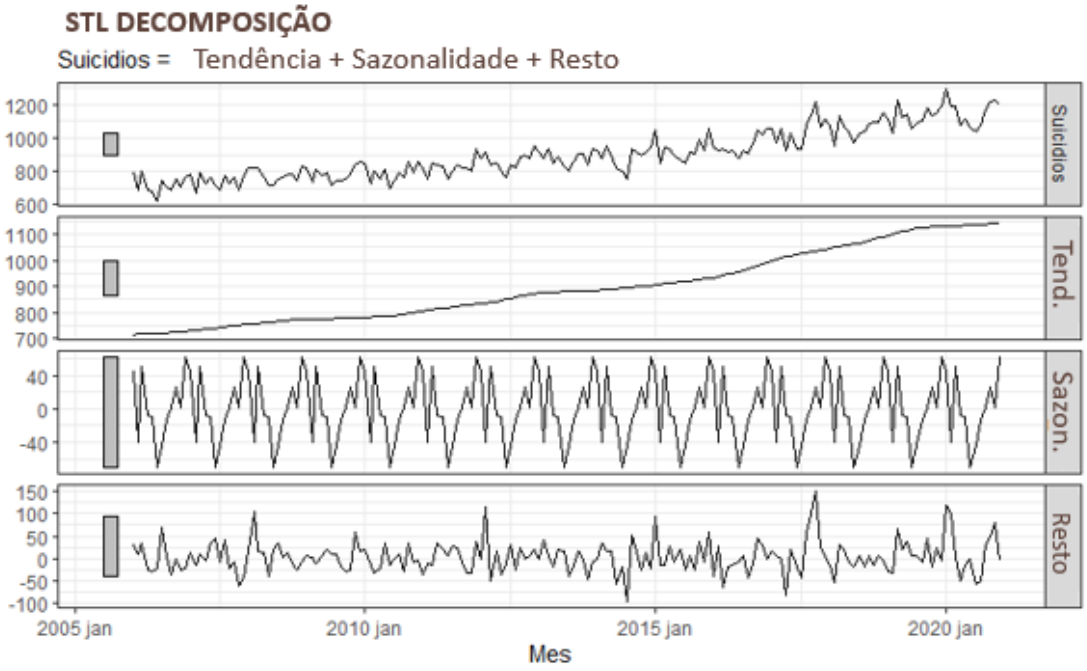


Fonte: O autor (2022)

Para decompor a série temporal em tendência, sazonalidade e erro foi utilizado o método STL, para Hyndman o mesmo é um método robusto e versátil que estima os parâmetros a partir de suas relações não lineares tendo algumas vantagens se comparados a outros métodos (SEATS e X11) como lidar com qualquer tipo de sazonalidade, permitir que os componentes sazonais possam mudar ao longo do tempo e ser mais robusto na presença de outliers. (2006)

A figura 16 apresenta o resultado da decomposição da série temporal estudada. A mesma apresenta indicativos de um componente de tendência crescente aditiva, sazonalidade com ciclo anual (12 meses) aditiva e um erro que a princípio aparenta seguir um modelo de ruído Branco.

Figura 14 - Decomposição da Série de Suicídios no Brasil



Fonte: O autor (2022)

Para analisar o impacto dos componentes sazonais e de tendência da série foi utilizado o método de Wang (Wang, Smith, & Hyndman, 2006) que calcula a força de tendência e a força sazonal utilizando as equações 26 e 27 respectivamente:

$$Força\ de\ tendência: MAX(0{,}1 - \frac{Var(Rt)}{Var(St + Rt)}) \quad (26)$$

$$Força\ de\ sazonalidade: MAX(0{,}1 - \frac{Var(Rt)}{Var(St + Rt)} \quad (27)$$

Onde:

- Rt = Componente resto isolada
- St = Componente da sazonalidade isolada
- Tt= Componente da tendência isolada

Para Wang, uma tendência ou sazonalidade forte devem possuir dados ajustados sazonalmente e de tendência com uma alta variação em relação aos

componentes restantes, logo quanto mais próximo de 1 for sua força, maior o impacto do componente na série estudada.

O método tem as seguintes classificações para tendência e sazonalidade:

- 0 a 0.19= Uma tendência ou Sazonalidade muito fraca
- 0.1 até 0.39 = Uma Tendência ou Sazonalidade Fraca
- 0.40 até 0.69 = Uma tendência ou Sazonalidade Moderada
- 0.70 até 0.89 = Uma Tendência ou Sazonalidade Forte
- 0.90 até 1= Uma Tendência ou Sazonalidade muito Forte.

Utilizando o software R, os valores encontrados foram:

$$Força\ de\ Tendência = 0{,}945$$

$$Força\ de\ sozonalidade = 0{,}62$$

Por conseguinte, a série estudada foi classificada com uma tendência muito forte e uma sazonalidade moderada, nesse sentido sendo um forte indicativo para a escolha do modelo de suavização exponencial de Holt-Winters pois o mesmo considera tanto a tendência quanto a sazonalidade.

4.4 DEFINIÇÂO DO MELHOR MODELO PARA PREVISÃO DE SUICÍDIOS NO BRASIL

4.4.1 Escolha de um modelo pelo método de suavização exponencial:

Utilizando a função ets() de otimização do R que estima os parâmetros de erro, tendência e sazonalidade a partir do menor AICc, maximizando a verossimilhança. (HYNDMAN et al., 2002)

O modelo estimado foi um ETS com erro multiplicativo, tendência aditiva e sazonalidade aditiva (M, A, A), seus parâmetros podem ser vistos na figura 17 logo abaixo.

Figura 15 - Saída do Software R: Modelo De Suavização exponencial

```
Series: Suicidios
Model: ETS(M,A,A)
  Smoothing parameters:
    alpha = 0.1426251
    beta  = 0.00497555
    gamma = 0.0001141694

  Initial states:
       l        b      s1        s2      s3        s4       s5        s6
 718.8198 1.740885 63.7531 -7.767268 30.2146 -7.742747 -8.04581 -44.12898
       s7        s8        s9      s10       s11      s12
 -71.38888 -12.01925 -12.66893 37.12647 -15.51388 48.18158

  sigma^2:  0.0021

     AIC     AICc      BIC
1785.223 1790.080 1835.710
```

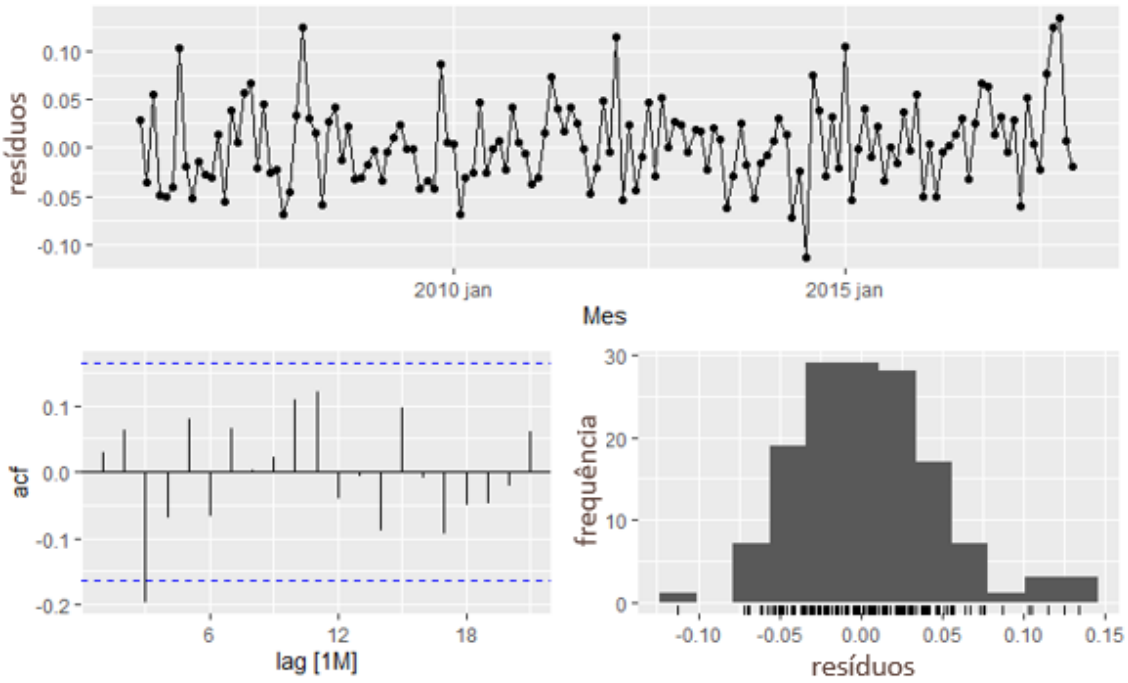
Fonte: O autor (2022)

O valor de alpha de 0.1426 indica que o modelo escolhido considera uma alta influência de observações mais distantes da série histórica. Já o valor da equação de tendência (beta) de 0,00497 indica que a inclinação da série de suicídios dificilmente mudará ao longo do tempo, o valor para gamma indica que dificilmente o componente sazonal irá mudar ao longo do tempo.

4.4.1.1 Análise dos resíduos

Analisando os resíduos, o gráfico ACF na figura 16 mostra que quase todos os lags não são significativos, o que implica que os resíduos não possuem autocorrelação entre si, sua série histórica aparenta seguir uma distribuição normal e variar com média 0.

Figura 16 – Plot dos resíduos gerados pelo modelo de Suavização exponencial.



Fonte: O Autor (2022)

Foi utilizado o teste de Ljung Box onde a hipótese nula é de que os dados não possuem auto correlação para um nível de confiança de 95%, aplicando o teste chegou-se no seguinte resultado:

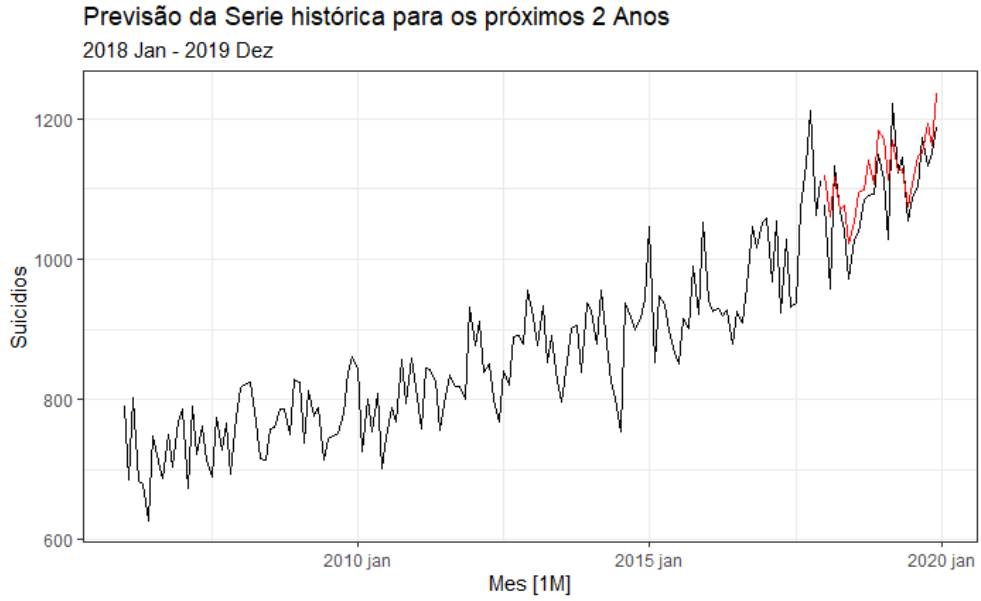
```
        Box-Ljung test

data:  residuos$.resid
X-squared = 11.705, df = 10, p-value = 0.3053
```

Com um p-value de 0.3053 não é possível rejeitar a hipótese nula de que os resíduos não possuem auto correlação.

4.4.1.2 Previsão com dados de teste

O gráfico 17 apresenta a comparação entre os dados estimados pelo modelo de suavização exponencial para os próximos dois anos (2018 e 2019) e as observações reais presentes na base de teste.

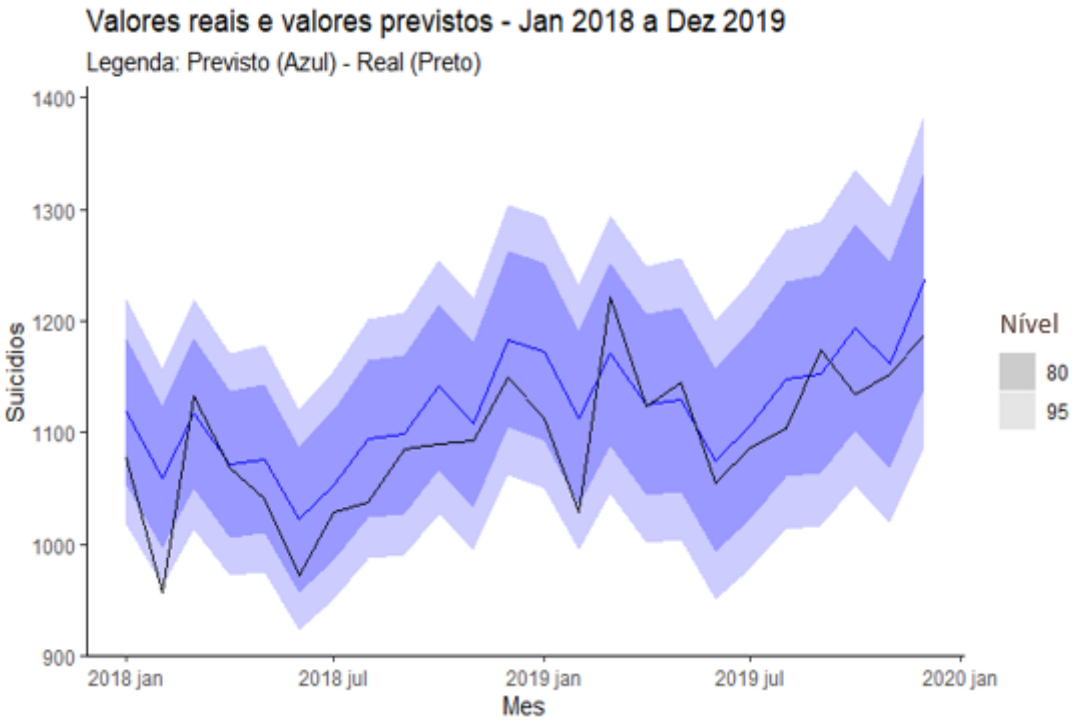
Figura 17 - Previsão de Séries temporais: Método de Suavização Exponencial



Fonte: O autor (2022)

Avaliando os resultados a partir da figura 18 percebe-se que o modelo conseguiu se ajustar bem aos dados e trouxe previsões precisas podendo o mesmo ser utilizado em políticas públicas para avaliar qual será o comportamento da série nos próximos anos. Além disso, todas as observações reais ficaram dentro dos intervalos de confiança de 80% e 95% da previsão estimada.

Figura 18 - Comparação entre valores estimas e reais para os níveis de confiança de 80% e 95%

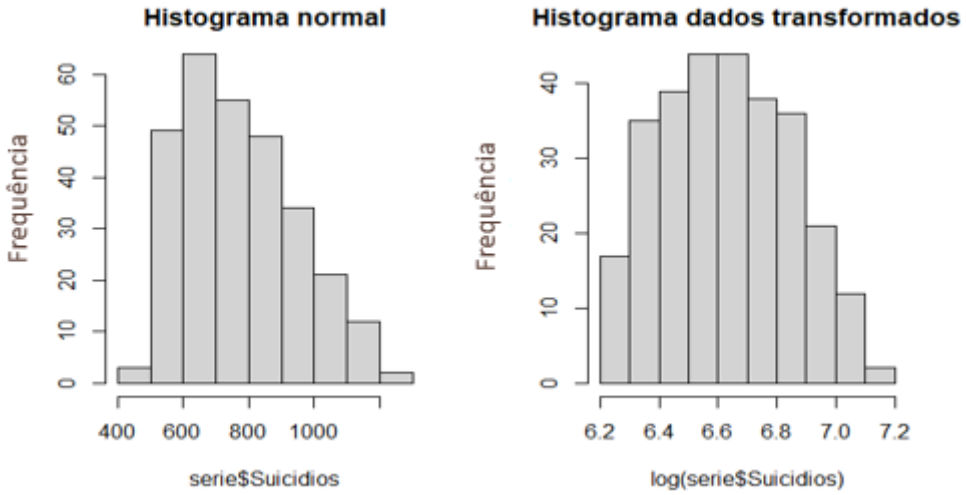


Fonte: O Autor (2022)

4.4.2 Escolha de um modelo utilizando o método ARIMA:

Os dados foram transformados utilizando uma escala logaritmica com objetivo de deixar mais normalizado a base tendo em vista que os modelos ARIMA se adaptam melhor a dados desse tipo.

Figura 19 - Histograma entre dados não transformados e transformados



Fonte: O Autor (2022)

Analisando o histograma da figura 19 percebe-se que o ajuste deixou os dados mais normalizados, porém utilizando o teste de Jarque Bera para avaliar a normalidade o valor p encontrado foi de 0.0083, sendo assim, rejeitamos a hipótese nula de que os dados são normais.

- Base não transformada

Figura 20 - Plot do Rstudio para o teste Jarque Bera para serie não transformada

```
        Jarque Bera Test

data:  serie$Suicidios
X-squared = 16.117, df = 2, p-value = 0.0003164
```

Fonte: O Autor (2022)

- Base transformada

Figura 21 - Plot do Rstudio para o teste Jaque Bera para série transformada

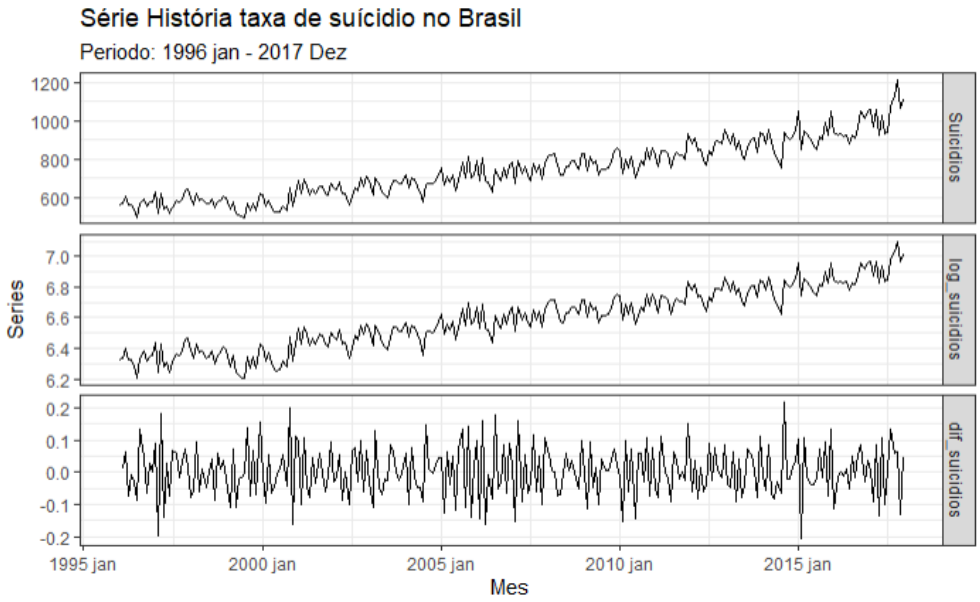
```
        Jarque Bera Test

data:  log(serie$Suicidios)
X-squared = 9.5829, df = 2, p-value = 0.0083
```

Fonte: O Autor (2022)

Foi utilizado o Teste de Kwiatkowski-Phillips-Schmidt-Shin para encontrar o número de diferenciações necessárias com objetivo de tornar a série estacionária, removendo o fator tendência e sazonalidade. O resultado encontrado indicou a

necessidade de apenas uma diferenciação ordinária. A figura 22 mostra a série histórica inicial e a final depois de todas as transformações realizadas.

Figura 22 - Comparação entre a série original e todas as transformações feitas.



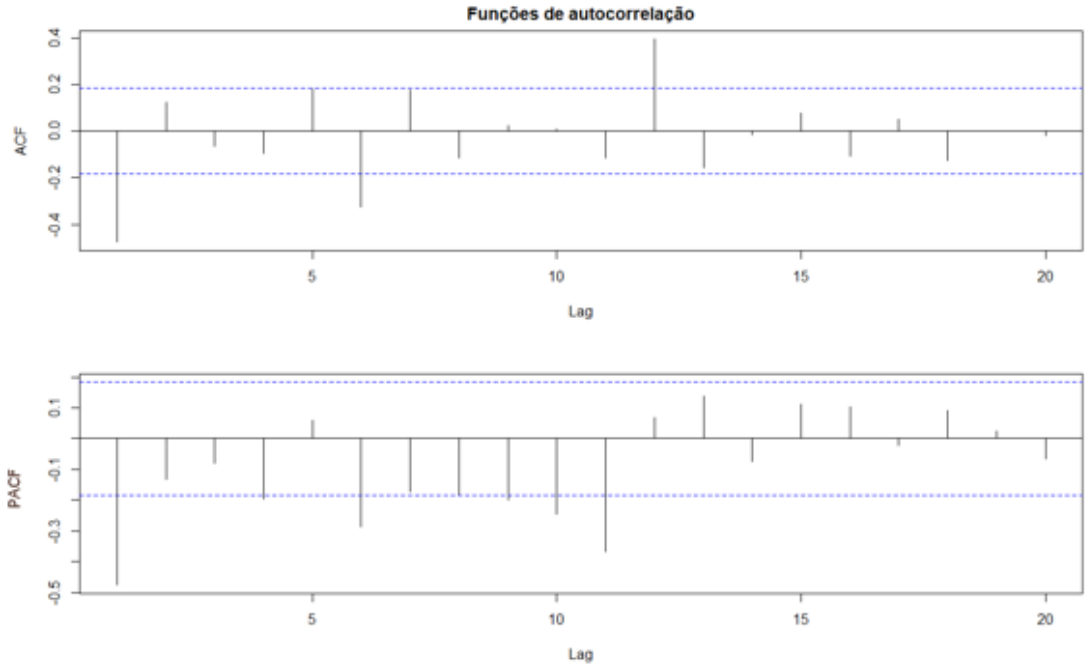
Fonte: O Autor (2022)

4.4.2.1 Função de autocorrelação e autocorrelação parcial

A figura 23 apresenta as funções de autocorrelação e autocorrelação parcial, pode-se supor alguns modelos a partir das seguintes informações observadas:

- A presença de lags significantes em ambos (6 e 12) indicam um provável modelo sazonal de ordem 6 ou 12.
- O pico significativo no lag 1 no ACF sugere um componente MA(1)
- O pico significativo no lag 1 a 11 no PACF pode indicar um componente AR(1) , AR(4) ou AR(11)

Figura 23 - ACF e PACF da série histórica de suicídios



Fonte: O Autor (2022)

Com base nos gráficos ACF e PACF foi sugerido os seguintes modelos para ser testado:

- ARIMA (1,1,1)
- ARIMA (4,1,1)
- ARIMA (11,1,1)
- ARIMA (4,1,1) (1,0,1) [12]
- ARIMA (4,1,1) (1,0,2) [6]

Além dos modelos definidos pelas funções de autocorrelação e autocorrelação parcial, foi utilizado a função do R auto.arima, que estima o modelo a partir do menor AIC. (HYNDMAN, 2008)

O modelo sugerido pelo comando do programa R foi:

- ARIMA (0,1,1) (0,0,2) [12]

4.4.2.2 Modelos estimados

Os modelos foram treinados utilizando dados de suicídio entre os períodos de 1996 a 2017, utilizando como critério de escolha o valor do AIC chegou-se no seguinte resultado:

Tabela 2 - Valores dos AICs para os modelos estimados

MODELO	AIC
ARIMA (1,1,1)	1.251
ARIMA (4,1,1)	1251
ARIMA (11,1,1)	1238
ARIMA (4,1,1) (1,0,1) [12]	1197
ARIMA (4,1,1) (1,0,2) [6]	1215
ARIMA (0,1,3) (0,0,2) [12]	1222

Fonte: O Autor (2022)

Logo o melhor modelo segundo o AIC foi ARIMA (4,1,1) (1,0,1) [12], a figura 24 mostra os parâmetros estimados pelo R para o modelo escolhido:

Figura 24 - Parâmetros estimados para o modelo ARIMA (4,1,1) (1,0,1) [12]

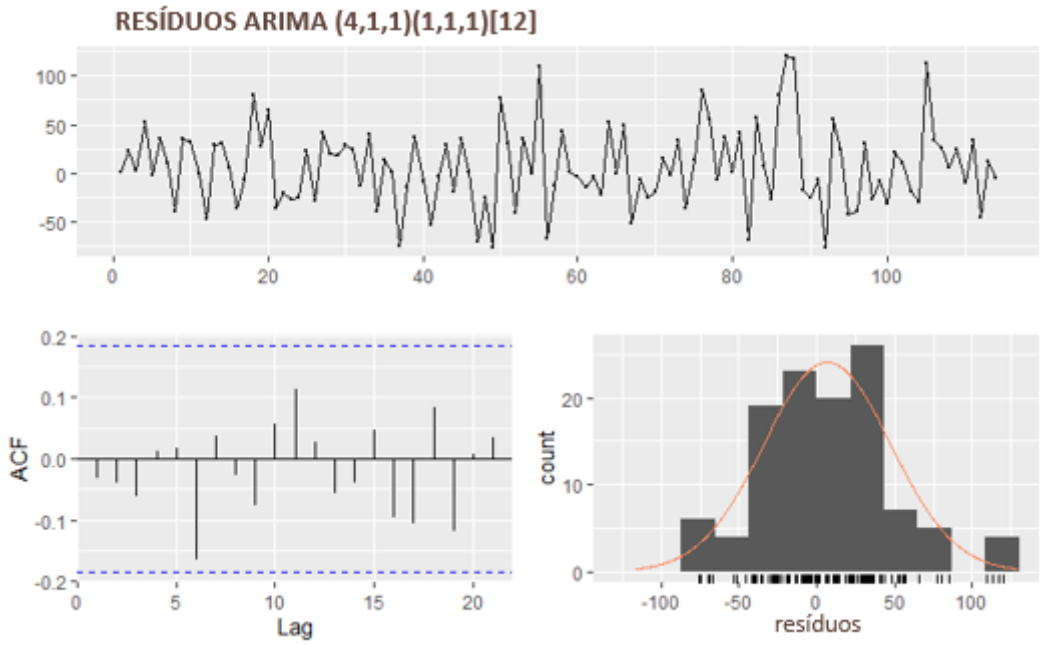
```
Call:
arima(x = treino_arima$Suicidios, order = c(4, 1, 1), seasonal = c(1, 0, 1,
    12))

Coefficients:
         ar1     ar2      ar3      ar4      ma1    sar1     sma1
      0.0152  0.1596  -0.1443  -0.1157  -0.8085  0.9922  -0.8657
s.e.  0.1226  0.1073   0.1018   0.1039   0.0850  0.0241   0.2015

sigma^2 estimated as 1722:  log likelihood = -591.38,  aic = 1196.75
```

Fonte: O Autor (2022)

4.4.2.3 Análise dos resíduos

Para garantir que todos os parâmetros do modelo são estatisticamente significativos, foi analisado se os resíduos é um ruído branco. Os gráficos abaixo apresentam um resumo do comportamento dos resíduos gerados:

Figura 25 - Plot dos resíduos gerados pelo modelo ARIMA.



Fonte: O Autor (2022)

Analisando os resíduos, o gráfico ACF mostra que todos os lags não são significativos, o que implica que os resíduos não possuem autocorrelação entre si. Foi utilizado o teste de Ljung Box onde a hipótese nula é de que os dados não possuem auto correlação para um nível de confiança de 95%, aplicando o teste chegou-se no seguinte resultado:

Figura 26 - Teste de autocorrelação dos resíduos gerados

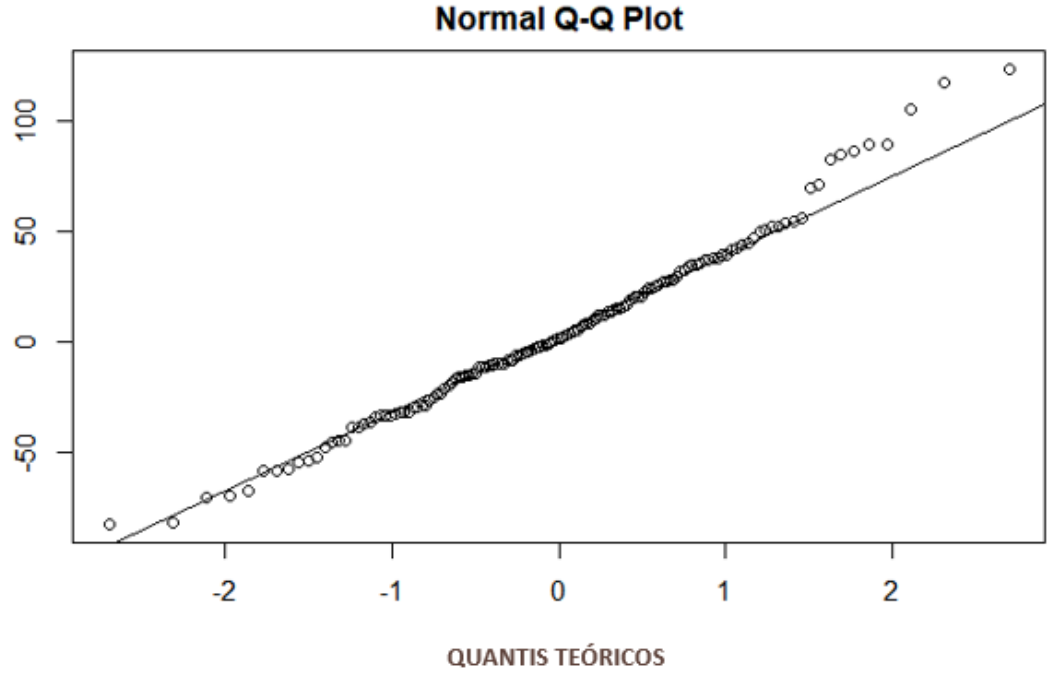
```
        Ljung-Box test

data:  Residuals from ARIMA(4,1,1)(1,1,1)[12]
Q* = 5.5364, df = 3, p-value = 0.1365

Model df: 7.   Total lags used: 10
```

Fonte: O Autor (2022)

Com um p-value de 0.1365 não é possível rejeitar a hipótese nula de que os resíduos são aleatórios e independentes. Logo o modelo sugerido é estatisticamente significativo.

O gráfico abaixo apresenta os resíduos em relação aos quartis teóricos, o mesmo mostra evidências de que os resíduos seguem uma distribuição normal embora exista uma assimetria a direita quando comparada ao valor esperado.

Figura 27 - Quantis teóricos dos resíduos

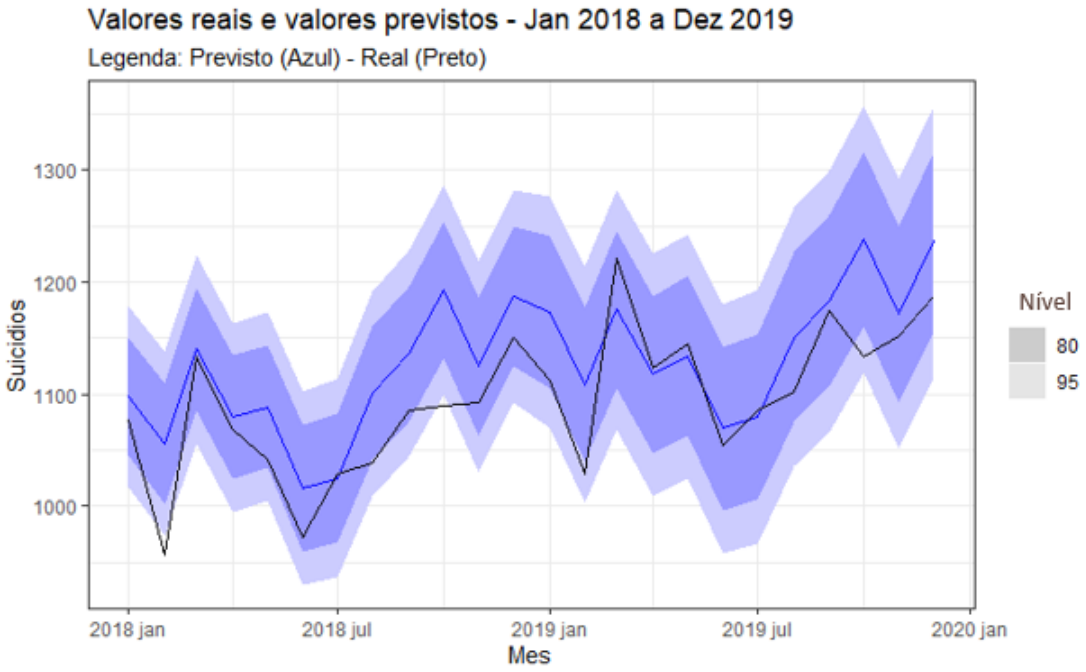


Fonte: O Autor (2022)

4.4.2.4 Previsão com dados de teste

Foi utilizado como dados de teste os números de suicídios durante o período de 2018 a 2019, a figura 28 apresenta uma comparação entre a previsão feita pelo modelo ARIMA e os dados reais. O modelo aderiu bem aos dados de forma que quase todas as observações reais estão dentro do intervalo de confiança de 95%.

Figura 28 - Comparação entre valores estimados e reais do modelo ARIMA.



Fonte: O Autor (2022)

4.4.3 Definindo o melhor modelo, ARIMA vs ETS

Para definir o melhor modelo foi considerado a métrica de acuracidade RMSE (Erro quadrático médio), onde segundo Hyndman é uma das medidas mais utilizadas para comparar series temporais. (2008)

Foi levado em consideração o erro das previsões nos dados de teste, número de suicídios durante o período de 2018 a 2019, em relação aos valores previstos pelo modelo para o mesmo período.

A tabela abaixo apresenta os resultados de acuracidade do modelo ARIMA (4,1,1) (1,1,1) [12] e o modelo de suavização exponencial (M, A, A).

Figura 29 - Comparação entre as métricas RMSE do modelo ARIMA e ETS

ACURACIDADE	
MODELO	RMSE
ARIMA	38,17
ETS	37,40

Fonte: O Autor (2022)

Embora os dois modelos tenham apresentados valores próximos, o modelo de suavização exponencial apontou um erro médio quadrático menor, o que significa que suas previsões estão mais próximas dos valores reais e apresentam menos erros de maior magnitude, tendo em vista que o RMSE é sensível a outliers.

4.5 IMPACTOS DO SETEMBRO AMARELO NOS NÚMEROS DE SUICÍDIOS

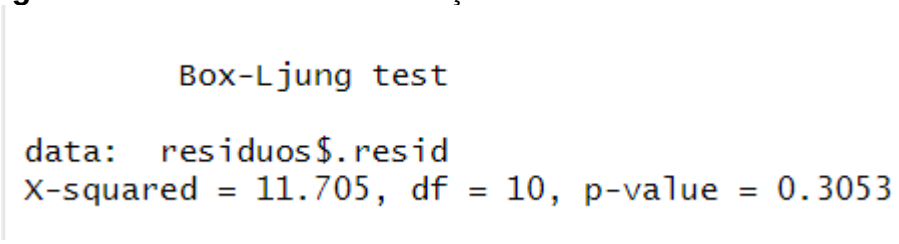
Com objetivo de conscientizar a população brasileira e atuar como uma campanha de prevenção a ideações suicidas foi criado em 2015 pelo Centro de Valorização a Vida, Conselho Federal de Medicina e pela Associação Brasileira de Psiquiatria o movimento “Setembro Amarelo”. (OLIVEIRA et al., 2020)

Tal movimento se utiliza de diversas ações como informativos, palestras, divulgações de taxa de suicídios para alertar as pessoas de que o suicídio não pode ser um tabu na sociedade e que pessoas com comportamento suicida devem procurar ajuda.

Nesse sentido, resolve-se utilizar o modelo de suavização exponencial ajustado anteriormente para avaliar se a campanha trouxe efeitos positivos no número de óbitos por suicídio.

A figura 30 apresenta o teste de aleatoriedade de Ljung Box, com um p-value maior de 0.30. Logo pode-se considerar a hipótese nula de que os resíduos seguem uma serie de ruído branco.

Figura 30 - Teste de autocorrelação dos resíduos do modelo ARIMA

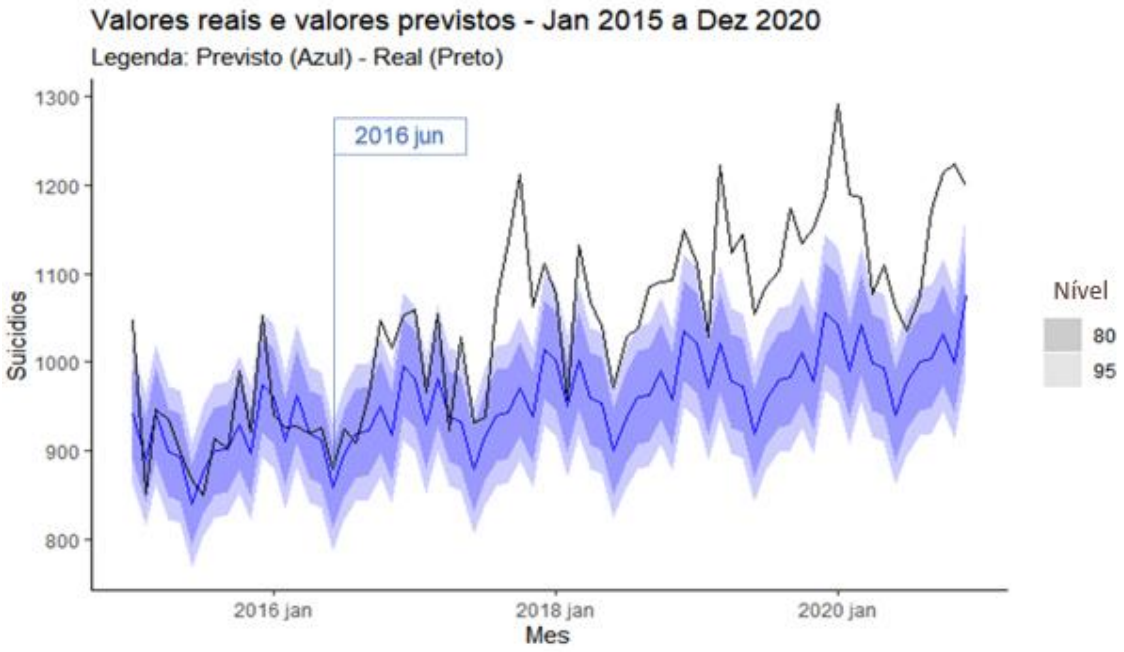


Fonte: O Autor (2022)

A figura 31 compara as previsões do modelo ajustado em relação aos dados reais para os períodos de 2015 a 2020. Para as observações anteriores a junho de 2016 o modelo gerou previsões aceitáveis com a maioria dos valores dentro do

intervalo de confiança de 95%, contudo após esse período os dados reais apresentaram um crescimento fora do esperado.

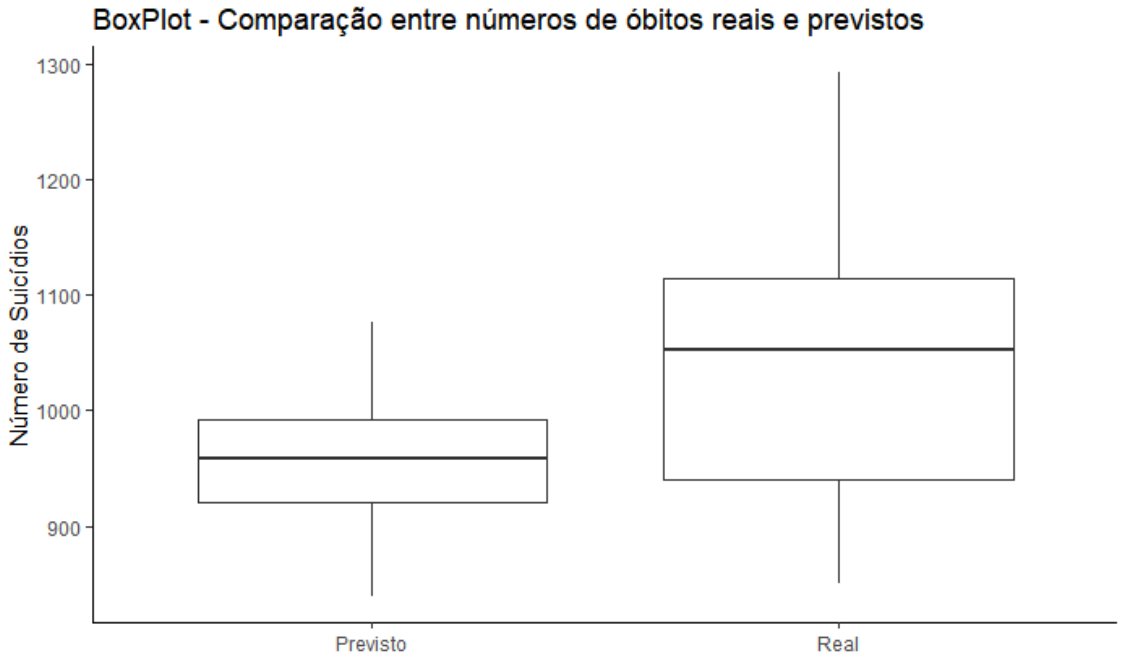
Figura 31 - Comparação entre os valores reais e previstos entre 2015 a 2020



Fonte: O Autor (2022)

A diferença entre os valores reais e previstos fica mais evidente observando o boxplot na figura 32, a maioria dos valores previstos estão abaixo do segundo quartil das observações reais, sendo este um forte indício de que a diferença entre as duas distribuições é estatisticamente significativa.

Figura 32 - BoxPlot dos valores previstos e esperados de 2015 a 2020



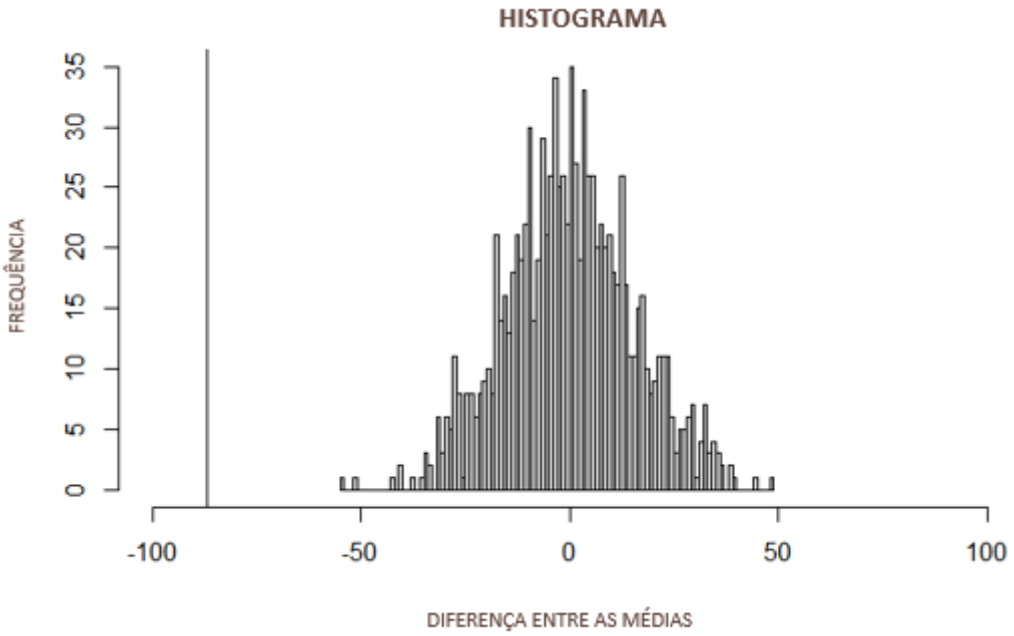
Fonte: O Autor (2022)

Foi testado a seguinte hipótese a partir do método de bootstrap para um nível de confiança de 95%:

- Ho = A diferença entre valor médio previsto e valor médio Real não é estatisticamente significativa.
- H1 = A diferença entre valor médio previsto e valor médio Real é estatisticamente significativa.

O gráfico abaixo apresenta a simulação feita reamostrando 1000 vezes os dados, o histograma não só mostra que a diferença é estatisticamente significativa a partir de um p-value = 0 , mas também evidencia que os valores reais são maiores do que os valores previstos.

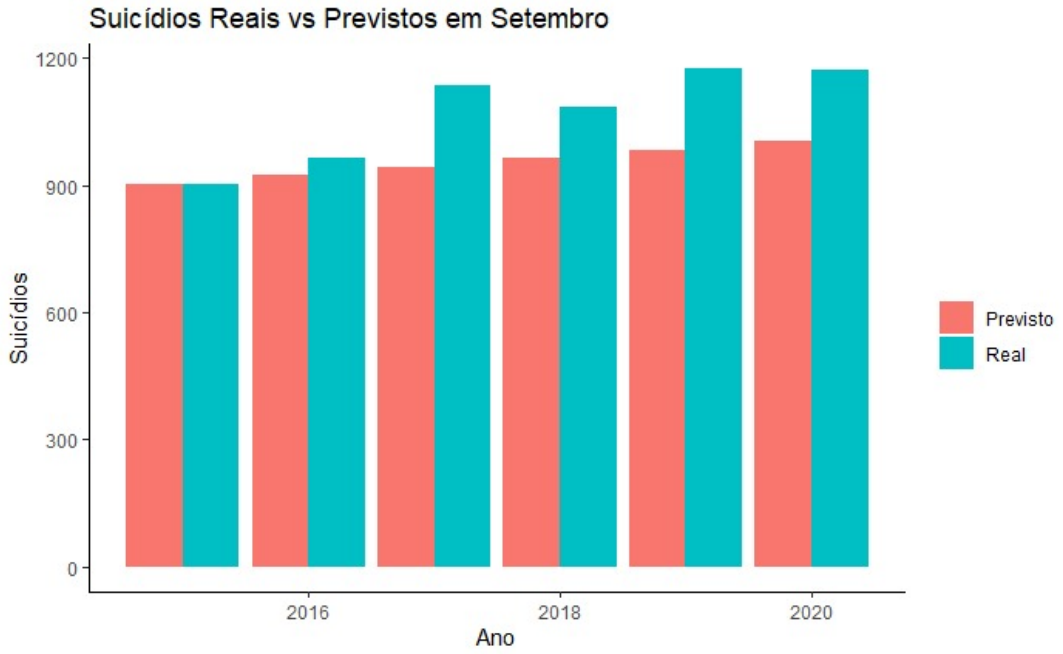
Figura 33 - Simulação Bootstrap da diferença entre as médias



Fonte: O Autor (2022)

Por conseguinte, o gráfico de barras na figura 34 compara para cada ano os números de óbitos reais e previstos apenas no mês de setembro. Não só houve um crescimento se comparado aos dados estimados pelo modelo, mas também em relação ao mesmo período do ano anterior.

Figura 34 - Comparação anual entre os valores previstos e reais durante o mês de setembro.



Fonte: O Autor (2022)

Embora esses resultados não comprovem uma falta de impacto positivo da campanha Setembro Amarelo em relação a diminuição do número de óbitos por suicídios, levanta-se o questionamento sobre a sua eficácia.

Tais achados corroboram com os estudos de Oliveira et al.(2020) que identificou um aumento nos números de notificações de suicídios após o início do CSA. Outro estudo feito no Brasil por Oliveira MB (2018) também identificou um aumento no número de suicídios após a campanha do setembro amarelo, atribuindo como causa o efeito contágio, onde acredita-se que a divulgação de notícias sobre suicídios pode reforçar comportamentos suicidas, sendo está a terceira principal causa de suicídios.

4.6 IMPACTOS DO COVID-19 NOS ANOS DE 2020

O COVID-19 trouxe não só impactos na saúde física, mas também na saúde mental. Para Trakur (2020) a pandemia acarretou diversos fatores de risco como:

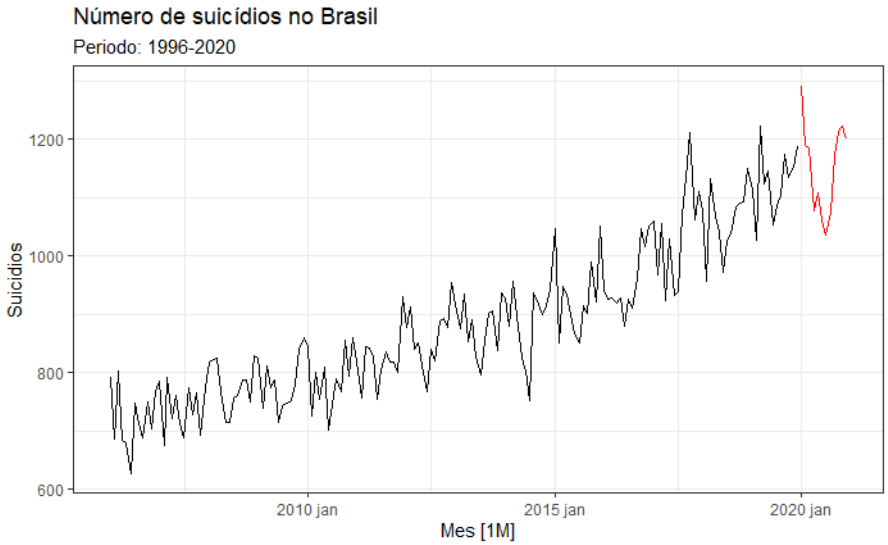
- Distanciamento Social: Induzem a sintomas de ansiedade, o interrompimento de uma vida social mais ativa pode ser o gatilho para pensamentos suicidas. O primeiro caso de suicídio relatado devido a

pandemia ocorreu na Índia em 12 de fevereiro de 2020, onde um homem de 50 anos com suspeita de COVID-19 se colocou em quarentena para proteger a família, perturbado com as informações sobre número de mortes em sua região acabou tirando a própria vida. (GOYAL et. Al., 2020).

- Recessão econômica: Uma crise econômica gera pânico, desemprego, pobreza. Esses fatores podem ser um gatilho por fazerem o indivíduo não ter esperança sobre o futuro, como foi o caso do ministro alemão de Finanças, Thomas Schaefer, profundamente preocupado com os impactos da pandemia na economia acabou cometendo suicídio. (ESTADO DE MINAS INTERNACIONAL, 2020)
- Boicote Social e a Discriminação: Foi registrado alguns casos em que um indivíduo com sintomas foi isolado socialmente ou sofreu algum tipo de discriminação, na Índia até agosto de 2020 foram relatados 4 casos pelo motivo citado de 69 suicídios registrado. (DSOUZA et al., 2020)

Para conferir os impactos da pandemia do COVID-19 no Brasil foi coletado os dados de suicídios em 2020 pelo DATASUS, a figura 35 apresenta a série histórica de mortes autoprovocadas, o período destacado em vermelho são as informações relacionadas ao ano de 2020.

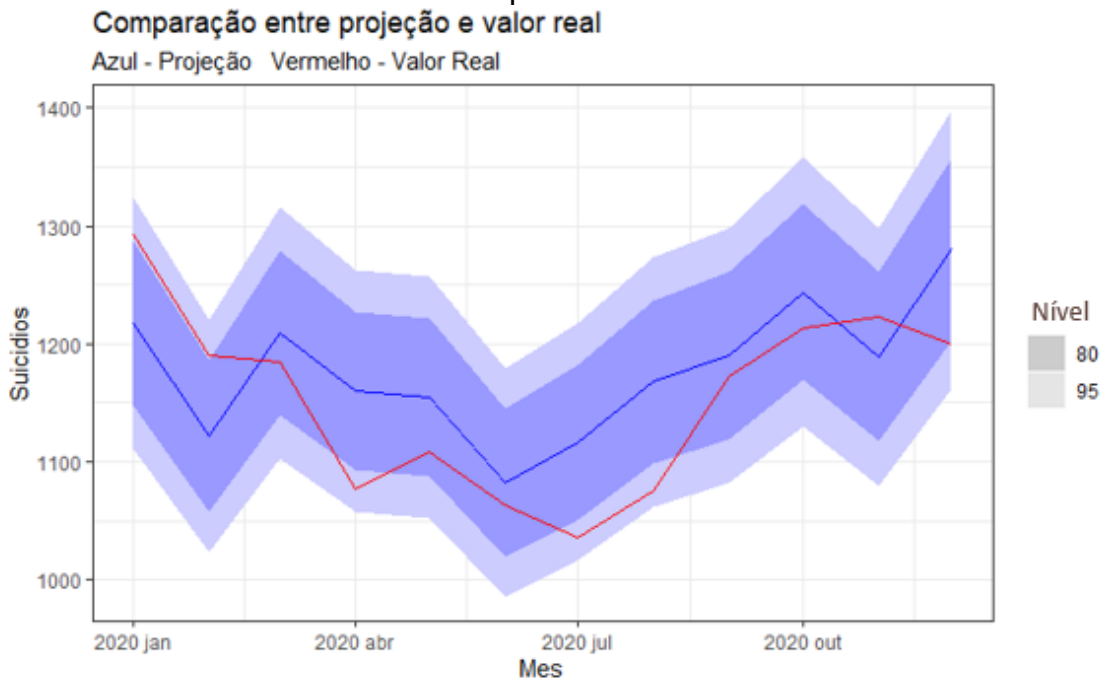
Figura 35 - Número de suicídios no Brasil



Fonte: O Autor (2022)

O gráfico abaixo apresenta uma comparação entre a projeção esperada de mortes autoprovocadas sem os efeitos da pandemia a partir do modelo de suavização exponencial criado anteriormente e os dados reais.

Figura 36 – Comparação entre o número real de suicídios em 2020 e os valores esperados

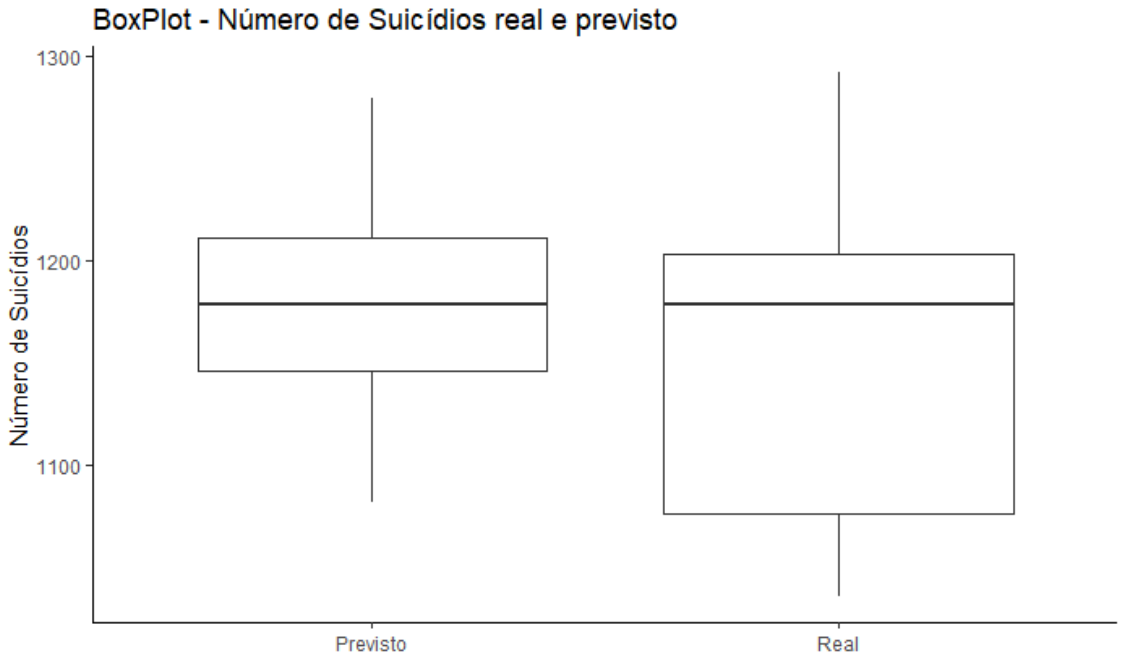


Fonte: O Autor (2022)

Os números reais apresentam um padrão diferente do que é esperado pelo modelo de suavização exponencial, mas a maior parte das observações ficam dentro do intervalo de confiança de 95%, fornecendo indícios de que o ano de 2020 não possuiu um aumento ou decrescimento no número de mortes.

O gráfico abaixo identifica que os dados reais possuem uma alta variância no 1ºe 2º quartil em relação ao esperado que possui um comportamento mais homogêneo, já no 3º e 4º quartil os valores reais seguem um padrão esperado ao que o modelo prever.

Figura 37 - BoxPlot entre os valores reais e previstos para 2020



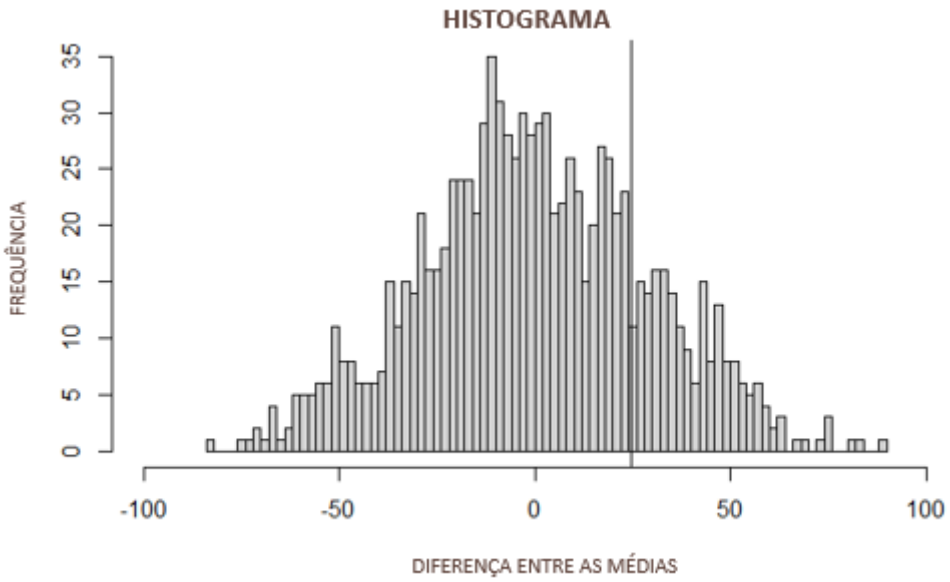
Fonte: O Autor (2022)

Para garantir que a diferença nas duas distribuições é estatisticamente significativa foi feito um teste de hipótese a partir do método de bootstrap onde foi reamostrado diversas vezes a diferença entres as médias dos valores previstos e reais. Foi definido as seguintes hipóteses para um nível de confiança de 95%:

- Ho = A diferença entre valor médio previsto e valor médio Real não é estatisticamente significativa.
- H1 = A diferença entre valor médio previsto e valor médio aparado Real é estatisticamente significativa.

A figura 38 apresenta um histograma com a distribuição das diferenças simuladas a partir do bootstrap e a linha vertical que identifica a diferença média entre as duas distribuições.

Figura 38 - Simulação bootstrap em relação as diferenças entre as médias para 2020

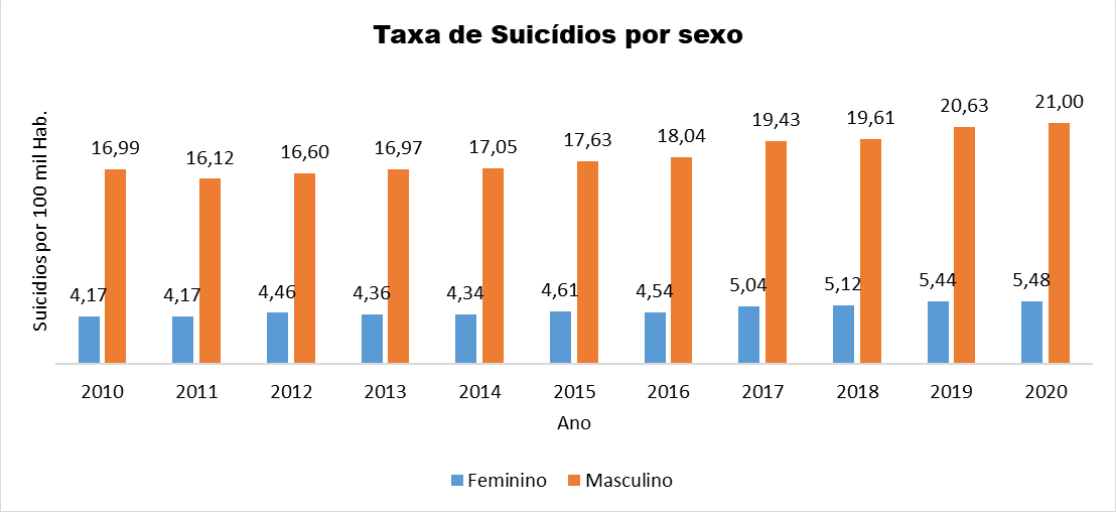


Fonte: O Autor (2022)

Com um p-value de 0,194 não é possível rejeitar a hipótese nula de que a diferença encontrada nas duas distribuições, real e previsto é estatisticamente significativa.

O número de suicídios não teve uma diferença significativa também quando a análise é feita por sexo, comparando a diferença entre a taxa de mortes autoprovocada durante a pandemia em 2020 e o ano anterior a pandemia, 2019, foi em média 1% e 2% de crescimento para o sexo masculino e sexo feminino respectivamente.

Figura 39 - Taxa de suicídios entre sexo de 2010 a 2020



Fonte: O Autor (2022)

5 CONCLUSÃO

No início do trabalho partiu-se da ideia de que realizar a pesquisa era relevante para averiguar os benefícios do uso de métodos de séries temporais para encontrar padrões de comportamento suicida, sazonalidade, tendência e modelos de previsão.

Desse modo constata-se após a realização da pesquisa que sua relevância é ainda maior, tendo em vista que o método seguido para avaliar questões sobre a saúde mental pode ser utilizado para guiar e mensurar políticas públicas na saúde para qualquer tema desejado, por exemplo, entender o comportamento temporal das internações ou óbitos da doença X e se a política pública Y teve êxito em diminuir os impactos desta comorbidade na sociedade.

No projeto estabeleceu-se como objetivo geral avaliar a evolução e o impacto da Pandemia da COVID-19 na distribuição dos suicídios no Brasil considerando a série história entre os anos de 1996 a 2020. Percebe-se neste momento que tal meta foi atingida tendo em vista que a partir da utilização de modelos de series temporais chegou-se como conclusão que não houve evidências que a pandemia teve um forte impacto para aumento do número de mortes por suicídios.

Em relação a identificar as características dos indivíduos que cometeram suicídios no Brasil e seus padrões temporais. Detectou-se aspectos epidemiológicos como:

- Pessoas do Sexo masculino se suicidam mais que pessoas do sexo feminino, embora as mulheres tem um número maior de tentativas em comparação aos homens.
- Enforcamento é o principal método escolhido, ou ao menos o mais efetivo para cometer suicídios.
- No mês de março e julho, tende-se a ter o maior e o menor número de registro de óbitos no Ano respectivamente.

Quanto a desenvolver modelos de previsão utilizando métodos de suavização exponencial e ARIMA chegou-se em dois modelos com bons ajustes, um modelo de suavização exponencial de erro multiplicativo com tendência e sazonalidade aditiva e um modelo ARIMA de ordem pdq (4,1,1) e PDQ (1,0,1) com sazonalidade de 12 meses.

Com base na métrica da raiz do erro quadrático médio (RMSE) identificou-se como melhor modelo o desenvolvido pelo método de suavização exponencial (M, A, A) se comparado ao modelo ARIMA.

Sobre o objetivo de esclarecer se o COVID-19 trouxe impactos nos números de óbitos por suicídios no Brasil. Constatou-se que embora a variação observada durante o ano de 2020 tenha um padrão diferente do estimado pelo modelo ETS(M, A, A), de certo, a diferença entre eles não é significativa. Concluindo-se que não é possível considerar que a pandemia aumentou o número de mortes por suicídios.

De mesmo modo, se comparado a taxa de suicídios por sexo entre o ano de 2019 e 2020, identifica-se que não houve um aumento significativo em ambos os sexos mesmo com o fator pandemia.

Contudo, de maneira alguma deve ser diminuído a atenção da sociedade quanto ao tema de suicídios, tendo em vista que uma das limitações desse projeto foi apenas avaliar o impacto da pandemia de COVID-19 no número de óbitos sem considerar o número de tentativas mal sucedidas e doenças mentais que tem uma alta correlação com futuros casos de mortes autoprovocadas como ansiedade e depressão. Ressalta-se também como uma limitação o curto período de dados disponível em relação a pandemia, 2020, existe uma possibilidade de um aumento no número de suicídios devido ao COVID-19 nos anos de 2021 e 2022.

Ao analisar a eficácia da campanha Setembro amarelo no combate dos suicídios no Brasil, chegou-se na conclusão de que não existem evidências que comprovem a eficiência da mesma, tal achado pode ser causado pelo efeito de contágio com a alta divulgação desse conteúdo pela mídia.

Com base nos dados sazonais pode-se sugerir como melhor mês para a campanha Setembro Amarelo como o mês de março por possuir no geral o maior pico de suicídios nos anos observados.

Com relação às dificuldades para o desenvolvimento do trabalho, destaca-se a etapa de levantamento e tratamento dos dados. O layout de como as informações são apresentadas e a necessidade de extrai-las separadamente em vários arquivos na plataforma do DATASUS foram os principais obstáculos enfrentados.

Para este projeto houve-se algumas limitações, quanto aos dados históricos existe um gap de 2 anos, os anos de 2021 e 2022 ainda não possuem números oficiais publicados pelo DATASUS.

Além disso, para algumas análises descritivas comparativas não foi possível trabalhar com a taxa de suicídios por não encontrar estimativas populacionais no ano ou mês sobre a característica comparada. (cor, faixa etária e raça).

Para trabalhos futuros recomenda-se a inclusão de outras bases de dados como variáveis de controles aos modelos ajustados como índice de desemprego, número de internações por outras doenças mentais como ansiedade e depressão. Ainda mais, é importante desenvolver uma análise dos impactos do COVID-19 e outras informações discretizadas por estado, afinal, pode-se existir comportamentos diferentes para cada estado ou região.

REFERÊNCIAS

ARAUJO, M. F. C. et al. Precipitação pluviométrica mensal no estado do rio de janeiro: sazonalidade e tendência. **Bioscience Journal**, Uberlândia, v. 25, n. 4, p. 90–100, 2009. Disponível em: https://sumarios.org/artigo/precipita%C3%A7%C3%A3o-pluviom%C3%A9trica-mensal-no-estado-do-rio-de-janeiro-sazonalidade-e-tend%C3%AAncia. Acesso em: 11 mar. 2022

ARRUDA , Vilmeyze Larissa *et al*. Suicídio em adultos jovens brasileiros: série temporal de 1997 a 2019. **Ciênc. Saúde Colet**, [*s. l.*], 2 jul. 2021. DOI https://doi.org/10.1590/1413-81232021267.08502021. Disponível em: https://www.scielosp.org/article/csc/2021.v26n7/2699-2708/. Acesso em: 11 mar. 2022.

BABALOLA, O.; RAZZAQUE A.; BISHAI, D. Temperature extremes and infant mortality in Bangladesh: Hotter months, lower mortality. **PLoS ONE [online]**, SanFrancisco, v. 13, n. 1, 2018. Disponível em: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189252#:~:text=Hotter%20months%20were%20associated%20with,%2C%20p%3C0.05)%20points. Acesso em: 29 mar. 2022.

BRANAS, Charles C *et al*. The impact of economic austerity and prosperity events on suicide in Greece: a 30-year interrupted time-series analysis. **Bmjopen**, [*S. l.*], p. 1-30, 2 fev. 2015. DOI 10.1136/bmjopen-2014-005619. Disponível em: https://bmjopen.bmj.com/content/5/1/e005619. Acesso em: 7 mar. 2022.

BRASIL. Ministério da Saúde. Suicídio: tentativas e óbitos por intoxicação exógena no Brasil, 2007 a 2016. **Boletim Epidemiológico**, v. 50, n. 15, 2019a. Disponível em: https://antigo.saude.gov.br/images/pdf/2019/julho/17/2019-014-Publicacao-02-07.pdf. Acesso em: 2 mar. 2022.

BRASIL (Senado Federal). Agencia Senado. **Casos de ansiedade, depressão e suicídio aumentam durante pandemia**. [*S. l.*], 10 set. 2020. Disponível em: https://www12.senado.leg.br/noticias/noticias/audios/2020/09/casos-de-ansiedade-depressao-e-suicidio-aumentam-durante-pandemia. Acesso em: 16 mar. 2022.

BOTEGA, Neury José. Comportamento suicida: epidemiologia. **Dossiê Suicídio - Psicol**, [*s. l.*], 2014. DOI https://doi.org/10.1590/0103-6564D20140004. Disponível em: https://www.scielo.br/j/pusp/a/HBQQM7PGMRLfr76XRGVYnFp/abstract/?lang=pt. Acesso em: 13 jan. 2022.

Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. Time series analysis forecasting and control, **3ed. Jhon Wiley & Song. 1994** New Jersey: Prentice Hall. Disponível em: https://www.worldcat.org/title/time-series-analysis-forecasting-and-control/oclc/28888762. Acesso em: 30 mar. 2022

BUENO, Rodrigo de Losso da Silveira. Econometria de Séries Temporais. Cengage Learning, São Paulo, v.3 , 2012.

DSOUZA, Deena Dimple *et al*. Aggregated COVID-19 suicide incidences in India: Fear of COVID-19 infection is the prominent causative factor. **ELSERVIER**, [*s. l.*], 1 ago. 2020. DOI https://doi.org/10.1016/j.psychres.2020.113145. Disponível em: https://www.sciencedirect.com/science/article/pii/S016517812031489X. Acesso em: 15 abr. 2022.

FAVA, V. L.; Análise de Séries de Tempo. In: VASCONCELLOS, M. A. S.: ALVES, D. Manual de Econometria: nível intermediário. São Paulo: Ed. **Atlas**, 2000.

GONÇALVES, Victor Henrique. **Previsão de séries temporais econômicas usando redes neurais caóticas**. 2018. Dissetação (Mestrado de pós-graduação em Ciências de Computação) - Universidade de São Paulo, [*S. l.*], 2018. Disponível em: https://teses.usp.br/teses/disponiveis/55/55134/tde-01022018-164230/publico/VictorHenriqueGoncalves_revisada.pdf. Acesso em: 10 mar. 2022.

GUJARATI, D.N.; PORTER, D. C. Econometria básica. 5. ed. São Paulo: **AMGH**, 2008. 24p.

HOLT, C.E. **Forecasting trends and seasonal by exponentially weighted averages. ONR Memorandum**, v.52, Carnegie Institute of Technology, Pittsburgh, USA,1957.

HYNDMAN, Rob J; ATHANASOPOULOS, George. Forecasting: Principles and Practice. 3. ed. Melbourne, Austrália: **OTexts**, 2021. Disponível em: https://otexts.com/fpp3/. Acesso em: 1 jan. 2022.

HYNDMAN, Rob J.; KOEHLER, A.B; SNYDER, R.D; GROSE, S. A state space framework for automatic forecasting using exponential smoothing methods. **International J. Forecasting**, [*S. l.*], p. 439-454, 1 maio 2002. Disponível em: https://robjhyndman.com/papers/hksg.pdf. Acesso em: 1 jan. 2022.

ISNAR, S.; OREMUS, M. Examining the association between the COVID-19 pandemic and self-harm death counts in four Canadian provinces. **Psychiatry Research**, v. 310, p. 114433, abr. 2022.

KOENIG, A. Previsão de demanda em séries temporais intermitentes mediante a utilização do Método de Croston. **Universidade Federal de Santa Catarina**, [*s. l.*], 2014. Disponível em: https://repositorio.ufsc.br/xmlui/handle/123456789/132948. Acesso em: 18 mar. 2022.

KIM, Agnus M. The short-term impact of the COVID-19 outbreak on suicides in Korea. **PubMed Central**, [*s. l.*], 5 dez. 2020. DOI 10.1016/j.psychres.2020.113632. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7718107/. Acesso em: 17 mar. 2022.

LEMOS, Fernando de Oliveira. Metodologia para Seleção de Métodos de Previsão de Demanda. **LUME**, [*s. l.*], 1 jan. 2006. Disponível em: http://hdl.handle.net/10183/5949. Acesso em: 7 jan. 2022.

LESKE, S. et al. Real-time suicide mortality data from police reports in Queensland, Australia, during the COVID-19 pandemic: an interrupted time-series analysis. **The Lancet Psychiatry**, 16 nov. 2020. Disponível em: https://www.thelancet.com/journals/lanpsy/article/PIIS2215-03662030435-1/fulltext. Acesso em: 05 fev. 2022.

MAKRIDAKIS, Spyros G.; WHEELWRIGHT, Steven C.; HYNDMAN, Rob J. **Methods and Applications**. 3. ed. New York: Jhon Wiley & Sons, 1998. 656 p.

MENDENHALL, W. **A Second Course in Business Statistics: Regression Analysis**. SecondSale Prentice Hall College, v1., 1993

MINAYO, Maria Cecília de Souza; MENEGHEL, Stela Nazareth; CAVALCANTE, Fátima Gonçalves. Suicídio de homens idosos no Brasil. **Ciênc. saúde coletiva**, [*s. l.*], 17 out. 2012. DOI https://doi.org/10.1590/S1413-81232012001000016. Disponível em: https://www.scielo.br/j/csc/a/nBzq9tgsv5LZDmkrjtgY5Bs/abstract/?lang=pt. Acesso em: 14 jan. 2022.

MINISTRO REGIONAL ALEMÃO PREOCUPADO COM A COVID-19 COMETE SUICÍDIO**. Estado de Minas.** 29/03/2020. Disponível em: https://www.em.com.br/app/noticia/internacional/2020/03/29/interna_internacional,1133554/ministro-regional-alemao-preocupado-com-a-covid-19-comete-suicidio.shtml. Acesso em: 15/04/2021

MINISTÉRIO DA SAÚDE (Brasil). Secretaria de Vigilância em Saúde. Mortalidade por suicídio e notificações de lesões autoprovocadas no Brasil. **Boletim Epidemiológico**, [*S. l.*], v. 55, n. 33, p. 1-10, 17 set. 2021. Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/edicoes/2021/boletim_epidemiologico_svs_33_final.pdf. Acesso em: 19 jan. 2022.

MORETTIN, Pedro A.; TOLOI, Clélia M.C. **Análise de Séries Temporais**. 2. ed. São Paulo: Egard Blucher, 2006.

NEPOMUCENO, Thyago Celso C.; MOURA , Jadielson Alves; SILVA, Lúcio Câmara; COSTA, Ana Paula Cabral Seixas. Alcohol and violent behavior among football spectators: An empirical assessment of Brazilian‘s criminalization. **International Journal of Law, Crime and Justice**, [*s. l.*], v. 51, p. 33-44, 2017. DOI https://doi.org/10.1016/j.ijlcj.2017.05.001. Disponível em: https://www.sciencedirect.com/science/article/pii/S175606161630204X. Acesso em: 9 mar. 2022.

OLIVEIRA, Milena Edite Casé *et al*. Série temporal do suicídio no Brasil: o que mudou após o Setembro Amarelo?. **Revista Eletrônica Acervo Saúde**, [*s. l.*], 14 maio 2020. DOI https://doi.org/10.25248/reas.e3191.2020. Disponível em:

https://acervomais.com.br/index.php/saude/article/view/3191. Acesso em: 11 mar. 2022.

OLIVEIRA MB. Suicídio, setembro amarelo e efeito contágio: um estudo ecológico em Santa Catarina. **Repositório Universitário da Ânima**, [s. l.], 2018. Disponível em: https://repositorio.animaeducacao.com.br/handle/ANIMA/9476. Acesso em: 2 maio 2022.

OLUKANMI, SEUN O.; NELWAMONDO, FULUFHELO V.; NWULU, D NNAMDI I. Who used the ARIMA model to model google search data to predict influenza-like illness in South Africa. **IEEE Access**, [*s. l.*], 7 set. 2021. DOI 10.1109/ACCESS.2021.3110972. Disponível em: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9530708. Acesso em: 8 mar. 2022.

OMS. OPAS. **Pandemia de COVID-19 desencadeia aumento de 25% na prevalência de ansiedade e depressão em todo o mundo**. [*S. l.*], 2 mar. 2022. Disponível em: https://www.paho.org/pt/noticias/2-3-2022-pandemia-covid-19-desencadeia-aumento-25-na-prevalencia-ansiedade-e-depressao-em#:~:text=Pandemia%20de%20COVID%2D19%20desencadeia,Organiza%C3%A7%C3%A3o%20Pan%2DAmericana%20da%20Sa%C3%BAde. Acesso em: 2 mar. 2022.

OMS. OPAS. **Após 18 meses de pandemia de COVID-19, OPAS pede prioridade para prevenção ao suicídio**. [*S. l.*], 9 set. 2021. Disponível em: https://www.paho.org/pt/noticias/9-9-2021-apos-18-meses-pandemia-covid-19-opas-pede-prioridade-para-prevencao-ao-suicidio#:~:text=Cerca%20de%2050%25%20das%20pessoas,sua%20sa%C3%BAde%20mental%20havia%20piorado. Acesso em: 13 jan. 2022.

OMS. Depression and Other Common Mental Disorders: Global Health Estimates. **World Health Organization**, [*s. l.*], 2017. Disponível em: https://apps.who.int/iris/bitstream/handle/10665/254610/WHOMSD?sequence=1. Acesso em: 15 mar. 2022.

SUICÍDIO. *In*: OMS. OPAS. **OMS**. [*S. l.*], 2018. Disponível em: https://www.paho.org/pt/topicos/suicidio. Acesso em: 12 jan. 2022.

Pauliana Valéria Machado Galvão et al. MORTALIDADE POR SUICÍDIO EM PERNAMBUCO: SÉRIE TEMPORAL DE 1991 A 2010. In: ANAIS DO CONGRESSO BRASILEIRO DE EPIDEMIOLOGIA, 2017, **. Anais eletrônicos... Campinas**, Galoá, 2017. Disponível em: https://proceedings.science/epi/trabalhos/mortalidade-por-suicidio-em-pernambuco-serie-temporal-de-1991-a-2010. Acesso em: 07 abr. 2022.

PELLEGRINI, F. R.; FOGLIATTO, F. S. Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso. **Revista Produção**., São Paulo , v. 11, n. 1, p. 43-64, June 2001

PIRKIS, Jane; JOHN, Ann; SHIN, Sangsoo; DELPOZO-BANOS, Marcos. Suicide trends in the early months of the COVID-19 pandemic: an interrupted time-series

analysis of preliminary data from 21 countries. **IntraMed**, [*s. l.*], 7 abr. 2021. DOI https://doi.org/10.1016/S2215-0366(21)00091-2. Disponível em: https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(21)00091-2/fulltext. Acesso em: 15 mar. 2022.

ROSA , Natalina Maria *et al*. Tendência de declínio da taxa de mortalidade por suicídio no Paraná, Brasil: contribuição para políticas públicas de saúde mental. **J. bras. psiquiatr.** , [*s. l.*], 3 jul. 2017. DOI https://doi.org/10.1590/0047-2085000000165. Disponível em: https://www.scielo.br/j/jbpsiq/a/snrzqtgV6csffGBZZJz83jb/?format=html. Acesso em: 11 mar. 2022

ROSTAMI, Mehran *et al*. Time Series Analysis of Monthly Suicide Rates in West of Iran, 2006-2013. **International Journal of Preventive Medicine**, New York, 1 maio 2019. DOI 10.4103/ijpvm.IJPVM_197_17. Disponível em: https://www.researchgate.net/publication/333167489_Time_Series_Analysis_of_Monthly_Suicide_Rates_in_West_of_Iran_2006-2013. Acesso em: 2 fev. 2022.

SANTOS, Erick Daniel Gomes de Melo; RODRIGUES, Gabriela Oliveira Lira; SANTOS, Lhays Melos; ALVES, Mateus Egilson da Silva; ARAÚJO, Ludgleydson Fernandes; SANTOS, José Victor de Oliveira. Suicídio entre idosos no Brasil: uma revisão de literatura dos últimos 10 anos. **Psicología**: Conocimiento y Sociedad, Universidad de la República, 2018. Disponível em: https://www.redalyc.org/journal/4758/475859262013/html/. Acesso em: 16 mar. 2022.

SECRETARIA DE PLANEJAMENTO (Brasil, BA). Superintendência de estudos econômicos e sociais da Bahia. **Textos para discussão**. [*S. l.*], 7 out. 2020. Disponível em: https://www.sei.ba.gov.br/images/publicacoes/download/textos_discussao/texto_discussao_27.pdf. Acesso em: 15 mar. 2022.

SilvaA. V. da; GonzagaG. L. P.; TorresL. A. M.; SilvaL. R. da; AciolyG. M. A.; ReisI. C. C. V.; RibeiroM. C.; BrumE. H. M. de; BarbosaK. G. N. Violência suicida entre os gêneros na região metropolitana de Maceió/Alagoas, Brasil: 20 anos de análise. **Revista Eletrônica Acervo Saúde**, v. 13, n. 9, p. e8780, 17 set. 2021.

SINAN. Boletim epidemiológico. SINAN, 2019. Disponível em: https://antigo.saude.gov.br/images/pdf/2019/setembro/13/BE-suic--dio-24-final.pdf. Acesso em : 03/11/2021

TSAI, J. F.; CHO, W. Re-examination of the seasonality of suicide in Taiwan during 1991- 2008: A population-based study. **Psychiatry Research** [online], Amsterdam, v. 186, n. 1, p. 147-149, 2011. Disponível em: Acesso em: 1 mar. 2022.

WANG, Xiaozhe; SMITH, Kate; HYNDMAN, Rob J. Characteristic-Based Clustering for Time Series Data. **Data Mining and Knowledge Discovery**, [*s. l.*], p. 335-364, 2006. DOI 10.1007/s10618-005-0039-x. Disponível em: https://www.researchgate.net/publication/220451959_Characteristic-Based_Clustering_for_Time_Series_Data. Acesso em: 9 mar. 2022.

WHO. **Suicide prevention and special programmes**. World Health Organization; 2018

Yoonhee k.; Ho Kim; Dong-silk Kim. Association between daily environmental temperature and suicide mortality in Korea (2001–2005). 2010. DOI: https://doi.org/10.1016/j.psychres.2010.08.006. https://www.sciencedirect.com/science/article/abs/pii/S0165178110005147?via%3Dihub#s0010. Acesso em: 07 mar. 2022;

APÊNDICE A – CÓDIGO FONTE

```
#Pacotes

install.packages("dplyr")

library(dplyr)

install.packages("tidyr")

library(tidyr)

install.packages("readr")

library(readr)

install.packages("ggplot2")

library(ggplot2)

install.packages("stringr")

library(stringr)

#install.packages("lubridate")

library(lubridate)

#install.packages("forecast")

library(forecast)

#install.packages("fpp3")

library(fpp3)

# Tratamento dos dados

## Importação de arquivo

df=read.csv("C:/.../Documents/Portfolio/TCC/Dados/SucidioPorRegiao.csv",
sep = ";")

## Função geral de tratamento

trataDados <- function(df,x){

  #Reunindo os estados em uma coluna

  df=gather(df,UF,Suicidios,2:x)
```

```
  names(df)[1]= "Ano.Mes"
  #Retirando textos que estão no meio da tabela
  df=df[!is.na(df$Suicidios),]
  df=df[!is.null(df$Suicidios)]
  df=df[!(df$Suicidios==""),]
  #Retirando valores de totais anuais
  df=filter(df,!df$Ano.Mes %in%
c(1996,1997,1998,1999,2000,2001,2002,2003,2004,2005,
2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020)
)
  #Retirando a linha que contem o total
  df=filter(df, !df$Ano.Mes %in% c("Total"))
  #Separando a coluna Ano.Mês.do.Obito em ano e mês
  chave=paste(df$Ano.Mes,df$UF,sep="_")
  df=separate(df,Ano.Mes,into=c("Mes","Ano"), sep="/")
  #Colocando uma coluna de chave para fazer a união das futuras tabelas
  df$chave=chave
  #Retira os dois pontos que ficou na frente de texto
  df$Mes=gsub("\\.\\.","",df$Mes)
  #Transforma dados em inteiros, isso vai gerar NAS quando o valor for "-"
  df$Suicidios=as.integer(df$Suicidios)
  #Substituindo os valores NAs gerados por "0"
  df[is.na(df$Suicidios),"Suicidios"]=0
  return(df)
}
df= trataDados(df,29)
df=filter(df, !df$UF %in% c("Total"))
setwd("~/Portfolio/TCC/Dados")
### Tratando os os dados relacionado aos obitos por sexo
```

```
SuicidioFem=
read.csv("C:/.../Portfolio/TCC/Dados/SuicidioPorSexoFem.csv",sep=";")

SuicidioMasc=
read.csv("C:/.../Portfolio/TCC/Dados/SuicidioPorSexoMasc.csv",sep=";")

SuicidioIgn=
read.csv("C:/.../Portfolio/TCC/Dados/SuicidioPorSexoIgn.csv",sep=";")


SuicidioFem=trataDados(SuicidioFem,29)

SuicidioMasc=trataDados(SuicidioMasc,29)

SuicidioIgn= trataDados(SuicidioIgn,16)

SuicidioFem$Sexo="Feminino"

SuicidioMasc$Sexo="Masculino"

SuicidioIgn$Sexo="Ign"

sexo=rbind(SuicidioFem,SuicidioMasc)

sexo=rbind(sexo,SuicidioIgn)

### Tratando dados Relacionados a Cor

SuicidioPorCor<-
read.csv("C:/.../Documents/Portfolio/TCC/Dados/SuicidioPorCor.csv", header
= T, sep=";")

SuicidioPorCor=trataDados(SuicidioPorCor,8)

names(SuicidioPorCor)[3]="Cor"

SuicidioPorCor=filter(SuicidioPorCor, !SuicidioPorCor$Cor %in% c("Total"))

SuicidioPorCor$chave=NULL

### Tratando dados Relacionados a Cor padrão IBGE

PretoPardo=group_by(filter(SuicidioPorCor, SuicidioPorCor$Cor %in%
c("Preta","Parda")),Mes,Ano)%>%summarise(Suicidios=sum(Suicidios))

PretoPardo$Cor="Pretos e Pardos"

SuicidioPorCorIBGE=rbind(filter(SuicidioPorCor, !SuicidioPorCor$Cor %in%
c("Preta","Parda")), PretoPardo)

### Tratando dados Relacionados a suicidio por Local
```

```
SuicidioPorLocal<-
read.csv("C:/.../Portfolio/TCC/Dados/SuicidioPorLocal.csv", header = T,
sep=";")

SuicidioPorLocal=trataDados(SuicidioPorLocal,8)

names(SuicidioPorLocal)[3]=c("Local")

SuicidioPorLocal=filter(SuicidioPorLocal,!SuicidioPorLocal$Local %in%
c("Total"))

### Tratamento Dados Relacionados a suicidio por Faixa Etaria

SuicidioPorFaixaEtaria<-
read.csv("C:/.../Documents/Portfolio/TCC/Dados/SuicidioPorFaixaEtaria.csv",
header = T, sep=";")

SuicidioPorFaixaEtaria=trataDados(SuicidioPorFaixaEtaria,15)

names(SuicidioPorFaixaEtaria)[3]="FaixaEtaria"

SuicidioPorFaixaEtaria=filter(SuicidioPorFaixaEtaria,!SuicidioPorFaixaEtari
a$FaixaEtaria %in% c("Total"))

### Tratamento Dados Relacionados a suicidio por Estado Civil

SuicidioPorEstadoCivil<-
read.csv("C:/.../Documents/Portfolio/TCC/Dados/SuicidioPorEstadoCivil.csv",
header = T, sep=";")

SuicidioPorEstadoCivil=trataDados(SuicidioPorEstadoCivil,29)

SuicidioPorEstadoCivil$Ano=NULL

names(SuicidioPorEstadoCivil)[1]="EstadoCivil"

### Tratamento de Dados Relacionado a suicidio por Tipo de Causa

SuicidioPorCausa<-
read.csv("C:/.../Documents/Portfolio/TCC/Dados/SuicidioPorCausa.csv",
header = T, sep=";")

SuicidioPorCausa=trataDados(SuicidioPorCausa,5)

SuicidioPorCausa$Ano=NULL

names(SuicidioPorCausa)[1]="Causa"

names(SuicidioPorCausa)[2]="Sexo"
```

```
SuicidioPorCausa=filter(SuicidioPorCausa,!SuicidioPorCausa$Sexo %in%
c("Total"))

### Tratamento de dados relacionado a suicidio por cor e idade

SuicidioCorIdade<-
read.csv("C:/.../Documents/Portfolio/TCC/Dados/SuicidioCorIdade.csv",
header = T, sep=";")

SuicidioCorIdade= trataDados(SuicidioCorIdade,15)

SuicidioCorIdade$Ano=NULL

names(SuicidioCorIdade)[1]="Cor"

names(SuicidioCorIdade)[2]="Idade"

SuicidioCorIdade=filter(SuicidioCorIdade,!SuicidioCorIdade$Idade %in%
c("Total"))

SuicidioCorIdade$Cor=as.factor(SuicidioCorIdade$Cor)

SuicidioCorIdade$Idade=as.factor(SuicidioCorIdade$Idade)

SuicidioCorIdade$chave=NULL

### Tratamento PopEstimada

PopEstimadaEstado<- read_xlsx("C:/.../Documents/Portfolio/Trabalho-
SeriesTemporais/Dados/EstiPopEstAno.xlsx", sheet = 4)

SuicidioPorEstado= df %>% group_by(Ano, UF) %>% summarise(Suicidios =
sum(Suicidios))

PopEstimadaEstado$Ano = as.character(PopEstimadaEstado$Ano)

SuicidioPorEstado = left_join(SuicidioPorEstado, PopEstimadaEstado , by =
c('Ano', 'UF'))

SuicidioPorEstado$taxa_suicidio =
(SuicidioPorEstado$Suicidios/SuicidioPorEstado$Pop) * 100000

### Tratamento PopEstimada Por faixa etaria

PopEstFaixaEta<- read_xlsx("C:/.../Documents/Portfolio/Trabalho-
SeriesTemporais/Dados/FaixaEtariaPop2010a2020.xlsx", sheet = 4)

PopEstFaixaEtaAno = PopEstFaixaEta %>% group_by(Ano, FaixaEtaria) %>%
summarise(Pop = sum(Pop))
```

```
SuicidioPorFaixaEtaria$FaixaEtaria= gsub('X',
'',SuicidioPorFaixaEtaria$FaixaEtaria ) %>% str_replace_all('\\.', ' ');

SuicidioPorFaixaEtaria= SuicidioPorFaixaEtaria %>% group_by(Ano,
FaixaEtaria) %>% summarise(Suicidios = sum(Suicidios))

SuicidioPorFaixaEtaria= SuicidioPorFaixaEtaria %>% filter(FaixaEtaria !=
'Idade ignorada')

SuicidioPorFaixaEtaria = left_join(PopEstFaixaEtaAno,
SuicidioPorFaixaEtaria, by =c('Ano', 'FaixaEtaria'))

SuicidioPorFaixaEtaria$taxa_suicidio  =
(SuicidioPorFaixaEtaria$Suicidios/SuicidioPorFaixaEtaria$Pop) *100000

### Tratamento PopEstimada Por SExo

sexo_taxa = sexo %>% filter(Sexo != "Ign")

PopPorSexo = read_xlsx("C:/.../Documents/Portfolio/Trabalho-
SeriesTemporais/Dados/FaixaEtariaPop2010a2020.xlsx", sheet = 4);

PopPorSexo$Sexo= gsub("Mulher", "Feminino", PopPorSexo$Sexo)

PopPorSexo$Sexo= gsub("HOMEM", "Masculino",PopPorSexo$Sexo)

PopPorSexo= PopPorSexo %>% group_by(Ano,Sexo) %>% summarise(Pop = sum(Pop))

sexo_taxa = sexo_taxa %>% group_by(Ano, Sexo) %>% summarise(Suicidios =
sum(Suicidios)) %>% left_join(PopPorSexo, by=c("Ano", "Sexo")) %>%
filter(!is.na(Pop))

sexo_taxa$taxa_suicidio = (sexo_taxa$Suicidios/sexo_taxa$Pop) * 100000

sexo_taxa

taxa_suicidios_Ano=sexo_taxa %>% group_by(Ano) %>% summarise(taxa_suicidio=
sum(taxa_suicidio))

### Dados gerais normalizados pela população estimada

PopAnoMes = read_xlsx("C:/.../Documents/Portfolio/Trabalho-
SeriesTemporais/Dados/EstiPopAnoMes.xlsx", sheet = 1);

PopAnoMes$Ano = as.character(PopAnoMes$Ano)

df_Normalizado = df %>% group_by(Ano,Mes) %>% summarise(Suicidios=
sum(Suicidios)) %>%
```

```
  left_join(PopAnoMes, by=c("Ano","Mes")) %>% filter(!is.na(Pop)) %>%
mutate(Taxa_Suicidio = (Suicidios/Pop)*100000)

## Análise Descritiva Geral

### Histograma e boxPlot de taxa de suicidios

hist(df_Normalizado$Taxa_Suicidio*10, main = "Histograma taxa de suicídios
entre 2010 até 2020", xlab = "Suicidio/1m. Hab.")

summary(df_Normalizado$Taxa_Suicidio)

shapiro.test(df_Normalizado$Taxa_Suicidio)

ggplot(df_Normalizado,aes_string(x="Ano", y="Taxa_Suicidio",group="Ano"))+

  geom_boxplot() + theme_classic() + ggtitle("Variação de suicidios por
Ano") + ylab("Suicidios/100mil Hab.")

## Análise descritiva por categoria

### Suicidio Por estado

x=aggregate(SuicidioPorEstado$taxa_suicidio,
list(local=SuicidioPorEstado$UF), sum)

x=x[order(x[,2],decreasing = T),]

x$indice=c(1:27)

ggplot(x, aes(x=reorder(local,indice), y=x)) + geom_col() + theme_classic()
+ labs(title ="Taxa de Suicídios por Estado") + ylab("taxa de Suicídios por
100 mil Hab.") + xlab("")

### Suicidios Por Causas no geral

x=aggregate(SuicidioPorCausa$Suicidios, list(local=SuicidioPorCausa$Causa),
sum)

x=x[order(x[,2],decreasing = T),]

x$indice=c(1:25)

x=x[x$indice<11,]

ggplot(x, aes(x=reorder(local,-indice), y=x)) + geom_col() + coord_flip() +
theme_classic() + labs(title ="Principais formas de Suicídio no Brasil") +
ylab("Número de Suicídios") + xlab("")

### Suicídios Por causa / sexo masculino
```

```
SuicidioPorCausaMasc=subset(SuicidioPorCausa,SuicidioPorCausa$Sexo=="Masc")

x=aggregate(SuicidioPorCausaMasc$Suicidios,
list(local=SuicidioPorCausaMasc$Causa), sum)

x=x[order(x[,2],decreasing = T),]

x$indice=c(1:25)

x=x[x$indice<11,]

ggplot(x, aes(x=reorder(local,-indice), y=x)) + geom_col() + coord_flip() +
theme_classic() + labs(title ="Homens cometerem Suicídio no Brasil") +
ylab("Número de Suicídios") + xlab("")

### Suicídios Por causa / sexo feminino

SuicidioPorCausaFem=subset(SuicidioPorCausa,SuicidioPorCausa$Sexo=="Fem")

x=aggregate(SuicidioPorCausaFem$Suicidios,
list(local=SuicidioPorCausaFem$Causa), sum)

x=x[order(x[,2],decreasing = T),]

x$indice=c(1:25)

x=x[x$indice<11,]

ggplot(x, aes(x=reorder(local,-indice), y=x)) + geom_col() + coord_flip() +
theme_classic() + labs(title ="Mulheres cometerem Suicídio no Brasil") +
ylab("Número de Suicídios") + xlab("")

### SUicidio por cor

x= aggregate(SuicidioPorCorIBGE$Suicidios,
list(local=SuicidioPorCorIBGE$Cor), sum)

x=x[order(x[,2], decreasing = T),]

x$indice= c(1:5)

ggplot(x, aes(x=reorder(local,indice), y=x)) + geom_col() + theme_classic()
+ labs(title ="N° de Suicidios por Cor (Padrão IBGE)") + ylab("Número de
Suicídios") + xlab("")


### Suicidios Por Cor e Ano

SuicidioPorCorIBGE$Ano= as.factor(SuicidioPorCorIBGE$Ano)
```

```
SuicidioPorCorIBGE$Cor=as.factor(SuicidioPorCorIBGE$Cor)

ggplot(SuicidioPorCorIBGE, aes(x=Cor, y=Suicidios, fill=Cor)) + geom_col()
+ labs(title ="N° de Suicidios por Cor e por Ano(Padrão IBGE)") +
ylab("Número de Suicídios") + xlab("") + theme_classic()+
theme(legend.direction = "vertical", axis.text.x = element_text(angle=-90))
+ facet_wrap(SuicidioPorCorIBGE$Ano)

### Suicídos por Faixa etaria e ano

SuicidioCorIdadeIBGE=filter(SuicidioCorIdadeIBGE,!SuicidioCorIdadeIBGE$Idad
e %in% c("X1.a.4.anos", "Menor.1.ano"))

ggplot(SuicidioCorIdadeIBGE, aes(x=Cor, y=Suicidios, fill=Cor))+ geom_col()
+ labs(title ="N° de Suicidios por Cor e por Idade(Padrão IBGE)") +
ylab("Número de Suicídios") + xlab("") +
facet_wrap(SuicidioCorIdadeIBGE$Idade) + theme_classic()+
theme(legend.direction = "vertical", axis.text.x = element_text(angle=-90))

### Suicidio por Sexo

sexo_taxa$Ano= as.factor(sexo_taxa$Ano)

ggplot(sexo_taxa, aes(x=Sexo, y=taxa_suicidio, fill=Sexo))+ geom_col() +
labs(title ="Taxa de suicidio por Sexo") + ylab("Taxa de Suicídios para
cada 100 mil Hab.") + xlab("") + theme_classic() +scale_y_continuous(labels
= function(x) format(x, scientific = FALSE))

sexo_taxa$Ano= as.factor(sexo_taxa$Ano)

ggplot(sexo_taxa, aes(x=Sexo, y=taxa_suicidio, fill=Sexo))+ geom_col() +
labs(title ="Taxa de Suicidios por Sexo por Ano") + ylab("Suicídios 100 mil
hab.") + xlab("") + theme_classic() +scale_y_continuous(labels =
function(x) format(x, scientific = FALSE)) + facet_wrap(sexo_taxa$Ano) +
coord_flip()+theme(legend.direction = "vertical", axis.text.x =
element_text(angle=-90))

### Graficos relacionado ao estado civil

x= aggregate(SuicidioPorEstadoCivil$Suicidios,
list(local=SuicidioPorEstadoCivil$EstadoCivil), sum)

x=x[order(x[,2], decreasing = T),]

x$indice= c(1:6)
```

```
ggplot(x, aes(x=reorder(local,indice), y=x)) + geom_col() + theme_classic()
+ labs(title ="Suícidio por Estado Civil") + ylab("Número de Suicídios") +
xlab("")

# Análise de séries temporais

###Criando Serie

dados=group_by(df,Ano,Mes)%>%summarise(Suicidios=sum(Suicidios))

dados$Mes=yearmonth(ymd(paste(dados$Ano,dados$Mes,"1",sep="/")))

serie=dados[,c("Mes","Suicidios")] %>%
  as_tsibble(index = Mes)

### Grafico de tempo

autoplot(serie,Suicidios) +
  labs(title="Numero de suicídios no Brasil",
       subtitle = "Periodo: 1996-2020") + theme_bw()

### Parcelas Sazonais

serie %>%
  gg_season(Suicidios, labels = "both") +
  labs(title = "Gráfico Sazonal: Suícidos no Brasil",
       y="Suicidios")+
  theme_bw()

### Gráfico de Subséries Temporais

ts(dados$Suicidios, start = 1996, frequency = 12) %>%
  ggsubseriesplot(ts) +
  labs(title= "Média de Suicídios por Mês",
       y="Suicídios") +theme_bw()

### Transformaçao e ajuste

serie %>%
  autoplot(Suicidios) +
  labs(title="Serie ajustada",
```

```
       y="Suicídios/população") + theme_bw()
serie %>%
model(STL(Suicidios~season(window="periodic")+trend(window=24),robust=TRUE)
) %>%
  components() %>%
  autoplot() + theme_bw()


### Força sazonal e de Tendência
serie %>%
  features(Suicidios, feat_stl) %>%
  select(c("trend_strength", "seasonal_strength_year"))
## Estimação modelo de Suavização Exponencial
### Divisão em dados de treino e teste
dados_treino = serie %>% filter(year(serie$Mes)<=2017)
dados_teste = serie %>% filter(year(serie$Mes)>2017) %>%
filter(year(Mes)<2020)
dados_2020 = serie %>% filter(year(serie$Mes)==2020)
### Escolha do Modelo e previsão
fit<-dados_treino %>%
  model(Previsao=ETS(Suicidios))
report(fit);
previsao= fit %>% forecast(h=24)
previsao
### Analise de residuos
fit %>%
  gg_tsresiduals()
residuos=treino_fit %>% residuals()
Box.test(residuos$.resid, lag=10, type="Ljung-Box")
```

```
### Visualizando Previsões

###Visualizando os resultados

autoplot(dados_treino, .vars = Suicidios, color="black") +

  autolayer(dados_teste, .vars=Suicidios, color="black") + theme_bw()+

  autolayer(previsao, level=NULL, color="red") +

  labs(title = "Previsão da Serie histórica para os próximos 2 Anos",
y="Suicidios", subtitle = "2018 Jan - 2019 Dez") + guides(colour=
guide_legend(title = "Forecast", label = TRUE))

previsao

prev= cbind(previsao$.mean, dados_teste)

names(prev)<- c("Previsto","Periodo","Real")

ggplot(prev, aes(x=Periodo,y=Real, color="black")) + geom_line() +
geom_line(aes(x=Periodo, y=Previsto, color="red")) + theme_bw() +
labs(title = "Valores reais e valores previstos ",y="Suicidios", subtitle =
"Periodo:2018 Jan - 2019 Dez") +
guides(colour=guide_legend(title="Series")) +

  scale_color_manual(labels = c("Real", "Previsto"), values = c("black",
"red"))

autoplot(previsao) +

  autolayer(dados_teste, .vars = Suicidios) + theme_classic() + labs(title
= "Valores reais e valores previstos - Jan 2018 a Dez 2019 ", subtitle =
"Legenda: Previsto (Azul) - Real (Preto)")

## Modelo ARIMA

### Teste de normalidade

par(mfrow=c(1,2))

hist(serie$Suicidios, main="Histograma normal")

hist(log(serie$Suicidios), main="Histograma dados transformados")

tseries::jarque.bera.test(serie$Suicidios)

tseries::jarque.bera.test(log(serie$Suicidios))

treino_arima=dados_treino %>%
```

```
  mutate(log_suicidios=log(Suicidios)) %>%
  mutate(dif_suicidios= difference(log(Suicidios)))
### Estacionariedade da Serie
dados_treino %>%
  features(Suicidios, unitroot_kpss)
dados_treino %>%
  features(Suicidios, unitroot_ndiffs)
###ACF E ACF
par(mfrow=c(2,1))
acf(diff(treino_arima$Suicidios), main="Funções de autocorrelação")
pacf(diff(treino_arima$Suicidios), main="")
### Estimação
ARIMA111=arima(treino_arima$Suicidios, c(1,1,1))
ARIMA411=arima(treino_arima$Suicidios,c(4,1,1))
ARIMA1111=arima(treino_arima$Suicidios,c(11,1,1))
ARIMA411_101_12=arima(treino_arima$Suicidios,c(4,1,1),c(1,0,1,12))
ARIMA411_102_6=arima(treino_arima$Suicidios,c(4,1,1),c(1,0,2,6))
ARIMA_AUTO_ARIMA=arima(treino_arima$Suicidios,c(1,1,3),c(0,0,2,12))
AIC(ARIMA111,ARIMA411,ARIMA1111,ARIMA411_101_12,ARIMA411_102_6,ARIMA_AUTO_A
RIMA)
### Diagnosticos
checkresiduals(ARIMA411_101_12)
qqnorm(residuals(ARIMA411_101_12))
qqline(residuals(ARIMA411_101_12))
### prevendo valores de teste
dados_treino = serie %>% filter(year(serie$Mes)<=2017)
dados_teste = serie %>% filter(year(serie$Mes)>2017) %>%
filter(year(Mes)<2020)
```

```
dados_2020 = serie %>% filter(year(serie$Mes)==2020)

fit_arima = dados_treino %>%

  model(ARIMA(Suicidios ~ 0 + pdq(4,1,1) + PDQ(1,1,1,12)))

prev_fit_arima = fit_arima %>% forecast(h=24)

autoplot(prev_fit_arima) + autolayer(dados_teste) + theme_bw() + labs(title
= "Valores reais e valores previstos - Jan 2018 a Dez 2019 ", subtitle =
"Legenda: Previsto (Azul) - Real (Preto)")

### Selecionando o melhor modelo

rbind(accuracy(fit_arima), accuracy(fit))

## Análise dos impactos do COVID-19 em 2020

###Retirando valor de 2020 da serie

serie = filter(serie, Mes< yearmonth('Jan 2020'))

serie

autoplot(serie,Suicidios) +

  labs(title="Número de suicídios no Brasil",

       subtitle = "Periodo: 1996-2020") +

  autolayer(dados_2020, col="red") + theme_bw()

###Aplicando modelo de suavização Exp


treino = serie

treino_fit <- treino %>%

  model(Previsao=ETS(Suicidios))

previsao = treino_fit %>%

  forecast(h=12)

### Comparando a previsão com os valores reais

autoplot(previsao,dados_2020) + theme_classic() + labs(title="Comparação
entre projeção e valor real ",subtitle = "Azul - Projeção   Vermelho -
Valor Real") + autolayer(dados_2020, col="red") + theme_bw()

### Teste de hipotese
```

```
### Junção dos dados de previsao e reais

comparacao=as.data.frame(previsao$.mean)

comparacao$Real = dados_2020$Suicidios

comparacao=`colnames<-`(comparacao,c("Previsto","Real"))

comparacao=pivot_longer(comparacao, cols = Previsto:Real, names_to =
"Tipo",values_to = "Suicidios")

comparacao$Tipo=as.factor(comparacao$Tipo)

ggplot(data=comparacao, aes_string(x=comparacao$Tipo,
y=comparacao$Suicidios, group=comparacao$Tipo)) + geom_boxplot() +
theme_classic() + labs( x="",y="Número de Suicídios", title = "BoxPlot -
Número de Suicídios real e previsto")

###Teste de hipotese

####Calculando a diferença entre as médias

media_previsto=mean(filter(comparacao,comparacao$Tipo=="Previsto")$Suicidio
s)

media_real=mean(filter(comparacao,comparacao$Tipo=="Real")$Suicidios)

diff_media= media_previsto - media_real

diff_media

####Criando funcao de permutação aleatoria

set.seed(123)

#Criando funçãod e permutação considerado hipotese h0

perm_fun = function(x,n1,n2){

  n= n1 + n2

  idx_b= sample(1:n, n1)

  idx_a= setdiff(1:n, idx_b)

  mean_diff= mean(x[idx_b])- mean(x[idx_a])

  return (mean_diff)

}

perm_diffs= rep(0,1000)
```

```
for(i in 1:1000)
  perm_diffs[i]= perm_fun(comparacao$Suicidios, 12,12)
hist(perm_diffs, xlim=c(-100,100),nclass = 80)
abline(v= (diff_media))
####Calculando o valor p
p=mean(perm_diffs>=diff_media)
p
## Analise de Suicidios Setembro Amarelo
### Tratamento dos dados
dados=group_by(df,Ano,Mes)%>%summarise(Suicidios=sum(Suicidios))
dados$Mes=yearmonth(ymd(paste(dados$Ano,dados$Mes,"1",sep="/")))
serie=dados[,c("Mes","Suicidios")] %>%
  as_tsibble(index = Mes)
dados_treino = serie %>% filter(year(serie$Mes)<=2014)
dados_teste = serie %>% filter(year(serie$Mes)>=2015)
### Criação do modelo
fit<-dados_treino %>%
  model(Previsao=ETS(Suicidios ~ error("M") + trend("A") + season("A")))
report(fit);
previsao= fit %>% forecast(h=72)
previsao
### Analise de residuos
fit %>%
  gg_tsresiduals()
residuos=treino_fit %>% residuals()
Box.test(residuos$.resid, lag=10, type="Ljung-Box")
###Visualizando os resultados
autoplot(dados_treino, .vars = Suicidios, color="black") +
```

```
  autolayer(dados_teste, .vars=Suicidios, color="black") + theme_bw()+

  autolayer(previsao, level=NULL, color="red") +  labs(title = "Previsão de
2015 a 2020", y="Suicidios", subtitle = "2015 Jan - 2020 Dez") +
guides(colour= guide_legend(title = "Forecast", label = TRUE))

previsao

prev= cbind(previsao$.mean, dados_teste)

names(prev)<- c("Previsto","Periodo","Real")

ggplot(prev, aes(x=Periodo,y=Real, color="black")) + geom_line() +
geom_line(aes(x=Periodo, y=Previsto, color="red")) + theme_bw() +
labs(title = "Valores reais e valores previstos ",y="Suicidios", subtitle =
"Periodo:2018 Jan - 2019 Dez") +
guides(colour=guide_legend(title="Series")) + scale_color_manual(labels =
c("Real", "Previsto"), values = c("black", "red"))

autoplot(previsao) + autolayer(dados_teste, .vars = Suicidios) +
theme_classic() + labs(title = "Valores reais e valores previstos - Jan
2015 a Dez 2020 ", subtitle = "Legenda: Previsto (Azul) - Real (Preto)")

### Teste de hipotese

comparacao=as.data.frame(previsao$.mean)

comparacao$Real = dados_teste$Suicidios

comparacao=`colnames<-`(comparacao,c("Previsto","Real"))

comparacao=pivot_longer(comparacao, cols = Previsto:Real, names_to =
"Tipo",values_to = "Suicidios")

comparacao$Tipo=as.factor(comparacao$Tipo)

ggplot(data=comparacao, aes_string(x=comparacao$Tipo,
y=comparacao$Suicidios, group=comparacao$Tipo)) + geom_boxplot() +
theme_classic() + labs( x="",y="Número de Suicídios", title = "BoxPlot -
Comparação entre números de óbitos reais e previstos")

####Calculando a diferença entre as médias

media_previsto=mean(filter(comparacao,comparacao$Tipo=="Previsto")$Suicidio
s)

media_real=mean(filter(comparacao,comparacao$Tipo=="Real")$Suicidios)

diff_media= media_previsto - media_real
```

```
diff_media

#####Criando funcao de permutação aleatoria

set.seed(123)

#Criando funçãod e permutação considerado hipotese h0

perm_fun = function(x,n1,n2){

  n= n1 + n2

  idx_b= sample(1:n, n1)

  idx_a= setdiff(1:n, idx_b)

  mean_diff= mean(x[idx_b])- mean(x[idx_a])

  return (mean_diff)

}

perm_diffs= rep(0,1000)

for(i in 1:1000)

  perm_diffs[i]= perm_fun(comparacao$Suicidios, 72,72)

hist(perm_diffs, xlim=c(-400,400),nclass = 80)

abline(v= (diff_media))

###Calculando o valor p

p=mean(perm_diffs<=diff_media)

p

#### Visualização de apenas Setembro

comparacao$ano =
c(2015,2015,2016,2016,2017,2017,2018,2018,2019,2019,2020,2020)

comparacao$Suicidios = round(comparacao$Suicidios,0)

ggplot(comparacao, aes(x=ano, y=Suicidios, fill=factor(Tipo))) +
geom_col(position ='dodge') + theme_classic() +labs(y = "Suicídios", x =
"Ano", title = "Suicídios Reais vs Previstos em Setembro") +
scale_fill_discrete(name="", labels=c("Previsto", "Real"))
```

Fonte: Autoria Própria