

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

**Um estudo sobre os impactos de algumas medidas
sanitárias na incidência e mortalidade por COVID-19**

Rhayani Aparecida Paiuta

Trabalho de Conclusão de Curso

Rhayani Aparecida Paiuta

Um estudo sobre os impactos de algumas medidas sanitárias na
incidência e mortalidade por COVID-19

Este exemplar corresponde à redação final do trabalho de conclusão de curso devidamente corrigido e defendido por Rhayani Aparecida Paiuta e aprovado pela banca examinadora.

Aprovado em 07 de abril de 2022

Banca Examinadora:

- Maria Sílvia de Assis Moura
- Michel Helcias Montoril
- Teresa Cristina Martins Dias

FEDERAL UNIVERSITY OF SÃO CARLOS
EXACT AND TECHNOLOGY SCIENCES CENTER
DEPARTMENT OF STATISTICS

A study on the impacts of some health measures on the incidence
and mortality of COVID-19

Rhayani Aparecida Paiuta

Advisor: Maria Sílvia de Assis Moura

Bachelors dissertation submitted to the Department of Statistics, Federal University of São Carlos - DEs-UFSCar, in partial fulfillment of the requirements for the degree of Bachelor in Statistics.

São Carlos

May 2022

Agradecimentos

Agradeço primeiramente a Deus, pois sem ele não seria possível chegar até aqui.

Agradeço aos meus pais Aleandra e Raimundo que sempre me apoiaram e deram suporte para que eu conseguisse alcançar todas as minhas conquistas e realizações profissionais.

Agradeço ao meu namorado Gabriel e a todos meus amigos por todo apoio, amizade e por caminharem ao meu lado durante todo esse período da faculdade.

Agradeço minha orientadora Maria Sílvia por todo suporte para construção desse trabalho, além dos conselhos e incentivos durante toda a graduação.

Por fim, agradeço a Universidade Federal de São Carlos, em especial a todo o Departamento de Estatística e seu corpo docente.

Resumo

Desde março de 2020 o Brasil, assim como o mundo todo, vem enfrentando uma grande pandemia causada pela COVID-19 (SARS-CoV-2). A COVID-19 é uma infecção respiratória com elevada transmissibilidade e por esse motivo governantes do mundo todo tomaram medidas sanitárias para controlar os casos da doença e assim frear a pandemia. O *lockdown* foi uma medida sanitária utilizada em algumas cidades para evitar a circulação de pessoas nas ruas e assim reduzir a contaminação de pessoas com a doença. Além disso, depois de quase um ano do início da pandemia, alguns países começaram a aplicar vacinas também a fim de diminuir os impactos da pandemia. Diante da incerteza sobre o que de fato seria eficaz para o controlar o números de casos e de mortes por COVID-19, pesquisadores e pessoas do senso comum começaram a ter interesse sobre essas medidas sanitárias e começaram se questionar se o *lockdown* ou se o início da vacinação ajudou no controle da pandemia. Nesse contexto, este trabalho estudou os efeitos do *lockdown* e da vacinação no número de casos novos e no número de mortes por COVID-19 através de técnicas de séries temporais. Após a aplicação da metodologia, observamos que o *lockdown* ajudou na redução de casos de COVID-19 na cidade de Araraquara e que a vacinação impactou na redução de casos e de mortes pela doença.

Palavras-chave: *Análise de Intervenção, ARIMA, COVID-19, Lockdown, Suavização Exponencial, Vacinação.*

Abstract

Since March 2020, Brazil, like the whole world, has been facing a major pandemic caused by COVID-19 (SARS-CoV-2). COVID-19 is a highly transmissible respiratory infection and for this reason governments around the world have taken health measures to control cases of the disease and thus stop the pandemic. The *lockdown* was a sanitary measure used in some cities to prevent the movement of people on the streets and thus reduce the contamination of people with the disease. In addition, after almost a year since the beginning of the pandemic, some countries have started to apply vaccines as well in order to reduce the impacts of the pandemic. Faced with the uncertainty about what would actually be effective to control the number of cases and deaths from COVID-19, researchers and people of common sense began to be interested in these sanitary measures and began to question whether the *lockdown* or if the start of vaccination helped to control the pandemic. In this context, this work studied the effects of *lockdown* and vaccination on the number of new cases and the number of deaths by COVID-19 through time series techniques. After applying the methodology, we observed that the *lockdown* helped to reduce cases of COVID-19 in the city of Araraquara and that vaccination impacted the reduction of cases and deaths.

Keywords: *ARIMA, COVID-19, Exponential smoothing, Intervention analysis, Lockdown, Vaccination.*

Lista de Figuras

3.1	Incidência de COVID-19 em Araraquara e São Carlos.	32
3.2	Incidência de COVID-19 em Araraquara, São Carlos, Campinas e São Paulo.	35
3.3	Mortalidade de COVID-19 em Araraquara, São Carlos, Campinas e São Paulo.	38

Lista de Tabelas

3.1	Ajuste ARIMA e suavização exponencial considerando o efeito do <i>Lockdown</i> .	34
3.2	Estimativas da variável de intervenção do modelo ARIMA.	34
3.3	Ajuste ARIMA e suavização exponencial considerando o efeito da vacinação na série de incidência de COVID-19.	37
3.4	Estimativas da variável de intervenção do modelo ARIMA com relação a série de incidência.	38
3.5	Ajuste ARIMA e suavização exponencial considerando o efeito da vacinação na série de mortalidade de COVID-19.	39
3.6	Estimativas da variável de intervenção do modelo ARIMA com relação a série de mortalidade.	40

Sumário

1	Introdução	17
2	Análise de intervenção com base em modelos ARIMA e de Suavização Exponencial	19
2.1	Modelos ARIMA	20
2.2	Métodos de alisamento exponencial	23
2.2.1	Alisamento exponencial simples	23
2.2.2	Alisamento exponencial com tendência	24
2.3	Análise de Intervenção	25
2.4	Modelo de espaço de estados	26
2.5	Medidas de qualidade dos modelos	29
2.5.1	Erro Quadrático Médio (EQM) e Critério de Informação de Akaike (AIC)	29
3	Resultados	31
3.1	Incidência da COVID-19 e efeitos do <i>lockdown</i> e início da vacinação.	31
4	Conclusão	41
	Referências Bibliográficas	43
A	Símbolos matemáticos	45
B	Códigos utilizados	47

Capítulo 1

Introdução

Nos últimos dois anos o Brasil e o mundo todo vem enfrentando uma pandemia causada pelo novo coronavírus SARS-CoV-2, em que a doença causada pelo vírus ficou conhecida como COVID-19. A COVID-19 é uma infecção respiratória aguda potencialmente grave, de elevada transmissibilidade e de distribuição global (da Saúde, 2020). Com a necessidade de entender a real situação do país, o governo criou um site¹ onde é possível acompanhar o número de casos novos e o número de mortes por COVID-19 desde 21 de março de 2020 até os dias atuais.

Diante o aumento de casos e de mortes por COVID-19, durante toda a pandemia o mundo todo tomou medidas sanitárias a fim de controlar a transmissão da doença. A cidade Arararaquara, que fica no interior do estado de São Paulo, foi uma das poucas cidades do Brasil a decretar *lockdown* como tentativa de amenizar a pandemia (de Arararaquara, 2021). Dessa forma, surgiram diversas questões como, por exemplo, se de fato o *lockdown* ajudou a controlar o número de casos de COVID-19.

Além disso, após quase um ano de pandemia, o Brasil iniciou a vacinação de pessoas seguindo diversas restrições que priorizavam médicos, enfermeiros, idosos e pessoas com comorbidades (da Saúde, 2020). Nesse momento, diante de um cenário crítico da pandemia, em que muitas pessoas já haviam perdido a vida pela doença, muitos pesquisadores e até mesmo pessoas do senso comum começaram a ter interesse sobre o efeito da vacina no número de casos e também na mortalidade.

Nesse contexto, nesse trabalho vamos estudar os impactos causados pelo *lockdown* e pelo início da vacinação no número de casos e no número de mortes por COVID-19. Para essa análise vamos utilizar técnicas de séries temporais que em resumo é uma área

¹<https://covid.saude.gov.br/>

direcionada ao estudo de observações coletadas ao longo do tempo, já que dados desse tipo naturalmente apresentam uma dependência entre si ([Morettin e Tolo, 2018](#)).

É importante ressaltar que podemos utilizar séries temporais para modelar diversos tipos de acontecimentos como, por exemplo, a taxa de desemprego, o preço de ativos negociados na bolsa de valores, fenômenos naturais e também, como nesse caso, modelar os números de casos e mortes por uma determinada doença. Além disso, nas últimas décadas, a difusão da internet e o avanço computacional colaboraram para o desenvolvimento de técnicas na área de séries temporais como, por exemplo, os modelos ARIMA propostos por [Box et al. \(2015\)](#) e os modelos de Suavização Exponencial [Gardner Jr \(1985\)](#).

No próximo capítulo será introduzido o conceito formal de série temporal além da apresentação dos métodos que serão utilizados nesse trabalho. No capítulo [3](#) a metodologia apresentada será aplicada no conjunto de dados da incidência e mortalidade de COVID-19 e, por fim, no capítulo [4](#) faremos uma conclusão a respeito dos resultados encontrados nesse trabalho.

Capítulo 2

Análise de intervenção com base em modelos ARIMA e de Suavização Exponencial

O objetivo desse trabalho é analisar o impacto de algumas medidas sanitárias tomadas a fim de controlar a pandemia de COVID-19 no número de casos novos da doença e também no número de óbitos. Antes de abordar a metodologia que será utilizada nesse trabalho, vamos entender o que é uma série temporal que é o principal objeto de estudo desse trabalho de conclusão de curso.

Para estudar séries temporais, é importante compreender o que é um processo estocástico visto que esse processo é uma coleção de variáveis aleatórias utilizadas para estudar a evolução de fenômenos observados ao longo do tempo. [Morettin e Tolo \(2018\)](#) define um processo estocástico como:

Definição 2.1 *Seja T um conjunto arbitrário, um processo estocástico é uma família $Y = \{Y_t, t \in T\}$, tal que, para cada $t \in T$, Y_t é uma variável aleatória.*

Em outras palavras, um processo estocástico é uma família de variáveis aleatórias, dentro do mesmo espaço de probabilidade, sendo que o conjunto T é usualmente definido como o conjunto dos inteiros $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ ou o conjunto dos reais \mathbb{R} . Nesse trabalho usaremos a definição de $T \in \mathbb{Z}$.

Uma série temporal formalmente é uma realização de um processo estocástico ([Morettin e Tolo, 2018](#)), ou seja, uma série temporal é qualquer conjunto de observações geradas sequencialmente no tempo, tendo como característica a dependência entre as ob-

servações e tem a finalidade de monitorar o comportamento do evento em estudo durante esse período.

Uma determinada série de tamanho n pode ser denotada da seguinte forma: $Y = (Y_1, Y_2, \dots, Y_n)$. Assim, Y_i representa um determinado ponto da série Y no instante i , sendo $i = 1, 2, \dots, n$.

Séries temporais podem ser ditas estacionárias quando flutuam em torno da mesma média e possuem variabilidade constante ao longo do período analisado. Em alguns casos, a série pode conter tendência, o que significa apresentar mudança crescente ou decrescente em sua média. Outrossim, a série pode ter efeito sazonal, que é a repetição do mesmo comportamento de m em m observações em que m é o período da série (Morettin e Toloi, 2018).

Para análise de séries temporais podemos ajustar modelos que levam em consideração a dependência existente entre as observações já que elas são coletadas ao longo do tempo. Na literatura temos diversas classes de modelos capazes de ajustar dados com dependência, como por exemplo a classe ARIMA e a classe de suavização exponencial. Além disso, pode ser do interesse do pesquisador considerar variáveis exógenas no modelo, como por exemplo variáveis de intervenção para capturar o efeito de algum evento sobre a série em estudo.

2.1 Modelos ARIMA

Uma metodologia para análise de séries temporais e construção de modelos paramétricos é conhecida como metodologia de Box e Jenkins (Box *et al.*, 2015). Esse método se resume no ajuste de modelos autorregressivos integrados de médias móveis, na qual chamamos de ARIMA, a uma série temporal observada.

Podemos escrever um modelo ARIMA de ordem (p, d, q) da seguinte forma (Morettin e Toloi, 2018):

$$\phi(B)\Delta^d Y_t = \theta(B)a_t,$$

em que Y_t é o valor da série no tempo t , a_t é uma componente aleatória de média zero, variância σ_a^2 em que a_t e a_{t-i} são não correlacionados para todo $i \neq 0$ (ruído branco), B

é o operador um passo atrás que é definido por:

$$BY_t = Y_{t-1},$$

$\phi(B)$ é o operador autorregressivo de ordem p , $\theta(B)$ é o operador médias móveis de ordem q e são descritos por:

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q,\end{aligned}$$

Δ é o operador diferença, definido como:

$$\Delta Y_t = Y_t - Y_{t-1},$$

e d é a quantidade de diferenças realizadas na série temporal Y_t para que se torne estacionária.

A primeira etapa para a construção do modelo ARIMA é determinar a ordem do modelo, ou seja, escolher as quantidades (p, d, q) . Esta escolha é feita com base na análise das autocorrelações e autocorrelações parciais estimadas, que devem representar essas quantidades teóricas que são desconhecidas (Box *et al.*, 2015).

Em resumo, a autocorrelação no instante τ utilizada na identificação do modelo pode ser definida pela razão entre a autocovariância no instante τ sobre a variância de Y_t , em que a autocovariância da série em um instante τ , é a covariância entre Y_t e $Y_{t+\tau}$ (Morettin e Toloï, 2018).

Após termos identificado um modelo para a série temporal, a próxima etapa é a estimação dos seus parâmetros. O método escolhido para estimação dos parâmetros é o Método da Máxima Verossimilhança, que em resumo, encontra os valores de parâmetros que maximizam a plausibilidade dos dados (Hyndman e Athanasopoulos, 2021). Neste trabalho, para estimar os parâmetros utilizaremos o R Core Team (2020).

Tendo estimado os parâmetros, devemos verificar se o modelo representa adequadamente os dados através do que chamamos de análise de diagnóstico. Em resumo, o objetivo da análise de diagnóstico é verificar se os resíduos do modelo (a_t) constituem um ruído branco, ou seja, se os resíduos tem média zero, variância constante e não apresentam correlação serial. Podemos fazer essa verificação através da análise gráfica da autocorrelação

e autocorrelação parcial dos resíduos e também conduzindo testes de hipóteses (Moretтин e Toloι, 2018).

Caso o modelo não esteja adequado, as etapas de identificação, estimação e análise de diagnóstico são repetidas até que o melhor modelo seja selecionado. Nesse caso, o que chamamos de melhor modelo, é aquele que tem um comportamento residual mais semelhante ao de um ruído branco.

Após o ciclo iterativo de identificação, estimação e diagnóstico, podemos utilizar o modelo ARIMA encontrado para fazer previsões para o futuro, ou seja, podemos ter o interesse de conhecer Y_{t+h} , em que $h \geq 1$ e considerando que temos observações até o instante t . A previsão para modelos ARIMA é feita de forma que o erro quadrático médio (EQM) das previsões seja mínimo, além disso, para calcular \hat{y}_{t+h} (previsão para Y_{t+h}) precisamos de $\hat{y}_{t+h-1}, \hat{y}_{t+h-2}, \dots$ que são calculados recursivamente (Box *et al.*, 2015). No mais, a notação utilizada nesse trabalho para modelos ARIMA é $ARIMA(p, d, q)$, em que p é o grau do polinômio autorregressivo, d é o número de diferenças para que a série se torne estacionária e q é o grau do polinômio médias móveis.

É importante ressaltar que caso a série apresente comportamento sazonal, ou seja, repetições de um comportamento de m em m observações, devemos incorporar ao modelo ARIMA polinômios referentes a parte sazonal. Esse polinômios assim como os do modelo ARIMA sem sazonalidade dependem de quantidades que devem ser estimadas. Além disso, a estratégia para construção de modelos SARIMA (ARIMA com sazonalidade) é a mesma do modelo ARIMA, seguindo iterativamente as etapas de identificação, estimação e diagnóstico.

Logo, podemos ajustar modelos da classe ARIMA utilizando funções prontas de softwares como o R Core Team (2020). Nesse trabalho, utilizaremos a função `auto.arima` do pacote Hyndman e Khandakar (2008), que antes do ajuste, faz uma seleção de modelos considerando algumas medidas como o Critério de Akaike (AIC). Para mais detalhes da função `auto.arima`, veja Hyndman e Athanasopoulos (2021). Após encontrar o melhor modelo, será feita uma análise residual do ajuste selecionado para verificar a adequabilidade do modelo.

2.2 Métodos de alisamento exponencial

Os métodos de suavização exponencial segundo [Hyndman e Athanasopoulos \(2021\)](#) são métodos de previsão e são conhecidos por dar pesos diferentes as observações de uma série, de forma que os pesos vão caindo exponencialmente a medida que as observações ficam antigas. Agora vamos apresentar dois métodos, o Método Inocente e o Método Médio que introduzem os métodos mais sofisticados que veremos a seguir.

Defina \hat{y}_{t+h} a previsão para a variável Y no horizonte h e origem t , ou seja, a previsão h passos a frente considerando os dados observados até o tempo t . No método Método Inocente uma forma de encontrar a previsão \hat{y}_{t+h} é dada por:

$$\hat{y}_{t+h} = Y_t,$$

para $h = 1, 2, 3, \dots$. Com isso, o Método Inocente considera apenas a última informação como relevante para obter informações sobre o futuro. Em contrapartida, temos o Método Médio, onde as previsões do futuro são iguais a média dos n últimos valores da série observada:

$$\hat{y}_{t+h} = \frac{1}{n} \sum_{j=1}^n Y_{t-j},$$

para $h = 1, 2, 3, \dots$.

Dessa forma, o Método Médio assume que todas as observações são relevantes e tem impactos semelhantes na previsão para o futuro. O que de fato é utilizado é uma mistura desses dois métodos, de forma que as últimas observações sejam mais relevantes, mas que todas as observações sejam consideradas na previsão ([Hyndman e Athanasopoulos, 2021](#)).

2.2.1 Alisamento exponencial simples

O método de Suavização Exponencial Simples (SES) é uma mistura do Método Inocente e do Método Médio e assume que a média ponderada das observações, de forma que as observações mais recentes tenham pesos maiores, seja razoável para a previsão de valores futuros. Logo, para $0 < \alpha < 1$, podemos escrever a função de previsão da seguinte

forma:

$$\hat{y}_{t+h} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \cdots + \alpha(1 - \alpha)^n Y_0,$$

em que Y_t é o valor observado da série temporal no t -ésimo tempo e $\alpha \in [0, 1]$ é o parâmetro de suavização. Outra forma de escrever os modelos de suavização exponencial simples é dada por

$$\begin{aligned}\hat{y}_{t+h} &= \ell_t, \\ \ell_t &= \alpha Y_t + (1 - \alpha)\ell_{t-1},\end{aligned}$$

tomando $a_t = Y_t - \hat{y}_t$, temos:

$$\begin{aligned}Y_t &= \ell_t + a_t, \\ \ell_t &= \ell_{t-1} + \alpha a_t,\end{aligned}$$

em que Y_t e α são denominados como anteriormente e ℓ_t é o nível da série no momento t . Essa última maneira de representar um modelo SES não é tão útil, mas será a melhor forma para representarmos outros componentes como tendência e sazonalidade em um modelo de alisamento.

2.2.2 Alisamento exponencial com tendência

Os modelos de suavização exponencial simples como o próprio nome diz são simples de serem implementados, além de serem automatizados. Nesse momento vamos incrementar ao modelo de suavização um componente de tendência que trará ainda mais utilidade para essa metodologia, ressaltando que em problemas reais a tendência é muito comum de ser encontrada. Utilizando a forma de recorrência podemos escrever um modelo de suavização com tendência da seguinte forma ([Hyndman e Athanasopoulos, 2021](#)):

$$\begin{aligned}\hat{y}_{t+h} &= \ell_t + hb_t, \\ \ell_t &= \alpha Y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}), \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1},\end{aligned}$$

ou ainda:

$$\begin{aligned} Y_t &= \ell_{t-1} + b_{t-1} + a_t, \\ \ell_t &= \ell_{t-1} + b_{t-1} + \alpha a_t, \\ b_t &= b_{t-1} + \alpha\beta a_t, \end{aligned}$$

em que ℓ_t denota a estimativa do nível da série no instante t , b_t é uma estimativa da tendência da série no momento t , $\alpha \in [0, 1]$ é o parâmetro de suavização para o nível e $\beta \in [0, 1]$ é o parâmetro de suavização da tendência. Neste trabalho, para denotar um modelo de suavização exponencial com tendência utilizaremos a sigla A .

É importante ressaltar que em modelos de suavização exponencial também é possível incrementar um componente de sazonalidade. Da mesma forma que nos modelos de suavização exponencial com tendência, ao incrementar a sazonalidade devemos estimar s_t e γ que são, respectivamente, o componente no instante t e o parâmetro referente a sazonalidade. Além disso *a priori* devemos conhecer m que é o número de observações por ciclo sazonal.

2.3 Análise de Intervenção

Segundo [Morettin e Tolo \(2018\)](#) uma Análise de Intervenção tem por objetivo estudar o impacto de um determinado evento, *a priori* conhecido, em uma série temporal. Esse evento pode impactar temporariamente ou permanentemente a série em questão e é usualmente representado por uma variável de intervenção que pode assumir alguma das seguintes formas:

$$\begin{aligned} \text{(a)} \quad \omega B^c S_t^{(T)}, & \quad \text{(b)} \quad \omega B^c P_t^{(T)}, & \text{(2.2)} \\ \text{(c)} \quad \frac{\omega B^c}{1 - \delta B} S_t^{(T)}, & \quad \text{(d)} \quad \frac{\omega B^c}{1 - \delta B} P_t^{(T)}, \end{aligned}$$

em que B é o operador um passo atrás, T é o tempo em que ocorre a intervenção, c é o atraso do tempo antes do efeito da intervenção e deve ser conhecido, ω e $-1 < \delta < 1$ são parâmetros que devem ser estimados. $S_t^{(T)}$ e $P_t^{(T)}$ são variáveis indicadoras e são definidas

por:

$$S_t^{(T)} = \begin{cases} 0, & \text{se } t < T, \\ 1, & \text{se } t \geq T. \end{cases} \quad P_t^{(T)} = \begin{cases} 0, & \text{se } t \neq T, \\ 1, & \text{se } t = T. \end{cases}$$

Assim, ao incorporar uma variável de intervenção no modelo devemos estimar os coeficientes ω e δ para as variáveis de intervenção S_t e P_t . Além disso, como observamos, podemos ter diferentes formatos e impactos de intervenção. Por exemplo, se tomarmos $c = 1$, a intervenção (a) resulta em uma mudança permanente na função média com um atraso de uma unidade de tempo após a intervenção. O formato apresentado em (b) caracteriza uma mudança temporária no primeiro instante após a intervenção. Os formatos apresentados por (c) e (d) afetam o modelo de forma gradual. Para mais detalhes desses formatos de intervenção veja [Seong e Lee \(2021\)](#).

Para representar as intervenções em um modelo, temos uma classe geral de modelos que considera diversas ocorrências. Essa classe é descrita por ([Morettin e Toloi, 2018](#)):

$$Y_t = \sum_{j=1}^k \nu_j(B) Z_{j,t} + N_t,$$

em que:

- Y_t é a variável aleatória referente a série temporal no tempo t ;
- $Z_{j,t}$, $j = 1, 2, \dots, k$ são as variáveis de intervenção;
- B é o operador um passo atrás, onde $BY_t = Y_{t-1}$;
- $\nu_j(B)$, $j = 1, 2, \dots, k$ é a função de transferência que muda de acordo com a forma da intervenção;
- N_t é a série temporal livre dos efeitos da intervenções.

2.4 Modelo de espaço de estados

A modelagem de espaço de estados fornece uma metodologia para o tratamento de vários problemas de séries temporais. Podemos utilizar essa metodologia para fins de previsão, estimação de parâmetros e construção de intervalos de confiança, uma vez que

nessa metodologia assumimos alguma distribuição para os erros (Durbin e Koopman, 2012).

Nessa classe, os modelos são descritos por uma equação que caracteriza os dados observados e por equações que descrevem estados não observáveis como o nível, tendência e sazonalidade que são parâmetros frequentemente utilizados em séries temporais (Hyndman e Athanasopoulos, 2021). Os modelos de espaço de estados podem ser escritos da seguinte forma (Durbin e Koopman, 2012):

$$\begin{aligned} Y_t &= W'x_t + \varepsilon_t, & \varepsilon_t &\sim N(0, R_t), \\ x_t &= Fx_{t-1} + G\nu_t, & \nu_t &\sim N(0, Q_t), \end{aligned} \quad (2.3)$$

em que W , F e G são matrizes de parâmetros com dimensões $1 \times k$, $k \times k$ e $k \times 1$ respectivamente, sendo k o número de parâmetros considerado no modelo. Neste caso, como o resíduo é calculado com base em x_{t-1} ao invés de x_t , as equações 2.3 são reescritas da seguinte forma:

$$\begin{aligned} Y_t &= W'x_{t-1} + a_t, \\ x_t &= Fx_{t-1} + Ga_t, & a_t &\sim N(0, \sigma^2), \end{aligned}$$

em que $a_t = y_t - \hat{y}_t$.

É importante ressaltar que a classe de modelos de espaço de estados engloba uma quantidade de sub-classes de modelos, como os de suavização exponencial, o que significa que os modelos de suavização podem ser reescritos na forma de um modelo de espaço de estados. Vamos exemplificar utilizando o método mais simples, sem tendência e sem sazonalidade:

$$\begin{aligned} \hat{y}_{t+h} &= \ell_t \\ \ell_t &= \alpha y_t + (1 - \alpha)\ell_{t-1} \\ &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha(y_t - \hat{y}_t), \end{aligned}$$

chamando $a_t = y_t - \hat{y}_t$ (resíduo), temos:

$$\ell_t = \ell_{t-1} + \alpha a_t,$$

logo:

$$Y_t = \ell_t + a_t$$

$$\ell_t = \ell_{t-1} + \alpha a_t.$$

Tomando $W = 1$, $F = 1$, $G = \alpha$ e $\mathbf{x}_t = \ell_t$, a expressão acima pode ser reescrita como:

$$Y_t = W' \mathbf{x}_{t-1} + a_t,$$

$$\mathbf{x}_t = F \mathbf{x}_{t-1} + G a_t,$$

podendo atribuir uma densidade para a_t , que neste caso será uma normal com média 0 e desvio padrão σ .

Outrossim, podemos incrementar ao modelo de espaço de estados covariáveis que colaboram para o conhecimento da variável Y que está sendo estudada, como por exemplo variáveis de intervenção. Para acrescentar a intervenção, basta fazer algumas modificações na equação do modelo de espaço de estados acrescentando o formato de intervenção desejado. Esse novo modelo (2.4) pode ser chamado de Modelo Subjacente e é facilmente expresso como:

$$Y_t = W' \mathbf{x}_{t-1} + z_t + a_t, \tag{2.4}$$

em que z_t denota variáveis de intervenção. A forma acima é similar a análise de intervenção baseada em modelos ARIMA, mas com uma modificação no vetor de espaço de estados.

Para o ajuste dos modelos de suavização exponencial escritos na forma de um modelo de espaço de estados, utilizaremos a função `se` do pacote [Svetunkov \(2020\)](#). Essa função permite incluir variáveis exógenas no modelo, como variáveis de intervenção nos formatos 3.1 (a) e (b).

2.5 Medidas de qualidade dos modelos

Quando temos vários modelos sendo analisados precisamos de alguma estratégia para selecionar o melhor. Neste trabalho vamos utilizar duas medidas: o Erro Quadrático Médio (EQM) e o Critério de Informação de Akaike (AIC) que serão definidos na próxima subseção.

2.5.1 Erro Quadrático Médio (EQM) e Critério de Informação de Akaike (AIC)

O erro quadrático médio ou simplesmente EQM é definido como

$$\text{EQM} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

em que n é o número de observações da série que foram utilizados na estimação dos parâmetros. Como gostaríamos que o modelo ajustado se ajuste bem ao conjunto de dados, quanto menor o EQM, melhor o ajuste, uma vez que dessa forma a distância entre y_i real e ajustado (\hat{y}_i) é pequena ([Morettin e Toloí, 2018](#)).

O Critério de Informação de Akaike ou simplesmente AIC é definido como

$$\text{AIC} = n \log \left(\frac{\text{SSE}}{n} \right) + 2(k + 2),$$

em que n é o número de observações da série que foram utilizados na estimação dos parâmetros, k é o número de preditores no modelo e $\text{SSE} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$. O modelo com o valor mínimo de AIC é muitas vezes o melhor modelo ([Hyndman e Athanasopoulos, 2021](#)).

Capítulo 3

Resultados

3.1 Incidência da COVID-19 e efeitos do *lockdown* e início da vacinação.

Para ilustrar análise de intervenção baseada em modelos ARIMA e de suavização exponencial, consideramos uma série temporal da incidência de COVID-19 em algumas cidades do estado de São Paulo. A base de dados foi retirada do site Coronarívus Brasil ¹ no dia 05 de janeiro de 2021, quando o site se encontrava desatualizado e por esse motivo a base tem informações desde o início da pandemia (27 de março de 2020) até dia 21 de dezembro de 2021.

Entre muitas variáveis, o banco de dados tem informações diárias do número de casos e óbitos de COVID-19 para todos os municípios do Brasil, além do número da população de cada cidade de acordo com o Tribunal de Contas da União (TCU) de 2019. Neste trabalho, decidimos agrupar os dados em semanas começando no domingo e terminando no sábado, pois dados diários são voláteis além de conter sazonalidade. Outro motivo que nos levou a trabalhar com dados semanais é que algumas modificações eram feitas pelo próprio site para corrigir o número de casos e de mortes e isso diminuía a qualidade das análises diárias. Com isso, para cada semana calculamos o coeficiente de incidência que mede o número de casos novos da doença em relação a população exposta. Em resumo, essa medida representa o risco de ocorrência de uma doença na população (Vieira, 1997).

Uma das medidas para controlar o aumento de casos de COVID-19 e conseqüentemente o número de internações e mortes foi o *Lockdown*. Essa medida consiste em restringir a

¹<https://covid.saude.gov.br/>

circulação de pessoas em lugares públicos, permitindo apenas de forma limitada em lugares essenciais como mercados e farmácias. A cidade de Araraquara, no interior paulista, foi uma das poucas cidades brasileiras a decretar o *Lockdown*.

Com isso, decidimos analisar a incidência de COVID-19 na cidade de Araraquara e através da análise de intervenção com base em modelos ARIMA e de suavização exponencial estimamos o efeito do *Lockdown* na incidência da doença. Para fins de comparação, também consideramos dados da cidade de São Carlos, que fica à 45 quilômetros de distância de Araraquara e não decretou *Lockdown*. A série temporal da incidência de

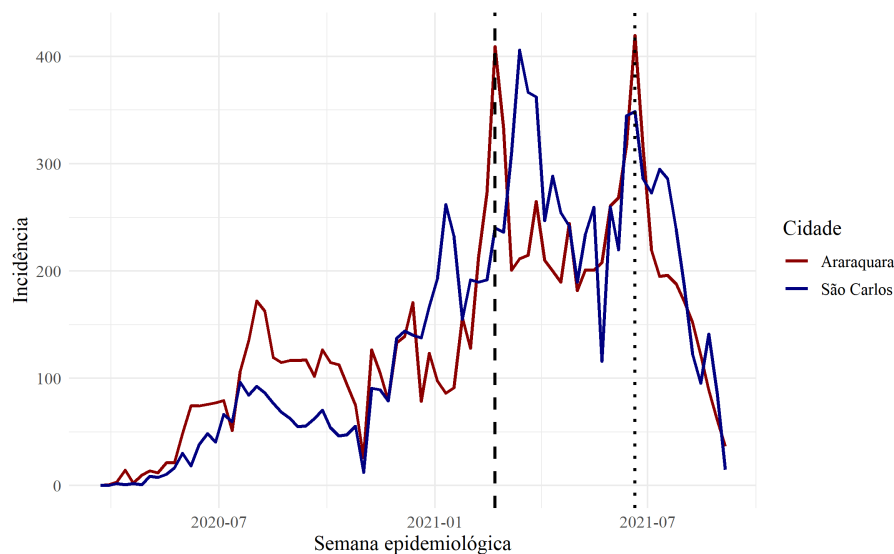


Figura 3.1: Incidência de COVID-19 em Araraquara e São Carlos.

COVID-19 das cidades Araraquara e São Carlos é apresentada na Figura 3.1. É possível observar pela análise gráfica que para ambas as cidades as séries não são estacionárias pois apresentam uma tendência de decrescimento, principalmente nas últimas semanas. Além disso, a linha vertical tracejada indica o início do primeiro *Lockdown* que ocorreu em Araraquara no fim de fevereiro de 2021. A linha vertical pontilhada representa o segundo *Lockdown* que também ocorreu em Araraquara no fim de junho de 2021.

Após o *Lockdown* que aconteceu em fevereiro, podemos observar um caimento brusco na incidência de COVID-19 em Araraquara, trazendo indícios de que a medida de segurança ajudou a controlar o risco de ocorrência da doença na população da cidade. Por outro lado, no mesmo período em São Carlos observamos que a incidência continua a crescer. Tal fato pode ter acontecido pela ausência de medidas para controlar a doença, mas também pode ter sido causado pela vinda de pessoas de Araraquara que precisavam trabalhar e que eventualmente contaminaram pessoas residentes em São Carlos.

Com relação ao *Lockdown* que ocorreu no fim de junho, observamos um caimento brusco na incidência de COVID-19 em ambas as cidades, porém dias depois a incidência em São Carlos começa a subir novamente enquanto que em Araraquara continua caindo. A redução da incidência nas cidades também pode ter sido influenciada pelo avanço da vacinação. Ainda assim, observamos que na cidade que houve *Lockdown* a redução é mais acentuada.

Para investigar se de fato os *Lockdowns* trouxeram impactos na incidência de COVID-19 em São Carlos e em Araraquara, ajustamos um modelo ARIMA em que a ordem de seus parâmetros foi escolhida conforme apresentamos no Capítulo 2. Além disso, ajustamos um modelo de suavização exponencial A, já que as séries apresentam tendência.

Nessa análise consideraremos o formato de intervenção apresentado em 3.1 (a). No caso do *Lockdown* que ocorreu no fim de fevereiro, tomamos $c = 1$ para a série de Araraquara, já que pela análise gráfica na Figura 3.1 notamos que na semana seguinte a incidência de COVID-19 passou a cair. No caso da série de São Carlos tomamos $c = 4$ pois observando a Figura 3.1 percebemos que os casos de COVID-19 começam a cair após algumas semanas. Analiticamente podemos representar o primeiro *Lockdown* em Araraquara e em São Carlos respectivamente da seguinte forma:

$$Z_{ar1} = \omega B^1 S_t^{(T)}, \quad Z_{sc1} = \omega B^4 S_t^{(T)},$$

em que T representa a semana do dia 21 de fevereiro de 2022, quando ocorreu o evento. Com relação ao *Lockdown* que ocorreu no fim de junho tomamos $c = 1$ para ambas as séries. Analiticamente podemos representar o segundo *Lockdown* em Araraquara e em São Carlos da seguinte forma:

$$Z_2 = \omega B^1 S_t^{(T)},$$

em que T representa a semana do dia 21 de junho de 2022, quando ocorreu o evento. Da Tabela 3.1, é evidente que os modelos ARIMA mostram um melhor ajuste do que os modelos de suavização exponencial independente da variável de intervenção ou da cidade, uma vez que todos os modelos ARIMA tem valores EQM (erro quadrático médio) e AIC menores que os modelos de suavização. A Tabela 3.2 mostra as estimativas das variáveis de intervenção e os respectivos erros padrão. Morettin e Toloï (2018) detecta observações atípicas no instante t verificando se $\hat{\lambda}_t > 2$, em que $\hat{\lambda}_t$ é o máximo em valor absoluto das

Tabela 3.1: Ajuste ARIMA e suavização exponencial considerando o efeito do *Lockdown*.

(a) Araraquara			(b) São Carlos		
Modelo + Intervenção	EQM	AIC	Modelo + Intervenção	EQM	AIC
ARIMA(0, 1, 0) + Z_{ar1}	1926,98	794,52	ARIMA(1, 1, 0) + Z_{sc1}	1915,87	796,13
ARIMA(0, 1, 0) + Z_2	1857,75	791,74	ARIMA(1, 1, 0) + Z_2	1862,82	794,00
$A + Z_{ar1}$	2032,04	809,88	$A + Z_{sc1}$	2032,08	809,84
$A + Z_2$	1957,23	806,95	$A + Z_2$	1987,41	808,13

Tabela 3.2: Estimativas da variável de intervenção do modelo ARIMA.

(a) Araraquara		
Efeitos	ARIMA(0, 1, 0) + Z_1	ARIMA(0, 1, 0) + Z_2
ω	-75,40	-104,63
Erro padrão	43,89	43,10
(b) São Carlos		
Efeitos	ARIMA(1, 1, 0) + Z_1	ARIMA(1, 1, 0) + Z_2
ω	20,36	-64,84
Erro padrão	43,72	41,98

estatísticas do teste da razão de verossimilhança e 2 é o nível de significância. Podemos extrapolar essa ideia para verificar se uma intervenção é significativa, de forma que se o erro da estimativa da intervenção for maior que a própria estimativa temos evidências de que aquele efeito não seja significativo, caso contrário, com algum nível de significância o contrario acontece.

Começando a analisar a Tabela 3.2 pela cidade de Araraquara, observamos que para ambas as intervenções o erro padrão é inferior a estimativa indicando que os efeitos são significativos. Ou seja, tanto o primeiro *Lockdown* no fim de fevereiro quanto o segundo *Lockdown* no fim de junho foram significativos para a redução da incidência de casos de COVID-19. Com relação a cidade de São Carlos, observamos que o erro padrão da primeira variável de intervenção é o dobro da estimativa dessa variável, logo não temos indícios de que o primeiro *Lockdown* de Araraquara tenha trazido impactos na incidência de COVID-19 em São Carlos. Em contra partida, o erro padrão da segunda intervenção é menor que a estimativa, trazendo indícios de que o segundo *Lockdown* tenha impactado na incidência de COVID-19 da cidade.

Outro evento que pode ter sido muito importante para o controle da pandemia foi o início da vacinação, que no estado de São Paulo começou no dia 17 de janeiro de 2021, quando a primeira pessoa foi vacinada. O esquema de vacinação começou priorizando

trabalhadores na linha de frente de combate à COVID-19 e pessoas mais idosas por fazerem parte do grupo de risco. Conforme os meses foram passando, a vacinação foi alcançando as demais faixas etárias até chegar nos adolescentes e crianças.

A importância da vacinação se dá pelo fato de ser a forma mais eficaz de frear a contaminação e o surgimento de novas variantes. Dessa forma, a vacinação em massa colabora para a diminuição do risco de contágio, porém caso a infecção ocorra, a vacina evitará sua evolução para quadros mais graves e principalmente a morte (TRE, 2021).

Nesse contexto, vamos investigar o efeito do início da vacinação no estado de São Paulo (17 de janeiro de 2021) na incidência e também na mortalidade de COVID-19. Como mencionado anteriormente, o esquema de vacinação foi conduzido em algumas etapas priorizando grupos de idade por fazerem parte do grupo de risco. Por esse motivo, também vamos investigar o efeito do início da vacinação em pessoas acima de 60 anos que ocorreu em 01 de abril de 2021, que de forma geral indica que a vacinação está mais avançada.

Para realizar o estudo, além dos dados da cidade de Araraquara e São Carlos, também consideramos os dados de São Paulo e Campinas que assim como as duas cidades do interior, são cidades próximas. Assim como na primeira análise de intervenção, vamos analisar o impacto da vacinação na incidência de COVID-19 e também na mortalidade, em que o coeficiente de mortalidade é a proporção de pessoas expostas ao risco na população que morreram da doença em questão (Vieira, 1997).

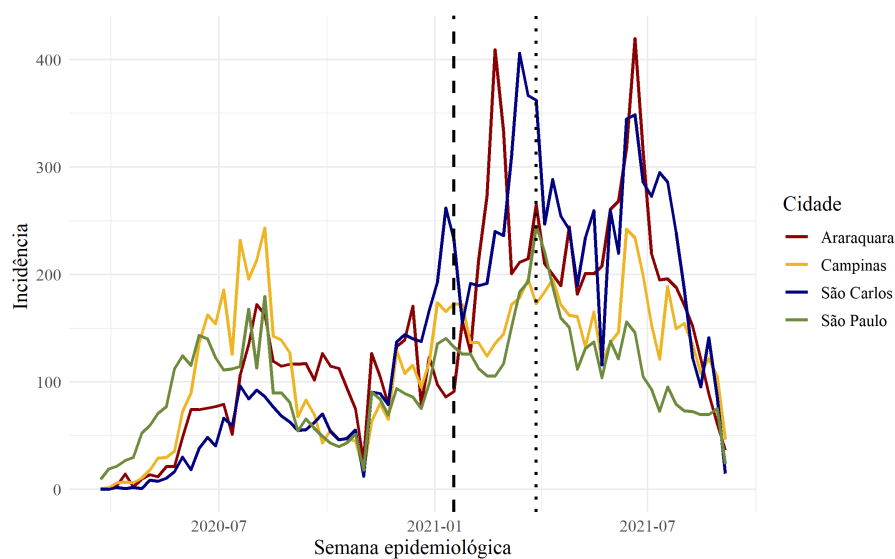


Figura 3.2: Incidência de COVID-19 em Araraquara, São Carlos, Campinas e São Paulo.

A Figura 3.2 mostra a incidência de COVID-19 nas quatro cidades em estudo. Observamos que após o início da vacinação representado no gráfico pela linha pontilhada vertical tracejada, a incidência de casos de COVID-19 oscila em todas as cidades, ou seja no caso de São Carlos, Campinas e São Paulo observamos que a incidência vinha caindo, continua caindo após a intervenção, porém semanas depois volta a subir. Com relação a Araraquara, a incidência vinha crescendo e após a intervenção continua crescendo. Esse comportamento indica que talvez fosse muito cedo para que o início da vacinação causasse algum impacto na incidência de COVID-19, uma vez que apenas pequenos grupos de pessoas estavam sendo vacinados.

Por outro lado, analisamos que, após o início da vacinação em pessoas acima de 60 anos em todas as quatro cidades em estudo (linha vertical pontilhada no gráfico da Figura 3.2), a incidência de covid-19 começa a cair, indicando que após boa parte da população estar vacinada, os número de casos novos de COVID-19 começa a cair e que de fato o início da vacinação em pessoas acima de 60 anos controlou o número de casos da doença. Algumas semanas depois da segunda intervenção observamos que a incidência volta a subir. Tal fato pode ter acontecido devido a chegada do inverno em que as pessoas eventualmente sentem sintomas de gripe, fazem mais testes de COVID-19 e conseqüentemente temos mais diagnósticos da doença.

Para analisar o impacto da vacinação, continuamos considerando o formato de intervenção apresentado em 3.1 (a). Com relação ao início da vacinação no estado de São Paulo, tomamos $c = 1$ para todas as cidades e representamos essa variável matematicamente da seguinte forma:

$$Z_3 = \omega B^1 S_t^{(T)},$$

em que T é o dia 17 de janeiro de 2021, quando o evento ocorreu. Com relação ao início da vacinação de pessoas acima de 60 anos, tomamos $c = 1$ para Araraquara, São Carlos e São Paulo uma vez que logo após a intervenção pela análise gráfica da Figura 3.2, observamos a incidência cair nessas cidades. Em Campinas por sua vez tomamos $c = 2$ por observamos uma demora de 2 semanas para que a incidência de casos de covid na cidade comece a cair. Essas intervenções foram representadas da seguinte forma:

$$Z_4 = \omega B^1 S_t^{(T)}, \quad Z_{cam4} = \omega B^2 S_t^{(T)},$$

em que T é o dia 01 de abril de 2021, quando o evento ocorreu. Da Tabela 3.3, fica

Tabela 3.3: Ajuste ARIMA e suavização exponencial considerando o efeito da vacinação na série de incidência de COVID-19.

(a) Araraquara			(b) São Carlos		
Modelo + Intervenção	EQM	AIC	Modelo + Intervenção	EQM	AIC
ARIMA(0, 1, 0) + Z_3	1926,98	795,23	ARIMA(1, 1, 0) + Z_3	1840,84	793,10
ARIMA(0, 1, 0) + Z_4	1962,50	795,90	ARIMA(1, 1, 0) + Z_4	1761,59	789,75
$A + Z_3$	2024,85	808,62	$A + Z_3$	2012,17	810,00
$A + Z_4$	2070,38	811,27	$A + Z_4$	1915,11	806,20

(c) Campinas			(d) São Paulo		
Modelo + Intervenção	EQM	AIC	Modelo + Intervenção	EQM	AIC
ARIMA(2, 1, 2) + Z_3	908,83	746,19	ARIMA(0, 1, 2) + Z_3	596,65	709,66
ARIMA(2, 1, 2) + Z_{cam4}	904,16	745,78	ARIMA(0, 1, 2) + Z_4	593,89	709,29
$A + Z_3$	1110,63	763,32	$A + Z_3$	682,09	725,78
$A + Z_4$	1102,77	762,77	$A + Z_4$	676,60	725,16

novamente evidente que os modelos ARIMA mostram um melhor ajuste do que os modelos de suavização exponencial independente da variável de intervenção ou da cidade, uma vez que todos os modelos ARIMA tem valores da EQM e AIC menores que os modelos de suavização.

A Tabela 3.4 apresenta as estimativas do efeito da intervenção e o respectivo erro padrão. Observamos que apenas em (a) e (b) a estimativa do efeito do início da vacinação em janeiro de 2021 (intervenção Z_3) é maior que o erro padrão, indicando que esse efeito pudesse ser significativo. Porém, como mencionamos anteriormente, ainda era muito cedo para que de fato a vacinação pudesse ter causado algum efeito sobre a incidência de COVID-19. Com relação ao início da vacinação em Abril de 2021, também apenas em Araraquara e São Carlos observamos a estimativa maior que o erro padrão, indicando que esse evento impactou na redução da incidência de COVID-19 uma vez que as estimativas são negativas.

Tabela 3.4: Estimativas da variável de intervenção do modelo ARIMA com relação a série de incidência.

(a) Araraquara		
Efeitos	ARIMA(0, 1, 0) + Z_3	ARIMA(0, 1, 0) + Z_4
ω	65,65	-54,64
Erro padrão	44,10	44,30

(b) São Carlos		
Efeitos	ARIMA(1, 1, 0) + Z_3	ARIMA(1, 1, 0) + Z_4
ω	-76,11	-107,64
Erro padrão	41,77	41,13

(c) Campinas		
Efeitos	ARIMA(2, 1, 2) + Z_3	ARIMA(2, 1, 2) + Z_{cam4}
ω	-3,20	-18,41
Erro padrão	30,99	28,40

(d) São Paulo		
Efeitos	ARIMA(0, 1, 2) + Z_3	ARIMA(0, 1, 2) + Z_4
ω	-8,61	-16,34
Erro padrão	22,96	22,99

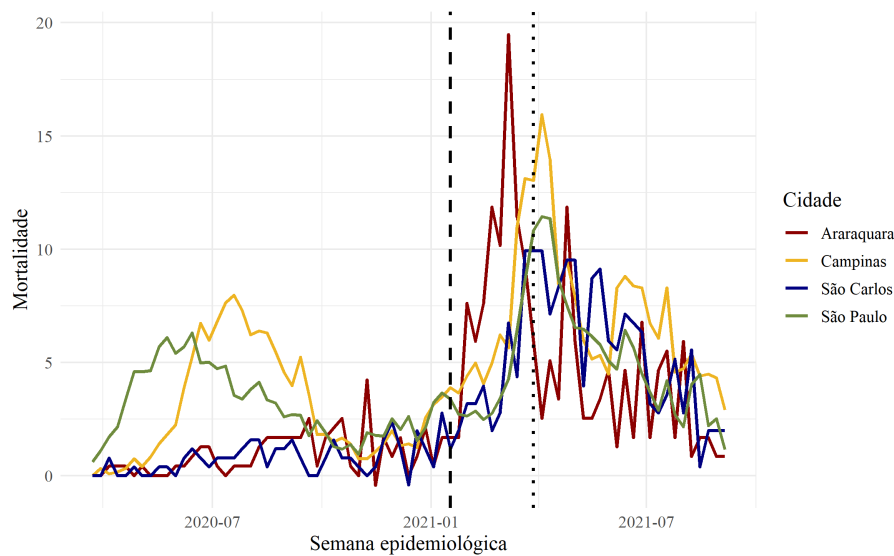


Figura 3.3: Mortalidade de COVID-19 em Araraquara, São Carlos, Campinas e São Paulo.

A Figura [Figura 3.3](#) mostra a mortalidade de COVID-19 nas quatro cidades em estudo. Observamos que após o início da vacinação no estado de São Paulo (linha vertical tracejada) a mortalidade continua em uma tendência de crescimento, acreditamos que esse comportamento se dá devido ao fato da vacinação ter avançado aos poucos obedecendo o

esquema de vacinação criado pelo governo do estado de São Paulo. O oposto é observado após o início da vacinação em pessoas acima de 60 anos (linha vertical pontilhada) no qual observamos uma tendência de decréscimo na mortalidade por COVID-19.

Para analisar o impacto da vacinação na mortalidade também vamos utilizar o formato de intervenção apresentado em 3.1 (a). No caso do início da vacinação do estado de São Paulo, vamos tomar $c = 1$ para todas as cidades, uma vez que observamos essa tendência de crescimento após a intervenção em todos os municípios. Matematicamente representaremos essa intervenção da seguinte forma:

$$Z_5 = \omega B^1 S_t^{(T)},$$

em que T é o dia 17 de janeiro de 2021, quando o evento ocorreu. Com relação a segunda intervenção, tomamos $c = 1$ para Araraquara e $c = 2$ para São Carlos, Campinas e São Paulo uma vez que pela análise gráfica (Figura 3.3) observamos que a série demora 2 semanas para ter alguma modificação. Matematicamente vamos representar o início da vacinação em pessoas maiores que 60 anos da seguinte forma:

$$Z_6 = \omega B^2 S_t^{(T)}, \quad Z_{ar6} = \omega B^1 S_t^{(T)},$$

em que T é o dia 01 de abril de 2021, quando o evento ocorreu. Da Tabela 3.5, observa-

Tabela 3.5: Ajuste ARIMA e suavização exponencial considerando o efeito da vacinação na série de mortalidade de COVID-19.

(a) Araraquara			(b) São Carlos		
Modelo + Intervenção	EQM	AIC	Modelo + Intervenção	EQM	AIC
ARIMA(4, 1, 1) + Z_5	5,10	354,48	ARIMA(0, 1, 1) + Z_5	2,27	284,52
ARIMA(4, 1, 1) + Z_{ar6}	4,28	342,47	ARIMA(0, 1, 1) + Z_6	2,26	284,00
$A + Z_5$	6,30	365,17	$A + Z_5$	2,40	290,85
$A + Z_6$	6,10	362,65	$A + Z_6$	2,38	290,12
(c) Campinas			(d) São Paulo		
Modelo + Intervenção	EQM	AIC	Modelo + Intervenção	EQM	AIC
ARIMA(3, 1, 2) + Z_5	1,96	277,19	ARIMA(2, 0, 0) + Z_5	0,77	211,15
ARIMA(3, 1, 2) + Z_6	1,93	276,06	ARIMA(2, 0, 0) + Z_6	0,77	210,77
$A + Z_5$	2,10	279,77	$A + Z_5$	0,91	216,56
$A + Z_6$	2,07	279,69	$A + Z_6$	0,90	214,75

mos que o EQM e o AIC dos modelos ARIMA e de suavização são bem próximos, mas novamente no modelo ARIMA obtemos os menores valores independente da variável de

intervenção ou da cidade. Por esse motivo analisamos os efeitos das intervenções considerando o modelo ARIMA. A Tabela 3.6 apresenta as estimativas do efeito das intervenções

Tabela 3.6: Estimativas da variável de intervenção do modelo ARIMA com relação a série de mortalidade.

(a) Araraquara		
Efeitos	ARIMA(4, 1, 1) + Z_5	ARIMA(4, 1, 1) + Z_{ar6}
ω	5,36	-2,42
Erro padrão	2,08	2,19
(b) São Carlos		
Efeitos	ARIMA(0, 1, 1) + Z_5	ARIMA(0, 1, 1) + Z_6
ω	1,17	0,30
Erro padrão	1,31	1,32
(c) Campinas		
Efeitos	ARIMA(3, 1, 2) + Z_5	ARIMA(3, 1, 2) + Z_6
ω	0,73	-3,29
Erro padrão	1,15	1,09
(d) São Paulo		
Efeitos	ARIMA(2, 0, 0) + Z_5	ARIMA(2, 0, 0) + Z_6
ω	-0,23	-2,18
Erro padrão	0,88	0,86

na série de mortalidade de COVID-19 para as quatro cidades em estudo. Com relação a intervenção do início da vacinação em janeiro, observamos que apenas em Araraquara a estimativa foi maior que o erro padrão, indicando que nessa cidade houve um impacto do início da vacinação na mortalidade. Porém mais uma vez reforçamos que ainda pudesse ser muito cedo para que a vacinação trouxesse algum impacto e que tal resultado pode estar relacionado por outro fatores.

Com relação ao início da vacinação em pessoas acima de 60 anos, observamos que em Campinas e em São Paulo a estimativa é maior que o erro padrão e por tal fato temos evidências de que esse evento foi significativo para a redução da mortalidade por COVID-19. Vale reforçar que o comportamento da mortalidade pode estar relacionado com diversos fatores externos que não observamos nessa análise, como o caso de Araraquara que passou por um decreto de *Lockdown* que como vimos foi significativo para a diminuição de novos casos de COVID-19.

Capítulo 4

Conclusão

O objetivo desse trabalho foi analisar os impactos causados pelo *Lockdown* e pelo início da vacinação no número de casos novos de COVID-19 e também na mortalidade. Os dados utilizados foram coletados desde o início da pandemia em março de 2020 até dezembro de 2021.

Para atender o objetivo ajustamos duas classes de modelos de séries temporais: ARIMA e suavização exponencial. Para considerar o efeito do *Lockdown* e início da vacinação incorporamos nos modelos variáveis de intervenção. No caso dos modelos de suavização exponencial, para inclusão da variável de intervenção, o reescrevemos na forma de um modelo de espaço de estados.

Após os ajustes dos modelos e a escolha do formato mais adequado de variável de intervenção, observamos através do AIC e do EQM que independente da intervenção ou da série que estava sendo analisada os modelos da classe ARIMA eram melhores, e por isso foram utilizados para analisar o efeito das variáveis de intervenção no número de casos e de mortes por COVID-19.

Com relação ao *Lockdown* na cidade de Araraquara, observamos que tanto o *Lockdown* que ocorreu em fevereiro quanto o de junho tiveram efeito sobre o número de casos de COVID-19. Em São Carlos observamos que apenas o *Lockdown* que ocorreu em junho teve efeito sobre o número de casos. É importante ressaltar que muitas variáveis que não foram consideradas nesse estudo podem trazer impactos nos casos de COVID-19 e estendemos a necessidade de estudá-las para ter afirmações mais certas.

Analisando a intervenção do início da vacinação em janeiro de 2021, observamos que apenas em Araraquara e São Carlos o número de casos de COVID-19 foi impactado por essa intervenção. Além disso, temos que essa intervenção também teve impacto no

número de mortes em Araraquara. Entretanto como o esquema de vacinação foi seguindo restrições que priorizavam pequenos grupos, consideramos a hipótese de que essa mudança no número de casos (em Araraquara e São Carlos) e no número de mortes (apenas em Araraquara) possa ter acontecido devido a outros fatores que não foram considerados nesse estudo.

Por fim, considerando o impacto do início da vacinação em pessoas acima de 60 anos no número de casos, observamos novamente que apenas em Araraquara e São Carlos o número de casos sofreu redução. Analisando o impacto dessa intervenção agora no número de mortes, observamos que tal medida colaborou na redução da mortalidade em Campinas e em São Paulo.

Dessa forma, podemos concluir que o *Lockdown* decretado pela prefeitura de Araraquara reduziu o número de casos de COVID-19 na cidade. Além disso, temos indícios de que o início da vacinação em pessoas acima de 60 anos colaborou para redução de casos em Araraquara e São Carlos e reduziu o número de óbitos por COVID-19 em Campinas e em São Paulo.

Em trabalhos futuros poderíamos considerar outras variáveis que eventualmente impactam na incidência e mortalidade por COVID-19, como por exemplo a idade, uma vez que a *priori* é sabido que a doença é mais letal em pessoas com mais idade. Além disso, como a vacinação seguiu um esquema que priorizava em especial pessoas com mais idade, também poderíamos pensar em analisar as séries da incidência e da mortalidade estratificada pela idade.

Referências Bibliográficas

Box, G. E., Jenkins, G. M., Reinsel, G. C. e Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

da Saúde, M. (2020). Disponível em: <<https://www.gov.br/saude/pt-br/coronavirus/o-que-e-o-coronavirus>>. Acessado em: 16 mar. 2022.

de Araraquara, P. M. (2021). Disponível em: <<https://www.araraquara.sp.gov.br/noticias/2021/fevereiro/19/novo-decreto-aumenta-restricoes-de-circulacao-para-frear-disseminacao-da-covid-19>>. Acessado em: 18 abril 2022.

Durbin, J. e Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.

Gardner Jr, E. S. (1985). Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1), 1–28.

Hyndman, R. J. e Athanasopoulos, G. (2021). *Forecasting: principles and practice*. OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on August 2021.

Hyndman, R. J. e Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3), 1–22.

Morettin, P. A. e Toloi, C. M. (2018). *Análise de séries temporais: modelos lineares univariados*. Editora Blucher.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Seong, B. e Lee, K. (2021). Intervention analysis based on exponential smoothing methods: Applications to 9/11 and covid-19 effects. *Economic Modelling*, 98, 290–301.

Svetunkov, I. (2020). *smooth: Forecasting Using State Space Models*. R package version 2.6.0.

TRE (2021). Disponível em: <<https://www.tre-se.jus.br/imprensa/noticias-tre-se/2021/Julho/importancia-e-beneficios-da-vacinacao>>.

Acessado em: 16 mar. 2022.

Vieira, S. (1997). *Introdução à bioestatística*. Elsevier Brasil.

Apêndice A

Símbolos matemáticos

- T conjunto arbitrário do tempo;
- t é o t -ésimo tempo;
- Y_t variável aleatória referente a série no instante t ;
- \hat{y}_{t+h} é a previsão para o horizonte h ;
- ℓ_t é o nível no tempo t ;
- α é o parâmetro de suavização;
- b_t é a estimativa para tendência;
- β é o parâmetro de suavização da tendência;
- s_t é a estimativa para sazonalidade;
- γ é o parâmetro para a sazonalidade;
- m é o número de observações por ciclo sazonal;
- S_t e P_t são intervenções;
- c é o atraso do tempo antes do efeito da intervenção;
- $Z_{j,t}$ são intervenções;
- B operador passo atrás;
- ν_j função de transferência;

- N_t é série livre de intervenção;
- x_t é o vetor de espaço de estados;
- W vetor de medição;
- F matriz de transição;
- G matriz de parâmetros;
- a_t é o ruído branco;

Apêndice B

Códigos utilizados

```
library(dplyr)
library(tidyr)
library(lubridate)
library(ggplot2)
library(aweeek)
library(forecast)

hist_2020_1 <- read.csv("C:/Projetos/TCC/TG2/HIST_PAINEL_COVIDBR_2020_
Parte1_21dez2021.csv", sep=";")
hist_2020_2 <- read.csv("C:/Projetos/TCC/TG2/HIST_PAINEL_COVIDBR_2020_
Parte2_21dez2021.csv", sep=";")
hist_2021_1 <- read.csv("C:/Projetos/TCC/TG2/HIST_PAINEL_COVIDBR_2021_
Parte1_21dez2021.csv", sep=";")
hist_2021_2 <- read.csv("C:/Projetos/TCC/TG2/HIST_PAINEL_COVIDBR_2021_
Parte2_21dez2021.csv", sep=";")

sc_2020_1=hist_2020_1 %>% filter(municipio %in% "São Carlos") %>%
filter(regiao %in% "Sudeste")
sc_2020_2=hist_2020_2 %>% filter(municipio %in% "São Carlos")%>%
filter(regiao %in% "Sudeste")
sc_2021_1=hist_2021_1 %>% filter(municipio %in% "São Carlos")%>%
filter(regiao %in% "Sudeste")
sc_2021_2=hist_2021_2 %>% filter(municipio %in% "São Carlos")%>%
filter(regiao %in% "Sudeste")

dados_sc = rbind(sc_2020_1[1:nrow(sc_2020_1),],sc_2020_2,sc_2021_1,sc_2021_2)
week=substring(date2week(dados_sc$data , week_start = 7),1,8)
dados_sc = cbind(dados_sc,week)
```

```

#média da população
pop_sc=group_by(dados_sc,week)%>%summarise(Media=mean(populacaoTCU2019))

#agrupando novos casos
dados_agrup_sc=group_by(dados_sc,week)%>%summarise(Total=sum(casosNovos))
#agrupando novos obitos
dados_agrup_scm=group_by(dados_sc,week)%>%summarise(Total=sum(obitosNovos))

incidencia = dados_agrup_sc$Total/pop_sc$Media*10^5
tmortalidade = dados_agrup_scm$Total/pop_sc$Media*10^5

cidade=rep("São Carlos",nrow(dados_agrup_sc))

dados_agrup_sc = cbind(dados_agrup_sc,tmortalidade,incidencia,cidade)

dados_agrup_sc = dados_agrup_sc[1:77,]

ar_2020_1=hist_2020_1 %>% filter(municipio %in% "Araraquara")
ar_2020_2=hist_2020_2 %>% filter(municipio %in% "Araraquara")
ar_2021_1=hist_2021_1 %>% filter(municipio %in% "Araraquara")
ar_2021_2=hist_2021_2 %>% filter(municipio %in% "Araraquara")

dados_ar = rbind(ar_2020_1[1:nrow(ar_2020_1),],ar_2020_2,ar_2021_1,ar_2021_2)
week=substring(date2week(dados_ar$data , week_start = 7),1,8)
dados_ar = cbind(dados_ar,week)

#média da população
pop_ar=group_by(dados_ar,week)%>%summarise(Media=mean(populacaoTCU2019))

#agrupando novos casos
dados_agrup_ar=group_by(dados_ar,week)%>%summarise(Total=sum(casosNovos))
#agrupando novos obitos
dados_agrup_arm=group_by(dados_ar,week)%>%summarise(Total=sum(obitosNovos))

incidencia = dados_agrup_ar$Total/pop_ar$Media*10^5
tmortalidade = dados_agrup_arm$Total/pop_ar$Media*10^5

cidade=rep("Araraquara",nrow(dados_agrup_ar))

dados_agrup_ar = cbind(dados_agrup_ar,tmortalidade,incidencia,cidade)

```

```
dados_agrup_ar = dados_agrup_ar[1:77,]
```

```
cam_2020_1=hist_2020_1 %>% filter(municipio %in% "Campinas")
```

```
cam_2020_2=hist_2020_2 %>% filter(municipio %in% "Campinas")
```

```
cam_2021_1=hist_2021_1 %>% filter(municipio %in% "Campinas")
```

```
cam_2021_2=hist_2021_2 %>% filter(municipio %in% "Campinas")
```

```
dados_cam = rbind(cam_2020_1[1:nrow(cam_2020_1)], cam_2020_2, cam_2021_1, cam_2021_2)
```

```
week=substring(date2week(dados_cam$data , week_start = 7),1,8)
```

```
dados_cam = cbind(dados_cam, week)
```

```
#média da população
```

```
pop_cam=group_by(dados_cam, week)%>%summarise(Media=mean(populacaoTCU2019))
```

```
#agrupando novos casos
```

```
dados_agrup_cam=group_by(dados_cam, week)%>%summarise(Total=sum(casosNovos))
```

```
#agrupando novos obitos
```

```
dados_agrup_camm=group_by(dados_cam, week)%>%summarise(Total=sum(obitosNovos))
```

```
incidencia = dados_agrup_cam$Total/pop_cam$Media*105
```

```
tmortalidade = dados_agrup_camm$Total/pop_cam$Media*105
```

```
cidade=rep("Campinas", nrow(dados_agrup_cam))
```

```
dados_agrup_cam = cbind(dados_agrup_cam, tmortalidade, incidencia, cidade)
```

```
dados_agrup_cam = dados_agrup_cam[1:77,]
```

```
sp_2020_1=hist_2020_1 %>% filter(municipio %in% "São Paulo")
```

```
sp_2020_2=hist_2020_2 %>% filter(municipio %in% "São Paulo")
```

```
sp_2021_1=hist_2021_1 %>% filter(municipio %in% "São Paulo")
```

```
sp_2021_2=hist_2021_2 %>% filter(municipio %in% "São Paulo")
```

```
dados_sp = rbind(sp_2020_1[1:nrow(sp_2020_1)], sp_2020_2, sp_2021_1, sp_2021_2)
```

```
week=substring(date2week(dados_sp$data , week_start = 7),1,8)
```

```
dados_sp = cbind(dados_sp, week)
```

```
#média da população
```

```
pop_sp=group_by(dados_sp, week)%>%summarise(Media=mean(populacaoTCU2019))
```

```
#agrupando novos casos
```

```

dados_agrup_sp=group_by(dados_sp,week)%>%summarise(Total=sum(casosNovos))
#agrupando novos obitos
dados_agrup_spm=group_by(dados_sp,week)%>%summarise(Total=sum(obitosNovos))

incidencia = dados_agrup_sp$Total/pop_sp$Media*10^5
tmortalidade = dados_agrup_spm$Total/pop_sp$Media*10^5

cidade=rep("São Paulo",nrow(dados_agrup_sp))

dados_agrup_sp = cbind(dados_agrup_sp,tmortalidade,incidencia,cidade)

dados_agrup_sp = dados_agrup_sp[1:77,]

idados_cidades = cbind("semana"=dados_agrup_sc$week,"sc"=dados_agrup_sc$incidencia,
"ar"=dados_agrup_ar$incidencia,"cam"=dados_agrup_cam$incidencia,
"sp"=dados_agrup_sp$incidencia)

mdados_cidades = cbind("semana"=dados_agrup_sc$week,"sc"=dados_agrup_sc$tmortalidade,
"ar"=dados_agrup_ar$tmortalidade,"cam"=dados_agrup_cam$tmortalidade,
"sp"=dados_agrup_sp$tmortalidade)

# LOCKDOWN ARARAQUARA
# são carlos
inci_sc=as.numeric(idados_cidades[,2])
selecao_arima <- auto.arima(inci_sc, stepwise = FALSE) #ARIMA(1,1,0)

# começa a cair só em 52 - c=4
#21-02-21 - linha 49
int_1sc=c(rep(0,52),rep(1,25))
m1<-arima(inci_sc,order=c(1,1,0),xreg=int_1sc)
summary(m1)

#20-06-21 - linha 66
int_2=c(rep(0,66),rep(1,11))
m2<-arima(inci_sc,order=c(1,1,0),xreg=int_2)

# araraquara
inci_ar=as.numeric(idados_cidades[,3])
selecao_arima <- auto.arima(inci_ar, stepwise = FALSE) #ARIMA(0,1,0)

#21-02-21 - linha 49

```

```

int_1ar=c(rep(0,49),rep(1,28))
m1<-arima(inci_ar,order=c(0,1,0),xreg=int_1ar)
summary(m1)

int_2=c(rep(0,66),rep(1,11))
m2<-arima(inci_ar,order=c(0,1,0),xreg=int_2)
summary(m2)

# VACINAÇÃO ESTADO DE SP
# são carlos
inci_sc=as.numeric(idados_cidades[,2])
selecao_arima <- auto.arima(inci_sc, stepwise = FALSE) #ARIMA(1,1,0)

#17-01-21 - linha 44
int_1sc=c(rep(0,44),rep(1,33))
m1<-arima(inci_sc,order=c(1,1,0),xreg=int_1sc)
summary(m1) #intervenção não significativa

#01-04-21 - linha 54
int_2=c(rep(0,54),rep(1,23))
m2<-arima(inci_sc,order=c(1,1,0),xreg=int_2)
summary(m2)

# araraquara
inci_ar=as.numeric(idados_cidades[,3])
selecao_arima <- auto.arima(inci_ar, stepwise = FALSE) #ARIMA(0,1,0)

#17-01-21 - linha 44
int_1ar=c(rep(0,44),rep(1,33))
m1<-arima(inci_ar,order=c(0,1,0),xreg=int_1ar)
summary(m1)

#01-04-21 - linha 54
int_2=c(rep(0,54),rep(1,23))
m2<-arima(inci_ar,order=c(0,1,0),xreg=int_2)
summary(m2)

# campinas
inci_cam=as.numeric(idados_cidades[,4])
selecao_arima <- auto.arima(inci_cam, stepwise = FALSE) #ARIMA(2,1,2)

```

```
#17-01-21 - linha 44
```

```
int_1ar=c(rep(0,44),rep(1,33))
m1<-arima(inci_cam,order=c(2,1,2),xreg=int_1ar)
summary(m1)
```

```
#01-04-21 - linha 66
```

```
int_2=c(rep(0,56),rep(1,21))
m2<-arima(inci_cam,order=c(2,1,2),xreg=int_2)
summary(m2)
```

```
# sao paulo
```

```
inci_sp=as.numeric(idados_cidades[,5])
selecao_arima <- auto.arima(inci_sp, stepwise = FALSE) #ARIMA(0,1,2)
```

```
#17-01-21 - linha 44
```

```
int_1ar=c(rep(0,44),rep(1,33))
m1<-arima(inci_sp,order=c(0,1,2),xreg=int_1ar)
summary(m1)
```

```
#01-04-21 - linha 54
```

```
int_2=c(rep(0,54),rep(1,23))
m2<-arima(inci_sp,order=c(2,1,2),xreg=int_2)
summary(m2)
```

```
##### mortalidade
```

```
# são carlos
```

```
mort_sc=as.numeric(mdados_cidades[,2])
selecao_arima <- auto.arima(mort_sc, stepwise = FALSE) #ARIMA(0,1,1)
```

```
int3 = c(rep(0,44),rep(1,33))
m3<-arima(mort_sc,order=c(0,1,1),xreg=int3)
summary(m3)
```

```
int4 = c(rep(0,56),rep(1,21))
m4<-arima(mort_sc,order=c(0,1,1),xreg=int4)
summary(m4)
```

```
#araraquara
```

```
mort_ar=as.numeric(mdados_cidades[,3])
selecao_arima <- auto.arima(mort_ar, stepwise = FALSE) #ARIMA(4,1,1)
```

```

int3 = c(rep(0,44),rep(1,33))
m3<-arima(mort_ar,order=c(4,1,1),xreg=int3)
summary(m3)

int4 = c(rep(0,54),rep(1,23))
m4<-arima(mort_ar,order=c(4,1,1),xreg=int4)
summary(m4)

#campinas
mort_cam=as.numeric(mdados_cidades[,4])
selecao_arima <- auto.arima(mort_cam, stepwise = FALSE) #ARIMA(3,1,2)

int3 = c(rep(0,44),rep(1,33))
m3<-arima(mort_cam,order=c(3,1,2),xreg=int3)
summary(m3) #intervenção não significativa

int4 = c(rep(0,60),rep(1,17))
m4<-arima(mort_cam,order=c(3,1,2),xreg=int4)
summary(m4)

#São Paulo
mort_sp=as.numeric(mdados_cidades[,5])
selecao_arima <- auto.arima(mort_sp, stepwise = FALSE) #ARIMA(2,0,0)

int3 = c(rep(0,44),rep(1,33))
m3<-arima(mort_sp,order=c(2,0,0),xreg=int3)
summary(m3) #intervenção não significativa

int4 = c(rep(0,56),rep(1,21))
m4<-arima(mort_sp,order=c(2,0,0),xreg=int4)
summary(m4)

# gráficos

s1=ggplot(dados_cidades[1:154,], aes(x = Semana, y = Incidencia)) +
  geom_line(aes(color = Cidade), size = 1) +
  scale_color_manual(values = c("red4", "navyblue")) +
  geom_vline(xintercept=as.numeric(dados_cidades$Semana[c(49)]),
            linetype=2, colour="black",size=1)+
  geom_vline(xintercept=as.numeric(dados_cidades$Semana[c(66)]),
            linetype=3, colour="black",size=1)+

```

```

labs(x = "Semana epidemiológica", y = "Incidência",title = "")+theme_minimal()+
theme(text = element_text(size = 13,
                           family ="serif"),
      plot.title = element_text(hjust = 0.5))
s1
ggsave(s1, filename = "incidencia_sc_ar.png",
      path =
        "C:/Projetos/TCC",
      width = 20,
      height = 13,
      units = "cm")

s2=ggplot(dados_cidades, aes(x = Semana, y = Incidencia)) +
  geom_line(aes(color = Cidade), size = 1) +
  scale_color_manual(values = c("red4", "goldenrod2","navyblue","darkolivegreen4")) +
  geom_vline(xintercept=as.numeric(dados_cidades$Semana[c(44)]),
            linetype=2, colour="black",size=1)+
  geom_vline(xintercept=as.numeric(dados_cidades$Semana[c(54)]),
            linetype=3, colour="black",size=1)+
  labs(x = "Semana epidemiológica", y = "Incidência ",title = "")+theme_minimal()+
  theme(text = element_text(size = 13,
                            family ="serif"),
        plot.title = element_text(hjust = 0.5))
s2

ggsave(s2, filename = "incidencia_sc_ar_cam_sp.png",
      path =
        "C:/Projetos/TCC",
      width = 20,
      height = 13,
      units = "cm")

s3=ggplot(dados_cidades, aes(x = Semana, y = Mortalidade)) +
  geom_line(aes(color = Cidade), size = 1) +
  scale_color_manual(values = c("red4", "goldenrod2","navyblue","darkolivegreen4")) +
  geom_vline(xintercept=as.numeric(dados_cidades$Semana[c(44)]),
            linetype=2, colour="black",size=1)+
  geom_vline(xintercept=as.numeric(dados_cidades$Semana[c(54)]),
            linetype=3, colour="black",size=1)+
  labs(x = "Semana epidemiológica", y = "Mortalidade",title = "")+theme_minimal()+
  theme(text = element_text(size = 13,

```



```
        family = "serif"),  
  plot.title = element_text(hjust = 0.5))  
s3  
  
ggsave(s3, filename = "mortalidade_sc_ar_cam_sp.png",  
  path =  
    "C:/Projetos/TCC",  
  width = 20,  
  height = 13,  
  units = "cm")
```