

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE CIÊNCIAS ECONÔMICAS
DEPARTAMENTO DE ECONOMIA E FINANÇAS

PABLO LIMA CARREIRA

Estudo do isolamento social no período pandêmico em São Paulo

Orientador: Prof. Dr. Eduardo Simões de Almeida

JUIZ DE FORA – MG

2025

PABLO LIMA CARREIRA

Estudo do isolamento social no período pandêmico em São Paulo

Projeto de monografia apresentado ao Curso de Ciências Econômicas da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Eduardo Simões de Almeida

JUIZ DE FORA – MG

2025

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Carreira, Pablo Lima.

Estudo do isolamento social no período pandêmico em São Paulo / Pablo Lima Carreira. -- 2025.

43 p. : il.

Orientador: Eduardo Simões de Almeida

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Faculdade de Economia, 2025.

1. COVID-19. 2. Pandemia. 3. Séries Temporais. 4. Modelo SARIMA. I. Almeida, Eduardo Simões de, orient. II. Título.



UNIVERSIDADE FEDERAL DE JUIZ DE FORA
REITORIA - FACECON - Depto. de Economia

FACULDADE DE ECONOMIA / UFJF

ATA DE APROVAÇÃO DE MONOGRAFIA II (MONO B)

NA DATA DE 12/03/2025, A BANCA EXAMINADORA, COMPOSTA PELOS PROFESSORES:

1 – EDUARDO SIMÕES DE ALMEIDA - ORIENTADOR; E

2 – RAFAEL MORAIS DE SOUZA,

REUNIU-SE PARA AVALIAR A MONOGRAFIA DO ACADÊMICO **PABLO LIMA CARREIRA**, INTITULADA "**ESTUDO DO ISOLAMENTO SOCIAL NO PERÍODO PANDÊMICO EM SÃO PAULO**".

APÓS PRIMEIRA AVALIAÇÃO, RESOLVEU A BANCA SUGERIR ALTERAÇÕES AO TEXTO APRESENTADO, CONFORME RELATÓRIO SINTETIZADO PELO ORIENTADOR. A BANCA, DELEGANDO AO ORIENTADOR A OBSERVÂNCIA DAS ALTERAÇÕES PROPOSTAS, RESOLVEU APROVAR A REFERIDA MONOGRAFIA.

ASSINATURA ELETRÔNICA DOS PROFESSORES AVALIADORES



Documento assinado eletronicamente por **Eduardo Simoes de Almeida, Professor(a)**, em 13/03/2025, às 19:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rafael Morais de Souza, Professor(a)**, em 14/03/2025, às 16:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2291843** e o código CRC **1E4066AB**.

RESUMO

O presente trabalho tem como objetivo realizar uma análise descritiva e exploratória dos dados de isolamento social durante o período pandêmico no estado de São Paulo. A análise descritiva e exploratória visa auxiliar e otimizar o processo de modelagem posterior, destacando-se como etapas cruciais no fluxo de análise de dados. Analisando os gráficos, observou-se que o modelo apresentou ótimo ajuste aos dados, captando a tendência de crescimento no início da pandemia. Concluiu-se que a série temporal tem um comportamento de sazonalidade forte, ocorrendo semanalmente. O modelo SARIMA sugeriu uma componente autorregressiva de baixa ordem, uma componente de média móvel que inclui tanto um termo de curto prazo quanto um termo sazonal semanal, e um componente de erro aleatório. Este modelo validou-se como adequado e bem ajustado aos dados, oferecendo contribuições valiosas para a compreensão do comportamento do isolamento social durante a pandemia.

Palavras-chave: COVID-19, Pandemia, Séries Temporais, Modelo SARIMA

ABSTRACT

The present work aims to perform a descriptive and exploratory analysis of social isolation data during the pandemic period in the state of São Paulo. The descriptive and exploratory analysis aims to assist and optimize the subsequent modeling process, standing out as crucial steps in the data analysis flow. By analyzing the graphs, it was observed that the model presented an excellent fit to the data, capturing the growth trend at the beginning of the pandemic. It was concluded that the time series exhibits a strong seasonality pattern, occurring weekly. The SARIMA model suggested a low-order autoregressive component, a moving average component that includes both a short-term term and a weekly seasonal term, and a random error component. This model proved to be adequate and well-fitted to the data, offering valuable contributions to understanding the behavior of social isolation during the pandemic.

Keywords: COVID-19, Pandemic, Time Series, SARIMA Model

LISTA DE FIGURAS

Figura 1 - Estado de São Paulo separado por mesorregiões.....	16
Figura 2 – Isolamento por dia, separado em mesorregiões.....	21

LISTA DE TABELAS

Tabela 1 - Mesorregiões com as populações estimadas.....	12
Tabela 2 - Índice médio de Isolamento Social diário.....	25
Tabela 3 - Teste ADF.....	28
Tabela 4 - Teste KPSS.....	29
Tabela 5 - ADF aplicado após diferenciação.....	33
Tabela 6 - KPSS aplicado após diferenciação.....	33
Tabela 7 - Aplicação do Teste Ljung-Box.....	37
Tabela 8 - Coeficientes Estimados.....	37

LISTA DE GRÁFICOS

Gráfico 1 - Série por Mesorregião.....	17
Gráfico 2 - Acumulado de isolamento mensal/região.....	18
Gráfico 3 - Média Estadual.....	19
Gráfico 4 - Gráfico Sazonal método STL.....	20
Gráfico 5 - Gráfico da Série (Variável Locais de Trabalho).....	22
Gráfico 6 - Estimação por STL da variável Locais de Trabalho.....	23
Gráfico 7 - Ida a Locais de Trabalho por dia em SP.....	24
Gráfico 8 - Isolamento Social por Dia em SP.....	26
Gráfico 9 - Funções de autocorrelação e autocorrelação parcial.....	27
Gráfico 10 - Estimação por STL da variável Isolamento Social.....	30
Gráfico 11 - Método de diferenciação.....	31
Gráfico 12 - Resíduos do Modelo.....	36
Gráfico 13 e 14 - Gráfico Ajustado Modelo SARIMA.....	39

LISTA DE ABREVIATURAS E SIGLAS

STL - Decomposição sazonal e de tendências usando Loess

ACF - Autocorrelation Function

KPPS – Teste de Kwiatkowski–Phillips–Schmidt–Shin

ADF - Teste de Dickey–Fuller Aumentado

SUMÁRIO

1. Introdução.....	12
1.1 Problemática.....	13
1.2 Justificativa.....	13
1.3 Objetivos.....	14
1.3.1 Objetivo Geral.....	14
1.3.2 Objetivos Específicos.....	14
2. Organização dos Dados.....	15
3. Descrição e Exploração.....	17
4. Análise via Google Mobility.....	21
5. Análise Inicial da Série.....	22
6. Análise e Definição Iniciais dos Modelos.....	23
7. Modelagem.....	29
8. Análise de Resíduos.....	33
9. Análise do Modelo Final e Conclusão.....	34
Referências.....	38

1. Introdução

A pandemia de COVID-19, declarada pela Organização Mundial da Saúde (OMS) em março de 2020, trouxe desafios sem precedentes para a saúde pública global. Entre as diversas medidas adotadas para conter a disseminação do vírus, o isolamento social destacou-se como uma das mais eficazes (NIU; XU, 2020). No Brasil, e especificamente no estado de São Paulo, essas medidas envolveram desde o fechamento de escolas e comércios até restrições de mobilidade e aglomerações (SILVA et al., 2021).

Para monitorar e avaliar a eficácia dessas medidas, o estado de São Paulo implementou o SIMI-SP (Sistema de Monitoramento Inteligente de São Paulo). Viabilizado por meio de um acordo com as operadoras de telefonia Vivo, Claro, Oi e TIM, através da Associação Brasileira de Recursos em Telecomunicações (ABR) e do Instituto de Pesquisas Tecnológicas (IPT), o SIMI-SP permite ao estado acessar informações agregadas e anônimas sobre o deslocamento dos cidadãos nos municípios paulistas mapeados. Em respeito à proteção de dados, as informações são anonimizadas e aglutinadas, assegurando a privacidade dos usuários enquanto fornecem dados georreferenciados essenciais para a elaboração de políticas públicas eficazes no combate ao coronavírus, como referenciado no próprio site do governo de São Paulo.

O índice de isolamento social calculado a partir desses dados tornou-se um indicador crucial para avaliar a adesão da população às medidas restritivas. Estudos indicam que fatores como a percepção de risco, confiança nas autoridades e situação socioeconômica influenciam o comportamento da população em relação ao isolamento social (BEZERRA et al., 2020). Compreender como esse índice variou ao longo do tempo e em diferentes regiões do estado é fundamental para analisar a eficácia das políticas públicas implementadas e planejar ações futuras em situações similares.

A escolha de estudar o índice de isolamento social em São Paulo se justifica pela importância econômica e populacional do estado, além de sua alta densidade demográfica, que potencializa os riscos de contágio. São Paulo é o estado mais populoso do Brasil,

concentrando aproximadamente 46 milhões de habitantes (IBGE, 2023), além de ser um dos principais polos econômicos do país. Estudos indicam que fatores como mobilidade urbana e alta densidade populacional podem influenciar significativamente a disseminação de doenças infecciosas, como observado durante a pandemia da COVID-19 (IPEA, 2021; BEZERRA et al., 2020). Analisar os dados de isolamento social durante a pandemia permite não apenas compreender os padrões de comportamento da população, mas também oferece insights valiosos para gestores públicos e pesquisadores sobre a eficácia das intervenções e o impacto das mesmas na contenção da COVID-19. Esse estudo contribui para a literatura existente ao fornecer uma visão abrangente e detalhada do comportamento do isolamento social em São Paulo durante a pandemia, auxiliando na formulação de estratégias mais eficazes para enfrentar crises sanitárias futuras (SILVA et al., 2021).

1.1 Problemática

A série temporal do índice de isolamento social no estado de São Paulo durante o período pandêmico foi analisada a partir de métodos de séries temporais. Busca-se então identificar e descrever os componentes da série e determinar quais modelos da classe SARIMA (BOX; JENKINS; REINSEL, 2015) fornecem as melhores previsões sobre o comportamento do isolamento social, visando entender a eficácia das medidas de contenção e orientar futuras políticas públicas.

1.2 Justificativa

O índice de isolamento social, medido a partir dos deslocamentos da população, é um ótimo indicador para avaliar a eficácia das medidas adotadas durante a pandemia de COVID-19. No estado de São Paulo, um dos mais populosos e economicamente relevantes do Brasil, compreender os padrões de isolamento social é essencial para a formulação de políticas públicas mais eficientes. Este trabalho utiliza métodos de séries temporais, como o modelo SARIMA (BOX; JENKINS; REINSEL, 2015), para analisar e prever o comportamento do isolamento social, fornecendo uma base empírica robusta para avaliar a aderência da população às restrições e a eficácia das intervenções. A análise detalhada dos dados históricos permite identificar tendências e sazonalidades que podem informar futuras estratégias de

contenção, contribuindo significativamente para a gestão de crises sanitárias e a melhoria das políticas de saúde pública. Portanto, este estudo é relevante porque oferece uma compreensão aprofundada do comportamento do isolamento social em São Paulo durante a pandemia, auxiliando na tomada de decisões mais informadas e eficazes (BEZERRA et al., 2020).

1.3 Objetivos

1.3.1 Objetivo geral

Analisar os dados de isolamento social no estado de São Paulo durante o período pandêmico, com o intuito de compreender a eficácia das medidas de isolamento e o comportamento da população frente às restrições impostas, utilizando técnicas de análise descritiva, exploratória e modelagem em séries temporais.

1.3.2 Objetivos específicos

- a) Realizar uma análise descritiva e exploratória dos dados de isolamento social:
 - Identificar padrões e tendências nos dados ao longo do tempo.
 - Avaliar variações regionais e temporais no índice de isolamento social.
- b) Desenvolver um modelo SARIMA para os dados de isolamento social:
 - Ajustar o modelo aos dados históricos para capturar tendências e sazonalidades.
 - Validar o modelo, verificando seu desempenho e adequação aos dados observados.
- c) Interpretar os resultados obtidos com o modelo SARIMA:
 - Analisar a componente autorregressiva, a componente de média móvel e a componente de erro aleatório do modelo.
 - Identificar e compreender a sazonalidade semanal presente nos dados.

2 Organização dos Dados

A seguinte pesquisa tem como objetivo realizar uma análise descritiva e exploratória nos dados de isolamento social durante o período pandêmico no estado de São Paulo, disponibilizados publicamente pelo Sistema de Monitoramento Inteligente do Governo de São Paulo. Os dados possuem um recorte temporal entre 26-02-2020 e 31-12-2021. As etapas de análise descritiva e exploratória tem como objetivo auxiliar e otimizar o processo de modelagem posterior, e assim são etapas importantíssimas no fluxo de análise de dados.

Os dados disponíveis são divididos por cidades no estado de SP. Buscando facilitar a análise, os dados foram agrupados por mesorregiões, passando a ter 15 mesorregiões finais para análise. As mesorregiões com suas populações estimadas para 2021 são:

Tabela 1 – Mesorregiões com as populações estimadas

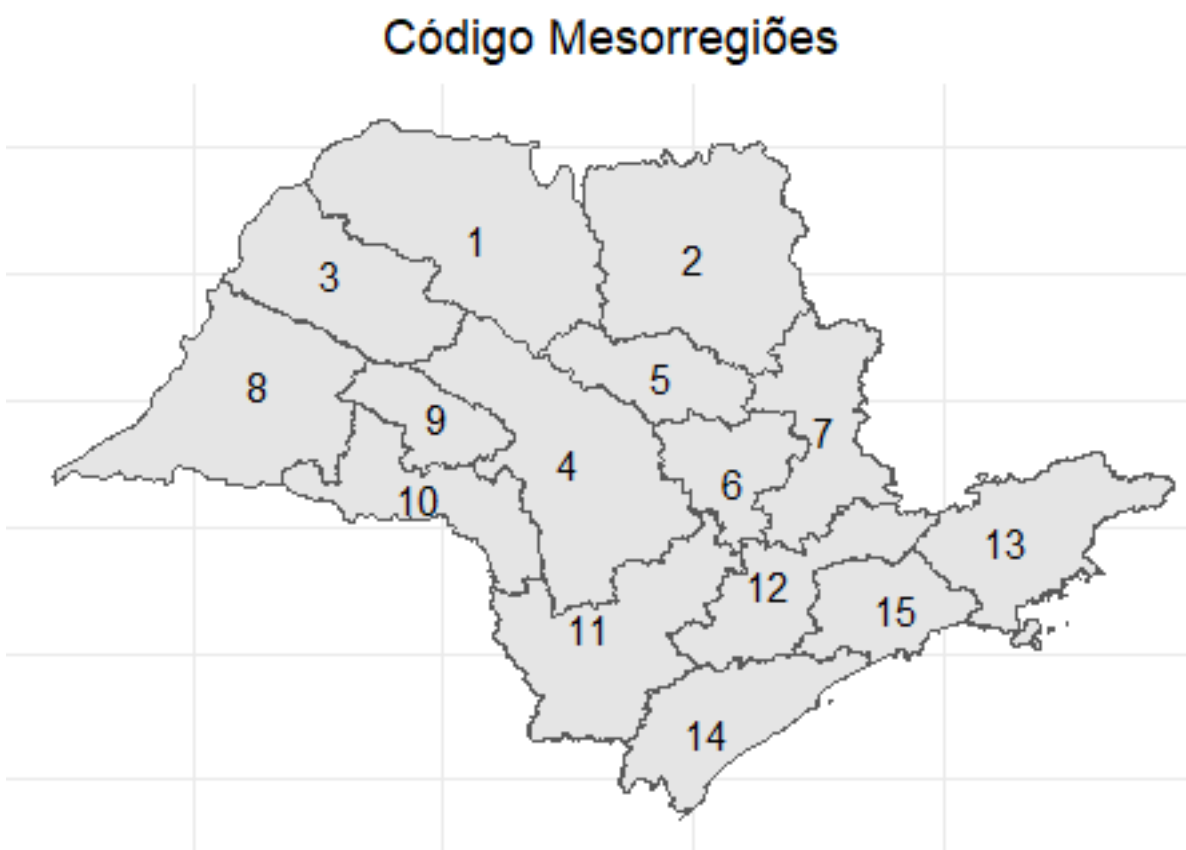
cod_mesoreg	mesoreg	pop_estim
1	sao_jose_do_rio_pr eto	867,653
2	ribeirao_preto	1,696,343
3	aracatuba	443,971
4	bauru	918,033
5	araraquara	637,049
6	piracicaba	1,219,973
7	campinas	3,803,313
8	presidente_prudent e	230,371
9	marilia	306,160
10	assis	219,439
11	itapetininga	496,109
12	macro_metropolitan a_paulista	2,518,376
13	vale_do_paraiba_p aulista	2,200,365

cod_mesoreg	mesoreg	pop_estim
14	litoral_sul_paulista	286,144
15	metropolitana_de_s ao_paulo	23,398,671

Fonte: Elaboração própria.

Possuindo a seguinte representação geográfica:

Figura 1 – Estado de São Paulo separado por Mesorregiões



Fonte: Elaboração própria.

Para o agrupamento, a estatística utilizada foi a mediana. Tal método foi selecionado buscando reduzir a influência de valores extremos atípicos, uma vez que a mediana é uma medida robusta menos sensível a outliers (TUKEY, 1977).

Assim, a estatística é calculada como:

$$\text{Valor Isolamento Região no Dia T} = \text{Mediana}(\text{Cidade 1 no Dia T}, \\ \text{Cidade 2 no Dia T}, \dots, \text{Cidade N no Dia T})$$

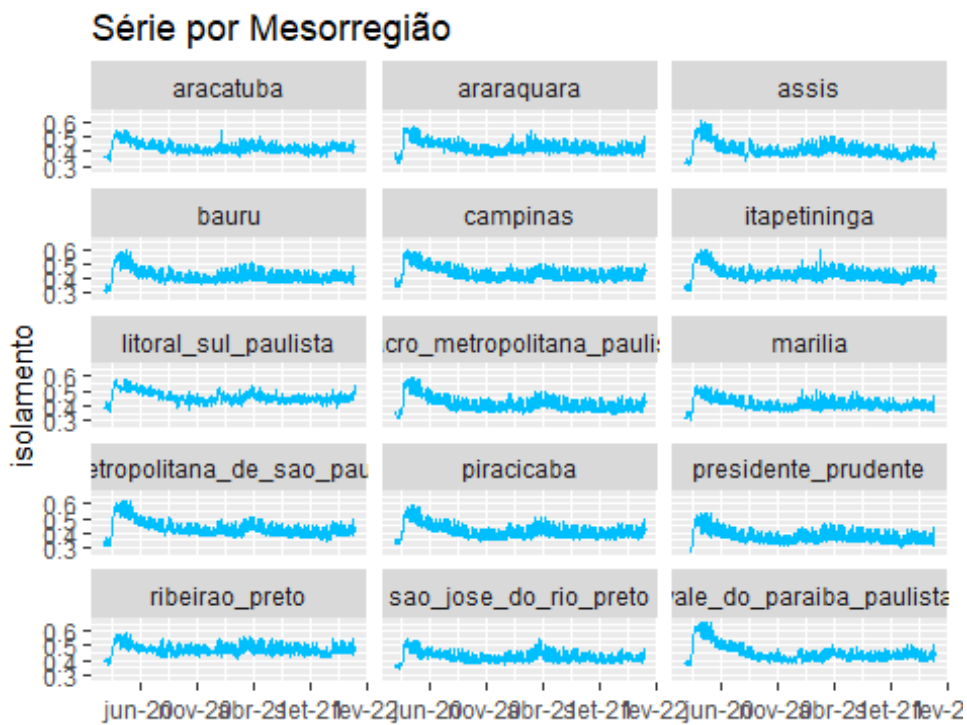
3 Descrição e Exploração dos Dados

Para a análise exploratória, foram utilizadas as bibliotecas:

- dplyr (limpeza e organização de dados)
- tidyr (limpeza e organização de dados)
- tseries (testes de hipótese)
- fpp3 (modelo STL)

Buscando observar o comportamento do isolamento no estado, o seguinte gráfico foi construído:

Gráfico 1 – Série por Mesorregião



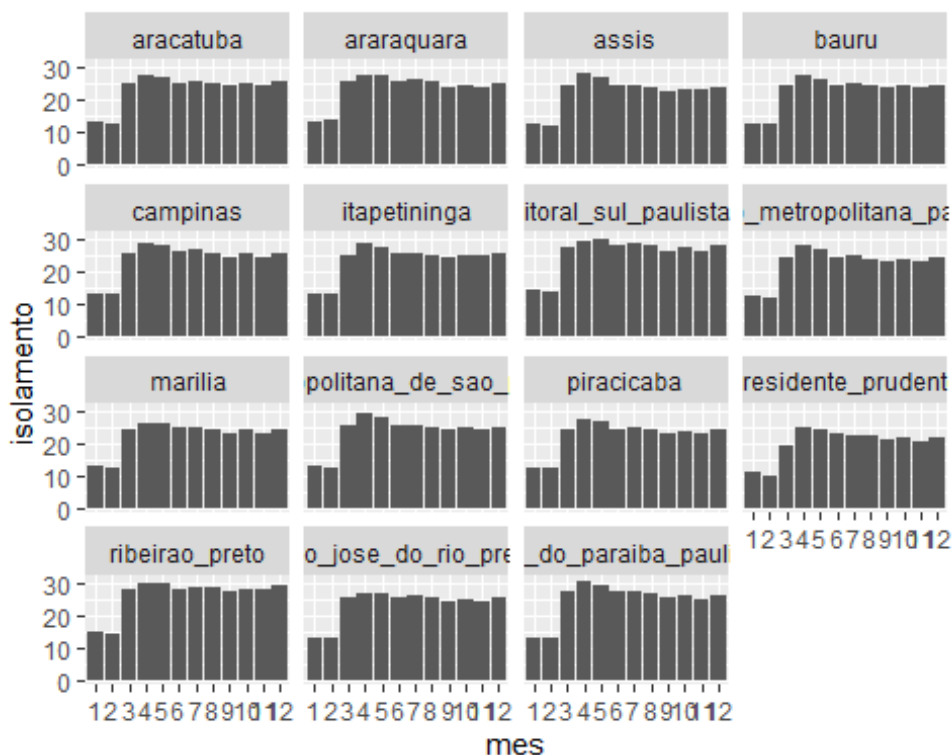
Fonte: Elaboração própria utilizando o software RStudio.

Vemos que todas as regiões apresentaram seus picos de isolamentos nos primeiros meses da pandemia. As regiões com maiores índices de isolamento foram Vale do Paraíba Paulista, Metropolitana de São Paulo e Assis.

Além disso, é possível observar um ponto de mudança nos valores de isolamento a partir emissão de estado pandêmico no Brasil no dia 11/03/2020, onde valores que até então se apresentavam próximos de 0.35, passaram a registrar números próximos a 0.6.

Os valores acumulados de isolamento por mês e por mesorregião podem ser observados a partir do seguinte gráfico:

Gráfico 2 – Acumulado de isolamento mensal/região



Fonte: Elaboração própria utilizando o software RStudio.

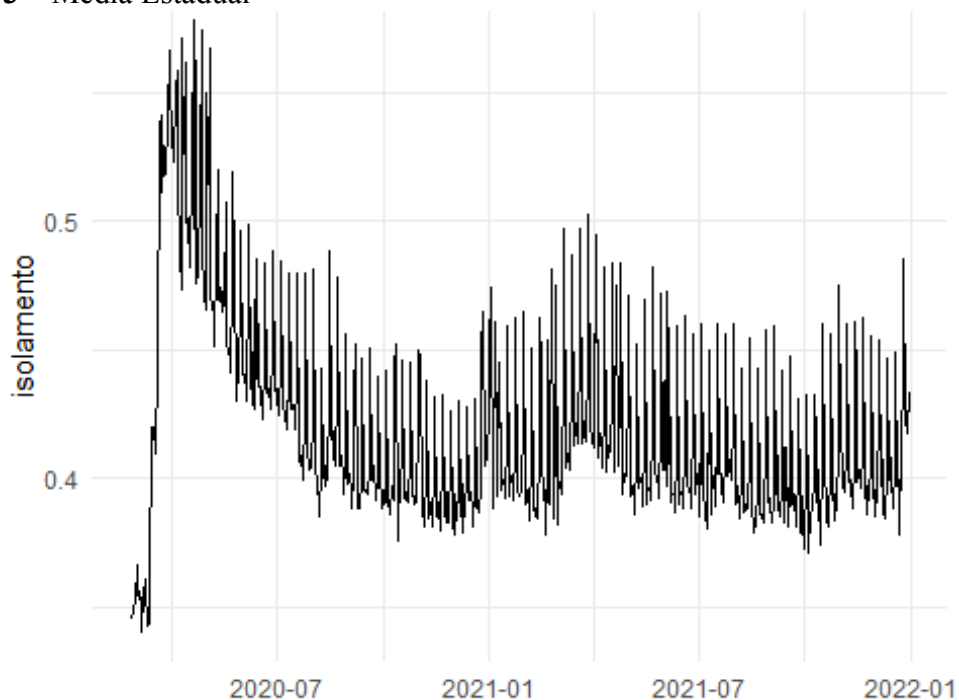
Como citado anteriormente, é notável um ponto de mudança no mês 3, onde esse mês marcou o início do lockdown no Brasil.

O decaimento no nível ao longo dos meses está associado a diversos fatores, como a flexibilização do lockdown, desgaste de população em se manter em casa e a busca por renda

como descrito na reportagem do IPEA - Lockdown. A partir do gráfico Série por Mesorregião é possível notar a presença de certos ciclos na série, podendo indicar a presença de sazonalidade.

A série da média estadual possui o seguinte comportamento:

Gráfico 3 – Média Estadual



Fonte: Elaboração própria utilizando o software RStudio.

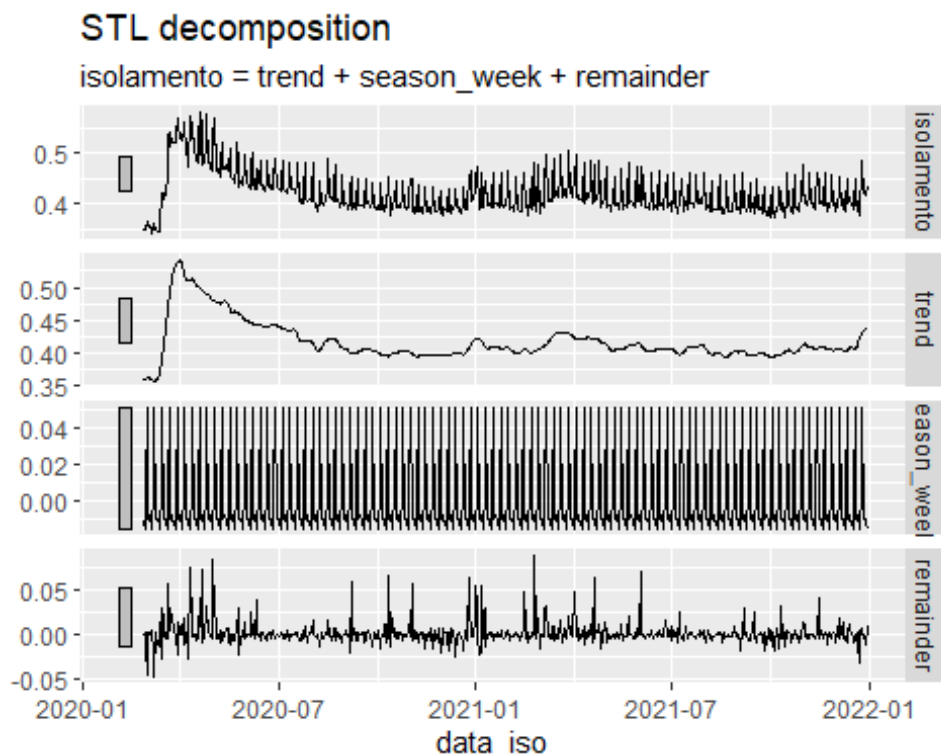
Os ciclos aparentam estar presentes dentro de 1 semana. Assim, o método de STL foi utilizado para melhor estudar a possível parcela sazonal dos dados.

STL é um método versátil e robusto para decomposição de séries temporais. O acrônimo refere-se à ‘Decomposição Sazonal e de Tendências usando Loess’, sendo que Loess é um método para estimar relações não lineares. O método STL foi desenvolvido por Cleveland et al. (1990) e permite uma decomposição flexível e robusta de séries temporais, sendo amplamente utilizado em análise de dados temporais.

O método se baseia em uma decomposição dos dados em três parcelas distintas: tendência, sazonalidade e ruído. A parcela de tendência busca modelar o comportamento geral de

crescimento e decrescimento dos dados, retirando a parcela sazonal presente. Já a parcela sazonal busca modelar o componente sazonal, removendo a tendência. O ruído é definido como a diferença entre o valor real e a soma das componentes de tendência e sazonalidade, ou seja, um resíduo do modelo de decomposição (MONTGOMERY; JENNINGS; KULAHCI, 2015).

Gráfico 4 – Gráfico Sazonal método STL

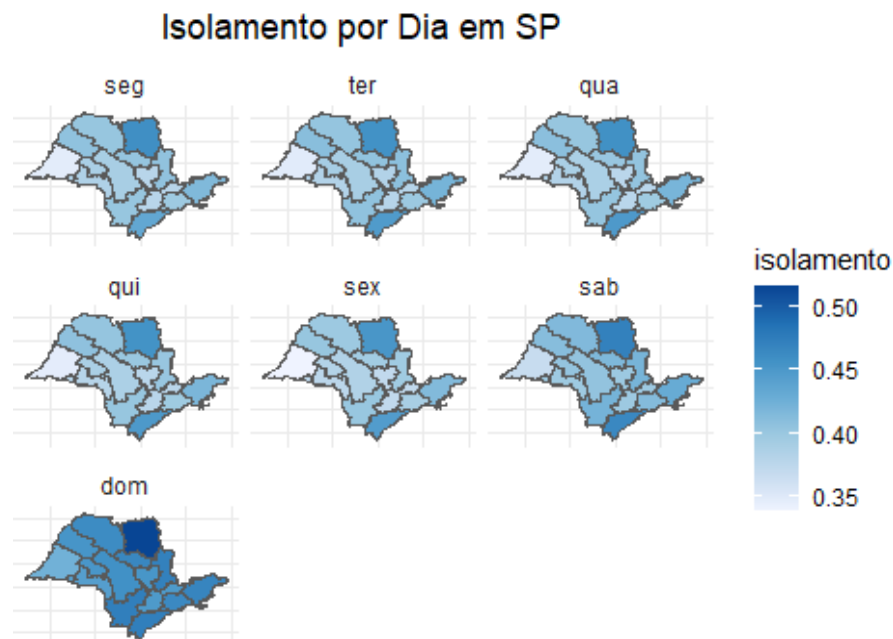


Fonte: Elaboração própria utilizando o software RStudio.

Via método STL definindo um período sazonal de 7 dias, vemos que a gráfico sazonal apresentou valores diferentes de 0, indicando que a sazonalidade realmente é semanal.

Confirmada a sazonalidade semanal, o seguinte gráfico foi construído:

Figura 2 – Isolamento por dia, separado em mesorregiões



Fonte: Elaboração própria utilizando o software RStudio.

O gráfico trás valores da mediana por dia da semana e por mesorregião no estado de SP. A partir dele vemos a sazonalidade indicada pela decomposição STL, onde um nível de isolamento maior ocorre aos domingos. Tal fato pode estar associado a jornada de trabalho semanal.

Para melhor investigar tal sazonalidade, dados de mobilidade urbana do Google Mobility foram utilizados.

4 Análise via Google Mobility

Os dados utilizados nesta seção foram retirados da plataforma de análise de dados epidemiológicos JFSalvandoTodos, onde os dados do Google Mobility já estão organizados e tratados em níveis municipais brasileiros (JFSALVANDOTODOS, 2024).

O mobility utiliza uma métrica própria de variação, calculada da seguinte forma:

- O valor base é a mediana do dia da semana correspondente, durante o período de cinco semanas de 3 de janeiro a 6 de fevereiro de 2020.

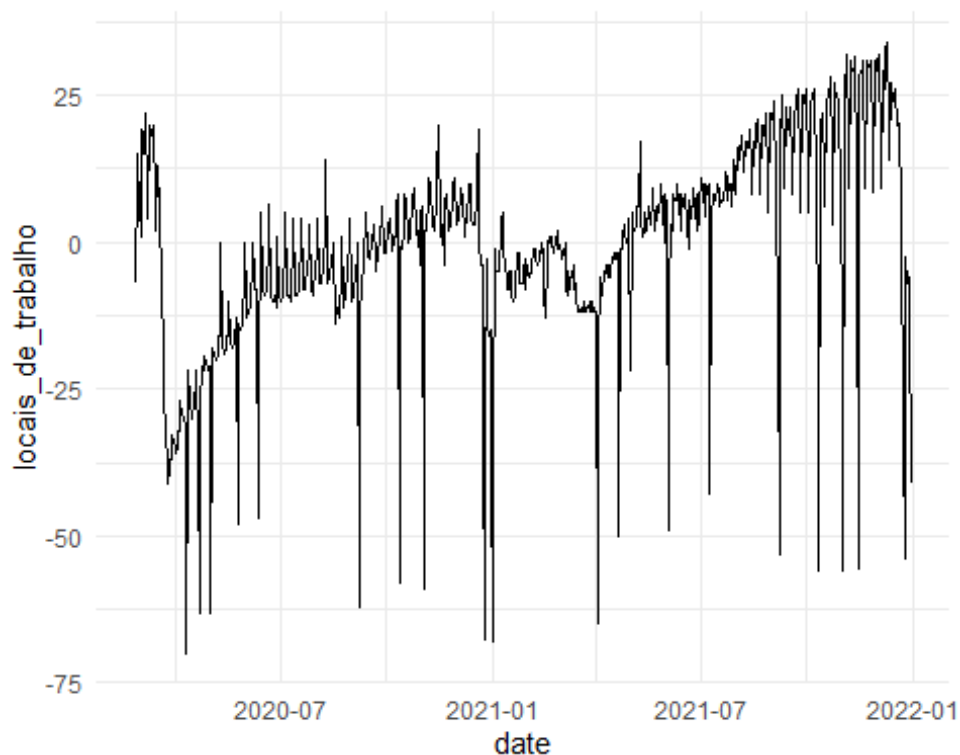
- Os conjuntos de dados mostram tendências ao longo de vários meses, e os dados mais recentes representam aproximadamente 2 a 3 dias atrás, que é o tempo necessário para produzir os conjuntos de dados.

Além disso, o google mobility disponibiliza 6 variáveis de estudo, sendo elas: Varejo e Recreação, Mercearia e Farmácia, Parques, Estações de Trânsito, Locais de Trabalho e Residencial.

Assim como a série de isolamento, os dados do mobility analisados vão de 26/02/2020 até 31/12/2021.

A variável de interesse é a de Locais de Trabalho, onde seus valores podem ser vistos através do seguinte gráfico:

Gráfico 5 – Gráfico da Série (Variável Locais de Trabalho)



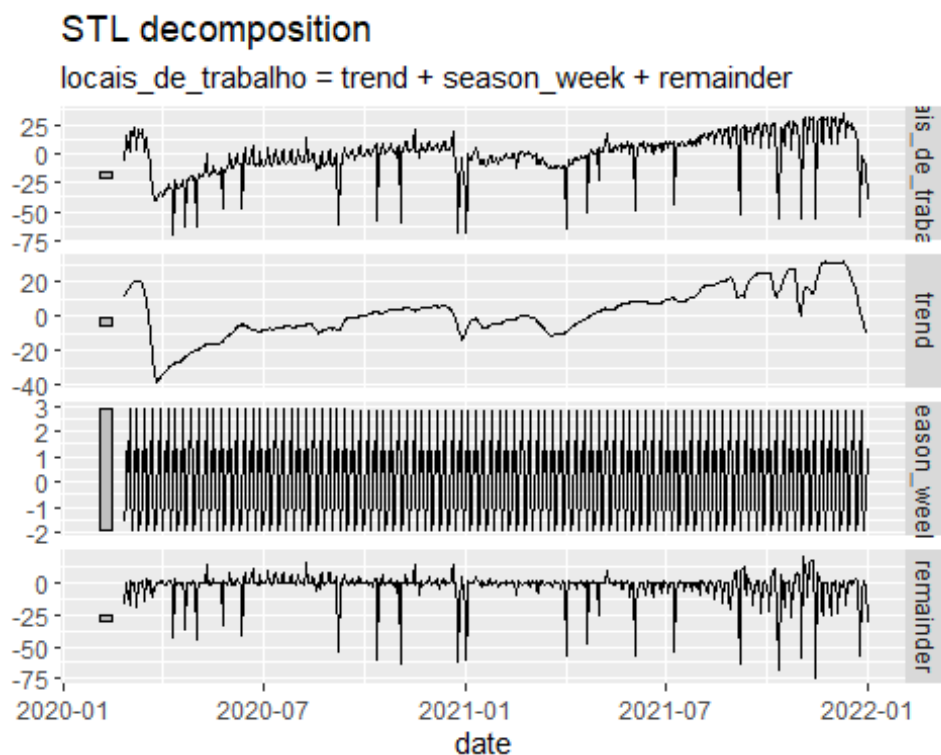
Fonte: Elaboração própria utilizando o software RStudio.

Através do gráfico da série, vemos uma queda significativa de mobilidade para locais de trabalho a partir do início da pandemia, onde tal mobilidade subiu gradativamente durante o ano de 2020.

Além disso, há a presença de ciclos nos dados estudados, onde o próximo tópico buscou estimar se havia sazonalidade em tais ciclos.

Para estimação da sazonalidade, novamente o método STL foi utilizado.

Gráfico 6 – Estimação por STL da variável Locais de Trabalho



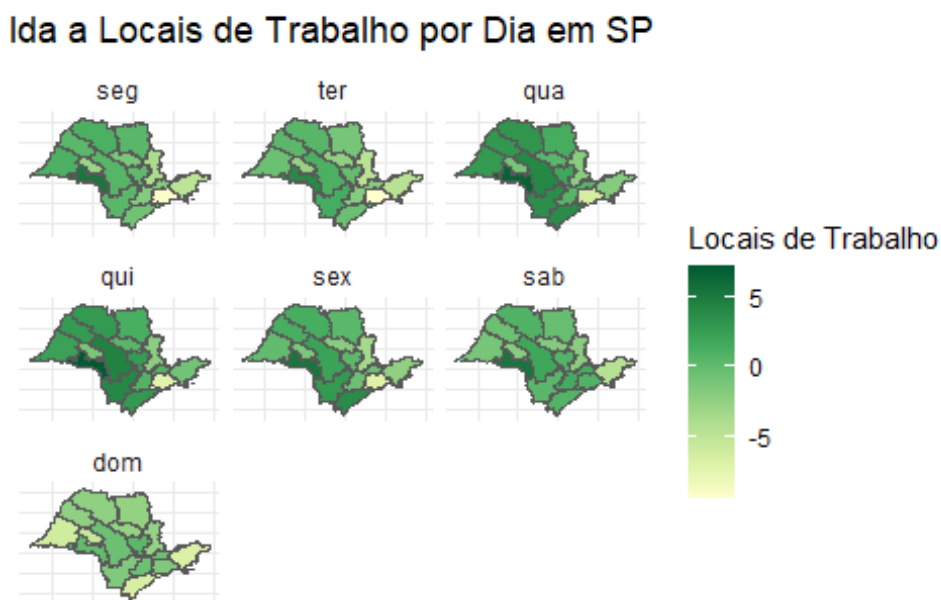
Fonte: Elaboração própria utilizando o software RStudio.

O método que busca dividir os dados em 3 parcelas distintas apontou na parcela de tendência uma queda significativa no início da pandemia, com um aumento gradual durante o ano de 2020. Ainda assim, o ano de 2021 apresentou outras 2 quedas fortes, tais quedas podem estar ligadas a feriados de fim de ano, picos de casos confirmados e maiores restrições de isolamento social.

Além disso, o método apontou novamente uma sazonalidade semanal, onde diferente dos dados de isolamento social que possuíam um pico todos os domingos, os dados de mobilidade a locais de trabalho indicam uma queda todos os domingos.

A partir do seguinte gráfico vemos como a mobilidade a locais de trabalho se distribui semanalmente:

Gráfico 7 – Ida a Locais de Trabalho por dia em SP



Fonte: Elaboração própria utilizando o software RStudio.

Assim como descrito anteriormente, vemos que os dados de variação de pessoas em locais de trabalho possuem uma sazonalidade semanal, onde a cada domingo, há uma queda de presença de pessoas nesses locais. De tal forma, a mobilidade a locais de trabalho se mostra como uma variável importante para o estudo do isolamento social, onde elas apresentam uma relação negativa entre si.

Agora, vamos seguir para a parte de modelagem dos dados de isolamento social no estado de São Paulo. O período escolhido será o mesmo, durante os anos pandêmicos de 2020 a 2022. A análise nesta modelagem utilizou uma abordagem diferente daquela utilizada na parte exploratória, onde buscou-se realizar uma análise separando SP em 15 microrregiões. Desta vez, o estado será analisado de maneira conjunta, tal abordagem buscou diminuir o número

finais de modelos a serem estimados. A diminuição de 15 para apenas 1 modelo final auxiliou também na comparação dos candidatos a melhor modelo, onde modelos com índices sazonais e com variáveis exógenas puderam ser detalhadamente estudados.

5 Análise Inicial da Série

Os dados coletados possuem índice temporal diário, ou seja, cada observação nessa série temporal representa valores para cada dia durante o período de análise. Além disso, originalmente os dados eram divididos por cidade, onde esses foram agrupados para uma análise estadual.

O conjunto de dados final teve a seguinte estrutura:

Tabela 2 – Índice médio de Isolamento Social diário

date	ind media iso
2020-02-26	0.3531727
2020-02-27	0.3527916
2020-02-28	0.3502064
2020-02-29	0.3573418
2020-03-01	0.3783417
2020-03-02	0.3578321

Fonte: Elaboração própria.

Possuindo as seguintes variáveis:

- date - Dia no qual os valores de isolamento social foram observados
- ind_media_iso - Índice Médio de isolamento social naquele dia

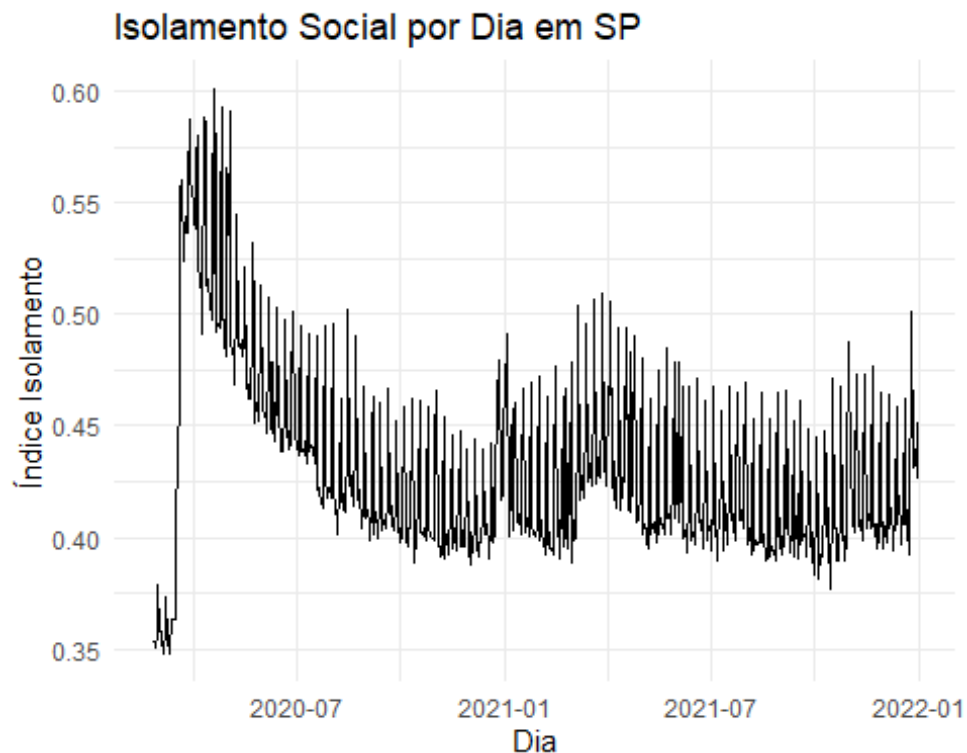
É importante destacar a mudança na variável final utilizada, enquanto os relatórios anteriores focaram na utilização da mediana como medida de tendência central, esse trabalho utilizou a média. Tal mudança foi ocasionada devido à similaridade dos modelos quando estimados utilizando como estatística para construção dos valores finais as funções de média e mediana. Dada tal similaridade, a decisão de se utilizar a média foi tomada por conta de ela ser

estimador de máxima verossimilhança para o parâmetro μ de uma distribuição Normal, possuindo as propriedades de consistência e eficiência.

6 Análise e Definição Iniciais dos Modelos

A série possui o seguinte comportamento quando plotada:

Gráfico 8 – Isolamento Social por Dia em SP

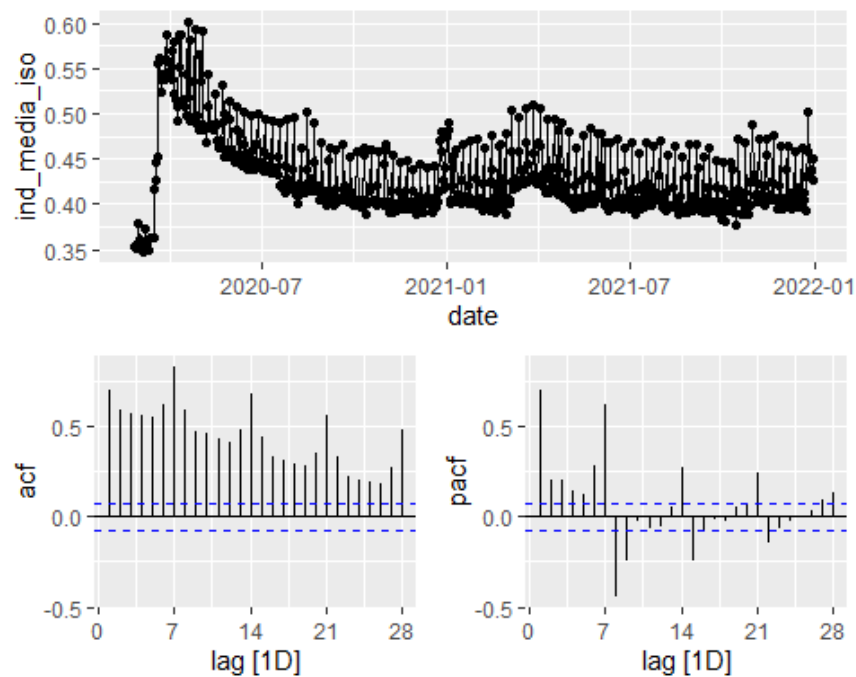


Fonte: Elaboração própria utilizando o software RStudio.

A partir desse simples gráfico podemos observar uma possível presença de sazonalidade, e uma possível tendência de decréscimo.

Ao construir os gráficos das funções de autocorrelação e autocorrelação parcial, o seguinte padrão foi visto:

Gráfico 9 - Funções de autocorrelação e autocorrelação parcial



Fonte: Elaboração própria utilizando o software RStudio.

Analisando o gráfico da função de autocorrelação parcial (ACF - Autocorrelation Function), é possível observar picos de correlação a cada 7 dias, indicando uma sazonalidade semanal. Além disso, os valores de correlação para todos os lags se mostraram significativos, sugerindo uma possível não-estacionariedade da série (BOX; JENKINS; REINSEL, 2015; HYNDMAN; ATHANASOPOULOS, 2018). A estacionariedade de uma série é um dos pressupostos para a aplicação de um modelo ARIMA, utilizado neste trabalho. Para verificar a estacionariedade dos dados, dois testes de hipótese foram empregados: o Teste de Dickey-Fuller Aumentado (ADF) e o Teste KPSS (DICKEY; FULLER, 1979; KWIATKOWSKI et al., 1992).

Primeiramente, o Teste de Dickey-Fuller Aumentado (ADF) foi aplicado. Esse teste tem como objetivo verificar a presença de uma raiz unitária na série temporal, testando a hipótese nula de não estacionariedade contra a hipótese alternativa de estacionariedade (DICKEY; FULLER, 1979). O teste possui as seguintes hipóteses:

H_0 :Série é não Estacionária

H_1 :Série é Estacionária

Tabela 3 - Teste ADF

Parameter1	Parameter2	z	df	p
Dados_Isolamento_SP		-3.57	7	0.036

Alternative hypothesis: stationary

Fonte: Elaboração própria utilizando o software RStudio.

A partir do teste aplicado, considerando um nível de confiança de 95%, a hipótese nula de raiz unitária é rejeitada, onde portanto é razoável assumir que a série trabalhada é estacionária.

Buscando maior precisão a respeito de tal resultado, um segundo teste foi aplicado. O Teste de Kwiatkowski–Phillips–Schmidt–Shin(KPPS) (possui hipóteses diferentes do primeiro aplicado, enquanto o teste ADF tinha a estacionariedade dos dados como hipótese alternativa, o teste KPPS traz como hipótese nula a série ser estacionaria em cima de uma tendência estacionária determinística contra a hipótese de raiz unitária (DICKEY; FULLER, 1979; KWIATKOWSKI et al., 1992).

Ao contrário da maioria dos testes de raiz unitária, a presença de uma raiz unitária não é a hipótese nula, mas a alternativa. Além disso, no teste KPSS, a ausência de uma raiz unitária não é uma prova de estacionariedade, mas, por definição, de estacionariedade de tendência (KWIATKOWSKI et al., 1992). Esta é uma distinção importante, uma vez que é possível que uma série temporal seja não estacionária, não tenha raiz unitária e ainda seja estacionária em termos de tendência.

Tanto nos processos de raiz unitária quanto nos de tendência estacionária, a média pode crescer ou diminuir ao longo do tempo. No entanto, na presença de um choque, os processos de tendência estacionária revertem à média, ou seja, são transitórios e tendem a convergir novamente para a média crescente, que não foi afetada pelo choque. Já os processos de raiz

unitária possuem um impacto permanente sobre a média, indicando que a série não apresenta convergência ao longo do tempo (HAMILTON, 1994; ENDERS, 2014).

Postulando as hipóteses, temos que:

H_0 :Série é Estacionária em Tendência

H_1 :Série não é Estacionária

Tabela 4 – Teste KPSS

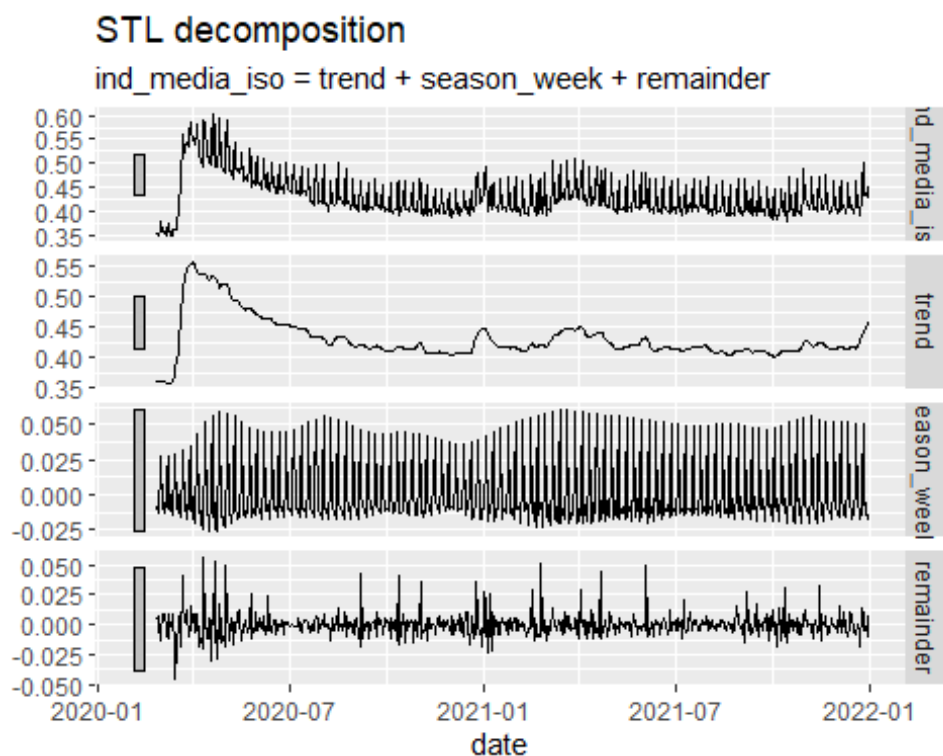
Parameter1	Parameter2	z	df	p
Dados_Isolamento_SP		2.37	6	0.010

Fonte: Elaboração própria utilizando o software RStudio.

Através da aplicação dos testes, os resultados obtidos foram parcialmente conflitantes. Enquanto o Teste ADF indicou que a série trabalhada era estacionária, o Teste KPSS rejeitou a hipótese de tendência estacionária, indicando uma série não estacionária (DICKEY; FULLER, 1979; KWIATKOWSKI et al., 1992).

Assim, para verificar se a série trabalhada é estacionária ou não, um método mais robusto foi utilizado: a Decomposição de Séries Temporais via STL. STL é um acrônimo para Seasonal and Trend Decomposition Using Loess, um método que se baseia na decomposição da série em três parcelas: tendência, sazonalidade e ruído. Loess, por sua vez, é um método de estimação de relações não lineares (CLEVELAND et al., 1990).

Gráfico 10– Estimação por STL da variável Isolamento Social

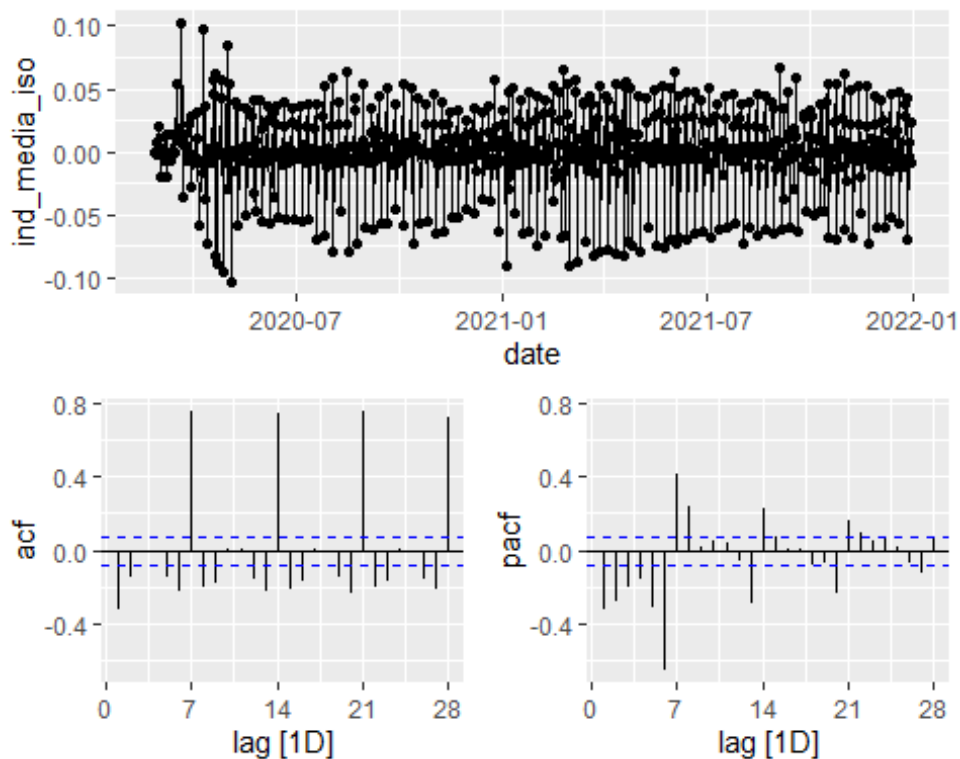


Fonte: Elaboração própria utilizando o software RStudio.

Via STL, podemos observar que a série apresentou mudança em sua tendência apenas no seu início, com uma rápida ascensão, e posteriormente manteve uma tendência com média constante ao longo dos dias analisados. O método também apontou uma sazonalidade semanal nos dados, conforme já indicado na análise exploratória (CLEVELAND et al., 1990).

Assim, apesar da série não apresentar uma tendência marcante em datas mais recentes, ambas as funções de autocorrelação (FAC e FACP) não exibiram um decaimento exponencial característico de séries estacionárias, indicando a necessidade de diferenciação (BOX; JENKINS; REINSEL, 2015). Dessa forma, buscando contornar a possível não estacionariedade da série trabalhada, o método de diferenciação de primeira ordem com lag igual a 1 foi aplicado (HYNDMAN; ATHANASOPOULOS, 2018). Após a aplicação da técnica, os dados apresentaram o seguinte comportamento:

Gráfico 11 – Método de diferenciação



Fonte: Elaboração própria utilizando o software RStudio.

A série diferenciada apresentou um comportamento desejado, onde é possível ver um decaimento exponencial na FACP e um truncamento de significância no lag 2 na FAC, indicando que um possível modelo MA(2) apresentaria um bom desempenho (BOX; JENKINS; REINSEL, 2015). Além disso, outra característica marcante dos dados é o pico de autocorrelação nos lags 7, 14, 21 e adiante, reforçando a presença de uma forte sazonalidade semanal (HYNDMAN; ATHANASOPOULOS, 2018).

Para concluir que a diferenciação foi corretamente aplicada aos dados, os testes ADF e KPSS foram novamente aplicados com uma significância de 5% (DICKEY; FULLER, 1979; KWIATKOWSKI et al., 1992).

O teste ADF obteve o seguinte resultado:

Tabela 5 – ADF aplicado após diferenciação

Parameter1	Parameter2	z	df	p
Dados_Isolamento_SP		-9.13	7	0.010

Alternative hypothesis: stationary

Fonte: Elaboração própria utilizando o software RStudio.

O teste teve um p-valor calculado igual a 0.01, sendo menor que a significância definida de 0.05, assim a hipótese nula de raiz unitária é rejeitada, onde é razoável supor a estacionariedade da série.

Já o teste KPSS obteve o seguinte resultado:

Tabela 6 – KPSS aplicado após diferenciação

Parameter1	Parameter2	z	df	p
Dados_Isolamento_SP		0.08	6	0.100

Fonte: Elaboração própria utilizando o software RStudio.

O teste teve um p-valor calculado igual a 0.1, sendo maior que a significância definida de 0.05, assim a hipótese nula da série ser estacionária em tendência não é rejeitada.

Portanto, dado a conclusão de ambos os testes, é razoável assumir a estacionariedade da série diferenciada e, portanto, utilizaremos o parâmetro $d = 1$ de um modelo ARIMA, indicando que uma diferenciação foi aplicada aos dados.

7 Modelagem

A partir das análises descritiva e exploratória realizadas nos dados, foi observada uma forte sazonalidade semanal, além da necessidade da aplicação de uma diferenciação de primeira ordem para tornar a série estacionária. Modelos da família SARIMA são amplamente utilizados para lidar com esse tipo de comportamento, permitindo a modelagem da dependência temporal e sazonal em séries temporais (BOX; JENKINS; REINSEL, 2015).

Para a modelagem, foram utilizadas as bibliotecas:

- fable(modelo Sarima)
- tseries (testes de hipóteses)
- fabletools (gráficos temporais)

Assim, o modelo base a ser utilizado deve comportar esses dois aspectos: sazonalidade e diferenciação. Para isso, o modelo ARIMA sazonal, chamado de SARIMA, foi selecionado. Tal modelo é uma extensão do ARIMA, incluindo uma parcela sazonal para lidar com padrões periódicos nos dados (BOX; JENKINS; REINSEL, 2015).

Assim como outros modelos de séries temporais, o SARIMA é representado por seus parâmetros, onde no caso do ARIMA, os parâmetros são: Autorregressivo (p), Diferenciação (d) e Médias Móveis (q), formando assim um modelo ARIMA(p,d,q) (HYNDMAN; ATHANASOPOULOS, 2018). Esses parâmetros indicam a estrutura do modelo:

(i) **Autorregressivo (AR)**: Representa a relação linear entre uma observação e um número fixo de observações passadas, capturando a influência das próprias observações anteriores no valor atual;

(ii) **Integrado (I)**: Indica o número de diferenciações necessárias para tornar a série estacionária;

(iii) **Médias Móveis (MA)**: Modela a relação entre o erro atual e os erros passados, permitindo capturar padrões que não foram explicados pela componente autorregressiva (BROCKWELL; DAVIS, 2016).

Juntas, essas parcelas fornecem uma estrutura flexível para modelar uma ampla gama de séries temporais.

O modelo tem a seguinte estrutura básica:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + e_t$$

onde y_t é o valor da série no tempo t , ϕ é o termo autorregressivo possuindo p elementos, e_t é o resíduo do modelo do tempo t e θ é o termo de médias móveis possuindo q elementos (BOX; JENKINS; REINSEL, 2015).

O modelo pode ser reescrito em forma de somatório, facilitando sua escrita.

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t$$

Já um modelo SARIMA comporta a adição de mais uma parcela autorregressiva e de médias móveis, buscando modelar a parcela sazonal. A representação das ordens sazonais é acompanhada do modelo, onde é escrito como SARIMA(p, d, q)(P, D, Q) $_{[m]}$, onde (p, d, q) representam a ordem não sazonal e (P, D, Q) representam a ordem sazonal do modelo (BOX; JENKINS; REINSEL, 2015). Ele pode ser escrito como:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + \sum_{i=1}^P \Phi_i y_{t-m-(i-1)} + \sum_{i=1}^Q \Theta_i e_{t-m-(i-1)} + e_t$$

onde Φ é o termo autorregressivo sazonal possuindo P elementos, Θ é o termo de médias móveis sazonal possuindo Q elementos e m representa o índice de sazonalidade. Para uma série diária possuindo sazonalidade semanal, assim como os dados trabalhados nesse estudo, deve-se utilizar o índice $m = 7$ (HYNDMAN; ATHANASOPOULOS, 2018).

Nesse contexto Hyndman e Khandakar (2008) sugerem uma maneira automática de seleção de ordens de parâmetros via critérios de informação: AIC ou BIC.

Critérios de informação são ferramentas estatísticas usadas para comparar e selecionar modelos que melhor descrevem um conjunto de dados. Os critérios de Akaike (AIC) e Bayesiano (BIC) são amplamente utilizados na seleção de modelos estatísticos, permitindo avaliar a qualidade relativa dos modelos ajustados e balancear a complexidade e o ajuste aos dados (AKAIKE, 1974; SCHWARZ, 1978).

Tanto o AIC quanto o BIC são usados para comparar diferentes modelos ajustados aos mesmos dados. O modelo com o menor valor de AIC ou BIC é considerado o melhor. Ambos

os critérios penalizam a inclusão de mais parâmetros. O BIC tem uma penalização maior que o AIC quando o número de observações é grande.

Para definição das ordens, tal método foi utilizado. 2 modelos iniciais foram construídos, um para cada métrica de critério de informação, onde o método Stepwise foi utilizado para seleção dos parâmetros (p, d, q) e (P, D, Q) . O método stepwise é uma abordagem sistemática para a seleção de ordens de parâmetros em modelos de séries temporais, como os modelos SARIMA. Este método envolve a adição e remoção iterativa de parâmetros do modelo com base nos critérios de informação, com o objetivo de identificar o modelo que melhor descreve os dados com a menor complexidade possível. (HYNDMAN; KHANDAKAR, 2008).

O modelo que minimizou o AIC via método stepwise foi um SARIMA(1, 0, 1)(0, 1, 1)₇. Já o modelo que minimizou o BIC via método stepwise foi um SARIMA (1, 0, 1)(0, 1, 1)₇, ou seja, o mesmo modelo minimizou o AIC e BIC. Um ponto interessante é que o algoritmo utilizado não aplicou uma diferenciação aos dados, onde tivemos o parâmetro $d = 0$. Foi visto via análise exploratória que a aplicação de uma diferenciação seria bem-vinda aos dados, tornando-os estacionários (BOX; JENKINS; REINSEL, 2015), e assim um terceiro modelo foi construído, sendo um SARIMA (1, 1, 1) (0, 1, 1)₇.

O modelo estimado teve um AIC e um BIC menor que os modelos construídos via stepwise (HYNDMAN; KHANDAKAR, 2008). Tal descoberta indica que o algoritmo em si não selecionou o melhor modelo, mas apenas indicou uma direção de onde este deve estar. Assim, diversos outros modelos próximos a um SARIMA (1, 1, 1)(0, 1, 1)₇ foram construídos e analisados, buscando aquele que apresentasse o menor valor de AIC e BIC, sendo eles:

- SARIMA (2, 1, 1)(0, 1, 1)₇
- SARIMA (1, 1, 1)(1, 1, 1)₇
- SARIMA (2, 1, 1)(2, 1, 1)₇

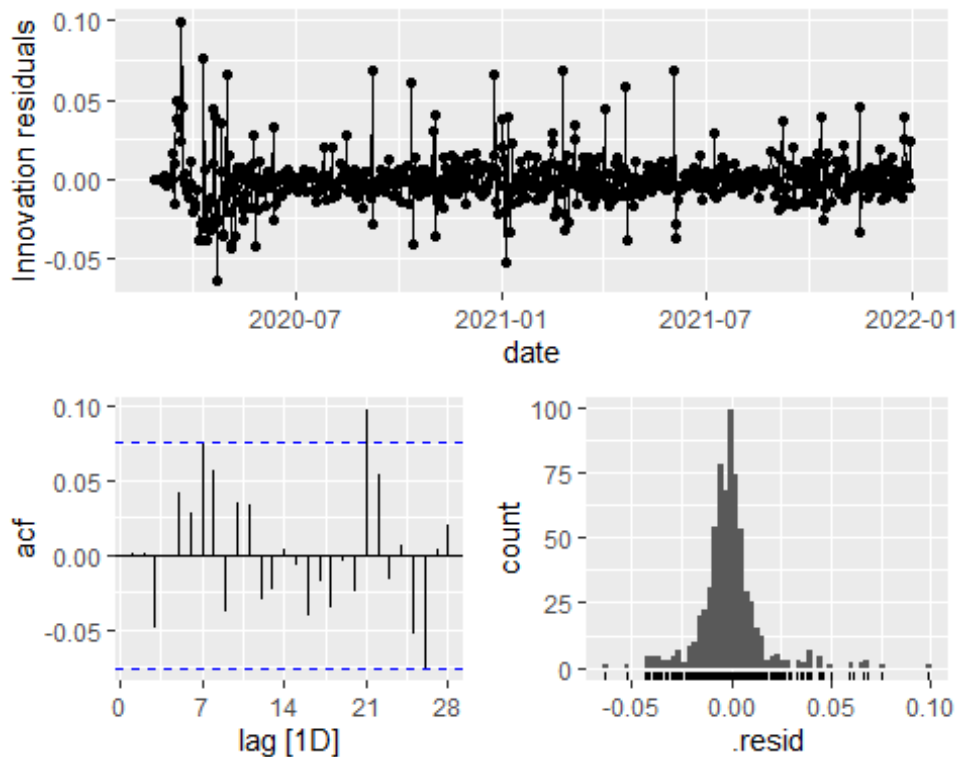
Dado os valores de AIC e BIC calculados para cada modelos, vemos que aquela a melhor minimizar tais métrica foi o SARIMA(2, 1, 1)(0, 1, 1)₇.

8 Análise de Resíduos

Apesar do modelo SARIMA(2,1,1)(0,1,1)₇ ter sido aquele a apresentar melhores valores de critério de informação, isso não garante que seja o melhor modelo se ele violar as suposições básicas de um modelo SARIMA. Como destacado na literatura, além de minimizar critérios como AIC e BIC, é essencial que o modelo respeite as suposições fundamentais para que suas previsões sejam confiáveis (BURNHAM; ANDERSON, 2002). Essas suposições incluem:

- Os resíduos devem ser normalmente distribuídos;
- Os resíduos devem ser não correlacionados (BOX; JENKINS; REINSEL, 2015; HYNDMAN; ATHANASOPOULOS, 2018).

Gráfico 12 - Resíduos do Modelo



Fonte: Elaboração própria utilizando o software RStudio.

Analisando o gráfico de autocorrelação dos resíduos, é possível ver que esse não possui muitos lags significativos, indicando que os resíduos são não correlacionados. Para testar tal

hipótese, o teste de Ljung-Box foi aplicado (BOX; JENKINS; REINSEL, 2015), possuindo a seguintes hipóteses:

H_0 :Resíduos são não-correlacionados

H_1 :Resíduos são correlacionados

Tabela 7 - Aplicação do Teste Ljung-Box

Parameter1	Parameter2	z	df	p
Resíduos		11.31	10	0.334

Fonte: Elaboração própria utilizando o software RStudio.

A partir do teste aplicado, considerando um nível de confiança de 95%, a hipótese nula não foi rejeitada, sendo razoável assumir que os resíduos são não correlacionados (LJUNG; BOX, 1978). Já para a distribuição, o histograma mostra claramente um comportamento próximo a uma distribuição Normal, sem evidências de assimetria ou curtose que fujam da normalidade (SHAPIRO; WILK, 1965).

Portanto, o modelo construído, SARIMA(2,1,1)(0,1,1)₇, é válido, pois esse não quebrou nenhuma suposição do modelo SARIMA, como estacionariedade, ausência de autocorrelação nos resíduos e normalidade (BOX; JENKINS; REINSEL, 2015; HYNDMAN; ATHANASOPOULOS, 2018).

9 Análise do Modelo Final e Conclusão

O modelo final selecionado foi um SARIMA(2,1,1)(0,1,1)₇, possuindo as seguintes estimativas autorregressivas e de médias móveis:

Tabela 8 – Coeficientes Estimados

term	estimate	std.error	statistic	p.value
ar1	0.09	0.10	0.87	0.38

term	estimate	std.error	statistic	p.value
ar2	0.06	0.06	0.92	0.36
ma1	-0.59	0.09	-6.70	0.00
sma1	-0.91	0.03	-27.45	0.00

Fonte: Elaboração própria utilizando o software RStudio.

Analisando os coeficientes estimados, vemos que os dois termos autorregressivos não foram significativos, enquanto os termos de média móvel sazonal e não sazonal foram. Isso é consistente com a literatura, que destaca a importância da escolha criteriosa de termos significativos para evitar modelos superparametrizados e pouco interpretáveis (BOX; JENKINS; REINSEL, 2015).

O modelo final possui a seguinte equação:

$$y_t = 0.09y_{t-1} + 0.06y_{t-2} - 0.59e_{t-1} - 0.91e_{t-7} + e_t$$

Apesar dos termos autorregressivos não terem se mostrado significativos, podemos interpretá-los da mesma maneira que os termos significativos de média móvel (HYNDMAN; ATHANASOPOULOS, 2018).

Dependência Autorregressiva: O valor atual y_t é influenciado em 9% pelo valor da série no período imediatamente anterior y_{t-1} e 6% pelo valor da série dois períodos atrás y_{t-2} . Isso sugere que os valores recentes da série temporal têm um impacto positivo em nível moderado sobre o valor atual.

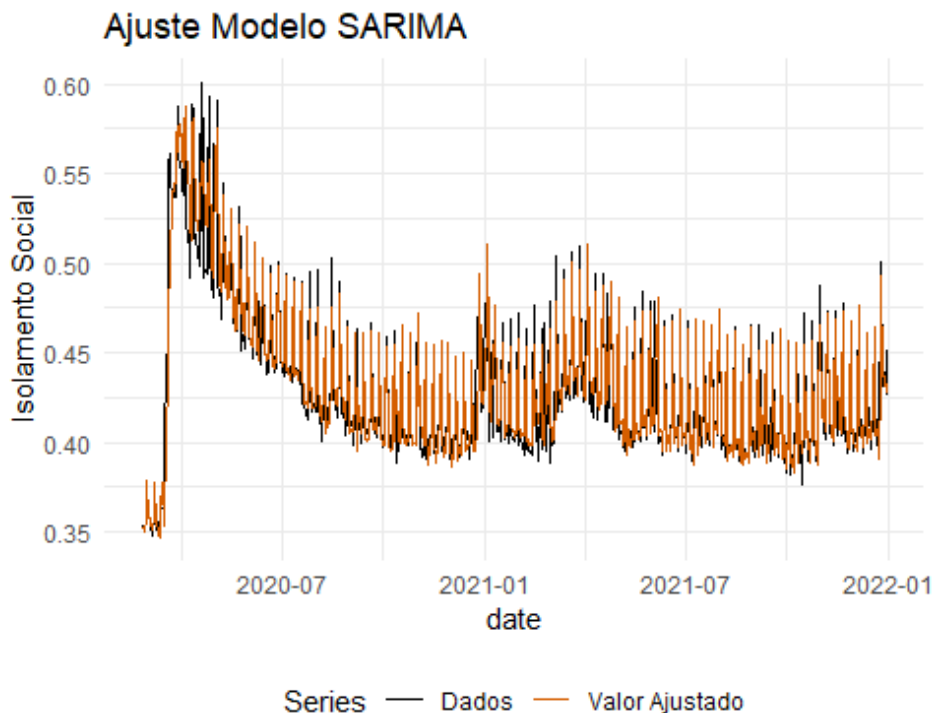
Dependência de Média Móvel: O valor atual y_t é influenciado negativamente em 59% pelo erro no período imediatamente anterior e_{t-1} e em 91% pelo erro sete períodos atrás e_{t-7} . O impacto negativo significativo do erro passado sugere que os desvios anteriores têm um efeito corretivo forte sobre o valor atual.

Portanto, este modelo SARIMA sugere que a série temporal tem uma componente autorregressiva de baixa ordem, uma componente de média móvel que inclui tanto um termo de curto prazo quanto um termo sazonal semanal, e um componente de erro aleatório.

Esses números nos contam como o modelo está estruturado. Notamos que os termos autorregressivos AR(1) e AR(2) acabaram não sendo estatisticamente diferentes de zero, o que implica que, do ponto de vista do modelo, incluir até o segundo dia passado não trouxe muita melhora no ajuste comparado a não os incluir. Já os termos de média móvel, especialmente o sazonal, foram cruciais para o ajuste. Isso faz sentido, pois a série tem muita aleatoriedade diária que é melhor captada como “choques” (termos MA) do que como persistência determinística (AR), e principalmente tem aquele comportamento semanal que é captado pelo erro de 7 dias atrás (LJUNG; BOX, 1978).

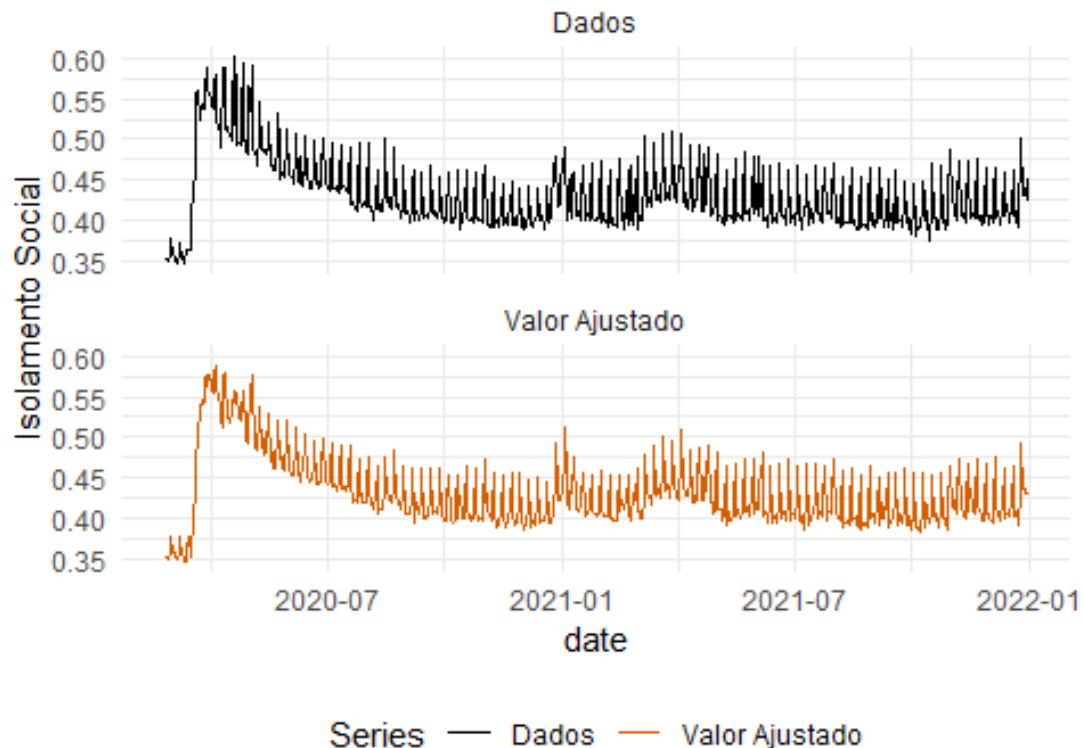
O modelo teve o seguinte desempenho nos dados observados:

Gráficos 13 e 14 – Gráfico Ajustado Modelo SARIMA



Fonte: Elaboração própria utilizando o software RStudio.

Ajuste Modelo SARIMA



Fonte: Elaboração própria utilizando o software RStudio.

Analisando os gráficos, vemos que o modelo apresentou ótimo ajuste aos dados, captando a tendência de crescimento no começo de 2020 e a sazonalidade semanal dos dados. Assim, o modelo construído é válido e teve bom ajuste aos dados. Todos os pressupostos de um modelo SARIMA foram atendidos: estacionariedade (após as diferenciações), resíduos independentes e aproximadamente normais, e parâmetros estatisticamente significativos nos componentes essenciais (BOX; JENKINS; REINSEL, 2015). Dessa forma, do ponto de vista de diagnóstico pós-ajuste, não há indicações de que devêssemos tentar um modelo mais complexo ou diferente – o SARIMA selecionado cumpre bem o papel de explicar a série (HYNDMAN; ATHANASOPOULOS, 2018). Este conjunto de verificações dá respaldo para utilizarmos o modelo tanto para compreender a série historicamente (nosso foco aqui) quanto, futuramente, para possíveis previsões, sabendo que ele se baseia em premissas satisfeitas pelos dados.

Referências

CLEVELAND, R. B.; CLEVELAND, W. S.; McRAE, J. E.; TERPENNING, I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, v. 6, n. 1, p. 3-73, 1990.

Montgomery, D.C., Jennings, C.L. e Kulahci, M. *Introduction to Time Series Analysis and Forecasting*. Nova York: Wiley. (seção 5.9)

Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1), 134–141. [DOI]

Hyndman, R. J., & Fan, S. (2010). Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, 25(2), 1142–1153. [DOI]

Engle, R. and Granger, C. (1987) Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica*, 55, 251-276.

Granger, C.W.J. and Weiss, A.A. (1983) Time Series Analysis of Error-Correcting Models. In: Karlin, S., Amemiya, T. and Goodman, L.A., Eds., *Studies in Econometrics, Time Series, and Multivariate Statistics*, Academic Press, New York, 255-278.

BEZERRA, Anselmo César Vasconcelos et al. Fatores associados ao comportamento da população durante o isolamento social na pandemia de COVID-19. *Ciência & Saúde Coletiva* [online]. v. 25, suppl 1, p. 2411-2421, 2020. Disponível em: <https://www.scielo.br/j/csc/a/NLmBdNTqPzJmt6Ft4gYqvVH/?lang=pt>. Acesso em: 3 jan. 2024.

NIU, Y.; XU, F. Deciphering the power of isolation in controlling Covid-19 outbreaks. *Lancet*, v. 8, n. 4, p. 452-453, 2020.

SILVA, F. C. DA et al. Social isolation and the speed of covid-19 cases: measures to prevent transmission. *Revista Gaúcha de Enfermagem*, v. 42, n. spe, 2021. Disponível em: <https://seer.ufrgs.br/RGENF/article/view/110765>. Acesso em: 3 jan. 2024.

IPEA - Instituto de Pesquisa Econômica Aplicada. Regras de distanciamento não conseguiram conter queda do nível de isolamento social. 2021. Disponível em: <https://www.ipea.gov.br/portal/categorias/45-todas-as-noticias/noticias/2188-regras-de-distanciamento-nao-conseguiram-conter-queda-do-nivel-de-isolamento-social>. Acesso em: 3 jan. 2024.

JFSALVANDOTODOS. Plataforma de análise de dados epidemiológicos. Universidade Federal de Juiz de Fora (UFJF), 2024. Disponível em: <http://jfsalvandotodos.ufjf.br/>. Acesso em: 8 jan. 2024

<https://www.saopaulo.sp.gov.br/coronavirus/isolamento/>. Acesso em: 11 dez. 2023.

TUKEY, J. W. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, v. 27, n. 3, p. 1–22, 2008. Disponível em: <https://www.jstatsoft.org/article/view/v027i03>. Acesso em: 17 mar. 2024.

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716-723, 1974.

SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, v. 6, n. 2, p. 461-464, 1978.

DICKEY, D. A.; FULLER, W. A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, v. 74, n. 366a, p. 427-431, 1979.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2. ed. Melbourne: OTexts, 2018. Disponível em: <https://otexts.com/fpp2/>. Acesso em: 14 out. 2024.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. *Time Series Analysis: Forecasting and Control*. 5. ed. Hoboken: Wiley, 2015.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, v. 65, n. 2, p. 297-303, 1978.

BURNHAM, K. P.; ANDERSON, D. R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2. ed. New York: Springer, 2002.

KWIATKOWSKI, D.; PHILLIPS, P. C. B.; SCHMIDT, P.; SHIN, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, v. 54, n. 1-3, p. 159-178, 1992.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, v. 52, n. 3-4, p. 591-611, 1965.

ENDERS, W. Applied Econometric Time Series. 4. ed. Hoboken: Wiley, 2014.

STOCK, J. H.; WATSON, M. W. Introduction to Econometrics. 3. ed. Boston: Pearson, 2015.

HAMILTON, J. D. Time Series Analysis. Princeton: Princeton University Press, 1994.

BROCKWELL, P. J.; DAVIS, R. A. *Introduction to Time Series and Forecasting*. 3. ed. New York: Springer, 2016.