

Previsão das Síndrome Respiratorias Agudas Graves (SRAG)

As SRAGs

A Síndrome Respiratória Aguda Grave (SRAG) é uma doença respiratória viral causada por diversos tipos de vírus. Foi relatado pela primeira vez no sul da China em 2002 e rapidamente se espalhou globalmente. A SRAG é caracterizada por sintomas como febre alta, tosse e falta de ar.

Dentre as SRAGs, a mais recente e com maior índice de disseminação dos últimos 3 anos foi a Covid-19. Sendo primeira vez datada ao final de 2019 na China, o coronavírus rapidamente se espalhou pelo mundo.

O surto de COVID-19 no Brasil tem sido um dos maiores e mais complexos do mundo. Em 2021, o Brasil registrou mais de 15 milhões de casos confirmados e mais de 420.000 mortes, tornando-se um dos países com o maior número de casos e mortes globalmente. A disseminação do COVID-19 no Brasil foi atribuída a vários fatores, incluindo alta densidade populacional do país, baixas taxas de vacinação e acesso limitado aos cuidados de saúde.

A previsão das SRAGs são importante para uma tomada de decisão eficaz. Envolve a análise de dados e o uso de modelos para prever tendências futuras, fornecendo informações valiosas para alocação de recursos e medidas de saúde pública. A previsão também ajuda no planejamento para futuras pandemias e ajuda a abordar as preocupações do público. Investir e melhorar as capacidades de previsão é crucial para uma melhor preparação para os desafios futuros.

Assim esse documento busca retratar o passo a passo para a construção do modelo de previsão utilizado na Plataforma JFST (O modelo final, assim como suas estimativas e erros se encontram na parte final do documento)

Dados utilizados

Os dados utilizados são disponibilizados pelo Ministério da Saúde através do seguinte link:

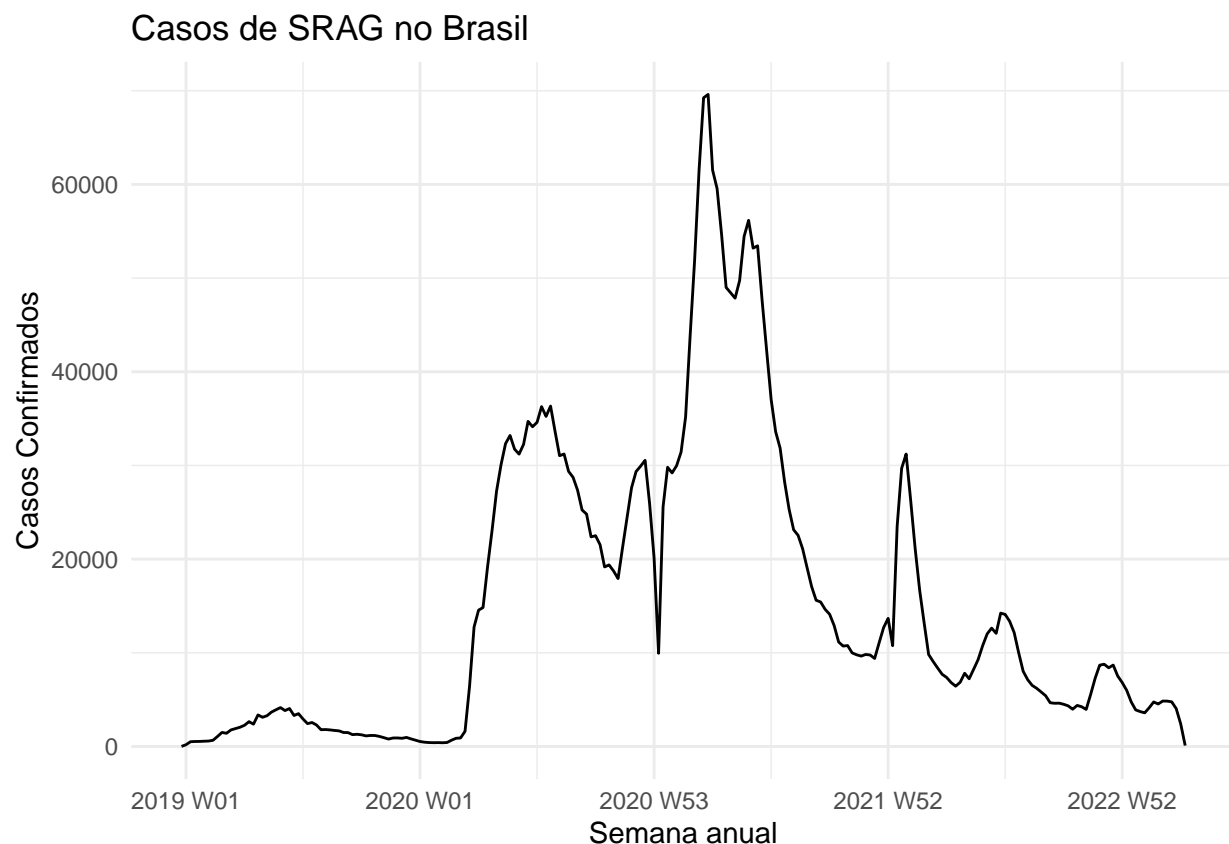
<https://opendatasus.saude.gov.br/gl/dataset/srag-2021-a-2023>

E se referem aos casos mais graves das SRAGs

E a análise desses dados em forma de gráficos já está disponível na plataforma <http://jfsalvandotodos.ufjf.br/#!/pesquisar/SRAG>

O modelo construído busca prever números de casos confirmados. Os dados utilizados possuem intervalo de Janeiro de 2019 a Julho de 2023

A seguir, é apresentado um gráfico simples de número de casos confirmados por semana.



Nota-se que se trata de uma série com um período de análise curto, de possível tendência de decrescimento sem uma sazonalidade marcante.

O Efeito Covid

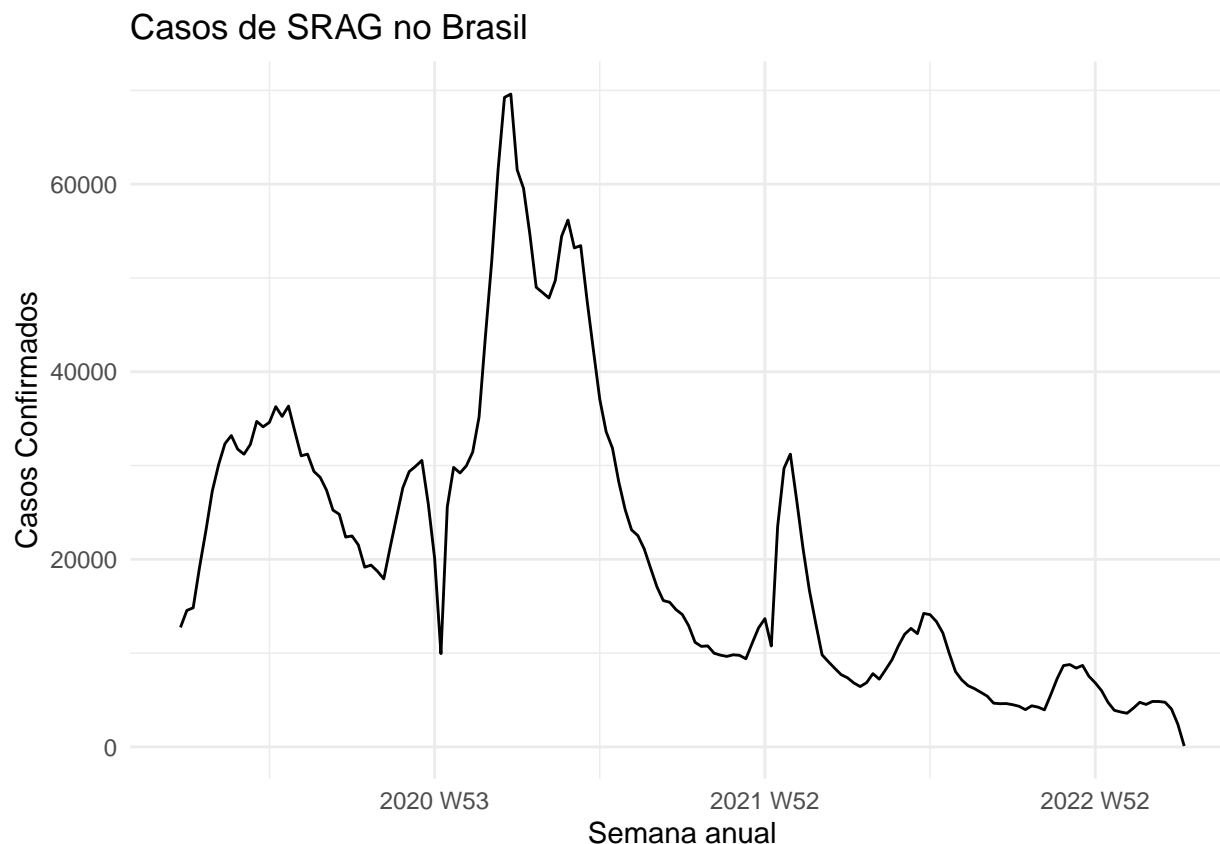
Assim como descrito em: Niquini et al. (2020), ao observar os dados é possível notar um comportamento antes, durante e após a pandemia de Covid-19.

Até o início de 2020, poucos casos graves de SRAGs eram registrados. Com o início da disseminação do vírus da Covid-19 no Brasil, em março de 2020,

rapidamente o número de casos confirmados semanais ultrapassaram a marca de 30.000 registros

Assim, buscando um melhor comportamento da previsão, os dados utilizados para treinamento do modelo tem data iniciada em março de 2020

O gráfico de casos confirmados por semana após esse corte é dado por:



Construção de Modelo

O próximo tópico teve como objetivo a construção de um modelo de previsão para casos confirmados semanais de SRAGS graves.

Para isso, utilizou a base teórica estatística de séries temporais. Construiu-se o seguinte diagrama para facilitar o entendimento do passo a passo da modelagem descrito no trabalho

Teste de Tendencia e Ciclos

Para a construção é necessária a análise de Tendencia e Sazonalidade da serie temporal

Tendência: Refere-se à direção geral ou padrão de crescimento ou decréscimo ao longo do tempo. É uma componente da série temporal que mostra a mudança sistemática ao longo do tempo, independentemente das flutuações aleatórias e sazonalidades presentes

Ciclo: Referem-se a padrões de oscilação ou flutuação que ocorrem em um período não fixado, geralmente com duração de 2 anos. Esses ciclos podem ser causados por fatores econômicos, sociais ou naturais que se repetem ao longo do tempo.

Para se testar a hipótese da tendência, utilizou-se o **Teste de Dickey-Fuller Aumentado** (Mushtaq 2011) com 5% de significância, onde tem-se as seguintes hipóteses:

H_0 : Tendência não estacionária ; H_1 : Tendência estacionária

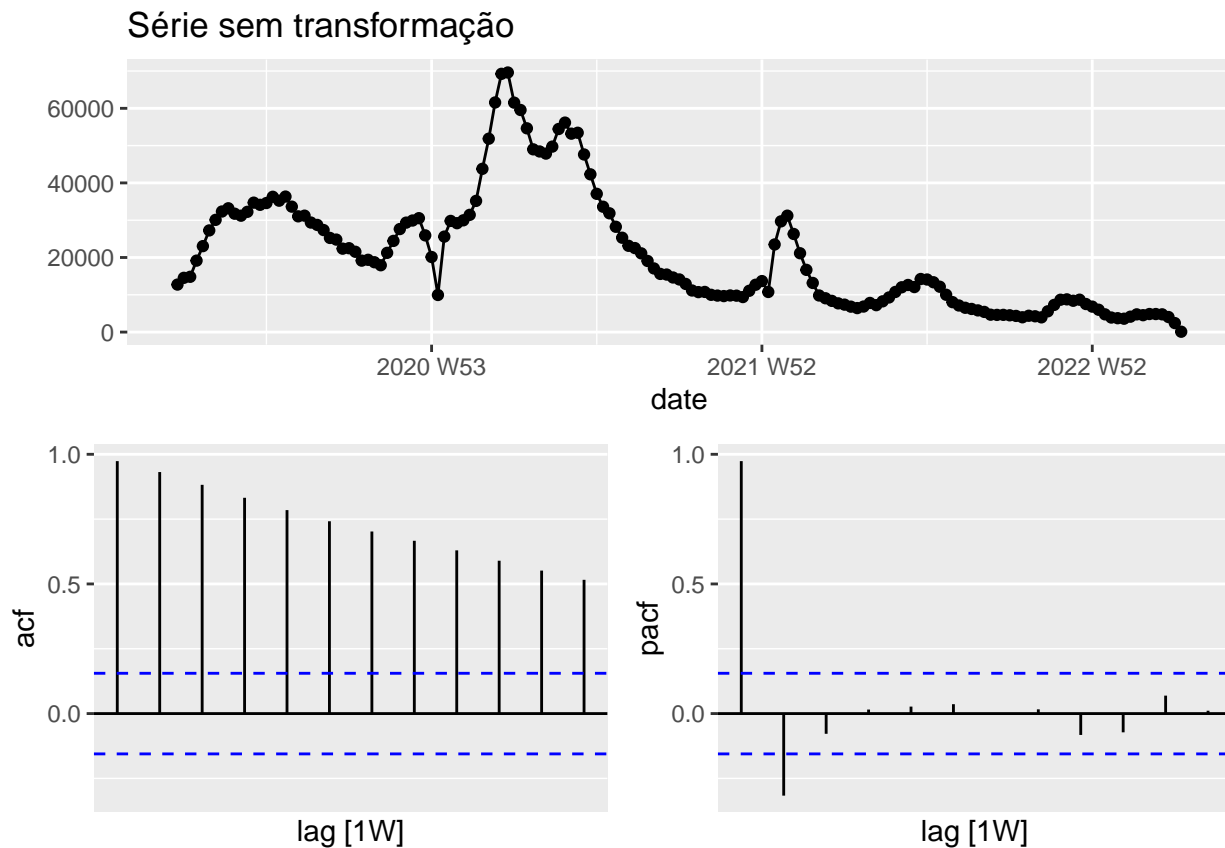
p.value	metodo
0.364236207606296	Augmented Dickey-Fuller Test

Assim, dado um p-valor acima da significância, não rejeitou-se a hipótese nula, ou seja, não rejeitou-se a hipótese da série ser não estacionária com 5% de significância.

Após a utilização de técnicas como transformação logarítmica e diferenciação, foi possível diminuir a variância da série e torná-la estacionária

Nos seguintes gráficos, tem-se a sequência de transformação utilizada

Gráfico da correlação parcial e total da Serie Original

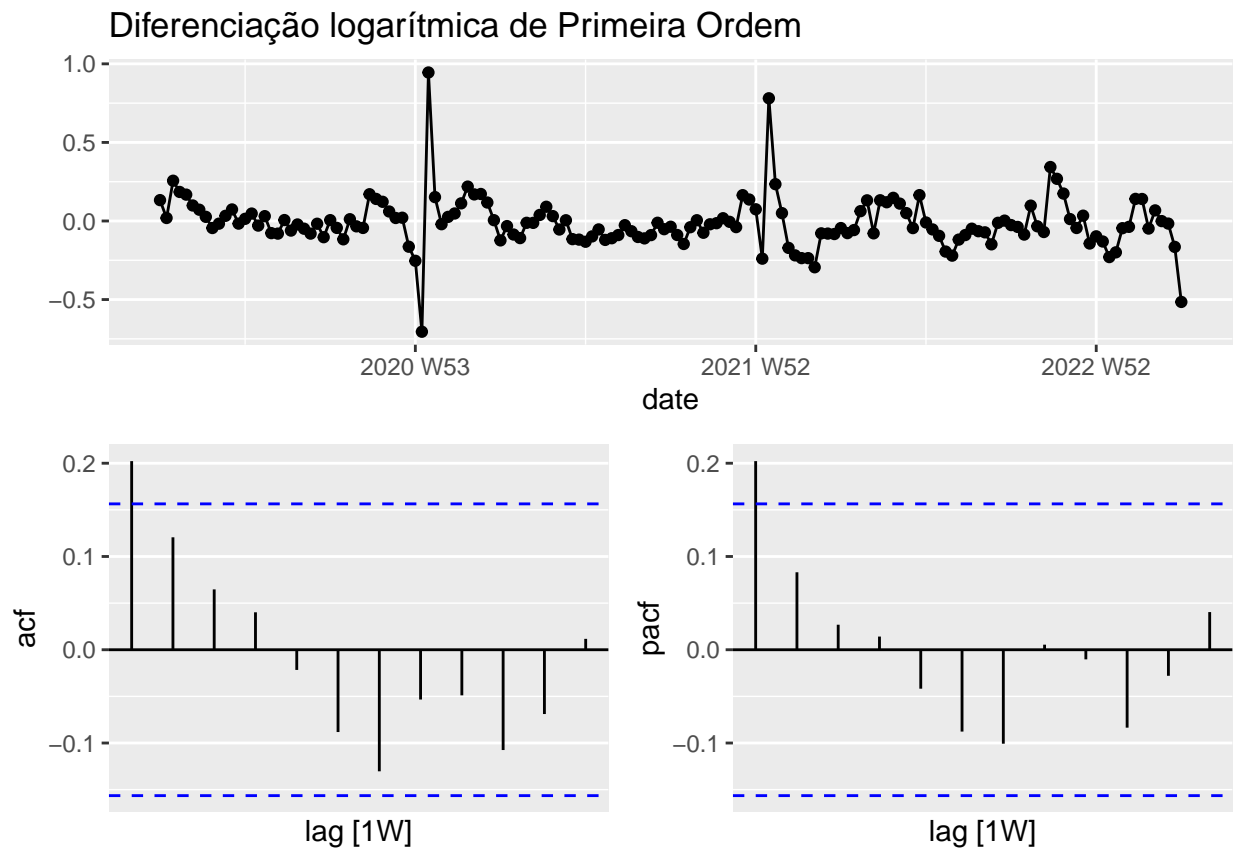


Observando os valores da correlação e da correlação parcial, notou-se um valor significativo para ambas as medidas.

No gráfico de correlação, é possível observar um decréscimo lento da correlação entre os lags, indicando novamente que os dados em questão são não-estacionários.

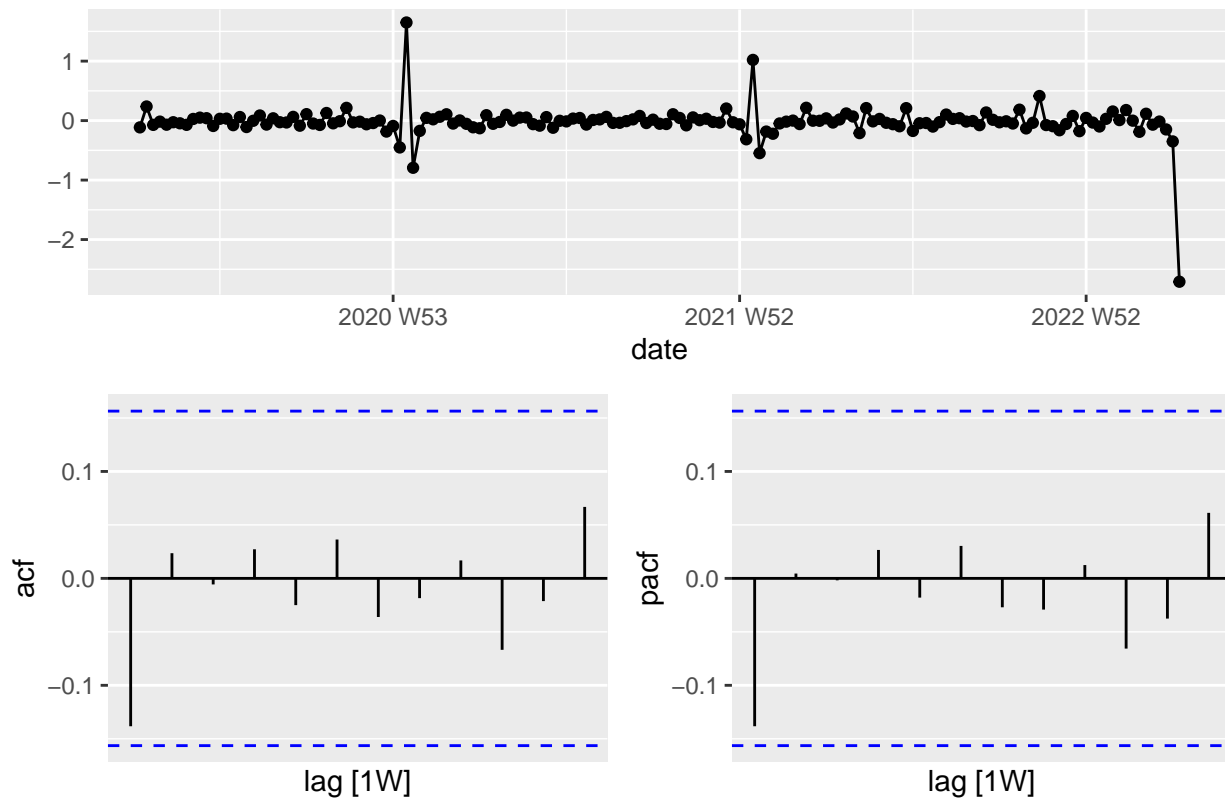
Assim, foi necessário a utilização da técnica de diferenciação. Foram testadas diferenciação de primeira e segunda ordem.

Gráfico da correlação parcial e total da Serie Trasformada



Via diferenciação de primeira ordem, a correlação entre os Lags próximos se mostrou significativo, onde além disso, o decrescimento não apresentou um decaimento rápido e aleatório. Assim, aplicou-se novamente o método de diferenciação

Diferenciação logarítmica de Segunda Ordem



A partir do gráfico anterior, nota-se que os valores de ACF e PACF não apresentam valores significativos de correlação entre cada Lag, além disso o decrescimento nos valores de correlação foi aleatório.

p.value	metodo
0.01	Augmented Dickey-Fuller Test

Aplicando novamente o teste, nota-se que a transformação utilizada alterou a tendencia da série, tornando-a estacionaria

Modelos

Foram considerados os modelos de ETS (Suavização Exponencial) e ARIMA (Auto-Regressivo Integrado de Medias Moveis). Para cada tipo de modelo foram apresentados os 3 melhores, apresentando seus valores escores de BIC, AIC e AICc

ETS

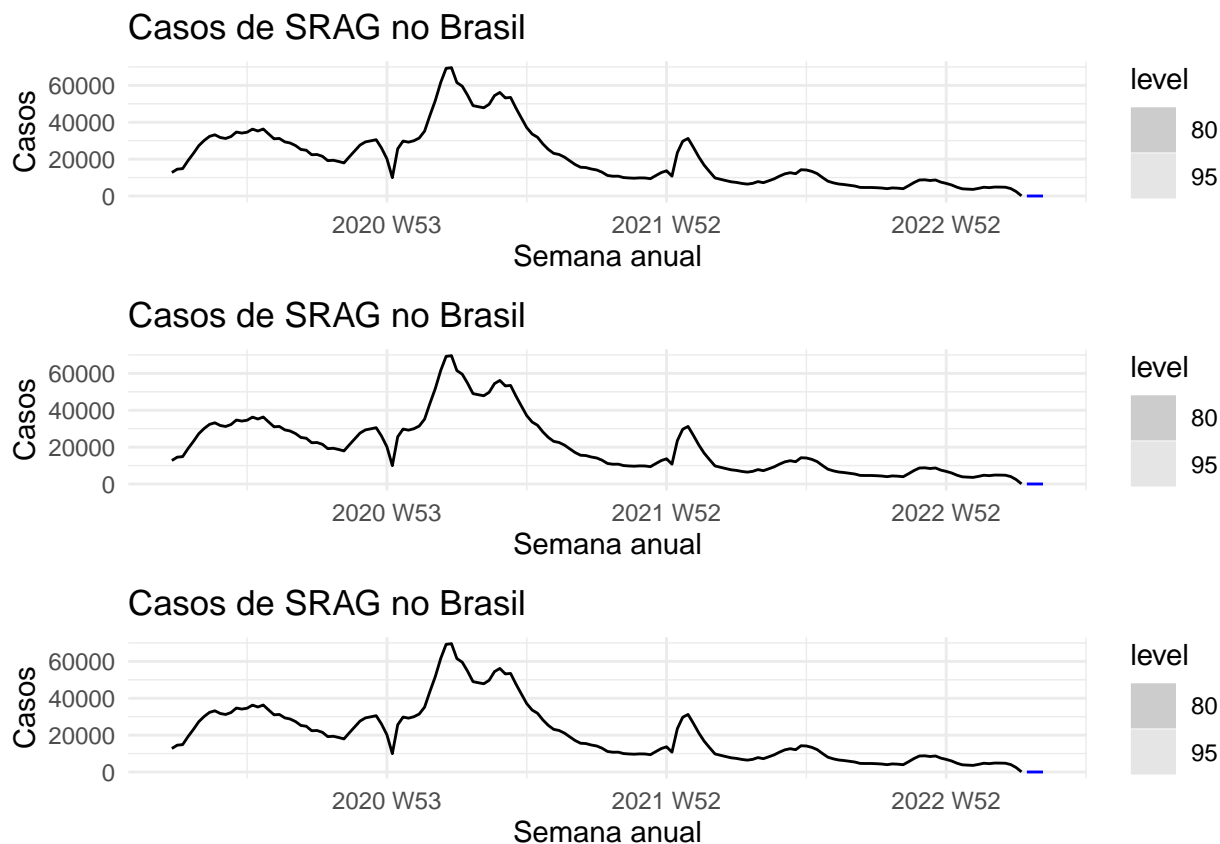
Foram testados os seguintes modelos:

$ETS(M, Ad)$,obtido atraves do algoritmo de Croston

$ETS(M, A)$

$ETS(A, Ad)$

Apresentando a seguinte previsão para o intervalo de 2 semanas



E apresentando os seguintes valores de Escores e do Erro Quadrático Médio

modelo	AIC	AICc	BIC	MSE
ets_m_ad	420.2661	420.6583	435.6107	0.0830250
ets_m_a	480.9270	481.3192	496.2716	0.0831958
ets_a_ad	420.2661	420.6583	435.6107	0.0830250

ARIMA

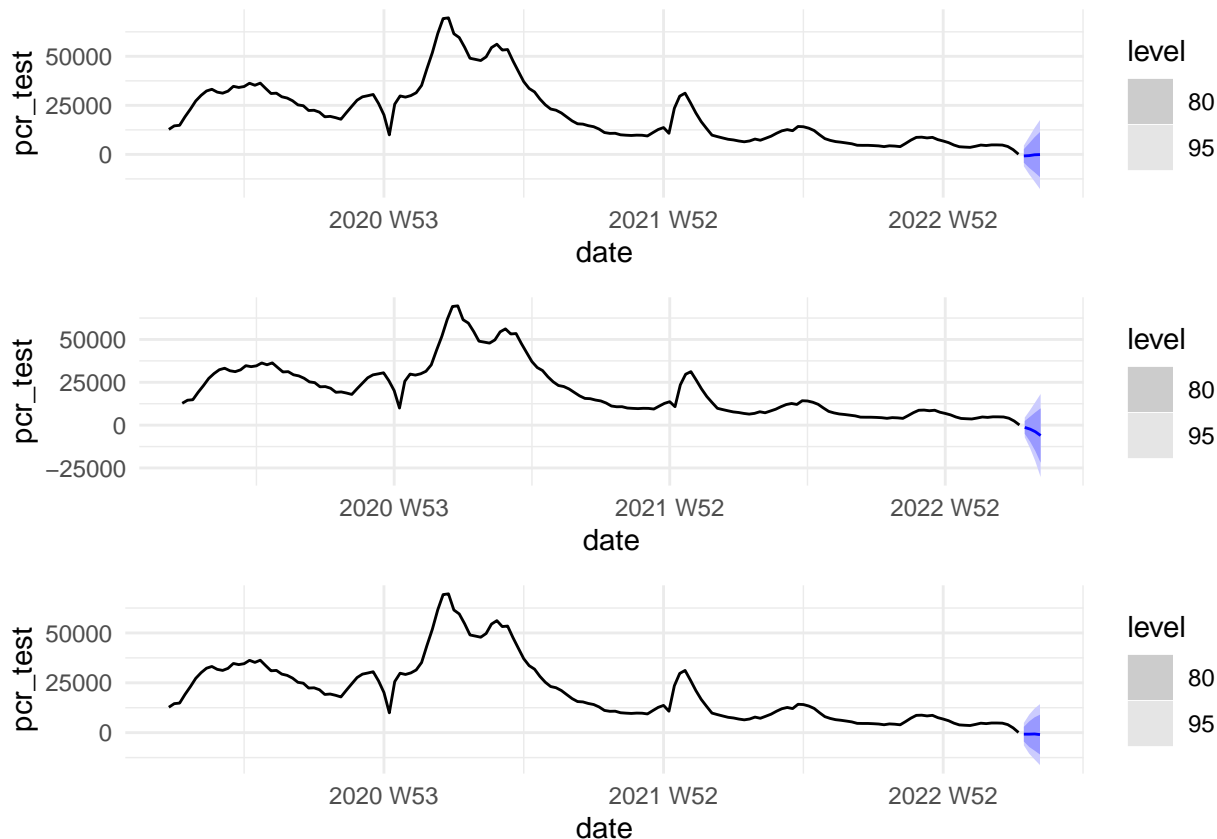
Foram testados os seguintes modelos:

$ARIMA(3, 1, 0)$, obtido atraves do Algoritmo de Hyndman-Khandakar

$ARIMA(0, 2, 1)$

$ARIMA(0, 2, 2)$

Apresentando a seguinte previsão para o intervalo de 1 mês



E apresentando os seguintes valores de Escores

modelo	AIC	AICc	BIC
arima_310	2962.797	2963.764	2987.298
arima_021	2965.789	2966.052	2978.014
arima_022	2957.241	2957.639	2972.522

Considerações Finais e Modelos Seleccionados

A previsão de um série temporal trata-se de uma forma de tentar prever o futuro com a maior precisão possível, considerando todas as informações disponíveis, incluindo dados históricos e conhecimento de quaisquer eventos futuros que possam afetar as previsões.

Os dados utilizados se tratam de uma série curta, assim o objetivo da previsão se engloba no escopo de Previsões de curto prazo de tempo.

Assim para a previsão de 2 semanas epidemiológicas será utilizado o modelo ETS(Multiplicativo, Aditivo Amortecido) dado por:

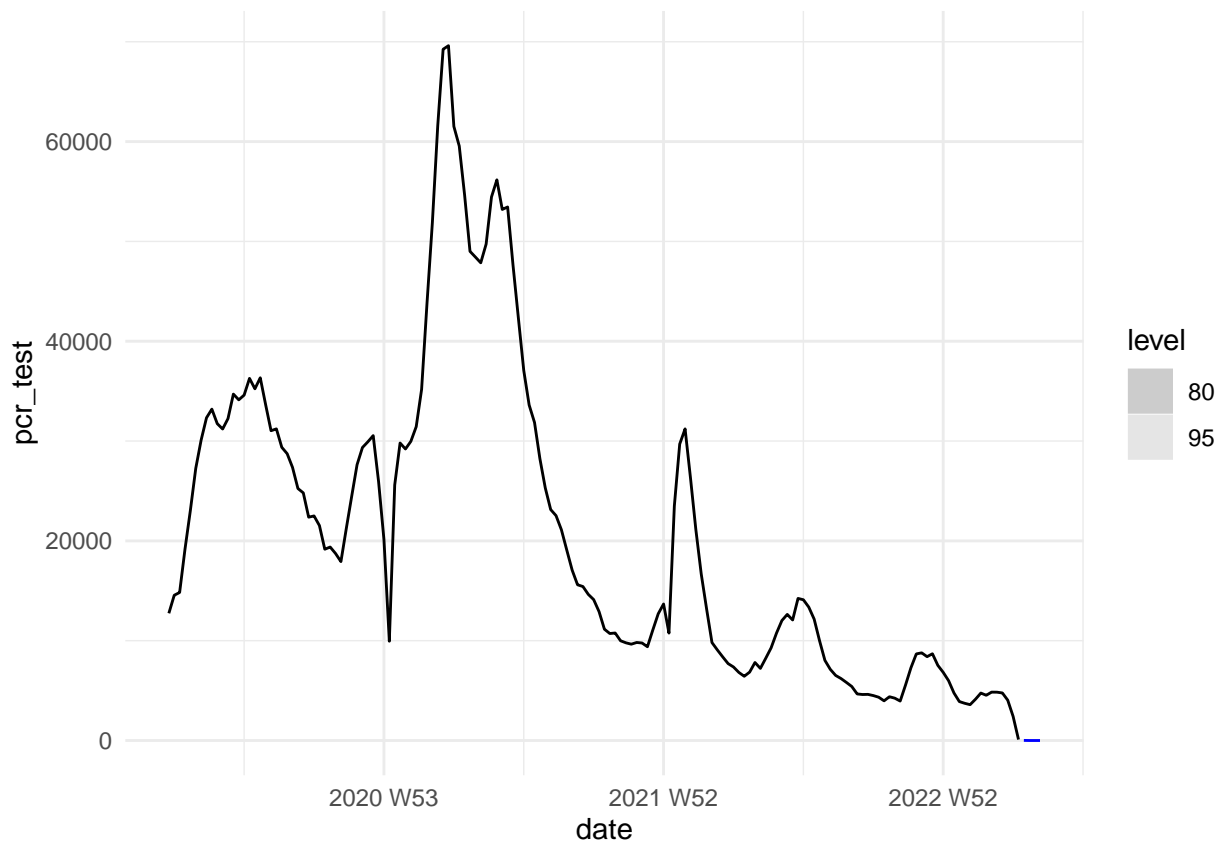
$$y_{t+h|t} = \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t, \text{onde}$$

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$$

$$b_t = \beta * (\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}.$$

Com $\alpha = 0.99$, $\beta = 0.535$, $\phi = 0.8$

Com a seguinte previsão para 2 semanas



Com a seguinte acurácia obtida

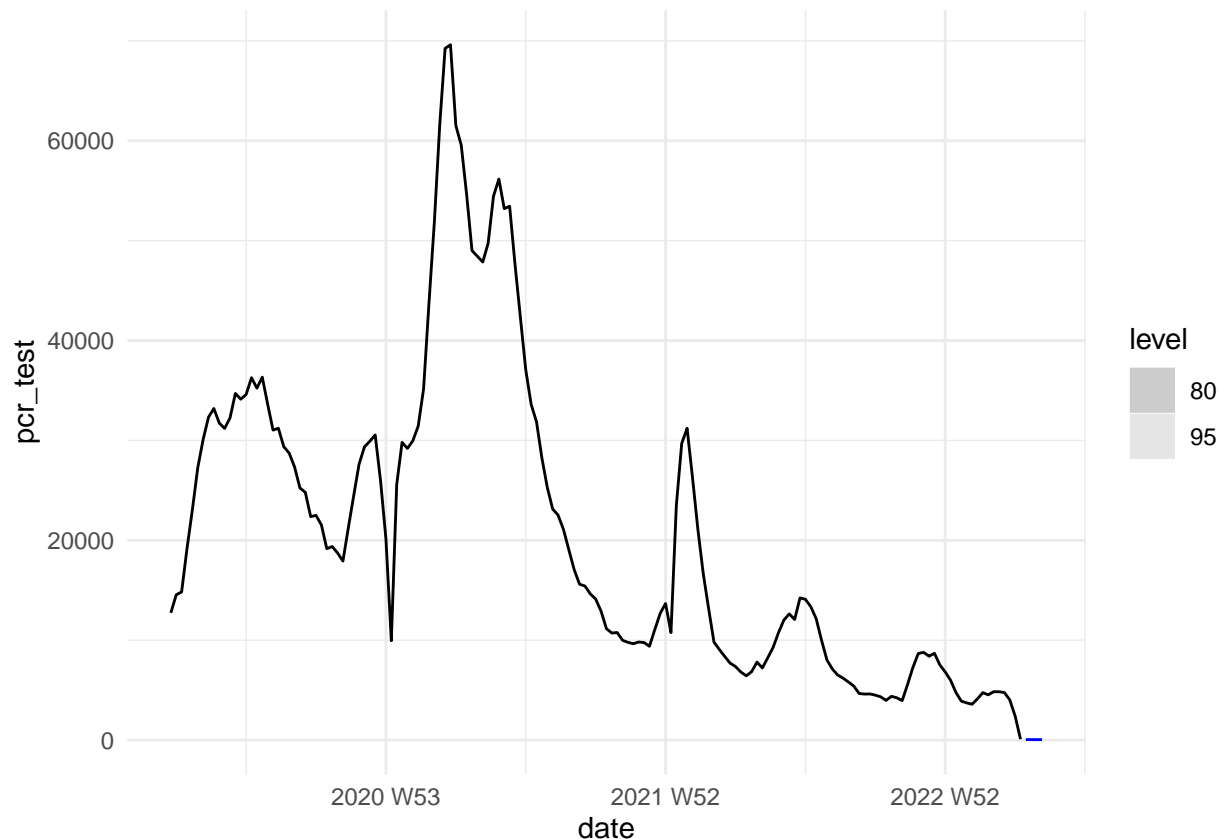
RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
3267.217	1788.814	-12.08728	19.41809	0.0995382	0.1374072	-0.0025977

Ja para a previsão de 1 mês, o modelo ETS apresenta um intervalo de confiança menos preciso que o modelo ARIMA. Assim para a previsão máxima a ser utilizada na plataforma de 1 mês, utilizou-se o modelo ARIMA(0,2,2), dado por:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-1}$$

Com $\theta_1 = -0.3225$, $\theta_2 = -0.3756$

Com a seguinte previsão para 1 mês



Com a seguinte acurácia obtida

RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
1038.263	704.5591	-21.01641	27.42611	0.1097113	0.1431524	-0.0995403

Para o modelo utilizado, uma transformação foi necessária para manter o intervalo de confiança acima de 0 (o intervalo não poderia ultrapassar a linha de 0 casos, pois é impossível um número de casos ser negativo)

Utilizou-se a seguinte transformação LOGIT Escalonada:

$$y = \log\left(\frac{x - a}{b - x}\right)$$

Onde, $a = 50$ e $b = 20000$

Disponibilização dos Dados e Referências

Ao longo do documento, os dados previsionados foram visualizados apenas com gráficos, assim o seguinte link disponibiliza os números para consultas e download

https://github.com/JFSalvandoTodos/public_data_rep/blob/main/forecast_values.csv

Mushtaq, Rizwan. 2011. "Augmented Dickey Fuller Test."

Niquini, Roberta Pereira, Raquel Martins Lana, Antonio Guilherme Pacheco, Oswaldo Gonçalves Cruz, Flávio Codeço Coelho, Luiz Max Carvalho, Daniel Antunes Maciel Villela, Marcelo Ferreira da Costa Gomes, and Leonardo Soares Bastos. 2020. "SRAG Por COVID-19 No Brasil: Descrição e Comparação de Características Demográficas e Comorbidades Com SRAG Por Influenza e Com a População Geral." *Cadernos de Saúde Pública* 36.