

Relatório 1 - Séries Temporais 2

Tharso Monteiro

5 de novembro de 2021

1 Introdução

Neste relatório, utilizaremos diversos modelos de séries temporais para tentar descrever o comportamento do número de casos confirmados de COVID-19 por dia na China e na Suíça. Os dados foram obtidos através do *Oxford COVID-19 Government Resopnse Tracker*, disponibilizados através do pacote **COVID19**, na linguagem R.

2 China

Os dados abaixo se referem à China Continental, com exclusão das regiões autônomas de Hong Kong e Macau.

O registro de casos confirmados de COVID-19 está ilustrado na Figura ??.

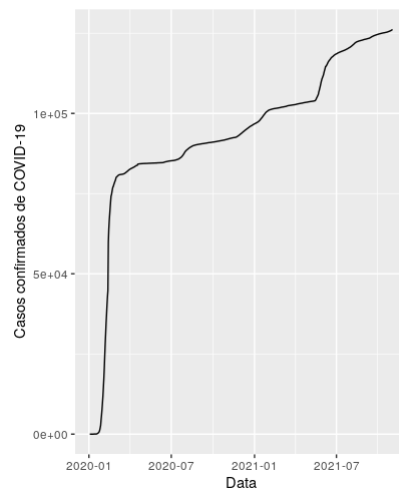


Figura 1: Casos confirmados de COVID-19 na China

Diferenciamos esta série temporal e obtivemos a série ilustrada na figura 2.

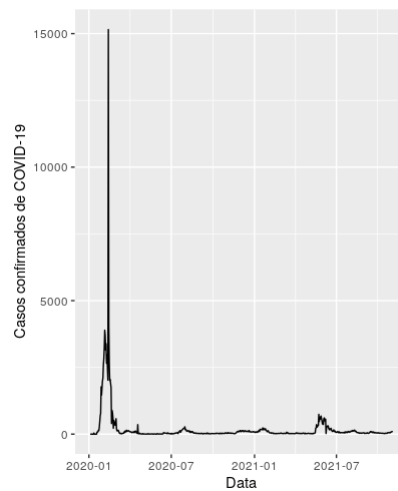


Figura 2: Casos diários de COVID-19 na China

Observa-se uma alta quantidade de casos confirmados no primeiro trimestre de 2020 e um aumento de casos no segundo trimestre de 2021. No momento de escrita deste relatório, está sendo registrado um aumento de casos na província de Heilongjiang, no nordeste da China.

Na figura 3, observa-se a decomposição STL da série temporal em seus componentes de tendência e sazonalidade.

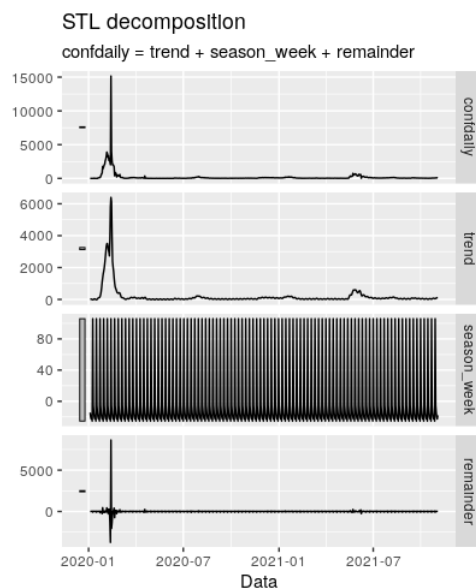


Figura 3: Decomposição STL : Casos diários de COVID-19 na China

A sazonalidade visível na decomposição certamente não advém de um comportamento da proliferação do vírus em si, mas sim da maneira em que os testes são realizados e seus resultados, introduzidos no banco de dados.

2.1 Modelo ETS

O modelo de suavização exponencial utiliza a decomposição da série em componentes de tendência, sazonalidade e erro, que podem ser multiplicativos, aditivos ou "amortecidos" (*damped*) para prever o comportamento futuro dos dados.

Em uma série temporal com presença de zeros (como é o caso desta), os modelos com erros multiplicativos são numericamente instáveis (Hyndman and Athanasopoulos, 2018). Assim, testamos os modelos restantes, comparamos seu erro médio quadrático, AIC e BIC. Verificamos que o modelo ETS mais razoável é o que apresenta todos os componentes aditivos.

A previsão de casos confirmados para os próximos 15 dias com base nesse modelo está ilustrada na Figura 4.

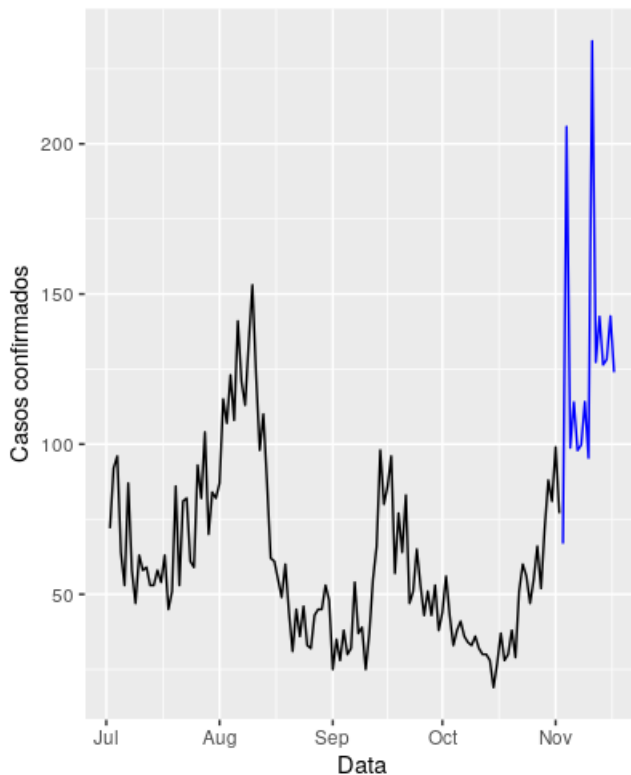


Figura 4: Previsão de novos casos, de acordo com modelo ETS aditivo

À primeira vista, este modelo pode parecer razoável, mas seu intervalo de previsão é bastante amplo e inclui valores negativos, como ilustra a Figura 5. Isso porque o modelo ETS não leva em consideração o fato de que a série com a qual se está trabalhando é uma de contagem, logo, não pode assumir valores negativos.

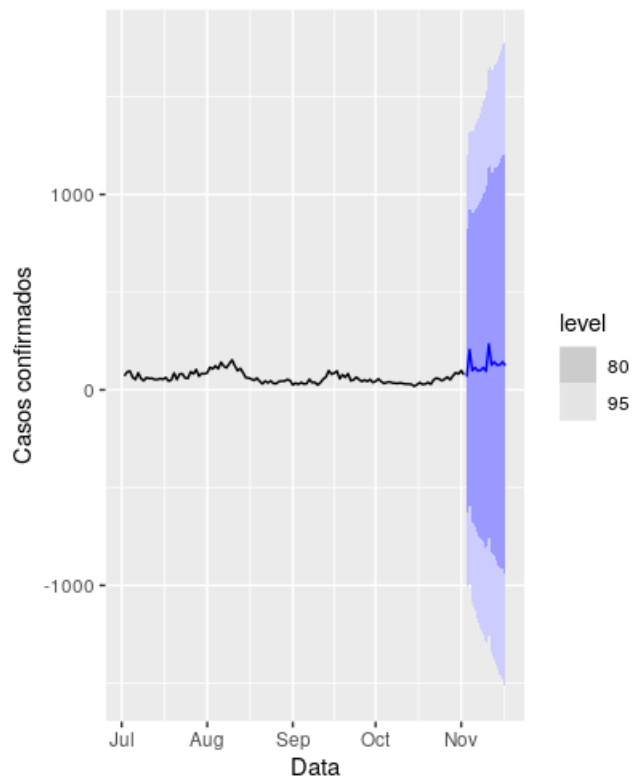


Figura 5: Intervalo de previsão de novos casos, de acordo com modelo ETS aditivo

2.2 Modelo ARIMA

Para a modelagem ARIMA, é necessário que o modelo seja estacionário.

Vamos trabalhar com a série já diferenciada uma vez, como ilustrada na Figura 2.

Fizemos o teste KPSS de raízes unitárias, que retornou um valor-p de 0.01. O teste de número de diferenças apontou a necessidade de mais uma diferenciação. Diferenciamos a série mais uma vez e analisamos seus gráficos de autocorrelação e autocorrelação parcial nas Figuras 6 e 7.

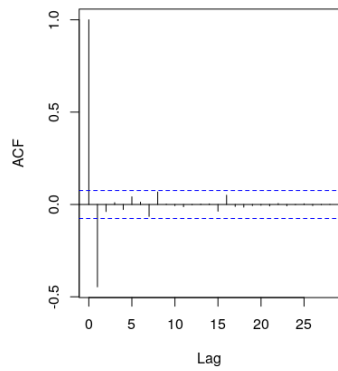


Figura 6: Autocorrelação da série com 2 diferenças

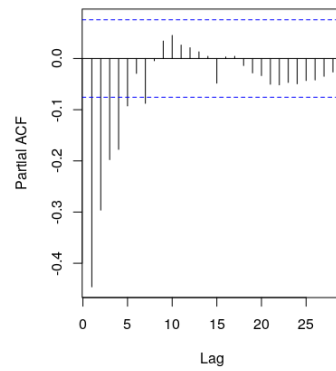


Figura 7: Autocorrelação parcial da série com 2 diferenças

Os gráficos sugerem um modelo MA(1) ou AR(5) na série diferenciada duas vezes, o que é equivalente a um modelo ARIMA(0,1,1) ou ARIMA(5,1,0) na série diferenciada uma vez. Comparando o AIC entre os dois modelos, vemos que o ARIMA(0,1,1) resulta num valor menor.

O gráfico de resíduos do modelo está representado na Figura 8. Vê-se que os resíduos parecem seguir um modelo de ruído branco. Confirmamos essa hipótese com o teste de Ljung-Box, que retornou um valor-p de 0.81.

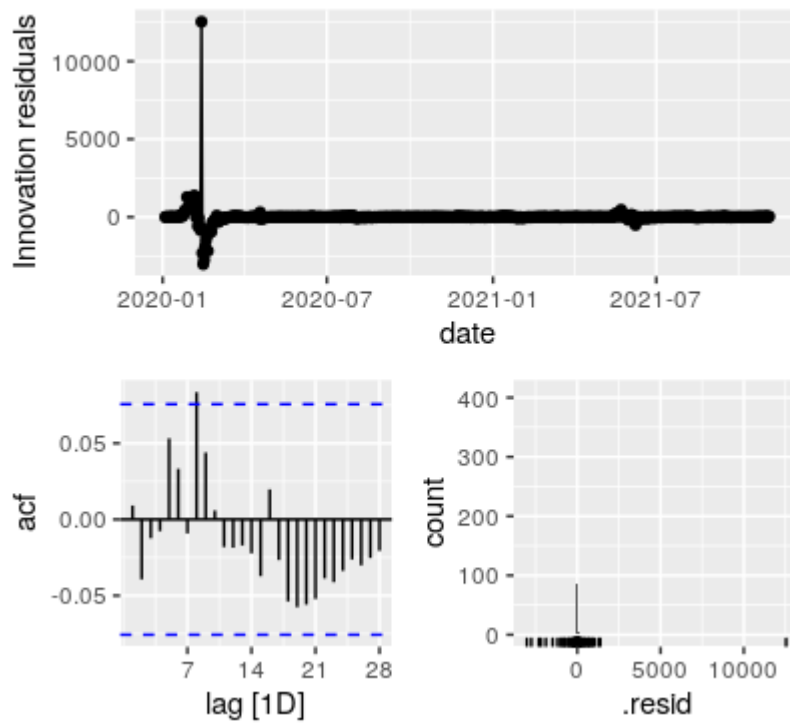


Figura 8: Gráfico de resíduos do modelo ARIMA

As previsões sem e com intervalo de previsão estão ilustradas nas Figuras 9 e 10, respectivamente.

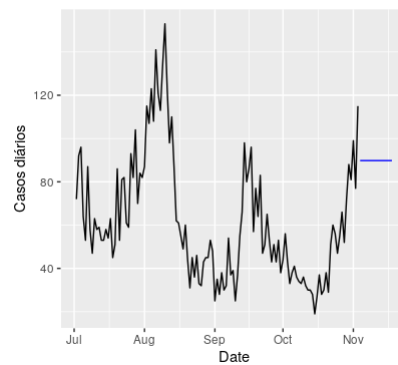


Figura 9: Previsões do modelo ARIMA para casos de COVID-19 na China

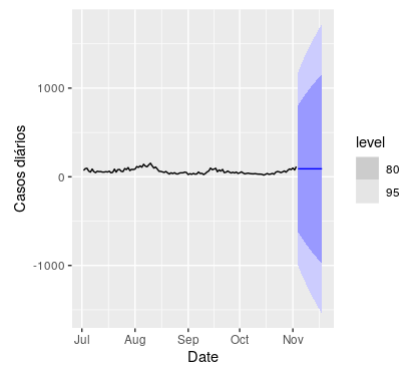


Figura 10: Intervalos de previsão do modelo ARIMA para casos de COVID-19 na China

Novamente, assim como o modelo ETS, o modelo ARIMA assume a possibilidade de valores negativos em seu intervalo de confiança, o que não é possível dado a natureza dos dados.

3 Suíça

Os gráficos de série temporal para os casos confirmados acumulados e diários estão disponíveis nas Figuras 11 e 12, respectivamente.

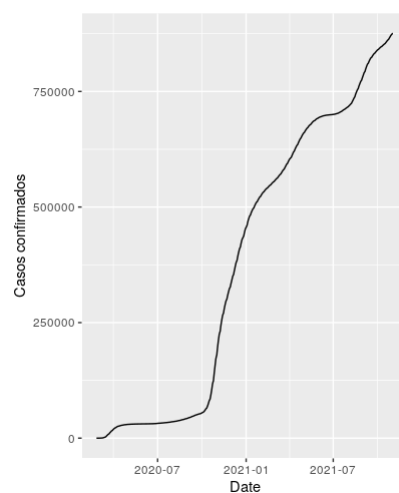


Figura 11: Casos confirmados de COVID-19 na Suíça

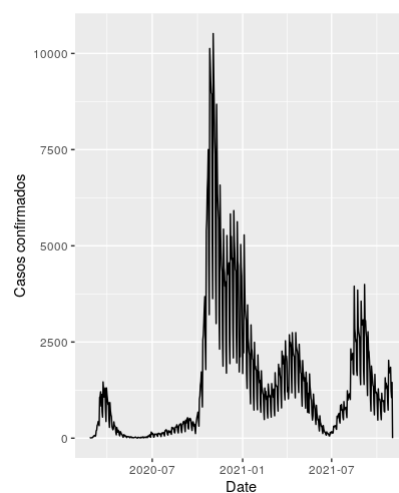


Figura 12: Casos diários de COVID-19 na Suíça

Observa-se um surto durante o último trimestre de 2020, assim como um aumento de casos confirmados em abril e agosto de 2021.

A decomposição STL da série está disponível na Figura 13.

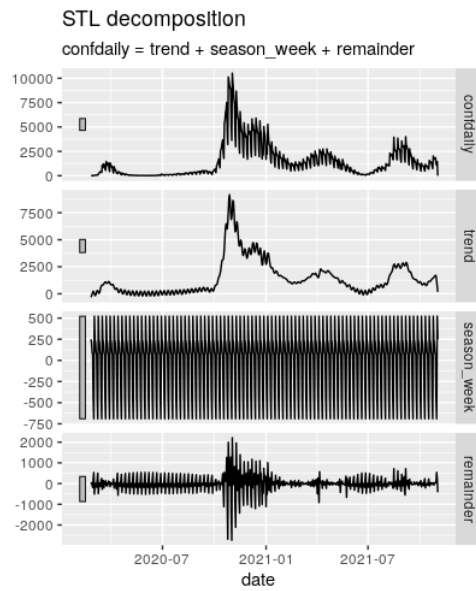


Figura 13: Decomposição STL : Casos diários de COVID-19 na Suíça

3.1 Modelo ETS

Repetimos o processo descrito para os dados da China. O modelo com menor AIC é o que possui erros e componente sazonal aditivos e componente de tendência amortecido.

A previsão para os próximos 15 dias com base neste modelo está ilustrada na Figura 14. Este modelo também tem intervalos de previsão que incluem valores negativos. Omitimos a visualização destes intervalos.

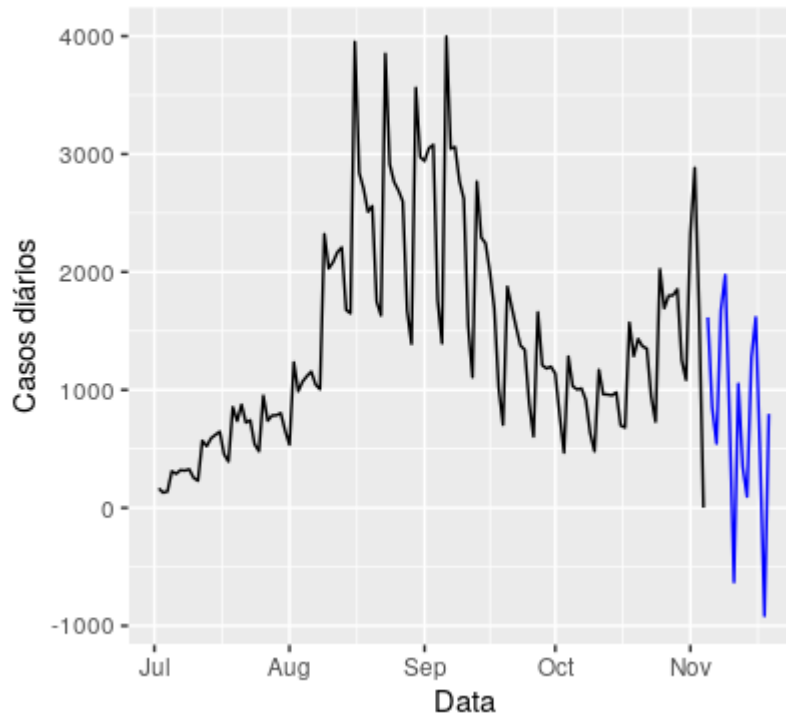


Figura 14: Previsão do modelo ETS para casos de COVID-19 na Suíça

3.2 modelo ARIMA

Repetimos o mesmo processo descrito para os dados da China. Os gráficos da ACF e PACF estão ilustrados nas Figuras 15 e 16, respectivamente.

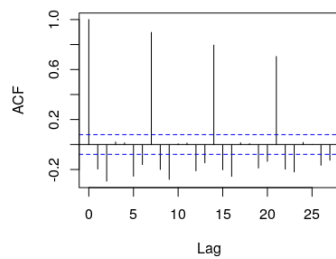


Figura 15: Autocorrelação da série com 2 diferenças

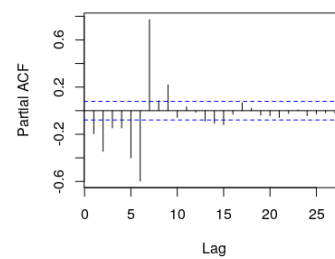


Figura 16: Autocorrelação parcial da série com 2 diferenças

Percebe-se uma autocorrelação sazonal semanal, mas não é trivial definir a ordem do modelo ARIMA apenas observando os gráficos. Utilizando a função **ARIMA**, escolhemos o modelo com menor AIC, que é o $\text{SARIMA}(2, 0, 3)(0, 1, 0)_7$.

Os gráficos referentes aos resíduos do modelo estão ilustrados na Figura 17. Os resíduos parecem seguir um modelo de ruído branco. Confirmamos esta hipótese com o teste de Ljung-Box, que retornou um valor-p de 0.78.

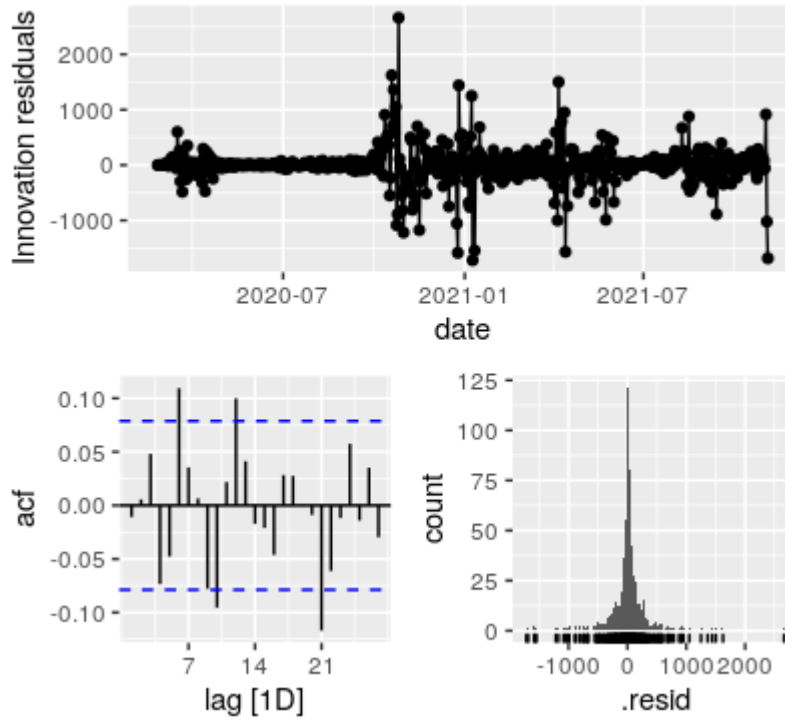


Figura 17: Gráfico de resíduos do modelo SARIMA

As previsões para os próximos 15 dias estão ilustradas na Figura 18. Assim como o modelo anterior, o modelo SARIMA gera intervalos de previsão que incluem valores negativos.

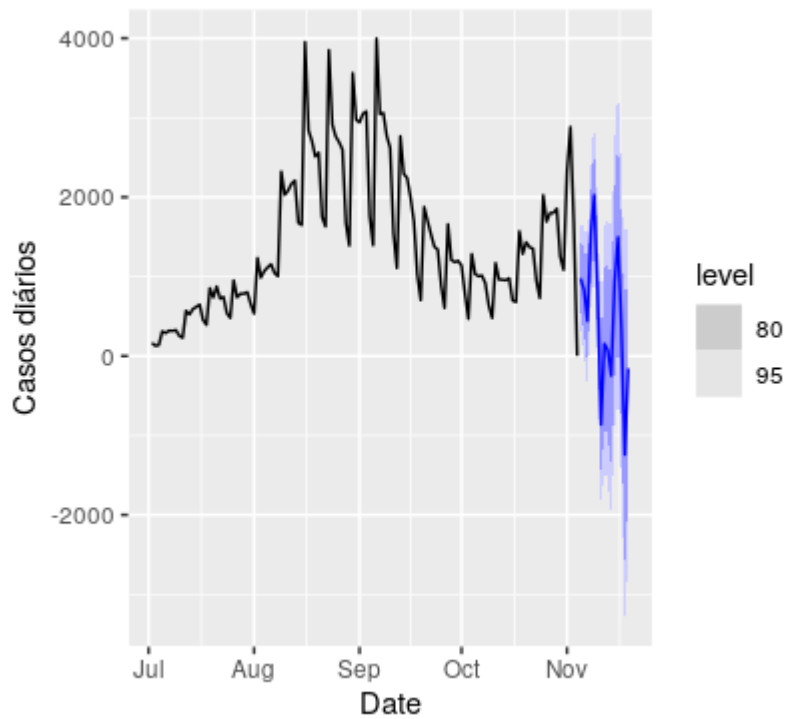


Figura 18: Previsão do modelo SARIMA para casos de COVID-19 na Suíça

Referências

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.