

# Um breve estudo utilizando série temporal, sobre os casos confirmados de COVID-19 no Brasil

Manuel Ferreira Junior, Universidade Federal da Paraíba

julho 06, 2021 - 18:18:50

## Resumo

O banco utilizado foi retirado de um repositório localizado no github, sendo mantido por uma iniciativa do dono do repositório. Os dados estão localizados de forma simples e prática, coletados manualmente do Ministério da Saúde. O conjunto de dados é continuamente atualizado, sendo realizado sempre próximo a meia noite.

- Fonte: Casos confirmados (<https://github.com/elhenrico/covid19-Brazil-timeseries/blob/master/confirmed-new.csv>) | Casos confirmados, filtrando apenas o Brasil ([https://github.com/Manuelfjr/ST/blob/main/R/project/R/data/confirmed/data\\_brasil\\_new\\_confirmed.csv](https://github.com/Manuelfjr/ST/blob/main/R/project/R/data/confirmed/data_brasil_new_confirmed.csv))

## Análise Descritiva

### Tratando o banco

```
library(urca)
url1 = 'https://raw.githubusercontent.com/Manuelfjr/ST/main/R/project/R'
url2 = 'data/confirmed/data_brasil_new_confirmed.csv'
url.c = paste(url1, url2, sep='/')
data.c = read.csv(url.c, header = TRUE)[-1,]
rownames(data.c) = seq(dim(data.c)[1])
colnames(data.c) = c('Data', 'BR')
data.c['BR'] = as.integer(data.c$BR)
```

### Descrição:

```
summary(data.c)
```

```
##      Data      BR
## Length:496    Min.   :    0
## Class :character 1st Qu.: 17391
## Mode  :character Median : 33775
##                      Mean  : 37888
##                      3rd Qu.: 55710
##                      Max.   :154664
```

```
str(data.c)
```

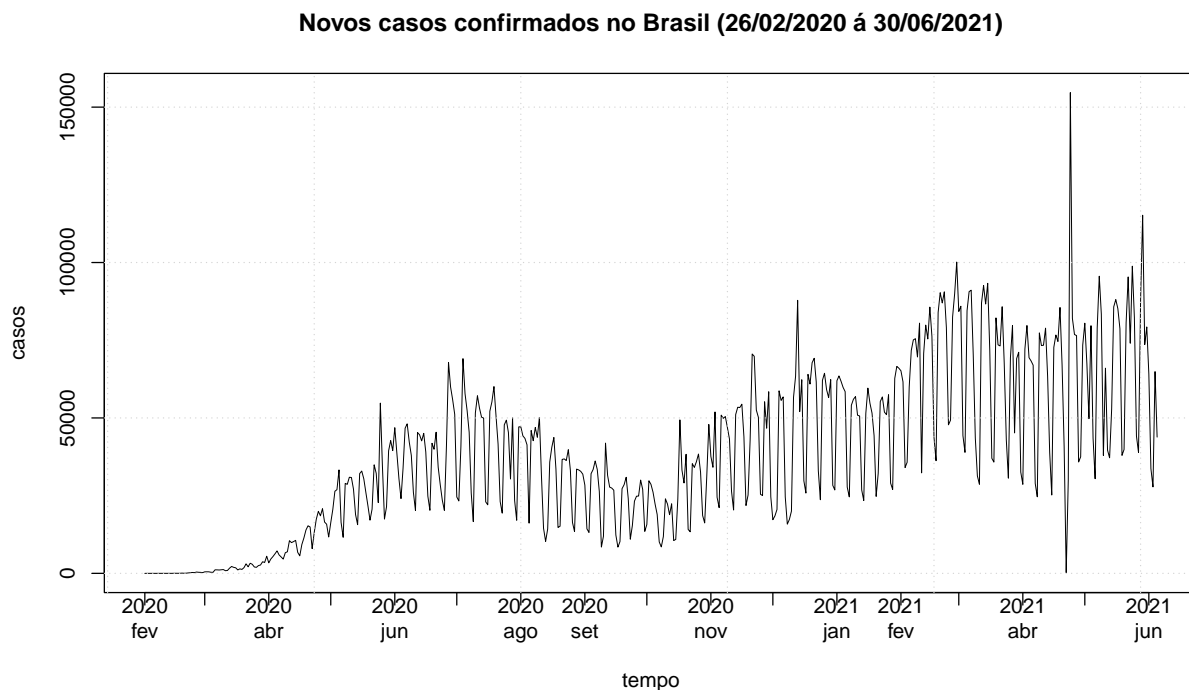
```
## 'data.frame':   496 obs. of  2 variables:
## $ Data: chr  "26/02/2020" "27/02/2020" "28/02/2020" "29/02/2020" ...
## $ BR : int  0 0 0 1 0 0 0 1 4 6 ...
```

Para o seguinte estudo, iremos considerar os casos diários de COVID-19 para o Brasil como um todo, realizando os técnicas estudadas durante o decorrer da disciplina de Séries Temporais. Além disso, o banco de dados foi filtrado para obtermos apenas os resultados referente ao número de casos diários confirmados no BRASIL, segue o link referente em anexo.

Com relação aos valores da série, ela não apresenta valores negativos e não foi preciso retirar nenhuma observação do conjunto de dados, ou seja, possuímos os valores observados dos casos confirmados de COVID-19 no Brasil desde o dia 26 de fevereiro de 2020.

## Série temporal

```
y.c = as.integer(data.c$BR[-seq(492,496,1)])
d <- seq(as.Date("2020-02-26"), as.Date("2021-06-30"), "day")
months <- seq(min(d), max(d), "month")
plot(y.c ~ d, type = 'l', xaxt = "n", xlab = 'tempo', ylab = 'casos', lwd = 0.5)
title('Novos casos confirmados no Brasil (26/02/2020 á 30/06/2021)')
axis(1, months, format(months, "\n%Y\n%b"))
grid()
```



A série utilizada contempla o período de 26 de fevereiro até o último dia atualizado sobre o conjunto de dados, sendo atualizado diariamente. Iremos contemplar dos dias 26 de fevereiro de 2020 até 30 de junho de 2021. Ao analisarmos a série, podemos observar uma queda no número de confirmados entre os períodos de agosto a outubro de 2020, evidenciado pelo possível resultado da criação da lei Nº 14.046, de 24 de agosto de 2020, onde fica explícito o adiamento e cancelamento de serviços, de reservas e de eventos dos setores de turismo e de cultura, em função do estado de calamidade pública, reconhecida pelo Decreto Legislativo nº 6, de 20 de março de 2020, e da emergência de saúde pública de importância internacional decorrente da pandemia da Covid-19.

Após a diminuição das medidas de afastamento, como lockdown e quarentena, em algumas regiões do país, mantendo apenas medidas de afastamentos previstas pela lei da quarentena, alguns estados começaram

a “relaxar”, e houve outro aumento no número de contaminados entre os meses de novembro e dezembro, tendo uma diminuição no aumento de contaminados no final de dezembro. Contudo, é possível notar que para o ano de 2021, já para o mês de fevereiro, podemos encontrar altos números de contaminados, dado pelo relaxamento das medidas de enfrentamento ao covid, por parte de alguns estados, sendo evidenciado uma possível terceira onda de ataque do vírus.

Por fim, podemos notar que até então, o Brasil encontra-se na sua terceira onda, a primeira sendo a mais duradoura, dando início seus altos números por volta do final de abril, e tendo uma queda no mês de agosto; A segunda onde é possível notar em um intervalo menor sendo entre o final de outubro e, mantendo uma constância no final de dezembro; E, atualmente, encontramos em uma terceira onda, vinda do início do ano de 2021.

Além disso, ao analisarmos a série temporal como um todo, podemos notar um comportamento estacionário, onde a série aparenta oscilar para baixo e para cima em torno de um valor esperado, ao desconsiderarmos o início dos registros de casos, por apresentar grande número de zeros inicialmente; Ademais, também é possível notar uma variância constante sobre as mesmas circunstâncias citadas anteriormente.

## Teste de estacionariedade

```
summary(ur.df(y.c, type='none', lags=0))

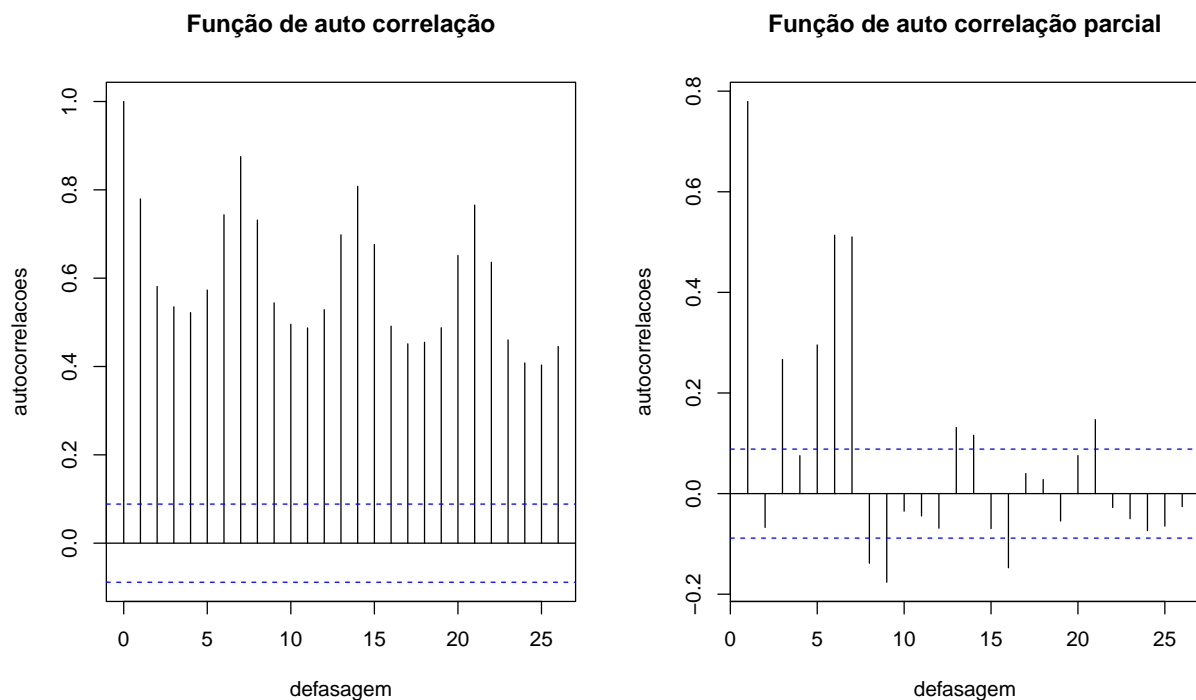
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -61723  -2964    660   6179 126857
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1 -0.07049      0.01679  -4.198  3.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17120 on 489 degrees of freedom
## Multiple R-squared:  0.03479,    Adjusted R-squared:  0.03282
## F-statistic: 17.62 on 1 and 489 DF,  p-value: 3.197e-05
##
##
## Value of test-statistic is: -4.1982
##
## Critical values for test statistics:
##          1pct   5pct 10pct
## tau1 -2.58 -1.95 -1.62
PP.test(y.c)
```

```
##
## Phillips-Perron Unit Root Test
##
## data: y.c
## Dickey-Fuller = -11.526, Truncation lag parameter = 5, p-value = 0.01
```

Ao considerarmos tanto o teste de de Phillips Perron quanto o de Dickey Fuller, tomando um nível de significância de 5%, temos que para ambos é possível notar um p-valor menor que o nível de significância, respectivamente equivalentes a 0.01 e 0.000867, ou seja, rejeitamos a hipótese nula de que a série é não estacionária (Testamos se  $\phi = 1$ , sobre  $h_0$ ), ou seja, há evidência suficientes para afirmar que a série temporal do número de casos de COVID-19 confirmados diariamente é estacionária.

## Função de autocorrelação e função de autocorrelação parcial

```
par(mfrow=c(1,2))
acf(y.c, main='Função de auto correlação',
    xlab="defasagem",ylab="autocorrelacoes")
pacf(y.c, main='Função de auto correlação parcial',
    xlab="defasagem",ylab="autocorrelacoes")
```



Considerando o gráfico de autocorrelação, podemos notar um decaimento das autocorrelações ao longo das defasagens lento, ou seja, uma diminuição de uma autocorrelação entre os valores observados ao longo dos períodos, decaindo de forma lenta. Da mesma forma, podemos notar que a função de autocorrelação parcial encontra-se em torno de 0, apresentando apenas algumas defasagens superior ao intervalo calculado para as suas autocorrelações, porém algumas das suas autocorrelações são superiores a 0.4. Além disso, ao repararmos que para ambos os gráficos, função de autocorrelação e função de autocorrelação parcial, algumas autocorrelações extrapolam os limites calculados, evidenciando uma violação da suposição que  $\rho = 0$ , ou seja, que as correlações ao longo das defasagens é diferente de 0.