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

Manuel Ferreira Junior, Universidade Federal da Paraíba

abril 28, 2021 - 12:14:44

Resumo

O banco utilizado foi retirado de um repositorio localizado no github, sendo mantido por uma iniciativa do dono do repositorio. 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 proximo a meia noite.

• Fonte: Casos confirmados (https://github.com/elhenrico/covid19-Brazil-timeseries/blob/master/confir med-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)

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)
                              BR
##
        Data
##
    Length: 424
                       Min.
    Class : character
                       1st Qu.: 15268
   Mode :character
                       Median : 30872
##
##
                        Mean
                               : 33746
##
                       3rd Qu.: 50634
##
                       Max.
                               :100158
str(data.c)
## 'data.frame':
                    424 obs. of 2 variables:
                 "26/02/2020" "27/02/2020" "28/02/2020" "29/02/2020" ...
    $ Data: chr
  $ BR : int 0 0 0 1 0 0 0 1 4 6 ...
```

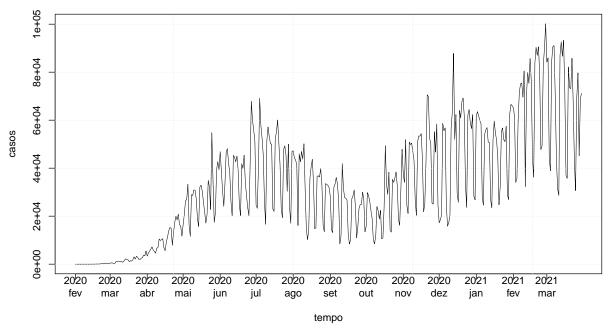
Para o seguinte estudo, iremos considerar os casos diários de COVID-19 para o Brasil como um todo, realizando os técnicas estudadas durante o decorrer da disciplina de Séries Temporais. Além disso, o banco de dados foi filtrado para obtermos apenas os resultados referente ao número de casos diários confirmados 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, possuimos 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)
d <- seq(as.Date("2020-02-26"), as.Date("2021-04-24"), "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 24/04/2021)')
axis(1, months, format(months, "\n%Y\n%b"))
grid()</pre>
```

Novos casos confirmados no Brasil (26/02/2020 á 24/04/2021)



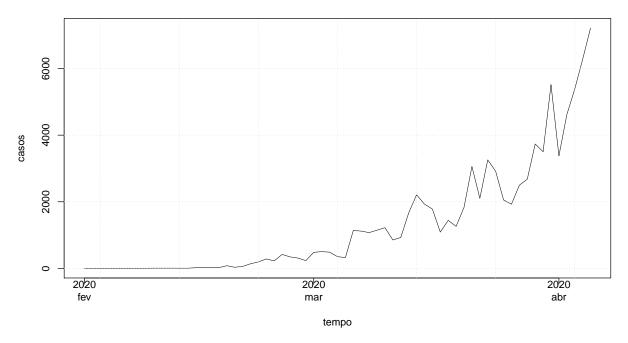
```
head(data.c, 65)
##
             Data
                    BR
      26/02/2020
                     0
## 1
  2
      27/02/2020
                     0
      28/02/2020
                     0
##
  3
##
  4
      29/02/2020
                     1
## 5
      01/03/2020
                     0
## 6
      02/03/2020
                     0
## 7
      03/03/2020
                     0
      04/03/2020
## 8
```

```
## 9 05/03/2020
                    4
## 10 06/03/2020
                    6
## 11 07/03/2020
                    6
## 12 08/03/2020
                    6
## 13 09/03/2020
                    5
## 14 10/03/2020
                    4
## 15 11/03/2020
                   18
## 16 12/03/2020
                   25
## 17 13/03/2020
                   21
## 18 14/03/2020
                   23
## 19 15/03/2020
                   79
## 20 16/03/2020
                   34
## 21 17/03/2020
                   57
## 22 18/03/2020
                  137
## 23 19/03/2020
                  193
## 24 20/03/2020
                  283
## 25 21/03/2020
                  224
## 26 22/03/2020
                  418
## 27 23/03/2020
                  345
## 28 24/03/2020
                  310
## 29 25/03/2020
                  232
## 30 26/03/2020
## 31 27/03/2020
                  502
## 32 28/03/2020
                  487
## 33 29/03/2020
                  352
## 34 30/03/2020
                  323
## 35 31/03/2020 1138
## 36 01/04/2020 1119
## 37 02/04/2020 1074
## 38 03/04/2020 1146
## 39 04/04/2020 1222
## 40 05/04/2020
                  852
## 41 06/04/2020
                  926
## 42 07/04/2020 1661
## 43 08/04/2020 2210
## 44 09/04/2020 1930
## 45 10/04/2020 1781
## 46 11/04/2020 1089
## 47 12/04/2020 1442
## 48 13/04/2020 1261
## 49 14/04/2020 1832
## 50 15/04/2020 3058
## 51 16/04/2020 2105
## 52 17/04/2020 3257
## 53 18/04/2020 2917
## 54 19/04/2020 2055
## 55 20/04/2020 1927
## 56 21/04/2020 2498
## 57 22/04/2020 2678
## 58 23/04/2020 3735
## 59 24/04/2020 3503
## 60 25/04/2020 5514
## 61 26/04/2020 3379
## 62 27/04/2020 4613
```

```
## 63 28/04/2020 5385
## 64 29/04/2020 6276
## 65 30/04/2020 7218

y.c = as.integer(head(data.c, 65)$BR)
d <- seq(as.Date("2020-02-26"), as.Date("2020-04-30"), "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 24/04/2021)')
axis(1, months, format(months, "\n%Y\n%b"))
grid()</pre>
```

Novos casos confirmados no Brasil (26/02/2020 á 24/04/2021)



A série utilizada contempla o périodo de 26 de fevereiro até o último dia atualizado sobre o conjunto de dados, sendo atualizado diariamente Iremos contemplar dos dia 26 de fevereiro de 2020 até 24 de abril de 2021. Ao analisarmos a série, podemos observar uma queda no número de confirmados entre os periodos de agosto a outubro de 2020, evidenciado pelo possivel resultado da criação da lei Nº 14.046, de 24 de agosto de 2020, onde fica explicito o adiamento e cancelamento de serviços, de reservas e de eventos dos setores de turismo e de cultura, em função do estado de calamidade publica, 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 possivel notar que para o ano de 2021, ja 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 possivel 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 possivel 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 inicio do ano de 2021.

Além disso, ao analisarmos a série temporal como um todo, podemos notar um comportamento estácionario, onde a série aparenta oscilar para baixo e para cima em torno de um valor esperado, ao desconsiderarmos o inicio dos registros de casos, por apresentar grande número de zeros inicialmente; Ademais, também é possivel notar uma variância constance sobre as mesmas circustancias 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 \sim z.lag.1 - 1)
## Residuals:
##
       Min
                1Q
                    Median
                                3Q
                                       Max
  -2480.99
                      0.34
##
            -80.17
                             103.40
                                   1791.19
##
## Coefficients:
##
         Estimate Std. Error t value Pr(>|t|)
## z.lag.1 0.06275
                    0.03646
                              1.721
                                     0.0902 .
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 570 on 63 degrees of freedom
## Multiple R-squared: 0.0449, Adjusted R-squared: 0.02974
## F-statistic: 2.962 on 1 and 63 DF, p-value: 0.09015
##
##
## Value of test-statistic is: 1.721
## Critical values for test statistics:
       1pct 5pct 10pct
## tau1 -2.6 -1.95 -1.61
PP.test(y.c)
##
##
   Phillips-Perron Unit Root Test
##
## data: y.c
## Dickey-Fuller = -0.39482, Truncation lag parameter = 3, p-value =
```

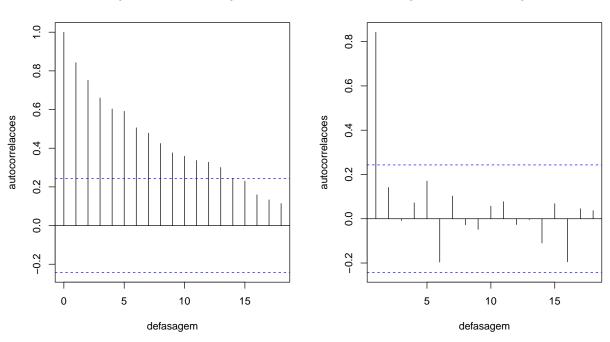
Ao considerarmos tanto o teste de de Phillips Perron quanto o de Dickey Fuller, tomando um nível de significância de 5%, temos que para ambos é possivel notar um p-valor menor que o nivel de significância,

respectivamente equivalentes a 0.01 e 0.000867, ou seja, rejeitamos a hipótese nula de que a série é não estácionaria (Testamos se $\phi = 1$, sobre h_0), ou seja, há evidência suficientes parar afirmar que a série temporal do número de casos de COVID-19 confirmados diariamente é estácionaria.

Função de autocorrelação e função de autocorrelação parcial

Função de auto correlação

Função de auto correlação parcial



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 osbervados ao longo dos periodos, 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 autocorrelações efunção de autocorrelação parcial, algumas autocorrelações extrapolam os limites calculados, evidenciando uma vioção da suposição que $\rho=0$, ou seja, que as correlações ao longo das defasagens é diferente de 0.