"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518, USA), Vol.7, Issue 3, 109-121 | Science Signpost Publishing

Covid-19: Correlation Between Confirmed Cases and Deaths – Greece & Cyprus

Nicholas G. Berketis PhD*

What is the daily correlation of **Confirmed** versus **Death** Cases in Covid-19? In other words, people who have passed away, on average, how many days ago they were reported (i.e. "Confirmed") as Covid-19 new cases?

To answer this question, we can take the correlation between the Daily Confirmed vs Daily Deaths and trying different lag values of the confirmed cases, since the assumption is that it will take some days for someone to pass away since he / she has been diagnosed with Covid-19.

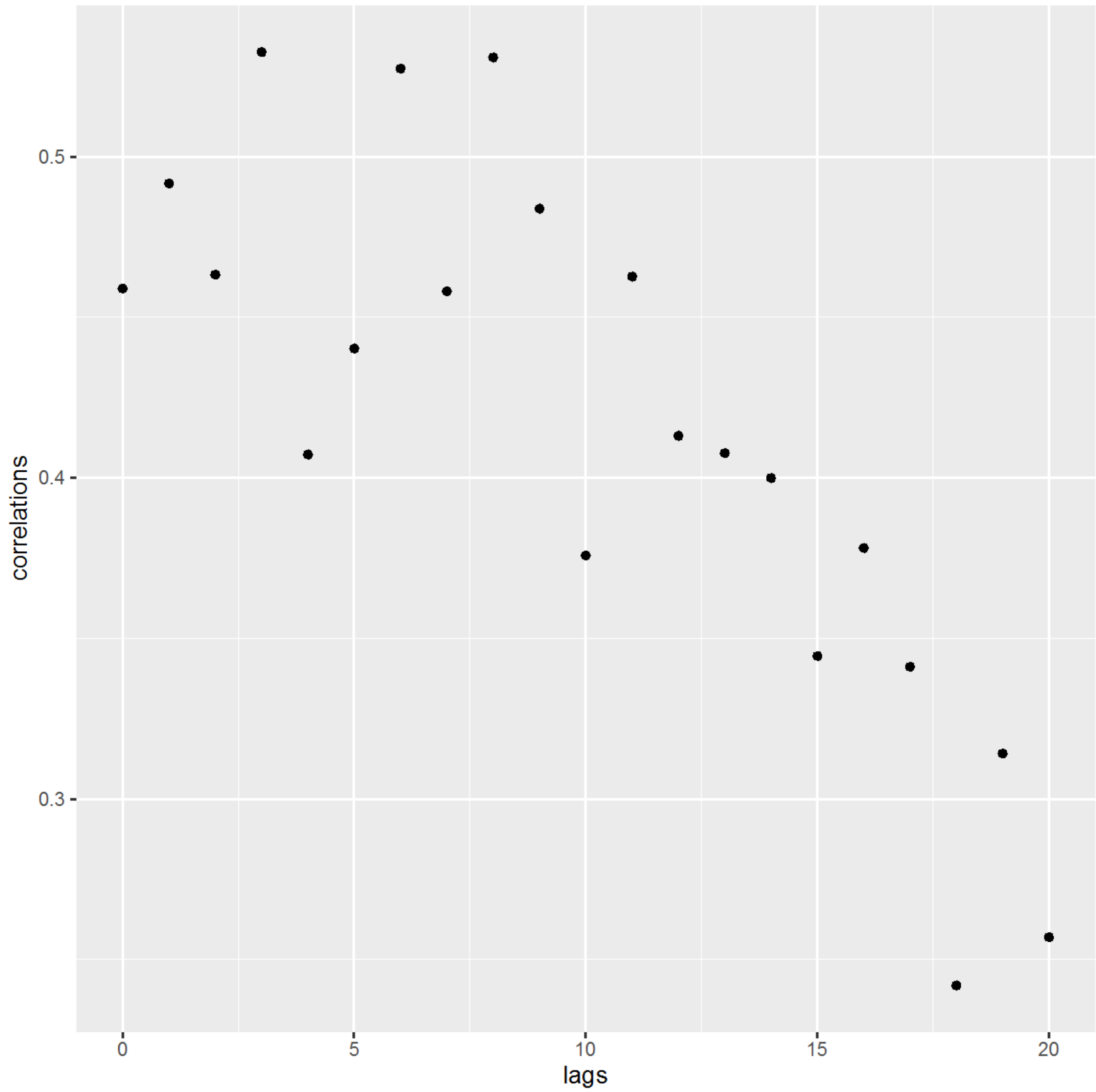
The problem with the data is that are affected by the number of tests and also during some days like weekends they do not report all the cases. This implies that our analysis is not valid, but we shall try to check the results.

The analysis is based on *R* (Version 4.0.3 – 10th October, 2020), an integrated language and environment for statistical computing and graphics. *R* provides a wide variety of statistical and graphical techniques.

```
df<-coronavirus%>%filter(country=='Greece', date>='2020-02-15')%>%select(date, country, type, cases)%>%
  group_by(date, country, type) %>%pivot_wider(names_from =type,
values_from=cases) %>%ungroup()

correlations<-c()
lags<-c(0:20)

for (k in lags) {

  tmp<-df%>%mutate(lagk=lag(confirmed,k))%>%select(death,lagk)%>%na.omit()

  correlations<-c(correlations,cor(tmp$death, tmp$lagk))
}

data.frame(lags, correlations)

data.frame(lags, correlations)%>%ggplot(aes(x=lags, y=correlations))+geom_point()
```

* Manager for J.Kouroutis & Co. Ltd. Insurance and Reinsurance Brokers, Piraeus, Greece (1982 – to-date). Lecturer of Marine Insurance for undergraduates at Frederick University, Cyprus (2008 – to-date) and for postgraduates at the MSc in International Shipping, Finance and Management of the Athens University of Economics and Business, Athens, Greece (2016 – to-date) & Course Leader for the Certificate in Marine Insurance at Lloyd's Maritime Academy (November 2019 – to-date).

lags	correlations
0	0.4589764
1	0.4916617
2	0.4632576
3	**0.5325860**
4	0.4072753
5	0.4402747
6	0.5274265
7	0.4580970
8	0.5308636
9	0.4838212
10	0.3758587
11	0.4626586
12	0.4130963
13	0.4077379
14	0.3999310
15	0.3445465
16	0.3781832
17	0.3411893
18	0.2420049
19	0.3141870
20	0.2569894



As we see, the argmax[†] correlation is at k=3, which implies (if the data were accurate), that from the people who have passed away, most of them diagnosed with Covid-19 **3 days ago**.

Let's do the same analysis, but this time by taking into consideration the Simple Moving Average[‡] of 4 days.

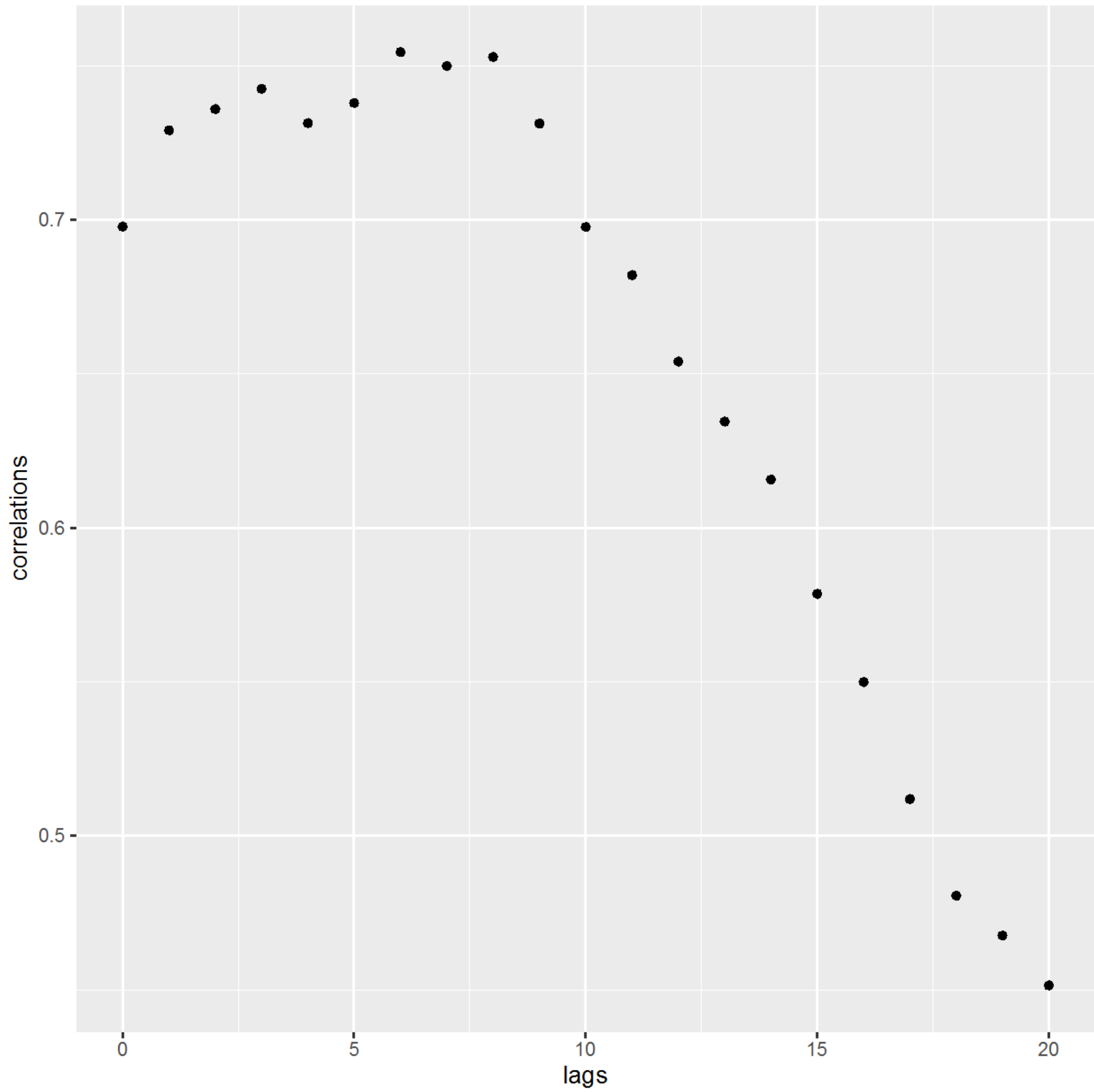
df<-coronavirus%>%filter(country=='Greece', date>='2020-02-15')%>%select(date, country, type, cases)%>%group_by(date, country, type) %>%pivot_wider(names_from =type,

[†] **Argmax**: the arguments of the maxima are the points, or elements, of the domain of some function at which the function values are maximized.

[‡] A **simple moving average** (SMA) calculates the **average** of a selected range of prices, usually closing prices, by the number of periods in that range.

```
values_from=cases) %>%ungroup()%>%mutate(confirmed = stats::filter(confirmed, rep(1 / 4, 4),
sides = 1), death = stats::filter(death, rep(1 / 4, 4), sides = 1))%>%na.omit()

correlations<-c()
lags<-c(0:20)

for (k in lags) {

  tmp<-df%>%mutate(lagk=lag(confirmed,k))%>%select(death,lagk)%>%na.omit()

  correlations<-c(correlations,cor(tmp$death, tmp$lagk))
}

data.frame(lags, correlations)

data.frame(lags, correlations)%>%ggplot(aes(x=lags, y=correlations))+geom_point()
```

lags	correlations
0	0.6978539
1	0.7291375
2	0.7359924
3	0.7425944
4	0.7314337
5	0.7379589
6	**0.7545403**
7	0.7499734
8	0.7529553
9	0.7313076
10	0.6977352
11	0.6821124
12	0.6540264
13	0.6345362
14	0.6156951
15	0.5786246
16	0.5499869
17	0.5119829
18	0.4805990
19	0.4676890
20	0.4514885



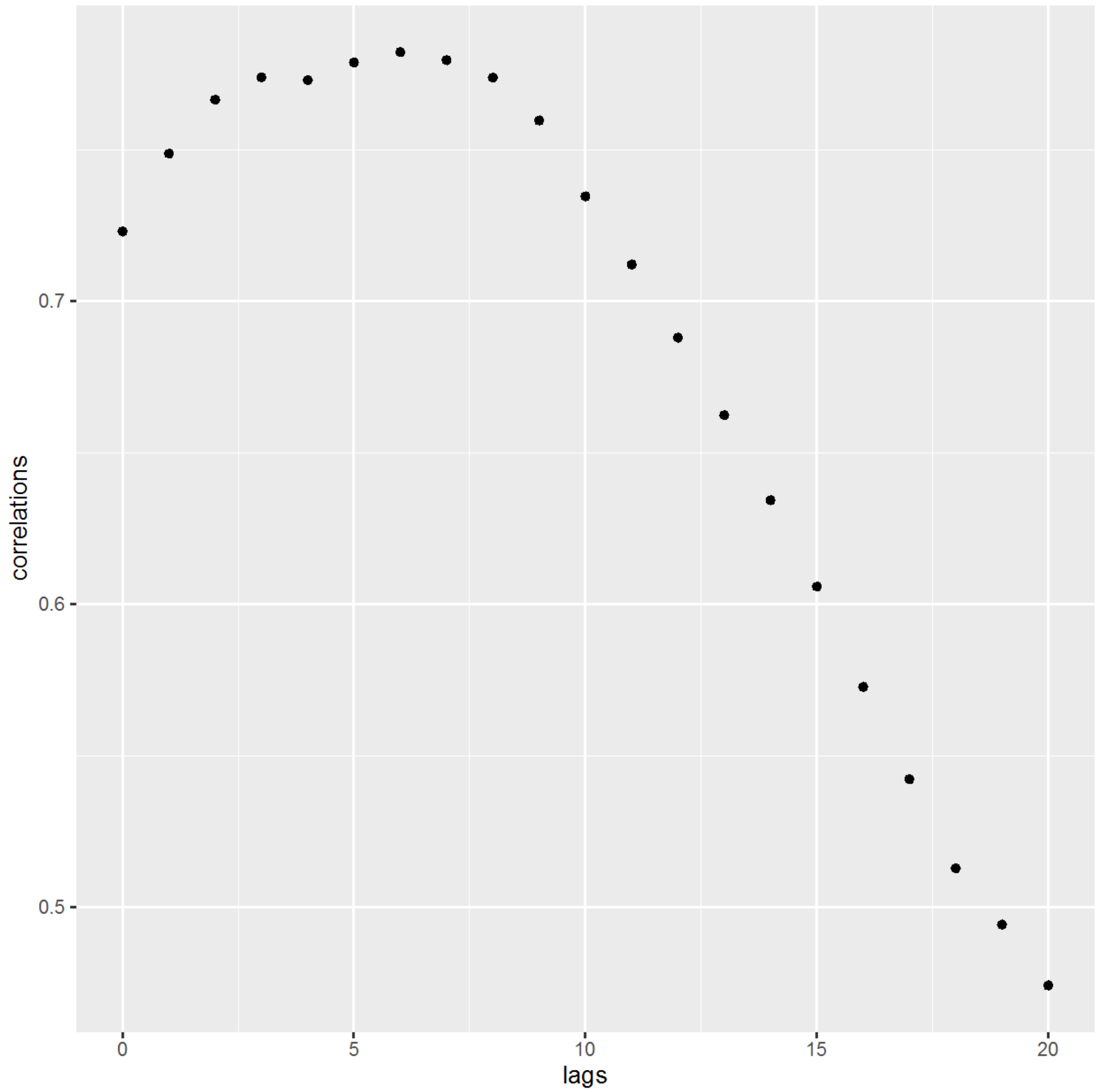
Where we consider the SMA of 4 days the maximum correlation **is at day 6**.

The conclusion is the same even if we consider the SMA of 5 days, i.e.:

```
df<-coronavirus%>%filter(country=='Greece', date>='2020-02-15')%>%select(date, country, type, cases)%>%
  group_by(date, country, type) %>%pivot_wider(names_from =type, values_from=cases) %>%ungroup()%>%
  mutate(confirmed = stats::filter(confirmed, rep(1 / 5, 5), sides = 1), death = stats::filter(death, rep(1 / 5, 5), sides = 1))%>%na.omit()

correlations<-c()
lags<-c(0:20)

for (k in lags) {

  tmp<-df%>%mutate(lagk=lag(confirmed,k))%>%select(death,lagk)%>%na.omit()
```

```
  correlations<-c(correlations,cor(tmp$death, tmp$lagk))
}

data.frame(lags, correlations)
data.frame(lags, correlations)%>%ggplot(aes(x=lags, y=correlations))+geom_point()
```

lags	correlations
0	0.7230774
1	0.7487569
2	0.7665641
3	0.7739573
4	0.7730262
5	0.7788614
6	**0.7822957**
7	0.7796162
8	0.7738220
9	0.7596860
10	0.7346517
11	0.7121598
12	0.6880434
13	0.6624295
14	0.6343464
15	0.6058305
16	0.5727766
17	0.5422709
18	0.5128789
19	0.4942831
20	0.4742360

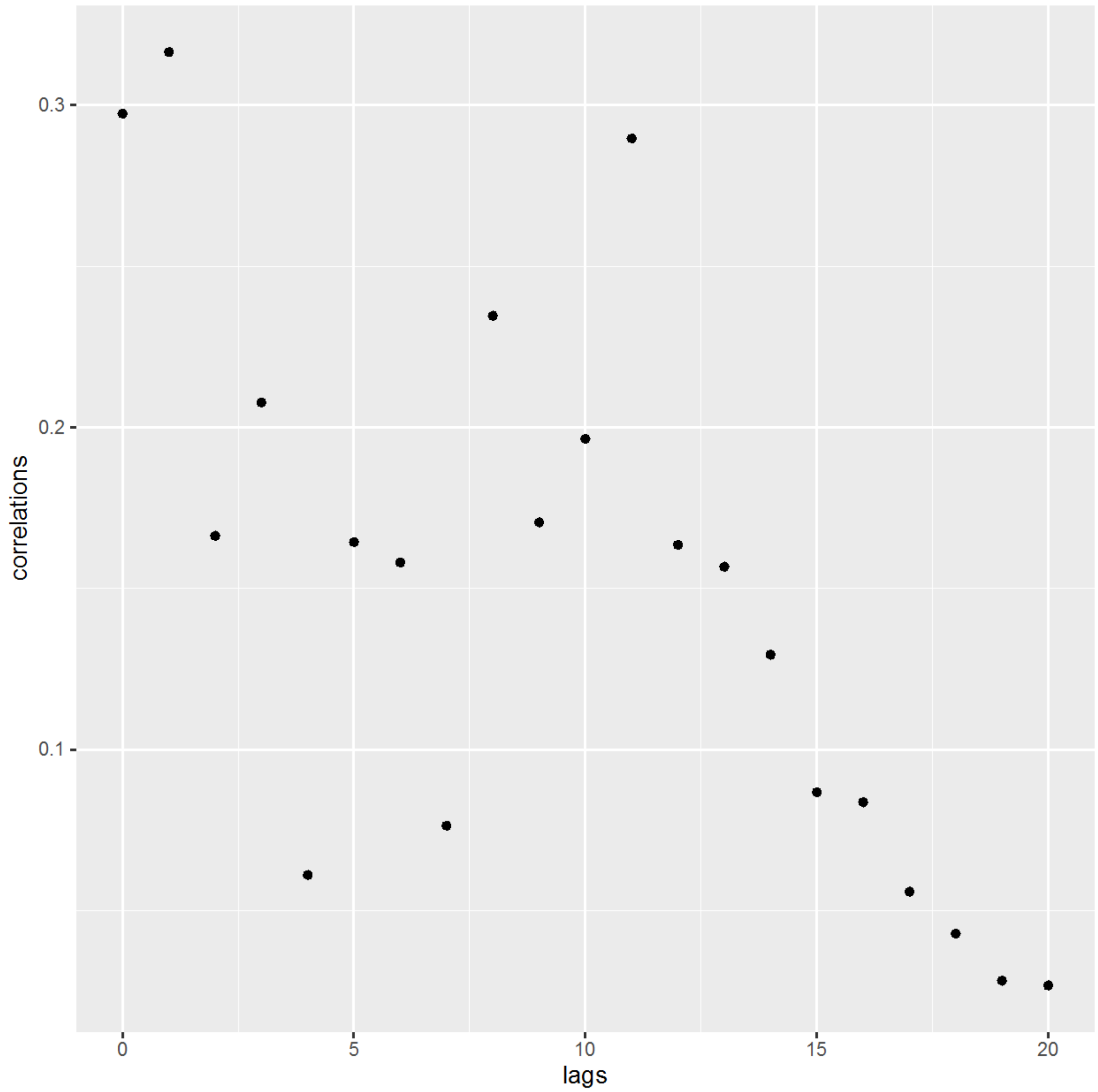


We shall carry out the same analysis for Cyprus:

```
df<-coronavirus%>%filter(country=='Cyprus', date>='2020-02-15')%>%select(date, country, type, cases)%>%
  group_by(date, country, type) %>%pivot_wider(names_from =type, values_from=cases) %>%ungroup()

correlations<-c()
lags<-c(0:20)

for (k in lags) {

  tmp<-df%>%mutate(lagk=lag(confirmed,k))%>%select(death,lagk)%>%na.omit()

  correlations<-c(correlations,cor(tmp$death, tmp$lagk))
}
```

```
data.frame(lags, correlations)

data.frame(lags, correlations)%>%ggplot(aes(x=lags, y=correlations))+geom_point()
```

lags	correlations
0	0.29734297
1	**0.31647552**
2	0.16631271
3	0.20770078
4	0.06107268
5	0.16435202
6	0.15805462
7	0.07636157
8	0.23461277
9	0.17050754
10	0.19640859
11	0.28963665
12	0.16354502
13	0.15674253
14	0.12941090
15	0.08675688
16	0.08364955
17	0.05586444
18	0.04283773
19	0.02825868
20	0.02679216



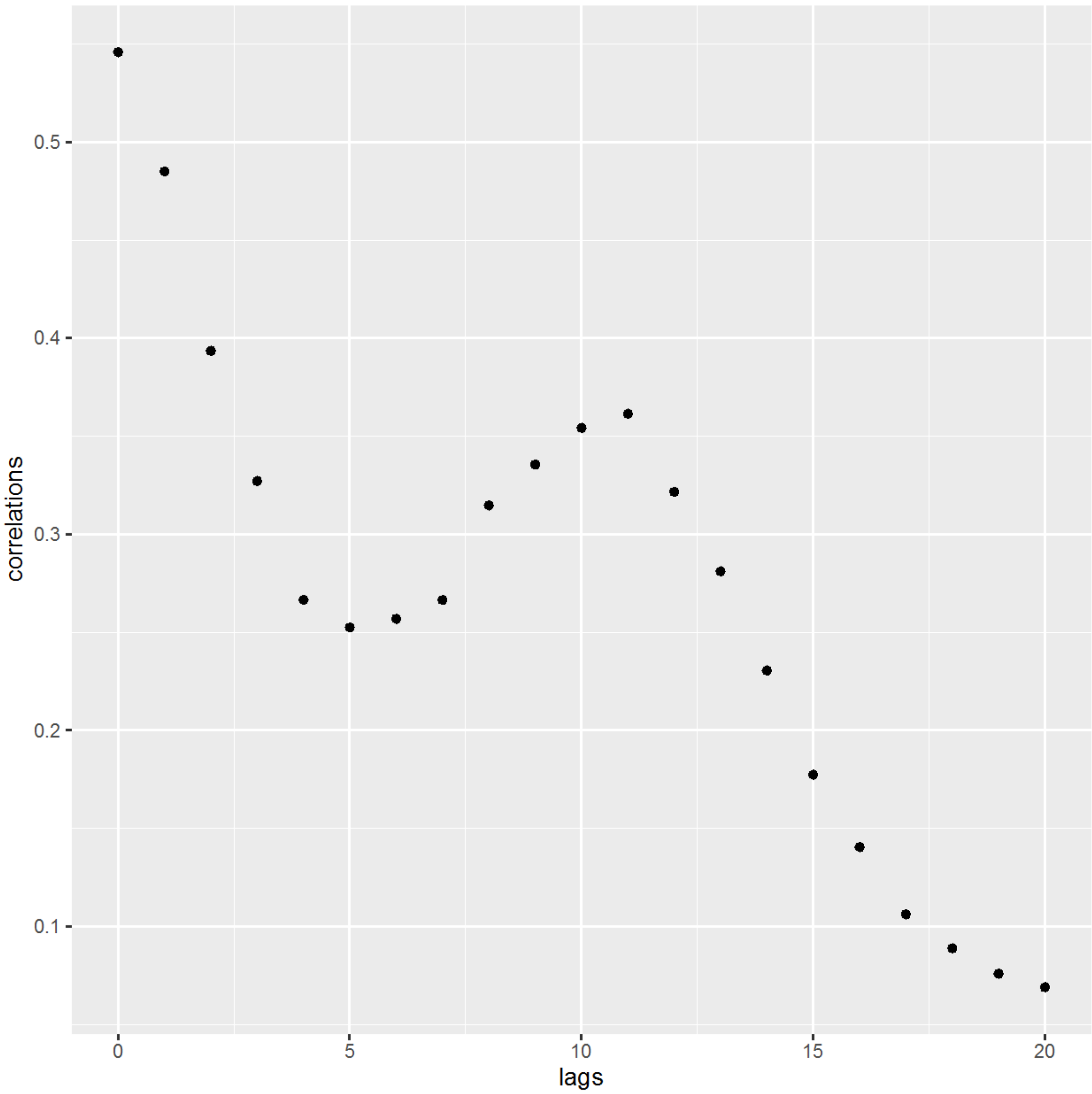
Again, in Cyprus, the highest correlation between Confirmed cases and Deaths, occurs after **1 day** that people have been reported as new cases.

Let's run the same analysis by taking into consideration the SMA 4.

```
df<-coronavirus%>%filter(country=='Cyprus', date>='2020-02-15')%>%select(date, country, type, cases)%>%
  group_by(date, country, type) %>%pivot_wider(names_from =type, values_from=cases) %>%ungroup()%>%
  mutate(confirmed = stats::filter(confirmed, rep(1 / 4, 4), sides = 1), death = stats::filter(death, rep(1 / 4, 4), sides = 1))%>%na.omit()

correlations<-c()
lags<-c(0:20)

for (k in lags) {

  tmp<-df%>%mutate(lagk=lag(confirmed,k))%>%select(death,lagk)%>%na.omit()
```

```
  correlations<-c(correlations,cor(tmp$death, tmp$lagk))
}

data.frame(lags, correlations)
data.frame(lags, correlations)%>%ggplot(aes(x=lags, y=correlations))+geom_point()
```

lags	correlations
0	**0.54587388**
1	0.48505075
2	0.39353514
3	0.32716838
4	0.26651528
5	0.25255863
6	0.25687756
7	0.26644043
8	0.31471813
9	0.33556331
10	0.35426576
11	0.36144415
12	0.32167056
13	0.28110514
14	0.23054363
15	0.17743614
16	0.14050113
17	0.10622712
18	0.08889594
19	0.07590604
20	0.06907297



Here, the maximum correlation observed on **the 0th day**.

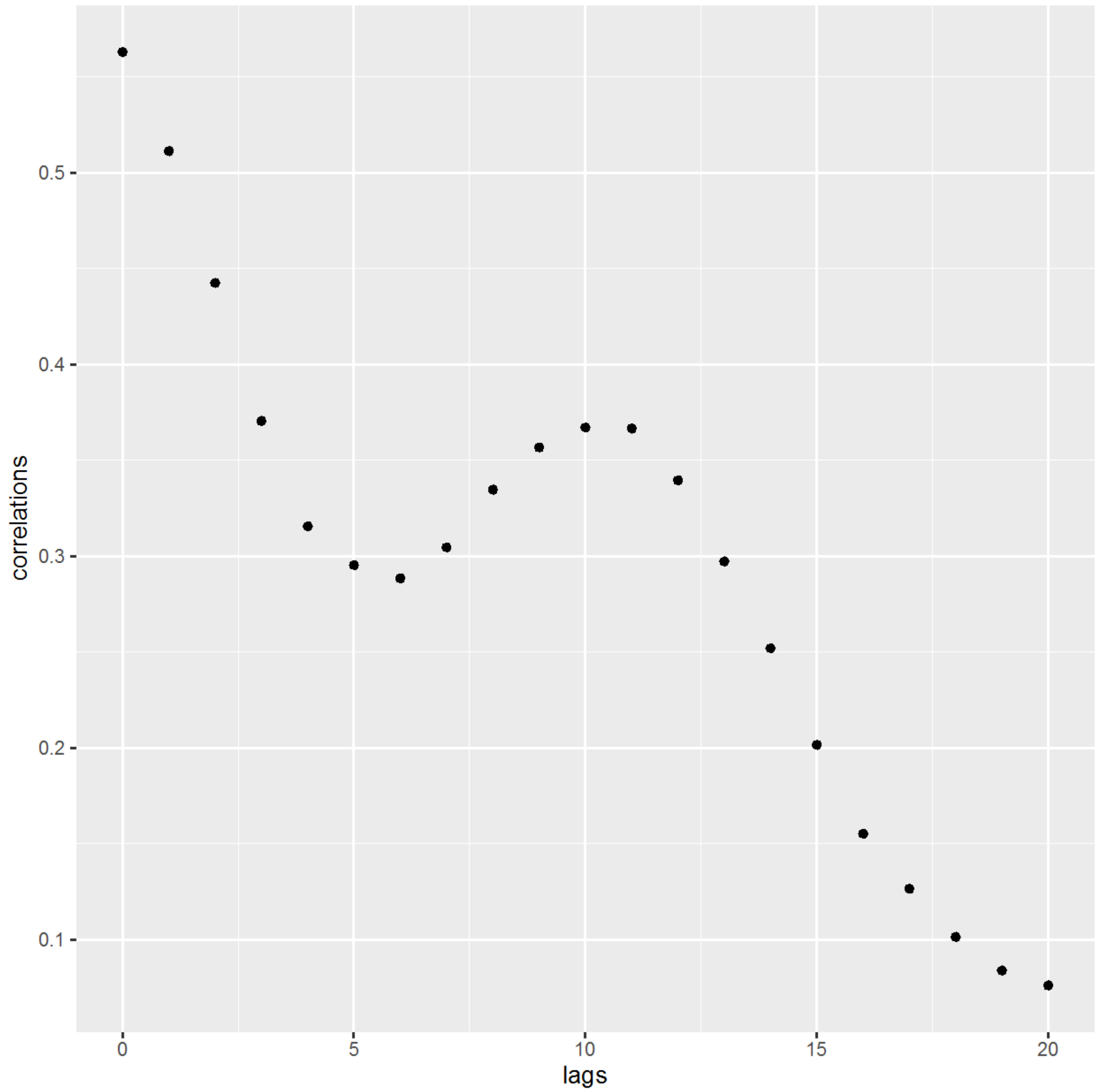
Finally, if we run the same analysis by taking into consideration the SMA 5.

```
df<-coronavirus%>%filter(country=='Cyprus', date>='2020-02-15')%>%select(date, country, type, cases)%>%
  group_by(date, country, type) %>%pivot_wider(names_from =type, values_from=cases) %>%ungroup()%>%
  mutate(confirmed = stats::filter(confirmed, rep(1 / 5, 5), sides = 1), death = stats::filter(death, rep(1 / 5, 5), sides = 1))%>%na.omit()

correlations<-c()
lags<-c(0:20)

for (k in lags) {

  tmp<-df%>%mutate(lagk=lag(confirmed,k))%>%select(death,lagk)%>%na.omit()
```

```
  correlations<-c(correlations,cor(tmp$death, tmp$lagk))
}

data.frame(lags, correlations)
data.frame(lags, correlations)%>%ggplot(aes(x=lags, y=correlations))+geom_point()
```

The result is the same, i.e. that the maximum correlation occurs on the 0^{th} day.

lags	correlations
0	**0.56288763**
1	0.51122698
2	0.44248618
3	0.37057832
4	0.31562168
5	0.29529844
6	0.28848165
7	0.30459523
8	0.33467750
9	0.35673843
10	0.36710924
11	0.36655890
12	0.33961842
13	0.29727072
14	0.25195790
15	0.20158851
16	0.15526497
17	0.12656945
18	0.10150166
19	0.08400634
20	0.07622187



This analysis is not valid because we lack much of the information about the way of collecting and reporting the data. However, it is clear that there is a lag between the Confirmed cases and Deaths but we cannot specify the number accurately.

Sources:

1. Johns Hopkins Report on COVID-19 (run daily), based on data of 31/10/2020.
2. Pipis, George, (2020), "Covid-19: Correlation Between Confirmed Cases And Deaths", 6, September.
3. The Actuary, (2020), "Huge Pandemic Insurance Gap Uncovered", October.