Harinandan Srikanth

Stat 153 - Professor Fisher Dated 4/26/2020

Time Series Analysis of Covid-19 Infections in Timeseria

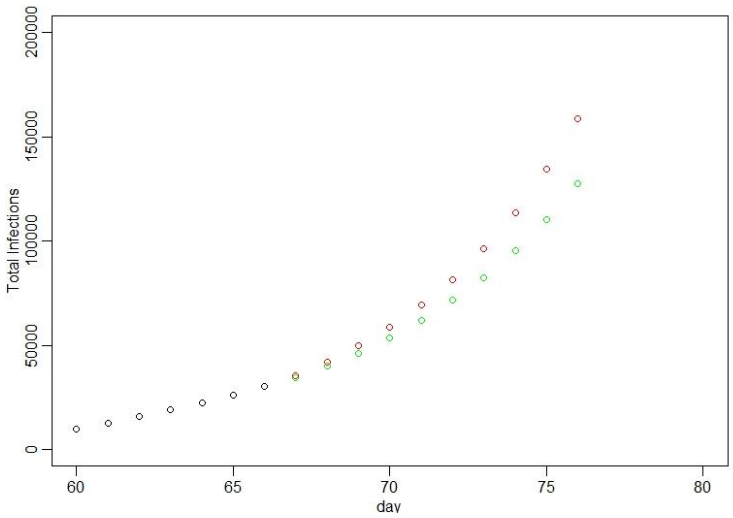
The COVID-19 outbreak that has blown up over the past few months, massively disrupted society, and claimed hundreds of thousands of lives, is a once-in-a-century type of event. Predicting the total infections with reasonable accuracy is critical for decision makers and the administration so that the appropriate steps can be taken to protect the population. In this study, I use the tools of data science to find the appropriate model to predict how the total infections will change in the near future. Given that the trajectory of total infections for the next days is highly dependent on the recently prevailing conditions, it is important that we use data that is most recent for the prediction. In my investigation, I found that the best model, meaning the one with the least possible prediction errors, is the one that uses the most recent data. In my analysis, I conclude that the latter model is superior. I predict that over the next ten days, on the average the total infections will increase to 158,443 which represents an increase of 128,327 from today. This represents the five-fold increase in the next ten days.

While this appears to be a shocking rate of increase, it is not all that surprising given that pathogens spread through the community at an exponential rate. Let me explain exponential growth if one person infects two individuals in a given day. This is known as the reproduction rate per day. If one person has the infection on day 0 and gives the infection to two people by day 1, the next day these two people will infect two other people each by day 2, and so on. At this rate, on day 20 a million people will have the infection, and on day 30 a billion people will it. As you can see, community spread of any pathogen will blow up tremendously and rapidly. If Timeseria has to prevent a catastrophe, we need to act now and reduce the rate of reproduction by ordering a lockdown. Based on the currently prevailing dynamics, the rate of reproduction per day in Timeseria stands at 1.12. If this number can be brought down below 1, then we can be assured that the spread of the disease will be successfully contained.

Why is it important to use the most recent data? I demonstrate this on my model by using the entire data available as opposed to using the last 30 days of data. It is well-known that in the early days, the disease spread was local to a few families. But in the last 30 days, it has turned into a communal spread. Shown below is the prediction for two different conditions of the model:

Harinandan Srikanth

Stat 153 - Professor Fisher Dated 4/26/2020



Currently, we are on day 66 of the disease spread, show by the last black circle. From then on, we see the predictions for the next ten days by two models: one that uses the most recent data (red circles) and one that uses all of the data available (green circles). The more appropriate model predicts 30,000 more infections and this provides a more conservative number for decision makers.