Pandemic Modeling Inquiry

Simon Lee

1 Motivation

Everywhere you go on the news as of late, we have heard that the spike in Coronavirus cases will drop once we reach the end of winter and head into the spring. Having been in the pandemic for nearly two years with access to vaccinations for over a year, there is a level of interest to see if this bold statement is indeed true. Because there has been plenty of press claiming an end to the pandemic and a rise to a seasonal "flu-like" endemic, Gov. Gavin Newsom of California has announced the lifting of mask mandates indoors. Whats felt like an absolute whirlwind these past two years might actually be coming to an end?

In this research proposal we want to test if seasons do in fact play a role in COVID-19 cases. It has been proven by scientists that ultra violet lighting kills airborne particles of the Coronavirus. It has also been studied by scientists that viruses thrive in colder temperatures. This leaves us scientists to wonder to see whether seasons and temperature do play integral parts in case numbers. Therefore we are interested in the COVID-19 cases of a northern hemisphere region (California, USA) and a southern hemisphere region (South Africa) to see if there are variations in spikes based on their individual seasons.

2 Methods and Deliverable

Luckily for us, News outlets like *NY times* and *LA Times* have made public data sets of all types readily available for use on GitHub. Therefore we intend to use *LA Times* data set to track number of cases in California starting from March 1st 2021, to Feb 10th 2022. (https://github.com/datadesk/california-coronavirus-data/blob/master/cdcr-state-totals.csv). In terms of our southern hemisphere region, we were lucky enough to also find a public dataset for South Africa which tracked total case numbers and with some simple spreadsheet manipulation to calculate case by day we are able to construct data that we can compare to our California dataset with the same dates. (https://github.com/dsfsi/covid19za/blob/master/data/covid19za_provincial_cumulative

In this research I wish to run a few statistical tests like curve fitting using Poisson or Gaussian Distribution to see trends within the data. In addition to trying to curve fit, I intend to also take some classic statistical measures like arithmetic mean, standard deviation, & variance to get more telling data to analyze in the paper. My intentions by curve fitting is to see whether indeed there has been some seasonal effect on COVID-19. These statistical tests have also been used to predict future trends so I think this is my main focus in trying to fit this using these distribution.

In terms of data that will be produced, I intend to generate 2 primary graphs with four subplots that contain the cases data by days with its respective seasonal distributions. I also plan to include a table of the other statistics (arithmetic mean, variance, etc.) to further analyze this in the paper. These are the main graphs that I wish to produce but I may also do a graph comparing trends with the two countries with the same seasons as well to see if the trends themselves are consistent across seasons to determine whether season does play a role. This I intend to do in four different graphs split up by season or just one big plot where season is color coded.

Through this data set and by experimenting with these statistical measures, I think I can have a good measure to answer the existing question of whether seasons play a role in COVID-19 cases. By taking two different samples from two different hemispheres, although not encompassing at all, may give a generalization of whether seasons play a role. I am rather excited and motivated to be running a set of experiments of my choice, and I cannot wait to get further into this question.