Shanni Lam and Xintong Wang
Math 189Z
24 April 2020

Final Project Proposal

Introduction

For our final project, we will model the true number of COVID-19 infections in the United States; in the United States, we seem to be undertesting our population. We have seen from other parts of the world that many people who have COVID-19 are asymptomatic, so they are undercounted, and in the United States, even people with light symptoms are discouraged from testing themselves to put less work on the hospitals, so this lack of ability for most people to test for COVID-19 means that the current number of confirmed cases in the United States does not truly capture the actual number of people infected.

Before creating a model, we will consider the prevalence rate vs. the positive test rate, and we will do preliminary research on why less cases are reported than there actually are, such as the delay in news reporting. Using different transmission rates proposed by various literature, we would like to explore how many actual cases of COVID-19 each confirmed case represents so that we can more accurately model the growth of COVID-19 in the United States. In addition, we will look into data from countries that have imposed widespread testing to get a better sense of the true growth rate of infections, while adjusting for factors that may influence the growth rate, such as the strength of social distancing policies and healthcare policies even among different states.

Methods

Topic modelling does not seem necessary to estimate the number of COVID-19 cases because we are looking for numerical data, not topics; instead, we should use the techniques we learned from Homework 1 to fit linear regression models. We would also like to explore other types of regressions, such as polynomial, logistic, and ridge, to see what type of regression best fits our data.

Helpful Sources

Articles:

High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2
How Bad Is the Coronavirus Outbreak? Here's a Key Number.
Estimates of the severity of coronavirus disease 2019: a model-based analysis
COVID-19: Study estimates rate of 'silent transmission'
Coronavirus data in South Korea is world's most comprehensive: Its lessons so far

Official Counts Understate the U.S. Coronavirus Death Toll

Data:

mrc-ide/COVID19_CFR_submission: Repository for all scripts required to replicate the CFR analysis for paper submission.

Coronavirus Disease (COVID-19) - Statistics and Research

• South Korea: coronavirus infections, recoveries, and test cases 2020