Covid19: Analysis of Accumulated Cases in Japan with insufficient Test Cases

Shiori Uji
May 27th, 2020

1 Introduction

1.1 Background

In January 2020, the world suddenly knew about the existence and its effect of COVID 19, so called coronavirus. Its case was mainly reported in Wuhan China in the early stage of pandemic, and spread to the whole world immediately. It killed hundreds and thousands of people, damages the world economy. Hospital quickly got full and lots of lives were taken because of shortage of tantalizer. Citizen rush to purchase masks, foods and even toilet papers from the fear of shortage. In order to handle this situation, hospital managers have to plan which patient to let hospitalized in their facility, government need to prepare financial support for small businesses, and firms have to forecast demands of tantalizers and masks to supply markets. How the number of infected people increases is one of the key parameters for everyone who would like to predict future governmental and social policy. For instance, policymakers and managers who had to prepare for the production of medical resources and allocation, restaurants, and shop owners have to calculate how much cash they have, and how long they can keep their business under a curfew of not going out. Predicting the change of infected cases is critical to saving patients, policymakers, and business owners to keep their lives, social stability, and their lives.

1.2 Business Problem

Decision-makers such as business leaders, doctors, and policymakers would like to know how patients of coronavirus rise for both medical and business purposes. Others who are seeking their job, students who are watching for their entrance exam, family members who have to take care of other members will also be interested in their decision making. The target of this research is to expect a theoretical number of infected persons.

1.3 Data set

The target of this research is to expect a theoretical number of infected persons.Data that might contribute to determining the number of new patients of coronavirus per day include reported cases per day per country and fatalities
per day per country. This is a time-series data set, starting from January 21st, 2020. I will use foursquare API to acquire locational data to visualize the inter-country difference. All data used for this analysis was taken COVID19 Global Forecasting (Week2) competition held at Kaggle. This train.csv contains the actual number of confirmed cases and fatalities in each countries.

1.4 Necessity of this research

If the PCR testing kits are in shortage, the government cannot always grasp the actual number of infected people anymore. I have a strong interest in Japan, where there are not enough test kits as Korea and Singapore. As of April 31st, the Japanese government recognizes 14088 cases and 415 Fatalities, which is not as large as Italy and France, however, only 3531 people could take PCR testing at one day. In order to estimate trends of patients increase, I will model the data set of Japan. Also, I will investigate other major country to compare their trends with Japanese case.

2 Data acquisition and cleaning

2.1 Data sources

Time-series data of confirmed cases and fatalities by coronavirus per day per country (2020) from Kaggle dataset, data scraped from COVID 19 Global Forecasting(Week2).This train.csv in this competition data set contains the actual number of confirmed cases and fatalities.

2.2 Data cleaning

There was almost no missing values, however, some countries like US counts their cases regionally whereas other countries count by country. In order to compare differences in nations, I summed up all the regional data to one country.

2.3 Feature selection

After cleaning this data set, three are 20580 rows and 6 columns. It contains 5 features(Id, date, Country, Confirmed Cases, Fatalities, confirmed cases, and fatalities) in 173 countries from January 22nd to March 31st(70 days).

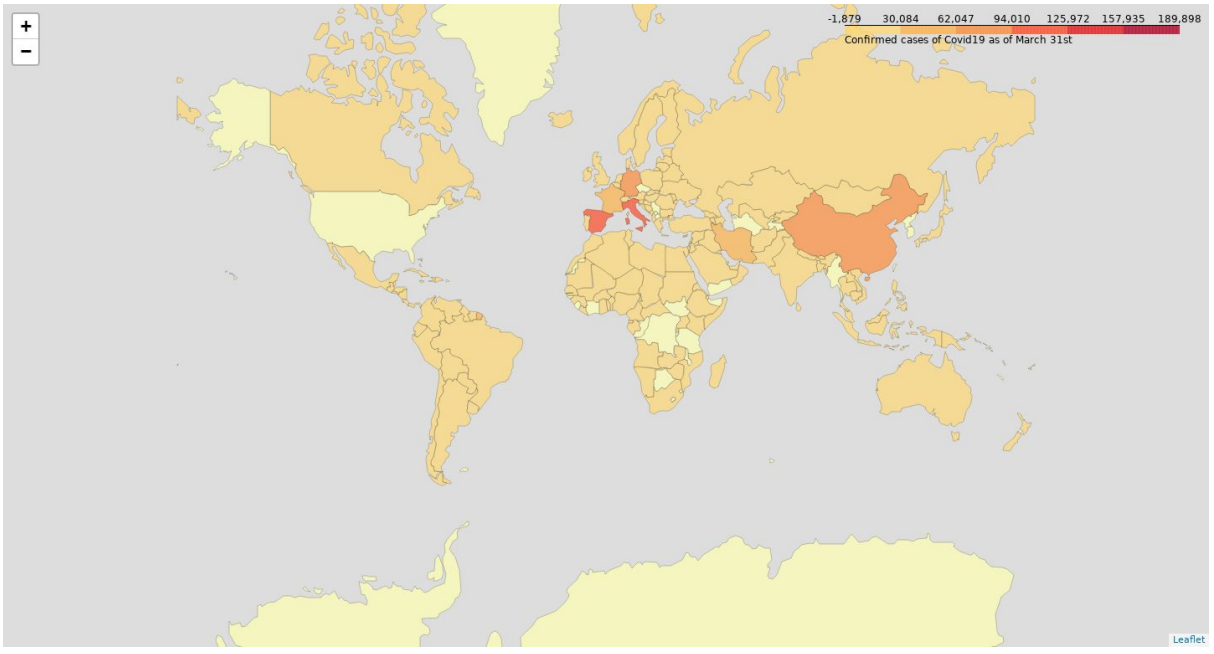
3 Exploratory Data analysis

3.1 Purpose of explanatory data analysis

As of March 31st, over hogehoge countries reported confirmed cases, and hoeghoge countries have fatalities. In order to grasp the overall trends of Covid-19 spreads and suitable countries to compare with Japan, I visualized the important information first, and dig deep into several countries.

3.2 Major countries struggling for Covid-19

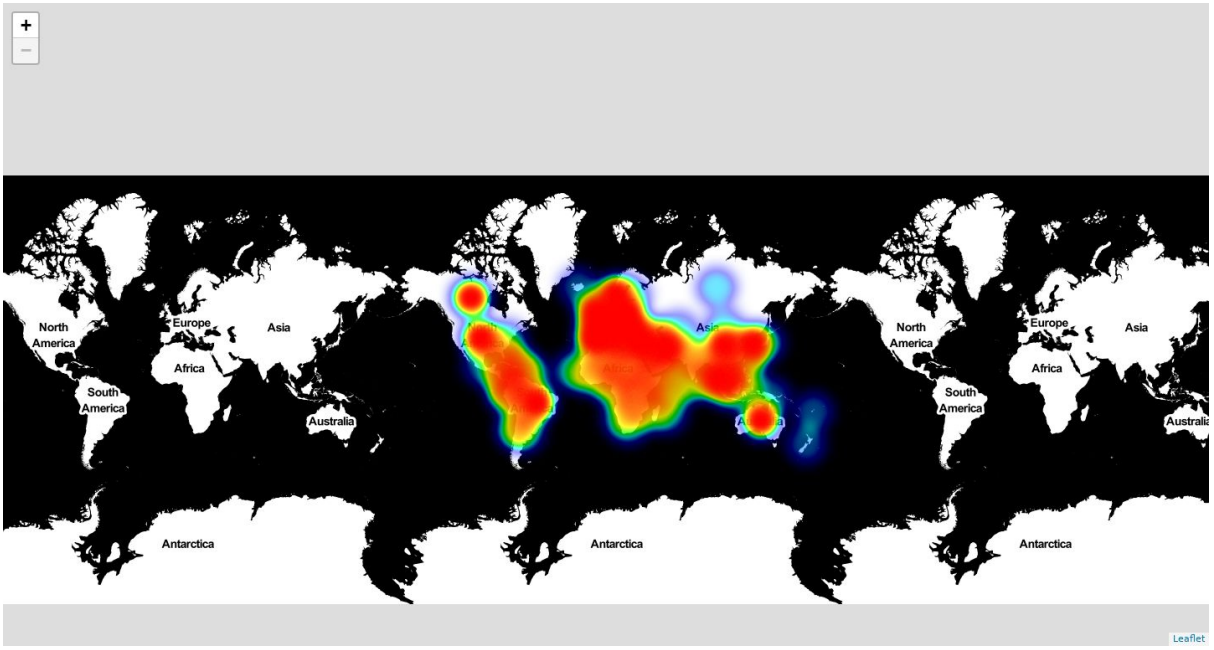
It was widely reported that coronavirus spread rapidly in China and Italy by the end of March. However, the result of this visualization shows coronavirus spreads all over the world.



3.3 Inter-nation comparison of mortality rate

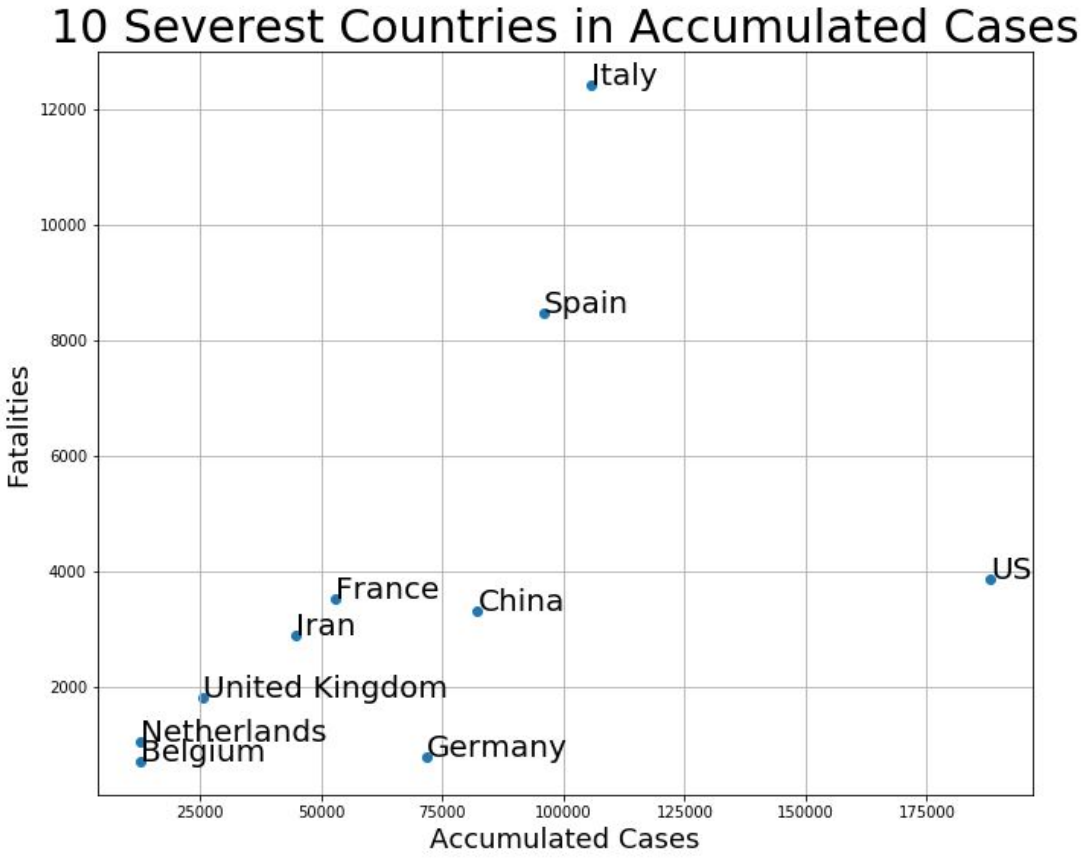
I hypothesized that the mortality rate will be high in some countries where there

are poor medical resources even if the nations' average age is low and that was mostly true. Italy experiences a fairly high mortality rate that will be rooted in the rapid spread of the virus.



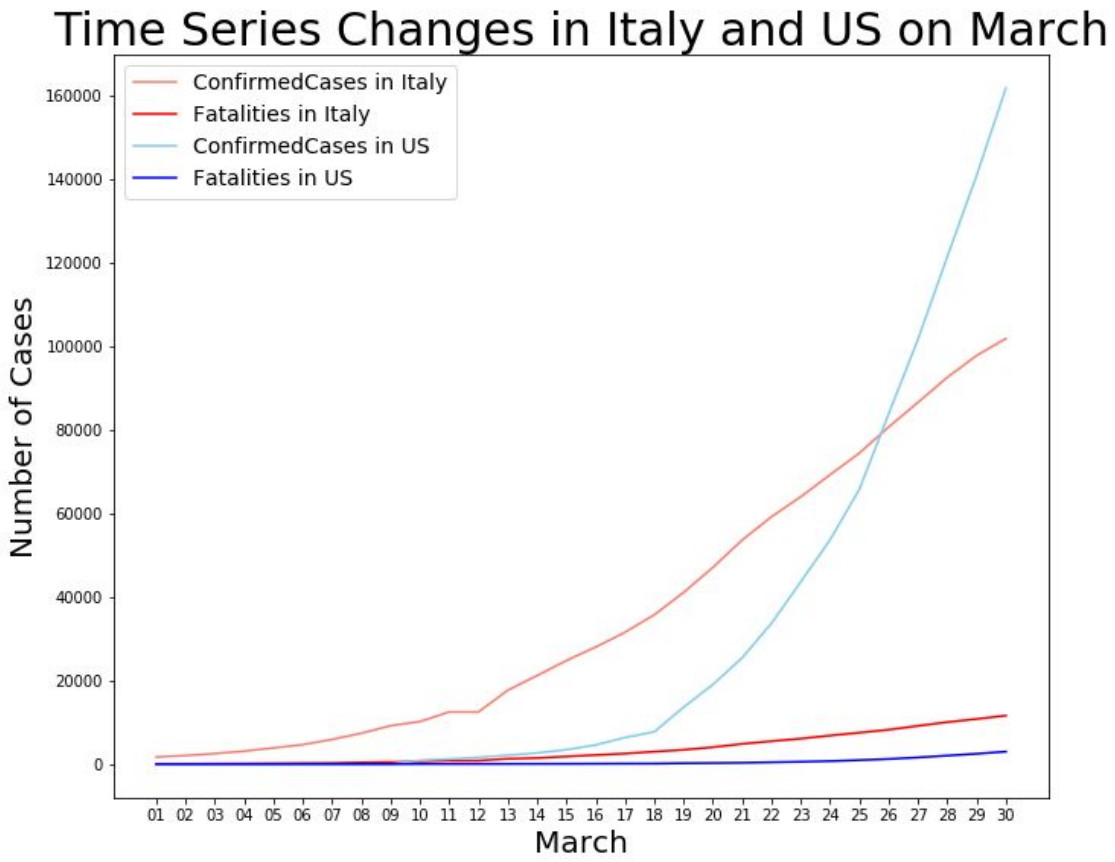
3.4 Comparison of 10 countries

Since I would like to choose countries to compare with Japan, I focused on Italy and Korea, where there are fairly enough medical resources and have experienced an increase in patients. Especially Korea and Japan are close in economy and culture, Korea has a sophisticated medical system against infectious diseases such as preparations of enough test kit and administrative system to stop virus spreads after experiencing SARS in 2003.



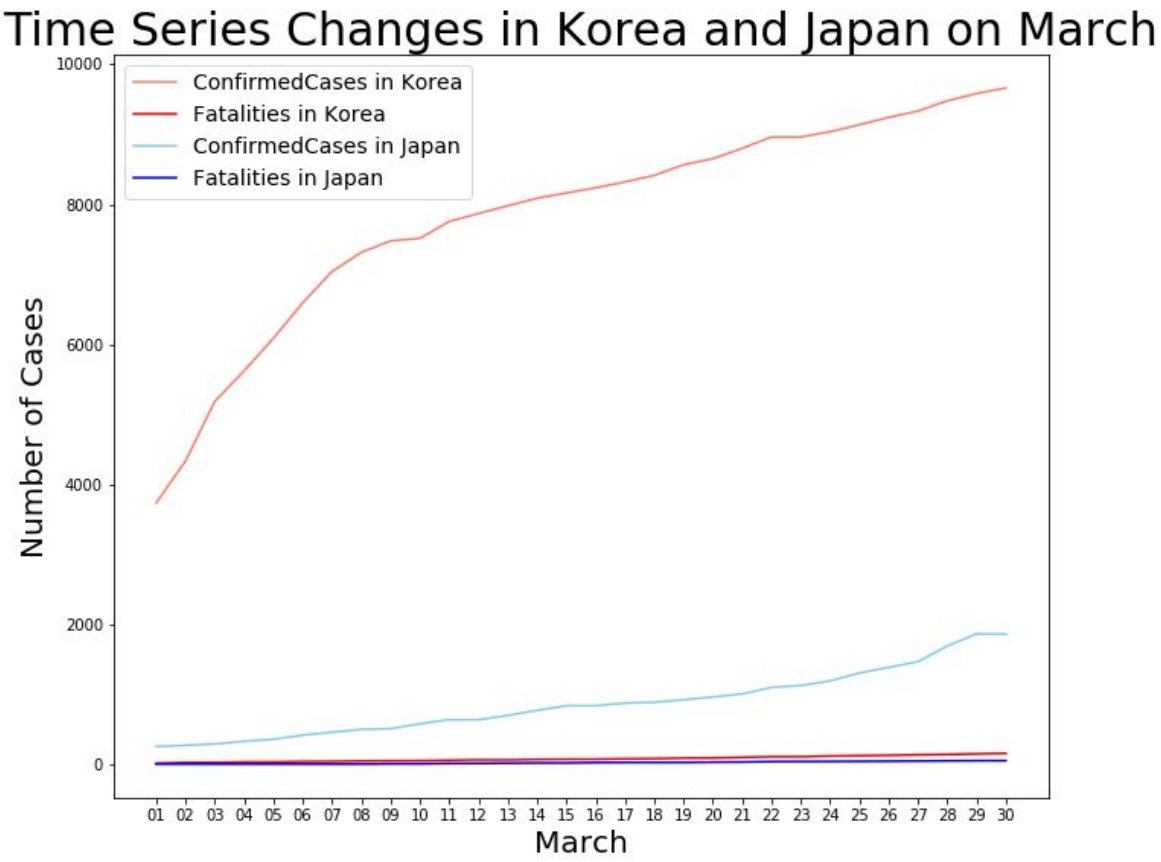
3.5 Italy and the US

Then, I focused on two severest countries, that is the US and Italy. This two countries experience exponential increase in Confirmed Cases.



3.6 Japan and Korea

Although the number of fatalities in South Korea and Japan has generally followed a similar trend, the number of reported infections is nearly ten times different.



4 Predictive modeling

There are two types of models, regression and classification. In this study, I carried out regression modeling to predict the future change of cases. I consider simulating the speed of the virus spreads with several models such as sigmoid and exponential model and compared their performance with RMSE(root mean squared error) as tuning and evaluation metrics.

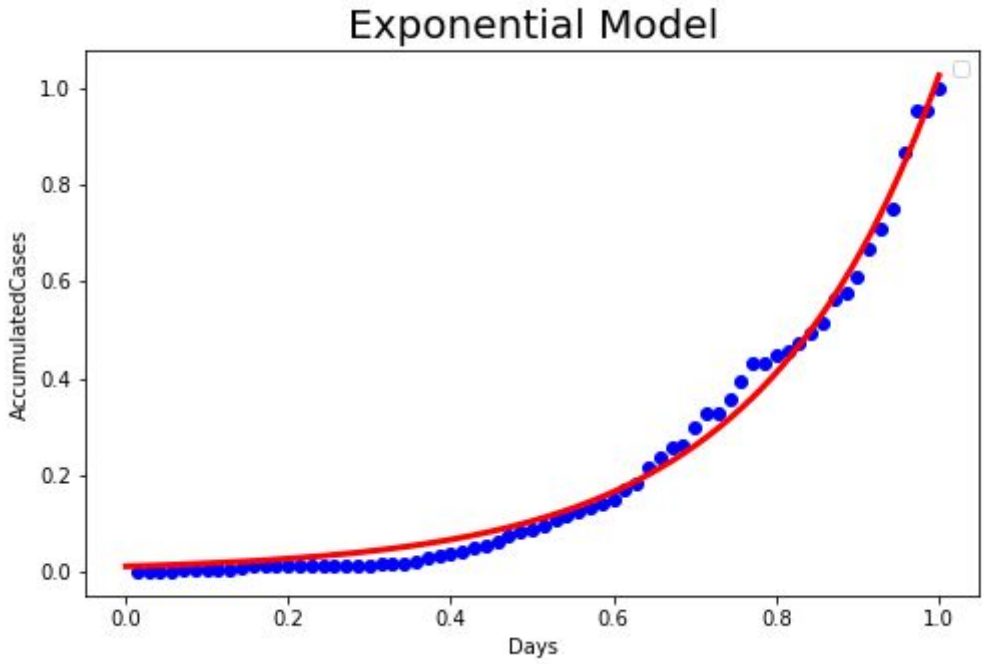
4.1 Performances of different models for Japan

Here is the comparison of 4 models, this implies that the exponential modeling performs best than other models in terms of R2 score and mean absolute error

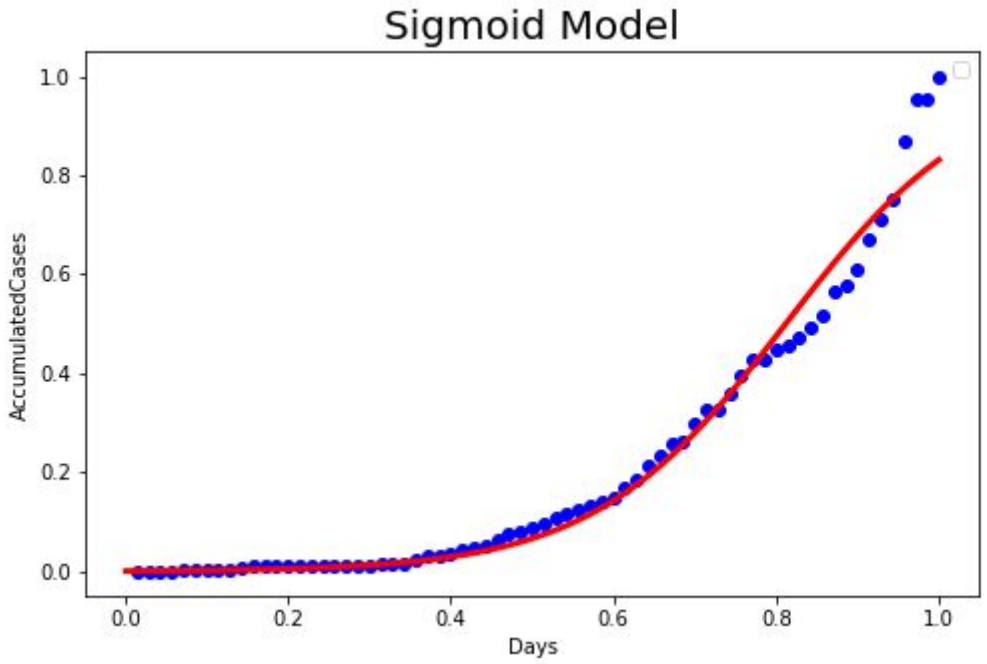
	Exponential Model	Sigmoid Model	Single RG	Polynomial RG
Mean absolute error	0.02	0.03	214.14	70.61

Residual sum of squares (MSE)	0.00	0.00	94996.57	11700.58
R2-score	0.99	0.00	0.32	0.96

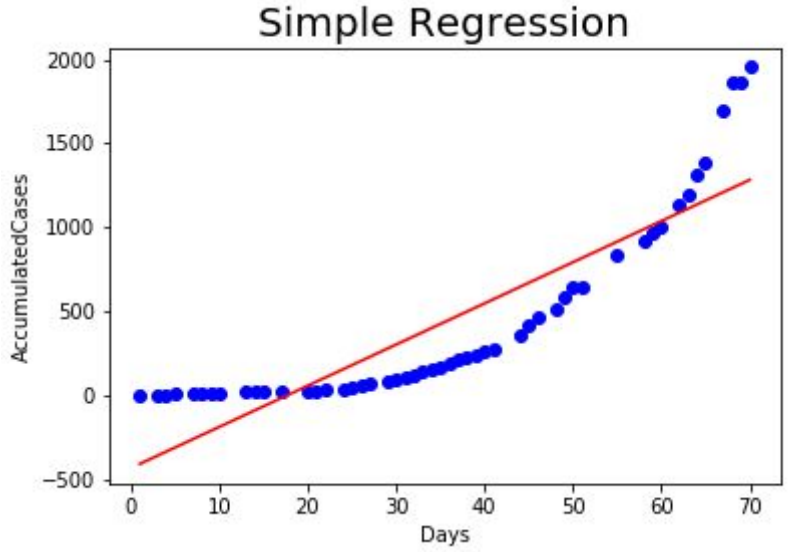
4.1.1 Exponential Modeling



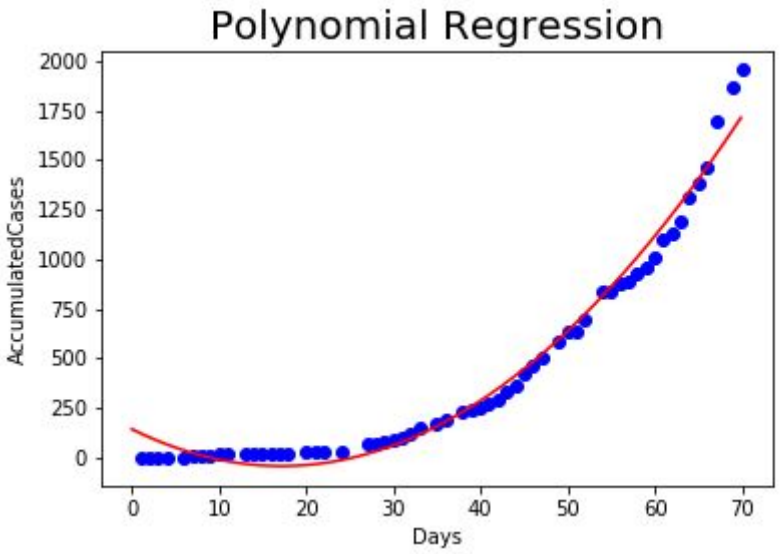
4.1.2 Sigmoid Modeling



4.1.3 Single Regression



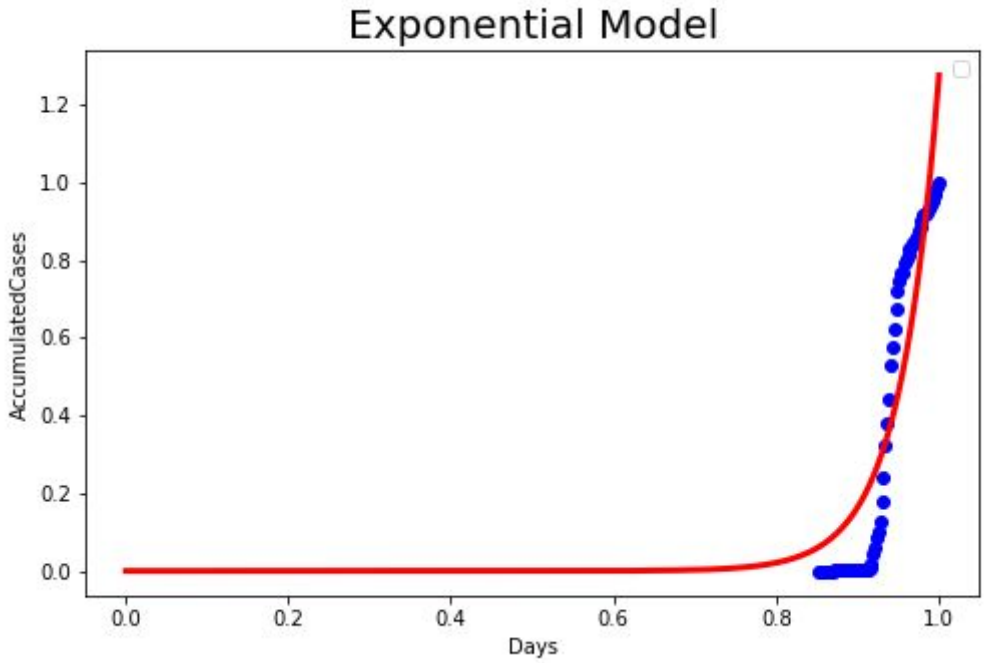
4.1.4 Polynomial Regression



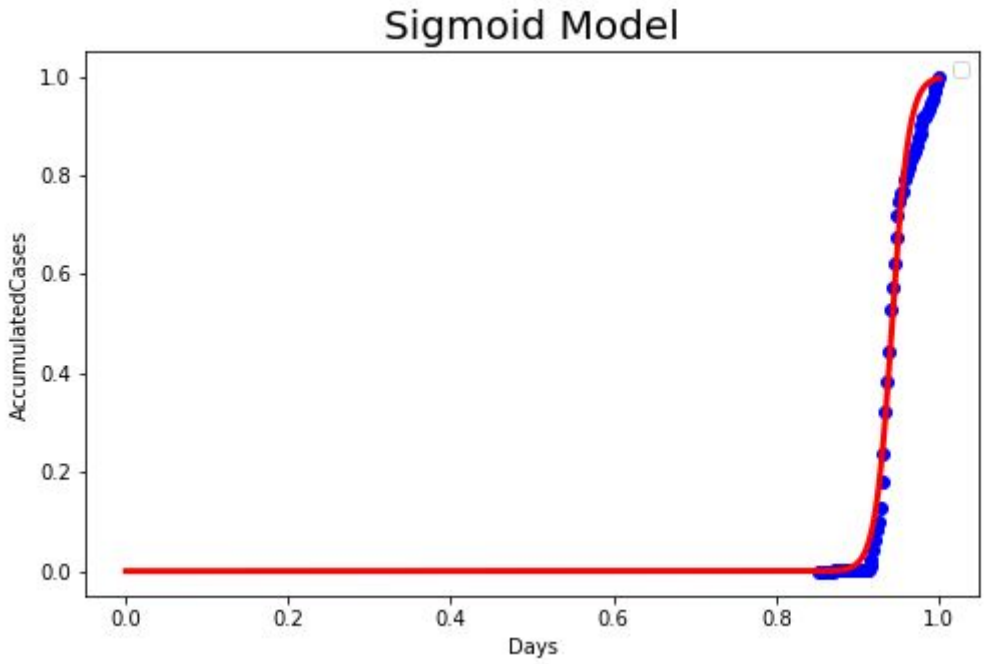
4.2 Performances of different models for Korea

A model of the number of infected people in Korea was also created for comparison.

4.2.1 Exponential Modeling



4.2.2 Sigmoid Modeling



4.2.3 Single Regression

	Exponential Model	Sigmoid Model	Single RG
Mean absolute error	0.16	0.03	1022.42
Residual sum of squares (MSE)	0.03	0.00	1930033.46
R2-score	0.62	0.99	0.84

5 Result

Despite the coincidence of death trends in South Korea and Japan, there are extremely few cases reported in Japan.
While it is extremely difficult to predict the exact number of infected people, Japan and South Korea may be staying pretty close.

6 Discussion

In response to media criticism of the low number of tests in Japan and strong public demand, the government and medical professionals are working together. While the lack of testing is a problem that should be resolved, the expansion of testing should not make already exhausted medical personnel even more exhausted, and the fact that the lethality rate is very low while the global epidemic is exploding should be a credit to the efforts of Japanese medical personnel and the medical system.

7 Summary

While Japan is trending closer to South Korea in the number of deaths, the number of tests is very low and it is estimated that there were many potential patients missing out on tests as of this March.

8 Conclusion and future direction

Future developments include the comparison of April's estimates with actual results and the use of specialized medical services such as the SIR model.
While it is important to increase the number of tests and make predictions in order to prevent a pandemic, medical systems vary greatly from country to country, so it is essential to provide support tailored to each country's situation.
For this reason, it will be necessary to carefully review the information and reports of each country, rather than mathematical modeling.