EPIDEMIOLOGIC MODELS OF COVID-19 DYNAMICS IN NEW JERSEY COUNTIES

MORANCHEL, PEDRO
LAHBAN, COURAGE
HUDSON COUNTY COMMUNITY COLLEGE
JERSEY CITY, NEW JERSEY

LI, AIHUA
DEPARTMENT OF MATHEMATICAL SCIENCE
MONTCLAIR STATE UNIVERSITY
MONTCLAIR, NEW JERSEY

Mr. Pedro Moranchel
Mr. Courage Lahban
Hudson County Community College
Jersey City, New Jersey

Aihua Li
Department of Mathematical Science
Montclair State University
Montclair, New Jersey

Epidemiologic Models of COVID-19 Dynamics in NJ Counties

Synopsis:

This research focuses on the dynamics of COVID-19 spreading in New Jersey counties. Two popular models were applied to simulate the original data, recover the missing part, and predict future situations: the SIR and the logarithmic model. Differences between the projected data and the actual data is analyzed. The constructed models help to better understand the virus spreading and to gain insight of the shared or distinct patterns presented through these counties.

Epidemiologic Models of COVID-19 Dynamics

Courage Lahban, Pedro Moranchel, Hudson County Community College, NJ, USA
Aihua Li, Montclair State Università, NJ, USA

Abstract.

The project focuses on the dynamics of COVID-19 spreading in New Jersey counties. Data were collected from March 15, 2020 to July 15, 2020. Due to various reasons, the original data has missing parts. Two epidemiologic models, SIR and Logarithmic models, are developed to simulate the original data, recover the missing part, and predict the future situations. For each method, the difference between the projected data and the actual data is analyzed. The error functions are calculated and the comparison study between the two methods are performed.

1. Introduction

1.1. Background
In 2019, an outbreak of a severe acute respiratory syndrome described as (SARS-CoV-2, and later Covid-19) was identified in Wuhan city, Hubei Province, China. On January 22, 2020, the World Health Organization confirmed the human-to-human spread of the novel coronavirus [1]. This has since been the most prominent mode of spread of the pandemic. As of 2020 methods of deterrence involve the implementation of distancing and isolation procedures for exposed or infected individuals. Numerous infections of this variant were recorded in the United States, with almost all states recording significant numbers of infections as early as February 2020. In the state of New Jersey, numbers varied county by county, with some recording different disparities in the number of infections and deaths.

An infection often involves a cycle of three stages: sources or host, susceptible, and transmission. Source is a place where bacteria or viruses live (host) such as human, surfaces, and so on. For an infection to occur, germs must enter a susceptible person's body and invade tissues, multiply, and cause a reaction [2]. The motivation for this research is from the general interests in the dynamics of COVID-19 spreading in New Jersey counties. The result of this research is from an REU program held in summer 2020 at Montclair State University. Due to the corona virus situation, this research was conducted remotely.

1.2. Research Questions and Goals
Throughout, local corona virus data were available on the internet. But due to various reasons, the original data has certain missing parts. It is our goal to apply some mathematics modeling methods for infectious disease to simulate the original data and to project real situations. The aim of this project is to study the growth and features of COVID-19 using the 21 counties of New Jersey as a case study.

The research goals are summarized as below:
- Use different epidemic models to simulate the Covid-19 dynamics in New Jersey.
- Obtain a better picture of the coronavirus situations in New Jersey counties;
- Predict future values or missing situations by the constructed models;
- Measure the error function between the projected data and the real situation.
- Compare the performance of different models.

1.3. Methodology and Main Development
We first apply the popular SIR model to simulate the original data, recover the missing part, and predict future situations. Second, we use a logarithmic modeling method on the same set of data. For each method, the difference between the projected data and the actual data is analyzed. The constructed models help to better understand the virus spreading and to gain insight of the shared or distinct patterns presented through these counties. The error functions are calculated and the comparison between the two methods are performed.

2. Data and Resources

The data was collected through a publicly available COVID-19 tracker website [3] that had detailed daily and weekly information on COVID-19 throughout the United States. It showcased COVID-19 data through each of New Jersey's 21 counties in a convenient matter, which made it our preferred data source for the duration of the project. Through the generosity of some of their developers, we were able to access the files in a more readily available format.

We organized all the data through Microsoft Excel. For each county we collected data from March 4th, 2020, to July 15th, 2020, mainly centered around the total confirmed infected cases and recovered cases. The data such as race were also collected for potential future work. We utilized the software Cocalc's Python features to perform the calculations and modeling process and to provide the graphical representations of the results. Data on total deaths, daily deaths, 7-day average confirmed and 7-day average deaths were also examined in the similar way for all New Jersey counties. Note that data of recovered and susceptible cases are not available in New Jersey.

The following table shows partial data of the first 20 days for Hudson County.

Dates	Total Confirmed	Total Deaths	Daily Confirmed	Daily Deaths	7 Day Avg Confirmed	7 Day Avg Deaths
03/15/2020	5	0	0	0	0.6	0
03/16/2020	19	0	14	0	2.6	0
03/17/2020	24	0	5	0	3.3	0
03/18/2020	34	1	10	1	4.7	0.1
03/19/2020	55	1	21	0	7.7	0.1
03/20/2020	66	1	11	0	8.9	0.1
03/21/2020	97	1	31	0	13.1	0.1
03/22/2020	126	1	29	0	17.3	0.1
03/23/2020	190	1	64	0	24.4	0.1
03/24/2020	234	1	44	0	30.0	0.1
03/25/2020	260	1	26	0	32.3	0.0
03/26/2020	441	2	181	1	55.1	0.1
03/27/2020	594	2	153	0	75.4	0.1
03/28/2020	771	4	177	2	96.3	0.4
03/29/2020	974	6	203	2	121.1	0.7
03/30/2020	1314	6	340	0	160.6	0.7
03/31/2020	1606	11	292	5	196.0	1.4
04/01/2020	1910	29	304	18	235.7	4.0
04/02/2020	2270	44	360	15	261.3	6.0
04/03/2020	2835	59	565	15	320.1	8.1
04/04/2020	3491	78	656	19	388.6	10.6
04/05/2020	3924	87	433	9	421.4	11.6

3. Simulations by Two Mathematical Models

Two traditional mathematical modeling methods were applied to simulate the real situation of the covid-19 spreading in New Jersey counties: the SIR model and the logarithmic modeling. Python codes were developed to run and simulate data by two modeling methods shown below.

A. The SIR Model.
Following the pioneering work by Kermack and McKendrick [6], the SIR model has been widely used as an effective method in predicting the outbreak period of infectious diseases. In this study, the SIR model is used to analyze critical transitions and parameters of the covid 19-epidemic and

to study the change of state variables from a time series parameters. An SIR model involves three-state variables, susceptible individuals $S(t)$, infected individuals $I(t)$, and recovered individuals $R(t)$. For a population in consideration of size N, $S(t) + I(t) + R(t) = N$. We need to determine meaningful model parameters for these state variables, which provide propagation of the disease transmission across the given population segments (the New Jersey counties).

To determine the dynamic of the module, we use a system of mathematical expressions to show the relationship between the state variables, (the transition from susceptible individuals to the infected or recovered individuals). Since there are only two transitions to be observed, only the rates of two changes are modeled for the population: the rate of infection for susceptible individuals and rate of recovery for recovered individuals. The model assumption is that individuals in all groups have an equal probability of contact or exposure.

Our model is demonstrated by 3 differential equations:

$$\frac{dS}{dt} = -\beta S(t)I(t) \qquad (1)$$
$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t) \qquad (2)$$
$$\frac{dR}{dt} = \gamma I(t), \qquad (3)$$

where β is the infection rate and γ is the recovery rate.
- In Equation (1), the number of susceptible individuals decreases at a rate proportional to the number of susceptible individuals multiplied by the infected individuals.
- Equation (2) shows two components. Firstly, the rate of increase for the number of the infected individuals is the same as that of decrease in the number of susceptible individuals. The second term of the equation shows a decrease in the number of infected individuals due to the increase in the number of recovered individuals or recovery groups by $-\gamma I(t)$.
- From Equation (3), the rate of change of the recovered group is directly proportional to the rate of the infected group. Here the content of proportionality, γ, shows the transition from the infected group to the recovery group over a given time series.

We demonstrate this model for Hudson County. Figures 1 shows the graph of culminative daily infected population in Hudson County. Figure 2 demonstrates the SIR model for Hudson County on daily removed individuals (red, with an interior maximum), daily susceptible individuals (purple, decreasing with a steady ending), and cumulative daily infected cases (green, increasing with a steady ending). The rate factors we applied here are $\beta = 0.37$ and $\gamma = 0.07$. These two numbers are determined by the current transmission rate and experimental simulations from the data.

Figure 1

Figure 2
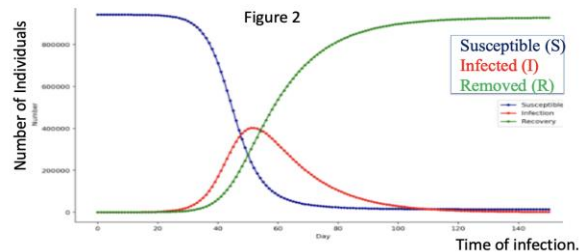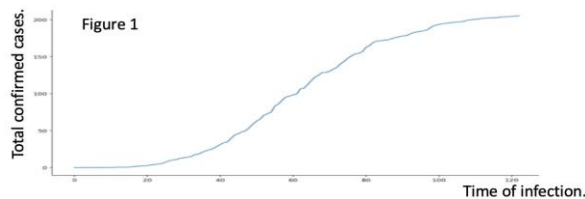
B. The Logarithmic Model

A more classical example of growth is modeled by the logarithmic model, through which the percentage difference of the projected data from the actual data is shown. The model is given by

$$y = ae^{-bx} + c; \qquad a = 0.3703, \, b = 0.0622, \, c = 0.0107.$$

Here a, b, c are determined by applying the least square model for the simulations of the real data. The model is demonstrated by Figures 3 and 4.

The logarithmic model was used to showcase Hudson County's progression over the duration of the data we collected. Notably, as seen by Figure 3, the projected data improved over time, having a lower percentage difference. This would be due in part to the smaller variation of the total confirmed case information over time, as the shorter time frame allows for larger daily variations of cases that could affect the overall percentage difference.
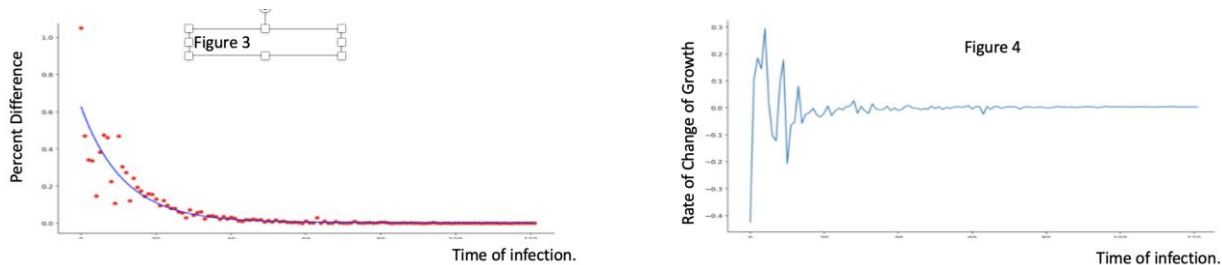

Figure 3 focuses on the relative increase rate in Hudson County, which is

$$\frac{\text{Current Day's Rate} - \text{Previous Day's Rate}}{\text{Previous Day's Rate}}.$$

It serves as a difference graph of the projected data and the actual data for Hudson County. Figure 4 demonstrates the error between projected values and the original logarithm values of daily relative increase of infected cases in Hudson County.

4. **Performance and Comparison of the Two Models**

The constructed models help to better understand the virus spreading dynamics and to gain insight of the shared or distinct patterns presented through these counties. Because the data of the recovered cases and the susceptible cases are not available in New Jersey, the logarithmic method shows a better simulation, especially after the first two weeks.

We also calculated the correlation matrix between every pair of New Jersey counties. It shows that all the correlations are mostly positive. Some pairs of counties have almost linear relation such as Hudson County and Bergen County. There are curves showing concave up or concave down pattern. For example, the correlation curve between Hudson and Atlanta shows a strong concave down pattern which means that the growth of the infected cases of Atlanta is much slower than that of Hudson's. For both modeling methods, Python codes were produced to simulate the data and produce the graphical presentations. The following table shows the correlation matrix between ten New Jersey counties.

	hudson	bergen	Essex	Monmouth	Union	Atlantic	Cumberland	Salem	Burlington	Gloucester
hudson	1.000000	0.638387	0.604216	0.669602	0.718935	-0.200125	-0.414515	-0.244025	0.430927	0.149091
bergen	0.638387	1.000000	0.831404	0.727767	0.801883	-0.390162	-0.497249	-0.250183	0.272423	0.034760
Essex	0.604216	0.831404	1.000000	0.647623	0.766622	-0.300523	-0.402962	-0.232953	0.282134	0.153795
Monmouth	0.669602	0.727767	0.647623	1.000000	0.706524	-0.277759	-0.396371	-0.172342	0.276949	0.088075
Union	0.718935	0.801883	0.766622	0.706524	1.000000	-0.370912	-0.412844	-0.218187	0.294151	0.093405
Atlantic	-0.200125	-0.390162	-0.300523	-0.277759	-0.370912	1.000000	0.377854	0.279880	0.133207	0.516531
Cumberland	-0.414515	-0.497249	-0.402962	-0.396371	-0.412844	0.377854	1.000000	0.343413	-0.072441	0.315771
Salem	-0.244025	-0.250183	-0.232953	-0.172342	-0.218187	0.279880	0.343413	1.000000	0.117054	0.254895
Burlington	0.430927	0.272423	0.282134	0.276949	0.294151	0.133207	-0.072441	0.117054	1.000000	0.290270
Gloucester	0.149091	0.034760	0.153795	0.088075	0.093405	0.516531	0.315771	0.254895	0.290270	1.000000

This research was done during the pandemic lockdowns. It was a unique experience for the student participants to conduct research remotely on a topic that attracted many people. For the faculty advisor, it was a valuable experiment in directing student research through zoom communications.

Future directions include the study what effects geography and economic factors can have on COVID-19 and other related disease spread. In other words, investigating impact by the ethnic, income, culture factors on the virus spreading.

Reference

[1] https://www.who.int/news/item/05-05-2022-14.9-million-excess-deaths-were-associated-with-the-covid-19-pandemic-in-2020-and-2021

[2] https://www.cdc.gov/infectioncontrol/spread/index.html

[3] https://covid.cdc.gov/covid-data-tracker/

[4] Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, et al. (2015). "A Household-Based Study of Contact Networks Relevant for the Spread of Infectious Diseases in the Highlands of Peru", PLoS ONE 10(3): e0118457. doi:10.1371/journal.pone.0118457.

[5] Ravi Maheswaran et al. A graph-theory method for patten identification in geographical epidemiology – a preliminary application of deprivation and mortality

[6] M. J. Keeling and P. Rohani, "Modeling Infectious Diseases in Humans and Animals", *Princeton University Press* (2007). ISBN: 9780691116174.

[7] W. O. Kermack and A. G. McKendrick, "Contributions to the Mathematical Theory of Epidemics-I", *Bulletin of Mathematical Biology*, **53** (1–2): 33–55 (1991).

Acknowledgement