Team Project: Mathematical Modeling

Zebedee Damrau Scarlett Knott Alazar Hailemeskel

Contents

Part I: Background Research

The world health organization has confirmed that the coronavirus outbreak first originated in Wuhan, China. The virus may have begun in bats and then spread to humans via a snake. According to a CDC report a Washington state resident was the first person in the United States with a confirmed case of coronavirus, having returned from Wuhan. Worldwide, Covid-19 fatality rate is 1.6 percent which makes it more deadly than the flu. Covid-19 illnesses have ranged from mild to severe to the point of requiring hospitalization, intensive care, and in some cases a ventilator. All ages groups can be infected with Covid-19 but the risk for complications increases with ages and underlying health conditions (such as diabetes, obesity, and lung disease).

The United states strictly followed the World Health Organization Covid-19 protocol of discouraging all public gatherings to achieve social distancing, hand washing and wearing face masks. Despite the implementing travel restriction from Covid-19 epicenter (China) and having high initial infections, the United States robust healthcare system made it to handle the pandemic successfully compared to other countries. The primary facilitators of the success in addressing the epidemic was an existing high hospital bed capacity as well as well-equipped hospitals. Most countries that experienced high Covid-19 related mortality lacked enough hospital space for those infected with the coronavirus. Unlike most other nations who waited for the people to get sick to begin containment measures, the United States began the process of early and effective testing.

Studying the spread of the virus will help us to conduct contact tracing and let people know if they need to self-isolate and get tested. By doing so, we can flatten the curve and result in fewer patients during this period, and hospitals would be better equipped to manage the influx of patients who are infected with COVID-19.

Part II: Brainstorming

Ideas

Infections are caused when an infected person and a non-infected person ‘interact’.

- Physical
  - Hand-to-Hand
  - Sharing Items
  - Sharing Surfaces
- Airborne
  - Coughing
  - Sneezing
  - Breathing

Infections can be prevented directly by masks and indirectly by quarantines.

- Quarantine
  - Schools
  - Businesses
  - Personal Gatherings
- Masks
  - Type
  - Availability
  - Compliance

Once the pandemic is underway, a strong government response can greatly reduce the severity.

- Response
  - Good Communication
  - Contact Tracing
  - PPE Distribution
  - Public Testing

Mind Map

Part III: Simplified SIR Model

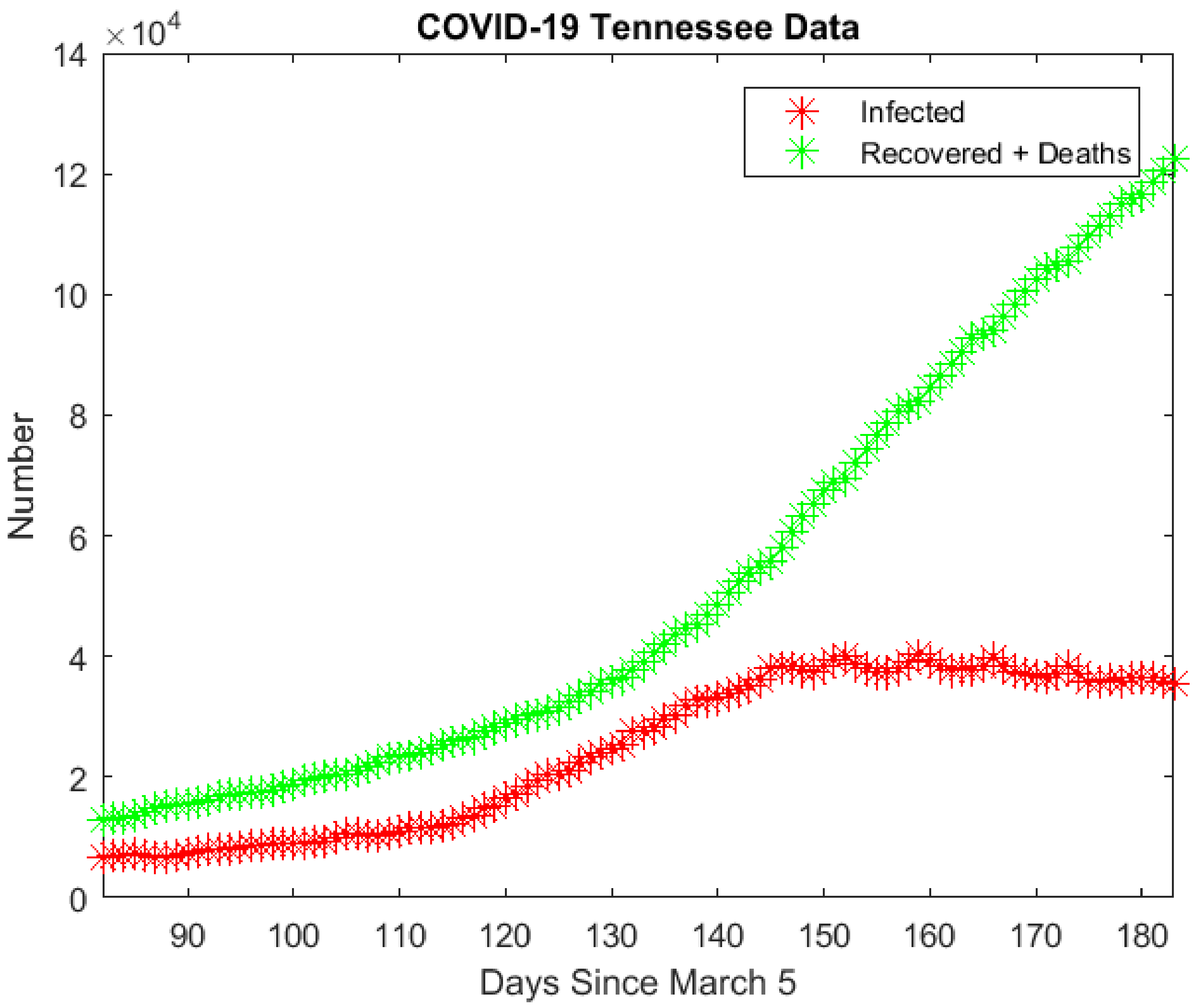
A SIR model is an infectious disease model in which there is a susceptible population, and infected population, and a recovered population. The susceptible population is those people that have not gotten the disease yet and are not immune for any reason. The infected population is the individuals who are currently infected and contagious. The recovered population includes those who are no longer contagious because they have recovered from the disease and those who are no longer contagious because they are deceased. In a simplified SIR model, people pass from susceptible to infected only through contact with those that are infected, and people only pass from infected to recovered if they recover or die from the disease, and there are no other sources of contagion, death, or birth.

The differential equations in an SIR model are derived by looking at the mind map of all the groups of people (susceptible, infected, and recovered + dead) and their interaction between each other. The change in the susceptible population is proportional to the interactions between the susceptible population and the infected population; this decreases the susceptible population. The change in the infected population is increased by interactions between the infected and susceptible populations and decreased once infected individuals recover. There are also proportionality constants in front of both of these terms to more accurately describe the changes. The change in the recovered and dead population increases and is proportional to the number of infected individuals, since a certain percent (roughly) recover or die after each unit of time.

When we consider a simplified SIR model, we are assuming several things. First, we are assuming that once someone has recovered from COVID-19, they are no longer susceptible and cannot return to the susceptible population. This is not true, both because immunity can diminish overtime and due to various strains of COVID-19. Second, we are not differentiating between different degrees of interaction people may have with the infected population. Some passing interactions are often insignificant in terms of spreading COVID-19, but longer interactions (especially those without masks) are much more likely to result in spreading COVID-19. Also, we are neglecting alternate sources, though they are small, of transmission of COVID-19, such as airborne particles and surfaces.

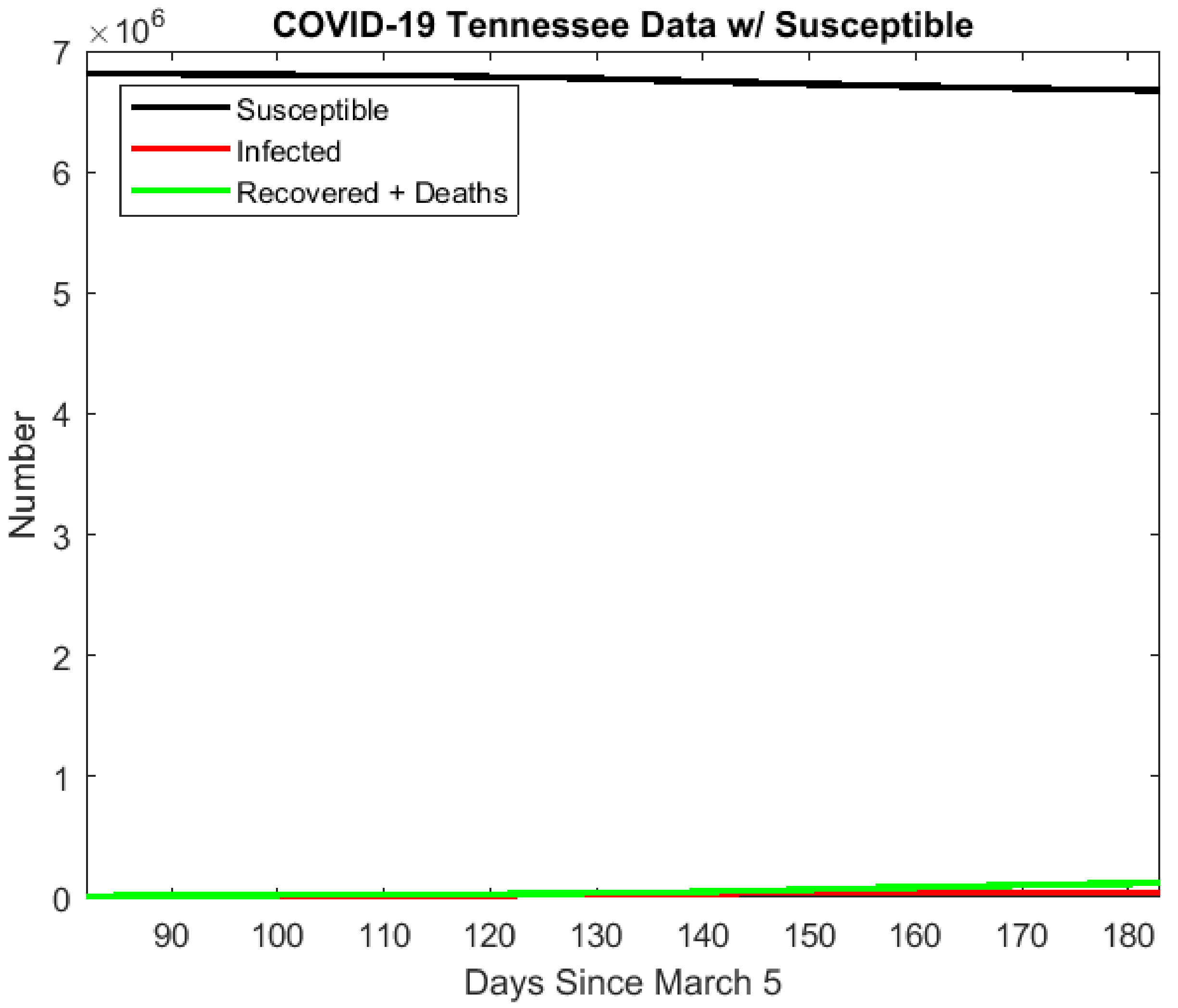
Part IV: Data Exploration

Infected and Former Infected



Here we can see the infected population in red and the former infected - both those who have recovered and those who have died - in green. Around day 145 the infected population levels off, but the former population continues to rises fairly uniformly. It seems like the infection rate fell to approximately that of the recovery rate. This is the end of July/beginning of August, so it is possible is that increasing quarantine restrictions, the proliferation of PPE, and greater access to tests caused a drop in infection rates.

Total Population



In addition to infected and formerly infected populations, this graph includes the total susceptible population of Tennessee. In a worst-case scenario where we are unable to prevent COVID-19 from spreading to everyone, we will see the green line rise to 7 million while the black line falls to 0.

One thing this graph makes clear however is how much further we have to go. By this frame of reference barely anyone has been infected, and any trends are nearly imperceptible.

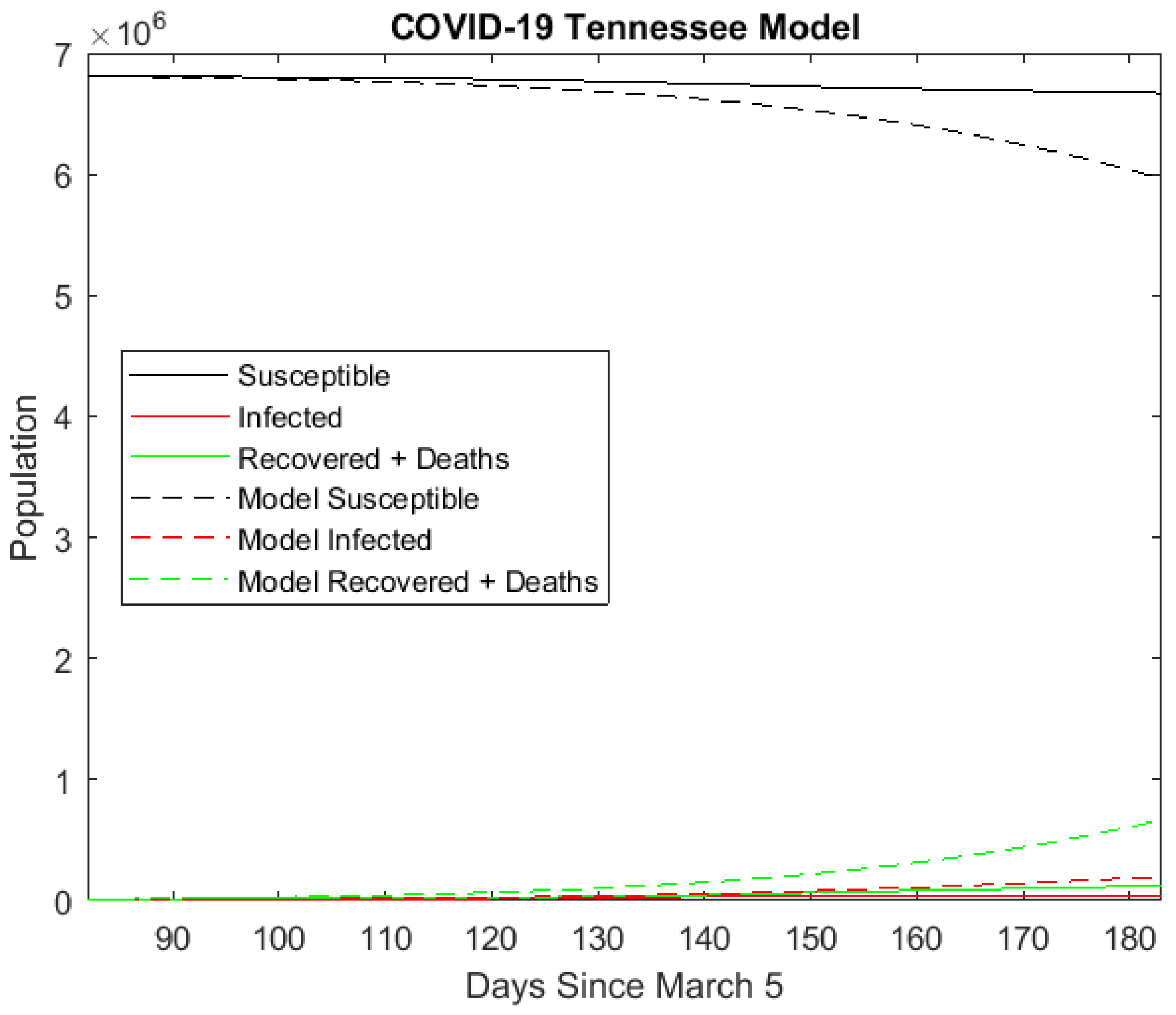
Part V: Developing the Model

Simple β Estimation

In this model, β is estimated by the formula

$$\beta_0 = \frac{I_1 - I_0 + \gamma_0 I_0}{I_0 * S_0}$$

where I is the infected population and S is the susceptible population.



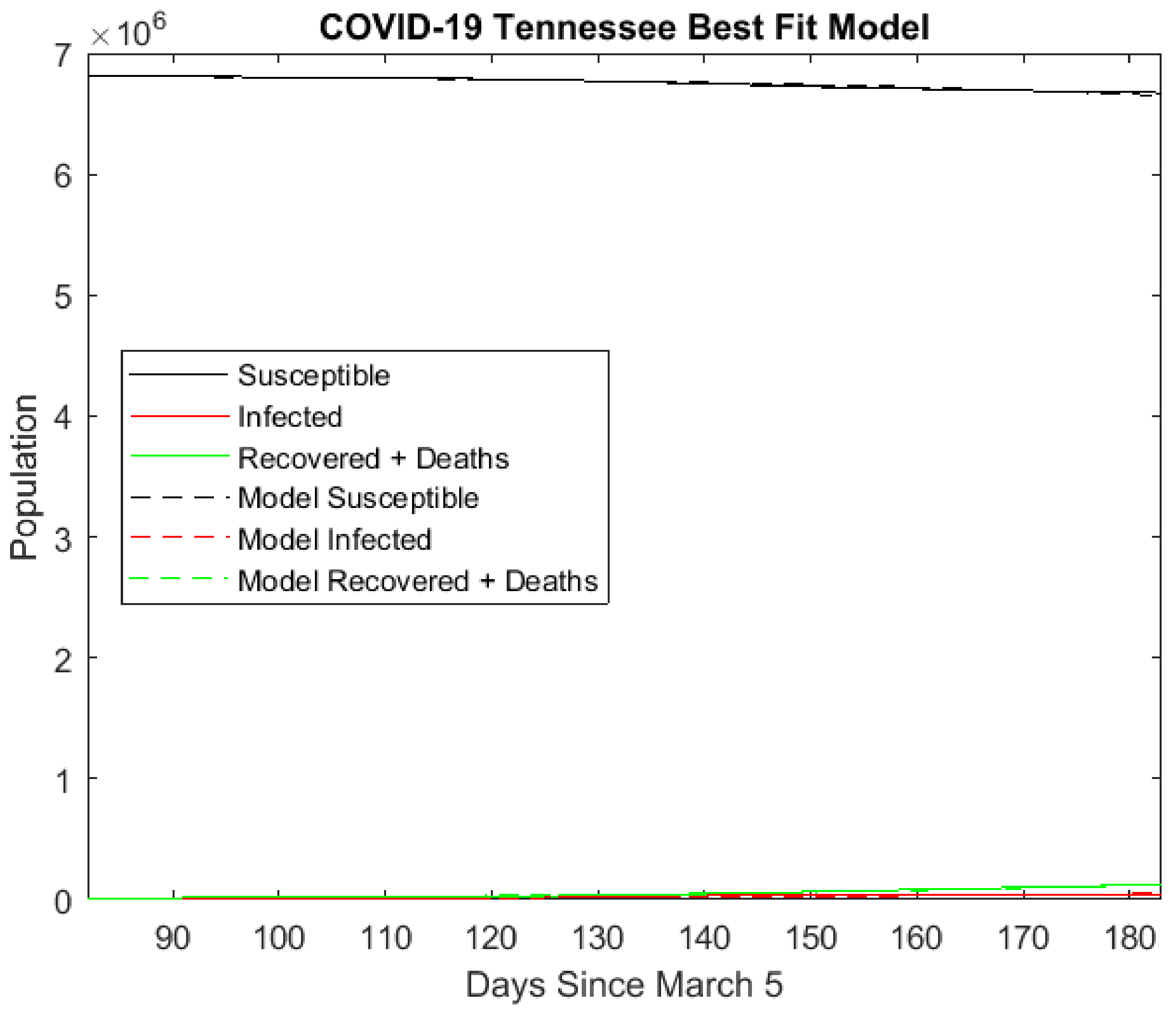
Here everything is happening much more quickly than in the actual data. The infection rate β too high, so by day 180 several times as many people have been infected. It seems that the first two days of the outbreak are not indicative of the entire period.

Ordinary Least Squares Analysis

To improve the fit of the model, we perform ordinary least squares analysis. This uses the formula

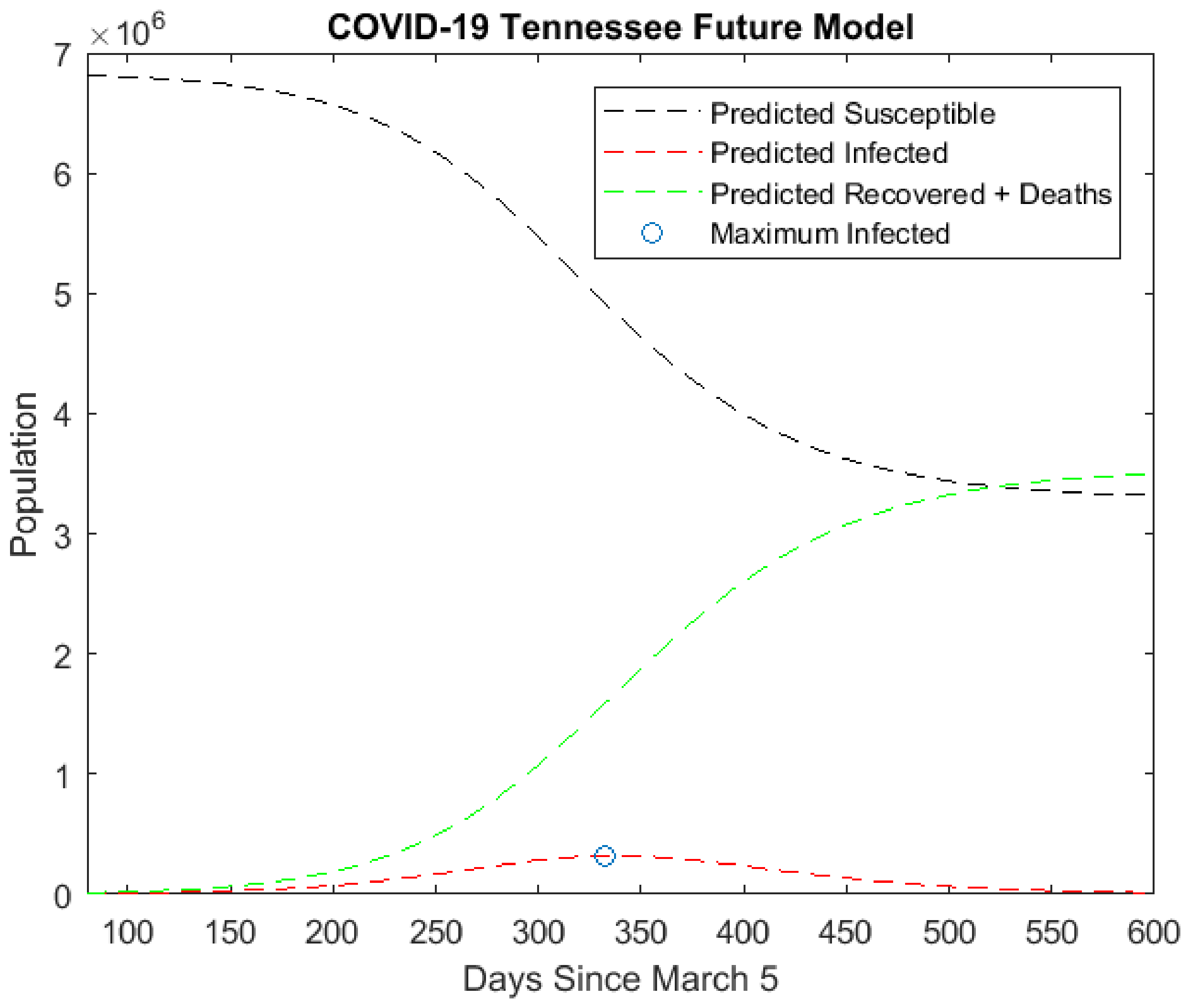
$$\hat{\mathrm{q}} = \min_{q \in Q} \frac{1}{N} \sum_{i=1}^{N} \left(\hat{y}_i - f\left(t_i, q\right)\right)^2$$

where q = $[\beta\ \gamma]$.This formula tries to minimize the difference between the predicted $f(t_i, q)$ and the actual $\hat{y}_i$. This results in parameters that should fit the data much better than the previous model.



As expected, this model has a much tighter fit. In the graph, the model is nearly indistinguishable from the data. Should no other changes to the infection rate or recovery rate occur, this model will do well in predicting the future.

Modeling the Future



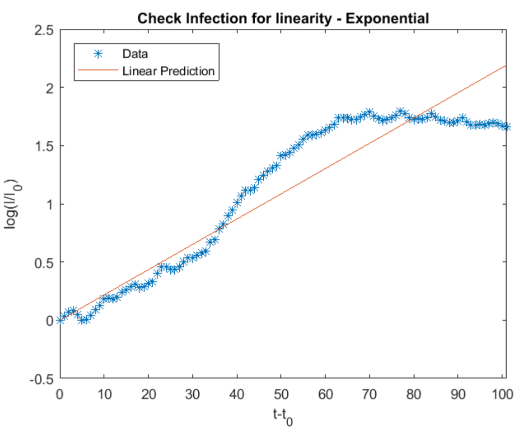
Our best fit model predicts that the infected population will peak around day 330 with 320,000 individuals sick at the same time - only about 5% of the total population. However, more than half of the population will catch COVID-19 before herd immunity stamps it out, and that takes about 600 days.

There are two reasons why this model may not represent reality, however. First, it does not model changes in behavior. Over time, people become less strict with adherence to safety protocols, increasing infection rates. Also, seasonal changes like reopening schools or holidays like Thanksgiving and Christmas increase the infection rate. Second, it does not model the distribution of vaccines. The vaccines coming out should (hopefully) transition the susceptible population directly into the recovered population, drastically reducing infections after 360 days.

This model would probably be better if it was broken up into more categories. For example, many sick people are isolated, so they would have a much lower infection rate. It may not be relevant yet, but a category for vaccinated people that can rarely get sick is better than adding them directly to recovered. Also, temporary boosts to the infection rate around holidays will simulate super-spreader events.

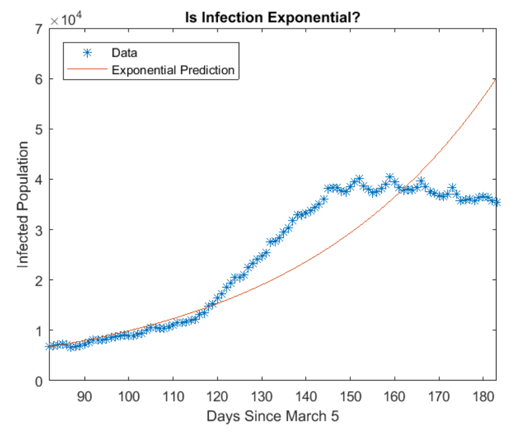
Part VI: Infected Population

Exponential: Check for Linearity



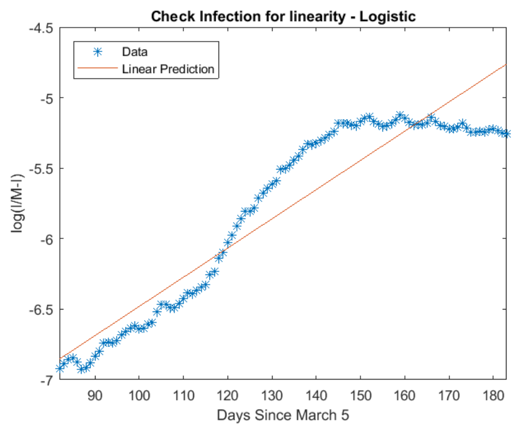
This graph compares $\log\left(\frac{I}{I_0}\right)$ vs $t - t_0$ where I is then infected population and includes a line of best fit. If the data was linear, we would know that the infected population is growing exponentially. In addition, the rate of exponential growth would be the slope of the best-fit line, $k = 0.0217$. However, while it is close to linear for the first 40 days, the infected population then rapidly rises before leveling off. It does not approximate exponential growth beyond those 40 days.

Exponential Model



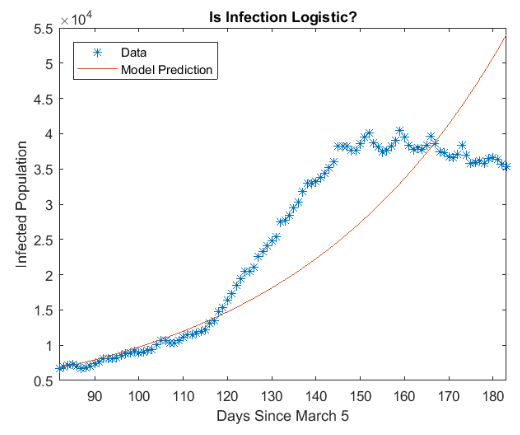
Here we see the data plotted alongside the best estimate for exponential growth. The model is $I = I_0 e^{k(t-t_0)}$, where k is the slope from the previous graph. As expected, the first 40 days fit the model fairly well. After that, however, the data and model greatly diverge. The model continues its smooth ascent while that data jumps up then holds steady. It is clear that the infected population is not growing exponentially.

Logistic: Check for Linearity



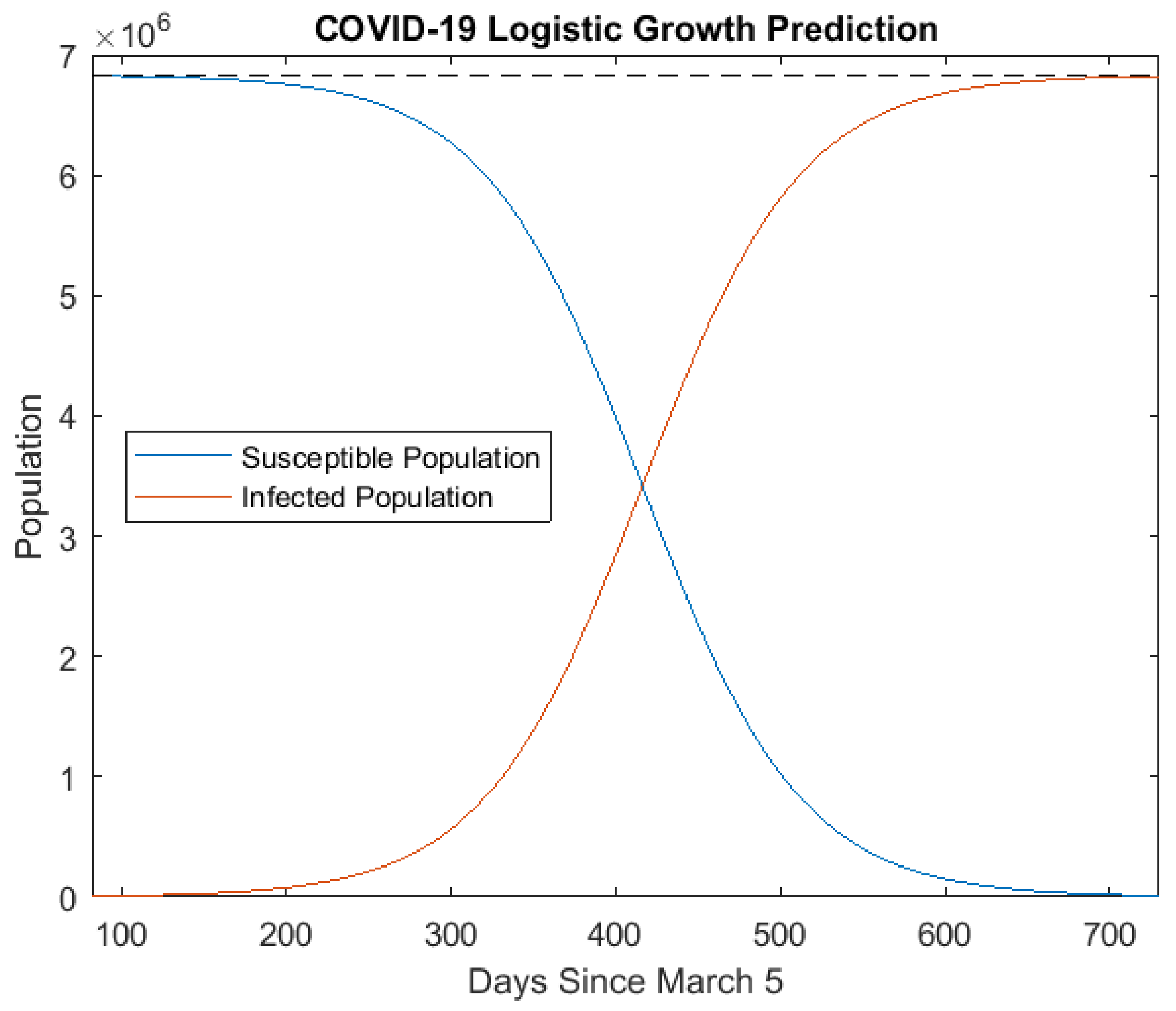
This graph compares $\log\left(\frac{I}{M-I}\right)$ vs t where I is the infected population, M is the total population (or maximum infected), and it includes a line of best fit. If the data was linear, we would know that the infected population was growing logistically. In addition, the rate of logistic growth would be the slope of the best-fit line, $r = 3.0332 * 10^{-9}$. However, just like with the exponential model, the data only appears linear for the first 40 days or so. After that, it follows the same pattern of rising then holding that we see for exponential growth.

Logistic Model

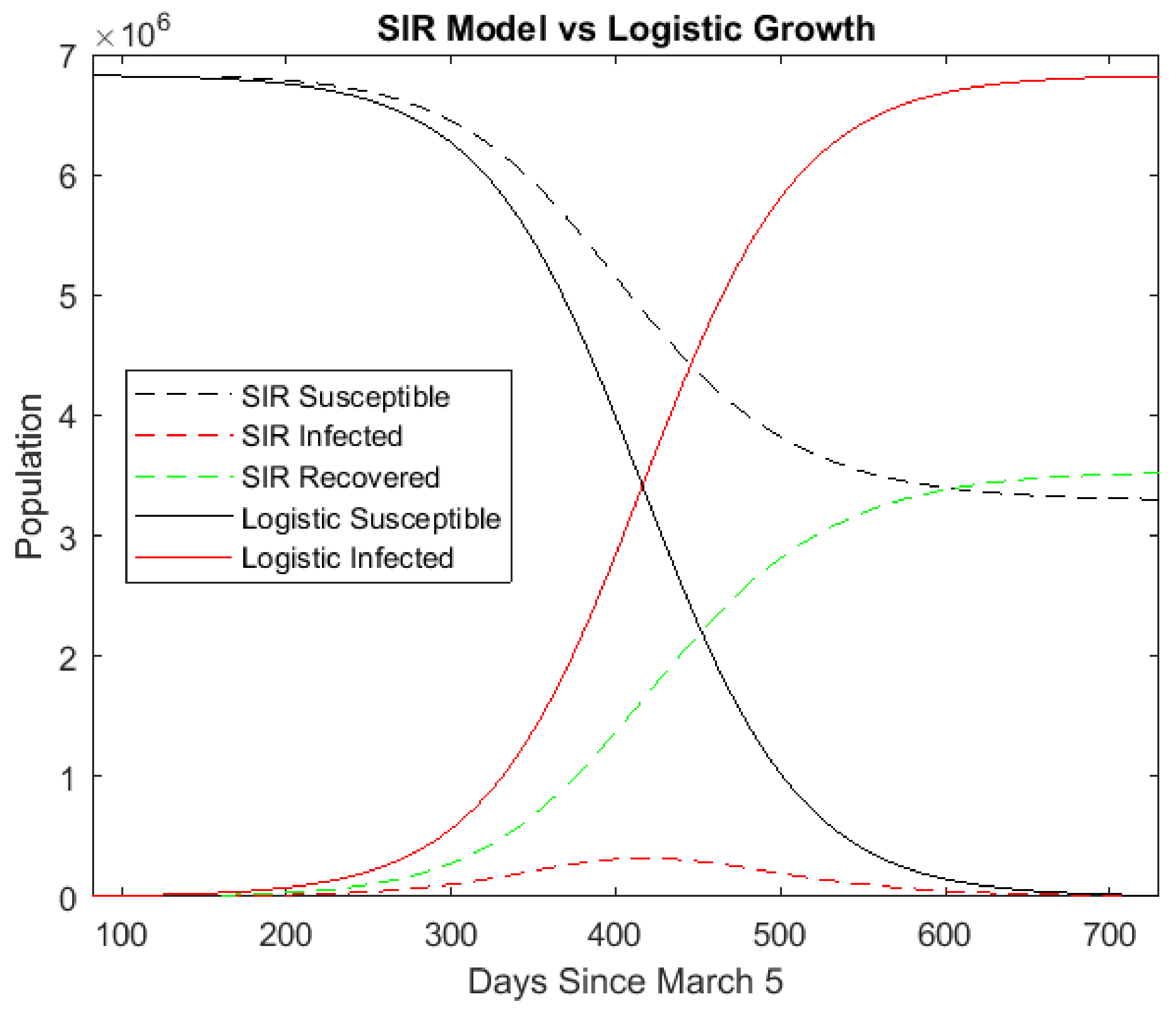


The model here is $I = \frac{I_0 M e^{rM(t-t_0)}}{M - I_0 + I_0 e^{rM(t-t_0)}}$ where r is the slope from the previous graph. This model confirms what the linearity check suggested: The infected population is not growing logistically. Just like in the exponential case, the data and the model diverge after 40 days, and it becomes clear that this isn't logistic growth.

Logistic Growth Prediction



This graph extends the logistic model out to two years after March 5. By this point, nearly the entire population has been infected. However, this doesn't make a lot of sense. There is no category for recovered people; Everyone who gets sick stays sick forever.

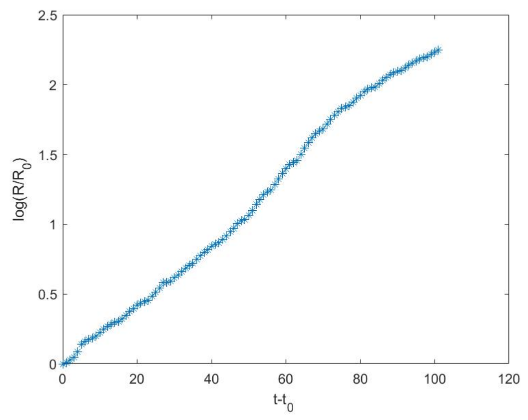


This comparison with the SIR model shows how extreme the differences are. In the SIR model, at most 5% of the population is infected at once, and nearly half of the population never gets sick. The infected population grows much slower and peaks much sooner, and is overall a much less bleak prediction. The logistic model by comparison does not seem like an accurate prediction.

There is one major reason to believe that neither exponential nor logistic growth apply in this situation: People get better. Both types of growth assume that the population cannot shrink, but with a disease like COVID-19 we know that people will recover. Even if everyone susceptible becomes infected, the infected population will inevitably shrink, so we can say with certainty that the infection rate is neither exponential nor logistic. The phrase "The number of infected is rising exponentially" can only be true in the short-term, when a tiny percentage of the population is sick and few have recovered.

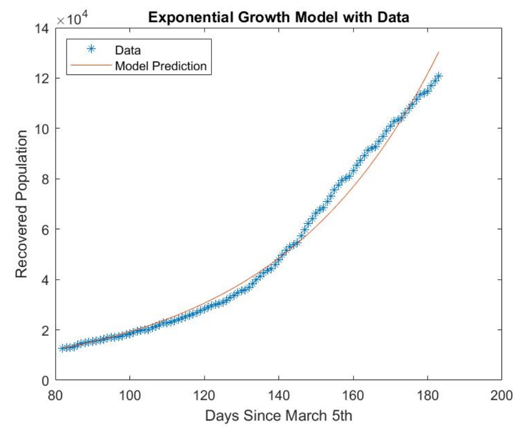
Part VII: Recovered Population

Exponential: Check for Linearity



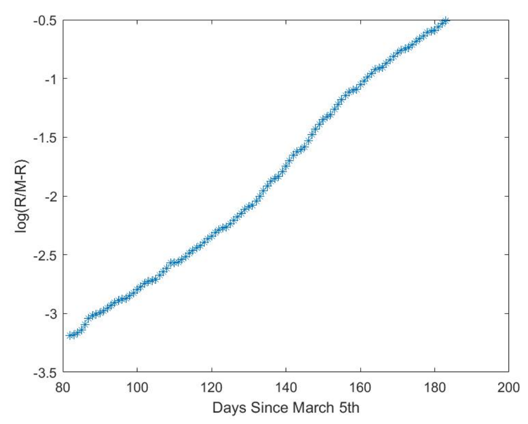
On the left, we have a graph of $\log\left(\frac{R}{R_0}\right)$ vs. $t - t_0$, where the variable R describes the recovered plus deaths population. This graph is somewhat linear. However, it seems mostly linear up to about 45 days after March 5th where the line drifts higher and then becomes roughly linear again starting at about 70 days after March 5th. This indicates that though the graph follows an overall linear trend, it might not be well described by a single linear approximation. Thus, an exponential model might not work best for this data.

Exponential Model



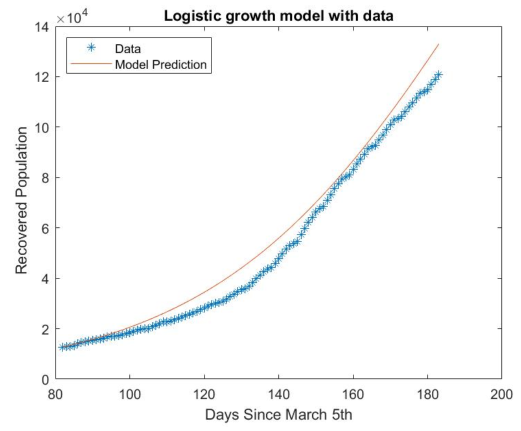
Utilizing a linear regression, we find that the exponential growth constant, k is approximately $k = 0.0230$. This model, overall, follows that data. However, there are definite portions of the graph where the model does not fit well. Starting around 110 days after March 5th, the model is higher than the actual data. Then, starting around 140 days after march 5th, the model is lower than the data, before returning to being higher than the data around 170 days after March 5th.

Logistic: Check for Linearity

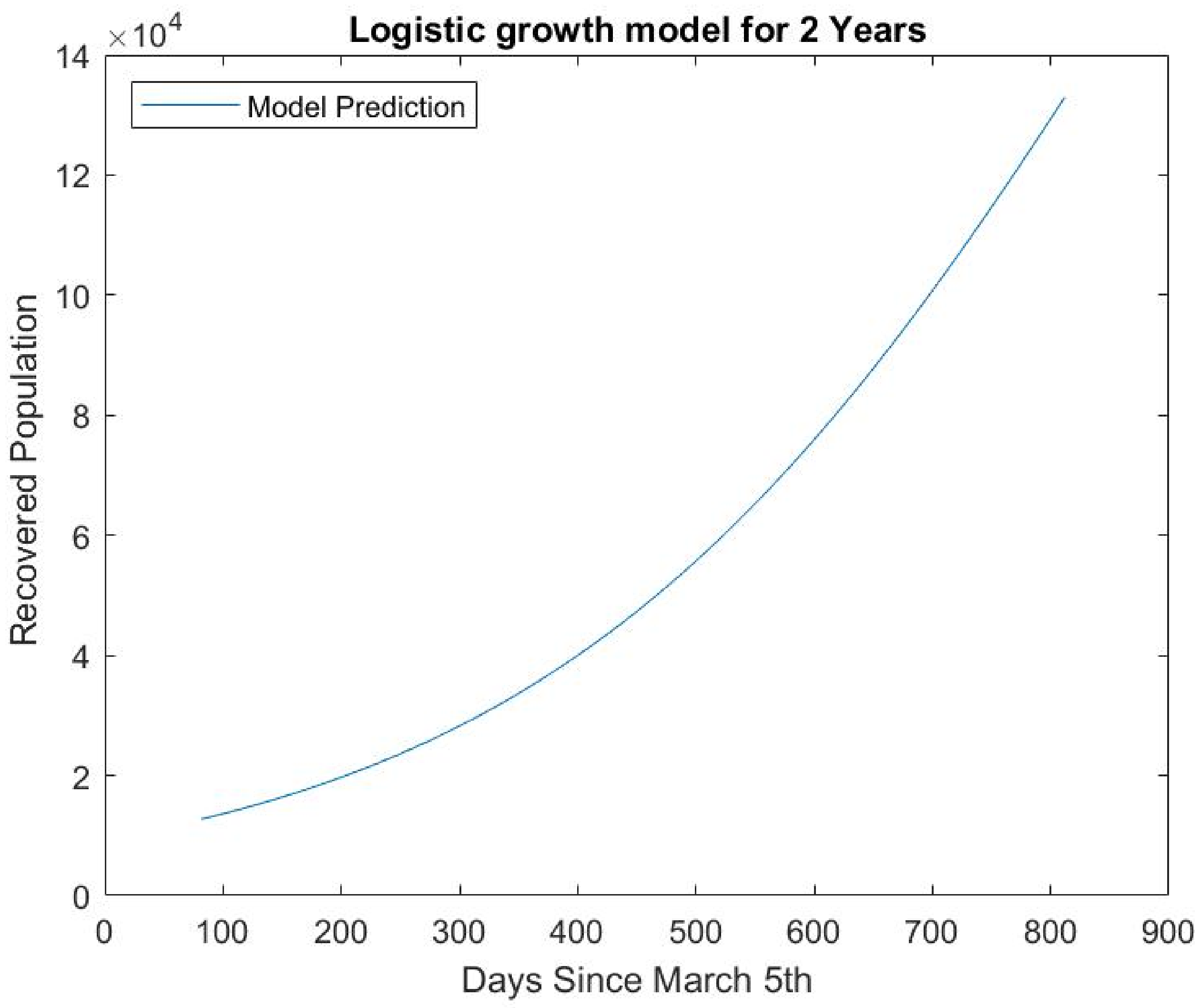


On the left is the graph of $\log\left(\frac{R}{M-R}\right)$ vs t. This graph looks approximately linear. It tends to trend linear until about 130 days after March 5th. Then the line shifts upward until about 160 days after March 5th. The graph is then linear from 160 days after March 5th onward. The estimated rate of growth r is $r = 8.7313 * 10^{-8}$.

Logistic Model



Here we have our logistic growth model with our data. This model does an okay job modeling the data. Until about 100 days since March 5th, the model is fairly accurate. Between 100 and 160 days after March 5th the model overestimates the recovered population; at points, this overestimate is just under 10,000 people higher than the data. The model comes closer to the data around 160 days after March 5th. The model then continues to overestimate the recovered population as compared to the remainder of the data. Overall, this model has a tendency to overestimate the recovered population, despite being quite close to the data itself in places.



The logistic model predicts that the recovered population will simply continue increasing for two years. I would estimate that the recovered population would level out after about 2 more years than shown. This model does not account for decreased interactions (because there would be less susceptible and infected people overall) like the SIR model. So, since about half the population is already recovered after about 2.5 years, another 2 years should mean that everyone has been infected and has recovered from COVID-19. The recovered population levels out much faster in the SIR model, after about 2 years. I think the SIR model makes the most accurate prediction, since it accounts for things like decreased transmission over time.

Part VIII: Conclusions

Major Findings

One major finding we have is that the infected population is not modeled well by a purely exponential or logistic model. This is likely because a variety of factors, including public policies and individual behavior, suppressed the infection rate initially. Additionally, factors like these mean that the rate of infection is not accurately expressed as a constant. The recovery rate seems to be well described and modelled by a logistic model. The model fits well, and this model allows for things like a limiting size (the number of recovered cannot exceed the state's population) to be included. Also, the SIR model developed in part V seems to be the best and most sensible model used throughout this report.

Future Work

In order to improve the model, we would add several factors. First, starting around January 2021 (or whenever proves to be best based off of real-world events) we would add a vaccinated population. A large percentage of this population would become not susceptible upon entering this population, and this population would increase as more people receive the vaccine. We would also make sure that the infection rate is not static. On top of public policies and individual choices, the seasonality of the infection rate is something to consider (both because the virus may actually be seasonal and because of more time indoors). Also, the recovery rate should go up over time. This is due to improved treatments for the virus. Additionally, the variable γ can change as well, since improved treatments can affect how quickly someone recovers from COVID-19. Lastly, this model could also factor in how mask mandates can affect the rate of infection and the recovery time of the virus. This could be accomplished at the county level or by looking at what percentage of Tennesseans are under a mask mandate at any given point in time.