Alexander Ochoa
Baixi Guo
James Wei
Nathan Downey
Victoria Mendoza

Using a SIR Model to Predict Coronavirus Cases

Introduction:

For this project, our group chose to use the SIR Model to predict coronavirus cases because we could apply what we learned to real world data.Through the dynamics of the SIR model, numbers of infection rate and recovery rate will be conveyed. The analysis of the data will help us approximate the time of those rate changes and possible inferences of the causes. Utilizing the SIR model will allow us to determine a likelihood behavior of the disease spread. With the recent pandemic, the SIR model is important to help analyze the expansion rate of COVID 19. Utilizing the concept of Fourth Order Runge Kutta Method from Math 131, we could apply this to the SIR model as it would help us determine a likelihood pattern for the rate of change between infections and recovery rate. Being able to adjust the parameters we can see how different scenarios of the pandemic could have happened.

Background:

For this project, a SIR model was used in order to describe Coronavirus data in which S is the number of people susceptible to get infected, I is the number of people infected, and R is the number of people that are removed all within the given time(Tolles). Through this, three assumptions were made. First, the population is a fixed number with no additives meaning we are not counting births and deaths. We are also assuming that once a person has recovered they cannot get infected again. Finally, there is a fixed transition rate between the susceptible to infected rate and infected to recovered rate. These assumptions were made due to the idea of how much it would complicate the code. For example we would have to consider how fast a person can get infected again which varies person to person. As for the total population, we understand that in the real world the population would vary, but given the idea that the total population is constant as the number of susceptible, infected, and recovered people in total is proportional to that constant, and the derivatives of each of these variables with respect to time is zero, we assumed that our population is fixed.

With this model comes several limitations, one being the model does not show the given time rate of when the person is infected versus when they're actually infectious. In our case for this project, it correlates with time given where people with coronavirus are asymptomatic and are typically not infectious and are not shown within the model. Other issues include how the model assumes that everyone has the same capability of getting infected, meaning the model does not take into account those who are social distancing and consider those who are social distancing just as likely to get infected as those who are not. Lastly, the model shows a fixed population meaning those who are not migrating, born, or pass away due to the disease are not factors in the model (Tolles). This also means that deaths by the disease are recorded within the recovery portion of the model therefore the recovered portion consists of those who actually did recover and those who passed away due to infection.

Below are three equations which were used for this project (Tsai). The first equation describes the derivative of susceptibles over time. In this equation, beta is the infection rate which correlates to the number of people infected per day and indicates a decrease in susceptibles over time while recovered individuals would increase over time. Equation two describes the derivative of infected individuals over time. In this case, this equation describes the difference between the rate of infected individuals and the rate of recovery also known as gamma. The third and final equation explains the rate of change of the recovery with respect to time. These three equations provide a better understanding of how the graphs obtained are portrayed.

Equations:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{1}$$

$$\frac{dI}{dt} = -\frac{\beta SI}{N} - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

Results:

From the coding section, there are four Matlab files. SIR_RK4_system.m constructed the system of SIR functions and developed an algorithm to solve for the system of differential equations by using fourth order Runge-kutta method. SIR_Visualization.m used the pre-defined inputs to implement the Runge-kutta method and visualize the numbers of susceptible, infectious and recovered populations with respect to time. SIR_RK4_built_in.m uses the same predefined inputs to implement Matlab built-in RK4 solver for system of equations. Since one of the inputs required a function input. SIR.m can be used as the input for SIR_RK4_built_in.m to construct the system of SIR functions (same SIR function constructed in SIR_RK4_system but this is a reference code from external source)

The SIR model prediction in Matlab can be used to find the pattern and rate that COVID-19 has on infection spread and people's recovery. However, the accuracy of the results are significant. The result generated from SIR_Visualization.m and SIR_RK4_built_in.m can be used to compare because "Note that by default MATLAB's ode45 solver is Dormand-Prince method, which is also a member of the Runge-Kutta family of ODE solvers. As ode 45 is an adaptive step size integration algorithm, it can give a more accurate numerical solution" (Tsai). An example would be calculating the maximum value of the number of infectious population when infection rate is 1.5 infectious people per day and recovery rate is 0.5 recovered people per day. By using the max() function on the I column vector we can get the results that maximum values from SIR_RK4_system.m is 33.5582 while maximum value from SIR_RK4_built_in.m is 33.5742. The relative error is $\frac{|33.5582 - 33.5742|}{33.5742} \approx 0.0004765$ = 0.04765%. We can see that the difference between the maximum value of an infectious population with the same conditions are nearly negligible.

There are still sources to cause inaccuracy of the values of the SIR model. One of the sources is the step size of the time interval because the more step size between the intervals the more accurate the results. Another source of inaccuracies is the formula of fourth order Runge-Kutta Method because I had seen the formula with slight difference comparing with this one in other sources (Dr.Math)

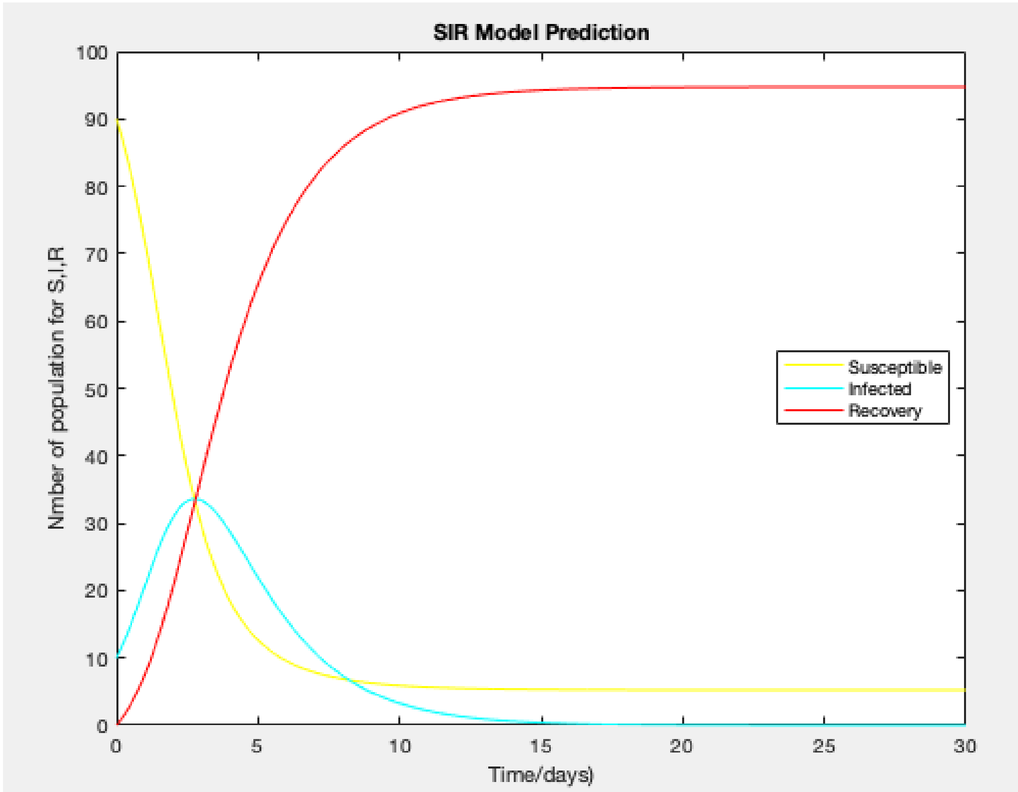

Figure 1. When $\beta = 1.5, \gamma = 0.5$

For the purpose of the project, we decided to insert 3 graphed predictions of the SIR model where the rate of infection, beta, is greater than, equal to, and less than the rate of recovery, gamma. These predictions were made as a test with 100 people over the span of 30 days. Above is **Figure 1**, a prediction where the value of beta is larger than gamma. Here we see that as the days progress, the number of recovered people increases and the number of susceptibles decreases at exponential rates within the span of days 5 through 10. The number of infected people on the other hand, shows a bell shaped curve that peaks at approximately day 3.

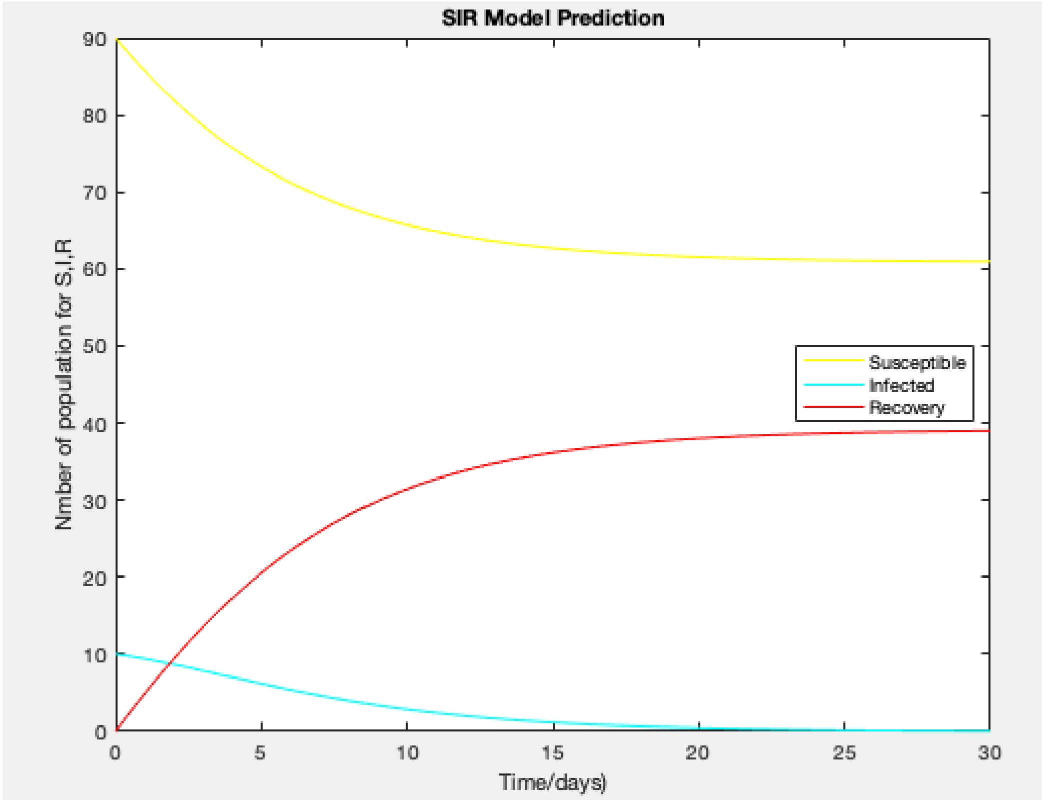

Figure 2. When $\beta = 0.5$, $\gamma = 0.5$

Figure 2 displays the prediction model of when the rate of recovery is equal to the rate of infection. In this figure we see how the number of infected slowly decreases until it reaches a value of zero where no one is infected at around day 20. For this case, those infected decrease and those recovered increase at exponential rates at the same time.

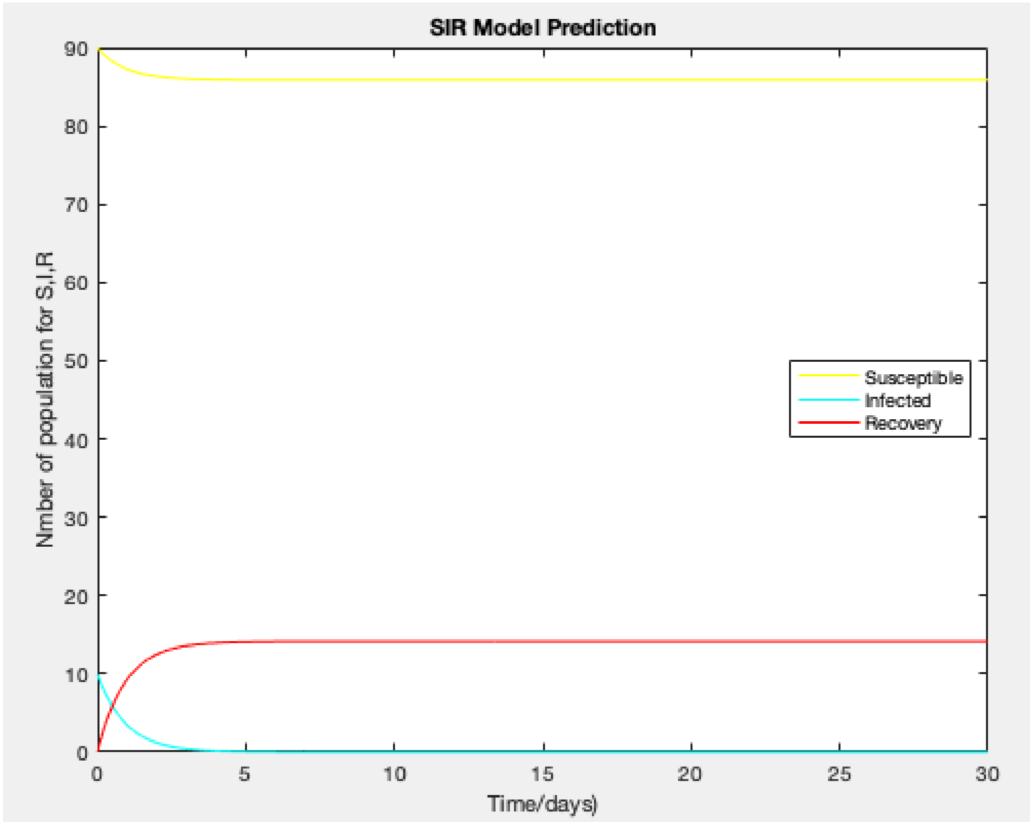

Figure 3. When $\beta = 0.5,\ \gamma = 1.5$

The last prediction is displayed above as **Figure 3** in which the rate of infection is smaller than the rate of recovery. For this graph the number of those recovered is greater than those susceptible to infection, and the number of those infected decreases exponentially. These three variables hit their minimum and maximum values within the first 5 days of the experiment and remain the same until day 30, the last day for our prediction results.

Conclusion:

In light of our experience, the group had decided to conduct their research on the infected rate and the recovery rate of covid. It was decided to conduct this research by using the SIR Model; the group was able to examine both rates compared to a group of subjects being tested. This was able to be done by using matlab to run a code simultaneously gathering data on all 3 variables and having the experiment run over a period of time. All while also displaying accurate graphs and tables of said data point. This was able to be done by using the RK4 and Euler's to help gather our data and as well as display data using the RK4. The data itself was able to display an accurate portrayal of the effect Coronavirus has had on people. After conducting our experiment we could have also added in more factors as more data is presented to us. Such as the rate of infecting asymptomatic people and well as people who were unable to recover or had passed due to other circumstances besides Coronavirus. You could also make the argument that we could have gone more in depth exploring the vaccination rates and the outcomes that may have on our rates of infections and recovery as vaccinations are becoming more available.

Works Cited

Juliana Tolles, MD. “Modeling Epidemics With Compartmental Models.” *JAMA*, JAMA Network, 23 June 2020, jamanetwork.com/journals/jama/fullarticle/2766672.

“Compartmental Models in Epidemiology.” *Wikipedia*, Wikimedia Foundation, 6 May 2021,en.wikipedia.org/wiki/Compartmental_models_in_epidemiology#Transition_rates.

Tsai, Tony. “Tony Tsai.” *RK4 Method for Solving SIR Model*, blog.tonytsai.name/blog/2014-11-24-rk4-method-for-solving-sir-model/.

Dr. Maths. *Lecture 20 : Numerical Solution of SIR model*, Jul 22, 2020. https://www.youtube.com/watch?v=qTUhLrwpiDs