

Epidemiology Project Report on SARS-CoV-2 West Virginia, USA

The SIR Model:

The **SIR model** is one of the simplest compartmental models, and many models are derivatives of this basic form.

The model consists of three compartments:

S: The number of susceptible individuals.

I: The number of infectious individuals.

R: The number of removed (and immune) or deceased individuals.

The Model Variables:

S(t) - Proportion of Susceptible individuals on day t.

I(t) - Proportion of infectious individuals on day t.

R(t) - Proportion of recovered individuals on day t.

The Constants:

N - Total population of the area **West Virginia** (1792147)

a - Proportion of infected individuals recovering per day ≈ 0.1365

r - Number of people each person infects per day ≈ 0.2575

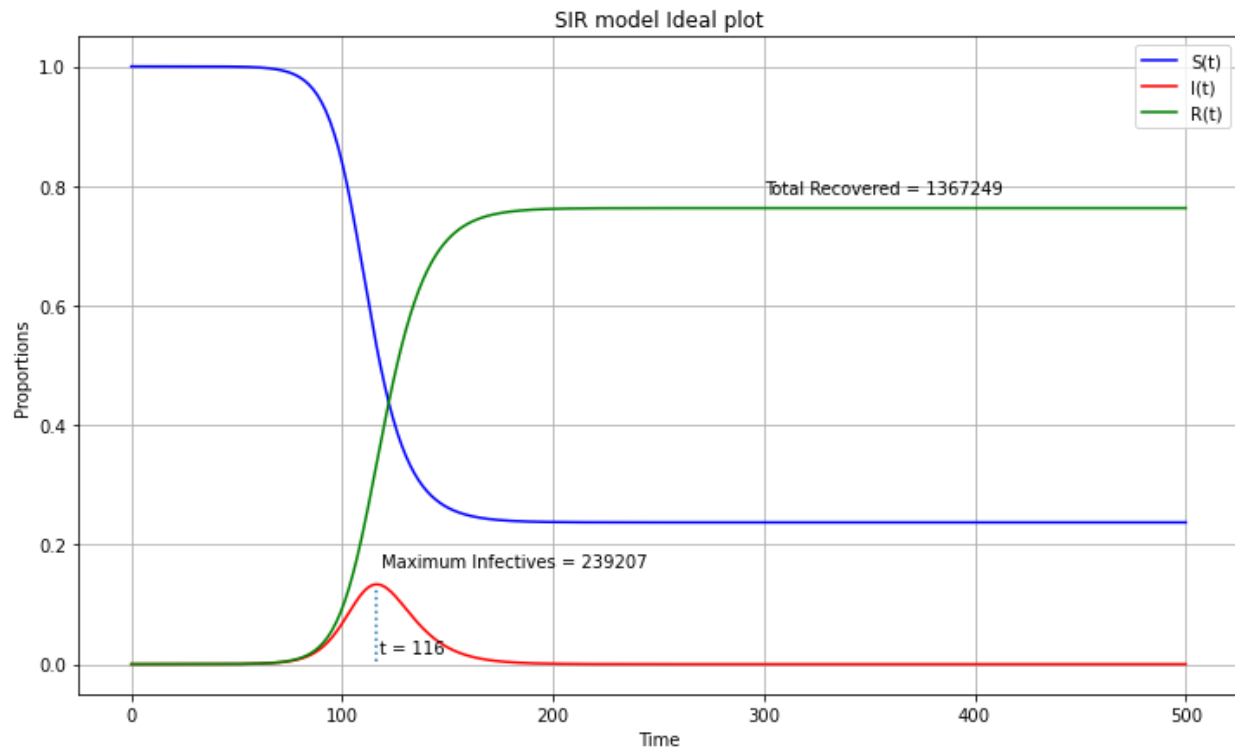
(a and r are calculated in the program)

Equations Used:

- $\frac{dS}{dt} = -r * S * I$
- $\frac{dI}{dt} = r * S * I - a * I$
- $\frac{dR}{dt} = a * I$

Assumption Used For Data Collection:

The values of the no. of deaths and the number of confirmed cases for the territory of West Virginia was available online. $d_{avg} = (\text{no. of deaths}) / (\text{no. of deaths} + \text{no. of cured}) = 90/1816$. (The no. of cured doesn't always come out to be an integer and therefore needs to be round off) Therefore we can find the no. of cured patients on each day.



$R_0 = (r/a) * \text{initial proportion of Susceptible} \approx 1.8866$

Maximum no. of infectives = 239207

The time location of the maximum number of Infectives = 116 days

The total number of recovered = 1367249

The duration of the epidemic = 116 days (The epidemic stops when the number of Infections start decreasing, in the case of this model, it's the maxima)

The first case of SARS-CoV-2 virus in West Virginia was seen on 18th March 2020 (Day 0).

Approach used for the creation of the model:

When we integrate the equation dI/dS , we get an expression of ρ on a particular day on the basis of the data we've collected (see in the program file).

The ρ value is then taken to be the average of that of ρ on each day.

Also, the expression for $R(t)$ helps us estimate the value of a on each day and the resultant a is the average of all these.

Now, we know the relation between a and r , and thus, we can estimate the value of r .

The *scipy* module has been used to integrate the basic equations of the SIR model (mentioned on page 1) and then plotted the graphs using *matplotlib*.

Results And Predictions:

The pandemic ideally should end for the state after 116 days i.e. 12th July.

While we keep the value of a for the graph in between 0 and $a_{actual} + 0.1$, the shape of the graph flattens as the value of a increases. As we increase the value from $a_{actual} - 1$, the graph shifts to the right slowly till it reaches a_{actual} . After that with every small increase in the value of a , the graph shifts to the right rapidly.

Whereas, as we increase the value of r in between 0 and $r_{actual} + 0.1$, the graph rises and also shifts to the left.

This shows us the dependencies of the model on the transmission rate as well as rate of recovery. We know how the graph shifts based on these parameters.

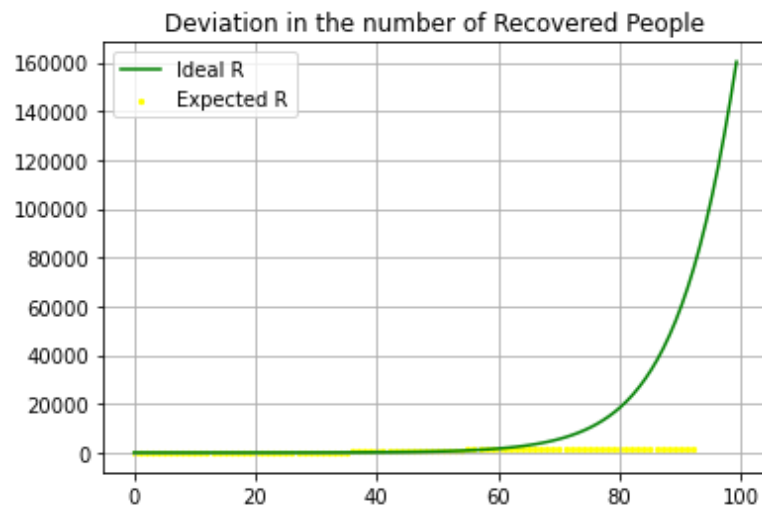
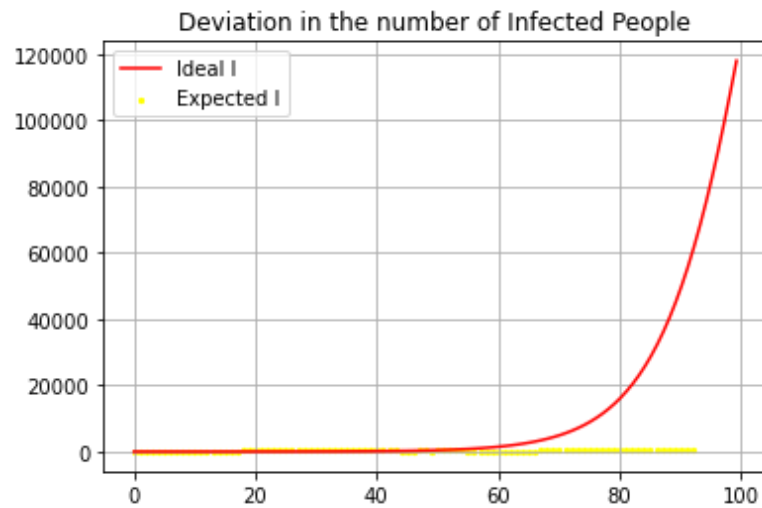
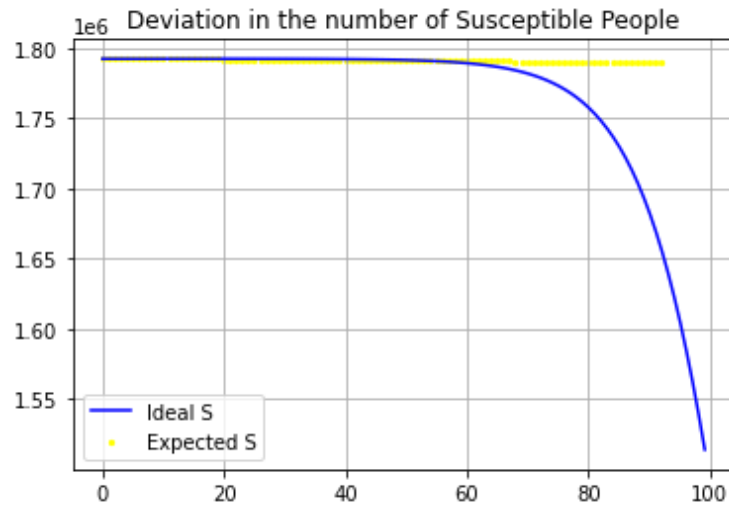
If we keep on changing the values of a and r beyond all this, the graph almost flattens down, the proportion of infected individuals is very low.

(I've manually checked the model graph by putting in various different values of a and r and changing them).

Why Not Choose The Machine Learning Approach?

Using the concepts of Linear Regression, Curve-fitting, error calculation and other advanced algorithms can be a fair way to do this project, but the true data actually does not follow the ideal expressions derived from the integration of the basic SIR model equations. So, the curve formed by using the approach would be far from accurate too. Also, this requires a deeper level of understanding as to why it works this way. We're not yet that familiar with the underlying logistics. So, we chose something completely based on our understanding and that we can completely justify.

Comparison:



Assumptions Taken During SIR Modelling:

- Rate of transmission remains constant.
- Every person is moving and has an equal chance of contact with every other person.
- Recovered individuals are immune to the disease.
- Population remains constant throughout the period of modelling (any other natural ways of death and birth are neglected).
- Each individual has an equal probability of getting infected and recovered.

Limitations of the SIR Model:

- No consideration of the decreased rate of infection by and to the quarantined individuals.
- Oversimplified
- All the assumptions can be considered as limitations of the model as they're not always 100% accurate.
- The actual scenario might not always come to equilibrium when the time approaches infinity. We're ignoring those disturbances. (eg. at the point where most of the infection dies out, there might be a second wave of the epidemic where the infection again rises rapidly).

References:

- ❑ <https://github.com/CSSEGISandData/COVID-19>
- ❑ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4996659/>
- ❑ <https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model>
- ❑ <https://dhhr.wv.gov/Pages/default.aspx>
- ❑ <https://covidactnow.org/state/WV?s=54069>
- ❑ Stack overflow for bugs
- ❑ Lectures

And various others