Abstract

For this case study, we sought to model the dynamics of Covid-19 within a population. We created a base infection model to plot the spread of a simple disease, but the Covid-19 data was anything but simple. We found that the many waves of Covid-19 could not simply be graphed with a lone disease model. By plotting our data as a single model, we received a linear result which did not accurately portray Covid-19's spread over time. To plot the data realistically, we had to make several models for each of the major periods of Covid's spread. Only through this method did we receive data that better fit the original data. Once we completed this, we then applied a policy to reduce infections and deaths to a parallel set of data so that we could compare the effects of the policy to the original result. After this, we finally went to work on a set of mock data to predict how much of a population was vaccinated and how many breakthrough cases were happening based on infection and death rates. We managed to create an algorithm which was able to predict and plot these necessary pieces of data, which shows us how data can help data scientists make predictions of certain conditions.

Conclusion

We found out that studying the spread of a disease isn't as simple as applying a single model or equation. We created an SIRD model which can accurately model how a simple disease spreads through a population over a short period of time. When we applied this model to actual data, we found our results to be too rigid and linear compared to the original data. It didn't accurately reflect the waves of Covid-19 spread, and how the rate of infection changes over time. To fix this we identified seven periods of varying infection rates and made individual models of these periods. By doing this, we managed to create a far more accurate result which nearly perfectly fit the original data. This showed us how large complex sets of data may be better processed separately and combined afterwards. From there, we were able to apply policies that could impact these results and give us lower infection and death rates. With this out of the way, we went on to process our mock data representing infections and deaths over time. We know for this data that a vaccine for Covid-19 was rolled out at day one hundred. Our objective was to predict how much of the population was vaccinated, and how many individuals were experiencing breakthrough infections over time. We created an algorithm to do just that, and it helped show us that data could be used to make predictions on related data.

We found our shortcomings in creating the multi-model set to accurately portray Covid-19 data. With the death counts being so small in the actual data, any little inconsistency could create an abnormal result. It was a significant challenge to be able to perfect these results. Additionally, we had challenges with predicting the number of vaccinated individuals and breakthrough cases. We went through many calculations, many algorithms to try to find the most accurate results, and even then we are left with the uncertainty of knowing that we will never truly know the exact answers.

If we were to expand upon this case study, we would investigate other historical breakouts of diseases and see how their spread is different from Covid-19. Perhaps it could be an opportunity to discern which disease is which based on how quickly it spreads across a population.