Jacob Yoke Hong Si
Student #: 1005815806

Title: STA302 Final Project Part 1

Research question: How do the stringency index and vaccination rate affect COVID-19 infection in Malaysia

The originality of the research comes from the fact that research on Malaysia's COVID-19 situation has not been thoroughly conducted, especially how the stringency index (lockdown measures) affects COVID-19 infections. As a Malaysian, we were unable to obtain vaccines as soon as possible and also had poor lockdown measures due to the lack of proactiveness from the Malaysian government. Thus, the importance of the research would indicate the benefits of vaccination as well as implementing lockdown measures on reducing the spread of COVID-19. My interest in this area is from the fact that due to the pandemic, there has been a myriad of negative effects such as online schooling, unemployment, increased hospitalizations and deaths. Hence, with my research, I hope to highlight the importance of having an effective National Pandemic Strategy in order to be prepared for any future virus outbreaks.

Background/Literature:

A search on the National Center for Biotechnology Information and The New England Journal of Medicine yielded three important research papers titled "The impact of vaccination on COVID-19 outbreaks in the United States", the "Effect of Vaccination on Household Transmission of SARS-CoV-2 in England" and the "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine". The first paper indicated that vaccination reduced the overall infection rate from 9.0% to 4.6% over 300 days for the Pfizer and Moderna vaccines. On the other hand, the second paper indicated that the likelihood of household transmission was roughly 40 to 50% lower in households that have been vaccinated when compared to households that are unvaccinated. Lastly, in the third paper, the results highlight a 95% protection against COVID-19 infections.

This would indicate that in our data analysis, we would expect the overall infection rate to decrease in Malaysia as vaccination rates increased. The results of these papers relate to this question because they explore the effects of vaccination on COVID-19 infections. The research question I will study relates to this because I will be exploring vaccines affect the COVID-19 situation in Malaysia. Additionally, I will be exploring how lockdown measures (stringency index) affect COVID-19 infections which are different to the research papers mentioned.

A point to note is that although not specified in the dataset I will be exploring, it is known that 51% of vaccines used in Malaysia is the Sinovac vaccine as of Sept 9, 2021, which could affect our case rates differently when compared to the results of the research conducted in the United States and the United Kingdom as mentioned previously.

Data source:
The main variables of my analysis include “data_MYS_vaccine”, “data_MYS_cases_v”, “data_MYS_strin” and “data_MYS_cases_s”.

“data_MYS_vaccine” represents the total number of people who received at least one vaccine dose per 100 people in the total population. “data_MYS_cases_v” represents the new COVID-19 cases per million people corresponding to “data_MYS_vaccine”. “data_MYS_strin” represents the stringency index which is a composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest response). Lastly, “data_MYS_cases_s” represents the new COVID-19 cases per million people corresponding to “data_MYS_strin”.

I have included the “data_MYS_vaccine” and “data_MYS_strin” variables as they will allow me to determine how the vaccination rate and government response affect COVID-19 infections in Malaysia. Furthermore, I had to determine the COVID cases (“data_MYS_cases_v” and “data_MYS_cases_s”) that correspond to the specific independent variable of vaccination and stringency since NA observations were present in their data.

Exploratory data analysis:

Table 1: Descriptive Statistics including Mean and Standard Deviation

Variable	Mean	Standard Deviation
data_MYS_vaccine	27.734	28.0658
data_MYS_cases_v	270.95	209.122
data_MYS_strin	64.48	16.67272
data_MYS_cases_s	72.2919	108.3967

Figure 1: Histograms of Variables

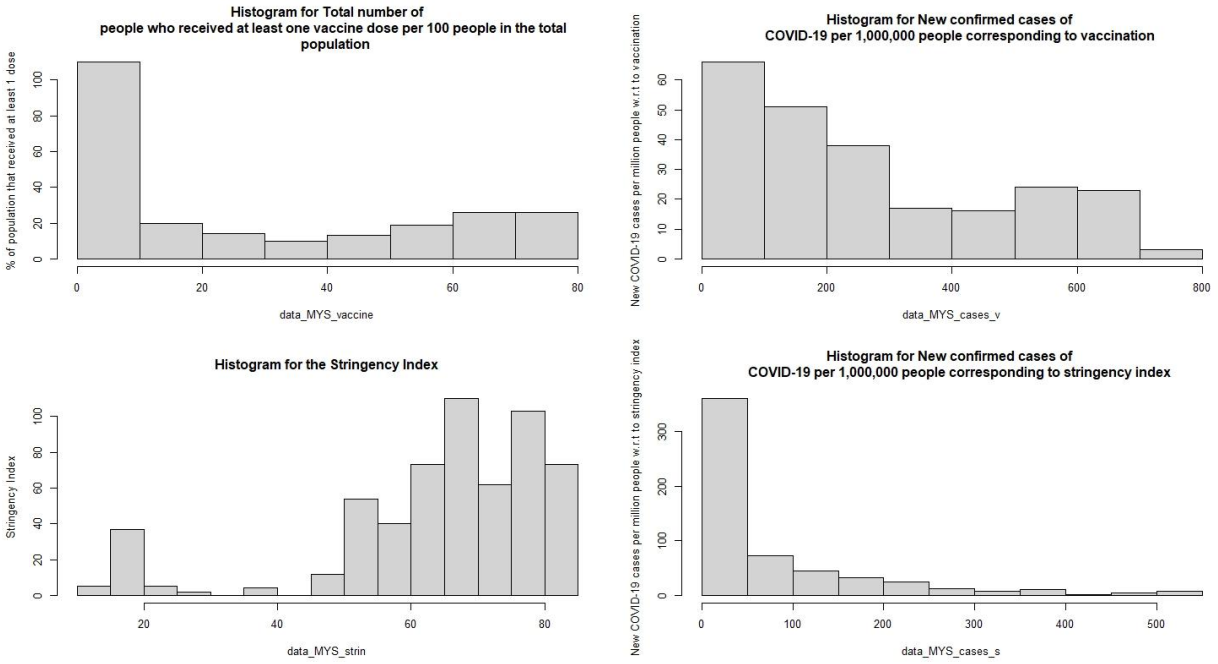


Figure 2: Boxplots of Variables

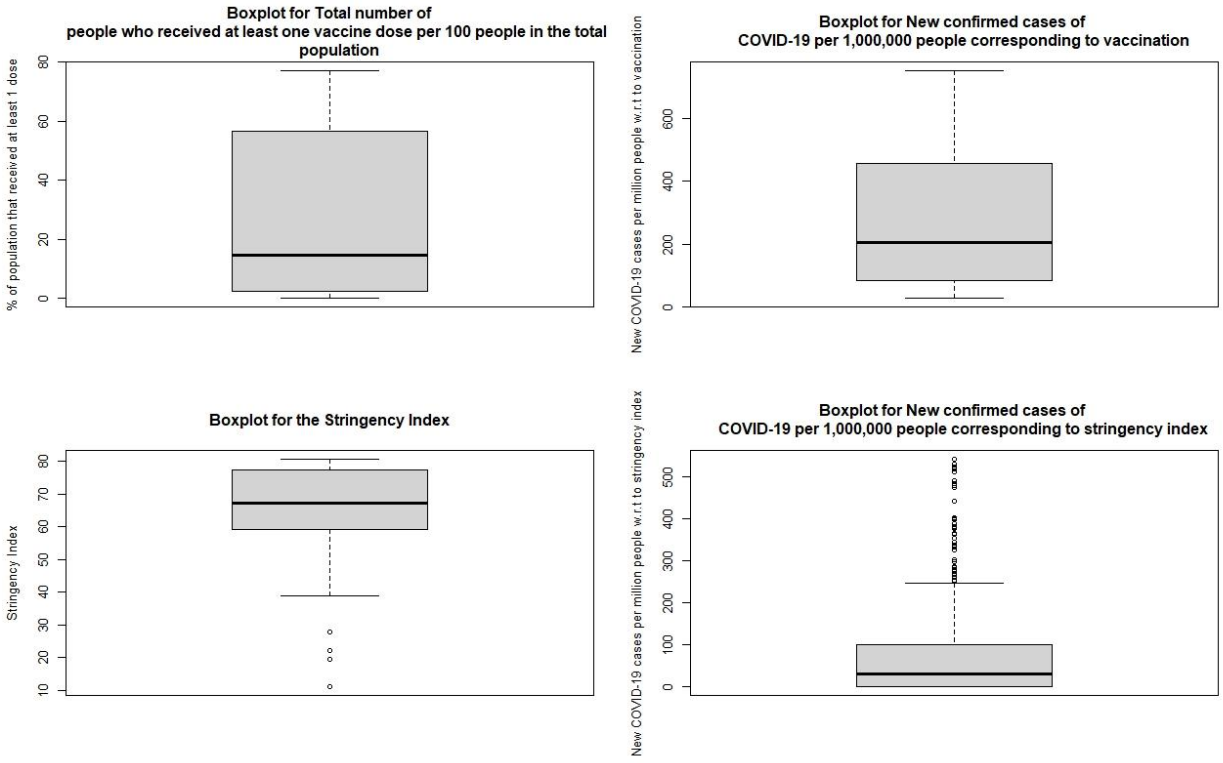
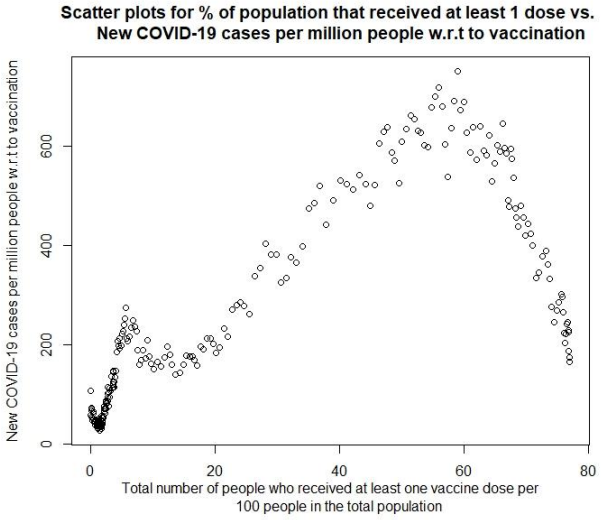
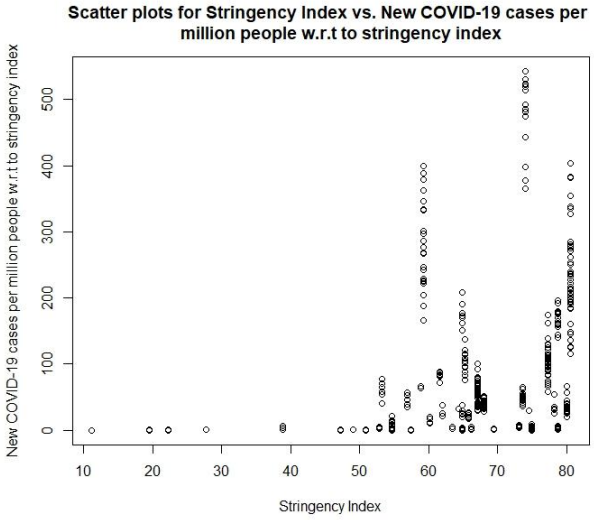


Figure 3: Scatter Plots of Variables





Discussion of what is observed in the EDA:

The sample comprises of data such as the mean new COVID-19 cases per million people corresponding to vaccination rate being 270.95 individuals as well as the mean new COVID-19 cases per million people corresponding to the stringency index being 72.2919 individuals. These descriptive statistics can be found in table 1.

When doing my EDA, I also observed much missing data thus, I omitted these NA variables and made new columns that have the independent variables of vaccination rate and stringency index correspond to the COVID-19 cases per million people.

In figure 1, the histograms depict that the total number of people who received at least one vaccine dose per 100 people in the total population (data_MYS_vaccine), new COVID-19 cases per million people corresponding to “data_MYS_vaccine” (data_MYS_cases_v) and the new COVID-19 cases per million people corresponding to “data_MYS_strin” (data_MYS_cases_s) have distributions that are right-skewed. Hence, we expect the mean and median to be greater than the mode. On the other hand, the histogram depicts that the histogram for the stringency index (data_MYS_strin) to be left-skewed. Thus, we expect the mean and median to be less than the mode.

On the other hand, in figure 2, the boxplots of the stringency index (data_MYS_strin) depicted outliers at the bottom whereas the boxplot of the new COVID-19 cases per million people corresponding to the stringency index (data_MYS_cases_s) have outliers at the top. The boxplots of data_MYS_vaccine and data_MYS_cases_v have no outliers present.

Linear model:
My research question can be answered using a linear model because my independent variables, the vaccination rate and stringency index, are linearly related to my dependent variable, the COVID-19 cases. This can be observed in figure 3 where, for instance, there is a negative correlation between the COVID-19 cases and the percentage of people who received at least one vaccine dose when the percentage is greater than approximately 60%.

I will be using data_MYS_cases_v and data_MYS_cases_s as my response and use the predictors data_MYS_vaccine and data_MYS_strin because I am exploring the relationship between vaccination rate and stringency index against COVID-19 infections.

Some anticipated issues might be that in the left scatter plot of figure 3, we observe both a positive correlation and negative correlation. Thus, we may have to determine two separate lines of best fit, one for the 0 to 60 percent of people who received at least one vaccine dose (positive correlation) and another for when the percentage is greater than 60% (negative correlation). Some things I plan to keep an eye on when working on my data analysis are any potential confounders that should be controlled for when determining the relationship between my independent and dependent variables.

References

Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., Appel, C., Giattino, C., & Rodés-Guirao, L. (2021, May 10). *A global database of COVID-19 vaccinations*. Nature News. Retrieved October 22, 2021, from https://www.nature.com/articles/s41562-021-01122-8.

Yahoo! (n.d.). *Covid-19 deaths among vaccinated rare, mostly Sinovac recipients*. Yahoo! News. Retrieved October 22, 2021, from https://malaysia.news.yahoo.com/covid-19-deaths-among-vaccinated-060400766.html.

Moghadas, S. M., Vilches, T. N., Zhang, K., Wells, C. R., Shoukat, A., Singer, B. H., Meyers, L. A., Neuzil, K. M., Langley, J. M., Fitzpatrick, M. C., & Galvani, A. P. (2021, January 2). *The impact of vaccination on covid-19 outbreaks in the United States*. medRxiv : the preprint server for health sciences. Retrieved October 22, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709178/.

Others, D. M. and, Others, M. B. and, & S. D. Anker and Others. (2021, October 6). *Effect of vaccination on household transmission of SARS-COV-2 in England: Nejm*. New England Journal of Medicine. Retrieved October 22, 2021, from https://www.nejm.org/doi/full/10.1056/NEJMc2107717.

Polack, F. P., Al., E., for the C4591001 Clinical Trial Group*, Author AffiliationsFrom Fundacion INFANT (F.P.P.) and iTrials-Hospital Militar Central (G.P.M.), Longo, E. J. R. and D. L., Others, A. J. S. and, Others, J. M. and, E. Mendes de Oliveira and Others, Others, D. M. and, Others, M. B. and, & S. D. Anker and Others. (2020, December 31). *Safety and efficacy of the BNT162B2 mrna covid-19 vaccine: Nejm*. New England Journal of Medicine. Retrieved October 22, 2021, from https://www.nejm.org/doi/full/10.1056/nejmoa2034577.