I. Introduction

Motivation

The originality of the research stems from the reason where research on Malaysia's COVID-19 situation has not been conducted thoroughly. As a Malaysian, we were unable to obtain vaccines promptly and had poor lockdown measures due to the lack of proactiveness from the government. Furthermore, there has been a myriad of negative effects such as unemployment, increased hospitalizations, and deaths.

The research would indicate the benefits of vaccination as well as implementing lockdown measures on reducing the spread of COVID-19. Hence, I hope to highlight the importance of having an effective National Pandemic Strategy 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 research papers titled "The impact of vaccination on COVID-19 outbreaks in the United States" (Moghadas, n.d.), the "Effect of Vaccination on Household Transmission of SARS-CoV-2 in England" (Harris, n.d.) and the "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine" (Polack, n.d.). These papers indicated that vaccination rates reduced the infection rate of COVID-19 in their samples.

This would indicate that in my data analysis, I would expect the overall infection rate to decrease in Malaysia as vaccination rates increased. The results of these papers relate to my research because they explore the effects of vaccination on COVID-19 infections. Furthermore, my research question differs from the background since I will be exploring how lockdown measures (stringency index) in addition to vaccines affect the COVID-19 situation in Malaysia.

II. Methods

The data analysis started off with viewing and cleaning the data where I filtered for data on Malaysia as well as checked for missing observations and omitting them.

I proceeded to begin my analysis by first randomly dividing up the dataset into train and test datasets with 50/50 proportions. I conduct most of my analysis using the training dataset only by first obtaining summary statistics and conducting an EDA. I then constructed my model and begin assessing the linear model assumptions (linearity of the relationship, uncorrelated errors, constant error variance, normality of the errors) formally by first checking that the additional conditions are satisfied. Since condition 2 failed, I applied the Box-Cox transformation to satisfy the condition. In my transformed model, the additional conditions are satisfied thus, I begin to interpret the residual plots ensuring that my assumptions are satisfied.

After fulfilling the assumptions, I identified problematic observations including leverage and influential points as well as outliers using their measures and cut-offs. I determined the VIF of my model to ensure that there is no multicollinearity present in my predictors. Lastly, I conducted an F-test to test whether a significant linear relationship exists overall as well as t-tests on my model predictors.

Now that I have a potential model, I will fit the model in my test data set and compare its properties to those in the training dataset. If the models look very similar to how they performed in the test dataset, I can conclude that my model is validated. (Daignault, STA302 Module Lectures, 2021)

III. Results

Data source:

The main variables of my analysis include vaccination rate ("data_MYS_vaccine"), COVID-19 cases ("data_MYS_cases"), stringency index ("data_MYS_strin") and vaccine effectiveness ("data_vaccine_effect").

The vaccination rate represents the total number of people who received at least one vaccine dose per 100 people in the total population. COVID-19 cases represent the new COVID-19 cases per million people in the population. The stringency index represents 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). The vaccine effectiveness denotes a binary variable that represents a cut-off at 60% of the vaccination rate where the cases begin to decrease. Since the effect of the vaccinations is not immediate as most of the population is still unvaccinated and susceptible to contracting the virus, it requires a threshold of the population to be vaccinated before cases start to decrease. Hence the vaccine effectiveness acts as an interaction term with the vaccination rate.

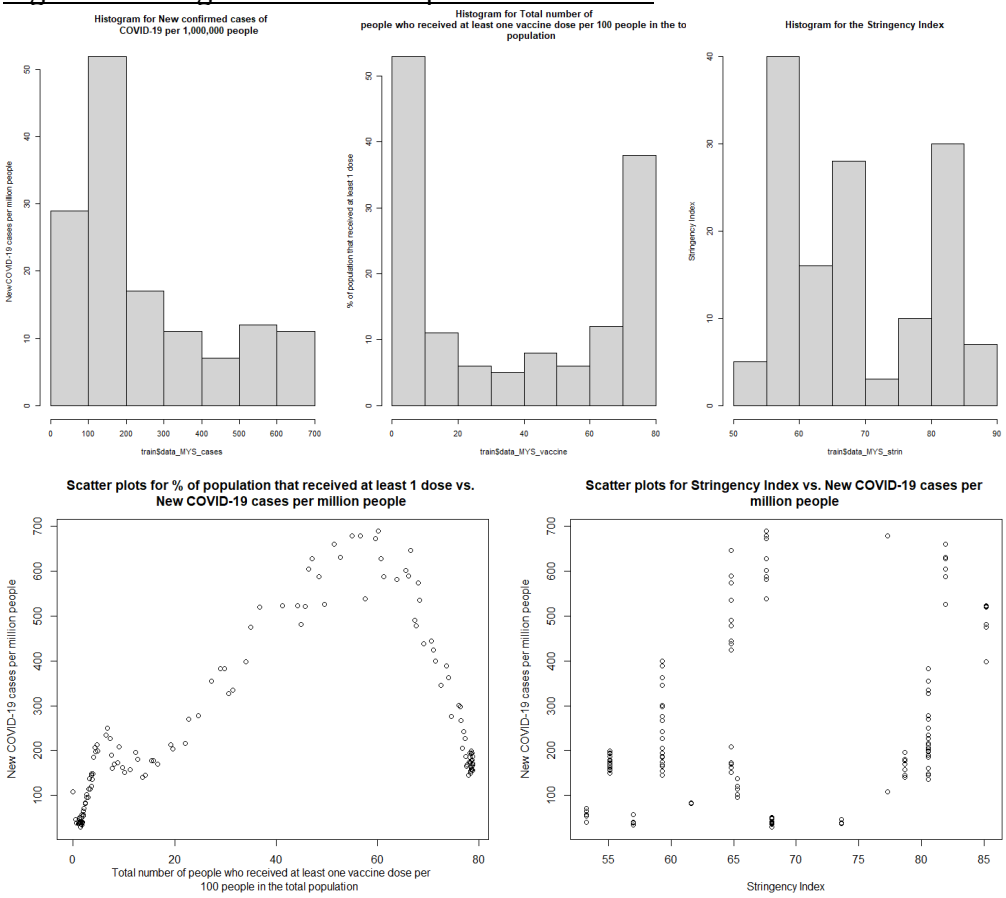
Exploratory data analysis:

Table 1: Descriptive Statistics including Mean and Standard Deviation

Variable	Mean	Standard Deviation
COVID-19 Cases	251.832	188.374
Vaccination Rate	36.473	31.949
Stringency Index	68.315	10.191
Vaccine Effectiveness	0.360	0.482

The descriptive statistics can be found in table 1. When conducting my EDA, I observed many missing data from some of my variables. Thus, I omitted the rows containing the NA values. The limitation of omitting missing variables will be found in the discussion.

Figure 1: Histograms and Scatterplots of Variables



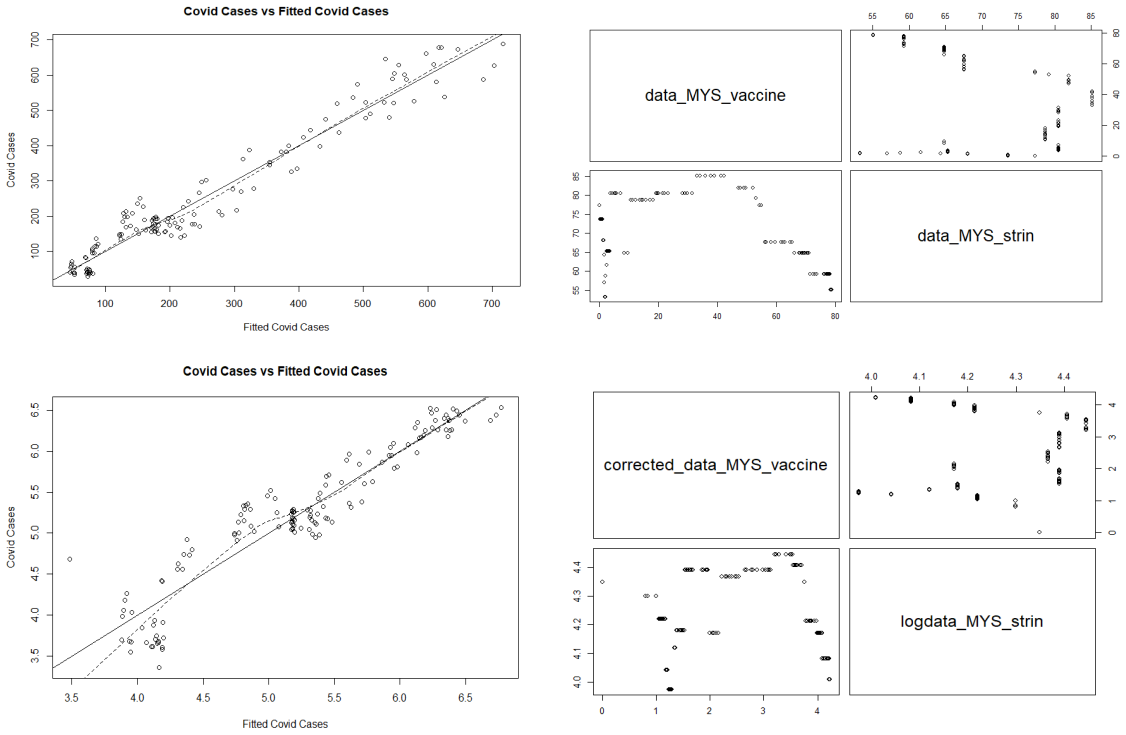
In figure 1, the histograms depict skewness, and the scatterplots indicate non-linearity. Therefore, we observe that there may be issues with non-linearity and non-normality which violates our model assumptions.

The model that is chosen for the analysis is as follows:

$$COVID\ Cases = Vaccination\ Rate + Vaccination\ Effect + Stringency\ Index + Vaccination\ Rate \cdot Vaccination\ Effect$$

Before using the residual plots, I checked for the additional conditions to ensure that my residual plots provide me with reliable results.

Figure 2: Checking additional conditions to ensure reliability of residual plots (top: before transformation, bottom: after transformation)

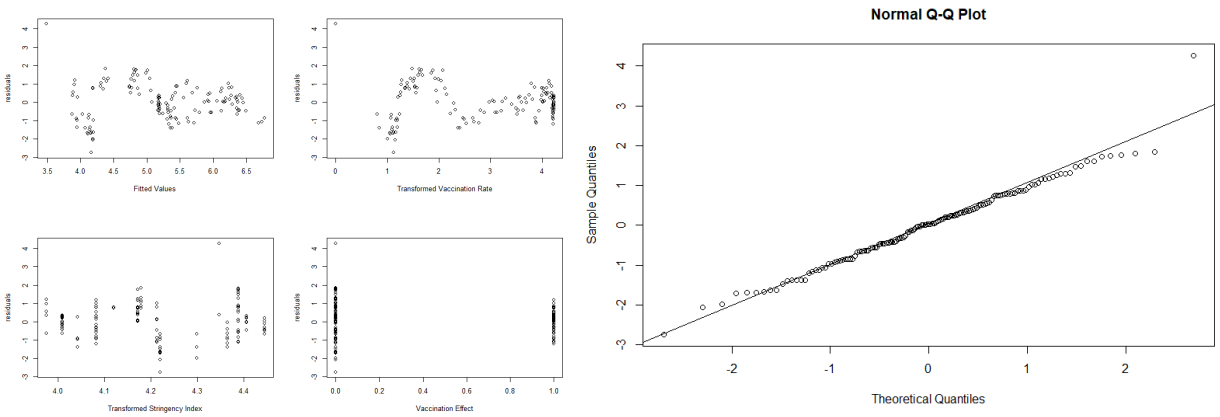


As observed in the plot of the cases against the fitted cases, the points are randomly scattered around the line of best fit which indicates that condition 1 holds. In the pairwise plots between the predictors, we do not observe a linear relationship between them hence, condition 2 does not hold. Since condition 2 does not hold, we apply the Box-Cox transformation to obtain a more desirable model satisfying the conditions.

Upon applying the transformation, the model obtained is as follows satisfying both conditions.

$$\log(COVID\ Cases) = Vaccination\ Rate^{0.33} + Vaccination\ Effect + \log(Stringency\ Index) + Vaccination\ Rate^{0.33} \cdot Vaccination\ Effect$$

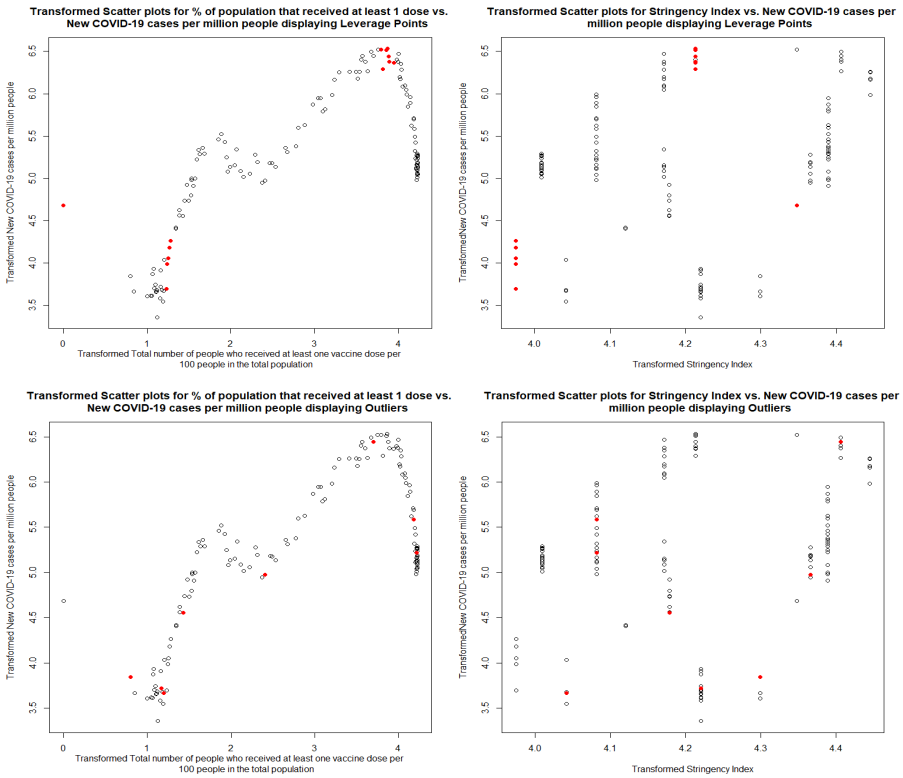
Figure 3: Residual Plots and Q-Q plot of our predictors and fitted values of the transformed model



Checking for the model assumptions, there is no discernible pattern seen in the residual plots and the residuals are uniformly scattered around 0. The Q-Q plot appears relatively without severe deviations at the end. The main concern is that the vaccination rate still depicts a slight pattern indicating non-linearity which I am unable to improve further despite attempting numerous transformations thus, the impact of it will be discussed in the limitations.

Checking for multicollinearity in our model using the VIF, we obtained 3.965169 for the vaccination rate and 2.765220 for the stringency index. Therefore, there is some amount of multicollinearity between the vaccination rate and stringency index, but it is not severe.
I obtained very high values of multicollinearity between the vaccination effect and the vaccination rate interacted with the vaccination effect. This is expected since the vaccination effect acts as an interaction term with the vaccination rate.

Figure 4: Scatterplots of predictors vs. response displaying leverage points (top) and outliers (bottom)



In figure 4, I observe leverage points, influential points, and outliers in my transformed model. Since I do not have contextual reason to remove these points and outliers, I leave them in the model and discuss its limitations in the following discussion.

Conducting a stepwise selection process would lead me to select the same model previously with an AIC of -332.08 and an adjusted R-squared of 0.8797 indicating a good model fit. Therefore, the final model is as follows.

$$\log(COVID\ Cases) = 0.77842\ VaccinationRate^{0.33} + 17.32577\ VaccinationEffect + 1.52095 \log(StringencyIndex) - 4.35800\ VaccinationRate^{0.33} \cdot VaccinationEffect - 3.12730$$

With this final model, I conducted an F-test which provided a result of 253.4 with a p-value of $2.2 \times 10^{-}16$ thus, I reject the null hypothesis meaning that at least one of the predictors is linearly related to COVID cases. Furthermore, I conducted t-tests on the predictors which indicate statistical significance at the 0.001 level. Therefore, I can conclude that the predictors are all linearly related to COVID cases.

Lastly, using the preferred transformed model, I fit the model in my test dataset and proceed to compare their properties to those in my training dataset. Comparing both the models, I obtained similar results for both datasets (Appendix: Test Dataset Results). Lastly, I checked my linear model assumptions, and no alarming new patterns are found. Therefore, I can conclude that my model is validated.

IV. Discussion

If we change vaccination rate to the power of 0.33 by 1 unit, we expect the COVID cases to change by $[(0.77842 \times 0.33\, VaccinationRate^{-0.67} - 4.35800 \times 0.33\, VaccinationRate^{-0.67} \cdot VaccinationEffect) \times 100]$ percent (log-level). If we change vaccination effect by 1 unit, we expect the COVID cases to change by $(17.32577 - 4.35800\, VaccinationRate^{0.33}) \times 100$ percent (log-level). Lastly, if we change the stringency index by one percent, we expect COVID cases to change by 1.52% (log-log).

Since we require the interaction term of vaccination effect to denote when the COVID cases begin to decrease, we are only able to analyse the partial effects of vaccination rate on covid cases and vaccination effect on covid cases. For instance, if there is an effect of vaccination on the covid cases where the vaccination rate is greater than 60%, the vaccination effect will equal 1 thus, there is a negative relationship between vaccination rate and covid cases. Additionally, at high levels of covid cases, restrictions tend to tighten denoting the positive relationship between the stringency index and cases.

Therefore, for Malaysia to curb the effects of a pandemic efficiently, the government should have vaccines administered to the population immediately to achieve the vaccine effectiveness cut-off, as well as impose lockdown measures for a continuous period without easing the restrictions until vaccination rate reaches 60%.

Based on my model results, there are several limitations. Firstly, there are missing values in the dataset therefore, it would lead to a reduction in statistical power as well as the representativeness of the sample. Furthermore, I am unable to satisfy linearity fully since there is a slight pattern in the vaccination rate residual plot despite using transformations. Hence, I utilised the interaction term of vaccine effect with the vaccination rate to correct this issue. Lastly, the problematic observations in the model could potentially affect the interpretation and coefficients since it displaces the regression line.

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.

Daignault, K. (2021). STA302 Module Lectures.

Jacob Yoke Hong Si
Student #: 1005815806

Appendix

Figure 5: Plots to help validate the training dataset using the test dataset