**Preliminary Report**

**1. Introduction**

*1.1 Background*

COVID-19 pandemic has been one of the most severe issues of the world as it appears to be equivalent to a world war. There are tons of adverse ramifications physically and mentally affected by this pandemic. Obviously, it leads to serious illnesses and death without any effective mitigation approaches at the beginning. In another aspect, COVID-19 has been emerging to destroy world growth in economics which drives all activities and developments to the world. This interests the authors to come up with the question of “How does COVID-19 affect the growth of world economics?”.

One of the most apparent indicators of world economics is the unemployment rate which represents whether the world is moving forward effectively and, in turn, people have the ability to pay for their own livings. High unemployment rate refers to the problem that companies have less capability in taking care of their employees leading in both situations of unemployment and laying off while it slows down the pace of world growth as it should actually be. According to the Congressional Research Service research, it is surprising that the highest unemployment rate occurred in the era of COVID-19 [1], much higher than any pandemic in the history of the U.S. This can be initially confirmed that there are some significant relationships between the COVID-19 pandemic, which allows us to perform data analysis regarding establishing the model to explain the variables and confirming that the relationships exist.

*1.2 Objective*

1) To refine a broad question to a more specific question that can be answered by data analysis.

2) To perform an exploratory data analysis to examine the structure and components of the COVID-19 and unemployment data set as well as the distributions and relationships.

3) To construct the statistical model to describe the variables and relationships as well as to perform hypothesis testing to confirm the existence of relationships.

4) To interpret the results and suggest the implications that should be taken.

**2. Methodologies**

*2.1 Refining the question*

According to the broad question we have set, “How does COVID-19 affect the growth of world economics?”. Our first objective is to refine this question into a more specific question which possesses the characteristics of a good question. After a thorough discussion, the refined question is

***“Are there any significant relationships between COVID-19 and the unemployment rate in the U.S.?”***

This is an exploratory question which requires the use of statistical modeling to describe the data and variables. Furthermore, this can be an inferential question since we aim to test for the significance of the relationship between these variables. The following criteria also explains why the authors decide to select this refined question.

(1) It is certain that people these days have been paying attention to this large-scale pandemic and are always aware of it. Therefore, it is **of interest** to individuals as well as the organizations in the U.S. not only the medical industry but also every industry as they all confront this difficult situation. According to the research, we have also known that the unemployment rates sharply increased to the peak in every sector. Additionally, the analysis provides the applications of the policies development in the sense that when there is another severe pandemic, the protective policies will be released in a timely manner which alleviate both physical and mental loss.

(2) The question has been clearly **established in the plausible framework** since it is grounded by the actual situation of COVID-19 followed by the news and many figures showing that unemployment rate has been skyrocketing. Additionally, there are a number of data sources regarding the number of cases, deaths, vaccinations etc. as well as the unemployment rate by states in the U.S. This refined question is **answerable** and although there is some research attempting to answer the question, it has not been clearly confirmed in a statistical approach.

(3) This question is much less difficult to answer as we specified to focus on the unemployment rate representing the world economic indicator and the variables from COVID-19 figures which **possessed the specificity** after refining from the very general question.

*2.2 Designs of experiment*

In this section, the authors sequentially proceed the steps of the analysis to ultimately answer the question we have set. The experiment, in the full analysis, is divided into five parts ranging from data collection process, exploratory data analysis, statistical model fitting, testing the hypothesis of the relationship to interpreting the results and answering the question. However, in this preliminary report, we are proposing the procedure and feasibility of the result we would obtain.

*2.2.1. Data collection phase*

According to the study, we are searching for four different sources of data. Our response data, an unemployment rate, is collected from [www.kff.org](https://www.kff.org/other/state-indicator/unemployment-rate/?dataView=0&currentTimeframe=17&selectedRows=%7B%22states%22:%7B%22all%22:%7B%7D%7D,%22wrapups%22:%7B%22united-states%22:%7B%7D%7D%7D&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D) (Kaiser Family Foundation) each month for every state in the U.S. For the predictors, we mainly look for the data of COVID-19 figures and number of cases such as new cases and death from [www.cdc.gov](https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm), vaccination rates in the U.S. from [ourworldindata.org](https://ourworldindata.org/us-states-vaccinations). We also consider lockdown days from [www.usatoday.com](https://www.usatoday.com/storytelling/coronavirus-reopening-america-map/) as the significant variable which might affect the unemployment rate since locking down leads to the suspension of the economic activities. All of these data are consolidated into one data set with some manipulations. We also perform an exploratory data analysis to initially identify the variables patterns and their distributions.

*2.2.2. Statistical modeling phase*

According to the data and the question we set, we roughly perform an exploratory data analysis, and we consider utilizing the modeling techniques to describe the variables in the statistical approach, beginning with the multiple linear regression with the model

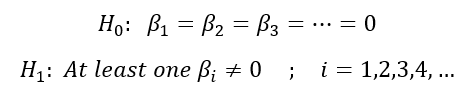


Note that nCase is the number of new cases in each month, nDeath is the number of deaths, vaccRate is the vaccination rate and lockdownDay is the days of lockdown in each month.

This is our primary model so that we could continue building the second or third model which also describes the variables and patterns such as the multiple linear regression with interaction terms or adding some categorical variables such as season, if we later found out that there might be an effect of seasonality in our data.

*2.2.3. Hypothesis testing and model selection phase*

Nevertheless, by merely fitting the model itself, does not explain the significance of the relationship, we, therefore, continue to perform the hypothesis testing to confirm whether the relationship between the unemployment rate and these predictors is significant. The hypothesis is set as follow:



According to the hypothesis setting, we aim to perform F-test in these multiple linear regression models to see if there are significant relationships while we are comparing these models using ANOVA F-test to ultimately select the best model that best describes the data which also corresponds to the real life situation.

*2.2.4. Possible answers*

The authors identify two possible outcomes of the analysis which are:

(1) There is no relationship between COVID figures and an unemployment rate. This would be one evidence against the assumption that COVID-19 significantly affects the growth of the world economy.

(2) There is a significant relationship between the above-mentioned variables which would be the part of evidence that can confirm the assumption of COVID-19 effect on the overall growth of world economics.

**3. Reference**

[1] Gene Falk, Romero P., Nicchitta I., Nyhof E. Unemployment rates during the COVID-19 pandemic. The Congressional Research Service, U.S.A.