****

**ITEC 621 Predictive Analytics Project**

**Project Name: The Financial Impact of COVID-19**

**Class Section:** [**ITEC-621-002.2021S**](https://american.instructure.com/courses/14747)

**Team Number: 2**

**Team Members: Chang, Ding, Lu, Xu**

**Last updated: 4/28/2021**

**Deliverable Number: 4**

**(1) The business case**

Since the COVID-19 pandemic outbreak, the economy has been affected seriously. This study considers the Nasdaq Composite Index as an important indicator of market valuation and confidence in public companies, allowing us to evaluate the US economy during the pandemic era. We would like to compare the number of new positive cases and deaths, daily vaccination, cumulative cases and deaths with the close value of Nasdaq Composite Index to know whether the severity of the outbreak has affected the judgment of people in financial markets about the future expectations of public companies. In short, the most striking impact of the unpredictable pandemic on the world economy is capital markets. However, to further predict the alteration of the market, we need to first understand the specific changes of the economy during the pandemic period, and seek to investigate if the trend of the pandemic affects the United States capital market growth.

**(2) The business question**

How does the status of the pandemic in the United States affect the Nasdaq's performance?If there is somehow some kind of relationship between these two objects, then people can consider the change of the pandemic as an index to reflect the fluctuation of the Nasdaq. This study is aimed to look at the stock index to understand the impact of the epidemic on the confidence of the entire US financial market. It is crucial to study the economic changes in the pandemic era so better future investment decisions can take place.

**(3) The analytics question**

Which predictors have the most impact on the closing price of Nasdaq Composite?

The first step is to find useful information about the pandemic as our predictor, such as daily positive cases, cumulative cases, daily deaths and cumulative deaths. In addition, for the Nasdaq, we use the closing price as our response variable, assuming that the closing price represents the performance of Nasdaq. In addition, we added a variable of daily vaccination into our calculations to see if the vaccine had any effect on the stock index.

**(4) Method**

Because our response variable is quantitative, OLS regression modeling with variable selection, variable transformation specifications will be used for our analysis for the first step. Then, Ridge and LASSO methods will be used to further evaluate the data.

**(5) Dataset Description**

Our dataset has the Covid-19 daily cases, cumulative confirmed cases, daily deaths, cumulative deaths, daily vaccinations and the close market price of NASDAQ Composite, the data is from 1/2/2020 to 4/1/2021. Because the stock market is closed on weekends or vacation, our dataset does not include these closing days so the total observation is 315.

Date: From 1/2/2020 to 4/1/2021

Daily\_cases: Number of COVID-19 confirmed cases for each day

Cumulative\_cases: Number of cumulative COVID-19 confirmed cases

Daily\_deaths: Number of death cases from COVID-19 for each day

Cumulative\_deaths: Number of cumulative death cases from COVID-19

Daily\_vaccinations: Number of COVID-19 vaccinations for each day

Open: Value of stocks when stock market opening

Close: Value of stocks when stock market closing

We get relevant data sets from the Humanitarian Data Exchange and Yahoo Finance website.

**(6) Descriptive Analytics**

According to the histograms, we found that daily data, such as daily cases, daily deaths, and daily vaccinations are not normally distributed. The histograms showed right-skewed. These data may cause our analysis to have a normality problem. Therefore, we used WLS to fix our OLS model. Then we used the Durbin-Waston test to know that we had a time series problem, so we lagged by 1 to fix this problem. Last, we used variable selection to get our final model.

**(7) Initial Set of Predictors**

At first, we selected all variables: Daily cases, Daily deaths, Daily vaccinations, Cumulative cases, Cumulative deaths. The reasons are as follow:

Our group considers the number of daily positive confirmed cases of COVID-19 as a predictor due to the virus contagiousness, if the number of daily cases increases, meaning the spread speed will increase and have further damages for society. The Daily death is also able to indicate the severity of the pandemic. For the cumulative cases and deaths, we assume that the total number of the cases and deaths makes investors have a negative and passive image about the US future and further affect the caprial market. Because the innovation of vaccines accelerated the country to re-open, we consider this predictor could also influence the capital market of the United States.

**(8) OLS Tentative Variable Selection**

After we fited the OLS model, we checked the heteroskedasticity problem and found our model has this issue. Thus, we applied the WLS model, and we proceeded to variable selection to avoid multicollinearity problems among the variables used in our model. We apply a stepwise selection method on WLS and the result is suggesting us using lagged one day closed price of the Nasdaq, daily cases, cumulative deaths as our predictors.

**(9) Inspection of Plots and Model Assumptions**

For the data analysis, we take Close as our response variable. For predictors, we take daily cases, daily death, daily vaccinations, accumulative cases, and accumulative death as our predictors. Our OLS model is lm(Close~Daily\_cases +Daily\_deaths +Daily\_vaccinations +Cumulative\_cases + Cumulative\_deaths). To conduct further analysis, we test this model’s residuals and its normality. The overall model and residuals are normally distributed. The next step is to examine whether heteroskedasticity problems happen in the model or not. The BP Test’s p value is significant, meaning that there is a heteroskedasticity problem in the data. We weighed the variables by absoluting the residuals of OLS and fitting the OLS’s residuals then obtained the weight value with formula :1/fitted(lm.abs.res)^2 to fix the heteroskedasticity problem. We eventually weighed the OLS and got the WLS as our model. We confirmed that this model suffered a time-series problem after applying the Durbin-Walson test. We lagged the close price for one day to solve the time-series issue and create a new lagged model. Then, we use the stepwise method to process variable selection in order to find the best predictors. The final linear regression model is lm(formula = Close ~ Close.L1 + daily\_cases + Cumulative\_deaths, data = na, weights = wts). Our study also applies Ridge Regression and LASSO to make a better fit of prediction. We obtain the best lambda of ridge model is 1.1 and the LASSO model ‘s best lambda is 0.57. To find out the best prediction model, we compare the cross-validation of each model’s RMSE, and the ridge regression model has the lowest RMSE among these three methods. Therefore, the ridge regression is our best prediction method for our model.

**(10) Results**

The best model is Ridge Regression with predictors Close.L1, daily positive cases and cumulative death count. The coefficient of the daily positive cases is positive and so does cumulative death count. We examined the normality and heteroskedasticity problem, and decided a weighted linear model for our final model. Then we selected variables to avoid multicollinearity in the model, we made our model to have R-squared in 98%, which is a high explanatory power within the model. According to the WLS model result, we found that the lagged Close one day variable is the most significant variable to our response. However, the significance of this predictor is due to it being a lagged variable to our response, and therefore, we consider the daily case is the most influential predictor to our dependent variable-- the close price of Nasdaq.

According to our analysis result, the interesting findings come to our view. The coefficient of daily cases and cumulative deaths are positive to the close price of Nasdaq. This is out of our expectations about the result. That is, on average, hold everything else constant, the additional daily cases increases, the Nasdaq’s close price increases 0.0006357 US dollars, and if the cumulative death increases one count, the close price of the Nasdaq increases 0.0004215 US dollars. Besides, we also found that the positive relationship of the Nasdaq close price relates to the previous day’s close price. On average, holding everything else constant, the previous day’s Nasdaq close price increased one US dollars, the Nasdaq close price increased 0.943 US dollars.

**(11) Conclusions**

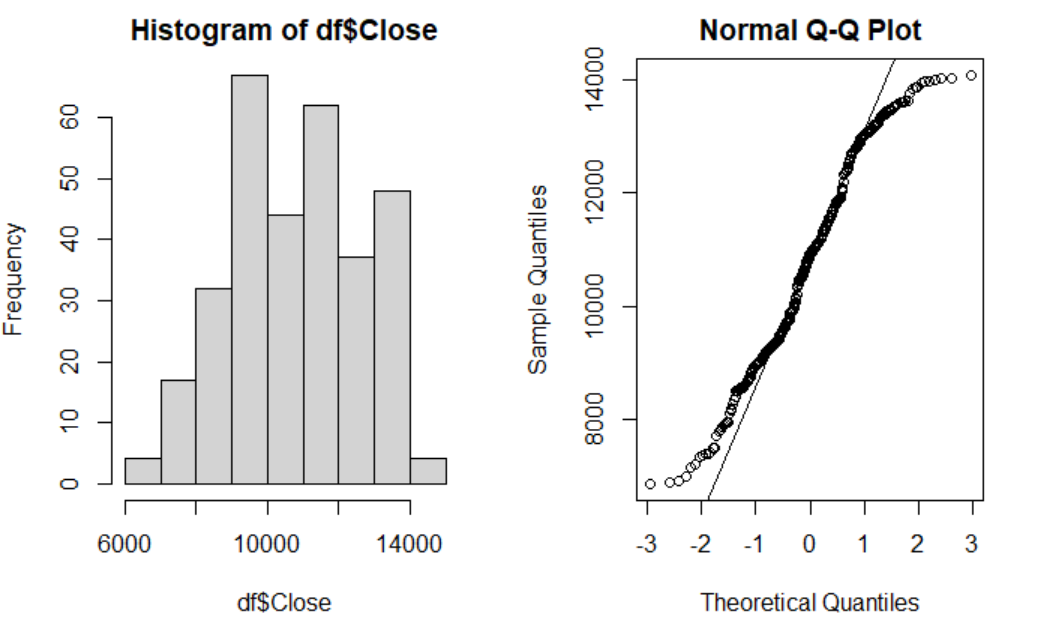
The results we got were quite different from what we had originally imagined. Before we started to analyze the data, we assumed that there may be a negative relationship between the

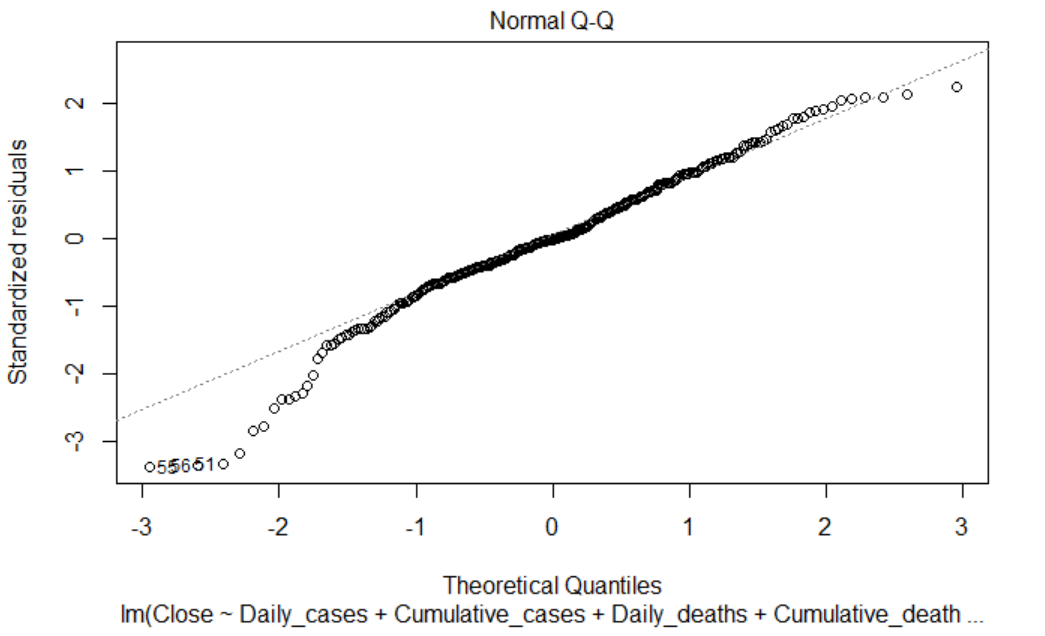
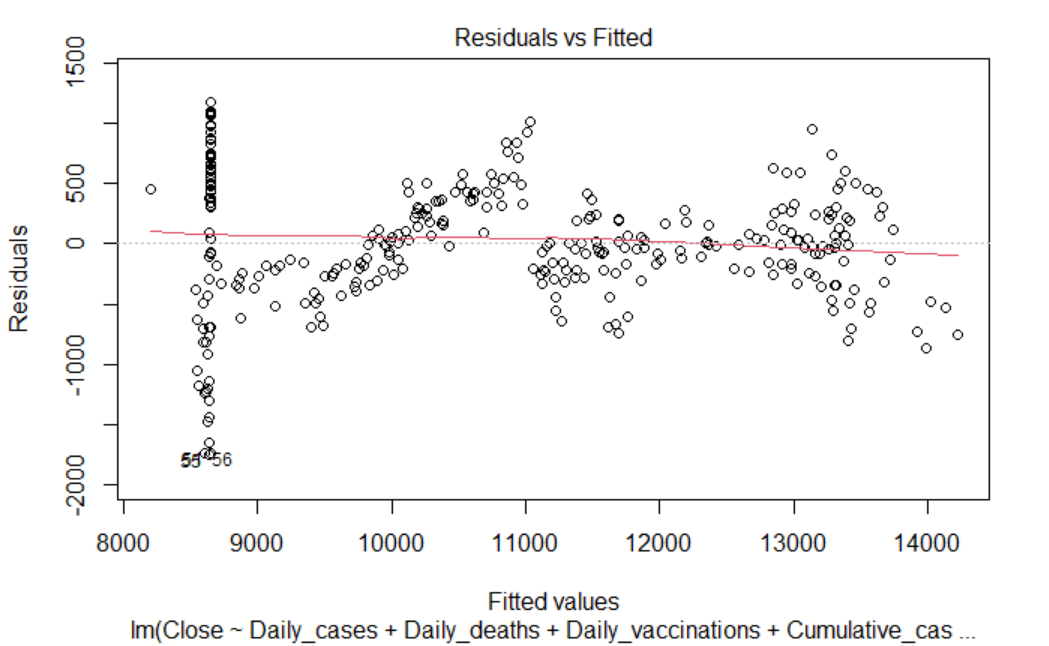
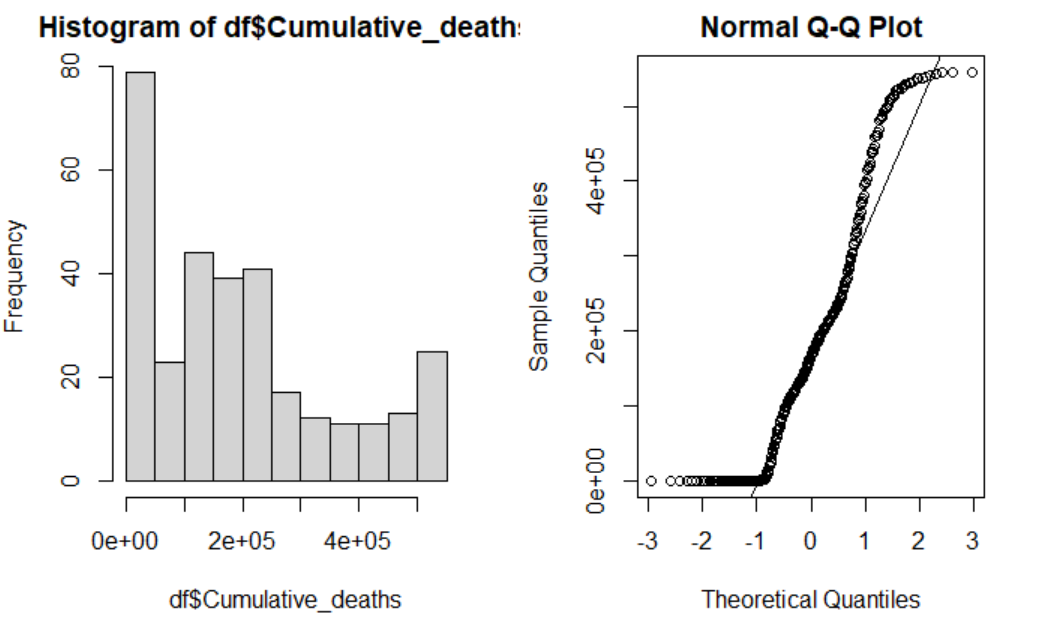
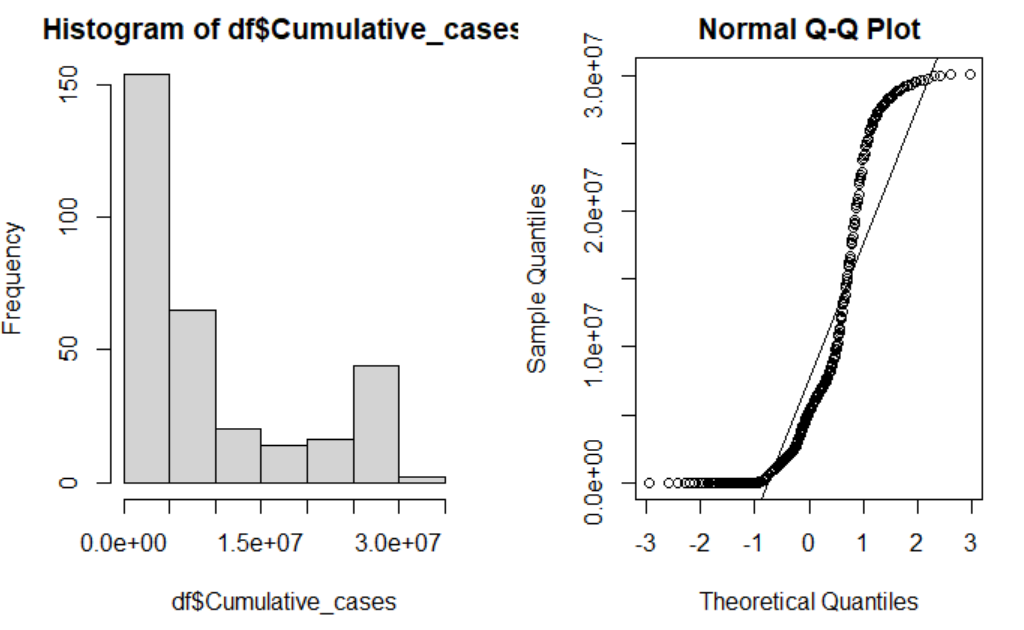
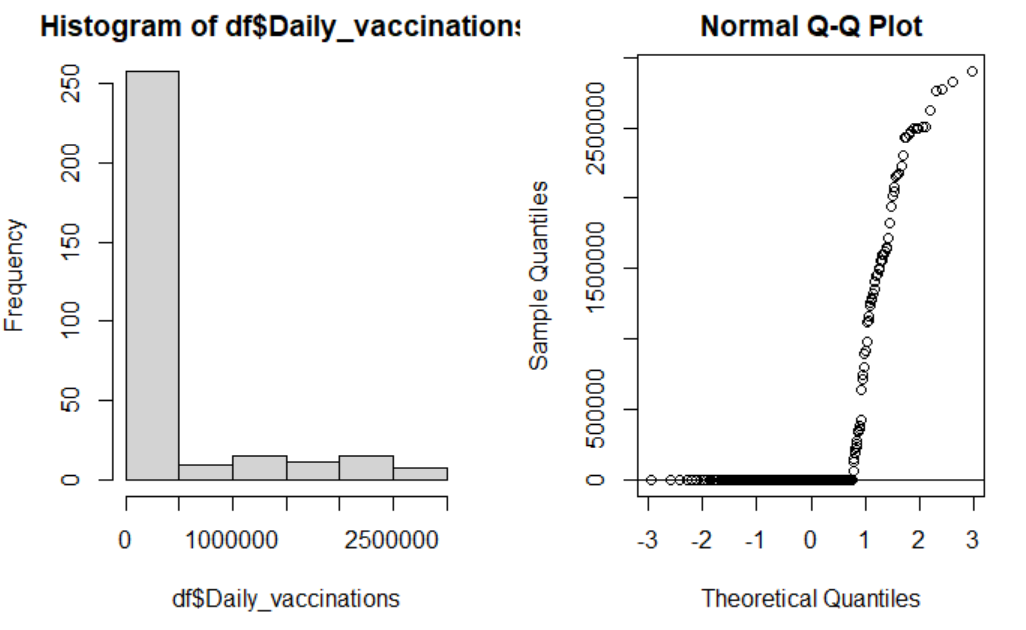
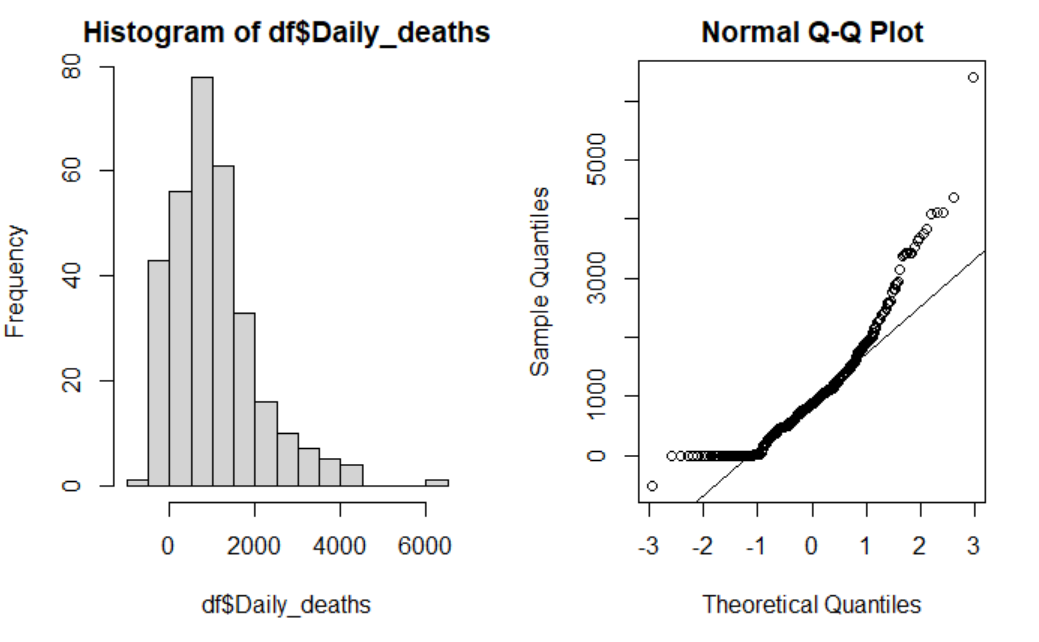
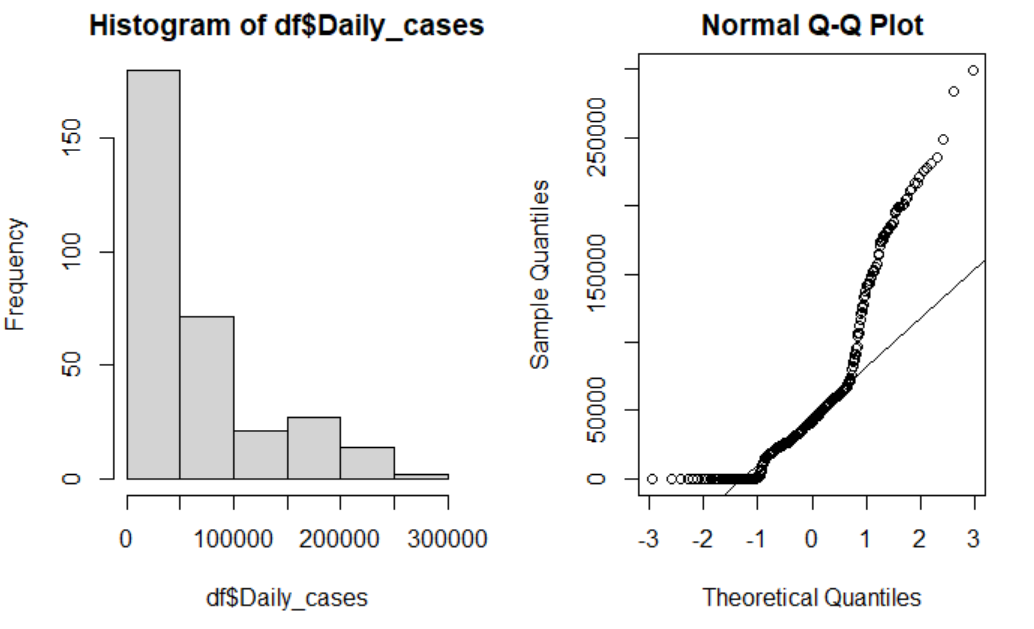
COVID-19 data and the close price of Nasdaq Index. But the fact that there is a positive relationship among them. According to our best model, the ridge regression having Close.L1, daily case and cumulative death count as predictors, the more positive COVID-19 confirmed cases and cumulative death counts, the close price increases accordingly. In other words, we can say the worse the pandemic is, the higher the stock is. However, this analysis result is a paradox to reality. The reasonable explanation is the worse the pandemic the worse the stock performance will be, but the result is not held in a reasonable way. There may be some factors that did not take into account in our analysis and may need further studies to support our result or to refute our result.

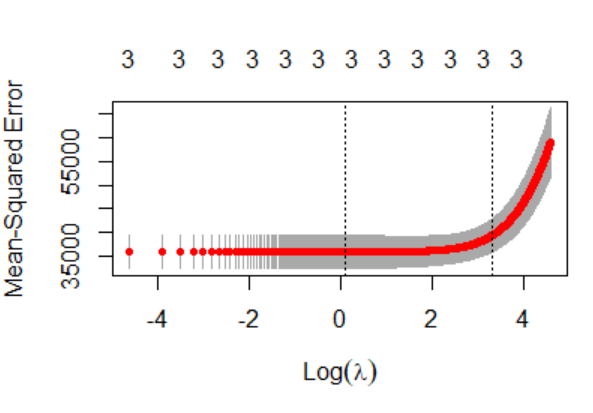
One assumption that makes this result is there is the influence of the government's economic stimulus plan. Since the pandemic has affected the economy, the US President Biden and the former president Trump have repeatedly carried out trillion-dollar stimulus plans in order to restore the economy.These stimulus packages have given financial markets a positive outlook for the future.That's why we concluded that the more severe the pandemic, the higher the stock market.

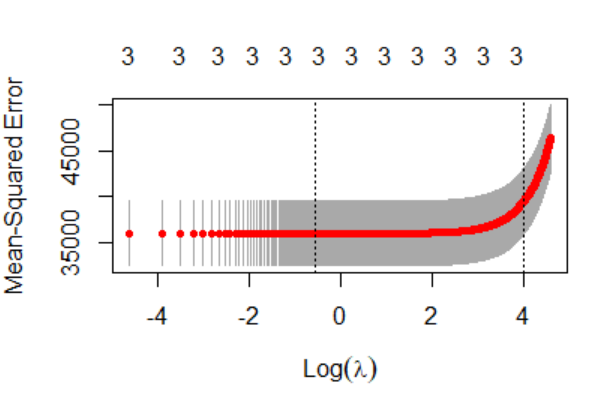
**Appendices**

**A.** **Visuals, Graphs and Plots**



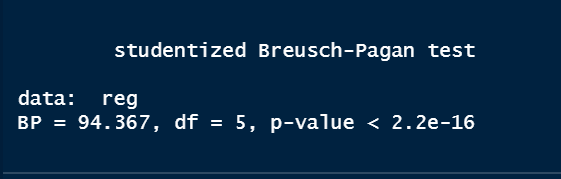


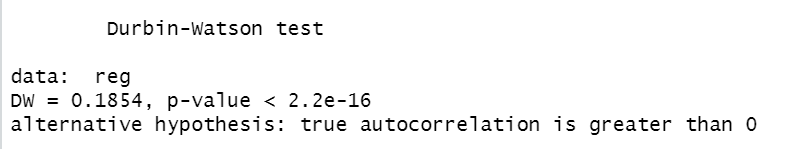


# Ridge 

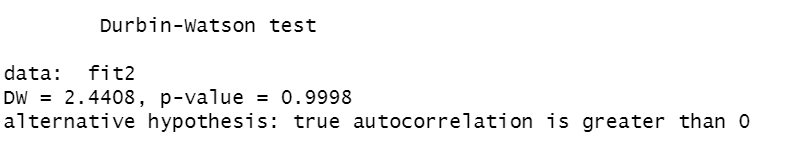
# LASSO

**B.** **Quantitative R Output**

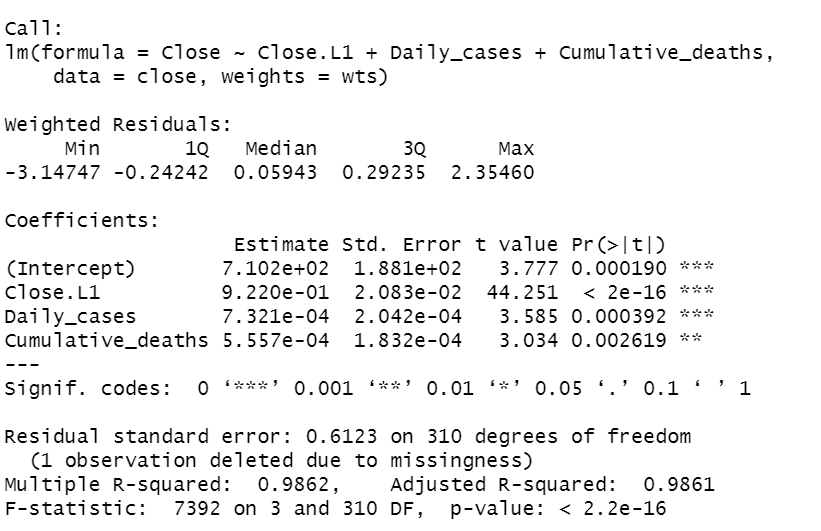


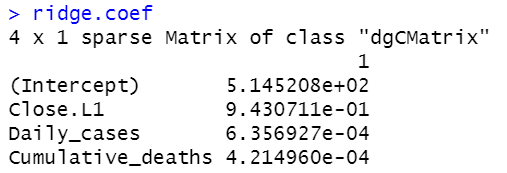


# WLS Model without lagged one day of Close



#Lagged Model: Adding lagged one day of close into predictors.





**C.** **Other**







**D.** **References**

<https://finance.yahoo.com/quote/%5EIXIC/history?p=%5EIXIC>

<https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>