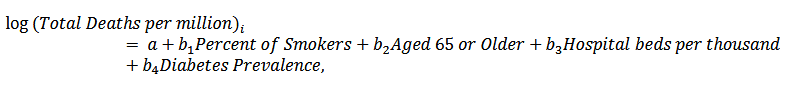
**Assignment 2**

**Business Intelligence (U21807) 2020/21**

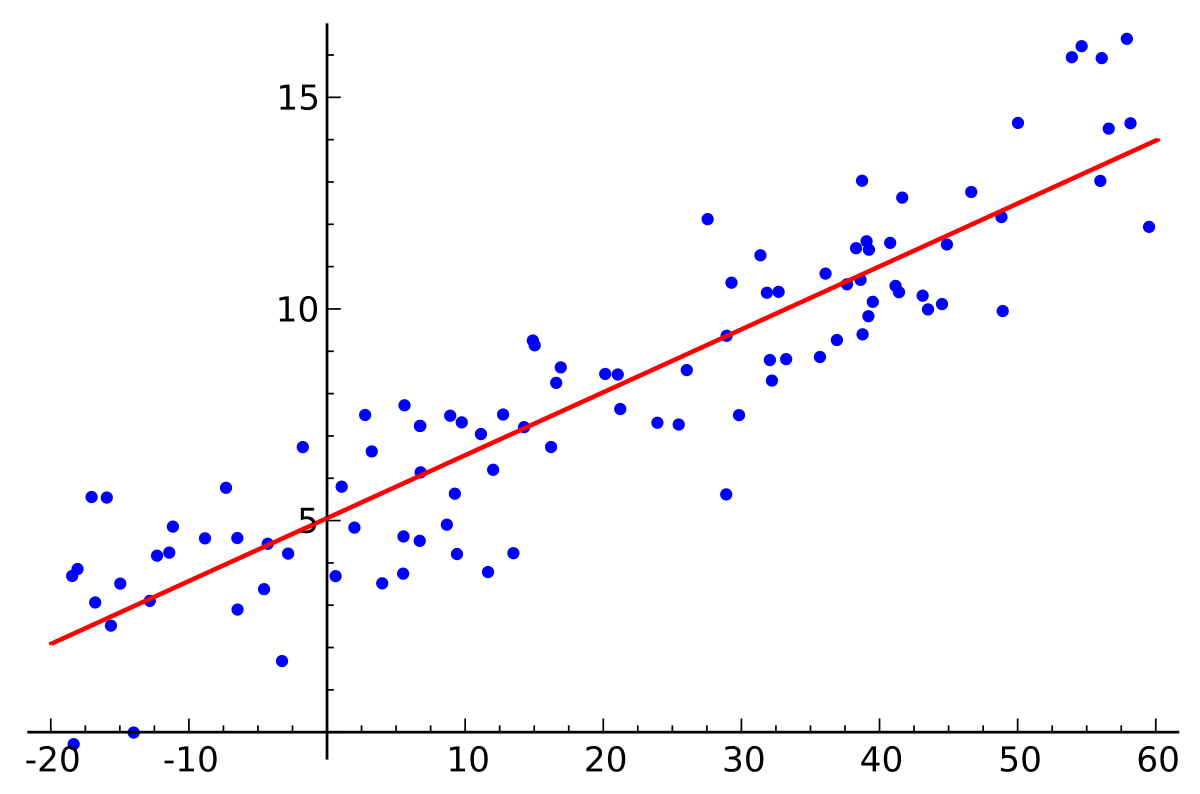
**Task 1**

1. The regression model formula which looks like below:



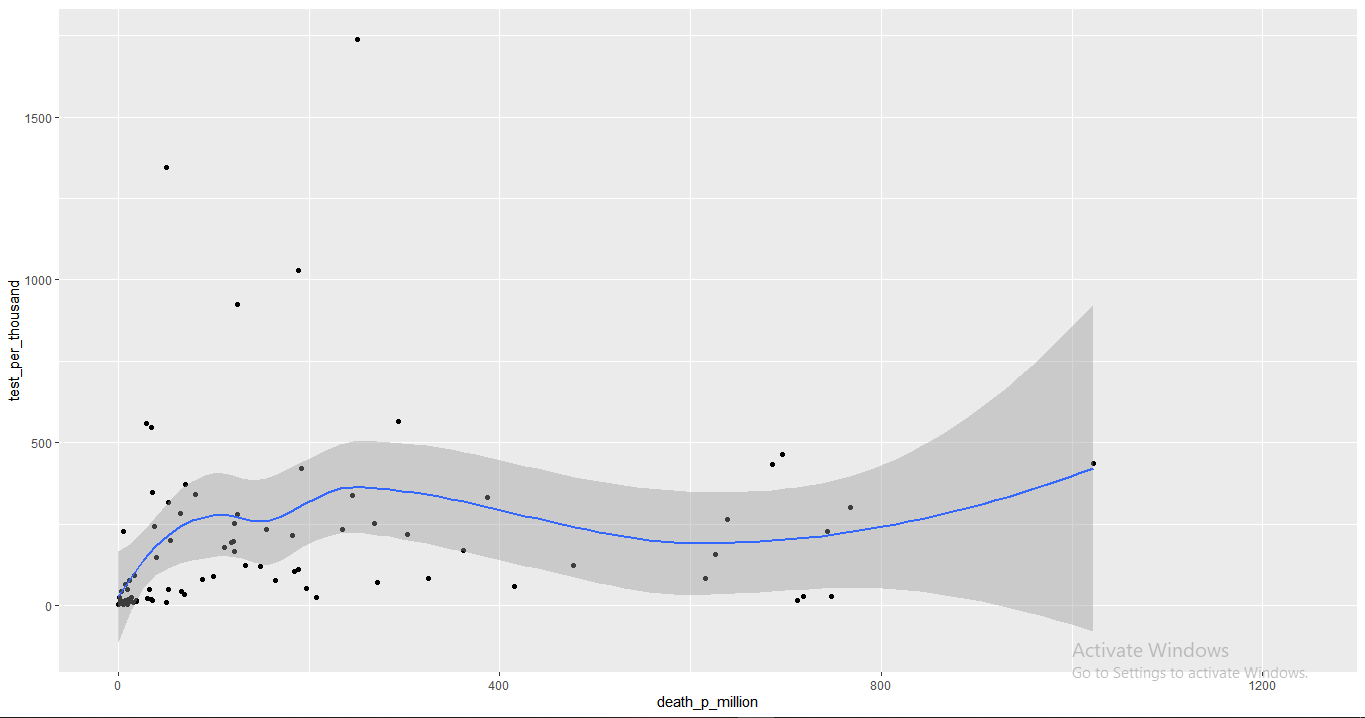
However, before doing so, conduct preliminary analysis on the variables to be included.

Answer: The linear regression model looks like below,

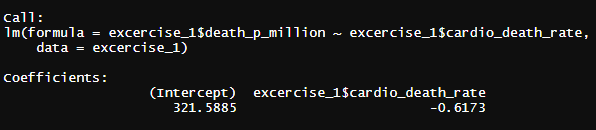


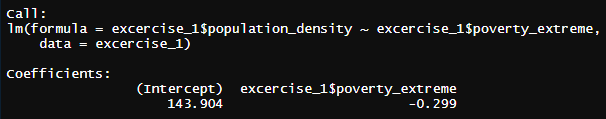
The dots in the model are the elements of the dataset which tells us the accuracy of the dataset and the line tells the ideal condition where the dots on the line tells the elements of the accuracy and those dots on the line are closed to ideal condition while the dots which are at the farthest distance are the least accurate.

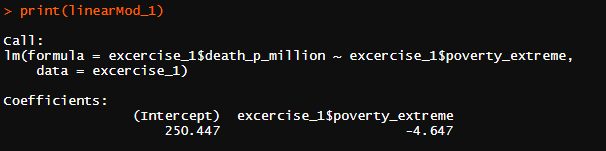
The model looks like below:

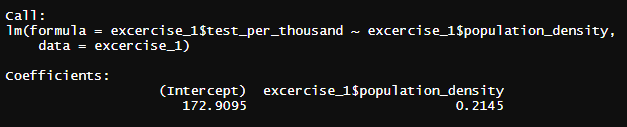


The output of the regression model looks like below:

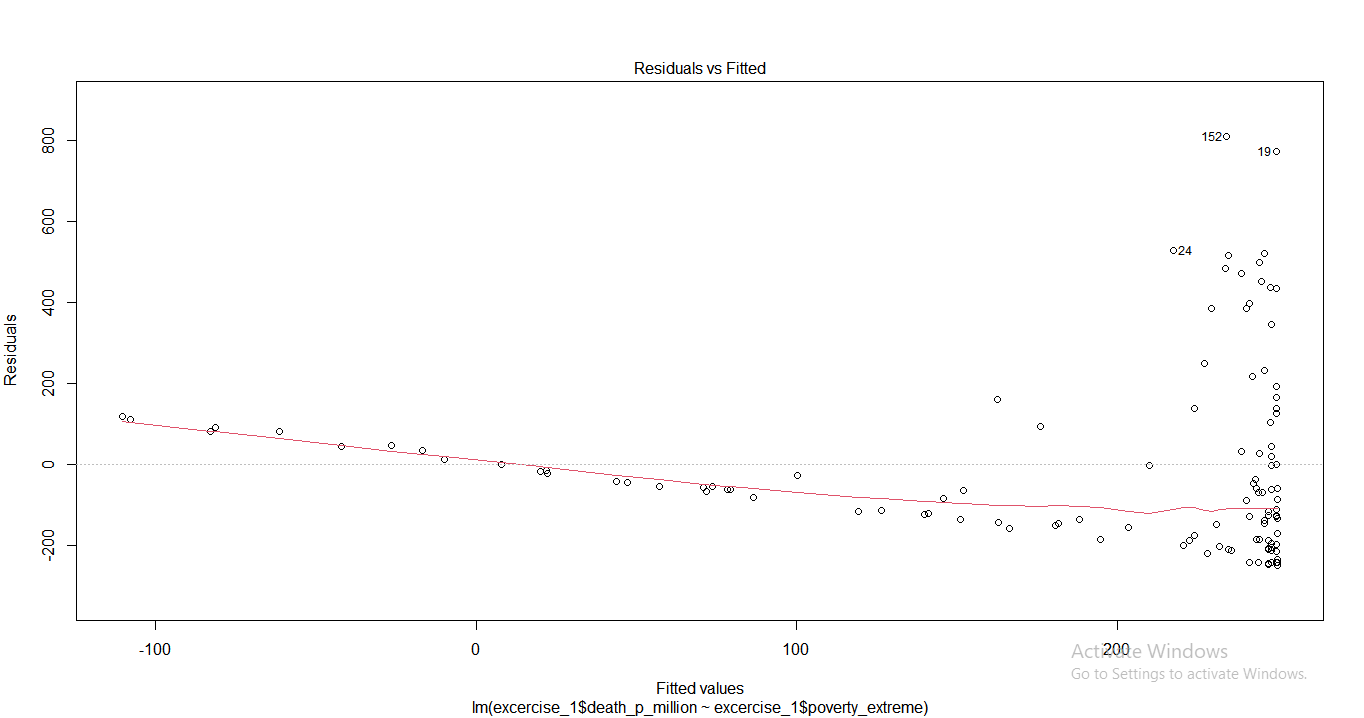


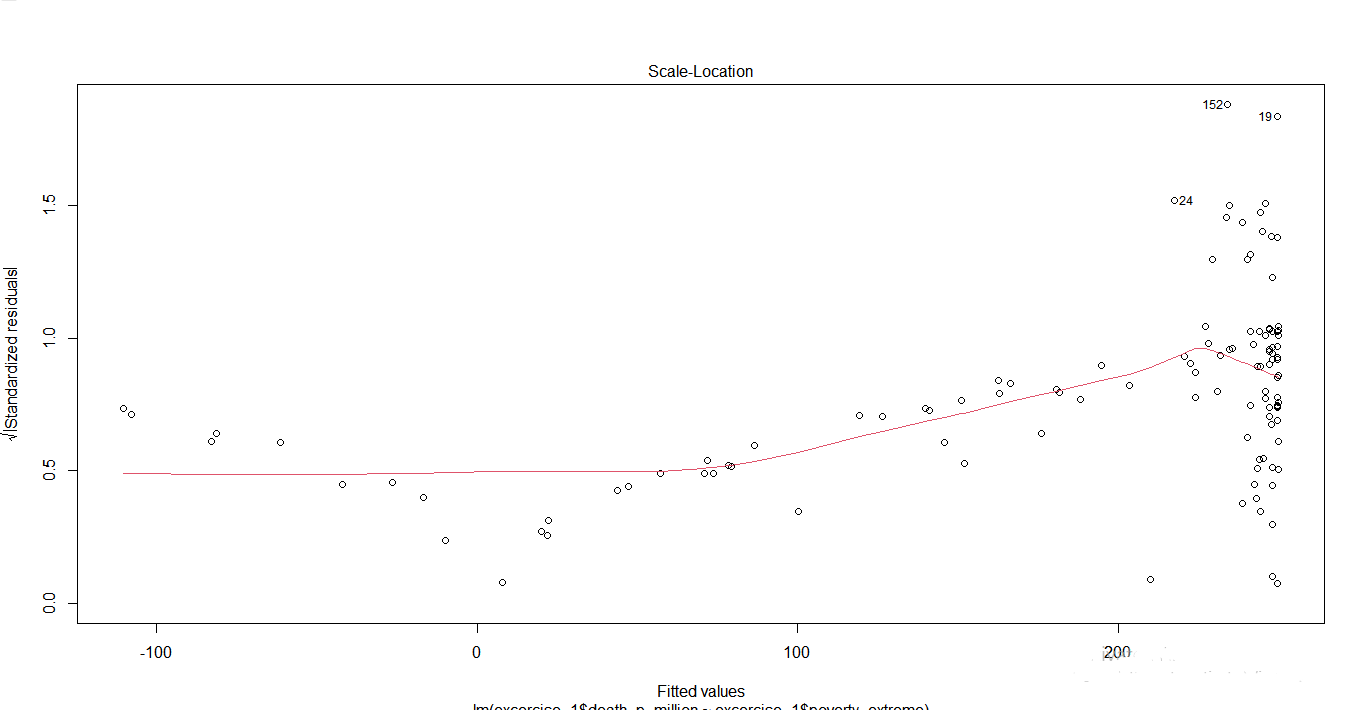






The plot of the regression model looks like below:

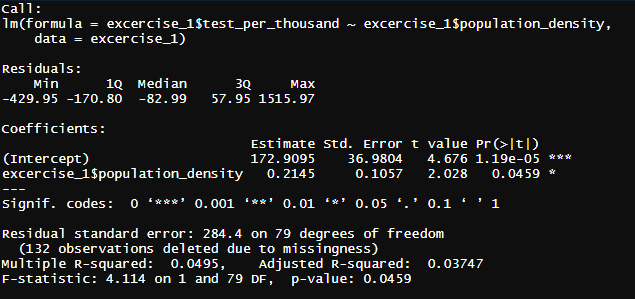


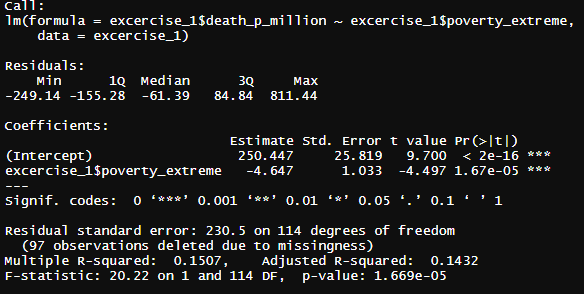


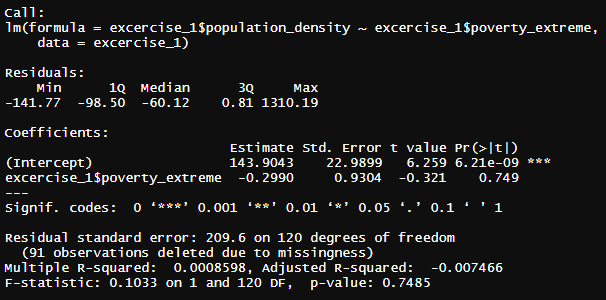
2. Comment on the overall validity of the model. Are there any caveats to note when it comes to the sample?

Answer: To check the validity of the model we will run the code which will run the summary of the linear model and tell us whether the model is valid or not, by this we can tell how much valid our model is and what is the accuracy of the model.

The output of the summary will look like below;





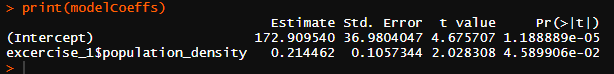


As we can see in the summary we can say that there are no caveats in the model and the p values and adjusted R squared are definite values and not in the fuzzy range. In the summary we can see the model which is valid for the population density against the extremity of poverty of the country with values of the minimum and maximum residuals, mean and median with their standard errors.

1. Comment on the variables’ coefficients and uncover which of them seem to matter -in statistical terms-for the death rate across countries and quantify their effects. Discuss the meaning of at least one. Can you convey more confidence in your estimations? Show how

Answer: In the linear regression model we can tell there are no fuzziness in the results and the variables’ coefficients about the death rate across countries and their effects across the countries to check the strength of the variables’ coefficients.

The model coefficients give us following values about the death rate across the countries;



In the above output we can see the coefficients with the estimate and its standard errors with t-value.

The estimation of the model also known beta estimations will look like below and it will tell us more confidently about what is the death rate across the countries against the poverty extreme,



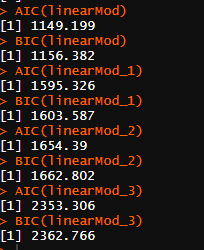
The standard error of the model and the estimations will be below and it will tell us the possible error which can happen;



1. What is the ‘goodness of fit? Comment on the goodness of fit of this model and explain its conceptual meaning. Theoretically speaking, state at least 2 ways that you could improve it on this occasion.

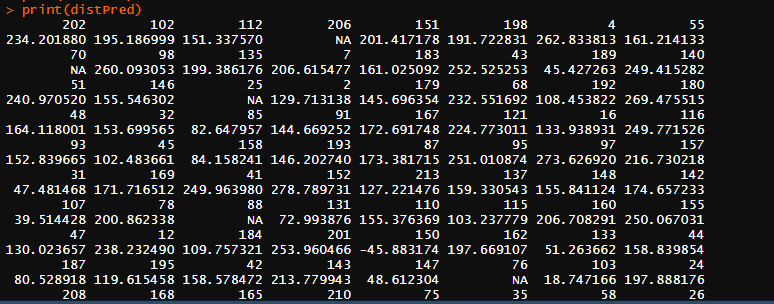
Answer: The goodness of fit test is a statistical hypothesis test to see how well sample data fit a distribution from a population with a normal distribution. Put differently, this test shows if your sample data represents the data you would expect to find in the actual population or if it is somehow skewed. Goodness-of-fit establishes the discrepancy between the observed values and those that would be expected of the model in a normal distribution case.

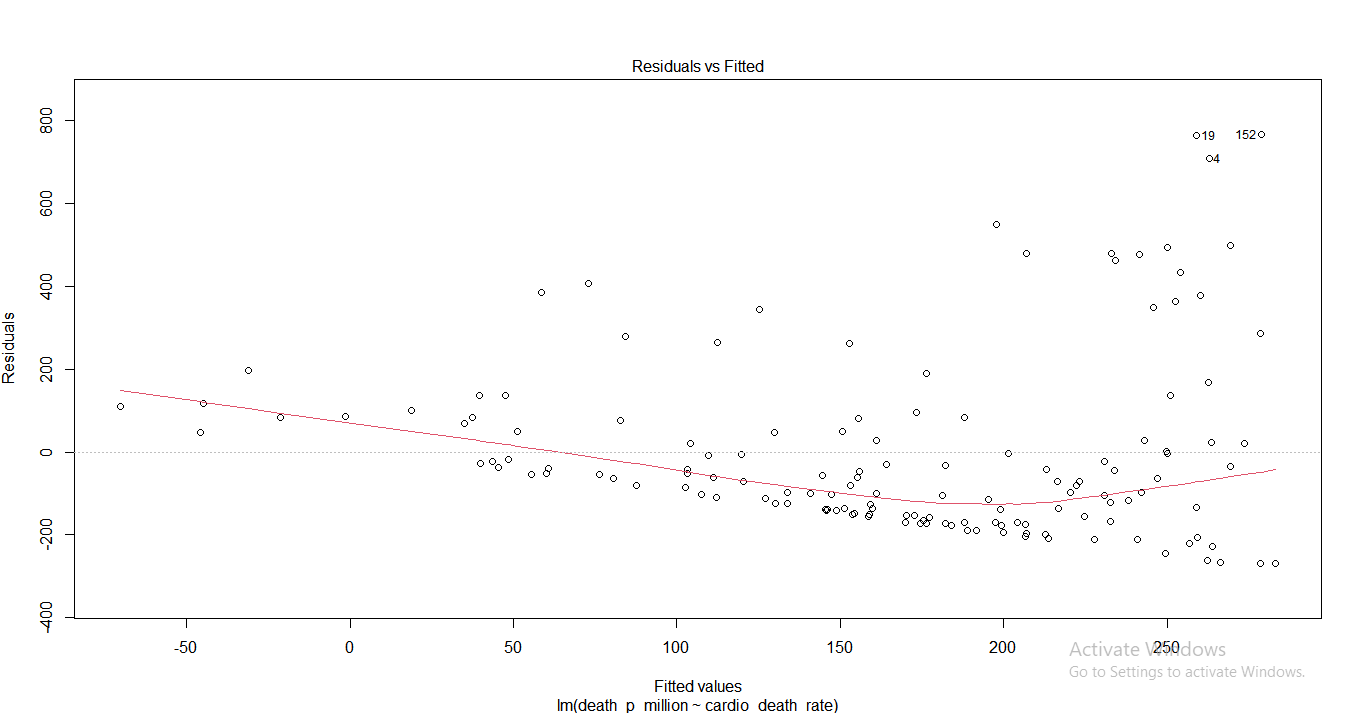
To check the goodness of fit model we will run the AIC and BIC values in the model and we will the predictive values and statistical hypothesis model and plot in the IDE and the AIC, BIC values and predictive value and linear model values plot look like below;



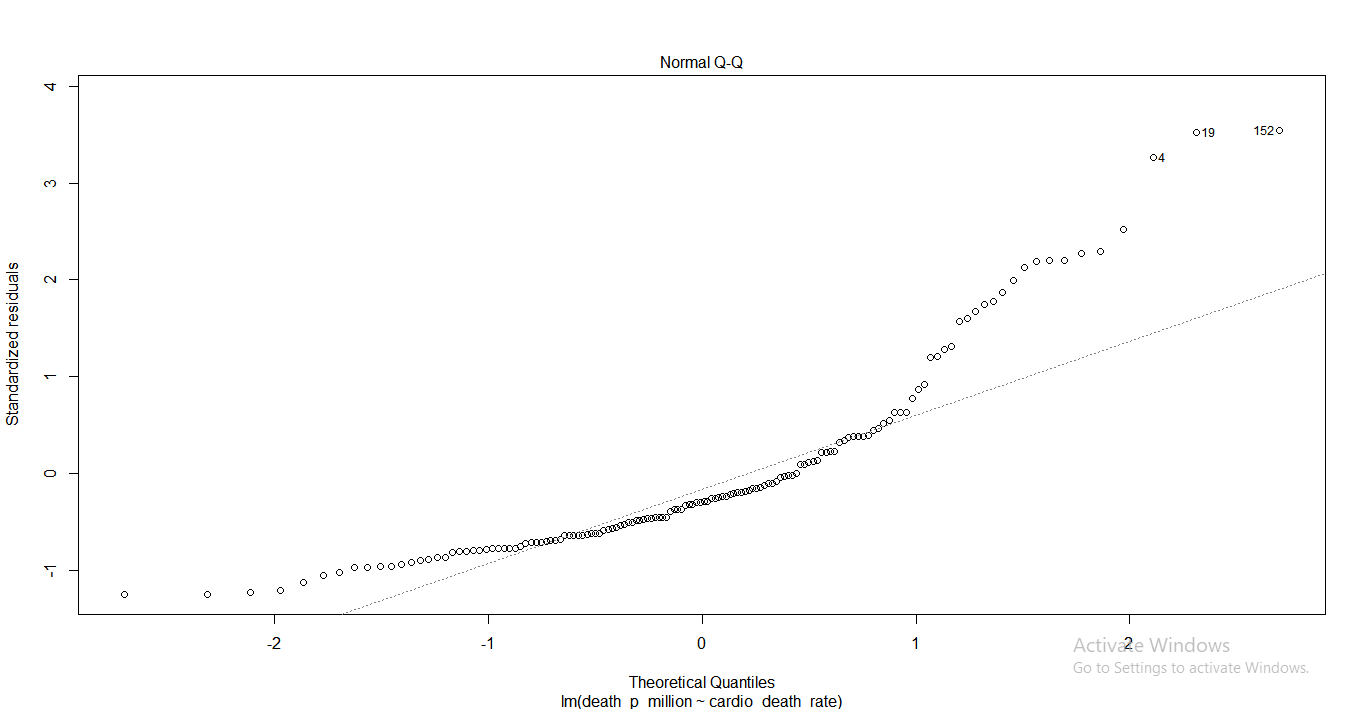
The predictive values plots look like below which will tell us residual values and other values etc.

**The prediction matrix:**





The plot of the death rate with its quantiles look like below;



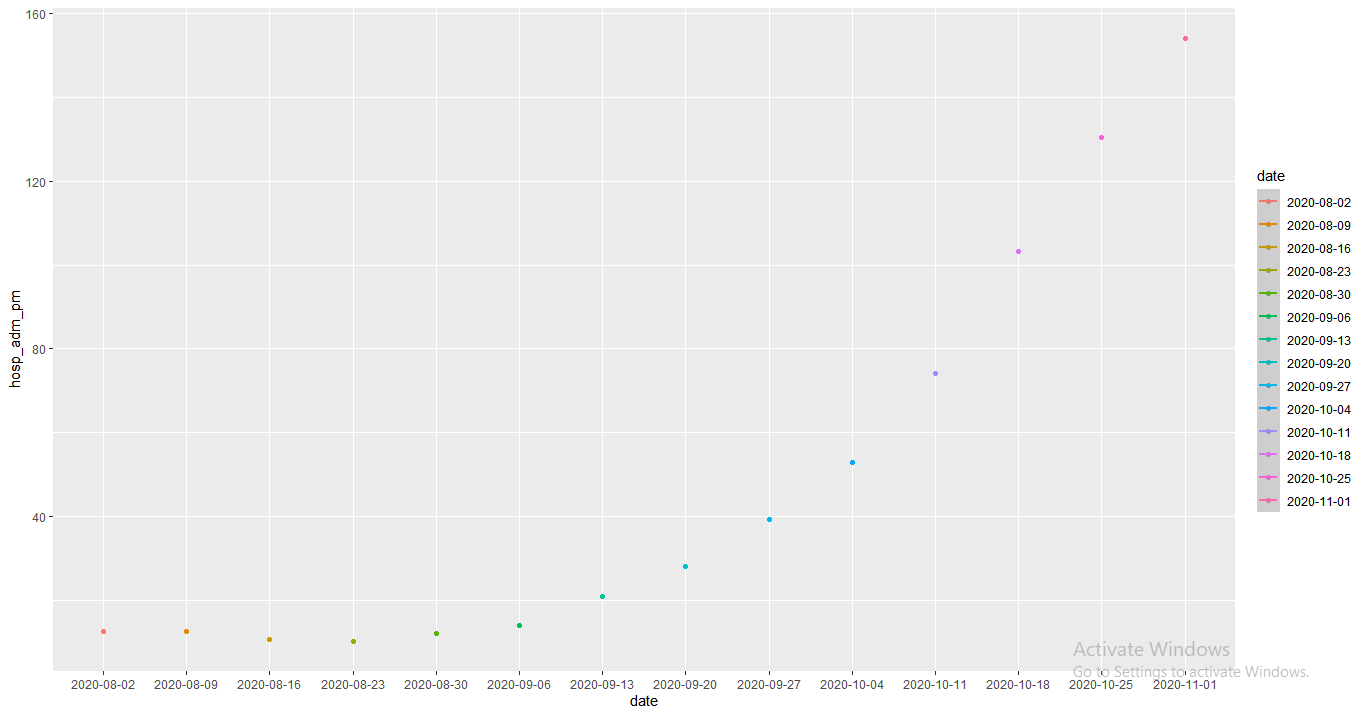
Hence this is how we can check the goodness of the fit model and test the statistical hypothesis of the death rate and poverty extreme across the country.

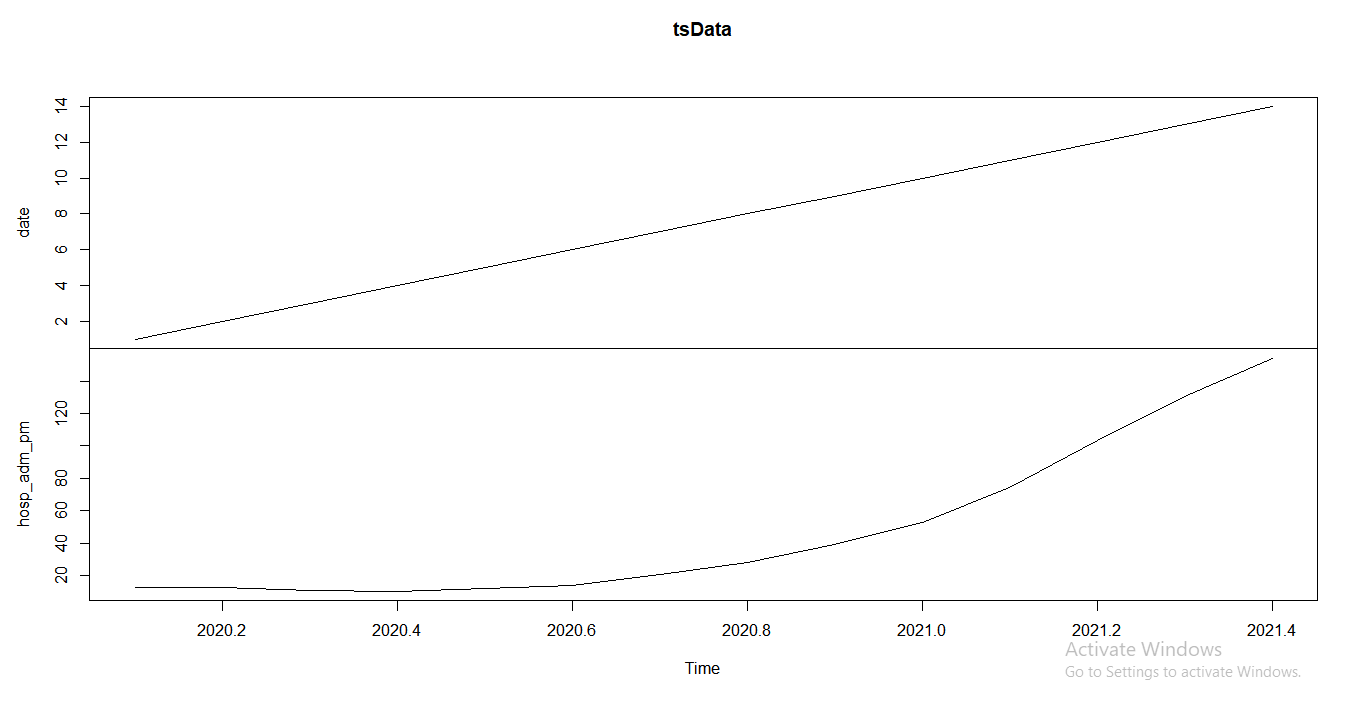
**Task 2**

1. Provide a plot of the time-series and comment on it.

Answer: To create the time series plot we will import the library ggplot2 in the ide and then we can find the time series plot according to the dates and hospitals per adm.

The time series plot will look like below;

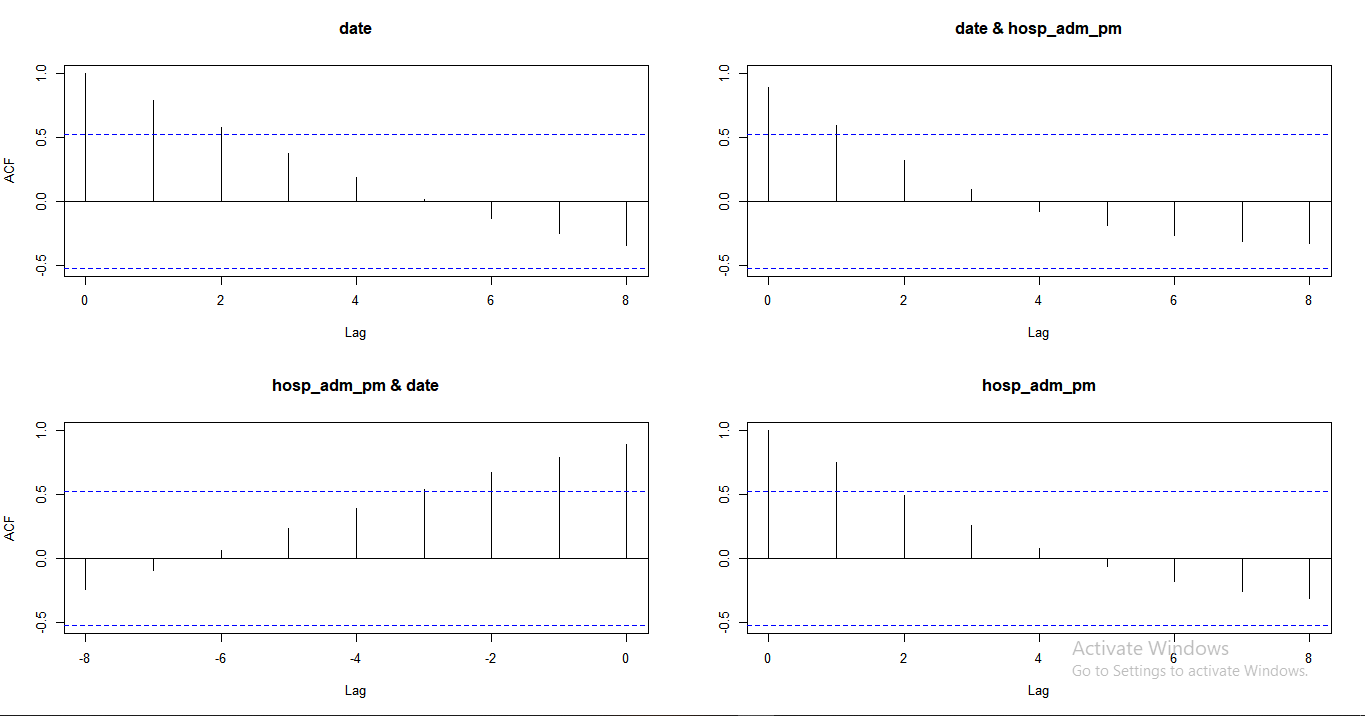




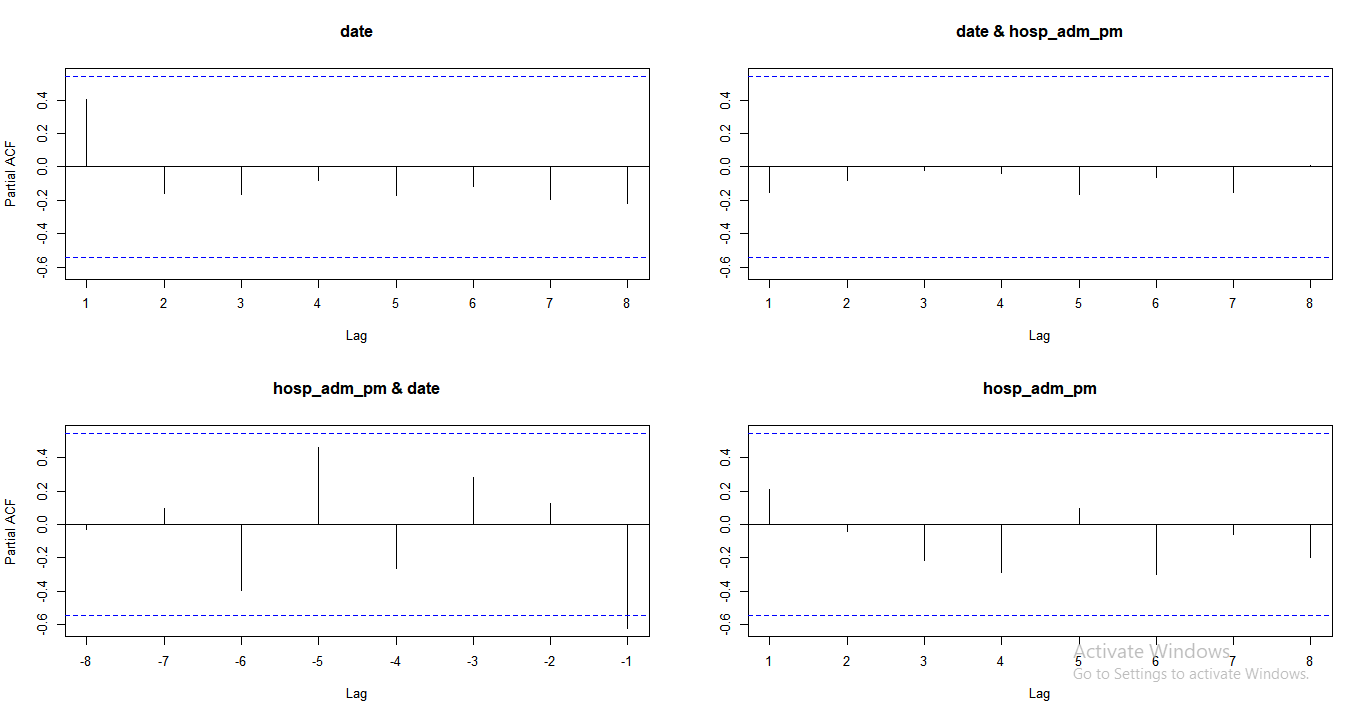
The tsData(i.e. Time series data) with the help of the package called forecast tells us the values and prediction till 2021.

1. The data set contains figures up to 11th November. Provide forecasts for the next 10days based on the Auto-Regressive Integrated Moving Average (ARIMA) model. Show the forecast output graphically and numerically, along with the 95% confidence intervals. What is the point estimate prediction on the 22nd November (10thday forecast)? Any caveats to note with this prediction horizon?

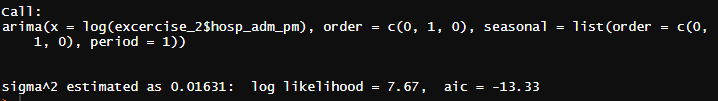
Answer: The data set contains the dates till 11th November and to provide forecast using ARIMA model we will import the library called “Forecast” which predicts the values of the fitted model and its accuracy till the coming date with the p-value of the 95.



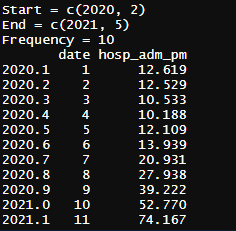
The above plot tells us the acf values and ARIMA forcast with the 95% of confidence and the Partial acf plot will look like this;



The fitted ARIMA forecast output will look like below;

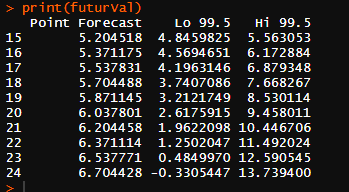


The time series data of the next 10 days will give the following output it will look like below;



1. Fit another ARIMA model and forecast the weekly hospital admissions across the UK in the next 5 weeks (NB: the data set is expressed in weekly figures). What is the point estimate prediction?

Answer: The another ARIMA model fitted in the dataset tells the predicted values with the 99.5 % of the prediction accuracy, after fitting another arima model the output looks like below;

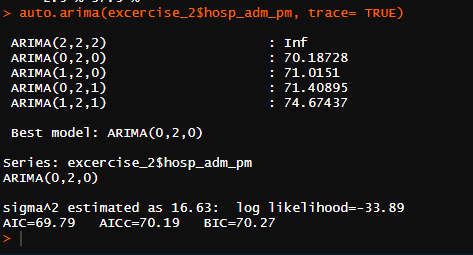


1. What is the typical error of your prediction? Discuss its meaning.

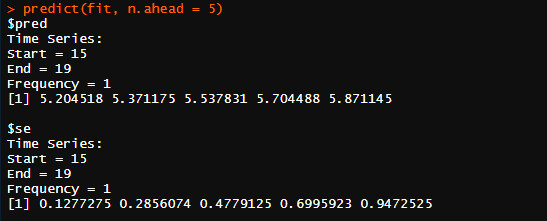
Answer: The **standard error** of the estimate is a measure of the accuracy of **predictions**. The regression line is the line that minimizes the sum of squared deviations of **prediction** (also called the sum of squares **error**), and the **standard error** of the estimate is the square root of the average squared deviation.

The error of prediction in the dataset will tell us the low and high prediction accuracy for the next 10 days and tells us the accuracy for the upcoming cases.



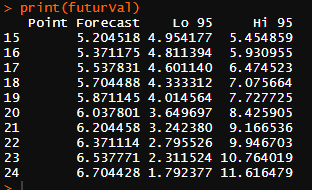


The prediction accuracy for the given dataset will give us following output;



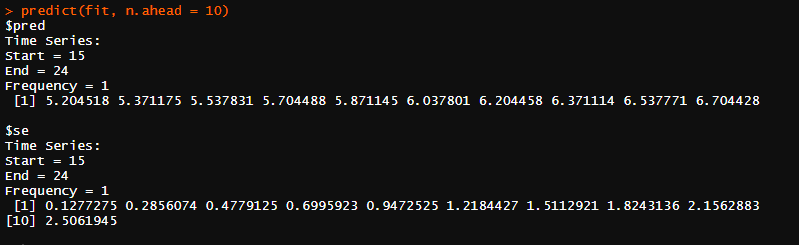
1. Name 2 ways that could significantly change your predictions either upwards or downwards?

Answer: There are two ways we can significantly change our which will tell us the prediction about the dataset for the next weeks and will tell us whether how much accurate will prediction be using top down and bottom up prediction using forecast library and ARIMA model.



The upwards and downwards predictions for the next 10 days will look like below, with the 95 percent accuracy;

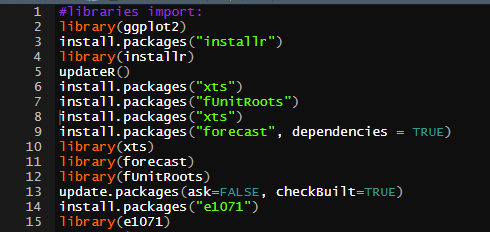
Here we can clearly see the p-value is taken 0.95 which means the values which are predicted or forecasted are 95% accurate by the method of upward and downward forecasting using ARIMA model and hence it tell the future values of the next 10 days.



Hence this is how we will tell the prediction of the next 10 days of the data using upwards and downwards forecasting algorithms, thus we will tell the current trends of the sick people in the UK and how much will they take.

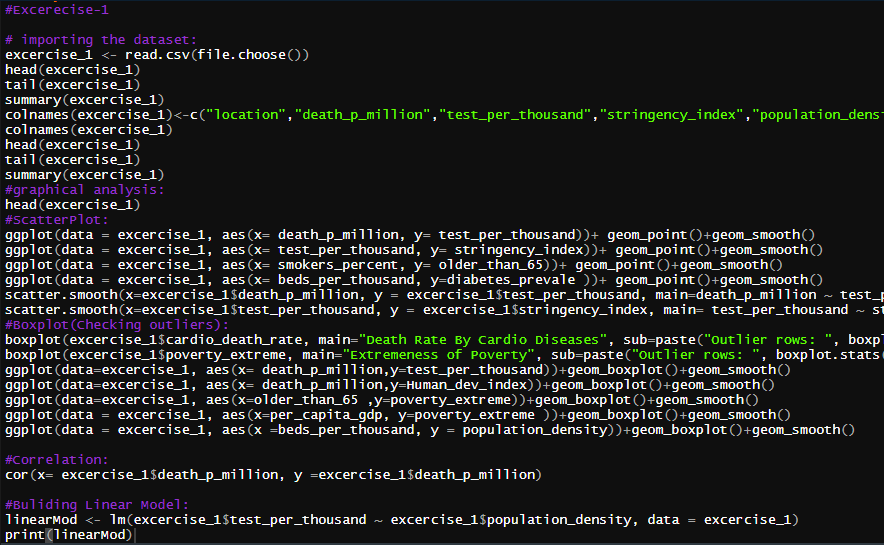
**Appendix**

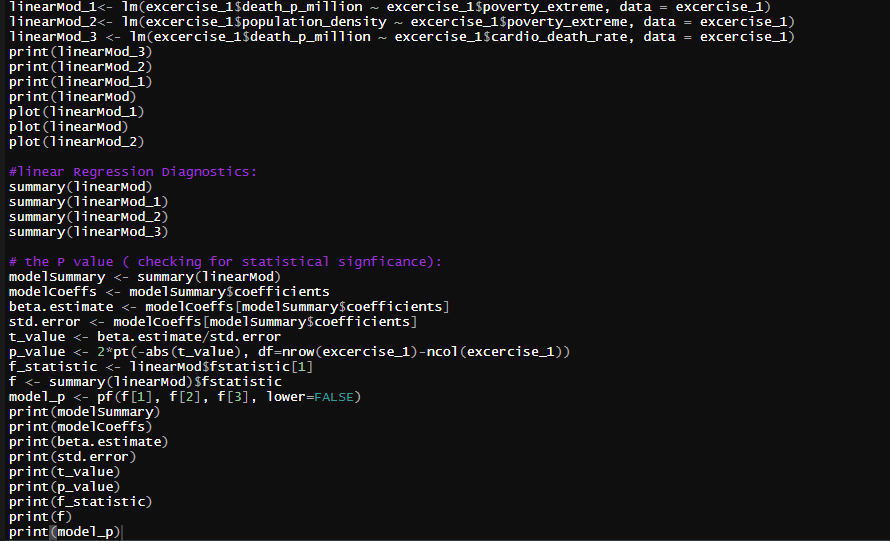
**Libraries and Packages used for forecasting**

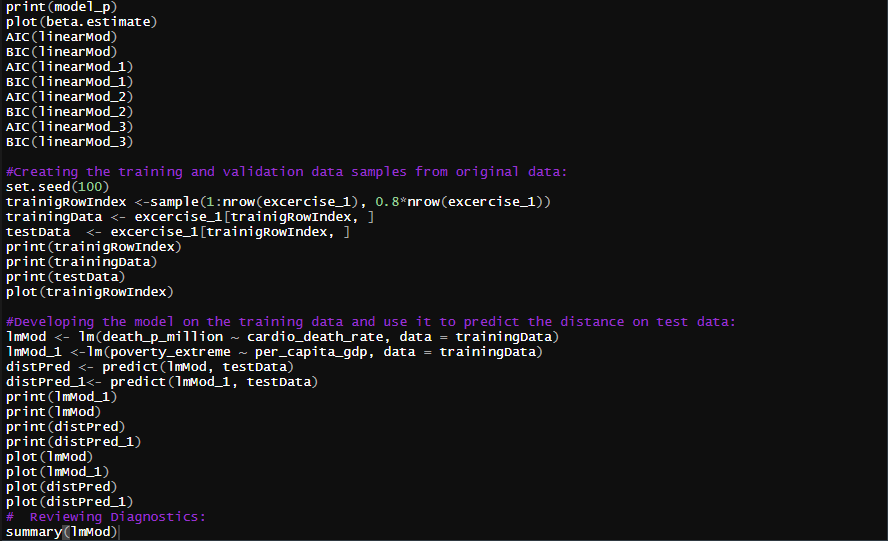
****

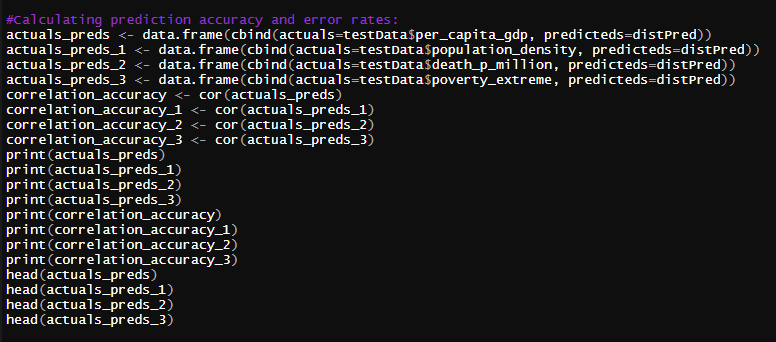
**Task-1**

For the implementation of the task 1 we will import the dataset exercise-1.xlsx and run the operations further like testing the model fitting the regression model etc.









**Task-2**

For the implementation of the task 2 we will implement the operations of forecasting by importing the package called “Forecast” with the help of Auto-Regressive Integrated Moving Average (ARIMA) model.

For this operation we will run the following codes;