**PORTFOLIO-5**

**Time series methods**

# Introduction

In the following portfolio we have been provided with Covid cases data from 1st January 2020 to 14th June 2020 and we are asked to predict the number of covid cases from 15-21 june 2020. We will be appropriate ARIMA model and estimate the model describing number of Covid19 cases per day in SPSS Statistics. The method we will be using for our analysis is Time-series method

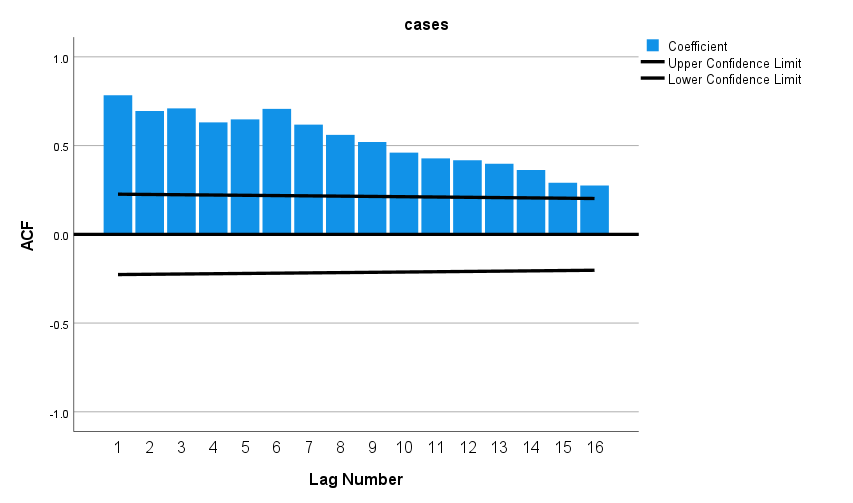
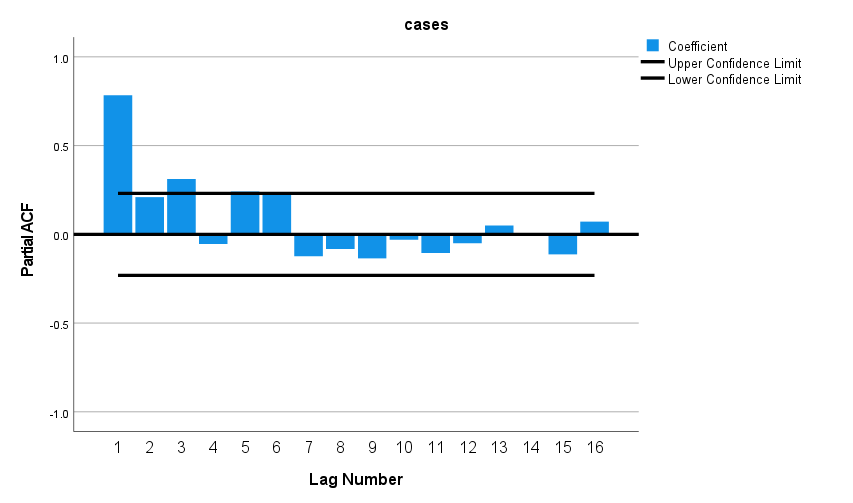
# Analysis

From the excel file provided to us, we will be estimating whether our time series is stationary or non-stationary.

As there is not a significant change here, so we are taking our data from 1st April 2020 to get a better understanding and results.

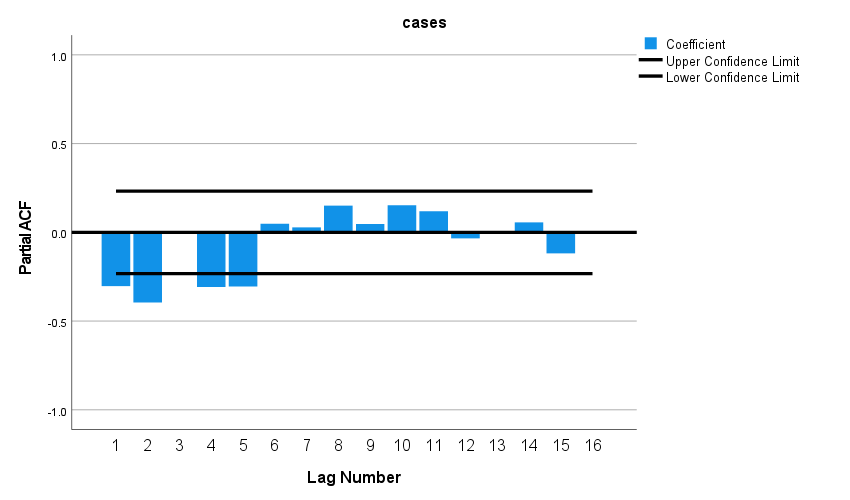
From the graph above, We can see that the mean of the data over time is not constant therefore, our time series is Non-stationary. Also, the presence of trends, and seasonality makes a series non-stationary.

We can also prove that the time series data is non-stationary by applying the autocorrelation method.



As we can In the ACF plot, all spikes are significant and there is a gradual decay in the spikes for the ACF plot. Also, if we look at the PACF plot the first spike is very significant (coefficient ~ 1) and the next lags are almost all 0. Hence, from both ACF and PACF plots, we can conclude that our time series is non-stationary.

Now, as our time series is non-stationary we need to difference the data to make our non-stationary time series into a Stationary time series.



Here, after considering difference of 1 we are see from the plot that in PACF lag 1,2, 4 and 5 are lags are significant and other are almost zero. This data is showing us the behaviour of ARIMA model and the ACF plot is showing us the number of terms in moving average(MA) part therefore (q=6) and PACF is providing us with autoregressive term (p=5). Therefore we know that for ARIMA model we require,

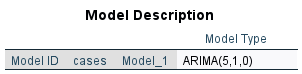
ARIMA (p,q,d) model

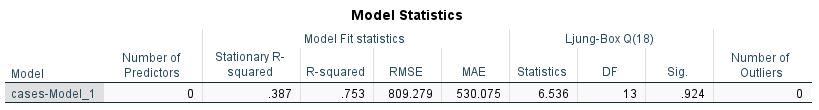
p = number of non-zero points of PACF

q = number of non-zero points of the ACF

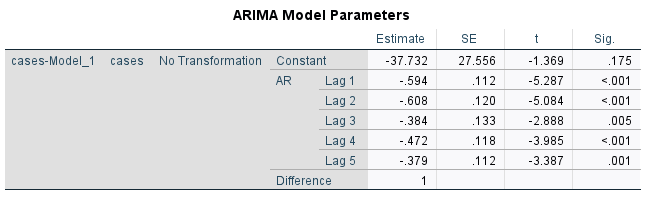
d = number of differencing

We will be considering an ARIMA model of (5,1,0)

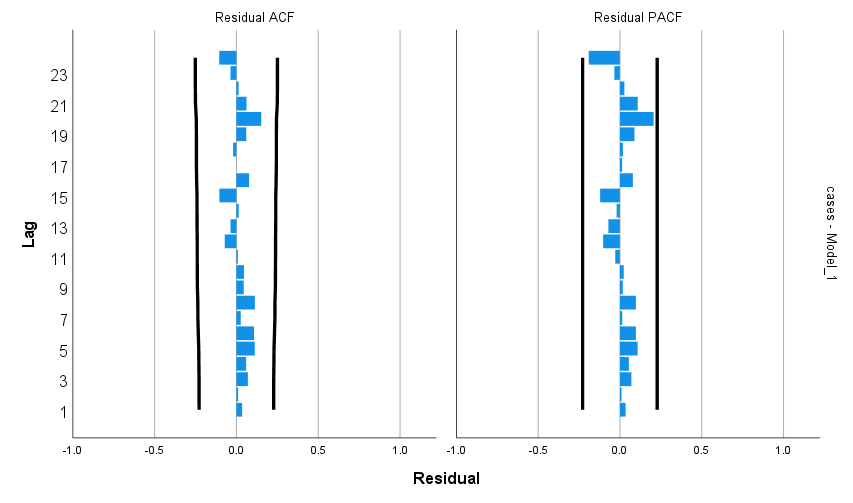


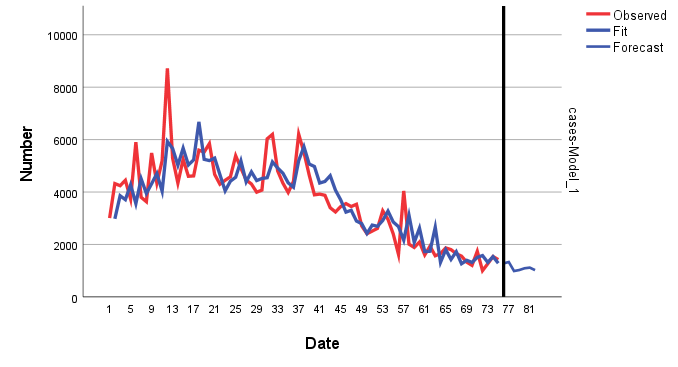
Here our significance value for Ljung-Box Q values is 0.924 (>5%) therefore we can conclude that our model is adequate model.

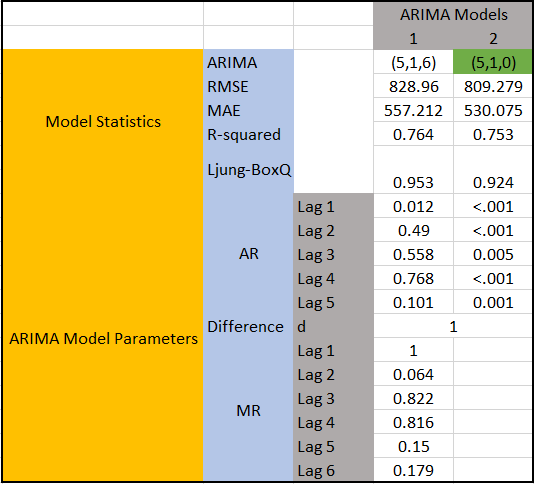
Now, Looking at our Lags for AR we can see that all the lags are significant which shows that our ARIMA model is good.



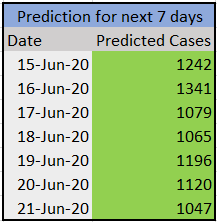
Looking at the residual we can see that all lags are within the significance interval for ACF and PACF.



 The graph above shows us the observed values against the fit values and we can see that the graphs almost coincide which shows that our prediction which is shown by the blue graph is very much close to the observed values for the Covid19 cases.

From the ACF and PACF model above we considered our first ARIMA model as (5,1,6) but we see that the AR and MR Lags are of very high values. Therefore, the most parsimonious model is (5,1,0) as the model has fewer errors and all the Lags are significant.

# Conclusion

From the findings above we get the predicted Covid19 cases for the next 7 days from 15-Jun-2020 to 21-Jun-2020.