Name: Sanika Hande

PRN: 22060641035

TSA Minitab Assignment

Define:

The data given is of covid cases and death. The question here asks for the appropriate smoothing technique for data. Analysis of ACF and PACF plots to find the appropriate order for the Arima model and transformations i.e., order of differencing to fit the Arima model.

The data must first be de-cumulated because it is cumulative, and after it has been done, some of the death numbers are negative, which is not feasible. This necessitates data manipulation before further analysis can be carried out.

Analysis:

- Tools Used:
  Here ACF and PACF tools are used to plot the graph of the autocorrelation function and partial autocorrelation function respectively. Whereas Descriptive Statistics and the graphical summary tool is used to perform EDA. ARIMA function is used to fit the time series model. Trend analysis is also used to check which type of curve fits the model. Single exponential and Double exponential function is also used to check with the technique smoothens the data more.
- Overview:
  In the dataset, there are 49 observations and 2 variables. The observations are measured along a specified and equal time interval and thus the dataset is a time series dataset. The given dataset is a non-Stationary dataset so first-order differencing is used here. The plots of ACF and PACF show the number of lags based on which we decide the order of our Arima model. Three techniques of smoothing namely exponential-smoothing, holts-smoothing and holts-winter smoothing are applied to check which smoothing technique gives the best result. The transformations are used when the dataset is non-stationary and since our dataset is non-stationary, we will use transformations such as box-cox transformation. Because Non-stationarity is present differencing will be of order 1 and not of order 0 Thus, the ARIMA model will be fitted
- Software used:
  Minitab software is used to perform all the analyses and generate the output.
- Outputs (Tables/Charts/Summary):

1)Descriptive Statistics:

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
cummulative cases	48	0	1387	161	1116	66	506	1156	1852	4213

Fig 1

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Cumulative deaths	48	0	45.96	5.71	39.58	1.00	13.75	39.00	70.00	195.00

Fig 2

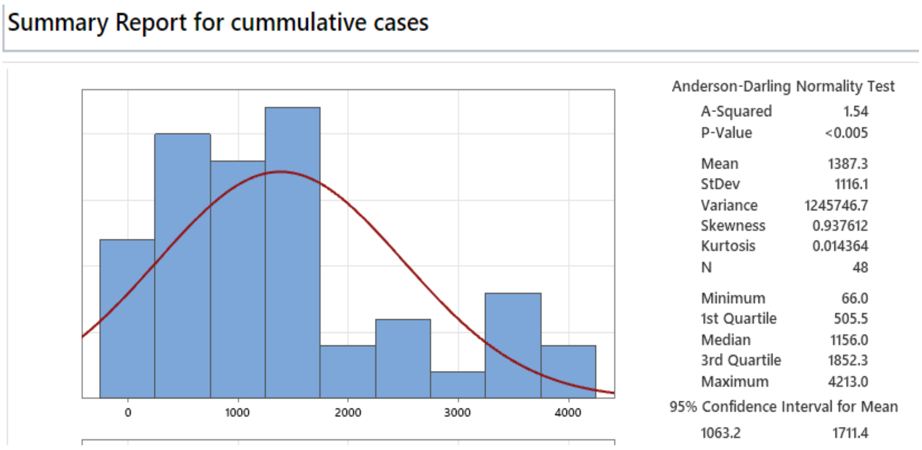

Fig 3

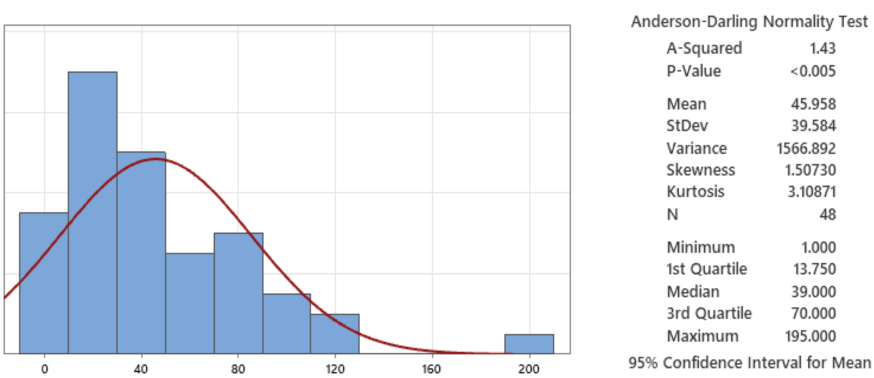

Fig 4

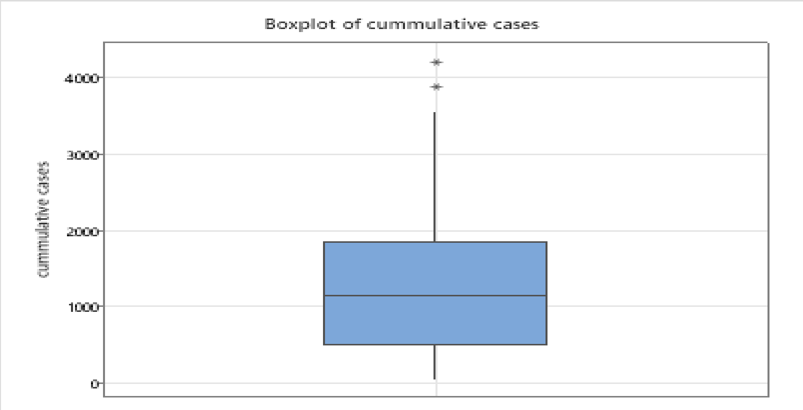

Fig 5

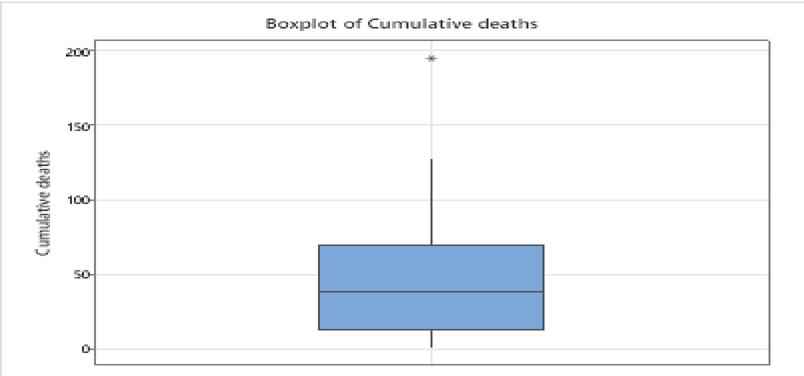

Fig 6

2) Trend analysis:

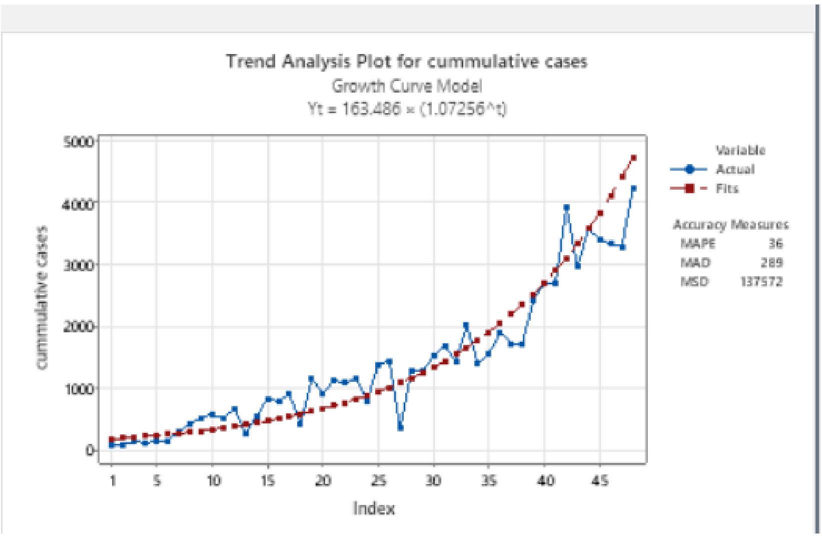

Fig 7

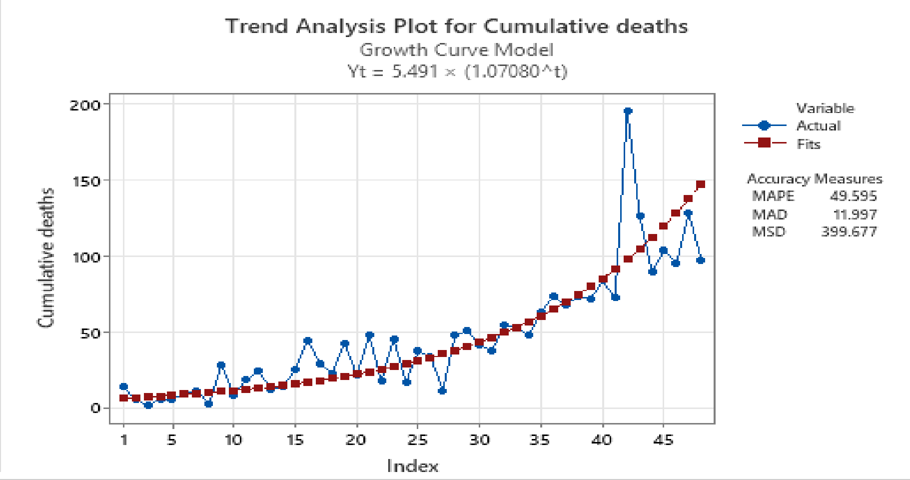

Fig 8

3) Exponential smoothing:

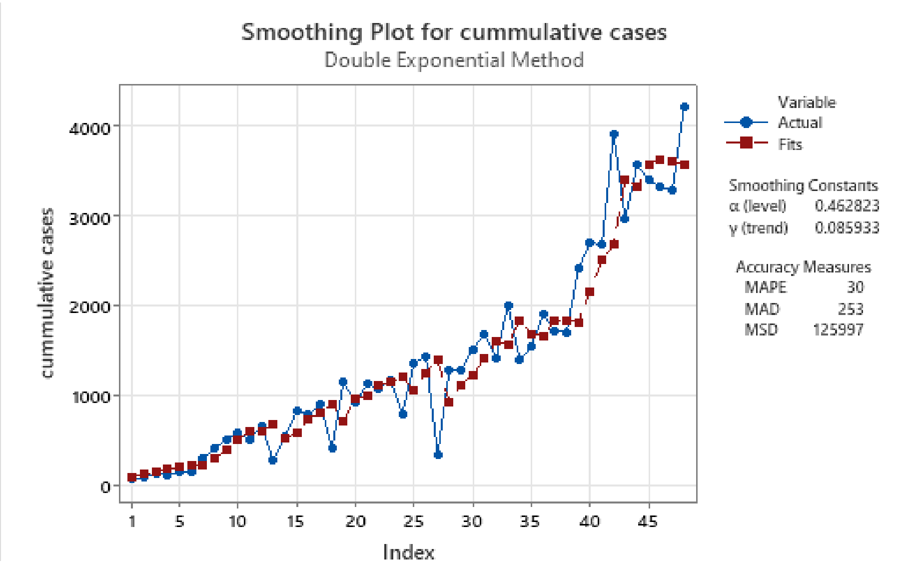

Fig 9

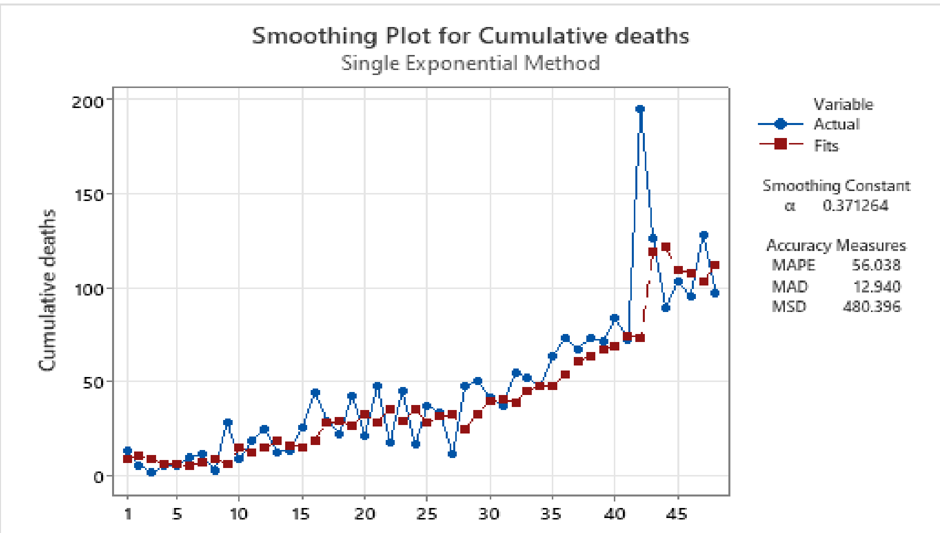

Fig 10

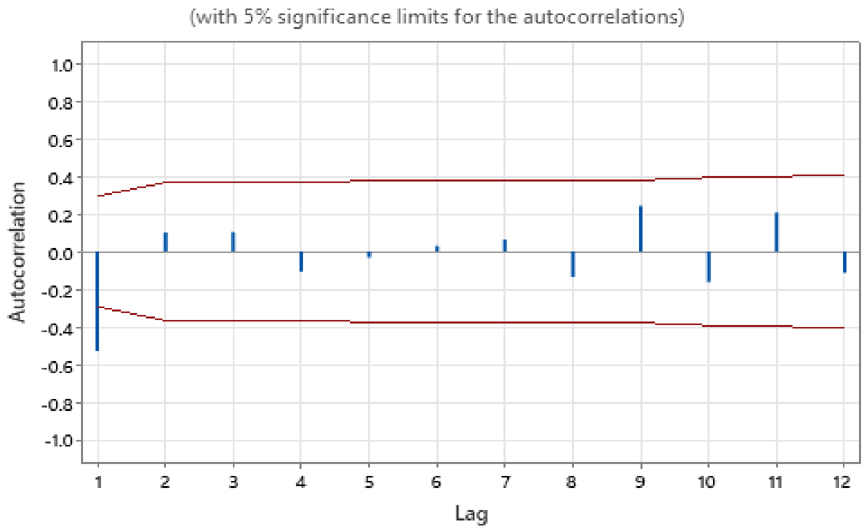

Fig 11

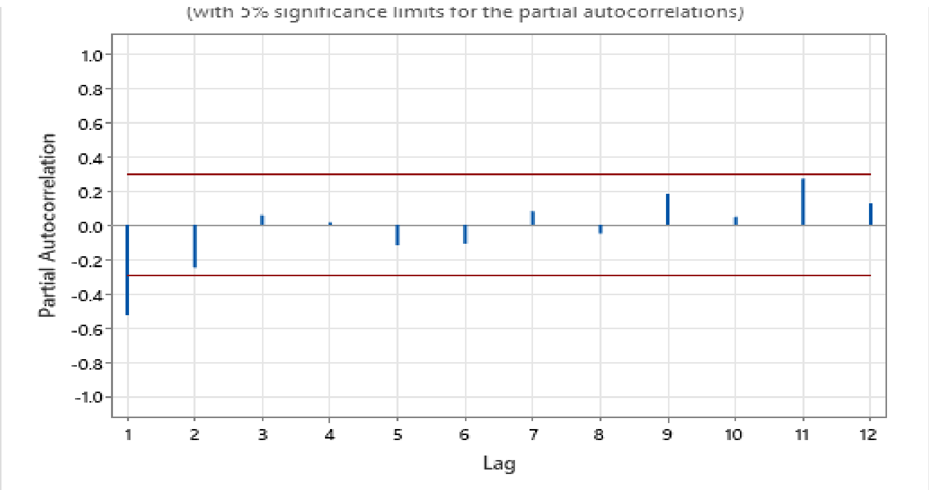

Fig 12

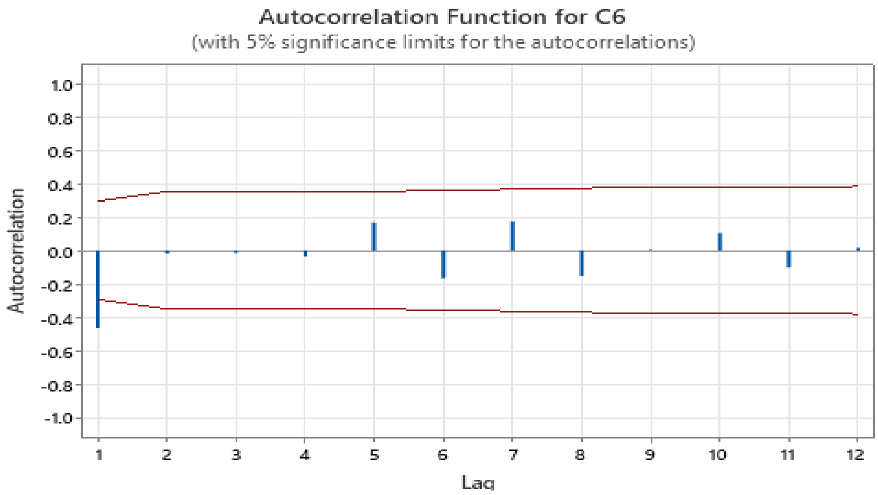

Fig 13

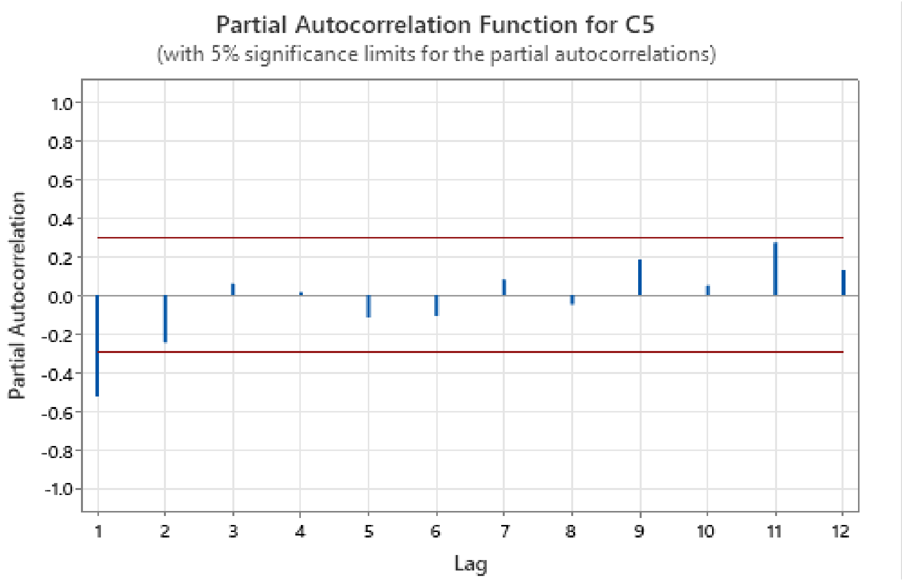

Fig 14

ARIMA Model: cummulative cases

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
MA 1	0.606	0.123	4.94	0.000
Constant	80.7	21.1	3.82	0.000

Differencing: 1 regular difference

Number of observations: Original series 48, after differencing 47

Residual Sums of Squares

DF	SS	MS
45	5898604	131080

Back forecasts excluded

Fig 15

ARIMA Model: Cumulative deaths

Type	Coef	SE Coef	T-Value	P-Value
MA 1	0.7883	0.0912	8.65	0.000
Constant	2.379	0.688	3.46	0.001

Differencing: 1 regular difference

Number of observations: Original series 48, after differencing 47

Residual Sums of Squares

DF	SS	MS
45	20405.7	453.460

Back forecasts excluded

Fig 16

- Exploratory Data Analysis:

  The descriptive statistics function and graphical summary function are used to perform EDA. Similarly, the first row from the dataset is removed as when the data is decumulated there is a loss of information. Also, the data is decumulated by subtracting the succeeding observations from the preceding observations.

  1)Covid- cases-

  The mean number of covid cases for the given period is 1387. The minimum number of covid cases observed is 66 and the maximum number is 4213. Covid cases below 506 lie in the first quartile and cases between 1200 to 1852 lie in the third quartile. Data for the covid cases is positively skewed as the measure of skewness is 0.9. Covid cases above 3600 imply outliers. The data is platykurtic for cases.

  2)Covid-Deaths-

  The mean number of covid deaths for the given period is 45.958. The minimum number of deaths is 1 which is observed on 27/03/2020 and the maximum number of deaths due to covid is 195 which are observed on 5/5/2020. Number of deaths below 14 lies in the first quartile while the deaths between 45 to 70 lie in the third quartile. Outlier is present as the maximum number of deaths i.e., 95 is considered as an outlier. The data is mesokurtic for deaths.

Interpretation:

Fig 1 and Fig 2 show the mean, median, minimum, and maximum values for cumulative cases and cumulative values. Whereas, fig 3 and Fig 4 show that our data is positively skewed. Also, Fig 5 and Fig 6 show the presence of outliers as few values lie outside the interquartile range.

Fig 7 and Fig 8 show the trend analysis for covid cases and covid deaths respectively. Trend analysis shows that the MAPE of covid cases is 64 for the linear growth curve, 44 for the quadratic curve, and 36 for the exponential curve. Since the exponential curves give the minimum APE it is the best-fitted trend for covid cases. Similarly, Covid deaths MAPE is 69.753 for the linear trend curve, 49.595 For the exponential curve, and 63.156 for the quadratic curve. Here also the exponential curve gives the minimum APE therefore exponential curve best fits the covid deaths.

Fig 9 and Fig 10 shows which smoothing techniques gives the best result. For covid cases, the double smoothing technique is the best suited as the value of alpha for double smoothing is 0.4628 and for the single smoothing technique is 0.5685. Since the value of alpha is minimum for the double smoothing technique the best smoothing technique for covid cases is double smoothing. Fig 10 shows that the single smoothing technique is best suited as the value of alpha for the single smoothing technique is 0.371 and for double smoothing it is 0.6744. Since alpha is minimum for single smoothing it is best suited for covid deaths.

Fig 11 and 12 shows the ACF and PACF plot for covid cases it is seen that there is a lag of only one for both ACF and PACF and the lag of ACF gives the order of the MA component whereas the lag of PACF gives the order of AR component and since the data is non-stationary the differencing is of order 1. Therefore, the ARIMA model is best fitted for ARIMA (0,1,1), ARIMA (1,1,1), and ARIMA (1,1,0) the p-values are significant for (0,1,1) and MSS is lower for ARIMA (1,1,1).

Fig 13 and Fig 14 show the ACF and PACF plot for covid deaths it is seen that here also the lag Is of one for both ACF and PACF. Thus, the ARIMA model fitted is of order (0,1,1), (1,1,1), (1,1,0) for covid deaths. ARIMA (0,1,1) shows the p-value is significant and MSS is also less for (0,1,1). Therefore ARIMA (0,1,1) is best fitted for covid deaths.

Fig 15 and 16 show the ARIMA (0,1,1) for covid cases and deaths respectively.

Conclusion:

From the trend analysis, it can be concluded that both the number of covid cases and the number of deaths follow an exponential growth curve indicating that both the numbers are expected to rise exponentially. The double exponential smoothing technique is appropriate for covid cases and single exponential smoothing for covid deaths. From the ACF and PACF graph for covid cases, it can be concluded that ARIMA (0,1,1) has more significant p-values whereas ARIMA (1,1,1) has less MSS but since the significance of p-values is of great importance we can say that ARIMA (0,1,1) is a good fit for cases and ARIMA (0,1,1) for deaths has both p-value significant and minimum MSS thus ARIMA (0,1,1) is best fit for covid deaths.