**Module Assignment**

**Time Series**



Write a simple R script to execute the following data preprocessing and statistical analysis. Where required show analytical output and interpretations.

**Preprocessing**

1. Load the file "6304 Module 7 Data.xlsx" into R. This data shows the number of visitors to the United States from the People's Republic of China on a quarterly basis from years 1998 to 2012. The data shown is scaled in thousands of people.
2. Create a new "index" variable in the data frame which will be an identifying sequential numbering of rows from 1 to the number of rows in the data frame.

**Analysis**

1. Show a plot of the data using the number of visitors as the "y" variable in the plot.
2. Using all the data parameterize a base time series simple regression model using "index" as the independent variable. Show the summary of your regression output.
3. Drawing on Analysis Part 1 above, show a properly titled plot of the time series data with the simple regression line layered on the graph in a contrasting color.
4. Execute and interpret a Durbin-Watson test on your model results.
5. Note the original data appears to have a pronounced cyclical pattern. Assuming the complete cycles are four quarters long, construct a set of seasonal indices which describe the typical annual fluctuations in visitors. Use these indices to deseasonalize the visitors data. Store this deseasonalized data in a column in the original data frame.
6. Using the deseasonalized data parameterize four different regression models. A simple regression model will be the base case to be followed by second order, third order, and fourth order polynomial models which attempt to describe the longer-term secular fluctuations in the deseasonalized data.
7. Reseasonalize the fitted values for each of the four models, storing the reseasonalized values in separate columns in the original data frame. Drawing on Analysis Part 3 above, construct a plot showing the original data and the fitted values for each of the four regression models. Show the four sets of fitted values plots in contrasting colors and title the graph appropriately.
8. Select the model which in your view is the best fit to the deseasonalized data. Give a brief justification as to why you believe your selection is the best fit.

Your deliverable will be a single MS-Word file showing 1) the R script which executes the above preprocessing and analysis instructions and 2) the results of those instructions and needed written interpretations. The first line of your script file should be a “#” comment line showing your name as it appears in Canvas. Results should be presented in the order in which they are listed here. Deliverable due time will be announced in class and on Canvas. This is an individual assignment to be completed before you leave the classroom. No collaboration of any sort is allowed on this assignment.