**QMB-6304 Analytical Methods for Business**

**Module Assignment**

**Time Series Regression**

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 Time Series Assignment 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. This will be your master (and only) data set for this assignment. **No random sampling is needed.**
2. Create a new index variable in the data frame which will be a 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 the number of visitors 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 a simple regression model. Show a plot of the original visitors variable and the fitted values from this model.
7. Using your seasonal indices and the model you built in Step 6, reseasonalize the fitted values, storing the reseasonalized values in separate columns in the original data frame. Construct a plot showing the original data with the deseasonalized and reseasonalized fitted values. Do the reseasonalized fitted values appear to have a better fit to the original data than the deseasonalized fitted values?
8. Calculate the standardized errors for the reseasonalized fitted values. Does there appear to be a pattern or any outliers in the errors?

Your deliverable will be a single MS-Word file showing 1) the R script which executes the above instructions, 2) the results of those instructions, and 3) any interpretations required. 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 and submitted by the time stated on Canvas. No collaboration of any sort is allowed on this assignment.**