**6304 Assignment 5**

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 monthly production of beer in Australia from January of 1956 to December 1978. The data shown is scaled in megaliters(1 Megaliter = 1 million liters).
2. Rename the following three columns in this fashion:

|  |  |
| --- | --- |
| **Old name** | **New name** |
| X | index |
| Month | date |
| Monthly.beer.production | production |

1. Create two new columns in the data frame accounting for the “year” and the “month” of the production, respectively. Use the method demonstrated in the lecture on this topic to extract information in the new columns from the existing “date” column.

**Analysis**

1. Show a line plot of the data using the index as ‘x’ and production as "y" variable. Show appropriate main titles and axis titles on the graph.
2. Using all the rows parameterize a base time series simple regression model using "index" as the independent variable and production as dependent variable. Show the summary of your regression output. From this state the slope of your regression line and the correlation coefficient between actual and predicted values for production.
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 12 months long, construct a set of seasonal indices which describe the typical annual fluctuations in production. Use these indices to deseasonalize the production data. Store this deseasonalized data in a column in the original data frame.
6. Using the deseasonalized to data parameterize two different regression models. A simple regression model will be the base case and a second order polynomial (x2) model which attempts to describe the non-linear secular fluctuations in the deseasonalized data.
7. Reseasonalize the fitted values for each of the two models. Drawing on Analysis Parts 1 and 3 above, construct a plot showing the original data and the fitted values for each of the two regression models. From a visual review, which model appears to have the better fit to the original beer production data?