**Module Assignment**

**Module 4**

**QMB-6304 Analytical Methods for Business**

Write a simple R script to execute the following:

**Preprocessing**

1. Load into R the data included in “Assignment 4 Data.xlsx”. This set includes data on 50 small US cities involving crime, and police expenditures. This is your master data set.
2. Using the numerical portion of your U number as a random number seed, take a random sample of 12 cases from the master data set using the method presented in class. This is your primary data set.

**Analysis**

1. With the data in your primary data set, use R to calculate and report the correlation coefficient between the two variables. Report and interpret the p values for the correlation coefficients.
2. Use your primary data set to create a scatterplot of your data. Show police funding on the x-axis and reported crimes per 100,000 on the y-axis. Make certain the x-axis is scaled between 0 and 100 and the y-axis is scaled between 300 and 2000. Apply appropriate axis labels to identify the variables.
3. With your primary data set use R to conduct a simple linear regression on the data with reported crimes per 100,000 as the dependent variable and police funding as the independent variable. As a part of this be sure to:
   1. Report the beta coefficients and associated p values and confidence intervals from your model.
   2. Give a written interpretation of your beta coefficients.
   3. Assess your model’s conformance with the LINE assumptions of regression.
   4. For a given small city spending $41 per resident on police protection use your model to predict crime rate per 100,000 residents. Include a 95% prediction interval and a written interpretation of both the prediction and the accompanying interval. If you only looked at this interval, what would it potentially indicate about model fit?
4. New York City’s budgeted police expenditures in 2020 are $10.9 billion. Its estimated population for the same year is approximately 8,550,000. Give two reasons why it would be wrong to use your model to predict New York’s crime rate per 100,000 residents.

Your deliverable will be a single MS-Word file showing 1) the R script which executes the above instructions and 2) the results of those instructions. 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. Please remember the prohibition on using screenshots in your deliverable.**