**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 “6304 Assignment 4 Data.xlsx”. This set includes data on 638 Southwest Airlines flights. The data set includes the following variables:
   1. Origin airport.
   2. Destination airport.
   3. Air miles traveled by the flight.
   4. The most common fare charged for the flight.
2. Using the numerical portion of your U number as a random number seed, take a random sample of 40 cases from this full data set using the method presented in class. This is the reduced data set you will use to execute this assignment.

**Analysis**

1. With the data in your reduced data set, use R to calculate and report the correlation coefficient between the two continuous variables. Report and interpret the p values for the correlation coefficients.
2. With your reduced data set use R to conduct a simple linear regression on the data with fare charged as the dependent variable and air miles traveled 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 flight traveling 900 air miles use your model to predict the fare that would be charged for such a flight. 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 could you conclude about model fit?
3. One of the documented longest commercial flights in the world is a Delta Airlines flight from Johannesburg, South Africa to Atlanta, Georgia. The flight's typical path traverses 8,439 air miles. Conduct a prediction for the fare of this flight using your model and explain why your prediction may or may not be valid.

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 before you leave the classroom. No collaboration of any sort is allowed on this assignment.**