**Module Assignment**

**Module 3**

**QMB-6304 Analytical Methods for Business**



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 3 Assignment Data.xlsx” into R. This file contains information on the times required for each of 46,484 instances of a vehicle listed for sale on Craig’s List in the United States.
2. Split the data by condition. You can use any of several approaches to this, including the “split” command which will create four data objects nested inside a list. The command is of the form:

*xx=split(master.data.frame,master.data.frame$split.variable)*

1. Pull out data on the condition levels “excellent”, “good”, “fair”, “like new”, and “new”. Place better classed vehicles with “excellent”, “like new”, and “new” conditions in one data frame and those lesser classed as “good” and “fair” in another. To do this use the R command below. The rbind() command may also be useful in this, though there are several ways to accomplish this step.

*condition.data.object=xx[[order.in.xx.list]]*

1. Using the numerical portion of your U number as a random number seed and the method demonstrated in class, take a random sample of 75 cases from each of the two condition data frames. Store the sampled observations from each condition group in separate data frames.

**Analysis**

1. Using your sample, construct a 90% confidence interval on the population mean odometer reading of the better classed vehicles.
2. Assuming the data on all better classed vehicles represents the population, does your 90% confidence interval include the true population mean odometer reading?
3. Use R and your reduced 75-case data set for lesser classed vehicles. Can you say (α = .05) that the population mean price of these vehicles is less than $9,000? How about less than $12,000?
4. Referencing Part 3 above, what “test against” (mu) value in a two-tailed hypothesis test would yield p = .05 in a test on the lesser vehicles’ population mean price?
5. Using R and your reduced 75-case data sets, show comparative notched boxplots of the two groups’ price variable. Your boxplots should be displayed side by side in a single graphic with an appropriate title and x-axis labels. Do these plots indicate a possible difference between the transaction times for the two facilities? Do these plots indicate a difference in skewness or number of potential outliers between the better and lesser classed vehicles?
6. Using R and your reduced 75-case data sets, does there appear to be a statistically significant difference (α = .05) between the mean odometer readings for the better and lesser classed vehicles? How about a statistically significant difference on price?

Your deliverable will be a single MS-Word file created using R Markdown. Your file will show 1) the R script which executes the above instructions and 2) the results of those instructions. The first two lines of your deliverable will state this is “Module 3 Assignment” of our course and your name as it appears in Canvas. Your code chunks and analysis 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.