**Module Assignment**

**Module 5**

**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 “Module 5 Assignment.xlsx” into R. This file contains information on 1338 instances of an adult being hospitalized somewhere in the United States. The variables included the patient’s age, body mass index (bmi), whether or not they were a smoker, and the total final charges for hospital care submitted to the patient or a third-party payer. Title this data object “healthcare”.
2. Using the numerical portion of your U number as a random number seed, take a random sample of 150 cases from the full data set using the method presented in class. Title this data object “my.healthcare”.

**Analysis**

1. Conduct a full regression analysis including all variables in the my.healthcare data set using the “charges” variable as the dependent variable in the regression.
2. Show your model output. Interpret the beta coefficients in your output in terms of the variable’s estimated impact on the y. Include an appropriate discussion of the p value for each beta coefficient (including β0).
3. Report the confidence interval for each beta coefficient in your model.
4. Determine and state whether your model appears to be in conformity with the LINE assumptions of regression. Show appropriate graphics where needed to justify your conclusions.
5. Using the tools presented in Module 4 of this course determine whether any of your my.healthcare data points have a high leverage in influencing the plot of the regression. Show appropriate graphics to support your conclusion. Also, report the observations from your my.healthcare data (if any) which have such high leverage.

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 “Assignment 5” 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.