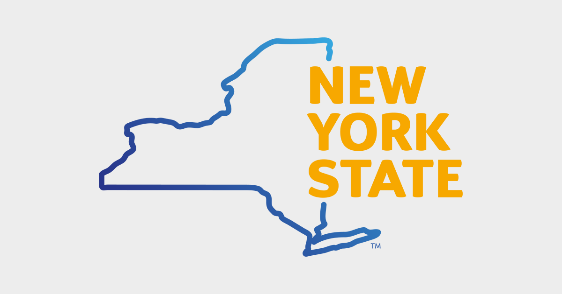
**Module Assignment**

**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 public information on 426,391 employees of the State of New York. This will be your master data set.
2. There are multiple categories in the “Authority Name” variable of your master data set. Among those categories are the “Capital District Transportation Authority” and the “Central New York Regional Transportation Authority”. Using the numerical portion of your U number as a random number seed and the method presented in class take a random sample of n=500 employees from each of these categories and store in two separate data frames. There are multiple ways this can be accomplished. These will be your final data sets for analysis.

**Analysis**

1. Using your sample data for the “Capital District Transportation Authority” construct a 90% confidence interval on the population mean Base Annualized Salary. Give a clear incorrect interpretation of your confidence interval.
2. Assuming Base Annualized Salary of all 2657 listed employees of the “Capital District Transportation Authority” represents the population, does your 90% confidence interval include the true population mean Base Annualized Salary?
3. Again using your sample “Capital District Transportation Authority” data construct a 95% confidence interval on the population mean for Base Annualized Salary. How much wider is the 95% confidence interval than the 90% confidence interval?
4. Using your sample “Central New York Regional Transportation Authority” data can you say (α = .05) that the population mean Base Annualized Salary is less than $50,000? How about less than $40,000?
5. Using your sample “Central New York Regional Transportation Authority” data what “test against” (mu) value in a two-tailed hypothesis test would yield p = .05 in a test on Base Annualized Salary?
6. Using both sample data sets, does there appear to be a statistically significant difference (α = .05) between the population mean Base Annualized Salary for employees of the two agencies?

Your deliverable will be a single MS-Word file showing 1) the R script which executes the above instructions, 2) the results of those instructions, and 3) any interpretations asked for. 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.**