

Name _____

STT 3850 Test 1

Directions: Start R Studio and enter the command **attach(mtcars)** on the R console. This will allow you to access the contents of the data frame **mtcars**. Enter the command **help(mtcars)** to get a description of this data frame then answer the following questions:

1. How many cases and how many variables are recorded in this data frame?
Is there a categorical variable among the variables in this data frame? If so, identify one and determine what percent (rounded to the nearest whole number) of the cars belong to each of the categories.

2. Describe the shape of the distribution of the **mpg** data. Calculate two measures of center and discuss what the values for these measures tell us about mpgs of the cars in this data frame.

3. Calculate two measures of spread or variability of the cars' **mpg** and discuss what their values tell us about mpgs of the cars in this data frame.

4. Calculate the first quartile and the 90th percentile of the **mpg** data and discuss what their values tell us about mpgs of the cars in this data frame.

5. Is there an outlier among the cars with respect to **mpg**?
Given the shape of the distribution of the **mpg** data, is it surprising that the mean is higher than the median? Explain.

6. What percent of the cars have **mpgs** that are more than 2 standard deviations higher than the mean **mpg**?
Explain why this percentage is not close to the percentage predicted by the empirical rule and state this predicted percentage.
7. Suppose the 32 cars on the **mtcars** data frame are treated as the population in one study, what is the population standard deviation of the **mpgs** of these cars?
8. Write a short paragraph comparing the **mpg** of cars with manual transmission (**am** = 1) with those with automatic transmission (**am** = 0).
9. Suppose cars in the data frame that weigh over the median weight (**wt**) are classified as heavy and cars that weigh less than or equal to the median weight are classified as light. What can we say about the relationship between **wt** and **mpg** of cars based on the scatter plot of the data? What is the difference in the average **mpg** between light cars and heavy cars?
10. Suppose another set of 32 cars from the 1980s have mpgs that are perfectly normally distributed with a mean of 22 and a standard deviation of 6. What percent of the 1980s cars have mpgs of least 30?
What is the 95th percentile of the mpgs of these cars from the 1980s?