**OEB 137 – Assignment 3**

**Due September 30, 2020 at 11:59 am (Eastern)**

**Data Distributions**

The following tasks are meant to establish your understanding of data distributions and how to manipulate and visualize them in R.

1. Load in the data frame attached with the assignment and put it in an object of your own naming. Plot separate histograms for each column in this new object. Based on looking at the histograms and thinking about various properties of the data that you see, indicate which distribution the data in each column represent and provide a few words as to why you suspect that data distribution. **(2pt)**
2. Create three objects called *n1*, *n2*, and *n3*. Have each object hold a vector of random numbers pulled from a normal distribution. Make *n1* a vector of 10 numbers from a distribution that has a mean of 13 and a standard deviation of 4, *n2* a vector of 100 numbers from a distribution that has a mean of 3 and standard deviation of 0.85, and *n3* a vector of 1000 numbers from a distribution that has a mean of 8 and a standard deviation of 7. \*\*Use a set.seed() command to allow reproducibility of these objects\*\* **(1pt)**
3. With these objects, plot 6 histograms displayed together in a 3 row x 2 column paneled figure. Two histograms for *n1*, two for *n2*, and two for *n3* should make up the top, middle, and bottom rows, respectively. For the left-hand histogram in each row, plot the frequency histogram with 8 bins. For the right-hand histogram in each row, plot the probability density histogram with 50 bins. **(2pt)**
4. Conduct a Shapiro-Wilks test on your three objects to check for normality. What do the P-values for this test mean? **(1pt)**
5. Create an object that contains a dataframe with two columns that are each 500 rows long. The first column should contain a vector of 500 observations taken randomly from a Poisson distribution with a lambda of 0.28. The second column should contain a vector of 500 observations that represent a binomial distribution with a probability of 0.72. **(1pt)**
6. For your own research question – Indicate what the primary dependent/response variable(s) will be and what the primary independent/driver variables will be. For each variable, indicate whether the data will be categorical or numeric, nominal or ordinal (for categorical data), discrete or continuous (for numeric data), what type of distribution you expect the data to follow, and why you expect this distribution. **(3pt)**