**Excel Project #3: Assessing Normality**

**Due Date:**

Introduction: In this project, you will be using Excel to analyze several datasets to determine whether or not they satisfy the criteria for normality. You will be provided with the datasets as an Excel spreadsheet. Go to *D2L > Contents > Excel Projects > Project #3* to find the folder containing the datasets. You will find a file called '*EP3\_yourname.xls*'. Replace *yourname* with your own name and save the workbook. The workbook contain a spreadsheets that can be accessed through the tabs on the bottom of the workbook. Follow the instructions for each part and then upload the finished product to D2L before the due date. If you have any questions, email me at: [grant.moore@frontrange.edu](mailto:grant.moore@frontrange.edu) .

Instructions: For each of the datasets contained in the spreadsheet, perform the following steps.

1. Sort the dataset from the smallest observation to largest observation.

2. Calculate the following summary statistics. Save and label your answer.

a. The number of observations.

b. The sample mean.

c. The sample standard deviation.

3. In a new column, calculate the Z-Score of each observation. Recall the formula for an observation's Z-score is,

where is the i-th observation,is the sample mean and *s* is the sample standard deviation.

4. In a new column, estimate the percentile of each observation using the approximation,

where *i* is the # of the ordered observation whose percentile you are calculating and *n* is the total number of observations.

5. Calculate the theoretical Z-score of each percentile by using the inverse normal distribution. In other words, suppose the percentile is stored in the cell *H3*. Then, you should calculate the theoretical Z-score using the formula,



6. Using the actual Z-score and the theoretical Z-score as your dependent and independent variable, create a scatter plot. Insert the line of best fit and display R-squared on the graph.

i. Based on your normal quantile plot (i.e. the scatter plot of actual versus theoretical Z-scores), does this dataset appear to be normal? Why or why not?

*Hint*: If a dataset is normal, it's quantile plot should be a straight line! In other words, the actual Z-scores should match up with the theoretical Z-scores!