**ENV 503: Statistics for Bioinformatics**

**Homework Set #1**

**Due: September 5, 2018**

*Instructions:*

*Use any form of technology you wish (Excel, R) to complete this assignment.*

*Assignment is to be submitted via Blackboard.*

Use the dataset **airq\_june.csv** to answer all questions.

1. List each variable in the dataset and characterize it in terms of the following:
   1. quantitative vs. categorical
   2. if categorical: is it on an ordinal or nominal scale?
   3. if quantitative: is it on an interval or ratio scale? Continuous or discrete?

Ans. List of variables:

Temperature=Quantitative, interval, Continuous.

Wind speed=Quantitative, interval, Continuous.

Solar radiation=Quantitative, interval, Continuous.

1. Plot a histogram for temperature, wind speed, and solar radiation. Describe each distribution in terms of symmetry/skewness, how flat/peaked, presence of outliers.

Figure 1: Histogram for temperature.

* + Not symmetric
  + Skewed left
  + Peaked
  + No outliers.

Figure 2: Histogram for wind speed

* Not symmetric
* Skewed left
* Peaked
* No outliers.

Figure 3: Histogram for Solar radiation

* Not symmetric
* Skewed left
* Peaked
* No outliers.

1. Calculate the five number summary plus mean and standard deviation of temperature, wind speed, and solar radiation.

Ans.

Temperature Wind speed Solar radiation

Five-number summary Max 93 20.7 332

Q3 82.75 11.5 270.75

Median 78 9.7 188.5

Q1 76 8 127

Min 65 1.7 31

Mean 79.1 10.26667 190.1667

Standard deviation 6.598589 3.769234 92.88298

1. Calculate 95% and 99% confidence intervals for mean temperature, using the z formula. In your own words, what does a confidence interval mean? Which confidence interval is narrower, and why?

95% confidence interval=2.463954 width=4.927907

99% confidence interval=3.320706 width=6.641413

Confidence intervals represent an estimated range of value in which the true value of the mean will fall.

We know the more accuracy means the higher confidence level, and the higher confidence level has the higher width. In this case,95% is the lower confidence level than 99% confidence level; therefore, it has a narrower width.

Hint:

Excel can generate histograms if the Data Analysis ToolPak is installed. Select “Histogram” on the Data Analysis menu.