Math 425/525 - Computer Homework 1
Due Wednesday 03/27

The data sets required for this homework can be found on the course website in the SPSS section:

http://math.iit.edu/~mccomic/525/spss

In addition to the output from SPSS, I would certainly appreciate some terse comments about your analysis of the data. If you choose to use a software package other than SPSS, please present your results in an orderly fashion.

Also note that as with all real world data, more information may be present than is required to answer these questions. It is up to you to determine what variables are superfluous.

1. For this problem you need data set OhioCountyPop.csv and we will be studying the population of the 88 counties in Ohio.

   (a) Using descriptive statistics determine the mean, median, standard deviation, skew (positive or negative), range.

   (b) Produce a histogram for this data set.

   (c) Explain why standard deviation may be somewhat confusing for this data set.

   (d) **BONUS:** In SPSS you can do operations on whole columns of data using the Compute option under the Transform tab. Using the SUBSTR command, create a column which is just the first letter of each county in Ohio.
   Using Descriptive Statistics determine the "Median Letter" of the counties in Ohio; the "Median Letter" is the letter for which 50% of the population lives in a county with a letter earlier in the alphabet, and 50% lives in a county with a letter later in the alphabet. If it doesn't come out to exactly 50% just use the nearest letter and the corresponding percentile.

2. In the SPSS tutorials section, check out the tutorial called "Analyzing a Frequency Table". It will teach you how to apply frequency table structure to the data. For this problem you need data set SchoolEnrollment.xls and we will be studying the school enrollment of people ages 3-24.

   (a) Determine the mean and median age of the population between 3 and 24.

   (b) Determine the mean and median age of the school enrolled population between 3 and 24.

3. This problem requires using the Bear Body Measurements data set. The age listed is in months, the month listed is the month in which the bear was captured (1=January, 2=February,...), the sex listed is 1=Male and 2=Female.

   (a) Using the independent samples t-test, determine which of the six listed measurements have a significant difference by sex. List the confidence intervals for any significant results, and as always we will assume equal variance. *(Hint: it will be faster if you put all the columns in the same t-test and let SPSS run them at the same time.)*

(b) In my extensive study of mammals, I have learned that bears fully mature by 2.5 years old. Using the cut-point option in the independent samples t-test, determine if there is a difference in weight for bears under 30 months old and over 30 months old. *(Hint: just put 30 in the cut point box after clicking define groups under the Grouping Variable/Factor box)*

(c) It is my belief that the results of this data set may be somewhat biased by the time of year it was conducted. Determine if the bears captured in October-December had a larger chest size then bears captured other times of the year. Try to explain why a significant difference would make me worried.

4. This problem requires the use of the Systolic Blood Pressure data set. The stress values listed are in blood pressure units (whatever they are) and some identifying material is also provided.

(a) Using the independent samples t-test, determine which of the stress conditions has a significant difference by race. Make sure to note whether or not you can assume equal variance, and to choose the appropriate t-test results.

(b) Determine the median blood pressure for someone before exercise with no stress. Now test whether being above or below that median produces a significant difference for the post exercise with no stress group.

(c) Create a "Classifier" column with values

Classification	Value
White Male	1
Black Male	2
White Female	3
Black Female	4

You may want to look at the tutorial online about Transformations to help you perform this. Using this column, determine if there is a statistical difference between white and black female blood pressures for pre-exercise with math stress. What about post-exercise with math stress?

(d) Come up with some other test that you would like to run, explain what it is, and why it is interesting, and then explain the results of your test.