Homework 1

Due: Wednesday February 1 at 11:59pm

See general homework tips and submit your files via the course website.

For all exercises, use the iris data set from the SAS help (e.g. **data=sashelp.iris**).

Exercise 1:

a) Obtain box plots (see SAS help for **proc boxplot**) for **petallength** by **species** and comment on any differences you notice between the different species' petal lengths.
b) Obtain basic descriptive statistics for the petal lengths for all species together. Comment on any general features of the data (e.g. typical petal lengths, spread of values, skewness, etc.). Also visually and quantitatively check if an assumption of normality would be reasonable for the underlying population.
c) Obtain basic descriptive statistics and visually and quantitatively check the assumption of normality for **petallength** by **species**. Comment on how the species-wise statistics differ from those for all species combined, and comment on the conclusions of the species-wise normality tests.

Exercise 2:

a) Test the null hypothesis that the true mean or median iris petal length is **45** against the alternative that it is not **45**. Use the normality tests from Exercise 1 to determine which location test should be used and comment on what we conclude about the true mean or median petal length of the population.
b) Of the three species, **Virginica** has the highest mean and median petal length, but is it significantly greater than the general population? Use the sample median for petal length of all species as the null value and perform a t, sign, or signed rank test to test whether **Virginica** has significantly larger petal length compared to the general population. (Note: you should have the median value in results from Exercise 1, and tests from Exercise 1 should also tell you whether to use a t test or one of the rank-based tests).
c) Consider the **Versicolor** and **Virginica** petal lengths. Test for differences of the two populations (e.g. test for a difference of mean or median if appropriate or test if one population is stochastically greater if testing the difference of means would not be appropriate), and state your conclusion.

Exercise 3:

a) Obtain the Pearson correlation matrix for the **Versicolor** and **Virginica** data combined together. Comment on what the results tell us about significant relationships between the four length and width measurements for the population of these two species together.
b) Obtain the Pearson correlation matrices for **Versicolor** and **Virginica** separately. Comment on what the results tell us about significant relationships between the four length and width measurements for each of these two species and how these relationships compare with those noted when these two species were combined.