Homework 1

Due: Friday February 3 at noon.

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

In this homework assignment, we will examine and compare some attributes of fish. For all exercises, use the dataset **fish** defined in **HW1Data.sas** in the course space. **Fish** data set in the **HW1Data.sas** file is from the SAS Help library. The variables are:

- **Species:** bream, perch, roach
- **weight**
- **length1**
- **length2**
- **length3**
- **height**
- **width**
- **id:** identification number

Exercise 1:

a) Obtain basic descriptive statistics for **weight** for all of the data. Provide basic descriptive statistics (eg, mean, median, std, skewness, and range).
Hint: you are expected to provide your interpretations, simply print out SAS output tables are not good enough.

b) Repeat the analysis in part **a)** by **Species**. Comment on what this tells us about similarities and differences of weight between the three species of fish.

Exercise 2:

a) For **weight**, visually and quantitatively check if an assumption of normality would be reasonable and state your conclusion.

b) Repeat the analysis in part **a)** by **Species**. Based on the results, comment on whether these three species of fish show significant differences from normality.

Exercise 3:

a) Perform a hypothesis test of whether bream fish species have significantly heavier weight than perch, and state your conclusions. Which test did you use and why?
Hints: consider the normality tests in **Exercise 2** when choosing your test for difference in **weight**. You also need to subset your data using the "where" statement.

Exercise 4:

Scatter plots are useful tools in examining relationship between variables. A pairwise scatter plot can be used to create scatter plots for all pairs of variables in a data set against each other. This can provide us

with a visualization of relationships between pairs of variables (it's like a visualization of a correlation matrix). In SAS a pairwise scatter plot can be created using **proc sgscatter** with the **matrix** statement.

a) Create a pairwise scatter plot for all continuous variables (except **id**) of the bream fish. Visually inspect and comment on which pairs of variables might be correlated.

Exercise 5:

a) Perform correlation analyses for all continuous variables of the bream fish. Use both Pearson and Spearman correlations. Which pairs of variables have statistically significant correlations? For each correlation measure, state the magnitude of the correlation of any statistically significantly correlated pairs. Do Pearson and Spearman correlations give us similar results? If so, why?