Statistics 135 – Assignment 1
Due: Wednesday, October 5, 2005

Please submit your solutions in a word processed format (I don't care which one you use). Also as part of your solutions, please give the commands used to generate the output. I'm curious to the approaches you take as some of the questions have multiple valid ways of getting the answers. For example, I can think of two ways of generating the set of summary statistics desired in one of the early questions. Also in your figures, where appropriate, please give informative axis label, figure titles, etc.

1. Download the data iris.dat from the course web site (either from the Datasets page or the Assignments page) and read the file into **R** using the `read.table` command into a data frame named `Iris`. This data set is the famous Fisher Iris data. It contains 4 variables (Sepal Length, Sepal Sepal Width, Petal Length, and Petal Width) on 50 flowers from each of 3 species (Setosa, Versicolor, and Virginica) of iris. All size measurements are given in centimeters. For a more complete description of the dataset, check the description file which is also available at the same locations as the dataset.

2. Calculate the standard summary statistics (mean, standard deviation, median, 1st and 3rd quartiles, min, and max) for the four variables for the combined data set and for each species separately.

3. Create a new data frame with the measurements given in inches instead of centimeters (use the conversion 1 inch = 2.54 cm). Show the first 5 rows of the two data frames.

4. Draw histograms of the four numeric variables. Please combine the four histograms into a single figure (using `par(mfrow=c(2,2))`). When creating figure, make sure it is clear which variable is being plotted in each histogram with an informative label.

5. For the two Petal variables, draw side by side boxplots for each species. Please combine these two plots into a single figure.

6. Create a scatterplot of Sepal Width (x axis) vs Sepal Length (y axis), superimposing the least square regression line on the plot. In addition, redo this scatterplot, but with different plotting symbols for each species. For this second plot, add the least squares regression lines for each of species. Use different line types for the different species. Use the legend command to indicate which symbol and line correspond to the different species.

7. It is possible to show that one way to generate an exponential random variable (RV) with mean μ is to generate a Uniform(0,1) RV, U and make the transformation $V = -\mu \log U$.

   (a) Generate 100 uniform random variables and store them in a vector `u`. (`u <- runif(100)`). Produce a histogram of the vector `u`.

   (b) Generate 100 exponential random variable with mean 2 using the **R** function `rexp` and store them in a vector `v`. Check the help page to see how **R** parameterizes the exponential distribution.

(c) Using the 100 uniforms generated in part (a), generate 100 exponentials with mean 2 and store them in a vector `w`. Examine whether the vectors `v` and `w` appear to really have the same distribution by examining summaries and graphs of the data. One useful graphic for this examination is a QQ plot (see help for `qqplot` function).