36-401 Homework #1 Due 9/7/00

Don't include S-plus code in your homework unless specifically instructed to do so. You may include code as an appendix so that the grader can point out the source of any errors.

1. **10 pts** Create a random vector using "ch_rchisq(40,1)". Plot a histogram, choosing a number of bars that displays the data well. Turn in your plot and the mean and median of the vector.

2. **10 pts** Load the data in the file "widgets.dat". Turn in the list of variable names, the number of rows, and the range (max-min) of each column.

3. **10 pts** Not all data comes in a ready-to-use format! Enter the data from problem 8 in RwG (p. 61). Use a text editor such as xemacs to complete the entry of the data in "seapart.dat". Add a one-word identifier to the top of each column. Use read.table() to load your data into S-plus. (The correct means are 965.3636 and 2452.818.)

   Construct a scatter plot showing area on the x-axis and pairs on the y-axis using something like "plot(seabird$area,seabird$pairs)". Fix up the plot so it has x and y labels, a title, and your name as the subtitle (at the bottom).

   Find the median and sd of each column.

   Turn in the scatter plot, the medians and sds (with an appropriate number of decimal places) and one or two sentences qualitatively describing the apparent relationship between the two variables.

4. **20 pts** Download "cane.dat" and "cane.txt" from the class website. Download the ndhist.q, rwgbox.q, symplot.q, qplot.q, qqn.q, and scatbox.q from the website. Check that you have all of the files using the Unix command "ls".

   Print "cane.txt" or read it with the Unix "more" command. Use read.table() to load the data in the "cane.dat" file into an object named "cane". (Be sure to use the header=T argument, since the first line of the file has the column names.) Check the dim(cane) and names(cane). View columns 6 to 12 for the first 10 cases using "cane[1:10, 6:12]".

   Load each of the functions in the .q files listed above using the source function. Try the first 5 functions on the Sugar variable, e.g. ndhist(cane$Sugar). (Note: the message "Error: getvec: unrecognized mode (means that you misspelled what comes after the dollar sign.) Also run "scatbox(cane$Sugar,cane$Area)" to compare these variables. Use the "mfrow=" argument of the par() command to create a single graph containing all 6 subgraphs. Print it, being certain to include your name as a subtitle.

   Turn in only the printout with the 6 plots on one page along with 1 or 2 sentences per graph describing the most important features revealed by each of the 6 graphs.