Assignment #2

Due on Wednesday September 17, 2008

Read Sections 1.1 on *Displaying Distributions with Graphs,* and Section 1.2 on *Describing Distributions with Numbers* in Moore, McCabe and Craig.

Read Appendix A in the class notes, *Calculations and Analysis Using R,* in the course webpage at `http://pages.pomona.edu/~ajr04747`

Do the following problems

1. Read the data contained in the MS Excel file Westvaco.xls, which may be downloaded from`http://pages.pomona.edu/~ajr04747`, into a dataframe that you may name `WestvacoD`. Define variables `Age` and `RIF` as described in Appendix A in the class notes at `http://pages.pomona.edu/~ajr04747`. Obtain vectors `laidoff` and `kept` containing ages of laid off workers and kept workers, respectively. Print a box plot showing the distributions of the two variables and write a paragraph comparing the two distributions. Does the picture show enough evidence to conclude that the company discriminated against older workers?

2. Write an R script which picks out those workers that get paid hourly from the `WestvacoD` dataframe and puts their ages into a vector called `hourly`. Test the script and explain the procedure you used. Do the same for the salaried workers putting their ages into a vector named `salaried`. Plot box plots of the distributions of the two variables in the same graph and compare. Is there any significant difference between the two distributions?

3. Write R scripts that pick out the workers that get laid off in a rounds 1, 2, 3, 3, 4 and 5, and puts them into vectors `round1`, `round2`, ..., `round5`, respectively. Plot box plots of the distributions of all the variables in the same graph. Discuss the picture.

4. In a study reported by Roseann M. Lyle et al. in the *Journal of the American Medical Association,* Vol. 257 (1987), pp. 1772-1776, titled *Blood Pressure and metabolic effects of calcium supplementation in normotensive white and black man,* results of an experiment testing the effects of calcium supplementation on the decrease of systolic blood pressure were reported. The experiment involved 21 black men who were randomly divided into a group of 10 and a group of 11. The first group of men was given calcium supplement pills for 12 weeks, and the second group (the *control* group) received a placebo pill identical to the calcium supplement also for 12 weeks (this is an example of a *randomized comparative experiment*). The men were all instructed to take one pill a day.

The men involved in the experiment did not know which pill they were taking. Also, the people administering the pills and those measuring the subjects' blood pressure did not know which pill the men were taking (this is known as a *double blind* experiment). Systolic blood pressures were measured at the beginning and end of the experiment. The data from the experiment are recorded in the MS Excel file, `CalciumBloodPressureData.xls`, which may be downloaded from `http://pages.pomona.edu/~ajr04747` (follow the link to Data Files in the Math 58 section). Decease in the systolic blood pressure are also recorded in the last column of the data set (the one labeled `dec`). Note that a negative value in the decrease columns indicates that the blood pressure actually went up.

Import the table in the spreadsheet to a dataframe in R which you may name `CalciumBloodPressureD`. Operate on the dataframe to obtain variables `group` and `dec`. Pick out the `dec` values for those subjects in the `calcium` group, and those for the `placebo` group. Put them into vectors called `calcium` and `placebo`, respectively.

Compute number summaries for each of the variable `calcium` and `placebo`. Is there any difference in the decrease in blood pressure in the two groups?

5. Plot box plots of `calcium` and `placebo` in the same graph. Print the plot and discuss your observations. Would you say that calcium supplementation has an effect in the decrease of blood pressure?