Course EPIB-613 - Introduction to Statistical Software

Assignment 1

1. The table below gives data on blood pressure before and after treatment for two groups of subjects participating in a clinical trial. One group took a daily calcium supplement, while the other group received a placebo.

Calcium Group		Placebo Group	
before	after	before	after
107	100	123	124
110	114	109	97
123	105	112	113
129	112	102	105
112	115	98	95
111	116	114	119
107	106	119	114
112	102	112	114
136	125	110	121
102	104	117	118
111	101	130	133

Write commands to create a data frame in R from these data, in three different ways:

(a) Directly using the R command line. Call the data frame calciumA.

(b) By using the script facility in R, placing the commands into the script file, and then running in R. Call the data frame calciumB.

(c) By creating and saving a text file containing the required R commands, and using the `source` command in R to read in the commands. Call the data frame calciumC.

In each case, check whether the data have been entered correctly as a data frame by typing the name of the data frame from the R command line. Note that the data

frame should have three variables, which you can name bp.before, bp.after, and group. For the group variable, enter the data as a character data, with "calcium" and "placebo".

(d) Create a variable called diff defined by the within subject (after - before) differences. Add this variable to the data frame calciumA, so that it now will have a total of four variables.

2. In this question we will work with various different data types:

(a) Create a scalar called a with value 5, and a vector called b consisting of the numbers from 1 to 5. What happens if you multiply a by b?

(b) Use the `help` command to look up how to use the `round` command.

(c) Create a 5 row by 10 column matrix consisting of random normal variables with mean 5 and variance 10, rounded to one decimal place. Call this object mat.

(d) Multiply the result in (c) with the scalar a from part (a). Look at the result.

(e) Multiply the result in (c) with the vector b from part (a). Look at the result. What pattern do you notice?

(f) Create a list consisting of objects a, b, calciumA, calciumB, calciumC, and mat. Call this object mylist. Name all of your objects. Check your result by typing mylist from the command line.

(g) Using the $ notation associated with lists, pull out only the mat object from mylist.

(h) Using the $ notation and the [ ] subscripting facility, pull out only the number in the second row, third column of the mat object from the list.

(i) Using the $ notation and the [ ] subscripting facility, pull out all numbers greater than 5 from the mat object from the list.