Homework 1

36-350, Statistical Computing

Due at 10:30 am on Wednesday 7 September 2011

Submit a single plain text file, whose name clearly includes both your Andrew ID and the assignment number. Inside the file, clearly indicate which parts of your responses go with which problems. Raw R output is not acceptable, and will be marked down accordingly; you are communicating with a human being, and need to write in a human language.

- 1. Read chapters 1 and 2 (pp. 1–32) of Braun and Murdoch.
- 2. Working with data (75 points total) The data set at http://www.stats. uwo.ca/faculty/braun/data/rnf6080.dat records hourly rainfall at a certain location in Canada, every day from 1960 to 1980.
 - (a) Load the data set into R and make it a data frame called rain.df. What command did you use? (6 points)
 - (b) How many rows and columns does rain.df have? How do you know? (If there are not 5070 rows and 27 columns, you did something wrong in the first part of the problem.) (6 points)
 - (c) What command would you use to get the names of the columns of rain.df? What are those names? (7 points)
 - (d) What command would you use to get the value at row 2, column 4? What is the value? (7 points)
 - (e) What command would you use to display the whole second row? What is the content of that row? (7 points)
 - (f) What does the following command do? (7 points)

names(rain.df) <- c("year","month","day",seq(0,23))</pre>

- (g) Create a new column called daily which is the sum of the 24 hourly columns. (7 points)
- (h) Give the command you would use to create a histogram of the daily rainfall amounts. Submit this histogram as a separate PDF file, named with your Andrew ID, the assignment number, and Fig1. (7 points)
- (i) Explain why the histogram from the previous question cannot possibly be right. (7 points)

- (j) Give the command you would use to fix the data frame. (7 points)
- (k) Create a corrected histogram and submit a PDF file of it, named as before but with Fig2 instead of Fig1. Explain why it is more reasonable than the previous histogram. (7 points)
- 3. *Data types* (25 points total) Make sure your answers to different parts of this problem are compatible with each other.
 - (a) Which of the following commands will produce an error message? For each command, Either explain why they should be errors, or explain the non-erroneous result. (9 points)

x <- c("5","12","7")
max(x)
sort(x)
sum(x)</pre>

(b) For the next two commands, either explain their results, or why they should produce errors. (8 points)

y <- c("5",7,12) y[2] + y[3]

(c) For the next two commands, either explain their results, or why they should produce errors. (8 points)

z <- data.frame(z1="5",z2=7,z3=12)
z[1,2] + z[1,3]</pre>