Statistics 133

Assignment #1 – R Expressions

**Due Thursday, Sept 11 at 11:59 pm on bSpace**

Look at the lecture notes for Lecture 2-3 under Resources on bspace for help. Many of you will be able to do Part 1 and installation of software (hopefully done already) without any guidance.

**Part 1: Questionnaire, saving to pdf**

1. Download the file FirstQuestionnaire.doc from bSpace. Open it up in Microsoft Word (or equivalent) and add your answers to the file.
2. Take a photo of yourself. (You can use the PhotoBooth application on the Macs in the lab to do this, or use a photo you already have.) Add it to the document with your questionnaire answers. Your picture should make it easy to recognize you, it will be used to help memorize your names.
3. Save the file as a .pdf document. (If you're working on a Mac this option is available via the Print menu. If you are working on a Windows or UNIX machine you can use free software such as CutePDF or cups-pdf.) Rename the .pdf file with your name, putting your last name first, e.g. Barack Obama’s file would be called ObamaBarack.pdf.
4. Submit your .pdf document on bspace.

**Part 2: Intro to R**

Start an interactive R session with RStudio, and use the File menu to open the script “HW1\_Rexpressions.R” file. This file should appear in the top left window of RStudio. You will add your answers to the questions there. This document is to be turned in to bspace.

Begin by entering your name after # Name: in the file

Load the data for this assignment into your R session with the following command:

load(url("http://www.stat.berkeley.edu/users/nolan/data/stat133/SFTemps.rda"))

To run this command, cut and paste into the upper left pane of RStudio, place your cursor at the beginning of the line of code, and then click on the Run button in the tool bar for this window. You must have an Internet connection for this to work.

This command loaded five objects into your workspace: temp, dayOfMonth, month, year, dates. The object called “temp” holds the daily temperature for each day for about 15 years.

Use the objects() command to confirm that you have these five objects in your workspace.

objects()

Use the length() function to find out how many observations there are in temp.

length(temp)

Answer the following questions by using functions in R. NA is not an acceptable answer to any of these questions.

1. Find the average daily temperature.

2. Find the 10% trimmed average daily temperature

3. Find the 50% trimmed average daily temperature

4. Compute the median daily temperature. How does it compare to the 50% trimmed mean? Explain.

5. We would like to convert the temperature from Farenheit to Celsius. Below are several attempts to do so. Each of these should fail. Try to run each expression in R.

Record the error message in a comment. Explain what it means. Be sure to directly relate the wording of the error message with the problem you find in the expression.

(temp – 32)\*5/9

(temp - 32)5/9

5/9(temp - 32)

[temp - 32]\*5/9

6. Provide a well-formed expression that correctly performs the calculation that we want. Assign the converted values to tempC

For the following questions, use one of: head(), summary(), class(), min(), max(), hist(), quantile() to answer the questions.

7. What were the warmest and coldest temperatures recorded in this time period?

8. What does the distribution of temperatures look like, e.g, Are there roughly as many warm as cold days? Are the temps clustered around one value or spread evenly across the range of observed temperatures?, etc.

9. Examine the first few values of the dates variable. These are a special type of data. Confirm this with class().

10. Run the following code to make a plot. Don't worry right now about what this code is doing. We’ll cover that soon.

plot(temp~dates, col = rainbow(12)[month], type="p", pch=19, cex = 0.3)

Use the Zoom button in the Plots window to enlarge the plot. Resize the plot so that it is long and short.

Make an observation about temperature in the Bay Area based on this plot (something that you couldn't see with the calculations so far.)

For the remainder of this assignment we will work with one of the random number generators in R. The rnorm() function is very handy for generating random values from the normal curve.

11. Use the following information about you to generate some random values:

* 1. Use your year of birth for the mean of the normal
  2. Use the day of the month you were born for the sd of the normal curve.
  3. Generate either 3, 4, or 5 random values, depending whether you were born in the first four months of the year, middle four months, or last four months.
  4. Assign the values to a variable matching your first name.
  5. Provide the values generated

12. Generate a vector called “normsamps” containing 1000 random samples from a normal distribution with mean 1 and SD 2.

13. Calculate the mean and sd of the 1000 values.

14. Use implicit coercion of logical to numeric to calculate the fraction of the values in normsamps that are less than 3. Implicit coercion means that you treat one data type as though it were another data type without explicitly changing the data type.

15. Look up the help for rnorm. You will see a few other functions listed. Use one of them to figure out about what answer you should expect for the previous problem. That is, find the area under the normal(1, 2) curve to the left of 3. This should be the chance of getting a random value less than 3.