

Psych 240 Lab 2

The purpose of this assignment is to show you how you can use R to get some basic descriptive statistics and plot a histogram for a distribution.

Setup Instructions (please complete these steps before the TA begins their presentation for the day)

- 1) Log on to your lab computer with your OIT username and password. Alert the TA if you have any trouble with this.
- 2) Go to the course Moodle page. In the “Labs” section, click on the file called “Lab2”. In the dialog box that opens, select “Save File” and hit OK. If you are asked to select a directory, choose the “Downloads” folder. If you are not asked to select a directory, the file should automatically be saved in “Downloads”.
- 3) On the Start Menu, select “All Programs” – “Statistics” – “R” – “R x64 3.1.0”. You should see R launch on your computer screen.
- 4) In the R console, type the following text: `setwd("C:/Users/your-OIT-username/Downloads")`. For example, if your OIT username is “astudent”, you would type in `setwd("C:/Users/astudent/Downloads")`.
- 5) Next, enter this at the R console: `source("Lab2.txt")`. If all goes correctly, you should be asked to enter your student ID number. If you get a “file not found” message, then you either didn’t save the Lab2 file from Moodle in the Downloads directory or didn’t set R to your Downloads directory. You can get help from the TA or another student.
- 6) If you see a prompt that says “Student ID not found”, check your ID number and try to enter it again. If it won’t work after several tries, you can just enter 0 instead of your ID and the program will let you continue. If you do this, you will not get a completion code at the end of the assignment, so you have to show the TA your R screen after you complete the assignment and before you leave.
- 7) Wait for further instruction from the TA.

Lab Demonstration

The TA will guide you through this section. Please wait until you are asked to begin and follow along with the TA.

1) One of the things that R is very useful for is helping you understand a data set. As we mentioned in class, two ways that we explore a data set are with descriptive statistics and histograms.

2) The Lab2 file has (made up) data from a hypothetical study in which 50 college students were asked to report how much they spent on textbooks in the current semester (stored in a vector called “txt”) and how much they spent on entertainment in the last month (stored in a vector called “ent”). Type in “txt” and hit ENTER, and you will see the text costs for each of the 50 students. Type in “ent” and hit ENTER, and you will see the entertainment spending for each of the 50 students.

3) One thing we need to know to understand this data set is the central tendency of the scores. The most popular measures for central tendency are the mean and the median, and R has functions that return both. Type in “mean(txt)” and hit ENTER, and you will see the mean of the textbook costs across all 50 students. Type in “median(txt)” and hit ENTER, and you will see the median.

4) Another thing we need to know to understand the data set is the variability. A popular measure of variability is the standard deviation. Type in “sd(txt)” and hit ENTER, and you will see the standard deviation of the textbook costs across the 50 students.

5) Finally, we need to know the distribution shape to understand a variable. R has a very convenient function for plotting histograms. As a most basic option, you can type “hist(txt)” and hit ENTER. R will set default ranges of scores and give you a plot showing the number of scores in each range. Please try this now, and let the TA know if you don’t have a plot show up on your screen.

6) The “hist” function can also take a number of useful arguments to control how your plot will look. One nice feature is that you can specify where you want to place the boundaries that define the different ranges of scores. Because you always want evenly spaced boundaries, another nice R function to use here is the “seq” function, which gives you a sequence of evenly-spaced values between two extremes. For example, say we want to look at how many students are in ranges of \$25 increments between \$100 and \$400 for textbook costs. We can define range boundaries by typing “seq(100,400,25)” and hitting ENTER. Try this now. In the “seq” function, the first argument is where to start, the second argument is where to stop, and the third argument is the width of each range.

7) Now we can integrate the “seq” function with the “hist” function to specify the ranges we want to plot on our histogram. The “breaks” argument in the “hist” function controls the ranges of scores (that is, how we “break up” the variable into different ranges). Try typing

`"hist(txt,breaks=seq(100,400,25))"` and hitting ENTER. You should see a different histogram appear. This one displays the same data as the first histogram, but now you see the number of scores in each \$25 range instead of each \$50 range.

8) Now let's try to repeat what we did above with the entertainment spending variable (ent). Get the mean and median for entertainment spending in these 50 students. Which is higher? Based on the relationship between the mean and median, what would you guess is the shape of the distribution for entertainment spending?

9) Now plot a histogram of entertainment spending with range boundaries from 100 to 1300 in increments of 100. To do this, you need to type `"hist(ent,breaks=seq(100,1300,100))"` and hit ENTER. How would you describe the shape of the distribution?

10) Based on your histogram, how many students spent between \$300 and \$400 on entertainment in the last month? How many students spent between \$700 and \$800?

11) Now get the standard deviation of the entertainment costs for these 50 students.

Lab Assignment

Complete this section on your own. You can ask the TA for help. Record all of your answers on a sheet of paper with your first and last name on the top. The correct answers are different for each student. R will tell you whether or not your answer is correct, as detailed below. When you enter all of the correct answers, R will tell you that you are finished and give you your completion code to write on your answer sheet. You can either leave when you are finished, or you can use any extra time to get help on your homework or the lecture content from the TA.

1) The "Lab2" file created a vector called "amherst" that holds the 2014 property taxes for 100 houses in Amherst and a vector called "hadley" that hold the property taxes for 100 houses in Hadley. What is the mean property tax amount for the 100 Amherst houses?

[Type in `"q1(your-answer-here)"` and hit ENTER to see if you are correct. For example, if you thought the answer was 12, you would type in `"q1(12)"`.]

2) What is the mean property tax amount for the 100 Hadley houses?

[Type in `"q2(your-answer-here)"` and hit ENTER to see if you are correct.]

3) What is the standard deviation in property taxes for the 100 Amherst houses?

[Type in `"q3(your-answer-here)"` and hit ENTER to see if you are correct.]

4) What is the standard deviation in property taxes for the 100 Hadley houses?

[Type in “q4(*your-answer-here*)” and hit ENTER to see if you are correct.]

5) Look at a histogram of the Amherst property tax amounts with range boundaries from 0 to 12000 in increments of 1000. How many houses in the Amherst sample paid property taxes between \$4000 and \$5000 in 2014?

[Type in “q5(*your-answer-here*)” and hit ENTER to see if you are correct.]

6) How many houses in the Amherst sample paid property taxes between \$6000 and \$7000 in 2014?

[Type in “q6(*your-answer-here*)” and hit ENTER to see if you are correct.]

7) Look at a histogram of the Hadley property tax amounts with range boundaries from 0 to 12000 in increments of 1000. How many houses in the Hadley sample paid property taxes between \$4000 and \$5000 in 2014?

[Type in “q7(*your-answer-here*)” and hit ENTER to see if you are correct.]

8) How many houses in the Hadley sample paid property taxes between \$6000 and \$7000 in 2014?

[Type in “q8(*your-answer-here*)” and hit ENTER to see if you are correct.]