STAT 217: Homework 1

Due Friday, Aug 28th beginning of class

Background: A study of the perceived stress levels of prisoners was conducted that involved measuring the Perceived Stress score that ranges from 0 to 56 points with higher scores related to more perceived stress. The prisoners were measured at the beginning and end of the study, with some of the prisoners involved in an exercise program and others not. We are interested in the changes in perceived stress levels over the course of 22 weeks of imprisonment for offenders. Specifically, we will calculate a Difference in stress (after 22 weeks – before 22 weeks) to generate a variable named Difference in the stress data frame.

Your task is to summarize those results using the provided code. In order to get started with R using R-studio, you will need to open R-studio and follow the instructions below. Some of the code is given to you, and some you will have to write yourself. The goal is to produce numerical summaries of the Differences, a histogram of the Differences, and a boxplot of the Differences.

Steps:

* Open a new R-script in RStudio. Go to File - New File - R Script. Save the script to your computer. Call it something meaningful like “Homework 1”
* Import the data into RStudio. The data file, stress.csv, is posted on D2L under DataSets. Save this file in an appropriate folder on your computer and follow the instructions on page 11 of the textbook.
* Notice a new tab with a data spreadsheet opens after import. Exit out of this tab and return to your “Homework 1” R script.

1. Your writeup starts here. In your writeup, print the code that you run and the output that you get for each of the following. Write the code in boldface font and the output in normal font. Follow the example I show in part (a).
2. At this point, you should have a new data.frame loaded into R called stress. Let's look at the data. Type stress into your R script. Put your cursor on this line and press Ctrl-Enter to run the code. You should see the dataset pop up in your console.

*Example Writeup*

Code: stress

Output:

Subject Group PSSbefore PSSafter

1 A1 Sport 25 13

2 A3 Sport 17 15

3 A4 Sport 12 13

4 A5 Sport 21 22

5 A6 Sport 29 25

6 A7 Sport 28 24

7 A8 Sport 21 19

8 B1 Sport 18 19

9 B2 Sport 20 24

10 B3 Sport 29 21

11 B4 Sport 27 22

12 B5 Sport 44 33

13 B6 Sport 23 8

14 B7 Sport 18 12

15 B8 Sport 27 30

16 A2 Control 0 16

17 C1 Control 30 27

18 C2 Control 12 31

19 C3 Control 29 21

20 C4 Control 25 33

21 C6 Control 2 9

22 C7 Control 6 26

23 C8 Control 20 20

24 C9 Control 27 28

25 C10 Control 14 21

26 C11 Control 15 29

1. Now, follow the instructions of page 12 of the textbook to lget R to print the first six lines of the stress dataset. Print your code and output in your writeup.
2. Now, follow the instructions to get R to print the last six lines of your dataset. Make sure to print your code and output in your writeup.
3. Recall that we are interested in the difference in stress levels (after-before) for each subject. The following line of code adds a new column to your dataset called Differences. Run this line of code and then look at the stress dataset again. Print the new stress dataset in your writeup.

stress$Differences<-stress$PSSafter-stress$PSSbefore

1. For i-v, follow the instructions on page 14 of the text.
2. Print only the Differences column
3. Print the mean and standard deviation of the differences.
4. Print the five number summary of the differences. The book says to use the function favstats, but instead use the function summary.
5. Make a histogram of the differences
6. Make a boxplot of the differences.
7. In a short written report, summarize the results, focusing on describing the shape of the distribution of the differences and whether you think the perceived stress levels changed over the course of the study (was the observed change very different from 0?). Refer to the figures and summary measures you found in part (e).

According to the results gained from section (e), the mean of the differences in the level of the prisoner’s stress is quite low which is only 0.8461538 while the deviation is much higher with 9.194313. Thus, we also obtained the histogram of stress differences which has a pretty wide spread with many different frequencies changing throughout the course of the study but has the highest frequency concentration from -5 to 5.