Psych 610, Homework 1

Due Wednesday, September 13, 2017, @ 5pm by email to Mitch (mcampbell9@wisc.edu)

**Part 1: Working with Data in R**

Save your answers to questions 1-6 in an R script named Lastname\_HW1.R (e.g., Campbell\_HW1.R). You *must* provide comments in your R script describing the command. See homework template for more info on this. Save your answers to questions 7-11 in a Microsoft Word document (or similar format) named Lastname\_HW1.docx.

**The Data**

For this first homework, we’re using fake data to practice your new skills. This hypothetical study is based on work in science and math learning (e.g., Schwartz, Chase, Oppezzo, & Chin, 2011). In this hypothetical study, suppose participants have been randomly assigned to one of two conditions, Explore or Teach, to learn about a new physics concept. All participants enrolled in the study completed a pretest, to assess their prior knowledge of the physics concept. Then, participants in Explore condition attempted to solve related problems on their own, and participants in Teach condition received a lesson about the concept directly. At the end of the study, participants’ learning was assessed using three posttest measures assessing conceptual understanding, procedural competence, and perceptual problem encoding. Thus, the data file includes the following 6 variables:

* Subject ID number (SubID).
* PreTest
* Condition (0 = Teach; 1 = Explore)
* PostCon
* PostProc
* PostPerc

1.  Read in "HW1\_Data.dat", from the course webpage. Name the data frame “d.” Use one of the functions we learned in lab to inspect the data, and then write out a list of all of the DV(s) and, separately, all of the IV(s)/predictor variables present in this dataset as a comment.

2. Obtain the following basic information about your data:

a. Get a summary of the data.

b. Get descriptive statistics for the data (mean, median, standard deviation, etc.).

c. Get descriptive statistics for just the Pretest variable.

d. Generate a histogram for the Pretest variable and give the graph a good main title (don’t export the graph, just generate it)

e. Using one line of code, find the difference between the post-test conceptual understanding score and perceptual problem encoding score of participant number 9.

3. Mean centering PreTest:

a. Create a new variable that is the mean-centered score for pretest knowledge. Give this variable a good name.

b. Check that your new variable is indeed mean-centered (include in your answer the code you could use).

4. Center the Condition variable, which is currently coded as 0 for Teach and 1 for Explore.

5. Average the posttest measures to create a single global measure of overall learning:

a. Create a variable that is the average of the three posttest variables (ignore missing data). Name this variable “d$MeanPostTest”

b. What are the mean and standard deviation of this average post-test score?

c. Generate descriptive statistics of this average post-test score for each of the experimental conditions.

6. Standardizing PostCon score

a. Create a standardized variable for the conceptual understanding measure.

b. How can you check your work to make sure this is the standardized score? Do it.

c. You have now created several new variables. Write out a command that shows all the variables in the “d” data frame (it’s okay if the command shows you more information than just the column names)

d. Generate descriptive stats for all variables in the d data frame.

Please answer Questions 7-11 in the word document named IDNumHW1.docx.

7. Create a table of descriptive stats (mean, standard deviation, minimum, maximum, skew) in publication quality format for the following variables: raw pretest score, each of the three raw posttest measures, and the averaged posttest measure. Give your variables understandable names. Please see the end of this document for an example of a publication quality table. *Feel free to use this table and edit it for the purpose of this homework.*

8. Make a scatter plot with raw pretest score as x-axis, mean posttest score as y-axis. Label the axes and the plot properly, using understandable and precise names. Add a regression line to this plot to reflect how mean posttest measure changes in response to pretest scores. Export this plot and drop it into your word document. Include the code for this plot in your R file.

9. Use the command we learned in lab to generate the average PostTest score for each Condition; what do you observe (answer in one or two sentences)? This code should appear in your script.

**Part 2: Reading Questions (Chapters 1 and 2)**

In the same Word document as Part 1, answer the following questions in a few sentences.

10. When fitting a model to data, the data analyst faces two conflicting goals. What are those two goals and why are they conflicting?

11. What are two consequences of using sum of squared errors rather than the sum of errors as a summary measure of model error?

Congratulations on finishing your first HW in 610!

In the Word document, please tell us how long it took you to complete this assignment.

Table 1

*Descriptive statistics for Variable 1 and Variable 2.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean Rating | *M* | *SD* | *Min* | *Max* | *Skew* |
| Variable 1 | 3.06 | 3.84 | 0.00 | 19.00 | 1.36 |
| Variable 2 | 1.36 | 1.85 | 0.00 | 9.00 | 1.05 |