**Lab Assignment #4**

**Instructions:** Complete the tasks outlined in this document. You should upload your completed assignment (with screenshots of your R code and output) online as a .pdf.

**Problem 1 : Cool Problems 2 : Electric Boogaloo**

Load the dataset “2cool4school.csv” into R. We used these data in our last lab assignment: an imaginary study of Berkeley and Stanford students who volunteered on their campus to answer questions about their love for coffee, and were then rated on how cool they were.

**1A. Data Cleaning (again)**

Last time, we did some data cleaning: (A) converted variable SEX to be a factor variable with three levels (1 = female; 2 = male; 3 = other), (B) changed the names of the levels of the factor variable SCH so Cal = Public and Stanf = Private, and (C) created a scale score for the variable coolness called coolness2. Copy and paste your code from the last lab to repeat these three steps (look at the key for the lab assignment if you want to make sure you are cleaning the data correctly). Use the write.csv() function to save this cleaned data object as “2cool4school\_clean.csv” so you don’t have to repeat these steps again.

**1B. Sex and Coolness**

Sex was measured as a three-level categorical factor (Female, Male, Other). Predict coolness from the variable sexF that you created in 1A. Then graph and interpret the results of this model, making sure to report the predicted values and 95% Confidence Intervals for each level, as well as the slopes, 95% Confidence Interval for the slope, t-test, and p-values.

**1C. Sex and Coolness, Again.**

Create a new variable called sexF2 that relevels the order of the levels in variable sexF where the baseline group is “other”. Then, predict coolness from this releveled variable. Graph the results of this model. What changed about the model output (compared to the model in 1B)?

**1D. Coffee and Coolness**

Love for coffee (coffee.love) was measured with a single item on a scale from 1 (Strongly Disagree) to 10 (Strongly Agree). Calculate the descriptive statistics for this variable. Then, build a linear model predicting coolness from coffee love. Graph (with a red line to illustrate the slope & intercept) and interpret the results of this model, as you would for a research paper (or your exam).

**1E. Coffee and Coolness (again)**

Now, test the same model in 1D, but z-score both the DV and IV. Show that the slope of this new model is equivalent to the correlation coefficient between coffee and coolness. How is the rest of this model output different from the output of 1D?

**1F. …by hand**

Show that R (and you) know that you are doing by calculating R2 and the F-test for your models in 1B and 1D by hand. (Remember you need to use the number of participants without missing data in order to get the exact same answer. See the key in the last lab assignment for help.)

**1G. …aaaaand power.**

Calculate the power for the estimated effects from your models in 1B and 1D. Yup.

**Problem 2 : Power Problems**

Hooray, more practice calculating the probability of power, Type I error, and Type II error for the following tests! After calculating these statistics, use these results to explain (with words) how power is influenced by sample size, variance, the size of the difference, and the directionality of the test.

* Population Mean = 80; Population SD = 20; Sample Mean = 71; Sample Size = 5; Directional Test
* Population Mean = 80; Population SD = 20; Sample Mean = 71; Sample Size = 5; Non-Directional Test
* Population Mean = 80; Population SD = 20; Sample Mean = 71; Sample Size = 15; Non-Directional Test
* Population Mean = 80; Population SD = 30; Sample Mean = 71; Sample Size = 15; ; Non-Directional Test
* Population Mean = 80; Population SD = 30; Sample Mean = 61; Sample Size = 15; ; Non-Directional Test
* Population Mean = 80; Population SD = 30; Sample Mean = 99; Sample Size = 15; ; Non-Directional Test

**Problem 3 : Midterm Two Problems Too**

For the second midterm, I’ll give you a dataset and a research question, and ask you to answer the research question (and interpret / graph the results) based on the data that I give you. Your answer should include the following components, organized with a label to make it easy for your GSI to grade.

* **Sample size and participant demographics:** report the key demographic variables available to you (what do you know about the number, age, and sex of participants?)
* **Descriptive statistics:** report the key descriptive statistics (Mean, Median, Standard Deviation, and Range) for all the variables that you will include in your model, and include graphs of the histograms. Comment on anything important about the data. If there are outliers, remove them (using R) and describe why you removed them.
* **Inferential statistics:** test the research question using a statistical model. Interpret (and graph) the results of this model as if you were reporting the results for a research paper. Make sure you include the predicted values of the estimates and 95% Confidence Intervals, as well as the slopes (and 95% Confidence Intervals, t-tests, p-values, effect sizes, and power for the slopes).

**Dataset:** fear\_study.csv (download from bCourses)

**Research Question:** Dr. Gomi Chung is interested in understanding what explains differences in people’s fear response to spiders. She holds two separate predictions:

* H1: people who are more anxious on average will be more fearful of spiders than people who are less anxious than average.
* H2 : people who are given an intranasal spray of adrenaline will be more fearful of spiders than people who are given no spray or a placebo spray.

Prepare a report to answer Dr. Chung’s question.

**Sample Size and Participant Demographics:**

**Descriptive Statistics:**

**Inferential Statistics:**