Psyc 3031, Winter 2018
Christopher D. Green
Assignment #2
Due: March 29, 7:00 pm

Include all R code and output, where relevant, as well as your own descriptions and conclusions.

1. Imagine a creativity test is given to 50 men and 50 women. Create these two data sets using the R function "rnorm". (Example: x <- rnorm(30, 100, 15) will give you 30 normally distributed scores with a mean of approximately 100 and a sd of approximately 15). For the men, the mean on the creativity test should be (approximately) 50 and the sd (approximately) 10. For the women, the mean should be (approximately) 54 and the sd (approximately) 20. Use the R function "describe" (in the "psych" package) on each of the two datasets, once you have created them, to ensure that your means are close to where they should be. If the men's mean is off by more than 1 full point, or the women's mean is off by more than 2 full points, create the dataset anew until you get a mean between 49 and 51 for the men, and a mean between 52 and 56 for the women (yes, the women's range is wider because they have a higher sd).

Now, do an independent groups *t* test using the R function "t.test". Report the result and describe your conclusion in ordinary words.

Note that the variances in your two groups are quite different (~100 vs. ~400), but also note that t.test automatically did the Welch adjustment of the df to account for the heterogeneity of variances. What should the nominal df be? What did the Welch adjust them to?

Now, re-run the t-test but this time turn off the Welch adjustment by including "var.equal=TRUE" in the argument to your t.test function. What effect did this have on your p-value? How much did it move, and what direction? Why?

2. Use "rnorm" to create a new creativity test dataset for the women: 50 data points again, but now with a mean of 53 and an sd of 10 (i.e., equal to the men's sd). Use "describe" to make sure the mean of this new dataset is between 52 and 54. Re-do it, if necessary, until you get a dataset with a mean in this range. Do a *t* test (with the Welch correction) between your original men's data and this new women's data. Describe the statistical outcome and your substantive (English) conclusion.

Now change the mean of the women's data to 54 (make sure it's between 53 and 55), but leave everything else the same. Do a *t* test (with Welch) between your original men's data and this new women's data. Describe the outcome. How much of a change do you see in the p-value from the last *t* test?

Change the women's mean to 55 (make sure it's between 54 and 56). Do the *t* test again. Describe the outcome. How much of a change do you see in the p-value from the earlier *t* tests?

Why do you think the p-value responded in this way to just a one-point change in the mean of one group?

3. Compute Cohen's d for each of the three tests in Question #2 (assume that your means really were exactly 50 for men, and exactly 53, 54, 55 for women, and that your sd was exactly 10). Now use the R function pwr.t.test (in the "pwr" package) to compute the power for each of those three t-tests. How much did power respond to these single point changes in the mean? Why?

4. Using "rnorm" again, create 20 new datasets for the women's scores, all with 50 scores each, mean = 54, and sd = 10. (This time, do ***not*** ensure that the mean is close to 54; just keep whatever rnorm gives you each time.)

Now, run 20 *t* tests with each of these 20 datasets against the original men's dataset. How many of these tests came up "significant" and how many did not? Create a barplot or histogram of the p-values, using the following ranges: $<.01$, .01-.05, .05-.10, $>.10$. Thinking about Geoff Cumming's video, "Dance of the p-values" (rewatch it if you like), why do you think you got this result?

5. Imagine you wanted to know whether university faculty are more politically conservative than university students. You drew samples of each and gave them a questionnaire that measures political orientation. You ran an independent groups *t*-test on your data, and you got a highly significant result ($p = .01$) and a Cohen's d of about 0.5. Then, you thought it might be a good idea to re-analyze the data using a Bayesian procedure. You found that the Bayes Factor was about 6.5. First, what is the proper interpretation of that Bayes Factor (i.e., what does it mean)? Second, does the Bayes Factor change the way you view the decisiveness of the *p*-value and, if so, in what way? (Hint: Consider Jeffreys' rule of thumb about Bayes Factors, given in the JASP tutorial that was posted on course's Moodle site.) Third, using the Held nomogram, what would your estimate of the posterior probability of the null hypothesis be after running this experiment?

6. Use "rcorr" (in the "Hmisc" package) to run a correlation between neuroticism and openness in the NeopiIQ dataset. (Unlike "cor" in the Base R package, "rcorr" gives you a significance test as well as the correlation coefficient.) Describe the statistical outcome and your substantive conclusion.

Now, using "cor", produce a matrix of all correlations among the variables in the NeopiIQ dataset. (You will have to remove the "sex" column from the dataset for this to work. To do this. To do this, use NoSexNeopiIQ <- NeopiIQ[,-1], which removes Column 1 from the data.) Examine the correlation matrix closely. Why might it lead you to doubt the true significance of the correlation between neuroticism and openness that you found in the first part of this problem?