Statistical Methods in Education Research
Vanderbilt University
Prof. Sean P. Corcoran

Due November 2, 2023
34 total points

Problem Set 8

Instructions: Answer the following questions in their entirety in a separate document. Submit your completed problem set as a PDF via email to `sean.corcoran@vanderbilt.edu`. Use your last name and problem set number as the filename. Working together is encouraged, but it is expected that all submitted work be that of the individual student.

---

1. Data are gathered on 16 students. Half of the students are randomly assigned to a new tutoring program and half have their usual schooling experience. A study finds that test scores for the tutoring program students are 10 points higher on average than those for the other students (*p=0.3*). The mean test score overall is 200 with a standard deviation of 40. (**6 points**)

   (a) Is the study's finding statistically significant? Explain why or why not.

   (b) Is the study's finding practically significant, in your opinion? Explain what practical significance means here.

   (c) Would you conclude that the tutoring program is effective based on these results? Ineffective? Briefly explain.

2. Two studies were commissioned to evaluate an intensive program designed to enhance social and emotional learning (SEL) among adolescents. The index used to measure SEL has a mean of 50 and a standard deviation of 10. Study failed to find a statistically significant improvement in SEL, with a 95% confidence interval for the gain in SEL of $(-7, 17)$. Study 2 also failed to find a statistically significant improvement in SEL, with a 95% confidence interval of $(-2.5, 1.5)$. Which of these two studies (if any) is more valuable to a policymaker, in your opinion, and why? (**4 points**)

3. Simulate the multiple hypothesis testing problem in Stata by doing the following steps: (**24 points**)

   (a) Begin by creating a dataset with 1,000 observations. Create a variable called *tstat1* consisting of random draws from the $N(0, 1)$ distribution. Think of *tstat1* as test statistics from a sampling distribution in which H_0 is true. Create a second variable called *pval1* that contains $Pr(z > tstat1)$. (That is, the probability of drawing a test statistic at least as large as *tstat1*). Create a third variable called *reject1* that equals 1 if $pval1 < 0.05$ and 0 otherwise. In what proportion of "samples" does $reject1 = 1$ (a Type I error)? (**4 points**)

(b) Now create four additional variables (*tstat2, tstat3, tstat4,* and *tstat5*) defined in the same way as *tstat1*. Create analogous variables *pval2-pval5* and *reject2-reject5* based on *tstat2-tstat5*. Think of each row in your dataset as 5 independent hypothesis tests conducted using *tstat1-tstat5* in the same sample (where H_0 is true in all cases). In what proportion of "samples" does $reject = 1$ for 1 or more hypothesis tests? (**4 points**)

(c) Use the family-wise error rate formula shown in class to calculate the proportion of samples that one would expect to reject H_0 in one or more of these hypothesis tests. How does this compare to your answer in (b)? (**4 points**)

(d) Create five new *reject* variables that use the Bonferroni-corrected thresholds for rejection. Based on these new variables, in what proportion of "samples" does $reject = 1$ for 1 or more hypothesis tests? (**4 points**)

(e) Use the following syntax to create five new *tstat* variables. These test statistics are still random draws from the N(0,1) distribution, but they are now dependent. That is, when one *tstat* is higher than average, the others tend to be higher than average (and when one *tstat* is lower than average, the other tend to be lower than average).

```
clear
matrix c=(1, 0.8, 1, 0.8, 0.8, 1, 0.8, 0.8, 0.8, 1, 0.8, 0.8, 0.8, 0.8, 1)
drawnorm tstat1 tstat2 tstat3 tstat4 tstat5, n(1000) corr(c) cstorage(lower)
```

As in parts (a)-(b), create *pval* and *reject* variables based on these new test statistics. In what proportion of "samples" does $reject = 1$ for 1 or more hypothesis tests? Is this proportion higher or lower than that in (b)? What is your intuition about why these differ from b (if they do)? (**4 points**)

(f) If you apply the Bonferroni correction to the variables created in part (e), how often will you reject 1 or more of these tests? (**4 points**)