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

Due October 3, 2024
40 total points

Problem Set 5

Instructions: Answer the following questions in their entirety in a separate document. Submit your completed problem set as a PDF document 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. (**2 points**) When a recent survey asked whether the government should impose strict laws to make industry do less damage to the environment, a 95% confidence interval for the population proportion responding *yes* was (0.87, 0.90). Would a 99% confidence interval be wider or narrower? Carefully explain why.

2. (**4 points**) An education school wants to estimate the mean annual salaries of the school's former students 5 years after graduation. A random sample of 25 such graduates found a sample mean of $51,288 and a sample standard deviation of $5,736. Assuming that the population distribution is normal, find a 90% confidence interval for the population mean. Show your work.

3. (**6 points**) A graduate school admissions officer has determined that historically, applicants have undergraduate grade point averages that are normally distributed with standard deviation 0.45. From a random sample of 25 applications for the current year, the sample mean grade point average is 3.20.

   (a) Find a 95% confidence interval for the population mean.

   (b) Based on this sample, a statistician computes a confidence interval for the population mean extending from 3.06 to 3.34. What is the *confidence level* associated with this interval? Show your work or explain how you found your answer.

4. (**3 points**) Find the t value that would be multiplied by the standard error to form a:

   (a) 95% confidence interval with a sample size of 5

   (b) 95% confidence interval with a sample size of 15

   (c) 95% confidence interval with a sample size of 25

   (d) 95% confidence interval with degrees of freedom of 25

(e) 99% confidence interval with degrees of freedom of 25

(f) 90% confidence interval with a sample size of 500

5. (**6 points**) An estimate is needed of the mean travel time from home to school in a large urban school district. The estimate should be correct to within 2 minutes in 95% of random samples. A previous study of school commuting time suggests that 15 minutes is a reasonable approximation for the standard deviation (σ) of commuting time.

   (a) How large of a sample of families is needed to meet this requirement?

   (b) A random sample is selected of the size you reported in part (a). The sample has a standard deviation of 9 minutes, rather than 15. What is the margin of error for a 95% confidence interval for the mean travel time to school?

6. (**6 points**) A question in the General Social Survey asks whether respondents agree or disagree with the following statement: "It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family." The sample proportion agreeing was 0.36 in 2004 (n=883).

   (a) Show that the estimated standard error for the sample proportion was 0.016.

   (b) Show that the margin of error for a 95% confidence interval was 0.03.

   (c) Construct the 95% confidence interval and interpret it.

7. (**8 points**) For this question, use the Stata dataset *Grade4_classrooms.dta* on Github. This file represents a random sample of 4th grade classrooms located in urban school districts in Texas. Each observation is a classroom, and the variables either describe the teacher (e.g., total teaching experience, teacher race/ethnicity), or are an average for the classroom (e.g., reading z-scores, % of students who are economically disadvantaged). All data are from 2006.

```
use https://github.com/spcorcor18/LPO-8800/raw/main/data/Grade4_classrooms.dta, clear
```

   (a) Inspect the histograms for the variables *totexp* (the teacher's total years of experience), *maplus* (an indicator of whether the teacher holds a master's degree or higher), and *readz_class* (the mean z-score in reading for the students in the class). How would you describe the shape of these distributions? Note: the mean z-score is based on students' position in the statewide distribution of test-takers.

(b) Provide 90% confidence intervals for the population means of each variables listed in part (a), and interpret each in words. Note that *maplus* is a binary (dichotomous) variable.

(c) For purposes of constructing confidence intervals, are you concerned about the normality (or lack thereof) in the distributions viewed in part (a)? Explain why or why not.

(d) Consider your confidence interval for reading z-scores. Is it consistent with the hypothesis that students in the average urban school district classroom perform about the same as the statewide average in reading? Explain your reasoning.

8. (**5 points**) In Lecture 5 you learned how to construct a confidence interval for the population *mean*, but not for the population *median*. There are methods for doing this, but an alternative approach would be to bootstrap. Continue using the *Grade4_classrooms.dta* dataset from #7. Use the `bootstrap` prefix to draw 1,000 samples with replacement from your data and capture the median (p50) teacher experience (*totexp*) on each draw. The sample size should be the same as your dataset. Save these results to a separate file. Using your saved results, between what two values does the median fall 95% of the time? The `bootstrap` results report a "normal-based" 95% confidence interval. How does this compare to your first answer? Does a "normal-based" confidence interval make sense to use here? Why or why not?