SOC 4015/5050: Lab-06 - Foundations for Inference

Christopher Prener, Ph.D.

Fall 2018

Directions

Please complete all steps below. Your your work "by hand" should be scanned and included in your `Lab-05` assignment submission. All work should be uploaded to your GitHub assignment repository by 4:15pm on Monday, October 8th, 2018.

Use the following scenario: You are working with census data of Saint Louis University students. Assume that these data are a complete representation of the student body and this represent a *population*. The student body average grade percentage is a 90 with a standard deviation of 10.

Analysis Development: Create a Project Folder System

1. Using RStudio, add an R Project to the *existing* directory in your assignments repository named `Lab-05`.

2. Add a new folder named `docs` to you project.

3. Create a new text file for your `README.md`. In the body of your `README.md` file, use Markdown formatting to write a sentence or two describing the purpose of this project. Then create an outline using bullets of the contents of the project itself.[1]

4. Create a new notebook with an expanded YAML heading.

5. Make sure your notebook has *completed* introductory, package loading, and data loading sections before proceeding with the parts below.

6. Be sure to "knit" your notebook at the end of the assignment!

This initial section follows the project workflow that is available in the `lecture-03` repo!

[1] See my write-up of the Markdown syntax in *Sociospatial Data Science* for details on creating lists.

Part 1: Sampling Distributions

7. Assume you draw repeated random samples of *n*=500 students. What is the standard error of these repeated samples?

8. Assume you draw repeated random samples of *n*=1000 students. What is the standard error of these repeated samples?

9. If you were to draw repeated random samples of n=500 students, what proportion of these samples will have sample means greater-than or equal to 95?

10. If you were to draw repeated random samples of n=500 students, what proportion of these samples will have sample means less-than or equal to 80?

11. What sample size would we need to have a sample mean that is within 1 percentage point of the population's?

12. What sample size would we need to have a sample mean that is within 3 percentage points of the population's?

13. What sample size would we need to have a sample mean that is within 5 percentage points of the population's?

Part 2: Predictive Intervals

14. Calculate and interpret a 95% predictive interval for x.

15. Calculate and interpret a 99.9% predictive interval for x.

16. Calculate and interpret a 95% predictive interval for $\bar{x}$ with a random sample of n=500 students.

17. Calculate and interpret a 99% predictive interval for $\bar{x}$ with a random sample of n=500 students.

Part 3: Confidence Intervals

18. Calculate and interpret a 95% confidence interval assuming we draw a random sample of n=500 students.

19. Calculate and interpret a 99% confidence interval assuming we draw a random sample of n=500 students.

20. Calculate and interpret a 95% confidence interval assuming we draw a random sample of n=1000 students.

21. Calculate and interpret a 99% confidence interval assuming we draw a random sample of n=1000 students.