

STAT 3201 Final Project

Due Monday, December 3rd at the beginning of class

All aspects of this project must be completed independently; you may not discuss the problems with any other than the instructor. In addition, you are responsible for ensuring that others do not have access to your project. Any violation of these rules, including but not limited to electronically receiving or transmitting of any work, is a violation of the OSU Student Code of Conduct and will be reported to the Committee on Academic Misconduct.

Directions:

This project is designed to introduce you to the idea of bootstrapping/resampling. This is a method you will explore in more detail in STAT 3202. There are 3 parts to this project.

Part 1:

Each of the distributions below could be used to model the time spent studying for an exam. Take 1,000 random samples of size 25 from each of the distributions below. In each case (a,b,c), plot the empirical distribution of the sample mean, estimate the mean of the sample mean, and estimate the standard deviation of the sample mean. Compare the results to the theoretical results.

- a. $N(5, 1.5^2)$
- b. $Unif(0, 10)$
- c. $Gamma(5, 1)$

Part 2:

Each of the distributions below could be used to model the time spent studying for an exam. Take one random sample of size 25 from each of the distributions below. Then, take 1,000 resamples (i.e., sample with replacement) of size 25 from your sample. In each case (a,b,c), plot the empirical distribution of the sample mean, estimate the mean of the sample mean, and estimate the standard deviation of the sample mean. Compare the results to the theoretical results. In addition, compare your results from part 2 to the corresponding results from part 1.

- a. $N(5, 1.5^2)$
- b. $Unif(0, 10)$
- c. $Gamma(5, 1)$

Part 3:

The Project Data 1 and Project Data 2 files on Carmen contain a data set of 25 and 250, respectively, randomly chosen study times. For each data set, use the resampling technique from part two to obtain the empirical distribution of the sample mean. Plot the empirical distribution of the sample mean, estimate the mean of the sample mean, and estimate the standard deviation of the sample mean. In addition, find the estimated 5 number summary and the estimated 2.5th and 97.5th percentiles. Compare and contrast your results for the two sample sizes.

The written report should be roughly 1 to 2 pages (excluding figures/plots) and should clearly explain your results. It should be well organized and grammatically correct. Please use a readable 11 or 12 point font. If need be, include a works cited or reference page. In addition, include your R code as an appendix (use either Courier New or Consolas font for the R code). Print your report double sided and staple it together.