

Econ 184b, Econometrics, Assignment 3

The raw .rmd file for this assignment are available at: https://raw.githubusercontent.com/flkidd/Econ184/main/Assignment/Assignment_3.Rmd

Note: For the math problems, you can either (1) type your solution and compile it with RMarkdown; (2) type your solution use another word-processing software and convert it into a PDF file; (3) write your solution on paper, scan it and make it into a legible PDF file. TAs can decide, at their discretion, that the submitted file is illegible and thus give zero credit to a question or the entire problem set. If you type your solutions (options 1 and 2 above), you will get **a bonus equal to 5% of your performance** on the problem set. No credit will be given if you only report the final answers without showing intermediate steps whenever appropriate.

For the programming problems, you need to submit both the compiled pdf/html file and the RMarkdown (.rmd) file on LATTE. You will get **an additional bonus equal to 5% of your performance** if your RMarkdown file is completely reproducible with minimal alteration (install packages, etc.). That is, our team (or anyone else) should be able to recompile your Rmarkdown file and reach *the same* result. You need to explicitly write out your answers - just showing programming outputs will receive zero credit.

You will be receiving a total of 15% bonus if you submit one single pdf/html compiled directly from Rmarkdown, along with the .rmd file.

Question 1:

You want to study whether or not academic achievement is related among different subjects. For example, do students who perform well on reading also tend to perform well on math? To answer this question, you collected data on 420 school districts in California. The sample is divided into two groups:

* High-reading group is defined as school districts with reading score above 655. Let μ_1 denote the population mean of math scores in this group. In the sample, this group consists of 259 school districts. For this group, the sample average of the math score is 664.11 and the standard deviation of the math score is 14.28. * Low-reading group is defined as school districts with reading score below 655. Let μ_2 denote the population mean of math scores in this group. In the sample, this group consists of 161 school districts. For this group, the sample average of the math score is 636.02 and the standard deviation of the math score is 10.13.

- construct a 95% confidence intervals for $\mu_1 - \mu_2$
- can we conclude that the two groups (defined by the reading score) have different math scores? Use 1% significance level. Please write out the null and alternative hypothesis, and use both the t-test and the p-value approach to test the hypothesis.