

Empirical Exercises

You must choose one of four empirical exercises to complete this semester. The exercises are listed and described in more detail below (or on our class website):

- [Exercise 1: Difference-in-differences](#)
- [Exercise 2: Instrumental variables](#)
- [Exercise 3: Regression Discontinuity](#)
- [Exercise 4: Hospital markets + demand estimation](#)

Please submit your answers as PDF on GradeScope (via Canvas) no later than **11:59pm on April 26**. In your PDF, please include a link to your GitHub repository. Be sure to also include in your repository all of your supporting documentation, including your code files (Stata, R, Python, SAS, etc.), all figures/tables, and some instructions (e.g., as part of your ReadMe file) that introduce the reader to your data and the sequence in which your code should be run. Practice writing good code and showing me only what I would need to recreate your results.

The exercise is worth 30 points toward your final grade, with 20 points allocated to the accuracy of your work (2 points for each of 10 questions) and 10 points allocated to the replicability of your work (i.e., can I easily recreate your answers from your repository?)

In your project, please be sure to organize your folders in a useful way. The way I organize things (though certainly not the only way to do it) is to keep a folder for each new project and named accordingly. I typically have the following subfolders:

- *Data*: This is where I keep the raw data files and any additional data files I create as part of my analysis. I also keep a “Research Data” folder on my computer that has raw data that I access regularly, in which case the “Data” folder in any given project includes symbolic links to the original raw data. I usually split this folder into two: “input” for raw data; and “output” for final or intermediate datasets.
- *Data Code*: This is where I keep the code for data management (merging, clean-up, wrangling, etc.). Final analytic datasets created by this code go to the “Data/output” subdirectory.
- *Analysis*: This is where I keep my analysis code files and log files (if relevant).

- *Results*: Output from the analysis code files. I tend to separate this into two subfolders - one for tables and one for figures.
- *Papers*: If I'm writing the paper in **R Markdown** or **Quarto**, this is where I keep all of the markdown files. With most co-authored work, though, we write in Overleaf. In that case, I'll still download different versions of the paper into this folder when necessary (e.g., for different journal submissions)
- *Presentation*: This is where I keep my slides and underlying code files.

It's good to start developing some organization practices that work best for you. It's extremely easy to forget what you were doing on a project once you have several things going at once, especially when you wait for 6-8 months after submitting a paper for publication. The last thing you want is to not be able to replicate your own work!