503 Final Paper Guidelines

Please send us a one-page outline or introduction to your final project by **Monday, March 23rd**. We will give you feedback on these projects by **Friday, March 27th**.

If you already know what your project is, please send the outline/introduction earlier rather than later.

Your final papers are due on **May 6th**. The last week of class will be dedicated to presenting these projects.

We recommend that your project take one of two forms. If your idea does not currently take these forms, please get in touch with Peter, Angele, or Cleo ASAP. More generally, the text below is designed to comprise guidelines not rules.

1. **Replication paper (approximately 8-10pp, single-spaced):** Identify a paper to replicate. We expect you to go over:
    a. Research question and literature review: What are the authors trying to answer? How do they situate themselves in the literature?
    b. Assumptions: what assumptions do the authors make in their identification strategy? Are these assumptions valid, particularly as it relates to the data and setting at hand? Why or why not? Can you relax these assumptions, or assess their plausibility? Bounds, sensitivity analysis, or modification of the approach are all feasible.
    c. Estimation strategy: given the authors' identification strategy, are the statistical methods used by the authors appropriate? If not, propose alternative approaches. E.g., if the original paper used OLS after invoking ignorability, you might want to try the DR estimator.
    d. Replicate findings: choose one of the below options.
        i. *Paper's data:* using the paper's replication files, examine whether the findings stay the same or differ using different identification strategies and assumptions, and sometimes-unstated data cleaning decisions. For example: how does the way they impute missing values affect their results? Break down their code and try to figure out (i) what they're doing and (ii) why you might want to change some of their decisions.
        ii. *Original data:* proposing the use of new data, see if the findings replicate in a different context. Why might they replicate? Why not? Theorize what might be different with a new set of data. For example, if the original project were a survey experiment, you might propose replicating the experiment on a new sample with different treatments or outcome measure.

2. **Original project or research design proposal (approximately 8-10pp, single-spaced):**
    a. Research question (ideally with some literature review to give us some context as to where the question came from).
    b. Proposed research design and identification strategy to answer that question. Discuss the assumptions you need for your identification strategy to be valid, and discuss tests you could perform to back up your arguments.
    c. Discuss data availability: do you need to collect data? Build new variables? Deal with missing data? Discuss the viability of your approach given data constraints.
    d. Data analysis (if possible). If data is available, you should conduct a preliminary analysis.
    e. Power analysis (if data not possible). If data is not available, you should conduct a power analysis with either purely simulated data, or with real data from another source (so that all of the outcomes are Y_i(0)) and simulated treatment effects.