POLS 641: Introductory Political Analysis
Prof. Kuehl - Fall 2023

Final Project

Overview The aim of your final project is to complete a shortened version of a journal article with a heavy emphasis on the methodology and results section. You will complete minimal versions of introduction, literature review and theory sections, but will complete full versions of methodology, description of data, and analysis of results. The focus will be on your ability to complete your analysis using the R programming language and to communicate the results of your analysis to a professional audience. Your are not expected to conduct novel research nor make a clear theoretical contribution, but you are expected to uphold political science standards of quantitative analysis. You will also present your results to the class during the final week. This project is worth 50% of your final grade.

Final Paper Due: December 13th

Grade Breakdown 50% of Course grade

1. Research Question(Oct 13th)- 5 pts
2. Abstract and Key Descriptive Statistics (Nov 10th) - 5 pts
3. Class Presentation (Dec 5th and Dec 7th) - 5 pts
4. Final Paper(Dec. 13th) - 35 pts

Component Tasks

1. **Research Questions** due October 13th

   Finding an interesting and realistic research question is often the hardest part of the research process. To that end, you will develop a basic research question and proposal for **two** potential projects. Each proposal will ask you to answer the following questions:

   - What is the research question?
   - What is your dependent variable? How can you operationalize it?
   - What is your independent variable? How can you operationalize it?
   - What is your hypothesis?
   - What are two articles/books that are related to this question? (provide APA citations)
   - Where will you get the data?(See below for data ideas)

   Your responses will be on Blackboard. It is up to you to decide which of these projects, or an alternative, you will pursue for your final paper.

2. **Abstract and Initial Data** due November 10th

   The next step will be to submit a 200 word abstract for your intended project and a knitted Rmarkdown file with descriptive statistics for your key variables. At that point you should have a clear idea of your project and most, if not all, of the necessary data collected. The abstract should be professionaly written as if you were going to submit it to a conference.

   The idea here is to show me that you have truly "gotten to know" your data. At the minimum your Rmarkdown file should include:

   - For your primary independent variable and dependent variables:
     - *Description* Written narrative of how the variables are operationalized. If your using a non-intuitive measure please describe in detail how it is calculated. It should be clear how your measure captures the concept(ie valid estimate of the estimand)
     - *Descriptive Statistics* Range, Mean, Median, Standard Deviation
     - *Distributions* Histogram and/or Box plots
   - *Bivariate Relationship* Scatter plot of DV and primary IV
   - *Control Variables* Basic descriptives as necessary.

   Think of this as the document you would send to a co-author or supervisor giving a broad overview of the quantitative portion of your paper. It does not have to be super polished, but should be clean enough that the concepts are clear to a non-expert.

3. **Presentation**Currently scheduled for December 5th and December 7th

   All students will give a 10 minute powerpoint style presentation of their results. This should reflect the basic components of the paper as discussed. Peers will then ask questions and provide feedback that will improve the final paper. Think of this as a practice mini-conference. You will be expected to integrate feedback into your final paper.

4. **Final Paper Submission** Due Friday, 13 December by Midnight on Blackboard.

   Please turn in the 1) final document(in your chosen format, 2) the R code (LaTeXcode as well if used) and 3) the final clean dataset. Late assignments will not be accepted without prior approval of the instructor.

Topic and Data With some limits the choice of topic(and dataset) is yours. The topic and data need to be suitable for bivariate regression analysis(ie both the dependent and independent variables need to be continuous) and dataset must contain other variables(either binary or continuous) that can be used as control variables. You should plan on using an existing dataset. You will not have the time to build a dataset of adequate size in our limited time period this semester. However, you may want to combine pre-existing datasets(we will learn how to merge in R). Ideally, this work will contribute to a future research project and further familiarize you with meaningful quantitative work in your chosen field. I encourage you to be strategic in your topic selection. Below are some suggestions to get your started:

Data sources:

- **Replication Data** Find a quantitative article you like and see if you can find the the data. It is increasingly common for authors to post their data on their personal websites, journal websites, or in data repositories(like ICPSR below or Harvard's dataverse). If not publicly available, many scholars will likely be happy to help given a polite email.
- **ICSPSR** www.icpsr.umich.edu Central collection of political science data.
- **Organizational Datasets** - World Bank, UN, WTO, etc and their subsidiaries provide reams of quality data.
- **Political Science Classics** - Correlates of War, National Elections Survey, World Values Survey, General Social Survey, and many more

I'm happy to help in topic selection and **highly encourage** you to check in with me on your dataset choice and hypothesis. Do not underestimate the time required to gather and clean your data.

Organization The final paper will be organized much like a standard journal article. You are expected to complete the following sections. Order and section titles may be changed to fit your analysis, but all components must be present. Your final presentation will include roughly the same components.

1. Introduction

   Engage the reader and convince them of the importance of your research. Think carefully about how you frame your research question. The intro should include a clear research question and should include an overview of the rest of the paper.

2. Literature Review

   Provide a brief introduction to the state of knowledge on your given issue. It should include the citations of at least 5 academic sources(one page). It does need to be comprehensive for this assignment, just a brief overview of where your research question fits within the literature. Make sure your literature review is well organized and that you use proper citation standards.

3. Theory and Hypothesis

   Use the literature and your own knowledge to provide a reasonable explanation why you expect to see the hypothesized relationship. Although you are not going to make causal claim(unless you have experimental data) think about this as explaining the plausable causal mechanisms. You do not need to fully immerse yourself in the literature, but do use some of the literature to make your argument. At the end of this section lay out your hypothesis or hypotheses formally (i.e. *H1:*).

4. Methodology

   Provide an overview of the the methodological approach and the choices made. Clearly articulate the reasons why you chose a particular procedure or technique given the limited background from the class.

5. Data

   Describe the dataset you are using. I.e. source, how collected, timing, structure, anything that is relevant. If combining datasets discuss that here. In addition, this is where you describe the measures relied on and the coding scheme. Convince the reader that the measures used are reasonable operationalizations of the abstract concept you are interested in. This section is your opportunity to convince reviewers/scholars that the data was collected or generated in a way that is consistent with accepted practice in the field of study. Be sure to cite where appropriate.

6. Descriptive Statistics

   Use tables, graphs, and text to describe your key variables. Be creative, but comprehensive. All tables and figures should be professional in appearance, be labeled(table 1 and figure 2 etc.), and have descriptive titles. Be creative and use your curiosity.

   (a) Dependent Variable
   (b) Primary Independent Variable
   (c) Control Variables

7. Model

   Describe your model choice and use mathematical script to lay out your bivariate and multivariate models(based on our core model). ie $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

8. Results

   Walk the reader through your findings. Regression output should be printed out in table form. Coefficient plots, regression plots, and predicted probabilities are also encouraged. Discuss substantive as well as statistical impacts.

9. Discussion

   Discuss your results in a broader context. What do your results mean? Also acknowledge limitations in your approach. I.e. problematic assumptions, data and measurment issues, generalizability problems, etc. Please be extensive in this section. What night an improved research design would look like? Why might a cross-sectional linear model be limited? What might a better model(that you haven't learned yet look like)? What would be multiple method project?This is also an opportunity to point out future research questions. This is a chance to really show your understanding of this class and your research project.

10. Conclusion Provide a brief summary of your contribution and remind the reader why the research project is so important.

Formatting In all aspects, including appearance, this project resemble an article you would be sending for possible publication. Accordingly, it should look presentable with professional looking tables etc. An Rmarkdown created html will not suffice. Feel free to use Word or your preferred documentation software. A couple bonus points will be awarded to those who prepare their project using LaTeX, however we will spend minimal class time practicing LaTeX. For formatting and citation questions I recommend consulting the APSA Style Manual

You will also will be expected to turn in the relevant R code and your data. Your code should be replicable and should include notes on what your code does and be in a similar order to how it is reported. The code should be clean. I.e. practice code or lines of code that didn't work should be eliminated. You should be able to return to this code years from now and be able to re-use it. That means including notes and annotations that explain what you are doing.

Suggested Slides for Presentation

1. Title Page
2. Research Puzzle
3. Existing Literature
4. Hypotheses
5. Descriptive Statistics
6. Primary Regression Results
7. Robustness Checks
8. Contributions and Next Steps
9. Conclusion and Thank you

Authorship and Plagiarism You may use your notes, book, and any other non-human (printed or web-based) resource to complete this final project. However, **you may not work together.** I will gladly help with general tasks such as topic and hypothesis selection, but will not provide help on specifics. I will not help with specific coding tasks, particularly those that were covered in class or on problem sets. If you are really stuck reach out, but the expectation is that datacamp and problem sets have prepared you to complete the code.

Combining this with another research project from another class is OK, but you need to clear it with me, and the other instructor if concurrent, before you start. Working on the same topic can be a smart long-term strategy, but don't expect a meaninfful reduction in the work you'll need to do.

Unsolicited Advice

1. **Start early!!!**
2. Take problem sets seriously, take good notes, and do your own work
3. Don't procrastinate. Go above and beyond on each component task
4. Get comfortable using online Resources
5. Build in extra time for Rguments
6. Be cuRious - take the time to teach yourself something new in R
7. Give yourself time to edit and revise
8. Channel your inner social scientist and have fun