Regression Analysis for Political Science II

true

January 18, 2023

Jonathan Cervas
Email: cervas@cmu.edu
Location: POS 147 Posner Hall
Time: Wednesdays 7:00p-9:50p Eastern
Office Hours by appointment (arrange via email)
CMU Academic Calendar

THIS SYLLABUS IS A WORK IN PROGRESS. CHANGES ARE LIKELY TO AFFECT THE SCHEDULE AFTER WEEK 1

"*One can learn data analysis only by doing, not by reading*" - Kosuke Imai

Summary

How do we evaluate empirical claims about events we see around the world? Can we measure discrimination in job hiring? What is the best way to predict election outcomes? What factors drive the onset of civil wars? Is it possible to determine what members of Congress are more or less liberal given their voting record? These are just a few of the numerous question that social scientists are tackling with quantitative data. Beyond academia, companies and non-profits have invested heavily in data science techniques to learn about their users, platforms, and programs. Data scientists at these institutions are essentially applied social scientists and employ many of the same techniques you will learn in this course. The goal is to provide students with the foundation necessary to analyze data in their own research and to become critical consumers of statistical claims made in the news media, in policy reports, and in academic research.

Course Description

This course is required for all CMU MS-IRP students. It is the second half of a two-part sequence on methods. The work in this course reinforces that of RAPS I, and will move beyond the basics into more advance technics for understanding data. In addition to Ordinary Least Squares regression, we will consider models for different types of dependent variables (logit, probit, etc). We will also address missing data problems, text as data, spatial data, experimental designs, regression discontinuity designs, simulations, and causality more generally. We will also spend some time on prediction models and probability. Some of the material for this course will inevitably be repetitive of RAPS I, but repeated exposure can lead to a deeper understanding of complex methods.

Objectives

A major objective of this course is to prepare you to complete your required thesis. By the end of the course, you should have developed a toolkit of methods that will allow you to test empirical questions underlying your hypotheses. Moreover, this class should help you develop skills that can be transferable to your future careers as data analysts.

Prerequisite Knowledge

I assume that everyone enrolled in this course has satisfactorily completed RAPS I. I also assume a basic understanding of the R programming language.

Book

(**QSS**) Imai, Kosuke. 2018. Quantitative Social Science: An Introduction. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691167039/quantitative-social-science

We may also use various material found on the internet. I offer some additional resources you may want to explore on your own if you want more advanced or specific methods.

Material

Our primary text will be Quantitative Social Science: An Introduction by Kosuke Imai. Imai's book comes in three versions: "In Introduction", "An Introduction in tidyverse", and "An Introduction in Stata"; you may use any version best suited for your background. I will upload (or link) all assignments and Problem Sets to Canvas, but you can also download all the files as a ZIP at https://press.princeton.edu/student-resources/quantitative-social-science

Unfortunately, I can not find QSS online, so you likely will have to buy a copy. You can use this link to access a pre-release version of the book. I highly recommend purchasing a new copy, which has subsequently been updated. Also, the Hunt Library has placed the book on reserve for our use this semester.

Grading

I care the most about what you learn, not what numerical/letter summary of that learning you get at the end of the semester. So I would love to not have grades at all, but the university does not agree. Thus, we have grades to help encourage you to put effort into learning the course material, and to satisfy the university overlords.

Assessment

The course grade will be a weighted average of the following components:

Category	Percent of Final Grade
Participation	15%
Problem Sets (~10)	40%
Exam Problem Sets (2)	10%
Final Project	35%

Swirl

Upon completion of each chapter, users should try the review questions before attempting to solve exercises that appear at the end of each chapter. These review questions are available as swirl lessons at https://github.com/kosukeimai/qss-swirl and can be answered within R.

To start the review questions, users must first install the swirl package (see Section 1.3.7) and then the lessons for this book using the following three lines of commands. Note that this installation needs to be done only once at the beginning.

```
install.packages("swirl", repos='http://cran.us.r-project.org') # install the package
library(swirl) # load the package
install_course_github("kosukeimai", "qss-swirl") # install the course
```

Problem Sets

To only read about data science is about as useful as reading the entire DMV handbook and memorizing all state driving laws, and then to show up to a Formula 1 race expecting to win. You need to drive.

Thus, in this course, you will have problem sets to complete throughout the semester that will give you an opportunity to apply the statistical techniques you are learning. They will usually be focused on data analysis in general and will often involve a real dataset. I encourage students to rely on peer working groups as they work on these questions, but each student will submit their own work individually. We will have class time to work on these. These problems can be found the the end of each chapter in QSS.

Exam Problem Sets

Similar to the problem sets that we will work on as a class, you will also on two occasions be assigned problem sets to work on individually.

Final Project

Present original research findings, either by extending a paper found in a social science journal OR by writing a research paper on another topic. Students may work individually, or can work in groups of up to 3 students. There are three prototypes of projects you can choose between (i.e., choose just one of these):

1) A replication of a published social science paper accompanied by an extension (e.g. with a different estimation approach or a different quantity of interest).

2) An analysis of a new dataset that asks a similar or related question as a published paper in a social science journal.

3) An analysis of an entirely different social science question using similar quantitative research designs, upon permission of the instructor. This is a good option if you have some ideas for your thesis and want to get a head-start.

Proposal	Wed, April 5th 11:59pm ET
Draft Analyses	Wed, April 19th 11:59pm ET
Final Report	Wed, May 3rd 11:59pm ET

Schedule

Below is the schedule for the semester with the topic, reading, and swirl tutorials due in each week of course. I will update this with problem sets, found on Canvas under 'Assignments'. You should generally:

- complete the readings by Wednesday;
- complete the tutorials by Wednesday evening at 7pm; and
- submit the problem set or exam by Friday at 11:59pm.

Week	Date	Topic	Reading	Swirl
0	1/18	Introduction	QSS 1.1-1.4	1,2
1	1/25	Causality: Randomized Experiments	QSS 2.1-2.4	3
2	2/1	Causality: Observational Studies	QSS 2.5-2.6	4
3	2/8	Measurement: Descriptive Statistics	QSS 3.1-3.3	5
4	2/15	Measurement: Sampling & Bivariate Statistics	QSS 3.4-3.6	6
5	2/22	Prediction: Elections & Regression	QSS 4.1-4.2.3	7
6	3/1	Prediction: More Regression Prediction: Interactions & Nonlinearities	QSS 4.2.4-4.3	8,9
7	**3/8**	**NO CLASS**		

Week	Date	Topic	Reading	Swirl
8	3/15	Discovery: Textual & Network Data	QSS 5.1-5.2	10,11
9	3/22	Discovery: Spatial Data	QSS 5.3	12
10	3/29	Probability: Basics	QSS 6.1-6.2	13
11	4/5	Probability: Random Variables & Large Samples	QSS 6.3-6.4	14
12	4/12	Inference: Hypothesis Testing & Estimation	QSS 7.1-7.2	15
13	4/19	Inference: Uncertainty in Regression	QSS 7.3	16
14	4/26	Review and Wrap-up		17

Additional Resources:

- Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2020. Regression and Other Stories. Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/regression-and-other-stories/DD20DD6C9057118581076E54E40C372C

- Alvarez, R. Michael, ed. 2016. Computational Social Science: Discovery and Prediction. Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/computational-social-science/FB97BD1704D957183899DE120BEE2E4B

- Druckman, James N. 2022. Experimental Thinking: A Primer on Social Science Experiments. Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/experimental-thinking/C43F73D2255BAD1CB47E39C05E51B399

- Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691207544/text-as-data

- Dunning, Thad. 2012. Natural Experiments in the Social Sciences: A Design-Based Approach. Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/natural-experiments-in-the-social-sciences/96A64CBDC2A2952DC1C68AF77DE675AF

- Best, Henning, and Christof Wolf. 2023. The SAGE Handbook of Regression Analysis and Causal Inference. London. https://methods.sagepub.com/book/regression-analysis-and-causal-inference

R Tutorials

- Introduction to R and R Markdown: https://www.miacosta.net/teaching/r-tutorials