FALL 2020
POIR 611 Introduction to Regression Analysis
Department of Political Science and International Relations
University of Southern California
Schedule: Wednesdays 2 to 4.50pm PT, Zoom link

Instructor: Prof. **Pablo Barberá**
email: pbarbera@usc.edu
Office Hours: Mondays 9am-10am PT and Thursdays 5pm-6pm PT
(To book office hours, click here for Monday slots and here for Thursday slots)

Link to lecture videos

Course Objectives

This course introduces PhD students to quantitative analysis in the social sciences. At the end of the semester, you will be able to:

1. Read and evaluate quantitative research in Political Science and IR
2. Test hypotheses about relationships between variables using quantitative methods, including regression analysis
3. Read and manipulate data in multiple formats for large-n research projects
4. Understand what additional training and skills you will need to conduct research, and the thorough grounding necessary for self-teaching
5. Provide you with a working knowledge of R to facilitate 2-4.

The course is roughly divided in three parts. Weeks 1-6 focus on learning description and inference for a single variable. We will cover the basics of probability theory and hypothesis testing. Weeks 7-11 introduce the workhorse of quantitative analysis – linear regression. This part will focus on the derivation, estimation, and interpretation of the linear model, and then solutions to violations of the linear regression assumptions. The final weeks of the semester will discuss more advanced topics, including time series analysis and data visualization.

The course will follow a "learning-by-doing" approach and will place emphasis on gaining experience in analyzing data with the R programming language. Students are expected to do the required readings for each week and submit completed problem sets ahead of the following session. The lectures will build upon the content of the readings with a series of short in-class assignments that will introduce new statistical and programming concepts, which will then be

applied to the analysis of data from published research papers or common tasks in quantitative social science. Most of the applications will be related to Political Science and International Relations questions, but the course should be of interest to social science students more generally.

Course Requirements and Grading

Class participation: 10%
Problem sets: 60%
Replication paper: 30%

Class participation

Students are expected to attend every session and do the assigned readings before each session. You should come to class with questions and ready to engage in a discussion about that week's topic.

Problem sets

There will be a total of twelve problem sets, each focusing on applying different statistical and programming concepts covered in the course to new datasets in R. Students are encouraged to collaborate but in the submission they must write up the code and answers on their own, and report the name of other students that they worked with in a footnote on the first page. Each problem set will be graded as check plus, check or check minus. Problem sets with a check minus can be resubmitted to increase their grade to check. Late problem sets will be penalized with a 10% grade penalty for each 24 hours after the deadline.

Replication paper

Students are required to submit a research paper that replicates and extends an existing piece of scholarship, either a published article or book chapter. The goal of this exercise is to demonstrate that you have the ability to conduct quantitative research. Note that it will not be sufficient with just reproducing the tables and figures in the original paper, you also need to extend the analysis in some way. You can either conduct additional robustness checks, add new variables or observations, test a new hypothesis with the author's dataset, etc. The submission should include a written report of around 8,000 words in length, as well as the R code and data required to run the analyses in the report. The report should contain a summary of the published research, a description of the methodology employed, the datasets used, a comparison of the published

results and the student's replication, and a discussion of any issues that arose during the replication exercise. The paper or chapter you replicate must be chosen in consultation with the instructor – please send me a PDF copy of the research you will replicate and a one-paragraph summary explaining your choice by the end of week 10 of the semester. The deadline to submit the full replication paper will be December 8 at 11.59pm PT.

Student presentation

Students will present (and give feedback on) a draft of their replication papers during the last session of the semester. Presentations will be less than 10 minutes in length (5-10 slides) and should include a summary of: theory and hypotheses, research design, comparison of original and replicated results, and discussion of potential issues. The final research report should incorporate the feedback received during this session.

Required Readings and Software

The primary software for the course is R. Make sure you install R and RStudio before the first day of class. Additional instructions will be shared before that. No previous programming experience is required.

The main textbooks for the course are:

· Imai, Kosuke (2016) *A First Course in Quantitative Social Science*. Princeton University Press. (QSS in the course outline below)
· Wackerly, Dennis; Mendenhall, William and Scheafer, Richard (2008) *Mathematical Statistics with Applications*. Thomson. (WMS in the course outline below)
· Gujarati, Damodar and Porter, Dawn (2008) *Basic Econometrics*, 5th Edition. McGraw-Hill. (GP in the course outline below)

QSS will be the main textbook of the class, and we will read most of it, and work through the code provided by the author. I recommend buying it.

WMS covers the same set of topics with greater depth. It is a good alternative to QSS if you don't find it useful.

We will rely on GP as additional reading for some topics, especially during the second half of the course. GP is an excellent introduction to econometrics that will be useful beyond the course.

Old editions (and the international edition) should be cheaper to buy and will also suffice for the course.

There are many other excellent textbooks and articles that address the topics from the course in greater detail. The following is a list of recommended readings that you can rely on to supplement the required books:

· Diez, David M., Christopher D. Barr, and Mine Çetinkaya-Rundel. 2015. *Open-Intro Statistics*. 3rd edition. https://www.openintro.org/
· Gelman, Andrew and Hill, Jennifer (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press. (GH in the course outline below)
· Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Nelson Education.
· Freedman, D., Pisani, R., & Purves, R. (2007). Statistics (4th edition). WW Norton and Company.
· Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
· Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Course Outline

Week 1. August 19. Course overview.

Introduction to quantitative data analysis.

R programming: arithmetic operations, object types, data files, packages, if/then conditions, subsetting data

Readings:
[required] QSS Chapters 1, 2.1, 2.2 (after class)

Week 2. August 26. Introduction to probability.

Probability theory: frequentist vs Bayesian frameworks. Set notation. Calculating the probability of an event using the sample-point method. Combinatorial analysis: permutations and combinations. Conditional, marginal, and joint probability. Bayes' rule.

R programming: functions and loops

<u>Readings</u>:
[required] QSS Chapters 6.1, 6.2.1, 6.2.2, and 6.2.3
[recommended] WMS Chapter 2
[application] Westwood, Sean; Solomon Messing and Yphatch Lelkes (2020) "Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilizes the Public" *Journal of Politics*, forthcoming.

Problem set 1 deadline: August 25, 11.59pm PT

Week 3. September 2. Discrete and continuous probability distributions

Random variables. The probability distribution for discrete random variables: Bernoulli, and binomial distributions. The probability distribution for a continuous random variable: uniform and normal probability distributions.

R programming: density, distribution function, quantile function and random generation from a random distribution.

<u>Readings</u>:
[required] QSS Chapter 6.3 (excluding 6.3.5)
[recommended] WMS Chapters 3.1-3.4, 4.1-4.5, and 8.1-8.4

Problem set 2 deadline: September 1, 11.59pm PT

Week 4. September 9. Uncertainty (I)

Expected value and variance of a random variable. The law of large numbers. Central limit theorem. Unbiasedness and consistency of point estimators.

R programming: Monte Carlo simulations.

<u>Readings</u>:

[required] QSS Chapters 6.3.5, 6.4 and 7.1.1 (pages 314-317)
[recommended] WMS Chapter 7.1-7.3
[application] Beber, Bernd and Alexandra Scacco (2012) "What the Numbers Say: A Digit-Based Test for Election Fraud" *Political Analysis*, 20 (2): 211- 234.

Problem set 3 deadline: September 8, 11.59pm PT

Week 5. September 16. Uncertainty (II)

Standard errors. Confidence intervals for proportions and averages. Margin of error and sample size calculation in polls.

R programming: confidence intervals

Readings:
[required] QSS Chapters 2.6 and 7.1 (excluding 7.1.1)
[recommended] WMS Chapter 8.5
[application] Rozenas, Arturas (2017) "Inferring Election Fraud from Distributions of Vote Proportions" *Political Analysis*, 25 (1): 41-56.

Problem set 4 deadline: September 15, 11.59pm PT

Week 6. September 23. Hypothesis testing

Elements of a statistical test. One-sample tests. Two-sample tests. Interpreting p-values. Power of tests.

R programming: hypothesis testing

Readings:
[required] QSS Chapter 7.2
[recommended] WMS Chapter 10.1-10.7
[application] Franco, A., N. Malhotra, and G. Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502–5.

Problem set 5 deadline: September 22, 11.59pm PT

Week 7. September 30. Bivariate linear regression.

Motivation for regression analysis. The method of least squares. Assumptions of linear regression. The Gauss-Markov theorem. Measuring model fit: R-squared and adjusted R-squared. Standard errors of estimated coefficients. Confidence intervals for regression coefficients.

R programming: estimating linear regression models.

Readings:
[required] QSS Chapter 4.2
[recommended] WMS Chapter 11.1-11.9 or GP Chapters 1, 2, and 3
[application] Munger, K. 2018 "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." *Political Behavior.*

Problem set 6 deadline: September 29, 11.59pm PT

Week 8. October 7. Multivariate linear regression

Multiple regression analysis: estimation and interpretation. Omitted variable bias. Standard errors in a multivariate regression. Testing the overall significance of a multiple regression using the F test.

R programming: multivariate regression

Readings:
[required] QSS Chapters 4.3.1, 4.3.2 and 7.3
[recommended] WMS Chapter 11.10-11.14 or GP Chapters 6, 7, and 8
[application] Qian, N. 2008. "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance." *Quarterly Journal of Economics*, 123(3): 1251–85
[application] Broockman, D., and J Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352(6282): 220-224.

Problem set 7 deadline: October 6, 11.59pm

Week 9. October 21. Relaxing the linearity assumption

Polynomial regression. Scaling and units of measurement. Interactive models. Analytical computation of marginal effects in regression models.

R programming: models with interaction effects

Readings:
[required] QSS Chapter 4.3.3

Problem set 8 deadline: October 20, 11.59pm

Week 10. October 28. Marginal effects. Simulation

The nature of multicollinearity. Theoretical and practical consequences of multicollinearity. The nature of heteroscedasticity. Detecting heteroscedasticity. Robust standard errors. Summarizing linear regressions using simulation.

R programming: Monte Carlo simulations applied to regression analysis; robust standard errors

Readings:
[required] GP Chapters 10 and 11
[recommended] GP Chapter 13

Problem set 9 deadline: October 27, 11.59pm

Week 11. November 4. Regression with binary dependent variables

Overview of regression with binary dependent variables: logit and probit regressions. Measurement error in the dependent or independent variables. Principles for effective data visualization.

R programming: logit regression models. Introduction to ggplot2

Readings:
[recommended] GP Chapter 15

[required] Kastellec, J. and E. Leoni (2007) "Using graphs instead of tables in political science" *Perspectives on Politics*

[recommended] Hughes, A. (2015) "Visualizing inequality: How graphical emphasis shapes public opinion" *Research and Politics*.

Problem set 10 deadline: November 3, 11.59pm **[Extension: November 10th]**

Week 12. November 11. Time series analysis. Good coding practices

OLS estimation in the presence of autocorrelation. Detecting autocorrelation. Unit root. Stationarity. Good coding practices.

R programming: time series models in R.

Readings:
[recommended] GP Chapter 12

Problem set 11 deadline: November 10, 11.59pm

Week 13. December 3rd. Student presentations.

Details TBD

Replication project deadline: December 8, 11.59pm

Problem set 12 deadline (optional): November 25, 11.59pm

Digital tools used for this course

Zoom

All lectures and office hours will occur via Zoom.
- Zoom links are available via Blackboard (see the Zoom tab)

- Though not required, I encourage you all to attend our synchronous sessions with video capability on
- All Zoom sessions will be recorded and shared with the class each week
- For office hours, the calendar invite you receive will contain a Zoom link. I will use the 'waiting room' option and give you access at the time you requested.

Slack

To supplement and enhance in-person classroom interactions, we will also use Slack, a collaboration hub where the right people are always in the loop and key information is always at your fingertips.

- I will share regular updates over Slack. Important updates will also be shared via email.
- As a student in this course, you will be able to use our #fall20-poir-611-52302 channel to collaborate with your peers and myself. Feel free to use it to ask any questions related to the class and to initiate chats with other students.
- Prior to the start of classes, please be sure to set up your Slack account at https://usc.enterprise.slack.com/ and familiarize yourself with our class channel. Additional reference guides and resources can be found on the Keep Teaching website.
- As a reminder, all Slack conduct must be in line with USC policies and USC Policies Student Issues. Any behavior that breaks outlined policies will be subject to discipline.

Academic Conduct

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Support Systems

Counseling and Mental Health - (213) 740-9355 – 24/7 on call

studenthealth.usc.edu/counseling

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call

suicidepreventionlifeline.org

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call

studenthealth.usc.edu/sexual-assault

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) - (213) 740-5086 | Title IX – (213) 821-8298

equity.usc.edu, titleix.usc.edu

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298

usc-advocate.symplicity.com/care_report

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity |Title IX for appropriate investigation, supportive measures, and response.

The Office of Disability Services and Programs - (213) 740-0776

dsp.usc.edu

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Campus Support and Intervention - (213) 821-4710

campussupport.usc.edu

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101

diversity.usc.edu

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu, emergency.usc.edu
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
ombuds.usc.edu
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.