

POLS-Y 575: Political Data Analysis I

Prof. Caleb Lucas

Fall 2024

Instructor:	Caleb Lucas
Office:	403 Woodburn
e-mail:	lucascaj@iu.edu
Class:	W 17:30 – 20:00
OH:	M 09:00 – 10:00

This syllabus is subject to change

Course Description

POLS-Y 575 introduces students to quantitative methods used in social science research. It focuses on foundational concepts such as descriptive statistics, probability theory, sampling, hypothesis testing, and linear regression. These concepts inform more advanced statistical methods used to study important political phenomena, but are also individually useful – they are commonly used in published quantitative social science research. Time permitting, the course will also introduce students to causal inference and the importance of research design.

This course is computational: students will become proficient in R, a statistical programming language, during the course.

Materials

- Agresti, Alan. *Statistical Methods for the Social Sciences*. Pearson, 2018. (AA)
 - I will use the **fifth edition**
- Wilke, Claus. *Fundamentals of Data Visualization*. O'Reilly, 2019. (FDV)
 - Available for free: <https://clauswilke.com/dataviz/>
- Gujarati, Damodar & Porter, Dawn. *Basic Econometrics*. McGraw-Hill, 2019. (GP)
- (optional) Llaudet, Elena & Imai, Kosuke. *Data Analysis for Social Science: A Friendly Introduction*. Princeton University Press, 2021. (DSS)
- Various journal articles & additional readings noted in the course schedule

Course R Package

I created an R package for use in this course. We will use the yesPLS package to share data and customized functions. I may update the package throughout the course, meaning you may need to update it periodically to ensure it is current with the content. The package is not published on CRAN, so you can download it using the devtools library. Download and update it with the following code:

```
devtools::install_github("caleblucas/yesPLS")
```

Policies

Computing

I will teach this course using the R programming language. It is open-source and completely free. All the packages we will use are also free. It is not only increasingly common in social science research, but gaining popularity among data scientists as well. Assignments will generally require using the R programming language. Instructions regarding the functional aspects of submitting them, such as file types, formatting, etc., will be elaborated on when they are assigned.

Download R at <https://cran.r-project.org/> and RStudio, the integrated development environment (IDE) that I recommend, at <https://rstudio.com/products/rstudio/download/>.

Attendance

Class attendance is strongly encouraged, but not required. There is no need to inform me of sicknesses or personal emergencies unless they will require missing more than two consecutive class meetings. However, please reach out to me if you need assistance staying current with the course because of these sorts of issues. This is especially important because a portion of your grade includes participation.

Academic Integrity

Please review and adhere to Indiana University's rules regarding academic integrity:

<https://studentlife.indiana.edu/care-advocacy/student-conduct/policies/index.html>

Violations of the university's academic code of conduct will be reported and could lead to probation or expulsion. A partial list of activities that I consider academically dishonest includes:

- Systematically copying and pasting answers from another student's work, even if you collaborated on that work in a manner allowed by the course (see below).
- Receiving answers to assignments from colleagues, former students, or the internet.

Collaboration

You are encouraged to collaborate with your classmates on assignments in this course. However, you must cite peers that you collaborated with on a question-by-question basis when submitting assignments. Citations by peers will be considered in participation grades. Anything you submit must be your own work. Please respect the academic integrity policy detailed above.

Methods Workshop + Extra Credit

I am planning a Methods Workshop for the POLS department. After it kicks off, you are strongly encouraged to attend for the purposes of this course and your career. Details about when these occur are your responsibility, but I will do my best to circulate information about that workshop's progress and the timing of events. Attendance is the only way to earn extra credit in this course. Email me with proof of attendance (include the word "IUmehods" somewhere in your email to ensure I can retrieve it later) and receive 0.5% of your grade as extra credit. The maximum possible increase in your final grade is 2%.

Students with Disabilities

If any student requires assistance or accommodations for a disability, please get in touch with me at your earliest convenience. You can reach me via email or by attending office hours. You must have established your eligibility for disability support services through the Office of Disability Services for Students in Wells Library, W302, 812-855-7578.

Assignments

Assignment Weights

- Problem Sets (50%)
- Participation (10%)
- Mid-term (20%)
- Final (20%)

Problem Sets

No credit will be given for late assignments. There are no exceptions to this rule unless they are prearranged with me. I will grade problem sets using a (✓−, ✓, ✓+) scale. You are encouraged to discuss problem sets with each other. However, all words, numbers, and code *must* be your own. You are also *strongly* encouraged to attempt all problem sets independently before working collaboratively. See the Collaboration policy above. The point of problem is the *learn*, not to 'finish' them or even to simply achieve a high grade. You *must* use R to complete these problem sets when programming is necessary.

Participation

Your participation in this course is important. You can earn credit towards your participation grade by being cited by your peers on problem sets, sharing insights relevant to the course with the class, attending departmental methods workshops, and actively engaging during class meetings. This list is not exhaustive.

Preliminary Schedule

Description and Inference

Week 1: August 28

Topic: Class Introduction, Statistical Inference, Programming Orientation

Week 2: September 4

Topic: Descriptive Statistics and Data Exploration through Visualization

Reading: AA 1-3, skim FDV 1-7, (optional) DSS 3

Week 3: September 11

Topic: Probability

Reading: AA 4, (optional) DSS 6, these articles in order (read abstract, intro, conclusion of full length one):

- <https://www.pnas.org/content/116/32/15877>,
- <https://www.pnas.org/content/117/3/1261>,
- <https://www.pnas.org/content/117/30/18130>,
- <https://spia.princeton.edu/news/cautionary-tale-about-measuring-racial-bias-policing>

Week 4: September 18

Topic: Statistical Inference: Estimation

Reading: AA 5, (optional) DSS 7

Week 5: September 25

Topic: Statistical Inference: Significance

Reading: AA 6, [Jeff Gill \(1999\): *The Insignificance of Null Hypothesis Significance Testing*](#)

Week 6: October 2

Topic: Comparing Two Groups, Associations Between Categorical Variables

Reading: AA 7, 8

Linear Regression

Week 7: October 9

Topic: Linear Regression & Correlation

Reading: AA 9, Optional: DSS 4, [Statistical Tests are Linear Models](#)

Week 8: October 16

Topic: Correlation and Causation, Multivariate Relationships

Reading: TBD, (optional) DSS 2, 5

Week 9: October 23

Topic: Finish OLS introduction, Predictions, Confidence intervals

Reading: AA 10

Week 10: October 30

Topic: Gauss-Markov, OLS assumptions, More applied regression

Reading: GP pgs 2-11, sections 2.1-3.4, 4.1-4.2

Week 10: November 6

Topic: Correlation and Causation, Multivariate Relationships

Reading: AA 11, 12

Week 11: November 13

Topic: Model Specification, Dichotomous variables, Confounding, Interactions

Reading: GP 13, [Regression Checklist](#)

Week 12: November 20

Topic: Multicollinearity, Heteroscedasticity, Diagnostics

Reading: GP 10-12 (skim as needed)

Week 14: December 4

Topic: Finish prior content. Simulation and advanced regression analysis if time allows

Reading: TBD

Week 15: December 11

Topic: Final Exam

I reserve the right to modify this syllabus at any point in the course to suit the needs of the class.