HARVARD
UNIVERSITY

Syllabus for Gov 2003 at Harvard University

Monday, & Wednesday, 1:30pm-2:45pm

Head Instructor
Jeff Gill
jgill@american.edu
JeffGill.org
Office Hours: TBD

Teaching Fellow
Xinyuan Yang
xy4066a@american.edu
Th 10:30 - 11:45 AM CGIS Knafel K109 (FAS)
Th 3:00 PM - 4:15 PM Northwest Bldg B107

Description. This course extends what you did in previous methods courses by focusing more on nonlinear model forms. These are typically called "generalized linear models," although for historical reasons people in political science call them "maximum likelihood models." The principle we will care about is how to adapt the standard linear model that you know so that a broader class of outcome variables can be accommodated. These include: counts, dichotomous outcomes, bounded variables, and more. There is a strong theoretical basis for the models that we will use. Because we live in data century there will also be some coverage of machine learning and neural nets. Also, the bulk of the learning in the course will take place outside of the classroom by reading, practicing using statistical software, replicating the work of others, and doing problem sets. Keep in mind that the skills attained in this course are those that the discipline of political science expects of any self-declared data-oriented researcher. In this course you will be expected to: complete 6 problem sets, submit on `Canvas`, participate in the course via lectures, sections, and discussion forums, complete a final group research project, and provide feedback on another group's research project.

Course objectives. After taking this course you will:

- Understand the basic issues and approaches to general inference and modeling in the social sciences.
- Be able to critically evaluate the assumptions that justify modeling decisions and techniques in applied settings.
- Have the ability to implement the most common social science regression inference methods with applied data.

Prerequisites. Students in the class should plan to work hard at understanding some difficult material. You'll need to have some proficiency with `R`. You should have taken GOV 2001 and GOV 2002 or the equivalent. Of course there are always exceptions to this, so feel free to talk with me about your background to see if the fit is right. Qualified undergraduates and graduate students from other departments are welcome to join the class.

Grading.

- Biweekly homework assignments (60% of final grade)
- Final collaborative project (20% of final grade)
- Participation (10% of final grade)
- 2-3 pages peer feedback on projects (10% of final grade).

Course structure. Lectures will cover the basic theoretical issues in lecture along with some applied examples. We'll try to foster discussions when we can and perhaps even attempt some "breakout" time to work through issues. Your attendance at lectures is very strongly recommended and your participation counts toward your grade.

Section. At section, your Teaching Fellow will review the material for the week and demonstrate how to implement the methods we describe in lecture.

Reading. There are readings for each topic and they mostly cover the theory of the method along with some applications. Obviously, read the required readings and any others that pique your curiosity. In addition, though, engage with the readings: take notes, write down your impressions or confusions, talk with your classmates. All of your classes should be pushing your research forward and you will be more creative the more you actively read.

Course assignments. Participation is crucial component to the class and we will grade your participation based on your engagement during class meetings along with your contributions to discussions and feedback via `Canvas`. Methods are tools and it is

not very instructive to read a lot about hammers or watch someone else wield a hammer. You need to get your hands on a hammer or two. Thus, in this course, you will have 6 problem sets. They will be a mix of conceptual questions, analytic problems, computer simulations, and data analysis. These problem sets should be typed and well-formatted, using `RMarkdown`. Each problem set will be equally weighted. I encourage you to work in groups on the homework, but you always need to write your own solutions including your computer code. Also, it is hugely beneficial to attempt the problems sets on your own before working in groups. The schedule for the problem sets is:

#	Release Date	Due Date
1	Thu, September 12th 12:00pm ET	Wednesday, September 25th 11:59pm ET
2	Thu, September 26th 12:00pm ET	Wednesday, October 2nd 11:59pm ET
3	Thu, October 3rd 12:00pm ET	Wednesday, October 16th 11:59pm ET
4	Thu, October 17th 12:00pm ET	Wednesday, October 30th 11:59pm ET
5	Thu, October 31st 12:00pm ET	Wednesday, November 13th 11:59pm ET
6	Thu, November 14th 12:00pm ET	Tuesday, November 26th 11:59pm ET

Student Project. In lieu of a final exam, this course requires students to write a short paper applying or extending the methods we learn in this class. These papers can be substantive (applying the methods to a problem of interest to you) or methodological (deriving a new method that innovates over existing ones). It should be no longer than 20 double-spaced pages and focus on the research design, data, methodology, results, and analysis. Literature reviews or double-spaced pages and focus on the research design, data, methodology, results, and analysis. Literature reviews or background material should be omitted or included in an appendix. Co-authored projects are strongly encouraged and working as an individual requires approval. Working with 3-4 collaborators will be the cornerstone of your career from now on, so it's crucial to get to know this process sooner rather than later. The timetable for the projects is:

- September 25th: Find collaborators or obtain permission from the instructor to work on an individual project.
- October 2nd: Submit a short (half-page) description of your proposed project and a feasible plan for carrying out the research.
- November ember 14th: Submit a brief (no longer than 5 page) page memo of

your main results, including tables, figures, and brief analysis. For methodological projects, this should include a description of the method and any analytical/simulation results. You will be required to give feedback on another group's project, which will be counted toward the overall grade based on attentiveness and usefulness of the feedback provided.

- December 3rd: Presentations in class.
- December 11th: Peer Review Due.
- December 11th: Submit your final version of the project.

Books. We will draw heavily on the following texts for this class:

- Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models.* Second Edition 2016.
- Gill and Torres, *Generalized Linear Models: A Unified Approach.* Second Edition. 2019.
- Cattaneo, Idrobo, and Titiunik, *A Practical Introduction to Regression Discontinuity Designs.* 2024. Available at `https://arxiv.org/pdf/2301.08958`.

Course Plan

- Week 1 Everything You Need to Know About the Linear Model. Wednesday, September 4th
  Reading: Gill and Torres Chapter 1, Faraway Chapter 1
- Week 2 Binomial and Logit Regression. Monday, September 9th and Wednesday, September 11th
  Reading: Faraway Chapter 2-4
- Week 3 Contingency Tables and Count Regression. Monday, September 16th and Wednesday, September 28rd
  Reading: Faraway Chapter 5-6
- Week 4 Multinomial Data. Monday, September 23th and Wednesday, September 25th
  Reading: Faraway Chapter 7
- Week 5 Generalized Linear Models and Estimation. Monday, October 30th and Wednesday, October 2nd
  Reading: Gill and Torres Chapter 2-6

- Week 6 Extensions to Generalized Linear Models. Monday, October 7th and Wednesday, October 9th
  Reading: Faraway Chapter 8-9
- Week 7 Fixed Effects. Random Effects and Hierarchies. Monday, October 14th and Wednesday, October 16th
  Reading: Faraway Chapter 10
- Week 8 Repeated Measures and Longitudinal Data. Monday, October 21st and Wednesday, October 23rd
  Reading: Faraway Chapter 11
- Week 9 Nonparametric Regression. Monday, October 28th (Zoom) and Wednesday, October 30th
  Reading: Faraway Chapter 14
- Week 10 Additive Models. Monday, November 4th and Wednesday, November 6th
  Reading: Faraway Chapter 15
- Week 11 Regression Disconinuity Designs. Monday, November 11th and Wednesday, November 13th
  Reading: Cattaneo, Idrobo, and Titiunik
- Week 12 A Brief Introduction to Machine Learning Including Trees. Monday, November 18th and Wednesday, November 20th
  Reading: Faraway Chapter 17
- Week 13 Neural Networks and Deep Learning. Monday, November 25th
  Reading: Handout
- Week 14 Brief Presentation of Projects. Monday, December 2
  Reading: None