PSC 8101: Introduction to Empirical Political Analysis
Department of Political Science ∼ George Washington University
Fall 2024 ∼ Wednesdays, 12:45-3:15 ∼ 1957 E St., 316

Professor Brandon Bartels (bartels@gwu.edu)   Office: Monroe 477
Office hours: Fridays, 1:30-3:00 or by appointment (in person or Zoom)

Course Description

The purpose of this Ph.D. course is to introduce you to probability, statistics, and data analysis, particularly with respect to their application in political science. It places particular emphasis on linear regression analysis ("ordinary least squares," or OLS), the most common statistical technique employed in the social sciences. Toward the end of class, we will also venture into: (1) maximum likelihood estimation (MLE) for nonlinear models (e.g., binary dependent variables using logit and probit), and (2) models for panel and time-series cross-sectional (TSCS) data.

In the course, you will learn fundamental principles of probability and statistical inference, how to summarize data and make statistical inferences, and how to manipulate and analyze data in statistical software. The class serves as a prerequisite for Causal Inference in the spring and will prepare you for more advanced classes in statistical methods.

Learning Objectives

As a result of completing this course, students will be able to:

- read a wider range of literature in political science and other social sciences,
- have a deeper understanding of what conclusions can be reached, given different research designs,
- develop basic facility analyzing data with statistical software using reproducible research methods, and
- be prepared to independently undertake empirical research projects.

Required Books

The following two books will serve as our main texts throughout the course:

- Diez, David, Christopher Barr, and Mine Çetinkaya-Rundel. 2019. *OpenIntro Statistics, Fourth Edition.* This book is free online (download as .pdf). Also note that the OpenIntro page has links to a number of other good free statistics texts. We'll rely on this book mostly during Weeks 2-5. The book is a great reference for the future

- Wooldridge, Jeffery M. *Introductory Econometrics: A Modern Approach.* Mason, OH: Thomson South-Western. This book is currently in its 7th edition, and it's expensive. I recommend buying the 5th edition, which you can get for pretty cheap. There isn't much change in content across editions. Even older than 5th is fine (e.g., 3rd or 4th). Abebooks has lots of cheap (less than $10) editions. Amazon is an option, too. We'll rely on this book once we get into regression, namely Weeks 6-14. This is one of the most important long-term methods reference books you'll have on your bookshelf, so

it's a great investment.

Supplemental Texts
In addition to the main texts, we will draw on the following texts for tips on data visualization and R guides. I will assign portions of these texts to read for select weeks. They are also excellent resources for the future. All of these books are available for free online (click on links to access).

- Ismay, Chester, and Albert Y. Kim. 2023. *Statistical Inference via Data Science: A ModernDive into R and the Tidyverse*. CRC Press.

- Wickham, Hadley, Mine Cetinkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science (2e)*. O'Reilley.

- Rafael A. Irizarry has two fantastic online books that echo and expand on some of the material in both Ismay and Kim and Wickham et al. (1) *Introduction to Data Science: Data Wrangling and Visualization with R* and (2) *Advanced Data Science: Statistics and Prediction Algorithms Through Case Studies*

- Heiss, Florian. 2016. *Using R for Introductory Econometrics*. This book contains R implementations of all of the empirical examples in the Wooldridge text. It's getting a little dated, but it remains useful. Once again, this is a great reference for class and for the future.

Statistical Software: R, RStudio, and Quarto
In this class, we will use RStudio, which is a streamlined software package for running the statistical software, R. You will download both R Studio and R, but you will only ever open RStudio (RStudio requires R to be downloaded). Both are free. R has now become perhaps the most popular statistical programming language in political science and applied statistics. Stata is also widely used in Economics, Political Science and in industries that lean on those disciplines.

If you take Causal Inference with Mike Miller in the Spring of 2024, the primary software there will be Stata. If you continue on to take PSC 8124 (Multilevel) with Chris Warshaw in the Fall of 2024, the primary software in that class will again be R. It's good to know both, which can open up opportunities for working with a wider set of people and projects. Upon request, we can occasionally use Stata in this class to show comparisons, though we'll prioritize R.

We will also use a powerful feature of RStudio called "Quarto." Quarto is a software tool that integrates your results from R analyses with your narrative for explaining your results. It is especially valuable for communicating your results in a paper, report, analysis memo, or even a book or dissertation. It's like a more intricate yet more powerful version of Microsoft Word. It generates documents in both .pdf and .html format. You can also use Quarto to create a webpage or presentation slides (akin to Powerpoint). Quarto also incorporates and accommodates the Latex infrastructure. And it works with other programs, such as Python

and Julia, which make it extremely useful and flexible beyond R. You will use Quarto to complete all of your assignments. As a sidenote, Quarto is a more generalized version of a similar program within R Studio called "R Markdown."

R and Quarto are very popular in academia, data science, and the programming industry (private sector and government). Importantly, you will attain valuable analytical, programming, data analytic, and statistical skills in this class that can make you more marketable.

I will devote part of each class to "lab sessions" where we illustrate how to do procedures in R. We'll assume that each student will bring a computer to class with R and RStudio loaded on it. We will review the basics of using RStudio in Weeks 1 and 2.

Grades

Your course grades will be based on a set of seven homework assignments (70%), a final exam (20%), and class participation (10%).

The homework assignments (70% of your grade) will be due on the following dates:

1. Fri., Sept. 13
2. Fr., Sept. 27
3. Mon., Oct. 14 [Fall Break covers Fri., Oct. 11]
4. Fri., Oct. 25
5. Fri., Nov. 8
6. Fri., Nov. 22
7. Mon., Dec. 9

You are permitted and encouraged to work on the problem sets with your classmates, but each student must independently write up his or her assignments.

The final exam will be open book and open note, part of which will be modeled on part I of the methods PhD comp. The final exam (20% of your grade) will be due on the day during finals week.

Participation, Reading, and Online Help Forum (Piazza)

This is a fairly small class, so your participation will help determine the success of the class. If you have a laptop, bring it to every class; we will do computer work during each class session. The reading load will not be too heavy, so you should be able to keep up with the readings throughout the semester. Also, we will be using Piazza for additional class discussion. The system is highly catered to getting you help fast and efficiently from classmates and from me. Rather than emailing questions to just me, I encourage you to post your questions on Piazza. If you have non-technical questions that are not appropriate for the class, just email me alone. To sign up on our Piazza course page, click here: https://piazza.com/gwu/fall2024/psc8101.

Additional Resources

There are a tremendous number of free statistics books online, videos with demonstrations

of solving statistics problems at sites like Khan Academy, pages and websites dedicated to particular statistical software packages, and omnibus pages gathering data, code, and learning modules for a host of different data analytic problems.

Most of the time, the best way to solve a vexing problem is to "Google it." You will rarely be the first person to encounter your problem in the wild. You will most likely find an answer. That's my preferred method, actually. The rise of ChatGPT also provides an additional opportunity for attaining tips on R code. The RStudio community has a friendly vibe where you can often find answers to your questions.

Course Schedule

1. Aug. 28: Course Introduction, Research Design, and Data Analysis; Lab: Downloading and Using R, RSTudio, and Quarto.

   - Ismay and Kim, Ch. 1
   - Wickham et al., Ch. 1
   - Begin reading OpenIntro, Ch. 1

2. Sept. 4: Data Visualization; Lab: Graphing (ggplot2), Maps, Data Wrangling, and Quarto

   - OpenIntro, Ch. 1 (contd.) and Ch. 2
   - Ismay and Kim, Chs. 2 and 3
   - Wickham et al., Chs. 2 and 4 (covers similar ground as Ismay and Kim)
   - For your reference: Irizarry (*Intro*), Chs. 7-10 (covers similar ground as Ismay and Kim and Wickham et al.)

3. Sept. 11: Probability; Lab: Data Wrangling (contd.)

   - OpenIntro, Ch. 3
   - Ismay and Kim, Ch. 3 (contd.)

4. Sept. 18: Distributions; Lab: Additional R Topics

   - OpenIntro, Ch. 4

5. Sept. 25: Inference, Including Difference in Means

   - OpenIntro. Ch. 5; Ch. 7 [focus on 7.1 and 7.2]

6. Oct. 2: Simple Regression

   - Wooldridge, Ch. 2

- For your reference: OpenIntro, Ch. 8

7. Oct. 9: Multiple Regression I

   - Wooldridge, Ch. 3
   - For your reference: OpenIntro, Ch. 9

8. Oct. 16: Multiple Regression II

   - Wooldridge, Ch. 4

9. Oct. 23: Multiple Regression III

   - Wooldridge, Ch. 4

10. Oct. 30: Model Specification: Functional Forms and Interactions

    - Wooldridge, Ch. 6
    - Brambor, Clark, and Golder (2006, Political Analysis)

11. Nov. 6: Interactions (contd.) and "Qualitative" Independent Variables (e.g., dummy variables)

    - Wooldridge, Ch. 7

12. Nov. 13: Regression Diagnostics

    - Wooldridge, Chs. 8-9

13. Nov. 20: Binary Dependent Variables: Logit, Probit, and the Linear Probability Model; Maximum Likelihood Estimation

    - OpenIntro, Ch. 6 and Ch. 9 (section 9.5)

   No class on Wed., Nov. 27, Thanksgiving break

14. Dec. 4: Introduction to Models for Panel and Time-Series Cross-Sectional (TSCS) Data

    - Wooldridge, Chs. 13 and 14

University Policies

Academic Integrity Code
Academic integrity is an essential part of the educational process, and all members of the GW community take these matters very seriously. As the instructor of record for this course, my role is to provide clear expectations and uphold them in all assessments. Violations of academic integrity occur when students fail to cite research sources properly, engage in unauthorized collaboration, falsify data, and otherwise violate the Code of Academic Integrity. If you have any questions about whether particular academic practices or resources are permitted, you should ask me for clarification. If you are reported for an academic integrity violation, you should contact Conflict Education and Student Accountability (CESA), formerly known as Student Rights and Responsibilities (SRR), to learn more about your rights and options in the process. Consequences can range from failure of assignment to expulsion from the University and may include a transcript notation. For more information, refer to the CESA website at students.gwu.edu/code-academic-integrity or contact CESA by email cesa@gwu.edu or phone 202-994-6757.

University policy on observance of religious holidays
Students must notify faculty during the first week of the semester in which they are enrolled in the course, or as early as possible, but no later than three weeks prior to the absence, of their intention to be absent from class on their day(s) of religious observance. If the holiday falls within the first three weeks of class, the student must inform faculty in the first week of the semester. For details and policy, see provost.gwu.edu/policies-procedures-and-guidelines.

Use of Electronic Course Materials and Class Recordings
Students are encouraged to use electronic course materials, including recorded class sessions, for private personal use in connection with their academic program of study. Electronic course materials and recorded class sessions should not be shared or used for non-course related purposes unless express permission has been granted by the instructor. Students who impermissibly share any electronic course materials are subject to discipline under the Student Code of Conduct. Contact the instructor if you have questions regarding what constitutes permissible or impermissible use of electronic course materials and/or recorded class sessions. Contact Disability Support Services at disabilitysupport.gwu.edu if you have questions or need assistance in accessing electronic course materials.

Academic Support

Academic Commons
Academic Commons is the central location for academic support resources for GW students. To schedule a peer tutoring session for a variety of courses visit go.gwu.edu/tutoring. Visit academiccommons.gwu.edu for study skills tips, finding help with research, and connecting with other campus resources. For questions email academiccommons@gwu.edu.

GW Writing Center
GW Writing Center cultivates confident writers in the University community by facilitating collaborative, critical, and inclusive conversations at all stages of the writing process. Work-

ing alongside peer mentors, writers develop strategies to write independently in academic and public settings. Appointments can be booked online at gwu.mywconline.

Support for Students In and Outside the Classroom

Disability Support Services (DSS) 202-994-8250
Any student who may need an accommodation based on the potential impact of a disability should contact Disability Support Services at disabilitysupport.gwu.edu to establish eligibility and to coordinate reasonable accommodations.

Student Health Center 202-994-5300, 24/7
The Student Health Center (SHC) offers medical, counseling/psychological, and psychiatric services to GW students. More information about the SHC is available at healthcenter.gwu.edu. Students experiencing a medical or mental health emergency on campus should contact GW Emergency Services at 202-994-6111, or off campus at 911.

GW Campus Emergency Information

GW Emergency Services: 202-994-6111 For situation-specific instructions, refer to GW's Emergency Procedures guide.

GW Alert
GW Alert is an emergency notification system that sends alerts to the GW community. GW requests students, faculty, and staff maintain current contact information by logging on to alert.gwu.edu. Alerts are sent via email, text, social media, and other means, including the Guardian app. The Guardian app is a safety app that allows you to communicate quickly with GW Emergency Services, 911, and other resources. Learn more at safety.gwu.edu.

Protective Actions
GW prescribes four protective actions that can be issued by university officials depending on the type of emergency. All GW community members are expected to follow directions according to the specified protective action. The protective actions are Shelter, Evacuate, Secure, and Lockdown (details below). Learn more at safety.gwu.edu/gw-standard-emergency-statuses.

Shelter

- Protection from a specific hazard
- The hazard could be a tornado, earthquake, hazardous material spill, or other environmental emergency.
- Specific safety guidance will be shared on a case-by-case basis.

Action:

- Follow safety guidance for the hazard.

Evacuate

- Need to move people from one location to another.
- Students and staff should be prepared to follow specific instructions given by first responders and University officials.

Action:

- Evacuate to a designated location.
- Leave belongings behind.
- Follow additional instructions from first responders.

Secure

- Threat or hazard outside of buildings or around campus.
- Increased security, secured building perimeter, increased situational awareness, and restricted access to entry doors.

Action:

- Go inside and stay inside.
- Activities inside may continue.

Lockdown

- Threat or hazard with the potential to impact individuals inside buildings.
- Room-based protocol that requires locking interior doors, turning off lights, and staying out of sight of corridor window.

Action:

- Locks, lights, out of sight
- Consider Run, Hide, Fight
- Classroom emergency lockdown buttons: Some classrooms have been equipped with classroom emergency lockdown buttons. If the button is pushed, GWorld Card access to the room will be disabled, and GW Dispatch will be alerted. The door must be manually closed if it is not closed when the button is pushed. Anyone in the classroom will be able to exit, but no one will be able to get in.

AI Policy (written by Profs. Robert Betz and Eric Lawrence)

We are entering a new technological era with the rise of generative Artificial Intelligence (AI), such as GPT, LLaMA, laMDA and other large language models, that are driving an ongoing conversation about their academic uses. Writing aid products like Grammarly, QuillBot,

Caktus.ai, etc. advertise their AI features (GrammarlyGO, etc.). We are also learning about the potential benefits and misuse of AI and how it can be applied in the classroom. Learning to use generative AI is an emerging skill, but we must use generative AI tools effectively and responsibly.

Generative AI has been discussed at length within academia, but other sectors are also grappling with its use due to its rapid rise and increased access to the tools. In the private sector, e.g., many workplaces are considering banning its use, partly due to security risks (see "Most businesses to ban ChatGPT, generative AI apps on work devices," August 8, 2023, CSOonline). We are all trying to figure out the right way to use GAI in the long and medium run, but below I have outlined the expectations in our class of its permitted and prohibited use.

Permitted:

- A student types a prompt into an AI tool and reviews the generated content to help them study for a quiz or exam (i.e., a study guide).
- A student types a prompt into an AI tool and uses the generated content to help them brainstorm ideas for a paper or research project.
- A student types a prompt into an AI tool and uses the generated content to help them create a citation for a source and/or reference list.
- A student types a prompt into an AI tool and uses the generated content to help them with small group discussion.

Citing GAI and Verifying its Accuracy:

- By submitting work for evaluation in this course, you represent it as your own intellectual product. If you include content (e.g., ideas, text, code, images) that was generated, in whole or in part, by generative AI tools (including, but not limited to, ChatGPT and other large language models) in work submitted for evaluation in this course, you must document and credit your source. Material generated using other tools should be cited accordingly.
- If you include material generated by a generative AI tool and it is substantively incorrect you will lose points as appropriate. You should verify the accuracy of all content you include in your work.

Sample citation:

- "ChatGPT-4. (YYYY, Month DD of query). 'Text of your query.' Generated using OpenAI. https://chat.openai.com/.
- "ChatGPT-4 (2023, August 9) 'What is a pressing policy issue in the District of Columbia?" Generated using OpenAI. https://chat.openai.com/.

Prohibited:

- While taking an out-of-class ("take-home") or an in-class quiz, a student types a prompt

into a generative AI tool and incorporates some or all of the generated content into their submitted answer.

Be aware of the limits of GAI:

- Generative AI is a tool, but you need to cite it when you use it. Always. No exception. And you are prohibited from using it as stated above.
- It may stifle your own independent thinking, creativity, and understanding of class concepts. Minimum effort into both generative AI prompts and your assignments will produce low quality results. Effectively and correctly using AI in academic work takes time and effort.
- Don't trust anything or everything AI says. If it gives you a number or fact, assume it is wrong un- less you either know the answer or can check it with another non-AI source. This is an opportunity for you to practice your critical analysis skills. As noted above, you will be responsible for any errors of omissions provided by the tool.
- AI tools are based on data that can include biases and reflect historical or social inequities and thus the AI tool can replicate those biases and inequities. Be aware that it can also produce problematic and potentially offensive answers.