INSH 6500: Statistical Analysis

Beatrice Magistro

Fall 2025

E-mail: b.magistro@northeastern.edu
Web: **beatricemagistro.com**
Office Hours: Tue 5-6 p.m/Wed 12-1 p.m.
Zoom: See Canvas page for link
24h response on e-mail
Optional Lab: Thur 5–6 p.m. (Zoom)

Course Description

This course introduces graduate students to quantitative methods essential for rigorous social science research. Through hands-on application of statistical techniques, students will develop the skills to evaluate data-driven claims, design research studies, and conduct their own empirical analyses. The course emphasizes practical data analysis using R, beginning with experimental design as the foundation for causal inference, then extending to observational studies and statistical modeling.

We will approach statistics not as abstract mathematics, but as a practical toolkit for answering substantive research questions. Each concept will be illustrated through real-world examples from political science and related fields, with extensive practice in data manipulation, visualization, and interpretation. By semester's end, students will have the quantitative foundation necessary for advanced coursework and independent research.

Learning Outcomes

Upon successful completion of this course, students will be able to:

- Evaluate the quality and validity of quantitative research in academic publications and policy reports
- Design appropriate research strategies for different types of social science questions
- Manipulate, explore, and visualize data using modern computational tools
- Apply regression analysis and statistical inference to test hypotheses
- Interpret statistical results and communicate findings to both technical and general audiences
- Recognize the assumptions and limitations of various statistical methods
- Implement reproducible research practices using R and RStudio

Required Textbook

- Llaudet, Elena, and Kosuke Imai. *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton University Press, 2022.

Additional Resources

I will also draw material from these free online textbooks:

- Ismay, Chester, and Albert Y. Kim. *Statistical Inference via Data Science: A ModernDive into R and the tidyverse*. CRC Press, 2019: https://moderndive.com/v2/
- Çetinkaya-Rundel, Mine, and Johanna Hardin. *Introduction to Modern Statistics*. OpenIntro, 2021: https://openintro-ims.netlify.app/
- Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. *R for Data Science*. 2nd Edition. O'Reilly Media, Inc, 2023: https://r4ds.hadley.nz/

Prerequisites

The course is suitable for students with a large range of prior exposure to statistics and mathematics. No prior statistical, mathematical, or programming experience is necessary beyond arithmetic, algebra, and elementary calculus. The most important prerequisite is a willingness to work hard on possibly unfamiliar material.

Course Structure

This course is delivered through **weekly Canvas modules**. Each module opens on **Tuesday 09:00am ET** and contains:

1. Video lectures (60 minutes)
2. A 5-question auto-graded mini-quiz (takes about 5 minutes) to be completed by **Sunday 11:59pm ET**

Live Support

- **Lab session (optional):** Thursdays 5–6 om ET on Zoom. Sessions are recorded and will be uploaded within 24 hours. Both the lab skeleton code that you can fill out during the session and the completed version are available in the DSS folder on Canvas.
- **Office hours:** Tuesdays 5–6 pm ET or Wednesdays 12-1 pm ET on Zoom. Outside these windows I answer e-mail within **24 hours** (weekdays).
- **Questions & Answers Help Forum:** Available 24/7 for peer support on Canvas discussion board.

Assessments and Weighting

Component	Notes	Weight
Homework ($4 \times 10\%$)	Short data-analysis assignments	40%
Weekly mini-quizzes	11 quizzes, lowest 1 dropped	5%
Project proposal video + write-up	5-min Panopto video (Week 6) + 500–750-word plan (Week 8)	10%
Peer feedback on proposal videos	Comment on $\geq$ 2 classmates' videos (Week 6)	5%
Final project presentation video	10-min Panopto video (Week 14)	10%
Final project paper	due Tue 9 Dec	30%
Total		100%

Late work : with prior notice no penalty (max. 2); otherwise 50% deduction up to 48 hours, then a zero.

Homework Timeline

Homework #1 : 11:59pm ET Sun 28 Sep 2025 — Data wrangling & visualization

Homework #2 : 11:59pm ET Sun 19 Oct 2025 — Linear regression foundations

Homework #3 : 11:59pm ET Sun 16 Nov 2025 — Probability & inference

Homework #4 : 11:59pm ET Tue 02 Dec 2025 — Multiple regression & visualization

Project Workflow

Tue 07 Oct: Upload a **5–minute Panopto screencast** that pitches your question, data, and intended methods. Post it to the *Proposal Showcase* Canvas discussion board.

Sun 12 Oct: Provide **two substantive comments** on peers' pitches.

Sun 26 Oct: Submit the written proposal (500–750 words).

Tue 02 Dec: Upload a **10–minute** final presentation video.

Tue 09 Dec: Submit the full paper (including a code appendix).

Statistical software

We will use R—a free software environment for statistical computing and graphics—to perform data analysis in this class. To be more efficient, we will be using the RStudio interface. Learning R will be one of the biggest challenges in this course, and you should be prepared to devote some time to this. The only way to learn how to use R is to try it! The only way to learn how to code is to start coding! An optional lab session will be held weekly to help you do your homework as

well as learn coding in R and do data analysis. You are also recommended to read Wickham and Groleman's R for Data Science from cover to cover while coding along every single example in that book.

Course Policies

Academic Integrity

All students enrolled in this course must adhere to Northeastern University's policy on "Academic Integrity." Those of you who are unfamiliar with this policy are expected to read it online at: https://osccr.sites.northeastern.edu/academic-integrity-policy/. Violations of the academic integrity policy will not be tolerated, and will be reported to university officials.

Title IX

Title IX protects individuals from sex or gender-based discrimination in educational programs. Northeastern's Title IX Policy prohibits sexual harassment, sexual assault, relationship or domestic violence, and stalking. In emergencies, call 911. For reporting options and resources: https://www.northeastern.edu/ouec. Please note that I am a university-mandated reporter, meaning that if I am made aware of an incident related to sexual assault, sexual harassment, gender-based harassment, dating or domestic violence, sexual exploitation, or stalking, I am required to report it to the Northeastern University's Office of University Equity and Compliance/Title IX Coordinator.

Support Services (*indicates confidential resource)

- ***University Health and Counseling Services:** Confidential medical and mental health services
- ***Find@Northeastern:** 24/7 mental health support (877.233.9477 US / +1.781.457.7777 International)
- **We Care:** Support network for various student concerns
- **Disability Resource Center:** Ensures equal access for students with disabilities
- ***OPEN:** Confidential support for alcohol, drugs, and sexual violence
- ***Center for Spirituality, Dialogue, and Service:** Spiritual support and guidance
- **Office for University Equity and Compliance:** Investigates discrimination reports

Challenges and Accommodations

Learning is most easily accomplished when you are physically and emotionally at your best. If you are struggling, know that there are resources on campus to help you. If you'd like to share any concerns with me, I'm happy to listen, support you, and help direct you to resources.

The Disability Access Services (DAS) at Northeastern ensures that students with disabilities have equal access to the academic experience at Northeastern and advocates for students with

learning differences. Please share any formal DAS-approved accommodations with me as early in the semester as possible so I can make sure to put those accommodations in place. You can do so by requesting a "professor notification letter" through your DAS portal and following up with me via email. More information can be found here: https://disabilityaccessservices.northeastern.edu/. If you do not have formal accommodations through DAS but would like to request them, you can register here https://disabilityaccessservices.northeastern.edu/incomingandsunregisteredstudents/.

Generative AI Policy

Students are encouraged to use generative AI tools (Claude, ChatGPT, Gemini, etc.) as a complement to learning, not as a substitute. You may use AI to debug code, learn more efficient programming techniques, improve visualizations, and clarify your writing. However, any code you submit must be thoroughly understood and well-commented (ideally every line) to demonstrate comprehension. AI can help you write better code and create better plots, but you cannot use it to complete entire assignments or projects for you—this defeats the learning purpose and is immediately recognizable. For written work, AI may assist with grammar and clarity, but the analysis and ideas must be your own. Remember: you are ultimately responsible for all submitted work and must be able to explain every aspect of your code and analysis. Using AI to avoid engaging with the material constitutes academic dishonesty.

Lecture Outline

Week 1, 09/02: Getting Started

- Welcome video; syllabus tour
- Install R / RStudio
- Post introduction on Canvas discussion board
- Complete Syllabus Quiz (counts as Mini-Quiz 1)

Lab Session: No lab session

Week 2, 09/09: Course Overview and Introduction to R

- Course introduction
- Introduction to R and RStudio
- Llaudet & Imai: 1-1.6
- ModernDive, Chapter 1

Lab Session: Basic R commands

Week 3, 09/16: Data Wrangling and Variables

- Data wrangling fundamentals
- Observations and variables
- Llaudet & Imai: 1.7-1.10
- ModernDive: 3.1-3.6

Lab Session: Data manipulation

Week 4, 09/23: Data Visualization and Randomized Experiments

- Introduction to data visualization
- Estimating causal effects with randomized experiments
- Llaudet & Imai: 2-2.7
- ModernDive: 2.1, 2.2, 2.5, 2.8

Lab Session: Creating visualizations

⋆ **Assignment #1 Due: Sunday, Sep 28**

Week 5, 09/30: Survey Research and Single Variables

- Survey research methods
- Exploring single variables
- Llaudet & Imai: 3-3.4

Lab Session: Exploratory data analysis

Week 6, 10/07: Data Analysis Project Proposals

- Upload presentations of project proposals
- Peer feedback of project proposals

Lab Session: No lab. Work on data analysis proposals

⋆ **Video Presentation of Project Proposal Due: Tuesday, Oct 7**

⋆ **Peer Feedback on Video Presentations: Sunday, Oct 12**

Week 7, 10/14: Relationships Between Variables and Linear Regression

- Exploring relationships between two variables
- Introduction to linear regression
- Llaudet & Imai: 3.5-3.6 and 4-4.4.1 and 4.5
- ModernDive: 2.3, 2.4, 2.6

Lab Session: No lab. Two asynch lectures.

⋆ **Assignment #2 Due: Sunday, Oct 19**

Week 8, 10/21: More Linear Regression

- Predicting binary outcomes using linear regression
- Estimating causal effects with observational data
- Llaudet & Imai: 4.6-4.7 and 5-5.3.1

Lab Session: Linear regression in R

⋆ **Research Project Written Proposal Due: Sunday, Oct 26**

Week 9, 10/28: Multiple Regression and Validity

- Controlling for confounders using multiple linear regression
- Internal and external validity
- Llaudet & Imai: 5.3.2-5.7

Lab Session: Regression diagnostics and interpretation

Week 10, 11/04: Probability and Sampling

- Sampling and probability fundamentals
- Llaudet & Imai: 6-6.6
- ModernDive, Chapter 7

Lab Session: Drop-in hour for research projects

Week 11, 11/11: Nov 11: NO NEW MODULE RELEASED - Veterans Day
Lab Session: No lab. Work on Final Projects
⋆ **Assignment #3 Due: Sunday, Nov 16**

Week 12, 11/18: Statistical Inference

- Confidence intervals and hypothesis testing
- Llaudet & Imai: 7-7.5
- ModernDive, Chapters 8-9

Lab Session: Bootstrap.

Week 13, 11/25: Multiple Regression Applications

- Advanced Visualization with 'marginaleffects'
- Magistro and Wack (2023) Racial Bias in Fans and Officials: Evidence from the Italian Serie A

Lab Session: Lab on applied regression will be pre-recorded. Wednesday-Friday is Fall Break

Week 14, 12/02: Final Project Presentations

- Student Video Presentations of Final Projects

Lab Session: No lab. Work on Final Projects
⋆ **Video Presentation of Final Project Due: Tuesday, Dec 2**
⋆ **Assignment #4 Due: Tuesday, Dec 2**

Week 15, 12/09: Finals Week

⋆ **Final Research Project Paper Due: Tuesday, December 9**