

## **SOC 506 Applied Regression**

### **Spring 2021 Syllabus**

Time: Wednesdays 2:40p - 5:40p ET

Location: <https://emory.zoom.us/j/98343368241>

Instructor: Heeju Sohn, Ph.D.

Email: [heeju.sohn@emory.edu](mailto:heeju.sohn@emory.edu)

Office Hours: Thursdays 4:00-5:00 pm ET or by appointment. [Sign up here.](#)  
(Regular office hours will start on Thursday, February 4)

### **Course Description**

---

This course builds upon the statistical toolkit from SOC 500 Linear Regressions and provides a foundation for conducting and evaluating regression-based works in the social sciences. The first part of the course will cover the topics in conducting transparent and reproducible research. Students will be expected to adopt these research practices throughout the semester. The second part of the course will cover generalized linear models (GLM) that examine non-linear outcome variables. The readings, lectures, and in-class discussion will address each method's mathematical justification, execution, and interpretation using statistical software and application in published articles. The third component of the course will focus on students' in-class presentations and discussions of their research projects.

This course's primary goal is for students to gain fluency in the foundational statistical methods in the social sciences. Fluency denotes the ability to 1) assess the methods' appropriateness to address sociological questions, 2) provide thoughtful reviews to works using these methods, and 3) actively engage in collaborations that use statistical methods. This course aims to provide a broad survey of the most commonly used generalized linear models rather than expert knowledge in any particular approach; each topic is worthy of its own semester-long course.

The success of this course depends critically on active participation. As such, students will be evaluated on their intellectual engagement during class discussions and presentations. Please complete the assigned readings before class. This course also requires students to complete an extended abstract or a preliminary analysis of an empirical paper. The final project will be evaluated on its transparency and reproducibility rather than its methodological sophistication. Stage assignments throughout the semester will allow for opportunities for feedback and revision.

Lastly, the primary focus of this course is not statistical programming. I do not expect you to memorize the commands and options for all the statistical methods that we will cover in this course. Instead, I expect you to understand the theory and assumptions behind the methods so that you will be able to figure out the programming part in the future. We will still heavily rely

on software for analysis and documentation throughout the course. I will provide support for Stata and R.

### **Learning Objectives**

1. Gain fluency in the application of generalized linear models in social science research
2. Adopt practices for transparent and reproducible research
3. Practice thoughtful feedback and collaboration in a working group setting
4. Complete an extended abstract of an empirical article

### **Technology requisitions and expectations**

---

#### **Synchronous class meetings on Zoom**

<https://emory.zoom.us/j/98343368241>

Meeting ID: 983 4336 8241

One tap mobile

+14703812552,,98343368241# US (Atlanta)

+14702509358,,98343368241# US (Atlanta)

Dial by your location

+1 470 381 2552 US (Atlanta)

+1 470 250 9358 US (Atlanta)

+1 301 715 8592 US (Washington D.C)

+1 312 626 6799 US (Chicago)

+1 646 558 8656 US (New York)

+1 253 215 8782 US (Tacoma)

+1 346 248 7799 US (Houston)

+1 669 900 6833 US (San Jose)

Meeting ID: 983 4336 8241

Find your local number: <https://emory.zoom.us/u/aipigSGt4>

We will meet once a week for 180 minutes (with intermittent breaks) via Zoom. Please be prepared to participate in meetings using video and a microphone.

- Guide to get started on zoom: <https://canvas.emory.edu/courses/74713/pages/zoom-user-guide-for-students>
- Zoom tips and etiquette: <https://canvas.emory.edu/courses/74713/pages/zoom-tips-and-etiquette/>

#### **Statistical software**

Students will be using statistical software to complete the assignments and to produce dynamic documents of their research project. Basic familiarity (data manipulation for analysis, producing summary statistics, and running linear regressions) with Stata or R is a prerequisite for this course. You can use either Stata or R based on your experience and comfort level.

## Email

I will aim to respond to emails within two business days. Please include the course number in the subject line. Please submit all assignments through Canvas and not via email.

## Required textbook and readings

---

**Christensen, G., J. Freese and E. Miguel. 2019.** Transparent and Reproducible Social Science Research: How to Do Open Science: University of California Press.

**Hoffmann, J.P. 2016.** Regression Models for Categorical, Count, and Related Variables: An Applied Approach: University of California Press.

In addition to these two books, I will assign journal articles each week. I will post them on Canvas.

## Assignments

---

### Day 1 in-class presentation

On the first day of the course, each student will give an in-class presentation on their general research interests, career goals, and the proposed empirical research project for this course. Please be prepared to present for no more than 10 minutes. Most of the time should be dedicated to discussing your project. You may continue to develop the project that you started in SOC 500 or another statistical course. The purpose of this in-class presentation is for me to get to know you as a scholar and your current research interests. I understand you may change your mind about your project; you do not have to do any of the things you propose on the first day of class. Please consider including the following elements in your presentation.

- Your research interests
- The research question you are thinking of exploring in this class
- The motivation and theory underlying your chosen research topic
- What data you will be using
- Types analyses you are considering
- How you will use the results: part of a book chapter, journal article etc.
- What statistical software you are planning to use in the course (R or Stata)

### Weekly readings

You are expected to complete all assigned readings and attend class ready to discuss the material. The quality of class meetings for both you and your classmates depend on your contribution.

We will be relying on Hoffman (2016) to learn about non-linear regression models. The book uses Stata code examples and output throughout the text. If you are an R user, the equivalent R

code is included in the Appendix. You are, however, not expected to memorize code. As you are reading the book, please focus on the following:

- When is this statistical method appropriate to use?
- What are the underlying assumptions of the method?
- What are the method's strengths and limitations?
- What outputs do they produce, and how do you interpret them?

Each week, you will be assigned journal articles that use the methods in Hoffman 2016. For each article, please prepare to discuss the following topics:

- What is the underlying theory or process that the article aims to study?
- How does the article contribute to the literature?
- How did the authors balance the strengths and limitations of their data and statistical approach?
- How do you interpret the output of their results (tables or figures)?
- How can the authors, editors, and reviewers improve the transparency and reproducibility of the article?

I suggest that you read the textbook first and then read the articles. You may find yourself referring back to the textbook as you analyze the article. Revisiting the textbook will likely help you absorb the material more effectively.

### **Empirical research**

Your term assignment will be centered on an empirical research project. You are not required to use any of the methods covered in this class. You can also use the methods you learned in SOC 500. In this course, I want you to focus on developing ethical and transparent research practices. The fundamental concepts are applicable to both quantitative and qualitative research. While we will be focusing on quantitative research methods in this course, I want you to adapt these practices for your own research approach. The empirical research assignment has several required components:

***Reproducible research notebook:*** This is a document that contains everything that is needed to fully replicate your work, including source to the raw data, code to clean data for analysis, the analysis itself, and code that creates your tables and figures. The document must also detail justifications on why you made certain choices throughout the analysis process (how you dealt with missing data and why, why you chose the model etc.). The completed research notebook will be due at the end of the semester.

***Peer evaluation on research notebook:*** You will exchange your research notebook with a classmate's half-way through the semester. You will provide (and receive for your own work) feedback on the transparency and reproducibility of your classmate's research notebook.

**Working paper presentation:** You will give an in-class presentation on your project. This is a valuable opportunity to share your work with your classmates and receive feedback on your work. The presentation can address any stage of the research process: linking theory to statistical models, interpreting results, running sensitivity analyses. Please prepare a 20-30 minute presentation, and we will dedicate another 15-20 minutes for Q&A and feedback. Working paper presentations will be scheduled each week, starting on about week 10 of the semester.

**Extended abstract:** You will write up an extended abstract of your research findings that includes the motivation for your research question, its contribution to the literature, data and methods, and the preliminary results. You may use this assignment for a conference submission or a grant proposal. The extended abstract will be due at the end of the semester.

## Resources

---

### Resources for Inclusive Learning

We all learn differently, and sometimes we need accommodations. Please let me know if any aspect of the course prevents you from learning or being fully engaged. If you need official accommodations for accessibility or alternative course materials, please utilize the University's services through the Department of Accessibility Services. Link to Emory's Accessibility Service: <http://accessibility.emory.edu/index.html> The site also has an informative section for self-advocacy. (<http://accessibility.emory.edu/students/new-to-oas/self-advocacy.html>)

### Additional programming references

For R users

- Gandrud, C. 2016. Reproducible Research with R and R Studio: CRC Press.
- Cheatsheets: <https://rstudio.com/resources/cheatsheets/>
- R Markdown in RStudio: [rmarkdown.rstudio.com](http://rmarkdown.rstudio.com)
- Citation styles: <https://github.com/citation-style-language/styles>
- <https://yihui.org/knitr/>

For Stata users

- Long, J. Scott. 2009. The Workflow of Data Analysis Using Stata: Stata Press.
- Long, J.S., J. Freese and StataCorp LP. 2006. Regression Models for Categorical Dependent Variables Using Stata, Second Edition: Taylor & Francis.
- Writing Dynamic Markdown Documents Using Stata (Doug Hemken) <https://www.ssc.wisc.edu/~hemken/Stataworkshops/dyndoc%20review/Review.html>

## Honor Code

---

We will follow Emory's code for academic integrity and conduct in this course. Please read Appendix I for properly paraphrasing and quoting another writer. You can also see common

forms of misconduct in Appendices II and III. Please ask if you have any doubts about whether something will violate the policy. Link to Emory's Honor Code:

<http://catalog.college.emory.edu/academic/policies-regulations/honor-code.html>

## **Grading Policy**

---

The expectation is that all students in the course will satisfy all the course requirements and receive an A. If you feel that you are falling behind, please reach out to me sooner than later. I will also let you know if you are not on track to receiving an A.