

INTRODUCTION TO STATISTICAL METHODS (SOC-GA 2332)

FALL 2017

Lecture	Monday 9:30am-12pm, Puck 4156.
Lab	Thursday 9:30am-12pm, Puck 4156.
Instructor	Siwei Cheng (siwei.cheng@nyu.edu) Office hours: Thursday 3pm-5pm
Teaching Assistant	Gerard Torrats-Espinosa (gte214@nyu.edu) Office hours: Thursday 12pm-2pm.

COURSE DESCRIPTION

This course provides a practical introduction to quantitative research methods and basic statistics. Students will gain an appreciation for the logic and techniques of statistical data analysis used in social science research. While the course will cover the goals and mathematical assumptions behind statistical techniques, the emphasis will be on skills for analyzing data and interpreting results.

We will meet for a two-and-a-half-hour lecture session once per week, followed by a lab session. The lecture session is a lecture and discussion of methodological issues and concrete examples of their application. Some of the weeks will have in-class mini-lectures, in which we talk about some practical tips for quantitative data analysis. In the lab session, students will have the opportunity to work on computer data analysis with teaching assistant present. Attendance to the lab sessions is mandatory.

COURSE REQUIREMENT AND GRADING

Course requirements include weekly assignments that usually involve computer data analysis (10% of final grade), in-class exams (40% of final grade, with lowest grade weighted downward), and a final replication project report makes use of statistical analysis (50% of final grade). Students will be assigned to work in groups on lab assignments, but each student must put together their own codes and turn in their own assignments.

Weekly Assignments:

The weekly assignments are designed to give you an intuitive and practical understanding of the theoretical concepts developed during the lecture. The weekly assignments will also be an opportunity for you to start working on the final project. The assignments will have two parts. The first part will ask you to use R to develop the intuition behind the key ideas of that week's lecture (using a simulation, for example). The second part will ask you to

complete a task for the final project. By setting earlier deadlines for smaller parts of the final project, we will ensure that you don't leave the final project for the last week, and most importantly, we will be able to assist you with any problems you may encounter along the way.

Working in teams on the weekly assignments is not only encouraged; it will be mandatory. The profession is increasingly becoming a collaborative environment in which scholars with different skill-sets work together on articles, grant applications, and so on. We want you to start developing this collaborative spirit early in the PhD program. The teams will have 2-3 students, and they will change a few times during the semester. As stated before, you will work in teams, but you will submit your own assignments.

TEXTBOOK AND SOFTWARE

Main Textbooks:

1. Agresti, A. & Finlay, B. *Statistical Methods for the Social Sciences*, 4th or 5th Edition. Upper Saddle River, New Jersey: Prentice Hall.
2. Fox, John (2015). *Applied regression analysis and generalized linear models*, 3th Edition. Sage Publications.

Other Useful Textbooks (for this class and future)

- Wooldridge, Jeffrey M. (2015). *Introductory Econometrics: A Modern Approach*, 6th Edition. Cengage Learning.
- Allison, Paul. *Multiple Regression: A Primer*. Pine Forge Press.
- Angrist, Joshua D. & Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Long, J. Scott. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press.
- Morgan, Stephen L. & Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- Powers, Daniel A. & Yu Xie *Statistical Methods for Categorical Data Analysis*. Emerald Group Publishing.
- Treiman, Donald J. *Quantitative Data Analysis: Doing Social Research to Test Ideas*. John Wiley & Sons.

ADDITIONAL RESOURCES

R

- R is a free software environment for statistical computing and graphics. We will be using R for lab assignments and the final project. Your TA and I will provide guidance on getting started with R, but you should also look through the introductory materials for R after class.
- An Introduction to R: <https://cran.r-project.org/doc/manuals/R-intro.pdf>
- RStudio: <https://www.rstudio.com/>
- DataCamp: <https://www.datacamp.com/>

L^AT_EX

- L^AT_EX (often pronounced as LAH-tek or LAY-tek) is a document preparation system. It is widely used in academia for the communication and publication of scientific documents in many fields.
- To get started on installing and using L^AT_EX, you can go to the online resources at the UCLA statistics site: <http://www.ats.ucla.edu/stat/latex/>.
- If you do not want to install L^AT_EX on your computer, you can use an online LaTeX editor, such as [ShareLaTeX](#), which allows online compiling of projects to PDF format. ShareLaTeX has a free version with which you can create, edit, and save your own LaTeX documents.

FINAL PROJECT DETAILS

Replication is an important part of social science research. It also provides a hands-on method to train your quantitative analysis skills. The final project will contain two parts. First, we ask everyone to replicate the basic descriptive and inferential results from a research article published in a leading sociology journal using publicly available data. You will see the replication article when we introduce the project in class. This part includes downloading and cleaning the data, generating descriptive tables and figures, estimating regression models, and summarizing these results into publication-ready tables. We will ask you to submit your codes and output. Second, you can either replicate more complicated results in that same research paper, or extend its analysis in an interesting and meaningful way. You will also be required to submit a short report that describes the motivations for your analysis and interprets your findings in relevant theoretical context. Details about the final project will be introduced in Week 3.

SUGGESTED TEXTBOOK PROBLEMS FOR PRACTICE

AF Ch.1: 1.7

AF Ch.3: 3.11, 3.17, 3.21, 3.35

AF Ch.4: 4.15, 4.31

AF Ch.5: 5.13, 5.27

AF Ch.6: 6.9, 6.37

AF Ch.7: 7.13

AF Ch.8: 8.7, 8.11

AF Ch.9: 9.9, 9.27, 9.29, 9.43

AF Ch.10: 10.9,10.15,10.33

AF Ch.11: 11.7,11.11,11.15

AF Ch.12: 12.4, 12.6

AF Ch.14: 14.7, 14.11, 14.35, 14.17

AF Ch.15: 15.1, 15.5, 15.9

TENTATIVE SCHEDULE

Week	Main Topic	Suggested Textbook Readings	Exam/Misc
W1 (9/11)	Introduction	AF Ch.1-4; Fox Ch.1	
W2 (9/18)	Statistical inference I	AF Ch.5&6	<i>Exam 1</i>
W3 (9/25)	Statistical inference II	AF Ch.7&8	Introducing the Final Project
W4 (10/2)	Regression I	AF Ch.9; Fox Ch.2	<i>Exam 2</i>
W5 (10/9)	Fall Recess - No Class		
W6 (10/16)	Regression II	AF Ch.10-12; Fox Ch.5 & 6	<i>Exam 3</i>
W7 (10/23)	Regression extended; I	AF Ch.14; Fox Ch.7 & 8	
W8 (10/30)	Regression extended; II	AF Ch.14; Fox Ch.9	<i>Exam 4</i>
W9 (11/6)	Regression extended; III Causality	TBD	
W10 (11/13)	Regression with discrete dependent variables	AF Ch.15; Fox Ch.14	
First part of final project is due on Friday November 17 at 4pm			
W11 (11/20)	Longitudinal data analysis I	AF Ch.16.1-16.3; Fox Ch. 23	<i>Exam 5</i>
W12 (11/27)	Longitudinal data analysis II	AF Ch.16.1-16.3; Fox Ch. 24	
W13 (12/4)	Presenting and Interpreting Statistical Results	TBD	
W14A (12/11)	Extended Topics	TBD	
W14B (12/12)	Extended Topics	Legislative Day - TBD	

****Final project report is due on Wednesday December 20 at 4pm****