

Sociology 362: Statistics for Sociologists III
University of Wisconsin-Madison
Spring 2019

Who and Where?

Lecture Time: Tuesday & Thursday 2:30-3:45PM

Lecture Location: Van Vleck B135

Instructor: Prof. Felix Elwert

Office: 4426 Social Science Building

Office Hours: By electronic appointment via:
<https://felix-elwert.youcanbook.me>

Tuesday 10-11am; Thursday 4-5pm

Email: elwert@wisc.edu

Teaching Assistant: Jungmyung Kim

Office: 7110 Social Science

Office Hours: Wednesday & Friday 9-10am.

Email: jungmyung.kim@wisc.edu

Course Website: <https://canvas.wisc.edu/courses/140490>

Administration Information

Credits: 3

Credit hours are met by 150 minutes of lecture plus 90 minutes of lab per week, plus a minimum of 7 hours of out-of-class student work per week.

Course Designations:

Level – Advanced; L&S Credit - Counts as Liberal Arts and Science credit in L&S; Grad 50% - Counts toward 50% graduate coursework requirement; Honors - Accelerated Honors.

Prerequisites: Junior standing, and Soc 361 or equivalent, or consent of instructor.

Instructional Model: All Face-to-face.

Course Description

This is the third course in the statistics sequence for sociologists. It is required for CAR and sociology Ph.D. students; it is elective for others.

This course aims to give students a keen understanding of the purposes and methods of modern quantitative research in sociology. We will engage with two overarching topics.

1. Learn the key strategies for drawing causal inferences from experimental and observational data (“identification”).
2. Develop a deeper technical understanding of the main statistical techniques for empirical research in sociology (“estimation”).

We will move back and forth between identification and estimation throughout the course, since different identification strategies may require different estimation approaches. For example, when

analysts believe that they can identify causal effects by *adjusting* or controlling for observed covariates, then they will resort to *regression* and *matching* estimators. When adjustment for observed covariates is not sufficient (perhaps because of unobserved confounding factors), they may defend specific *functional form assumptions* and resort to *difference-in-difference* or *fixed-effects estimators*. Alternatively, when they can pinpoint the source of *random variation* in treatment assignment, they may use *regression-discontinuity* or *instrumental variables estimators*.

Throughout, *linear regression* will serve as the traditional point of reference for all estimation topics. In contrast to Soc 361, however, move beyond OLS by considering important extensions for non-metric outcomes (*general linear models*), nested data structures (*hierarchical or mixed models*), and censored data (*tobit*).

We will cover our topics at various levels of technical depth. We will emphasize concepts and intuitions where these are sufficient, and we will additionally develop more formal notation where it is useful for linking to more advanced courses.

Since quantitative research is an interdisciplinary enterprise, we will liberally draw on methodological contributions from econometrics, statistics, biostatistics, political methodology, and computer science.

Learning Outcomes

By the end of this course, you will be able to recognize what statistical techniques are appropriate to extract answers for common sociological questions from data in a variety of realistic scenarios.

Students who want to *produce* quantitative empirical research, will learn the tools to execute competent work and be able to decide which of the more specialized methodological courses they should take next. (*You will* have to take more courses.)

Students who will primarily consume quantitative empirical research will acquire the necessary tools to become critical readers of the majority of quantitative research in sociology.

Grading

Exams (55%): Students will complete 2 midterm exams and one cumulative final exam. Exams cover all material presented in lecture and strongly build on the problem sets in content and form. To respect that everybody can have a bad day, we will give greater weight to your better midterm (15%) and less weight to your weaker midterm (10%). The final exam counts 30%.

There is no possibility of make up midterms. If you have to miss a midterm, please provide a compelling reason *before the beginning of the exam*, in which case your other midterm will count more. If both midterms are missed, the final exam will constitute the entire exam grade.

For each exam, you may bring *one single-sided sheet* of your own handwritten notes (for formulas etc.).

Problem sets (40%): Students will complete nine computational and conceptual problem sets, some of which contain data-analysis exercises. Problem sets are typically due one week after they have been distributed. Late assignments receive a score of zero. You can drop one assignment

without penalty. Assignments will be graded on effort for partial credit and for correctness to achieve full credit.

Attendance (5%): Lecture and section attendance are required and accounts for 5% of the grade. Students may miss a total of 2 lecture and 2 section meetings and still receive full attendance credit. Attendance credit drops to 0% with the fifth absence.

Readings

We post required and recommended readings on the course website. Students must complete all required readings before lecture. Students should revisit their readings, and the lecture script again after lecture. *I cannot stress this enough:* in order to learn statistics, you must read the material *repeatedly*.

This course does not follow any one textbook. Instead, we are assembling the readings from a journal articles, published textbooks, and the *de facto* draft of a textbook (Chris Winship's lecture scripts) to cover the material you need at an appropriate technical level.

Students are encouraged to buy at least one of two excellent and highly accessible books on the logic of empirical research, which focus on foundational insights and guiding intuition that will be useful for both producers and consumers of statistics, whatever their level of technical sophistication.

1. Rosenbaum, Paul. 2017. *Observation and Experiment. An Introduction to Causal Inference*. Harvard University Press: Cambridge, MA.
2. Angrist, Joshua D., and Jörn-Steffen Pischke. 2015. *Mastering Metrics: The Path from Cause to Effect*. Princeton University Press: Princeton, NJ.

Additionally, graduate and CAR students who plan to produce quantitative research should obtain Angrist and Pischke's justly famous *Mostly Harmless Econometrics* (the graduate version of #2 above). Some of the material in this book is presented at a higher technical level than Soc 362, but it is the minimal level to which you will need to rise eventually to produce state-of-the-art quantitative research on your own.

3. Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press: Princeton, NJ.

Several good texts on social science statistics in Stata exist. I recommend:

4. Baum, Christopher F. 2006. *An Introduction to Modern Econometrics Using Stata*. Stata Press: College Station, TX.

How much should you read? My advice is to *read more than required, especially if you struggle with the material*. Reading the same text twice is good; but reading the same material twice from different authors may be even better.

Do not be tricked by apparently low page counts (e.g. 30 pages per week). Reading statistics is different from reading prose. It is important to read slowly and carefully. The devil is in the details. Take note of symbols, formulas, and definitions. Highlight and annotate your text. Excerpt your readings by hand and on paper. It is perfectly normal to spend 1-3 or hours on 10 pages of technical reading.

	24	L2: Potential Outcomes Causality & Randomized Experiments MHE: Chapter 2 * OE: Chapters 1-2
	29	L3: Matching I Readings for the entire week: Stuart (2010) *OE: Chapter 5 *Gerber and Green (2012), Chapter 4
	31	L4: Matching II (Same readings as L3)
February	3	L5: Matching III & Estimation Error * Imai, Stuart, King (2008) pp481-493 Read one of the following two: Iacus, King, Porro (2011) (read for ideas, skip math if you must) Blackwell et al. (2009) (Stata manual for Iacus et al.)
	7	L6: Path Models Pearl 2013: 155-159 *Elwert 2014: 245-51, 256-8.
	12	L7: Technical Tools 1: Properties of Estimators Winship, Lecture Script R1
	14	L8: Technical Tools 2: Relationships Between Distributions Winship Lecture Script R2
	19	L9: Technical Tools 3: Matrix Algebra Winship Lecture Script L4
	21	L10: Regression in Matrix Notation Winship Lecture Script L5
	26	L11: Identification and Misspecification I Winship Lecture Script L6 ** MHE, CH3 pp. 68-77
	28	L12: Identification and Misspecification II Pearl 2013 (remaining pages)
March	5	Review
	7	Midterm I
	12	L13: Robust Inference and GLS Winship Lecture Script 12 *Baum 2006, Chapter 6 * Abadie et al. 2017: pages 1-7.

	14	L14: Mixed Models I Note: The following two required readings overlap in substance, but differ greatly in presentation. You should read all for L14 and L15. Raudenbush and Bryk 2002, Ch. 2 Winship Lecture Script 13 Raudenbush and Bryk 2002, Ch. 3
	19	Spring break (no class)
	21	Spring break (no class)
	26	L15: Mixed Models II Same readings as L14
	28	L16: Limited Dependent Variables Winship Lecture Script 10 * Long, pp34-54
April	2	L17: Logistic Regression Long, pp 61-79, *79-83 Mood 2010 **MHE, Section 3.4.2
	4	L18: Instrumental Variables I MM Chapter 3 *MHE Sections 4.1 & 4.2
	9	L19: Instrumental Variables II TBD
	11	Review
	16	Midterm II
	18	L20: Regression Discontinuity Design MM Chapter 4
	23	L21: Panel Models: DiD and Fixed Effects I MM Chapter 5
	25	L22: Panel Models: DiD and Fixed Effects II TBD
	30	L23: Mediation Analysis * VanderWeele 2015, Ch. 1 VanderWeele 2015, Ch.2
May	2	Review
	8	Final Exam: 2:45-4:45 (Place: TBD)

