

Emory University

RESEARCH METHODS & MODELS: STATISTICS

SOC 500; Fall 2020

Wednesdays, Fridays 9:40 – 12:40

Dr. Timothy J. Dowd
tdowd@emory.edu

Office Hours
Mondays, 1:30 – 3:30

COURSE DESCRIPTION

This an introduction to descriptive and inferential statistics for bivariate and multivariate analyses. The course will help you understand statistics reported in social science publications and in the news media, as well as help you conduct original research. The overall goal is to increase your statistical literacy – your ability to create, interpret, present, and critically evaluate statistical evidence. This is a set of skills that you will find highly useful in your current academic life and in your future career. It is also a valuable set of skills for virtually everyone in modern society, as statistical knowledge (and numerical literacy more broadly) is key for making sense of the growing amounts of information that we encounter in a digital world.

THE ORGANIZATION OF THIS COURSE

I approach this course with certain principles. First, I firmly believe that the use of statistics without a theoretical framework is highly problematic. Indeed, the best sociological scholarship weaves together theoretical concerns with empirical investigation. The language of the syllabus emphasizes the empirical aspects of scholarship, well, because this is a statistics course. But, please know that while we grapple with such things as data, sampling, regression – that is due to the nature of this course rather than my privileging of statistics over theory. (I also teach theory courses, by the way).

Second, I likewise do not privilege quantitative approaches over qualitative approaches. Both are extremely important for scholarship across the disciplines. The quantitative-heavy nature of this course is due to the fact, well, that this is a statistics course.

Third, I recognize that most practicing sociologists rely upon software packages when conducting statistical analyses (I sure do). As a result, this introductory course on statistics also is paired with a particular package – that of *R*. As two of the books emphasize, *R* has many advantages, but it also has a learning curve. We handle that curve in two ways. On the one hand, we ease into *R* by way of descriptive statistics and diagnostics before getting into data visualization and linear / logistic regression. Thus, this course picks up speed, so to speak, as you become more adept at both statistics and the use of *R*.

Finally, I believe that the best way to master statistics is use them in the context of your own research (rather than simply doing the exercises found in a given text). I have set up this class, then, with the opportunity for us to engage in statistical analysis collectively—with several intensive workshops spurring along your progress. My goal is that, rather than produce papers and exams as “busy work” for this course, you all will emerge with one or more projects on which to employ your statistical skills.

COURSE REQUIREMENTS

This course is a seminar, which entails, not only a fair amount of reading, but also class discussion and in-class exercises / analyses. Given that, you are expected to attend each class and to participate.

A) Attendance

If you must miss a class, please inform me ahead of time so that we can make arrangements. Note that unexcused absences will negatively affect your final grade.

Of course, given the current pandemic, some students might become sick or will need to go into isolation or quarantine. If you are sick, understand that **I will be flexible about attendance**. Please make sure to email me so that we can discuss your individual circumstances. For students in quarantine who are well, we have provided ways that you can keep up with your schoolwork, whether our class is delivered online or in person. Please also contact me via email if you are in quarantine.

B) Class Participation and Discussion

Active participation requires adequate preparation. You therefore must read the assigned material before class and practice the skills necessary for statistical analysis and visualization.

Such careful preparation will improve the quality of class discussion. Of course, class discussion should be both informed and respectful; moreover, it should be a forum wherein all can raise questions, explore ideas, and express misgivings.

C) Assignments

You will not be doing assignments for this class in the traditional sense. Of course, you are still expected to develop the necessary knowledge and skills. Indeed, in each of our classes, you will be demonstrating your mastery of the week’s materials. So come prepared!

The actual assignments, instead, will connect course material to your substantive interests. The dates that these assignments are distributed and due are listed below.

D) Final Grade

My starting assumption is that you will earn an “A” in this seminar. If you do the work and contribute to the class, then all is fine. If you fall below expectations, I will let you know, so that you can adjust accordingly. Of course, if you are having difficulties, please contact me.

COURSE RESOURCES

We will be drawing upon five books for the bulk of our readings, books that I will provide for the class. They are as follows:

Paul D. Allison. 1999. *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge.

Whitney Battle-Baptiste and Britt Rusert, Editors. 2018. *W.E.B. Du Bois’s Data Portraits: Visualizing Black America*. NY: Princeton Architectural Press.

John Fox. 2016. *Applied Regression Analysis and Generalized Linear Models*. Thousand Oaks, CA.

Kieran Healy. 2019. *Data Visualization: A Practical Introduction*. Princeton, NJ: Princeton University Press.

Wendy Zeitlin and Charles Auerbach. 2019. *Basic Statistics for the Behavioral and Social Scientists Using R*. New York: Oxford University Press.

We will read additional papers, articles or chapter in addition to our five books. Those will be posted on our Canvas site for SOC 500.

If you have any special needs, please contact me at the beginning of the semester, and we will discuss the necessary arrangements (for additional information, visit the Emory Office of Accessibility Services website; see <http://equityandinclusion.emory.edu/access/index.html>).

COURSE SCHEDULE (Subject to Revision)

August 19: Introductions

August 26: Situating Data, Statistics, & Research

Eduardo Bonilla-Silva and Tukufu Zuberi. 2008. “Toward a Definition of White Logic and White Methods.” Pages 3-27 in *White Logic, White Methods: Racism and Methodology*, edited by Tukufu Zuberi and Eduardo Bonilla-Silva. Lanham, MD: Rowman & Littlefield.

Fox 2016. Chapter One.

Mary Romero. 2020. "Sociology Engaged in Social Justice." *American Sociological Review* 85: 1-30.

September 2: Getting Started with Statistics and Statistical Packages

Allison 1999: Chapter 4.

Maryann Erigha. 2018. "On the Margins: Black Directors and the Persistence of Racial Inequality in Twenty-First Century Hollywood." *Ethnic and Racial Studies* 41: 1217-1234.

Zeitlin and Auerbach 2019. Chapters 1 through 4 and Appendix C.

September 9: Workshop on R

Assignment #1 Distributed

September 16: From Inferential Statistics to Simple Linear Regression

Allison 1999: Chapters 1 and 5.

Timothy J. Dowd & Ju Hyun Park. 2020. "She *Still* Works Hard for the Money: Women Composers, Precarious Work and the Gender Pay Gap." Unpublished paper; Department of Sociology, Emory University.

Zeitlin and Auerbach 2019. Chapters 5 through 9.

September 23: Getting into Multiple Regression (OLS)

Allison 1999: Chapters 2 and 6.

Fox. 2016. Chapters 5 and 6.

Assignment #2 Due.

September 30: When OLS Regression Goes Awry

Allison 1999. Chapters 3, 7, & 8.0 through 8.4.

Fox. 2016. Chapters 12 and 13.

October 7: Depicting Patterns Amidst All That Data

Battle-Baptiste and Britt Rusert 2018.

Healy 2019. Chapters 1 through 5.

October 14: Workshop on Data Visualization

Assignment #2 Distributed

October 21: Depicting Spatial Patterns Amidst All That Data

Healy 2019. Chapters 7 & 8.

Timothy J. Dowd, Trent Ryan, Vaughn Schmutz, Dionne Parris, Ashlee Bledose, and Dan Semenaza. 2020. "Retrospective Consecration Beyond the Mainstream: The Creation of a Progressive Rock Canon." *American Behavioral Scientist*. Forthcoming.

October 28: Categorical Independent Variables and Interactions

Allison 1999. Chapters 8.5 through 8.7.

Fox. 2016. Chapters 12 and 13.

Robert Friedrich. 1982. "In Defense of Multiplicative Terms in Multiple Regression Equations." *American Journal of Political Science* 26:797-833.

Assignment #2 Due

November 4: Outliers and Missing Data

Paul Allison. 2002. *Missing Data*. Thousand Oaks, CA: Sage. Selections.

Fox 2016. Chapter 11.

November 11: Workshop on Missing Data and Imputation

Assignment #3 Distributed

November 18: Getting into Logistic Regression

Fox 2016: Chapter 14

Hemmert, Giselmair A.J., Laura M. Schons, Jan Wieseke, and Heiko Schimmelpfennig. 2018. "Log-Likelihood-Based Pseudo-R² in Logistic Regression: Deriving Sample-Sensitive Benchmarks." *Sociological Methods & Research* 47: 507-531.

Menard, Scott. 2001. *Applied Logistic Regression Analysis*. Thousand Oaks, CA: Sage. Selections.

November 24: Assignment #3 Due