Introduction to Empirical Methods
Ralph Bunche Summer Institute 2018

Professor: David A. Siegel
TAs: Kaitlyn Webster (head) and Gabriel Madson
Course Website: Sakai

Contact Information:
E-mail: david.siegel@duke.edu
Tel: 660-4306
Office: 294H Gross Hall
TA E-mails: kaitlyn.webster@duke.edu and gabriel.madson@duke.edu

Course Description

This course will introduce you to the tools needed to assess and perform empirical research in the social sciences. Its intent is twofold: (i) to get you to be able to read original research in political science; and (ii) to help you produce a piece of original research on your own. The course's level is pitched at somewhere between a good undergraduate class in research methods and an introductory graduate class in quantitative methods. As such, it will undoubtedly be difficult. Many worthwhile things are, though, and I believe learning the material in this course is worthwhile. Not only is the material you will learn central to most graduate programs in the social sciences, it is increasingly in demand across a range of career options. Pretty much any time data can be exploited, these tools will be useful, and the quantity of data to be exploited grows exponentially.

We will begin the course by discussing the logic of science, with a focus on causal relationships and hypothesis generation. Then we will move on to issues of empirical measurement. In so doing we will also have a quick primer on the fundamentals of probability and probabilistic thinking in the social sciences, as well as linear algebra. Next will we tackle the thorny question of statistical inference, which helps in connecting correlations to causal relationships. This part of the course, which lasts two weeks, culminates in the introduction of ordinary least squares and bivariate regression.

The second part of the course expands on the concepts covered in the first. In the third week, we will expand our notion of inference to more than one explanatory variable. While doing so we will discuss the meaning of statistical control, and also have a quick primer on the fundamentals of differential calculus. Then, in the course's last week, we will turn to complications that arise in inference, of the type you might encounter in working on your projects. This includes the difficult question of assessing causal inference, as well as what to do when the assumptions that underlie our models prove to be false. As part of the last topic, we will discuss the concept of a latent variable, and introduce regression with binary dependent variables.

These topics can be abstract, so we will ground them in lab sessions with the TAs. Lab sessions will get you working with real data from the outset, and allow you to put into practice what you have learned during lectures. Taking advantage of these sessions, as well as other opportunities to get help and feedback from both me and the course TAs, will help you get the most out of the course. Three weekly problem sets as well as a final will help to cement understanding as well. Should you begin to feel overwhelmed, though, please come talk as soon as possible. It is entirely reasonable, even expected, to feel that way at some point when learning new and difficult material for the first time, and I and the TAs are here to help. There are lots of additional resources that can help as well, and your patience and investment in the course material will pay you dividends later.

Readings

The course uses two books:

- **(KW)** Kellstedt, Paul M. and Guy D. Whitten. 2013. *The Fundamentals of Political Science Research, 2nd Edition.* Cambridge University Press.
- **(MS)** Moore, Will H. and David A. Siegel. 2013. *A Mathematics Course for Political and Social Research.* Princeton University Press.

We will go through much of **KW**, and use **MS** for our math primers. The latter has an associated set of videos and additional problems you can do located at `http://people.duke.edu/~das76/MooSieBook.html`. These will go into more depth on all topics than we necessarily can in class, and may help in clarifying some topics. There are a multitude of other online sources for both stats and math you can explore as well, should you want more sources. My advice for readings is to skim all assigned material before class, then go back and read more carefully all material that we covered in class, which will necessarily be a subset of that in the readings.

Course Requirements

There are two components to the methods grade. (There is also a methods component of the grade on your final papers, but this is not included here.)

- Problem Sets (50%): There will be three problem sets during the course. Each will be posted to Sakai on Monday by noon, and due the following Monday by noon. They will each cover material from that week's class, as well as the associated labs. Each will be turned in to Sakai's dropbox, and there will be no late problem sets accepted. You should feel free to work together on problem sets, but each person's work must be written up (on a computer, not by hand) independently, and all work must represent an understanding of all problems in the assignment. Within a few hours of noon, solutions to each problem set will be posted to Sakai under the Resources link. After this point you will have two days to correct your own problem set. The intent of this self-grading exercise is to ensure that you go carefully through the solutions. By Wednesday at noon you must post your corrected problem sets to Sakai's dropbox. Note you only need to correct your problem sets, not assign yourself a grade. I will look at both initial and final problem sets and assign a grade then, which I will post to Sakai's gradebook. As long as you evince strong effort on the first submission and are thorough in correcting any misunderstandings in the second submission, you will score high marks on these. *Do not put assignments off to the last minute!* The earlier you start, the more help you can expect.
- Final Exam (50%): There will be a final exam the Tuesday after the last full week of class. It'll be a take home exam in exactly the format of the problem sets, save that you will not grade it yourself, you will have only one work day to do it, and you must do the work yourself. It will cover material from the entire course. You will get access to it by 9 am that morning, and turn it into Sakai's dropbox by 5 pm that day.

Course Outline (Subject to Change with Advance Notice):

Week:

1. Topics: Science, Theory, and Causation; Measurement and Research Design
   Lab Topics: What is Data, and Where to Find It?; Introduction to R
   Readings:
   - (Wed) **KW** Chapters 1-3
   - (Fri) **KW** Chapter 4-5, **MS** 12 (275-282 only)
2. Topics: Statistical Inference and Bivariate Hypothesis Testing; Ordinary Least Squares
   Lab Topics: Data Cleaning and Management; OLS in R
   Readings:
   - (Tue) **KW** Chapters 6-7, **MS** Chapters 9 (175-182 only), 10 (198-230 only)
   - (Thu) **KW** Chapter 8
3. Topics: Introduction to Multiple Regression
   Lab Topics: Interpreting and Graphing Results in R
   Readings:
   - (Thu) **KW** Chapter 9
4. Topics: Differentiation and Matrices; Multiple Regression Model Specification
   Lab Topics: Research Paper Troubleshooting
   Readings:
   - (Tue) **MS** Chapter 5, Chapter 12 (282-300 only), Ch 13 (321-323 only)
   - (Wed) **KW** Chapter 10
   - (Thu) **KW** Chapter 11