Introduction to Empirical Methods
Ralph Bunche Summer Institute

Professor: David A. Siegel
E-mail: david.siegel@duke.edu

Course Description

This course is the online version of a course taught as part of the Ralph Bunche Summer Institute (RBSI). Its purpose is to introduce students to the tools needed to read, assess, and perform empirical research in the social sciences. The course's level is pitched at a level somewhere between a good undergraduate class in research methods and an introductory graduate class in quantitative methods. As such, it will be difficult at times. Many worthwhile things are, though, and I believe learning the material in this course is worthwhile. Not only is the material students 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. Upon completion of the course, students will:

- understand the logic of science and the importance of theory;
- have thought carefully about issues of measurement and causality; and
- be familiar with basic descriptive and inferential statistics.

Course Format

The course material comprises four video lectures, each roughly corresponding to a week of classes in the RBSI. Each of these lectures is composed of a series of shorter modules of varying lengths. While each module is intended to be watched in one sitting, students are encouraged to pause, rewind, and rewatch as desired. All these videos are posted on YouTube and free to access; a link to the channel is provided at the top of this syllabus. You may want to subscribe in case additional content is added later. More resources for this course can be found on the webpage linked to above, including slides for each lecture. The YouTube channel also contains a complete online math course for social scientists which may be particularly helpful the summer before you enter graduate school.

The first lecture starts by discussing the logic of science, with a focus on causal relationships and hypothesis generation. Then it moves on to issues of empirical measurement and research design. In the middle of that there will be a very brief primer on some fundamentals of linear algebra. The first lecture is the longest in terms of time, despite the fact that it has the least math, because it covers the most important topics. Without strong theory, measurement, and research design, statistical tools are of limited use.

The second lecture tackles the thorny question of statistical inference, which helps in connecting correlations to causal relationships. This lecture culminates in the introduction of ordinary least squares and bivariate regression. It also contains a primer on probability and probabilistic thinking in the social sciences.

The third lecture expands our notion of inference to more than one explanatory variable. It also discuss the meaning of statistical control and provides some needed math: a quick primer on the fundamentals of differential calculus and matrix algebra.

The fourth and final lecture turns to complications that arise in statistical inference. That 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, the lecture also discuss the concept of a latent variable, and introduces regression with binary dependent variables.

The topics in this course can be abstract, so I have tried to ground the material in real-world examples. I have also included in each lecture a few practice modules, in which I work through a handful of problems of the type you might encounter in a course. Students should try to do these themselves before watching my solutions. Finally, I have included a few practice problems in pdf form on the course website, along with their solutions. Working through these will help cement course material as well. However, I want to stress that even students doing everything right may at times feel overwhelmed during this course. That is the nature of learning new and difficult material for the first time. My advice is to take a break and come back later when that happens, rewatching material as needed. Becoming comfortable with research methods and statistics can difficult, but your patience and investment in the course material will pay you dividends later.

I have also included a bonus fifth lecture that provides a brief introduction to game theory. This introduction is pitched at the undergraduate level, does not employ calculus, and has a behavioral flavor in that it addresses, albeit very briefly, behavioral game theory, bounded rationality, and lab experiments. While this lecture is not part of what is typically taught in the RBSI, it does expand greatly on the material in the fourth module of the first lecture, and may be useful to some students who want to delve more deeply into theory generation. It is at present not tied to any readings.

Readings

While this online course is intended to be self-contained, it draws its material heavily from two books:

- **(KW)** Kellstedt, Paul M. and Guy D. Whitten. 2018. *The Fundamentals of Political Science Research, 3rd Edition.* Cambridge University Press.
- **(MS)** Moore, Will H. and David A. Siegel. 2013. *A Mathematics Course for Political and Social Research.* Princeton University Press.

The course outline below suggests readings that connect to each of the lectures. For **MS**, used for our math primers, there is 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 the course, 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.

Course Outline (with Links and Readings):

Introduction: `https://youtu.be/N6RRu2kfsBE`

Lecture 1: Science, Theory, and Causation; Measurement and Research Design; Vectors

Lecture 1 Readings: **KW** Chapters 1-5, **MS** 12 (275-282 only)

Lecture 1 Slides Part 1: `https://www.daveasiegel.com/files/2020/07/RBSI_Lecture_1_Part_1.pdf`

Lecture 1, Module 1 (A Science of Politics): https://youtu.be/BG-pQVuxIKY

Lecture 1, Module 2 (Models and Hypotheses): https://youtu.be/k8otYUjB7wk

Lecture 1, Module 3 (Theory Building): https://youtu.be/OyMl_5YthVU

Lecture 1, Module 4 (Formal Theories): https://youtu.be/0haACoNlCpE

Lecture 1, Module 5 (Theory Evaluation): https://youtu.be/_2neTvGTpwM

Lecture 1, Module 6 (Theories and Hypotheses): https://youtu.be/tWIW3hQH4_c

Lecture 1, Module 7 (Causation): https://youtu.be/-pxEseGEZxU

Lecture 1 Slides Part 2: https://www.daveasiegel.com/files/2020/07/RBSI_Lecture_1_Part_2.pdf

Lecture 1, Module 8 (Research Design): https://youtu.be/ey0H3620Ofg

Lecture 1, Module 9 (Observational Studies): https://youtu.be/b7U7hColNK8

Lecture 1, Module 10 (Measurement): https://youtu.be/Zby5NtyFxXs

Lecture 1, Module 11 (Reliability and Validity): https://youtu.be/l-Xg_41OZBQ

Lecture 1, Module 12 (Level of Measurement): https://youtu.be/VgkK25t5xq0

Lecture 1, Module 13 (Population and Samples): https://youtu.be/RZMiPEzgniI

Lecture 1, Module 14 (Scalars and Vectors): https://youtu.be/gU76-NGxb0o

Lecture 1, Module 15 (Descriptive Statistics): https://youtu.be/KlVhdcn_ANc

Lecture 1, Module 16 (Dispersion): https://youtu.be/GOXppjzYGok

Lecture 1, Module 17 (Problem Session): https://youtu.be/pUfjKjiM9Gc

Lecture 2: Statistical Inference and Bivariate Hypothesis Testing; Ordinary Least Squares; Probability

Lecture 2 Readings: **KW** Chapters 6-9, **MS** Chapters 9 (175-182 only), 10 (198-230 only)

Lecture 2 Slides Part 1: https://www.daveasiegel.com/files/2020/07/RBSI_Lecture_2_Part_1.pdf

Lecture 2, Module 1 (Statistical Inference): https://youtu.be/eV8HjDaQ9q0

Lecture 2, Module 2 (Estimates and Uncertainty): https://youtu.be/_OyoTHTcwGo

Lecture 2, Module 3 (Probability): https://youtu.be/g-RyMmqhFl4

Lecture 2, Module 4 (Distributions): https://youtu.be/qpx9o49NK1o

Lecture 2, Module 5 (Hypothesis Testing): https://youtu.be/b1vM60QQnKc

Lecture 2, Module 6 (Categorical Data and the χ^2): https://youtu.be/vr7aUHWz6vc

Lecture 2, Module 7 (Continuous Data and the Correlation r): https://youtu.be/Okb8Ix25wuM

Lecture 2, Module 8 (Difference of Means and Student's t): https://youtu.be/fInBGkc308M

Lecture 2 Slides Part 2: https://www.daveasiegel.com/files/2020/07/RBSI_Lecture_2_Part_2.pdf

Lecture 2, Module 9 (Ordinary Least Squares (OLS) Regression 1): https://youtu.be/EseOI8LUYd8

Lecture 2, Module 10 (OLS Regression 2): https://youtu.be/JW0xvG7Cy0s

Lecture 2, Module 11 (Regression Assumptions): https://youtu.be/EMfB5aKCMFM

Lecture 2, Module 12 (Prediction, Residuals, and Fit): https://youtu.be/uddy13r4DxY

Lecture 2, Module 13 (Problem Session): https://youtu.be/46u01DCPi2U

Lecture 3: Introduction to Multiple Regression; Differentiation and Matrices

Lecture 3 Readings: **KW** Chapter 10, **MS** Chapter 5, Chapter 12 (282-300 only), Ch 13 (321-323 only)

Lecture 3 Slides: https://www.daveasiegel.com/files/2020/07/RBSI_Lecture_3.pdf

Lecture 3, Module 1 (Controlling for Confounding Variables): https://youtu.be/pPYM8c-73WE

Lecture 3, Module 2 (Differentiation): https://youtu.be/nLmErNfDtN0

Lecture 3, Module 3 (Matrices): https://youtu.be/wnw04GsSSh0

Lecture 3, Module 4 (Model Assessment): https://youtu.be/PjotwDEVyp4

Lecture 3, Module 5 (Problem Session): https://youtu.be/xzXPVm6nUJU

Lecture 4: Multiple Regression Model Specification

Lecture 4 Readings: **KW** Chapter 11-12

Lecture 4 Slides: https://www.daveasiegel.com/files/2020/07/RBSI_Lecture_4.pdf

Lecture 4, Module 1 (Variant Regression Specifications): https://youtu.be/u5G1VhSHupA

Lecture 4, Module 2 (Dummy Variables): https://youtu.be/o9NVzS2e9Zo

Lecture 4, Module 3 (Logit): https://youtu.be/rSgsXTmmae4

Lecture 4, Module 4 (Looking Forward): https://youtu.be/hRMCInldvZg

Lecture 4, Module 5 (Problem Session): https://youtu.be/H9LITfNlm3s

Lecture 5: A Brief Introduction to Game Theory

Lecture 5 Slides: https://www.daveasiegel.com/files/2020/08/RBSI_Lecture_5.pdf

Lecture 5 Introduction: https://youtu.be/AGq9WhquNhg

Lecture 5 Module 1 (Building Blocks of Formal Models): https://youtu.be/xfZjn9WeCmA

Lecture 5 Module 2 (Preferences and Rational Choice): https://youtu.be/YFfdrkppcHg

Lecture 5 Module 3 (Nash Equilibrium and Simultaneous Games): https://youtu.be/xgX-vaL8TXQ

Lecture 5 Module 4 (Subgame Perfection and Sequential Games): https://youtu.be/WNb9rtu-Pi4

Lecture 5 Module 5 (Uncertainty and Mixed Strategies): https://youtu.be/9hpgkhekRaE

Lecture 5 Module 6 (Discounting and Repeated Games): https://youtu.be/1sJ-Yj-IC00

Lecture 5 Module 7 (Behavioral Game Theory): https://youtu.be/H4wNhX6O_8c

Lecture 5 Module 8 (Incomplete Information and Bayesian Nash Equilibrium): https://youtu.be/FxRP6haVDAw

Lecture 5 Module 9 (Bounded Rationality): https://youtu.be/QIC2VPcs6IM

Lecture 5 Module 10 (Games and Experiments): https://youtu.be/kh3k-h4N0Jc