

Introduction to Political Science Data Analysis: Political Science 7085
University of Colorado-Boulder

Dr. Vanessa A. Baird, Professor
Office Ketchum 131D; Email Vanessa.Baird@Colorado.edu
Office Hours: Tuesday and Thursday 1-2
(also by appointment)

Course Objectives

The objective of this course is to introduce the methods by which social scientists apply statistics to substantive areas of research. We will examine the elements of research inference, which apply both to qualitative and quantitative designs, but we will focus on being particularly critical of quantitative designs. This class is separated into three themes: 1) causal theory, 2) measurement, 3) multiple variable analysis.

1. Causal theory. The most important yet most difficult job of a social scientist is to explain why things happen. What causes war? What causes people to participate in politics? What causes democratic stability? One way to answer such questions is to look around for other abstract concepts that go together with what you are interested in explaining. Causal theories ought to be as exhaustive as possible; however, your model of the world must not (and actually should not) match the world precisely. Though we will discuss statistical ways of assessing the balance between accuracy and parsimony later in the seminar, developing sound causal theory based on an exhaustive investigation of previous scholarly work is of primary importance in this class.

2. Measurement theory. Since we are interested in explaining why some phenomena vary or change, across people, time or geographical units, we will talk about that variance is distributed across numerical values. Matching an abstract concept to a quantitative coding scheme must always be defended, so we will learn how to devise and assess theories of measurement. We will also learn to evaluate the reliability and validity of the correspondence between the concept and quantitative operationalization.

3. Multiple variable analysis. The question that we continue to ask as we develop causal theories is whether there are alternative explanations for our variable of interest that we have not considered. Therefore, we will be spending much of our time looking at statistical methods that assess the manner and extent to which variables correspond while controlling for other variables. Since the objective of any scientist is to disprove their causal argument, an important part of what we learn is how to estimate the error in our estimation. This will be a major part of the focus in this section of the class. In the later part of the class, we will learn how to manipulate explanatory variables to incorporate the context into our models. In other words, rather than asking what causes something, we will begin to ask whether there is anything about the particular context that ___ or mitigates that causal relationship.

Course Structure

The structure of the course is applied; you should be spending about 50% of your time learning and understanding statistics and the other 50% of your time will be associated with applying those

statistics to real questions. Some of the former will be in class (though you will need to reinforce your understanding of statistics on your own) and most of the latter will happen outside of class, in the Class Assignments and on your paper. I expect that you will spend at least 15 hours per week on this class. Each lecture builds substantially on previously covered material. You cannot afford to miss class. If you do not understand something from the class, ask questions. If you still do not understand, you should take the initiative to visit me or the TAs during office hours. It is your responsibility to ask questions if you do not understand something from the lecture. This is a difficult class and I do not expect you to understand everything the first time. I have extra readings that I will make available in case the text is not clear on some issues.

Course Requirements

It should go without saying that attendance is absolutely mandatory. Your Class Assignments will consist of 25% of your grade (graded in terms of \checkmark , $\checkmark+$, $\checkmark-$ and 0). If I note that students are not doing the reading, then pop quizzes will accompany the lab grades. The final exam will count for 25% of your grade. The paper is 50% of your grade. The paper is divided into three sections, the research design, measurement, and analysis. Each subsequent section includes revisions of previous sections. 10% of your paper grade will consist of your comments on another student's paper. Each portion of the paper will be handed into me by email.

Texts

I find that people learn in many different ways, and so I have readings on the topics from multiple textbooks in my office that I will make available in addition to the required texts for the course. You are responsible for the material and if you have no problem understanding the material from my lectures and the texts, then you are not required to read the additional available materials. However, if you are having any difficulties at all, I expect you to read the extra material.

McClendon, McKee. 2002. *Multiple Regression and Causal Analysis*. Peacock.

McIver, John P., and Edward G. Carmines. 1981. *Unidimensional Scaling*. Beverly Hills: Sage Publications.

Carmines, Edward G., and Richard A. Zeller. *Reliability and Validity Assessment*. Beverly Hills, Calif.: Sage Publications.

Pollock, Philip H. *The Essentials of Political Analysis* Congressional Quarterly.

Pollock, Philip H. *A Stata Companion to Political Analysis*, Congressional Quarterly.

Course Outline

Topic	Assignment
Introduction	Overview of the course.
History of causality, overview of the discipline	<p>Class Assignment Due: Search through APSA program and PROceedings, and ICPSR for potential research problems and respective data. You will report your findings in class.</p> <p>Reading: McClendon, 1-8; Pollock, Essentials 77-86</p>
Spurious relationships	Reading: McClendon, 8-18; Pollock, Essentials 77-86
Types of variables	<p>Class Assignment Due: Come up with your own example of a spurious relationship.</p> <p>Discussion of potential paper topics. Each person is expected to come to class to report on either an interesting data set or research project that you found in a literature search.</p>
Central tendency, variance, standard deviation	<p>Class Assignment Due: Download a dataset from ICPSR as well as the codebook. Make a list of concepts that would be available in that data set. You should have a paper topic at this stage and optimally, should be downloading a dataset that you can use for your paper topic.</p>
Normality, probability, central limit theorem	Reading: McClendon, 20-28; Pollock, Essentials, Chapter 5
Crosstabs Analysis	Reading: Stone and Davis: An Introduction to Quantitative Methods, Pollock, Essentials Chapters 3 and 4. You should have a topic outline and an exhaustive bibliography for your research design at this stage.
Introduction to Bivariate Regression: Covariance, Correlation, Expected Values	Reading: Pollock Chapter 2
Bivariate Regression: Fitting a Line	Reading: McClendon 28-41, Pollock, Essentials 154-162
Monday September 24th	FIRST PAPER DUE BY NOON BY EMAIL. 15 PAGE LIMIT
Properties of Estimators	Reading: McClendon 28-41; Pollock, Essentials, chapter 6

Elements of Computing Bivariate Analysis	Reading: Pollock, Essentials, Chapter 1
Reporting Bivariate Results	
Creating Indices; Guttman Scales	Reading: <i>Unidimensional Scaling</i>
Validity	Reading: <i>Reliability and Validity</i> , Chapters 1 and 2
Reliability	Reading: <i>Reliability and Validity</i> , Chapters 3 and 4
Missing data and recoding	Reading: Cohen and Cohen chapter
Introduction to Multivariate Regression	Reading: McClendon 60-80, Pollock, Essentials, 168-178
The Logic of Controlling	Reading: Re-read McClendon 60-80; Pollock, Essentials, 168-178
Multivariate Regression: Residuals	Reading: McClendon 94-118
Monday October 29th	SECOND PAPER DUE AT NOON BY EMAIL. 30 PAGE LIMIT
Interpreting Multivariate Output	Reading: Re-read McClendon 94-118
Review Multivariate Regression	
Dummy Variables	Reading: McClendon 198-214
Quadratic Terms	Reading: McClendon 230-256
Interactions I: Interpreting Dummy Variable Interactions	Reading: McClendon 271-281
Interactions II: Interpreting Ordinal and Categorical Interactions	Reading: McClendon 281-287

Calculating Interactions	
Evaluating and Correcting for Missing Data	
Evaluating Regression Models: I	READING TBA
Evaluating Regression Models: II	READING TBA
Monday December 11th	FINAL EXAM DUE AT NOON BY EMAIL. 50 PAGE LIMIT.
Evaluating Regression Models: III	DECEMBER 12TH, 14TH, PAPER PRESENTATIONS

FINAL EXAM DUE Monday, DECEMBER 10, 5pm
FINAL PAPER DUE Monday, DECEMBER 17, 5pm