Introduction to Applied Statistics in Political Science: Description, Comparison, Estimation, Inference

Jake Bowers
jwbowers@illinois.edu
Online: http://jakebowers.org/

Fall 2016

Overview

When/Where We meet 3:30 to 5:50pm in 404 David Kinley Hall. Moodle https://learn.illinois.edu/course/view.php?id=19118

Office Hours Jake will have office hours 1:30–2:30 Mondays and Thursdays. Please make an appointment if you want to come to office hours to ensure that we can meet and talk.

Introduction This is the start of your practice with the skills and concepts that are commonly used by political scientists to make arguments about how observation can teach us about specific implications of theory. This is a practice because the field constantly changes and because the subject is so deep and important that no one can ever truly master it. We are all always learning. So this course is to help you start the learning that you will continue for your whole life as an academic.

As a practice, statistics involves skills and concepts. If you don't have the skills, then the concepts are not concrete and are difficult to understand. In most statistics PhD programs the skills involve the mathematics of linear algebra and calculus and mathematical problem solving skills involved in deductive proofs and algebraic manipulation. In this course, we are going to use a flexible computer programming system in lieu of math in order to demonstrate to ourselves and make concrete the concepts that we must internalize and use in order to apply statistics to help us learn about the world and about our theories. (Math, after all, is a language. R, the language we will be using, is another way to express abstract ideas.) A by product of using R to engage with statistical concepts is that you'll also practice how to use R to solve data problems. That is, you will start to practice some of the basics of "data science" on the way to practice some of the basics of statistics (which is the discipline devoted to learning how to do science).

This course is also a graduate course in applied statistics or political methodology. This means that it exists within a series of continuous developments within multiple disciplines. The contents of this course will change over time because the disciplines change over time. What we thought worked well in the past, may not be what we think works well today. What we teach in this course today may be seen by future scholars (we hope) as old fashioned and suboptimal. That is, this course is just like any other phd level graduate class: we engage with the past in order to do things differently in the future. This class does not aim, therefore, to teach you to do political science as it was done in the 1950s, 1960s, 1970s, or even last year. It is a tradition-based but future oriented class just like all of your other classes.

Goals and Expectations

This class aims to help you learn to think about what it means to do statistical inference for both descriptive and causal claims.

The point of the course is to position you to do the future learning that is at the core of your work as an academic analyzing data.

I also hope that this course will help you continue to develop the acumen as a reader, writer, programmer and social scientist essential for your daily life as a social science researcher.

Expectations First and foremost, I assume you are eager to learn. Eagerness, curiosity and excitement will impel your energetic engagement with the class throughout the term. If you are bored, not curious, or unhappy about the class you should come and talk with me immediately. Energetic engagement manifests itself in meeting with your classmates outside of the class, in asking questions during the class, and in taking the term paper seriously.

Second, I assume you are ready to work. Learning requires work. As much as possible I will encourage you to link practice directly to application rather than merely as a opportunity for me to rank you among your peers. Making work about learning rather than ranking, however, will make our work that much more difficult and time consuming. You will make errors. These errors are opportunities for you to learn — some of your learning will be about how to help yourself and some will be about statistics. If you have too much to do this term consider dropping the course. Graduate school is a place for you to develop and begin to pursue your own intellectual agendas: this course may be important for you this term, or it may not. That is up for you to decide.

Third, I assume you are willing to go along with my decisions about the material and sequence. I will be open to constructive and concrete suggestions about how to teach the class as we go along, and I will value such evaluations at any point in the class. I have made changes to this course in the middle of the term upon hearing great and useful ideas from students. I am happy to do so. That said, if you do not think you need to take this course, then don't take it.

Fourth, I assume some previous engagement with high school mathematics.

Rules There aren't many rules for the course, but they're all important. First, ask questions when you don't understand things; chances are you're not alone. Second, don't miss class or section. Third, do the work. This does not mean divide the work up among your classmates so that you only do part of the work. Each person should engage with all of the work even if the people who writes it up changes from week to week.

All papers written in this class will assume familiarity with the principles of good writing in Becker (1986).

All final written work will be turned in as pdf files unless we have another specific arrangement.[1] I will not accept Microsoft, Apple, or any other proprietary format.

Late Work I do not like evaluation for the sake of evaluation. Evaluation should provide opportunities for learning. If you think that you and/or the rest of the class have a compelling reason to change the due date on one of those assignments, let me know in advance and I will probably just change the due date for the whole class.

Incompletes Incompletes are fine in theory but terrible in practice. I urge you to avoid an incomplete in this class. If you must take an incomplete, you must give me *at least* 2 months from the time of turning in an incomplete before you can expect a grade from me. This means that if your fellowship, immigration status, or job depends on erasing an incomplete in this class, you should not leave this incomplete until the last minute.

Participation We will be doing hands-on work nearly every class meeting. I will lecture very little and instead will use the assignments given out the previous week to raise questions about statistical theory, research design, and data, which will require us to confront and apply the reading that prepared us for the day's work. I will break away to draw on the board or demonstrate on my own computer now and then if everyone is running into the same problem or is asking a question that raised by the work.

Explorations Every week or so, I will ask you to complete a short assignment that encourages you to engage creatively with the topics of interest. I anticipate that you will work on these assignments in groups and that each of you will come to class prepared to discuss them. I don't think that the groups should have more than 3 people in them. However, I'm willing to have larger groups if you talk with me about it.

You will need to turn it in the day before class at 5pm so that I can read it prepare for class.

[1]For example, if you have some reason why pdf files make your life especially difficult, then of course I will work with you find another format.

Daily R Five days per week, you will need to practice writing R code. I will require that you make a Github Gist https://help.github.com/articles/about-gists/ each day written in either R or R markdown format in which you load a dataset from the web, and learn something of interest to you about the units represented by those data. I'm imagining 2 to 10 lines of R code. Then you will paste the url to that gist into a Moodle Journal or a Google Spreadsheet (not sure which yet). Each gist must be written so that it runs from start to finish on any computer — not just yours.

Each week I will choose at random two gists to discuss in class.

Grades are Feedback Humans need feedback to close the intention to action gap. They also need feedback to feel good about their progress and to motivate them. In this class I will use grades as feedback. All grades except for the final grade will be satisfactory, unsatisfactory (with the possibility of "outstanding"), and fail. These map roughly onto A=satisfactory, C=unsatisfactory, and F=fail (i.e. you didn't try).

I'll calculate your grade for the course this way: 20% daily R (you have 5 days out of every 7 to turn it in, no late work accepted, satisfactory if you turned it in, fail if you didn't turn it in); 10% daily R in class discussion (satisfactory if the gist runs from start to finish on my computer or on a randomly chosen classmate's computer, unsatisfactory otherwise, fail if there is no gist from you when your gist is chosen, no late work accepted); 30% explorations (everyone in the group receives the same grade, satisfactory if you are creative and thoughtful and diligent, unsatisfactory if you are not or if you don't seem to be getting the concepts, no late work accepted); 30% in class participation (satisfactory if you ask good questions that show that you have thought about the material, a good question can be a simple question; unsatisfactory if your questions show that you are not doing the reading and/or are not actively involved in the explorations; fail if you are not in class); 10% attendance (satisfactory if you show up, fail if not).

You can miss two classes without grade penalty although I suspect you'll be sad to miss the discussion.

Because moments of evaluation are also moments of learning in this class, I do not curve. If you all perform at 100%, then I will give you all As.

Involvement Quality class participation does not mean "talking a lot." It includes coming to class; turning in assignments on time; thinking and caring about the material and expressing your thoughts respectfully and succinctly in class. As much as possible, we will be working in groups. This work will require that you come prepared to meetings with your classmates and also to the class itself and that you are an active collaborator. Involvement also means meeting with your classmates at least once per week outside of class to complete the work begun in the class.

This class is an opportunity to practice courage: I expect you to make a guess when I ask a question (in writing or in person), I expect that you will ask a question when you have a problem.

Computing We will be using R in class so those of you with laptops available should bring them. Of course, I will not tolerate the use of computers for anything other than class related work during active class time. Please install R (http://www.r-project.org) on your computers before the first class session.

Computing is an essential part of modern statistical data analysis — both for turning data into information and for conveying that information persuasively (and thus transparently and reliably) to the scholarly community. In this course we will pay attention to computing, with special emphasis on understanding what is going on behind the scenes. You will be writing your own routines for a few simple and common procedures: your own likelihood functions, your own least squares solvers, your own bootstrapping and permutation statistical inference routines, and your own posterior distribution for Bayesian statistical inference.

Most applied researchers use two or three computing packages at any one time because no single language or environment for statistical computing can do it all. In this class, I will be using the R statistical language. You are free to use other languages, although I suspect you will find it easier to learn R unless you are already a code ninja in some other language that allows matrix manipulation, optimization, and looping.

As you work on your papers, you will also learn to write about data analysis in a way that sounds and looks professional by using either R markdown or Sweave (R+LaTeX). No paper will be accepted without a code appendix or reproduction archive attached (or available to me online). No paper will be accepted unless it is in Portable Document Format (pdf).[2] No paper will be accepted with cut and pasted computer output in the place of well presented and replicable figures and tables. Although good empirical work requires that the analyst understand her tools, she must also think about how to communicate effectively: ability to reproduce past analyses and clean and clear presentations of data summaries are almost as important as clear writing in this regard.

Books

I'm am not requiring any particular books this term. The readings will be drawn from a variety of sources. I will try to make most of them available to you as we go if you can't find them easily online yourselves.

Recommended No book is perfect for all students. I suggest you ask around, look at other syllabi online, and just browse the shelves at the library and used bookstores to find books that make things clear to you. I will be adding some recommendations here. Let me know now if you have favorites.

Self-Help

If you discover any books or websites that are particularly useful to you, please alert me and the rest of the class about them. Thanks!

Schedule

Note: This schedule is preliminary and subject to change. If you miss a class make sure you contact me or one of your colleagues to find out about changes in the lesson plans or assignments.

The idea behind the sequencing here is to start as simple as possible and complicate later.

Data: I'll be bringing in data that I have on hand. This means our units of analysis will often be individual people or perhaps political or geographic units, mostly in the United States. I'd love to use other data, so feel free to suggest and provide it to me — come to office hours and we can talk about how to use your favorite datasets in the class.

Theory: This class is about description, estimation, comparisons, and inference. Yet, statistics as a discipline exists to help us understand more than why the linear model works as it does. Thus, social science theory cannot be far from our minds as we think about what makes a given data analytic strategy meaningful. That is, while we spend a term thinking a lot about how to make meaningful statements about statistical inference, we must also keep substantive significance foremost in our minds.

0—August 23— Overview, Statistics, Data, Variables

Introductions.

Introduction to the class and my thoughts about statistics. Your thoughts about statistics. Your thoughts about learning difficult topics?

To start Daily R we will need datasets. We need to talk about which datasets might be of interest to the class and get them online by the end of Wednesday.

What is the point of statistics? What is data? What is a variable

I Description

1—August 30—Description in One Dimension

What makes a description useful or not useful? What is a good description? How would we know whether we have a good one or a bad one? Are there descriptions that are particularly robust to observations we'd like to ignore or mostly ignore because they are apt to mislead us?

[2] Actually, I'm willing to consider HTML or Postscript although practice with pdf will help you most in submitting papers to journals and other forms of scholarly communication.

Reminder: Bring laptops if you have them. Those bringing laptops should have R installed.

Read: **Henceforth, "*" means "recommended" or "other useful" reading. The readings not marked with "*" are required.**

*Wilcox, 2012, Chap 1–3

*Kaplan, 2012, Chap 2–3

2—September 6—Description in Two Dimensions I

Topics: Linear data models and fitting.

Questions: Why use straight lines to describe relationships? On what basis should we choose a straight line (why choose one straight line over others)? How to interpret slopes and intercepts (i.e. the descriptors of a line) given different ways of choosing lines? What is the relationship between a straight line and differences between two groups? When might we want to identify and perhaps diminish the influence of particular individual observations on an overall linear description?

Read: *Achen, 1982, Chap 2

*Kaplan, 2012, Chap 4,5–8

*James et al., 2013, Chap 2–3

*Berk, 2008, Chap 1

*Berk, 2010 (especially level 1 regression)

*Wilcox, 2012, Chap 5 and 10

3—September 13—Description in Two Dimensions II

Topics: Nonlinear data models and fitting

Questions: What other criteria for line choice might we have? Re-interpreting linear description as both smooth and comparison. How to argue that you have smoothed the data appropriately? How does the use of dummy variables or indicator variables and/or interaction terms allow us to engage with theoretical expectations that are not linear? What about transformations of predictors like simple polynomials? Or piecewise fits (smooth ones using splines, or not smooth ones using indicators and interactions)? Or locally smooth fits?

Read: *James et al., 2013, Chap 3,7[3]

See also `http://www.med.upenn.edu/beat/docs/Jacoby2000.pdf` and `https://socserv.socsci.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Nonparametric-Regression.pdf`

II Creating Interpretable Comparisons

4—September 20—Conditional description and Causal Counterfactuals, Statistical Adjustment I

Topics: Counterfactual approaches to formalizing causal statements; potential outcomes

[3] See `http://www-bcf.usc.edu/~gareth/ISL/` for materials

Question: What do we mean when we say "Z caused Y"? What is the counterfactual interpretation of this statement? What is the role of variables that we observe but which are not Z or Y? How might we use them to get more clear on the Z to Y relationship?

Read: *Brady, 2008 for an excellent overview and then a discussion of Neyman's "average treatment effects" engagement with the fundamental problem of causal inference.

*Rosenbaum, 2010, Chap 1 and 2

*Rosenbaum, 2002, Chap 1 and 2

*Imbens and Rubin, 2015, Chap 1 and 2

*Dunning, 2012, Chap 1 and 2

5—September 27—Experiments, Adjustment by Design

Topics Randomization as a way to exclude alternative explanations. Randomization as a method for producing a reasoned basis for statistical inference about causal effects / hypotheses.

Questions Why randomize? What should we do if we discover that one variable out of many is imbalanced? Why rely on averages? Why not use other measures comparing experimental treated to control units.

Reading *Kinder and Palfrey, 1993

*Gerber and Green, 2012, Chap 1 to 3

*Angrist and Pischke, 2014, Chap 1

*Fisher, 1935, Chap 2

*Fujiwara and Wantchekon, 2013 See also `http://egap.org/content/brief-13-reducing-clientelism-benin`

6—October 4—Statistical Adjustment II: Matching Methods

Topics: Matching methods as a way to do the subsetting/stratification-based approaches but when one wants to adjust for multiple background variables in a way that also allows for assessment of the question: "Did you adjust enough?"

Reading: *Rosenbaum, 2010, Chap 1,3,7,8,9,13 (`http://www.springerlink.com/content/978-1-4419-1212-1/contents/`)

*Gelman and Hill, 2007, Chap 9.0–9.2 (on causal inference and especially interpolation and extrapolation)

*Hansen, 2004 on full matching for adjustment

*Hansen and Bowers, 2008 on assessing balance.

8–October 11–Statistical Adjustment III

Topics: Covariance adjustment, adjustment by linear model (interaction terms for stratification, or direct covariance adjustment aka "controlling for"), What does "controlling for" do? When does it make most sense? When does it make least sense?

Reading. *Berk, 2008, Pages 1–8[4]

*Berk, 2004, Chap 6–7 (skipping stuff on standardized coefs)

*Achen, 2002 and *Achen, 2005 (on the problem of kitchen sink regressions)

*Fox, 2008, Chap 11 on Overly Influential Points

[4] `http://www.library.uiuc.edu/proxy/go.php?url=http://dx.doi.org/10.1007/978-0-387-77501-2`

7—October 18—No Class

9—October 25—Statistical Adjustment IV

Topic: Difference in differences; instrumental variables[5]

Read: Find a resource you like on difference in differences. (For example, Angrist and Pischke, 2009, Chap 5.2 or Angrist and Pischke, 2014, Chap 5)

*Gelman and Hill, 2007, Chap 10 (`http://www.stat.columbia.edu/~gelman/arm/chap10.pdf`)

*Imai and Kim, 2016 (`http://imai.princeton.edu/research/files/FEmatch.pdf`)

*Bertrand et al., 2004

*Sovey and Green, 2011

*Gerber and Green, 2012, Chap 5

Other Topics: The Placebo-Control Design for the CACE/LATE (see for example, Gerber and Green, 2012, Chap 5.9 and Nickerson, 2005 and Nickerson, 2008 and the recent application in Broockman and Kalla, 2016).

12—November 1—Statistical Adjustment V

Topics: Regression discontinuity designs

Read: Find a resource you like on regression discontinuity designs. (For example, Angrist and Pischke, 2009, Chap 6).

*Sales and Hansen, 2016

*Skovron and Titiunik, 2015

*Cattaneo et al., 2016

*Sekhon and Titiunik, 2016

Other Topics: Natural Experiments (see Dunning, 2012)

III A Taste of Statistical Inference: Hypothesis Tests

10—November 8—Statistical Inference I: Hypothesis Tests

Topics: What is a p-value? What did R.A. Fisher think a p-value was for? How did he propose to create a p-value? How should one interpret at p-value? What do we need to assume in order to calculate the p-value for the sharp null hypothesis of no treatment effects in a randomized experiment?

Read: *Fisher, 1935, Chap 2

*Gerber and Green, 2012, Chap 3

*Rosenbaum, 2010, Chap 1 & 2

*Rosenbaum, 2002, Chap 2

[5] We will talk about placebo control designs and regression discontinuity designs later if the class is up for it.

*Imbens and Rubin, 2015, Chap 2 & 5

IV Applying what you learned: Your Own Data

11—November 15—Data Day I

Topics: Apply the tools and concepts of this term to a substantively interesting data example. Come to class ready to distribute and/or present your ideas.

November 22—No Class

12—November 29—Data Day II

Topics: Apply the tools and concepts of this term to a substantively interesting data example. Come to class ready to distribute and/or present your ideas.

13—December 6—Class Choice

Topics: How to prepare for PS 531? How to spend Winter break?

13—December 13—Class Choice, Optional Extra Class

Topics:

Read:

V References

Achen, C. H. (1982). *Interpreting and Using Regression*. Sage, Newbury Park, CA.

Achen, C. H. (2002). Toward A New Political Methodology: Microfoundations and Art. *Annual Review of Political Science*, 5:423–450.

Achen, C. H. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science*, 22(4):327–339.

Angrist, J. and Pischke, J. (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton Univ Pr.

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'metrics: the path from cause to effect*. Princeton University Press.

Becker, H. S. (1986). *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article*. University of Chicago Press.

Berk, R. (2004). *Regression Analysis: A Constructive Critique*. Sage.

Berk, R. (2008). *Statistical learning from a regression perspective*. Springer.

Berk, R. (2010). What you can and can't properly do with regression. *Journal of Quantitative Criminology*, pages 1–7.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Brady, H. E. (2008). Causation and explanation in social science. *Oxford handbook of political methodology*, pages 217–270.

Broockman, D. and Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282):220–224.

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., and Keele, L. (2016). Interpreting regression discontinuity designs with multiple cutoffs. *The Journal of Politics*, 78(3):000–000.

Dunning, T. (2012). *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press.

Fisher, R. (1935). *The design of experiments. 1935*. Oliver and Boyd, Edinburgh.

Fox, J. (2008). *Applied regression analysis and generalized linear models*. Sage.

Fujiwara, T. and Wantchekon, L. (2013). Can informed public deliberation overcome clientelism? experimental evidence from benin. *American Economic Journal: Applied Economics*, 5(4):241–255.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gerber, A. and Green, D. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Hansen, B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467):609–618.

Hansen, B. and Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23:219.

Imai, K. and Kim, I. S. (2016). When should we use linear fixed effects regression models for causal inference with panel data?

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.

Kaplan, D. (2012). *Statistical Modeling A Fresh Approach*. Daniel Kaplan, Macalester College, St. Paul, MN, second edition.

Kinder, D. and Palfrey, T. (1993). On behalf of an experimental political science. *Experimental foundations of political science*, pages 1–39.

Nickerson, D. (2005). Scalable protocols offer efficient design for field experiments. *Political Analysis*, 13(3):233.

Nickerson, D. (2008). Is voting contagious? evidence from two field experiments. *American Political Science Review*, 102(01):49–57.

Rosenbaum, P. R. (2002). *Observational Studies*. Springer-Verlag, second edition.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer.

Sales, A. and Hansen, B. B. (2016). Limitless regression discontinuity. *arXiv preprint arXiv:1403.5478*.

Sekhon, J. S. and Titiunik, R. (2016). On interpreting the regression discontinuity design as a local experiment. *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics*, 38.

Skovron, C. and Titiunik, R. (2015). A practical guide to regression discontinuity designs in political science. Technical report, working paper, University of Michigan.

Sovey, A. J. and Green, D. P. (2011). Instrumental variables estimation in political science: A readers guide. *American Journal of Political Science*, 55(1):188—200.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing*. Academic Press.