Political Science 532
Quantitative Political Analysis II (aka "the regression class")

Jake Bowers
jwbowers@uiuc.edu
http://www.umich.edu/~jwbowers/ps532.html

Fall 2007

Info Class meets in Lincoln 398 1:00-3:20 Tuesdays. I will hold office hours in Lincoln 495 4:00-6:00 Wednesdays.

Goals This course aims to help you develop the reasoning and skills necessary to use linear models and extensions thereof in social science research. Some previous engagement with probability and statistical computing is assumed (see, for example, the syllabus for PS531).

Reasoning and skills here mean reading, writing, mathematics, statistics, computation, typesetting, and social scientific intuition. All of these abilities are important in this class.

There aren't many rules for the course, but they're all important. First, read the assigned readings before you come to class. Second, turn everything in on time. Third, ask questions when you don't understand things; chances are you're not alone. Fourth, don't miss class or section.

GRADING

I'll calculate your grade for the course this way: 24% on 4 problem sets; 55% for an original research paper; 15% on course involvement; 6% from a confidential evaluation of your contributions to your problem set workgroup.

The Paper

This class offers you the opportunity to apply the concepts and skills you learn here to the social science topic of your choice. The goal is to draft a paper that has some potential to be published in a scholarly journal. Most such papers start out life with a bunch of mostly informal methodological appendices (that many scholars don't even think of as appendices, but just as auxiliary files). In this class, these appendices will be formal and explicit — thus you'll have a paper that you could submit for publication, but in these appendices, you'll also have written extensively about your methodological choices, showed the many alternate methods you tried, and displayed how you thoroughly convinced yourself that you had covered every possible objection to your research design and data analysis. You'll even provide all of the code and data required to completely replicate your entire article!

There will be a series of paper assignments throughout the class which are to be completed and turned in as individuals (of course, because they will be about your own paper). The overall paper grade will consist, in part, of the grades on those assignments.

Problem Sets

How are you going to learn how to write such a paper? To execute such analyses? To typeset such a manuscript? You will learn mainly from the problem sets and from each other. Each of you will work in a group of roughly 4 people. The groups will turn in problem sets, the individuals will not. However, each individual in each group will be required to take a turn at being the spokesperson for that group (i.e. the person who compiles and typesets the final report and uploads it and/or prints and hands it to me). So, your grade on the problem sets will be identical to your group's grade.

Working in groups in this way will force you to learn some social and technological skills related to collaboration (we'll talk about version control systems and file sharing). It will also require you to convince each other of the answers to different problems — teaching others is always the best way to learn.

Finally, it makes my job easier as your one and only grader. I would prefer to spend my limited time reading drafts of your papers, your outlines, your paper proposals, and all of the other intermediate steps leading to your paper than to spend it grading many problem sets analyzing datasets chosen by me, with answers that I already know (rather than having you teach me something new about the political world).

Course Involvement

Quality seminar participation does not mean "talking a lot." It includes turning in assignments on time; thinking and caring about the material and expressing your thoughts respectfully and succinctly in class. It also means doing well on quizzes and providing excellent and constructive written comments on the drafts of your colleagues' work.

Confidential Peer Evaluations of Workgroup Performance

At the end of the term, each member of each group will submit a grade representing the group involvement of the other members of the group. I will keep these peer evaluations strictly confidential. These grades are one more small way to encourage everyone to try to pull more than their own weight during problem set sessions.

Books

I'm erring here on requiring fewer books rather than more, and listing several others as recommended.

Required books:

> Fox, John. 1997. *Applied Regression Analysis, Linear Models, and Related Methods.* (see `http://socserv.mcmaster.ca/jfox/Books/Applied-Regression/index.html`)
>
> Achen, Chris. 1982. *Interpreting and Using Regression.*

And recommended books in alphabetical order:

> Becker, Howard S. 1986. *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article.*
>
> Berk, Richard. *Regression Analysis: A Constructive Critique*
>
> Faraway, Julian. 2004. *Linear Models with R*
>
> Faraway, Julian. 2005. *Extending the Linear Model with R*
>
> Fox, John. 2002. *An R and S-PLUS Companion to Applied Regression* (see `http://socserv.mcmaster.ca/jfox/Books/Companion/index.html`)
>
> Freedman, David, Robert Pisani, and Roger Purves. *Statistics, Fourth Edition* For an excellent basic introduction to statistics in general.

Gelman, Andrew and Jennifer Hill. 2006 *Data Analysis Using Regression and Multilevel/Hierarchical Models* (see `http://www.stat.columbia.edu/~gelman/arm/`).

Gonick, Larry. *Cartoon Guide to Statistics*

Lancaster, Tony. *An Introduction to Modern Bayesian Econometrics*

Rice, John. *Mathematical Statistics and Data Analysis, 3rd Edition* For a more theoretical coverage of mathematical statistics.

Venables, William N. and Ripley, Brian D. *Modern Applied Statistics With S-PLUS, 4th ed* I learned S and R from earlier editions of this book.

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.

You should also have at least one math book on your shelves. Some general recommendations for books that combine linear algebra and calculus among other topics: Carl Simon and Lawrence Blum's *Mathematics for Economists* is a nice math book as is Alpha Chiang's *Fundamental Methods of Mathematical Economics*, and Jeff Gill just came out with *Essential Mathematics for Political and Social Research*.

Computing Computing is an essential part of modern statistical data analysis — both for producing persuasive information from data and for conveying that information to the scholarly community. So 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. All homeworks will be turned in with an appendix that I can run (not cut and paste, but submit as a batch job to an interpreter or compiler) to replicate your analyses. 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 which allows matrix manipulation, optimization, and looping. You will also learn to write about data analysis in a way that sounds and looks professional by using either a WYSIWYG system like Word, OpenOffice, or Wordperfect, or a typesetting system like LaTeX, to produce documents that are be suitable for correspondence, collaboration, and publication.

Schedule

8/28, Tuesday—Preview and Review

Homework: Hand out Paper assignment 1.

9/4, Tuesday—Least Squares as Minimization and as Projection

Reading: Fox Chap 5,6

Homework: Paper assignment 1 due. Hand out Problem Set 1.

9/11, Tuesday—OLS as MLE, Classical Assumptions required for Probabilistic Inference

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

Homework: Sign up for paper meeting with Jake.

9/18, Tuesday—Review of Hypothesis Testing

Reading: Gonick, L. and Smith, W. (1993). *The Cartoon Guide to Statistics*. HarperPerennial, New York, NY, Chapters 5-8 *and* Fox, Appendix D.

Homework: Problem Set 1 Due. Hand out Mini-Computing Assignment 1, and Paper Assignment 2.

Paper Assignment 2 is due in two weeks. The mini-computing assignment is due in class in one week.

9/25, Tuesday—Probabilistic Inference for Plain Old Ordinary Least Squares (i.e. hypothesis testing and confidence intervals for $\hat{\beta}$

Reading: Fox 6,9,10

Homework: Mini-computing assignment 1 due.

10/2, Tuesday—Power of Hypothesis Tests, Goodness of Fit and Interpretation of the Linear Model

Reading: Re-read Fox where he talks about the F-test and Residual Sums of Squares, etc..; Re-read the Cartoon Guide about Hypothesis Testing and Confidence Intervals.

Homework: Paper Assignment 2 due.

10/9, Finish Goodness of Fit and Begin: What to do when we don't believe our predictions? Introducing Factors, Dummies and Interaction Terms

Reading: Fox Chap 7, 14.2.2 (optional 14.2.3 and 14.4); Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14:63–82 (see also `http://homepages.nyu.edu/~mrg217/interaction.html`).

10/16, Tuesday–How to detect and handle Unusual and Influential Data

Reading: Fox, Chapter 11, Chapter 14.3.

Homework: Hand out Problem Set 3.

10/23, Tuesday—(a) Specification and "Controlling For" and (b) What to do when we don't believe our standard errors (or our t-statistics or our p-values)? (hint: Try Bootstrapping and Adjust with WLS, GLS)

Reading: Fox 12, 14.1, 16; *and* Freedman, D. A. (2006). On the So-called "Huber Sandwich Estimator" and "Robust Standard Errors". *The American Statistician*, 60(4):299–302.

10/30, Tuesday—What to do when we don't believe our standard errors (or our t-statistics or our p-values)? (hint: Model the dependence structure with a multilevel model.)[1]

Reading: Faraway, J. (2006). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. CRC Press, Chapter 8 and 9 *and* Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Chapter 11 and 12 and 13.1 - 13.5

Homework: Problem Set 3 due.

11/6, Tuesday— GLMs 1 (and the Linear Probability Model): What to do when we know $Y \in \{0, 1\}$?

Reading: Fox 15, Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Chapter 5

11/13, Tuesday—GLMs 2: What to do when we believe that $Y \sim$ Poisson or Negative-Binomial or something else?

Reading: Fox 15.4, Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Chapter 6

Homework: Hand out Problem Set 4.

11/20, Tuesday—NO CLASS—Thanksgiving Break.

11/27, Tuesday—Instrumental Variables and 2SLS

Reading: Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Chapter 9 and 10; Berk (2004), Chapter 5 and 11.

[1] There is also at least one other way other than modeling the dependence structure but we won't cover it in this class: Rosenbaum, P. R. (2002). Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, 17(3):286–327

Homework:

12/4, Tuesday—Conclusion

Reading: Does anyone in the class have reading or topics that they'd like us to engage with for this last day of class?

Homework: Problem Set 4 due

Final Papers due 12/12