

Syllabus

GOV 1000/2000 Quantitative Methods for Political Science I

Professor: Adam Glynn
TFs: Iain Osgood and Maya Sen
Fall Semester 2009

Class Room

Room N-354, CGIS
2-4 PM
Tuesday

Office

Adam Glynn
Room N-305, CGIS
Phone: (617) 496-2426
Email: aglynn@fas.harvard.edu

Iain Osgood
Email: osgood2@fas.harvard.edu

Maya Sen
Email: msen@fas.harvard.edu

Preliminaries

Overview and Class Goals

This is a course on quantitative political methodology, by which we mean a course on the application of statistical methods to problems in political science. In this course, we focus on regression models (primarily linear regression). However, the principles learned in this course provide a foundation for a general understanding of quantitative political methodology. If you ever want to collect data, analyze data, critically read an article which presents a data analysis, or think about the relationship between theory and the real world, then you will (hopefully) find this class useful.

You can only learn statistics by doing statistics. In recognition of this fact, the reading for this course will be minimal, while the homework for this course will be extensive. However, you may find it helpful to read the material multiple times (before and after the corresponding lecture).

This class has two course numbers available. Gov 2000 is designed explicitly for graduate students in the Government Department. Gov 1000 is appropriate for undergraduates that have taken a course in introductory statistics. All students will receive the same lectures, but Gov 2000 students will be responsible for extra homework and examination problems.

Class Requirements

Grades will be based on

- weekly homework assignments (50 % of final grade)
- a take-home early term exam (10 % of final grade)
- a cumulative take-home final exam (30 % of final grade)
- participation and presentation (10 % of final grade).

The weekly homework assignments will consist of analytical problems, computer simulations, and data analysis. They will be assigned on Tuesday night and due the following Tuesday, prior to lecture. No late homework will be accepted except in the case of a documented emergency. All sufficiently attempted homework¹ will be graded on a (+,√,-) scale, and may be re-written and re-graded once. The re-write is due

¹The instructor will determine sufficiency in borderline cases. All sufficiently attempted homework will be typed and well organized with all problems attempted.

before the Monday lecture one week after the assignment is returned. I encourage students to work together on the assignments, but you should write your own solutions (this includes R code), and I strongly suggest that you make a solo effort at all the problems before consulting others.² The take-home early term exam will be handed out one week after we finish the lectures on elementary probability and univariate statistics. The take home final exam will be handed out one week before the last Friday of reading period. Ten percent of the grade will be awarded for class participation and for the quality of presentation on the homework assignments and the exams.

You should bring pencil and paper to class because I will give small in class assignments throughout the course. These assignments are intended provide you with feedback on your level of understanding. Although the assignments will not be graded, your participation in this activity is required.

I will not give incompletes in this course.

How To Do Statistical Homework (Efficiently)

Do not leave statistical homework until the last minute! You will find that some aspects of the homework can take hours to resolve by yourself (e.g. your code won't run), but will take minutes to resolve with the help of a TF. Plan to spend some time on the homework before the discussion section and before each of the TF office hours so you can ask questions if you get stuck.

Discussion Sections

Sections will be held on Thursday evenings from 6:00-7:30pm in K-354. The section will cover a review of the theoretical material and then focus on computing issues. Iain Osgood and Maya Sen will run the sections and can provide more details.

Computation

We will use the R computing language in this course. The math pre-fresher course gave a good introduction to the language, and the website for the pre-fresher course at: <http://people.hmdc.harvard.edu/~mathpre/> provides a number of valuable resources.

The R language is completely open-source, so you can download it for free from <http://www.r-project.org/>. Chris Green's R primer <http://www.stat.washington.edu/cggreen/rprimer/> provides explicit downloading instructions for all operating systems. R will also be installed at HMDC and on the FAS computing cluster. It is recommended that students use VNC to connect to the FAS cluster.

Course Website

The course website is located at the following URL:

<http://www.courses.fas.harvard.edu/2281>. This site will provide homework assignments, datasets, and supplementary materials.

Course Mailing List

Information about the course mailing list is available at the course web-site under the "WWW Links" section. All students should subscribe to the list. This an ideal forum for posting questions regarding the course material and/or R. I encourage students to reply to each other's questions, and a student's respectful and constructive participation on the mailing list will count toward his/her class participation grade.

Office Hours and Availability

My office hours will be Wednesday from 10:00am - noon this semester in room N-305 of CGIS, and I am also available at other times if you would like to make an appointment. Iain Osgood will hold office hours from 3:00pm-5:00pm every Friday. Maya Sen will hold office hours from 10:00am-noon every Monday. If you have questions about the course material, computational issues, or other course-related issues please do not

²We also ask that you write the names of your co-workers on your assignments.

hesitate to set up an appointment with either me, Iain, or Maya.

If you have a question, you should try to send it first to the course mailing list. This is almost always the fastest way to get an answer. However, you can also email me directly at aglynn@fas.harvard.edu. If the question is of general interest, I will forward the question and my answer to the list. Make sure to tell me explicitly in your email if you'd like to stay anonymous.

Required Books

Gelman, Andrew and Hill, Jennifer. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

We will not cover the entire book, but the portions of the book that we will cover are close in spirit to the lectures, and the uncovered portions of the book will be a good resource for future study.

Gill, Jeff. 2006. *Essential Mathematics for Political and Social Research*. 1st Edition. 2nd printing. New York: Cambridge University Press.

This book was required for the prefrasher, and we will use some chapters.

Suggested Books

In the past, we have used the 1997 Fox textbook (see below) because this text has the most thorough coverage of the required material and because the companion R text (Fox 2002) can be helpful for the computing portions of the course. However, some past students have preferred the presentation in Wooldridge (see below). You are also welcome to utilize any of the many other available regression or econometrics texts, but I will only provide suggested reading assignments from Fox and Wooldridge.

Fox, John. 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.

Fox, John. 2002. *An R and S-PLUS Companion to Applied Regression*. Thousand Oaks, CA: Sage.

Wooldridge, Jeffrey. 2000. *Introductory Econometrics*. New York: South-Western.

Weisberg, Sanford. 2005. *Applied Linear Regression*. 3rd Edition. Hoboken, NJ: John Wiley.

In some cases, more recent editions may be available, and these will be acceptable for the course. In what follows I refer to Fox's Applied Regression Analysis book as Fox97 and the R companion as Fox02.

Optional Books

The following books are optional but may prove useful to students looking for additional coverage of some of the course topics. Freedman et al. will be extremely valuable to students with little or no background in statistics. Many of the concepts that Fox and Wooldridge take for granted are discussed clearly and in detail (with very little math) in Freedman et al.

Freedman, David; Robert Pisani; and Roger Purves. 1998. *Statistics*. 3rd Edition. New York: Norton.

Agresti, Alan and Finlay, Barbara. 1997. *Statistical Methods for the Social Sciences* Upper Saddle River, NJ: Prentice Hall.

Cleveland, William S. 1993. *Visualizing Data*. Summit, NJ: Hobart Press.

- Achen, Christopher. 1982. *Interpreting and Using Regression*. Thousand Oaks, CA: Sage.
- Weisberg, Sanford. 2005. *Applied Linear Regression*. 3rd Edition. Hoboken, NJ: John Wiley.
- Simon, Carl and Blume, Lawrence. 1994. *Mathematics for Economists*. New York: Norton.
- Kennedy, Peter. 2003. *A Guide to Econometrics*. 5th Edition. Malden. Blackwell.
- Wonnacott, Thomas H. and Ronald J. Wonnacott. 1990. *Introductory Statistics*. 5th Edition. New York: Wiley.
- Dalgaard, Peter. 2002. *Introductory Statistics with R*. New York: Springer.
- Venables, W.N. and B.D. Ripley. 2002. *Modern Applied Statistics with S-PLUS*. New York: Springer
- Gonick, Larry and Smith, Woollcott. 1993. *The Cartoon Guide to Statistics* New York: Harper.
- Tufte, Edward. 2001. *The Visual Display of Quantitative Information, 2nd Edition*. Cheshire, CN: Graphics Press.
- Hastie, Trevor; Robert Tibshirani; and Jerome Friedman. 2001. *The Elements of Statistical Learning*. New York: Springer.

Preliminary Schedule

The “**Required Reading**” should be completed prior to the lecture for which it is assigned. The “Suggested Reading” presents alternative presentations of the required material. The “Optional Reading” provides supplementary material that may be helpful or interesting. *This schedule is subject to change.*

1 Introduction - Sept. 8

Topics covered

- Overview and Course Requirements
- Examples of Statistical Inference in Political Science
- Course Outline

Required Reading

Gelman/Hill Chapter 1

Gill, Chapter 7

Suggested Reading

Fox97, Chapter 1

Fox02, Preface, Sections 1.1 and 1.3

Wooldridge, Chapter 1

Optional Reading

Agresti and Finlay, Chapters 1 and 2

Freedman, et al., Chapters 1 and 2

2 Some Probability Useful for Statistics - Sept. 15

Topics covered

- Elementary Probability Theory
- Random Variables and Functions of Random Variables

Required Reading

Gelman/Hill Section 2.1

Gill, Chapter 8 (with the exception of 8.3.5, 8.3.8, 8.9)

Suggested Reading

Wooldridge Appendix B

Fox97 Appendix D.1-D.4

Optional Reading

Gonick and Smith, Chapters 2-5

Freedman et al., Chapters 13-15

Wonnacott & Wonnacott, Sections 3.1-3.6, 4.1-4.5, 5.1-5.4, 6.1-6.3

Diaconis, Persi and Mosteller, Frederick. 1989. "Methods for Studying Coincidences." *JASA* 84: 853-861.

Gelman, Andrew; Gary King; and John Boscardin. 1998. "Estimating the Probability of Events that Have Never Occurred: When is Your Vote Decisive?." *JASA*. 93: 1-9. (available via JSTOR)

3 Univariate Statistical Inference - Sept. 22

Topics covered

- Sampling Distributions
- Point Estimation
- Interval Estimation
- Small and Large Sample Properties of Estimators
- Hypothesis Testing

Required Reading

Gelman/Hill Sections 2.2-2.5, 7.1

Excerpts from Wooldridge Appendix C

Suggested Reading

Agresti and Finlay, Sections 4.3-5.3, 6.1 - 6.5

Fox97 Appendix D.5-D.5.2

American National Election Studies website (<http://www.electionstudies.org/>)

Optional Reading

Gonick and Smith, Chapters 6-8

Freedman et al., Chapters 16-26

4 What is Regression Analysis? - Sept. 29

Topics covered

- Summarizing and Plotting Bivariate Data
- Review of Joint and Conditional Distributions
- Review of Conditional Expectation
- Non-parametric Regression
- Bias-Variance Tradeoff
- Three Uses for Linear Regression

Required Reading

Excerpts from Fox97, Chapter 2

Tatem, Andrew J; Carlos A. Guerra; Peter M. Atkinson; and Simon I. Hay. 2004. "Momentous Sprint at the 2156 Olympics." *Nature* 431 (30 September): 525. (available at course website)

Suggested Reading

Cleveland, Chapter 2 pp. 16 - 33 and Chapter 3 pp. 86-101

Optional Reading

Achen, Chapter 2

Tufte, Edward. 2001. *The Visual Display of Quantitative Information, 2nd Edition*. Cheshire, CN: Graphics Press.

Cleveland, William S. and Robert McGill. 1987. "Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data." (with discussion) *JRSS A*. 150: 192-229. (available via JSTOR)

Cleveland, William S. and Robert McGill. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *JASA*. 79: 531-554 (available via JSTOR)

Cleveland, William S.; Persi Diaconis; and Robert McGill. 1982. "Variables on Scatterplots Look More Highly Correlated When the Scales are Increased." *Science*. 216: 1138-1141 (available via JSTOR)

5 Simple Linear Regression in Scalar Form - Oct. 6

Topics covered

- Simple Correlation and Least Squares Regression
- Simple Dummy Variable Regression
- Derivation and Formulas
- Properties of Point and Interval Estimators
- Hypothesis Tests
- Simple Regression Diagnostics and Transformations

Required Reading

Gelman/Hill Chapters 3 and 4 (but especially Sections 3.1,3.4,3.6,4.1,4.3-4.5), Section 7.1

Suggested Reading

Fox97, Sections 5.1 and 6.1

Fox02, Section 4.1.1

Wooldridge, Chapter 2

Fox97, Sections 4-4.4

Fox02, Section 3.4

Osberg, Lars and Timothy Smeeding. 2004. " 'Fair' Inequality? An International Comparison of Attitudes to Pay Differentials" *American Sociological Review*, 71: 450-473.

Optional Reading

Freedman et al., Chapters 8-12

Wonnacott & Wonnacott, Chapters 11 and 12

Tufte, Edward. 1974. *Data Analysis for Politics and Policy*. Englewood Cliffs, NJ: Prentice-Hall. Chapter 3.

Stigler, Steven M. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, MA: Harvard University Press. Chapters 1,8

6 Linear Regression with Two Explanatory Variables - Oct. 13

Topics covered

- Dummy Variables and Continuous Regressors
- Interaction Terms
- Correlation and Least Squares Regression
- Added Variable Plots
- Formulas and Intuition

Required Reading

Gelman/Hill Chapters 3 and 4, Section 7.2

Suggested Reading

Fox97 5.2, 5.3

Fox97, Sections 7-7.3.3

Fox02, Section 4.1, 4.2

Wooldridge, Chapter 3

Treisman, Daniel. 2000. "The causes of corruption: a cross-national study" *Journal of Public Economics* 76: 399-457.

7 Causality with Least Squares Regression - Oct. 20

Topics covered

- Neyman-Rubin Causal Model
- Nonparametric Structural Equation Models
- Linear Regression Causal Model
- Identification of Causal Effects

Required Reading

Gelman/Hill Chapter 9

Suggested Reading

Fox97 6.2-6.5

Wooldridge 3.1-3.3

Optional Reading

Holland, Paul. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81: 945-960.

Glymour, Clark. 1986. "Comment on Statistics and Causal Inference: Statistics and Metaphysics." *Journal of the American Statistical Association* 81: 964-966.

Pearl, Judea. 2000. "Epilogue: The Art and Science of Cause and Effect." in *Causality*. Cambridge University Press.

8 Multiple Regression in Matrix Form - Oct. 27

Topics covered

- Matrix Algebra
- Vector Calculus
- Derivation of the Least Squares Estimator

Required Reading

Gill Chapters 3, 4

Suggested Reading

Fox97, Appendices C.1, C.3

Fox97, Sections 9.1 pp. 204-206, 9.2

Wooldridge Appendix D

Wooldridge Appendix E

Optional Reading

Freedman et al., Chapters 26-29

Achen, Chapter 3

Simon and Blume. 1994. *Mathematics for Economists*. Sections 8.1-8.4

Cleveland. Chapter 4

9 Statistical Inference for Least Squares Regression - Nov. 3

Topics covered

- Properties of the Least Squares Estimator
- Gauss-Markov Theorem
- Testing and Intervals for Linear Combinations of the Parameters
- Confidence Intervals and Ellipses

Required Reading

Fox97, Sections 9.4

Wooldridge, Sections 4.4-4.5

Suggested Reading

Fox02, Sections 4.5 and 4.6

Braumoeller, Bear. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization*. 58: 807-820.

Optional Reading

Kennedy, Appendix A, 418-422

Freedman et al., Chapters 26-29

Achen, Chapter 4

10 Diagnosing and Fixing Problems I - Nov. 10

Topics covered

- Nonconstant Error Variance
- Weighted Least Squares
- Heteroskedasticity-Robust Standard Errors
- Correlated Errors
- Generalized Least Squares

Required Reading

Gelman/Hill Chapter 11

Fox97, 12.2, 14.1

Fox02, Web Appendix: "Time-Series Regression and Generalized Least Squares"

Suggested Reading

Fox02, Sections 6.3

Wooldridge, Chapters 8

Beck, Nathaniel and Katz, Jonathan. 1995. "What to do (and not to do) with Time Series Cross-Sectional Data." *APSR*. 89: 634-647. (available via JSTOR)

Optional Reading

Berk, Kenneth. 1998. "Regression Diagnostic Plots in 3-D" *Technometrics*. 40: 39-47.

Weisberg, Sanford. 2005. *Applied Linear Regression*. 3rd Edition. Hoboken, NJ: John Wiley.

Wooldridge, Chapters 10-12

11 Diagnosing and Fixing Problems II - Nov. 17

Topics covered

- Leverage Points
- Outliers
- Influence Points
- Nonnormality
- Nonlinearity

Required Reading

Gelman/Hill Chapter 8

Excerpts from Weisberg Chapter 5

Suggested Reading

Fox97, Chapter 11

Fox02, Section 6.1

Wooldridge, Chapters 9.4

Jackman, Robert W. 1987. "The Politics of Economic Growth in the Industrial Democracies, 1974-80: Leftist Strength or North Sea Oil?" *The Journal of Politics*, Vol. 49, No. 1, pp. 242-256. (available via JSTOR)

Wand, Jonathan; Kenneth Shotts; Jasjeet Sekhon; Walter Mebane; Michael Herron; and Henry Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *APSR*. 95: 793-810.

Fox97, 12.1,12.3,12.4, 12.7

Fox02, Sections 6.2, 6.3, and 6.4

Beck, Nathaniel and Simon Jackman. 1998. "Beyond Linearity by Default: Generalized Additive Models." *AJPS*. 42: 596-627. (available via JSTOR)

Optional Reading

Weisberg, Sanford. 2005. *Applied Linear Regression*. 3rd Edition. Hoboken, NJ: John Wiley.

Hastie, Trevor; Robert Tibshirani; and Jerome Friedman. 2001. *The Elements of Statistical Learning*. New York: Springer.

Hastie, Trevor and Rob Tibshirani. 1990. *Generalized Additive Models*. London: Chapman & Hall.

12 Preview of Advanced Causal Inference - Nov. 24

Topics covered

- Balance and Overlap
- Matching
- Regression Discontinuity

Required Reading

Gelman/Hill Chapter 10.1-10.4

Eggers, Andrew; Hainmueller, Jens. 2009 "MPs for Sale? Returns to Office in Postwar British Politics."
APSR, in press.

Suggested Reading

Fox97, Section 6.4

Wooldridge, Chapter 9.3, 15