

# Syllabus

## GOV 1000/2000/2000e/E-2000 Quantitative Methods for Political Science I

Professor: Adam Glynn  
TFs: Andy Hall and Konstantin Kashin  
Fall Semester 2012

**Class Room**  
Room S-020, CGIS  
2-4 PM  
Tuesdays

**Office**  
Adam Glynn  
Room N-305, CGIS  
*Email:* aglynn@fas.harvard.edu

Andy Hall  
*Email:* hall@fas.harvard.edu

Konstantin Kashin  
*Email:* kkashin@fas.harvard.edu

## Preliminaries

### Overview and Class Goals

How can we detect voting irregularities? What causes individuals to vote? How do electoral institutions affect the number of political parties? Quantitative political scientists address these questions and many others by using and developing statistical methods that are informed by theories in political science. In this course, we provide an introduction to the tools used in basic quantitative political methodology. The first half of the course covers descriptive inference with univariate statistics and linear regression from a sampling perspective, with some discussion of missing data. The second half of the course covers causal inference with linear regression and standardization, also from a sampling and missing data perspective. The principles learned in this course provide a foundation for the future study of more advanced topics in quantitative political methodology.

While the tools of statistical inference are worth studying in their own right, the primary goal of this course is to provide graduate students (and some undergraduates) with the necessary skills to critically read, interpret, and replicate the quantitative content of many political science articles. As such, the statistical methods covered in this course will be presented within the context of a number of articles. Throughout the term, we will reanalyze the data and revisit the conclusions from:

- “Racial Prejudice and Attitudes Toward Affirmative Action,” by James H. Kuklinski, Paul M. Sniderman, Kathleen Knight, Thomas Piazza, Philip E. Tetlock, Gordon R. Lawrence, and Barbara Mellers, *The American Journal of Political Science*, 1997.
- “Measuring Political Preferences,” by Lee Epstein and Carol Mershon, *American Journal of Political Science*, 1996.
- “Female Socialization: How Daughters Affect Their Legislator Fathers’ Voting on Women’s Issues,” by Ebonya L. Washington, *The American Economic Review*, 2008.
- “Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment,” by Alan S. Gerber, Donald P. Green, and Christopher W. Larimer, *The American Political Science Review*, 2008.

- “Law and Data: The Butterfly Ballot Episode,” by Henry E. Brady, Michael C. Herron, Walter R. Mebane Jr., Jasjeet Sekhon, Kenneth W. Shotts, Jonathan Wand, *PS: Political Science and Politics*, 2001.
- “Electoral Institutions, Cleavage Structures, and the Number of Parties,” by Amorim Neto and Gary W. Cox, *American Journal of Political Science*, 1997.
- “MPs for Sale? Returns to Office in Postwar British Politics,” by Andrew Eggers and Jens Hainmueller, *The American Political Science Review*, 2009.

### The 1000, 2000, 2000e, and E-2000 Course Numbers

**GOV 2000** is designed for students who already have some background in statistics/mathematics/computing, or for beginners who are looking for a challenge. Students taking this section of the course will learn to be flexible data analysts, capable of tailoring standard methods to the unique specifications of each task. As such, these students will be asked in the problem sets to write/adjust the code necessary to replicate and critique results from the literature. This section of the course will be taught within the R statistical computing environment.

**GOV 2000e** is designed for students with a limited background in statistics/mathematics/computing. Students in this section of the course will focus on the analysis and critique portions of the assignments. This section of the course will be taught with the Stata statistical software, and the students will be provided with the additional code necessary to replicate and critique results from the literature.

**GOV 1000** is intended for undergraduate students and will be taught with the Stata statistical software as the default option. Undergraduate students may choose to use R instead, but will be responsible for some additional questions on problem sets.

**GOV E-2000** is designed for Harvard Extension School students. This section of the course will be taught with the R statistical computing environment as the default with the belief that concepts such as statistical simulation, which are heavily used in GOV E-2001, are important skills that students should take away from the course. However, we will consider accommodating requests to complete the problem sets using Stata statistical software.

**Prerequisites and Recommended Preparation** The prerequisites differ across type of student. For graduate students in the Government Department, there are no prerequisites. For other graduate students, undergraduate students, and Extension School students, the prerequisite is GOV 50, GOV E-1005, or the equivalent.

For any student who meets the prerequisites yet is concerned with his or her preparedness for the course, we strongly encourage the following in advance of the semester. First, we recommend reading and working through the exercises in David Freedman, Robert Pisani, and Roger Purves, *Statistics*, 2007 (any of the older editions should suffice as well). Next, we encourage familiarization with the appropriate statistical package - R or Stata - for the section of the course the student intends on taking. Moreover, if the student plans on typesetting problem set answers in  $\text{\LaTeX}$ , familiarity with the  $\text{\LaTeX}$  markup language would be helpful. Resources on R, Stata, and  $\text{\LaTeX}$  are available under the “Resources” tab on the class website: <http://projects.iq.harvard.edu/gov2000>.

## Class Requirements

Grades will be based on

- weekly homework assignments (50 % of final grade)
- a midterm exam (10 % of final grade)
- a cumulative take-home final exam (30 % of final grade)
- participation, presentation, annotation, and reading comments (10 % of final grade).

*I will not give incompletes in this course.*

## Homework

The weekly homework assignments will consist of analytical problems, computer work, and data analysis. The 2000 section of the course will have additional problems. For all sections, the homework will be assigned on Tuesday night and due the following Tuesday at 1:00pm. Solutions will be posted on Tuesday night, and students will have one week to “self correct” their homework on the basis of the solution key (due the following Tuesday at 1:00pm). These corrections should take the form of comments added to the original homework that indicate where mistakes were made and that demonstrate an understanding of those mistakes.

The homework write-up must be word processed (MS Word is fine), with tables and figures incorporated in the document. No late homework will be accepted except in the case of a documented emergency. All sufficiently attempted homework will be graded on a (+,√,-) scale, and all sufficiently student corrected homework will recover half credit (e.g., homework that receives a √ and is sufficiently corrected with receive a final grade of √/+).<sup>1</sup>

We encourage students to work together on the homework assignments, but you must write your own solutions (this includes computer work), and you must write the names of your collaborators on your assignment. I also strongly suggest that you make a solo effort at all the problems before consulting others. The midterm and the final will be very difficult if you have no experience working on your own.

## Midterm

The midterm will be a short checkout exam (5hrs), that should only take a few hours to complete, and only involves short analytical problems. This exam will be available for checkout one week after we finish the material on univariate descriptive statistics, and it is designed to ensure that all students understand the foundational material. **There is no collaboration allowed on the midterm.**

## Take-home Final

The take-home final exam will be handed out on Wednesday, December 5, one week before the last day of reading period. It will be due at 5:00pm on Wednesday, December 12, the last day of reading period. The take-home final is an exercise in guided replication and primarily involves data analysis and interpretation. Note that the format and goals for the take-home exam are very different from the format and goals for the midterm exam. **There is no collaboration allowed on the final.**

---

<sup>1</sup>All sufficiently attempted homework will be typed and well organized with all problems attempted, and all sufficiently corrected homework will include typed and well organized comments integrated into the original homework. The instructor will determine sufficiency in borderline cases.

### Participation, Presentation, Annotation, and Reading Comments

Ten percent of the grade will be awarded for class participation, quality of presentation on the homework, annotation, and reading comments. A preliminary version of the lecture slides and some of the readings will be posted at NB on Friday evening with references to pages of the textbook on the slides. Students will do the assigned reading and use the annotation tool to append any questions/comments to the slides by Sunday evening. Additionally, students are required to answer a very short set of questions regarding the assigned reading by Sunday evening. The annotations and reading questions provide feedback for tailoring the Tuesday lecture to the needs of the students in the course.

### Discussion Sections

There will be two discussion sections for this course, held back-to-back on Wednesdays from 6-7pm and 7-8pm. The 2000 section will discuss the concepts and R code needed to complete the Gov 2000 homework. The 2000e/1000 section will discuss the concepts and Stata procedures needed to complete the Gov 2000e and 1000 homework. Both sections will be recorded and all students are welcome to sit in both sections of the course.

### Course Website

The course website is located at the following URL:

<http://projects.iq.harvard.edu/gov2000/>. This site will provide homework assignments, datasets, and supplementary materials.

### Course Mailing List

The course mailing list is [gov2000-list@lists.fas.harvard.edu](mailto:gov2000-list@lists.fas.harvard.edu). Please subscribe to the list at <https://lists.fas.harvard.edu/mailman/listinfo/gov2000-list>. If you have trouble subscribing to the list, please email Andy and Konstantin within the first week of the course. This is an ideal forum for posting questions regarding the course material and/or computing. 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 immediately follow lecture on Tuesdays from 4:00 - 6:00pm (with the exception of 9/4, 9/11, and 11/27). I am also available by appointment.

The office hours for Andy Hall and Konstantin Kashin will be determined during the first week of the course. Both will be held in the HMDC computer lab. If you have questions about the course material, computational issues, or other course-related issues please do not hesitate to set up an appointment with either me, Andy, or Konstantin.

If you have a general question, you can also send it 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](mailto: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 would like to stay anonymous.

### Required Books

(ALZ) Ashenfelter, Orley, Levine, Philip, and Zimmerman, David. 2003. *Statistics and Econometrics: Methods and Applications*. John Wiley & Sons.

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

*Note: Most of the required material from ALZ can also be found in the Wooldridge text listed first in the optional books section.*

**Optional Books**

*The following books are optional but may prove useful to students looking for additional coverage of some of the course topics.*

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

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.

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

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

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.

Simon, Carl and Blume, Lawrence. 1994. *Mathematics for Economists*. New York: Norton.

Kennedy, Peter. 2003. *A Guide to Econometrics*. 5th Edition. Malden. Blackwell.

Venables, W.N. and B.D. Ripley. 2002. *Modern Applied Statistics with S-PLUS*. New York: Springer

Gonick, Larry and Smith, Woolcott. 1993. *The Cartoon Guide to Statistics* New York: Harper.

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

Freund, John E. 2004. *Modern Elementary Statistics, 11th Edition*. Upper Saddle River, NJ: Prentice Hall.

Lohr, Sharon L. 2009. *Sampling Design and Analysis, 2nd Edition*. Boston, MA: Brooks/Cole.

Hernán, Miguel and Robins, James. 2011. *Causal Inference (Part I)*.

<http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/>

## Preliminary Schedule

The “Required Reading” should be completed prior to completing the annotation and questions on Sunday evening. The “Suggested Reading” presents alternative presentations of the required material or alternative presentations of the lecture material. The “Optional Reading” provides supplementary material that may be helpful or interesting. *This schedule is subject to change.*

### 1 Descriptive Inference

#### 1.1 Introduction - Sept. 4

##### Topics Covered

- Overview and Course Requirements
- Course Outline
- Introductory Sampling Activity

#### 1.2 Descriptive Questions - Sept. 11

##### Topics Covered

- Describing Univariate Populations
- Describing Bivariate and Multivariate Populations
- Summarization with Bivariate and Multivariate Regression

##### Required Reading

Agresti and Finlay, Ch. 3

Fox 97, Ch. 5

Freund, Sec 2.5

Pages 261-272 of, “Measuring Political Preferences,” by Lee Epstein and Carol Mershon, *American Journal of Political Science*, 1996.

##### Optional Reading

Freedman et al., Chapters 8-12

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.

### 1.3 Randomly Sampled Observations and Basic Probability - Sept. 18

#### Topics Covered

- Elementary Probability Theory
- Random Variables and Functions of Random Variables (Expectation, Variance, ...)
- Joint and Conditional Distributions

#### Required Reading

ALZ, Chs. 2-5

“Racial Prejudice and Attitudes Toward Affirmative Action,” by James H. Kuklinski, Paul M. Sniderman, Kathleen Knight, Thomas Piazza, Philip E. Tetlock, Gordon R. Lawrence, and Barbara Mellers, *The American Journal of Political Science*, 1997.

#### Suggested Reading

Wooldridge Appendix B

Gill, Chapters 7 and 8 (with the exception of 8.3.5, 8.3.8, 8.9)

Fox97 Appendix D.1-D.4

#### Optional Reading

Gonick and Smith, Chapters 2-5

Freedman et al., Chapters 13-15

### 1.4 Random Samples and Descriptive Inference (Univariate) - Sept. 25

#### Topics Covered

- Simple Random Sampling (with and without replacement)
- Distribution of the Sample as an Estimate of the Population Distribution
- Sample Statistics
- Sampling Distributions
- Point Estimation
- Interval Estimation (i.e., confidence intervals)
- Hypothesis Testing

#### Required Reading

ALZ, Chs. 6-8

#### Suggested Reading

Wooldridge Appendix C

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

Fox97 Appendix D.5-D.5.2

#### Optional Reading

Gonick and Smith, Chapters 6-8

Freedman et al., Chapters 16-26

## 1.5 Random Samples and Descriptive Inference (Regression) - Oct. 9

### Topics Covered

- Simple Random Sampling (with and without replacement)
- Stratified Random Sampling (with and without replacement)
- Distribution of the Sample as an Estimate of the Population Distribution
- Sample Statistics and Sampling Distributions
- Point Estimation and Interval Estimation
- Hypothesis Testing

### Required Reading

ALZ, Chs. 9-12

### Suggested Reading

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

Wooldridge, Chapter 2

Fox97, Sections 4-4.4, 5.1 and 6.1

Fox02, Section 3.4, 4.1.1

## 1.6 Diagnosing and Fixing Problems (Part 1) - Oct. 16

### Topics Covered

- Nonlinearity
- Nonconstant Error Variance and Correlated Errors
- Weighted Least Squares and Generalized Least Squares
- “Robust” Standard Errors
- Nonnormality
- Unusual Observations (leverage points, outliers, and influence points)

### Required Reading

ALZ, Ch. 14

Fox11, Ch. 6 (for R students)

<http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter2/statareg2.htm> (for Stata students)

“Law and Data: The Butterfly Ballot Episode,” by Henry E. Brady, Michael C. Herron, Walter R. Mebane Jr., Jasjeet Sekhon, Kenneth W. Shotts, Jonathan Wand, *PS: Political Science and Politics*, 2001.



## 1.7 Diagnosing and Fixing Problems (Part 2) - Oct. 23

### Topics Covered

- Data Missing at Random (conditional on observed data)
- Bounding and Sensitivity Analysis for Data Not Missing at Random

### Required Reading

Lohr, Ch. 8

GH 7.1-7.2, 25

### Suggested Reading

Little, R.A. and Rubin, D.B. 1989. "The Analysis of Social Science Data with Missing Values." in *Sociological Methods & Research*.

## 2 Causal Inference

### 2.1 Introduction - Oct. 30

#### Topics Covered

- Potential Outcomes and Causal Effects
- Causal Inference as a Missing Data Problem
- Introductory Causal Inference Activity

#### Required Reading

Gerber and Green, Ch. 2

Hernán and Robins, Secs. 1.1-1.3, 1.5,

### 2.2 Causal Questions - Oct. 30

#### Topics Covered

- Describing Univariate Distributions of Potential Outcomes
- Describing Univariate Distributions of Causal Effects
- Conditional Distributions of Potential Outcomes and Causal Effects
- Causal Questions Not Addressed in the Course

#### Required Reading

Hernán and Robins, Secs. 2.1-2.3, 3.1, 4.1, 5.1

**Suggested Reading**

Sovey, Allison J. and Green, Donald P. 2011. "Instrumental Variables Estimation in Political Science: A Readers' Guide." *American Journal of Political Science*

Yamamoto, Teppei. 2012. "Understanding the Past: Statistical Analysis of Causal Attribution." *American Journal of Political Science*.

Menzies, Peter. 2008. Section 4.3 of "Counterfactual Theories of Causation" in the *Stanford Encyclopedia of Philosophy*.  
<http://stanford.library.usyd.edu.au/entries/causation-counterfactual/>

**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.

**2.3 Randomized Treatment Assignment - Nov. 6****Topics Covered**

- Identification with Randomized and Conditionally Randomized Treatment Assignment
- Estimation, CIs, and Testing with Randomized and Conditionally Randomized Treatment Assignment

**Required Reading**

Re-read Hernán and Robins, Secs. 2.1-2.3

Re-read ALZ, Secs 9.3, 11.1-11.3, Ch. 12

"Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment," by Alan S. Gerber, Donald P. Green, and Christopher W. Larimer. *The American Political Science Review*, 2008.

**Suggested Reading**

Wooldridge 3.3

Fox97 6.2-6.5

**2.4 Observational Studies with Measured Confounding - Nov. 13****Topics Covered**

- The Assumption of No Unmeasured Confounding
- Relation to Classical Econometric Assumptions
- Choosing Conditioning Variables
- Regression Based Estimation (Additive and Interactive)

**Required Reading**

Re-read Hernán and Robins, Sec. 3.1

GH, Ch. 9

ALZ, Secs. 13.1 - 13.3

“Female Socialization: How Daughters Affect Their Legislator Fathers’ Voting on Women’s Issues,” by Ebonya L. Washington, *The American Economic Review*, 2008.

**2.5 Diagnosing and Fixing Problems - Nov. 20****Topics Covered**

- Assessing Overlap and Balance
- Revising the Question of Interest
- Bounding and Sensitivity Analysis for Unmeasured Confounding

**Required Reading**

GH, Secs. 10.1-10.4

“MPs for Sale? Returns to Office in Postwar British Politics.” Eggers, Andrew and Hainmueller, Jens, *The American Political Science Review*, 2009.

“Electoral Institutions, Cleavage Structures, and the Number of Parties,” by Amorim Neto and Gary W. Cox, *American Journal of Political Science*, 1997.

**2.6 Overflow/Review Session - Nov. 27****Topics Covered**

- Remaining Topics
- Review of Course Material