# Quantitative Methods for Political Science I
# Government 1000

Professor Jasjeet Singh Sekhon
Jacob Kline, TF

Class: 4-6pm Mondays
CBRSS Conference Room (34 Kirkland St.)

**Assistant Professor Jasjeet Singh Sekhon**
jasjeet_sekhon@harvard.edu
HTTP://jsekhon.fas.harvard.edu/
Phone: 496-2426, Fax: 496-5149
Office: CBRSS, 34 Kirkland St., Room #4
Office hours: 2−3pm Wednesdays, or by appointment, or by email anytime


**Jacob Kline, Teaching Fellow**
jkline@fas.harvard.edu
Phone: TBA
Office: TBA
Office hours: TBA
Sections: TBA

## Catalog Description

Half Course. Introduction to major quantitative techniques used in political science. Covers exploratory data analysis, as well as descriptive and causal statistical inference of many types. The course emphasizes probability theory, least squares, robust estimation, matching and propensity scores. Note: Frequently taken by graduate students satisfying department requirements. Undergraduates looking for an introductory class should consider government 1001.

## Discussion

We use an email list in this class quite frequently: `mailto:gov1000-list@fas.harvard.edu`. If students have a question about the course material, they are advised to email the question to the entire list and ask for help. Letting everyone see each other's questions and the answers improves everyone's work.The primary focus of this course is to teach students (primarily undergraduates) enough statistical methods so they may conduct original quantitative empirical work on their own.

## Prerequisites

High school algebra.

If you need a refresher you should try to get an introductory algebra book, such as from the Schaum's Outline Series, and use it as a reference while doing homework assignments.

## Evaluation

The only way to learn statistics is by doing. There is no substitute for answering questions. Therefore, there will be regular class assignments and two exams. Grades will be based on a weighted average of the midterm, final exam, regular class assignments and participation in both class and section. More information will be provided in class.

It is highly recommended that students form study groups in order to complete the homework assignments. Although it is recommended that people work together in order to complete the homework assignments, each individual student must hand in their own answers. Photocopies (or equivalent reproductions) of someone else's answers are not acceptable.

## Course Books and Software

Students may use any computer package to complete the assignments they wish. However, we recommend that students use one of the two main variants of the $S$ statistical programming language: $R$ or *Splus*. Both are installed on Harvard-MIT Data Center computers. $R$ is available for a large variety of platforms for download: `http://www.r-project.org/`. $R$ is open source software (released under the GNU public license) and is available at no charge. *Splus* is a commercial product and is available at no cost to Harvard students.

The five books listed below are required and available at various online bookstores. The first book is the main text book for the course. The second book is the secondary text and offers more intuition but less detail than the first. The third is a computer manual which does a nice job of linking theory with practice. The last two books supplement our discussion of least squares and logistic regression. The core course material will be communicated in lectures and associated notes and handouts. The textbooks are important reference guides which help communicate the concepts.

- Wonnacott, Thomas H. and Roland J. Wonnacott. 1990. *Introductory Statistics*. 5th ed. Wiley, John & Sons. ISBN: 0471615188.

- Freedman, David, Robert Pisani and Roger Purves. 1997. *Statistics*. 3rd ed. W. W. Norton & Company. ISBN: 0393970833.

- Dalgaard, Peter. 2002. *Introductory Statistics with R*. 1st ed. Springer Verlag. ISBN: 0387954759.

- Achen, Chris. 1982. *Interpreting and Using Regression*. Sage Publications. ISBN: 0803919158.

- Pampel, Fred C. 2000. *Logistic Regression: A Primer*. Sage Publications. ISBN: 0761920102.

# Course Plan and Tentative Outline

We will not be able to cover all of this material. But we will cover as much as we can. There is no serious time constraint because any material left over will be covered in Government 2000. We do not read many articles, but the articles we do read will be read carefully over several weeks.

1. The Nature of Statistics

    (a) Observational studies versus randomized experiments

    (b) Causal inference versus description

    (c) Methods of inductive inference

    Readings:

    - Ch. 1 of Wonnacott and Wonnacott
    - Ch. 1, 2 of Freedman et al.

    These readings will be examined closely over several weeks. They will get clearer as we learn more math.

    - Jasjeet S. Sekhon. "Quality Meets Quantity: Case Studies, Conditional Probability and Counterfactuals." *Perspectives on Politics*. Forthcoming.
    - Reiter, Jerome. 2000. "Using Statistics to Determine Causal Relationships." *American Mathematical Monthly*. Pages 24–32.
    - Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis* 2:131–150.

2. Descriptive Statistics (in Section)

    (a) Frequency tables and graphs

    (b) Center of a distribution

    (c) Spread of a distribution

    (d) Linear Transformations

    (e) Calculations using relative frequencies

    Readings:

    - Ch. 2 of Wonnacott and Wonnacott
    - Ch. 3, 4 of Freedman et al.
    - Ch. 1 of Dalgaard

3. Probability

    (a) Probability models

    (b) Compound events

    (c) Conditional probability

    (d) Independence

    (e) Conditional Independence

(f) Bayes Theorem

(g) Other views of probability

Readings:

- Ch. 3 of Wonnacott and Wonnacott
- Ch. 13, 14, 15 of Freedman et al.

4. Probability Distributions

(a) Discrete random variables

(b) Mean and variance

(c) The binomial distribution

(d) Continuous distributions

(e) The normal distribution

(f) Functions of random variables

Readings:

- Ch. 4 of Wonnacott and Wonnacott
- Ch. 5, 6 Freedman et al.
- Ch. 2, 3 of Dalgaard

5. Two Random Variables

(a) Distributions

(b) A function of two random variables

(c) Covariance

(d) Correlation (see section 15-1 of Wonnacott and Wonnacott)

(e) Linear combination of two random variables

Readings:

- Ch. 5 of Wonnacott and Wonnacott
- Ch. 8, 9 Freedman et al.
- Ch. 4 of Dalgaard

6. Rules of Variances and Expectations (in Section)
   See lecture notes.

7. Experimental Manipulation.
   Readings: These readings will be examined closely over several weeks. They will get clearer as we learn more math.

   - Holland, Paul. 1986. "Statistics and Causal Inference (with comments by Rubin, Clark Glymour, Clive Granger)." *Journal of the American Statistical Association* 81(396): 945–970

- Gerber, Alan S. and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94(3): 653–663.
- Imai, Kosuke. "Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review*. Forthcoming.

8. Sampling

   (a) Random Sampling
   (b) Moments of the sample mean
   (c) The shape of the sampling distribution
   (d) Proportions (percentages)

   Readings:

   - Ch. 6 of Wonnacott and Wonnacott
   - Ch. 19–21 Freedman et al.
   - Ch. 7 of Dalgaard
   - Reading: Michael P. McDonald and Samuel Popkin. 2001. "The Myth of the Vanishing Voter." *American Political Science Review* 95(4): 963-974.

9. Point Estimation

   (a) Unbiasedness and Consistency
   (b) Ordinary least squares (OLS)
   (c) Robustness, least median of squares

   Readings:

   - Ch. 7 of Wonnacott and Wonnacott
   - Ch. 7, 10 Freedman et al.
   - Ch. 5 of Dalgaard

10. Confidence intervals (Chapter 8)

    (a) A single mean
    (b) Difference in two means, independent samples
    (c) Proportions
    (d) Confidence intervals and tests for $\beta$
    (e) Small-sample $t$

11. Hypothesis Testing (Chapter 9)

    (a) Hypothesis testing using confidence intervals
    (b) $p-$values
    (c) Classical hypothesis tests
    (d) Bayesian confidence sets

12. Point Estimation and Simple Regression (Chapters 11–12)

    (a) The regression model
    (b) Sampling variability
    (c) Predicting $Y$ at a given level of $X$

    Reading: Achen, pages 1-37

13. Multiple Regression and Regression and Correlation Compared (Chapters 13 and 15)

    (a) The regression model and its OLS fit
    (b) Regression coefficients as multiplication factors
    (c) Simple and multiple regression compared
    (d) Correlation and regression
    (e) Two regression lines
    (f) Correlation in multiple regression
    (g) Multicollinearity

    Readings:

    - Achen, pages 37ff
    - MacKuen, Michael B., Robert S. Erikson and James A. Stimson. 1992. "Peasants or Bankers? The American Electorate and the U.S. Economy" *American Political Science Review* 86(3): 597–611.

14. Binary Regression
    Readings:

    - Pampel, *Logistic Regression: A Primer.*
    - Bartels, Larry M. 1996. "Uninformed Votes: Information Effects in Presidential Elections" *American Journal of Political Science* 40(3): 905–942.

15. Propensity Scores
    Readings:

    - Rosenbaum, Paul R. and Rubin, Donald B. (1985) "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score" *American Statistician* 39: 33–38.
    - We revisit Gerber and Green (2000) and Imai.

16. Regression Extensions (Chapter 14)

    (a) Non-spherical errors

17. Instrumental Variables (lecture notes).
    Reading: Steven D. Levitt and James M. Snyder, Jr. 1997. "The Impact of Federal Spending on House Election Outcomes." *Journal of Political Economy* 105(1): 30–53.

18. Simulating Quantities of Interest: A Cautionary Tale
    Reading: King, Gary and Michael Tomz and Jason Wittenberg. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation."

19. Robust Estimation (lecture notes).
    Readings:

    - Western, Bruce. 1995. "Concepts and Suggestions for Robust Regression Analysis." *American Journal of Political Science* 39: 786–817.
    - Jonathan N. Wand, Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael C. Herron and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95 (4): 793–810.
    - Walter R. Mebane, Jr. and Jasjeet S. Sekhon. forthcoming. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science.*

# Supplementary Reading Material

The following books are certainly not required, but they may be of interest during the course. It is often very useful to read the same material covered by a variety of authors. Within each section, books are approximately ordered by increasing sophistication.

## Computer Books

- Krause, Andreas and Melvin Olson. 2002. *The Basics of S-PLUS.* 3rd ed. New York: Springer-Verlag. ISBN: 0387954562.

- Spector, Phil. 1995. *An Introduction to S and S-Plus.* Wadsworth Publishing Company. ISBN: 053419866X.

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

- Venables, W.N and Brian D. Ripley. 2002. *Modern Applied Statistics with S.* New York: Springer-Verlag. ISBN: 0387954570

## Mathematical Review

- Spiegel, Murray R. and Robert E. Moyer. 1997. *Schaum's Outline of College Algebra.* 2nd edition. McGraw-Hill .

- Kleppner, Daniel and Norman Ramsey. 1972. *Quick Calculus.* Wiley.

## Probability and Statistics

- Freedman, David, Robert Pisani, and Roger Purves. 1997. *Statistics.* 3rd edition. Norton, W.W. & Company.

- Lee J. Bain and Max Engelhardt. 1987. *Introduction to Probability and Mathematical Statistics.* Duxbury.

- Mendenhall, William and Robert J. Beaver. 1994. *Mathematical Statistics with Applications.* Duxbury.

- DeGroot, Morris H. 1986. *Probability and Statistics.* Addison-Wesley.

## Linear Regression and Applications

- Achen, Christopher H. 1982. *Interpreting and Using Regression.* Beverly Hills Sage Publications.

- Gujarati, Damodar N. 1995. *Basic Econometrics.* 3rd edition. New York McGraw-Hill.

- Gujarati, Damodar N. 1998. *Essentials of Econometrics.* 2nd edition. New York McGraw-Hill.

- Achen, Christopher H. 1986. *Statistical Analysis of Quasi-Experiments.* Berkeley University of California Press.

- Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists.* New York Academic Press.

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

- Pindyck, Robert S. and Daniel L. Rubinfeld. 1997. *Econometric Models and Economic Forecasts.* 4th ed. Boston, MA: Irwin/McGraw-Hill.

- Greene, William H. 1990. *Econometric Analysis.* New York MacMillan Publishing Company.