Conduct of Inquiry II
Government 613

Ryan T. Moore*

20 January 2016 at 14:18

Course Information

Government GOVT 613
Conduct of Inquiry II
Thursday, 2.35pm–5.15pm, Hurst Hall 209

Instructor Information

Ryan T. Moore, Ph.D.
Assistant Professor of Government
Office: Ward 226
Telephone: 202.885.6470
Fax: 202.885.2967
Homepage: http://www.ryantmoore.org
Email: rtm (at) american (dot) edu
Office Hours: Wednesday, 10am-noon or by appointment
(Please use https://calendly.com/ryantmoore to schedule times.)

Ghazal Nadi
Office Hours Location: Hurst Building 206
Email: gp4917a (at) american.edu
Office Hours: Tuesday, 10.30am-12.30pm, or by appointment

Course Description

This course covers the use of bivariate and multivariate regression analysis in the social sciences. We examine the estimation, properties, and interpretation of linear models. We also introduce nonlinear regression, conditioning via non-regression estimators, and regression approaches that are robust to violations of the assumptions of the typical linear model. Throughout, we give attention to the distinction between causal and descriptive analyses.

---

*Department of Government, American University, Ward Circle Building 226, 4400 Massachusetts Avenue NW, Washington DC 20016-8130. tel: 202.885.6470; fax: 202.885.2967; rtm (at) american (dot) edu; http://ryantmoore.com.

Learning Objectives

By the end of the course, you should be able to

- Translate political phenomena into mathematical notation

- Differentiate causal from descriptive statistical analyses

- Understand the value and limitations of regression analyses

- Test substantive hypotheses using regression

- Interpret a variety of types of regression coefficients (on indicator variables, interaction terms, higher-order terms, etc.)

- Describe assumptions of linear models, determine the degree to which they validate a technique (such as ordinary least squares), and address violations of them where possible

- Given a data generating process, select an appropriate statistical model and method

- Write an original research paper that uses a technique from the course to answer a relevant social science question

- Use R to import and manipulate data, implement regression and other conditioning techniques, calculate diagnostics, and produce publication-quality graphics

- Typeset social scientific methods and results legibly, likely using LaTeX

Readings

Readings should be completed before the course meeting under which they are listed below. The course readings are a mixture of academic articles and book chapters. Many readings are available on the Blackboard page under the "Course Reserves" link. Note that in some cases, the required readings represent only a subset of the pages posted on the reserves page.

As Conduct II begins, we will use the last chapter of the textbook from Conduct I,

> Imai, Kosuke. *A First Course in Quantitative Social Science.* Princetion University Press, Princeton, NJ, 2016.

Thereafter, the primary textbook for the course is

> John Fox. *Applied Regression Analysis and Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 3rd edition, 2016.

The companion guide for R is

> John Fox and Sanford Weisberg. *An R Companion to Applied Regression.* Sage Publications, Inc., Thousand Oaks, CA, 2011.

If you took Conduct I during fall 2015, the *Companion* guide is optional, given your exposure to R. The most useful chapters of the *Companion* guide for you at this stage are likely to be Chapters 4 and 6. If you are encountering R for the first time in this course, I recommend the *Companion* guide.

For the statistical topics of Conduct II, I can also recommend

Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach.* South-Western, Mason, OH, 5th edition, 2012.

which has recently been the text for Conduct III. For a social scientific introduction to causal inference, see

Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference.* Cambridge University Press, New York, 2007.

which has recently been the text for Conduct IV. For two other texts that provide introductions to statistical material, see

Sean Gailmard. *Statistical Modeling and Inference for Social Science.* Cambridge U Press, New York, NY, 2014.

David Freedman, Robert Pisani, and Roger Purves. *Statistics.* WW Norton & Company, Ltd, New York, NY, 4th edition, 2007.

Requirements and Evaluation

Students are required to do the weekly reading, attend class, complete all assignments and exams, contribute significantly to in-class group work, and participate in course discussions about the material. Using the course email list to ask and answer questions is strongly encouraged, and will contribute to your participation evaluation.

The research paper should represent original data analysis, and should address a question of substantive interest in your field. Your proposal for the paper (due before Spring Break) should succintly answer several questions: What is the literature your paper will participate in? What data will your paper utilize? How do you expect to analyze the data? What conclusions will you draw, given certain results?

A summary of the course assessments is in Table 1.

Assignment	Weight	Due date
Problem Sets	30%	Weekly
Paper Proposal		1 March
Midterm Exam	15%	3 March
Paper Presentation	5%	19-21 April
Final Exam	15%	28 April
Final Paper	25%	3 May
Participation	10%	Weekly

Table 1: Course Assessment Summary

No late work will be accepted. If you cannot submit an assignment on time, arrange to submit it early. We encourage you to use office hours to discuss any specific assignments, difficulties, or questions about the course.

Academic integrity is a core value of institutions of higher learning. It is your responsibility to avoid and report plagiarism, cheating, and dishonesty. Please (re-)read the University policy on academic integrity at http://www.american.edu/academics/integrity/code.cfm, particularly Sections I and II.

Software and Statistics Support

The primary software for the course is R. See http://www.ryantmoore.org/files/ht/htr.pdf for help getting started. Support for statistical software is available through CTRL. See http://j.mp/ZrBr2Z for CTRL's workshop schedule. The workshop introducing dissertation templates on February 4 might be of particular interest.

The Department of Mathematics and Statistics offers statistical consulting services, with extensive hours. For the schedule and contact information, see http://j.mp/1EmVqkY.

The library itself offers support for various software. For example, they can help you troubleshoot issues with your LATEX installation.

Graduate Certificate in Applied Statistics

If you are interested in political methodology or applied statistics, consider completing the Graduate Certificate in Applied Statistics. In addition to learning exciting new techniques, certification can help you distinguish your skills and commitment to methods from those of others applying for the same positions you are. More information is available at http://www.american.edu/cas/mathstat/CERT-GAS.cfm.

Intellectual Property

Course content is the intellectual property of the instructor or student who created it, and may not be recorded or distributed without consent.

Course Evaluation

The course evaluation will take place in class towards the end of the semester. Students who submit the evaluation will earn one percentage point toward the participation grade.

Calendar

I. Introduction. Preliminary topics in inference. Causality.

14 January

Introduction. Course policies and requirements. Computing. Probability and distributions.

Required reading.

This syllabus.
(Review Imai, §7.1–7.1.3)

21 January

Randomization inference. The p-value. Statistical inference for means and proportions.
PS 1 due.

Required reading.

> Kosuke Imai. *A First Course in Quantitative Social Science*. Princetion University Press, Princeton, NJ, 2016, **§7.1.4–7.2**

Supplementary reading.

> Daniel E. Ho and Kosuke Imai. "Randomization Inference with Natural Experiments: An Analysis of Ballot Effects in the 2003 California Recall Election". *Journal of the American Statistical Association*, 101(475):888–900, September 2006.

28 January

Causal inference. Causal identification. Matching methods.
PS 2 due. For PS 2, complete Imai §7.5.2 on the "filedrawer" problem, questions 1, 2, 4.1, 5, 6, and 7 (question 3 and the second half of 4 are optional, for bonus credit).

Required reading.

> David A. Freedman. Statistical models and shoe leather. *Sociological Methodology*, 21:291–313, 1991.

> David Freedman. From Association to Causation: Some Remarks on the History of Statistics. *Statistical Science*, 14(3):243–258, 1999.

Supplementary reading.

> Stefano M. Iacus, Gary King, and Giuseppe Porro. Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1):1–24, Winter 2012.

> Donald B. Rubin. The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies. *Biometrics*, 29(1):185–203, Mar 1973.

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

> Alexis Diamond and Jasjeet S. Sekhon. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 95(3):932–945, July 2013.

II. Linear Modeling.

4 February

Conditional expectation functions. Bivariate regression.
PS 3 due.

Required reading.

> John Fox. *Applied Regression Analysis and Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 3rd edition, 2016, **Chapter 2 and §5.1**

11 February

Linear algebra.
PS 4 due.

Required reading.

> Jeff Gill. *Essential Mathematics for Political and Social Research.* Cambridge U Press, New York, 2006, **Chapter 3**.

> Jeff Gill. *Essential Mathematics for Political and Social Research.* Cambridge U Press, New York, 2006, **Chapter 4**.

18 February

Least squares regression.
PS 5 due.

Required reading.

> John Fox. *Applied Regression Analysis and Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 3rd edition, 2016, **§5.2, §6.1–6.3**

> Arthur S. Goldberger. *"A Course in Econometrics".* Harvard University Press, Cambridge, MA, 2000, pp. 176–179.

25 February

Properties of least squares regression.
PS 6 due.

Required reading.

> John Fox. *Applied Regression Analysis and Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 3rd edition, 2016, **§9.3**

(1 March)

Paper proposal due.

3 March

Midterm Exam.

10 March

Spring Break. No class meeting.

17 March

Properties of least squares, continued. Inference for multiple regression.

Required reading.

John Fox. *Applied Regression Analysis and Generalized Linear Models*. Sage Publications, Thousand Oaks, CA, 3rd edition, 2016, **§9.4**

Gelman, Andrew and Hal Stern. "The Difference Between 'Significant' and 'Not Significant' is not Itself Statistically Significant". *The American Statistician*, 60(4):328–331, 2006.

Supplementary reading.

Gill, Jeff. "The Insignificance of Null Hypothesis Significance Testing". *Political Research Quarterly*, 52(3):647–674, 1999.

24 March

Interpreting regression. Indicator variables. Interaction terms. Higher-order terms.
PS 7 due.

Required reading.

Bear F. Braumoeller. Hypothesis testing and multiplicative interaction terms. *International Organization*, 58:807–820, Fall 2004.

Peter M. Aronow and Cyrus Samii. Does regression produce representative estimates of causal effects? *American Journal of Political Science*, 60(1):250–267, January 2016.

III. Problems (and Solutions) in Ordinary Least Squares.

31 March

Assumptions and violations. Linearity. Error variance. Error distributions. Multicollinearity.
PS 8 due.

Required reading.

> John Fox. *Applied Regression Analysis and Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 3rd edition, 2016, **pp. 296–323**.

> Gary King and Margaret E. Roberts. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It". *Political Analysis*, August 2014.

> Gary King. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science". *American Journal of Political Science*, 30:666–687, 1986.

> Arthur S. Goldberger. *"A Course in Econometrics"*. Harvard University Press, Cambridge, MA, 2000, pp. 245–252.

Supplementary reading.

> John Fox. *Applied Regression Analysis and Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 3rd edition, 2016, Chapter 13.

7 April

Outlier, leverage, and influence diagnostics. Robust and resistant regression.
PS 9 due.

Required reading.

> John Fox. *Applied Regression Analysis and Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 3rd edition, 2016, **Chapter 11 and Chapter 19**.

14 April

Regression with non-continuous outcomes. The maximum likelihood approach.
PS 10 due.

Required reading.

> John Fox. *Applied Regression Analysis and Generalized Linear Models.* Sage Publications, Thousand Oaks, CA, 3rd edition, 2016, §**14.1**.

21 April

Final topics.
Paper presentations.

IV. Assessments

28 April

Final exam, 2.35pm-5.05pm, Location TBA.

3 May

Final paper due, noon.