
PSC 404

The Art and Practice of Data Analysis

Fall 2022
15:25-16:40 T/Th
Fenno Room

Kevin A. Clarke
Harkness 317
Office Hours: just stop by.
kevin.clarke@rochester.edu

TA: Joe Essig
Harkness 305
Recitation: TBA
wqi3@ur.rochester.edu

PURPOSE

This course in applied data analysis provides graduate students with the tools to estimate useful and empirically adequate linear models that address substantive concerns in political science. Estimating models is trivial; estimating models that provide useful information about political outcomes is far more challenging. Our focus will be on how estimates are produced and what those estimates mean with the goal of leaning which estimates we can trust.

PREREQUISITES

The math “boot camp” is the only course prerequisite, as familiarity with calculus is necessary to understand the material.

COURSE REQUIREMENTS

Evaluation is based on homework assignments (20%), a midterm exam (25%), a final exam (25%), and a data analysis project (30%). In addition to office hours, the teaching assistant will hold a weekly R recitation on Thursdays from 1-2 PM. Attendance is mandatory. Students are responsible for material covered in lecture, recitation, and the required readings.

READINGS

All reading are online either through the library[†], JSTOR[‡], or on Blackboard. Students looking for additional information could consult the following resources.

Mathematical statistics

T. Amemiya (1994). *Introduction to Statistics and Econometrics*. Harvard University Press

M.H. DeGroot and M.J. Schervish (2012). *Probability and Statistics*. Addison-Wesley

G. Casella and R.L. Berger (2002). *Statistical Inference*. Thomson Learning
D.J. Poirier (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press

Econometrics

J.M. Wooldridge (2008). *Introductory Econometrics: A Modern Approach*. Cengage Learning

Joshua Angrist and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton, NJ: Princeton

A.C. Cameron and P.K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press

J.M. Wooldridge (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press

COURSE SCHEDULE

Week 1: Introduction and the origins of regression

Topics: Regression to the mean, least squares

Week 2: Useful matrix algebra

Topics: Special matrices, useful identities, matrix calculus

- Fox 2009.[†] Chapter 1.
- Namboodiri 1984.[†] Entire.
- Rencher and Schaalje 2008. Chapters 2, 3, 5

Week 3: Fitting a line to data and being responsive to theory

Topics: Assumptions, dummy variables, nonlinearities, interactions

- Berk 2004.[†] Chapter 3.
- Berk 2008.[†] Pages 1-8.
- Fox 2016. Chapter 5.
- Rencher and Schaalje 2008. Chapter 7.

Week 4: Understanding regression uncertainty

Topics: Sampling distributions, standard errors, inference

- Achen 1982.[†] Entire.
- Berk 2004.[†] Chapters 4, 6-7.

- Berk 2010.[‡]
- Fox 2016. Chapter 21.
- Gelman, Hill, and Vehtari 2020, Chapters 4-5

Week 5: Hypothesis testing and the linear model

Topics: Wald tests, confidence intervals, ANOVA

- Braumoeller 2004.[‡]
- Imbens and Ruben 2015. Chapters 5 and 6.
- Kaplan 2009. Chapter 15 and 16.
- Rencher and Schaalje 2008. Chapter 8.

Week 6: Endogeneity and other problems

Topics: Omitted variables, selection, measurement error

- Achen 2005.[‡]
- Clarke 2005.[‡]
- Clarke 2009.[‡]

Week 7: Regression in practice: a case study .

- Achen and Bartels 2012.[†]
- Fowler and Hall 2018.[†]
- Achen and Bartels 2018.[†]
- Berk 2004.[†] Chapter 11.

Week 8: Review and midterm

Week 9: Maximum likelihood and the linear model

Topics: Likelihood inference, role of large samples, distributions

- Cox 2006. Chapter 2.
- Fox 2009.[†] Chapter 3.

Week 10: Generalized linear models

Topics: Probit, logit, Poisson

- Fox 2016. Chapters 14-15.
- Gelman, Hill, and Vehtari 2020. Chapters 13-15.

Week 11: Instrumental variables

Topics: Two-stage least squares

- J. D. Angrist and Krueger 2001[‡]
- Sovey and Green 2011.[‡]

Week 12: Causality and the potential outcomes model

- Gelman, Hill, and Vehtari 2020. Chapters 18-19.
- Imbens and Ruben 2015. Chapter 2.

Week 13: Difference-in-differences and regression discontinuity

- TBA.

Week 14: Review and final

References

- Achen, Christopher H. (1982). *Interpreting and Using Regression*. Newbury Park, CA: Sage.
- (Winter 2005). “Let’s Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong”. *Conflict Management and Peace Science* 20.4.
- Achen, Christopher H. and Larry M. Bartels (2012). “Blind Retrospection: Why Shark Attacks are Bad for Democracy”.
- (Oct. 2018). “Statistics as If Politics Mattered: A Reply to Fowler and Hall”. *The Journal of Politics* 80.4.
- Amemiya, T. (1994). *Introduction to Statistics and Econometrics*. Harvard University Press.
- Angrist, Joshua and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: an empiricist’s companion*. Princeton, NJ: Princeton.
- Angrist, Joshua D. and Alan B. Krueger (2001). “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments”. *The Journal of Economic Perspectives* 15.4, pp. 69–85.
- Berk, Richard A. (2004). *Regression Analysis: A Constructive Critique*. Sage.
- (2008). *Statistical learning from a regression perspective*. Springer.
- (2010). “What You Can and Can’t Properly Do with Regression”. *Journal of Quantitative Criminology* 26.4, pp. 481–487.
- Braumoeller, Bear (2004). “Hypothesis testing and multiplicative interaction terms”. *International Organization* 58.4.
- Cameron, A.C. and P.K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Casella, G. and R.L. Berger (2002). *Statistical Inference*. Thomson Learning.
- Clarke, Kevin A. (Winter 2005). “The Phantom Menace: Omitted Variable Bias in Econometric Research”. *Conflict Management and Peace Science* 20.4.
- (Feb. 2009). “Return of the Phantom Menace: Omitted Variable Bias in Econometric Research”. *Conflict Management and Peace Science* 26.1.

- Cox, D.R. (2006). *Principles of Statistical Inference*.
- DeGroot, M.H. and M.J. Schervish (2012). *Probability and Statistics*. Addison-Wesley.
- Fowler, Anthony and Andrew B. Hall (Oct. 2018). “Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent Finding on Voter Competence”. *The Journal of Politics* 80.4.
- Fox, John (2009). *A mathematical primer for social statistics*. Sage.
- (2016). *Applied regression analysis and generalized linear models*. Sage.
- Gelman, A., J. Hill, and A. Vehtari (2020). *Regression and Other Stories*. Cambridge University Press.
- Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference*. Cambridge.
- Kaplan, Daniel T. (2009). *Statistical Modeling: A Fresh Approach*.
- Namboodiri, Krishnan (1984). *Matrix Algebra: An Introduction*. Sage.
- Poirier, D.J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press.
- Rencher, Alvin C. and G. Bruce Schaalje (2008). *Linear Models in Statistics*. Wiley.
- Sovey, Allison J. and Donald P. Green (2011). “Instrumental Variables Estimation in Political Science: A Readers’ Guide”. *American Journal of Political Science* 55.1, pp. 188–200.
- Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- (2008). *Introductory Econometrics: A Modern Approach*. Cengage Learning.