SS 228a: Applied Data Analysis for the Social Sciences

Fall 2018-19

Prof. Jonathan N. Katz
Office: Baxter 325
Phone: x4191
Email: jkatz@caltech.edu

The purpose of this seminar is to familiarize the students with the applied topics in data analysis and causal inference. Given that this is an advanced graduate seminar, it will be run somewhat informally with a combination of some lecturing as well as group discussion and presentations.

I will periodically give exerises that you are expected to complete, but the main work for the class is to work towards an empirical analysis. This need not be a completed paper, but I am more interested in seeing the analysis and code for the project.

The main texts for the class are:

Hernán, Miguel A. and James M. Robins. 2016. *Causal Inference*. Forthcoming. New York: Cambridge University Press. Available at:
`https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/`

Imai, Kosuke. 2017. *Quantitative Social Sciences: An Introduction*. Princeton: Princeton University Press.

Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd Edition. New York: Cambridge Univeristy Press.

Rosenbaum, Paul R. 2009. *Design of Observational Studies*. New York: Springer.

Trochim, William M.K. 2016. Research Methods Knowledge Base. Available at:
`http://www.socialresearchmethods.net/kb/`

Wickham, Hadley and Garett Grolemund. 2016. *R for Data Science*. Avaibalbe at:
`http://r4ds.had.co.nz`

Course Schedule

October 3: Introduction

October 10: Measurement, Construct Validity, and Design

Required:

Adcock, Robert and David Collier. 2001. "Measurement validity: A shared standard for qualitative and quantitative research." *American Political Science Review*, 95(03): 529–546.

Imai (2017), Chapter 3

Kaufmann, Daniel and Aart Kraay. 2007. Governance Indicators: Where Are We, Where Should We Be Going? The World Bank Policy Research Working Paper 4370

Rosenbaum (2009), Chapter 1.

Trochim (2016), Chapters 3 (Measurement) and 4 (Design).

Recommended:

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Deigning Social Inquiry: Scientific Inference in Qualitative Research* Princeton, NJ: Princeton University Press.

October 17: Introduction to R and Best Programing Practices

Required:

Wickham and Grolemund (2016). *R for Data Science*, Entire Book.

Imai (2017), Chapter 1.

Recommended:

Google's R Style Guide, available at:
`https://google.github.io/styleguide/Rguide.xml`

R Coding Conventions (RCC), available at: `http://www.aroma-project.org/developers/RCC`

October 24: Counterfactuals, Potential Outcomes, Causal Graphs

Required:

Imai (2017), Chapters 2 and 4.3.

Morgan and Winship (2015), Chapters 1 to 4.

Hernán and Robins (2016), Chapters 1 to 3 and 6 to 10.

Recommended:

Imbens, Guido W. and Donald R. Rubin. 2015. *Casual Inference for Statistics, Social, and Biomedical Science.* New York: Cambridge University Press. Chapters 1 to 3.

Keele, Luke 2015. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis*, 23(3):313–335.

October 31: Regression

Required:

Imai (2017), Chapter 4.2.

Morgan and Winship (2015), Chapters 6 and 7.

Aronow, Peter M. and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science*, 60(1):250–267.

Beck, Nahaniel and Jonathan N. Katz. 2011. "Modeling Dynamics in Time-Series–Cross-Section Political Economy Data". *Annual Review of Political Science.* 14:331–352.

Recommended:

Fox, John. 2016. *Applied Regression Analysis and Generalized Linear Models.* Thousands Oaks, CA: SAGE Publications, Chapters 11 to 13.

Imbens and Rubin (2015), Chapters 13 to 19.

November 7: Matching Estimators

Required:

Morgan and Winship (2015), Chapters 5.

Rosenbaum (2009), Chapters 2,3, 7, 8, 9, 13.

November 14: Panel & Time-Series-Cross-Section Data

Required:

Morgan and Winship (2015), Chapters 11

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2016. "How Much Should We Trust Differences in Differences Estimates" *Quarterly Journal of Economics* 119(1):249–275.

Beck, Nathaniel and Jonathan N. Katz 1995. "What To Do (and Not To Do) with Times-SeriesCross-Section Data in Comparative Politics". *American Political Science Review* 89(3): 634-647. 8)

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of Californias Tobacco Control Program" *Journal of the American Statistical Association*, 105(490): 493–505.

November 21: Dynamics Required:

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker 1998. " Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable". *American Journal of Political Science* 42(4):331–352.

Beck, Nathaniel and Jonathan N. Katz 2011. "Modeling Dynamics in Time-SeriesCross-Section Political Economy Data". *Annual Review of Political Science* 14:331–352

November 28: Analyzing Text Required:

Silge, Julua and David Robinson 2017. *Text Mining with R: A Tidy Approach* Sebastopol, CA: O'Reilly Media, Inc.

December 5: Student Presentations

Version: December 5, 2018