

MGMT 737: Applied Empirical Methods (Quantitative Storytelling) Professor Olav Sorenson - Fall 2019

Mondays 1:00pm to 4:00pm (165 Whitney Ave, Room 2230)

Course overview

This course has been designed as a Ph.D.-level survey of empirical methods used in current research on organizations and in sociology. It has three main goals: (1) to introduce students to practical issues that arise in analyzing data; (2) to develop an intuition for why and how various estimation choices matter; and (3) to provide students with a better sense of ways in which one might provide a compelling set of analyses in support of some theoretical story. The course will also give particular attention to common methods in organizations and sociology that often receive less attention in courses on econometrics, such as the modeling of binary and count dependent variables and of survival rates.

The subtitle refers to "storytelling" for two reasons. First, in many cases, the nature of the question of interest or of the data available preclude the identification of a causal effect. Data analyses in these cases remain descriptive. They nevertheless often attempt to convince the reader of the plausibility of one particular interpretation of the relationships observed. Second, even in cases that allow for identification of a causal effect, that estimation often serves as a necessary but not a sufficient condition to providing empirical support for the underlying theoretical mechanism of interest. One should not read "storytelling" to mean that the treatment of methods in the course will not be serious.

Note that the teaching of the course will presume that students arrive with some familiarity with general linear models and causal inference, at least at a theoretical level (e.g., ECON 550a, PLSC 503b, SOCY 501b). In the spirit of a "flipped" class, most of the time spent together will be spent on reading and critiquing articles and on analyzing data and discussing the results. Although I am happy to answer any questions that may arise, I will generally not lecture on the content of the background readings in class. Because of the learning-by-doing nature of the course, the workload will likely end up being on the higher side, even relative to other Ph.D.-level offerings. Plan accordingly.

Assignments and exercises for the course will use Stata as the default but students may use other statistical programming languages with the understanding that the instructor may have limited ability to support them in those languages.

Assessment

Evaluation for this course consists of three components: class participation, assignments, and a final paper.

Component	Weight
Participation	15%
Assignments	60%
Final paper	25%

Class participation: Each week will consist mostly of discussion and working through analyses. I expect that students will have read the assigned readings prior to the lecture and come prepared with any questions they may have.

Assignments: Much of the course involves learning-by-doing. Almost every week therefore has an assignment to put the ideas covered and discussed into practice.

Final paper: As a final paper, you should find a published paper for which you can obtain the data used (either because the authors have made the data available or because they use a data source that you can access). First, try to replicate their results exactly. Second, explore the robustness of the results to alternative modelling choices. Third, propose at least one modification or extension to the story and explore that proposition empirically.

Instructor

You can reach me by e-mail at olav.sorenson@yale. If you wish to meet with me outside of class, please see my calendar at https://olavsorenson.youcanbook.me. It will show my availability each week and allow you to book a time slot.

Background reading

Although I do not plan to spend much time on discussing workflows, I would recommend reading one or both of the following guides (particularly if you have not already developed an almost-fully-automated approach to analyzing data and producing output and papers):

- Recommended: Gentzkow, Matthew, and Jesse M. Shapiro (2014). Code and Data for the Social Sciences: A Practitioner's Guide. Available at: https://web.stanford.edu/gentzkow/research/CodeAndData.pdf
- Recommended: Healy, Kieran (2017). The Plain Person's Guide to Plain Text Social Science. Available at: https://kieranhealy.org/resources/
- Extended reference: Long, J. Scott (2009). The Workflow of Data Analysis Using Stata. College Station, TX: Stata Press

Schedule

- 1. August 30: What makes (quantitative) analyses compelling?
 - **Discussion paper:** Sorenson, Olav, and Pino G. Audia (2000). "The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States, 1940-1989." *American Journal of Sociology*, 106: 424-462
 - **Discussion paper:** Desmond, Matthew, Andrew V. Papachristos, and David S. Kirk (2016). "Police violence and citizen crime reporting in the black community." *American Sociological Review*, 81: 857-876
 - **Discussion paper:** Ferguson, John-Paul, and Rembrand Koning (2018). "Firm turnover and the return of racial establishment segregation" *American Sociological Review*, 83: 445-474
 - Discussion paper: Srivastava, Sameer B., Amir Goldberg, V. Govind Manian, and Christopher Potts (2018). "Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations" *Management Science*, 64: 1348-1364

- 2. September 9: Causal interpretations versus mechanisms
 - **Pre-class assignment:** Draw graphs of the causal structure for two of the four discussion papers from August 30. What would you consider the primary threat(s) to a causal interpretation of the results.
 - Background: Knight, Carly R., and Christopher Winship (2013). "The causal implications of mechanistic thinking: Identification using directed acyclic graphs (DAGs)." Chapter 14 in S.L. Morgan (ed.), *Handbook for Causal Analysis for Social Research*. Dordrecht: Springer Science+Business
 - **Background:** Imbens, Guido W. (2019). "Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics." Working paper, Stanford University.
 - Background: Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres (2018). "How conditioning on post-treatment variables can ruin your experiment and what to do about it." *American Journal of Political Science*, 62: 760-775
 - Recommended: Angrist, Joshua D., and Jörn-Steffen Pischke (2008). "The experimental ideal." Chapter 2 of Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press
 - Recommended: Pearl, Judea (2012). "The mediation formula: A guide to the assessment of causal pathways in nonlinear models" Chapter 12 of Causality: Statistical Perspectives and Applications. West Sussex: John Wiley & Sons
 - Extended reference: Morgan, Stephen L., and Christopher Winship (2007). "Causal graphs, identification, and models of causal exposure." Chapter 3 of *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* Cambridge: Cambridge University Press
- 3. September 16: Matching methods
 - **Discussion paper:** Eggers, J.P., and Lin Song (2015). "Dealing with failure: Serial entrepreneurs and the costs of changing industries between ventures." *Academy of Management Journal*, 58: 1785-1803
 - **Discussion paper:** Feldman, Maryann P., Serdan Ozcan, and Toke Reichstein (2019). "Falling not far from the tree: Entrepreneurs and organizational heritage." *Organization Science*, 30: 337-360
 - Background: Iacus, Stefano M., Gary King, and Giuseppe Porro (2012). "Causal inference without balance checking: Coarsened exact matching." *Political Analysis*, 20: 1-24
 - **Background:** King, Gary, Richard Nielsen, Carter Coberly, James E. Pope, and Aaron Wells (2011). "Comparative effectiveness of matching methods for causal inference." Working paper, Harvard University.

- *Recommended:* Rosenbaum, Paul R., and Donald B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects." *Biometrika*, 70: 41-55
- 4. September 23: Fixed effects and differences-in-differences
 - **Discussion paper:** Zhang, Letian (2017). "A fair game? Racial bias and repeated interaction between NBA coaches and players." *Administrative Science Quarterly*, 62: in press
 - **Discussion paper:** McDonnell, Mary-Hunter, and Brayden King (2013). "Keeping up appearances: Reputational threat and impression management after social movement boycotts." *Administrative Science Quarterly*, 58: 387-419
 - **Discussion paper:** Kang, Hyo, and Lee Fleming (2019). "Non-competes, business dynamism, and concentration: Evidence from a Florida case study." Working paper, University of Southern California
 - Background: Angrist, Joshua D., and Jörn-Steffen Pischke (2008). "Individual fixed effects." Chapter 5.1 of *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press
 - Background: Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics*, 119: 249-276
 - *Extended reference:* Athey, Susan, and Guido Imbens (2006). "Identification and inference in nonlinear difference-in-differences models." *Econometrica*, 74: 431-497
 - *Extended reference:* Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association*, 105: 493-505
- 5. September 30: Instrumental variables I
 - **Discussion paper:** Ananat, Elizabeth Oltmans (2011). "The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality." *American Economic Journal: Applied Economics*, 3: 34-66
 - Discussion paper: Kang, Jingoo, and Andy Y. Han Kim (2017). "The relationship between CEO media appearances and compensation." Organization Science, 28: 379-394
 - Background: Angrist, Joshua D., and Alan B. Krueger (2001). "Instrumental variables and the search for identification: From supply and demand to natural experiments." *Journal of Economic Perspectives*, 15: 69-85

- Background: Murray, Michael P. (2006). "Avoiding invalid instruments and coping with weak instruments." *Journal of Economic Perspectives*, 20: 111-132
- Background: Young, Alwyn (2019). "Consistency without inference: Instrumental variables in practical application." Working paper, London School of Economics
- Extended reference: Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2011). "Plausibly exogenous." Review of Economics and Statistics, 94: 260-272
- 6. October 7: Instrumental variables II
 - **Discussion paper:** Emran, M. Shahe, and Zhaoyang Hou (2013). "Access to markets and rural poverty: Evidence from household consumption in China." *Review of Economics and Statistics*, 95: 682-697
 - **Discussion paper:** Gupta, Abhinav, Forrest Briscoe, and Donald C. Hambrick (2017). "Red, blue, and purple firms: Organizational political ideology and corporate social responsibility." *Strategic Management Journal*, 38: 1018-1040
 - Background: Bond, Stephen R. (2002). "Dynamic panel data models: A guide to micro data methods and practice." *Portuguese Economic Journal*, 1: 141-162
 - Background: Wooldridge, Jeffrey M. (2015). "Control function methods in applied econometrics." *Journal of Human Resources*, 50: 420-445
 - **Background:** Quiroga, Bernardo F. (2018). "Addressing endogeneity without strong instruments: A practical guide to heteroskedasticity-based instrumental variables." Working paper, Pontifica Universidad
 - Recommended: Lewbel, Arthur (2012). "Using heteroscedasticity to identify and estimated mismeasured and endogenous regressor models." Journal of Business & Economic Statistics, 30: 67-80
- 7. October 21: Regression discontinuity
 - Discussion paper: Liu, Ka-Yuet, Marissa King, and Peter S. Bearman (2010). "Social influence and the autism epidemic" American Journal of Sociology, 115: 1387-1434
 - **Discussion paper:** Kerr, William R., Josh Lerner, and Antoinette Shoar (2014). "The consequences of entrepreneurial finance: Evidence from angel financings." *Review of Financial Studies*, 27: 20-55
 - Background: Imbens, Guido, and Thomas Lemieux (2008). "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142: 615-635

- Background: Keele, Luke J., and Rocio Titiunik (2015). "Geographic boundaries as regression discontinuities." *Political Analysis*, 23: 127-155
- **Background:** Kelly, Morgan (2019). "The standard errors of persistence." Working paper, University College Dublin
- Recommended: Angrist, Joshua D., and Jörn-Steffen Pischke (2008). "Getting a little jumpy: Regression discontinuity designs." Chapter 6 of Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press
- Recommended: Gelman, Andrew, and GUido Imbens (2019). "Why highorder polynomials should not be used in regression discontinuity designs." Journal of Business & Economic Statistics, 37: 447-456
- 8. October 28: Telling the story with graphics
 - **Pre-class assignment:** Find at least one graph from a published paper that you find convincingly illustrates the central claim of the paper. Please e-mail a PDF of the article to olav.sorenson@yale.edu by noon on October 27.
 - Recommended: Tufte, Edward R. (2001). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press
- 9. November 4: Qualitative dependent variables (LPMs, logit, count, survival)
 - **Discussion paper:** Lutter, Mark (2015). "Do women suffer from network closure? The moderating effect of social capital on gender inequality in a project-based labor market, 1929 to 2010" *American Sociological Review*, 80: 329-358
 - **Discussion paper:** Ter Wal, Anne L.J., Oliver Alexy, Jörn Block, and Phillip G. Sandner (2016). "The best of both worlds: The benefits of openspecialized and closed-diverse syndication networks for new ventures' success" *Administrative Science Quarterly*, 61: 393-432
 - Background: Angrist, Joshua D., and Jörn-Steffen Pischke (2008). "Limited dependent variables and marginal effects." Chapter 3.4.2 of *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press
 - Background: Mood, Carina (2010). "Logistic regression: Why we cannot do what we think we can do, and what we can do about it." *European Sociological Review*, 26: 67-82
 - Background: Pötter, Ulrich, and Götz Rohwer (2007). Introduction to Event History Analysis. Ruhr University

- *Recommended:* Ai, Chunrong, and Edward C. Norton (2003). "Interaction terms in logit and probit models." *Economics Letters*, 80: 123-129
- Recommended: King, Gary (1989). "Variance specification in event count models: From restrictive assumptions to a generalized estimator." American Journal of Political Science, 33: 762-784
- Extended reference: Blackburn, Mckinley L. (2015). "The relative performance of Poisson and negative binomial regression estimators." Oxford Bulletin of Economics and Statistics, 77: 605-616
- Extended reference: Muris, Chris (2017). "Estimation in the fixed-effects ordered logit model." Review of Economics and Statistics, 99: 465-477
- Extended reference: Cleves, Mario, William W. Gould, and Yulia V. Marchenko (2016). An Introduction to Survival Analysis Using Stata, Revised Third Edition. Stata Press
- 10. November 11: Analyzing social relatonships
 - **Discussion paper:** Sorenson, Olav, and Toby E. Stuart (2008). "Bringing the context back in: Settings and the search for syndicate partners in venture capital investment networks" *Administrative Science Quarterly* 53: 266-294
 - **Discussion paper:** Hasan, Sharique, and Surendrakumar Bagde (2013). "The mechanics of social capital and academic performance in an Indian college" *American Sociological Review*, 78: 1009-1032
 - Discussion paper: Goodreau, Steven M., James A. Kitts, and Martina Morris. (2009). "Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks." *Demography* 46: 103-125
 - Background: Manski, Charles F. (1993). "Identification of endogenous social effects: The reflection problem. "*Review of Economics and Statistics* 60: 531-542
- 11. NOVEMBER 18: ANALYZING SPATIAL DATA
 - Discussion paper: Tolnay, Stewart E., Glenn Deane, and E.M. Beck (1996). "Vicarious violence: Spatial effects on Southern lynchings, 1890-1919." American Journal Sociology, 102: 788-815
 - **Discussion paper:** Kabo, Felichism, Yongha Hwang, Margaret Levenstein, and Jason Owen-Smith (2015). "Shared paths to the lab: A sociospatial network analysis of collaboration." *Environment and Behavior*, 47: 57-84
 - Background: Anselin, Luc, Anil K. Bera, Raymond Florax, and Mann J. Yoon (1996). "Simple diagnostic tests for spatial dependence." *Regional Science and Urban Economics*, 26: 77-104.

- Background: Beck, Nathaniel, Kristian Skrede Gleditsch, and Kyle Beardsley (2006). "Space is more than geography: Using spatial econometrics in the study of political economy." *International Studies Quarterly*, 50: 27-44.
- *Recommended:* Fotheringham, A. Stewart and David W.S. Wong (1991). "The modifiable areal unit problem in multivariate statistical analysis." *Environment and Planning A*, 23: 1025-1034.
- Extended reference: Strang, David, and Nancy Brandon Tuma (1993). "Spatial and temporal heterogeneity in diffusion." American Journal of Sociology, 99: 614-639
- 12. December 2: Analyzing unstructured data (Text)
 - Discussion paper: Bail, Christopher A., Taylor W. Brown, and Marcus Mann (2017). "Channeling hearts and minds: Advocacy organizations, cognitive-emotional currents, and public conversation." *American Sociological Review*, 82: 1188-1213
 - **Discussion paper:** Fligstein, Neil, Jonah Stuart Brundage, and Michael Schultz (2017). "Seeing like the Fed: Culture, cognition, and framing in the failure to anticipate the financial crisis of 2008" *American Sociological Review*, 82: 879-909
 - **Discussion paper:** Vilhena, Daril A., Jacob G. Foster, Martin Rosvall, Jevin D. West, James Evans, and Carl T. Bergstrom (2017). "Finding cultural holes: How structure and culture diverge in networks of scholarly communication" *Sociological Science*, 1: 221-258
 - Background: Grimmer, Justin, and Brandon M. Stewart (2013). "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis*, 21: 267-297.
 - Background: Evans, James and Pedro Aceves (2016). "Machine translation: Mining text for social theory." Annual Review of Sociology, 42: 21–50.
 - Background: Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Rogers, and Brandon M. Stewart (2017). "How to make causal inferences using text." Working paper, Princeton University
 - *Recommended:* Landauer, Thomas K., Peter W. Foltz, and Darrell Laham (1998). "Introduction to latent semantic analysis." *Discourse Processes*, 25: 259-284
 - Recommended: Chang, Jonathan, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei (2009). "Reading tea leaves: How humans interpret topic models." Proceedings of Neural Information Processing Systems