Advanced Applied Statistics

Sune Welling Hansen
Robert Klemmensen
Erik Gahner Larsen

September 2, 2015

1 Course overview

Social scientists study complex phenomena. In order to study these phenomena using quantitative information, they have to use suitable statistical methods that take into account the properties of the research question, research design and the data. The purpose of this course is to give students a detailed understanding of select advanced statistical methods and give them the skills required to implement these methods in either Stata or R, two very popular statistical programs. In many cases, the theoretical and empirical interests of the students result in both interesting and complex topics of study, but using only basic statistical methods would make it impossible to analyze these questions, or would result in less valid (perhaps even invalid) results and conclusions. The main objective of this course is to teach students how they can overcome these problems by providing them with the necessary skills to employ state-of-the-art statistical methods.

Upon completion, students are expected to gather and develop the following knowledge, skills, and competence related aspects:

Knowledge

1. Understand the principles of statistical analysis and the assumptions behind select statistical methods
2. Discuss and understand assumptions behind these methods
3. Evaluate whether these assumptions are correct, or realistic, and correct for breaches with viable solutions (through the advanced methods discussed in the course) for complex problems or data-related issues
4. Interpret results from these select statistical methods

Skills

1. Select appropriate statistical methods for own research questions
2. Analyze and critically assess scientific articles that use the advanced statistical methods taught in the course

Competences

1. Implement advanced statistical methods using either Stata or R
2. Comfortably use statistical software for own analyses
3. Correctly select and interpret the quantities of interest resulting from the application of these methods and link these results to research goals

2 Prior knowledge, mandatory assignments and the exam

Students are expected to have taken at least a course on social scientific method (i.e. be familiar causality, measurement, research design, validity and reliability etc.) and a prior quantitative methods class (i.e. what is a variable, significance testing and confidence intervals, correlation and regression etc.). While having experience with either Stata or R will be to the benefit of the student, it is not a mandatory prerequisite.

In order to be able to take the exam, students *have to complete and pass* four assignments during the semester. All the assignments are empirical applications, and are graded as passed/not passed. Those who do not pass all four assignments by the date stipulated below will not be able to attend the exam.

The exams consist of a one week take-home written exam (max. 10 pages) graded as passed/not passed or on the 7-point scale. The exam will be an empirical application of one or more theoretical arguments.

No.	Topic	Teacher	Date	Time	Room
1	Introduction and refresher to basic statistics	Sune	2 September	16-18	U49E
2	Refresher of linear regression	Robert	9 September	16-18	U49E
3	How to study causality in non-experimental studies	Sune	16 September	16-18	U49E
4	Binary logistic regression	Sune	23 September	16-18	U49E
5	Ordered and multinomial logistic regression	Sune	30 September	16-18	U49E
6	Time & space: Models for independent pooled cross sections and panel data	Sune	7 October	16-18	U49E
7	Time & space: Fixed effects and random effects models	Sune	21 October	16-18	U49E
8	Hierarchical models I	Robert	28 October	16-18	U49E
9	Hierarchical models II	Robert	4 November	16-18	U49E
10	Experiments and Causal Inference	Erik	11 November	16-18	U49E
11	Matching and Propensity Scores	Erik	18 November	16-18	U49E
12	Regression-Discontinuity Designs	Erik	25 November	16-18	U49E
13	Instrumental Variable Regression	Erik	2 December	16-18	U49E
14	Factor analysis	Robert	9 December	16-18	U49E
15	Communicating results	Robert	16 December	16-18	U49E

4 Overview of labs

Students are required to bring their own laptops to the lab sessions. All details related to the practical parts (data and script) will be uploaded to Blackboard.

No.	Topic	Date	Time	Room
1	Linear regression	10 September	14-16	U49C
2	Data simulation and binary logistic regression	24 September	14-16	U49C
3	Ordered and multinomial logistisc regression	1 October	14-16	U49C
4	Data over time & space	22 October	14-16	U49C
5	Hierarchical models	5 November	14-16	U49C
6	Experimental data I	19 November	14-16	U49C
7	Experimental data II	26 November	14-16	U49C

5 Overview of mandatory assignments

The mandatory assignments are basically graded homeworks and all the information necessary to complete the assignment will be supplied on the hand-out date. Assignments will be handed out through Blackboard, are due the following week and must be submitted electronically through Blackboard, using the examination number of the student as identifier.

No.	Topic	Hand-out date	Due date	Feedback by	Resubmission by
1	Linear regression	10 September	16 September	23 September	30 September
2	Logistic regression	1 October	7 October	21 October	28 October
3	Data over space and/or time	5 November	11 November	18 November	25 November
4	Experimental data	26 November	2 December	7 December	10 December (!)

Please consult the above table **very carefully** and plan accordingly. All assignments are designed to evaluate the progress towards the learning outcomes stipulated in the course description.

6 Readings for lectures

The curriculum consists of a number of texts (books, chapters from books, articles and online documents), and the curriculum for each lecture is described below. Texts marked with an asterisk ('*') are found in the compendium for the course, and you have to download (or access) the rest via the university library's website (the exception being texts with a double asterisk ('**') which are made available for copying before the lecture.)

Note that books from the Quantitative Applications in the Social Sciences series by Sage (or Little Green Books) - such as Davis (1985) and Fox (1991) - are available online through the university library collection, and hence they are not included in the compendium.

1. Introduction and refresher to basic statistics (Sune)

Review literature and notes from your BA course in statistics. Examples of textbooks:

Agresti, A. & Finlay, B. (2009). *Statistical methods for the social sciences*. New Jersey: Pearson. Chapters 4-8.

Donnelly, R. A. (2007). *Complete idiot's guide statistics*. New York: Alpha Books (a very pedagogical refresher).

Gravetter, F. & Wallnau, L. (2008). *Essentials of statistics for the behavioral sciences*. Cengage Learning.

Make also sure that Stata and/or R are installed on your computer.

2. Refresher on linear regression (Robert)

Required readings:

** Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press. Chapter 3 and 4.

Achen, C. H. (1982). *Interpreting and using regression*. Sage. Pages 46-51.

* Darcy, R. (1991). "More latitude for women in national parliaments". *Journal of Irreproducible Results*, 37, 7–8.

Brambor, T., Clark, W. R., & Golder, M. (2006). "Understanding interaction models: Improving empirical analyses". *Political Analysis*, 14(1), 63–82.

A detailed look at http://mattgolder.com/interactions.

Recommended readings:

Jasso, G. (1986). "Is it outlier deletion or is it sample truncation? Notes on science and sexuality". *American Sociological Review*, 738–742. (Jasso's discussion with Kahn and Udry is an entertaining discussion of whether or not to include outliers in an empirical analysis.)

Kahn, J. R. & Udry, J. R. (1986). "Marital coital frequency: Unnoticed outliers and unspecified interactions lead to erroneous conclusions". *American Sociological Review*, 734–737.

Fox, J. (1991). *Regression diagnostics: An introduction*. Sage.

3. How to study causality in non-experimental studies (Sune)

Required readings:

* Kellstedt, P. & Whitten, G. (2009). *The fundamentals of political science research*. Cambridge University Press. Pages 77-83.

* Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press. Pages 89-94.

Davis, J. A. (1985). *The logic of causal order*. Sage.

4. Binary logistic regression (Sune)

Required readings:

* Agresti, A. & Finlay, B. (2009). *Statistical methods for the social sciences*. New Jersey: Pearson. Pages 483-495.

* Rose, L. E. (2002). "Municipal size and local nonelectoral participation: findings from Denmark, the Netherlands, and Norway". *Environment and Planning C: Government and Policy*, 20, 1-23.

Recommended readings:

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press. Chapter 5. (for labs, good source for logistic regression in R)

Sønderskov, K. M. (2015). *Stata: A Practical Introduction*. Hans Reitzel. Chapter 11. (for labs, good source for logistic regression in Stata)

5. Ordered and multinomial logistic regression (Sune)

Required readings:

* Agresti, A. & Finlay, B. (2009). *Statistical methods for the social sciences*. New Jersey: Pearson. Pages 496-503.

Hobolt, S. B. (2012). "Citizen Satisfaction with Democracy in the European Union". *Journal of Common Market Studies*, 50(1), 88-105.

Schoen, H. & Schumann, S. (2007). "Personality Traits, Partisan Attitudes, and Voting Behavior. Evidence from Germany". *Political Psychology*, 28(4), 471-492.

6. Time & space: Models for independent pooled cross sections and panel data (Sune)

Required readings:

* Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press. Chapter 13.

Serritzlew, S. (2005). "Breaking Budgets: An Empirical Examination of Danish Municipalities". *Financial Accountability & Management*, 21(4), 413-435.

Hansen, S. W. (2013). "Polity Size and Local Political Trust: A Quasi-experiment Using Municipal Mergers in Denmark". *Scandinavian Political Studies*, 36(1), 43-66.

7. Time & space: Fixed effects and random effects models (Sune)

Required readings:

* Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press. Chapter 14.

8. Hierarchical models I (Robert)

Required readings:

* Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. Chapters 11 and 12.

Steenbergen, M. R. & Jones, B. S. (2002). "Modeling multilevel data structures". *American Journal of Political Science*, 218–237.

Recommended readings:

King, G. & Roberts, M. E. (2015). "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It". *Political Analysis*, 23(2), 159-179.

9. Hierarchical models II (Robert)

Required readings:

** Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. Chapter 13.

Pittau, M. G., Zelli, R., & Gelman, A. (2010). "Economic disparities and life satisfaction in European regions". *Social Indicators Research*, 96(2), 339–361.

Stegmueller, D. (2013). "How many countries for multilevel modeling? A comparison of frequentist and bayesian approaches". *American Journal of Political Science*, 57(3), 748–761.

Recommended readings:

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*.

Cambridge University Press. Chapters 20 and 21.

Enders, C. K. & Tofighi, D. (2007). "Centering predictor variables in cross-sectional multilevel models: A new look at an old issue". *Psychological Methods*, 12(2), 121–138.

Ghitza, Y. & Gelman, A. (2013). "Deep interactions with mrp: Election turnout and voting patterns among small electoral subgroups". *American Journal of Political Science*, 57(3), 762–776.

10. Experiments and Causal Inference (Erik)

Required readings:

** Gelman, A. & J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. Sections 9.1, 9.2, 9.3 and 9.6.

Druckman, J. N. & A. Lupia (2012). "Experimenting with Politics". *Science*, 335(6073): 1177-1179.

* Gerber, A. S. & D. P. Green (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton & Company. Chapter 2.

Recommended readings:

Holland, P. W (1986). "Statistics and Causal Inference". *Journal of the American Statistical Association*, 81(396), 945-960.

Morgan, S. L. & C. Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Morton, R. B. & K. C. Williams (2010). Experimental Political Science and the Study of Causality: From Nature to the Lab. New York: Cambridge University Press.

Mutz, D. (2011). *Population-Based Survey Experiments*. New Jersey: Princeton University Press.

Rubin, D. B. (2008). "For Objective Causal Inference, Design Trumps Analysis". *The Annals of Applied Statistics*, 2(3), 808-840.

Recommended readings in Danish

Blom-Hansen, J. & S. Serritzlew (2014). "Endogenitet og eksperimenter – forskningsdesignet som løsning". *Politica*, 46(1), 5-23.

11. Matching and Propensity Scores (Erik)

Required readings:

* Gelman, A. & J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Mod-*

els. New York: Cambridge University Press. Sections 10.1, 10.2 and 10.3.

Kam, C. D. & C. L. Palmer (2008). "Reconsidering the Effects of Education on Political Participation". *Journal of Politics*, 70(3), 612-631.

Sekhon, J. S. (2009). "Opiates for the Matches: Matching Methods for Causal Inference". *Annual Review of Political Science*, 12, 487-508.

Recommended readings:

Henderson, J. & S. Chatfield (2011). "Who Matches? Propensity Scores and Bias in the Causal Effects of Education on Participation". *Journal of Politics*, 73(3), 646-658.

Mayer, A. K. (2011). "Does Education Increase Political Participation?" *Journal of Politics*, 73(3), 633-645.

Ho, D. E., K. Imai, G. King & E. A. Stuart (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference". *Political Analysis*, 15(3), 199-236.

Guo, S. & M. W. Fraser (2010). *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks: Sage.

Recommended readings in Danish

Justesen, M. K. & R. Klemmensen (2014). "Sammenligning af sammenlignelige observationer: kausalitet, matching og observationsdata". *Politica*, 46(1), 60-78.

12. Regression-Discontinuity Designs (Erik)

Required readings:

* Gelman, A. & J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. Section 10.4.

*/** Dunning, T. (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York: Cambridge University Press. Chapter 3 and section 5.2 (last-mentioned will be handed out prior to the lecture).

Recommended readings:

Imbens, G. W. & T. Lemieux (2008). "Regression discontinuity designs: A guide to practice". *Journal of Econometrics*, 142(2), 615-635.

Lee, D. S. & T. Lemieux (2010). "Regression Discontinuity Designs in Economics". *Journal of Economic Literature*, 48(2), 281-355.

Shadish, W. R., T. D. Cook & D. T. Campbell (2002). *Experimental and Quasi-Experimental*

Designs for Generalized Causal Inference. Belmont, CA: Wadsworth Cengage learning. Chapter 7.

Recommended readings in Danish

Olsen, A. L. (2014). "Tærskelvariable og tærskelværdier: en introduktion til regressionsdiskontinuitetsdesignet". *Politica*, 46(1), 42-59.

13. Instrumental Variable Regression (Erik)

Required readings:

* Gelman, A. & J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. Section 10.5 and 10.6.

* Dunning, T. (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York: Cambridge University Press. Chapter 4.

Sovey, A. J. & D. P. Green (2011). "Instrumental Variables Estimation in Political Science: A Readers' Guide". *American Journal of Political Science*, 55(1), 188-200.

Recommended readings:

Angrist, J. D. & Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. New Jersey: Princeton University Press.

Bollen, K. A. (2012). "Instrumental Variables in Sociology and the Social Sciences". *Annual Review of Sociology*, 38, 37-72.

Dunning, T. (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York: Cambridge University Press.

Dunning, T. (2008). "Model Specification in Instrumental-Variables Regression". *Political Analysis*, 16(3), 290-302.

Angrist, J. D. & A. B. Krueger (2001). "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments". *Journal of Economic Perspectives*, 15(4), 69-85.

Recommended readings in Danish:

Hariri, J. G. (2014). "Statskundskabens sammenfiltrede virkelighed og et bud på en løsning: IV-estimation". *Politica*, 46(1), 79-94.

14. Factor analysis (Robert)

Required readings:

TBA

Recommended readings:

TBA

15. Communicating results (Robert)

Weingast, B. R. (2010). *Caltech rules for writing papers: How to structure your paper and write an introduction*. Available here.

* Miller, J. E. (2005). *The Chicago guide to writing about multivariate analysis*. Chicago University Press. Chapters 1, 2, 3 and 14.