Syllabus for STV4020B

Silje Synnøve Lyder Hermansen
s.s.l.hermansen@jus.uio.no

Øyvind Stiansen
oyvind.stiansen@jus.uio.no.

Fall 2019

Course Overview

This course provides a toolkit for students looking to execute their own research projects using quantitative methods. We go through the statistical theory behind models, but we emphasize understanding when models are useful and how to employ them. The course helps students 1) identify appropriate statistical models that describe different data types and 2) design studies that enable causal inference (as opposed to merely establishing correlation).

Many of the phenomena political scientists take interest in can be classified into categories or events. This includes variables like voters' choice of party, number of violent incidences in a time span or time between violent events. The course introduces students to a number of models which are specifically designed to describe the underlying phenomenon that generates such data. The focus of the first part of the course is on correlations. Our purpose is to help students gear the statistical analysis towards a realistic description of the data.

Beyond statistical description, researchers' interest in political phenomena is often motivated by a wish to identify their causes. Causal claims focus on whether we would continue to observe y even in the absence of x. This is a typical question when studying the effect of policies, for example. The potential outcomes framework provides a useful heuristic for assessing causal relationships. The second part of the course is devoted to a set of quantitative research designs that can be used to make causal inferences.

Readings

We will use two textbooks corresponding to the two parts of the course:

Ward and Ahlquist (2018) will introduce students to maximum likelihood as a framework for statistical modelling, different generalized linear models suited for various data types frequently encountered in political science, model selection strategies, and methods for interpreting and presenting empirical results.

Angrist and Pischke (2014) will introduce students to the potential outcomes framework for causal inference and research designs that may allow causal interpretations of empirical results.

Some students may find Ward and Ahlquist (2018) to be more technically demanding than Angrist and Pischke (2014). If so: Do not despair! We will use the lectures to clarify and provide intuition. However, we do encourage you to not skip the equations. A useful approach is to engage actively with the examples. Working through as much as possible of the maths will give you a much better grasp of the material!

Recommended readings Sometimes we understand things better when they are explained in different ways. You may find it useful to consult the following books for alternative explanations throughout the course: Fox (2008), Stock and Watson (2012) and Hermansen (2019).

Teaching and evaluation

The teaching is structured around two plenary lectures and one seminar per week. The curriculum is relatively limited in terms of page count, and students are expected to read all assigned readings carefully prior to the relevant session. The seminars will be devoted to implementing the models – their estimation and interpretation – in the statistical package R. That is, while the lectures will focus on identifying and understanding different modelling strategies, the seminars will provide the students with the practical skill set to execute them.

The evaluation consists of one compulsory and one optional (but highly recommended) part.

Weekly home assignment (optional) Each week you will be asked to complete a home assignment. It will consist of analyzing a dataset relevant for the topic of the week and to critically evaluate and interpret the results. The home assignment is made available after the Thursday lecture and is due 23:59 the next Monday. Completing the assignments is optional, but strongly encouraged. It forms the basis for the term paper, and only students submitting their home assignments on time can expect feedback.

The length of the home assignment should be between 800 and 1000 words. Figures, tables, and references do not count towards the word limit but all tables and figures included should be discussed in the main text. You will also need to enclose the R code needed to reproduce your results. We will not enforce any length limit on the code (in terms of lines), but the code should be organized neatly, be easy to read, and include appropriate comments that describe what you are doing.

Term paper (compulsory) The term paper should be a compilation of all the home assignments. You should incorporate the feedback received based on the home assignments and resubmit everything as a single document. R code for each part should be appended to the document.

The word limits for each component of the term paper are the same as for the home assignments.

Helpful advice

Ward and Ahlquist (2018: 101) provide some helpful advice for how to manage your workflow and avoid errors. In particular, they suggest that you should be working exclusively with scripts (i.e. R scripts) to manipulate data, estimate models, and generate outputs in terms of graphs and tables. They also caution against relying on copy/paste when writing up your results. Instead, it is useful to generate all tables and figures directly from your R scripts and choose a typesetting system where you can automatically include the most recent version of the relevant table/figure. For this reason, RMarkdown or LaTeX are superior to Microsoft Word or similar programs. While these programs have a slight learning curve, they will make your life much easier in the long run. A variety of R packages make it very easy to export BUTONs[1] or other output to LaTeX. With this helpful advice, we leave the decisions about whether to make the transition entirely up to you!

Plenary Sessions

[1]See Ward and Ahlquist (2018: 98) for an explanation of the acronym!

Date	Time	Topic	Readings	Lecturer
Tu. 12. Nov.	12:15–14:00	Introduction and course overview, review of linear regression, and introduction to maximum likelihood	Ward and Ahlquist (2018: Chapters 1,2,5,6,7)	Øyvind
Th. 14. Nov.	10:15–12:00	Models of outcome and choice (logistic regression)	Ward and Ahlquist (2018: Chapters 3)	Silje
Tu. 19. Nov.	12:15–14:00	Models of outcome and choice (ordered and multinomial logit)	Ward and Ahlquist (2018: Chapters 8, 9)	Silje
Th. 21. Nov.	10:15–12:00	Modelling count outcomes	Ward and Ahlquist (2018: Chapter 10)	Silje
Tu. 26. Nov.	12:15–14:00	Duration models	Ward and Ahlquist (2018: Chapter 11)	Øyvind
Th. 28. Nov.	10:15–12:00	Duration models	Ward and Ahlquist (2018: Chapter 11)	Øyvind
Tu. 3. Dec.	12:15-14:00	Causal inference, the potential outcomes framework, experimental designs	Angrist and Pischke (2014: Chapter 1)	Silje
Th. 5. Dec.	10:15-12:00	Causal inference with observational data: regression and matching	Angrist and Pischke (2014: Chapter 2)	Øyvind
Tu. 10. Dec.	12:15-14:00	Causal inference with observational data: Regression discontinuities and differences-in-differences	Angrist and Pischke (2014: Chapters 4, 5)	Silje
Th. 12. Dec.	10:15-12:00	Causal inference with observational data: instrumental variables + Course summary	Angrist and Pischke (2014: Chapters 3, 6)	Øyvind

References

Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering'metrics: The path from cause to effect*. Princeton University Press.

Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*. 2 ed. London: Sage.

Hermansen, Silje Synnøve Lyder. 2019. *Lær deg R - En innføring i statistikkprogrammets muligheter*. Bergen: Fagbokforlaget.

Stock, James H. and Mark M. Watson. 2012. *Introduction to Econometrics*. Pearson series in economics ed. Essex: Pearson.

Ward, Michael D and John S Ahlquist. 2018. *Maximum Likelihood for Social Science: Strategies for Analysis*. Cambridge University Press.