Department of Psychology, Stockholm University
Doctoral program in Psychology, spring semester 2024

Syllabus: Statistics 2 (7.5 hp)

This course will follow the book "Statistical Rethinking" by Richard McElreath (2020, 2nd ed.) The book is an excellent introduction to statistical modeling in general, and modeling from a Bayesian perspective in particular. The approach of the book is to push the reader to perform step-by-step calculations that are usually automated, to ensure sufficient understanding of the details to make reasonable choices and interpretations in the reader's own statistical work. Calculations are done in the software R, so the course will also lead to improved R programming skills. For those new to R, support will be provided before each seminar.

The course include all parts of McElreath's book up to, but not including the last chapters on Multilevel modeling. **NOTE**: There is a follow up course in April, called Statistics 2.5 that covers part of chapter 13 and 14 of the book.

Stat 2 provide the basics of statistical modeling from the perspective of the Generalized Linear Model. This provide a good foundation to understand the complexities of multilevel modeling, and students are encouraged to continue on their own with the final chapters of the book (Chs. 13-16), or take Stat 2.5 or do both!

Prior knowledge
The course assumes prior knowledge corresponding to the course Statistics 1. Statistics 2 rely heavily on the statistical software R and its help-program R-studio. These are free open source software that can be downloaded from the web.

Learning outcomes
- Understanding of the main ideas of Bayesian statistics and its strengths and weaknesses in relation to conventional approaches to data analysis.
- Understanding the difference between prediction and causal inference and its consequences for selection of data analytic strategy.
- Skills in causal inference using directed acyclical graphs (DAGs).
- Understanding regression modeling, including issues related to over-and underfitting, model comparisons (including resampling methods for cross-validation), interaction effects, and generalized linear models with dichotomous or count data as outcome variables.
- Practical skills in R programming.
- Practical data analysis of own data (real or simulated) using the approach advocated by McElreath (2020).

Course content

- Probability from a Bayesian perspective
- Multivariate linear models, including dummy-variable regression
- Prediction and cross-validation
- Causal inference, basics
- Interpreting interaction effects
- Generalized Linear model: Binomial regression and Poisson regression
- Generalized Linear model: Overdispersion, Ordinal regression
- (The basics of Bayesian multi-level modelling is covered in Stat 2.5)
- R programming

Activities

A series of 14 seminars. The seminars will start with a theoretical discussion of a selection of topics covered by specific book chapters, followed by student presentations of solutions to selected exercises. Much of the seminar discussions will concern how to address problems and illustrate phenomena using R. It is therefore a good idea to bring a laptop with R and R-studio installed to each seminar.

In addition, there will be opportunities for help getting started with R before the first seminar, and help catching up during the course.

Examination

The course is graded *Pass* or *Fail*. *Pass* requires passing both of the two examination parts described below.

1. Individual assignment: Problems. Solving a set of my exercises, specified in chapter 16 of my notes (https://matni.bitbucket.io/stat2solutions/docs/).

2. Individual assignment: Paper. A report of analyses of data of the students own choice (real or simulated data). The analyses should follow analytic approaches discussed during the course. The student will present an outline of the planned analyses at the last seminar.

Please submit both assignments no later than 2 months after the last seminar. If revision is needed, the revision should be delivered no later than 3 months after the last seminar.

Literature

Abbrvtn	Books
SR	McElreath, R. (2020). *Statistical Rethinking. A Bayesian Course with Examples in R and Stan (2nd ed)*. New York: CRC Press. **Chapters 1-12**. [Available in electronic format from Stockholm University Library.]
Abbrvtn	Articles/websites
R	Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*. 1, 27-42.
T	Textor, J. (2021). *DAGitty*. Retrieved February 8, 2021, from dagitty.net
G	Gelman, A, Hill, J., & Vehtari, A. (2021). *Regression and Other Stories*. Cambridge, UK: Cambridge University Press. Chapter 18: Causal inference and randomized experiments (pp. 339-361).

The additional sources (R, T, G) are about causal inference. McElreath relies heavily on Directed Acyclical Graphs (DAGs), also covered in R and T. DAGs are based on the potential outcome model of causality, covered in G.

Schedule

Time: 13.00 – 16.00, Mondays and Thursdays from Feb 1 to Mar 28..
R-support provided upon request on seminar days (before or after seminar).
Place: Room *Smultron* at Department of Psychology, SU.

	Date	Topic	Literature	Exercises*
0	Feb 1	Help getting started with R, and libraries rethinking and rstan		
1	Feb 5	Introduction, Probability	SR1-2	MN 1, 2 SR 2M3-6 SR 2H1-3
2	Feb 8	Probability	SR2-3	MN 3 SR 2M1-2 SR 3M6
3	Feb 12	Linear models	SR4	SR 4M1-3 SR 4H1-2
4	Feb 15	Linear models	SR5	SR 5E1-4
5	Feb 19	Causal inference	SR5, R, T, G	MN 5 SR 5M1-3 SR 5H3
6	Feb 22	Causal inference	SR6, R, T, G	MN 6 SR 6E1-4 SR 6M1-3
	Feb 26	*Cath-up (if needed)*		
7	Feb 29	Model comparison	SR6-7	MN 7 SR 6H3-5 SR 7H4
8	Mar 4	Interactions	SR8	SR 8E1-3 SR 8H4
---	*Mar 7*	*Cath-up (if needed)*		
9	Mar 11	MCMC	SR9	SR 9M1-2 SR 9H5
10	Mar 14	Generalized linear models (Binomial)	SR10-11	SR 11M7 SR 11H4-5
11	Mar 18	Generalized linear models (Poisson)	SR10-11	MN 9 SR 11M8 SR 11H3
12	Mar 21	Generalized linear models (Overdispersed Binomial, Poisson)	SR12	MN 10
13	Mar 25	Generalized linear models (Ordinal)	SR12	MN 11
14	Mar 28	Presentations, individual assignment	---	

* *Note:* SR: Exercises from McElreath's book, notation: 2E1, 2M1, 2H1 for chapter 2, Easy, Moderate, and Hard problem, respectively. MN 1-11: see Chapter 15 in https://matni.bitbucket.io/stat2solutions/docs/ (MN4 and MN8 are reserved for the Individual Assignment: Problems)