Department of Psychology, Stockholm University
Doctoral program in Psychology, spring semester 2019

Syllabus: Statistics 2 (7.5 hp)

This course will follow the book "Statistical Rethinking" by Richard McElreath (2016) (we may also use parts of a draft of the 2nd edition, to be published in 2020). 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. The course provide the basics of statistical modeling needed to understand the complexities of multilevel modeling, and students are encouraged to continue on their own with the final chapters of the book (Chs. 12-15).

The course will also cover good practice in data management and documentation of analysis for reproducible science.

Prior knowledge

The course assumes prior knowledge corresponding to the course Statistics 1 (Master course, main book: Howell, 2012, *Statistical Methods for Psychology*, 8th International edition, ISBN: 978-1-111-84085-3). 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
- Skills in data management and documentation of data analysis for reproducible science.
- Understanding of the main ideas of Bayesian statistics and its strengths and weaknesses in relation to conventional approaches to data analysis.
- Understanding regression modeling, including issues related to over-and underfitting, model comparisons, 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 the course book (McElreath, 2016

Course content
- Data management and reproducible data analysis
- Probability from a Bayesian perspective
- Multivariate linear models, including dummy-variable regression
- Interpreting interaction effects
- Model comparison
- Generalized Linear model: Binomial regression and Poisson regression
- R programming

Activities

A series of 11 seminars. The seminars will start with a theoretical discussion of the topic covered by a specific book chapter, followed by student presentations of solutions to the selected exercises. Much of the seminar discussions will concern how to address problem and illustrate phenomena using R. It is therefore a good idea to bring a laptop with R and R-studio installed to each seminar.

1. **Introduction.** Ch.1 of McElreath (2016),
2. **Probability 1.** Ch. 2; selected exercises from Ch. 2 of McElreath (2016)
3. **Probability 2.** Ch. 3; selected exercises from Ch. 3 of McElreath (2016)
4. **Linear models.** Ch. 4; selected exercises from Ch. 4 of McElreath (2016)
5. **Multivariat linear models.** Ch. 5; selected exercises from Ch. 5 of McElreath (2016)
6. **Causal inference**. Directed Acyclical Graphs and related tools (cf. **McElreath, 2019**, Ch. 6)
7. **Overfitting and underfitting.** Ch. 6; selected exercises from Ch. 6 of McElreath (2016)
8. **Interactions.** Ch. 7; selected exercises from Ch. 7 of McElreath (2016)
9. **Markov Chain Monte Carlo**. Chapter 8; selected exercises from Ch. 8 of McElreath (2016)
10. **Generalized Linear models.** McElreath, Chs. 9-10; selected exercises from Ch. 10
11. **Presentations individual assignment.**

Seminar leader: Mats E. Nilsson.

Note: The plan above is tentative, and will be discussed at the first semianr.

Examination

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

1. Solving a set of exercises from McElreath (2016) not discussed during the course. Selected exercises will be handed out at the start of the course. Solutions should be delivered no later than 1 week after the last seminar. If revision is needed, the revision should be delivered no later than 2 weeks after the last seminar.

2. A report of analyzes of data of the students own choice (real or simulated data). The analyzes should follow the approach discussed in the course. The student will present the result of the analyzes at the last seminar, and should submit a report no later than 1 month after the last seminar. If revision is needed, the revision should be delivered no later than 2 months after the last seminar.

Literature

McElreath, R. (2016). *Statistical Rethinking. A Bayesian Course with Examples in R and Stan.* New York: CRC Press. Chapters 1-7, 9-10.

McElreath, R. (2019). *Statistical Rethinking. A Bayesian Course with Examples in R and Stan.* 2nd Ed. New York: CRC Press. To be made available as pdf (**many thanks to the author for sharing his draft version!**).

The first edition of the book (2016) is available in electronic format from Stockholm University Library.

Schedule

Dates (it's a Wednesday or Thursday): *April* 4, 11, 17, 25; *May* 2, 9, 16, 23, 29; and *June* 5, 13.

Time: 13.00 – 16.00. Possibilities for R-support will be provided 11.00-12.00 on the seminar days.

Place: Seminar room at Gösta Ekman Laboratory.