Syllabus: Statistics 2 (7.5 hp)

This course will follow the book “Statistical Rethinking” by Richard McElreath (2016). 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, a satellite course in basic R will be offered that runs in parallel with the main course.

The course will include all parts of McElreath’s book except those related to Multilevel modeling. These chapters are instead included in the course “Bayesian data analysis and multilevel modeling” (BAMM, 4 hp) given directly after Statistics 2. Statistics 2 covers the basics of modeling needed to understand the complexities of multilevel modeling. Students planning to take BAMM are therefore strongly recommended to first take Statistics 2.

Prior knowledge

Learning outcomes
• 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
• Probability from a Bayesian perspective
• Multivariate linear models, including dummy-variable regression
• Analyzing interaction effects
• Model comparison
• Generalized Linear model: Binomial regression and Poisson regression
• R programming
Activities
A series of nine 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.** McElreath (2016), Ch.1
2. **Probability 1.** McElreath (2016), Ch. 2; selected exercises from Ch. 2
3. **Probability 2.** McElreath (2016), Ch. 3; selected exercises from Ch. 3
4. **Linear models.** McElreath (2016), Ch. 4; selected exercises from Ch. 4
5. **Multivariat linear models.** McElreath (2016), Ch. 5; selected exercises from Ch. 5
6. **Overfitting and underfitting.** McElreath (2016), Ch. 6; selected exercises from Ch. 6
7. **Interactions.** McElreath (2016), Chapter 7; selected exercises from Ch. 7
8. **Generalized Linear models.** McElreath (2016), Chs. 9-10; selected exercises from Ch. 10
9. **Presentations individual assignment.**

Seminar leader: Mats E. Nilsson

Note: The plan above is tentative, if there is an interest in MCMC (Ch. 8), we may spend time on that, or even schedule an extra seminar for those interested.

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

The book is available in electronic format from Stockholm University Library

Schedule
Time: 13.00 – 16.00
Place: Seminar room at Gösta Ekman Laboratory.