Department of Psychology, Stockholm University
Doctoral program in Psychology, spring semester 2020

Syllabus: Statistics 2 (7.5 hp)

This course will use two books:
(1) "Statistical Rethinking" by Richard McElreath (2016). This book is an excellent introduction to statistical modeling in general, and modeling from a Bayesian perspective in particular. The course include selected parts of McElreath's book up to, but not including the last chapters on Multilevel modeling. As a complement to McElreath, there will be seminars devoted to causal inference from the perspective of potential outcome model of causality and directed acyclical graphs.
(2) Selected chapters from the book "An Introduction to Statistical Learning" (James, et al., 2013), to give examples of statistical approaches focusing on prediction and classification rather than causal inference.

Both books and their exercises use 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 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 basic statistical learning methods for prediction and classification.
- 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 approaches advocated by McElreath (2016) or James, et al., (2013).

Course content

- Data management and reproducible data analysis
- Probability from a Bayesian perspective (McElreath, 2016)
- Causal inference (Elwert, 2013, Elwert & Winship, 2014; Textor, 2019)
- Multivariate linear models, including dummy-variable regression (McElreath, 2016; James et al., 2013)
- Interpreting interaction effects (McElreath, 2016)
- Prediction and classification (James et al., 2013)
- Resampling methods: cross-validation and bootstrapping (James et al., 2013)
- Generalized Linear model: Binomial regression and Poisson regression (McElreath, 2016)
- R programming (McElreath, 2016; James et al., 2013)

Activities

A series of 17 seminars. The seminars will start with a theoretical discussion of the topic covered by a specific book chapter or article, followed by student presentations of solutions to 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; Data management and reproducible data analysis**
2. **Probability 1.** McElreath
3. **Probability 2.** McElreath
4. **Linear models 1.** McElreath, James et al.
5. **Linear models 2.** McElreath, James et al.
6. **Linear models 3**. McElreath, James et al.
7. **Causal inference 1**. Directed Acyclical Graphs and related tools
8. **Causal inference 2.** Directed Acyclical Graphs and related tools
9. **Interactions**. McElreath
10. **Overfitting and underfitting.** McElreath, James et al.
11. **Cross-validation.** James et al., McElreath
12. **Bootstrapping**. James et al.
13. **Classification.** James et al.
14. **Markov Chain Monte Carlo.** McElreath
15. **Generalized Linear models 1.** McElreath
16. **Generalized Linear models 2**. McElreath
17. **Presentations individual assignment.**

Note: The plan above is very tentative and subject to change (it will be finalized after discussions at the first seminar).

Seminar leader: Mats E. Nilsson.

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) and James, et al. (2013) that were 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 analytic approaches discussed during the course. The student will present an outline of the planned analyzes at the last seminar, and should submit a report no later than 2 month after the last seminar. If revision is needed, the revision should be delivered no later than 3 months after the last seminar.

Literature

Books:

McElreath, R. (2016). *Statistical Rethinking. A Bayesian Course with Examples in R and Stan.* New York: CRC Press. **Chapters 1-10**. [Available in electronic format from Stockholm University Library.]

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning.* New York: Springer. **Chapters 1-6.** [Available in electronic format from Stockholm University Library.]

Articles/websites:

Elwert, F. (2013). Graphical causal models. In *Handbook of Causal Analysis for Social Research* (pp. 245–273). Berlin: Springer.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: the problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53.

Textor, J. (2019). *DAGitty*. Retrieved August 17, 2019, from dagitty.net

Schedule

Dates (it's a Monday or Thursday): *April* 2, 6, 9, 13, 16, 20, 23, 27; *May* 4, 7, 11, 14, 18, 25, 28; and *June* 1, 4.

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

Place: Seminar room at Gösta Ekman Laboratory.