

STATISTICAL METHODS IN LINGUISTIC RESEARCH

Time: September 29-30 and October 27-28 2011, 9-16.

Venue: Copenhagen Business School, room tba

Credits: 3.5 ECTS for participation, 5 ECTS if also handing in a 5 page final report at the end of the course (see below).

Organiser: Department of International Language and Computational Linguistics, Copenhagen Business School.

Registration: by e-mail to hbm.isv@cbs.dk no later than September 13. For questions relating to the content of the course, please contact the course coordinator Laura Winther Balling at lwb.isv@cbs.dk.

Purpose

The purpose of the course is to introduce the participants to statistical tools and methodologies that will enable them to conduct statistical analyses in their PhD dissertations and further linguistic research. The course focuses on the practical applications of statistics and includes a substantial amount of hands-on examples and exercises. The course is of interest to all students who are doing or plan to do quantitative research.

Prerequisites

The course builds to some extent on concepts and methods introduced in "Statistical methods in linguistic research. Module 1: Elementary methods". Participants who have not attended that course should read the course book (S. Rasinger, *Quantitative Research in Linguistics. An Introduction*, Continuum, 2008) for that module and do the exercises in the book. Further details about Module I may be found on <http://gradeast.dk/kurser/statistik2010/>. Some time will be devoted at the beginning of the course to brush up the main points of the first module.

The course does not assume any prior knowledge about mathematics or statistics. The course builds mainly on examples from linguistics, but a background in linguistics is not a prerequisite.

Learning outcome

The participants will learn how to conduct statistical analyses on real-world data sets from a range of linguistic domains, using the R statistics software package. In particular, they will learn how to utilize graphical data exploration, statistical hypothesis testing, and regression analysis with a focus on mixed-effect models that are well suited for addressing many questions in linguistics. Datasets from different

subfields will be provided, but participants who have their own data will also get the chance to work on them.

Participants who want full credit for the course must write a final report in which they use the methods from the course to address a research problem within their own research area. The final report must describe the linguistic problem the participant wants to examine, the type of conclusions the participant would like to draw, and the data collection and statistical analysis that is needed in order to support the conclusions.

Form and language

The course will consist of a mixture of lectures, discussions and practical exercises. The course will be taught in English.

Course book

R.H. Baayen, *Analyzing Linguistic Data. A Practical Introduction to Statistics using R*, Cambridge University Press, 2008.

Preparation

The participants should bring a laptop with the R statistics package installed; the R package can be downloaded for free from www.r-project.org and runs under Windows, Mac and Linux. Participants who do not have access to a laptop should contact the organisers. For the first of the two modules (September 29-30), please read chapters 1 to 4 of Baayen's book and try all the examples in the first two chapters in the R statistics package. The goal of the preparation is to make the participant familiar with R, and to allow the participant to formulate questions to the first two chapters. The remaining chapters will be discussed in the second of the two modules, more detail will be provided later.

Teachers

Laura Winther Balling (course coordinator, assistant professor of experimental psycholinguistics, CBS), Søren Feodor Nielsen and/or Peter Dalgaard (both professors of statistics, CBS).