PhD COURSE

Workshop on Statistics for linguistics

PhD course, University of Bergen, August 31 to September 2, 2015

Course instructor: Melanie Bell, Anglia Ruskin University

Registration deadline: August 15, 2015

To register for the course fill out the registration form **here**.

Contents of the course

Melanie Bell, Senior Lecturer at Anglia Ruskin University, will give a Workshop on Statistics for Linguistics August 31 to September 2, 2015. The workshop will be a follow-up to her introductory statistics course at the LingPhil summer school in Fevik and will cover more advanced statistical techniques. The workshop will include both lectures on various topics and comments on the individual PhD projects of course participants. It will be possible to participate in the workshop either with or without a presentation of your project.

Those PhD candidates who want their projects to be commented on will send in a manuscript ahead of time. This will include a brief description of the PhD project, as well as an explicit statement of the hypotheses that are going to be tested, and as much methodological detail as possible. The manuscript should be accompanied by some data, which could be made up if real data are not available yet. Each candidate who has supplied data will give a brief oral presentation of their project, and the data will be used to exemplify the statistical techniques covered in the course.

There is some scope to adapt the course content according to the projects presented, but the following areas will provisionally be included:

Testing hypotheses
Comparing groups
Correlation
Multiple Regression
Logistic regression
Mixed effects modelling

Prerequisites

All participants will be expected to have a working knowledge of the topics covered in the introductory course: types of data, patterns of distribution and descriptive statistics. If you have not taken an introductory course in statistics, you can still

participate, but in that case you should prepare by reading (and doing the exercises for) Chapters 1–4 in Chris Butler's book *Statistics in Linguistics*. This book is out of print but has been made available by the author online, see:

http://www.uwe.ac.uk/hlss/llas/statistics-in-linguistics/bkindex.shtml

The author has corrected some errors in the printed version and therefore urges people to use the online version.

All participants should also bring their own laptop to the workshop and should have the following software installed before the workshop starts:

R: http://www.r-project.org/

RStudio: http://rstudio.org/

In addition, you should ensure that you know how to get data into R and how to manipulate dataframes in R by working through Chapter 1 of the following book:

R. H. Baayen *Analyzing Linguistic Data: A Practical Introduction to Statistics using R* (http://www.cambridge.org/no/academic/subjects/languages-linguistics/grammar-and-syntax/analyzing-linguistic-data-practical-introduction-statistics-using-r)

Recommended reading:

The reading for the course comes from the book by Chris Butler, mentioned above, and Keith Johnson's *Quantitative Methods In Linguistics*: (http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1405144246.html)

Everyone should read at least Chapter 1 of Johnson before the workshop starts. Based on the provisional list of topics, the other relevant sections of the books are as follows:

Testing hypotheses
Butler: Chapters 5 and 6
Johnson: Sections 2.1 to 2.3

Comparing groups
Butler: Chapters 7 and 8
Johnson: Section 3.1

Correlation
Butler: Chapter 11
Johnson: Section 2.4

Multiple Regression
Johnson: Section 3.2

Logistic regression
Butler: Chapter 9
Johnson: Sections 5.1 to 5.4

Mixed effects modelling
Johnson: Sections 7.1 to 7.3

Credits: 3 ECTS credit points for participating with a presentation, 1 ECTS point for participating without

Venue: University of Bergen

Program: Daily classes from 10:00 to 15:00 with a one hour lunch break at 12:00

Melanie Bell: Statistics

Synopsis

The course assumes no previous experience of statistics or statistical software. It will provide a basic introduction to statistical principles and methods, with a view to building participants' confidence in reading and understanding linguistic studies that use such methods. Topics covered will include: types of data, sampling, patterns of distribution, describing data, testing hypotheses, levels of significance, comparing groups for significant differences, and identifying significant relationships between variables. All examples will use data and problems related to linguistics, and will be presented using the statistical software R. Students will be expected to undertake set reading in support of the classes.

Lecture topics

- Describing information
- Taking samples and testing hypotheses
- Comparing groups and investigating relationships

Syllabus

Lecture	**Topics**	**Reading**
1	Basic concepts Describing data	chapter 01_fundamental_concepts chapter 02_frequency_distributions chapter 03_central_tendency_and_variability chapter 04_normal_distribution
2	Sampling Testing hypotheses	chapter 05_sampling_and_estimation chapter 06_hypothesis_testing
3	Comparing groups Correlation	chapter 07_parametric_tests_of_significance chapter 11_correlation

The reading listed above, which is available on the summer school server, comes from the following textbook:

Butler, Christopher. 1985. Statistics in linguistics. Oxford: Blackwell.
http://www.uwe.ac.uk/hlss/llas/statistics-in-linguistics/bkindex.shtml

This book assumes no prior knowledge of statistics and presents the necessary mathematical concepts without assuming knowledge of any computer software. As far as possible, please read the chapters indicated before each session.

Additional material and a more extensive bibliography will be provided in class.