INFORMATION ON STATISTICS FOR LINGUISTICS

Empirical research in contemporary Linguistics often involves quantitative analysis. For example, we might want to compare the frequency of occurrence of a word or phrase in two or more different corpora (e.g. of British and American English), how often men and women use particular discourse markers, or we might want to evaluate data collected from questionnaires. Unfortunately, simply counting raw numbers often doesn't tell us very much and that is why in recent years Statistics have become indispensable for the interpretation of quantitative results. Measures such as normalising, representing raw frequencies as percentages, significance tests and correspondence analysis help us identify meaningful patterns in our data that would be very difficult to spot with the naked eye.

Various general courses and workshops on statistical software and methods are offered on a regular basis at Leuphana. These are usually advertised on *myStudy*. We are also pleased to announce the workshop "Statistics for Linguistics with *R*" which will take place on 19-21 February 2013. For further information, please visit
http://www.leuphana.de/forschung/services/alma/aktuell/termine/ansicht/datum/2013/02/19/alma-workshop-multivariate-statistics-in-linguistics.html).

In the following, we list some useful statistical web tools and further reading intended specifically for linguists. For help or advice please refer to Lina-Lab (linalab@leuphana.de) or the Methodenzentrum (Gebäude 4).

Useful literature:

- Baayen, R. Harald (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge, New York: Cambridge University Press.
- Gries, Stefan Thomas (2009): *Statistics for Linguistics with R. A Practical Introduction*. Berlin: Mouton De Gruyter.
- Larson-Hall, Jenifer (2010). *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York: Routledge.
- McEnery, T, Xiao, R. & Tono, Y. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. Oxon and New York: Routledge. 52-58.
- Oakes, Michael P. (1998): *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Tagliamonte, Sali A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Useful web calculators:

- A chi-square calculator with a detailed illustration of the procedures involved in the test: http://www-user.uni-bremen.de/~anatol/qnt/qnt_dist.html
- A chi-square calculator with a slightly different layout: http://graphpad.com/quickcalcs/chisquared1.cfm
- A log-likelihood score calculator and some further reading: http://ucrel.lancs.ac.uk/llwizard.html
- Critical p-values for the chi-squared test and the log-likelihood test: http://www.unc.edu/~farkouh/usefull/chi.html