Statistics for Linguists

- an Introduction with Practical Exercises.

Stefan Evert

Tutorial at DGKL 6

Monday, 29 September 2014, 11:00 – 18:00

Bismarckstr. 12, 91054 Erlangen, room 0.320 (ground floor)

This tutorial gives a basic introduction to statistical hypothesis testing in linguistics, with a particular focus on corpus frequency data. Participants will be familiarized with fundamental notions such as the null hypothesis, p-values and confidence intervals.

Several essential hypothesis test will be introduced and participants will learn how to apply these procedures using the free, open-source statistical software package R (www.r-project.org).

This tutorial is based on the SIGIL course by Marco Baroni & Stefan Evert. Handouts, exercises and additional course units can be found on the course homepage at http://SIGIL.r-forge.r-project.org/.

Provisional schedule:

11:00 – 12:30   Corpus frequency data & statistical inference

                                         lunch break

13:30 – 14:30   First steps in R

14:30 – 15:30   Methods for continuous data

                                         coffee break

16:00 – 17:00   Practical exercises

17:00 – 18:00   The problem of non-randomness