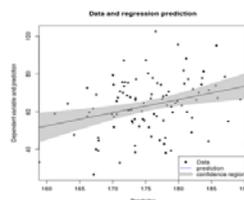


Statistics for Linguists

Type of course: MA.AA.SW05
 Time: Friday 10–12 a.m. (10:15–11:45)
 Venue: Ernst-Abbe-Platz 8, SR 216 (PC room)
 Workload: 5 ECTS



Course description

In the last two decades, linguistics has undergone a major ‘empirical turn’, to the effect that the collection and analysis of quantitative data has become increasingly important for linguistic research projects. This development goes hand in hand with a need for the sound statistical analysis of linguistic data, whether they are frequency data from corpora, questionnaire data from second language acquisition or experimental data from psycholinguistics or sociolinguistics. Accordingly, the major goal of this course is to provide a comprehensive introduction to typical data structures in linguistics and their statistical analysis. Based on a wide variety of authentic data sets, we will acquaint ourselves with basic principles of quantitative analysis (e.g. variables, coding, descriptive statistics), the general logic of inferential statistical testing and model building and, finally, with specific statistical models for commonly encountered data in linguistics (e.g. correlation and regression analysis (including ANOVA, linear and logistic regression), measures of association, exploratory methods like cluster analysis and MDS). We will be using the freely available and very powerful software *R* (see www.r-project.org) and the interface *RStudio* (see www.rstudio.com); all texts, materials and data sets will be provided in class.

Requirements and marking system

- regular attendance, active participation and thorough weekly preparation (see below)
- final exam** (90 minutes, 15 February 2019, 10–12 a.m.)

Weekly preparation

An essential part (and requirement) of the course is your in-depth preparation of each session at home. This usually comes in the form of **preparatory texts** and accompanying **homework**. The course will be based on the major chapters in:

Levshina, Natalia (2015). *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam, Philadelphia: John Benjamins.

Further useful references will be provided for each topic of the course.

The course management system

The course management system of our department, called **WORDWISE** (www.wordwise.uni-jena.de), will be the online platform of this class. All class materials (including PPT presentations, in-class worksheets, follow-up handouts, references, etc.) will be made available there. In addition, important organisational information (e.g. announcements etc.) will also be distributed via WORDWISE (to your email address). Therefore, please make sure your email address on WORDWISE is valid and that you check it regularly!

Course programme

DATE	TOPIC AND OBLIGATORY LITERATURE	THEMATIC AND METHODOLOGICAL ISSUES
	Foundations of statistics for linguists	
1 19.10.18	<i>Introduction to the course: Aims and structure of the course. Organisation. Logic and structure of an empirical study.</i>	Understanding the ‘quantitative turn’ in linguistics and what it entails. Practicalities (textbook, software, interface, website, etc.).
2 26.10.18	<i>Introduction to statistical thinking: Variables, hypotheses and the general logic of inferential statistical testing.</i> Levshina 2015: Ch. 0+1 (pp. 1–19)	Spreadsheet software, types of variables, status of variables in hypothesis testing. Samples and populations, descriptive and inferential statistics. Statistical models, test statistics and <i>p</i> -values.
3 02.11.18	<i>Introduction to R and RStudio.</i> Levshina 2015: Ch.2	Issues of data collection and storage. Data structures in <i>R</i> . Accessing and manipulating information in data frames.
4 09.11.18	<i>Descriptive statistics for quantitative variables.</i> Levshina 2015: Ch.3	Measures of the central tendency and dispersion. Visualising and exploring distributions.
5 16.11.18	<i>Exploring qualitative variables.</i> Levshina 2015: Ch.4	Visualising proportions. Operations on factors in data frames.
	Inferential statistical methods	
6 23.11.18	<i>Comparing (the means or medians of) two groups.</i> Levshina 2015: Ch.5	<i>t</i> -tests and their non-parametric equivalents. Dependent and independent samples. Confidence intervals and standard error.
7 30.11.18	<i>Relationships between two quantitative variables: Correlation and basics of linear regression modelling.</i> Levshina 2015: Ch.6	Parametric and non-parametric correlation coefficients and their associated significance tests. General goals and logic of regression.
8 07.12.18	<i>Simple linear regression.</i> No reading, but a few exercises.	Regression lines and residuals. Assessing the model and individual predictors. Effects plots.
9 14.12.18	<i>Multiple linear regression.</i> Levshina 2015: Ch.7	Expanding regression into multiple linear regression. Principles of statistical model building.
10 21.12.18	<i>Multiple linear regression continued. Analytical practice and revision questions.</i> No reading, but revision.	More on model building and model selection.
11 11.01.19	<i>Measuring associations between two qualitative (categorical) variables.</i> Levshina 2015: Ch.9	Analysis of contingency tables: Chi-squared test, Fisher exact test, associated effect sizes and visualization tools.
12 18.01.19	<i>Binomial logistic regression</i>	Fitting, interpreting and validating complex statistical models.
13 25.01.19	Levshina 2015: Ch. 12	
	Extension and outlook	
14 01.02.19	<i>Exploratory methods for large data sets.</i> Szmrecsanyi and Kortmann 2009	Basic principles of cluster analysis, multidimensional scaling and principal component analysis.
15 08.02.19	<i>Practice. Summary and wrap-up. Course evaluation.</i> Gries 2012 (Statistical modelling)	Bring any remaining questions!
15.02.19	Final exam (10–12 a.m., EAP8, SR 216)	
20.03.19	Resit exam (10–12 a.m., EAP8, SR 216)	