

## STATISTICS FOR LINGUISTIC APPLICATIONS

Informal syllabus by Maciej Karpiński, Institute of Applied Linguistics AMU  
(academic year 2019/20)

1. In our classes we will learn how to collect linguistic data, produce and share datasets, understand and take measurements, and analyse data properties.
2. It is highly recommended to bring your computer to the class. You will be instructed on which applications you should install.
3. The data for the classes will be made available online, brought to the class by the lecturer, or prepared by the participants of the course.
4. Paper and pens still may be useful in many cases :-)
5. After first three meetings, you should be able to formulate the topic of your project. It may evolve later on, but sticking to the initial version will have some advantages. The project must involve linguistic data collection and processing. The topic must be accepted by the lecturer.
6. The results of the projects should be shared with the group and the lecturer. The last (two) meeting(s) will be confessed to the presentation of your results.

### Grading rules:

During the classes, participants will design and develop their project based on the linguistic data of their interest. The project will be presented to the entire group and the lecturer will estimate it for its formal and conceptual soundness.

### Content of the classes:

1. **Statistics:** What is it for? Why for linguists? What can we do with it? What are the tools that we can use for doing statistics?
2. **Data in linguistics. Basic notions:** Specimen (example, unit, item), sample (set, collection, group), population. Qualitative vs. quantitative studies.
3. **Collecting linguistic data:** Random, representative, sufficient? Major data sources: experiments and corpora. Scrutinizing data sets. Data quality. Expected and unexpected values. Range and variability.
4. **Counting and measuring. Variables.** Examples from corpus linguistics, phonetics, psycholinguistics, etc. Tables and visual data representation.
5. **Frequency, frequency distributions,** what they say about the variables, and why they are important.
6. **General properties of data collections.** Descriptive statistics of samples. Central tendencies. Dispersion. Visual representations of basic descriptive statistics.
7. **Estimating properties of populations.** Inferential statistics. Samples vs. populations. Comparing mean values: Basic parametric and non-parametric tests and their applications.
8. **Relations** between and among variables, and how to detect them.
9. Statistics-based **language modelling** and its applications.

### Reading \*

Eddington, D. 2015. *Statistics for Linguists: A Step-by-Step Guide for Novices*. Cambridge Scholar Publishers.

Woods, A., Fletcher, P., & Hughes, A. (1986). *Statistics in language studies*. Cambridge University Press.

Rietveld, T., & Van Hout, R. (2010). *Statistics in language research: Analysis of variance*. Walter de Gruyter.

\*) More texts may be provided by the lecturer during the course or on demand. Any introductory coursebook, including “for the Dummies” series, may prove useful.