

## Lent Term

Contact: Mr. Sam Oliver, Graduate Administrative Assistant (RFB 102). Email Address: [gradtal@mml.cam.ac.uk](mailto:gradtal@mml.cam.ac.uk)

Time/Venue: Tuesdays in Weeks 3 – 6 Lent Term, 4.00 pm – 5.00 pm, CALL Facility

Course convenor: Dr Barbara McGillivray

Overview: This module aims to further consolidate and deepen the statistics knowledge that students gained in basic courses at SSRMC. It will do so by first introducing basic hypothesis testing techniques (i.e. chi-square test and t-test), effect size measures and confidence intervals (Weeks 1-2) and then covering linear regression models (Weeks 3-4). The course will focus on the conceptual and intuitive understanding of statistical methods rather than the mathematical details underlying them. For the analysis, we will use the R programming language through R Studio, although the ideas covered in the course generally apply when analyzing data with other software packages (e.g., SPSS) as well. Each session will include a lecture followed by hands-on practical work. The course presupposes the knowledge of descriptive statistics (e.g., mean, standard deviation) and basic inferential statistical techniques, as well as the basic idea of null-hypothesis significance testing. Familiarity with R is desirable but not essential.

## Course outline

Lent Term:

- |  |             |
|--|-------------|
| 1. Hypothesis testing (chi-square test and t-test) | 5 February  |
| 2. Effect sizes and confidence intervals           | 12 February |
| 3. Regression with continuous variables            | 19 February |
| 4. Regression with categorical variables           | 26 February |

## Computer packages

R and R Studio are recommended for the analysis and visualization of statistical data, although basic calculation and visualization can also be performed with Excel.

## Textbooks

There are a large number of books on statistical data analysis that are accessible to linguists and other social scientists. Three particularly useful books are listed below. All of them are held by the MML Library [Field et al.: classmark Z1.D.13; Levshina: classmark Z1.D.21; Baayen: classmark Baayen: L14.B.3].

- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage Publications.
- Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press. [An older version of this book with some typos and small errors is available from the author's website: <http://www.sfs.uni-tuebingen.de/~hbaayen/publications.html>]