

STATISTICS FOR LANGUAGE STUDIES

Calendari	Dates: Tuesdays, from April 14 th to May 12 th Time: 3pm-5pm
Aula	52.213
Durada	10 hours (in-person classes)
Llengua d'impartició	English
Format	Asynchronous (previous individual work) + Synchronous (in-person classes)
Destinataris	Master students

Ponents

- Gemma Repiso Puigdelliura – gemma.repiso@uab.cat

Competències

- Methodological competence (CM)
- Digital competence (CDD)
- Others: Analytical competence

Objectius

- This course will introduce both programming in R and statistical methods in linguistics. No prior knowledge is assumed, as this course constitutes a starting point for attendees so that they can continue to use R for both manipulating and dealing with data as well as fitting regression models to analyze data.
- In addition to providing practical experience with R during the classes, the last part of every session will be devoted to providing attendees with recommended materials for further study, ensuring that those who want to learn more after the course is completed will be able to do so.
- At the end of this course, you will be able to:
 1. Perform quantitative data analysis using the R environment
 2. Fit the most common statistical models used in linguistics
 3. Interpret and communicate results of statistical analyses
 4. Visualize results and data of various natures
 5. Know where to find high-quality materials for further study

Continguts

Week	Content
1	<ul style="list-style-type: none">• Getting started with R<ul style="list-style-type: none">○ Why use R?○ R Basics○ Simple data manipulation○ Types of variables○ Organizing your data in an R-friendly way (tidyverse)• Descriptive statistics<ul style="list-style-type: none">○ Measures of central tendency and dispersion○ Empirical distributions• Basic data visualization (basic R and ggplot2)
2	<ul style="list-style-type: none">• An introduction to inferential statistics• Effect size and power• Linear regression (simple and multiple regression)<ul style="list-style-type: none">○ Checking assumptions○ Contrast coding○ Interpreting regression output through model predictions
3	<ul style="list-style-type: none">• Categorical data analysis and logistic regression<ul style="list-style-type: none">○ Odds and odds ratios○ Simple and multiple logistic regression○ Goodness of fit
4	<ul style="list-style-type: none">• Mixed models for continuous data: Linear regression<ul style="list-style-type: none">○ Fixed vs. random effects○ Random intercepts and slopes○ Model comparison and reporting
5	<ul style="list-style-type: none">• Mixed models for categorical data: Logistic regression<ul style="list-style-type: none">○ Fixed vs. random effects○ Random intercepts and slopes○ Model comparison and reporting

Metodologia

- This course does not require any previous knowledge of R or RStudio. The necessary basic knowledge will be presented in a gradual manner, and recommended resources will be provided for future study.
- This course is intended to provide students with a conceptual and practical introduction so that they can begin to analyze linguistic data. The lectures will provide both theoretical explanations of statistical and programming content, as well as having time for questions and practice. There will be a small amount of work that attendees must cover independently before each class so that we can use our time together more effectively, which will be posted to the Aula Global.

Recommended reading: Sonderegger, M. (2023). *Regression modeling for linguistic data*. MIT Press.

Avaluació

- The course will not have a formal evaluation.

Característiques de l'aula

- Moodle
- Students need access to a computer and the internet