Universitat Pompeu Fabra *Barcelona*

CLIK
Centre per a la Innovació
en Aprenentatge i Coneixement

CLIK Formació

STATISTICS FOR LANGUAGE STUDIES

Calendari	Dates: May 8th, 15th, 22nd, 29th Time: 15h-17h (local Barcelona time)
Aula	TBD
Durada	8 hours (live in-person sessions) + 8 hours of preparatory individual work
Llengua d'impartició	English
Format	Asynchronous (previous individual work) + Synchronous (live in-person sessions)
Número màxim d'assistents	25
Destinataris	Graduate students

Ponents

- Scott James Perry (University of Alberta) – sperry1@ualberta.ca
- Adriana Soto-Corominas (Universitat Autònoma de Barcelona) – Adriana.Soto@uab.cat

Competències

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

Objectius

- To analyze your data, you do not need previous knowledge of programming or mathematics. We will give you the tools and knowledge to analyze your data, whether you have already collected it or not. It is expected that many students will have taken Tècniques Experimentals I d'Observació with Sílvia Perpiñán and Scott Perry in the fall of 2023. Therefore, the first week will allow for those who are not familiar with statistical analysis or R to catch up and for those who are familiar with those topics to refresh their knowledge before we build on that knowledge with more advanced content. Videos and other materials that cover the previous content will be provided before the course begins.
- At the end of this course, you will be able to:
  1. Perform quantitative data analysis using the R environment
  2. Justify the use of one statistical analysis over another
  3. Interpret and communicate the results of statistical analyses
  4. Visualize results and data of various natures

Universitat Pompeu Fabra Barcelona

CLIK
Centre per a la Innovació
en Aprenentatge i Coneixement

CLIK Formació

Continguts

Week	Content
1	This week is intended to review the content we assume people will be familiar with before moving to more advanced topics. It will be heavily self-guided, with each student watching videos to fill the gaps in their knowledge. • Basic functionality of R • Descriptive statistics and visualizations • Null hypothesis significance testing • Linear regression – theory and practice
2	This week provides an in-depth look at three common generalized linear models in linguistics. We discuss how non-Gaussian distributions allow us to model non-continuous data with the help of link functions. There is an emphasis on describing the components of these more complicated models and understanding how they let us model various types of data. • Logistic regression o The binomial distribution and logit link function o Dealing with separation in logistic regression o Different interpretations (log-odds, odds ratios, etc.) • Poisson regression o The Poisson distribution and log link function o Dealing with variability in count data (overdispersion) o Dealing with too many zeros (zero inflation) • Ordinal regression o Cumulative logit model o Interpreting thresholds in ordinal regression o Latent variable interpretation of coefficients in ordinal regression
3	This week, we unpack mixed-effects models and understand what problems they solve for us. We describe how random effects are estimated and what differentiates them from fixed effects. We also look at generalized mixed models and explore how differences can arise in interpreting conditional vs. marginal effects when there is a non-linear link function. • Mixed effects models o No pooling, complete pooling, and partial pooling o Random intercepts and slopes o Interpreting conditional effects o Marginal effects – how are they different?
4	In this last week, we look at a different kind of model that can be used for modelling non-linear trends in data: the generalized additive model. We focus on understanding how the non-linear trend is estimated, how to visualize these trends, and how to account for repeated measures in the data. • Generalized additive models o How the models fit a curve o Visualizing non-linear effects o Adding random effects to generalized additive models

Metodologia
- This course does not require any previous knowledge of R or RStudio. However, we expect that most students will have previously taken a course where they were exposed to R and some of the content in this course. For a student who does not know R or statistics, we expect that preparation for the first week may take longer than the two hours described in this document.
- This course follows the flipped classroom methodology: attendees are expected to attend the synchronous classes after viewing the lecture videos, reading the provided materials, and completing the assigned activities.
- Both instructors will be available to all students to resolve any doubts and support the course's length. To receive individual attention, attendees:
  - Must complete asynchronous content (videos + activities)
  - Must attend the in-person sessions

Avaluació
- The course will not have a formal evaluation, and student success will be assessed based on their ability to analyze data and report findings at the end of the course

Característiques de l'aula

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