Regression for Linguists

WiSe23/24

Daniela Palleschi

Humboldt-Universität zu Berlin

2023-10-10

Welcome!

- the language of instruction is English
- Block course:
  - October 10, 11, 12 (10am-4pm)
  - January 12th (4-8pm)
  - January 26th (4-8pm)
  - February 9th (4-8pm)

Course description

- via hands-on exercises in RStudio with the R programming language, we will develop skills and know-how related to

  - analysing linguistic data with *linear models*

  - the *theory* behind linear models

  - implement *(generalised) linear (mixed) models*

  - learn how to **communicate** our findings

Course materials

- Moodle: Regression for linguists
  - Enrolment Key: reg4ling
- Course website (work-in-progress)
  - https://daniela-palleschi.github.io/reg4ling
- communal terms and concepts Google Sheet
  - https://docs.google.com/spreadsheets/d/17CqdxKL9lyy-PbTB2ZnfWNWs4oV–CcBvrqlh_aEPGQ/edit?usp=sharing
  - please keep this open during class, and write down any words/concepts that you think are important
  - you don't need to write the definition, this is a collaborative document
  - by keeping the 'Lecture topic' column accurate I can easily update the website with our terms and definitions

4

Course credits

This course is part of Modul 9 (Forschungsmodul) and is worth 4 LP. A break down of the course credits:

LP	type	task
1LP	preparation	readings
1LP	participation	lecture materials/exercises
1LP	preparation	homework due December 20th
1LP	preparation	final homework due March 29th

- Office hours: Wednesdays, 15.00-16.00 (by appointment)

Reading and preparation

- this course mainly follows Winter (2019)
- by now, you should be familiar with the concepts in Chaters 1-3
    - Intro to R
    - the tidyverse and reproducible workflow
    - Descriptive statistics, models, and distributions

Syllabus

Further readings

- there are lots of useful resources out there, specifically:
  - Bodo Winter's tutorials on linear (mixed) models (Winter, 2013, 2014)
  - Sonderegger (2023): a recently published book, I'd say more intermediate than Winter (2019)
  - the PsyTeachR website is a *great* resource for hands-on stats and/or data analysis in R from the University of Glasgow School of Psychology and Neuroscience

A word on reproducibility

- I have no expectations in this course for whether you use .R scripts, Rmarkdown, or Quarto (all of which available under `File > New File`)

- If you're using Rmarkdown or Quarto (recommended), I suggest you have the following code at the end of every script:

```
1  sessionInfo()
```

- this will print your session info at the end of each document to print your current packages and R version (etc.)

- In this course you'll see some examples of tools for a reproducible workflow

  - but I won't explicitly discuss them (unless it's requested)

Session Info

```
R version 4.3.0 (2023-04-21)
Platform: aarch64-apple-darwin20 (64-bit)
Running under: macOS Ventura 13.2.1

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;  LAPACK version 3.11.0

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: Europe/Berlin
tzcode source: internal

attached base packages:
```

References

Sonderegger, M. (2023). *Regression Modeling for Linguistic Data*.

Winter, B. (2013). *Linear models and linear mixed effects models in R: Tutorial 1*. https://bodowinter.com/tutorial/bw_LME_tutorial1.pdf

Winter, B. (2014). *A very basic tutorial for performing linear mixed effects analyses (Tutorial 2)*.
https://bodowinter.com/tutorial/bw_LME_tutorial2.pdf

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Routledge. https://doi.org/10.4324/9781315165547