Regression for Linguists

WiSe23/24

Daniela Palleschi

2023-10-10

Table of contents

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)

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

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

Course credits

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

- 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:

```
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;  L

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:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
 [1] jsonlite_1.8.7    dplyr_1.1.3       compiler_4.3.0    tidyselect_1.2.0
 [5] webshot_0.5.4     xml2_1.3.4        stringr_1.5.0     systemfonts_1.0.4
 [9] scales_1.2.1      yaml_2.3.7        fastmap_1.1.1     readr_2.1.4
[13] R6_2.5.1          generics_0.1.3    curl_5.0.1        knitr_1.43
[17] tibble_3.2.1      kableExtra_1.3.4  munsell_0.5.0     svglite_2.1.1
[21] pillar_1.9.0      tzdb_0.4.0        rlang_1.1.1       utf8_1.2.3
[25] rbbt_0.0.0.9000   stringi_1.7.12    xfun_0.39         fs_1.6.2
[29] viridisLite_0.4.2 cli_3.6.1         magrittr_2.0.3    digest_0.6.33
[33] rvest_1.0.3       rstudioapi_0.14   hms_1.1.3         lifecycle_1.0.3
[37] vctrs_0.6.3       evaluate_0.21     glue_1.6.2        fansi_1.0.4
[41] colorspace_2.1-0  rmarkdown_2.22    httr_1.4.6        tools_4.3.0
[45] pkgconfig_2.0.3   htmltools_0.5.5
```

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