Shrinkage and partial pooling

Mixed Models 3

Daniela Palleschi
Humboldt-Universität zu Berlin

2024-01-12

Learning Objectives

Today we will learn…

- about no/complete/partial pooling
- about shrinkage

Resources

- this lecture covers
  - Blog post "Plotting partial pooling in mixed-effects models" from Tristin Mahr (2017)
  - Section 15.9 'Shrinkage and Individual Differences' in Winter (2019)
  - Box 8.2 'Broader Context: Shrinkage and Partial Pooling' in Sonderegger (2023)
- we will be using the data from Biondo et al. (2022)

Set-up

```r
# suppress scientific notation
options(scipen=999)
```

Load packages

```r
# load libraries
pacman::p_load(
               tidyverse,
               janitor,
               here,
               lmerTest)
```

```r
lmer <- lmerTest::lmer
```

Load data

- data from Biondo et al. (2022)

```
df_biondo <-
  read_csv(here("data", "Biondo.Soilemezidi.Mancini_dataset_ET.csv"),
           locale = locale(encoding = "Latin1") ## for special characters in Spanish
           ) |>
  clean_names() |>
  mutate(gramm = ifelse(gramm == "0", "ungramm", "gramm")) |>
  mutate_if(is.character,as_factor) |> # all character variables as factors
  droplevels() |>
  filter(adv_type == "Deic")
```

Set contrasts

```
contrasts(df_biondo$verb_t) <- c(-0.5,+0.5)
contrasts(df_biondo$gramm) <- c(-0.5,+0.5)
contrasts(df_biondo$adv_type) <- c(-0.5,+0.5)
```

```
contrasts(df_biondo$verb_t)
```

```
       [,1]
Past   -0.5
Future  0.5
```

```
contrasts(df_biondo$gramm)
```

```
        [,1]
gramm   -0.5
ungramm  0.5
```

```
contrasts(df_biondo$adv_type)
```

```
         [,1]
Deic     -0.5
Non-deic  0.5
```

Run models

- random-intercepts only

```r
fit_fp_1 <-
  lmer(log(fp) ~ verb_t*gramm +
         (1 |sj) +
         (1|item),
       data = df_biondo,
       subset = roi == 4)
```

- by-item varying tense slopes

```r
fit_fp_item <-
  lmerTest::lmer(log(fp) ~ verb_t*gramm +
         (1 |sj) +
         (1 + verb_t|item),
       data = df_biondo,
       subset = roi == 4)
```

Pooling

- do the random effects represent the exact average of participants?
  - below we see the mean logged first-pass reading time per participant (`mean`) and the by-participant intercepts from `fit_fp_1` and `fit_fp_item`
- to understand what's happening, we first have to understand pooling

▶ Code

```
# A tibble: 6 × 5
  sj     mean population_mean intercept_1 intercept_item
  <chr> <dbl>           <dbl>       <dbl>          <dbl>
1 1      6.42            5.96        6.40           6.40
2 2      5.79            5.96        5.79           5.80
3 07     5.87            5.96        5.87           5.87
4 09     5.78            5.96        5.78           5.78
5 10     6.67            5.96        6.62           6.62
6 11     5.91            5.96        5.91           5.92
```

No pooling

- no pooling refers to separate regression lines fit e.g., per participant
  - each regression line is fit ignoring the population-level information
  - the intercepts are the true mean from each participant

```r
head(df_no_pooling)
```

```
       model sj intercept    verb_t1      gramm1 verb_t1:gramm1
1 No pooling  1  6.422811 0.16094962  0.07844247     0.12950513
2 No pooling  2  5.792669 0.10115512 -0.10571656    -0.23199316
3 No pooling 07  5.870556 0.15344172 -0.25264603    -0.29866189
4 No pooling 09  5.780839 0.16938275  0.14074977    -0.07324559
5 No pooling 10  6.664530 0.04786447 -0.13824470     0.21824110
6 No pooling 11  5.912309 0.07573670 -0.06469794     0.35318406
```

```r
sum_shrinkage |> head(6)
```

```
# A tibble: 6 × 5
  sj     mean population_mean intercept_1 intercept_item
  <chr> <dbl>           <dbl>       <dbl>          <dbl>
1 1      6.42            5.96        6.40           6.40
2 2      5.79            5.96        5.79           5.80
3 07     5.87            5.96        5.87           5.87
4 09     5.78            5.96        5.78           5.78
5 10     6.67            5.96        6.62           6.62
6 11     5.91            5.96        5.91           5.92
```

Complete pooling

- complete pooling refers to ignoring grouping factors
  - i.e., fixed-effects only models (e.g., with `lm()` or `glm()`)
  - one regression line fit ignoring the individual-level information
  - the intercepts are the same as the population-level mean

```r
head(df_pooled)
```

```
# A tibble: 6 × 6
  model            sj    intercept verb_t1  gramm1 `verb_t1:gramm1`
  <chr>            <fct>     <dbl>   <dbl>   <dbl>            <dbl>
1 Complete pooling 1          5.96  0.0612 0.00310          -0.0152
2 Complete pooling 2          5.96  0.0612 0.00310          -0.0152
3 Complete pooling 07         5.96  0.0612 0.00310          -0.0152
4 Complete pooling 09         5.96  0.0612 0.00310          -0.0152
5 Complete pooling 10         5.96  0.0612 0.00310          -0.0152
6 Complete pooling 11         5.96  0.0612 0.00310          -0.0152
```

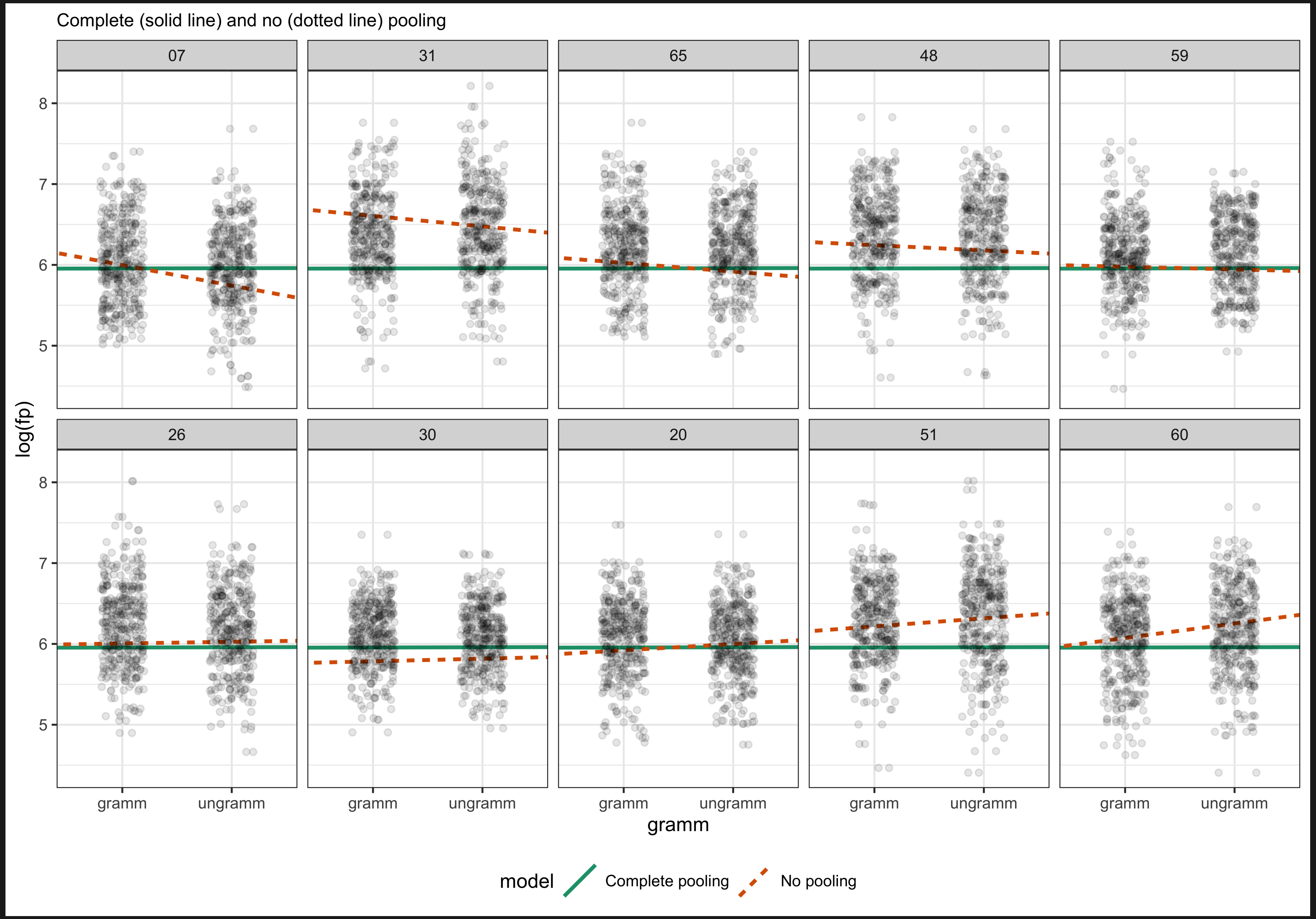
```r
sum_shrinkage |> head(6)
```

```
# A tibble: 6 × 5
  sj     mean population_mean intercept_1 intercept_item
  <chr> <dbl>           <dbl>       <dbl>          <dbl>
1 1      6.42            5.96        6.40           6.40
2 2      5.79            5.96        5.79           5.80
3 07     5.87            5.96        5.87           5.87
4 09     5.78            5.96        5.78           5.78
5 10     6.67            5.96        6.62           6.62
6 11     5.91            5.96        5.91           5.92
```

Complete vs. no pooling

- complete pooling (green solid line) and no pooling (orange dotted line) of grammaticality effects for 10 participants
  - describe what you see in terms of intercept and slopes across the participants

Figure 1: Observations (black dots) with complete pooling regression line (solid green) and no pooling line (dotted orange) per 10 participants



Partial pooling: mixed models

Shrinkage

- turns out the estimates are pulled towards the population-level estimates
  - all the information in the model is taken into account when fitting varying intercepts and slopes

Figure 2: Elaine Benes learns about shrinkage of random effect estimates towards the population-level estimates

Shrinkage

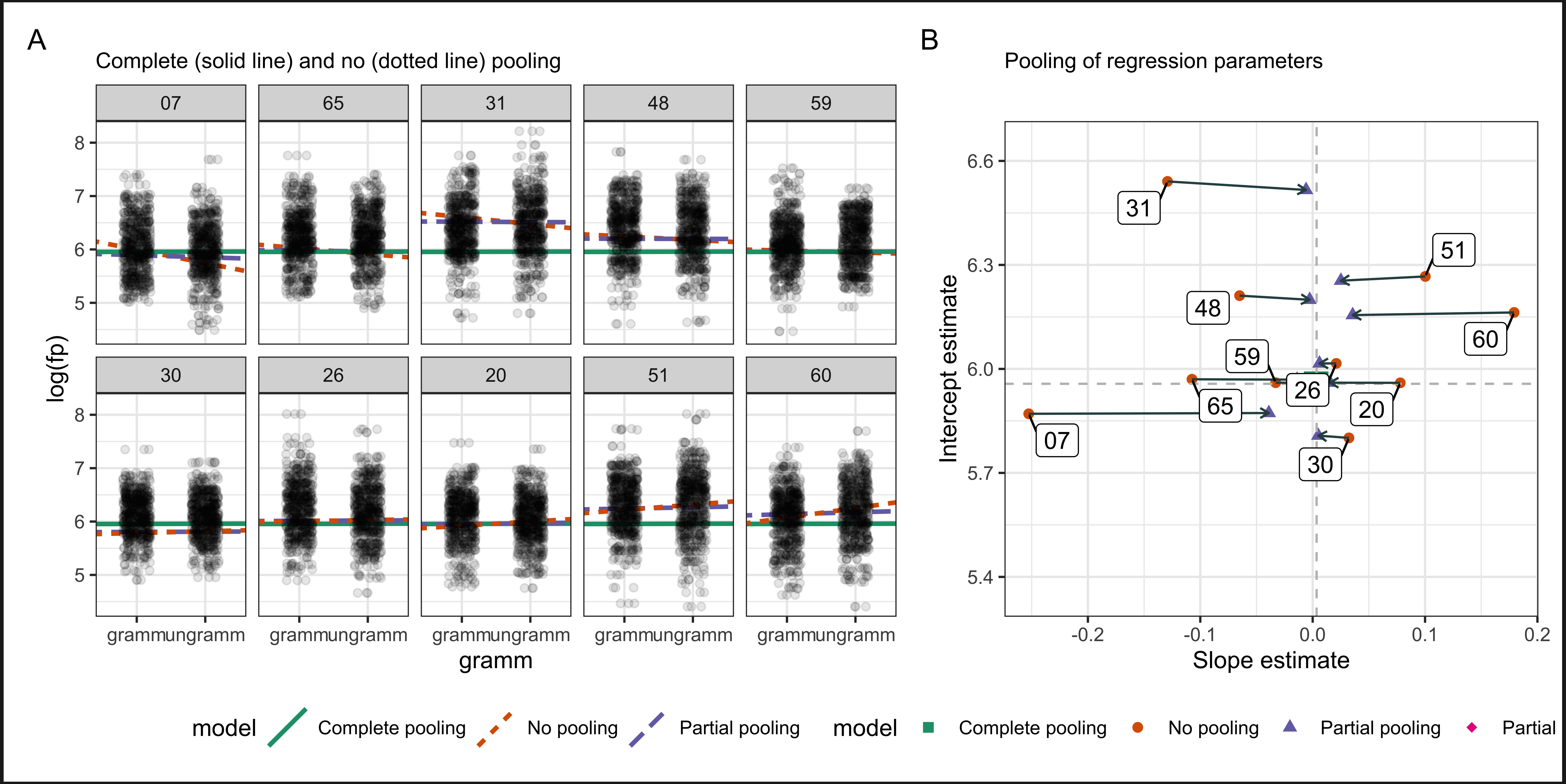


Figure 3: Shrinkage of 10 participants

Centre of gravity

- why are some points not being pulled directly to the 'centre of gravity'?
  - they're being pulled to a higher confidence region

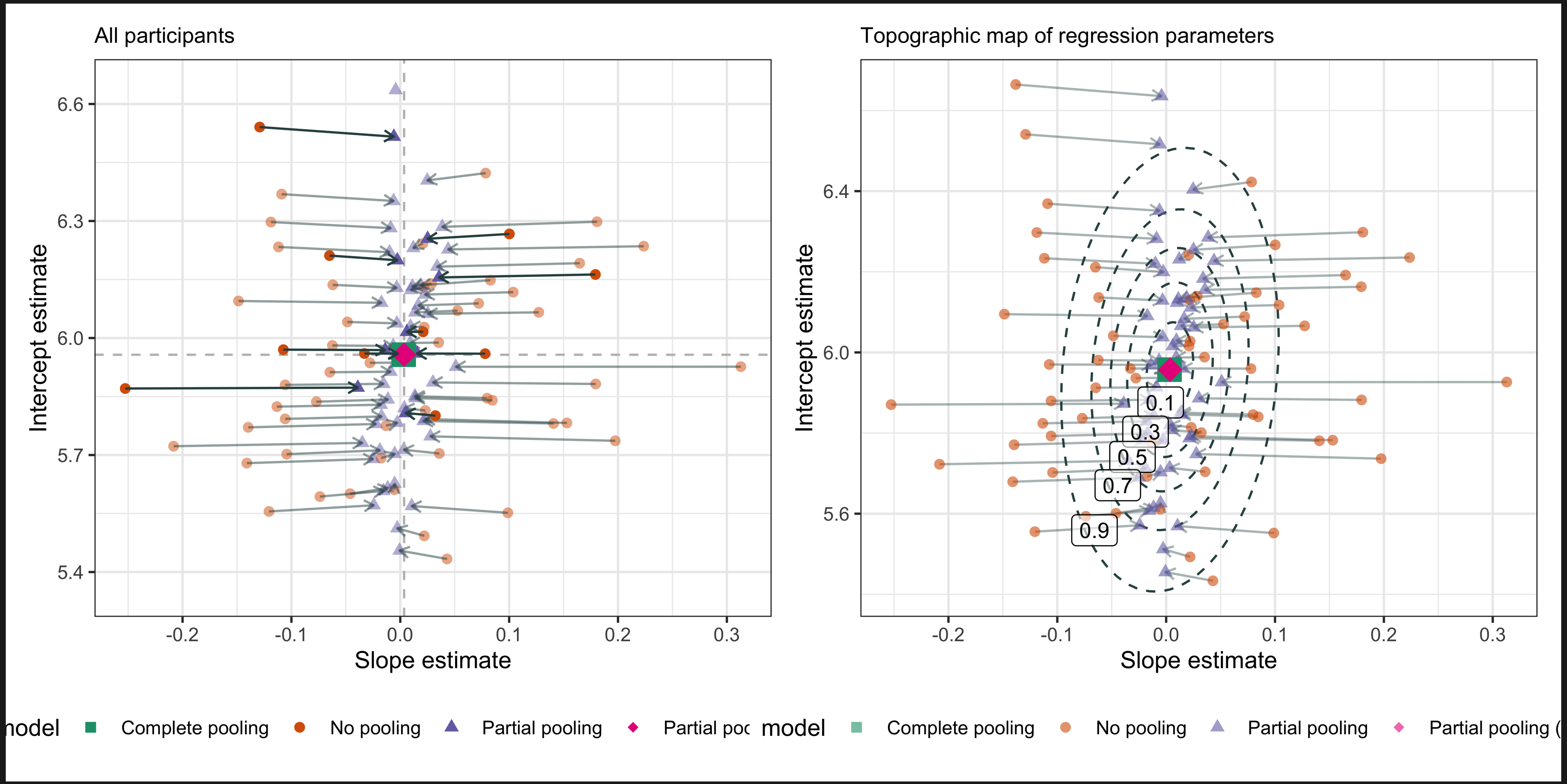


Figure 4: Shrinkage for all participants: each ellipsis represents a confidence level (really, a quantile: q1, q3, q5, q7, and q9); The inner ellipsis contains the centre 10% of the data, the outer ellipsis 90%

Why shrinkage?

- with partial pooling, each random effect is liek a weighted average
  - it takes into account the effect for one group level (e.g., one participant) *and* the population-level estiamtes
  - the empirical effect for a group level is weighted by the number of observations
  - so if one participant has fewer observations than another, then more weight is given to the population-level estimates, and vice versa
- the implications (benefits) of this:
  - imbalanced data are not a problem for linear mixed models
  - the model can make predictions for unseen levels, i.e., it can generalise to new data

Learning objectives 🏁

Today we learned…

- what linear mixed models are ✅
- how to fit a random-intercepts model ✅
- how to inspect and interpret a mixed effects model ✅

Important terms

Term	Definition	Equation/Code
linear mixed (effects) model	NA	NA

References

Biondo, N., Soilemezidi, M., & Mancini, S. (2022). Yesterday is history, tomorrow is a mystery: An eye-tracking investigation of the processing of past and future time reference during sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(7), 1001–1018. https://doi.org/10.1037/xlm0001053

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

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