Multilevel/hierarchical models: Overview

Silje Synnøve Lyder Hermansen

2026-02-19

Where are we in the course?

Recap from Monday

When observations are not i.i.d. (i.e. they share a group identity), we will often consider alternatives to the ordinary linear model

- negative take: the assumptions of the linear model are not met.
  - non-normal residuals,
  - heteroscedastic residuals
  - correlation between x and residuals
- positive take: we have variation that we want to leverage strategically
  - within-group variation
  - between-group variation
  - more correct estimation of the standard errors

$\Rightarrow$ *see this as an opportunity*

I pick my models as part of my research design

What are the most relevant correlations/variation given my theory?

- in experiments: you can create that variation and randomize the rest (cut out confounders)
- in observational studies: you'll have to "hunt" for the variation you want and control away the rest

Confounders

- Control variables that – if absent lead to omitted variable bias – satisfy three criteria:
  - z correlates with y
  - z correlates with x
  - z causes x and y (not intermediate/post-treatment)

  $\rightarrow$ *even when 3 is not satisfied, it might be a sign of a common group identity (e.g. nationality)*
- Group identities: observations done in the same context share many potential confounders
  - you might kill several birds with one stone

The principle

The principle

We make the assumption that the residuals are drawn from a normal distribution

- **pooled models:** a single distribution

$$y_i = a + bx_i + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

- **hierarchical models:** add a hierarchy
  - assume groups are drawn from different distributions
  - their mean is drawn from a single distribution that "rules them all"

$$y_i = a + bx_i + \epsilon_{ji}$$
$$\epsilon_j \sim N(\alpha_j, \sigma_j^2)$$
$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

Untangling the parameters/variation

This allows me to untangle different sources of variation

$$y_i = a + bx_i + \epsilon_{ji}$$

$$\epsilon_j \sim N(\alpha_j, \sigma_j^2)$$

$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

- α_j: grouped mean of residuals: group intercept
- σ_α^2: between-group variation
- σ_j^2: group-level (within) variation

The promises of a hierarchical structure

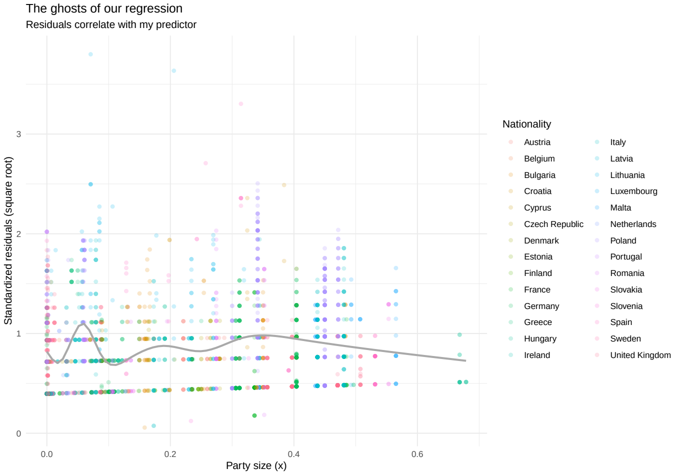
This allows me to leverage different sources of variation

- leverage **within-group** variation:
  - by factoring out/control for between-group variation (σ_j^2)
- leverage **between-group** variation:
  - by running a second regression on the group means (α_α^2)
    - adjusts the standard errors
    - data augmentation: add variables from other sources that vary by group
    - predict out of sample even for new groups
- leverage **both sources** of variation
  - by borrowing from the more informative variation ("pooling"/"shrinkage")

Labeling the errors: grouped residuals

Labeling the errors: grouped residuals

Our residuals have group identities that we can “label” as such.

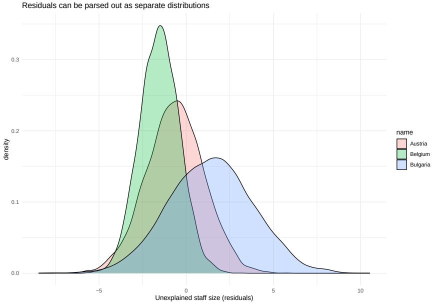


Group means and group-level variation

Our residuals have group identities that we can "label" as such.

- each group of residuals has a distribution with a mean and a spread

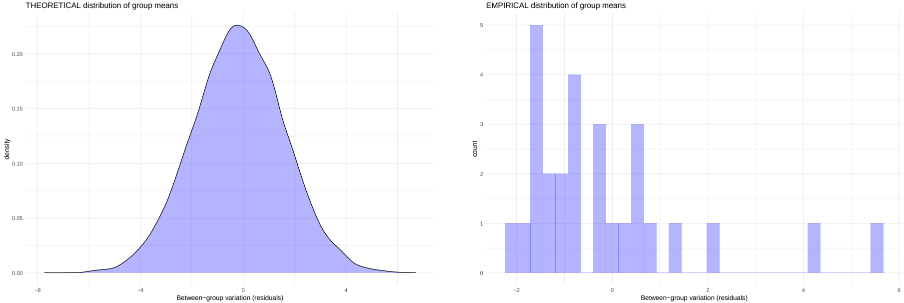
```
## # A tibble: 28 x 3
##    Nationality    y_bar_j sigma2_alpha
##    <chr>            <dbl>        <dbl>
##  1 Austria         -0.665        1.65
##  2 Belgium         -1.54         1.15
##  3 Bulgaria         1.41         2.44
##  4 Croatia          0.549        4.10
##  5 Cyprus          -0.253        1.89
##  6 Czech Republic  -0.206        1.84
##  7 Denmark         -1.48         1.30
##  8 Estonia         -1.33         0.950
##  9 Finland         -1.47         0.919
## 10 France          -1.11         1.26
## # i 18 more rows
```



⇒ *I can reconstruct their theoretical distribution by calculating the group mean and standard deviation*

Between-group variation

The group means are drawn from a common normal distribution with a mean and a spread



$\Rightarrow$ *I am treating the residuals as if they were a variable, so statistical theory can be applied*

Varying-intercepts regression: within-group variation

Varying-intercepts regression: within-group variation

The random/varying-intercept model:

- a common slope for all predictors
- separate intercepts for all group identities
- a common intercept (grand mean)

From labelled errors to varying intercepts

Instead of hiding the groupings in the residuals, we can report them as a series of intercepts (i.e. report their group means)

$$y_i = a + bx_i + \alpha_j$$
$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

- a: the **grand mean** (mean of α means)
- α_j: **varying intercepts** (deviations from this grand mean)

$\Rightarrow$ *useful for interpretation in R*

Varying-intercepts

Now, it is clear that I parse out (control for) between-group variation

- **within-group variation** the b coefficients report the effect of observation-level variables
- **group-level variation** is reported in the varying intercepts, it is the variation that:
  - has not been accounted for by my main effects
  - that can be attributed to group identities

Estimation in R: Varying national intercepts

Estimation in R: Varying national intercepts

Let's regress MEPs' investment in their district (y) on...

- x: their party's size in the national parliament (as a proxy for state funding).
- ... while controlling away between-national variation

Equation:

$Staff\ size = a + b \times Party\ size + \alpha_{Nationality}$

$y_i = a + bx_i + \alpha_{ij}$

Estimation:

```
library(lme4)
mod.ran.int <- lmer(y ~ x + (1|Nationality),
                    df)
```

Reading the R output

Reading the R output

```r
summary(mod.ran.int)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ x + (1 | Nationality)
##    Data: df
##
## REML criterion at convergence: 31355.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1127 -0.5387 -0.1435  0.3598 15.2357
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Nationality (Intercept) 3.125    1.768
##  Residual                5.240    2.289
## Number of obs: 6948, groups:  Nationality, 28
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.6799     0.3386   7.915
## x            -1.6722     0.1678  -9.965
##
## Correlation of Fixed Effects:
##   (Intr)
## x -0.117
```

R refers to the residuals as “random effects”

σ^2_α: remaining **between-group variance**: 3.12

- standard deviation: 1.77
- the unexplained variation between groups

`Residual`: remaining **within-group** variance: 5.24

- standard deviation of within-group distribution: 2.29
- the unexplained variation within all groups

R refers to regression coefficients as “fixed effects”

a: intercept/**grand mean**: 2.68

- a hypothetical intercept for interpretation (mean of means)

b: **slope**: -1.67

- the marginal effect of party size (x)

Interpretation

Interpretation

Interpretation follows normal principles, but there are some complications:

a. we now have two intercepts per scenario:

- the grand mean (a): for focus on general effect of x
- the group-level mean (α_j): for description and prediction
- sum of the grand mean (a) and group-level mean (α_j): for prediction

b. all effects are linear

- so first-difference and marginal effects are the same

Interpreting marginal effects

The interpretation of the marginal effect is as with any linear model:

Table 1: Effect of state funding for parties on MEPs' local staff size

	Dependent variable:
	y
x	-1.672^{***}
	(0.168)
Constant	2.680^{***}
	(0.339)
Observations	6,948
Log Likelihood	−15,677.610
Akaike Inf. Crit.	31,363.210
Bayesian Inf. Crit.	31,390.600
Note:	*p$<$0.1; **p$<$0.05; ***p$<$0.01

$\Rightarrow$ *A 10% decrease in the national party's seat share would lead every 6th MEP to compensate by hiring an additional local staffer.*

Prediction

The varying intercepts are reported as deviations from the grand mean

```
fixef(mod.ran.int); ranef(mod.ran.int)
```

```
## (Intercept)           x
##    2.679887   -1.672226
```

```
##                (Intercept)
## Austria        -0.49518857
## Belgium        -1.52249566
## Bulgaria        1.54657524
## Croatia         0.68267309
## Cyprus         -0.05313986
## Czech Republic -0.10587832
```

Predicted local staff in Austria when national party is not in Parliament:

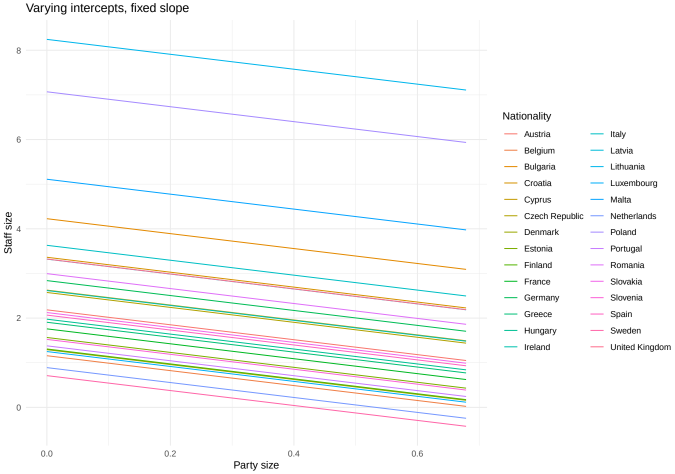
- $2.68 + \text{-}0.5 \times 0 = 2.18$

Predicted local staff in Austria when national party holds 10% of the seats

- $2.68 + \text{-}0.5 + \text{-}1.67 \times 0.1 = 2.02$

Visualization

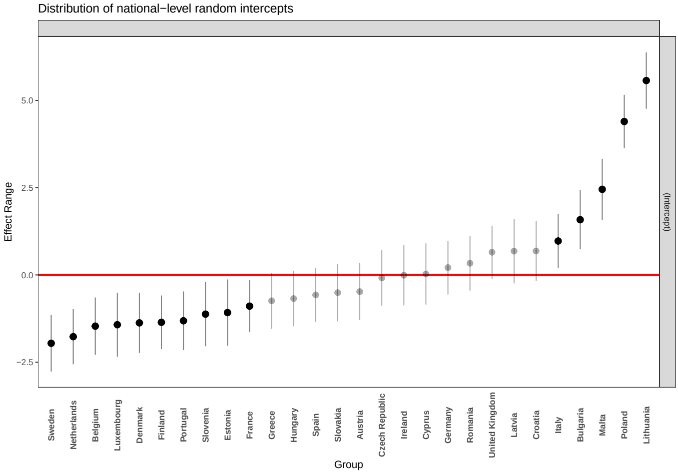
Effect of x, the slope coefficient



$\Rightarrow$ *the slope is constant, but the intercept changes across nationalities*

Visualization: as distributions

The intercepts are distributions in their own right



⇒ *each varying intercept has a point estimate (regression coefficient) and a distribution. They vary around a normal distribution with mean of 0*

Varying slopes, varying intercepts

Defintion

We can let the effect of z vary by group

$$y_i = a + b_1 x_i + c_j z_i + \alpha_j$$

- c_j: **varying slope** (the effect of z varies by group)
- α_j: **varying intercepts**
- we can rewrite to make this explicit

$$y_i = a + bx_i + \epsilon_{ij}$$

$$\epsilon_j \sim N(\alpha_j, \sigma_\alpha)$$

$$\alpha_j = \lambda_j + c_j z_j$$

- λ_j: **varying intercepts**

$\Rightarrow$ *a series of regressions within the regression*

Estimation in R

the estimation is done as if it was an interaction effect

- fixed-effects model with cross-level interaction

```
mod.ran.slope <- lm(y ~ x + ProxNatElection * Nationality, df)
```

- random-effects model with varying slope

```
mod.ran.slope <- lmer(y ~ x + (ProxNatElection | Nationality),
```

Interpretation

Marginal effects

We can read these coefficients as if they were from separate models

```
ranef(mod.ran.slope)
```

```
##                (Intercept) ProxNatElection
## Austria         -0.3201686     0.001073577
## Belgium         -1.3944093    -0.018299157
## Bulgaria         1.8027887     0.086786758
## Croatia          0.9352286     0.058233060
## Cyprus           0.1020429    -0.003477477
## Czech Republic   0.1112017     0.024050888
```

MEPs from Austria hire on average 0.004 (= 0.001 * 4) assistants more immediately before an election compared to immediately after, while MEPs from Belgium hire on average 0.073 (= 0.018 * 4) fewer assistants.

- These are negligible marginal effects.

Prediction

The prediction is done per group, but follows normal rules

- two intercepts:
  grand mean + group-level intercept
- one slope per group

```
fixef(mod.ran.slope); ranef(mod.ran.slope)
```

```
## (Intercept)           x
##    2.513035   -1.691321

##                (Intercept) ProxNatElection
## Austria         -0.3201686     0.001073577
## Belgium         -1.3944093    -0.018299157
## Bulgaria         1.8027887     0.086786758
## Croatia          0.9352286     0.058233060
## Cyprus           0.1020429    -0.003477477
## Czech Republic   0.1112017     0.024050888
```

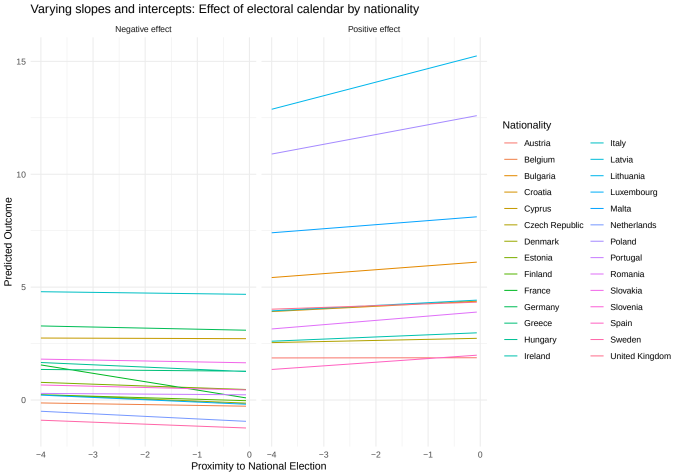
Austria after election:

- $2.51 + -0.32 + 0.001 \times -4 = 2.189$

Austria before election:

- $2.51 + -0.32 + 0.001 \times 0 = 2.193$

Visualization



Level-2 regression: between-group variation

Definition

Definition

We can think of the residuals/group intercepts as a variable in their own right

$$y_i = bx_i + \epsilon_{ji}$$

- they are generated by draws from J number of distributions:

$$\epsilon_{ji} \sim N(\alpha_j, \sigma_\alpha^2)$$

- ... and therefore we can model them

$$\alpha_j = a + dz_j$$

- a: a **single intercept**
- d: a **single slope coefficient**

⇒ *we run a second regression on the residuals*

Implications

We explicitly model between-group variation

- z, the level-2 predictor only varies at the group level
  - standard errors for z reflect the number of groups
  - the more groups, the more the approach makes sense
- data augmentation
  - we can add information from other to the model
  - contextual elements
  - improves prediction

Estimation in R: Electoral system

Estimation in R: Electoral system

Let's add electoral system (z) as a predictor

- it never changes in a country (in this study)

R handles this automatically

- same data frame
  - all variables that don't vary within groups are regressed as a level 2
- coefficients reported the same way
- estimation of coefficients and standard errors is different

```
mod.two.levels <- lmer(y ~ x + z + (1|Nationality), df)
```

Reading the R output

Reading the R output

The R output looks exactly the same as for the varying-intercept model.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ x + z + (1 | Nationality)
##    Data: df
##
## REML criterion at convergence: 31353.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1145 -0.5388 -0.1434  0.3599 15.2339
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Nationality (Intercept) 3.235    1.799
##  Residual                5.240    2.289
## Number of obs: 6948, groups:  Nationality, 28
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   2.5268     0.6030   4.191
## x            -1.6719     0.1678  -9.962
## z             0.2263     0.7311   0.310
##
## Correlation of Fixed Effects:
##   (Intr) x
## x -0.077
## z -0.821  0.013
```

The level-2 regression coefficient appears as "fixed effects"

a: **grand mean:** 2.53

- the "mean of means"

d: **slope:** 0.23

- the marginal effect of electoral system (z)

Check the **change in between-group variance**:

- the between-group variance (σ^2_α, 3.23) should normally decrease
- it is not the case here ($3.12 \leq 3.23$)

→ *increase in variance indicates "complexities" between levels (interactions)*

`Correlation of Fixed Effects:`

- negative correlation between predictor (z) and intercept (-0.82): high level of z correlates with low base-line value of y.

Pooling

Pooling

What is the difference between a fixed-effects and a random-effects model, then?

- the fixed-effects model only compares within groups

```
mod.fix <- lm(y ~ a + Nationality, df)
```

- the random-effects (hierarchical) model borrows information between and within groups → *pools*

```
mod.fix <- lmer(y ~ a + (1|Nationality), df)
```

⇒ *both are varying-intercepts models*

What is pooling?

The hierarchical model calculates a weighted average of between- and within-group variation for each coefficient

$$\frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

- the denominator is there to normalize $(\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}) \rightarrow$ *ignore it*
- $\bar{y_{all}}$: the pooled mean
  - its weight $(\frac{1}{\sigma_\alpha^2})$
    - σ_α^2: between-group variation
- $\bar{y}_j$: the group mean
  - its weight $(\frac{n_j}{\sigma_y^2})$
    - n_j: size of the group (number of observations)
    - σ_y^2: residual variation not explained by the between-group variation

The weights in pooling

The hierarchical model calculates a weighted average of between- and within-group variation for each coefficient

$$\frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

- σ_α^2: as the **between-group variation** increases, the weight of the pooled mean decreases
- n_j: as **the size of the group** (number of observations) increases, the weight of the non-pooled (within-) group mean increases

What to do?

Sooo... what do I choose?

Condition	Fixed	Random	Advantage	Limitation
plenty of within-group variation	x		stringent comparison	no weighing of groups
		x	weighing by group size	groups should be distinct (between-group variation is high)
variables only vary by group		x	standard errors are corrected	fixed effects will be non-identified
mix of between- and within-group variation		x	pooling/borrows information	no idea where the info comes from
data augmentation/prediction		x	infers from group-level predictors	fixed effects don't perform out of sample

How many groups and how many observations?

Random/hierarchical model

- if you want level 2 variables:
  - **many groups** → *you run a second regression*
- if you want within-group variation:
  - **distinct groups** (large between-group variation, size matters less) → *similar to fixed-effects*
  - not distinct groups (little between-group variation) → *similar to pooled model*
- if you think the smaller groups are less representative
  - larger groups count more for within-group variation → *unbalanced panels*

Fixed-effects model

- only the observations with variation within the groups count towards the estimate → *your N may be deceptive*

Recap

Recap

Hierarchical models leverage variation according to the structure in the data (groupings)

- varying-intercepts models (fixed and random effects)
  - *one slope*, but control for group identities
- varying-intercept, varying slope (fixed and random effects)
  - one intercept and one slope *per group*,
- level-2 regression (random effects)
  - one slope per group predictor, but adjusts *standard errors*,
- pooling (all random effects models)
  - regression coefficients are a weighted average of between- and within-group variation

$\Rightarrow$ *Pick the variation you want, then pick the model you need.*