

Hierarchical modelling



Nicole Cruz

SMLP 2025

Last update: August 27, 2025

Contents

Introduction

Why multilevel?

Notation

Example with `brms`

Introduction

Method with several names

- ▶ multilevel models
- ▶ hierarchical models
- ▶ mixed effects models

Multilevel

- ▶ Hierarchical / multilevel models because they allow data analysis at several levels of aggregation.
- ▶ E.g. level 1: single data points of a student, level 2: student, level 3: school, level 4: city, level 5: country...
- ▶ The levels are numbered from bottom to top, and from specific to general.

Multilevel

- ▶ Hierarchical / multilevel models because they allow data analysis at several levels of aggregation.
- ▶ E.g. level 1: single data points of a student, level 2: student, level 3: school, level 4: city, level 5: country...
- ▶ The levels are numbered from bottom to top, and from specific to general.

Multilevel

- ▶ Hierarchical / multilevel models because they allow data analysis at several levels of aggregation.
- ▶ E.g. level 1: single data points of a student, level 2: student, level 3: school, level 4: city, level 5: country...
- ▶ The levels are numbered from bottom to top, and from specific to general.

Mixed

- ▶ Mixed models, because they allow the analysis not only of means but also of variances.
- ▶ Parameters for means: "fixed effects. parameters for variances: "random effects".

Mixed

- ▶ Mixed models, because they allow the analysis not only of means but also of variances.
- ▶ Parameters for means: "fixed effects. parameters for variances: "random effects".

Why multilevel?

Multilevel data structures

- ▶ Generalization of the simple linear regression and ANOVA to complexer data with nested groups
- ▶ When hierarchical data is analysed only at one level, this can lead to under- or overestimation of effects, which undermines the interpretability of the findings.
 - Ecological and atomistic fallacy, Simpson's paradox

Multilevel data structures

- ▶ Generalization of the simple linear regression and ANOVA to complexer data with nested groups
- ▶ When hierarchical data is analysed only at one level, this can lead to under- or overestimation of effects, which undermines the interpretability of the findings.
 - Ecological and atomistic fallacy, Simpson's paradox

Multilevel data structures

- ▶ Generalization of the simple linear regression and ANOVA to complexer data with nested groups
- ▶ When hierarchical data is analysed only at one level, this can lead to under- or overestimation of effects, which undermines the interpretability of the findings.
- Ecological and atomistic fallacy, Simpson's paradox

Ecological fallacy

An ecological fallacy is a fallacy in the interpretation of statistical data. It occurs when a data pattern that was observed at the group level (e.g. the group mean) is generalised to the level of individual data points. This is a fallacy because the pattern at one level of aggregation does not necessarily coincide with the pattern at another level of aggregation.

Ecological fallacy in which the direction of an effect switches between the group and the individual levels: [Simpson's Paradoxon](#).

(Eid et al., 2010; Robinson, 1950).

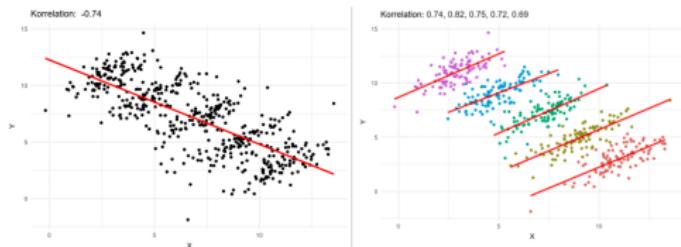
Ecological fallacy

An ecological fallacy is a fallacy in the interpretation of statistical data. It occurs when a data pattern that was observed at the group level (e.g. the group mean) is generalised to the level of individual data points. This is a fallacy because the pattern at one level of aggregation does not necessarily coincide with the pattern at another level of aggregation.

Ecological fallacy in which the direction of an effect switches between the group and the individual levels: [Simpson's Paradoxon](#).

([Eid et al., 2010](#); [Robinson, 1950](#)).

Ecological fallacy: example



A research team studies the relation between autonomy at the work place and work satisfaction. It conducts interviews in 5 companies, correlates autonomy with satisfaction and finds an overall negative correlation between the two. But the same correlation was positive in each individual company. They conclude that autonomy correlates positively with satisfaction in general, but that other variables such as income also influence satisfaction scores.

Heteroscedasticity

- ▶ Hierarchical data violate the assumption of homoscedasticity (i.e. variance homogeneity) from linear regression.
- ▶ Variance within groups tends to be smaller than variance between groups. (e.g. learn environment of students in the same class is more similar than of students in different classes).
- ▶ In multilevel analyses this is no problem. On the contrary: differences between variances can themselves become the subject of new research hypotheses.

Heteroscedasticity

- ▶ Hierarchical data violate the assumption of homoscedasticity (i.e. variance homogeneity) from linear regression.
- ▶ Variance within groups tends to be smaller than variance between groups. (e.g. learn environment of students in the same class is more similar than of students in different classes).
- ▶ In multilevel analyses this is no problem. On the contrary: differences between variances can themselves become the subject of new research hypotheses.

Heteroscedasticity

- ▶ Hierarchical data violate the assumption of homoscedasticity (i.e. variance homogeneity) from linear regression.
- ▶ Variance within groups tends to be smaller than variance between groups. (e.g. learn environment of students in the same class is more similar than of students in different classes).
- ▶ In multilevel analyses this is no problem. On the contrary: differences between variances can themselves become the subject of new research hypotheses.

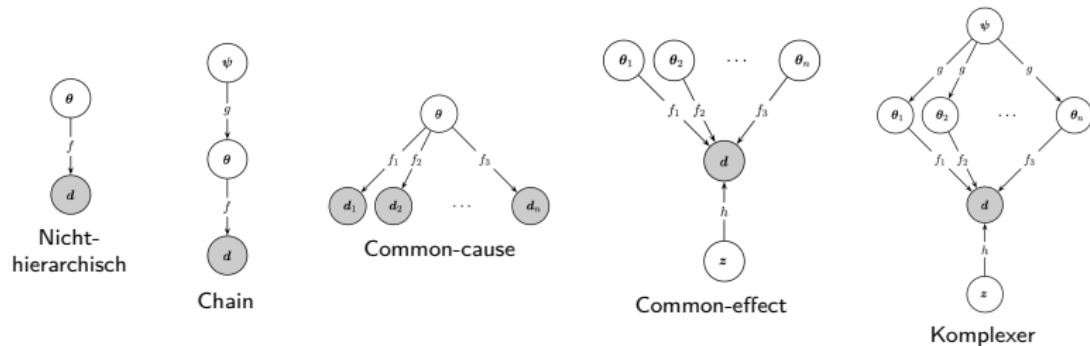
Advantages for theory development

- ▶ Development of more differentiated, abstract theories
- ▶ Models in which one latent variable can be linked to several observations, and one observation to several latent variables (e.g. common-cause, common-effect networks).

(Lee, 2011).

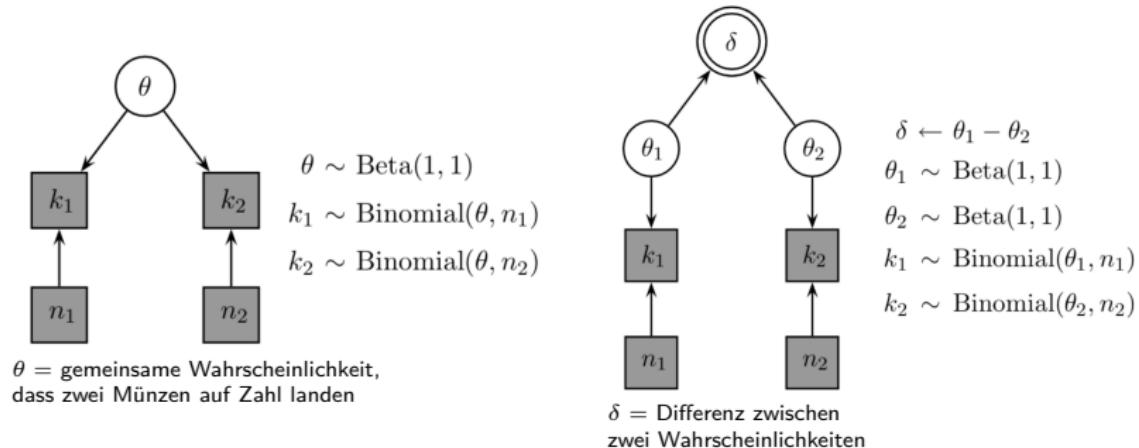
Notation

Notation for causal models



(Lee, 2011).

General graphical notation



- **squared:** discrete vs. **round:** continuous
- **shaded:** observed vs. **unshaded:** latent
- **double border:** variable value is fully determined by the values of the variables that link to it.

(Lee & Wagenmakers, 2013).

Mathematical notation

$$\begin{aligned}\mu_{ij} &= \beta_{0i} + x_j \beta_{1i} \\ Y_{ijk} &\sim \text{Normal}(\mu_{ij}, \sigma^2)\end{aligned}$$

- The data Y_{ijk} are estimated by model parameters.
- Model parameters at one level are estimated by model parameters at the next highest level.
- At the highest level prior distributions are defined.

(Haaf, 2025; [Veenman et al., 2023](#)).

Mathematical notation

$$\begin{aligned}\sigma_0^2, \sigma_1^2 &\sim \text{Normal}_+(0, .001), \sigma^2 \sim \text{Normal}_+(0, .001) \\ \beta_0 &\sim \text{Normal}(1, 1), \beta_1 \sim \text{Normal}(0, 0.1^2) \\ \beta_{0i} &\sim \text{Normal}(\beta_0, \sigma_0^2), \beta_{1i} \sim \text{Normal}(\beta_1, \sigma_1^2) \\ \mu_{ij} &= \beta_{0i} + x_j \beta_{1i} \\ Y_{ijk} &\sim \text{Normal}(\mu_{ij}, \sigma^2)\end{aligned}$$

- The data Y_{ijk} are estimated by model parameters.
- Model parameters at one level are estimated by model parameters at the next highest level.
- At the highest level prior distributions are defined.

(Haaf, 2025; Veenman et al., 2023).

Notation with brms

1. `m1 <- brm(y ~ x1 + x2, data = dat)`
2. `m1 <- brm(y ~ x1 + x2 + (1 | id), data = dat)`
3. `m1 <- brm(y ~ x1 + x2 + (1 | classroom / id),
data = dat)`
4. `m1 <- brm(y ~ x1 + x2 + (1 | classroom) +
(1 | classroom:id), data = dat)`
5. `m1 <- brm(y ~ x1 + x2 + (1 | id) +
(1 | materials), data = dat)`
6. `m1 <- brm(y ~ x1 + x2 + (x1 || id),
data = dat)`

(Bürkner, 2017, 2018).

Notation mit brms (2)

1. One level
2. Two levels, grouping level: id
3. Three levels, trials are nested in individuals (id) and individuals are nested in classrooms
4. Equivalent to previous example
5. Two crossed (instead of nested) levels: trials are nested in individuals and in experimental materials
6. Double stroke "||": Variances, but no covariances are estimated.

(Bürkner, 2017, 2018).

Model components in brms

- ▶ Dataset
- ▶ Model specification
 - Model structure
 - Priors
- ▶ MCMC Parameter
 - Number of iterations
 - Number of chains
 - Number of warmup / burnin samples
 - Save parameters?
 - Parallel Processes? (number of cores)

Which sample size?

- ▶ Data analysed with multilevel models do not have a single sample size, but separate sample sizes at each level of aggregation (e.g. study with 20 trials per participant, 35 participants per school, 10 schools per country).
- ▶ These different sample sizes make it more difficult to determine degrees of freedom for statistical tests, and there is no single agreed on method for doing so.
- ▶ This is less of a concern in the Bayesian framework, where posterior distributions can be estimated consistently regardless of the sample sizes involved.

Which sample size?

- ▶ Data analysed with multilevel models do not have a single sample size, but separate sample sizes at each level of aggregation (e.g. study with 20 trials per participant, 35 participants per school, 10 schools per country).
- ▶ These different sample sizes make it more difficult to determine degrees of freedom for statistical tests, and there is no single agreed on method for doing so.
- ▶ This is less of a concern in the Bayesian framework, where posterior distributions can be estimated consistently regardless of the sample sizes involved.

Which sample size?

- ▶ Data analysed with multilevel models do not have a single sample size, but separate sample sizes at each level of aggregation (e.g. study with 20 trials per participant, 35 participants per school, 10 schools per country).
- ▶ These different sample sizes make it more difficult to determine degrees of freedom for statistical tests, and there is no single agreed on method for doing so.
- ▶ This is less of a concern in the Bayesian framework, where posterior distributions can be estimated consistently regardless of the sample sizes involved.

How many levels?

Different criteria have been proposed for when to model an additional level of aggregation:

- ▶ When there are 5 or more data points per cell in the design
- ▶ When there are several measurements per individual (within design).
- ▶ When individuals are in groups (e.g. classrooms in schools in districts; doctors in hospitals; plants in fields)
- ▶ When we want to generalise beyond the specific individuals / experimental conditions measured.
- ▶ When we want to estimate not only means but also variances.
- ▶ When the variance at the group level makes up a certain percentage (e.g. 10% of the total sample variance (ICC)).

(Gelman, 2005).

Interclass correlation

- The intraclass correlation (**ICC**) is the proportion of level 2 variance out of the total variance.
- For trials nested in individuals: proportion of the total variance attributable to interindividual differences.

$$ICC = \frac{\sigma_{level2}^2}{\sigma_{level1}^2 + \sigma_{level2}^2}$$

Rule of thumb: open an aggregation level when $ICC \geq 10\%$.

(Tabachnick & Fidell, 2007).

Interclass correlation

- ▶ The intraclass correlation (**ICC**) is the proportion of level 2 variance out of the total variance.
- ▶ For trials nested in individuals: proportion of the total variance attributable to interindividual differences.

$$ICC = \frac{\sigma_{level2}^2}{\sigma_{level1}^2 + \sigma_{level2}^2}$$

Rule of thumb: open an aggregation level when $ICC \geq 10\%$.

(Tabachnick & Fidell, 2007).

Interclass correlation

- The intraclass correlation (**ICC**) is the proportion of level 2 variance out of the total variance.
- For trials nested in individuals: proportion of the total variance attributable to interindividual differences.

$$ICC = \frac{\sigma_{level2}^2}{\sigma_{level1}^2 + \sigma_{level2}^2}$$

Rule of thumb: open an aggregation level when $ICC \geq 10\%$.

(Tabachnick & Fidell, 2007).

Model selection

- ▶ Forward selection: from simple to complex - e.g. population level (fixed) effects, then group level (random) intercepts, then group level (random) slopes
- ▶ Backward selection: from complex to simple - e.g. remove covariances between intercepts and slopes, then random slopes, then random intercepts
- ▶ Theory based

Example with `brms`

Penguin research



```
1 > head(penguins)
2
3   species     island bill_len bill_dep flipper_len body_mass
4 1  Adelie  Torgersen     39.1      18.7        181     3750
5 2  Adelie  Torgersen     39.5      17.4        186     3800
6 3  Adelie  Torgersen     40.3      18.0        195     3250
```

Can a penguins' flipper length be predicted from their bill length?
To what extent does this vary between species, and/or between islands?

Models to compare

```
1 m1 <- brm( flipper_len ~ bill_len + (1 | species)
2   , data = penguins
3   , family = gaussian()
4   , sample_prior = TRUE
5   , warmup = 1000
6   , iter = 10000
7   , chains = 4
8   , cores = 4
9   , save_pars = save_pars(all = TRUE) # required to compute
  BFs
```

```
1 m2 <- brm( flipper_len ~ bill_len + (1 | island)
```

```
1 m3 <- brm( flipper_len ~ bill_len + bill_dep + (1 | species)
  + (1 | island)
```