# Introduction to mixed-effects models

Last modified: 2026-01-20

# Outline

1. Introduction mixed-effects models

2. Longitudinal data analysis

# 1 Introduction mixed-effects models

# Mixed-effects models

- Mixed-effects models are a class of statistical models that include fixed effects as well as random effects
- Fixed effects vs. random effects[1]
  - For fixed effects, only effects of the factor levels used in the present study are considered (manipulated conditions, e. g., assigned groups, but also sex, or other variables . . . )

    → Of interest is how these levels differ
  - For random effects, the factor levels considered in a study are regarded as a (random) sample from some population (e. g., words, raters, subjects, . . . )

    → Of interest are conlusions about the underlying population and its variation

[1] Some critical discussion on these definitions: http://andrewgelman.com/2005/01/25/why_i_dont_use/

# Linear mixed-effects model

- The linear mixed-effects model has the general form

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\upsilon}_i + \varepsilon_i$$

  with fixed effects $\boldsymbol{\beta}$, random effects $\boldsymbol{\upsilon}_i$, and the design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ and the assumptions

$$\boldsymbol{\upsilon}_i \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_\upsilon), \qquad \varepsilon_i \overset{iid}{\sim} N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$$

- This implies for the marginal covariance matrix

$$Cov(\mathbf{y}_i) = \boldsymbol{\Sigma}_i = \mathbf{Z}_i \boldsymbol{\Sigma}_\upsilon \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{n_i}$$

# Linear mixed-effects model

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{Np} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} z_{10} & z_{11} & \dots & z_{1q} & \dots \\ z_{20} & z_{21} & \dots & z_{2q} & \dots \\ z_{30} & z_{31} & \dots & z_{3q} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{N0} & z_{N1} & \dots & z_{Nq} & \dots \end{pmatrix} \cdot \begin{pmatrix} \upsilon_{10} \\ \vdots \\ \upsilon_{1q} \\ \upsilon_{20} \\ \vdots \\ \upsilon_{Nq} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

## 2 Longitudinal data analysis

# Longitudinal data

- Consist of repeated measurements on the same subject taken over time
- Are a frequent use case for mixed-effects models
- Contain time as a predictor: time trends within and between subjects are of interest

```
library("lme4")
data("sleepstudy")
?sleepstudy
str(sleepstudy)
summary(sleepstudy)
head(sleepstudy)
```

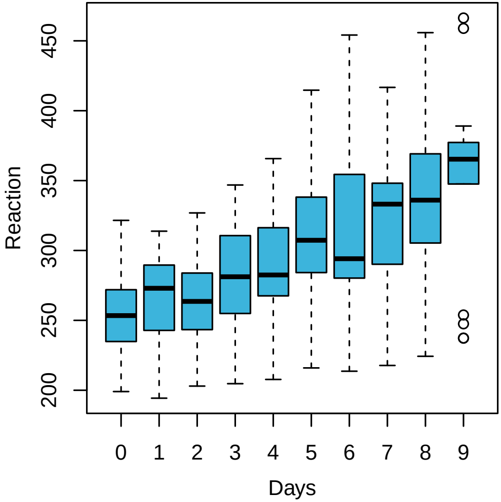
# Sleep study

- Average reaction time per day for subjects in a sleep deprivation study
- On day 0, the subjects had their normal amount of sleep
- Starting that night they were restricted to 3 hours of sleep per night
- Observations represent the average reaction time on a series of tests given each day to each subject

A data frame with 180 observations on the following 3 variables

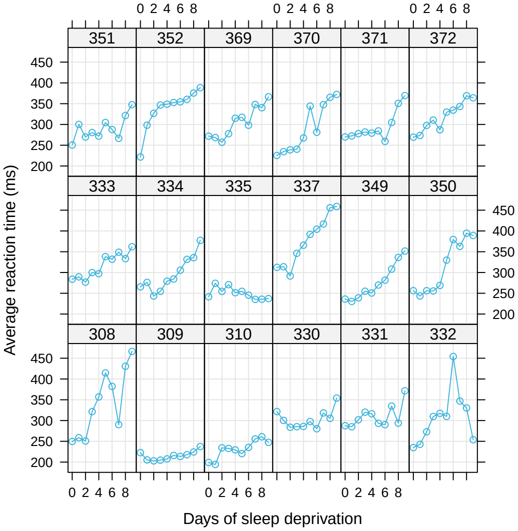
| | |
|---|---|
| `Reaction` | Average reaction time (ms) |
| `Days` | Number of days of sleep deprivation |
| `Subject` | Subject number on which the observation was made |

# Visualization of data




```
boxplot(Reaction ~ Days, sleepstudy)
```

# Visualization of individual data




```
library(lattice)

xyplot(Reaction ~ Days | Subject,
  data = sleepstudy,
  type = c("g", "b"),
  xlab = "Days of sleep deprivation",
  ylab = "Average reaction time (ms)",
  aspect = "xy")
```

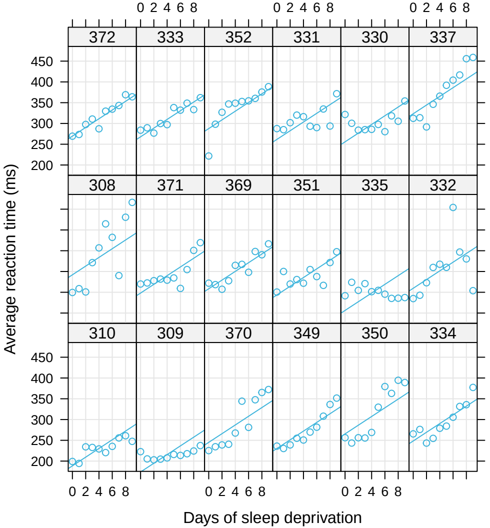
# Random intercept model

- The random intercept model adds a random intercept for each subject

$$y_{ij} = \beta_0 + \beta_1 \, Days_{ij} + \upsilon_{0i} + \varepsilon_i$$

with $\upsilon_{0i} \overset{iid}{\sim} N(0, \sigma_\upsilon^2)$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

- The slope is identical for each subject (and the population)



# Random slope model

- The random slope model adds a random intercept and a random slope for each subject

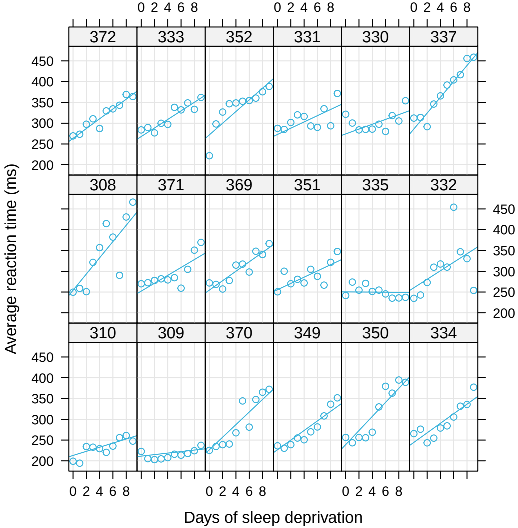
$$y_{ij} = \beta_0 + \beta_1 \, Days_{ij} + \upsilon_{0i} + \upsilon_{1i} \, Days_{ij} + \varepsilon_{ij}$$

with

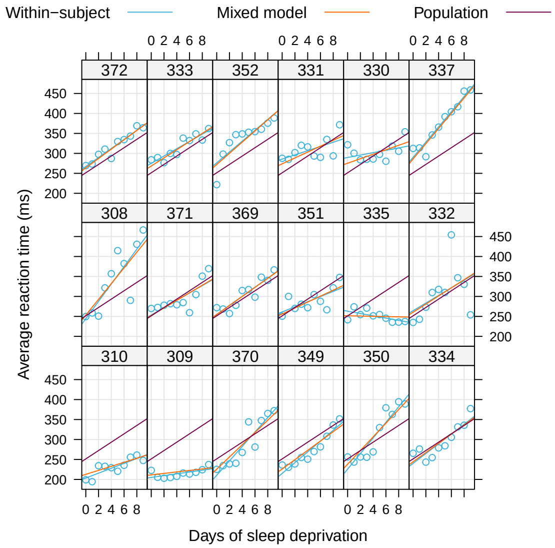
$$\begin{pmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{pmatrix} \overset{iid}{\sim} N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_\upsilon = \begin{pmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0 \upsilon_1} \\ \sigma_{\upsilon_0 \upsilon_1} & \sigma^2_{\upsilon_1} \end{pmatrix} \right)$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$$

- Individual slopes for each subject



# Partial pooling




- Within-subject regression line shows regression line fitted to data for each individual
- Population regression line shows fixed effects for mixed-effects model
- Mixed model regression line shows individual regression lines as predicted by mixed-effects models

# References

Bates, D. (2010). *lme4: Mixed-effects modeling with R (book draft).* Retrieved from https://lme4.r-forge.r-project.org/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. Retrieved from https://CRAN.R-project.org/package=lme4/vignettes/lmer.pdf doi: 10.18637/jss.v067.i01