

# Longitudinal data

Nora Wickelmaier

Last modified: October 17, 2024

# Outline

- ① Introduction mixed-effects models
- ② Sleep study
- ③ Parameter estimation

## ① Introduction mixed-effects models

## Mixed-effects models

- Mixed-effects models take account of dependencies in hierarchical, longitudinal, and other dependent data
- Mixed-effects models have been developed in a variety of disciplines with varying names and terminology
  - Random-effects models (statistics, econometrics)
  - Variance and covariance component analysis (statistics)
  - Hierarchical linear models (education)
  - Multilevel models (sociology)
  - Contextual-effects models (sociology, political science)
  - Random-coefficient models (econometrics)
  - Repeated-measures models (statistics, psychology)

Fox (2016)

## Data schema for dependent data

Person	Time	Observ.	Covariates		
1	1	$y_{11}$	$x_{111}$	$\dots$	$x_{11p}$
1	2	$y_{12}$	$x_{121}$	$\dots$	$x_{12p}$
.	.	.	.	$\dots$	.
1	$n_1$	$y_{1n_1}$	$x_{1n_11}$	$\dots$	$x_{1n_1p}$
.	.	.	.	$\dots$	.
.	.	.	.	$\dots$	.
$N$	1	$y_{N1}$	$x_{N11}$	$\dots$	$x_{N1p}$
$N$	2	$y_{N2}$	$x_{N21}$	$\dots$	$x_{N2p}$
.	.	.	.	$\dots$	.
$N$	$n_N$	$y_{Nn_N}$	$x_{Nn_N1}$	$\dots$	$x_{Nn_Np}$

- $i = 1, \dots, N$  persons
- $j = 1, \dots, n_i$  time points for person  $i$
- All observations:  $\sum_i^N n_i$
- Vector of all observations for person  $i$   
 $(\mathbf{y}_i)_{n_i \times 1}$
- Vector of covariates for person  $i$  at time point  $j$   
 $(\mathbf{x}_{ij})_{p \times 1}$
- All covariates of person  $i$   
 $(\mathbf{X}_i)_{n_i \times p}$

## 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<sup>1</sup>
  - 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 conclusions about the underlying population and its variation

---

<sup>1</sup>Some critical discussion on these definitions:

[http://andrewgelman.com/2005/01/25/why\\_i\\_dont\\_use/](http://andrewgelman.com/2005/01/25/why_i_dont_use/)

## Crossed random effects

- In many experiments in psychology the reaction of each subject ( $j = 1, \dots, N$ ) to a complete set of stimuli or items ( $k = 1, \dots, K$ ) is measured

$$y_{ijk} = \beta_0 + \beta_i x_i + v_{0j} + \eta_{0k} + \varepsilon_{ijk}$$

with  $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$ ,  $v_{0j} \stackrel{iid}{\sim} N(0, \sigma_v^2)$ , and  $\eta_{0k} \stackrel{iid}{\sim} N(0, \sigma_\eta^2)$

- Data are completely crossed: all subjects work on all items

		Subject				
		1	2	3	...	20
Item	1	1	1	1	...	1
	2	1	1	1	...	1
	3	1	1	1	...	1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	10	1	1	1	...	1

## Crossed random effects

```
> head(dat, 12)
```

	id	cond	item	av
1	1	cond1	item01	105
2	1	cond1	item02	116
3	1	cond1	item03	104
4	1	cond1	item04	81
5	1	cond1	item05	99
6	1	cond1	item06	109
7	1	cond1	item07	100
8	1	cond1	item08	103
9	1	cond1	item09	89
10	1	cond1	item10	94
11	2	cond1	item01	107
12	2	cond1	item02	100

```
> xtabs( ~ item + id, dat)
```

	id																			
item	1	2	3	4	5	6	7	8	9	...	20									
item01	1	1	1	1	1	1	1	1	1	...	1									
item02	1	1	1	1	1	1	1	1	1	...	1									
item03	1	1	1	1	1	1	1	1	1	...	1									
item04	1	1	1	1	1	1	1	1	1	...	1									
item05	1	1	1	1	1	1	1	1	1	...	1									
item06	1	1	1	1	1	1	1	1	1	...	1									
item07	1	1	1	1	1	1	1	1	1	...	1									
item08	1	1	1	1	1	1	1	1	1	...	1									
item09	1	1	1	1	1	1	1	1	1	...	1									
item10	1	1	1	1	1	1	1	1	1	...	1									



## Nested random effects

- We talk about nested random effects, when certain levels of one factor are combined only with certain levels of another factor (factors are nested within each other)
- The standard example for a nested design are students in classes in schools

$$y_{ijk} = \beta_0 + v_{0i} + \eta_{0ij} + \varepsilon_{ijk}$$

with  $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$ ,  $v_{0i} \stackrel{iid}{\sim} N(0, \sigma_v^2)$ , and  $\eta_{0ij} \stackrel{iid}{\sim} N(0, \sigma_\eta^2)$

		Classes									
		1	2	3	4	5	6	7	8	9	10
Schools	1	$n_1$	$n_2$	.	.	.	.	.	.	.	.
	2	.	.	$n_3$	$n_4$	.	.	.	.	.	.
	3	.	.	.	.	$n_5$	$n_6$	.	.	.	.
	4	.	.	.	.	.	.	$n_7$	$n_8$	.	.
	5	.	.	.	.	.	.	.	.	$n_9$	$n_{10}$

## Nested random effects

```
> head(dat, 12)
   id  class school av
1   1 class01 school1 105
2   2 class01 school1  93
3   3 class01 school1 119
4   4 class01 school1  94
5   5 class01 school1 107
6   6 class01 school1 100
7   7 class01 school1  98
8   8 class01 school1 108
9   9 class01 school1 108
10 10 class01 school1  91
11 11 class01 school1  75
12 12 class01 school1  76
```

```
> xtabs(~school + class, dat, sparse = TRUE)
school1 19 17  .  .  .  .  .  .  .  .
school2  .  . 23 21  .  .  .  .  .  .
school3  .  .  .  . 18 20  .  .  .  .
school4  .  .  .  .  .  . 19 21  .  .
school5  .  .  .  .  .  .  .  . 23 19
```

## Linear model

- We have observations  $(y_i, x_{i1}, \dots, x_{ip})$  with  $i = 1, \dots, N$  and the stochastic model

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \dots + \beta_p \cdot x_{ip} + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2) \text{ i.i.d.}$$

- In matrix notation

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

which corresponds to

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

## Linear mixed-effects model

- The linear mixed-effects model has the general form

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{v}_i + \boldsymbol{\varepsilon}_i$$

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

$$\mathbf{v}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_v) \text{ i.i.d.}, \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}) \text{ i.i.d.}$$

- This implies for the marginal covariance matrix

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

# Linear mixed-effects model

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

## Linear mixed-effects model

- Random intercept model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_{0i} + \varepsilon_i$$

with  $v_{0i} \sim N(0, \sigma_v^2)$  i.i.d.,  $\varepsilon_{ij} \sim N(0, \sigma^2)$  i.i.d.,  $v_{0i}$  and  $\varepsilon_{ij}$  i.i.d.

- Random slope model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_{0i} + v_{1i} x_{ij} + \varepsilon_i$$

with

$$\begin{pmatrix} v_{0i} \\ v_{1i} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_v = \begin{pmatrix} \sigma_{v_0}^2 & \sigma_{v_0 v_1} \\ \sigma_{v_0 v_1} & \sigma_{v_1}^2 \end{pmatrix} \right) \text{ i.i.d.} \\ \varepsilon_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}) \text{ i.i.d.}$$

## ② Sleep study

## Longitudinal data

- Longitudinal data consist of repeated measurements on the same subject taken over time
  - are special cases of mixed-effects models
  - contain a time covariate
  - time trends within and between subjects are of interest
- We will look at an example from the `lme4` package<sup>2</sup>

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

---

<sup>2</sup>The example can be found in a book draft by Douglas Bates: <http://lme4.r-forge.r-project.org/> or the JSS paper on `lme4`: <https://www.jstatsoft.org/article/view/v067i01>



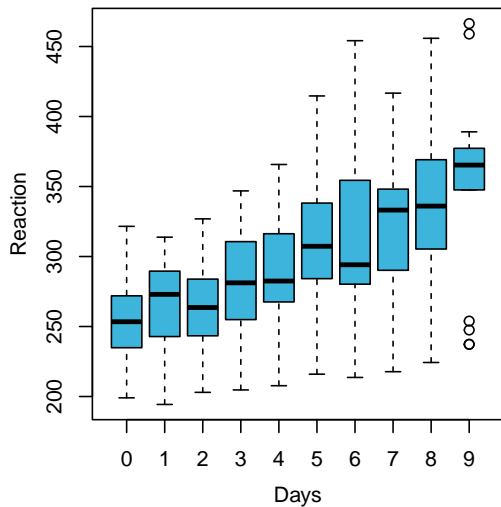
## Sleep study

- 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

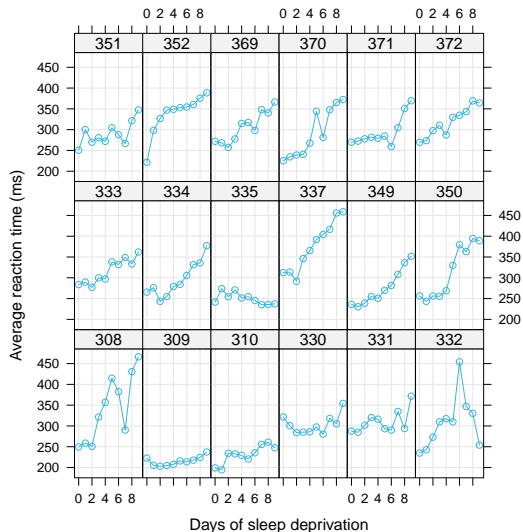
Reaction	Average reaction time (ms)
Days	Number of days of sleep deprivation
Subject	Subject number on which the observation was made

## 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")
```

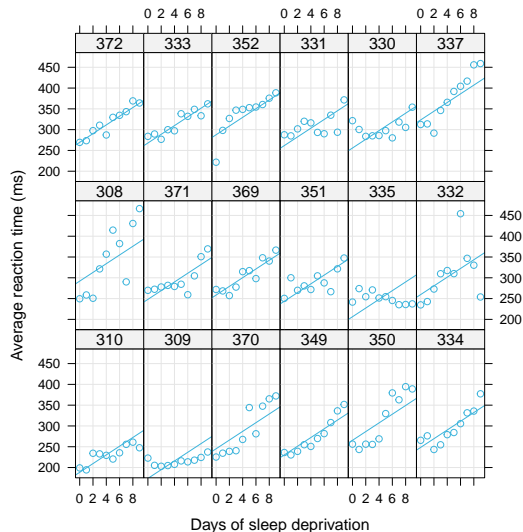
# Random intercept model

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

$$y_{ij} = \beta_0 + \beta_1 \text{Days}_{ij} + v_{0i} + \varepsilon_i$$

with  $v_{0i} \stackrel{iid}{\sim} N(0, \sigma_v^2)$ ,  $\varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$

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



## Random intercept model

```
lme0 <- lmer(Reaction ~ Days + (1 | Subject), sleepstudy)
summary(lme0)

# model matrices
X <- model.matrix(~ Days, sleepstudy)
Z <- model.matrix(~ 0 + Subject, sleepstudy)

# coefficients
coef(lme0)
fixef(lme0)
ranef(lme0)
```

# Random slope model

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

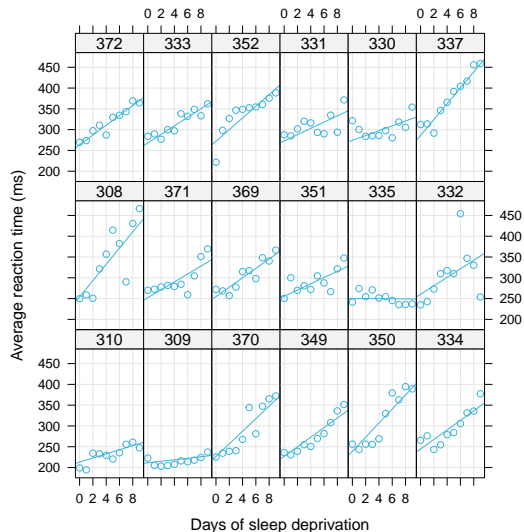
$$y_{ij} = \beta_0 + \beta_1 \text{Days}_{ij} + v_{0i} + v_{1i} \text{Days}_{ij} + \varepsilon_{ij}$$

with

$$\begin{pmatrix} v_{0i} \\ v_{1i} \end{pmatrix} \stackrel{iid}{\sim} N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma}_v = \begin{pmatrix} \sigma_{v_0}^2 & \sigma_{v_0 v_1} \\ \sigma_{v_0 v_1} & \sigma_{v_1}^2 \end{pmatrix} \right)$$

$$\varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

- Individual slopes for each subject



## Random slope model

```
lme1 <- lmer(Reaction ~ Days + (Days | Subject), sleepstudy)
summary(lme1)

# model matrices
X <- model.matrix(~ Days, sleepstudy)
Z <- model.matrix(~ 0 + Subject + Subject:Days, sleepstudy)

# coefficients
coef(lme1)
fixef(lme1)
ranef(lme1)
```

## Model with uncorrelated random effects

- We will now consider a model without correlated random effects

$$y_{ij} = \beta_0 + \beta_1 \text{Days}_{ij} + v_{0i} + v_{1i} \text{Days}_{ij} + \varepsilon_{ij}$$

with

$$\mathbf{v} \stackrel{iid}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_v = \begin{pmatrix} \sigma_{v_0}^2 & 0 \\ 0 & \sigma_{v_1}^2 \end{pmatrix}\right) \quad \text{and} \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

```
lme2 <- lmer(Reaction ~ Days + (Days || Subject), sleepstudy)
summary(lme2)

# likelihood-ratio test
anova(lme1, lme2)

# confidence intervals
confint(lme2)
```



## Confidence intervals and interpretation

- The results indicate that the extra parameter  $\sigma_{v_0v_1}$  does not produce a significantly better fit
- Results show that we have a significant effect for days with an average increase in reaction time of 10.47 ms for each day of sleep deprivation
- We get an estimate of  $\sigma_{v_0} = 24.17$  for the standard deviation of reaction time for subjects and a standard deviation of  $\sigma_{v_1} = 5.80$  for the dependence of reaction time on days of sleep deprivation

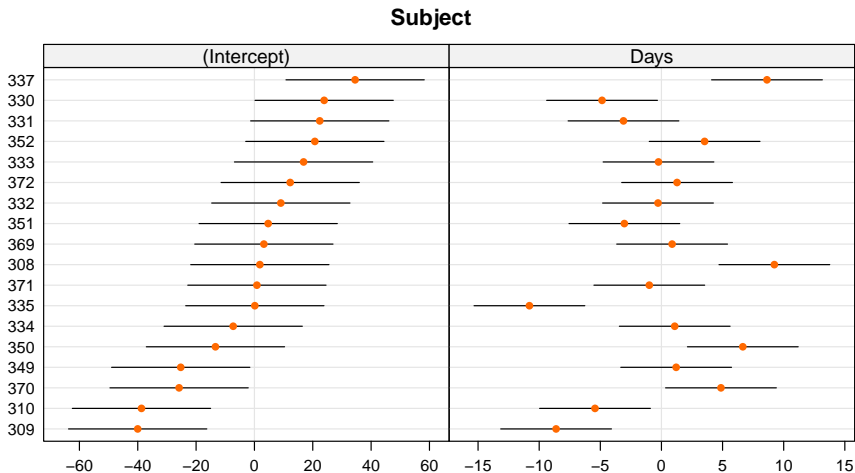
## Examining random effects and predictions

```
coef(lme2)
fixef(lme2)
ranef(lme2)

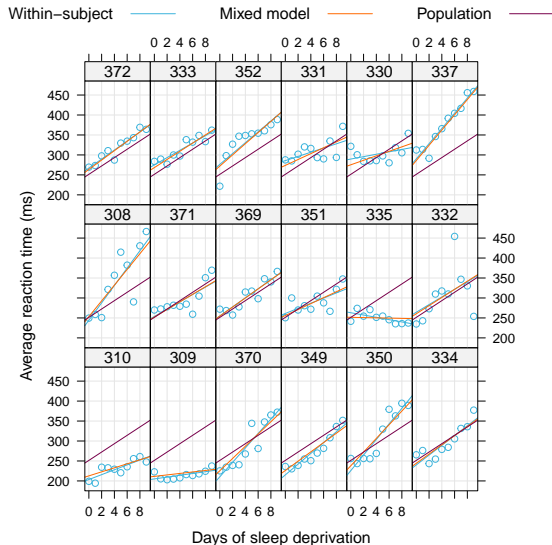
# correlational structure for intercepts and slopes
dotplot(ranef(lme2, condVar = TRUE),
        scales = list(x = list(relation = "free")))[[1]]

# predictions for all subjects
predict(lme2)
```

# Examining random effects and predictions



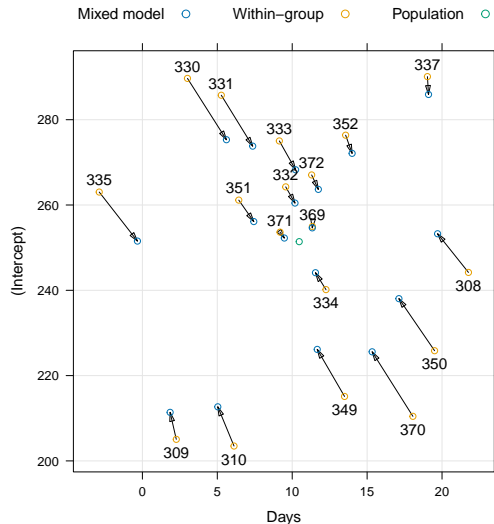
## Examining random effects and predictions



- 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

# Shrinkage

- When per-subject slopes and intercepts calculated from a mixed-effects model are compared to estimated slopes and intercepts within subjects, estimates from mixed-effects model are closer to the population estimates (the fixed effects)
- This pattern is sometimes described as *shrinkage* of coefficients toward the population values
- The more within-subject variance, the stronger parameters *shrink* towards the population parameters



# Assumptions

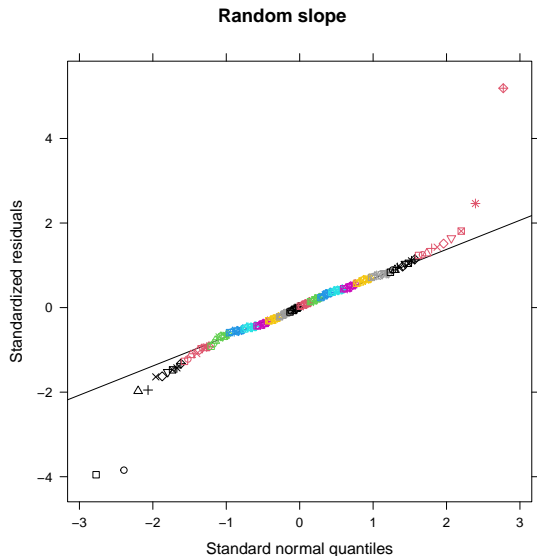
- Assumptions can be checked visually
- Normality assumption

```
qqmath(lme2,  
       col = sleepstudy$Subject,  
       pch = sleepstudy$Days)
```

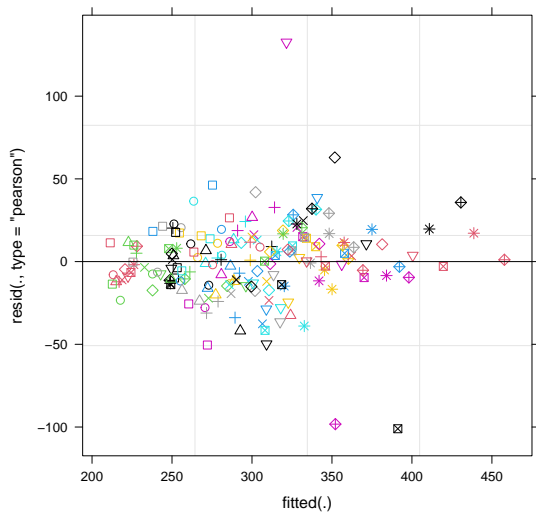
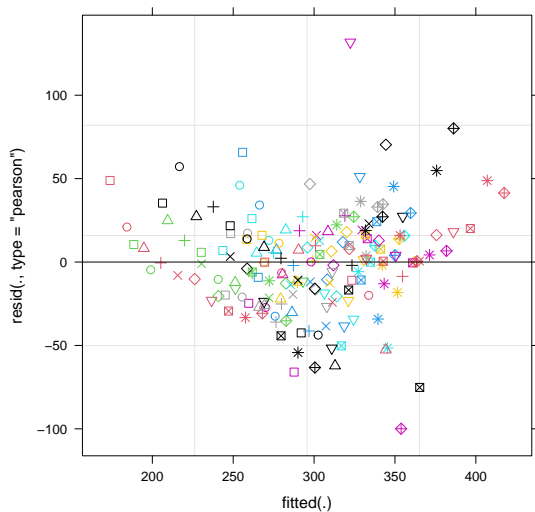
- Independence assumption

```
plot(lme2,  
     col = sleepstudy$Subject,  
     pch = sleepstudy$Days)
```

[https://bbolker.github.io/morelia\\_2018/notes/mixedlab.html](https://bbolker.github.io/morelia_2018/notes/mixedlab.html)



# Assumptions

**Random slope****Random intercept**

### ③ Parameter estimation



# Maximum Likelihood Estimation

- The maximum likelihood principle was introduced by R. A. Fisher
  - It can be used for almost all estimation problems, also complex ones
  - The resulting estimation functions have many desirable properties
- Obtaining the MLE  $\hat{\vartheta}$  takes the following steps

1. Generate the log-likelihood

$$\log L(\vartheta)$$

2. Take the derivative of the log-likelihood with respect to parameter  $\vartheta$

$$(\log L)'(\vartheta) = \frac{d \log L(\vartheta)}{d\vartheta}$$

3. Set the derivative equal to 0

$$(\log L)'(\hat{\vartheta}) = 0$$

4. Solve the resulting equation for the estimator  $\hat{\vartheta}$

## Restricted Maximum Likelihood Estimation

- The restricted maximum likelihood (REML) approach is a form of Maximum Likelihood Estimation
- In particular, REML is used as a method for fitting linear mixed-effects models
- In contrast to traditional Maximum Likelihood Estimation, REML can produce unbiased estimates of variance and covariance parameters
- MLE and REML do not outmatch each other, both can be used
- REML is the default in `lme4`, but if we want to conduct likelihood ratio tests, only MLE estimates are valid

## References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
<https://doi.org/10.18637/jss.v067.i01>
- Bates, D. M. (2010). lme4: Mixed-effects modeling with R.
- Fox, J. (2016). *Applied regression analysis and generalized linear models*. Sage Publications.