

Multilevel (mixed-effects) modelling in R

Matteo Lisi

<https://mattelisi.github.io/>

May 21, 2019

A linear model

- The dependent variable y is modeled as a weighted combination of the independent variables, plus an additive error ϵ

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_n x_{ni} + \epsilon_i$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

A linear model

- The dependent variable y is modeled as a weighted combination of the independent variables, plus an additive error ϵ

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

A linear model

Matrix notation

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{m1} & \dots & x_{mn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 1 \begin{pmatrix} a \\ c \end{pmatrix} + 2 \begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} a + 2b \\ c + 2d \end{pmatrix}$$

Linear mixed-effects models

- Classical' linear models are *fixed-effects* only:

Linear mixed-effects models

- Classical' linear models are *fixed-effects* only:
 - independent variables are all experimental manipulation (they are not random).

Linear mixed-effects models

- Classical' linear models are *fixed-effects* only:
 - independent variables are all experimental manipulation (they are not random).
 - the only random source of variation is the residual error $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

Linear mixed-effects models

- Classical' linear models are *fixed-effects* only:
 - independent variables are all experimental manipulation (they are not random).
 - the only random source of variation is the residual error $\epsilon \sim \mathcal{N}(0, \sigma^2)$.
- However *observations* (e.g. trials) are often grouped according to *observational clusters* (e.g. subjects), random samples from a larger population, on which we'd like to make inferences.

Linear mixed-effects models

- Classical' linear models are *fixed-effects* only:
 - independent variables are all experimental manipulation (they are not random).
 - the only random source of variation is the residual error $\epsilon \sim \mathcal{N}(0, \sigma^2)$.
- However *observations* (e.g. trials) are often grouped according to *observational clusters* (e.g. subjects), random samples from a larger population, on which we'd like to make inferences.
- To generalize from the sample to the population, these variables should be treated as *random-effects*. Mixed-effects models allow to do that by explicitly modelling the population distribution.

Linear mixed-effects models

- A model with both fixed and random effects is called a *mixed-effects model* (but also: multilevel, hierarchical)

Linear mixed-effects models

- A model with both fixed and random effects is called a *mixed-effects model* (but also: multilevel, hierarchical)
- Mixed-effects models describe the relationship between a dependent variable and some independent variables that are grouped according to one (or more) classification factor (e.g., subject, stimulus, session)

Linear mixed-effects models

- A model with both fixed and random effects is called a *mixed-effects model* (but also: multilevel, hierarchical)
- Mixed-effects models describe the relationship between a dependent variable and some independent variables that are grouped according to one (or more) classification factor (e.g., subject, stimulus, session)
- A typical example is *repeated-measures data*, where observations are grouped according to the subject: common random effects are associated with observations made on the same subject (i.e., sharing the same level of the classification factor)

Linear mixed-effects models

- A model with both fixed and random effects is called a *mixed-effects model* (but also: multilevel, hierarchical)
- Mixed-effects models describe the relationship between a dependent variable and some independent variables that are grouped according to one (or more) classification factor (e.g., subject, stimulus, session)
- A typical example is *repeated-measures data*, where observations are grouped according to the subject: common random effects are associated with observations made on the same subject (i.e., sharing the same level of the classification factor)
- Random effects are treated as random variations around a *population* mean. These variations are (usually) assumed to have a Gaussian distribution.

Random intercept

A simple example: a *random-intercept* model. Regressions have the same slope in each of the J groups ($j = 1, \dots, J$), but *random* (subject-specific) variations in intercept

$$y_{ij} = \beta_0 + b_j + \beta_1 x_{ij} + \epsilon_i$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$b \sim \mathcal{N}(0, \sigma_b^2)$$

Random slopes

A *random-slope* model. Regressions have the same intercept in each of the J groups ($j = 1, \dots, J$), but random (subject-specific) variations in slope

$$y_{ij} = \beta_0 + (\beta_1 + b_j) x_{ij} + \epsilon_i$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$b \sim \mathcal{N}(0, \sigma_b^2)$$

Mixed-effects model

- General formulation in matrix notation

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}b + \boldsymbol{\epsilon},$$

where \mathbf{X} and \mathbf{Z} are the known fixed-effects and random-effects regressor matrices.

Mixed-effects model

- General formulation in matrix notation

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}b + \epsilon,$$

where \mathbf{X} and \mathbf{Z} are the known fixed-effects and random-effects regressor matrices.

- The components of the residual error vector $\epsilon \sim \mathcal{N}(0, \sigma^2)$ are assumed to be i.i.d. (independent and identically distributed).

Mixed-effects model

- General formulation in matrix notation

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}b + \epsilon,$$

where \mathbf{X} and \mathbf{Z} are the known fixed-effects and random-effects regressor matrices.

- The components of the residual error vector $\epsilon \sim \mathcal{N}(0, \sigma^2)$ are assumed to be i.i.d. (independent and identically distributed).
- The random-effect components, $b \sim \mathcal{N}(0, \Omega)$ are assumed to be normally distributed with mean 0, however they are not necessarily independent (the components b_j can be correlated, and correlations can be estimated).

Mixed-effects model

- Example, a model with correlated random slopes and random intercepts:

$$y_{ij} = \beta_0 + b_{j1} + (\beta_1 + b_{j1}) x_{ij} + \epsilon_i$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Omega = \begin{bmatrix} \text{Var}(b_0) & \text{cov}(b_0, b_1) \\ \text{cov}(b_0, b_1) & \text{Var}(b_1) \end{bmatrix} \right)$$

Likelihood function

- Parameters (β , σ^2 and Ω) are estimated by maximizing the likelihood function, which is the probability of the data, given the parameters¹.

¹but treated as a function of the parameters, keeping the data fixed

Likelihood function

- Parameters (β , σ^2 and Ω) are estimated by maximizing the likelihood function, which is the probability of the data, given the parameters¹.
- The likelihood *conditional to the random effects* is integrated with respect to the distribution of the random effects, to obtain the *marginal density* of the data

$$L(\beta, \sigma^2, \Omega \mid \text{data}) = \int \underbrace{p(\text{data} \mid \beta, \sigma^2, b)}_{\text{likelihood}} p(b \mid \Omega) db$$

¹but treated as a function of the parameters, keeping the data fixed

ML vs RML

- The R library lme4 provides two methods for estimating parameters: Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML).

ML vs RML

- The R library lme4 provides two methods for estimating parameters: Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML).
- ML tends to be biased and underestimates the variance components (e.g. Ω).

ML vs RML

- The R library lme4 provides two methods for estimating parameters: Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML).
- ML tends to be biased and underestimates the variance components (e.g. Ω).
- REML provides less biased variance estimates: conceptually similar to Bessel's correction for sample variance (using $n - 1$ instead of n in the denominator).

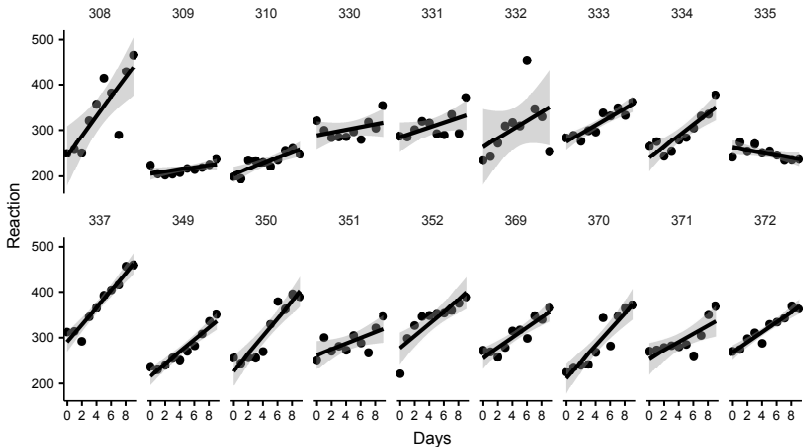
Advantages of multilevel models

- Improved estimates for repeated sampling (e.g. repeated measures designs).
- Particularly for imbalanced datasets (e.g. unequal number of trials across subjects).
- Avoid averaging (pre-averaging of data remove variation and can manufacture false confidence).
- Subject-specific standard error is taken into account in group-level estimates.
- Variation among group or individuals is modelled explicitly.
- Outperform classical methods in predictive ability.

- **sleepstudy** is a dataset in the lme4 package, with reaction times data from 18 subjects that were restricted to 3 hours of sleep for 10 days.
 - How reaction times changes with each sleep-deprived night?
 - Are individual difference in baseline response times related to individual differences in the effect of sleep deprivation?

```
> library(lme4)
> str(sleepstudy) 'data.frame': 180 obs. of 3 variables:
 $ Reaction: num 250 259 251 321 357 ...
 $ Days : num 0 1 2 3 4 5 6 7 8 9 ...
 $ Subject : Factor w/ 18 levels "308","309","310",...: 1 1 1 1 ...
```

```
ggplot(sleepstudy, aes(x=Days,y=Reaction))  
+geom_point() +geom_smooth(method="lm",color="black",lwd=1)  
+facet_wrap(~Subject,ncol=9) +nice_theme
```



Fitting the model (function `lmer()`)

```
library(lme4) # load the library
m.1 <- lmer(Reaction ~ Days + (Days|Subject), sleepstudy)
```

See here for more details on lme4 formula interface:

<https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>

equation	formula
$\beta_0 + \beta_1 X_i + e_{si}$	n/a (Not a mixed-effects model)
$(\beta_0 + b_{S,0s}) + \beta_1 X_i + e_{si}$	<code>~ X + (1 Subject)</code>
$(\beta_0 + b_{S,0s}) + (\beta_1 + b_{S,1s})X_i + e_{si}$	<code>~ X + (1 + X Subject)</code>
$(\beta_0 + b_{S,0s} + b_{I,0i}) + (\beta_1 + b_{S,1s})X_i + e_{si}$	<code>~ X + (1 + X Subject) + (1 Item)</code>
As above, but S_{0s}, S_{1s} independent	<code>~ X + (1 Subject) + (0 + X Subject) + (1 Item)</code>
$(\beta_0 + b_{S,0s} + b_{I,0i}) + \beta_1 X_i + e_{si}$	<code>~ X + (1 Subject) + (1 Item)</code>
$(\beta_0 + b_{I,0i}) + (\beta_1 + b_{S,1s})X_i + e_{si}$	<code>~ X + (0 + X Subject) + (1 Item)</code>

Summaries on model results obtained with the function `summary()`

```
> summary(m.1)
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (Days | Subject)
Data: sleepstudy

REML criterion at convergence: 1743.6

Scaled residuals:
   Min       1Q   Median       3Q      Max
-3.9536 -0.4634  0.0231  0.4633  5.1793

Random effects:
 Groups   Name                Variance Std.Dev. Corr
 Subject (Intercept)    611.90   24.737
           Days              35.08    5.923   0.07
 Residual                654.94   25.592
Number of obs: 180, groups: Subject, 18

Fixed effects:
              Estimate Std. Error t value
(Intercept)  251.405     6.824   36.843
Days         10.467     1.546    6.771

Correlation of Fixed Effects:
(Intr)
Days -0.138
```

Model information, value of fitting criterion (ML or RML), summary statistics on the distribution of residuals.

```
Linear mixed model fit by REML ['lmerMod']  
Formula: Reaction ~ Days + (Days | Subject)  
Data: sleepstudy
```

REML criterion at convergence: 1743.6

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9536	-0.4634	0.0231	0.4633	5.1793

Random effects parameters

(that is the matrix $\Omega = \begin{bmatrix} \text{Var}(b_0) & \text{cov}(b_0, b_1) \\ \text{cov}(b_0, b_1) & \text{Var}(b_1) \end{bmatrix}$)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	611.90	24.737	
	Days	35.08	5.923	0.07
Residual		654.94	25.592	

Number of obs: 180, groups: Subject, 18

Subject-specific coefficients obtained with

```
> ranef(m.1)
$Subject
  (Intercept)      Days
308  2.2575329   9.1992737
309 -40.3942719 -8.6205161
310 -38.9563542 -5.4495796
330  23.6888704 -4.8141448
331  22.2585409 -3.0696766
```

(...)

```
351  4.5777099 -3.0152825
352 20.8614523  3.5364062
369  3.2750882  0.8722876
370 -25.6110745  4.8222518
371  0.8070591 -0.9881730
372 12.3133491  1.2842380
```

with conditional variances for \Subject"

Fixed-effects coefficients (group-level estimates).

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.405	6.824	36.843
Days	10.467	1.546	6.771

Fixed-effects coefficients (group-level estimates).

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.405	6.824	36.843
Days	10.467	1.546	6.771

No p values!

Fixed-effects coefficients (group-level estimates).

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.405	6.824	36.843
Days	10.467	1.546	6.771

No p values!

- Except for some special case (classical balanced designs) it is not in general clear that the null distribution of the sums of squares is really an F distribution and what the degrees of freedom should be.
- Some packages provide approximations (e.g. Satterthwaite in `lmerTest`) that can be used in case in case of balanced designs.
- A quick-and-dirty way to test for effects are likelihood ratio test. The golden standard is bootstrapping.

Likelihood ratio test

If L_1 and L_2 are the maximised likelihoods of two *nested* models with $k_1 < k_2$ parameters, the test statistic is $2\log(L_2/L_1)$, which is approximately χ^2 with $k_2 - k_1$ degrees of freedom.

Likelihood ratio test

If L_1 and L_2 are the maximised likelihoods of two *nested* models with $k_1 < k_2$ parameters, the test statistic is $2\log(L_2/L_1)$, which is approximately χ^2 with $k_2 - k_1$ degrees of freedom.

- Example: test if the covariance $\text{cov}(b_0, b_1)$ is different from zero (e.g. are people with faster reaction times at baseline less affected by sleep deprivation?).

Likelihood ratio test

If L_1 and L_2 are the maximised likelihoods of two *nested* models with $k_1 < k_2$ parameters, the test statistic is $2\log(L_2/L_1)$, which is approximately χ^2 with $k_2 - k_1$ degrees of freedom.

- Example: test if the covariance $\text{cov}(b_0, b_1)$ is different from zero (e.g. are people with faster reaction times at baseline less affected by sleep deprivation?).

```
> anova(m.1, m.2)
refitting model(s) with ML (instead of REML)
Data: sleepstudy
Models:
m.2: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
m.1: Reaction ~ Days + (Days | Subject)
      Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
m.2   5 1762.0 1778.0 -876.00   1752.0
m.1   6 1763.9 1783.1 -875.97   1751.9 0.0639      1    0.8004
```

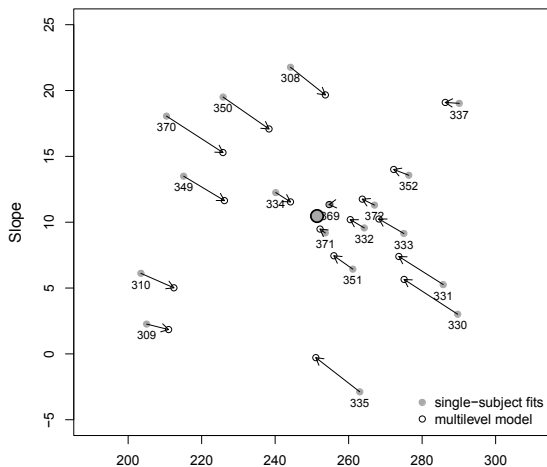
Bootstrap confidence intervals.

```
CI_fixef <- confint(m.1, method="boot", nsim=500, oldNames=F)
> print(CI_fixef, digits=2)
```

	2.5 %	97.5 %
sd_(Intercept) Subject	12.81	35.63
cor_Days.(Intercept) Subject	-0.53	0.81
sd_Days Subject	3.57	8.32
sigma	22.61	28.44
(Intercept)	237.02	264.76
Days	7.54	13.20

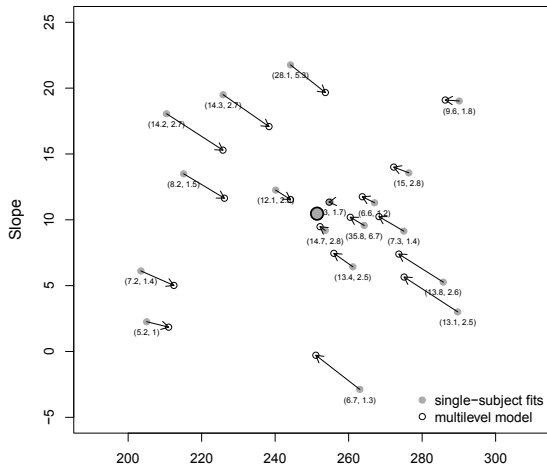
Shrinkage

The predicted \hat{b}_j (conditional modes of the random effects) can be seen as a "compromise" between the within-subject estimates and the population mean.



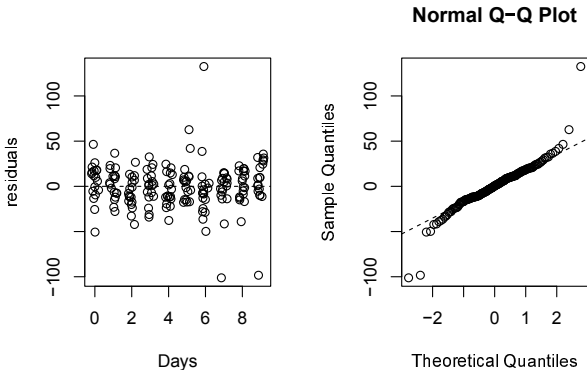
Shrinkage

The extent by which individual estimates are shrunk toward the mean depends on their value, on Ω , and on the subject-specific standard errors [now indicated, (SE_{b_0}, SE_{b_1})].



As for any linear model, it is important to check that residual errors are well-behaved.

```
par(mfrow=c(1,2))
plot(jitter(sleepstudy$Days),resid(m.1), xlab="Days",
      ylab="residuals");
abline(h=0,lty=2)
qqnorm(resid(m.1)); qqline(resid(m.1),lty=2)
```



Some useful references

- Bates, D. M. ("in progress") lme4: Mixed-effects modeling with R. freely available at <http://lme4.r-forge.r-project.org>
- Knoblauch, K., & Maloney, L. T. (2012). Modeling psychophysical data in R
- Pinheiro, J. C., & Bates, D. M. (2000). Mixed-Effects Models in S and S-PLUS.
- Moscatelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, 12(11)(26)
- Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2010). Experimental Effects and Individual Differences in Linear Mixed Models: Estimating the Relationship between Spatial, Object, and Attraction Effects in Visual Attention. *Frontiers in Psychology*, 1, 238.
- <http://www.r-bloggers.com>
- <https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>
- <http://stats.stackexchange.com>