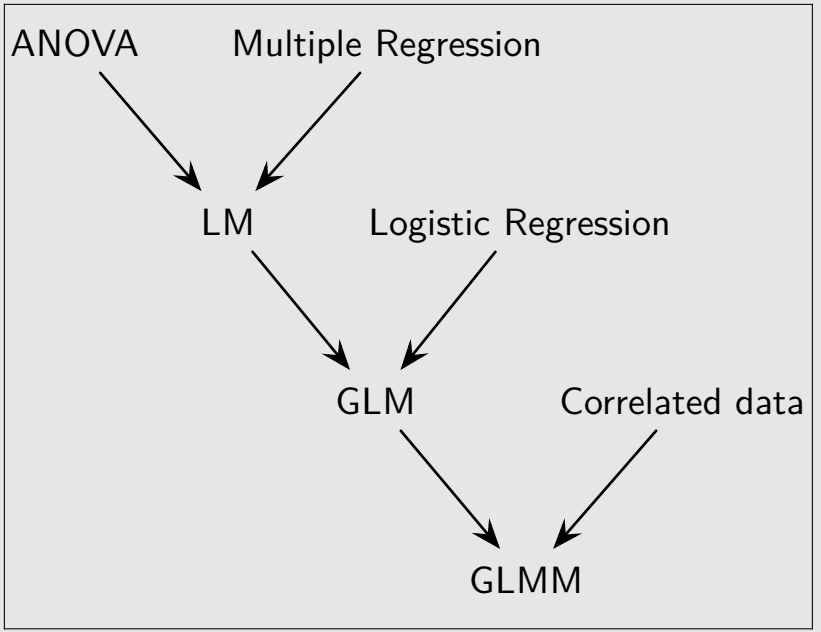
GLMM tutoR

Outline

Highlights the connections between different class of widely used models in psychological and biomedical studies.



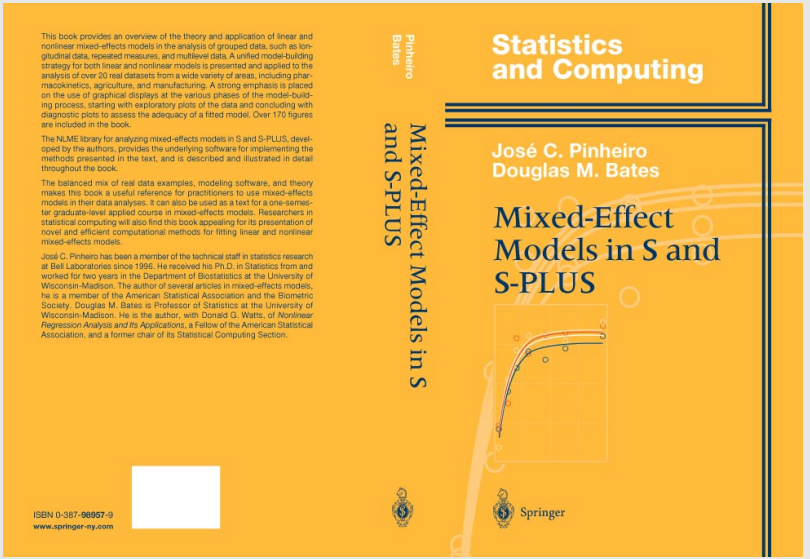
Overview of GLMM

Compared to standard (generalized) linear models, mixed-effect models further include random-effect terms that allow to reflect correlation between statistical units. Such clustered data may arise as the result of grouping of individuals (e.g., students nested in schools), within-unit correlation (e.g., longitudinal data), or a mix thereof (e.g., performance over time for different groups of subjects) [1, 2, 3, 4, 5].

Typically, this approach fall under the category of conditional models, as opposed to marginal models, like GLS or GEE, where we are interested in modeling population-averaged effects by assuming a working (within-unit) correlation matrix.

Mixed-effect models are not restricted to a single level of clustering, and they give predicted values for each cluster or level in the hierarchy.

Available tools in R

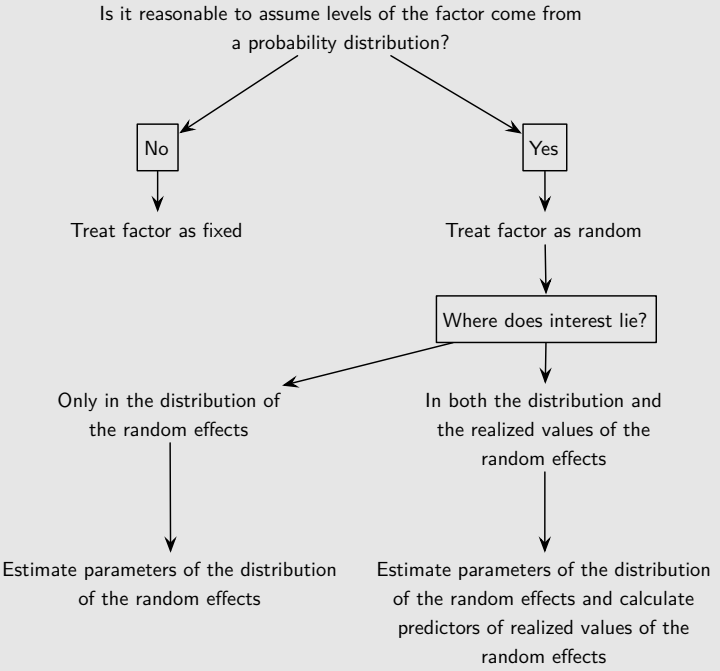


In preparation: Bates and coll., *lme4: Mixed-effects Modeling with R*, `http://lme4.r-forge.r-project.org/book/`

Fixed vs. random effects

There seems to be little agreement about what fixed and random effects really means: `http://bit.ly/Jd0EiZ`, but see also [6].

As a general decision workflow, we can ask whether we are interested in just estimating parameters for the random-effect terms, or get predictions at the individual level [7, p. 277].



Analysis of paired data

The famous "sleep" study [8] is a good illustration of the importance of using pairing information when available.

Test assuming independent subjects	`t.test(extra ~ group, data=sleep)`
Test acknowledging pairing	`t.test(extra ~ group, data=sleep, paired=TRUE)`

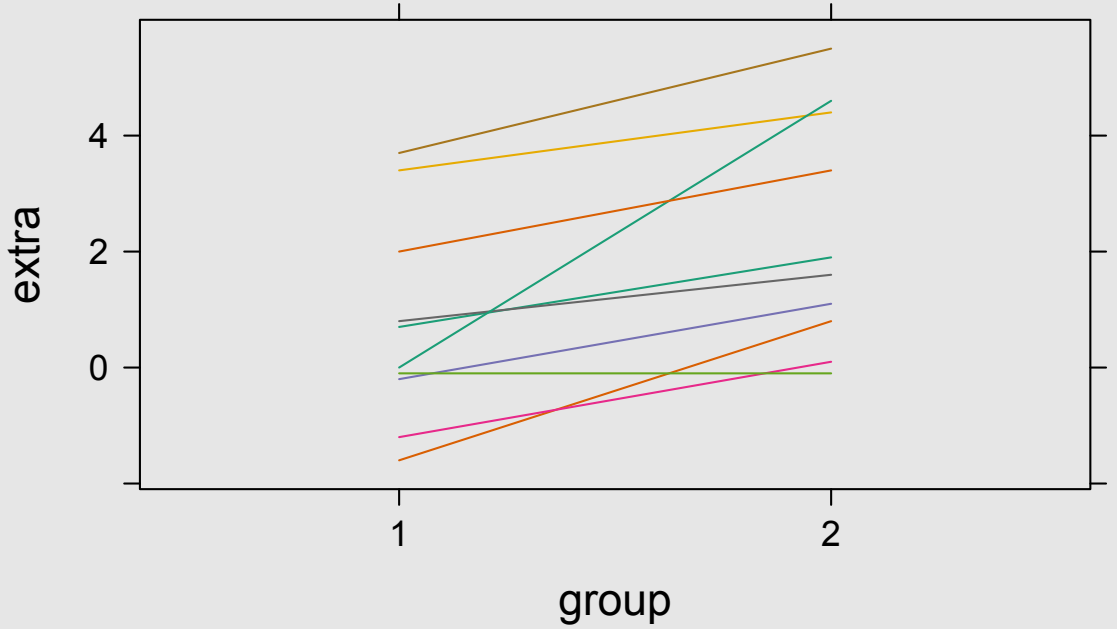
Generally, ignoring within-unit correlation results in a less powerful test.

Why? We know that

$$\mathsf{Var}(X_1 - X_2) = \mathsf{Var}(X_1) + \mathsf{Var}(X_2) - 2\mathsf{Cov}(X_1, X_2)$$

($\mathsf{Cov}(X_1, X_2) = \rho\sqrt{\mathsf{Var}(X_1)}\sqrt{\mathsf{Var}(X_2)}$).
Considering that $\mathsf{Cov}(X_1, X_2) = 0$ amounts to over-estimate variance of the differences, since $\mathsf{Cov}(X_1, X_2)$ will generally be positive.

A positive covariance in this case means that subjects having higher values on the first level will also have higher values on the second level.



Analysis of repeated-measures

ID	Pill type				Subject average
	None	Tablet	Capsule	Coated	
1	44.5	7.3	3.4	12.4	16.9
2	33.0	21.0	23.1	25.4	25.6
3	19.1	5.0	11.8	22.0	14.5
4	9.4	4.6	4.6	5.8	6.1
5	71.3	23.3	25.6	68.2	47.1
6	51.2	38.0	36.0	52.6	44.5
Pill type average	38.1	16.5	17.4	31.1	25.8

Lack of digestive enzymes in the intestine can cause bowel absorption problems, as indicated by excess fat in the feces. Pancreatic enzyme supplements can be given to ameliorate the problem [7, p. 254].

There is only one predictor, Pill type, which is attached to subject and period of time (subsumed under the repeated administration of the different treatment).

Variance components

Two different ways of decomposing the total variance:

One-way ANOVA	`aov(fecfat ~ pilltype, data=fat)`
Two-way ANOVA	`aov(fecfat ~ pilltype + subject, data=fat)`
RM ANOVA	`aov(fecfat ~ pilltype + Error(subject), data=fat)`

Source	DF	SS	MS	M1	M2*/M3
`pilltype`	3	2009	669.5	669.5/359.7 p=0.169	669.5/107.0 p=0.006
`subject`	5	5588	1117.7	—	1117.7/107.0 p=0.000*
Residuals	15	1605	107.0	—	—

The first model, which assumes independent observations, does not remove variability between subjects (about 77.8% of residual variance).

The last two models incorporate subject-specific effects:

$$y_{ij} = \mu + subject_i + pilltype_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$$

In the third model, we further assume $subject_i \sim \mathcal{N}(0, \sigma^2_s)$, independent of ε_{ij}.

Variance components (Con't)

The inclusion of a random effect specific to subjects allows to model several types of within-unit correlation at the level of the outcome.

What is the correlation between measurements taken from the same individual? We know that

$$\text{Cor}(y_{ij}, y_{ik}) = \frac{\text{Cov}(y_{ij}, y_{ik})}{\sqrt{\text{Var}(y_{ij})}\sqrt{\text{Var}(y_{ik})}}.$$

Because μ and $pilltype$ are fixed, and $\varepsilon_{ij} \perp subject_i$, we have

$$\begin{aligned}\text{Cov}(y_{ij}, y_{ik}) &= \text{Cov}(subject_i, subject_i) \\ &= \text{Var}(subject_i) \\ &= \sigma_s^2,\end{aligned}$$

and variance components follow from $\text{Var}(y_{ij}) = \text{Var}(subject_i) + \text{Var}(\varepsilon_{ij}) = \sigma_s^2 + \sigma_\varepsilon^2$, which is assumed to hold for all observations.

So that, we have

$$\text{Cor}(y_{ij}, y_{ik}) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_\varepsilon^2}$$

which is the proportion of the total variance that is due to subjects. It is also known as the intraclass correlation, ρ, and it measures the closeness of observations on different subjects (or within-cluster similarity).

- Subject-to-subject variability simultaneously raises or lowers all the observations on a subject.
- The variance-covariance structure in the above model is called compound symmetry:

$$\begin{bmatrix} \sigma_s^2 + \sigma_\varepsilon^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 + \sigma_\varepsilon^2 & \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 & \sigma_s^2 + \sigma_\varepsilon^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 + \sigma_\varepsilon^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}.$$

$$(\sigma^2 = \sigma_s^2 + \sigma_\varepsilon^2)$$

Estimating ρ

Observations on the same subject are modeled as correlated through their shared random subject effect. Using the random intercept model defined above, we can estimate ρ as follows:

Load required package	`library(nlme)`
Fit a random intercept model	`lme.fit <- lme(fecfat ~ pilltype, data=fat, random= ~ 1 | subject)`
ANOVA table	`anova(lme.fit)`
Get approximate 95% CIs	`intervals(lme.fit)`
$\hat{\sigma}_s$	`sigma.s <- as.numeric(VarCorr(lme.fit)[1,2])`
$\hat{\sigma}_\varepsilon$	`sigma.eps <- as.numeric(VarCorr(lme.fit)[2,2])`
$\hat{\rho}$=0.7025066	`sigma.s^2/(sigma.s^2+sigma.eps^2)`

It should be noted that REML method is used by default: random effects are estimated after having removed fixed effects. Tests for random effects ($H_0 : \sigma^2 = 0$) with LRT are conservative.

Estimating ρ (Con't)

From the ANOVA table, we can also extract the relevant variance components to get the desired $\hat{\rho}$:

Extract variance components

```
ms <- anova(lm(fecfat ~ pilltype + subject,
               data=fat))[[3]]
```

Subjects variance

```
vs <- (ms[2] - ms[3])/nlevels(fat$pilltype)
```

Residual variance

```
vr <- ms[3]
```

$\hat{\rho}$=0.7025066

```
vs / (vs+vr)
```

Notice that we could also use Generalized Least Squares, imposing compound symmetry for errors correlation:

Fit GLS model

```
gls.fit <- gls(fecfat ~ pilltype, data=fat,
    corr=corCompSymm(form= ~ 1 | subject))
```

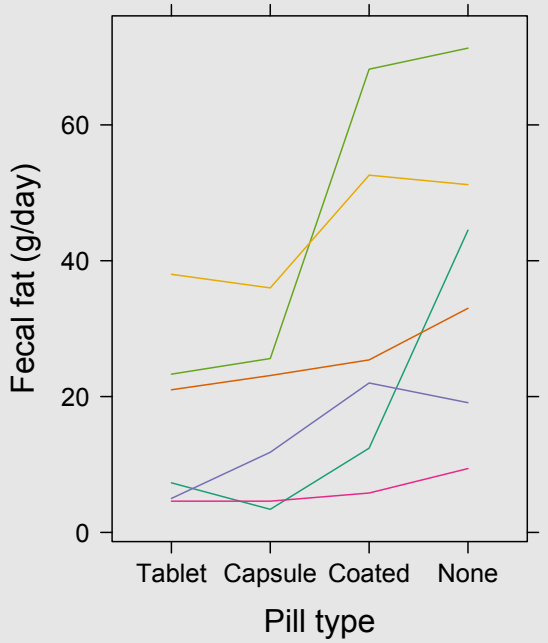
ANOVA table

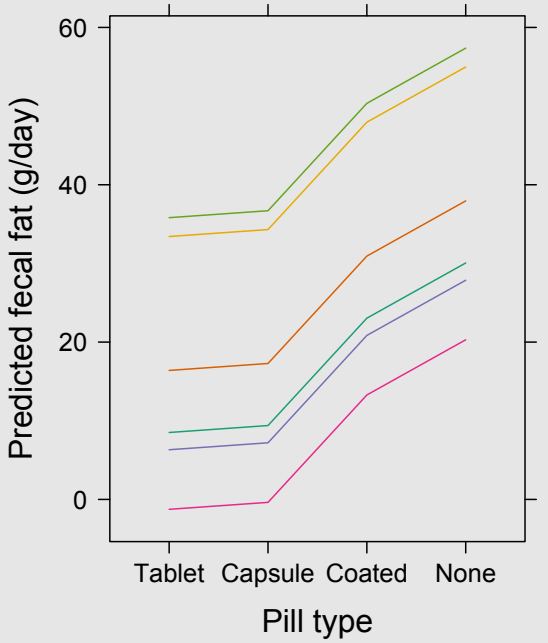
```
anova(gls.fit)
```

$\hat{\rho}$=0.7025074

```
intervals(gls.fit)
```

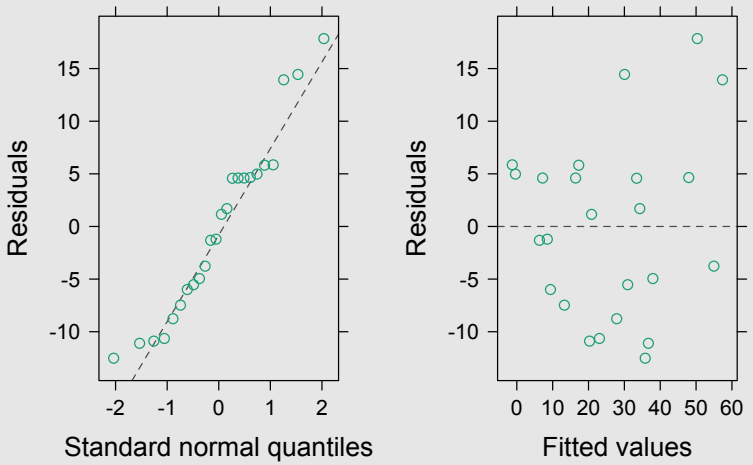
The imposed variance-covariance structure is clearly reflected in the predicted values.





Model diagnostic

Inspection of the distribution of the residuals and residuals *vs.* fitted values plots are useful diagnostic tools. It is also interesting to examine the distribution of random effects (intercepts and/or slopes).



Some remarks

- For a balanced design, the residual variance for the within-subject ANOVA and random-intercept model will be identical (the REML estimator is equivalent to ANOVA MSs). Likewise, Pill type effects and overall mean are identical.
- Testing the significance of fixed effects can be done using ANOVA (F-value) or by model comparison. In the latter case, we need to fit model by ML method (and not REML) because models will include differents fixed effects.

ANOVA table	`anova(lme.fit)`
Refit base model by ML	`lme.fit <- update(lme.fit, method="ML")`
Intercept-only model	`lme.fit0 <- update(lme.fit, fixed= . ~ - pilltype)`
LRT=14.6 (p=.0022)	`anova(lme.fit, lme.fit0)`

- Parametric bootstrap may be used to compute a more accurate p-value for the LRT (here a $\chi^2(3)$), see [9, pp. 158–161].

Variance-covariance structure

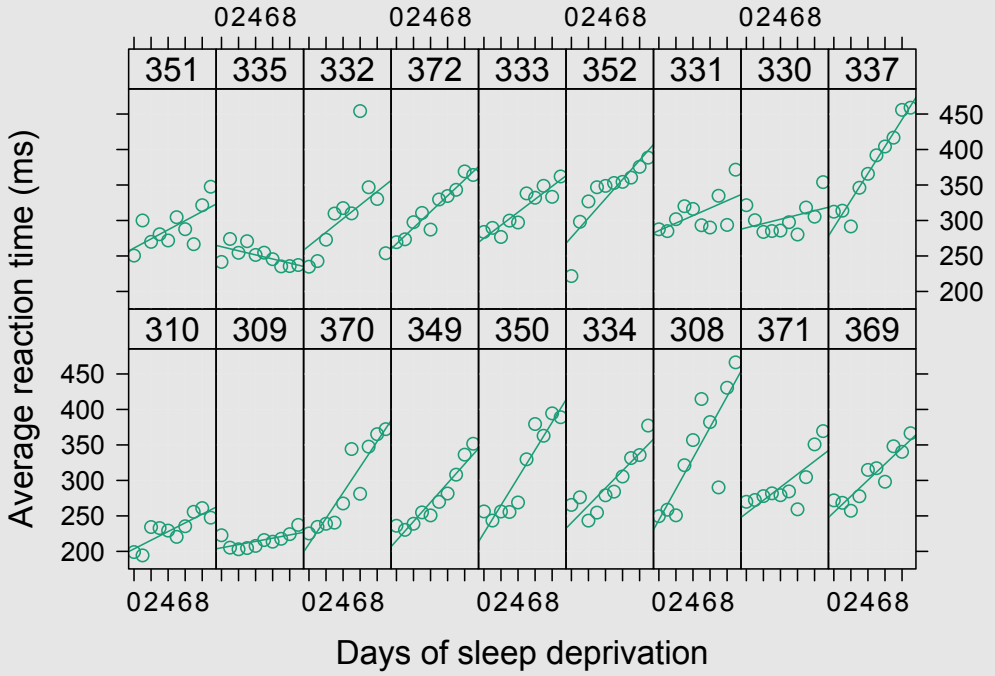
Other VC matrices can be choosen, depending on study design [10, pp. 232–237], e.g. Unstructured, First-order auto-regressive, Band-diagonal, AR(1) with heterogeneous variance.

The random-intercept model ensures that the VC structure will be constrained as desired. With repeated measures ANOVA, a common strategy is to use Greenhouse-Geisser or Huynh-Feldt correction to correct for sphericity violations, or to rely on MANOVA which is less powerful, e.g. [11, 12] (see also `http://homepages.gold.ac.uk/aphome/spheric.html`).

Mixed-effect models are more flexible as they allow to make inference on the correlation structure, and to perform model comparisons.

Longitudinal data

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. The observations represent the average reaction time on a series of tests given each day to each subject. D. Bates, Lausanne 2009, `http://bit.ly/Kj8VVj`



Results from simple linear regression

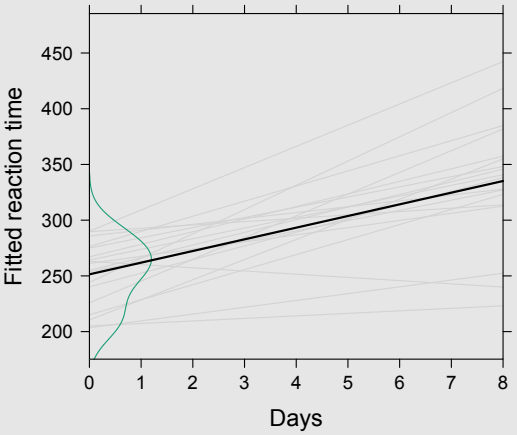
Individual regression lines ($N = 18$):

```
    (Intercept)      Days
25%    229.4167  6.194548
75%    273.0255 13.548395
```

Subjects exhibit different response profiles, and overall regression line is:

$$\tilde{y} = 251.4 + 10.5x.$$

How well does this equation capture observed trend (across subjects)?



- OLS estimates are correct, but their standard errors aren't (because independence is assumed).
- Individual predictions will be incorrect as well.

Some plausible models

- Random-intercept model

  `Reaction ~ Days + (1 | Subject)`

- Random-intercept and slope model

  `Reaction ~ Days + (Days | Subject)`

- Uncorrelated random effects (intercept and slope)

  `Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)`

```
Models:
lme.mod1: Reaction ~ Days + (1 | Subject)
lme.mod3: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
lme.mod2: Reaction ~ Days + (Days | Subject)
         Df    AIC    BIC  logLik   Chisq Chi Df Pr(>Chisq)
lme.mod1  4 1802.1 1814.9 -897.05
lme.mod3  5 1762.0 1778.0 -876.02 42.0525      1  8.885e-11 ***
lme.mod2  6 1764.0 1783.1 -875.99  0.0609      1     0.8051
```

Predicting random effects

In a model with random effects, the corresponding regression coefficients are no longer parameters and cannot be estimated as is done in usual linear models. Moreover, their expectation is zero. However, it is possible to rely on their posterior distribution (Bayesian approach). Combining conditional modes of the random effects and estimates of the fixed effects yields conditional modes of the within-subject coefficients. See also `lme4::mcmcsamp`.

It is easy to check that predicted random effects are related to fixed effects.

Two-way ANOVA

```
lm.mod1 <- aov(Reaction ~ Days + Subject,
               data=sleepstudy)
```

Extract fixed-effects

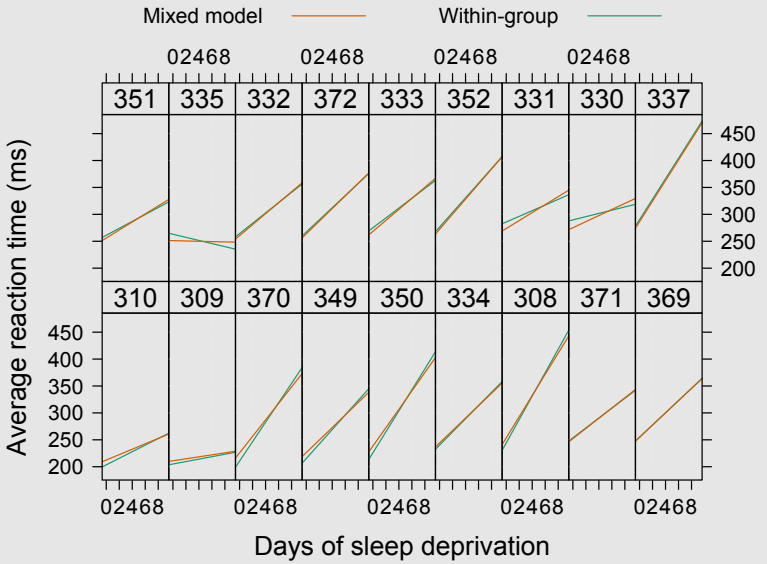
```
subj.fixef <- model.tables(lm.mod1,
       cterms="Subject")[[1]]$Subject
```

Constant ratio (0.9349347)

```
ranef(lme.mod1)$Subject/subj.fixef
```

Predicted values

$$\tilde{y}_i = (\underbrace{\hat{\beta}_0}_{\text{Fixed}} + \underbrace{\hat{u}_{0i}}_{\text{Random}}) + (\hat{\beta}_1 + \hat{u}_{1i})x$$

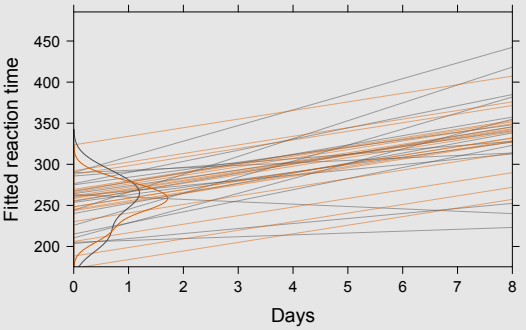


Shrinkage of within-unit coefficients

Predicted values from random-effect models are shrinkage estimators. For simple cases, it amounts to

$$\tau = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2/n_i},$$

where n_i is the sample size for the ith cluster. Here, $\tau = 37.1^2/(37.1^2 + 31.0^2/10) = 0.935$. There will be little skrinkage when statistical units are different, or when measurements are accurate, or with large sample.

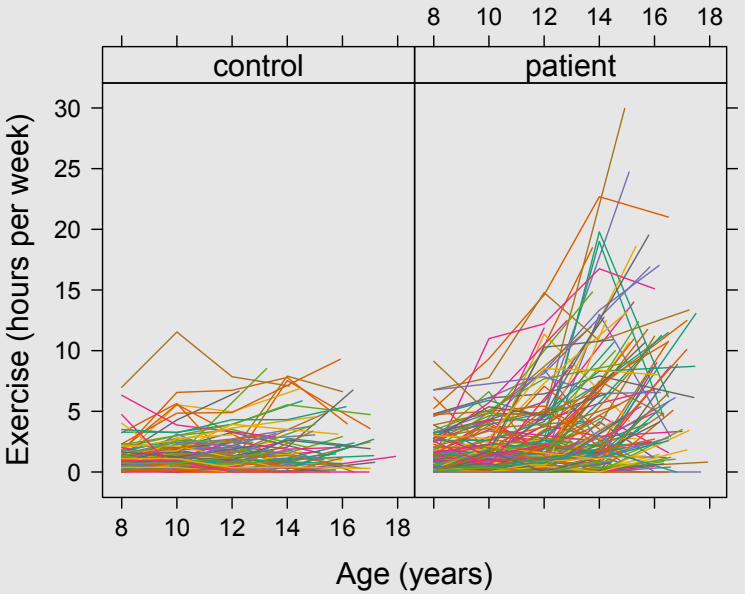



```
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 1378.18  37.124
 Residual              960.46  30.991
Number of obs: 180, groups: Subject, 18

Fixed effects:
            Estimate Std. Error t value
(Intercept) 251.4051     9.7459   25.80
Days         10.4673     0.8042   13.02
```

```
     Reaction Days Subject     pred
1   249.5600    0     308 292.1888
2   258.7047    1     308 302.6561
3   250.8006    2     308 313.1234
4   321.4398    3     308 323.5907
...
177 334.4818    6     372 332.3246
178 343.2199    7     372 342.7919
179 369.1417    8     372 353.2591
180 364.1236    9     372 363.7264
```

Categorical and continuous predictors



Blackmoor and Davis's data on exercise histories of 138 teenaged girls hospitalized for eating disorders and 98 control subjects. (John Fox, `car::Blackmoor`)

Modeling an interaction term

Random-intercept and/or slope models [13]:

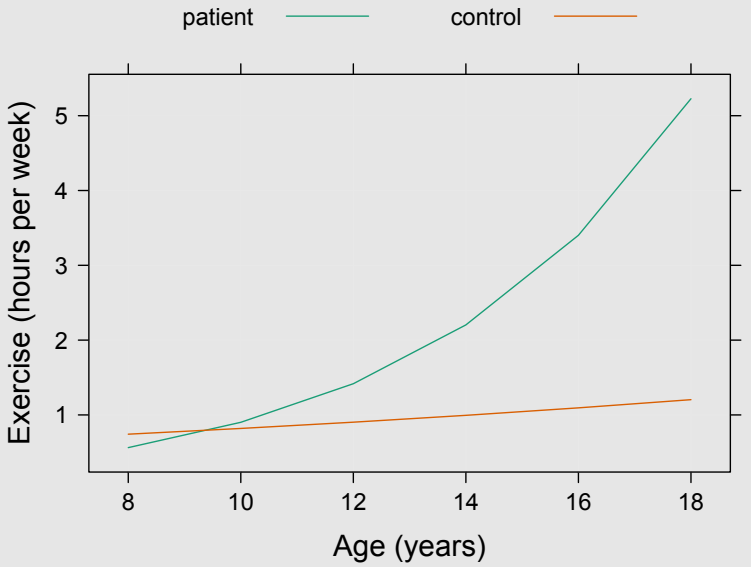
Load required dataset
```
data(Blackmoor, package="car")
```
Transform outcome
```
log.ex <- log(Blackmoor$exercise + 5/60, 2)
```
Random intercepts and slopes
```
lme.mod0 <- lme(log.ex ~ I(age-8)*group,
                random= ~ I(age-8) | subject,
                data=Blackmoor)
```
Random intercepts only
```
lme.mod1 <- update(lme.mod0,
                   random= ~ 1 | subject)
```
Uncorrelated random effects
```
lme.mod2 <- update(lme.mod0,
        random= ~ I(age-8) - 1 | subject)
```
LRT 1 vs. 0 ($p<.0001$)
```
anova(lme.mod0, lme.mod1)
```
LRT 2 vs. 0 ($p<.0001$)
```
anova(lme.mod0, lme.mod2)
```
Coefficients estimates
```
summary(lme.mod0)
```

Here are the results for the random intercept and slope model:

```
Fixed effects: log.exercise ~ I(age - 8) * group
                            Value  Std.Error  DF   t-value p-value
(Intercept)             -0.2760170 0.18236870 712 -1.513511  0.1306
I(age - 8)               0.0640222 0.03136052 712  2.041491  0.0416
grouppatient            -0.3539943 0.23529124 229 -1.504494  0.1338
I(age - 8):grouppatient  0.2398585 0.03940734 712  6.086646  0.0000
```

Predicted values

The interaction between the grouping factor and continuous predictor appears clearly when predicting outcomes from the random intercept and slope model. Fox [13] also demonstrates that adding serial correlation for the errors will improve goodness of fit.



Take-away message

- Conditional and marginal models extend classical generalized linear models by modeling within-unit correlation, and possibly errors structure.
- Mixed-effect models are flexible tools that allow to draw inference on such correlation structures, and also compute predicted values at different levels of clustering. Inference on random effects require sophisticated computational methods, though.
- Incorporating random effects and testing working hypothesis on within-unit correlation overcome the classical paradigm of repeated measures ANOVA.

References

[1] CE McCulloch and SR Searle. *Generalized, Linear, and Mixed Models*. Wiley, 2001.

[2] JK Lindsey. *Models for Repeated Measurements*. Oxford University Press, 2nd edition, 1999.

[3] SW Raudenbush and AS Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks CA: Sage, 2nd edition, 2002.

[4] A Gelman and J Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2007.

[5] GM Fitzmaurice. *Longitudinal Data Analysis*. CRC Press, 2009.

[6] A Gelman. Analysis of variance–why it is more important than ever. *Annals of Statistics*, 33(1):1–53, 2005.

[7] E Vittinghoff, DV Glidden, SC Shiboski and McCulloch. *Regression Methods in Biostatistics. Linear, Logistic, Survival, and Repeated Measures Models*. Springer, 2005.

[8] Student. The probable error of a mean. *Biometrika*, 6(1):1-25, 1908.

[9] JJ Faraway. *Extending the linear model with R*. Chapman & Hall/CRC, 2006.

[10] JC Pinheiro and DM Bates. *Mixed-Effects Models in S and S-PLUS*. Springer, 2000.

[11] H Abdi. The greenhouse-geisser correction. In N Salkind, editor, *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage, 2010.

[12] JH Zar. *Biostatistical Analysis*. Pearson, Prentice Hall, 4th edition, 1998.

[13] J Fox. Linear mixed models. App. to An R and S-PLUS Companion to Applied Regression. 2002