01NAEX - Lecture 10
Introduction to Mixed Linear Models

Jiri Franc

Czech Technical University
Faculty of Nuclear Sciences and Physical Engineering
Department of Mathematics

Introduction to Linear Mixed Models

Throughout the previous lectures we have assumed that all factors in an experiment were treated as **fixed factors**, so statistical inference was restricted to the specific levels observed in the study.

In many applications, some factor levels (e.g. subjects, schools, laboratories, batches) can be viewed as a random sample from a larger population of possible levels, and we wish to draw conclusions about that population, not only about the particular levels used in the experiment.

Such factors are called **random factors**, and the corresponding terms in the model are called **random effects**.

In this lecture we introduce **linear mixed-effects models**, which combine fixed effects and random effects in a single regression framework. Nested and split-plot designs, where random effects arise naturally, will be presented in following lectures.

Introduction to Linear Mixed Models

What is a linear mixed-effects model?

- Linear mixed-effects models extend ordinary linear models by including both **fixed effects** and **random effects**.
- They allow us to model correlation and heterogeneous variability that arise from grouping, clustering or repeated measurements on the same experimental units.
- In this lecture we focus on **linear** mixed models for continuous (approximately Gaussian) responses.

Common random-effects structures in LMMs

- **Random-intercept models:** each cluster (subject, school, block, . . .) has its own intercept drawn from a population distribution; this adds an extra source of between-cluster variation.
- **Random-intercept and random-slope (random-coefficient) models:** both intercepts and selected regression coefficients (e.g. slopes over time) vary between clusters; this induces correlation and flexible covariance structures for repeated measurements.

Mathematical Definition of the General Framework

A Linear Mixed Model (LMM) is defined as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon$$

Key Components:

- $\mathbf{y}$: $N \times 1$ vector of observed responses.
- $\mathbf{X}$: $N \times p$ design matrix for **Fixed Effects**.
- β: $p \times 1$ vector of fixed effect coefficients (unknown constants).
- $\mathbf{Z}$: $N \times q$ design matrix for **Random Effects**.
- $\mathbf{u}$: $q \times 1$ vector of random effects (random variables).
- ε: $N \times 1$ vector of residuals.

Assumptions:

$$\begin{pmatrix} \mathbf{u} \\ \varepsilon \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix} \right)$$

Here, $\mathbf{G}$ is the variance-covariance matrix of random effects, and $\mathbf{R}$ is the variance-covariance matrix of errors (often $\sigma^2\mathbf{I}$).

Mixed-effects models

Mixed-effects models describe the relationship between a response variable and one or more covariates when observations are grouped or measured repeatedly.

- **Fixed effects:** coefficients associated with levels or values of covariates for which we want to estimate specific effects (e.g. treatment groups, dose levels, time trend). Inference is about these particular levels or about prespecified contrasts between them.
- **Random effects:** terms associated with experimental units sampled from a larger population (e.g. subjects, clusters, blocks, schools). They are modelled as random variables drawn from a probability distribution, and inference concerns the variability between units and the population distribution of these effects.
- **Mixed-effects models:** models that contain both fixed effects and random effects, allowing us to estimate treatment effects while properly accounting for between-unit variation and within-unit correlation.

Introduction to Linear Mixed Models

Why use mixed-effects models?

- To obtain valid standard errors and tests when observations within clusters or subjects are correlated.
- To get more appropriate and efficient estimates of fixed effects by correctly partitioning variability into between- and within-cluster components.
- To broaden inference from the observed units (subjects, schools, laboratories, . . .) to a wider population.
- To naturally handle unbalanced designs and certain types of missing data within a likelihood-based framework.
- To flexibly model correlation structures in longitudinal and repeated-measures data.
- To accommodate heteroscedasticity between treatment groups or over time via random effects and/or structured residual covariance.
- To analyse complex designs (nested, crossed, split-plot, . . .) in a unified regression framework.

Why is the terminology so varied?

Different names arose partly because similar models were developed independently in many fields (psychology, education, agronomy, biostatistics, economics, ...), each with its own notation and emphasis.

You may encounter, for essentially the same class of models:

- Linear mixed-effects models (LMM)
- Hierarchical linear models
- Multilevel models
- Variance-component models
- Random-effects models
- Random-coefficients (random-slope) regression models
- ...

In most applications these terms refer to closely related or equivalent models; the differences are mainly in perspective and terminology rather than in the underlying mathematics.

Fixed vs. random factors difference

Comparison of Fixed and Mixed model

	Fixed Model	Random Model
Levels	Given number of possibilities	Selected at random from a population
New experiment	Use same levels	Use different levels from same population
Goal	Estimates means of fixed levels	Estimate variance of population of means
Inference	$H_0: \ \mu_1 = \mu_2 = \ldots = \mu_k$ $H_1: \ \exists i, j \in \{1, \ldots, k\} \ \mu_i \neq \mu_j$	$H_0: \ \sigma^2_\mu = 0$ $H_1: \sigma^2_\mu > 0$

Mathematical Consequence: Shrinkage Estimation

Why not just estimate everything as fixed effects?

Fixed Effects Estimation (OLS):

$$\hat{\mu}_i = \bar{y}_i$$

Each group mean is estimated solely from its own data. This is unbiased but has high variance for small groups.

Random Effects Estimation (BLUP/Empirical Bayes):

$$\tilde{\mu}_i = w\bar{y}_i + (1 - w)\mu_{global}$$

where $w = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2/n_i}$ is the "shrinkage factor".

Interpretation:

- If n_i is small (little data) or σ^2 is large (noise), $w \to 0$: We "shrink" the estimate towards the global mean.
- If n_i is large (lots of data), $w \to 1$: We trust the group mean.
- This **borrowing of strength** is the mathematical power of mixed models.

Packages for Linear Mixed Effects models in R

There are several packages in R that deal with linear mixed models, most widely used are:

- **nlme** - Non-Linear Mixed Effects, Part of R, `library(nlme)`
  Fit only Gaussian outcomes, it is possible to specify the variance-covariance matrix for the random effects.
- **lme4** - Linear Mixed Effects, `library(lme4)`
  Can be used to fit generalized mixed-effects regression models, it is not possible to specify the variance-covariance matrix for the random effects, but can handle with diagonal covariance structures or unstructured covariance matrices.
- **Other useful packages:** `merTools, glmmTMB, brms, modelr`

All of them has some advantages and some disadvantages. Notations, specifications, special functions and classes are different.
For power analysis for random effects in mixed models is possible to use package called **pamm**.

Packages for Linear Mixed Effects models in Python

Statsmodels supports some of the Linear Mixed Effects Models
Statsmodels Linear Mixed Effects Models

- Random intercepts models
- Random slopes models
- Variance components models

Connection to R Lme4 **Statsmodels and lme4**

Python Implementation Example (statsmodels)

Using `statsmodels.formula.api` to fit LMMs similar to R's `lmer`.

```
import statsmodels.formula.api as smf
import pandas as pd

# Load data
data = pd.read_csv("oats.csv")

# Model Definition
# Fixed: yield ~ variety * nitrogen
# Random: random intercept for block (groups='block')
model = smf.mixedlm("yield ~ variety * nitrogen",
                    data,
                    groups=data["block"])

# Fitting (REML is default)
result = model.fit()

# Output
print(result.summary())
```

Simple example:

```
library(nlme)
library(lme4)
library(MASS)

data(oats)
names(oats)   = c('block', 'variety', 'nitrogen', 'yield')
oats$mainplot = oats$variety
oats$subplot  = oats$nitrogen
attach(oats)

# SIMPLE MIXED EFFECTS MODEL
model1a=lme( yield~variety*nitrogen, random = ~ 1|block)
model2a=lmer(yield~variety*nitrogen + (1|block))

# NESTED MIXED EFFECTS MODEL (will be covered next lecture)
model1b=lme( yield~variety*nitrogen, random=~1|block/mainplot)
model2b=lmer(yield~variety*nitrogen + (1|block/mainplot))
```

Simple example - comparison of classical and Mixed Models approach

Description of data set:

In a pharmaceutical company the use of NIR (Near Infrared Reflectance) spectroscopy was investigated as an alternative to the more cumbersome (and expensive) HPLC method to determine the content of active substance in tablets.

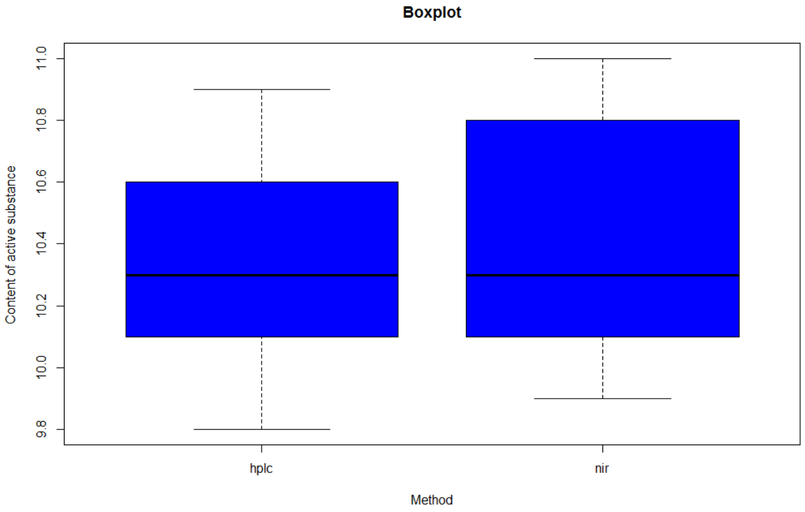
Source of data set:

Brockhoff and Thierry-Carstensen 2003, Test set validation using simple statistical methods.

	hplc	nir	difference
Tablet_1	10.4	10.1	0.3
Tablet_2	10.6	10.8	-0.2
Tablet_3	10.2	10.2	0.0
Tablet_4	10.1	9.9	0.2
Tablet_5	10.3	11.0	-0.7
Tablet_6	10.7	10.5	0.2
Tablet_7	10.3	10.2	0.1
Tablet_8	10.9	10.9	0.0
Tablet_9	10.1	10.4	-0.3
Tablet_10	9.8	9.9	-0.1

The aim is to study the method differences.

Simple example - comparison of classical and Mixed Models approach



Simple analysis of the pharmaceutical data by the paired t-test

```
> mean(d)  -0.05
> var(d)    0.08722222
> sd(d)     0.2953341
t.test(d)
# alternatively:  t.test(hplc, nir, paired = TRUE)
Paired t-test
data:  hplc and nir
t = -0.5354, df = 9, p-value = 0.6054
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:  -0.2612693  0.1612693
sample estimates: mean of the differences -0.05
```

The standard error of the mean $\bar{d}$ (ie. the uncertainty of the estimated difference) :
$SE_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{0.295}{\sqrt{10}} = 0.0934$

t-statistics: $t = \frac{\bar{d}}{SE_{\bar{d}}} = \frac{-0.05}{0.0934} = -0.5354$

Final regression model: $d_i = \mu + \varepsilon_i \quad \varepsilon_i \sim N(o, \sigma^2)$

Estimated model parameters: $\hat{\mu} = \bar{d}, \; \hat{\sigma} = s_d$

Simple example - comparison of classical and Mixed Models approach

Simple analysis of the pharmaceutical data by the ANOVA

We have randomized balanced design with two treatments - methods and 10 blocks - tablets.

Used regression model: $y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$

```
summary(aov(y ~ method+tablet))

          Df Sum Sq Mean Sq F value Pr(>F)
method     1 0.0125 0.01250   0.287 0.6054
tablet     9 2.0005 0.22228   5.097 0.0118
Residuals  9 0.3925 0.04361
```

The uncertainty of the average method difference:

$$SE_{\bar{y}_2 - \bar{y}_1} = \sqrt{MS_E \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{0.0436 (\frac{1}{10} + \frac{1}{10})} = 0.0934$$

We have the same p-value as in the t-test approach and the uncertainty since the t statistic for testing $H_0 : \mu_1 = \mu_2$ is

$$t = \frac{\bar{y}_2 - \bar{y}_1}{\sqrt{MS_E \left(\frac{1}{n_1} + \frac{1}{n_j} \right)}}.$$

Misleading information from the classical ANOVA approach

The ANOVA approach compare treatments means and is based on knowledge obtained only from certain measurements:

In pharmaceutical data example, the **ANOVA approach** is valid only for statements about the 10 specific tablets in the experiment, not for tablets in general.

The uncertainty of the average NIR value is given by using only the 10 NIR measurements: $SE_{\bar{y}_1} = \frac{s_1}{\sqrt{10}} = 0.127$, $s_1 = sd(NIR) = 0.4012$.

On the other hand, the **Mixed Models approach** is valid for tablets (blocks) in general, since we consider the 10 tablets as a random sample.

Let us consider Mixed Model with tablet as a random effects:

$$y_{ij} = \mu + a_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2),\ a_i \sim N(0, \sigma_T^2).$$

Simple example - comparison of classical and Mixed Models approach

Comparison of Fixed and Mixed model

The observations in Mixed Model are no longer independent and the tablet differences become a part of the variance structure.

1. The expected value of the observation y_{ij}.
2. The variance of the observation y_{ij}.
3. The relation between two observations.

Comparison table of fixed and mixed model

		Fixed Model	Mixed Model
1	$E\left[y_{ij}\right]$	$\mu + \alpha_i + \beta_j$	$\mu + \beta_j$
2	$var\left[y_{ij}\right]$	σ^2	$\sigma_T^2 + \sigma^2$
3	$cov(y_{i_1 j_1}, y_{i_2 j_2})\ j_1 \neq j_2$	0	σ_T^2 (if $i_1 = i_2$), 0(if $i_1 \neq i_2$)

Simple example - comparison of classical and Mixed Models approach

Analysis of the pharmaceutical data by the Mixed Models

Used mixed model:

$$y_{ij} = \mu + a_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2),\ a_i \sim N(0, \sigma_T^2).$$

```
# ways how to estimate mixed models
# function lme  (need library nlme)
# function lmer (need library lme4)
> model<-lme( y~method, random = ~1|tablet,     data=temp)
> model<-lme( y~method, random=list(tablet=~1), data=temp)
> model<-lmer(y~method+(1|tablet),              data=temp)
> summary(model)
```

Simple example - comparison of classical and Mixed Models approach

Analysis of the pharmaceutical data by the Mixed Models

Used mixed model: $y_{ij} = \mu + a_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2), \; a_i \sim N(0, \sigma_T^2).$

```
> summary(model)
Linear mixed model fit by REML
Formula: y ~ method + (1 | tablet)
Random effects:
 Groups   Name        Variance Std.Dev.
 tablet   (Intercept) 0.089333 0.29889
 Residual             0.043611 0.20883
Number of obs: 20, groups: tablet, 10

Fixed effects:
            Estimate Std.
 Error t value
(Intercept) 10.39000    0.11530   90.11
methodhplc  -0.05000    0.09339   -0.54
Correlation of Fixed Effects:
           (Intr)
methodhplc -0.405

anova(model)   numDF denDF  F-value p-value
(Intercept)        1     9 9666.573  <.0001
method             1     9    0.287  0.6054
```

Simple example - comparison of classical and Mixed Models approach

Analysis of the pharmaceutical data by the Mixed Models

Used mixed model:

$$y_{ij} = \mu + a_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2),\ a_i \sim N(0, \sigma_T^2),$$

with estimated Random effects components:

```
 Groups   Name        Variance Std.Dev.
 tablet   (Intercept) 0.089333 0.29889
 Residual             0.043611 0.20883
```

$$\hat{\sigma}^2 = 0.043611, \quad \hat{\sigma}_T^2 = 0.089333,$$

and Fixed effects components:

```
            Estimate Std.
 Error t value
(Intercept) 10.39000    0.11530   90.11
methodhplc  -0.05000    0.09339   -0.54
```

$$\hat{\mu} = 10.39000, \quad \hat{\beta}_1 = 0, \quad \hat{\beta}_2 = -0.05000$$

Example with missing values

We have the same pharmaceutical data set, but we add 5 non-paired measuremnt from each method.

n	tablet	method	y	n	tablet	method	y	n	tablet	method	y
1	1	hplc	10.4	14	7	nir	10.2	27	14	hplc	NA
2	1	nir	10.1	15	8	hplc	10.9	28	14	nir	10.3
3	2	hplc	10.6	16	8	nir	10.9	29	15	hplc	NA
4	2	nir	10.8	17	9	hplc	10.1	30	15	nir	9.7
5	3	hplc	10.2	18	9	nir	10.4	31	16	hplc	10.3
6	3	nir	10.2	19	10	hplc	9.8	32	16	nir	NA
7	4	hplc	10.1	20	10	nir	9.9	33	17	hplc	9.6
8	4	nir	9.9	21	11	hplc	NA	34	17	nir	NA
9	5	hplc	10.3	22	11	nir	10.8	35	18	hplc	10.0
10	5	nir	11.0	23	12	hplc	NA	36	18	nir	NA
11	6	hplc	10.7	24	12	nir	9.8	37	19	hplc	10.2
12	6	nir	10.5	25	13	hplc	NA	38	19	nir	NA
13	7	hplc	10.3	26	13	nir	10.5	39	20	hplc	9.9
								40	20	nir	NA

The mixed model in a direct way may give information about the key issues in a data set, that a straightforward fixed ANOVA does not.

Example with missing values

Analysis by fixed effects ANOVA

The fixed effect analysis only uses the information in the first 10 tablets.

```
> model1<-lm(y~tablet+method)
> anova(model1)
Analysis of Variance Table

Response: y
          Df Sum Sq  Mean Sq F value  Pr(>F)
tablet    19 3.7230 0.195947  4.4931 0.01288 *
method     1 0.0125 0.012500  0.2866 0.60537
Residuals  9 0.3925 0.043611
```

Note that only the Tablets row of the table has changed compared to the previous analysis.

Example with missing values

Analysis by mixed model

```
> model2<-lme(y~method,random=~1| tablet,data=temp2t)
> anova(model2)
            numDF denDF   F-value p-value
(Intercept)     1    19 15715.981  <.0001
method          1     9     0.687  0.4285
> summary(model2)
Linear mixed-effects model fit by REML
Random effects:
 Formula: ~1 |
 tablet
        (Intercept)  Residual
StdDev:   0.3192429 0.2085067
Fixed effects: y ~ method
                Value  Std.Error DF   t-value p-value
(Intercept) 10.283857 0.09259174 19 111.06668  0.0000
methodhplc  -0.072111 0.08697180  9  -0.82913  0.4285
Correlation:
           (Intr)
methodhplc -0.47
```

Example with missing values

Compariosn of Fixed and Mixed model approach

We use the `R` function `estimable`.

FIXED:

```
       Estimate Std.
 Error    t value DF  Pr(>|t|)   Lower.CI   Upper.CI
LS HPLC  10.2125 0.06177356 165.321550  9 0.0000000 10.0727585 10.3522415
LS NIR   10.2625 0.06177356 166.130957  9 0.0000000 10.1227585 10.4022415
LS DIF   -0.0500 0.09339284  -0.535373  9 0.6053664 -0.2612693  0.1612693
```

MIXED:

```
          Estimate Std.
 Error      t value DF  Pr(>|t|)   Lower.CI   Upper.CI
LS HPLC 10.21174590 0.09259174 110.2878719  9 0.0000000 10.0022888 10.4212030
LS NIR  10.28385689 0.09259174 111.0666778 19 0.0000000 10.0900602 10.4776536
LS DIF  -0.07211099 0.08697180  -0.8291307  9 0.4284719 -0.2688549  0.1246329
```

Note that apart from giving a slightly different value, the Mixed model estimation is also more precise, than the one only based on tablets 1-10. This is the kind of analysis that the mixed model for this situation leads to.

Simple Introduction to Theory of Mixed Models

Recall simple linear regression model of **fixed effects** approach:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2),\ i \in \{1, 2\}\,,\ j \in \{1, 2, 3\}\,,$$

$$\underbrace{\begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \cdot \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} e_{11} \\ e_{21} \\ e_{12} \\ e_{22} \\ e_{13} \\ e_{23} \end{pmatrix}}_{\mathbf{e}} .$$

equivalently,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e},$$

where **y** is an vector of all observations - response variables, **X** is an known matrix of predictors (usually called design matrix), β is a vector of unknown coefficients - fixed effects parameters and **e** is an vector of unknown independent measurement errors $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

Regression model of **Mixed linear model** approach:

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij} \quad b_j \sim N(0, \sigma_B^2),\ \varepsilon_{ij} \sim N(0, \sigma^2) \quad i \in \{1,2\}\,,\ j \in \{1,2,3\}\,,$$

$$\underbrace{\begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \cdot \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{Z}} \cdot \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}}_{u} + \underbrace{\begin{pmatrix} e_{11} \\ e_{21} \\ e_{12} \\ e_{22} \\ e_{13} \\ e_{23} \end{pmatrix}}_{\mathbf{e}},$$

equivalently,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}u + \mathbf{e},$$

where **y** is an vector of all observations - response variables, **X** is an known matrix of predictors (usually called design matrix), β^0 is a vector of unknown coefficients - fixed effects parameters, **Z** is the design matrix for random effects, **u** is the vector of random effects $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, $cov(u_i, u_j) = G_{ij}$ and **e** is an vector of unknown independent measurement errors $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, $cov(e_i, e_j) = R_{ij}$, typically **R** is diagonal.

Conditional vs. Marginal Moments

It is crucial to distinguish between the conditional distribution (given random effects) and the marginal distribution (unconditional).

1. Conditional on u:

$$E[\mathbf{y}|\mathbf{u}] = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$$

$$Var(\mathbf{y}|\mathbf{u}) = \mathbf{R}$$

2. Marginal (integrating out u): Since $E[\mathbf{u}] = \mathbf{0}$,

$$E[\mathbf{y}] = E[E[\mathbf{y}|\mathbf{u}]] = \mathbf{X}\beta$$

And using the law of total variance:

$$Var(\mathbf{y}) = E[Var(\mathbf{y}|\mathbf{u})] + Var(E[\mathbf{y}|\mathbf{u}])$$

$$Var(\mathbf{y}) = E[\mathbf{R}] + Var(\mathbf{X}\beta + \mathbf{Z}\mathbf{u})$$

$$\mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^T$$

This matrix $\mathbf{V}$ captures the correlation structure induced by the shared random effects.

Simple Introduction to Theory of Mixed Models

Distribution of response variable **y** in Mixed Models:

Let us consider linear mixed model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}u + \mathbf{e}, \quad \text{where } \mathbf{u} \sim N(\mathbf{0}, \mathbf{G}),\ \mathbf{e} \sim N(\mathbf{0}, \mathbf{R}).$$

Distribution of **y** is multivariate normal

$$\mathbf{y} \sim N(\mu, \mathbf{V})$$

with

$$\mu = E\left[\mathbf{X}\beta + \mathbf{Z}u + \mathbf{e}\right] = \mathbf{X}\beta$$

and

$$\mathbf{V} = var\left[\mathbf{X}\beta + \mathbf{Z}u + \mathbf{e}\right] = var\left[\mathbf{Z}u\right] + var\left[\mathbf{e}\right] = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}.$$

$$\mathbf{y} \sim N(\mathbf{X}\beta, \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R})$$

Notice that if **R** is diagonal and we have random block effect model then **V** is block diagonal matrix.

Simple Introduction to Theory of Mixed Models

Example of a **y** distribution for simple linear Mixed Effects Models:

Let us consider linear mixed effect model:

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij} \quad b_j \sim N(0, \sigma_B^2),\ \varepsilon_{ij} \sim N(0, \sigma^2) \quad i \in \{1, 2\}\,,\ j \in \{1, 2, 3\}\,,$$

$$\mu = \begin{pmatrix} \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_1 \\ \mu + \alpha_2 \end{pmatrix} \quad \mathbf{V} = \begin{pmatrix} \sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & 0 & 0 & 0 \\ \sigma_B^2 & \sigma^2 + \sigma_B^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_B^2 & \sigma^2 + \sigma_B^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 + \sigma_B^2 & \sigma_B^2 \\ 0 & 0 & 0 & 0 & \sigma_B^2 & \sigma^2 + \sigma_B^2 \end{pmatrix}$$

Notice that two observations from the same block are correlated !!

The likelihood function L for mixed effect models:

For given parameter values, the likelihood function L returns a measure of the probability of observing response variables $\mathbf{y}$.

The negative log likelihood function for a mixed model is given by:

$$l(\mathbf{y},\beta,\gamma) = \frac{1}{2}\left(n\ln(2\pi) + \ln|\mathbf{V}(\gamma)| + (\mathbf{y}-\mathbf{X}\beta)^{\mathbf{T}}\mathbf{V}(\gamma)^{-\mathbf{1}}(\mathbf{y}-\mathbf{X}\beta)\right)$$

In the simple one way ANOVA with random block effect model is

$$\gamma = (\sigma^2, \sigma_b^2)^T \text{ and } \beta = (\alpha_1, \alpha_2)^T$$

The maximum likelihood estimation is given by

$$(\hat{\beta}^{(ML)}, \hat{\gamma}^{(ML)}) = \text{argmin}_{(\beta,\gamma)} l(\mathbf{y},\beta,\gamma)$$

Assume γ is known then ML estimation of $\hat{\beta}(\gamma)$ is given by weighted least squares estimation

$$\hat{\beta}^{ML}(\gamma) = \text{argmin}_{(\beta)}(\mathbf{y}-\mathbf{X}\beta)^{\mathbf{T}}\mathbf{V}(\gamma)^{-\mathbf{1}}(\mathbf{y}-\mathbf{X}\beta),$$

by differentiate and equal to zero we obtain:

$$\hat{\beta}^{ML}(\gamma) = (\mathbf{X}^{\mathbf{T}}\mathbf{V}(\gamma)^{-\mathbf{1}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\mathbf{V}(\gamma)^{-\mathbf{1}}\mathbf{y}$$

Henderson's Mixed Model Equations (MME)

How do we estimate β and predict $\mathbf{u}$ simultaneously? Henderson (1950) maximized the joint density of $\mathbf{y}$ and $\mathbf{u}$, leading to this system of equations:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Solution properties:

- $\hat{\beta}$ is the Best Linear Unbiased Estimator (BLUE).
- $\hat{\mathbf{u}}$ is the Best Linear Unbiased Predictor (BLUP).
- The term $\mathbf{G}^{-1}$ acts as a "penalty" or regularization term (Ridge Regression analogy), shrinking random effects toward zero.

Theory of Mixed Models

The restricted likelihood method for mixed effect models:

Since the maximum likelihood estimation is biased, we want to modify it to obtain unbiased estimator.

Idea of Restricted (residual) maximum likelihood (REML) is in linear transform of data which eliminates mean.

REML method is given by modification of classical ML function by

$$l_{REML}(\mathbf{y}, \beta, \gamma) = \frac{1}{2}\left(n \ln(2\pi) + \ln|\mathbf{V}(\gamma)| + (\mathbf{y} - \mathbf{X}\beta)^{\mathbf{T}}\mathbf{V}^{-1}(\gamma)(\mathbf{y} - \mathbf{X}\beta) + \ln\left|\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}(\gamma)\mathbf{X}\right|\right).$$

This REML method gives (at least in balanced case) the unbiased estimates and is generally preferred in mixed models.

Mathematical Derivation of REML

REML works by maximizing the likelihood of "error contrasts" rather than the data $\mathbf{y}$ itself.

Let $\mathbf{K}$ be an $N \times (N - p)$ matrix such that:

$$\mathbf{K}^T \mathbf{X} = \mathbf{0} \quad \text{(Orthogonal to fixed effects)}$$

We transform the data: $\mathbf{y}^* = \mathbf{K}^T \mathbf{y}$.

The distribution of $\mathbf{y}^*$ does not depend on $\boldsymbol{\beta}$:

$$\mathbf{y}^* = \mathbf{K}^T (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}) = \mathbf{0} + \mathbf{K}^T \mathbf{Z}\mathbf{u} + \mathbf{K}^T \boldsymbol{\varepsilon}$$

$$\mathbf{y}^* \sim N(\mathbf{0}, \mathbf{K}^T \mathbf{V} \mathbf{K})$$

Maximizing the likelihood of $\mathbf{y}^*$ gives estimates of variance components $(\mathbf{G}, \mathbf{R})$ that are unbiased because the degrees of freedom used for estimating $\boldsymbol{\beta}$ are accounted for by the projection $\mathbf{K}$.

Theory of Mixed Models

Restricted maximum likelihood (REML)

- The default parameter estimation criterion for linear mixed models in `lme` and `lmer` functions is REML.
- Maximum likelihood (ML) estimates (sometimes called full maximum likelihood) can be requested by specifying `REML=FALSE`.
- Generally REML estimates of variance components are preferred. ML estimates are known to be biased. Although REML estimates are not guaranteed to be unbiased, they are usually less biased than ML estimates.

For more info and REML proof see:

- **lmer vignettes**
- **proof by Kevin Liu**

Summary Key Takeaways

1. **Generalization:** Mixed models allow inference to a larger population of units (e.g., all tablets, not just the 10 observed).
2. **Correlation:** They explicitly model the correlation structure ($\mathbf{V}$) arising from repeated measures or clustering.
3. **Shrinkage:** Random effect estimates (BLUPs) are "shrunk" towards the population mean, providing protection against overfitting small groups.
4. **Estimation:** REML is preferred for unbiased variance estimation; Henderson's MME provides a unified way to solve for β and $\mathbf{u}$.

Example of Mixed effect model analysis

Drying of beech wood planks

To investigate the effect of drying of beech wood on the humidity percentage, the following experiment was conducted. Each of 20 planks was dried in a certain period of time. Then the humidity percentage was measured in 5 depths (1,3,5,7,9) and 3 widths (1,2,3) for each plank.

Variables:

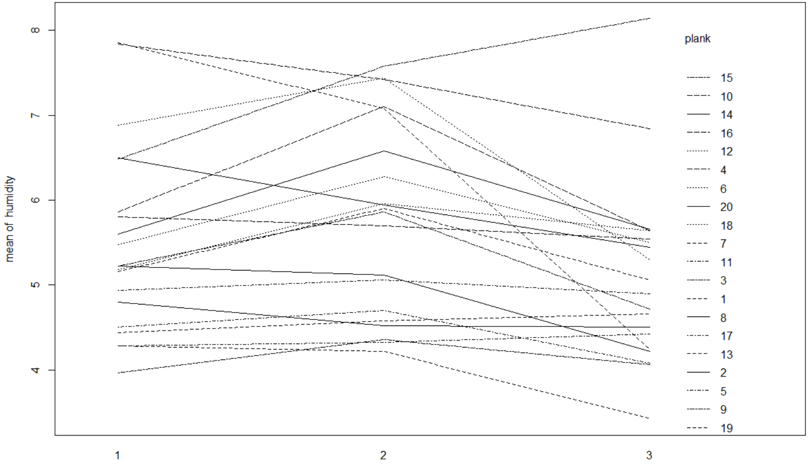
- plank - Numbered 1-20
- width - Numbered 1,2,3
- depth - Numbered 1,3,5,7,9
- humidity - Humidity percentage

Source: The Royal Veterinary and Agricultural University, Denmark.
Number of observations: 300 (20 planks)

```
depth 1:  close to the top
depth 5:  in the center
depth 9:  close to the bottom
depth 3:  between 1 and 5
depth 7:  between 5 and 9
width 1:  close to the side
width 3:  in the center
width 2:  between 1 and
 3
```

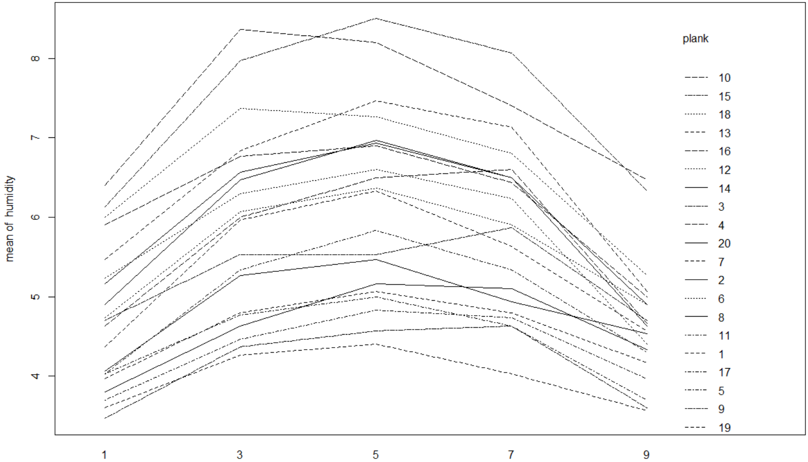
Exercise - Drying of beech wood planks

```
with(planks, interaction.plot(width,plank,humidity,legend=T))
```



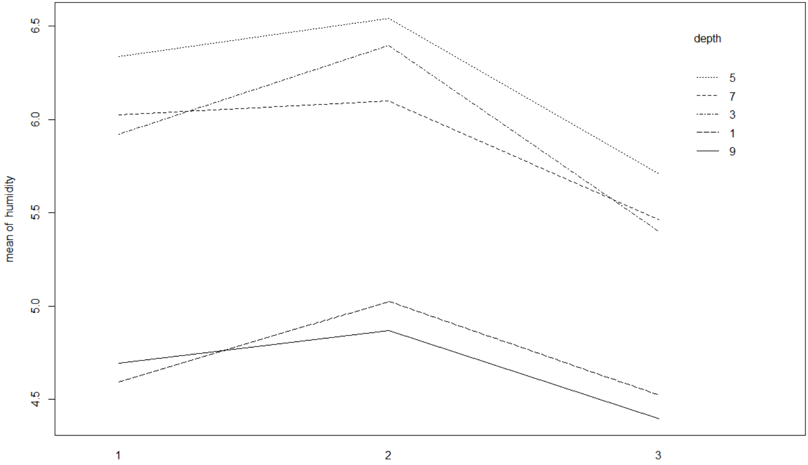
Exercise - Drying of beech wood planks

```
with(planks, interaction.plot(depth,plank,humidity,legend=T))
```



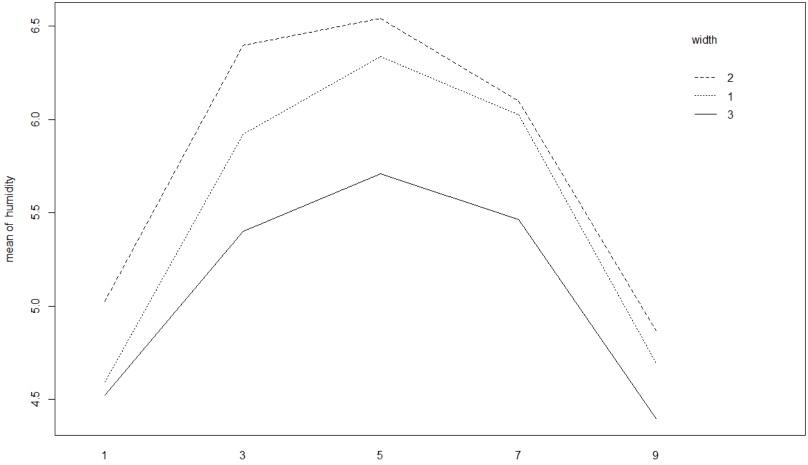
Exercise - Drying of beech wood planks

```
with(planks, interaction.plot(width,depth,humidity,legend=T))
```



Exercise - Drying of beech wood planks

```
with(planks, interaction.plot(depth,width,humidity,legend=T))
```



Exercise

Analyze data from the **Drying of beech wood planks** experiment.

- Plot four average humidity profiles:
  2 interaction plots for width and 2 for depth (done).
- Carrying out the fixed effects model analysis.
- Carry out the mixed model analysis.
- Run the post hoc analysis
- Compare the fixed parameters and use the p-value correction (TukeyHSD).
- Summarize results.