Modelos Matemáticos e Aplicações

Modelos lineares mistos

Linear mixed models

Elsa Gonçalves
ISA/UL

2022-2023

Linear Mixed Models

Summary:

1. Motivation: some typical examples involving mixed models

2. The general linear mixed model: description and properties; estimation of covariance parameters; estimation of fixed effects and prediction of random effects; hypotheses tests for covariance parameters, fixed and random effects; model selection (model comparison via likelihood ratio tests and via information criteria); validation of model assumptions

3. Some particular cases and applications: exercises

Fitting linear mixed models : *Packages usually used*

➢ *Package* **nlme**, ⓡ (*http://www.R-project.org*)

➢ *Package* **lme4**, ⓡ (*http://www.R-project.org*)

➢ Package **lmerTest,** ⓡ (*http://www.R-project.org*), provides tests in Linear Mixed Effects Models fitted with *Package* **lme4**

➢ **Proc mixed** do SAS (*SAS Institute, Inc.*)

➢ *Package* **ASREML-R,** ⓡ (*VSN International*) (mainly focused on aninal and plant breeding)

➢ *Package* **rrBLUP,** ⓡ (*http://www.R-project.org*) (focused on genomic selection)

➢ *Package* ***sommer ,*** ⓡ (*http://www.R-project.org*) focused on **GWAS** (genome-wide association), genomic selection

Bibliography

- Demidenko, E. (2013). *Mixed Models. Theory and Applications with R*. John Wiley & Sons, Inc., New Jersey.

- Fitzmaurice, G.M., Laird, N.M:, Ware, J.H. (2004). *Applied Longitudinal Analysis*. John Wiley & Sons, Inc.

- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., Schabenberger, O. (2006). *SAS system for mixed models*. 2nd ed.. SAS Institute, Cary, NC.

- McCulloch, C.E., Searle, S.R. (2001). *Generalized linear and mixed models*. John Wiley & Sons, New York.

- Pinheiro, J.C., Bates, D.M. (2000). *Mixed-effects models in S and S-plus*. Springer-Verlag, New York.

- Searle, S.R., Casella, G., McCulloch, C.E. (1992). *Variance components*. John Wiley & Sons, New York.

- Stroup, W.W. (2013). *Generalized linear mixed models. Modern concepts, methods and applications*. CRC Press Taylor & Francis Group, Boca Raton, FL, USA. 529 pp.

- Verbeke, G., Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer-Verlag, Springer Series in Statistics, New York.

- Zuur, A., Ieno E., · Walker, N. , Saveliev, A. , Smith, G. (2009). Mixed Effects Models and Extensions in Ecology with R. Springer, New York.

What is a mixed model?

- Models in which the only effects are fixed effects are called fixed effects models.

- **Models that contain both fixed and random effects are called mixed models.**

- A special case of a mixed model is when there are no fixed effects (except a general mean common to all observations, μ), that is, a model having only random effects. It is called a **random model.**

What are random effects?

When the levels of the factor have been selected at random from a population of possible levels and we want to obtain information about the parameters of the distribution of those levels.

The main goals of the analysis of random effects models are:
➢ Estimate covariance parameters
➢ Test hypotheses about the parameters or functions of the parameters
➢ Calculate predictors (BLUP) of the realized values of the random effects

Motivation

Some typical examples involving mixed models
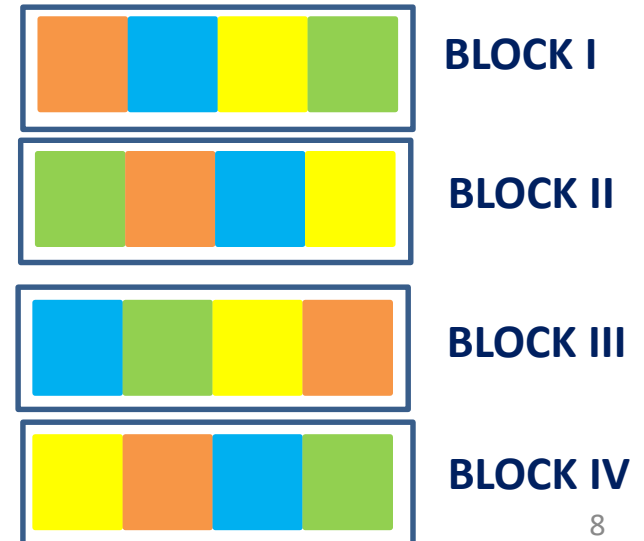
The Randomized Complete Block design (RCB)

The randomized block design is just the begining with mixed models.

Blocking is a research technique that is used to reduce the effects of variation among experimental units. The units can be people, plants, animals, etc..

> ➢ The RCB is the standard design for agricultural experiments.
> ➢ The field is divided into units to account for any variation in the field (by accounting for spatial effects). Blocks are groups of units that are formed so that units within the blocks are as nearly homogeneous as possible.
> ➢ The levels of the factor being investigated, called treatments, are randomly assigned to units within the blocks (each treatment once per block).
> ➢ The number of blocks is the number of replications.
> ➢ Any treatment can be adjacent to any other treatment, but not to the same treatment within the block.

Example: each row represents a block. There are 4 blocks (I-IV) and 4 treatments (different colors)

BLOCK I

BLOCK II

BLOCK III

BLOCK IV

The Randomized Complete Block (RCB) design

- Usually, the primary objectives are to estimate and compare treatment means. In most cases, the treatment effects are considered fixed because the treatments in the experiment are the only ones to which inference is to be made.

- Block effects are usually considered random because the blocks in the experiment constitute only a small subset of the larger set of blocks over which inferences about treatment means are to be made.

The model for data from a randomized blocks design usually contains a fixed effects factor and random effects for blocks, making it a mixed model

Designs belonging to the family of incomplete block designs

When the number of treatments used becomes large, field trials are laid out on a rectangular grid of plots.

Under these conditions, experimental designs with incomplete but smaller blocks are preferred, such as alpha designs, which constitute a particular class of generalized lattice designs.

Incomplete blocks nested within a complete block

Incomplete block

Complete block

Incomplete block (column)

Incomplete block (row)

Complete block
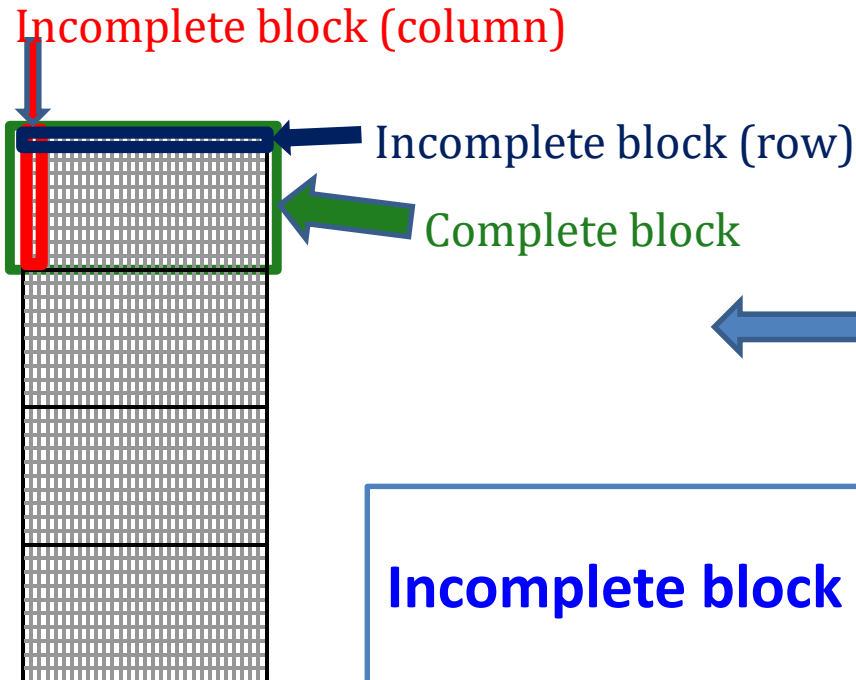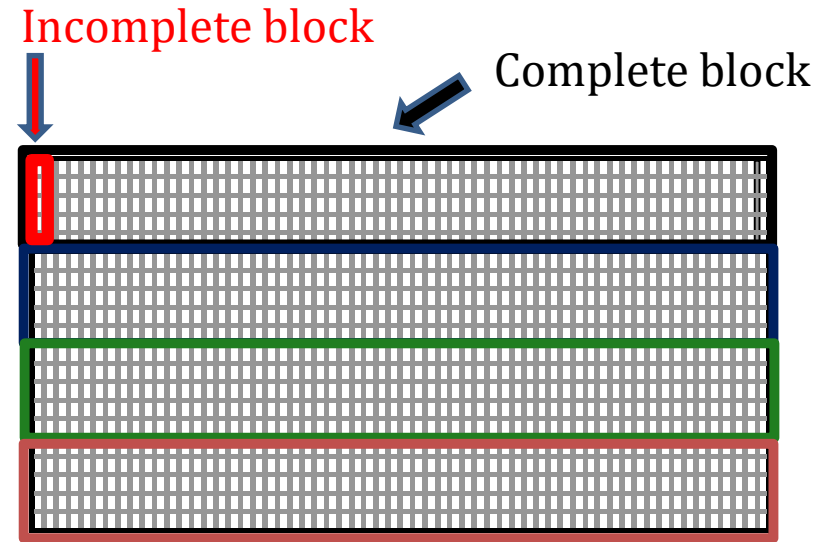
Alternatively, to control for environmental gradients in both row and column directions by blocking, row-column designs with blocks in both rows and columns are used, which correspond to groups of more complex Latin square designs.
Rows nested within a complete block
Columns nested within a complete block

Incomplete block effects are considered random

RCB repeated at more than one location

• Blocks are laid out at more than one location. Treatments are assigned at random to those blocks.

• Treatments are assigned at random within blocks, each treatment once per block.

• The number of blocks is the number of replications.

• Any treatment can be adjacent to any other treatment, but not to the same treatment within the block.

➢ Usually, the treatment effects are considered fixed.

➢ Usually, location effects are considered random because the locations constitute only a small subset of the larger set of locations over which inferences about treatment means are to be made.

➢ Block effects nested within location are usually considered random.

The model for data from this type of design contains a fixed effects factor and random effects factors,

it is a mixed model

Example: there are 4 treatments (different colors), 4 locations and 4 blocks per location

Location I
BLOCK I
BLOCK II
BLOCK III
BLOCK IV

Location II
BLOCK I
BLOCK II
BLOCK III
BLOCK IV

Location III
BLOCK I
BLOCK II
BLOCK III
BLOCK IV

Location IV
BLOCK I
BLOCK II
BLOCK III
BLOCK IV

11
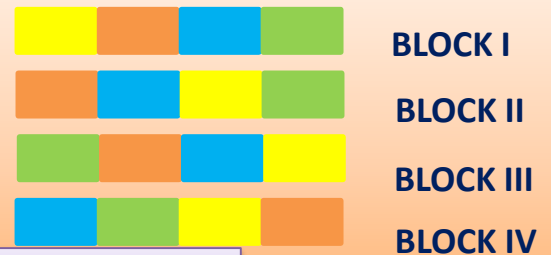
The split-plot design

- The split-plot design has an agricultural heritage, with the whole plots usually being large areas of land and the subplots being smaller areas of land within the large areas.

Example: several varieties of a crop could be planted in different fields (whole plots), one variety to a field. Then each field could be divided into several subplots, and each subplot could be treated with a different type of fertilizer.

- Despite its agricultural basis, the split-plot design is useful in many other scientific areas.

The split-plot design on a RCB

• Main treatments (levels of factor A) are assigned at random within blocks, each treatment once per block; they are divided further into additional independent units (subplots) to which another set of treatments (levels of factor B) are randomly assigned.
• The number of blocks is the number of replications.
• Any main treatment can be adjacent to any other treatment, but not to the same treatment within the block.

Example:

Different colors represent different main treatments; each row represents a block. There are 4 blocks (I-IV) each of 4 main treatments (colors) divided into 4 further sub-plot treatments (symbols).

13
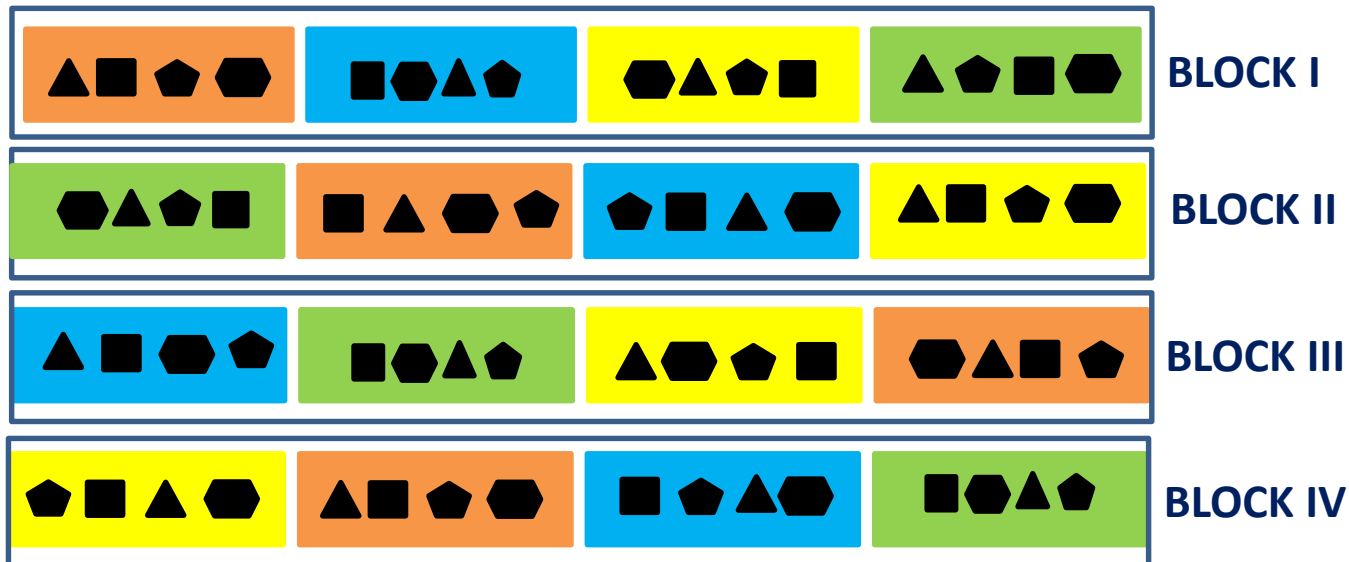
The split-plot design on a RCB

- The effects of factor A, the effects of factor B and interaction A*B can be considered as fixed or random.
- Block effects are usually considered random effects.
- The effects of interaction Block*Factor A (whole plot error) and Block*Factor B are assumed random effects

The model for data from this type of design contains fixed and random effects, it is a mixed model

❑ Random models (a particular case of mixed model) and mixed models are historically applied in plant and animal breeding. In this context, the objectives are focused on the covariance parameters estimates and on the preditors of random effects.

❑ Other numerous experimental designs produce data for which mixed models are appropriate. Some examples are nested designs, repeated measures designs, longitudinal models, etc..

The Linear Mixed Model

In matrix formulation, the linear mixed model can be written as (the bold will be used for matrix notation):

$$Y = X\beta + Zu + e$$

$Y_{(n \times 1)}$ is the vector of observations

$\beta_{(p \times 1)}$ is the vector of fixed effects

$X_{(n \times p)}$ is the model matrix for fixed effects

$u_{(q \times 1)}$ is the vector of random effects

$Z_{(n \times q)}$ is the model matrix for random effects

$e_{(n \times 1)}$ is a vector random errors

The predictors can be factors or numeric variables.

Example: model with one fixed effects factor and
one random effects factor (when X and Z are incidence matrices)

$$
\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{u} + \mathbf{e}
$$

$$
\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1q} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2q} \\ \vdots \\ y_{p1} \\ y_{p2} \\ \vdots \\ y_{pq} \end{bmatrix}
=
\begin{bmatrix}
1 & 1 & 0 & \cdots & 0 \\
1 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & 0 & \cdots & 0 \\
1 & 0 & 1 & \cdots & 0 \\
1 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 0 & \cdots & 1 \\
1 & 0 & 0 & \cdots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}
+
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_q \end{bmatrix}
+
\begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1q} \\ e_{21} \\ e_{22} \\ \vdots \\ e_{2q} \\ \vdots \\ e_{p1} \\ e_{p2} \\ \vdots \\ e_{pq} \end{bmatrix}
$$

When X is a singular matrix, the procedure is similar to the already described for ANOVA with fixed effects

Linear mixed model

$$Y = X\beta + Zu + e$$

❑ The vectors u and e are assumed mutually independent with multivariate normal distribution with vector of mean values $\mathbf{0}_{(n \times 1)}$ and covariance matrices G and R, respectively:

$$\text{cov}[u, e] = 0$$

$$u \cap \mathcal{N}_q(0, G)$$

$$e \cap \mathcal{N}_n(0, R)$$

where $G_{(q \times q)}$ and $R_{(n \times n)}$ are symmetric and positive-definite matrices

❑ As a consequence,

$$E[Y] = X\beta$$

$$\mathrm{var}[Y] = ZGZ^T + R = V,$$

where Z^T is the transpose of Z.

❑ The distribution of Y is multivariate normal with mean value $X\beta$ and variance-covariance matrix V, symmetric and positive-definite,

$$Y \cap \mathcal{N}_n(X\beta, V).$$

There are many variants for linear mixed models

- **In the classical linear mixed model is assumed that:**

- The elements of the vector $\boldsymbol{u}$ are independent and identically distributed random variables (i.i.d.), that is, the covariance matrix is $\boldsymbol{G} = \sigma_u^2 \boldsymbol{I_q}$, where $\boldsymbol{I_q}$ is the identity matrix $q \times q$;

- The elements of the vector $\boldsymbol{e}$ are i.i.d. random variables, that is, the covariance matrix is $\boldsymbol{R} = \sigma_e^2 \boldsymbol{I_n}$, where $\boldsymbol{I_n}$ is the identity matrix $n \times n$.

- **A particular case is the traditional random model:**

It means that $\boldsymbol{X} = \mathbf{1_n}, \boldsymbol{\beta} = \mu, \boldsymbol{G} = \sigma_u^2 \boldsymbol{I_q}, \boldsymbol{R} = \sigma_e^2 \boldsymbol{I_n}$

Example: the particular case of ANOVA with one factor with random effects, balanced
(G and R diagonal matrices)

$$Y_{ij} = \mu + u_i + e_{ij}$$

for $i = 1, \ldots, a, j = 1, \ldots, b.$

Y_{ij} is the jth observation in the ith level of factor A;

μ is a general mean (population);

u_i is the effect of the level i of the factor A (random effects);

e_{ij} is the random error associated to the observation Y_{ij}.

In the traditional random model it is assumed:

- $u_i \cap \mathcal{N}\left(0, \sigma^2{}_u\right), \forall i$

- $e_{ij} \cap \mathcal{N}\left(0, \sigma^2{}_e\right), \forall ij$

- $\mathrm{cov}\left(e_{ij}, e_{i'j'}\right) = 0$, except for $i = i'$ and $j = j'$

- $\mathrm{cov}(u_i, u_{i'}) = 0, \forall i \neq i'$

- $\mathrm{cov}\left(u_i, e_{i'j'}\right) = 0, \forall i, i' \text{e j}'$

- Although u_i and e_{ij} are uncorrelated, the Y_{ij}s are not. In the same random effect the observations are correlated (a difference from the fixed effects model):

$$\text{cov}(Y_{ij}, Y_{ij\prime}) = cov\left(\mu + u_i + e_{ij}, \mu + u_i + e_{ij\prime}\right) = \sigma^2{}_u \text{, for } j \neq j\prime$$

$$\text{cov}(Y_{ij}, Y_{i\prime j\prime}) = 0, for\ i \neq i\prime$$

- As a consequence,

$$E(Y_{ij}) = \mu$$

$$var(Y_{ij}) = \text{var}(\mu + u_i + e_{ij}) = \sigma^2{}_u + \sigma^2{}_e$$

$$var(\bar{Y}_{i.}) = \sigma^2{}_u + \sigma^2{}_e \Big/ b$$

Many other situations can arise according to:

➢ $X\beta$

- one or more factors of fixed effects (including interactions, nested factores, etc.)

➢ Zu

- one ou more factors of random effects (including interactions, nested factores, etc.)

➢ **The structure of the covariance matrices *G* and *R***

As mentioned above, in the traditional linear mixed model the **covariance matrices G and R are defined as** $G = \sigma_u^2 I_q$ and $R = \sigma_e^2 I_n$, respectively:

$$G_{(q \times q)} = \begin{bmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & \sigma_u^2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_u^2 \end{bmatrix} \quad R_{(n \times n)} = \begin{bmatrix} \sigma_e^2 & 0 & \cdots & 0 \\ 0 & \sigma_e^2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_e^2 \end{bmatrix}$$

But modeling the covariance struture is arguably the most powerful and important single feature of mixed models, and what sets it apart from conventional linear models. For exemple, it permits to include correlation among observations, heterogeneous variances.

➤ *ZU* and matrix *G*

For example, frequently:

❑ **vector** u **consists in** k **sub-vectors**, that is,

$$u = \left(u_1^T, \cdots, u_k^T\right)^T$$

where u_i is a vector $q_i \times 1$, thus $q = \sum_{i=1}^{k} q_i$.

The model matrix associated to vector u is:

$$Z = [Z_1 \ Z_2 \ \cdots \ Z_k].$$

❑ Generalizing for k sub-vectors of random effects,

$$Zu = [Z_1 \ Z_2 \ \cdots \ Z_k] \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix} = \sum_{i=1}^{k} Z_i u_i$$

In the simplest case,

❑ each sub-vector of random effects, represented by u_i, has the properties:

$$E[u_i] = 0 \quad \text{var}[u_i] = \sigma^2_{u_i} I_{q_i} = G_i,$$

❑ vectors u_i e $u_{i'}$ are assumed mutually independent,

$$\text{cov}[u_i, u_{i'}] = 0, \text{ for } i \neq i',$$

and, consequently,

$$\text{var}[u] = \bigoplus_{i=1}^{k} G_i = G,$$

that is, G is the direct sum of matrices G_i.

$$G_{(q \times q)} \begin{bmatrix} G_1 & 0 & \cdots & 0 & 0 \\ 0 & G_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & G_{k-1} & 0 \\ 0 & 0 & \cdots & 0 & G_k \end{bmatrix}$$

Some other common covariance structures

- Unstructured (the most complex)

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}$$

- Compound symmetry

$$\begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

- Autoregressive

$$\begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \sigma^2\rho^3 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho^3 & \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 \end{bmatrix}$$

- Toeplitz

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

...

✓ Depending on the context, there are numerous structures for covariance matrices G and R. Some frequently used are inspired in time series and spatial analysis.

The linear mixed model in R

The linear mixed model can be fitted in R using functions **lme** (*library* nlme), **lmer** (*library* lme4)

For example:

> library(nlme)

lme(fixed, …,random, correlation, method, control, ….)

lme(response~fixed effects, random=~random effects…)

Example 1: fitting the classical random model (one random effects factor, $G=\sigma_u^2 I_q$, $R=\sigma_e^2 I_n$)

library(nlme)

There is a single fixed effect, the intercept

Indicates that there is a single random effect for each group and the grouping is given by the variable clone

```
> arlme1<-lme(rend~1, random=~1|clone, data=arinto)
> arlme1
Linear mixed-effects model fit by REML
  Data: arinto
  Log-restricted-likelihood: -714.2113
  Fixed: rend ~ 1
(Intercept)
   1.389713

Random effects:
 Formula: ~1 | clone
        (Intercept)  Residual
StdDev:   0.3288653 0.4273398
```

Example 2: fitting a classical linear mixed model (one factor of fixed effects, one factor of random effects factor, $G=\sigma_u^2 I_q$, $R=\sigma_e^2 I_n$)

library(nlme)

Fixed effect

random effect

- ➤ **> terrenolme1<-lme(rend~variedade, random=~1|terreno, data=terrenos)**
- ➤ **> terrenolme1**
- ➤ **Linear mixed-effects model fit by REML**
- ➤ **Data: terrenos**
- ➤ **Log-restricted-likelihood: -21.71354**
- ➤ **Fixed: rend ~ variedade**
- ➤ **(Intercept) variedadeB variedadeC variedadeD**
- ➤ **1.55600000 -0.02384615 -0.38907692 -0.37784615**

- ➤ **Random effects:**
- ➤ **Formula: ~1 | terreno**
- ➤ **(Intercept) Residual**
- ➤ **StdDev: 0.1604919 0.3123811**

- ➤ **Number of Observations: 52**
- ➤ **Number of Groups: 13**

Example 2: fitting a classical linear mixed model (one factor of fixed effects, one factor of random effects factor, $G=\sigma_u^2 I_q$, $R=\sigma_e^2 I_n$)

library(lme4)

Fixed effect   random effect

```
>  terrenoslmer1<-lmer(rend~ variedade + (1 | terreno), data = terrenos)
terrenoslmer1
Linear mixed model fit by REML ['lmerMod']
Formula: rend ~ variedade + (1 | terreno)
   Data: terrenos
REML criterion at convergence: 43.4271
Random effects:
 Groups   Name       Std.Dev.
 terreno  (Intercept) 0.1605
 Residual            0.3124
Number of obs: 52, groups:  terreno, 13
Fixed Effects:
(Intercept)  variedadeB  variedadeC  variedadeD
   1.55600    -0.02385    -0.38908    -0.37785
```

Example 3: fitting a random model with

$$G = \bigoplus_{i=1}^{k} G_i \quad , G_i = \sigma_{u_i}^2 I_{q_i}, \quad R = \sigma_e^2 I_n$$

Library(nlme)

> arlme2<-lme(rend~1, **random=list(clone=pdDiag(~regiao-1))**, data=arinto)
> arlme2
Linear mixed-effects model fit by REML
 Data: arinto
 Log-restricted-likelihood: -712.0479
 Fixed: rend ~ 1
(Intercept)
  1.414257

Random effects:
 Formula: ~regiao - 1 | clone
 Structure: Diagonal
        regiaoBairrada regiaoLafoes regiaoOeste regiaoVverdes Residual
StdDev: 0.3319987   0.3038772 0.4048056   0.272041 0.4273398

Number of Observations: 988
Number of Groups: 247

Methods for estimating covariance parameters

Covariance components estimation in linear mixed models mainly use three approaches:

❑ **restricted (or residual) maximum likelihood (REML)**
It is the method of estimation currently most used (by default, it is the estimation method used in all packages). The restricted maximum likelihood method in the context of linear mixed models was introduced by Patterson e Thompson (1971)[1].

❑ maximum likelihood (ML)
ML is sometimes discouraged, because the variance component estimates are biased downward.

❑ procedures based on expected mean squares from the analysis of variance (ANOVA): involves equating various quadratic forms of the observations to their expected values, then solving for the variance components (Herderson's methods). It is the classical approach, but has drawbacks: the large number of times that negative estimates of variance are obtained; is only well applied for simple models with balanced data and when G and R are diagonal matrices.

[1] Patterson, H.D., Thompson, R. (1971) - Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**:545-554.

Some considerations about REML/ML/ANOVA estimators

- For unbalanced data each of ML and REML are to be preferred over procedures based on expected mean squares from the analysis of variance (ANOVA).
- REML and ML have the same merits of being based on the maximum likelihood principle that is known to have useful properties of consistency and asymptotic normality of the estimators and the asymptotic sampling dispersion matrix of the estimators is also known.
- For simple models and balanced data REML solutions are identical to ANOVA estimators.
- For complex models, ML and REML are computationally intensive.
- REML and ML estimators are based on assuming normality of the data, but in many circumstances that assumption is unlikely to be seriously wrong. Therefore, the asymptotic variance-covariance properties are valid only in the large-sample sense.

Some considerations about REML/ML/ANOVA estimators

- ML provides estimators of fixed effects, whereas REML , of itself, does not.
- ML variance component estimates are biased downward. REML estimators are based on taking into account the degrees of freedom for the fixed effects in the model (losing 1 degree of freedom for each). For this reason, REML estimators for covariance parameters are unbiased. The use of REML method instead of ML is particularly important when the rank of matrix X is large in relation to the sample size.
- REML estimators do not seem to be as sensitive to outliers in the data as are ML estimators.

Restricted (or residual) maximum likelihood (REML) method

General model $Y = X\beta + Zu + e$, with $Y \sim \mathcal{N}_n(X\beta, V)$,

- Estimation of the covariance parameters included in matrix V:
$$V(\varphi) = ZG(v)Z^T + R(\phi),$$
v is the vector of parameters included in matrix G, ϕ is the vector of parameters included in matrix R, $\varphi = \left(v^T, \phi^T\right)^T$.

The likelihood function is
$$L(\beta, \varphi|y) = \frac{1}{(2\pi)^{\frac{n}{2}}|V|^{1/2}} exp\left(-\frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta)\right),$$

where $V^{-1} = R^{-1} - R^{-1}Z(Z^T R^{-1}Z + G^{-1})^{-1}Z^T R^{-1}$, $|V| = |R|^{-1}|I_n + R^{-1}ZGZ^T|$

Thus the log-likelihood is
$$l(\beta, \varphi|y) = -\frac{1}{2}n \ln 2\pi - \frac{1}{2}\ln|V| - \frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta).$$

The fundamental idea of REML is to maximize the likelihood after accounting for the model's fixed effects. Instead of maximizing the likelihood of $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{X\beta}, \boldsymbol{V})$, maximize the likelihood of $\boldsymbol{K^T Y}$ where $\boldsymbol{K}$ is any matrix such that $E[\boldsymbol{K^T Y}] = \boldsymbol{0}$ and hence $\boldsymbol{K^T Y} \cap \mathcal{N}_n(\boldsymbol{0}, \boldsymbol{K^T V K})$. This removes the fixed effects from the estimation of $\boldsymbol{\varphi}$.

Let

$$\boldsymbol{K^T Y} = \boldsymbol{K^T X \beta} + \boldsymbol{K^T Z u} + \boldsymbol{K^T e}.$$

$\boldsymbol{K^T}$ is a matrix $(n - r_X) \times n$ (where r_X is the rank of matrix $\boldsymbol{X}$), such that $\boldsymbol{E[K^T Y] = 0}$ (that is, $\boldsymbol{K^T X \beta = 0}$). $\boldsymbol{K}$ is also called a matrix of error contrasts (hence the alternative name of residual maximum likelihood method).

Typically, $\boldsymbol{K^T} = \boldsymbol{I} - \boldsymbol{X(X^T X)^{-1} X^T}$.

$$E[\boldsymbol{K^T Y}] = \left[\boldsymbol{I} - \boldsymbol{X(X^T X)^{-1} X^T}\right] E\,[\boldsymbol{Y}] = \left[\boldsymbol{I} - \boldsymbol{X(X^T X)^{-1} X^T}\right] \boldsymbol{X\beta} = 0$$

The transformed model is

$$\boldsymbol{K^T Y = K^T Z u + K^T e}$$

with $\boldsymbol{K^T Y} \cap \mathcal{N}_n \left(\boldsymbol{0, K^T V K}\right)$

- The likelihood function of this model is called the restricted likelihood function and is given by:

$$L_R(\boldsymbol{\varphi}|\boldsymbol{y}) = \frac{1}{(2\pi)^{\frac{1}{2}(n-r_X)} |\boldsymbol{K^T V K}|^{1/2}} exp\left(-\frac{1}{2}\left(\boldsymbol{K^T y}\right)^T \left(\boldsymbol{K^T V K}\right)^{-1} \left(\boldsymbol{K^T y}\right)\right)$$

And the REML log - likelihood is

$$l_R(\boldsymbol{\varphi}|\boldsymbol{y}) = -\frac{1}{2}(n - r_X)\,log\,2\pi - \frac{1}{2}log|\boldsymbol{K}^T\boldsymbol{V}\boldsymbol{K}| - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{V}\boldsymbol{K})^{-1}\boldsymbol{K}^T\boldsymbol{y}.$$

Denoting

$$\boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{V}\boldsymbol{K})^{-1}\boldsymbol{K}^T = \boldsymbol{P}\,,$$

$$l_R = -\frac{1}{2}(n - r_X)\log 2\pi - \frac{1}{2}\log|\boldsymbol{K}^T\boldsymbol{V}\boldsymbol{K}| - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{y}.$$

Note: since $\boldsymbol{K}^T$ has full row rank $n - r_X$ and $\boldsymbol{K}^T\boldsymbol{X} = \boldsymbol{0}$, can be shown that

$$\boldsymbol{K}(\boldsymbol{K}^T\boldsymbol{V}\boldsymbol{K})^{-1}\boldsymbol{K}^T = \boldsymbol{P} \text{ where } \boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}$$

To maximize l_R we differentiate with respect to φ_i and equate to zero and solve the resulting equations.

$$\frac{\partial l_R(\boldsymbol{\varphi}|\boldsymbol{y})}{\partial \varphi_i} = 0 \ .$$

Differentiate with respect to φ_i,

$$\frac{\partial l_R}{\partial \varphi_i} = -\frac{1}{2}\left(tr(\boldsymbol{P\dot{V}_i}) - \boldsymbol{y}^T\boldsymbol{P\dot{V}_iPy}\right),$$

where $\dot{V}_i$ is the derivative of V with respect to φ_i. Equating to zero, we obtain

$$tr(P\dot{V}_i) = y^T P\dot{V}_i Py.$$

Completing the maximization demands checking second derivatives and also demands checking the likelihood function on the boundary of the parameter space, since the maximization must be confined to the parameter space.

REML estimation proceeds iteratively. We implement REML using *Newton-Raphson or* Fisher scoring algorithms (among others).

The Hessian and information matrices derived from REML log-likelihood are:

Hessian ($ijth$ element)

$$\frac{\partial \nabla l_R(\varphi_i)}{\partial \varphi_j} = \frac{1}{2} tr\left(\boldsymbol{P\dot{V}_j P\dot{V}_i}\right) - \frac{1}{2} tr\left(\boldsymbol{P\ddot{V}_{ij}}\right) + \frac{1}{2} \boldsymbol{y}^T \boldsymbol{P\ddot{V}_{ij} Py} - \boldsymbol{y}^T \boldsymbol{P\dot{V}_i P\dot{V}_j Py},$$

where $\boldsymbol{\nabla l_R}$ is the gradient vector of l_R and $\boldsymbol{\ddot{V}}_{ij} = \frac{\partial^2 \boldsymbol{V}}{\partial \varphi_i \partial \boldsymbol{\varphi}_j}$.

Information matrix ($ijth$ element):

$$I_e\left(\varphi_i, \varphi_j\right) = E\left[-\frac{\partial^2 l_R}{\partial \varphi_i \partial \varphi_j}\right]$$

It can be proved that $I_e\left(\varphi_i, \varphi_j\right) = \frac{1}{2} tr\left(\boldsymbol{P\dot{V}_i P\dot{V}_j}\right)$.

Properties of REML estimators

The REML estimator of $\widehat{\boldsymbol{\varphi}}$ is consistent and, under regularity conditions, is asymptotic normal with vector of mean values $\boldsymbol{\varphi}$ and asymptotic covariance matrix $[\boldsymbol{I}_e(\boldsymbol{\varphi})]^{-1}$, where $\boldsymbol{I}_e$ is the information matrix (expected information matrix):

$$\widehat{\boldsymbol{\varphi}} \overset{a}{\sim} \mathcal{N}\left(\boldsymbol{\varphi}, [\boldsymbol{I}_e(\boldsymbol{\varphi})]^{-1}\right).$$

- **Restricted maximum likelihood estimation for variance components**

Example: the particular case of ANOVA with one factor with random effects, balanced

$(\boldsymbol{G}$ and $\boldsymbol{R}$ diagonal matrices$)$

$$Y_{ij} = \mu + u_i + e_{ij}$$

for $i = 1, \dots, a, j = 1, \dots, b.$

Y_{ij} is the jth observation in the ith level of factor A;

μ is a general mean (population);

u_i is the effect of the level i of the factor A (random effects);

e_{ij} is the random error associated to the observation Y_{ij}.

- **Restrited maximum likelihood estimation for variance components**

For a model with one factor of random effects and balanced, the restricted log-likelihood (l_R) is given by:

$$l_R = -\frac{1}{2}(ab-1)\ln 2\pi - \frac{1}{2}\ln ab - \frac{1}{2}a(b-1)\ln \sigma_e^2 - \frac{1}{2}(a-1)\ln \lambda - \frac{SQRE}{2\sigma_e^2} - \frac{SQA}{2\lambda}.$$

with $\quad \lambda = \sigma_e^2 + b\sigma_u^2 \quad SQA = \sum_{i=1}^{a} b\,(\bar{Y}_{i.} - \bar{Y}_{..})^2 \quad SQRE = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{i.})^2$

Equating to zero the partial derivatives of l_R with respect to σ_e^2 and σ_u^2, the following solutions are obtained:

$$\dot\sigma_e{}^2 = \frac{SQRE}{a(b-1)} = QMRE$$

$$\dot\sigma_u{}^2 = \frac{\left(\frac{SQA}{a-1}\right) - QMRE}{b}$$

The restricted maximum likelihood estimators for variance components are:

$$\hat{\sigma}_u{}^2 = \frac{QMA - QMRE}{b}, \quad \text{if } QMA \geq QMRE,$$

$$\hat{\sigma}_u{}^2 = 0, \quad \text{otherwise}$$

$$\hat{\sigma}_e{}^2 = QMRE, \quad \text{if } QMA \geq QMRE,$$

$$\hat{\sigma}_e{}^2 = \frac{SQT}{ab - 1}, \quad \text{otherwise}$$

Asymptotic covariance matrix for REML estimators

$$\text{var}\begin{bmatrix} \hat{\sigma}_e^{\ 2} \\ \hat{\sigma}_u^{\ 2} \end{bmatrix}$$

$$\approx \begin{bmatrix} \dfrac{2\sigma_e^{\ 4}}{a(b-1)} & \dfrac{-2\sigma_e^{\ 4}}{ab(b-1)} \\ & \dfrac{2\sigma_e^{\ 4}}{b^2}\left[\dfrac{(\sigma_e^{\ 2}+b\sigma_u^{\ 2})^2/\sigma_e^{\ 4}}{a} + \dfrac{1}{a(b-1)}\right] \end{bmatrix}$$

In R **Library(lme4)**

REML estimation

> **terrenoslmer1<-lmer(rend~ variedade + (1 | terreno), data = terrenos)**

> **summary(terrenoslmer1)**

Linear mixed model fit by REML ['lmerMod']

Formula: rend ~ variedade + (1 | terreno)

   Data: terrenos

REML criterion at convergence: 43.4

Scaled residuals:

    Min   1Q Median   3Q   Max
-1.91520 -0.72001 0.03862 0.54535 2.82304

Random effects:

 Groups   Name   **Variance** Std.Dev.
 terreno (Intercept) **0.02576** 0.1605
 Residual   **0.09758** 0.3124

Number of obs: 52, groups: terreno, 13

...

ML estimation

> **terrenoslmer1ML<-update(terrenoslmer1, REML=FALSE)**

> **summary(terrenoslmer1ML)**

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula: rend ~ variedade + (1 | terreno)

   Data: terrenos

Scaled residuals:

    Min   1Q Median   3Q   Max
-1.9934 -0.7494 0.0402 0.5676 2.9383

Random effects:

 Groups   Name   **Variance** Std.Dev.
 terreno (Intercept) **0.02378** 0.1542
 Residual   **0.09008** 0.3001

Number of obs: 52, groups: terreno, 13

In ⓇR Library(nlme)

REML estimation

> **terrenolme1<-lme(rend~variedade, random=~1|terreno, data=terrenos)**
> **summary(terrenolme1)**
Linear mixed-effects model fit by REML
 Data: terrenos
      AIC   BIC   logLik
  55.42708 66.65429 -21.71354
Random effects:
 Formula: ~1 | terreno
          (Intercept) Residual
StdDev: 0.1604919 0.3123811

….

> **VarCorr(terrenolme1)**

            Variance   StdDev
(Intercept) **0.02575765** 0.1604919
Residual   **0.09758196** 0.3123811

ML estimation

> **terrenolme1ML<-update(terrenolme1, method="ML")**
> **summary(terrenolme1ML)**

Linear mixed-effects model fit by maximum likelihood
 Data: terrenos
      AIC   BIC   logLik
  43.76901 55.47647 -15.8845
Random effects:
 Formula: ~1 | terreno
          (Intercept) Residual
StdDev: 0.1541963 0.3001259

….

> **VarCorr(arlme2ML)**
            Variance   StdDev
(Intercept) **0.02377651** 0.1541963
Residual   **0.09007553** 0.3001259

In R

Estimated covariance (asymptotic) matrix for REML estimators

$$\widehat{\mathrm{var}(\boldsymbol{\hat{\varphi}})}$$

> vcov(terrenosvarcomp1, "varComp")

```
                terreno         error
terreno  0.0004522882 -0.0001322531
error    -0.0001322531  0.0005290124
```

Estimating fixed effects and predicting random effects: the mixed model equations

- **The estimators of fixed effects and predictors of random effects are obtained through *mixed model equations* (Henderson, 1975[4]).**

[4]Henderson, C.R. (1975) - Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**:423 - 447.

The joint density of $\boldsymbol{u}$ and $\boldsymbol{Y}$ is

$$f(\boldsymbol{Y}, \boldsymbol{u}) = f(\boldsymbol{Y}|\boldsymbol{u})f(\boldsymbol{u}) =$$

$$= \frac{exp\left(-\frac{1}{2}\left((\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u})^T \boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{u}) + \boldsymbol{u}^T \boldsymbol{G}^{-1}\boldsymbol{u}\right)\right)}{(2\pi)^{(n+q)/2}|\boldsymbol{R}|^{1/2}|\boldsymbol{G}|^{1/2}}$$

The mixed models equations are derived by maximizing $\ln f(\boldsymbol{Y}, \boldsymbol{u})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{u}$:

$$\ln f(\boldsymbol{Y}, \boldsymbol{u})$$
$$= -\frac{1}{2}(n+q)\log(2\pi) - \frac{1}{2}(\ln|\boldsymbol{R}| + \ln|\boldsymbol{G}|)$$
$$-\frac{1}{2}\left(\boldsymbol{Y}^T \boldsymbol{R}^{-1}\boldsymbol{y} - 2\boldsymbol{Y}^T \boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{Y}^T \boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{u} + 2\boldsymbol{\beta}^T \boldsymbol{X}^T \boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\beta}^T \boldsymbol{X}^T \boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}^T \boldsymbol{Z}^T \boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{u}$$
$$+ \boldsymbol{u}^T \boldsymbol{G}^{-1}\boldsymbol{u}\right).$$

Equating to zero:

$$\frac{\partial \ln f(\boldsymbol{Y}, \boldsymbol{u})}{\partial \boldsymbol{\beta}} = 0 \Leftrightarrow \boldsymbol{X}^T \boldsymbol{R}^{-1}\boldsymbol{Y} - \boldsymbol{X}^T \boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}^T \boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{u} = 0$$

$$\frac{\partial \ln f(\boldsymbol{Y}, \boldsymbol{u})}{\partial \boldsymbol{u}} = 0 \Leftrightarrow \boldsymbol{Z}^T \boldsymbol{R}^{-1}\boldsymbol{Y} - \boldsymbol{Z}^T \boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}^T \boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{u} - \boldsymbol{G}^{-1}\boldsymbol{u} = 0 \,.$$

Which results in

$$\begin{bmatrix} X^T R^{-1} X \beta + X^T R^{-1} Z u \\ Z^T R^{-1} X \beta + (Z^T R^{-1} Z + G^{-1}) u \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix},$$

and, consequently, the mixed models equations are defined as:

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

- *The estimator of fixed effects ($\widehat{\beta}$) and the predictor of random effects ($\tilde{u}$) are:*

$$\widehat{\beta} = \left(X^T V^{-1} X \right)^{-1} X^T V^{-1} Y$$

$$\tilde{u} = G Z^T V^{-1} \left(Y - X \widehat{\beta} \right).$$

Proof in next slides, additional information

Additional information

Proof

Para chegar às soluções das equações do modelo misto, podemos começar por verificar, a partir da segunda equação de $\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & (Z^T R^{-1} Z + G^{-1}) \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X^T R^{-1} Y \\ Z^T R^{-1} Y \end{bmatrix}$ que

$$u = (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} (Y - X\beta),$$

e substituindo este resultado na primeira equação,

$$X^T R^{-1} X\beta + X^T R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} (y - X\beta) = X^T R^{-1} Y,$$

logo,

$$X^T R^{-1} Y - X^T R^{-1} X\beta - X^T R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} (Y - X\beta) = 0.$$

Depois de algum rearranjo da expressão anterior, chega-se a

$$X^T \left(R^{-1} - R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \right) X\beta = X^T \left(R^{-1} - R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \right) Y,$$

o que conduz à solução

$$\hat{\beta} = \left(X^T \left(R^{-1} - R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \right) X \right)^{-1} X^T \left(R^{-1} - R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} \right) Y$$

Mas, sabendo que $V^{-1} = R^{-1} - R^{-1} Z (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}$, obtém-se

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

Proof (cont.)

Voltando novamente a $\boldsymbol{u}$, verifica-se que

$$
\begin{aligned}
\widetilde{\boldsymbol{u}} = \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} & Z^T R^{-1} V V^{-1}\left(Y - X\widehat{\boldsymbol{\beta}}\right) \\
&= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1}\left(ZGZ^T + R\right)V^{-1}\left(Y - X\widehat{\boldsymbol{\beta}}\right) \\
&= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1}\left(Z^T R^{-1} ZGZ^T + Z^T\right)V^{-1}\left(Y - X\widehat{\boldsymbol{\beta}}\right) \\
&= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1}\left(Z^T R^{-1} Z + G^{-1}\right)GZ^T V^{-1}\left(Y - X\widehat{\boldsymbol{\beta}}\right),
\end{aligned}
$$

isto é,
$$
\widetilde{\boldsymbol{u}} = GZ^T V^{-1}\left(Y - X\widehat{\boldsymbol{\beta}}\right).
$$

- Vector $\widehat{\boldsymbol{\beta}}$ has multivariate normal distribution with vector of mean values $\boldsymbol{\beta}$ and covariance matrix $\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}$

$$\widehat{\boldsymbol{\beta}} \cap \mathcal{N}\left(\boldsymbol{\beta}, \left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\right)$$

- Vector $\widetilde{\boldsymbol{u}}$ has multivariate normal distribution with vector of mean values zero and covariance matrix $\boldsymbol{G}\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{Z}\boldsymbol{G}$,

$$\widetilde{\boldsymbol{u}} \cap \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{G}\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{Z}\boldsymbol{G}\right),$$

with $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{V}^{-1}$

Exercise: prove the previous results.

❑ When matrices G e R are known, $\widehat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$, and $\widetilde{\boldsymbol{u}}$ is the best linear unbiased predictor (BLUP) of $\boldsymbol{u}$.

❑ However, in general, matrices G e R are unknown, and only their estimates are available, $\widehat{G}$ e $\widehat{R}$. In this case, we have the empirical best linear unbiased estimator (EBLUE) and the empirical best linear unbiased predictor (EBLUP):

$$\widehat{\boldsymbol{\beta}}_{EBLUE} = \left(X^T\widehat{V}^{-1}X\right)^{-1}X^T\widehat{V}^{-1}Y$$

and

$$\widetilde{\boldsymbol{u}}_{EBLUP} = \widehat{G}Z^T\widehat{V}^{-1}\left(Y - X\widehat{\boldsymbol{\beta}}_{EBLUE}\right).$$

Example: for a random model with one factor of random effects, balanced (factor with a levels, b observations per level), the empirical best linear unbiased predictor of u_i (for the level i) is:

$$EBLUP(u_i) = \frac{b\hat{\sigma}^2{}_u}{b\hat{\sigma}^2{}_u + \hat{\sigma}^2{}_e}(\bar{Y}_{i.} - \bar{Y}_{..})$$

Given the mixed model equations:

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

denote by C the matrix:

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix}^{-1} = C.$$

The elements of matrix C:

$$C_{11} = \left(X^T V^{-1} X \right)^{-1},$$

$$C_{21} = -G Z^T V^{-1} X C_{11},$$

$$C_{22} = \left(Z^T R^{-1} Z + G^{-1} \right)^{-1} - C_{21} X^T V^{-1} Z G = G - G Z^T P Z G$$

$$C_{12} = C_{21}^T .$$

The matrix C is the covariance matrix of $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widetilde{\boldsymbol{u}} - \boldsymbol{u})$,

$$C = \begin{bmatrix} \mathrm{var}[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}] & \mathrm{cov}[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widetilde{\boldsymbol{u}} - \boldsymbol{u}] \\ \mathrm{cov}[\widetilde{\boldsymbol{u}} - \boldsymbol{u}, \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}] & \mathrm{var}[\widetilde{\boldsymbol{u}} - \boldsymbol{u}] \end{bmatrix}.$$

$\mathrm{var}[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}] = \mathrm{var}[\widehat{\boldsymbol{\beta}}]$, is the covariance matrix of the fixed effects estimators

$\mathrm{var}[\widetilde{\boldsymbol{u}} - \boldsymbol{u}]$, is the prediction variance matrix of the random effects

Proof (additional information): Segundo a teoria de matrizes, a inversa da matriz não singular subdividida em

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

assumindo que todas as inversas necessárias existem, é

$$A^{-1} = \begin{bmatrix} T & -TA_{22}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}T & A_{22}^{-1} + A_{22}^{-1}A_{21}TA_{12}A_{22}^{-1} \end{bmatrix}$$

onde $T = \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}$

Aplicando ao nosso caso, chega-se a

$$C_{11} = \left((X^T R^{-1} X) - (X^T R^{-1} Z)(Z^T R^{-1} Z + G^{-1})^{-1}(Z^T R^{-1} X)\right)^{-1}$$
$$= \left(X^T \left(R^{-1} - R^{-1} Z(Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}\right) X\right)^{-1},$$

e, atendendo a que $V^{-1} = R^{-1} - R^{-1} Z(Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}$, $C_{11} = (X^T V^{-1} X)^{-1}$.

$$C_{21} = -(Z^T R^{-1} Z + G^{-1})^{-1}(Z^T R^{-1} X)(X^T V^{-1} X)^{-1}$$

$$= -(Z^T R^{-1} Z + G^{-1})^{-1}(Z^T R^{-1} V V^{-1} X)(X^T V^{-1} X)^{-1}$$

$$= -(Z^T R^{-1} Z + G^{-1})^{-1}(Z^T R^{-1} V)V^{-1} X(X^T V^{-1} X)^{-1}$$

$$= -(Z^T R^{-1} Z + G^{-1})^{-1}(Z^T R^{-1}(ZGZ^T + R))V^{-1} X(X^T V^{-1} X)^{-1}$$

$$= -(Z^T R^{-1} Z + G^{-1})^{-1}(Z^T R^{-1} ZGZ^T + Z^T R^{-1} R)V^{-1} X(X^T V^{-1} X)^{-1}$$

$$= -(Z^T R^{-1} Z + G^{-1})^{-1}(Z^T R^{-1} ZGZ^T + G^{-1}GZ^T)V^{-1} X(X^T V^{-1} X)^{-1}$$

$$= -(Z^T R^{-1} Z + G^{-1})^{-1}(Z^T R^{-1} Z + G^{-1})GZ^T V^{-1} X(X^T V^{-1} X)^{-1} = -GZ^T V^{-1} X C_{11}.$$

Finally,

$$C_{22} = \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} - \left(-\left(Z^T R^{-1} Z + G^{-1}\right)^{-1}\left(Z^T R^{-1} X\right)\left(X^T V^{-1} X\right)^{-}\right)\left(X^T R^{-1} Z\right)\left(Z^T R^{-1} Z + G^{-1}\right)^{-1}$$

$$= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} - C_{21}\left(X^T V^{-1} V R^{-1} Z\right)\left(Z^T R^{-1} Z + G^{-1}\right)^{-1}$$

$$= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} - C_{21}\left(X^T V^{-1}\left(Z G Z^T + R\right) R^{-1} Z\right)\left(Z^T R^{-1} Z + G^{-1}\right)^{-1}$$

$$= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} - C_{21} X^T V^{-1}\left(Z G Z^T R^{-1} Z + R R^{-1} Z\right)\left(Z^T R^{-1} Z + G^{-1}\right)^{-1}$$

$$= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} - C_{21} X^T V^{-1}\left(Z G Z^T R^{-1} Z + Z G G^{-1}\right)\left(Z^T R^{-1} Z + G^{-1}\right)^{-1}$$

$$= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} - C_{21} X^T V^{-1} Z G\left(Z^T R^{-1} Z + G^{-1}\right)\left(Z^T R^{-1} Z + G^{-1}\right)^{-1}$$

$$= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} - C_{21} X^T V^{-1} Z G.$$

Proof (additional information)

A matriz C_{11}, corresponde à matriz de covariâncias dos estimadores $\widehat{\beta}$, já que

$$Var[\widehat{\beta} - \beta] = Var[\widehat{\beta}].$$

A matriz C_{12}, corresponde à matriz de covariâncias entre os erros de estimação e de predição, uma vez que

$$Cov[\widehat{\beta} - \beta, \widetilde{u} - u] = E[\widehat{\beta}\widetilde{u}^T] - E[\widehat{\beta}u^T]$$

Mas,

$$E[yy^T] = V + X\beta\beta^T X^T \,,$$
$$= (X^T V^{-1} X)^- X^T V^{-1} E[yy^T] PZG,$$

e sendo

$$E[\widehat{\beta}\widetilde{u}^T] = E[(X^T V^{-1} X)^- X^T V^{-1} yy^T PZG]$$

então,

$$E[\widehat{\beta}\widetilde{u}^T] = (X^T V^{-1} X)^- X^T V^{-1} (V + X\beta\beta^T X^T) PZG$$

$$= (X^T V^{-1} X)^- X^T PZG + (X^T V^{-1} X)^- X^T V^{-1} X\beta\beta^T X^T PZG$$

$$= 0,$$

porque $X^T P = 0$ $(PX = V^{-1} X - V^{-1} X(X^T V^{-1} X)^- X^T V^{-1} X = 0)$.

Proof (additional information)

Por outro lado,

$$E[\hat{\beta}u^T] = E[(X^TV^{-1}X)^-X^TV^{-1}yu^T]$$

$$= (X^TV^{-1}X)^- X^TV^{-1}E[yu^T]$$

$$= (X^TV^{-1}X)^- X^TV^{-1}ZG$$
$$= C_{11}\,X^TV^{-1}ZG,$$

consequentemente,

$$Cov[\hat{\beta} - \beta, \tilde{u} - u] = -C_{11}X^TV^{-1}ZG,$$

isto é,

$$Cov[\hat{\beta} - \beta, \tilde{u} - u] = C_{12}.$$

Relativamente a C_{22}

$$Var[\,\tilde{u} - u] = G - GZ^TPZG$$

$$= G - GZ^T(V^{-1} - V^{-1}X(X^TV^{-1}X)^-X^TV^{-1})ZG$$

$$= G - GZ^TV^{-1}ZG - GZ^TV^{-1}X(X^TV^{-1}X)^-X^TV^{-1}ZG,$$

Proof (additional information)

mas, tendo em conta que

$$GZ^TV^{-1}X(X^TV^{-1}X)^-X^TV^{-1}ZG = C_{21}X^TV^{-1}ZG,$$

resta provar que

$$G - GZ^TV^{-1}ZG = (Z^TR^{-1}Z + G^{-1})^{-1}.$$

Sabendo-se que

$$GZ^TV^{-1} = (Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1},$$

então

$$G - (Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1}ZG - (Z^TR^{-1}Z + G^{-1})^{-1} =$$
$$= G - (Z^TR^{-1}Z + G^{-1})^{-1}(Z^TR^{-1}ZG + I_q)$$
$$= G - (Z^TR^{-1}Z + G^{-1})^{-1}(Z^TR^{-1}Z + G^{-1})G$$
$$= G - G = 0,$$

logo,

$$Var[\tilde{u} - u] = (Z^TR^{-1}Z + G^{-1})^{-1} - C_{21}X^TV^{-1}ZG = C_{22}.$$

The residuals

In the linear mixed model , $Y = X\beta + Zu + e,$ with $Y \sim \mathcal{N}_n(X\beta, V)$

- **The conditional residuals that measure deviations from the conditional mean are defined as :**

$$\tilde{e} = Y - X\widehat{\beta} - Z\widetilde{u}.$$

Alternatively, it is common appear in the form:
$$\tilde{e} = RPY,$$
with $P = V^{-1} - V^{-1}X(X^T V^{-1} X)^{-1} X^T V^{-1}$ and $V - ZGZ^T = R$.

Proof in next slide (additional information)

Vector $\tilde{e}$ has multivariate normal distribution with vector of mean values $\mathbf{0}_{(n\times 1)}$ and covariance matrix RPR,
$$\tilde{e} \sim \mathcal{N}_n(0, RPR),$$
with $P = V^{-1} - V^{-1}X(X^T V^{-1} X)^{-1} X^T V^{-1}$ and $V - ZGZ^T = R$.

➢ Proof: $\tilde{e} = RPY$ (additional information)

With the results obtained for $\hat{\beta}$ e $\tilde{u}$

$$\tilde{e} = Y - X(X^TV^{-1}X)^{-1}X^TV^{-1}Y - ZGZ^TPY$$
$$= \left(I_n - X(X^TV^{-1}X)^{-1}X^TV^{-1} - ZGZ^TP\right)Y$$
$$= \left(V\left(V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}\right) - ZGZ^TP\right)Y.$$

Using the result $P = V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}$,

$$\tilde{e} = VPY - ZGZ^TPY$$
$$= (V - ZGZ^T)PY,$$

but $V - ZGZ^T = R$, therefore, $\tilde{e} = RPY$

In 🄬

❑ **The empirical best linear unbiased estimator (EBLUE) of fixed effects**

```
> fixed.effects(terrenolme1)
  (Intercept)    variedadeB  variedadeC  variedadeD
 1.55600000 -0.02384615 -0.38907692 -0.37784615
```

❑ **The empirical best linear unbiased predicted (EBLUP) of random effects**

```
> ranef(terrenolme1)
         (Intercept)
I     0.042982693
II    0.123871539
III  -0.009659254
IV   -0.047150592
IX   -0.165530776
V     0.208483839
VI    0.056207377
VII   0.010883945
VIII  0.023081469
X    -0.211881368
XI    0.032325909
XII  -0.122903638
XIII  0.059288857
```

> fitted(terrenolme1)

I	II	III	IV	V	VI	VII	VIII
1.5989827	1.6798715	1.5463407	1.5088494	1.7644838	1.6122074	1.5668839	1.5790815

IX	X	XI	XII	XIII	I	II	III
1.3904692	1.3441186	1.5883259	1.4330964	1.6152889	1.5751365	1.6560254	1.5224946

...

> predict(terrenolme1)

I	II	III	IV	V	VI	VII	VIII
1.5989827	1.6798715	1.5463407	1.5088494	1.7644838	1.6122074	1.5668839	1.5790815

IX	X	XI	XII	XIII	I	II	III
1.3904692	1.3441186	1.5883259	1.4330964	1.6152889	1.5751365	1.6560254	1.5224946

...

> resid(terrenolme1) (conditional residuals)

I	II	III	IV	V	VI
0.201017307	0.029128461	-0.269340746	0.166150592	0.049516161	0.283792623

VII	VIII	IX	X	XI	XII
-0.488883945	0.160918531	-0.190469224	0.155881368	0.343674091	-0.264096362

...

In ®

Estimated covariance (asymptotic) matrix for estimators $\widehat{\beta}$

$$\widehat{\mathrm{var}}(\widehat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}^T \widehat{\boldsymbol{V}}^{-1} \boldsymbol{X}\right)^{-1}$$

> vcov(terrenolme1)

```
                (Intercept)   variedadeB   variedadeC   variedadeD
(Intercept)     0.009487662  -0.007506305 -0.007506305 -0.007506305
variedadeB     -0.007506305   0.015012610  0.007506305  0.007506305
variedadeC     -0.007506305   0.007506305  0.015012610  0.007506305
variedadeD     -0.007506305   0.007506305  0.007506305  0.015012610
```

Tests of Hypotheses for covariance parameters, fixed and random effects

Model selection (model comparison via likelihood ratio tests and via information criteria)

Tests of Hypotheses for covariance parameters

- Inference concerning covariance parameters of a linear mixed model usually relies on approximate distributions for the RE(ML) estimators derived from asymptotic results.
  In the current context of mixed models, the most important formal test for covariance components is the likelihood ratio test (the most used is the REML likelihood ratio test).

Likelihood ratio tests for covariance components

- Hypotheses: $H_0: \varphi_i = 0 \; (reduced \; model) \quad vs \quad H_1: \varphi_i \neq 0 \; (full \; model)$

  ($H_0: \varphi_i = 0 \quad vs \quad H_1: \varphi_i > 0$, if φ_i is a variance component)

- The REML likelihood ratio statistic :

$$\Lambda = 2\left(l_{R_1} - l_{R_0}\right) \sim \chi_\nu^2$$

being l_{R_1} the REML log-likelihood of the more general model (full model) and l_{R_0} the REML log-likelihood of the reduced model (that is, the REML log-likelihood under the null hypothesis). Under regularity conditions and under the null hypothesis, the likelihood ratio statistic, has an approximate χ_ν^2 distribution with the degrees of freedom (ν) equal to the difference in the number of parameters between the two models. However, when we test a variance component, under the null hypothesis the parameter falls on the boundary of the parameter space. Theoretically it can be shown that for a single variance component, the asymptotic distribution of the REMLRT is a mixture of χ^2 variables , where the mixing probabilities are 0.5, one with 0 degrees of freedom and the other with one degree of freedom. As a consequence we can perform the likelihood ratio test as if the standard conditions apply, and divide the resulting p-value by two.

The REML likelihood ratio test is only valid if the fixed effects are the same for both models.

- Significance level: α

- Rejection region: upper (right-hand) tail, Reject H_0 if $\Lambda_{calc} > \chi^2{}_{\alpha(\nu)}$

For ML likelihood ratio test the procedure is similar, using the log-likelihood instead of REML log-likelihood.

Likelihood ratio tests for covariance parameters

- The REML likelihood ratio test is implicitly two-sided, and must be adjusted when the test involves an hypothesis with the parameter on the boundary of the parameter space. When we test a variance component, under the null hypothesis the parameter falls on the boundary of the parameter space. Theoretically it can be shown that for a single variance component, the asymptotic distribution of the REMLRT is a mixture of χ^2 variates , where the mixing probabilities are 0.5, one with 0 degrees of freedom and the other with one degree of freedom. As a consequence we can perform the likelihood ratio test as if the standard conditions apply, and divide the resulting p-value by two.

- The distribution of the REMLRT for the test involving more complex situations (for example, test that k variance components are zero), involves a mixture of χ^2 variates from 0 to k degrees of freedom (Self and Liang, 1987; Stram and Lee, 1994; Verbeke and Molenberghs, 2003, etc.).

  In **R** *the library* '**RLRsim**' is dedicated to the simulation of the empirical distribution of the REML likelihood statistic when variance components testing is involved. The *library* '**lme4**' also has this functionality.

Likelihood ratio tests for covariance parameters

- The naïve approach to use a χ^2 distribution with degrees of freedom determined by the difference in the number of parameters in the models is currently implemented by several packages. However, one should be aware that the p-values obtained may be conservative (that is, the reported p-value may be greater than the true p-value for the test).

In

Performing a REML likelihood ratio test for a variance component

➢ **TerrenosH1<-varComp(rend~variedade, random=~terreno, data=terrenos)**
> logLik(TerrenosH1)
'log Lik.' -202.4013 (df=2)

> TerrenosH0<-varComp(rend~variedade, data=terrenos)
> logLik(TerrenosH0)
'log Lik.' -203.6992 (df=1)

> 2*(logLik(TerrenosH1)-logLik(TerrenosH0))
'log Lik.' 2.595781

> 1-pchisq(2.595781,1)
[1] 0.1071486
> (1-pchisq(2.595781,1))/2
[1] 0.0535743

Some considerations

1. For its historical importance, we will also study tests concerning variance components in balanced designs, with G and R diagonal matrices, derived from the usual analysis of variance in the applications section (for example: random model with one random effects factor; linear mixed model with one factor of fixed effects, one factor of random effects, without interaction and with interaction; models with several random effects factors; analysis of split plot experiments).

2. For random or complex mixed models there are no exact statistical tests for certain model effects (the numerator and denominator of the F statistics are linear combinations of mean squares). In these cases, approximate F tests are performed. One of the classic methods most used for this approach is the method of Satterthwaite (1941). However, other methods are implemented in more complex mixed models frequently reported in the literature and commonly used in several packages, for example, the methods of Giesbrecht and Burns (1985) and Kenward and Roger (1997 (additional information in the applications section).

3. A note about Wald statistic

One common statistic is the Wald Z, which is computed as the parameter estimate divided by its estimated asymptotic standard error computed from the inverse of the second derivative matrix of the log-likelihood with respect to covariance parameters. The Wald Z test is valid for large samples, but it can be unreliable for small data sets and for parameters such as variance components that are known to have a bounded sampling distribution.

Not recommended for variance components

Tests of hypotheses for fixed
(and random effects)

Tests of hypotheses for linear combinations of the effects of the mixed model ($L^T \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix}$), being L a non random vector

- Hypotheses : $\quad H_o: L^T \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix} = \mathbf{0} \qquad$ vs. $\qquad H_1: L^T \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix} \neq \mathbf{0}$

- Test statistic: $T = \dfrac{L^T \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{u}} \end{bmatrix}}{\sqrt{(L^T \widehat{C} L)}} \sim t_{\upsilon_2}$, under H_0

Under the assumed normality of $\boldsymbol{u}$ and $\boldsymbol{e}$, T has an exact t-distribution only for data exhibiting certain types of balance and for some special unbalanced cases. In general, it is only approximately t-distributed, and its degrees of freedom must be estimated (for example, using *Satterthwaite approximation).* This not happen only for particular cases for data exhibiting certain types of balance and for some special unbalanced cases with the elements of the vectors $\boldsymbol{u}$ e $\boldsymbol{e}$ being *i.i.d.* random variables. In these cases, $\upsilon_2 = n - r(\boldsymbol{W})$, where $r(\boldsymbol{W})$ is the rank of the matrix $\mathbf{W}$ which contains the columns of matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$.

$\sqrt{(L^T \widehat{C} L)}$ is a scalar, is the standard error of the estimator of the parameter being tested, matrix $\boldsymbol{C}$ was defined in slides 61 and 62.

Tests of hypotheses for linear combinations of the effects of the mixed model ($L^T \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix}$)**, being** L **a non random vector (cont.)**

- Significance level: α

- Rejection region: two-tailed

$$\text{Reject } H_0 \text{ if } |T_{calc}| > t_{\alpha/2\,(v_2)}$$

The current procedure is mainly focused on the test of hypotheses for the fixed effects of the model. In this case $\widehat{C} = \left(X^T \widehat{V}^{-1} X\right)^{-1}$.

It follows from previous slides that a confidence interval $(1 - \alpha) \times 100\%$ for $L^T \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix}$ is given as:

$$\left] L^T \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{u}} \end{bmatrix} - t_{\alpha/2, v_2} \sqrt{L^T \widehat{C} L}, L^T \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{u}} \end{bmatrix} + t_{\alpha/2, v_2} \sqrt{L^T \widehat{C} L} \right[$$

Tests of hypotheses for linear combinations of the fixed effects of the mixed model ($L^T[\beta]$), when L is a matrix (rank of L greater than 1)

- Hypotheses: $H_o : L^T[\beta] = 0$ vs. $H_1 : not\ all\ L^T[\beta] = 0$

- Test statistic : $F = \dfrac{[\hat{\beta}]^T L (L^T \widehat{C_{11}} L)^{-1} L^T [\hat{\beta}]}{rank(L)} \sim \mathcal{F}_{v_1, v_2}$, sob H_0

F in general has an approximate F-distribution, with $v_1 = rank(L)$ and v_2 must be estimated (for example, using *Satterthwaite approximation)*. This not happen only for particular cases for data exhibiting certain types of balance and for some special unbalanced cases with the elements of the vectors u e e being *i.i.d.* random variables. In these cases, $v_2 = n - r(W)$, where $r(W)$ is the rank of the matrix W which contains the columns of matrices X and Z.

Tests of hypotheses for linear combinations of the effects of the mixed model ($L^T \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix}$), when L is a matrix (rank of L greater than 1) (cont.)

- Significance level : α

- Rejection region: upper (right-hand) tail

$$\text{Reject } H_0 \text{ if } F_{calc} > f_{\alpha(v_1, v_2)}$$

Some considerations

- For random or complex mixed models there are no exact statistical tests for certain model effects (the numerator and denominator of the F statistics are linear combinations of mean squares). In these cases, approximate F tests are performed. One of the classic methods most used for this approach is the method of Satterthwaite (1941). However, other methods are implemented in more complex mixed models frequently reported in the literature and commonly used in several packages, for example, the methods of Giesbrecht and Burns (1985) and Kenward and Roger (1997). (next slide, additional information)

Additional information

❑ **Example: Satterthwaite Degrees of freedom Approximation**

Satterthwaite showed that given the ratio

$$\frac{X^2_{num}/v_1}{X^*_2/v^*_2}$$

where $X^2_{num} \cap \chi^2_{v_1}$ and X^*_2 is a linear combination of chi-square random variable all independent of X^2_{num}, the $X^*_2 \cap \chi^2_{v^*_2}$, where

$$v^*_2 \cong \frac{\left(\sum_m l_m X^2_m\right)^2}{\sum_m (l_m X^2_m)/df_m},$$

X^2_m denotes the $\chi^2_{df_m}$ random variables, l_m denote the constants in the linear combination, df_m the degrees of freedom for the respective X^2_m.

No R

> **summary(arlme2)**
Linear mixed-effects model fit by REML
 Data: arinto
     AIC   BIC   logLik
 1152.537 1181.887 -570.2686

Random effects:
 Formula: ~1 | clone
          (Intercept) Residual
StdDev: 0.3512633 0.3488315

Fixed effects: rend ~ bloco

	Value	Std.Error	DF	t-value	p-value
(Intercept)	1.3938704	0.03149894	738	44.25134	0.0000
blocoB3	-0.3075385	0.03138934	738	-9.79754	0.0000
blocoB4	-0.0085951	0.03138934	738	-0.27382	0.7843
blocoB5	0.2995020	0.03138934	738	9.54152	0.0000

> **anova(arlme2)**

	numDF	denDF	F-value	p-value
(Intercept)	1	738	3101.4982	<.0001
bloco	3	738	124.6915	<.0001

> intervals(arlme2)
Approximate 95% confidence intervals

 Fixed effects:

	lower	est.	upper
(Intercept)	1.33203224	1.393870445	1.4557087
blocoB3	-0.36916150	-0.307538462	-0.2459154
blocoB4	-0.07021818	-0.008595142	0.0530279
blocoB5	0.23787898	0.299502024	0.3611251

attr(,"label")
[1] "Fixed effects:"

 Random Effects:
 Level: clone

	lower	est.	upper
sd((Intercept))	0.3144	0.3512633	0.3924488

Model selection

- ▪ Model comparison via likelihood ratio tests

- REML and ML likelihood ratio tests can be used to compare nested models. One model is said to be nested within another model if it represents a special case of the other model.

- The REML likelihood ratio tests should only be used for models with same fixed-effects specification.

(procedure already described in slide 75)

Model selection

- **Model comparison via information criteria**

- Information criteria provide an alternative to formal testing. The two most widely used information criteria are the AIC (Akaike information criterion) and BIC (Bayesian information criterion). There are others, for example: sample corrected version of AIC, AICC; Information criterion of Hannan-Quinn, HQIC; etc..

- Both the AIC and BIC contain two terms that measure the fit of the model and the complexity of the model.

- With REML, AIC and BIC criteria can only be used for comparing models with same fixed-effects specification (with equal $X\beta$)

- They can be used to compare nested models and non-nested models

Akaike information criterion (AIC_R) is defined as:

$$AIC_R = -2l_R + 2n_{par} \, ,$$

l_R is the REML log-likelihood of the model and n_{par} is the number of covariance parameters in the model.

- AIC is calculated for each model. The model with the smallest value is chosen as the preferred model.
- Penalizes model complexity (models with a higher number of parameters).

Note: For ML estimation method, REML log-likelihood is replaced by log-likelihood and n_{par} is the number of parameters in the model (thus, includes fixed effects).

Bayesian information criterion (BIC_R) is defined as,

$$BIC_R = -2l_R + n_{par}ln(n - r_X),$$

l_R is the REML log-likelihood of the model and n_{par} is the number of covariance parameters in the model, n is the number of observations and r_X is the rank of matrix $\boldsymbol{X}$.

- BIC is calculated for each model. The model with the smallest value is chosen as the preferred model.
- Penalizes model complexity (models with a higher number of parameters).
- The number of observations is also taken into account.

Note: For ML estimation method, the REML log-likelihood is replaced by log-likelihood, n_{par} is the number of parameters in the model (thus, includes fixed effects) and $ln(n - r_X)$ is replaced by $ln(n)$.

In ®R

```
>siria.lme1<-lme(rend~subpop+bloco, data=siria, random=list(clone=pdIdent(~subpop-1)))
> summary(siria.lme1)
Linear mixed-effects model fit by REML
 Data: siria
     AIC    BIC   logLik
 2163.291 2207.831 -1072.646
Random effects:
 Formula: ~subpop - 1 | clone
 Structure: Multiple of an Identity
      subpopALE subpopALG subpopPINHEL  Residual
StdDev: 0.4550345 0.4550345   0.4550345 0.5784932
```

```
> siria.lme2<-lme(rend~subpop+bloco, data=siria, random=list(clone=pdDiag(~subpop-1)))
> summary(siria.lme2)
Linear mixed-effects model fit by REML
 Data: siria
     AIC        BIC        logLik
 2153.235 2207.673 -1065.618
Random effects:
 Formula: ~subpop - 1 | clone
 Structure: Diagonal
      subpopALE subpopALG subpopPINHEL  Residual
StdDev:  0.245502 0.4622154   0.5355529 0.5784956
```

> **anova(siria.lme1,siria.lme2)**

Attention to AIC and BIC values!

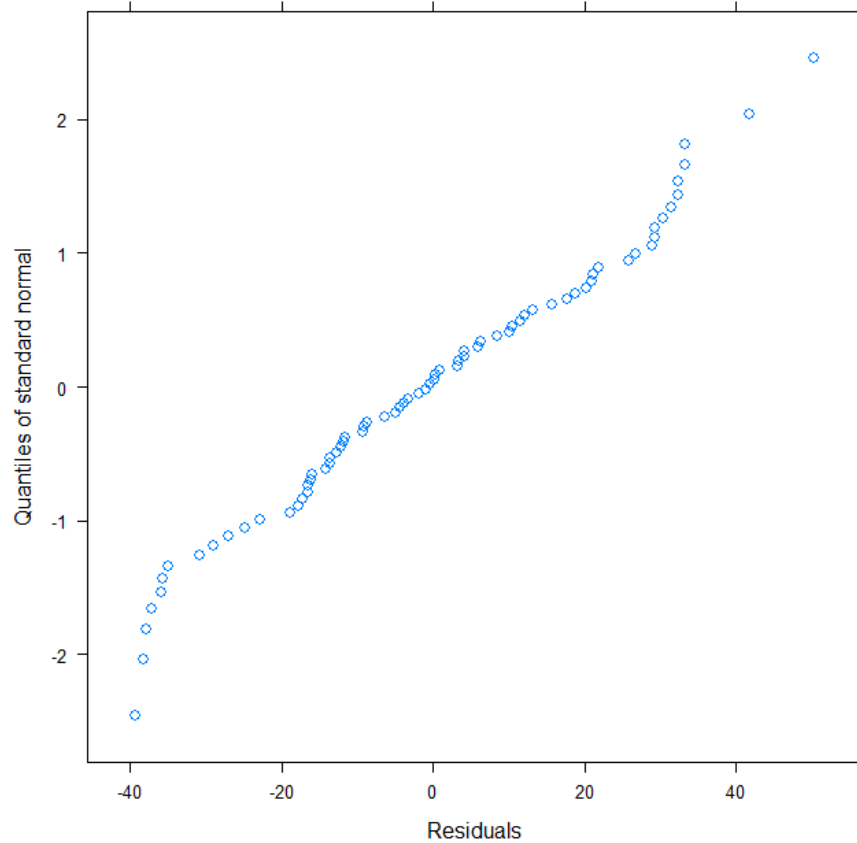|	Model	df	**AIC**	**BIC**	**logLik**	Test	**L.Ratio**	p-value
siria.lme1	1	9	**2163.291**	**2207.831**	**-1072.646**			
siria.lme2	2	11	**2153.235**	**2207.673**	**-1065.618**	1 vs 2	**14.05585**	9e-04

96

Validation of the model assumptions

❑ As in classical linear model, validation of model assumptions is made essentially using graphic tools.

❑ One key assumption is the normality. Now it is necessary to validate this hypothesis in the residuals (by tools already mentioned in the linear model) and in the predictors of random effects (using the same diagnostic tools).

❑ In the classical linear mixed model, the homogeneity of variances is validated using diagnostic plots. In certain contexts, when this assumption fails, the usual technique is to fit a new linear mixed model that assumes heterogeneous variances. Both models are compared using a formal test (Likelihood ratio test) or by AIC and BIC criteria.

❑ In the classical linear mixed model, the classical assumption of independence can also be validated by graphic tools. In certain contexts, the semivariogram of the residuals (a tool of spatial analysis) is used for validation of this assumption. When this assumption fails, the usual technique is to fit a new linear mixed model with covariance matrices which include correlation among observations.
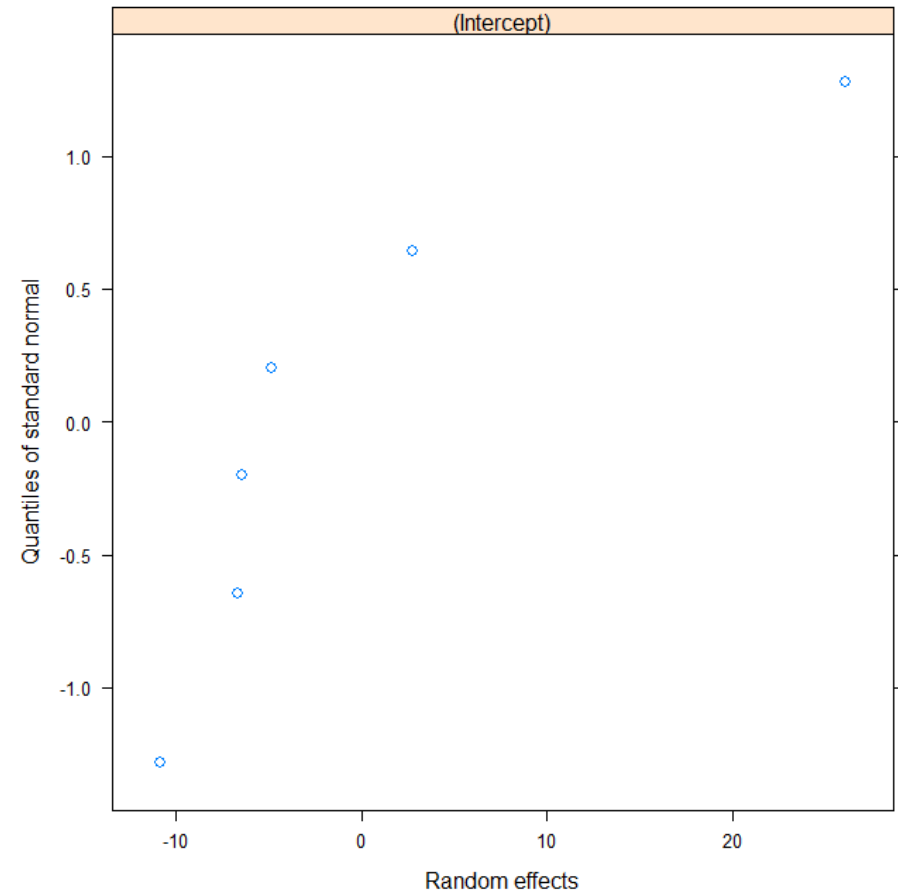
In R

> **oats.lme1<-lme(yield~Variety, data=oats,random=~1|Blocks)**
> **plot(oats.lme1)**

In ![R]
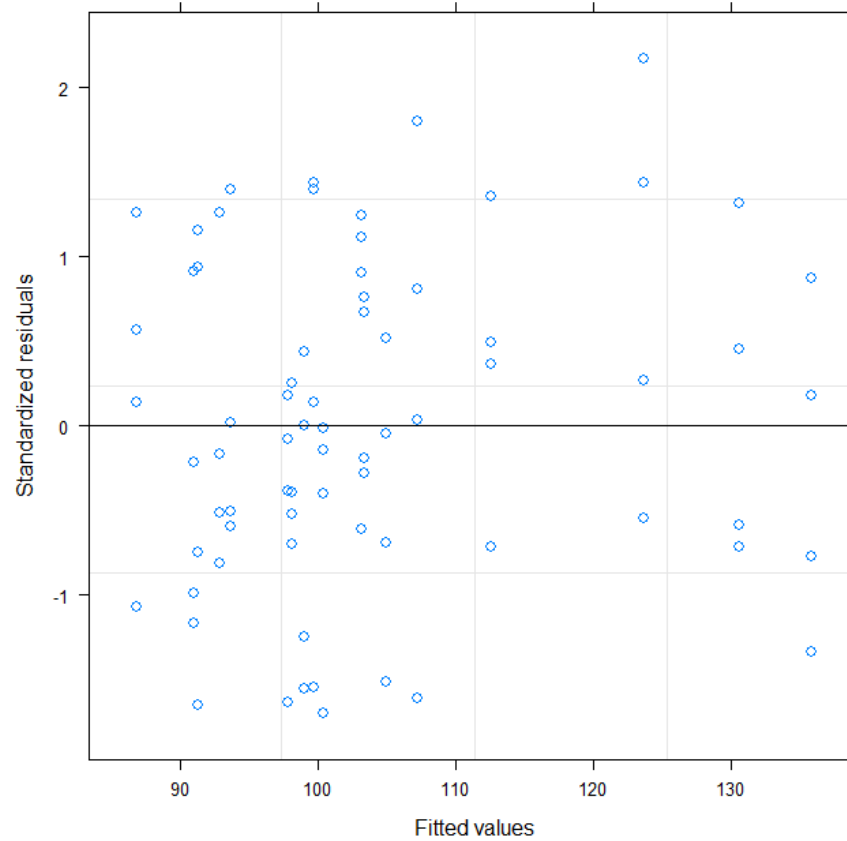
>qqnorm(oats.lme1, ~resid(.))   **>qqnorm(oats.lme1, ~ranef(.))**

☐ Cook's Distance can be studied using package "*influence.ME*" (requires to fit the model with the function *lmer* of the package *lme4).*