



Linear and Nonlinear Mixed Effects Models

Yuedong Wang

© All Rights Reserved

Department of Statistics and Applied Probability
University of California-Santa Barbara

Linear Mixed Effects Models



Outline I

1 Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2 Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3 Hierarchical Models and Bayes Estimation

4 lme, lmer and proc mixed

5 Examples

- Personnel data
- Spectrophotometer data
- Machine data



Outline II

- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Notations

Convention of notations:

- scalar: lower-case. Example, $x, y, z, \beta, b, \epsilon, \dots$
- vector: lower-case and boldface. Example, $\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\beta}, \mathbf{b}, \boldsymbol{\epsilon}, \dots$
- matrix: upper-case. Example, X, Z, W, G, \dots



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

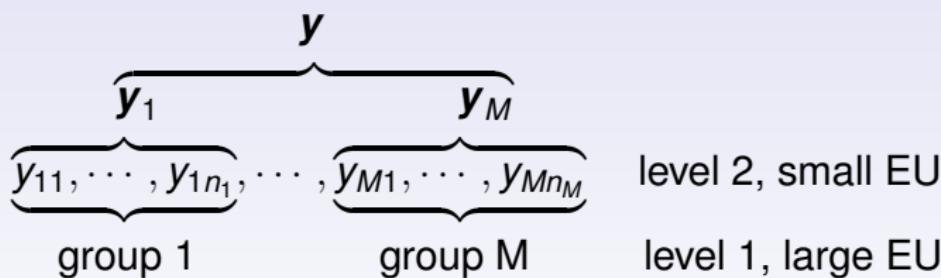
Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Single level of grouping

We first consider linear mixed effects (LME) models for a *single level of grouping*. Suppose that there are M groups and group i has n_i observations.





Scalar form

$$y_{ij} = \sum_{v=1}^p x_{ijv} \beta_v + \sum_{v=1}^k z_{ijv} b_{iv} + \epsilon_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, n_i$$

- y_{ij} : j th response in group i
- x_{ijv} : j th observation in group i on the v th covariate for the fixed effects
- β_v , $v = 1, \dots, p$: p fixed effects
- Z_{ijv} : j th observation in the group i on the v th covariate for random effects
- b_{iv} , $v = 1, \dots, k$: k random effects for group i .
Distributions will be specified jointly
- ϵ_{ij} : *within-group* random errors. Distributions will be specified jointly
- both fixed and random effects enter the model *linearly*



Vector form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M$$

- $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$ are responses in group i
- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, a p -dimensional vector of *fixed effects*
- $\mathbf{b}_i = (b_{i1}, \dots, b_{ik})^T$, a k -dimensional vector of *random effects*. $\mathbf{b}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 D)$. Note that they are group specific
- $\mathbf{X}_i = \{x_{ijv}\}$ is a $n_i \times p$ design matrix for the fixed effects
- $\mathbf{Z}_i = \{z_{ijv}\}$ is a $n_i \times k$ design matrix for the random effects
- $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T \sim \mathcal{N}(0, \sigma^2 R_i)$ are *within-group* random errors
- \mathbf{b}_i 's and $\boldsymbol{\epsilon}_i$'s are mutually independent



Stacked form

Let

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_M \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_M \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_M \end{pmatrix},$$

and

$$Z = \text{diag}(Z_1, \dots, Z_M) = \begin{pmatrix} Z_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & Z_M \end{pmatrix}$$



Stacked form

Then

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

- \mathbf{y} : a $n = \sum_{i=1}^M n_i$ dimensional vector of all responses
- \mathbf{X} : $n \times p$ design matrix for the fixed effects
- $\boldsymbol{\beta}$: p *fixed effects*
- \mathbf{Z} : $n \times q$ design matrix for the random effects where $q = Mk$
- \mathbf{b} : q *random effects*. $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 G)$ where

$$G = \text{diag} \underbrace{(\mathbf{D}, \dots, \mathbf{D})}_M$$

- $\boldsymbol{\epsilon}$: random errors. $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 R)$ where $R = \text{diag}(R_1, \dots, R_M)$



Example: personnel data continued

A company wants to compare ratings given by its personnel officers to potential employees. Four prospective employees were assigned at random to each of 5 randomly selected officers. The response is `rate`.

Officer	Candidate			
	1	2	3	4
A	76	64	85	75
B	58	75	81	66
C	49	63	62	46
D	74	71	85	90
E	66	74	81	79



Example: personnel data continued

4 rates from each of the 5 officers form a group. Thus $M = 5$, $n_i = 4$. We considered the one-way random effect model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

- μ : fixed effect
 - α_i : random effects $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_a^2)$
 - ϵ_{ij} : random errors, $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$
 - α_i and ϵ_{ij} are mutually independent
-
- Level 1: officer
 - Level 2: replications within officers



Example: personnel data continued

4 rates from each of the 5 officers form a group. Thus $M = 5$, $n_i = 4$. We considered the one-way random effect model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

- μ : fixed effect
 - α_i : random effects $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_a^2)$
 - ϵ_{ij} : random errors, $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$
 - α_i and ϵ_{ij} are mutually independent
-
- Level 1: officer
 - Level 2: replications within officers



Example: personnel data continued

The one-way random effect model can be written as

$$\mathbf{y}_i = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \alpha_i + \boldsymbol{\epsilon}_i.$$

Thus

$$\mathbf{X}_i = \mathbf{Z}_i = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

$\alpha_i \stackrel{iid}{\sim} N(0, \sigma_1^2)$, and $\boldsymbol{\epsilon}_i \stackrel{iid}{\sim} N(\mathbf{0}, \sigma^2 I_4)$.



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

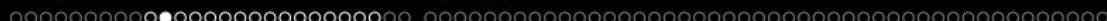
Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

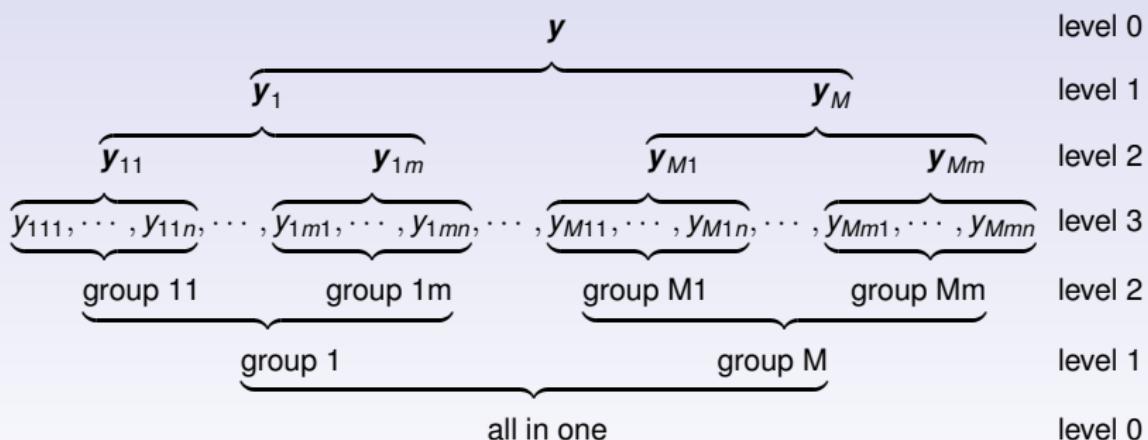
Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Two nested levels of grouping

The following diagram is for a balanced design with $n_{ij} = n$, $m_i = m$.

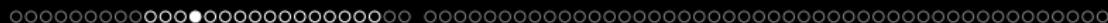




Vector form

$$\mathbf{y}_{ij} = X_{ij}\boldsymbol{\beta} + Z_{i,j}\mathbf{b}_i + Z_{ij}\mathbf{b}_{ij} + \epsilon_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, M_i$$

- $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijn_{ij}})^T$ are responses of the second-level group j within the first-level i
- $\boldsymbol{\beta}$, a p -dimensional vector of *fixed effects*
- \mathbf{b}_i , a k_1 -dimensional vector of the *first-level random effects*
 $\mathbf{b}_i \sim N(\mathbf{0}, \sigma^2 D_1)$
- \mathbf{b}_{ij} , a k_2 -dimensional vector of the *second-level random effects*.
 $\mathbf{b}_{ij} \sim N(\mathbf{0}, \sigma^2 D_2)$. Note that they are nested within i



Vector form

- X_{ij} is a $n_{ij} \times p$ design matrix for the fixed effects
- $Z_{i,j}$ is a $n_{ij} \times k_1$ design matrix for the first-level random effects
- Z_{ij} is a $n_{ij} \times k_2$ design matrix for the second-level random effects
- $\epsilon_{ij} = (\epsilon_{ij1}, \dots, \epsilon_{ijn_{ij}})^T \sim N(0, \sigma^2 R_{ij})$ are *within-group* random errors
- \mathbf{b}_i 's, \mathbf{b}_{ij} 's and ϵ_{ij} 's are mutually independent

oooooooooooo●oooooooooooooooooooo ooooooooooooooooooooooooooooooooooooo

Stacked form

Let

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_{11} \\ \vdots \\ \mathbf{y}_{1m_1} \\ \vdots \\ \mathbf{y}_{M1} \\ \vdots \\ \mathbf{y}_{Mm_M} \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} \mathbf{b}_{11} \\ \vdots \\ \mathbf{b}_{1m_1} \\ \vdots \\ \mathbf{b}_{M1} \\ \vdots \\ \mathbf{b}_{Mm_M} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_{11} \\ \vdots \\ \boldsymbol{\epsilon}_{1m_1} \\ \vdots \\ \boldsymbol{\epsilon}_{M1} \\ \vdots \\ \boldsymbol{\epsilon}_{Mm_M} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} X_{11} \\ \vdots \\ X_{1m_1} \\ \vdots \\ X_{M1} \\ \vdots \\ X_{Mm_M} \end{pmatrix},$$

$$\mathbf{u}_1 = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_M \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}, \quad Z_i = \begin{pmatrix} Z_{i,1} \\ \vdots \\ Z_{i,m_i} \end{pmatrix}, \quad i = 1, \dots, M$$

and

$$Z = (\text{diag}(Z_1, \dots, Z_M), \text{diag}(Z_{11}, \dots, Z_{1m_1}, \dots, Z_{M1}, \dots, Z_{Mm_M}))$$



Stacked form

Then

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

- \mathbf{y} : a $n = \sum_{i=1}^M \sum_{j=1}^{m_i} n_{ij}$ dimensional vector of all responses
- \mathbf{X} : $n \times p$ design matrix for the fixed effects
- $\boldsymbol{\beta}$: p *fixed effects*
- \mathbf{Z} : $n \times q$ design matrix for the random effects where

$$q = k_1 M + k_2 \times \sum_{i=1}^M m_i$$

- Note that the dimension of the random effects q can get large very quickly



Stacked form

- \mathbf{b} : q random effects. $\mathbf{b} \sim N(\mathbf{0}, \sigma^2 G)$ where

$$G = \text{diag}(\underbrace{D_1, \dots, D_1}_M, \underbrace{D_2, \dots, D_2}_{\sum_{i=1}^M m_i})$$

- ϵ : random errors. $\epsilon \sim N(\mathbf{0}, \sigma^2 R)$ where

$$R = \text{diag}(R_{11}, \dots, R_{1m_1}, \dots, R_{M1}, \dots, R_{Mm_M})$$



Example: machine data continued

A company wanted to replace the machines used to make a certain component. Three different brands of machines were available, so the investigators designed an experiment to evaluate the productivity of the machines when operated by the company's own personnel. 6 workers were randomly selected to participate in the experiment, each of whom was to operate each machine 3 different times. The response is an overall productivity score.



Example: machine data continued

Machine	Person	Score				
		All scores-Balanced			Partial scores-Unbalanced	
1	1	52.0	52.8	53.1	52.0	
1	2	51.8	52.8	53.1	51.8	52.8
1	3	60.0	60.2	58.4	60.0	
1	4	51.1	52.3	50.3	51.1	52.3
1	5	50.9	51.8	51.4	50.9	51.8
1	6	46.4	44.8	49.2	46.4	44.8
2	1	62.1	62.6	64.0		64.0
2	2	59.7	60.0	59.0	59.7	60.0
2	3	68.6	65.8	69.7	68.6	65.8
2	4	63.2	62.8	62.2	63.2	62.8
2	5	64.8	65.0	65.4	64.8	65.0
2	6	43.7	44.2	43.0	43.7	44.2
3	1	67.5	67.2	66.9	67.5	67.2
3	2	61.5	61.7	62.3	61.5	61.7
3	3	70.8	70.6	71.0	70.8	70.6
3	4	64.1	66.2	64.0	64.1	66.2
3	5	72.1	72.0	71.1	72.1	72.0
3	6	62.0	61.4	60.5	62.0	61.4



Example: machine data continued

There are two factors: machine (factor A) and worker (factor B). We considered the two-way mixed effects model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

- μ and α_i : fixed effects with $\sum_{i=1}^a \alpha_i = 0$
- β_j and $(\alpha\beta)_{ij}$: random effects with $\beta_j \stackrel{iid}{\sim} N(0, \sigma_b^2)$ and $(\alpha\beta)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{ab}^2)$
- ϵ_{ijk} : random errors with $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$
- β_j , $(\alpha\beta)_{ij}$ and ϵ_{ijk} are mutually independent



Example: machine data continued

- Level 1: worker
- Level 2: machine within worker (denoted by worker/machine)
- Level 3: replications

3 replicates from each worker/machine form a group at the second level, all 9 observations from each worker form a group at the first level.



Example: machine data continued

$$\mathbf{y}_{ij} = (y_{ij1}, y_{ij2}, y_{ij3})^T, \quad \boldsymbol{\epsilon}_{ij} = (\epsilon_{ij1}, \epsilon_{ij2}, \epsilon_{ij3})^T, \quad i = 1, \dots, 3; \quad j = 1, \dots, 6$$

Then

$$\mathbf{y}_{ij} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \alpha_i + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \beta_j + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (\alpha\beta)_{ij} + \boldsymbol{\epsilon}_{ij}$$

$$Z_{i,j} = Z_{ij} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Combine all deterministic parameters μ and α_i into a vector β , X_{ij} can be defined as in LMs.



Example: wheat data continued

An experiment was conducted to investigate how fertility (factor A) and variety (factor B) affect wheat yield. A field was divided into two blocks, each with four whole plots. Each of the four fertility levels was randomly assigned to one whole plot within each block. Each whole plot was split into two sub-plots, and each variety of wheat was randomly assigned to one sub-plot within each whole plot. Measurements at two levels of factor B in each whole plot are listed in each cell of the following table.

	A			
block	1	2	3	4
1	35.4 37.9	36.7 38.2	34.8 36.4	39.5 40.0
2	41.6 40.3	42.7 41.6	43.6 42.8	44.5 47.6



Example: wheat data continued

We assumed the following model

$$\begin{aligned}y_{ijk} = & \mu + \rho_i + \alpha_j + e_{ij} && \text{(whole plot part of model)} \\& + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk} && \text{(sub-plot part of model)}\end{aligned}$$

- y_{ijk} : observation from the i th block, fertility j and variety k
- α_j , β_k and $(\alpha\beta)_{jk}$: fixed effects, same as those in the two-way fixed effects models
- ρ_i (block effect) and e_{ij} (whole plot errors): random effects,
 $\rho_i \stackrel{iid}{\sim} N(0, \sigma_b^2)$ and $e_{ij} \stackrel{iid}{\sim} N(0, \sigma_e^2)$
- ϵ_{ijk} : sub-plot errors, $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$
- ρ_i , e_{ij} and ϵ_{ijk} are mutually independent



Example: wheat data continued

- Level 1: block
- Level 2: whole-plot
- Level 3: sub-plot

2 replicates from each block and fertility combination form a group at the second level, all 8 observations from each block form a group at the first level.



Example: wheat data continued

$$\mathbf{y}_{ij} = (y_{ij1}, y_{ij2})^T, \quad \boldsymbol{\epsilon}_{ij} = (\epsilon_{ij1}, \epsilon_{ij2})^T, \quad i = 1, 2; \quad j = 1, 4$$

Then

$$\mathbf{y}_{ij} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rho_i + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \alpha_j + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mathbf{e}_{ij} + \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} (\alpha\beta)_{j1} \\ (\alpha\beta)_{j2} \end{pmatrix} + \boldsymbol{\epsilon}_{ij}$$

$$Z_{i,j} = Z_{ij} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Combine all deterministic parameters μ, α_i, β_j and $(\alpha\beta)_{ij}$ into a vector $\boldsymbol{\beta}$, X_{ij} can be defined as in LMs.



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

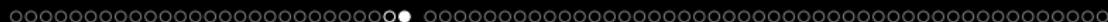
Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors

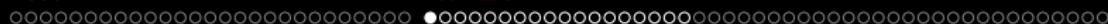


The general model

- All models we have discussed so far can be represented in the stacked form:

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

- The stacked form is very general and contains many more models
- The stacked form provides a unified framework for estimation, inference and software implementation



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

Review

- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Multivariate normal distribution

We will first review multivariate normal distribution, MLE, and linear model.

Suppose that \mathbf{y} is an n -dimensional multivariate normal random vector with mean μ and covariance matrix Σ , denoted by

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

\mathbf{y} has the density function

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left\{ \frac{1}{2} (\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu) \right\}$$



Multivariate normal distribution

- A is a $m \times n$ matrix, \mathbf{x} is a $m \times 1$ vector. Then

$$A\mathbf{y} + \mathbf{x} \sim N(A\mu + \mathbf{x}, A\Sigma A^T)$$

- Suppose that \mathbf{y} , μ and Σ are partitioned as follows:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Then both \mathbf{y}_1 and \mathbf{y}_2 are multivariate normal, and

$$\mathbf{y}_1 | \mathbf{y}_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$



Likelihood and maximum likelihood estimate

Denote $f(\mathbf{y}|\theta)$ as the density function of an observation vector \mathbf{y} with parameters θ . Regarded as a function of θ ,

$$L(\theta|\mathbf{y}) = f(\mathbf{y}|\theta)$$

is called the likelihood, and

$$l(\theta|\mathbf{y}) = \log L(\theta|\mathbf{y})$$

is called the log likelihood.

The maximum likelihood estimate (MLE) of θ , $\hat{\theta}$, is the maximizer of the likelihood.



Likelihood-based statistics

Score statistic: $u(\theta) = \frac{\partial l}{\partial \theta}$

Likelihood ratio statistic: let $\hat{\theta}$ and $\hat{\theta}_0$ be the MLE of θ in parameter spaces Θ and Θ_0 respectively, where $\Theta_0 \subset \Theta$. The likelihood ratio statistic is defined as

$$2\{l(\hat{\theta}|\mathbf{y}) - l(\hat{\theta}_0|\mathbf{y})\}$$

The Fisher information is

$$i(\theta) = \text{Var}\left(\frac{\partial l}{\partial \theta}\right) = -\mathbb{E}\left(\frac{\partial^2 l}{\partial \theta \partial \theta^T}\right)$$



Asymptotic results

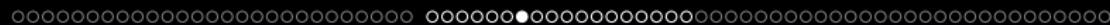
Under regularity conditions:

- $u(\theta) \sim N(\mathbf{0}, i(\theta))$
- $\hat{\theta} \sim N(\theta, i^{-1}(\theta))$
- the likelihood ratio statistic

$$2\{I(\hat{\theta}|\mathbf{y}) - I(\hat{\theta}_0|\mathbf{y})\} \sim \chi_d^2$$

where $d = \dim(\Theta) - \dim(\Theta_0)$

The symbol \sim means “approximately distributed as”. These asymptotic results are often used to conduct hypotheses tests and construct confidence intervals. Unknown parameters are usually replaced by their MLEs.



Likelihood ratio test

To test $H_0 : \theta \in \Theta_0$, the likelihood ratio statistic

$$2\{I(\hat{\theta}|\mathbf{y}) - I(\hat{\theta}_0|\mathbf{y})\} \stackrel{d}{\sim} \chi_d^2$$

where $d = \dim(\Theta) - \dim(\Theta_0)$.



Wald test

To test $H_0 : \theta = \theta_0$, the Wald statistic

$$(\hat{\theta} - \theta_0)^T i(\hat{\theta})(\hat{\theta} - \theta_0) \stackrel{d}{\sim} \chi_d^2$$

where d equals to the number of parameters being tested. θ does not have to include all parameters. In particular, when $d = 1$, the square root of the Wald test statistic is

$$z = \frac{(\hat{\theta}_k - \theta_{0k})}{se(\hat{\theta}_k)}$$



Score test

To test $H_0 : \theta = \theta_0$, the score statistic

$$u(\theta_0)^T i^{-1}(\theta_0) u(\theta_0) \stackrel{d}{\sim} \chi_d^2$$

where d equals to the number of parameters being tested.

The following link provides an excellent comparison of these three tests:

<https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/>

faqhow-are-the-likelihood-ratio-wald-and-lagrange-multiplier-score-tests-different-and-or-sim



Confidence intervals

The test procedures can be inverted to produce confidence sets. For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region at level α for the hypothesis $H_0 : \theta = \theta_0$. Then

$$C(\mathbf{y}) = \{\theta_0 : \mathbf{y} \in A(\theta_0)\}$$

is a $100(1 - \alpha)\%$ confidence set. Basically, it contains all parameters θ_0 such that $H_0 : \theta = \theta_0$ would not be rejected. For a single parameter θ_k , the confidence interval based on Wald test is

$$\hat{\theta}_k \pm z_{1-\alpha/2} se(\hat{\theta}_k)$$

Let $I(\theta_k | \mathbf{y})$ be the *profile likelihood* for θ_k , i.e. all other parameters are set to the MLE. The confidence interval based on the profile likelihood is

$$\{\theta_k : 2(I(\hat{\theta}_k | \mathbf{y}) - I(\theta_k | \mathbf{y})) < \chi^2_{1-\alpha}\}$$



Nuisance parameters

Suppose that θ is partitioned into two components

$$\theta = \begin{pmatrix} \phi \\ \lambda \end{pmatrix}$$

Suppose ϕ are parameters of interest and λ are nuisance parameters. Partition the joint Fisher information matrix and its inverse as follows

$$i(\theta) = \begin{pmatrix} i_{\phi\phi} & i_{\phi\lambda} \\ i_{\lambda\phi} & i_{\lambda\lambda} \end{pmatrix}, \quad i(\theta)^{-1} = \begin{pmatrix} i_{\phi\phi}^{-1} & i_{\phi\lambda} \\ i_{\lambda\phi} & i_{\lambda\lambda} \end{pmatrix}$$

Then

$$\{i_{\phi\phi}\}^{-1} = i_{\phi\phi} - i_{\phi\lambda} i_{\lambda\lambda}^{-1} i_{\lambda\phi}$$

is the approximate inverse covariance matrix of $\hat{\phi}$.



Linear model

We collect n observations on dependent variable y and independent variables x_1, \dots, x_p :

$$(y_i, x_{i1}, \dots, x_{ip}), \quad i = 1, \dots, n$$

The LM assumes that

$$y_i = \sum_{j=1}^p \beta_j x_{ij} + \epsilon_i,$$

where ϵ_i are random errors with mean zero and variance σ^2 .



The matrix form of the LM

Let

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{x}_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix},$$

and

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p) = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Then the LM model can be represented in the matrix form

$$\mathbf{y} = \sum_{j=1}^p \beta_j \mathbf{x}_j + \boldsymbol{\epsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$



LS estimation

We estimate parameters β by minimizing the least squares (LS)

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 = \|\mathbf{y} - \mathbf{X}\beta\|^2$$

The LS estimates of β is

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



Properties of the LS estimates

The LS estimate is unbiased

$$\mathbb{E}(\hat{\beta}) = \beta$$

with variance

$$\text{Var}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$$

BLUE (Gauss-Markov theorem)

For any p -dimensional vector \mathbf{c} , $\mathbf{c}^T \hat{\beta}$ is the unique best linear unbiased estimator (BLUE) of $\mathbf{c}^T \beta$ in the sense that it has the minimum variance (equivalently minimum mean squared error $\mathbb{E}(\mathbf{c}^T \hat{\beta} - \mathbf{c}^T \beta)^2$) in the class of linear unbiased estimators for $\mathbf{c}^T \beta$.



Properties of the LS estimates

The LS estimate is unbiased

$$E(\hat{\beta}) = \beta$$

with variance

$$\text{Var}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$$

BLUE (Gauss-Markov theorem)

For any p -dimensional vector \mathbf{c} , $\mathbf{c}^T \hat{\beta}$ is the unique best linear unbiased estimator (BLUE) of $\mathbf{c}^T \beta$ in the sense that it has the minimum variance (equivalently minimum mean squared error $E(\mathbf{c}^T \hat{\beta} - \mathbf{c}^T \beta)^2$) in the class of linear unbiased estimators for $\mathbf{c}^T \beta$.



Some remarks

- An unbiased estimator of σ^2 is

$$S^2 = \text{RSS}/(n - p)$$

- The LS estimation does not require random errors to be Gaussian. Inference (confidence interval and hypothesis test) do require random errors to be Gaussian
- When $\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$, the LS estimates are the MLE
- The MLE of σ^2 is

$$S^2 = \text{RSS}/n$$



t-tests and F-tests

For non-zero \mathbf{a} ,

$$t = \frac{\mathbf{a}^T \hat{\boldsymbol{\beta}} - \mathbf{a}^T \boldsymbol{\beta}}{\sqrt{S\{\mathbf{a}^T(X^T X)^{-1} \mathbf{a}\}}^{1/2}} \sim t_{n-p}$$

Replace $\mathbf{a}^T \boldsymbol{\beta}$ by c for testing the hypothesis

$$H_0 : \mathbf{a}^T \boldsymbol{\beta} = c$$

The F-statistic

$$F = \frac{(RSS_{H_0} - RSS)/q}{RSS/(n-p)} \stackrel{H_0}{\sim} F_{q,n-p}$$

is used to test the hypothesis

$$H_0 : A\boldsymbol{\beta} = \mathbf{c}$$

where A is a known $q \times p$ matrix of rank q , \mathbf{c} is a known $q \times 1$ vector, and RSS_{H_0} is the RSS under H_0 .



t-tests and F-tests

For non-zero \mathbf{a} ,

$$t = \frac{\mathbf{a}^T \hat{\boldsymbol{\beta}} - \mathbf{a}^T \boldsymbol{\beta}}{\sqrt{S\{\mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}\}}^{1/2}} \sim t_{n-p}$$

Replace $\mathbf{a}^T \boldsymbol{\beta}$ by c for testing the hypothesis

$$H_0 : \mathbf{a}^T \boldsymbol{\beta} = c$$

The F-statistic

$$F = \frac{(RSS_{H_0} - RSS)/q}{RSS/(n-p)} \stackrel{H_0}{\sim} F_{q,n-p}$$

is used to test the hypothesis

$$H_0 : A\boldsymbol{\beta} = \mathbf{c}$$

where A is a known $q \times p$ matrix of rank q , \mathbf{c} is a known $q \times 1$ vector, and RSS_{H_0} is the RSS under H_0 .



Weighted LS

Suppose now $E(\epsilon) = \mathbf{0}$ and $\text{Var}(\epsilon) = \sigma^2 W^{-1}$, where W is known. Then we estimate β using the Weighted LS (WLS)

$$WLS = (\mathbf{y} - X\beta)^T W (\mathbf{y} - X\beta)$$

Normal equation:

$$X^T W X \beta = X^T W \mathbf{y}$$

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



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Estimation and prediction for LME

The general form of linear models provides a unified framework for estimation, inference, and software implementation. The methods (e.g. LS, F test) and software (e.g. `proc glm, lm`) are applicable to many models, including regression, ANOVA, and ANCOVA.

We will consider the general form (*stacked form*) of linear mixed effects models, which provide a unified framework for estimation (e.g. MLE, REML, BLUP), inference (e.g. likelihood ratio test, F-test) and software implementation (e.g. `proc mixed` and `lme`). They are applicable to all mixed effects models discussed so far and many, many more.



Estimation and prediction for LME

The general form of linear models provides a unified framework for estimation, inference, and software implementation. The methods (e.g. LS, F test) and software (e.g. `proc glm, lm`) are applicable to many models, including regression, ANOVA, and ANCOVA.

We will consider the general form (*stacked form*) of linear mixed effects models, which provide a unified framework for estimation (e.g. MLE, REML, BLUP), inference (e.g. likelihood ratio test, F-test) and software implementation (e.g. `proc mixed` and `lme`). They are applicable to all mixed effects models discussed so far and many, many more.



Estimation and prediction for LME

The stacked form is clean which makes it easy to introduce methods for estimation and inference. However, it hides special structures in design matrices and covariance matrices. Thus computation and software implementation is usually based the vector form. Read section 2.2 of PB for formulations aimed at computation.

Stacked form of LME

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

- \mathbf{y} : a n dimensional vector of all responses
- \mathbf{X} : $n \times p$ design matrix for the fixed effects
- $\boldsymbol{\beta}$: p *fixed effects*
- \mathbf{Z} : $n \times q$ design matrix for the random effects
- \mathbf{b} : q *random effects*. $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 G)$
- $\boldsymbol{\epsilon}$: random errors. $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 R)$
- \mathbf{b} and $\boldsymbol{\epsilon}$ are independent

Suppose that G and R depend on a vector of parameters θ . No specific structure is assumed. We will discuss how to model the covariance structure, namely G and R , later.

We need to estimate the following parameters

$$\boldsymbol{\beta}, \ \sigma^2, \ \theta$$

Stacked form of LME

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

- \mathbf{y} : a n dimensional vector of all responses
- \mathbf{X} : $n \times p$ design matrix for the fixed effects
- $\boldsymbol{\beta}$: p *fixed effects*
- \mathbf{Z} : $n \times q$ design matrix for the random effects
- \mathbf{b} : q *random effects*. $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 G)$
- $\boldsymbol{\epsilon}$: random errors. $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 R)$
- \mathbf{b} and $\boldsymbol{\epsilon}$ are independent

Suppose that G and R depend on a vector of parameters θ . No specific structure is assumed. We will discuss how to model the covariance structure, namely G and R , later.

We need to estimate the following parameters

$$\boldsymbol{\beta}, \ \sigma^2, \ \theta$$



Likelihood

Note that

$$\mathbf{y} \sim \mathcal{N}(X\beta, \sigma^2 W^{-1})$$

where

$$W^{-1} = W^{-1}(\theta) = ZGZ^T + R$$

When necessary, the dependence of W on θ will be expressed explicitly. But most of the time we use W for simplicity.

The log-likelihood

$$\begin{aligned} I(\beta, \sigma^2, \theta | \mathbf{y}) &= \text{constant} - \frac{n}{2} \log \sigma^2 + \frac{1}{2} \log |W| \\ &\quad - \frac{1}{2\sigma^2} (\mathbf{y} - X\beta)^T W (\mathbf{y} - X\beta) \end{aligned}$$



Estimation of fixed effects

We first fix θ and solve for β and σ^2 :

$$\frac{\partial l}{\partial \beta} = X^T W (\mathbf{y} - X\beta) = \mathbf{0}$$

leads to

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

Note: $\hat{\beta}$ is the same as the WLS estimate.



Estimation of variance and covariance parameters

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T W (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

leads to

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T W (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Plugging back, we have the profiled likelihood

$$l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \theta | \mathbf{y}) = \text{constant} + \frac{1}{2} \log |W| - \frac{n}{2} \log [(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T W (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})]$$

The MLE of θ , $\hat{\theta}$, is the maximum of $l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \theta | \mathbf{y})$.

Usually $\hat{\theta}$ does not have a closed form. One needs to compute it numerically, which is usually achieved through an iterative procedure such as EM algorithm, Newton-Raphson algorithm or a combination of these two. Read Ch.8 of SCM.



MLE estimates

Because

$$\mathbb{E} \left(\frac{\partial^2 I}{\partial \beta \partial \sigma^2} \right) = \mathbf{0}, \quad \mathbb{E} \left(\frac{\partial^2 I}{\partial \beta \partial \theta^T} \right) = \mathbf{0},$$

the information matrix is block diagonal, thus $\hat{\beta}$ is asymptotically independent of $\hat{\sigma}^2$ and $\hat{\theta}$



EM algorithm

It is an iterative procedure for computing MLE with incomplete data such as missing or censored data. It is a broadly applicable technique rather than a specific algorithm. It is based on the observation (or assumption) that computing MLE with the complete data is simple, often with a closed form.

We now describe the EM algorithm in general setting, and then we show how it applies to LME.

Let \mathbf{z}_o be the observed data, \mathbf{z}_m be the missing ("missing" in the general sense which could be censored or unobservable) data, and $\mathbf{z}^T = (\mathbf{z}_o^T, \mathbf{z}_m^T)$ be the complete data. Let $g_c(\mathbf{z}|\xi)$ be the density function of \mathbf{z} with parameters ξ . Then the complete-data log-likelihood

$$l_c(\xi|\mathbf{z}) = \log g_c(\mathbf{z}|\xi)$$



EM algorithm

Often maximizing I_c is simple, but not applicable since part of \mathbf{z} is not observed. The trick is to replace I_c by its conditional expectation given \mathbf{z}_o , using the current estimates of parameters.

Let $\xi^{(k)}$ be the estimates at the k iteration. At iteration $k + 1$, the EM algorithm consists of two steps:

E-Step or Expectation step. Compute

$$Q(\xi, \xi^{(k)}) = E_{\xi^{(k)}}(I_c(\xi | \mathbf{z}) | \mathbf{z}_o)$$

M-Step or Maximization step. Choose $\xi^{(k+1)}$ that maximizes $Q(\xi, \xi^{(k)})$

E- and M-steps are alternated repeatedly until convergence.

Basic property: The likelihood increases with iterations and the likelihood values converge to stationary values.



Application of the EM algorithm to LME

In the stacked form model, random effects \mathbf{b} are unobservable.
Consider it as missing, we have

$$\mathbf{z}_o = \mathbf{y}, \quad \mathbf{z}_m = \mathbf{b}, \quad \mathbf{z} = (\mathbf{y}^T, \mathbf{b}^T)^T$$

Let $\xi = (\beta^T, \sigma^2, \theta^T)^T$. Since

$$\mathbf{z} \sim N\left(\begin{pmatrix} X\beta \\ \mathbf{0} \end{pmatrix}, \sigma^2 \begin{pmatrix} W^{-1} & ZG \\ GZ^T & G \end{pmatrix}\right)$$

then

$$I_c(\xi|\mathbf{z}) = \text{constant} - \frac{1}{2} \log |\Sigma| - (\mathbf{y} - X\beta, \mathbf{b})^T \Sigma^{-1} (\mathbf{y} - X\beta, \mathbf{b})$$

where

$$\Sigma = \sigma^2 \begin{pmatrix} W^{-1} & ZG \\ GZ^T & G \end{pmatrix}$$



Application of the EM algorithm to LME

The conditional expectation at the E-step requires the following quantities

$$E_{\xi^{(k)}}(\mathbf{b}|\mathbf{y}) \text{ and } E_{\xi^{(k)}}(\mathbf{b}\mathbf{b}^T|\mathbf{y})$$

Since both \mathbf{y} and \mathbf{b} are multivariate normal, $\mathbf{b}|\mathbf{y}$ is multivariate normal with mean and covariance matrix depend on current estimates.

Note that the density $g_c(\mathbf{z})$ belongs to the exponential family, the computation is usually based on conditional expectations of the complete-data sufficient statistics for σ^2 and θ . The computation depends on special covariance structures. More detail can be found in Laird and Ware (1982).



Newton-Raphson method

Let $I(\theta) = I(\hat{\beta}, \hat{\sigma}^2, \theta | \mathbf{y})$. We need to solve

$$u(\theta) \stackrel{\triangle}{=} \partial I(\theta) / \partial \theta = \mathbf{0}$$

Let $\theta^{(k)}$ be the estimates at the k iteration. At iteration $k + 1$, approximate the left hand side of above equation by a linear Taylor series expansion about current fit $\theta^{(k)}$

$$u(\theta) \approx u(\theta^{(k)}) + H(\theta^{(k)})(\theta - \theta^{(k)}) = \mathbf{0}$$

where

$$H(\theta) = \partial^2 I(\theta) / \partial \theta \partial \theta^T$$

Then we have a new update

$$\theta^{(k+1)} = \theta^{(k)} - H^{-1}(\theta^{(k)})u(\theta^{(k)})$$



Comparison and hybrid algorithm

EM algorithm

- computations at each iteration are simple and usually fast
- converge to the region of the optimum very quickly
- progress to the optimum slow when near the optimum

Newton-Raphson

- quadratic convergence rate
- computation of Hessian at each iteration may be expensive
- may be unstable when the initial values are far away from the optimum

Hybrid algorithm

Use EM algorithm for a few iterations and then switch to the Newton-Raphson algorithm. The default in `lme` is 25 EM iterations and this number can be modified by the `control` option.



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- **Restricted maximum likelihood estimation**
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

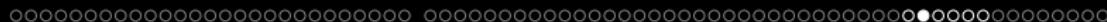
Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Restricted maximum likelihood (REML)

One problem with the MLEs of variance components, σ^2 and θ , is that they may have large bias when the number of fixed effects p is large compare to n .

The easiest way to appreciate this problem is through the standard LM where random errors are iid $N(0, \sigma^2)$. Recall that the MLE of σ^2 is

$$\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta})/n$$

$\hat{\sigma}^2$ is biased downward since

$$E(\hat{\sigma}^2) = \frac{n-p}{n} \sigma^2.$$

The bias may be large when p is relatively large. A unbiased estimate is

$$\hat{\sigma}_R^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta})/(n-p)$$

which turns out to be a REML estimate.



Restricted maximum likelihood (REML)

The reason for the bias in MLE is that the loss of degrees of freedom due to estimating β, p , is ignored. Thus to correct the bias problem one needs to account for the loss of degrees of freedom resulting from estimating fixed effects.

One way to think of REML estimation is the following two steps

- ① eliminate fixed effects part of contribution to the likelihood
- ② estimate variance components based on the reduced data



Restricted maximum likelihood (REML)

Recall that

$$\mathbf{y} \sim N(X\beta, \sigma^2 W^{-1})$$

Let O be a $(n - p) \times n$ matrix of full column rank and

$$OX = \mathbf{0}$$

Raws of O are $n - p$ linearly independent contrasts. Then

$$O\mathbf{y} \sim N(\mathbf{0}, \sigma^2 OW^{-1} O^T)$$

which is independent of β .

$O\mathbf{y}$ is the reduced data. For different choices of O matrix, the likelihood function of $O\mathbf{y}$ differ by a constant independent of variance components. Thus any matrix satisfying above conditions will work.



Restricted maximum likelihood (REML)

An obvious choice of O is any $n - p$ raws of the projection matrix

$$P = I - X(X^T X)^{-1} X^T$$

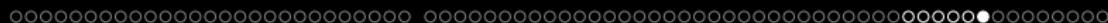
which has full column rank. Let

$$X = (Q_1 \ Q_2) \begin{pmatrix} R \\ \mathbf{0} \end{pmatrix}$$

be the QR decomposition of X . Another obvious choice of O is Q_2^T .

It can be verified that the log-restricted likelihood of Oy for σ^2 and θ is

$$\begin{aligned} l_R(\hat{\sigma}^2, \theta | \mathbf{y}) &= \text{constant} - \frac{n-p}{2} \log \sigma^2 + \frac{1}{2} \log |W| \\ &\quad - \frac{1}{2} \log |X^T W^{-1} X| + \frac{1}{2\sigma^2} (\mathbf{y} - X\beta)^T W (\mathbf{y} - X\beta) \end{aligned}$$



Restricted maximum likelihood (REML)

- For details read Diggle, Liang and Zeger (1994)
- Compare to the full likelihood, a correction term

$$\frac{1}{2} \log |X^T W^{-1} X|$$

is added

- Taking derivative with respect to σ^2 and setting to zero, we have

$$\hat{\sigma}_R^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})^T W(\mathbf{y} - \mathbf{X}\hat{\beta})/(n - p)$$

- Both MLE and REML estimates are consistent and asymptotically normal



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- **Prediction of random effects**

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Prediction of random effects

So far we considered estimations of the deterministic parameters β , σ^2 and θ only. Sometimes we are interested in quantities involving random effects. For example, for the orthodont data, each subject (or their parents or doctors) is more interested in his/her own trajectory. Other applications.

- animal breeding - breeding values of individual animals for selection of candidates for the future breeding program
- ranking plants or animals using their genetic traits
- predicting quantities where no observations were taken, especially spatially - Kriging is essentially the same as BLUP

Traditionally, estimation of the random effects is called prediction.



Henderson likelihood

Assume that σ^2 and θ are known for the moment. We want to estimate β and \mathbf{b} . Regard \mathbf{b} as parameters. Consider the joint likelihood of \mathbf{b} and \mathbf{y}

$$\begin{aligned} h(\beta, \mathbf{b} | \mathbf{b}, \mathbf{y}) &= \text{constant} + \frac{1}{2} \log |W| - \frac{1}{2\sigma^2} [\mathbf{b}^T G^{-1} \mathbf{b} \\ &\quad + (\mathbf{y} - X\beta - Z\mathbf{b})^T R^{-1} (\mathbf{y} - X\beta - Z\mathbf{b})] \end{aligned}$$

Differentiating with respect to β and \mathbf{b} , we have the following mixed model equations

$$\begin{aligned} X^T R^{-1} X \beta + X^T R^{-1} Z \mathbf{b} &= X^T R^{-1} \mathbf{y} \\ Z^T R^{-1} X \beta + (Z^T R^{-1} Z + G^{-1}) \mathbf{b} &= Z^T R^{-1} \mathbf{y} \end{aligned}$$



Henderson likelihood

Solutions to the mixed model equations are

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

$$\hat{\mathbf{b}} = (Z^T R^{-1} Z + G^{-1}) Z^T R^{-1} [I - X(X^T W X)^{-1} X^T W] \mathbf{y}$$

Remarks

- Henderson likelihood is not the standard likelihood: \mathbf{b} appeared as both parameter and observations, and is actually not observable at all
- $\hat{\beta}$ is the same as the MLE
- $G^{-1} \rightarrow \mathbf{0}$ corresponds to the usual normal equations for β and \mathbf{b} when \mathbf{b} are fixed effects. See the Bayesian formulation later



BLUP: extension of the Gauss-Markov theorem

For any p -dimensional vector \mathbf{c}_1 and q -dimensional vector \mathbf{c}_2 ,
 $\mathbf{c}_1^T \hat{\beta} + \mathbf{c}_2^T \hat{\mathbf{b}}$ is the *best linear unbiased prediction* (BLUP) of
 $\mathbf{c}_1^T \beta + \mathbf{c}_2^T \mathbf{b}$ in the sense that it has the minimum mean squared error

$$\mathbb{E}(\mathbf{c}_1^T \hat{\beta} + \mathbf{c}_2^T \hat{\mathbf{b}} - \mathbf{c}_1^T \beta - \mathbf{c}_2^T \mathbf{b})^2$$

in the class of linear unbiased estimators for $\mathbf{c}_1^T \beta + \mathbf{c}_2^T \mathbf{b}$.

Remarks

- *linear* means that they are linear in \mathbf{y}
- *unbiased* means that

$$\mathbb{E}(\mathbf{c}_1^T \hat{\beta} + \mathbf{c}_2^T \hat{\mathbf{b}}) = \mathbb{E}(\mathbf{c}_1^T \beta + \mathbf{c}_2^T \mathbf{b})$$

Especially, $\mathbb{E}(\hat{\mathbf{b}}) = \mathbb{E}(\mathbf{b})$, not $\mathbb{E}(\hat{\mathbf{b}}|\mathbf{b}) = \mathbf{b}$ for all \mathbf{b} . On average, we get the average of the moving target, not specific realizations



BLUP: extension of the Gauss-Markov theorem

For any p -dimensional vector \mathbf{c}_1 and q -dimensional vector \mathbf{c}_2 , $\mathbf{c}_1^T \hat{\beta} + \mathbf{c}_2^T \hat{\mathbf{b}}$ is the *best linear unbiased prediction* (BLUP) of $\mathbf{c}_1^T \beta + \mathbf{c}_2^T \mathbf{b}$ in the sense that it has the minimum mean squared error

$$E(\mathbf{c}_1^T \hat{\beta} + \mathbf{c}_2^T \hat{\mathbf{b}} - \mathbf{c}_1^T \beta - \mathbf{c}_2^T \mathbf{b})^2$$

in the class of linear unbiased estimators for $\mathbf{c}_1^T \beta + \mathbf{c}_2^T \mathbf{b}$.

Remarks

- *linear* means that they are linear in \mathbf{y}
- *unbiased* means that

$$E(\mathbf{c}_1^T \hat{\beta} + \mathbf{c}_2^T \hat{\mathbf{b}}) = E(\mathbf{c}_1^T \beta + \mathbf{c}_2^T \mathbf{b})$$

Especially, $E(\hat{\mathbf{b}}) = E(\mathbf{b})$, not $E(\hat{\mathbf{b}}|\mathbf{b}) = \mathbf{b}$ for all \mathbf{b} . On average, we get the average of the moving target, not specific realizations



Proof



$$\begin{aligned} E(\mathbf{y}\mathbf{y}^T) &= X\beta\beta^T X^T + \sigma^2 W^{-1} \\ E(\mathbf{b}\mathbf{y}^T) &= \sigma^2 G Z^T \end{aligned}$$

- Linear unbiased estimates of zero are of the form $\mathbf{a}^T \mathbf{y}$ where \mathbf{a} satisfies $X^T \mathbf{a} = 0$
- $\mathbf{a}^T \mathbf{y}$ are uncorrelated with $\hat{\beta}$ and $\hat{\mathbf{b}}$.

$$\begin{aligned} &E(\hat{\beta} - \beta)\mathbf{y}^T \mathbf{a} \\ &= (X^T W X)^{-1} X^T W E(\mathbf{y}^T \mathbf{y}) \mathbf{a} - \beta E(\mathbf{y}^T) \mathbf{a} \\ &= \beta \beta^T X^T \mathbf{a} + \sigma^2 (X^T W X)^{-1} X^T \mathbf{a} - \beta \beta^T X^T \mathbf{a} \\ &= \mathbf{0} \\ &E(\hat{\mathbf{b}} - \mathbf{b})\mathbf{y}^T \mathbf{a} \\ &= (Z^T R^{-1} Z + G^{-1}) Z^T R^{-1} [I - X(X^T W X)^{-1} X^T W] E(\mathbf{y}\mathbf{y}^T) \mathbf{a} - E(\mathbf{b}\mathbf{y}^T) \mathbf{a} \\ &= (Z^T R^{-1} Z + G^{-1}) Z^T R^{-1} [I - X(X^T W X)^{-1} X^T W] (X^T \beta \beta^T X + \sigma^2 W^{-1}) \mathbf{a} - \sigma^2 G Z^T \mathbf{a} \\ &= (Z^T R^{-1} Z + G^{-1}) Z^T R^{-1} (X^T \beta \beta^T X + \sigma^2 W^{-1} - X^T \beta \beta^T X - \sigma^2 X(X^T W X)^{-1} X^T) \mathbf{a} - \sigma^2 G Z^T \mathbf{a} \\ &= \sigma^2 (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} (Z^T G Z + R) \mathbf{a} - \sigma^2 G Z^T \mathbf{a} \\ &= \sigma^2 (Z^T R^{-1} Z + G^{-1})^{-1} (Z^T R^{-1} Z + G^{-1}) G Z^T \mathbf{a} - \sigma^2 G Z^T \mathbf{a} \\ &= \mathbf{0} \end{aligned}$$



Proof

- Any linear unbiased estimator of $\mathbf{c}_1^T \boldsymbol{\beta} + \mathbf{c}_2^T \mathbf{b}$ must be of the form $\mathbf{c}_1^T \hat{\boldsymbol{\beta}} + \mathbf{c}_2^T \hat{\mathbf{b}} + \mathbf{a}^T \mathbf{y}$
- The MSE of $\mathbf{c}_1^T \hat{\boldsymbol{\beta}} + \mathbf{c}_2^T \hat{\mathbf{b}} + \mathbf{a}^T \mathbf{y}$

$$\begin{aligned} & \mathbb{E}[\mathbf{c}_1^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \mathbf{c}_2^T(\hat{\mathbf{b}} - \mathbf{b}) + \mathbf{a}^T \mathbf{y}]^2 \\ = & \mathbb{E}[\mathbf{c}_1^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \mathbf{c}_2^T(\hat{\mathbf{b}} - \mathbf{b})]^2 + \mathbb{E}(\mathbf{a}^T \mathbf{y} \mathbf{y}^T \mathbf{a}) \end{aligned}$$

The second part is always greater or equal to zero. Thus $\mathbf{c}_1^T \hat{\boldsymbol{\beta}} + \mathbf{c}_2^T \hat{\mathbf{b}}$ has smaller MSE than $\mathbf{c}_1^T \hat{\boldsymbol{\beta}} + \mathbf{c}_2^T \hat{\mathbf{b}} + \mathbf{a}^T \mathbf{y}$.



Variance-covariance

Assume that X is of full column rank. Then

$$\mathbb{E} \left(\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{b}} - \mathbf{b} \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{b}} - \mathbf{b} \end{pmatrix}^T \right) = \sigma^2 \begin{pmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{pmatrix}^{-1}$$

- The variance of $\hat{\mathbf{b}} - \mathbf{b}$, not the variance of $\hat{\mathbf{b}}$, will be used to assess the estimation error in $\hat{\mathbf{b}}$ because it also accounts for the variation in \mathbf{b}
- The variance-covariance formula may be used to construct confidence intervals
- Unknown variance components are replaced by their MLE or REML estimates



Hierarchical model

Mixed effects models are closely related to hierarchical models which are strongly tied to Bayesian methods.

A *hierarchical model* is specified in stages, with each stage building upon another. Advantages:

- model in each stage is relatively simple and easy to understand
- the entire model can be rather complicated, thus flexible
- there is no limit to the number of stages

See Chapter 9 of SCM for more details.



Bayes method

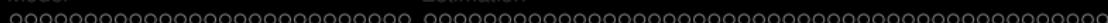
Suppose we observe a vector of random variables \mathbf{y} with distribution $f(\mathbf{y}|\xi)$ where ξ is a vector of parameters. The Bayes approach specifies a prior distribution, $\pi(\xi)$, for ξ and make inference based on the posterior distribution

$$f(\xi|\mathbf{y}) = \frac{f(\mathbf{y}|\xi)\pi(\xi)}{\int f(\mathbf{y}|\xi)\pi(\xi)d\xi}$$

For example, one may use the posterior mean

$$\text{E}(\xi|\mathbf{y}) = \frac{\int \xi f(\mathbf{y}|\xi)\pi(\xi)d\xi}{\int f(\mathbf{y}|\xi)\pi(\xi)d\xi}$$

or posterior mode as a point estimate of ξ .



Bayes method

The specification of sampling distribution and prior distribution is an example of a hierarchical model with two levels of hierarchy:

$$\begin{aligned} \mathbf{y}|\boldsymbol{\xi} &\sim f(\mathbf{y}|\boldsymbol{\xi}) && \text{first level} \\ \boldsymbol{\xi} &\sim \pi(\boldsymbol{\xi}) && \text{second level} \end{aligned}$$

One may have more than two levels of hierarchy:

$$\begin{aligned} \mathbf{y}|\boldsymbol{\xi} &\sim f(\mathbf{y}|\boldsymbol{\xi}) && \text{first level} \\ \boldsymbol{\xi} &\sim \pi(\boldsymbol{\xi}|\boldsymbol{\lambda}) && \text{second level} \\ \boldsymbol{\lambda} &\sim \pi(\boldsymbol{\lambda}|\boldsymbol{\delta}) && \text{third level} \\ \boldsymbol{\delta} &\sim \pi(\boldsymbol{\delta}|\boldsymbol{\phi}) && \text{fourth level} \\ &\vdots && \end{aligned}$$

The mixed model hierarchy

For the LME, consider the following hierarchical model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \quad \text{first level}$$

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma^2 \mathbf{B}), \quad \mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{G}), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R}) \quad \text{second level}$$

- the first level (stage) model assumes that given $\boldsymbol{\beta}$ and \mathbf{b} , the responses follow a general linear model
- at the second level, $\boldsymbol{\beta}$ as well as \mathbf{b} and $\boldsymbol{\epsilon}$ are assumed random and mutually independent. Thus there is no distinction between *fixed* and *random*. It only matters whether a quantity is *observable* (e.g. \mathbf{y}) or *unobservable* (e.g. $\boldsymbol{\beta}$, \mathbf{b} and $\boldsymbol{\epsilon}$).

$$\begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{b} \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \\ X\boldsymbol{\beta}_0 \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{B} & \mathbf{0} & \mathbf{BX}^T \\ \mathbf{0} & \mathbf{G} & \mathbf{GZ}^T \\ XB & ZG & XBX^T + ZGZ^T + R \end{pmatrix}\right)$$



Estimating β

Assume σ^2 and θ are known.

$$\beta | \mathbf{y} \sim N((X^T W X + B^{-1})^{-1} (X^T W \mathbf{y} + B^{-1} \beta_0), (X^T W X + B^{-1})^{-1})$$

Typically we don't have prior information about β . A reasonable strategy is to let $B^{-1} = \mathbf{0}$, a noninformative prior. In this case,

$$E(\beta | \mathbf{y}) = (X^T W X)^{-1} X^T W \mathbf{y} = \hat{\beta}$$

Note: The posterior mean with the noninformative priors the same as the BLUE.

Estimating (predicting) \boldsymbol{b}

Assume σ^2 and θ are known. It can be shown that

$$\boldsymbol{b}|\boldsymbol{y} \sim N((Z^T K Z + G^{-1})^{-1} Z^T K (\boldsymbol{y} - X\beta_0), (Z^T K Z + G)^{-1})$$

where

$$K = (R + XBX^T)^{-1}$$

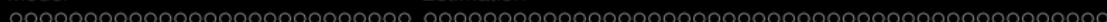
Again, assuming vague prior $B^{-1} = \mathbf{0}$, it may be shown that

$$E(\boldsymbol{b}|\boldsymbol{y}) = GZ^T W(\boldsymbol{y} - X\hat{\beta}) = \hat{\boldsymbol{b}}$$

Note: The posterior mean with noninformative prior is the same as the BLUP. This provides another justification for the BLUP.

Proof (details read Ch.9 of SCM)

- 1 $X\hat{\beta} = X(X^T WX)^{-1}X^T W\boldsymbol{y}$
- 2 $(Z^T R^{-1}Z + G^{-1})GZ^T W = Z^T R^{-1}(ZGZ^T + R)W = Z^T R^{-1}$. Thus
 $GZ^T W = (Z^T R^{-1}Z + G^{-1})^{-1}Z^T R^{-1}$.
- 3 $\hat{\boldsymbol{b}} = (Z^T R^{-1}Z + G^{-1})^{-1}Z^T R^{-1}(\boldsymbol{y} - X(X^T WX)^{-1}X^T W\boldsymbol{y}) = GZ^T W(\boldsymbol{y} - X\hat{\beta})$



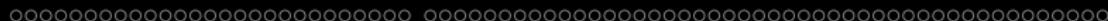
Estimating σ^2 and θ

A full Bayesian approach would specify priors for these parameters and then proceed to compute posterior means or modes.

- Difficult to specify priors
- Difficult to compute posteriors
- Computations usually involves high-dimension integrations which do not have closed forms. Advanced techniques such as Gibbs sampling and MCMC are used

Empirical Bayes

Empirical Bayes here refers to using marginal distributions to estimate parameters and substitute these parameters for their corresponding parameters in a formal Bayes estimator.



Empirical Bayes

Outline of the general strategy for obtaining the empirical Bayes estimate of parameter τ :

- ① Specify a distribution $\pi(\tau|\eta)$ for τ where η represents parameters known as *hyperparameters*
- ② Calculate the posterior of τ

$$\pi(\tau|y, \eta) = \frac{f(y|\tau, \eta)\pi(\tau|\eta)}{\int f(y|\tau, \eta)\pi(\tau|\eta)d\tau}$$

and use it to estimate τ , say posterior mode $\hat{\tau}(\eta)$ where the dependence on η is made explicit

- ③ Estimate the hyperparameter η using the marginal distribution

$$m(y|\eta) = \int f(y|\tau, \eta)\pi(\tau|\eta)d\tau$$

- ④ Denote the estimate as $\hat{\eta}$
- ④ The empirical Bayes estimate of τ is $\hat{\tau}(\hat{\eta})$



Empirical Bayes for LME

For the hierarchical model, let $\tau = (\sigma^2, \theta^T)^T$. Assume

- a flat (noninformative) prior

$$\pi(\tau|\eta) = 1$$

- $f(\mathbf{y}|\tau, \eta) = f(\mathbf{y}|\tau)$ and $\int f(\mathbf{y}|\tau)d\tau < \infty$

Then the posterior

$$\pi(\tau|\mathbf{y}, \eta) = \frac{f(\mathbf{y}|\tau)\pi(\tau|\eta)}{\int f(\mathbf{y}|\tau)\pi(\tau|\eta)d\tau} \propto f(\mathbf{y}|\tau)$$

Thus the modes of $\pi(\tau|\mathbf{y}, \eta)$ and $f(\mathbf{y}|\tau)$ are the same.

$$f(\mathbf{y}|\tau) = \int f(\mathbf{y}|\beta, \tau)f(\beta)d\beta$$

The form $f(\mathbf{y}|\tau)$ depends on the specification of the prior $f(\beta)$. We now consider two different specifications which lead to MLE and REML respectively.



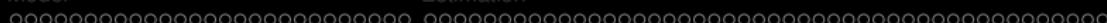
Connection to MLE

Assume a point-mass prior density for β . This is equivalent to

$$\begin{aligned} \mathbf{y} | \boldsymbol{\beta}, \mathbf{b} &\sim N(X\boldsymbol{\beta} + Z\mathbf{b}, \sigma^2 R) \\ \mathbf{b} &\sim N(\mathbf{0}, \sigma^2 G) \end{aligned}$$

which is the standard LME. Thus $\mathbf{y} \sim N(X\boldsymbol{\beta}, \sigma^2 W^{-1})$ and $f(\mathbf{y} | \tau)$ is the likelihood function.

Note: The MLEs are empirical Bayes estimators with a point-mass prior.



Connection to REML

Assume a flat prior for β

$$\begin{aligned}\mathbf{y}|\beta, \mathbf{b} &\sim N(X\beta + Z\mathbf{b}, \sigma^2 R) \\ \beta \sim f(\beta) = 1, \quad \mathbf{b} &\sim N(\mathbf{0}, \sigma^2 G)\end{aligned}$$

It can be shown that (for details read SCM section 9.2)

$$\log f(\mathbf{y}|\tau) = I_R(\tau|\mathbf{y})$$

Note: The REML estimates of variance components are also empirical Bayes estimates with a noninformative prior. This provides another justification for REML.



Examples

In this section we show how to fit various mixed effects models using R functions `lme` and `lmer`, and the SAS procedure `proc mixed`.

We will illustrate using previous examples as well as more complicated examples.



The lme function

Linear Mixed-Effects Models

Description:

This generic function fits a linear mixed-effects model in the formulation described in Laird and Ware (1982) but allowing for nested random effects. The within-group errors are allowed to be correlated and/or have unequal variances.

Usage:

```
lme(fixed, data, random, correlation, weights, method)
```



The lme function

Arguments:

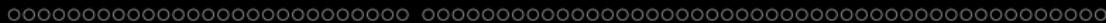
fixed: a two-sided linear formula object describing the fixed-effects part of the model, with the response on the left of a `~` operator and the terms, separated by `+` operators, on the right, an `lmList` object, or a `groupedData` object. The method functions `lme.lmList` and `lme.groupedData` are documented separately.

data: an optional data frame containing the variables named in `fixed`, `random`, `correlation`, `weights`, and `subset`. By default the variables are taken from the environment from which `lme` is called.



The lme function

random: optionally, any of the following: (i) a one-sided formula of the form $\sim x_1 + \dots + x_n \mid g_1 / \dots / g_m$, with $x_1 + \dots + x_n$ specifying the model for the random effects and $g_1 / \dots / g_m$ the grouping structure (m may be equal to 1, in which case no $/$ is required). The random effects formula will be repeated for all levels of grouping, in the case of multiple levels of grouping; (ii) a list of one-sided formulas of the form $\sim x_1 + \dots + x_n \mid g$, with possibly different random effects models for each grouping level. The order of nesting will be assumed the same as the order of the elements in the list; (iii) a one-sided formula of the form $\sim x_1 + \dots + x_n$, or a pdMat object with a formula (i.e. a non-NULL value for `formula(object)`), or a list of such formulas or pdMat objects. In this case, the grouping structure formula will be derived from the data used to fit the linear mixed-effects model, which should inherit from class



The lme function

groupedData; (iv) a named list of formulas or pdMat objects as in (iii), with the grouping factors as names. The order of nesting will be assumed the same as the order of the order of the elements in the list; (v) an reStruct object. See the documentation on pdClasses for a description of the available pdMat classes. Defaults to a formula consisting of the right hand side of fixed.

method: a character string. If ''REML'' the model is fit by maximizing the restricted log-likelihood.
If ''ML'' the log-likelihood is maximized.
Defaults to ''REML''.

Note: if random is not specified, then it is the same as given by the fixed statement. Two arguments, correlation and weights will be explained later.



The lmer function

Description:

Fit a linear mixed model or a generalized linear mixed model or a nonlinear mixed model.

Usage:

```
lmer(formula, data, family = NULL, REML = TRUE)
```

Arguments:

formula: a two-sided linear formula object describing the fixed-effects part of the model, with the response on the left of a `~` operator and the terms, separated by `+` operators, on the right. The vertical bar character `'|'` separates an expression for a model matrix and a grouping factor.

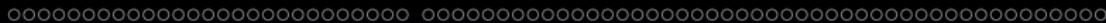


The lmer function

`data`: an optional data frame containing the variables named in `formula`. By default the variables are taken from the environment from which `lmer` is called.

`family`: a GLM family, see `glm` and `family`. If `family` is missing then a linear mixed model is fit; otherwise a generalized linear mixed model is fit.

`REML`: logical argument to `lmer` only. Should the estimates be chosen to optimize the REML criterion (as opposed to the log-likelihood)? Defaults to `TRUE`.



nlme VS lme4

Details about `nlme` can be found in the book by Pinheiro and Bates. Details about `lme4` can be found in the book by Bates, "lme4: Mixed-effects modeling with R", and the paper "Fitting Linear Mixed-Effects Models Using lme4" by Bates, Mächler, Bolker and Walker. According to the paper,

Advantage of `lme4` over `nlme`

- more efficient linear algebra tools, giving improved performance on large problems
- simpler syntax and more efficient implementation for fitting models with crossed random effects
- the implementation of profile likelihood confidence intervals on random-effects parameters
- the ability to fit generalized linear mixed models

Advantage of `nlme` over `lme4`

- a user interface for fitting models with structure in the residuals (various forms of heteroscedasticity and autocorrelation) and in the random-effects covariance matrices (e.g., compound symmetric models)

More information about `lme4` can be found at

<https://lme4.r-forge.r-project.org/>

<https://github.com/lme4/lme4/>

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

SAS proc mixed

```
PROC MIXED options;
CLASS variable-list;
MODEL dependent=fixed effects/ options;
RANDOM random effects / options;
REPEATED repeated effects / options;
CONTRAST 'label' fixed-effect values |
            random-effect values/ options;
ESTIMATE 'label' fixed-effect values |
            random-effect values/ options;
LSMEANS fixed-effects / options;
MAKE 'table' OUT= SAS-data-set < options >;
RUN;
```

Components in lme, lmer and proc mixed

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

- \mathbf{y} : a n dimensional vector of all responses
- \mathbf{X} : $n \times p$ design matrix for the fixed effects
- $\boldsymbol{\beta}$: p *fixed effects*
- \mathbf{Z} : $n \times q$ design matrix for the random effects
- \mathbf{b} : q *random effects*. $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{G})$
- $\boldsymbol{\epsilon}$: random errors. $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R})$
- \mathbf{b} and $\boldsymbol{\epsilon}$ are independent
- lme: fixed **specifies \mathbf{y} and \mathbf{X}** ; random **specifies Z and G** ; and correlation and weights **specifies R**
- lmer: formula **specifies \mathbf{y} , \mathbf{X} , \mathbf{Z} and \mathbf{G}** ; weights **specifies weights as in lm**;
- proc mixed: model **specifies \mathbf{y} and \mathbf{X}** ; random **specifies Z and G** ; and repeated **specifies R**



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: personnel data continued

Input and plot data

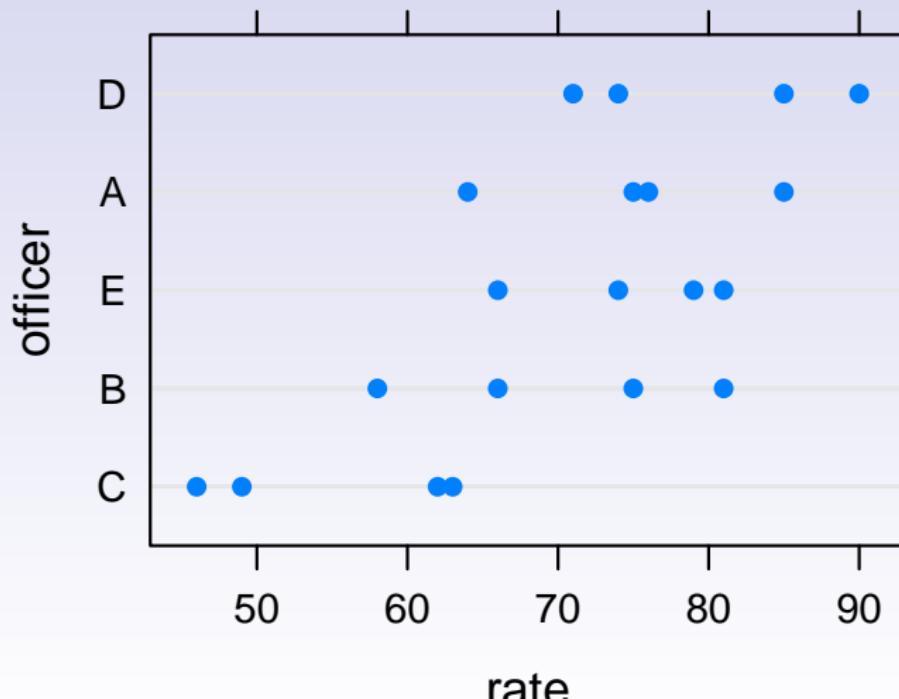
```
> library(nlme)

# create grouped data
> personnel <- groupedData( rate ~ 1 | officer,
  data = data.frame(rate=c(76,64,85,75,58,75,
    81,66,49,63,62,46,74,71,85,90,66,74,81,79),
    officer=rep(c("A","B","C","D","E"),rep(4,5))),
  labels = list(x="rate",y="officer"))

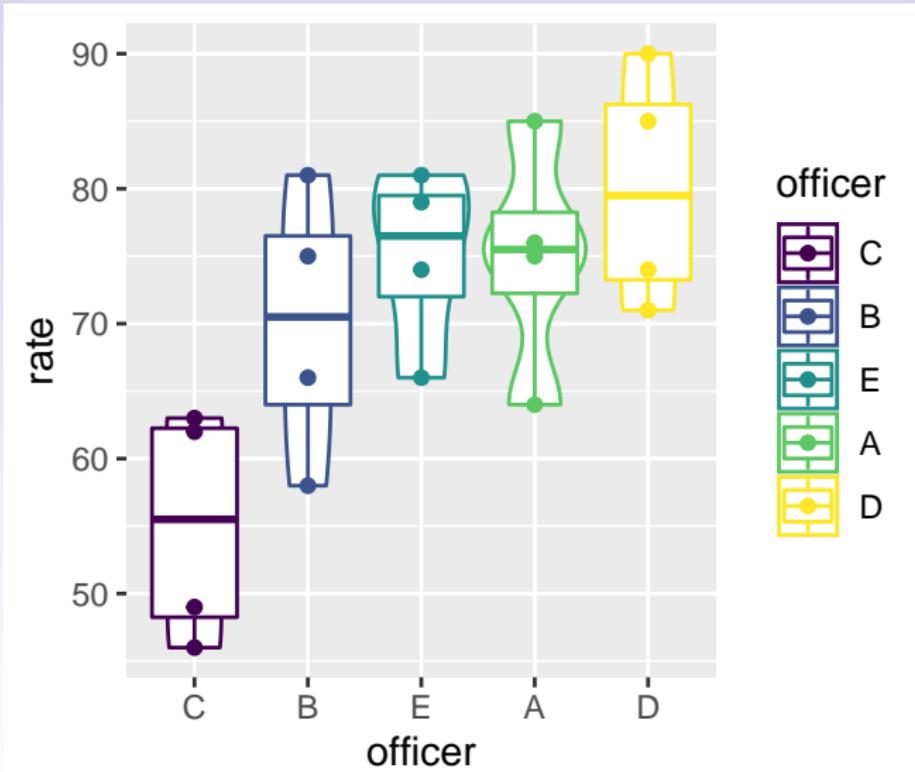
# display data
> plot(personnel,xlab='rate')

> ggplot(personnel, aes(officer, rate, col = officer)) +
  geom_violin() +
  geom_boxplot() +
  geom_point()
```

Plot of the personnel data



Plot of the personnel data



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Summary statistics

```
# grouping factor is ordered according to their maximum value
# plots usually are more informative
> personnel$officer
[1] A A A A B B B C C C C D D D D E E E E
Levels: C < B < E < A < D

> isBalanced(personnel)
[1] TRUE

> summary(personnel)
      rate      officer
Min.   :46.00   C:4
1st Qu.:63.75   B:4
Median  :74.00   E:4
Mean    :71.00   A:4
3rd Qu.:79.50   D:4
Max.   :90.00

> gsummary(personnel)
      rate officer
C     55      C
B     70      B
E     75      E
A     75      A
D     80      D
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme

```
> pers.lme <- lme( rate ~ 1, random = ~1 | officer,  
                     data=personnel)
```

```
> summary(pers.lme)
```

Linear mixed-effects model fit by REML

Data: personnel

AIC	BIC	logLik
151.4513	154.2846	-72.72563

Random effects:

Formula: ~1 | officer

(Intercept) Residual

StdDev: 8.579045 8.694826

Fixed effects: rate ~ 1

	Value	Std.Error	DF	t-value	p-value
(Intercept)	71	4.301163	15	16.50716	<.0001

Standardized Within-Group Residuals:

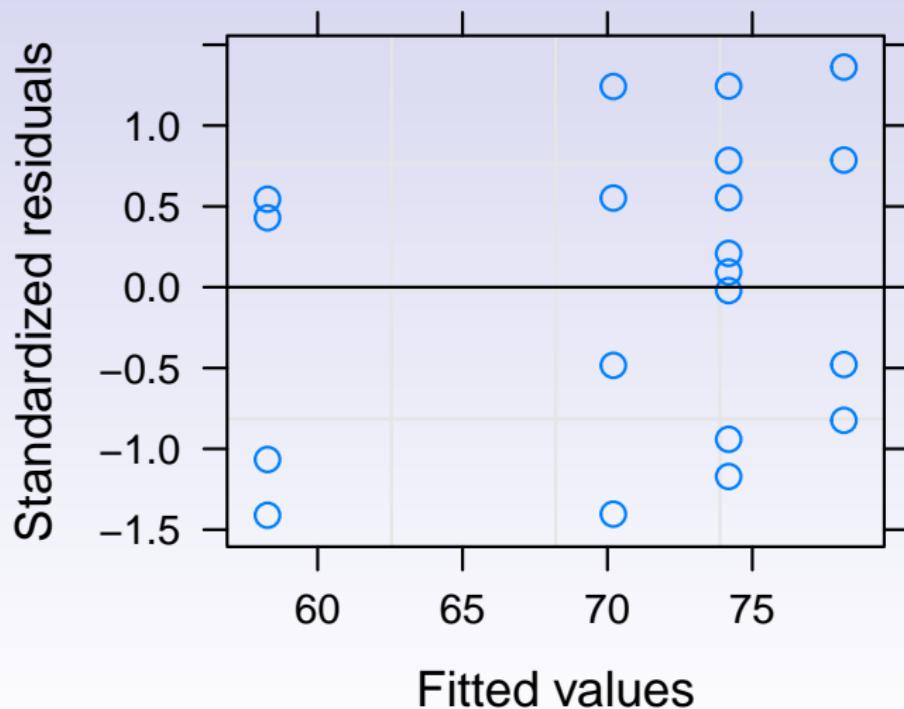
Min	Q1	Med	Q3	Max
-1.4110908	-0.8529770	0.1515036	0.6115473	1.3616050

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

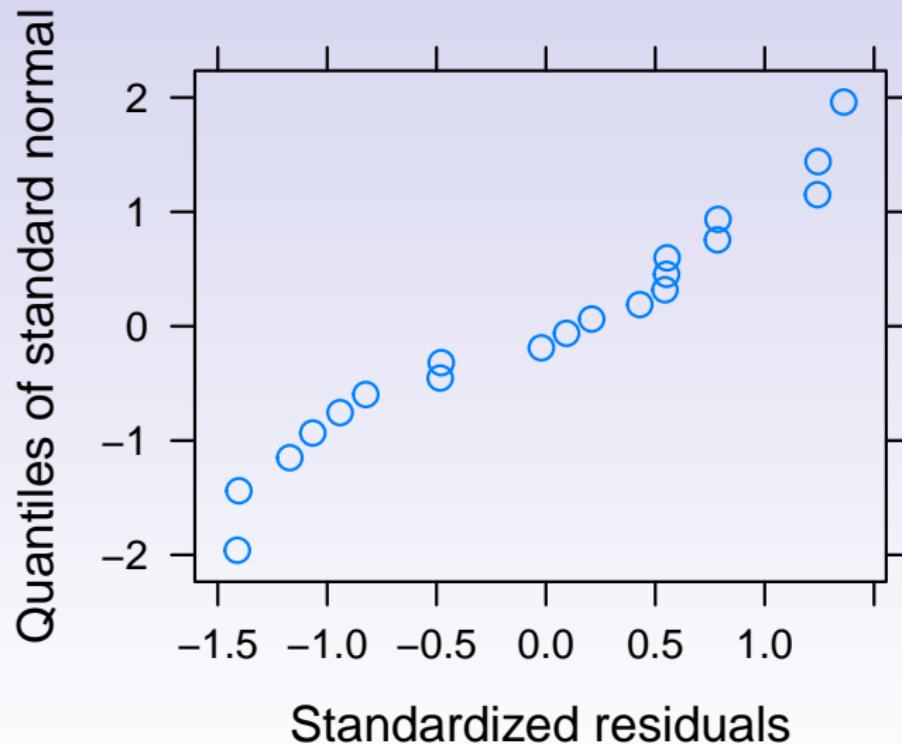
Diagnostic plots

```
# residuals vs fitted  
> plot(pers.lme)  
  
# check normality of random errors  
> qqnorm(pers.lme)  
  
# check normality of random effects  
> qqnorm(pers.lme, ~ranef(.), strip=F)
```

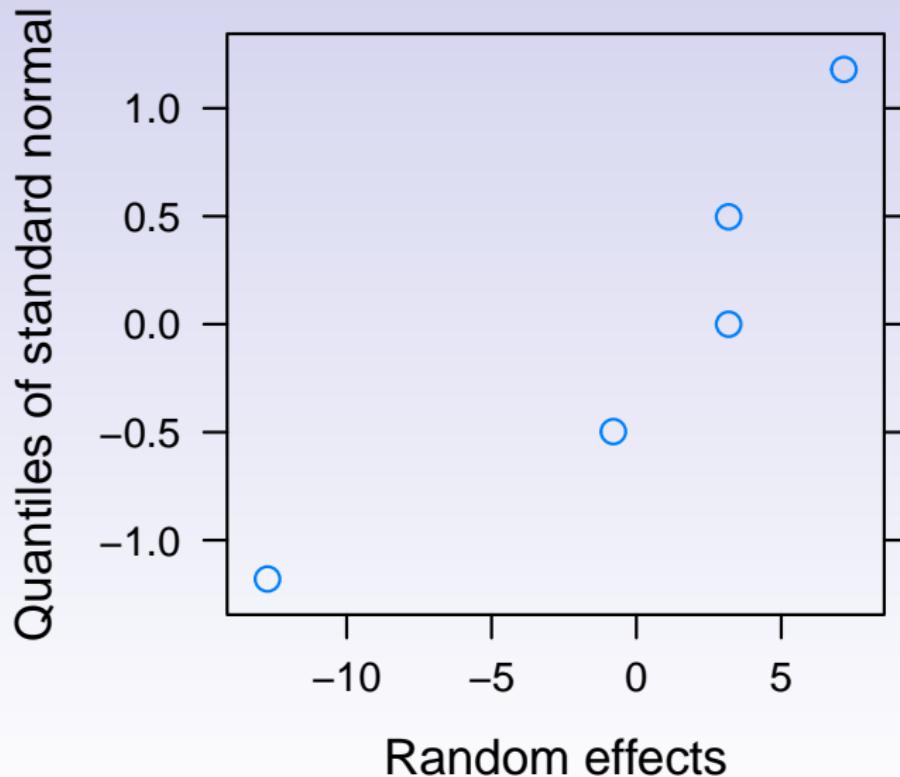
Diagnostic plots



Diagnostic plots



Diagnostic plots



oooooooooooooooooooooooooooo

Estimates of parameters

```
> VarCorr(pers.lme)
officer = pdLogChol(1)
      Variance StdDev
(Intercept) 73.60008 8.579049
Residual     75.59998 8.694825
```

```
> intervals(pers.lme)
Approximate 95% confidence intervals
```

Fixed effects:

	lower	est.	upper
(Intercept)	61.83228	71	80.16772
attr(,"label")			
[1] ``Fixed effects:''			

Random Effects:

Level: officer

	lower	est.	upper
sd((Intercept))	3.57368	8.579049	20.59504

Within-group standard error:

	lower	est.	upper
	6.079302	8.694825	12.435636

oooooooooooooooooooooooooooo

ML instead of REML

```
> lme( rate ~ 1, random = ~1 | officer, data=personnel,  
      method = 'ML')
```

Linear mixed-effects model fit by maximum likelihood

Data: personnel

Log-likelihood: -75.04559

Fixed: rate ~ 1

(Intercept)

71

Random effects:

Formula: ~1 | officer

(Intercept) Residual

StdDev: 7.42294 8.694825

Number of Observations: 20

Number of Groups: 5

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lmer

```
> pers.lmer <- lmer(rate ~ 1+(1|officer), data=personnel)
> summary(pers.lmer)
Linear mixed model fit by REML
Formula: rate ~ 1 + (1 | officer)
Data: personnel
AIC   BIC logLik deviance REMLdev
151.5 154.4 -72.73    150.2   145.5
Random effects:
 Groups     Name        Variance Std.Dev.
 officer  (Intercept) 73.6      8.5790
 Residual           75.6      8.6948
Number of obs: 20, groups: officer, 5

Fixed effects:
            Estimate Std. Error t value
(Intercept) 71.000     4.301   16.51
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Diagnostic plots

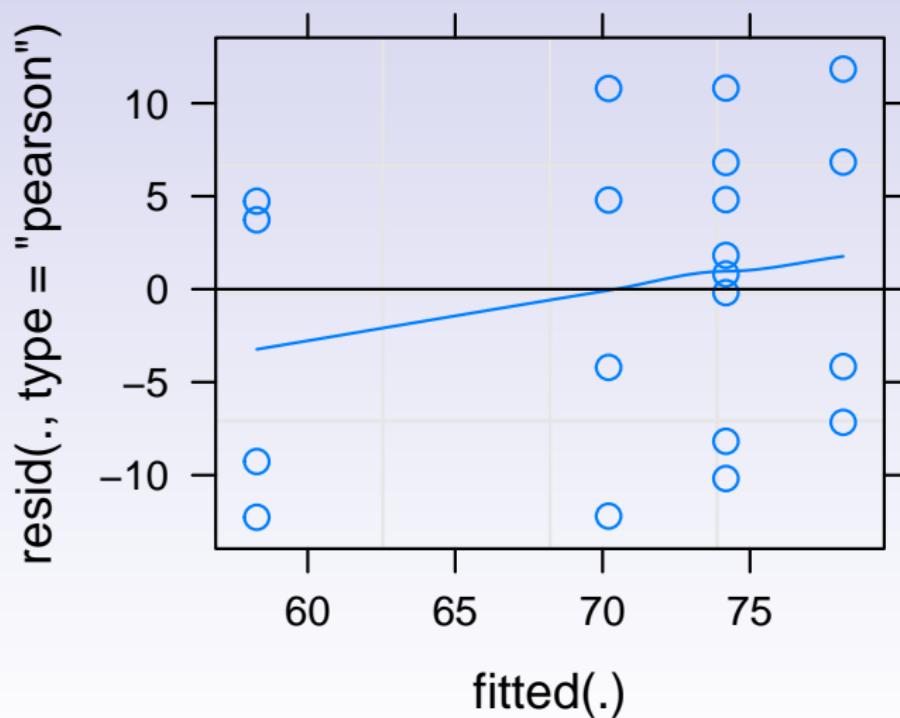
```
# residual vs fitted
plot(pers.lmer, type = c("p", "smooth"))

# scale-locaion plot
plot(pers.lmer, sqrt(abs(resid(.))) ~ fitted(.),
      type = c("p", "smooth"))

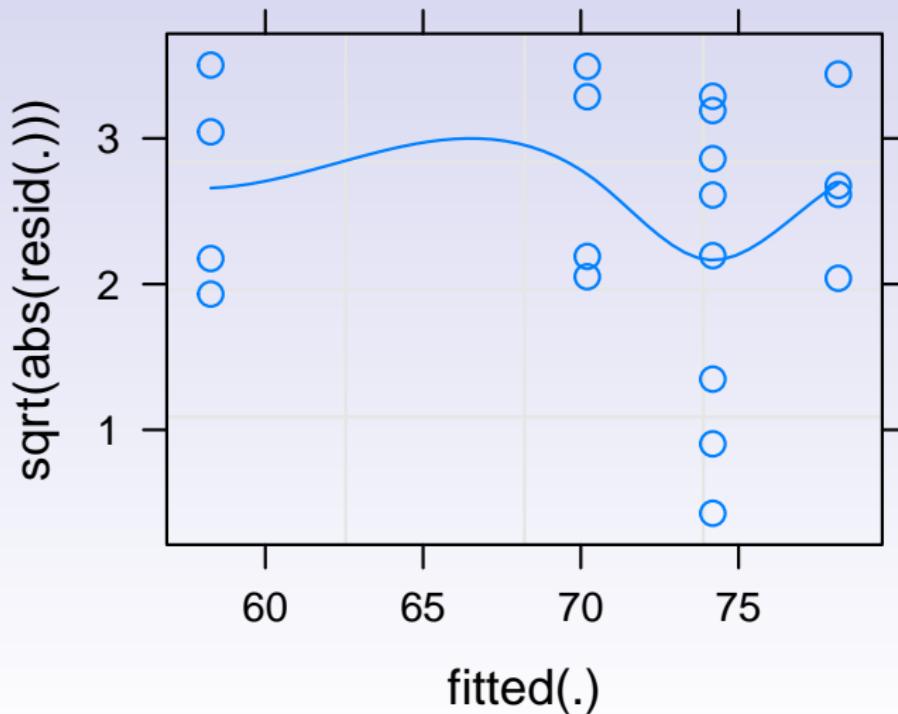
# QQ plot of residuals
qqmath(pers.lmer, id = 0.05)

# Q plot of random effects
qqnorm(ranef(pers.lmer)$officer[[1]])
qqline(ranef(pers.lmer)$officer[[1]])
```

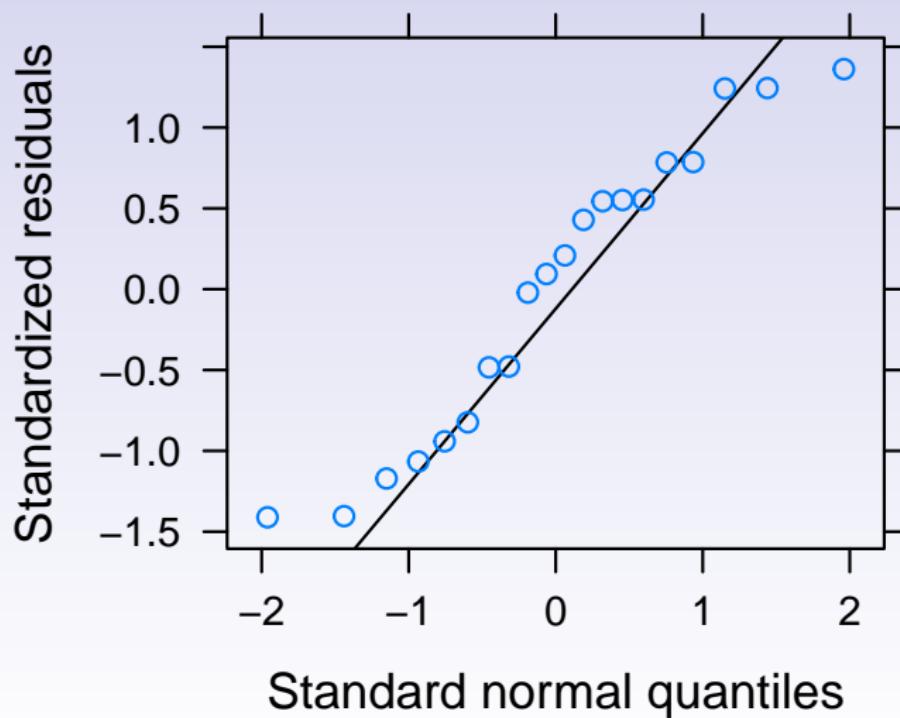
Diagnostic plots



Diagnostic plots

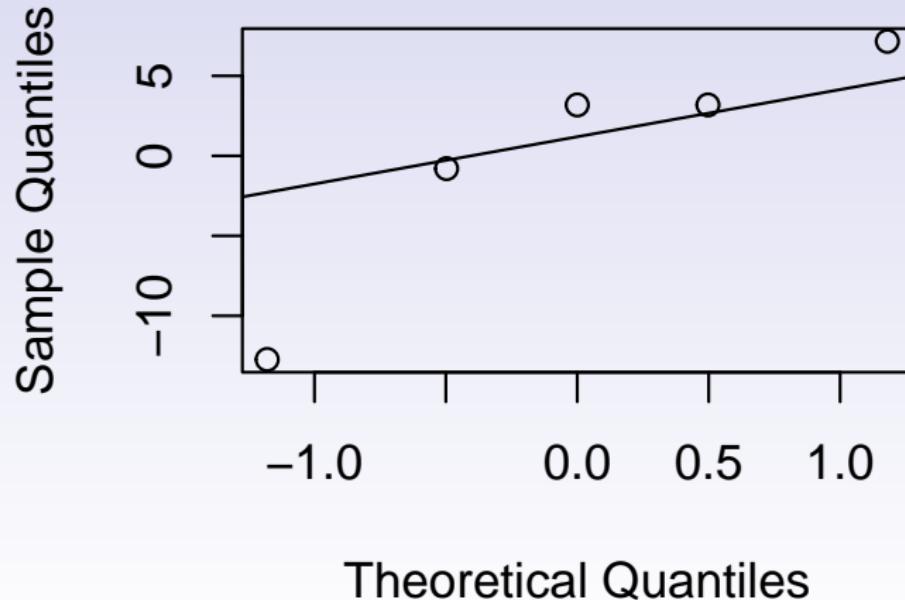


Diagnostic plots



Diagnostic plots

Normal Q-Q Plot



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lmer

```
# use ML for profile plots
> pers.lmer.ml <- lmer( rate ~ (1 | officer),
                           data=personnel, REML=F)
> confint(pers.lmer.ml, method="profile", oldNames = FALSE)

                                2.5 %    97.5 %
sd_(Intercept)|officer   1.816894 18.23627
sigma                      6.315070 13.05072
(Intercept)                 61.750709 80.24929

> confint(pers.lmer.ml, method="Wald", oldNames = FALSE)
                                2.5 %    97.5 %
sd_(Intercept)|officer      NA        NA
sigma                      NA        NA
(Intercept)                 63.45987 78.54013
> confint(pers.lmer.ml, method="boot", oldNames = FALSE)

                                2.5 %    97.5 %
sd_(Intercept)|officer  0.00000 12.71823
sigma                      5.29343 11.45120
(Intercept)                 63.76427 78.90970
```



Profile plots

Details of these methods can be found in "Fitting Linear Mixed-Effects Models Using lme4" by Bates et al. Profile plots can be used to check the adequacy of these methods. The `profile` function systematically varies the parameters in a model and compare to the globally optimal.

Write the deviance (negative twice the log-likelihood) as $d(\theta_i, \theta_{(-i)} | \mathbf{y})$ where θ_i is the parameter we want to profile on and $\theta_{(-i)}$ is the vector of the rest parameters. The profile deviance

$$\tilde{d}_i(\theta_i) = \min_{\theta_{(-i)}} d(\theta_i, \theta_{(-i)} | \mathbf{y})$$

are computed at different values of θ_i . The signed square root scale (zeta)

$$\zeta_i(\theta_i) = \text{sgn}(\theta_i - \hat{\theta}_i) \sqrt{\tilde{d}_i(\theta_i) - d(\hat{\theta})}.$$

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

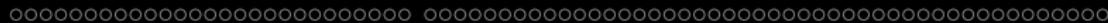
Profile plots

```
> (prof=profile(pers.lmer.ml))
      .zeta     .sig01     .sigma (Intercept)     .par
 1 -2.03087442  0.0000000 11.432410    71.00000  .sig01
 2 -2.02665177  0.2976079 11.418749    71.00000  .sig01
 3 -1.84554394  1.9509271 10.869263    71.00000  .sig01
 4 -1.54033270  3.1494533 10.112204    71.00000  .sig01
 5 -1.15700299  4.2406682  9.454268    71.00000  .sig01
 6 -0.71085693  5.3969871  9.006920    71.00000  .sig01
 7 -0.24346490  6.6806448  8.765075    71.00000  .sig01
 8  0.00000000  7.4229387  8.694826    71.00000  .sig01
 9  0.02319014  7.4971681  8.689568    71.00000  .sig01
10  0.24381913  8.2394619  8.649660    71.00000  .sig01
11  0.61251031  9.6559829  8.614647    71.00000  .sig01
12  0.96360908 11.2735786  8.605603    71.00000  .sig01
13  1.30963076 13.2133493  8.610291    71.00000  .sig01
14  1.65036450 15.5735927  8.621848    71.00000  .sig01
15  1.98624323 18.4900214  8.635871    71.00000  .sig01
16  2.31750359 22.1457888  8.649681    71.00000  .sig01
17  2.64433677 26.7922126  8.661821    71.00000  .sig01
18  2.96691418 32.7777395  8.671685    71.00000  .sig01
19  3.28540119 40.5900257  8.679219    71.00000  .sig01
20  3.59995952 50.9175478  8.684677    71.00000  .sig01
21 -3.78280254 8.2377424  4.955645    71.00000  .sigma
22 -3.31487323 8.1917179  5.251957    71.00000  .sigma
23 -2.84465294 8.1368200  5.582890    71.00000  .sigma
24 -2.37193670 8.0706957  5.954452    71.00000  .sigma
25 -1.89649509 7.9901763  6.374037    71.00000  .sigma
26 -1.41808202 7.8909189  6.850811    71.00000  .sigma
27 -0.93658961 7.7668908  7.396056    71.00000  .sigma
28 -0.45409918 7.6102754  8.020900    71.00000  .sigma
29  0.00000000  7.4229387  8.694826    71.00000  .sigma
30  0.05431983 7.3973045  8.781774    71.00000  .sigma
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Profile plots

31	0.44698500	7.1866150	9.455700	71.00000	.sigma
32	0.81980319	6.9354554	10.178303	71.00000	.sigma
33	1.19211175	6.6167290	10.994344	71.00000	.sigma
34	1.56282122	6.2044563	11.917168	71.00000	.sigma
35	1.93194512	5.6546956	12.965249	71.00000	.sigma
36	2.29943839	4.8855534	14.160702	71.00000	.sigma
37	2.66526278	3.7016829	15.530299	71.00000	.sigma
38	3.02938378	0.9172319	17.106566	71.00000	.sigma
39	3.40438232	0.0000000	18.929176	71.00000	.sigma
40	-3.58537807	30.8046419	8.694826	41.10308	(Intercept)
41	-3.26718622	24.6345203	8.694826	47.51044	(Intercept)
42	-2.94334235	19.9904272	8.694826	52.43883	(Intercept)
43	-2.61303540	16.4631331	8.694826	56.30528	(Intercept)
44	-2.27527514	13.7657574	8.694826	59.40707	(Intercept)
45	-1.92885703	11.6976973	8.694826	61.95920	(Intercept)
46	-1.57230892	10.1206639	8.694826	64.12048	(Intercept)
47	-1.20380702	8.9430896	8.694826	66.01213	(Intercept)
48	-0.82103895	8.1105217	8.694826	67.73189	(Intercept)
49	-0.42096291	7.6006406	8.694826	69.36607	(Intercept)
50	0.00000000	7.4229387	8.694826	71.00000	(Intercept)
51	0.41736518	7.5975984	8.694826	72.61972	(Intercept)
52	0.81763953	8.1047068	8.694826	74.25366	(Intercept)
53	1.20054277	8.9344146	8.694826	75.97230	(Intercept)
54	1.56915800	10.1088015	8.694826	77.86206	(Intercept)
55	1.92580164	11.6820254	8.694826	80.02052	(Intercept)
56	2.27230108	13.7453051	8.694824	82.56863	(Intercept)
57	2.61013111	16.4364365	8.694826	85.66480	(Intercept)
58	2.94049835	19.9554216	8.694826	89.52347	(Intercept)
59	3.26439469	24.5882460	8.694826	94.44103	(Intercept)
60	3.58263267	30.7428734	8.694826	100.83327	(Intercept)



Profile plots

- $.zeta$: as defined
- $.sig01$: standard deviation of random effect σ_1
- $.sigma$: standard deviation of random error σ
- $(Intercept)$: fixed effect μ
- $.par$: the parameter being varied

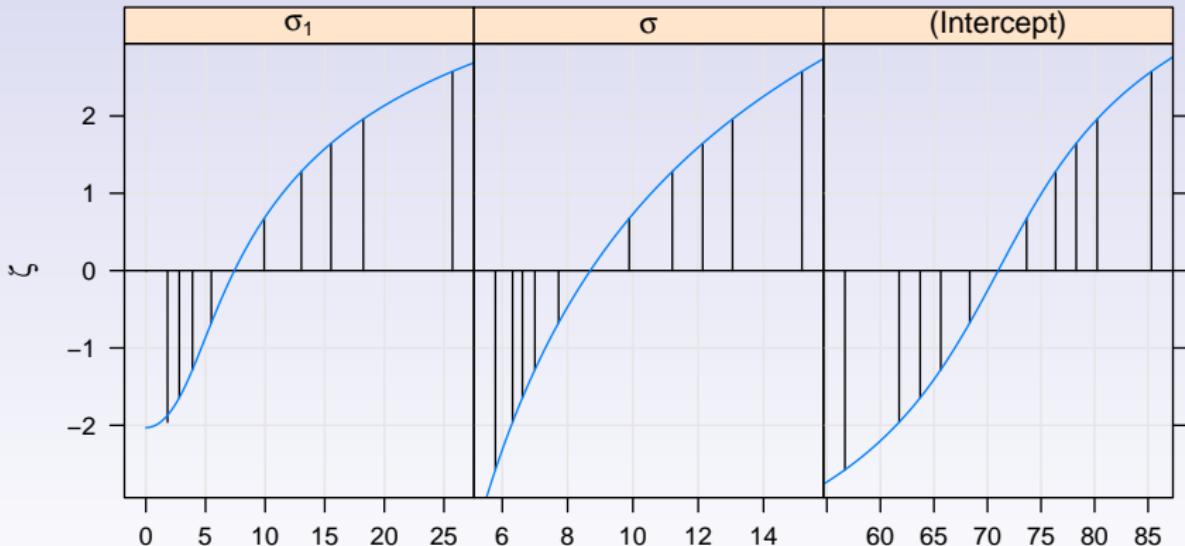


Profile zeta plots

```
xyplot(prof, aspect=1.3, layout = c(3,1))
```

- Plot signed square root of the likelihood ratio test statistic for each of the parameters in model
- The vertical lines in the panels correspond to the 50%, 80%, 90%, 95% and 99% confidence intervals, if can be calculated
- Read like a normal QQ plot:
 - an S-shape indicates heavy or light tail (the confidence intervals will be conservative or liberal)
 - a U-shape indicates skewness of the distribution (the confidence intervals will be asymmetric)
 - a straight line indicates that the confidence intervals based on the quantiles of the standard normal distribution are reasonably accurate

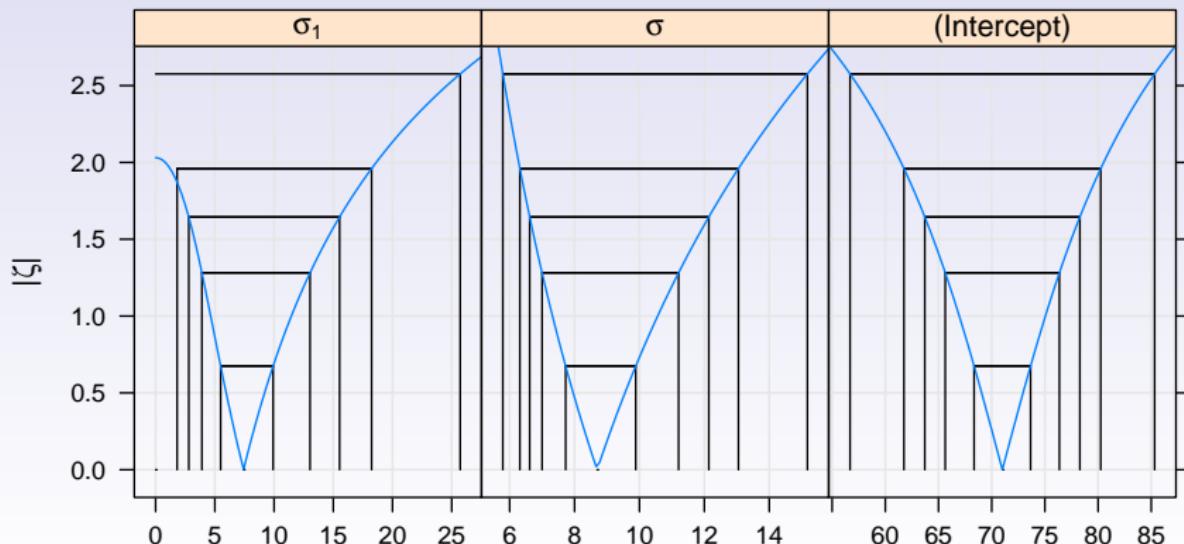
Profile zeta plot



oooooooooooooooooooooooooooo

Profile zeta plot on the absolute scale

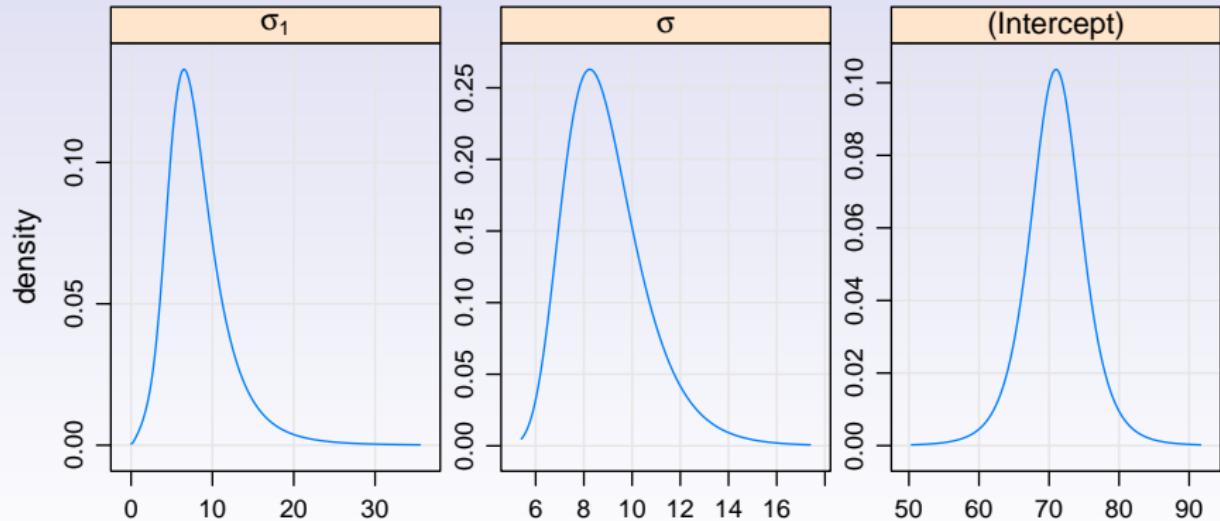
```
xyplot(prof, aspect=1.3, layout = c(3,1), absVal=T)
```



ooooooooooooooooooooooooooooo oooooooooooooooooooooooooooooo oooooooooooooooooooooo

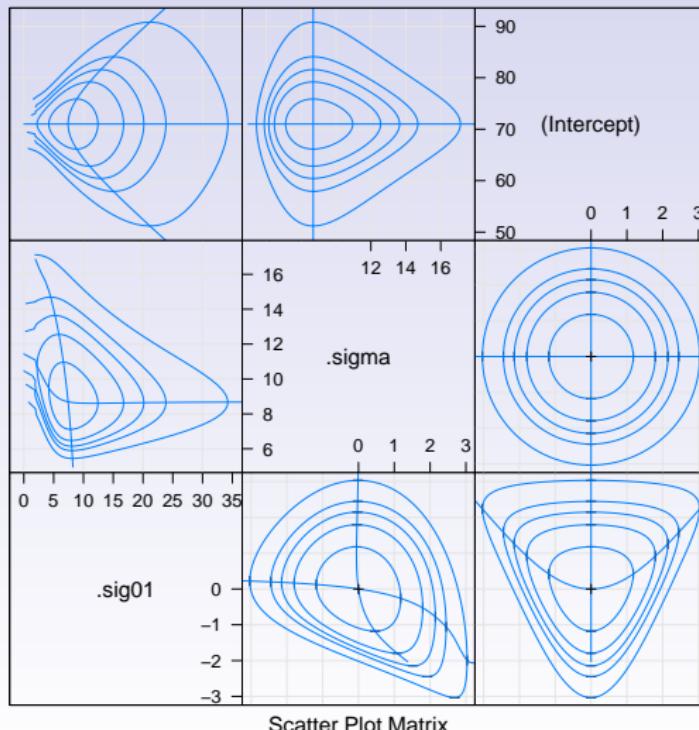
Profile density plot

```
# map the zeta scale through the cdf of the standard normal  
densityplot(prof, aspect=1.3, layout = c(3,1))
```



Profile pairs plot

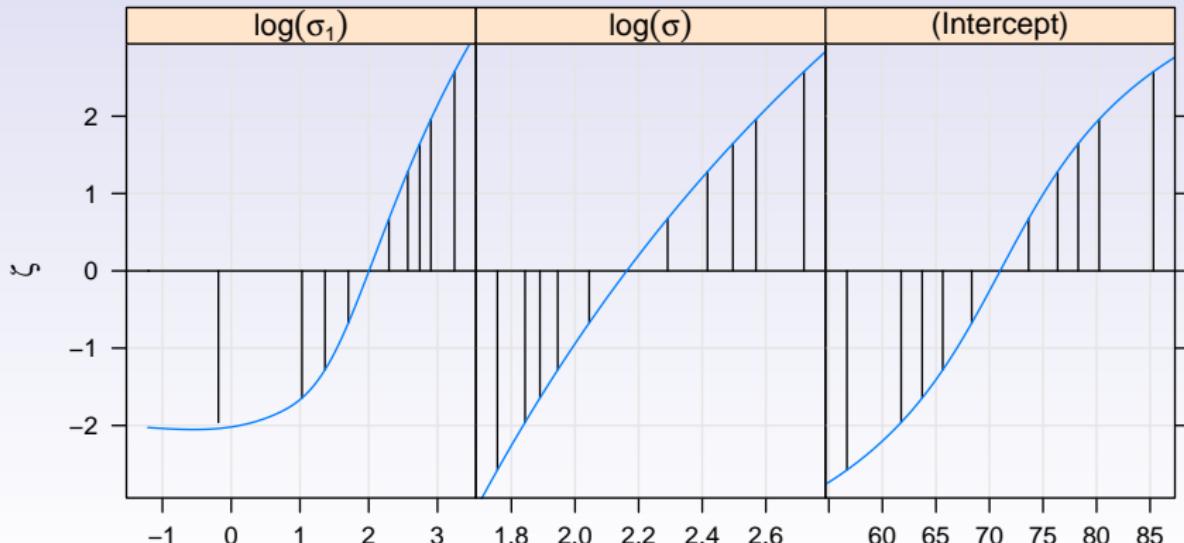
```
splom(prof)
```



ooooooooooooooooooooooooooooo oooooooooooooooooooooooooooooo oooooooooooooooooooooo

Profile zeta plot on log scale

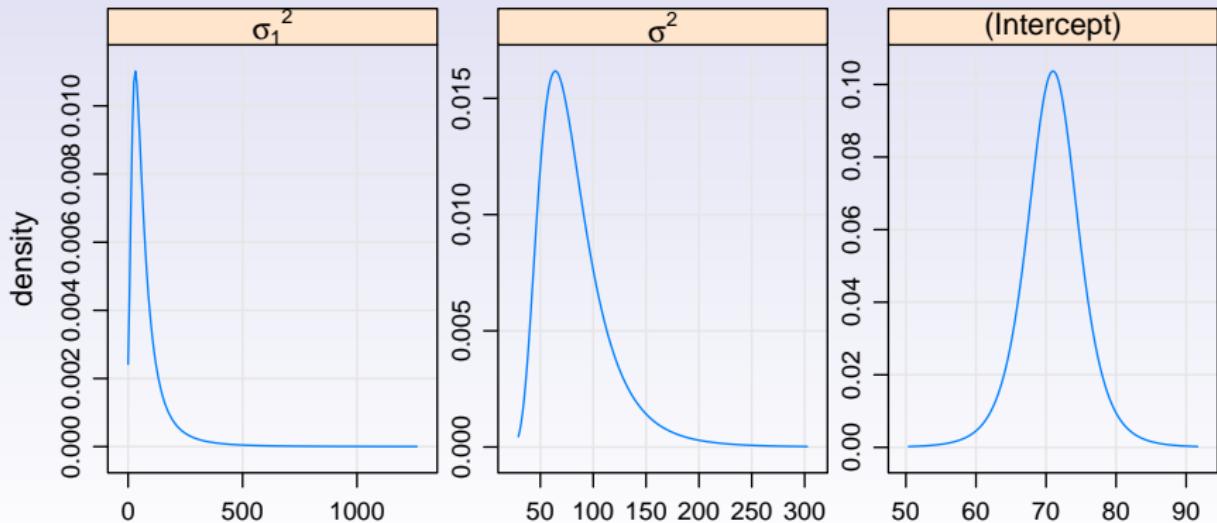
```
xyplot(logProf(prof), aspect=1.3, layout = c(3,1))
```



ooooooooooooooooooooooooooooo oooooooooooooooooooooooooooooo oooooooooooooooooooooo oooooooooooooo

Profile density plot for variances

```
densityplot(varianceProf(prof), aspect=1.3, layout = c(3,1))
```

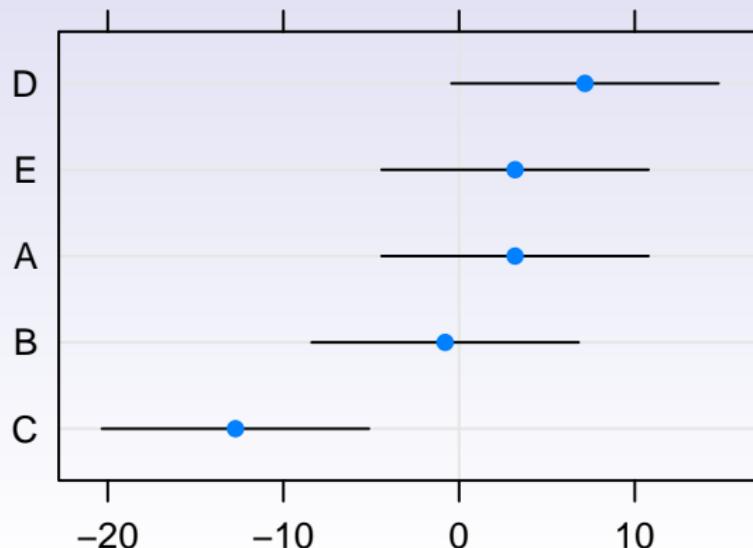


ooooooooooooooooooooooooooooo oooooooooooooooooooooooooooooo oooooooooooooooooooooo

Caterpillar plot

```
# plot of estimates of random effects with CIs  
> library(lattice)  
> dotplot(ranef(pers.lmer), strip=F)
```

officer



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using SAS proc mixed

```
options nocenter ps=64 ls=76;
data a;
  input officer $ @;
  do cand=1 to 4;
    input rate @;
    output;
  end;
  cards;
A 76 64 85 75
B 58 75 81 66
C 49 63 62 46
D 74 71 85 90
E 66 74 81 79
;

proc mixed method=reml CL;
  class officer;
  model rate = / CL ALPHA=.05;
  random officer;
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Output from proc mixed

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
officer	73.6000	0.05	21.9413	1544.15
Residual	75.6000	0.05	41.2538	181.09

Fit Statistics

-2 Res Log Likelihood	145.5
AIC (smaller is better)	149.5
AICC (smaller is better)	150.2
BIC (smaller is better)	148.7

The Mixed Procedure

Solution for Fixed Effects

Standard

Effect	Estimate	Error	DF	t Value	Pr > t
Alpha					
Intercept	71.0000	4.3012	4	16.51	<.0001
0.05					

Solution for Fixed Effects

Effect	Lower	Upper
Intercept	59.0581	82.9419



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- **Spectrophotometer data**
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors

oooooooooooooooooooooooooooo

Example: spectrophotometer data continued

4 machines were randomly selected from the pilot assembly production. 8 replicate serum samples were prepared each day from the same stock reagents. 2 serum samples were randomly assigned to each of the 4 machines on each of the 4 randomly selected days. Therefore, we have a completely randomized design with 2 replications of each treatment and day combination. The same technician prepared the serum samples and operated the machines throughout the experiment.

Example: spectrophotometer data continued

The observations on triglyceride levels (mg/dl) in the serum samples are shown in the following table.

Day	Machine			
	1	2	3	4
1	142.3, 144.0	148.6, 146.9	142.9, 147.4	133.8, 133.2
2	134.9, 146.3	145.2, 146.3	125.9, 127.6	108.9, 107.5
3	148.6, 156.5	148.6, 153.1	135.5, 138.9	132.1, 149.7
4	152.0, 151.4	149.7, 152.0	142.9, 142.3	141.7, 141.2



Example: spectrophotometer data continued

We considered the following two-way random effects model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

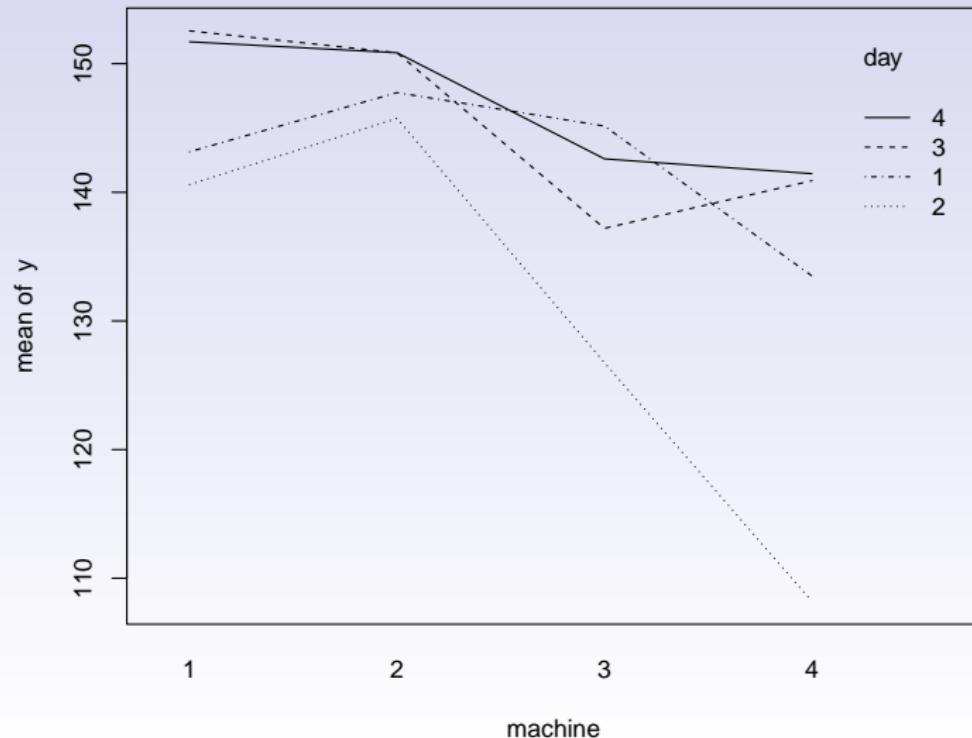
- μ : fixed effects
- α_i, β_j and $(\alpha\beta)_{ij}$: random effects, $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_a^2)$,
 $\beta_j \stackrel{iid}{\sim} N(0, \sigma_b^2)$, $(\alpha\beta)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{ab}^2)$
- ϵ_{ijk} : random errors, $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$
- $\alpha_i, \beta_j, (\alpha\beta)_{ij}$ and ϵ_{ijk} are mutually independent

oooooooooooooooooooooooooooo

Example: spectrophotometer data continued

```
> y <- c(142.3,144.0,134.9,146.3,148.6,156.5,152.0,151.4,  
       148.6,146.9,145.2,146.3,148.6,153.1,149.7,152.0,  
       142.9,147.4,125.9,127.6,135.5,138.9,142.9,142.3,  
       133.8,133.2,108.9,107.5,132.1,149.7,141.7,141.2)  
> spectrophotometer <- data.frame(y=y,  
           machine = gl(4,8,32), day = gl(4,2,32),  
           repl = gl(16,2), grp <- rep(1,32))  
> attach(spectrophotometer)  
> interaction.plot(machine, day, y)
```

Interaction plot of spectrophotometer data



oooooooooooooooooooooooooooo ooooooooooooooooooooo ooooooooooooo ooooooooooooo

Fit using lme

```
> library(nlme)
> fit1 <- lme(y~1, random=list(
+     grp=pdBlocked(list(pdIdent(~machine-1),
+         pdIdent(~day-1), pdIdent(~repl-1)))),
+     data=spectrophotometer,
+     control=list(msTol = 1e-8))
```

```
> summary(fit1)
```

Linear mixed-effects model fit by REML

Data: spectrophotometer

AIC BIC logLik

225.0384 232.2083 -107.5192

Random effects:

Composite Structure: Blocked

Block 1: machine1, machine2, machine3, machine4

Formula: ~machine - 1 | grp

Structure: Multiple of an Identity

 machine1 machine2 machine3 machine4

StdDev: 7.597367 7.597367 7.597367 7.597367

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme

Block 2: day1, day2, day3, day4

Formula: ~day - 1 | grp

Structure: Multiple of an Identity

day1 day2 day3 day4

StdDev: 6.684682 6.684682 6.684682 6.684682

Block 3: repl1, repl2, repl3, repl4, repl5, repl6, repl7,
repl8, repl9, repl10, repl11, repl12, repl13, repl14,
repl15, repl16

Formula: ~repl - 1 | grp

Structure: Multiple of an Identity

repl1 repl2 repl3 repl4 repl5

repl6 repl7 repl8 repl9 repl10 repl11

StdDev: 5.892446 5.892446 5.892446 5.892446 5.892446

5.892446 5.892446 5.892446 5.892446 5.892446 5.892446

 repl12 repl13 repl14 repl15 repl16 Residual

StdDev: 5.892446 5.892446 5.892446 5.892446 5.892446 4.230287

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using lme

Fixed effects: y ~ 1

	Value	Std.Error	DF	t-value	p-value
(Intercept)	141.1844	5.322643	31	26.52524	0

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-1.84281300	-0.35581370	0.03484132	0.20699414	2.31766114

Number of Observations: 32

Number of Groups: 1

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme

```
> intervals(fit1)
```

```
Approximate 95% confidence intervals
```

Fixed effects:

	lower	est.	upper
(Intercept)	130.3288	141.1844	152.04
attr(,"label")			
[1] ``Fixed effects:''			

Random Effects:

Level: grp

	lower	est.	upper
sd(machine - 1)	2.922098	7.597367	19.75293
sd(day - 1)	2.454304	6.684682	18.20678
sd(repl - 1)	3.273430	5.892446	10.60689

Within-group standard error:

	lower	est.	upper
2.991564	4.230287	5.981931	

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lmer

```
> library(lme4)
> fit2 <- lmer(y~1+(1|machine)+(1|day)+(1|machine:day),
+                 data=spectrophotometer)
> summary(fit2)
Linear mixed model fit by REML
Formula: y ~ 1 + (1 | machine) + (1 | day) + (1 | machine:day)
Data: spectrophotometer
AIC    BIC logLik deviance REMLdev
225 232.4 -107.5    220.2      215
Random effects:
Groups      Name        Variance Std.Dev.
machine:day (Intercept) 34.721   5.8925
day          (Intercept) 44.685   6.6847
machine      (Intercept) 57.719   7.5973
Residual                17.895   4.2303
Number of obs: 32, groups: machine:day, 16; day, 4; machine, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept) 141.184     5.322   26.53
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

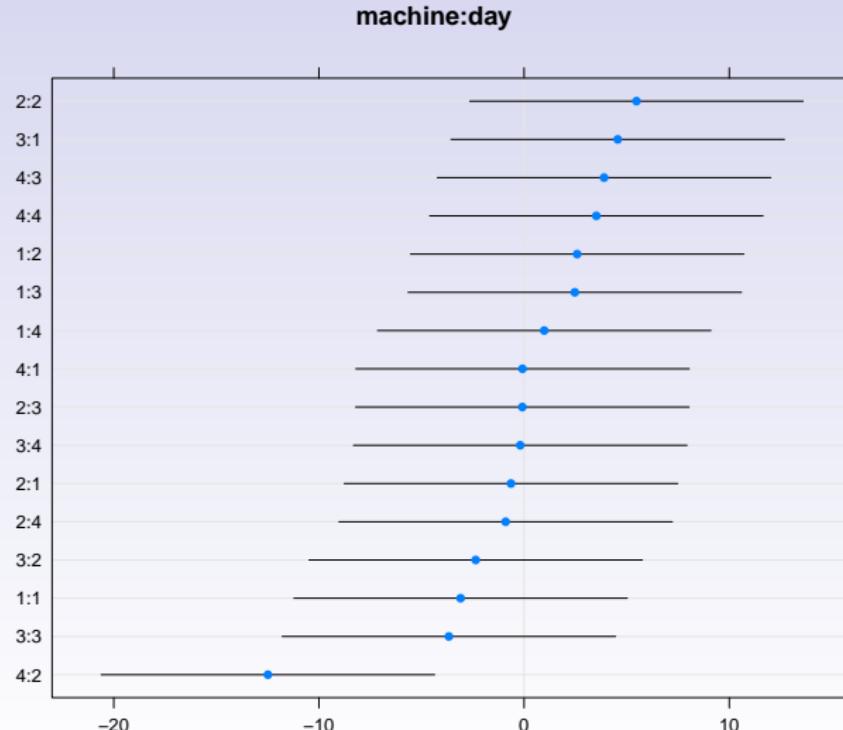
R analysis of the spectrophotometer data

```
> confint(fit2, method="profile", oldNames = FALSE)
```

	2.5 %	97.5 %
sd_(Intercept) machine:day	2.956017	11.29686
sd_(Intercept) day	0.000000	16.96878
sd_(Intercept) machine	0.000000	18.48160
sigma	3.100604	6.25775
(Intercept)	129.938470	152.43024

Caterpillar plot of interaction

```
dotplot(ranef(fit2, condVar=T), strip=F) [ [ 'machine:day' ] ]
```



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using SAS proc mixed

```
data a;
  do machine=1 to 4;
    do day=1 to 4;
      do rep=1 to 2;
        input y @;
        output;
      end;
    end;
  end;
  cards;
142.3 144.0
134.9 146.3
148.6 156.5
152.0 151.4
148.6 146.9
145.2 146.3
148.6 153.1
149.7 152.0
142.9 147.4
125.9 127.6
135.5 138.9
142.9 142.3
133.8 133.2
108.9 107.5
132.1 149.7
141.7 141.2
;
```



Fit using SAS proc mixed

```
proc mixed method=reml CL;  
  class day machine;  
  model y = / CL ALPHA=.05;  
  random day | machine;
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using SAS proc mixed

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
day	44.6855	0.05	11.8836	2037.90
machine	57.7194	0.05	15.9896	1951.29
day*machine	34.7210	0.05	14.0432	183.52
Residual	17.8953	0.05	9.9262	41.4503

Fit Statistics

-2 Res Log Likelihood	215.0
AIC (smaller is better)	223.0

The Mixed Procedure

Fit Statistics

AICC (smaller is better)	224.6
BIC (smaller is better)	220.6

Solution for Fixed Effects

Standard

Effect	Estimate	Error	DF	t Value	Pr > t
Alpha					
Intercept	141.18	5.3226	3	26.53	0.0001
0.05					

Solution for Fixed Effects

Effect	Lower	Upper
Intercept	124.25	158.12

Spectrophotometer data - unbalanced

```
data b;  
  set a;  
  if _N_=1 then delete;  
  
proc mixed method=reml CL;  
  class day machine;  
  model y = / CL ALPHA=.05;  
  random day | machine;
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Spectrophotometer data - unbalanced

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
day	45.4137	0.05	12.1580	1968.77
machine	59.6119	0.05	16.6145	1931.72
day*machine	33.1811	0.05	13.1065	190.17
Residual	18.8973	0.05	10.3377	45.0326

The Mixed Procedure

Fit Statistics

-2 Res Log Likelihood	209.4
AIC (smaller is better)	217.4
AICC (smaller is better)	219.0
BIC (smaller is better)	215.0

Solution for Fixed Effects

Effect	Estimate	Error	DF	t Value	Pr > t	Standard
Alpha						
Intercept	141.28	5.3808	3	26.26	0.0001	
0.05						

Solution for Fixed Effects

Effect	Lower	Upper
Intercept	124.15	158.40



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- **Machine data**
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: machine data continued

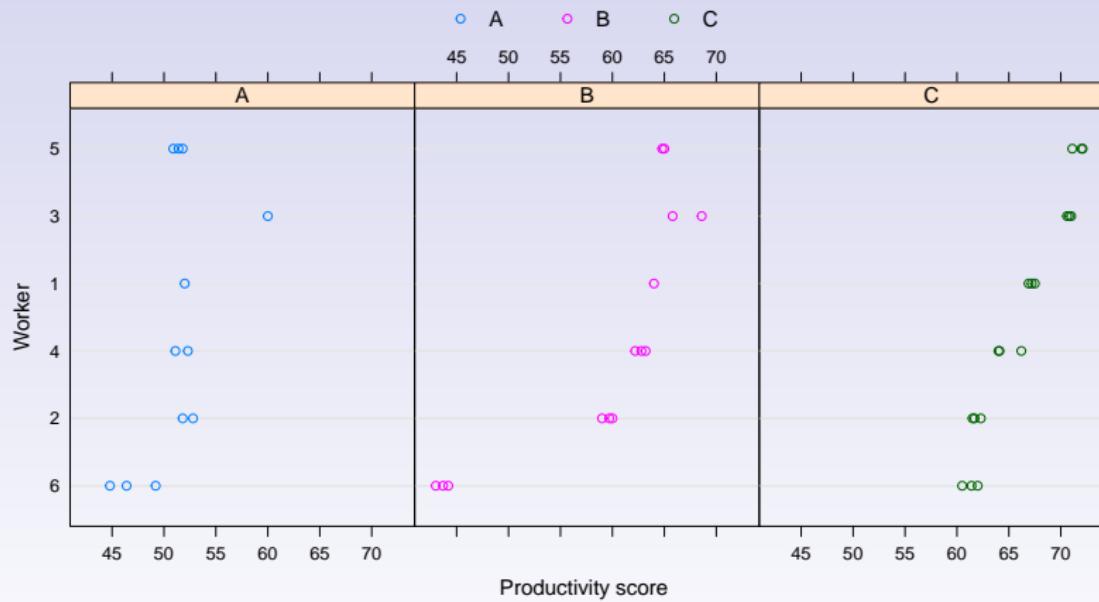
We show the fit to unbalanced data only. The same codes can be used to fit the balanced data.

```
> library(nlme)
> data(Machines)
> isBalanced(Machines)
[1] TRUE

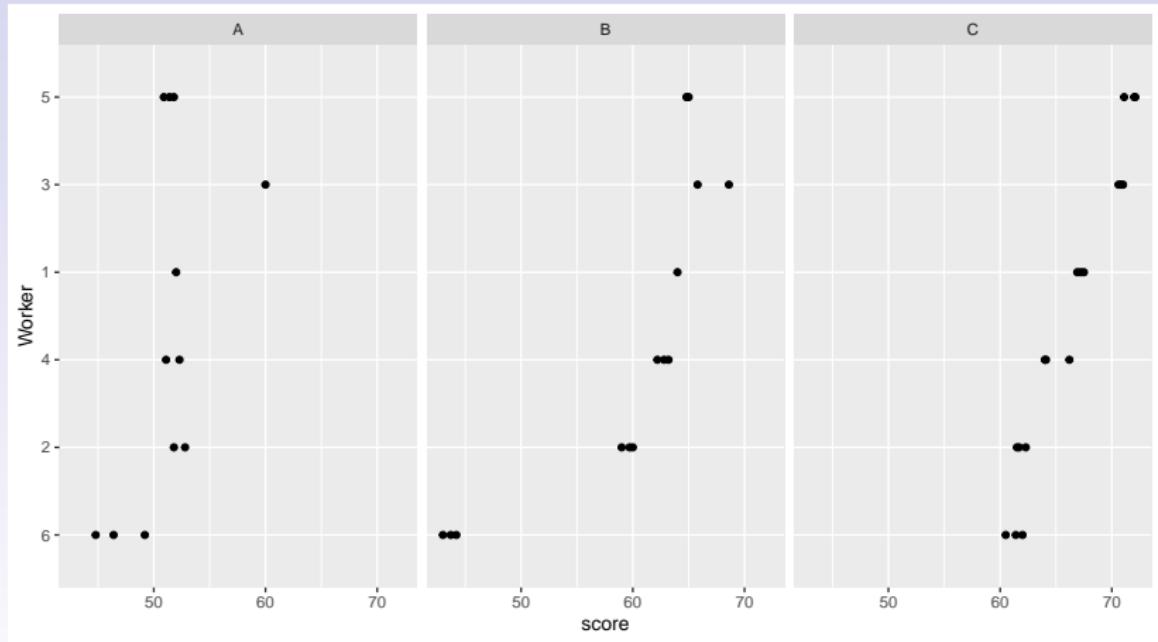
> MachinesUnbal <- Machines[-c(2,3,6,8,9,12,19,20,27,33),]
> isBalanced(MachinesUnbal)
[1] FALSE

> plot(MachinesUnbal,outer=~Machine,layout=c(3,1), cex=.5)
> qplot(score,Worker,data=MachinesUnbal,facets=~Machine)
> ggplot(MachinesUnbal,
  aes(x = Machine, y = score, group = Worker, col = Worker)) +
  stat_summary(fun = mean, geom = "line") +
  theme_bw()
> with(MachinesUnbal, interaction.plot(x.factor = Machine,
  trace.factor = Worker,
  response = score))
```

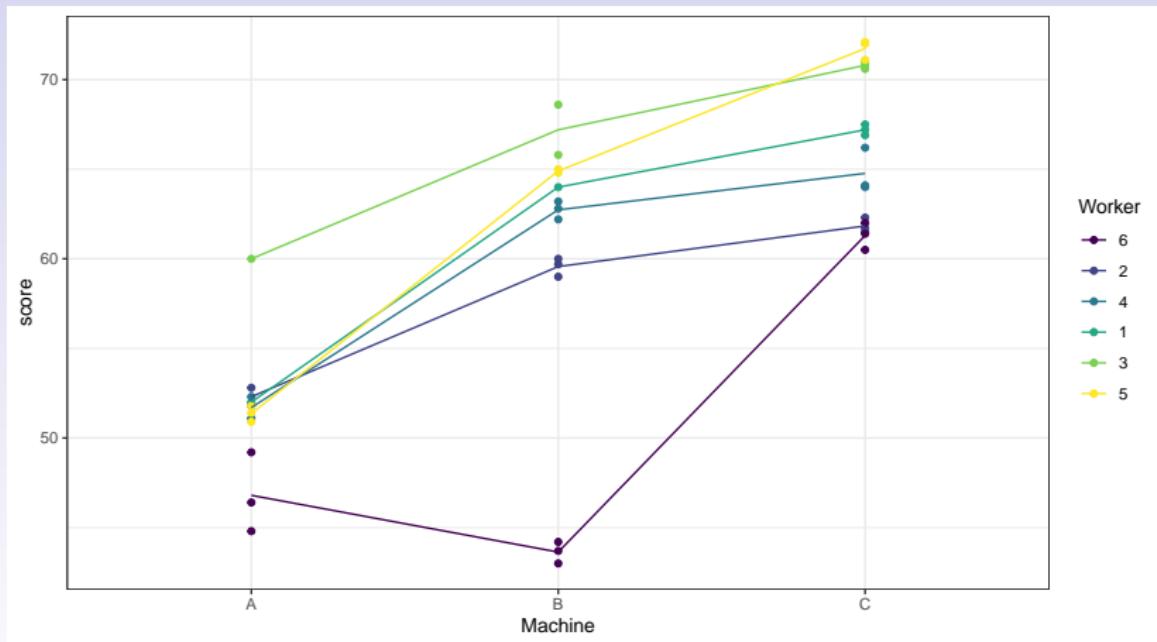
Plot of raw data



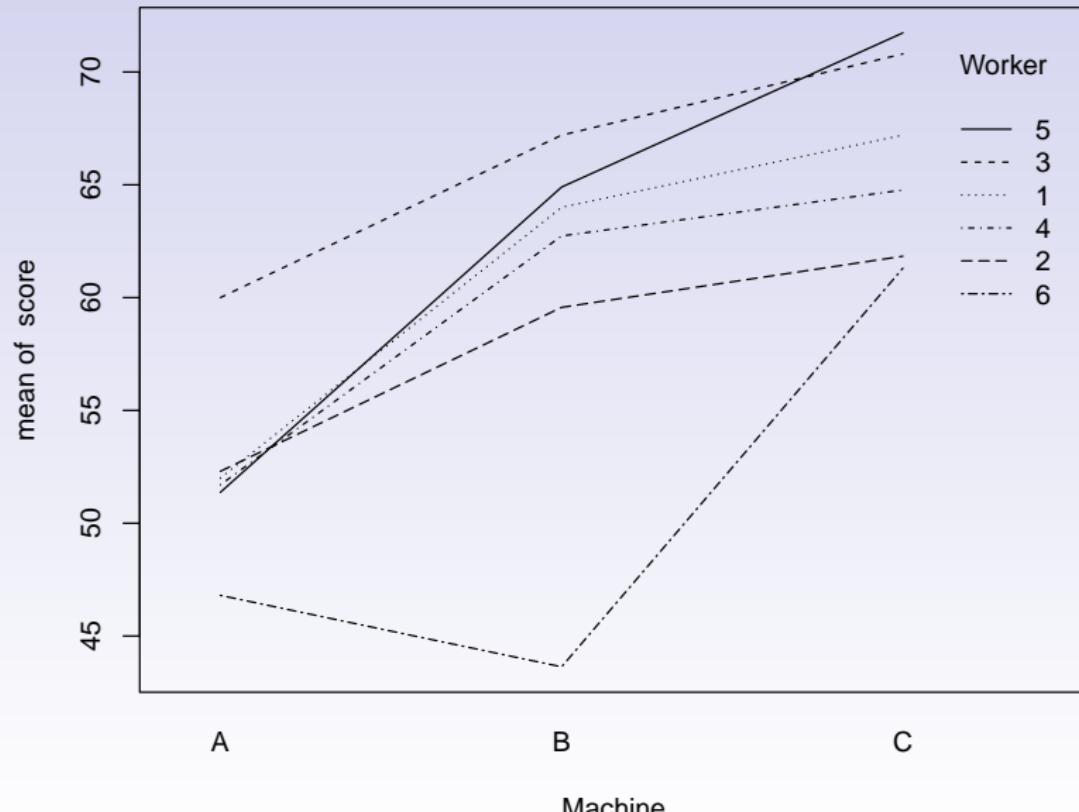
Plot of raw data from qplot



Interaction plot from ggplot



Interaction plot



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme - unbalanced design

```
> mach3 <- lme( score ~ Machine, random = ~1 | Worker/Machine,  
                  data = MachinesUnbal)
```

```
> summary(mach3)
```

Linear mixed-effects model fit by REML

Data: MachinesUnbal

AIC BIC logLik

193.8715 204.1529 -90.93575

Random effects:

Formula: ~1 | Worker

(Intercept)

StdDev: 4.738733

Formula: ~1 | Machine %in% Worker

(Intercept) Residual

StdDev: 3.77281 0.9332026

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme - unbalanced design

Fixed effects: score ~ Machine

	Value	Std.Error	DF	t-value	p-value
(Intercept)	52.35400	2.490613	26	21.020531	0.0000
MachineB	7.96245	2.214705	10	3.595262	0.0049
MachineC	13.91822	2.209363	10	6.299655	0.0001

Correlation:

	(Intr)	MachnB
MachineB	-0.446	
MachineC	-0.447	0.503

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.10231214	-0.49907035	-0.00960014	0.39503219	2.61263385

Number of Observations: 44

Number of Groups:

Worker	Machine	%in% Worker
6		18

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme - unbalanced design

```
> intervals(mach3)
Approximate 95% confidence intervals

Fixed effects:
            lower      est.      upper
(Intercept) 47.234473 52.354000 57.47353
MachineB     3.027775  7.962446 12.89712
MachineC     8.995455 13.918222 18.84099
attr(),"label")
[1] "Fixed effects:"
```



```
Random Effects:
  Level: Worker
            lower      est.      upper
sd((Intercept)) 2.216302 4.738733 10.13201
  Level: Machine
            lower      est.      upper
sd((Intercept)) 2.40915  3.77281  5.908348
```



```
Within-group standard error:
            lower      est.      upper
0.7118781 0.9332026 1.2233375
```

oooooooooooooooooooooooooooo ooooooooooooooooooooo ooooooooooooo ooooooooooooo

Fit using lmer

```
> mach4 <- lmer(score ~ Machine + (1 | Worker) +
+                   (1 | Worker:Machine), data = MachinesUnbal)
> summary(mach4)
Linear mixed model fit by REML ['lmerMod']
Formula: score ~ Machine + (1 | Worker) + (1 | Worker:Machine)
Data: MachinesUnbal

REML criterion at convergence: 181.9
```

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.1023	-0.4991	-0.0096	0.3950	2.6126

Random effects:

Groups	Name	Variance	Std.Dev.
Worker:Machine	(Intercept)	14.2353	3.7730
Worker	(Intercept)	22.4504	4.7382
Residual		0.8709	0.9332

Number of obs: 44, groups: Worker:Machine, 18; Worker, 6

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lmer

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	52.354	2.490	21.022
MachineB	7.962	2.215	3.595
MachineC	13.918	2.209	6.299

Correlation of Fixed Effects:

	(Intr)	MachnB
MachineB	-0.446	
MachineC	-0.447	0.503

```
> confint(mach4, oldNames = FALSE)
```

	2.5 %	97.5 %
sd_(Intercept) Worker:Machine	2.3791085	5.498943
sd_(Intercept) Worker	1.8798870	9.354293
sigma	0.7274193	1.256267
(Intercept)	47.3967260	57.316833
MachineB	3.6645453	12.266842
MachineC	9.6280610	18.212152

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using proc mixed

```
options nocenter ps=64 ls=76;
data a;
  do machine=1 to 3;
    do person=1 to 6;
      do rep=1 to 3;
        input y @;
        output;
      end;
    end;
  end;
cards;
52.0 52.8 53.1
51.8 52.8 53.1
60.0 60.2 58.4
51.1 52.3 50.3
50.9 51.8 51.4
46.4 44.8 49.2
62.1 62.6 64.0
59.7 60.0 59.0
68.6 65.8 69.7
63.2 62.8 62.2
64.8 65.0 65.4
43.7 44.2 43.0
67.5 67.2 66.9
61.5 61.7 62.3
70.8 70.6 71.0
64.1 66.2 64.0
72.1 72.0 71.1
62.0 61.4 60.5
;
```

```
oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo
```

Fit using proc mixed

```
proc mixed method=reml CL;
  class machine person;
model y = machine / CL ALPHA=.05;
random person machine*person / CL ALPHA=.05;
estimate 'BLUE of machine 1' intercept 1 machine 1 0 0;
estimate 'BLUE of diff between machine 1 and 2' machine 1 -1 0;
estimate 'BLUP of person 1 average over all machines'
            intercept 3 machine 1 1 1 | person 3 0 0 0 0 0
            machine*person 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
            0/divisor=3;
estimate 'BLUP of person 1 using machine 1'
            intercept 1 machine 1 0 0 | person 1 0 0 0 0 0
            machine*person 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0;
contrast 'machine 1 vs 2' machine 1 -1 0;
lsmeans machine;
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using proc mixed

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
person	22.8584	0.05	7.6910	251.49
machine*person	13.9095	0.05	6.7031	44.2384
Residual	0.9246	0.05	0.6115	1.5601

Fit Statistics

-2 Res Log Likelihood	215.7
AIC (smaller is better)	221.7
AICC (smaller is better)	222.2

The Mixed Procedure

Fit Statistics

BIC (smaller is better)	221.1
-------------------------	-------

Solution for Fixed Effects

Standard

Effect	machine	Estimate	Error	DF	t Value	Pr > t
Alpha						
Intercept		66.2722	2.4858	5	26.66	<.0001
0.05						
machine	1	-13.9167	2.1770	10	-6.39	<.0001
0.05						
machine	2	-5.9500	2.1770	10	-2.73	0.0211
0.05						
machine	3	0

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using proc mixed

Solution for Fixed Effects

Effect	machine	Lower	Upper
Intercept		59.8822	72.6623
machine	1	-18.7673	-9.0661
machine	2	-10.8006	-1.0994
machine	3	.	.

Solution for Random Effects

Effect	machine	person	Estimate	Pred	DF	t Value	Pr > t	Std Err
person		1	1.0445	2.6610	36	0.39	0.6970	
person		2	-1.3759	2.6610	36	-0.52	0.6083	
person		3	5.3608	2.6610	36	2.01	0.0515	
person		4	-0.05982	2.6610	36	-0.02	0.9822	
person		5	2.5446	2.6610	36	0.96	0.3453	
person		6	-7.5143	2.6610	36	-2.82	0.0077	
machine*person	1	1	-0.7501	2.3875	36	-0.31	0.7552	
machine*person	1	2	1.5526	2.3875	36	0.65	0.5196	
machine*person	1	3	1.7776	2.3875	36	0.74	0.4614	
machine*person	1	4	-1.0394	2.3875	36	-0.44	0.6659	
machine*person	1	5	-3.4569	2.3875	36	-1.45	0.1563	
machine*person	1	6	1.9163	2.3875	36	0.80	0.4275	
machine*person	2	1	1.5000	2.3875	36	0.63	0.5338	
machine*person	2	2	0.6069	2.3875	36	0.25	0.8008	
machine*person	2	3	2.2994	2.3875	36	0.96	0.3419	
machine*person	2	4	2.4174	2.3875	36	1.01	0.3181	
machine*person	2	5	2.1521	2.3875	36	0.90	0.3734	
machine*person	2	6	-8.9757	2.3875	36	-3.76	0.0006	
machine*person	3	1	-0.1142	2.3875	36	-0.05	0.9621	
machine*person	3	2	-2.9966	2.3875	36	-1.26	0.2175	
machine*person	3	3	-0.8149	2.3875	36	-0.34	0.7348	
machine*person	3	4	-1.4144	2.3875	36	-0.59	0.5573	
machine*person	3	5	2.8532	2.3875	36	1.20	0.2399	
machine*person	3	6	2.4870	2.3875	36	1.04	0.3045	

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using proc mixed

Solution for Random Effects

Effect	machine	person	Alpha	Lower	Upper
person		1	0.05	-4.3522	6.4413
person		2	0.05	-6.7726	4.0209
person		3	0.05	-0.03597	10.7575
person		4	0.05	-5.4566	5.3369
person		5	0.05	-2.8521	7.9414
person		6	0.05	-12.9110	-2.1175
machine*person	1	1	0.05	-5.5923	4.0920
machine*person	1	2	0.05	-3.2896	6.3947
machine*person	1	3	0.05	-3.0646	6.6198
machine*person	1	4	0.05	-5.8815	3.8028
machine*person	1	5	0.05	-8.2991	1.3852
machine*person	1	6	0.05	-2.9259	6.7584
machine*person	2	1	0.05	-3.3422	6.3422
machine*person	2	2	0.05	-4.2353	5.4490
machine*person	2	3	0.05	-2.5428	7.1416
machine*person	2	4	0.05	-2.4248	7.2595
machine*person	2	5	0.05	-2.6901	6.9943
machine*person	2	6	0.05	-13.8179	-4.1335
machine*person	3	1	0.05	-4.9564	4.7279
machine*person	3	2	0.05	-7.8388	1.8455
machine*person	3	3	0.05	-5.6571	4.0272
machine*person	3	4	0.05	-6.2566	3.4278
machine*person	3	5	0.05	-1.9889	7.6954
machine*person	3	6	0.05	-2.3552	7.3291

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using proc mixed

Type 3 Tests of Fixed Effects

	Num	Den	F Value	Pr > F
Effect machine	2	10	20.58	0.0003

Estimates

Standard

Label	Estimate	Error	DF	t Value	Pr > t
BLUE of machine 1	52.3556	2.4858	10	21.06	<.0001
BLUE of diff between machine 1 and 2	-7.9667	2.1770	10	-3.66	0.0044
BLUP of person 1 average over all machines	60.9064	0.3200	10	190.32	<.0001
BLUP of person 1 using machine 1	54.0357	4.2058	10	12.85	<.0001

Contrasts

	Num	Den	F Value	Pr > F
Label machine 1 vs 2	1	10	13.39	0.0044

Least Squares Means

Standard

Effect	machine	Estimate	Error	DF	t Value	Pr > t
machine	1	52.3556	2.4858	10	21.06	<.0001
machine	2	60.3222	2.4858	10	24.27	<.0001
machine	3	66.2722	2.4858	10	26.66	<.0001



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- **Wheat data**
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors

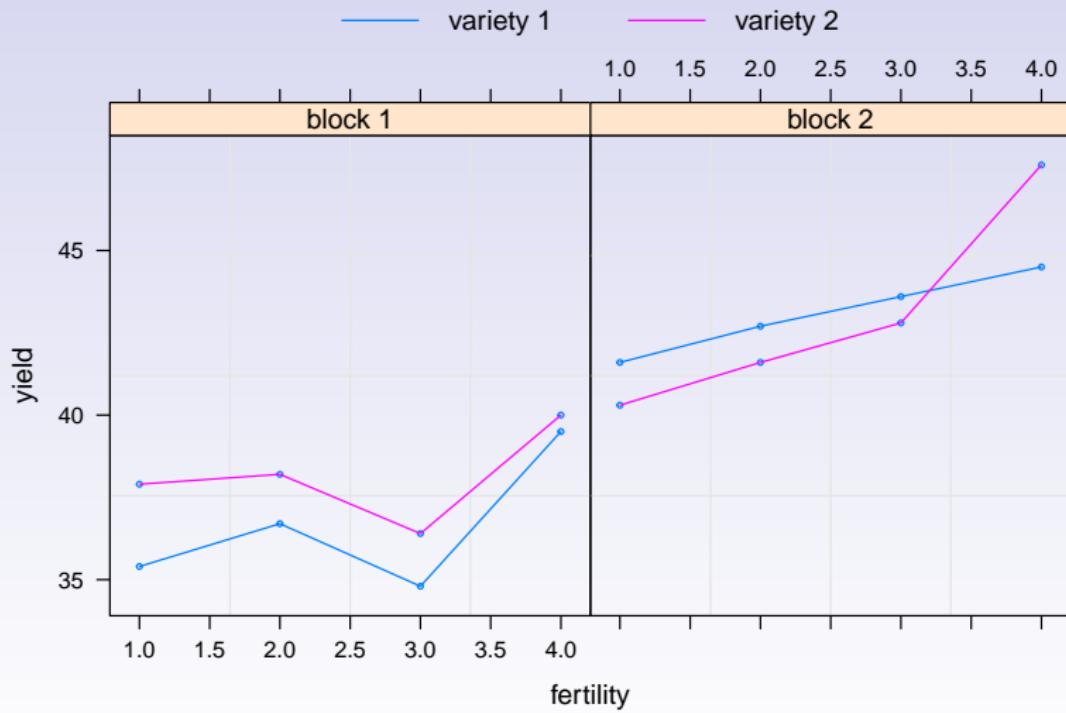
oooooooooooooooooooooooooooo

Example: wheat data continued

Input data and plot

```
> library(nlme)
> wheat <- groupedData( yield ~ fertility | block,
  data = data.frame(yield=c(35.4,37.9,36.7,38.2,34.8,
  36.4,39.5,40.0,41.6,40.3,42.7,41.6,43.6,42.8,
  44.5,47.6),
  block=rep(c(`block 1','block 2'),rep(8,2)),
  fertility=rep(rep(1:4,rep(2,4)),2),
  variety=rep(c(`variety 1','variety 2'),8)))
> plot(wheat,inner=~variety,layout=c(2,1),cex=.4,asp=1)
```

Example: wheat data continued



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme

```
> wheat$fertility <- rep(rep(as.character(1:4), rep(2,4)),2)
> wheat$variety <- rep(c(''1'', ''2''), 8)
> fit1 <- lme(yield~fertility*variety, data=wheat,
+               random=~1|block/fertility)
> summary(fit1)
Linear mixed-effects model fit by REML
Data: wheat
      AIC      BIC      logLik
 60.61684 61.4907 -19.30842

Random effects:
Formula: ~1 | block
            (Intercept)
StdDev:    4.012377

Formula: ~1 | fertility %in% block
            (Intercept) Residual
StdDev:    0.3175427 1.451723
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme

Fixed effects: yield ~ fertility * variety

	Value	Std.Error	DF	t-value	p-value
(Intercept)	38.5	3.025517	4	12.725100	0.0002
fertility2	1.2	1.486046	3	0.807512	0.4785
fertility3	0.7	1.486046	3	0.471049	0.6697
fertility4	3.5	1.486046	3	2.355243	0.0998
variety2	0.6	1.451723	4	0.413302	0.7006
fertility2:variety2	-0.4	2.053047	4	-0.194832	0.8550
fertility3:variety2	-0.2	2.053047	4	-0.097416	0.9271
fertility4:variety2	1.2	2.053047	4	0.584497	0.5903

Correlation:

	(Intr)	frtlt2	frtlt3	frtlt4	varty2	frt2:2	frt3:2
fertility2	-0.246						
fertility3	-0.246	0.500					
fertility4	-0.246	0.500	0.500				
variety2	-0.240	0.488	0.488	0.488			
fertility2:variety2	0.170	-0.691	-0.345	-0.345	-0.707		
fertility3:variety2	0.170	-0.345	-0.691	-0.345	-0.707	0.500	
fertility4:variety2	0.170	-0.345	-0.345	-0.691	-0.707	0.500	0.500

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-1.070631e+00	-3.436653e-01	-2.456368e-15	3.436653e-01	1.070631e+00

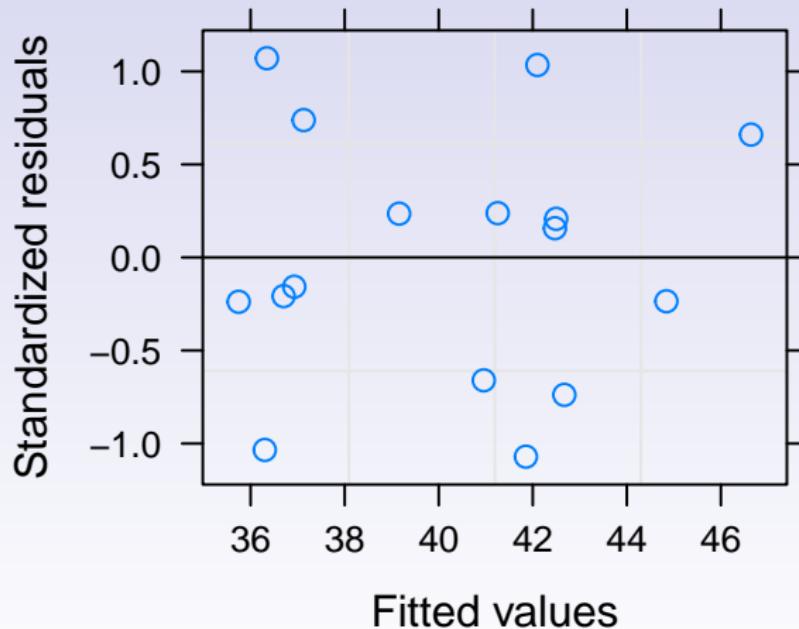
Number of Observations: 16

Number of Groups:

block	fertility	%in%	block
2			8

Residual plot

```
> plot(fit1)
```



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme

```
> intervals(fit1)
Approximate 95% confidence intervals

Fixed effects:
            lower   est.   upper
(Intercept) 30.099819 38.5 46.900181
fertility2  -3.529262  1.2  5.929262
fertility3  -4.029262  0.7  5.429262
fertility4  -1.229262  3.5  8.229262
variety2    -3.430630  0.6  4.630630
fertility2:variety2 -6.100171 -0.4  5.300171
fertility3:variety2 -5.900171 -0.2  5.500171
fertility4:variety2 -4.500171  1.2  6.900171
attr(),"label")
[1] "Fixed effects:"
```



```
Random Effects:
  Level: block
            lower   est.   upper
sd((Intercept)) 0.9787476 4.012377 16.44874
  Level: fertility
            lower   est.   upper
sd((Intercept)) 2.581763e-06 0.3175427 39056.01
```



```
Within-group standard error:
            lower   est.   upper
0.7250793 1.4517231 2.9065786
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lmer

```
> library(lme4)
> wheat$wholeplot <- rep(as.character(1:8),rep(2,8))
> fit2 <- lmer(yield~fertility*variety+(1|block)+(1|wholeplot),
+                 data=wheat)
> summary(fit2)
Linear mixed model fit by REML
Formula: yield ~ fertility * variety + (1 | block) + (1 | wholeplot)
Data: wheat
AIC    BIC logLik deviance REMLdev
60.62 69.12 -19.31     55.05   38.62
Random effects:
Groups      Name        Variance Std.Dev.
wholeplot  (Intercept) 0.10083  0.31754
block       (Intercept) 16.09917 4.01238
Residual                2.10750 1.45172
Number of obs: 16, groups: wholeplot, 8; block, 2
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using lmer

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	38.500	3.025	12.726
fertility2	1.200	1.486	0.808
fertility3	0.700	1.486	0.471
fertility4	3.500	1.486	2.355
variety2	0.600	1.452	0.413
fertility2:variety2	-0.400	2.053	-0.195
fertility3:variety2	-0.200	2.053	-0.097
fertility4:variety2	1.200	2.053	0.584

Correlation of Fixed Effects:

	(Intr)	frtlty2	frtlty3	frtlty4	varty2	frt2:2	frt3:2
fertility2	-0.246						
fertility3	-0.246	0.500					
fertility4	-0.246	0.500	0.500				
variety2	-0.240	0.488	0.488	0.488			
frtlty2:vr2	0.170	-0.691	-0.345	-0.345	-0.707		
frtlty3:vr2	0.170	-0.345	-0.691	-0.345	-0.707	0.500	
frtlty4:vr2	0.170	-0.345	-0.345	-0.691	-0.707	0.500	0.500

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lmer

```
> confint(fit2)
```

	2.5 %	97.5 %
.sig01	0.0000000	1.350558
.sig02	1.2973570	11.974888
.sigma	0.6735066	1.596461
(Intercept)	31.5312431	45.468071
fertility2	-1.0183154	3.418315
fertility3	-1.5183154	2.918315
fertility4	1.2816846	5.718315
variety2	-1.5971771	2.797179
fertility2:variety2	-3.5072777	2.707280
fertility3:variety2	-3.3072777	2.907280
fertility4:variety2	-1.9072777	4.307280

```
oooooooooooooooooooooooooooo
```

Fit using proc mixed

```
options nocenter ps=64 ls=76;
data a;
do block=1 to 2;
  do fertility=1 to 4;
    do variety=1 to 2;
      input yield @;
      s = (block-1)*2+variety;
      output;
    end;
  end;
end;
cards;
35.4 37.9 36.7 38.2 34.8 36.4 39.5 40.0
41.6 40.3 42.7 41.6 43.6 42.8 44.5 47.6
;
```

```
oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo
```

Fit using proc mixed

```
/* reorder to match results in R */
proc format;
  value fertfmt 1 = 'f1' 2 = '2' 3 = '3' 4 = '4';

proc format;
  value vartfmt 1 = 'v1' 2 = '2';

proc mixed method=reml CL;
  format fertility fertfmt. variety vartfmt.;
  class block fertility variety;
  model yield = fertility | variety / CL ALPHA=.05;
  random block block*fertility;
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using proc mixed

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
block	16.0992	0.05	3.1529	20880
block*fertility	0.1008	0.05	0.04510	4.11E224
Residual	2.1075	0.05	0.7565	17.4023

Fit Statistics

-2 Res Log Likelihood	38.6
AIC (smaller is better)	44.6

The Mixed Procedure

Fit Statistics

AICC (smaller is better)	50.6
BIC (smaller is better)	40.7

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using proc mixed

Solution for Fixed Effects

Standard

Effect	fertility	variety	Estimate	Error	DF	t Value
Intercept			38.5000	3.0255	1	12.73
fertility	2		1.2000	1.4860	3	0.81
fertility	3		0.7000	1.4860	3	0.47
fertility	4		3.5000	1.4860	3	2.36
fertility	f1		0	.	.	.
variety		2	0.6000	1.4517	4	0.41
variety		v1	0	.	.	.
fertility*variety	2	2	-0.4000	2.0530	4	-0.19
fertility*variety	2	v1	0	.	.	.
fertility*variety	3	2	-0.2000	2.0530	4	-0.10
fertility*variety	3	v1	0	.	.	.
fertility*variety	4	2	1.2000	2.0530	4	0.58
fertility*variety	4	v1	0	.	.	.
fertility*variety	f1	2	0	.	.	.
fertility*variety	f1	v1	0	.	.	.

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit using proc mixed

Effect	fertility	variety	Pr > t	Alpha	Lower	Upper
Intercept			0.0499	0.05	0.05717	76.9428
fertility	2		0.4785	0.05	-3.5293	5.9293
fertility	3		0.6697	0.05	-4.0293	5.4293
fertility	4		0.0998	0.05	-1.2293	8.2293
fertility	f1	
variety		2	0.7006	0.05	-3.4306	4.6306
variety		v1
fertility*variety	2	2	0.8550	0.05	-6.1002	5.3002
fertility*variety	2	v1
fertility*variety	3	2	0.9271	0.05	-5.9002	5.5002
fertility*variety	3	v1
fertility*variety	4	2	0.5903	0.05	-4.5002	6.9002
fertility*variety	4	v1
fertility*variety	f1	2
fertility*variety	f1	v1

Type 3 Tests of Fixed Effects

	Num DF	Den DF	F Value	Pr > F
Effect				
fertility	3	3	5.80	0.0914
variety	1	4	1.07	0.3599
fertility*variety	3	4	0.25	0.8612



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data**
- Orthodont data

6

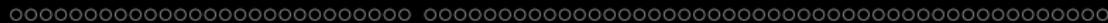
Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Example: lab data continued

In a cooperative trial in analytical chemistry, seven specimens were sent to six laboratories, each three times a month apart for duplicated analysis. The response is the concentration of (unspecified) analyte in g/kg. The purpose of the study was to assess components of variation in cooperative trials. We use the data from Specimen 1 shown in the following table.

Batch	Laboratory					
	1	2	3	4	5	6
1	0.29	0.40	0.40	0.90	0.44	0.38
	0.33	0.40	0.35	1.30	0.44	0.39
2	0.33	0.43	0.38	0.90	0.45	0.40
	0.32	0.36	0.32	1.10	0.45	0.46
3	0.34	0.42	0.38	0.90	0.42	0.72
	0.31	0.40	0.33	0.90	0.46	0.79



Nested model

We considered the following model

$$\begin{aligned}y_{ijk} &= \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}, \\ i &= 1, \dots, 6; j = 1, \dots, 3; k = 1, \dots, 2\end{aligned}$$

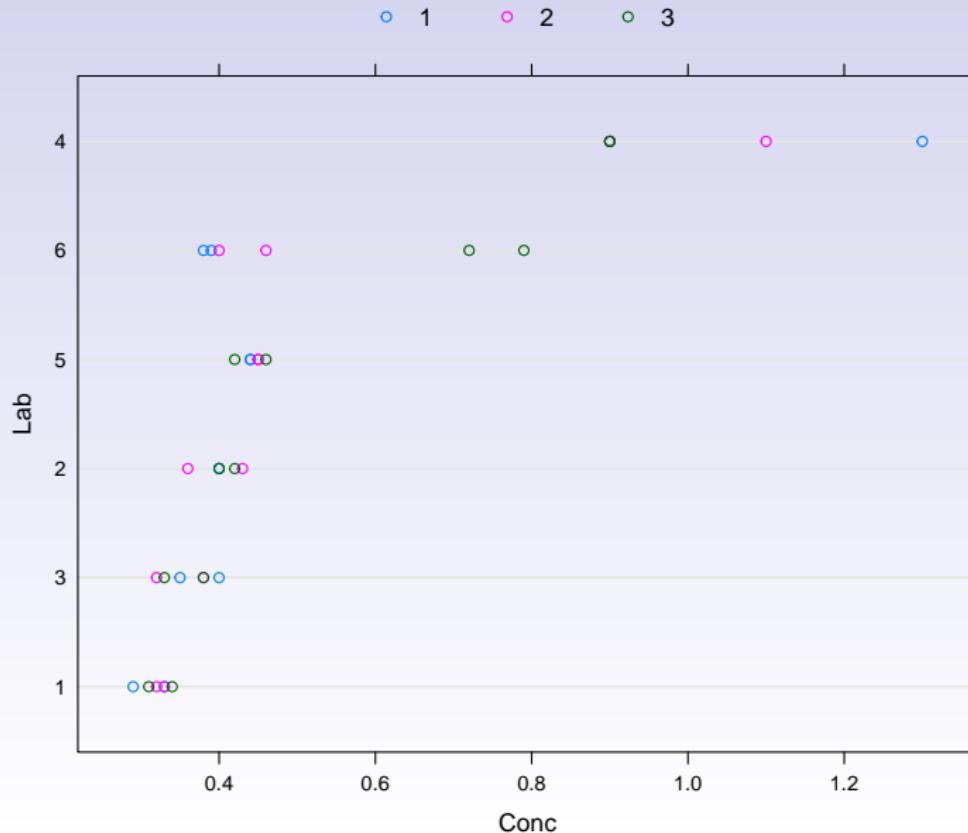
- y_{ijk} : concentration for the k th replicate of batch j in lab i
- μ : overall mean
- α_i : effects of Lab, $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_a^2)$
- $\beta_{j(i)}$: effects of Bat $\beta_{j(i)} \stackrel{iid}{\sim} N(0, \sigma_b^2)$
- ϵ_{ijk} : random errors, $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$
- α_i , $\beta_{j(i)}$ and ϵ_{ijk} are mutually independent

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

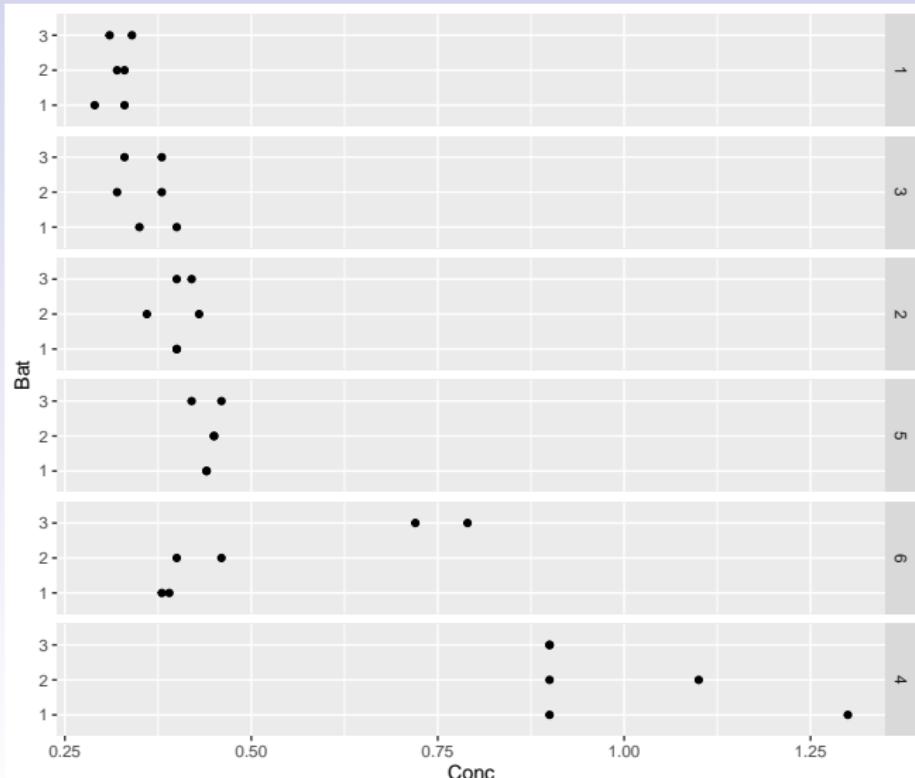
Fit using lme

```
> library(MASS)
> lab <- groupedData( Conc ~ Bat | Lab,
  data = data.frame(Conc=coop$Conc[coop$Spc=='S1'],
  Lab=rep(as.character(1:6), rep(6,6)),
  Bat=rep(rep(as.character(1:3), rep(2,3)), 6)))
> plot(lab, inner=~Bat)
> ggplot(lab, aes(y = Bat, x = Conc)) + geom_point() +
  facet_grid(Lab ~ .)
```

Plot of lab data



Plot of lab data



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme

```
> fit1 <- lme(Conc~1, random=~1|Lab/Bat, data=lab)
> summary(fit1)
```

Linear mixed-effects model fit by REML

Data: lab

AIC	BIC	logLik
-34.04316	-27.82177	21.02158

Random effects:

Formula: ~1 | Lab

(Intercept)

StdDev: 0.2452922

Formula: ~1 | Bat %in% Lab

(Intercept) Residual

StdDev: 0.07326702 0.07935504

Fixed effects: Conc ~ 1

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.5080556	0.1024753	18	4.957834	1e-04

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-1.95117382	-0.36261669	-0.03009679	0.18536155	3.08946386

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lme

```
> intervals(fit1)
```

Approximate 95% confidence intervals

Fixed effects:

	lower	est.	upper
(Intercept)	0.2927629	0.5080556	0.7233482

```
attr(,'label')
```

```
[1] ``Fixed effects:'''
```

Random Effects:

Level: Lab

	lower	est.	upper
sd((Intercept))	0.1281407	0.2452922	0.4695484

Level: Bat

	lower	est.	upper
sd((Intercept))	0.03775365	0.07326702	0.1421864

Within-group standard error:

	lower	est.	upper
0.05724107	0.07935504	0.11001230	

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lmer

```
> library(lme4)
> fit2 <- lmer(Conc~1+(1|Lab)+(1|Lab:Bat), data=lab)
> summary(fit2)
Linear mixed model fit by REML
Formula: Conc ~ 1 + (1 | Lab) + (1 | Lab:Bat)
Data: lab
      AIC      BIC logLik deviance REMLdev
-34.04 -27.71  21.02   -44.78   -42.04
Random effects:
Groups     Name        Variance Std.Dev.
Lab:Bat   (Intercept) 0.0053681 0.073267
Lab       (Intercept) 0.0601682 0.245292
Residual           0.0062972 0.079355
Number of obs: 36, groups: Lab:Bat, 18; Lab, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.5081    0.1025   4.958

> confint(fit2, oldNames = FALSE)

              2.5 %    97.5 %
sd_(Intercept)|Lab:Bat 0.00000000 0.1355194
sd_(Intercept)|Lab     0.13123985 0.4563944
sigma                 0.05910359 0.1144730
(Intercept)          0.29104449 0.7250667
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using proc mixed

```
data a;
do Lab=1 to 6;
  do Bat=1 to 3;
    do rep=1 to 2;
      input y @;
      output;
    end;
  end;
end;
cards;
0.29 0.33 0.33 0.32 0.34 0.31
0.40 0.40 0.43 0.36 0.42 0.40
0.40 0.35 0.38 0.32 0.38 0.33
0.90 1.30 0.90 1.10 0.90 0.90
0.44 0.44 0.45 0.45 0.42 0.46
0.38 0.39 0.40 0.46 0.72 0.79
;
```



Fit using proc mixed

```
proc mixed method=reml CL;  
  class Lab Bat;  
  model y = / CL ALPHA=.05;  
  random Lab Bat (Lab);
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using proc mixed

Covariance Parameter Estimates					
Cov Parm	Estimate	Alpha	Lower	Upper	
Lab	0.06017	0.05	0.02267	0.4098	
Bat(Lab)	0.005368	0.05	0.001992	0.03880	
Residual	0.006297	0.05	0.003595	0.01377	
Fit Statistics					
-2 Res Log Likelihood			-42.0		
AIC (smaller is better)			-36.0		
AICC (smaller is better)			-35.3		
The Mixed Procedure					
Fit Statistics					
BIC (smaller is better)			-36.7		
Solution for Fixed Effects					
Standard					
Effect	Estimate	Error	DF	t Value	Pr > t
Intercept	0.5081	0.1025	5	4.96	0.0043
Solution for Fixed Effects					
Effect	Lower		Upper		
Intercept	0.2446		0.7715		



Example: orthodont - a longitudinal data

Investigators at the University of North Carolina Dental School followed the growth of 27 children (16 males, 11 females) from age 8 until age 14. Every two years they measured the distance between the pituitary and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head.



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Example: orthodont data

The Orthodont data frame contains the following columns:

- `distance`: a numeric vector of distances from the pituitary to the pterygomaxillary fissure (mm). These distances are measured on x-ray images of the skull.
- `age`: a numeric vector of ages of the subject (yr).
- `Subject`: an ordered factor indicating the subject on which the measurement was made. The levels are labelled 'M01' to 'M16' for the males and 'F01' to 'F13' for the females. The ordering is by increasing average distance within sex.
- `Sex` a factor with levels 'Male' and 'Female'

oooooooooooooooooooooooooooo

Example: orthodont data

```
> library(nlme)
> data(Orthodont)
> Orthodont
Grouped Data: distance ~ age | Subject
  distance age Subject Sex
1       26.0    8     M01 Male
2       25.0   10     M01 Male
...
> sapply(Orthodont,data.class)
distance      age   Subject      Sex
"numeric" "numeric" "ordered" "factor"
```

oooooooooooooooooooooooooooo

Example: orthodont data

```
> gsummary(Orthodont)
   distance age Subject    Sex
M16     23.000  11      M16 Male
M05     23.000  11      M05 Male
M02     23.375  11      M02 Male
M11     23.625  11      M11 Male
M07     23.750  11      M07 Male
M08     23.875  11      M08 Male
M03     24.250  11      M03 Male
M12     24.250  11      M12 Male
M13     24.250  11      M13 Male
M14     24.875  11      M14 Male
M09     25.125  11      M09 Male
M15     25.875  11      M15 Male
M06     26.375  11      M06 Male
M04     26.625  11      M04 Male
M01     27.750  11      M01 Male
M10     29.500  11      M10 Male
F10     18.500  11      F10 Female
F09     21.125  11      F09 Female
F06     21.125  11      F06 Female
F01     21.375  11      F01 Female
F05     22.625  11      F05 Female
F07     23.000  11      F07 Female
F02     23.000  11      F02 Female
F08     23.375  11      F08 Female
F03     23.750  11      F03 Female
F04     24.875  11      F04 Female
F11     26.375  11      F11 Female
```

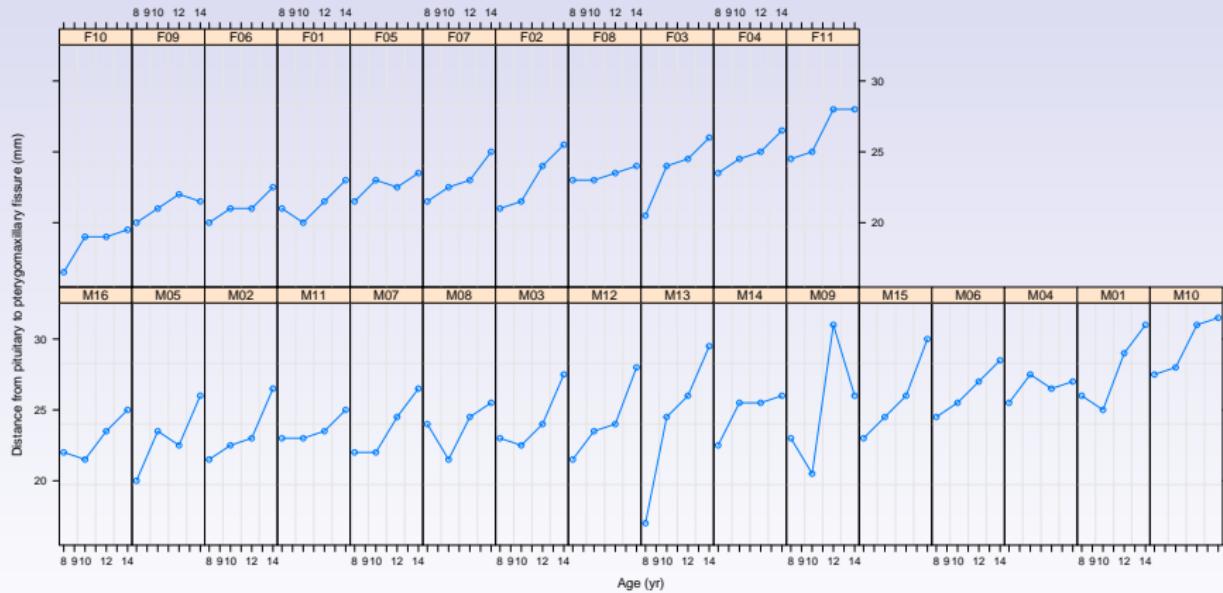
```
oooooooooooooooooooooooooooo
```

Example: orthodont data

```
> str(Orthodont)
Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame':
108 obs. of  4 variables:
 $ distance: num  26 25 29 31 21.5 22.5 23 26.5 23 22.5 ...
 $ age      : num  8 10 12 14 8 10 12 14 8 10 ...
 $ Subject  : Ord.factor w/ 27 levels "M16"<"M05"<"M02"<...: 15 15 15 15 3 3 3 3 7 7 ...
 $ Sex      : Factor w/ 2 levels "Male","Female": 1 1 1 1 1 1 1 1 1 1 ...
- attr(*, "outer")=Class 'formula' language ~Sex
... ...- attr(*, ".Environment")=<environment: R_GlobalEnv>
- attr(*, "formula")=Class 'formula' language distance ~ age | Subject
... ...- attr(*, ".Environment")=<environment: R_GlobalEnv>
- attr(*, "labels")=List of 2
..$ x: chr "Age"
..$ y: chr "Distance from pituitary to pterygomaxillary fissure"
- attr(*, "units")=List of 2
..$ x: chr "(yr)"
..$ y: chr "(mm)"
- attr(*, "FUN")=function (x)
..- attr(*, "source")= chr "function (x) max(x, na.rm = TRUE)"
- attr(*, "order.groups")= logi TRUE
```

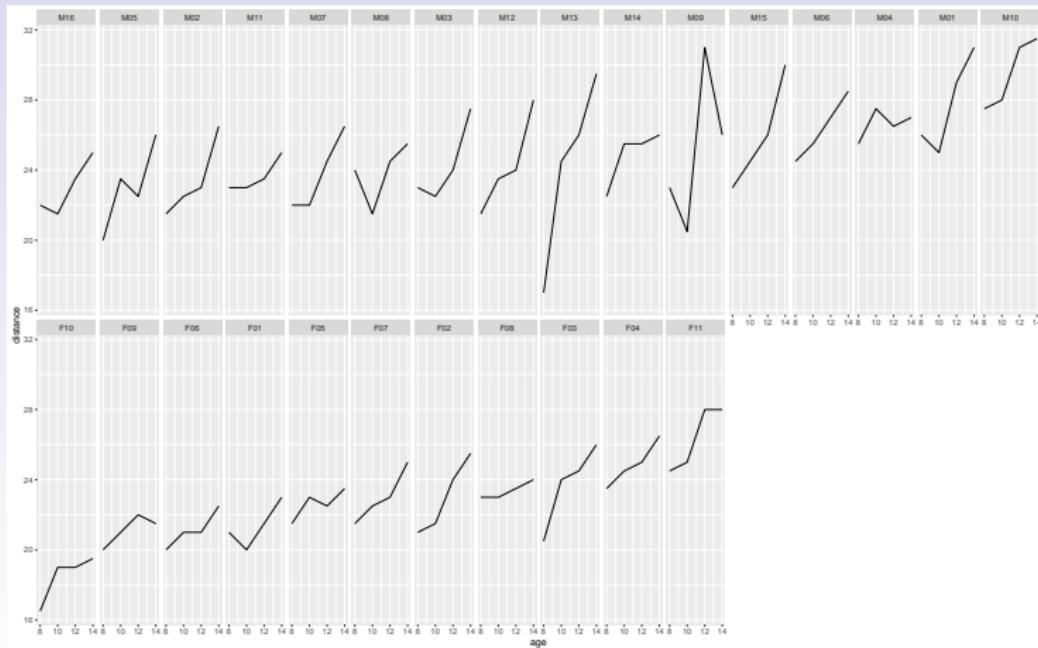
Plot of orthodont data

```
plot(Orthodont, layout=c(16, 2))
```



Plot of orthodont data - ggplot

```
qplot(age, distance, data=Orthodont, geom="line") +  
  facet_wrap(~Subject, ncol=16)
```

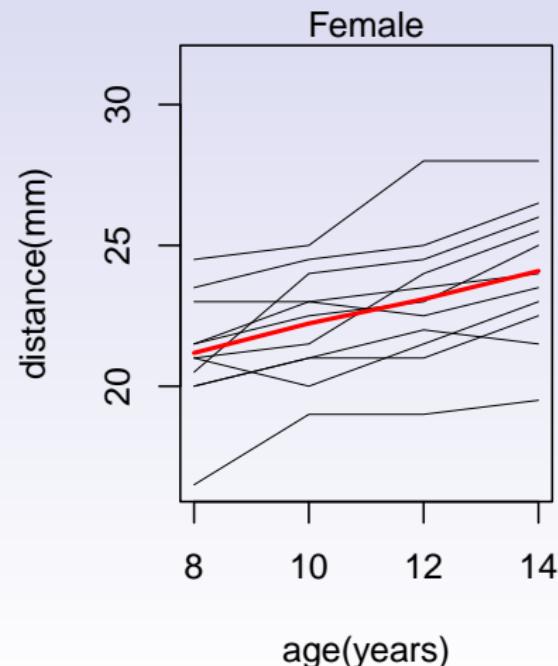
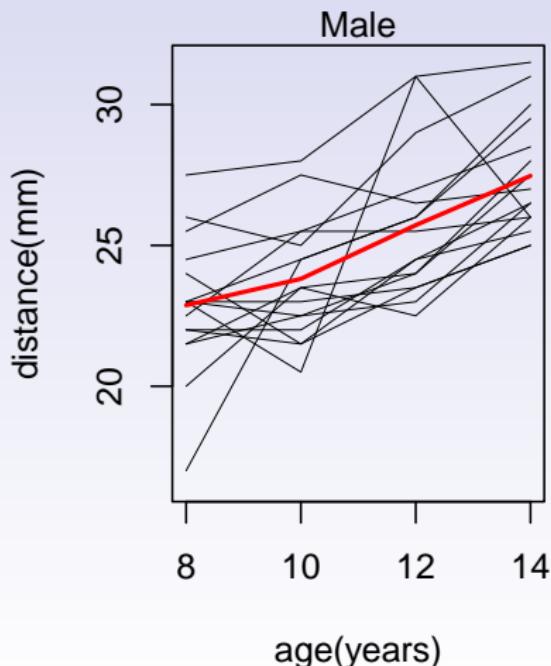


oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Spaghetti plot

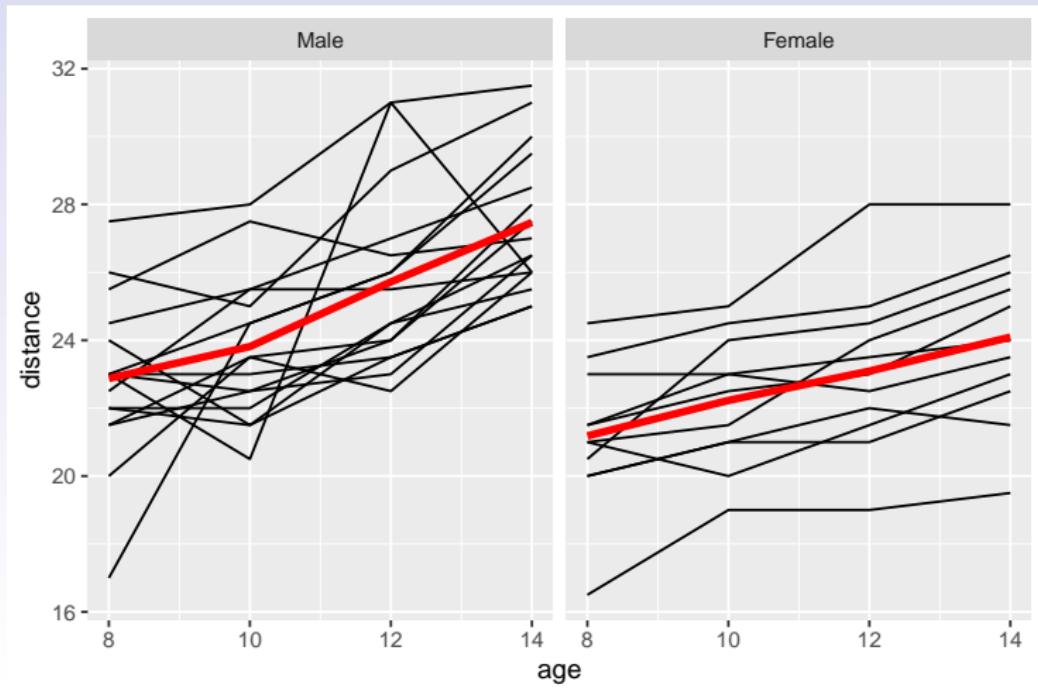
```
attach(Orthodont)
par(mfrow=c(1,2))
t <- c(8,10,12,14)
sub <- unique(Subject)
plot(age, distance, xlab='age (years)', ylab='distance (mm)',
      type='n',axes=F)
box(); axis(2); axis(1,at=t)
for (i in 1:16) lines(age[Subject==sub[i]],
                      distance[Subject==sub[i]],lwd=.2)
lines(unique(age),apply(matrix(distance[Sex=='Male'],
                                 16,4,byrow=T),2,mean),lwd=2,col='red')
mtext('Male')
plot(age, distance, xlab='age (years)', ylab='distance (mm)',
      type='n',axes=F)
box(); axis(2); axis(1,at=t)
for (i in 17:27) lines(age[Subject==sub[i]],
                      distance[Subject==sub[i]],lwd=.2)
lines(unique(age),apply(matrix(distance[Sex=='Female'],
                                 11,4,byrow=T),2,mean),lwd=2,col='red')
mtext('Female')
```

Spaghetti plot



Spaghetti plot - ggplot

```
ggplot(Orthodont, aes(x=age, y=distance, group=Subject)) +  
  geom_line() + stat_summary(aes(group=1), geom='line', fun.y=mean,  
  color='red', size=1.5) + facet_grid(.~Sex)
```





Observations from the plots

- Overall, distance increases with age
- Distances are larger for boys on average
- The rate of change seems similar
- Large variation between subjects. Each child has his/her own trajectory. For a given child and on average, the trajectory looks roughly linear



Questions of interest

- ➊ Growth patterns for boys and girls
- ➋ Are distances larger for boys than for girls?
- ➌ Do boys and girls have similar rates of changes?

We have 3 factors: Sex, Subject and age. Sex and Subject discrete factors. Note that Subject is nested within Sex. Consider these subjects as random samples from a population. Then the factor Subject is random. Other two factors are fixed. Since all observations are taken at the same ages, we may treat the age factor as either discrete or continuous.



Questions of interest

- ➊ Growth patterns for boys and girls
- ➋ Are distances larger for boys than for girls?
- ➌ Do boys and girls have similar rates of changes?

We have 3 factors: Sex, Subject **and** age. Sex **and** Subject discrete factors. Note that Subject is nested within Sex. Consider these subjects as random samples from a population. Then the factor Subject is random. Other two factors are fixed. Since all observations are taken at the same ages, we may treat the age factor as either discrete or continuous.

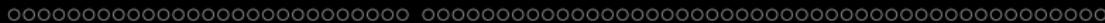


Classical nested model

Treating age as a discrete factor, we may consider the following nested model:

$$y_{ijk} = \mu + \xi_k + \alpha_{i(k)} + \beta_j + (\xi\beta)_{kj} + (\alpha\beta)_{i(k)j} + \epsilon_{ijk}$$

- y_{ijk} : distance at age j of subject i in group k where $k = 1$ corresponds boys and $k = 2$ corresponds girls
- ξ_k : main effect of Sex, $\sum \xi_k = 0$
- $\alpha_{i(k)}$: main effect of Subject, $\alpha_{i(k)} \stackrel{iid}{\sim} N(0, \sigma_1^2)$
- β_j : main effect of age, $\sum \beta_j = 0$
- $(\xi\beta)_{kj}$: interaction between Sex and age,
 $\sum_k (\xi\beta)_{kj} = \sum_j (\xi\beta)_{kj} = 0$
- $(\alpha\beta)_{i(k)j}$: interaction between Subject and age,
 $(\alpha\beta)_{i(k)j} \stackrel{iid}{\sim} N(0, \sigma_2^2)$
- ϵ_{ijk} are random errors, $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$
- all random effects and random errors are mutually independent



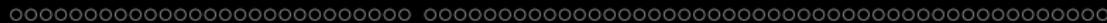
Growth curve model

Now treat age as a continuous variable. Assume a linear relationship between distance and age. We have the following growth curve (also called random coefficient) model:

$$y_{ijk} = \mu + \xi_k + \alpha_{i(k)} + (\beta + \delta_k + \gamma_{i(k)})x_j + \epsilon_{ijk}$$

- y_{ijk} : distance at age j of subject i in group k where $k = 1$ corresponds boys and $k = 2$ corresponds girls
- ξ_k : main effect of Sex, $\sum \xi_k = 0$
- $\alpha_{i(k)}$: main effect of Subject
- β : average slope for all boys and girls
- δ_k : departure of the slope for Sex k from the average slope, $\sum \delta_k = 0$
- $\gamma_{i(k)}$: departure of the slope for Subject i in group k from the average slope.

$$(\alpha_{i(k)}, \gamma_{i(k)}) \stackrel{iid}{\sim} \mathbf{N}(\mathbf{0}, \Sigma_k)$$

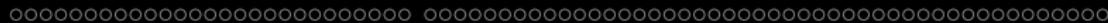


Growth curve model

- ϵ_{ijk} are random errors. $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$
- all random effects and random errors are mutually independent
- x_j : age at level j

Remarks

- ① A linear model with different intercepts and slopes for each subject in a sex group. The intercept depends on Sex (ξ_k) and Subject ($\alpha_{i(k)}$), and the slope depends Sex (δ_k) and Subject ($\gamma_{i(k)}$)
- ② A mixed effects model with two random effects
- ③ Higher order terms may be added to the model, if necessary



Growth curve model

For simplicity, consider female subjects only for now. Thus $k = 2$ and the subscript k is dropped.

$$y_{ij} = \mu + \alpha_i + (\beta + \gamma_i)x_j + \epsilon_{ij}$$

- y_{ij} : distance at age j of subject i
- α_i : main effect of Subject
- β : average slope for all girls
- γ_i : departure of the slope for Subject i from the average slope.

$$(\alpha_i, \gamma_i) \stackrel{iid}{\sim} N(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix})$$

- ϵ_{ij} are random errors. $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$
- all random effects and random errors are mutually independent
- x_j : age at level j

oooooooooooooooooooooooooooo ooooooooooooooooooooo ooooooooooooo ooooooooooooo

Growth curve model

```
> OrthoF <- Orthodont[Orthodont$Sex=='Female',]
> OrthF2 <- lme(distance ~ age, data=OrthoF, random=~age | Subject)
> summary(OrthF2)
Linear mixed-effects model fit by REML
Data: OrthoF
      AIC      BIC      logLik
 149.4287 159.8547 -68.71435

Random effects:
Formula: ~age | Subject
Structure: General positive-definite, Log-Cholesky parametrization
          StdDev   Corr
(Intercept) 1.8841866 (Intr)
age         0.1609278 -0.354
Residual    0.6682746

Fixed effects: distance ~ age
              Value Std.Error DF  t-value p-value
(Intercept) 17.372727 0.7606027 32 22.840737     0
age          0.479545 0.0662140 32  7.242353     0

Correlation:
  (Intr)
age -0.637

Standardized Within-Group Residuals:
      Min        Q1        Med        Q3        Max
-1.85438224 -0.46784889  0.06779759  0.42976633  1.59215841

Number of Observations: 44
Number of Groups: 11
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Growth curve model

It seems that different girls have different intercepts, but similar slopes. So we may also consider a model with random intercepts only:

```
> OrthF1 <- lme(distance ~ age, data=OrthoF, random=~1 | Subject)
> summary(OrthF1)
Linear mixed-effects model fit by REML
Data: OrthoF
      AIC      BIC      logLik
 149.2183 156.169 -70.60916

Random effects:
Formula: ~1 | Subject
          (Intercept) Residual
StdDev:     2.06847  0.7800331

Fixed effects: distance ~ age
              Value Std.Error DF   t-value p-value
(Intercept) 17.372727 0.8587419 32 20.230440      0
age         0.479545 0.0525898 32   9.118598      0

Correlation:
  (Intr)
age -0.674

Standardized Within-Group Residuals:
      Min        Q1        Med        Q3        Max
-2.2736479 -0.7090164  0.1728237  0.4122128  1.6325181

Number of Observations: 44
Number of Groups: 11
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit using lmer

```
> lmer(distance ~ age+(1+age|Subject), data=OrthoF)
Linear mixed model fit by REML
Formula: distance ~ age + (1 + age | Subject)
Data: OrthoF
AIC  BIC logLik deviance REMLdev
149.4 160.1 -68.71    134.6   137.4
Random effects:
Groups      Name        Variance Std.Dev. Corr
Subject (Intercept) 3.550158  1.88419
          age         0.025898  0.16093 -0.354
Residual                 0.446591  0.66827
Number of obs: 44, groups: Subject, 11
```

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	17.37273	0.76060	22.841
age	0.47955	0.06621	7.242

Correlation of Fixed Effects:

(Intr)	age
-0.637	



Inference

After fitting a model and diagnosis, we want to assess the precision of the estimates (confidence intervals), test the significance of various terms in the model (hypothesis test), and compare different models (model selection). Read Section 2.4 of PB.



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors

oooooooooooooooooooooooooooo

Confidence intervals for variance components

lme use the `intervals` function:

```
intervals(fittedobj, which="var-cov")
```

Use the approximate normal distributions of the LME and REML estimates. Some transforms are used in `lme`

lmer use the `confint` function:

```
confint(fittedobj, method =  
c("profile", "Wald", "boot"))
```

proc mixed use option CL:

```
proc mixed CL
```



Hypothesis test

The likelihood ratio test (LRT) is often used to compare nested models.

Caution: the likelihood function is invariant to one-to-one reparameterization of the fixed effects (e.g. different forms of contrasts), REML is not. Thus one cannot compare LME models with different fixed effects structure using REML or construct likelihood ratio statistics using REML



Example: machine data continued

Full model:

$$y_{ijk} = \mu + \alpha_j + \beta_i + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

- y_{ijk} : k th observation from Worker i operating on Machine j
- μ and α_j are fixed effects with $\sum \alpha_j = 0$
- β_i and $(\alpha\beta)_{ij}$ are random effects with $\beta_i \stackrel{iid}{\sim} N(0, \sigma_b^2)$ and $(\alpha\beta)_{ij} \stackrel{iid}{\sim} N(0, \sigma_{ab}^2)$
- ϵ_{ijk} : random errors, $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$
- β_j , $(\alpha\beta)_{ij}$ and ϵ_{ijk} are mutually independent



Example: machine continued

Suppose we are interested in testing the hypothesis that there is no interaction between Worker and Machine

$$H_0 : \sigma_{ab}^2 = 0 \quad \text{vs} \quad H_1 : \sigma_{ab}^2 > 0$$

Reduced model under H_0 :

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

Reduced model is nested within the full model. Thus LRT may be used.

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: machine data continued

```
> mach.main <- lme( score ~ Machine, random = ~1 | Worker,
+                      data = MachinesUnbal)
> mach.inter <- lme( score ~ Machine, random = ~1 | Worker/Machine,
+                      data = MachinesUnbal)
> (LR.app <- anova( mach.main, mach.inter ))
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mach.main     1 5 249.4794 258.0473 -119.73970
mach.inter     2 6 193.8715 204.1529  -90.93575 1 vs 2 57.60791 <.0001
```

The p-value was calculated based on the chi-squared distribution with 1 degree of freedom. *It is conservative.*



Example: orthodont data continued

Full model (OrthF2):

$$y_{ij} = \mu + \alpha_i + (\beta + \gamma_i)x_j + \epsilon_{ij}$$

- y_{ij} : distance at age j of subject i
- x_j : age at level j
- α_i : main effect of Subject
- β : average slope for all girls
- γ_i : departure of the slope for Subject i from the average slope
-

$$(\alpha_i, \gamma_i) \stackrel{iid}{\sim} N(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix})$$

- ϵ_{ij} are random errors. $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$
- all random effects and random errors are mutually independent



Example: orthodont data continued

Suppose we are interested in testing the hypothesis that there is no interaction between Subject and age,

$$H_0 : \sigma^2_{\alpha\beta} = 0 \quad \text{vs} \quad H_1 : \sigma^2_{\alpha\beta} > 0$$

Reduced model under H_0 (random intercepts only):

$$y_{ij} = \mu + \alpha_i + \beta x_j + \epsilon_{ij}$$

Note:

- $\sigma^2_{\alpha\beta} = 0$ implies $\sigma_{\alpha\beta} = 0$
- Reduced model (OrthF1) is nested within the full model.
Thus LRT may be used

oooooooooooooooooooooooooooo

Example: orthodont data continued

```
> OrthoF <- Orthodont[Orthodont$Sex=="Female",]  
> OrthF1 <- lme(distance ~ age, data=OrthoF,  
                     random=~1 | Subject)  
> OrthF2 <- lme(distance ~ age, data=OrthoF,  
                     random=~age | Subject)  
> (LR.app <- anova( OrthF1, OrthF2 ))  
      Model df      AIC      BIC    logLik     Test L.Ratio p-value  
OrthF1     1  4 149.2183 156.1690 -70.60916  
OrthF2     2  6 149.4287 159.8547 -68.71435 1 vs 2 3.789622  0.1503
```

The p-value was calculated based on the chi-squared distribution with 2 degrees of freedom. *It is conservative.*



Violation of a regularity condition

- One of the key assumptions for the LRT to be asymptotically chi-square is that parameters under the null hypothesis is an interior point of the parameter space Θ
- *Situation A.* For the machine example, $\theta = (\sigma_a^2, \sigma_{ab}^2)^T$, $\Theta = [0, \infty) \times [0, \infty)$. Under the null hypothesis $H_0 : \sigma_{ab}^2 = 0$, $\theta = (\sigma_a^2, 0)^T$ is on the boundary of Θ
- *Situation B.* For the Orthodont example, $\theta = (\sigma_1^2, \sigma_2^2, \sigma_{12})^T$, $\Theta = [0, \infty) \times [0, \infty) \times (-\infty, \infty)$. Under the null hypothesis $H_0 : \sigma_2^2 = 0$, $\theta = (\sigma_1^2, 0, 0)^T$ is again on the boundary of Θ
- Such problem is common for hypotheses on variance components. Variances are constraint to be non-negative. Thus they are on the boundary when we want to test the hypothesis that some of them equal zero



Null distribution of the LRT

- Liang and Self (JASA, 1987) showed that when the true parameters are on the boundary of the parameter space, the asymptotic distribution of LRT is a mixture of chi-squared distributions
- Stram and Lee (Biometrics, 1994) applied this general theory to test hypotheses on variance components. They have shown that
 - *Situation A*, the asymptotic distribution of the LRT is a 50:50 mixture of χ_0^2 and χ_1^2
 - *Situation B*, the asymptotic distribution of the LRT is a 50:50 mixture of χ_1^2 and χ_2^2
- A $p : (1 - p)$ mixture of two distributions F and G is the CDF $pF + (1 - p)G$. It is easy to compute critical values from mixture of chi-squared distributions. The form of the mixture depends on parameter space geometry, which is not simple in general



Evaluate approximations

The `simulate.lme` function in `nlme` is a very useful tool to assess the accuracy of these chi-square approximations.

```
> help(simulate.lme)
Description
```

The model object is fit to the data. Using the fitted values of the parameters, `nsim` new data vectors from this model are simulated. Both `m1` and `m2` are fit by maximum likelihood (ML) and/or by restricted maximum likelihood (REML) to each of the simulated data vectors.

Usage

```
simulate(object, nsim, seed, m2, method, niterEM, useGen, ...)
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: machine data continued

Simulation and PP/QQ plots

```
> machLRTsim <- simulate.lme(mach.main, mach.inter, seed=3473, nsim=1000)

# PP plot
plot(machLRTsim, df=c(0,1), layout=c(4,1), between=list(x=c(0,.5)))

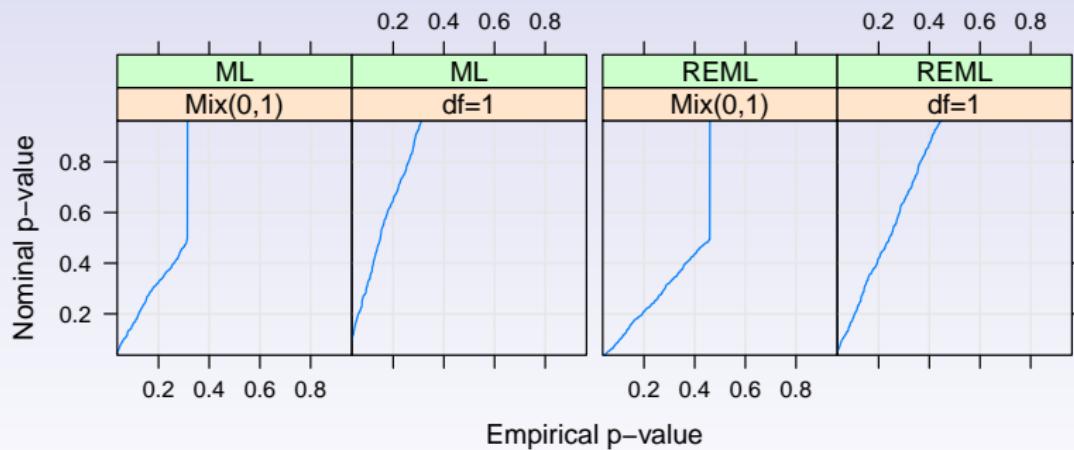
# QQ plot
library(assist)
okML <- machLRTsim>null$ML[,"info"]<8&machLRTsim$alt$ML[,"info"]<8
okREML <- machLRTsim>null$REML[,"info"]<8&machLRTsim$alt$REML[,"info"]<8
MLx <- sort(2*pmax(0,machLRTsim$alt$ML[okML,"logLik"]-machLRTsim>null$ML[okML,"logLik"]))
REMLx <- sort(2*pmax(0,machLRTsim$alt$REML[okML,"logLik"]-machLRTsim>null$REML[okML,"logLik"])

frm1 <- data.frame(x=c(MLx, MLx, REMLx, REMLx),
                     y=c(qchisq(ppoints(MLx),1), qchisq(2*ppoints(MLx)-1,1),
                          qchisq(ppoints(REMLx),1), qchisq(2*ppoints(REMLx)-1,1)),
                     df=rep(rep(c("df=1","mix(0,1)"), rep(length(MLx),2)), 2),
                     method=rep(c("ML","REML"),rep(2*length(MLx),2)))

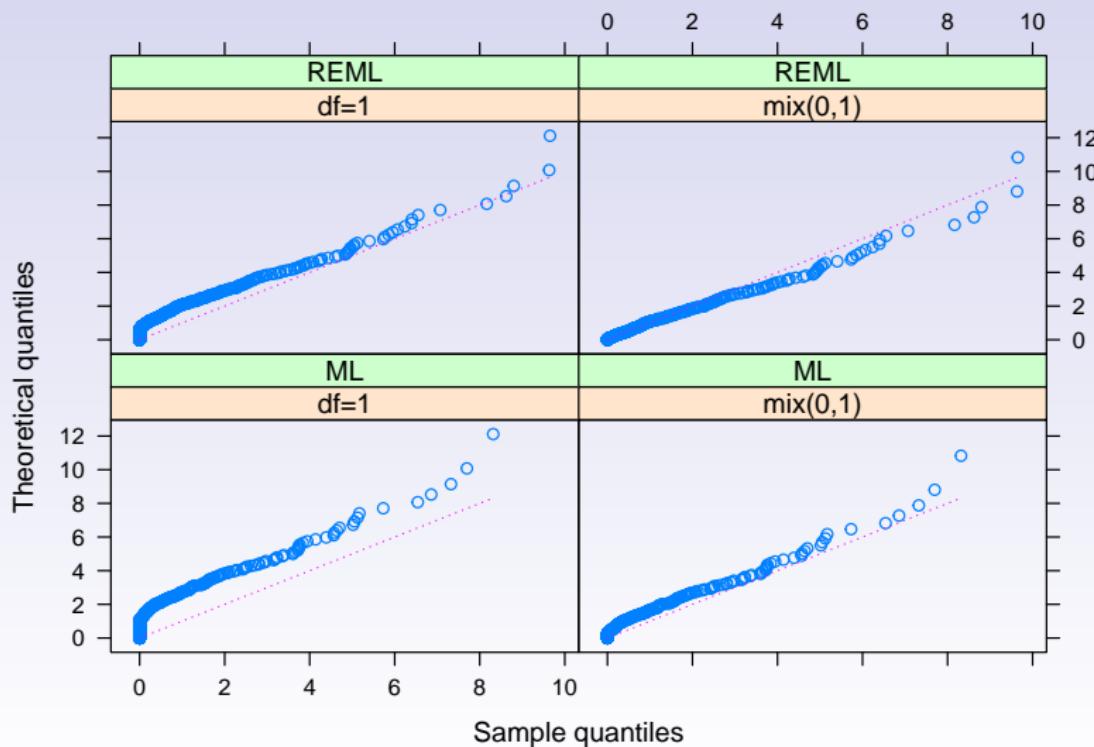
frm2 <- data.frame(x=c(MLx, MLx, REMLx, REMLx),
                     y=c(MLx, MLx, REMLx, REMLx),
                     df=rep(rep(c("df=1","mix(0,1)"), rep(length(MLx),2)), 2),
                     method=rep(c("ML","REML"),rep(2*length(MLx),2)))

xyplot2(y~x|df*method, data=list(frm1, frm2), type=c("p","l"),
        xlab="Sample quantiles", ylab="Theoretical quantiles")
```

Probability-probability plot



QQ plot





How to read the plots?

Look at small p values:

- conservative if above the diagonal (dotted) line in the PP (QQ) plot
- anti-conservative if below
- good approximation if close to the diagonal (dotted) line

For the machine example, the mixture approximation is accurate for the REML, but is not accurate for the ML (conservative). Using χ_1^2 would be too conservative.

Example: orthodont data continued

Simulation and PP/QQ plots

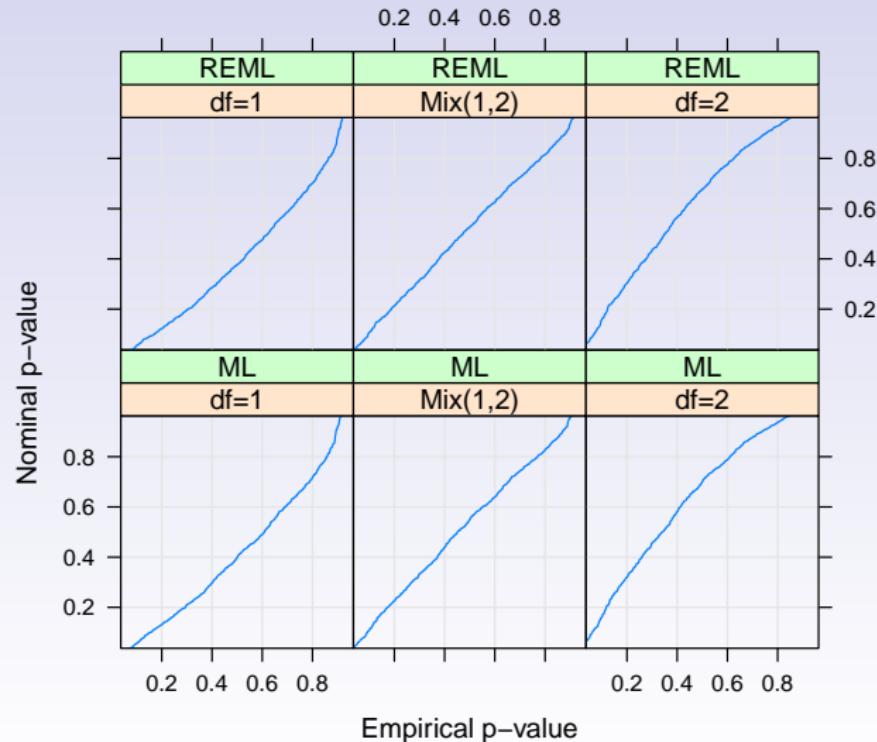
```
> OrthoF <- Orthodont[Orthodont$Sex=="Female",]
> OrthF1 <- lme(distance ~ age, data=OrthoF, random=~1 | Subject)
> OrthF2 <- lme(distance ~ age, data=OrthoF, random=~age | Subject)
> orthLRTsim <- simulate.lme(OrthF1, OrthF2, nsim=1000, seed=2394)

# PP plot
> plot(machLRTsim, df=c(1,2))

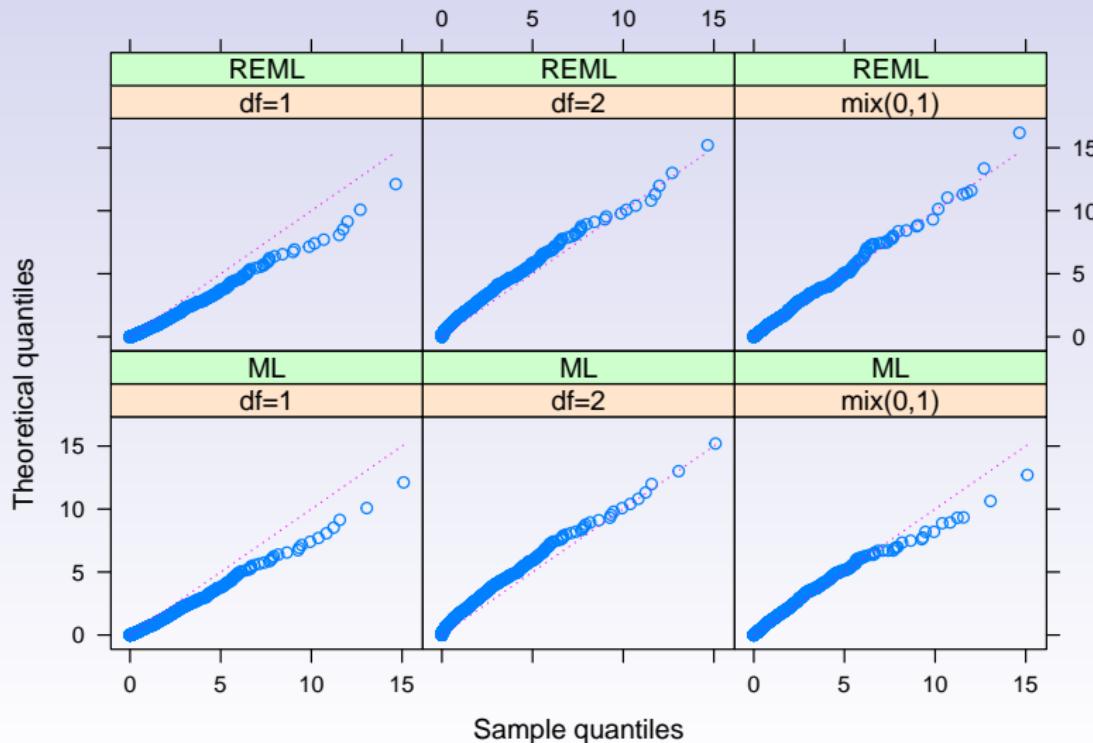
# QQ plot
okML <- orthLRTsim>null$ML[, "info"] < 8 & orthLRTsim$alt$ML[, "info"] < 8
okREML <- orthLRTsim>null$REML[, "info"] < 8 & orthLRTsim$alt$REML[, "info"] < 8
MLx <- sort(2*pmax(0,orthLRTsim$alt$ML(okML,"logLik")-
    orthLRTsim>null$ML(okML,"logLik")))
REMLx <- sort(2*pmax(0,orthLRTsim$alt$REML(okML,"logLik")-
    orthLRTsim>null$REML(okML,"logLik")))
frm1 <- data.frame(x=c(MLx, MLx, MLx, REMLx, REMLx, REMLx),
y=c(qchisq(ppoints(MLx),1), qchisq(ppoints(MLx),2),
    sort(c(rchisq(500,df=1),rchisq(500,df=2))),
    qchisq(ppoints(REMLx),1), qchisq(ppoints(REMLx),2),
    sort(c(rchisq(500,df=1),rchisq(500,df=2)))),
df=rep(rep(c(''df=1'', ''df=2'', ''mix(0,1)''), rep(length(MLx),3)),2),
method=rep(c(''ML'', ''REML''),rep(3*length(MLx),2)))
frm2 <- data.frame(x=c(MLx, MLx, MLx, REMLx, REMLx, REMLx),
y=c(MLx, MLx, MLx, REMLx, REMLx, REMLx),
df=rep(rep(c(''df=1'', ''df=2'', ''mix(0,1)''), rep(length(MLx),3)), 2),
method=rep(c(''ML'', ''REML''),rep(3*length(MLx),2)))

xyplot2(y~x|df*method, data=list(frm1, frm2), type=c("p","l"),
xlab="Sample quantiles", ylab="Theoretical quantiles")
```

PP plot



QQ plot





Results based on χ^2 distribution

- For the machine example, χ_1^2 is conservative and the mixture is not very accurate for ML
- For the orthodont example, χ_1^2 is anti-conservative, χ_2^2 is conservative, the mixture is quite accurate
- In general, it is difficult to come up with a rule works for all situations. The `anova` function in `nlme` uses the conservative χ^2 distribution

oooooooooooooooooooooooooooo

Bootstrap p-value

Inside `simulate.lme`, samples generated under H_0 may be regarded as bootstrap samples. Therefore, we may compute the bootstrap p-value as follows:

```
# for machine example  
> print(mean(sort(REMLx) > LR.app$L.Ratio[2]))  
[1] 0
```

```
# for orthodont example  
> print(mean(sort(REMLx) > LR.app$L.Ratio[2]))  
[1] 0.096
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Test using lmer

```
> mach.main.lmer <- lmer(score ~ Machine + (1 | Worker),
+                           data = MachinesUnbal)
> mach.inter.lmer <- lmer(score ~ Machine + (1 | Worker) +
+                            (1 | Worker:Machine), data = MachinesUnbal)
> anova(mach.main.lmer, mach.inter.lmer)
refitting model(s) with ML (instead of REML)
Data: MachinesUnbal
Models:
mach.main.lmer: score ~ Machine + (1 | Worker)
mach.inter.lmer: score ~ Machine + (1 | Worker) + (1 | Worker:Machine)
                npar    AIC    BIC   logLik deviance Chisq Df
mach.main.lmer     5 257.07 265.99 -123.536   247.07
mach.inter.lmer     6 203.51 214.21  -95.753   191.51 55.566  1
                Pr(>Chisq)
mach.main.lmer
mach.inter.lmer  9.035e-14 ***
---
---
```

The p-value was computed using a χ^2_1 distribution which is conservative. Note that the models were refitted using ML. Other R packages may exist for better approximations.



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- **Inference for fixed and random effects**
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Inference for fixed and random effects

lme for fixed effects, use the intervals function:

```
intervals(fittedobj, which="fixed")
```

lme does not provide confidence intervals for random effects

lmer use the confint function:

```
confint(fittedobj, method =
```

```
c("profile", "Wald", "boot"))
```

For random effects, use the ranef function:

```
ranef(fittedobj, condVar=T)
```

proc mixed use option CL in the model and random statements for fixed and random effects respectively



Hypothesis tests

- LRT may also be used. But it tends to be “anti-conservative”, thus is *not* recommended
- Instead, the conditional t-tests or F-tests are recommended
- Conditional tests are in the following sense: when θ (ie W matrix) is known, $\hat{\beta}$ are WLS estimates, thus standard t-tests and F-tests can be constructed. Now replace θ by $\hat{\theta}$ and pretend it was known. Thus the tests are *conditional* on the estimates of θ . Note the uncertainty in $\hat{\theta}$ is ignored in such conditional tests
- Figuring out the degrees of freedom for a t-test and the denominator degrees of freedom for an F-test is quite tricky and cumbersome. Fortunately, the program does this for us, if the model is specified correctly



Degrees of freedom

- The degrees of freedom depends on the factor being tested is *inner* or *outer* to the grouping factor
- A term is *inner* relative to a grouping factor if its value can change within a group level of the grouping factor
- A term is *outer* to a grouping factor if its value does not change within a group level of the grouping factor
- A term is estimated at the level i if it is inner to the $(i - 1)$ st grouping factor and outer to the i th grouping factor
- Suppose that there are Q levels of grouping. And denote level 0 as the group of all responses, and level $Q + 1$ as the group of individual responses



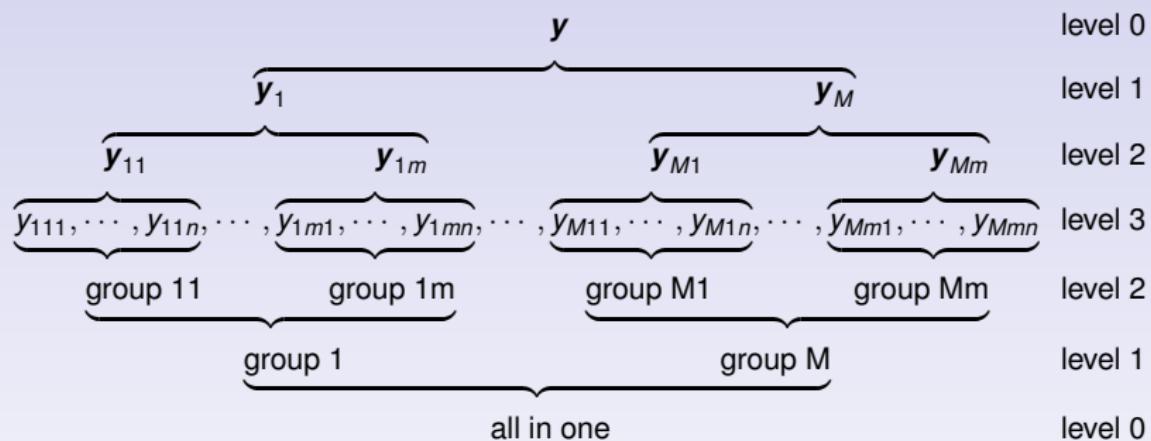
Degrees of freedom

- m_i : the total number of groups in level i . Convention: $m_0 = 1$ when the fixed effects model includes an intercept and $m_0 = 0$ otherwise. $m_{Q+1} = n$
- p_i : the sum of the degrees of freedom corresponding to terms estimated at level i
- The i th level denominator degrees of freedom is defined as

$$\text{denDF}_i = m_i - (m_{i-1} + p_i)$$

- The intercept is treated differently. As a parameter, it is estimated at level 0 since it is outer to all the grouping factors. However, its denominator degrees of freedom are calculated as if it is estimated at level $Q + 1$ since it pools information from all observations

Example: machine data continued



We have $Q = 2$ levels of grouping. Level 1: Worker. Level 2: Machine within Worker (Worker/Machine). We have a balanced design with $n_{ij} = 3$, $m_i = 3$. Suppose that we are interested in testing the fixed factor Machine. It is easy to see that Machine is inner to Worker (level 1) but is outer to Worker/Machine (level 2). Thus it is estimated at level 2.

Example: machine data continued

Level	number of groups (m_i)	sum of df (p_i)	denDF _i
3	54	0	54-18-0=36
2	18	2	18-6-2=10
1	6	0	6-1-0=5
0	1	1	NA

Since Machine effects are estimated at level 2, thus the denominator degrees of freedom is 10. Note this is the same as the F-test using MSAB as the denominator with
 $df = (a - 1)(b - 1) = 2 \times 5 = 10$.

```
> anova(mach2)
      numDF denDF   F-value p-value
(Intercept)     1      36  773.5708 <.0001
Machine         2      10   20.5761  3e-04
```



Example: wheat data continued

We have $Q = 2$ levels of grouping. Level 1: block. Level 2: fertility within block (block/fertility)

Suppose that we are interested in testing the fixed factors fertility, variety and fertility*variety. It is easy to see that variety and fertility*variety are inner to block/fertility. Thus they are estimated at level 3. fertility is inner to block but is outer to block/fertility. Thus it is estimated at level 2.

Level	number of groups (m_i)	sum of df (p_i)	denDF _i
3	16	1+3=4	16-8-4=4
2	8	3	8-2-3=3
1	2	0	2-1-0=1
0	1	1	NA

oooooooooooooooooooooooooooo

Example: wheat data continued

```
> anova(fit1)
      numDF denDF   F-value p-value
(Intercept)      1      4 197.46998 0.0001
fertility        3      3  5.80152 0.0914
variety          1      4  1.06762 0.3599
fertility:variety 3      4  0.24516 0.8612
```



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- **Model selection**

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Model selection

For comparing non-nested models, we may use

$$\text{AIC} = -2 \times \text{maximum log likelihood} + 2 \times \text{number of parameters}$$

$$\text{BIC} = -2 \times \text{maximum log likelihood} + \log(n) \times \text{number of parameters}$$

Models with the lowest AIC and/or BIC are preferred. Likelihood may be replaced by the restricted likelihood.

```
> anova(mach.main.lmer, mach.inter.lmer)
refitting model(s) with ML (instead of REML)
Data: MachinesUnbal
Models:
  mach.main.lmer: score ~ Machine + (1 | Worker)
  mach.inter.lmer: score ~ Machine + (1 | Worker) + (1 | Worker:Machine)
      npar     AIC     BIC   logLik deviance Chisq Df
mach.main.lmer     5 257.07 265.99 -123.536   247.07
mach.inter.lmer    6 203.51 214.21  -95.753   191.51 55.566  1
Pr(>Chisq)
mach.main.lmer
mach.inter.lmer  9.035e-14 ***
---
```



Linear mixed effects model

Recall LME:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \quad \mathbf{b} \sim N(\mathbf{0}, \sigma^2 \mathbf{G}), \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{R})$$

Thus

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{W}^{-1})$$

where

$$\mathbf{W}^{-1} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$$



Mean and covariance structures

- The *mean* structure is modeled by $X\beta$. Methods for LM can be used to build this part of the model
- The covariance structure consists two components: contribution from random effects, ZGZ^T , and contribution from within-group random errors, R
- These two components may compete to each other: similar covariance may be introduced from either component. Care must be exercised to prevent nonidentifiability
- We now learn how to model each of these two components



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Modeling random effects

We use single level of grouping for illustration. Methods can be generalized to multiple levels of grouping. Recall the vector form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M$$

where

$$\mathbf{b}_i \stackrel{iid}{\sim} N(\mathbf{0}, \sigma^2 D)$$

Let $D = \{d_{jv}\}$ where $j = 1, \dots, k$ and $v = 1, \dots, k$. Note that $d_{vv} = 0$ implies that $d_{vj} = d_{jv} = 0$ for all j , that is, the v th raw and v th column are zero. Thus $d_{vv} = 0$ implies that the v th covariate z_v for the random effects is not needed. Therefore, the key to modeling the random effects is to explore the covariance matrix D .



Explore structure in D

- If we observe \mathbf{b}_i , then a common estimate of $\sigma^2 D$ is its sample covariance matrix

$$\sum_{i=1}^M (\mathbf{b}_i - \bar{\mathbf{b}}_i)^T (\mathbf{b}_i - \bar{\mathbf{b}}_i) / M$$

- Of course \mathbf{b}_i are unobservable. Thus we need some preliminary estimates. One way to get these estimates is to fit the first stage model for each group
- For each fixed group i , fit

$$\mathbf{y}_i = X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \boldsymbol{\epsilon}_i$$

with \mathbf{b}_i as fixed parameters. The estimates $\tilde{\mathbf{b}}_i$ may be regarded as M realizations from $N(\mathbf{0}, \sigma^2 D)$. Then an estimate of $\sigma^2 D$ is

$$\sum_{i=1}^M (\tilde{\mathbf{b}}_i - \bar{\tilde{\mathbf{b}}}_i)^T (\tilde{\mathbf{b}}_i - \bar{\tilde{\mathbf{b}}}_i) / M$$

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

lmList

Tools in R: lmList and related plots are helpful.

```
> help(lmList)
```

List of lm Objects with a Common Model

Description:

'Data' is partitioned according to the levels of the grouping factor 'g' and individual 'lm' fits are obtained for each 'data' partition, using the model defined in 'object'.

Usage:

```
lmList(object, data, level, subset, na.action, pool)
```



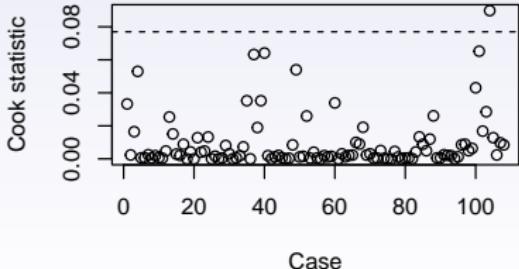
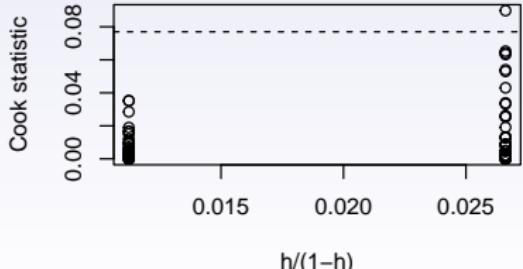
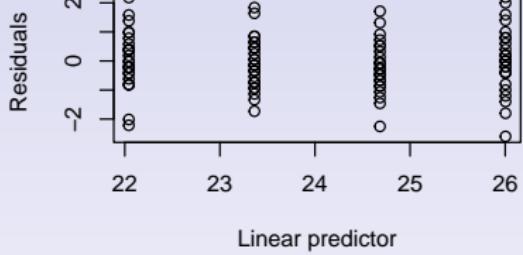
Example: orthodont data continued

For illustration we first fit a simple linear regression model for all subjects:

```
> library(boot)
> fm1Orth.lm <- glm(Orthodont$distance ~ Orthodont$age )
> glm.diag.plots( fm1Orth.lm )
```

- Plots are different from the book
- Large unexplained variability in residuals
- Next we add the Sex variable

Diagnostic plots for the simple linear regression fit



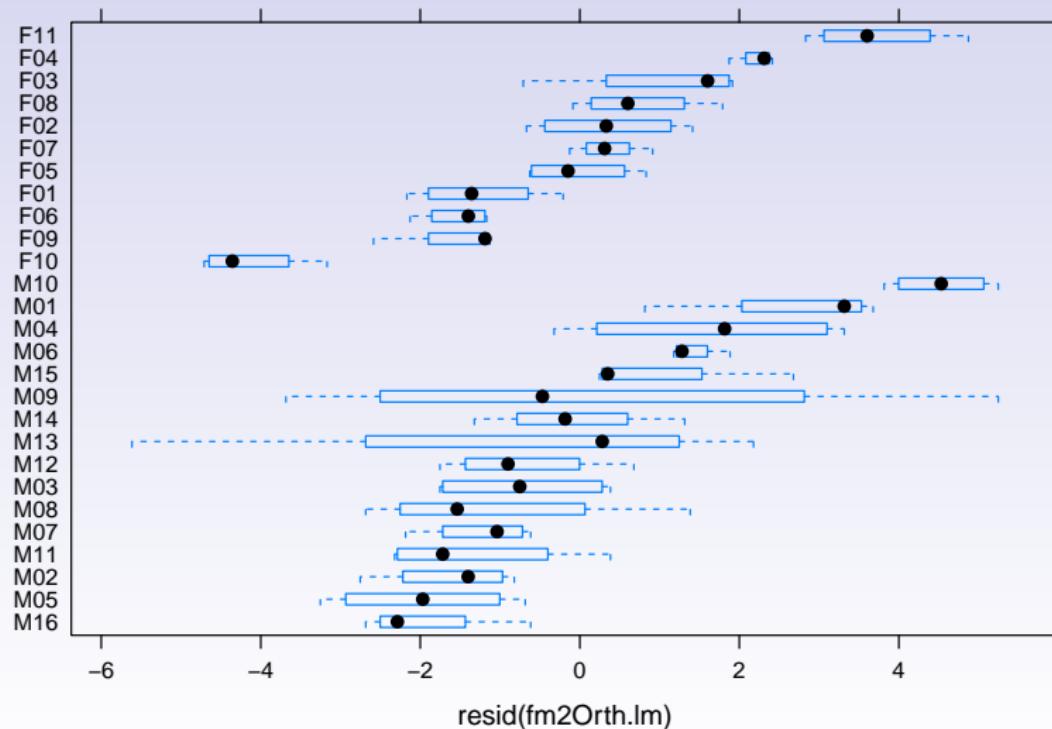
oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Add Sex variable to fixed effects

```
> fm1Orth.lm <- lm(distance ~ age, Orthodont )
> fm2Orth.lm <- update ( fm1Orth.lm,
                           formula = distance ~ Sex*age)
> anova(fm1Orth.lm,fm2Orth.lm)
Analysis of Variance Table
Model 1: distance ~ age
Model 2: distance ~ Sex + age + Sex:age
Res.Df   RSS Df Sum of Sq    F    Pr(>F)
1     106 682.34
2     104 529.76  2      152.58 14.977 1.923e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# bwplot is one function in lattice package
> library(lattice)
> bwplot( getGroups(Orthodont)~resid(fm2Orth.lm) )
```

Residual plots by subject





Subject effects

- Sex*age are significant
- Residuals depend on subject: some are positive (above average growth) and some are negative (below average growth)
- In other words, residuals (observations) from the same subject are correlated
- We need to model subject effects. Since subjects represent a random sample from a population, Subject is random factor. Thus all terms involve Subject are random effects
- Which terms involve Subject should be included?



Subject effects

- To be specific, assume the relationship between growth and age can be well approximated by a linear model. Then we need to decide if the intercept and slope depend on Subject. That is, do we need subject-specific random intercepts and slopes?
- To answer this question, we need to know the variations in intercepts and slopes among subjects. A small variation indicates that we don't need to include the random effects
- Are intercepts and slopes correlated?
- To address these questions, we fit a line to each subject. That is, to build the first stage model. Then explore variation and correlation in the estimated intercepts and slopes

oooooooooooooooooooooooooooo

Fit a line to each subject

```
> detach("package:lme4", unload=TRUE)
> library(nlme)
> fm1Orth.lis <- lmList ( distance ~ age | Subject, Orthodont )
> fm1Orth.lis
Call:
  Model: distance ~ age | Subject
  Data: Orthodont

Coefficients:
      (Intercept)    age
M16          16.95  0.550
M05          13.65  0.850
M02          14.85  0.775
M11          20.05  0.325
M07          14.95  0.800
M08          19.75  0.375
M03          16.00  0.750
M12          13.25  1.000
M13          2.80   1.950
M14          19.10  0.525
M09          14.40  0.975
M15          13.50  1.125
M06          18.95  0.675
M04          24.70  0.175
M01          17.30  0.950
M10          21.25  0.750
F10          13.55  0.450
F09          18.10  0.275
F06          17.00  0.375
F01          17.25  0.375
```

```
oooooooooooooooooooooooooooo
```

Fit a line to each subject

```
F05      19.60 0.275  
F07      16.95 0.550  
F02      14.20 0.800  
F08      21.45 0.175  
F03      14.40 0.850  
F04      19.65 0.475  
F11      18.95 0.675
```

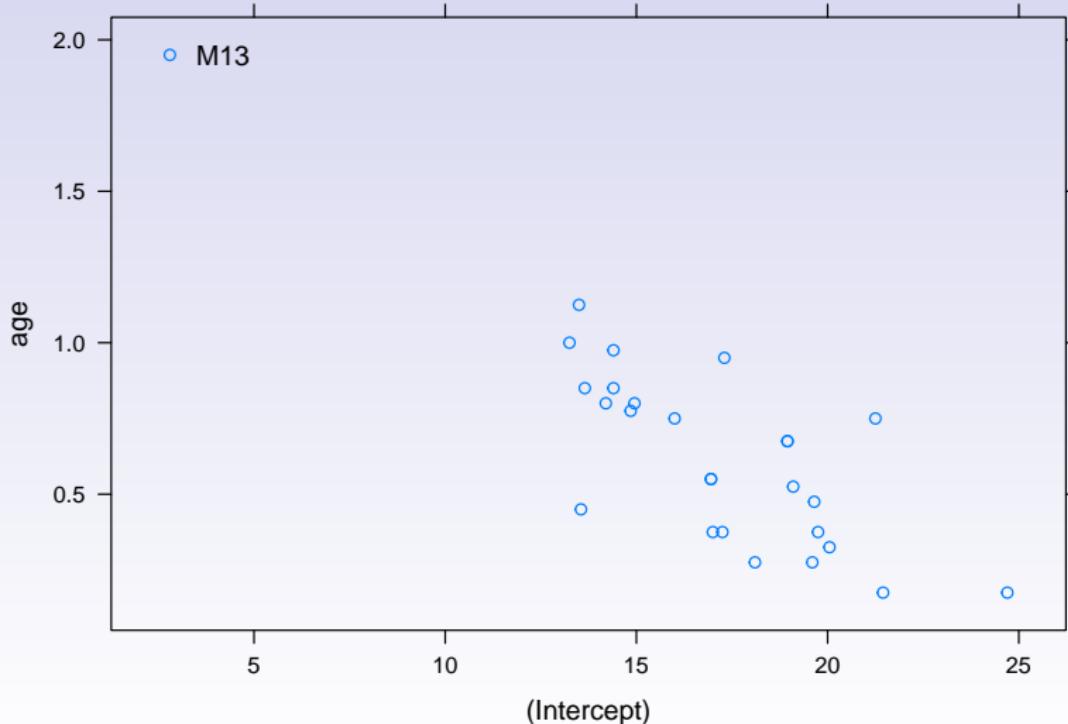
Degrees of freedom: 108 total; 54 residual

Residual standard error: 1.310040

```
# id to identify outliers. adj for location of the id label.  
> pairs( fm1Orth.lis, id=.01, adj=-.5)
```

- high negative correlation between intercepts and slopes
- this is because that observations were made from age 8-14
- center `age`: create a new variable=`age-11`

Plot intercepts vs slopes



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit a line to each subject with centered age

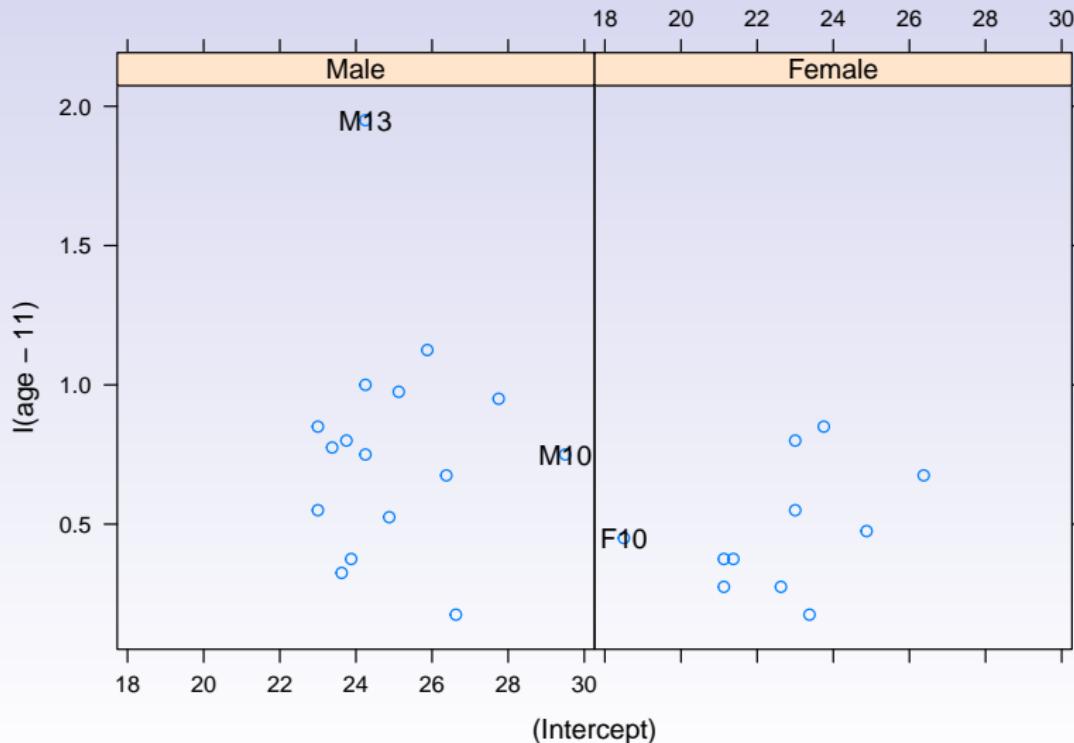
```
> fm2Orth.lis <- update (fm1Orth.lis, distance ~ I(age-11))

> pairs(fm2Orth.lis, ~coef(.) | Sex, id = 0.1)

> cor(coef(fm2Orth.lis))
      (Intercept) I(age - 11)
(Intercept)    1.0000000  0.2938604
I(age - 11)    0.2938604  1.0000000

# plot confidence intervals for intercepts and slopes
> plot(intervals( fm2Orth.lis ))
```

Plot intercepts vs slopes



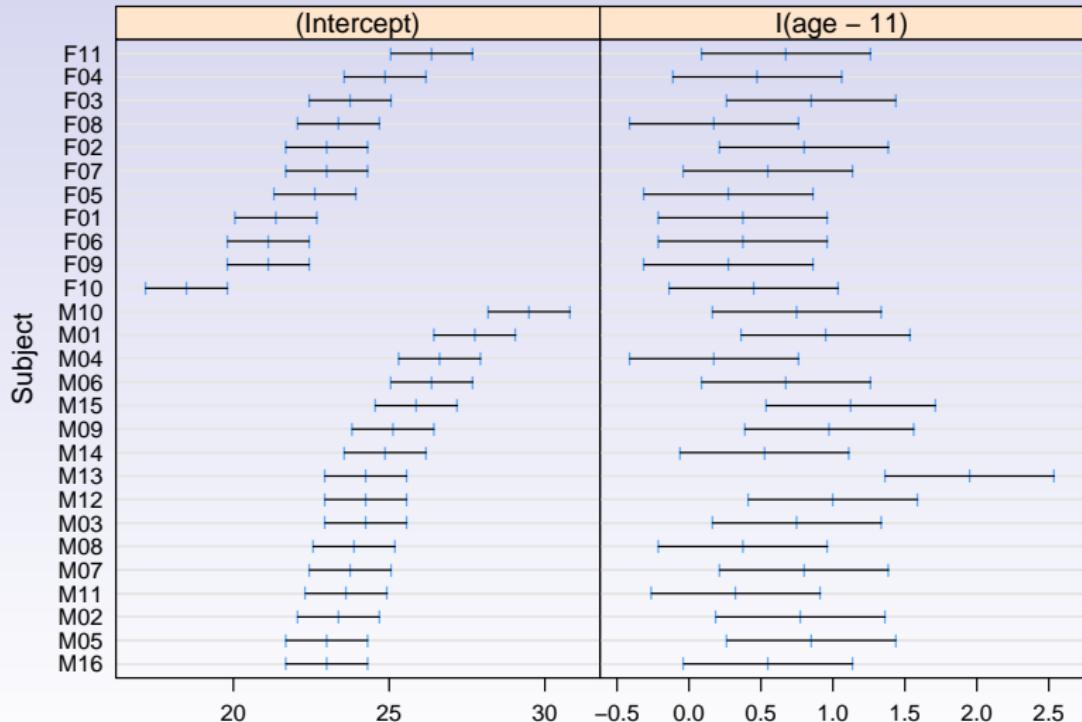


Comments

From plot of coefficients:

- no evidence of correlation
- large between subjects variations for both intercepts and slopes
- F10 has unusual small distance at age 11, even though her growth rate is average. M10 has unusual large distance at age 11, even though his growth rate is average. M13 has unusual growth rate, even though his distance at age 11 is average
- intercepts tend to be lower for girls
- excluding identified extremes, variations seems similar for boys and girls

Plot confidence intervals for intercepts and slopes





Comments

From plot of confidence intervals:

- between subject variation in the intercept is large
- except for subject M13, all confidence intervals for the slope overlap, so random slope effect may not needed. We will test this formally later

Specify D in lme

lme uses pdMat classes to specify *patterned variance-covariance matrices*, i.e. the D matrix.

pdBlocked	block-diagonal
pdCompSymm	compound-symmetry structure
pdDiag	diagonal
pdIdent	multiple of an identity
pdSymm (<i>default</i>)	unstructured positive-definite matrix

Table: Standard pdMat classes.

More detail can be found using

```
> help(pdClasses)
```



Special cases of D

block-diagonal (`pdBlocked`)

$$D = \begin{pmatrix} D_1 & & & \\ & \ddots & & \\ & & D_d & \end{pmatrix}$$

compound-symmetry (CS) structure (`pdCompSymm`)

$$D = \sigma_1^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \rho & 1 \end{pmatrix}$$



Special cases of D

diagonal (`pdDiag`)

$$D = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_q^2 \end{pmatrix}$$

identity (`pdIdent`)

$$D = \sigma_1^2 I$$

unstructured (`pdSymm`): any positive-definite matrix

Some useful utility functions for lme

`anova` compare likelihoods of fitted objects

`augPred` obtain predicted values at the specified values of the primary variable

`compareFits` compare fitted objects. Note that no extension ".lme" is needed

`comparePred` compare predictions. Note that no extension ".lme" is needed

`coef` extract lme coefficients

`fitted` extract lme fitted values

`intervals` provide confidence intervals on parameters

`pairs` diagnostic pairs plots

`plot` diagnostic plots

`predict` obtain predictions

Some useful utility functions for lme

`qqnorm` diagnostic plots for assessing the normality of residuals and random effects

`ranef` extract random effects

`residual` extract residuals

`summary` provide a summary of the fit

`Variogram` compute semivariogram for residuals from an `lme` object

More detail can be found using

```
> help("function-name".lme)
```

oooooooooooooooooooooooooooo

Example: orthodont data continued

(a) Random intercept and slopes for all children. Unstructured covariance matrix for random effect (default)

```
> fm1Orth.lme <- lme( distance ~ I(age-11), data=Orthodont,
  random = ~ I(age-11) | Subject )
> summary(fm1Orth.lme)
Linear mixed-effects model fit by REML
Data: Orthodont
      AIC      BIC      logLik
 454.6367 470.6173 -221.3183
```

oooooooooooooooooooooooooooo

Example: orthodont data continued

Random effects:

Formula: ~I(age - 11) | Subject

Structure: General positive-definite, Log-Cholesky parametrization

StdDev Corr

(Intercept) 2.1343295 (Intr)

I(age - 11) 0.2264293 0.503

Residual 1.3100382

Fixed effects: distance ~ I(age - 11)

	Value	Std.Error	DF	t-value	p-value
--	-------	-----------	----	---------	---------

(Intercept)	24.023148	0.4296602	80	55.91197	<.0001
-------------	-----------	-----------	----	----------	--------

I(age - 11)	0.660185	0.0712534	80	9.26532	<.0001
-------------	----------	-----------	----	---------	--------

Correlation:

(Intr)

I(age - 11) 0.294

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-----	----	-----	----	-----

-3.223089596	-0.493761902	0.007316477	0.472149144	3.916037455
--------------	--------------	-------------	-------------	-------------

Number of Observations: 108

Number of Groups: 27

oooooooooooooooooooooooooooo

Example: orthodont data continued

Get estimates of random effects

```
> ranef(fm1Orth.lme)
  (Intercept) I(age - 11)
M16   -0.9451481 -0.068853701
M05   -0.8950650  0.025602098
...
...
```

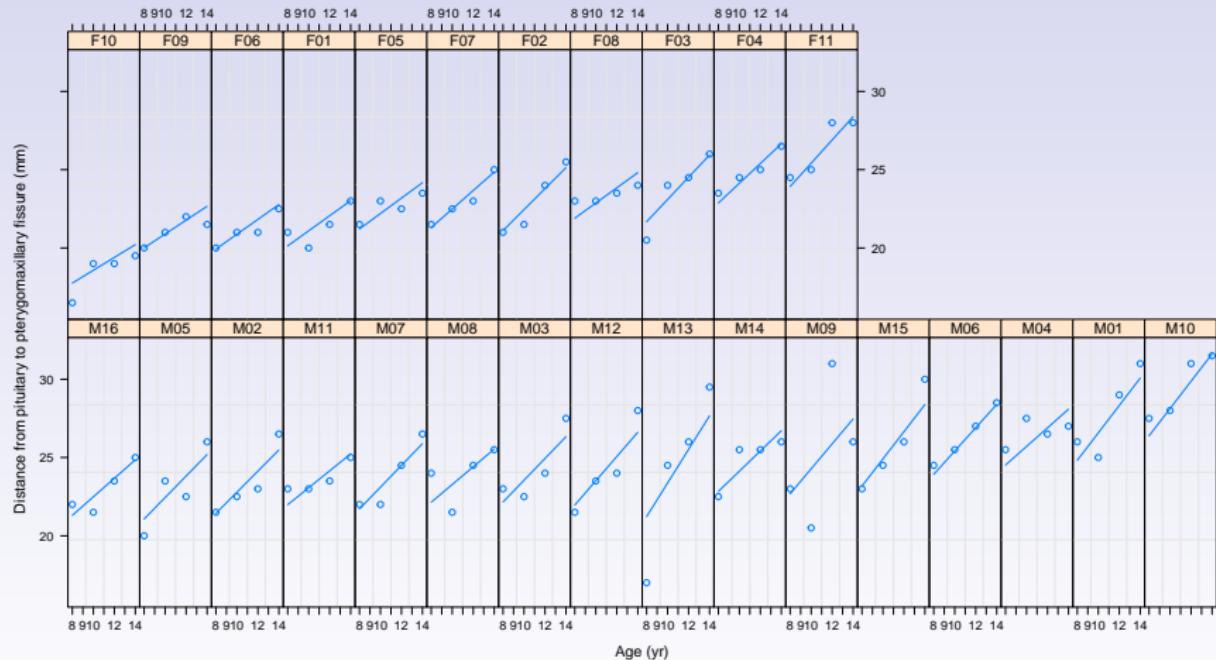
Get estimated intercepts and slopes (fixed+random)

```
> coef(fm1Orth.lme)
  (Intercept) I(age - 11)
M16      23.07800  0.5913315
M05      23.12808  0.6857873
...
...
```

Plot data and fits

```
> plot( augPred(fm1Orth.lme), aspect=''xy'',  
       layout=c(16,2))
```

Plot of fits



oooooooooooooooooooooooooooo

Example: orthodont data continued

(b) Add Sex variable to the fixed effects

```
> fm2Orth.lme <- update(fm1Orth.lme, fixed = distance ~ Sex*I(age-11))
> summary(fm2Orth.lme)
Linear mixed-effects model fit by REML
Data: Orthodont
      AIC      BIC      logLik
448.5817 469.7368 -216.2908

Random effects:
Formula: ~I(age - 11) | Subject
Structure: General positive-definite, Log-Cholesky parametrization
           StdDev     Corr
(Intercept) 1.8303405 (Intr)
I(age - 11) 0.1802607 0.207
Residual    1.3101001
```

oooooooooooooooooooooooooooo

Example: orthodont data continued

Fixed effects: distance ~ Sex + I(age - 11) + Sex:I(age - 11)

	Value	Std.Error	DF	t-value	p-value
(Intercept)	24.968750	0.4860065	79	51.37534	<.0001
SexFemale	-2.321023	0.7614258	25	-3.04826	0.0054
I(age - 11)	0.784375	0.0859913	79	9.12156	<.0001
SexFemale:I(age - 11)	-0.304830	0.1347225	79	-2.26265	0.0264

Correlation:

	(Intr)	SexFml	I(-11)
SexFemale	-0.638		
I(age - 11)	0.102	-0.065	
SexFemale:I(age - 11)	-0.065	0.102	-0.638

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-3.168695060	-0.386063208	0.006982768	0.445259220	3.849301277

Number of Observations: 108

Number of Groups: 27

Note that the interaction is significant

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Extract estimates and prediction

```
#extract fitted values at both population and subject levels
> fitted(fm2Orth.lme, level=0:1)
      fixed   Subject
1    22.61563 24.84563
2    24.18438 26.57645
...
# extract residuals, subject level
> resid(fm2Orth.lme, level=1)
      M01        M01        M01        M01        M02
1.154371823 -1.576447332  0.692733514  0.961914359  0.225097263
...
# to get standardized residuals
> resid(fm2Orth.lme, level=1, type="pearson")

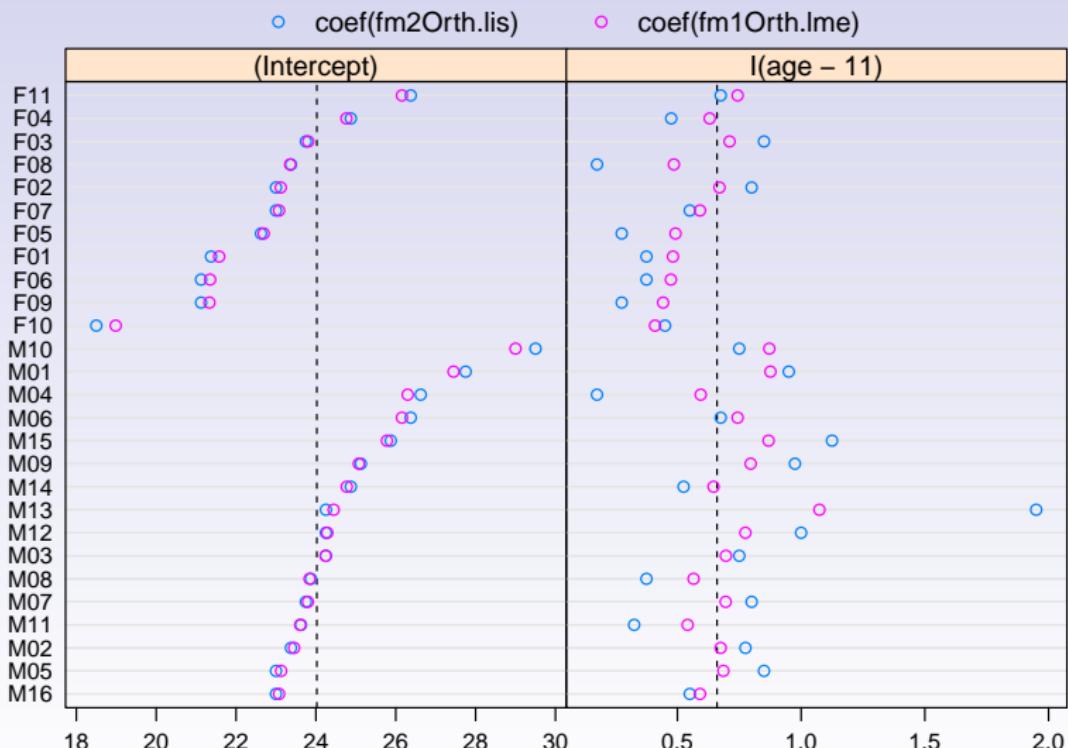
# to predict distance for boy M11 and girl F03 at ages 16, 17 and 18
> newOrth <- data.frame( Subject=rep(c("M11","F03"),c(3,3)),
                           Sex=rep(c("Male","Female"),c(3,3)),
                           age=rep(16:18,2))
> predict(fm2Orth.lme, newdata=newOrth)
      M11        M11        M11        F03        F03        F03
26.96843 27.61236 28.25629 26.61330 27.20635 27.79940
```

oooooooooooooooooooooooooooo

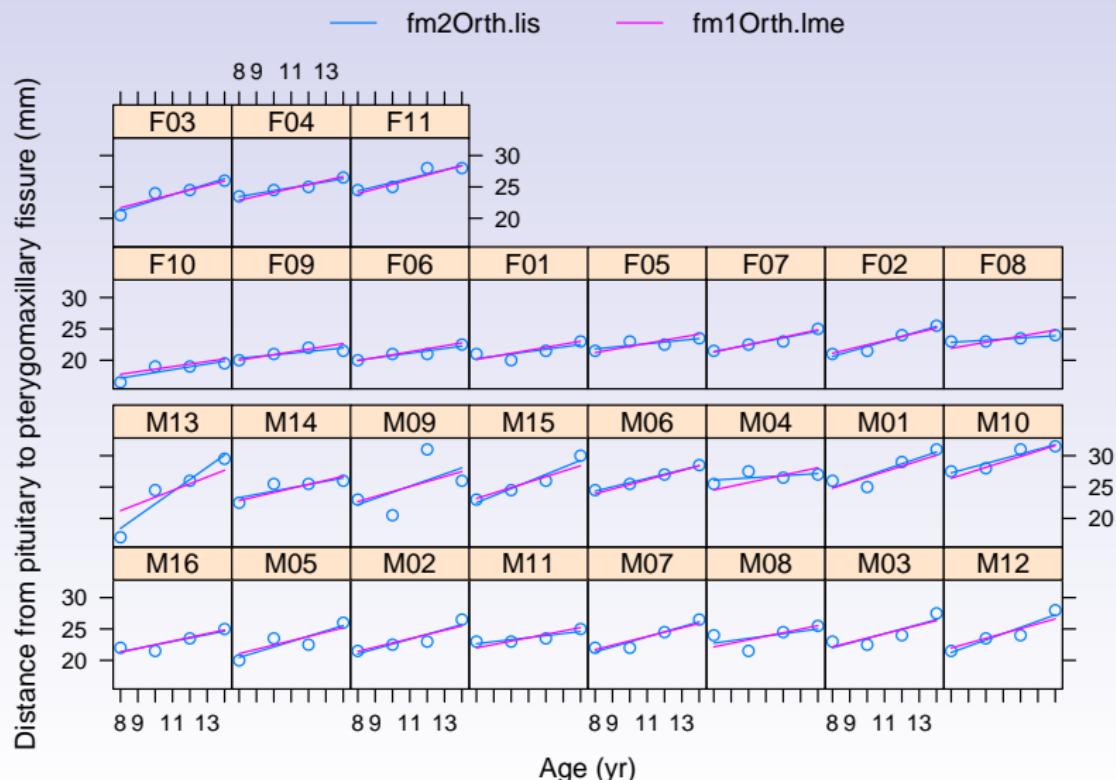
lme vs lmList and REML vs ML

```
> compOrth <- compareFits(coef(fm2Orth.lis), coef(fm1Orth.lme))
> plot( compOrth, mark=fixef(fm1Orth.lme), layout=c(2,1))
> plot( comparePred(fm2Orth.lis,fm1Orth.lme,length.out=2),
       layout=c(8,4), between=list(y=c(0,.5)))
> fm2Orth.lmeM <- update(fm2Orth.lme, method="ML")
> plot( compareFits(ranef(fm2Orth.lme),ranef(fm2Orth.lmeM)),
       mark=c(0,0), layout=c(2,1))
```

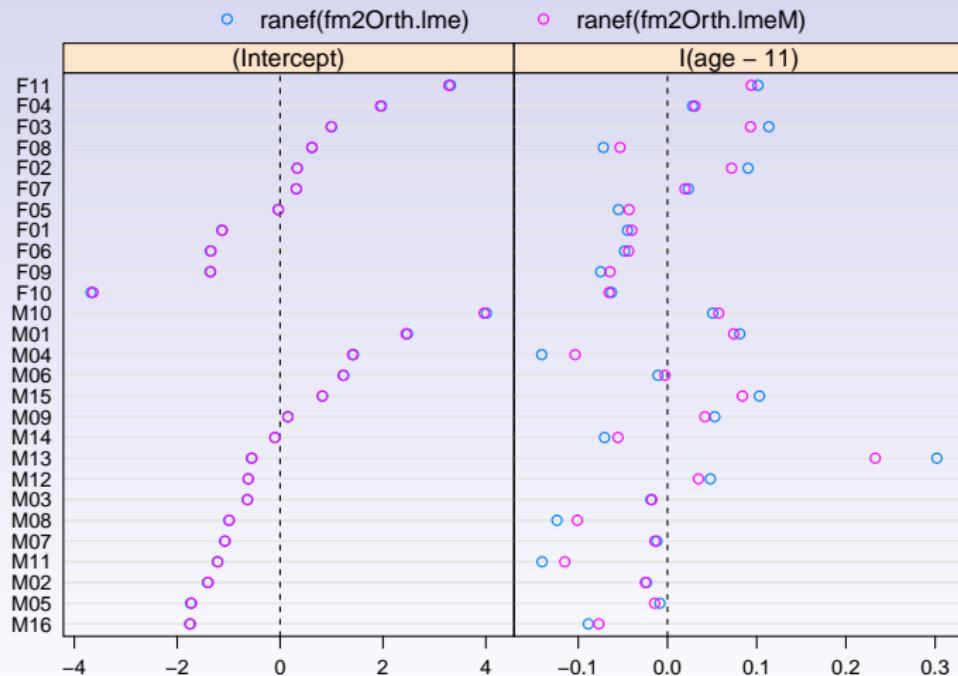
Comparing coefficients from lmList and lme fits



Comparing fits from lmList and lme fits



Comparing estimates of random effects using REML and ML





lme vs lmList and REML vs ML

- LME estimates are pulled toward the fixed effect. That is, they are shrunken toward the population average
- Because they are pulled toward the population average, lme fits are less sensitive to outliers. This is clear from the fits to boy M13
- Usually the REML produces larger estimates for the random-effects variances, which leads to less shrinkage toward zero for random effects

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Correlation between intercept and slope

(c) Is correlation between intercept and slope significant? (no)

Fit with pdDiag:

```
> fm3Orth.lme <- update(fm2Orth.lme,
  random=list(Subject=pdDiag(~ I(age-11))))
# or simply
> fm3Orth.lme <- update(fm2Orth.lme, random=pdDiag(~ I(age-11)))
> summary(fm3Orth.lme)
Linear mixed-effects model fit by REML
Data: Orthodont
      AIC      BIC      logLik
446.8426 465.3533 -216.4213

Random effects:
Formula: ~I(age - 11) | Subject
Structure: Diagonal
            (Intercept) I(age - 11) Residual
StdDev:     1.830330   0.1803439  1.31004
```

oooooooooooooooooooooooooooo

Correlation between intercept and slope

```
Fixed effects: distance ~ Sex + I(age - 11) + Sex:I(age - 11)
                Value Std.Error DF t-value p-value
(Intercept)      24.968750 0.4860015 79 51.37587 <.0001
SexFemale        -2.321023 0.7614181 25 -3.04829 0.0054
I(age - 11)       0.784375 0.0859993 79  9.12071 <.0001
SexFemale:I(age - 11) -0.304830 0.1347350 79 -2.26244 0.0264
```

Correlation:

	(Intr)	SexFml	I(-11)
SexFemale	-0.638		
I(age - 11)	0.000	0.000	
SexFemale:I(age - 11)	0.000	0.000	-0.638

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-3.06659814	-0.39982780	0.02559551	0.43693902	3.85940241

Number of Observations: 108

Number of Groups: 27

```
> anova(fm2Orth.lme, fm3Orth.lme)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
fm2Orth.lme	1	8	448.5817	469.7368	-216.2908			
fm3Orth.lme	2	7	446.8426	465.3533	-216.4213	1 vs 2	0.2609316	0.6095

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Variances of intercept and slope

(d) Do intercept and slope random effects have the same variance? (no)

Fit with pdIdent:

```
> fm4Orth.lme <- update(fm2Orth.lme, random=pdIdent(~ I(age-11)))  
> summary(fm4Orth.lme)
```

Linear mixed-effects model fit by REML

Data: Orthodont

AIC BIC logLik

493.2222 509.0886 -240.6111

Random effects:

Formula: ~I(age - 11) | Subject

Structure: Multiple of an Identity

 (Intercept) I(age - 11) Residual

StdDev: 1.116357 1.116357 1.398877

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Variances of intercept and slope

Fixed effects: distance ~ Sex + I(age - 11) + Sex:I(age - 11)

	Value	Std.Error	DF	t-value	p-value
(Intercept)	24.968750	0.3293429	79	75.81383	<.0001
SexFemale	-2.321023	0.5159812	25	-4.49827	0.0001
I(age - 11)	0.784375	0.2898380	79	2.70625	0.0083
SexFemale:I(age - 11)	-0.304830	0.4540888	79	-0.67130	0.5040

Correlation:

	(Intr)	SexFml	I(-11)
SexFemale	-0.638		
I(age - 11)	0.000	0.000	
SexFemale:I(age - 11)	0.000	0.000	-0.638

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.58766910	-0.40427596	0.04511455	0.38598268	3.54421578

Number of Observations: 108

Number of Groups: 27

> anova(fm3Orth.lme, fm4Orth.lme)

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
fm3Orth.lme	1	7	446.8426	465.3533	-216.4213		
fm4Orth.lme	2	6	493.2222	509.0886	-240.6111	1 vs 2	48.37964 <.0001

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo

Are slopes the same for different subjects?

(e) Is random slope necessary? (no)

```
> fm5Orth.lme <- update(fm2Orth.lme, random=~1)
> anova(fm2Orth.lme,fm5Orth.lme)
      Model df     AIC     BIC   logLik   Test  L.Ratio p-value
fm2Orth.lme     1  8 448.5824 469.7376 -216.2912
fm5Orth.lme     2  6 445.7572 461.6236 -216.8786 1 vs 2 1.174804  0.5558
```

Read Section 4.2.2 of PB for more examples of pdMat classes



Diagnostics for random effects

Assumptions:

- ① normality
- ② mean zero and covariance matrix independent of the group
- ③ independent for different groups

Tools:

- `ranef` to extract estimates of random effects
- `qqnorm` to check marginal normality
- `pairs` to identify outliers and check homogeneity

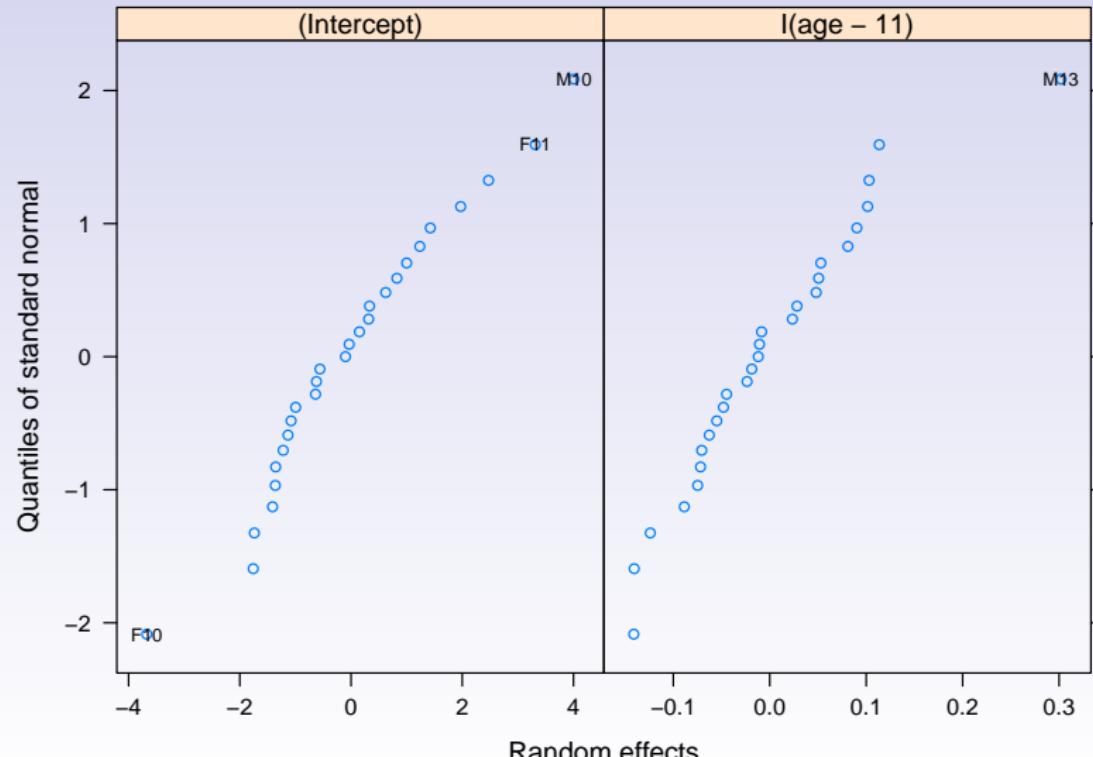
```
oooooooooooooooooooooooooooo
```

Diagnostics for random effects

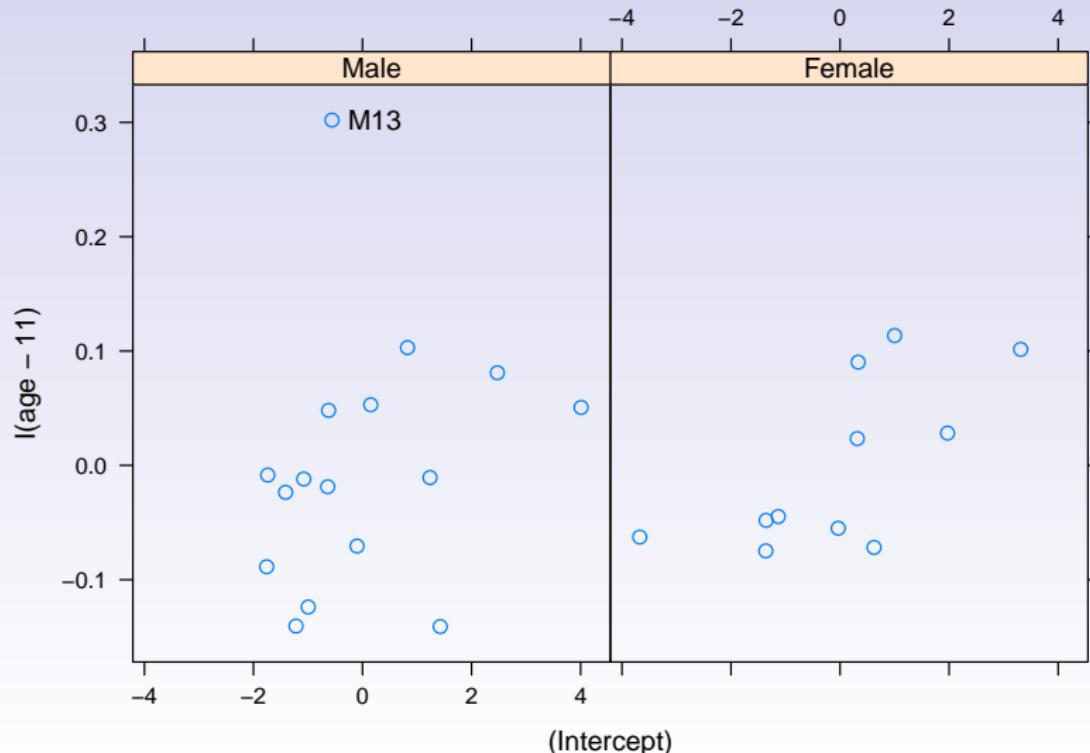
```
# diagnosis: normality  
qqnorm( fm2Orth.lme, ~ranef(.), id=.1, cex=.7)  
  
# diagnosis: homogeneity between boy and girls  
pairs( fm2Orth.lme, ~ranef(.) | Sex,  
      id=~Subject=="M13", adj=-.3)
```

- some outliers: F10, F11 and M10
- normality assumption seems reasonable
- Except for subject M13, pair plot does not reveal violations to homogeneity

QQ plots of estimated random effects



Estimates of random effects by sex



oooooooooooooooooooooooooooo

Fitting similar models using lmer

```
fm1Orth.lmer <- lmer( distance ~ 1 + I(age-11) +
                      (1 + I(age-11) | Subject),
                      data=Orthodont )
fm2Orth.lmer <- update(fm1Orth.lmer, .~. + Sex + Sex:I(age-11))
fm3Orth.lmer <- lmer(distance ~ 1 + Sex*I(age-11) +
                      (1 | Subject) + (0+I(age-11) | Subject),
                      data=Orthodont )
```



Outline

1

Model

- Single level of grouping
- Two nested levels of grouping
- The general model

2

Estimation

- Review
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Prediction of random effects

3

Hierarchical Models and Bayes Estimation

4

lme, lmer and proc mixed

5

Examples

- Personnel data
- Spectrophotometer data
- Machine data
- Wheat data
- Lab data
- Orthodont data

6

Inference

- Inference for variance components
- Inference for fixed and random effects
- Model selection

7

Modeling the Covariance Structure

- Modeling random effects
- Modeling within-group random errors



Modeling within-group random errors

Recall the vector form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, M$$

where

$$\boldsymbol{\epsilon}_i \stackrel{iid}{\sim} N(\mathbf{0}, \sigma^2 R_i)$$

Thus we need to model R_i , and check other assumptions such as normality.

oooooooooooooooooooooooooooo

Decomposition of R

Ignoring the subscript i , we can always write

$$\begin{aligned} R &= \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \cdots & \sigma_2\sigma_n\rho_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_1\sigma_n\rho_{1n} & \sigma_2\sigma_n\rho_{2n} & \cdots & \sigma_n^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \\ &= VCV \end{aligned}$$



Modeling within-group random errors

- V decides the variance
- C decides the correlation
- R is decomposed into a *variance structure* and a *correlation structure*
- we will model each structure separately



Modeling variance structure

V allow us to model heteroscedasticity, when necessary. In general, the variance function is assumed to have the form

$$\text{Var}(\epsilon_{ij}|\boldsymbol{b}_i) = \sigma^2 g^2(\mu_{ij}, \mathbf{v}_{ij}, \boldsymbol{\delta}), \quad i = 1, \dots, M, \quad j = 1, \dots, n_i$$

- $\mu_{ij} = E(y_{ij}|\boldsymbol{b}_i)$, the conditional mean. Thus the variance depends on the random effects \boldsymbol{b}_i if g depends on μ_{ij}
- \mathbf{v}_{ij} is a vector of *variance covariates*
- $\boldsymbol{\delta}$ is a vector of *variance parameters*
- g is a known function which is assumed to be continuous in $\boldsymbol{\delta}$
- This form is very general: it allows the variance to depend on the conditional mean and other covariates in a complicated manner

Example $\text{Var}(\epsilon_{ij}|\boldsymbol{b}_i) = \sigma^2 |\mu_{ij}|^{2\delta}$ corresponds to $g(x, y) = |x|^y$.



Problem with the general variance function

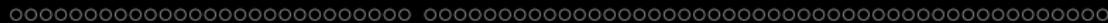
The variance function depends on the conditional mean which in turn depends on the random effects. Thus random errors are not independent of random effects. This poses some theoretical and computational difficulties. In `lme`, an approximate variance model

$$\text{Var}(\epsilon_{ij} | \mathbf{b}_i) \approx \sigma^2 g^2(\hat{\mu}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\delta}), \quad i = 1, \dots, M, j = 1, \dots, n_i$$

where $\hat{\mu}_{ij}$ is the BLUP of μ_{ij} . In the calculation, the iteratively reweighted scheme is used: given current estimates of parameters,

- ① compute BLUP $\hat{\mu}_{ij}$
- ② produce new estimates of parameters

Iterate until convergence. Note that if g does not depend on μ_{ij} , then no approximation is needed.



Specify variance structure

`lme` uses `varFunc` classes to specify *variance structure*, i.e. the function g

<code>varFixed</code>	fixed variances
<code>varIdent</code>	different variances per stratum
<code>varPower</code>	power of covariate
<code>varExp</code>	exponential of covariate
<code>varConstPower</code>	constant plus power of covariate
<code>varComb</code>	combination of variance functions

Table: Standard `varFunc` classes.

Specify variance structure

Most `varFunc` constructors have three arguments:

- `form`: a one-sided formula specifying the variance covariate v

```
form = ~ v
```

If we want different parameters for each level of a stratification variable, say s , then

```
form = ~ v | s
```

To use the fitted values as the variance covariate, one uses

```
form = ~ fitted(.)
```

- `value`: specifies the value of the variance parameters δ .
Useful for specifying initial values
- `fixed`: to fix some parameters

More detail can be found using

```
> help(varFunc)
```



Standard varFunc classes

- varFixed. $g(v) = \sqrt{|v|}$. No parameters.
 $\text{varFixed}(\sim v)$
- varIdent. Let s be a stratification variable taking values in the set $\{1, 2, \dots, S\}$. Then

$$g(s, \delta) = \delta_s$$

where s is the level for an observation. So each stratum has its own variance. Together with σ^2 , there are $S + 1$ parameters representing S variances. For identifiability, $\delta_1 = 1$. Thus δ_i represents the ratio between the standard deviations of the i th stratum and the first stratum.

`varIdent(form=~1 | s)`



Standard varFunc classes

- varPower. $g(v, \delta) = |v|^\delta$.

`varPower(form=~v)`

If we want different power parameters for a stratification variable s , then $g(v, s, \delta) = |v|^{\delta_s}$, and

`varPower(form=~v | s)`

- varExp. $g(v) = \exp(\delta v)$.

`varExp(form=~v)`

- varConstPower. $g(v, \delta) = \delta_1 + |v|^{\delta_2}$.

`varExp(form=~v)`

- varComb. Allow the combination of two or more variance models which multiply together. $g = g_1 \cdots g_K$ for any K , and each g_k , $k = 1, \dots, K$ can be any one of above varFunc classes.

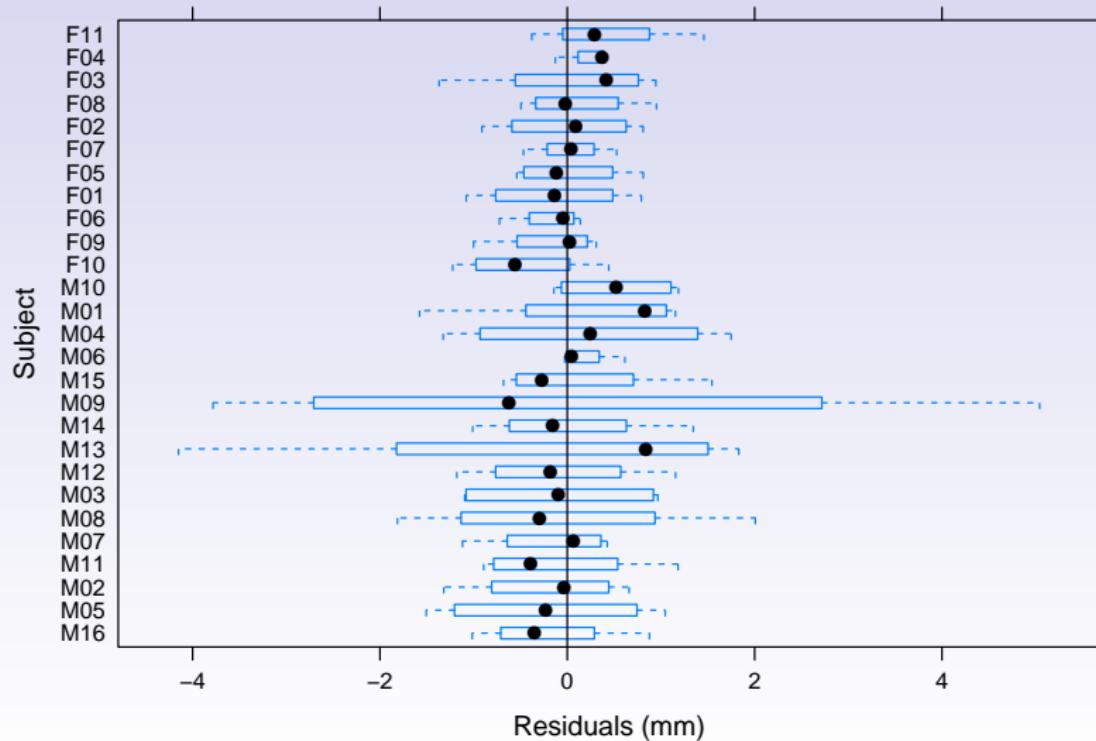
oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: orthodont data continued

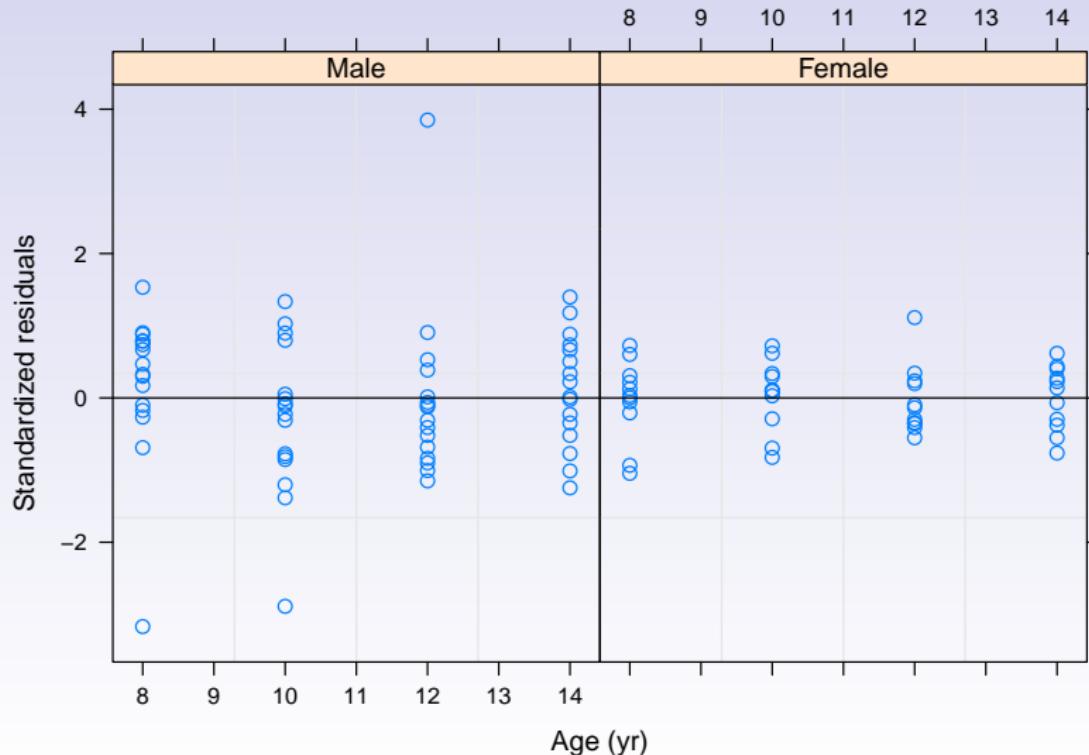
(f) Diagnosis for within-group random errors. So far we have assumed iid normal random errors. Now we want to check normality and constant variance assumptions. We will learn how to check the independence assumption later.

```
# boxplot of residuals for each subject  
> plot(fm2Orth.lme, Subject~resid(.), abline=0)  
  
# does variance depend on age?  
> plot(fm2Orth.lme, resid(.,type=''p'')~age | Sex, abline=0)  
  
# does variance depend on Sex?  
> plot(fm2Orth.lme, resid(.,type=''p'')~fitted(.) | Sex,  
      id=.05, ajd=-.3)
```

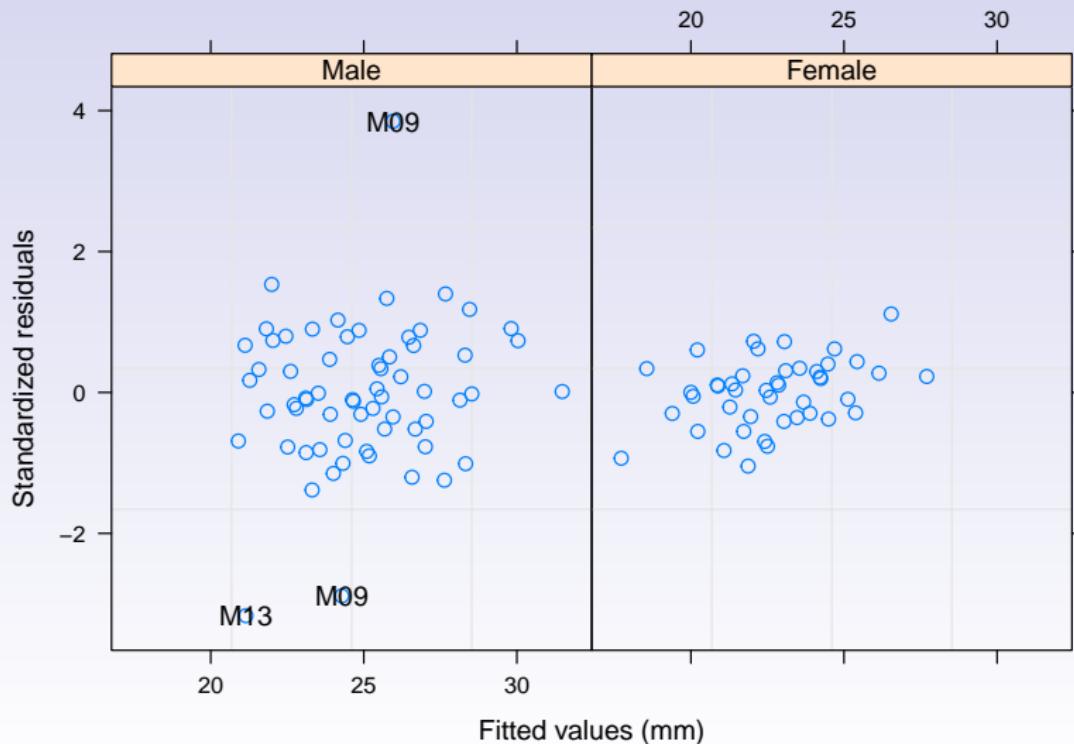
Box plots of residuals



Standardized residuals vs age



Standardized residuals vs fitted values



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit with unequal variances

(g) Fit with unequal variances for boys and girls. Variance structures are specified in `lme` using the `weights` argument.

```
> fm6Orth.lme <- update( fm2Orth.lme, weights=varIdent(form=~1|Sex))

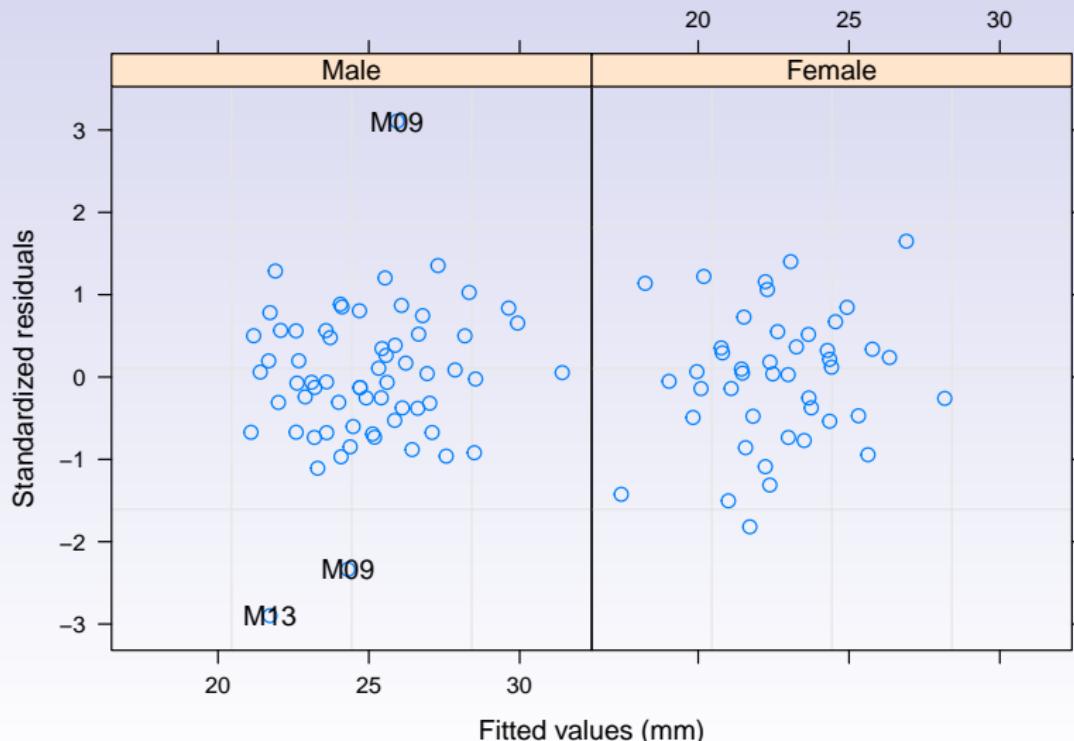
> anova(fm2Orth.lme, fm6Orth.lme)
      Model df     AIC     BIC   logLik   Test  L.Ratio p-value
fm2Orth.lme     1 8 448.5817 469.7368 -216.2908
fm6Orth.lme     2 9 429.5227 453.3222 -205.7613 1 vs 2 21.05899 <.0001

# difference disappeared
> plot(fm6Orth.lme, resid(.,type='p')~fitted(.) | Sex, id=.05, adj=-.3)

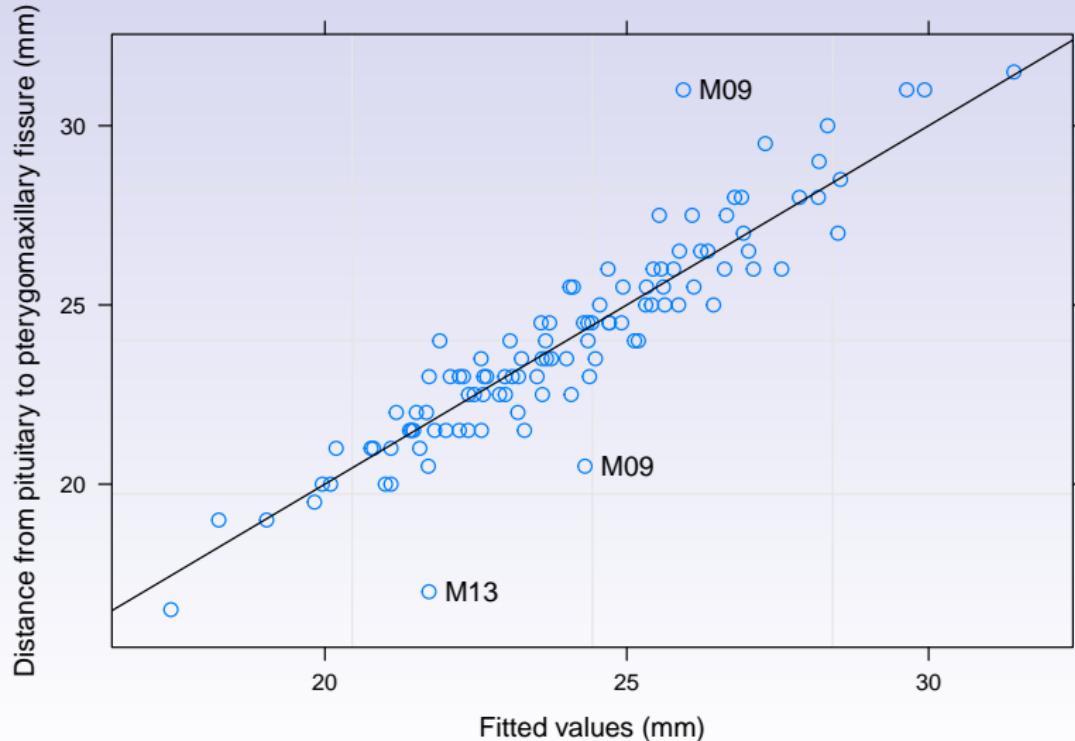
# check overall fit
> plot( fm6Orth.lme, distance ~ fitted(.), abline=c(0,1), id=.05, adj=-.3 )

# check normality
> qqnorm( fm6Orth.lme, ~resid(.) | Sex)
```

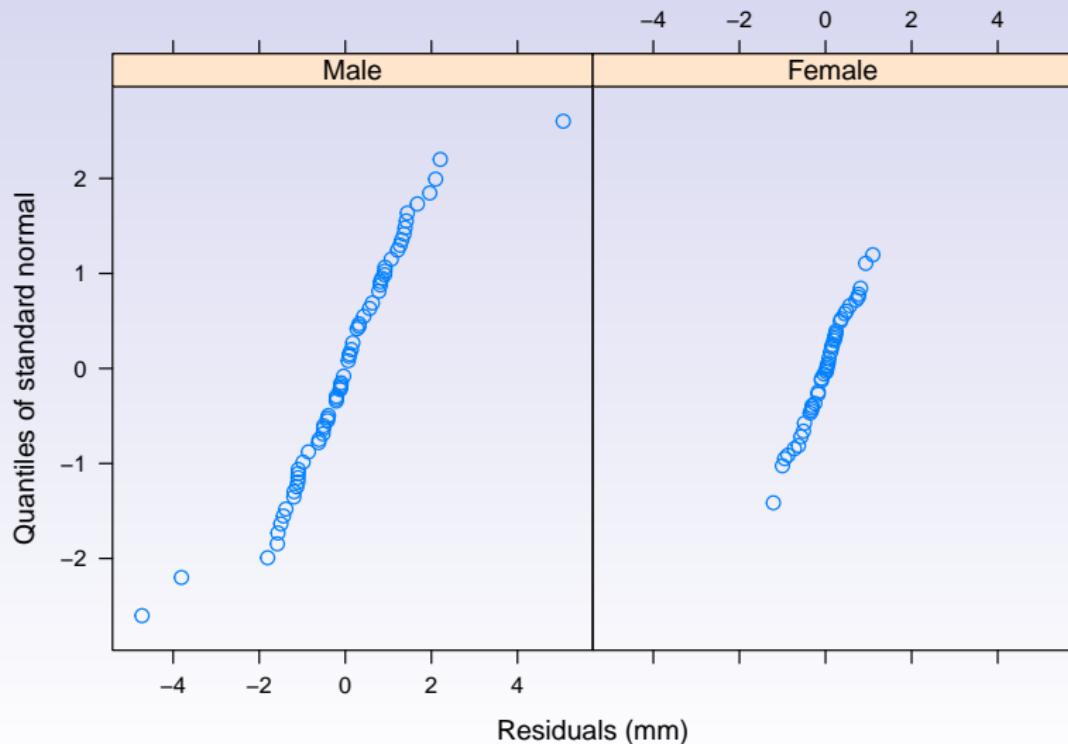
Standardized residuals vs fitted values for the new fit



Observed vs fitted



QQ plot of residuals





Example: dialyzer data

Vonesh and Carter (1992) described data measured on high-flux hemodialyzers to assess their in vivo ultrafiltration characteristics. The ultrafiltration rates (in mL/hr) of 20 high-flux dialyzers were measured at seven different transmembrane pressures (in dmHg). The in vitro evaluation of the dialyzers used bovine blood at flow rates of either 200 dL/min or 300 dL/min.

This data frame `Dialyzer` contains the following columns:

`Subject` an ordered factor giving the unique identifier for each subject

`QB` a factor with levels ‘200’ and ‘300’ giving the bovine blood flow rate (dL/min).

`pressure` a numeric vector giving the transmembrane pressure (dmHg).

`rate` the hemodialyzer ultrafiltration rate (mL/hr).

`index` index of observation within subject-1 through 7.

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: dialyzer data

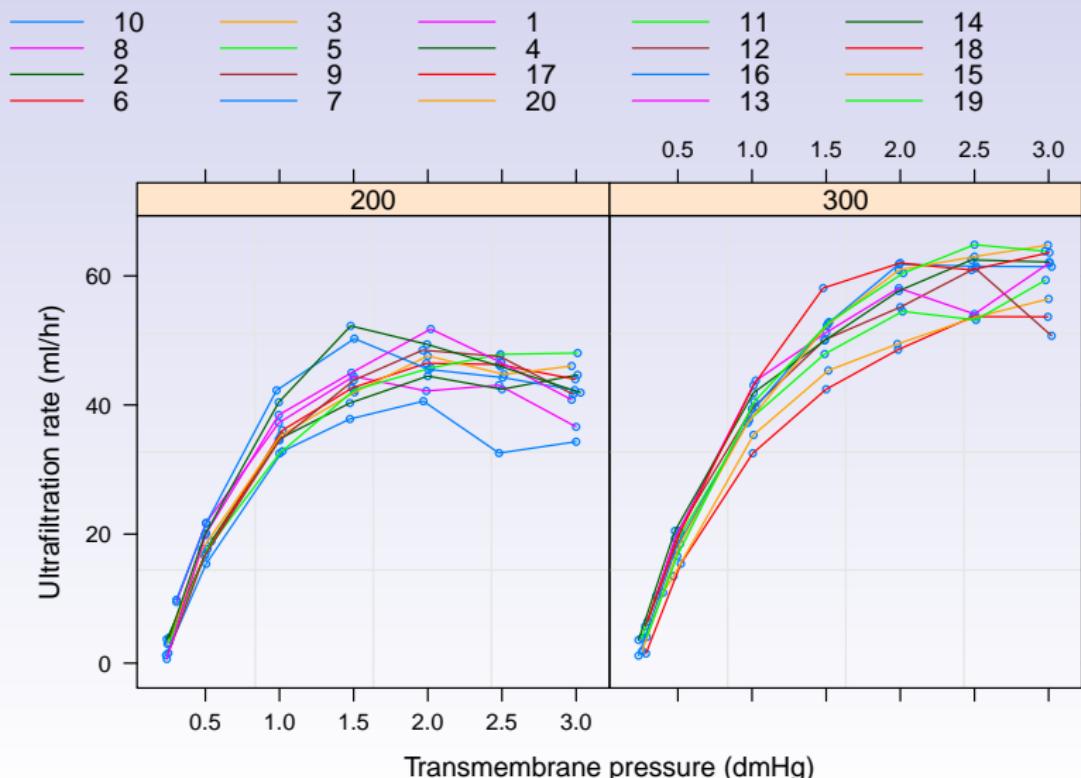
(a) Plot data and fit a polynomial model

```
> plot(Dialyzer,outer=T,layout=c(2,1),aspect=1)

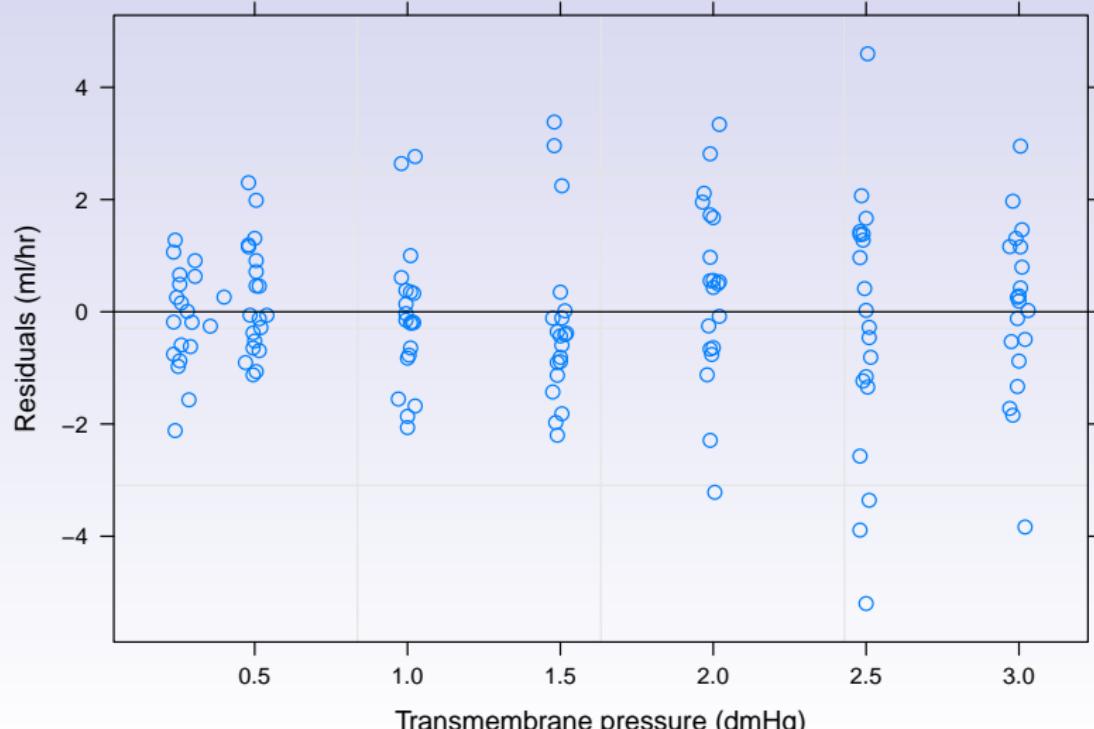
# random coefficients to the 2nd power
> fm1Dial.lme <- lme(rate ~ (pressure+I(pressure^2)+I(pressure^3)
+I(pressure^4))*QB,
Dialyzer, ~ pressure + I(pressure^2))

# plot residuals
> plot( fm1Dial.lme, resid(.)~pressure, abline=0)
```

Plot of dialyzer data



Residuals vs pressure



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit with unequal variances

(b) Variance increases with pressure. Fit a power varFunc

```
> fm2Dial.lme <- update( fm1Dial.lme, weights=varPower(form=~pressure))
> fm2Dial.lme
Linear mixed-effects model fit by REML
  Data: Dialyzer
  Log-restricted-likelihood: -306.0399
  Fixed: rate ~ (pressure + I(pressure^2) + I(pressure^3) + I(pressure^4)) *
    (Intercept)           pressure      I(pressure^2)      I(pressure^3)      QB
    -16.7589669        92.3577109     -49.6662342     11.7535207
    I(pressure^4)          QB300      pressure:QB300  I(pressure^2):QB300
    -1.0963009        -1.8415868      2.7061956      0.9606231
  I(pressure^3):QB300  I(pressure^4):QB300
    0.9827223        -0.2925454

Random effects:
  Formula: ~pressure + I(pressure^2) | Subject
  Structure: General positive-definite, Log-Cholesky parametrization
            StdDev   Corr
  (Intercept) 1.857003 (Intr) pressr
  pressure    5.327563 -0.522
  I(pressure^2) 1.648061  0.362 -0.954
  Residual     1.262651

Variance function:
  Structure: Power of variance covariate
  Formula: ~pressure
  Parameter estimates:
    power
  0.7492193
  Number of Observations: 140
  Number of Groups: 20
```

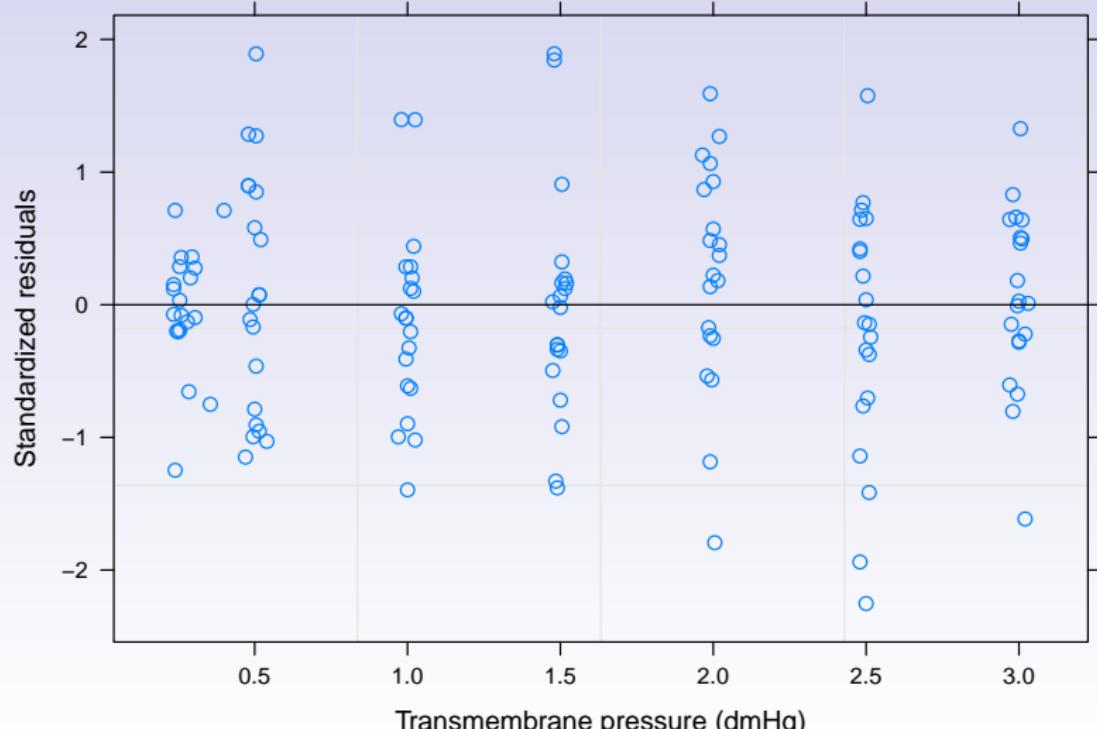
oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit with unequal variances

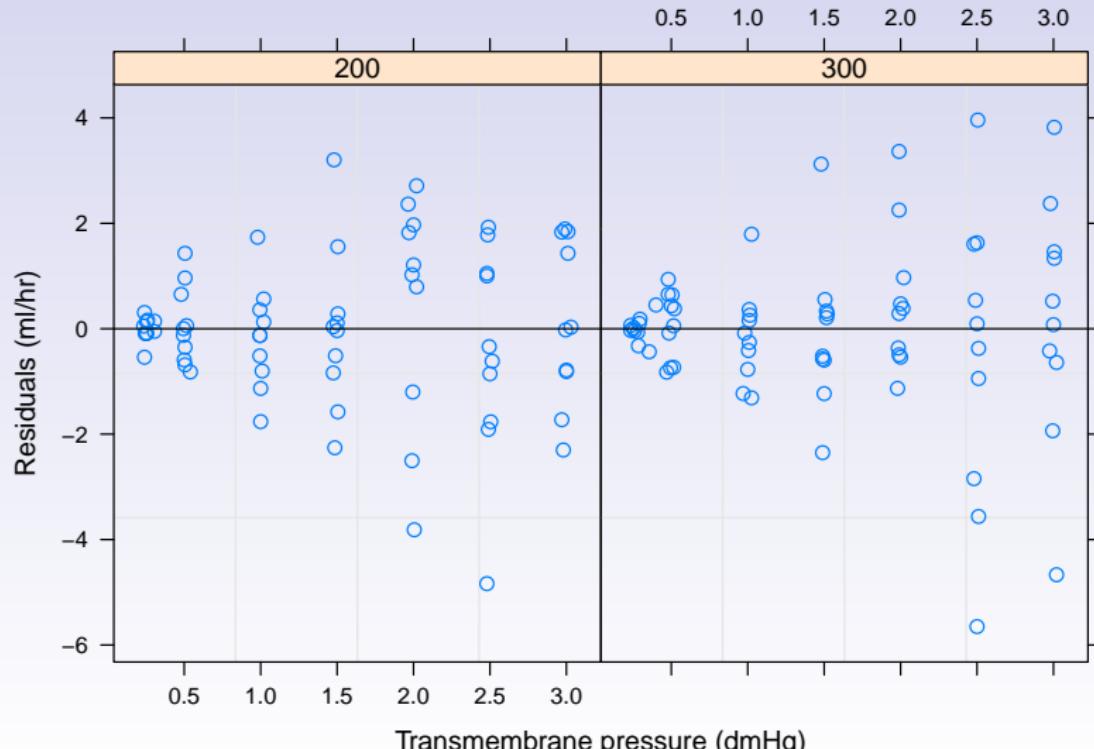
```
> anova( fm1Dial.lme, fm2Dial.lme)
      Model df     AIC     BIC   logLik   Test  L.Ratio p-value
fm1Dial.lme     1 17 679.8500 728.5980 -322.9250
fm2Dial.lme     2 18 648.0798 699.6955 -306.0399 1 vs 2 33.77011 <.0001

> intervals( fm2Dial.lme)
...
Variance function:
    lower      est.      upper
power 0.5079076 0.7492193 0.990531
...
# plot standardized residuals
> plot( fm2Dial.lme, resid(.,type="p")~pressure, abline=0)
#plot residuals by QB
> plot( fm2Dial.lme, resid(.)~pressure|QB, abline=0)
```

Standardized residuals vs pressure



Residuals vs pressure for different flow levels



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooooooooooo

Fit with unequal variances

(c) Heteroscedastic patterns seem to be the same for both levels. To test formally, we fit

```
> (fm3Dial.lme <- update(fm2Dial.lme, weights=varPower(form=~pressure|QB)))
Linear mixed-effects model fit by REML
  Data: Dialyzer
  Log-restricted-likelihood: -305.6845
  Fixed: rate ~ (pressure + I(pressure^2) + I(pressure^3) + I(pressure^4)) * QB
          (Intercept)           pressure      I(pressure^2)      I(pressure^3)
          -16.6781090        91.9198618       -49.0362343      11.4332175
          I(pressure^4)         QB300      pressure:QB300  I(pressure^2):QB300
          -1.0440399        -2.0338647        3.6791667      -0.3888335
  I(pressure^3):QB300 I(pressure^4):QB300
          1.6544754        -0.4007027
```

Random effects:

```
Formula: ~pressure + I(pressure^2) | Subject
Structure: General positive-definite, Log-Cholesky parametrization
          StdDev   Corr
(Intercept) 1.824399 (Intr) pressr
pressure     5.242070 -0.502
I(pressure^2) 1.637342  0.338 -0.951
Residual     1.259087
```

Variance function:

```
Structure: Power of variance covariate, different strata
```

```
Formula: ~pressure | QB
```

```
Parameter estimates:
```

```
    200      300
0.6477326 0.8378068
```

```
Number of Observations: 140
```

```
Number of Groups: 20
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

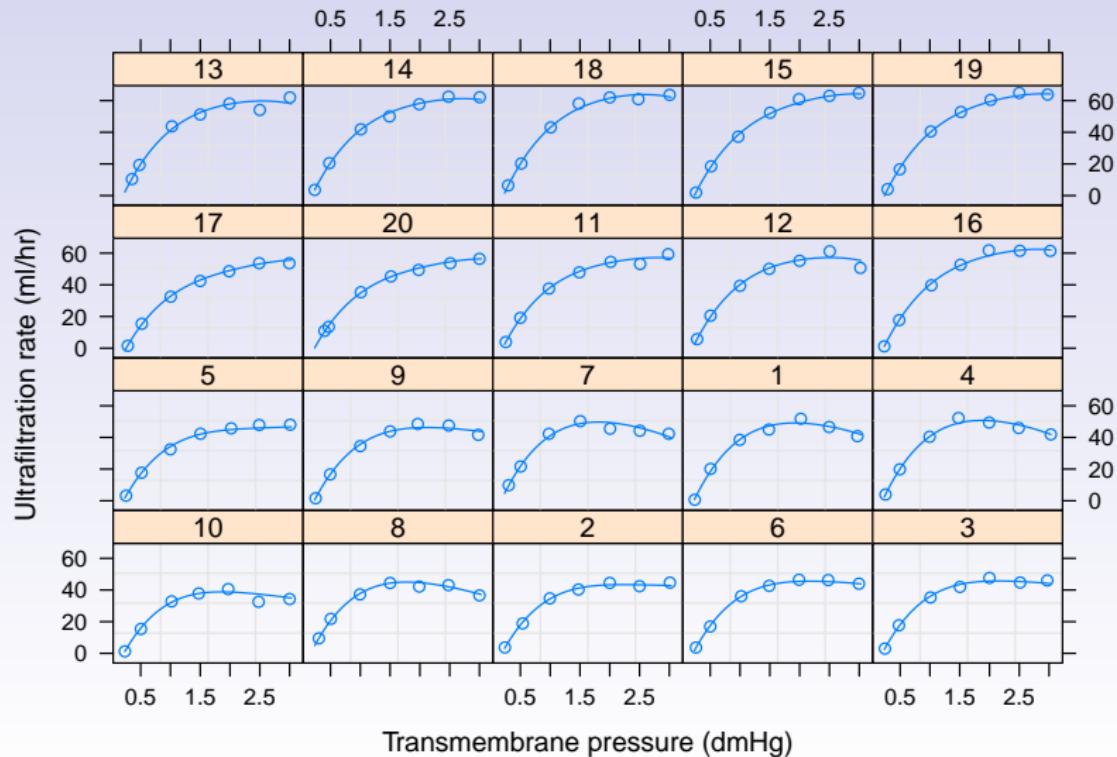
Fit with unequal variances

```
> anova(fm2Dial.lme, fm3Dial.lme)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fm2Dial.lme     1 18 648.0798 699.6955 -306.0399
fm3Dial.lme     2 19 649.3690 703.8521 -305.6845 1 vs 2 0.710896 0.3991

# constant plus power
> fm4Dial.lme <- update(fm2Dial.lme, weights=varConstPower(form=~pressure))
> anova(fm2Dial.lme, fm4Dial.lme)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fm2Dial.lme     1 18 648.0798 699.6955 -306.0399
fm4Dial.lme     2 19 650.0595 704.5427 -306.0298 1 vs 2 0.02032585  0.8866

# plot the fits
> plot( augPred(fm2Dial.lme), grid=T)
```

Plot of fits



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Fit with unequal variances

```
> anova( fm2Dial.lme )
      numDF denDF   F-value p-value
(Intercept)       1     112  552.7042 <.0001
pressure          1     112 2328.8701 <.0001
I(pressure^2)     1     112 1174.8448 <.0001
I(pressure^3)     1     112  359.8875 <.0001
I(pressure^4)     1     112   12.4689 0.0006
QB                1      18    4.8283 0.0413
pressure:QB       1     112   80.1173 <.0001
I(pressure^2):QB  1     112    1.3507 0.2476
I(pressure^3):QB  1     112    2.2436 0.1370
I(pressure^4):QB  1     112    0.1666 0.6839

# we can test terms jointly
> anova( fm2Dial.lme, Terms = 8:10)
F-test for: I(pressure^2):QB, I(pressure^3):QB, I(pressure^4):QB
      numDF denDF   F-value p-value
1         3     112  1.253624  0.2939
```



Modeling correlation structure

C allow us to model correlation structure for within-group random errors. Most of the correlation structures were developed for *time series* and *spatial data*.

time series : observations are indexed by an integer-valued *time* variable. For example, observations are made on equally spaced time points

spatial data : observations are indexed by a two-dimensional *spatial location* variable. For example, longitude and latitude, x-y locations in an image

As before, we use single level of grouping for illustration.
Methods can be generalized to multiple levels of grouping.



Modeling correlation structure

In general, we assume within-group random errors ϵ_{ij} are associated with a *position* vector \mathbf{p}_{ij} . Position could be time point, spatial location or other variables.

In most situations the correlation structure is *isotropic*

$$\text{cor}(\epsilon_{ij}, \epsilon_{ij'}) = h(d(\mathbf{p}_{ij}, \mathbf{p}_{ij'}), \rho)$$

- $d(\mathbf{p}_{ij}, \mathbf{p}_{ij'})$ is some *distance* between \mathbf{p}_{ij} and $\mathbf{p}_{ij'}$
- h is a known function with range between -1 and 1
- ρ is a vector of parameters
- modeling correlation structure then boils down to modeling the function h



Modeling correlation structure

We distinguish two situations:

- ① p is a scalar with domain $N = \{1, 2, \dots\}$. This is usually for time series data. In this case we denote p as ρ
- ② $p \in R^d$. This is usually for spatial data and time series with unequally spaced observations



Situation 1: general

For $p \in N$, no structure is assumed:

$$\text{cor}(\epsilon_{jj}, \epsilon_{jj'}) = \rho_{jj'}$$

- also called unstructured. It is non-isotropic
- the number of parameters increases quadratically
- useful when the number of observations per group is small
- useful as a exploratory tool for more parsimonious model



Situation 2: serial

Serial (isotropic) correlation structure assumes

$$\text{cor}(\epsilon_{ij}, \epsilon_{ij'}) = h(|p_{ij} - p_{ij'}|, \rho)$$

Since $|p_{ij} - p_{ij'}|$ takes non-negative integer values, we only need to define $h(k, \rho)$ for all $k = 1, 2, \dots$. Note that $h(0, \rho) = 1$.



Compound symmetry

$$h(k, \rho) = \rho, \quad k = 1, 2, \dots$$

where ρ is called *intraclass* correlation coefficient. It has a close relationship with one-way random effect model.



Relation to the random intercept

Consider the simple one-way random effect model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

where μ is the overall mean, $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_1^2)$, $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$, and α_i and ϵ_{ij} are independent of each other.

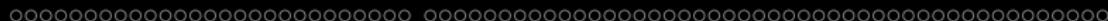
If instead we write the model as

$$y_{ij} = \mu + e_{ij}$$

where $e_{ij} = \alpha_i + \epsilon_{ij}$, then it is easy to check that

$$\text{cor}(e_{ij}, e_{ij'}) = \sigma_1^2 / (\sigma_1^2 + \sigma^2) = \rho$$

Thus e_{ij} have a compound symmetry structure. The minor difference is that this ρ is non-negative, where ρ defined via h can take negative values.



Autoregressive-moving average

Autoregressive model of order p , AR(p), assumes that

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \cdots + \phi_p \epsilon_{t-p} + a_t$$

where a_t is independent of previous observations with mean zero.

For AR(1),

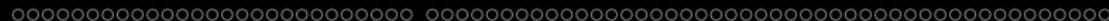
$$h(k, \phi) = \phi^k, \quad k = 0, 1, 2, \dots$$

Continuous time AR(1), CAR(1), extends above to

$$h(s, \phi) = \phi^s, \quad s \geq 0, \quad \phi \geq 0$$

When $p > 1$, there is no simple form for h . Instead, it is defined recursively

$$h(k, \phi) = \phi_1 h(|k - 1|, \phi) + \cdots + \phi_p h(|k - p|, \phi)$$



Autoregressive-moving average

Moving average model of order q , MA(q), assume that

$$\epsilon_t = \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + a_t$$

$$h(k, \theta) = \begin{cases} \frac{\theta_k + \theta_1\theta_{k-1} + \cdots + \theta_{k-q}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & k = 1, \dots, q, \\ 0, & k = q+1, q+2, \dots \end{cases}$$

Autoregressive-moving average model ARMA(p,q) assume that

$$\epsilon_t = \sum_{i=1}^p \phi_i \epsilon_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + a_t$$

h is more complicated, but can be found in PB, Section 5.3.1.



Situation 3: spatial correlation structure

Assume that $p \in R^d$. Note that $d = 1$ is a special case. Thus these structures can also be used for time series, especially for unequally spaced observations.

Spatial correlation structures are usually represented by their *semivariogram*, instead of correlation function.



Semivariogram

For any two observations $\epsilon \mathbf{p}_1$ and $\epsilon \mathbf{p}_2$ taken at positions \mathbf{p}_1 and \mathbf{p}_2 , the *semivariogram* is defined as

$$\gamma(d(\mathbf{p}_1, \mathbf{p}_2), \lambda) = \frac{1}{2} \text{Var}(\epsilon \mathbf{p}_1 - \epsilon \mathbf{p}_2) = \frac{1}{2} E(\epsilon \mathbf{p}_1 - \epsilon \mathbf{p}_2)^2$$

- The last equality follows from the assumption $E(\epsilon \mathbf{p}_1) = E(\epsilon \mathbf{p}_2) = 0$
- $\frac{1}{2} \text{Var}(\epsilon \mathbf{p}_1 - \epsilon \mathbf{p}_2) = \text{Var}(\epsilon \mathbf{p}_1)(1 - \text{cor}(\epsilon \mathbf{p}_1, \epsilon \mathbf{p}_2))$. For identifiability, set $\text{Var}(\epsilon \mathbf{p}_1) = 1$. Then

$$\gamma(s, \rho) = 1 - h(s, \rho)$$



Nugget effect

From above relationship, $\gamma(0, \rho) = 1 - h(0, \rho) = 0$. It is sometimes desirable to account for abrupt changes at very small distances. This is usually done by allowing

$$\gamma(s, \rho) \xrightarrow{s \rightarrow 0} c_0$$

where $0 < c_0 < 1$. This is called the *nugget effect*. For any continuous correlation function h_{cont} , we can incorporate a nugget effect by creating a new correlation function

$$h_{nugg}(s, c_0, \rho) = \begin{cases} (1 - c_0)h_{cont}(s, \rho), & s > 0 \\ 1, & s = 0 \end{cases}$$

(note the typo in PB)

Some commonly used isotropic variogram models

Exponential	$\gamma(s, \rho) = 1 - \exp(-s/\rho)$
Gaussian	$\gamma(s, \rho) = 1 - \exp[-(s/\rho)^2]$
Linear	$\gamma(s, \rho) = 1 - (1 - s/\rho)I(s < \rho)$
Rational quadratic	$\gamma(s, \rho) = (s/\rho)^2/[1 + (s/\rho)^2]$
Spherical	$\gamma(s, \rho) = 1 - [1 - 1.5(s/\rho) + .5(s/\rho)^3]I(s < \rho)$



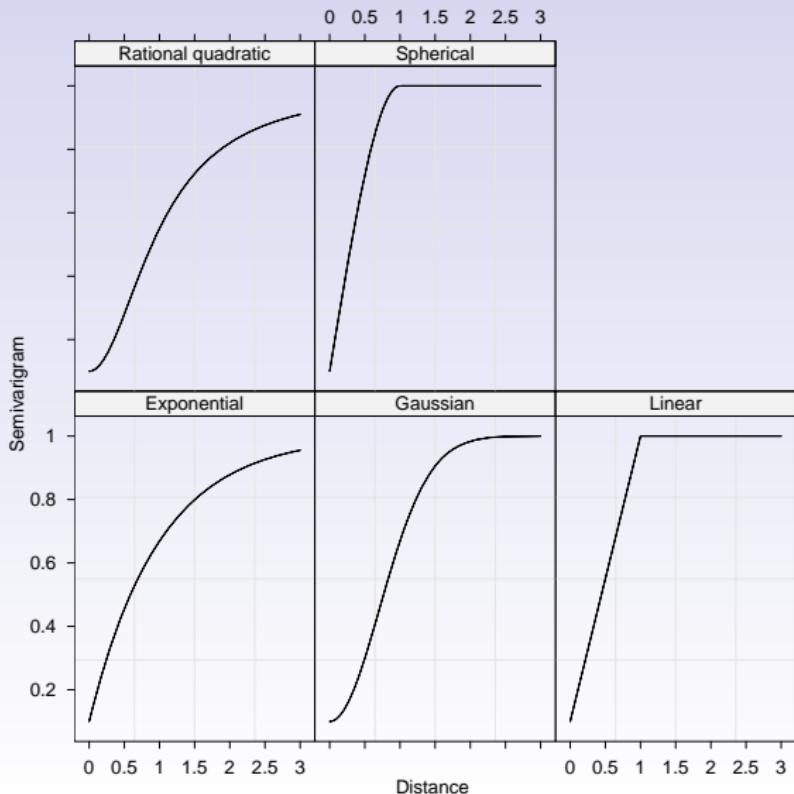
$$I(s < \rho) = \begin{cases} 1, & s < \rho \\ 0, & s \geq \rho \end{cases}$$

- the exponential spatial correlation is equivalent to CAR(1) since

$$h(s, \rho) = \exp(-s/\rho) = \phi^s$$

where $\phi = \exp(-1/\rho)$

Semivarigrams





Nugget effect

A nugget effect is added using

$$\gamma_{nugg}(s, c_0, \rho) = \begin{cases} c_0 + (1 - c_0)\gamma(s, \rho), & s > 0 \\ 0, & s = 0 \end{cases}$$

The parameters ρ is called *range*



Standard covStruct classes

lme uses corStruct classes to specify *correlation structure*, i.e. the function h

corCompSymm	compound symmetry
corSymm	general
corAR1	AR(1)
corCAR1	CAR(1)
corARMA	ARMA
corExp	exponential
corGaus	Gaussian
corLin	linear
corRatio	rational quadratic
corSpher	spherical

Specify covStruct

- **form:** a one-sided formula specifying the *position* vector p

$$\text{form} = \sim p_1 + \dots + p_d$$

If we want the correlation structure to apply only to observations within the same grouping level for a group variable, say s , then

$$\text{form} = \sim p_1 + \dots + p_d \mid s$$

- **value:** specifies the value of the correlation parameters ρ .
Useful for specifying initial values
- **fixed:** to fixed some parameters

Correlation structures are specified in `lme` using the correlation argument.

oooooooooooooooooooooooooooo

Example: counts of ovarian follicles

The data frame `Ovary` contains the following columns:

`Mare` an ordered factor indicating the mare on which the measurement is made.

`Time` time in the estrus cycle. The data were recorded daily from 3 days before ovulation until 3 days after the next ovulation. The measurement times for each mare are scaled so that the ovulations for each mare occur at times 0 and 1.

`follicles` the number of ovarian follicles greater than 10 mm in diameter.

oooooooooooooooooooooooooooo

Example: counts of ovarian follicles

(a) compare 3 models and check autocorrelations in residuals

```
> plot(Ovary)

> fm1Ovar.lme <- lme( follicles ~ sin(2*pi*Time)+cos(2*pi*Time), data=Ovary)
> fm2Ovar.lme <- lme( follicles ~ sin(2*pi*Time)+cos(2*pi*Time), data=Ovary,
+                      random=pdDiag(~sin(2*pi*Time)+cos(2*pi*Time)))
> fm3Ovar.lme <- lme( follicles ~ sin(2*pi*Time)+cos(2*pi*Time), data=Ovary,
+                      random=pdDiag(~sin(2*pi*Time)))
> anova(fm1Ovar.lme,fm2Ovar.lme,fm3Ovar.lme)
      Model df     AIC     BIC   logLik   Test  L.Ratio p-value
fm1Ovar.lme     1 10 1630.034 1667.237 -805.0168
fm2Ovar.lme     2  7 1633.616 1659.658 -809.8078 1 vs 2  9.581807  0.0225
fm3Ovar.lme     3  6 1638.082 1660.404 -813.0409 2 vs 3  6.466236  0.0110
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: counts of ovarian follicles

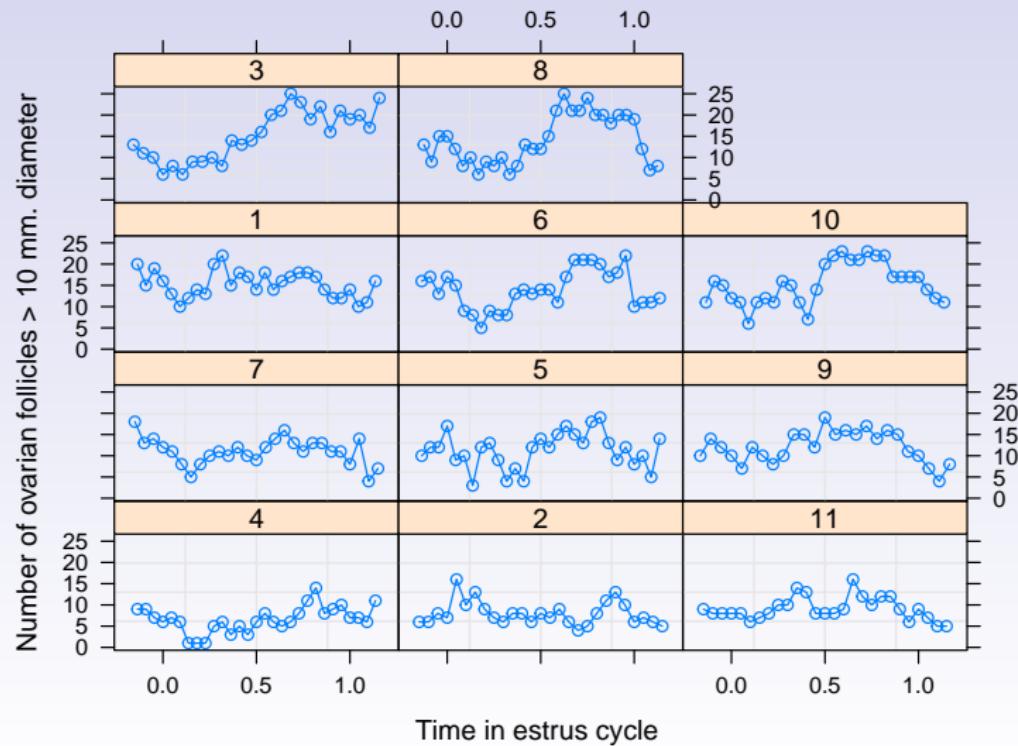
```
# use fm3Ovar.lme as in book
> fm3Ovar.lme
Linear mixed-effects model fit by REML
Data: Ovary
Log-restricted-likelihood: -813.0409
Fixed: follicles ~ sin(2 * pi * Time) + cos(2 * pi * Time)
      (Intercept) sin(2 * pi * Time) cos(2 * pi * Time)
      12.1820241          -3.2985373          -0.8623725

Random effects:
Formula: ~sin(2 * pi * Time) | Mare
Structure: Diagonal
      (Intercept) sin(2 * pi * Time) Residual
StdDev:    3.052136          2.079312 3.112854

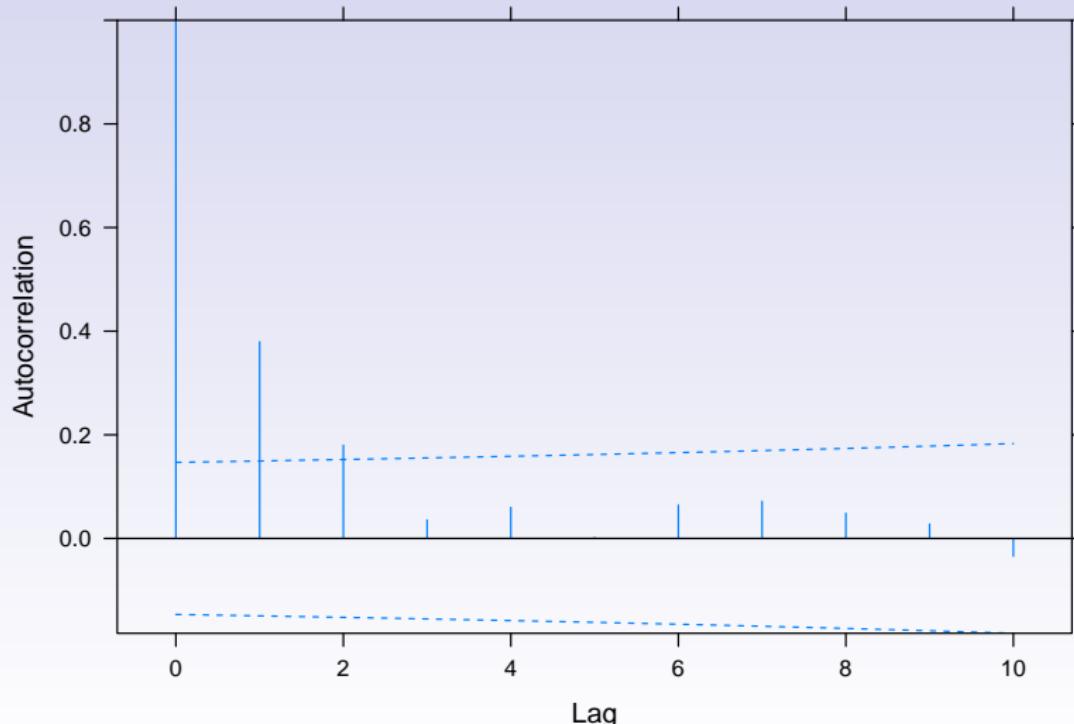
Number of Observations: 308
Number of Groups: 11

> plot(ACF(fm3Ovar.lme, maxLag=10), alpha=.01)
```

Plot of ovary data



Empirical autocorrelation function



oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: counts of ovarian follicles

(b) fit with AR(1) within-group random error structure

```
> fm4Ovar.lme <- update(fm3Ovar.lme, correlation=corAR1())
> fm4Ovar.lme
Linear mixed-effects model fit by REML
Data: Ovary
Log-restricted-likelihood: -774.724
Fixed: follicles ~ sin(2 * pi * Time) + cos(2 * pi * Time)
        (Intercept) sin(2 * pi * Time) cos(2 * pi * Time)
        12.1880886           -2.9852974          -0.8777618

Random effects:
Formula: ~sin(2 * pi * Time) | Mare
Structure: Diagonal
        (Intercept) sin(2 * pi * Time) Residual
StdDev:      2.858385           1.257977 3.507053

Correlation Structure: AR(1)
Formula: ~1 | Mare
Parameter estimate(s):
    Phi
0.5721866
Number of Observations: 308
Number of Groups: 11
```

oooooooooooooooooooooooooooo

Example: counts of ovarian follicles

```
> intervals(fm4Ovar.lme)
Approximate 95% confidence intervals
Fixed effects:
      lower      est.      upper
(Intercept) 10.330929 12.1880886 14.04524774
sin(2 * pi * Time) -4.177135 -2.9852974 -1.79345990
cos(2 * pi * Time) -1.818055 -0.8777618  0.06253155
attr(,"label")
[1] ``Fixed effects:''

Random Effects:
Level: Mare
      lower      est.      upper
sd((Intercept)) 1.6994209 2.858385 4.807734
sd(sin(2 * pi * Time)) 0.3676751 1.257977 4.304086

Correlation structure:
      lower      est.      upper
Phi 0.4304938 0.5721866 0.6863558
attr(,"label")
[1] ``Correlation structure:''

Within-group standard error:
      lower      est.      upper
3.020523 3.507053 4.071952
```

oooooooooooooooooooooooooooo ooooooooooooooooooooooooooooo ooooooooooooo

Example: counts of ovarian follicles

(c) Try other ARMA structures

```
# try MA(2)
> fm5Ovar.lme <- update(fm3Ovar.lme, correlation=corARMA(q=2))

# try CAR(1)
> fm6Ovar.lme <- update(fm3Ovar.lme, correlation=corCAR1(form=~Time))

# try ARMA(1,1)
> fm7Ovar.lme <- update(fm3Ovar.lme,
+ correlation=corARMA(value=c(.78716,-.27957),p=1,q=1))
> anova(fm4Ovar.lme,fm5Ovar.lme,fm6Ovar.lme,fm7Ovar.lme, test=F)

          Model   df      AIC      BIC    logLik
fm4Ovar.lme     1    7 1563.448 1589.490 -774.7240
fm5Ovar.lme     2    8 1571.231 1600.993 -777.6154
fm6Ovar.lme     3    7 1565.535 1591.577 -775.7676
fm7Ovar.lme     4    8 1559.894 1589.657 -771.9471

# plot ACF of normalized residuals
> plot(ACF(fm7Ovar.lme, maxLag=10, resType='n'), alpha=.01)
```

Empirical autocorrelation function

