

## Outline

### Statistical inference for linear mixed models

Rasmus Waagepetersen  
Department of Mathematics  
Aalborg University  
Denmark

October 2, 2019

- ▶ general form of linear mixed models
- ▶ examples of analyses using linear mixed models
- ▶ prediction of random effects
- ▶ estimation (including restricted maximum likelihood estimation)

1/1

2/1

### One-way ANOVA in matrix-vector form

One observation:

$$Y_{ij} = \mu + U_i + \epsilon_{ij}$$

Vector of observations

$$Y = \mu \mathbf{1}_n + ZU + \epsilon$$

where  $Y$ ,  $U$  and  $\epsilon$  vectors of  $Y_{ij}$ 's,  $U_i$ 's and  $\epsilon_{ij}$ 's.  $\mathbf{1}_n$  vector of 1's and  $Z$   $n \times k$  matrix with  $Z_{(ij)q} = 1$  if  $q = i$  and zero otherwise.

### Linear regression with random effects in matrix-vector form

Consider mixed model:

$$Y_{ij} = \beta_1 + U_i + [\beta_2 + V_i]x_{ij} + \epsilon_{ij}$$

May be written in matrix vector form as

$$Y = X\beta + ZU + \epsilon$$

where  $\beta = (\beta_1, \beta_2)^T$ ,  $U = (U_1, \dots, U_k, V_1, \dots, V_k)^T$ ,  $\epsilon = (\epsilon_{11}, \epsilon_{12}, \dots, \epsilon_{km})^T$ ,  $X$  is  $n \times 2$  and  $Z$  is  $n \times 2k$ .

3/1

4/1

## Linear mixed model: general form

Consider model

$$Y = X\beta + ZU + \epsilon$$

where  $U \sim N(0, \Psi)$  and  $\epsilon \sim N(0, \Sigma)$  are independent.

All previous models special cases of this.

Then  $Y$  has multivariate normal distribution

$$Y \sim N(X\beta, Z\Psi Z^T + \Sigma)$$

5/1

## Linear mixed models using lmer

General lmer model formulation

```
y ~ 'fixed formula' + ('rand formula_1' | Group_1) + ...
      + ('rand. formula_n' | Group_n)
```

translates into linear mixed model with independent sets of random effects for each grouping variable and e.g.

(z | Group\_i)

corresponds to

$$U_{ij} + V_{ij}z$$

i.e. model with random intercept and random slope for covariate  $z$  within each level  $l$  of grouping factor  $\text{Group}_i$ .

NB independence between levels of  $\text{Group}_i$  but intercept and slope dependent within level.

Only random intercept respectively slope: (1 |  $\text{Group}_i$ ) resp. (-1 + z |  $\text{Group}_i$ )

7/1

## Hierarchical version

1.  $U \sim N(0, \Psi)$
2.  $Y | U = u \sim N(X\beta + Zu, \Sigma)$

Crucial assumption:  $Y$  normal given  $U$  !

Often  $\Sigma = \sigma^2 I$  so further assumption is individual observations  $Y_i$  independent given  $U$ .

Hierarchical version useful for extension to binary and count data.

6/1

## Linear mixed model for orthodont data - independent random slope and intercept

```
> ort6=lmer(distance~age*Sex+(1|Subject)+(-1+age|Subject))
> summary(ort6)
```

Linear mixed model fit by REML

Formula: distance ~ age \* Sex + (1 | Subject) + (-1 + age |

	AIC	BIC	logLik	deviance	REMLdev
	447.2	465.9	-216.6	428.1	433.2

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	2.4168043	1.554607
Subject	age	0.0077469	0.088017
Residual		1.8645950	1.365502

Number of obs: 108, groups: Subject, 27

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	16.34062	0.94087	17.368
age	0.70100	0.07011	9.999
Sex	0.07011	0.07011	1.000

8/1

## Linear mixed model for orthodont data - correlated random slope and intercept

```
> ort7=lmer(distance~age*Sex+(age|Subject))
> summary(ort7)
Linear mixed model fit by REML
Formula: distance ~ age * Sex + (age | Subject)
      AIC BIC logLik deviance REMLdev
448.6 470 -216.3   427.9   432.6
Random effects:
Groups   Name      Variance Std.Dev. Corr
Subject (Intercept) 5.786427 2.40550
          age       0.032524 0.18035 -0.668
Residual                1.716205 1.31004
Number of obs: 108, groups: Subject, 27

Fixed effects:
              Estimate Std. Error t value
(Intercept)    16.3406     1.0185  16.043
age             0.7044     0.0000   0.101
Sex             0.0000     0.0000   0.000
```

9/1

## SPSS

Choose Analyze → Mixed models → Linear.

Need to specify 'Subject' variables - these correspond to the grouping variables for lmer.

With SPSS one can choose to model correlation in residuals ( $\Sigma \neq \sigma^2 I$ ) - then one also need to specify a 'Repeated' variable (e.g. residuals for each subject may be correlated in time).

Specify fixed part of model using item 'fixed' and random part using item 'random' in menu.

Under random: several sets of random effects can be specified (corresponding to several (1) in R).

11/1

## Comparison of models for orthodont data

Fixed part: age+Sex+age:sex

Random part:

Model	AIC	BIC	logLik	Number of parameters
$a$	445.8	461.9	-216.9	4+2
$bx$	448.7	464.8	-218.4	4+2
$a + bx, \text{Cov}(a, b) = 0$	447.2	465.9	-216.6	4+3
$a + bx$	448.6	470	-216.3	4+4

Larger logLik and smaller AIC or BIC means better model.

AIC and BIC takes into account number of parameters - penalizes complex models

The simplest one (just random intercept) seems better.

When REML (see last slide) is used (is default) for parameter estimation, need same mean structure in the models compared.

10/1

## SPSS - continued

Under random: various options for covariance matrix of random effects within subject. Use covariance structure 'Variance Components' to get independent random effects or 'unstructured' to get dependent random effects.

Remember to include intercept.

Output: Type III F-tests for fixed effects.

See also power-point slides regarding SPSS.

12/1

## Tests for fixed effects

With lmer: to test the effect of a covariate or a factor you fit models with and without the covariate (or factor) and compute F-test using procedure KRmodcomp from pbkrtest package:

```
> ort4=lmer(distance~age*Sex+(1|Subject))
> ort4.1=lmer(distance~age+Sex+(1|Subject))#remove interaction
> library(pbkrtest)
> KRmodcomp(ort4,ort4.1)
F-test with Kenward-Roger approximation; computing time: 0.17 se
large : distance ~ age * Sex + (1 | Subject)
small : distance ~ age + Sex + (1 | Subject)
      stat      ndf      ddf F.scaling p.value
Ftest 6.3027  1.0000 79.0000         1 0.0141 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: interaction now significant at 5% level.

SPSS: F-tests available in output.

13/1

## Nested two-way analysis of variance

For five cardboards we have 4 replications at 4 positions.

Hierarchical model (nested random effects)

$$Y_{ipj} = \mu + U_i + U_{ip} + \epsilon_{ipj}$$

$$\text{Var } Y_{ipj} = \tau^2 + \omega^2 + \sigma^2$$

14/1

## Covariance structure for nested random effects model

$$Y_{ipj} = \mu + U_i + U_{ip} + \epsilon_{ipj}$$

$$\text{Cov}(Y_{ipj}, Y_{lqk}) = \begin{cases} 0 & i \neq l \\ \tau^2 & i = l, p \neq q \text{ same card} \\ \tau^2 + \omega^2 & i = l, p = q \text{ same card and pos.} \\ \tau^2 + \omega^2 + \sigma^2 & i = l, p = q, k = j \quad (\text{Var } Y_{ipj}) \end{cases}$$

15/1

## Nested two-way analysis of variance

```
> out2=lmer(Reflektans~(1|Pap.nr.)+(1|Pap.nr.*Sted))
> summary(out2)
Linear mixed model fit by REML
  AIC   BIC logLik deviance REMLdev
-432 -422.4  220  -443.8  -440
Random effects:
Groups      Name      Variance  Std.Dev.
Pap.nr.      (Intercept) 1.6560e-02 0.1286843
Pap.nr. * Sted (Intercept) 9.4539e-04 0.0307472
Residual                                6.3494e-05 0.0079683
Number of obs: 80, groups: Pap.nr. * Sted, 20; Pap.nr., 5
```

Largest part of variance is between cardboard variance !

16/1

Explanation of  $\text{Reflektans} \sim (1|\text{Pap.nr.}) + (1|\text{Pap.nr.} * \text{Sted})$ :

- ▶ no fixed formula: intercept always included as default
- ▶  $(1|\text{Pap.nr.})$  random intercepts for groups identified by variable  $\text{Pap.nr.}$  (card board effects)
- ▶  $(1|\text{Pap.nr.} * \text{Sted})$  random intercepts for groups identified by cross of variables  $\text{Pap.nr.}$  and  $\text{Sted}$  (positions within cardboard)
- ▶ random effects specified by different terms independent.

17 / 1

Classical anova table:

```
> fit1=lm(intensity~treat*factor(exon)+factor(patient)+
          factor(patient):treat,data=genel)
```

```
> anova(fit1)
```

Analysis of Variance Table

Response: intensity

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treat	1	3.242	3.242	14.4796	0.0002199
factor(exon)	7	254.343	36.335	162.2852	< 2.2e-16
factor(patient)	9	15.405	1.712	7.6449	6.703e-09
treat:factor(exon)	7	2.238	0.320	1.4278	0.1998234
treat:factor(patient)	9	8.190	0.910	4.0643	0.0001345
Residuals	126	28.211	0.224		

$$\hat{\sigma}^2 = 0.224 \quad \hat{\sigma}_{P \times T}^2 = (0.91 - 0.224)/8 = 0.08575$$

$$\hat{\sigma}_P^2 = (1.712 - 0.91)/16 = 0.050125$$

19 / 1

## A more complicated example: gene-expression

Gene (DNA string) composed of substrings (exons) which may be more or less expressed according to treatment.

Expression measured as intensities on micro-array (chip). One chip pr. patient-treatment.

Factors: E (exon 8 levels), P (patient, 10 levels), T (treatment, 2 levels)

Y: vector of intensities (how much is exon expressed).

Model:

$$y_{ept} = \mu + \alpha_e + \beta_t + \gamma_{et} + U_p + U_{pt} + \epsilon_{ept}$$

$U_{pt}$  and  $U_p$  random chip and patient effects.

Main question: are exons differentially expressed - i.e. are  $\gamma_{et} \neq 0$  or not ?

18 / 1

F-test for no treatment-exon interaction: 1.4278 with  $p$ -value 0.1998.

I.e. interaction not significant - no evidence of differential exon usage.

Classical ANOVA:

- ▶ not straightforward to obtain estimates of variance components from table of sums of squares (I will not go into detail with this).
- ▶ in the presence of random effects not straightforward to compute F-tests for fixed effects (which sums of squares should be used ?)
- ▶ exact F-tests (only) available in balanced case (equal number of observations for each combination of factor levels)

20 / 1

Using lmer:

```
> fit1=lmer(intensity~treatment*factor(exon)+(1|patient)
+ (1|factor(patient):treatment),data=g
> summary(fit1)
  AIC   BIC logLik deviance REMLdev
298.9 357.3 -130.5   232.1   260.9
Random effects:
Groups                Name          Variance Std.Dev.
factor(patient):treatment (Intercept) 0.085762 0.29285
patient                (Intercept) 0.050100 0.22383
Residual                                0.223894 0.47317
Number of obs: 160, groups:
      factor(patient):treatment, 20; patient, 10
```

We directly obtain estimates of variance components.

21/1

## With 12.5% missing data

20 of out 160 missing at random.

```
Random effects:
Groups                Name          Variance Std.Dev.
factor(patient):treatment (Intercept) 0.055072 0.23467
patient                (Intercept) 0.074383 0.27273
Residual                                0.226591 0.47602
Number of obs: 140, groups: factor(patient):treatment, 20;
patient, 10
```

23/1

## Tests of fixed effects

Test for no treatment-exon interaction:

```
> fit1=lmer(intensity~treatment*factor(exon)+
(1|patient)+(1|factor(patient):treatment),data=g
> fit2=lmer(intensity~treatment+factor(exon)+
(1|patient)+(1|factor(patient):treatment),data=
> KRmodcomp(fit1,fit2)
F-test with Kenward-Roger approximation; computing time: 0.
large : intensity ~ treatment * factor(exon) + (1 | patient
(1 | factor(patient):treatment)
small : intensity ~ treatment + factor(exon) + (1 | patient
(1 | factor(patient):treatment)
      stat      ndf      ddf F.scaling p.value
Ftest  1.4278    7.0000 126.0000      1 0.1998
```

Treatment-exon interaction not significant !

22/1

## Adjusted F-test

```
> KRmodcomp(fit1,fit2)
F-test with Kenward-Roger approximation; computing time: 0.
large : intensity ~ treatment * factor(exon) + (1 | patient
(1 | factor(patient):treatment)
small : intensity ~ treatment + factor(exon) + (1 | patient
(1 | factor(patient):treatment)
      stat      ndf      ddf F.scaling p.value
Ftest  1.4725    7.0000 107.3559    0.99999 0.1847
```

anova says 106 denominator df but this assumes balanced design.

KRmodcomp adjusts df to 107.35 to account for unbalanced data.

24/1

## Predictions/Residuals

The random effects  $U$  in a linear mixed model can be predicted using 'best linear unbiased prediction' (BLUP) - useful if we want to look at subject specific characteristics.

In the context of linear mixed models, BLUP  $\hat{U}$  is the conditional mean of the random effects given the data:

$$\hat{U} = \mathbb{E}[U|Y = y]$$

Typically we assume  $\epsilon_{ij}$  independent and  $N(0, \sigma^2)$ . To check this we can consider residuals:

$$\hat{\epsilon} = Y - X\hat{\beta} - Z\hat{U}$$

and perform the usual residual diagnostics.

With lmer: use `ranef`, `fitted` and `residuals` to extract BLUPS, fitted values and residuals.

SPSS: save predicted values and residuals under 'Predicted values and residuals'

25 / 1

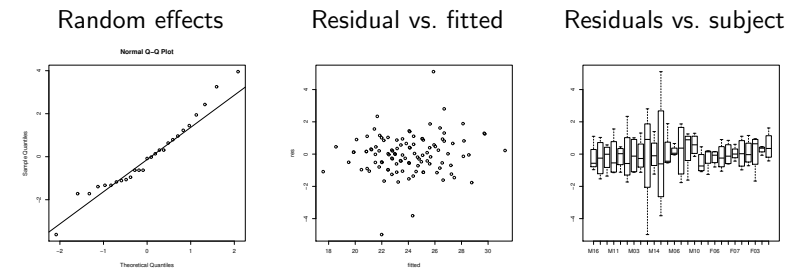
26 / 1

## Example: orthodont data

Extract BLUPS, fitted values and residuals

```
> childeffects=ranef(ort4)$Subject
> qqnorm(childeffects[[1]])
> qqline(childeffects[[1]])
> res=resid(ort4)
> fitted=fitted(ort4)
> plot(fitted,res)
> boxplot(res~Subject)
```

## Plots



Outliers for two subjects !

27 / 1

28 / 1

## Estimation

For linear mixed model two sets of parameters:  $\beta$  (fixed effects) and  $\psi$  (random effects variances).

Maximum likelihood estimation: parameter estimates are those parameter values that make data most likely under the given model:

$$(\hat{\beta}, \hat{\psi}) = \underset{\beta, \psi}{\operatorname{argmax}} f(y; \beta, \psi)$$

where  $f(y; \beta, \psi)$  is the normal probability density of the data  $y$ .

Given  $\psi$ ,  $\hat{\beta}$  is the generalized least squares estimate:

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

which minimizes the generalized sum of squares

$$(y - X\beta)^T V(\psi)^{-1} (y - X\beta).$$

29 / 1

In general  $\psi$  needs to be obtained by iterative maximization of

$$L(\psi) = f(y; \hat{\beta}(\psi), \psi)$$

One issue: MLE of  $\psi$  in general biased.

30 / 1

## MLE's of variances biased

Consider simple normal sample  $Y_i \sim N(\mu, \sigma^2)$ .

MLE's:

$$\hat{\mu} = \bar{Y}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Bias of  $\hat{\sigma}^2$ :

$$\mathbb{E}\hat{\sigma}^2 = \sigma^2(n-1)/n$$

Bias arise from estimation of  $\mu$  ( $\sum_i (Y_i - \mu)^2$  vs  $\sum_{i=1}^n (Y_i - \bar{Y})^2$ ).

Often we use instead unbiased estimate

$$s^2 = \frac{1}{n-1} \sum_i (Y_i - \bar{Y})^2$$

Similarly: maximum likelihood estimate of between subject variance in one-way anova is biased due to estimation of mean.

31 / 1

## REML (restricted/residual maximum likelihood)

Idea: use linear transform of data which eliminates mean. Suppose design matrix  $X : n \times p$  and let  $A : n \times (n-p)$  have columns spanning the orthogonal complement  $L^\perp$  of  $L = \operatorname{span} X$ . Then  $A^T X = 0$ .

Transformed data  $((n-p) \times 1)$

$$\tilde{Y} = A^T Y = A^T ZU + A^T \epsilon$$

has mean 0 and covariance matrix  $A^T V(\psi) A$  where  $V = Z\Psi Z^T + \Sigma$  covariance matrix of  $Y$  and  $\psi$  covariance parameters. Then proceed as for MLE.

Default choice for estimation of variance parameters in both `lmer` and Mixed model in SPSS.

$s^2$  is one example of REML. Classical ANOVA variance estimates also REML.

32 / 1



## Summary

- ▶ Linear mixed models flexible class of models for continuous observations.
- ▶ incorporates classical ANOVA models and random coefficients models
- ▶ Useful for modeling of correlated observations, for decomposition of variance and for estimation of population variances.
- ▶ Userfriendly software available

33 / 1

## Exercises

1. Use `lmer` or Mixed models in SPSS to fit a one-way ANOVA model with random operator effects for the pulp data. Compare with results from previous exercise (classical anova for pulp data).
2. Install the R-package `faraway` which contains the data set `penicillin`. The response variable is yield of penicillin for four different production processes (the 'treatment'). The raw material for the production comes in batches ('blends'). The four production processes were applied to each of the 5 blends. Fit anova models with production process as a fixed factor and blend as random factor. Compute an F-test for the effect of production process.

34 / 1

3. The rats data has variables (1) `obs`: observation number (2) `treat`: treatment group ('con': control; 'hig': high dose; 'low': low dose) (3) `rat`: rat identification number (4) `age`: age of the rat at the moment the observation is made (5) `respon`: the response measured (height of skull) (6) `logage`: log-transformed age.

The treatment is a drug that inhibits production of testosterone. The scientific question is whether/how the drug affects the growth rate of the rats.

- 3.1 take a look at data by plotting response against age and `logage` (with separate curves for each rat).
- 3.2 fit a linear regression model for the response with `logage` as the independent variable and an interaction between `logage` and treatment. Is the interaction between `logage` and treatment significant? Is treatment significant?

35 / 1

3. (continued) fit a linear mixed model by extending the previous models with random rat specific intercepts.
  - 3.3 what is the proportion of variance explained by the random intercepts?
  - 3.4 What are the conclusions regarding interaction and treatment effects based on this model? Compare with the previous model.
  - 3.5 Check the fitted linear mixed model using residuals and predicted random effects.

36 / 1

4. Write out  $X$  and  $Z$  matrix for model on slide 4.
5. (slide 8, independent random intercepts and slopes) What are the proportions of variance of an observation due to respectively random slopes and random intercepts when age=8,10,12 or 14 ?
6. (slide 9, correlated random intercepts and slopes) What are the proportions of variance of an observation due to respectively random slopes and random intercepts when age=8,10,12 or 14 ?