

Outline for today

To serve and project. A geometric approach to balanced mixed models.

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

February 20, 2020

- ▶ Fitting linear mixed models using R
- ▶ One-way ANOVA using orthogonal projections

1 / 31

2 / 31

Rep: specification of linear models in R

$$\begin{array}{ll} y = \alpha + \beta x + \epsilon & \tilde{y} \sim x \\ y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} & \tilde{y} \sim A+B \\ y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} & \tilde{y} \sim A+B+A:B \\ & \tilde{y} \sim A*B \\ y_{ij} = \mu + \alpha_i + \beta_i x_{ij} + \epsilon_{ij} & \tilde{y} \sim A+A*x \\ & \text{etc.} \quad \dots \end{array}$$

NB: A is a factor/categorical variable identifying groups of observations. The i th group of observations is assigned the parameter value α_i .

NB: replace A with `factor(A)` if A not already declared a factor.

Hierarchical principle

$\tilde{y} \sim A+B+A:B$ $\tilde{y} \sim A+A:B$ $\tilde{y} \sim B+A:B$ and $\tilde{y} \sim A:B$ all fit the same model for the mean vector.

I.e. in presence of interaction A:B it does not make sense to attempt to omit main effects A or B.

If you really want to 'remove' main effects then you need to mess with design matrix - and results depend crucially on choice of parametrization constraints.

3 / 31

4 / 31

Rep: Multiple linear regression in R I

```
#fit model with sex specific intercepts and slopes
> ort1=lm(distance~age+age:factor(Sex)+factor(Sex))
> summary(ort1)
...
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    16.3406     1.4162  11.538 < 2e-16 ***
age              0.7844     0.1262   6.217 1.07e-08 ***
factor(Sex)Female  1.0321     2.2188   0.465  0.643
age:factor(Sex)Female -0.3048     0.1977  -1.542  0.126
...
Residual standard error: 2.257 on 104 degrees of freedom
> #compute F-tests respecting hierarchical principle
> drop1(ort1,test="F")
Single term deletions

...
              Df Sum of Sq  RSS   AIC F value    Pr(F)
<none>                529.76 179.75
age:factor(Sex)  1      12.11 541.87 180.19  2.3782 0.1261
age:Sex not significant !
```

5 / 31

Multiple linear regression in R II

```
> ort2=lm(distance~age+factor(Sex))
> drop1(ort2,test="F")
Single term deletions

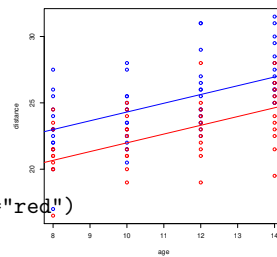
Model:
distance ~ age + factor(Sex)
              Df Sum of Sq  RSS   AIC F value    Pr(F)
<none>                541.87 180.19
age                  1    235.36 777.23 217.15  45.606 8.253e-10 **
factor(Sex)         1     140.46 682.34 203.09  27.218 9.198e-07 **

both age and sex significant
```

6 / 31

Multiple linear regression in R III

```
#plot data and two regression lines
col=rep("blue",length(Sex))
col[Sex=="Female"]="red"
plot(distance~age,col=col)
abline(parm[1:2],col="blue")
abline(c(parm[1]+parm[3],parm[2]),col="red")
```

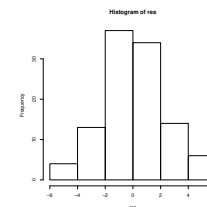


7 / 31

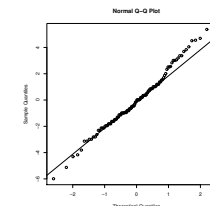
Multiple linear regression in R IV

```
res=residuals(ort2)
```

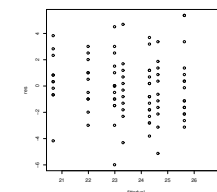
```
hist(res)
```



```
qqnorm(res)
qqline(res)
```



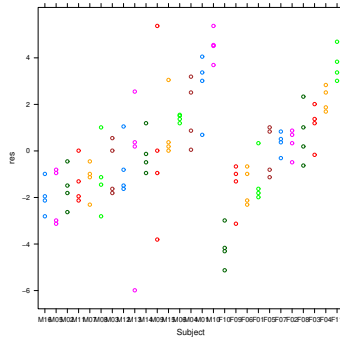
```
fittedval=fitted(ort)
plot(res~fittedval)
```



8 / 31

Multiple linear regression in R V

```
> library(lattice)
> xyplot(res~Subject, groups=Subject)
```



Oups - residuals not independent and identically distributed ! Hence computed F -tests not valid.

Problem: subject specific intercepts (and possibly subject specific slopes too)

9 / 31

Fitting linear mixed models in R

General procedures for linear mixed models: `lme()` from the `nlme` package and `lmer()` from the `lme4` package.

Quote from internet (Ben Bolker):

“lmer is newer, much faster, handles crossed random effects well (and generalized linear mixed models), has some support for producing likelihood profiles (in the development version), and is under rapid development. It does not attempt to estimate residual degrees of freedom and hence does not give p -values for significance of effects. lme is older, better documented (Pinheiro and Bates 2000), more stable, and handles ‘R-side’ structures (heteroscedasticity, within-group correlations)”

I will mainly use `lmer()` in this course: specification of model for random effects fairly straightforward. `lme()` is covered in M&T Chapter 5.11. Package `lmerTest` adds p -values to output of `lmer()`

10 / 31

Linear mixed models using lmer

General `lmer` model formulation

```
y~‘fixed formula’+(‘rand formula_1’|Group_1)+ ...
      +(‘rand. formula_n’|Group_K)
```

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

$(z|Group_1)$

corresponds to

$$U_i + V_i z$$

i.e. model with random intercept and random slope for covariate z within each level i of grouping factor `Group_1`.

NB independence between levels of `Group_1` but intercept and slope dependent within level.

Only random intercept respectively slope: $(1|Group_1)$ resp. $(-1+z|Group_1)$

11 / 31

Orthodont with random subject intercepts

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

```
Random effects:
Groups Name      Variance Std.Dev.
Subject (Intercept) 3.299    1.816
Residual          1.922    1.386
Number of obs: 108, groups: Subject, 27

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  16.3406    0.9813 103.9864  16.652 < 2e-16 ***
age           0.7844    0.0775  79.0000  10.121 6.44e-16 ***
SexFemale    1.0321    1.5374 103.9864   0.671  0.5035
age:SexFemale -0.3048    0.1214  79.0000  -2.511  0.0141 *
```

Now interaction significant ($p=0.0141$) assuming t -value approximately standard normal.

What is interpretation of interaction ? Does it make sense ?

12 / 31

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

Note: corresponding model without random effects has much inflated residual variance $5.09 = 2.257^2$ vs. 1.922 for mixed model.

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

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	2.416451	1.55449
Subject.1	age	0.007748	0.08802
Residual		1.864634	1.36552

Number of obs: 108, groups: Subject, 27

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	16.34062	0.94087	67.09150	17.368	< 2e-16 ***
age	0.78438	0.07944	67.09021	9.873	1.06e-14 ***
SexFemale	1.03210	1.47405	67.09150	0.700	0.4862
age:SexFemale	-0.30483	0.12446	67.09021	-2.449	0.0169 *

13 / 31

14 / 31

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

Formula: `distance ~ age * Sex + (age | Subject)`

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	5.77441	2.4030	
	age	0.03245	0.1801	-0.67
Residual		1.71661	1.3102	

Number of obs: 108, groups: Subject, 27

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	16.34063	1.01824	25.00829	16.048	1.12e-14 ***
age	0.78437	0.08598	25.01351	9.123	1.97e-09 ***
SexFemale	1.03210	1.59528	25.00829	0.647	0.5235
age:SexFemale	-0.30483	0.13471	25.01351	-2.263	0.0326 *

15 / 31

Comparison of models for orthodont data

Fixed part: `age+Sex+age:sex`

Random part:

Model	AIC	BIC	logLik	Number of parameters
U	445.8	461.9	-216.9	4+2
V_x	448.7	464.8	-218.4	4+2
$U + V_x, \text{Cov}(U, V) = 0$	447.2	465.9	-216.6	4+3
$U + V_x$	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 is used (is default) for parameter estimation, **need same mean structure** in the models compared.

16 / 31

Random intercepts with MLE

```
ort35=lmer(distance~age*Sex+(1|Subject),data=Orthodont,REML=F)
Formula: distance ~ age * Sex + (1 | Subject)
```

```
Random effects:
Groups Name Variance Std.Dev.
Subject (Intercept) 3.030 1.741
Residual 1.875 1.369
Number of obs: 108, groups: Subject, 27
```

```
Fixed effects:
Estimate Std. Error df t value Pr(>|t|)
(Intercept) 16.34062 0.96309 107.88346 16.967 < 2e-16 ***
age 0.78438 0.07654 80.99936 10.248 2.77e-16 ***
SexFemale 1.03210 1.50886 107.88346 0.684 0.4954
age:SexFemale -0.30483 0.11991 80.99936 -2.542 0.0129 *
```

Slightly different variance estimates. Fixed effects estimates in this case same as REML (since balanced dataset)

Due to nice data structure same fixed effects estimates for all covariance structures !

17 / 31

One-way anova

Let F be a factor with k levels and consider the model

$$y_{ij} = \xi + U_i + \epsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i$$

or

$$y = \xi \mathbf{1}_n + Z_F U + \epsilon$$

where n is total number of observations and Z_F is the design matrix corresponding to F : ij , q th entry of Z_F is 1 if y_{ij} belongs to the q th group and zero otherwise.

F is balanced if common number $n_i = m$ of observations at each of the k levels (whereby $n = mk$).

In this case, P_F (orthogonal projection on $L_F = \text{span} Z_F$) is

$$P_F = \frac{1}{m} Z_F Z_F^T$$

Action of P_F : replaces y_{ij} by \bar{y}_i . (averages within each group).

19 / 31

Analysis of variance

Analysis of variance (ANOVA) models are specified in terms of grouping variables or factors.

A factor F is a variable that assigns a grouping label to each observation.

E.g. $F_i = q$ means that observation y_i (or index i) is assigned to group/level q for the factor F .

Suppose F generates k groups. The design matrix Z_F corresponding to F is $n \times k$ and the iq th entry of Z_F is 1 if i is assigned to group q and 0 otherwise.

(NB: i could be a multi-index $i = i_1 i_2 \dots i_p$)

18 / 31

Two special factors: unit factor I has a unique level for each observation $L_I = \mathbb{R}^n$ and $P_I = I$ (with an abuse of notation I is used both for factor and identity matrix). Factor 0 assigns all observations to the same group and $L_0 = \text{span}(\mathbf{1}_n)$, $P_0 = \mathbf{1}_n \mathbf{1}_n^T / n$.

Then $L_0 \subset L_F \subset L_I$.

Orthogonal decomposition of \mathbb{R}^n :

$$\mathbb{R}^n = V_0 \oplus V_F \oplus V_I$$

where $V_0 = L_0 = \text{span}(\mathbf{1}_n)$, $V_F = L_F \ominus V_0$ and $V_I = \mathbb{R}^n \ominus L_F$. Dimensions of V_0 , V_F and V_I are 1, $k - 1$ and $n - k$.

Sum of orthogonal subspaces: $L_1 \oplus L_2 = \{x + y | x \in L_1, y \in L_2\}$ for orthogonal subspaces L_1 and L_2 .

Orthogonal complement: $L_2 \ominus L_1 = \{x \in L_2 | x^T y = 0 \forall y \in L_1\}$ for $L_1 \subseteq L_2$.

Orthogonal projections on V_0 , V_F and V_I are $Q_0 = P_0$, $Q_F = P_F - P_0$ and $Q_I = I - P_F$.

20 / 31

Decomposition of covariance matrix into orthogonal projections:

$$\text{Cov } Y = m\tau^2 P_F + \sigma^2 I = \lambda P_F + \sigma^2 Q_I$$

where $\lambda = m\tau^2 + \sigma^2$ (recall $I = P_F + Q_I$).

Note: one-to-one correspondence between (λ, σ^2) and (τ^2, σ^2) .

Orthogonal decomposition of data vector:

$$Y = P_F Y + Q_I Y$$

Note $P_F 1_n = Q_I 1_n = 1_n$ and $Q_I 1_n = 0$. Moreover $Q_I P_F = 0$.

Hence

$$\begin{bmatrix} P_F \\ Q_I \end{bmatrix} Y \sim N \left(\begin{pmatrix} 1_n \xi \\ 0_n \end{pmatrix}, \begin{bmatrix} \lambda P_F & 0 \\ 0 & \sigma^2 Q_I \end{bmatrix} \right)$$

We can thus base maximum likelihood estimation of (ξ, λ) on $P_F Y$ and σ^2 on $Q_I Y$.

21 / 31

22 / 31

More precisely,

$$|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(Y - 1_n \xi)^T \Sigma^{-1}(Y - 1_n \xi)\right) = \lambda^{-k/2} \exp\left(-\frac{1}{2\lambda} \|P_F Y - 1_n \xi\|^2\right) \times (\sigma^2)^{-k(m-1)/2} \exp\left(-\frac{1}{2\sigma^2} \|Q_I Y\|^2\right)$$

$$(\Sigma^{-1} = \sigma^{-2} Q_I + \lambda^{-1} P_F \text{ and } |\Sigma| = \lambda^k (\sigma^2)^{m-k})$$

Note: the two factors in the above likelihood are 'generalized' densities of the 'degenerate' normal vectors $P_F Y$ and $Q_I Y$.

Consider e.g. the factor $\lambda^{-k/2} \exp\left(-\frac{1}{2\lambda} \|P_F Y - 1_n \xi\|^2\right)$ involving the parameters λ and ξ . We can maximize this with respect to λ and ξ in exactly the same way as when we previously considered the likelihood of $N_n(X\beta, \tau^2 I)$ (see slide 'Estimation using orthogonal projections' in second set of handouts). Thus we obtain

$$\widehat{1_n \xi} = P_0 P_F Y = P_0 Y \text{ and } \widehat{\lambda} = \|P_F Y - P_0 Y\|^2 / k = SSB / k$$

Proceeding in the same way for the second factor (where there is no mean parameter), we obtain

$$\widehat{\sigma}^2 = \|Q_I Y\|^2 / (k(m-1)) = SSE / (k(m-1))$$

Note $Q_F Y \sim N_n(0, \lambda Q_F)$. By Exercise 9, $\|P_F Y - P_0 Y\|^2 = \|Q_F Y\|^2 \sim \lambda \chi^2(k-1)$ which has mean $\lambda(k-1)$. Thus $\widehat{\lambda}$ is biased.

Unbiased estimate: $\tilde{\lambda} = \|P_F Y - P_0 Y\|^2 / (k-1) = SSB / (k-1)$ (REML)

23 / 31

24 / 31

Implementation in R

For cardboard/reflectance data, $k = 34$ and $m = 4$. `anova()` procedure produces table of sums of squares corresponding to orthogonal decomposition.

```
> anova(lm(Reflektans~factor(Pap.nr.)))
Analysis of Variance Table

Response: Reflektans
          Df Sum Sq Mean Sq F value #SSB V_F
factor(Pap.nr) 33  0.9009 0.0273  470.7 #SSB V_F
Residuals      102 0.0059 0.00006      #SSE V_I
---
```

Hence $\hat{\sigma}^2 = 0.00006$, $\hat{\lambda} = 0.90088/34$ (or $\tilde{\lambda} = 0.90088/33 = 0.0273$) and $\hat{\tau}^2 = (\hat{\lambda} - 0.00006)/4 = 0.00661$ (or $\tilde{\tau}^2 = (\tilde{\lambda} - 0.00006)/4 = 0.00681$).

Biggest part of variation is between cardboard.

25 / 31

Implementation using `lmer()`

One-way anova is special case of linear mixed model so we can use `lmer()`:

```
> out1=lmer(Reflektans~(1|Pap.nr.),REML=F)
> summary(out1)
Linear mixed model fit by maximum likelihood
Formula: Reflektans ~ (1 | Pap.nr.)
      AIC      BIC logLik deviance REMLdev
-726.5 -717.8  366.3  -732.5  -725.8
Random effects:
Groups Name      Variance Std.Dev.
Pap.nr. (Intercept) 6.6096e-03 0.0812994
Residual              5.7997e-05 0.0076156
Number of obs: 136, groups: Pap.nr., 34
```

26 / 31

REML results with `lmer()`

```
> out1=lmer(Reflektans~(1|Pap.nr.))#default is REML
> summary(out1)
...
Random effects:
Groups Name      Variance Std.Dev.
Pap.nr. (Intercept) 6.8103e-03 0.0825247
Residual              5.7997e-05 0.0076156
number of obs: 136, groups: Pap.nr., 34

Fixed effects:
          Estimate Std. Error t value
(Intercept) 0.61690    0.01417  43.54
```

27 / 31

`anova(lm())` vs `lmer()`

`lmer()`: easy model specification in accordance with specification as general linear mixed model. Directly provides estimates of original variance parameters. Balanced factor not required.

`anova(lm())`: classical method. Computationally very efficient (no need for numerical optimization). Requires some skill/care to obtain original variance estimates. Balanced factor required.

28 / 31

Exercises

1. fit a linear model to the orthodontic data with covariates age, sex and Subject. What happens ?
2. Consider the mixed model for the orthodontic data with uncorrelated random intercept and random slope for each child. What are the proportions of variances due to respectively noise, random intercepts, and random slopes ? How do the results depend on age ?
3. Let $L_1 \subseteq L_2$ with orthogonal projections P_1 and P_2 . Show that $P_2 - P_1$ is the orthogonal projection on $L_2 \ominus L_1$.
4. Show that an orthogonal projection only has eigen values 1 or 0.
5. For a symmetric matrix A show that $|A|$ is the product of A 's eigen values.
6. Show $Q_I P_F = 0$
7. Let $S = aP + bQ$ where P and Q are orthogonal projections with $P + Q = I$ and $a, b \neq 0$. Show that the eigen values of S are the non-zero eigen values a and b of aP and bQ . Show that $S^{-1} = a^{-1}P + b^{-1}Q$.
8. Show that $\|Y\|^2 \sim \sigma^2 \chi^2(d)$ if $Y \sim N(0, \sigma^2 P)$ and P is an orthogonal projection on a subspace of dimension d (hint: use spectral decomposition and the result above regarding the eigen values of P).
9. Install the R-package `faraway` which contains the data set `pulp` (brightness of paper pulp in groups given by different operators). Analyze the data using a one-way anova with random operator effects. Estimate variance components and the intra-class correlation. Try both `lmer` and `anova`.