

Introduction to General and Generalized Linear Models

Mixed effects models - I

Henrik Madsen
Poul Thyregod
Anders Nielsen

March 25, 2012

About random effect models

- In purely fixed effects models we have
 - Random variables we observe
 - Model parameters we want to estimate
- In random effects models we have
 - Random variables we observe
 - Random variables we do **NOT** observe
 - Model parameters we want to estimate
- This model class is very useful and goes by many names: **random effects models**, **mixed models**, **latent variable models**, **state-space models**, **frailty models**, **hierarchical models**, ...
- Many tools (e.g. R) can handle linear Gaussian models.
- AD Model Builder has a neat way to handle non-linear non-Gaussian random effect models.

Motivating example: Paired observations

- Two methods A and B to measure blood cell count (to check for the use of doping).
- Paired study.

Person ID	Method A	Method B
1	5.5	5.4
2	4.4	4.9
3	4.6	4.5
4	5.4	4.9
5	7.6	7.2
6	5.9	5.5
7	6.1	6.1
8	7.8	7.5
9	6.7	6.3
10	4.7	4.2

- It must be expected that two measurements from the same person are correlated, so a paired t-test is the correct analysis
- The t-test gives a p-value of 5.1%, which is a borderline result...
- But more data is available

- In addition to the planned study 10 persons were measured with only one method

- Want to use all data, which is possible with random effects

- Assume these 20 are randomly selected from a population where the blood cell count is normally distributed

- Consider the following model:

$$C_i = \alpha(M_i) + B(P_i) + \varepsilon_i, \quad i = 1 \dots 30$$

$\alpha(M_i)$ the 2 fixed method effects

$B(P_i) \sim \mathcal{N}(0, \sigma_P^2)$ the 20 rand. eff.

$\varepsilon_i \sim \mathcal{N}(0, \sigma_R^2)$ measurement noise

All $B(P_i)$ and ε_i are independent

- This model uses all data
- Allows us to test method difference

ID	Meth. A	Meth. B
1	5.5	5.4
2	4.4	4.9
3	4.6	4.5
4	5.4	4.9
5	7.6	7.2
6	5.9	5.5
7	6.1	6.1
8	7.8	7.5
9	6.7	6.3
10	4.7	4.2
11		3.4
12		4.7
13		3.9
14		2.5
15		4.1
16	4.0	
17	6.3	
18	6.0	
19	6.4	
20	3.5	

The paired t-test

```
> t.test(obs[,1],obs[,2],paired=TRUE)
```

```
Paired t-test
```

```
data: obs[, 1] and obs[, 2]
```

```
t = 2.2558, df = 9, p-value = 0.05052
```

```
alternative hypothesis: true difference in means is not equal
```

```
95 percent confidence interval:
```

```
-0.000616709  0.440616709
```

```
sample estimates:
```

```
mean of the differences
```

```
0.22
```

Fit random effects model for the entire data set

- Same data but organized as one line per observation

```
> dim(dat)
```

```
[1] 30  3
```

```
> head(dat)
```

```
   id met  x
1  1  1  5.5
2  1  2  5.4
3  2  1  4.4
4  2  2  4.9
5  3  1  4.6
6  3  2  4.5
```

- Call to fit the random effects model

```
> library(nlme)
```

```
> fit1<-lme(x~met, random=~1|id, data=dat)
```

Summary of model fit

```
> summary(fit1)
```

```
Linear mixed-effects model fit by REML
```

```
Data: dat
```

```
      AIC      BIC   logLik  
82.9282 88.25702 -37.4641
```

```
Random effects:
```

```
Formula: ~1 | id
```

```
(Intercept) Residual
```

```
StdDev:    1.333150 0.2183855
```

```
Fixed effects: x ~ met
```

	Value	Std.Error	DF	t-value	p-value
(Intercept)	5.479360	0.3344155	19	16.384887	0.0000
met	-0.236767	0.0970331	9	-2.440068	0.0374

```
Correlation:
```

```
(Intr)
```

```
met -0.435
```

```
Standardized Within-Group Residuals:
```

Min	Q1	Med	Q3	Max
-1.71559896	-0.29261631	-0.07249575	0.32413719	1.65810172

Testing for no method effect

```
> anova(fit1)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	19	289.66135	<.0001
met	1	9	5.95393	0.0374

Reproducing the t-test as a random effect model

Just to show that this is in fact the same logic that was behind the t-test

```
> datcut<-dat[dat$id<11,]  
> fit2<-lme(x~met, random=~1|id, data=datcut)  
> anova(fit2)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	9	252.85716	<.0001
met	1	9	5.08877	0.0505

Could not do it without random effects

```
> # Notice wrongly ignoring the pairwise nature of  
> # (some of) the data  
> fit3<-lm(x~id+met,data=dat)  
> drop1(fit3, test='F')
```

Single term deletions

Model:

$x \sim id + met$

	Df	Sum of Sq	RSS	AIC	F value	Pr(F)
<none>			44.665	17.940		
id	1	3.4605	48.125	18.178	2.0919	0.1596
met	1	4.2006	48.865	18.636	2.5393	0.1227

2. Motivating example

Unprocessed (baled) wool contain varying amounts of fat and other impurities that need to be removed before further processing. The price - and the value of the baled wool depends on the amount of pure wool that is left after removal of fat and impurities. The purity of the baled wool is expressed as the mass percentage of pure wool in the baled wool.

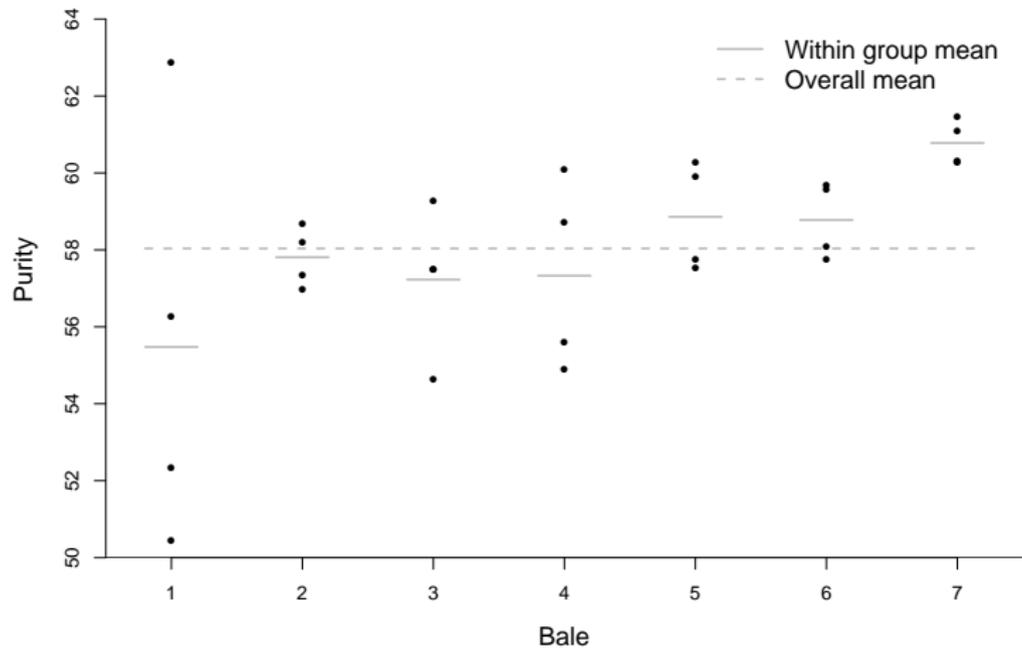
As part of the assessment of different sampling plans for estimation of the purity of a shipment of several bales of wool has U.S. Customs Laboratory, Boston selected 7 bales at random from a shipment of Uruguyan wool, and from each bale, 4 samples were selected for analysis.

Data

	Bale no.						
Sample	1	2	3	4	5	6	7
1	52.33	56.99	54.64	54.90	59.89	57.76	60.27
2	56.26	58.69	57.48	60.08	57.76	59.68	60.30
3	62.86	58.20	59.29	58.72	60.26	59.58	61.09
4	50.46	57.35	57.51	55.61	57.53	58.08	61.45
Bale average	55.48	57.81	57.23	57.33	58.86	58.78	60.78

Table: The purity in % pure wool for 4 samples from each of 7 bales of Uruguyan wool.

Data



One-way ANOVA

The variation in the data can be split up into variation between groups and variation within groups.

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$$

$$SSB = \sum_{i=1}^k n_i (\bar{y}_i - \bar{\bar{y}})^2$$

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

One-way ANOVA

The variation in the data can be split up into variation between groups and variation within groups.

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

The total variation in data

$$SSB = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$$

Variation between groups

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

Variation within groups

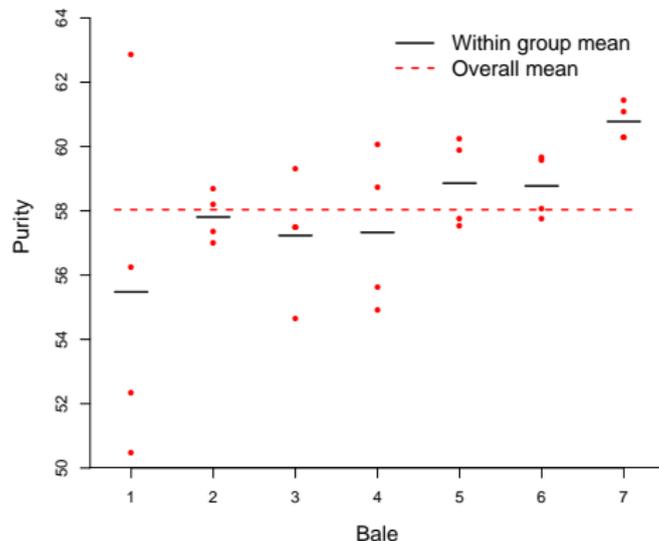
$$SST = SSB + SSE$$

SST

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

$$SSB = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

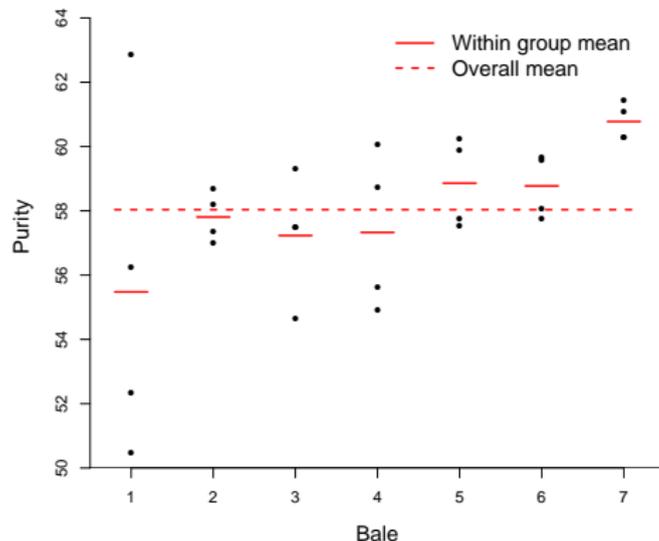


SSB

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

$$SSB = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

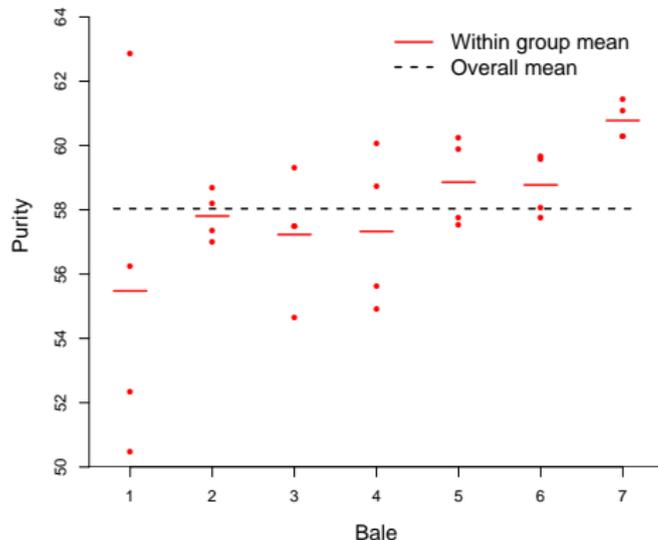


SSE

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

$$SSB = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$



Model with fixed effects

We could formulate a one-way model as discussed in Chapter 3:

$$\mathcal{H}_1 : Y_{ij} \sim N(\mu_i, \sigma^2) \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, n_i$$

and obtain the ANOVA table:

Variation	Sum of Squares	f	$s^2 = SS/f$	F-value	Prob > F
Between bales	SSB 65.9628	6	10.9938	1.76	0.16
Within bales	SSE 131.4726	21	6.2606		
Total	SST 197.4348	27			

The test statistic for $\mathcal{H}_0 : \mu_1 = \mu_2 = \dots = \mu_k$ is $F = 10.99/6.26 = 1.76$.

Model with fixed effects

Such a model would be relevant, if we had selected seven specific bales, eg the bales with identification labels “AF37Q”, “HK983”, . . . , and “BB837”.

Thus, $i = 1$ would refer to bale “AF37Q”, and the probability distributions would refer to repeated sampling, but under such imaginative repeated sampling, $i = 1$ would always refer to this specific bale with label “AF37Q”.

Model with random effects

However, although there is not strong evidence against \mathcal{H}_1 , we will not consider the bales to have the same purity. The idea behind the sampling was to *describe the variation* in the shipment, and the purity of the seven selected bales was not of interest in it self, but rather as representative for the variation in the shipment.

Therefore, instead of the *fixed effects* model in Chapter 3, we shall introduce a *random effects* model.

Formulation of the random model

Definition (One-way model with random effects)

Consider the random variables $Y_{i,j}$, $i = 1, 2, \dots, k$; $j = 1, 2, \dots, n_i$
A *one-way random effects model* for $Y_{i,j}$ is a model such that

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

with $\epsilon_{ij} \sim N(0, \sigma^2)$ and $U_i \sim N(0, \sigma_b^2)$, and where ϵ_{ij} are mutually independent, and also the U_i 's are mutually independent, and finally the U_i 's are independent of ϵ_{ij} .

We shall put

$$N = \sum_{i=1}^k n_i$$

When all groups are of the same size, $n_i = n$, we shall say that the model is *balanced*.

Parameters in the one-way random effects model

Parameters in the one-way random effects model

Consider a one-way random effects model. The parameters of the model are $(\mu, \sigma^2, \sigma_b^2)$.

Sometimes, the *signal to noise ratio*

$$\gamma = \frac{\sigma_b^2}{\sigma^2}$$

is used instead of σ_b^2 .

Thus, the parameter γ expresses the inhomogeneity between groups in relation to the internal variation in the groups. We shall use the term *signal/noise ratio* for the parameter γ .

The one-way model as a hierarchical model

The one-way model as a hierarchical model

Putting $\mu_i = \mu + U_i$ we may formulate it as a *hierarchical model*

We shall assume that

$$Y_{ij} | \mu_i \sim N(\mu_i, \sigma^2) ;$$

in contrast to the *systematic model*, the bale level μ_i is modeled as a realization of a random variable,

$$\mu_i \sim N(\mu, \sigma_b^2),$$

where the μ_i 's are assumed to be mutually independent, and Y_{ij} are *conditionally independent*, i.e. Y_{ij} are mutually independent in the conditional distribution of Y_{ij} for given μ_i .

Random/fixed/mixed

- The random effects model will be a reasonable model in situations where the interest is not restricted alone to the experimental conditions at hand.
- The random effects model will be a reasonable model when the experimental conditions are considered as representative for a larger collection (population) of varying experimental conditions, in principle selected at random from that population.
- The analysis of the systematic model (fixed) puts emphasis on the assessment of the results in the individual groups, μ_i , and possible differences, $\mu_i - \mu_h$, between the results in specific groups.
- The analysis of the random effects model aims at describing the variation between the groups, $\text{Var}[\mu_i] = \sigma_b^2$.
- Models with both fixed and random effects are mixed effects models.

Marginal distribution

Theorem (Marginal distribution in the random effects model for one-way ANOVA)

The marginal distribution of Y_{ij} is a normal distribution with

$$\begin{aligned} E[Y_{ij}] &= \mu \\ \text{Cov}[Y_{ij}, Y_{hl}] &= \begin{cases} \sigma_b^2 + \sigma^2 & \text{for } (i, j) = (h, l) \\ \sigma_b^2 & \text{for } i = h, j \neq l \\ 0 & \text{for } i \neq h \end{cases} \end{aligned}$$

We note that there is a positive covariance between observations from the same group. This positive covariance expresses that observations within the same group will deviate in the same direction from the mean, μ , in the marginal distribution, namely in the direction towards the group mean in question.

The coefficient of correlation

The coefficient of correlation

The coefficient of correlation,

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} = \frac{\gamma}{1 + \gamma}$$

that describe the correlation within a group, is often termed the *intraclass correlation*.

Distribution of individual group averages

Distribution of individual group averages

We finally note that the simultaneous distribution of the group averages is characterized by

$$\text{Cov}[\bar{Y}_{i.}, \bar{Y}_{h.}] = \begin{cases} \sigma_b^2 + \sigma^2/n_i & \text{for } i = h \\ 0 & \text{otherwise} \end{cases}$$

That is, that the k group averages $\bar{Y}_{i.}$, $i = 1, 2, \dots, k$ are mutually independent, and that the variance of the group average

$$\text{Var}[\bar{Y}_{i.}] = \sigma_b^2 + \sigma^2/n_i = \sigma^2(\gamma + 1/n_i)$$

includes the variance of the random component, $\sigma_b^2 = \sigma^2\gamma$, as well as the effect of the residual variance on the group average.

Thus, an increase of the sample size in the individual groups will improve the precision by the determination of the group mean α_i , but the variation between the individual group means is not reduced by this averaging.

Observation vector for a group

When we consider the set of observations corresponding to the i 'th group as a n_i -dimensional column vector,

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix}$$

The set of observations \mathbf{Y}_i , $i = 1, 2, \dots, k$ may be described as k independent observations of a n_i dimensional variable $\mathbf{Y}_i \sim N_{n_i}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_{n_i} + \sigma_b^2 \mathbf{J}_{n_i})$, i.e. that the dispersion matrix for \mathbf{Y}_i is

$$\begin{aligned} \mathbf{V}_i &= D[\mathbf{Y}_i] \\ &= E[(\mathbf{Y}_i - \boldsymbol{\mu})(\mathbf{Y}_i - \boldsymbol{\mu})^T] \\ &= \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & \dots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 & \dots & \sigma_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \dots & \sigma_b^2 + \sigma^2 \end{pmatrix} \end{aligned}$$

where \mathbf{J}_{n_i} is a $n_i \times n_i$ -dimensional matrix consisting solely by 1's.

Covariance structure for the whole set of observations

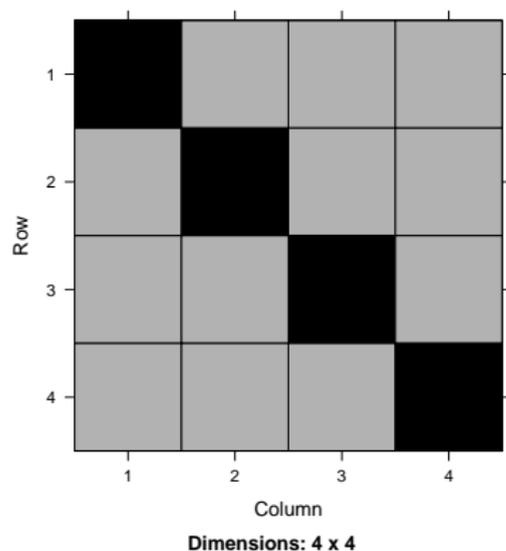
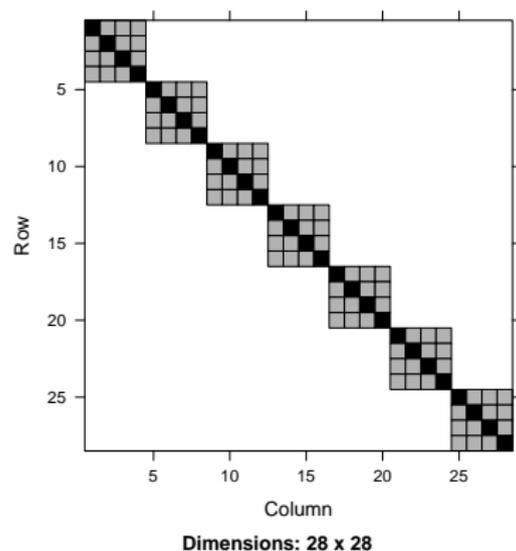
Covariance structure for the whole set of observations

If we organize all observations in one column, organized according to groups, we observe that the $N \times N$ -dimensional dispersion matrix $D[\mathbf{Y}]$ is

$$\mathbf{V} = D[\mathbf{Y}] = \text{Block diag}\{\mathbf{V}_i\}$$

This illustrates that observations from *different groups are independent*, whereas observations from the *same group are correlated*.

Covariance structure for the whole set of observations



Test of the hypothesis of homogeneity in the random effects model

Under the random effect model, the hypothesis that the varying experimental conditions do not have an effect on the observed values, is formulated as

$$\mathcal{H}_0 : \sigma_b^2 = 0.$$

Theorem (Test of the hypothesis of homogeneity in the random effects model)

Under the one-way random effects model the likelihood ratio test for the hypothesis has the test statistic

$$Z = \frac{\text{SSB} / (k - 1)}{\text{SSE} / (N - k)}$$

Large values of z are considered as evidence against the hypothesis.

Under the hypothesis, Z will follow a $F(k - 1, N - k)$ -distribution.

The mixed linear model

Consider now the one way ANOVA with random block effect:

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

The matrix notation is:

$$\underbrace{\begin{pmatrix} Y_{11} \\ Y_{21} \\ Y_{12} \\ Y_{22} \\ Y_{13} \\ Y_{23} \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix}}_{\mathbf{U}} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

Notice how this matrix representation is constructed in exactly the same way as for the fixed effects model — **but separately** for fixed and random effects.

A general linear mixed effects model

A general linear mixed model can be presented in matrix notation by:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{U} + \varepsilon, \quad \text{where } \mathbf{U} \sim N(\mathbf{0}, \mathbf{G}) \text{ and } \varepsilon \sim N(\mathbf{0}, \mathbf{R}).$$

- \mathbf{Y} is the observation vector
- \mathbf{X} is the design matrix for the fixed effects
- β is the vector containing the fixed effect parameters
- \mathbf{Z} is the design matrix for the random effects
- \mathbf{U} is the vector of random effects
 - It is assumed that $\mathbf{U} \sim N(\mathbf{0}, \mathbf{G})$
 - $\text{cov}(U_i, U_j) = G_{i,j}$ (typically \mathbf{G} has a very simple structure (for instance diagonal))
- ε is the vector of residual errors
 - It is assumed that $\varepsilon \sim N(\mathbf{0}, \mathbf{R})$
 - $\text{cov}(\varepsilon_i, \varepsilon_j) = R_{i,j}$ (typically \mathbf{R} is diagonal, but we shall later see some useful exceptions for repeated measurements)

Examples using `lme` in R

Taken from Mixed Effects Models in S and S-PLUS
J.C. Pinheiro and D.M. Bates

The `lme` function is a part of the `nlme` package

One-way random model: Rail data

The data comes from a study of travel time for a certain type of wave that results from longitudinal stress of rails used for railroad track.

```
> str(dat)
```

```
'data.frame':  18 obs. of  2 variables:
 $Rail: Factor w/ 6 levels "1","2","3","4",...: 1 1 1 2 2 2 3 3 3 4
 $travel: int  55 53 54 26 37 32 78 91 85 92 ...
```

The model can be written as

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

with $\epsilon_{ij} \sim N(0, \sigma^2)$ and $U_i \sim N(0, \sigma_b^2)$, and where ϵ_{ij} are mutually independent, and also the U_i 's are mutually independent, and finally the U_i 's are independent of ϵ_{ij} .

groupedData object

A `groupedData` object contains the data values themselves, stored as a data frame and a formula that specifies the *response*, the *primary covariate* and the *grouping factor* that divides the observations into groups. It is given as:

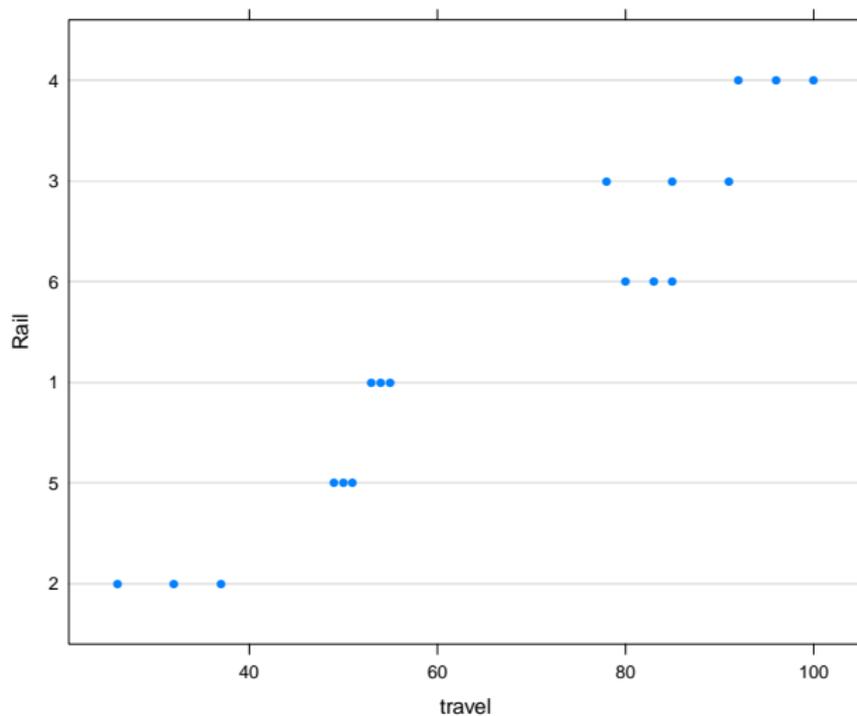
```
response ~ primary | grouping
```

In the Rail case we write:

```
library(nlme)  
dat2<-groupedData(travel ~ 1 | Rail,dat)
```

Plotting groupData objects

```
> plot(dat2)
```



The `lme()` function

The `lme` function can be used to fit linear mixed-effects models by maximum likelihood or by restricted maximum likelihood (default). The `lme` function is in the `nlme` package. Several arguments can be used with the function but a typical call is:

```
lme( fixed, data, random)
```

The first argument is a two-sided linear formula specifying the dependent variable and the fixed effects in the model (as in `lm`). The third argument is typically given as a one-sided linear formula specifying the effects and the grouping structure in the model.

For the Rail data the call is:

```
fit1<-lme(travel ~ 1, data = dat2, random = ~ 1|Rail)
```

The `lme()` function

```
fit1<-lme(travel ~ 1, data = dat2, random = ~ 1|Rail)
```

The first argument indicates that the response is `travel`, and that there is a single fixed effect, the intercept.

The second argument indicates that the data can be found in the `dat2` object.

The third argument indicates that there is a single random effect for each group and that the grouping is given by the variable `Rail`.

Some lme methods

<code>anova</code>	Likelihood ratio or conditional test
<code>coef</code>	Estimated coefficients for different levels of grouping
<code>fitted</code>	Fitted values for different levels of grouping
<code>fixef</code>	fixed-effects estimates
<code>intervals</code>	Confidence intervals on model parameters
<code>logLik</code>	Log-likelihood at convergence
<code>pairs</code>	Scatter-plot matrix of coefficients or random effects
<code>predict</code>	Predictions for different levels of grouping
<code>print</code>	Brief information about the fit
<code>qqnorm</code>	Normal probability plots
<code>ranef</code>	Random-effect estimates
<code>resid</code>	Residuals for different levels of grouping
<code>summary</code>	More detailed information about the fit
<code>update</code>	Update the lme fit

summary()

```

> summary(fit1)
Linear mixed-effects model fit by REML
Data: dat2
      AIC      BIC   logLik
128.177 130.6766 -61.0885
Random effects:
Formula: ~1 | Rail
      (Intercept) Residual
StdDev:    24.80547 4.020779
Fixed effects: travel ~ 1
              Value Std.Error DF   t-value p-value
(Intercept)  66.5    10.17104 12  6.538173     0
Standardized Within-Group Residuals:
      Min           Q1           Med           Q3           Max
-1.61882658 -0.28217671  0.03569328  0.21955784  1.61437744
Number of Observations: 18
Number of Groups: 6

```

The REML estimates are:

$$\hat{\beta} = 66.5, \quad \hat{\sigma}_b = 24.805, \quad \hat{\sigma} = 4.021.$$

AIC, BIC

The summary output included the values of the *Akaike Information Criterion*, AIC, and *Bayesian Information Criterion*, BIC.

They are both model comparison criteria defined as

$$AIC = -2 \log \text{Lik} + 2n_{par}$$

$$BIC = -2 \log \text{Lik} + n_{par} \log(N)$$

where n_{par} denotes the number of parameters in the model and N the total number of observations used to fit the model.

Under these definitions smaller is better!

Testing the model

```
# test for fixed part
> anova(fit1)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	12	42.7477	<.0001

```
# test for random part
fit2<-lm(travel ~ 1,data = dat2)
```

```
> anova(fit1,fit2)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
fit1	1	3	128.1770	130.6766	-61.08850			
fit2	2	2	162.6815	164.3479	-79.34075	1 vs 2	36.50451	<.0001

Confidence intervals on parameters

```

> intervals(fit1)
Approximate 95% confidence intervals

Fixed effects:
              lower est.      upper
(Intercept) 44.33921 66.5 88.66079
attr(,"label")
[1] "Fixed effects:"

Random Effects:
Level: Rail
              lower      est.      upper
sd((Intercept)) 13.27433 24.80547 46.35346

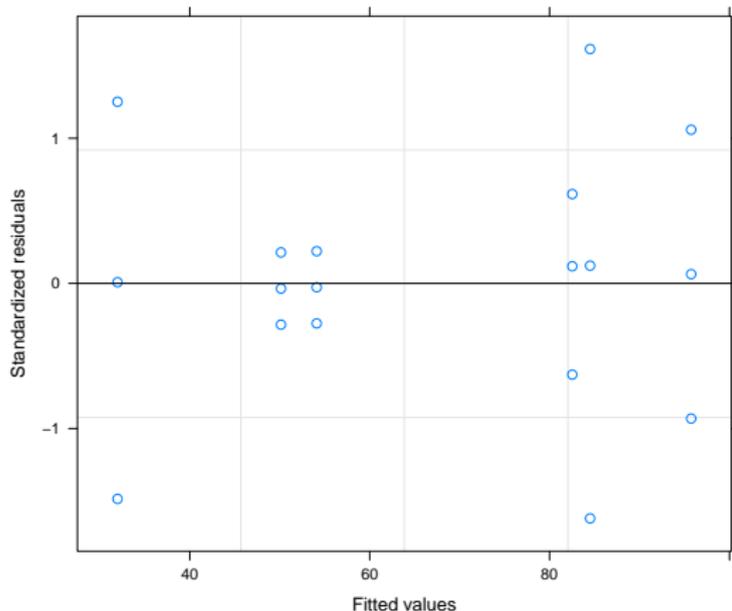
Within-group standard error:
  lower      est.      upper
2.695003 4.020779 5.998757

```

Assessing the model fit

The fitted model should be examined using graphical summaries:

```
> plot(fit1)
```



and more plots ...

Randomized block design: Chair data

The data comes from an experiment for assessing the effects of a different chair types on muscular activity.

The effort of required by each of nine different subjects to arise from each of four types of chairs is recorded. We want to compare the four particular chair types so we model Type as a fixed factor. The nine subjects represent a sample from the population we wish to make inferences so we model Subject as a random factor.

```
> str(dat)
```

```
'data.frame':  36 obs. of  3 variables:
 $effort : int  12 15 12 10 10 14 13 12 7 14 ...
 $Type: Factor w/ 4 levels "T1","T2","T3",...: 1 2 3 4 1 2 3 4 1 2 ...
 $Subject:Factor w/ 9 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3
```

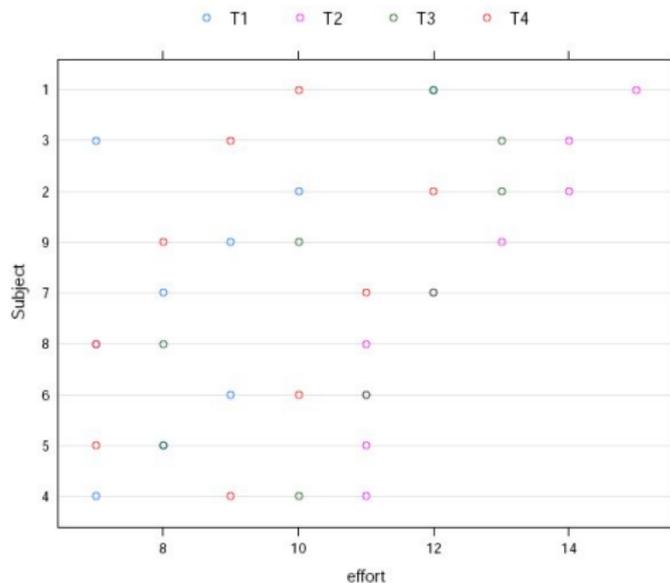
The model can be written as:

$$Y_{ij} = \mu + a_i + U_j + \epsilon_{ij} ,$$

with $\epsilon_{ij} \sim N(0, \sigma^2)$ and $U_j \sim N(0, \sigma_b^2)$, and where ϵ_{ij} are mutually independent, and also the U_j 's are mutually independent, and finally the U_i 's are independent of ϵ_{ij} .

plot()

```
> dat2<-groupedData(effort ~ Type | Subject,data=dat)
> plot(dat2)
```



lme - summary

```
> fit1<-lme(effort~Type, data = dat2, random = ~1|Subject)
> summary(fit1)
```

Linear mixed-effects model fit by REML

Data: dat2

	AIC	BIC	logLik
	133.1308	141.9252	-60.5654

Random effects:

Formula: ~1 | Subject

(Intercept) Residual

StdDev: 1.332465 1.100295

Fixed effects: effort ~ Type

	Value	Std.Error	DF	t-value	p-value
(Intercept)	8.555556	0.5760123	24	14.853079	0.0000
TypeT2	3.888889	0.5186838	24	7.497610	0.0000
TypeT3	2.222222	0.5186838	24	4.284348	0.0003
TypeT4	0.666667	0.5186838	24	1.285304	0.2110

...

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-1.80200345	-0.64316591	0.05783115	0.70099706	1.63142054

Number of Observations: 36

Number of Groups: 9

Testing the model

```
# Fixed effects
```

```
> anova(fit1)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	24	455.0075	<.0001
Type	3	24	22.3556	<.0001

```
# Random effect
```

```
> fit2<-lm(effort~Type, data = dat)
```

```
> anova(fit1,fit2)
```

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
fit1	6	133.1308	141.9252	-60.56539			
fit2	5	144.6081	151.9367	-67.30403	1 vs 2	13.47728	2e-04

Confidence intervals

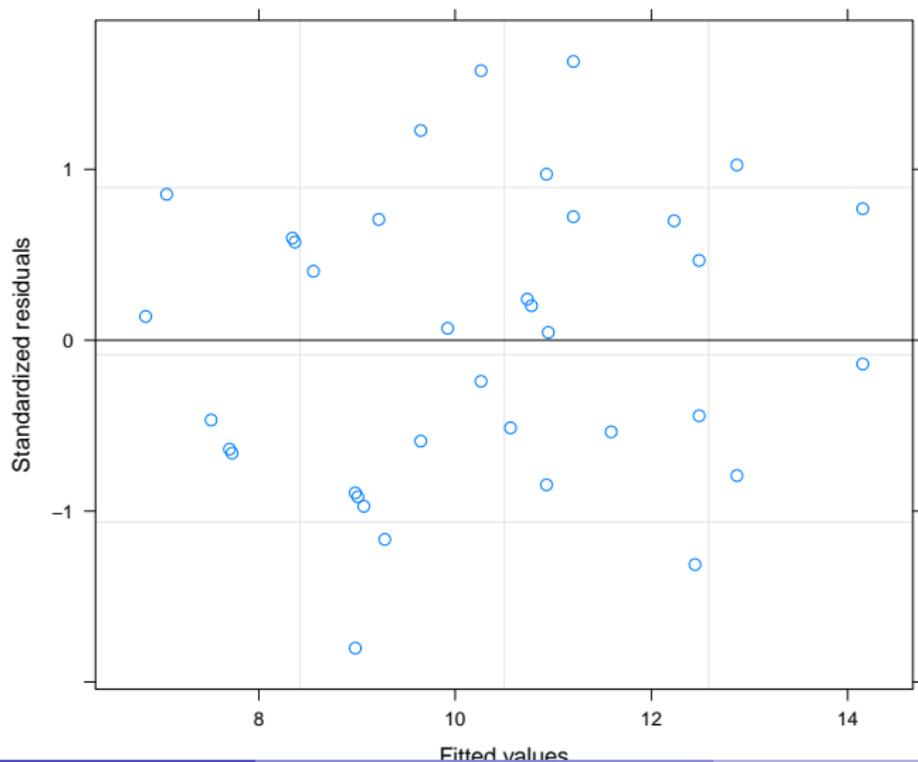
```

> intervals(fit1)
Approximate 95% confidence intervals
Fixed effects:
              lower      est.    upper
(Intercept)  7.3667247  8.5555556  9.744386
TypeT2       2.8183781  3.8888889  4.959400
TypeT3       1.1517114  2.2222222  3.292733
TypeT4      -0.4038442  0.6666667  1.737177
attr("label")
[1] "Fixed effects:"
Random Effects:
  Level: Subject
              lower      est.    upper
sd((Intercept)) 0.7495964  1.332465  2.368559
Within-group standard error:
      lower      est.    upper
0.8292498  1.1002946  1.4599319

```

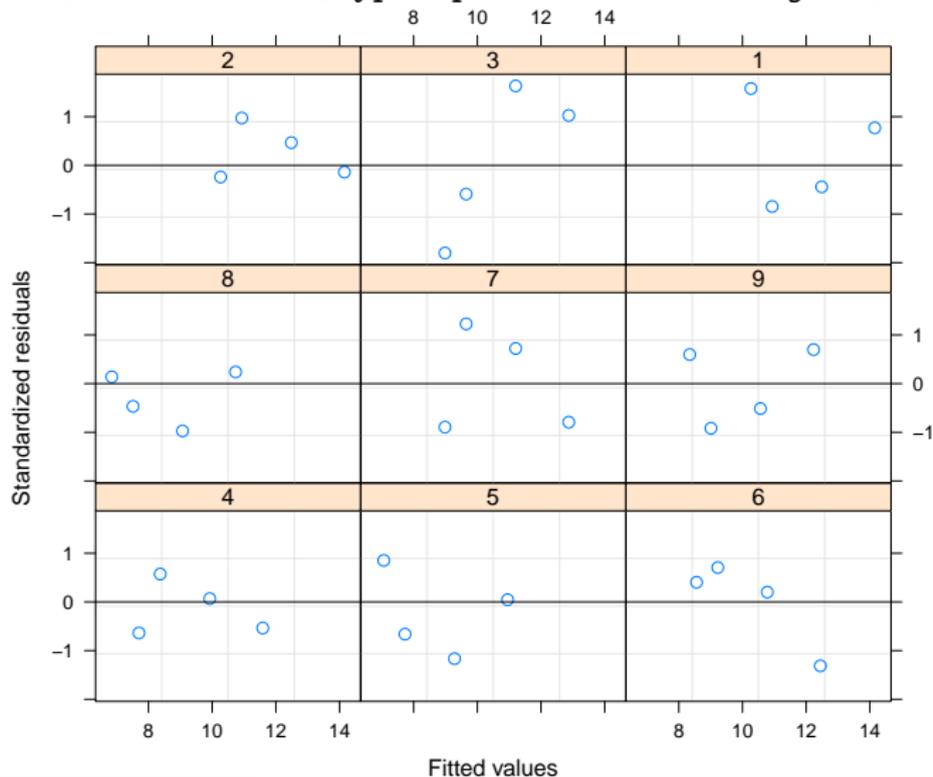
`plot()`

```
> plot(fit1)
```



plot()

```
> plot(fit1, form=resid(., type='p')~fitted(.)|Subject, abline=0)
```



```
plot()
```

```
> plot(fit1,form=resid(.,type='p')~fitted(.)|Type,abline=0)
```

