Intruction to General and Generalized Linear Models
Mixed Effects Models IV

Henrik Madsen
Anna Helga Jónsdóttir

hm@imm.dtu.dk

April 30, 2012

This lectures

- The extended linear mixed effects model

- Nested models

- Hierarchical random effects

- The extended linear mixed effects model

  - Modelling heteroscedasticity

  - Modelling dependence among within-group errors

Linear mixed effects model

Recall from Section 5.4 p. 179:

Definition (Linear mixed effects model)

The model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{\epsilon}$$

with $\boldsymbol{X}$ and $\boldsymbol{Z}$ denoting known matrices, and where $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{\Psi})$ are independent is called a *mixed general linear model*. In the general case may the variance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ depend on some unknown parameters, $\boldsymbol{\theta}$, that also needs to be estimated.

The parameters $\boldsymbol{\beta}$ are called *fixed effects* or *systematic effects*, while the quantities $\boldsymbol{U}$ are called *random effects*.

In previous examples, we most often assumed that

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_N$$

$$\boldsymbol{\Psi} = \sigma_b^2 \boldsymbol{I}_k$$

We will relax the assumption regarding the $\boldsymbol{\Sigma}$-matrix later in the lecture.
We will use $\boldsymbol{V}$ for the dispersion matrix in the marginal distribution of $\boldsymbol{Y}$, i.e.
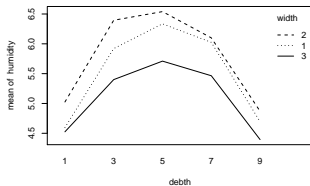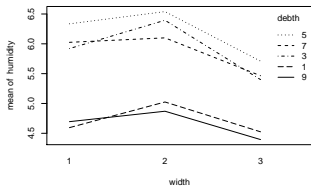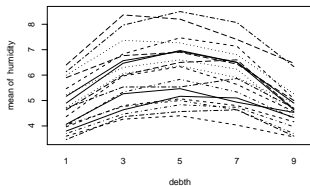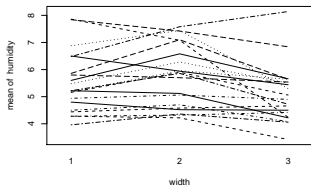
Example: Drying of beech wood

To investigate the effect of drying of beech wood on humidity, the following experiment was conducted:
Each of 20 planks was dried in a certain period of time. Then the humidity was measured in 5 depths (1,3,5,7,9) and 3 widths (1,2,3) for each plank: depth 1: close to the top, depth 5: in the center, depth 9: close to the bottom, depth 3: between 1 and 5, depth 7: between 5 and 9. width 1: close to the side, width 3: in the center, width 2: between 1 and 3.

```
> str(dat)
'data.frame':    300 obs. of  4 variables:
$ plank : Factor w/ 20 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 1 ...
$ width : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 2 2 2 2 2 ...
$ depth : Factor w/ 5 levels "1","3","5","7",..: 1 2 3 4 5 1 2 3 4 5 ...
$ humidity: num  3.4 4.9 5 4.9 4 4.1 4.7 5.2 4.6 4.3 ...
```

Example: Drying of beech wood

```
with(dat, interaction.plot(width,plank,humidity,legend=F))
with(dat, interaction.plot(depth,plank,humidity,legend=F))
with(dat, interaction.plot(width,depth,humidity,legend=T))
with(dat, interaction.plot(depth,width,humidity,legend=T))
```

Example: Drying of beech wood

```
options(contrasts=c("contr.sum", "contr.poly"))
> fit1<-lme(humidity~depth+width+depth:width,
random=~1|plank,data=dat)
> anova(fit1)
            numDF denDF  F-value p-value
(Intercept)     1   266 593.6487  <.0001
depth           4   266  78.2592  <.0001
width           2   266  29.6463  <.0001
depth:width     8   266   1.0840  0.3745

> fit2<-lme(humidity~depth+width,random=~1|plank,data=dat)
> anova(fit2)
            numDF denDF  F-value p-value
(Intercept)     1   274 593.6487  <.0001
depth           4   274  78.0676  <.0001
width           2   274  29.5738  <.0001
```

Example: Drying of beech wood

```
> summary(fit2)
Linear mixed-effects model fit by REML
 Data: dat
     AIC      BIC   logLik
  687.24 720.3615 -334.62
Random effects:
 Formula: ~1 | plank
        (Intercept) Residual
StdDev:   0.9897981 0.6361904
Fixed effects: humidity ~ depth + width
                Value  Std.Error  DF   t-value  p-value
(Intercept)  5.466333 0.22435273 274 24.364907  0.0000
depth1      -0.751333 0.07346094 274 -10.227658 0.0000
depth2       0.438667 0.07346094 274  5.971428  0.0000
depth3       0.728667 0.07346094 274  9.919104  0.0000
depth4       0.397000 0.07346094 274  5.404233  0.0000
width1       0.047667 0.05194473 274  0.917642  0.3596
width2       0.319667 0.05194473 274  6.153977  0.0000
...
```

Hierarchical random effects

In some cases the experimental design leads to a natural hierarchy of the random effects.

Consider, for example the following experimental design to determine the fat content in milk.

- Two farms were selected randomly from all milk producing farms in some area.

- From each selected farms, two cows were randomly selected.

- From each cow, two cups of milk were taken randomly from the daily production and analysed in a lab.

Single level of grouping

To analyse the cow data using single level of grouping, the information about farm is ignored but the information about the cow is included.

In that case, the model is

$$y_{ij} = \mu + B_i + \epsilon_{ij}$$
$$B_i \in N(0, \sigma_1^2), \quad \epsilon \in N(0, \sigma^2)$$

and the dispersion matrix in the marginal distribution is:

$$\boldsymbol{V} = \begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_1^2 + \sigma^2 & \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_1^2 & \sigma_1^2 + \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_1^2 + \sigma^2 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_1^2 & \sigma_1^2 + \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{pmatrix}$$

Two level of grouping

The natural model for the cow data is the two level grouping model, with `farm` as the top layer.

In that case, the model is

$$y_{ijk} = \mu + B_i + B_{ij} + \epsilon_{ijk}$$
$$B_i \in N(0, \sigma_2^2), \quad B_{ij} \in N(0, \sigma_1^2), \quad \epsilon \in N(0, \sigma^2)$$

with B_i representing the farms and B_{ij} the cows nested within the farms. The dispersion matrix in the marginal distribution is:

$$\begin{pmatrix} \sigma_2^2 + \sigma_1^2 + \sigma^2 & \sigma_2^2 + \sigma_1^2 & \sigma_2^2 & \sigma_2^2 & 0 & 0 & 0 & 0 \\ \sigma_2^2 + \sigma_1^2 & \sigma_2^2 + \sigma_1^2 + \sigma^2 & \sigma_2^2 & \sigma_2^2 & 0 & 0 & 0 & 0 \\ \sigma_2^2 & \sigma_2^2 & \sigma_2^2 + \sigma_1^2 + \sigma^2 & \sigma_2^2 + \sigma_1^2 & 0 & 0 & 0 & 0 \\ \sigma_2^2 & \sigma_2^2 & \sigma_2^2 + \sigma_1^2 & \sigma_2^2 + \sigma_1^2 + \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_2^2 + \sigma_1^2 + \sigma^2 & \sigma_2^2 + \sigma_1^2 & \sigma_2^2 & \sigma_2^2 \\ 0 & 0 & 0 & 0 & \sigma_2^2 + \sigma_1^2 & \sigma_2^2 + \sigma_1^2 + \sigma^2 & \sigma_2^2 & \sigma_2^2 \\ 0 & 0 & 0 & 0 & \sigma_2^2 & \sigma_2^2 & \sigma_2^2 + \sigma_1^2 + \sigma^2 & \sigma_2^2 + \sigma_1^2 \\ 0 & 0 & 0 & 0 & \sigma_2^2 & \sigma_2^2 & \sigma_2^2 + \sigma_1^2 & \sigma_2^2 + \sigma_1^2 + \sigma^2 \end{pmatrix}$$

Example: Lactase measurements in piglets

As part of a larger study of the intestinal health in newborn piglets, the gut enzyme lactase was measured in 20 piglets taken from 5 different litters. For each of the 20 piglets the lactase level was measured in three different regions. At the time the measurement was taken the piglet was either unborn (status=1) or newborn (status=2).

The hierarchical structure for the first two litters is:
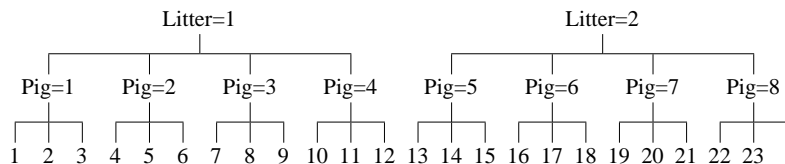
Example: Lactase measurements in piglets

The model can be written as

$$y_{ijkl} = \mu + r_i + s_j + ri_{ij} + B_k + B_{kl} + \epsilon_{ijkl}$$
$$B_k \in N(0, \sigma_1^2), \quad B_{kl} \in N(0, \sigma_2^2), \quad \epsilon \in N(0, \sigma^2)$$

where r_i denotes the region, s_j the status, B_k denotes the litter, B_{kl} the piglets within litters.

```
> str(dat)
'data.frame':    60 obs. of  5 variables:
$ litter : Factor w/ 5 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
$ pig    : Factor w/ 20 levels "1","2","3","4",..: 1 1 1 2 2 2 3 3 3 4 ...
$ reg    : Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ...
$ status : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 2 2 2 2 ...
$ loglact: num  25 32 19 45 9 41 18 22 2 40 ...
```

Example: Lactase measurements in piglets

```
> fit1<-lme(loglact~reg+status+reg:status,data=dat,
random=~1|litter/pig,na.action=na.omit)
> summary(fit1)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  97.91436 115.647 -39.95718
Random effects:
 Formula: ~1 | litter
         (Intercept)
StdDev: 2.269151e-05
Formula: ~1 | pig %in% litter
         (Intercept) Residual
StdDev:    0.3345345 0.3673081
Fixed effects: loglact ~ reg + status + reg:status
                 Value Std.Error DF   t-value p-value
(Intercept)  2.1292914 0.1877797 35 11.339305  0.0000
reg2         0.0093629 0.1963344 35  0.047688  0.9622
reg3        -0.0086600 0.1963344 35 -0.044108  0.9651
status2     -0.1211776 0.2329121 14 -0.520272  0.6110
reg2:status2 0.1098179 0.2435229 35  0.450955  0.6548
reg3:status2 -0.6550188 0.2458415 35 -2.664395  0.0116
```

Example: Lactase measurements in piglets

```
> anova(fit1)
            numDF denDF  F-value p-value
(Intercept)     1    35 478.7082  <.0001
reg             2    35  10.2922  0.0003
status          1    14   2.5496  0.1326
reg:status      2    35   5.6265  0.0076
```

Example: Lactase measurements in piglets

```
fit2<-lme(loglact~reg+status+reg:status,data=dat,
random=~1|pig,na.action=na.omit)
fit3<-lm(loglact~reg+status+reg:status,data=dat,
na.action=na.omit)

> anova(fit1,fit2)
Model df  AIC      BIC      logLik     L.Ratio      p-value
fit1 9 97.91436 115.6470 -39.95718
fit2 8 95.91436 111.6767 -39.95718 1.349653e-08 0.9999
> anova(fit2,fit3)
Model df  AIC      BIC      logLik     L.Ratio      p-value
fit2 8  95.91436 111.6767 -39.95718
fit3 7 103.46331 117.2554 -44.73165 9.548944     0.002
```
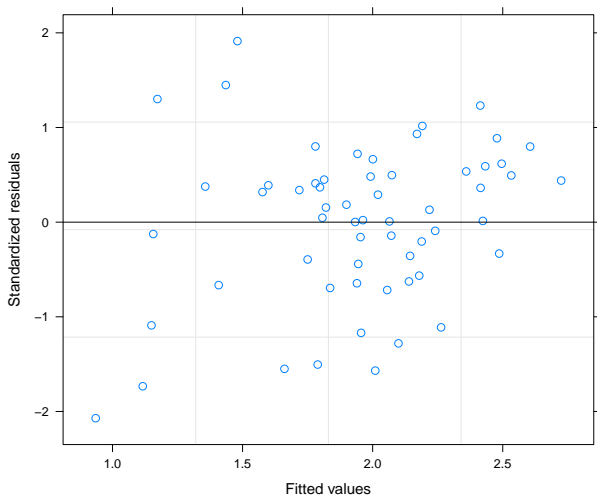
Example: Lactase measurements in piglets

`plot(fit2)`

The extended linear mixed effects model

- In the models we have considered so far, considerably flexibility is allowed in the random effects structure but we have restricted the within group errors ϵ to be independent, identically distributed random variables with mean zero and constant variance.

- There are many applications involving grouped data for which the within-group errors are *heteroscedastic* (i.e. have unequal variances) or are *correlated*, or even both.

- We will now extend the linear mixed effect model to allow for heteroscedastic correlated within group errors and use lme to fit those kind of models.

The extended linear mixed effects model

- Within-group errors allowed to be correlated and heteroscedastic

$$\epsilon \sim N(0, \sigma^2 \mathbf{\Sigma})$$

- The within-group variance-covariance matrix $\mathbf{R}$ can be decomposed into

$$\mathbf{\Sigma} = \mathbf{V}\mathbf{C}\mathbf{V},$$

- $\mathbf{V}$ describes the variance and $\mathbf{C}$ describes the correlation

- Convenient since the heteroscedasticity and the correlation structure can be modeled separately

Modeling heteroscedasticity - `varFunc` classes in R

- Variance functions in R
- Arguments: `value` and `form`
- `value` specifies the initial value for the variance parameter(s). As default it is optimized in the iteration process, but held fixed if specified
- `form` is a one-sided formula specifying a covariate and optionally a grouping variable

$$\texttt{form = } \sim \texttt{ covariate | group}$$

- If the variance should depend on $\hat{\mu}_{ij}$, it is formulated

$$\texttt{form = } \sim \texttt{ fitted(.) | group}$$

Modeling heteroscedasticity - `varFunc` classes in R

- `varFixed`

$$\mathsf{Var}(\epsilon_{ij}) = \sigma^2 \mathsf{covariate}_{ij}$$

within-group variance proportional to a covariate. No grouping variables or the symbol `.`, i.e. no fitted values are allowed

- `varIdent`

$$\mathsf{Var}(\epsilon_{ij}) = \sigma^2 \delta_{s_{ij}}^2$$

different within-group variances for each level of a grouping variable. Variance parameters is the standard deviation of the kth group relative to the first group, i.e. $\delta_1 = 1$ and $\delta_k > 0, k = 2, ..., S$

Modeling heteroscedasticity - `varFunc` classes in R

- `varPower`

$$\text{Var}(\epsilon_{ij}) = \sigma^2 |v_{ij}|^{2\delta}$$

the absolute value of a covariate raised to a power δ. When values of the covariate are close 0, `varConstPower` may be more more appropriate

- `varConstPower`

$$\text{Var}(\epsilon_{ij}) = \sigma^2(\delta_1 + |v_{ij}|^{\delta_2})^2$$

Similar to `varPower` but with a constant added to the power term. δ_1 is restricted to be positive. May provide a more realistic model when the values of the covariate are close to 0

Modeling heteroscedasticity - `varFunc` classes in R

- `varExp`

$$\mathsf{Var}(\epsilon_{ij}) = \sigma^2 \mathsf{exp}(2\delta v_{ij})$$

  the within-group variance increases or decreases exponentially with a covariate

- `varComb`
  Combines different variance functions by multiplying them together

Modeling dependence among within-group errors

- Historically developed for time-series data (serial correlation) and spatial data (spatial correlation)

- We will only consider serial correlation here.

- Use `corStruct` classes in R to model the correlation

- Syntax in R is similar to the `varFunc` classes
  Arguments: `value` and `form`

- `value`: initial value for the correlation parameter(s). Held fixed if specified

- `form`: one-sided formula specifying the position vector and grouping variable

$$\text{form} = \sim \text{position} \mid \text{group}$$

Modeling dependence - serial `corStruct` classes

- `corCompSymm`

$$h(k, \rho) = \rho, \quad k = 1, 2, ...$$

All within-group errors in the same group are correlated equally regardless the distance between them

- This is the correlation structure for a mixed-effects model with a random intercept and iid within-group errors with the correlation parameter

$$\rho = \sigma_b^2 / (\sigma_b^2 + \sigma^2)$$

- Useful when there are few observations per group but may be too simple otherwise

Modeling dependence - serial `corStruct` classes

- `corSymm`

$$h(k, \rho) = \rho_k, \quad k = 1, 2, ...$$

- General correlation structure with different parameters for each lag
- Often yields overparameterized models and mostly useful with few observations per group

Modeling dependence - serial `corStruct` classes

- `corARMA`: Possible to specify autoregressive models, moving average models or both
- The autoregressive (AR) model of order p

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

  The current observation (Y_t) is a linear function of prior observations plus an iid noise term ϵ_t
- Since Y_t depends on prior observations that again depends on prior observations etc, the correlation structure is infinite

Modeling dependence - serial `corStruct` classes

- The moving average (MA) model of order q

$$Y_t = \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

The current observation is a linear function of prior errors plus an iid noise term ϵ_t

- Y_t depends only on prior errors and the correlation structure is finite. Observations with lag $> q$ are uncorrelated

- In R the arguments p and q specifies the order of the autoregressive model and the moving average, respectively

Modeling dependence - serial `corStruct` classes

- `corAR1`

$$h(k, \rho) = \phi^k, \quad k = 0, 1, ...$$

- Special case of the general autoregressive function described above
- A simple and often useful correlation model with only one parameter
- `corCAR1`

$$h(s, \rho) = \phi^s, \quad s \leq 0, \quad \rho \leq 0$$

- `corCAR1` is `corAR1` extended to accommodate continuous time measurements

Example on specification in 'lme'

Specification of a mixed-effects model - also with heteroscedastic error structure

```
m.0 <- lme(dds.log ~ dbh.log+dbh+BAL+BA+SITES,
            random=~1|PLot/MP, # measured periods (MP) are nes
            data=tree, method="ML")
m.AR1 <- update(m.0, correlation = corAR1())
m.0d <- update(m.AR1, weights = varPower(form = ~dbh))
m.0w <- update(m.AR1, weights = varPower())
m.0w.REML <- update(m.0w, method="REML")
```

Example: Orthodont

The Orthodont data frame has 108 rows and 4 columns of the change in an orthdontic measurement over time for several young subjects.

```
> str(Orthodont)
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 ...
```

Example: Orthodont

From a previous exercise ...

```
fit1<-lme(distance ~ I(age-11),data=Orthodont,random= ~1|Subject)
```

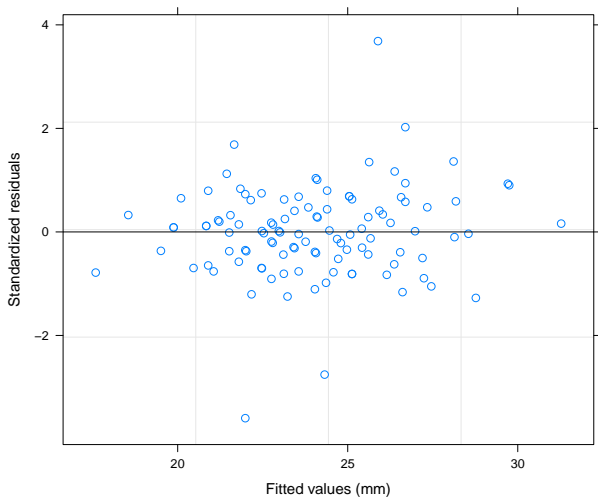We would like to see if boys and girls have different growth patterns:

```
fit2<-update(fit1,fixed=distance~Sex*I(age-11))

> anova(fit2)
               numDF denDF  F-value p-value
(Intercept)        1    79 4123.156  <.0001
Sex                1    25    9.292  0.0054
I(age - 11)        1    79  122.450  <.0001
Sex:I(age - 11)    1    79    6.303  0.0141
```
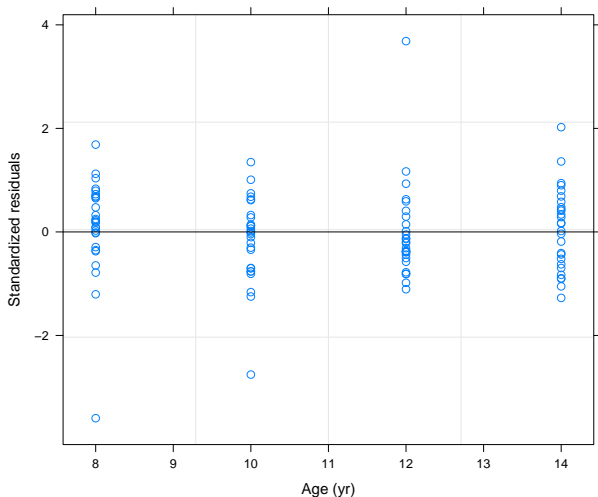
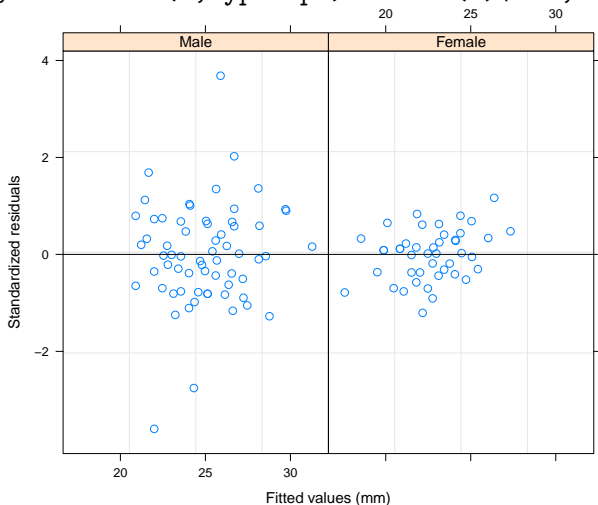# Example: Orthodont

`plot(fit2)`

Example: Orthodont

```
plot(fit2,form=resid(.,type='p')~age,abline=0)
```
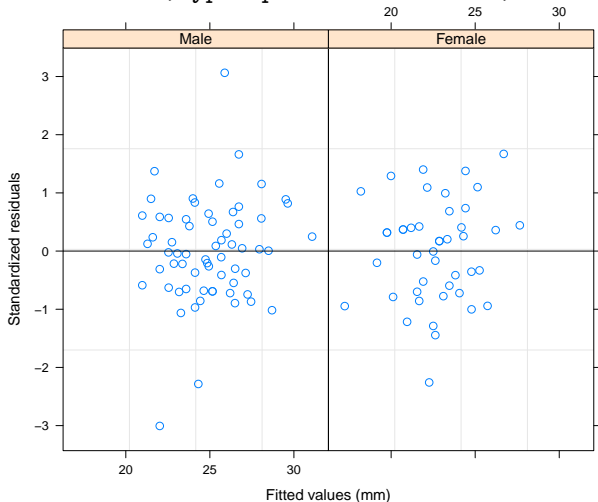
Example: Orthodont

```
plot(fit2,form=resid(.,type='p')~fitted(.)|Sex,abline=0)
```

Example: Orthodont

```
fit3<-update(fit2,weights=varIdent(form = ~ 1 | Sex))
plot(fit3,form=resid(.,type='p')~fitted(.)|Sex,abline=0)
```

Example: Orthodont

```
> anova(fit2,fit3)
     df AIC       BIC       logLik   Test L.Ratio p-value
fit2  6 445.7572 461.6236 -216.8786
fit3  7 429.2205 447.7312 -207.6102 1 vs 2 18.53677  <.0001

> summary(fit3)
....
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | Sex
 Parameter estimates:
    Male     Female
1.0000000 0.4678944
....
```