559

Mixed Effects Modeling-II

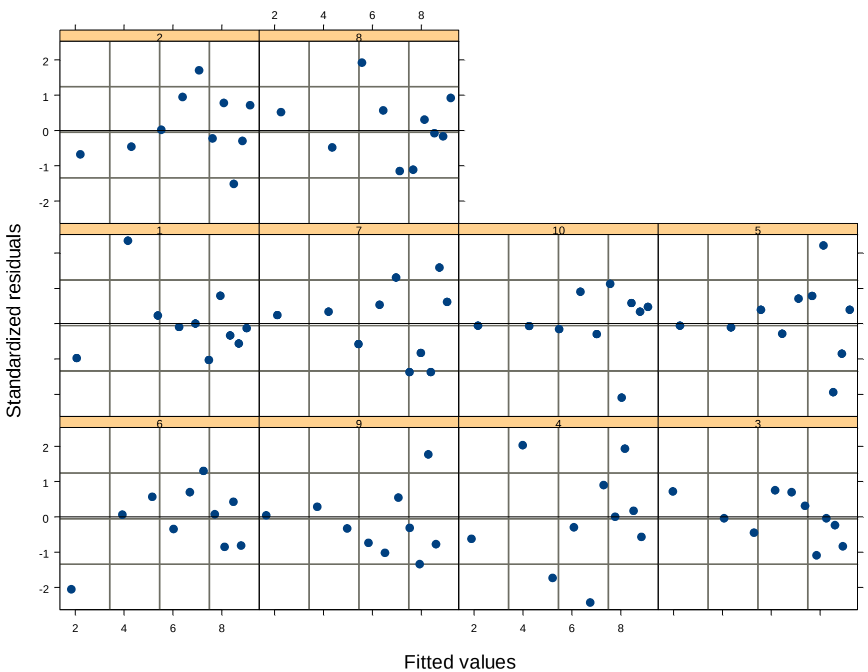
Fish 559; Lecture 7

Tricks and Traps-II

- Residuals and the like:
  - To plot Pearson residuals:

    plot(lmout,form=resid(.,type="p")~fitted(.)|Subject,abline=0,pch=16)

- You can also have interactions (random effects within random effects)
- You can also have different variances for different groups.



Tricks and Traps-III

- It is possible to allow for differences in within-group variability, i.e.:

$$\mathrm{Var}(\varepsilon_{i,j} \mid \underline{b}_i) = \sigma^2 g^2(E[y_{i,j} \mid \underline{b}_i], v_{i,j}, \underline{\delta}); \quad = \sigma^2 v_{i,j}$$

Variance function

Variance covariate

- It is also possible to allow for correlation among the within-group residuals.
- R: Use the varFunc and corStruct classes. This also allow weighting of different data points.

Further R Tricks
(Variance formulae)

- The varFunc class:
  - varFixed: $var(\varepsilon_{i,j}) = \sigma^2 v_{i,j}$
  - varIdent: $\mathrm{var}(\varepsilon_{1,j}) = \sigma^2;\ \mathrm{var}(\varepsilon_{i\neq 1,j}) = \sigma^2 \delta_i^2$
  - varPower: $\mathrm{var}(\varepsilon_{i,j}) = \sigma^2 \mid v_{i,j} \mid^{2\delta}$
  - varExp: $\mathrm{var}(\varepsilon_{i,j}) = \sigma^2 \exp(2\delta v_{i,j})$
  - varConstPower: $\mathrm{var}(\varepsilon_{i,j}) = \sigma^2 (\delta_1 + \mid v_{i,j} \mid^{\delta_2})$
  - varComb: Combinations of variance functions
- Notes:
  - varPower(initial value, form, fixed)
  - vf1 <- varPower(0.2, form = ~age|Sex)
  - vf1 <- varPower(0.2, form = ~ fitted()|Sex, fixed=list(Male=0.5, Female=0))

Packages

- The function lmer (in package lme4) also implements linear mixed models (including generalized linear models).
- lmer doesn’t have all the functionality of lme – however, both packages should be considered in any application.

Back to an Example

The file Lect2a.txt contains data on the profit of 20 boats. Use these data to explore the support for the following models (assume a boat-specific intercept):

- Latitude as a main effect; boat-specific variances
- Latitude*Longitude as main effects; boat-specific variances
- Latitude*Longitude as main effects; constant variance

Extensions to the Linear Mixed Model

1. A non-linear relationship between the response variable and the co-variates:

$$\underline{y}_i = f_i(\underline{\beta}, \underline{b}_i) + \varepsilon_{i,j} \qquad \underline{b}_i \sim N(0; \mathbf{\Psi}) \qquad \varepsilon_{i,j} \sim N(0; \sigma^2)$$

2. A non-normal distribution for the random effects.
3. A non-normal distribution for the within-group variability.

- We only discuss extension 1) here. Extensions 2) and 3) will be discussed when we talk about STAN.

Non-linear Mixed Effects Models
(Some Examples)

1. Modeling age and length data for animals that are measured multiple times.
2. A meta-analysis for the parameters of the stock-recruitment relationship for many stocks.
3. A surplus production model with process and observation error**.
4. An age-structured population dynamics model with recruitment variability (Stock Synthesis)**.

** These can be dealt with more straightforwardly by adopting a Bayesian rather a classical estimation framework.

Theoretical Aspects-I

Function relating the parameters and covariates to the observations

$$\phi_{i,j} = \mathbf{A}_{i,j}\underline{\beta} + \mathbf{B}_{i,j}\underline{b}_i$$

Fixed effects

Random effects

$$\underline{y}_i = f(\boldsymbol{\varphi}, v_{i,j}) + \underline{\varepsilon}_i \quad \underline{\varepsilon}_i \sim N(0, \sigma^2\mathbf{I})$$

$$\underline{b}_i \sim N(0, \boldsymbol{\Psi})$$

$$\sigma^2\boldsymbol{\Psi}^{-1} = \boldsymbol{\Delta}^T\boldsymbol{\Delta}$$

Observations for group *i*

This matrix will be a function of parameters, Θ

Time Out-I
(First Example – length and age)

- We shall assume that:
  - Growth follows a von Bertalanffy growth equation.
  - t_0 and κ are the same for all animals in the population while the asymptotic size differs among individuals.
  - There is normally distributed measurement error.

Time Out-II
(First Example – length and age)

Now for the general case:

$$\underline{y}_i = f(\boldsymbol{\varphi}, v_{i,j}) + \underline{\varepsilon}_i \qquad \phi_{i,j} = \mathbf{A}_{i,j}\,\underline{\beta} + \mathbf{B}_{i,j}\,\underline{b}_i$$

For the specific case of a von Bertalanffy growth equation:

$$\ell_{i,j} = \ell_{\infty,i}(1 - e^{-\kappa(a_{i,j} - t_0)}) + \varepsilon_{i,j}$$

The response variable

The fixed effects

The random effect

The covariate

Theoretical Aspects

As before:

$$L(\mathbf{y} \mid \underline{\beta}, \underline{\theta}, \sigma^2) = \prod_{i=1}^{M} L(\underline{y}_i \mid \underline{\beta}, \underline{\theta}, \sigma^2)$$

$$= \prod_{i=1}^{M} \int L(\underline{y}_i \mid \underline{\beta}, \underline{b}_i, \sigma^2)\, p(\underline{b}_i \mid \underline{\theta}, \sigma^2)\, d\underline{b}$$

This simplifies to:

$$L(y \mid \underline{\beta}, \underline{\theta}, \sigma^2) = \frac{|\boldsymbol{\Delta}|^M}{(2\pi\sigma^2)^{(N+Mq)/2}} \prod_{i=1}^{M} \int \exp\left\{ -\frac{\left\| \underline{y}_i - f_i(\underline{\beta}, \underline{b}_i, \mathbf{v}) \right\|^2 + \left\| \Delta \underline{b}_i \right\|^2}{2\sigma^2} \right\} d\underline{b}_i$$

The integrals cannot be evaluated analytically and so numerical approximation methods are needed to find the Maximum Likelihood (or REML) estimates for the model parameters.

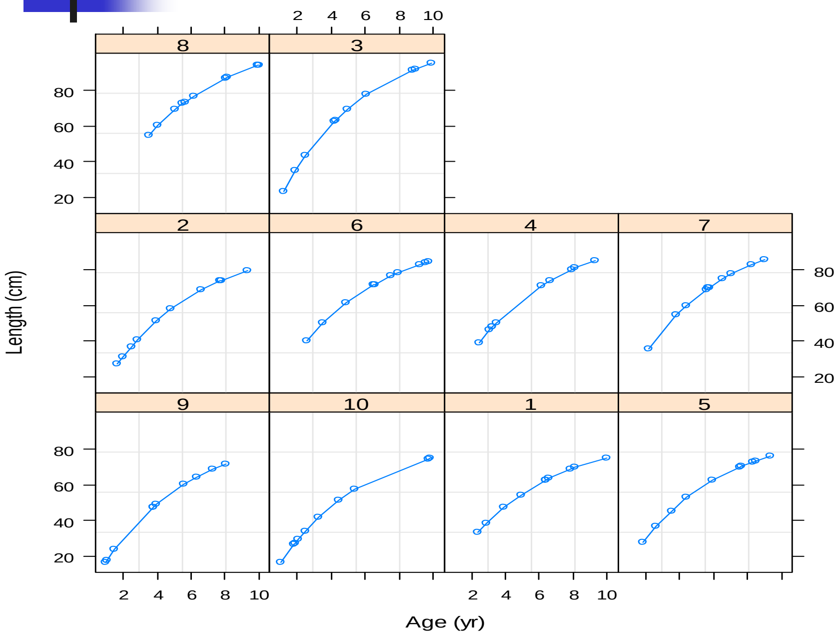
Non-linear Mixed Effects models (Fitting techniques)

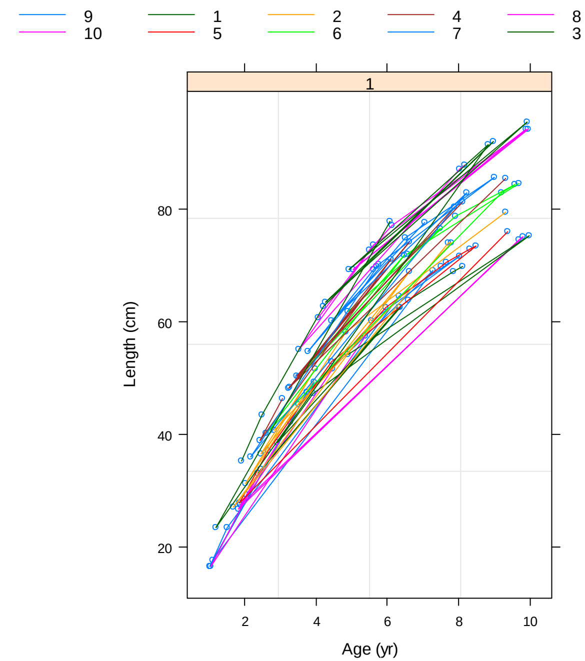
- R can be used to fit non-linear mixed effects models (function NLME). However, SAS appears to be more powerful (e.g. can handle non-normal error distributions).
- To fit non-linear mixed effects models using TMB / EXCEL you need to evaluate the integrally numerically / apply a method such as the Laplace transformation.

Fitting Length and Age Data (Background)

- We have 10 subjects (animals).
- Each animal is measured 10 times between age 1 and age 10.
- Growth follows a von Bertalanffy growth equation where the asymptotic size is subject-specific and the growth rate and age at zero size are the same across subjects.

Fitting Length and Age Data (The basic data)



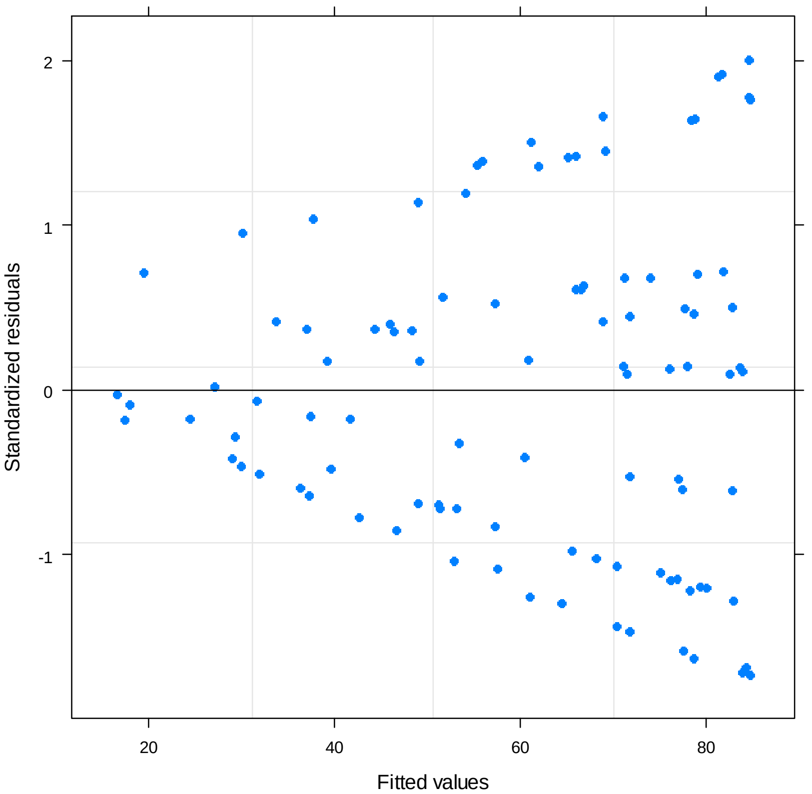



```
FileName <- “LECT2B.TXT"
 TheData <- scan(FileName, what=list(Subject=0,Age=0,Length=0,NULL), skip=1, n=4*100)
xx <- as.data.frame(cbind(Subject=TheData$Subject,Age=TheData$Age,Length=TheData$Length))
AgeLen <- groupedData(Length~Age|Subject, data=as.data.frame(TheData),
labels=list(x="Age",y="Length"), units=list(x="(yr)",y="(cm)"))
plot(AgeLen)
plot(AgeLen,outer=~1)
```

Fitting Length and Age Data (Standard non-linear model-I)

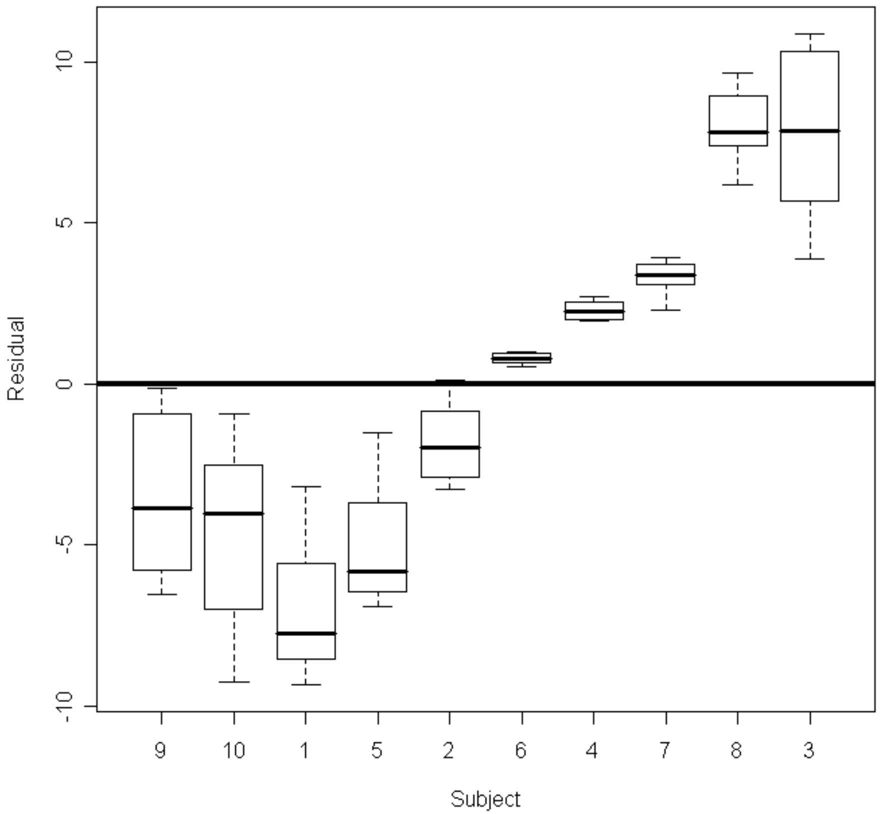
```
lm1 <- nls(formula=Length~Linf*(1-exp(-1*Kappa*(Age-Tzero))),
      data=AgeLen, start=c(Linf=100,Kappa=0.2,Tzero=0))
```

There is evidence in the residuals for model-mispecification. This is not heteroscedastic errors -rather it is correlation among observations for the same animal.



Fitting Length and Age Data (Standard non-linear model-II)

boxplot(split(residuals(lm1),AgeLen$Subject),ylab="Residual", xlab="Subject",csi=0.2)

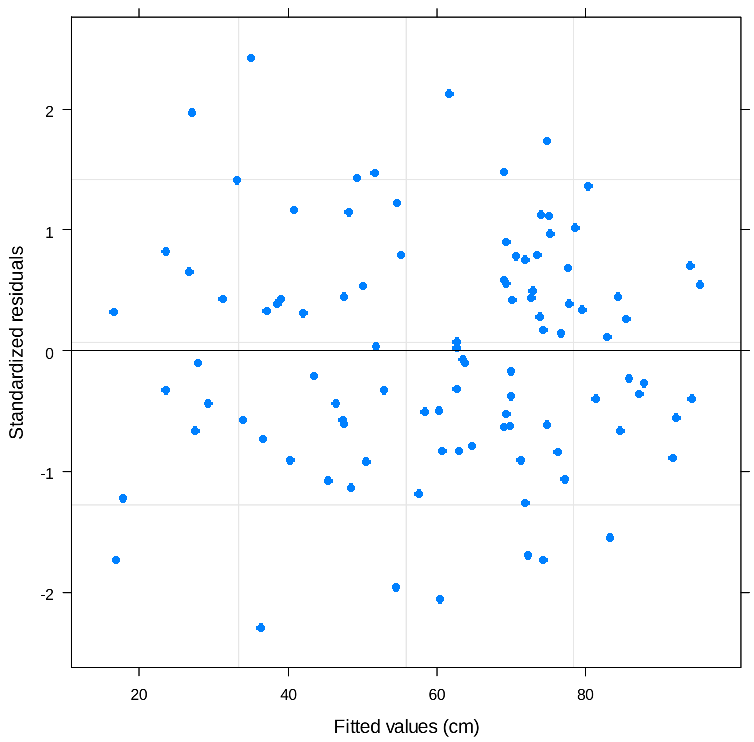


Fitting Length and Age Data (Non-linear mixed model-I)

```
lm2 <- nlme(model=Length~Linf*(1-exp(-1*Kappa*(Age-Tzero))), data=AgeLen,
    random=Linf~1, fixed=Linf+Kappa+Tzero~1,start=c(Linf=100,Kappa=0.2,Tzero=0))
Print(lm1$modelStruct)
```

Random effect:
- Linf by subject

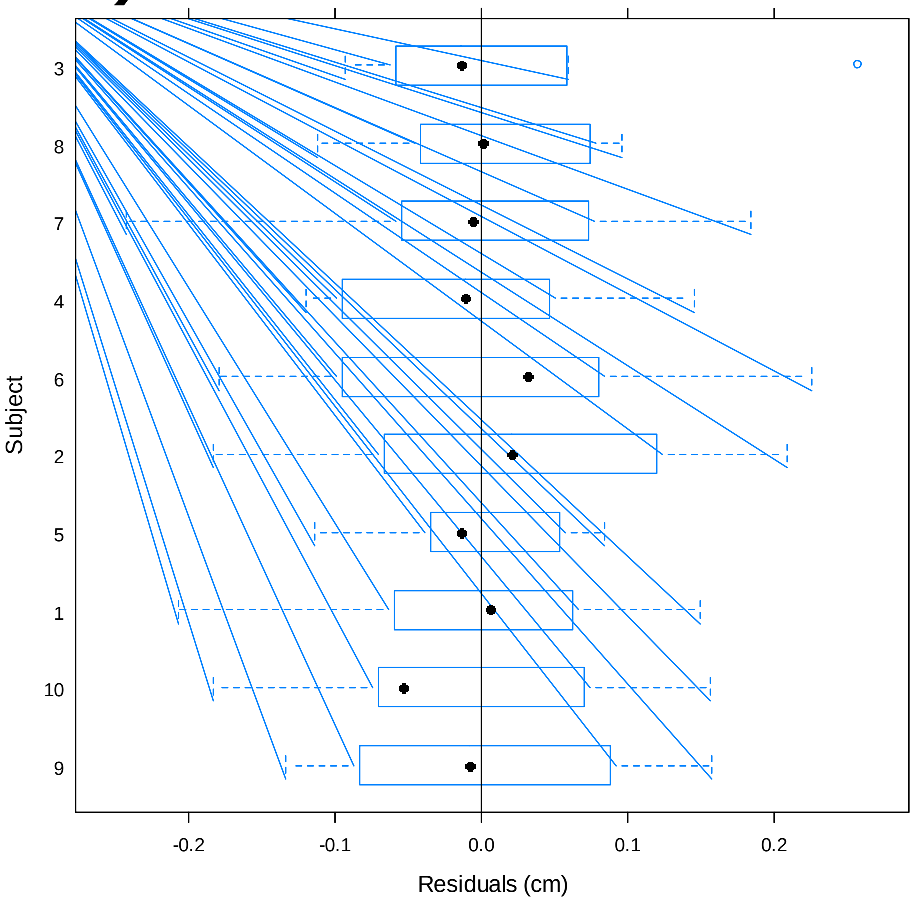
Fixed effects:
- Linf population mean
- Kappa
- t_0



Fitting Length and Age Data (Non-linear mixed model-II)

plot(lm2,Subject ~ resid(.),abline=0)

Try "intervals (lm2)" to see the confidence intervals

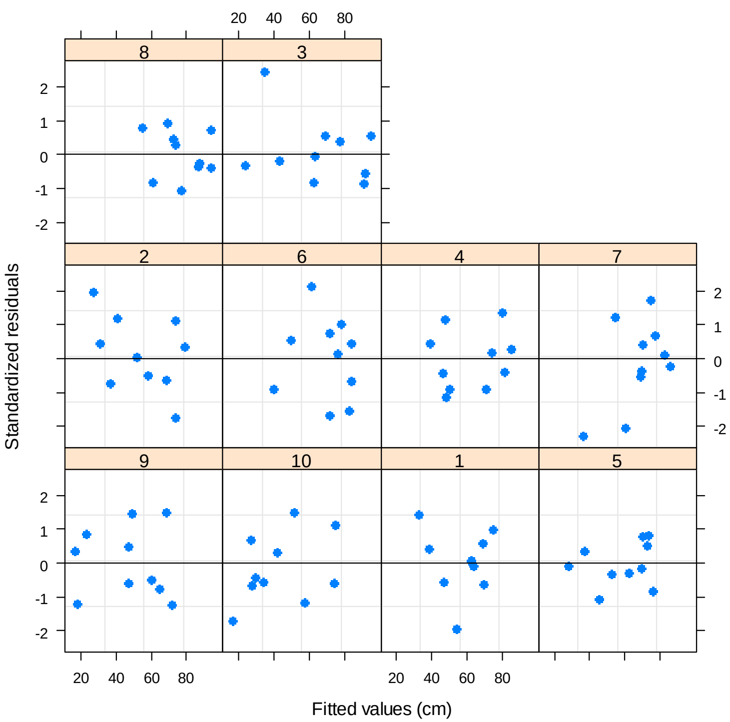


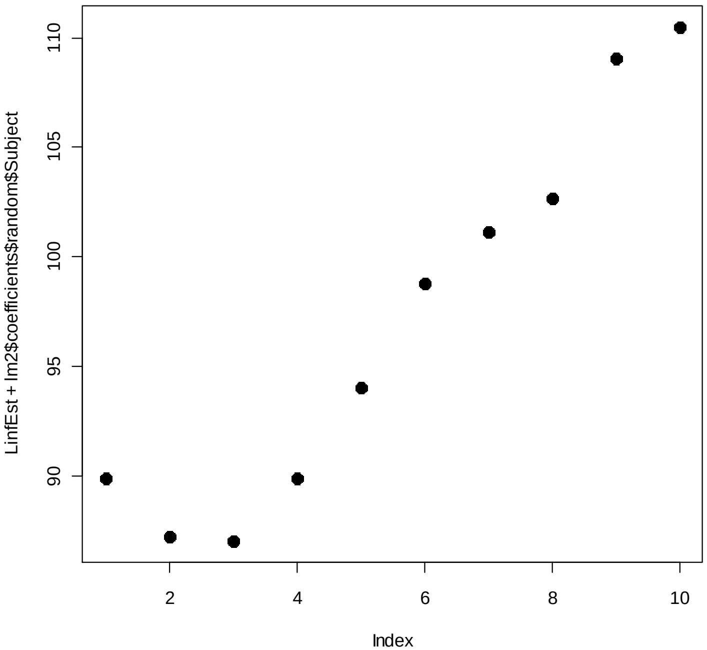
Fitting Length and Age Data (Non-linear mixed model-III)

```
Nonlinear mixed-effects model fit by maximum likelihood
  Model: Length ~ Linf * (1 - exp(-1 * Kappa * (Age - Tzero)))
 Data: AgeLen
       AIC       BIC   logLik
  -53.64647 -40.62062 31.82323

Random effects:
 Formula: Linf ~ 1 | Subject
           Linf  Residual
StdDev: 8.28837 0.1060163

Fixed effects: Linf + Kappa + Tzero ~ 1
         Value Std.Error DF  t-value p-value
Linf  97.02240 2.6629267 88  36.4345  0.0000
Kappa  0.19960 0.0005324 88 374.8752  0.0000
Tzero -0.00453 0.0040570 88  -1.1176  0.2668
 Correlation:
      Linf   Kappa
Kappa -0.034
Tzero -0.027  0.876
```

Fitting Length and Age Data (Non-linear mixed model-IV)

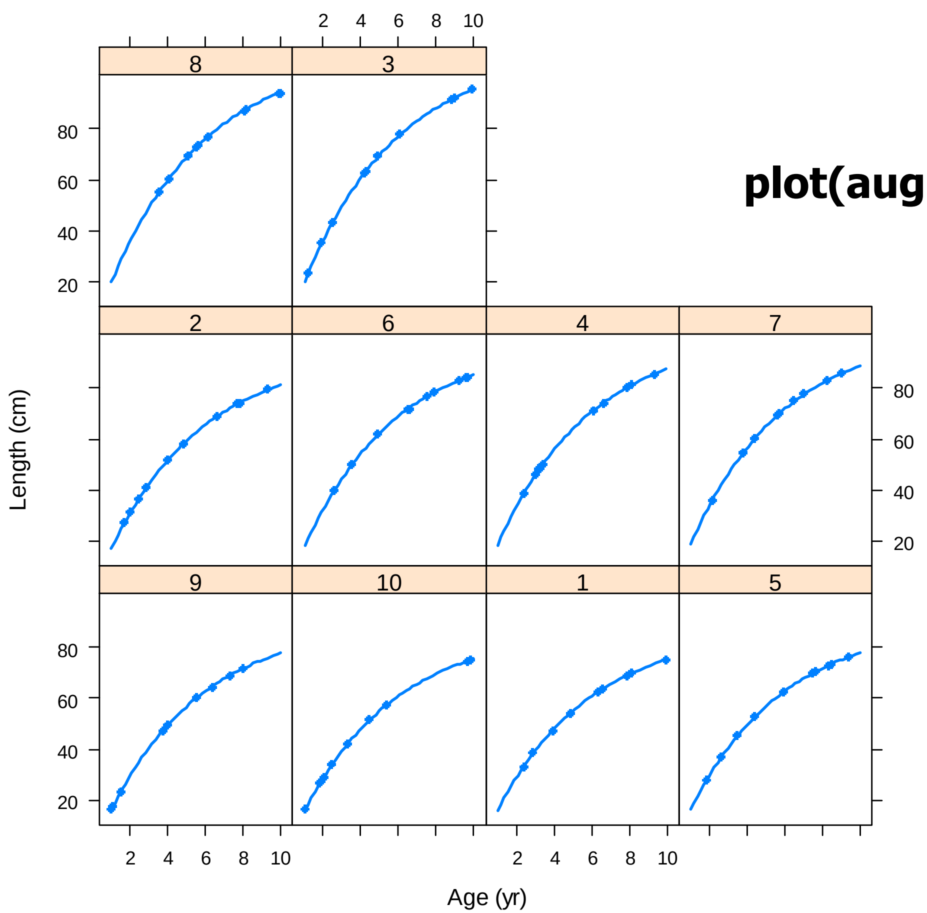
plot(lm2, form = resid(., type = "p") ~ fitted(.) | Subject, abline = 0,pch=16)





plot(LinfEst+lm2$coefficients$random$Subject, pch=16,csi=0.3)

Fitting Length and Age Data (Non-linear mixed model-V)



plot(augPred(lm2),csi=0.3,pch=16,lwd=2)

Plots the data and the model predictions

Non-linear Mixed Models
(Population Average estimates-I)

- If you fit a mixed effects model and need to know the value of some (non-linear) function averaged across the population, you have to integrate over the random effects, i.e.:

$$E[H(\underline{\beta})] = \int H(\underline{\beta}, \underline{b}) d\underline{b}$$

- For a linear mixed effects model, this integration is not needed – why?

Non-linear Mixed Models
(Population Average estimates-II)

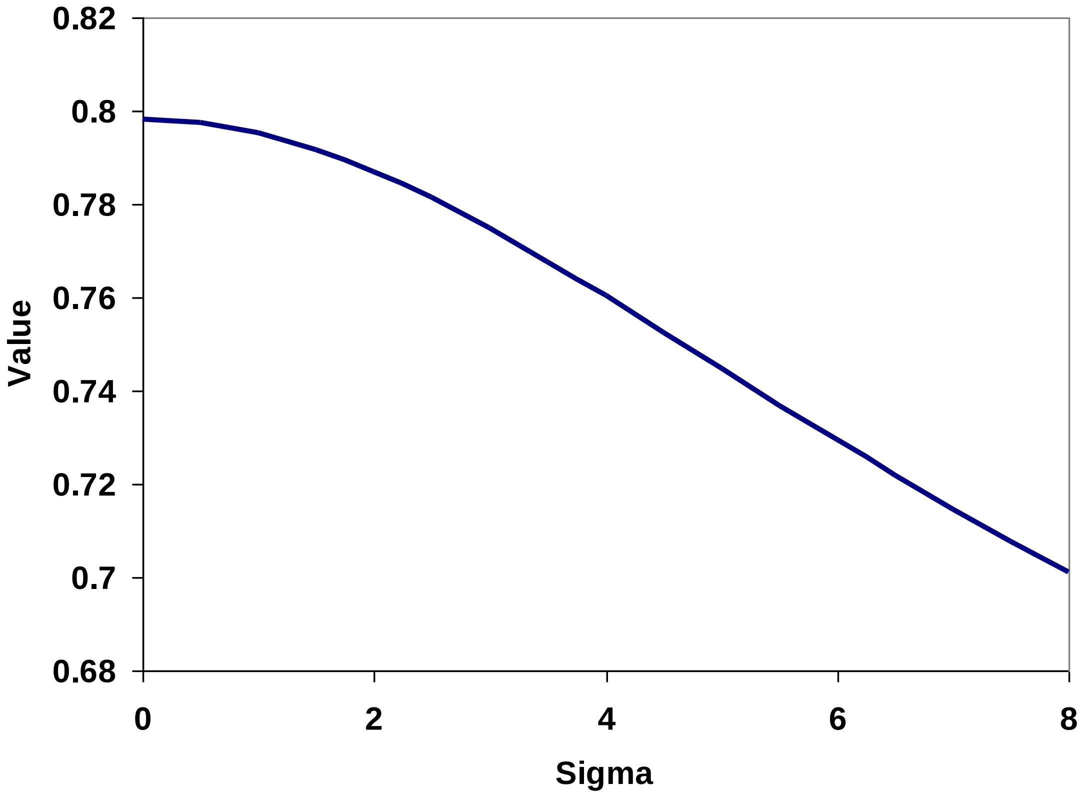
- Consider the case in which selectivity-at-age has been computed from data collected from *n* sites and the parameters estimated by fitting a mixed effects model, i.e.:

$$S_a = [1 + \exp\{-1 * (x - (5 + b_i)) / 4\}]^{-1}; \qquad b \sim N(0; \sigma^2)$$

- The quantity of interest is the selectivity for age 5.5.
- How does this age change with the value of σ.

Non-linear Mixed Models
(Population Average estimates-III)

The selectivity of age 5.5 decreases with increasing σ



The NLME Function-I

- The NLME function has several options:

nlme(model, data, fixed, random, groups, start, correlation, weights, subset, method,na.action,naPattern, control,verbose)

- model: "response ~fn(fixed effects)"
  - you can also supply a function (possibly with its derivatives) – this will be discussed when we deal with non-linear regression methods.
- fixed: "fixed = c(fix1~var1,fix2~var2)".
  - There are several ways to specify the fixed effects structure.

The NLME Function-II

nlme(model, data, fixed, random, groups, start, correlation, weights, subset, method, control)

- random: "random=c(rnd1 ~ var1, rnd1 ~ 1)". Leaving out a "random" specification implies all fixed effects have an associated random effect.
- start: "start = c(fix1=x1,fix2=x2,..)" – the initial values for the fixed parameters of the model.

- Be wary of using NLME to make predictions when some of the parameters are non-linear.

The NLME Function-III

You can use many of the functions defined for linear models with nlme, e.g.:

- update
- logLik
- plot
- predict
- intervals
- anova
- residuals

References

- Bolker et al. (2008). Generalized linear mixed models: a practical guide for ecology and evolution. TREE: 24: 127-35.
- Pinheiro, J.C. & Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag.
- Zuur *et al*. (2009). *Mixed-effects Models and Extensions in Ecology with R*. New York: Springer-Verlag.