

# Modeling random effects and mixed effect models part I

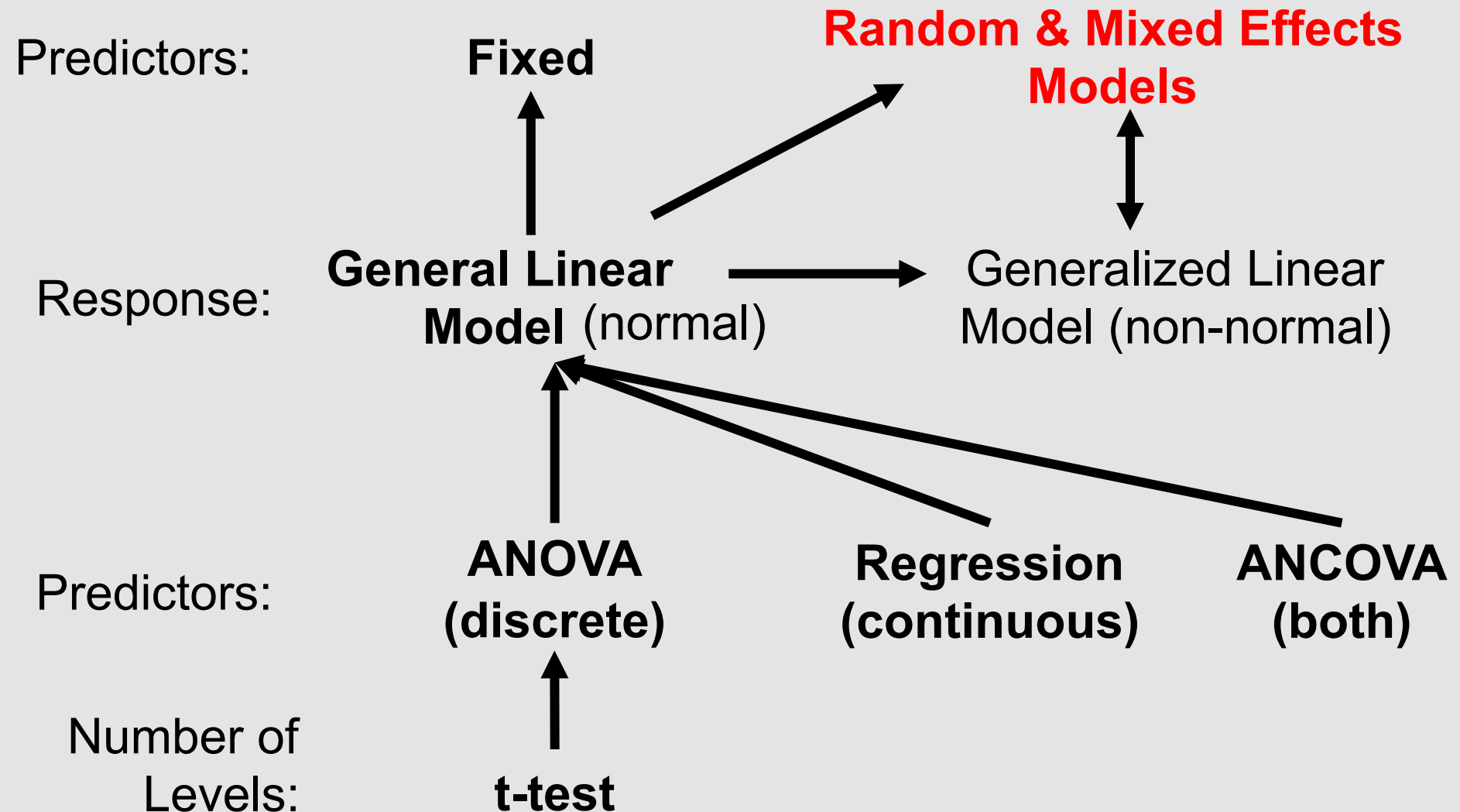
SMEE ZOL851

Tuesday Nov 27<sup>th</sup> 2012

$$Y \sim N(\mathbf{X}\beta, \Sigma)$$

# Continuity of Statistical Approaches

## Process Models



# Readings

In the Mixed model folder in ANGEL I have put a number of potentially useful readings:

Bolker book: pp. 316-328

Gelman and Hill: Chapters 11-13 (Start with 11)

Hadfield: Course notes for Mixed Models using MCMCglmm

Some papers

Bolker et al. 2008. Generalized linear mixed models: a practical guide for ecology and evolution. TREE.

Bates, D. [R] lmer, p-values and all that. (Why p-values are can be so difficult to determine for mixed models, along with df).

# Readings

In addition there are a number of books which may be useful to you.

Pinheiro, J. & Bates, D. 2000. Mixed-effects models in S and S-Plus. (E-book , link available in the resource section).

Zuur, A. et al. 2009. Mixed effects Models and Extensions in Ecology with R.

Note: While this book has a lot of potential, it is also full of errors. So please use additional sources for understanding the basics.

# The problem

- So far we have assumed that are observations are IID.
- Identically distributed.: The observations come from the same underlying error distribution.
- Independently distributed: The observations are all independent of one another , i.e.  $\text{cov}(y_i, y_j) = 0$

# The problem

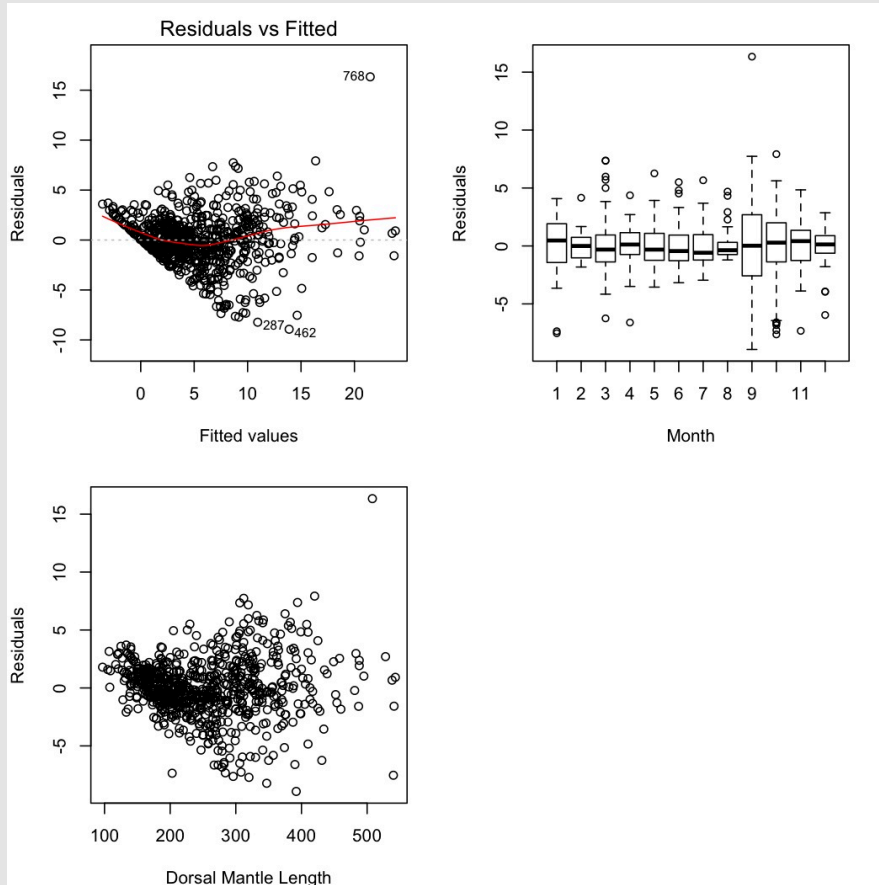
- However both of these assumptions are violated with many data sets, and can lead to many issues with estimation and inference.
- Thus we need to stop thinking about the residual error as the “dumping ground” for unfit variance, and start incorporating structure into this part of the model as well!!

# Outline

- Heterogeneity in residual variances
- Replication and Independence of errors
- Accounting for grouping in observations
- Fixed vs Random effects
- The addition of random effects to our general linear model
  - Familiar Approaches
    - Blocking factor
    - Repeated measures
    - Nested Designs
- Mixed-Effects Models in R

# Heterogeneity in variance

Fitting a model for relating squid size and month to testis size as a measure of sexual maturity.





# Heterogeneity in variance

- When dealing with such heterogeneity we want to model the residual variance.

$$\varepsilon_i = N(0, \sigma^2 \times DML_i)$$

A fixed variance structure where the variance is multiplied by one of the continuous variables.

```
design.M.1 <- model.matrix(~ DML, data=Squid) # Design matrix

NLL.M.1.HET.VAR <- function(b0, b1, sig){
  het.sig <- sig*sqrt(design.M.1[,2]) # sigma as a function
  det <- b0 + b1*design.M.1[,2] # deterministic part like before.
  -sum(dnorm(Squid$Testisweight, mean=det, sd=het.sig, log=T))}

M.1.Het <- mle2(NLL.M.1.HET.VAR , start=list(b0=0, b1=0, sig=0.01))
```

# Heterogeneity in variance

- When dealing with such heterogeneity we want to model the residual variance.

$\varepsilon_i = N(0, \sigma^2 \times DML_i)$     A fixed variance structure where the variance is multiplied by one of the continuous variables.

In R we would use the `gls()` in the `nlme` library

```
vf1Fixed <- varFixed(~DML) #specifying the variance structure
M.1.gls <- gls(Testisweight ~ DML*fMONTH, weights = vf1Fixed,
data=Squid) # only difference is incorporating it into the
model.
```

# Heterogeneity in variance

- There are many other variance structures that can be used

$$\varepsilon_{ij} = N(0, \sigma^2_j)$$

Different variance for each treatment level of a factor. varIdent

In R we would use the gls() in the nlme library

```
vf2 <- varIdent(~ 1| fMONTH) #specifying the variance structure
M.1.gls <- gls(Testisweight ~ DML*fMONTH, weights = vf2,
data=Squid) # only difference is incorporating it into the
model.
```

# Heterogeneity in variance

- There are many other variance structures that can be used

Name of the function in nlme	What it does
VarFixed	Fixed variance based on the covariate
VarIdent	Different variances per treatment levels
VarPower	Power of the variance covariate
VarExp	Exponential of variance covariate
VarConstPower	Constant + power of variance covariate
VarComb	combination

After Zuur et al. 2009

# Heterogeneity in variance

- There are many other variance structures that can be used
- For non-linear models there is the `gnls()` function in `nlme`
- `?varClasses`

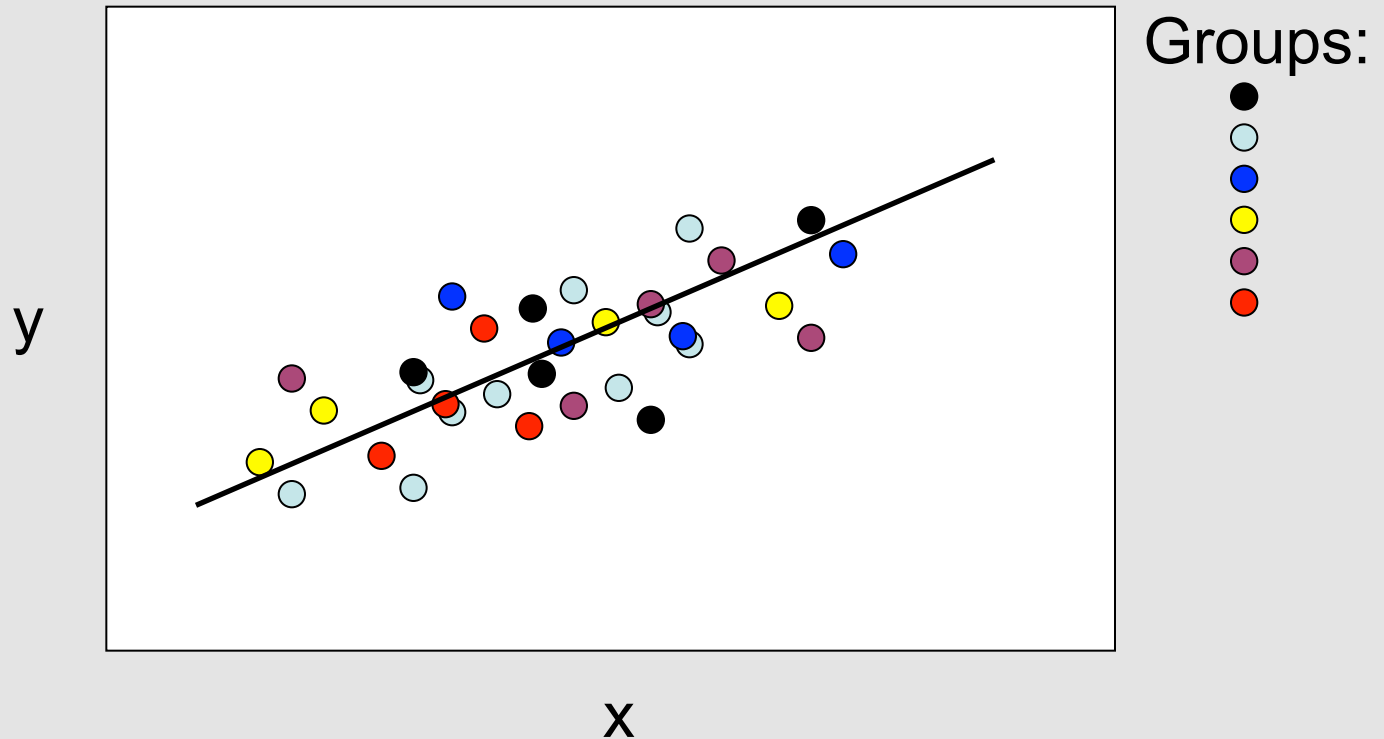
# Replication and Independence

- Statistics is based on collecting independent samples from a population. The degree to which those samples reflect the true population depends on the number of samples collected
- So, all inferential statistics depend on independence
- Replicates must be more than just unbiased, they must be independent
  - Pairs of samples treated alike are on average no more similar or dissimilar than those treated differently - from Hurlbert, 1984
    - Shading may cause plants to grow taller and two plants experiencing the same degree of shading might both be tall but they are independent if the source of the shading differs between the two plants

# Independence of Errors

- Independence of errors is the key assumption
- Lack of independence comes from some sort of “**grouping**” of the data.
- Examples:
  - Multiple measurements made per individual
  - Observations clustered in space or time
  - Treatment randomly applied to groups of replicates rather than individual replicates
  - Some individuals come from the same genetic strain

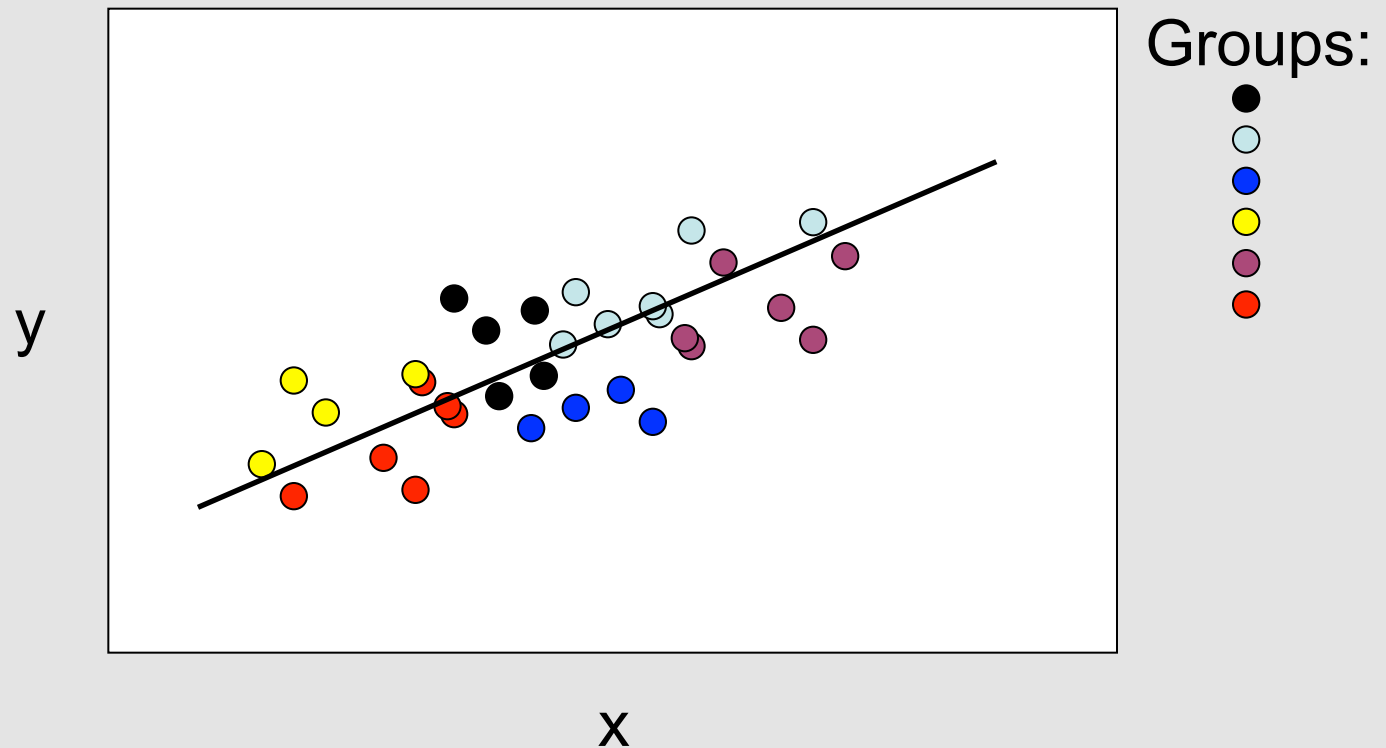
# Underlying Structure



Errors are independent of group structure



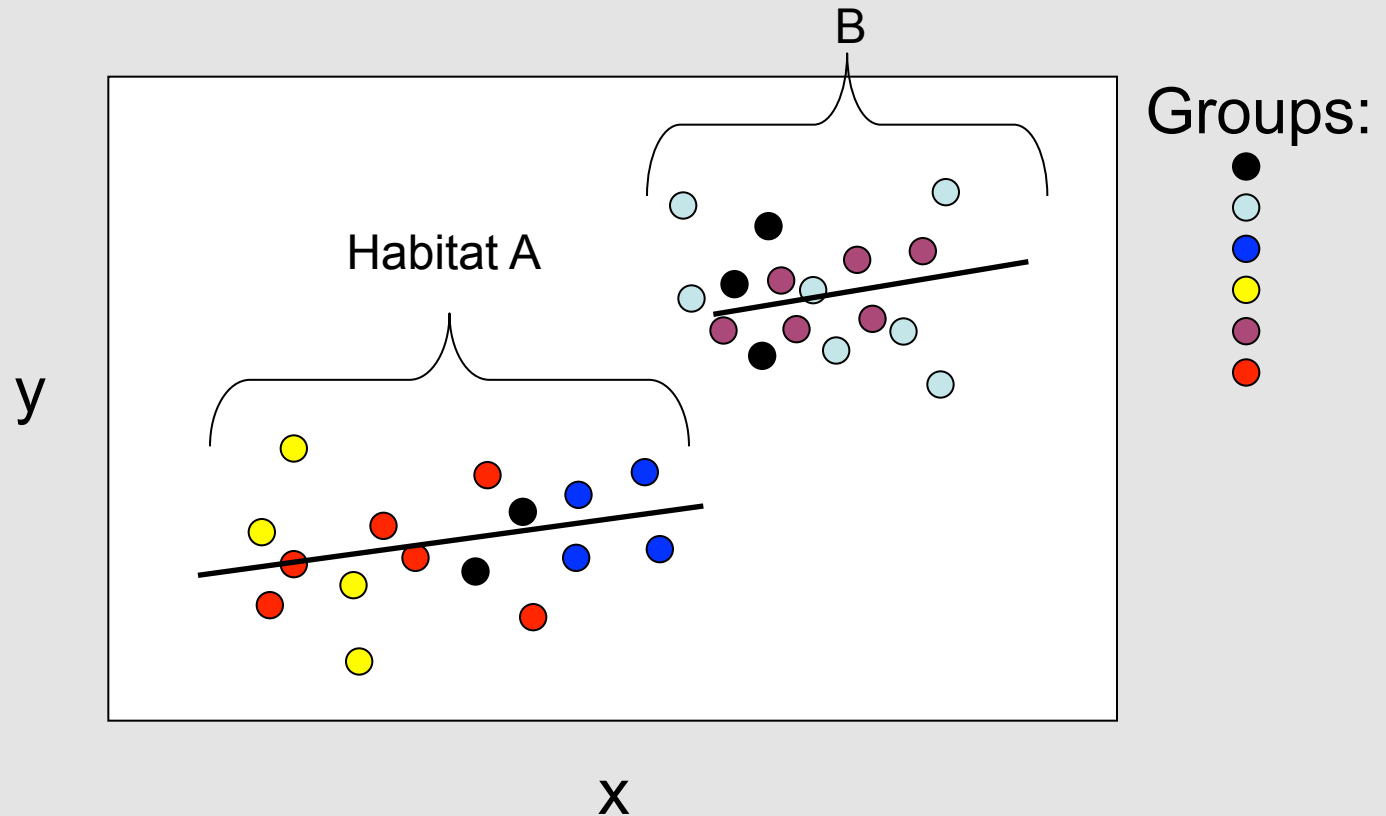
# Underlying Structure



# Accounting for Grouping

- Samples might not be independent but we can and need to account for this lack of independence in our analysis.
- Independence as “virtue” and “truth”

# Underlying Structure



# Covariance structures for time series

Under the assumption of independence we assume that there is no covariance between years.

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

$$Var(\varepsilon) = \sigma_{\varepsilon}^2 \times \mathbf{I} = \begin{bmatrix} \sigma_{\varepsilon}^2 & 0 & 0 & 0 \\ 0 & . & 0 & 0 \\ 0 & 0 & . & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon}^2 \end{bmatrix}$$

For n observations

# Covariance structures for time series

- Clearly this assumption is not valid so we should try some other types of covariance structures.
- The simplest is the compound symmetry structure
- Assumes that there is a constant co-variance between errors, regardless of time.

# Compound symmetry

- The simplest is the compound symmetry structure
- Assumes that there is a constant co-variance between errors, regardless of time.

$$Var(\varepsilon) = \begin{matrix} & \theta + \sigma_{\varepsilon}^2 & \theta & \theta & \theta \\ \begin{matrix} \theta \\ \theta \\ \theta \\ \theta \end{matrix} & \begin{matrix} \theta \\ \theta \\ \theta \\ \theta \end{matrix} & \begin{matrix} . \\ \theta \\ \theta \\ \theta \end{matrix} & \begin{matrix} \theta \\ \theta \\ . \\ \theta \end{matrix} & \begin{matrix} \theta \\ \theta \\ \theta \\ \theta + \sigma_{\varepsilon}^2 \end{matrix} \end{matrix}$$

```
# Compound symmetry structure in R using gls()  
M.1 <- gls(Birds ~ Rainfall + Year, data=Hawaii,  
correlation=corCompSymm(form =~Year) )
```

This adds just one more parameter, theta. It is still pretty simplistic though.

# Auto-regressive (AR)-1 auto-correlation

- Auto-regressive order model 1
- Models residual at time  $s$  as a function of the residual of time  $s-1$  plus error.

$$Cor(\varepsilon) = \begin{matrix} & 1 & \rho^1 & \rho^2 & \rho^3 \\ \begin{matrix} \rho^1 \\ \rho^2 \\ \rho^3 \end{matrix} & \begin{matrix} \rho^1 \\ \rho^2 \\ \rho^3 \end{matrix} & \begin{matrix} . \\ \rho^1 \\ \rho^2 \end{matrix} & \begin{matrix} \rho^1 \\ . \\ \rho^1 \end{matrix} & \begin{matrix} \rho^2 \\ \rho^1 \\ 1 \end{matrix} \end{matrix}$$

```
# Compound symmetry structure in R using gls()  
M.1 <- gls(Birds ~ Rainfall + Year, data=Hawaii,  
correlation=corAR1(form =~Year) )
```

The further you are in time, the less covariation.

# ARMA

- Auto-regressive moving average.. Yet more complex models...





# Mixed-Effects Models

- General (-ized) Linear Models that contain both fixed and random effects
- So far we have considered only fixed effects
- Models to predict the response to some predictors when the data are grouped according to one or more classification variables (random effects)

# Random & Mixed-effects models

- Such models are also referred to as multilevel models (mostly in the social sciences and psychology, but increasingly in biology).
- Also called hierarchical models (which will often refer to a nesting structure).

There may actually be some subtle distinctions implied by these names, but the amount of variation in how people use these terms is far greater than the variation that may be implied by the models...

# Fixed effects VS. Random effects

- Break into small groups and come up with a definition of each.

# Fixed VS. Random effects

- Break into small groups and come up with a definition of each.

# Fixed VS. Random effects:

## Some definitions

- ***Fixed*** effects are constant across individuals, and ***random*** effects vary. (Kreft & De Leeuw 1998 p.12).
- Effects are ***fixed*** if they are interesting in themselves, or ***random*** if there is interest in the underlying population (Searle, Casella & McCulloch 1992).
- When a sample (close to) exhausts the population, the corresponding variable is ***fixed***; when the sample is a small part of the population the corresponding variable is ***random*** (Green and Tukey 1960).

# Fixed VS. Random effects: Some definitions

- If an effect is assumed to be a realized value of a random variable, it is called a ***random*** effect (LaMotte 1983).
- ***Fixed*** effects are estimated using LS or ML and ***random*** effects are estimated with shrinkage (“linear unbiased predictions”). This is the common definition for multilevel models.

# Fixed vs. Random Factors

## Fixed

- Specific levels of factor have some interest / significance
- Levels are fixed and repeatable
- Interested in estimating the means for these levels
- Examples: experimental treatment, habitat, age class, sex, species, year?

## Random

- Levels of factor do not have significance
- Repeatability of levels isn't important
- Levels “drawn” at random from a population of possible levels
- Interested in measuring or correcting for variation among levels
- Examples: Subject, plot, year?



Thus one of the most difficult parts about mixed effect models is figuring out which variables should be fixed and which should be random.

Gelman & Hill choose not to make a distinction in most cases. They treat all variables as random (which is consistent with a Bayesian philosophy of estimation).

# Examples of Common Designs with Random Effect

## 1. Blocking Factors

- Lump similar (usually spatially close) plots together
- Analogous to the pairwise t-test
- Called “within-subject” in psychology
  - Also equivalent to giving same plant two different treatments and measuring response each time
- Pulls variation into the block and out of the explained error term ( $\varepsilon$ )

Example: Interested in primary production. 5 fields in different locations. Randomly apply treatments within fields. Include “Field” as a blocking factor to account for variation among fields

# Examples of Common Designs with Random Effect

## 2. Repeated Measures - ANOVAR

- Also called within-subject, longitudinal study
- Record multiple measures on an individual under different conditions
- Similar to blocking but the “plot” is an individual organism and the blocking factor is time
- Many scenarios:
  - Using same individuals to reduce variance
  - Growth, recruitment or other variable changing over time
  - Before/After

Example: Measure body temperature in mice when non-reproductive, pregnant and lactating.

# Examples of Common Designs with Random Effect

## 3. Nested Designs

- Exploring variation with spatial scale
- Subplots within plots within sites
- When each level of factor A is only observed in one level of a different factor B then A is nested within B

### Examples:

- Growth of sea lamprey larvae within streams within lakes.
- Growth of nestlings within clutches within years
- Primary production within subplots within exclosures

Let's remind ourselves of the sex comb teeth data set from *Drosophila*.

- Trying to account for variation in number of SCT on male flies.
- We had numerous factors we were looking at:
- Genotype (2 treatment levels: *Dll* and *wt*).
- Temperature (2 treatment levels: 25 & 30°C)
- Line ( strain of flies: 27 levels)

Let's remind ourselves of the sex comb teeth data set from *Drosophila*.

- Work in groups to figure out which variables you think should be fixed or random, and **why**.

# Benefits of Mixed-Effects Models

- Allow you to correct for lack of independence of errors
  - Account for underlying group structure in the data
- Handle balanced and unbalanced designs with missing data.
- Utilizes all of the data when making “group” level predictors (I.e. the “best linear unbiased predictions, BLUPs” for random co-efficients).





# Let's look at some ways to write a pure random effects models.

In the simplest model we one term that some consider “fixed”, the overall population mean (or intercept, although we will see that term gets confusing).

$$y = a + e \quad e \sim N(0, \sigma_e^2) \quad \text{Just a population mean}$$

or

$$y \sim N(\mu_a, \sigma^2)$$

# Let's look at some ways to write a pure random effects models.

In the simplest Random effects model we have one term that some consider “fixed”, the overall population mean (or intercept, although we will see that term gets confusing). We also have a single **random** (grouping) variable.

$$y = \mu + a_j + e$$
$$e \sim N(0, \sigma_e^2)$$
$$a_j \sim N(\mu_a, \sigma_a^2)$$

Random effect

$$y_{ij} \sim N(\mu + a_j, \sigma_e^2)$$
$$a_j \sim N(0, \sigma_a^2)$$

The important point is that we are modeling some random variables, so that there are additional sources of variance.

In general these can all be combined into a model like

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

X & Z are incidence (X is still called the design matrix)  
Beta and u are what we are trying to estimate.

$$y_i \sim N(X_i\boldsymbol{\beta} + a_{j[i]}, \sigma_e^2)$$

$$a_j \sim N(0, \sigma_a^2)$$

# Sex comb teeth in *Drosophila* example.

In the simplest case let's say way we wanted to generate a pure random effects model where we estimate the varying intercepts model for line effects.



# Mixed Effects Models in R

- Using mixed models requires more complicated calculation approach
  - SAS uses PROC MIXED, glmmixed, nlmixed
  - R has MANY options
    - *lme()* is an older function from the *nlme* package that works well
    - *lmer()* is a new function in the package *lme4*
    - *lmer* handles both nested and non-nested random factors as well as non-normal error terms
    - MCMCglmm() – which we will use.
- Using *lmer*
  - *MCMCglmm()*, *lmer()* work in a similar fashion to *lm* and *glm*
    - Takes formula
    - Gives object that goes to summary, print, anova, etc.
    - Specify model in the same way, but random effects are indicated slightly differently
    - e.g. `Y ~ Fixed.1 + Fixed.2 + (1|Random.1)` *lmer*
    - `Y ~ Fixed.1. + Fixed.2, random=~ Random.1`

# How do we code this in R?

- Load lme4 package (install first)
- Start with original linear model
- Add random effect for “line”

```
>model2<-lmer(SCT~1 +(1|line), data=dll.data)
```

```
>model2 <- MCMCglmm(SCT ~1, random=~line, data=dll.data)
```

This model is conceptually similar to

```
>model3<-lm(SCT~ 0 + line, data=dll.data)
```


Why? Let's dissect this (on the board)



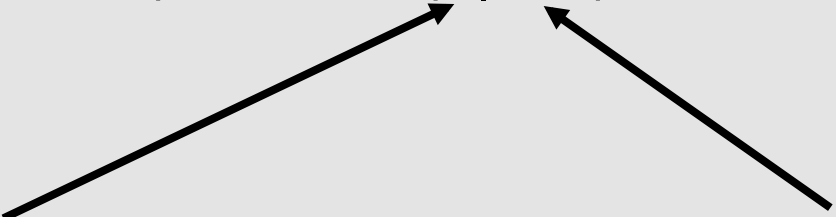
Coding models in lmer() (from lme4 library)

## How do we code this in R?

Population intercept



```
>model2<-lmer(SCT~1 +(1|line), data=dll.data)
```



This 1 says are grouping variable (line), only needs to vary for the intercept term (the 1).

Coding models in MCMCglmm()

## How do we code this in R?

Population intercept



```
>model2.MCMC<- MCMCglmm(SCT~1, random=~line, data=dll.data)
```



This says the grouping variable (line), only needs to vary for the implicit intercept term .

```
> model2.MCMC<- MCMCglmm(SCT~1, random=~us(1):line, data=dll.data)
```

```
> model2.MCMC<- MCMCglmm(SCT~1, random=~idh(1):line, data=dll.data)
```

All three of these are identical models, we will discuss why soon.

# The original package for such models was nlme, and it is still useful for a variety of things.

- It specifies things in a *similar* manner to MCMCglmm...

```
require(nlme) # original mixed model library for R/S  
lme(SCT ~ 1, data=dll.data, random=~ line)
```

# How might we code a more complicated model

- Let's say we wanted to model the effects of tarsus, and account for line level variation? How might we do it?

Add some figures to help this along plots of lines with varying slopes and intercepts

Coding models in MCMCglmm()

## How do we code this in R?

What you probably **DO NOT** want is:

```
>model2.MCMC<- MCMCglmm(SCT~1 + tarsus, random=~sct:line,  
data=dll.data)
```

# Coding models in MCMCglmm() and lme4

## How do we code this in R?

More likely you want to consider one of these.

```
model2.MCMC<- MCMCglmm(SCT~1 + tarsus,  
  random=~us(1 + tarsus):line, data=dll.data)
```

```
model2.lmer<- lmer(SCT~1 + tarsus + (1 + tarsus| line),  
  data=dll.data)
```

These two are  
equivalent  
representations

Or this:

```
model2.MCMC<- MCMCglmm(SCT~1+ tarsus,  
  random=~idh(1 + tarsus):line, data=dll.data)
```

```
model2.lmer<- lmer(SCT ~ 1 + tarsus  
  + (1 | line) + (0 + tarsus| line), data=dll.data)
```

These two are  
equivalent  
representations

These differ in whether they fit the co-variances between the parameter estimates for random effects

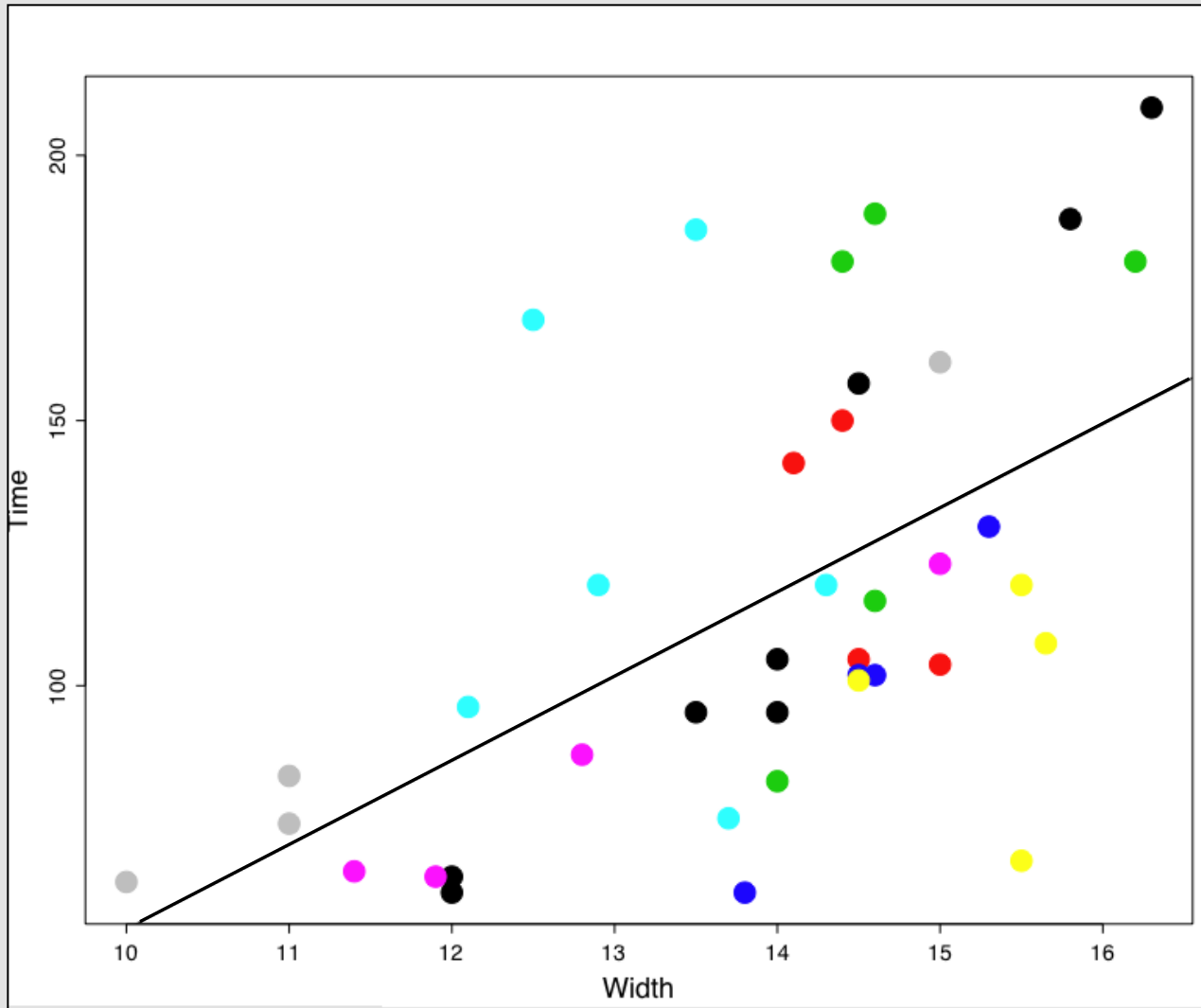
# Variance structures

lmer	MCMCglmm/asreml	No. Parameters	Variance	Correlation
(1 dam)	dam	1	$\begin{bmatrix} V & V & V \\ V & V & V \\ V & V & V \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
(sex-1 dam)	us(sex):dam	6	$\begin{bmatrix} V_{1,1} & C_{1,2} & C_{1,3} \\ C_{1,2} & V_{2,2} & C_{2,3} \\ C_{1,3} & C_{2,3} & V_{3,3} \end{bmatrix}$	$\begin{bmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{1,2} & 1 & r_{2,3} \\ r_{1,3} & r_{2,3} & 1 \end{bmatrix}$
(1 sex:dam)	sex:dam	1	$\begin{bmatrix} V & 0 & 0 \\ 0 & V & 0 \\ 0 & 0 & V \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
(1 dam)+(1 sex:dam)	dam+sex:dam	2	$\begin{bmatrix} V_1 + V_2 & V_1 & V_1 \\ V_1 & V_1 + V_2 & V_1 \\ V_1 & V_1 & V_1 + V_2 \end{bmatrix}$	$\begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix}$
-	idh(sex):dam	3	$\begin{bmatrix} V_{1,1} & 0 & 0 \\ 0 & V_{2,2} & 0 \\ 0 & 0 & V_{3,3} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
-	corh(sex):dam	4	$\begin{bmatrix} V_{1,1} & rV_{1,1}V_{2,2} & rV_{1,1}V_{2,2} \\ rV_{1,1}V_{2,2} & V_{2,2} & rV_{2,2}V_{2,3} \\ rV_{1,1}V_{3,3} & rV_{2,2}V_{3,3} & V_{3,3} \end{bmatrix}$	$\begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix}$
-	cor(sex):dam	3	$\begin{bmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{1,2} & 1 & r_{2,3} \\ r_{1,3} & r_{2,3} & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{1,2} & 1 & r_{2,3} \\ r_{1,3} & r_{2,3} & 1 \end{bmatrix}$

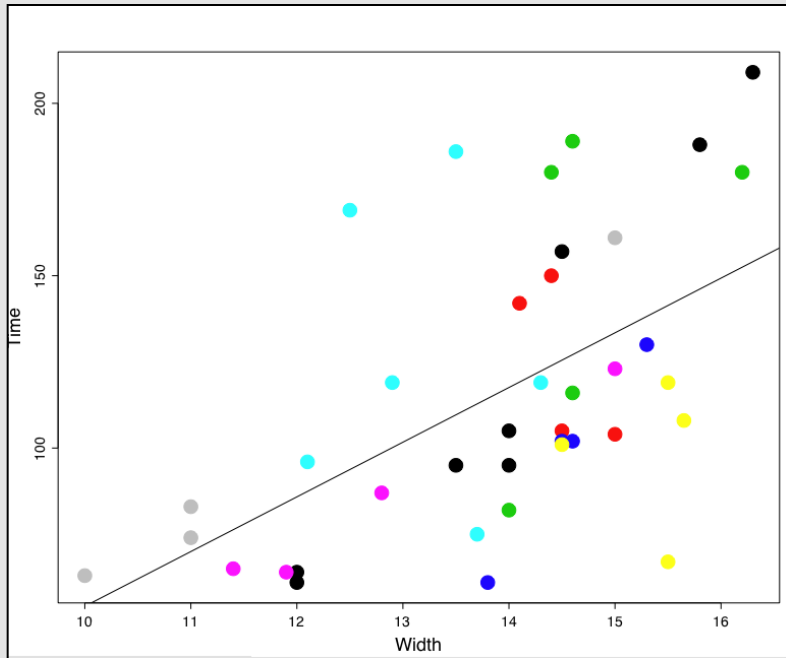
From Hadfield 2010, pg 68, course notes for MCMCglmm



# Squirrel Cone Handling



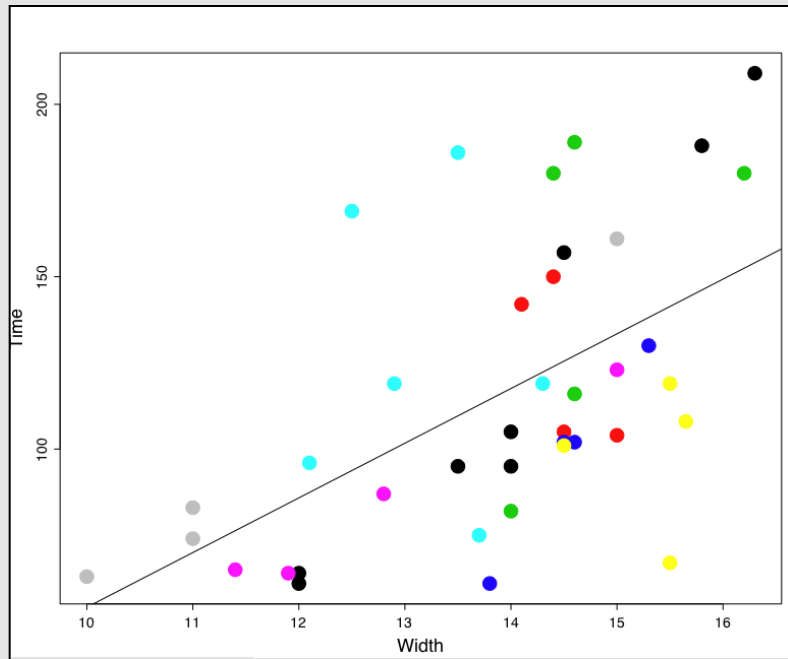
# Regression Model



$$Time_{ij} = \beta_0 + \beta_1 Width + \varepsilon_{ij}$$

Errors ( $\varepsilon_{ij}$ ) are not independent

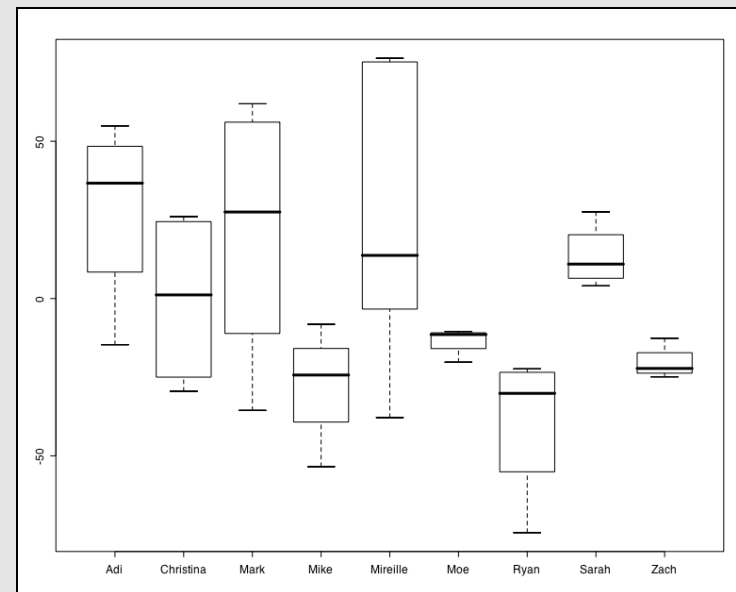
# Regression Model



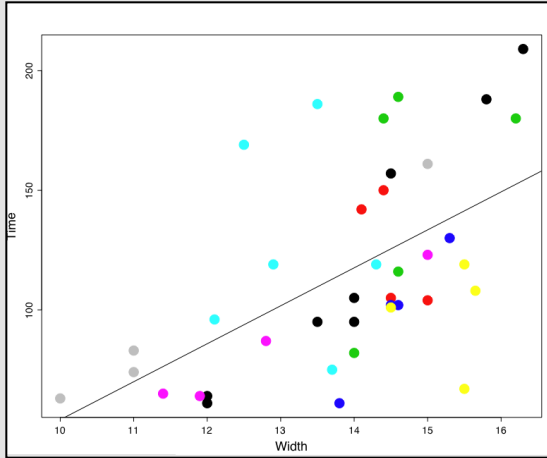
$$Time_{ij} = \beta_0 + \beta_1 Width + \varepsilon_{ij}$$

Errors ( $\varepsilon_{ij}$ ) are not independent

Account for this lack of independence by including a random “Squirrel” effect in the model.



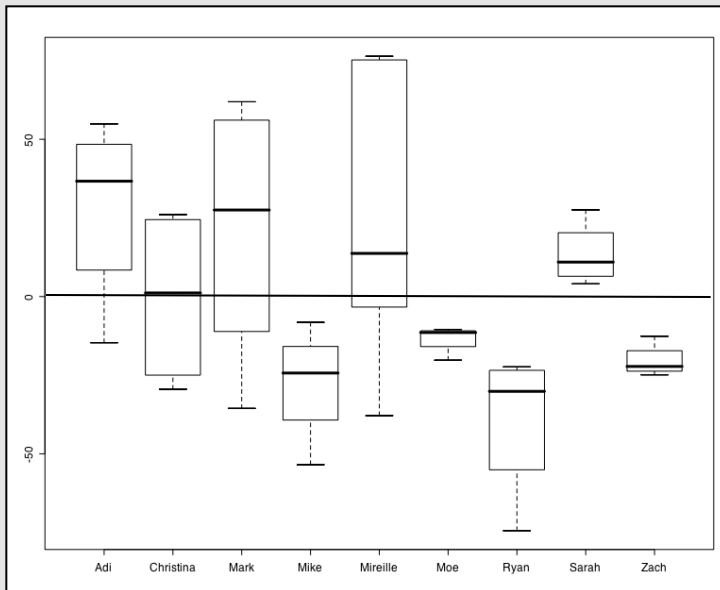
# What do the random effects do?



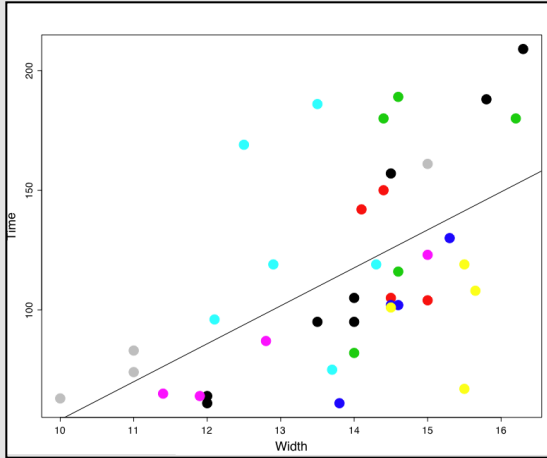
$$Time_{ij} = \beta_0 + \beta_1 Width + \varepsilon_{ij}$$

$$Time_{ij} = \beta_0 + \beta_1 Width + b_j + \varepsilon_{ij}$$

$b_j$  is the random effect



# What do the random effects do?

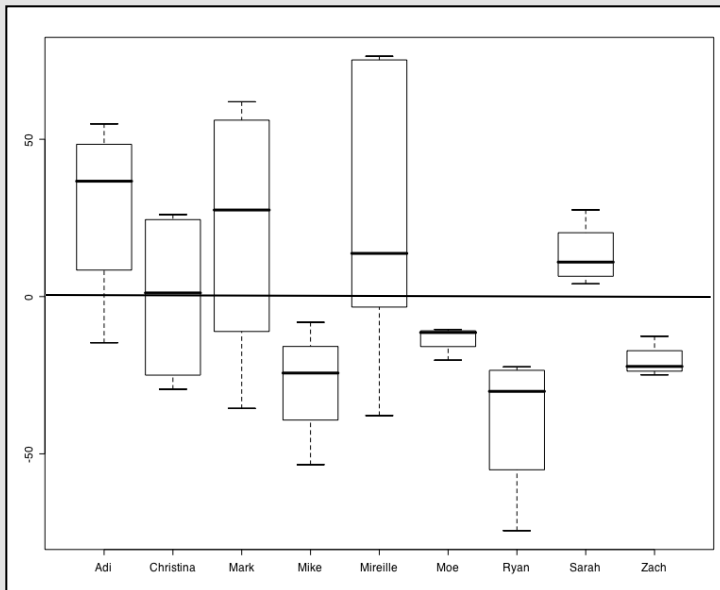


$$Time_{ij} = \beta_0 + \beta_1 Width + \varepsilon_{ij}$$

$$Time_{ij} = \beta_0 + \beta_1 Width + b_j + \varepsilon_{ij}$$

$b_j$  is the random effect

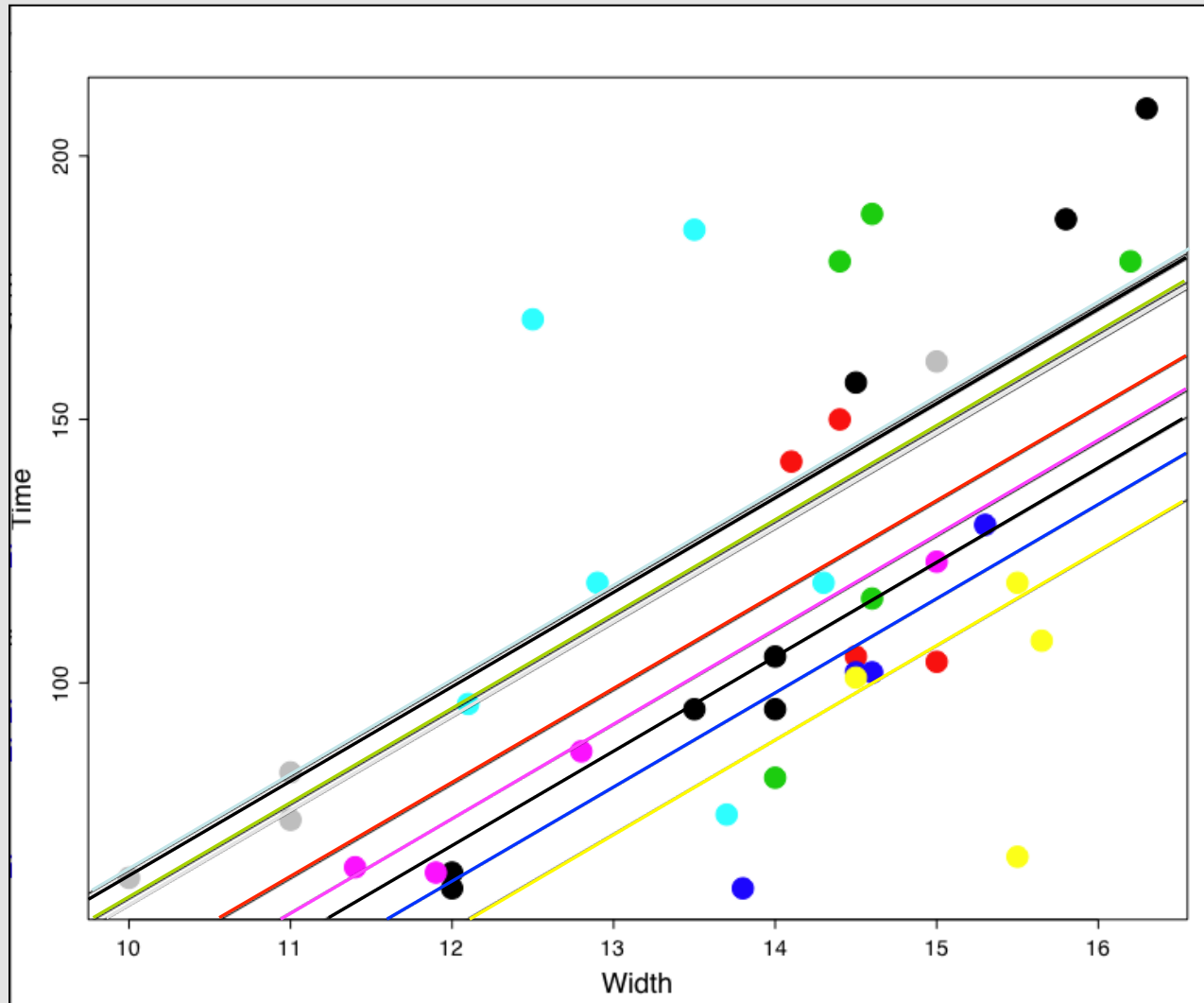
$$b_j \sim N(0, \sigma_b^2)$$



$b_{\text{Adi}} = 18.4$   
 $b_{\text{Christina}} = -0.1$   
 $b_{\text{Mark}} = 13.87$   
 $b_{\text{Mike}} = -18.5$

...

This is the same as fitting separate regressions for each squirrel using a common slope but different intercepts



# We now have within group errors and random effects

- Remember that our model is:

$$Time_{ij} = \beta_0 + \beta_1 Width + b_j + \varepsilon_{ij}$$

Random effects

$$b_j \sim N(0, \sigma_b^2)$$

Within group errors

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

# Yet more complex models

- What if we wanted to model group level variation for the effect of “width” (I.e. group level variation for the slope)?

```
>model3.MCMC <-MCMCglmm(Time~ Width+ random=~Squirrel, data=cone.handling)
```

```
>model3<-lmer(Time~ Width+(1|Squirrel), data=cone.handling)
```

How would we modify this?



# Yet more complex models

- What if we wanted to model group level variation for the effect of “width” (I.e. group level variation for the slope)?

```
>model4.MCMC <-MCMCglmm(Time~ Width +  
                        random=~idh(1+ Width):Squirrel, data=cone.handling)
```

```
>model4<-lmer(Time~ Width+(Width|Squirrel), data=cone.handling)
```



This tells us that width needs to incorporate group (random) level variation, like a separate slopes model. There is an implicit intercept for lmer (Width|Squirrel) = (1+Width|Squirrel).

# Variance structures

lmer	MCMCglmm/asreml	No. Parameters	Variance	Correlation
(1 dam)	dam	1	$\begin{bmatrix} V & V & V \\ V & V & V \\ V & V & V \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
(sex-1 dam)	us(sex):dam	6	$\begin{bmatrix} V_{1,1} & C_{1,2} & C_{1,3} \\ C_{1,2} & V_{2,2} & C_{2,3} \\ C_{1,3} & C_{2,3} & V_{3,3} \end{bmatrix}$	$\begin{bmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{1,2} & 1 & r_{2,3} \\ r_{1,3} & r_{2,3} & 1 \end{bmatrix}$
(1 sex:dam)	sex:dam	1	$\begin{bmatrix} V & 0 & 0 \\ 0 & V & 0 \\ 0 & 0 & V \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
(1 dam)+(1 sex:dam)	dam+sex:dam	2	$\begin{bmatrix} V_1 + V_2 & V_1 & V_1 \\ V_1 & V_1 + V_2 & V_1 \\ V_1 & V_1 & V_1 + V_2 \end{bmatrix}$	$\begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix}$
-	idh(sex):dam	3	$\begin{bmatrix} V_{1,1} & 0 & 0 \\ 0 & V_{2,2} & 0 \\ 0 & 0 & V_{3,3} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
-	corh(sex):dam	4	$\begin{bmatrix} V_{1,1} & rV_{1,1}V_{2,2} & rV_{1,1}V_{2,2} \\ rV_{1,1}V_{2,2} & V_{2,2} & rV_{2,2}V_{2,3} \\ rV_{1,1}V_{3,3} & rV_{2,2}V_{3,3} & V_{3,3} \end{bmatrix}$	$\begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix}$
-	cor(sex):dam	3	$\begin{bmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{1,2} & 1 & r_{2,3} \\ r_{1,3} & r_{2,3} & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{1,2} & 1 & r_{2,3} \\ r_{1,3} & r_{2,3} & 1 \end{bmatrix}$

From Hadfield 2010, pg 68, course notes for MCMCglmm



# Best linear unbiased predictor's (BLUP's)

- This is where the shrinkage comes in.

For a model where all we were estimating is the group level “means”, the use of a mixed model framework would result in.

$$\hat{\alpha}_j^{BLUP} \approx \frac{\frac{n_j}{\sigma_e^2} \bar{y}_j + \frac{1}{\sigma_\alpha^2} \bar{y}_{all}}{\frac{n_j}{\sigma_e^2} + \frac{1}{\sigma_\alpha^2}}$$

# Shrinkage in mixed models

$$\hat{\alpha}_j^{BLUP} \approx \frac{\frac{n_j}{\sigma_e^2} \bar{y}_j + \frac{1}{\sigma_\alpha^2} \bar{y}_{all}}{\frac{n_j}{\sigma_e^2} + \frac{1}{\sigma_\alpha^2}}$$

$\hat{\alpha}_j^{BLUP}$  The new predicted value for group j (instead of the mean)

$\bar{y}_j$  The (LS) mean for group j (unpooled estimate)

$\bar{y}_{all}$  The mean over all observations (pooled estimate)

$\sigma_e^2$  Within group variance

$\sigma_\alpha^2$  Between group variance

$n_j$  Sample size in group j

# Shrinkage in mixed models

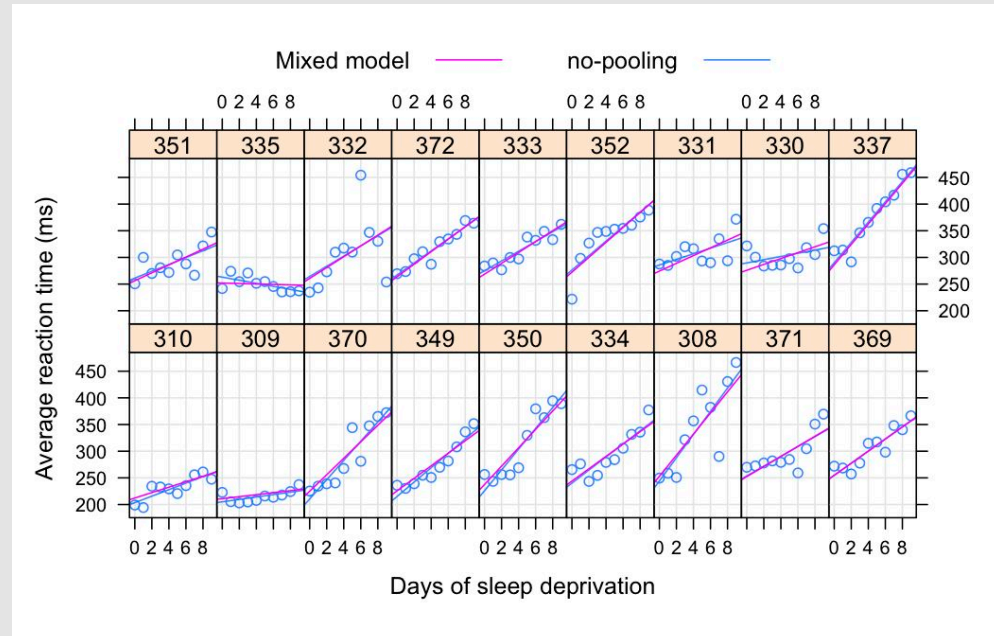
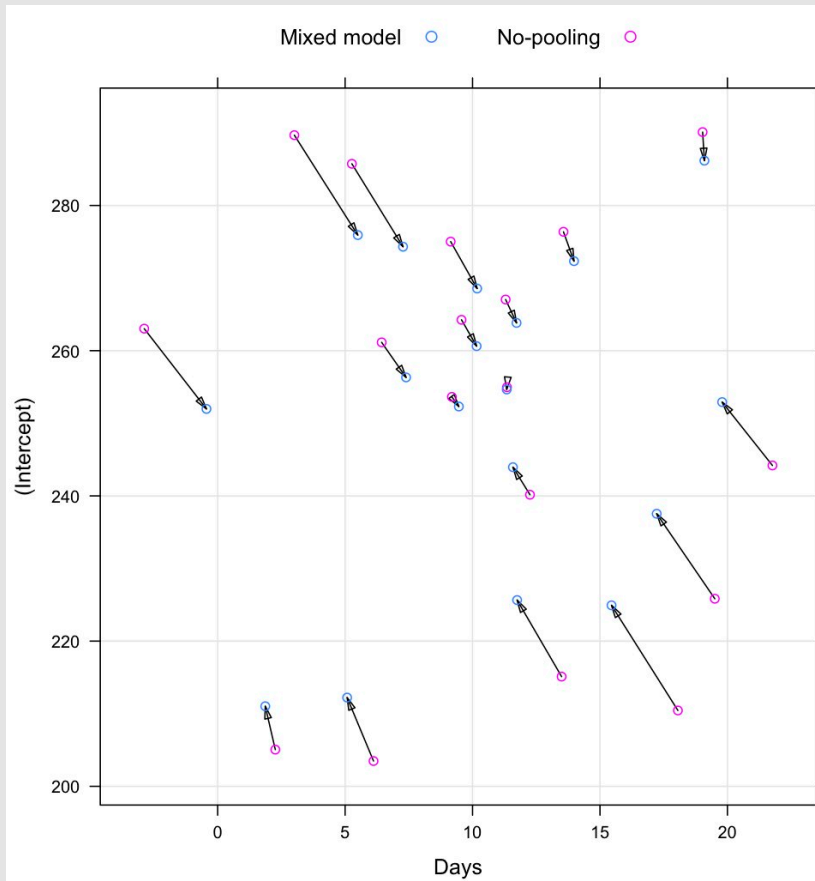
$$\hat{\alpha}_j^{BLUP} \approx \frac{\frac{n_j}{\sigma_e^2} \bar{y}_j + \frac{1}{\sigma_\alpha^2} \bar{y}_{all}}{\frac{n_j}{\sigma_e^2} + \frac{1}{\sigma_\alpha^2}}$$

What this means is that group averages  $E(y_j)$  with small sample sizes contain less information, and the weighting pulls the BLUP's closer to the overall average.

When sample size is larger for a group the estimate will be closer to the expected value  $E(y_j)$ .

You can consider this a kind of “partial pooling”

# The effects of shrinkage



Example for longitudinal (repeated measure) data

# *ML* vs. *REML* in Mixed-Effects Models

- Maximum likelihood estimates of variance components maximize the likelihood of the data
  - Do not take into account the fixed effect structure of a model
    - Assume that the fixed effects are known without error
  - Are, therefore, biased downward (similar to the bias in sample variance)
- Restricted ML address this bias by considering only the residuals of the fixed effects in the model
  - Maximize only the part of the likelihood that does not depend on the fixed effects
- This is helpful for coming up with good estimates of variance components, but it causes problems when we want to use **LRT** to assess a change in the fixed effects between two models.s



# Assumptions of Mixed-Effects Models

1. Within group errors ( $e_{ij}$ ) are independent, do not differ among groups, have a mean of zero and are normally distributed
2. Random effects ( $b_j$ ) are normally distributed, mean of zero and are independent of one another

# Mixed Effects Models Summary

- Fixed Effects estimate means for levels within a factor, random effects estimate variation among levels
- The inclusion of random effects allows us to account for structure in the data
  - Independence of errors
  - Corrects degrees of freedom
- Flexible framework for the analysis of blocked, nested and repeated measures designs
- Handle missing data
- In more advanced scenarios you can model both the response and the variance in random effects

# Benefits of Mixed-Effects Models

- Allow you to correct for lack of independence of errors
  - Account for underlying group structure in the data
- Handle balanced and unbalanced designs with missing data.
- Utilizes all of the data when making “group” level predictors (i.e. the “best linear unbiased predictions, BLUPs” for random coefficients).

# Software for (generalized) linear Mixed models

**Table 1. Capabilities of different software packages for GLMM analysis: estimation methods, scope of statistical models that can be fitted and available inference methods**

		Penalized quasilikelihood	Laplace	Gauss- Hermite quadrature	Crossed random effects	Wald $\chi^2$ or Wald F tests	Degrees of freedom	MCMC sampling	Continuous spatial/ temporal correlation	Overdispersion
SAS	PROC GLIMMIX	✓	✓ <sup>a</sup>	✓ <sup>a</sup>	✓	✓	BW, S, KR		✓	QL
	PROC NL MIXED			✓		✓	BW, S, KR			Dist
R	glmmPQL	✓				✓	BW		✓	QL
	glmmML		✓	✓						
	glmer		✓	(✓)	✓			(✓)		QL
	glmmADMB		✓							Dist
	GLMM	✓			✓?	✓			✓	QL
GenStat/ ASREML			✓	✓	✓			✓		Dist
AD Model Builder		✓	✓		✓					✓
HLM				✓						
GLLAAMM (Stata)								✓		Dist
WinBUGS					✓			✓		

Abbreviations: BW, between-within; dist, specified distribution (e.g. negative binomial); KR, Kenward-Roger; QL, quasilikelihood; S, Satterthwaite.

<sup>a</sup>Version 9.2 only.

Bolker et al. 2009

Also add RLRsim, Imm, MCMCglmm, sabreR in R and JAGS (Gibbs Sampler), wombat

For general linear mixed models SAS also has PROC MIXED

# Estimation

**Table I. Techniques for GLMM parameter estimation, their advantages and disadvantages and the software packages that implement them**

Technique	Advantages	Disadvantages	Software
Penalized quasilielihood	Flexible, widely implemented	Likelihood inference inappropriate; biased for large variance or small means	PROC GLIMMIX (SAS), GLMM (Genstat), glmmPQL (R), glmer (R)
Laplace approximation	More accurate than PQL	Slower and less flexible than PQL	PROC GLIMMIX [56], glmer (R), glmm.admb (R), AD Model Builder, HLM
Gauss-Hermite quadrature	More accurate than Laplace	Slower than Laplace; limited to 2–3 random effects	PROC GLIMMIX [56], PROC NLMIXED (SAS), glmer (R), glmmML (R)
Markov chain Monte Carlo	Highly flexible, arbitrary number of random effects; accurate	Very slow, technically challenging, Bayesian framework	WinBUGS, JAGS, MCMCpack, (R), AD Model Builder

MCMCglmm ( R )

# Making inferences for mixed models

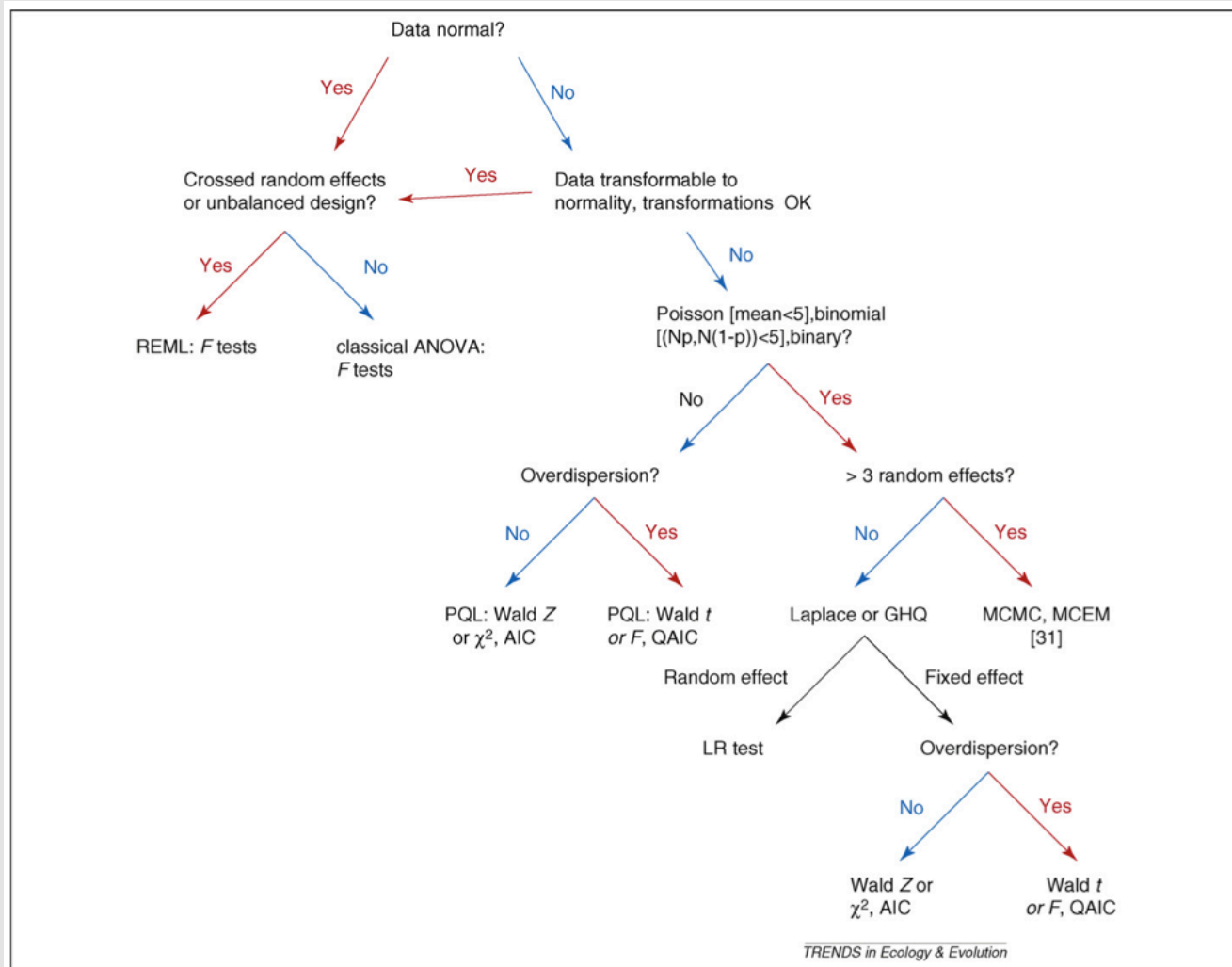
## Options

Parametric Bootstrap (Monte Carlo Simulations) – available in `lmer()`, maybe in others?

Non-Parametric Bootstrapping (must carefully consider sampling design, within and between random effects. In particular you want to keep the numbers of levels of your random effect the same as with the observed data. However you may also want to sample within groups as well... Depends on the data).

MCMC (options available in `MCMCglmm`, `glmmADMB` and to some extent `lmer` via `mcmc` and `arm`).

# Decision tree



**Figure 1.** Decision tree for GLMM fitting and inference. Conditions on the Poisson and binomial distributions along the right branch refer to penalized quasiliikelihood (PQL) rules of thumb [30]: to use PQL, Poisson distributions should have mean  $> 5$  and binomial distributions should have the minimum of the number of successes and failures  $> 5$ . MCEM = Monte Carlo expectation-maximization [40].