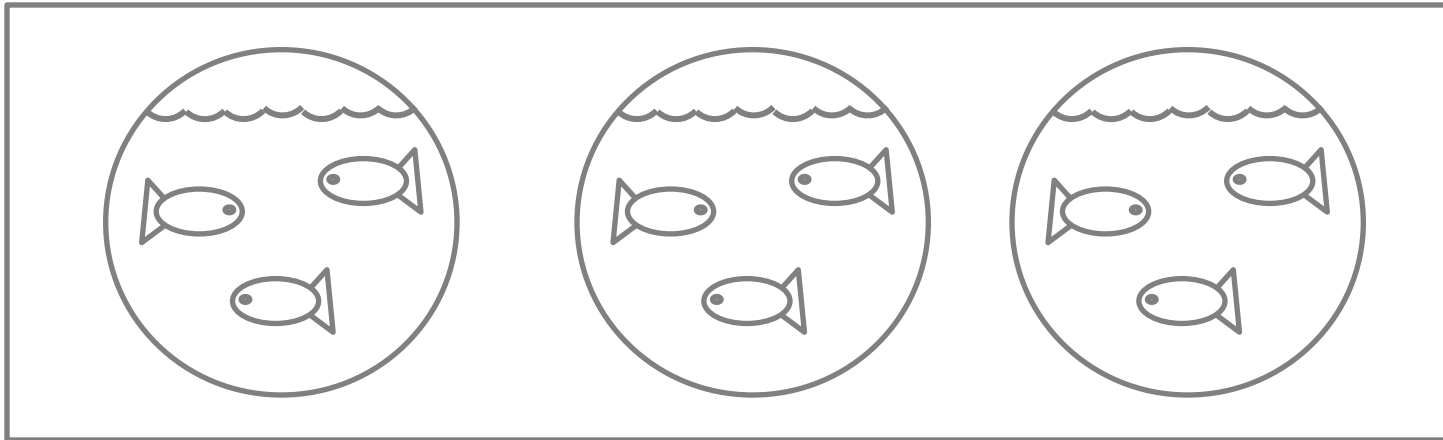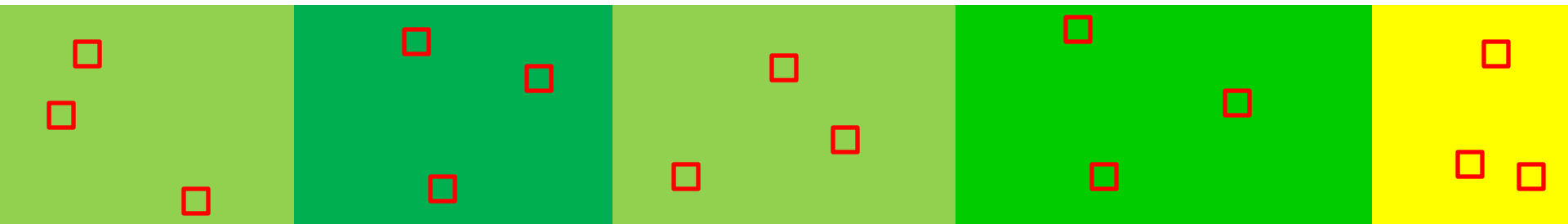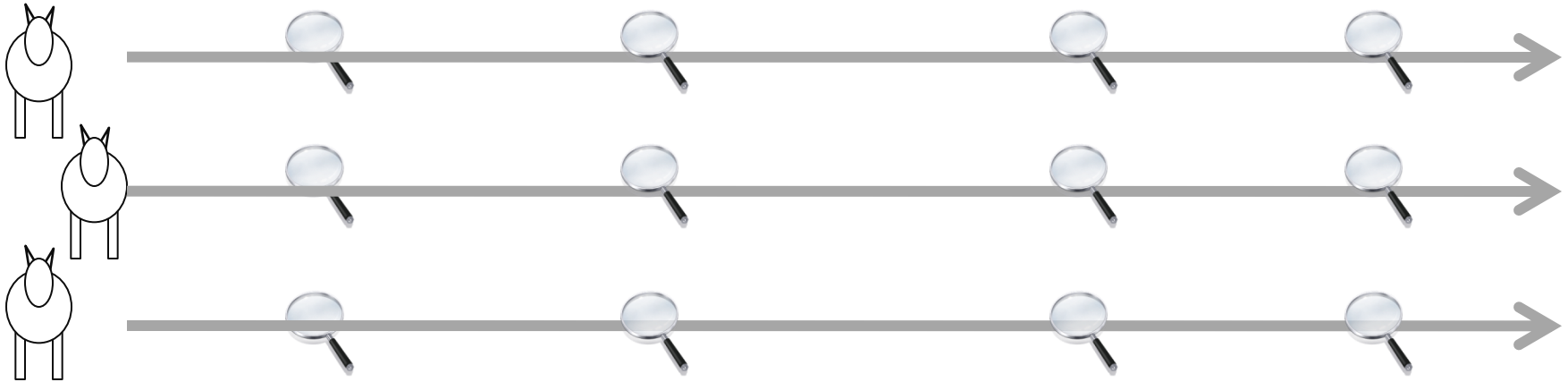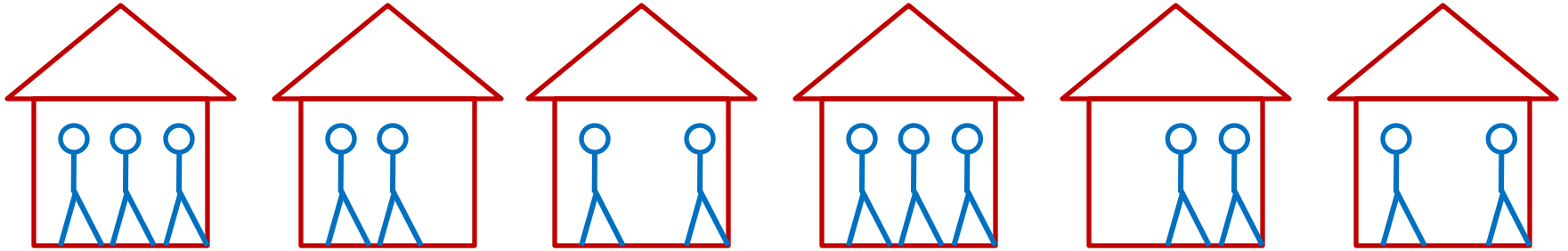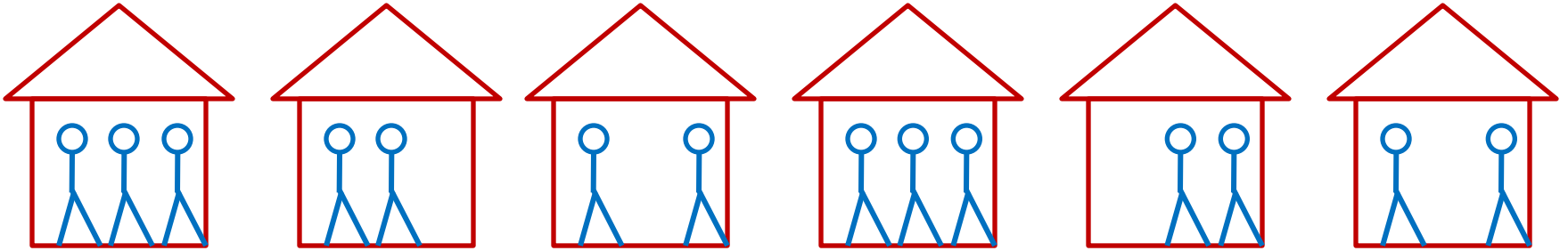# L7: Multilevel (Mixed) Models

Grouped data

Grouped data

Data = Mean + Treatment effect + Error

Data = Mean (+ Treatment effect) + Error

(+ Group treatment effect) + Group error

Up to now...

- We have assumed that factors in our GLMs have had their levels chosen, or "fixed"
  - we assumed that all error lies in our dependent variable
  - we deliberately choose Drugs A, B and C to compare – we don't pick three drugs at random

- So our results apply **only** to the exact levels we chose - Drugs A vs. B vs. C – no inference about "Drugs" in general.
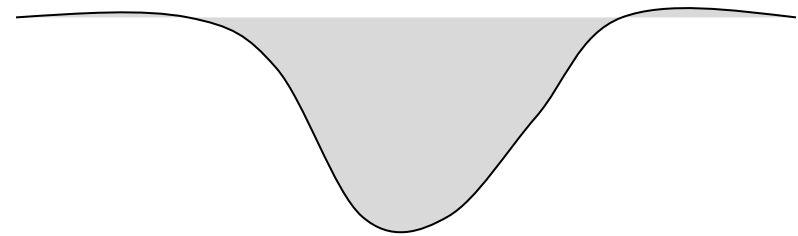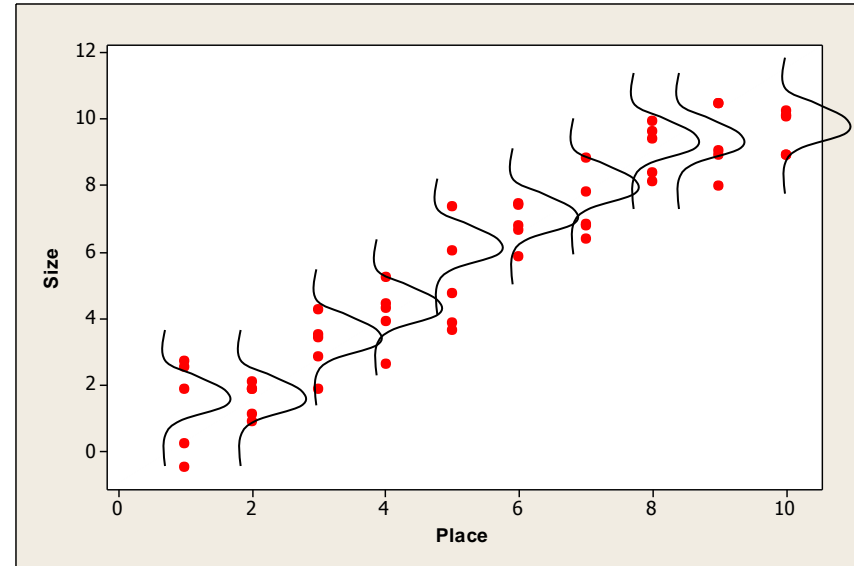
But suppose we want to know about effects of a random sample of treatments…

Fixed Effects

- the model estimates the mean effect of each factor level on the response

- for treatments you specified in your hypothesis

- coefficients will be means for each group

  "no pooling"

Random Effects

- the model can also estimate **variance** of responses among the factor levels

- for treatments you didn't specify or control
  – Species, Genotype, Country

- for groupings in your data
  – Location, Individual, Brood
  – Block, Field, Plot, Sub-plot

- estimates for each group ≠ means; drawn towards grand mean according to within-group variability

"partial pooling"

More formally…

- A model with a single fixed factor estimates:
  - $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ e.g. height = mean + sex + error
  - μ = mean, α_i = treatment effect, ε_{ij} = residual ($N(0, \sigma^2)$)

- A model with a single random factor predicts the variance from which each factor level is drawn:
  - $y_{ij} = \mu + A_i + \varepsilon_{ij}$ e.g. growth rate = mean + individual + error
  - *A* is the random effect and is assumed $N(0, \sigma^2)$; each individual i is drawn at random from this distribution
  - There is *sampling error* associated with the <u>choice of individual</u> (A) and with <u>measuring growth rate,</u> and the errors are independent

- A mixed model combines fixed and random effects, e.g:
  - $y_{ij} = \mu + \alpha_i + A_i + \varepsilon_{ij}$ e.g. growth rate = mean + sex + individual + error

Two uses of random factors

1. Where the variance of a factor's effects is more interesting than knowing the actual means.

   – e.g. Nussey *et al* (2005) J. Anim. Ecol. estimating among-individual variation in plasticity (response of calving date to climate)

2. Where you want to know mean effects of a factor's levels but want them to fit a distribution (central tendency)

   – e.g. known genotypes, species, other biological groups.

GxE
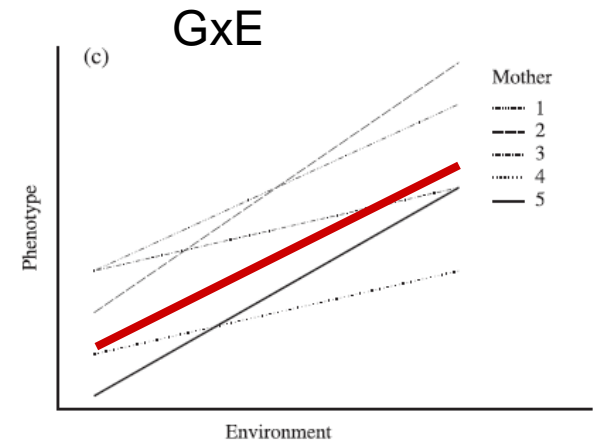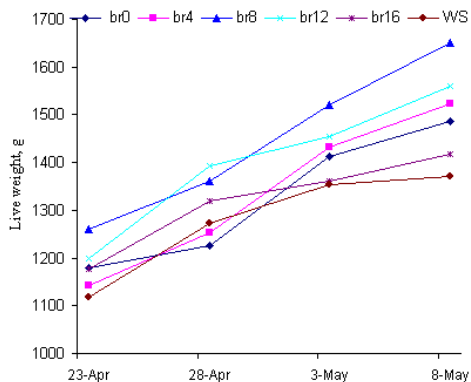
Examples of groups treated as adding random variation:

- Groups of individuals (additional error associated with group)
  - Broods/families
  - Populations
- Repeated measurements on individuals (additional error associated with choice of individual)
  - E.g. growth rate
- Grouping in time or space (additional error associated with choice of time or place)
  - Day of measurement/climate
  - Block (field/plot/greenhouse/incubator/waterbath/experimenter)
  - Site, Region, County, Country, etc
- Populations sampled at random, where you want to make inferences about the variance among populations as well as an overall mean

"Mixed Effects": Fixed + Random

Douglas Bates

If levels of a factor represent a **random sample** from the set of all possible levels, we model them using **random effects**. Conclusions apply to the population of levels from which we chose.

Sampling at random means there will be **sampling error** associated with a predictor variable, which violates the GLM assumption that X is measured without error.

> *Mixed-effects models* or, more simply, *mixed models* are statistical models that incorporate both fixed-effects parameters and random effects. Because of the way that we will define random effects, a model with random effects always includes at least one fixed-effects parameter. Thus, any model with random effects is a mixed model.
>
> There are two things to notice about this distinction between fixed-effects parameters and random effects. First, the names are misleading because the distinction between fixed and random is more a property of the levels of the categorical covariate than a property of the effects associated with them. Secondly, we distinguish between "fixed-effects parameters", which are indeed parameters in the statistical model, and "random effects", which, strictly speaking, are not parameters. As we will see shortly, random effects are unobserved random variables.

What is a "Random Effect"?

Andrew Gelman (2005: *Ann. Stat.* 33:1–53)

5 definitions of "random effects" (mutually inconsistent):

1. Fixed effects are constant across individuals, and random effects vary. For example, in a growth study, a model with random intercepts a_i and fixed slope b corresponds to parallel lines for different individuals i ($y_{it} = a_i + b_t$). Kreft and De Leeuw (1998).

2. Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population. Searle et al (1992).

3. "When a sample 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)

4. "If an effect is assumed to be a realized value of a random variable, it is called a random effect." (LaMotte, 1983)

5. Fixed effects are estimated using least squares (or, more generally, maximum likelihood) and random effects are estimated with shrinkage ("linear unbiased prediction" in the terminology of Robinson, 1991). This definition is standard in the multilevel modeling literature (see e.g. Snijders and Bosker, 1999, Section 4.2)

→ **Just refer to levels of grouping in multi-level / hierarchical models**

Hierarchical grouping ("nesting", stratified sampling)

You may have replicated samples within units of space or time: grouped replicates likely to be more similar

- Chicks in a nest
- Individuals within families
- Sizes at ages (repeated measures)
- Samples within plots within fields within farms within landscapes within regions…
- Locations within tables within glasshouses
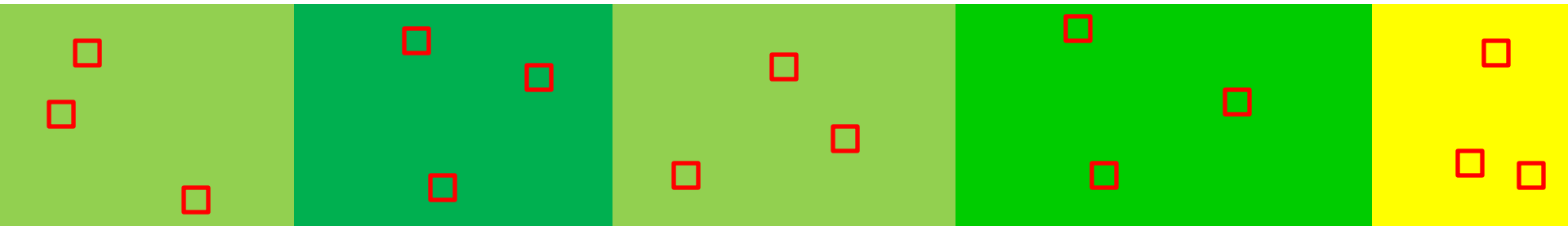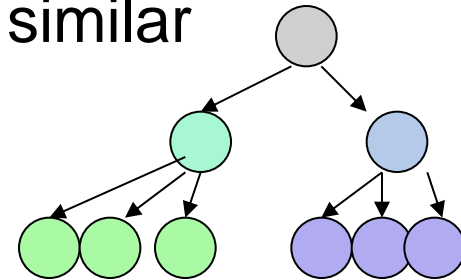
*"**Closer** replicates being **more** similar" will crop up for time series and spatial models in the concept of **autocorrelation***

An example of multiple error terms

- Observations = insect leaf damage on trees

Variation among woods

Variation among trees

Variation among branches

Variation among leaves = ε

Variation due to treatment

Unexplained variation among observations

among observations
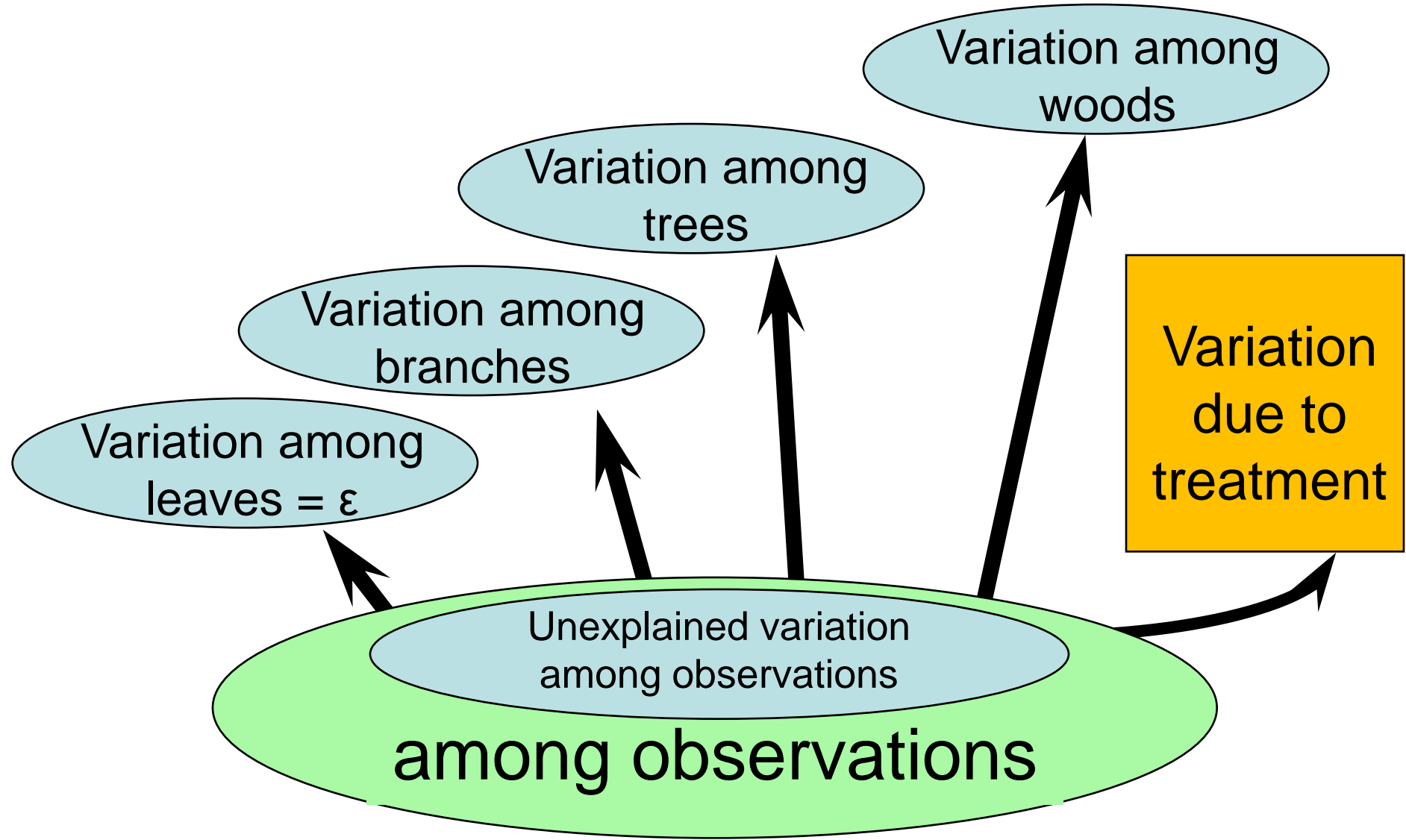
Example 1: Great tit nest boxes

Experiment to look at influence of heating or cooling on weights of chicks

- 12 nest boxes randomly assigned to 3 treatments (control, warm, cool)

- Fixed effects are <u>treatment</u> (of box) and <u>sex</u> (of individual)

- Random effect (grouping factor) is <u>nest box</u>

- Gives two "strata" for errors:
  - Nest box (group) and chicks within nest (individual)

Ignoring nest box "complete pooling"

`xyplot(WEIGHT ~ SEX|TREAT, auto.key= TRUE)`
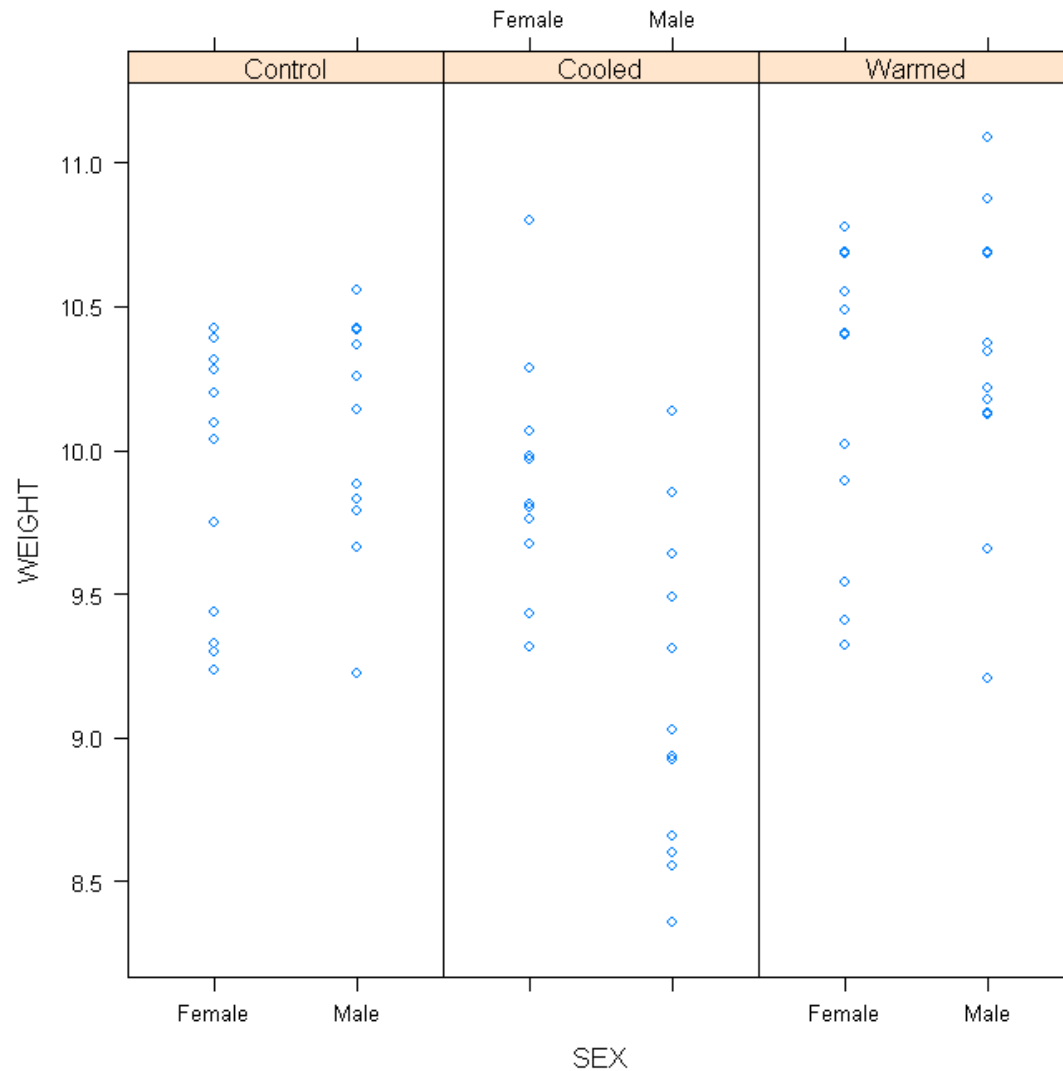
Ignoring nest box   "complete pooling"

```
> regstyle <- lm(WEIGHT~SEX*TREAT)
> summary(regstyle)

Call:
lm(formula = WEIGHT ~ SEX * TREAT)

Residuals:
     Min        1Q     Median        3Q       Max
-1.09080  -0.28797   0.06284   0.36663   1.01295

Coefficients:
                          Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)               9.899443    0.138885   71.278   < 2e-16 ***
SEXMale                   0.182225    0.196413    0.928  0.356911
TREATCooled               0.008162    0.196413    0.042  0.966978
TREATWarmed               0.283346    0.196413    1.443  0.153861
SEXMale:TREATCooled      -0.966490    0.277770   -3.479  0.000895 ***
SEXMale:TREATWarmed      -0.067498    0.277770   -0.243  0.808760
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

Residual standard error: 0.4811 on 66 degrees of freedom
Multiple R-Squared: 0.4067,       Adjusted R-squared: 0.3618
F-statistic:  9.05 on 5 and 66 DF,  p-value: 1.371e-06
```

Residual Variance = 0.48^2 = 0.23

Ignoring nest box "complete pooling"

Fixed Effects model

```
> anova(regstyle)
Analysis of Variance Table

Response: WEIGHT
            Df  Sum Sq Mean Sq F value     Pr(>F)
SEX          1  0.4749  0.4749  2.0519  0.156738
TREAT        2  6.5053  3.2527 14.0522 8.238e-06 ***
SEX:TREAT    2  3.4937  1.7468  7.5468  0.001118 **
Residuals   66 15.2770  0.2315
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

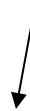Total df recorded as 71, with 66 in error:
too high, since chicks within each
nestbox are not independent

What should we do?

partial pooling

Sex of individual

Data = Mean + Treatment + Error
+ Group Treatment + Group error

Temperature of box

Using the correct error structure should give optimal estimates of each treatment's effects (as well as estimating variance among nest boxes).

```
Great.tits <- groupedData(WEIGHT~SEX|BOX, outer=TREAT)
plot(Great.tits)
```

NB: can specify grouping structure as a `groupedData()`
object before analysis to access complicated graphs simply

Factor that groups subjects
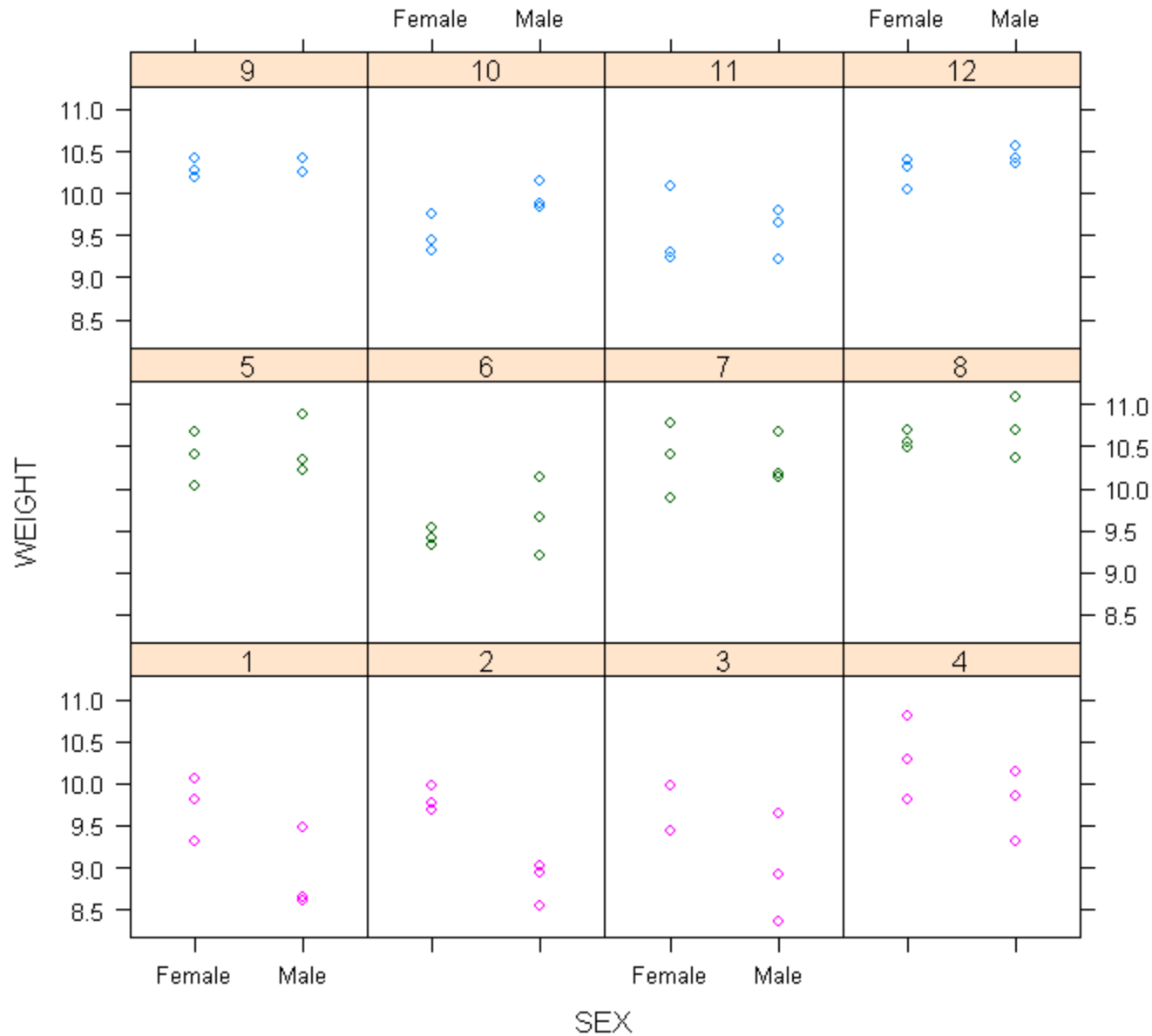
```
> model1 <- lme(WEIGHT~SEX*TREAT, random=~1|BOX, data=gg)
> summary(model1)
Linear mixed-effects model fit by REML
 Data: gg
       AIC       BIC      logLik
  89.34314 106.8604 -36.67157

Random effects:
 Formula: ~1 | BOX
         (Intercept)  Residual
StdDev:   0.3953251 0.3218725

Fixed effects: WEIGHT ~ SEX * TREAT
                       Value Std.Error DF  t-value p-value
(Intercept)          9.899443 0.2184124 57 45.32455  0.0000
SEXMale              0.182225 0.1314039 57  1.38676  0.1709
TREATCooled          0.008162 0.3088818  9  0.02643  0.9795
TREATWarmed          0.283346 0.3088818  9  0.91733  0.3829
SEXMale:TREATCooled -0.966490 0.1858332 57 -5.20085  0.0000
SEXMale:TREATWarmed -0.067498 0.1858332 57 -0.36322  0.7178

 Correlation:
                    (Intr) SEXMal TREATC TREATW SEXM:TREATC
SEXMale             -0.301
TREATCooled         -0.707  0.213
TREATWarmed         -0.707  0.213  0.500
SEXMale:TREATCooled  0.213 -0.707 -0.301 -0.150
SEXMale:TREATWarmed  0.213 -0.707 -0.150 -0.301  0.500

Standardized Within-Group Residuals:
       Min         Q1         Med          Q3         Max
-2.0153036 -0.7082614 -0.0123904  0.5418996  1.9656051

Number of Observations: 72
Number of Groups: 12
```

Fit by Restricted Maximum Likelihood (default)

Goodness of fit measures:

When comparing models, the model with the lower AIC and greater log-likelihood is a better fit

Estimate of the standard deviation of the distribution of random effects

Estimates of the standard deviation of the within-group error distribution:

residual variance = 0.32^2 = 0.103 (cf 0.23)

Coefficients (default "treatment contrasts")

Useful to look at to check the model was specified correctly

Importance of correct model specification

$F = MS_{effect}/0.103$

```
> anova(mixed effects model)
Analysis of Variance Table
          Df Sum Sq Mean Sq  Denom F value     Pr(>F)
SEX        1  0.475   0.475 66.000  4.5843    0.03596 *
TREAT      2  0.647   0.324 66.000  3.1237    0.05056 .
SEX:TREAT  2  3.494   1.747 66.000 16.8611 1.216e-06 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

```
> anova(fixed effects ignoring box)
Analysis of Variance Table

Response: WEIGHT
          Df   Sum Sq Mean Sq F value     Pr(>F)
SEX        1   0.4749  0.4749  2.0519   0.156738
TREAT      2   6.5053  3.2527 14.0522 8.238e-06 ***
SEX:TREAT  2   3.4937  1.7468  7.5468   0.001118 **
Residuals 66 15.2770  0.2315
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```
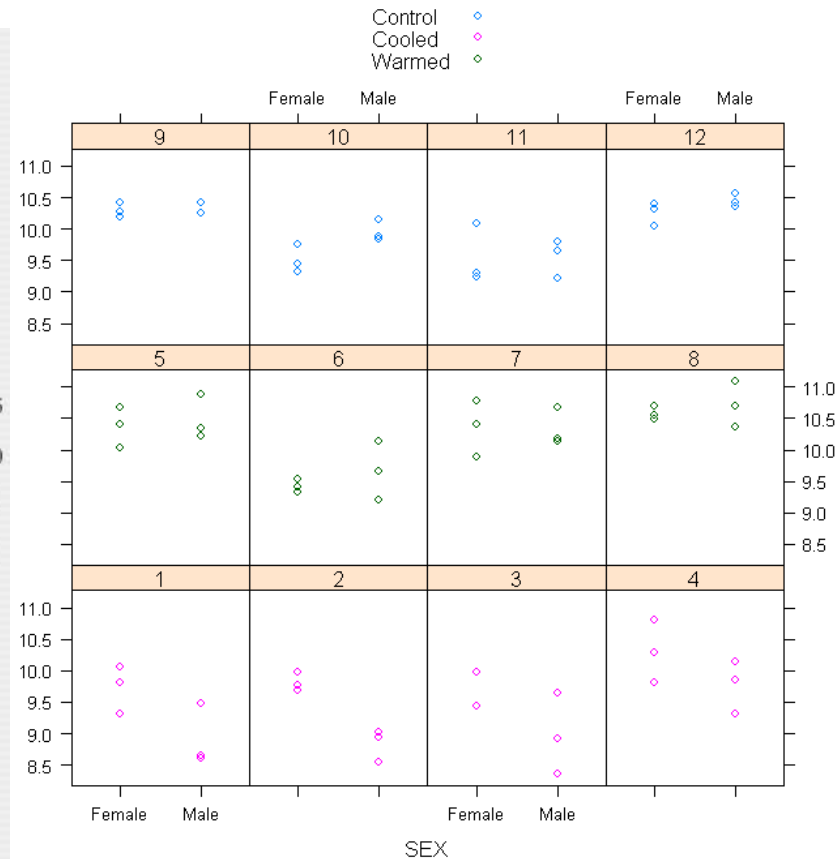
$F = MS_{effect}/0.2315$

So how well do the predictions fit?
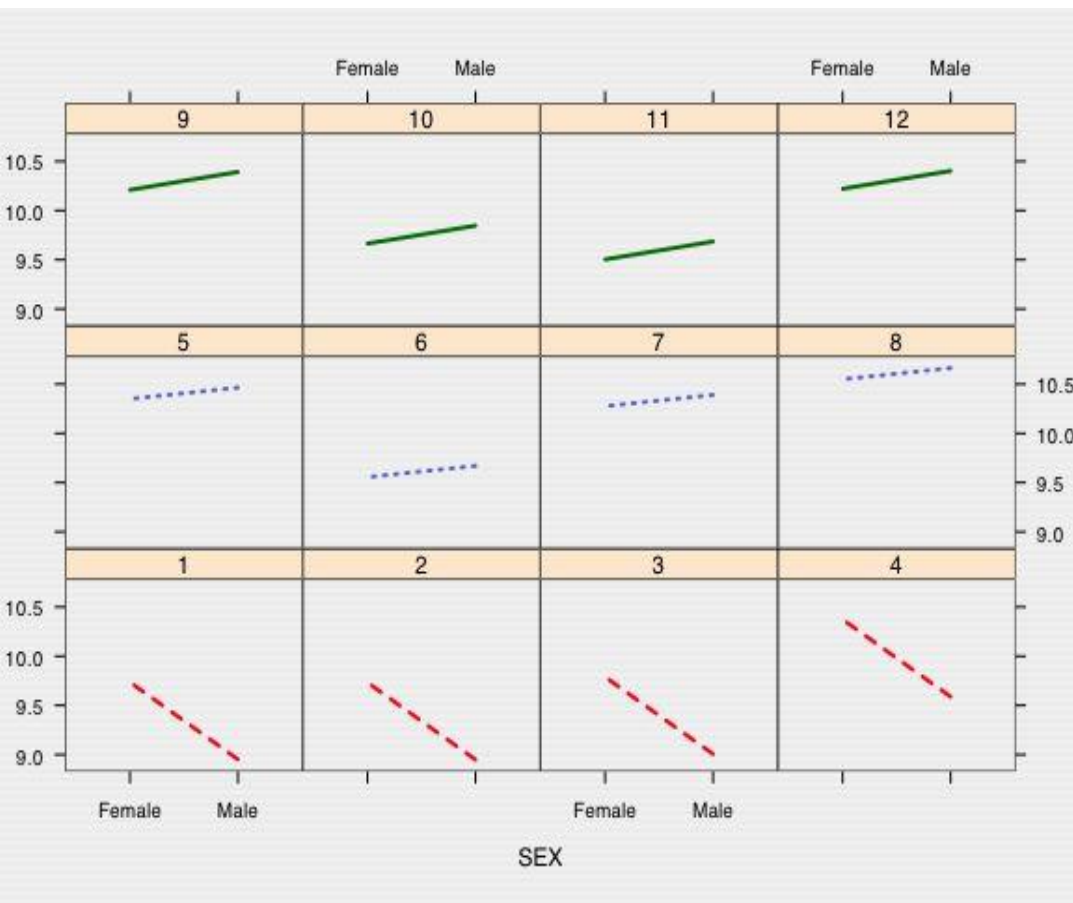
Checking assumptions
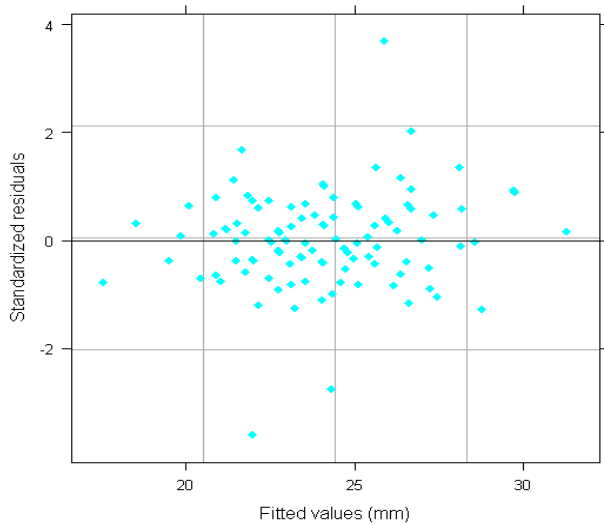
Homoscedasticity
`plot(model1)`

Normality of within-group error
`qqnorm(model1)`
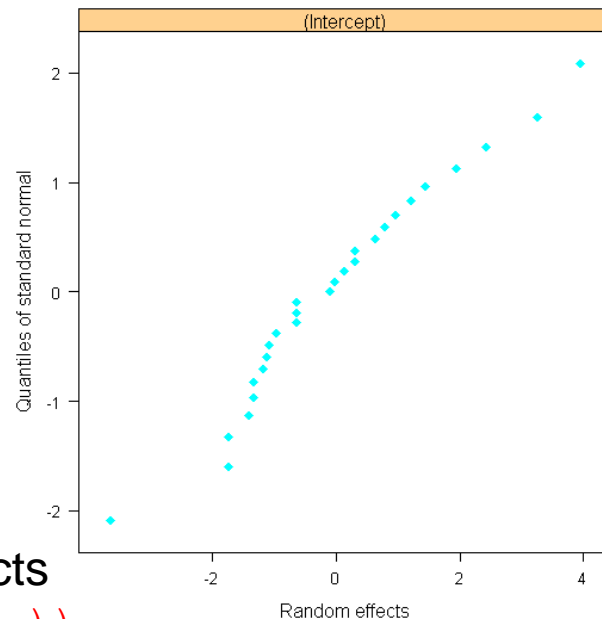
Normality of random effects
`qqnorm(model1,~ranef(.))`

Model outputs

Other useful functions:

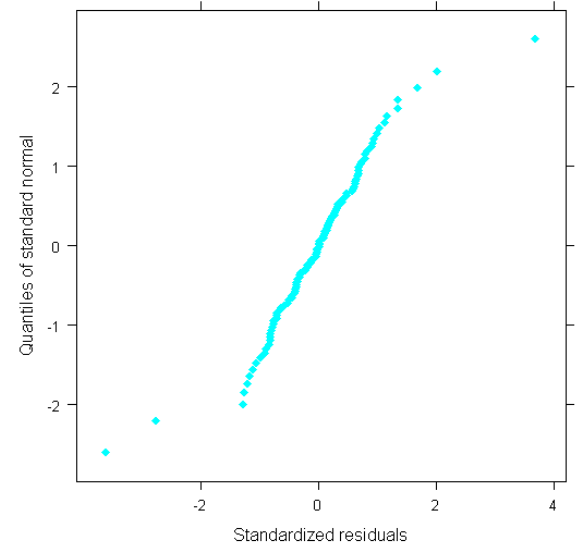- **VarCorr(model)** will give you the variance and standard deviations of the distributions of random effects and residuals (N.B. you already get the SDs in the `summary()` output)

- **ranef(model)** will give you estimated coefficients for whatever varies among factor levels. This is useful to check you have specified the model correctly.

- **intervals(model)** will provide 95% C.I. for coefficients of the fixed effects and the standard deviations of the random effects and residual distribution. This is useful to check the precision of the estimates; huge intervals can indicate misspecification of the model.

Packages in R

- library(nmle) for **lme()**
  - Older package
  - Allows autocorrelation structures…
- library(lme4) for **lmer()**
  - Faster, better?
  - Allows 2+ non-nested groupings
  - Allows generalised models

Fitting linear mixed

Using the lme4 package

by Douglas Bates

The lme function, which fits linear mixed models of the form described in Pinheiro and Bates (2000), has been available in the required R package nlme for several years. Recently my colleagues and I have been developing another R package, called lme4, and its lmer function which provides more flexible fitting of linear mixed models and also provides extensions to generalized linear mixed models.

The good news for lme users is that the lmer function fits a greater range of models, is more reliable, and is faster than the lme function. The bad news is that the model specification has been changed slightly. The purpose of this article is to introduce lmer, to describe how it can be used to fit linear mixed models and to highlight some of the differences between lmer and lme.

```
> lme1  <- lme( y~x, random= ~1|z )   is equivalent to
> lmer1 <- lmer( y~x + (1|z) )
```

Specifying models in **lmer()**

FIXED (treatment) effects **(+ ε)**

```
mod031 <- lmer( No ~ R + HC + M + CT + L +
 (1|Year:S) + (1|C) + (1|YF:FID) + (1|YF:FID:L), data=ins)
```

Survey within year

Cluster (neighbourhood)

Field within farm

Location within field within farm

RANDOM effects (grouping structure)

lmer() formula:
| introduces a grouping variable
/ indicates nesting
: indicates both an interaction and nesting;
a + b%in%a means a + a:b

Instead of ML, use REML for fitting

- **Maximum likelihood** is the mathematical technique of estimating the parameters that would make the data most likely *(e.g. if you toss a coin 1000 times and get 509 heads, a coin with a 0.509 chance of heads would be most likely to produce that).*

- **Restricted maximum likelihood** "was developed… because maximum likelihood estimates of variance components take no account of the degrees of freedom used in estimating treatment effects, [so] they have a downwards bias which increases with the number of fixed effects in the model. This in turn leads to under-estimates of standard errors for fixed effects."
  http://www.vsni.co.uk/products/genstat/htmlhelp/server/REML.htm

Hypothesis testing 1. Fixed effects

1. *F*-tests (recommended by Bates). Conservative

```
# Use anova() to test the significance of fixed terms by an F-test
> anova(model1)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	57	6797.970	<.0001
SEX	1	57	4.584	0.0366
TREAT	2	9	3.124	0.0933
SEX:TREAT	2	57	16.861	<.0001

2. Likelihood-Ratio tests (recommended by Crawley). Anti-conservative (note method=ML)

```
> model2 <- lme( WEIGHT~SEX*TREAT, random=~1|BOX, data=gg, method="ML" )
> model3 <- lme( WEIGHT~SEX+TREAT, random=~1|BOX, data=gg, method="ML" )
> anova(model2,model3)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
model2	1	8	78.25097	96.4643	-31.12548			
model3	2	6	102.13607	115.7961	-45.06803	1 vs 2	27.8851	<.0001

Hypothesis testing 2. Random terms

Compare models with different random effects by a likelihood ratio test

```
> model1 <- lme(WEIGHT~SEX*TREAT, random=~1|BOX, data=gg)
> model0 <- lm(WEIGHT~SEX*TREAT, data=gg)
> anova(model1,model0)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
model1	1	8	89.34314	106.8604	-36.67157			
model0	2	7	119.63090	134.9585	-52.81545	1 vs 2	32.28776	<.0001

The model with the random component provides a significantly better fit.

Treat box as random or fixed?

```
> model1 <- lme(WEIGHT~SEX*TREAT, random= ~1|BOX,
  data=cool, method="ML")
> model2 <- lm(WEIGHT~SEX*TREAT+BOX, data=cool)


> anova(model1,model2)
       Model df       AIC        BIC     logLik
model1     1  8  78.25097   96.4643 -31.12548
model2     2  8 102.66355  120.8769 -43.33177
```

Practical

- Day 4_mixedeffects