

## Outline for today

- Random vs fixed effects
- Two-factor example
- Why the calculations are different with random effects
- Unbalanced designs with random effects
- Examples of experiments with random effects
- Linear mixed-effects models
- Example: Estimating repeatability of a measurement
- Assumptions of linear mixed-effects models
- An example violating assumptions, with solutions

## What are fixed effects

Fixed effects are predetermined categories of a variable, of direct interest, and are repeatable. An experiment with the same treatment levels could be repeated.

For example:

- medical treatments in a clinical trial
- predetermined doses of a toxin
- diet or fertilization treatments
- age groups in a population
- habitat, season

Any conclusions reached in the study about differences among groups can be applied only to the groups included in the study. The results cannot be generalized to other treatments, habitats, etc. not included in the study.

## Example of factorial experiment with fixed effects

Field transplant experiment to investigate how herbivores affect the abundance of the red alga, *Mazzaella parksii* in the intertidal habitat of coastal Washington State using (Harley 2004). I analyzed a subset of treatments:

intertidal zone	herbivory treatment	
	present	absent
low	$n = 16$ plots	$n = 16$ plots
mid	$n = 16$ plots	$n = 16$ plots

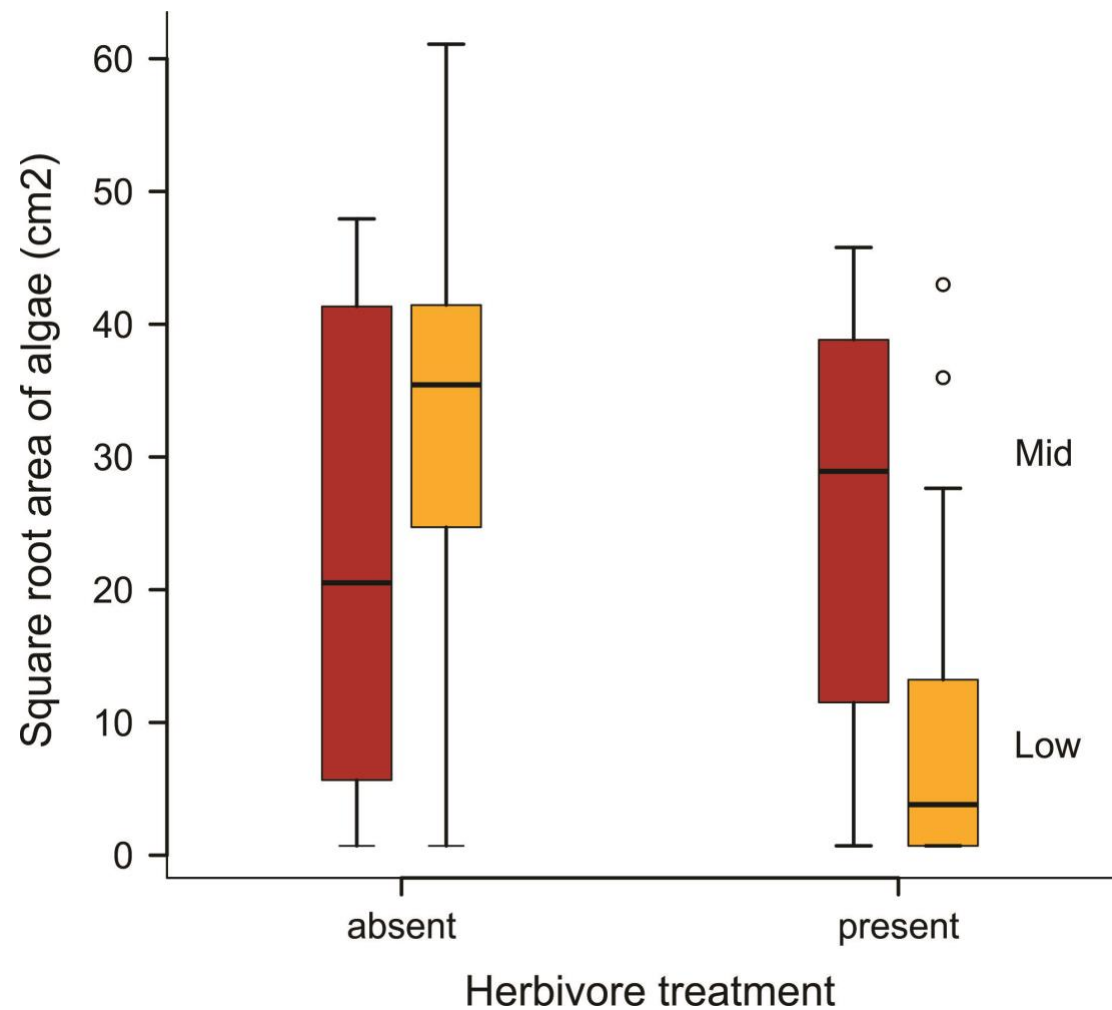


$N = 64$  plots in a completely randomized design

## Factorial experiment with fixed effects

Response variable: surface area of red alga, *Mazzaella parksii* in plots.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_2 X_1 X_2$$



## Linear model for factorial experiment with fixed effects

```
z <- lm(sqrtarea ~ herbivores * zone)
```

The denominator of the  $F$  statistic for the treatment effect is  $MS_{\text{residual}}$

Source	SS	df	MS	$F$	$P$
zone	89	1	89	$89/238 = 0.37$	0.54
herbivory	1512	1	1512	$1512/238 = 6.36$	0.01
zone*herbivory	2617	1	2617	$2617/238 = 11.0$	<0.01
residual (error)	14271	60	238		

The residuals represent the sole source of random variation in the analysis.

The residual mean square ( $MS_{\text{residual}}$ ) is the appropriate quantity to use as a reference when asking whether treatment mean squares are larger than random variation (chance), i.e., whether  $F \gg 1$ .

## What are random effects

Randomly sampled categories of a variable, representing *groups* or *clusters* of measurements or units. An experiment with the same treatment levels could not be repeated. For example:

- randomly sampled families made up of siblings
- randomly sampled subjects measured repeatedly
- randomly placed transects each of multiple quadrats in a sampling survey
- field plots each assigned multiple treatments
- environment chambers each containing many aquariums

Groups are assumed to be randomly sampled from a *population of groups*.

Therefore, conclusions reached about groups *can* be generalized to this population.

## What are random effects

In some cases, the random effects are of no interest themselves – a nuisance.

- field plots each assigned multiple treatments
- environment chambers containing aquariums

In other cases, measuring the variance associated with each level of random groupings is one purpose of the study.

- families of siblings (to estimate heritability, etc)
- subjects measured repeatedly (to estimate repeatability)
- transects with multiple quadrats (to estimate  $\alpha$ - and  $\beta$ -diversity)

Repeated measurements of groups are not independent.

Modeling random effects properly (using mixed effects models) avoids pseudoreplication.

## Factorial experiment when one of the factors is random



Futuyma and Philippi (1987)

Fall cankerworm, *Alsophila pometaria*

<http://cfs.nrcan.gc.ca/subsite/glfc-sugarbush/alsophila-pometaria-images>

Caterpillars of the fall cankerworm feed on the leaves of hardwood trees. Adult female moths are wingless. Many reproduce clonally, producing only daughters genetically identical to themselves.

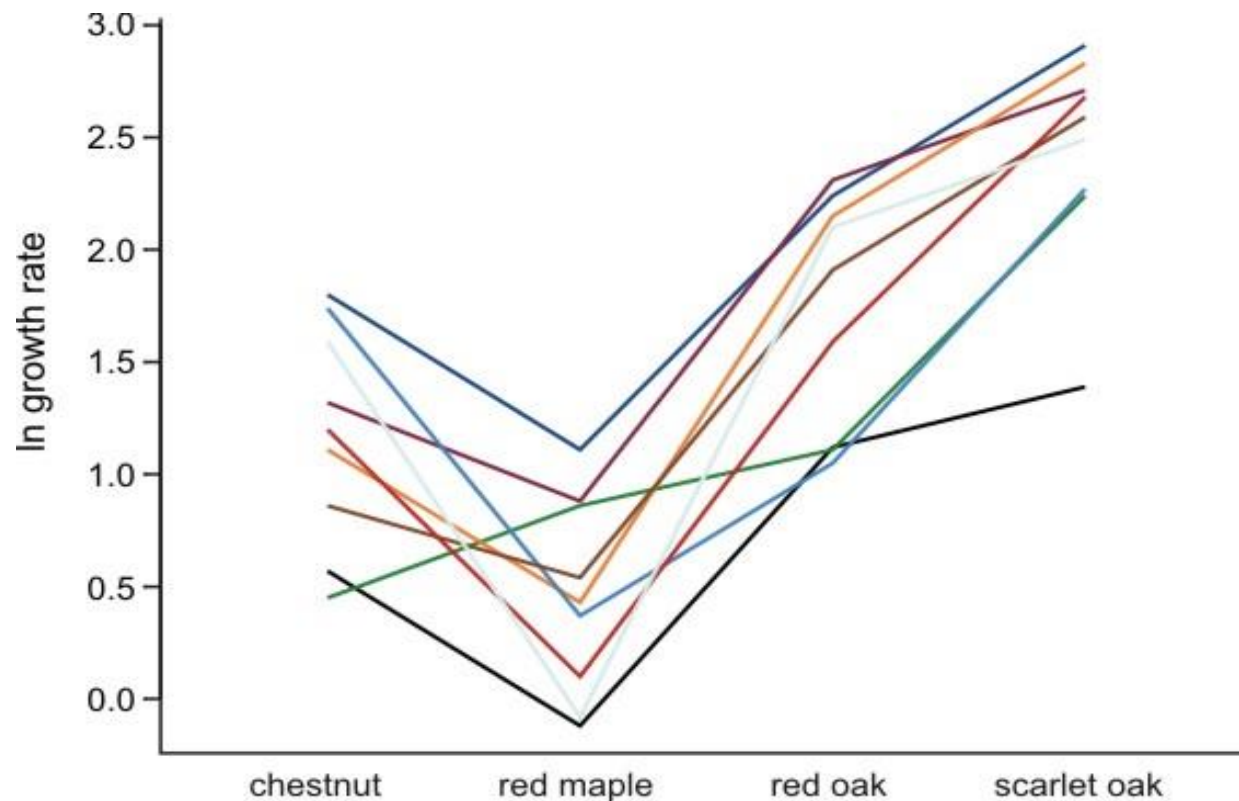
Research question: What is the effect of tree species on growth; and how much do clones vary in growth? (This is not a question about specific clones but about the population of clones.)



## Factorial experiment when one of the factors is random

Design: Sample 9 female moths from a population in NY. Raise larvae from 9 clones on leaves of 4 tree species. Measure individual growth after 15 days.

Two factors: **Tree species** (fixed), **Clone** (random)



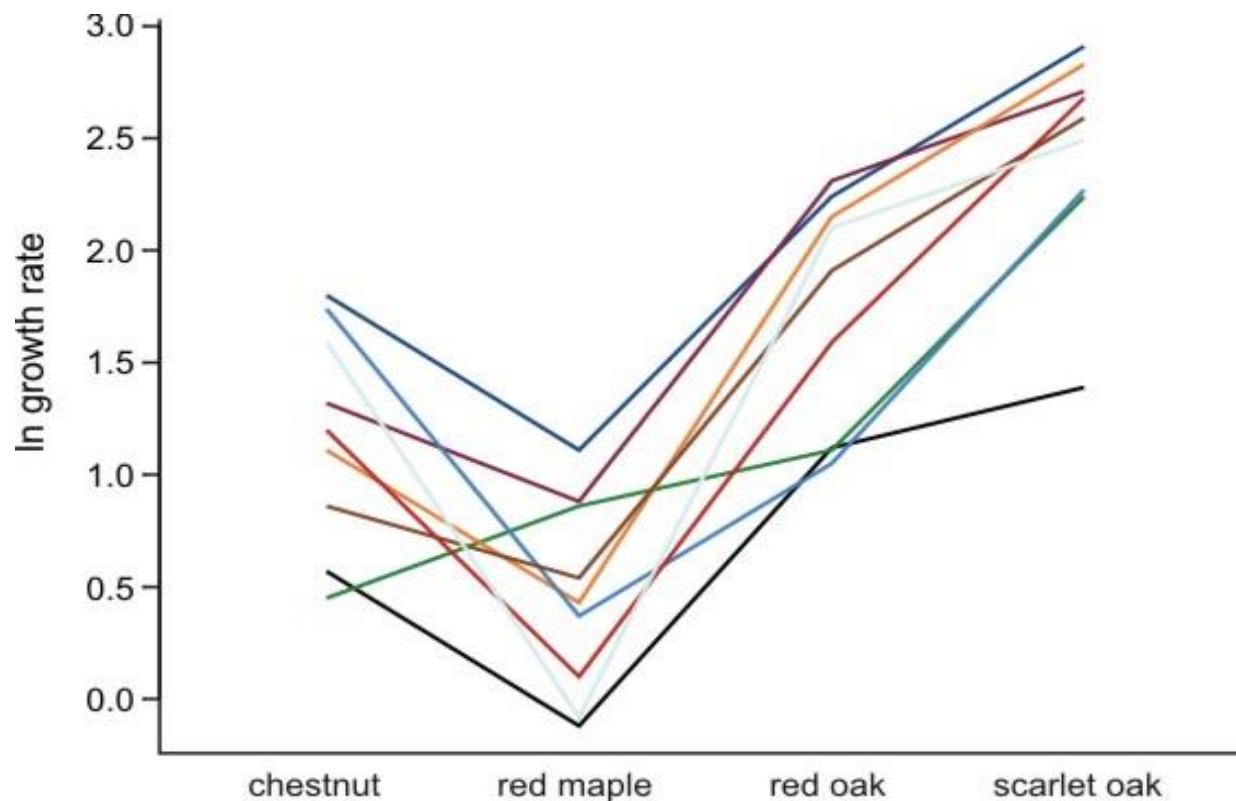
Interaction plot:

Mean growth of caterpillars from 9 families (clones) raised on four tree species.

$N = 326$  caterpillars total.

## Factorial experiment when one of the factors is random

Caterpillars from the same clone are not independent. There are 326 caterpillars but only 9 clones. In this case the residual mean square ( $MS_{\text{residual}}$ ) from a linear model fit is no longer the appropriate quantity to use as a reference when asking whether treatment mean squares are larger than expected by chance. To do so would commit pseudoreplication.



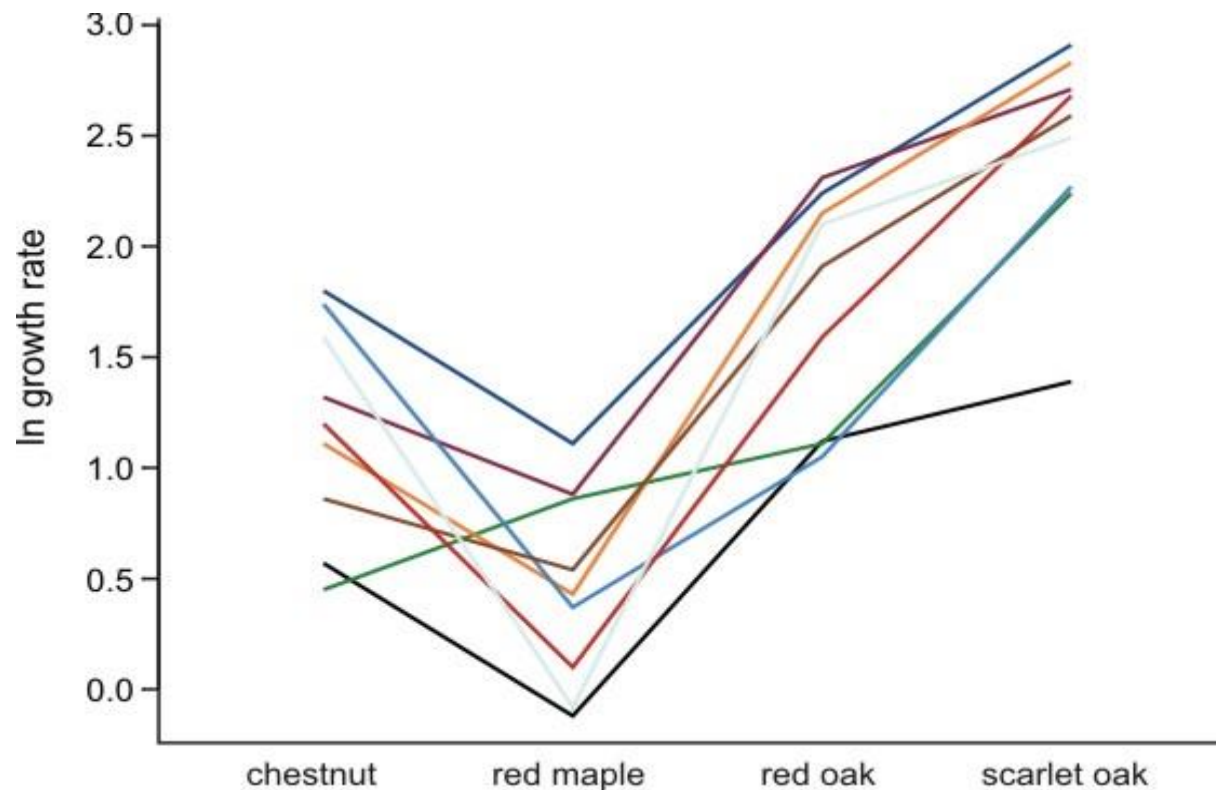
Instead, the variance among clones is the key to testing treatment effects.

(Always report *df* with *F* statistics in your papers to prove you analyzed such data correctly.)

## Factorial experiment when one of the factors is random

The presence of a random factor represents another source of random variation: the variation among clones.

The mean effect of tree species on caterpillar growth is modeled as the mean of clone means  $\pm$  variation among clones.



## More reasons why analysis is different with random effects

1. Unlike fixed groups, the means of the specific random groups (e.g., clones) are not of direct interest.
2. Interest is focused on the variance among the random groups. A goal of the Futuyma and Philippi experiment was to *estimate* these variances, to measure the amount of genetic variation in the population.
3. When a design including random effects is unbalanced, the standard formulas for  $F$  statistics (as calculated in the above ANOVA table of fixed effects) are not  $F$ -distributed. Standard ANOVA table calculations don't work with unequal sample sizes. Sorry!
4. With unbalanced designs, mixed models calculate approximate  $F$ -statistics and degrees of freedom for tests of fixed effects. Results are approximate.

## How to know when you have random effects in your study

You have random effects:

- whenever your sampling design is nested:  
quadrats within transects; transects within woodlots; woodlots within districts.
- whenever your measurements are clustered spatially or temporally within randomly sampled units i.e., in *blocks*, which are typically analyzed as random effects.
- whenever you divide up plots, or families, clones, ponds, etc, and apply separate fixed treatments to subplots, or siblings, pond-halves, etc.
- whenever you take measurements on related individuals.
- whenever you measure subjects or other sampling units *repeatedly*.

## Caution when analyzing random effects

Use linear mixed effects models to analyze random effects.

In R, `lm()` is for fixed effects only. Do not use if you have random effects.

In R, use `lmer()` in the `lme4` and `lmerTest` packages or `lme()` in the `nlme` package to analyze models containing random effects. These packages model the variance structure of random effects explicitly.

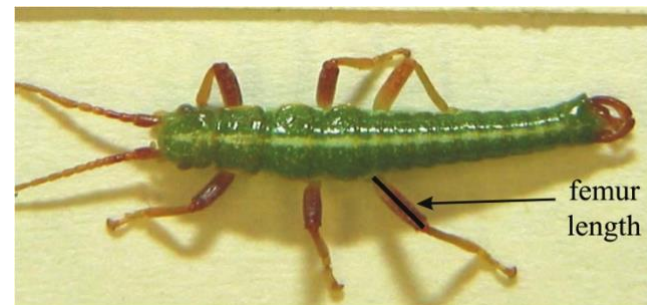
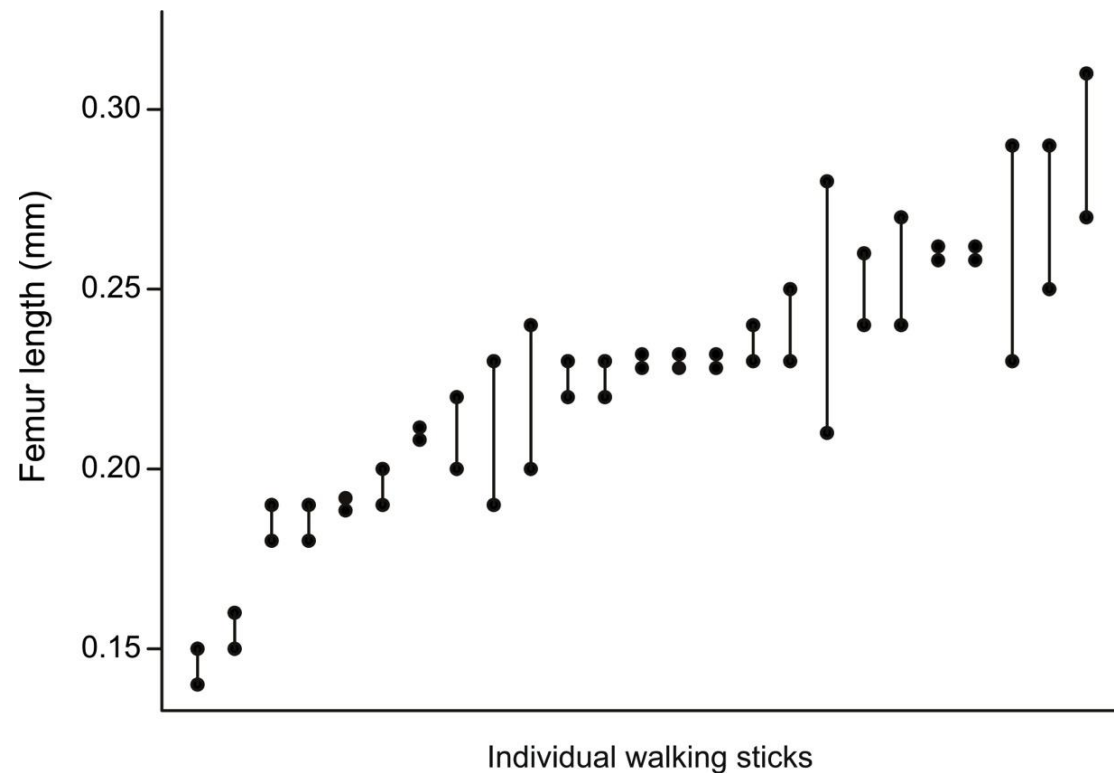
Tests of treatment effects using the Kenward-Roger or Satterthwaite approximations for degrees of freedom are most accurate (available in `lmerTest` package).

## Attributes of linear mixed-effects models

- They are linear models that include both fixed and random effects.
- They model different error variance for each level of random variation.
- Estimation and testing are based on restricted maximum likelihood (REML), which can handle unequal sample size.
- *P*-values for tests of fixed effects are conservative when designs are unbalanced.
- Implemented in the `lme4` and `lmerTest` and `nlme` packages in R.

## Example 1: Study of measurement repeatability (simple nested design)

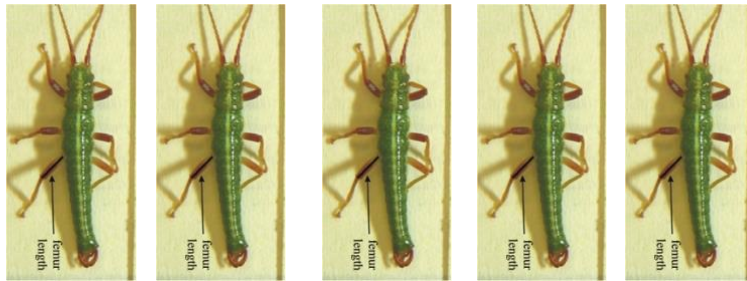
The walking stick, *Timema cristinae*, is a wingless herbivorous insect on plants in chaparral habitats of California. Nosil and Crespi (2006) measured individuals using digital photographs. To evaluate measurement repeatability they took two separate photographs of each specimen. After measuring traits on one set of photographs, they repeated the measurements on the second set.



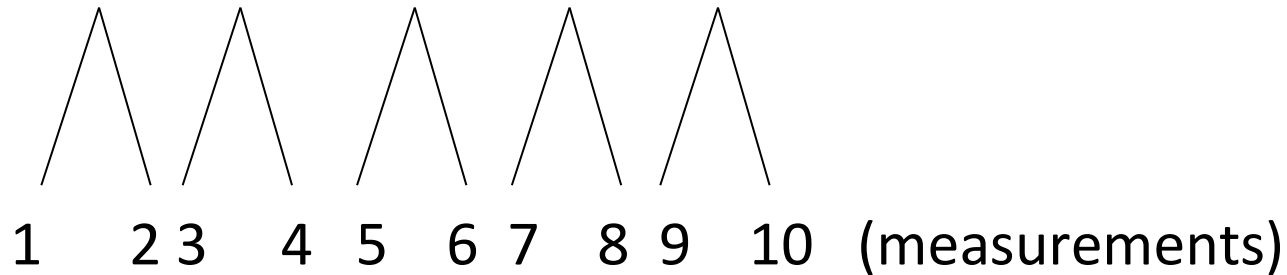


## Example 1: Study of measurement repeatability (simple nested design)

Linear mixed model:  $Y_i = \beta_0 + u_i + \text{error}$  for individual bugs  $i$



Including a random effect means that the analysis follows design



Individual bugs are the random groups in this study, with two repeated measurements per group.

Model has two parts, each with its own source of error variance:

- 1) Random part: the measurement of individual bug  $i$ :  $u_i \pm \text{random error}$ .
- 2) Fixed part: the mean of bug means:  $\beta_0 \pm \text{bug error}$  (i.e., variation among bugs)

## Example 1: Study of measurement repeatability (simple nested design)

```
library(lmerTest)
```

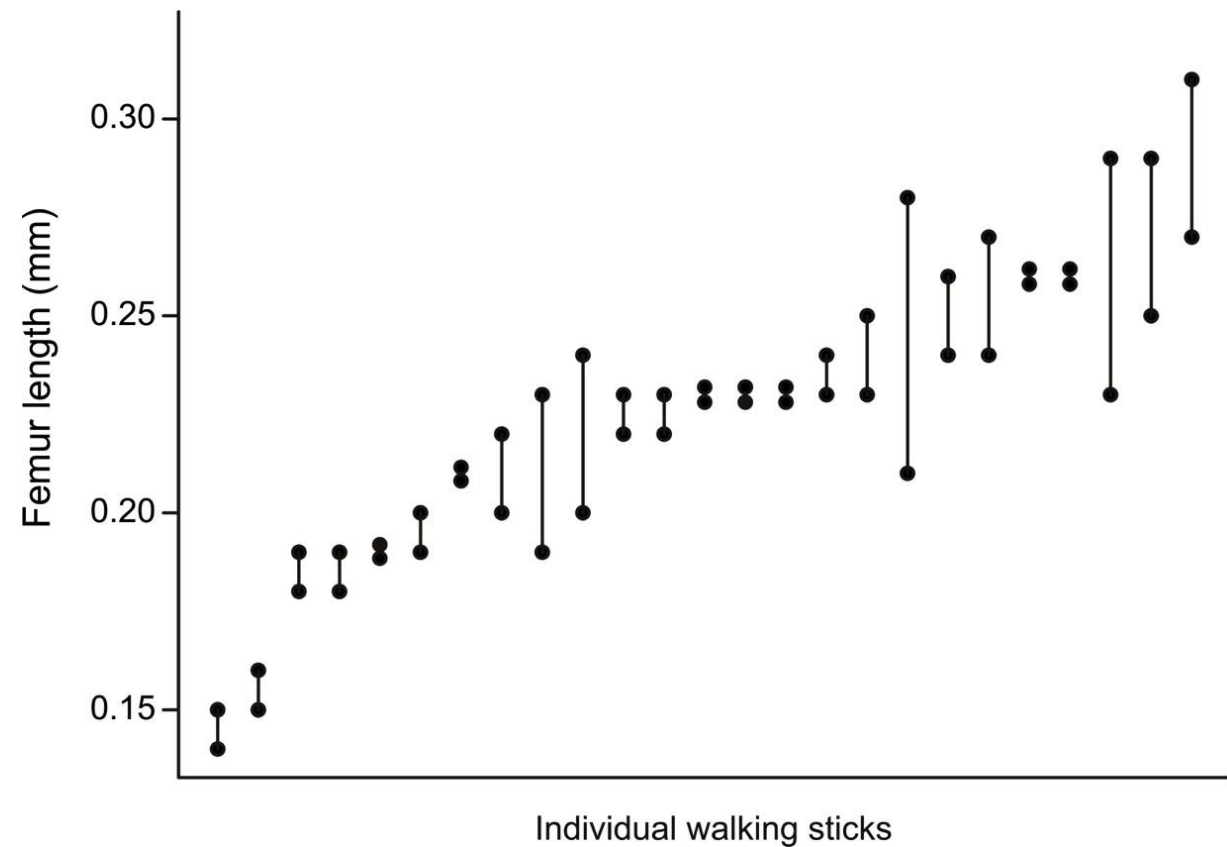
```
z <- lmer(femurlength ~ 1 + (1|individual))
```

**Fixed** part of the formula:

```
femurlength ~ 1
```

**Random** part of the formula

```
+ (1|individual)
```



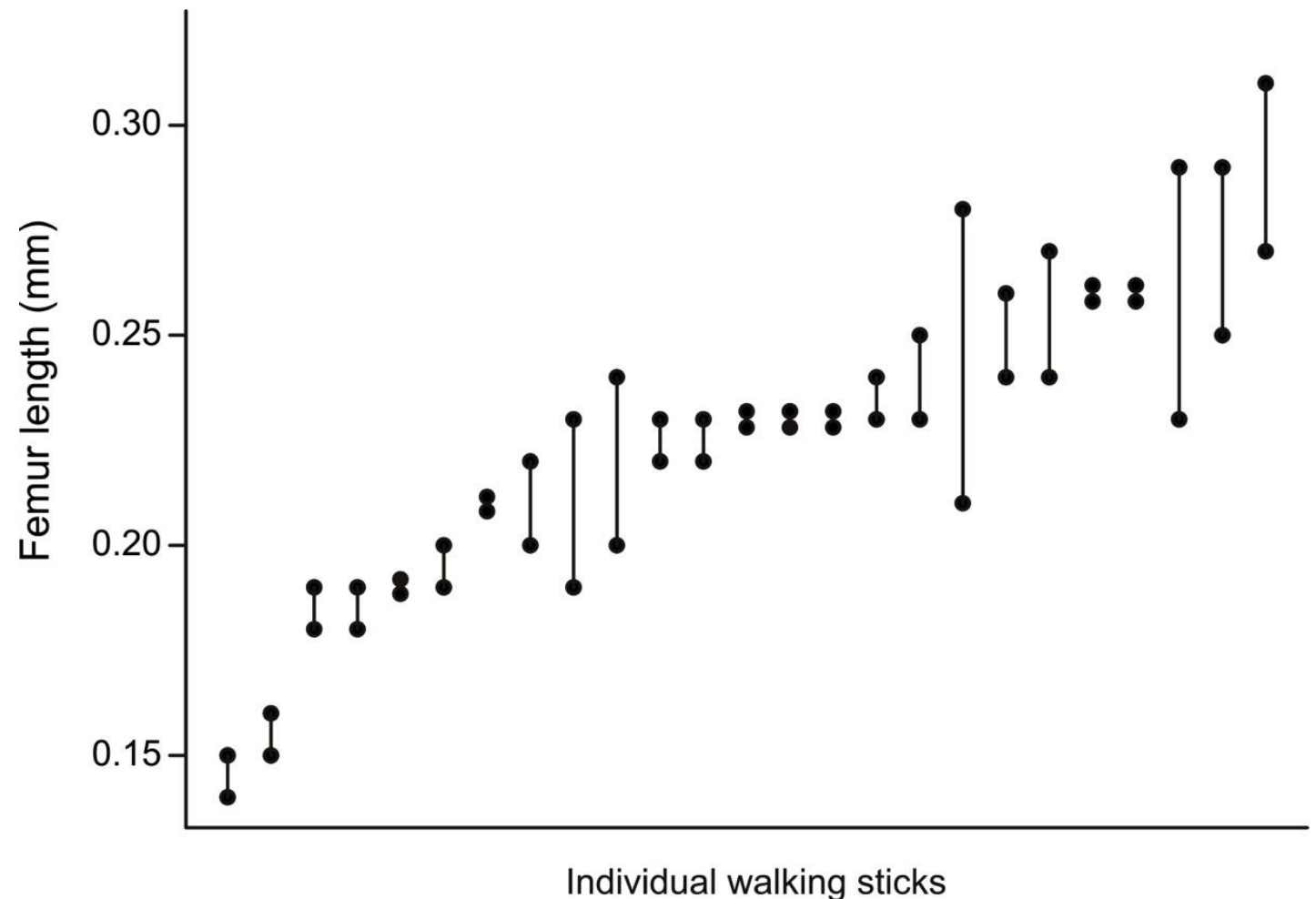
## Example 1: Study of measurement repeatability (simple nested design)

The **fixed** part of the formula instructs R to fit a constant (intercept) based on the fitted values of the random groups (individual bugs)

```
femurlength ~ 1
```

The **random** part of the formula instructs R to fit a constant (an intercept) to the two measurements within each individual. This yields a fitted value for each individual walking stick.

```
+ (1 | individual)
```



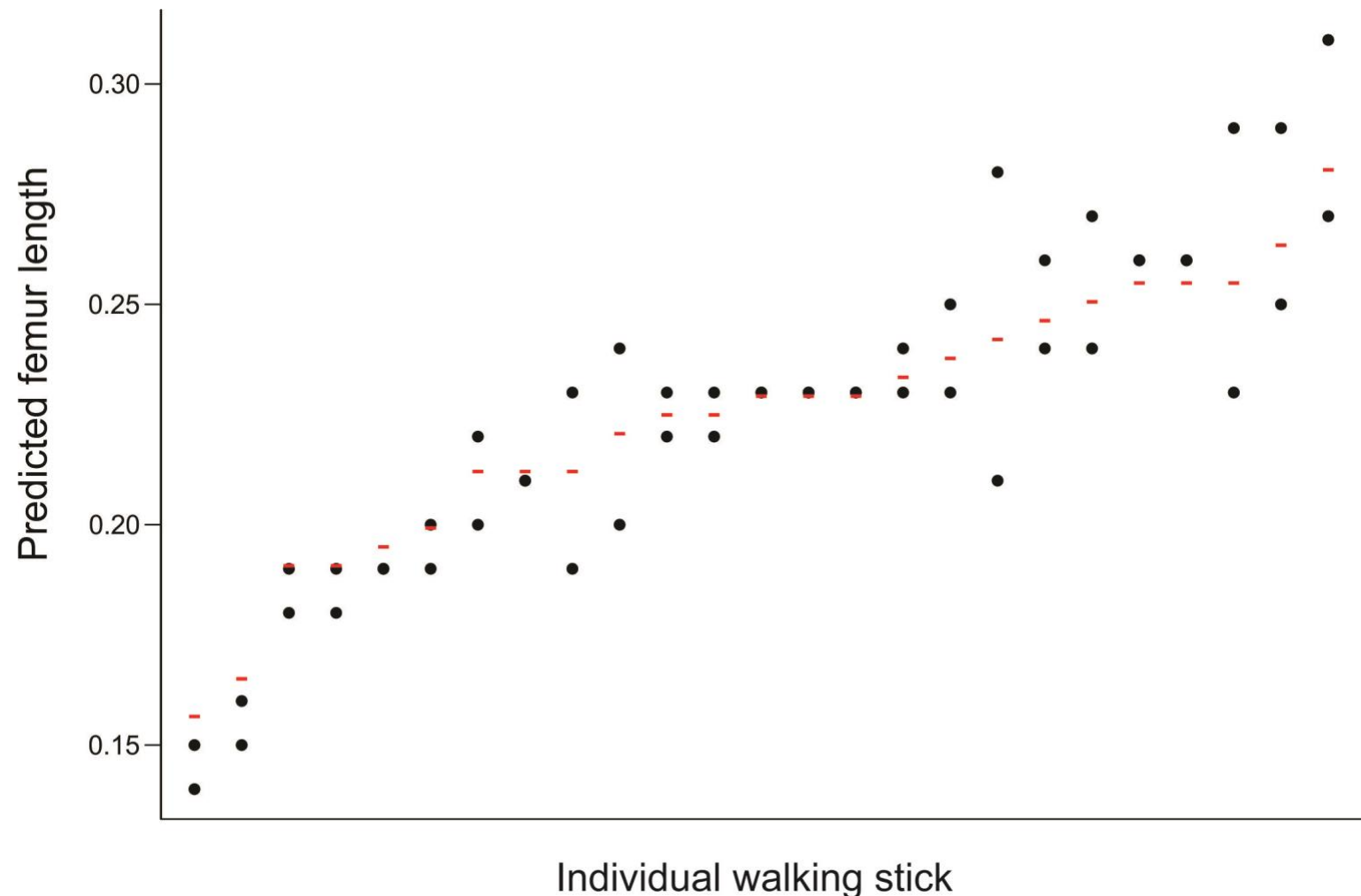
## Example 1: Study of measurement repeatability (simple nested design)

```
z <- lmer(femurlength ~ 1 + (1|individual))
```

```
fitted(z) # yields best linear unbiased predictors (BLUPs) shown in plot:
```

R fits them all together,  
rather than in two stages,  
yielding variance  
components and BLUPs.

The BLUPs are not the  
means for each insect. They  
are “shrunk” towards the  
centre compared with the  
individual insect means.



## Example 1: Study of measurement repeatability (simple nested design)

`VarCorr(z)` extracts the variance components (square the standard deviations to obtain the variances)

Groups	Name	Std.Dev.
individual	(Intercept)	0.032464 = variability among random groups (individuals)
Residual		0.018868 = variability within random groups (usual error)

We can use these quantities to calculate the fraction of variation that is among individuals (repeatability):

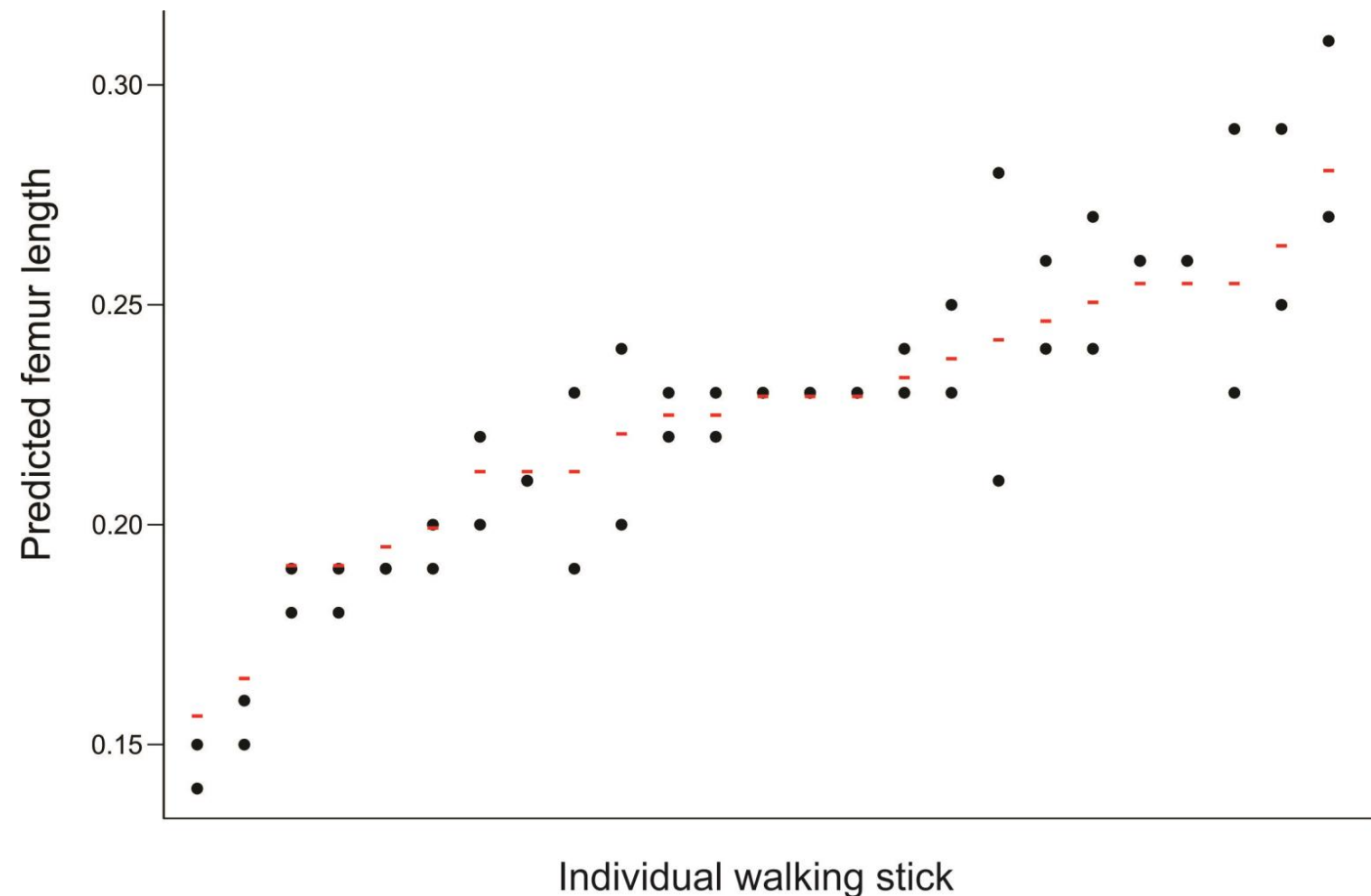
$$\text{repeatability} = \sigma^2_{\text{among}} / (\sigma^2_{\text{among}} + \sigma^2_{\text{within}})$$

$$\begin{aligned} \text{estimate of repeatability} &= 0.032464^2 / (0.032464^2 + 0.018868^2) \\ &= 0.75 \end{aligned}$$

## Example 1: Study of measurement repeatability (simple nested design)

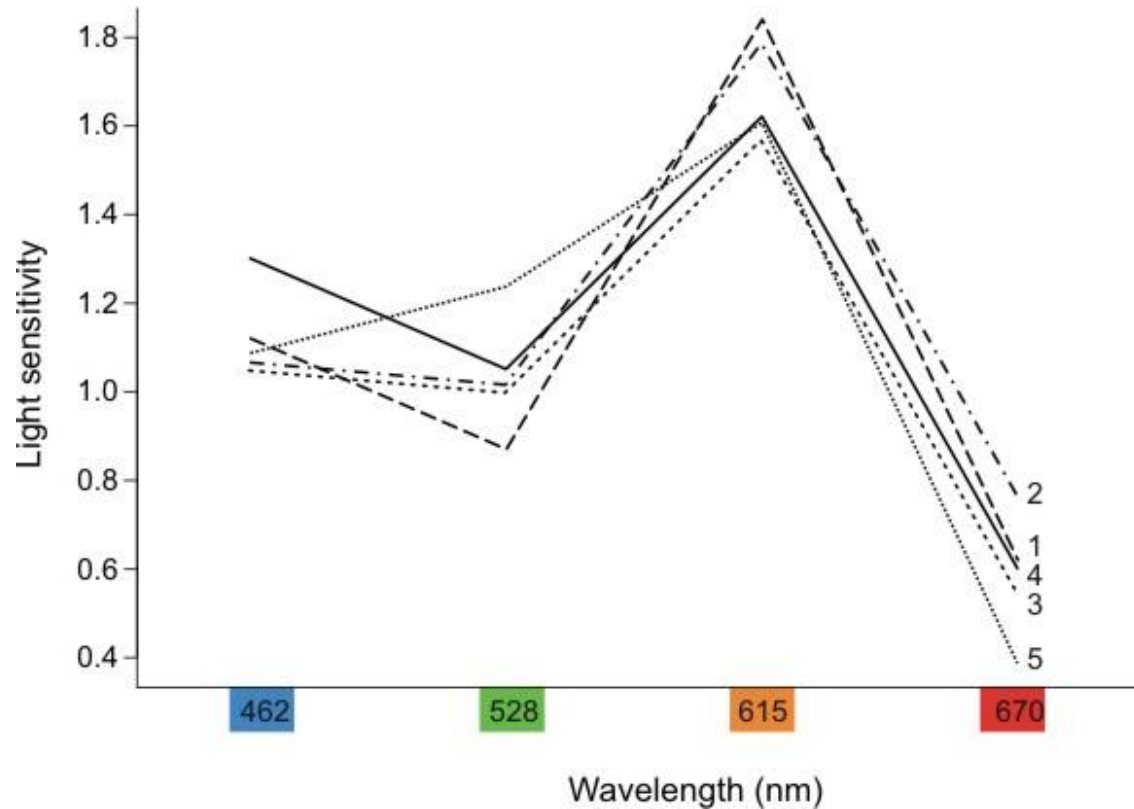
Estimate of repeatability = 0.75

That is, an estimated 75% of the variation among measurements is true variation among individual insects in their femur lengths, the rest is measurement error.



## Example 2: “Subjects by treatment” repeated measures design

Cronly-Dillon and Muntz (1965) used the optomotor response to measure color vision in the goldfish. Each fish was tested at different wavelength in random order.



A large value indicates that the fish has high sensitivity --- it can detect a low light intensity.

### Factors:

Wavelength (fixed, repeated measure)

Fish (random)



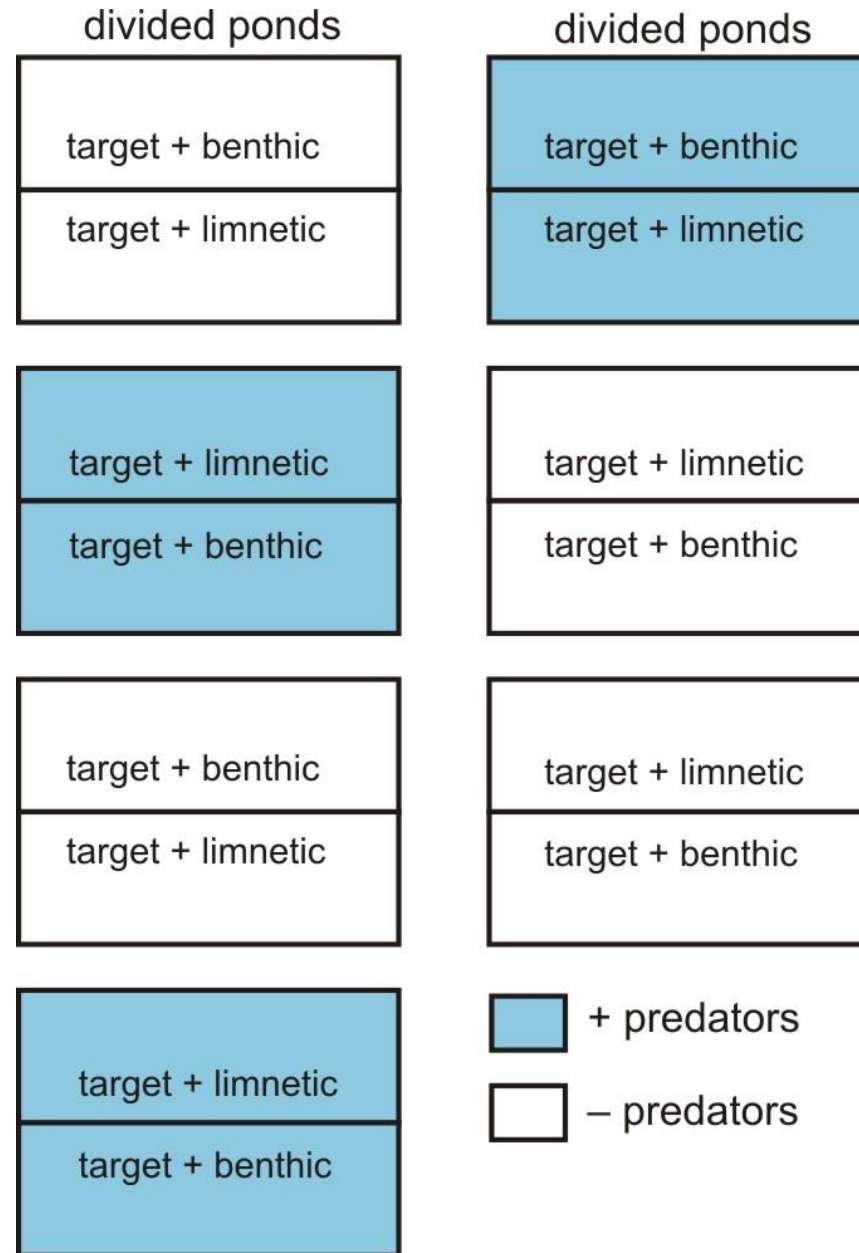
Light sensitivity of 5 goldfish to specific wavelengths of light.

### Example 3: Split plot design



#### Factors:

- Predation treatment (fixed), whole ponds
- Competition treatment (fixed), split ponds
- Pond (random)





## Assumptions of linear mixed-effects models

- As with all linear models: Residuals follow a normal distribution with equal variance.
- Groups are randomly sampled from a “population” of groups (i.e., are independent and sampled without bias).
- Group means have a normal distribution.
- Replicates within groups are also randomly sampled (i.e. independent and sampled without bias).
- No carry-over between repeated measurements on the same subject.
- Sphericity: the variances of the differences between all pairs of factor levels are equal. (Problems can arise when one of the factors is time.)

## Example 4: “Subjects by trials” repeated measures design

*Ecology*, 65(6), 1984, pp. 1780–1786  
© 1984 by the Ecological Society of America

### GRANIVORY IN A DESERT ECOSYSTEM: EXPERIMENTAL EVIDENCE FOR INDIRECT FACILITATION OF ANTS BY RODENTS<sup>1</sup>

D. W. DAVIDSON

*Department of Biology, University of Utah, Salt Lake City, Utah 84112 USA*

AND

R. S. INOUE<sup>2</sup> AND J. H. BROWN

*Department of Ecology and Evolutionary Biology, University of Arizona,  
Tucson, Arizona 85721 USA*

### Factors:

Rodent treatment (fixed)

Date (fixed)

*Plot* (random)

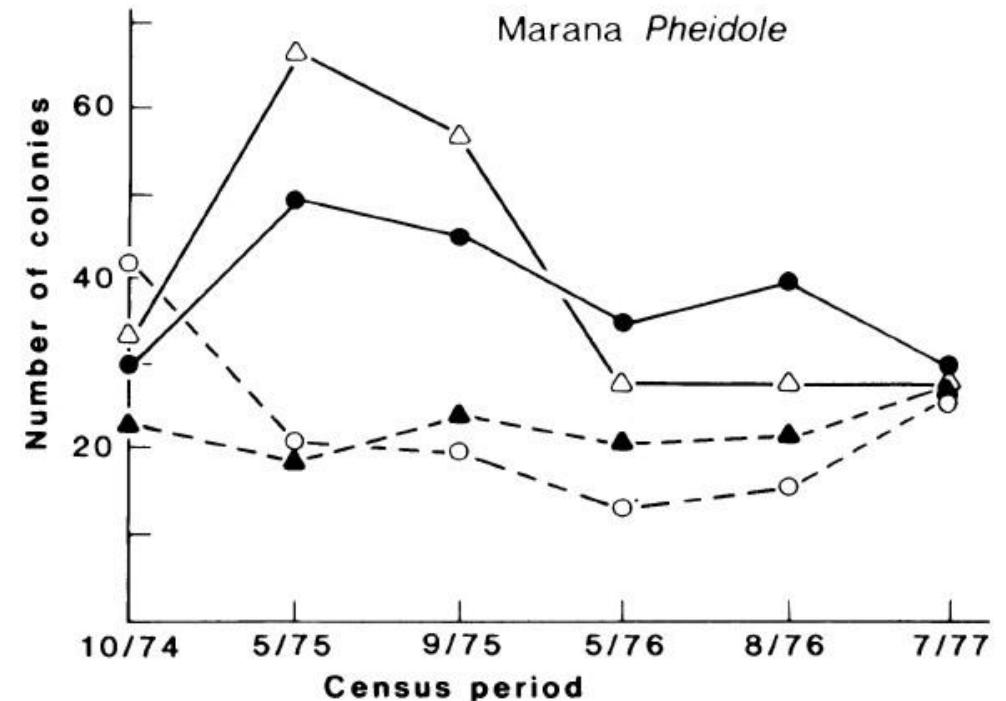


FIG. 3. Changes in density of *Pheidole* spp. (including *P. xerophila tucsonica*, *P. sitarches*, and *P. gilvescens*) on two rodent removal plots (—) and two control plots (— —) at Marana, Arizona over a 2<sup>3</sup>/<sub>4</sub>-yr period.

## Example 4: Warning about “subjects by trials” repeated measures design

### Factors:

Rodent treatment (fixed)

Census date (fixed)

Plot (random)

Tempting to analyze these data using a linear mixed-effects model:

Fixed part of formula:

```
ncolonies ~ treatment + date
```

Random part of formula:

```
+ (1|plot)
```

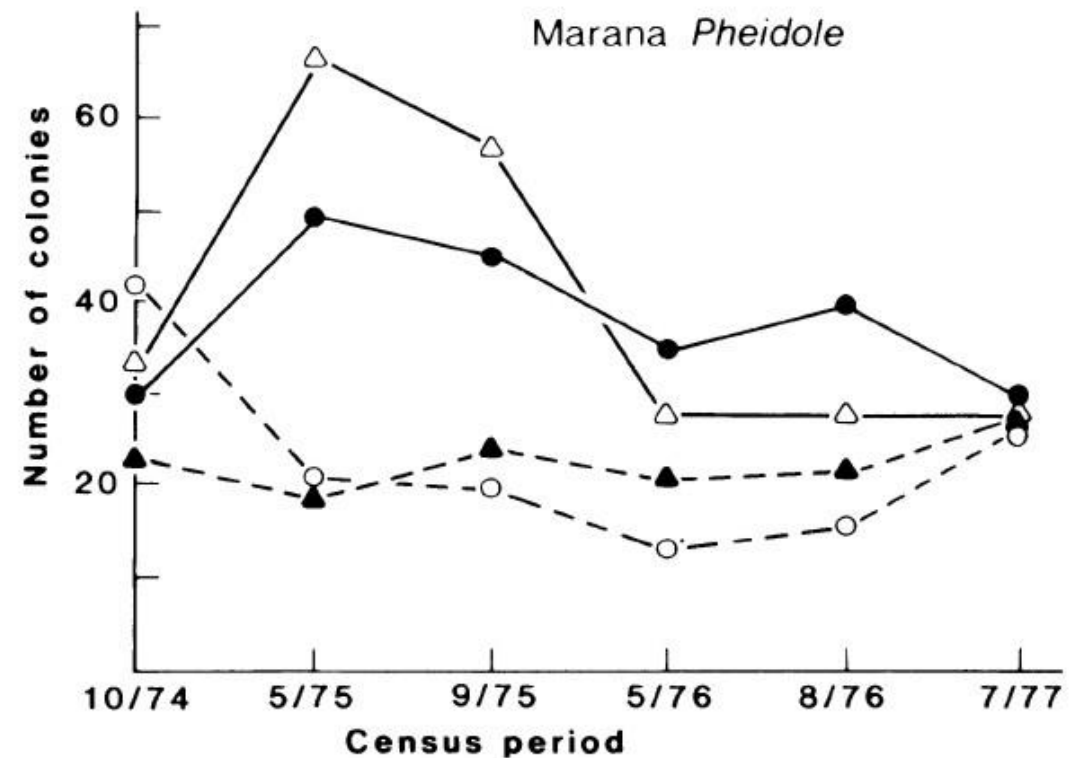


FIG. 3. Changes in density of *Pheidole* spp. (including *P. xerophila tucsonica*, *P. sitarches*, and *P. gilvescens*) on two rodent removal plots (—) and two control plots (— —) at Marana, Arizona over a 2<sup>3</sup>/<sub>4</sub>-yr period.

## Example 4: Warning about “subjects by trials” repeated measures design

Problem: Including census date as a factor would likely lead to a violation of the sphericity assumption: among plots, the variance of the pairwise difference between values of the response variable will be lower between nearby dates than dates that are farther apart.

Increases Type 1 errors  
(*P*-values inaccurate)

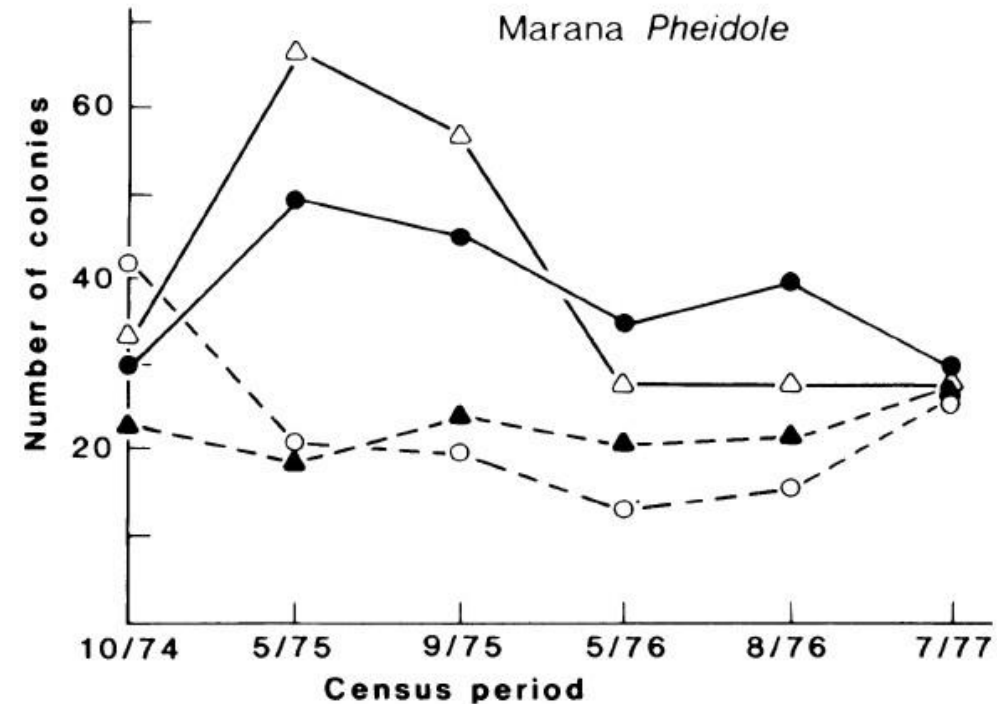


FIG. 3. Changes in density of *Pheidole* spp. (including *P. xerophila tucsonica*, *P. sitarches*, and *P. gilvescens*) on two rodent removal plots (—) and two control plots (— —) at Marana, Arizona over a 2<sup>3</sup>/<sub>4</sub>-yr period.

## Example 4: Warning about “subjects by trials” repeated measures design

Analyses of growth curves in time might have the same problem.

Any repeated measures experiment in which the treatment levels are given in the same sequence (i.e., not in random order) might have the same problem.

Sphericity correction is possible.

Anova ( ) in car package:

Mauchly test for Sphericity

Greenhouse-Geisser and Huynh-Feldt corrections to  $P$ -values.

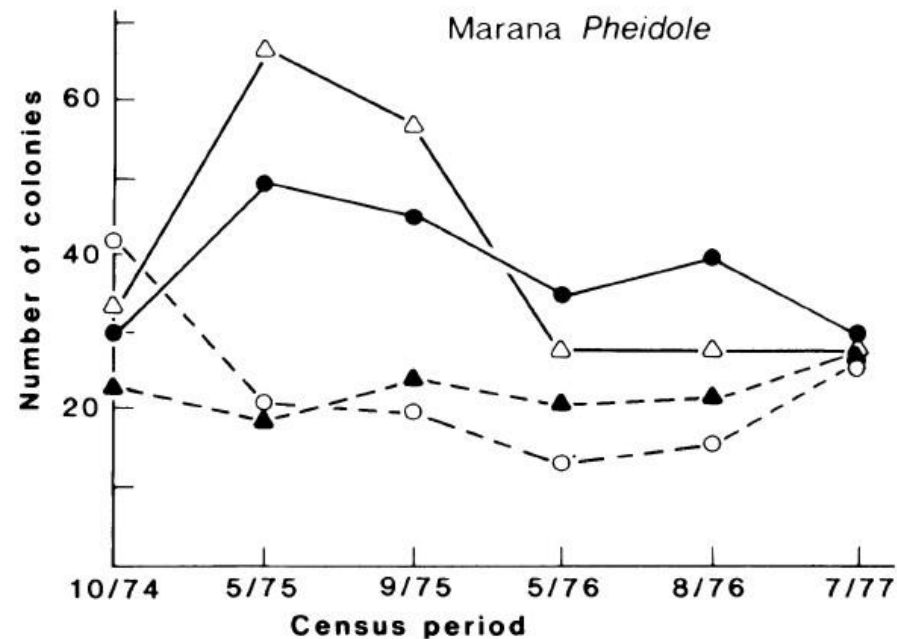


FIG. 3. Changes in density of *Pheidole* spp. (including *P. xerophila tucsonica*, *P. sitarches*, and *P. gilvescens*) on two rodent removal plots (—) and two control plots (---) at Marana, Arizona over a 2¾-yr period.

## Where to get further advice

I have found Quinn & Keough to be very useful in understanding design and assumptions of mixed-effects models (though it won't help you with modeling using maximum likelihood in R).

Quinn & Keough 2002. Experimental design and data analysis for biologists

Online books on mixed-effects models in R (see Books tab at course web site):

Pinheiro and Bates (2000). Mixed-effects models in S and S-PLUS.

Zuur et al (2009). Mixed effects models and extensions in ecology with R.

Article:

Bates, Douglas, et al. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1-48.

## **Discussion paper for next week:**

Murtaugh (2007) Simplicity and complexity in data analysis.  
Ecology 88: 56–62.

Presenters: Zandra & Settare

Moderators: Isidora & Michelle