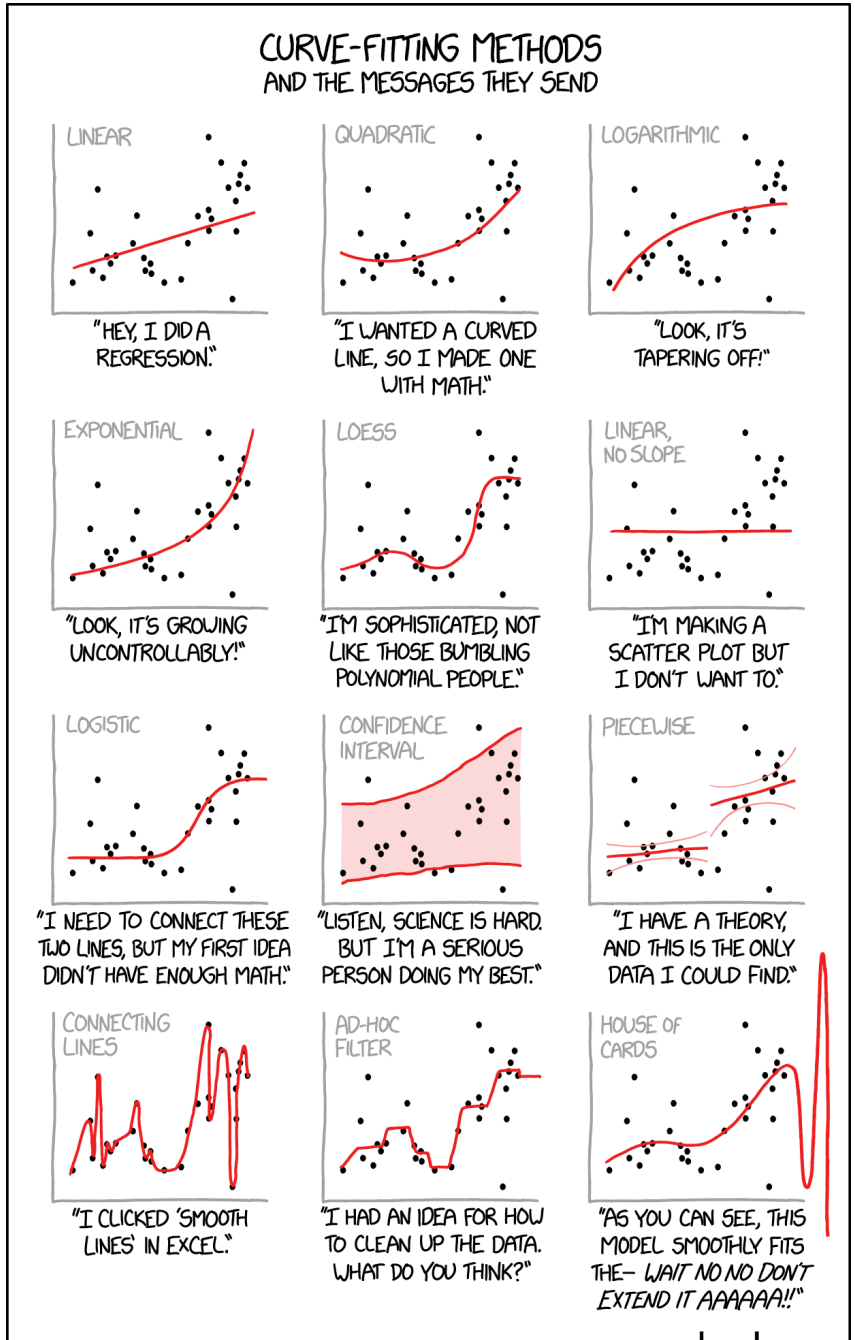
BISC-869, Mixed Effects Models

February 26, 2020


CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND
LINEAR
"HEY, I DID A REGRESSION."
QUADRATIC
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."
LOGARITHMIC
"LOOK, IT'S TAPERING OFF!"
EXPONENTIAL
"LOOK, IT'S GROWING UNCONTROLLABLY!"
LOESS
"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."
LINEAR, NO SLOPE
"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."
LOGISTIC
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."
CONFIDENCE INTERVAL
"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."
PIECEWISE
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND."
CONNECTING LINES
"I CLICKED 'SMOOTH LINES' IN EXCEL."
AD-HOC FILTER
"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"
HOUSE OF CARDS
"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE– WAIT NO NO DON'T EXTEND IT AAAAAA!!"

source: xkcd

What are fixed effects?

Predetermined categories of a variable of direct interest, repeatable.

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: 'hedgerows' in agriculture

Restoration experiment to investigate the effect of native plant introductions along field edges in California (data courtesy of Claire Kremen, UBC).

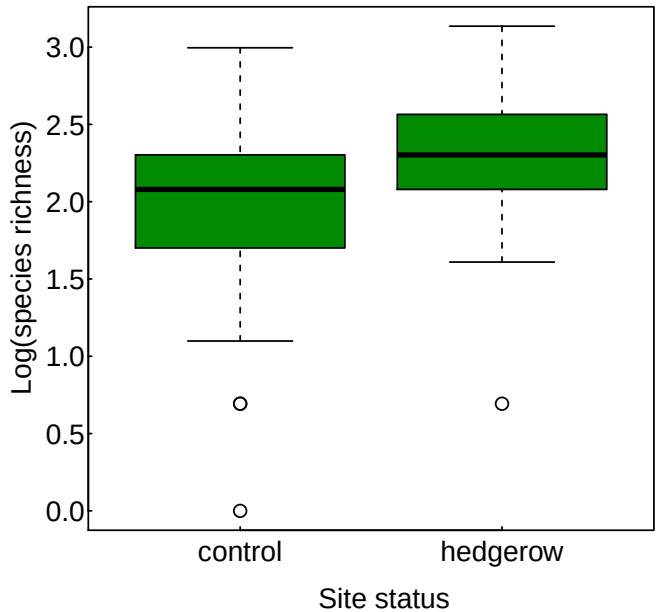
Before

After

Selected 15 sites. 5 sites were restored and 10 sites were not. All sites were sampled multiple times per year for $\sim$ 10 years. This is a "BACI" (before-after-control-impact) design.

Fixed effect: "restored" (aka "hedgerow") vs "not restored" (aka "control")

```
out <- lm(log.richness~status, data=hh)
summary(out)

Call:
lm(formula = log.richness ~ status, data = hh)

Residuals:
     Min       1Q   Median       3Q      Max
-1.95216 -0.20563  0.01751  0.35042  1.04357

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.95216    0.05570  35.046  < 2e-16 ***
statushedgerow  0.33291    0.09905   3.361  0.00105 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4982 on 115 degrees of freedom
Multiple R-squared:  0.08944, Adjusted R-squared:  0.08152
F-statistic:  11.3 on 1 and 115 DF,  p-value: 0.001055

drop1(out, test='F')

Single term deletions

Model:
log.richness ~ status
       Df Sum of Sq    RSS     AIC F value   Pr(>F)
<none>              28.546 -161.05
status  1    2.8039 31.350 -152.09  11.296 0.001055 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



It appears there is a significant difference between restored and non-restored sites in the observed level of species richness.

Note that we have 115 degrees of freedom here.

What are random effects?

Randomly sampled categories of a variable, representing groups of measurements or units. For example:

- families made up of siblings
- subjects measured repeatedly ("repeated measures")
- transects of quadrats in a sampling survey
- field plots of plants
- environment chambers containing aquaria

Groups are assumed to be randomly sampled from a population of groups. Therefore, conclusions reached about groups can be generalized to the population of groups.

What are random effects?

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

- field plots
- environment chambers containing aquaria

In other cases, measuring the variance associated with different levels of random groupings is a major point of the study.

- families made up of siblings
- subjects measured repeatedly ("repeated measures")
- transects of quadrats in a sampling survey

In either case, random effects must be incorporated into the model, because units within groups are not independent (e.g., repeated measures). Modeling random effects explicitly avoids pseudoreplication.

Most statistical packages assume that all factors are fixed unless you instruct otherwise.

Designating factors as random takes extra work and probably a read of the manual.

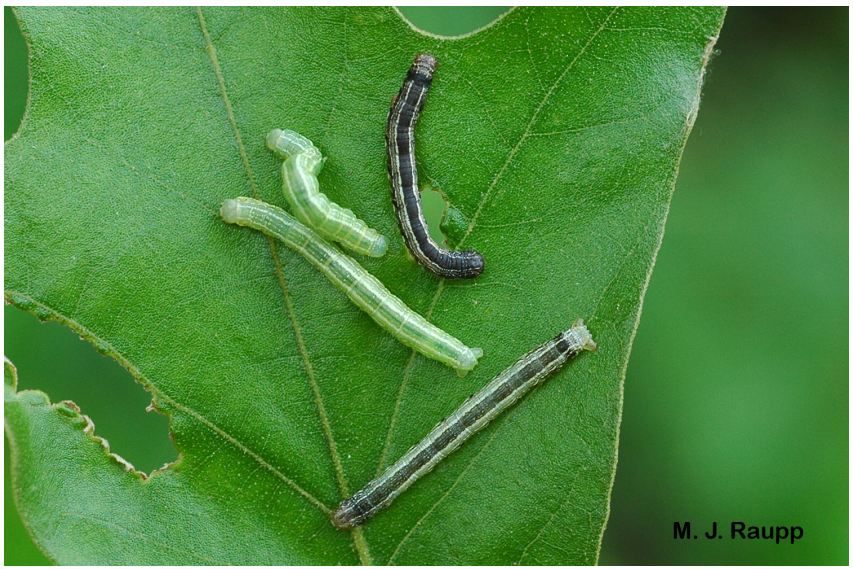
In R, `lm` assumes that all effects are fixed. Do not use `lm` if you have random effects. Instead, use linear mixed effects models to analyze random effects.

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

Note: "Mixed-effects" refers to models that include both fixed and random effects.

Futuyma and Philippi, *Evolution*, 1987

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



Research questions:

- What is the effect of tree species on growth?
- How much do clones vary in growth?

The latter is a not a question about specific clones but about the population of clones.

Interaction plot of responses

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

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

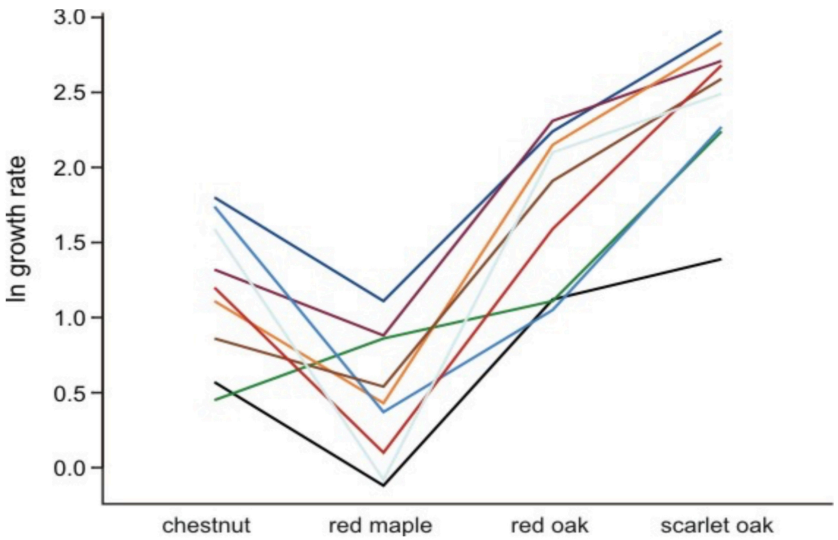


Figure shows mean growth of caterpillars from 9 families (clones) raised on four tree species.

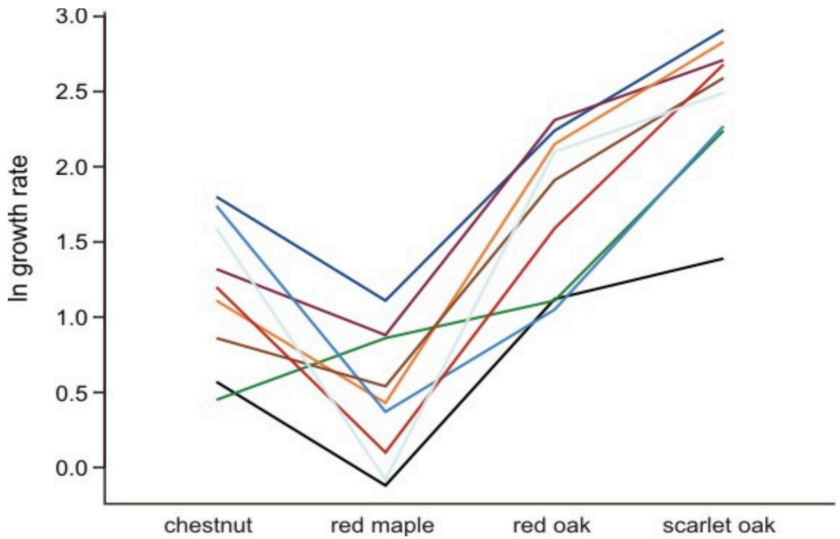
$N = 326$ caterpillars total.

Does this interaction plot "show the data?"

The unit of replication for the test of treatment effect is the clone not the caterpillar.

Why the analysis is different when there are random effects

If clones are random and they vary, then caterpillars from the same clone are not independent. There are only 9 clones, and it would be pseudoreplication to base our test of a treatment effect on the number of caterpillars.

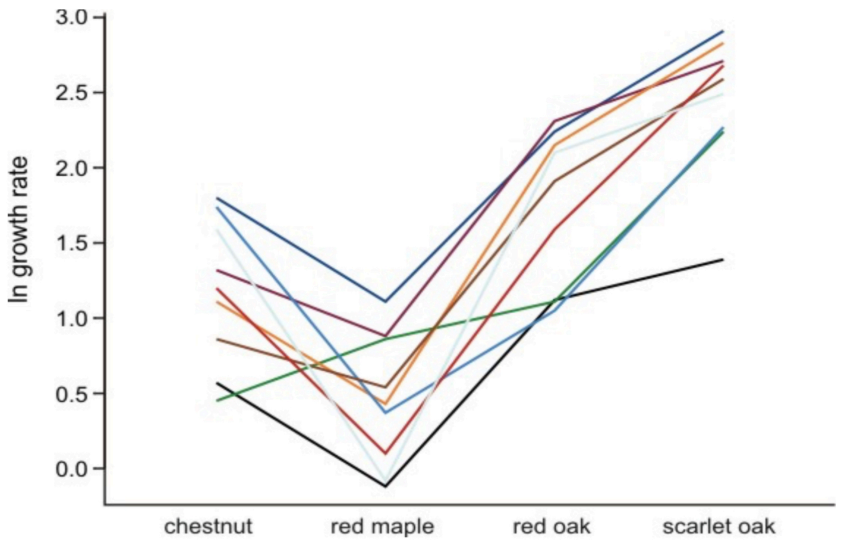


Always report df with F-statistic in your papers to prove you analyzed such data correctly.

Why the analysis is different when there are random effects

The presence of a random factor adds another layer of random variation. A linear model effectively then has two stages:

1. Within a clone, tree species affects mean growth + caterpillar error.
2. Mean effect of tree species in the population is modeled as the mean of clone means + clone error (not caterpillar error).



More reasons why analysis is different with random effects

1. Unlike fixed groups, the means of the random groups (e.g., clones) are not of direct interest. Instead, interest is focused on the variance among random groups (variance components). One purpose of the Futuyma & Philippi experiment was to estimate these variances.

2. When a design including random effects is unbalanced, the standard F-statistics as calculated above are not F-distributed. Standard ANOVA table calculations don't work with unequal sample sizes.

3. With unbalanced designs, the F-statistics and degrees of freedom for fixed effects in mixed models are approximations. `lmer` won't give P-values at all (`lmerTest` will do so).

You have random effects:

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

There is a different error variance for each random effect.

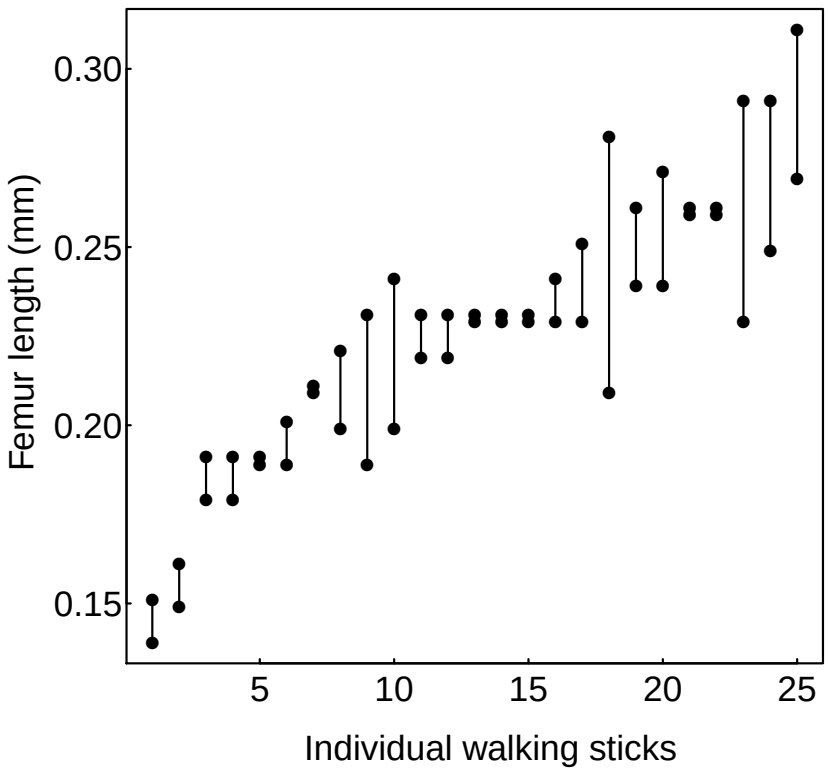
Estimation and testing are based on restricted maximum likelihood (REML) instead of maximum likelihood (more on likelihood next week).

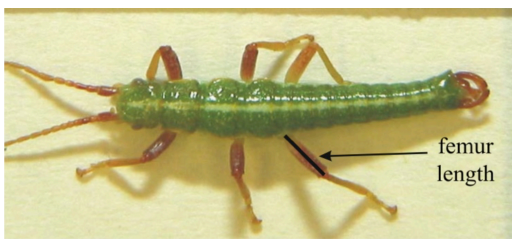
P-values for fixed effects are conservative when designs are unbalanced.

Implemented in the `lme4` and `nlme` packages in R.

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.





Linear model: $Y = \beta_0 + b_i +$ random error

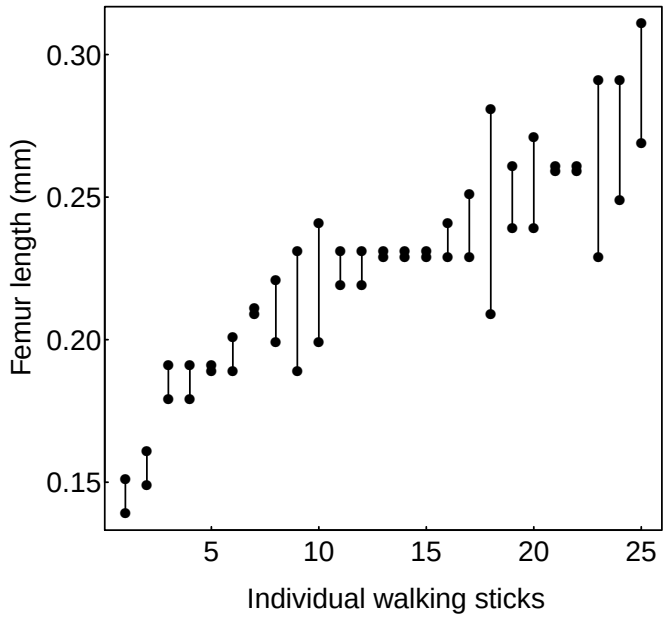
Individual bugs, $i \rightarrow$						...
Measurements $\rightarrow$	1 2	3 4	5 6	7 8	9 10	

The 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: $b_i \pm$ measurement error.
2. Fixed part: the mean of bug means: $\beta_0 \pm$ bug error (i.e., variation among bugs)

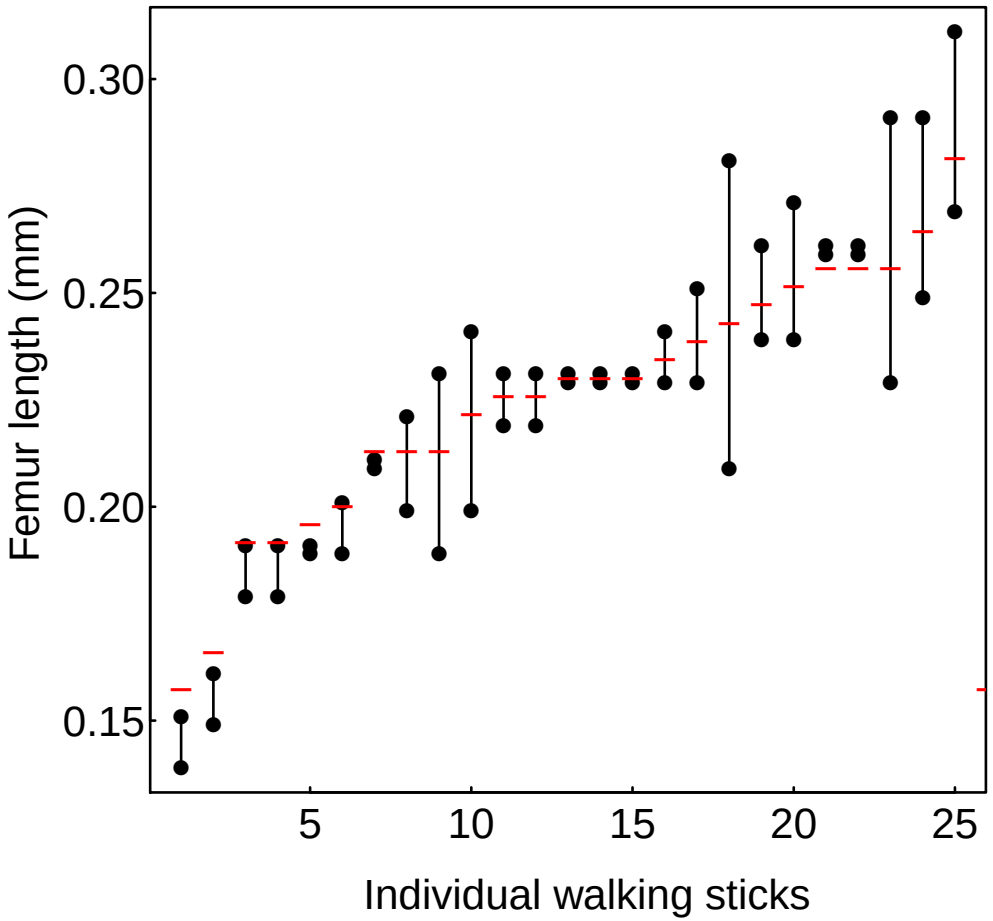
```
library(lme4)
out <- lmer(femurlength~1+(1|individual))
```



The fixed part of the formula " `~1`" instructs R to fit a constant (intercept) based on the fitted values of the random groups (individual bugs)

The random part of the formula " `+(1|individual)`" 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.

`fitted(out)` yields best linear unbiased predictors (BLUPs):



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 group means.

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

```
Groups     Name        Std.Dev.
individual (Intercept) 0.032464
Residual               0.018868
```

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

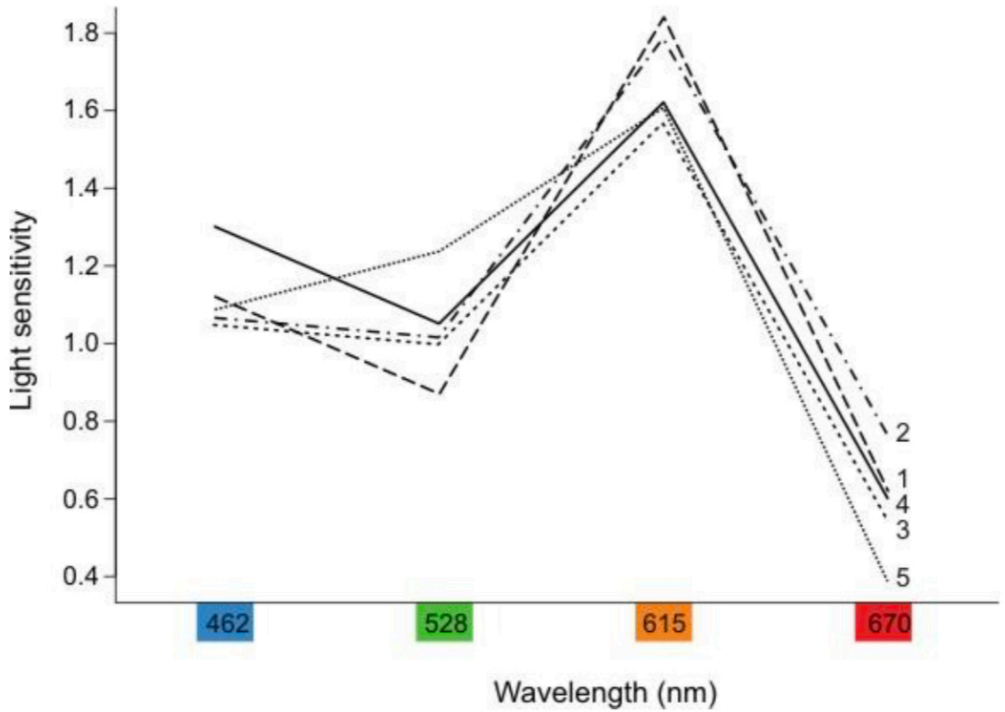
$$\text{repeatability} = \frac{\sigma^2_{\text{among}}}{\sigma^2_{\text{among}} + \sigma^2_{\text{within}}}$$

Thus, our estimate of repeatability is

$$\frac{0.032464^2}{0.032464^2 + 0.018868^2} = 0.75$$

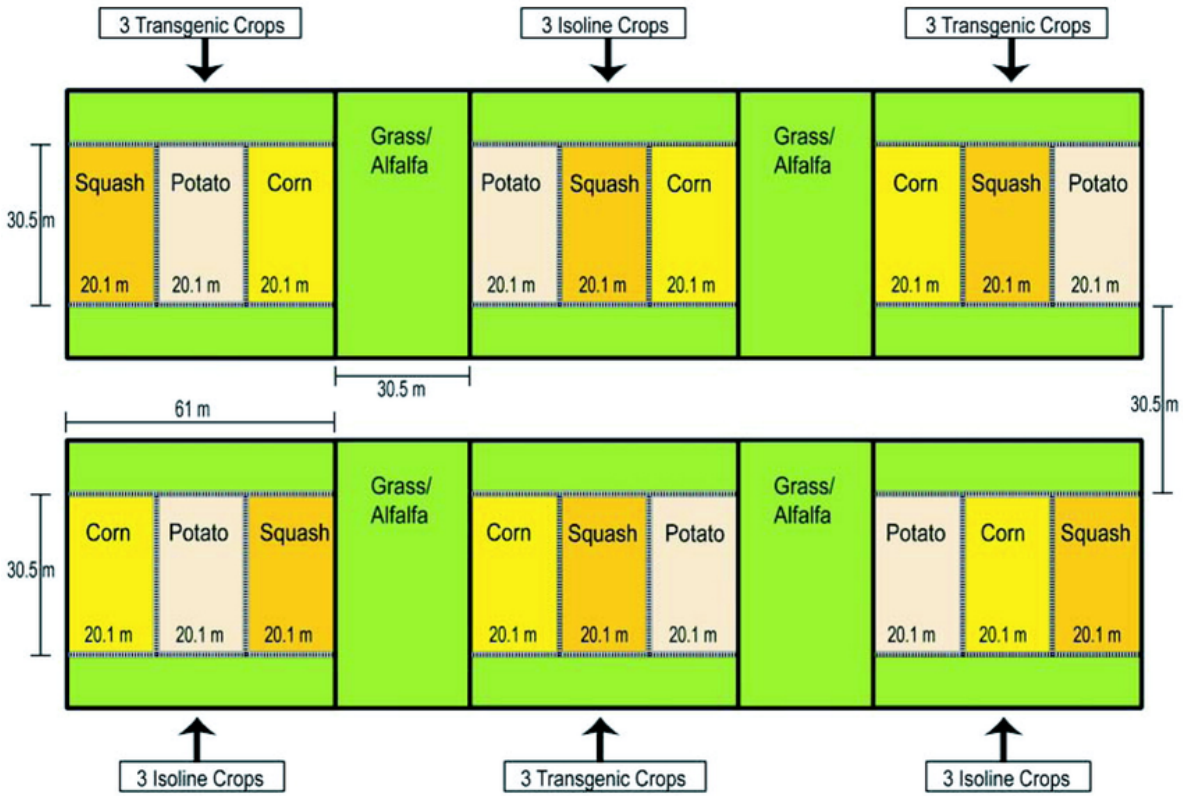
Cronly-Dillon and Muntz (1965) used the optomotor response to measure color vision in the goldfish. Each fish was tested at different wavelengths in random order. A large value indicates that the fish has high sensitivity - it can detect a low light intensity.

Factors: Wavelength (fixed, repeated measure), Fish (random)



Light sensitivity of 5 goldfish to specific wavelengths of light.

Split-plot designs were originally used in agricultural experiments and represent a randomized complete block design, with one or more factors applied to experimental units within each block. A second factor (or set of factors) is then applied to whole blocks, with replicate blocks for each level of this factor.



Variation within groups follows a normal distribution with equal variance among groups.

Groups are randomly sampled from a "population" of groups (i.e., are independent and sampled without bias).

Group effects follow 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.

GRANIVORY IN A DESERT ECOSYSTEM: EXPERIMENTAL EVIDENCE FOR INDIRECT FACILITATION OF ANTS BY RODENTS[1]

D. W. DAVIDSON
Department of Biology, University of Utah, Salt Lake City, Utah 84112 USA

AND

R. S. INOUYE[2] 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)

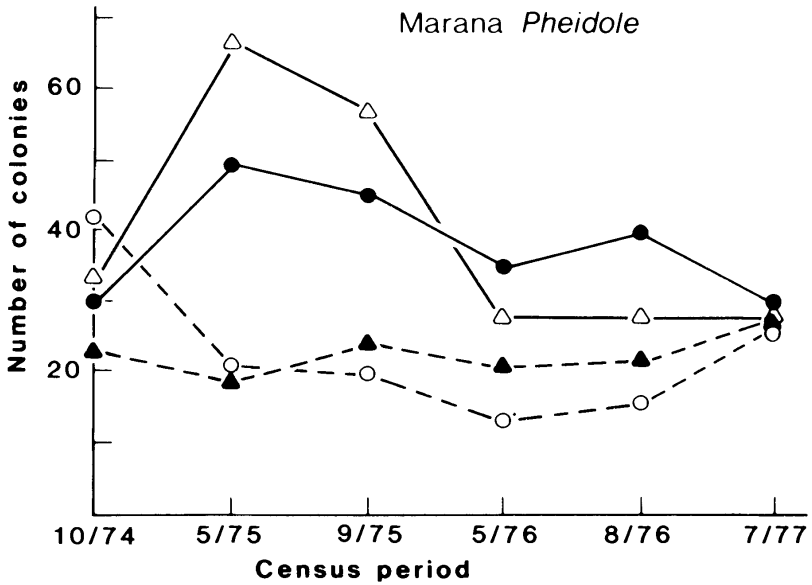


FIG. 3. Changes in density of *Pheidole* spp. (including *P. xerophila tucsonica, P. sitarches,* and *P. gilvescens*) on two rodent removal plots (——) and two control plots (— —) at Marana, Arizona over a 2¾-yr period.

At our Sonoran Desert study site, both of the major taxa exhibited short-term increase in density when the other taxon was experimentally removed. Over the longer term, density compensation continued at a relatively constant level for rodents in the absence of ants. In contrast, beginning ∼ 2 yr after initiation of experiments, ant populations on rodent removal plots showed a gradual but significant decline relative to densities on control plots.

It is tempting to analyze these data using a linear mixed-effects model:

```
lmer(ncolonies~treatment + date + (1|plot), data=...)
```

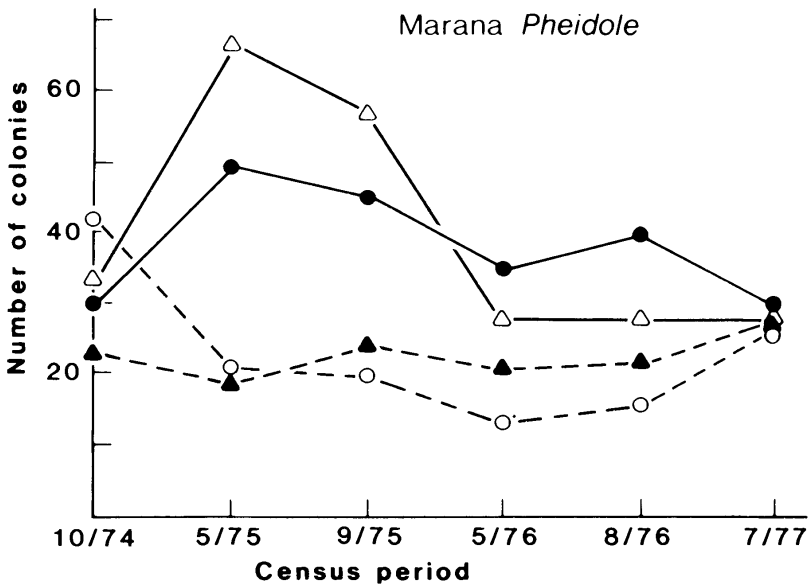


FIG. 3. Changes in density of *Pheidole* spp. (including *P. xerophila tucsonica, P. sitarches,* and *P. gilvescens*) on two rodent removal plots (——) and two control plots (— —) at Marana, Arizona over a 2¾-yr period.

Problem: Including date as a predictor 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.

This will increases Type I error rate (*P*-values inaccurate).

Warning about "subjects by trials" design

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

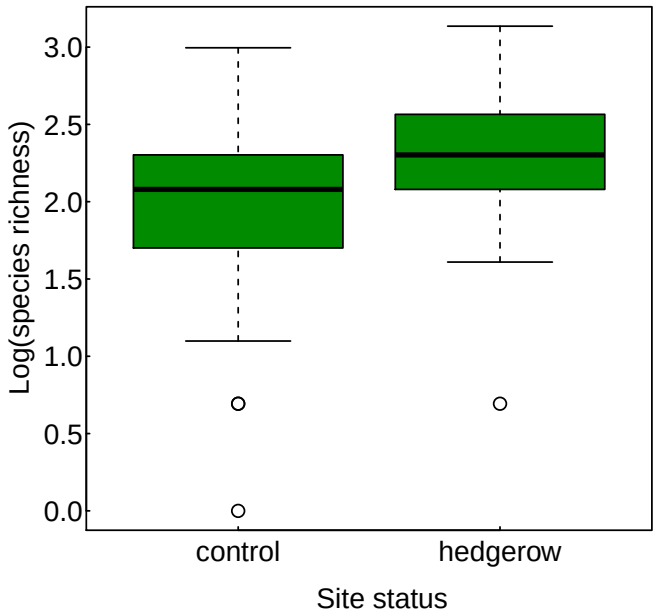
Any repeated measures experiment in which the treatment levels are given in the same sequence (i.e., not in random order) might have the same problem.

Sphericity correction is possible. `Anova()` in car package: Mauchly test for Sphericity.

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

Fixed effect: "restored" (aka "hedgerow") vs "not restored" (aka "control")

```
out <- lm(log.richness~status, data=hh)
summary(out)

Call:
lm(formula = log.richness ~ status, data = hh)

Residuals:
     Min       1Q   Median       3Q      Max
-1.95216 -0.20563  0.01751  0.35042  1.04357

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.95216    0.05570  35.046  < 2e-16 ***
statushedgerow  0.33291    0.09905   3.361  0.00105 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4982 on 115 degrees of freedom
Multiple R-squared:  0.08944, Adjusted R-squared:  0.08152
F-statistic:  11.3 on 1 and 115 DF,  p-value: 0.001055

drop1(out, test='F')

Single term deletions

Model:
log.richness ~ status
       Df Sum of Sq    RSS     AIC F value   Pr(>F)
<none>              28.546 -161.05
status  1    2.8039 31.350 -152.09  11.296 0.001055 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



It appears there is a significant difference between restored and non-restored sites in the observed level of species richness.

Note that we have 115 degrees of freedom here.

Remember: We selected 15 sites. 5 sites were restored and 10 sites were not. All sites were sampled multiple times per year for ~10 years.

```
out <- lm(log.richness~status, data=hh)
summary(out)
```

```
Call:
lm(formula = log.richness ~ status, data = hh)

Residuals:
     Min       1Q   Median       3Q      Max
-1.95216 -0.20563  0.01751  0.35042  1.04357

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.95216    0.05570  35.046  < 2e-16 ***
statushedgerow  0.33291    0.09905   3.361  0.00105 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4982 on 115 degrees of freedom
Multiple R-squared:  0.08944, Adjusted R-squared:  0.08152
F-statistic:  11.3 on 1 and 115 DF,  p-value: 0.001055
```

Notice the change in the SE estimate once we incorporate the random effects. This is because we have removed the *pseudoreplication* present in the initial analysis.

Note also that we have no *P*-values from `lmer`.

```
out <- lmer(log.richness~
          status+(1|site)+(1|year),
          data=hh)
summary(out)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: log.richness ~ status + (1 | site) + (1 | year)
   Data: hh

REML criterion at convergence: 152.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7561 -0.5109  0.1136  0.5323  2.3415

Random effects:
 Groups   Name        Variance Std.Dev.
 site     (Intercept) 0.06609  0.2571
 year     (Intercept) 0.03308  0.1819
 Residual             0.15924  0.3991
Number of obs: 117, groups:  site, 15; year, 9

Fixed effects:
             Estimate Std. Error t value
(Intercept)    1.9359     0.1067  18.144
statushedgerow   0.3682     0.1412   2.608

Correlation of Fixed Effects:
            (Intr)
statushdgrw -0.398
```