

Mixed Effects Models

Lecture 06.1: Linear Mixed Effects Models

Lauren Sullivan

Module: Linear, Nonlinear, and Mixed Effects Models

Readings

Required for class:

- ▶ NA

Optional:

- ▶ Coding Club - Mixed Effects Models
- ▶ Dr. Demos Mixed Models

Mixed Effects Models

Sometimes our linear models do not work because we do not have independence between all of our data points. For example:

1. You want to study how rodent population size varies with seabird density across a series of islands and the islands are different in some way that affects your data (different temperature, nutrient content, predators, etc).
2. You have different people scoring videos of chimp behavior in response to food rewards and each person scores a bit differently (but is still consistent within their own scoring).
3. You want to know how different diets affect milk production in cows, and there are with multiple feeding trials on a single animal.
4. You are interested in the effects of grazing on specific leaf area of plants in the tropics, but some plant families have larger leaves in general than others.

Mixed Effects Models

Mixed Effects Models have both fixed and random effects.

$Y \sim \text{Fixed Effects} + \text{Random Effects}$

By this, we mean that our predictor variables (our X variables) come from a fixed distribution, but the intercepts and/or slopes of our statistical equations are allowed to vary randomly based on our specified random effects.

Practically, I like to think about think about Fixed Effects as the dependent variables that you are interested in, while taking the Random Effects into account. You don't really care so much about random effects - you just want to account for their variability.

Breaking Down the Examples

1. You want to study how rodent population size varies with seabird density across a series of islands and the islands are different in some way that affects your data (different temperature, nutrient content, predators, etc).

Rodent Population Size \sim Seabird Density (Fixed) + Islands (Random)

Breaking Down the Examples

-
2. You have different people scoring videos of chimp behavior in response to food rewards and each person scores a bit differently (but is still consistent within their own scoring).

Chimp Behavior \sim Food Reward Type (Fixed) + Person Scoring Videos (Random)

Breaking Down the Examples

3. You want to know how different diets affect milk production in cows, and there are with multiple feeding trials on a single animal.

Milk Production \sim Diet (Fixed) + Individual Animal (Random)

Breaking Down the Examples

4. You are interested in the effects of grazing on specific leaf area of plants in the tropics, but some plant families have larger leaves in general than others.

Specific Leaf Area \sim Grazing (Fixed) + (Plant Family/Plant Individual) (Random)

Experimental Design - Blocking

One way to avoid heterogeneity is the use of blocking.

- ▶ A set of all treatment plots within an experiment are randomly placed in one of x locations to account for differences in fields. Then block becomes the random effect.

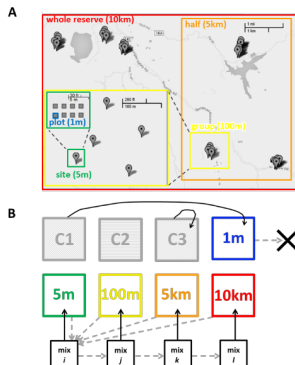
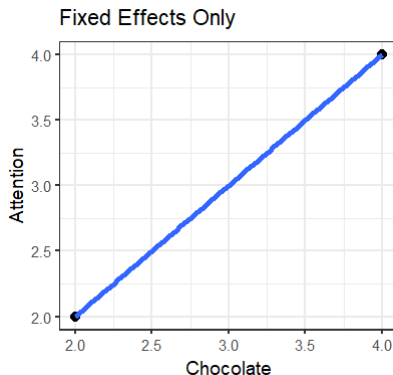
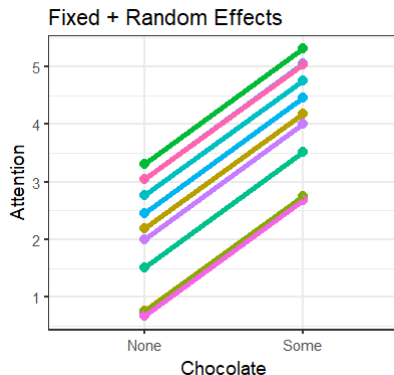


Figure 1: Germain et al. 2017

Random Intercepts

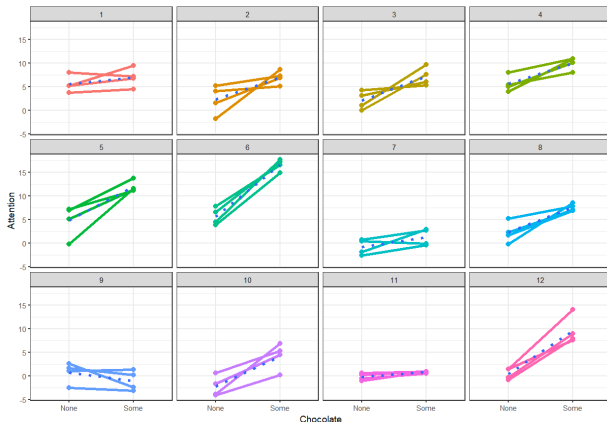
Say you are interested in the effect that **eating chocolate has on attention span**.



The overall trend may be that eating some chocolate increases attention, but you can see there is a clear difference in innate attention span among your subjects.

Random Slopes

But what if you ran the trials multiple times. You might find that the slope of the relationship between chocolate and attention varies as well.



Across trials people may vary in how much or how little their attention increases with eating chocolate.

Libraries

There are several libraries you will need for mixed models.

1. `library(lme4)` - runs the **mixed effects models**
 - ▶ Also works for running generalized linear mixed effects models `'glmer()'`, and non-linear mixed effects models `'nlmer()'`
 - ▶ If you have complicated variance/covariance structures, consider `'library(nlme)'`
2. `library(lmerTest)` - provides your **p-value** with Type I, II or III sums of squares.
3. `library(emmeans)` - allows you to specify **contrasts**
4. `library(MuMIn)` - describes conditional and marginal **variance components**

Model Statements

```
random_intercept <- lmer(Response Variable ~  
  Dependent 1 + Dependent 2 + Dependent 3 +  
  (1|random_intercept), data=your_data, REML=TRUE)  
  
random_slope_intercept <- lmer(Response Variable ~  
  Dependent 1 + Dependent 2 + Dependent 3 +  
  (1 + random_slope|random_intercept),  
  data=your_data, REML=TRUE)
```

Note: you use REML = TRUE for small sample sizes, as it optimizes your estimates using Restricted Maximum Likelihood (REML) instead of Log-likelihood. The log-likelihood estimates are biased low when your sample sizes are small.

Model Statements - Multiple and Nested Random Effects

Sometimes you have more than one random effect, or random effects are nested, like plot is nested within block.

```
random_intercept_multi <- lmer(Response Variable ~  
    Dependent 1 + Dependent 2 + Dependent 3 +  
    (1|random_intercept1) + (1|random_intercept2),  
    data=your_data, REML=TRUE)  
  
random_intercept_nested <- lmer(Response Variable ~  
    Dependent 1 + Dependent 2 + Dependent 3 +  
    (1|random_intercept_block/random_intercept_plot),  
    data=your_data, REML=TRUE)
```

Chocolate Example

For the chocolate and attention example, you could run two models, one with a random intercept, and one with a random slope **and** intercept and find out which fits best.

Test Random Intercepts & Slopes

- Model 2a: Random intercepts
- Model 2b: Random intercepts & Slopes (correlated)

```
Model.2a<-lmer(Attention~Chocolate
               +(1|SS),
               data=DataSim2g, REML=FALSE)

Model.2b<-lmer(Attention~Chocolate
               +(1+Chocolate|SS),
               data=DataSim2g, REML=FALSE)
```

Chocolate Example

Then compare your model results with `anova(model1, model2)`.

Test Random Slope

- Just as we did in MLM, we will test to see if the random slope improves the fit, `anova(Model.2a, Model.2b)`

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Model.2a	4	494.7545	505.0119	-243.3772	486.7545	NA	NA	NA
Model.2b	6	467.3984	482.7845	-227.6992	455.3984	31.35608	2	2e-07

The model with both the random intercept and slope (Model2b) fits better based on AIC (and many other comparison tests).

Type I, II, and III Sums of Squares

Specifying the sums of squares allows you to be specific about how you want to evaluate the `anova()` of your mixed effects model when you have *multiple dependent variables*.

- ▶ **Type I sums of squares** are sequential and are evaluated in the order they are written. $Y \sim X_1 + X_2 + X_3 + \text{random}$ may not necessarily equal $Y \sim X_3 + X_1 + X_2 + \text{random}$, etc.
- ▶ **Type III sums of squares** are partial, and basically every term in the model is evaluated in light of every other term, including interactions. So it evaluates all combinations of $X_1 + X_2 + X_3$.
 - ▶ If you have a perfectly balanced factorial design, then this is equivalent to Type III sums of squares.
- ▶ **Type II sums of squares** are similar to Type III, but they are marginal - they evaluate in the light of all other terms, but not the interaction terms (so just the additive ones)

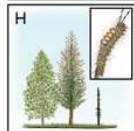
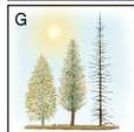
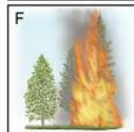
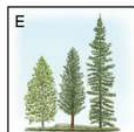
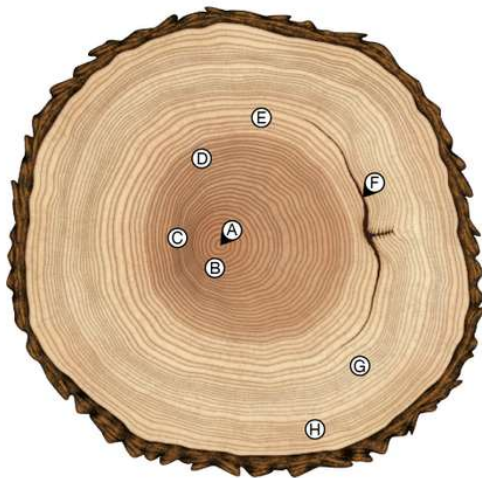
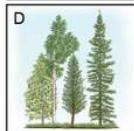
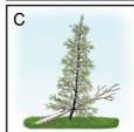
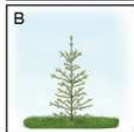
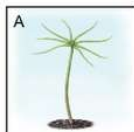
We will be looking at **tree ring data** and how growth of trees is affected by their climatic and soil conditions.

Soil nutrients influence growth response of temperate tree species to drought

Mathieu Lévesque^{1,2*}, Lorenz Walthert¹ and Pascale Weber¹

¹WSL Swiss Federal Institute for Forest, Snow and Landscape Research, Zuercherstrasse 111, Birmensdorf CH-8903, Switzerland; and ²Tree-Ring Laboratory, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA

Tree Rings



Tree Ring Data

Does the diameter at breast height (dbh), the mean temperature (tmean) and the soil nutrient level (C.N) influence the average tree ring width (trw.mean)?

```
tree <- read_csv("../data/LME_Data_TreeGrowth_Clim_Soil_Sites.csv")  
  
model_lme <- lmer(trw.mean ~ dbh + tmean + C.N + (1|site),  
                  data = tree, REML = TRUE)
```

We use site as a random intercept effect in case there is some inherent difference among sites that could obscure the nutrient and climate pattern.

Tree Ring Results

To get the model results, we use the `anova()` function. Don't forget to specify the type of sums of squares (here, `type = 3`)

```
anova(model_lme, type = 3)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##      Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## dbh      22687   22687      1 85.002  9.0511 0.003453 **
## tmean       569      569      1 40.500  0.2270 0.636306
## C.N      11186   11186      1 37.735  4.4629 0.041313 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Both dbh and C.N influence tree ring width. So trees with larger dbh tend to have larger rings, and soil with a lower C.N value have smaller rings.

Variance Components

You can also determine how much of the variance is explained by either the fixed effects alone (marginal - R2m below) or the fixed and random effects (conditional - R2c below).

```
r.squaredGLMM(model_lme)
```

```
##              R2m              R2c  
## [1,] 0.1886414 0.6942555
```

Here, the fixed effects explain 18% of the variance, and the fixed and random effects together explain 69% of the variance. So in this case, the random effect of site explains a lot of the variance.

Tree Ring Results

Let's plot the tree ring width against soil nutrient content and demonstrate the random effects. You can see a lot of differences among sites here.

