

GLAMM Workshop: Linear Mixed Effects Models (LMMs)

Robert Fitch and Mallory Rice
August 11th, 2020

Outline for Today

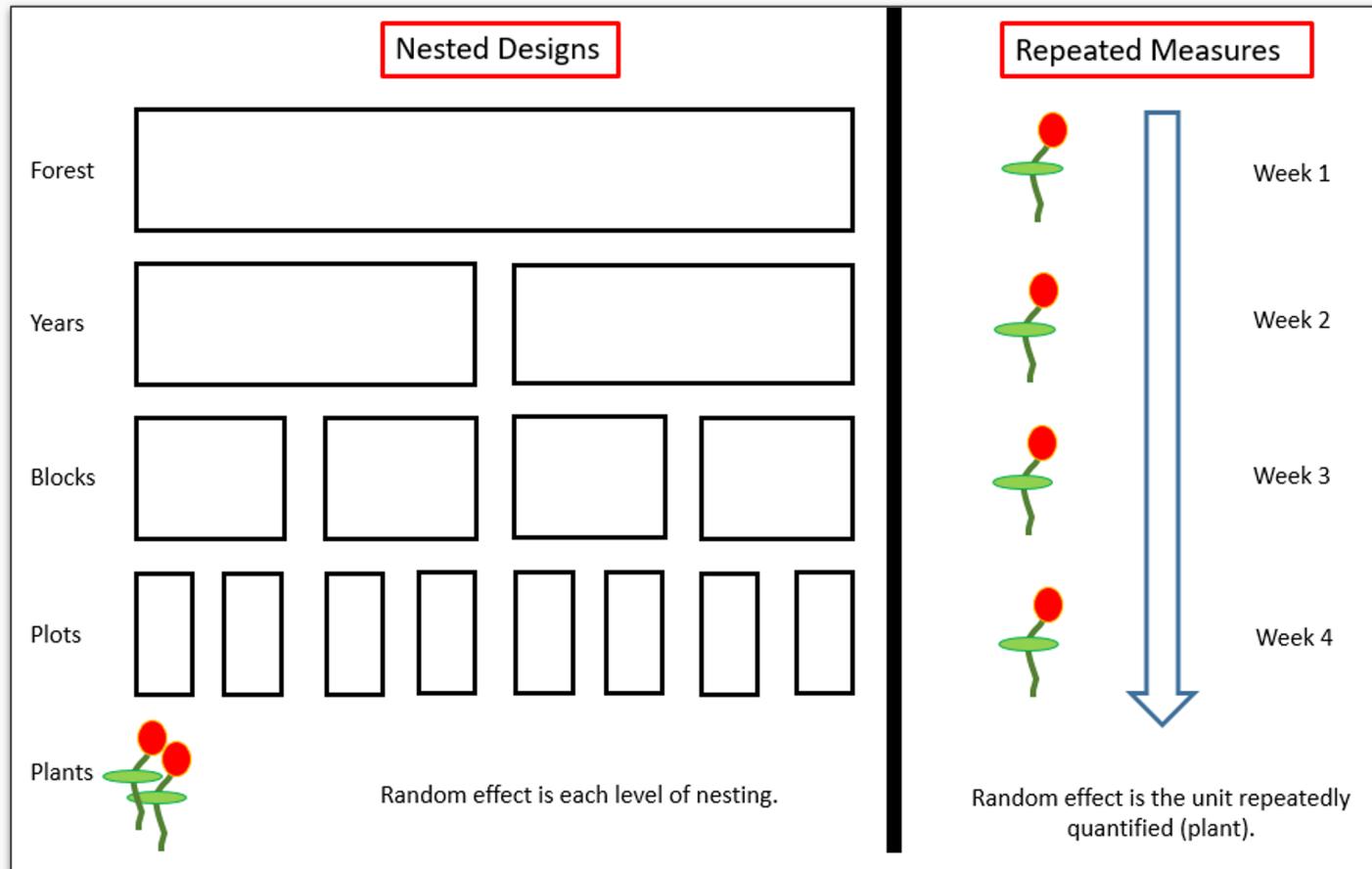
- What is a linear mixed effects model? Why use it?
 - What are benefits and limitations of this method compared to other regressions?
 - What are the assumptions of this method?
- What are fixed vs. random effects?
 - How do you decide which variables should be fixed vs. random?
 - How to incorporate repeated measures?
 - How do you decide which type of random effects to include?
 - How do you decide which random effects structure to include in the final model?
- How do you evaluate model assumptions?
- How do you interpret the summary output?
 - What's the difference between the ANOVA table and summary() output?
 - How can we do post-hoc tests with this method?

What is a linear mixed effects model? Why use it?

A regression analysis that uses fixed effects and ***random effects***

- Used to analyze nested and clustered data

Design Structure: Nested and Repeated Measures



What is a linear mixed effects model? Why use it?

Type in the chat box:

Think about your own data. Do you think there is inherent clustering, hierarchy, or repeated measures in your data?

Benefits and Assumptions

Benefits

- Using nested / clustered data, improves accuracy of your standard errors
 - If you do not account for nested data structure (repeated measures) = pseudoreplication
- Can be used with other distributions (GLMs, GAMs)
- Does not calculate sum of squares, but uses likelihood estimation
 - Makes meeting assumptions easier, good for unbalanced designs
 - Easier to specify than complex ANOVA designs...

Assumptions

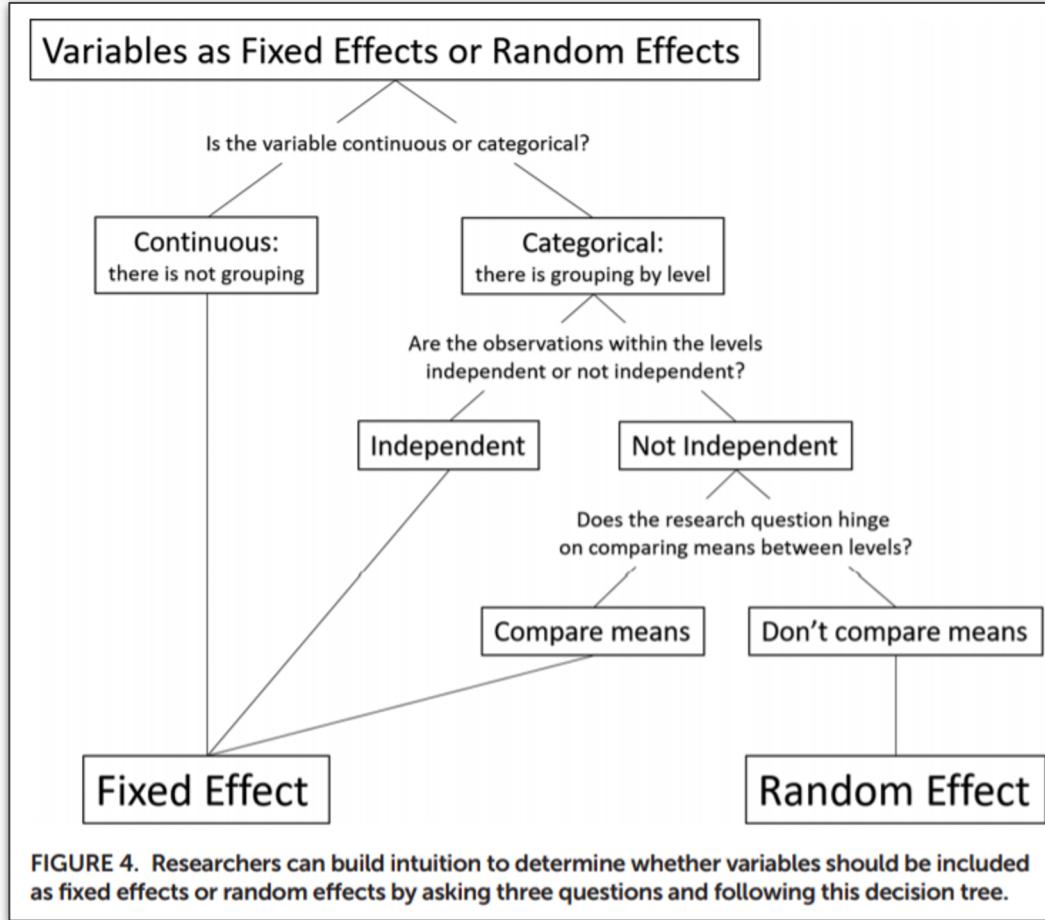
- Normality of residual error and constant variance (homogeneity of variance)

What are fixed and random effects?

Sorting Fixed and Random Effects:

You'll be collaborating in breakout rooms for ~5 minutes. Have one person share their screen and sort the different model variables into **fixed** and **random** effects.

What are fixed and random effects?



Incorporating Repeated Measures

Include the **experimental unit** being repeatedly measured as a random effect.

Examples:

In a **plot**, you measure percent cover monthly.

On a **fish**, you measure heart after several trials.

**Thus, it is important to code your data so each exp. unit has a unique identifying number.*

***Also, for LMM if you have multiple measurements, do not average the data include all observations.*

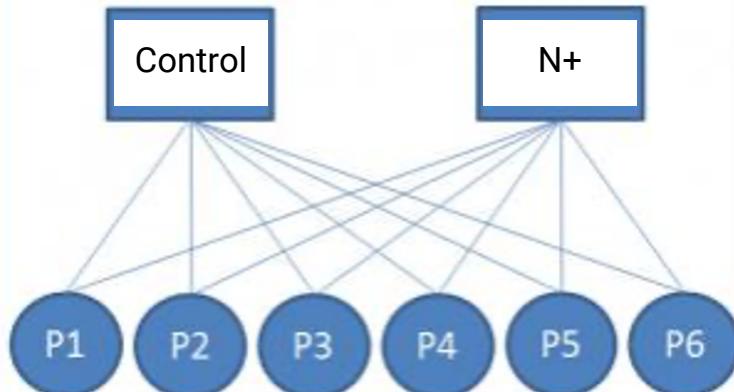
Treatment	plot_ID
A	1
A	2
B	3
B	4
A	5
A	6
B	7
B	8
A	9
A	10
B	11

Crossed vs Nested Random Effects

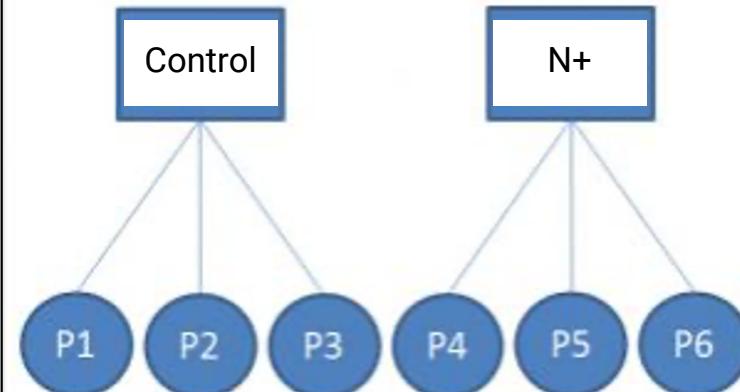
Crossed- all treatments occur in combination with each other.

Nested- certain individuals only occur in specific treatments.

Crossed example



Nested example



Crossed vs Nested Random Effects

Nested- certain female birds only occur in one woodland and not others.
Female is nested with Woodland.

$$\text{ClutchMass} \sim \text{Foraging Rate} + (1|\text{Woodland}/\text{Female ID})$$

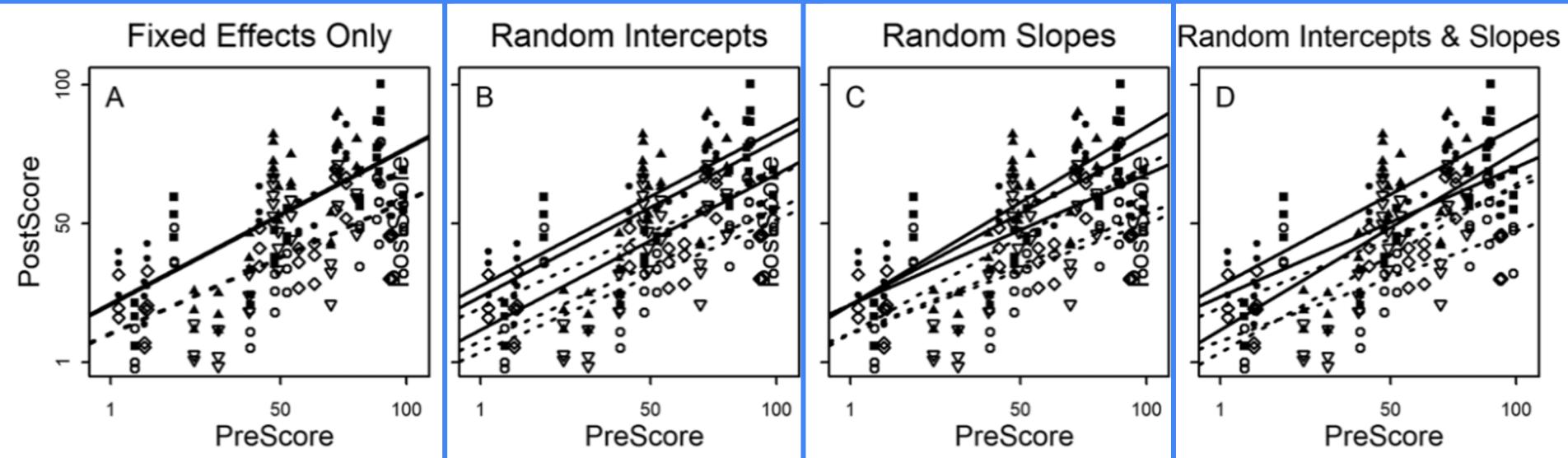
Crossed (Year)- the researcher takes data yearly (all woodlands and all females occur in each year, so year is crossed).

$$\text{ClutchMass} \sim \text{Foraging Rate} + (1|\text{Woodland}/\text{Female ID}) + (1|\text{Year})$$

*HOWEVER, if you code your data so all exp. units and blocks have a unique identifying number.... Then, the function “automatically sees” the nested structure, and $(1|\text{Block}/\text{Plot})$ is equivalent to $(1|\text{Block}) + (1|\text{Plot})$.

lme4 notation

How do you decide which random effects structure to include in the final model?



Post ~ Pre

Regression line for control
and treatments, “blocking”
factor is ignored

Control = dashed line
Treatment = solid line

Post ~ Pre + (1|Block)

Slope is the same for
all groups, intercepts
vary for each group

Post ~ Pre + (0 + Pre|Block)

Slopes vary for each
group, intercept is the
same for all groups

0 fixes the intercept

Groups have different
slopes and different
intercepts

Use : to separate random
effects

Theobald (2018)

How do you decide which random effects structure to include in the final model?

Random effects

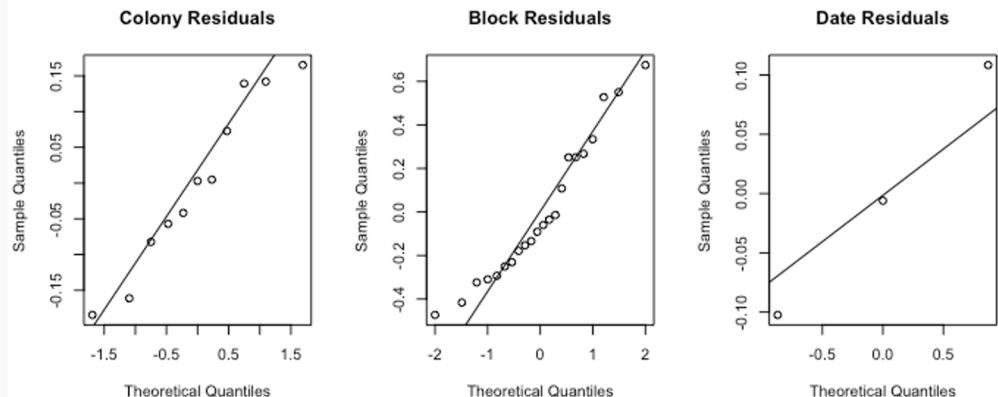
```
log.full.model = lmer(log.mortality ~ Snail * Nutrient * Sediment + Shell.length + (1|Colony) + (1|Block) + (1|Date), data = mortality, REML = TRUE)  
log.colony.model = lmer(log.mortality ~ Snail * Nutrient * Sediment + Shell.length + (1|Colony) + (1|Date), data = mortality, REML = TRUE)  
log.block.model = lmer(log.mortality ~ Snail * Nutrient * Sediment + Shell.length + (1|Block) + (1|Date), data = mortality, REML = TRUE)
```

Step 1: AICc for small sample sizes

```
> AICc(log.full.model, log.colony.model, log.block.model)  
df    AICc  
log.full.model 13 433.4458  
log.colony.model 12 474.4741  
log.block.model 12 431.6206
```

****NOTE:** if you want to compare models with **different fixed effects** using AIC you must specify **REML = FALSE**.

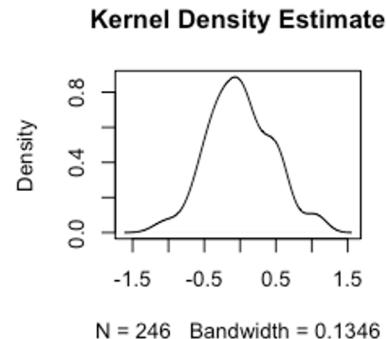
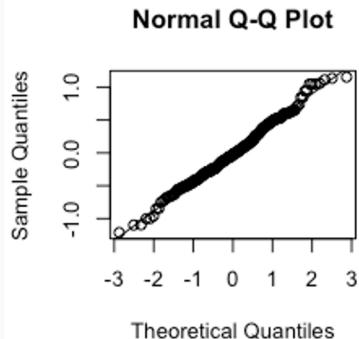
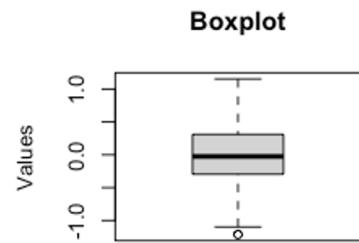
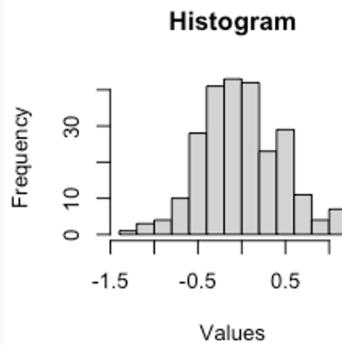
Step 2: Assess random effects residuals for normality



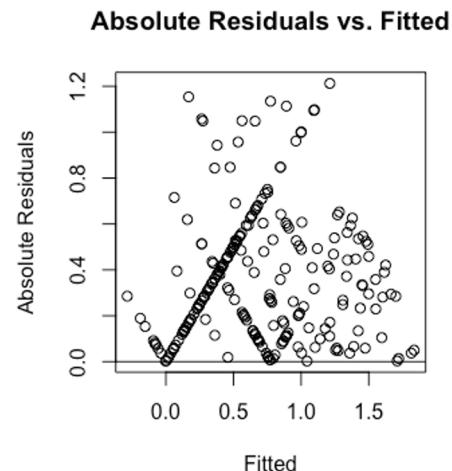
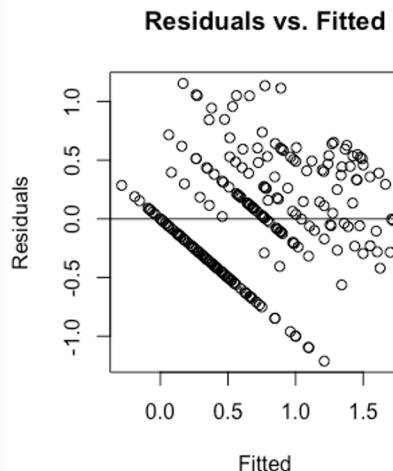
```
qqnorm(ranef(log.full.model)$Colony[,1], main = "Colony Residuals")  
qqline(ranef(log.full.model)$Colony[,1])
```

How do you evaluate model assumptions?

Assessing residuals for normality



Assessing residuals for constant variance



How do you interpret model output?

```
anova(log.block.model, type=3, ddf = "Kenward-Roger")
```

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Snail	0.62425	0.62425	1	231.143	2.7320	0.099716 .
Nutrient	0.23609	0.23609	1	20.063	1.0332	0.321504
Sediment	0.04065	0.04065	1	217.119	0.1779	0.673614
Shell.length	0.31639	0.31639	1	231.184	1.3846	0.240522
Snail:Nutrient	1.67648	1.67648	1	216.966	7.3369	0.007294 **
Snail:Sediment	1.03514	1.03514	1	217.888	4.5302	0.034425 *
Nutrient:Sediment	0.86735	0.86735	1	216.780	3.7959	0.052670 .
Snail:Nutrient:Sediment	0.07779	0.07779	1	216.969	0.3405	0.560169

Signif. codes:	0	***	0.001	**	0.01	**
					0.05	.'
					0.1	'
					1	

How do you interpret model output?

```
summary(log.block.model)
```

```
> summary(log.block.model)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: log.mortality ~ Snail * Nutrient * Sediment + Shell.length +
          (1 | Block) + (1 | Date)
Data: mortality

REML criterion at convergence: 406.3

Scaled residuals:
    Min      1Q  Median      3Q     Max 
-2.53610 -0.61604 -0.05252  0.64432  2.41371 

Random effects:
 Groups   Name        Variance Std.Dev. 
Block    (Intercept) 0.19669  0.4435 
Date    (Intercept) 0.01344  0.1159 
Residual           0.22850  0.4780 
Number of obs: 246, groups: Block, 22; Date, 3

Fixed effects:
(Intercept)               0.52348  0.17113 25.08134  3.059  0.00523 ** 
SnailSnail Predation       0.78217  0.48146 230.51480  1.625  0.10562 
NutrientNot Enriched      -0.05415  0.22476 32.46112 -0.241  0.81114 
SedimentSediment Addition  0.07792  0.11768 215.18795  0.662  0.50857 
Shell.length                -0.03270  0.02764 231.22142 -1.183  0.23812 
SnailSnail Predation:NutrientNot Enriched  0.27049  0.18213 217.93445  1.485  0.13894 
SnailSnail Predation:SedimentSediment Addition -0.34523  0.18437 218.23584 -1.872  0.06248 . 
NutrientNot Enriched:SedimentSediment Addition  0.17016  0.16913 215.62503  1.006  0.31550 
SnailSnail Predation:NutrientNot Enriched:SedimentSediment Addition  0.14576  0.24968 217.13493  0.584  0.55998 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
(Intr) SnlSnP NtrrNE SdmnSA Shll.l SSP:NE SSP:SA NNE:SA
SnlSnlPrdt -0.084
NtrntNtEnrc -0.645  0.049
SdmntSdmntA -0.344  0.122  0.262
Shell.Lngth  0.000 -0.962  0.015  0.000
SnlSPrd:NNE  0.222 -0.014 -0.363 -0.323 -0.189
SnlSPrd:SSA  0.219 -0.308 -0.165 -0.638  0.116  0.498
NtrntNE:SSA  0.239 -0.066 -0.388 -0.696 -0.020  0.483  0.442
SSP:NNE:SSA -0.162  0.090  0.264  0.471  0.058 -0.719 -0.722 -0.678
```

How can we do post-hoc tests with this method?

Multiple comparisons

```
emmeans(log.block.model, list(pairwise ~ Snail * Nutrient), adjust = "tukey")
```

```
> emmeans(log.block.model, list(pairwise ~ Snail * Nutrient), adjust = "tukey")
NOTE: Results may be misleading due to involvement in interactions
```

```
$`emmeans of Snail, Nutrient`  
  Snail      Nutrient    emmean   SE df lower.CL upper.CL  
  No Snail Predation Enriched    0.300 0.275 113  -0.396   0.996  
  Snail Predation Enriched     0.910 0.284 122   0.191   1.629  
  No Snail Predation Not Enriched 0.331 0.277 114  -0.369   1.031  
  Snail Predation Not Enriched  1.284 0.304 140   0.518   2.050
```

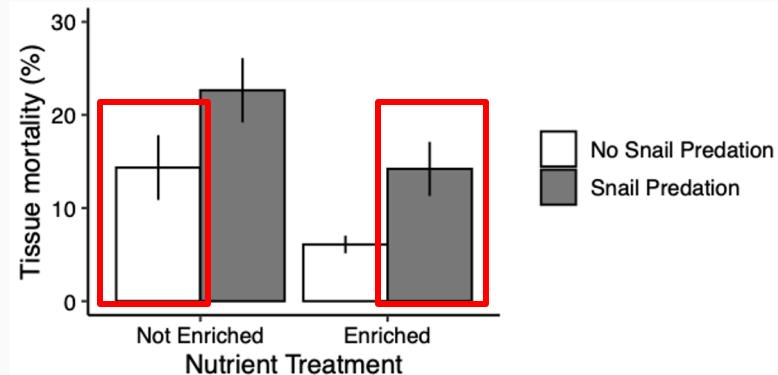
Results are averaged over the levels of: Sediment

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

Conf-level adjustment: sidak method for 4 estimates

```
$`pairwise differences of Snail, Nutrient`  
  contrast                      estimate   SE   df t.ratio p.value  
  No Snail Predation Enriched - Snail Predation Enriched -0.6096 0.464 230.9 -1.314  0.5547  
  No Snail Predation Enriched - No Snail Predation Not Enriched -0.0309 0.207 23.4 -0.149  0.9988  
  No Snail Predation Enriched - Snail Predation Not Enriched -0.9839 0.523 207.0 -1.881  0.2394  
  Snail Predation Enriched - No Snail Predation Not Enriched  0.5786 0.503 200.9  1.151  0.6583  
  Snail Predation Enriched - Snail Predation Not Enriched -0.3743 0.211 25.2 -1.773  0.3094  
  No Snail Predation Not Enriched - Snail Predation Not Enriched -0.9529 0.490 231.1 -1.946  0.2119
```



How can we do post-hoc tests with this method?

Within group comparisons

```
emmeans(log.block.model, pairwise ~ Snail | Nutrient)
```

```
> emmeans(log.block.model, pairwise ~ Snail | Nutrient)  
NOTE: Results may be misleading due to involvement in interactions
```

```
$emmeans  
Nutrient = Enriched:  
  Snail          emmean    SE  df lower.CL upper.CL  
No Snail Predation  0.300 0.275 113  -0.244   0.845  
Snail Predation     0.910 0.284 122   0.347   1.473
```

```
Nutrient = Not Enriched:  
  Snail          emmean    SE  df lower.CL upper.CL  
No Snail Predation  0.331 0.277 114  -0.217   0.879  
Snail Predation     1.284 0.304 140   0.684   1.884
```

```
Results are averaged over the levels of: Sediment  
Degrees-of-freedom method: kenward-roger  
Confidence level used: 0.95
```

```
$contrasts  
Nutrient = Enriched:  
  contrast           estimate    SE  df t.ratio p.value  
No Snail Predation - Snail Predation  -0.610 0.464 231 -1.314  0.1900
```

```
Nutrient = Not Enriched:  
  contrast           estimate    SE  df t.ratio p.value  
No Snail Predation - Snail Predation  -0.953 0.490 231 -1.946  0.0529
```

```
Results are averaged over the levels of: Sediment  
Degrees-of-freedom method: kenward-roger
```

