JEST @ 2022 Jornada Ecology Short Course

Introduction to ANOVA-type Linear Mixed Models in R

Wednesday June 29, 2022 1:00-2:30pm
Instructors: Darren James and Greg Maurer

https://jornada-im.github.io/JEST/workshops/20220629-jrn-ecology-short-course/index.html

Goals and objectives

- Review of basic hypothesis testing
  - Adopting a balanced view of p-values
- Fitting ANOVA Linear Models (LMs) in R
- Fitting ANOVA-type Linear Mixed Models (LMMs) in R

Please provide feedback: this is a practice run

- Insufficient time to cover our prepared content
- We intend to offer future workshops on topics of interest

Types of Statistical Analysis Objectives *(in approximate order of increasing difficulty)*

Descriptive – describe or summarize a set of data

Exploratory – pattern detection; find relationships not previously known

Inferential – use a relatively small sample of data to say something about the population at large

Predictive – use current and historical data to make predictions about future data

Causal – see what happens to one variable when we manipulate another variable

Mechanistic – understand the exact changes in variables that lead to exact changes in other variables

Credit: The Data Scientist's Toolbox, Johns Hopkins University, coursera.org

Hypothesis testing, reviewed

General goal: Assess "significant" differences by ruling out chance/sampling error as a plausible explanation

Null Hypothesis (H_0): Hypothesis of no/uninteresting statistical relationship
Alternative Hypothesis (H_A): Hypothesis of one or more relationship(s) of interest

Some examples of null and alternative hypotheses

Global F test for a fixed effect in ANOVA

H_0: all levels of the effect have equal means

H_A: at least one of the levels has a different mean than at least one other level

Shapiro-Wilk test for normality

H_0: data follow a normal distribution

H_A: data do not follow a normal distribution

Hypothesis testing: p-values

- Software returns a ***p-value*** for us to interpret

Behind the scenes:

After defining H_0 and H_A :

1. Identify a test statistic whose distribution under H_0 is known (*t, f, X* ^{2}, etc.)
2. Calculate the test statistic for the data
3. Compare the test statistic to its distribution under H_0
   - Calculate the probability of observing a test statistic more extreme: **p-value**

- If the p-value is sufficiently low (usually 5% or less), reject H_0 in favor of H_A
- If we fail to reject H_0: proceed as if null hypothesis is true
  - *But we haven't actually proven H_0 true; we just acknowledge that our data are consistent with H_0*

The p-value has a nuanced and clumsy definition that is easily misunderstood.

- **A p-value is the conditional probability of observing a statistic as extreme as or more extreme than the one computed from the current data, across hypothetical repetitions of the experiment.**
  - A p-value is not the probability of H_0 being true
  - A p-value is not the probability of falsely rejecting the null hypothesis (i.e. the probability of a Type I error)

Alternative definition: **a p-value is the probability of the data given that it was generated under the null hypothesis (H_0).**

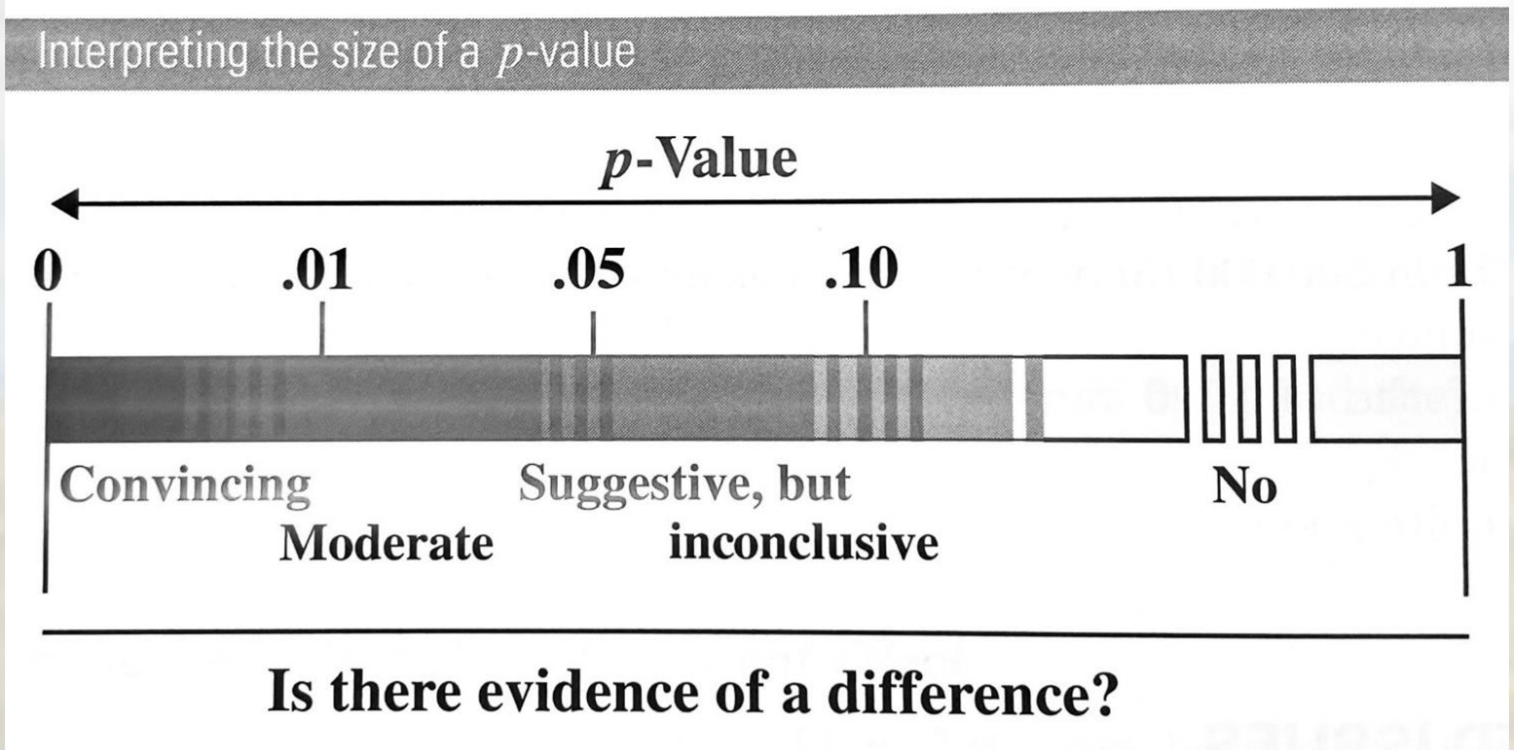
Toward a more nuanced view of p-values

Point #1: it's impossible to interpret a p-value without knowing the null hypothesis

- *Always make sure you have a clear understanding of the null hypothesis*

Point #2: p-values are conditional probabilities: they are the probability of an event occurring, given that other events have already occurred.

- *Conditioned on: the null hypothesis being true; the model; the model's assumptions; the sample size; the experimental design; the sampling methods, the researcher using the software correctly, etc.)*



Interpreting the size of a p-value
p-Value
0
.01
.05
.10
1
Convincing
Moderate
Suggestive, but inconclusive
No
Is there evidence of a difference?

Point #3: there is no "magic" associated with p ≤ 0.05.

We often treat p-values as binary (significant/non-significant), but their interpretation is better treated as "continuous" when p is about 0.10 or less.

*Credit: Ramsey, Fred L. and Daniel W. Schafer. 2013. **The Statistical Sleuth: A Course in Methods of Data Analysis, Third Edition**. Boston, MA: Brooks/Cole. Display 2.12 p.48*

First R exercise: Analysis of Variance (ANOVA)

In ANOVA, the outcome of hypothesis testing for differences in means depends on:

1. **n: how many samples per group**
2. **variance in each population: σ^2**
3. **effect size: how big is the difference between means?**

The three major assumptions of Analysis of Variance (ANOVA):

1. **Errors are normal**
   - This is the *least important* assumption; ANOVA is often robust to this
   - Assess normality of model residuals, not the raw data
2. **Equal variance in all groups**
   - *Second least important* assumption; some degrees of robustness to violations
3. **Independent observations**
   - Most important assumption

R Exercise: Use 2017 total estimated ANPP from 15 NPP sites

- 3 sites from each of 5 different vegetation zones
- Question: which vegetation zones are different from each other?

From Linear Models (LMs) to Linear Mixed Models (LMMs)

The three major assumptions of Analysis of Variance (ANOVA):

1. **Errors are normal**
2. **Equal variance in all groups**
3. **Independent observations**

Mixed models can handle data that violate #2 and #3:

- Heterogeneous variance (heteroskedascity)
- Correlations between observations (i.e. multi-level or hierarchical structure)

Mixed models accomplish this with **random effects**

Fixed Effects vs. Random Effects

Traditional definititions:

Fixed: The researcher(s) who who planned the experiment decided which levels to use.

Random: Each level can be regarded as a sample from a population of levels.

Key idea: the influence of the random factors are incorporated into the variance of the fixed factors.

- Fixed factors will have the same means as they would in a Linear Model
- But they will likely have higher standard errors because their variance estimates include the variance component(s) from the random effect(s) in addition to σ^2.

How do we choose between fixed and random?

- **Inference** considerations
- **Structural** considerations
- **Practical** considerations

Fixed Effects vs. Random Effects

- Inference considerations

Fixed: Inference is confined to the levels in the experiment

Random: Inference can be applied to levels not measured in the experiment

Example: study with multiple sites selected in southern NM

- Formulate site as a **fixed** effect: inference is limited to these sites only
- Formulate site as **random** effect: can be basis to apply inference to similar sites in a larger region such as the northern Chihuahuan Desert

- Structural considerations

Subsamples from the same plot (experimental unit) at the same time

➢ Must formulate plot as random effect to avoid pseudo-replication

Subsamples from the same plot at different times (repeated measures)

➢ Solution: formulate time as a random effect

- Practical considerations

Random effects don’t appear in ANOVA tables

Random effects don’t consume degrees of freedom

➢ In practice there can be some flexibility in assigning factors as fixed or random

MIXED EFFECT MODEL
HIERARCHICAL LINEAR MODEL
RANDOM EFFECTS MODELS
VARIANCE COMPONENT MODELS
MIXED MODELS
RANDOM INTERCEPTS/ SLOPES
REGULARIZED REGRESSION
MULTI LEVEL MODEL
NESTED DATA MODEL
RANDOM PARAMETER MODEL
imgflip.com

https://twitter.com/ChedidFabienne/status/1466163176718974979/photo/1

The unfortunate naming conventions of linear mixed models

“Mixed” → fixed effects combined with random effects

Always true: mixed models have at least one variance parameter in addition to the usual σ^2.

Linear Model (LM): only σ^2

Linear Mixed Model (LMM): σ^2 plus at least one more variance estimate.

Linear model

observation
(data point)
= group mean + statistical noise

Means model $$y = \mu_i + e$$

$i = \{\text{group1}, \text{group2}, \text{group3}\}$

overall mean + group effect

Effects model: $$y = \mu + \tau_i + e$$

overparamaterized:
4 parameters but only 3 groups.
Side condition: $\sum \tau_i = 0$*)*

In matrix form: $$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

What do these matrices look like?

Here's an example:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\overset{\boldsymbol{y}}{\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix}} = \overset{\boldsymbol{X}}{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}} \overset{\boldsymbol{\beta}}{\begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}} + \overset{\boldsymbol{\varepsilon}}{\begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \\ e_{21} \\ e_{22} \\ e_{23} \end{bmatrix}}$$

with data:

$$\overset{\boldsymbol{y}}{\begin{bmatrix} 2 \\ 0 \\ 1 \\ 4 \\ 6 \\ 8 \\ 9 \\ 10 \\ 5 \end{bmatrix}} = \overset{\boldsymbol{X}}{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}} \overset{\widehat{\boldsymbol{\beta}}}{\begin{bmatrix} 5 \\ -4 \\ 1 \\ 3 \end{bmatrix}} + \overset{\boldsymbol{\varepsilon}}{\begin{bmatrix} 1 \\ -1 \\ 0 \\ -2 \\ 0 \\ 2 \\ 1 \\ 2 \\ -3 \end{bmatrix}}$$

The Normal Equations:

$$\boldsymbol{X'X\beta} = \boldsymbol{X'y}$$

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{X'(X'X)^{-1}X'y}$$

Minimizes $\boldsymbol{\varepsilon}$

Incorporating random effects #1: the structure of R

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{R} = var\{\boldsymbol{\varepsilon}\}$$

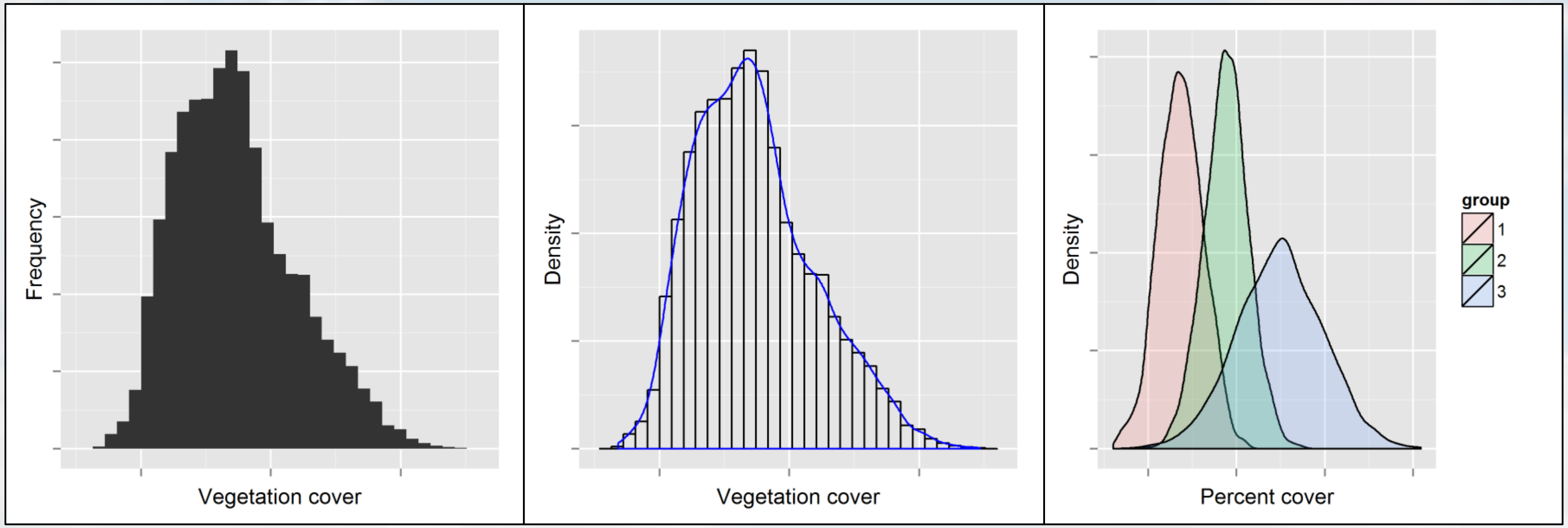
$\boldsymbol{R}$ is a square matrix that describes the variance of the errors according to the grouping structure you specify.

$\boldsymbol{R}$ is essentially a table that shows how the groups **vary with themselves (variance)** and **vary with each other (covariance)**

	group 1	**group 2**	**group 3**
group 1	🟨	🟦	🟦
group 2	🟦	🟨	🟦
group 3	🟦	🟦	🟨

Repeated measures models commonly use time as the grouping structure

Heterogeneous variances



$$\boldsymbol{R} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_2^2 \end{bmatrix}$$

In linear models, ***R*** looks like this:

- Equal variance in all groups
- No correlation between groups

$$\boldsymbol{R} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_1^2 \end{bmatrix}$$

The variance of groups 1 and 2 is the same (σ_1^2).

The variance of group 3 is different (σ_2^2).

In this case we're not modeling correlations: the off-diagonal values are 0.

The R matrix in 2 parts: variance and correlation

When fitting Repeated measures LMMs, we are often faced with heterogenous variance in addition to correlation between times

In the **nlme:gls()** function, the variance and correlation are spread across 2 arguments

Variance: weights = argument
weights=varIdent(form=~1|time)

	time1	time2	time3
time1	σ_1^2		
time2		σ_2^2	
time3			σ_3^2

Correlation: correlation = argument
correlation = corSymm(form = ~ 1 | EU)

	time1	time2	time3
time1	1	ρ 12	ρ 13
time2	ρ 12	1	ρ 23
time3	ρ 13	ρ 13	1

These two arguments produce an “Unstructured” covariance structure

- most general structure possible
- useful starting point for selecting covariance structures
- most possible parameters
- not not always estimable; depends on data

Many other correlation structures are available in the nlme package **?nlme::corClasses**

First-order autoregressive
correlation = corAR1 (form = ~ 1 | EU)

	time1	time2	time3
time1	1	ρ	ρ^2
time2	ρ	1	ρ
time3	ρ^2	ρ	1

- First order autoregressive: observations farther apart in time have lower correlation
- This structure is not appropriate for unequally spaced sampling times.

Compound symmetry
correlation = corCompSymm(form = ~ 1 | EU)

	time1	time2	time3
time1	1	ρ	ρ
time2	ρ	1	ρ
time3	ρ	ρ	1

- Compound symmetry: constant correlation across times

Using fit statistics to choose the best covariance structure

Akaike Information Criterion (AIC) : a fit statistic that measures “information loss” between the model and the data

AIC = - 2*log likelihood + 2*(#parameters)

Lesser values closer indicate better fit and greater parsimony.

→ Model with **lowest AIC** has “best” fit

→ Sometimes AIC is negative; “best” model has most negative AIC (not closest to 0)

Other fit statistics:

AICC

BIC

Linear Mixed Models

There are ***2 general ways*** to incorporate random effects into a linear model:

1. **Embed the random effects into the structure of the errors**

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad var\{\boldsymbol{\varepsilon}\} = \mathbf{R}$$

This involves structuring $\boldsymbol{R}$, the variance matrix of $\boldsymbol{\varepsilon}$

- This is called ***<u>R-side</u>*** modelling
- These models are called ***R-side*** or ***correlated errors*** models

2. **Model the random effects $\boldsymbol{b}$ explicitly**

$$\boldsymbol{y}|\boldsymbol{b} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon}$$

Conditional formulation: **y**, given the random effects $\boldsymbol{b}$

Design matrix for random effects

Solution for fixed effects

Linear Mixed Models

1. Embed the random effects into the structure of the errors

   (***R-side*** modeling)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad var\{\boldsymbol{\varepsilon}\} = \mathbf{R}$$

2. Model the random effects ***b*** explicitly

$$\boldsymbol{y}|\boldsymbol{b} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon} \qquad var\{\mathbf{Z}\} = \mathbf{G}$$

This involves specifying the random effects ***b***, their design matrix ***Z***, and the structure of ***G***, the variance matrix of ***Z***

- This is called ***G-side*** modelling
- These models are called ***G-side*** models

R-side vs. G-side models

R-side: Unmeasured sources of variation

Population-wide inferences

Easier for software to profile

SAS Proc Mixed: **repeated** statement

nlme::gls()

G-side: Models variation directly

Subject-specific inferences in addition to population-wide

SAS Proc Mixed: **random** statement

nlme::lme(); **lme4::lmer()**

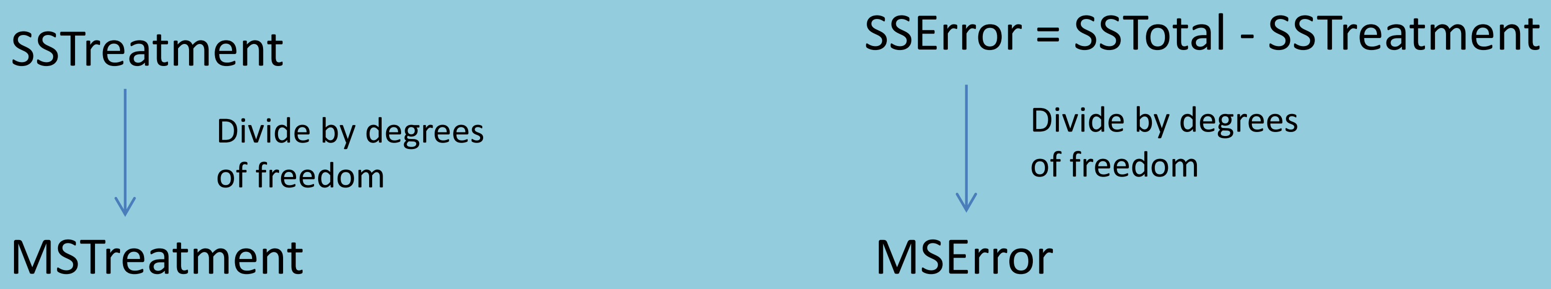
- In normal models (LMMs), the R-side vs. G-side difference has little consequence for inference
  - More of a technicality
  - But necessary to help understand model specification in software
  - Both the conditional and marginal distributions are normal
    - Closed under linear transformations

- **When we use non-normal distributions this is not the case!**

ANOVA: estimate variance using Sums of Squares

SSTotal: Sums of Squares under one overall mean

SSTreatment: Sums of Squares under individual group means



SSTreatment → Divide by degrees of freedom → MSTreatment

SSError = SSTotal - SSTreatment → Divide by degrees of freedom → MSError

$$\frac{variance\ between\ treatments}{variance\ within\ treatments} = \frac{Mean\ Square\ Treatment}{Mean\ Square\ \ Error} = F_{num,\, denom} \rightarrow \textit{p-value}$$

Mixed Models: estimate variance components using more complicated numerical procedures

- Most common: **Restricted Maximum Likelihood (REML)**
  - Iterative process that does not always converge
  - Involves a transformation of the data

Sums of Squares and F-tests

- The are several types of sums of squares
- **Darren recommends using F-tests derived from Type III Sums of Squares**
- In many R functions, the default is to use Type I

Type I Tests shows the additional effect of each variable in the model, so it changes depending on the order of the factors.

Type III Tests (aka Partial Sums of Squares) looks at the incremental effect of each term in the model after the other effects have been accounted for.

  - Order of the factors in the model is not important.
  - Type III are especially important with unbalanced data.
  - Appropriate for most use cases

The linear mixed model equations

$$\begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{R'R^{-1}Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix}$$

Use REML to obtain solutions for $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{b}$

Problem: variance estimates of $\widehat{\boldsymbol{\beta}}$ are biased downward because they are estimated after accounting for the random effects

- Standard practice: Some type of adjustment must be implemented to avoid biased variance estimates

- Gold standard: **Kenward and Roger adjustment**
  - (available for some models with emmeans package)

Models used for data typically encountered in ecological research

Response Variable Type	Commonly used probability distributions	Explanatory variables and error structure			
		fixed effects		random effects	
		categorical	continuous	conditional (normally distributed)	correlated errors
Continuous Symmetric	Normal (Gaussian)	Linear Models ANOVA ANCOVA	Regression	Linear Mixed Models *G-side*	*R-side*
Categorical	Bernoulli, Binomial, Multinomial	Logistic Regression Generalized Linear Models		Generalized Linear Mixed Models	
Counts	Poisson, Negative Binomial				
Continuous Proportion	Beta				
Time to Event	Exponential, Gamma				

Common R packages for fitting LMs, GLMs, LMMs, GLMMS

(not exhaustive)

Linear Models: ANOVA, regression, Analysis of Covariance (ANCOVA)

- **stats::lm()**
- stats:aov() may provide a more convenient interface for ANOVA

Generalized Linear Models: logistic regression, Poisson regression, etc.

- **stats::glm()**

Linear Mixed Models:

- **nlme::gls()**
- **nlme::lme()**
- **lme4::lmer()**
- many others

Generalized Linear Mixed Models:

- **lme4::glmer()**
- others