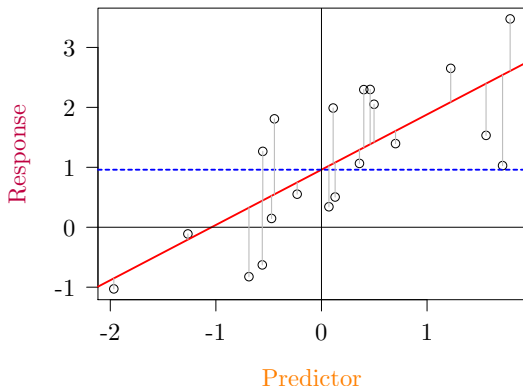# Linear mixed models
Why, what, how?

Timothée Bonnet with content from Terry Neeman
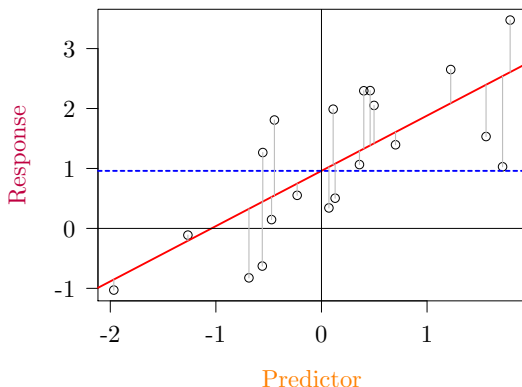
January 17, 2019

Statistical models: MEAN and VARIANCE components

Predictor

Response = **Intercept** + **Slope** × **Predictor**
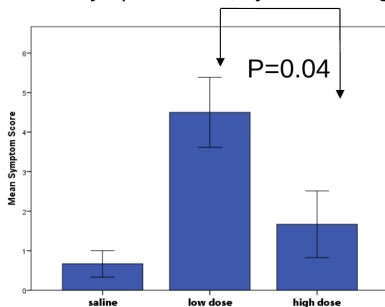
Statistical models: MEAN and VARIANCE components

$$\underbrace{Response = \underbrace{Intercept + Slope \times Predictor}_{\substack{\text{Mean Structure} \\ \text{Experimental factors}}} + \underbrace{Error}_{\substack{\text{Variance Structure} \\ \text{Unrelated to experiment factors} \\ \text{Unexplained "noise"}}}, \text{ with } \epsilon \sim N(0, \sigma)}$$

What is in ϵ? How can we tweak that? Why should we care?

Let's do exercises in section 1

Exercise 1:

Mean symptom score by treatment group

Vaccine challenge experiment:

- 6 mice/group (saline/low dose/high dose)
- All mice challenged with Shigella
- Followed for 14 days
- Outcome: Symptom score average Days 2 - 8

One-way ANOVA (post-hoc Bonferroni) p=0.04

Noise confounded with treatment

Experimental design

The observed difference in outcome could be the result of:

- Cage effects
- Mouse strain effects

These effects are CONFOUNDED with treatment effect

Cage 2:
Low Dose

Cage 1:
saline

Cage 3:
High Dose

Noise confounded with treatment

Experimental design

The observed difference in outcome could be the result of:

- Cage effects
- Mouse strain effects

These effects are CONFOUNDED with treatment effect

Cage 2: Low Dose

Cage 1: saline

Cage 3: High Dose

Solutions:

Mixed cages: can compare within cages

More cages: must compare between cages

Noise confounded with treatment

Mixed cages: can compare within cages

- **Share the noise among treatments**
- Few cages needed: Technically efficient
- But may be technically impossible

Noise confounded with treatment

Mixed cages: can compare within cages

- **Share the noise among treatments**
- Few cages needed: Technically efficient
- But may be technically impossible

More cages: must compare between cages

- **Redefine experimental unit**
- Noise among cages, instead of within
- Needs to re-scale the experiment

Exercise 2:

Week 1   Week 2   Week 3   Week 4   Week 5   Week 6

Wild-type   Mutant 1   Mutant 2   Mutant 3   Mutant 4   Mutant 5

10 eggs/tube   10 eggs/tube   10 eggs/tube   10 eggs/tube   10 eggs/tube   10 eggs/tube

x 6   x 6   x 6   x 6   x 6   x 6

What is wrong with this design?

What is wrong with this design?

- CONTROLS: not tested under identical conditions
- REPLICATION: only pseudo-replication
- BLOCKING: none
- RANDOMISATION: NA

What is wrong with this design?

- CONTROLS: not tested under identical conditions
- REPLICATION: only pseudo-replication
- BLOCKING: none
- RANDOMISATION: NA

This experiment is useless

What is going on conceptually?

$$\textit{Response} = \underbrace{\textit{Intercept} + \textit{Slope} \times \textit{Predictor}}_{\substack{\text{Mean Structure} \\ \text{Experimental factors}}} + \underbrace{\textit{Error}, \text{ with } \epsilon \sim N(0, \sigma)}_{\substack{\text{Variance Structure} \\ \text{Unrelated to experiment factors} \\ \text{Unexplained "noise"}}}$$

What is going on conceptually?

$$Response = \underbrace{Intercept + Slope \times Predictor}_{\substack{\text{Mean Structure} \\ \text{Experimental factors}}} + \underbrace{Error, \text{ with } \epsilon \sim N(0, \sigma)}_{\substack{\text{Variance Structure} \\ \text{Unrelated to experiment factors} \\ \text{Unexplained "noise"}}}$$

For robust models we need assumptions about the error:

① Gaussian error distribution

② Homoscedasticity (constant error variance)

③ **Independence of errors**

That what $\epsilon \sim N(0, \sigma)$ means

What is going on conceptually?

$$\text{Response} = \underbrace{\textit{Intercept} + \textit{Slope} \times \textit{Predictor}}_{\substack{\text{Mean Structure} \\ \text{Experimental factors}}} + \underbrace{\textit{Error}, \text{ with } \epsilon \sim N(0, \sigma)}_{\substack{\text{Variance Structure} \\ \text{Unrelated to experiment factors} \\ \text{Unexplained "noise"}}}$$

For robust models we need assumptions about the error:

1. Gaussian error distribution

2. Homoscedasticity (constant error variance)

3. **Independence of errors**

That what $\epsilon \sim N(0, \sigma)$ means

In mice and frog experiments, ϵ are non-independent
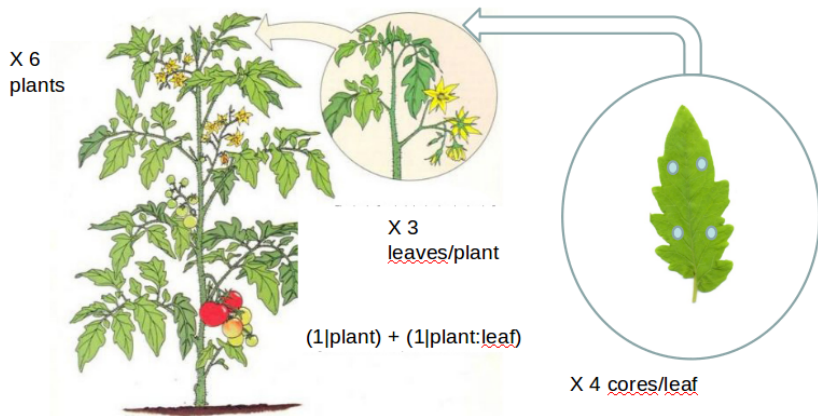
Fixed or random effect?

In general

- Doesn't change inference much. Random effect slightly more efficient.
- Summary cleaner with random effect, especially when many random levels
- Random shifts the focus from level values to variation among levels
- Variance parameters interesting in themselves
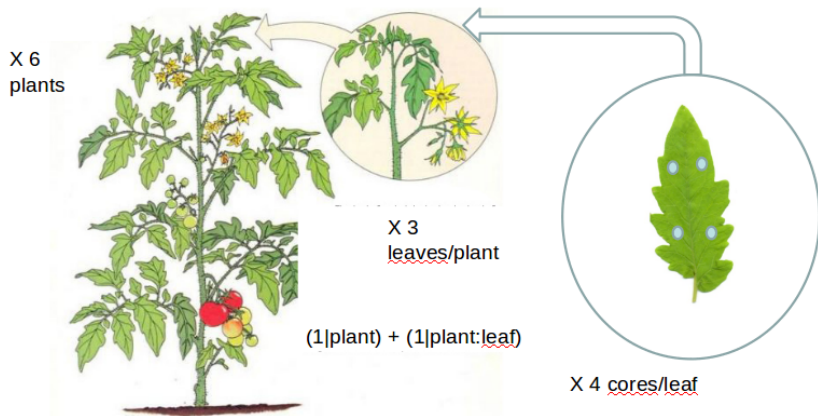- Are levels of interest (fixed) or are they some kind of noise (random)

Exercises with lme4 output

Understanding different variance structure

X 6 plants

X 3 leaves/plant

(1|plant) + (1|plant:leaf)

X 4 cores/leaf

Understanding different variance structure

X 6 plants

X 3 leaves/plant

(1|plant) + (1|plant:leaf)

X 4 cores/leaf

Understanding different variance structure

X 6 cages

X 4 time points/mouse
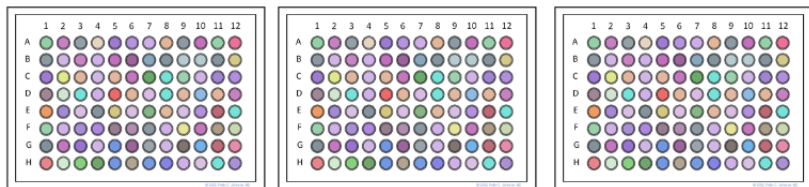
(1|cage) + (1|cage:mouse)

Understanding different variance structure: **Nested and Crossed structures**

Crossed: (1|plate) + (1|row) + (1|column)

Nested: (1|plate) + (1|plate:row) + (1|plate:column) = (1|plate/row/column)
What is the difference?

crossed random effects: one level of a random effect can appear in conjunction with more than one level of another random effect

Beyond random intercepts

Right-hand side = what groups observations

Nested, crossed et al. on the right hand side of the |: (1|something)

How are data related to each other, what groups them

Does not tell what parameter vary accoring to group

Beyond random intercepts

Right-hand side = what groups observations

Nested, crossed et al. on the right hand side of the |: (1|something)
How are data related to each other, what groups them
Does not tell what parameter vary accoring to group

Left-hand side = what varies according to grouping

The 1 stands for **intercept**
But many things can go to the left hand side.

Beyond random intercepts

Right-hand side = what groups observations

Nested, crossed et al. on the right hand side of the |: (1|something)
How are data related to each other, what groups them
Does not tell what parameter vary accoring to group

Left-hand side = what varies according to grouping

The 1 stands for **intercept**
But many things can go to the left hand side.

Random interactions, random regressions, random slopes. . .

e.g., $y \sim 1 + x + (1 + x|something)$

Everything you need to know about mixed models

- http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html
- Subscribe to mailing-list:
  https://stat.ethz.ch/mailman/listinfo/r-sig-mixed-models