

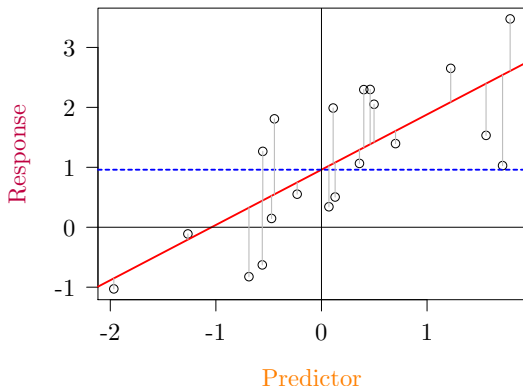
# Linear mixed models

Why, what, how?

Timothée Bonnet with content from Terry Neeman

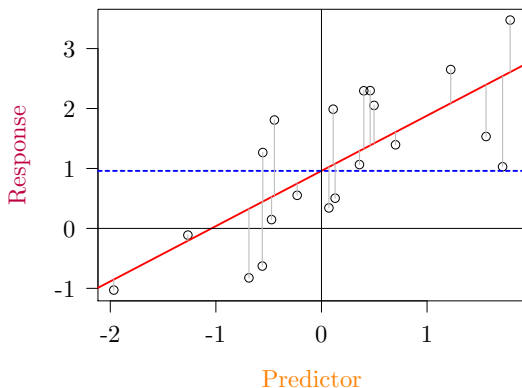
January 15, 2019

# Statistical models: MEAN and VARIANCE components



$$\text{Response} = \text{Intercept} + \text{Slope} \times \text{Predictor}$$

# Statistical models: MEAN and VARIANCE components



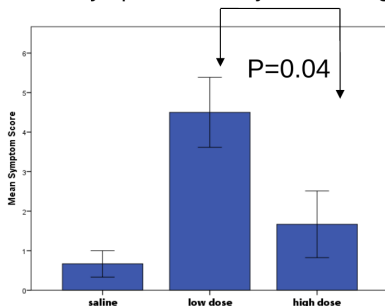
$$\text{Response} = \underbrace{\text{Intercept} + \text{Slope} \times \text{Predictor}}_{\substack{\text{Mean Structure} \\ \text{Experimental factors}}} + \underbrace{\text{Error}}_{\substack{\text{Variance Structure} \\ \text{Unrelated to experiment factors} \\ \text{Unexplained "noise" } \epsilon \sim N(0, \sigma^2)}}$$

What is in  $\epsilon$ ? How can we tweak that? Why should we care?

# Let's go to exercises 1 and 2

# 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 1:  
saline



Cage 2:  
Low Dose



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 1:  
saline



Cage 2:  
Low Dose



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



x 6

10 eggs/tube



x 6

10 eggs/tube



x 6

10 eggs/tube



x 6

10 eggs/tube



x 6

10 eggs/tube



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?

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

# What is going on conceptually?

$$\text{Response} = \underbrace{\text{Intercept} + \text{Slope} \times \text{Predictor}}_{\substack{\text{Mean Structure} \\ \text{Experimental factors}}} + \underbrace{\text{Error}}_{\substack{\text{Variance Structure} \\ \text{Unrelated to experiment factors} \\ \text{Unexplained "noise"}}$$

with  $\epsilon \sim N(0, \sigma)$

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?

$$\text{Response} = \underbrace{\text{Intercept} + \text{Slope} \times \text{Predictor}}_{\substack{\text{Mean Structure} \\ \text{Experimental factors}}} + \underbrace{\text{Error}}_{\substack{\text{Variance Structure} \\ \text{Unrelated to experiment factors} \\ \text{Unexplained "noise"}}$$

with  $\epsilon \sim N(0, \sigma)$

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,  $\epsilon$  are non-independent

# Fixed or random effect?

## In this example

- Doesn't change inference (same p-value for temperature)
- Summary cleaner with random effect



# Fixed or random effect?

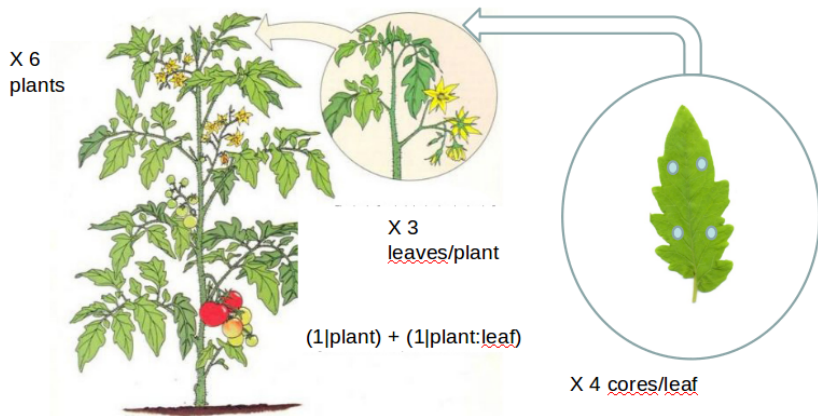
## In this example

- Doesn't change inference (same p-value for temperature)
- Summary cleaner with random effect

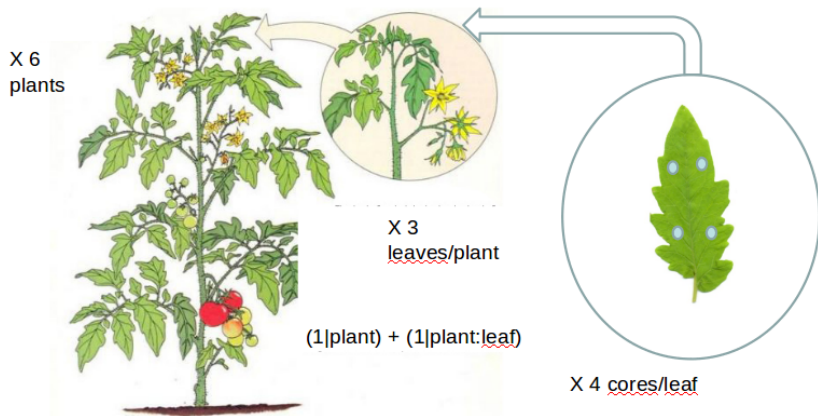
## In general

- Generally 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)

# Understanding different variance structure



# Understanding different variance structure



# Understanding different variance structure

X 6 cages



X 4 time  
points/mouse

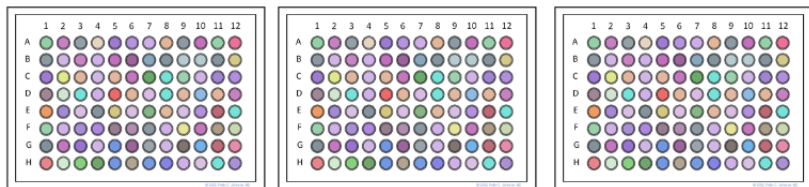
(1|cage) + (1|cage:mouse)

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

**Crossed:**  $(1|\text{plate}) + (1|\text{row}) + (1|\text{column})$

**Nested:**  $(1|\text{plate}) + (1|\text{plate}:\text{row}) + (1|\text{plate}:\text{column}) = (1|\text{plate}/\text{row}/\text{column})$

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*

# 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>