Statistical Modelling: Understanding Mean Structure

Chapter 3

Terry Neeman and Timothée Bonnet

November 20, 2018

Key components of a statistical model of an experiment

- Outcome measure
  - Response variable
  - Measure of interest
- Experimental factors
  - Conditions that can be manipulated
  - Conditions of interest (e.g. genotype, gender)
  - Main questions: do the conditions impact upon the outcome measure?
- Blocking factors
  - Conditions (not of interest) that may impact upon the outcome measure
  - Sources of variation in the experiment that need to be controlled for
  - Clustering of experimental units

ALWAYS BEGIN WITH A RESEARCH QUESTION

Example 1: Can drought tolerance in Arabidopsis be improved through genetic modification?

Context

Outcome measure: Leaf water retention LWR (%)
Experimental factors:

- Gene A, genotypes (`AA/aa`)
- Gene B, genotypes (`BB/bb`)

How many parameters?

Example 1: Can drought tolerance in Arabidopsis be improved through genetic modification?

Context

Outcome measure: Leaf water retention LWR (%)
Experimental factors:

- Gene A, genotypes (`AA/aa`)
- Gene B, genotypes (`BB/bb`)

How many parameters?

4 treatments		Gene A	
		AA	aa
Gene B	BB	C	$C + A$
	bb	$C + B$	$C + A + B + D$

Two different models

Additive model - 3 parameters

4 treatments		Gene A	
		AA	aa
Gene B	BB	C	$C + A$
	bb	$C + B$	$C + A + B$

Full factorial model / Interactive model - 4 parameters

4 treatments		Gene A	
		AA	aa
Gene B	BB	C	$C + A$
	bb	$C + B$	$C + A + B + D$

What is different? What does the additive model assume?

Which model to use?

Additive model - 3 parameters

4 treatments		Gene A	
		AA	aa
Gene B	BB	C	$C + A$
	bb	$C + B$	$C + A + B$

Full factorial model / Interactive model - 4 parameters

4 treatments		Gene A	
		AA	aa
Gene B	BB	C	$C + A$
	bb	$C + B$	$C + A + B + D$

Analysis in R

1. Import data "Prac3mockLWR.csv"
2. Visualize data
3. Model data
4. Assess model assumptions

Analysis in R

1. Import data ”Prac3mockLWR.csv”

```
LWR <- read.csv(\Prac3mockLWR.csv")
```

Analysis in R

2. Visualise the data

```
ggplot(LWR, aes(GeneB,LWR,colour=GeneA)) +
  geom_boxplot() + geom_point()
```

Full factorial or additive?

Analysis in R

3. Model data

```
lmadditive <- lm(LWR ~ GeneA + GeneB, data = LWR)
summary(lmadditive)
anova(lmadditive)
```

```
lminteraction <- lm(LWR ~ GeneA * GeneB, data = LWR)
summary(lminteraction)
anova(lminteraction)
emmeans(lminteraction, pairwise ~ GeneA|GeneB)
emmeans(lminteraction, pairwise ~ GeneB|GeneA)
```

What are the estimates for A, B, C, D under each models?

Analysis in R

4. Model assumptions

```r
plot(lminteraction)
```