

# **6. Experimental design: concept and terminology**

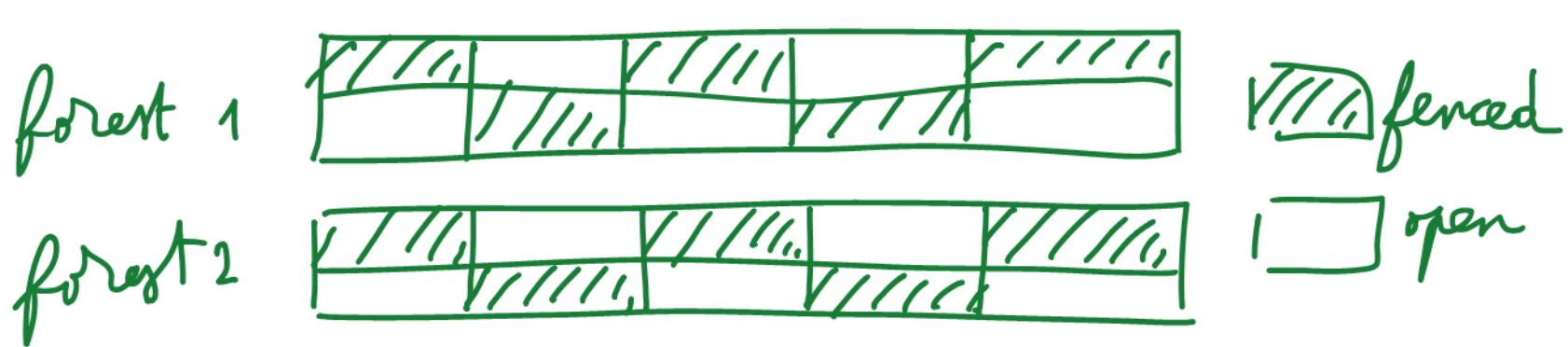
# Experimental Design vs. Observational analysis

## Sampling versus experiments

- In both cases: information about variables is gathered and analyzed
- experiments presuppose intervention through applying a ***treatment*** (an action or absence of an action) to a unit, called the ***experimental unit***. The experimental unit is an item on which the treatment is applied.
- The goal is to obtain results that indicate cause and effect.

# Example

- The effect of animal browsing on herbaceous plants is of interest. In each of two forest types, 10 areas are established at the beginning of the year. Five out of the 10 are fenced off, eliminating animal browsing. The rest are marked but left open to animals. The heights and coverages of plants are measured at the end of the year.



# Definitions of terms

- For each **experimental unit** = *forest area*, measures of the *variables of interest* (i.e., *response* or *dependent variables*) = *plant heights and coverages* are used to indicate treatment impacts.
- Treatments are randomly assigned to the experimental units = *forest areas*.  
*2 forest types x 2 animal browsing intensities = 4 treatments*
- *Replication* is the observation of two or more experimental units under identical experimental conditions.  
*5 replications for each treatment*
- A *factor* is a grouping of related treatments.  
*2 Factor(s):*
  - *animal browsing (experimental)*
  - *forest type (observational)*

# Factors

- An explanatory variable (in regression context: « predictor », « independent variable »)
- Experimental factor: assigned at random to the experimental unit
  - Ex. animal browsing in above example
- Observational factor: not under control of the investigator
  - Ex. forest type in above example
  - Still we can look at

$$H_0: \mu_{\text{forest type 1}} = \mu_{\text{forest type 2}}$$

$$H_1: \mu_{\text{forest type 1}} \neq \mu_{\text{forest type 2}}$$

!!can be found in experimental studies and should be recognized as such!!

!!cause-and-effect inferences cannot be made for these factors!!

- Ex. differences between forest types can be due to many reasons: different fauna, different trees, different weather conditions, ...

# Experimental factors can be

- Qualitative: levels differ by qualitative attribute
  - Ex. animal browsing and forest type in above example
- Quantitative: each level is described by a numerical quantity on an equal interval scale
  - Ex. Temperature in degrees °C, age in years, price in euros

# Factor level:

- A particular form of the factor
  - Ex. Study of browsing effect in two forest types:  
two levels of forest type factor (forest type 1, forest type 2)  
two levels of browsing type factor (fenced or open)

# Treatment

- In single-factor studies: a treatment corresponds to a factor level
- In multifactor studies: a treatment corresponds to a combination of factor levels (sometimes referred to as « treatment combination »)
- Definition of treatment can be difficult:  
Ex. study whether C or JAVA is a better language for an introductory computing course. Some teachers prefer C, others JAVA. How do we rule out effect of teacher?  
→ should instructors be randomized over two languages (some forced to teach in language they don't prefer? Should each teacher teach in both languages? Should language preference be added as second factor?)

# Experimental Units

- The smallest unit of experimental material to which a treatment can be assigned
- *Determined by the method of randomization*  
Ex. Effect of browsing intensity on tree height  
Random assignment of different browsing intensities to individual trees: impossible  
Smallest unit of experimental material to which a treatment (browsing intensity) can be randomly assigned is the forest plot  
→ Forest plot is experimental unit
- Ex. Effect of pay on productivity  
Random assignment of different pays to individual employees: impossible  
Smallest unit of experimental material to which a treatment (pay) can be randomly assigned is the factory  
→ Factory is experimental unit

# What is “*treatments are randomly assigned to experimental units*”?

- Randomization tends to average out the influence of factors not under direct control of the experimenter (whether anticipated or not)
- Haphazard vs. random allocation (to treatment)  
Choosing to treat the first 5 mice that are removed from a box of 10 may result in the fattest, oldest and slowest mice being treated. Random does not mean haphazard.
- Practical problems and implications
  - It is not possible to allocate disease status at random
  - It would be neither ethical nor feasible to consider randomizing mothers to smoking or non-smoking during pregnancy.

# Sample Size and Replication

- If each treatment is repeated twice: « there were two complete replicates of the experiment»
- Sample size is integer multiple of number of treatments (determined by number of complete replicates)
- Replication: difference in response for the same treatment under similar experimental conditions is due to experimental error
- Replications → Estimation of the experimental error variance:
  - Required for testing the presence of treatment effects
  - Required for establishing confidence interval estimates of these effects

- *Experimental error* is the measure of variance due to chance causes, among experimental units that received the same treatment.
- The impacts of treatments on the response variables will be detectable only if the impacts are measurably larger than the variance due to chance causes.
- To reduce the variability due to causes other than those manipulated by the experimenter, relatively homogenous experimental units are carefully selected.

- Random allocation of a treatment to an experimental unit helps insure that the measured results are due to the treatment, and not to another cause.

Example: Analysis of the effect of browsing:

if we have applied the no browsing treatment (fencing) to experimental units on south facing sites, whereas browsing treatments (no fencing) are applied only to north facing sites, we would not know if differences in average height growth were due to the browsing, the orientation of the sites, or both. The results would be *confounded* and very difficult to interpret.

# Variations in experimental design

The « Design of an experiment » refers to the structure of the experiment:

- The set of explanatory factors included in the study
- The set of treatments included in the study
- The set of experimental units included in the study
- The rules and procedures by which the treatments are randomly assigned to the experimental units (or vice versa)
- The outcome measurements that are made on the experimental units

# *Introduction of More Than One Factor:*

- Interested in the interaction among factors, and the effect of each factor.
- A treatment represents a particular combination of levels from each of the factors.
- When all factor levels of one factor are given for all levels of each of the other factors, this is a *crossed experiment*.
- When the levels of a factor are unique to a particular level of another factor, this is a *nested experiment*

# Example: yield of a chemical process

- Three levels of temperature
- Two levels of solvent concentration
- $3 \times 2 = 6$  factor level combinations

Factor combination	Temperature	Solvent concentration
1	Low	Low
2	Low	High
3	Medium	Low
4	Medium	High
5	High	Low
6	High	High

# Example: yield of a chemical process

- These factors can be represented by the two-way table:

		Temperature		
		Low	Medium	High
Solvent	Low	X	X	X
	High	X	X	X

→The two factors are crossed

## *Nesting of Factors*

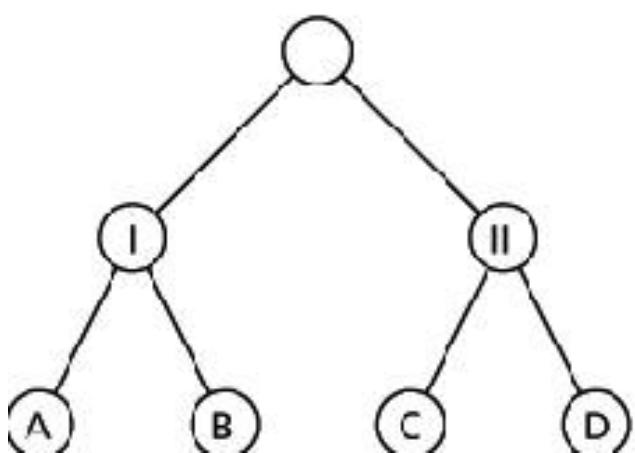
- Treatment levels for one factor may be particular to the level of another factor, resulting in nesting of treatments.

# Example: Effect of operators on production yield in three manufacturing plants

- Three levels of manufacturing plant
- In each plant, production yield of three selected operators is recorded  
→ Operators are nested within manufacturing plants

		Operator								
		1	2	3	4	5	6	7	8	9
Plant	1	X	X	X						
	2				X	X	X			
	3							X	X	X

**Nested design**



**Crossed design**

	A	B	C	D
I				
II				

# *Fixed, Random, or Mixed Effects:*

- Fixed factors: the experimenter would like to know the change that is due to the particular treatments applied; only interested in the treatment levels that are in the experiment (e.g., difference in growth between two particular genetic stocks) [*fixed effects*]
- Random factors: the variance due to the factor is of interest, not particular levels (e.g., variance due to different genetic stocks—randomly select different stock to use as the treatment) [*random effects*]
- Mixture of factor types: Commonly, experiments include a mixture of factors, some random and some fixed [*mixed effect*].

# Main questions in experiments

- Do the treatments affect the variable of interest?
- For fixed effects: Is there a difference between the treatment means of the variable of interest? Which means differ? What are the means by treatment and confidence intervals on these means?
- For random effects: Do the treatments account for some of the variance of the variables of interest? How much?

## Restricted Randomization Through Blocking:

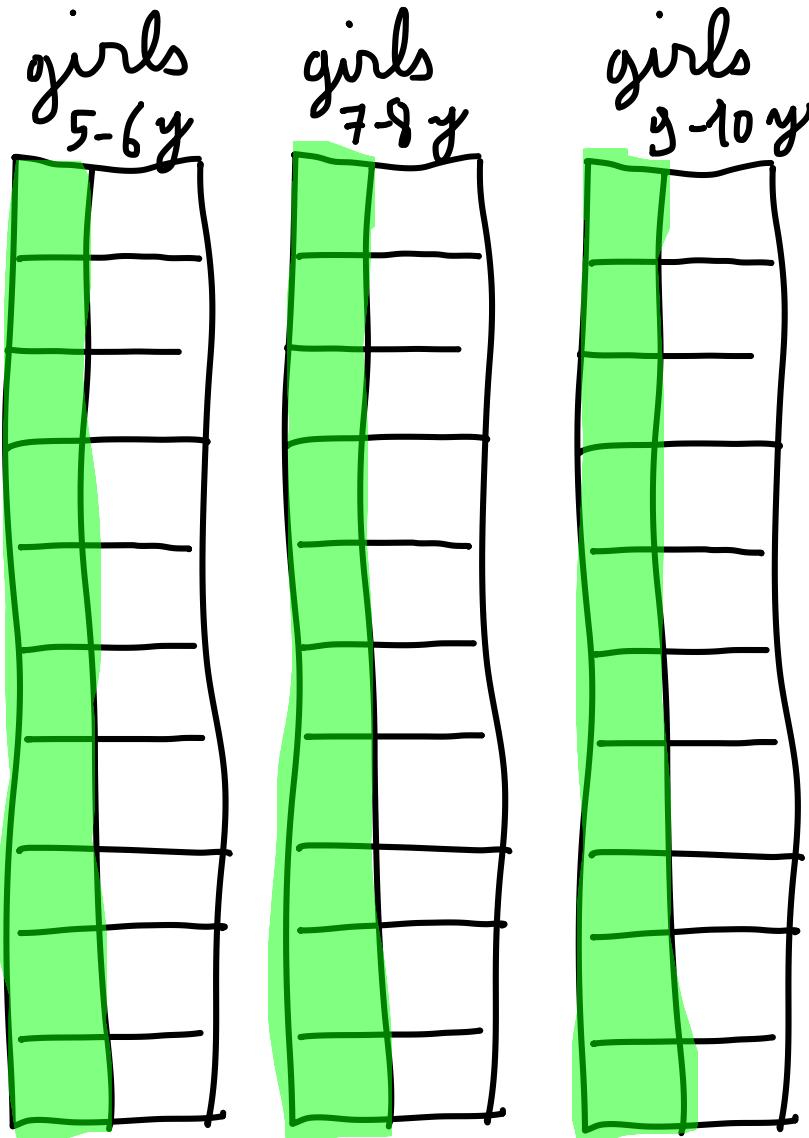
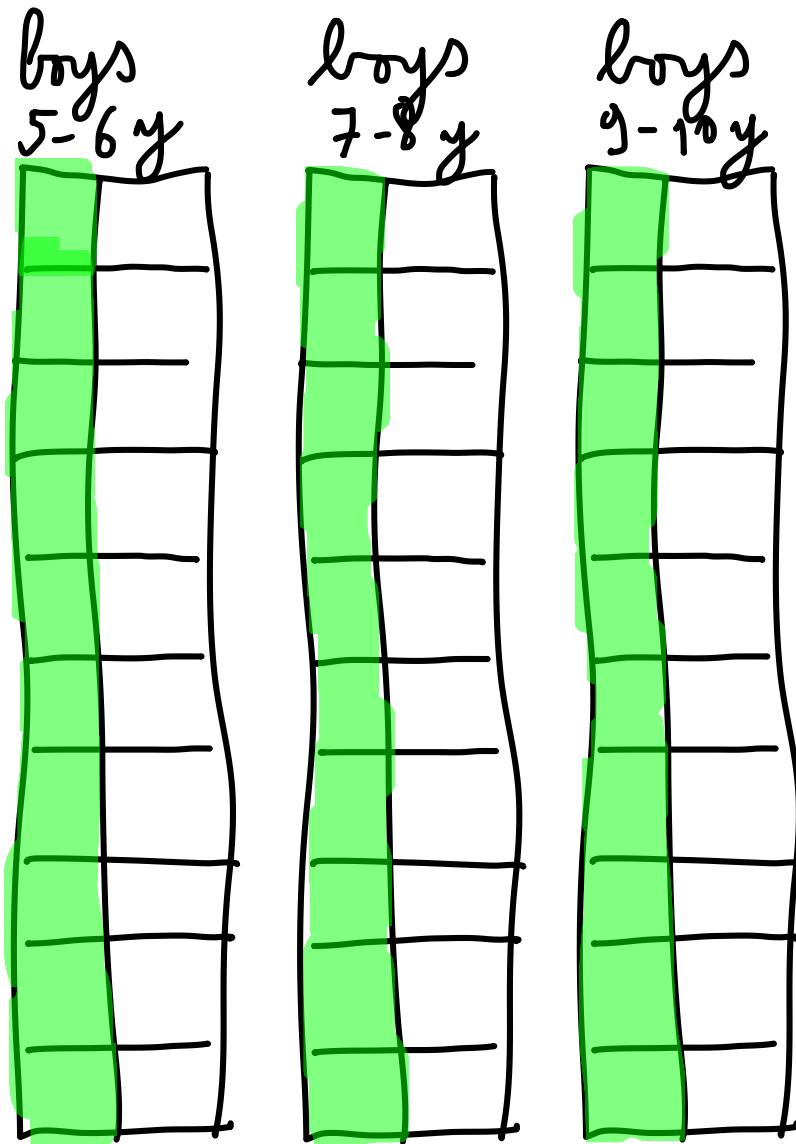
Example: Study of effect of vitamin C on number of colds

- Half of the children randomly assigned to treatment « receiving vit C », other half assigned to treatment « not receiving vit C »
- *Nuisance factors or confounding factors*: affect the response but are not of interest to the investigator:
  - Gender
  - Age
  - General health status
  - ...
- Solution 1: Hold nuisance factors constant: Limit the experiment to (for example) one gender, and one age window (ex. boys of 5-6 years old)

✓ Disadvantages:

- lower sample size → lower precision
- Problem for generalization

- Solution 2: Heterogeneous experimental units are divided into homogeneous subgroups called blocks



vit C

no vit C

- Solution 2: Heterogeneous experimental units are divided into homogeneous subgroups called *blocks*
  - ✓ separate randomizations of treatments to experimental units within each block
    - assignment of treatments can only be made to experimental units within the given block: « Six Restricted Randomizations »
  - ✓ Overall effect of the experimental factor is obtained by combining the estimated effects from each of the blocks
  - ✓ Advantages:
    - Generalizability
    - High precision because of small experimental errors within the blocks

# Restricted Randomization Through Blocking: Randomized Block (RCB), Latin Square, and Incomplete Blocks Designs:

- Technique to increase precision in experiment through reduction in experimental error variance
  - identify and control factors that contribute to variation in the experimental error
- Randomize treatments within blocks of experimental units
- Reduces the variance by taking away variance due to the item used in blocking (e.g., high, medium and low site productivity)
- Results in more homogeneous experimental units within each block.

## Restricted Randomization Through Splitting Experimental Units:

- Called “split plot”
- An experimental unit is split. Another factor is randomly applied to the split.

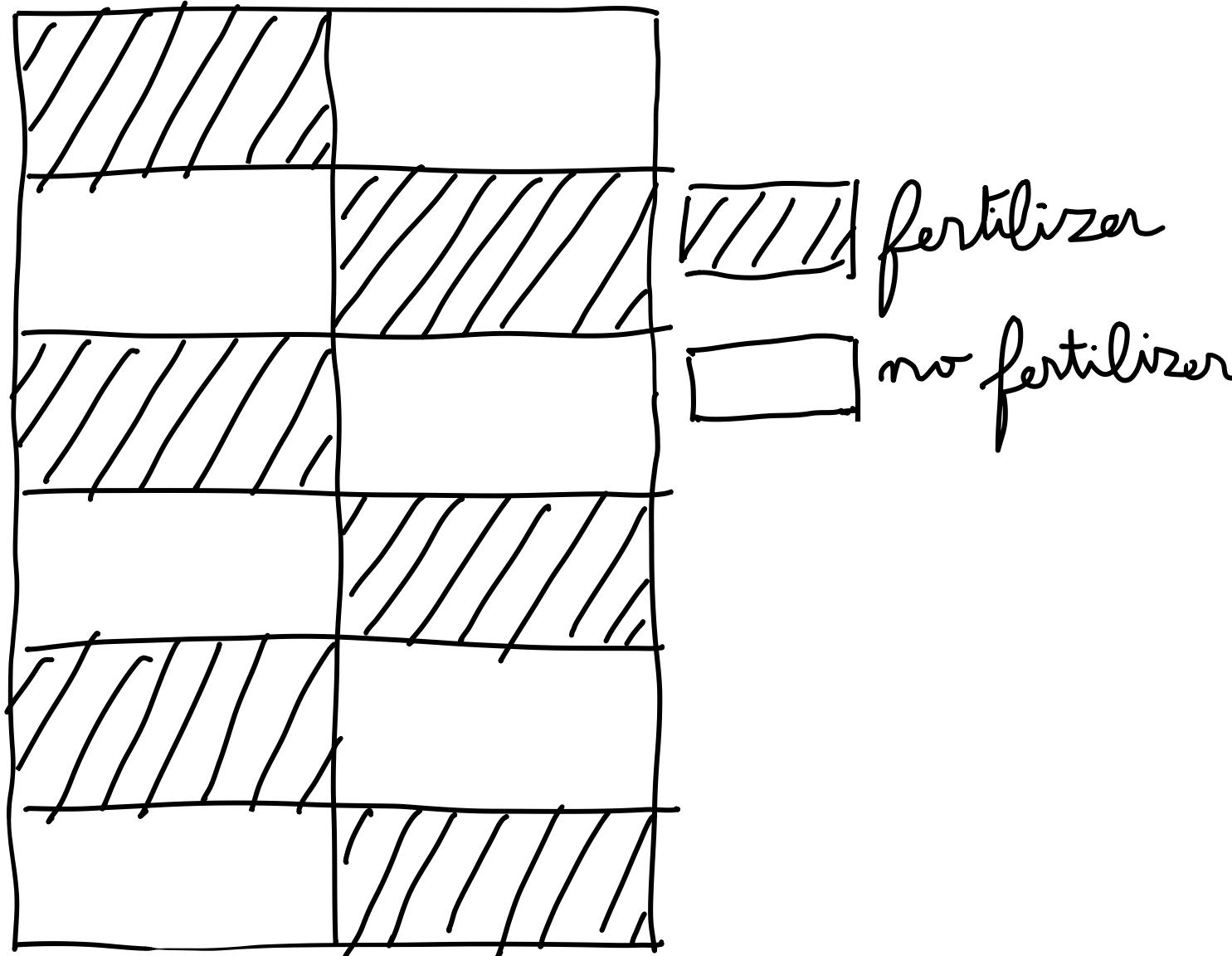
Example: The factor fertilizer is applied to 0.001 ha plots. Each of the 0.001 ha plot is then split into two, and two different species are planted in each. Fertilizer is applied to the whole plot, and species is applied to the split plot. Species is therefore randomly assigned to the split plot, not to the whole experimental unit.

# Restricted Randomization Through Splitting

## Experimental Units:

Example: The factor fertilizer is randomly applied to 0.001 ha plots.

Experimental unit = 0.001 ha plots



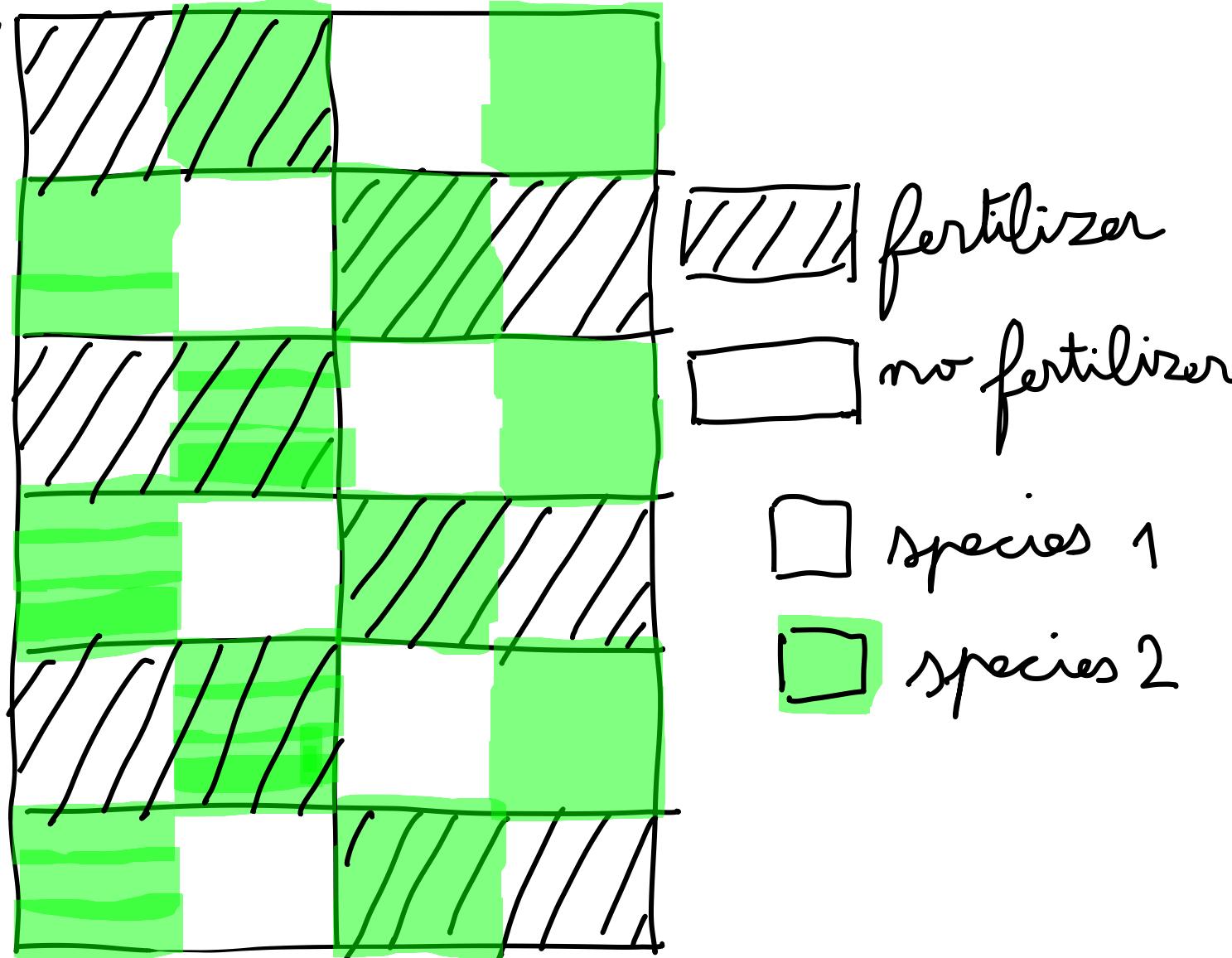
# Restricted Randomization Through Splitting

## Experimental Units:

Each of the 0.001 ha plot is then split into two, and two different species are planted in each.

**Fertilizer** is randomly applied to the whole experimental unit.

**Species** is randomly applied to the split experimental unit.



# Example:

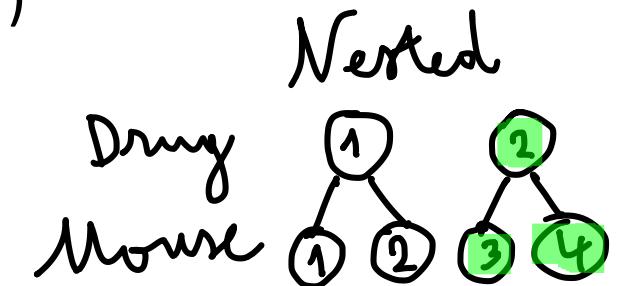
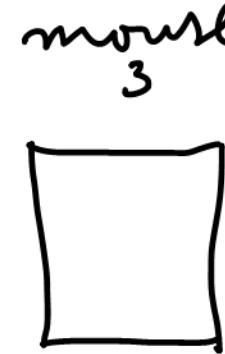
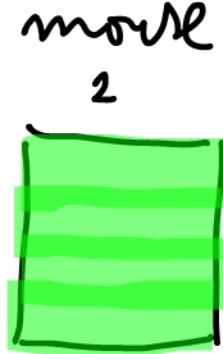
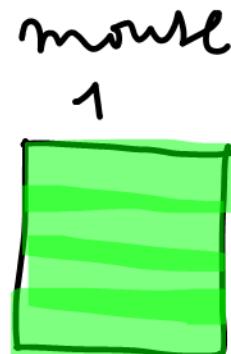
- The effect of two drugs is of interest. The drugs are tested on four mice
- 2 Factors: Drug (= Fixed Factor) and Mouse (= Random Factor)

1. If the same mouse cannot be re-used for different drugs:

→ Each level of mouse is unique for a certain level of drug

→ Levels of random factor (mouse) nested within levels of fixed factor (drug)

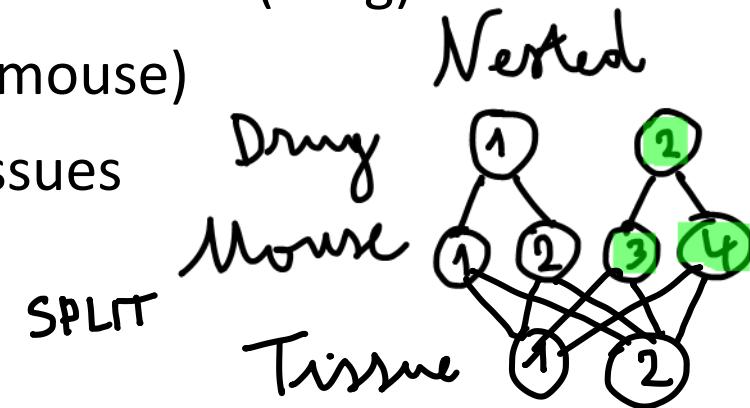
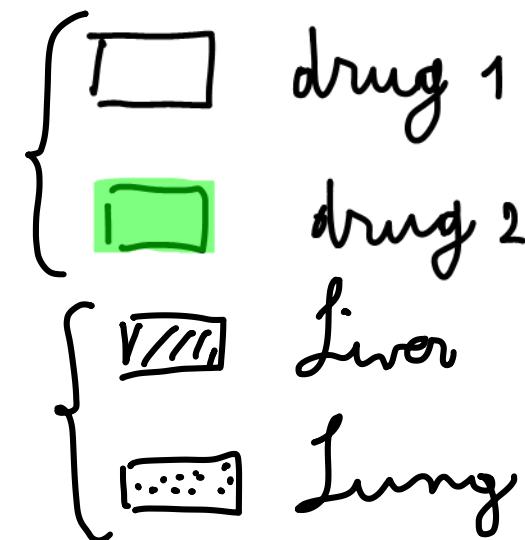
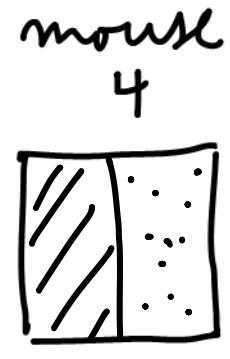
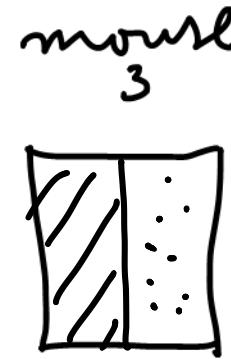
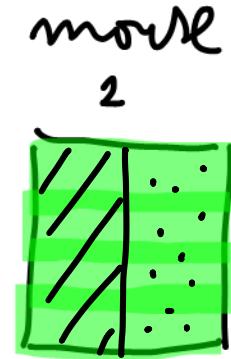
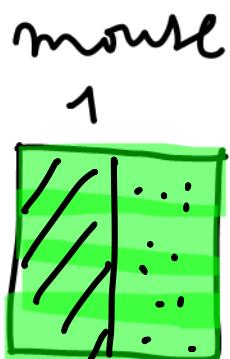
→ Experimental unit = mouse



2. If the same mouse cannot be re-used AND the effect of the drug on two different tissues is of interest:

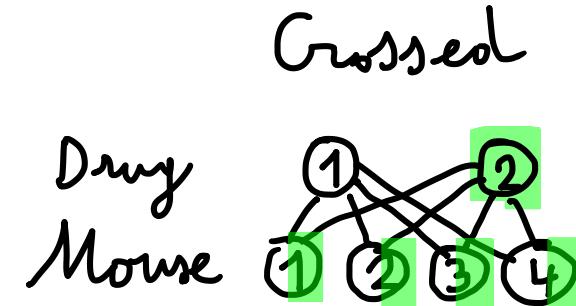
- Two fixed factors: Drug and Tissue
- Levels of random factor (mouse) nested within levels of fixed factor (drug)
- Drug is randomly applied to the whole experimental unit (mouse)
- Split plot design: The experimental unit is split into two tissues

Tissue is « applied » to the split experimental unit.

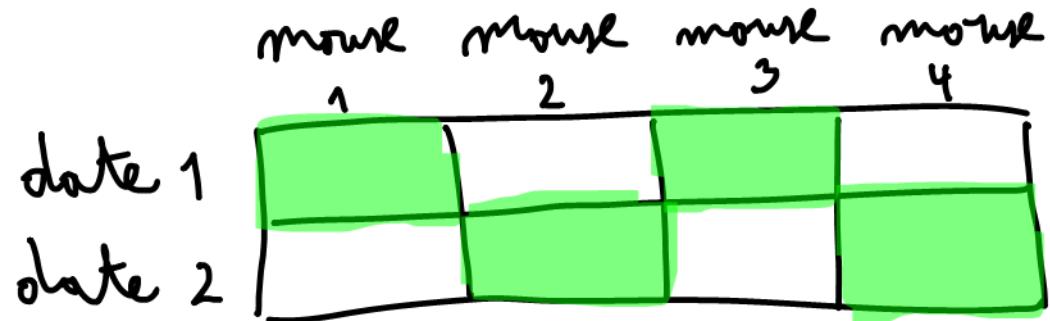


### 3. If the same mouse can be re-used for different drugs:

- Crossed experiment containing all combinations of drug levels x mouse levels
- 4 mouse levels x 2 drug levels = 8 treatments
- Random factor (mouse) designed as a block



*Overall effect of the fixed factor (drug) is obtained by combining the estimated effects from each of the random factor levels (mice = blocks)*



4. If the same mouse can be re-used for different drugs AND, in each mouse, two tissues are tested:

→ Random factor (mouse) designed as a block

→ Two fixed effects: Drug and Tissue

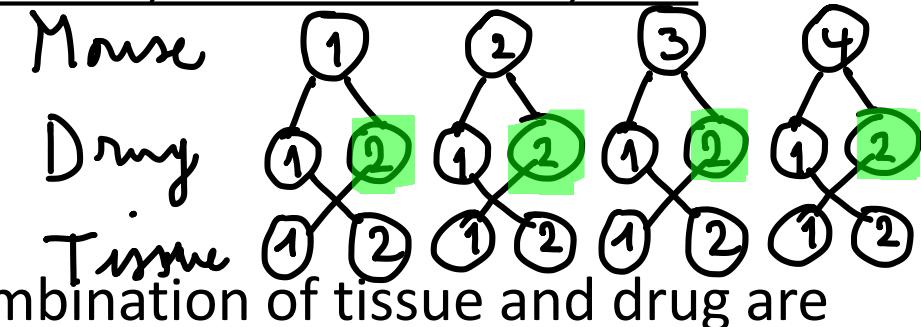
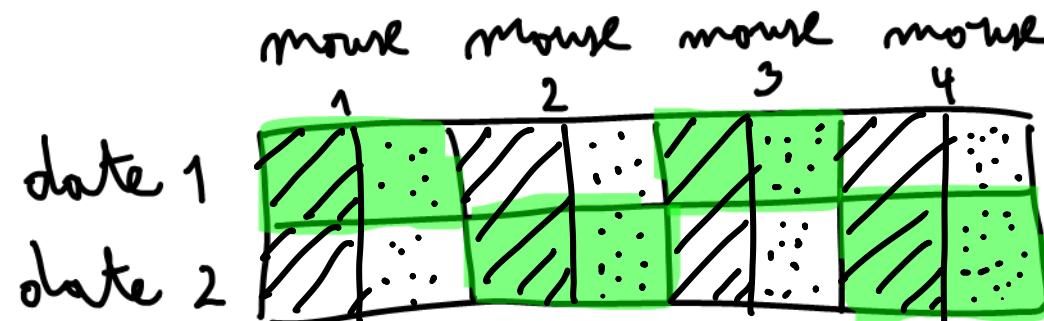
→ The fixed effects (tissue, drug) are **crossed** (each combination of tissue and drug are tested)

→ the treatments (tissue x drug) are themselves nested within **replicates** (= mice, = levels of random factor)

→ Split plot design: The experimental unit (mouse) is split into two tissues

Drug is applied to the whole experimental unit

Tissue is « applied » to the split experimental unit.



4 treatments

drug 1 - liver

drug 2 - liver

drug 1 - lung

drug 2 - lung

# *Hierarchical Designs and Sub-Sampling:*

- Commonly in experiments, the experimental unit represents a group of items that we measure.
  - ✓ A pot in which several plants germinate from seeds.
  - ✓ The liver tissue of a mouse, containing many cells.
- Treatments are randomly assigned to the larger unit
  - ✓ To each pot not to each seedling
  - ✓ To each mouse, not to each cell
  - ➔ The experimental unit is the larger sized unit.
- May want to quantify **variance due to the experimental unit**
  - ✓ Pots
  - ✓ Mice

**and variance due to units within experimental unit:**

  - ✓ Plants
  - ✓ Cells

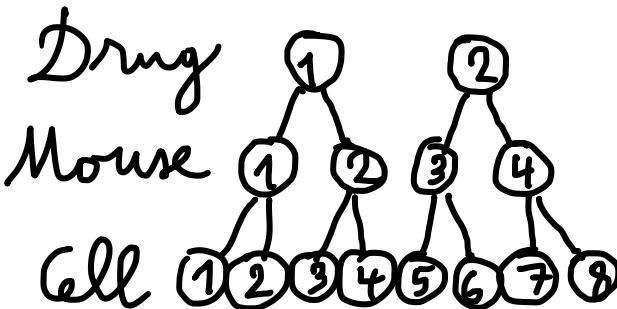
# Hierarchical Designs and Sub-Sampling:

- A common variation on hierarchical designs is measuring a sample of items, instead of measuring all items in an experimental unit.
- **Experimental units** (Pots, Mice) and **units with experimental units** (Plants, Cells) are
  - 1) random effects
  - 2) nested in the treatment (fixed effect)
  - 3) hierarchical

treatment : FIXED

experimental unit: RANDOM

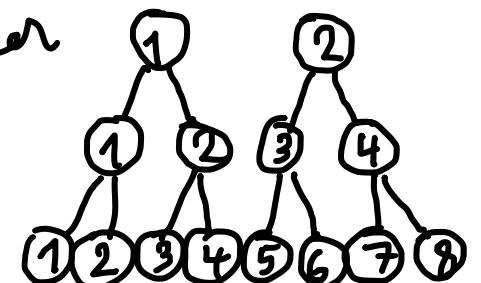
sub-samples: RANDOM



Fertiliser

Pots

Plants



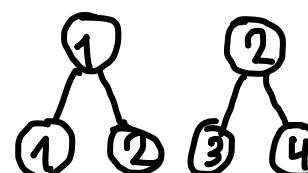
# *Introduction of Covariates* → ANCOVA

- The initial conditions for an experiment may not be the same for all experimental units, even if blocking is used to group the units.
- Site measures such as soil moisture and temperature, and starting conditions for individuals such as starting height, are then measured (called covariates) along with the response variable
- These covariates are used to reduce the experimental error.
- Covariates are usually interval or ratio scale (continuous).

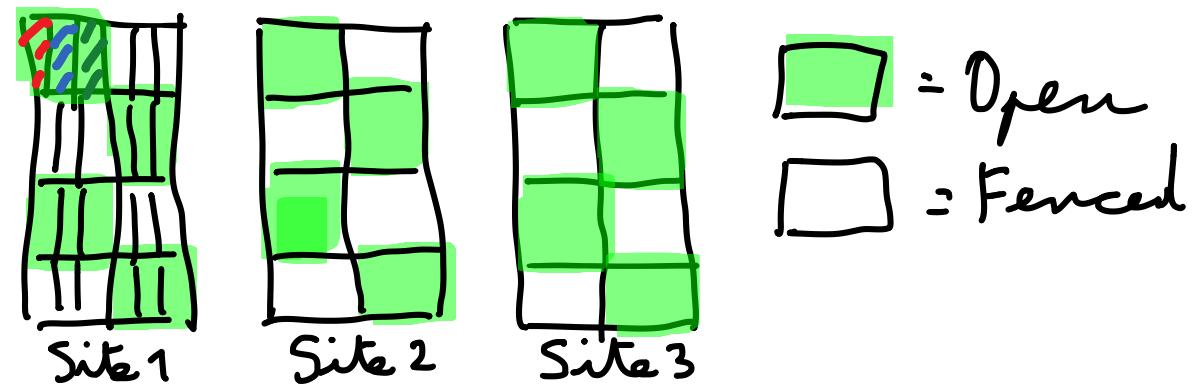
# Designs in use

- The most simple design is one fixed-effects factor, with random allocation of treatments to each experimental unit, with no 1) blocking; 2) sub-sampling; 4) splits; or 5) covariates

treatment : **FIXED** Fertilizer  
experimental unit: **RANDOM** Pots



## Designs in use



- Most designs use combinations of the different variations. For example, one fixed-effects factor, one random-effects factor, blocked into three sites, with trees measured within plots within experimental units (sub-sampling/hierarchical), and measures taken at the beginning of the experiment are used as covariates (e.g., initial heights of trees).

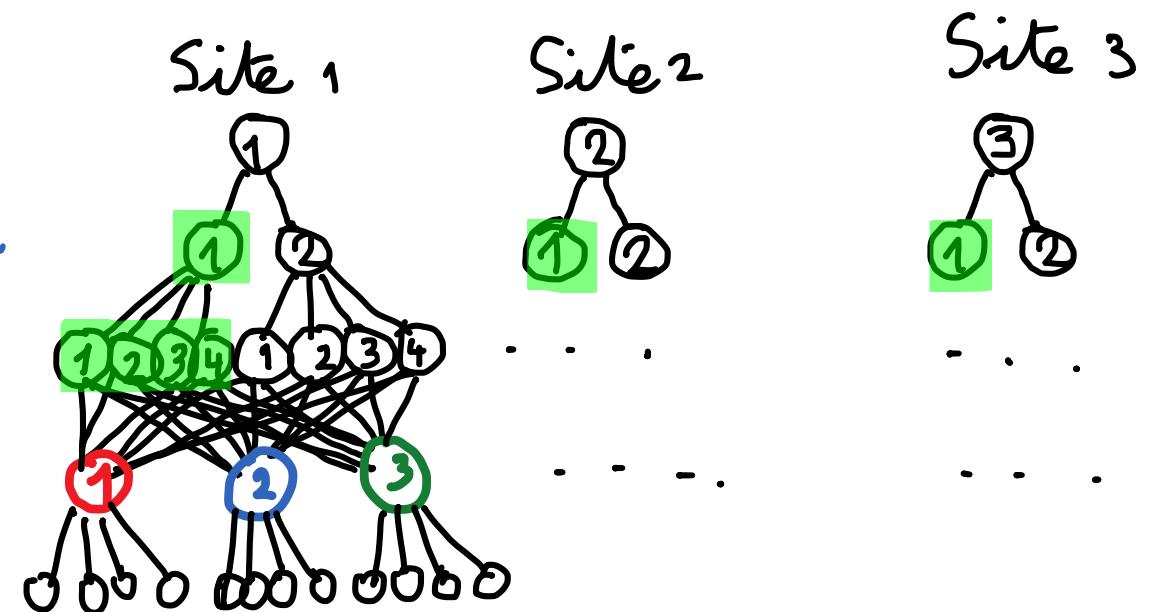
Sites = blocks: Random

Browsing(Open/Fenced) = FIXED factor

Plots = Experimental unit = RANDOM

SPLIT plots : 3 species = RANDOM

subsampling individual trees



*In this course: start with the simple designs and add complexity.*

- Completely Randomized Design (CRD)
  - Only fixed factors
  - Adding a random factor
- Restrictions in randomization
  - Randomized Complete Block Design (RCB)
  - Split-Plots
- Nested designs, Sub-sampling

## **7. Completely randomized design (CRD) – One factor, Fixed Effects**

# Completely Randomized Design (CRD)

- Treatments are randomly assigned to treatment units
- No blocking is used
- We measure a variable of interest for each experimental unit

# CRD: One Factor Experiment, Fixed Effects

## Main questions of interest

Are the treatment means different?

Which means are different?

What are the estimated means and confidence intervals for these estimates?

# Example Single factor study: quantitative factor

Effectiveness of different dosages of drug

- 30 patients
- 3 dosage levels
- 10 patients in each dosage level

= completely randomized design based on a single, three-level  
quantitative factor

= *balanced* design (each treatment replicated the same number of times)

# Single factor with $J$ levels: TWO approaches

## I. Regression model

For example 
$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_{11} x_{ij}^2 + \varepsilon_{ij}$$

where:

$x_{ij}$  = centered dosage level amount for the  $ij$ th case

ONLY possible for a quantitative factor

# Single factor with $J$ levels: TWO approaches

## II. Analysis of Variance model (ANOVA)

$J-1$  dummy variables as predictors

A regression model with *only dummy predictor variables* is called an *analysis of variance model*

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \varepsilon_{ij}$$

Example effectiveness of drug dosage:  
treatment  $j$  (1->3)  
replicate  $i$  (1->10)

where:

$$X_{ij1} = \begin{cases} 1 & \text{if treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

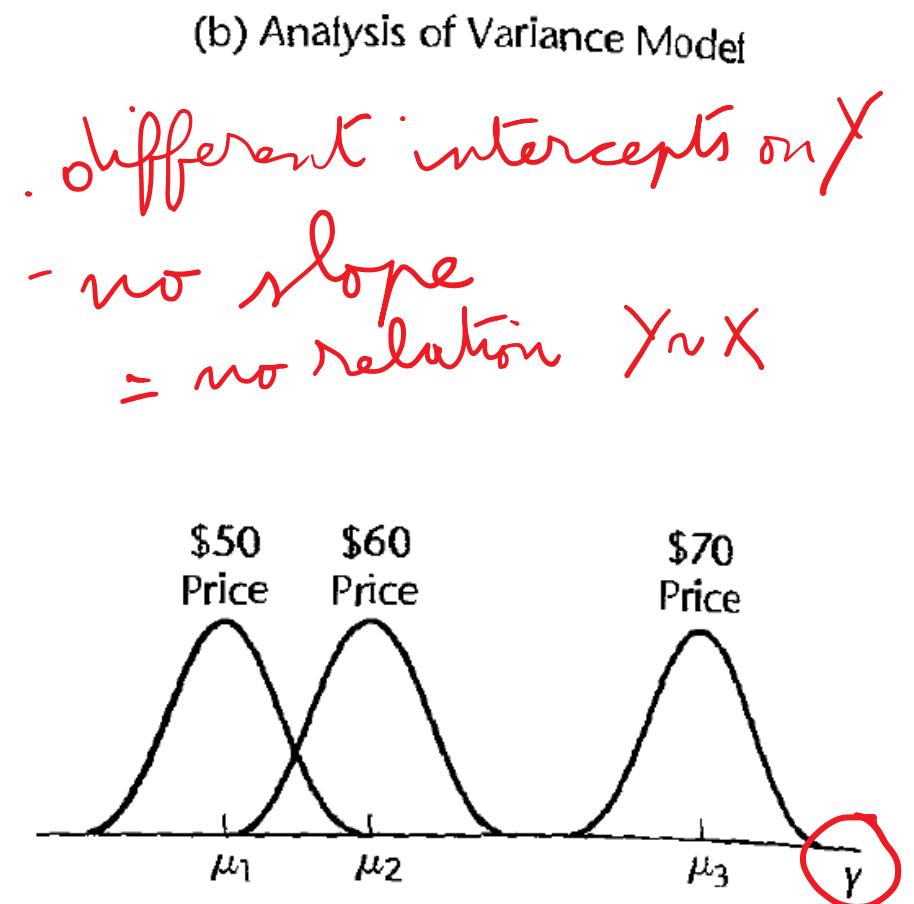
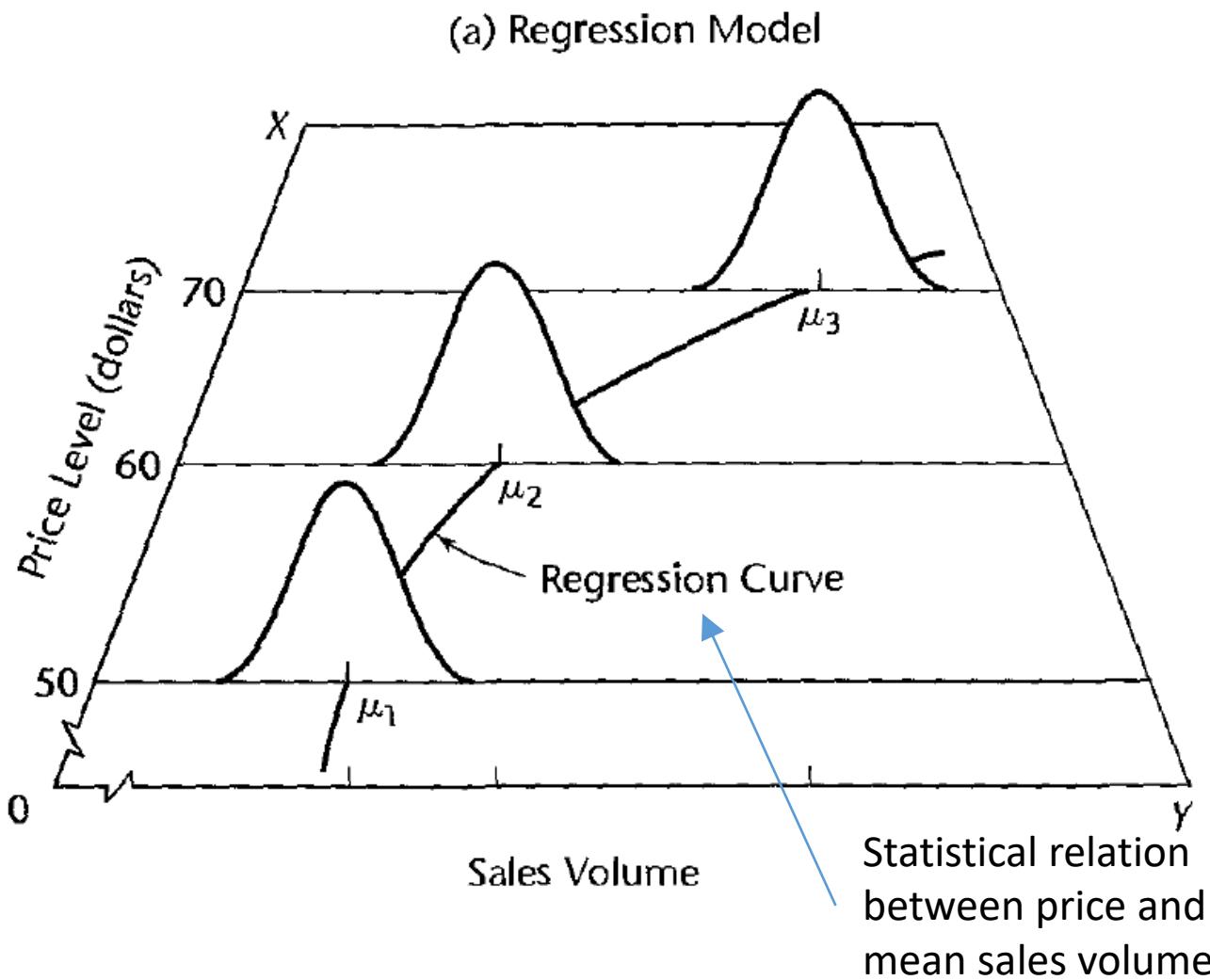
→ The intercept is simply the mean of the reference group. The coefficients for the other groups are the differences in the mean between the reference group and the other groups.

# Relation between Regression and ANOVA

Difference ANOVA vs regression:

- Predictor variable may be qualitative
- If predictor variables are quantitative, no assumption is made about the nature of the statistical relation between them and the response variable

# Illustration: effect of price levels on sales volume



# Treating the levels of the factor as qualitative vs. quantitative?

If you believe that the response is linear (and only linear) then treating the salt levels as quantitative (continuous) is appropriate : a regression analysis

Treating the levels as qualitative (e.g. nominal or ordinal scale) allows for ANY response function other than a straight line with 0 slope (which corresponds to zero effect).

→ detect a threshold effect, or a quadratic effect, or any other pattern in addition to a linear effect by doing a ANOVA rather than a REGRESSION.

# Single Factor ANOVA Model

- Corresponding to each factor level, there is a probability distribution of responses
- Assumptions of ANOVA model:
  1. Each probability distribution is normal
  2. Each probability distribution has the same variance (=homoscedasticity)
  3. The responses for each factor level are random selections from the corresponding probability distribution and are independent of the responses for any other factor level
- →Probability distributions differ only with respect to their means

# Single Factor ANOVA Model for CRD: Notation

Population:  $y_{ij} = \mu + \tau_j + \varepsilon_{ij}$  OR  $y_{ij} = \mu_j + \varepsilon_{ij}$

$y_{ij}$  = response variable measured on experimental unit  $i$  and treatment  $j$

$j=1$  to  $J$  treatments

$\mu$  = the grand or overall mean regardless of treatment

$\mu_j$  = the mean of all measures possible for treatment  $j$

$\tau_j$  = the difference between the overall mean of all measures possible from all treatments and the mean of all possible measures for treatment  $j$ , called the *treatment effect*

$\varepsilon_{ij}$  = the difference between a particular measure for an experimental unit  $i$ , and the mean for the treatment  $j$  that was applied to it

$$\varepsilon_{ij} = y_{ij} - \mu_j$$

# Single Factor ANOVA Model for CRD: Notation

For the experiment:

$$y_{ij} = \bar{y}_{..} + \hat{\tau}_j + e_{ij} \quad \text{OR} \quad y_{ij} = \bar{y}_{\bullet j} + e_{ij}$$

$\bar{y}_{..}$  = the grand or overall mean of all measures from the experiment regardless of treatment; under the assumptions for the error terms, this will be an unbiased estimate of  $\mu$

$\bar{y}_{\bullet j}$  = the mean of all measures for treatment  $j$ ; under the assumptions for the error terms, this will be an unbiased estimate of  $\mu_j$

$\hat{\tau}_j$  = the difference between the mean of experiment measures for treatment  $j$  and the overall mean of measures from all treatments; under the error term assumptions, will be an unbiased estimate of  $\tau_j$

$e_{ij}$  = the difference between a particular measure for an experimental unit  $i$ , and the mean for the treatment  $j$  that was applied to it

$$e_{ij} = y_{ij} - \bar{y}_{\bullet j}$$

# Single Factor ANOVA Model for CRD: Notation

$n_j$ = the number of experimental units measured in treatment  $j$

$n_T$ = the number of experimental units measured over all

$$\text{treatments} = \sum_{j=1}^J n_j$$

# Example: Fertilization Trial

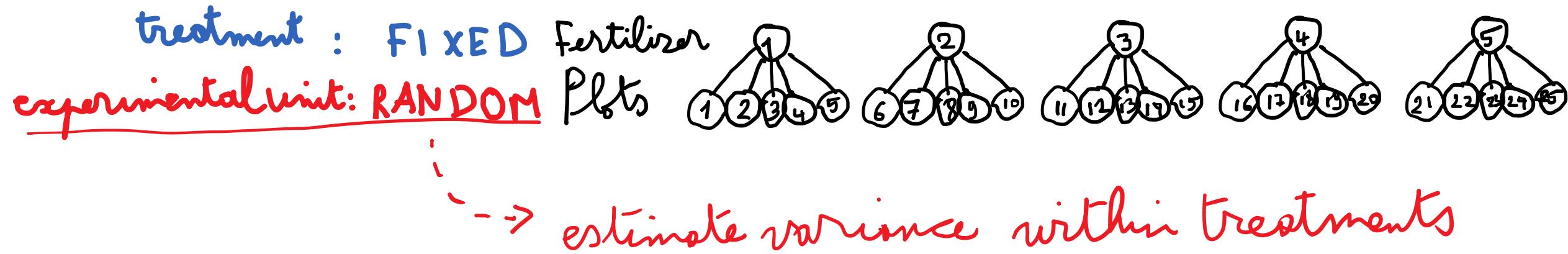
A forester would like to test whether different site preparation methods result in difference in heights. Twenty five areas each 0.02 ha in size are laid out over a fairly homogeneous area. Five site preparation treatments are randomly applied to the 25 plots. One hundred trees are planted (same genetic stock and same age) in each area. At the end of 5 years, the heights of seedlings in each plot were measured, and averaged for the plot.

# Schematic layout:

3	4	4	5	1
1	2	3	5	2
2	1	2	4	2
5	4	3	1	5
4	3	1	5	3

Five site preparation treatments are randomly applied to the 25 plots

- The most simple design:  
one fixed-effects factor,  
with random allocation of treatments to each experimental unit:



0.02 ha plot is the experimental unit

Factor: site preparation

Number of treatments:  $J=5$  site preparation methods (= 5 levels)

$i$  = a particular 0.02 ha area in treatment  $j$ , from 1 to 5

Response variable  $Y_{ij}$ : Plot average seedling height after 5 years

$n_T$  = the number of experimental units over all the treatments =  $\sum_{j=1}^5 n_j = 25$

$n_1 = n_2 = n_3 = n_4 = n_5 = 5$  experimental units measured for each treatment

# Data Organization and Preliminary Calculations

For easy calculations by hand, the data could be organized in a spreadsheet as:

NOTE: may not be the same number of observations for each treatment.

Obs: $i=1$ to $n_j$	Treatment, $j=1$ to $J$					
	1	2	3	...	$J$	
1	$y_{11}$	$y_{12}$	$y_{13}$	...	$y_{1J}$	
2	$y_{21}$	$y_{22}$	$y_{23}$	...	$y_{2J}$	
3	$y_{31}$	$y_{32}$	$y_{33}$	...	$y_{3J}$	
...	...	...	...	...	...	
$n$	$y_{n1}$	$y_{n2}$	$y_{n3}$	...	$y_{nJ}$	
Sum	$y_{\cdot 1}$	$y_{\cdot 2}$	$y_{\cdot 3}$	...	$y_{\cdot J}$	$y_{..}$
Averages	$\bar{y}_{\bullet 1}$	$\bar{y}_{\bullet 2}$	$\bar{y}_{\bullet 3}$		$\bar{y}_{\bullet J}$	$\bar{y}_{\bullet \bullet}$

$$y_{\bullet j} = \sum_{i=1}^{n_j} y_{ij} \quad \bar{y}_{\bullet j} = \frac{y_{\bullet j}}{n_j} \quad y_{\bullet \bullet} = \sum_{i=1}^J \sum_{j=1}^{n_j} y_{ij} \quad \bar{y}_{\bullet \bullet} = \frac{y_{\bullet \bullet}}{n_T}$$

## Plot Average Heights (m)

### Example

Observation	Treatments					Overall
	1	2	3	4	5	
1	4.6	4.9	4.0	3.4	4.3	
2	4.3	4.3	3.7	4.0	3.7	
3	3.7	4.0	3.4	3.0	3.7	
4	4.0	4.6	3.7	3.7	3.0	
5	4.0	4.3	3.0	3.4	3.4	
SUMS	20.600	22.100	17.800	17.500	18.100	96.100
Means	4.120	4.420	3.560	3.500	3.620	3.844
$n_j$	5	5	5	5	5	25

Example Calculations:

$$\bar{y}_{\bullet 1} = \frac{\sum_{i=1}^5 y_{ij}}{5} = (4.6 + 4.3 + 3.7 + 4.0 + 4.3) / 5 = 4.12$$

$$\bar{y}_{\bullet\bullet} = \frac{\sum_{j=1}^5 \sum_{i=1}^{n_j} y_{ij}}{\sum_{k=1}^5 n_j} = (20.6 + 22.1 + 17.8 + 17.5) / 25 = 3.844$$

# Using R

For entry into statistical programs like R, the data should be organized as:

Treatment	Obs:	Response
j=1 to J	i=1 to $n_j$	
1	1	$y_{11}$
1	2	$y_{21}$
1	3	$y_{31}$
...	...	...
1	$n_1$	$y_{(n1)1}$
2	1	$y_{12}$
2	2	$y_{22}$
2	3	$y_{32}$
...	...	...
2	$n_2$	$y_{(n2)2}$
...	...	...
J	1	$y_{1J}$
J	2	$y_{2J}$
J	3	$y_{3J}$
...	...	...
J	$n_J$	$y_{(nJ)3}$

For the example, we can put the data into an EXCEL file:

Treatment	Observation	AveHt
1	1	4.6
1	2	4.3
1	3	3.7
1	4	4.0
1	5	4.0
2	1	4.9
2	2	4.3
2	3	4.0
2	4	4.6
2	5	4.3
3	1	4.0
3	2	3.7
3	3	3.4
3	4	3.7
3	5	3.0
4	1	3.4
4	2	4.0
4	3	3.0
4	4	3.7
4	5	3.4
5	1	4.3
5	2	3.7
5	3	3.7
5	4	3.0
5	5	3.4

And save as comma-delimited text file:  
.CSV

# Bring in data from an outside comma delimited text file

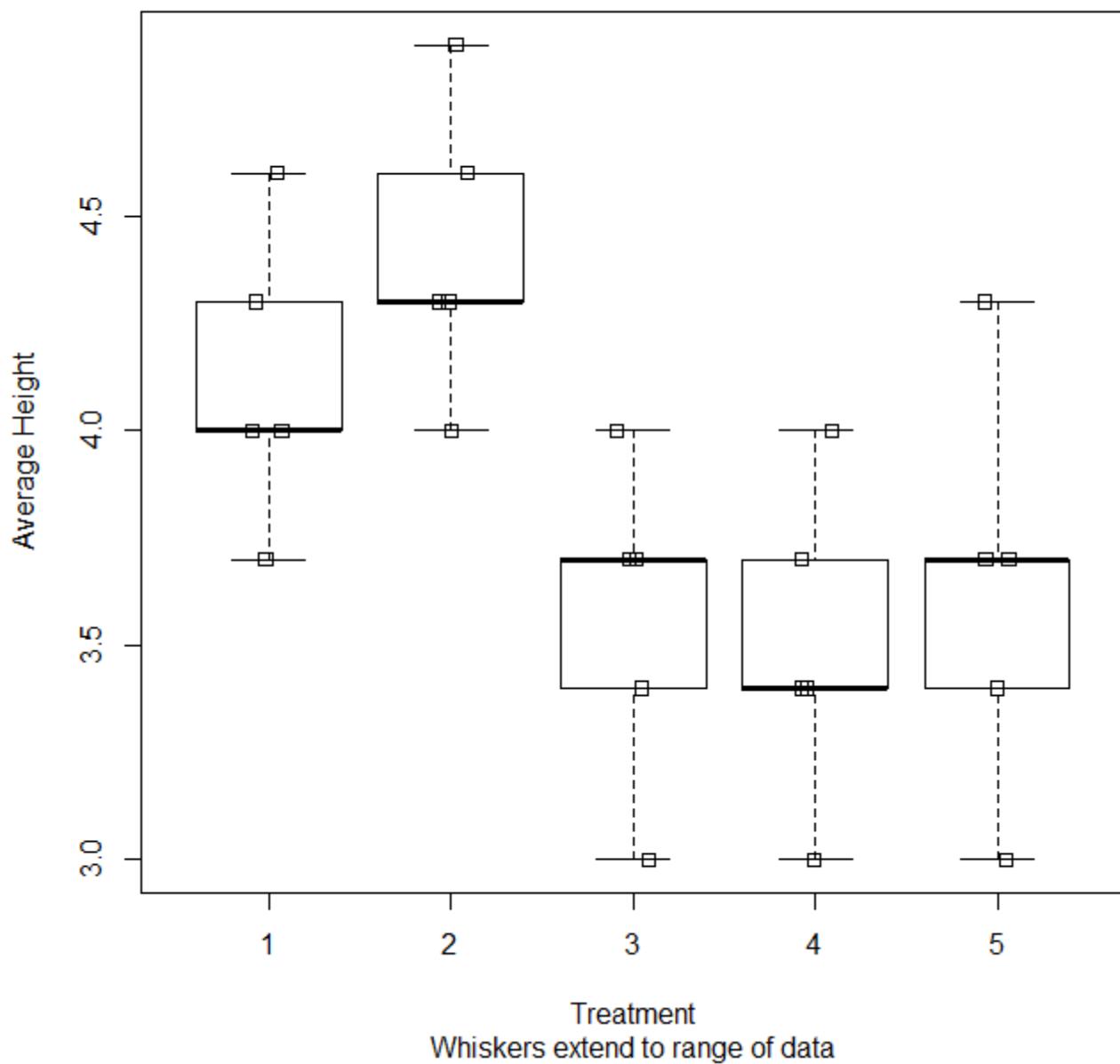
```
> fertdat <-  
read.csv("C:\\\\Users\\\\Rebecca\\\\Documents\\\\Cours\\\\201415\\\\TADE2015\\\\Fertilizationtr  
ial.csv", header=TRUE, as.is=TRUE)  
>  
> # check importation  
> names(fertdat)  
[1] "Treatment"    "Observation"   "AveHt"  
> dim(fertdat)  
[1] 25  3  
>
```

careful when reading in variables that are character based: R will often automatically translate these to “factors”.  
This default action is turned OFF when the as.is=TRUE argument is specified in the read.csv() function call.

begin by plotting the data using side-by-side box-plot and dot-plots to check for outliers

```
> boxplot( AveHt~Treatment, data=fertdat,  
+           range=0, #determines how far the plot whiskers extend. zero  
causes the whiskers to extend to the data extremes.  
+           main="Average Height by Treatment",  
+           sub='Whiskers extend to range of data',  
+           xlab="Treatment", ylab="Average Height")  
>  
> stripchart(AveHt~Treatment, data=fertdat, add=TRUE,  
+             vertical=TRUE, method="jitter", jitter=.1)
```

### Average Height by Treatment



# Calculate some summary statistics

```
> attach(fertdat)
> tapply(AveHt, Treatment, mean) #the mean for each treatment
  1   2   3   4   5
4.12 4.42 3.56 3.50 3.62
```

**In helperfunctions.R**

```
> library(plyr)
> report <- ddply(fertdat, "Treatment", sf.simple.summary, variable="AveHt",
  crd=TRUE)
```

	Treatment	n	nmis	mean	sd	se	lcl	ucl
1	1	5	0	4.12	0.3420526	0.1529706	3.695286	4.544714
2	2	5	0	4.42	0.3420526	0.1529706	3.995286	4.844714
3	3	5	0	3.56	0.3781534	0.1691153	3.090461	4.029539
4	4	5	0	3.50	0.3741657	0.1673320	3.035412	3.964588
5	5	5	0	3.62	0.4764452	0.2130728	3.028415	4.211585

$$= \frac{sd}{\sqrt{n}}$$

Because the data was collected using a CRD, it is sensible to compute individual standard errors and confidence intervals based on the raw data and not worrying about a more complex model

We can run this as either a regression or an ANOVA.

Regression: the categorical variable (treatment) is dummy coded  
= each category's intercept is compared to the reference group's intercept

ANOVA: the categorical variable (treatment) is effect coded  
= each category's mean is compared to the grand mean

# Regression approach: dummy coded

- have 5 treatments. Set up  $X_1, X_2, X_3$  and  $X_4$  as dummy variables:

Treatment	dummy variables			
	x1	x2	x3	x4
1	0	0	0	0
2	1	0	0	0
3	0	1	0	0
4	0	0	1	0
5	0	0	0	1

- Only need four dummy variables to represent the five treatments (levels-1 dummies).
- The dummy variables as a group represent the categorical variable (treatment).**
- Add the dummy variables to the equation – **this will alter the intercept**

Response:

Average height

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + b_4 x_{4i} + e_i$$

- For Treatment 1

$$y_i = b_0 + e_i$$

- For Treatment 2

$$y_i = \boxed{b_0 + b_1} + e_i$$

Difference between intercept for treatment 1  
and intercept for treatment 2

- For Treatment 3

$$y_i = \boxed{b_0 + b_2} + e_i$$

→ Different « intercept » for each  
treatment

- For Treatment 4

$$y_i = \boxed{b_0 + b_3} + e_i$$

- For Treatment 5

$$y_i = \boxed{b_0 + b_4} + e_i$$

→ Fit one equation using all data, but get different equations for different treatments

- In regression analysis, interactions between dummy variables and continuous variable(s) alter the slopes
- HERE: **There is NO continuous variable (only dummy variables), so no slope, only intercept**
- Since the intercept is defined as the mean value when all other predictors = 0, and there are no other predictors, the five intercepts are just means.

(All response values can be represented on the y-axis)

# Regression significant?

H0: Regression is not significant

H1: Regression is significant

Same as:

H0:  $\beta_1 = \beta_2 = \beta_3 = \dots = \beta_m = 0$  [all slopes are zero meaning no relationship with x's]

H1: not all slopes = 0 [some or all slopes are not equal to zero]

If H0 is true, then the equation is:

Average height  $y_i = \beta_0 + 0x_{1i} + 0x_{2i} + \dots + 0x_{mi} + \varepsilon_i$

$$y_i = \beta_0 + \varepsilon_i \quad \hat{y}_i = \beta_0 \quad \text{This is the Mean value for treatment 1!}$$

Where the x-variables have no influence over y; they do not help to better estimate y

→ The mean value for each treatment = the mean value for treatment 1

$$\begin{cases} m = 4 \\ n = 25 \end{cases}$$

Using an F test statistic:

$$F = \frac{SSreg/m}{SSE/(n - m - 1)} = \frac{MSreg}{MSE}$$

$$df_1 = 4 \quad df_2 = 20$$

SAME F-distribution as  $\frac{MS_{TR}}{MSE}$

- Under  $H_0$ , this follows an F distribution for a  $1-\alpha$  percentile with  $m$  and  $n-m-1$  degrees of freedom
- If the F for the fitted equation is larger than the F from the table, we reject  $H_0$  (not likely true). The regression is significant, in that one or more of the true slopes (the population slopes) are likely not equal to zero

→ The mean value for at least one treatment is different from the mean value for treatment 1

# Regression approach: Dummy coded

"factor"  
= limited number  
of possible values  
= "levels"  
= treatments

```
> fertdat$Treatment <- factor(fertdat$Treatment)
> modell<-lm(AveHt~Treatment, data=fertdat)
> summary(modell)
Call:
lm(formula = AveHt ~ Treatment, data = fertdat)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.62   -0.16  -0.10   0.18   0.68 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 4.1200    0.1725  23.883 3.56e-16 ***
Treatment2  0.3000    0.2440   1.230  0.2331    
Treatment3 -0.5600    0.2440  -2.295  0.0327 *  
Treatment4 -0.6200    0.2440  -2.541  0.0194 *  
Treatment5 -0.5000    0.2440  -2.049  0.0538 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

Residual standard error: 0.3857 on 20 degrees of freedom
Multiple R-squared:  0.5247,    Adjusted R-squared:  0.4297 
F-statistic: 5.52 on 4 and 20 DF,  p-value: 0.003674
```

# Regression approach: Dummy coded

```
> fertdat$Treatment <- factor(fertdat$Treatment)
> model1<-lm(AveHt~Treatment, data=fertdat)
> summary(model1)
Call:
lm(formula = AveHt ~ Treatment, data = fertdat)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.62	-0.16	-0.10	0.18	0.68

**mean for Treatment 1**

**Coefficients of dummies = The differences between the means of Treatments 2 to 5 and the reference mean (Treatment 1)**

Df of SSR = number of dummy variables (= m)  
 Df of SSE = number of experimental units – number of dummy variables – 1 (= n – m – 1)

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.1200	0.1725	23.883	3.56e-16	***
Treatment2	0.3000	0.2440	1.230	0.2331	*
Treatment3	-0.5600	0.2440	-2.295	0.0327	*
Treatment4	-0.6200	0.2440	-2.541	0.0194	.
Treatment5	-0.5000	0.2440	-2.049	0.0538	.
---					

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3857 on 20 degrees of freedom

Multiple R-squared: 0.5247, Adjusted R-squared: 0.4297

F-statistic: 5.52 on 4 and 20 DF, p-value: 0.003674

$$H_0: \beta_0 = 0$$

$$H_a: \beta_0 \neq 0$$

$$\text{t - statistic : } \frac{-0.3000 - 0}{0.2440} = 1.2331$$

if  $|t| > t_{critical}$ , we reject  $H_0$

this is the same as  $Pv < \alpha$

Are these differences significantly different from 0?

# ANOVA vs. Regression

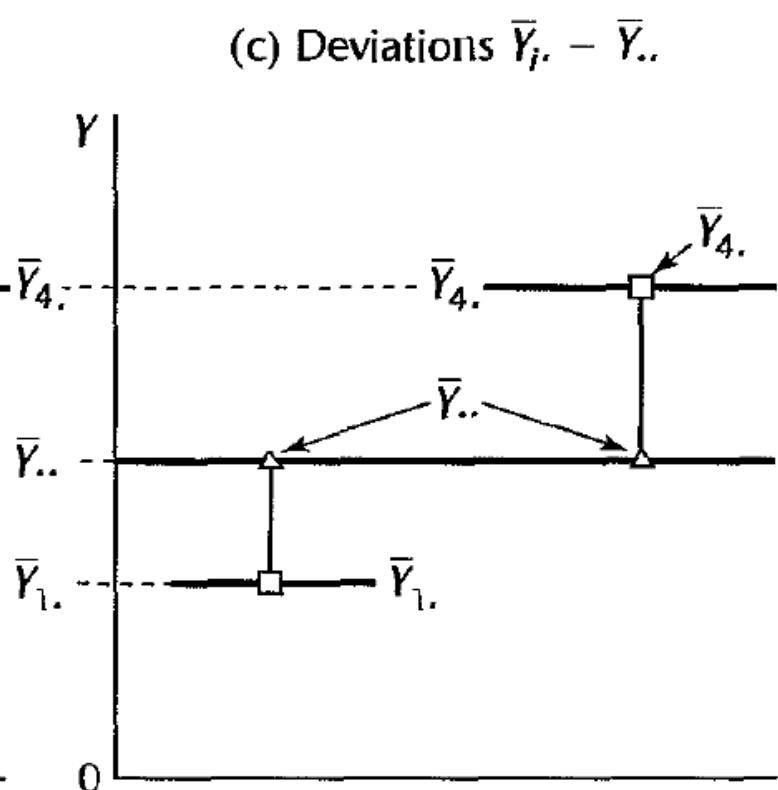
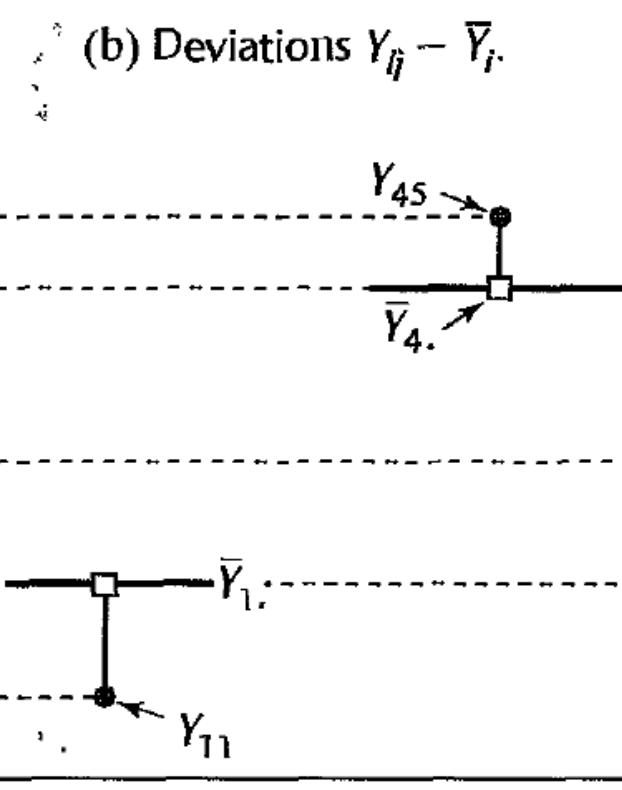
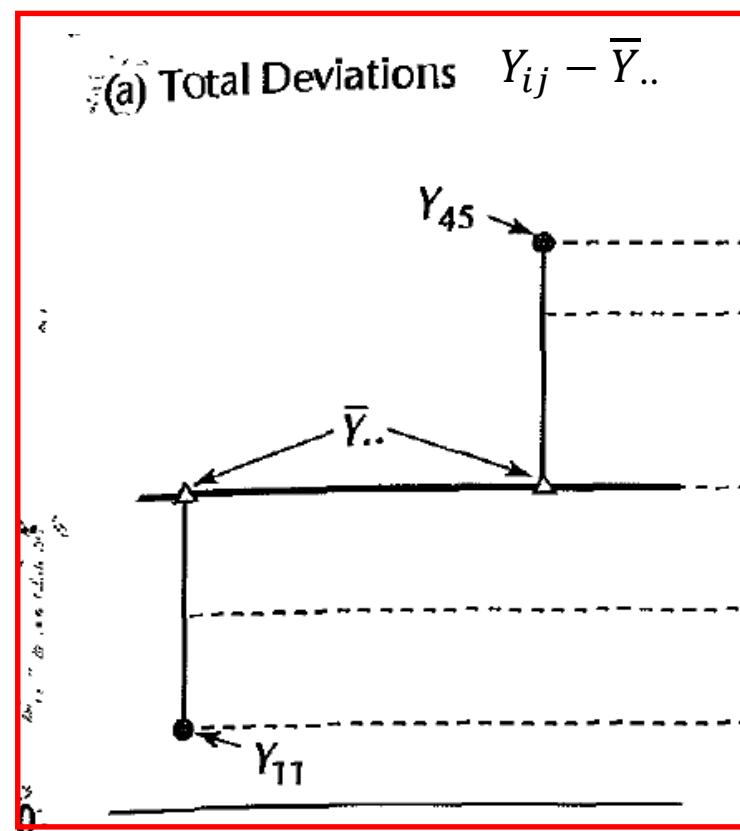
- Regression reports only one mean (as an intercept), and the differences between that one and all other means  
-> the p-values evaluate those specific comparisons
- ANOVA reports each mean and a p-value that says at least two are significantly different

# ANOVA

- regression model: ANOVA partitions the **total** sum of squares into **regression** sum of squares and **error** sum of squares
- ANOVA model: ANOVA partitions **total** sum of squares into **treatment** sum of squares and **error** sum of squares

The total variability of the  $Y_{ij}$  observations, not using any information about factor levels, is measured in terms of the total deviation of each observation, i.e., the deviation of  $Y_{ij}$  around the overall mean  $\bar{Y}_{..}$ :

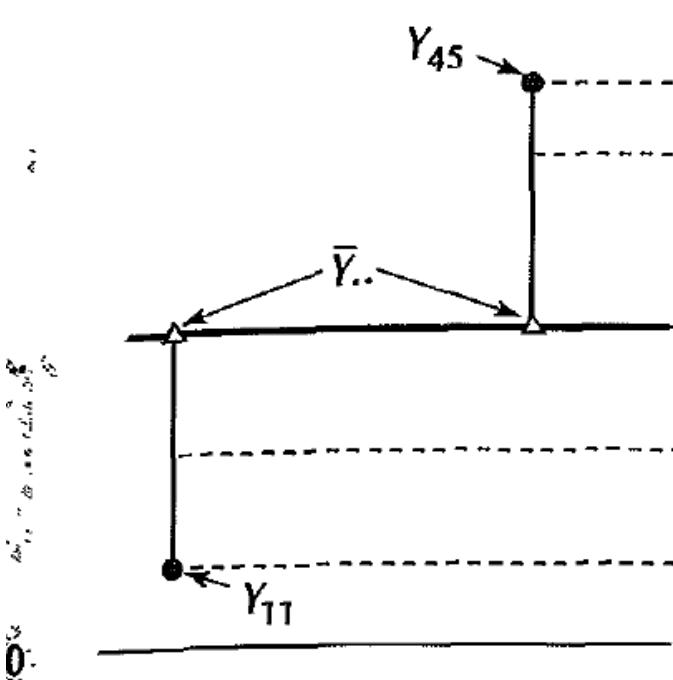
$$Y_{ij} - \bar{Y}_{..}$$



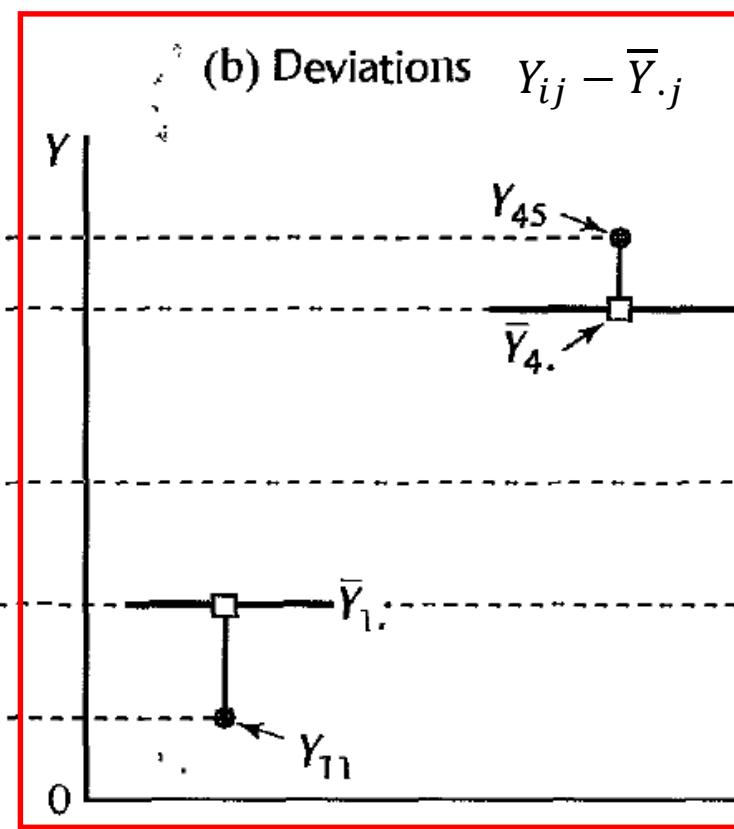
When we use information about the factor levels, the deviations reflecting the uncertainty remaining in the data are those of each observation  $Y_{ij}$  around it's estimated factor level mean  $\bar{Y}_{\cdot j}$ :

$$Y_{ij} - \bar{Y}_{\cdot j}$$

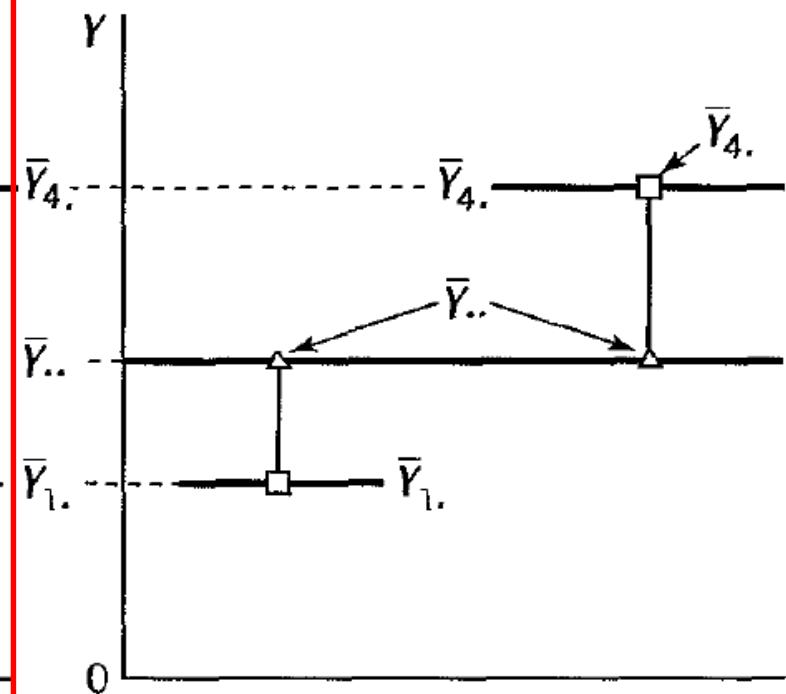
(a) Total Deviations  $Y_{ij} - \bar{Y}_{..}$



(b) Deviations  $Y_{ij} - \bar{Y}_{\cdot j}$

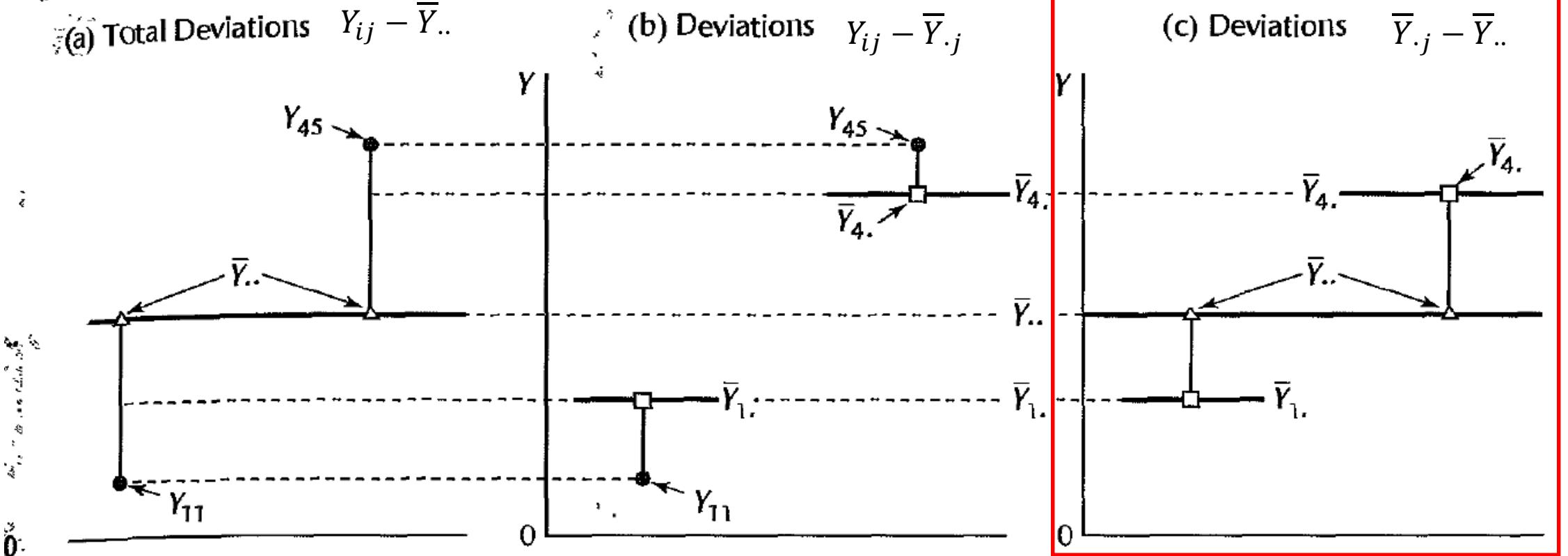


(c) Deviations  $\bar{Y}_{\cdot i} - \bar{Y}_{..}$



The difference between these deviations (a) and (b) reflects the difference between the estimated factor level mean and the overall mean:

$$(Y_{ij} - \bar{Y}_{..}) - (Y_{ij} - \bar{Y}_{.j}) = \bar{Y}_{.j} - \bar{Y}_{..}$$



We can thus decompose the total deviation  $Y_{ij} - \bar{Y}_{..}$  into two components:

$$(Y_{ij} - \bar{Y}_{..}) = (\bar{Y}_{.j} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{.j})$$

Total deviation	Deviation of estimated factor level mean around overall mean	Deviation around estimated factor level mean
-----------------	--	--

# Partitioning of SSy

When we square both sides and sum, we obtain

$$\sum_j \sum_i (Y_{ij} - \bar{Y}_{..})^2 = \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})^2$$

$$SSy = SS_{TR} + SSE$$

## SSE

- A measure of the random variation of the observations around the respective estimated factor level means
- If  $SSE = 0$ , the observations for any given factor level are all the same, and this holds for all factor levels
- The more the observations for each factor level differ among themselves, the larger will be  $SSE$

$SS_{TR}$

- A measure of the extent of differences between the estimated factor level means, based on the deviations of the estimated factor level means around the overall mean
- If all the estimated factor level means are the same, then  $SS_{TR} = 0$
- The more the estimated factor level means differ, the larger will be  $SS_{TR}$

# Breakdown of Degrees of Freedom

- $SS_y$  has  $n_T - 1$  degrees of freedom

$n_T$  deviations  $Y_{ij} - \bar{Y}_{..}$  and 1 df lost because they must sum to zero:

$$\sum \sum (Y_{ij} - \bar{Y}_{..}) = 0$$

- $SS_{TR}$  has  $J - 1$  degrees of freedom

$J$  estimated factor level mean deviations  $\bar{Y}_{.j} - \bar{Y}_{..}$  and 1 df lost because the weighted sum must equal zero:

$$\sum n_j (\bar{Y}_{.j} - \bar{Y}_{..}) = 0$$

Remember Regression:

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 0$$

# Breakdown of Degrees of Freedom

- $SSE$  has  $n_T - J$  degrees of freedom

Component  $SSE$  for  $j$ th factor level:

$$\sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{\cdot j})^2$$

This expression is equivalent of a total sum of squares considering only the  $j$ th factor level

Hence there are  $n_j - 1$  degrees of freedom associated with this sum of squares

Since  $SSE$  is a sum of component sum of squares, the df associated with  $SSE$  are:

$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_J - 1) = n_T - J$$

# Overview

always add up  
the same number  
of elements =  $n_T$ :  
 $\sum_{j=1}^J \sum_{i=1}^{n_j}$

- 1) Sum of squared differences between the observed values and the overall mean ( $SS_y$ ):

$$SS_y = \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{..})^2 \quad df = \sum_{j=1}^J n_j - 1$$

Also called, sum of squares total (same as in regression)

- 2) Sum of squared differences between the treatment means, and the grand mean, weighted by the number of experimental units in each treatment ( $SS_{TR}$ )

$$SS_{TR} = \sum_{j=1}^J \sum_{i=1}^{n_j} (\bar{y}_{\bullet j} - \bar{y}_{..})^2 = \sum_{j=1}^J n_j (\bar{y}_{\bullet j} - \bar{y}_{..})^2 \quad df = J - 1$$

- 3) Sum of squared differences between the observed values for each experimental unit and the treatment means ( $SSE$ )

$$SSE = \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet j})^2 \quad df = n_T - J$$

$$SS_y = SS_{TR} + SSE$$

Alternative formulae for the sums of squares that may be easier to calculate are:

$$SSy = \sum_{j=1}^J \sum_{i=1}^{n_j} y_{ij}^2 - \frac{\bar{y}_{\bullet\bullet}^2}{n_T}$$

$$SS_{TR} = \sum_{j=1}^J n_j \bar{y}_{\bullet j}^2 - \frac{\bar{y}_{\bullet\bullet}^2}{n_T}$$

$$SSE = SSy - SS_{TR}$$

This is an estimate of the real population  $\sigma^2$

## Example: differences from treatment means

Obs.	Treatments					Overall
	1	2	3	4	5	
1	0.480	0.480	0.440	-0.100	0.680	
2	0.180	-0.120	0.140	0.500	0.080	
3	-0.420	-0.420	-0.160	-0.500	0.080	
4	-0.120	0.180	0.140	0.200	-0.620	
5	-0.120	-0.120	-0.560	-0.100	-0.220	
SUMS	0.000	0.000	0.000	0.000	0.000	0.000
Sum of Squares Error	0.468	0.468	0.572	0.560	0.908	2.976
$n_j$	5	5	5	5	5	25
$s_{j1}^2$	0.117	0.117	0.143	0.140	0.227	

$$\begin{aligned} \text{SSE for treatment 1} &= \sum_{j=1}^5 (y_{1j} - \bar{y}_{\bullet 1})^2 \\ &= (4.6 - 4.1)^2 + (4.3 - 4.1)^2 + (3.7 - 4.1)^2 + (4.0 - 4.1)^2 + (4.0 - 4.1)^2 = 0.468 \end{aligned}$$

$$s_{j1}^2 = \frac{\text{SSE for treatment 1}}{n_1 - 1} = \frac{0.468}{5 - 1} = 0.117$$

$$\begin{aligned} \text{SSE} &= \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet j})^2 \\ &= \text{SSE for treatment 1} + \text{SSE for treatment 2} + \dots + \text{SSE for treatment 5} \\ &= 0.468 + 0.468 + 0.572 + 0.560 + 0.908 = 2.976 \end{aligned}$$

should be equal

$$MSE = \frac{2,976}{(25 - 5)} = 0,15$$

# Example: differences from grand mean

	Treatments					Overall
Obs.	1	2	3	4	5	
1	0.756	1.056	0.156	-0.444	0.456	
2	0.456	0.456	-0.144	0.156	-0.144	
3	-0.144	0.156	-0.444	-0.844	-0.144	
4	0.156	0.756	-0.144	-0.144	-0.844	
5	0.156	0.456	-0.844	-0.444	-0.444	
SUMS	1.380	2.880	-1.420	-1.720	-1.120	0.000
Sum of Squares Total	0.849	2.127	0.975	1.152	1.159	6.262
n <sub>j</sub>	5	5	5	5	5	25

$$SSy = \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet\bullet})^2$$

= SSy for treatment 1 + SSy for treatment 2 + ... + SSy for treatment 5

$$= 0.849 + 2.127 + 0.975 + 1.152 + 1.159 = 6.262$$

# Example: Difference between treatment means and grand mean

	Treatments					Overall
	1	2	3	4	5	
Mean	4.120	4.420	3.560	3.500	3.620	
Difference	0.276	0.576	-0.284	-0.344	-0.224	0.000
Sum of Squares Treatment	0.076	0.332	0.081	0.118	0.050	3.286
n <sub>j</sub>	5	5	5	5	5	25

$$\begin{aligned}
 SS_{TR} &= \sum_{j=1}^J n_j (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2 = (5 \times (4.120 - 3.844)^2) + (5 \times (4.420 - 3.844)^2) \\
 &\quad + (5 \times (3.560 - 3.844)^2) + (5 \times (3.500 - 3.844)^2) + (5 \times (3.620 - 3.844)^2) \\
 &= 3.286
 \end{aligned}$$

# Mean Squares

- Obtained by dividing sum of squares by associated degrees of freedom

$$MS_{TR} = \frac{SS_{TR}}{J - 1}$$

Variance between treatment means

$$MSE = \frac{SSE}{n_T - J}$$

Variance within treatments

Note: Mean Squares are not additive

# Expected Values of mean squares

= mean of the mean squares sampling distribution

= the average over repeated experiments

= « unbiased estimate of »

- Tells what is being estimated by the mean squares
- Statistical theory provides the following results:

$$E\{MSE\} = \sigma^2$$

(→ the variance of the error terms  $\varepsilon_{ij}$ )

$$E\{MS_{TR}\} = \sigma^2 + \phi_{TR} = \sigma^2 + \frac{\sum n_j (\mu_j - \mu.)^2}{J - 1}$$

« Second term »

$$\text{where } \mu. = \frac{\sum n_j \mu_j}{n_T}$$

(= the weighted mean)

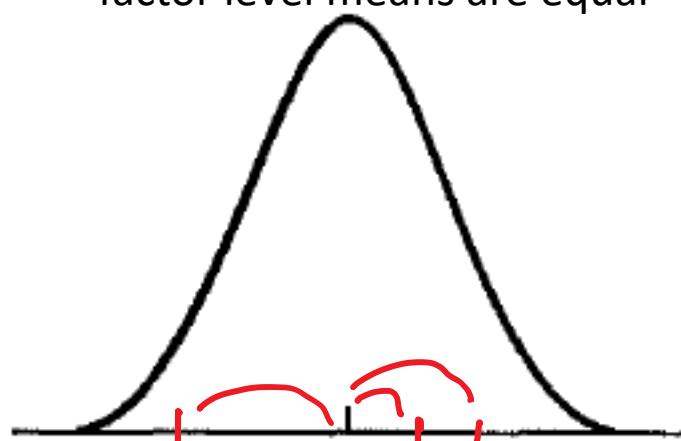
$\neq 0$  if at least 1 ( $= 2$ )  
factor level mean  
is different from  
the weighted mean

- $MS_{TR}$  will be big if the difference between the treatment means is big
- if the variation among the treatment means is small,  $MS_{TR}$  becomes equal to MSE

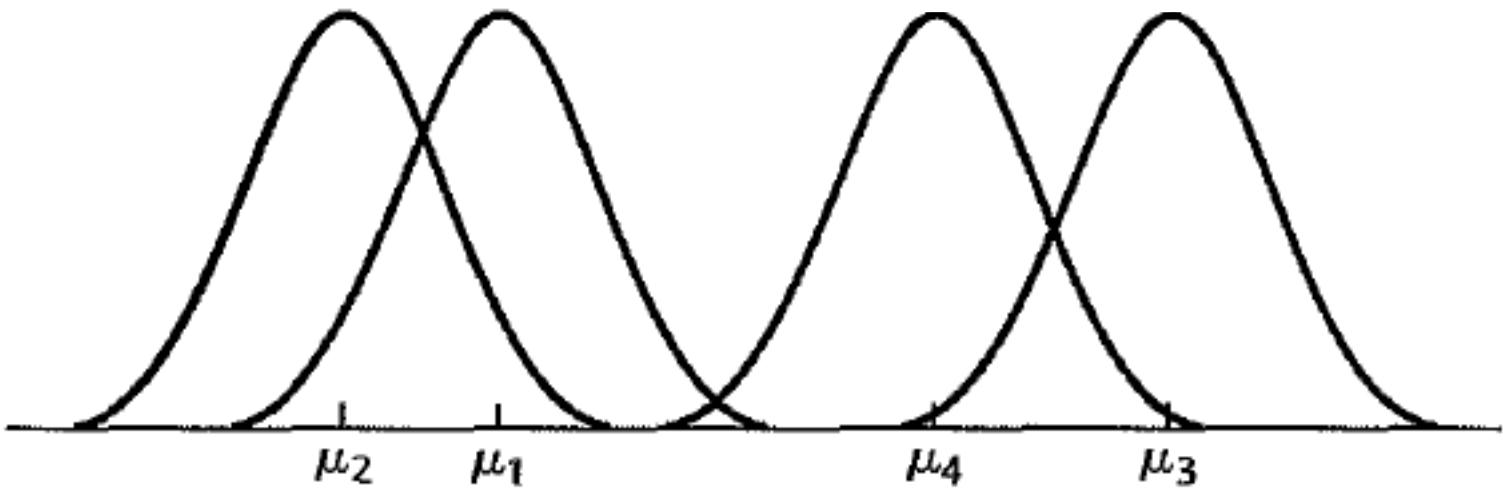
# Sampling distributions of $\bar{Y}_{\cdot j}$ for four treatments

Greek = population!

- (a)  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_c$   
factor level means are equal



- (b) factor level means  $\mu_j$  are not equal



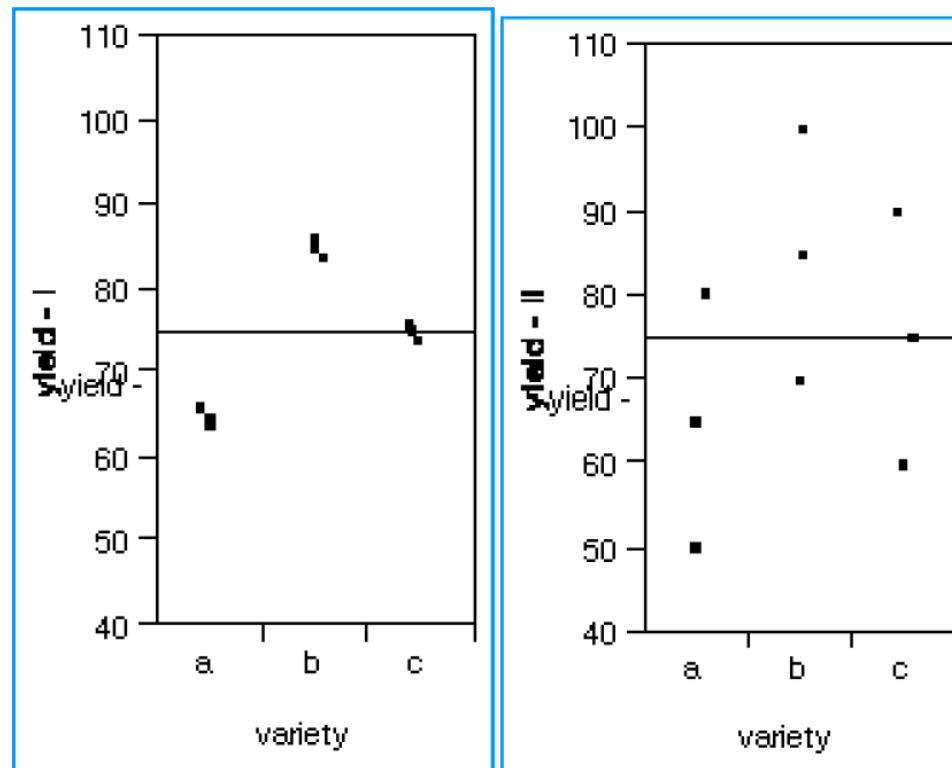
samples:  $\bar{Y}_1$   $\mu_c$   $\bar{Y}_2$   $\bar{Y}_3$

All factor level means are equal to the weighted mean  $\mu$ .  
 $\rightarrow E\{MS_{TR}\} = \sigma^2$  since the second term becomes zero  
 $\rightarrow MS_{TR}$  and  $MSE$  both estimate the error variance  $\sigma^2$

$\rightarrow E\{MS_{TR}\} > E\{MSE\}$  since the second term is positive

*different treatments  
= samples with different means  
coming from the same population*

- the idea behind the tests for equality of means (ANOVA) is to compare the relative variation among means ( $MS_{TR}$ ) to the variation within each group (MSE).



The first main question is: Are the treatment means different?

$$H_0: \mu_1 = \mu_2 = \dots = \mu_J$$

$H_1$ : not all the same

*OR:*

$$H_0: \tau_1 = \tau_2 = \dots = \tau_J = 0$$

Treatment effects are zero

$H_1$ : not all equal to 0

*OR:*

$$H_0: (\phi_{TR+} \sigma^2_\varepsilon) / \sigma^2_\varepsilon = 1$$

$\sigma^2_\varepsilon$  is the variance of the error terms;  
 $\phi_{TR}$  is the effect of the fixed treatments

$$H_1: (\phi_{TR+} \sigma^2_\varepsilon) / \sigma^2_\varepsilon > 1$$

If the treatment does not account for any of the variance in the response variable, then treatment effects are likely all = 0, and all the treatment means are likely all the same

# F test for Equality of factor level means

- The test statistic for choosing between the alternatives is:

$$F = \frac{SS_{TR} / (J - 1)}{SSE / \sum_{j=1}^J (n_j - 1)} = \frac{SS_{TR} / (J - 1)}{SSE / (n_T - J)} = \frac{MS_{TR}}{MSE}$$

Note:  $MS_{TR}$  plays the role of  $MSR$  for a regression model

- $MS_{TR}$  tends to exceed  $MSE$  when  $H_a$  holds  
→ large values of  $F$  support  $H_a$
- Both  $MS_{TR}$  and  $MSE$  have the same expected value when  $H_0$  holds  
→ values of  $F$  near 1 support  $H_0$

→ One-sided Upper tail test

# Distribution of $F$

- When  $H_0$  holds (all treatment means are equal, each response has the same expected value),

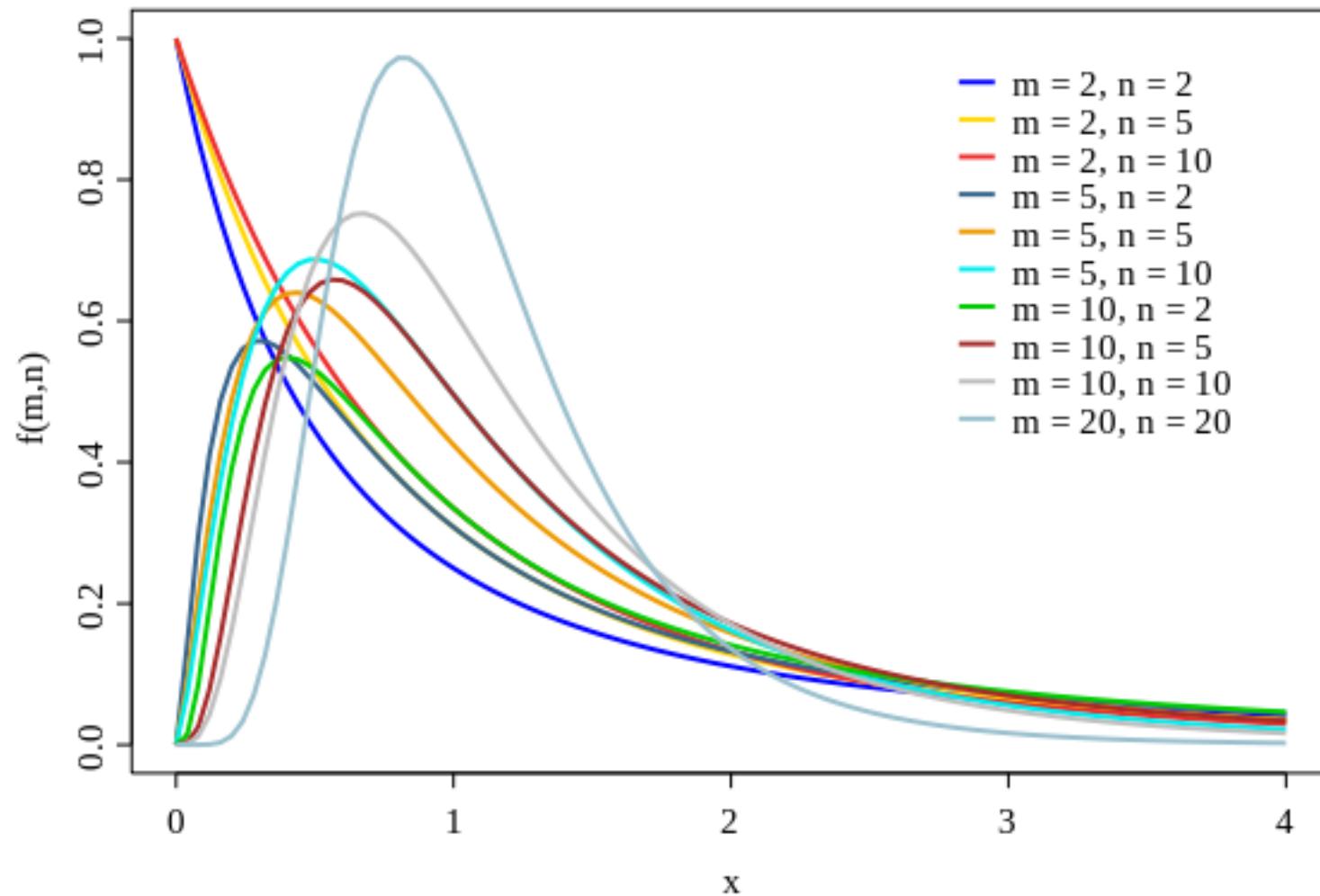
$F$  is distributed as  $F(J-1, n_T-J)$

- When  $H_a$  holds (not all treatment means are equal,  
 $F$  does not follow the  $F$  distribution  
( $F$  follows the *non-central F distribution*)

# F-distribution

- Is the ratio of 2 variables that each have a  $\chi^2$ -distribution eg.  
The ratio of 2 sample variances for variables that are each normally distributed.
  - Is a family of curves based on the d.f. of numerator and d.f. of denominator
  - Never negative
  - **Mean is approximately 1**
  - Need the percentile, and two degrees of freedom (one for the numerator and one for the denominator)
- Each alpha value involves a separate table

# F-distribution



# Errors of hypothesis testing

	H0 True	H0 False
Accept	$1-\alpha$	$\beta$ (Type II error)
Reject	$\alpha$ (Type I error)	$1-\beta$

- **Type I Error:** Reject H0 when it was true.  
Probability of this happening is  $\alpha$
- **Type II Error:** Accept H0 when it is false.  
Probability of this happening is  $\beta$
- **Power of the test:** Reject H0 when it is false.  
Probability of this is  $1-\beta$

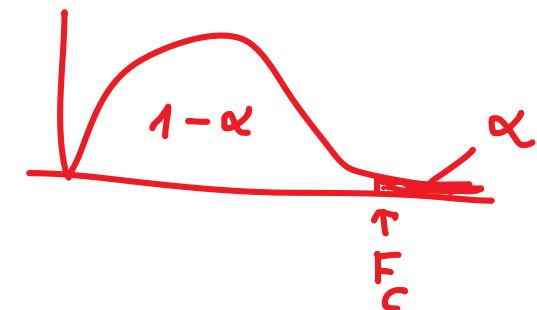
# Construction of Decision Rule

- This controls the risk of making a Type I error (*seeing differences where there are none*)

If  $F \leq F_{(J-1, n_T - J, 1-\alpha)}$ , conclude  $H_0$

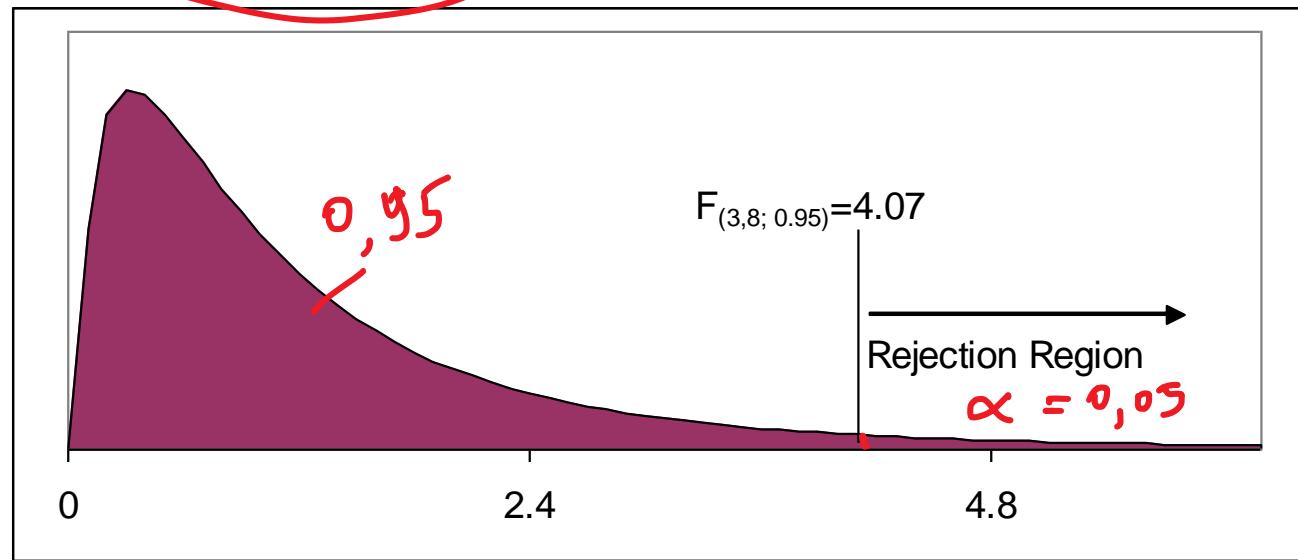
*a quantile , critical F-value*

If  $F > F_{(J-1, n_T - J, 1-\alpha)}$ , conclude  $H_a$  : there is a difference between the treatment means



Where  $F_{(J-1, n_T - J, 1-\alpha)}$  is the  $1 - \alpha$  quantile of the appropriate  $F$  distribution

- For example, if we have 4 treatments, and 12 experimental units, and we want  $\alpha=0.05$ :



- If the calculated F is larger than 4.07, we reject  $H_0$ : The treatments means are likely different, unless a 5% error has occurred.
- OR: We take our calculated F value from our experiment and plot it on this F curve. Then, find the area to the right of this value (p-value). We reject a hypothesis if the probability value (p-value) for the test is less than the specified significance level.

# Analysis of Variance table

Source	df	SS	MS	F	p-value
Treatment	$J-1$	$SS_{TR}$	$MS_{TR} =$ $SS_{TR}/(J-1)$	$F =$ $MS_{TR}/MSE$	Prob F > $F_{(J-1), (nT-J), (1-\alpha)}$
Error	$n_T - J$	$SSE$	$MSE =$ $SSE/(n_T - J)$		
Total	$n_T - 1$	$SS_y$			

# Example:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_1: \text{not all equal}$$

Analysis of Variance (ANOVA) Table:

Source	df	SS	MS	F	p-value
Treatment	5-1=4	3.286	0.821	5.51	0.004
Error	25-5=20	2.976	0.149		
Total	25-1=24	6.262			

Since the p-value is smaller than ( $\alpha=0.05$ ), we reject  $H_0$  and conclude that there is a difference in the treatment means.

BUT this is only a good test if the assumptions of analysis of variance have been met. Need to check these first (as with regression analysis).

# To calculate $F_{critical}$ and p-value for F in R

```
> #give quantile value based on probability  
> qf (0.95, 4, 20) #this gives the critical value for F  
[1] 2.866081  
>  
> #give probability based on quantile value  
> pf (5.51, 4, 20) #this gives the area to the left of  
this percentile value  
[1] 0.9962928  
> 1- pf (5.51, 4, 20) #this gives the area to the right of  
this percentile value  
[1] 0.00370715
```

# ANOVA approach

- Make sure the treatment is a "factor!" = limited number of values
- Use lm() and anova() to get the analysis of variance table and F test for treatments

```
> fertdat$Treatment <- factor(fertdat$Treatment)
> model1<-lm(AveHt~Treatment, data=fertdat)
> anova(model1) #anova approach
```

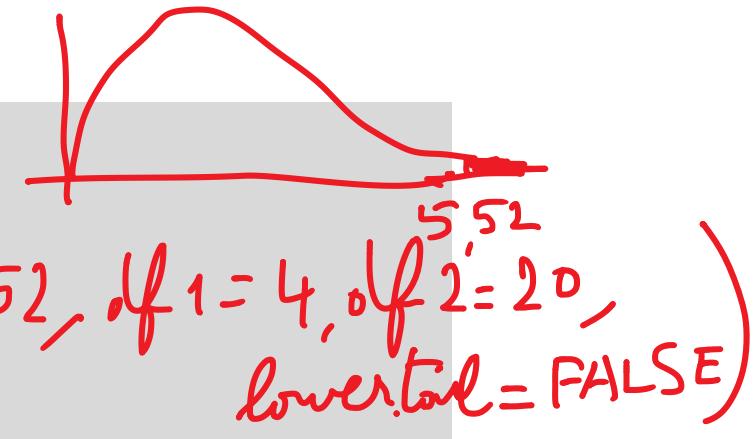
Analysis of Variance Table

Response: AveHt

	Df	Sum Sq	Mean Sq	F value	Pr (>F)
Treatment	4	3.2856	0.8214	5.5202	0.003674 **
Residuals	20	2.9760	0.1488		
	--				

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

"levels"



F-value and p-value are the same as for regression approach  
It's essentially the same thing!!

J-1

n\_T

# Regression approach: Dummy coded

```
> summary(modell)
Call:
lm(formula = AveHt ~ Treatment, data = fertdat)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.62   -0.16  -0.10   0.18   0.68 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 4.1200    0.1725  23.883 3.56e-16 ***
Treatment2  0.3000    0.2440   1.230  0.2331    
Treatment3 -0.5600    0.2440  -2.295  0.0327 *  
Treatment4 -0.6200    0.2440  -2.541  0.0194 *  
Treatment5 -0.5000    0.2440  -2.049  0.0538 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

Residual standard error: 0.3857 on 20 degrees of freedom
Multiple R-squared:  0.5247,    Adjusted R-squared:  0.4297 
F-statistic: 5.52 on 4 and 20 DF,  p-value: 0.003674
```

$$d.f. = 4$$

$$F = \frac{MSR}{MSE} = \frac{MSTR}{MSE}$$

$$d.f. = 20$$

Before interpreting the ANOVA table, we must check assumptions regarding the error term

For the estimated means for this experiment to be unbiased estimates of the means in the population, and the MSE to be an unbiased estimate of the variance within each experimental unit, the following assumptions must be met:

1. Observations are independent – not related in time nor in space [independent data]
2. There is normal distribution of the y-values [or the error terms] around each treatment mean [normally distributed]
3. The variances of the y's around each treatment mean [or the error terms] are the same (homogeneous) for all treatment means [equal variance]

# Similar to regression:

- a normal probability plot for the error terms can be used to check the assumption of normality, and
- a residual plot can be used to visually check the assumption of equal variance.

OR, these can be tested using (1) normality tests (as with regression); (2) Bartlett's test for equal variances (for more than one factor or for other designs with blocking, etc. this becomes difficult).

# Transformations to meet assumptions

Similar to regression:

- logarithmic transformations can be used to equalize variances
- arcsine transformation can be used to transform proportions into normally distributed variables
- rank transformation can be used when data are not normally distributed and other transformations do not “work” [nonparametric analysis of variance using ranks]

# Process

Unlike regression you must transform the y-variable  
(x variables are class variables (just names), we cannot change them)

- do your analysis with the measured response variable
- if assumptions of the error term are not met, transform the y-variable
- do the analysis again and check the assumptions; if not met, try another transformation

# Check assumptions:

## (1) Homoscedasticity

= equal variance across treatments

- Residuals plot

```
> names(model1)
[1] "coefficients"   "residuals"      "effects"       "rank"
[5] "fitted.values"  "assign"        "qr"           "df.residual"
[9] "contrasts"      "xlevels"       "call"          "terms"
[13] "model"
> model1$fitted.values
   1    2    3    4    5    6    7    8    9    10   11   12   13   14   15   16
4.12 4.12 4.12 4.12 4.12 4.42 4.42 4.42 4.42 4.42 3.56 3.56 3.56 3.56 3.56 3.50
   17   18   19   20   21   22   23   24   25
3.50 3.50 3.50 3.50 3.62 3.62 3.62 3.62 3.62
> model1$residuals
   1    2    3    4    5    6    7    8    9    10   11   12   13
0.48 0.18 -0.42 -0.12 -0.12 0.48 -0.12 -0.42 0.18 -0.12 0.44 0.14 -0.16
   14   15   16   17   18   19   20   21   22   23   24   25
0.14 -0.56 -0.10 0.50 -0.50 0.20 -0.10 0.68 0.08 0.08 -0.62 -0.22
> yhat<-model1$fitted.values
> resid<-resid(model1)
> plot(resid~yhat) # residual plot
```

Residual here is the observed value - mean for the treatment

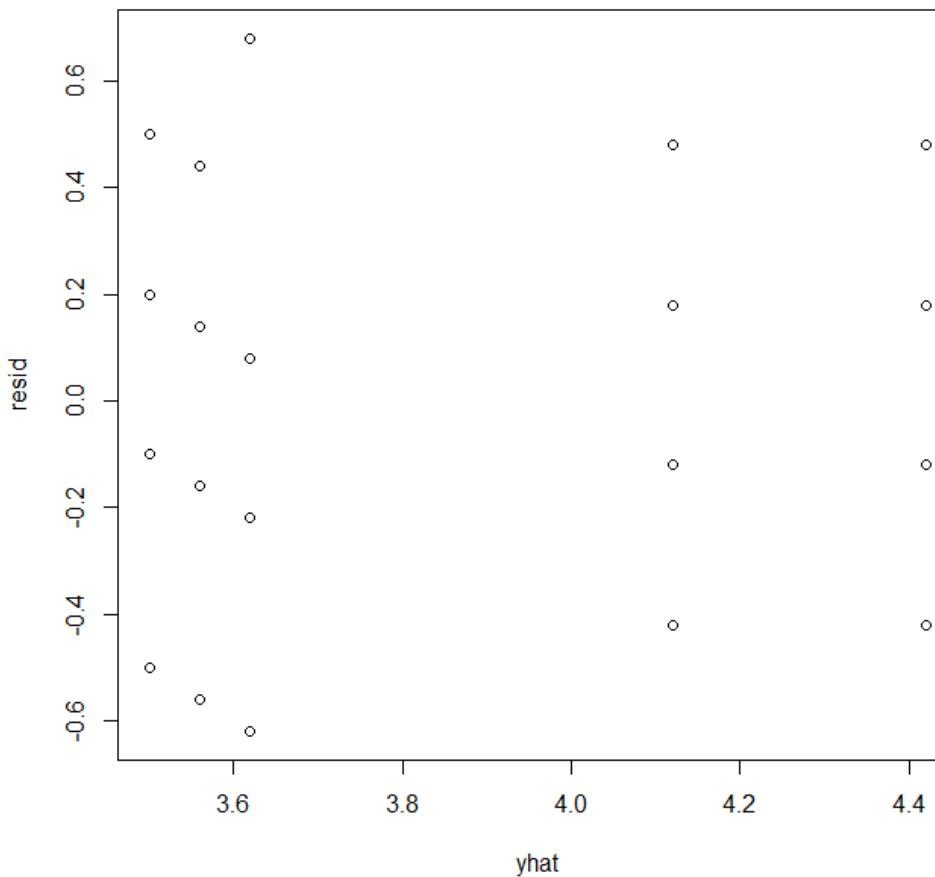
Check assumptions:

(1) Homoscedasticity

= equal variance across treatments

(estimated by MSE as 0.149 on our ANOVA table)

- Residuals plot



# Check assumptions:

## (1) Homoscedasticity

- Bartlett's test for Homogeneity of AveHt Variance

$$H_0: \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5$$

```
> bartlett.test(AveHt~as.factor(Treatment), data=fertdat)
```

Bartlett test of homogeneity of variances

```
data: AveHt by as.factor(Treatment)  
Bartlett's K-squared = 0.579, df = 4, p-value = 0.9654
```

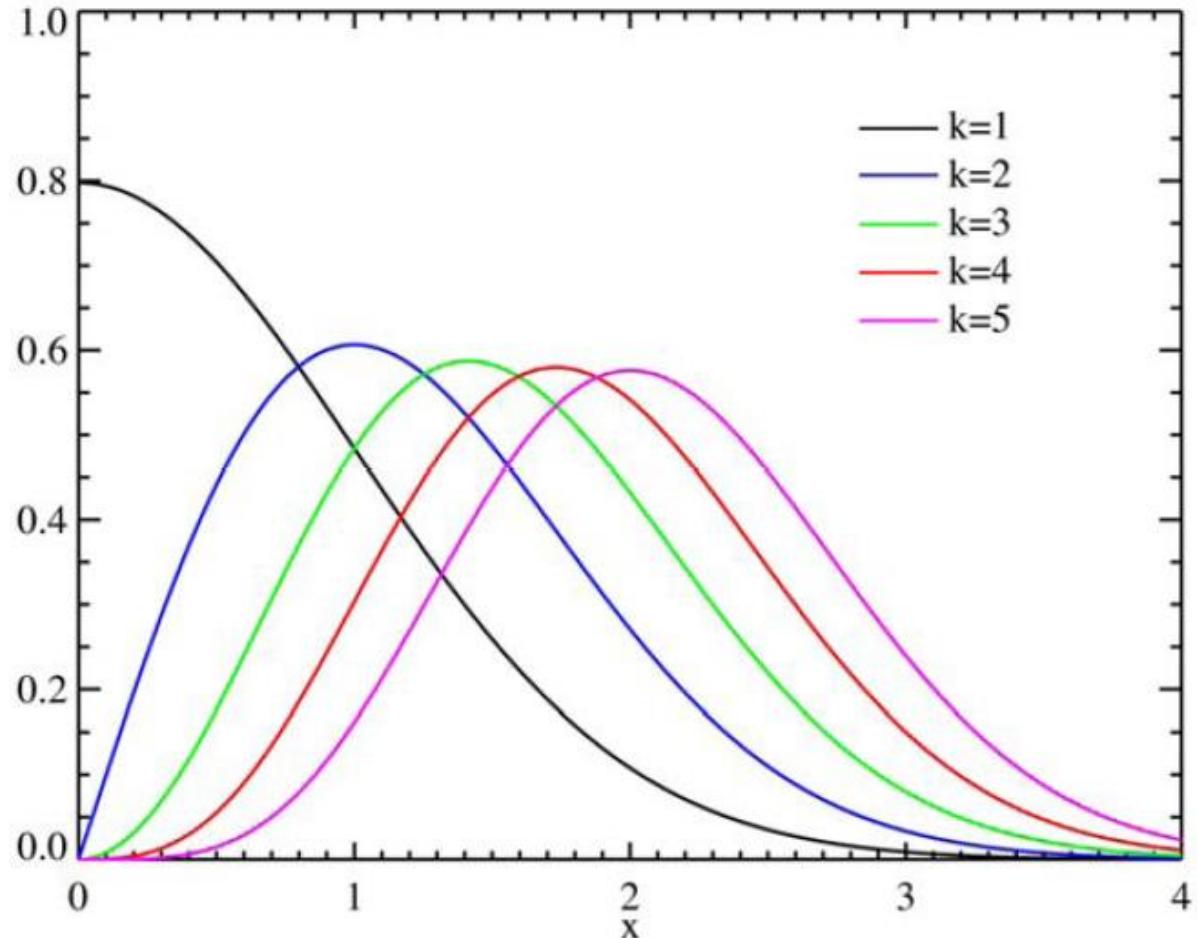
Chi-Square

Pr > Chi-Square (upper tail)



This probability is much higher than 0.05: we do not reject the null-hypothesis. The variances are equal

# $\chi^2$ -distribution



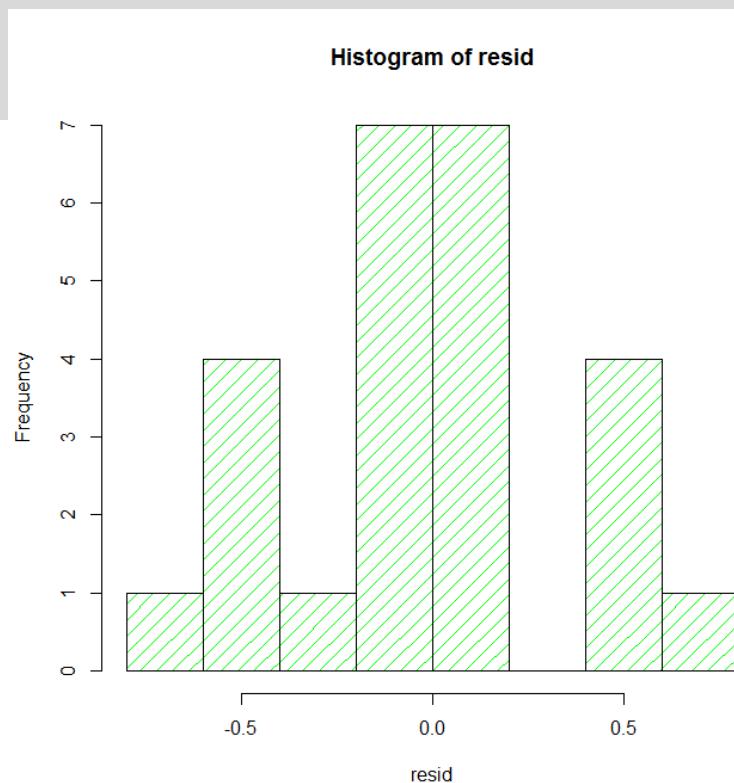
```
> #give probability based on quantile value: pchisq (q, df)
> pchisq(0.5790,4,lower.tail=FALSE) #we need the upper tail
[1] 0.9653684
```

# Check assumptions:

## (2) normality of residuals

- Residual here is the observed value - mean for the treatment

```
> resid<-resid(model1)
> hist(resid, breaks =6 , density=10,col="green", border="black") # draws a
histogram
```



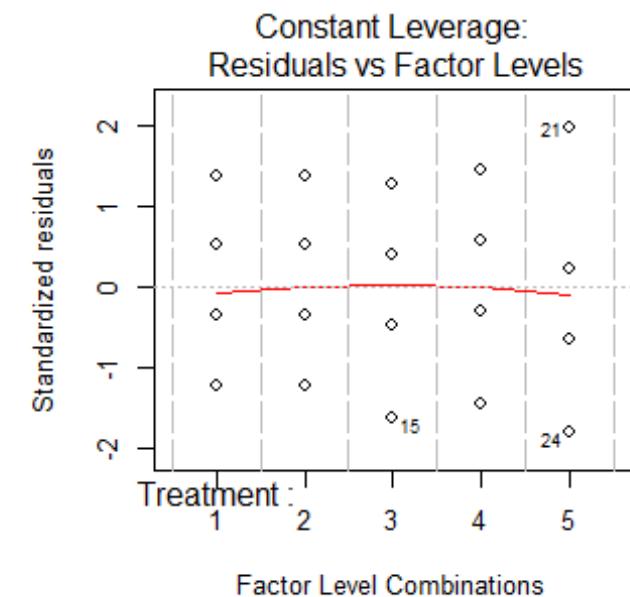
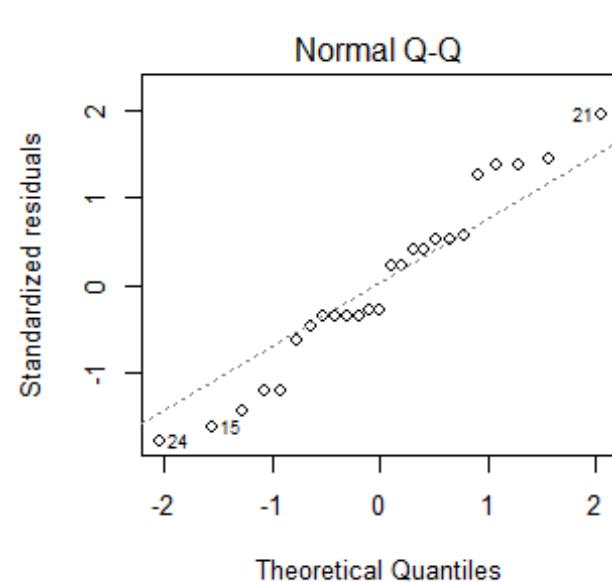
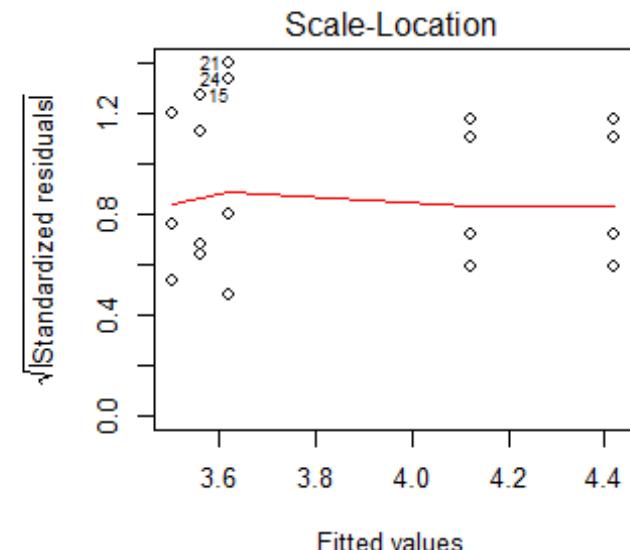
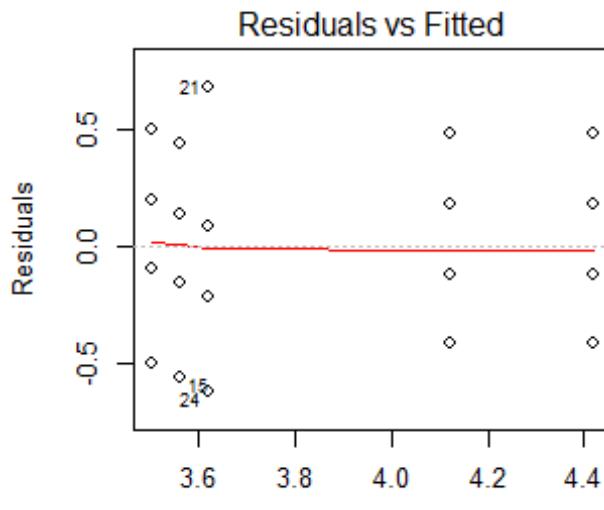
# Check the assumptions of the ANOVA model using Base R graphics

```
> layout(matrix(1:4, 2, 2))  
> plot(model1)  
> layout(1)
```

Standardized residuals are calculated by:

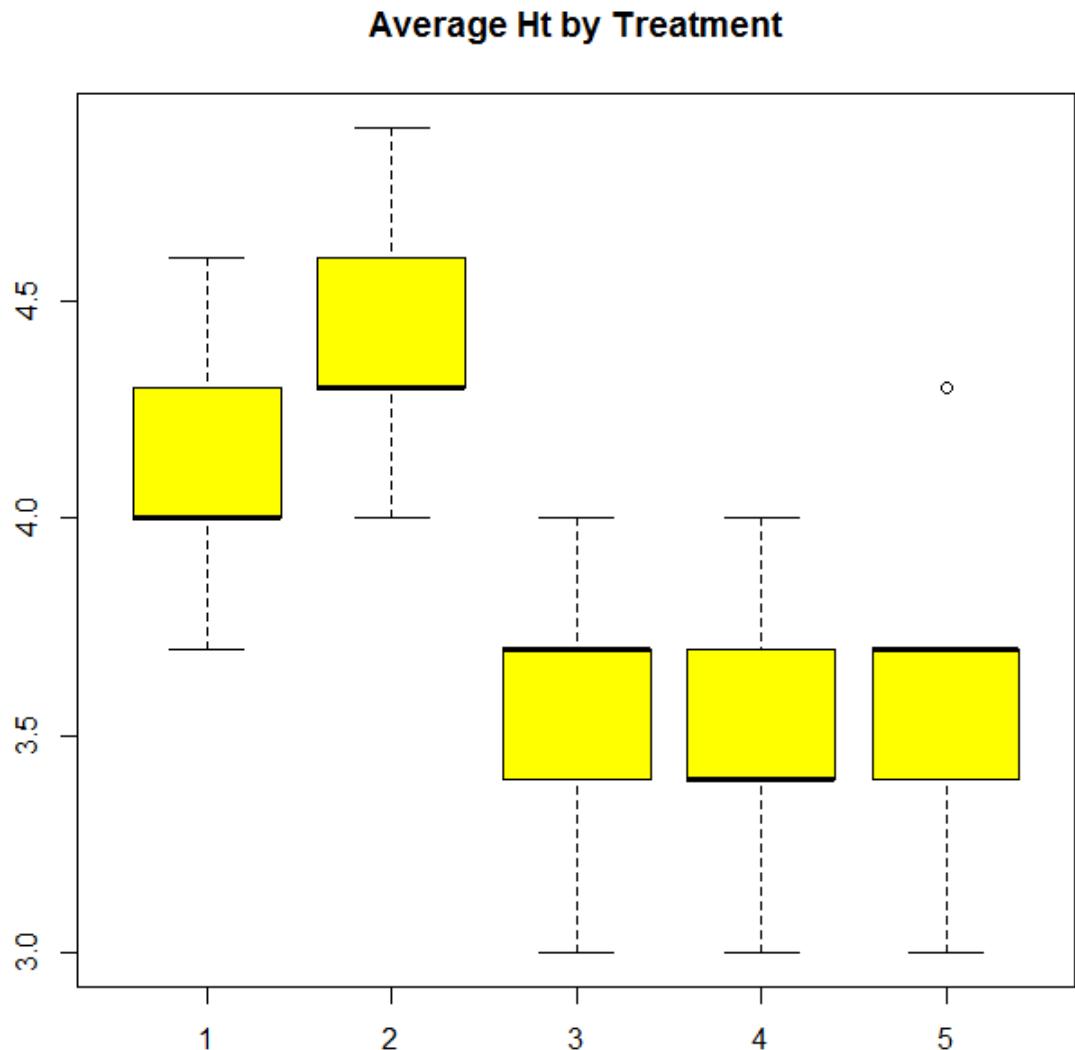
$$e_i(\text{standardized}) = \frac{e_i - 0}{\sqrt{MSE}}$$

Compare these to z-values for a standard normal distribution with a mean of zero and a variance of 1 ( $z(0,1)$ )



If treatment means vary, then test which pairs differ

```
> attach(fertdat)  
>  
boxplot(AveHt~Treatment,col='yellow',  
main="Average Ht by Treatment")  
> detach(fertdat)
```



# Multiple comparisons following ANOVA: Why is there a problem?

- If there are differences among means detected, which means differ?
  - To make specific comparisons, we could try using individual t-tests.
  - In the first instance above, we are comparing 5 means. Thus there are 10 pairs of means to test
  - This number increases rapidly. For example, amongst the 6 means there are 15 possible t-tests. With 20 means, there are 190 possible comparisons.
  - When you are performing all these t-tests, chances are that one of them will produce a significant p-value, even if there are no systematic differences amongst the groups just by random chance.
- must adjust the alpha level used by dividing by the number of pairs.

# Hundreds of multiple comparison procedures exist

Scheffe's multiple comparisons

Bonferroni's adjustments

Tukey adjustments

...

- Which one is best is a difficult question to answer.
- The key idea behind all of the multiple comparison procedures is to control the probability that you will make at least one Type I error somewhere among the entire set of comparisons examined.
- Don't get hung up on this topic - understand why these tests are needed, and take care if you are doing experiments with many, many treatments

Example: given that there is a difference among treatment means, which pairs of means differ?

t-test for pairs of means:

determine the number of pairs possible :

$$\binom{n}{k} = \frac{n!}{(n - k)! k!}$$
 where  
 $n$  = number of treatments  
 $k$  = number to compare (in t-test  $k=2$ )

$$\binom{5}{2} = \frac{5!}{(5 - 2)! 2!} = 10 \text{ possible pairs of means}$$

- In R:

```
> choose(5, 2)
[1] 10
```

# Comparing Treatments 2 (largest estimated mean) versus 4 (smallest estimated mean):

$$H_0: \mu_2 - \mu_4 = 0 \quad \text{OR} \quad H_0: \mu_2 = \mu_4$$

$$H_1: \mu_2 - \mu_4 \neq 0$$

$$t = \frac{(\bar{y}_{\bullet 2} - \bar{y}_{\bullet 4}) - 0}{\sqrt{MSE \left( \frac{1}{n_2} + \frac{1}{n_4} \right)}}$$

From the ANOVA table

Standard error of the mean  
 $s_{(\bar{y}_{\bullet 2} - \bar{y}_{\bullet 4})}$

$$t = \frac{(4.4 - 3.5)}{\sqrt{0.149 \times \left( \frac{1}{5} + \frac{1}{5} \right)}} = 3.686$$

**IMPORTANT FORMULA!**

ANOVA post-hoc

≠ like normal t-test  
uses info of other treatments

→ Better estimation of  $\sigma^2$   
→ More degrees of freedom

This is an estimate of the real population  $\sigma^2$

## Example: differences from treatment means

Obs.	Treatments					Overall
	1	2	3	4	5	
1	0.480	0.480	0.440	-0.100	0.680	
2	0.180	-0.120	0.140	0.500	0.080	
3	-0.420	-0.420	-0.160	-0.500	0.080	
4	-0.120	0.180	0.140	0.200	-0.620	
5	-0.120	-0.120	-0.560	-0.100	-0.220	
SUMS	0.000	0.000	0.000	0.000	0.000	0.000
Sum of Squares Error	0.468	0.468	0.572	0.560	0.908	2.976
$n_i$	5	5	5	5	5	25
$s_{ij}^2$	0.117	0.117	0.143	0.140	0.227	

$$\begin{aligned} \text{SSE for treatment 1} &= \sum_{j=1}^5 (y_{1j} - \bar{y}_{\bullet 1})^2 \\ &= (4.6 - 4.1)^2 + (4.3 - 4.1)^2 + (3.7 - 4.1)^2 + (4.0 - 4.1)^2 + (4.0 - 4.1)^2 = 0.468 \end{aligned}$$

$$s_{11}^2 = \frac{\text{SSE for treatment 1}}{n_1 - 1} = \frac{0.468}{5 - 1} = 0.117$$

$$\begin{aligned} \text{SSE} &= \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet j})^2 \\ &= \text{SSE for treatment 1} + \text{SSE for treatment 2} + \dots + \text{SSE for treatment 5} \\ &= 0.468 + 0.468 + 0.572 + 0.560 + 0.908 = 2.976 \end{aligned}$$

should be equal

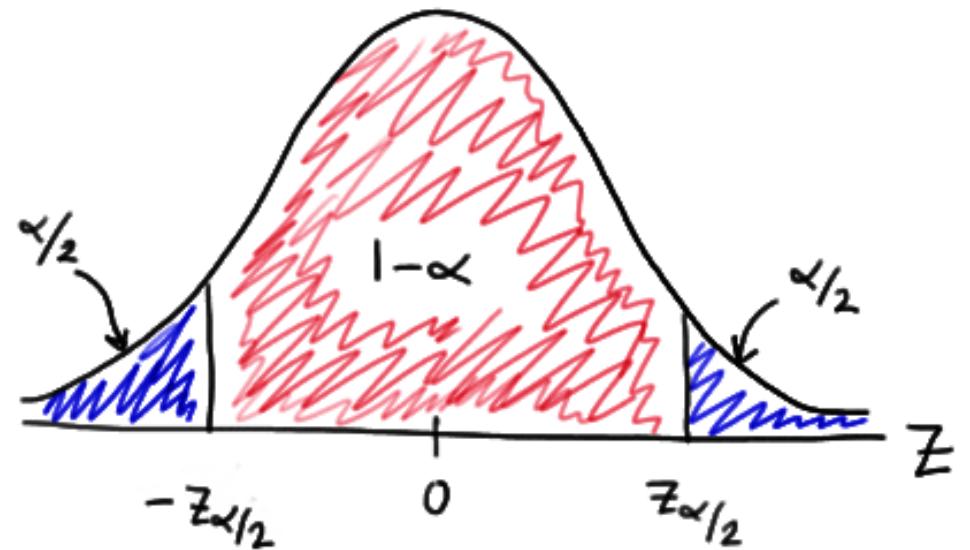
$$\text{MSE} = \frac{2,976}{(25 - 5)} = 0,15$$

d.f. MSE

For each t-test: Bonferroni adjustment:  $\alpha_{\text{adjustment}} = \frac{\alpha \text{ for } F\text{-test}}{\text{Number of } T\text{-tests}}$

Under  $H_0$ : This follows:

$$t_{\text{critical}} = t_{1 - \frac{\alpha}{2}, n_T - J}$$



Using alpha=0.005 ( $0.05/10=0.005$ ), for 5 treatments and 25 observations, the t-value is 3.153. Result?

```
> #give quantile value based on probability  
> qt(0.9975, 20)  
[1] 3.153401
```

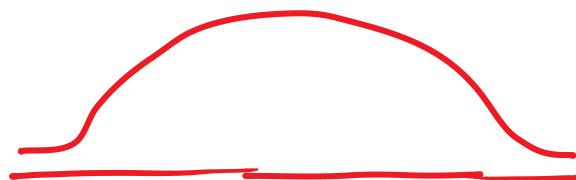
## T-test comparing two means

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{s_{\bar{y}_1 - \bar{y}_2}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Under H0:

this will follow a t-distribution with  $df = (n_1 + n_2 - 2)$

8



## ANOVA post-hoc test comparing 5 means

$$t = \frac{(\bar{y}_{\bullet 2} - \bar{y}_{\bullet 4}) - 0}{\sqrt{MSE \left( \frac{1}{n_2} + \frac{1}{n_4} \right)}}$$

MSE takes data of all treatments into account

→ BETTER estimation of  $\sigma^2$

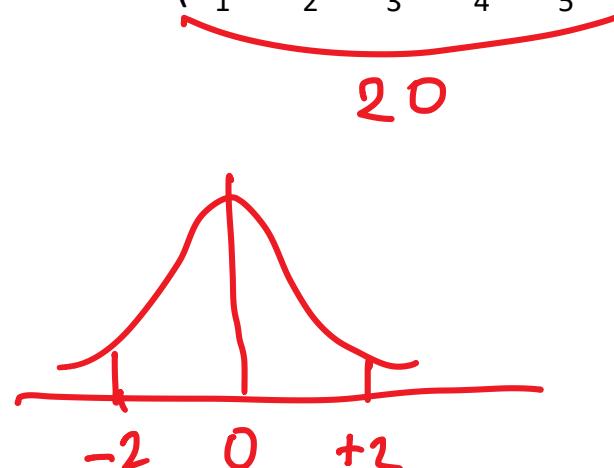
Under H0:

this will follow a t-distribution with  $df = (n_1 + n_2 + n_3 + n_4 + n_5 - 5)$

→ MORE degrees of freedom

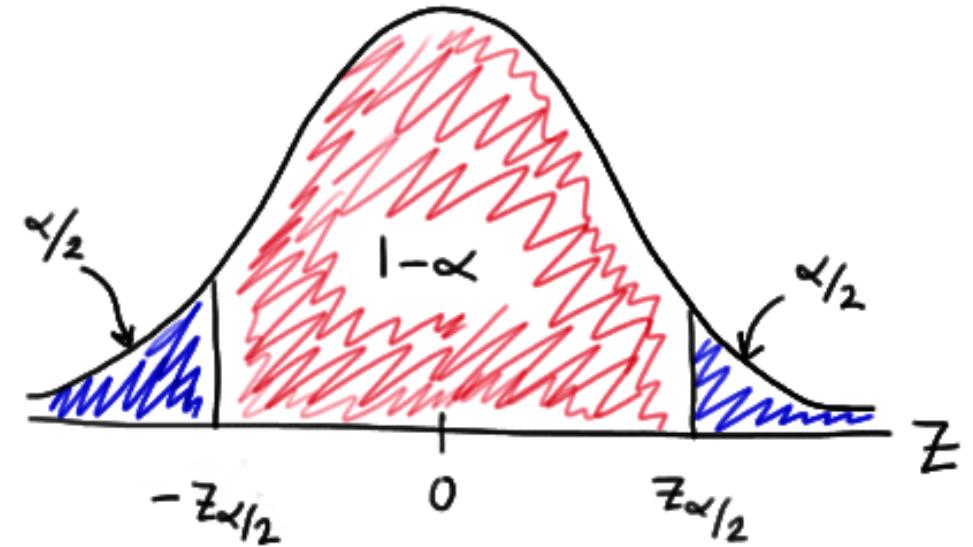
→ Narrower T-distribution

→ More powerful test



Another way to assess this is to obtain the p-value for  $t = 3.686$ , with 20 degrees of freedom (25-5).

```
> #give probability based on quantile value  
> 2*pt(3.686, 20, lower.tail=FALSE)  
[1] 0.001464178
```



This is 0.001464. Since this is less than 0.005, we reject  $H_0$  and conclude that these two means differ.

Can test

- the other pairs of means.
- could test for any size of difference between two means, for example:

$$H_0: \mu_2 - \mu_4 = c$$

$$H_1: \mu_2 - \mu_4 \neq c$$

$$t = \frac{(\bar{y}_{\bullet 2} - \bar{y}_{\bullet 4}) - c}{\sqrt{MSE \left( \frac{1}{n_2} + \frac{1}{n_4} \right)}}$$

# Scheffe's multiple comparison test – conservative

Can test

- any pair of means
- or other comparisons.

# Scheffe's multiple comparison test – conservative

Testing whether the means for Treatments 2 and 4 differ:

$$H_0 : 0\mu_1 + \frac{1}{2}\mu_2 + 0\mu_3 - \frac{1}{2}\mu_4 + 0\mu_5 = 0$$

$$H_0 : \frac{1}{2}\mu_2 - \frac{1}{2}\mu_4 = 0 \qquad \qquad H_0 : \mu_2 = \mu_4$$

The sum of the  $c_j$  (coefficient) values must add up to zero.

The test statistic is:

$$S = \frac{\hat{L}}{s(\hat{L})} \qquad \hat{L} = \sum_{j=1}^J c_j \bar{y}_{\bullet j} \qquad s(\hat{L}) = \sqrt{MSE \times \left( \sum_{j=1}^J c_j^2 \times \frac{1}{n_j} \right)}$$

Example:

$$c_1 = 0 \quad c_2 = \frac{1}{2} \quad c_3 = 0 \quad c_4 = -\frac{1}{2} \quad c_5 = 0$$

$$\hat{L} = \frac{1}{2} \times 4.4 - \frac{1}{2} \times 3.5 = 0.45$$

$$s(\hat{L}) = \sqrt{0.149 \times \left( \left( \frac{1}{2} \right)^2 \times \frac{1}{5} + \left( -\frac{1}{2} \right)^2 \times \frac{1}{5} \right)} = 0.122$$

$$S = \frac{0.45}{0.122} = 3.686$$

Under  $H_0$ , this follows:

$$S_{\text{critical}} = \sqrt{(J - 1)F_{1-\alpha, J-1, n_T - J}}$$

For  $J=5$ ,  $\alpha=0.05$ , and  $n_T=25$  observations:

$$S_{\text{critical}} = \sqrt{(5 - 1)2.87} = 3.38$$

Calculated  $|S| > 3.38$ , so we reject  $H_0$ , the treatment means differ.

NOTE: The means would have to be at least how much apart to reject?

How much the means would have to be at least apart to reject?

$$S_{critical} = 3.38$$

$$\Leftrightarrow \frac{\hat{L}_{critical}}{s(\hat{L})} = 3.38$$

$$\Leftrightarrow \hat{L}_{critical} = 3.38 \times 0.122$$

$$\Leftrightarrow \text{difference}_{\text{critical}} = (3.38 \times 0.122) \times 2$$

```
> 3.38*0.122*2  
[1] 0.82472
```

# Scheffe's can be used for many comparisons

For example: Test if treatments 3, 4 and 5 differ from treatments 1 and 2:

$$H_0: \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 - \frac{1}{3}\mu_5 = 0 \quad \text{OR}$$

$$H_0: \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4 + \mu_5}{3} = 0$$

NOTE: c's add up to 0.

# Scheffe's can be used for many comparisons

Test if treatments 3, 4 and 5 differ from treatments 1 and 2:

$$c_1 = \frac{1}{2} \quad c_2 = \frac{1}{2} \quad c_3 = -\frac{1}{3} \quad c_4 = -\frac{1}{3} \quad c_5 = -\frac{1}{3}$$

$$\hat{L} = \frac{1}{2} \times 4.1 + \frac{1}{2} \times 4.4 - \frac{1}{3} \times 3.6 - \frac{1}{3} \times 3.5 - \frac{1}{3} \times 3.6 = 0.68$$

$$s(\hat{L}) = 0.158$$

$$s(\hat{L}) = \sqrt{0.149 \times \left( \left(\frac{1}{2}\right)^2 \times \frac{1}{5} + \left(\frac{1}{2}\right)^2 \times \frac{1}{5} \left(-\frac{1}{3}\right)^2 \times \frac{1}{5} + \left(-\frac{1}{3}\right)^2 \times \frac{1}{5} + \left(-\frac{1}{3}\right)^2 \times \frac{1}{5} \right)} = 0.158$$

$$S = \frac{0.68}{0.158} = 4.30$$

Result: Greater than the critical value of 3.38;  
do reject  $H_0$ .

# Confidence limits for treatment means

Under the assumptions, confidence intervals for each treatment mean ( $\mu_j$ ) can be obtained by:

*use info of other treatments*

$$\bar{y}_{\bullet j} \pm t_{(n_T - J), 1-\alpha/2} \sqrt{\frac{MSE}{n_j}}$$

Estimated mean

standard error to estimated mean

good estimate of  $\sigma^2$

if  $n_j$  is constant  
this confidence interval is the SAME  
for all treatment levels

many degrees of freedom  
 $t_c$  is small value (close to 1,96)  
→ narrow confidence interval

For the example:

$$\bar{y}_{\bullet j} \pm t_{(n-k), 1-\alpha/2} \sqrt{\frac{MSE}{n_j}}$$

$$\bar{y}_{\bullet 1} = 4.1 \quad \bar{y}_{\bullet 2} = 4.4 \quad \bar{y}_{\bullet 3} = 3.6 \quad \bar{y}_{\bullet 4} = 3.5 \quad \bar{y}_{\bullet 5} = 3.6$$

All  $\sqrt{\frac{MSE}{n_j}}$  the same as  $n_j$  are all equal  $\sqrt{\frac{0.149}{5}} = 0.173$

$$t_{20, 0.975} = 2.09$$

For treatment 1:

$$4.1 \pm 2.09 \times 0.173$$

$$4.1 \pm 0.36$$

# If treatment means vary, then test which pairs differ

```
> attach(fertdat)
> pairwise.t.test(AveHt, Treatment, p.adj="bonferroni")
```

# t-test()

Pairwise comparisons using t tests with pooled SD

data: AveHt and Treatment

	1	2	3	4
2	1.000	-	-	-
3	0.327	0.021	-	-
4	0.194	0.012	1.000	-
5	0.538	0.038	1.000	1.000

P value adjustment method: bonferroni  
> detach(fertdat)

$$t = \frac{(\bar{y}_{\bullet 2} - \bar{y}_{\bullet 4}) - 0}{\sqrt{MSE \left( \frac{1}{n_2} + \frac{1}{n_4} \right)}}$$

follows t-distribution with n-5 d.f.

We can get estimates of the individual group means + Tukey multiple comparison using the `lsmeans()` function from the `lsmeans` package

```
> library(lsmeans)
> modell.lsmo <- lsmeans(modell, ~Treatment, adjust="tukey")
> summary(modell.lsmo, infer=TRUE)
```

	Treatment	lsmean	SE	df	lower.CL	upper.CL	t.ratio	p.value
1		4.12	0.1725109	20	3.760149	4.479851	23.883	<.0001
2		4.42	0.1725109	20	4.060149	4.779851	25.622	<.0001
3		3.56	0.1725109	20	3.200149	3.919851	20.636	<.0001
4		3.50	0.1725109	20	3.140149	3.859851	20.289	<.0001
5		3.62	0.1725109	20	3.260149	3.979851	20.984	<.0001

Requests also a Tukey multiple comparison procedure

T-tests for means being different of zero

Confidence level used: 0.95

$$\text{SE} = \sqrt{\frac{MSE}{n_j}} = \sqrt{\frac{MSE}{5}} = 0.173 \rightarrow MSE = \frac{SSE}{n_T - J} = \frac{SSE}{20} \leftarrow df$$

The standard errors for the estimated means use the pooled estimate of the within group variation to compute better standard errors. Because the group sizes are all equal, their standard errors will also be equal. For large group sizes, there won't be much of a difference between the standard errors reported here and those reported earlier.

# Compact-letter-display of multiple comparisons:

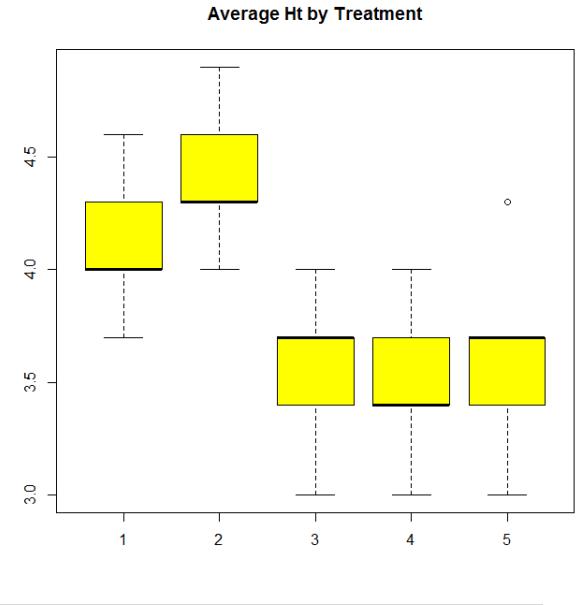
```
> # Get the compact letter display and a plot  
> modell1.cld <- cld(modell1.lsmo)  
> modell1.cld
```

Treatment	lsmean	SE	df	lower.CI	upper.CI	.group
4	3.50	0.1725109	20	3.140149	3.859851	1
3	3.56	0.1725109	20	3.200149	3.919851	1
5	3.62	0.1725109	20	3.260149	3.979851	1
1	4.12	0.1725109	20	3.760149	4.479851	12
2	4.42	0.1725109	20	4.060149	4.779851	2

Confidence level used: 0.95

P value adjustment: tukey method for comparing a family of 5 estimates

significance level used: alpha = 0.05



# Compact-letter-display

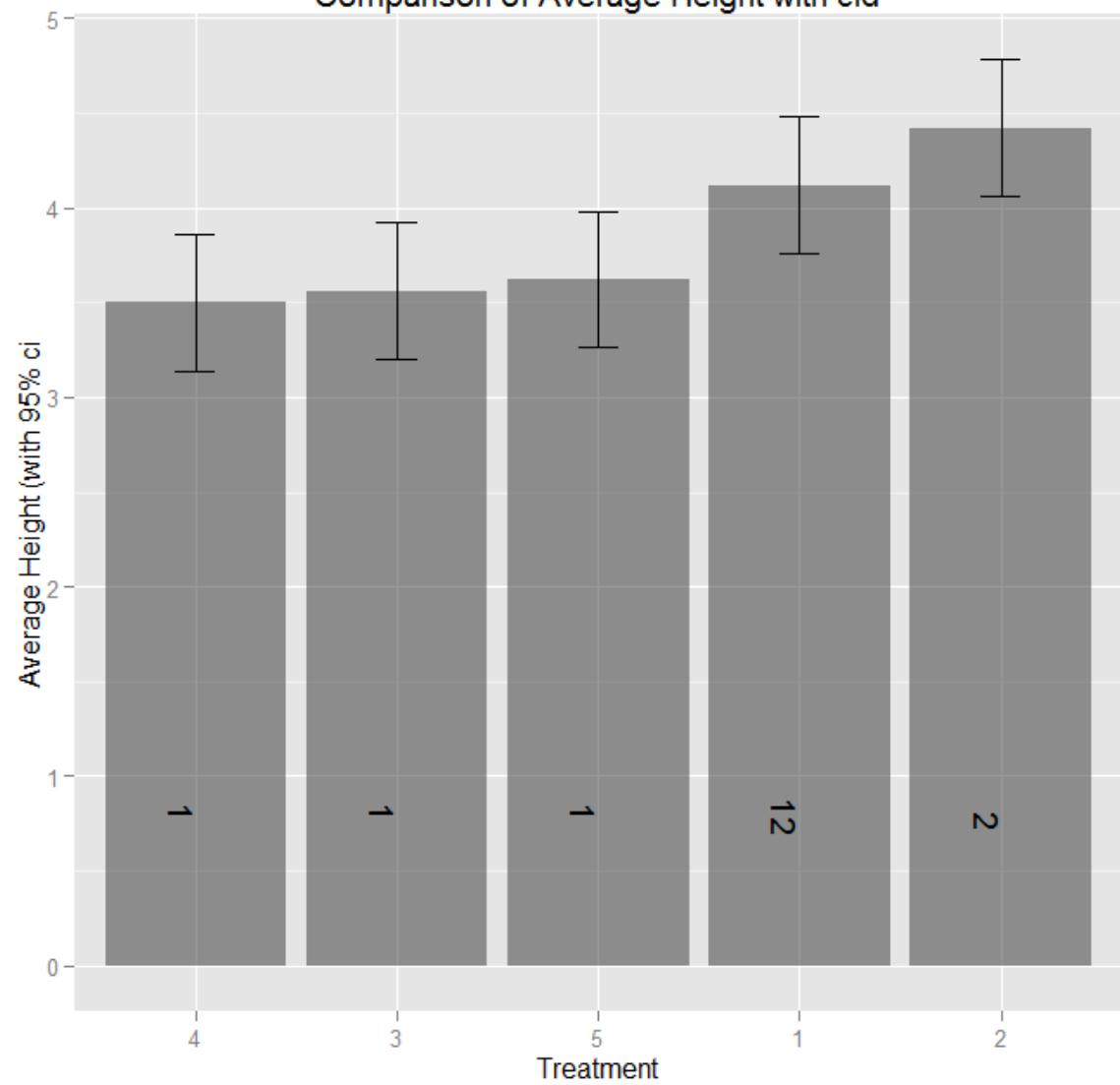
- The sample means are first sorted.
- Then starting with the lowest sample mean, which means cannot be distinguished from that mean are “joined” by the same symbol.
- This indicates that there is no evidence in the data to distinguish the mean of Treatments 4, 3, 5 and 1.
- Note that the mean of Treatment 2 is NOT joined with that of Treatment 4 by any letter. This indicates that there is sufficient evidence to conclude that the mean of Treatment 2 may be different from the mean of Treatment 4.
- Next, look at the second mean (that of Treatment 3) and repeat the process.

# It is also useful to draw plots displaying the results

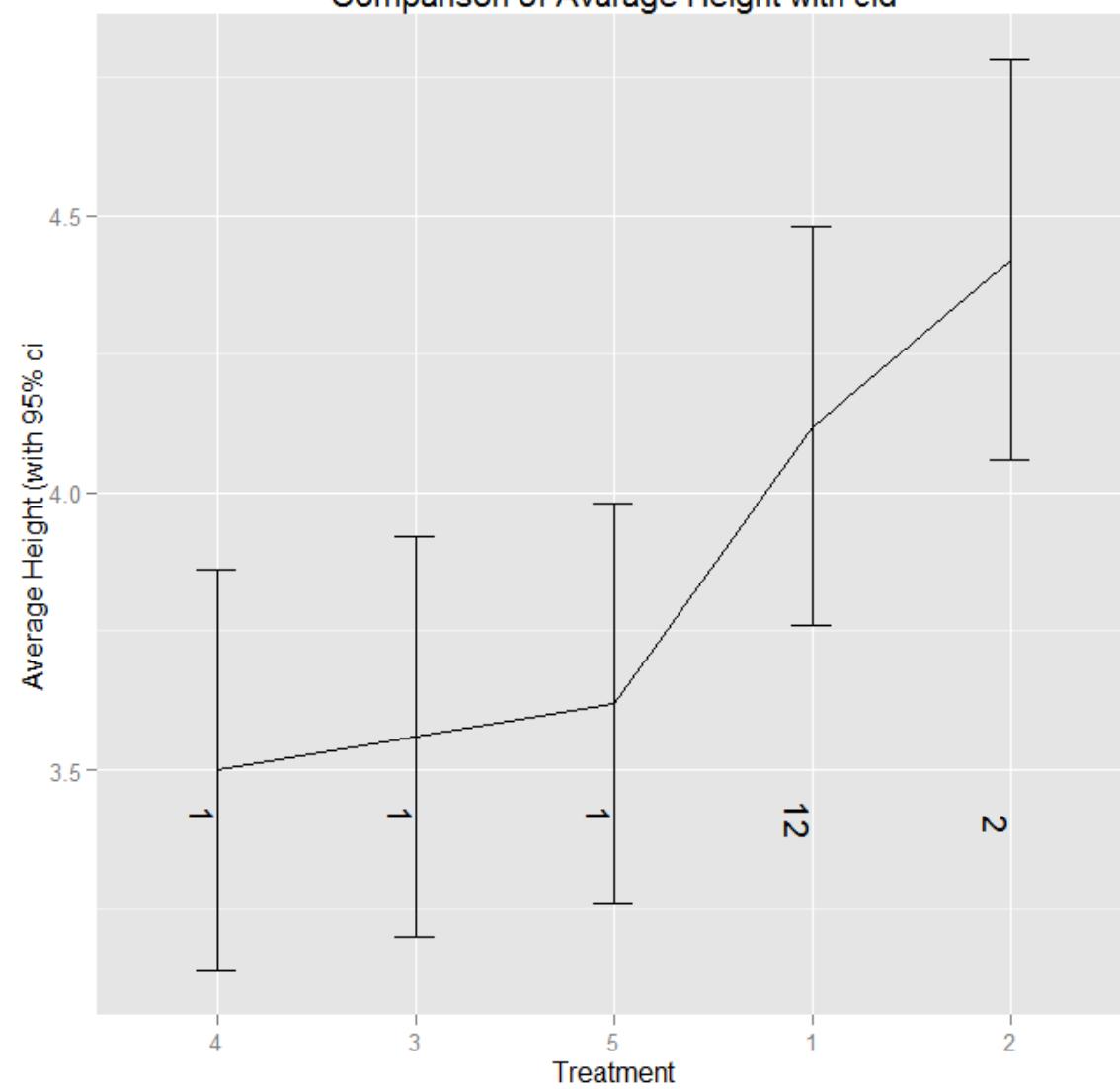
```
> # Make a bar plot of the cld display
> plotcld <- sf.cld.plot.bar(model1.cld, variable="Treatment")
> plotcld <- plotcld +
+   xlab("Treatment") +
+   ylab("Average Height (with 95% ci)") +
+   ggttitle("Comparison of Average Height with cld")
> plotcld

> # Make a line graph of the cld display
> plotcldb <- sf.cld.plot.line(model1.cld, variable="Treatment") #,
ciwidth=0.1)
> plotcldb <- plotcldb +
+   xlab("Treatment") +
+   ylab("Average Height (with 95% ci)") +
+   ggttitle("Comparison of Avarage Height with cld")
> plotcldb
```

Comparison of Average Height with cld



Comparison of Avarage Height with cld



It is also useful to compute all the estimates of the pairwise differences in means along with adjusted confidence intervals for the difference in the means

- This is done using the pairs() function applied to the lsmeans object

```

> # Find all the pairwise differences adjusting for multiplicity
> modell.pairs <- pairs(modell.lsmo, adjust="tukey")
> summary(modell.pairs, infer=TRUE)

contrast estimate      SE df lower.CL upper.CL t.ratio p.value
1 - 2     -0.30 0.2439672 20 -1.03004127 0.4300413 -1.230 0.7347
1 - 3      0.56 0.2439672 20 -0.17004127 1.2900413  2.295 0.1874
1 - 4      0.62 0.2439672 20 -0.11004127 1.3500413  2.541 0.1208
1 - 5      0.50 0.2439672 20 -0.23004127 1.2300413  2.049 0.2797
2 - 3      0.86 0.2439672 20  0.12995873 1.5900413  3.525 0.0162
2 - 4      0.92 0.2439672 20  0.18995873 1.6500413  3.771 0.0094
2 - 5      0.80 0.2439672 20  0.06995873 1.5300413  3.279 0.0275
3 - 4      0.06 0.2439672 20 -0.67004127 0.7900413  0.246 0.9991
3 - 5     -0.06 0.2439672 20 -0.79004127 0.6700413 -0.246 0.9991
4 - 5     -0.12 0.2439672 20 -0.85004127 0.6100413 -0.492 0.9872

```

Confidence level used: 0.95

Conf-level adjustment: tukey method for comparing a family of 5 estimates

P value adjustment: tukey method for comparing a family of 5 estimates

A plot of all of the pairwise differences (and confidence intervals) can be formed:

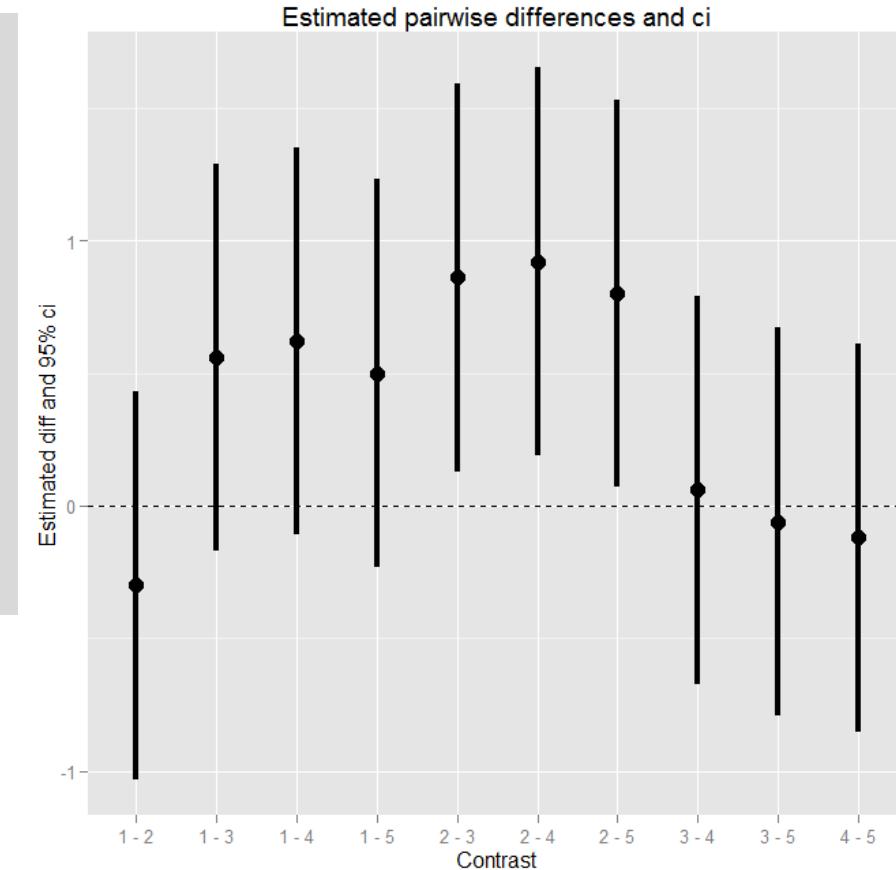
```
> modell.pairs.ci <- confint(modell.pairs) # extract  
the ci values  
> modell.pairs.ci  
 contrast estimate      SE  df lower.CI upper.CI  
1 - 2     -0.30 0.2439672 20 -1.03004127 0.4300413  
1 - 3      0.56 0.2439672 20 -0.17004127 1.2900413  
1 - 4      0.62 0.2439672 20 -0.11004127 1.3500413  
1 - 5      0.50 0.2439672 20 -0.23004127 1.2300413  
2 - 3      0.86 0.2439672 20  0.12995873 1.5900413  
2 - 4      0.92 0.2439672 20  0.18995873 1.6500413  
2 - 5      0.80 0.2439672 20  0.06995873 1.5300413  
3 - 4      0.06 0.2439672 20 -0.67004127 0.7900413  
3 - 5     -0.06 0.2439672 20 -0.79004127 0.6700413  
4 - 5     -0.12 0.2439672 20 -0.85004127 0.6100413
```

Confidence level used: 0.95

Conf-level adjustment: tukey method for comparing a family of 5 estimates

A plot of all of the pairwise differences (and confidence intervals) can be formed:

```
> plotdiff <- ggplot(modell1.pairs.ci, aes(contrast,
estimate, ymin = lower.CL, ymax = upper.CL)) +
+   geom_point(size=4) +
+   geom_linerange(size=1.5) +
+   geom_abline(intercept=0, slope=0, linetype=2) +
+   ylab("Estimated diff and 95% ci") +
+   xlab("Contrast") +
+   ggttitle("Estimated pairwise differences and ci")
> plotdiff
```



# Exercise 1:

- Word document: crd.doc
- Using the script found in crd.R

The script contains additional coding using other R functions

## **8. Completely randomized design (CRD) – Two factors, Fixed effects**

# CRD: Two Factor Factorial Experiment, Fixed Effects

## Introduction

- Treatments can be combinations of more than one factor
- For 2-factor experiment, have several levels of Factor A and of Factor B
- All levels of Factor A occur for Factor B and vice versa (called a *Factorial Experiment*, or *crossed treatments*)

# Example:

- Factor A, (three levels of fertilization: A1, A2, and A3)
- Factor B (four species: B1, B2, B3 and B4)
- Crossed: 12 treatments
- Four replications (pots) per treatment for a total of 48 experimental units
- Measured Responses: seedling height growth in mm

# *Schematic and Measured Response for the Example:*

A1B1=10	A3B2=25	A3B4=35	A2B2=23	A1B2=14	A2B3=24
A1B4=24	A2B2=22	A1B2=15	A2B4=28	A3B3=32	A3B2=25
A3B2=27	A1B4=23	A3B3=29	A3B2=26	A1B3=17	A1B1=11
A3B4=35	A1B2=13	A1B4=22	A1B1=11	A2B3=24	A3B3=30
A1B3=19	A2B1=18	A2B4=30	A3B3=31	A2B3=23	A1B4=22
A3B1=22	A2B4=29	A3B1=23	A2B1=18	A1B2=15	A3B1=23
A2B2=25	A3B4=37	A1B1=9	A3B1=24	A3B4=36	A2B4=28
A1B3=17	A2B1=18	A2B2=20	A2B1=18	A2B3=26	A1B3=18

A1B1=10 indicates that the response variable was 10 for this experimental unit that received Factor A, level 1 and Factor B, level 1. Treatments randomly assigned to the 48 experimental units.

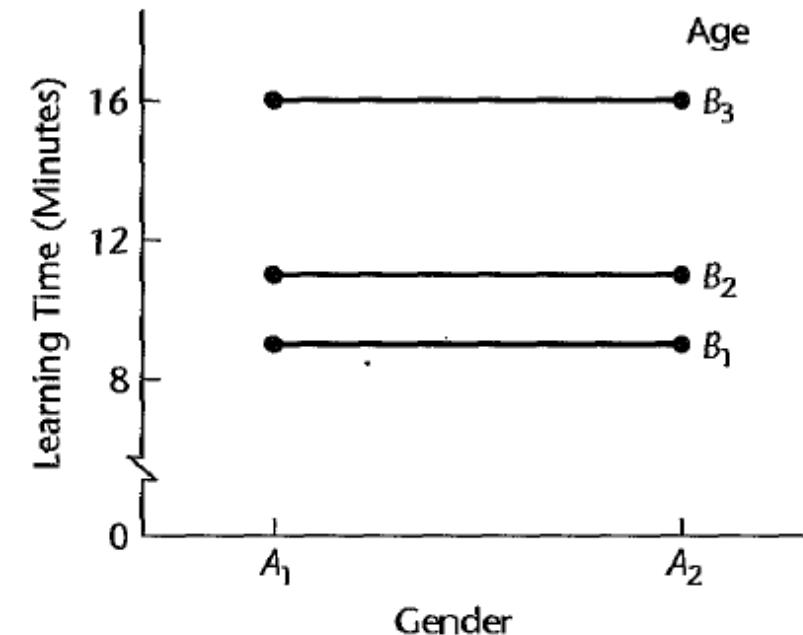
# *Organization of data for analysis using a statistics package:*

A	B	result
1	1	10
1	1	11
1	1	9
1	1	11
1	2	15
1	2	15
1	2	13
1	2	14
1	3	17
1	3	18
1	3	17
1	3	19
1	4	22
1	4	23
1	4	24
1	4	22
2	1	18
2	1	18
2	1	18
2	2	20
...		
3	3	32
3	4	35
3	4	36
3	4	37
3	4	35

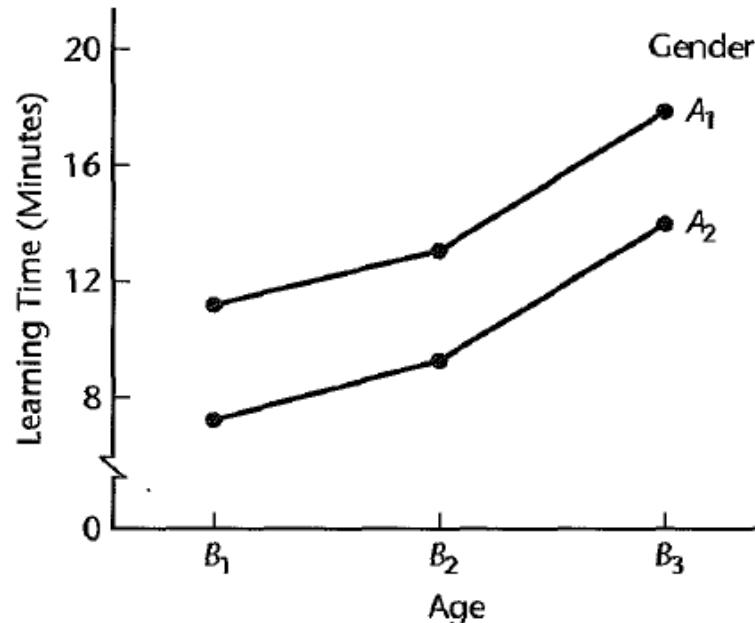
## *Main questions:*

1. Is there an interaction between Factor A and Factor B (fertilizer and species in the example)? Or do the means by Factor A remain the same regardless of Factor B and vice versa?
2. If there is no interaction, is there a difference
  - a) Between Factor A means?
  - b) Between Factor B means?
3. If there are differences:
  - a) If there is an interaction, which treatment means differ?
  - b) If there is no interaction, then which levels of Factor A means differ? Factor B means?

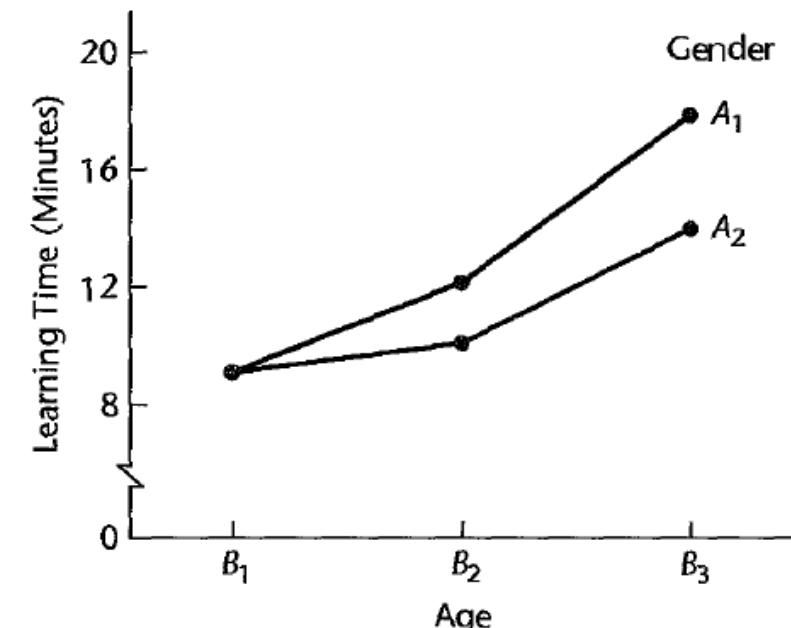
# Two factorial experiment



AGE EFFECT  
BUT NO GENDER EFFECT



AGE AND GENDER EFFECTS  
WITH NO INTERACTIONS



AGE AND GENDER EFFECTS  
WITH IMPORTANT INTERACTIONS

# Notation, Assumptions, and Transformations

Population:  $y_{ijk} = \mu + \tau_{A_j} + \tau_{B_k} + \tau_{AB_{jk}} + \varepsilon_{ijk}$

$y_{ijk}$  = response variable measured on experimental unit  $i$  and factor A level  $j$ , factor B level  $k$

$j=1$  to  $J$  levels for Factor A;  $k=1$  to  $K$  levels for Factor B

$\mu$  = the grand or overall mean regardless of treatment

$\tau_{A_j}$  = the *treatment effect* for Factor A, level  $j$

$\tau_{B_k}$  = the *treatment effect* for Factor B, level  $k$

$\tau_{AB_{jk}}$  = the *interaction* for Factor A, level  $j$  and Factor B, level  $k$

$\varepsilon_{ijk}$  = the difference between a particular measure for an experimental unit  $i$ , and the mean for a treatment:

$$\varepsilon_{ijk} = y_{ijk} - (\mu + \tau_{A_j} + \tau_{B_k} + \tau_{AB_{ij}})$$

For the experiment:  $y_{ijk} = \bar{y}_{...} + \hat{\tau}_{Aj} + \hat{\tau}_{Bk} + \hat{\tau}_{ABjk} + e_{ijk}$

$\bar{y}_{...}$  = the grand or overall mean of all measures from the experiment regardless of treatment; under the assumptions for the error terms, this will be an unbiased estimate of  $\mu$

$\bar{y}_{\bullet jk}$  = the mean of all measures from the experiment for a particular treatment  $jk$

$\bar{y}_{\bullet j\bullet}$  = the mean of all measures from the experiment for a particular level  $j$  of Factor A (includes all data for all levels of Factor B)

$\bar{y}_{\bullet\bullet k}$  = the mean of all measures from the experiment for a particular level  $k$  of Factor B (includes all data for all levels of Factor A)

$\hat{\tau}_{Aj}, \hat{\tau}_{Bk}, \hat{\tau}_{ABjk}$  = under the error term assumptions, will be unbiased estimates of corresponding treatment effects for the population

$e_{ijk}$  = the difference between a particular measure for an experimental unit  $i$ , and the mean for the treatment  $jk$  that was applied to it

$$e_{ijk} = y_{ijk} - \bar{y}_{\bullet jk}$$

$n_{jk}$  = the number of experimental units measured in treatment  $jk$

$n_T$  = the number of experimental units measured over all

$$\text{treatments} = \sum_{k=1}^K \sum_{j=1}^J n_{jk}$$

Treatment means  
Also called  
« predicted values »

Means for the example:

Factor A: 16 observations per level

A1=16.25, A2=23.38, A3=28.75

J = levels for factor A  
= 3

Factor B: 12 observations per level

B1=17.08, B2=20.83, B3=24.17, B4=29.08

K = levels for factor B  
= 4

Treatments (A X B): 4 observations per treatment

$n_{jk} = 4$   
 $J \times K = 3 \times 4 = 12$  treatments  
 $n_T = 48$

## *Sums of Squares:*

as with CRD: One Factor.

$$SSy = SS_{TR} + SSE$$

BUT is now divided into:

$$SS_{TR} = SSA + SSB + SSAB$$

SSy: The sum of squared differences between the observations and the grand mean:

$$SSy = \sum_{k=1}^K \sum_{j=1}^J \sum_{i=1}^{n_{jk}} (y_{ijk} - \bar{y}_{...})^2 \quad df = n_T - 1$$

SSA: Sum of squared differences between the level means for factor A and the grand mean, weighted by the number of experimental units for each treatment:

$$SSA = \sum_{k=1}^K \sum_{j=1}^J n_{jk} (\bar{y}_{\bullet j \bullet} - \bar{y}_{\dots})^2 \quad df = J - 1$$

*12 combinations*  
*x 4 replicates*

SSB: Sum of squared differences between the level means for factor B and the grand mean, weighted by the number of experimental units for each treatment:

$$SSB = \sum_{k=1}^K \sum_{j=1}^J n_{jk} (\bar{y}_{\bullet \bullet k} - \bar{y}_{\dots})^2 \quad df = K - 1$$

$$SSA = \sum_{k=1}^K \sum_{j=1}^J n_{jk} (\bar{y}_{\bullet j \bullet} - \bar{y}_{\dots})^2 \quad df = J - 1$$

Mean for each factor A level:

$$\bar{y}_{\cdot 1 \cdot}$$

$$\bar{y}_{\cdot 2 \cdot}$$

$$\bar{y}_{\cdot 3 \cdot}$$

$$\begin{aligned}
 k = 1, j = 1 &\xrightarrow{\text{yields}} (16.25 - \bar{y}_{\dots})^2 \times 4 \\
 k = 1, j = 2 &\xrightarrow{\text{yields}} (23.38 - \bar{y}_{\dots})^2 \times 4 \\
 k = 1, j = 3 &\xrightarrow{\text{yields}} (28.75 - \bar{y}_{\dots})^2 \times 4 \\
 k = 2, j = 1 \\
 k = 2, j = 2 \\
 k = 2, j = 3 \\
 \dots \\
 k = 4, j = 3
 \end{aligned}$$



Number of times  $j=1$  with  
k=1, with k=2, with k=3  
and with k=4

$$\begin{aligned}
 &(16.25 - \bar{y}_{\dots})^2 \times (4 + 4 + 4 + 4) \\
 &+(23.38 - \bar{y}_{\dots})^2 \times (4 + 4 + 4 + 4) \\
 &+(28.75 - \bar{y}_{\dots})^2 \times (4 + 4 + 4 + 4)
 \end{aligned}$$

weighted by the  
number of  
experimental units  
for each treatment

SSAB: Sum of squared differences between treatment means for  $jk$  and the grand mean, minus the factor level differences, all weighted by the number of experimental units for each treatment:

$SSAB$

$$= \sum_{k=1}^K \sum_{j=1}^J n_{jk} ((\bar{y}_{\bullet jk} - \bar{y}_{\dots}) - (\bar{y}_{\bullet \bullet k} - \bar{y}_{\dots}) - (\bar{y}_{\bullet j \bullet} - \bar{y}_{\dots}))^2$$

Since some of the estimated grand means cancel out we obtain:

$$SSAB = \sum_{k=1}^K \sum_{j=1}^J n_{jk} (\bar{y}_{\bullet jk} - \bar{y}_{\bullet \bullet k} - \bar{y}_{\bullet j \bullet} + \bar{y}_{\dots})^2 \quad df = (J-1)(K-1)$$

SSE: Sum of squared differences between the observed values for each experimental unit and the treatment means:

$$SSE = \sum_{k=1}^K \sum_{j=1}^J \sum_{i=1}^{n_{jk}} (y_{ijk} - \bar{y}_{\bullet jk})^2$$

$$df = \sum_{k=1}^K \sum_{j=1}^J (n_{jk} - 1) = n_T - JK$$

Replicates  
in each  
treatment

Number of  
treatments

*Alternative computational formulae:*

$$SSy = \sum_{k=1}^K \sum_{j=1}^J \sum_{i=1}^{n_{jk}} y_{ijk}^2 - \frac{\bar{y}_{\dots\dots\dots}^2}{n_T}$$

$$SSA = \sum_{k=1}^K \sum_{j=1}^J n_{jk} \bar{y}_{\bullet j \bullet}^2 - \frac{\bar{y}_{\dots\dots\dots}^2}{n_T}$$

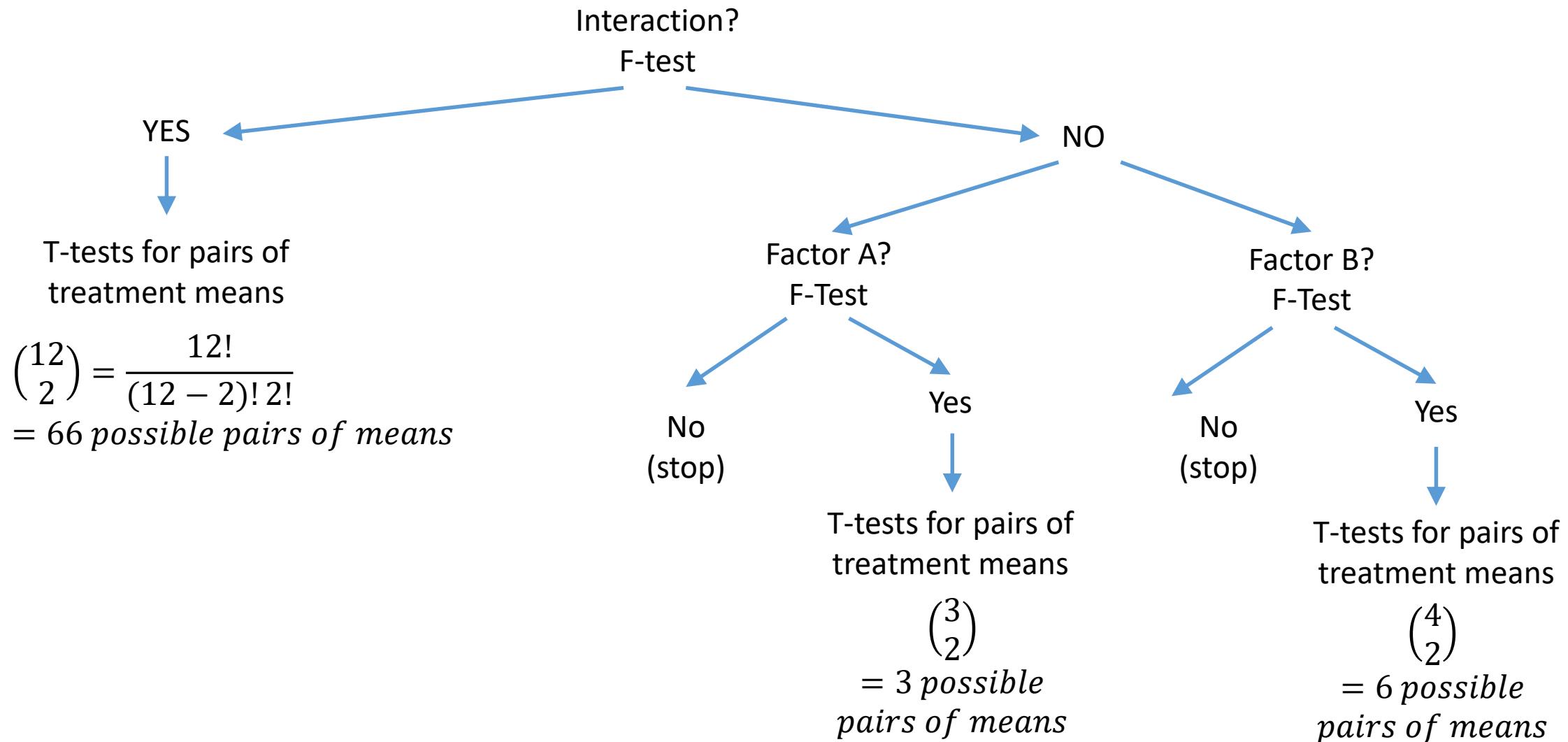
$$SS_{TR} = \sum_{k=1}^K \sum_{j=1}^J n_{jk} \bar{y}_{\bullet j k}^2 - \frac{\bar{y}_{\dots\dots\dots}^2}{n_T}$$

$$SSB = \sum_{k=1}^K \sum_{j=1}^J n_{jk} \bar{y}_{\bullet \bullet k}^2 - \frac{\bar{y}_{\dots\dots\dots}^2}{n_T}$$

$$SSAB = SS_{TR} - SSA - SSB$$

$$SSE = SSy - SS_{TR}$$

# Procedure



# Assumptions and Transformations:

## *Assumptions regarding the error term*

*(Here no sampling assumptions because it's experimental, not observational)*

- Must meet assumptions to obtain unbiased estimates of population means, and an unbiased estimate of the variance of the error term (same as CRD: One Factor)
  - independent observations (not time or space related)
  - normality of the errors,
  - equal variance for each treatment.
- Use residual plot and a plot of the standardized errors against the expected errors for a normal distribution to check these assumptions.

# *Transformations:*

As with CRD: One Factor, you must transform the y-variable

Process:

- First fit the model (means find the numbers!)
- Then test the assumptions of the error term
- If they are not met, transform the y-variable and fit the model again
- If they are met, go for hypothesis testing: Is there an interaction between A and B?

# Test for Interactions and Main Effects

The first main question is: Is there an interaction between the two factors?

$H_0$ : No interaction

$H_1$ : Interaction

*OR:*

$H_0: (\phi_{AB} s^2_\epsilon) / s^2_\epsilon = 1$

$H_1: (\phi_{AB} s^2_\epsilon) / s^2_\epsilon > 1$

Where  $s^2_\epsilon$  is the variance of the error terms;

$\phi_{AB}$  is the interaction effect of the fixed treatments.

# Using an analysis of variance table:

Source	df	SS	MS	F	p-value
A	J-1	SSA	MSA= $SSA/(J-1)$	F= MSA/ <b>MSE</b>	Prob F> $F_{(J-1),(dfE), 1-\alpha}$
B	K-1	SSB	MSB= $SSB/(K-1)$	F= MSB/ <b>MSE</b>	Prob F> $F_{(K-1),(dfE), 1-\alpha}$
A X B	(J-1)(K-1)	SSAB	MSAB= $SSAB/(J-1)(K-1)$	F= MSAB/ <b>MSE</b>	Prob F> $F_{dfAB, dfE, 1-\alpha}$
Error	$n_T - JK$	SSE	<b>MSE</b> = $SSE/(n_T - JK)$		
Total	$n_T - 1$	SSy			

Variance within treatments

→ if assumptions met, then this estimates  $\sigma_\varepsilon^2$  ("unbiased estimator")

$$E[MSE] = \sigma_\varepsilon^2$$

« expected value of MSE is  $\sigma_\varepsilon^2$  »

This means: the average over repeated experiments is  $\sigma_\varepsilon^2$

Source	df	MS	E[MS]
A	$J-1$	$MSA$	$\sigma_{\varepsilon}^2 + \phi_A$
B	$K-1$	$MSB$	$\sigma_{\varepsilon}^2 + \phi_B$
A X B	$(J-1)(K-1)$	$MSB$	$\sigma_{\varepsilon}^2 + \phi_{AB}$
Error	$n_T - JK$	$MSE$	$\sigma_{\varepsilon}^2 = \text{variance within combination of treatments}$
Total	$n_T - 1$		

Treatment effect

For the interactions:

$$F = \frac{SSAB/(J-1)(K-1)}{SSE/(n_T - JK)} = \frac{MSAB}{MSE}$$

Under  $H_0$ , this follows  $F_{df1, df2, 1-\alpha}$

where

- df1 is from the numerator  $(J-1)(K-1)$
- df2 is from the denominator  $(n_T - JK)$

If the  $F$  calculated is greater than the tabular  $F$ ,  
or if the p-value for  $F$  calculated is less than  $\alpha$ ,  
reject  $H_0$

- The means of Factor A are influenced by the levels of Factor B and the two factors cannot be interpreted separately.
- Graph the means of all treatments (boxplots)
- Conduct multiple comparisons of all treatments (rather than on means of each Factor)
- Not as much power if this occurs.

If there are no interactions between the factors, we can look at each factor separately  
→ fewer means, less complicated.

Factor A:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_J$$

OR:

$$H_0: (\phi_A + \sigma^2_\varepsilon)/\sigma^2_\varepsilon = 1$$

$$H_1: (\phi_A + \sigma^2_\varepsilon)/\sigma^2_\varepsilon > 1$$

Where  $\sigma^2_\varepsilon$  is the variance of the error terms;  
 $\phi_A$  is fixed effect for Factor A.

From the ANOVA table:

$$F = \frac{SSA/(J-1)}{SSE/(n_T - JK)} = \frac{MSA}{MSE}$$

Under  $H_0$ , this follows  $F_{df1, df2, 1-\alpha}$

where  $df1$  is from the numerator ( $J-1$ ) and  
 $df2$  is from the denominator ( $n_T - JK$ )

If the  $F$  calculated is greater than the tabular  $F$ ,  
or if the p-value for  $F$  calculated is less than  $\alpha$ ,  
reject  $H_0$ .

- The means of Factor A in the population are likely not all the same
- Graph the means of Factor A levels
- Conduct multiple comparisons between means for the  $J$  levels of Factor A, separately

The analysis and conclusions would follow the same pattern for Factor B.

# *Analysis of Variance Table Results for the Example:*

J=3  
K=4

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F	p
A	$J - 1 = 2$	1258.17	629.08	514.70	<0.0001
B	$K - 1 = 3$	934.75	311.58	254.93	<0.0001
A X B	$2 \times 3 = 6$	17.00	2.836	2.32	0.0539
Error	$48 - (4 \times 3) = 36$	44.00	1.22		
Total	47	2253.92			

If assumptions met, (residuals are independent, are normally distributed, and have equal variances among treatments), we can interpret the results.

## *Interpretation using $\alpha = 0.05$ :*

- No significant interaction ( $p=0.0539$ ); we can examine species and fertilizer effects separately.
- significant differences between the three fertilizer levels of Factor A ( $p<0.0001$ ), and between the four species of Factor B ( $p<0.0001$ ).
- The mean values based on these data are:

$A_1=16.25, A_2=23.38, A_3=28.75$

$B_1=17.08, B_2=20.83, B_3=24.17, B_4=29.08$

Did not have to calculate these for each of the 12 treatments since there is no interaction

# Further analyses, for each Factor separately:

- Scheffé's test for multiple comparisons, could then be used to compare and contrast Factor level means.
  - The number of observations in each factor level are: 16 for Factor A, and 12 for Factor B
  - Use the MSE for both Factor A and for Factor B (denominator of their F-tests)
- t-tests for each pair of means could be used instead.
  - Again, use MSE, and 16 observations for Factor A versus 12 for Factor B
  - Must split alpha level used in the F-tests by the number of pairs

## Factor A: t-tests for pairs of means

Determine the number of pairs possible

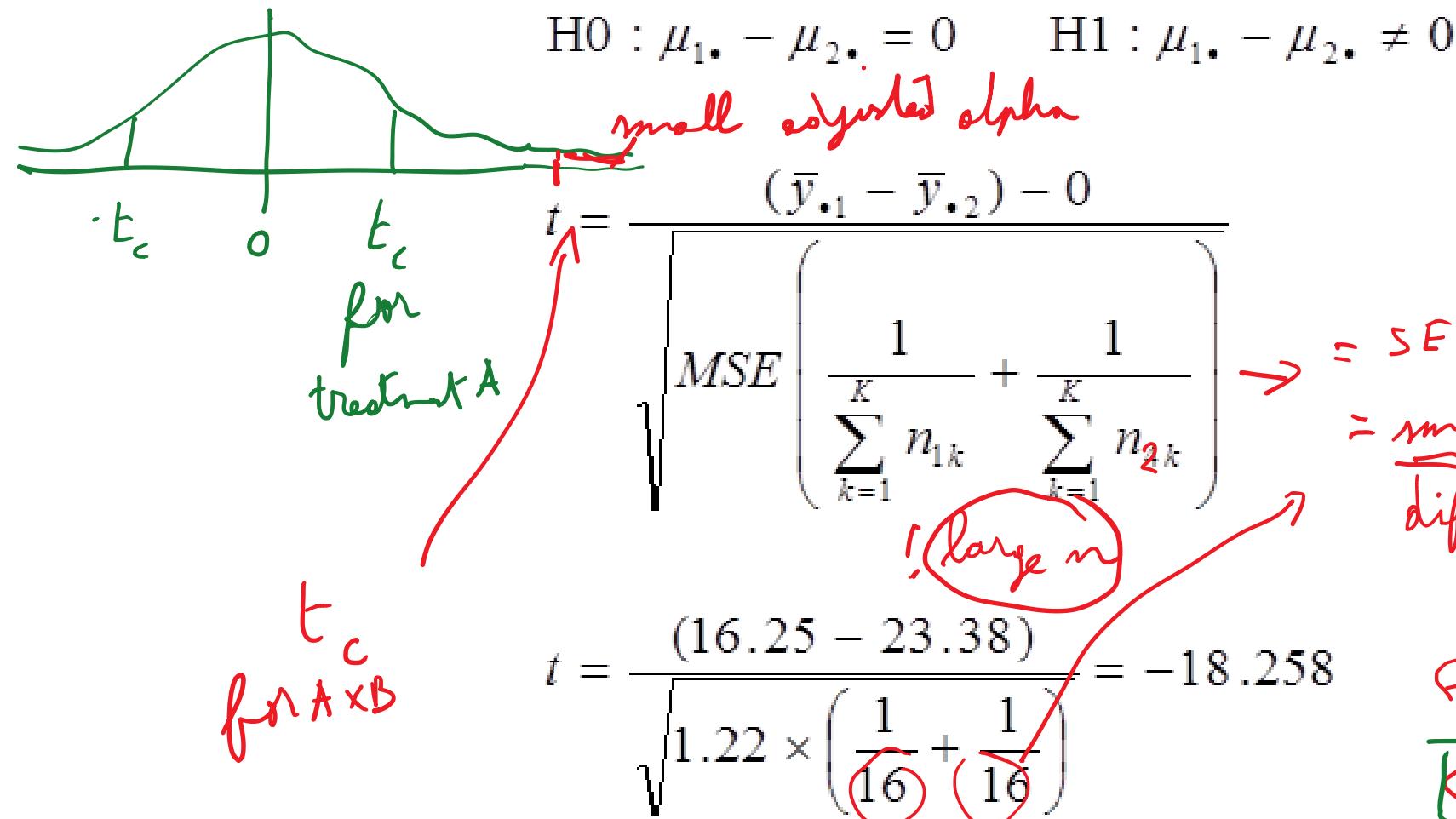
$$\binom{3}{2} = \frac{3!}{1!2!} = 3 \text{ possible pairs of means}$$

Use a significance level of  $0.05/3$  pairs=0.017 for each t-test

$$t_c = qt(\text{alpha} / \text{number of tests}, df = n_T - JK)$$

smaller alpha, larger  $t_c$  for treatment interactions

## Comparing Factor Levels 1 and 2: A1 vs. A2



some t-distribution  
for separate treatments and interactions

$= SE \text{ of } (\bar{y}_{\cdot 1} - \bar{y}_{\cdot 2})$   
 $= \text{smaller than } SE \text{ of }$   
 $\text{difference between means}$   
 $\text{of treatment interactions}$

$\Leftrightarrow t = \text{bigger and } t_c = \text{smaller}$   
 $\Leftrightarrow \text{more powerfull!}$

# Critical t value from a probability table for:

- $df(error) = 36$  based on ( $n_T - JK$ ), and 0.017 significance level (For  $\alpha = 0.05$  use  $0.05/3$  pairs for each t-test), 2-sided test

```
> #give quantile value based on probability  
> #the alpha value should be divided by two, since we look here at the lower tail  
> qt(0.017/2, 36)  
[1] -2.502785
```

- Since the absolute value of the calculated t is greater than |-2.50| we reject H0.

OR

- enter your t-value, df (error), and 2 (for 2-sided) into the EXCEL function =tdist(18.258,36,2)

```
> #give probability based on quantile value  
> #what is the chance to have this t-value or more extreme: a two-sided test!  
> 2*pt(-18.258, 36,lower.tail=TRUE) #area less than -18.258 plus area greater than 18.258  
[1] 8.735309e-20
```

- Since  $p < 0.017$ , we reject H0

→ The mean of treatment A1 differs from the mean of A2.

## For Factor B

- Recalculate the number of possible pairs for 4 factor levels (will be 6 pairs; divide alpha by this for each test )
- The observations per factor level is 12, rather than 16
- Df(error) and MSE are the same as for Factor A.

## *A Different Interpretation using $\alpha = 0.10$ :*

- There is a significant interaction ( $p=0.0539$ ) using  $\alpha = 0.10$ ; cannot interpret main effects (A and B) separately.
- The mean values based on these data are:

$A1B1=10.25$   $A1B2=14.25$   $A1B3= 17.75$   $A1B4= 22.75$

$A2B1=18.00$   $A2B2=22.50$   $A2B3= 24.25$   $A2B4=28.75$

$A3B1= 23.00$   $A3B2=25.75$   $A3B3=30.50$   $A3B4=35.75$

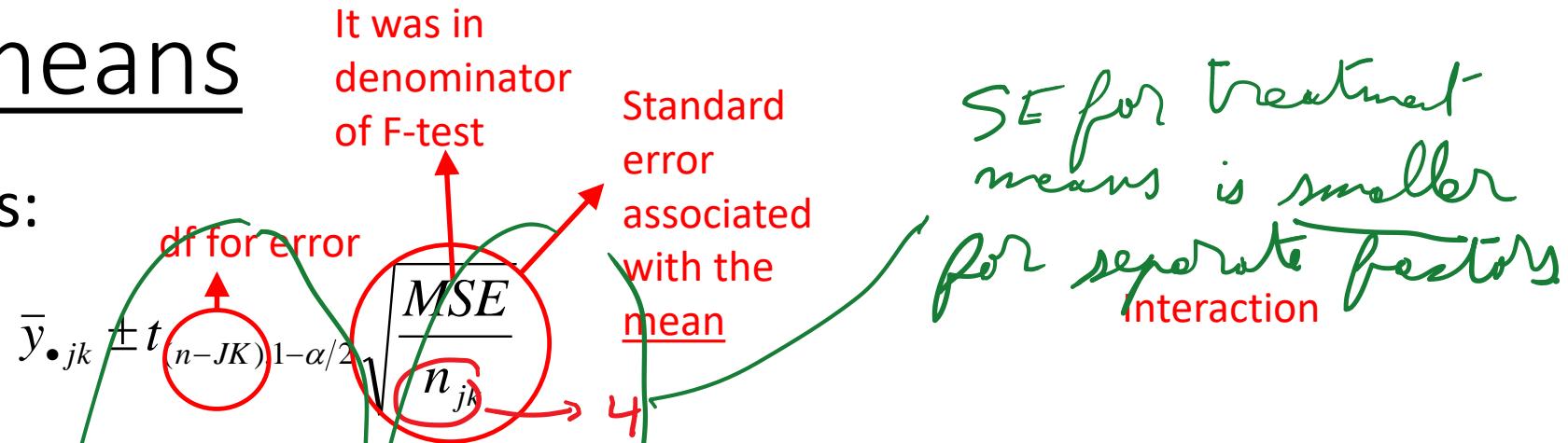
**12 mean values** as there is a significant interaction

# Further analyses:

- Scheffé's test for multiple comparisons (or others), could then be used to compare and contrast treatment means (pairs or other groupings of means). The number of observations in each treatment are 4 [lower power than if there was no interaction], and use the MSE.
- Using t-tests for pairs of means, the number of observations are 4 for each  $jk$  treatment, use the MSE, and recalculate the number of possible pairs out of 12 treatments (will be 66 pairs! Retaining  $\alpha = 0.10$ , we would use  $0.10/66 = 0.0015$  for each t-test )

# Confidence limits for factor level and treatment means

- Treatment means:



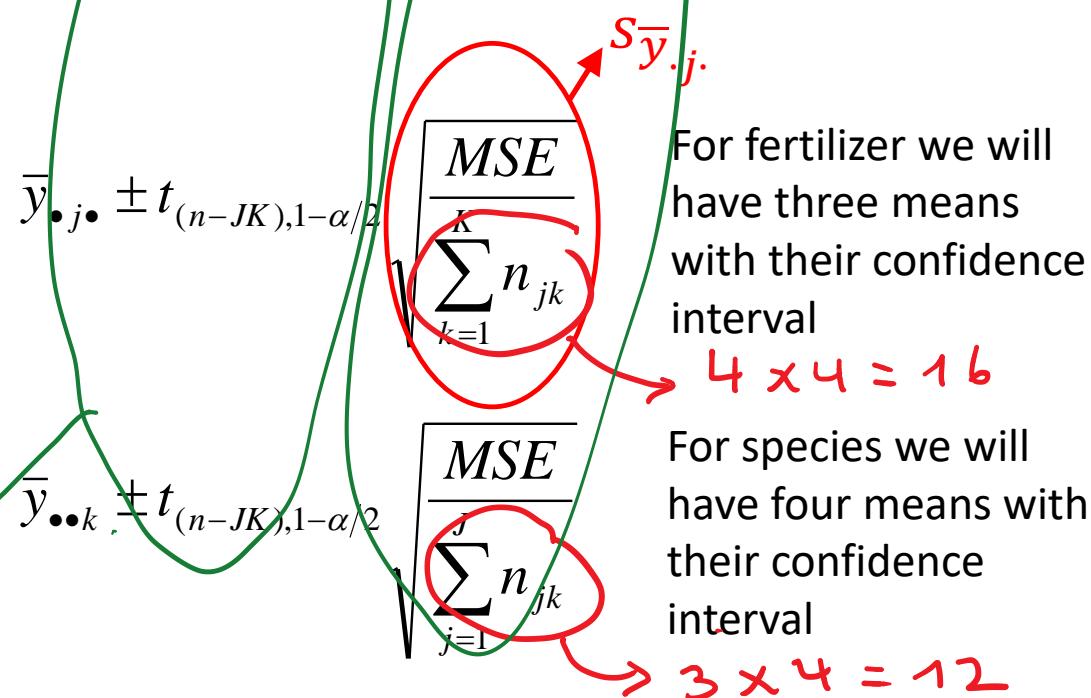
- Factor A means:

$$J = 3$$

$$K = 4$$

- Factor B means:

*t is the same*



```

> fertdat <-  

read.csv("C:\\\\Users\\\\Rebecca\\\\Documents\\\\Cours\\\\201415\\\\TADE2015\\\\Fertilizatio  

ntrial_two_factors.csv", header=TRUE)  

> names(fertdat)  

[1] "A"      "B"      "result"  

> dim(fertdat)  

[1] 48   3  

> fertdat$A <- as.factor(fertdat$A)  

> fertdat$B <- as.factor(fertdat$B)  

> model1<-lm(result~A+B+A*B, data=fertdat)  

> anova(model1)

```

Analysis of Variance Table

Response: result

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	2	1258.17	629.08	514.7045	< 2e-16 ***
B	3	934.75	311.58	254.9318	< 2e-16 ***
A:B	6	17.00	2.83	2.3182	0.05392 .
Residuals	36	44.00	1.22		
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

The `lm()` function can be used to fit this model followed by using the `anova()` function to investigate the effect tests.

The `lm()` function can be used for balanced and unbalanced data, but some care is needed in getting the correct ANOVA tables with unbalanced data – the `anova()` function does NOT give the correct sums-of-squares !! (it gives what are known as incremental or Type I sums of squares) in unbalanced data.

## Exercise :

- Word document: crd\_two\_factors.doc
- Using the script found in crd\_two\_factors.R

# Unbalanced data in CRD

- Equal replications per cell

		Factor B		
		b1	b2	b3
		-----	-----	-----
Factor A	a1	xx	xx	xx
	a2	xx	xx	xx
		-----	-----	-----

- Easiest!
- Like previous examples

- Unequal replications per cell, but replicates in every cell

		Factor B		
		b1	b2	b3
		-----	-----	-----
Factor A	a1	xxx	xx	xxx
	a2	xx	xx	xxxx
		-----	-----	-----

- computer package should be programmed correctly.
- R fits “incremental sums-of-squares” which may not test hypotheses that are of interest – see the examples below

- Unequal replications per cell, with some cells having only a single observation

		Factor B		
		b1	b2	b3
Factor A	a1	x	xx	xxx
	a2	xx	x	xx

- Like previous
- But must assume that the variability in the cells with replicates is an accurate representation of that in cells with only a single observation

- One observation per cell

		Factor B		
		b1	b2	b3
Factor A	a1	x	x	x
	a2	x	x	x

- Impossible to test interaction effect
- MUST ASSUME that no interaction exists
- fit a model without any interaction terms.

- One or more cells completely empty

		Factor B		
		b1	b2	b3
		+	-----+	-----+
Factor A	a1	xx	xx	
	a2	xx	xx	xx

- Tough problem! – seek help
- You may remove a column from the dataset to get rid of empty cells

# Example of a two factor CRD analysis of variance with unbalanced data:

## Stream residence time of salmon

- The stream residence of time was measured for individually tagged fish in a number of years.

```
Effect of Sex and Year on restime levels in Fish
```

```
Listing of part of the raw data
```

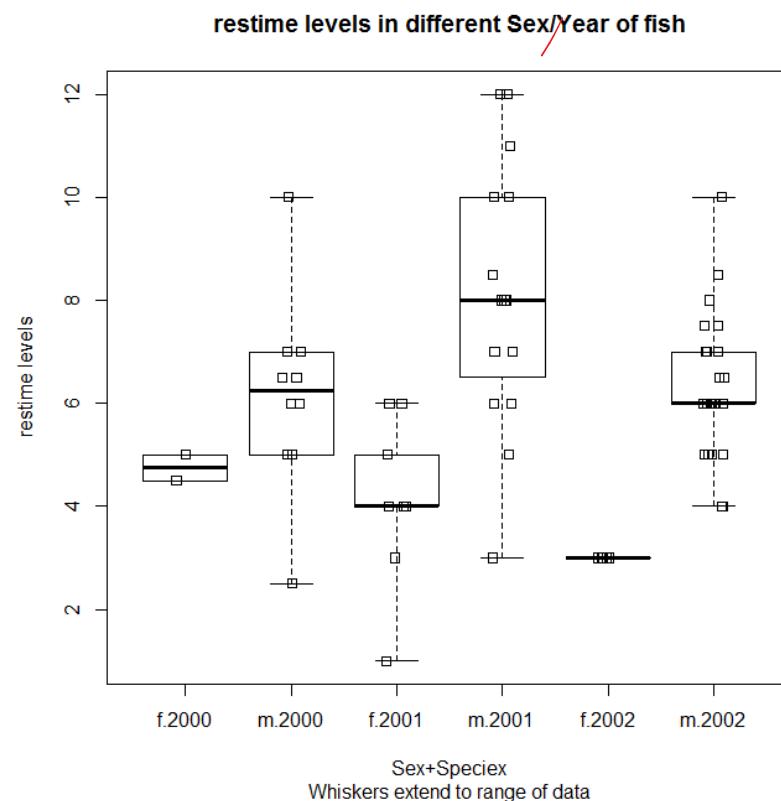
	ResTime	Sex	Year	trt
1	4.5	f	2000	f.2000
2	5.0	f	2000	f.2000
3	6.0	m	2000	m.2000
4	5.0	m	2000	m.2000
5	6.5	m	2000	m.2000
6	10.0	m	2000	m.2000
7	6.0	m	2000	m.2000
8	7.0	m	2000	m.2000
9	7.0	m	2000	m.2000
10	6.5	m	2000	m.2000

Script: residence.R

Data: residence.csv

	Sex	Year	n	nmiss	mean	sd	se	lcl	ucl
1	f	2000	2	0	4.750000	0.3535534	2.500000e-01	1.573449	7.926551
2	f	2001	9	0	4.111111	1.5365907	5.121969e-01	2.929983	5.292239
3	f	2002	6	0	3.000000	0.0000000	1.058756e-16	3.000000	3.000000
4	m	2000	10	0	6.150000	1.9010231	6.011563e-01	4.790090	7.509910
5	m	2001	15	0	8.100000	2.5787594	6.658328e-01	6.671931	9.528069
6	m	2002	26	0	6.288462	1.3279539	2.604332e-01	5.752089	6.824834

68 +



Unbalanced: unequal numbers of males and females in each year

Important assumption: missing values occur at random (not informative)!

J = levels for factor Sex

$$= 2$$

K = levels for factor Year

$$= 3$$

$$n_T = 68$$

$n_{j,k}$  = different for each treatment  $j,k$

# The statistical model

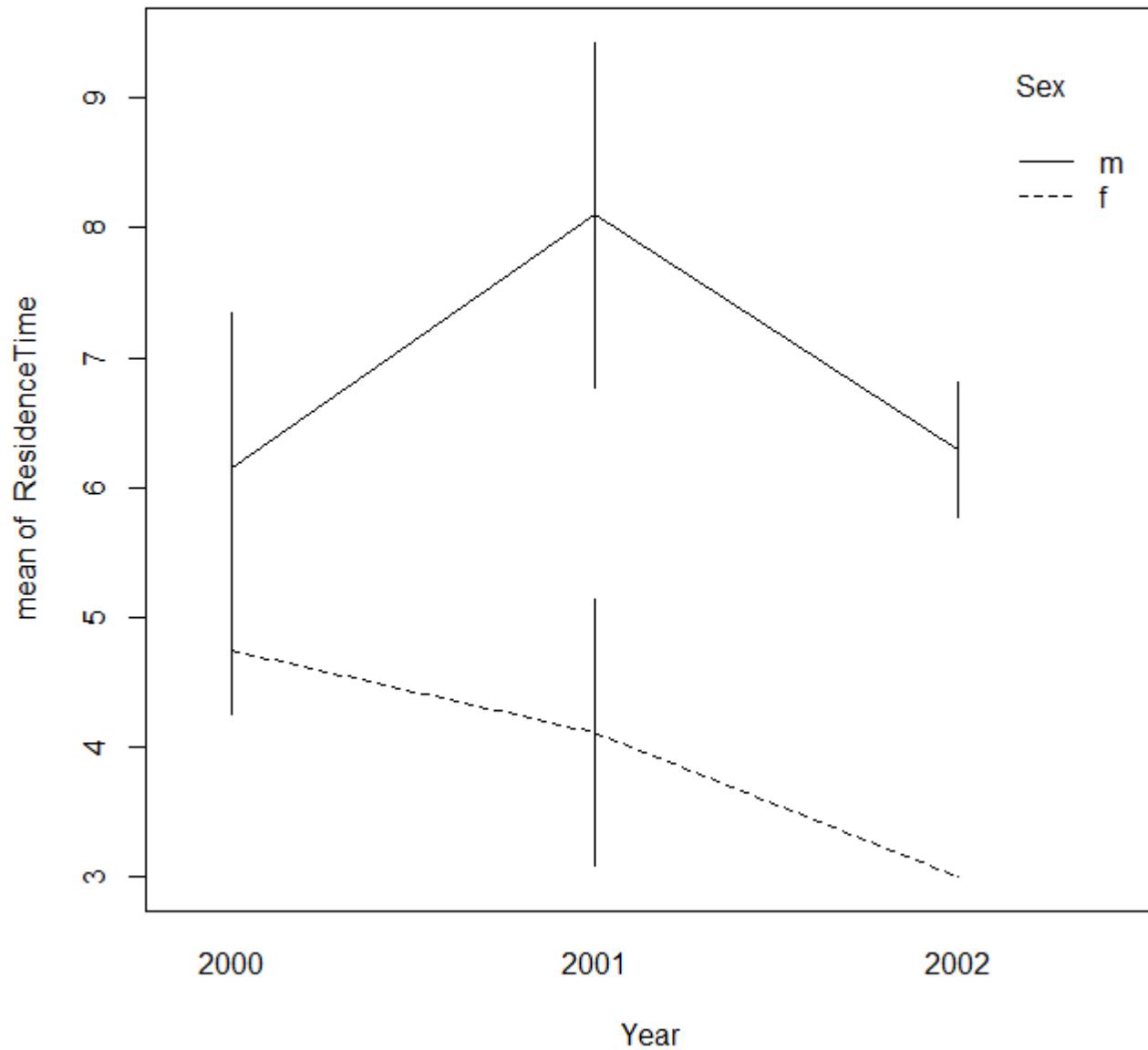
- Sex: fixed effect
- Year: fixed or random?

Reason to say it's random: You do want to extrapolate to all possible years

Reason to say it's fixed: not randomly selected from all possible years

→ let's say it's fixed

**Profile plot with approx 95% ci**



No evidence of interaction

# Hypothesis testing and estimation

- The `Im()` function can be used for balanced and unbalanced data, but GREAT care is needed in getting the correct ANOVA tables with unbalanced data – the `anova()` function does NOT give the correct sums-of-squares (it gives what are known as incremental or Type I sums of squares).
- The **Type III tests from the `Anova()` function from the car package** requires that you set the treatment contrasts to sum-to-zero rather than the default treatment BEFORE fitting the `Im()` model!
- This is done using the `options()` functions as noted in the code.  
→ So use the `Im()` with the default contrasts for everything EXCEPT the ANOVA table!

```
> #The sum of squares and F-tests from the anova() below are INCORRECT in  
unbalanced data  
> #because they are sequential and only adjust for effects that enter the  
model prior to the term in question.")  
> model <- lm( ResidenceTime ~ Sex + Year + Sex*Year, data=restime)  
> # Analysis of variance -- this is NOT CORRECT because design is unbalanced  
> anova(model)
```

Analysis of Variance Table

Response: ResidenceTime

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Sex	1	114.750	114.750	37.6976	6.45e-08	***
Year	2	34.240	17.120	5.6243	0.005693	**
Sex:Year	2	8.652	4.326	1.4211	0.249192	
Residuals	62	188.725	3.044			

---

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

$$n_T - JK = 68 - (2 \times 3)$$

```
> #Use the Type III tests from the Anova() function from the car package
> #but you need to set the treatment contrasts to sum rather than treatment BEFORE
fitting the lm() model!
> library(car)
> old.options <- options()
> options(contrasts=c(unordered="contr.sum", ordered="contr.poly"))
> options()$contrasts
  unordered      ordered
"contr.sum" "contr.poly"
> model2 <- lm( ResidenceTime ~ Sex + Year + Sex*Year, data=restime)
> Anova(model2,type=3) #you must specify type 3, this is not default!
```

Anova Table (Type III tests)

Response: ResidenceTime

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	1067.99	1	350.8552	< 2.2e-16 ***
Sex	76.61	1	25.1665	4.695e-06 ***
Year	22.31	2	3.6648	0.03131 *
Sex:Year	8.65	2	1.4211	0.24919
Residuals	188.73	62		

---

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

```
> options(old.options)
```

# Types of Sum of Squares\*

- The statistical significance and effect size of each predictor is determined in part by the sum of squares attributed to that predictor.
- However, there are different ways of calculating the SS for each predictor.

\*topic out of the scope of this course but required here

# Type I Sum of Squares

- Consider the model

$$y = b_0 + b_1 x_1 + \cdots + b_n x_n$$

- Each  $SS_M$  is calculated relative to the preceding model.

- $SS_{M1}$  is calculated relative to Model 0:  $y_0 = b_0$ :

$$SS_{M1} = \sum_{i=1}^n (y_{1i} - y_{0i})^2$$

- $SS_{M2}$  is calculated relative to Model 1:  $y_1 = b_0 + b_1 x_1$ :

$$SS_{M2} = \sum_{i=1}^n (y_{2i} - y_{1i})^2$$

- ...

- The total of the sums of squares for each predictor equals the total sum of squares for the model as a whole:

$$SS_M = SS_{M1} + \cdots + SS_{Mn}$$

# Type III Sum of Squares

- Consider the model

$$y = b_0 + b_1 x_1 + \cdots + b_n x_n$$

- Each  $SS_M$  is calculated relative to the entire model but excluding that predictor.

  - $SS_{M1}$  is calculated relative to  $y = b_0 + b_2 x_2 + \cdots + b_n x_n$

  - $SS_{M2}$  is calculated relative to  $y = b_0 + b_1 x_1 + b_3 x_3 + \cdots + b_n x_n$

  - ...

- Note that the individual  $SS_M$ s no longer sum to the total

$$SS_M: SS_M \neq SS_{M1} + \cdots + SS_{Mn}$$

*except when design is balanced  
( $\rightarrow$  Type I = Type III)*

# Type I / Type III

- SAS will print out a Type I, II, and III tests –
- in balanced data these will always be the same and any can be used.
- In unbalanced data, these three types of tests will have different results –
- ALWAYS use the Type III (also known as the marginal) tests.
  
- In R, the default method are Type I (incremental) tests and can be misleading (i.e. testing the wrong hypothesis) in cases of unbalanced data

See <http://r.789695.n4.nabble.com/Type-I-v-s-Type-III-Sum-Of-Squares-in-ANOVA-td1573657.html>

- First test for interaction

$H_0$ : no interaction between the two factors.

$H_1$ : some interaction between the two factors.

$F = 1.42$  with a p-value of 0.25: no interaction

- Then examine main effects of each factor

null hypothesis is no effect of the factor upon the mean response, when averaged over the levels of the other factor. ( $\rightarrow$ test for interaction between factors before testing for main effects!!)

- ✓ strong evidence of a Sex effect upon the mean residence time ( $F = 25.2$ ;  $p < 0.0001$ )
- ✓ evidence of a Year effect upon the mean residence time ( $F = 3.7$ ;  $p = 0.031$ )

# The sex effect: marginal means (Lsmean)

```
> library(dplyr) ← to calculate lsmeans  
> #Estimated marginal means  
> popMeans(model, eff="Sex")
```

	estimate	se	df	t.stat	p.value	Sex
1	3.953704	0.5128919	62	7.708649	1.268027e-10	f
2	6.846154	0.2633970	62	25.991770	4.837737e-35	m

remember:  
SE depends on

LSME

$n \leftarrow !$

$n \uparrow \leftrightarrow SE \downarrow$

SE are different due to unbalanced design!

- The “least-squares mean” for females is 3.95 days which differs from the raw mean of 3.79 days. This is an artifact of the unbalance in the experiment.
- The LSMEANS of 3.95 days is found as the “average of the averages”,  
i.e. = weight the different years equally

less female than male  
→ larger SE

$$3.95 = (4.75 + 4.11 + 3.00)/3$$

↑ 2002: more observations BUT  
should not contribute more to mean

# Estimate the sex effect:

```
> library(multcomp)
> model.mcp.Sex <- glht(model, linfct = mcp(Sex = "Tukey",
interaction_average=TRUE)
> summary(model.mcp.Sex)
```

The estimated difference in mean residence time (averaged across all three years) is about 2.89 days with a standard error of .58 days.

*linear function specifies the hypothesis to be tested:*

*here: multiple comparison in ANOVA model  
→ mcp( )*

*general linear hypothesis test*

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = ResidenceTime ~ Sex + Year + Sex \* Year, data = restime)

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t )
m - f == 0	2.8925	0.5766	5.017	4.7e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- single-step method)

```
> confint(model.mcp.Sex)
```

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = ResidenceTime ~ Sex + Year + Sex \* Year, data = restime)

Quantile = 1.999

95% family-wise confidence level

Linear Hypotheses:

	Estimate	lwr	upr
m - f == 0	2.8925	1.7399	4.0450

```
> model.mcp.Sex.cld <- cld(model.mcp.Sex) # joined line plot
```

```
> model.mcp.Sex.cld
```

f m

"a" "b"

# The year effect: marginal means (Lsmean)

```
> library(dplyr)
> #Estimated marginal means
> popMeans(model, eff="Year")
  estimate      se  df t.stat   p.value Year
1 5.450000 0.6757172 62 8.065504 3.046359e-11 2000
2 6.105556 0.3678139 62 16.599579 1.707527e-24 2001
3 4.644231 0.3950955 62 11.754705 2.004726e-17 2002
```

least obs. in 2000

- LSMEANS differ from the raw means because of the imbalance in the data.
- Here it makes sense to weight the sexes equally because of the (presumed) 50:50 sex ratio within each year.

year 2000 :  $\text{lsmean} = \frac{4,75 + 6,15}{2} = 5,45$

f: n = 2      M: n = 10

But male observations should not contribute more to the mean than female  $\rightarrow$  LSMEAN

# Estimate the year effect:

```
> model.mcp.Year <- glht(model, linfct = mcp(Year =
  "Tukey", interaction_average=TRUE))
> summary(model.mcp.Year)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = ResidenceTime ~ Sex + Year + Sex \* Year, data = restime)

Linear Hypotheses:

		Estimate	Std. Error	t value	Pr(> t )
2001 - 2000 == 0		0.6556	0.7693	0.852	0.6679
2002 - 2000 == 0		-0.8058	0.7827	-1.029	0.5562
2002 - 2001 == 0		-1.4613	0.5398	-2.707	0.0228 *
---					

Signif. codes: 0 '\*\*\*\*' 0.001 '\*\*\*' 0.01 '\*\*' 0.05 '\*' 0.1 '.' 1  
(Adjusted p values reported -- single-step method)

```
> confint(model.mcp.Year)
```

### Simultaneous Confidence Intervals

#### Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = ResidenceTime ~ Sex + Year + Sex \* Year, data = restime)

Quantile = 2.389

95% family-wise confidence level

#### Linear Hypotheses:

		Estimate	lwr	upr	
2001 - 2000 == 0		0.6556	-1.1824	2.4935	}
2002 - 2000 == 0		-0.8058	-2.6758	1.0643	
2002 - 2001 == 0		-1.4613	-2.7509	-0.1717	

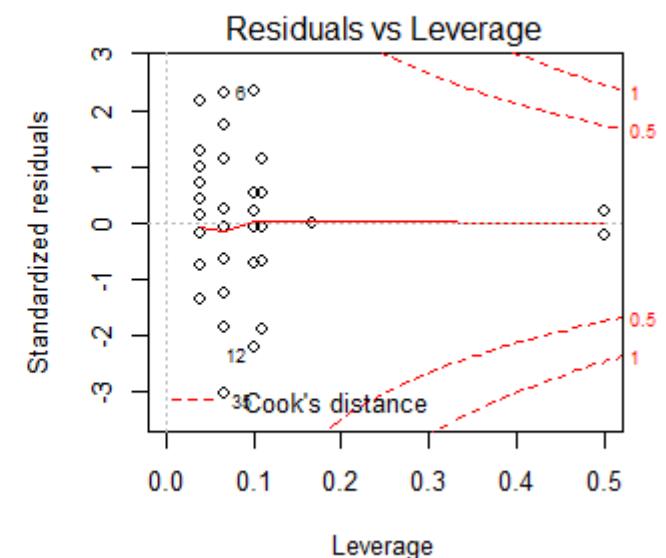
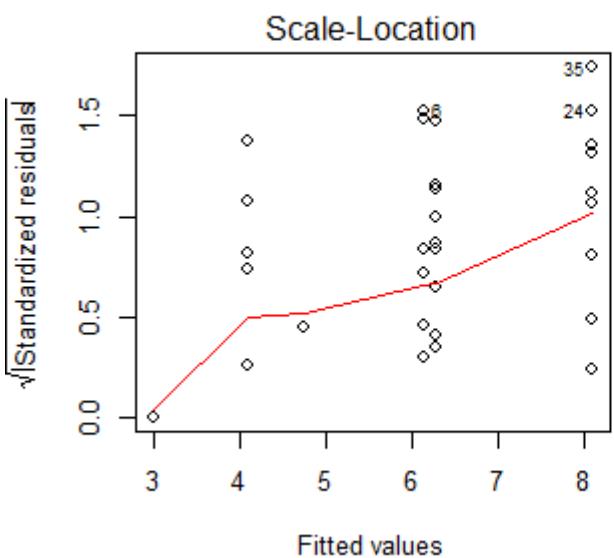
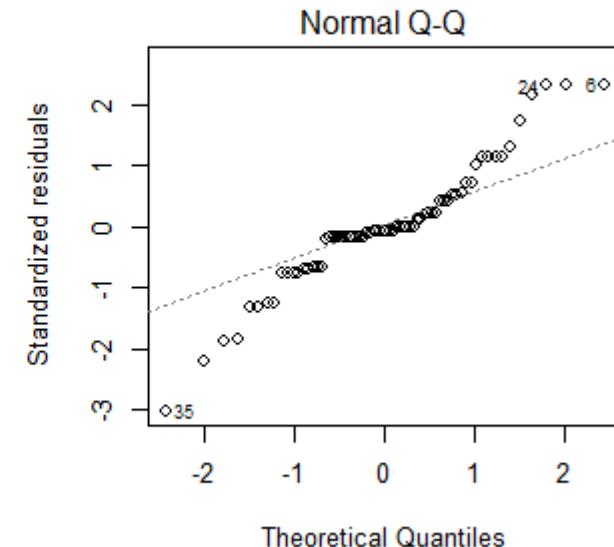
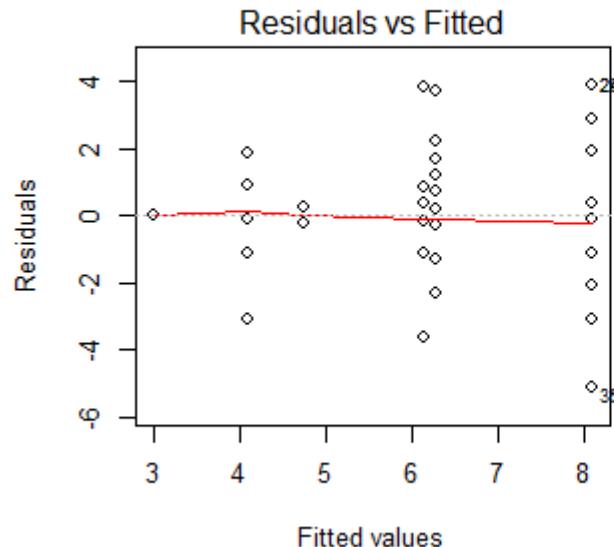
```
> model.mcp.Year.cld <- cld(model.mcp.Year) # joined line plot  
> model.mcp.Year.cld  
2000 2001 2002  
"ab" "b" "a"
```

95% Confidence Interval  
comprises 0  
 $\Leftrightarrow$  difference not significantly  
different from 0

# Diagnostic plots

```
> oldpar <- par(mfrow=c(2, 2))  
> plot(model)  
> par <- oldpar
```

- no evidence of any gross problems
- some evidence that the residuals are not equally variable, but the difference in variance is not large.



## **8. Completely randomized design (CRD) – Two factors, Mixed effects**

# CRD: Random and Mixed Effects

Factors in experiments can be:

- Fixed: all levels of interest are included in the experiment; we are mostly interested in testing differences and estimating means for factor levels
- Random: levels are randomly selected; not all levels of interest are included; we are mostly interested in the variance of the response variable that is DUE TO the factor
- Mixed: When there is more than one factor, there may be a mixture, with some factors that are fixed-effects and others that are mixed-effects

Often, it is difficult to make the distinction!

## Example 1:

We are interested in height growth for different families (genetic stock). We select 4 families from all possible families, and include these in the experiment. Then, we get an estimate of the variance in the height growth due to changes in genetics.

→ [One random-effect factor – family]

# Example 1:

Response variable  $y$  = height growth

One factor A = genetic stock, 4 levels

We are interested in the variance in  $y$  that is due to genetic stock

We are not interested in the means of different levels

$H_0: \sigma_{genetic\ stock}^2 = 0$  (genetic stock doesn't alter height growth)

$H_1: \sigma_{genetic\ stock}^2 > 0$  (genetic stock alters height growth)

→ F-test to test variance due to genetic stock (random factor)

→ No longer test of means ( $\mu_1 = \mu_2 = \mu_3 = \mu_4$ )

# Example 2: example from before with two factors

- Factor A, (three levels of fertilization: A1, A2, and A3)
- Factor B (four species: B1, B2, B3 and B4)
- Crossed: 12 treatments
- Four replications (pots) per treatment for a total of 48 experimental units
- Measured Responses: seedling height growth in mm
- We assumed both Factors were fixed – wanted to compare mean height growth between species and between fertilizers.

Now, we will assume that **species is random** -- these are a few of the species that we are interested in and we wish to look at the variance in height growth that is due to species.

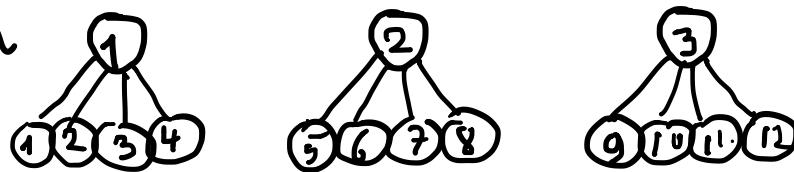
## Example 2:

We are interested in height, depending on species and fertilization level. We select 4 species out of many possible species, and include fertilization levels of low, medium, and high.

- The fertilization levels are fixed-effects (we are only interested in these specific levels that we might apply in a greenhouse to generate seedlings).
- The species are considered random-effects (we are interested estimating the variance in seedling success due to species).

- The most simple design:  
one fixed-effects factor,  
with random allocation of treatments to each experimental unit:

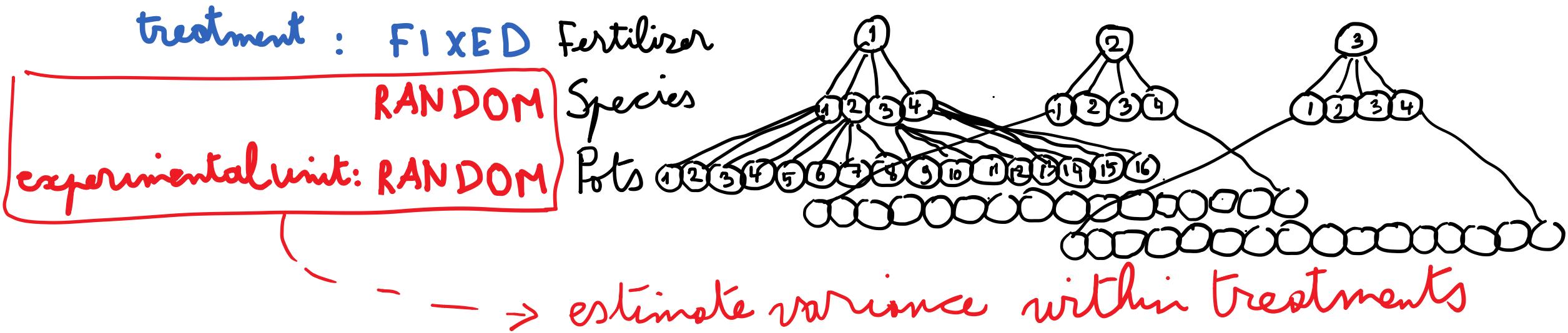
treatment : FIXED Fertilizer  
experimental unit: RANDOM Pots



:  
→ estimate variance within treatments

- Adding a RANDOM variable:

Giving structure to the variance within treatments



Some of the variance within treatment is due to  
the fact that different species are affected differently by Fertilization level  
= interaction effect FIXED x RANDOM

# Adding a random effect (example 2)

- essentially gives structure to the error term “ $\varepsilon$ ”
  - Unexplained variance of errors within treatment levels
  - Variance of errors within treatment due to interaction treatment x species
- resolves the non-independence that stems from having multiple responses for the same species
- Tell the model to assume an intercept that's different for each level of species

## Example 2:

Because there are two factors, we first test the interaction:

$H_0$ : no interaction between fertilization and species      F-test  
 $H_1$ : interaction between fertilization and species

*Interaction is random-effect if one of the factors is random!*

Then, we test

Factor B: species, 4 levels      Random  $\rightarrow$   $H_0: \sigma_{species}^2 = 0$       F-test  
 $H_1: \sigma_{species}^2 > 0$

Factor A: fertilization: 3 levels Fixed  $\rightarrow$   $H_0: \mu_{low} = \mu_{medium} = \mu_{high}$       F-test  
 $H_1: \text{not all the same}$

# Random effects will affect

- the **expected values of the Mean squares**, and then, the F-tests that are used
- Tests that are done following the overall F-test
- The conclusions that are made

For J levels of Factor A and K levels of Factor B, we have the following model:

$$y_{ijk} = \bar{y}_{...} + \hat{\tau}_{A_j} + \hat{\tau}_{B_k} + \hat{\tau}_{AB_{jk}} + e_{ijk}$$

Possibilities:

- Both are fixed (covered already)
- Both are random
- One is fixed and one is random

Variance within  
treatment

Factor effect

# Expected Mean Square Values Comparison:

Mean Square	Model I <b>Both A and B are Fixed</b>	Model II <b>Both A and B are Random</b>	Model III <b>A is Fixed B is Random</b>
A (MSA)	$\sigma_{\varepsilon}^2 + \phi_A$	$\sigma_{\varepsilon}^2 + nK\sigma_A^2 + n\sigma_{AB}^2$	$\sigma_{\varepsilon}^2 + \phi_A + n\sigma_{AB}^2$
B (MSB)	$\sigma_{\varepsilon}^2 + \phi_B$	$\sigma_{\varepsilon}^2 + nJ\sigma_B^2 + n\sigma_{AB}^2$	$\sigma_{\varepsilon}^2 + nJ\sigma_B^2$
A X B (MSAB)	$\sigma_{\varepsilon}^2 + \phi_{AB}$	$\sigma_{\varepsilon}^2 + n\sigma_{AB}^2$	$\sigma_{\varepsilon}^2 + n\sigma_{AB}^2$
Error (MSE)	$\sigma_{\varepsilon}^2$	$\sigma_{\varepsilon}^2$	$\sigma_{\varepsilon}^2$

= structured variance of  $\varepsilon$   
- explained  $\bar{G}_E^L = G_{AB}^2$   
- unexplained  $\bar{G}_E^U$   
variance due to the  
fact that different species  
may be effected differently  
by fertilizer

random  
effect gives  
structure  
to the error  
term  
This is the "new" 11th structure MSE  
that counts when testing  $\phi_A$

Fertilization: Fixed factor = Treatment

Species: Random factor

$$E\{MSE\} = \sigma^2$$

$\sigma^2$  unexplained variance of the error terms  $\varepsilon_{ij}$  **within each combination** Fertilization x Species

$$E\{MS_{AB}\} = \sigma^2 + n\sigma_{AB}^2$$

$n\sigma_{AB}^2$  variance of the error terms **within each combination** Fertilization x Species, **due to AxB interaction** (different fertilizer effects on different species)

- $MS_{AB}$  will be bigger than MSE if the fixed treatment has different effects on different species
- $MS_{AB}$  is the new, « structured » within-treatment variance, containing unexplained and explained variance

$$E\{MS_{TR}\} = \sigma^2 + n\sigma_{AB}^2 + \phi_{TR}$$

- If the treatment effects are larger than [the variance within each combination + the variance due to interaction],  **$MS_{TR}$  will be bigger than  $MS_{AB}$**

# F-tests

- Sums of squares, mean squares, etc are **calculated the same for all three types of models**
- Assumptions: Same as for fixed-effects models
- Change the F-test, so that the numerator differs from the denominator ONLY in the item that you are testing:
- e.g. in Model III: to look at effect of interaction :

$$\text{Test } F = \frac{MSAB}{MSE} \text{ with expected values } \frac{\sigma_{\varepsilon}^2 + n\sigma_{AB}^2}{\sigma_{\varepsilon}^2}$$

- But to look at effect of fixed factor A  $\emptyset_A$ :

$$\text{Test } F = \frac{MSA}{MSAB} \text{ with expected values } \frac{\sigma_{\varepsilon}^2 + \emptyset_A + n\sigma_{AB}^2}{\sigma_{\varepsilon}^2 + n\sigma_{AB}^2}$$

→ if F is big, the effect of A ( $\emptyset_A$ ) is significantly different from 0

- To look at effect of interaction :

Test  $F = \frac{MSAB}{MSE}$  with expected values  $\frac{\sigma_\varepsilon^2 + n\sigma_{AB}^2}{\sigma_\varepsilon^2}$

$$\begin{aligned} H_0: \sigma_{Fertilization * species}^2 &= 0 \\ H_1: \sigma_{Fertilization * species}^2 &> 0 \end{aligned}$$

- If there is no interaction, can look at effect of random factor B :

Test  $F = \frac{MSB}{MSE}$  with expected values  $\frac{\sigma_\varepsilon^2 + nJ\sigma_B^2}{\sigma_\varepsilon^2}$

$$\begin{aligned} H_0: \sigma_{species}^2 &= 0 \\ H_1: \sigma_{species}^2 &> 0 \end{aligned}$$

→ if F is big, the variance due to factor B ( $\sigma_B^2$ ) is significantly different from 0

- BUT to look at effect of fixed factor A :

Test  $F = \frac{MSA}{MSAB}$  with expected values  $\frac{\sigma_\varepsilon^2 + \emptyset_A + n\sigma_{AB}^2}{\sigma_\varepsilon^2 + n\sigma_{AB}^2}$

$$\begin{aligned} H_0: \mu_1 = \mu_2 = \mu_3 \\ H_1: \text{not all equal} \end{aligned}$$

→ if F is big, the effect of A ( $\emptyset_A$ ) is significantly different from 0

→ As interaction effect is included in the “within treatment variance”, absence of interaction effect [random X fixed] is NOT a prerequisite for testing effect of the fixed factor.

# T-tests

- For means tests, use the same denominator as used for the F-test (e.g., instead of MSE, use MSAB when testing for differences in Factor A means)

$$H_0: \mu_2 - \mu_4 = 0 \quad \text{OR} \quad H_0: \mu_2 = \mu_4$$

$$H_1: \mu_2 - \mu_4 \neq 0$$

Change this according to F-test!  
$$t = \frac{(\bar{y}_{\bullet 2} - \bar{y}_{\bullet 4}) - 0}{\sqrt{MSE \left( \frac{1}{n_2} + \frac{1}{n_4} \right)}}$$

- Not relevant to test for differences among means of a Random-effects factor as we are interested in the variance due to that factor

```

> fertdat <-  

read.csv("C:\\\\Users\\\\Rebecca\\\\Documents\\\\Cours\\\\201415\\\\TADE2015\\\\Fertilizat  

iontrial_two_factors.csv",header=TRUE)  

> fertdat$A <- as.factor(fertdat$A)  

> fertdat$B <- as.factor(fertdat$B)  

>  

> model1 <- lm(result~A + B + A*B,data=fertdat)  

> anova(model1)

```

Analysis of Variance Table

Response: result

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	2	1258.17	629.08	514.7045	< 2e-16 ***
B	3	934.75	311.58	254.9318	< 2e-16 ***
A:B	6	17.00	2.83	2.3182	0.05392 .
Residuals	36	44.00	1.22		

---

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

F values calculated with MSE in the denominator



NOT correct  
correct  
correct

Variance due to species is not zero, so species have an impact

Accept H0, no interaction

# Adding random effects in R:

- Many methods, but R is not really fully developed for mixed models! – lots of disagreement
- An intercept only model, where you just estimate the mean of the data:

```
lm(Y ~ 1, data=d)
```

- Lone random effect

```
library(lme4)
```

```
lmer(Y ~ 1 + (1 | A), data=d)
```

Random effects, like  $(1 | A)$ :

a conditional distribution of possible case level intercepts for each level or quantity of A.  
« assume a different intercept depending on the level of A »

# Adding random effects in R:

- A is fixed and B is random (not taking interaction into account):

```
lmer(Y ~ 1 + A + (1 | B), data=d)
```

```
lmer(Y ~ A + (1 | B), data=d) #equivalent, R adds the intercept  
automatically
```

- With an interaction term. The interaction term is random if one of the factors is random:

```
lmer(Y ~ A + (1 | B) + (1 | A*B), data=d)
```

# Now let's get the correct F-value for factor A

```
> anova (model1) $ "Mean Sq" [1] /anova (model1) $ "Mean Sq" [3]
[1] 222.0294
```

*Fixed effect*      *random effects*

```
>library(lme4)
> model<-lmer(result~A + (1 | B) + (1 | A:B) , data=fertdat)
> anova (model)
```

Analysis of Variance Table

	DF	Sum Sq	Mean Sq	F value
A	2	542.74	271.37	222.03

1 stands for the intercept:  
« assume a different intercept for each species »

Significance? P-values?  
→ Multiple approaches and discussion surrounding these...

# Significance of A: calculate P-value for F-value

```
> df_A <- anova(model1)$Df[1]
> df_AB <- anova(model1)$Df[3]
> F_A <- anova(model)$"F value"[1] #the correct F-value from mixed effects
model
> 1- pf(F_A, df_A, df_AB) #this gives the area to the right of this
percentile value
[1] 2.369441e-06
```

# Now let's get the correct F-value for factor A

```
> anova(model1)$"Mean Sq"[1]/anova(model1)$"Mean Sq"[3]
[1] 222.0294
> library(lmerTest)
> model<-lmerTest::lmer(result~A +(1|B)+(1|A:B), data=fertdat)
> anova(model) #this gives the correct F-value for A!!!
Analysis of Variance Table of type 3 with Satterthwaite
approximation for degrees of freedom
  Sum Sq Mean Sq NumDF DenDF F.value    Pr(>F)
A 542.74 271.37      2       6 222.03 2.369e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fixed effect random effects

→

→

1 stands for the intercept:  
« assume a different intercept for  
each species »

Significance? P-values?

→ Multiple approaches and  
discussion surrounding these...

# Significance of A: compare models using the likelihood ratio test

- two models to compare with each other:
- one with the effect in question, one without the effect
- you have to set REML=FALSE. This changes the likelihood estimator

# Significance of A: compare models using the likelihood ratio test

```
> model<-lmer(result~A +(1|B)+(1|A:B),data=fertdat, REML=FALSE)
> model.null <-lmer(result~1 +(1|B)+(1|A:B),data=fertdat, REML=FALSE)
> anova(model,model.null)

Data: fertdat
Models:
model.null: result ~ 1 + (1 | B) + (1 | A:B)
model:      result ~ A + (1 | B) + (1 | A:B)
              Df     AIC     BIC logLik deviance Chisq Chi Df Pr(>Chisq)
model.null   4 213.83 221.31 -102.915    205.83
model        6 183.29 194.52  -85.644    171.29 34.541      2  3.159e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

You would report this result the following way:

“... Fertilization affected result ( $\chi^2=34.541$ ,  $p= 3.159e-08$ )”

# T-tests

See script:

`crd_two_factors_mixed.R`

# **9. Randomized Complete Block design (RCB) – One Factor, fixed effects**

# Randomized Complete Block (RCB) With One Fixed-Effects Factor

## Introduction

Completely randomized design (CRD): complete randomization of experimental units to treatments

- The effects of all possible other variables that might affect the response are, on average, equal in all treatment groups.
- Differences in the group means can be attributed to the treatments.

# In some cases...

- known or suspected in advance, that a variable, not of primary interest to the experimenter, will affect the results
  - possible to group experimental units into clusters (or blocks or strata) where units within a cluster have similar values of this other variable.
- ➔ By changing the experimental design slightly, it is possible to design a more powerful experiment that adjusts for the potential effects of this additional explanatory variable.

# Example: investigate the effect of a drug in lowering blood pressure

## completely randomized design:

- 1/2 of the test subjects are assigned at random to the control group to receive a placebo, and 1/2 of the test subjects are assigned to the drug group
- By randomizing, the effects of other, uncontrolled variables such as amount of exercise, metabolism, diet, etc., would be equal, on average, between the two groups.
- However, these other uncontrolled variables would result in a *large variation in blood pressure within each group* making it harder to detect any changes.

[power for this type of experiment is related to the ratio of the difference in means to the standard deviation within groups.]

Paired design – (more generally: Blocked design: two or more levels randomized within each block)

- treating each subject with both the placebo and the drug (in random order).
- each subject serves as a “control” for these other variables and the difference in blood pressure readings will be free (we hope) of the effects of these other variables.
- This design is not perfect – one still has to worry about
  - ✓ carry-over effects (e.g. the response for the second treatment might be affected by what happened in the first treatment)
  - ✓ the interaction of the blocking factor with the treatment (i.e., perhaps people with high blood pressure react differently to the drug than people with low blood pressure).
- It is possible to block by more than one variable, e.g. the subjects could be further grouped by initial blood pressure levels – this is beyond the scope of this course.

# Randomized Complete Block (RCB) With One Fixed-Effects Factor

- In RCB, treatments are assigned randomly, but only within blocks of treatments
- Restricting randomization of treatments to within blocks (often called sites or trials) is used when the experimental units can be grouped by another variable that may impact the results
- In field experiments with large experimental units, blocking is often very useful in reducing error variance with only a small reduction in error degrees of freedom
- Blocks are most often random effects (we are interested in the variance due to blocks)
- The interest with RCB is with the factor, not with the blocks; the blocks are simply used to reduce the variability among experimental units

# examples of blocked designs:



Honey bee colonies are stacked on pallets, three per pallet. Investigators wish to determine which of three brands of a chemical treatment is most effective in killing a bee mite. They randomly assign the three treatments within each pallet ensuring that each pallet receives all three treatments.



- Blocks:
- Factor:
- Levels (treatments):



# examples of blocked designs:



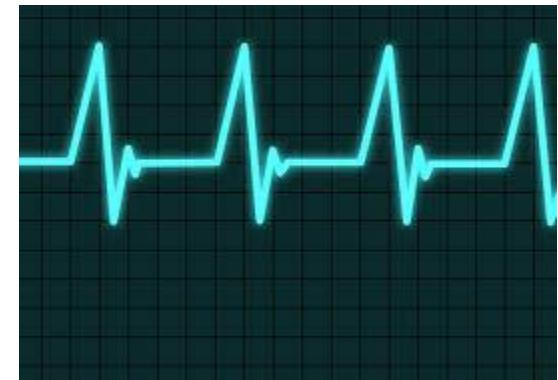
Honey bee colonies are stacked on pallets, three per pallet. Investigators wish to determine which of three brands of a chemical treatment is most effective in killing a bee mite. They randomly assign the three treatments within each pallet ensuring that each pallet receives all three treatments.



- Blocks: pallets
- Factor: the chemical treatment
- Levels (treatments): the three brands



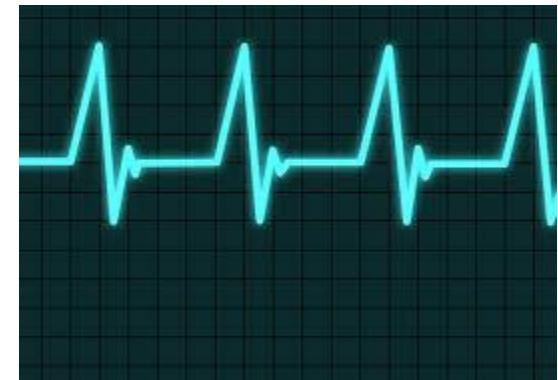
# examples of blocked designs:



Study of change in heart rate before and after exercise.

- Blocks:
- Factor:
- Levels (treatments):

# examples of blocked designs:



Study of change in heart rate before and after exercise.

- Blocks: people
- Factor: time
- Levels (treatments): before and after exercise

(Notice in this experiment, you can't randomize time)

# examples of blocked designs:



Compare the durability of different brands of tires. Because driving habits vary considerably among drivers, you may decide to mount all brands on the same car and do a direct comparison under the same driving conditions, rather than using different cars for each brand with different drivers and (presumably) different driving conditions.

- Blocks:
- Factor:
- Levels (treatments):

# examples of blocked designs:



Compare the durability of different brands of tires. Because driving habits vary considerably among drivers, you may decide to mount all brands on the same car and do a direct comparison under the same driving conditions, rather than using different cars for each brand with different drivers and (presumably) different driving conditions.

- Blocks: driver/car
- Factor: brand of tire
- Levels (treatments): the particular brands chosen in the experiment

# examples of blocked designs:

- Factor A: six levels of fertilization: A1 to A6
- two sites
- Randomization of Factor A is restricted to within sites.
- Response variable: biomass of grasses and herbs (kg)
- 2 observations per treatment – 1 in each site

Site 1		Site 2	
A1 = 9	A6=21	A4=25	A3=19
A3=15	A2=12	A1=12	A5=27
A5=20	A4=17	A2=16	A6=29

# *Organization of data for analysis using a statistics package:*

Site	Treatment	$y_{jk}$
1	A1	9
	A2	12
	A3	15
	A4	17
	A5	20
	A6	21
2	A1	12
	A2	16
	A3	19
	A4	25
	A5	27
	A6	29

# Main questions of interest:

- Are the treatment means different?
- Which means are different?
- What are the estimated means and confidence intervals for these estimates?

As for CRD with one factor

- The organization of the data is the same for CRD with **two** factors as with RCB, BUT the **interpretation** differs:
- It is assumed that there is no interaction between the blocks and the treatments.
- Blocks are usually considered random-effects; want to remove the effects of blocks from the analysis

# Remember this:

- One observation per cell

		Factor B				
		b1	b2	b3		
		+-----+	+-----+	+-----+		
Factor A		a1	x	x	x	
		+-----+	+-----+	+-----+	+-----+	
a2		x	x	x		
		+-----+	+-----+	+-----+	+-----+	

- Impossible to test interaction effect
- MUST ASSUME that no interaction exists
- fit a model without any interaction terms.

# What is the difference between treating blocks as simply blocks or treating blocks as another factor?

In some cases, the creation of blocks is dependent upon a variable that looks like a factor.  
example: blocks formed based upon (hypothesized) fertility differences among fields.

If blocks are formed upon (hypothesized) differences in fertility among plots:

- NOT necessary to measure the actual fertility levels
- NOT necessary to restrict the number of fertility levels to a small number of levels

If you treat fertility as a factor (two-factor design), you

- would be forced to measure the fertility of each plot of land
- would normally only have a few levels (e.g. low, medium, high)
- would like to have replicates of each level of fertility

# Randomization protocol (A single factor, randomized complete block design (RCB))

- within each block, experimental units are as similar as possible
- independently within each block, experimental units are completely randomized to treatments such that
  - ✓ every treatment occurs **once and only once in each block**
  - ✓ All treatments occur in each block

# Example: Importance of complete randomization within blocks:

- in a drug study to compare the blood pressure of a drug vs. the blood pressure when taking a placebo, the placebo may always be given before the drug. In this case, it is impossible to know if the drug caused any change in mean blood pressure; perhaps the different amount of sunlight between the two occasions caused the change.

# Examples of several possible designs for a problem

- A study of the effects of supplemental fertilization on the growth of seedlings after planting.
- Three levels of supplemental fertilization will be used (none, low, or high amounts of fertilization).
- The experiment will be conducted at six different test sites.
- At each location, three recently replanted forest plots are available for use in the experiment.

# Completely randomized design - no blocking

- The experimental units were completely randomized to the treatments ignoring the site groupings.
- There is no guarantee that each site receives each treatment level.
- extra variation because of different locations

Site	Plot number at each site		
	1	2	3
-----			
1	0	high	0
2	low	low	high
3	high	low	0
4	low	0	high
5	low	0	low
6	0	high	high
-----			

# Randomized complete block design - RCB design

- Within each block, all treatments occur once and only once
- within each block randomization was performed independently of the randomization in other blocks.

Site	Plot number at each site		
	1	2	3
-----			
1	0	high	low
2	low	0	high
3	high	low	0
4	low	0	high
5	high	0	low
6	0	high	low
-----			

# Generalized randomized complete block design

- the blocks are large enough to have some or all of the treatments replicated within each block.
- This provides additional information about the variability of treatments within blocks.
- the fact that you now know the variability of responses within blocks for the same treatment, allows a statistical test to examine if the block-treatment additivity assumption holds in this experiment.

Site	Plot number at each site			
	1	2	3	4
-----				
1	0	high	low	low
2	low	0	high	0
3	high	low	0	0
4	low	0	high	high
5	high	0	low	high
6	0	high	low	low
-----				

# Notation:

Population:  $y_{jk} = \mu + \tau_{Bj} + \tau_{Ak} + \varepsilon_{jk}$

$y_{jk}$  = response variable measured on block j and treatment k

j=1 to J blocks; k=1 to K treatments

$\mu$  = the grand or overall mean regardless of treatment or block

$\tau_{Ak}$  = the treatment effect for k

$\tau_{Bj}$  = the block effect for block j

$\varepsilon_{jk}$  = is actually an interaction term between block and treatment, defined as:

$$\varepsilon_{jk} = y_{jk} - (\mu + \tau_{Ak} + \tau_{Bj})$$

For the experiment:  $y_{jk} = \bar{y}_{..} + \hat{\tau}_{Bj} + \hat{\tau}_{Ak} + e_{jk}$

$\bar{y}_{..}$  = the grand or overall mean of all measures from the experiment regardless of treatment; under the assumptions for the error terms, this will be an unbiased estimate of  $\mu$

$\bar{y}_{j.}$  = the mean of all measures from the experiment for a particular block  $j$  (includes all data for all levels of the treatment)

$\bar{y}_{.k}$  = the mean of all measures from the experiment for a particular treatment  $k$  over all blocks

$\hat{\tau}_{Ak}, \hat{\tau}_{Bj}$  = under the error term assumptions, will be unbiased estimates of corresponding treatment effects for the population

$$\begin{aligned} e_{jk} &= (y_{jk} - \bar{y}_{..}) - (\bar{y}_{j.} - \bar{y}_{..}) - (\bar{y}_{.k} - \bar{y}_{..}) \\ &= y_{jk} - \bar{y}_{j.} - \bar{y}_{.k} + \bar{y}_{..} \end{aligned}$$

J= number of blocks **and also the number of measures (experimental units) for treatment k**

KJ = total number of experimental units on which the response was measured

Sum of squares:  $SSy = SS_{BLK} + SS_{TR} + SSE$

$SSy$ : The sum of squared differences between the observations and the grand mean:

$$SSy = \sum_{k=1}^K \sum_{j=1}^J (y_{jk} - \bar{y}_{..})^2 \quad df = JK - 1$$

$SS_{BLK}$ : Sum of squared differences between the block means, and the grand mean, weighted by the number of treatments (experimental units in each block)

$$SS_{BLK} = \sum_{j=1}^J K (\bar{y}_{j\bullet} - \bar{y}_{..})^2 \quad df = J - 1$$

$SS_{TR}$ : Sum of squared differences between the treatment means, and the grand mean, weighted by the number of blocks (experimental units in each treatment)

$$SS_{TR} = \sum_{k=1}^K J (\bar{y}_{\bullet k} - \bar{y}_{..})^2 \quad df = K - 1$$

$SSE$ : sum of squared differences between the observation and the grand mean minus the treatment and block effects.

$$SSE = SS_y - SS_{TR} - SS_{BLK} \quad df = (J-1)(K-1)$$

*Alternative computational formulae:*

$$SS_y = \sum_{k=1}^K \sum_{j=1}^J y_{jk}^2 - \frac{y_{\bullet\bullet}^2}{JK}$$

$$SS_{TR} = J \sum_{k=1}^K \bar{y}_{\bullet k}^2 - \frac{y_{\bullet\bullet}^2}{JK} \quad SS_{BLK} = K \sum_{j=1}^J \bar{y}_{j\bullet}^2 - \frac{y_{\bullet\bullet}^2}{JK}$$

$$SSE = SS_y - SS_{TR} - SS_{BLK}$$

# Assumptions

## 1. Additivity between blocks and treatments

THIS IS A CRUCIAL ASSUMPTION that allows the analysis to be interpreted!

- means are  $\mu + \alpha_j + \beta_k$
- This assumption states that the difference in the mean response between any two treatments is the same in all blocks.
- Another name for the assumption of additivity is **no interaction between treatments and blocks**.

- Example:

### Population means under assumption of additivity

Block	Treatment		
	a	b	c
1	10	20	15
2	35	45	40

- Note that the above table refers to POPULATION means - the sample means may not enjoy this strict additivity as an artifact of the sampling process.

- Example where additivity does not hold but is correctable:

Population means when the assumption of additivity is false but correctable

Block	Treatment		
	a	b	c
1	10	20	15
2	100	200	150

- the mean for treatment b is always twice the mean for treatment a in both blocks.
- This suggests that the effects of treatment are multiplicative rather than additive, and that the analysis should proceed on the log-scale.

- Indeed, consider the same values after a log-transformation:

Population means when the assumption of additivity is false but correctable. A log-transform

Block	Treatment		
	a	b	c
1	2.30	3.00	2.71
2	4.61	5.30	5.01

- The treatment effects are now additive on the log-scale

- In some cases, no transformation will correct non-additivity:

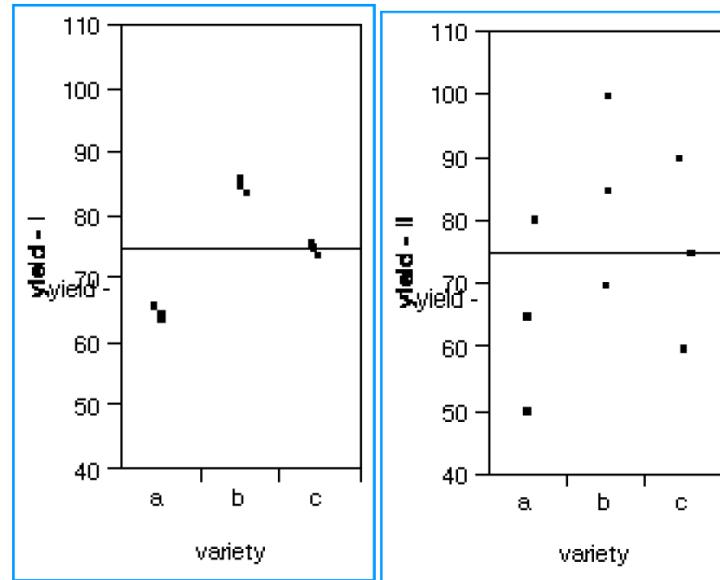
Population means when the assumption of additivity is false and not correctable.

Block	Treatment		
	a	b	c
1	10	20	15
2	35	30	25

- the response is generally higher in block 2 than in block 1
- but the difference between the mean response for treatment a and treatment b changes in the two blocks.  
= interaction between treatments and blocks.

## 2. No outliers should be present

- This is the same assumption as in the CRD
- the idea behind the tests for equality of means (ANOVA) is to compare the relative variation among means to the variation within each group.



- Outliers can severely distort estimates of the within-group variation and severely distort the results of the statistical test.

3. Equal treatment group standard deviations? (homoscedasticity)
4. Are the errors normally distributed within each treatment group?
5. Are the errors independent?

- These assumptions are also the same as in the CRD
- Use residual plot and a plot of the standardized errors against the expected errors for a normal distribution to check these assumptions.
- To meet assumptions you might have to transform the y-variable, as with other designs

# Differences among treatment means

The main question is:

Is there a difference between treatment means:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_K$$

OR:

$$H_0: \frac{(\phi_{TR} + \sigma_\varepsilon^2)}{\sigma_\varepsilon^2} = 1$$

$$H_1: \frac{(\phi_{TR} + \sigma_\varepsilon^2)}{\sigma_\varepsilon^2} > 1$$

Where  $\sigma_\varepsilon^2$  is the variance of the error terms;

$\phi_{TR}$  is fixed effect for the treatments.

# Using an analysis of variance table:

Source	df	SS	MS	F	p-value
Block	J-1	$SS_{BLK}$	$MS_A = SS_{BLK} / (J-1)$		
Treat.	K-1	$SS_{TR}$	$MS_{TR} = SS_{TR} / (K-1)$	$F = MS_{TR}/MSE$	$Prob F > F_{(K-1),(dfE),1-\alpha}$
Error	(J-1)(K-1)	SSE	$MSE = SSE/(J-1)(K-1)$		
Total	JK - 1	SSy			

Source	df	MS	$E[MS]$
Block	J-1	$MS_{BLK}$	$\sigma_\varepsilon^2 + K\sigma_{BLK}^2$
Treat.	K-1	$MS_{TR}$	$\sigma_\varepsilon^2 + \phi_{TR}$
Error	(J-1)(K-1)	MSE	$\sigma_\varepsilon^2$
Total	$n_T - 1$		

$\phi$  is used here to represent fixed effects

$\sigma^2$  is used here to represent random effects

# From the ANOVA table

$$F = \frac{SS_{TR} / (K - 1)}{SSE / (J - 1)(K - 1)} = \frac{MS_{TR}}{MSE}$$

Under  $H_0$ , this follows  $F_{df1, df2, 1-\alpha}$  where  $df1$  is from the numerator ( $K-1$ ) and  $df2$  is from the denominator ( $J-1$ ) ( $K-1$ )

If the  $F$  calculated is greater than the tabular  $F$ , or if the  $p$ -value for  $F$  calculated is less than  $\alpha$ , reject  $H_0$ , the means of treatments in the population are likely not all the same

# Further analyses

Can conduct multiple comparisons between means for the  $K$  treatments:

- using MSE and using **J (number of blocks) as the number of observations per treatment.**
- Can use t-tests of pairs of means -- must divide alpha by the number of possible pairs

*Confidence limits for treatment means*

$$\bar{y}_{\bullet k} \pm t_{(dfE),1-\alpha/2} \sqrt{\frac{MSE}{J}}$$

two seemingly different experimental procedures and analyses:

1. Paired design

There are two treatment levels

A paired t-test is used to analyze the data.

2. Blocked design

There are two or more treatment levels.

An ANOVA is used to analyze the data.

The paired t-test is a special case of a more general ANOVA approach and the two approaches will give identical results for designs with exactly 2 treatment levels.

In cases with more than 2 levels, the paired approach cannot be used.

# Comparing two means in a paired design - the Paired t-test

- a special case of a more general ANOVA approach :

the analyst realizes the variation among experimental units may be very large compared to the difference in response caused by the treatment

→ conduct the experiment by doing both treatments on each experimental unit.

The data are “paired” : a pair of observations from each experimental unit

# Examples:

- measuring every person's change in heart-rate using two different walking styles (in random order).
- measuring every person before and after a drug is administered.
- administering every person two drugs (in random order) and measuring the heart rate after each drug is administered.

Compare this to completely randomized design where separate experimental units are used for every treatment.

- R offers three basic approaches for the analysis of a simple paired experiment – illustrated below

# Example - effect of stream slope upon fish abundance

- Based upon paper:

Isaak, D.J. and Hubert, W.A. (2000). Are trout populations affected by reach-scale stream slope. Canadian Journal of Fisheries and Aquatic Sciences, 57, 468-477.



# effect of stream slope upon fish abundance: Introduction to the problem

- “stream reach”: a portion of a stream that exhibits consistent slope.
- The slope influences the speed of the water, which exerts an influence on the structure of the habitat.
- Reach-scale stream slope and the structure of associated habitats are thought to affect trout populations
- If measuring many streams distributed across space and time, inter-stream differences and temporal variation in trout populations result in extra variation which reduces the power of the survey to detect effects.
- For this reason, a paired approach was taken.
- A total of twenty-three streams were sampled from a large watershed.
- Within each stream, two reaches were identified in a “low slope” environment and a “high slope” environment.
- In each reach, fish abundance was determined and the numbers converted to a density per 100 m<sup>2</sup> of stream surface.

Raw data:  
paired\_stream.csv

= Observational study: randomization occurs by selecting streams at random from some larger population of potential streams.

Notice: density varies considerably among stream but appears to be fairly consistent within each stream

Stream	slope (%)	Slope class	Density (per 100 m <sup>2</sup> )
1	0.7	low	15
1	4	high	21
2	2.4	low	11
2	6	high	3.1
3	0.7	low	5.9
3	2.6	high	6.4
4	1.3	low	12.2
4	4	high	17.6
5	0.6	low	6.2
5	4.4	high	7
6	1.3	low	39.8
6	3.2	high	25
7	2	low	6.5
7	4.2	high	11.2
8	1.3	low	9.6
8	4.2	high	17.5
9	2	low	7.3
9	3.6	high	10
10	0.7	low	11.3
10	3.5	high	21
11	2.3	low	12.1
11	6	high	12.1

12	2.5	low	13.2
12	4.2	high	15
13	2.3	low	5
13	6	high	5
14	1.2	low	10.2
14	2.9	high	6
15	0.7	low	8.5
15	2.9	high	7
16	1.1	low	5.8
16	3	high	5
17	2.2	low	5.1
17	5	high	5
18	0.7	low	65.4
18	3.2	high	55
19	0.7	low	13.2
19	3	high	15
20	0.3	low	7.1
20	3.2	high	12
21	2.3	low	44.8
21	7	high	48
22	1.8	low	16
22	6	high	20
23	2.2	low	7.2
23	6	high	10.1

## Script: paired\_stream.R

stops the automatic conversion  
of character variables to factors

```
> fish <- read.csv("paired_stream.csv", header=TRUE, as.is=TRUE)
> fish[1:12,]
  stream slope slopeclass density
1       1   0.7        low    15.0
2       1   4.0       high    21.0
3       2   2.4        low    11.0
4       2   6.0       high     3.1
5       3   0.7        low     5.9
6       3   2.6       high     6.4
7       4   1.3        low    12.2
8       4   4.0       high    17.6
9       5   0.6        low     6.2
10      5   4.4       high     7.0
11      6   1.3        low    39.8
12      6   3.2       high    25.0
```

# 1. Using a Differences analysis:

- the difference in density is computed for each stream between the low and high slope sections.
- The relevant hypothesis is that the mean difference is zero.
- A one-sample t-test is used to test the hypothesis and to estimate the average difference

## 1. Specify the hypothesis

Let  $\mu_{diff}$  be the population mean of the difference in densities:

$$H_0: \mu_{diff} = 0$$

$$H_1: \mu_{diff} \neq 0$$

## 2. Collect the data: need to split the dataset

```
> low.slope <- fish[ fish$slopeclass == "low" , ]
> names(low.slope)[names(low.slope)== "density"]<- "low.density"
> high.slope <- fish[ fish$slopeclass == "high", ]
> names(high.slope)[names(high.slope)== "density"]<- "high.density"
>
> paired.slope <- merge(low.slope [,c("stream","low.density")],
+                         high.slope[,c("stream","high.density")],
by="stream")
> paired.slope
```

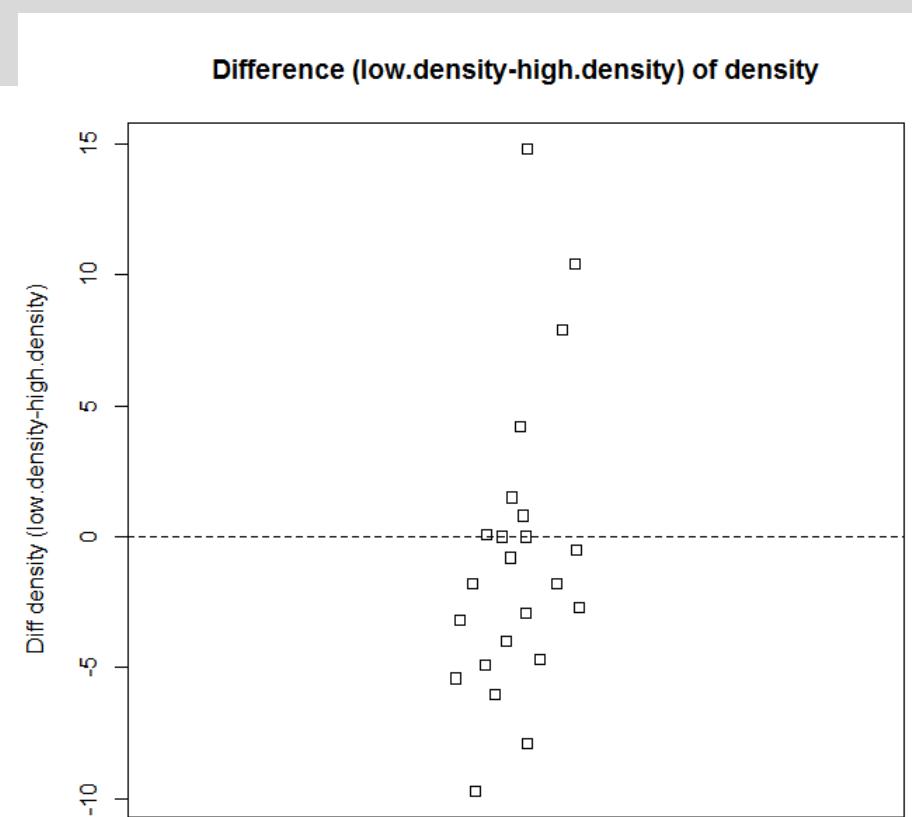
	stream	low.density	high.density
1	1	15.0	21.0
2	2	11.0	3.1
3	3	5.9	6.4
4	4	12.2	17.6
5	5	6.2	7.0
6	6	39.8	25.0
7	7	6.5	11.2
8	8	9.6	17.5
9	9	7.3	10.0
10	10	11.3	21.0
11	11	12.1	12.1
12	12	13.2	15.0
13	13	5.0	5.0
14	14	10.2	6.0
15	15	8.5	7.0
16	16	5.8	5.0
17	17	5.1	5.0
18	18	65.4	55.0
19	19	13.2	15.0
20	20	7.1	12.0
21	21	44.8	48.0
22	22	16.0	20.0
23	23	7.2	10.1

- Create a new variable in the dataframe for the difference between the two densities measured on each stream:

```
> paired.slope$diff <- paired.slope$low.density - paired.slope$high.density  
> paired.slope  
   stream low.density high.density diff  
1       1      15.0        21.0 -6.0  
2       2      11.0        3.1  7.9  
3       3       5.9        6.4 -0.5  
4       4      12.2        17.6 -5.4  
5       5       6.2        7.0 -0.8  
6       6      39.8        25.0 14.8  
7       7       6.5        11.2 -4.7  
8       8       9.6        17.5 -7.9  
9       9       7.3        10.0 -2.7  
10     10      11.3        21.0 -9.7  
11     11      12.1        12.1  0.0  
12     12      13.2        15.0 -1.8  
(.....etc.)
```

### 3. Compute a test statistic, p-value, and confidence intervals

```
> stripchart(paired.slope$diff,  
+    vertical=TRUE, method="jitter", jitter=.1,  
+    main="Difference (low.density-high.density) of density",  
+    xlab="", ylab="Diff density (low.density-high.density)")  
> abline(h=0, lty=2)
```



### 3. Compute a test statistic, p-value, and confidence intervals

```
> t.test(paired.slope$diff)

  One Sample t-test

data: paired.slope$diff
t = -0.6071, df = 22, p-value = 0.55
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-3.187395 1.743917
sample estimates:
mean of x
-0.7217391
```

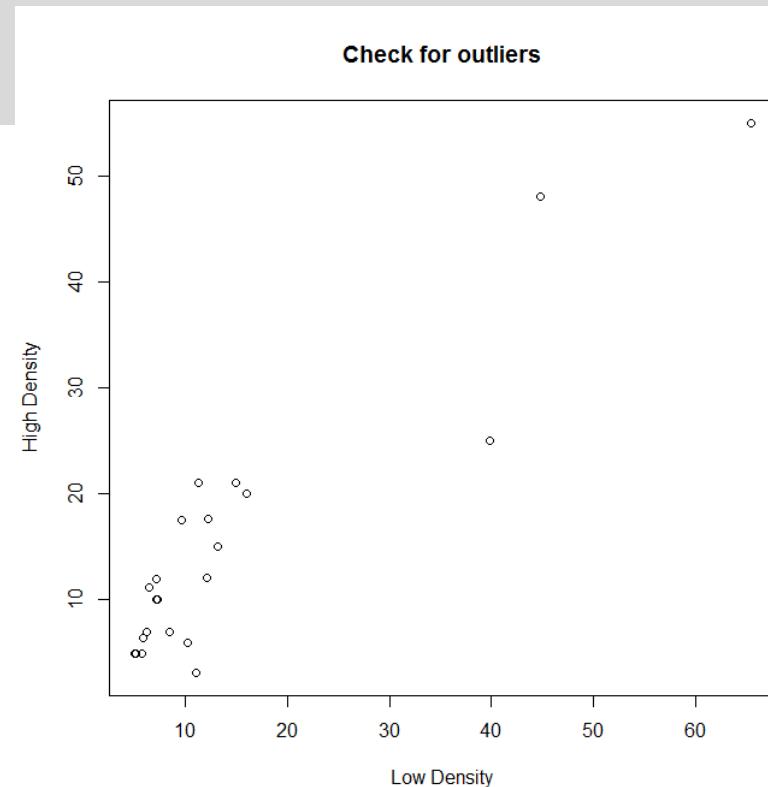
#### 4. Make a decision

- Because the confidence interval includes the value 0, there is no evidence of a difference in the mean density between high and low slope reaches.
- The p-value of 0.55 also indicates that we would conclude that there is no evidence of a difference in the mean density between the high and low slope reaches.

## 2. Using a Matched paired analysis

- Not necessary to create the variable « differences »
- Check for outliers like this:

```
> plot(paired.slope$low.density, paired.slope$high.density,  
+       main="Check for outliers",  
+       xlab="Low Density", ylab="High Density")
```



## 2. Using a Matched paired analysis

- Functionally identical to the analysis of the differences:

```
> t.test(paired.slope$low.density, paired.slope$high.density, paired=TRUE)

Paired t-test

data: paired.slope$low.density and paired.slope$high.density
t = -0.6071, df = 22, p-value = 0.55
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-3.187395 1.743917
sample estimates:
mean of the differences
-0.7217391
```

### 3. Using a General Modeling analysis

- Two sources of variation in the model:
  1. Variation that comes from treatment structure: The different slope classes (high or low gradient) in the stream
  2. Variation that comes from experimental unit structure:
    - Blocks (explicitly specified: the stream measured)
    - experimental units within blocks (implicitly assumed)
- The crucial assumption of additivity between treatments and blocks is implicit and is specified **by not specifying an interaction term.**

1. Specify the hypothesis

$$H_0: \mu_{low} = \mu_{high}$$

$$H_1: \mu_{low} \neq \mu_{high}$$

2. Collect the data:

- The data frame must have separate columns for
  - ✓ the factor slopeclass
  - ✓ the blocking variable stream
  - ✓ the response variable density
- Both variables must be factors.
- Factor conversion must be explicitly set using the `factor()` function.
- The `str()` function allows you to check that the variables have been properly set in the data frame.

*(if the Stream variable is NOT declared as a factor, you will fit a model where the stream number is used as a regression term (rather than an ANOVA term). )*

```
> fish$slopeclass <- factor(fish$slopeclass)
> fish$stream       <- factor(fish$stream)
> str(fish)
'data.frame':   46 obs. of  4 variables:
 $ stream      : Factor w/ 23 levels "1","2","3","4",...: 1 1 2 2 3 3 4 4 5 5
 ...
 $ slope       : num  0.7 4 2.4 6 0.7 2.6 1.3 4 0.6 4.4 ...
 $ slopeclass: Factor w/ 2 levels "high","low": 2 1 2 1 2 1 2 1 2 1 ...
 $ density     : num  15 21 11 3.1 5.9 6.4 12.2 17.6 6.2 7 ...
```

### 3. Compute a test statistic, p-value, and confidence intervals

- in R the blocking variable must come FIRST in the model and the factor variable second! (although in other programs terms can come in any order)

```
> model <- lm(density ~ stream + slopeclass, data=fish)
> anova(model)
Analysis of Variance Table

Response: density
                         Df Sum Sq Mean Sq F value    Pr(>F)
stream                 22 8229.4  374.06 23.0115 1.563e-10 ***
slopeclass              1    6.0     5.99  0.3685          0.55
Residuals                22  357.6   16.26
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$= 5.99/16.26$$

- Check the assumptions

Same as for CRD: residuals plot

The statistical significance of each effect in the model is summarized by a series of F-statistics.

- When there are only two levels in a factor, F-statistic is equivalent to the (T-statistic)<sup>2</sup>:  $0.368 = 0.6071^2$
- The p-value of 0.55 is identical to the previous results
- The test for a blocking variable is automatically computed, but is rarely of interest

- We obtain estimates of the individual group means using the `lsmeans()` function from the `lsmeans` package

```
# Create the lsmeans object that is used in subsequent computations and
# obtain basic estimates of the marginal means
model.lsmo <- lsmeans(model, ~slopeclass, adjust="tukey")
summary(model.lsmo, infer=TRUE)

slope.class lsmean SE df lower.CL upper.CL t.ratio p.value
high 15.43478 0.840689 22 13.69130 17.17826 18.360 <.0001
low 14.71304 0.840689 22 12.96956 16.45653 17.501 <.0001
Results are averaged over the levels of: Stream
Confidence level used: 0.95
> sqrt(16.26/23)
[1] 0.8408071
```

- The set of all pairwise differences in the means is found using the `pairs()` function applied to the `lsmeans` object

- The set of all pairwise differences in the means is found using the pairs() function applied to the lsmeans object

```
# Find all the pairwise differences adjusting for multiplicity
model.pairs <- pairs(model.lsmo, adjust="tukey")
summary(model.pairs, infer=TRUE)
contrast estimate SE df lower.CL upper.CL t.ratio p.value
high - low 0.7217391 1.188914 22 -2.765225 4.208704 0.607 0.5500
Results are averaged over the levels of: Stream
Confidence level used: 0.95
```

#### 4. Make a decision

→ The same decision

Which analysis is best? Doesn't matter

- Most important is the recognition that pairing (or blocking) has taken place. If this is missed, then all the subsequent output is incorrect.

# Single Factor - Randomized Complete Block (RCB) Design

- Is an extension of the paired design to the case of two or more treatments
- The basic idea of an analysis for a RCB is to again analyze the differences, rather than the individual observations.

# Example - Comparing effects of salinity in soil

- We wish to investigate the effect of salinity in the soil on the growth of salt marsh plants.
- Experimental fields of land were located at an agricultural field station, and each field was divided into six smaller plots.
- Each of the smaller plots was treated with a different amount of salt (measured in ppm)
- The biomass at the end of the experiment was recorded.



## Experimental layout:

Salinity				
Field 1	Field 2	Field 3	Field 4	
15	30	10	20	
25	35	15	25	
30	15	25	30	
20	10	20	10	
10	25	30	15	
35	20	35	35	

- every block (the field) has every treatment (all six salt concentrations)
- the randomization in each block was done independently of every other block

# But salinity is a numerical value?

The key difference between treating the levels of the factor as qualitative (e.g. nominal or ordinal scale) vs. quantitative (e.g. continuous scale) is how you think the response variable (yield) varies in relation to the factor.

If you believe that the response is linear (and only linear) then treating the salt levels as quantitative (continuous) is appropriate : a regression analysis

Treating the salt levels as qualitative (e.g. nominal or ordinal scale) allows for ANY response function other than a straight line with 0 slope (which corresponds to zero effect).

→ detect a threshold effect, or a quadratic effect, or any other pattern in addition to a linear effect by doing a ANOVA rather than a REGRESSION.

Raw data:

Salt	Yield			
	Field 1	Field 2	Field 3	Field 4
10	11.8	15.1	22.6	7.1
15	21.3	22.3	19.8	9.9
20	8.8	8.1	6.1	1
25	10.4	8.5	8.2	2.8
30	2.2	3.3	6.1	0.7
35	8.4	7.3	5.2	2.2

- every block (the field) has every treatment (all six salt concentrations)
- the randomization in each block was done independently of every other block

```
> marsh <- read.csv("salt.csv",
header=TRUE, as.is=TRUE)
> print(marsh[1:12,])
  salt biomass block
1    10     11.8 Field1
2    15     21.3 Field1
3    20      8.8 Field1
4    25     10.4 Field1
5    30      2.2 Field1
6    35      8.4 Field1
7    10     15.1 Field2
8    15     22.3 Field2
9    20      8.1 Field2
10   25      8.5 Field2
11   30      3.3 Field2
12   35      7.3 Field2
```

Raw data: salt.csv  
Script: salt.R

Both the salt and block variable must be defined as factors  
(otherwise you will do a regression rather than ANOVA, and no error message will tell you)

```
> marsh$salt    <- factor(marsh$salt)
> marsh$block   <- factor(marsh$block)
> str(marsh)
'data.frame':   24 obs. of  3 variables:
 $ salt      : Factor w/ 6 levels
 "10","15","20",...: 1 2 3 4 5 6 1 2 3 4
 ...
 $ biomass: num  11.8 21.3 8.8 10.4 2.2
 8.4 15.1 22.3 8.1 8.5 ...
 $ block     : Factor w/ 4 levels
 "Field1","Field2",...: 1 1 1 1 1 1 2 2 2 2
 ...
```

## 1. Specify the hypothesis

$$H_0: \mu_{10} = \mu_{15} = \cdots = \mu_{35}$$

$H_1$ : at least one mean is different from the rest

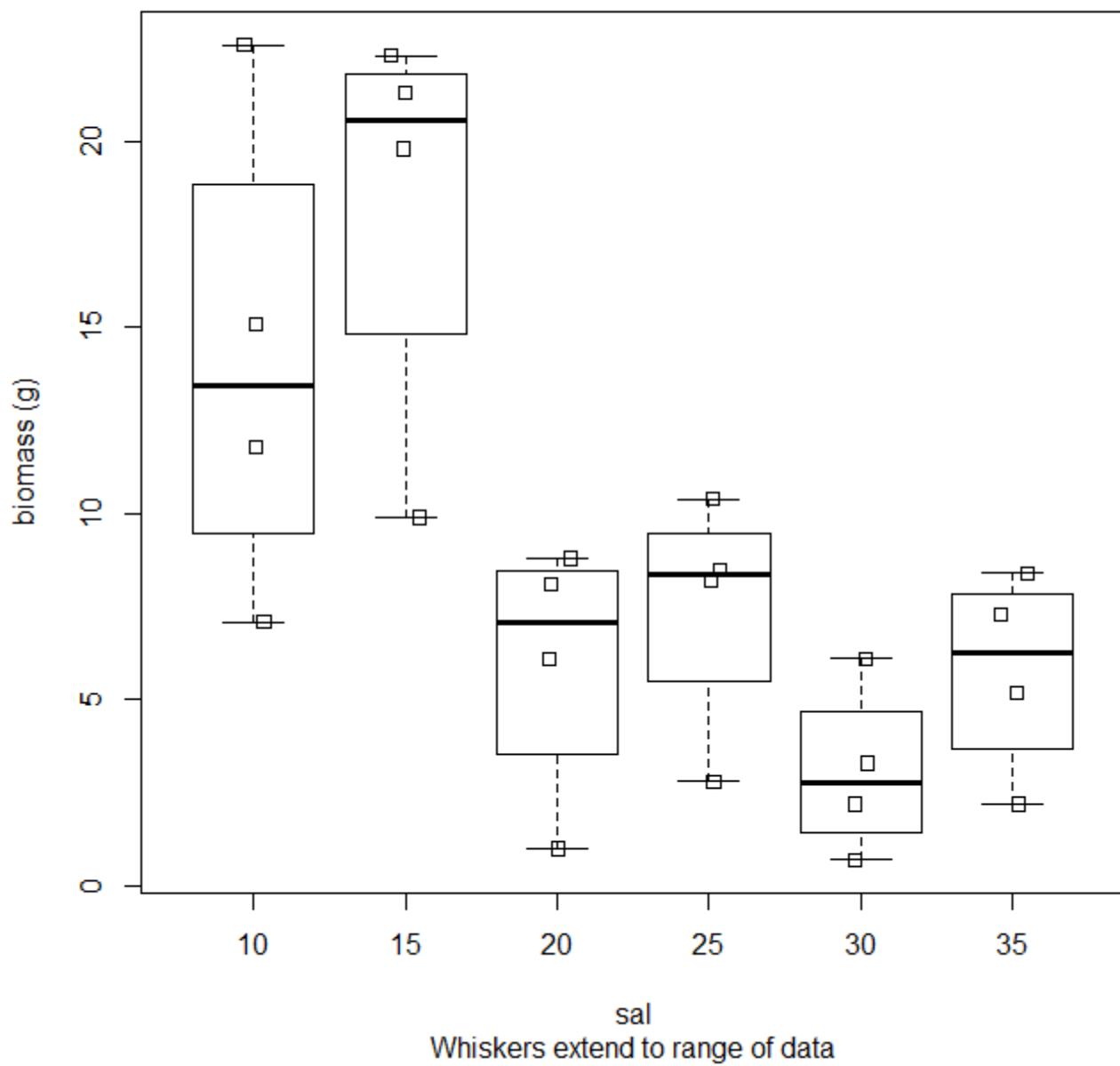
→ Same as in single factor CRD ANOVA

## 2. Collect the data and compute some summary statistics

Get a side by side boxplot

```
> boxplot(biomass ~ salt, data=marsh, range=0,
+   main="biomass at various salts",
+   sub="Whiskers extend to range of data",
+   xlab="sal", ylab="biomass (g)")
>
> stripchart(biomass ~ salt, data=marsh, add=TRUE,
+   vertical=TRUE, method="jitter", jitter=.1)
```

### biomass at various salts

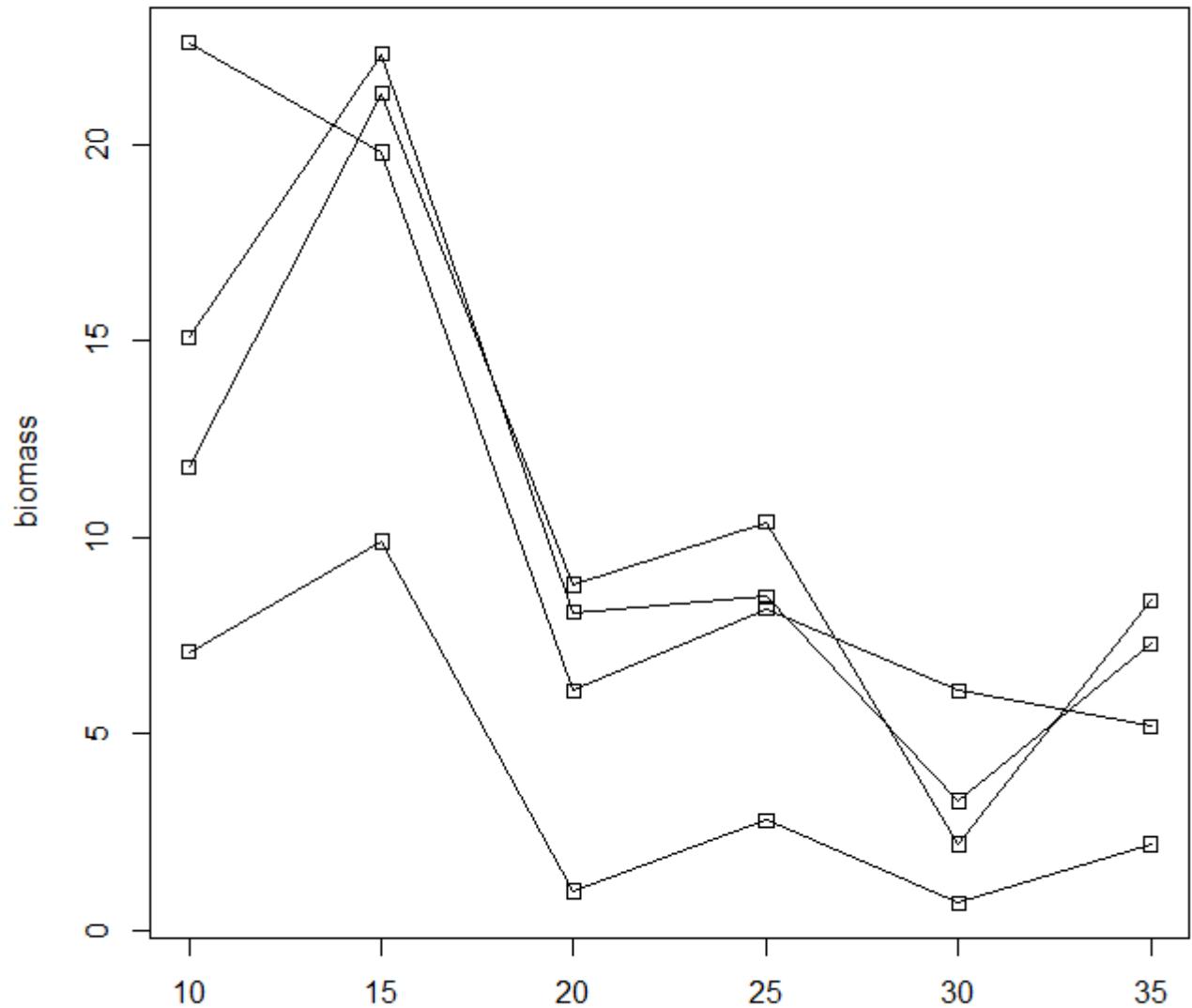


## Get a profile plot to check the additivity

```
> stripchart(biomass ~ salt, data=marsh, add=TRUE,  
+             vertical=TRUE, method="jitter", jitter=.1)  
> library(plyr)  
> d_ply(marsh,"block",  
+        function(x) {points(as.numeric(x$salt), x$biomass, type="l") })  
> # Do a plot of the biomass over the salts by sample to check for  
additivity  
> stripchart(biomass ~ salt, data=marsh, vertical=TRUE)  
>  
> library(plyr)  
> d_ply(marsh,"block",  
+        function(x) {points(as.numeric(x$salt), x$biomass, type="l") })
```

- There appears to be an effect of salt concentration, particularly when the concentration is 20 ppm or higher.
- It appears that there is not much a difference between Blocks 1-3, but Block 4 appears to have a consistently lower response than the other blocks.
- It appears that the treatment and block effects are additive as the lines are roughly parallel.

→ exactly situation for RCB



We get the means and standard deviations and check to see that the sample standard deviations are about equal in all factor levels:

```
> #first create the function "sf.simple.summary" using  
"helpoerfunctions.R"  
> report <- ddply(marsh, "salt", sf.simple.summary, variable="biomass")  
> report  
salt n nmiss      mean        sd  
1   10 4       0 14.150 6.519969  
2   15 4       0 18.325 5.709860  
3   20 4       0  6.000 3.524202  
4   25 4       0  7.475 3.265348  
5   30 4       0  3.075 2.280899  
6   35 4       0  5.775 2.728095
```

#### 4. Find the test-statistic and compute a p-value:

```
> model <- lm(biomass ~ block + salt, data=marsh)
> anova(model)
Analysis of Variance Table

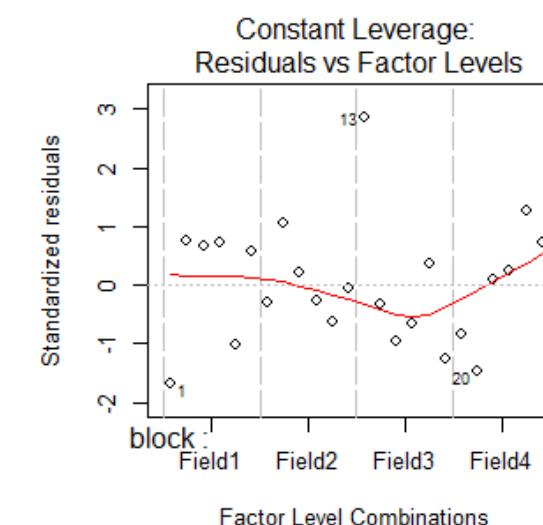
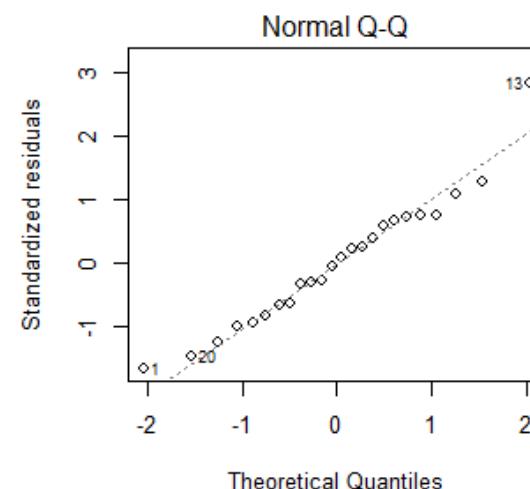
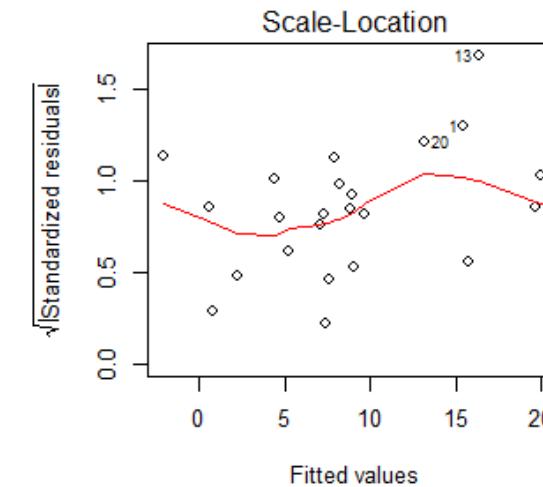
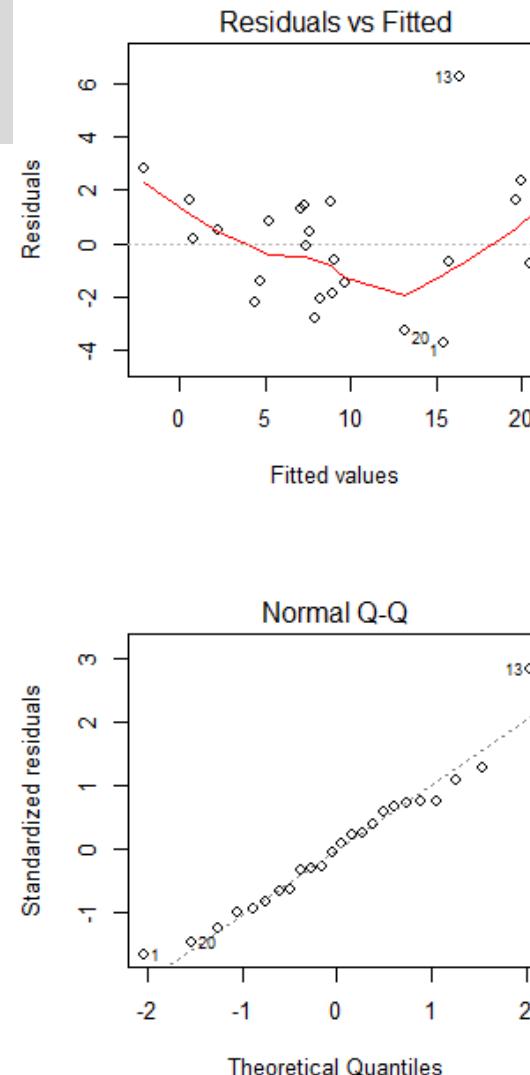
Response: biomass
            Df  Sum Sq Mean Sq F value    Pr(>F)
block        3 217.18  72.394  9.4152 0.0009603 ***
salt         5 680.81 136.163 17.7085 8.076e-06 ***
Residuals 15 115.34   7.689
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1
>
```

- The F-ratio is 17.7084 and the p-value is < 0.0001.
- We ignore the test for blocks as these are NOT an experimental factor in the design.

- Check the assumptions of equal variance and normality of residuals

```
> layout(matrix(1:4, nrow=2))
> plot(result)
> layout(1)
```

There is some evidence that the residuals are not evenly distributed over the salt concentrations, but this isn't too bad.



## 5. Make a decision

Because the p-value is small, we conclude that there is evidence that the mean biomass differs among the salt concentrations.

## 6. If you find evidence against the null hypothesis, use a multiple comparison procedure

We obtain estimates of the individual group means using the `lsmeans()` function from the `lsmeans` package

NOTE that the SE are NOT computed as simply  $s/\sqrt{n}$ .

Every design leads to a different formula for computing standard errors.

```
> library(lsmeans)
> model.lsmo <- lsmeans(model, ~salt, adjust="tukey") Requests also a Tukey multiple
> #not adjusted for simultaneous coverage: comparison procedure
> summary(model.lsmo, infer=TRUE)
```

salt	lsmean	SE	df	lower.CL	upper.CL	t.ratio	p.value
10	14.150	1.386462	15	11.1948255	17.105174	10.206	<.0001
15	18.325	1.386462	15	15.3698255	21.280174	13.217	<.0001
20	6.000	1.386462	15	3.0448255	8.955174	4.328	0.0006
25	7.475	1.386462	15	4.5198255	10.430174	5.391	0.0001
30	3.075	1.386462	15	0.1198255	6.030174	2.218	0.0424
35	5.775	1.386462	15	2.8198255	8.730174	4.165	0.0008

T-tests for means  
being different of  
zero

Results are averaged over the levels of: block

Confidence level used: 0.95

- subtle difference between the Least Square Mean and the Mean column that is a concern when the data is unbalanced (i.e., some blocks are missing some treatments).

## To see the results of the Tukey multiple comparison procedure: (The 'multcompView' package must be installed)

```
> library(multcompView)
Loading required package: grid
> # Get the compact letter display and a plot
> model.cld <- cld(model.lsmo)
> model.cld
   salt lsmean      SE df lower.CL upper.CL .group
 30    3.075 1.386462 15  0.1198255  6.030174     1
 35    5.775 1.386462 15  2.8198255  8.730174     1
 20    6.000 1.386462 15  3.0448255  8.955174     1
 25    7.475 1.386462 15  4.5198255 10.430174     1
 10   14.150 1.386462 15 11.1948255 17.105174     2
 15   18.325 1.386462 15 15.3698255 21.280174     2
```

Results are averaged over the levels of: block

Confidence level used: 0.95

P value adjustment: tukey method for comparing a family of 6 estimates

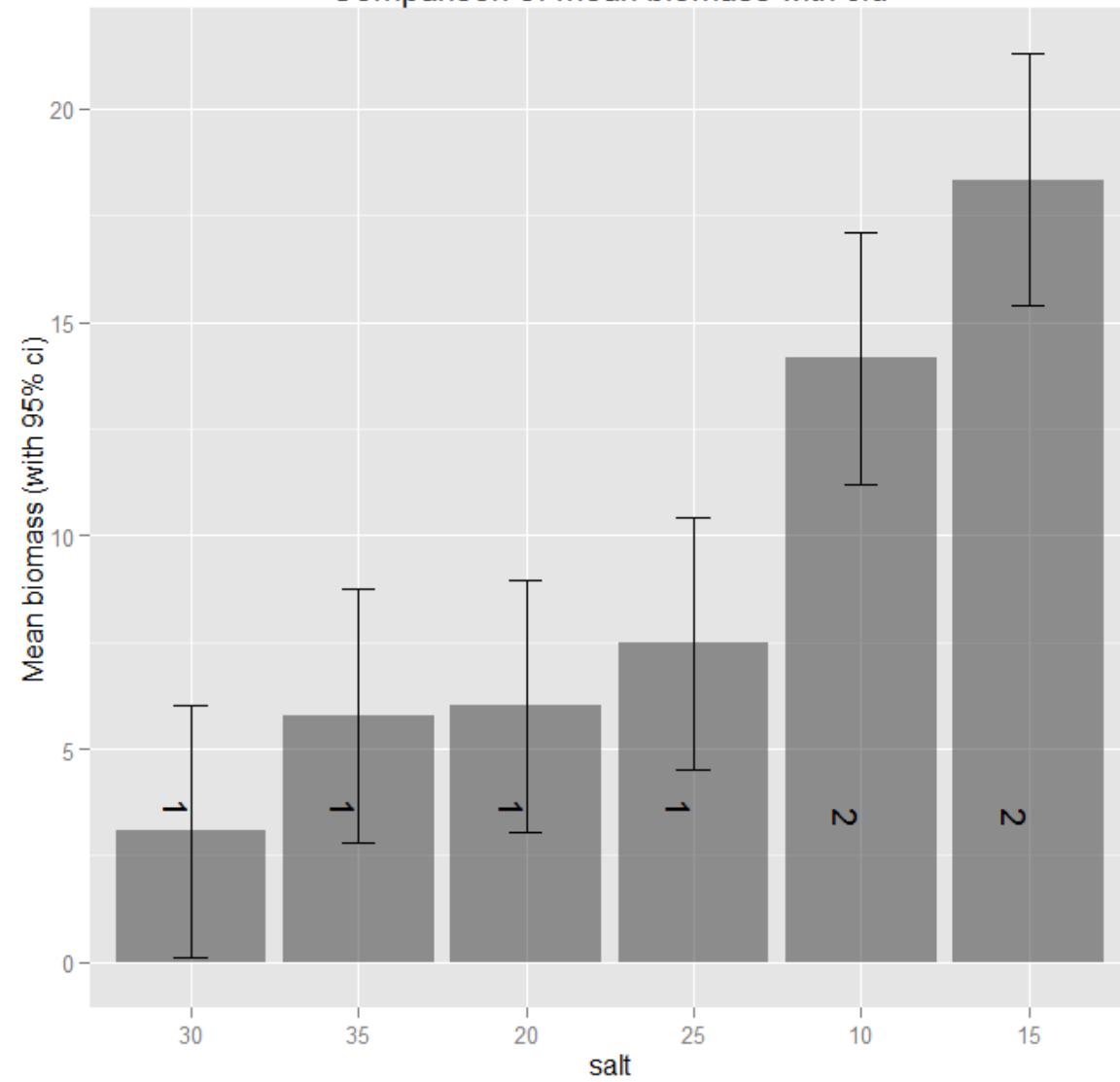
significance level used: alpha = 0.05

- Also interesting to draw plots displaying the results.
- Two common plots can be obtained using the helper functions (`helperfunctions.R`)
- Both display the same information (but in different formats)

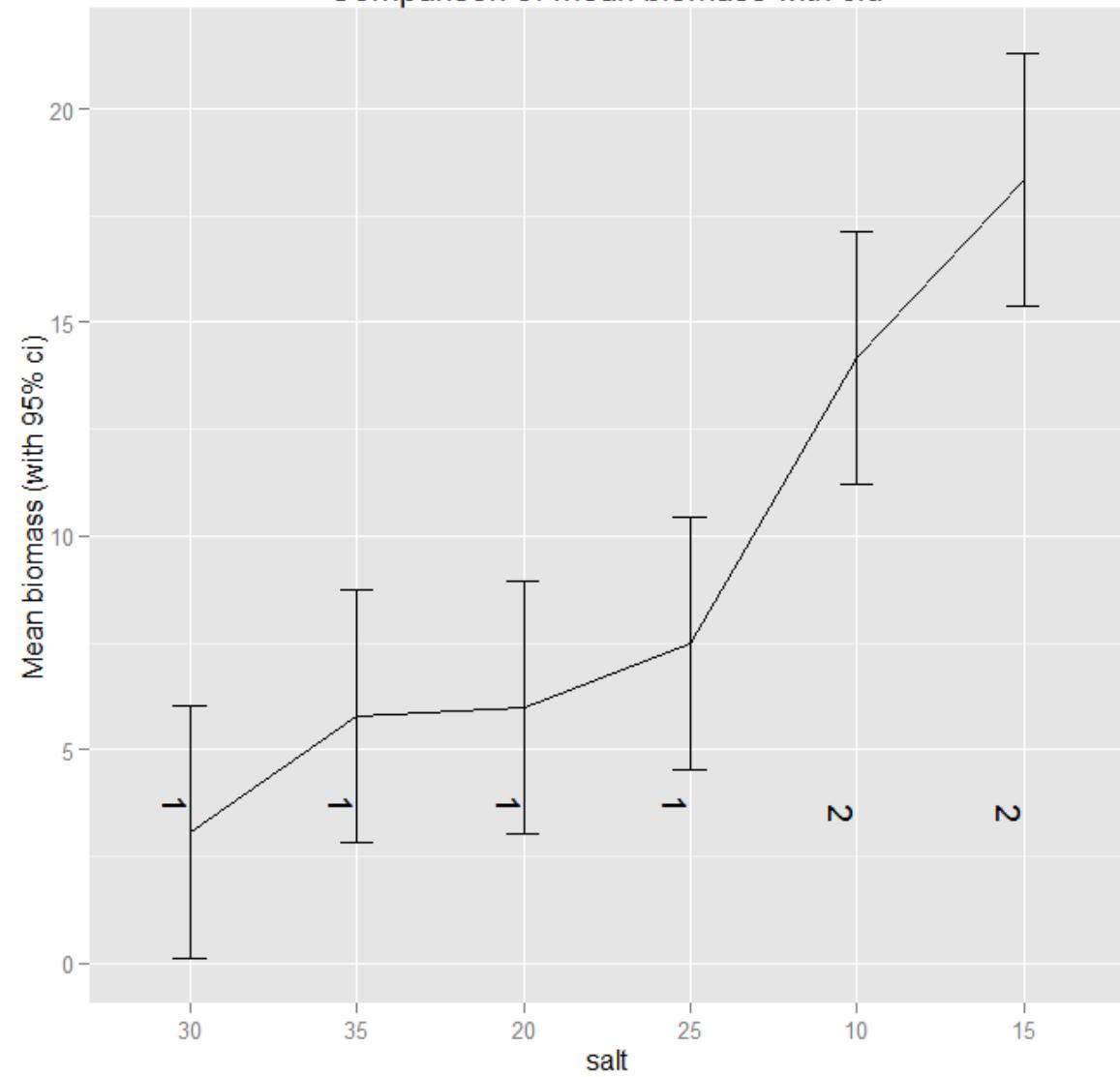
```
> # Make a bar plot of the cld display
> plotcld <- sf.cld.plot.bar(model.cld, variable="salt")
> plotcld <- plotcld +
+   xlab("salt") +
+   ylab("Mean biomass (with 95% ci)") +
+   ggtitle("Comparison of mean biomass with cld")
> plotcld

> # Make a line graph of the cld display
> plotcldb <- sf.cld.plot.line(model.cld, variable="salt")#, ciwidth=0.1)
> plotcldb <- plotcldb +
+   xlab("salt") +
+   ylab("Mean biomass (with 95% ci)") +
+   ggtitle("Comparison of mean biomass with cld")
> plotcldb
```

Comparison of mean biomass with cld



Comparison of mean biomass with cld



```

> # Find all the pairwise differences adjusting for multiplicity
> model.pairs <- pairs(model.lsmo, adjust="tukey")
> summary(model.pairs, infer=TRUE)

contrast estimate      SE df lower.CL upper.CL t.ratio p.value
10 - 15    -4.175 1.960754 15 -10.5454267  2.195427 -2.129  0.3242
10 - 20     8.150 1.960754 15  1.7795733 14.520427  4.157  0.0089
10 - 25     6.675 1.960754 15  0.3045733 13.045427  3.404  0.0374
10 - 30    11.075 1.960754 15  4.7045733 17.445427  5.648  0.0005
10 - 35     8.375 1.960754 15  2.0045733 14.745427  4.271  0.0072
15 - 20    12.325 1.960754 15  5.9545733 18.695427  6.286  0.0002
15 - 25    10.850 1.960754 15  4.4795733 17.220427  5.534  0.0007
15 - 30    15.250 1.960754 15  8.8795733 21.620427  7.778  <.0001
15 - 35    12.550 1.960754 15  6.1795733 18.920427  6.401  0.0001
20 - 25    -1.475 1.960754 15 -7.8454267  4.895427 -0.752  0.9715
20 - 30     2.925 1.960754 15 -3.4454267  9.295427  1.492  0.6742
20 - 35     0.225 1.960754 15 -6.1454267  6.595427  0.115  1.0000
25 - 30     4.400 1.960754 15 -1.9704267 10.770427  2.244  0.2749
25 - 35     1.700 1.960754 15 -4.6704267  8.070427  0.867  0.9488
30 - 35    -2.700 1.960754 15 -9.0704267  3.670427 -1.377  0.7394

```

Results are averaged over the levels of: block

Confidence level used: 0.95

Conf-level adjustment: tukey method for comparing a family of 6 estimates

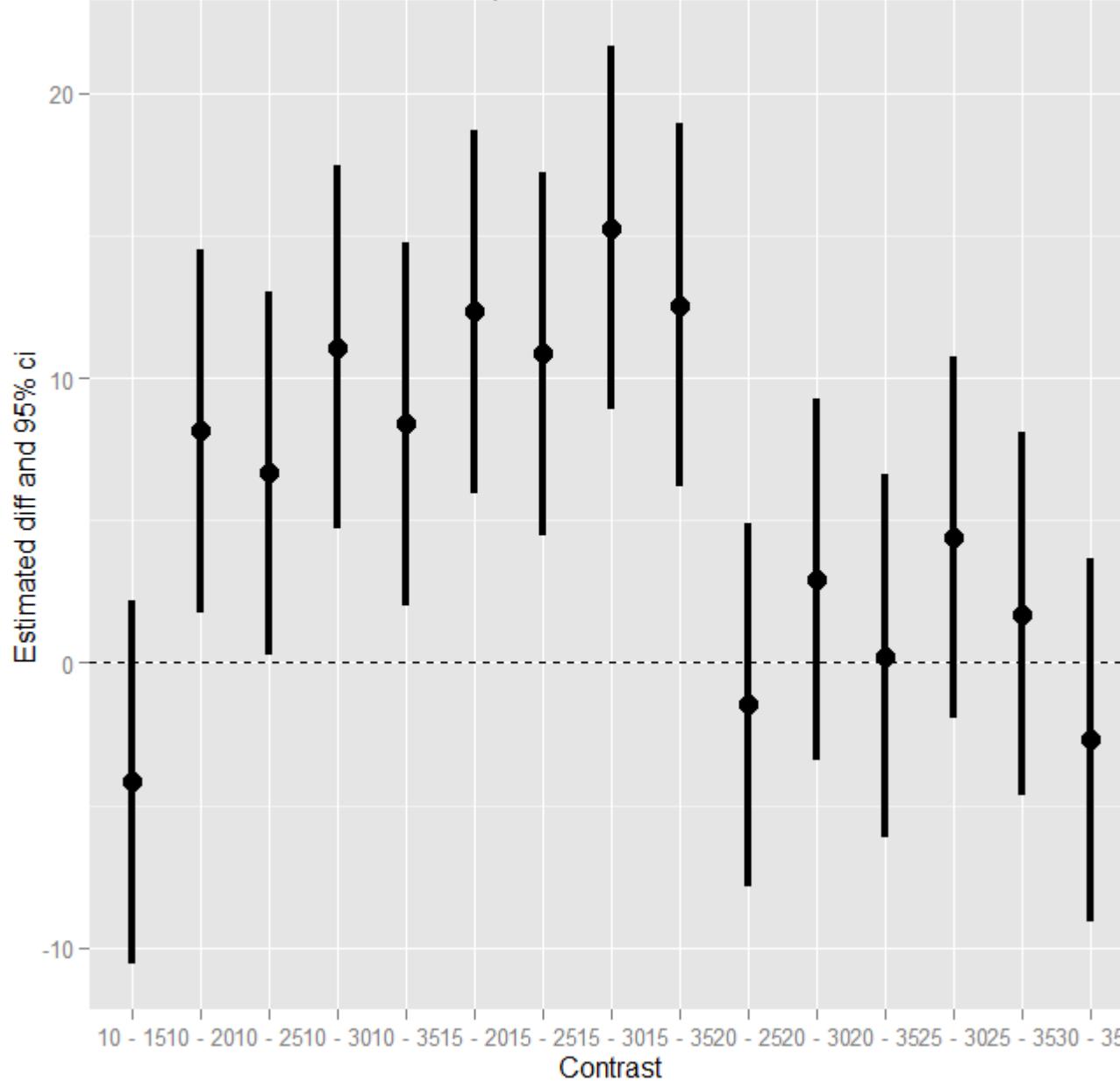
P value adjustment: tukey method for comparing a family of 6 estimates

The set of all pairwise differences in the means is found using the pairs() function applied to the lsmeans object

A plot of all pairwise differences and confidence intervals can be formed:

```
model.pairs.ci <- confint(model.pairs) # extract the ci values
model.pairs.ci
plotdiff <- ggplot(model.pairs.ci, aes(contrast, estimate, ymin = lower.CL, ymax = upper.CL)) +
  geom_point(size=4) +
  geom_linerange(size=1.5) +
  geom_abline(interecept=0, slope=0, linetype=2) +
  ylab("Estimated diff and 95% ci") +
  xlab("Contrast") +
  ggtitle("Estimated pairwise differences and ci")
plotdiff
```

Estimated pairwise differences and ci



Look which confidence intervals contain the value of zero? → no difference in the means

- Alternative functions and code: see salt.R

# Incomplete Block Designs

NOT every block has every treatment

# Example: Investigate differences in water quality

- TSS (Total Suspended Solids in water samples taken
  - upstream of a development of harmful algal bloom (the reference sample),
  - at the development (the mid-stream sample),
  - or downstream of the development (the ds sample).
- Samples are taken during storm events when water quality may be compromised by the development.



Location	Storm 1	Storm 2	Storm 3	Storm 4
Ref	.	.	25	20
Mid	51	.	100	.
DS	173	137	170	110

. represents data that is missing

We assume that the missing data are MCAR (Missing Completely at Random):  
missingness is unrelated to the value of TSS or any other measurable covariate

*this could be violated:*

*“Censored data”: missing values indicate that the TSS is above or below detection limit of measurement device.*

→ more advanced methods necessary

We assume that there is no interaction between storm events and locations:

- Water quality varies among the storm events in some unknown fashion, but it is thought that all locations should be influenced in the same way.
- Example: events with large amounts of precipitation may increase the TSS in all locations.

# How to analyze this?

Location	Storm 1	Storm 2	Storm 3	Storm 4
Ref	-	-	25	20
Mid	51	-	100	-
DS	173	137	170	110

- A comparison of the simple average of the values in each location is an unfair comparison because the different averages would compare different combinations of storm events.
- An incomplete-block analysis takes into account the pattern of missing values:
  - ✓ comparison ref vs. mid locations: use the data from Storm 3
  - ✓ comparison ref vs. ds locations: use the data from Storm 4
  - ✓ comparison mid vs. ds locations: use the data from Storms 1 and 3

*(Such a small set of data likely has very poor power to detect anything but very large differences in water quality.)*

```
> wq <- read.csv("water-quality.csv", header=TRUE,  
+                  na.strings = ".", as.is=TRUE)  
> wq$logTSS <- log(wq$TSS)  
> wq$Location <- as.factor(wq$Location) # convert to factors  
> wq$Event     <- as.factor(wq$Event)  
> wq
```

	Location	Event	TSS	logTSS
1	ds	storm	1	173 5.153292
2	mid	storm	1	51 3.931826
3	ref	storm	1	NA NA
4	ds	storm	2	137 4.919981
5	mid	storm	2	NA NA
6	ref	storm	2	NA NA
7	ds	storm	3	170 5.135798
8	mid	storm	3	100 4.605170
9	ref	storm	3	25 3.218876
10	ds	storm	4	110 4.700480
11	mid	storm	4	NA NA
12	ref	storm	4	20 2.995732

Data: water-quality.csv  
Script: incompleteblock\_waterquality.R

Wide range in data and the ratio among values is of interest

→ a log-transformation of the data is often preferred.

*Note that the log function is the natural logarithm (i.e. to base e) and not the common (i.e. to base 10) logarithm.*

The hypothesis of interest is that:

- $H_0 : \mu_{\text{location ds}} = \mu_{\text{location mid}} = \mu_{\text{location ref}}$
- $H_A : \text{not the above}$

where the  $\mu$  terms refer to the mean logTSS at each of the three locations.

- The programs automatically account for the missing values
- In case of extreme numbers of missing values: design is “non-connected”  
→ analysis fails

```
> model <- lm( logTSS ~ Event + Location, data=wq) Here, blocks are  
> anova(model) treated as fixed effects.  
Analysis of Variance Table  
  
Response: logTSS  
                Df Sum Sq Mean Sq F value Pr(>F)  
Event          3 0.9031 0.30104 3.7911 0.21572  
Location      2 3.9996 1.99979 25.1836 0.03819 * some evidence of a difference in  
Residuals    2 0.1588 0.07941 the mean logTSS among the three  
---  
Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

use the `lsmeans()` functions to fit the (Tukey) multiple comparison to estimate

- ✓ the marginal means
- ✓ the difference in mean log(TSS) among the Locations

use the `lsmeans()` function to fit the (Tukey) multiple comparison to estimate the marginal means and the difference in mean log(TSS) among the Locations:

```
> model.lsmo <- lsmeans(model, ~Location)
> model.cld <- cld(model.lsmo)
> model.cld
  Location    lsmean       SE  df lower.CL upper.CL .group
  ref        3.067368 0.2329431 2  2.065095  4.069642     1
  mid        4.152659 0.2329431 2  3.150385  5.154932    12
  ds         4.977388 0.1408976 2  4.371154  5.583621     2
```

Results are averaged over the levels of: Event

Confidence level used: 0.95

P value adjustment: tukey method for comparing a family of 3 estimates

significance level used: alpha = 0.05

population means (`lsmeans`) are different from the raw mean.

This is because of the adjustment by the procedure for the pattern of missing values.

use the `lsmeans()` function to fit the (Tukey) multiple comparison to estimate the marginal means and the difference in mean log(TSS) among the Locations:

```
> pairs(model.lsmo, adjust="tukey")
   contrast   estimate      SE df t.ratio p.value
   ds - mid  0.8247292 0.272240  2   3.029  0.1659
   ds - ref  1.9100195 0.272240  2   7.016  0.0358      WHAT DOES IT MEAN?
   mid - ref 1.0852903 0.325389  2   3.335  0.1409
```

Results are averaged over the levels of: Event

P value adjustment: tukey method for comparing a family of 3 estimates

$$\begin{array}{ccc} \times e^2 & \stackrel{e^3 \rightarrow 3}{\downarrow} & \Delta = 2 \\ & \downarrow & \\ e^5 \rightarrow 5 & & \end{array}$$

TSS at the ds location is estimated to be  $e^{1.91} = 6.75$  TIMES larger (on average) than at the ref site.

use the `lsmeans()` function to fit the (Tukey) multiple comparison to estimate the marginal means and the difference in mean log(TSS) among the Locations:

```
> confint(pairs(model.lsmo, adjust="tukey"))
   contrast   estimate      SE df lower.CL upper.CL
   ds - mid  0.8247292 0.272240  2 -0.7789692 2.428428
   ds - ref  1.9100195 0.272240  2  0.3063210 3.513718
   mid - ref 1.0852903 0.325389  2 -0.8314960 3.002077
```

Difference in TSS between ds and ref location is, with 95% confidence, between  $e^{0.3} = 1.35$  and  $e^{3.5} = 33.1$  FOLD

Results are averaged over the levels of: Event

Confidence level used: 0.95

Conf-level adjustment: tukey method for comparing a family of 3 estimates

```
> 0.272240 * 6.75
[1] 1.83762
```

SE of the anti-log of the estimate:

“delta-method”:

$$SE_{\text{anti-log}} = SE_{\log} \times e^{1.91} = 0.272240 \times 6.75$$

- The confidence intervals are much wider than the usual 2se because of the small sample size

A more refined analysis would treat the storm events as *random effects*

two sources of information about treatment effects:

- The major part of the information comes from the intra-block analysis (done above).
- Some small amount of additional information can be extracted (known as the inter-block information).

By specifying that blocks are a random effect (i.e. that you wish to extrapolate to other events other than the observed storms), it is possible to combine both analyses

```
> library(lmerTest)
> model.lmer <- lmer( logTSS ~Location + (1|Event), data=wq,
+                      na.action=na.omit) ← State explicitely
> anova(model.lmer, ddf="Kenward-Roger")
Loading required package: pbkrtest
Analysis of Variance Table of type 3 with Kenward-Roger
approximation for degrees of freedom
      Sum Sq Mean Sq NumDF DenDF F.value    Pr(>F)
Location 3.345  1.6725     2 2.9457  28.107 0.01205 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The *effect test* uses more information and so indicates more evidence of an effect of location upon the mean logTSS

# Estimate the population marginal means

Safe to specify here that you want the lsmeans function from the lsmeans package (because lmerTest package has the same function)

```
> model.lmer.lsmo <- lsmeans::lsmeans(model.lmer, ~Location)
> model.lmer.cld <- cld(model.lmer.lsmo)
> model.lmer.cld
   Location    lsmean       SE   df lower.CL upper.CL .group
   ref        3.090375 0.2334176 4.96 2.488895 3.691855     1
   mid        4.229486 0.2334176 4.96 3.628006 4.830966    12
   ds         4.977388 0.1376413 4.80 4.619205 5.335571     2

Confidence level used: 0.95
P value adjustment: tukey method for comparing a family of 3 estimates
significance level used: alpha = 0.05
```

estimated DIFFERENCE of the mean logTSS are slightly changed (no dramatic change) but the standard error is improved

```
> pairs(model.lmer.lsmo, adjust='tukey')
   contrast estimate      SE   df t.ratio p.value
ds - mid    0.7479019 0.2555202 2.88   2.927  0.1243
ds - ref   1.8870128 0.2555202 2.88   7.385  0.0119
mid - ref  1.1391109 0.3194795 3.22   3.566  0.0678

P value adjustment: tukey method for comparing a family of 3 estimates
> confint(pairs(model.lmer.lsmo, adjust='tukey'))
   contrast estimate      SE   df lower.CL upper.CL
ds - mid    0.7479019 0.2555202 2.88 -0.3503971 1.846201
ds - ref   1.8870128 0.2555202 2.88  0.7887137 2.985312
mid - ref  1.1391109 0.3194795 3.22 -0.1382610 2.416483

Confidence level used: 0.95
Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

CRD and an RCB with  
sub-sampling

# Introduction

Sub-samples or pseudo-replicates are NOT independent experimental units

If they are treated as independent experimental units, this typically results in a false positive or Type I error.

Example of a sub-sampling design:

multiple fish in a fish tank. A chemical is added to the tank. The subsequent growth of fish is measured.

Experimental unit: the tank

Observational unit: the fish.

*Not all fish within a tank are independent of each other!*

Basic idea of the analysis: average over sub-samples and analyze the average only.

For unbalanced data (NOT the same number of sub-samples in each experimental unit), the averaging approach is only an approximate analysis.

Experimental design: fewer experimental unit with more sub-samples, or more experimental units and fewer sub-samples?

→ optimal design (best precision for a fixed cost) can be determined based on

the ratio of the variation among experimental units to the variation among sub-samples.

# Example 1

## Fat levels in fish

-

balanced data in a CRD

- Four different species of fish (a, b, c, d)
- Three fish per species:
- From each fish of each species, three samples are taken of the muscle
- and the fat level is measured.

$$H_0: \mu_a = \mu_b = \mu_c = \mu_d$$

$H_1$ : at least one group's mean fat level differs from the rest

Species	Fish	Sample	Fat
a	1	1	11.2
a	1	2	11.6
a	1	3	12
a	2	1	16.5
a	2	2	16.8
a	2	3	16.1
a	3	1	18.3
a	3	2	18.7
a	3	3	19
b	1	1	14.1
b	1	2	13.8
b	1	3	14.2
b	2	1	19
b	2	2	18.5
b	2	3	18.2
b	3	1	11.9
b	3	2	12.4
b	3	3	12
c	1	1	15.3
c	1	2	15.9
c	1	3	16
c	2	1	19.5
c	2	2	20.1
c	2	3	19.3
c	3	1	16.5
c	3	2	17.2
c	3	3	16.9
d	1	1	7.3
d	1	2	7.8
d	1	3	7
d	2	1	8.9
d	2	2	9.4
d	2	3	9.3
d	3	1	11.2
d	3	2	10.9
d	3	3	10.5

Data: fat.csv  
Script: fat.R

NOTE: Fish are named 1, 2, 3 (not 1 to 36)  
Although Fish 1 in species a is obviously not the same as Fish 1 in species b

This is actually bad form, but common.

In general, always try and use unique codes for separate experimental units

# Method 1: Analysis based on sample means

usual “cure” for subsampling: analyze the averages over the subsamples

→ Create a “new” dataset containing the average fat concentration over the three subsamples (repetitions) of each fish:

    Use summaryBy() function in the doBy package

```
> library(doBy)
> summaryBy(Fat ~ Fish, FUN=mean, data=fishfat)# this doesn't make sense!
  Fish Fat.mean
1     1 12.18333
2     2 15.96667
3     3 14.62500
> #first divide by species and then by fish
> avg <- summaryBy(Fat ~ Species+ Fish, FUN=mean, data=fishfat)
> avg
  Species Fish Fat.mean
1         a    1 11.600000
2         a    2 16.466667
3         a    3 18.666667
4         b    1 14.033333
5         b    2 18.566667
6         b    3 12.100000
7         c    1 15.733333
8         c    2 19.633333
9         c    3 16.866667
10        d    1  7.366667
11        d    2  9.200000
12        d    3 10.866667
```

→ We will do now an ANOVA on these averages

# single-factor CRD ANOVA on the sample means

```
> model <- lm(Fat.mean ~ Species, data=avg)
> anova(model)
Analysis of Variance Table

Response: Fat.mean
            Df  Sum Sq Mean Sq F value    Pr(>F)
Species       3 114.734  38.245   4.9056 0.03205 *
Residuals     8  62.369   7.796
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-statistic for the hypothesis of no difference between the mean levels of fat among species is 4.91 with a p-value of 0.0321.

Fish = experimental unit

Here, because of averaging: each experimental unit corresponds to 1 observation (an average value)

Simple CRD: NO explicit term for the experimental unit in the model

# population marginal means (the LSmeans)

```
> lsmeans <- popMeans(model, eff="Species")
> lsmeans
  estimate      se  df   t.stat     p.value Species
1 15.577778 1.61205   8  9.663337 1.095464e-05    a
2 14.900000 1.61205   8  9.242892 1.522992e-05    b
3 17.411111 1.61205   8 10.800606 4.762894e-06    c
4  9.144444 1.61205   8  5.672558 4.690099e-04    d
```

The estimated mean fat level for species a is 15.58 with an estimated standard error of 1.61

# multiple-comparisons

The estimated difference in the mean fat levels between species a and species b is .678 with an estimated standard error of 2.28

Linear hypotheses to be tested

```
> library(multcomp)
> model.tukey <- glht(model, linfct = mcp(Species = "Tukey"))
> summary(model.tukey)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = Fat.mean ~ Species, data = avg)

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t )
b - a == 0	-0.6778	2.2798	-0.297	0.9902
c - a == 0	1.8333	2.2798	0.804	0.8508
d - a == 0	-6.4333	2.2798	-2.822	0.0857 .
c - b == 0	2.5111	2.2798	1.101	0.6986
d - b == 0	-5.7556	2.2798	-2.525	0.1299
d - c == 0	-8.2667	2.2798	-3.626	0.0276 *
---				

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- single-step method)

Magic function from multcomp package: glht

# multiple-comparisons (CI)

```
> confint(model.tukey)

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = Fat.mean ~ Species, data = avg)

Quantile = 3.2039
95% family-wise confidence level

Linear Hypotheses:
                         Estimate lwr      upr
b - a == 0   -0.6778  -7.9821  6.6265
c - a == 0    1.8333  -5.4710  9.1376
d - a == 0   -6.4333 -13.7376  0.8710
c - b == 0    2.5111  -4.7932  9.8154
d - b == 0   -5.7556 -13.0599  1.5487
d - c == 0   -8.2667 -15.5710 -0.9624
```

# multiple- comparisons (plot)

```
> model.tukey.cld <- cld(model.tukey)
> model.tukey.cld
      a      b      c      d
"ab" "ab"  "b"   "a"
> old.par <- par()
> layout(matrix(1:2,2,1))
> plot(model.tukey)
> plot(model.tukey.cld,
+       xlab="Species",
+       ylab="fat")
> par <- old.par
```

Compact letter display

Directly plot Result of glht function

Plot result of cld function  
(applied to result of glht  
function)

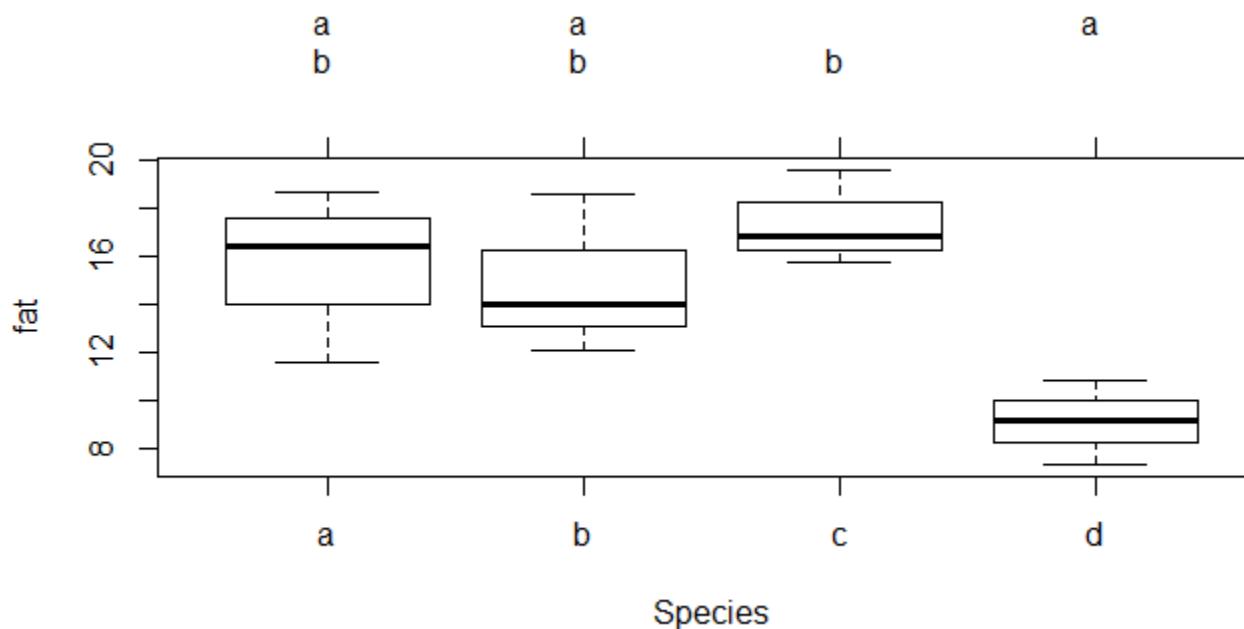
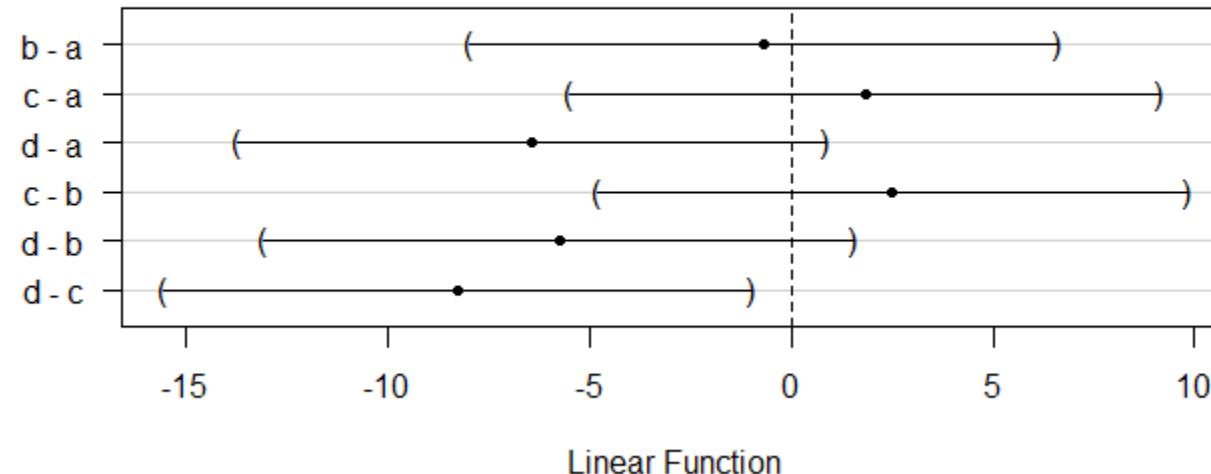
```
plot(model.tukey)
```

(Directly plot Result of  
glht function)

```
plot(model.tukey.cld)
```

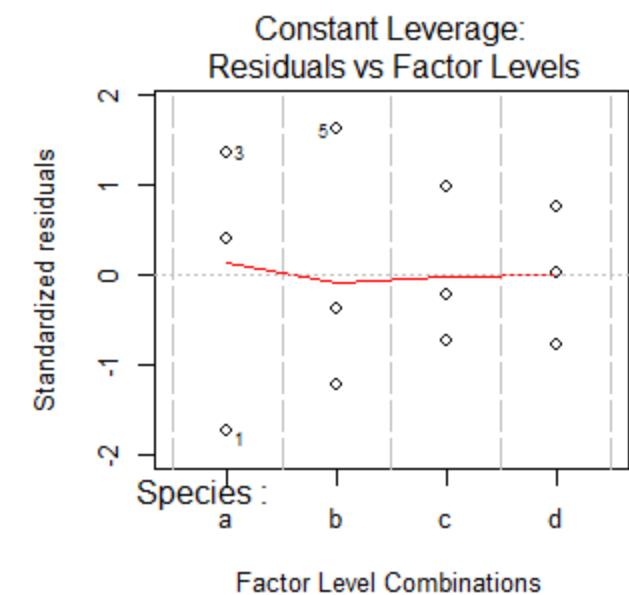
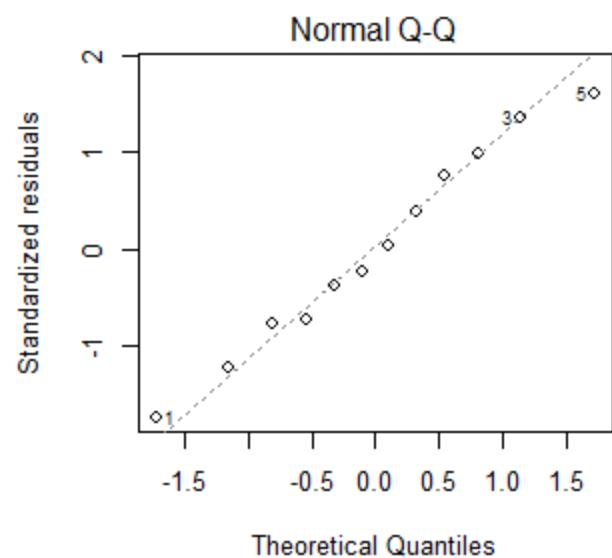
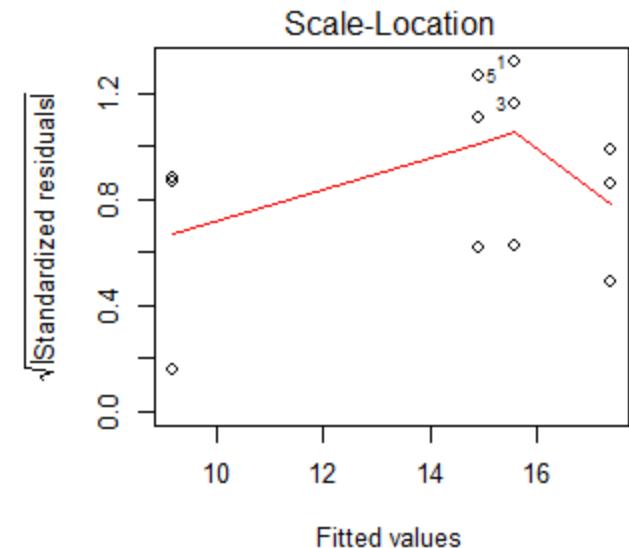
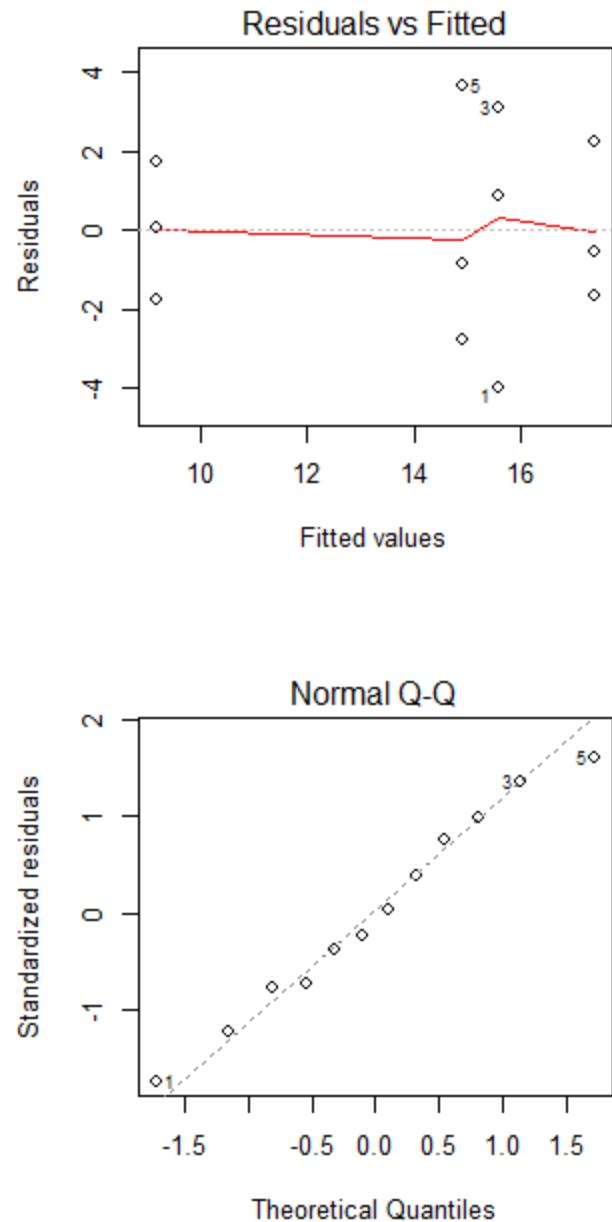
Plot result of cld function  
(applied to result of glht  
function)

95% family-wise confidence level



diagnostic plots  
doesn't indicate  
any problems  
with the fit

```
> plot(model)
```



# Method 2: Analysis using individual values

model must include the following terms:

- Species                      → treatment
- Fish(Species)              → the **experimental unit (Fish)**
  - a **nested** effect ....(Species))
  - a **random** effect

nested effect: the levels change as the levels of other factors change.

Here: fish 1 of Species a is a different fish than fish 1 of Species b, even though they are both numbered “1”.

Experimental units are always Random effects:

- if the experiment were to be repeated, different fish would be used
- we don't wish to limit inferences to these particular fish sampled in this particular experiment.

OR, if the fish are individually labelled (always a good idea):

model must include the following terms:

- Species → treatment
- Fish → the **experimental unit (Fish)**  
a **random effect**

**no explicit term for the observational unit (the sub-samples)**

nested effect: leave it up to the computer package:

In this case, the computer package should be able to figure out that fish are the experimental units and samples are the sub-sampling units.

To fit the model in R:

YOU MUST HAVE UNIQUE LABELS for each fish  
and just add it as a random effect

- Use the lme() function in the nlme package
- OR use lmer() function in lmerTest package

Both functions cannot deal with the nesting term

→ create unique labels for each fish using the interaction() function.

- Fish and Species variables must be declared as a factor rather than a continuous variable

```

> # Create a unique fish id for each combination of species and fish
> # to avoid having to specify the nesting of fish(species)
> fishfat$Fish.id <- interaction(fishfat$Species, fishfat$Fish)
>
> library(nlme)
> model2 <- lme( Fat ~ Species, random=~1 | Fish.id, data=fishfat)
> anova(model2)

```

	numDF	denDF	F-value	p-value
(Intercept)	1	24	312.92542	<.0001
Species	3	8	4.90559	0.0321

```

>
> library(lmerTest)
> model2b<-lmerTest::lmer(Fat ~ Species +(1|Fish.id), data=fishfat)
> anova(model2b)

```

Analysis of Variance Table of type 3 with Satterthwaite approximation for degrees of freedom

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Species	1.8151	0.60502	3	8	4.9056	0.03205 *
<hr/>						
Signif. codes:	0	'****'	0.001	'**'	0.01	'*'
	0.05	'.'	0.1	' '	1	

NO test for the effect of fish(species)  
 This is an experimental unit!  
 experimental units ALWAYS differ in their response to the same treatment.

Some of the output from R is missing or WRONG (!) because of “errors” in the R code where certain functions don’t work with lme() models.

```
> # Get the marginal means  
> popMeans(model2, effect="Species")  
Error in terms.default(model.frame(object)) :  
  no terms component nor attribute
```

Note that the marginal means cannot be computed in R, and I don't know how to fix this

```
> library(multcomp)
> model2.tukey <- glht(model2, linfct = mcp(Species = "Tukey"))
> summary(model2.tukey)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lme.formula(fixed = Fat ~ Species, data = fishfat, random = ~1 | Fish.id)

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z )
b - a == 0	-0.6778	2.2798	-0.297	0.99087
c - a == 0	1.8333	2.2798	0.804	0.85251
d - a == 0	-6.4333	2.2798	-2.822	0.02483 *
c - b == 0	2.5111	2.2798	1.101	0.68862
d - b == 0	-5.7556	2.2798	-2.525	0.05598 .
d - c == 0	-8.2667	2.2798	-3.626	0.00162 **
---				

Signif. codes: 0 '\*\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- single-step method)

```
> confint(model2.tukey)
```

### Simultaneous Confidence Intervals

#### Multiple Comparisons of Means: Tukey Contrasts

Fit: lme.formula(fixed = Fat ~ Species, data = fishfat, random = ~1 | Fish.id)

Quantile = 2.569

95% family-wise confidence level

#### Linear Hypotheses:

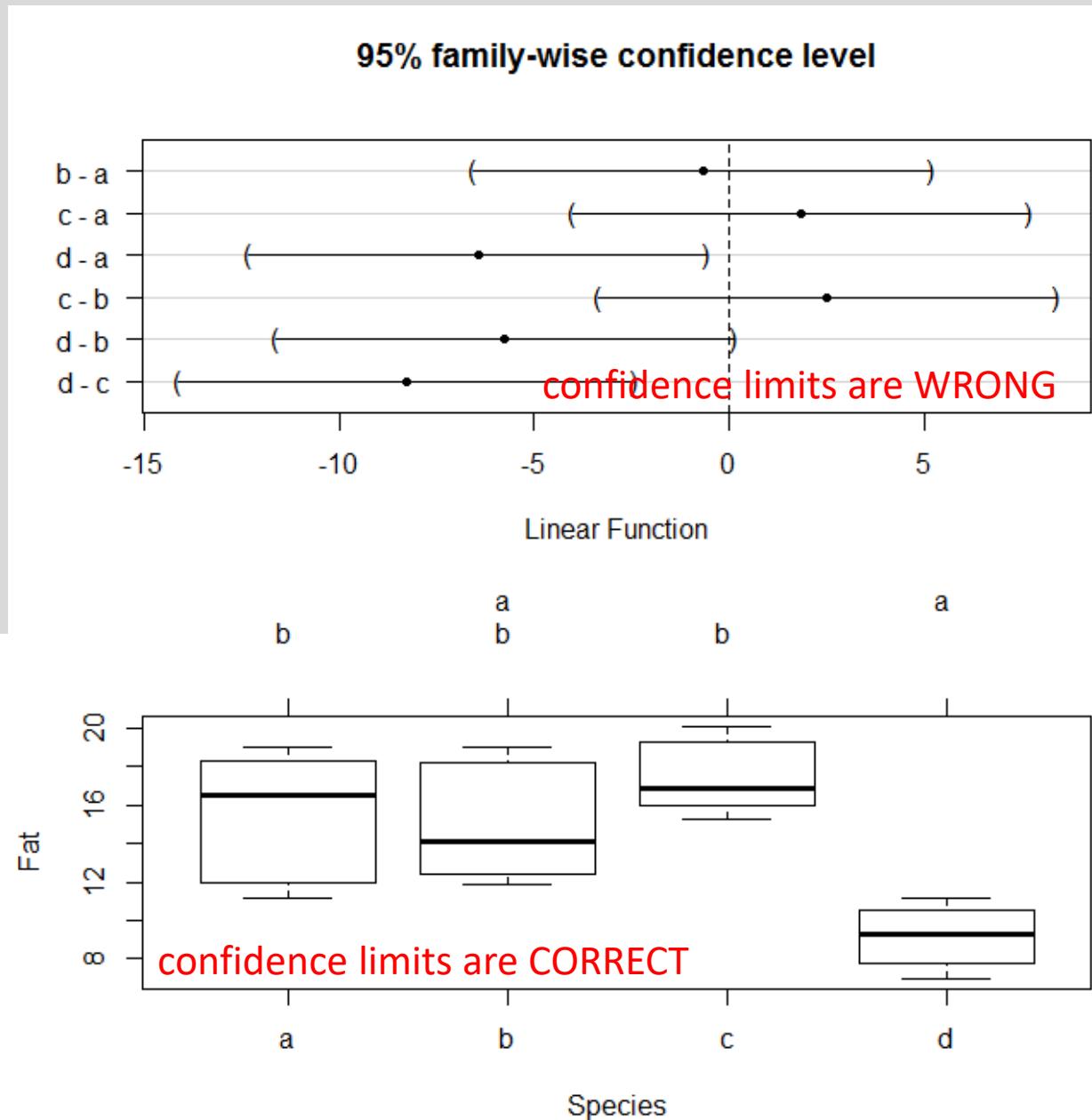
	Estimate	lwr	upr
b - a == 0	-0.6778	-6.5345	5.1790
c - a == 0	1.8333	-4.0234	7.6901
d - a == 0	-6.4333	-12.2901	-0.5766
c - b == 0	2.5111	-3.3456	8.3678
d - b == 0	-5.7556	-11.6123	0.1012
d - c == 0	-8.2667	-14.1234	-2.4099

Note that the confidence limits of  
the DIFFERENCES are WRONG, and I  
don't know how to fix them

```

> model2.tukey.cld <- cld(model2.tukey)
> model2.tukey.cld
  a   b   c   d
"b" "ab" "b"  "a"
> old.par <- par()
> layout(matrix(1:2, 2, 1))
> plot(model2.tukey)
>
> plot(model2.tukey.cld,
+       xlab="Species",
+       ylab="Fat")
> par <- old.par

```



# Comparison with “means of means approach”

## No difference:

- The F-statistic and p-value for the test of no difference in the mean fat levels among species
- The population marginal mean (the LSmeans)
- estimated differences among the species means

## BUT there is some useful information in the fish(species) term:

can be used to see the relative sizes of the variance components

✓ among fish

✓ among samples within fish

# Variance components

```
> vc <- VarCorr(model2) #calculates variance due to random effects in mixed models
> vc
Fish.id = pdLogChol(1)
      Variance StdDev
(Intercept) 7.7549983 2.7847798
Residual     0.1233333 0.3511885
```

Variance among fish  
Variance among samples

- The total variation is 7.878.
- Over 98% of the variation in the data is among fish, rather than among samples within a fish. (makes sense!)

*Remember: random factors are seen as a structuring of the variance, NOT as a treatment effect*

# the precision of the experiment

- is a function of

$$\frac{\sigma^2_{fish}}{n_{fish}} + \frac{\sigma^2_{sample}}{n_{fish} \times n_{sample/fish}}$$

- The smaller the value of this function, the “better” the experiment.

## → Examine the tradeoff between

- sampling more fish (increase the value of  $n_{fish}$  and decrease the value of  $n_{sample/fish}$ )
  - taking more samples per fish (increase the value of  $n_{sample/fish}$  and decrease the value of  $n_{fish}$ )
- while keeping the total number of samples ( $n_{fish} \times n_{sample/fish}$ ) constant.

# Example: precision

n_fish	n_sample/fish	n_total	Precision
12	1	12	0.657
6	2	12	1.303
4	3	12	1.949
3	4	12	2.595
2	6	12	3.888
1	12	12	7.765

Here:  $\sigma^2_{fish} \gg \sigma^2_{samples}$

→ The best option, is to sample as many fish as possible and only take one sample from each fish.

BUT this ignores the costs of obtaining a new fish (expensive) vs. the cost of taking an additional sample (cheap)!

To use cost information: can use SOLVER tool in Excel spreadsheet:

- costs of the various options are specified
- a total cost function is determined
- the variables  $n_{fish}$  and  $n_{sample/fish}$  are manipulated using the SOLVER tool to minimize the precision function while keeping the cost fixed at your budget level.

« optimal » number of sub-samples is

$$n_{samples} = \sqrt{\frac{c_{fish}}{c_{sample}} \times \frac{\sigma^2_{sample}}{\sigma^2_{fish}}}$$

→ ratios of costs and of variances are important

more sub-samples are taken if

- the cost of new fish is large relative to the cost of sampling
- the variance of the sub-samples is large relative to variance among fish.

→ Ignoring costs, then *unless the variation of the sub-samples is at least as large as the variation among experimental units, it never pays to sub-sample.*

# diagnostic plots doesn't indicate any problems with the fit

```
# lme() uses Trellis graphics, so the usual
plotting commands are not useful

plot1 <- plot(model2, resid(., type = "p") ~
fitted(.) | Species, abline = 0)

# box-plots of residuals by Subject

plot2 <- plot(model2, Species ~ resid(.))

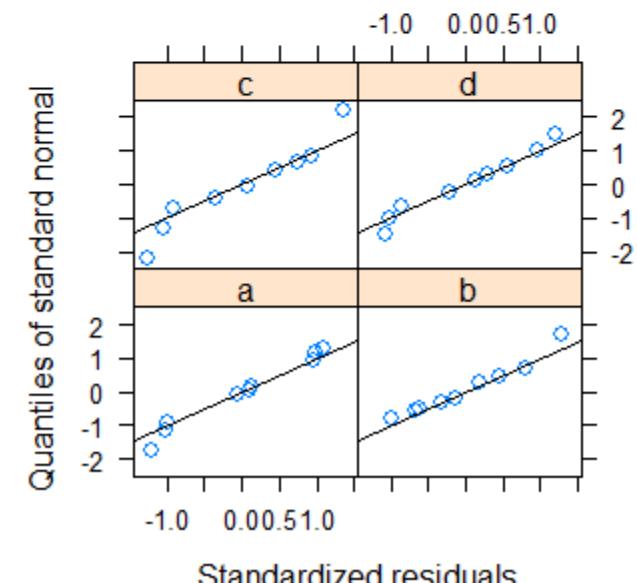
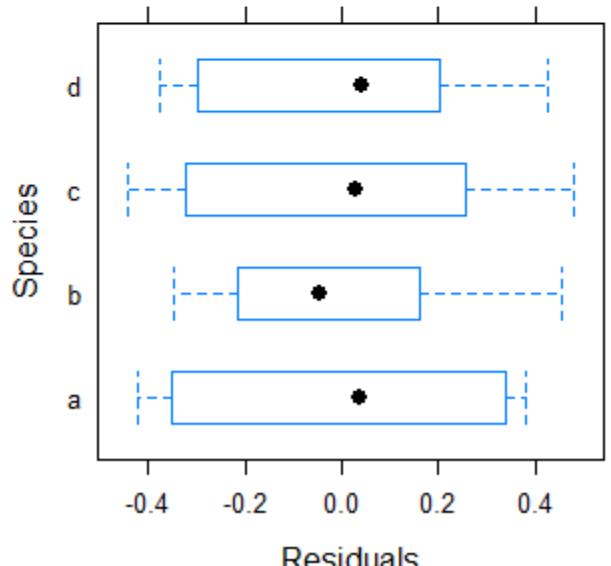
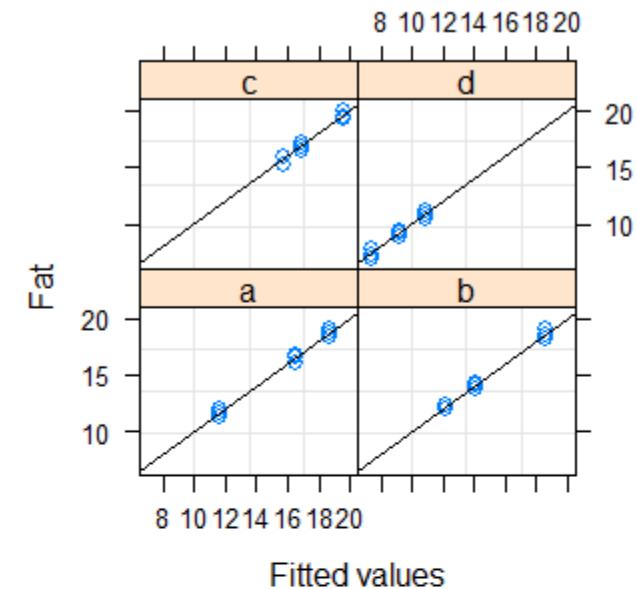
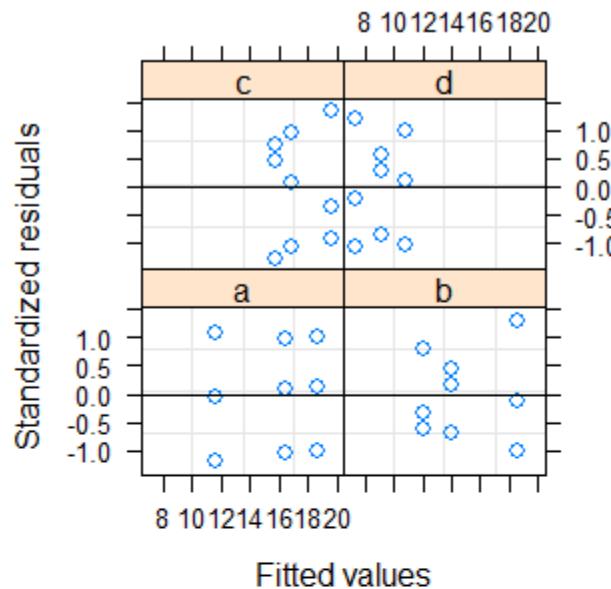
# observed versus fitted values by Subject

plot3 <- plot(model2, Fat ~ fitted(.) | Species,
abline = c(0,1))

# normak probability plot of the residuals.

plot4 <- qqnorm(model2, ~ resid(., type = "p") |
Species, abline = c(0, 1))

print(plot1, split=c(1,1,2,2), more=TRUE)
print(plot2, split=c(1,2,2,2), more=TRUE)
print(plot3, split=c(2,1,2,2), more=TRUE)
print(plot4, split=c(2,2,2,2))
```



# Example 2

## Fat levels in fish

-

unbalanced data in a CRD

number of sub-samples available differs for the experiment units

- The method of analyzing the averages will not be exact, but will only give an approximate answer.
  - The method of fitting the model including the nested terms proceeds exactly in the same fashion.
- NO exact test can be constructed
- BUT an approximate F-statistic (called a pseudo-F-statistic) can be constructed.
- not all the LSmeans have the same precision
  - Not all estimates of the differences between the species have the same precision.

Example: repeat the above analyses after randomly deleting some of the data.

# Example 2

## Effect of UV radiation

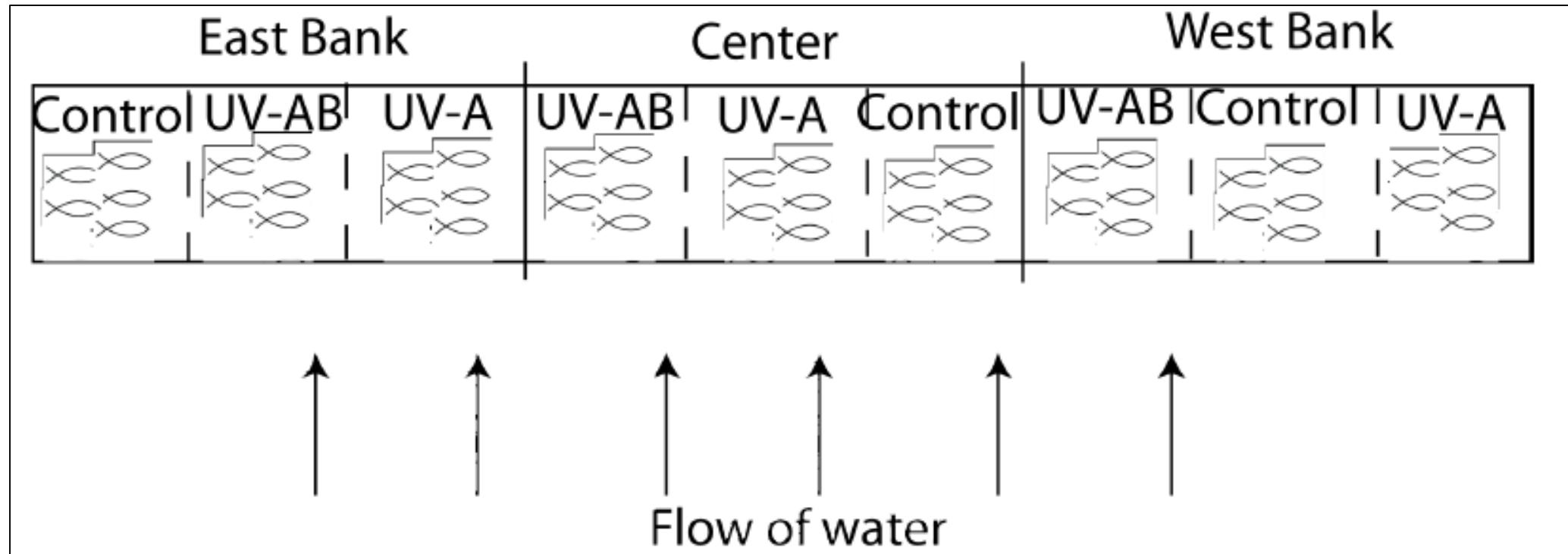
-

### balanced data in RCB

# Investigation of the effects of different amounts and types of UV radiation upon the growth of fish

- A single factor with three levels :
  - Control : normal amounts of sunlight
  - UVA : additional UVA radiation over the control
  - UVAB : additional UVA and UVB radiation over the control

- A total of 9 flumes were placed across a stream into blocks of 3 flumes.
  - Within each block of 3 flumes, the UV levels were randomly assigned to the three flumes.
  - Five individually marked fish were placed within each flume.



After about 150 days, the weight gain of each fish was measured.

Data: uvexp.csv

Script: uvexp.R

Block	Trt	Flume	Fish	WeightGain
1	Control	1	L<	1.1
1	Control	1	LF	1.3
1	Control	1	LR	1.9
1	Control	1	RR	0.9
1	Control	1	RV	2.1
1	UVAB	2	L<	1.8
1	UVAB	2	LF	1.7
1	UVAB	2	LR	1.8
1	UVAB	2	RF	2
1	UVAB	2	RR	1.8
1	UVA	3	LF	1.2
1	UVA	3	LR	2.7
1	UVA	3	RF	3.2
1	UVA	3	RR	1.7
1	UVA	3	RV	2.2
2	Control	4	LF	3.5
2	Control	4	LR	3
2	Control	4	RF	3.8
2	Control	4	RR	1.7
2	UVAB	5	LF	0.6
2	UVAB	5	LR	0.4
2	UVAB	5	RF	0.5
2	UVAB	5	RR	1.2
2	UVAB	5	RV	0.4
2	UVA	6	LF	1.4
2	UVA	6	LR	2.1
2	UVA	6	RF	0.9
2	UVA	6	RR	2.4
2	UVA	6	RV	2.1
3	Control	7	LF	4
3	Control	7	LR	4.6
3	Control	7	RF	2.6
3	Control	7	RR	2.7
3	Control	7	RV	4.7
3	UVAB	8	L<	1.2
3	UVAB	8	LF	1.8
3	UVAB	8	LR	1.1
3	UVAB	8	RF	1
3	UVAB	8	RR	1.5
3	UVA	9	L<	2.9
3	UVA	9	LF	3
3	UVA	9	LR	3.1
3	UVA	9	RF	3.2
3	UVA	9	RR	3.6

- The data are imported into R using the `read.csv()` function:

```
> uvexp <- read.csv("uvexp.csv", header=TRUE)
> uvexp[1:10, ]
```

	Block	Trt	Flume	Fish	WeightGain
1	1	Control	1	L<	1.1
2	1	Control	1	LF	1.3
3	1	Control	1	LR	1.9
4	1	Control	1	RR	0.9
5	1	Control	1	RV	2.1
6	1	UVAB	2	L<	1.8
7	1	UVAB	2	LF	1.7
8	1	UVAB	2	LR	1.8
9	1	UVAB	2	RF	2.0
10	1	UVAB	2	RR	1.8

### The treatment structure:

single factor design with three levels (Control, UVA, UVAB)

### The experimental unit structure:

- The **experimental units** are the **flumes**, as the treatment is applied to the entire flume and not to individual fish.
- These are arranged into blocks of three flumes.
- Each block is complete, i.e., each block contains all treatments.

(Note that the flumes were numbered so that the flume in each block assigned to the Control treatment had the lowest number, the flume assigned to the UVAB treatment had the next lowest number etc.— this makes the data appear to be ‘non-random’, but the experiment was randomized to flumes within each block.)

- The **observational units** are the **fish** within each flume.

**→This experiment is an example of sub-sampling as the observational unit is different from the experimental unit.**

The randomization occurred at two levels:

1. the treatments were completely randomized to flumes within each block.
2. the fish were completely randomized to each flume.

The hypothesis of interest is that:

- $H_0 : \mu_{control} = \mu_{UVB} = \mu_{UVAB}$
- $H_A : \text{at least one mean is different}$

Because the design is balanced, the analysis can be performed

- either on the averages of the five fish within each flume
- or on the individual fish

# 1. Analysis on sample means

- The `summaryBy()` function in the `doBy` package can be used to get the average weight gain for each flume:

```
> library(doBy)
Loading required package: survival
> avg <- summaryBy(WeightGain ~ Block + Trt+ Flume , FUN=mean,
data=uvexp)
> avg
```

	Block	Trt	Flume	WeightGain.mean
1	1	Control	1	1.46
2	1	UVA	3	2.20
3	1	UVAB	2	1.82
4	2	Control	4	2.96
5	2	UVA	6	1.78
6	2	UVAB	5	0.62
7	3	Control	7	3.72
8	3	UVA	9	3.16
9	3	UVAB	8	1.32

Now, the experiment looks like a randomized complete block design:  
1 « observation » per treatment level per block

- The `lm()` function can be used to fit a single-factor RCB ANOVA on the sample means (same way was seen in previous examples)
- Be sure that the blocking variable is declared as a factor!

```
> avg$Block <- as.factor(avg$Block)
> model <- lm(WeightGain.mean ~ Block + Trt, data=avg)
```

```
> anova(model)
```

Analysis of Variance Table

Response: WeightGain.mean

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	2	1.7198	0.85991	1.2967	0.3681
Trt	2	3.5121	1.75604	2.6479	0.1852
Residuals	4	2.6527	0.66318		

# Estimate the population marginal means (the LSmeans)

- Use the popMeans() function in the doBy package

```
> lsmeans <- popMeans(model, eff="Trt")
> lsmeans
   estimate      se  df t.stat    p.value      Trt
1 2.713333 0.4701694  4 5.770970 0.004475721 Control
2 2.380000 0.4701694  4 5.062005 0.007170453      UVA
3 1.253333 0.4701694  4 2.665706 0.056056064     UVAB
```

# Multiple comparisons (using the multcomp package - the usual way)

```
> library(multcomp)
Loading required package: mvtnorm
Loading required package: TH.data
>
> model.tukey <- glht(model, linfct = mcp(Trt = "Tukey"))
> summary(model.tukey)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = WeightGain.mean ~ Block + Trt, data = avg)

Linear Hypotheses:

			Estimate	Std. Error	t value	Pr(> t )	
UVA	-	Control	== 0	-0.3333	0.6649	-0.501	0.875
UVAB	-	Control	== 0	-1.4600	0.6649	-2.196	0.186
UVAB	-	UVA	== 0	-1.1267	0.6649	-1.694	0.313

(Adjusted p values reported -- single-step method)

```
> confint(model.tukey)
```

### Simultaneous Confidence Intervals

#### Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = WeightGain.mean ~ Block + Trt, data = avg)

Quantile = 3.5645

95% family-wise confidence level

#### Linear Hypotheses:

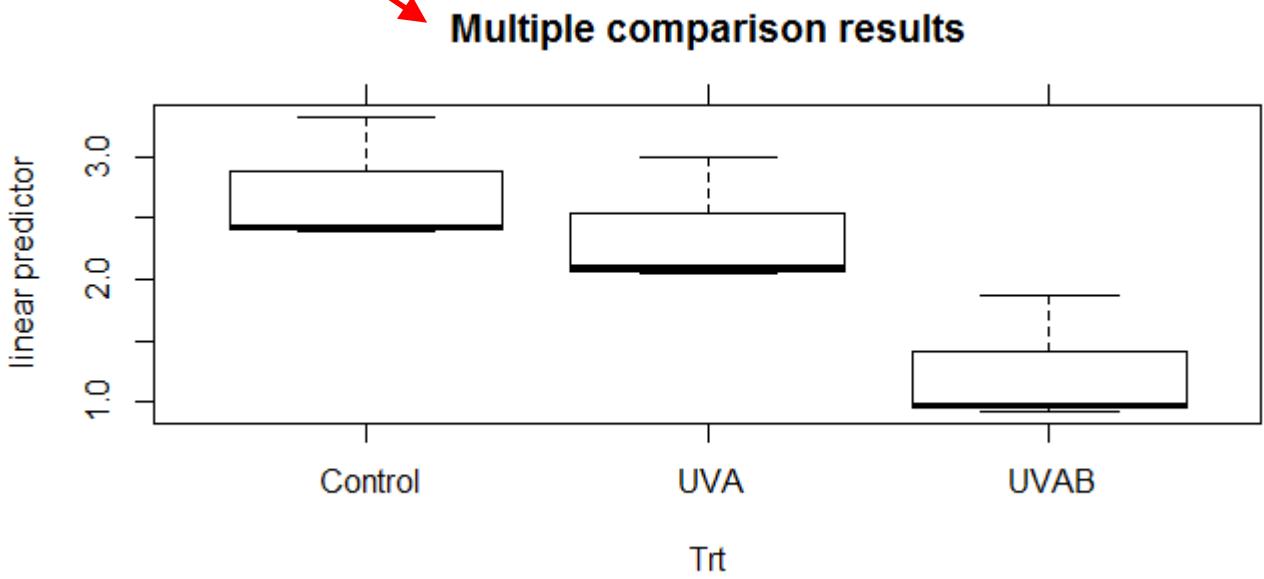
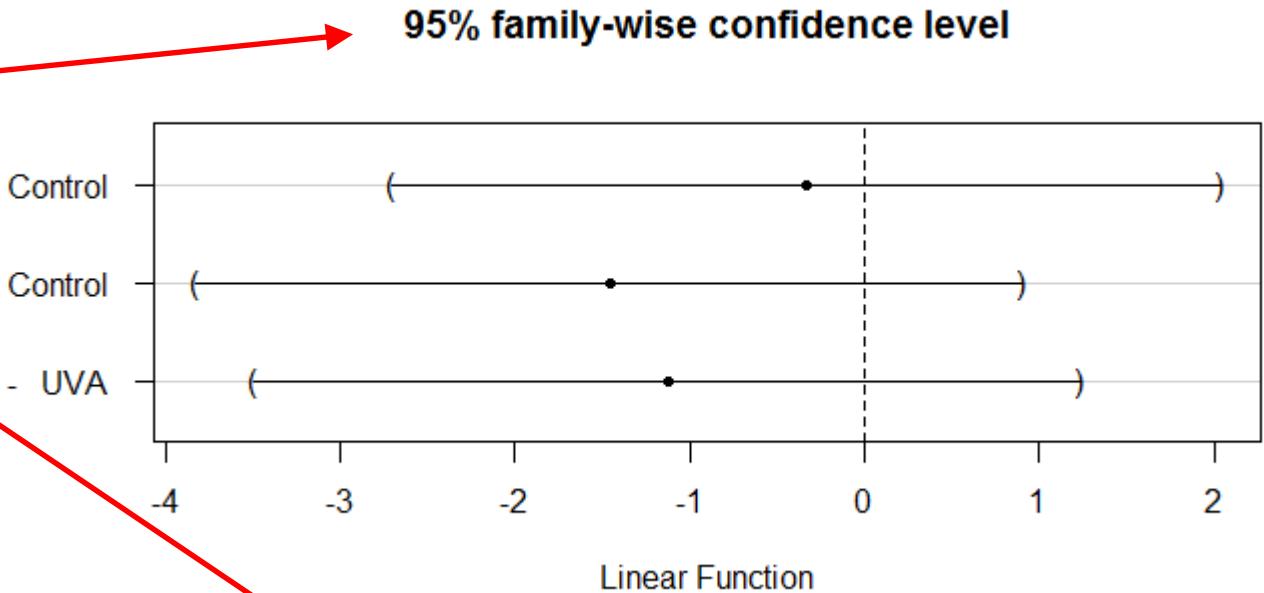
		Estimate	lwr	upr
UVA	- Control == 0	-0.3333	-2.7034	2.0368
UVAB	- Control == 0	-1.4600	-3.8301	0.9101
UVAB	- UVA == 0	-1.1267	-3.4968	1.2434

```
> model.tukey.cld <- cld(model.tukey)
```

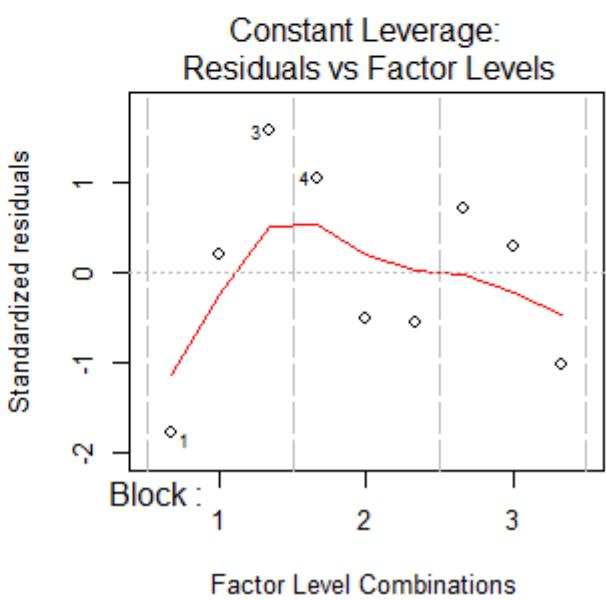
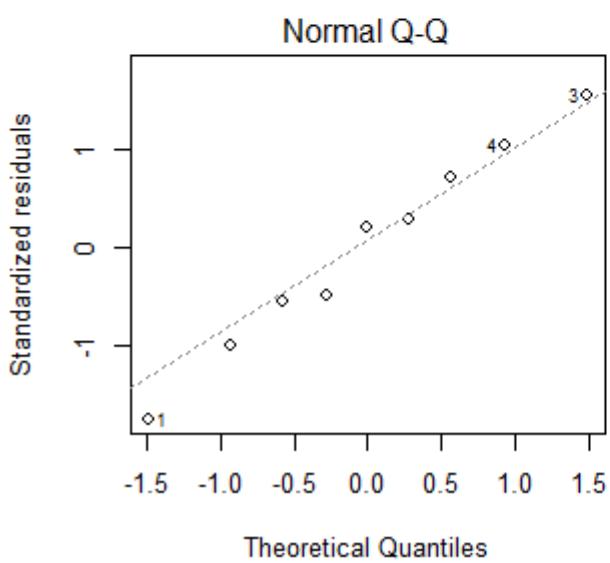
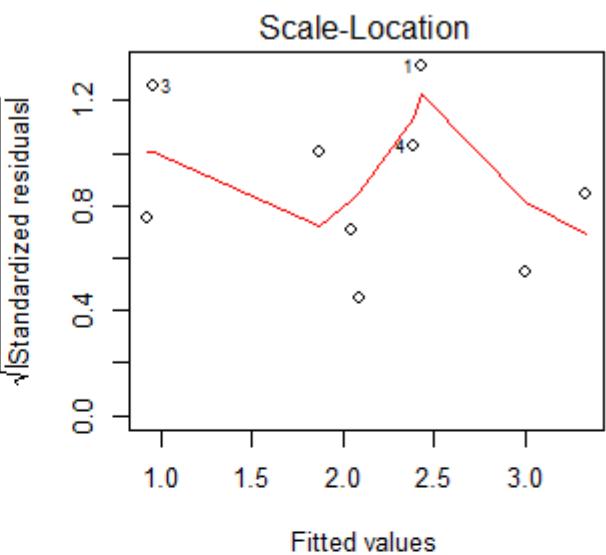
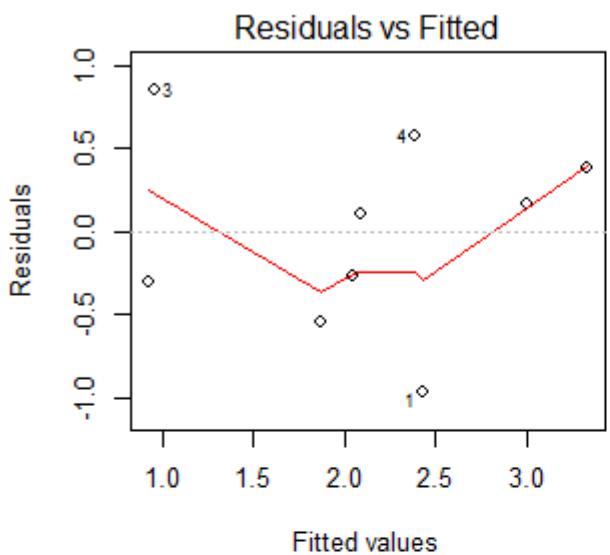
```
> model.tukey.cld
```

Control	UVA	UVAB
"a"	"a"	"a"

```
> old.par <- par()  
> layout(matrix(1:2,2,1))  
> plot(model.tukey)  
> plot(model.tukey.cld,  
main="Multiple comparison  
results",  
+       xlab="Trt",  
+       ylab="fat",  
+       notch=FALSE)  
> par <- old.par
```



```
> # Check the assumptions of the  
linear model on the averages  
> layout(matrix(1:4, 2, 2))  
> plot(model)
```



## 2. Analysis using individual values

Logically, based on CRD example, the model must include the following terms:

- Block → blocking factor
- Trt → treatment
- Flume(Trt) → the **experimental unit (Flume)**
  - a **nested** effect ....(Trt))
  - a **random** effect

**BUT THIS IS WRONG!**

# The third term is supposed to represent the experimental units of the flumes

- Unfortunately, simple syntax with representation of the experimental units as a nested term FLUMES(TRT) will fail for many packages. (this is for historical reasons)
- The standard historical convention is to represent experimental unit effects in a blocked design as an “interaction” between BLOCKS and TREATMENT.

Block\*Trt

*(because every treatment occurs once in every block, the combination of block and treatment levels is sufficient to identify the flume used)*

- EXTREMELY MISLEADING! Because a key assumption of a blocked design is the LACK of interaction between blocks and treatments!

# So, the correct model includes

- Block → blocking factor
- Trt → treatment
- Block\*Trt → the **experimental unit (Flume)**
  - a **random effect**

## Experimental units are always Random effects:

- if the experiment were to be repeated, different fish would be used
- we don't wish to limit inferences to these particular fish sampled in this particular experiment.

## Is Blocking factor fixed or random?

Not so clear!

Here the blocks are the sets of flumes across a river.

We could argue that if the experiment were to be repeated, the same sets of flumes would be used.

In this case the block effect is fixed

→ Let's say the Blocks are fixed

OR, if the experimental units are individually labelled (always a good idea):

- Block → blocking factor
- Trt → treatment
- Flume → the **experimental unit (Flume)**
  - a **random effect**

nested effect: leave it up to the computer package:

In this case, the computer package should be able to figure out that flumes are the experimental units and fish are the sub-sampling units.

To fit the model in R:

YOU MUST HAVE UNIQUE LABELS for each flume  
and just add it as a random effect

- Use the lme() function in the nlme package

The functions cannot deal with the nesting term

→ create unique labels for each FLUME using the interaction()  
function.

- Created unique labels for each flume using the **interaction()** function.
- Flume and Block variables must be declared as a factor rather than a continuous variable

```

> # Be sure to specify that Flume and Block are a factor
> uvexp$Flume <- as.factor(uvexp$Flume)
> uvexp$Block <- as.factor(uvexp$Block)
>
>
> library(nlme)
> # Create a unique FLUME id for each combination of treatment and flume
> # to avoid having to specify the nesting of flume(treatment)
> uvexp$Flume.id <- interaction(uvexp$Trt,uvexp$Flume)

```

```

> model2 <- lme( WeightGain ~ Block + Trt, random=~1 | Flume.id,
  data=uvexp)
> anova(model2)

```

	numDF	denDF	F-value	p-value
(Intercept)	1	36	60.73861	<.0001
Block	2	4	1.29666	0.3681
Trt	2	4	2.64795	0.1852

The F-statistics are identical to the simple analysis presented before.

(block effect is not of interest)

## Certain functions don't work with lme() models!

```
> # Get the marginal means
> popMeans(result2, effect='Trt')
Error in terms.default(model.frame(object)) :
  no terms component nor attribute
```

```
> # Multiple comparison
> library(multcomp)
>
> model2.tukey <- glht(model2, linfct = mcp(Trt = "Tukey"))
> summary(model2.tukey)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lme.formula(fixed = WeightGain ~ Block + Trt, data = uvexp, random = ~1 | Flume.id)

Linear Hypotheses:

		Estimate	Std. Error	z value	Pr(> z )
UVA	- Control == 0	-0.3333	0.6649	-0.501	0.8707
UVAB	- Control == 0	-1.4600	0.6649	-2.196	0.0719 .
UVAB	- UVA == 0	-1.1267	0.6649	-1.694	0.2073
---					

Signif. codes: 0 '\*\*\*\*' 0.001 '\*\*\*' 0.01 '\*\*' 0.05 '\*' 0.1 '.' 1  
(Adjusted p values reported -- single-step method)

```
> confint(model2.tukey)
```

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: lme.formula(fixed = WeightGain ~ Block + Trt, data = uvexp, random = ~1 | Flume.id)

Quantile = 2.3441

95% family-wise confidence level

Linear Hypotheses:

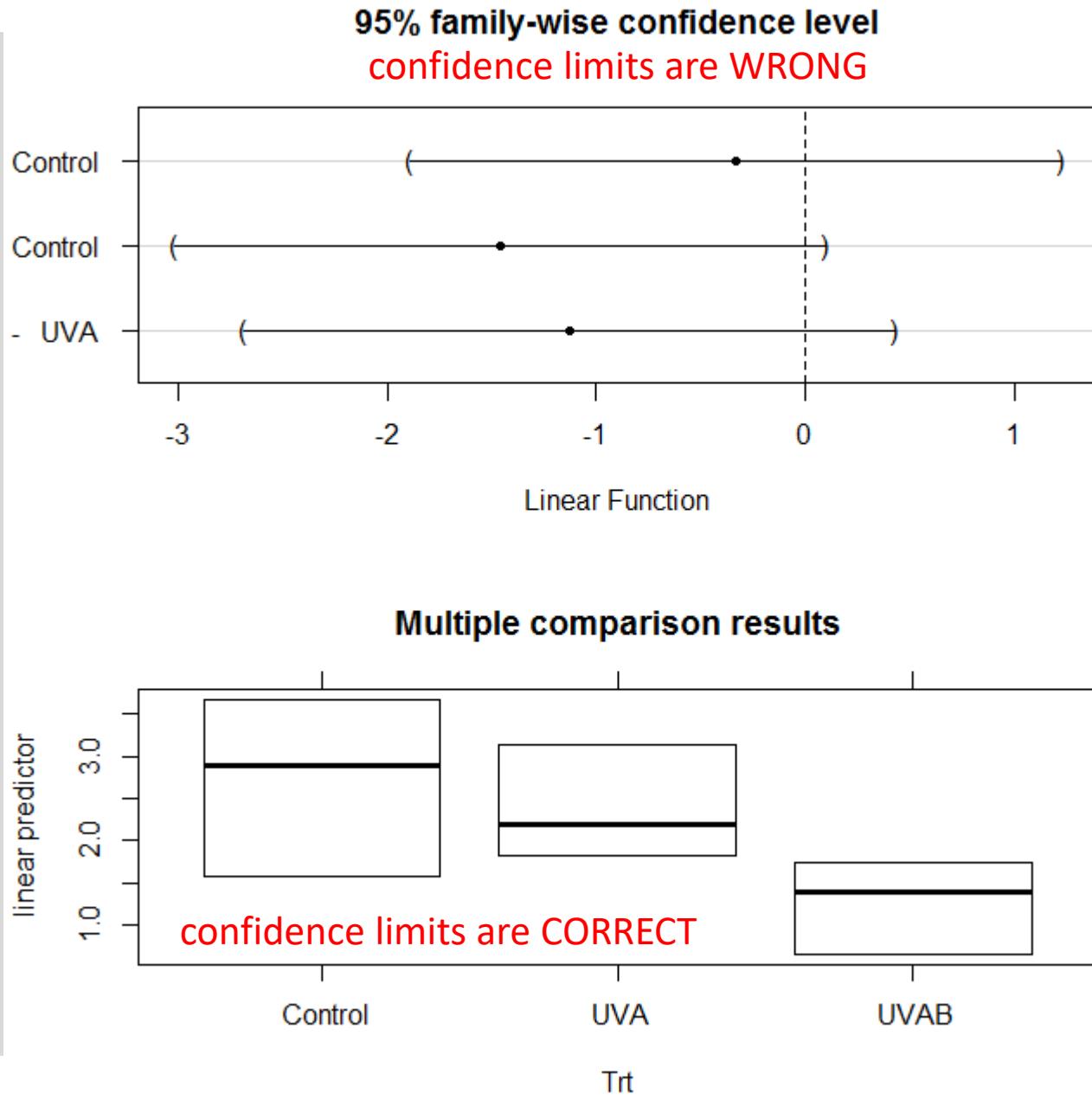
		Estimate	lwr	upr
UVA	- Control == 0	-0.33333	-1.89199	1.22532
UVAB	- Control == 0	-1.46000	-3.01866	0.09866
UVAB	- UVA == 0	-1.12667	-2.68532	0.43199

These confidence limits are WRONG.

```

> model2.tukey.cld <-
cld(model2.tukey)
> model2.tukey.cld
  Control          UVA        UVAB
    "a"           "a"       "a"
>
> old.par <- par()
> layout(matrix(1:2,2,1))
> plot(model2.tukey)
> #Note that the Joined Letter plot
isn't produced, and I don't know
how to fix them
>
> plot(model2.tukey.cld,
main="Multiple comparison results",
+       xlab="Trt",
+       ylab="WeightGain",
+       notch=FALSE)

```



Finally, this analysis also gives estimates of the variance components:

```
> vc <- VarCorr(model12)
> vc
Flume.id = pdLogChol(1)
          Variance StdDev
(Intercept) 0.5908614 0.7686751
Residual     0.3615559 0.6012952
```

Variance among experimental units: FLUME (random factor)

Variance among observational units (sub-samples): fish

→ The largest amount of variation is among the flumes but the fish within a flume also show substantial variation.

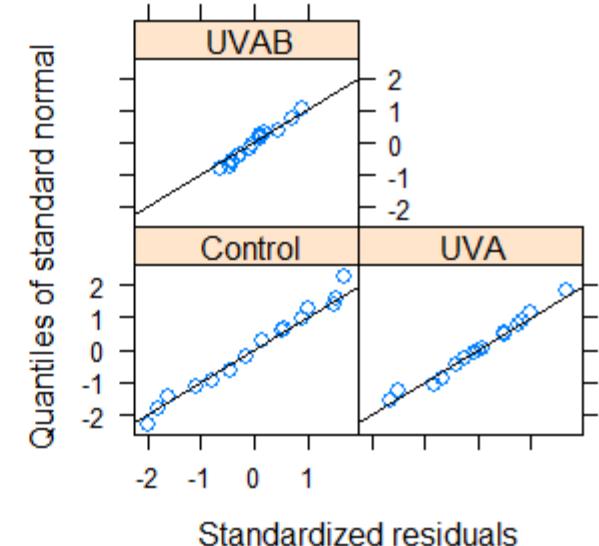
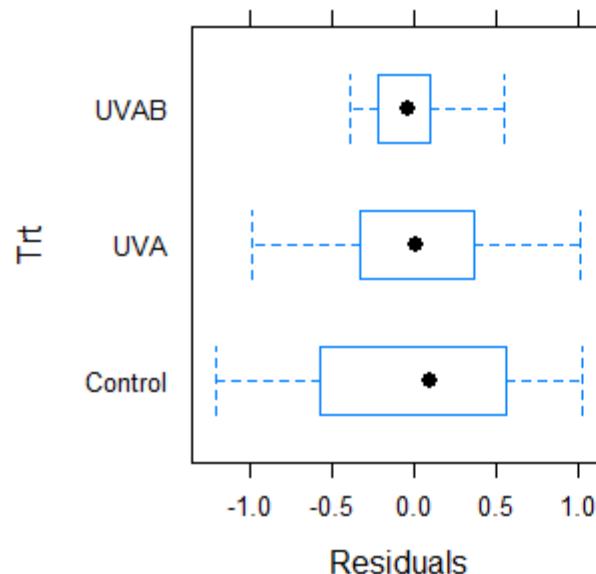
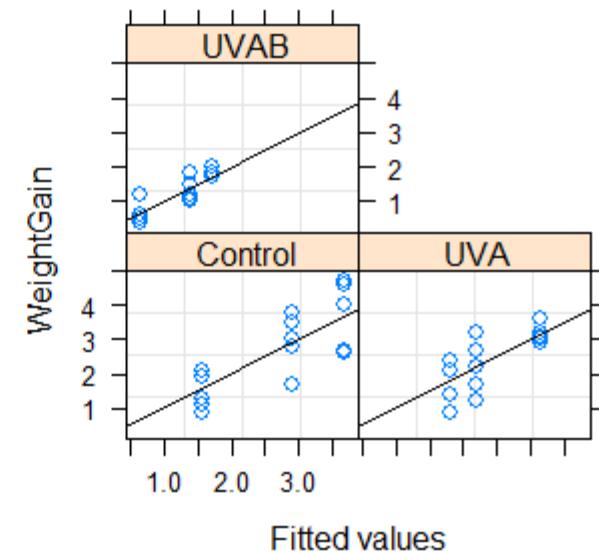
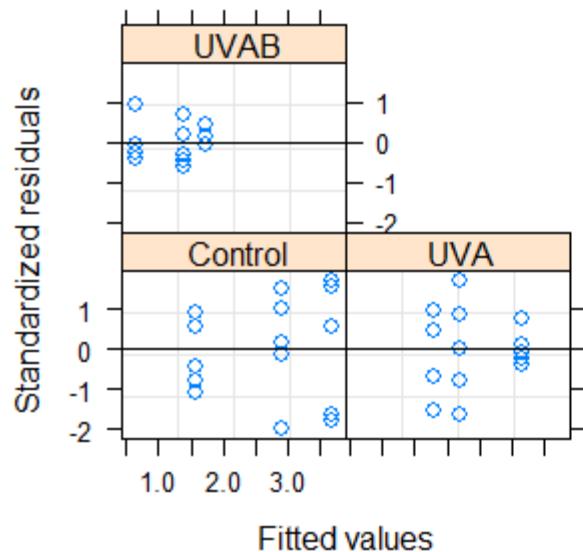
The variance components could be used in a more detailed cost-benefit analysis for planning an experiment as shown earlier.

# Model diagnostic plots: check the residuals

```
> # lme() uses Trellis graphics, so the usual plotting commands are not
useful
> plot1 <- plot(model2, resid(., type = "p") ~ fitted(.) | Trt, abline = 0)
> # box-plots of residuals by Subject
> plot2 <- plot(model2, Trt ~ resid(.))
> # observed versus fitted values by Subject
> plot3 <- plot(model2, WeightGain ~ fitted(.) | Trt, abline = c(0,1))
> # normak probability plot of the residuals.
> plot4 <- qqnorm(model2, ~ resid(., type = "p") | Trt, abline = c(0, 1))
>
> print(plot1, split=c(1,1,2,2), more=TRUE)
> print(plot2, split=c(1,2,2,2), more=TRUE)
> print(plot3, split=c(2,1,2,2), more=TRUE)
> print(plot4, split=c(2,2,2,2))
```

# The model diagnostic plots:

→ doesn't show a problem with the fit



# Example 3

## Monitoring Fry Levels

-

unbalanced data with sampling  
over time

# CAUTION when TIME is a factor!

Impossible to randomize time (i.e. readings in 2004 must be taken after readings in 2003)

BUT assumption of CRD and RCB: randomization of treatments to experimental units

- This randomization implies that the correlation of responses between any pair of units is equal.

In experiments with repeated measurements taken over time: “AUTOCORRELATION”:

- The **measurements closer together in time are more highly correlated than measurement far apart in time.**
- Problems with the computation of proper p-values and standard errors of effect sizes.

More refined analysis of experiments with repeated measurements over time exist:

repeated measures analysis

- (beyond the scope of this course)
  - Requires more experimental units than typically available
- not even feasible

model the autocorrelation structure in time

- e.g. Proc Mixed in SAS
  - (beyond the scope of this course)
  - Requires more experimental units than typically available
- not even feasible

BUT the methods presented below are a good first close approximation!

The experiment:  
monitoring the health of streams :  
measure the density of fry (small fish)  
over time



- repeated measurements are taken for a few years
- 5 to 20 locations per stream (depending on the length and size of the stream)

If density declines over time, it may be an indication that the “health” of the stream is declining.

# The key objectives of the study:

1. what are the fry densities in a given year
2. how have densities changed from one year to the next?
3. use the data to determine how many locations and sites per stream are required to detect a change in fry densities of  $\pm 25\%$  from one year to the next.

Data: fry.csv

Script: fry.R

- Salmon fry densities were monitored at 4 different rivers.
- Up to three sites were selected from each river.
- Not all sites were monitored in all years.

Location	Site	Fry Density in each year				
		2000	2001	2002	2003	2004
A	1	4	55	28	12	9
A	2	25	11	45	84	27
A	3	.	.	27	.	.
B	1	139	234	496	349	209
B	2	272	262	102	90	35
B	3	.	127	.	.	.
C	1	34	249	91	79	124
C	2	122	.	.	.	.
D	1	128	213	.	97	.
E	1	184	47	131	107	103
E	2	413	508	204	323	115
E	3	70	.	.	.	.
F	1	140	307	189	243	110
F	2	181	326	361	468	186

# Data in wide format

```
> fry.wide <- read.csv("fry.csv", header=TRUE, na.strings=".")  
> fry.wide
```

	Location	Site	X2000	X2001	X2002	X2003	X2004
1	A	1	4	55	28	12	9
2	A	2	25	11	45	84	27
3	A	3	NA	NA	27	NA	NA
4	B	1	139	234	496	349	209
5	B	2	272	262	102	90	35
6	B	3	NA	127	NA	NA	NA
7	C	1	34	249	91	79	124
8	C	2	122	NA	NA	NA	NA
9	D	1	128	213	NA	97	NA
10	E	1	184	47	131	107	103
11	E	2	413	508	204	323	115
12	E	3	70	NA	NA	NA	NA
13	F	1	140	307	189	243	110
14	F	2	181	326	361	468	186

These columns should  
be a single column  
« Density »

# Data in long (stacked) format

```
> fry <- reshape(fry.wide,
+ varying=c('x2000','x2001','x2002','x2003','x2004'),
+               v.names='Density',
+               times=2000:2004, timevar='Year',
+               direction="long")
> # Drop missing values
> fry <- fry[!is.na(fry[, "Density"]), ]
> fry <- cbind(fry, logDensity=log(fry[, "Density"]))
> fry[1:10,]

  Location Site Year Density id logDensity
1.2000      A   1 2000       4  1  1.386294
2.2000      A   2 2000      25  2  3.218876
4.2000      B   1 2000     139  4  4.934474
5.2000      B   2 2000     272  5  5.605802
7.2000      C   1 2000      34  7  3.526361
8.2000      C   2 2000     122  8  4.804021
9.2000      D   1 2000     128  9  4.852030
10.2000     E   1 2000     184 10  5.214936
11.2000     E   2 2000     413 11  6.023448
12.2000     E   3 2000      70 12  4.248495
```

### Locations: Blocks

- group readings of fry density that should be similar over time.
- For example, Location A looks as if it is relatively poor habitat for fry compared to Location B.
- *Random or fixed effect?*
  - *a random sample of many locations in the stream → random effect*
  - *chosen for accessibility → Fixed effect*
  - *as long as we are interested in the year-to-year changes, it really doesn't matter*

Sites within each Location: **sub-blocks** within each of the locations

- **sub-sampling!**
- *Chosen at random → Random effect*

In this experiment, the same sites were  
repeatedly measured over time  
→ Sites = sub-blocks

### Year: Treatment

- We are interested in how fry densities change over time in response to year effects
- *Randomization: problematic as time cannot be randomized.*

### Fry density: response variable

- measured at each site within a location at each year.

## Sub-sampling: Layers of experimental units

- In one sense, locations are one unit - six different locations were measured at this stream.
- At the next level, between 1 and 3 sites were measured at each location.
- Finally, each site-location pair was measured between 1 and 5 times over the course of the experiment.

# Preliminary plots: always wise

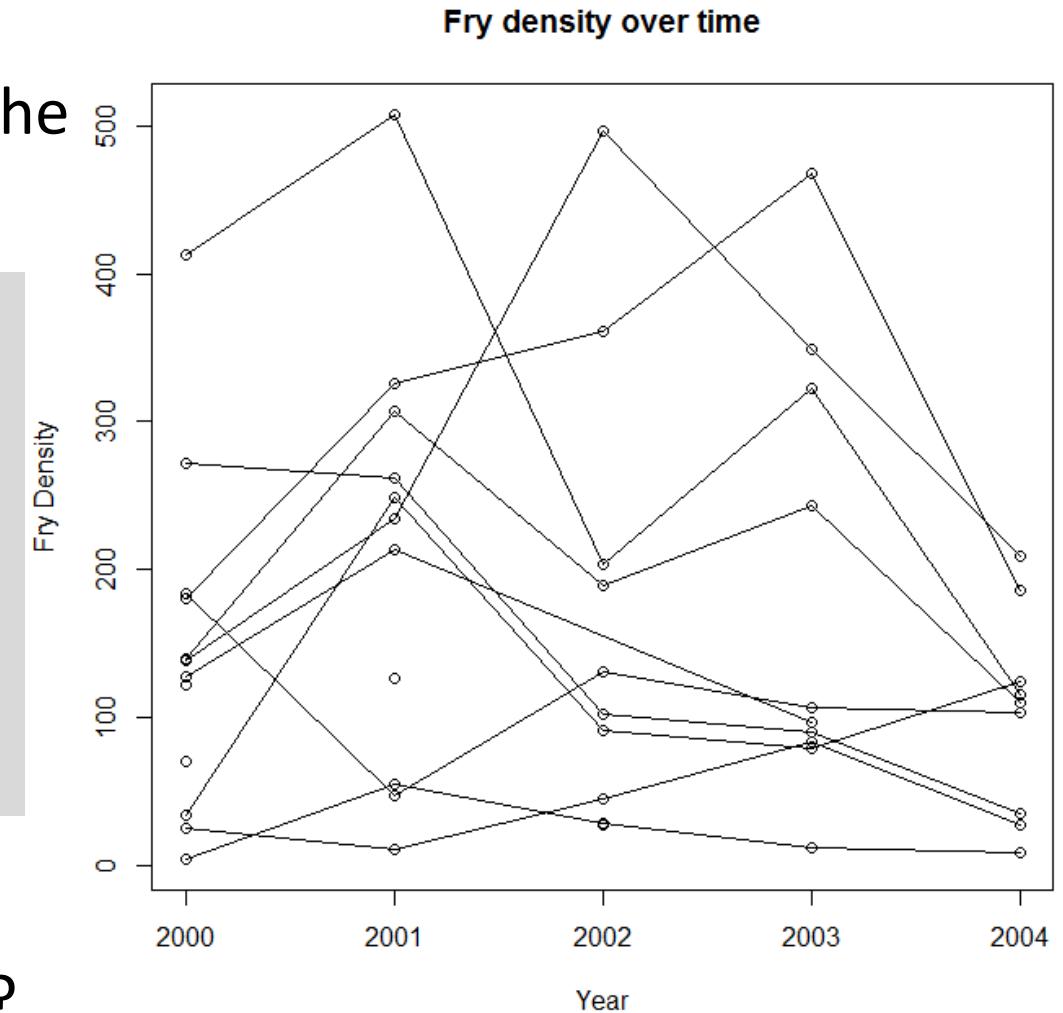
- how the response variable will change as the treatments (the different years) change?

```
> plot(fry[, "Year"], fry[, "Density"],
+       main='Fry density over time',
xlab='Year', ylab='Fry Density')

> temp <-
lapply(split(fry, list(Location=fry[, "Location"],
Site=fry[, "Site"])),
+     function(x) {
+       lines(x$Year, x$Density) })
```

→ a tremendous scatter!

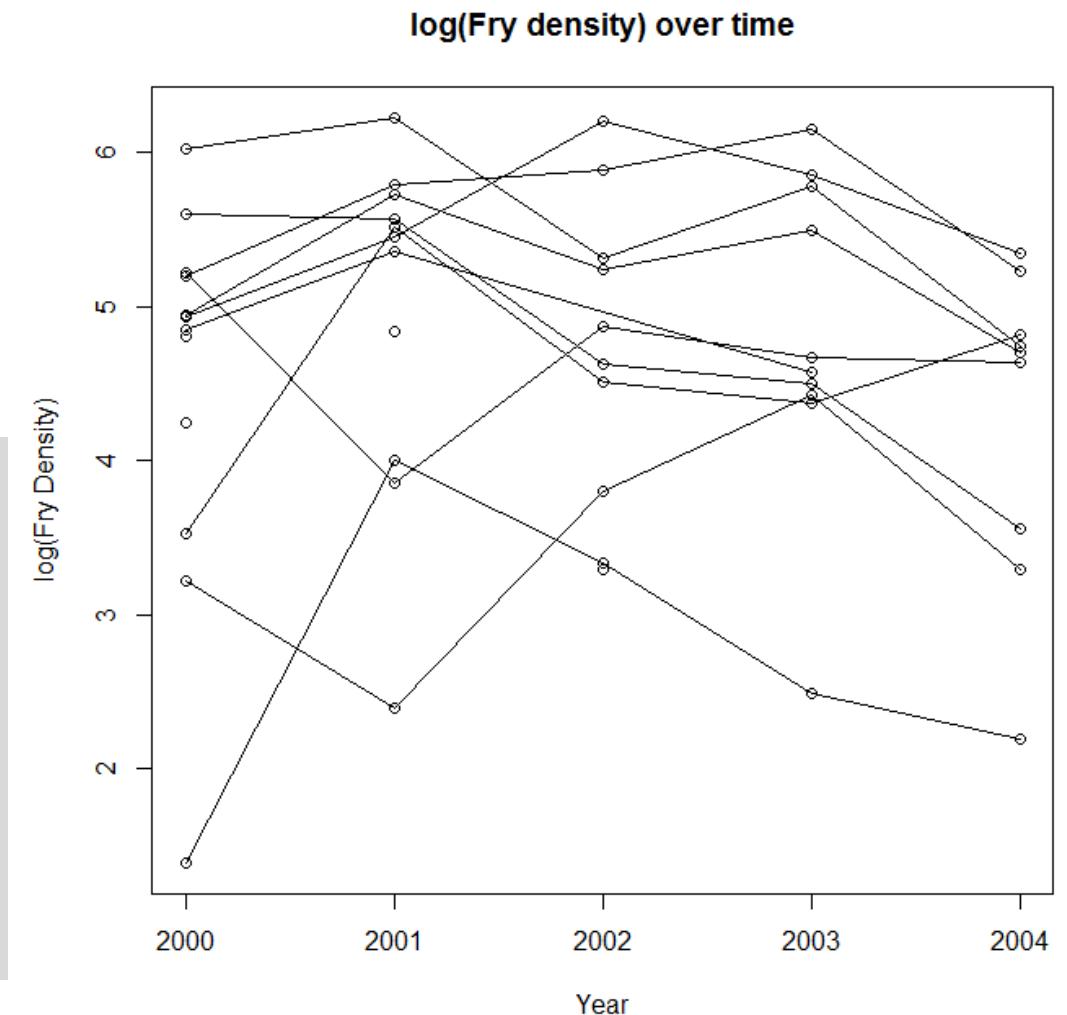
BUT Is the data measured on the right scale?



- It seems reasonable that yearly effects should be multiplicative rather than additive.
- For this reason, a logarithmic transform of the response variable is often used.

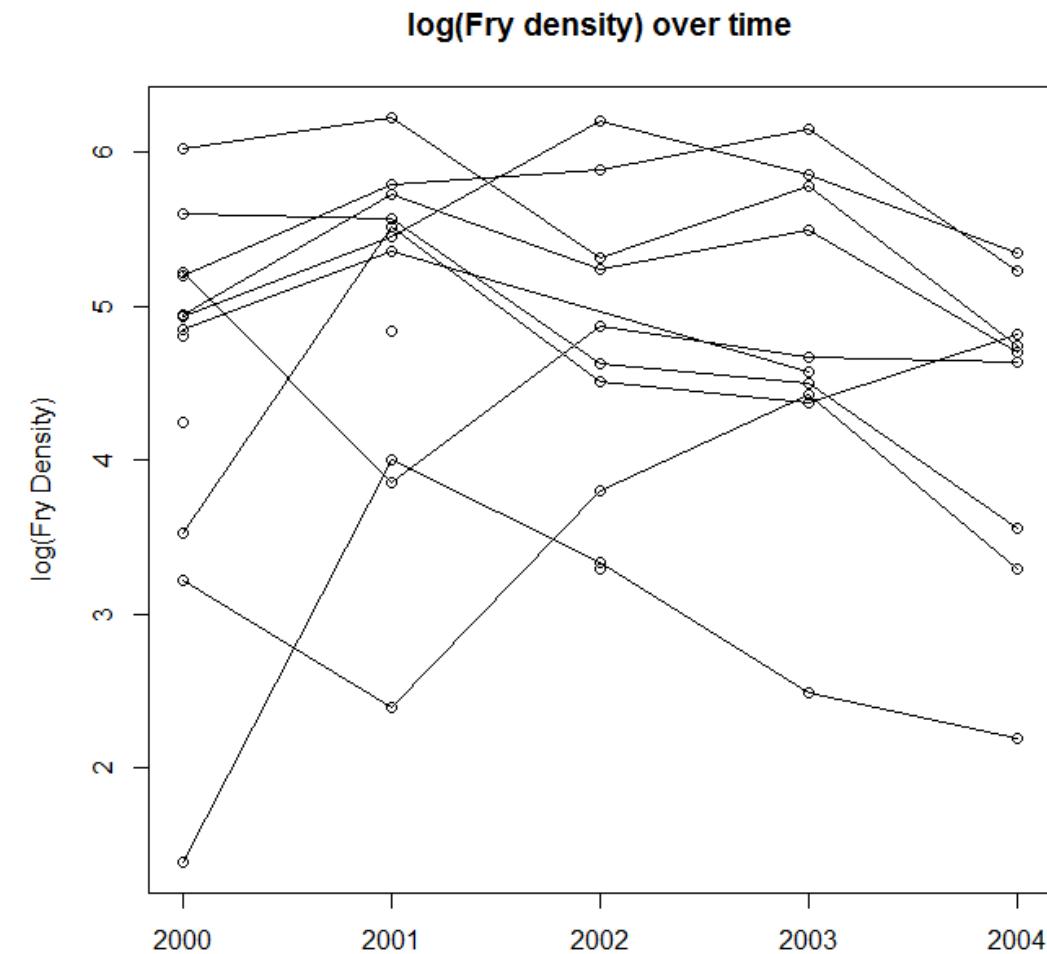
```
> plot(fry[, "Year"], fry[, "logDensity"],
+       main='log(Fry density) over time',
xlab='Year', ylab='log(Fry Density')

> temp <-
lapply(split(fry, list(Location=fry[, "Location"],
Site=fry[, "Site"])),
+     function(x) {
+       lines(x$Year, x$logDensity) })
```



A plot of  $\log(\text{density})$  over time is much easier to interpret:

- Most of the lines appear to be parallel
- Some evidence of a decline
- The variation appears to be “smaller” than in the previous graph.



The parallelism of response is important because one assumption of blocked designs is that responses should have a similar profile (i.e. parallel effects) and that blocks do not interact with treatments.

# 1. Approximate analysis of means

- Because the data are unbalanced, i.e. not every site in a location was measured every year, an analysis on the averages will only be approximate.
- A more exact analysis is presented in the next section.

The `summaryBy()` function in the `doBy` package can be used to get the average density of fry at each location over the multiple sites measured:

Create a new dataset

```
> library(doBy)
> avg <- summaryBy(logDensity ~ Location + Year , FUN=mean,
data=fry)
> avg
```

	Location	Year	logDensity.mean
1	A	2000	2.302585
2	A	2001	3.202614
3	A	2002	3.478235
4	A	2003	3.457862
5	A	2004	2.746531
6	B	2000	5.270138
7	B	2001	5.289284
8	B	2002	5.415774
9	B	2003	5.177441
10	B	2004	4.448841
11	C	2000	4.165191...

Location = BLOCKS

Year = Treatment

→ There should be ONE average value for each combination of year and location!

→ a single-factor randomized block ANOVA

- The `lm()` function can be used to fit a single-factor RCB ANOVA on the sample means (same way was seen in previous examples)
- Be sure that **Location** and **Year** variables are declared as factors!

```
> avg$Location <- as.factor(avg$Location)
> avg$Year <- as.factor(avg$Year)
>
> model <- lm(logDensity.mean ~ Location + Year, data=avg)
> anova(model)
Analysis of Variance Table

Response: logDensity.mean
            Df  Sum Sq Mean Sq F value    Pr(>F)
Location      5 18.1506  3.6301 29.6556 4.318e-08 ***
Year          4  1.6417  0.4104  3.3529  0.03226 *
Residuals   18  2.2034  0.1224
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ONE average value for each combination of year and location!



- Year is usually specified last in model because R gives Type I tests.
- In this case, the design is balanced (for each treatment level (year), 6 observations (locations)), so it doesn't matter.

→ Some evidence of a year effect:  $p$ -value just over 3%

# Estimate the population marginal means (the LSmeans)

- Use the popMeans() function in the doBy package

```
> lsmeans <- popMeans(model, eff="Year")
> lsmeans
   estimate      se  df t.stat    p.value Year
1 4.470384 0.1428339 18 31.29779 3.791976e-17 2000
2 5.027972 0.1428339 18 35.20153 4.721354e-18 2001
3 4.847952 0.1607542 18 30.15754 7.308444e-17 2002
4 4.770911 0.1428339 18 33.40182 1.197938e-17 2003
5 4.368304 0.1607542 18 27.17380 4.587596e-16 2004
```

→ 2001 was a good year for fry

# Multiple comparisons

(using the multcomp package - the usual way)

```

> library(multcomp)
> model.tukey <- glht(model, linfct = mcp(Year = "Tukey"))
> summary(model.tukey)

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = logDensity.mean ~ Location + Year, data = avg)

Linear Hypotheses:
Estimate Std. Error t value Pr(>|t|) 
2001 - 2000 == 0 0.55759 0.20200 2.760 0.0833 .
2002 - 2000 == 0 0.37757 0.21504 1.756 0.4270
2003 - 2000 == 0 0.30053 0.20200 1.488 0.5823
2004 - 2000 == 0 -0.10208 0.21504 -0.475 0.9887
2002 - 2001 == 0 -0.18002 0.21504 -0.837 0.9153
2003 - 2001 == 0 -0.25706 0.20200 -1.273 0.7101
2004 - 2001 == 0 -0.65967 0.21504 -3.068 0.0458 *
2003 - 2002 == 0 -0.07704 0.21504 -0.358 0.9961
2004 - 2002 == 0 -0.47965 0.22128 -2.168 0.2358
2004 - 2003 == 0 -0.40261 0.21504 -1.872 0.3658
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```

```
> confint(model.tukey)
```

### Simultaneous Confidence Intervals

#### Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = logDensity.mean ~ Location + Year, data = avg)

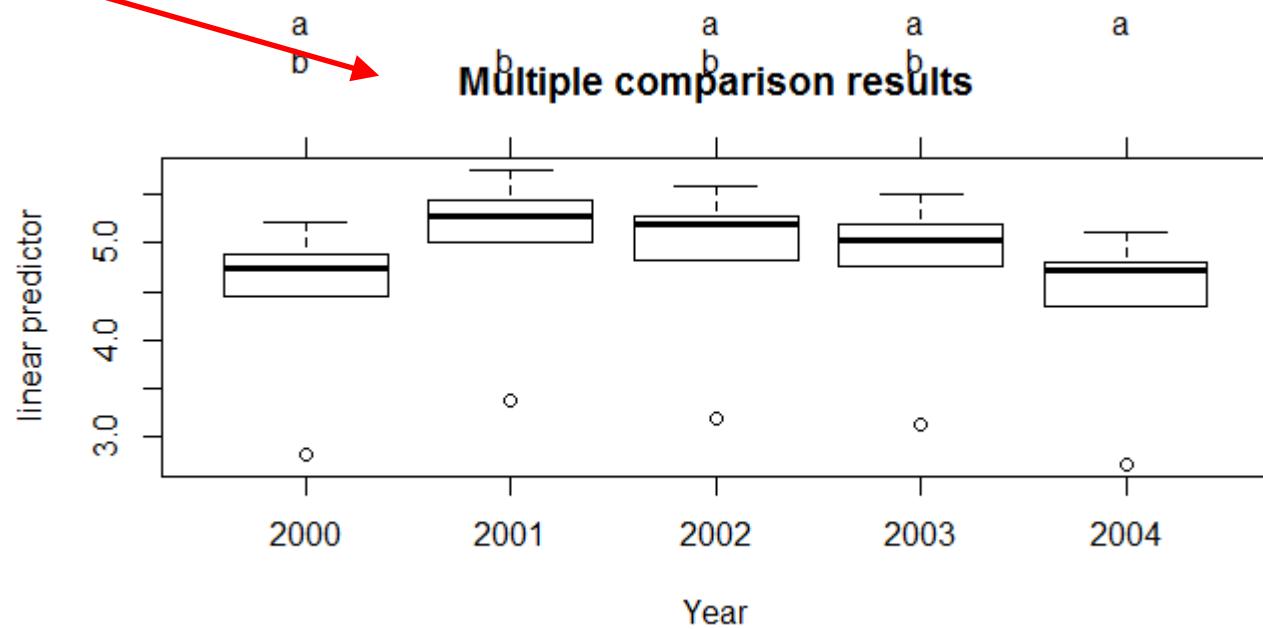
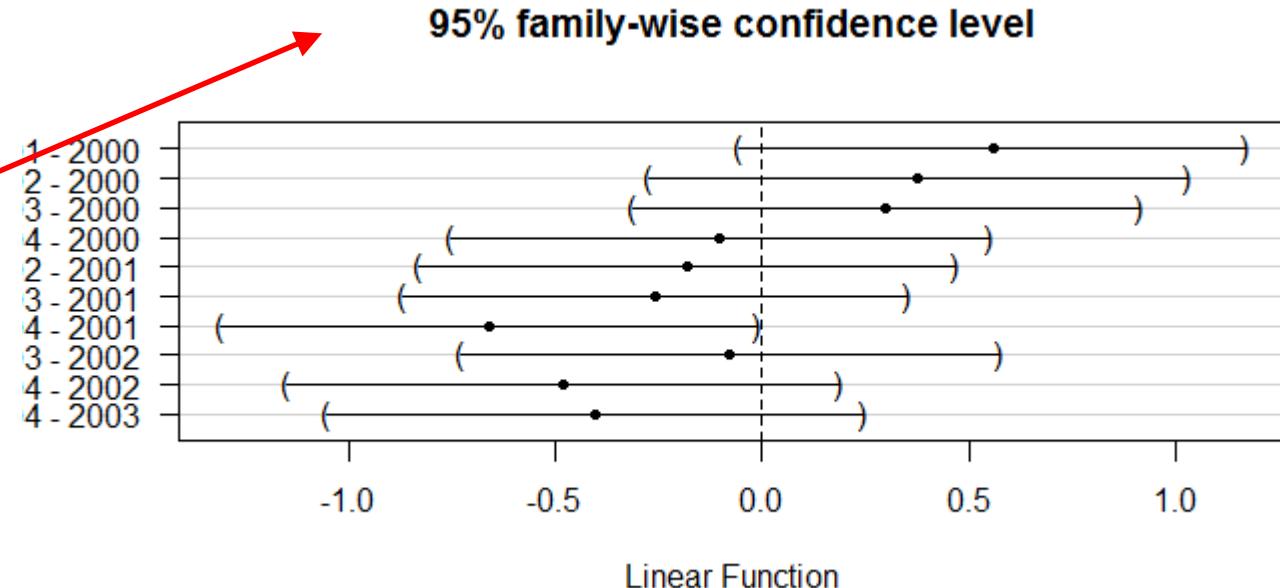
Quantile = 3.023

95% family-wise confidence level

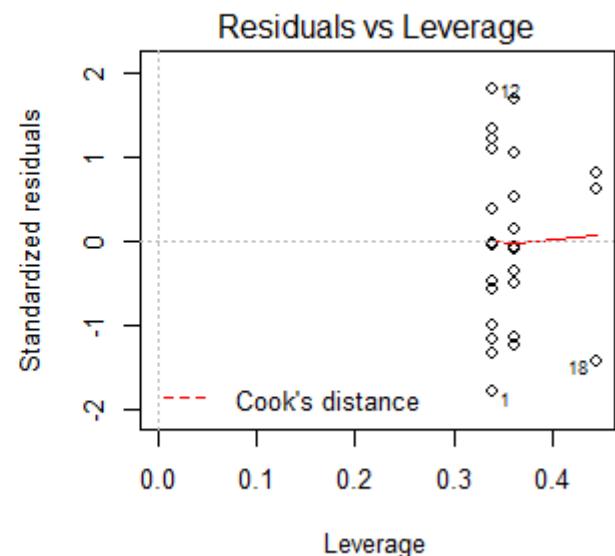
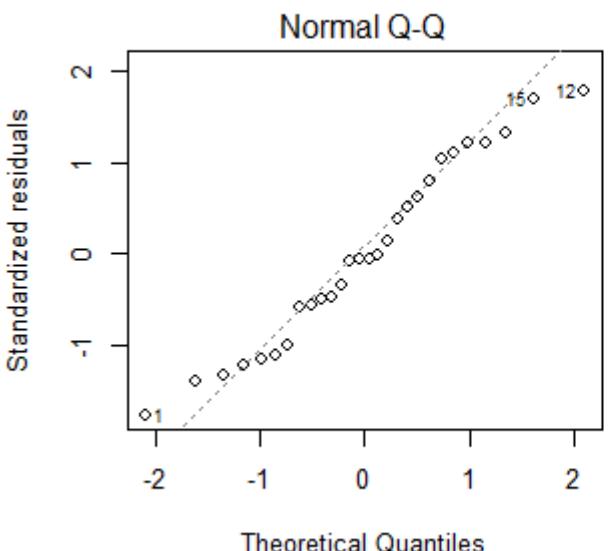
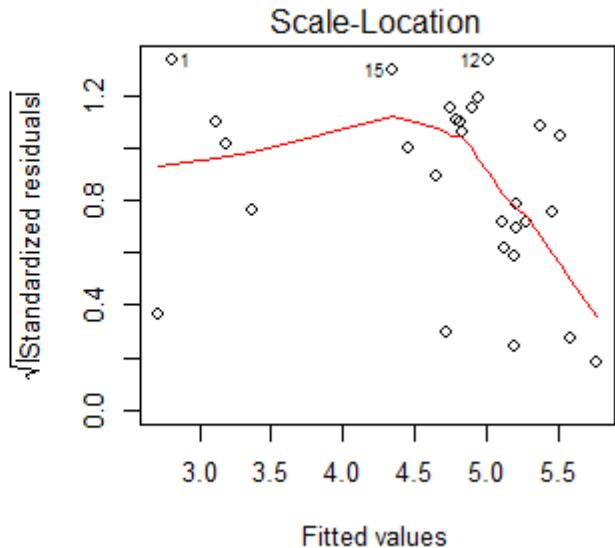
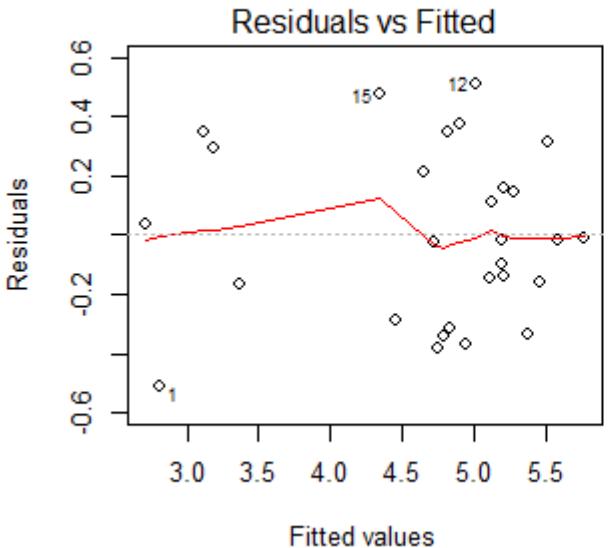
#### Linear Hypotheses:

		Estimate	lwr	upr	
2001	- 2000	== 0	0.55759	-0.05304	1.16822
2002	- 2000	== 0	0.37757	-0.27250	1.02764
2003	- 2000	== 0	0.30053	-0.31011	0.91116
2004	- 2000	== 0	-0.10208	-0.75215	0.54799
2002	- 2001	== 0	-0.18002	-0.83009	0.47005
2003	- 2001	== 0	-0.25706	-0.86769	0.35357
2004	- 2001	== 0	-0.65967	-1.30974	-0.00960
2003	- 2002	== 0	-0.07704	-0.72711	0.57303
2004	- 2002	== 0	-0.47965	-1.14856	0.18927
2004	- 2003	== 0	-0.40261	-1.05267	0.24746

```
> model.tukey.cld <-  
cld(model.tukey) > model.tukey.cld  
2000 2001 2002 2003 2004  
"ab" "b" "ab" "ab" "a"  
>  
> old.par <- par()  
> layout(matrix(1:2, 2, 1))  
> plot(model.tukey)  
> plot(model.tukey.cld,  
main="Multiple comparison results",  
+       xlab="Year",  
+       ylab="fat",  
+       notch=FALSE)  
> par <- old.par
```



```
> # Check the assumptions of the  
linear model on the averages  
> layout(matrix(1:4, 2, 2))  
> plot(model)
```



→ doesn't show anything unusual about the fit.

## 2. Analysis using individual values

The model must include the following terms:

- Location → blocking factor
- Year → treatment
- Site(Location) → the **sub-blocking** factor
  - a **nested** effect ....(Location))
  - a **random** effect

To fit the model in R:

YOU MUST HAVE UNIQUE LABELS for each SITE  
and add it as a random effect

- Use the lme() function in the nlme package

The functions cannot deal with the nesting term

→ create unique labels for each SITE using the **interaction()** function.

- Year and Location variables must be declared as a factor rather than a continuous variable

```

> # Be sure to specify that Year and Location are factors
> fry$Year <- as.factor(fry$Year)
> fry$Location <- as.factor(fry$Location)
>
>
> library(nlme)
> # Create a unique Site-Location id for each combination of Site and Location
> # to avoid having to specify the nesting of site(location)
> fry$Site.id <- interaction(fry$Location,fry$Site)

```

First the blocking factor  


```

> model2 <- lme( logDensity ~ Location + Year, random=~1 | Site.id, data=fry)
> anova(model2)
      numDF denDF   F-value p-value
(Intercept)     1     34 1125.3826  <.0001
Location        5      8    7.1799  0.0079  (block effect is not of interest)
Year            4     34    2.1225  0.0994

```

➔ This more refined analysis gives a *p*-value for the year effect of around 10%.

Previous analysis is very sensitive to the missing values.

If you remove the locations with a limited number of years, the results are in much closer agreement.

## Certain functions don't work with lme() models!

```
> # Get the marginal means
> popMeans(result2, effect='Trt')
Error in terms.default(model.frame(object)) :
  no terms component nor attribute
```

```
> # Multiple comparison
> library(multcomp)
>
> model2.tukey <- glht(model2, linfct = mcp(Trt = "Tukey"))
> summary(model2.tukey)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lme.formula(fixed = WeightGain ~ Block + Trt, data = uvexp, random = ~1 | Flume.id)

Linear Hypotheses:

		Estimate	Std. Error	z value	Pr(> z )
UVA	- Control == 0	-0.3333	0.6649	-0.501	0.8707
UVAB	- Control == 0	-1.4600	0.6649	-2.196	0.0719 .
UVAB	- UVA == 0	-1.1267	0.6649	-1.694	0.2073

---

Signif. codes: 0 '\*\*\*\*' 0.001 '\*\*\*' 0.01 '\*\*' 0.05 '\*' 0.1 '.' 1  
(Adjusted p values reported -- single-step method)

```
> confint(model2.tukey)
```

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: lme.formula(fixed = WeightGain ~ Block + Trt, data = uvexp, random = ~1 | Flume.id)

Quantile = 2.3441

95% family-wise confidence level

Linear Hypotheses:

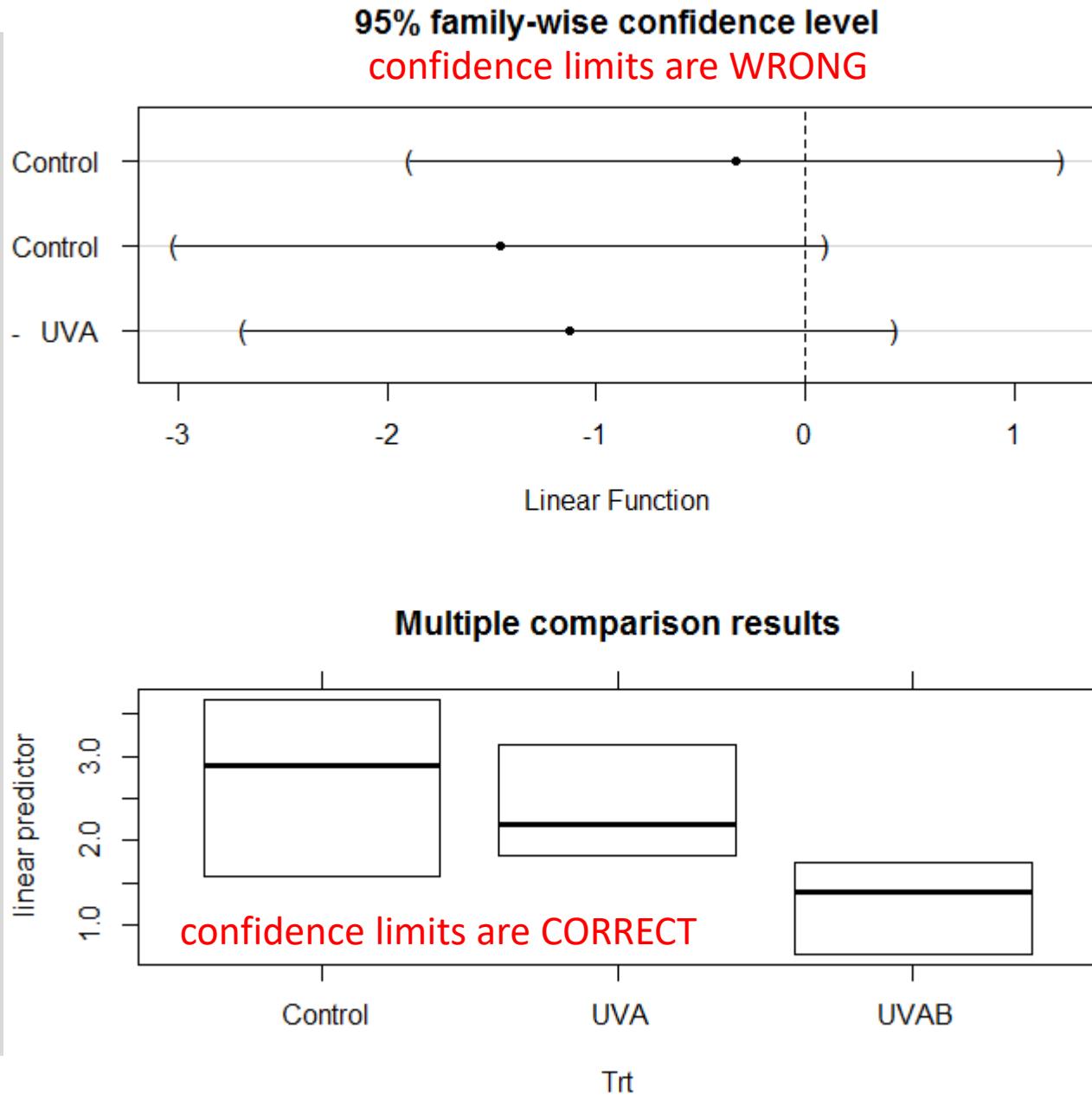
		Estimate	lwr	upr
UVA	- Control == 0	-0.33333	-1.89199	1.22532
UVAB	- Control == 0	-1.46000	-3.01866	0.09866
UVAB	- UVA == 0	-1.12667	-2.68532	0.43199

These confidence limits are WRONG.

```

> model2.tukey.cld <-
cld(model2.tukey)
> model2.tukey.cld
  Control          UVA        UVAB
    "a"           "a"       "a"
>
> old.par <- par()
> layout(matrix(1:2,2,1))
> plot(model2.tukey)
> #Note that the Joined Letter plot
isn't produced, and I don't know
how to fix them
>
> plot(model2.tukey.cld,
main="Multiple comparison results",
+       xlab="Trt",
+       ylab="WeightGain",
+       notch=FALSE)

```



# Finally, this analysis also gives estimates of the variance components:

```
> vc <- VarCorr(model12)
> vc
Site.id = pdLogChol(1)
      Variance StdDev
(Intercept) 0.1447539 0.3804653
Residual     0.3774972 0.6144080
```

Variance among the sites  
(random factor)

Unexplained variance: 70%

The total variation is 0.52

About 70% of the variation is yet unexplained

Only about 30% of the variation is found among sites within locations.

→ densities are not very consistent at the same site over time.

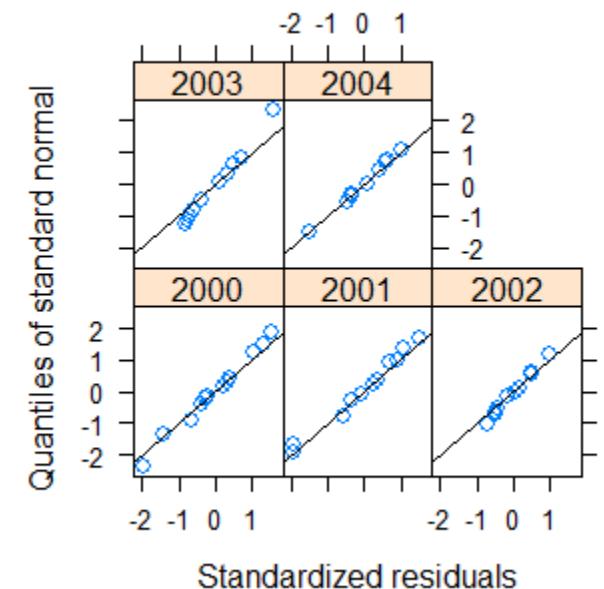
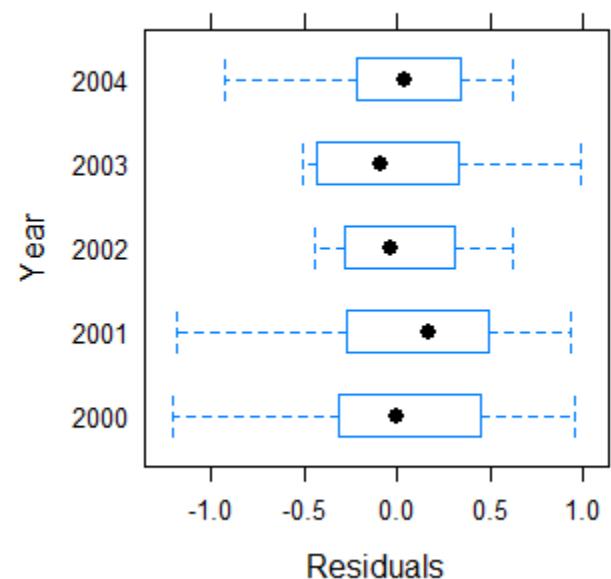
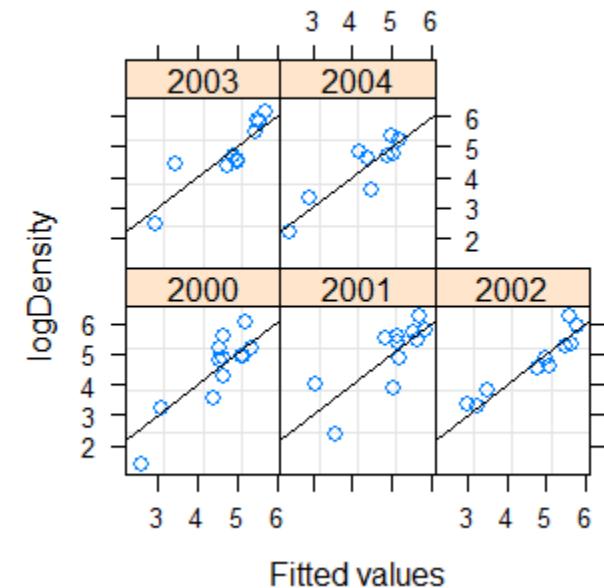
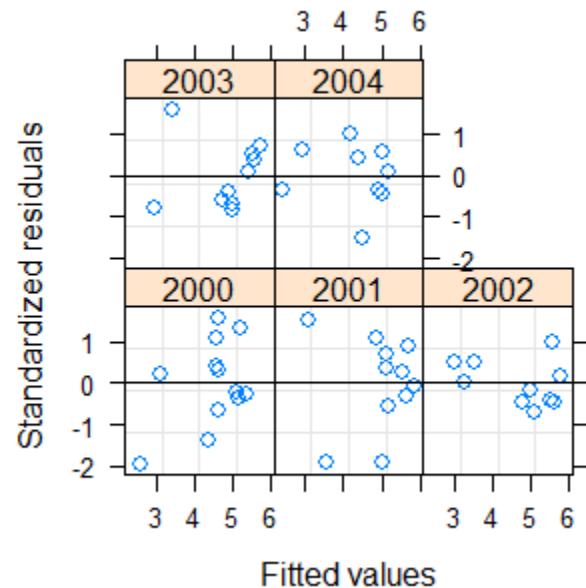
→ little advantage to always returning to the same site within each location over time.

# Model diagnostic plots: check the residuals

```
> # lme() uses Trellis graphics, so the usual plotting commands are not
useful
> plot1 <- plot(model2, resid(., type = "p") ~ fitted(.) | Year, abline = 0)
> # box-plots of residuals by Subject
> plot2 <- plot(model2, Year ~ resid(.))
> # observed versus fitted values by Subject
> plot3 <- plot(model2, logDensity ~ fitted(.) | Year, abline = c(0,1))
> # normak probability plot of the residuals.
> plot4 <- qqnorm(model2, ~ resid(., type = "p") | Year, abline = c(0, 1))
>
> print(plot1, split=c(1,1,2,2), more=TRUE)
> print(plot2, split=c(1,2,2,2), more=TRUE)
> print(plot3, split=c(2,1,2,2), more=TRUE)
> print(plot4, split=c(2,2,2,2))
```

# The model diagnostic plots:

→ doesn't show a problem with the fit



# Example 4

## Comparing mean flagella lengths

---

sub-sampling  
pseudo-replicates

# Pseudo-replicates

- Multiple measurements are taken from each experimental unit.
- should not be treated as independent measurements when **estimating the overall mean**

If this pseudo-replication is ignored, then

- the standard error of the mean (SE) is typically underreported  
*(underestimates the true uncertainty in the mean)*
- the confidence intervals will be too narrow

# Comparing the means across several populations

**The analysis needs to account for the pseudo-replicates.**

Two ways of proceeding:

1. “Average of averages”: First find the averages of the pseudo-replicates for each experimental unit. This reduces the data to a single value for each experimental unit. Then apply standard single-factor CRD ANOVA models.
2. Fit a more complex model that recognizes (and adjusts) for the pseudo-replication.

Balanced data (the same number of pseudo-replicates for each experimental unit):

The first option will give identical results to the second option

Unbalanced data:

The first option is only approximate

The key advantage of the second approach is that it is applicable in all cases (balanced or unbalanced) and also provides more information (the relative variation among- and within- experimental units).

# Compare the mean length of flagella among the variants of unicellular algae

- Microphotographs of algae are taken, and the length of one or two flagellum of each cell is measured.
  - The experimental unit is the cell (each row of the data)
  - pseudo-replicates are the multiple flagella measured on each cell.

Data: flagella.csv

Script: flagella.R

```

> lengths <- read.csv("flagella.csv", header=TRUE)
> lengths$Variant.Cell <-
paste(lengths$Variant,".",lengths$Cell...,sep="")
> lengths[1:10,]
  Variant Cell... Flagellar.1 Flagellar.2 Comment Variant.Cell
1      A      1       26.9        NA
2      A      2       29.8        NA
3      A      3       30.9        NA
4      A      4       29.0        NA
5      A      5       24.9       27.0
6      A      6       32.3        NA
7      A      7       28.3        NA
8      A      8       31.4        NA
9      A      9       28.5       35.0
10     A     10       25.8       32.4

```

Not every cell has two flagella measured because in some case the flagellum was hidden by other cells, broken, or not clearly visible in the micropograph

- A large variation among individual cells
  - BUT, if a cell tends to have longer flagella than other cells, then both flagella also tend to be longer.
- This lack-of-independence is what makes a naive analysis using all flagella lengths in a pooled sample inappropriate.

# First stack the data

```
> stack.length <- reshape(lengths, idvar="Variant.Cell",
+                         varying=list(c("Flagellar.1","Flagellar.2")),
+                         times=c("Flagellar.1","Flagellar.2"), timevar="Length",
+                         v.name="Length",
+                         direction="long")
> stack.length[1:10,]
      Variant Cell... Comment Variant.Cell Length
A.1.Flagellar.1     A     1                   A.1    26.9
A.2.Flagellar.1     A     2                   A.2    29.8
A.3.Flagellar.1     A     3                   A.3    30.9
A.4.Flagellar.1     A     4                   A.4    29.0
A.5.Flagellar.1     A     5                   A.5    24.9
A.6.Flagellar.1     A     6                   A.6    32.3
A.7.Flagellar.1     A     7                   A.7    28.3
A.8.Flagellar.1     A     8                   A.8    31.4
A.9.Flagellar.1     A     9                   A.9    28.5
A.10.Flagellar.1    A    10                  A.10   25.8
```

1. Average of averages approach

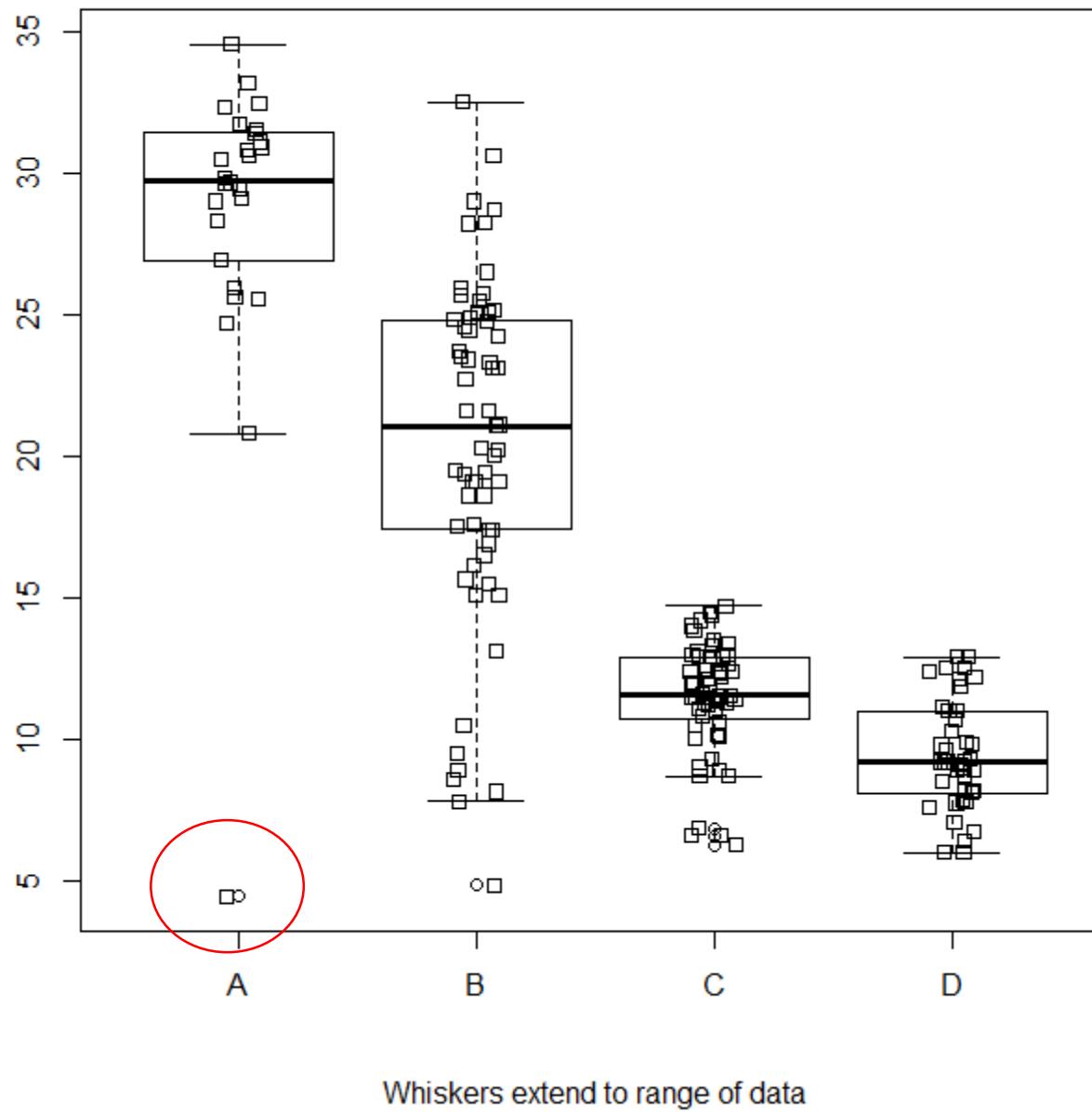
Average flagellum length for each cell is computed and the data is reduced to a single value for each cell.

```
> avg <- aggregate(Length ~ Variant+ Cell..., FUN=mean, data=stack.length)
> avg[1:10,]
  Variant Cell... Length
1          A      1  26.90
2          B      1  23.10
3          C      1  12.95
4          D      1  12.15
5          A      2  29.80
6          B      2  18.60
7          C      2   8.90
8          D      2  10.30
9          A      3 30.90
10         B     3   7.80
>
```

# Plot to check for outliers

```
> boxplot(Length ~ Variant, data=avg, main="Avg flagella length for each cell",
+   sub='Whiskers extend to range of data')
> stripchart(Length ~ Variant, data=avg, add=TRUE,
+   vertical=TRUE, method="jitter", jitter=.1,
+   main="Avg flagella length for each cell",
+   sub='Whiskers extend to range of data',
+   xlab='', ylab='Average length')
```

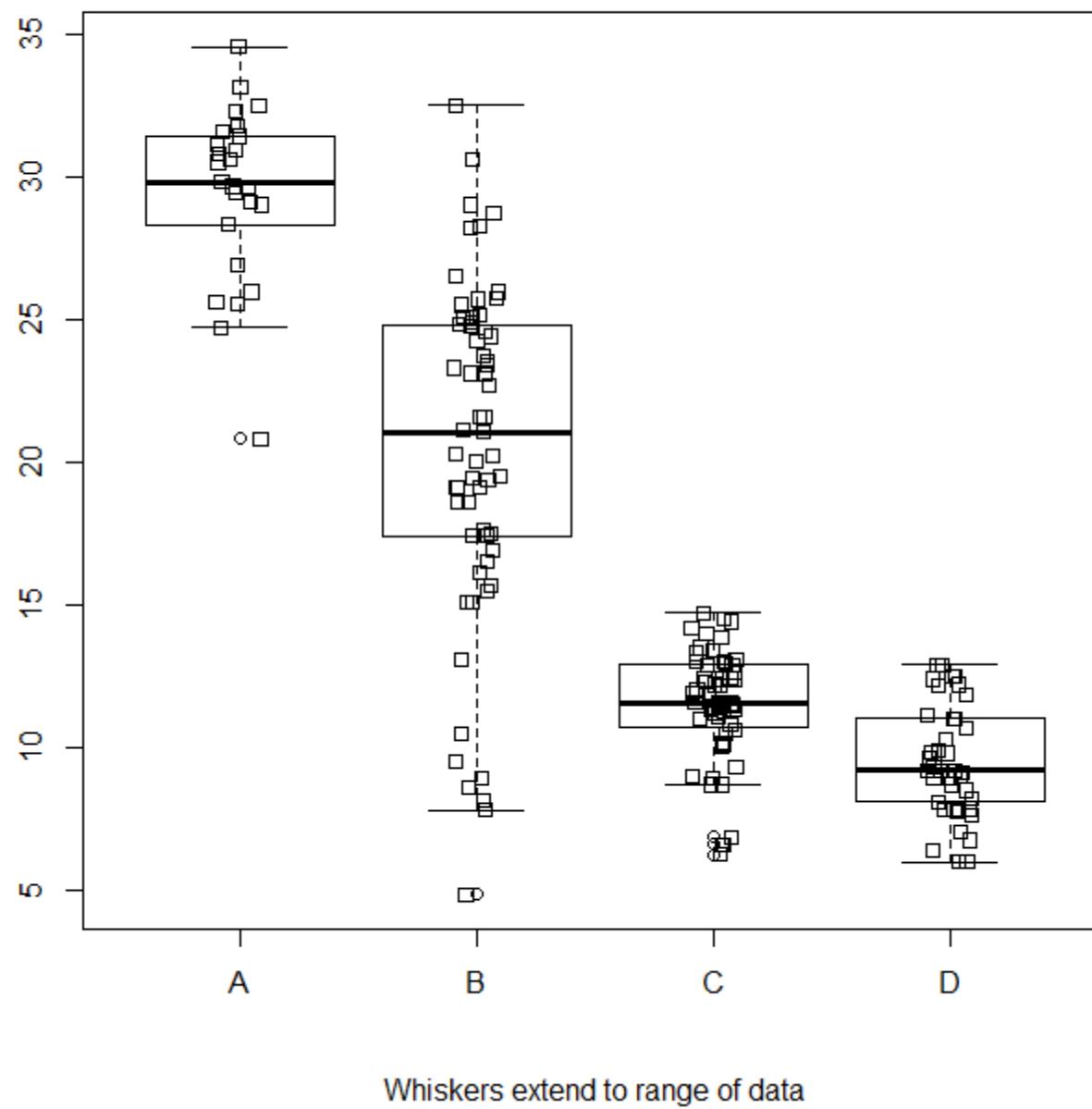
### Avg flagella length for each cell



- obvious outlier in the results for Variant A.
- Exclude that row and redo the graph

```
> avg <- avg[avg$Length > 4.5,]
> boxplot(Length ~ Variant, data=avg, main="Avg flagella length for each cell",
+   sub='Whiskers extend to range of data')
> stripchart(Length ~ Variant, data=avg, add=TRUE,
+   vertical=TRUE, method="jitter", jitter=.1,
+   main="Avg flagella length for each cell",
+   sub='Whiskers extend to range of data',
+   xlab='', ylab='Average length')
```

### Avg flagella length for each cell



# Check the assumption of equal standard deviations among groups:

- Find the sample size, mean, and standard deviation for each variant

```
> library(doBy)
Loading required package: survival
> report <- summaryBy(Length ~ Variant, data=avg, FUN=c(length,mean, sd))
> report$Length.se <- report$Length.sd/sqrt(report$Length.length)
> report
   Variant Length.length Length.mean Length.sd Length.se
1       A          25     29.41800   3.077976  0.6155951
2       B          61     20.38197   6.042482  0.7736605
3       C          56     11.46696   1.986071  0.2654000
4       D          41      9.44878   1.938443  0.3027339
```

The sample standard deviations vary among the groups by about a factor of 2 to 3 with variant B apparently having a much larger standard deviation. This amount of variation in the standard deviations is acceptable

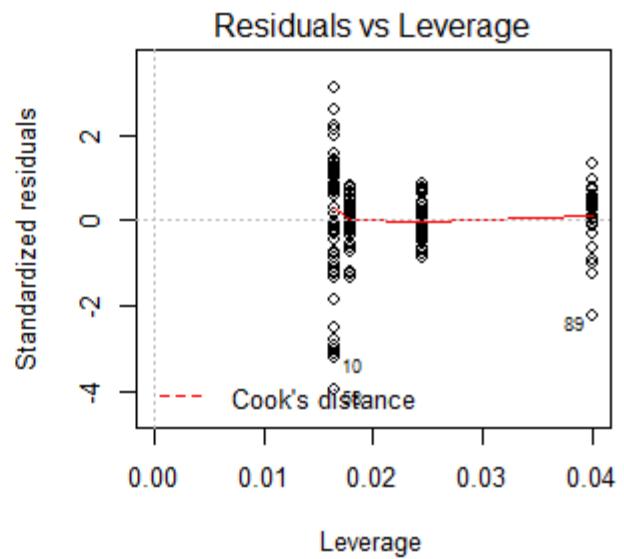
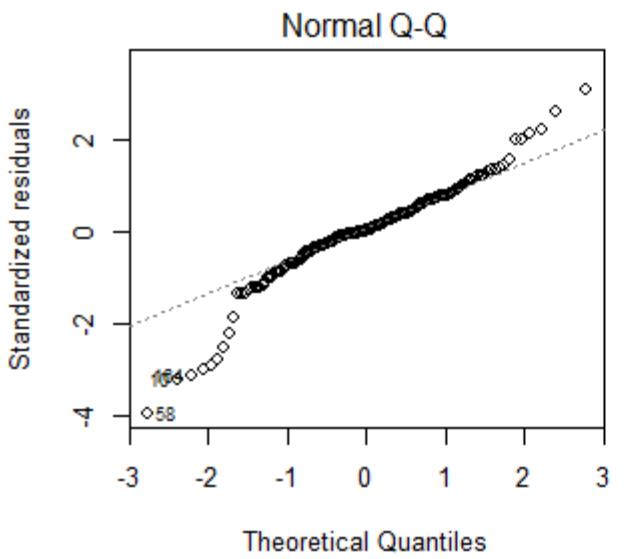
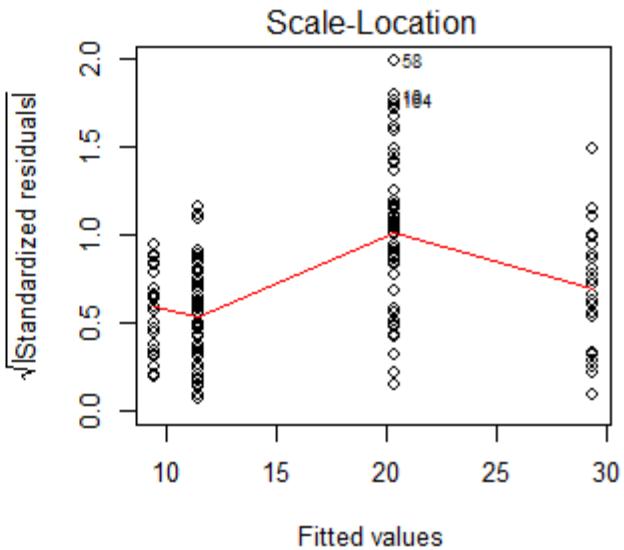
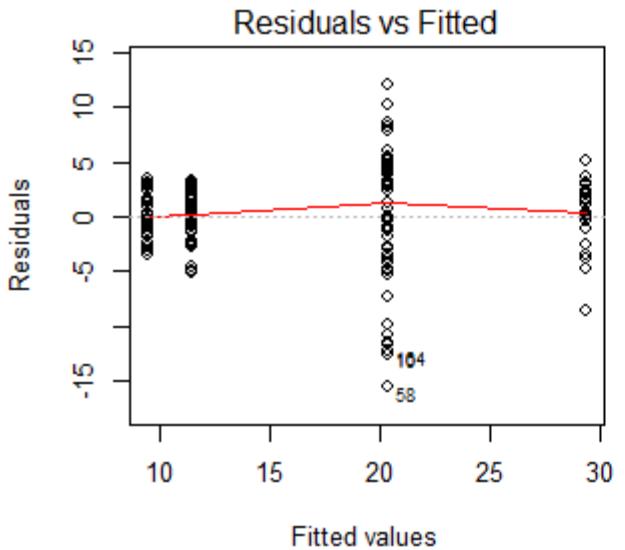
# a single-factor CRD ANOVA

```
> # fit the linear model and get the ANOVA table and test for effects
> model <- lm(Length ~ Variant, data=avg)
> anova(model)
Analysis of Variance Table

Response: Length
            Df  Sum Sq Mean Sq F value    Pr(>F)
Variant      3 8547.5 2849.17   183.1 < 2.2e-16 ***
Residuals 179 2785.3    15.56
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> # Check the assumptions of  
ANOVA model  
> layout(matrix(1:4, 2, 2))  
> plot(model)
```

→ doesn't show anything unusual about the fit.

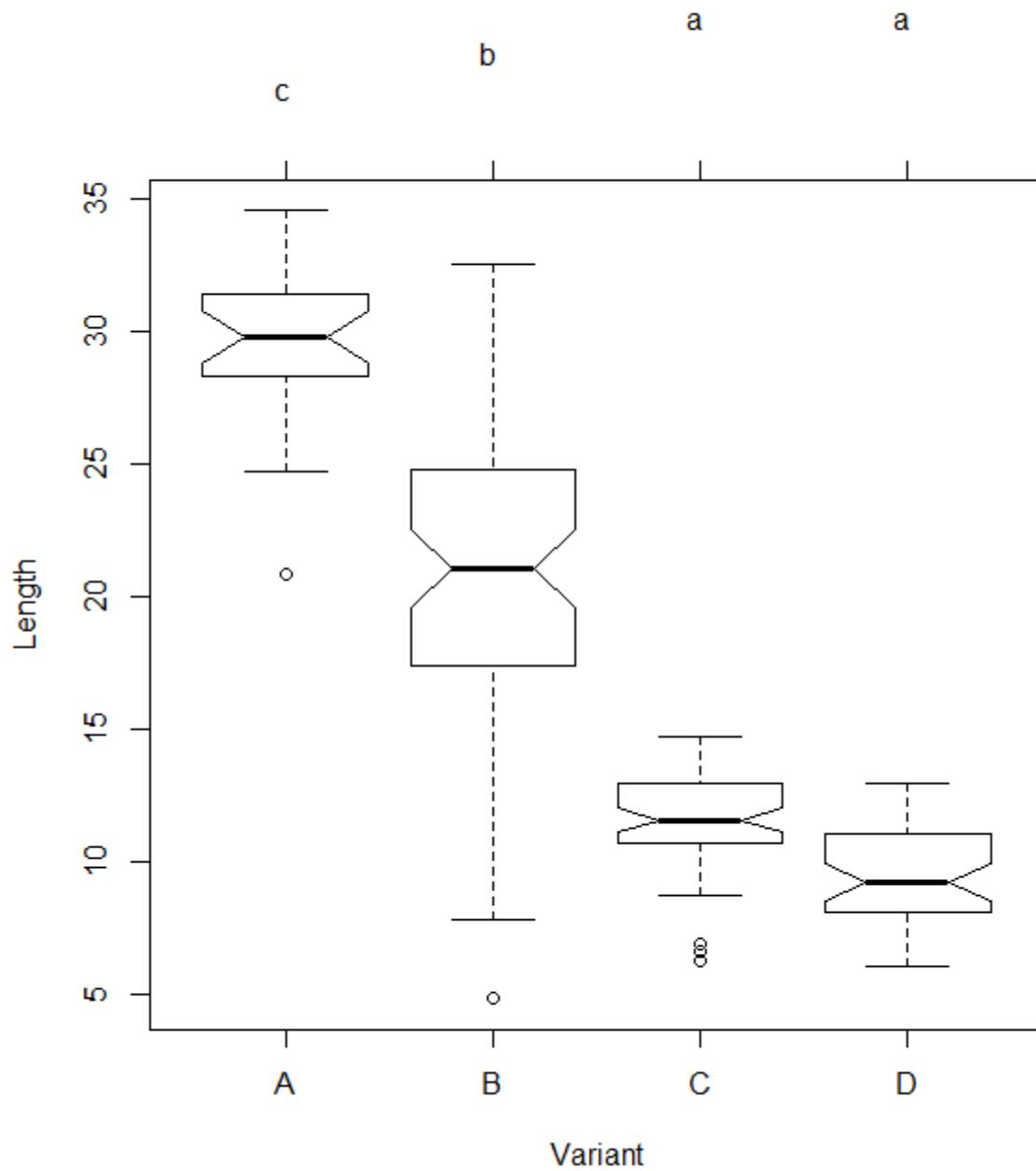


# Estimate the marginal means for each Variant:

```
> library(doBy)
> lsmeans <- popMeans(model, eff="Variant")
>
> lsmeans
   estimate      se   df t.stat     p.value Variant
1 29.41800 0.7889344 179 37.28827 2.586459e-86      A
2 20.38197 0.5050635 179 40.35526 8.281746e-92      B
3 11.46696 0.5271290 179 21.75362 3.852478e-52      C
4   9.44878 0.6160543 179 15.33758 1.914862e-34      D
```

# Multiple comparisons

```
> model.tukey <- glht(model, linfct = mcp(Variant = "Tukey") )
> model.tukey.cld <- cld(model.tukey) # joined line plot
> model.tukey.cld
  A   B   C   D
"c" "b" "a" "a"
> old.par <- par( mai=c(1,1,1.25,1)) # set top margin bigger
> plot(model.tukey.cld, # main="Multiple comparison results",
+       xlab="Variant",
+       ylab="Length",
+       notch=TRUE)
```



# 1. Analysis on individual measurements

# The model must include the following terms:

- Length → response variable
- Variant → treatment (Fixed)
- Cell(Variant) → the cell-to-cell variation
  - a **nested** effect ....(Variant))
  - a **random** effect

Nesting notation indicates that each cell number differs within each variant

→ it's a mixed model

**The within-cell variation is the lowest level in the experiment and is implicit (does not appear explicitly in the model)**

# Stack the data

```
> #import data with outlier deleted
> lengths <- read.csv("flagella2.csv", header=TRUE)
> stack.length <- reshape(lengths, idvar="Variant.Cell",
+                         varying=list(c("Flagellar.1","Flagellar.2")),
+                         times=c("Flagellar.1","Flagellar.2"), timevar="Length",
+                         v.name="Length",
+                         direction="long")
> stack.length[1:10,]
```

	Variant	Cell...	Comment	Length	Variant.Cell
1.Flagellar.1	B	15	NA	3.1	1
2.Flagellar.1	C	12	NA	5.8	2
3.Flagellar.1	D	3	NA	6.0	3
4.Flagellar.1	C	19	NA	6.6	4
5.Flagellar.1	C	56	NA	6.6	5
6.Flagellar.1	D	41	NA	6.7	6
7.Flagellar.1	D	38	NA	6.8	7
8.Flagellar.1	D	36	NA	6.9	8
9.Flagellar.1	B	49	NA	7.1	9
10.Flagellar.1	C	5	NA	7.4	10

# Fit the model

```
> library(lme4)
Loading required package: Matrix
Loading required package: Rcpp
> model2 <- lmer( Length ~ Variant + (1|Variant.Cell), data=stack.length)
```

```
> summary(model2)
Linear mixed model fit by REML ['lmerMod']
Formula: Length ~ Variant + (1 | Variant.Cell)
Data: stack.length
```

REML criterion at convergence: 1610.6

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.6916	-0.3635	0.0278	0.3815	3.2780

Random effects:

Groups	Name	Variance	Std.Dev.
Variant.Cell	(Intercept)	9.525	3.086
Residual		9.450	3.074

Number of obs: 285, groups: Variant.Cell, 183

....

```
> anova(model2) # note no p-value given
```

Analysis of Variance Table

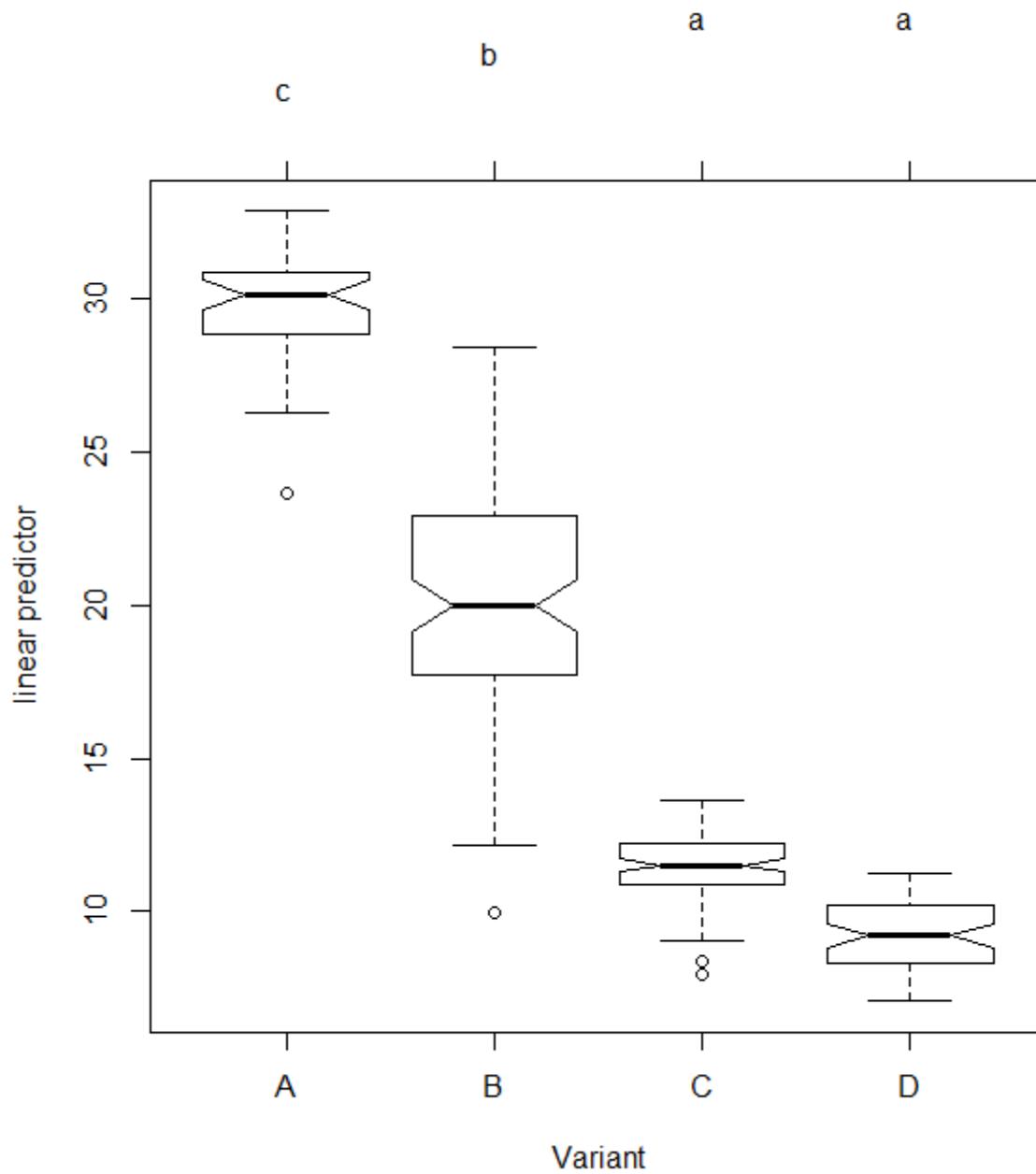
Df	Sum Sq	Mean Sq	F value	
Variant	3	5065.1	1688.4	178.67

variance component analysis:  
cell-to-cell variation (9.52) is roughly the same  
order of magnitude as the flagella-to-flagella  
variation (9.44)

```
> # Get the marginal means
> # There is no easy way to do this using lmer()
> pred.Variant <- expand.grid( Variant=unique(avg$Variant), Length=0)
> mm <- model.matrix(terms(model2),pred.Variant)
> means <- mm %*% fixef(model2)
> means.vcv <- mm %*% tcrossprod(vcov(model2),mm)
> means.se <- sqrt(diag(mm %*% tcrossprod(vcov(model2),mm)) )
> cbind(pred.Variant, means, means.se)
   Variant Length      means   means.se
1       A      0 29.428525 0.7912881
2       B      0 20.204547 0.5159561
3       C      0 11.466180 0.5284951
4       D      0  9.364474 0.6333776
>
```

# Multiple comparison

```
> library(multcomp)
>
> model2.tukey <- glht(model2, linfct = mcp(Variant = "Tukey"))
> model2.tukey.cld <- cld(model2.tukey) # joined line plot
>
> # create the display
> model2.tukey.cld
   A      B      C      D
"c" "b" "a" "a"
> old.par <- par( mai=c(1,1,1.25,1)) # set top margin bigger
> plot(model2.tukey.cld, # main="Multiple comparison results",
+       xlab="Variant",
+       ylab="Length",
+       notch=TRUE)
```



# THE END

Now that the classes are over, you might enjoy listening to some  
classical songs from

<http://www.glicko.net/music.html>

# ANOVA Man

Words: Mark Glickman

Music: Lennon/McCartney ("Nowhere Man")

He's a real ANOVA man  
Designing all his sampling plans  
Calculating mean-squared errors and p-values.

Wants to test for equal mu's  
Knows which tables he must use  
All his samples he will choose at random.

ANOVA man, please listen;  
Where's the data that you're missing?  
ANOVA man, what kinds of bias can you withstand?

Writes down two hypotheses;  
Hopes to reject the first of these;  
Needs to list out all degrees of freedom.

ANOVA man, try harder;  
Don't give up, you're smarter;  
ANOVA man, how come your students don't understand?

At 0.05 he rejects  
Ignores the size of his effects  
Now he's stuck -- he's got selection bias!

ANOVA man, please listen;  
Where's the data that you're missing?  
ANOVA man, what kinds of bias can you withstand?

He's a real ANOVA man  
Designing all his sampling plans  
Calculating mean-squared errors and p-values.