

Workshop in experimental design and applied statistics: Day 2

Presentation and workshop materials developed by Maegen
Simmonds

AIP App developed by Ian King Kyle, Jason Moore, and Maegen
Simmonds

Funded by USAID

Workshop Goals

- To understand experimental design and principles of statistical analysis to produce trustworthy results and interpretation.
- To learn how to use the free AIP computer app to analyze and visualize your data

What are the basic experimental designs that you use in your research?

Examples:

1. Complete Randomized Design
2. Randomized Complete Block Design
3. Split-plot Design

Note: “Factorial” is a treatment structure of one of the experimental designs, which we will also cover.

What are we doing today?

- Instruction 9:00 – 11:00 AM
 - Section 4: Testing assumptions of ANOVA and transforming data
- Tea break 11:00 – 11:30 AM
- Instruction 11:30 – 1:00 PM
 - Section 5: Presenting results, and RCBD with factorial treatment structure
- Lunch 1:00 – 2:00 PM
- Instruction 2:00 – 2:30 PM
 - Section 6: Split plot experimental design
- Tea break 2:30 – 3:00 PM
- Instruction 3:00 – 4:00 PM
 - Section 7: Multi-location studies, missing data

Have data. Want analysis.



Load and run a warm-up dataset.

- What are the results?
- Is everything OK with it?

valid

Have data. Want ⁿ analysis.



Do you recall the
assumptions of ANOVA?
Which ones have you
tested before?



Section 4 Learning Objectives

Have basic understanding of:

1. The assumptions of ANOVA and transformations
 - a) Know the assumptions, and how to check your data
 - b) What to do when assumption tests fail, and what transformation to choose for your data
 - c) How to do all of this with the app!

Follow these four key assumptions of ANOVA to ensure that your analysis and conclusions are valid and correct:

ANOVA Assumptions	When might I have a problem?
1. Errors are statistically independent	Did I forget to <i>randomly</i> assign treatments to experimental units?(e.g. fertilizer rate treatments assigned to plots)
2. Additive effects: the variables are <i>added</i> , not multiplied	Does the <i>effect</i> of the treatments <i>depend</i> on what block they're in?
3. Normal distribution of errors	Do the data represent a rare event, or are there outliers?
4. Equal variance	Is there a very strong treatment effect, or is this a multi-site study?

Four key assumptions are used in ANOVA to ensure that your analysis and conclusions are valid and

Ensure this by *randomizing* treatment assignments to experimental units!

ANOVA Assumptions

1. Errors are statistically independent

treatments to experimental units?(e.g. fertilizer rate treatments assigned to plots)

2. Additive effects: the variables are *added*, not multiplied

Does the *effect* of the treatments *depend* on what block they're in?

3. Normal distribution of errors

Do the data represent a rare event, or are there outliers?

4. Equal variance

Is there a very strong treatment effect, or is this a multi-site study?

Four key assumptions are used in ANOVA to ensure that your analysis and conclusions are valid and correct.

ANOVA Assumptions	When might I have a problem?
1. Errors are statistically independent	Did I forget to <i>randomly</i> assign treatments to experimental units?(e.g. fertilizer rate
2. Additive effects: the variables are <i>added</i> , not multiplied	Do I have a very strong treatment effect, or is this a multi-site study?
3. Normal distribution of errors	
4. Equal variance	

It's standard practice to perform statistical tests for #3 and #4, and include it in materials and methods section of research reports.

Four key assumptions are used in

ANOVA. The app will give you the test results, but YOU must conclude, interpret and address any problems.

ANOVA

1. Error

independ

(we will learn how)

treatments to experimental units?(e.g. fertilizer rate

2. Additive effects: the variables are *added*, not multiplied

Do

de

3. Normal distribution of errors

4. Equal variance

It's standard practice to perform statistical tests for #3 and #4, and include it in materials and methods section of research reports.

Is there a very strong treatment effect, or is this a multi-site study?

Four key assumptions are used in ANOVA to ensure that your analysis and conclusions are valid and correct.

ANOVA Assumptions

1. Errors are statistically independent

2. Additive effects: the variables are *added*, not multiplied

3. Normal distribution of errors

4. Equal variance

The app provides test results for this too!

*Only for basic RCBD with 1 replication per block.

Does the *effect* of the treatments *depend* on what block they're in?

Do the data represent a rare event, or are there outliers?

Is there a very strong treatment effect, or is this a multi-site study?

**Inspect *all* assumption
tests and
corresponding plots
from your app output
before looking at
ANOVA table**

Let's use the app to check the ANOVA assumptions on the wheat variety dataset (RCBD) from yesterday (wheat_yield_data.csv)

- ☐ Normality
- ☐ Equal variance
- ☐ Additivity (only for RCBD with 1 replication per block)

Let's use the app to check the ANOVA assumptions on the wheat variety dataset (RCBD) from yesterday (wheat_yield_data.csv)

- ☐ Normality
- ☐ Equal variance
- ☐ Additivity (only for RCBD with 1 replication per block)

Shapiro-Wilk Normality Test

Good (Wheat example)

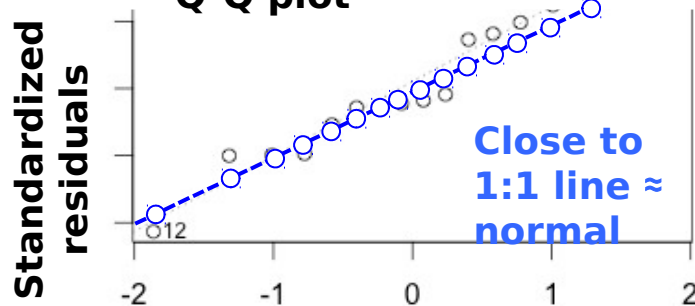
p-value > 0.05

CONCLUSION: Errors are normally distributed.

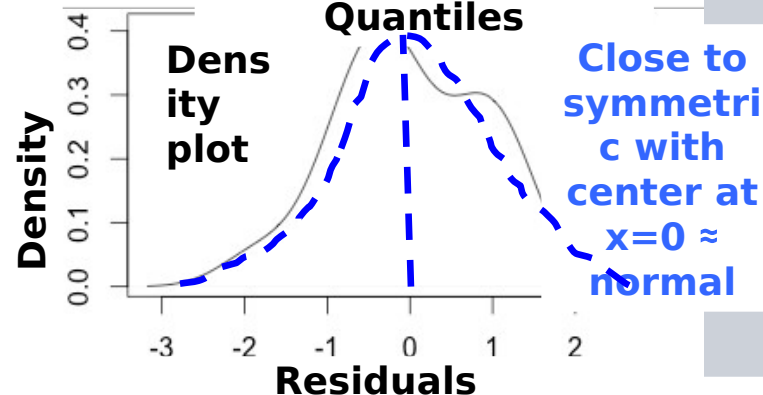
W = 0.9472 , p-value =

0.71

Q-Q plot



Density plot



Bad

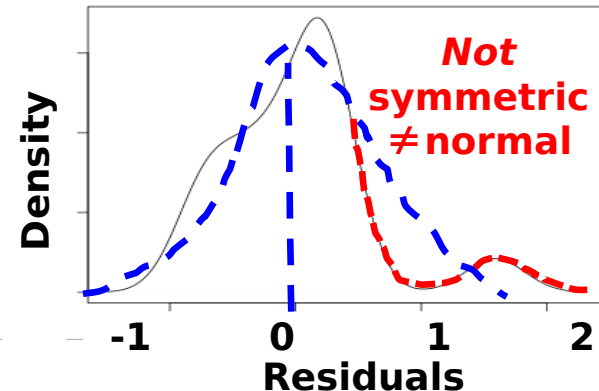
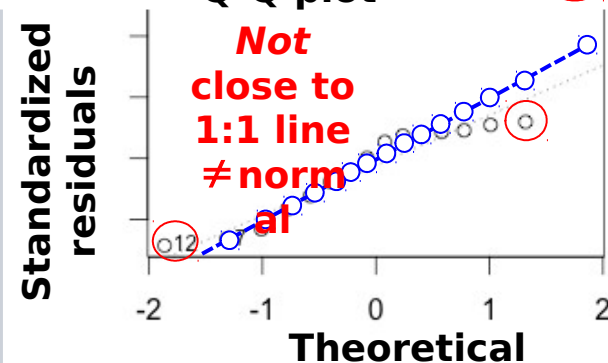
p-value ≤ 0.05

CONCLUSION: Errors are *not* normally distributed.

W = 0.8784 , p-value =

0.01

Q-Q plot



Wheat variety dataset:

- ☒ Normality
- ☐ Equal variance
- ☐ Additivity

Wheat variety dataset:

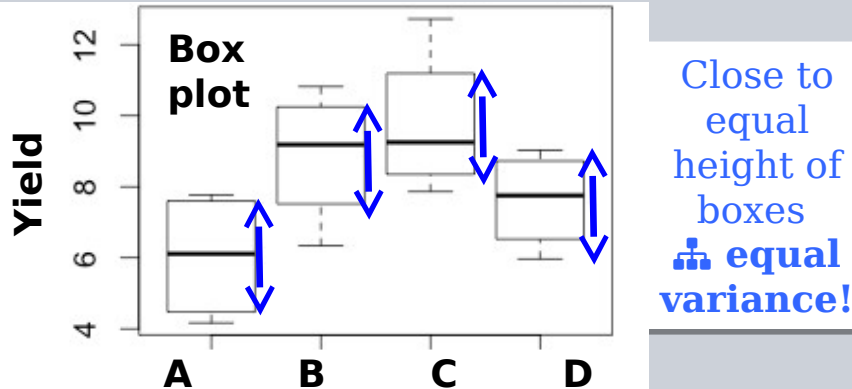
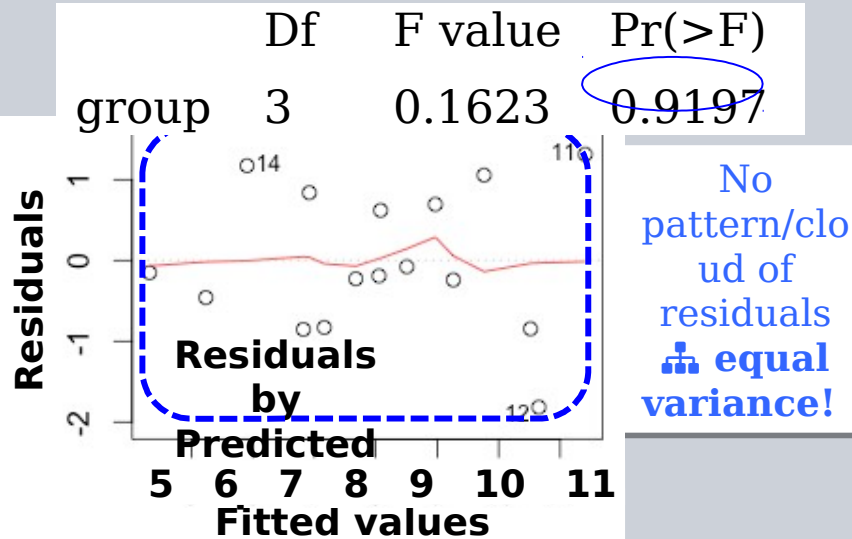
- ☒ Normality
- ☐ Equal variance
- ☐ Additivity

Levene's Test for Homogeneity of Variance

Good (Wheat example)

p-value > 0.05

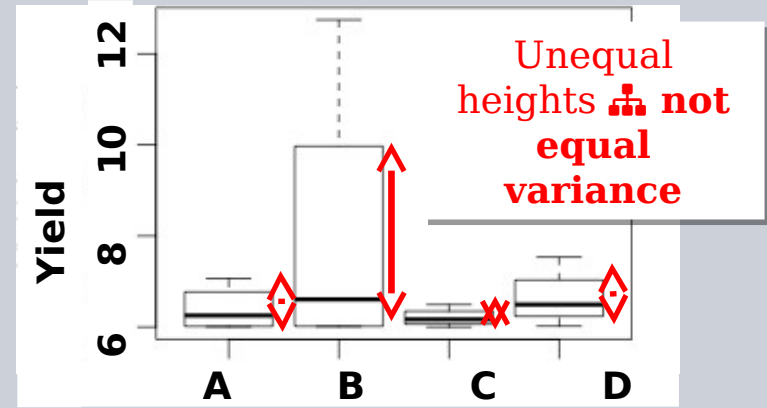
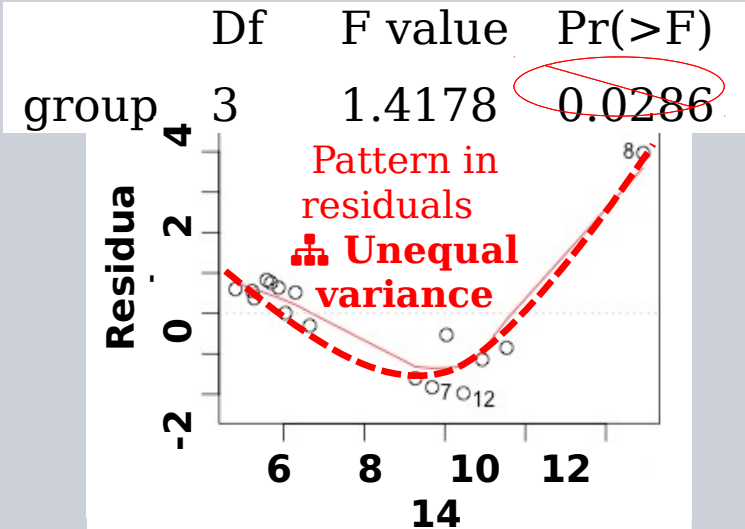
CONCLUSION: Equal variance across treatments.



Bad

p-value ≤ 0.05

CONCLUSION: Treatments variances are *not* equal.



Wheat variety dataset:

- ☒ Normally distributed errors
- ☒ Equal variance
- ☐ Additivity

Wheat variety dataset:

- ☒ Normally distributed errors
- ☒ Equal variance
- ☐ Additivity

Tukey's Test for Nonadditivity

Good (Wheat example)

p-value > 0.05
CONCLUSION: Additive
model.

Analysis of Variance Table

Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	31.5311	10.5104	7.2485	0.01141 *
Block	3	28.2354	9.4118	6.4891	0.01550 *
sq_preds	1	0.0015	0.0015	0.0011	0.97486
Residuals	8	11.6033	1.4504		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

Only look at p-value
for 'sq_preds'.

Bad

p-value ≤ 0.05
CONCLUSION: Nonadditivity
in model.

Analysis of Variance Table

Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	4.9189e+19	1.6396e+19	21.895	0.0003267 ***
Block	3	8.4309e+19	2.8103e+19	37.527	4.640e-05 ***
sq_preds	1	1.1685e+20	1.1685e+20	156.030	1.578e-06 ***
Residuals	8	5.9910e+18	7.4888e+17		

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

Wheat variety dataset:

- ✓ ☒ Normally distributed errors
- ✓ ☒ Equal variance
- ✓ ☒ Additivity

All our assumption tests were GOOD.
Thus, the ANOVA of the field study is
valid.



No



F



Ad

What if we
failed 1 or more
of the
assumption
tests?

All our assumption tests were GOOD.
Thus, the ANOVA of the field study is
valid.

Transformations

- † Perform ANOVA on the dataset that *best* meets assumptions
- † May not be the original scale, such as lb/ac
- Thus, when data do *not* meet assumptions, one approach is to transform your data to see if the assumption tests improve.
 - Note: the transformation is applied to the dependent variable only (Y_{ijk})

WARNING: We do not transform data to get the results we want. We transform data so that the analysis and results are *valid* and *correct*.

Transformations

If you decide to transform your data, you can choose from 2 transformations using the app:

1. Log transformation: $\log_{10}(Y_{ij}) = \mu + \tau_i + \beta_j + \varepsilon_{ij}$
2. Power transformation: $(Y_{ij})^a = \mu + \tau_i + \beta_j + \varepsilon_{ij}$

Try both transformations. Use the one that most improves the assumption tests.

Before choosing log transformation

CSV file and reload it to the app:

1. Experimental Design

2. Variable types

3. Dependent Variable

4. Independent variables

5. Transformations

transformations_panel

Select a transformation for the dependent variable:

Logarithmic



Run analysis

Not yet!

*Before choosing **log** transformation*

CSV file and reload it to the app:

1. Are all observations (Y_{ij}) greater than 0 and some (or all) between 0 and 1?

✓ If so, multiply your entire Y_{ij} dataset by a constant 'X' so that all values are ≥ 1 .

1. Experimental Design

2. Variable types

3. Dependent Variable

4. Independent variables

5. Transformations

transformations_panel

Select a transformation for the dependent variable:

Logarithmic

Run analysis



Not yet!

Before choosing log transformation

Example dataset

	A	B	C
1	Block	Independent	Dependent
2	1	A	0.60230465
3	2	A	0.16753883
4	3	A	12.0412073
5	4	A	26.4773239
6	1	B	29.8006572
7	2	B	0.54721707
8	3	B	0.2823309
9	4	B	168.656146
10	1	C	12.2879646
11	2	C	0.0048015
12	3	C	3.26729745
13	4	C	5.11869392
14	1	D	0.45601859
15	2	D	11.427267
16	3	D	13.288801
17	4	D	38.3352375
18			

CSV file and reload it to the app if:

1. Are all observations (Y_{ij}) greater than 0 and some (or all) between 0 and 1?

✓ If so, multiply your entire Y_{ij} dataset by a value 'X' so that all are ≥ 1 .

What X value would you choose for this dataset?

Before choosing **log** transformation

CSV file and reload it to the app if:

	A	B	C	D
1	Block	Independent	Dependent	
2	1	A	0.60230465	=C2*1000
3	2	A	0.16753883	
4	3	A	12.0412073	
5	4	A	26.4773239	
6	1	B	29.8006572	
7	2	B	0.54721707	
8	3	B	0.2823309	
9	4	B	168.656146	
10	1	C	12.2879646	
11	2	C	0.0048015	
12	3	C	3.26729745	
13	4	C	5.11869392	
14	1	D	0.45601859	
15	2	D	11.427267	
16	3	D	13.288801	
17	4	D	38.3352375	

At least
 $X = 1000$

1. Are all observations (Y_{ij}) greater than 0 and some (or all) between 0 and 1?

✓ If so, multiply your entire Y_{ij} dataset by a constant 'X' so that all values are ≥ 1 .

Before choosing log transformation

CSV file and reload it to the app:

	A	B	C	D
1	Block	Independent	Dependent	
2	1	A	0.60230465	602.304648
3	2	A	0.16753883	167.538834
4	3	A	12.0412073	12041.2073
5	4	A	26.4773239	26477.3239
6	1	B	29.8006572	29800.6572
7	2	B	0.54721707	547.217069
8	3	B	0.2823309	282.330903
9	4	B	168.656146	168656.146
10	1	C	12.2879646	12287.9646
11	2	C	0.0048015	4.80149538
12	3	C	3.26729745	3267.29745
13	4	C	5.11869392	5118.69392
14	1	D	0.45601859	456.01859
15	2	D	11.427267	11427.267
16	3	D	13.288801	13288.801
17	4	D	38.3352375	38335.2375

1. Are all observations (Y_{ij}) greater than 0 and some (or all) between 0 and 1?

✓ If so, multiply your entire Y_{ij} dataset by a constant 'X' so that all values are ≥ 1 .

All values were multiplied by 1000, and are greater than 1.

Before choosing log transformation

CSV file and reload it to the app:

	A	B	C	D
1	Block	Independent	Dependent	
2	1	A	0.60230465	602.304648
3	2	A	0.16753883	167.538834
4	3	A	12.0412073	12041.2073
5	4	A	26.4773239	26477.3239
6	1	B	29.8006572	29800.6572
7	2	B	0.54721707	547.217069
8	3	B	0.2823309	282.330903
9	4	B	168.656146	168656.146
10	1	C	12.2879646	12287.9646
11	2	C	0.0048015	4.80149538
12	3	C	3.26729745	3267.29745
13	4	C	5.11869392	5118.69392
14	1	D	0.45601859	456.01859
15	2	D	11.427267	11427.267
16	3	D	13.288801	13288.801
17	4	D	38.3352375	38335.2375

1. Are all observations (Y_{ij}) greater than 0 and some (or all) between 0 and 1?

- ✓ If so, multiply your entire Y_{ij} dataset by a constant 'X' so that all values are ≥ 1 .

Copy values

Before choosing **log** transformation

CSV file and reload it to the app:

	A	B	C
1	Block	Independent	Dependent
2	1	A	602.304648
3	2	A	167.538834
4	3	A	12041.2073
5	4	A	26477.3239
6	1	B	29800.6572
7	2	B	547.217069
8	3	B	282.330903
9	4	B	168656.146
10	1	C	12287.9646
11	2	C	4.80149538
12	3	C	3267.29745
13	4	C	5118.69392
14	1	D	456.01859
15	2	D	11427.267
16	3	D	13288.801
17	4	D	38335.2375

1. Are all observations (Y_{ij}) greater than 0 and some (or all) between 0 and 1?

Paste values

✓ If so, multiply your entire Y_{ij} dataset by a constant 'X' so that all values are ≥ 1 .

Before choosing log transformation

	A	B	C
1	Block	Independent	Dependent
2	1	A	602.304648
3	2	A	167.538834
4	3	A	12041.2073
5	4	A	26477.3239
6	1	B	29800.6572
7	2	B	547.217069
8	3	B	282.330903
9	4	B	168656.146
10	1	C	12287.9646
11	2	C	4.80149538
12	3	C	3267.29745
13	4	C	5118.69392
14	1	D	456.01859
15	2	D	11427.267
16	3	D	13288.801
17	4	D	38335.2375

CSV file and reload it to the app:

1. Are all observations (Y_{ij}) greater than 0 and some (or all) between 0 and 1?

✓ If so, multiply your entire Y_{ij} dataset by a constant 'X' so that all values are > 1 .

'Save as' new file name (e.g. yield_multiplied_1000.csv) and load the new file to the app.

Before choosing log transformation

CSV file and reload it to the app:

Note: you only do one or the other, and if all your data are greater than 1, no pre-transformation is needed

1. Are all observations (Y_{ij}) greater than 0 *and* some (or all) between 0 and 1?

✓ If so, multiply your entire Y_{ij} dataset by a constant 'X' so that all values are ≥ 1 .

2. Are any observations (Y_{ij}) ≤ 0 ?

✓ If so, add a constant 'X' to entire Y_{ij} dataset so that all values are ≥ 1 .

Before choosing **log** transformation

CSV file and reload it to the app if:

1. Are all observations (Y_{ij}) greater than 0 *and* some (or all) between 0 and 1?
 - ✓ If so, multiply your entire Y_{ij} dataset by a constant 'X' so that all values are ≥ 1 .
2. Are any observations (Y_{ij}) ≤ 0 ?
 - ✓ If so, add a constant 'X' so that all values are ≥ 1 .

1. Experimental Design

2. Variable types

3. Dependent Variable

4. Independent variables

5. Transformations

transformations_panel

Select a transformation for the dependent variable:

Logarithmi

C

Run analysis

Now your data are ready for the **log** transformation.

Section 4 Learning Objectives

You should now have a basic understanding of:

- ✓ The assumptions of ANOVA and transformations
 - a) What are the four assumptions of ANOVA?
 - b) How do you ensure your data meets the assumption of independent errors?
 - c) How do you use the app to check if your data meets the assumptions?
 - d) What can you do if your data fail an assumption test?
 - e) What types of transformations can you apply to your data using the app?
 - f) What range of values must your dataset be in before applying the log transformation?
 - What can you do if it is outside of that range?













Section 5 Learning Objectives

Have basic understanding of:

1. Factorial treatment structures
 - Linear model and ANOVA of factorial experiments
 - Interpreting ANOVA results and interactions
 - Post-hoc tests
2. Using the app to check assumptions *and* to transform data
3. Back-transformations and presenting results

Recall the CRD milking experiment. What was the independent variable(s)?

Observed milk production
(kg)













Replication	(1) Hand-milk 2x	(2) Hand-milk 3x	(3) Machine-milk 2x	(4) Machine-milk 3x
(1)	8.6 	11.2 	9.6 	11.5 
(2)	9.3 	10.8 	9.3 	12.0 
(3)	9.1 	11.7 	9.0 	11.8 
Treatment means:	9.0	11.2	9.3	11.8

Independent variable = milking practice, which consists of 2 factors:

(1) Factor A = 'milking method' with 2 levels (hand, machine)

(2) Factor B = 'milking frequency' with 2 levels (2x per day, 3x per day)

☎️ ♦️ We analyzed it as a single combined treatment ('milking practice') with $2 \times 2 = 4$ levels

n			milk 2x	milk 3x
(1)	8.6 	11.2 	9.6 	11.5 
(2)	9.3 	10.8 	9.3 	12.0 
(3)	9.1 	11.7 	9.0 	11.8 
Treatment means:	9.0	11.2	9.3	11.8

We used the following linear model to predict milk production based on milking practice:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

where Y_{ij} = observation of i^{th} treatment and j^{th} replication, μ = grand mean, τ_i = effect of i^{th} treatment (Hand-milk 2x, Hand-milk 3x, Machine-milk 2x, Machine-milk 3x), and ε_{ij} = random error.

We can get *more* information about the treatment effect by partitioning τ_i into: Factor A (milking method), Factor B (milking frequency), and AxB interaction. This linear model is given by:

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

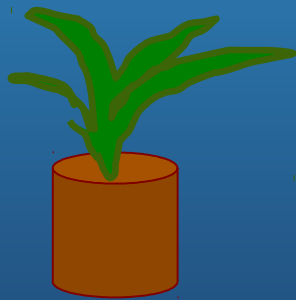
where α_i = effect of i^{th} level of Factor A (hand-milk, machine-

This is a *factorial* treatment structure. Each level of Factor A must be combined with Factor B

**Let's work on an
example of an
experiment with a
factorial treatment
structure...**

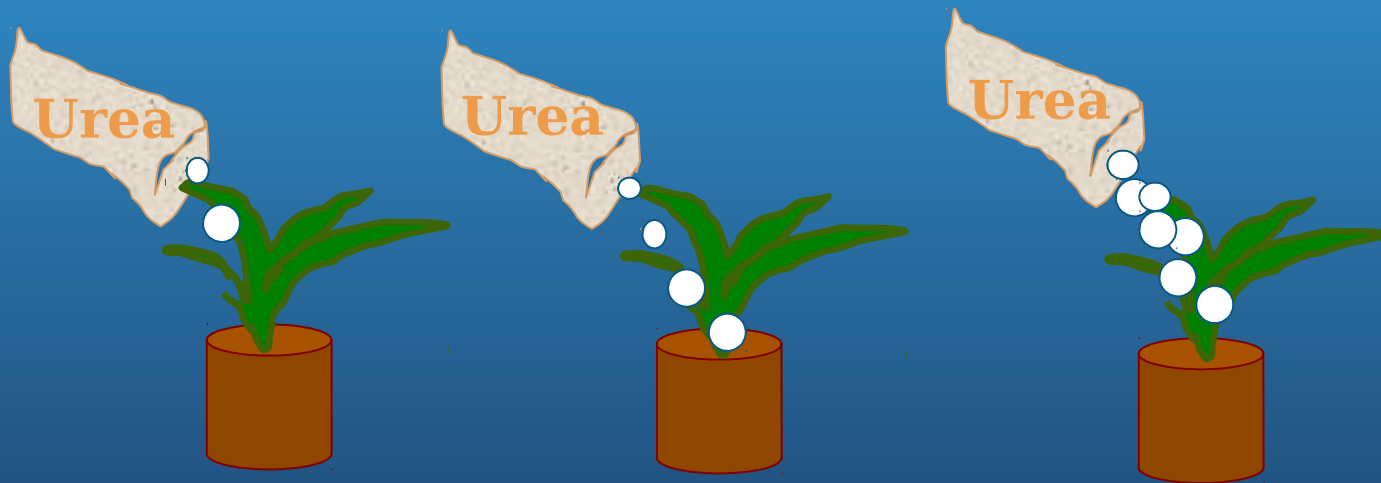
Research questions:

- Does yield differ among clones (3 types)?



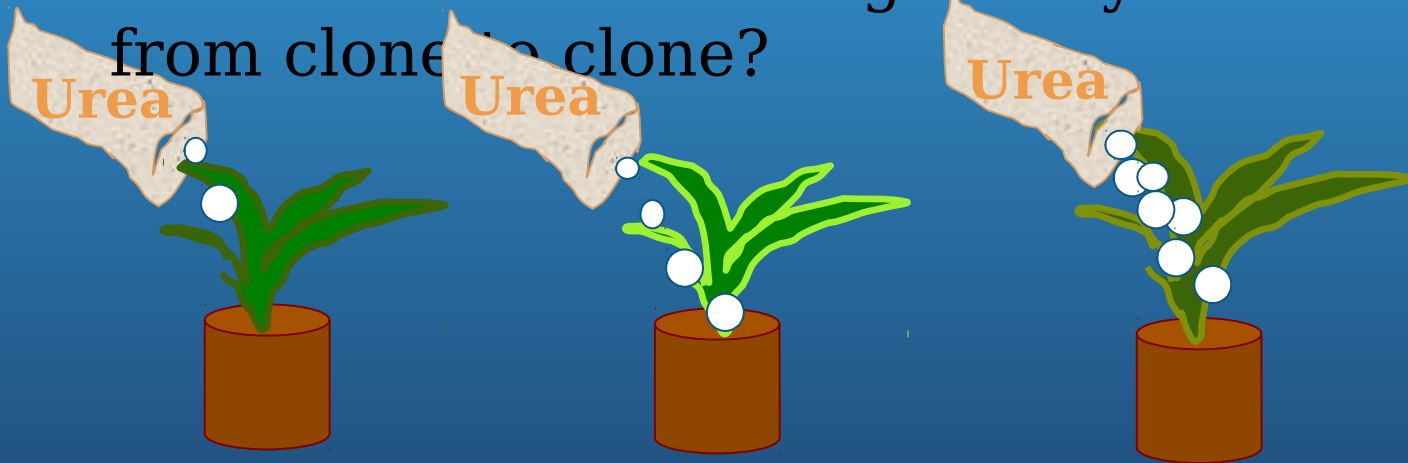
Research questions:

- Does yield differ among clones (3 types)?
- Does nitrogen fertilization affect yield (3 levels)?



Research questions:

- Does yield differ among clones (3 types)?
- Does nitrogen fertilization affect yield (3 levels)?
- Does the effect of nitrogen on yield change from clone to clone?



Research questions:

- Does yield differ among clones (3 types)?
- Does nitrogen fertilization affect yield (3 levels)?
- Does the effect of nitrogen on yield change from clone to clone?




In groups, outline a field experiment to answer the questions above.

Include:

- Dependent variable?
- Independent variable(s)?
- Experimental design and treatment structure?
- Field layout of experimental plots? (Assume 4 reps)

- Dependent variable: Yield
- Independent variable(s): Clone and Nitrogen
- Experimental design and treatment structure: RCBD experiment with 2x2 factorial treatment structure (Clone, Nitrogen), 3 levels each, 4 blocks.

3 levels for each factor
 (3 x 3) 9 unique combinations

	Clone	Nitrogen
1	c1	n1
2	c1	n2
3	c1	n3
4	c2	n1
5	c2	n2
6	c2	n3
7	c3	n1
8	c3	n2
9	c3	n3

What does the layout of your field experiment look like?

What is the experimental unit?

How many are there?

9 x 4 replications = 36

- Dependent variable: Yield
- Independent variable(s): Clone and Nitrogen
- Experimental design and treatment structure: RCBD experiment with 2x2 factorial treatment structure (Clone, Nitrogen), 3 levels each, 4 blocks.

← Plot = experimental unit
(n = 36)

- Dependent variable **Yield**
- Independent variable(s) **Clone and Nitrogen**
- Experimental design and treatment structure **RCBD experiment with 2x2 factorial treatment structure (Clone, Nitrogen), 3 levels each, 4 blocks.**

Block 1	Block 2	Block 3	Block 4
c3 n2	Randomly assign the 9 treatment combinations to each plot within a block		
c1 n1			
c3 n3			
c2 n1			
c1 n2			
c1 n3			
c3 n1			
c2 n3			
c2 n2			

- Dependent variable: Yield
- Independent variable(s): Clone and Nitrogen
- Experimental design and treatment structure: RCBD experiment with 2x2 factorial treatment structure (Clone, Nitrogen), 3 levels each, 4 blocks.

Repeat randomization of treatments for each block

Block 1	Block 2	Block 3	Block 4
c3 n2	c1 n2	c3 n1	c2 n3
c1 n1	c3 n1	c2 n3	c1 n3
c3 n3	c3 n2	c1 n3	c2 n1
c2 n1	c1 n1	c3 n2	c3 n2
c1 n2	c2 n2	c2 n1	c3 n1
c1 n3	c2 n3	c1 n1	c1 n1
c3 n1	c1 n3	c1 n2	c2 n2
c2 n3	c1 n1	c2 n2	c3 n3
c2 n2	c2 n1	c1 n1	c1 n2

- Dependent variable **Yield**
- Independent variable(s) **Clone and Nitrogen**
- Experimental design and treatment structure **RCBD experiment with 2x2 factorial treatment structure (Clone, Nitrogen), 3 levels each, 4 blocks.**

Layout of field experiment

Block 1	Block 2	Block 3	Block 4
c3 n2	c1 n2	c3 n1	c2 n3
c1 n1	c3 n1	c2 n3	c1 n3
c2 n1	c1 n1	c3 n2	c2 n2
c1 n2	c2 n2	c2 n1	c3 n1
c1 n3	c2 n3	c1 n1	c1 n1
c3 n1	c1 n3	c1 n2	c2 n2
c2 n3	c1 n1	c2 n2	c3 n3
c2 n2	c2 n1	c1 n1	c1 n2

Your yield data are ready

Let's use the app to analyze the clone x irrigation dataset!

- Load 'rcbd_two_var.csv' to the app
- This time, we will do all the analysis steps in the correct order:
 - 1) Check assumptions, and transform data if necessary
 - 2) Check ANOVA results, and perform Fisher's Protected LSD test

Does everything look okay?

```
yield ~ clone + nitrogen + clone:nitrogen + block
```

Model Fit Summary

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	497.7	248.86	209.567	6.37e-16	***
nitrogen	2	54.1	27.03	22.760	2.87e-06	***
block	3	20.8	6.92	5.825	0.003875	**
clone:nitrogen	4	43.3	10.82	9.111	0.000126	***
Residuals	24	28.5	1.19			

Does everything look okay?

```
yield ~ clone + nitrogen + clone:nitrogen + block
```

Model Fit Summary

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	497.7	248.86	209.567	6.37e-16	***
nitrogen	2	54.1	27.03	22.760	2.87e-06	***
block	3	20.8	6.92	5.825	0.003875	**
clone:nitrogen	4	43.3	10.82	9.111	0.000126	***
Residuals	24	28.5	1.19			

The model
looks good

Does everything look okay?

```
yield ~ clone + nitrogen + clone:nitrogen
```

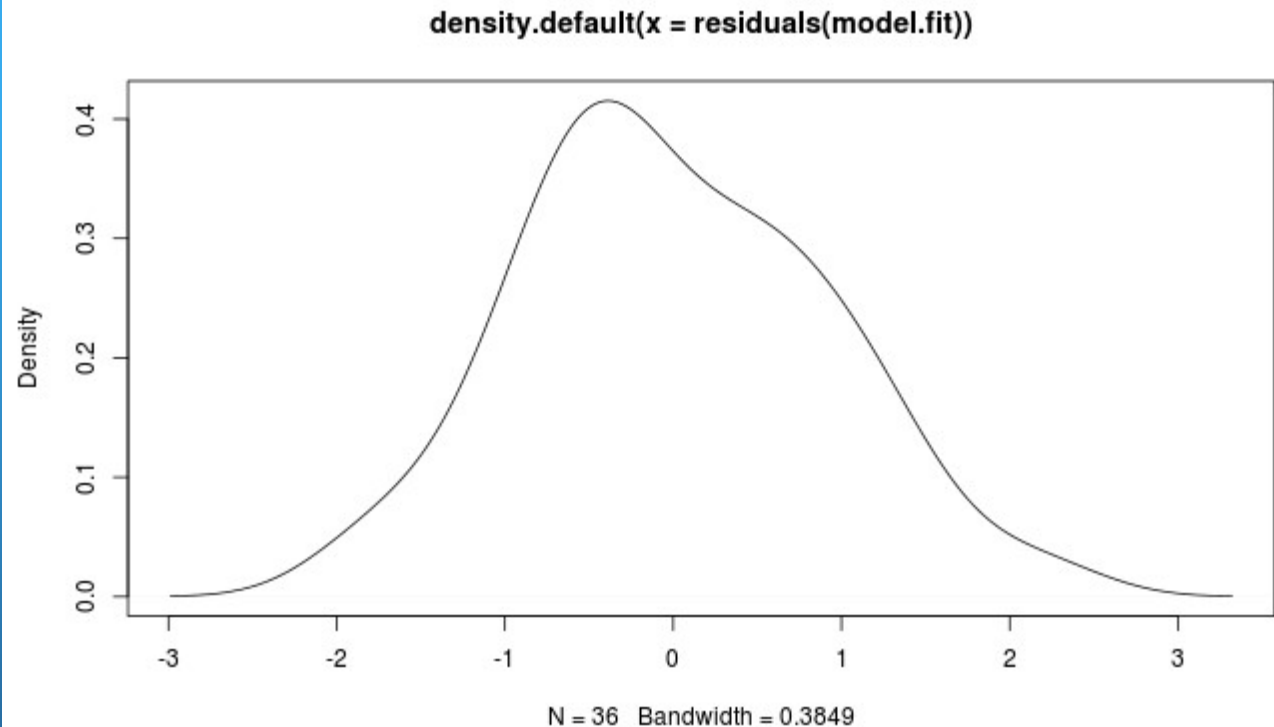
Model Fit Summary

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	497.7	248.86	209.567	6.37e-16	***
nitrogen	2	54.1	27.03	22.760	2.87e-06	***
block	3	20.8	6.92	5.825	0.003875	**
clone:nitrogen	4	43.3	10.82	9.111	0.000126	***
Residuals	24	28.5	1.19			

But we don't know yet if the ANOVA is valid! Look below this in the app output to check assumption tests and associated plots...

Kernel Density of the Residuals



Normality



Equal variance



Additive

Shapiro-Wilk Normality Test Results

model

Shapiro-Wilk normality test

data: residuals(fit)

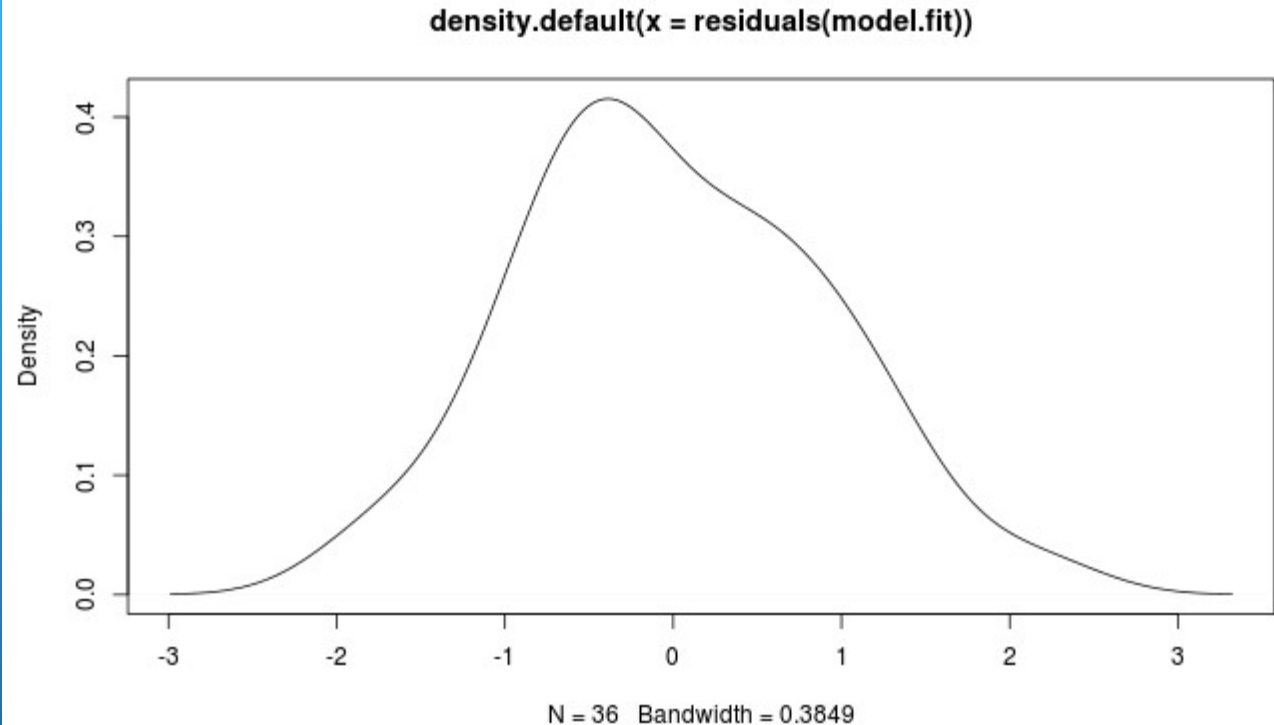
W = 0.98486, p-value = 0.893

Kernel Density of the Residuals



Normality

☐ Equal variance
☐ Additive



Shapiro-Wilk Normality Test Results

model

Shapiro-Wilk normality test

data: residuals(fit)

W = 0.98486, p-value = 0.893



P-value > 0.05 🏠 GOOD!

Fitted



Yield ~ Clone + Nitrogen + Clone:Nitrogen + Block

Normality

y



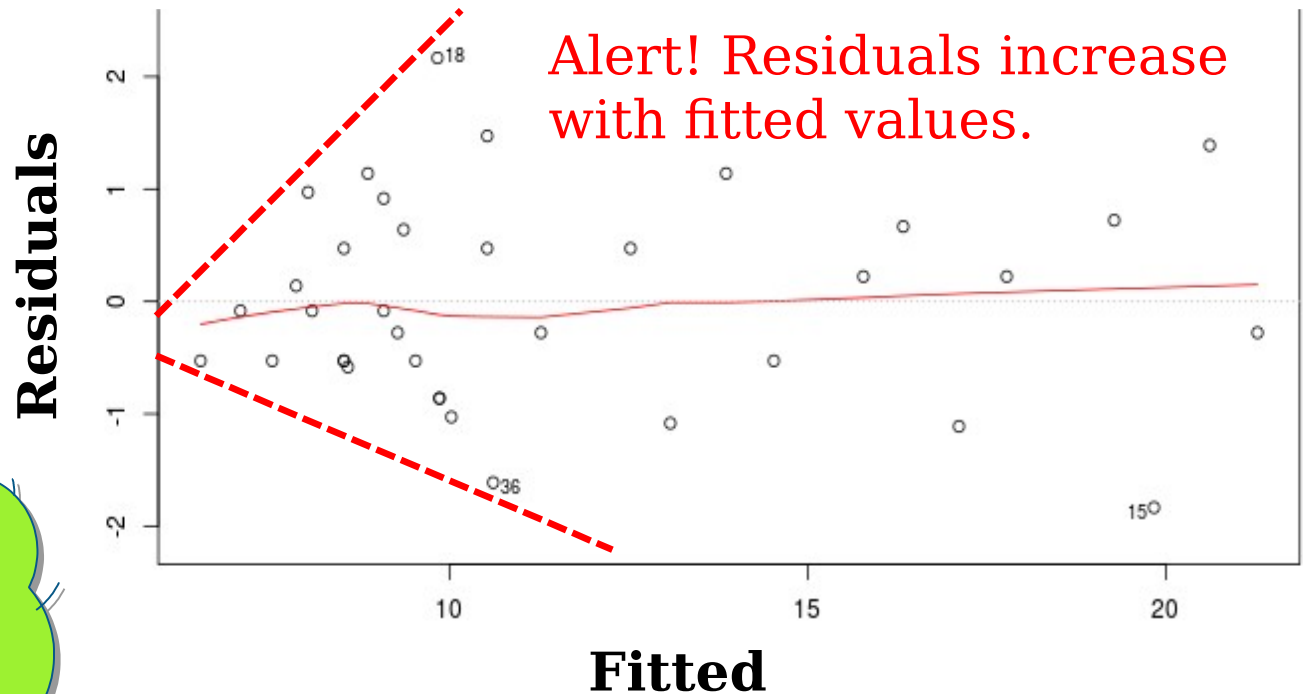
Equal
variance



Additive

Does the plot display any trend that may be a problem?

Residuals vs Fitted



Fitted



Yield ~ Clone + Nitrogen + Clone:Nitrogen + Block

Normality

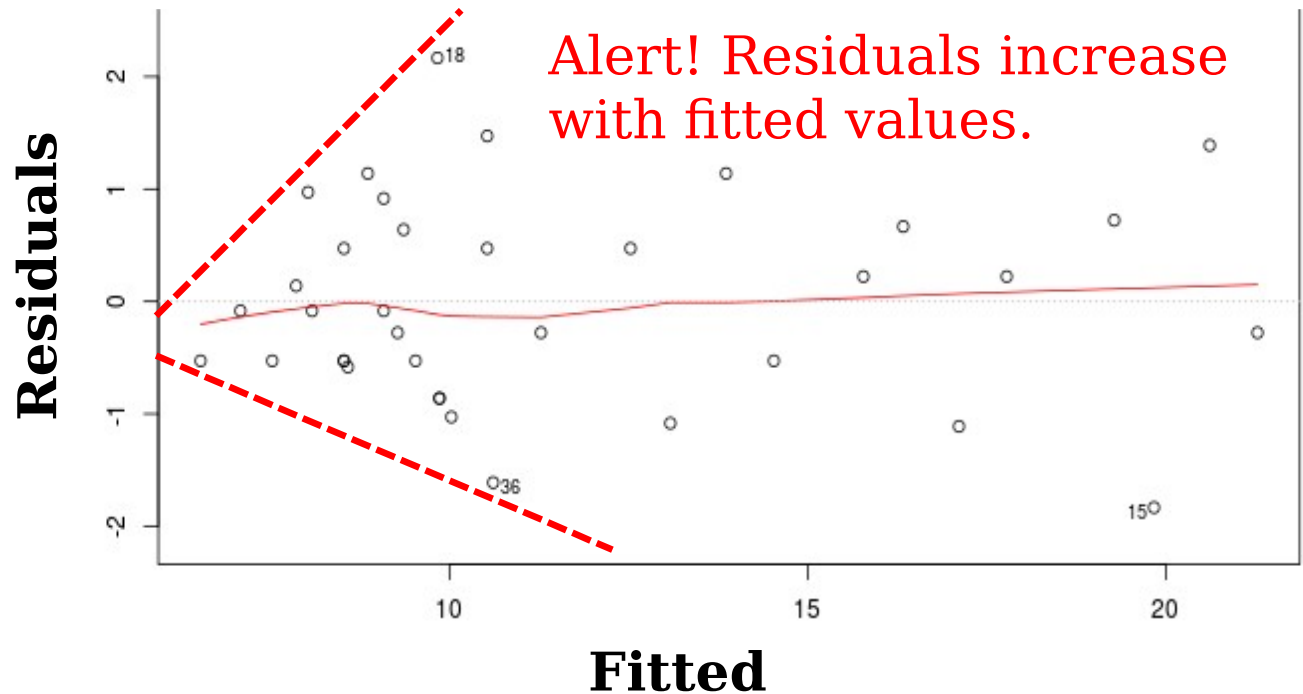
y

Equal variance

Additive

model

Residuals vs Fitted



Let's check the box plots to see how variance compares among clone and nitrogen treatment levels



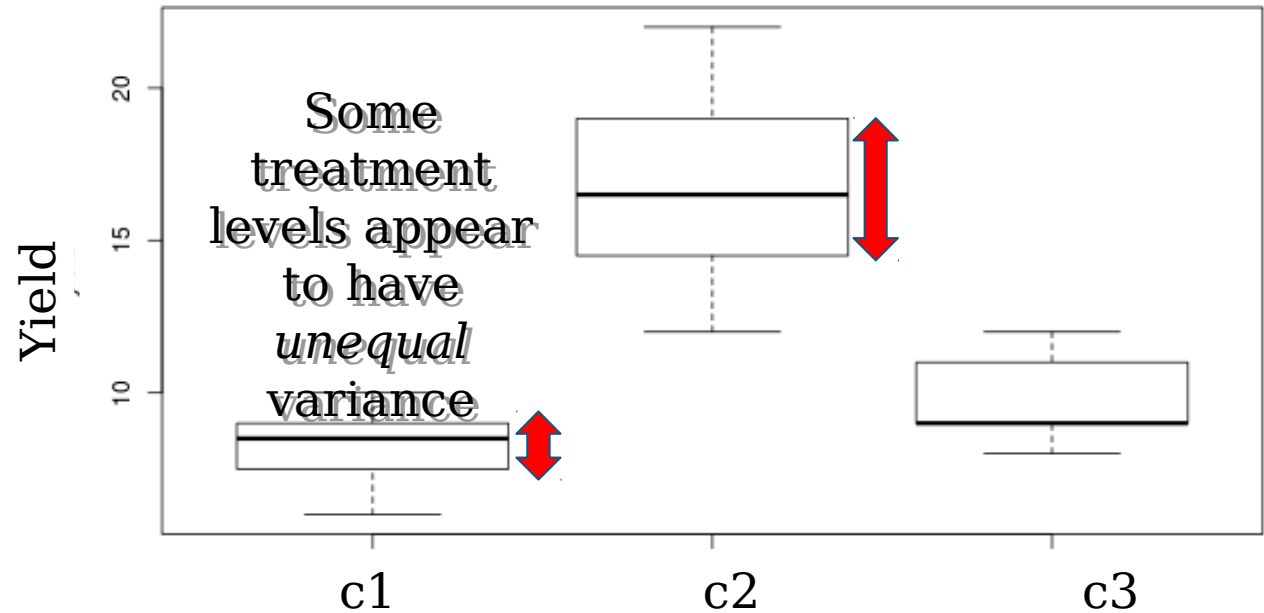
Normality

Equal variance

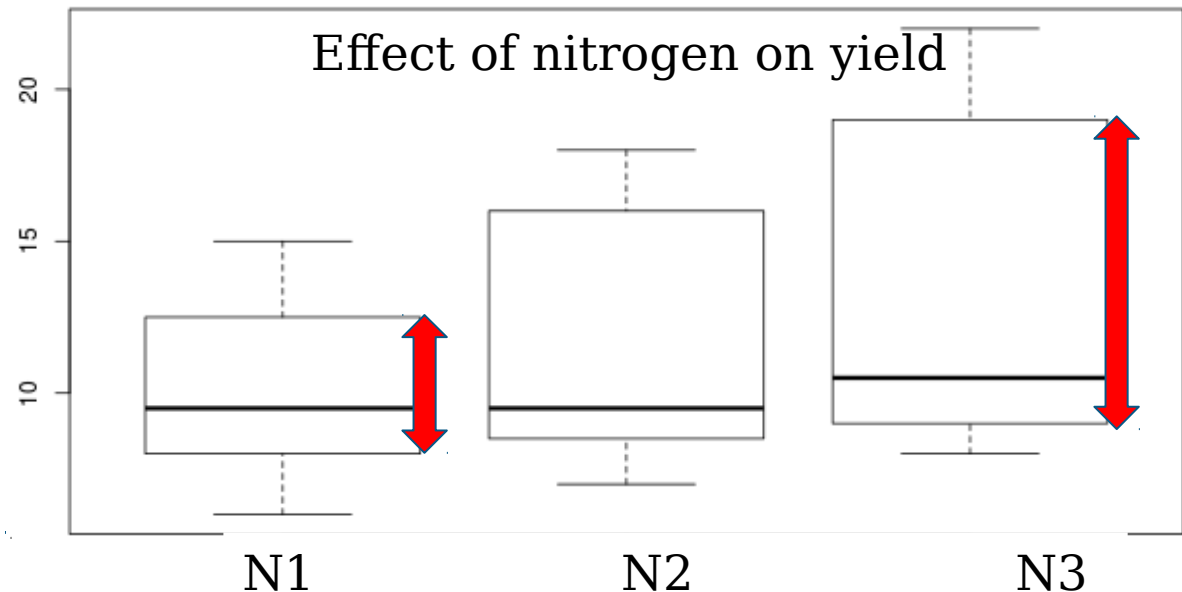
Additivity

What do these box plots tell you about variance across treatment levels?

Effect of clone on yield



Effect of nitrogen on yield





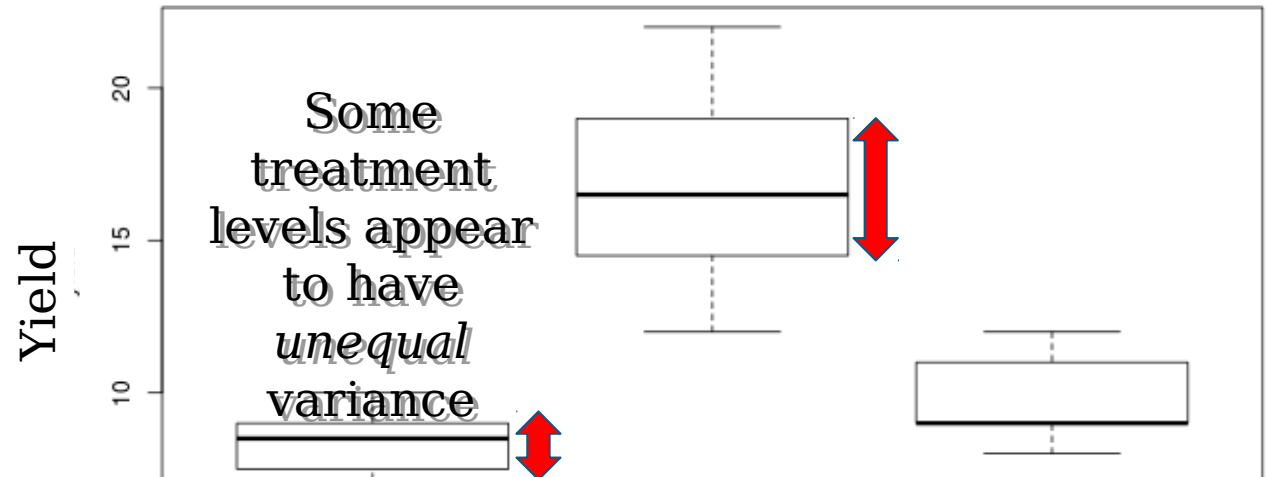
Normality

Equal variance

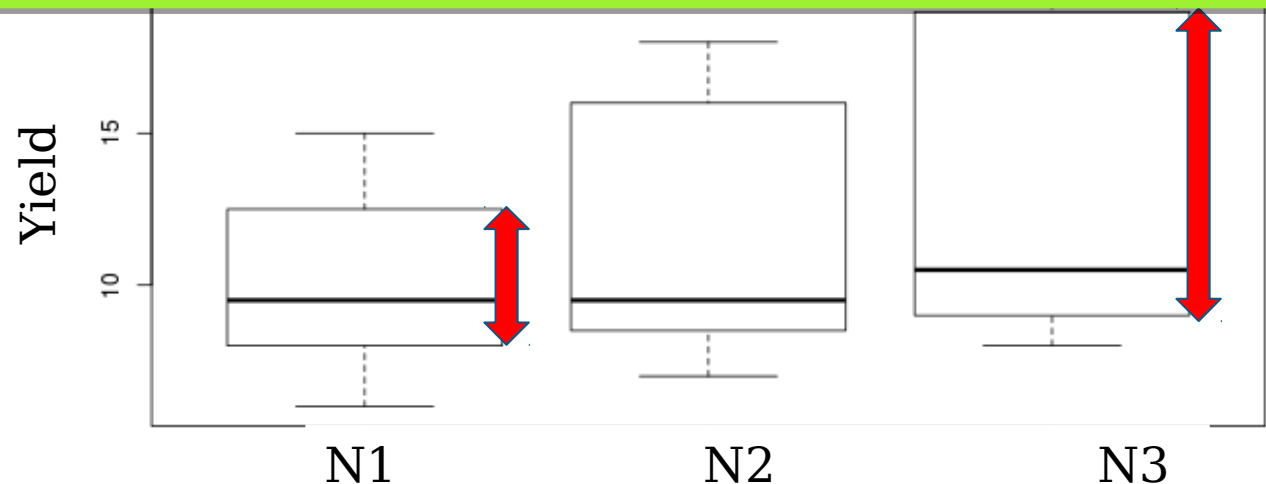
Additive

model

Effect of clone on yield



To confirm, let's check Levene's tests for (1) clone and (2) nitrogen





Normality



Equal
variance

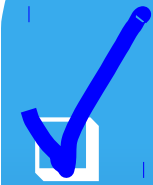


Assumptions

Does
everything
look
OK?

Levene's Test for Homogeneity of Variance

```
$`yield ~ clone`  
Levene's Test for Homogeneity of Variance (center = median)  
      Df F value  Pr(>F)  
group  2  5.2195 0.01073 *  
      33  
-----  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
$`yield ~ nitrogen`  
Levene's Test for Homogeneity of Variance (center = median)  
      Df F value Pr(>F)  
group  2  1.0005 0.3786
```

~~Normality~~


☐ Equal variance

☐ Additive model

Levene's Test for Homogeneity of Variance

```
$'yield ~ clone'
```


```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  5.2195 0.01073 *
      33
```

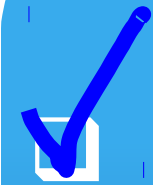
P-value < 0.05 
Unequal variance

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

```
$'yield ~ nitrogen'
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.0005 0.3786
```

P-value ≥ 0.05  equal variance



~~Normality~~

☐ Equal variance


☐ Additive

model

Levene's Test for Homogeneity of Variance

```
$'yield ~ clone'
```


```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  5.2195 0.01073 *
      33
```

P-value < 0.05 
Unequal variance

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

```
$'yield ~ nitrogen'
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.0005 0.3786
```

P-value ≥ 0.05  equal
variance

Let's check the last test, Tukey's
Test for Nonadditivity.



~~Normality~~

☐ Equal variance

☐ Additive

Does everything look OK?

Tukey's Test for Nonadditivity

Attention: Refer only to the 'yield.pred.sq' row in this table, ignore all other rows.

Analysis of Variance Table

Response: yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	497.72	248.861	201.4245	2.651e-15	***
nitrogen	2	54.06	27.028	21.8759	4.770e-06	***
block	3	20.75	6.917	5.5982	0.004935	**
yield.pred.sq	1	41.19	41.189	33.3376	6.993e-06	***
clone:nitrogen	4	2.17	0.543	0.4396	0.778695	
Residuals	23	28.42	1.236			

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



~~Normality~~

~~Equal variance~~

~~Additivity~~

Does everything look OK?

Tukey's Test for Nonadditivity


Attention: Refer only to the 'yield.pred.sq' row in this table, ignore all other rows.

Analysis of Variance Table

Response: yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	497.72	248.861	201.4245	2.651e-15	***
nitrogen	2	54.06	27.028	21.8759	4.770e-06	***
block	3	20.75	6.917	5.5982	0.004935	**
yield.pred.sq	1	41.19	41.189	33.3376	6.993e-06	***
clone:nitrogen	4	2.17	0.543	0.4396	0.778695	
Residuals	23	28.42	1.236			

Signif. codes: 0 '***' 0.001 '**'

P-value < 0.05 
Nonadditivity in model



- ☒ Normality
- ☒ Equal variance
- ☐ Additive

model





- ☒ Normality
- ☒ Equal variance
- ☐ Additive

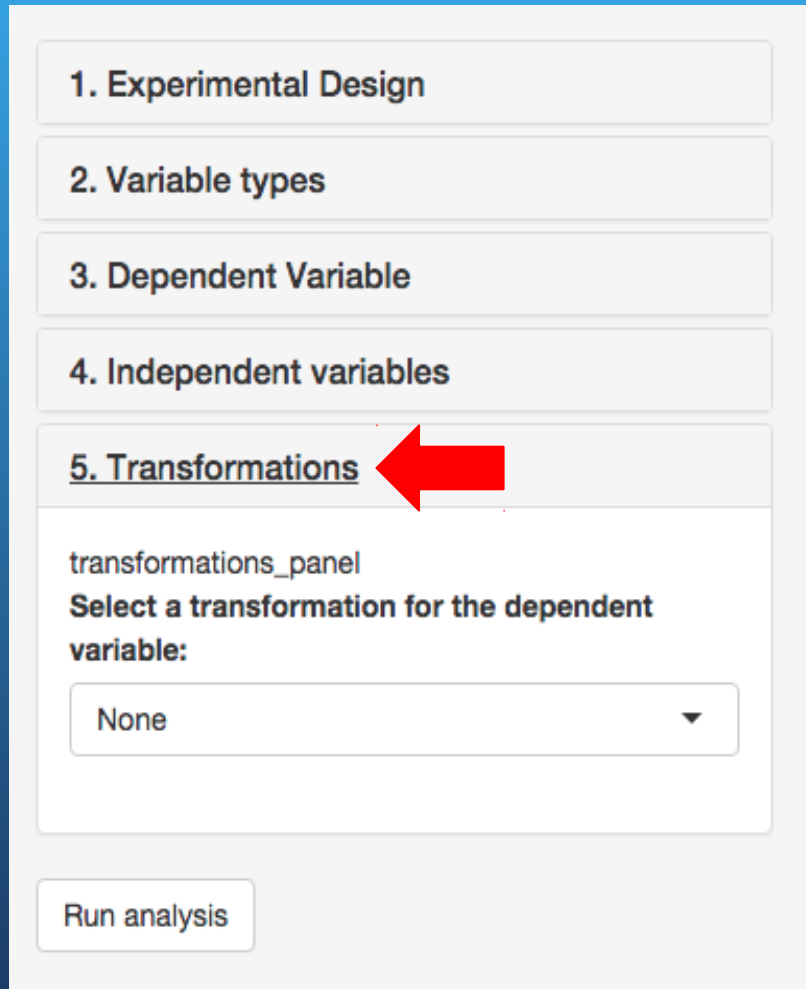
model

We need to transform the *yield data* to address
(1) unequal variance and
(2) nonadditivity.

Let's try both **log and power transformations.**

**First the power
transformation....**

Click Transformations and select Power.




1. Experimental Design

2. Variable types

3. Dependent Variable

4. Independent variables

5. Transformations 

transformations_panel

Select a transformation for the dependent variable:

None ▼

Run analysis

Click Transformations and select Power.

1. Experimental Design

2. Variable types

3. Dependent Variable

4. Independent variables

5. Transformations

transformations_panel

Select a transformation for the dependent variable:

Power ▼

Run analysis



Click Run analysis

1. Experimental Design

2. Variable types

3. Dependent Variable

4. Independent variables

5. Transformations

transformations_panel

Select a transformation for the dependent variable:

Power

Run analysis



Note the new model formula with the dependent variable, `yield.pow`

Model Formula

```
yield.pow ~ clone + nitrogen + clone:nitrogen + block
```

Exponent from Power Transformation

```
0.906523802529747
```

And the exponent used in the power transformation


Assumption of normality

Original data

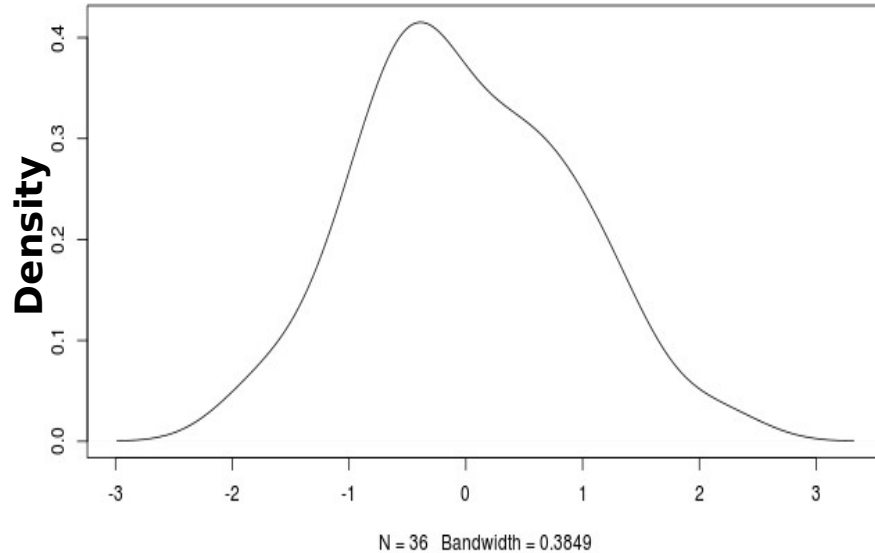
Shapiro-Wilk Test

Data: residuals(fit)

W = 0.98486, p-value = 0.893

Recall, $p > 0.05$  normally distributed residuals

density.default(x = residuals(model.fit))



Power-transformed

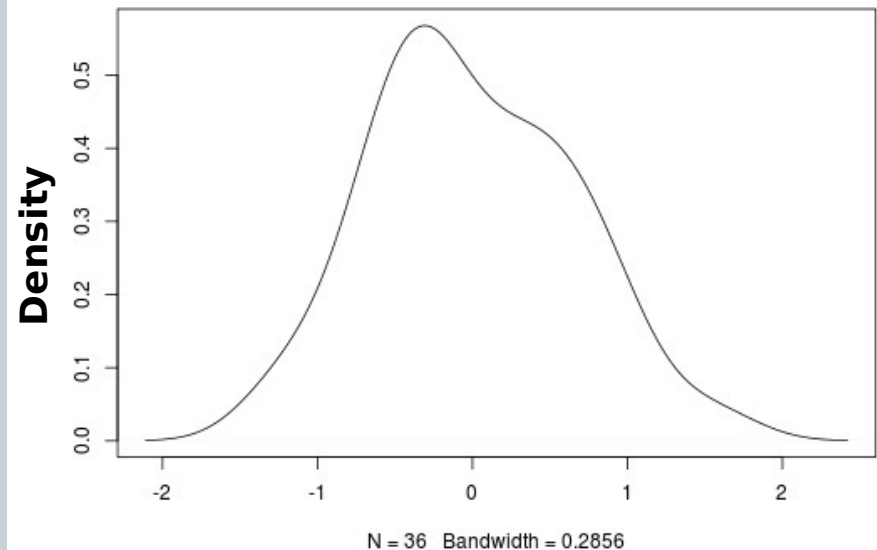
Shapiro-Wilk Test

Data: residuals(fit)

W = 0.98225, p-value = 0.8185

Also OK

density.default(x = residuals(model.fit))




Assumption of equal variance

Original data

Levene's Test

yield ~ clone

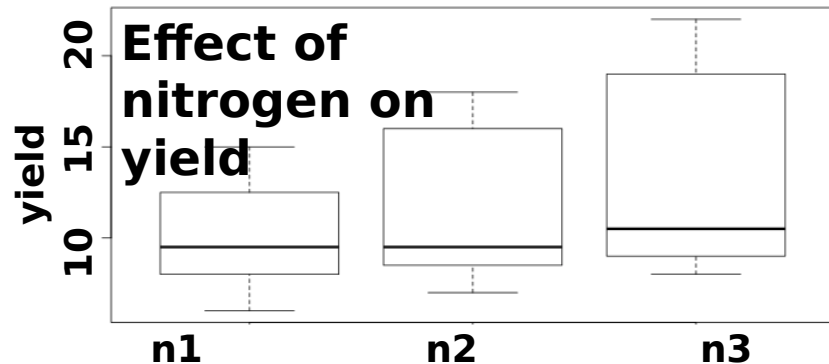
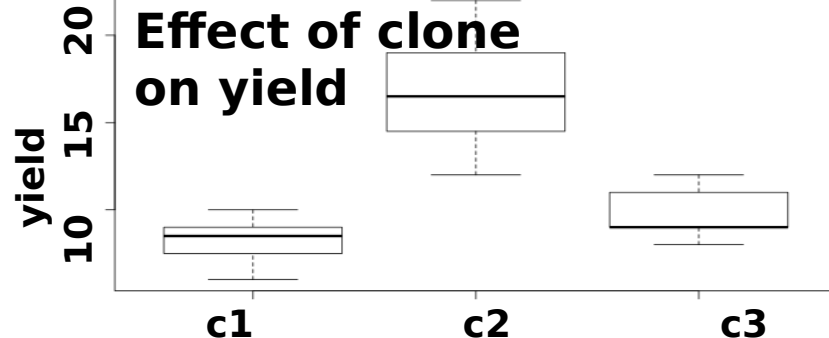
	Df	F value	Pr(>F)
clone	2	1.0005	0.3786

Recall, $p < 0.05$ 
unequal variance
among clone
levels

0.01073

Pr(>F)

0.3786



Power-transformed

Levene's Test

yield.pow ~ clone

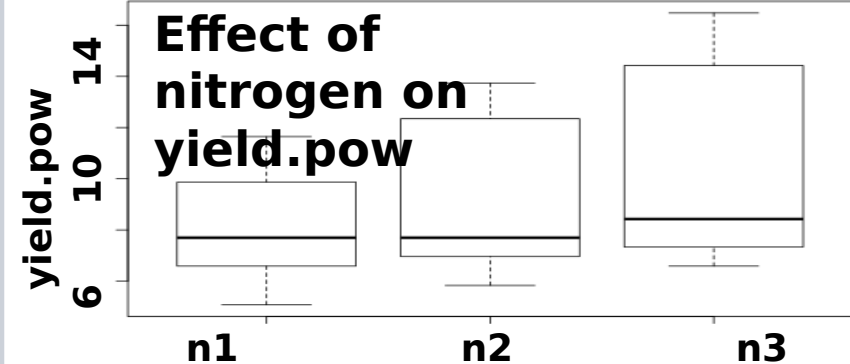
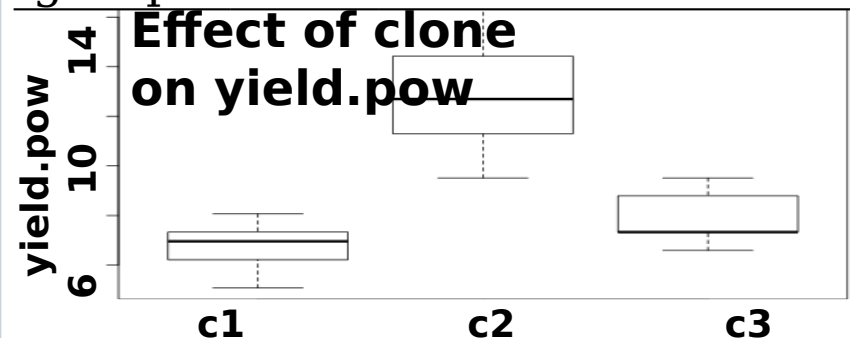
	Df	F value	Pr(>F)
clone	2	0.9421	0.4

Not much
better!

0.01707

Pr(>F)

0.4



Assumption of additivity

Original data

Tukey's Test for Nonadditivity

Attention: Refer only to the 'yield.pred.sq' row in this table, ignore all other rows.

Analysis of Variance Table

Response: yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	497.72	248.861	201.4245	2.651e-15	***
nitrogen	2	54.06	27.028	21.8759	4.770e-06	***
block	3	20.75	6.917	5.5982	0.004935	**
yield.pred.sq	1	41.19	41.189	33.3376	6.993e-06	***
clone:nitrogen	4	2.17	0.543	0.4396	0.78695	
Residuals	23	28.42	1.236			

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

Recall, $p\text{-value} < 0.05$
Nonadditivity in model

Power-transformed

Tukey's Test for Nonadditivity

Attention: Refer only to the 'yield.pred.sq' row in this table, ignore all other rows.

Analysis of Variance Table

Response: yield.pow

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	253.825	126.912	199.0748	3.012e-15	***
nitrogen	2	27.070	13.535	21.2308	5.971e-06	***
block	3	10.949	3.650	5.7251	0.004447	**
yield.pow.pred.sq	1	19.411	19.411	30.4483	1.303e-05	***
clone:nitrogen	4	1.255	0.314	0.4923	0.741386	
Residuals	23	14.663	0.638			

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

Not much better!
 $p < 0.05$ Nonadditivity in model

**Now let's try the
logarithmic (log)
transformation...**

Note the new model formula
with the dependent variable:
yield.log10

Model Formula

```
yield.log10 ~ clone + nitrogen + clone:nitrogen + block
```


Assumption of normality

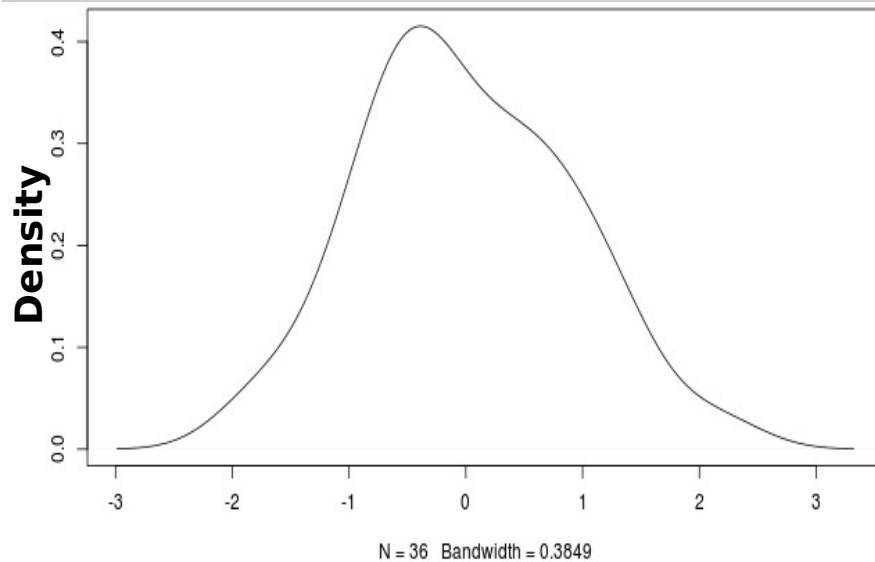
Original data

Shapiro-Wilk Test

Data: residuals(fit)

W = 0.98486, p-value = 0.893

Recall, $p > 0.05$  OK



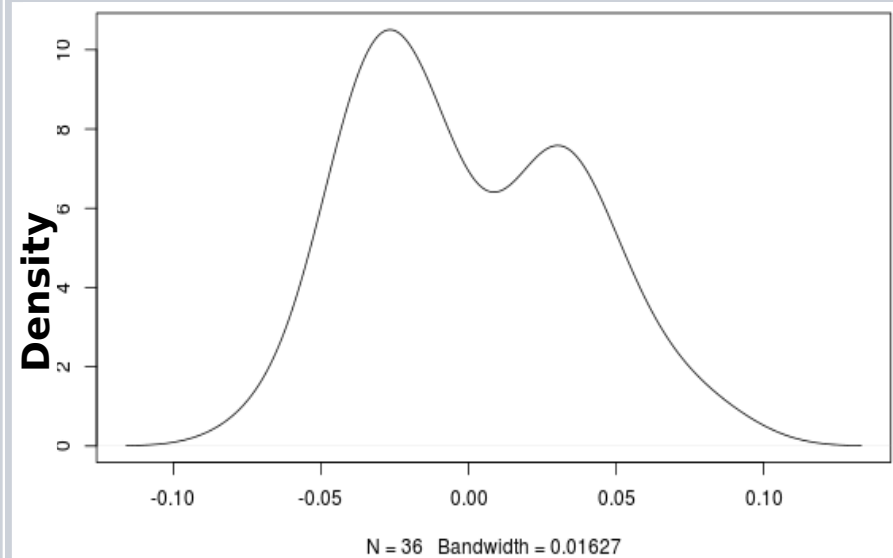
log-transformed

Shapiro-Wilk Test

Data: residuals(fit)

W = 0.94738, p-value = 0.08646

Also OK, but worse than original




Assumption of equal variance

Original data

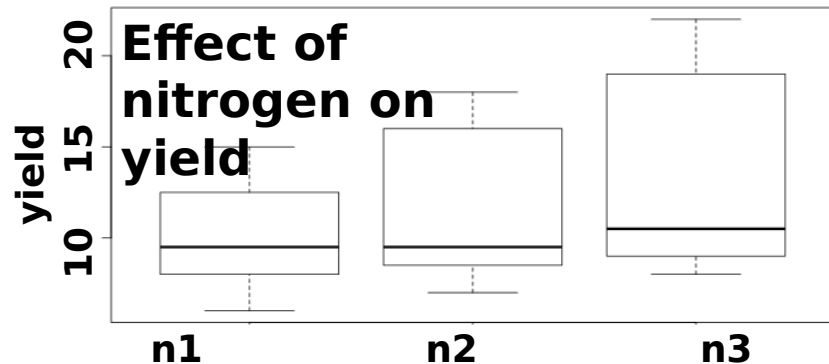
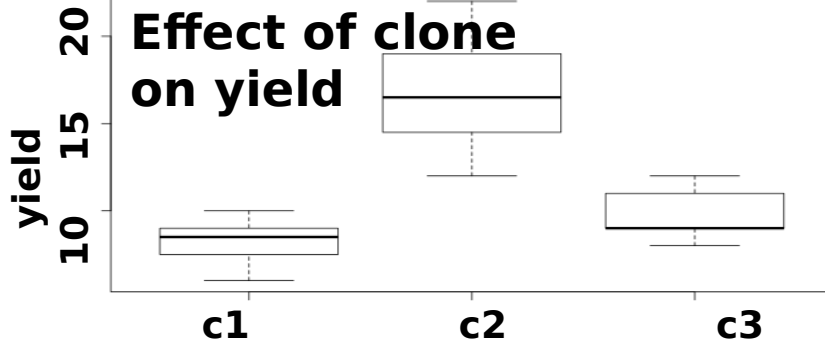
Levene's Test

yield ~ clone

	Df	F value	Pr(>F)
yield ~ clone	2	1.0005	0.01073
yield ~ nitrogen	2	1.0005	0.3786

Recall, $p < 0.05$ 
Bad (unequal
variance among
clone levels)

→ 0.01073



log-transformed

Levene's Test

yield.log10 ~ clone

	Df	F value	Pr(>F)
yield.log10 ~ clone	2	0.6221	0.6221
yield.log10 ~ nitrogen	2	0.6894	0.6894

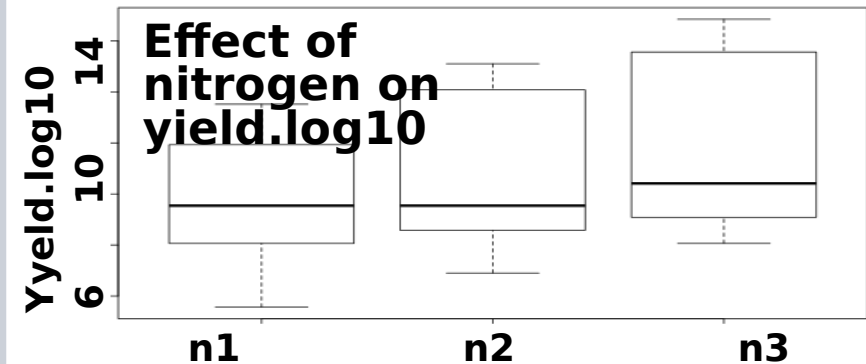
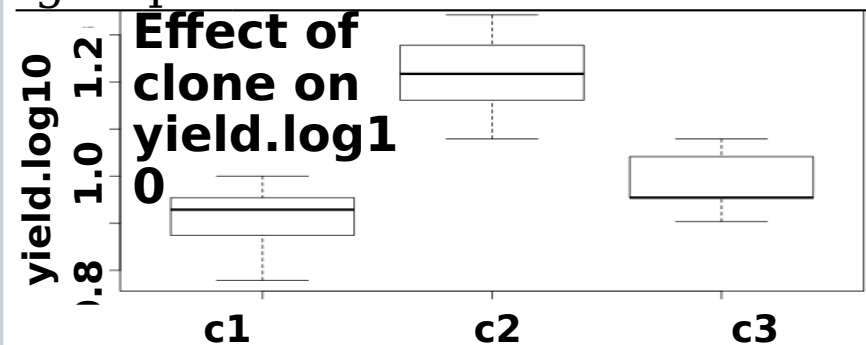
Improvement!
 $p > 0.05$  OK!

→ 0.6221

yield.log10 ~ nitrogen

Df F value Pr(>F)

group → 0.6894



Assumption of additivity

Original data

Tukey's Test for Nonadditivity

Attention: Refer only to the 'yield.pred.sq' row in this table, ignore all other rows.

Analysis of Variance Table

Response: yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	497.72	248.861	201.4245	2.651e-15	***
nitrogen	2	54.06	27.028	21.8759	4.770e-06	***
block	3	20.75	6.917	5.5982	0.004935	**
yield.pred.sq	1	41.19	41.189	33.3376	6.993e-06	***
clone:nitrogen	4	2.17	0.543	0.4396	0.78695	
Residuals	23	28.42	1.236			

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

Recall, $p\text{-value} < 0.05$ 🚩
Bad (nonadditivity in model)

log-transformed

Tukey's Test for Nonadditivity

Attention: Refer only to the 'yield.log10.pred.sq' row in this table, ignore all other rows.

Analysis of Variance Table

Response: yield.log10

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	0.60563	0.302815	165.9591	2.155e-14	***
nitrogen	2	0.05674	0.028371	15.5491	5.349e-05	***
block	3	0.03799	0.012662	6.9395	0.001713	**
yield.log10.pred.sq	1	0.01893	0.018925	10.3722	0.003787	**
clone:nitrogen	4	0.00993	0.002482	1.3605	0.278312	
Residuals	23	0.04197	0.001825			

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

Better, but $p < 0.05$ 🚩
Bad (Nonadditivity in model)

Overall, the log transformation improved the assumption tests the most.

Thus, we will use the log-transformed data for the final analysis.

What to do if data
are *still* failing the
Levene's test after
transformations?!



If your data transformation does not improve your Levene's test, you may want to use a Weighted Least Squares (WLS) approach, but this is not covered in the course or available (yet) with the app.

Output for log-transformed yield

Model Formula

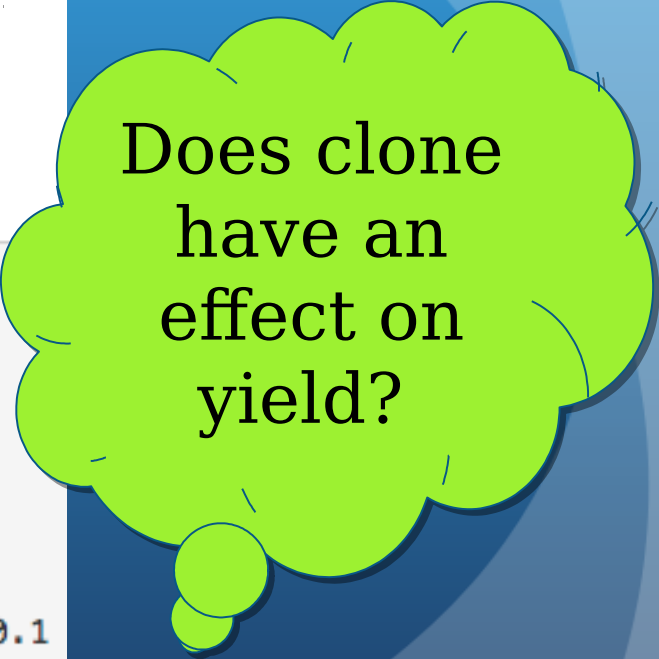
```
yield.log10 ~ clone + nitrogen + clone:nitrogen + block
```

Model Fit Summary

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	0.6056	0.30281	151.468	2.45e-14	***
nitrogen	2	0.0567	0.02837	14.191	8.56e-05	***
block	3	0.0380	0.01266	6.334	0.00256	**
clone:nitrogen	4	0.0228	0.00571	2.856	0.04554	*
Residuals	24	0.0480	0.00200			

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



Does clone
have an
effect on
yield?

Output for log-transformed yield

Model Formula

```
yield.log10 ~ clone + nitrogen + clone:nitrogen + block
```

Model Fit Summary

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	0.6056	0.30281	151.468	2.45e-14	***
nitrogen	2	0.0567	0.02837	14.191	8.56e-05	***
block	3	0.0380	0.01266	6.334	0.00256	**
clone:nitrogen	4	0.0228	0.00571	2.856	0.04554	*
Residuals	24	0.0480	0.00200			

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

The effect of clone on yield *depends* on the nitrogen level.

Output for log-transformed yield

Model Formula

```
yield.log10 ~ clone + nitrogen + clone:nitrogen + block
```

Model Fit Summary

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	0.6056	0.30281	151.468	2.45e-14	***
nitrogen	2	0.0567	0.02837	14.191	8.56e-05	***
block	3	0.0380	0.01266	6.334	0.00256	**
clone:nitrogen	4	0.0228	0.00571	2.856	0.04554	*
Residuals	24	0.0480	0.00200			

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

We know this because there is a **significant interaction**.
Interaction = the effect of one factor depends on the level of the other factor.

Output for log-transformed yield

Model Formula

yield.log10 ~ clone + nitrogen + clone:nitrogen + block

(interaction)

Model Fit Summary

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
clone	2	0.6056	0.30281	151.468	2.45e-14	***
nitrogen	2	0.0567	0.02837	14.191	8.56e-05	***
block	3	0.0380	0.01266	6.334	0.00256	**
clone:nitrogen	4	0.0228	0.00571	2.856	0.04554	*
Residuals	24	0.0480	0.00200			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0

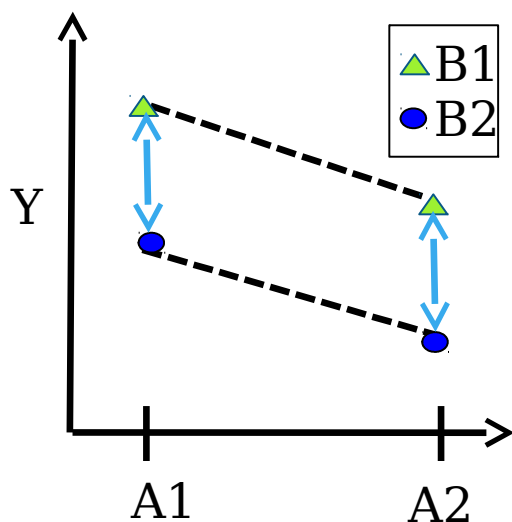
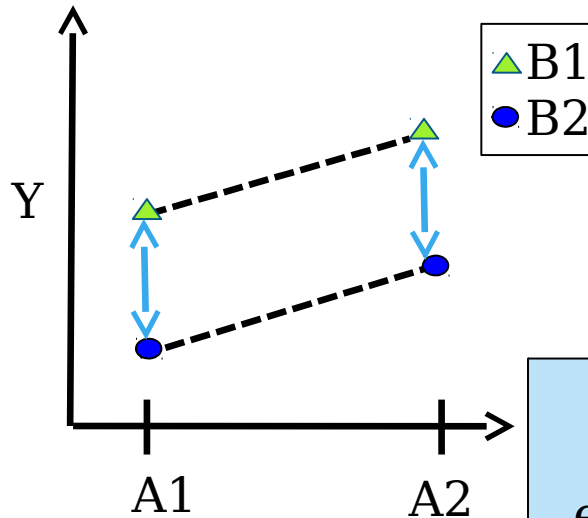
In a 2-factor factorial, there are 3 terms to look at in ANOVA table:


1. Factor A (clone)
2. Factor B (nitrogen)
3. AxB (clone:nitrogen)

If interaction is significant ($p < 0.05$), LSD test is performed on all treatment level combinations (simple effects). For example, we can't say what N level produced the highest yield (main effect) unless we refer

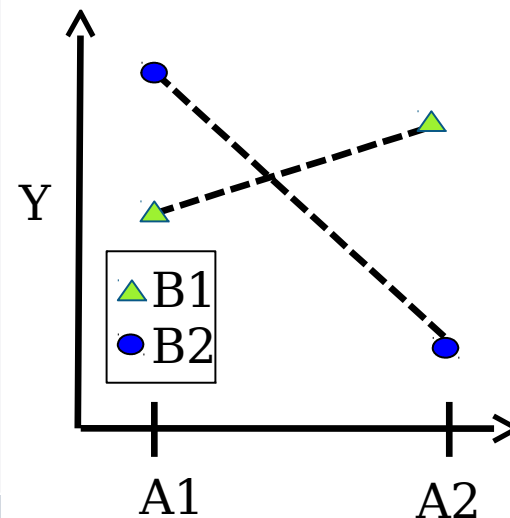
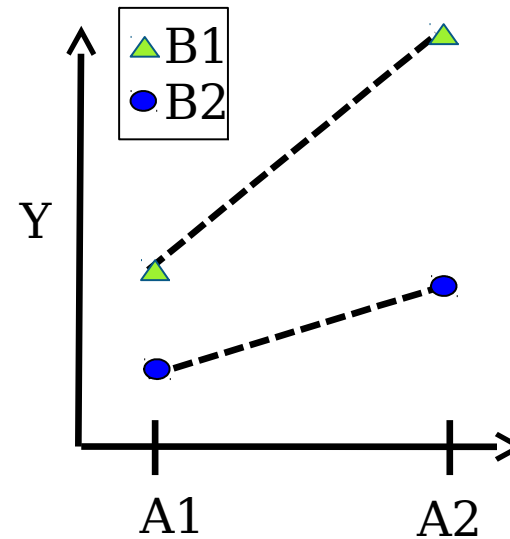
Interaction plots


No Interaction



Parallel lines 
effect of A
on Y is
independent
of B

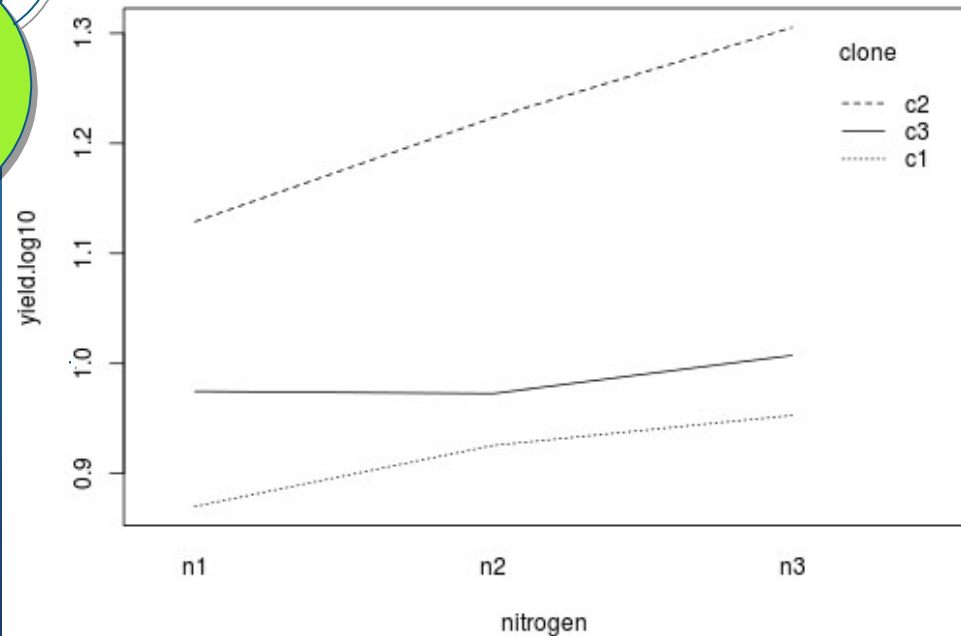
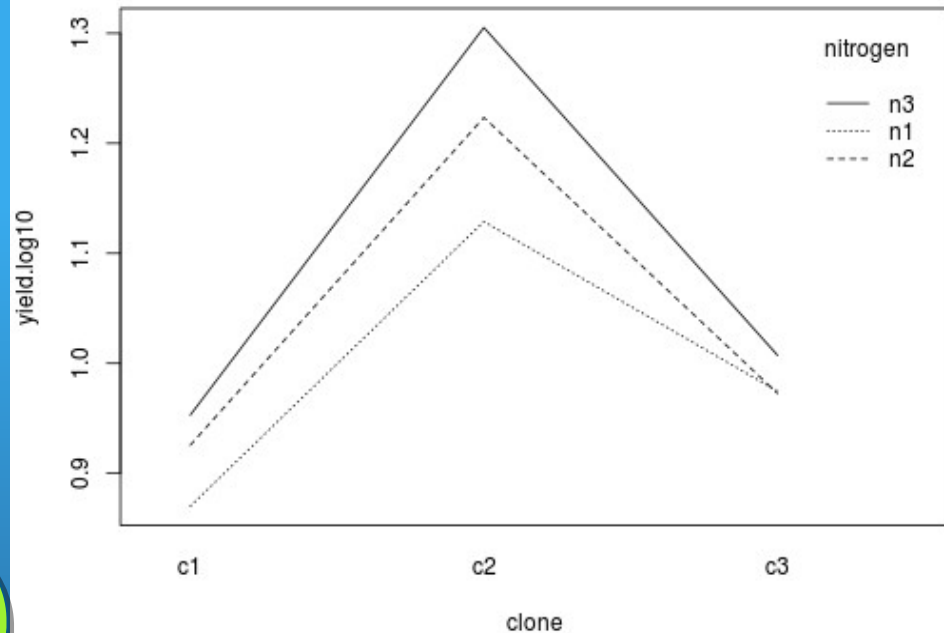
Interaction



Intersecting
or non-
parallel
lines 
effect of A
on Y is
dependent
on B

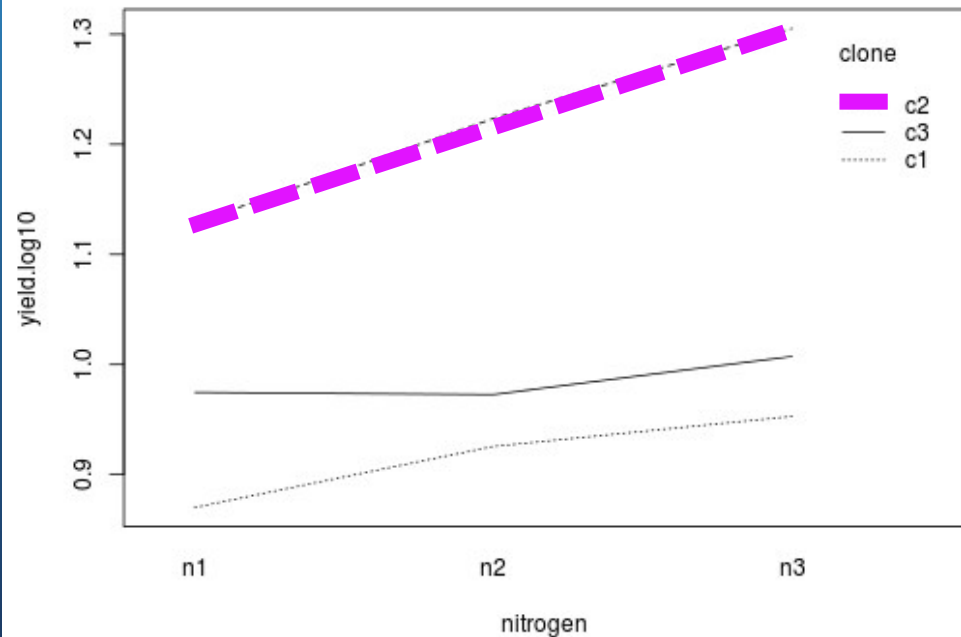
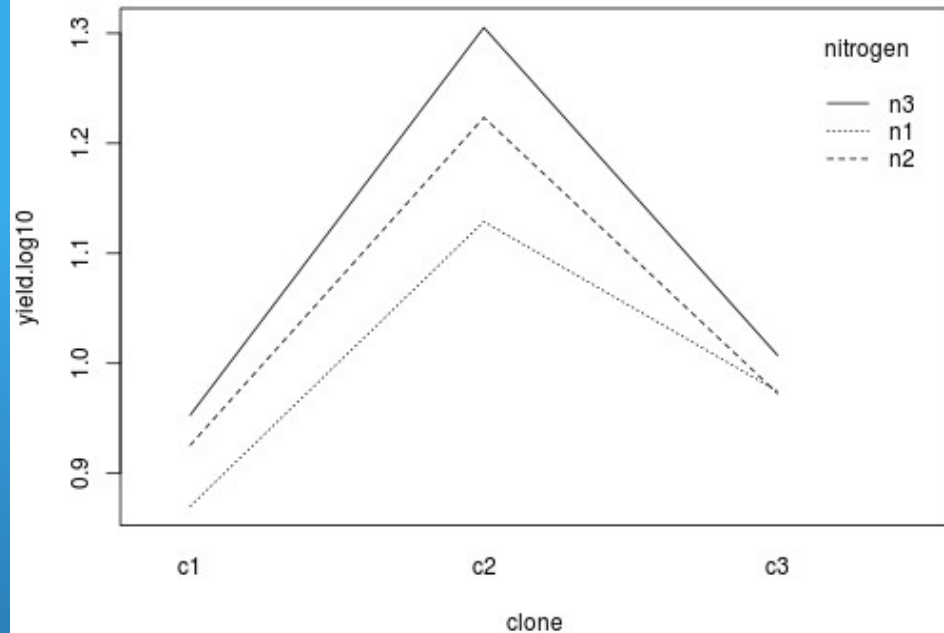
Interaction plots from log-transformed yield variable

What do these plots say about the nitrogen by clone interaction on yield?



Interaction plots from log-transformed yield variable

The effect of N rate differs for each clone, with the strongest N effect exhibited in clone 2.



Click: Post-hoc tests



Agricultural Field Trial Statistics Package

1. Load data

2. Analysis

3. Post-hoc tests

4. Report

About

1. Experimental Design

2. Variable types

3. Dependent Variable

4. Independent variables

5. Transformations

transformations_panel

Select a transformation for the dependent variable:

Logarithmic

Run analysis

```
1 # convert categorical variables to factors
2 my.data$block <- as.factor(my.data$block)
3 my.data$clone <- as.factor(my.data$clone)
4 my.data$nitrogen <- as.factor(my.data$nitrogen)
5
6 # transform the dependent variable
7 my.data$yield.log10 <- log10(my.data$yield)
8
9 # fit the model
```

Model Formula

```
yield.log10 ~ clone + nitrogen + clone:nitrogen + block
```

Model Fit Summary

ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
clone	2	0.6056	0.30281	151.468	2.45e-14 ***
nitrogen	2	0.0567	0.02837	14.191	8.56e-05 ***
block	3	0.0380	0.01266	6.334	0.00256 **
clone:nitrogen	4	0.0228	0.00571	2.856	0.04554 *
Residuals	24	0.0480	0.00200		

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

Click: Run post-hoc analysis

Agricultural Field Trial Statistics Package

1. Load data

2. Analysis

3. Post-hoc tests

4. Report

About

Post hoc analysis

Run post hoc analysis



Post hoc tests and figures

Post hoc tests and figures

The interaction, clone:nitrogen, is significant (alpha = 0.05). Both factors are significant.

The app chooses the correct post-hoc tests and figures based on the ANOVA.

Study: fit ~ lsd.vars

LSD t Test for yield.log10

Mean Square Error: 0.001999199

clone:nitrogen, means and individual (95 %) CI

	yield.log10	std	r	LCL	UCL	Min	Max
c1:n1	0.8701454	0.07582428	4	0.8240045	0.9162864	0.7781513	0.9542425
c1:n2	0.9256076	0.06669088	4	0.8794667	0.9717486	0.8450980	1.0000000
c1:n3	0.9528938	0.03959399	4	0.9067528	0.9990347	0.9030900	1.0000000
c2:n1	1.1288360	0.04171115	4	1.0826950	1.1749769	1.0791812	1.1760913
c2:n2	1.2234903	0.02455571	4	1.1773494	1.2696313	1.2041200	1.2552725
c2:n3	1.3052361	0.03735114	4	1.2590952	1.3513771	1.2552725	1.3424227
c3:n1	0.9746813	0.05953852	4	0.9285404	1.0208222	0.9030900	1.0413927
c3:n2	0.9726891	0.07497815	4	0.9265481	1.0188300	0.9030900	1.0791812
c3:n3	1.0072647	0.06313851	4	0.9611238	1.0534057	0.9542425	1.0791812

alpha: 0.05 ; Df Error: 24

Critical Value of t: 2.063899

Least Significant Difference 0.06525313

Means with the same letter are not significantly different.

Post hoc tests and figures

The interaction, clone:nitrogen, is significant ($\alpha = 0.05$). Both factors are significant.

Study: fit ~ lsd.vars

LSD t Test for yield.log10

Mean Square Error: 0.001999

NOTE: Output is in log-scale due to the transformation. We will learn how to present results in the original units in the next section.

clone:nitrogen, means and individual (95 %) CI

	yield.log10	std	r		LCL	UCL	Min	Max
c1:n	0.8701454	0.07582428	4	0.8240045	0.9162864	0.7781513	0.9542425	
c1:n	0.9256076	0.06669088	4	0.8794667	0.9717486	0.8450980	1.0000000	
c1:n	0.9528938	0.03959399	4	0.9067528	0.9990347	0.9030900	1.0000000	
c2:n	1.1288360	0.04171115	4	1.0826950	1.1749769	1.0791812	1.1760913	
c2:n	1.2234903	0.02455571	4	1.1773494	1.2696313	1.2041200	1.2552725	
c2:n	1.3052361	0.03735114	4	1.2590952	1.3513771	1.2552725	1.3424227	
c3:n	0.9746813	0.05953852	4	0.9285404	1.0208222	0.9030900	1.0413927	
c3:n	0.9726891	0.07497815	4	0.9265481	1.0188300	0.9030900	1.0791812	
c3:n	1.0072647	0.06313851	4	0.9611238	1.0534057	0.9542425	1.0791812	

alpha: 0.05 ; Df Error: 24

Critical Value of t: 2.063899

Least Significant Difference 0.06525313

Means with the same letter are not significantly different.

Post hoc tests and figures

The interaction, clone:nitrogen, is significant ($\alpha = 0.05$). Both factors are significant.

Study: fit ~ lsd.vars

LSD t Test for yield.log10

Mean Square Error: 0.001999199

clone:nitrogen, means and individual (95 %) CI

	yield.log10	std	r	LCL	UCL	Min	Max
c1:n	0.8701454	0.07582428	4	0.8240045	0.9162864	0.7781513	0.9542425
c1:n	0.9256076	0.06669088	4	0.8794667	0.9717486	0.8450980	1.0000000
c1:n	0.9528938	0.03959399	4	0.9067528	0.9990347	0.9030900	1.0000000
c2:n	1.1288360	0.04171115	4	1.0826950	1.1749769	1.0791812	1.1760913
c2:n	1.2234903	0.02455571	4	1.1773494	1.2696313	1.2041200	1.2552725
c2:n	1.3052361	0.03735114	4	1.2590952	1.3513771	1.2552725	1.3424227
c3:n	0.9746813	0.05953852	4	0.9285404	1.0208222	0.9030900	1.0413927
c3:n	0.9726891	0.07497815	4	0.9265481	1.0188300	0.9030900	1.0791812
c3:n	1.0072647	0.06313851	4	0.9611238	1.0534057	0.9542425	1.0791812

alpha: 0.05 ; Df Error: 24

Critical Value of t: 2.063899

Least Significant Difference 0.06525313

Means with the same letter are not significantly different.

Key output includes:

1. Least square means

Post hoc tests and figures

The interaction, clone:nitrogen, is significant ($\alpha = 0.05$). Both factors are significant.

Study: fit ~ lsd.vars

LSD t Test for yield.log10

Mean Square Error: 0.001999199

clone:nitrogen, means and individual (95 %) CI

	yield.log10	std	r	LCL	UCL	Min	Max
c1:n1	0.8701454	0.07582428	4	0.8240045	0.9162864	0.7781513	0.9542425
c1:n2	0.9256076	0.06669088	4	0.8794667	0.9717486	0.8450980	1.0000000
c1:n3	0.9528938	0.03959399	4	0.9067528	0.9990347	0.9030900	1.0000000
c2:n1	1.1288360	0.04171115	4	1.0826950	1.1749769	1.0791812	1.1760913
c2:n2	1.2234908	0.02455571	4	1.1773494	1.2696313	1.2041200	1.2552725
c2:n3	1.3052361	0.03735114	4	1.2590952	1.3513771	1.2552725	1.3424227
c3:n1	0.9746818	0.05953852	4	0.9285404	1.0208222	0.9030900	1.0413927
c3:n2	0.9726891	0.07497815	4	0.9265481	1.0188300	0.9030900	1.0791812
c3:n3	1.0072647	0.06313851	4	0.9611238	1.0534057	0.9542425	1.0791812

alpha: 0.05 ; Df Error: 24

Critical Value of t: 2.063899

Least Significant Difference 0.06525313

Means with the same letter are not significantly different.

Key output includes:

1. Least square means
2. Standard deviation

Post hoc tests and figures

The interaction, clone:nitrogen, is significant (alpha = 0.05). Both factors are significant.

Study: fit ~ lsd.vars

LSD t Test for yield.log10

Mean Square Error: 0.001999199

clone:nitrogen, means and individual (95 %) CI

	yield.log10	std r		LCL	UCL	Min	Max
c1:n1	0.8701454	0.07582428	4	0.8240045	0.9162864	.7781513	0.9542425
c1:n2	0.9256076	0.06669088	4	0.8794667	0.9717486	.8450980	1.0000000
c1:n3	0.9528938	0.03959399	4	0.9067528	0.9990347	.9030900	1.0000000
c2:n1	1.1288360	0.04171115	4	1.0826950	1.1749769	.0791812	1.1760913
c2:n2	1.2234903	0.02455571	4	1.1773494	1.2696313	.2041200	1.2552725
c2:n3	1.3052361	0.03735114	4	1.2590952	1.3513771	.2552725	1.3424227
c3:n1	0.9746813	0.05953852	4	0.9285404	1.0208222	.9030900	1.0413927
c3:n2	0.9726891	0.07497815	4	0.9265481	1.0188300	.9030900	1.0791812
c3:n3	1.0072647	0.06313851	4	0.9611238	1.0534057	.9542425	1.0791812

alpha: 0.05 ; Df Error: 24

Critical Value of t: 2.063899

Least Significant Difference 0.06525313

Means with the same letter are not significantly different.

Key output includes:

1. Least square means
2. Standard deviation
3. 95% confidence interval

Post hoc tests and figures

The interaction, clone:nitrogen, is significant ($\alpha = 0.05$). Both factors are significant.

Study: fit ~ lsd.vars

LSD t Test for yield.log10

Mean Square Error: 0.001999199

clone:nitrogen, means and individual (95 %)

	yield.log10	std	r	LCL	UCL	Min	Max
c1:n1	0.8701454	0.07582428	4	0.8240045	0.9162864	0.7781513	0.9542425
c1:n2	0.9256076	0.06669088	4	0.8794667	0.9717486	0.8450980	1.0000000
c1:n3	0.9528938	0.03959399	4	0.9067528	0.9990347	0.9030900	1.0000000
c2:n1	1.1288360	0.04171115	4	1.0826950	1.1749769	1.0791812	1.1760913
c2:n2	1.2234903	0.02455571	4	1.1773494	1.2696313	1.2041200	1.2552725
c2:n3	1.3052361	0.03735114	4	1.2590952	1.3513771	1.2552725	1.3424227
c3:n1	0.9746813	0.05953852	4	0.9285404	1.0208222	0.9030900	1.0413927
c3:n2	0.9726891	0.07497815	4	0.9265481	1.0188300	0.9030900	1.0791812
c3:n3	1.0072647	0.06313851	4	0.9611238	1.0534057	0.9542425	1.0791812

alpha: 0.05 ; Df Error: 24

Critical Value of t: 2.063899

Least Significant Difference 0.06525313

Means with the same letter are not significantly different.

Key output includes:

1. Least square means
2. Standard deviation
3. 95% confidence interval
4. Minimum and Maximum values

Post hoc tests and figures

The interaction, clone:nitrogen, is significant ($\alpha = 0.05$). Both factors are significant.

alpha: 0.05 ; Df Error: 24

Critical Value of t: 2.063899

Least Significant Difference 0.06525313

Means with the same letter are not significantly different.

Groups	Treatments and means	
a	c2:n3	1.305
b	c2:n2	1.223
c	c2:n1	1.129
d	c3:n3	1.007
de	c3:n1	0.9747
de	c3:n2	0.9727
de	c1:n3	0.9529
ef	c1:n2	0.9256
f	c1:n1	0.8701

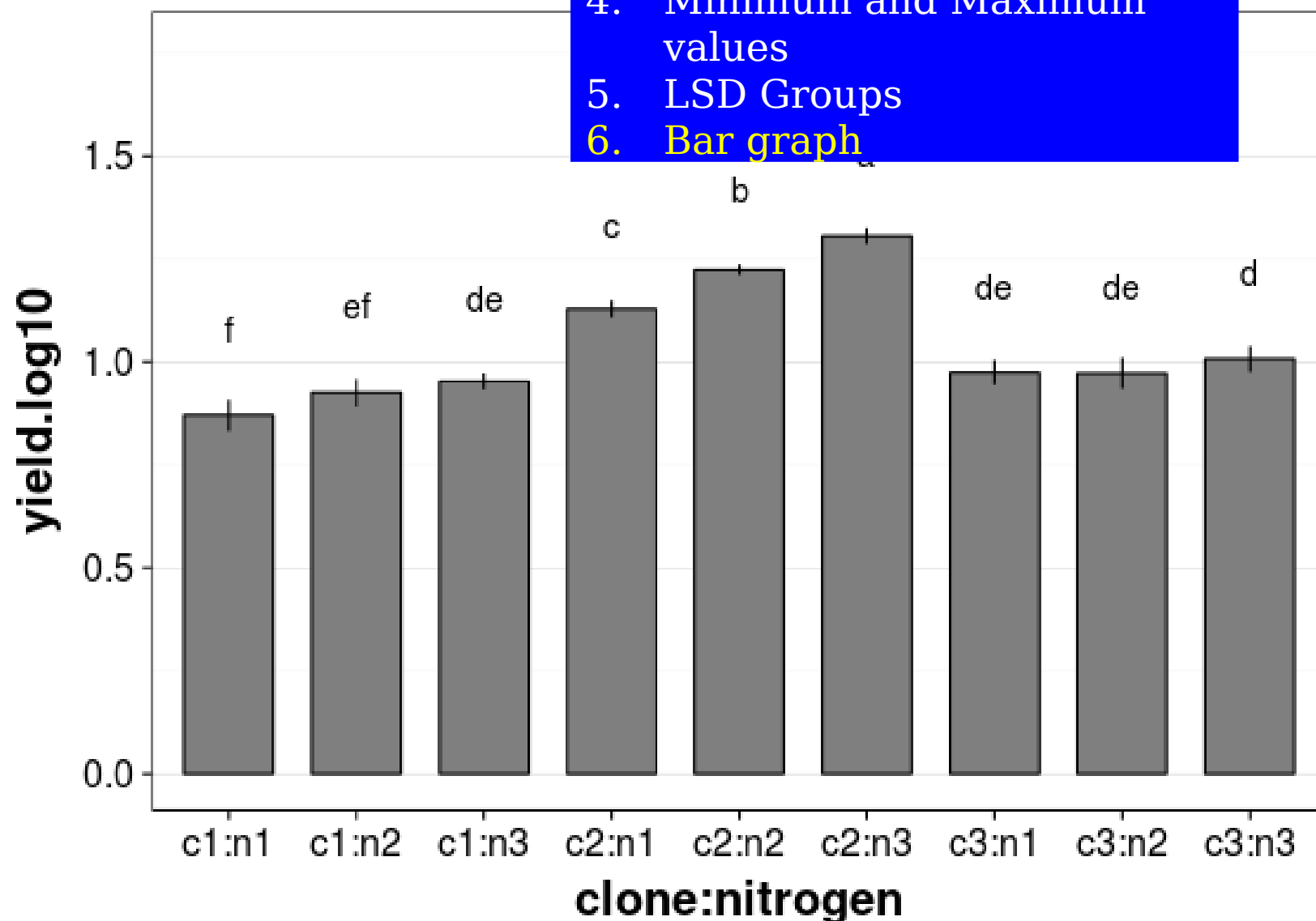
Key output includes:

1. Least square means
2. Standard deviation
3. 95% confidence interval
4. Minimum and Maximum values
5. LSD Groups

Post hoc tests and figures

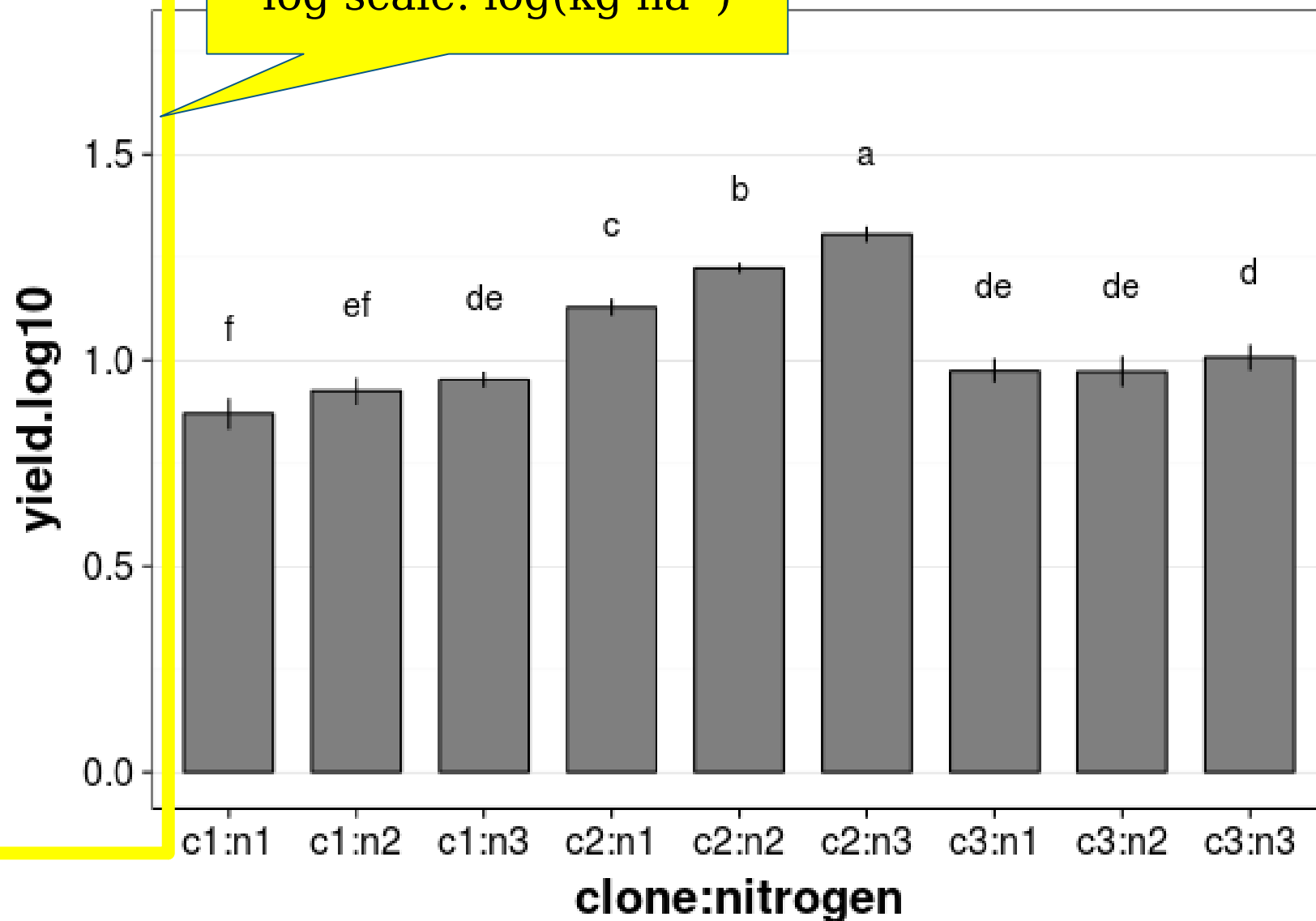
Key output includes:

1. Least square means
2. Standard deviation
3. 95% confidence interval
4. Minimum and Maximum values
5. LSD Groups
6. Bar graph



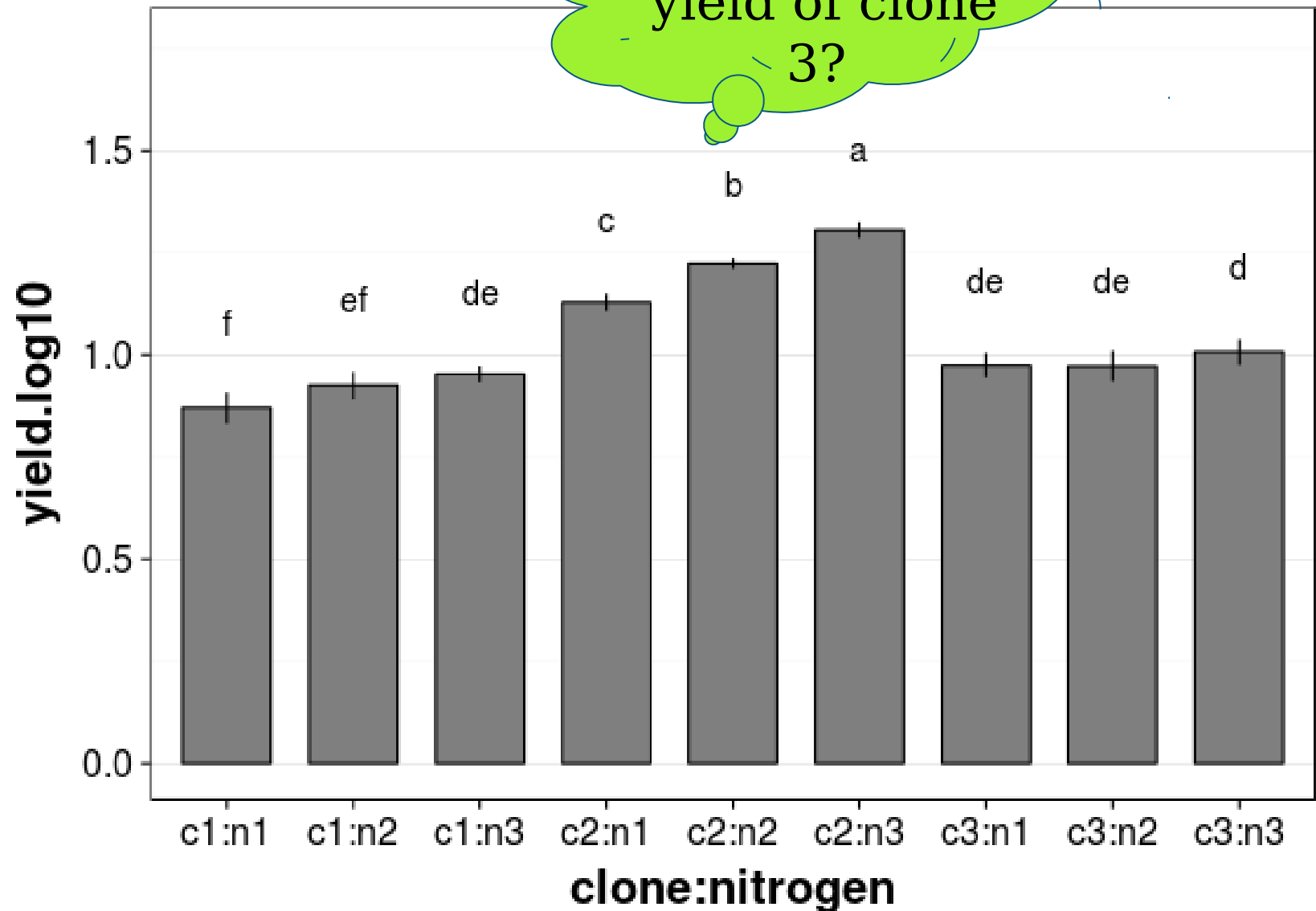
Post hoc tests and figures

Note that the y-axis is in log scale: $\log(\text{kg ha}^{-1})$



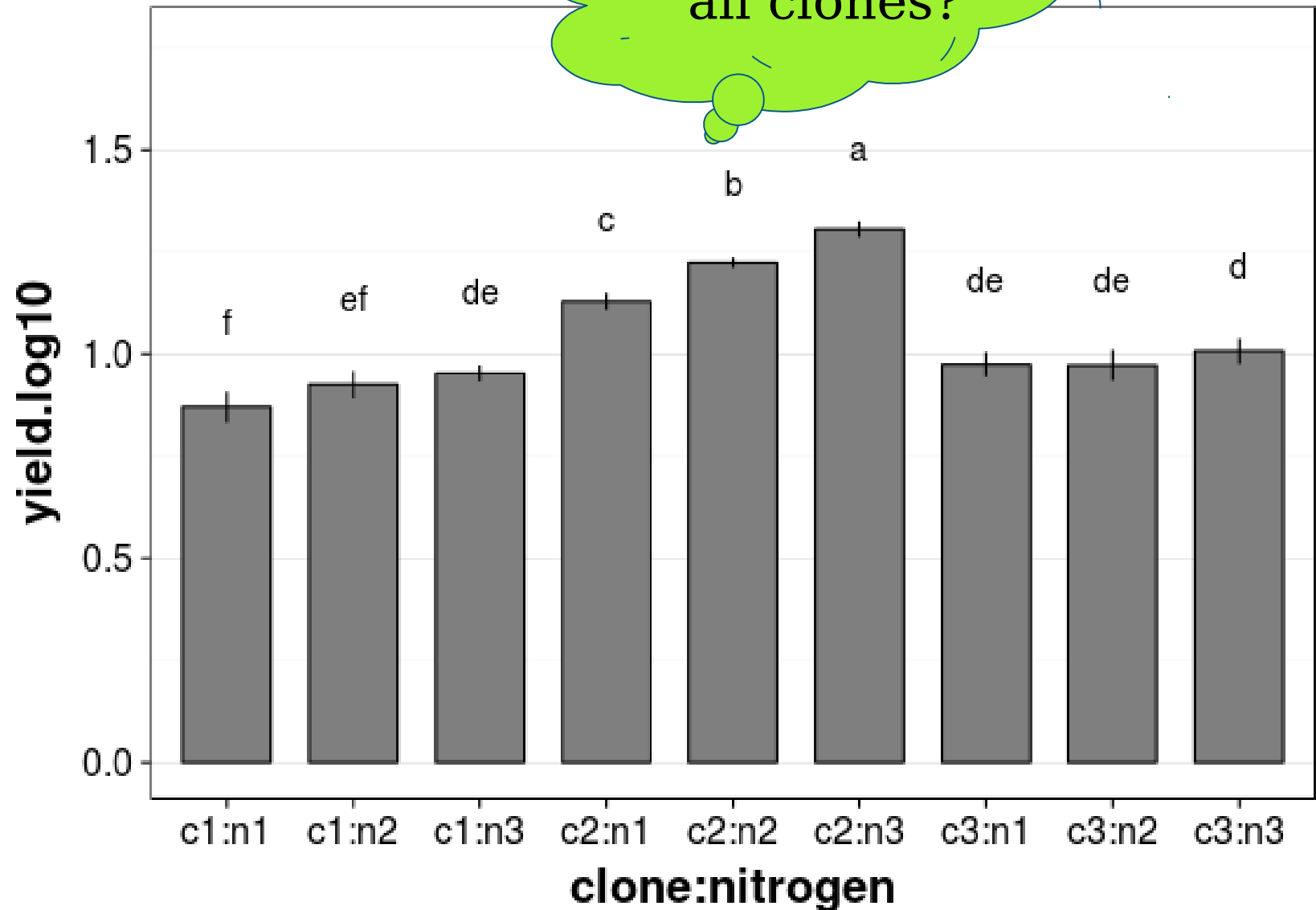
Post hoc tests and figures

How does
yield of clone
2 compare to
yield of clone
3?

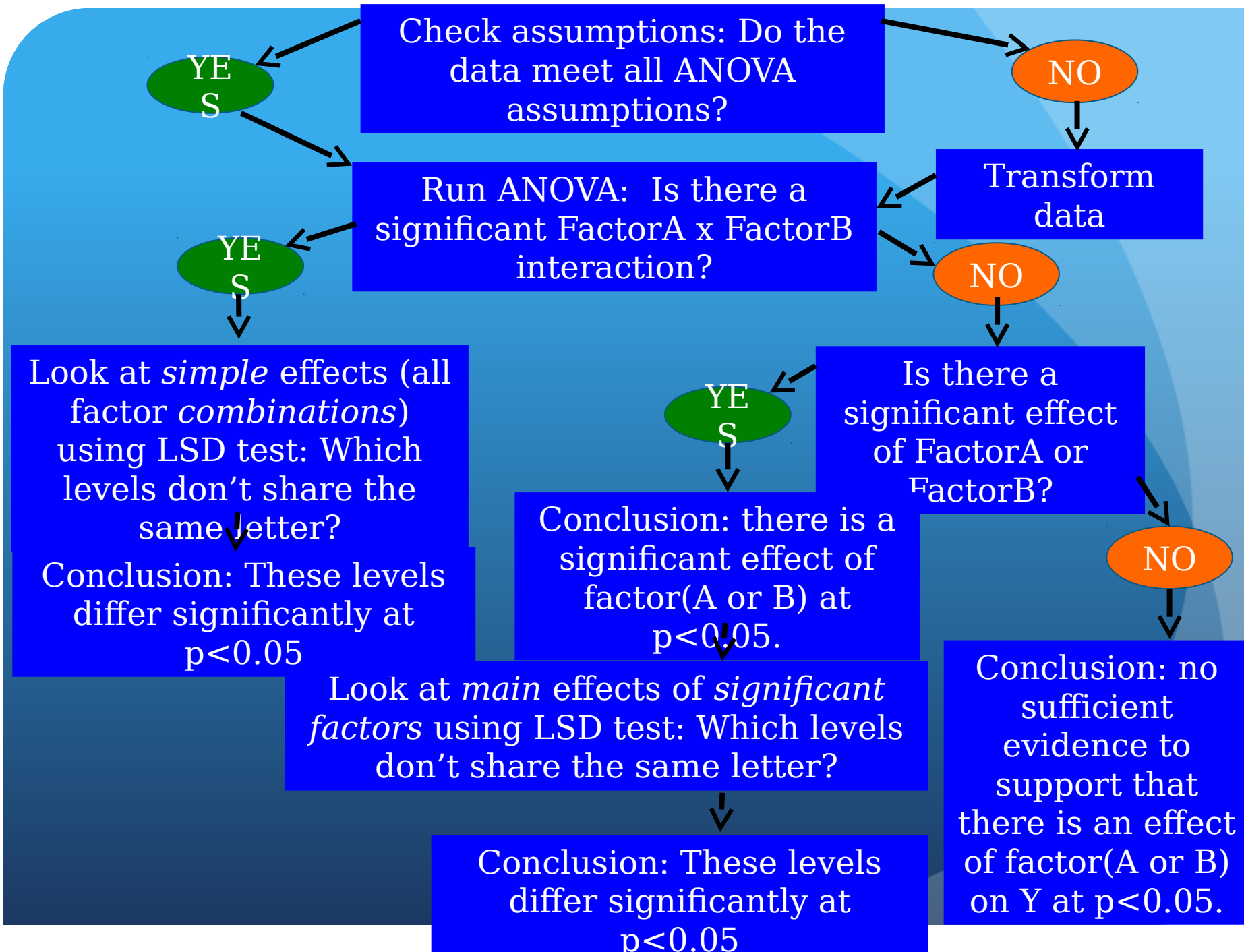


Post hoc tests and figures

Does nitrogen
affect yield of
all clones?



Next, a quick review...



(1) If the p-value = 0.03 for the Shapiro-Wilk test, can we proceed with the ANOVA?

Why or why not?

Example dataset

<i>Independent variable:</i> N fertilizer rate (kg nitrogen ha ⁻¹)	<i>Bloc</i> k	<i>Dependent variable:</i> Yield (kg grain ha ⁻¹)
0	1	7.1
50	1	9
100	1	9.5
0	2	6.2
50	2	8.8
100	2	9.2

(2) If the app determines the exponent a for power transformation of the dataset below:
 $a = 0.2$. What are the power-transformed values?

N fertilizer rate (kg nitrogen ha ⁻¹)	Block	Yield (kg grain ha ⁻¹)	Power-transformed Yield
0	1	7.1	
50	1	9	
100	1	9.5	
0	2	6.2	
50	2	8.8	
100	2	9.2	

(2) If the app determines the exponent a for power transformation of the dataset below:
 $a = 0.2$. What are the power-transformed values?

N fertilizer rate (kg nitrogen ha ⁻¹)	Bloc k	Yield (kg grain ha ⁻¹)	Power- transformed Yield: $(Y_{ij})^{0.2}$
0	1	7.1	$(7.1)^{0.2} = 1.48$
50	1	9	1.55
100	1	9.5	1.57
0	2	6.2	1.44
50	2	8.8	1.54
100	2	9.2	1.56

(3) What are the units of the power-transformed data?

N fertilizer rate (kg nitrogen ha ⁻¹)	Block	Yield (kg grain ha ⁻¹)	Power-transformed Yield
0	1	7.1	1.48
50	1	9	1.55
100	1	9.5	1.57
0	2	6.2	1.44
50	2	8.8	1.54
100	2	9.2	1.56

(3) What are the units of the power-transformed data?

N fertilizer rate (kg nitrogen ha ⁻¹)	Bloc k	Yield (kg grain ha ⁻¹)	Power- transformed Yield (kg grain ha ⁻¹) ^{0.2}
0	1	7.1	1.48
50	1	9	1.55
100	1	9.5	1.57
0	2	6.2	1.44
50	2	8.8	1.54
100	2	9.2	1.56

Presenting results from analyses of transformed data

When your data has been transformed, the units change too!

Original units	Log transformation	Log-transformed units
Ex. Yield (ton/ha)		Yield (log(ton/ha))

Tests of significance (e.g. F tests in ANOVA table) and multiple mean comparisons (e.g. LSD) are performed on the *transformed* data.

However, it's common practice to present results using the *original* units. This can be done in two ways: (1) present original means, or (2) back-transform means

IMPORTANT: When reporting results using original units, make sure to state that *the statistical analyses were performed on (power- or log-) transformed data.*

Example: If this was the app output for the LSD test, how would you formally present the results?

Fertilizer N Treatment (kg N ha ⁻¹)	Power-Transformed Mean Yield (kg grain ha ⁻¹) ^{0.2}	LSD Groupings of Power-Transformed Means
0	1.46	a
50	1.55	b
100	1.57	b

Two ways to present results in original units:

1. Back-transformation

a. Convert power-transformed mean yield (Y_{trans}) back to the original units (kg ha^{-1}) by: $(Y_{\text{trans}})^{(1/a)} = Y$

b. Convert log-transformed mean yield (Y_{trans}) back to the original units (kg ha^{-1}) by: $10^{Y_{\text{trans}}} = Y$

- If data were *multiplied* by constant 'x' *before* doing the log-transformation, first *divide* each transformed mean by x: $10^{(Y_{\text{trans}}/x)} = Y$
- If data were *added* to a constant 'x' *before* doing the log-transformation, first *subtract* each transformed mean by x: $10^{(Y_{\text{trans}} - x)} = Y$

Example: If this were your post-hoc output for the LSD test, how would you formally present the results?

Fertilizer N Treatment (kg N ha ⁻¹)	Power-Transformed Mean Yield (kg grain ha ⁻¹) ^{0.2}	LSD Groupings of Power-Transformed Means
0	1.46	a
50	1.55	b
100	1.57	b

Option 1:

Convert power-transformed mean yield (kg ha^{-1})^{1/a} back to the original units (kg ha^{-1}) by: $(Y_{i,\text{trans}})^{(1/a)} = Y_i$

Note: these are weighted means \neq original means

Fertilizer N Treatment (kg N ha^{-1})	Power-transformed Mean Yield (kg grain ha^{-1}) ^{0.2}	Back-Transformed Mean Yield (kg ha^{-1})	LSD Groupings of Power-Transformed Means
0	1.46	$(1.46)^{1/a}$	a
50	1.55	$(1.55)^{1/a}$	b
100	1.57	$(1.57)^{1/a}$	b

Option 1:

Convert power-transformed mean yield (kg ha^{-1})¹⁾^a back to the original units (kg ha^{-1}) by: $(Y_{i,\text{trans}})^{(1/a)} = Y_i$

Fertilizer N Treatment (kg N ha^{-1})	Power-transformed Mean Yield (kg grain ha^{-1}) ^{0.2}	<i>Back-Transformed</i> Mean Yield (kg ha^{-1})	LSD Groupings of Power-Transformed Means
0	1.46	$(1.46)^{1/0.2}$	a
50	1.55	$(1.55)^{1/0.2}$	b
100	1.57	$(1.57)^{1/0.2}$	b

Option 1:

Convert power-transformed mean yield (kg ha^{-1})^{*a*} back to the original units (kg ha^{-1}) by: $(Y_{i,\text{trans}})^{(1/a)} = Y_i$

Fertilizer N Treatment (kg N ha^{-1})	Power-transformed Mean Yield (kg grain ha^{-1}) ^{0.2}	<i>Back-Transformed</i> Mean Yield (kg ha^{-1})	LSD Groupings of Power-Transformed Means
0	1.46	6.63	a
50	1.55	8.80	b
100	1.57	9.39	b

A formatted table may look something like this:

Table 1. Yield response to N fertilizer additions.

Fertilizer N Rate (kg N ha ⁻¹)	Mean Yield ¹ (kg ha ⁻¹)
0	6.63a ²
50	8.80b
100	9.39b

¹Least square means after transformation back to the original measurement scale.

²Mean yields not sharing the same letter are significantly different at $p < 0.05$ based on Fisher's protected least significant difference test applied to power-transformed data.

Option 1:

Convert power-transformed mean yield (Y_{trans}) back to the original units (kg ha^{-1}) by:

$$(Y_{trans})^{(1/a)} = Y$$

Fertilizer N Treatment (kg N ha⁻¹)	Power-transformed Mean Yield (kg ha⁻¹)^{0.2}	<i>Back-Transformed</i> Mean Yield (kg ha⁻¹)	LSD Groupings Of Power-Transformed Means
0	1.46	6.63	a
50	1.55	8.80	b
100	1.57	9.39	b

Option 1:

Convert power-transformed mean yield (Y_{trans}) back to the original units (kg ha^{-1}) by: $(Y_{trans})^{(1/a)} = Y$

Option 2:

Calculate means of original dataset.

Fertilizer N Treatment (kg N ha⁻¹)	Power-transformed Mean Yield (kg ha⁻¹)^{0.2}	Back-Transformed Mean Yield (kg ha⁻¹)	LSD Groupings Of Power-Transformed Means
0	1.46	6.63	a
50	1.55	8.80	b
100	1.57	9.39	b

Option 1:

Convert power-transformed mean yield (Y_{trans}) back to the original units (kg ha^{-1}) by: $(Y_{trans})^{(1/a)} = Y$

Option 2:

Calculate means of original dataset.

Fertilizer N rate (kg N ha^{-1})	Block	Yield (kg ha^{-1})	Original dataset	Back-Transformed Mean Yield (kg ha^{-1})	Mean Yield (kg ha^{-1})	LSD Groupings of Power-Transformed Means
0	1	7.1	Original dataset			
50	1	9				
100	1	9.5				
0	2	6.2				
50	2	8.8				
100	2	9.2				
			Fertilizer N Treatment (kg N ha^{-1})			
			0	6.63		a
			50	8.80		b
			100	9.39		b

Option 1:

Convert power-transformed mean yield (Y_{trans}) back to the original units (kg ha^{-1}) by: $(Y_{trans})^{(1/a)} = Y$

Option 2:

Calculate means of original dataset.

Fertilizer N rate (kg N ha^{-1})	Block	Yield (kg ha^{-1})
0	1	7.1
50	1	9
100	1	9.5
0	2	6.2
50	2	8.8
100	2	9.2

Fertilizer N Treatment (kg N ha^{-1})	Back-Transformed Mean Yield (kg ha^{-1})	Mean Yield (kg ha^{-1})	LSD Groupings Of Power-Transformed Means
0	6.63	$(7.1 + 6.2) / 2$	a
50	8.80		b
100	9.39		b

Option 1:

Convert power-transformed mean yield (Y_{trans}) back to the original units (kg ha^{-1}) by: $(Y_{trans})^{(1/a)} = Y$

Option 2:

Calculate means of original dataset.

Fertilizer N rate (kg N ha^{-1})	Block	Yield (kg ha^{-1})
0	1	7.1
50	1	9
100	1	9.5
0	2	6.2
50	2	8.8
100	2	9.2

Fertilizer N Treatment (kg N ha^{-1})	Back-Transformed Mean Yield (kg ha^{-1})	Mean Yield (kg ha^{-1})	LSD Groupings of Power-Transformed Means
0	6.63	6.65	a
50	8.80	8.90	b
100	9.39	9.35	b

A formatted table may look something like this:

Table 1. Yield response to N fertilizer additions.

Fertilizer N Rate (kg N ha ⁻¹)	Mean Yield (kg ha ⁻¹)
0	6.7a ¹
50	8.9b
100	9.4b

¹Mean yields not sharing the same letter are significantly different at $p < 0.05$ based on Fisher's protected least significant difference test applied to power transformed data.

Section 6 Learning Objectives

Have basic understanding of:

1. Split-plot experimental designs

- What are they, and why are they used?
- What are the linear models for a split-plot CRD and split-plot RCBD?
- How to use the app to analyze both of these types of split-plot designs?

Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

- If only 2 irrigation methods could be established per field due to practical constraints (e.g. risk of seepage), how could you design an experiment to investigate these research questions if you had 3 fields?

Field 1

Field 2

Field 3



I. Factorial experiment in CRD = ONE stage randomization.

I. Factorial experiment in CRD = ONE stage randomization.

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

I. Factorial experiment in CRD = ONE stage randomization.

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

II.
b'

I. Factorial experiment in CRD = ONE stage randomization.

II.
b'

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2
A1- B2	A3- B2	A3- B1						
A2- B2	A2- B1	A1- B1						

I. Factorial experiment in CRD = ONE stage randomization.

II.
b'

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2
A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2			
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1			

I. Factorial experiment in CRD = ONE stage randomization.

II.
b'

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2
A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2	A3- B2	A1- B1	A3- B1
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1	A2- B1	A2- B2	A1- B2

I. Factorial experiment in CRD = ONE stage randomization.

II.
b'

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2	A3- B2	A1- B1	A3- B1
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1	A2- B1	A2- B2	A1- B2

I. Split-plot CRD

1. Factor A randomly assigned to Main-Plots

--	--	--	--	--	--	--	--	--

I. Factorial experiment in CRD = ONE stage randomization.

II.
b'

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2	A3- B2	A1- B1	A3- B1
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1	A2- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots

2. Factor B randomly assigned to Sub-Plots *within* Main Plots

A2	A3	A3	A2	A1	A1	A3	A2	A1
----	----	----	----	----	----	----	----	----

I. Factorial experiment in CRD = ONE stage randomization.

II.
b)

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2	A3- B2	A1- B1	A3- B1
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1	A2- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots

2. Factor B randomly assigned to Sub-Plots *within* Main Plots

A2- B1	A3- B2	A3- B2	A2- B2	A1- B1	A1- B2	A3- B2	A2- B1	A1- B1
A2- B2	A3- B1	A3- B1	A2- B1	A1- B2	A1- B1	A3- B1	A2- B2	A1- B2

I. Factorial experiment in CRD = ONE stage randomization.

II.
b'

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

III.

A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2	A3- B2	A1- B1	A3- B1
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1	A2- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots

2. Factor B randomly assigned to Sub-Plots *within* Main Plots

IV.
per block.

A2- B1	A3- B2	A3- B2	A2- B2	A1- B1	A1- B2	A3- B2	A2- B1	A1- B1
A2- B2	A3- B1	A3- B1	A2- B1	A1- B2	A1- B1	A3- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots *within* blocks

--	--	--	--	--	--	--	--	--

I. Factorial experiment in CRD = ONE stage randomization.

II.
b

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

III.

A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2	A3- B2	A1- B1	A3- B1
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1	A2- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots

2. Factor B randomly assigned to Sub-Plots *within* Main Plots

IV.
per block.

A2- B1	A3- B2	A3- B2	A2- B2	A1- B1	A1- B2	A3- B2	A2- B1	A1- B1
A2- B2	A3- B1	A3- B1	A2- B1	A1- B2	A1- B1	A3- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots *within* blocks

A1	A3	A2						
----	----	----	--	--	--	--	--	--

I. Factorial experiment in CRD = ONE stage randomization.

II.
b'

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

III.

A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2	A3- B2	A1- B1	A3- B1
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1	A2- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots

2. Factor B randomly assigned to Sub-Plots *within* Main Plots

IV.
per block.

A2- B1	A3- B2	A3- B2	A2- B2	A1- B1	A1- B2	A3- B2	A2- B1	A1- B1
A2- B2	A3- B1	A3- B1	A2- B1	A1- B2	A1- B1	A3- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots *within* blocks.

A1	A3	A2	A2	A1	A3			
----	----	----	----	----	----	--	--	--

I. Factorial experiment in CRD = ONE stage randomization.

II.
b

A1- B1	A2- B1	A2- B2	A3- B2	A1- B2	A3- B1	A2- B1	A2- B2	A2- B2
A1- B2	A3- B1	A1- B1	A1- B2	A3- B1	A1- B1	A3- B2	A2- B1	A3- B2

III.

A1- B2	A3- B2	A3- B1	A2- B2	A2- B1	A3- B2	A3- B2	A1- B1	A3- B1
A2- B2	A2- B1	A1- B1	A3- B1	A1- B2	A1- B1	A2- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots

2. Factor B randomly assigned to Sub-Plots *within* Main Plots

IV.
per block.

A2- B1	A3- B2	A3- B2	A2- B2	A1- B1	A1- B2	A3- B2	A2- B1	A1- B1
A2- B2	A3- B1	A3- B1	A2- B1	A1- B2	A1- B1	A3- B1	A2- B2	A1- B2

1. Factor A randomly assigned to Main-Plots *within* blocks.

A1	A3	A2	A2	A1	A3	A2	A3	A1
----	----	----	----	----	----	----	----	----

I. Factorial experiment in CRD = ONE stage randomization.

II.
b'

A1-B1	A2-B1	A2-B2	A3-B2	A1-B2	A3-B1	A2-B1	A2-B2	A2-B2
A1-B2	A3-B1	A1-B1	A1-B2	A3-B1	A1-B1	A3-B2	A2-B1	A3-B2
A1-B2	A3-B2	A3-B1	A2-B2	A2-B1	A3-B2	A3-B2	A1-B1	A3-B1
A2-B2	A2-B1	A1-B1	A3-B1	A1-B2	A1-B1	A2-B1	A2-B2	A1-B2

1. Factor A randomly assigned to Main-Plots

2. Factor B randomly assigned to Sub-Plots *within* Main Plots

IV. *per block.*

A2-B1	A3-B2	A3-B2	A2-B2	A1-B1	A1-B2	A3-B2	A2-B1	A1-B1
A2-B2	A3-B1	A3-B1	A2-B1	A1-B2	A1-B1	A3-B1	A2-B2	A1-B2

1. Factor A randomly assigned to Main-Plots *within* blocks

A1-B2	A3-B2	A2-B1	A2-B2	A1-B1	A3-B2	A2-B2	A3-B1	A1-B1
A1-B1	A3-B1	A2-B2	A2-B1	A1-B2	A3-B1	A2-B1	A3-B2	A1-B2

Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

- If only 2 irrigation methods could be established per field due to practical constraints (e.g. risk of seepage), how could you design an experiment to investigate these research questions if you had 3 fields?

1. Experimental design? **Split-plot design with RCBD**

main-plots

Field 1

Field 2

Field 3

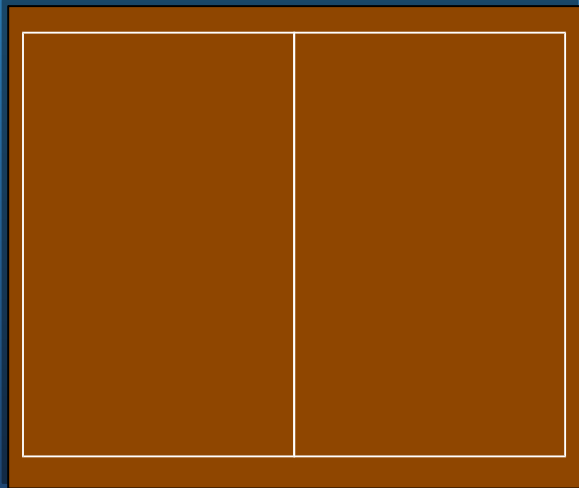
Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

- If only 2 irrigation methods could be established per field due to practical constraints (e.g. risk of seepage), how could you design an experiment to investigate these research questions if you had 3 fields?

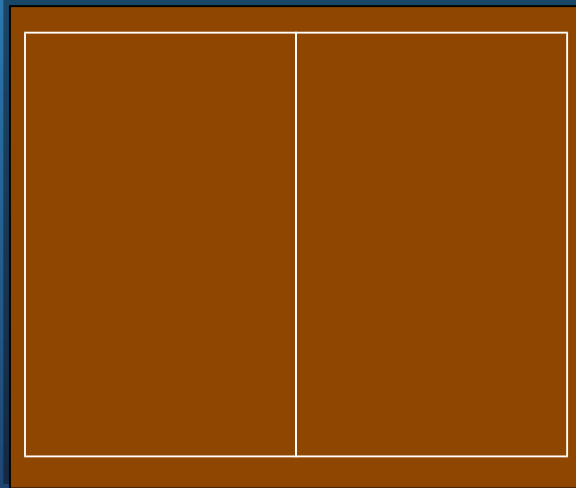
1. Experimental design? **Split-plot design with RCBD**
2. Dependent variable(s)? **main-plots
Methane emissions**
3. Independent variable(s)? **Irrigation method and N rate**
4. Main-plot? **Irrigation method**
5. Sub-plot? **N rate**
6. Blocking variable? **Field**

Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

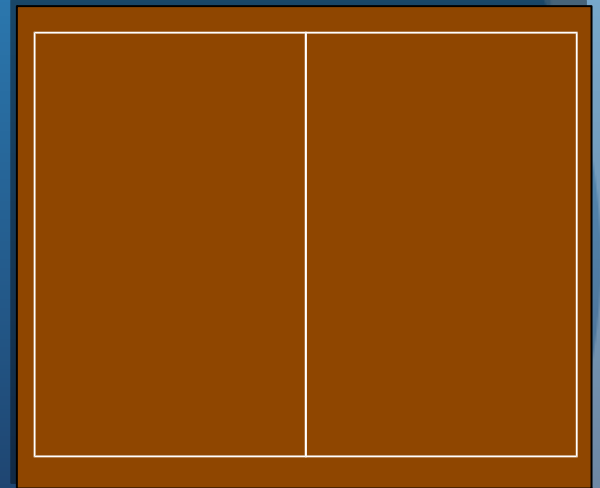
Step 1: Randomize main-plots *within* blocks



Field 1



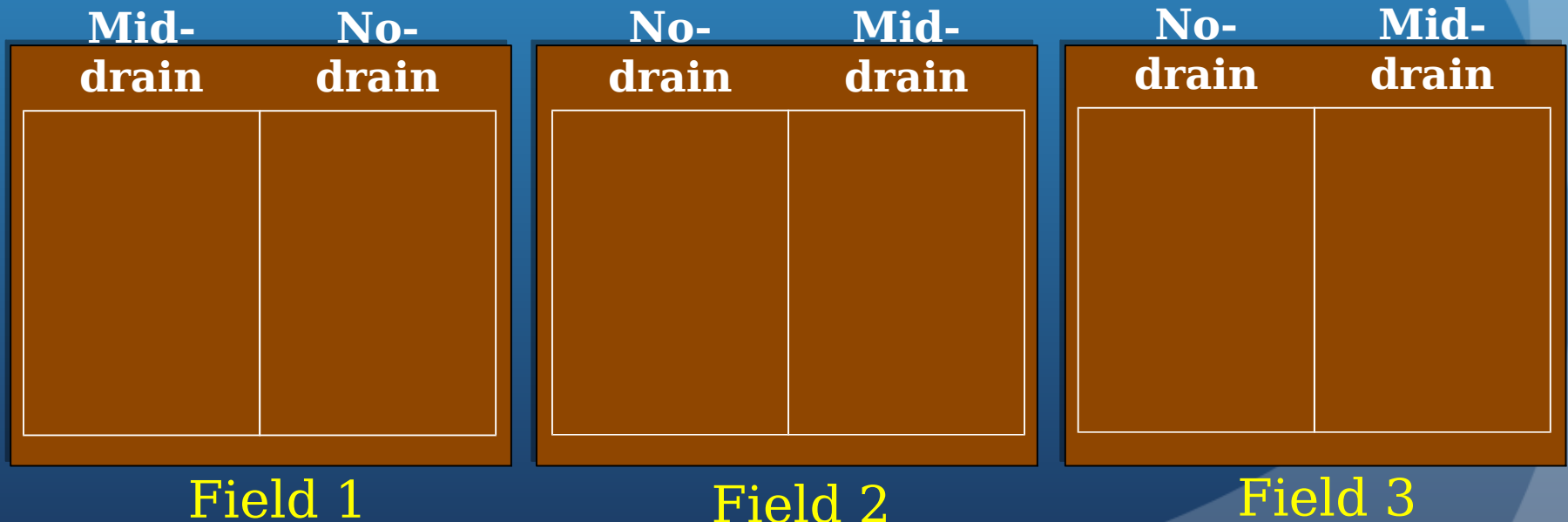
Field 2



Field 3

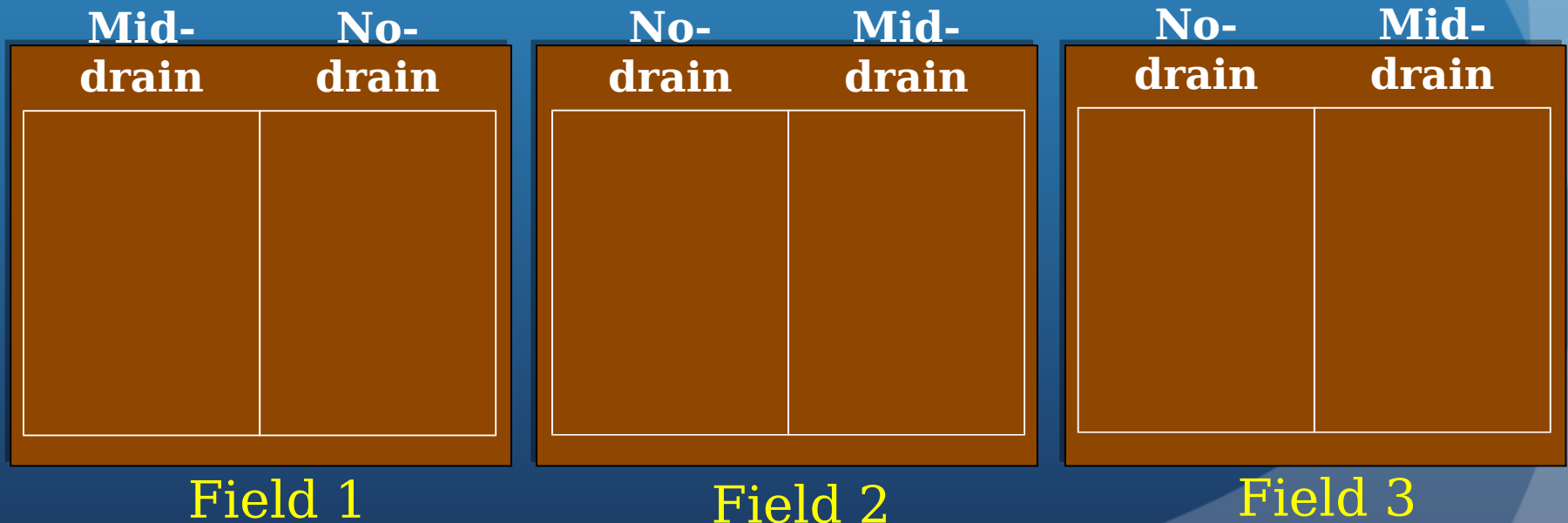
Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

Step 1: Randomize main-plots *within* blocks



Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

Step 2: Randomize sub-plots *within* main-plots



Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

Step 2: Randomize sub-plots *within* main-plots



Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

Step 2: Randomize sub-plots *within* main-plots

Mid-drain	No-drain	No-drain	Mid-drain	No-drain	Mid-drain
80N					
0N					
140N					
Field 1		Field 2		Field 3	

Research question: In a rice system, does the type of irrigation method (mid-season drain, no drain) and rate of nitrogen fertilizer application (0, 80, 140 kg N ha⁻¹) affect methane emissions?

Step 2: Randomize sub-plots *within* main-plots

Mid-drain	No-drain
80N	0N
0N	80N
140N	140N

Field 1

No-drain	Mid-drain
0N	80N
140N	140N
80N	0N

Field 2

No-drain	Mid-drain
140N	80N
80N	140N
0N	0N

Field 3

Split-Plot experimental designs

Why
use
them?

- Good for experiments where it's difficult to change one factor at a small scale
- Less precision in estimating main-plot effect, but more precision for estimating sub-plot effect and sub-plot x main-plot interaction

III. Factorial experiment in **split-plot CRD** = TWO stage randomization.

1. Factor A randomly assigned to Main-Plots
2. Factor B randomly assigned to Sub-Plots *within* Main Plots

A2-B1	A3-B2	A3-B2	A2-B2	A1-B1	A1-B2	A3-B2	A2-B1	A1-B1
A2-B2	A3-B1	A3-B1	A2-B1	A1-B2	A1-B1	A3-B1	A2-B2	A1-B2

IV. **split-plot RCBD** per block.

1. Factor A randomly assigned to Main-Plots *within* blocks

A1-B2	A3-B2	A2-B1	A2-B2	A1-B1	A3-B2	A2-B2	A3-B1	A1-B1
A1-B1	A3-B1	A2-B2	A2-B1	A1-B2	A3-B1	A2-B1	A3-B2	A1-B2

Split-Plot experimental designs

- 2 stages of randomization and 2 error terms
 - Make sure to choose “split-plot..” from experimental design in app!
 - ****Important note for split-plot CRD**** Your .csv data file must include a column for ‘replication’ identification (e.g. Factor 3)
- III. Factorial experiment in **split-plot CRD** = TWO stage randomization.
1. Factor A randomly assigned to Main-Plots
 2. Factor B randomly assigned to Sub-Plots *within* Main Plots

A2-B1	A3-B2	A3-B2	A2-B2	A1-B1	A1-B2	A3-B2	A2-B1	A1-B1
A2-B2	A3-B1	A3-B1	A2-B1	A1-B2	A1-B1	A3-B1	A2-B2	A1-B2

IV. **split-plot RCBD** per block.

1. Factor A randomly assigned to Main-Plots *within* blocks

A1-B2	A3-B2	A2-B1	A2-B2	A1-B1	A3-B2	A2-B2	A3-B1	A1-B1
A1-B1	A3-B1	A2-B2	A2-B1	A1-B2	A3-B1	A2-B1	A3-B2	A1-B2

Split-Plot experimental designs

- Linear model for split-plot with CRD for main-plots:

$$Y_{ijk} = \mu + \alpha_i + (\alpha*\gamma)_{ik} + \beta_j + (\alpha*\beta)_{ij} + \varepsilon_{ijk}$$

Main-plot effect (A)

Error A = random A x Replication interaction

Sub-plot effect (B)

AxB interaction

Error B, Error A*B

Split-Plot experimental designs

- Linear model for split-plot with RCBD for main-plots:

$$Y_{ijk} = \mu + \nu_k + \alpha_i + (\alpha*\nu)_{ik} + \beta_j + (\alpha*\beta)_{ij}$$

random
Block
effect

Main-
plot
effect
(A)

Error A=
random
AxBBlock
interaction

Sub-
plot
effect
(B)

AxB
interaction

Error
B,
Error
A*B

Let's practice using the app to analyze a new dataset¹!

Oat field trial measuring yield of 3 varieties at 4 nitrogen fertilizer application rates.

- Varieties are randomized within blocks, and 4 rates of fertilizer are randomized within varieties

Split-plot RCBD

- Experimental design?
- Main-plot?
- Sub-plot?
- Block?

¹Source: Pinheiro, J. C. and Bates, D. M. (2000), Mixed-Effects Models in S and S-PLUS, Springer, New York. (Appendix A.15)
Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. (4th ed), Springer, New York.

First let's create
the
experimental
layout

Block 1

Block 2

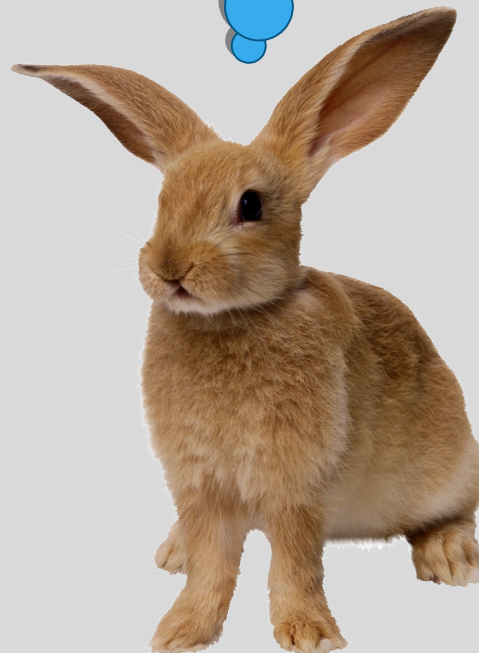
Block 3

Block 4

Block 5

Block 6

What is the
first step in
assigning
treatments?



Block 1

Block 2

Block 3

Block 4

Block 5

Block 6

Golden Rain



Marvellous



Victory



Block 1

Block 2

Block 3

Block 4

Block 5

Block 6

Randomize
varieties (main-
plot) within
blocks

Golden Rain



Marvellous



Golden Rain

Victory



Victory



Marvellous



Golden Rain



Victory



Victory



Marvellous



Golden Rain



Golden Rain



Victory



Marvellous



Victory



Golden Rain



Marvellous

Block 1

Block 2

Block 3

Block 4

Block 5

Block 6

What is the next step in assigning treatments?





Golden Rain	0	0.4	0.2	0.6	Block 1
Marvellous					
Golden Rain					Block 2
Victory					
Victory					Block 3
Marvellous					
Golden Rain					Block 4
Victory					
Victory					Block 5
Marvellous					
Golden Rain					Block 6
Golden Rain					
Victory					Block 7
Marvellous					
Golden Rain					Block 8
Marvellous					

Golden Rain	0	0.4	0.2	0.6	Block 1
Marvellous	0.6	0	0.4	0.2	
Golden Rain					Block 2
Victory					
Victory					Block 3
Marvellous					
Golden Rain					Block 4
Victory					
Victory					Block 5
Marvellous					
Golden Rain					Block 6
Golden Rain					
Victory					Block 7
Marvellous					
Golden Rain					Block 8
Marvellous					

Golden Rain	0	0.4	0.2	0.6	Block 1
Marvellous	0.6	0	0.4	0.2	
Golden Rain	0.4	0	0.2	0.6	
Victory					Block 2
Victory					
Marvellous					Block 3
Golden Rain					
Victory					Block 4
Victory					
Marvellous					
Golden Rain					Block 5
Golden Rain					
Victory					Block 6
Marvellous					
Victory					
Golden Rain					Block 6
Marvellous					

Golden Rain	0	0.4	0.2	0.6	Block 1
Marvellous	0.6	0	0.4	0.2	
Golden Rain	0.4	0	0.2	0.6	
Victory	0	0.4	0.6	0.2	Block 2
Victory	0	0.2	0.6	0.4	
Victory	0.4	0.2	0	0.6	
Marvellous	0.4	0.6	0.2	0	Block 3
Golden Rain	0	0.6	0.2	0.4	
Golden Rain	0.6	0	0.4	0.2	
Victory	0	0.4	0.6	0.2	Block 4
Victory	0.2	0.4	0	0.6	
Marvellous	0.4	0	0.6	0.2	
Golden Rain	0.6	0.4	0.2	0	Block 5
Golden Rain	0	0.2	0.4	0.6	
Victory	0.4	0.2	0.6	0	
Marvellous	0.6	0.2	0.4	0	Block 6
Golden Rain	0.6	0	0.4	0.2	
Marvellous	0.6	0.4	0.2	0	

Load oat data file to app

`'oats-split-plot-rcbd.csv'`

Section 6 Learning Objectives

You should now have a basic understanding of:

1. Split-plot designs

- How to randomize treatments in 2 stages for CRD and RCBD main-plots
- Benefit and cost of this experimental design
- Statistical models for split-plot design CRD and split-plot design RCBD
- Using the app to analyze both of these types of split-plot designs

Section 7 Learning Objectives

Have basic understanding of:

1. Multi-location studies
2. What to do when you have missing data

Experimental designs that we've discussed:

1. Complete Randomized Design (CRD) (*1 or 2 factors*)
2. Randomized Complete Block Design (RCBD) (*1 or 2 factors*)
3. Split-plot Design (*CRD or RCBD main-plots with 2*

Multi-site experiments

Experiments are often repeated at multiple locations, which may be analyzed as a single dataset using the app (mixed effects model). In the app, the variable you select as 'site' is designated as having a *random* effect, allowing for conclusions of the treatment effect to be extended to other representative sites (e.g. particular growing region for rice).

Be aware that this approach makes the following assumptions:

1. You are *not* particularly interested in differences among sites, or in these sites alone.
2. Sites were *randomly* selected from many potential sites of interest.

Linear model for RCBD experiment repeated at multiple locations:

$$Y_{ijk} = \mu + \tau_i + \nu(l)_{jk} + l_k + \varepsilon_{ijk}$$

- Y_{ijk} = observation of the i^{th} treatment level in the j^{th} block at the k^{th} location
- τ_i = effect of the i^{th} treatment
- $\nu(l)_{jk}$ = *random* effect of the j^{th} block at the k^{th} location
- l_k = (*random*) effect of the k^{th} location
- ε_{ijk} = random error

Linear model for RCBD experiment repeated at multiple locations:

$$Y_{ijk} = \mu + \tau_i + \nu(l)_{jk} + l_k + (\tau l)_{ij} + \varepsilon_{ijk}$$

- Y_{ijk} = observation of the i^{th} treatment level in the j^{th} block at the k^{th} location
- τ_i = effect of the i^{th} treatment
- $\nu(l)_{jk}$ = *random* effect of the j^{th} block at the k^{th} location
- l_k = *random* effect of the k^{th} location
- $(\tau l)_{ik}$ = *random* effect of ij^{th} location x treatment interaction (excluding it assumes that treatment effects are similar across sites)
- ε_{ijk} = random error

Missing data