LGN 5822 - Biometrical Genetics

L10 – Latin Square

Michele Jorge Silva Siqueira

2023

Introduction

- In some situations, there may be more than one source of heterogeneity among experimental units

Introduction

- The Latin Square Design is an appropriate design for environments heterogeneous experiments

- In the Latin Square Design, in addition to the principles of **repetition and randomization**, the principle of **local control is used twice** to control the effect of two factors

  - More control of experimental units

Introduction

- To control this variability, it is necessary to **divide the experimental units into homogeneous blocks of experimental units** in relation to each controlled factor

Latin Square

Design Characterization

- If there are two gradients, in perpendicular directions, blocking factors can be used to simultaneously control both sources of variation

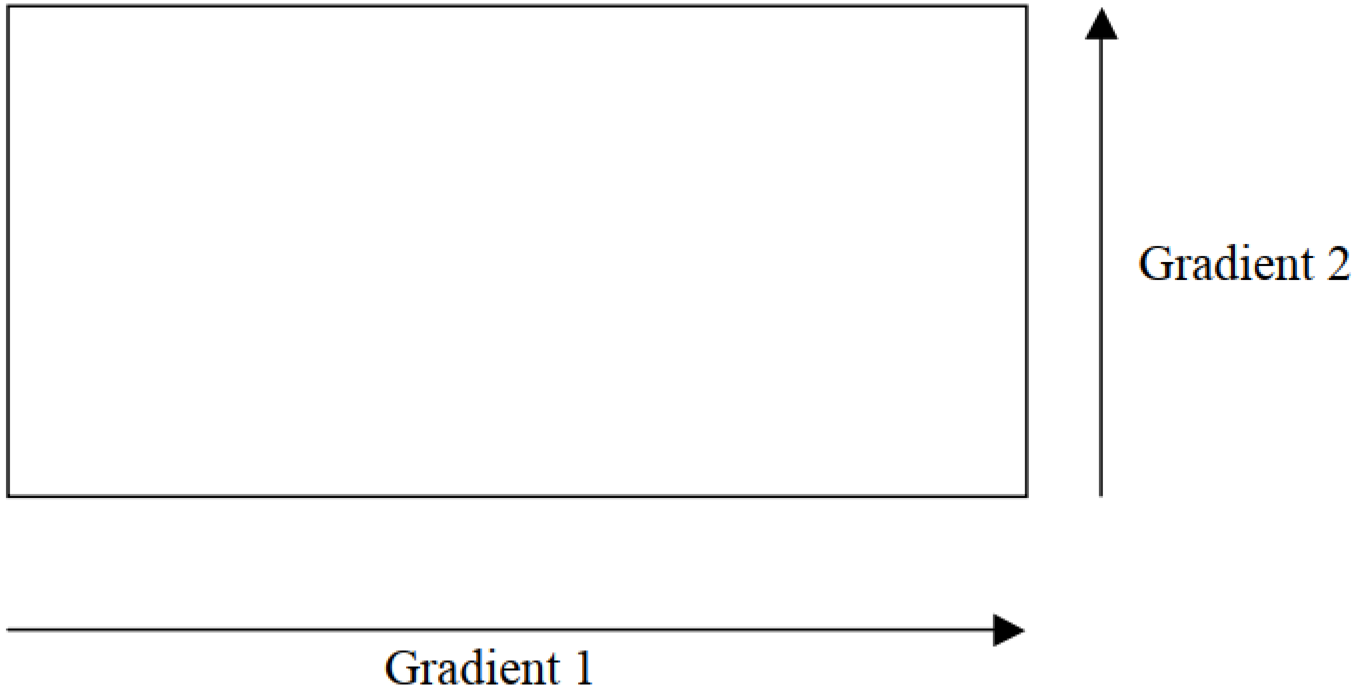
  - Physical gradients in the field plot
  - Other (orthogonal) experimental sources of variation
  - Treatments of interest

When there are two main sources of variation that can be controlled

Latin Square

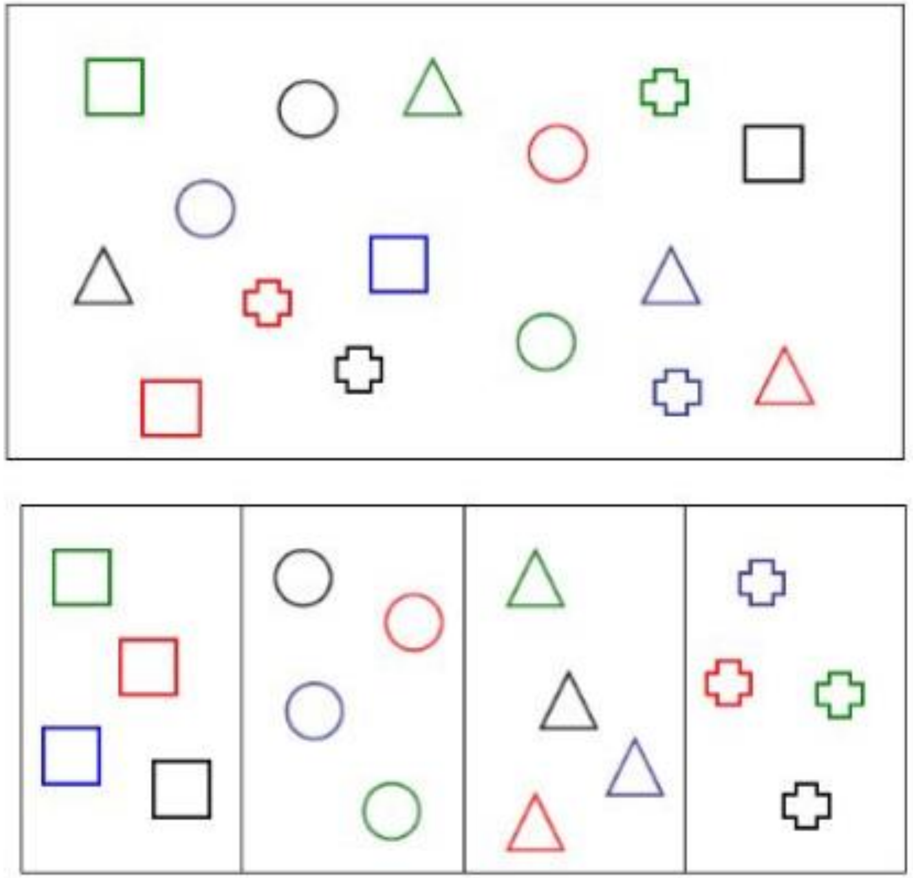
Example

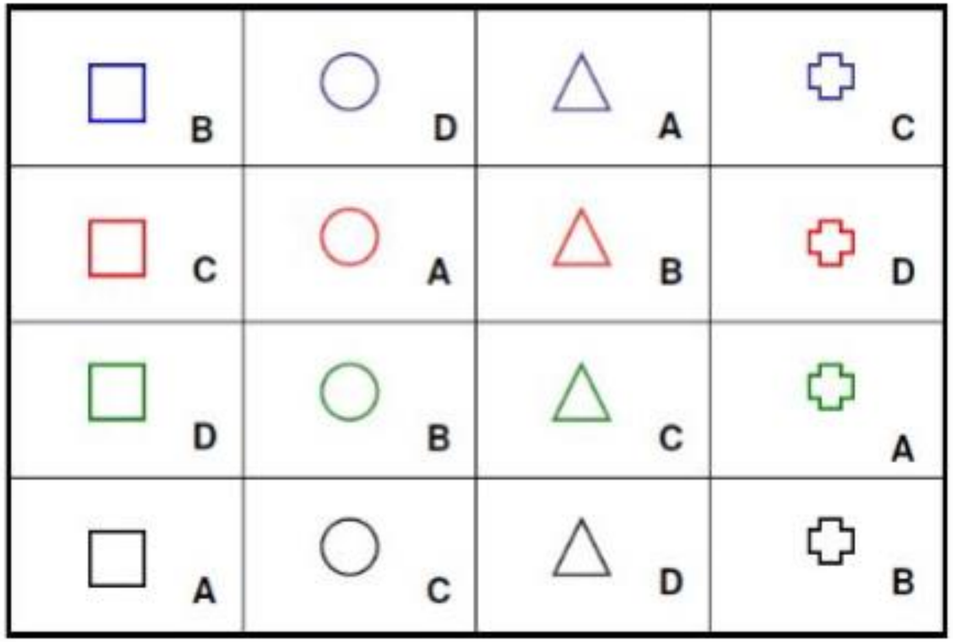
- Field trials in which the experimental error has two fertility gradients running perpendicular each other



Latin Square

Experimental Design Layout





Latin Square

Experimental Design Layout

- Consider a competition experiment with 4 sugarcane varieties in which the experimental area presents a soil fertility gradient in two directions.

	Column 1	Column 2	Column 3	Column 4
Row 1	T1	T2	T3	T4
Row 2	T2	T3	T4	T1
Row 3	T3	T4	T1	T2
Row 4	T4	T1	T2	T3

Latin Square

Description of the Design

- With the Latin Square design you are able to control variation in two directions
- Treatments are arranged in rows and columns
- Each row contains every treatment
- Each column contains every treatment
- The total number of experimental units is thus t^2 (t t is the number of treatments)
  - This is a ***square** design*

Latin Square

Description of the Design

- The number of treatments is equal to the number of repetitions

- This design is advisable when the number of treatments varies between 3 and 10

- But, for 3 and 4 treatments, only when the experiment can be repeated in several Latin squares

- It has **more efficient local control** than the randomized block design (horizontal and vertical control)

Latin Square

Casualization in the Latin Square Design

- The treatments are distributed within the rows, so that each column also contains all the treatments

Latin Square

Casualization in the Latin Square Design

- The treatments are distributed within the rows, so that each column also contains all the treatments

	Colunas				
Linhas	1	2	3	4	5
1	A	B	C	D	E
2	E	A	B	C	D
3	D	E	A	B	C
4	C	D	E	A	B
5	B	C	D	E	A

Latin Square

Casualization in the Latin Square Design

- Then, rows are then randomly distributed among themselves, and then the columns

E	A	B	C	D
C	D	E	A	B
B	C	D	E	A
A	B	C	D	E
D	E	A	B	C

Casualizing the lines (2, 4, 5, 1, 3)

Latin Square

Casualization in the Latin Square Design

B	D	E	C	A
E	B	C	A	D
D	A	B	E	C
C	E	A	D	B
A	C	D	B	E

→ Final square

Casualizing the columns (3, 5, 1, 4, 2)

Latin Square

Casualization in the Latin Square Design

- 3x3 Latin Square

	1	2	3
1	A	B	C
2	B	C	A
3	C	A	B

Standard square

Randomize columns

3	1	2
C	A	B
A	B	C
B	C	A

Randomize all but the first row

C	A	B
B	C	A
A	B	C

Advantages of the Latin Square Design

- You can control variation in two directions
- Hopefully you increase efficiency as compared to the RCBD

Disadvantages of Latin Square Design

- The number of treatments must equal the number of replicates
- The experimental error to increase with the size of the square
- Small squares have very few degrees of freedom for experimental error

Disadvantages of Latin Square Design

- Effect of the Size of the Square on Error Degrees of Freedom

SOV	Df	2x2	3x3	4x4	5x5	8x8
Rows	r-1	1	2	3	4	7
Columns	r-1	1	2	3	4	7
Treatments	r-1	1	2	3	4	7
Error	(r-1)(r-2)	0	2	6	12	42
Total	$r^2 - 1$	3	8	15	24	63

The experimental error is likely to increase with the size of the square

Latin Square

Statistical Model

- The statistical model for a latin square design is written as:

$$y_{ijk} = \mu + \tau_i + \gamma_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

where y_{ijk} is the observed response, μ is the overall mean, τ_i is the treatment effect, β_j is the row effect, γ_k is the column effect and ε_{ijk} is the experimental error

Latin Square

Statistical Model

- One or both of the blocking factors can be treated as random (Mixed Models)
- Treatment effects are usually considered fixed

Latin Square

Statistical Model

- If rows and columns are random, we assume that:

  We assume that:

  - $\beta_j \sim N\left(0, \sigma_\beta^2\right)$
  - $\gamma_k \sim N(0, \sigma_\gamma^2)$
  - $e_{ijk} \sim N(0, \sigma_e^2)$
  - β_j, γ_k and e_{ijk} are all independent of each other

Latin Square

Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Rows	$t-1$	SS_{Row}	$MS_{Row} = \frac{SS_{Row}}{t-1}$
Columns	$t-1$	SS_{Col}	$MS_{Col} = \frac{SS_{Col}}{t-1}$
Treatments	$t-1$	SS_{Trt}	$MS_{Trt} = \frac{SS_{Trt}}{t-1}$
Residuals	$(t-1)(t-2)$	SS_{Res}	$MS_{Res} = \frac{SS_{Res}}{(t-1)(t-2)}$
Total	t^2-1	SS_{Total}	

- There are $(t-1)(t-2)$ degrees of freedom for the residual
  - More treatment levels increase the residual degrees of freedom, but require a larger experiment

Latin Square

Analysis of Variance

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square	F-Statistic
Rows	$t-1$	MS_{Row}	$\sigma_\varepsilon^2 + t\sigma_\beta^2$	
Columns	$t-1$	MS_{Col}	$\sigma_\varepsilon^2 + t\sigma_\gamma^2$	
Treatments	$t-1$	MS_{Trt}	$\sigma_\varepsilon^2 + \frac{t}{t-1}\sum_i \tau_i^2$	$F = \frac{\text{MS}_{\text{Trt}}}{\text{MS}_{\text{Res}}}$
Residuals	$(t-1)(t-2)$	MS_{Res}	σ_ε^2	
Total	t^2-1			

The F-test for treatment effects is not affected if rows and/or columns are fixed

Let´s Practice 01!

- The first exercise data set consists of stem dry weight (in g, log10 scale) of different sunflower genotypes

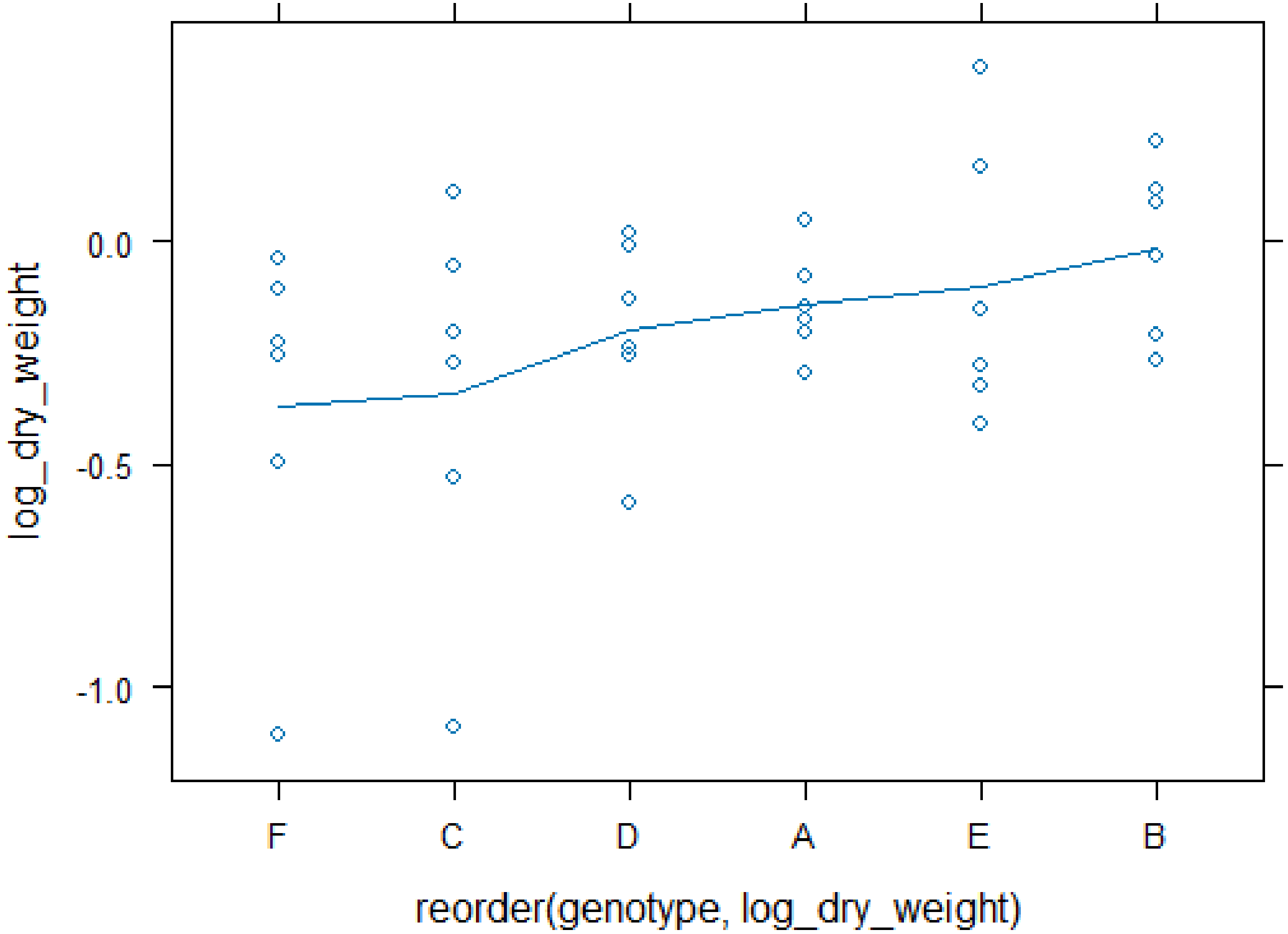
- Six genotypes, denoted A through F, were grown in a greenhouse according to a latin square design

- These data come from Povin, C (1993) ANOVA: Experiments in Controlled Environments. **Design and Analysis of Ecological Experiments**. Ed. SM Scheiner, Ed. J Gurevitch. Chapman & Hall. 46-67.

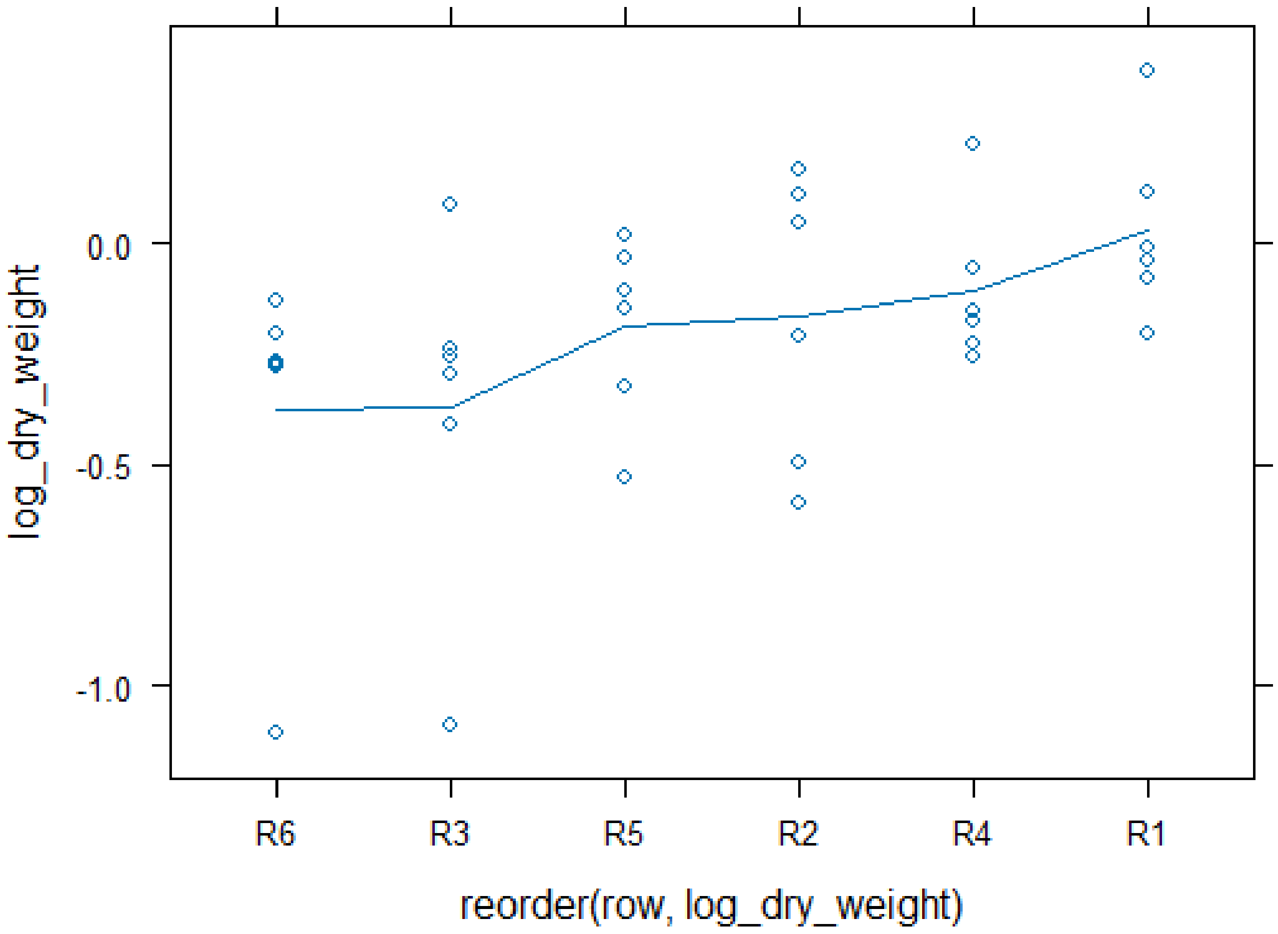
Let´s Practice 01!

- Use the R function read.csv to import the data
- Fit the model with fixed (and/or random) effects
- Check if model assumptions are met
- Build the ANOVA table and test the null hypothesis of no difference between genotypes
- Use multiple comparisons to assess pairwise differences

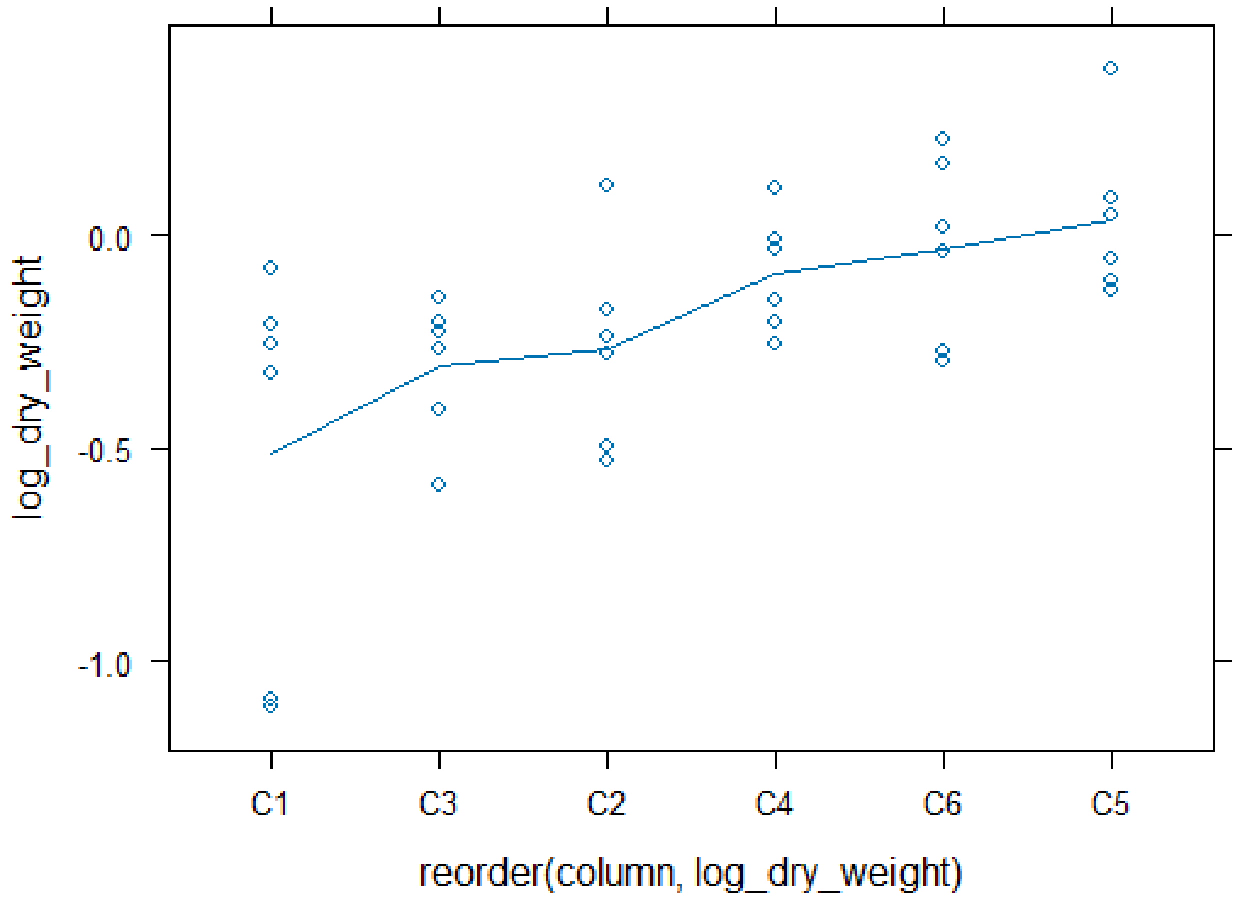
Let´s Practice 01!



Let´s Practice 01!



Let´s Practice 01!



Let´s Practice 01!

- Fit the model with fixed (and/or random) effects

```
#Fixed Model
fm <- lm(log_dry_weight ~ row + column + genotype, data = dados)
anova(fm)

Analysis of Variance Table

Response: log_dry_weight
          Df  Sum Sq  Mean Sq F value    Pr(>F)    
row        5 0.73072 0.146145  4.2691 0.0083606 ** 
column     5 1.25690 0.251380  7.3431 0.0004727 ***
genotype   5 0.57970 0.115939  3.3867 0.0223142 *  
Residuals 20 0.68467 0.034233                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

"fm" used to refer to an adjusted model

Let´s Practice 01!

- Fit the model with fixed (and/or random) effects
  - Only genotype as fixed

```
library(lmerTest)
fme <- lmer(log_dry_weight ~ genotype + (1 | row) + (1 | column), data = dados)
anova(fme)
```

```
> fme <- lmer(log_dry_weight ~ genotype + (1 | row) + (1 | column), data = dados)
> anova(fme)
Type III Analysis of Variance Table with Satterthwaite's method
         Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
genotype 0.5797 0.11594     5    20  3.3867 0.02231 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The package “lmerTest” is used to perform statistical hypothesis testing on linear mixed models fitted with the lme4 package

Let´s Practice 01!

- Multiple Comparisons

```
library(emmeans)
fm_means <- emmeans(fm, "genotype")
pairs(fm_means)
plot(fm_means, comparisons = TRUE)
cld(fm_means, adjust = "tukey", Letters = letters)
```

- The emmeans function, short for "Estimated Marginal Means": used to calculate and extract estimated marginal means after fitting a statistical model

Let´s Practice 01!

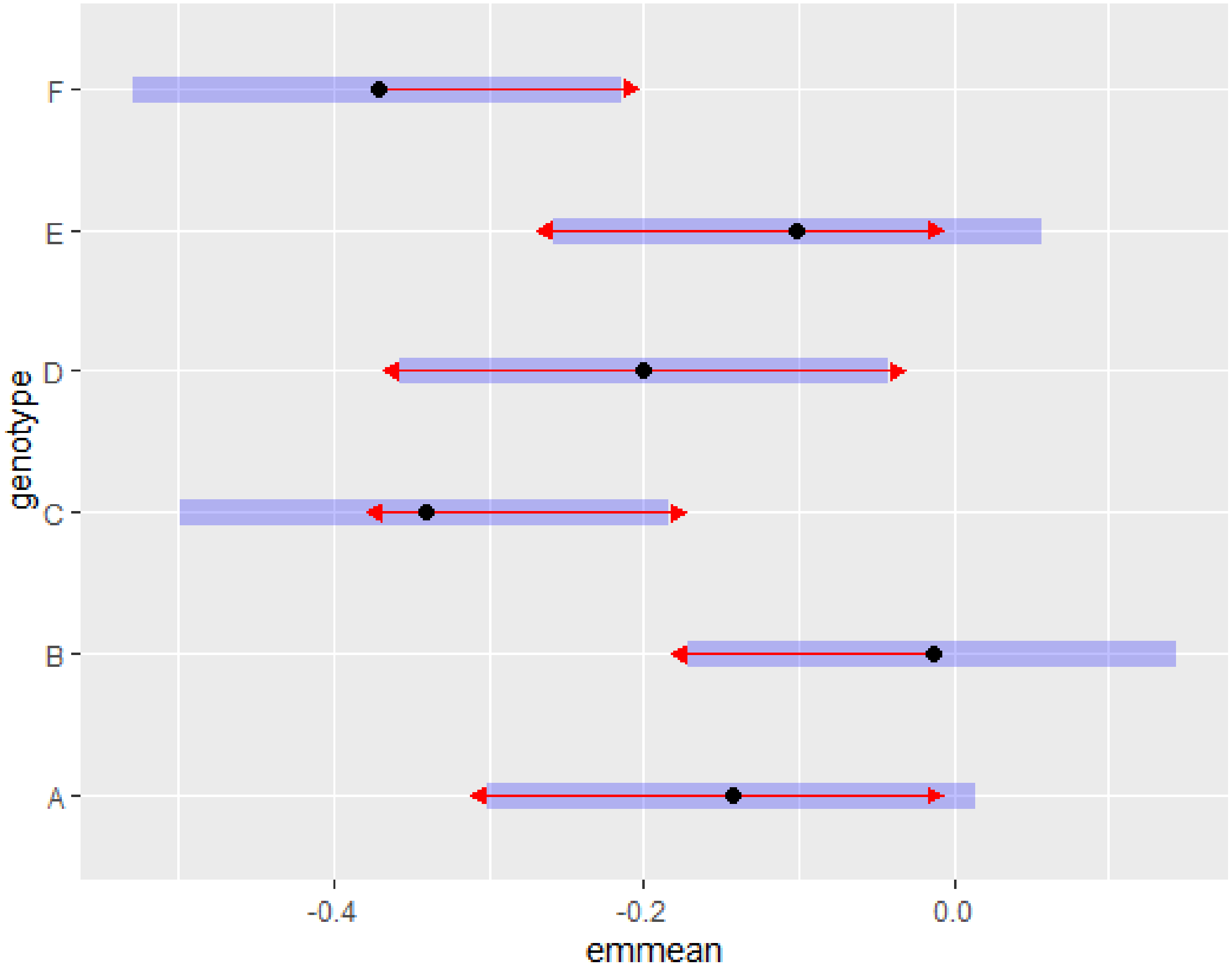
- Multiple Comparisons

```
> library(emmeans)
> fm_means <- emmeans(fm, "genotype")
> pairs(fm_means)
 contrast estimate    SE df t.ratio p.value
 A - B     -0.1290 0.107 20  -1.208  0.8280
 A - C      0.1979 0.107 20   1.853  0.4570
 A - D      0.0568 0.107 20   0.531  0.9942
 A - E     -0.0418 0.107 20  -0.391  0.9986
 A - F      0.2282 0.107 20   2.136  0.3095
 B - C      0.3270 0.107 20   3.061  0.0591
 B - D      0.1858 0.107 20   1.739  0.5235
 B - E      0.0873 0.107 20   0.817  0.9609
 B - F      0.3572 0.107 20   3.344  0.0330
 C - D     -0.1412 0.107 20  -1.321  0.7702
 C - E     -0.2397 0.107 20  -2.244  0.2622
 C - F      0.0302 0.107 20   0.283  0.9997
 D - E     -0.0985 0.107 20  -0.922  0.9361
 D - F      0.1714 0.107 20   1.604  0.6052
 E - F      0.2699 0.107 20   2.527  0.1632

Results are averaged over the levels of: row, column
P value adjustment: tukey method for comparing a family of 6 estimates
```

Let´s Practice 01!

- Multiple Comparisons



References

- Chapters 9.10, 9.11 (for a more classical view)[1]
- Section 3.6.3 and sections 5.1 to 5.3[2]

1. Steel, R. G. & Torrie, J. H. Principles and Procedures of Statistics: A Biometrical Approach. 2nd Edition. (1980).

2. Casella, G. Statistical Design. (2008).