LGN 5822 - Biometrical Genetics

L09 – Factorial Experiments

Michele Jorge Silva Siqueira

2023

Introduction

- You may be interested in comparing two different sets of treatments

Introduction

- You may be interested in comparing two different sets of treatments
- You can fix one level of a treatment and study the other, or you may **compare both simultaneously**
- Treatment designs that combine two (or more) treatment factors are factorial designs

Introduction

Factorial Experiments

- Factorial experiments involve simultaneously more than one factor and each factor is at two or more levels

Introduction

Factorial Experiments: Objective

- Effect of different factors that influence the variable under study and the relationship between them

- The main idea of a factorial experiment is that it allows you to evaluate not only the individual effect of each factor, but also how these factors interact

Introduction

- Several factors affect simultaneously the characteristic under study in factorial experiments

- The researcher is interested in the main effects and the interaction effects among different factors

Factorial Experiments

Experimental Design X Factorial Arrangement

- **Experimental Design:** It is the manner in which levels of a factor or treatments are distributed to experimental units
  - Experiment design defines the error structure

    Completely Randomized Design (CRD)
    - Randomized Complete Block Design (RCBD)
    - Latin Square (LS)

- **Factorial arrangement:** It is the way in which the combination of the levels of the factors under study is organized in experimental units

What is the difference?

Factorial Experiments

Experimental Design X Factorial Arrangement

- Factorial experiments are set up according to a type of design experimental, such as: CRD, RCBD, LS

- The choice of design depends on the homogeneity of the experimental unit

Factorial Experiments

Treatment Design

- In factorial experiments, treatments are obtained by combinations factor levels

- Complete factorial design, each level of a factor combines with all levels of the other factors

Factorial Experiments

Treatment Design

- The symbology commonly used for factorial experiments is to indicate the product of the levels of the factors under test ("x")

  - *e.g.* Factorial Experimente $2x4x6$
    - in the experiment were tested simultaneously 3 factors. The first has 2 levels, the second 4 levels and the third 6 levels

- When the number of levels is the same for all factors, the following symbols can be used: n^F
  - where **F** is the number of factors and **n** is the number of levels of each factor
    - *e.g.* 4^3*:* the experiment has 3 factors with 4 levels each

Factorial Experiments

Treatment Design

- Factorial designs can include multiple levels of a larger number of treatments
    - *E.g.* if three factors are under investigation, the design is an $m \times n \times o$ factorial arrangement

- If the numbers of factors and/or levels are moderately large, the size of the experiment can be a limiting factor

Factorial Experiments

Effects

- **Main Effect:** is the effect of each factor, independent of the effect of other factors

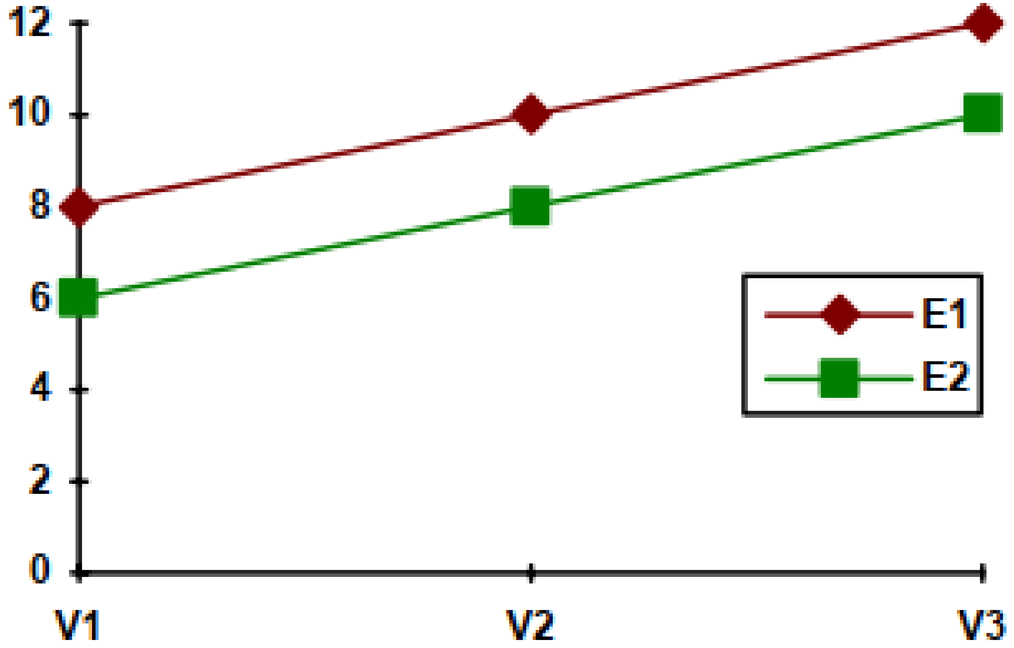
- **Interaction Effect:** is effect of factors on the variable in study
    - The interaction between factors occurs when the effects of the levels of one factor are modified by the levels of the other factor

Factorial Experiments

Effects

Is there interaction?

- Factorial: 3x2



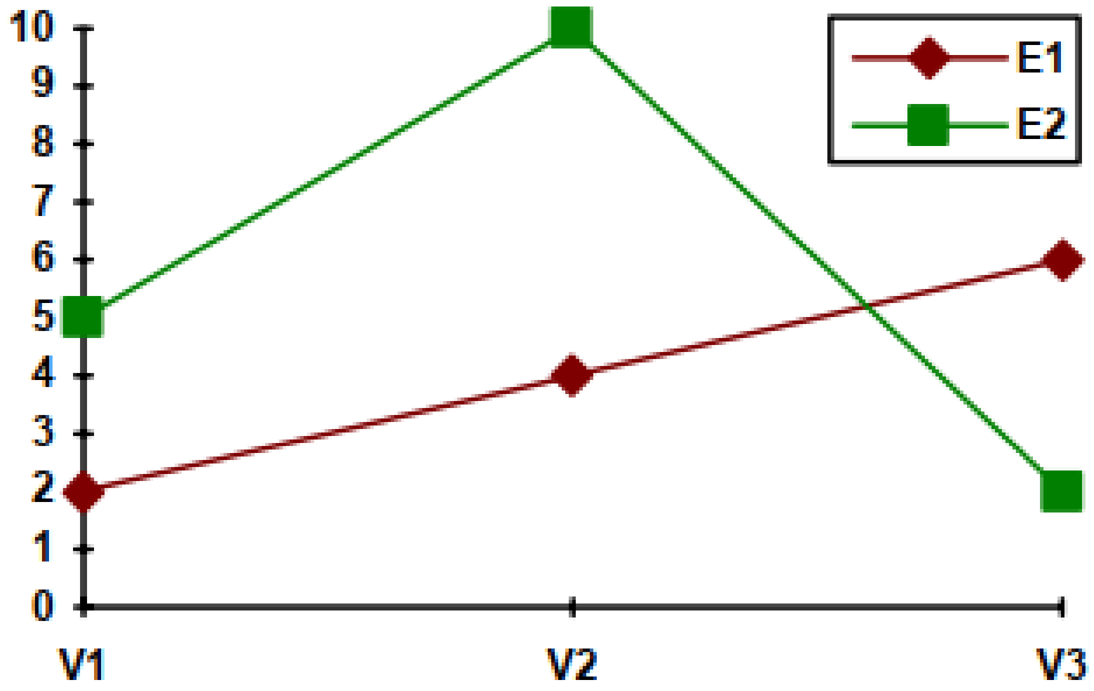
No interaction

Factorial Experiments

Effects

Is there interaction?

- Factorial: 3x2



Interaction

Factorial Experiments

Data Table

- Factors A and B, with I and levels, respectively, according to the CRD
  - K replicates

Repetition	A1				A2				...	AI			
	B1	B2	...	BJ	B1	B2	...	BJ	...	B1	B2	...	BJ
1	Y_{111}	Y_{121}	...	Y_{1J1}	Y_{211}	Y_{221}	...	Y_{2J1}	...	Y_{I11}	Y_{I21}	...	Y_{IJ1}
2	Y_{112}	Y_{122}	...	Y_{1J2}	Y_{212}	Y_{222}	...	Y_{2J2}	...	Y_{I12}	Y_{I22}	...	Y_{IJ2}
...	...	...	...	...	...	...	...	...	...	...	...	...	...
K	Y_{11K}	Y_{12K}	...	Y_{1JK}	Y_{21K}	Y_{22K}	...	Y_{2JK}	...	Y_{I1K}	Y_{I2K}	...	Y_{IJK}
Total	$Y_{11\bullet}$	$Y_{12\bullet}$	...	$Y_{1J\bullet}$	$Y_{21\bullet}$	$Y_{22\bullet}$	...	$Y_{2J\bullet}$	...	$Y_{I1\bullet}$	$Y_{I2\bullet}$	...	$Y_{IJ\bullet}$

Factorial Experiments

Data Table

- Number of experimental units: $\mathrm{N = I \times J \times K}$
- Total of the *ij*-th treatment: $AB_{\mathrm{ij}} = \sum_{\mathrm{k=1}}^{\mathrm{k}} \mathrm{Y}_{\mathrm{ij}k} = \mathrm{Y}_{\mathrm{ij.}}$
- Total of the *i*-th level of factor A: $A_{\mathrm{i}} = \sum_{j=1,\mathrm{k}=1}^{j,\mathrm{k}} \mathrm{Y}_{\mathrm{ij}k} = \mathrm{Y}_{\mathrm{i..}}$
- Total of the *j*-th level of factor B: $B_{\mathrm{j}} = \sum_{i=1,\mathrm{k}=1}^{i,\mathrm{k}} \mathrm{Y}_{\mathrm{ij}k} = \mathrm{Y}_{\mathrm{.j.}}$

Factorial Experiments

Data Table

- Mean of the *i*-th level of factor A: $\widehat{m_{Ai}} = \frac{A_i}{JK}$
- Mean of the *j*-th level of factor B: $\widehat{m_{Bj}} = \frac{B_j}{IK}$
- General mean: $\hat{m} = \frac{G}{N}$
    - Number of experimental units: $\mathrm{N = I \times J \times K}$

Factorial Experiments

Data Table: Total Treatments

Factor A	Factor B				Total
	B1	B2	...	BJ	
A1	$Y_{11.}$	$Y_{12.}$	...	$Y_{1J.}$	A_1
A2	$Y_{21.}$	$Y_{22.}$	...	$Y_{2J.}$	A_2
...	...	...	...	...	...
AI	$Y_{I1.}$	$Y_{I2.}$	...	$Y_{IJ.}$	A_I
Totais	B_1	B_2	...	B_j	G

Factorial Experiments

Statistical Model

Consider a factorial experiment, with two factors: factor T with i levels and factor G with j levels, installed according to the DIC, with K repetitions

Factorial Experiments

Statistical Model

Model

- In a CRD, data can be described with the following model:

$$y_{ijk} = \mu + \tau_i + \gamma_i + (\tau\gamma)_{ij} + \varepsilon_{ijk}$$

where τ_i is the effect of treatment T, γ_i is the effect of treatment G and $(\tau\gamma)_{ij}$ represents the interaction of the two factors

Factorial Experiments

Statistical Model

Model

- In a RCBD, data can be described with the following model:

$$y_{ijk} = \mu + \tau_i + \gamma_i + (\tau\gamma)_{ij} + \omega_k + \varepsilon_{ijk}$$

where τ_i is the effect of treatment T, γ_i is the effect of treatment G and $(\tau\gamma)_{ij}$ represents the interaction of the two factors

Factorial Experiments

Statistical Model

- The effects τ_i of treatment T and γ_i of treatment G are called *main effects*

Factorial Experiments

Statistical Model

We assume that:

- $\varepsilon_{ij} \sim N(0, \sigma^2)$, for $i = 1, \dots, t$ and $j = 1, \dots, g$ and $k = 1, \dots, r$
- $cov\left(\varepsilon_{ijk}, \varepsilon_{i^{\prime}j^{\prime}k^{\prime}}\right) = 0$

Errors are independent and identically distributed (i.i.d)

Factorial Experiments

Analysis of Variance

- The corresponding ANOVA table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
T	$t-1$	SS_T	$MS_T = \frac{SS_T}{t-1}$
G	$g-1$	SS_G	$MS_G = \frac{SS_G}{g-1}$
T × G	$(t-1)(g-1)$	$SS_{T \times G}$	$MS_{T \times G} = \frac{SS_{T \times G}}{(t-1)(g-1)}$
Within	$tg(r-1)$	SS_{Within}	$MS_{Within} = \frac{SS_{Within}}{tg(r-1)}$
Total	$tgr-1$	SS_{Total}	

Factorial Experiments

Analysis of Variance

- If treatments T and G are random:

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square
T	$t-1$	MS_T	$\sigma^2 + r\sigma_{tg}^2 + rg\sigma_t^2$
G	$g-1$	MS_G	$\sigma^2 + r\sigma_{tg}^2 + rt\sigma_g^2$
T × G	$(t-1)(g-1)$	$MS_{T\times G}$	$\sigma^2 + r\sigma_{tg}^2$
Within	$tg(r-1)$	MS_{Within}	σ^2
Total	$tgr-1$		

The expected mean squares are the expected values of these terms with the specified model

Factorial Experiments

Analysis of Variance

Mixed Models

- For mixed effects models, we will use REML estimates, the likelihood ratio test and model selection criteria such as AIC/BIC

- From now on, we deal with fixed effects only

Factorial Experiments

Analysis of Variance

- If T is fixed and G is random:

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square
T	$t-1$	MS_T	$\sigma^2 + r\sigma_{tg}^2 + \frac{rg}{t-1}\sum_i \tau_i^2$
G	$g-1$	MS_G	$\sigma^2 + rt\sigma_g^2$
T × G	$(t-1)(g-1)$	$MS_{T\times G}$	$\sigma^2 + r\sigma_{tg}^2$
Within	$tg(r-1)$	MS_{Within}	σ^2
Total	$tgr-1$		

Factorial Experiments

- If treatments T and G are fixed:

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square
T	$t-1$	$\mathrm{MS_T}$	$\sigma^2 + \frac{rg}{t-1}\sum_i \tau_i^2$
G	$g-1$	$\mathrm{MS_G}$	$\sigma^2 + \frac{rt}{g-1}\sum_j \gamma_j^2$
T × G	$(t-1)(g-1)$	$\mathrm{MS_{T\times G}}$	$\sigma^2 + \frac{r}{(t-1)(g-1)}\sum_{ij}(\tau\gamma)$
Within	$tg(r-1)$	$\mathrm{MS_{Within}}$	σ^2
Total	$tgr-1$		

Factorial Experiments

Hypothesis Testing

- The null hypothesis of no effect of treatment T is

  H_0: τ_i = 0 for all i

- Similarly, the null hypothesis of no effect of treatment G is

  H_0: τ_j = 0 for all j

- The null hypothesis of no interaction effects is

  H_0: $+(\tau\gamma)_{ij}$ = 0 for all i, j

Factorial Experiments

Hypothesis Testing

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square	F-Statistic
T	$t-1$	$\mathrm{MS_T}$	$\sigma^2 + \frac{rg}{t-1}\sum_i \tau_i^2$	$F = \frac{\mathrm{MS_T}}{\mathrm{MS_{Within}}}$
G	$g-1$	$\mathrm{MS_G}$	$\sigma^2 + \frac{rt}{g-1}\sum_j \gamma_j^2$	$F = \frac{\mathrm{MS_G}}{\mathrm{MS_{Within}}}$
T × G	$(t-1)(g-1)$	$\mathrm{MS_{T\times G}}$	$\sigma^2 + \frac{r}{(t-1)(g-1)}\sum_{ij}(\tau\gamma)_{ij}^2$	$F = \frac{\mathrm{MS_{T\times G}}}{\mathrm{MS_{Within}}}$
Within	$tg(r-1)$	$\mathrm{MS_{Within}}$	σ^2	
Total	$tgr-1$			

Factorial Experiments

Interaction

- Possible results for the F test of the interaction

Factorial Experiments

Interaction

- Possible results for the F test of the interaction

Note

- **Non-significant interaction**, the effects of the factors are independent

Factorial Experiments

Interaction

- Possible results for the F test of the interaction

Note

- A **significant interaction** means that the effect of one treatment is dependent on the levels of another

- Interpretation of treatment effects can be difficult, especially when there are many interactions

- The effect of one factor depends on the level of the other factor

Factorial Experiments

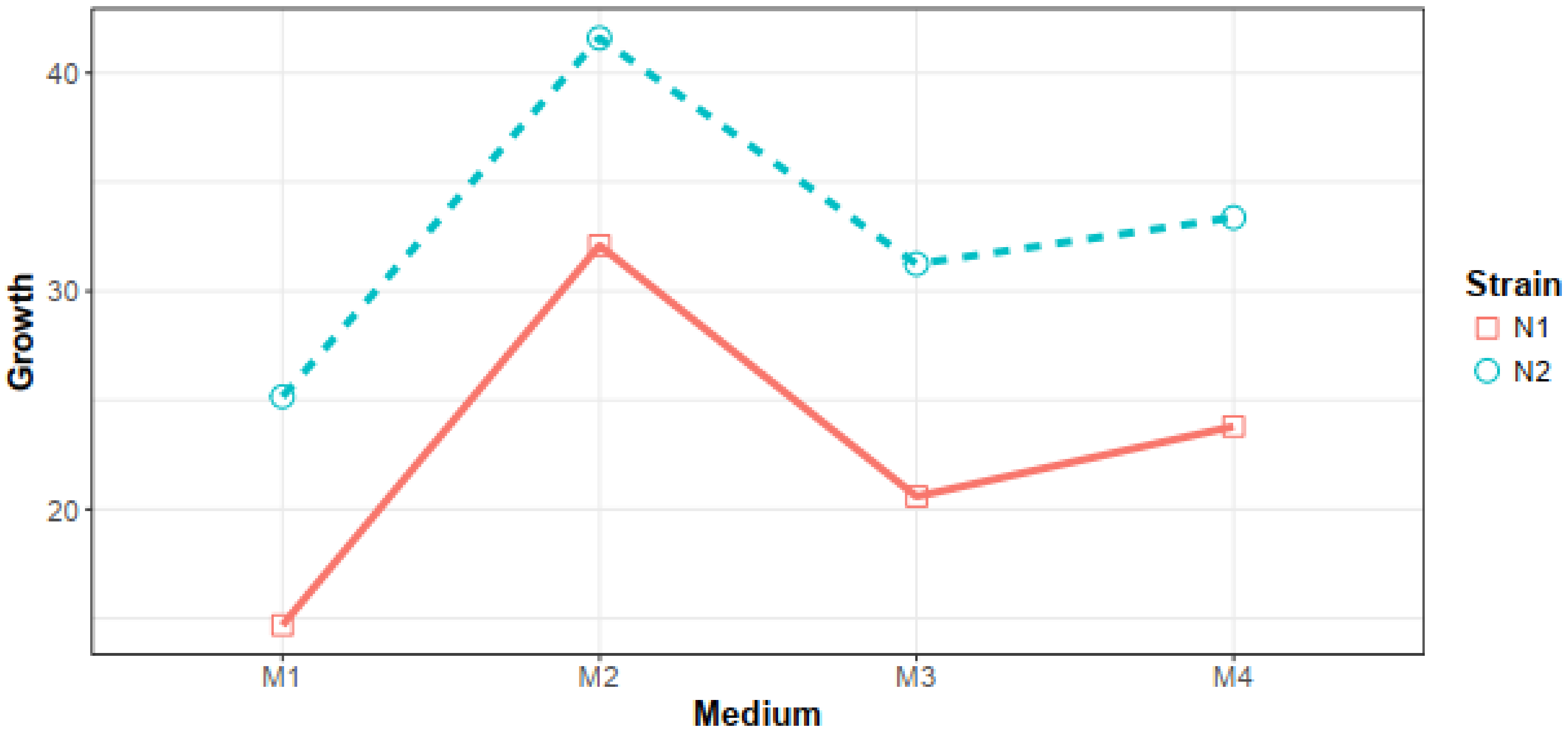
Example: Interaction

- Suppose we are investigating the effect of $m = 4$ different culture mediums (solution) with the growth rate of $n = 2$ strains

- We can plot cell means to construct the **interaction plot**

Factorial Experiments

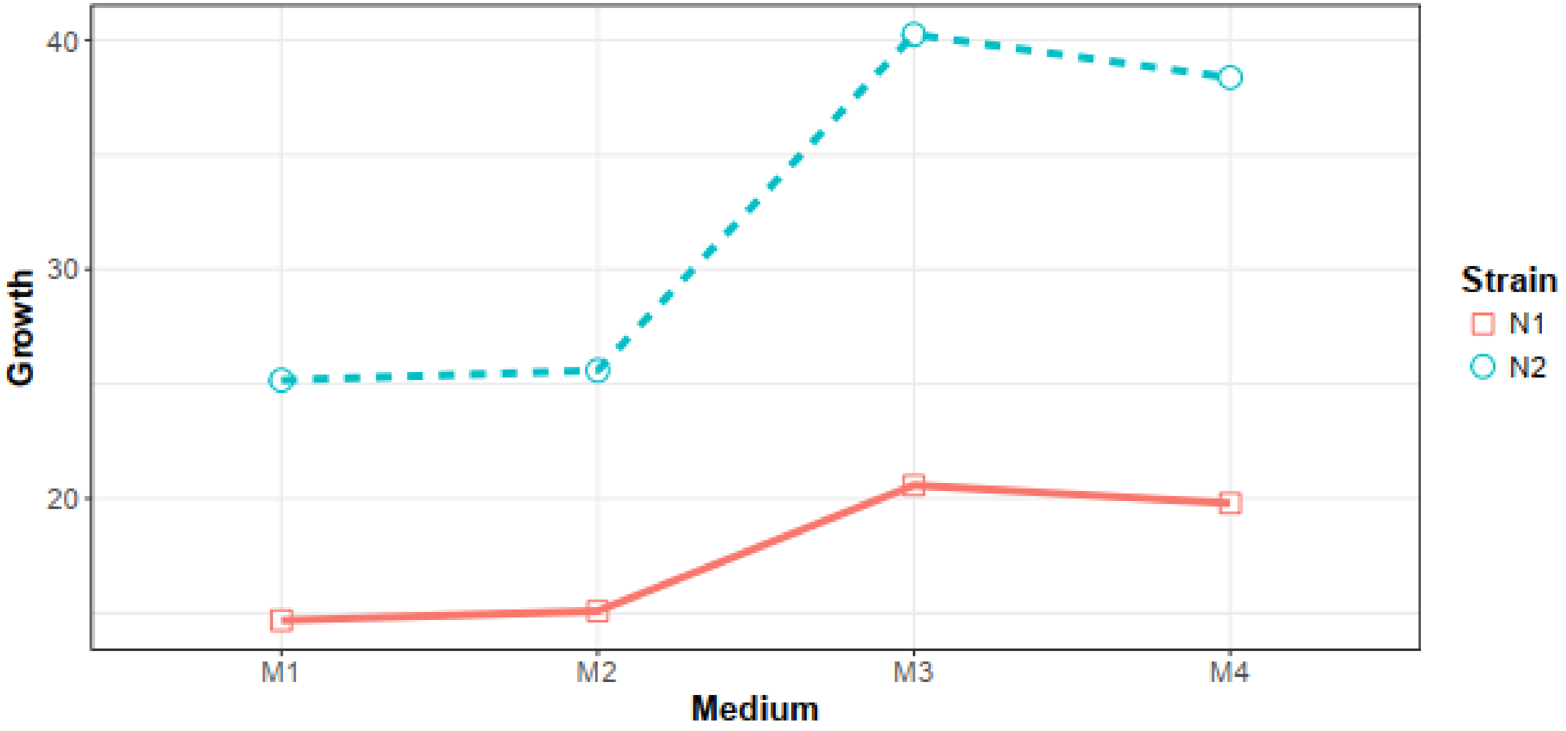
Example: Interaction



Absence of interaction

Factorial Experiments

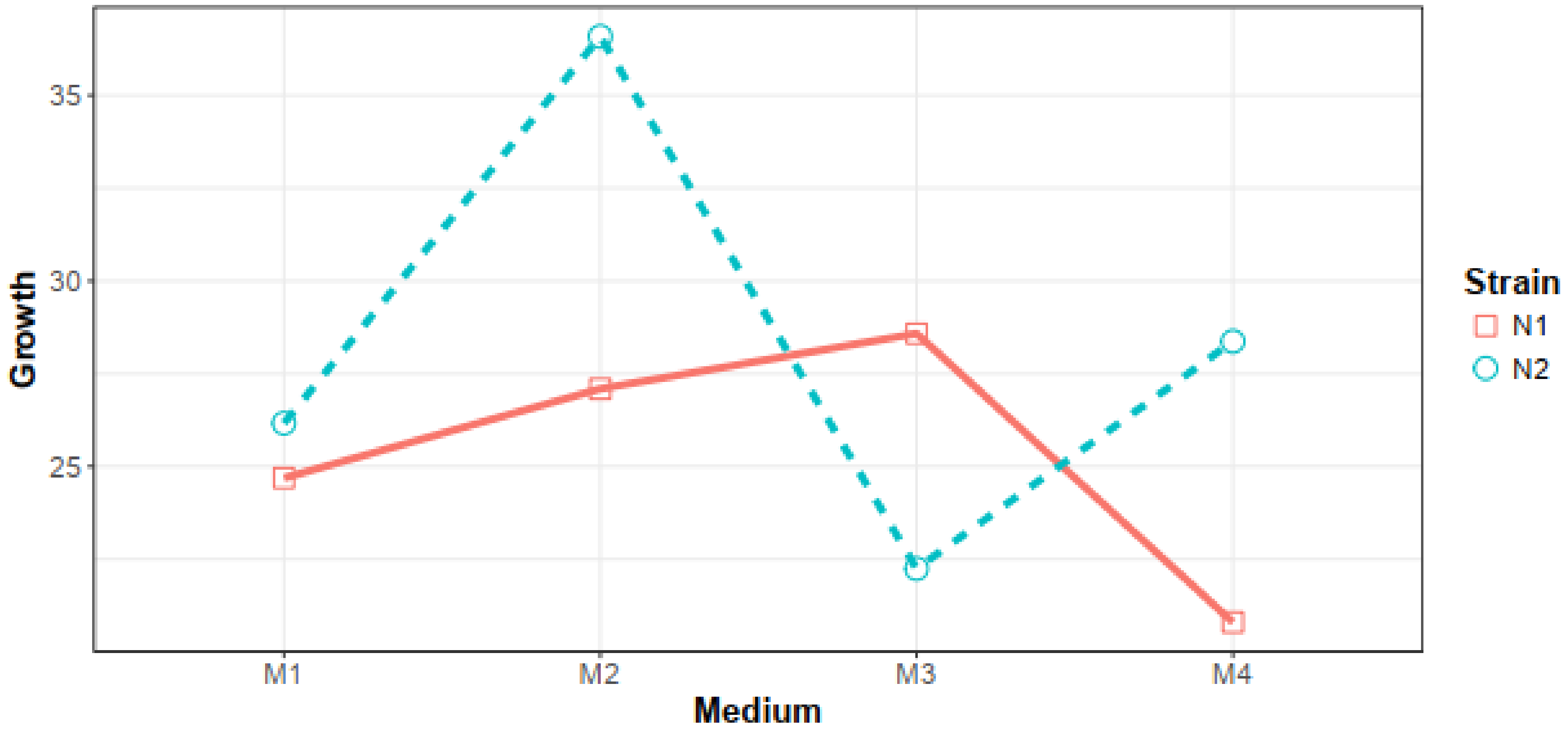
Example: Interaction



Quantitative interaction

Factorial Experiments

Example: Interaction



Qualitative interaction

Factorial Experiments

Multiple Comparasions

- In the absence of interaction, we can contrast the main (average) effects of both treatment factors
  - Compare the levels of treatment T and/or
  - Compare the levels of treatment G

Factorial Experiments

Multiple Comparasions

- **If interaction if significant**, we compare the levels of treatment T within each level of treatment G (and vice versa)

Factorial Experiments

Advantages of a factorial experiment

- It allows the study of main effects and the effect of the interaction between factors

- The number of degrees of freedom associated with the residue is high when compared to simple experiments on the same factors, **which contributes to reducing the residual variance, increasing the precision of the experiment**

Factorial Experiments

Disadvantage of a factorial experiment

- Requires a greater number of experimental units in relation to simple experiments

Let´s Practice 01!

- The data set consists of dry matter measurements (in mg) in *Aspergillus nidulans*
- Dry matter yield from **ten different strains** was measured in two contrasting conditions: mediums with and without nicotinic acid
- The experiment was conducted as an RCB design with two blocks
- These data come from Favraud & Azevedo. ESALQ/USP.

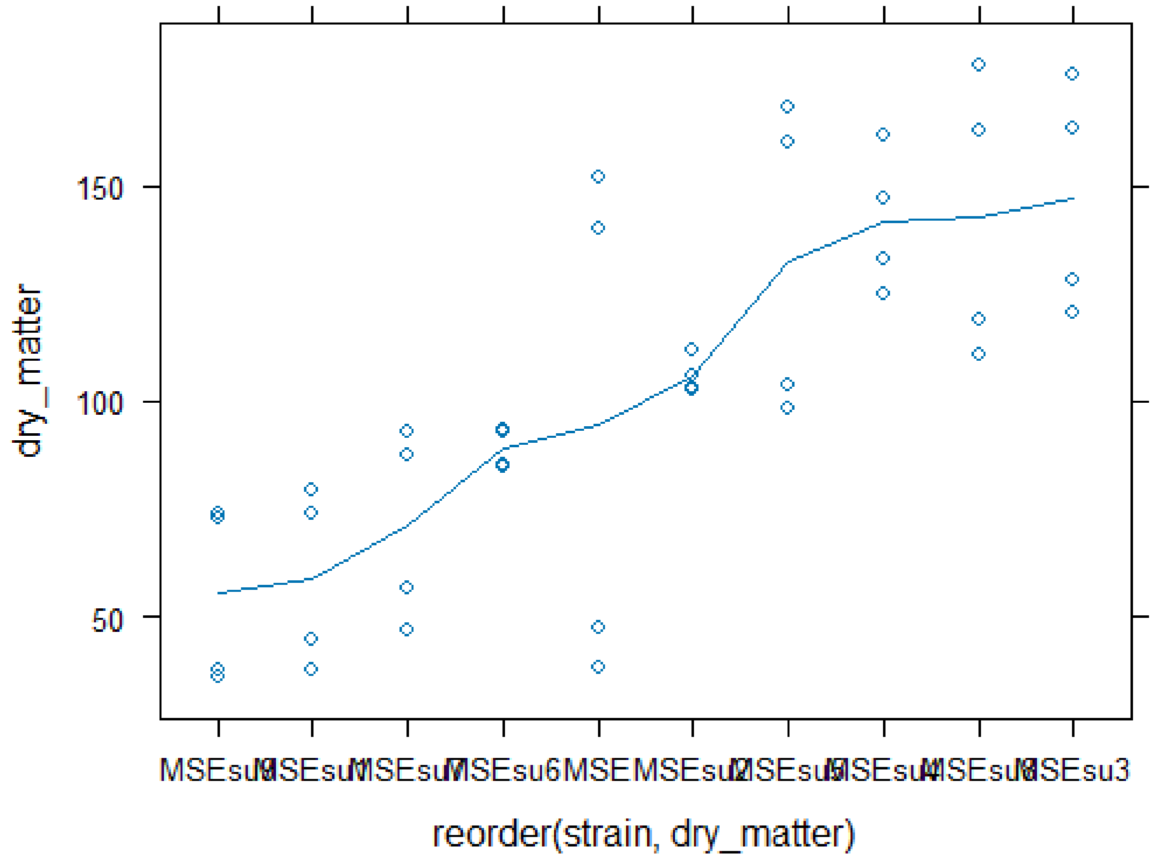
Let´s Practice 01!

- 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 hypotheses of no difference between strains, no difference between nicotinic acid treatments and no interaction between both factors
- Draw the interaction plot
- Use multiple comparisons to assess pairwise differences

Let´s Practice 01!

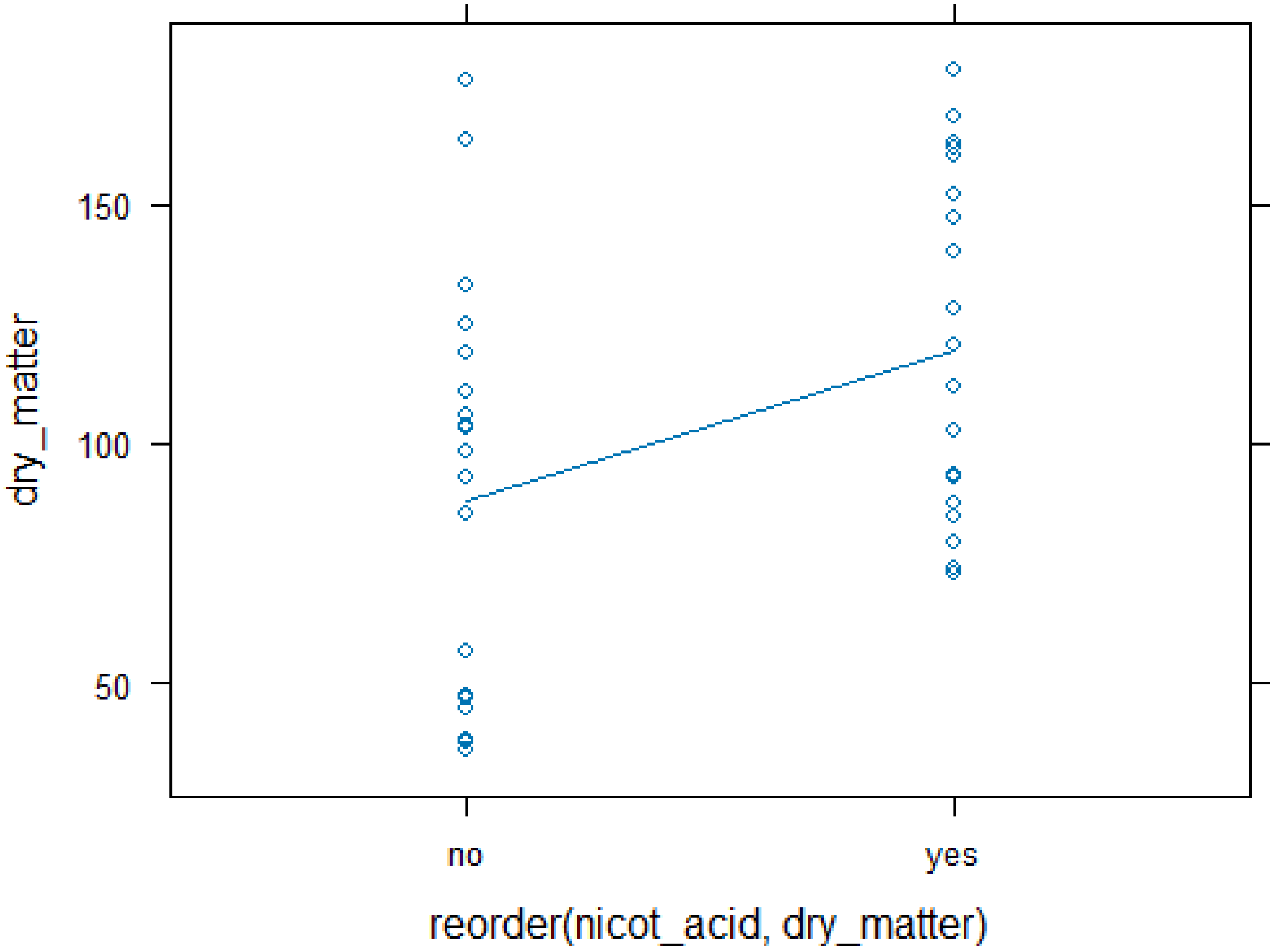
```
> dados
   strain nicot_acid  treatment block dry_matter
1     MSE         no     MSE_no    B1       47.2
2     MSE         no     MSE_no    B2       38.2
3     MSE        yes    MSE_yes    B1      140.0
4     MSE        yes    MSE_yes    B2      152.0
5  MSEsu1         no  MSEsu1_no    B1       37.6
6  MSEsu1         no  MSEsu1_no    B2       44.4
7  MSEsu1        yes MSEsu1_yes    B1       79.6
8  MSEsu1        yes MSEsu1_yes    B2       74.0
9  MSEsu2         no  MSEsu2_no    B1      103.2
10 MSEsu2         no  MSEsu2_no    B2      106.0
11 MSEsu2        yes MSEsu2_yes    B1      102.6
12 MSEsu2        yes MSEsu2_yes    B2      112.0
13 MSEsu3         no  MSEsu3_no    B1      163.6
```

Let´s Practice 01!

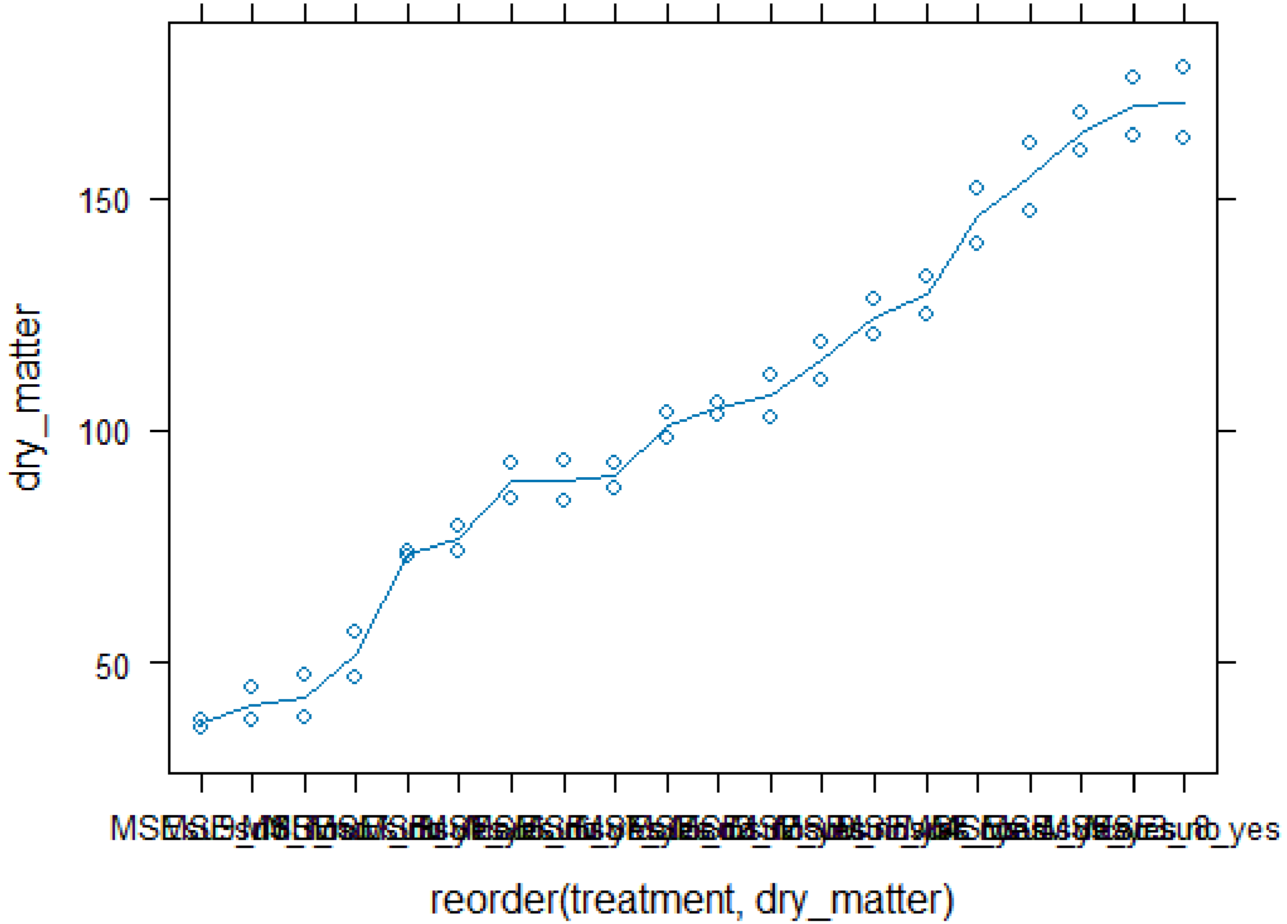


"xyplot" is a function from the "lattice" package. This function is used to create scatter plots

Let´s Practice 01!



Let´s Practice 01!



Let´s Practice 01!

Anova without interaction

```
fm1 <- lm(dry_matter ~ treatment, data = dados)
anova(fm1)

fm2 <- lm(dry_matter ~ strain * nicot_acid, data = dados)
anova(fm2)

fm3 <- lm(dry_matter ~ block + treatment, data = dados)
anova(fm3)
```

Let´s Practice 01!

\# Anova without interaction

```
> fm1 <- lm(dry_matter ~ treatment, data = dados)
> anova(fm1)
Analysis of Variance Table

Response: dry_matter
          Df Sum Sq Mean Sq F value    Pr(>F)
treatment 19  70179  3693.6  95.539 1.39e-15 ***
Residuals 20    773    38.7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Let´s Practice 01!

Anova without interaction

```
> fm2 <- lm(dry_matter ~ strain * nicot_acid, data = dados)
> anova(fm2)
Analysis of Variance Table

Response: dry_matter
                  Df Sum Sq Mean Sq F value    Pr(>F)    
strain             9  45653  5072.5 131.205 8.517e-16 ***
nicot_acid         1   9967  9966.6 257.796 6.819e-13 ***
strain:nicot_acid  9  14560  1617.8  41.845 4.938e-11 ***
Residuals         20    773    38.7                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Let´s Practice 01!

Anova without interaction

```
> fm3 <- lm(dry_matter ~ block + treatment, data = dados)
> anova(fm3)
Analysis of Variance Table

Response: dry_matter
          Df Sum Sq Mean Sq F value    Pr(>F)
block      1     64    64.0  1.7148     0.206
treatment 19  70179  3693.6 98.9537 4.403e-15 ***
Residuals 19    709    37.3
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Let´s Practice 01!

Anova with interaction

```
fm5 <- lm(dry_matter ~ block + strain + nicot_acid: strain + nicot_acid, data = dados)
anova(fm5)
```

```
> fm5 <- lm(dry_matter ~ block + strain + nicot_acid: strain + nicot_acid, data = dados)
> anova(fm5)
Analysis of Variance Table

Response: dry_matter
                  Df Sum Sq Mean Sq  F value    Pr(>F)    
block              1     64    64.0   1.7148     0.206    
strain             9  45653  5072.5 135.8943 2.552e-15 ***
nicot_acid         1   9967  9966.6 267.0099 1.210e-12 ***
strain:nicot_acid  9  14560  1617.8  43.3402 8.815e-11 ***
Residuals         19    709    37.3                       
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Let´s Practice 01!

Anova with interaction

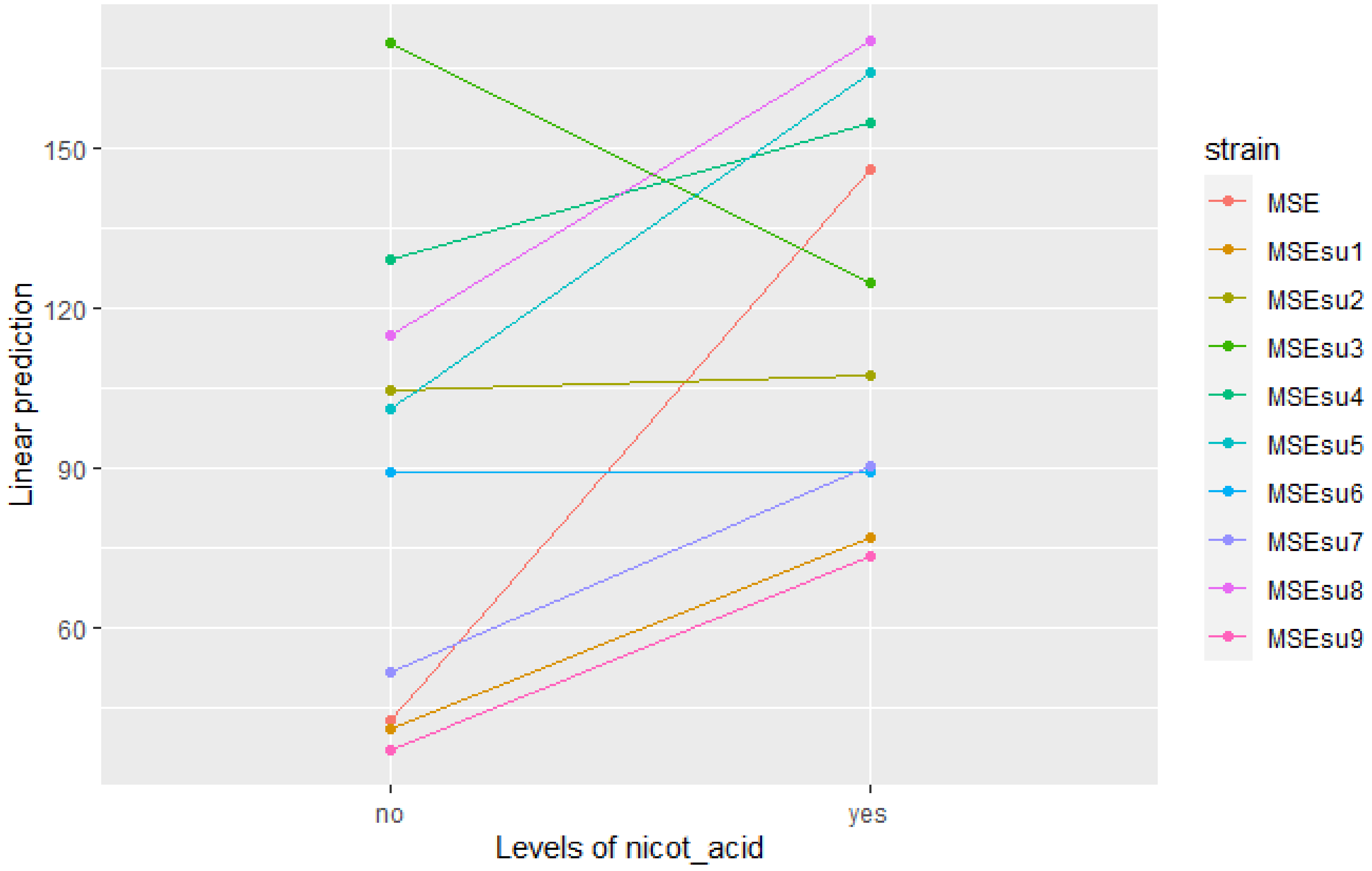
```
fm6 <- lm(dry_matter ~ block + strain * nicot_acid, data = dados)
anova(fm6)
```

```
> fm6 <- lm(dry_matter ~ block + strain * nicot_acid, data = dados)
> anova(fm6)
Analysis of Variance Table

Response: dry_matter
                  Df Sum Sq Mean Sq  F value    Pr(>F)    
block              1     64    64.0   1.7148     0.206    
strain             9  45653  5072.5 135.8943 2.552e-15 ***
nicot_acid         1   9967  9966.6 267.0099 1.210e-12 ***
strain:nicot_acid  9  14560  1617.8  43.3402 8.815e-11 ***
Residuals         19    709    37.3                       
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let´s Practice 01!

- Graph to analyze interaction



Let´s Practice 01!

- Multiple comparisons

```
> fm_means
 strain emmean   SE df lower.CL upper.CL
 MSE      94.3 3.11 20     87.9    100.8
 MSEsu1   58.9 3.11 20     52.4     65.4
 MSEsu2  106.0 3.11 20     99.5    112.4
 MSEsu3  147.2 3.11 20    140.7    153.6
 MSEsu4  141.8 3.11 20    135.3    148.3
 MSEsu5  132.6 3.11 20    126.1    139.1
 MSEsu6   89.1 3.11 20     82.6     95.6
 MSEsu7   71.0 3.11 20     64.5     77.5
 MSEsu8  142.6 3.11 20    136.1    149.0
 MSEsu9   55.2 3.11 20     48.8     61.7
```

Let´s Practice 01!

- Multiple comparisons

```
Results are averaged over the levels of: nicot_acid
Confidence level used: 0.95
> pairs(fm_means)
 contrast          estimate  SE df t.ratio p.value
 MSE - MSEsu1         35.45 4.4 20   8.063  <.0001
 MSE - MSEsu2        -11.60 4.4 20  -2.638  0.2628
 MSE - MSEsu3        -52.80 4.4 20 -12.009  <.0001
 MSE - MSEsu4        -47.45 4.4 20 -10.792  <.0001
 MSE - MSEsu5        -38.25 4.4 20  -8.700  <.0001
 MSE - MSEsu6          5.25 4.4 20   1.194  0.9650
 MSE - MSEsu7         23.35 4.4 20   5.311  0.0011
 MSE - MSEsu8        -48.20 4.4 20 -10.963  <.0001
 MSE - MSEsu9         39.10 4.4 20   8.893  <.0001
 MSEsu1 - MSEsu2     -47.05 4.4 20 -10.701  <.0001
 MSEsu1 - MSEsu3     -88.25 4.4 20 -20.072  <.0001
```

References

- Chapters 7 (Analysis of Variance II: Multiway Classifications) and 15 (Analysis of Variance III: Factorial Experiments)[1] (for a more classical view)

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