Working Paper- Design and Analysis of Experiments

Diego Vallarino - Not to Reprint

17/11/2021

Contents

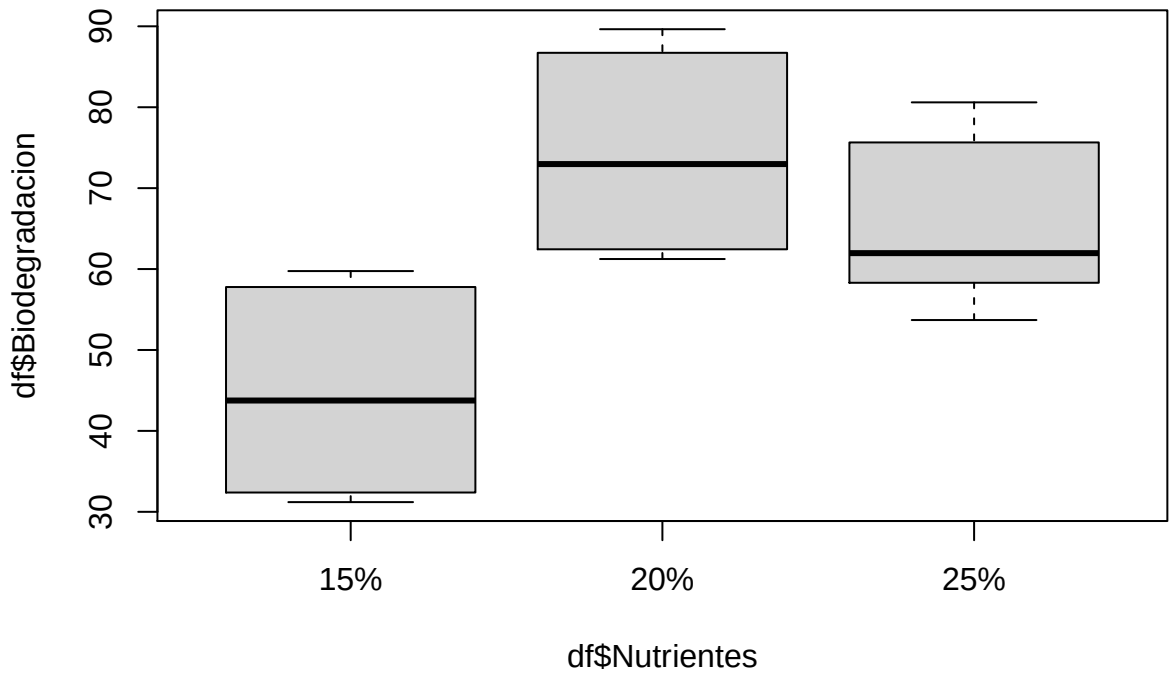
1 First analysis

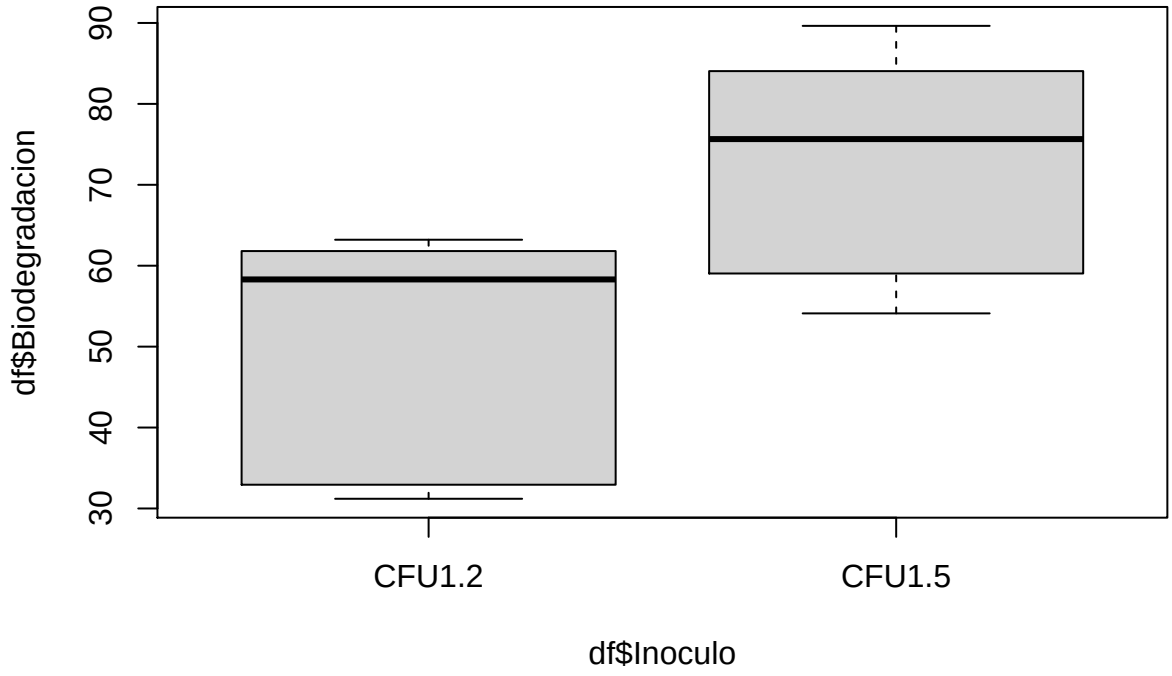
Considering the following experimental approach: "the data come from conducting an experiment with *completely randomized design and factorial structure* in which three replicates were performed per experimental condition", the following methodology will be used:

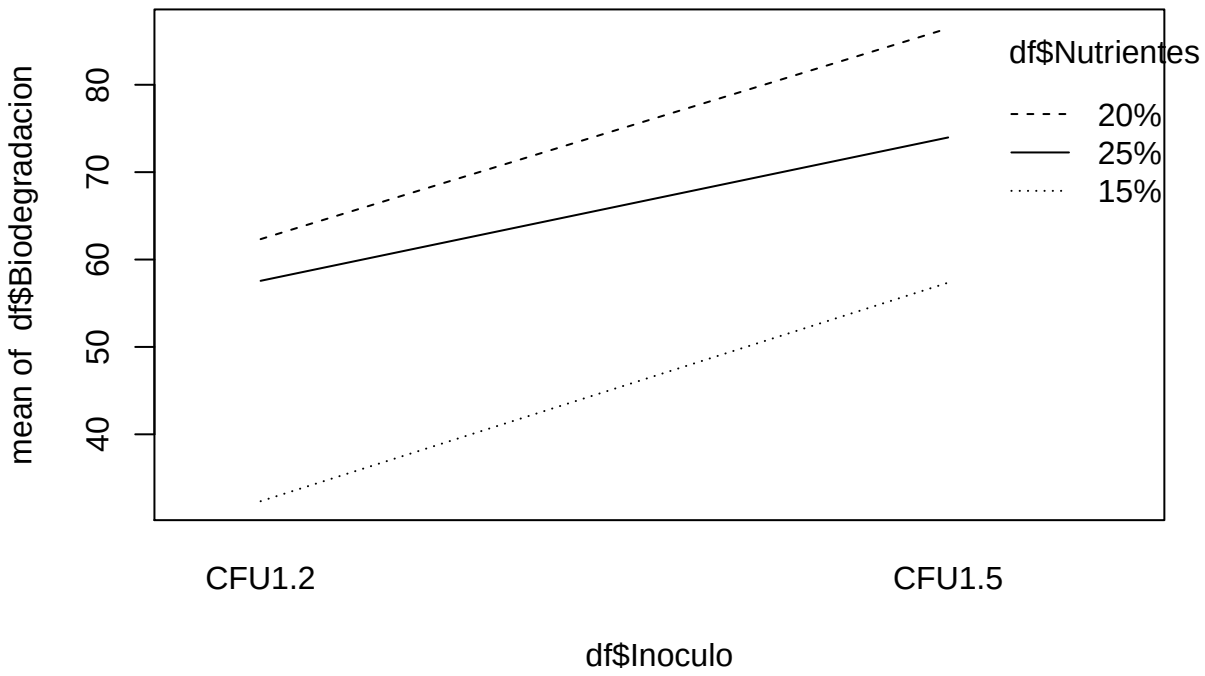
```
##   Nutrientes Inoculo Biodegradacion
## 1        15%  CFU1.2          32.47
## 2        15%  CFU1.2          32.30
## 3        15%  CFU1.2          33.40
## 4        15%  CFU1.2          31.20
## 5        15%  CFU1.5          57.23
## 6        15%  CFU1.5          54.11
```

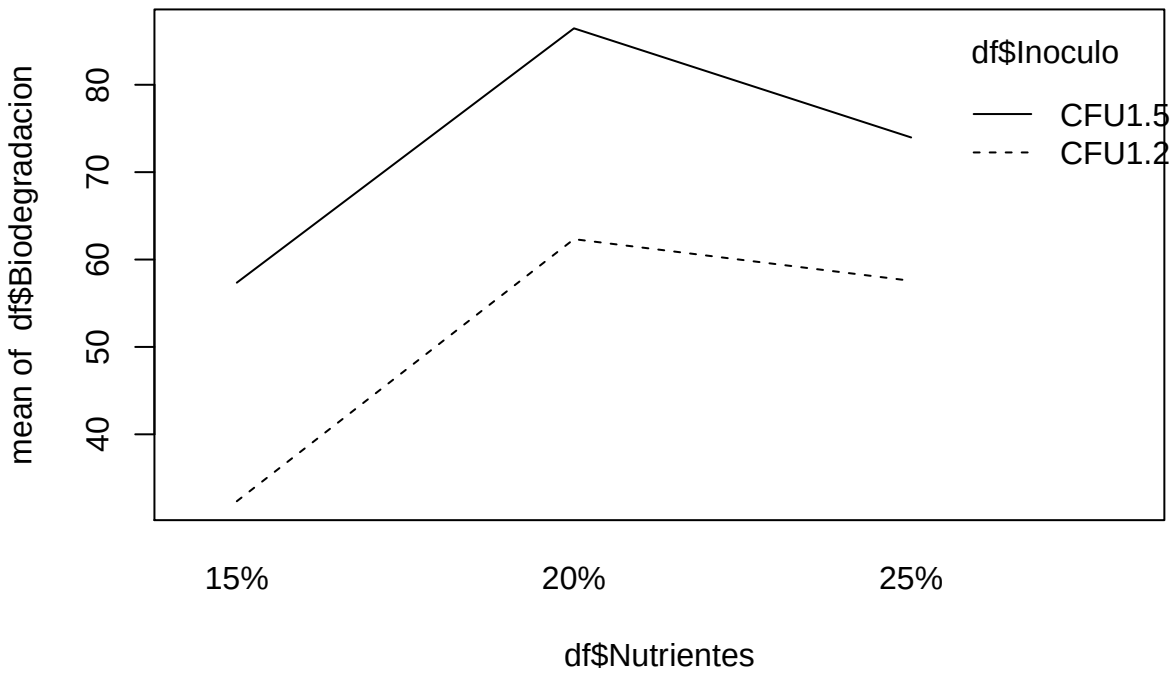
```
## 'data.frame':    24 obs. of  3 variables:
##  $ Nutrientes    : chr  "15%" "15%" "15%" "15%" ...
##  $ Inoculo       : chr  "CFU1.2" "CFU1.2" "CFU1.2" "CFU1.2" ...
##  $ Biodegradacion: num  32.5 32.3 33.4 31.2 57.2 ...
```

1.1 Description of each Factor and Interaction






mean of df$Biodegradacion
40
50
60
70
80
df$Nutrientes
20%
25%
15%
CFU1.2
CFU1.5
df$Inoculo



1.2 Modeling

```
modelo<-aov(Biodegradacion~Inoculo*Nutrientes, data = df)
summary(modelo)
```

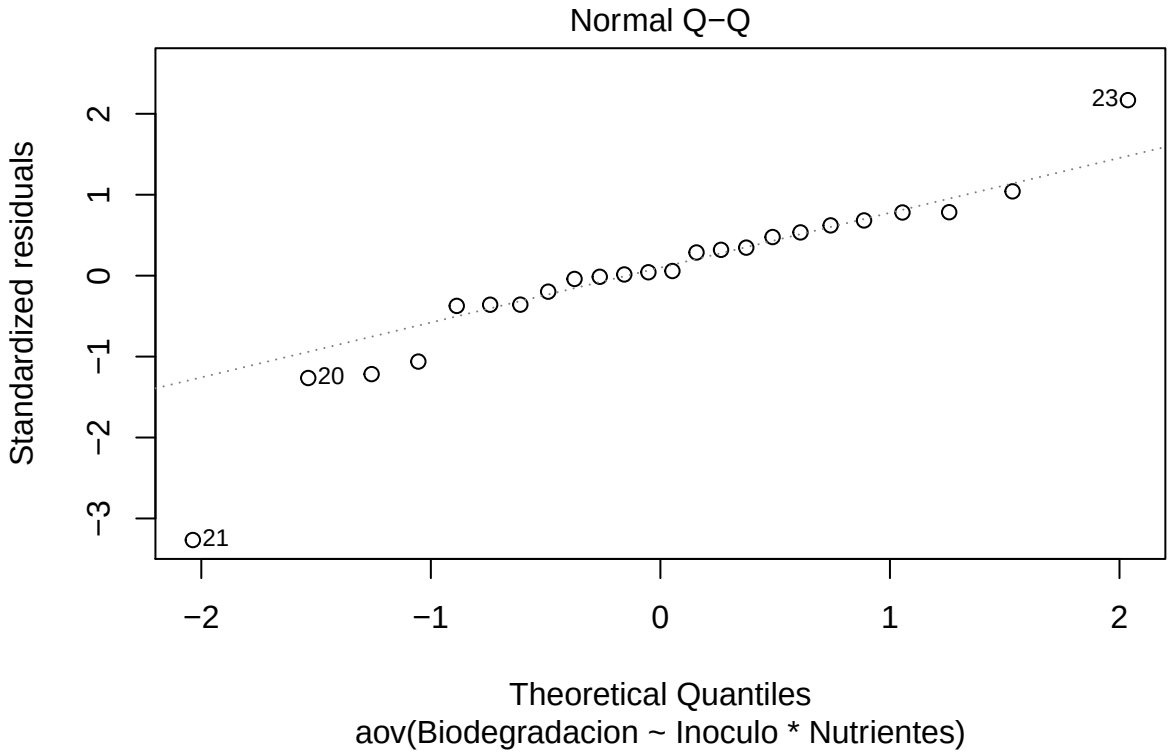
```
##                     Df Sum Sq Mean Sq F value   Pr(>F)
## Inoculo              1   2864  2863.9 230.002 1.07e-11 ***
## Nutrientes           2   3694  1846.8 148.323 6.56e-12 ***
## Inoculo:Nutrientes   2     90    44.8   3.597   0.0485 *
## Residuals           18    224    12.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1.3 Supuestos del Modelo

```
##  [1] CFU1.2.15% CFU1.2.15% CFU1.2.15% CFU1.2.15% CFU1.5.15% CFU1.5.15%
##  [7] CFU1.5.15% CFU1.5.15% CFU1.2.20% CFU1.2.20% CFU1.2.20% CFU1.2.20%
## [13] CFU1.5.20% CFU1.5.20% CFU1.5.20% CFU1.5.20% CFU1.2.25% CFU1.2.25%
## [19] CFU1.2.25% CFU1.2.25% CFU1.5.25% CFU1.5.25% CFU1.5.25% CFU1.5.25%
## 6 Levels: CFU1.2.15% CFU1.5.15% CFU1.2.20% CFU1.5.20% ... CFU1.5.25%
```

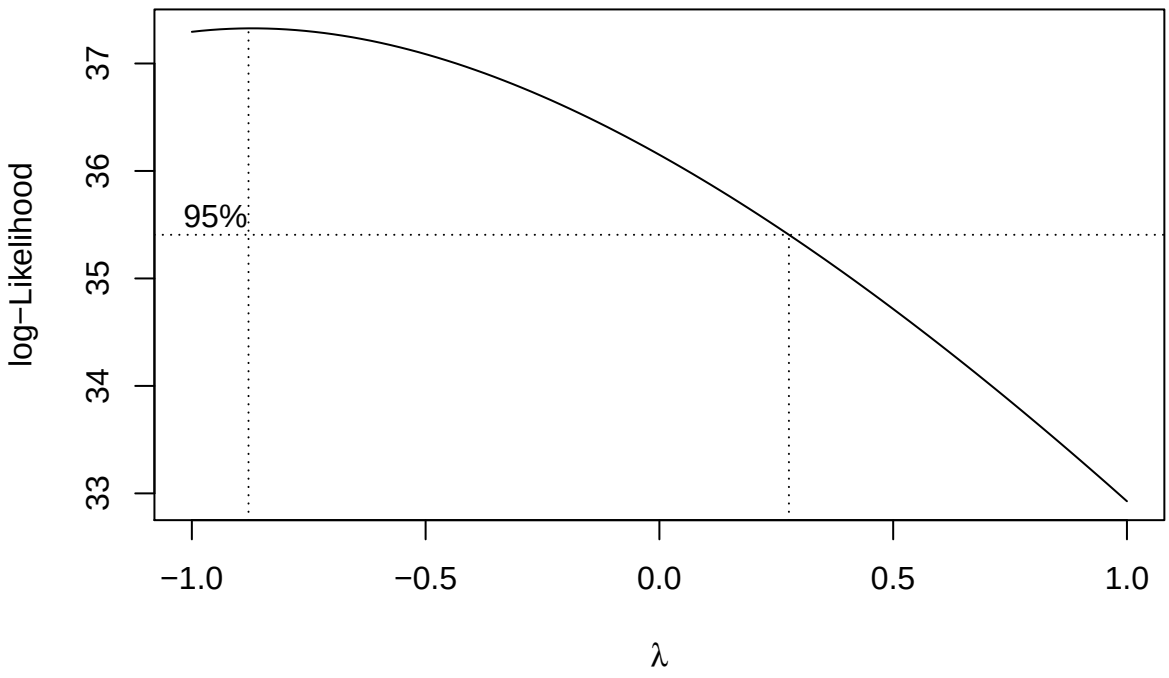
```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##       Df F value Pr(>F)
## group  5  1.2666 0.3207
##       18
```




```
##
##  Shapiro-Wilk normality test
##
## data:  modelo$residuals
## W = 0.89431, p-value = 0.01634
```

NOTE: since the normality of the residuals of the model is NOT fulfilled, but the homoscedasticity IS, we must adjust some of the variables, and model again.




```
## [1] -0.8787879
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)
## Inoculo              1   3691    3691  427.70 5.40e-14 ***
## Nutrientes           2   5755    2878  333.48 5.98e-15 ***
## Inoculo:Nutrientes   2   1241     620   71.89 2.61e-09 ***
## Residuals           18    155       9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo2$residuals
## W = 0.93457, p-value = 0.1233
```

The restriction of normality of the residuals was lifted, so we went on to analyze the new model *model2*

```
## Analysis of Variance Table
##
## Response: Bio
##                     Df Sum Sq Mean Sq F value    Pr(>F)
## Inoculo              1 3690.8  3690.8 427.702 5.400e-14 ***
## Nutrientes           2 5755.4  2877.7 333.477 5.977e-15 ***
## Inoculo:Nutrientes   2 1240.7   620.4  71.888 2.613e-09 ***
```

```
## Residuals          18  155.3     8.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1.4 Means by levels and by experimental conditions

```
## Tables of means
## Grand mean
##
## 2357.643
##
##  Inoculo
## Inoculo
## CFU1.2 CFU1.5
## 2345.2 2370.0
##
##  Nutrientes
## Nutrientes
##    15%    20%    25%
## 2336.0 2371.5 2365.4
##
##  Inoculo:Nutrientes
##         Nutrientes
## Inoculo  15%    20%    25%
##   CFU1.2 2313.5 2363.5 2358.7
##   CFU1.5 2358.6 2379.5 2372.1
```

And to obtain the estimates of the main effects and interactions, we will do the following:

```
## Tables of effects
##
##  Inoculo
## Inoculo
##  CFU1.2  CFU1.5
## -12.401  12.401
##
##  Nutrientes
## Nutrientes
##     15%     20%     25%
## -21.620  13.834   7.786
##
##  Inoculo:Nutrientes
##         Nutrientes
## Inoculo  15%     20%     25%
##   CFU1.2 -10.140   4.419   5.722
##   CFU1.5  10.140  -4.419  -5.722
```

Since the interaction has been significant, we must do one more test:

```
##
## Study: modelo3 ~ "condition"
##
```

```
## HSD Test for Bio
##
## Mean Square Error:  8.629445
##
## condition,  means
##
##                  Bio       std r      Min      Max
## CFU1.2.15% 2313.481 2.8098605 4 2309.874 2316.716
## CFU1.2.20% 2363.494 0.7402953 4 2362.496 2364.281
## CFU1.2.25% 2358.750 3.1269075 4 2354.535 2361.268
## CFU1.5.15% 2358.564 2.5839814 4 2355.022 2361.073
## CFU1.5.20% 2379.459 1.4977369 4 2377.605 2380.997
## CFU1.5.25% 2372.109 4.9634253 4 2364.959 2376.442
##
## Alpha: 0.05 ; DF Error: 18
## Critical Value of Studentized Range: 4.49442
##
## Minimun Significant Difference: 6.601386
##
## Treatments with the same letter are not significantly different.
##
##                  Bio groups
## CFU1.5.20% 2379.459      a
## CFU1.5.25% 2372.109      b
## CFU1.2.20% 2363.494      c
## CFU1.2.25% 2358.750      c
## CFU1.5.15% 2358.564      c
## CFU1.2.15% 2313.481      d
```

We can affirm that there are no relevant differences in the treatments with the following characteristics:

- CFU1.2.20% 2363.494
- CFU1.2.25% 2358.750
- CFU1.5.15% 2358.564

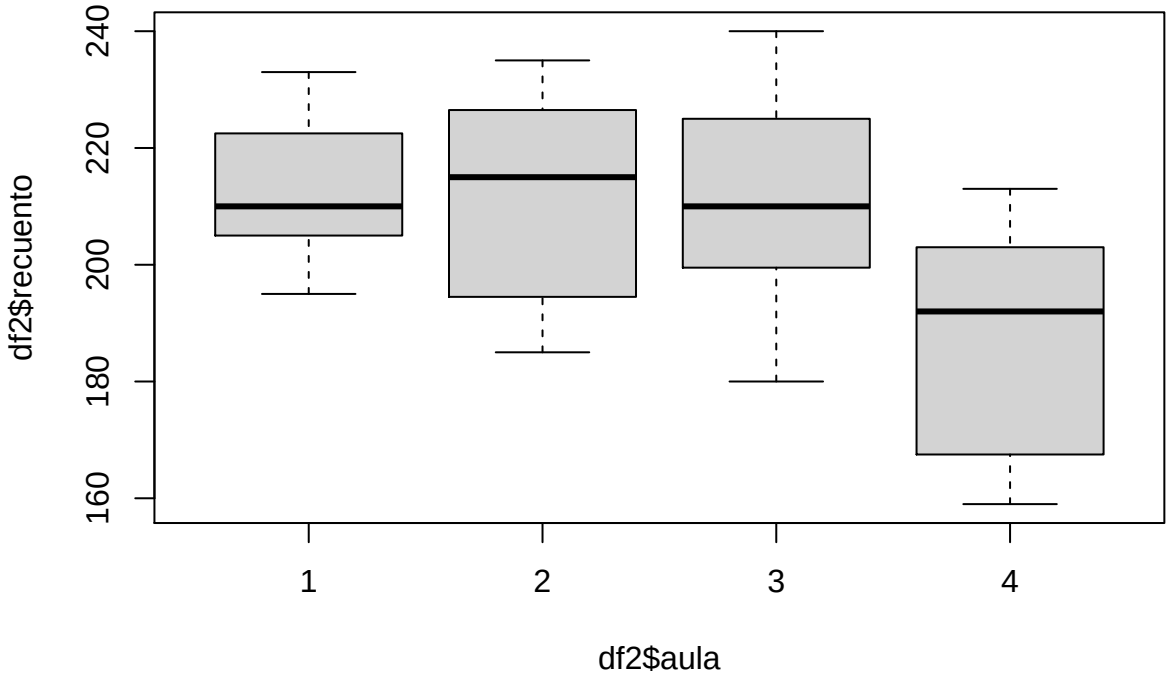
2 Second analysis

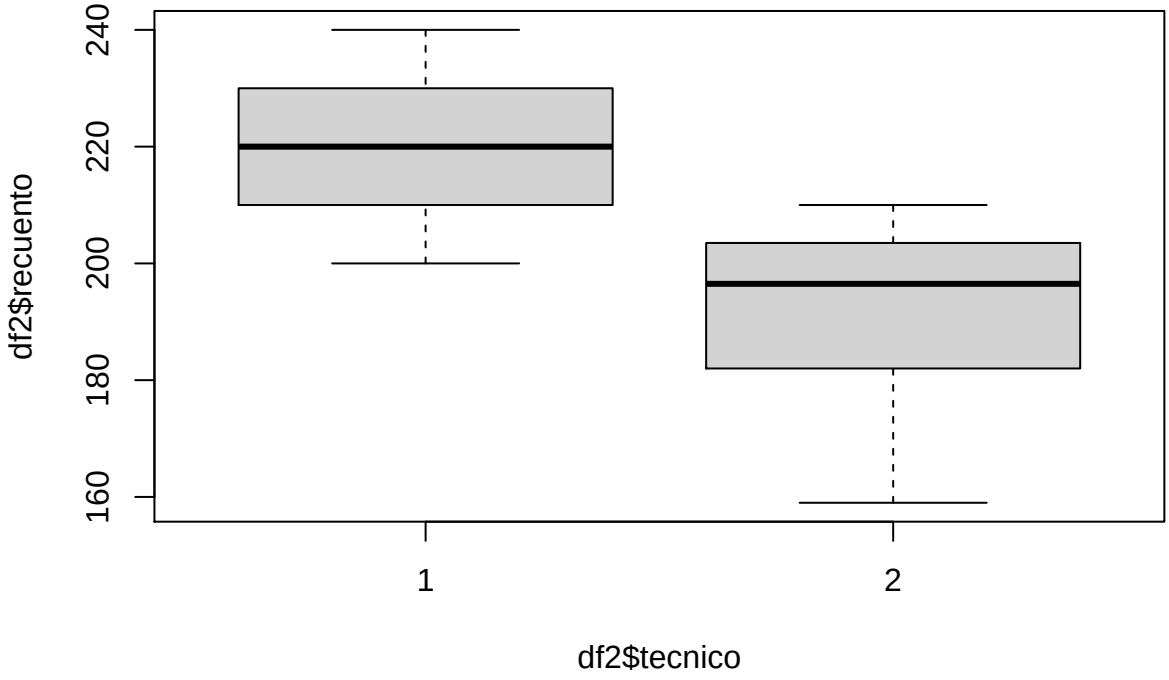
Considering that the experiment has the following design and implementation characteristics "the data presented refer to the determination of aerobic colonies in four *randomly chosen computer classrooms* of the University, whose samples were taken by *two laboratory technicians chosen at Random* of the Department of Microbiology", we will use the following methodology (or steps).

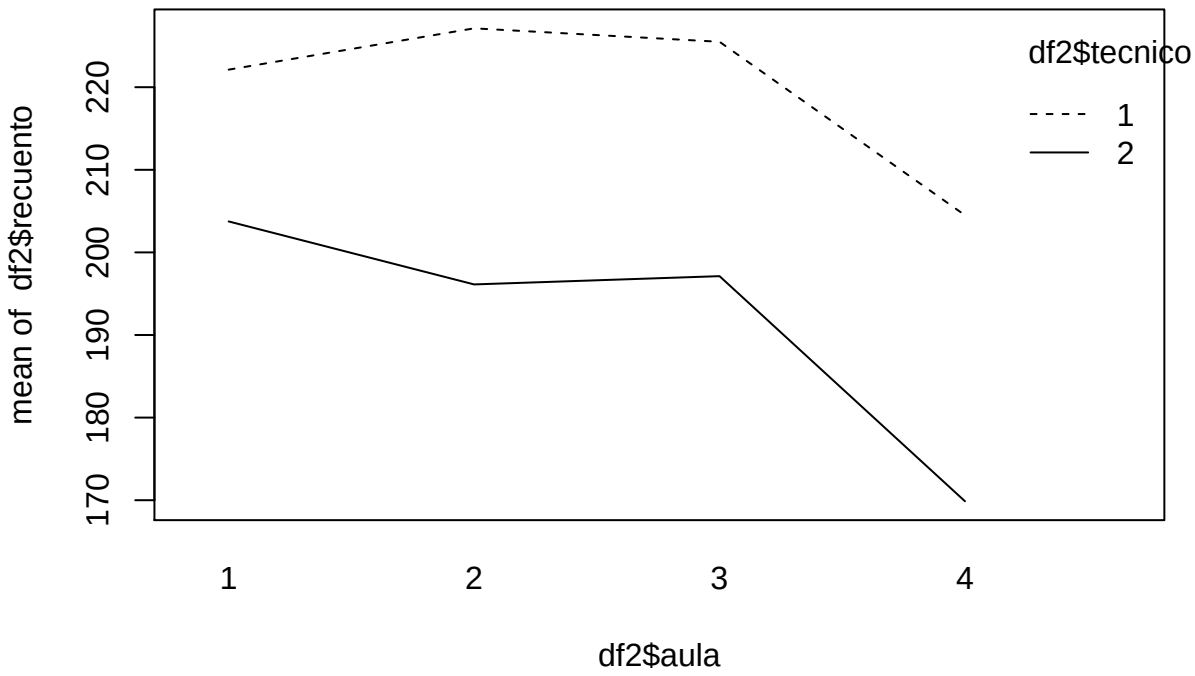
```
##   aula tecnico recuento
## 1    1       1      233
## 2    1       1      227
## 3    1       1      210
## 4    1       1      217
## 5    1       1      225
## 6    1       1      230
```

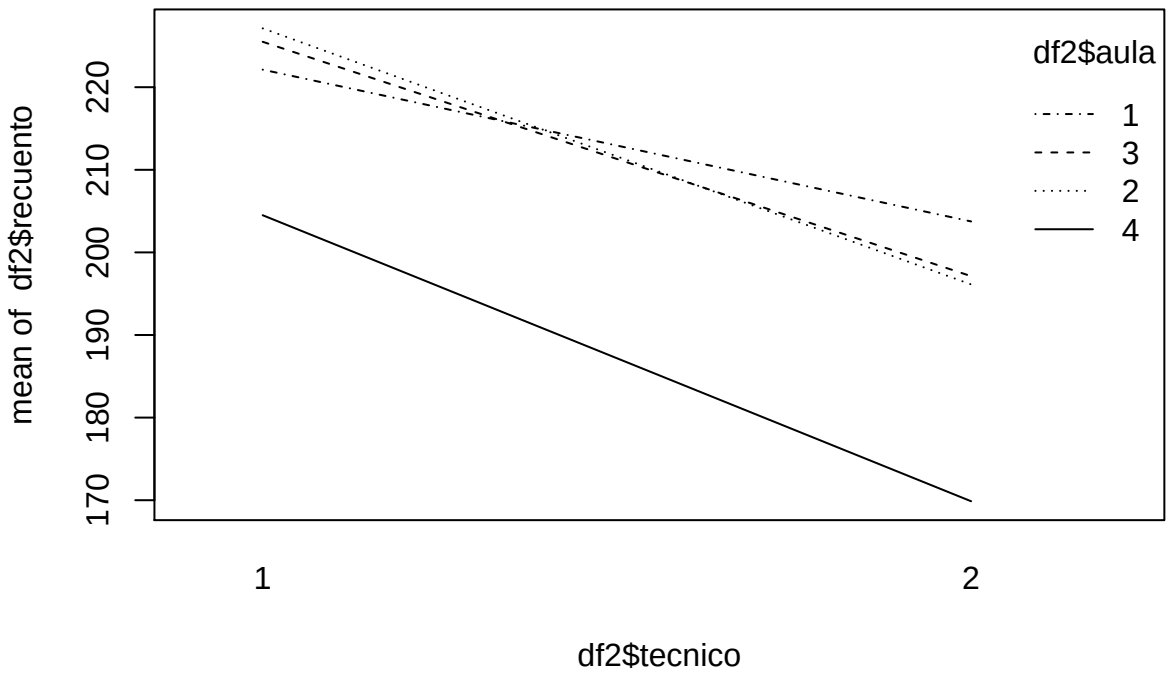
```
## 'data.frame':    64 obs. of  3 variables:
##  $ aula    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ tecnico : int  1 1 1 1 1 1 1 1 2 2 ...
##  $ recuento: int  233 227 210 217 225 230 215 220 209 205 ...
```

2.1 Description of each Factor and Interaction






mean of df2$recuento
170
180
190
200
210
220
1
2
3
4
df2$aula
df2$tecnico
1
2



2.2 Modeling

```
##
## Call:
## lm(formula = recuento ~ r(tecnico) * r(aula), data = df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.1250  -5.9375  -0.1875   4.9688  14.5000
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        205.7656     0.9893 207.986  < 2e-16 ***
## tecnico(1)          14.0469     1.7419   8.064 6.06e-11 ***
## aula(1)              7.1719     3.0170   2.377  0.02089 *
## aula(2)              5.8594     3.0170   1.942  0.05716 .
## aula(3)              5.5469     3.0170   1.839  0.07129 .
## tecnico(1):aula(1)  -4.8594     1.7136  -2.836  0.00635 **
## tecnico(1):aula(2)   1.4531     1.7136   0.848  0.40004
## tecnico(1):aula(3)   0.1406     1.7136   0.082  0.93489
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 7.915 on 56 degrees of freedom
## Multiple R-squared: 0.8545,
```

```
## Adjusted R-squared: 0.8363
## F-statistic: 46.97 on 7 and 56 DF,  p-value: < 2.2e-16
```

2.3 Model Assumptions

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  7  1.1014  0.375
##       56
```

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo4$residuals
## W = 0.98421, p-value = 0.5866
```

NOTE: We confirm that the model complies with the main assumptions, so there is no need to adjust the variables. Therefore, we will assume the assumptions of the modeling, and we will proceed to solve the contrasts of interest. The contrasts to resolve are:

- $H_0 : \sigma_A^2 = 0$
- $H_0 : \sigma_B^2 = 0$
- $H_0 : \sigma_{AB}^2 = 0$

where the alternative hypotheses, in each case, state that the variance is greater than zero.

```
## Analysis of variance (unrestricted model)
## Response: recuento
##               Mean Sq   Sum Sq Df F value Pr(>F)
## tecnico      12628.14 12628.14  1   65.03 0.0040
## aula          2462.31  7386.92  3   12.68 0.0328
## tecnico:aula   194.18   582.55  3    3.10 0.0339
## Residuals       62.64  3507.87 56       -      -
##
##                Err.term(s) Err.df VC(SS)
## 1 tecnico              (3)      3  388.6
## 2 aula                 (3)      3  141.8
## 3 tecnico:aula         (4)     56   16.4
## 4 Residuals              -      -   62.6
## (VC = variance component)
##
##             Expected mean squares
## tecnico       (4) + 8 (3) + 32 (1)
## aula          (4) + 8 (3) + 16 (2)
## tecnico:aula  (4) + 8 (3)
## Residuals     (4)
```

We conclude that the technical and classroom factors, and the interaction are significant. It is important to highlight that, unlike the fixed effects models, we will now NOT make multiple comparisons, we will proceed to evaluate the variance components individually.

a) Error variance: 62.64

b) Interaction variance: 16.4
c) Variance of the technical factor: 388.6
d) Variance of the classroom factor: 141.8

The estimate of the variance of the response variable is = 62.64 + 16.4 + 388.6 + 141.8 = **609.44** and therefore, the percentage of the variance components are:

a) Error Variance Component: 10.27%
b) Interaction Variance Component: 2.7%
c) Component of the Variance of the technical factor: 63.76%
d) Variance component of the classroom factor: 23.26%

We conclude that the *technical factor* is the main cause of variability in the response (63.76%%), followed by the *classroom factor* (23.26%).