

# PEC2 - Diseño y Análisis de Experimentos

Diego Vallarino

17/11/2021

## 1 Ejercicio UNO

Considerando el siguiente planteo de experimento: “los datos provienen de la realización de un experimento con *diseño completamente al azar y estructura factorial* en el que se realizaron tres réplicas por condición experimental”, se utilizará la siguiente metodología:

### 1.1 Levantar la Librerías

```
library(car)
library(MASS)
library(psych)
library(agricolae)
library(mixlm)
```

### 1.2 Cargamos los datos

```
datos_bio <- read.csv("D:/Master en BioEstadística/Materias/3.Análisis de Experimentos/PEC2/datos_bio.csv")
df<-datos_bio
head(df)
```

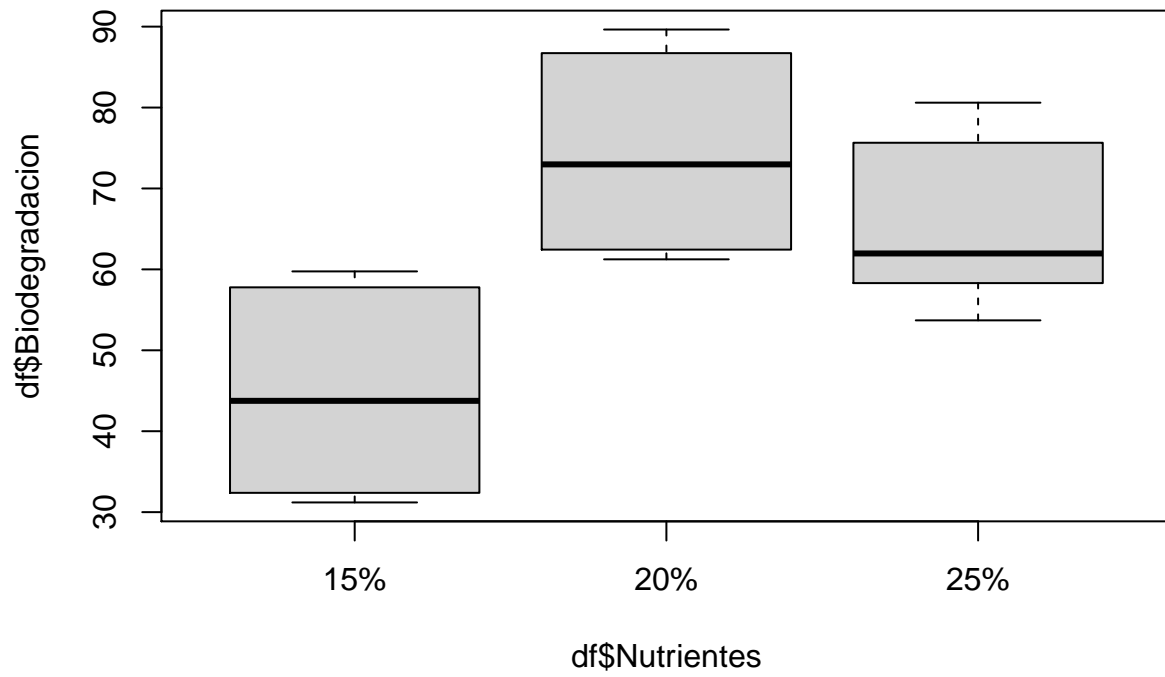
```
##   Nutrientes Inoculo Biodegradacion
## 1      15% CFU1.2      32.47
## 2      15% CFU1.2      32.30
## 3      15% CFU1.2      33.40
## 4      15% CFU1.2      31.20
## 5      15% CFU1.5      57.23
## 6      15% CFU1.5      54.11
```

```
str(df)
```

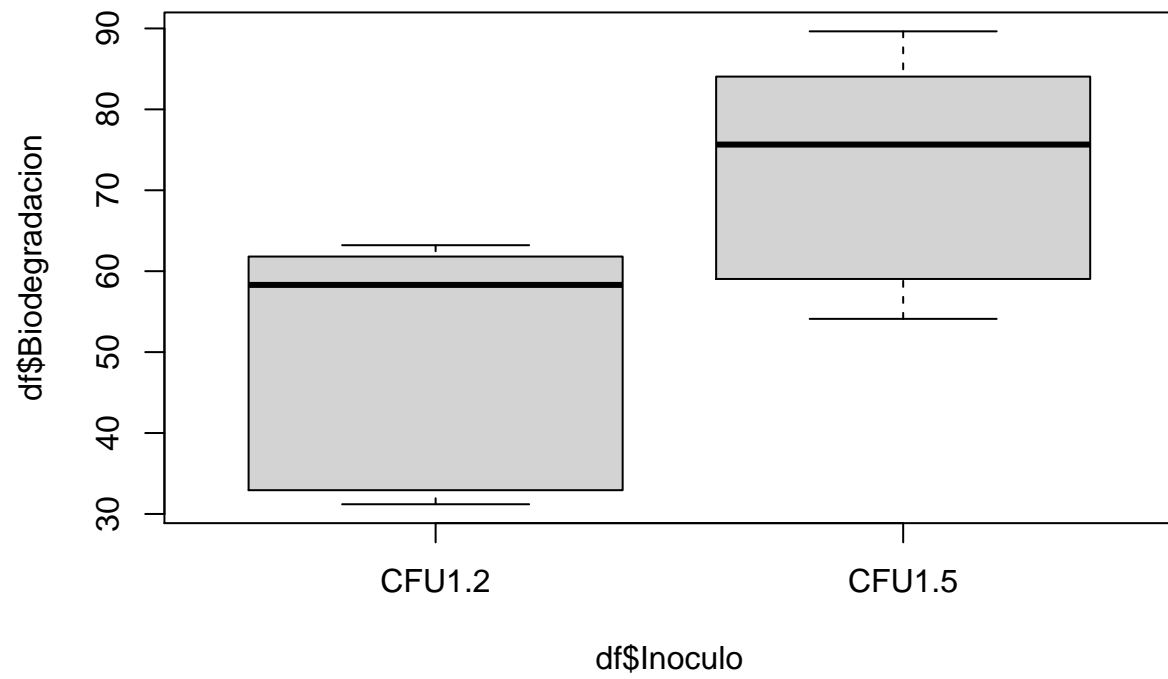
```
## 'data.frame':   24 obs. of  3 variables:
## $ Nutrientes    : chr  "15%" "15%" "15%" "15%" ...
## $ Inoculo       : chr  "CFU1.2" "CFU1.2" "CFU1.2" "CFU1.2" ...
## $ Biodegradacion: num  32.5 32.3 33.4 31.2 57.2 ...
```

### 1.3 Descriptiva de cada Factor y de la Interaccion

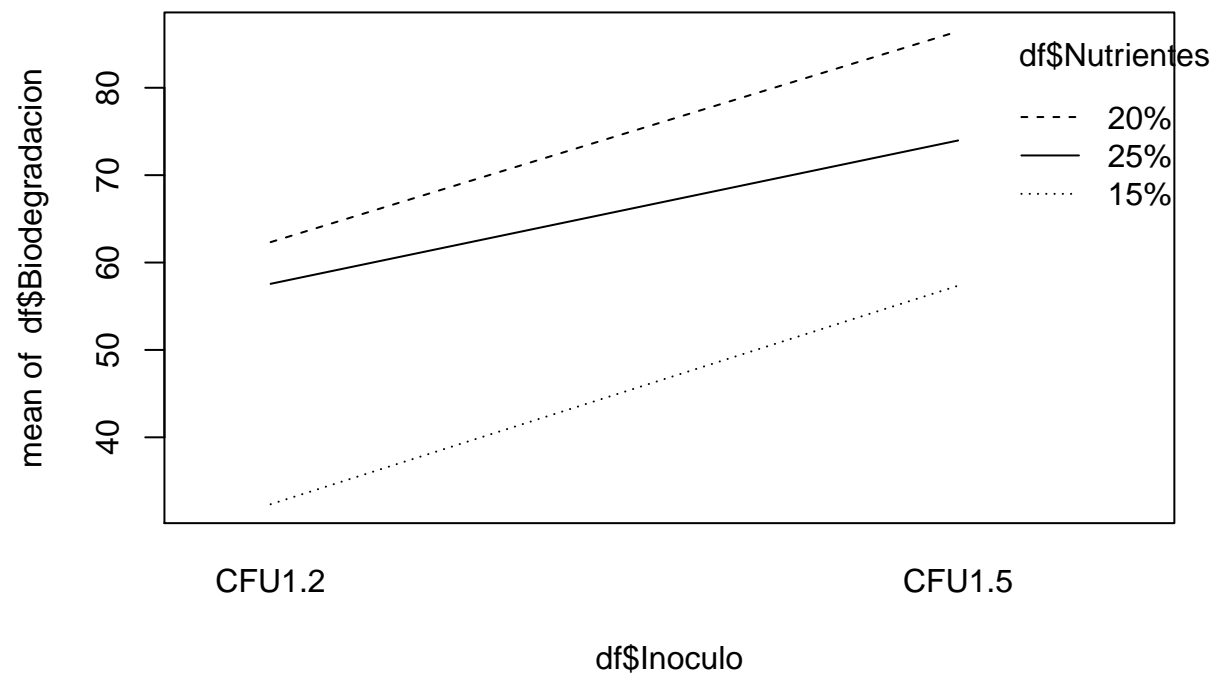
```
boxplot(df$Biodegradacion~df$Nutrientes)
```



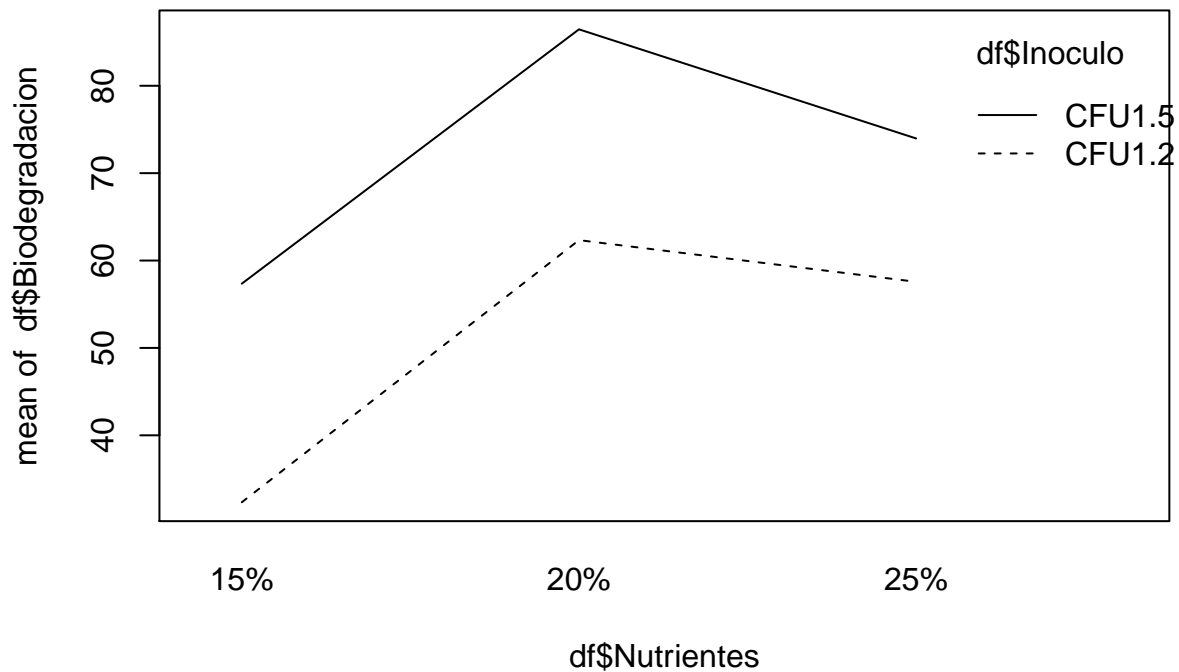
```
boxplot(df$Biodegradacion~df$Inoculo)
```



```
interaction.plot(df$Inoculo, df$Nutrientes, df$Biodegradacion)
```



```
interaction.plot(df$Nutrientes, df$Inoculo, df$Biodegradacion)
```



## 1.4 Modelamos

```
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
```

## 1.5 Supuestos del Modelo

```
condition <- with(df, interaction(Inoculo, Nutrientes))
condition
```

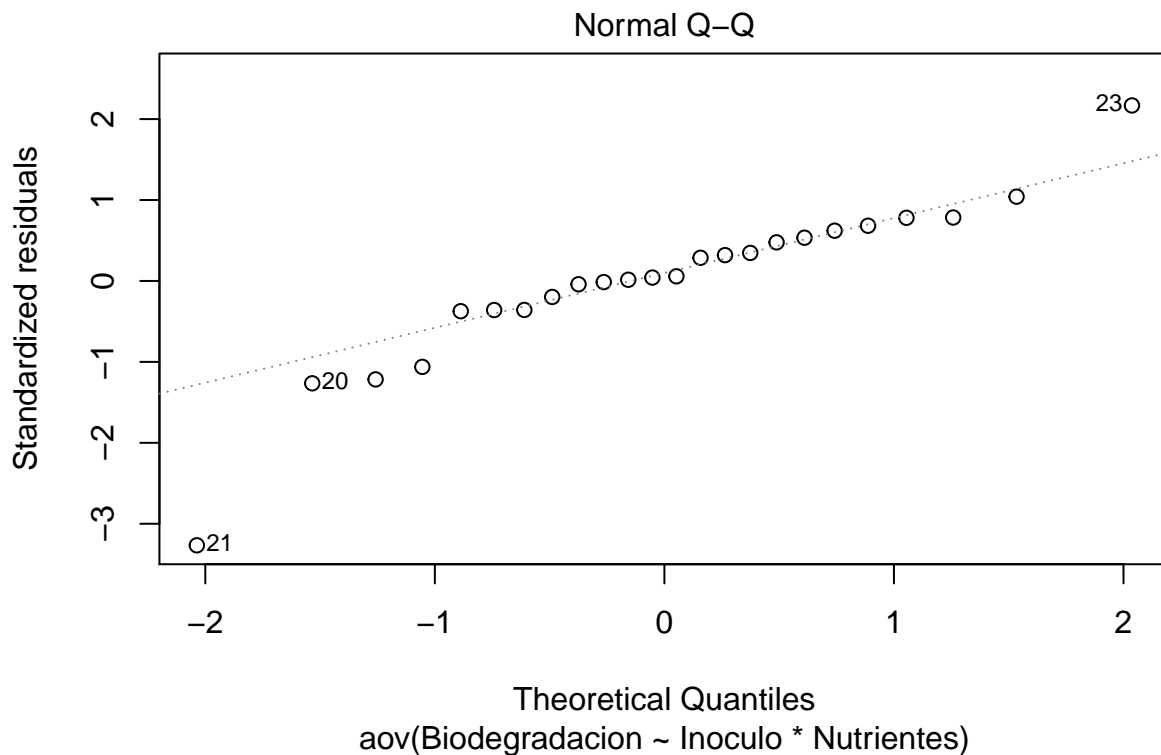
```
## [1] CFU1.2.15% CFU1.2.15% CFU1.2.15% CFU1.2.15% CFU1.5.15% CFU1.5.15%
## [7] CFU1.5.15% CFU1.5.15% CFU1.2.20% CFU1.2.20% CFU1.2.20% CFU1.2.20%
## [13] CFU1.5.20% CFU1.5.20% CFU1.5.20% CFU1.5.20% CFU1.2.25% CFU1.2.25%
```

```
## [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%
```

```
leveneTest(Biodegradacion ~ condition, data = df)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 5  1.2666 0.3207
##      18
```

```
plot(modelo, which = 2)
```

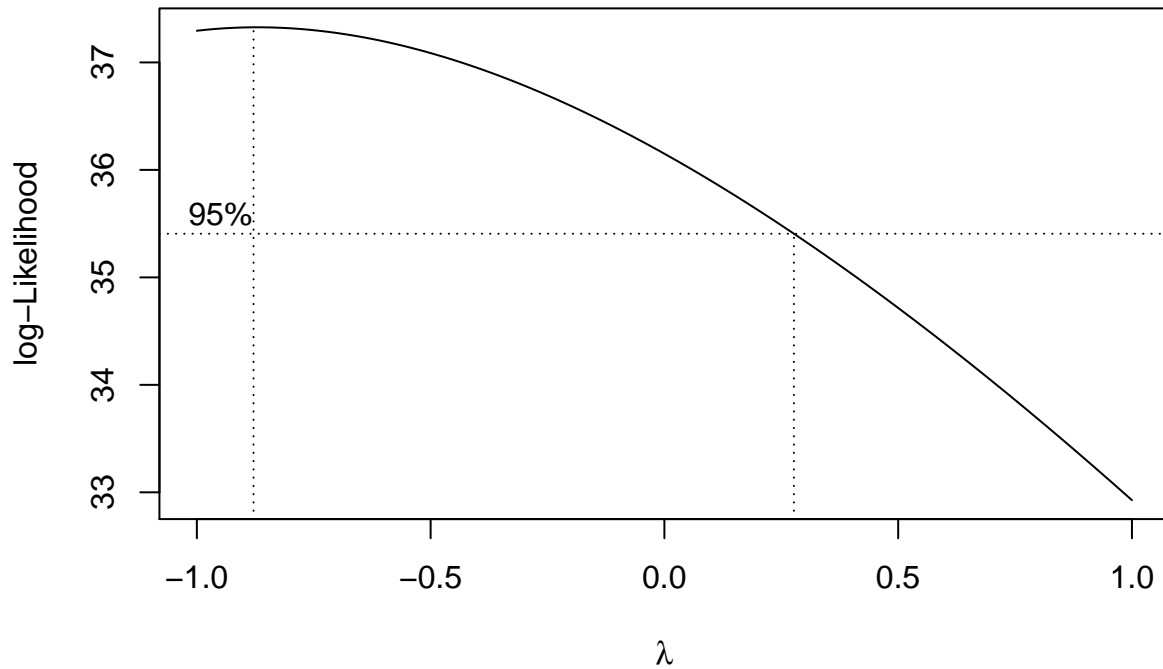


```
shapiro.test(modelo$residuals)
```

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

NOTA: como NO se cumple la normalidad de los residuos del modelo pero SI la homocedasticidad, debemos ajustar algunas de las variables, y modelar nuevamente.

```
bct<-boxcox(Biodegradacion~Inoculo*Nutrientes,lambda = seq(-1, 1, length = 10),data=df,plotit=T)
```



```
lambda<-bct$x[which.max(bct$y)]
lambda
```

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

```
gm<-geometric.mean(df$Biodegradacion)
lambda<- -0.88
df$Bio<-(df$Biodegradacion^lambda-1)/(lambda*gm^(lambda-1))

modelo2<-aov(Bio~Inoculo*Nutrientes, data = df)
summary(modelo2)
```

```
##               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.test(modelo2$residuals)
```

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

Se levantó la restricción de normalidad de los residuos, por lo que pasamos a analizar el nuevo modelo *modelo2*

```
anova(modelo2)
```

```
## 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
```

## 1.6 Las medias por niveles y por condiciones experimentales

```
model.tables(modelo2, type = "mean")
```

```
## 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
```

Y para obtener las estimación de los efectos principales e interacciones, haremos lo siguiente:



```
model.tables(modelo2, type = "effects")
```

```
## 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
```

Dado que la interacción ha sido significativa, debemos hacer un test más:

```
modelo3 <- aov(Bio ~ condition, data = df)
HSD.test(modelo3, "condition", console = TRUE)
```

```
##
## 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
##
## Minimum 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
```

Podemos afirmar que no existen diferencias relevante en los tratamientos con las siguientes características:

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

## 2 Ejercicio DOS

Considernado que el experimento cuenta con las siguientes características de diseño e implementación “los datos presentados se refieren a la determinación de colonias aeróbicas en cuatro *aulas informáticas elegidas al azar* de la Universidad, cuyas muestras fueron tomadas por *dos técnicos de laboratorio elegidos al azar* del Departamento de Microbiología”, utilizaremos la siguiente metodología (o pasos).

### 2.1 Cargamos los datos

```
micro <- read.csv("D:/Master en BioEstadistica/Materias/3.Analisis de Experimentos/PEC2/micro.csv", sep=";", as.is=T)
df2<-micro
head(df2)
```

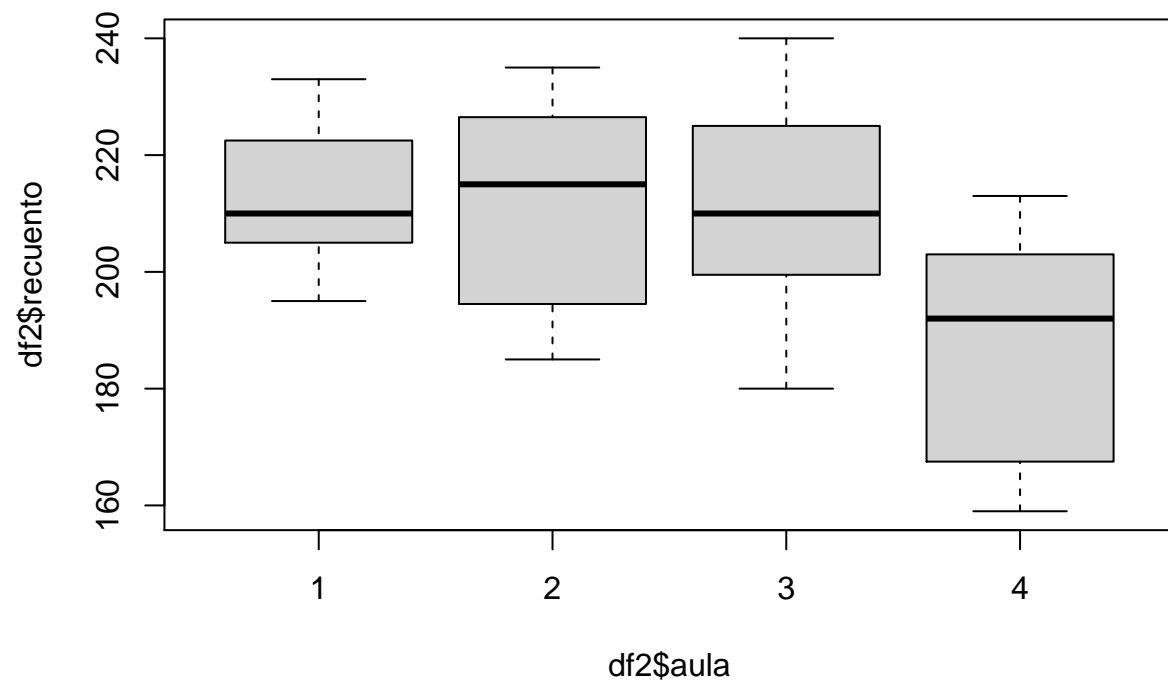
```
##   aula tecnico recuento
## 1    1         1       233
## 2    1         1       227
## 3    1         1       210
## 4    1         1       217
## 5    1         1       225
## 6    1         1       230
```

```
str(df2)
```

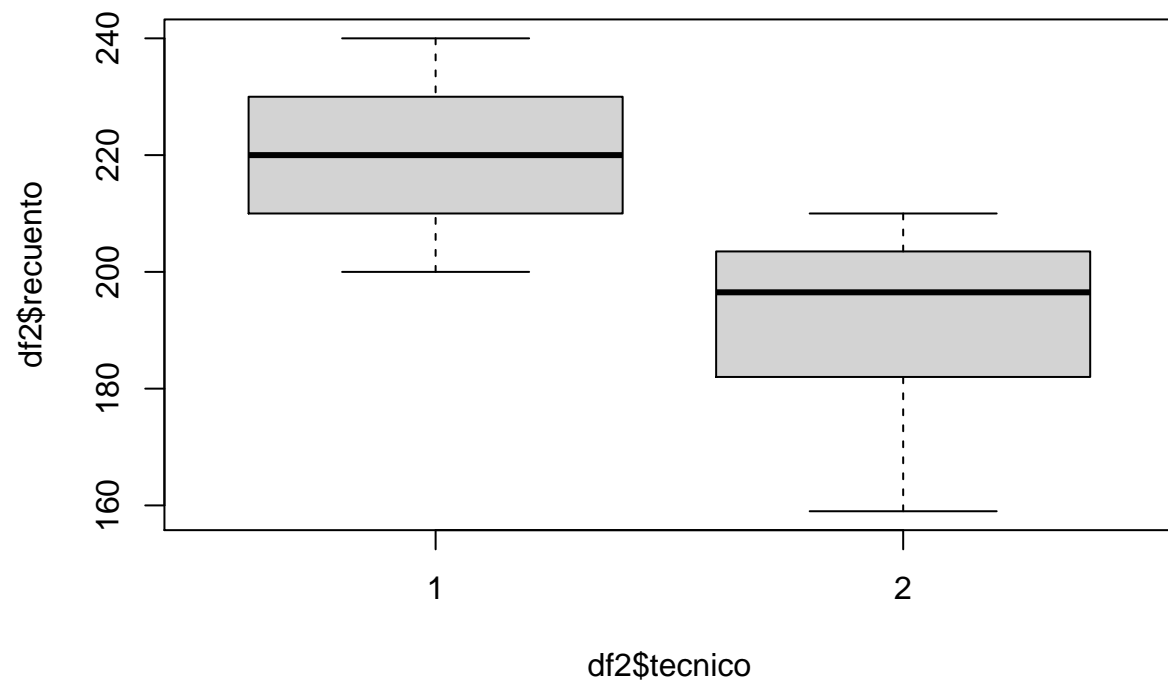
```
## 'data.frame':   64 obs. of  3 variables:
## $ aula      : int  1 1 1 1 1 1 1 1 1 1 ...
## $ tecnico   : int  1 1 1 1 1 1 1 1 2 2 ...
## $ recuento  : int  233 227 210 217 225 230 215 220 209 205 ...
```

### 2.2 Descriptiva de cada Factor y de la Interaccion

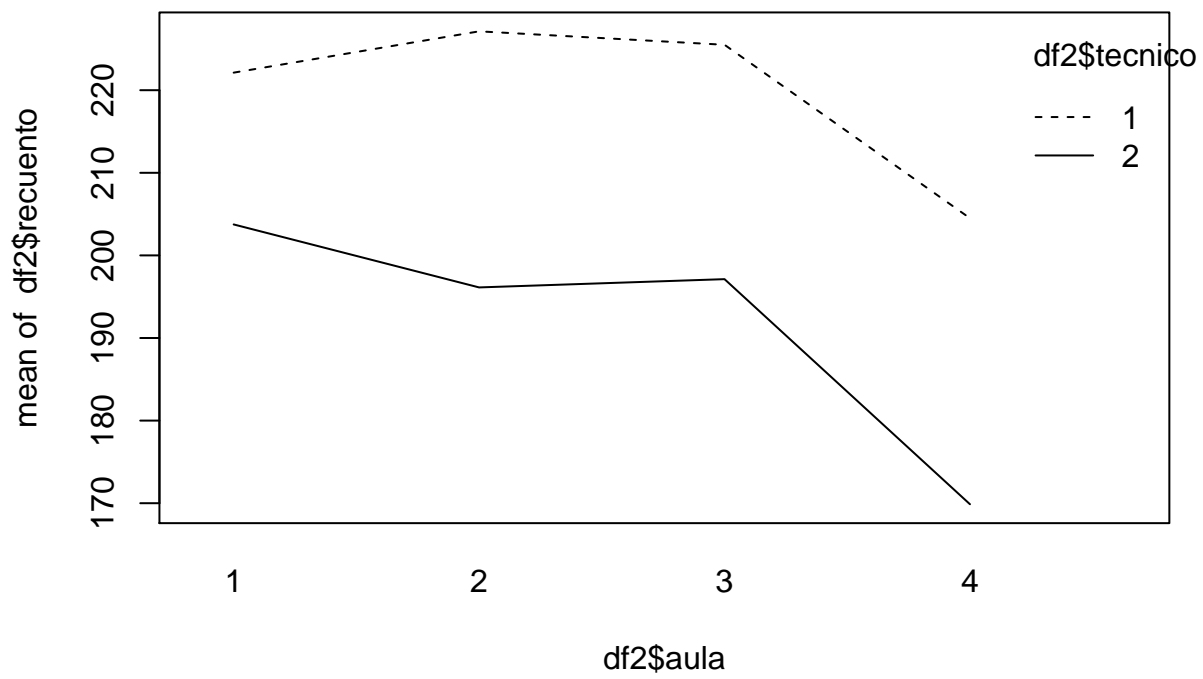
```
boxplot(df2$recuento~df2$aula)
```



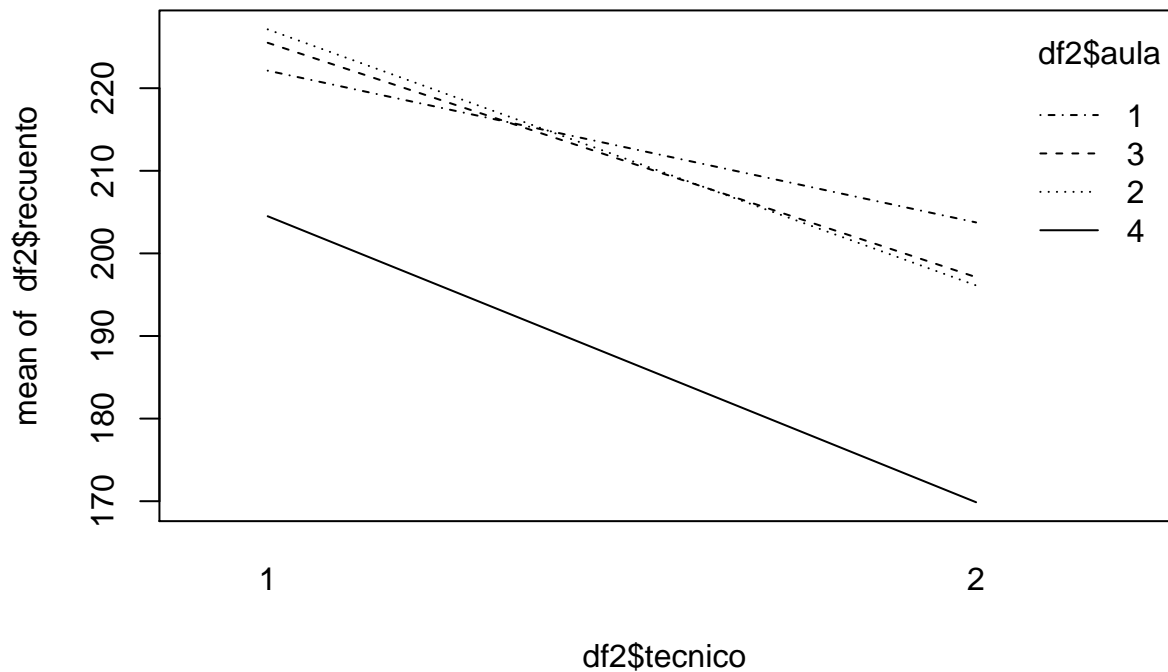
```
boxplot(df2$recuento~df2$aula)
```



```
interaction.plot(df2$aula, df2$tecnico, df2$recuento)
```



```
interaction.plot(df2$tecnico,df2$aula,df2$recuento)
```



## 2.3 Modelamos

```
df2$aula<-as.factor(df2$aula)
df2$tecnico<-as.factor(df2$tecnico)
modelo4<-lm(recuento ~ r(tecnico)*r(aula), data = df2)
summary(modelo4)
```

```
##
## 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
```

## 2.4 Supuestos del Modelo

```
condition2 <- with(df2, interaction(aula, tecnico))
leveneTest(recuento ~ condition2, data = df2)
```

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

```
shapiro.test(modelo4$residuals)
```

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

NOTA: confirmamos que el modelo cumple con los principales supuestos, por lo que no hay que ajustar las variables. Por ende, asumiremos las suposiciones de la modelización, y pasaremos a resolver los contrastes de interés. Los contrastes a resolver son:

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

donde las hipótesis alternativas, en cada caso, afirman que la varianza es mayor que cero.

```
Anova(modelo4, type="III")
```

```
## 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)
```

Concluimos que los factores tecnico y aula, y la interacción son significativos. Es importante, resaltar que a diferencia de los modelos de efectos fijos ahora NO haremos comparaciones múltiples, pasaremos a evaluar las componentes de la varianza individualmente.

- a) Varianza del error: 62.64
- b) Varianza de la interacción: 16.4
- c) Varianza del factor tecnico: 388.6
- d) Varianza del factor aula: 141.8

La estimación de la varianza de la variable respuesta es  $= 62.64 + 16.4 + 388.6 + 141.8 = \mathbf{609.44}$  y por tanto, el porcentaje de las componentes de la varianza son:

- a) Componente de la Varianza del error: 10.27%
- b) Componente de la Varianza de la interacción: 2.7%
- c) Componente de la Varianza del factor tecnico: 63.76%
- d) Componente de la Varianza del factor aula: 23.26%

Concluimos, que el *factor tecnico* es la causa principal de variabilidad en la respuesta (63.76%%) , seguido del *factor aula* (23.26%).