

Anova

Cleber Perez

2024-08-27

El rendimiento

En un instituto se han matriculado 36 estudiantes. Se desea explicar el rendimiento de ciencias naturales en función de dos variables: género y metodología de enseñanza. La metodología de enseñanza se analiza en tres niveles: explicación oral y realización del experimento (1er nivel) explicación oral e imágenes (2º nivel) y explicación oral (tercer nivel). En los alumnos matriculados había el mismo número de chicos que de chicas, por lo que formamos dos grupos de 18 sujetos; en cada uno de ellos, el mismo profesor aplicará a grupos aleatorios de 6 estudiantes las 3 metodologías de estudio. A fin de curso los alumnos son sometidos a la misma prueba de rendimiento. Los resultados son los siguientes:

Introduciendo Datos

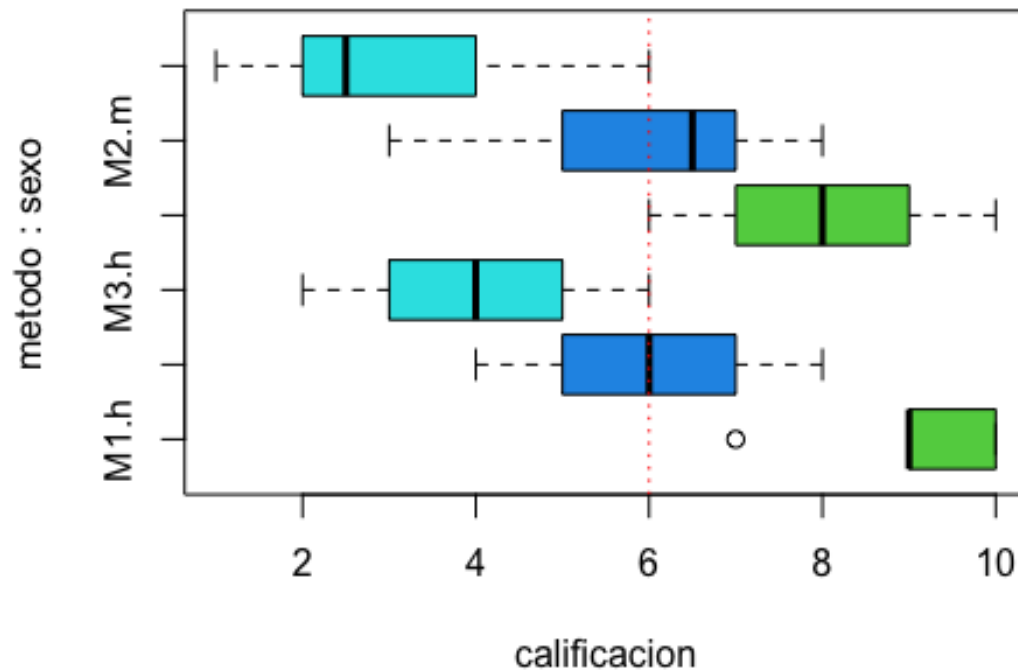
```
calificacion=c(10,7,9,9,9,10,5,7,6,6,8,4,2,6,3,5,5,3,9,7,8,8,10,6,8,3,5,6,7,7,2,6,2,1,4,3)
metodo=c(rep("M1",6),rep("M2",6),rep("M3",6),rep("M1",6),rep("M2",6),rep("M3",6))
sexo = c(rep("h", 18), rep("m",18))
metodo = factor(metodo)
sexo = factor(sexo)
m = tapply(calificacion, metodo : sexo, mean)
m

## M1:h M1:m M2:h M2:m M3:h M3:m
##    9    8    6    6    4    3
```

¿Existe alguna influencia de la metodología de enseñanza y el género de los estudiantes en el rendimiento de los estudiantes?

BoxPlot de la Evaluacion

```
boxplot(calificacion ~ (metodo : sexo), col = 3:5, horizontal = TRUE)
abline(v = mean(calificacion), lty = 3, col = "red")
```



El modelo tiene principalmente 2 factores: el metodo de enseñanza y el sexo

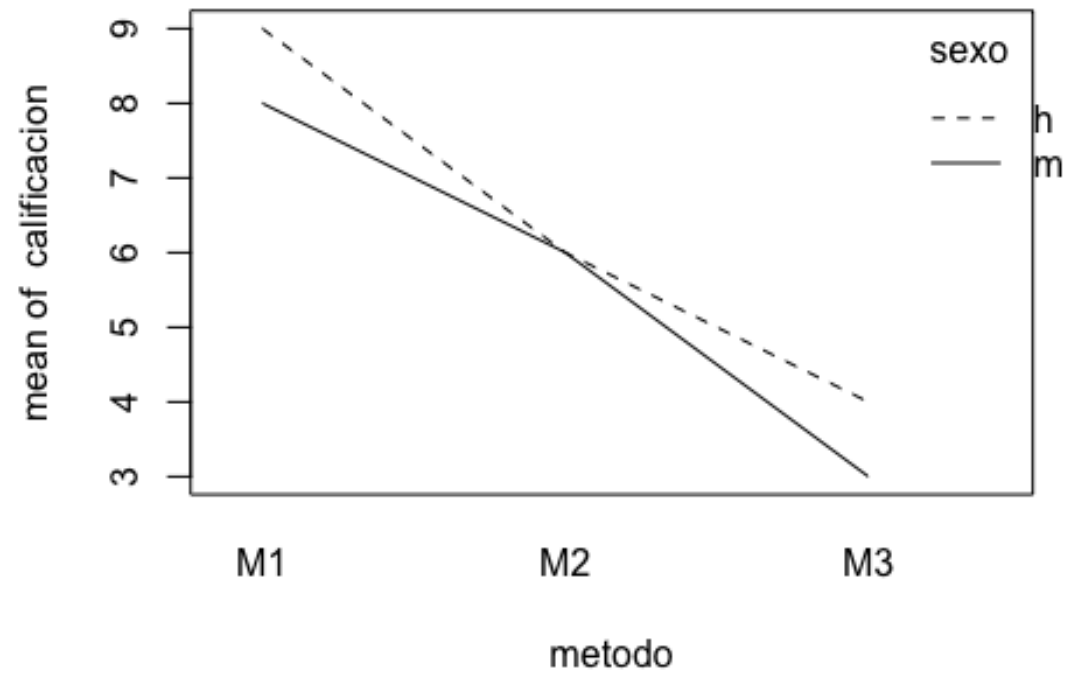
Anova con interaccion

```
A<-aov(calificacion~metodo*sexo)
```

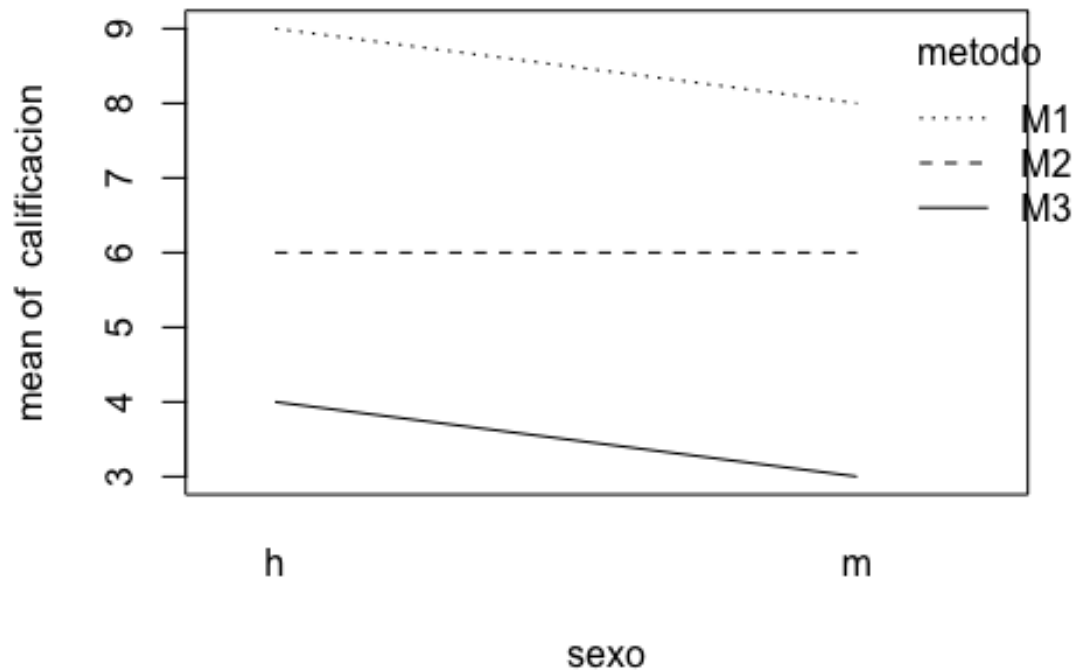
```
summary(A)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## metodo      2    150   75.00  32.143 3.47e-08 ***
## sexo        1     4    4.00   1.714  0.200
## metodo:sexo  2     2    1.00   0.429  0.655
## Residuals   30    70    2.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
interaction.plot(metodo, sexo, calificacion)
```



```
interaction.plot(sexo, metodo, calificacion)
```



Anova sin interaccion

```
B<-aov(calificacion~metodo+sexo)
summary(B)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## metodo         2    150   75.00   33.333 1.5e-08 ***
## sexo           1     4    4.00    1.778  0.192
## Residuals     32     72    2.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tapply(calificacion,sexo,mean)
```

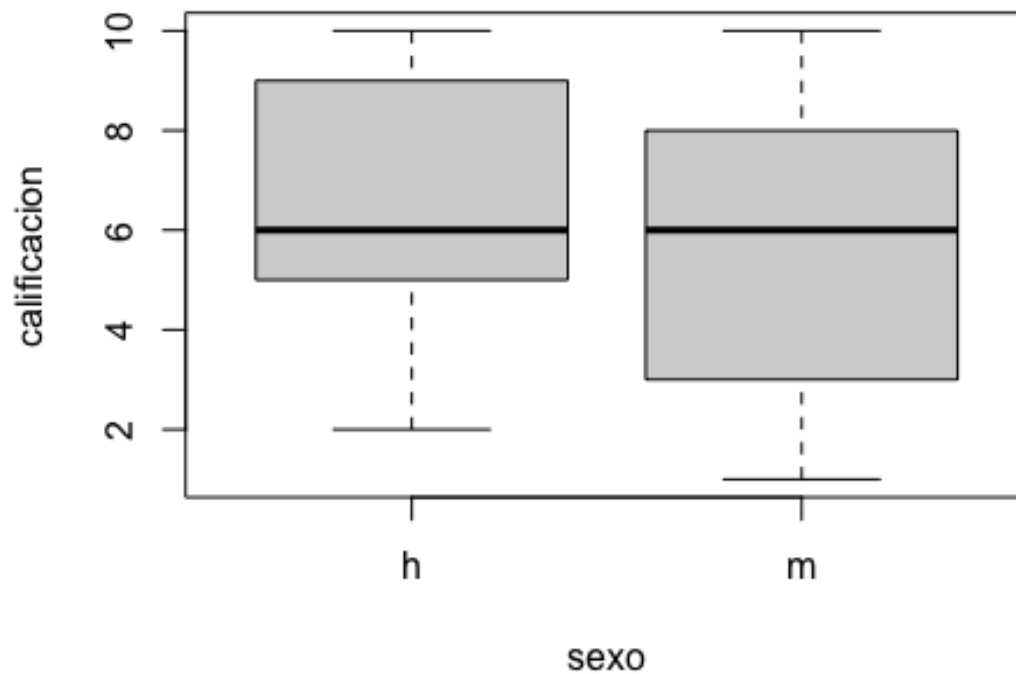
```
##          h          m
## 6.333333 5.666667
```

```
tapply(calificacion,metodo,mean)
```

```
##  M1  M2  M3
## 8.5 6.0 3.5
```

```
M=mean(calificacion)
M
```

```
## [1] 6
boxplot(calificacion ~ sexo)
```



Anova para un efecto principal

```
C<-aov(calificacion~metodo)
summary(C)

##           Df Sum Sq Mean Sq F value    Pr(>F)    
## metodo      2    150    75.0   32.57 1.55e-08 ***
## Residuals  33     76     2.3               
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

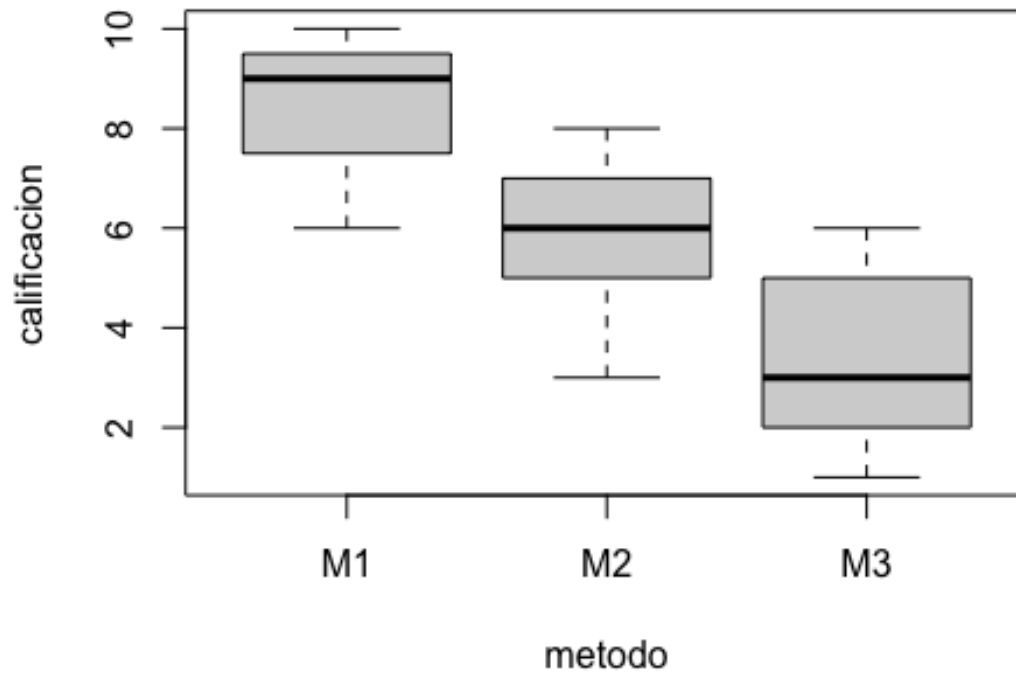
tapply(calificacion,metodo,mean)

##  M1  M2  M3
## 8.5 6.0 3.5

M2=mean(calificacion)
M2

## [1] 6
```

```
boxplot(calificacion ~ metodo)
```

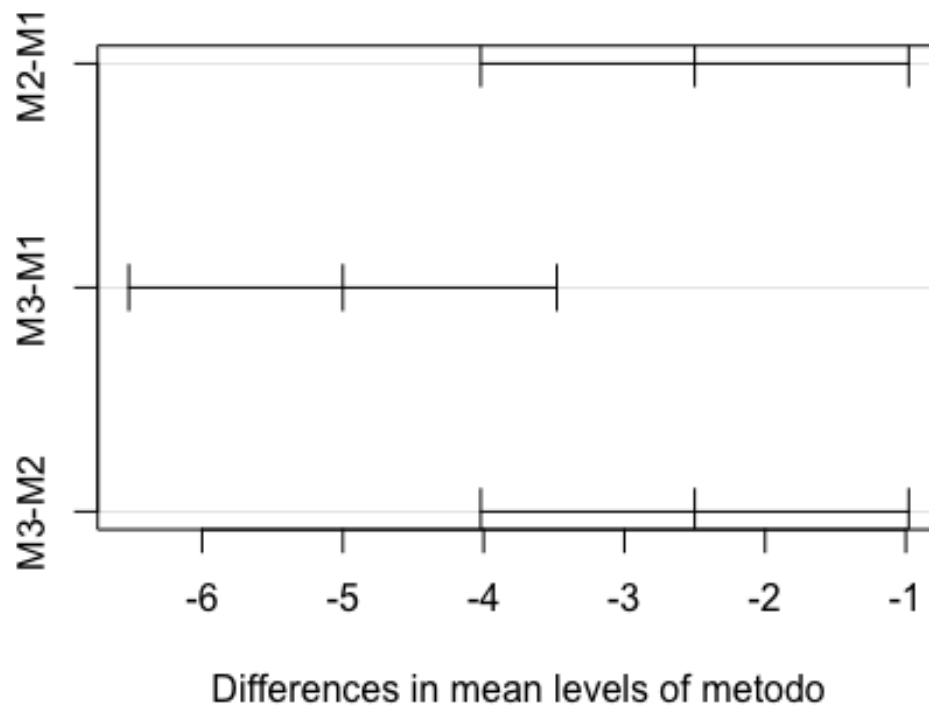


```
I = TukeyHSD(aov(calificacion ~ metodo))
I

##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = calificacion ~ metodo)
##
## $metodo
##      diff      lwr      upr    p adj
## M2-M1 -2.5 -4.020241 -0.9797592 0.0008674
## M3-M1 -5.0 -6.520241 -3.4797592 0.0000000
## M3-M2 -2.5 -4.020241 -0.9797592 0.0008674
```

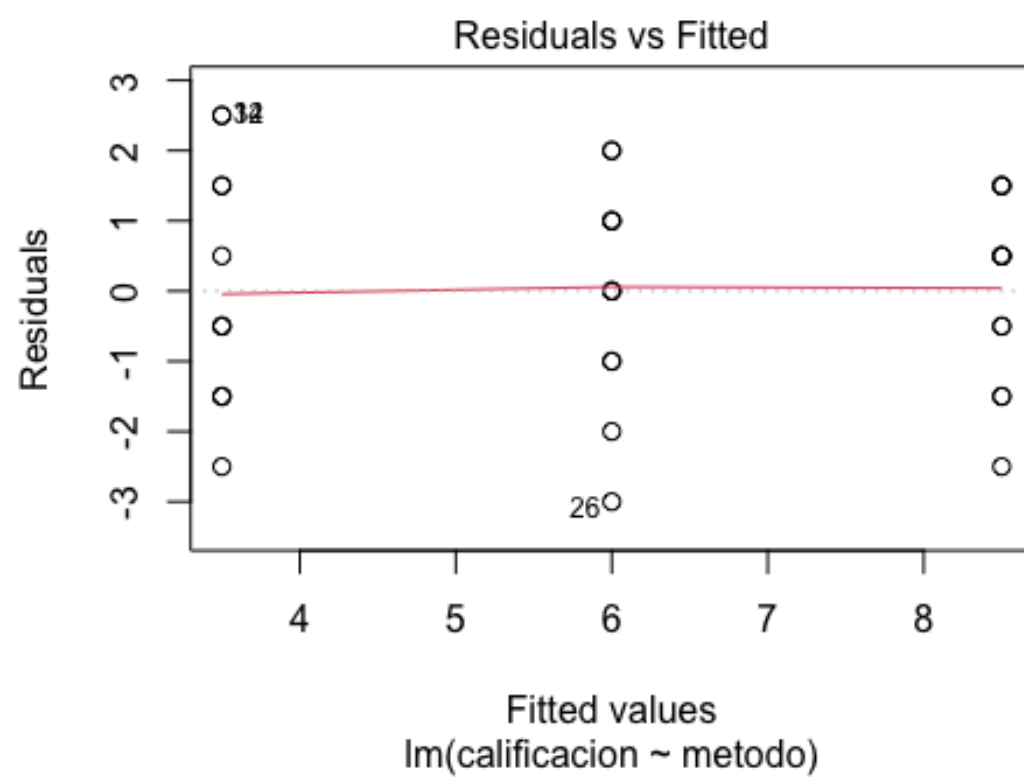
```
plot(I) #Los intervalos de confianza se observan mejor si se grafican
```

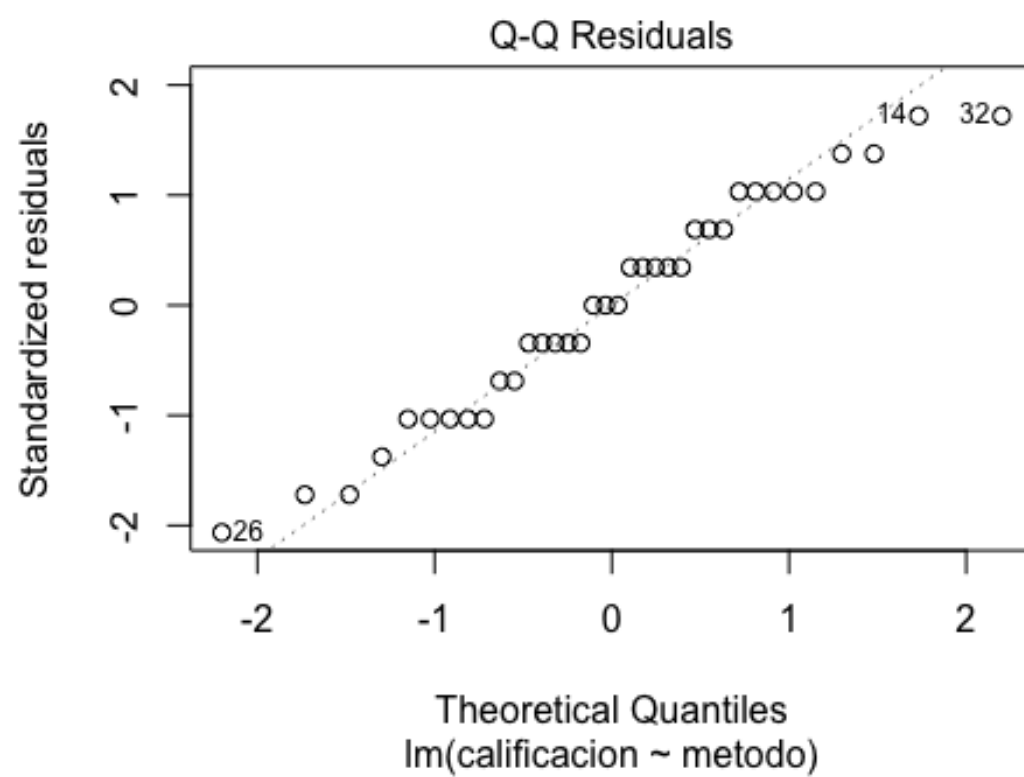
95% family-wise confidence level

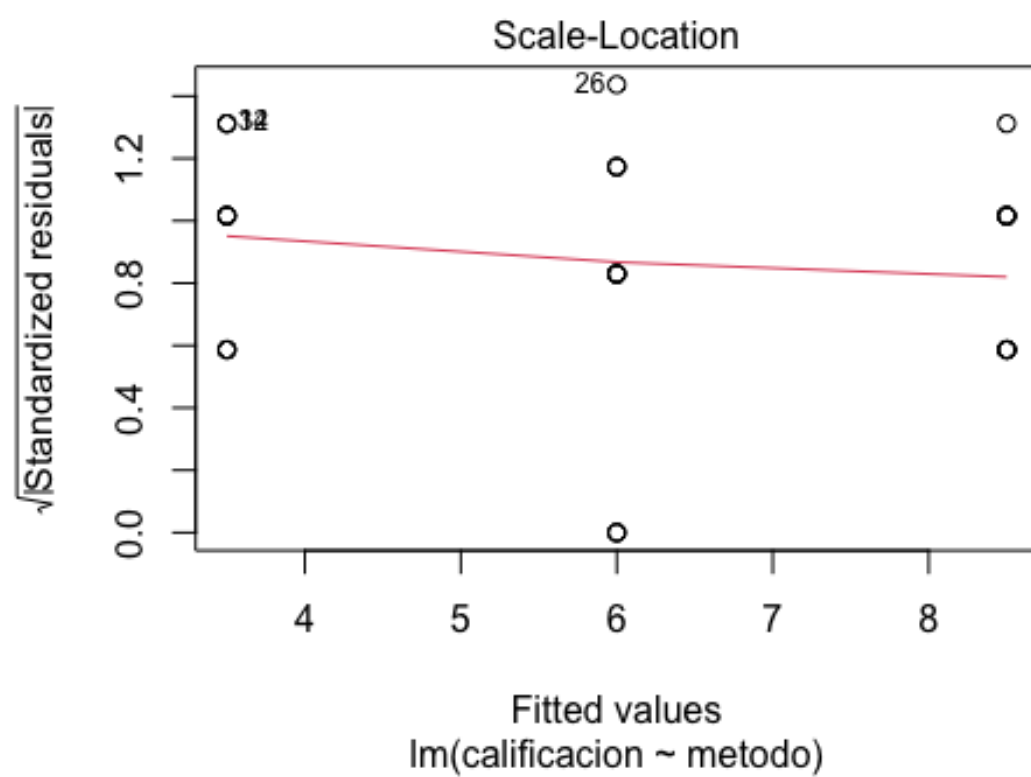


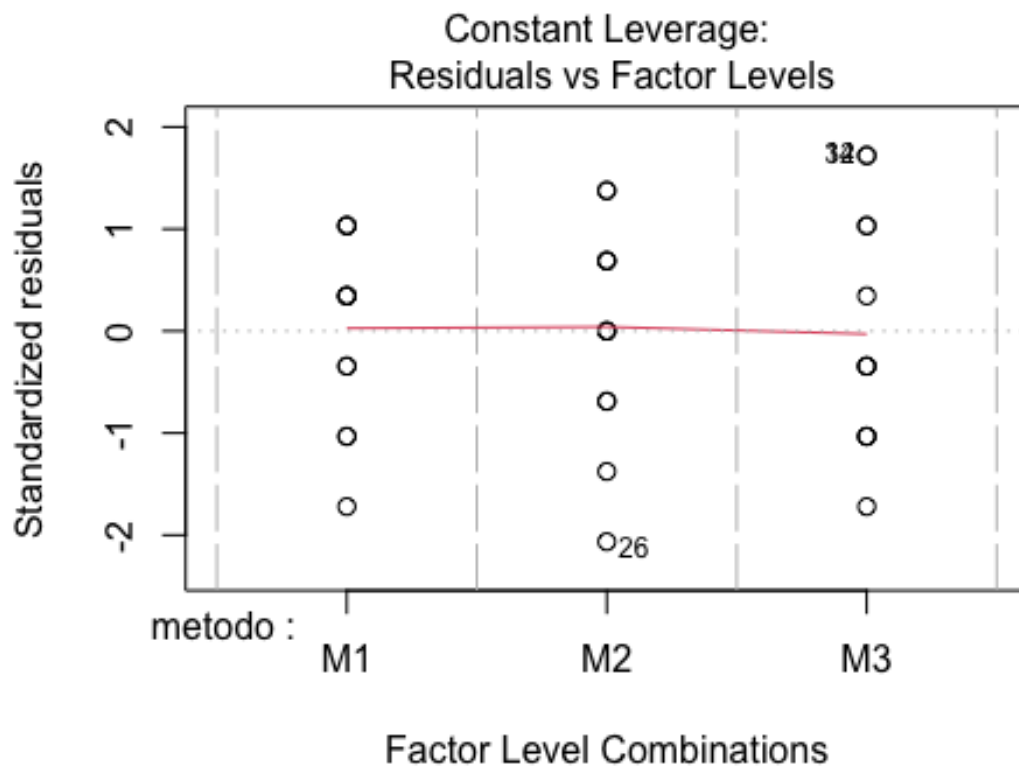
Analysis

```
plot(lm(calificacion~metodo))
```









```
CD= 150/(150+76) #coeficiente de determinación para el modelo.
```

Vibracion de motores

Un ingeniero de procesos ha identificado dos causas potenciales de vibración de los motores eléctricos, el material utilizado para la carcasa del motor (factor A) y el proveedor de cojinetes utilizados en el motor (Factor B). Los siguientes datos sobre la cantidad de vibración (micrones) se obtuvieron mediante un experimento en el cual se construyeron motores con carcasas de acero, aluminio y plástico y cojinetes suministrados por cinco proveedores seleccionados al azar.

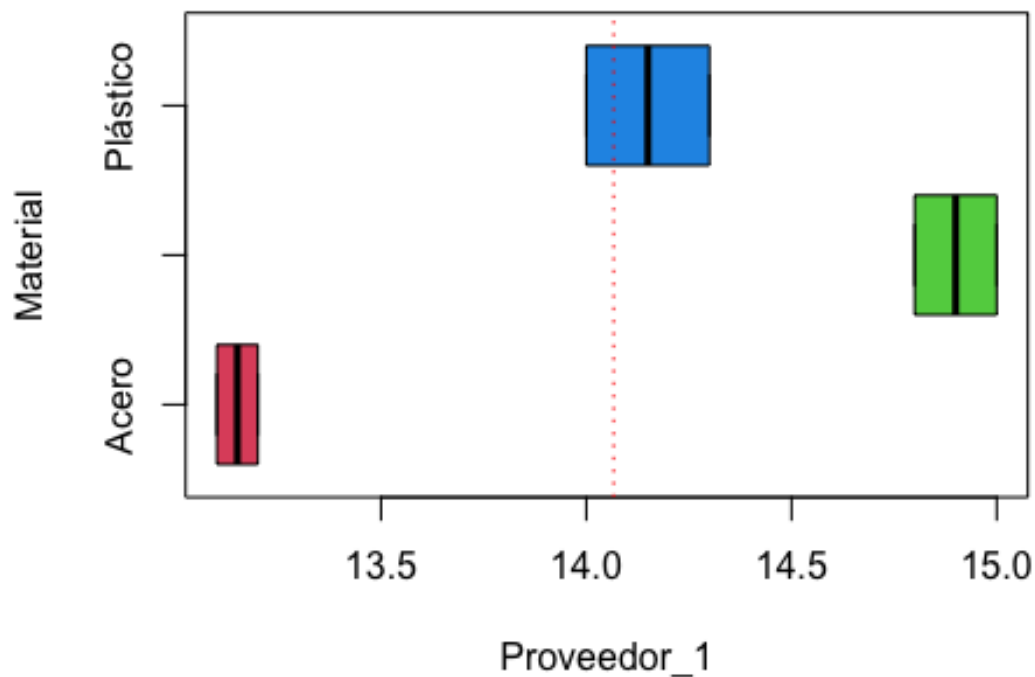
Introduciendo Datos

```
Material = c(rep("Acero", 2), rep("Aluminio", 2), rep("Plástico", 2))
Proveedor_1 = c(13.1, 13.2, 15.0, 14.8, 14.0, 14.3)
Proveedor_2 = c(16.3, 15.8, 15.7, 16.4, 17.2, 16.7)
Proveedor_3 = c(13.7, 14.3, 13.9, 14.3, 12.4, 12.3)
Proveedor_4 = c(15.7, 15.8, 13.7, 14.2, 14.4, 13.9)
Proveedor_5 = c(13.5, 12.5, 13.4, 13.8, 13.2, 13.1)
Resultados = c(13.1, 13.2, 15.0, 14.8, 14.0, 14.3, 16.3, 15.8, 15.7, 16.4,
17.2, 16.7, 13.7, 14.3, 13.9, 14.3, 12.4, 12.3, 15.7, 15.8, 13.7, 14.2, 14.4,
13.9, 13.5, 12.5, 13.4, 13.8, 13.2, 13.1)
```

¿Existe evidencia de que el proveedor y el material son causas de la vibración de los motores eléctricos?

BoxPlot de la evaluación del Proveedor 1

```
boxplot(Proveedor_1 ~ Material, col = 2:5, horizontal = TRUE)  
abline(v = mean(Proveedor_1), lty = 3, col = "red")
```



Anova con interaccion

```
A2<-aov(Proveedor_1~Material)  
summary(A2)
```

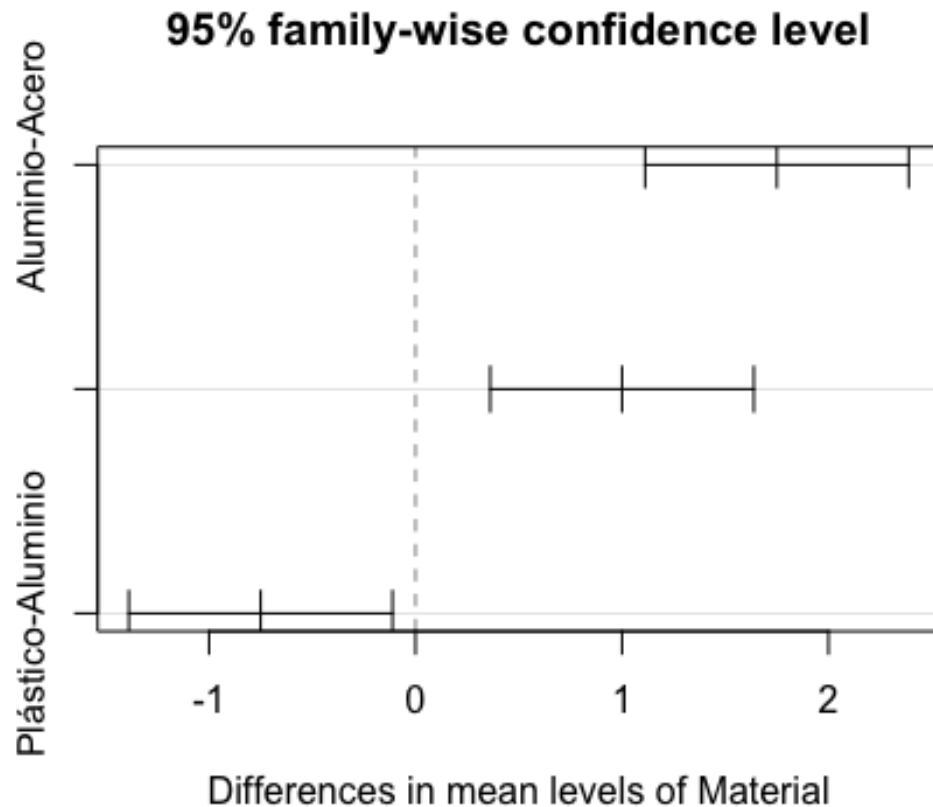
```
##              Df Sum Sq Mean Sq F value    Pr(>F)      
## Material      2   3.083   1.5417    66.07 0.00331 **    
## Residuals    3   0.070   0.0233                  
## ---   
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
I2 = TukeyHSD(aov(Proveedor_1 ~ Material))  
I2
```

```
##      Tukey multiple comparisons of means  
##      95% family-wise confidence level  
##
```

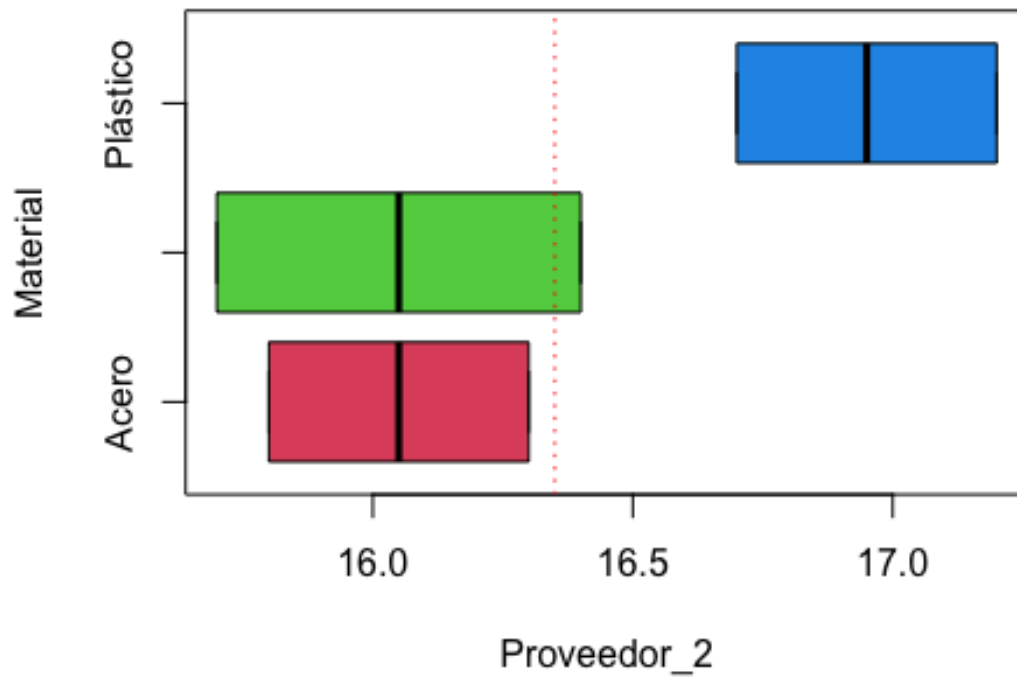
```
## Fit: aov(formula = Proveedor_1 ~ Material)
##
## $Material
##           diff          lwr          upr      p adj
## Aluminio-Acero    1.75    1.1116834    2.3883166 0.0029465
## Plástico-Acero     1.00    0.3616834    1.6383166 0.0147633
## Plástico-Aluminio -0.75   -1.3883166   -0.1116834 0.0326523
```

`plot(I2)`



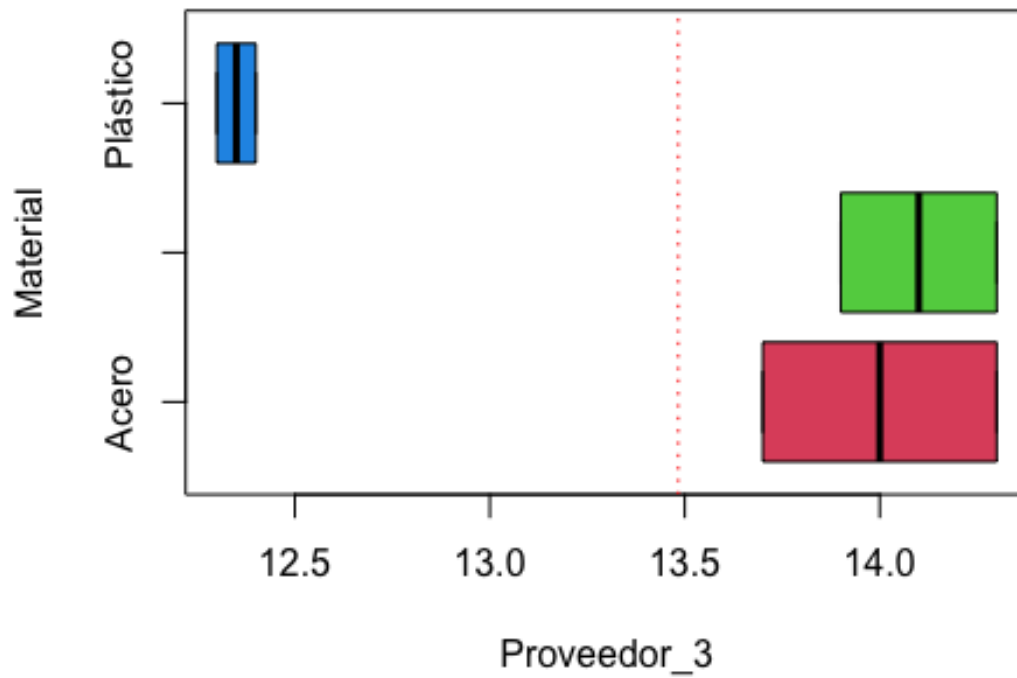
BoxPlot de la evaluacion del Proveedor 2

```
boxplot(Proveedor_2 ~ Material, col = 2:5, horizontal = TRUE)
abline(v = mean(Proveedor_2), lty = 3, col = "red")
```



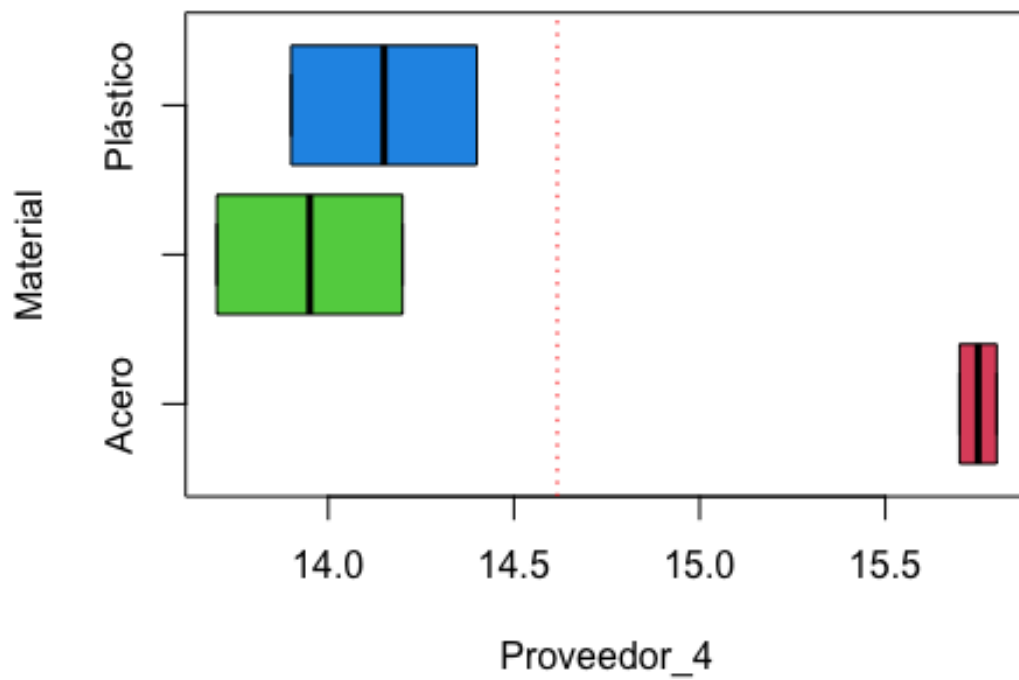
BoxPlot de la evaluación del Proveedor 3

```
boxplot(Proveedor_3 ~ Material, col = 2:5, horizontal = TRUE)
abline(v = mean(Proveedor_3), lty = 3, col = "red")
```



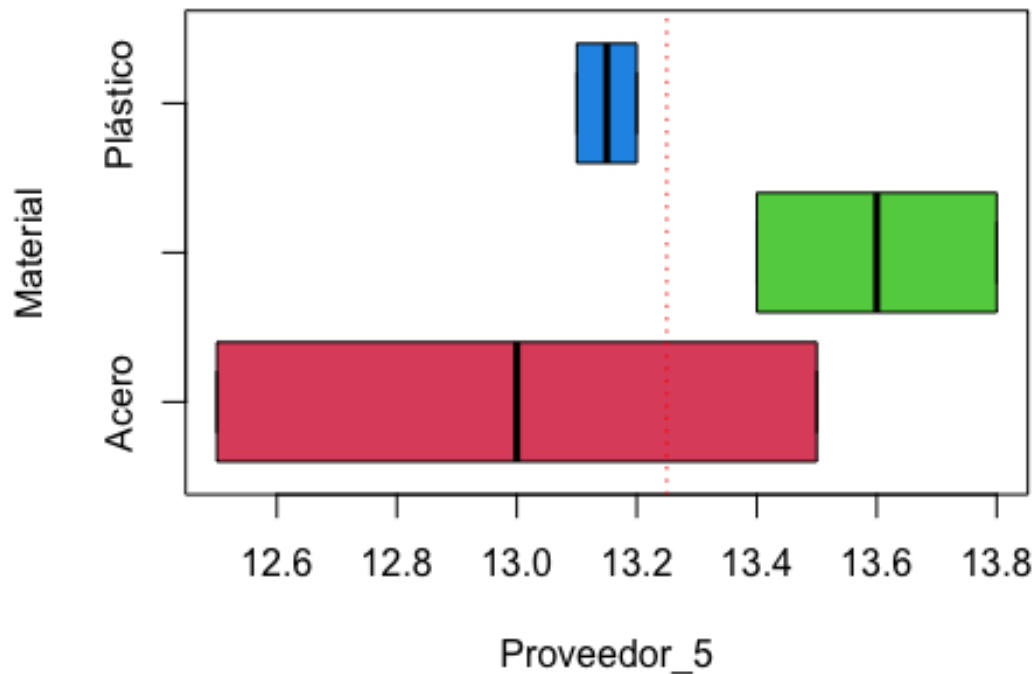
BoxPlot de la evaluacion del Proveedor 4

```
boxplot(Proveedor_4 ~ Material, col = 2:5, horizontal = TRUE)
abline(v = mean(Proveedor_4), lty = 3, col = "red")
```



BoxPlot de la evaluación del Proveedor 5

```
boxplot(Proveedor_5 ~ Material, col = 2:5, horizontal = TRUE)
abline(v = mean(Proveedor_5), lty = 3, col = "red")
```

Tras ver los diferentes boxplots de cada variable podemos observar que en unos proveedores específicos existe mayor varianza en los resultados de sus pruebas estos siendo los proveedores 2 y 5 que analizaremos mas a detalle.

Proveedor 2

```
C2<-aov(Proveedor_2~Material)
summary(C2)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Material    2  1.080   0.540   3.273  0.176
## Residuals   3   0.495   0.165
```

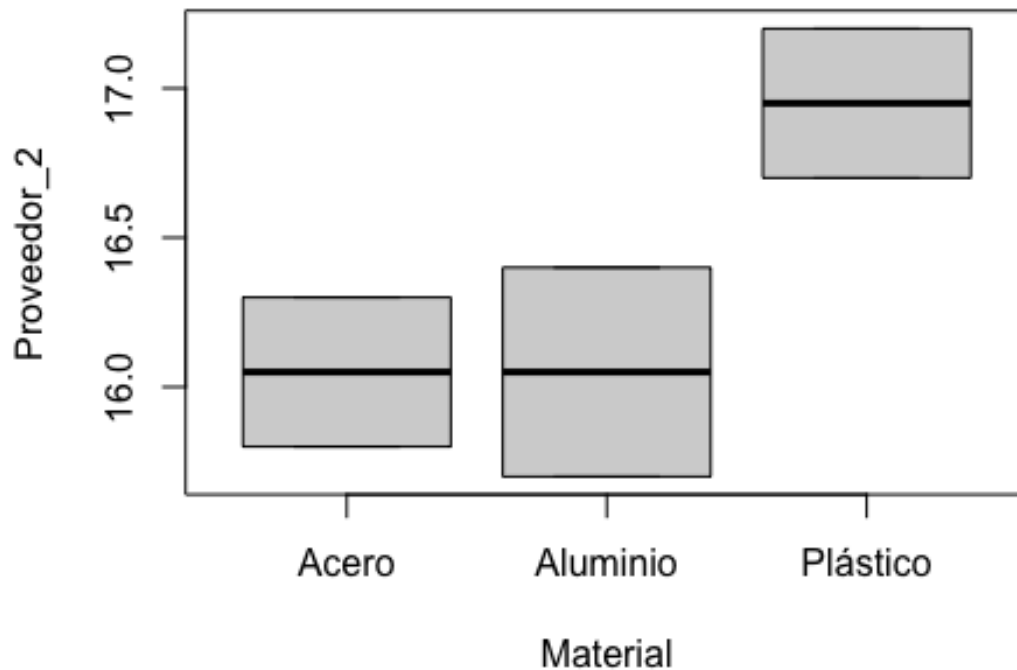
```
tapply(Proveedor_2,Material,mean)
```

```
##   Acero Aluminio Plástico
##  16.05  16.05   16.95
```

```
M2=mean(Proveedor_2)
M2
```

```
## [1] 16.35
```

```
boxplot(Proveedor_2 ~ Material)
```



```
I2 = TukeyHSD(aov(Proveedor_2 ~ Material))
```

```
I2
```

```
## Tukey multiple comparisons of means
```

```
## 95% family-wise confidence level
```

```
##
```

```
## Fit: aov(formula = Proveedor_2 ~ Material)
```

```
##
```

```
## $Material
```

```
## diff lwr upr p adj
```

```
## Aluminio-Acero -3.552714e-15 -1.6974215 1.697422 1.0000000
```

```
## Plástico-Acero 9.000000e-01 -0.7974215 2.597422 0.2141524
```

```
## Plástico-Aluminio 9.000000e-01 -0.7974215 2.597422 0.2141524
```

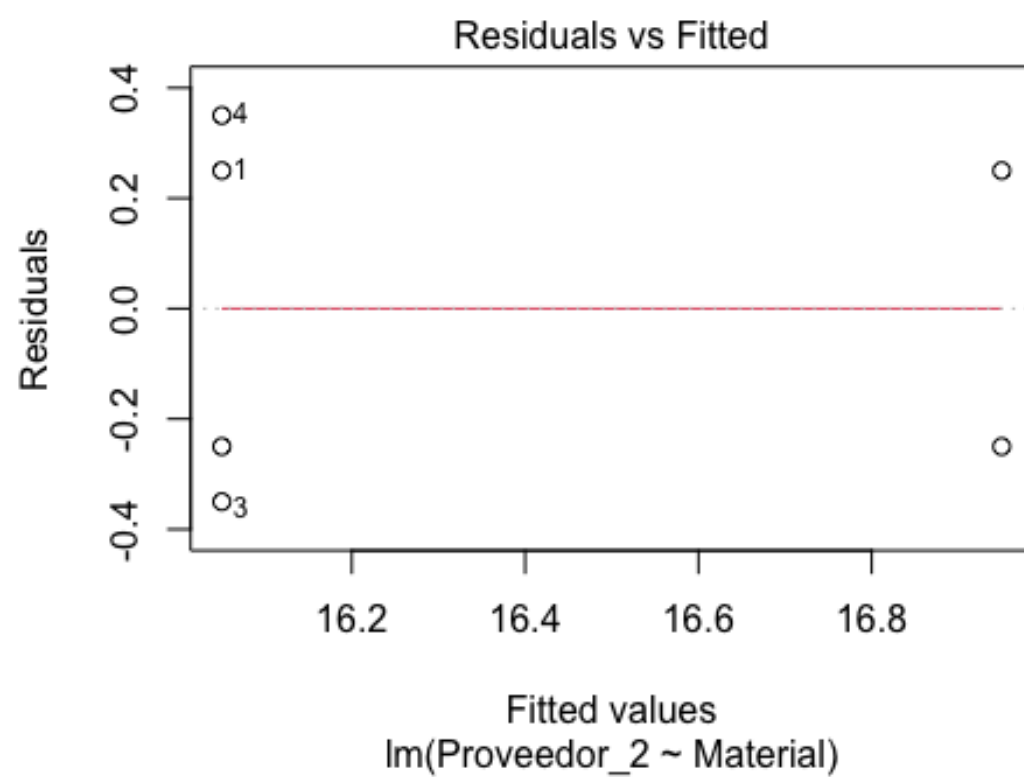
```
plot(I2) #Los intervalos de confianza se observan mejor si se grafican
```

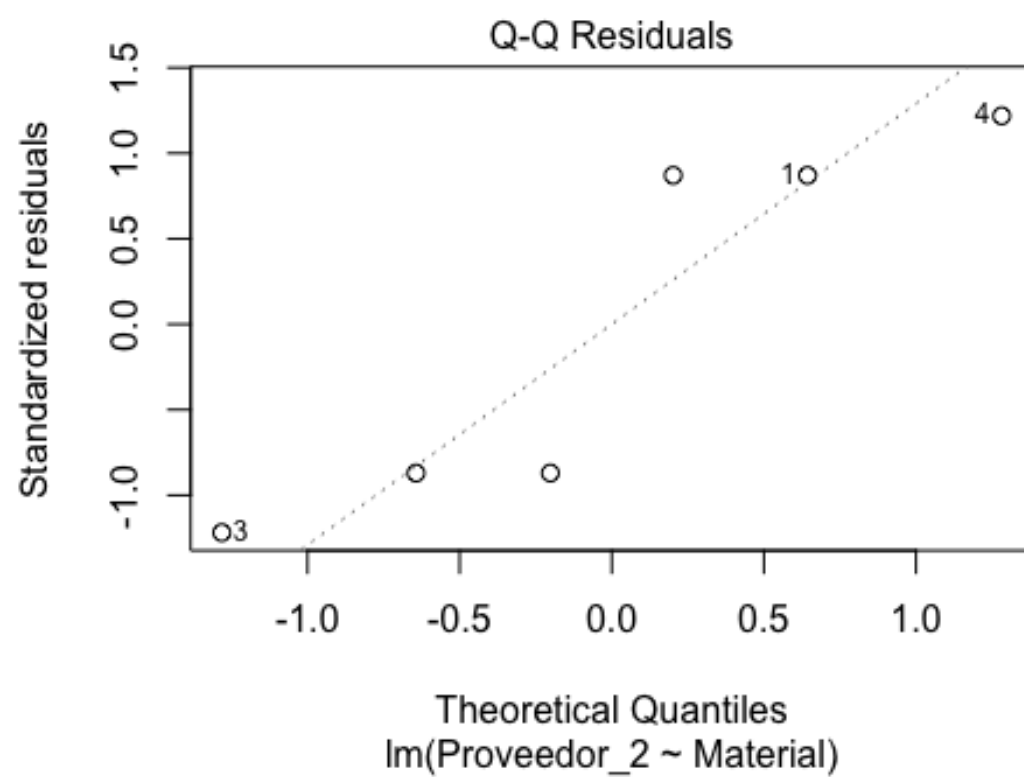


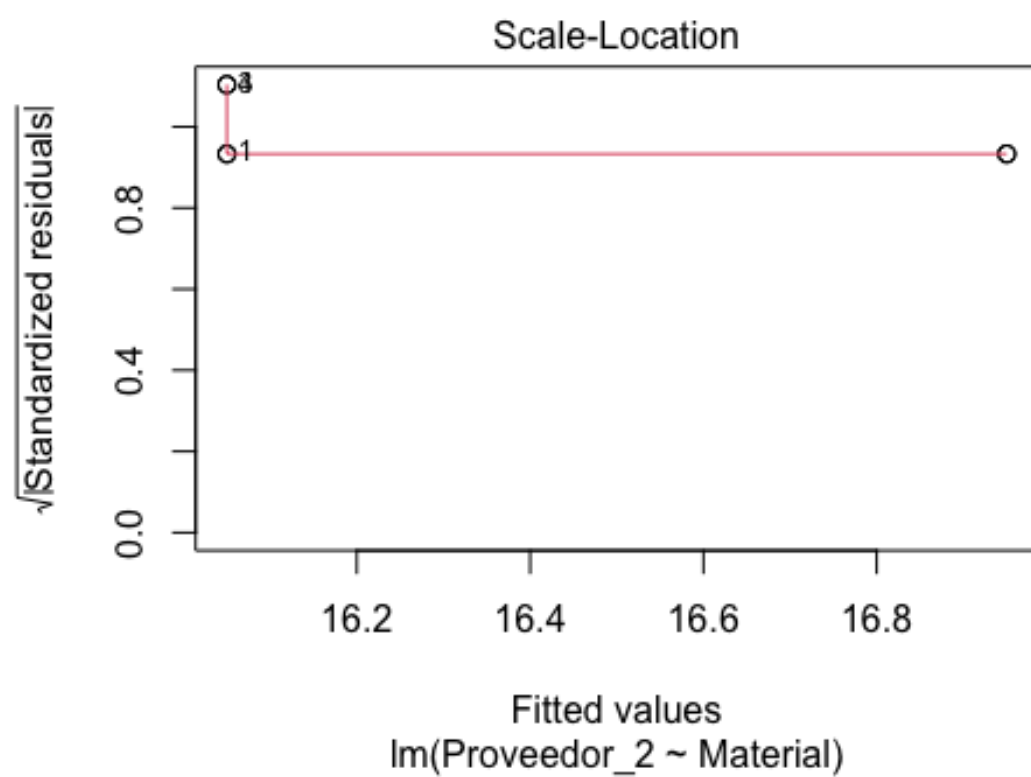
##

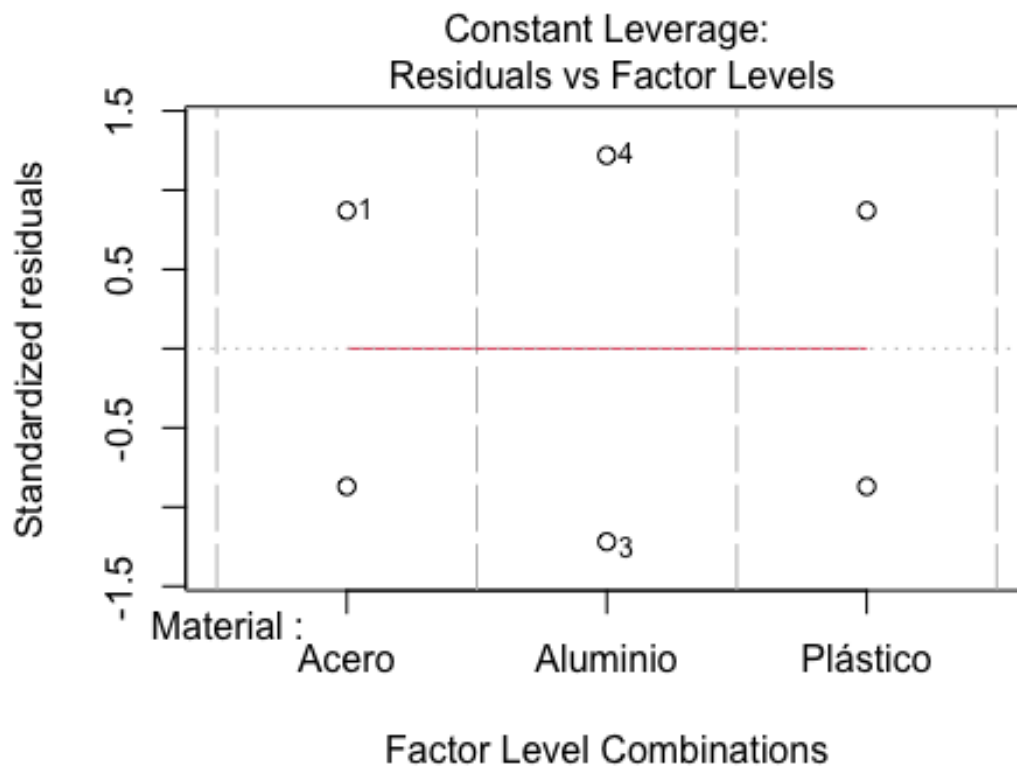
Analysis

```
plot(lm(Proveedor_2~Material))
```









Proveedor 5

```
C5<-aov(Proveedor_5~Material)
```

```
summary(C5)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Material    2  0.390   0.195      1  0.465
## Residuals   3  0.585   0.195
```

```
lapply(Proveedor_5,Material,mean)
```

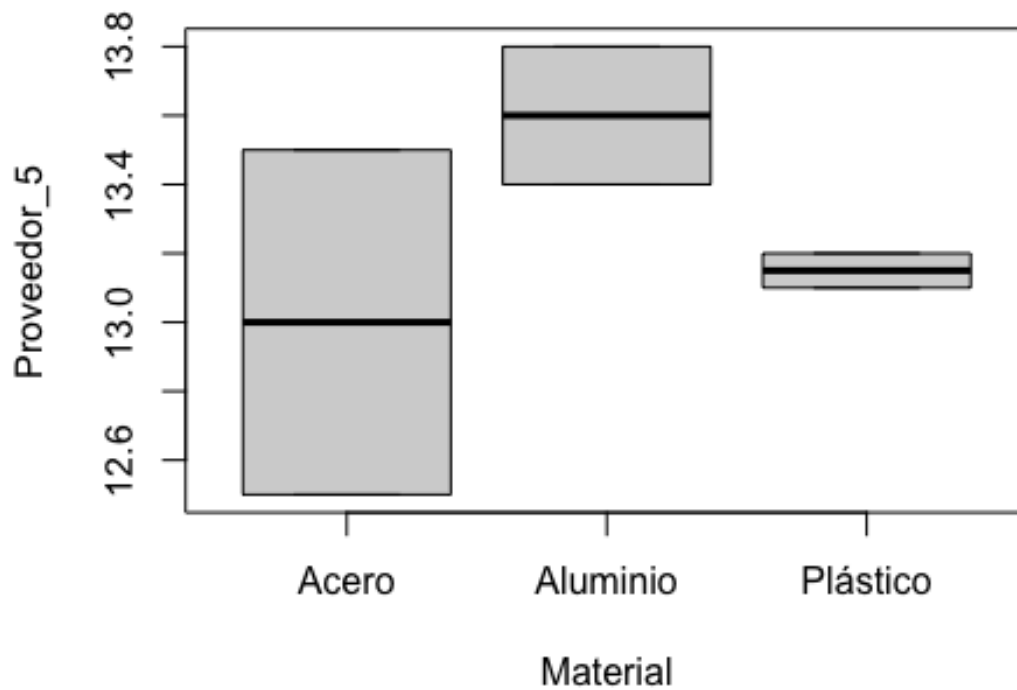
```
##   Acero Aluminio Plástico
##  13.00   13.60   13.15
```

```
M5=mean(Proveedor_5)
```

```
M5
```

```
## [1] 13.25
```

```
boxplot(Proveedor_5 ~ Material)
```



```
I5 = TukeyHSD(aov(Proveedor_5 ~ Material))
```

```
I5
```

```
## Tukey multiple comparisons of means
```

```
## 95% family-wise confidence level
```

```
##
```

```
## Fit: aov(formula = Proveedor_5 ~ Material)
```

```
##
```

```
## $Material
```

```
## diff lwr upr p adj
```

```
## Aluminio-Acero 0.60 -1.245292 2.445292 0.4608778
```

```
## Plástico-Acero 0.15 -1.695292 1.995292 0.9396205
```

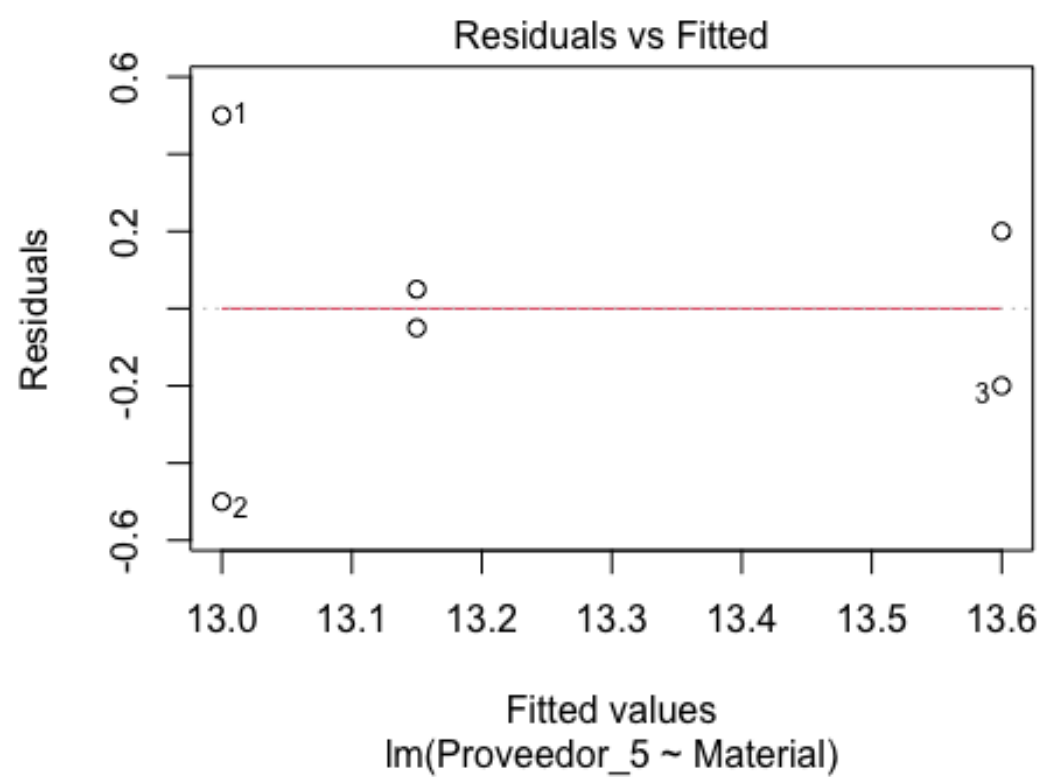
```
## Plástico-Aluminio -0.45 -2.295292 1.395292 0.6163837
```

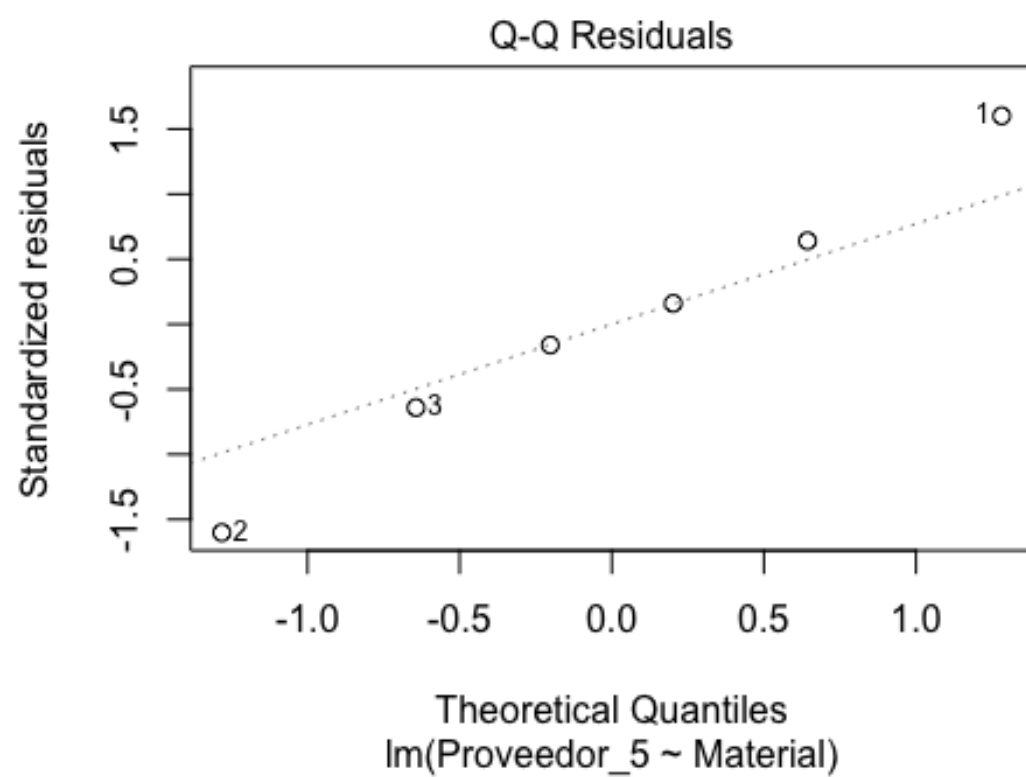
```
plot(I5) #Los intervalos de confianza se observan mejor si se grafican
```

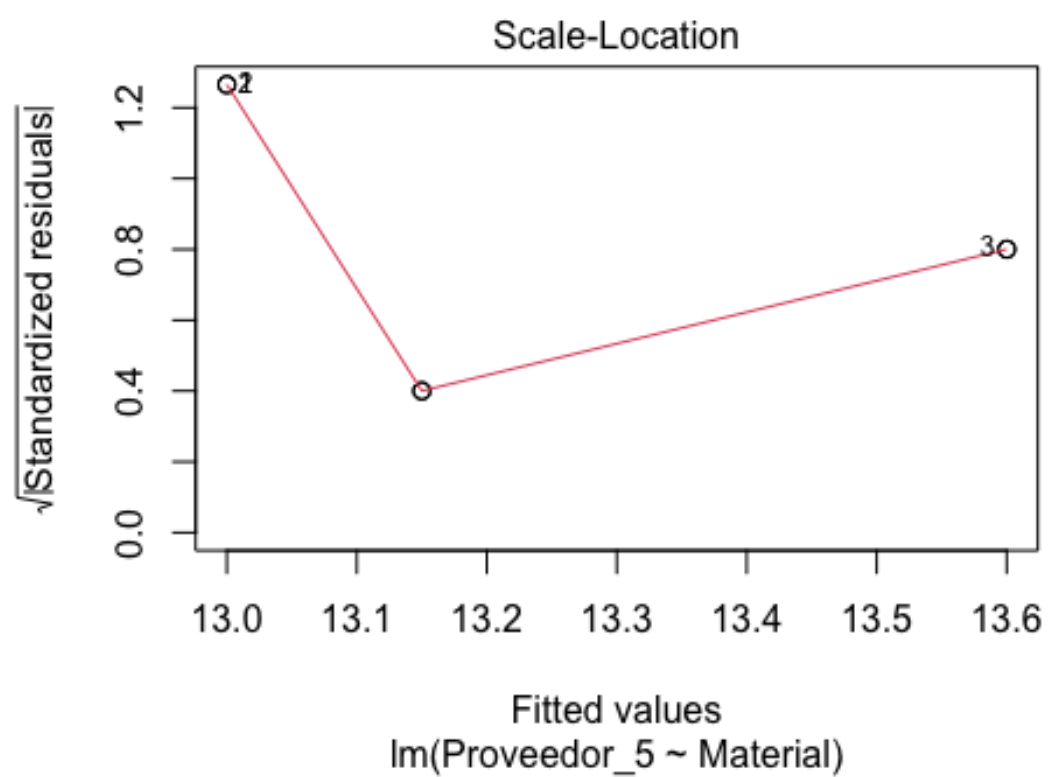


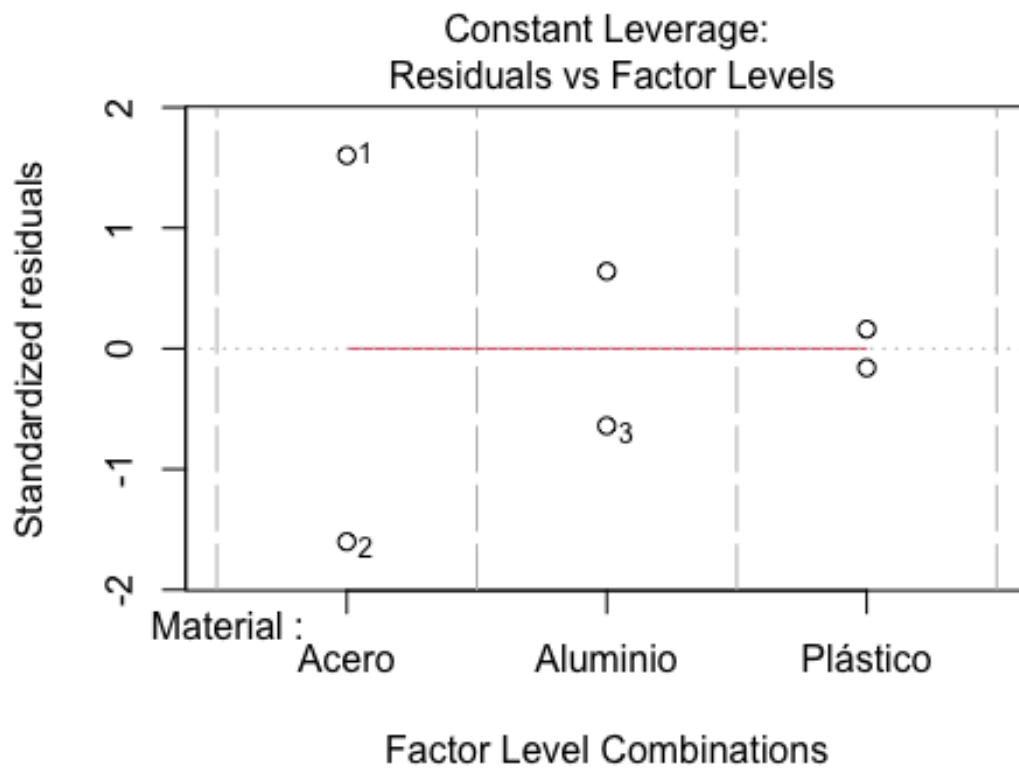

Analysis

```
plot(lm(Proveedor_5~Material))
```









Como podemos observar dependiendo del proveedor la varianza entre materiales es mayor en unos que en otros, se puede observar de mejor manera en las tablas de los analisis de cada proveedor-material.