EjemploAnova

Oskar Arturo Gamboa Reyes

2024-08-27

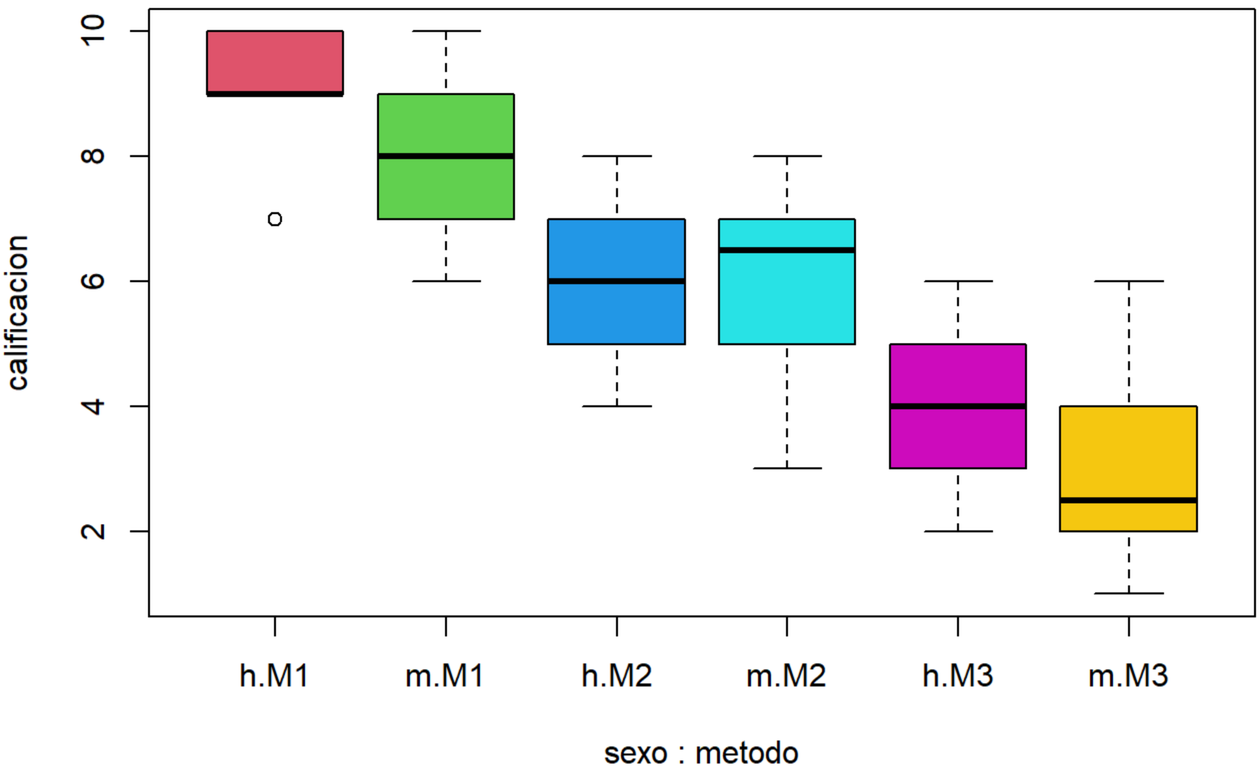
Problema 1

Cargar 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)

datos = data.frame(calificacion, metodo, sexo)
```

BoxPlot interacción

```
boxplot(calificacion~sexo:metodo, datos ,col = 2:8 )
```



Por lo que podemos ver en esta gráfica es que no hay tanta varianza entre los sexos, la mayor varianza se encuentra entre los métodos. Por lo que es seguro que el nivel de significancia no sea tan grande.

Hipótesis

Primera Hipótesis $H_0 : \tau_i = 0$ $H_1 :$ algún τ_i es distinto a 0

Segunda Hipótesis $H_0 : \alpha_j = 0$ $H_1 :$ algún α_j es distinto a 0

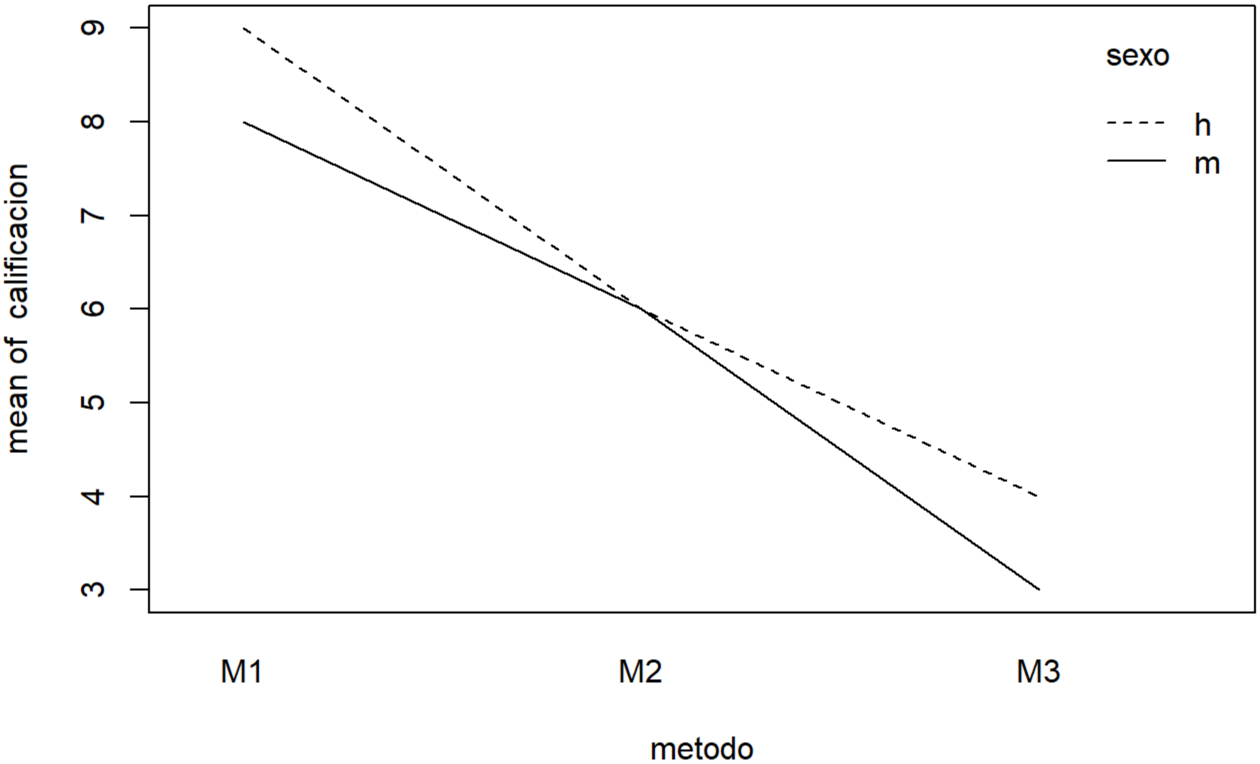
Tercera Hipótesis $H_0 : \tau_i\alpha_j = 0$ $H_1 :$ algún $\tau_i\alpha_j$ es distinto a 0

Anova con interacción

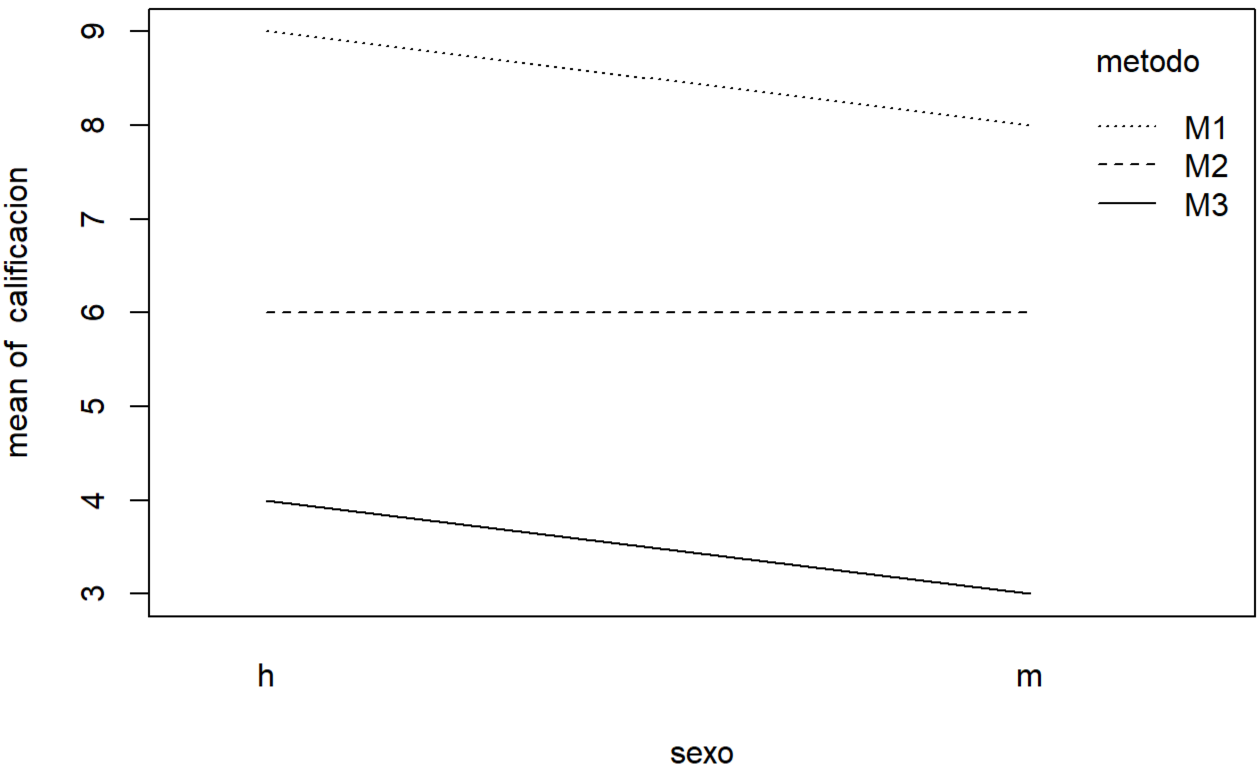
```
A<- aov(calificacion~metodo*sexo, datos)
summary(A)
```

```
##             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)
```



Podemos ver que la interacción entre sexo y metodo no afectan crear una varianza en la calificacion, ni tampoco el sexo, lo que tiene una mayor varianza es el metodo de enseñanza. ## Anova sin interacción

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

Boxplot sexo y medias

```
cat("Medias por sexo y método","\n")
```

```
## Medias por sexo y método
```

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

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

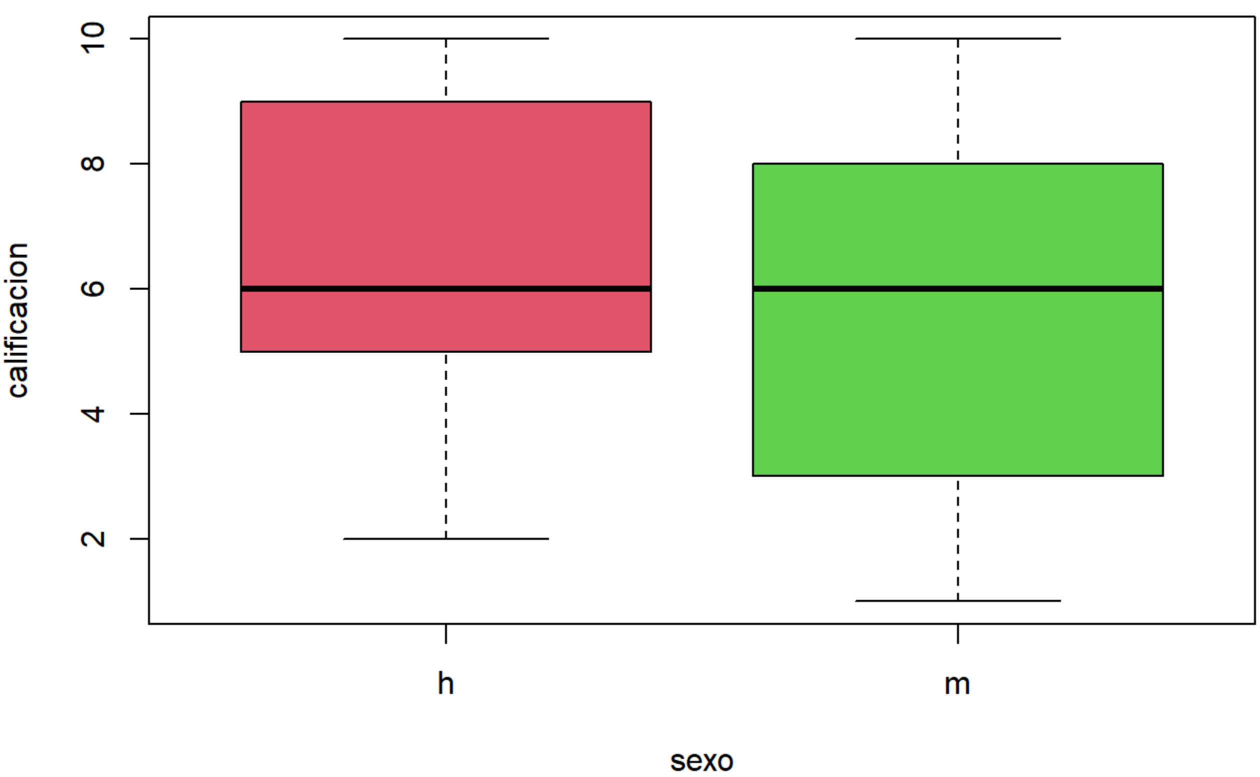
```
tapply(calificacion,metodo,mean)
```

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

```
M=mean(calificacion)
cat("Media General=", M)
```

```
## Media General= 6
```

```
boxplot(calificacion ~ sexo, col =2:3)
```

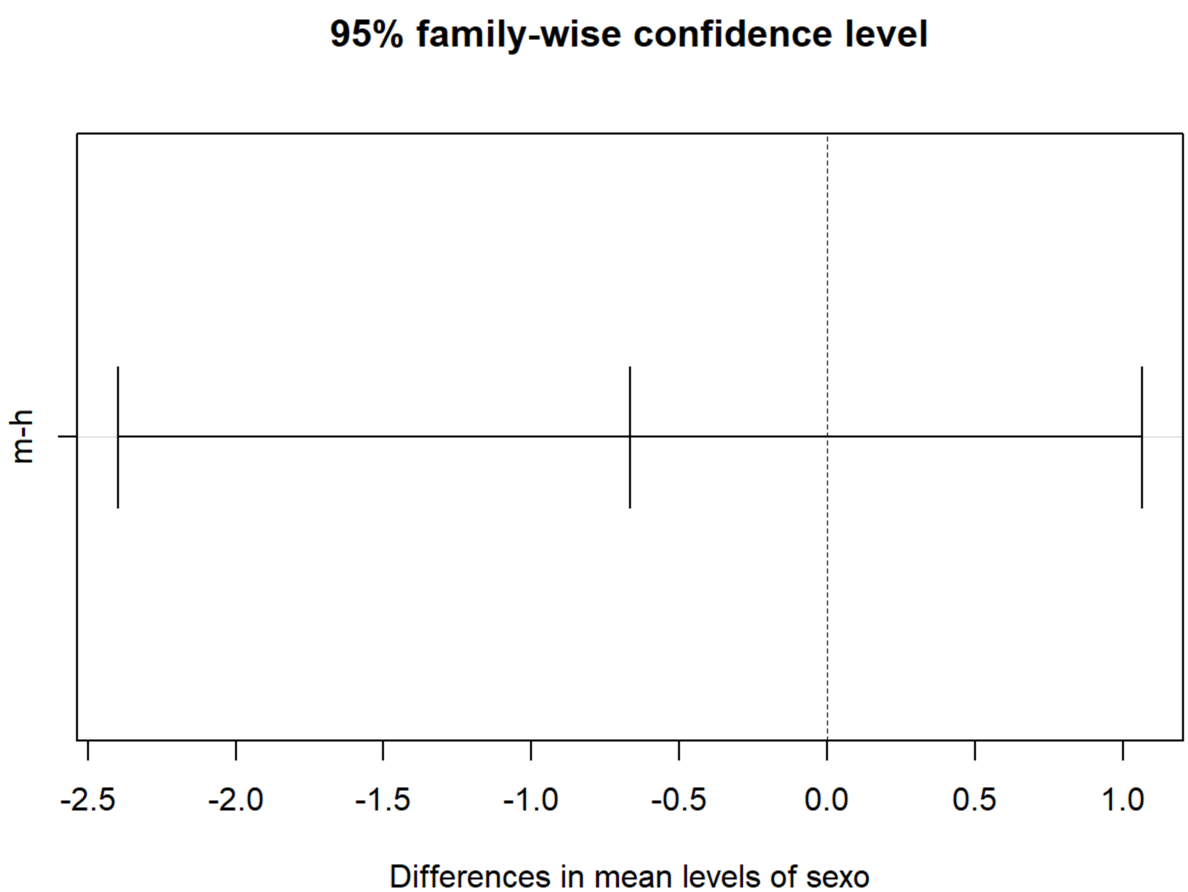


Niveles de confianza de diferencia de medias

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

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = calificacion ~ sexo)
##
## $sexo
##           diff       lwr      upr     p adj
## m-h -0.6666667 -2.397645 1.064312 0.4392235
```

```
plot(I)
```




```
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
```

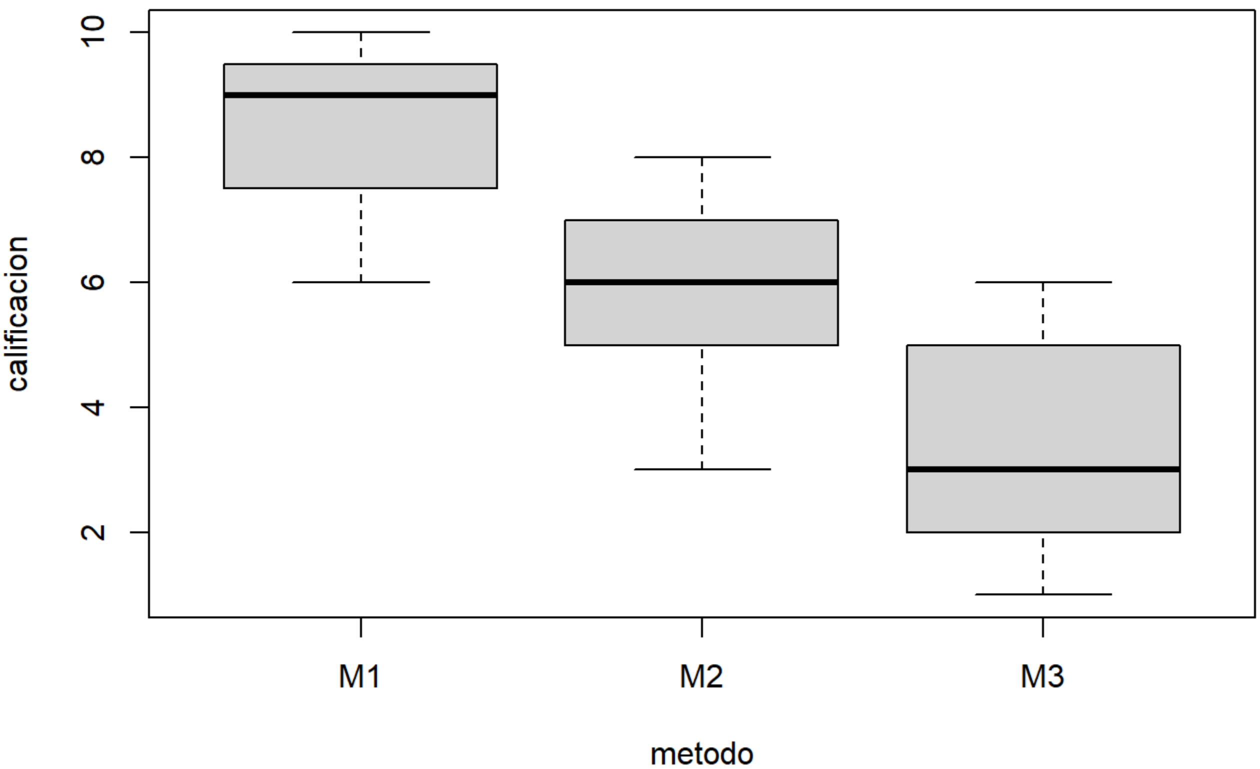
```
tapply(calificacion,metodo,mean)
```

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

```
mean(calificacion)
```

```
## [1] 6
```

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



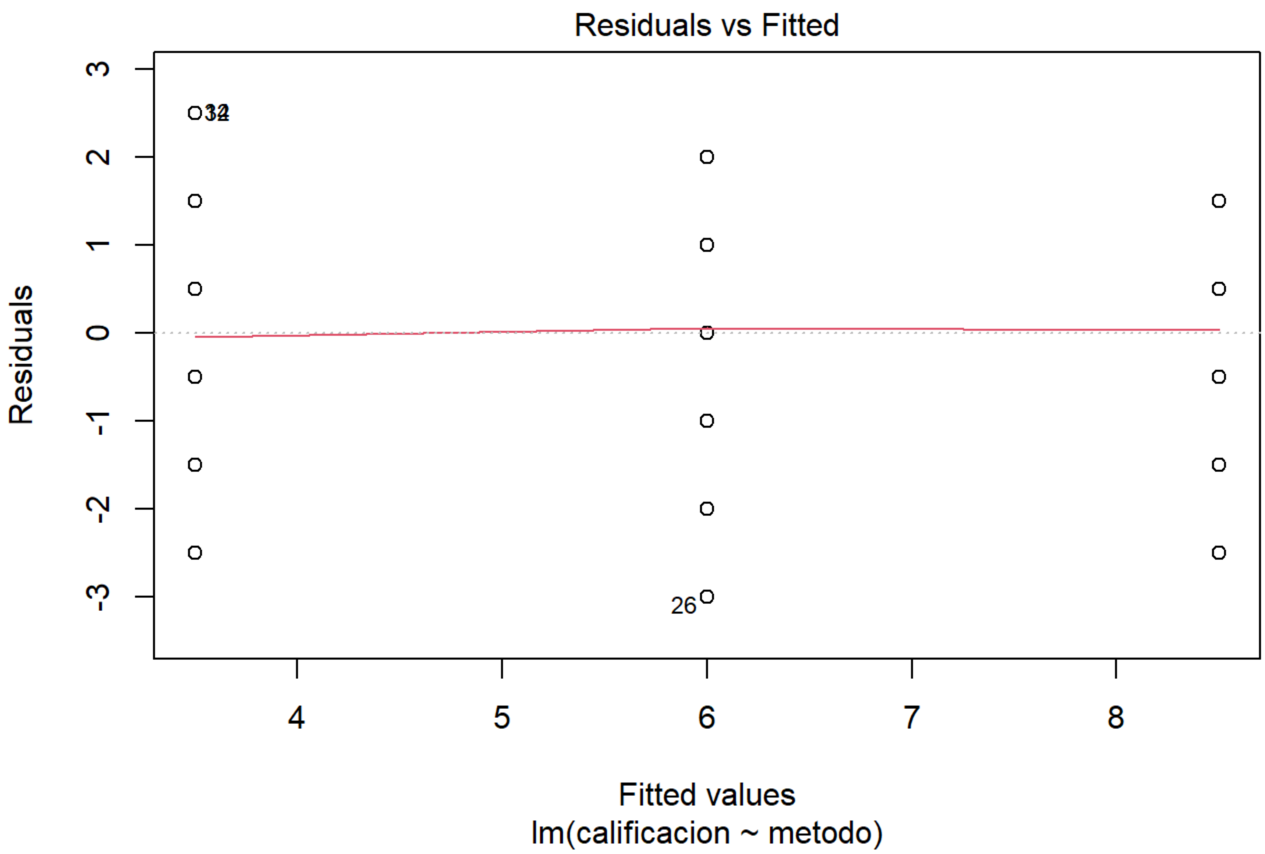

```
I = TukeyHSD(aov(calificacion ~ metodo))
I
```

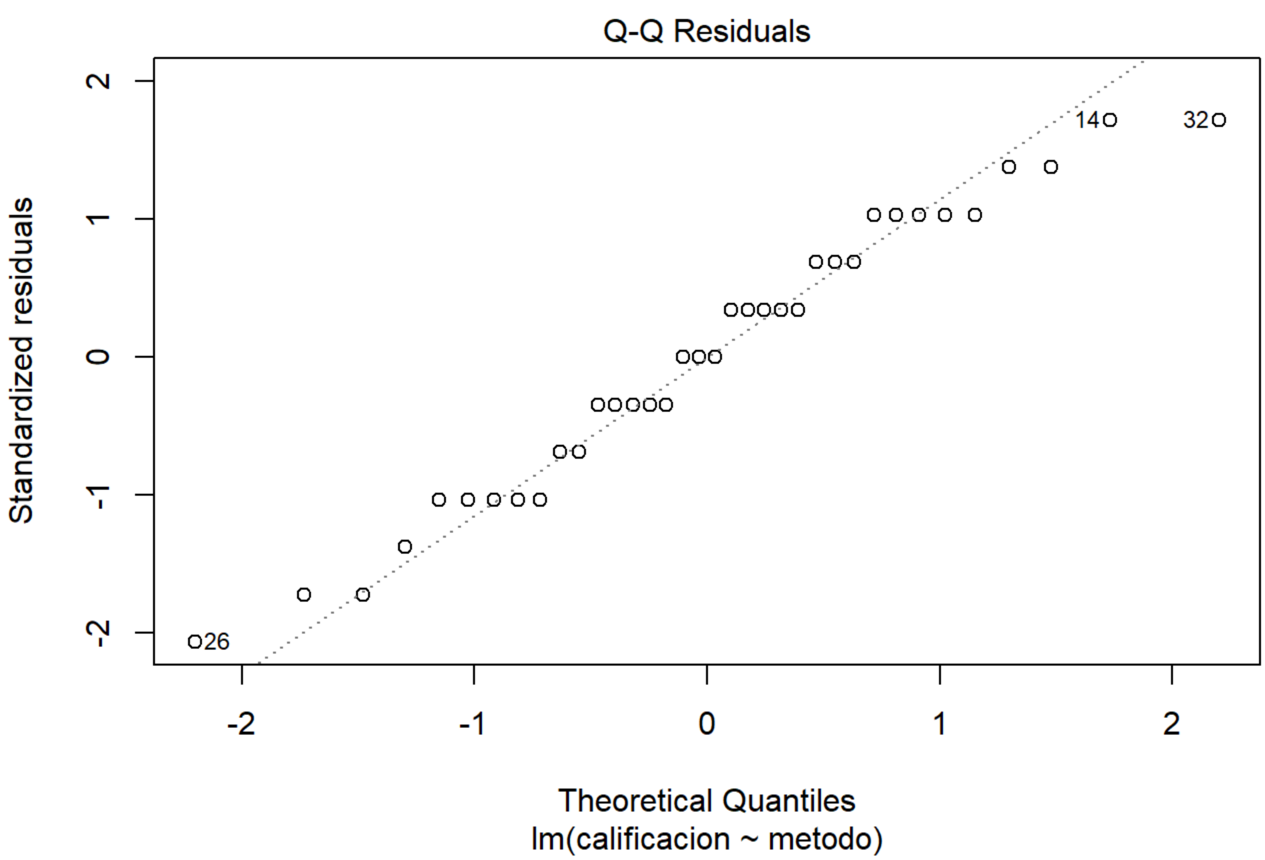
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = calificacion ~ metodo)
##
## $metodo
##        diff       lwr        upr     p adj
## M2-M1 -2.5 -4.020241 -0.9797592 0.0008674
## M3-M1 -5.0 -6.520241 -3.4797592 0.0000000
## M3-M2 -2.5 -4.020241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

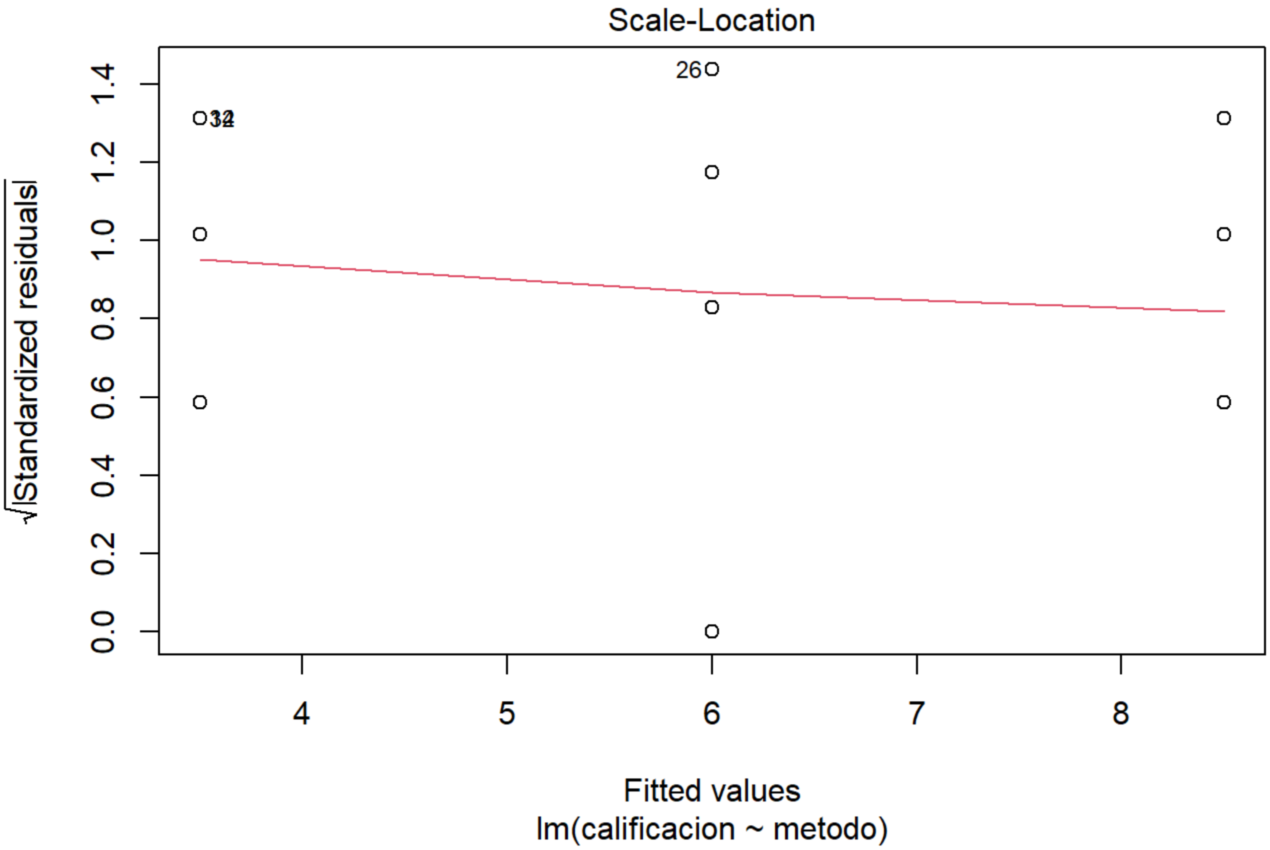
```
plot(I)
```

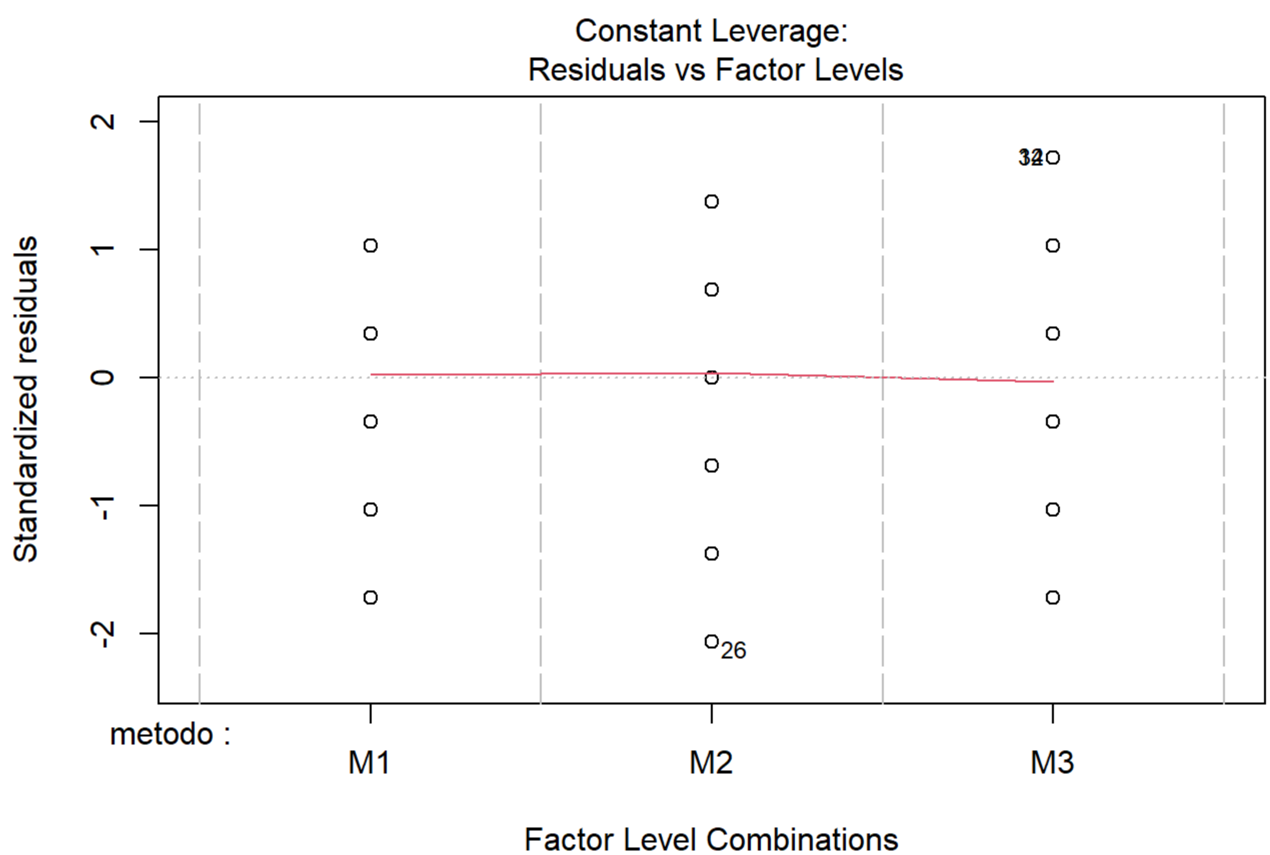
Podemos ver que el sexo tiene una media parecida, a un nivel de confianza del 95% las medias no tienen diferencia, mientras que los metodos tienen medias completamente diferentes. ## Análisis de residuales

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


Residuals vs Fitted
Residuals
Fitted values
lm(calificacion ~ metodo)
3
2
1
0
-1
-2
-3
4
5
6
7
8
26


Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(calificacion ~ metodo)
2
1
0
-1
-2
14
32
26


Scale-Location
26
3
1
2
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
√|Standardized residuals|
4
5
6
7
8
Fitted values
lm(calificacion ~ metodo)


Constant Leverage:
Residuals vs Factor Levels
32
1
2
1
0
-1
-2
26
Standardized residuals
metodo :
M1
M2
M3
Factor Level Combinations

```
CD= 150/(150+76)
cat("Coeficiente de determinación =", CD)
```

```
## Coeficiente de determinación = 0.6637168
```

Al elegir un modelo que solo toma en cuenta el metodo de ensañanza podemos ver que los residuos siguen un comportamiento que podriamos llamar aleatorio. Por lo que podemos decir que la primera hipotesis no se cumple, ya que tiene un alpha distinto a 0. Por lo que podemos concluir que que el metodo de enseñanza tiene una mayor repercusion en las calificaciones que el sexo, y al final esta variable es la que va a determinar el grado que obtenga el alumno.

Problema 2

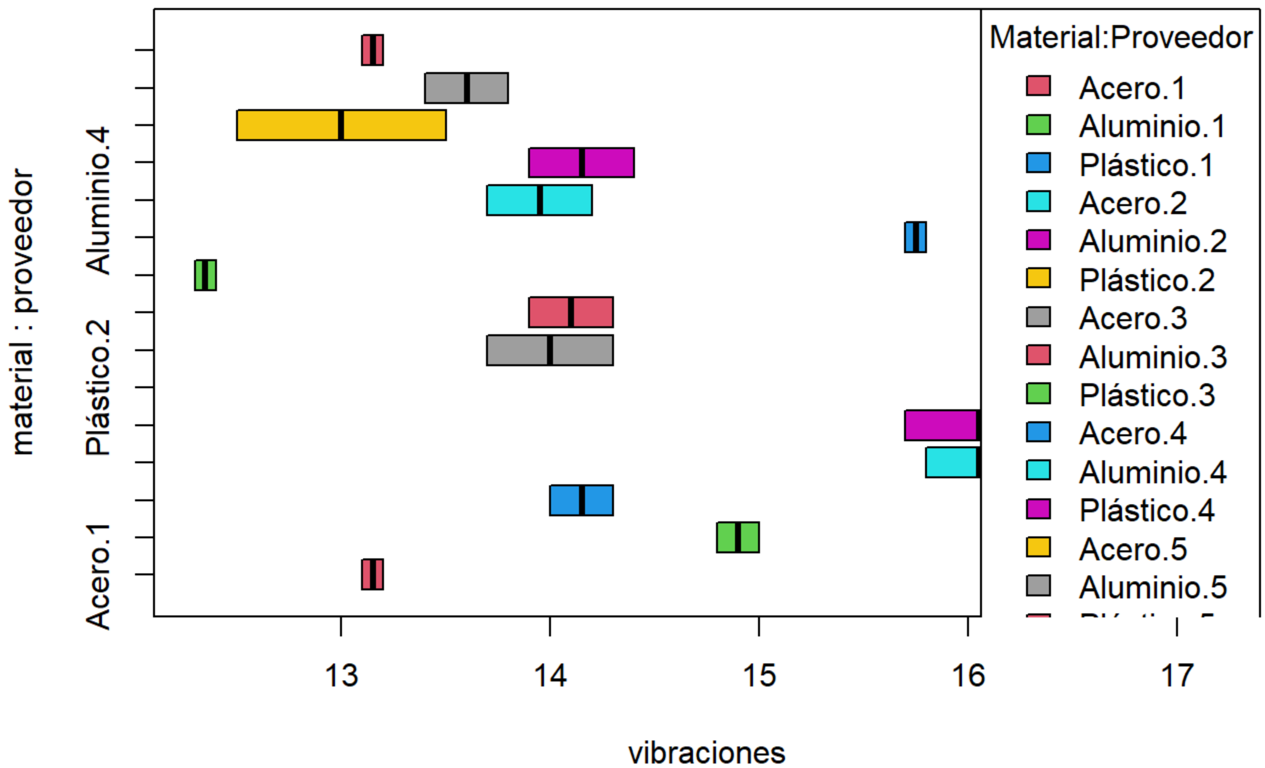
Cargar datos

```
vibraciones=c(13.1, 16.3, 13.7, 15.7, 13.5, 13.2, 15.8, 14.3, 15.8, 12.5, 15.0, 15.7, 13.9, 1
3.7, 13.4, 14.8, 16.4, 14.3, 14.2, 13.8, 14.0, 17.2, 12.4, 14.4, 13.2, 14.3, 16.7, 12.3, 1
3.9, 13.1)
material=c(rep("Acero",10),rep("Aluminio",10),rep("Plástico",10))
proveedor = c(rep("1", 1), rep("2",1), rep("3", 1), rep("4",1), rep("5", 1),rep("1", 1), rep
("2",1), rep("3", 1), rep("4",1), rep("5", 1),rep("1", 1), rep("2",1), rep("3", 1), rep
("4",1), rep("5", 1),rep("1", 1), rep("2",1), rep("3", 1), rep("4",1), rep("5", 1),rep("1",
1), rep("2",1), rep("3", 1), rep("4",1), rep("5", 1),rep("1", 1), rep("2",1), rep("3", 1), re
p("4",1), rep("5", 1))
material = factor(material)
proveedor = factor(proveedor)

datos = data.frame(vibraciones, material, proveedor)

print(datos)
```

```
##    vibraciones material proveedor
## 1         13.1    Acero         1
## 2         16.3    Acero         2
## 3         13.7    Acero         3
## 4         15.7    Acero         4
## 5         13.5    Acero         5
## 6         13.2    Acero         1
## 7         15.8    Acero         2
## 8         14.3    Acero         3
## 9         15.8    Acero         4
## 10        12.5    Acero         5
## 11        15.0 Aluminio         1
## 12        15.7 Aluminio         2
## 13        13.9 Aluminio         3
## 14        13.7 Aluminio         4
## 15        13.4 Aluminio         5
## 16        14.8 Aluminio         1
## 17        16.4 Aluminio         2
## 18        14.3 Aluminio         3
## 19        14.2 Aluminio         4
## 20        13.8 Aluminio         5
## 21        14.0 Plástico         1
## 22        17.2 Plástico         2
## 23        12.4 Plástico         3
## 24        14.4 Plástico         4
## 25        13.2 Plástico         5
## 26        14.3 Plástico         1
## 27        16.7 Plástico         2
## 28        12.3 Plástico         3
## 29        13.9 Plástico         4
## 30        13.1 Plástico         5
```

BoxPlot interacción

```
boxplot(vibraciones~material:proveedor, datos ,col = 2:8 , horizontal = TRUE)
legend("topright", legend = levels(interaction(material, proveedor)), fill = 2:8, title = "Ma
terial:Proveedor")
```



Es dificil encontrar alguna relacion entre los datos ya que hay mucha información y muchas categorias. A primera vista parece que existe menos varianza entre los materiales y lo que determina la vibración es el proveedor ya que estos estan mas agrupados.

Hipótesis

Primera Hipótesis $H_0 : \tau_i = 0$ $H_1 :$ algún τ_i es distinto a 0

Segunda Hipótesis $H_0 : \alpha_j = 0$ $H_1 :$ algún α_j es distinto a 0

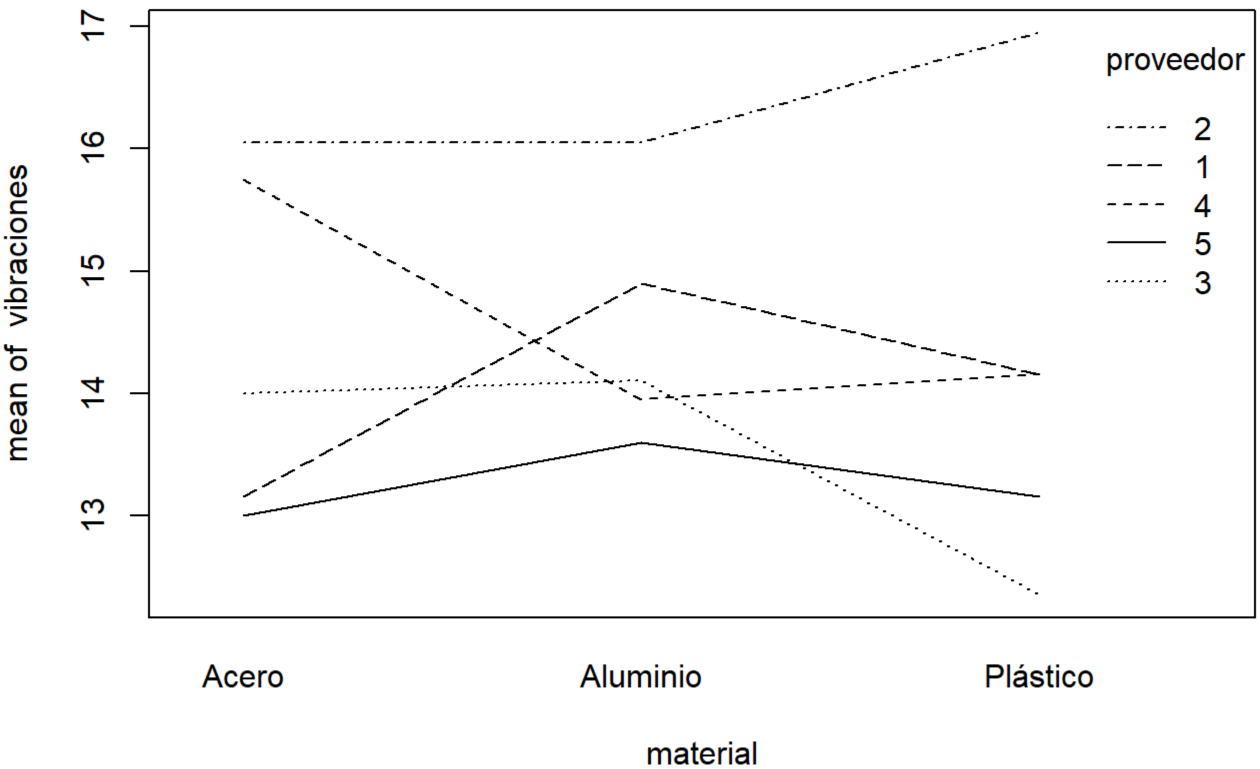
Tercera Hipótesis $H_0 : \tau_i \alpha_j = 0$ $H_1 :$ algún $\tau_i \alpha_j$ es distinto a 0

Anova con interacción

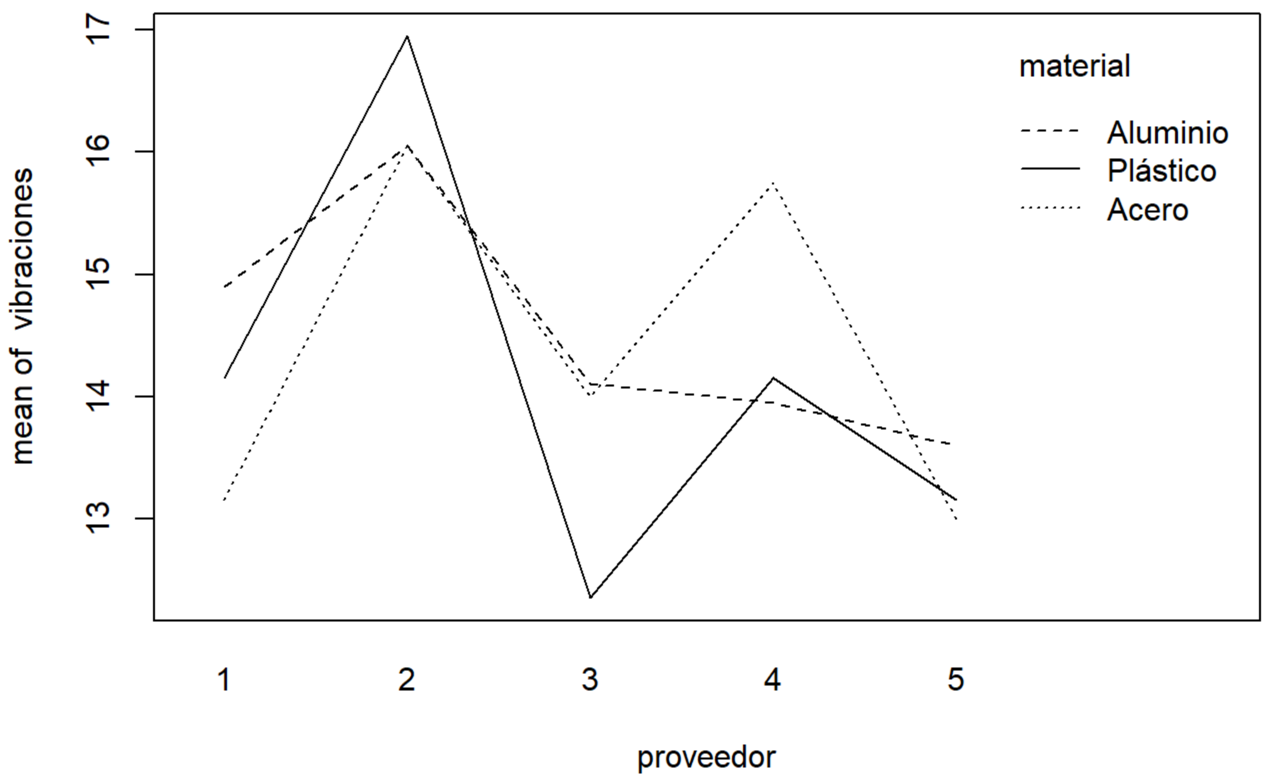
```
A<- aov(vibraciones~material*proveedor, datos)
summary(A)
```

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## material            2   0.70   0.352   3.165   0.0713 .
## proveedor           4  36.67   9.169  82.353 5.07e-10 ***
## material:proveedor  8  11.61   1.451  13.030 1.76e-05 ***
## Residuals          15   1.67   0.111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
interaction.plot(material, proveedor, vibraciones)
```




```
interaction.plot(proveedor, material, vibraciones)
```



La varianza del proveedor es muy grande, al igual que la de la interaccion entre material y proveedor. Por lo que podemos asumir que esto es lo que mayor efecto tiene sobre las vibraciones.

Anova sin interacción

```
B<- aov(vibraciones~material+proveedor, datos)
summary(B)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## material     2   0.70   0.352    0.61    0.552    
## proveedor    4  36.67   9.169   15.88 2.28e-06 ***
## Residuals   23  13.28   0.577                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Al eliminar la interaccion y pasar todo esta variable al error podemos notar lo mismo, la varianza de los materiales es casi inexistente, mientras que el proveedor cambia la variable de vibraciones. ## Boxplot sexo y medias

```
cat("Medias por material y proveedor","\n")
```

```
## Medias por material y proveedor
```

```
tapply(vibraciones,material,mean)
```

```
##    Acero Aluminio Plástico
##    14.39    14.52    14.15
```

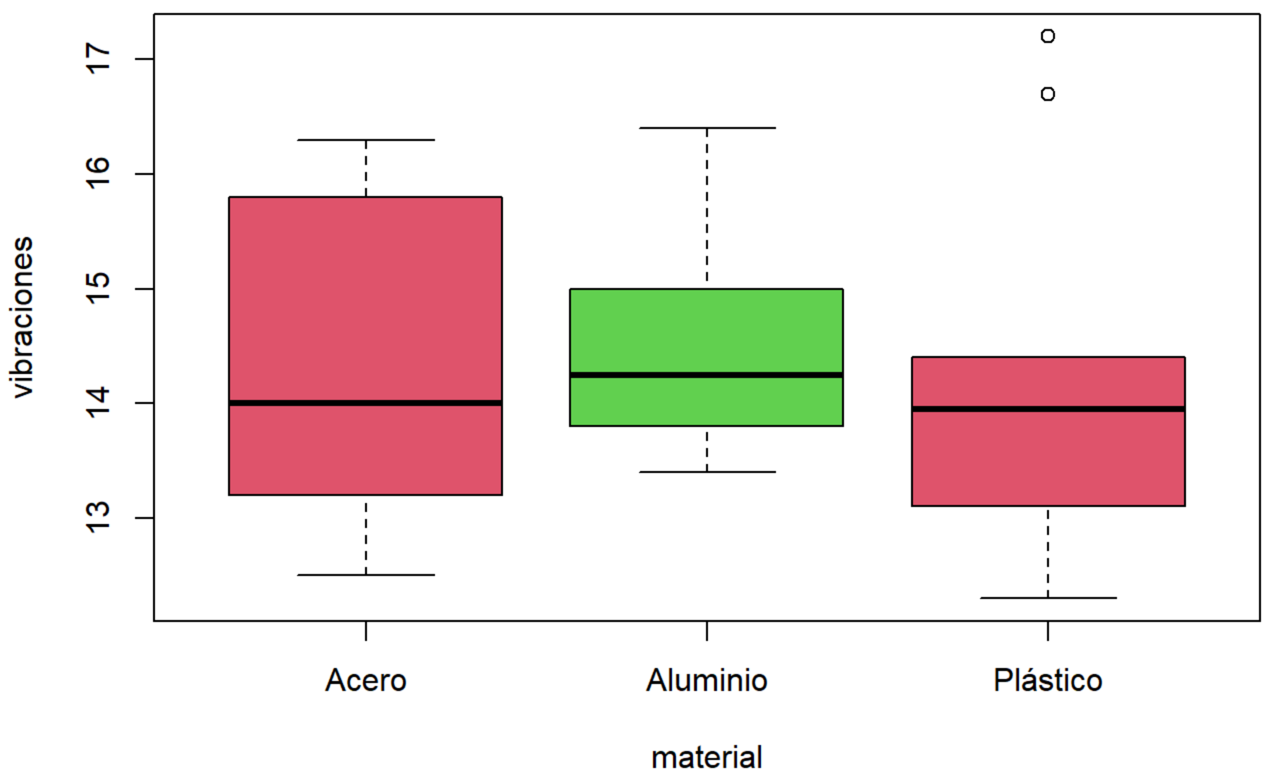
```
tapply(vibraciones,proveedor,mean)
```

```
##        1        2        3        4        5
## 14.06667 16.35000 13.48333 14.61667 13.25000
```

```
M=mean(vibraciones)
cat("Media General=", M)
```

```
## Media General= 14.35333
```

```
boxplot(vibraciones ~ material, col =2:3)
```

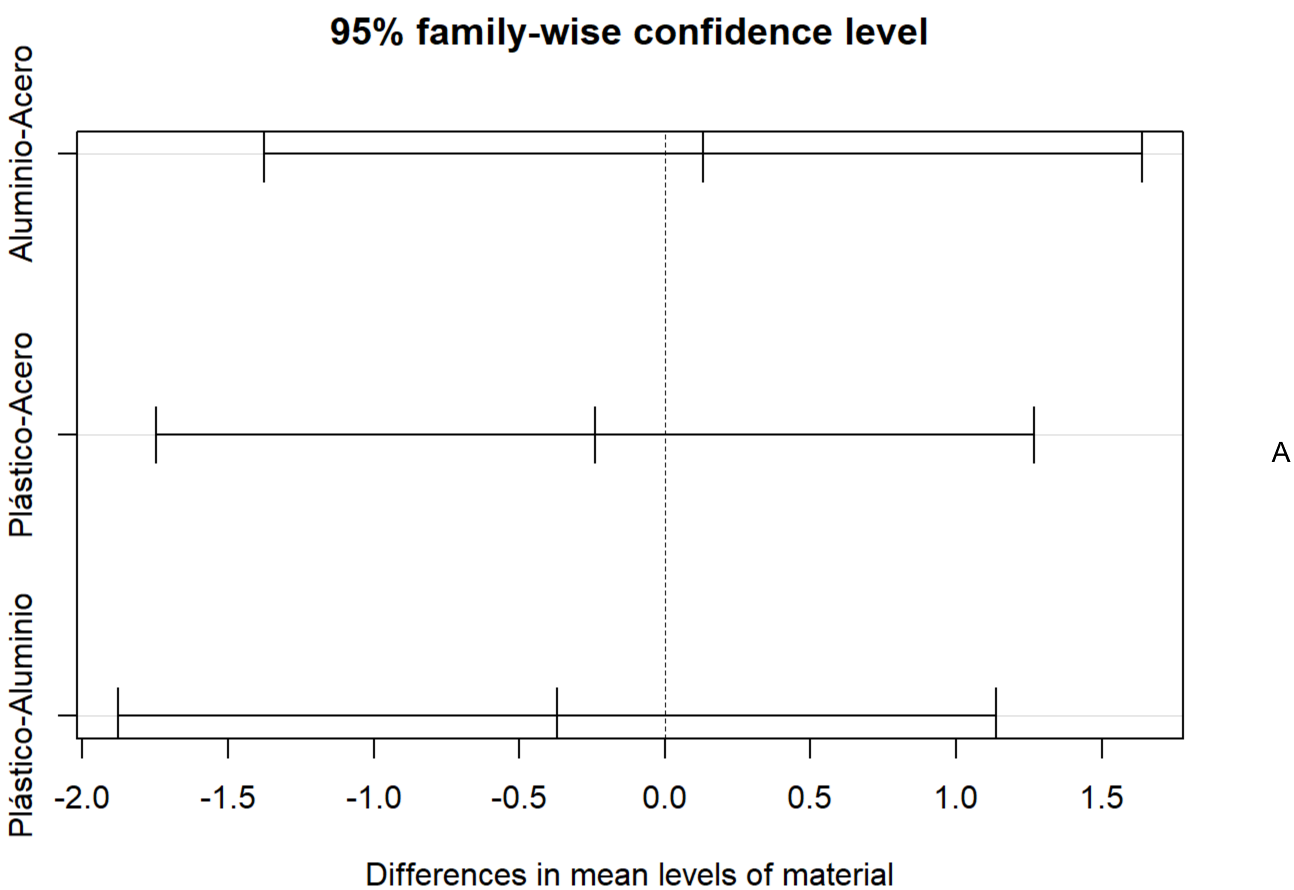


Niveles de confianza de diferencia de medias

```
I = TukeyHSD(aov(vibraciones ~ material))
I
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = vibraciones ~ material)
##
## $material
##                    diff       lwr      upr     p adj
## Aluminio-Acero     0.13 -1.378171 1.638171 0.9751575
## Plástico-Acero    -0.24 -1.748171 1.268171 0.9180284
## Plástico-Aluminio -0.37 -1.878171 1.138171 0.8168495
```

```
plot(I)
```



traves de estas graficas podemos determinar que no tiene gran significacia el material en el numero de vibraciones, las medias son practicamente iguales, ya que sus diferencias se encuentran practicamente en 0.

```
C<-aov(vibraciones~proveedor)
summary(C)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## proveedor    4  36.67   9.169    16.4 1.03e-06 ***
## Residuals   25  13.98   0.559                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

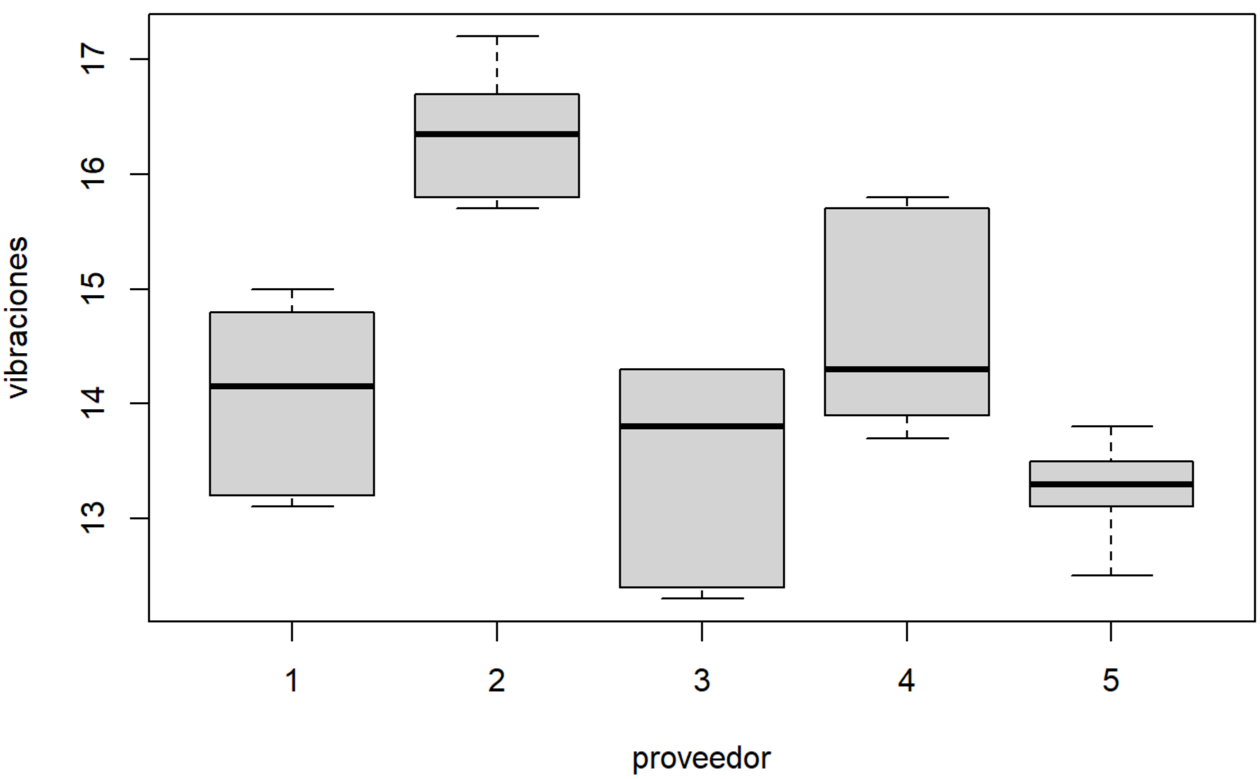
```
tapply(vibraciones,proveedor,mean)
```

```
##        1        2        3        4        5 
## 14.06667 16.35000 13.48333 14.61667 13.25000
```

```
mean(vibraciones)
```

```
## [1] 14.35333
```

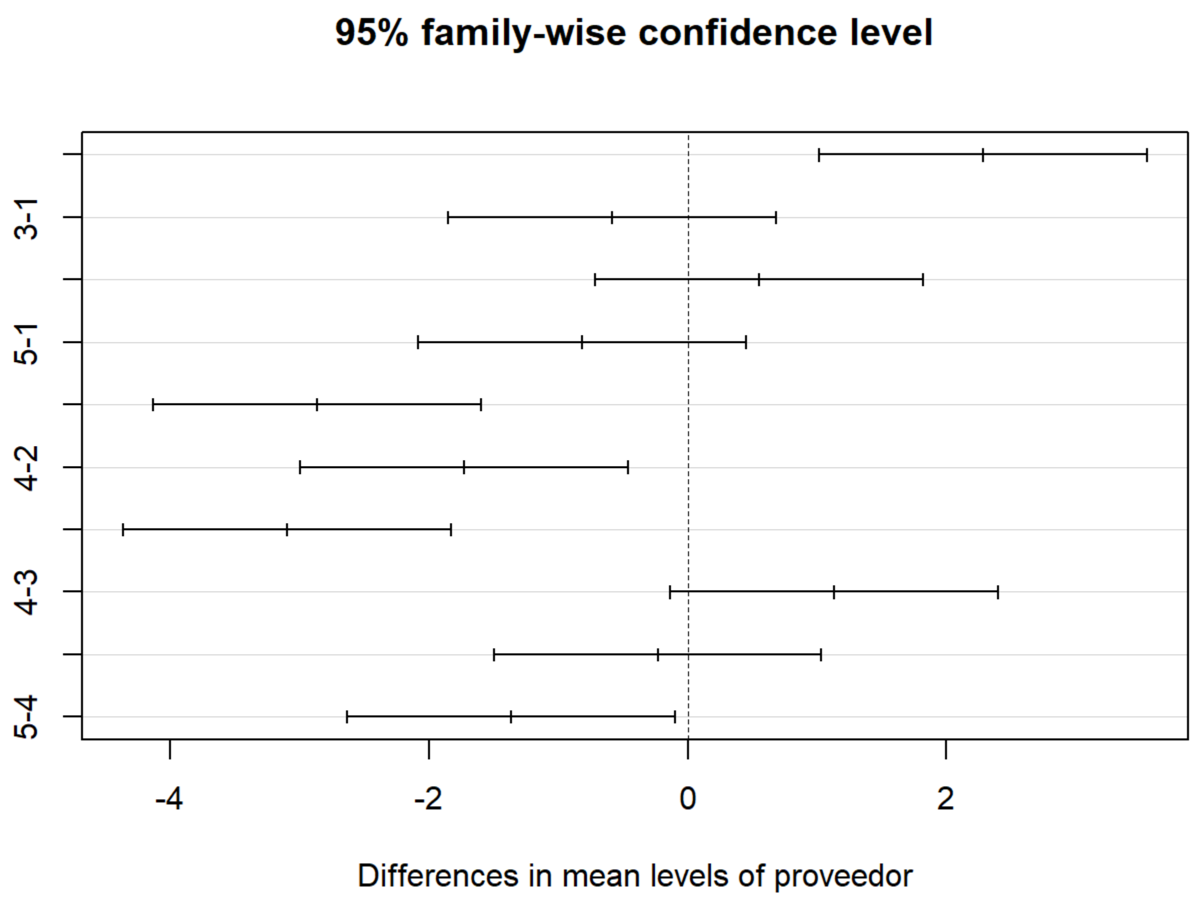
```
boxplot(vibraciones ~ proveedor)
```




```
I = TukeyHSD(aov(vibraciones ~ proveedor))
I
```

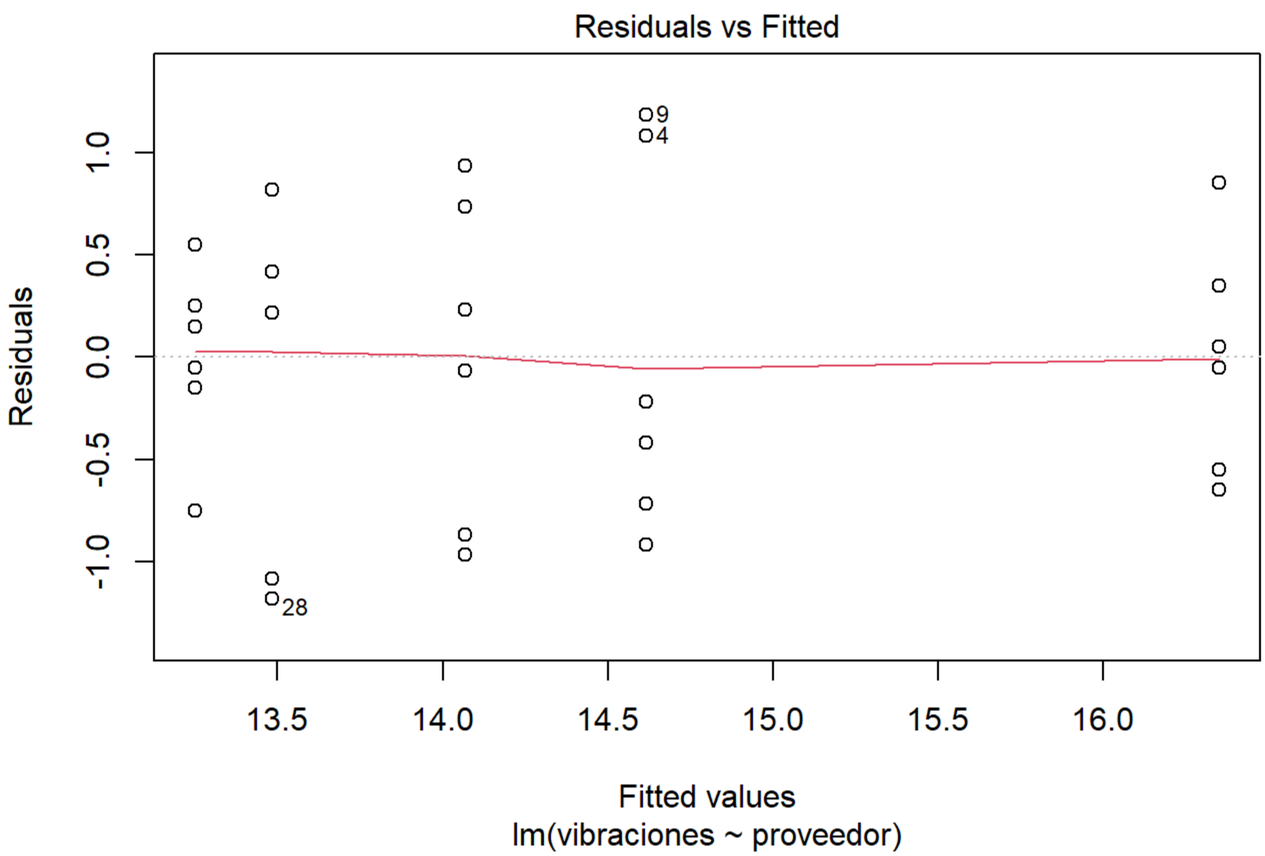
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = vibraciones ~ proveedor)
##
## $proveedor
##           diff        lwr         upr     p adj
## 2-1  2.2833333  1.0153666  3.55130006 0.0001595
## 3-1 -0.5833333 -1.8513001  0.68463339 0.6630108
## 4-1  0.5500000 -0.7179667  1.81796672 0.7089904
## 5-1 -0.8166667 -2.0846334  0.45130006 0.3474956
## 3-2 -2.8666667 -4.1346334 -1.59869994 0.0000055
## 4-2 -1.7333333 -3.0013001 -0.46536661 0.0039774
## 5-2 -3.1000000 -4.3679667 -1.83203328 0.0000015
## 4-3  1.1333333 -0.1346334  2.40130006 0.0959316
## 5-3 -0.2333333 -1.5013001  1.03463339 0.9821261
## 5-4 -1.3666667 -2.6346334 -0.09869994 0.0301318
```

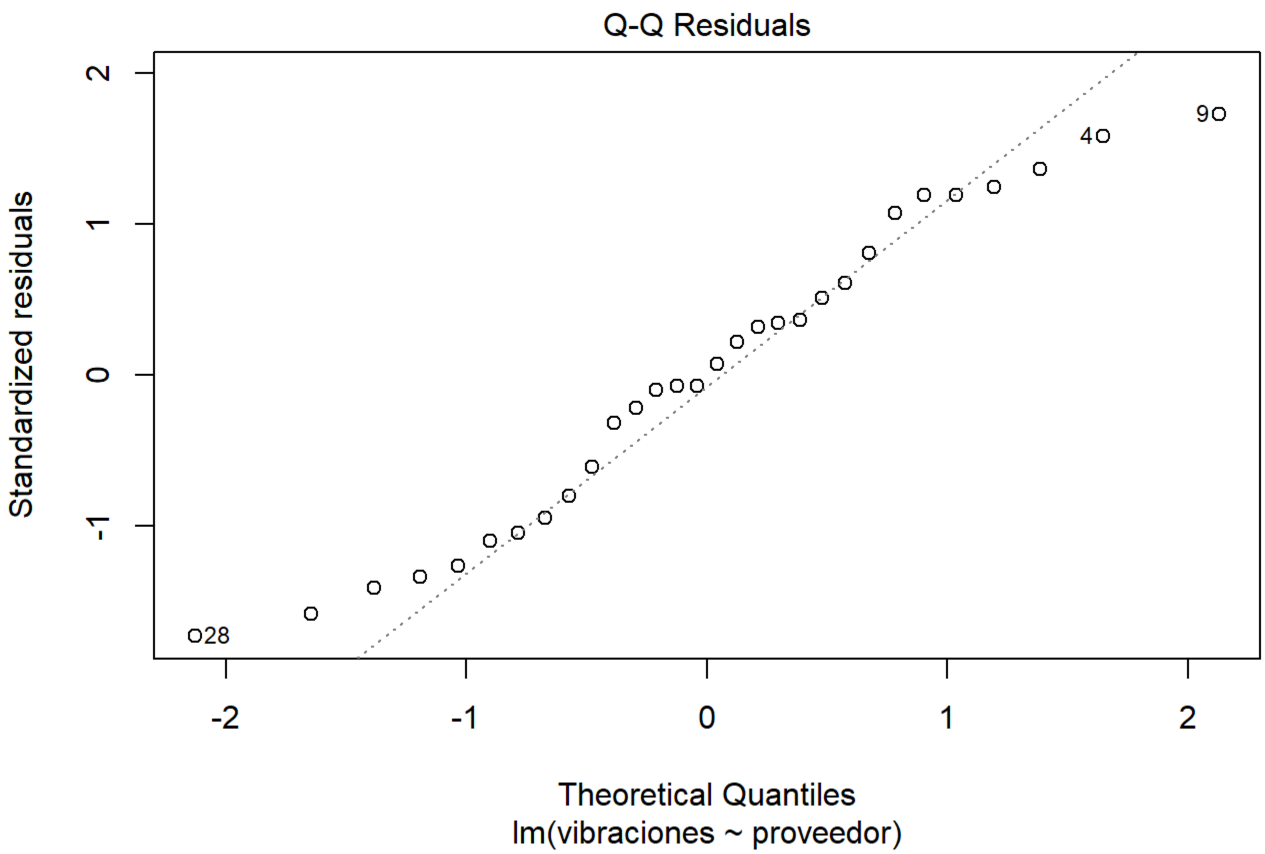
```
plot(I)
```

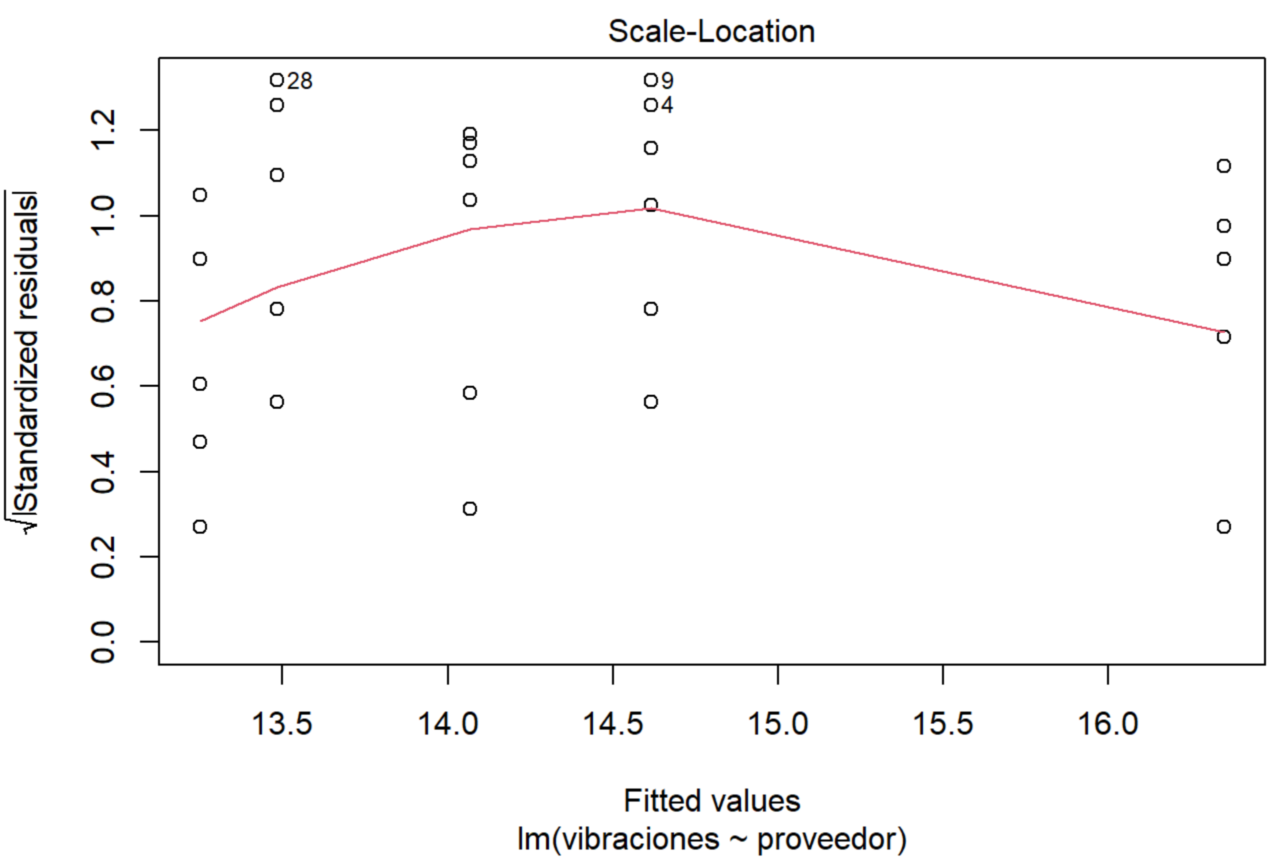


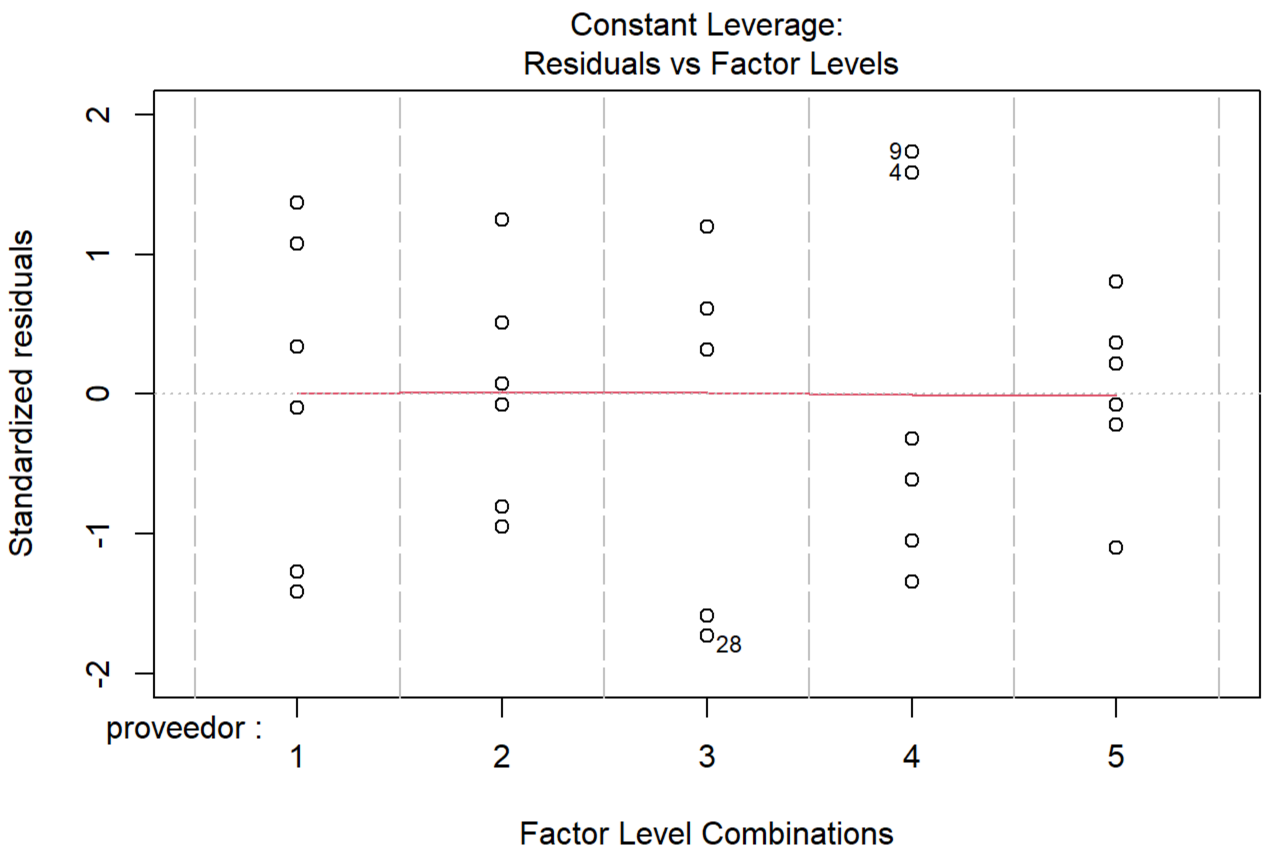
Con estas graficas vemos que el segundo y el cuarto proveedor tienen la mayor cantidad de vibraciones. ## Análisis de residuales

```
plot(lm(vibraciones~proveedor))
```




Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(vibraciones ~ proveedor)
28
4
9
-2
-1
0
1
2


Scale-Location
√|Standardized residuals|
Fitted values
lm(vibraciones ~ proveedor)
28
9
4
0.0
0.2
0.4
0.6
0.8
1.0
1.2
13.5
14.0
14.5
15.0
15.5
16.0




```
CD= (36.67)/(36.67+13.98)
cat("Coeficiente de determinación =", CD)
```

```
## Coeficiente de determinación = 0.7239882
```

Finalmente los residuos son normales por lo que ya no tienen información que brindar y podemos determinar que existe evidecia que el proveedor afecta mucho en cuantas vibraciones va a generar el motor.