Tarea 6. ANOVA

Francisco Castorena, A00827756

2023-09-04

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:

```r
# Chicos
M1H <- c(10,7,9,9,9,10)
M2H <- c(5,7,6,6,8,4)
M3H <- c(2,6,3,5,5,3)

#Chicas
M1M <- c(9,7,8,8,10,6)
M2M <- c(8,3,5,6,7,7)
M3M <- c(2,6,2,1,4,3)

data_df <- data.frame(
  M1H = M1H,
  M2H = M2H,
  M3H = M3H,
  M1M = M1M,
  M2M = M2M,
  M3M = M3M
)
```

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

1. Establece las hipótesis estadísticas (tienen que ser 3).

Primera hipótesis: $H_0 : \tau_i = 0$ $H_1 : \exists \tau_i \neq 0$

Segunda hipótesis: $H_0 : \alpha_j = 0$ $H_1 : \exists \alpha_j \neq 0$

Tercera hipótesis: $H_0 : \tau_i \alpha_j = 0$ $H_1 : \exists \tau_i \alpha_j \neq 0$

2. Realiza el ANOVA para dos niveles con interacción:

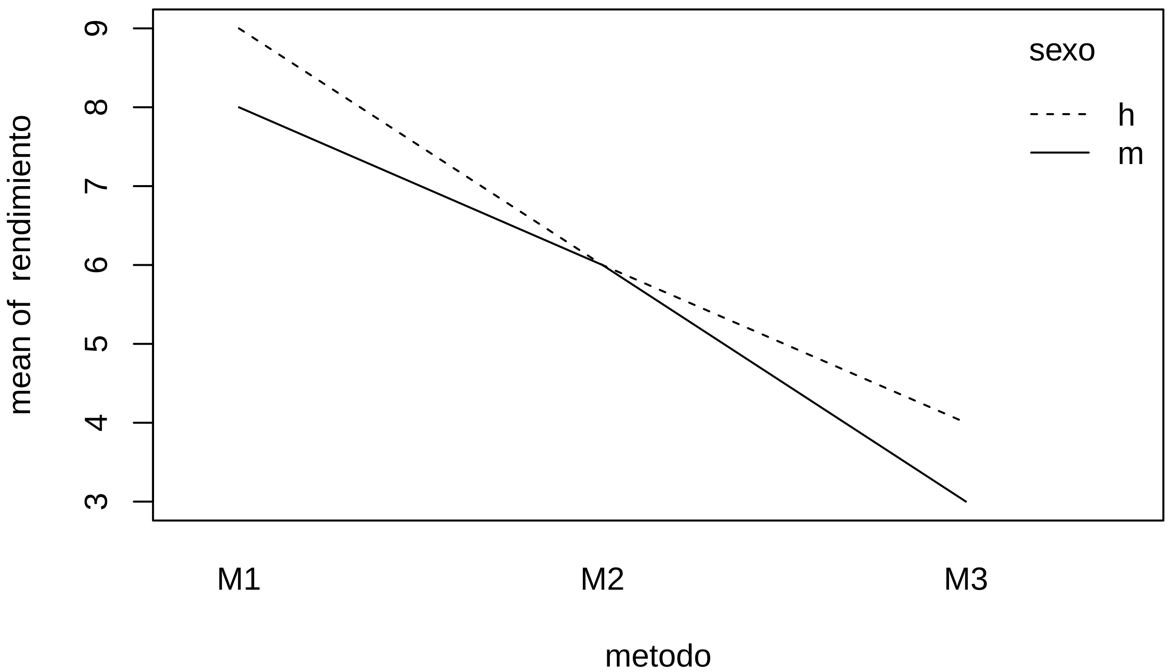
- Consulta el código en R en los apoyos de clase de "ANOVA".
- Haz la gráfica de interacción de dos factores.
- Interpreta el resultado desde la perspectiva estadística y en el contexto del problema
- Escribe tus conclusiones parciales

```r
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)
```

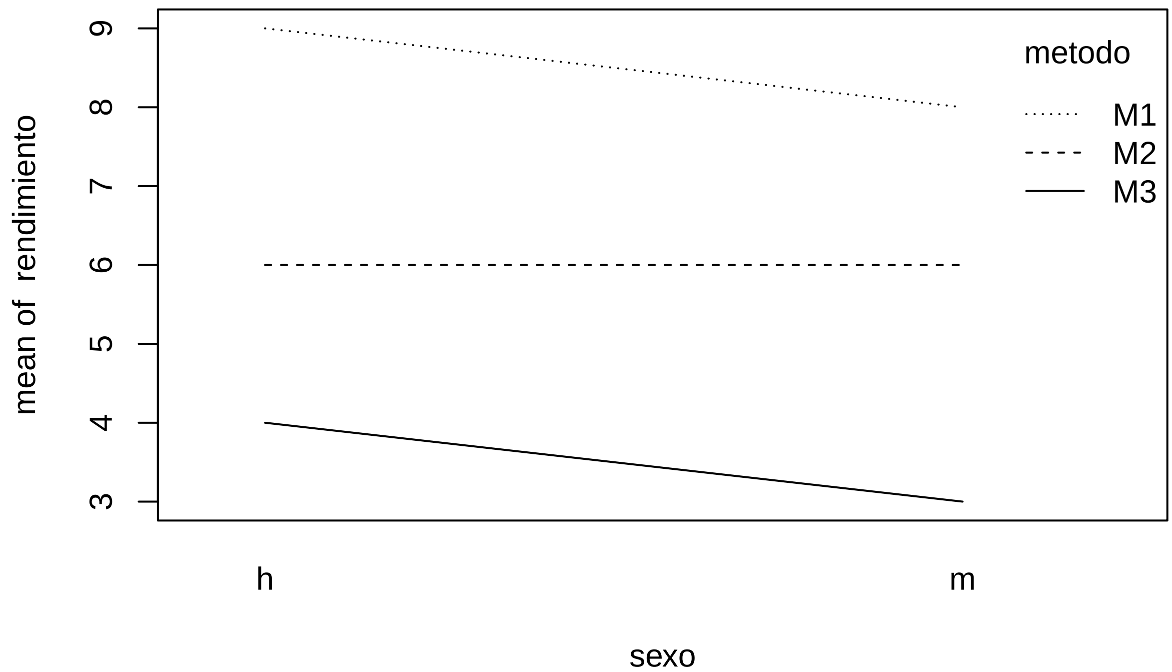
```r
rendimiento = calificacion
A<-aov(rendimiento~metodo*sexo)
summary(A)
```

```
##             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
```

```r
interaction.plot(metodo,sexo,rendimiento)
```




```r
interaction.plot(sexo, metodo, rendimiento)
```

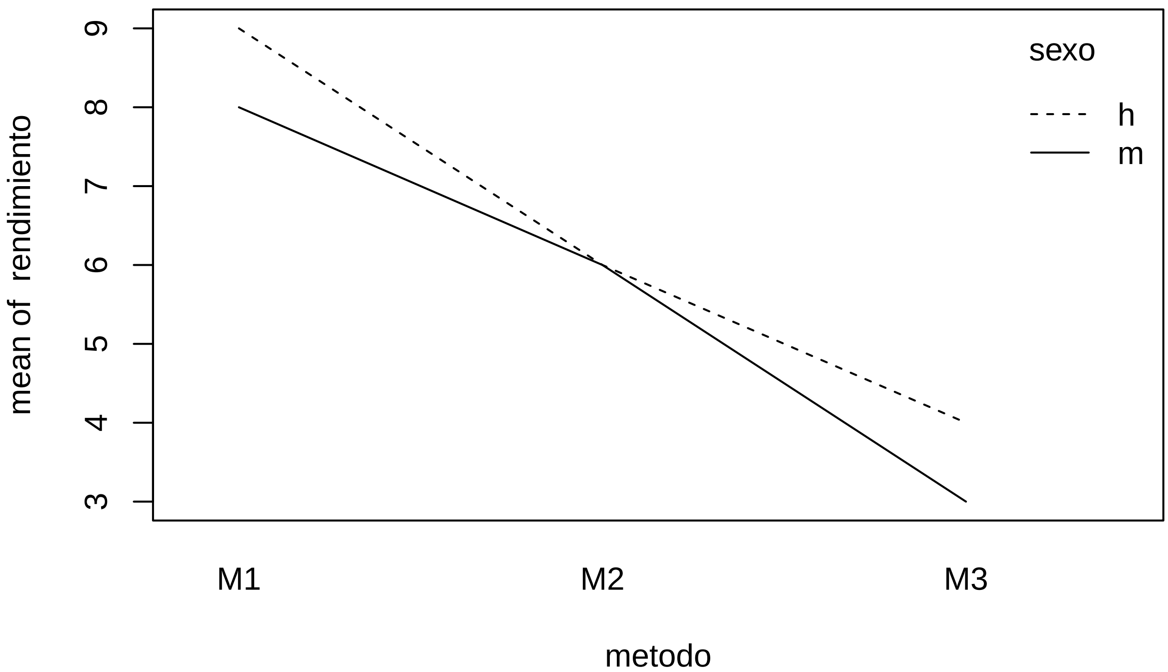


Aquí podemos observar que el método es el la única variable que parece afectar a la varianza del rendimiento, también vemos que no hay efecto significativo en el rendimiento cuando hay interacción entre sexo y método, tampoco hay efecto significativo sobre el rendimiento por sexo.

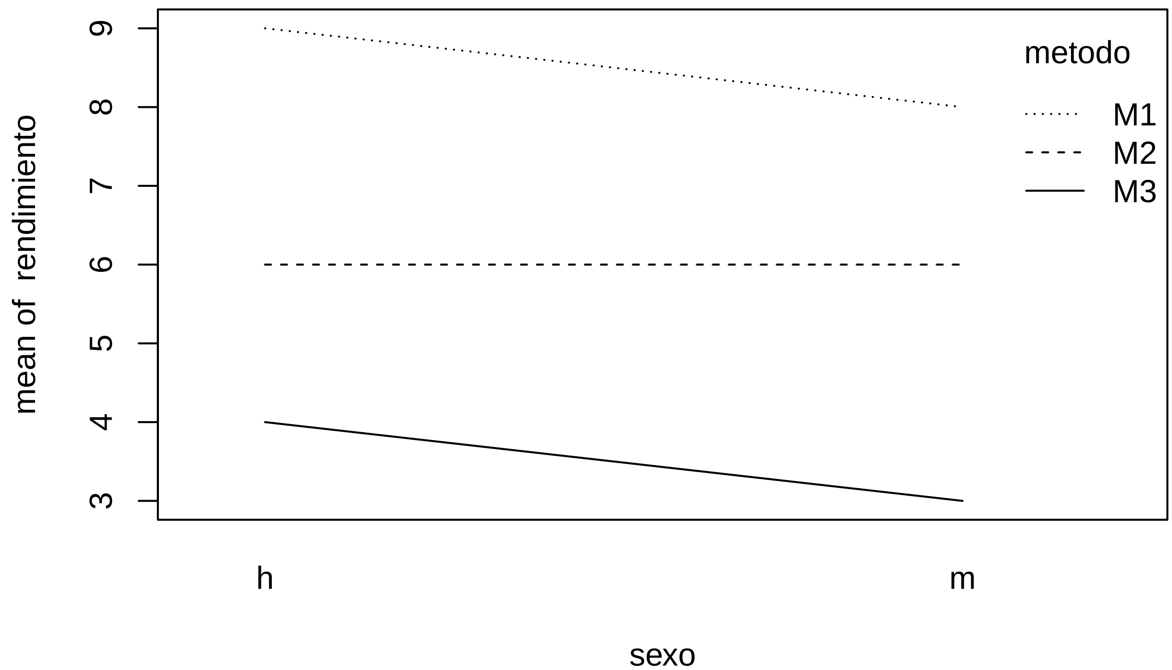
```
rendimiento = calificacion
A<-aov(rendimiento~metodo+sexo)
summary(A)
```

```
##             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
```

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




```r
interaction.plot(sexo, metodo, rendimiento)
```



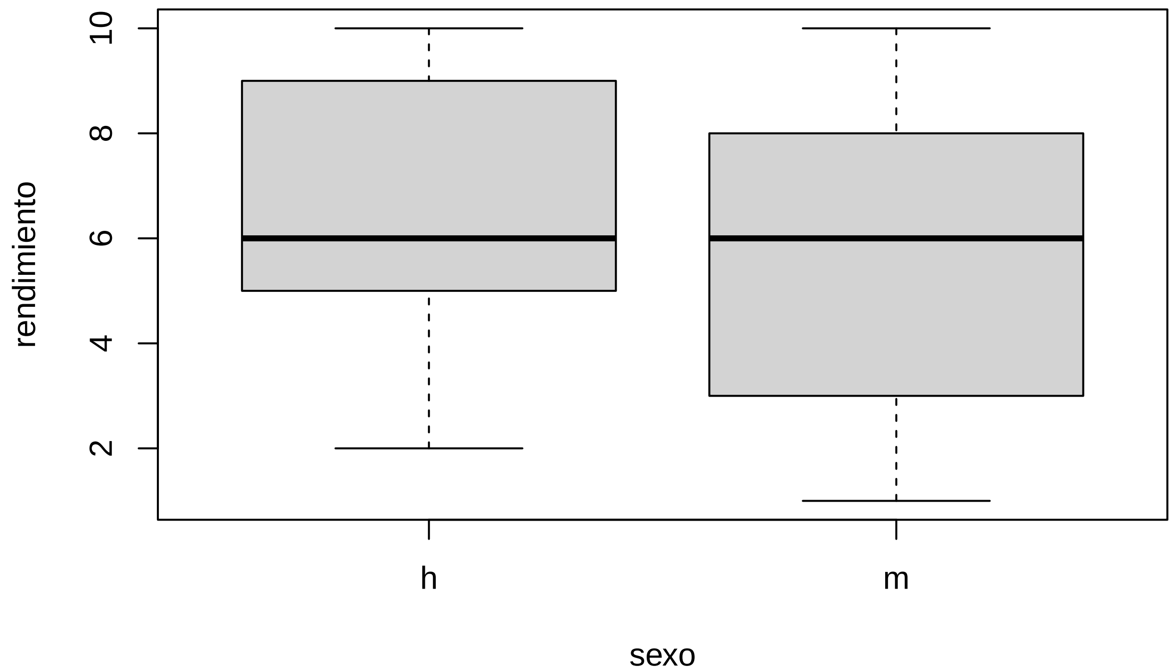

```
#apply(rendimiento,metodo,mean)
tapply(rendimiento,metodo,mean)
```

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

```
M=mean(rendimiento)
M
```

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

```
boxplot(rendimiento ~ sexo)
```



Podemos ver que las medias por grupo son significativas,como lo indica el anova hecho anteriormente, por otra parte, se ve que la mediana y los cuartiles extremos por sexo son muy similares, indicando que el sexo no es significativo en el desempeño.

```
C<-aov(rendimiento~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
```

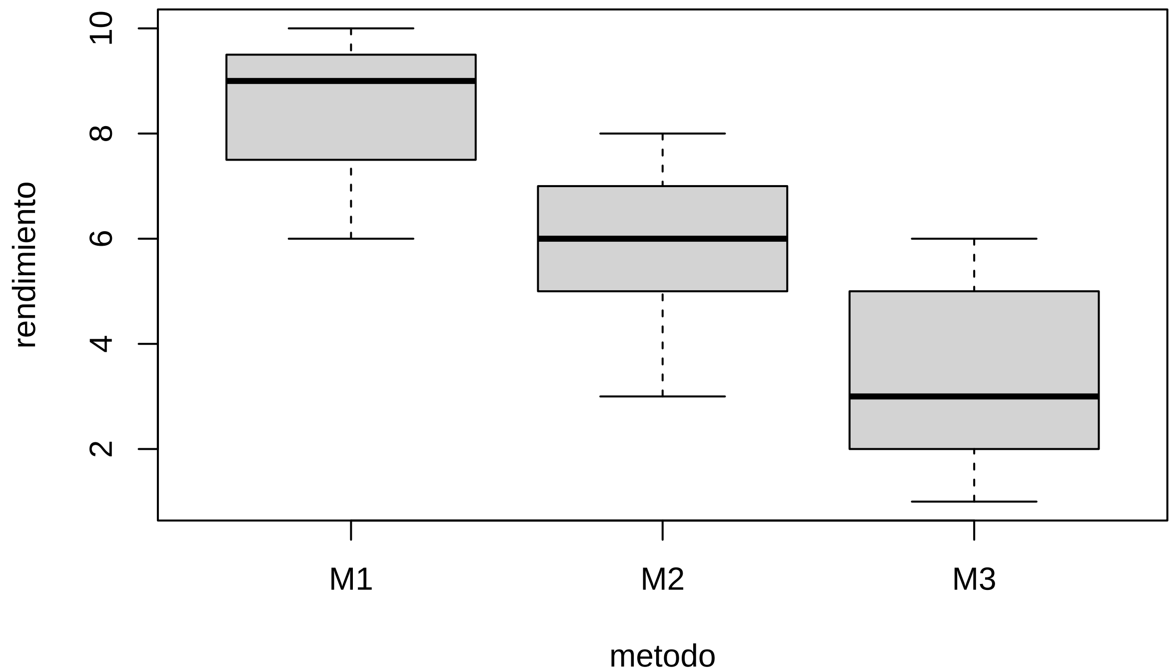
```
tapply(rendimiento,metodo,mean)
```

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

```
mean(rendimiento)
```

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

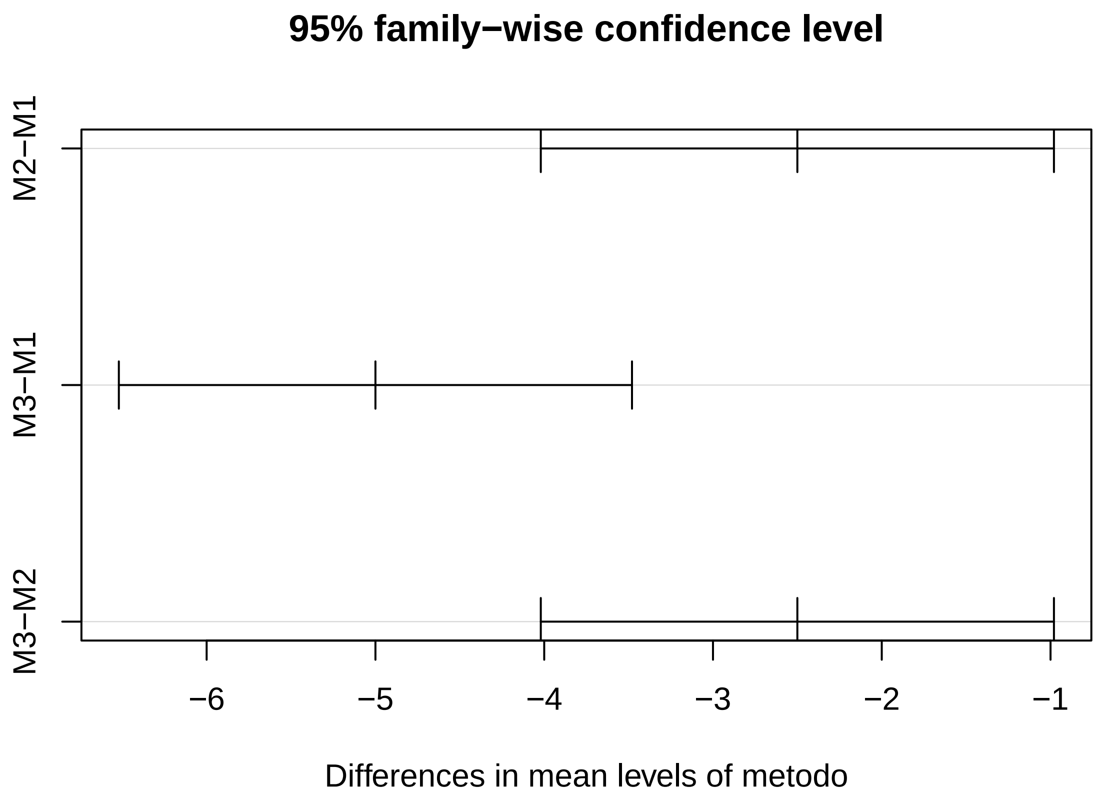
```
boxplot(rendimiento ~ metodo)
```




```r
I = TukeyHSD(aov(rendimiento ~ metodo))
I
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = rendimiento ~ 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
```

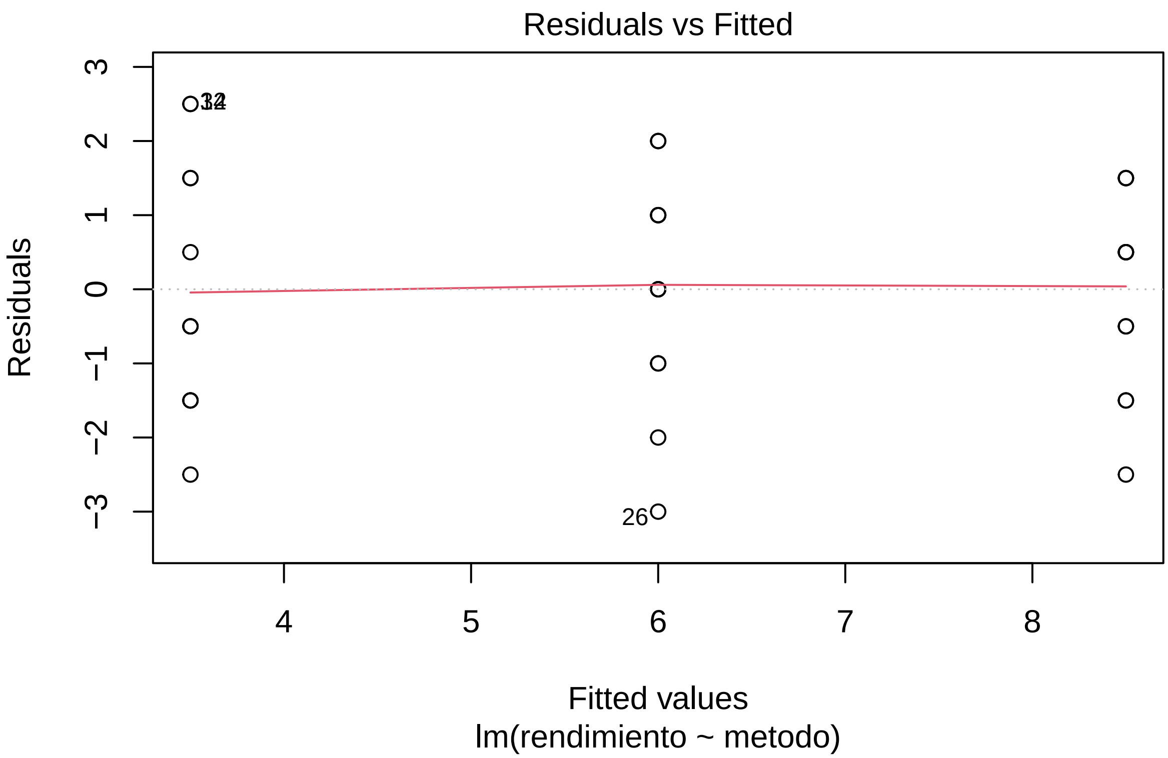
```r
plot(I) #Los intervalos de confianza se observan
```

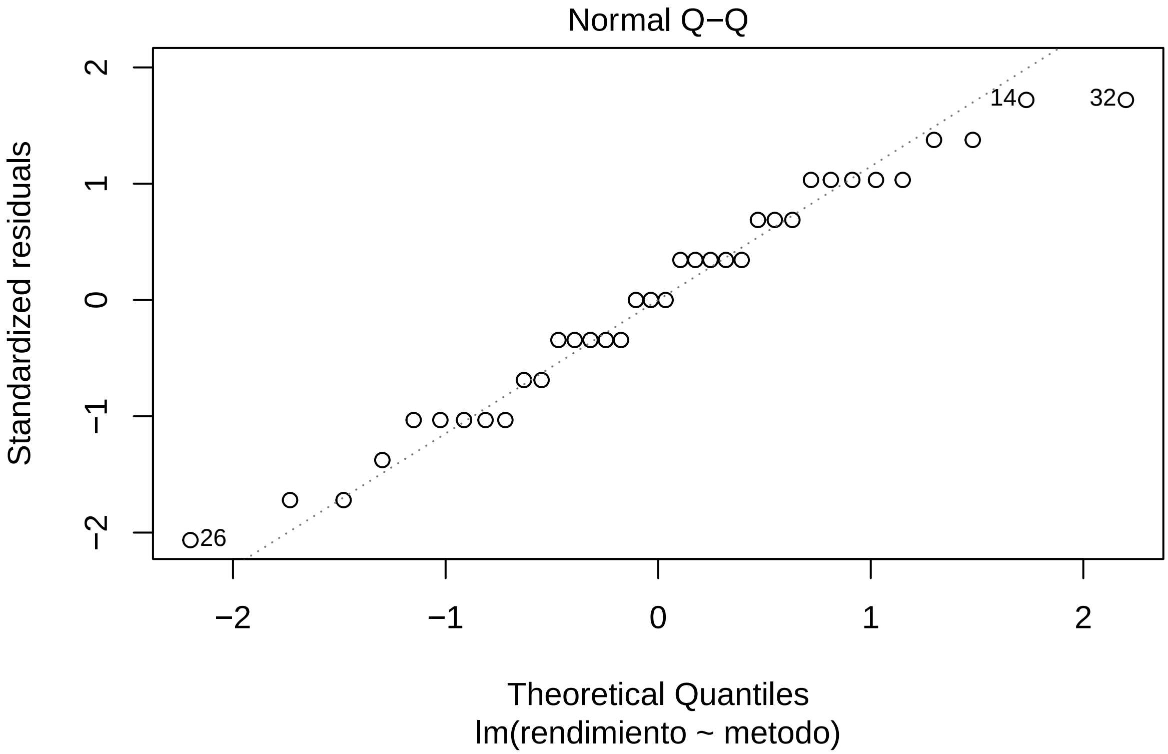


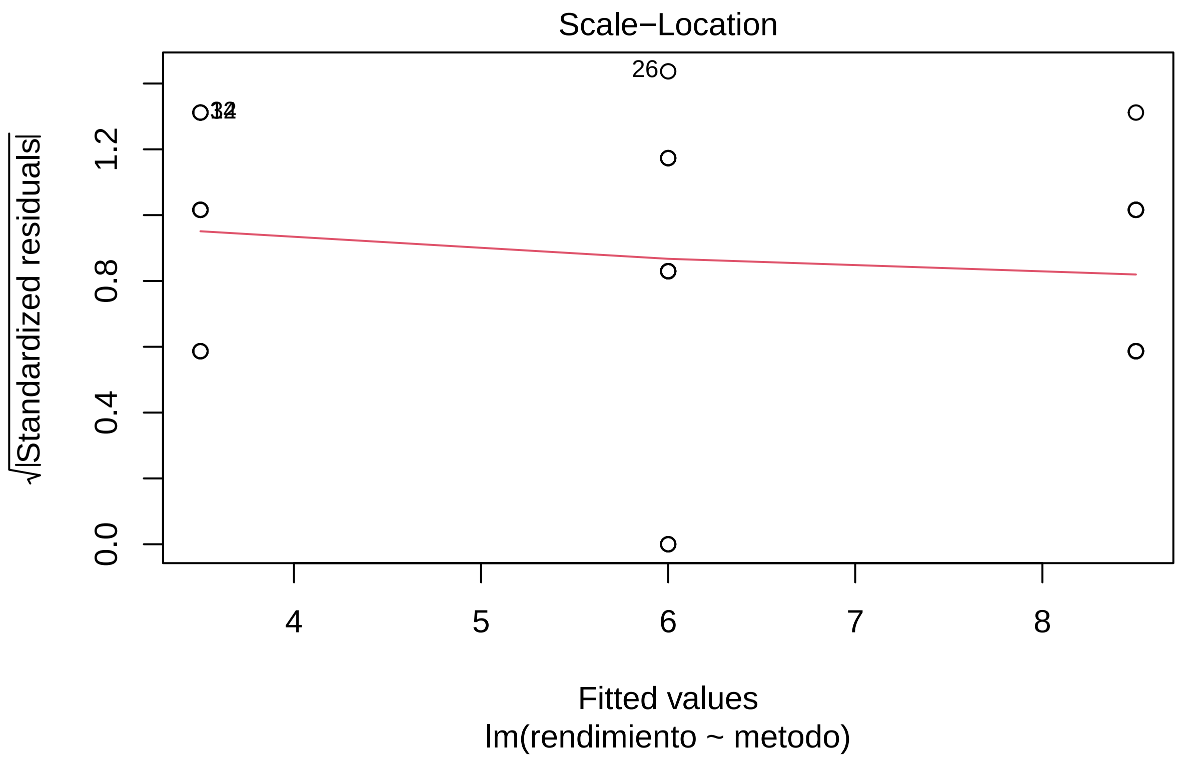

```r
# mejor si se grafican
```

Aquí observamos como con un 95% de confianza se definen los intervalos de las diferencias entre grupos de la variable metodo, vemos como siempre hay una diferencia menor a -1, lo cual es significativo, como lo indica el test de Tukey, todos los p-values entre grupos son menores a .05, con lo cual sabemos que hay una diferencia significativa entre las medias de desempeño por grupo.

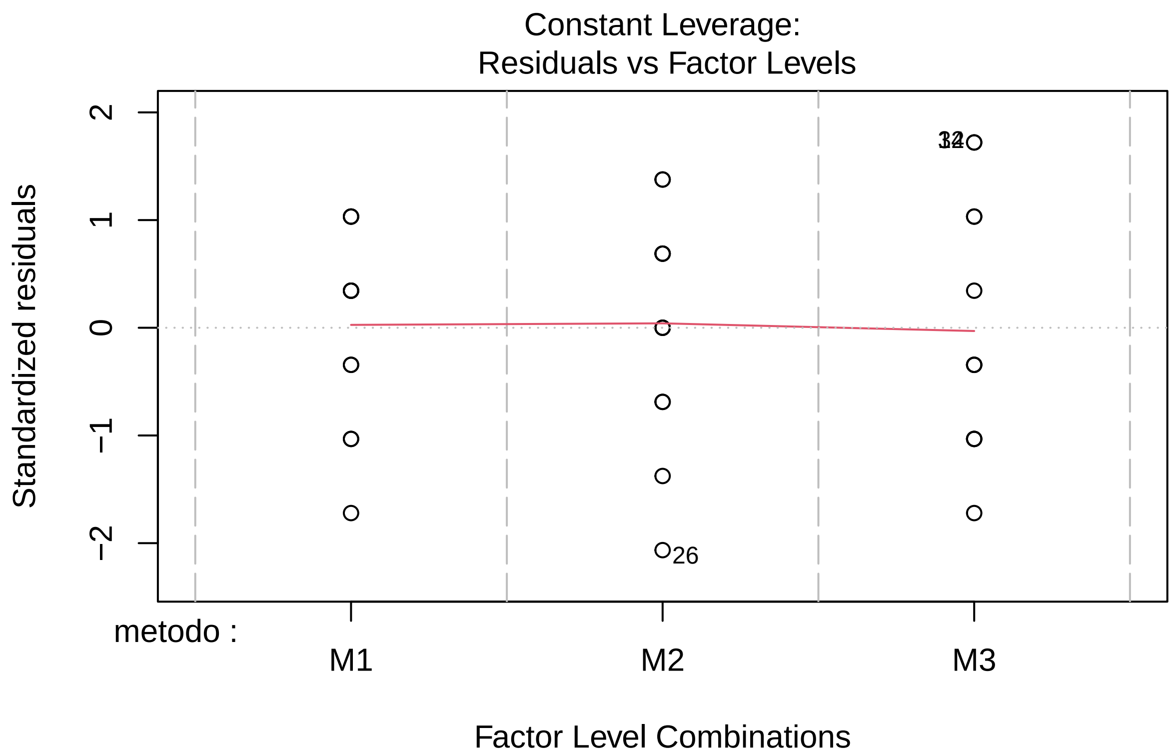
```r
model <- lm(rendimiento~metodo)
plot(lm(rendimiento~metodo))
```


Residuals vs Fitted
3
2
1
0
−1
−2
−3
Residuals
32
14
26
4
5
6
7
8
Fitted values
lm(rendimiento ~ metodo)


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
26
14
32
−2
−1
0
1
2
Theoretical Quantiles
lm(rendimiento ~ metodo)



Scale−Location
26
3
14
$\sqrt{|\text{Standardized residuals}|}$
0.0
0.4
0.8
1.2
4
5
6
7
8
Fitted values
lm(rendimiento ~ metodo)




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

```
##
## Call:
## lm(formula = rendimiento ~ metodo)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.000 -1.125  0.000  1.125  2.500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.5000     0.4381  19.403  < 2e-16 ***
## metodoM2     -2.5000     0.6195  -4.035 0.000304 ***
## metodoM3     -5.0000     0.6195  -8.070 2.59e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.518 on 33 degrees of freedom
## Multiple R-squared:  0.6637, Adjusted R-squared:  0.6433
## F-statistic: 32.57 on 2 and 33 DF,  p-value: 1.551e-08
```

Vemos como el 64% de la varianza del desempeño puede ser explicada por el modelo lineal implementado.

VIbración 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.

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