

# ANOVA

Diego Alberto Baños Lopez  
A01275100

03-09-2023

## Problema: 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.

Método 1	Método 2	Método 3
<b>Chicos</b>	<b>Chicos</b>	<b>Chicos</b>
10	5	2
7	7	6
9	6	3
9	6	5
9	8	5
10	4	3
<b>Chicas</b>	<b>Chicas</b>	<b>Chicas</b>
9	8	2
7	3	6
8	5	2
8	6	1
10	7	4
6	7	3

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

## Introducción de datos en R

```
rendimiento = 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)

```

## Hipotesis de este Problema

En base al problema generaremos las siguientes hipotesis y en base a ellas se realizará una comprobación de las siguientes:

- Primera hipotesis:
  - $H_0 : \tau_i = 0$  (no hay efecto del método de enseñanza)
  - $H_1 : \tau_i$  es distinto de cero
- Segunda hipotesis:
  - $H_0 : \alpha_j = 0$  (no hay efecto del sexo)
  - $H_1 : \alpha_j$  es distinto de cero
- Tercera hipotesis:
  - $H_0 : \tau_i \alpha_j = 0$  (No hay interacción entre el método de enseñanza y el sexo en el rendimiento)
  - $H_1 : \tau_i \alpha_j$  es distinto de cero

## Análisis

### ANOVA

#### Con interacción

```

A <- aov(rendimiento ~ metodo * sexo)
summary(A)

```

```

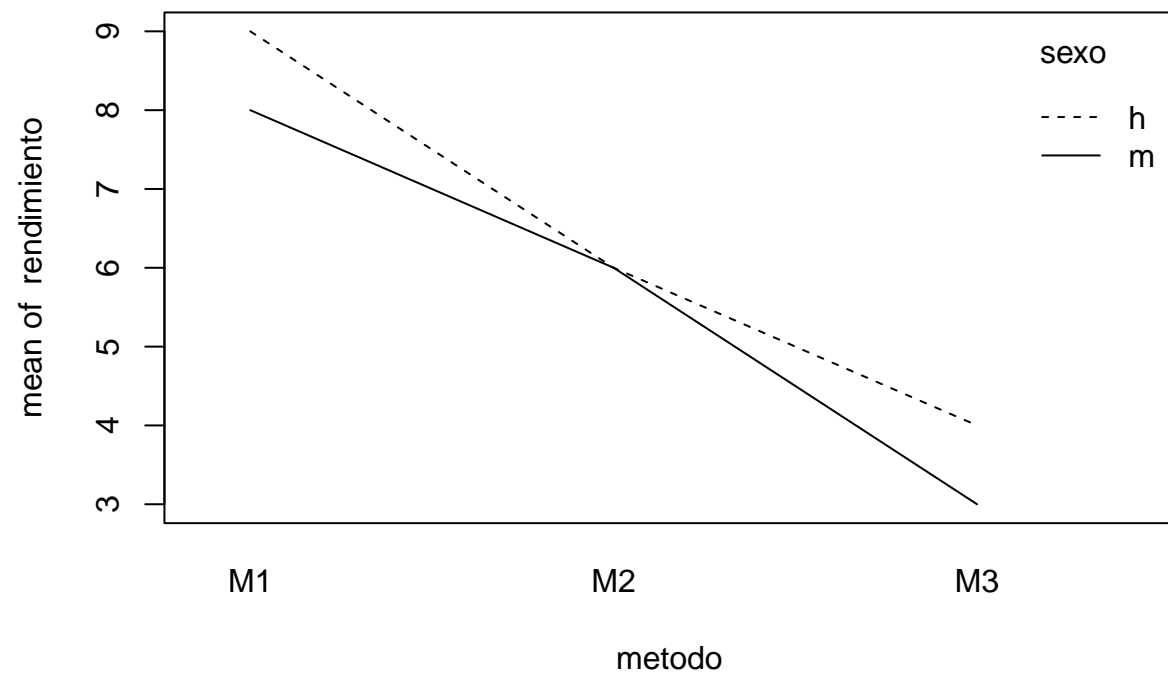
##              Df Sum Sq Mean Sq F value    Pr(>F)
## metodo         2    150   75.00  32.143 3.47e-08 ***
## sexo           1     4    4.00   1.714   0.200
## metodo:sexo     2     2    1.00   0.429   0.655
## Residuals     30     70    2.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05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

```

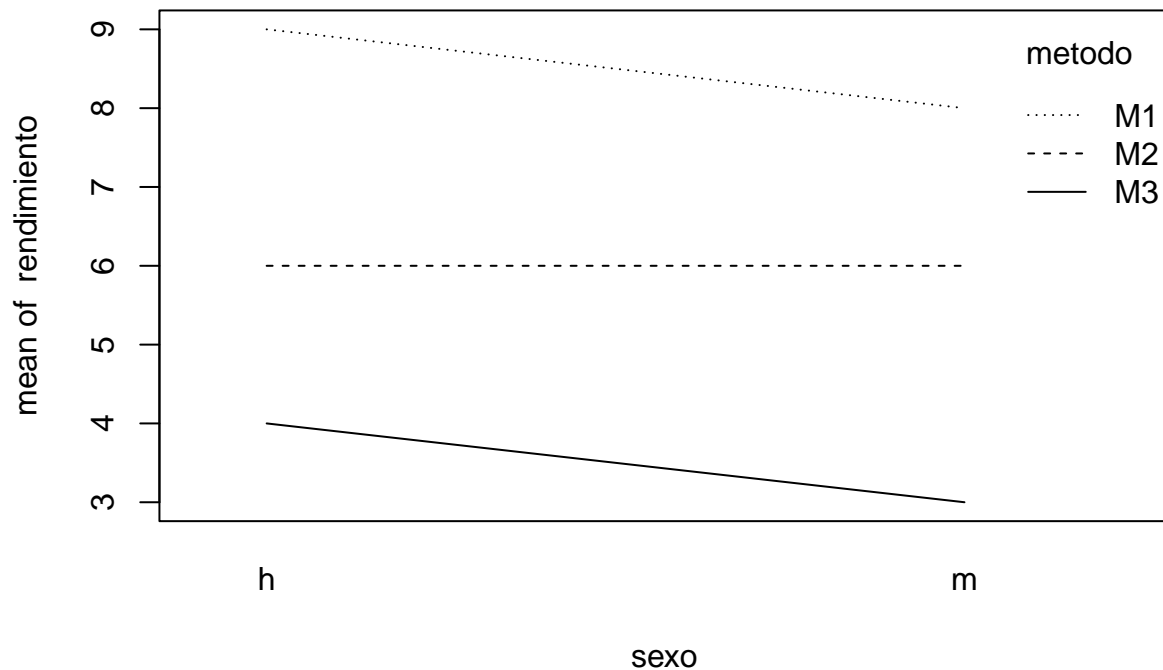
```

interaction.plot(metodo, sexo, rendimiento)

```



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



En ANOVA con interaccion se observa lo siguiente:

- Suponiendo que el valor  $P = \Pr(>F)$
- El valor P del metodo de enseñanza es menor a 0.05 ( $3.47e-08$ ), por lo que rechaza la hipótesis nula  $H_0$
- El valor P del sexo es mayor a 0.05 (0.200), por lo que no rechazamos la hipótesis nula  $H_0$ , o sea que no es significativo
- El valor P de la interaccion enseñanza - sexo es mayor a 0.05 (0.655) por lo Por lo tanto, no rechazamos la hipótesis nula  $H_0$ , o sea que no es significativa

### Sin interacción

```
B = aov(rendimiento ~ 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
```

Para observar mejor los efectos de los factores principales, se calcula la media por nivel y se grafica por nivel. También se calcula la media general.

```
tapply(rendimiento, sexo, mean)
```

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

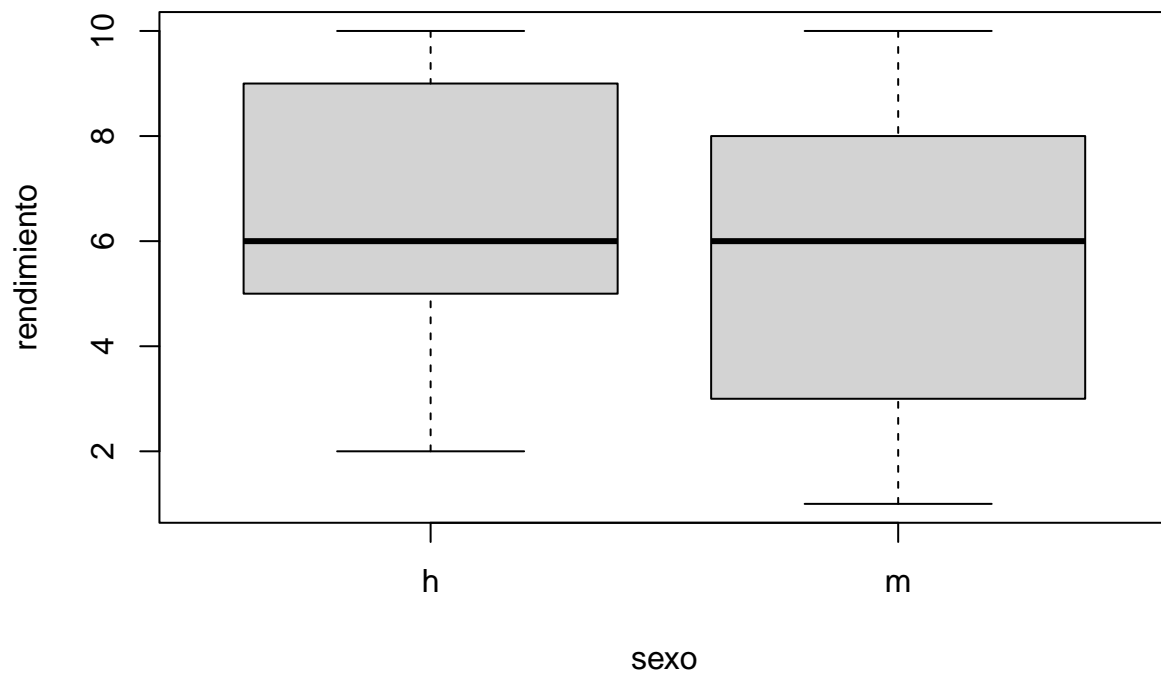
```
tapply(rendimiento, metodo, mean)
```

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

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

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

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



En ANOVA sin interaccion se observa lo siguiente:

- Para metodologia al igual que en el modelo anterior, el valor P es significativo (1.5e-08).
- Para sexo el valor P sigue siendo no significativo (0.192)

## ANOVA con un solo factor (el significativo)

En el modelo, se consideran sólo el efecto significativo.

```
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
```

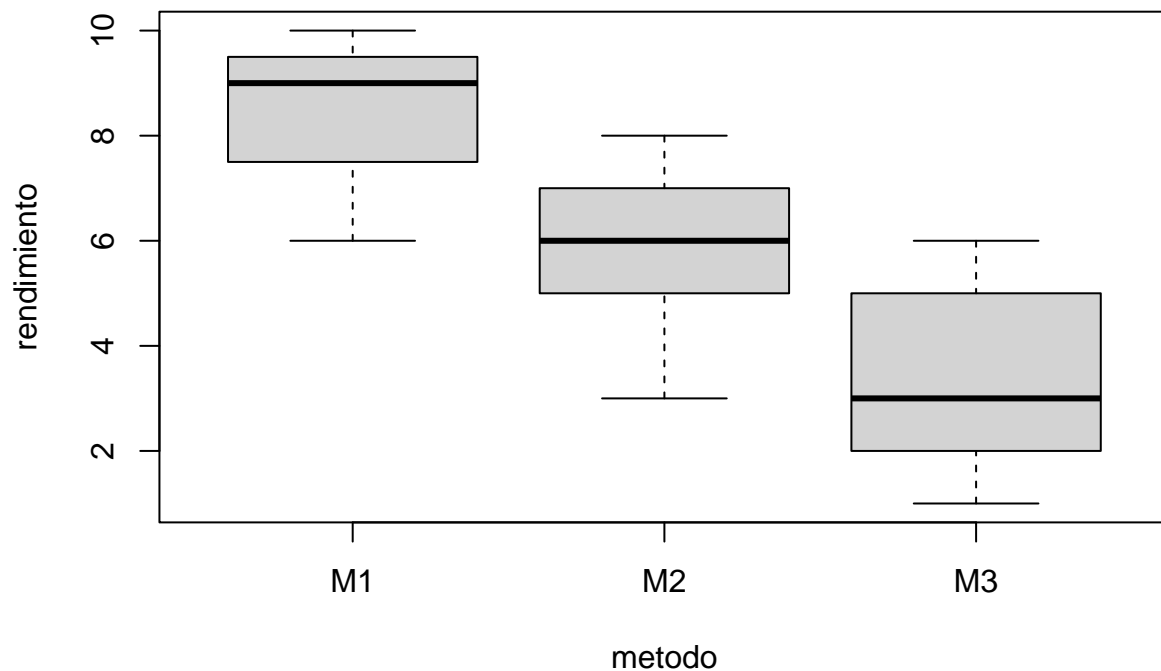
```
tapply(rendimiento, metodo, mean)
```

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

```
mean(rendimiento)
```

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

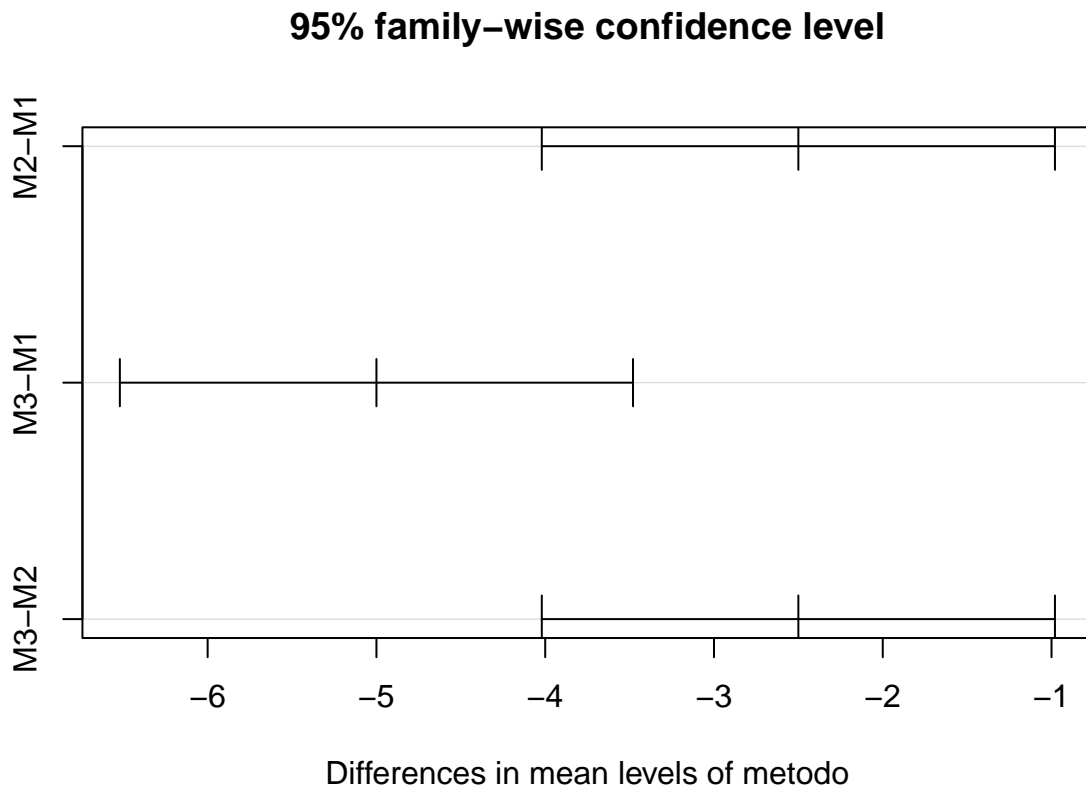
```
boxplot(rendimiento ~ metodo)
```



```
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
```

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



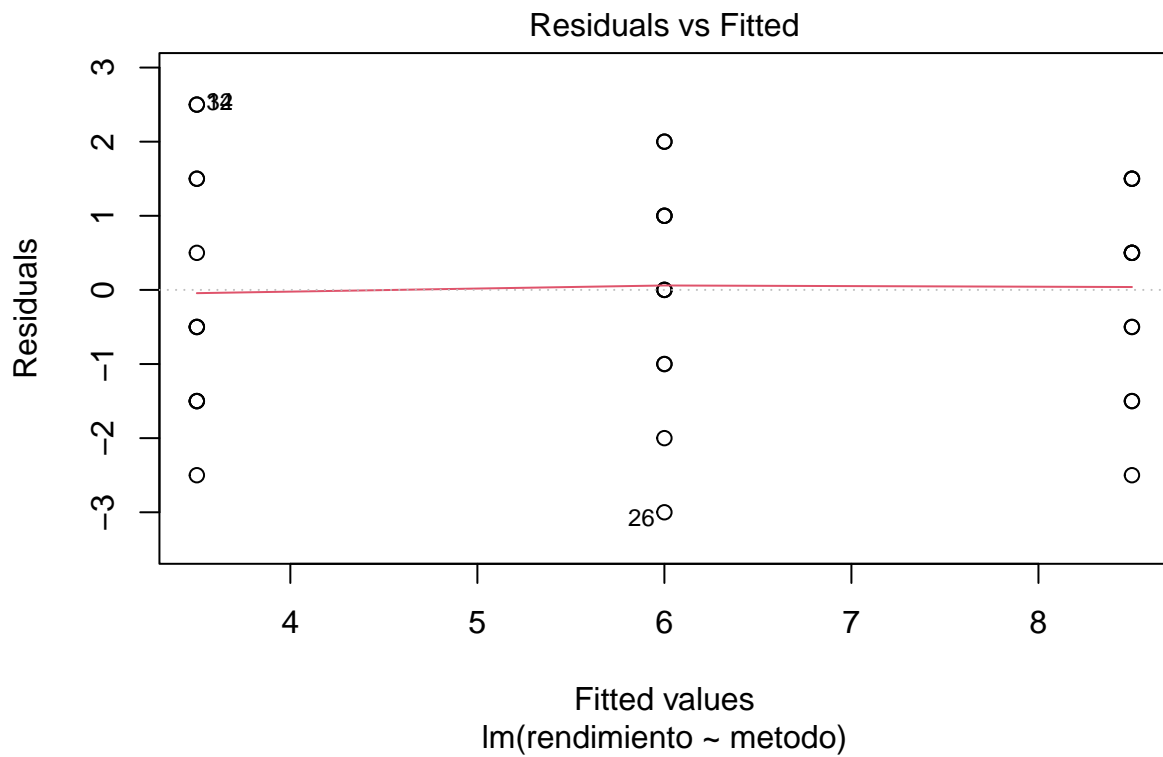
En ANOVA con un solo factor se observa lo siguiente:

- Efectivamente se afirma a través los resultados como el valor P ( $1.55 \times 10^{-8}$ ) que la metodología es el factor que mas juega en el rendimiento de los estudiantes
- En las pruebas de Tukey indica que hay diferencias significativas entre los niveles de la metodología de enseñanza. Específicamente, el método 1 tiene un rendimiento significativamente mayor que los métodos 2 y 3.

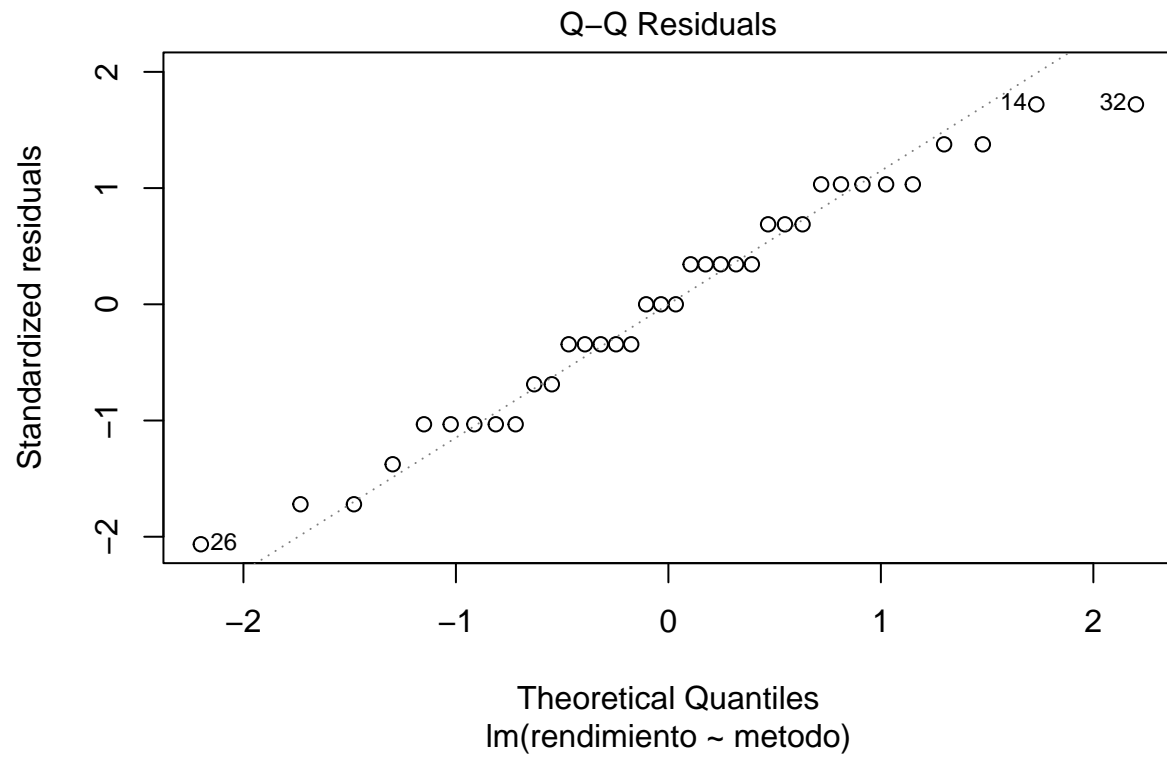
## Análisis del modelo

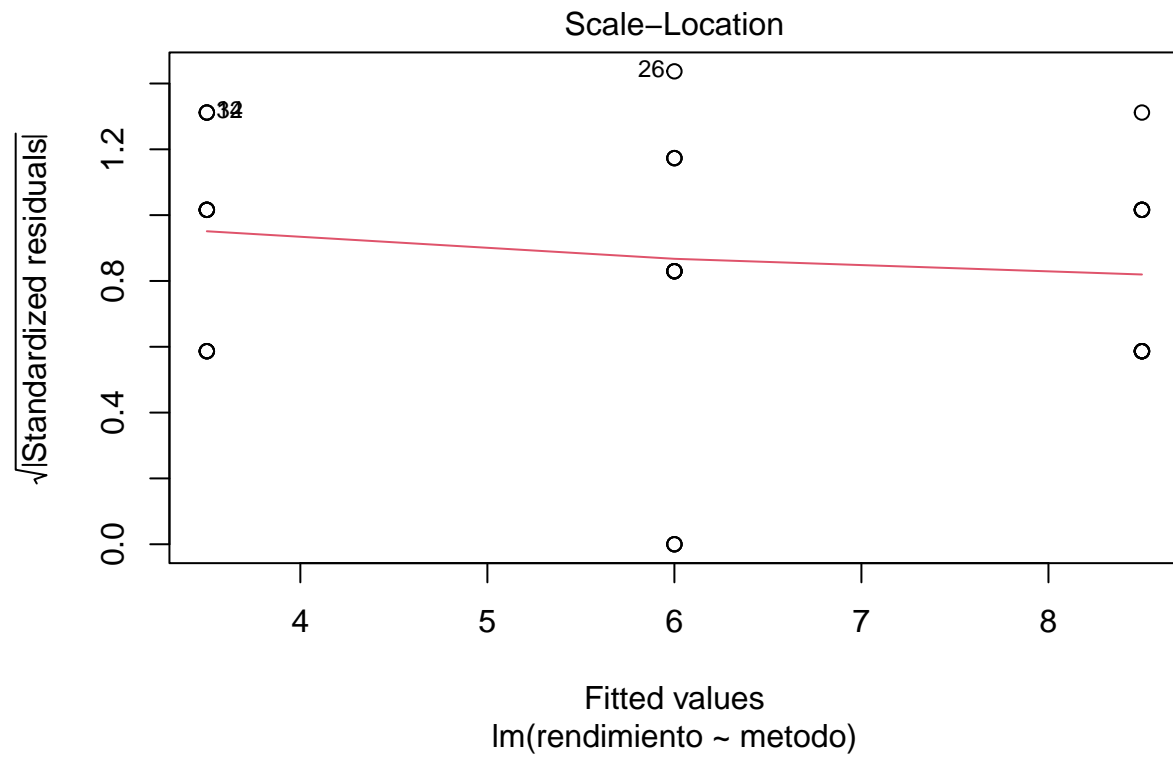
Se verifica la validez del modelo por medio de las gráficas de residuos y la gráfica de normalidad. También se pueden calcular los coeficientes de determinación del modelo para conocer la variación explicada por el modelo.

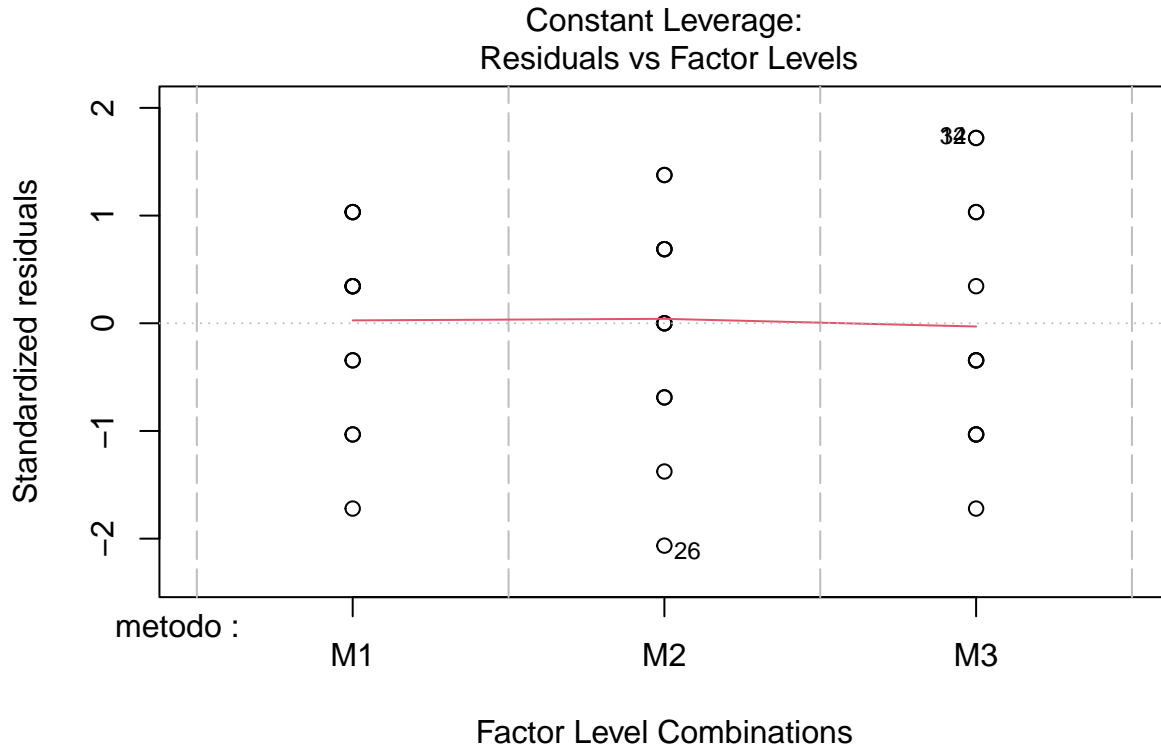
```
plot(lm(rendimiento ~ metodo))
```











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

```
## [1] 0.6637168
```

- Observando el resultado de los análisis se puede llegar a las siguientes conclusiones:
  - Acorde a los resultados de ANOVA si existe una influencia significativa en el rendimiento en cuanto a metodología de enseñanza, pero tanto en genero como en la interacción entre metodología y genero no hay ninguna indicación de que estos afecten al rendimiento
  - Las distintas metodologías tienen efecto en el Rendimiento, siendo el metodo 1 que tiene un mayor rendimiento que el metodo 2 y el metodo 3
  - El coeficiente de determinación nos muestra que el modelo explica la variabilidad del rendimiento en un 66.37%, haciendo que el porcentaje restante pueda ser explicado por otras variables o por la aleatoriedad