Actividad_Retroalimentacion

Iván L. Hernández Buda

2023-08-25

```r
Hm1 <- c(10, 7, 9, 9, 9, 10)
Hm2 <- c(5, 7, 6, 6, 8, 4)
Hm3 <- c(2, 6, 3, 5, 5, 3)
Mm1 <- c(9, 7, 8, 8, 10, 6)
Mm2 <- c(8, 3, 5, 6, 7, 7)
Mm3 <- c(2, 6, 2, 1, 4, 3)
```

Hipótesis

Nuestro modelo es: $Y_{ijk} = \mu + \tau_i + \alpha_j + \tau_i\alpha_j + \varepsilon_{ijk}$

Primera hipótesis:

H_0: τ_i=0 H_1: τ_i!=0

Segunda hipótesis:

H_0: α_j=0 H_1: α_j!=0

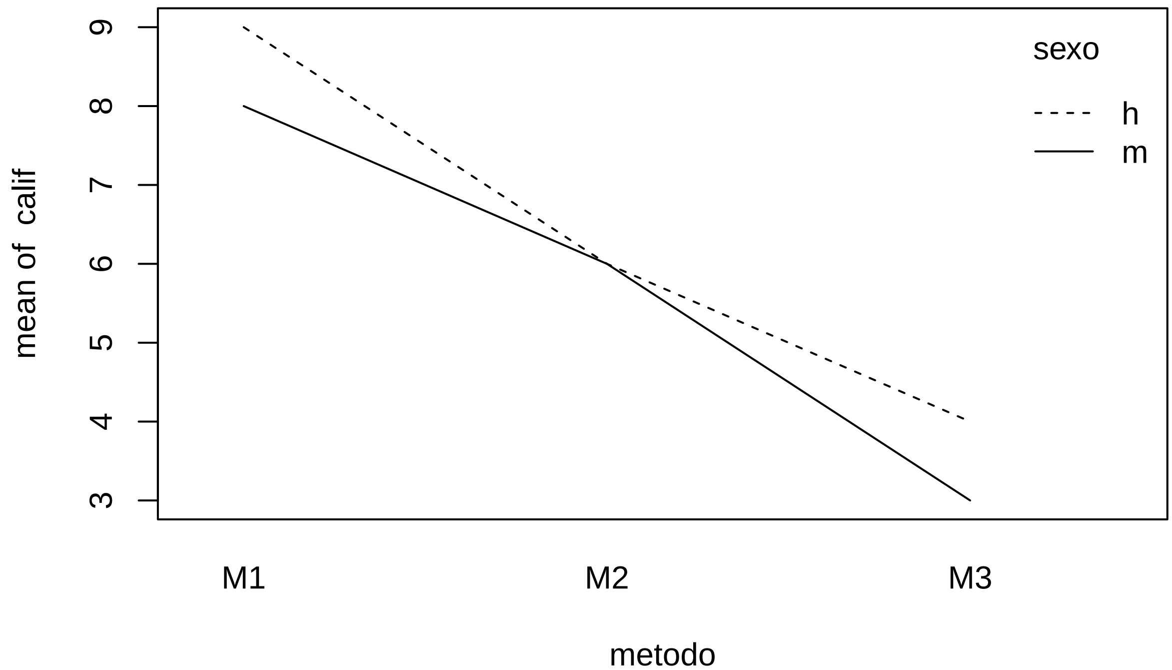
Tercera hipótesis:

H_0: $\tau_i\alpha_j$=0 H_1: $\tau_i\alpha_j$!=0

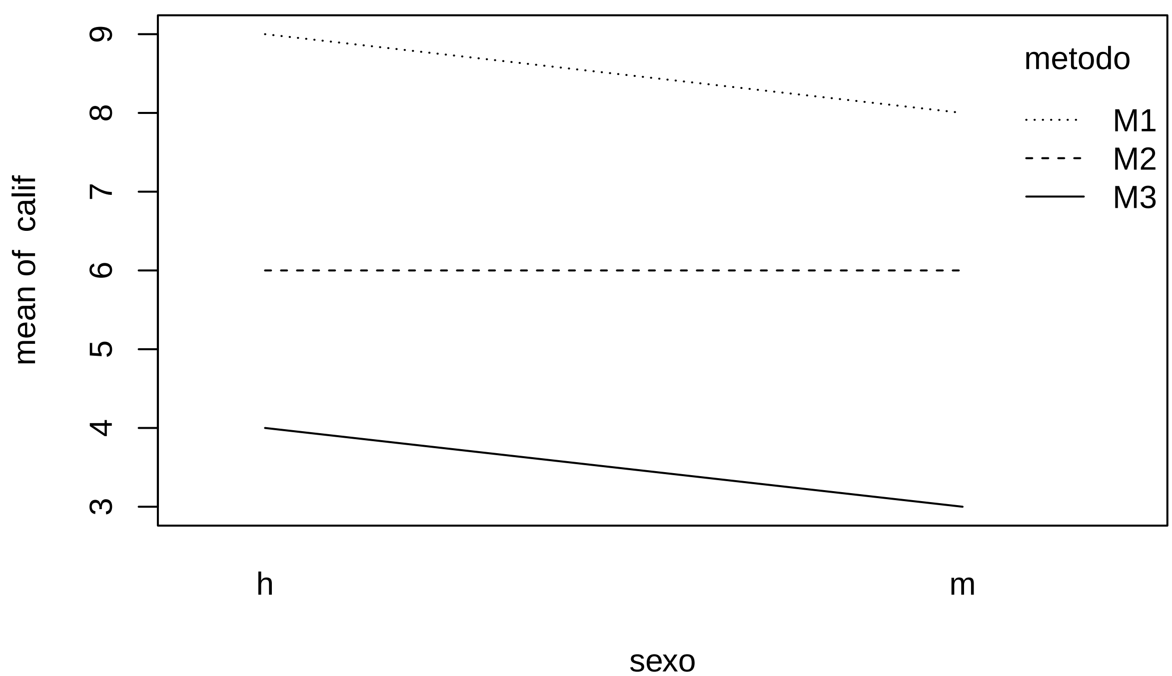
```r
calif = c(Hm1,Hm2,Hm3,Mm1,Mm2,Mm3)
metodo = c(rep("M1",6),rep("M2",6),rep("M3",6),rep("M1",6),rep("M2",6),rep("M3",6))
sexo = c(rep("h", 18), rep("m",18))
metodo = factor(metodo)
sexo = factor(sexo)

A <- aov(calif~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,calif)
```




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



En primera instancia, podemos observar varias cosas. Primeramente, se denota que el valor-p más bajo es el de 'método', el cuál es practicamente cero, lo que corresponde a una alta significancia, mientras que, las demás variables tienen un valor-p demasiado alto, por lo que posteriormente eliminaremos a 'método:sexo' del modelo por ser menos significante de los tres.

En las dos siguientes gráficas se observa el comportamiento que tiene el rendimiento debido a los métodos en cada sexo (Hombre y Mujer). debido a que el rendimiento mejora relativamente igual en ambos sexos, y las líneas no se cruzan, podemos concluir que no hay influencia del método sobre el sexo.

```r
B <- aov(calif~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
```

```r
tapply(calif, sexo, mean)
```

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

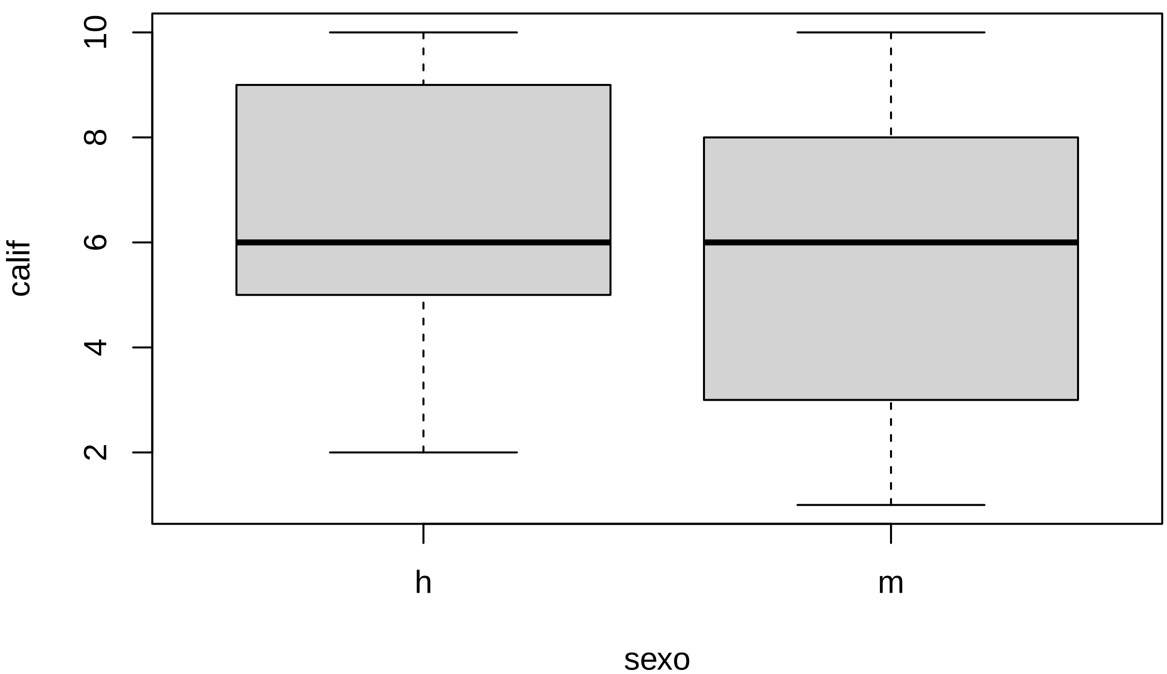
```r
tapply(calif, metodo, mean)
```

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

```r
M = mean(calif)
M
```

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

```r
boxplot(calif ~ sexo)
```



Nuestro modelo se reduce a: $Y_{ijk} = \mu + \tau_i + \alpha_j + \varepsilon_{ijk}$

En esta gráfica de cajas, se pueden comparar el promedio de rendimiento de cada sexo. Se observa que es casi igual para los dos. Sin embargo, es notorio que para las mujeres, los métodos, en general, muestran menos resultados.

También, en sumario, se puede observar que el valor-p de sexo cambió; bajó de 0.2 a 0.192, pero sigue siendo aún insignificante, por lo que se procederá a retirar del modelo. La variable método sigue siendo muy significante por su cercanía a cero.

```r
C <- aov(calif~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
```

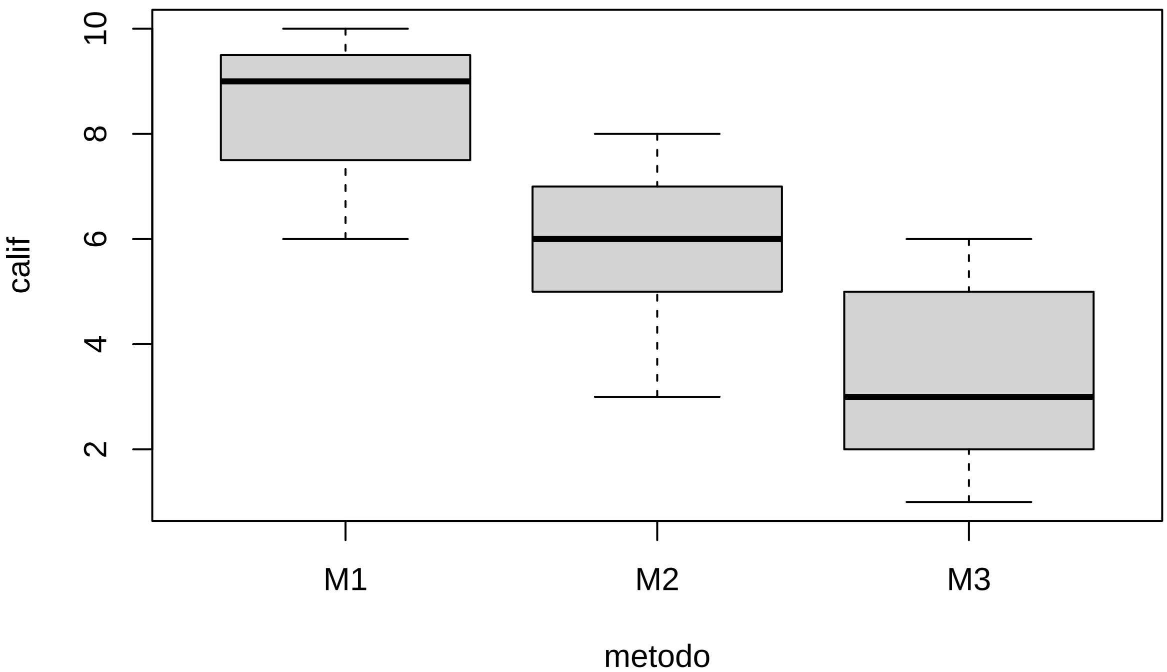
```
tapply(calif, metodo, mean)
```

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

```
mean(calif)
```

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

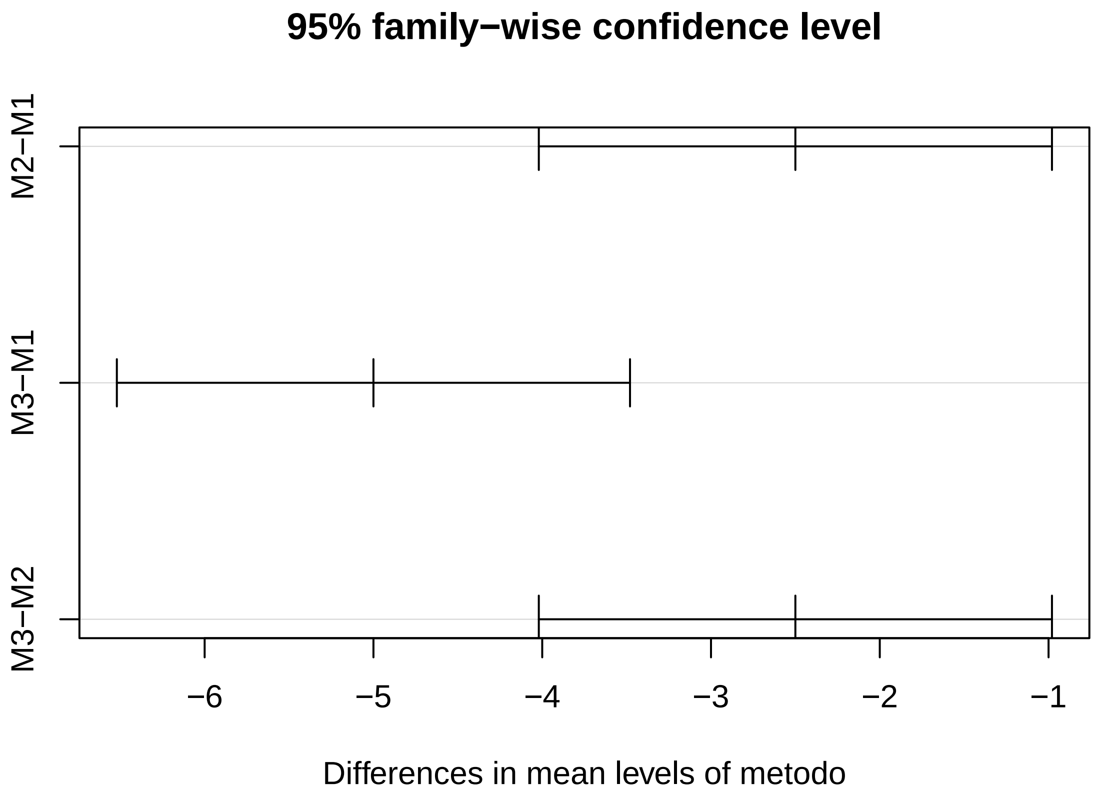
```
boxplot(calif ~ metodo)
```




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

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

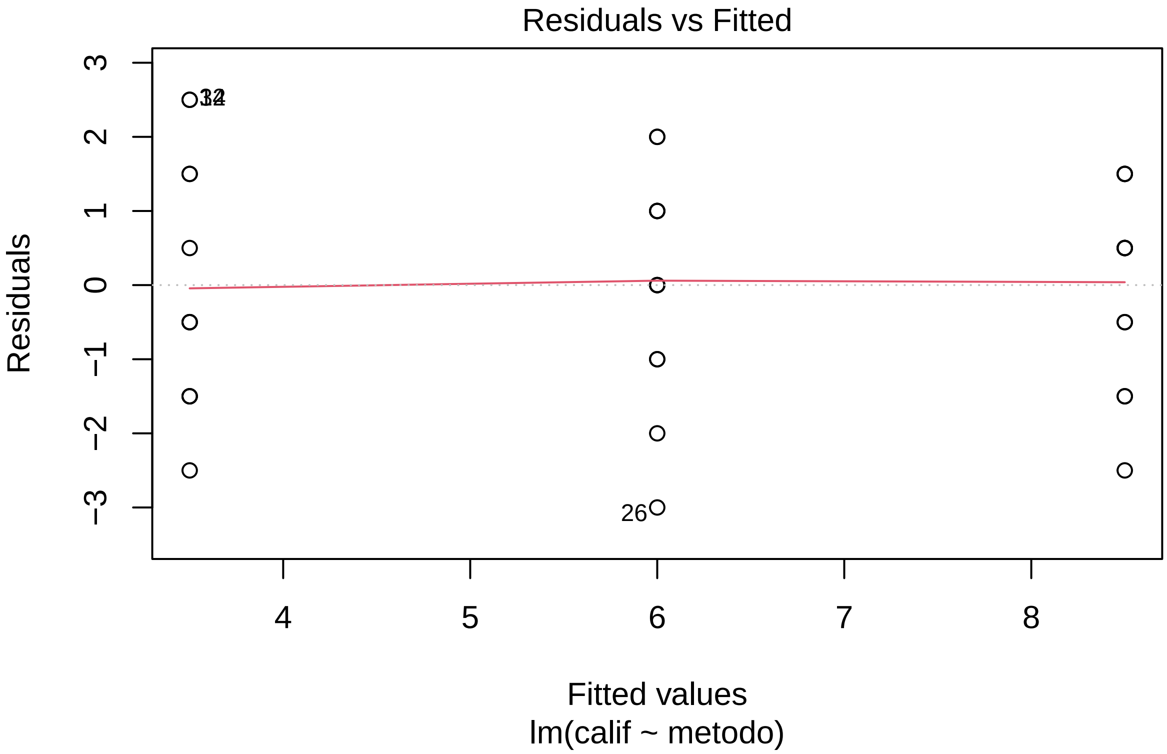


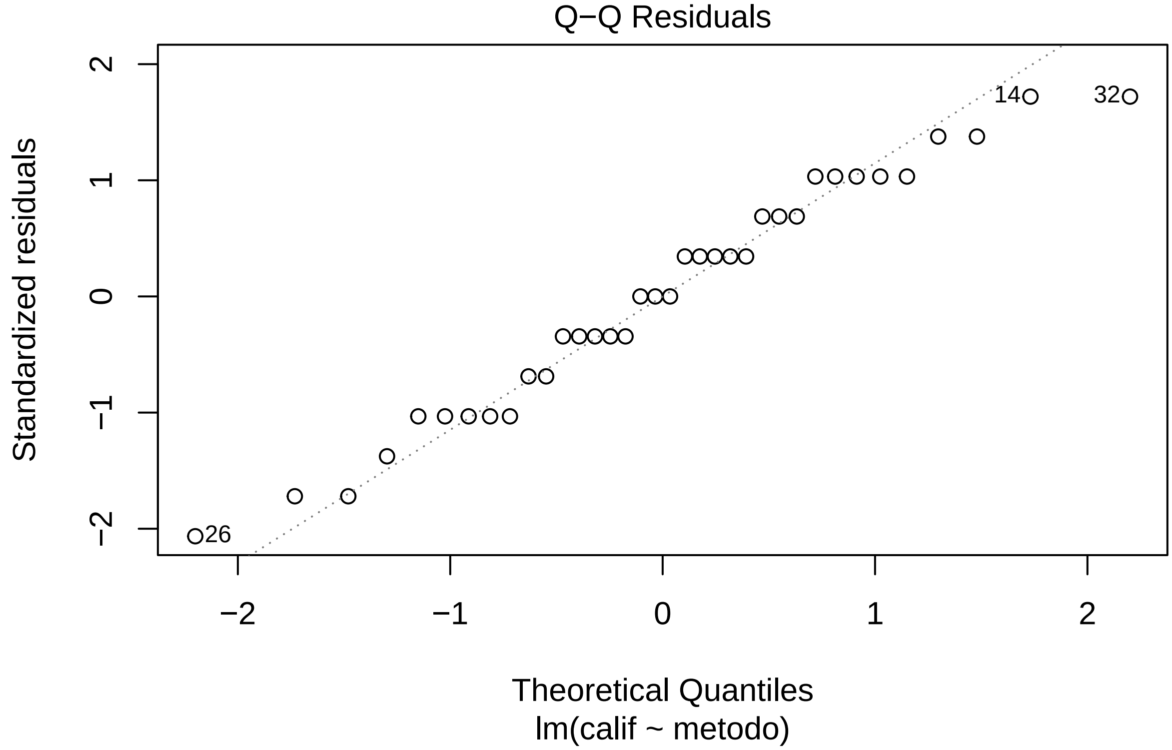
Nuestro modelo se reduce a: $Y_{ijk} = \mu + \tau_i + \varepsilon_{ijk}$

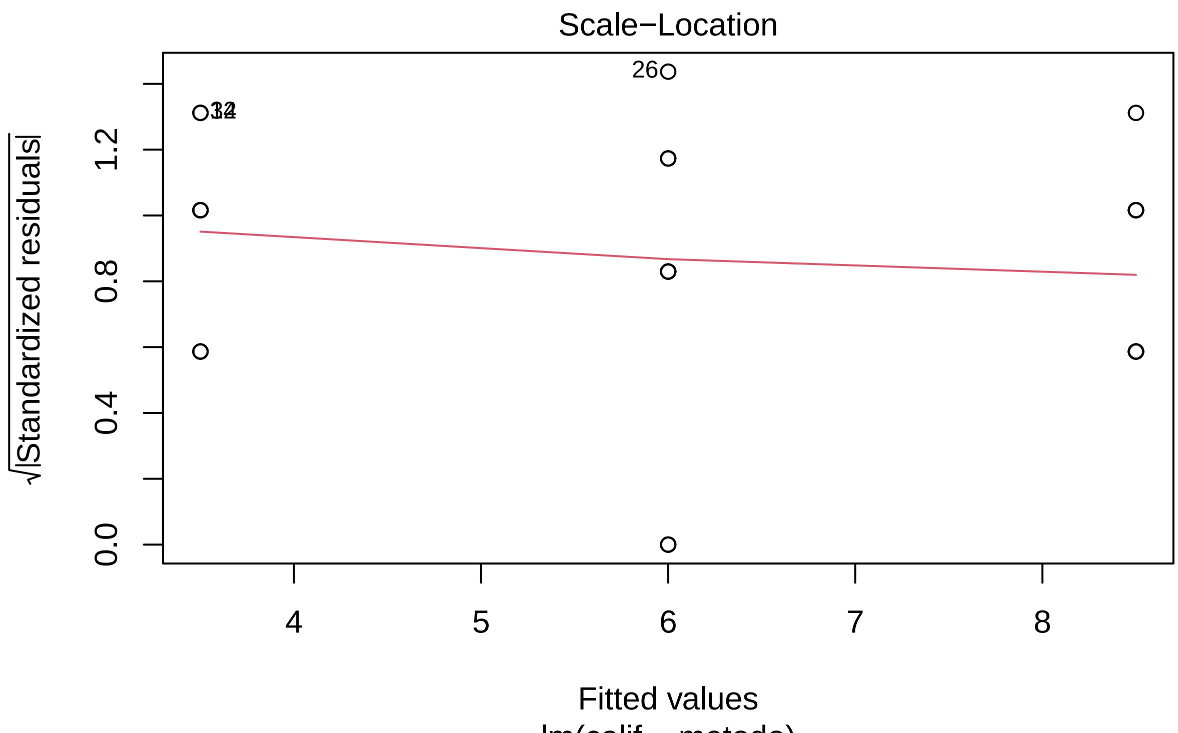
Se retiran ya todas las variables insignificantes del modelo y nos queda únicamente 'método' y, posteriormente, se hace un análisis comparativo de medias de Tukey.

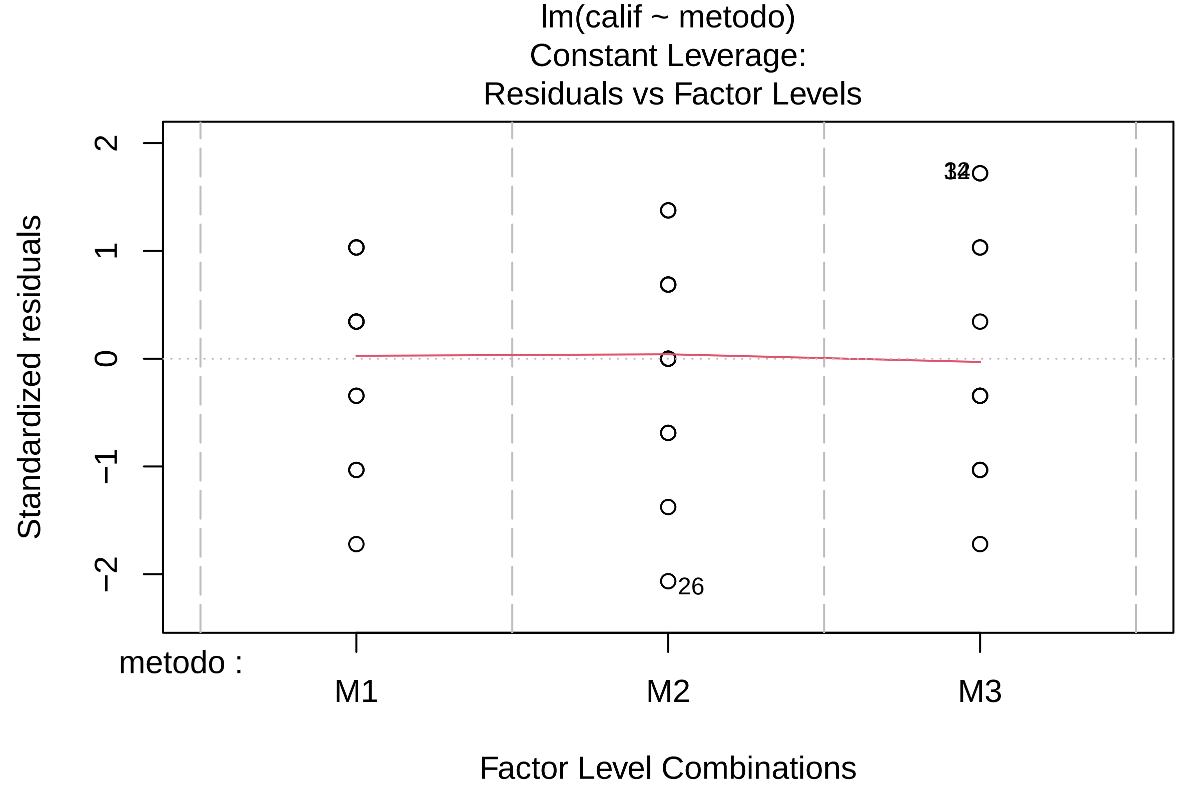
En este caso, en las gráficas anteriores, se aprecian las diferencias de los niveles de media entre cada método. Se puede ver que método 1 es mejor que el método 2, el método 1 es mejor que el método 3 y que el método 2 es mejor que método 3. Sin embargo, también se puede apreciar que los intervalos de las diferencias de la media prevalecen en los negativos, lo que se traduce a que ningún método es igual a otro.

```
plot(lm(calif~metodo))
```


Residuals vs Fitted
Residuals
Fitted values
lm(calif ~ metodo)
32
14
26


Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(calif ~ metodo)
14
32
26






```
SCtrat = 150
residual = 76
SCtotal = SCtrat + residual
CD = SCtrat/SCtotal
CD
```

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

El coeficiente de determinación ‘CD’ dice que tanto explica el modelo; este modelo explica el 66.37% de la variación, es decir, el método de enseñanza es un factor determinante o que influye significativamente el rendimiento de los estudiantes, además porque fue el único factor significativo del modelo. Pero, también hay que tomar en cuenta que el 33.73% restante que el modelo no explica puede deberse a la aleatoriedad intrínseca en el factor humano de la población de estudiantes.

El coeficiente de determinación ‘CD’ dice que tanto explica el modelo; este modelo explica el 66.37% de la variación, es decir, el método de enseñanza es un factor determinante o que influye significativamente el rendimiento de los estudiantes, además porque fue el único factor significativo del modelo. Pero, también hay que tomar en cuenta que el 33.73% restante que el modelo no explica puede deberse a la aleatoriedad intrínseca en el factor humano de la población de estudiantes.