

9. ANOVA

Ricardo Salinas

2024-08-27

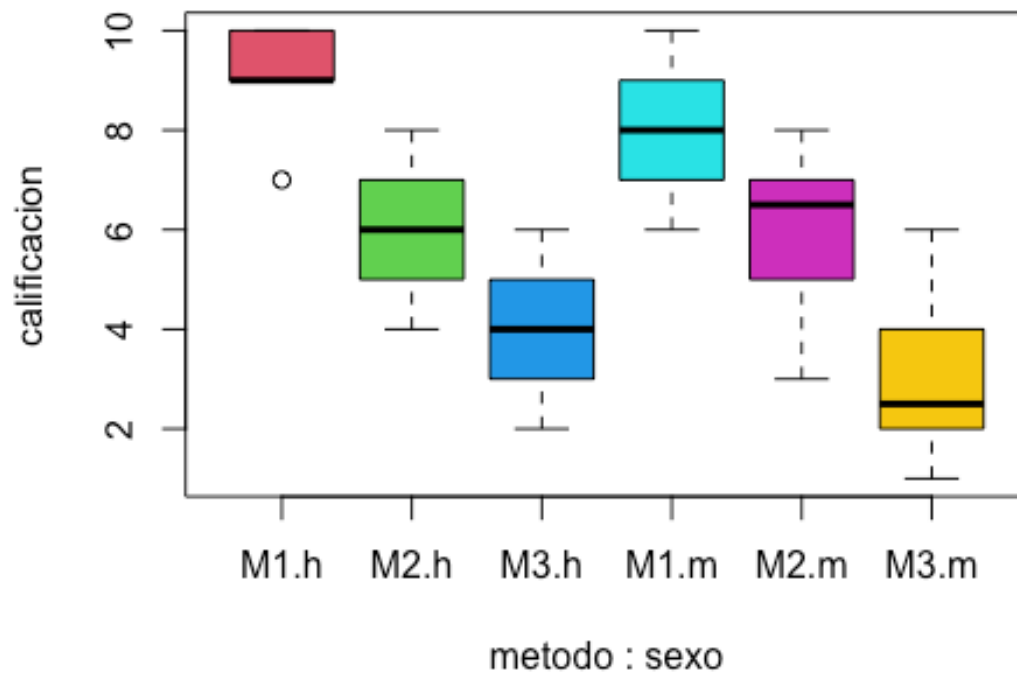
En un instituto se han matriculado 36 estudiantes. Se desea explicar el rendimiento de ciencias naturales en función de dos variables: género y metodología de enseñanza. La metodología de enseñanza se analiza en tres niveles: explicación oral y realización del experimento (1er nivel) explicación oral 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:

```
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)
```

1. Análisis exploratorio. Calcula la media para el rendimiento por método de enseñanza.

#Haz el boxplot de resistencia a la tensión por concentración de madera dura.

```
metodo = factor(metodo)
sexo = factor(sexo)
datos = data.frame(calificacion, metodo, sexo)
boxplot( calificacion ~ metodo:sexo, datos, col=2:8 )
```



#Boxplot de calificacion y metodo, calculando la media del rendimiento por metodo de enseñanza

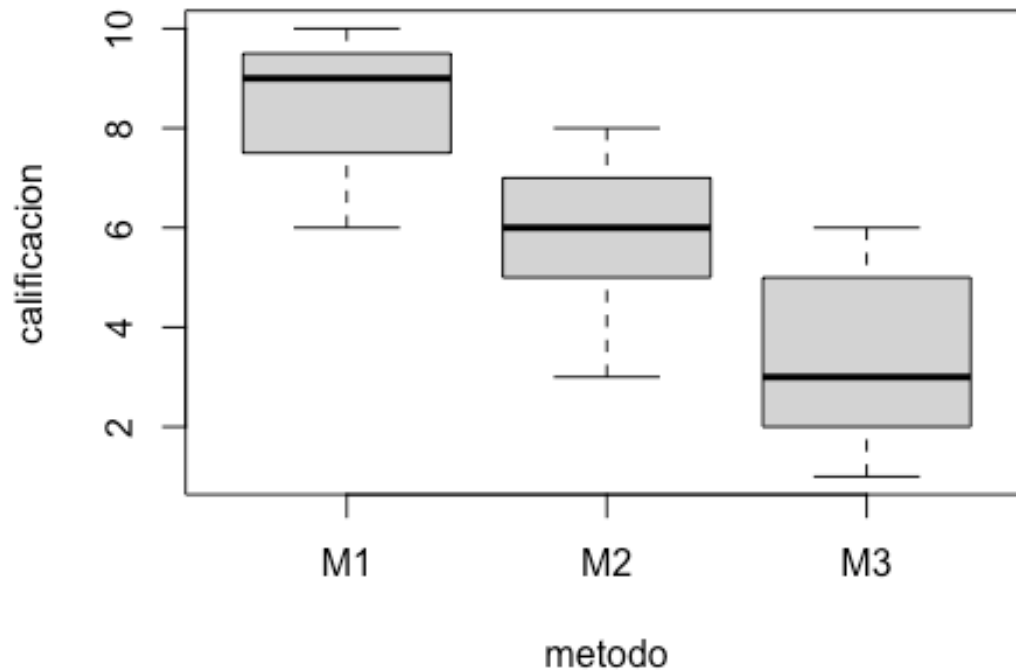
```
tapply(calificacion,metodo,mean)
```

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

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

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

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



#Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.

```
print('Como se puede ver dentro del boxplot de colores, ambos sexos tienen una tendencia de tener un valor mas bajo en el tercer metodo, viendo que la variabilidad de los datos y sus medias estan por debajo de los otros metodos')
```

```
## [1] "Como se puede ver dentro del boxplot de colores, ambos sexos tienen una tendencia de tener un valor mas bajo en el tercer metodo, viendo que la variabilidad de los datos y sus medias estan por debajo de los otros metodos"
```

#Escribe tus conclusiones parciales

```
print('Se puede ir concluyendo que el mejor metodo en general seria el primero, ya que muestra mejores calificaciones en general y dentro de cada sexo.')
```

```
## [1] "Se puede ir concluyendo que el mejor metodo en general seria el primero, ya que muestra mejores calificaciones en general y dentro de cada sexo."
```

2. Las hipótesis. Establece las hipótesis estadísticas (tienen que ser 3).

Primera hipotesis

$$\sum_{i=1}^n \tau_i = 0$$

Segunda hipótesis

$$\sum_{j=1}^n \alpha_j = 0$$

Tercera hipótesis

$$\sum_{i=1}^n \sum_{j=1}^n \tau_i \alpha_j = 0$$

3. Realiza el ANOVA para dos niveles con interacción:

#Consulta el código en R en Los apoyos de clase de "ANOVA":

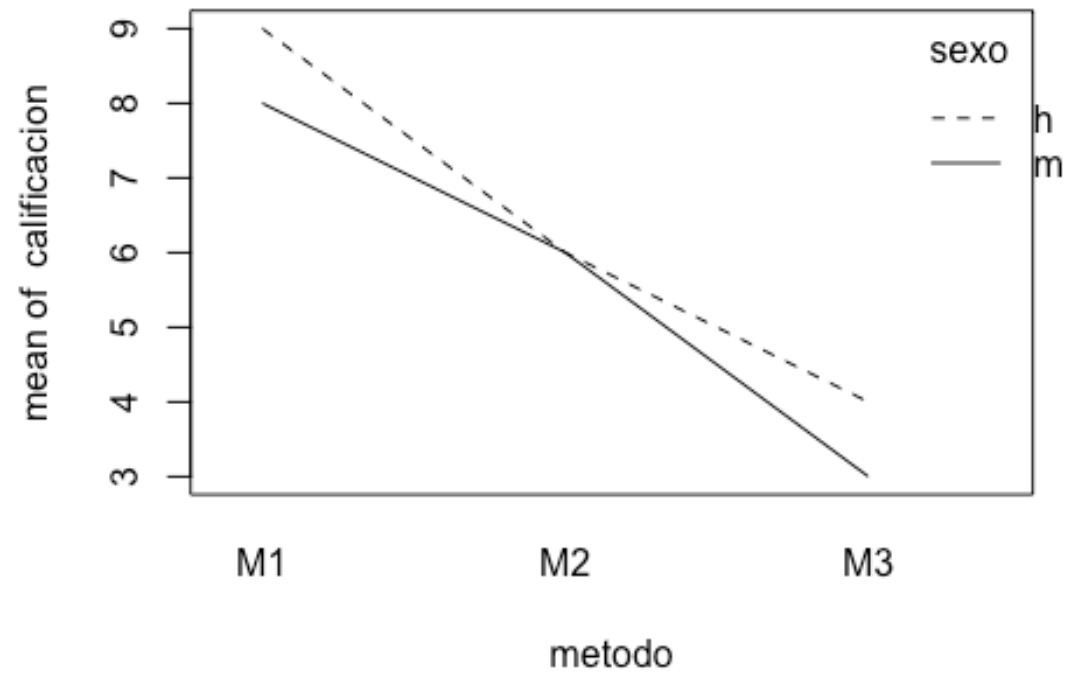
```
anova = aov(calificacion ~ metodo * sexo, datos)
```

```
summary(anova)
```

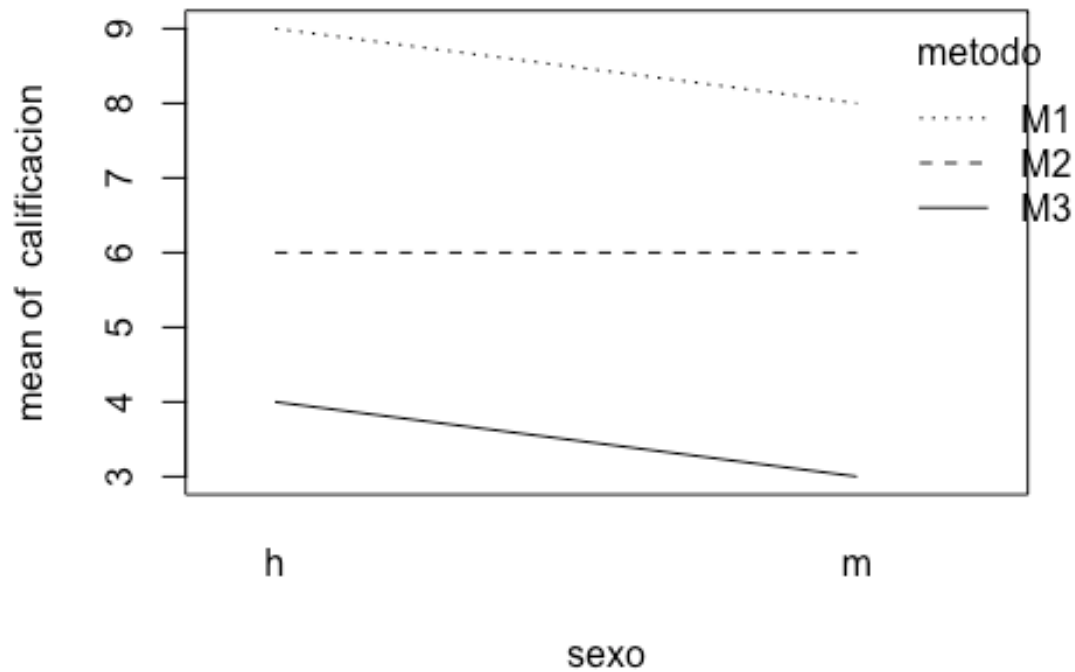
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## metodo         2    150    75.00   32.143 3.47e-08 ***
## sexo           1     4     4.00    1.714   0.200
## metodo:sexo     2     2     1.00    0.429   0.655
## Residuals     30     70     2.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01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

#Haz la gráfica de interacción de dos factores en ANOVA

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



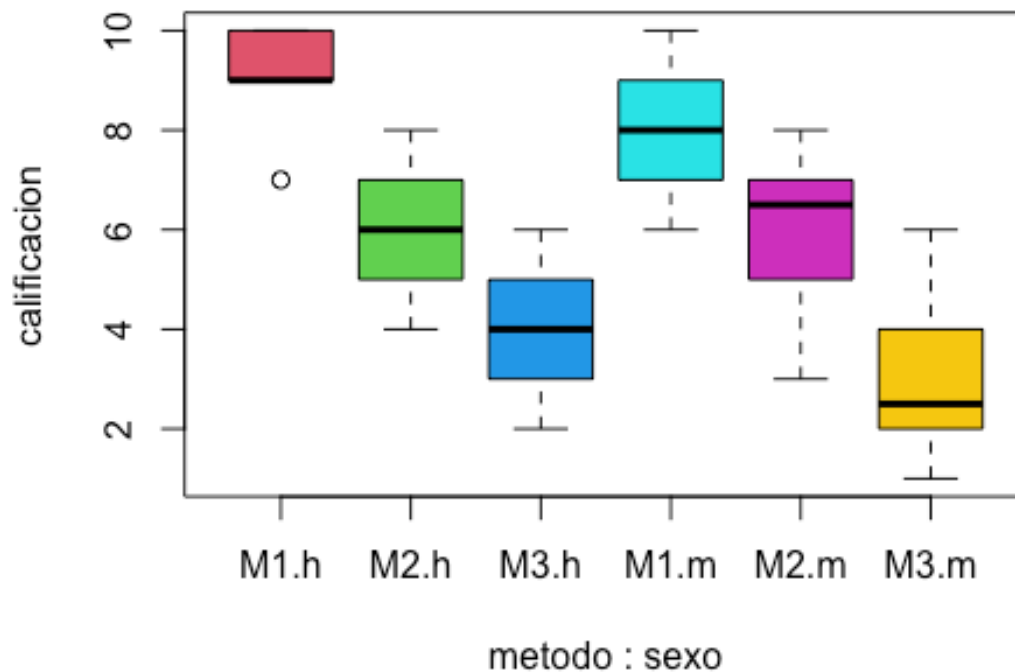
```
interaction.plot(sexo, metodo, calificacion)
```



#Haz el boxplot para visualizar la interacción de los factores, por ejemplo, peso por dieta interacción ejercicio:

*#boxplot(peso ~ dieta * ejercicio, data = data, col = c("lightblue", "lightgreen"), main = "Boxplot de Peso por Dieta y Ejercicio")*

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



#Interpreta el resultado desde la perspectiva estadística y en el contexto del problema

```
print('Tenemos 2 graficas, la primera grafica que representa mean of
calificacion en comparacion con metodo nos muestra como aunque los hombres
llegan a tener una mejor calificacion que las mujeres dentro del metodo 1,
ambos sexos tienen calificacion declinativas en el segundo y tercer metodo,
en el segundo grafico donde se compara con sexo demuestra que los hombres
tuvieron una mejor media de calificacion en el primer y tercer metodo, pero
en el segundo metodo ambos sexos tuvieron la misma media.')
```

```
## [1] "Tenemos 2 graficas, la primera grafica que representa mean of
calificacion en comparacion con metodo nos muestra como aunque los hombres
llegan a tener una mejor calificacion que las mujeres dentro del metodo 1,
ambos sexos tienen calificacion declinativas en el segundo y tercer metodo,
en el segundo grafico donde se compara con sexo demuestra que los hombres
tuvieron una mejor media de calificacion en el primer y tercer metodo, pero
en el segundo metodo ambos sexos tuvieron la misma media."
```

#Escribe tus conclusiones parciales

```
print('Se vuelve a confirmar la hipotesis pasada de que el mejor metodo para
ambos sexos es el primer metodo.')
```

```
## [1] "Se vuelve a confirmar la hipotesis pasada de que el mejor metodo para ambos sexos es el primer metodo."
```

4. Realiza el ANOVA para dos niveles sin interacción.

#Consulta el código de R en Los apoyos de clase de "ANOVA"

#Haz el boxplot de rendimiento por sexo. Calcula la media para el rendimiento por sexo y método.

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

```
##           h           m
```

```
## 6.333333 5.666667
```

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

```
##  M1  M2  M3
```

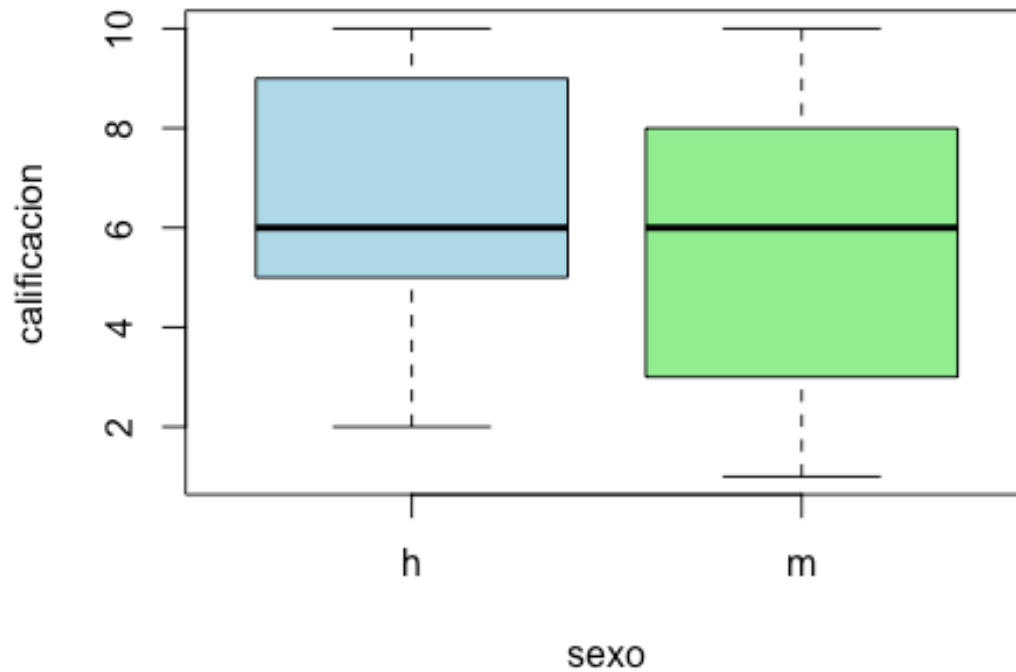
```
## 8.5 6.0 3.5
```

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

```
M
```

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

```
boxplot(calificacion ~ sexo, col = c("lightblue", "lightgreen"))
```

#Haz Los intervalos de confianza de rendimiento por sexo. Grafícalos

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

```
I
```

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

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

```
##
```

```
## Fit: aov(formula = calificacion ~ sexo)
```

```
##
```

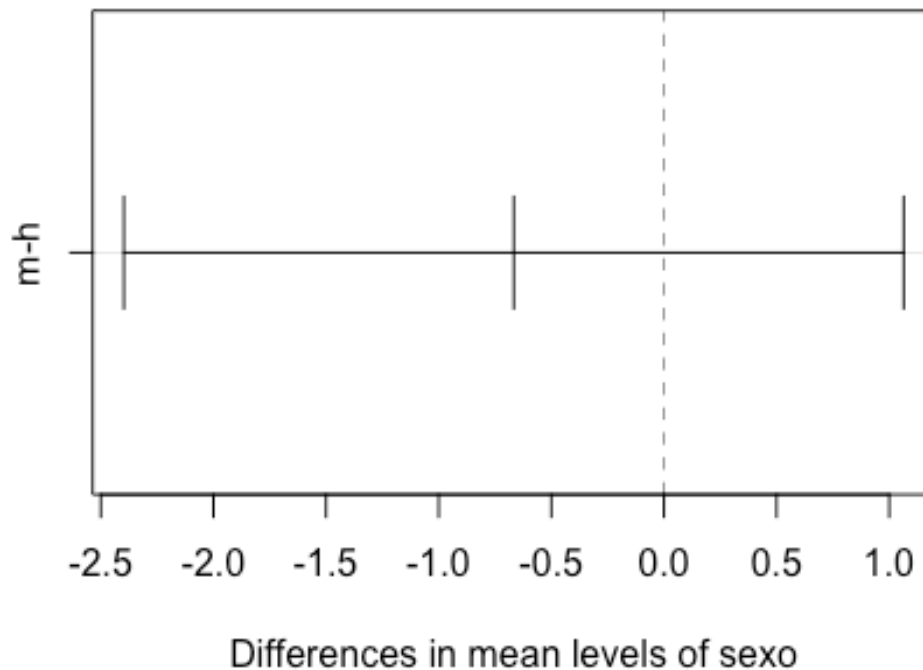
```
## $sexo
```

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

```
## m-h -0.6666667 -2.397645 1.064312 0.4392235
```

```
plot(I)
```

95% family-wise confidence level



```
s = sd(calificacion)
n = length(calificacion)
m = mean(calificacion)
```

```
sm = s/sqrt(n)
E=abs(qt(0.025,n-1))*sm
Infe = m-E
Sup = m+E
```

```
print(Infe)
```

```
## [1] 5.140219
```

```
print(Sup)
```

```
## [1] 6.859781
```

#Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.

```
print('Se puede ver dentro del boxplot que los hombres tienen una mayor media dentro de calificacion en general.')
```

```
## [1] "Se puede ver dentro del boxplot que los hombres tienen una mayor
media dentro de calificacion en general."

#Escribe tus conclusiones parciales

print('Los hombres tienen una myor calificacion que las mujeres tomando en
cuenta todos los metodos')

## [1] "Los hombres tienen una myor calificacion que las mujeres tomando en
cuenta todos los metodos"
```

5. Realiza el ANOVA para un efecto principal

#Consulta el código de R en Los apoyos de clase de "ANOVA"

#Haz los intervalos de confianza de rendimiento por método. Grafícalos

```
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

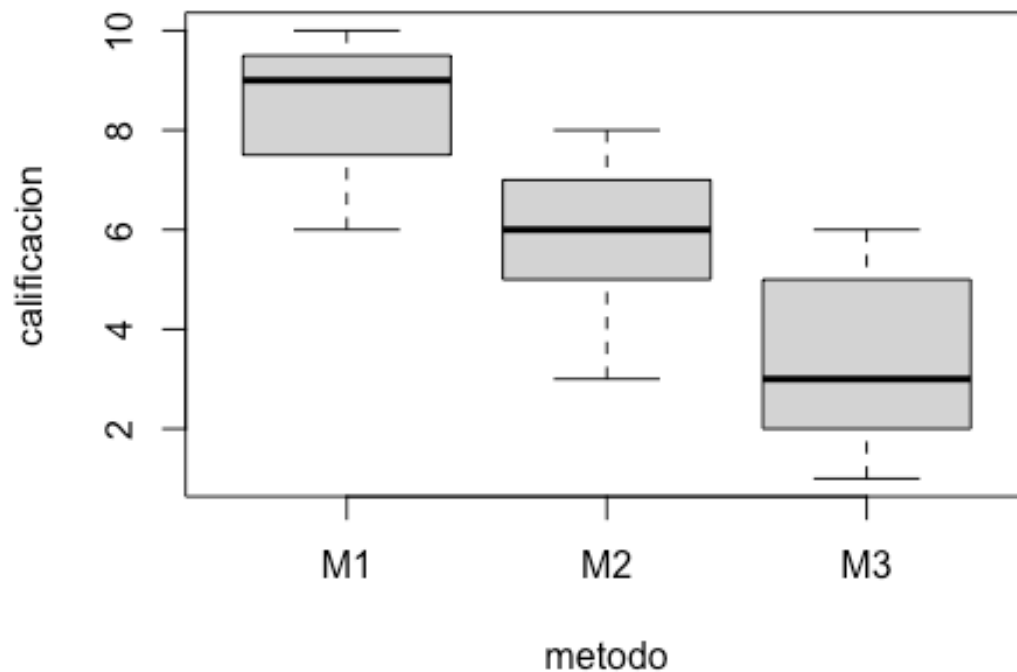
tapply(calificacion,metodo,mean)

##  M1  M2  M3
## 8.5 6.0 3.5

mean(calificacion)

## [1] 6

boxplot(calificacion ~ metodo)
```

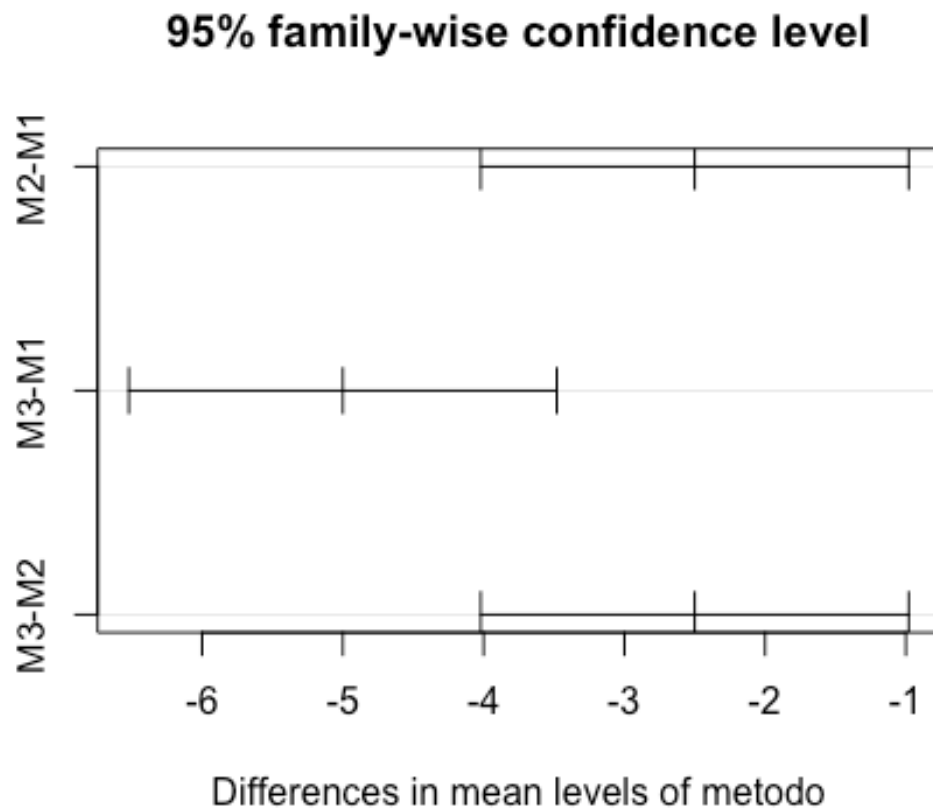


#Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey.

```
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)
```



#Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.

```
print('Ambos sexos tienen una mejor calificacion en el primer metodo, bajando en el segundo metodo y llegando lo mas bajo en el tercero, demostrando que el primer metodo es el mejor.')
```

```
## [1] "Ambos sexos tienen una mejor calificacion en el primer metodo, bajando en el segundo metodo y llegando lo mas bajo en el tercero, demostrando que el primer metodo es el mejor."
```

#Escribe tus conclusiones parciales

```
print('El primer metodo es el mejor metodo para ambos sexos.')
```

```
## [1] "El primer metodo es el mejor metodo para ambos sexos."
```

6. Comprueba la validez del modelo. Comprueba:

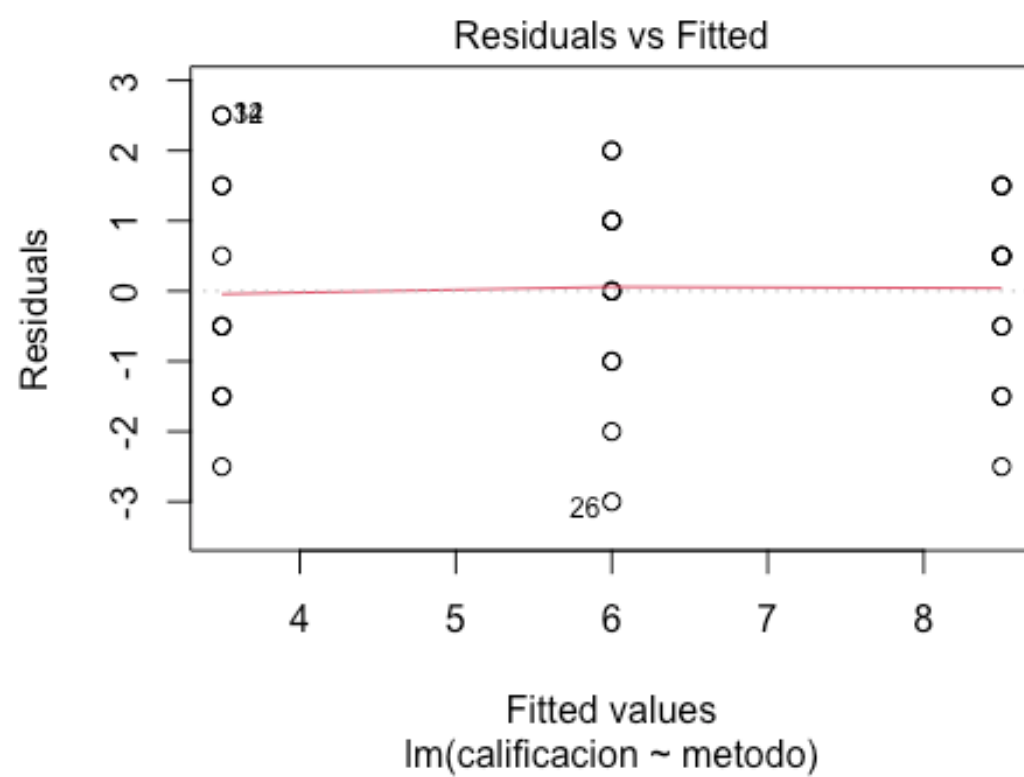
#Normalidad

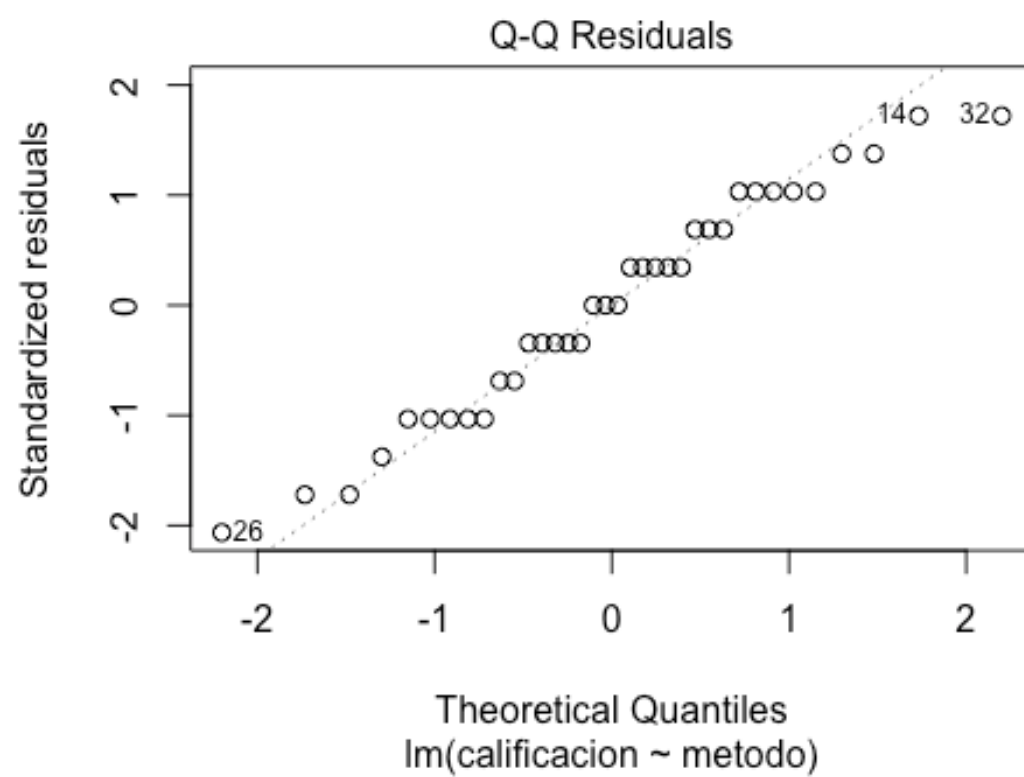
#Homocedasticidad

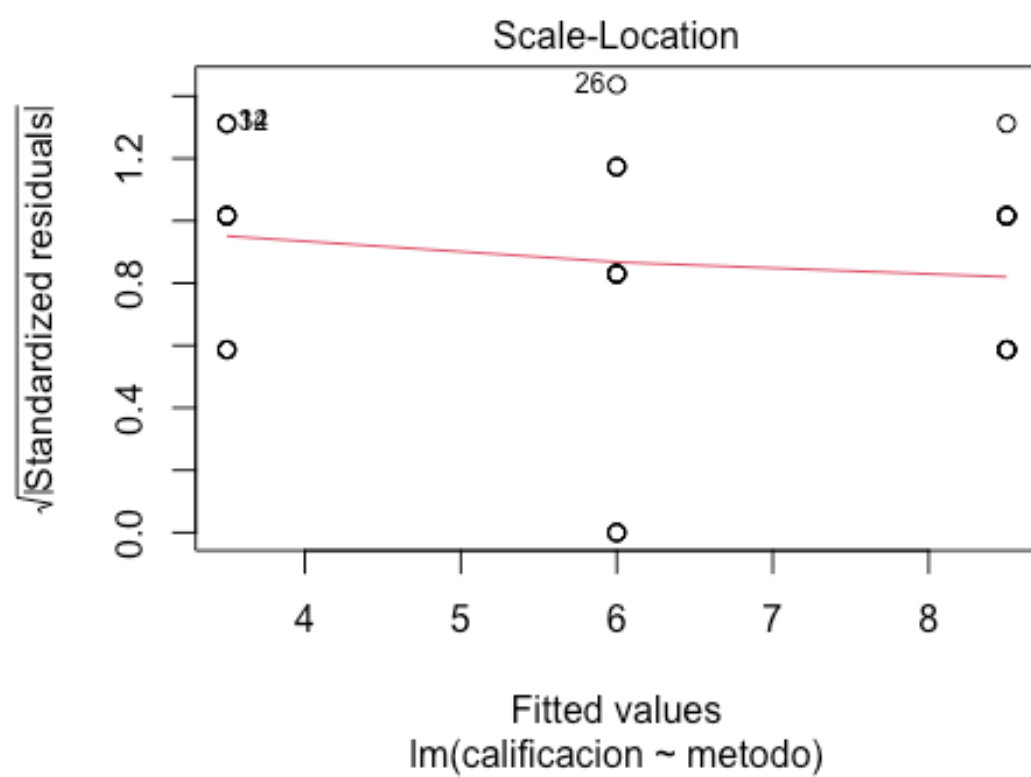
#Independencia

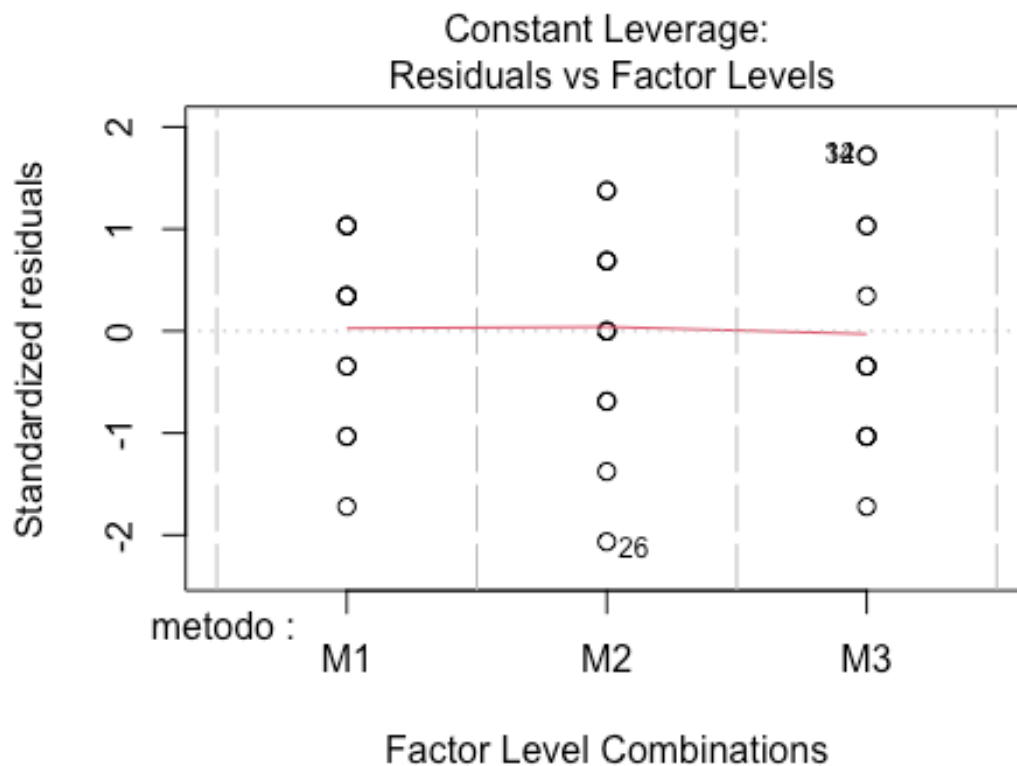
#Relación lineal entre las variables (coeficiente de determinación).

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









CD= $150 / (150 + 76)$

7. Concluye en el contexto del problema.

```
print('Se puede concluir que el mejor metodo de enseñanza es el primero,
siendo que ambos sexos tienen mejores calificaciones en este metodo, tambien
se puede concluir que los hombres tienen mejores calificaciones en general,
ya que se tienen mejores medias de calificacion en el primer y tercer metodo,
aunque en el segundo metodo ambos sexos tienen el mismo mean de
calificacion.')
```

```
## [1] "Se puede concluir que el mejor metodo de enseñanza es el primero,
siendo que ambos sexos tienen mejores calificaciones en este metodo, tambien
se puede concluir que los hombres tienen mejores calificaciones en general,
ya que se tienen mejores medias de calificacion en el primer y tercer metodo,
aunque en el segundo metodo ambos sexos tienen el mismo mean de
calificacion."
```

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.

```

calificacion = 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, 13.7, 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, 13.9, 13.1)

# Crear factor de materiales (cada material se repite 10 veces)
material = factor(rep(c("Acero", "Aluminio", "Plástico"), each = 10))

# Crear factor de proveedores (secuencia de 1 a 5 repetida por columnas)
proveedor = factor(rep(1:5, times = 6))
datos = data.frame(Material = material, Proveedor = proveedor, Calificacion =
calificacion)

# Mostrar la estructura de los datos
print(datos)

```

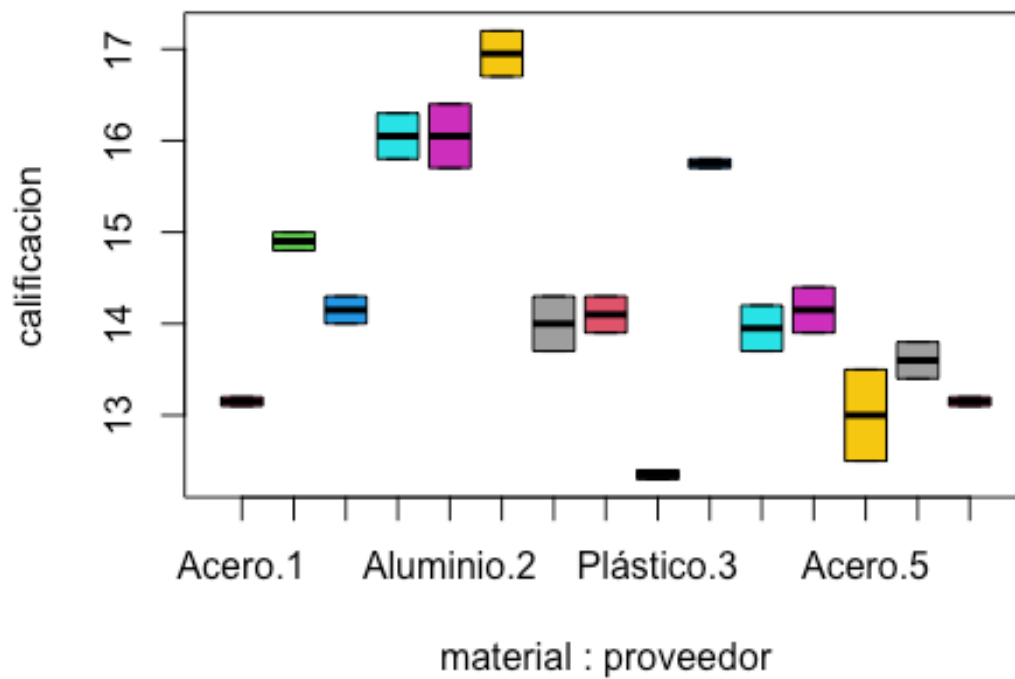
```

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

```

1. Análisis exploratorio. Calcula la media para el rendimiento por método de enseñanza.

```
#Haz el boxplot de resistencia a la tensión por concentración de madera dura.  
datos = data.frame(calificacion, material, proveedor)  
boxplot( calificacion ~ material:proveedor, datos, col=2:8 )
```



```
#Boxplot de calificacion y proveedor, calculando la media del rendimiento por  
metodo de enseñanza
```

```
tapply(calificacion, proveedor, mean)
```

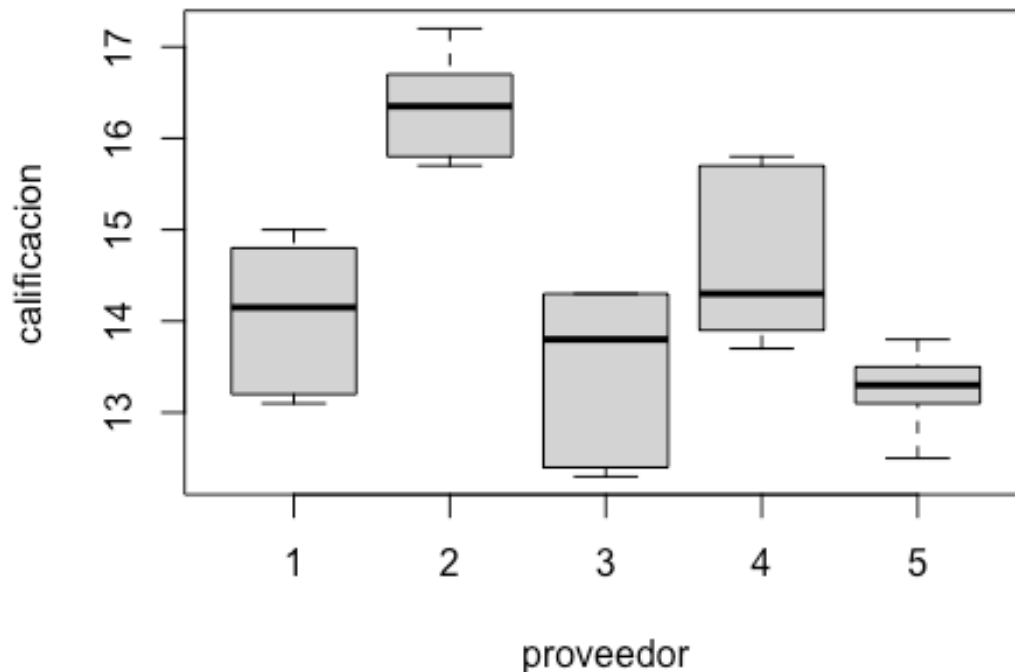
```
##      1      2      3      4      5  
## 14.06667 16.35000 13.48333 14.61667 13.25000
```

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

```
M
```

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

```
boxplot(calificacion ~ proveedor)
```



#Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.

```
print('Dentro del boxplot de proveedor y calificacion podemos ver que el
proveedor con la mejor media de calificacion es el segundo proveedor, despues
el cuarto proveedor, luego el primero, despues el tercero y el proveedor con
menor calificaciones el quinto.')
```

```
## [1] "Dentro del boxplot de proveedor y calificacion podemos ver que el
proveedor con la mejor media de calificacion es el segundo proveedor, despues
el cuarto proveedor, luego el primero, despues el tercero y el proveedor con
menor calificaciones el quinto."
```

#Escribe tus conclusiones parciales

```
print('El segundo proveedor puede ser que sea el que mejor calificacion tenga
en general.')
```

```
## [1] "El segundo proveedor puede ser que sea el que mejor calificacion
tenga en general."
```

2. Las hipótesis. Establece las hipótesis estadísticas (tienen que ser 3).

Primera hipotesis

$$\sum_{i=1}^n \tau_i = 0$$

Segunda hipótesis

$$\sum_{j=1}^n \alpha_j = 0$$

Tercera hipótesis

$$\sum_{i=1}^n \sum_{j=1}^n \tau_i \alpha_j = 0$$

3. Realiza el ANOVA para dos niveles con interacción:

#Consulta el código en R en Los apoyos de clase de "ANOVA":

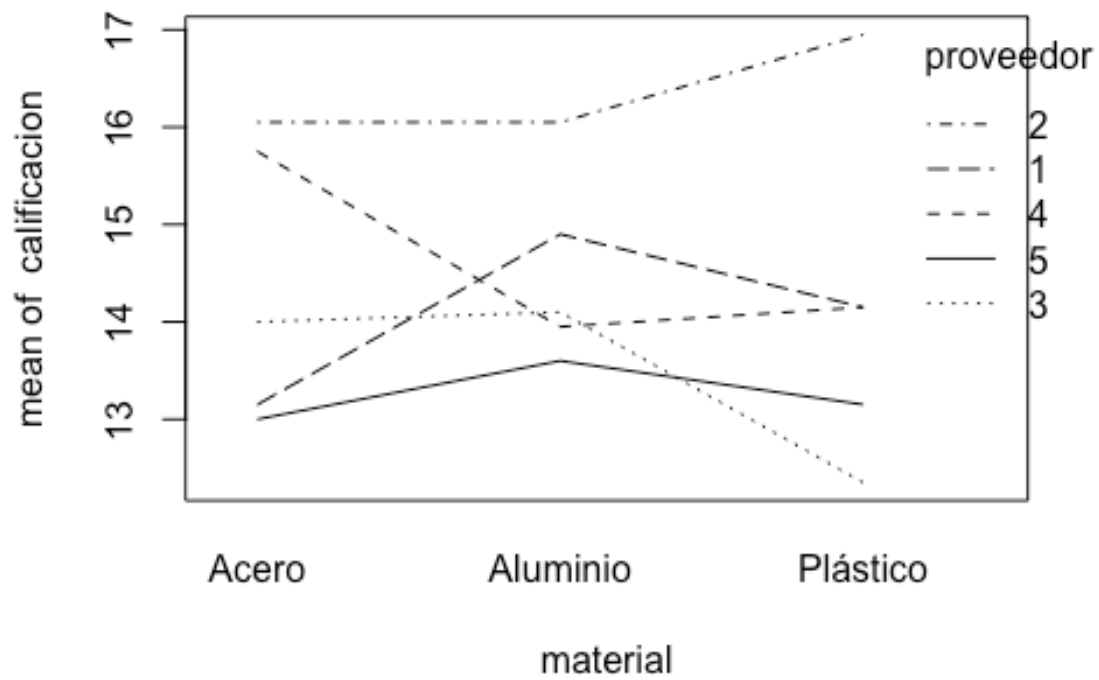
```
anova = aov(calificacion ~ material * proveedor, datos)
```

```
summary(anova)
```

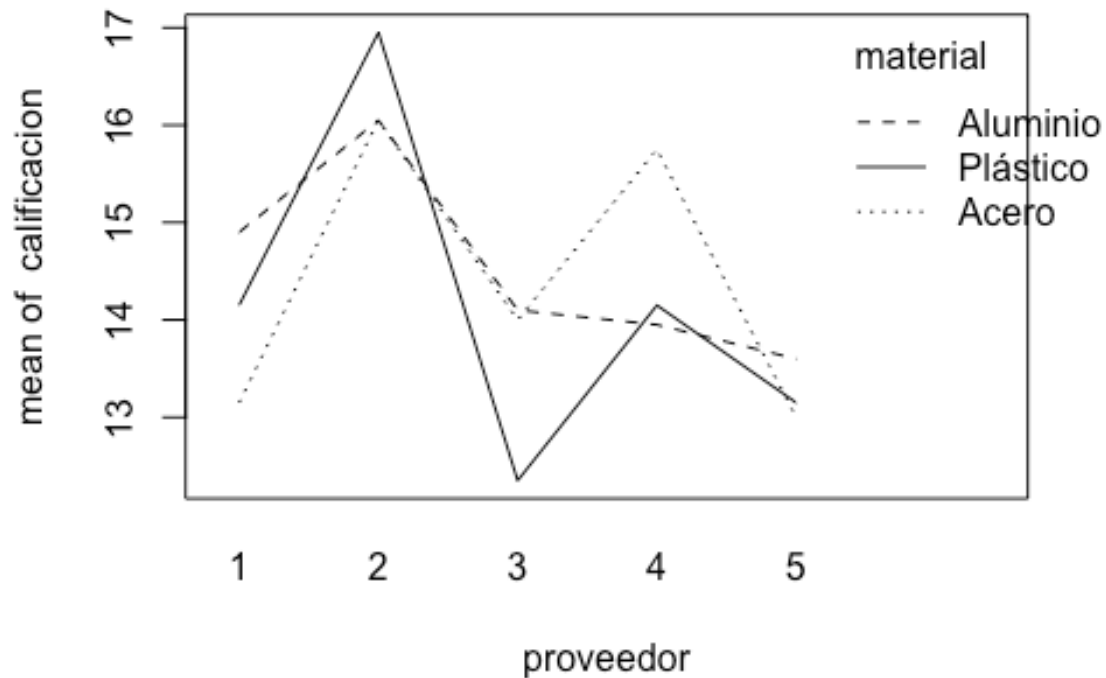
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## material          2    0.70    0.352     3.165    0.0713 .
## proveedor         4   36.67    9.169   82.353 5.07e-10 ***
## material:proveedor 8   11.61    1.451   13.030 1.76e-05 ***
## Residuals       15    1.67    0.111
## ---
## Signif. codes: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

#Haz la gráfica de interacción de dos factores en ANOVA

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



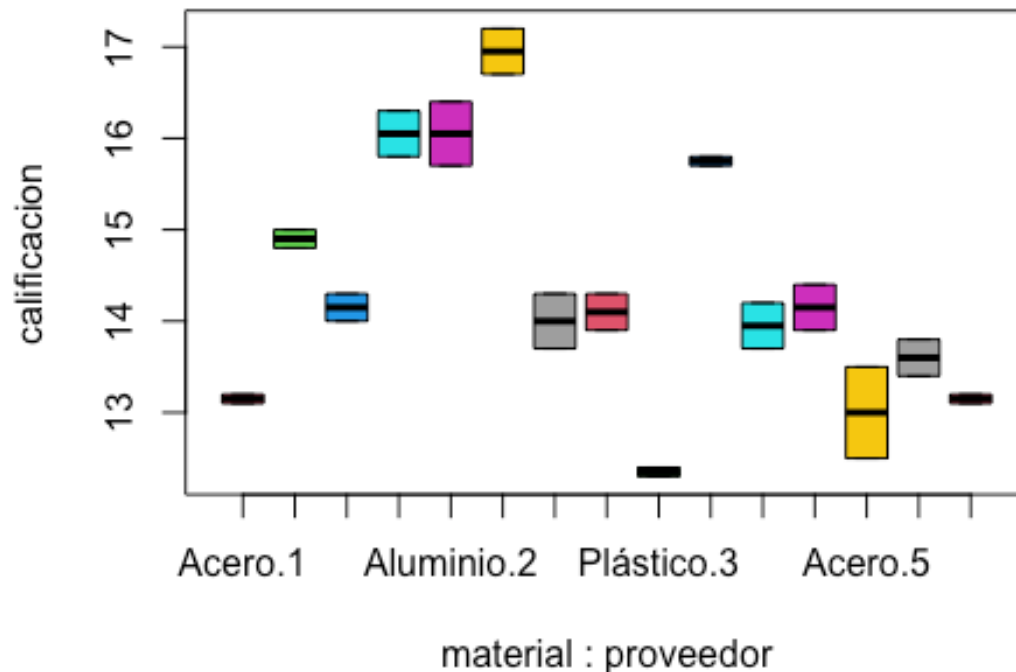
```
interaction.plot(proveedor, material, calificacion)
```



#Haz el boxplot para visualizar la interacción de los factores, por ejemplo, peso por dieta interacción ejercicio:

*#boxplot(peso ~ dieta * ejercicio, data = data, col = c("lightblue", "lightgreen"), main = "Boxplot de Peso por Dieta y Ejercicio")*

`boxplot(calificacion ~ material:proveedor, datos, col=2:8)`



#Interpreta el resultado desde la perspectiva estadística y en el contexto del problema

```
print('El segundo proveedor muestra mejores calificaciones por un gran valor que los demas, siendo que con plastico tiene una gran ventaja, tambien vemos que el material con mejor calificacion en general es el plastico de proveedor 2, teniendo gran varianza de calificacion con los demas proveedores.')
```

```
## [1] "El segundo proveedor muestra mejores calificaciones por un gran valor que los demas, siendo que con plastico tiene una gran ventaja, tambien vemos que el material con mejor calificacion en general es el plastico de proveedor 2, teniendo gran varianza de calificacion con los demas proveedores."
```

#Escribe tus conclusiones parciales

```
print('El segundo proveedor sigue demostrando mejores calificaciones que los demas.')
```

```
## [1] "El segundo proveedor sigue demostrando mejores calificaciones que los demas."
```

4. Realiza el ANOVA para dos niveles sin interacción.

#Haz el boxplot de rendimiento por sexo. Calcula la media para el rendimiento por sexo y método.

```
tapply(calificacion,proveedor,mean)
```



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

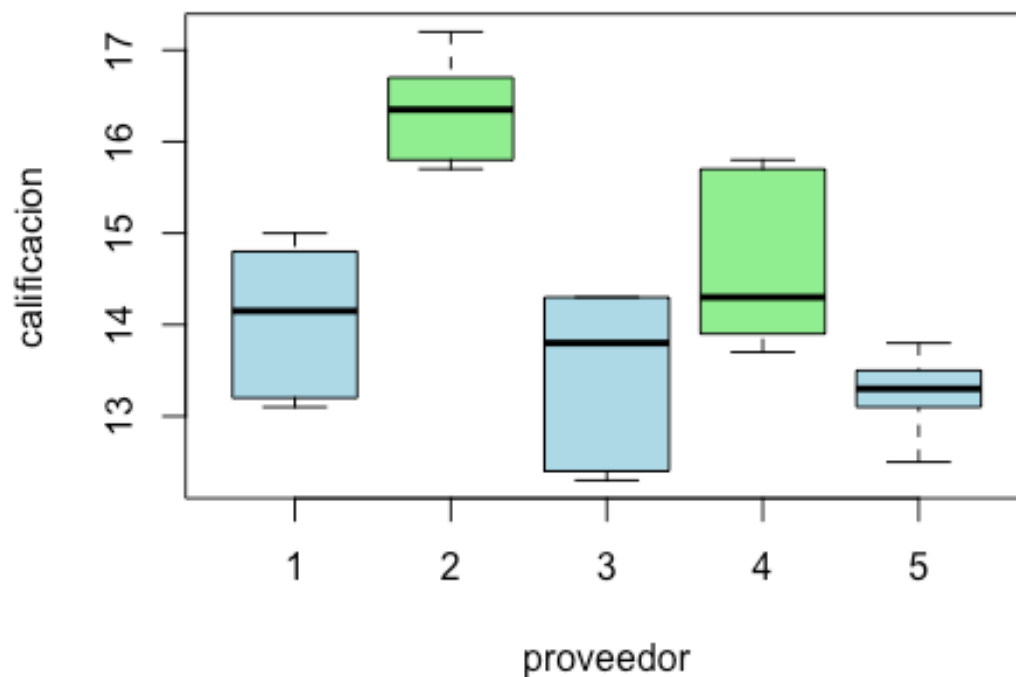
tapply(calificacion,material,mean)

##   Acero Aluminio Plástico
##   14.39   14.52   14.15

M=mean(calificacion)
M

## [1] 14.35333

boxplot(calificacion ~ proveedor, col = c("lightblue", "lightgreen"))
```

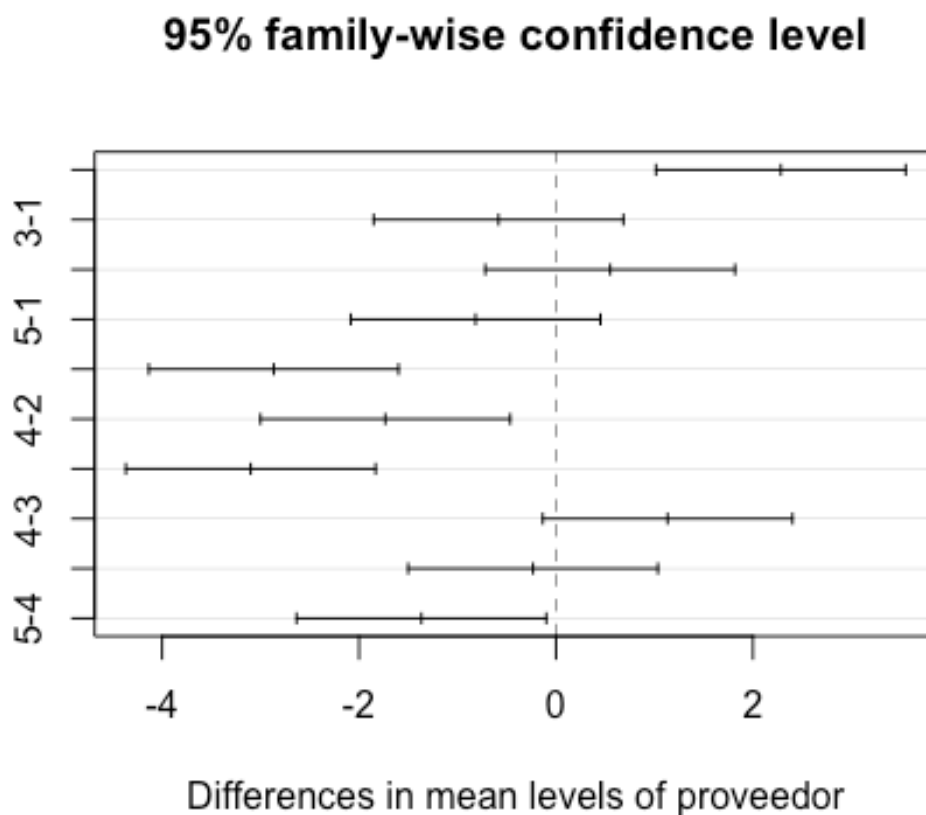


```
#Haz los intervalos de confianza de rendimiento por sexo. Grafícalos
I = TukeyHSD(aov(calificacion ~ proveedor))
I

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

```
##           diff      lwr      upr    p adj
## 2-1  2.2833333  1.0153666  3.5513006 0.0001595
## 3-1 -0.5833333 -1.8513001  0.6846339 0.6630108
## 4-1  0.5500000 -0.7179667  1.8179667 0.7089904
## 5-1 -0.8166667 -2.0846334  0.4513006 0.3474956
## 3-2 -2.8666667 -4.1346334 -1.5986994 0.0000055
## 4-2 -1.7333333 -3.0013001 -0.4653661 0.0039774
## 5-2 -3.1000000 -4.3679667 -1.8320328 0.0000015
## 4-3  1.1333333 -0.1346334  2.4013006 0.0959316
## 5-3 -0.2333333 -1.5013001  1.0346339 0.9821261
## 5-4 -1.3666667 -2.6346334 -0.0986994 0.0301318
```

```
plot(I)
```



#Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.

```
print('El segundo proveedor tiene una mayor media que los demas proveedores en cuano a calificacion, se demuestra como este tiene superioridad.')
```

```
## [1] "El segundo proveedor tiene una mayor media que los demas proveedores en cuano a calificacion, se demuestra como este tiene superioridad."
```

#Escribe tus conclusiones parciales

```
print('Se puede notas como se sigue comprobando que el segundo proveedor es mejor que los demas.')
```

```
## [1] "Se puede notas como se sigue comprobando que el segundo proveedor es mejor que los demas."
```

5. Realiza el ANOVA para un efecto principal

#Haz los intervalos de confianza de rendimiento por método. Grafícalos

```
C = aov(calificacion ~ material)
```

```
summary(C)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## material    2   0.70   0.3523    0.19  0.828
## Residuals   27  49.95   1.8500
```

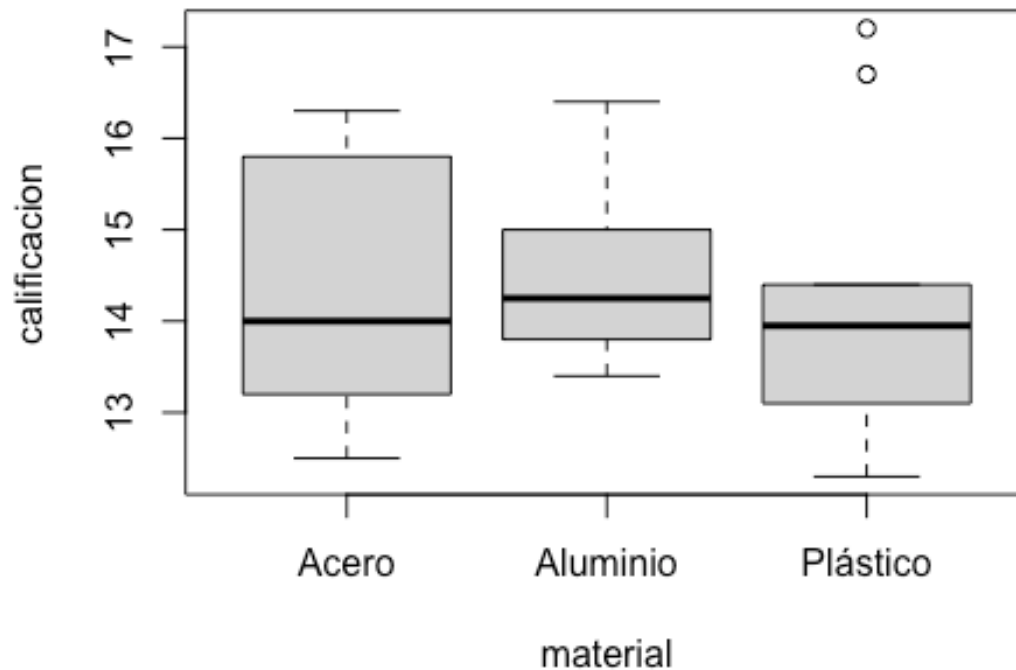
```
tapply(calificacion,material,mean)
```

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

```
mean(calificacion)
```

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

```
boxplot(calificacion ~ material)
```



#Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey.

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

```
I
```

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

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

```
##
```

```
## Fit: aov(formula = calificacion ~ material)
```

```
##
```

```
## $material
```

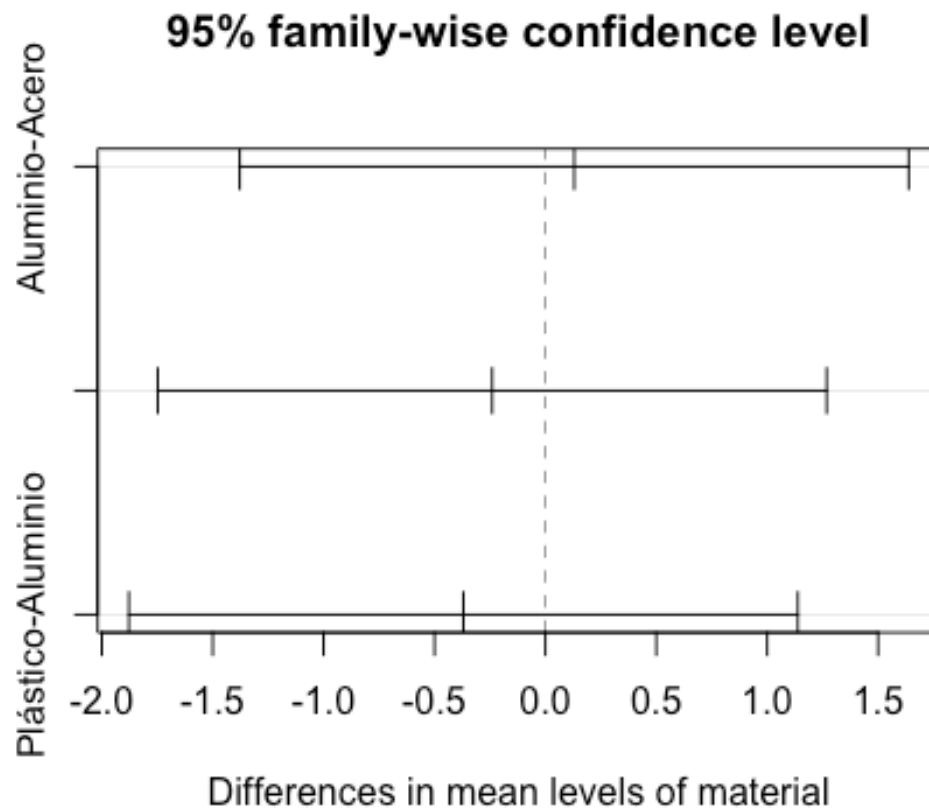
```
##          diff          lwr          upr          p adj
```

```
## 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)
```



#Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.

```
print('Podemos ver como el acero tiene las calificaciones mas altas, pero la
media de aluminio es mayor, y la media de plastivo llega a ser igual a la del
acero, pero ambas tienen menor varianza y valores maximos menores al acero.')
```

```
## [1] "Podemos ver como el acero tiene las calificaciones mas altas, pero la
media de aluminio es mayor, y la media de plastivo llega a ser igual a la del
acero, pero ambas tienen menor varianza y valores maximos menores al acero."
```

#Escribe tus conclusiones parciales

```
print('El acero tiene valores mas altos que los demas materiales, pero la
media mas alta es del aluminio.')
```

```
## [1] "El acero tiene valores mas altos que los demas materiales, pero la
media mas alta es del aluminio."
```

6. Comprueba la validez del modelo. Comprueba:

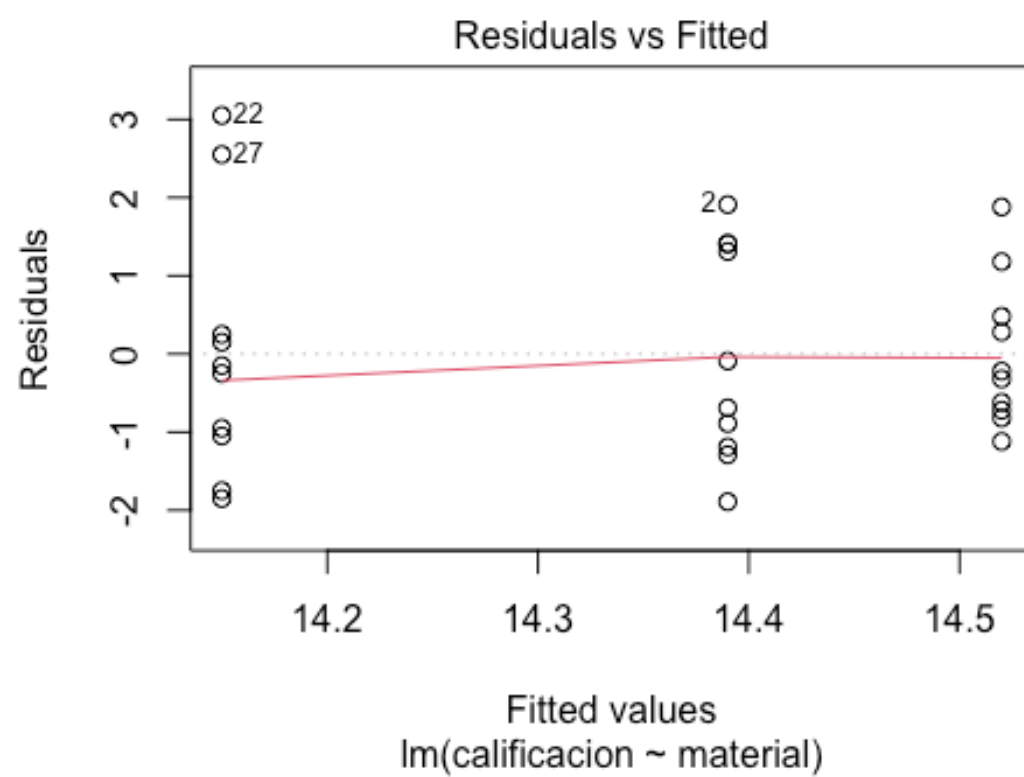
#Normalidad

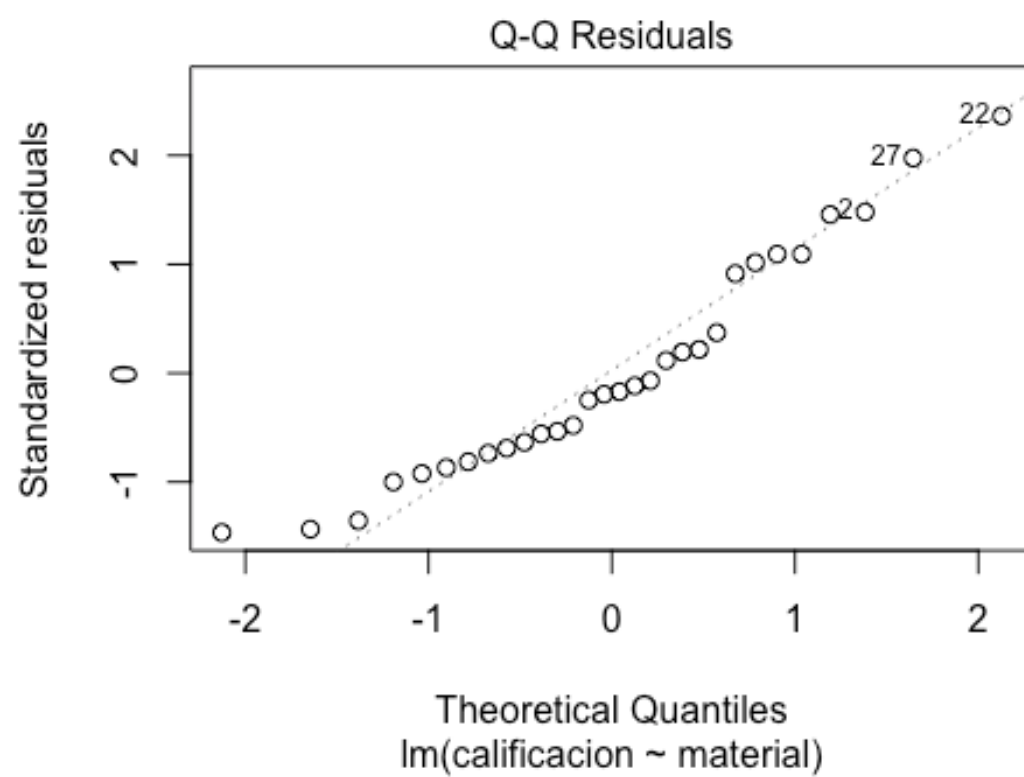
#Homocedasticidad

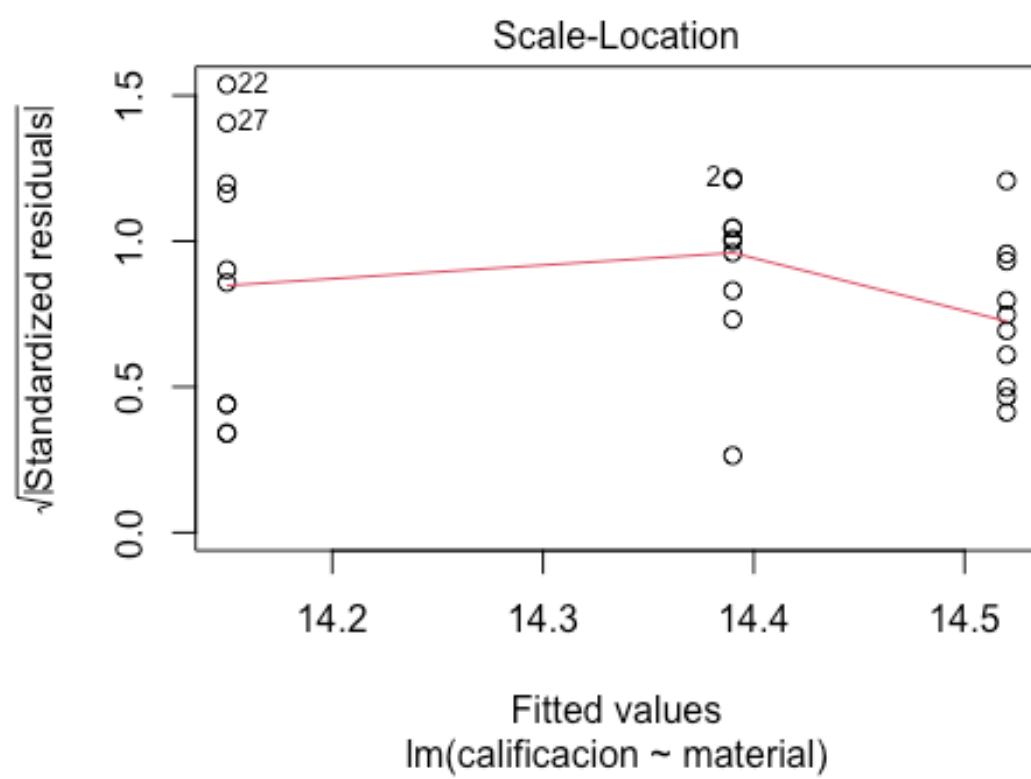
#Independencia

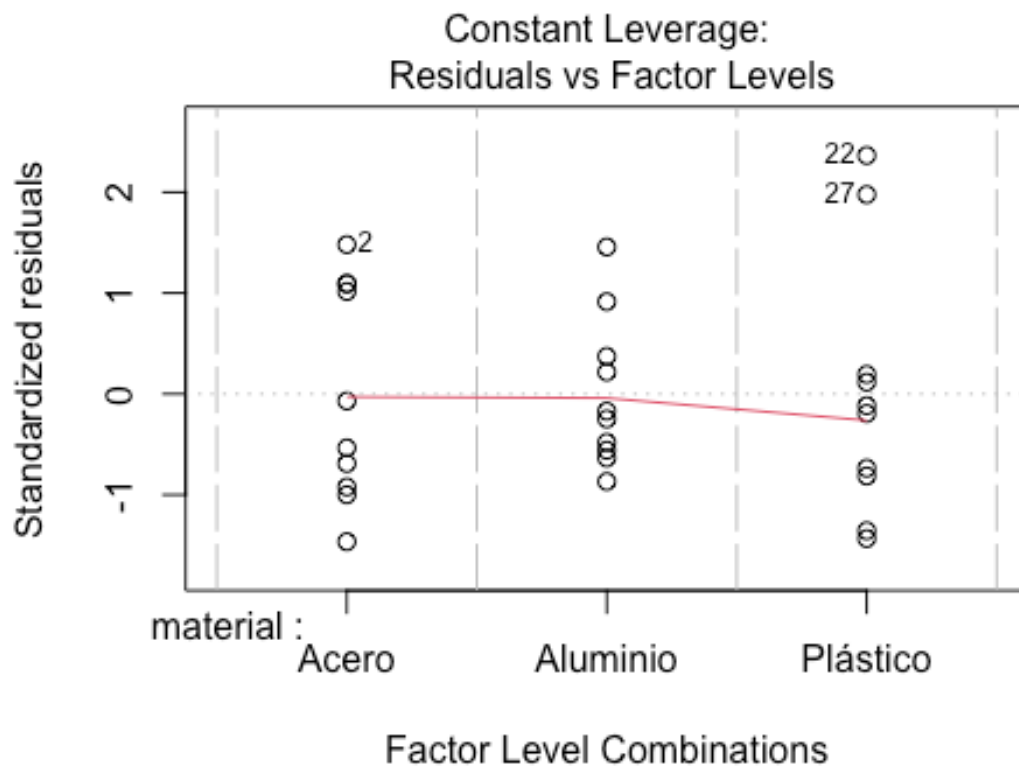
#Relación lineal entre las variables (coeficiente de determinación).

```
plot(lm(calificacion~material))
```









CD= $150 / (150 + 76)$

7. Concluye en el contexto del problema.

```
print('El mejor proveedor se demostro ser el segundo, con mejores resultados de calificacion, y el mejor material se puede considerar que es el acero, pero como el aluminio tiene una media mas alta que el acero, se puede demostrar mas confiable el aluminio.')
```

```
## [1] "El mejor proveedor se demostro ser el segundo, con mejores resultados de calificacion, y el mejor material se puede considerar que es el acero, pero como el aluminio tiene una media mas alta que el acero, se puede demostrar mas confiable el aluminio."
```