Actividad 9: ANOVA

Daniela Jiménez Téllez

2024-08-27

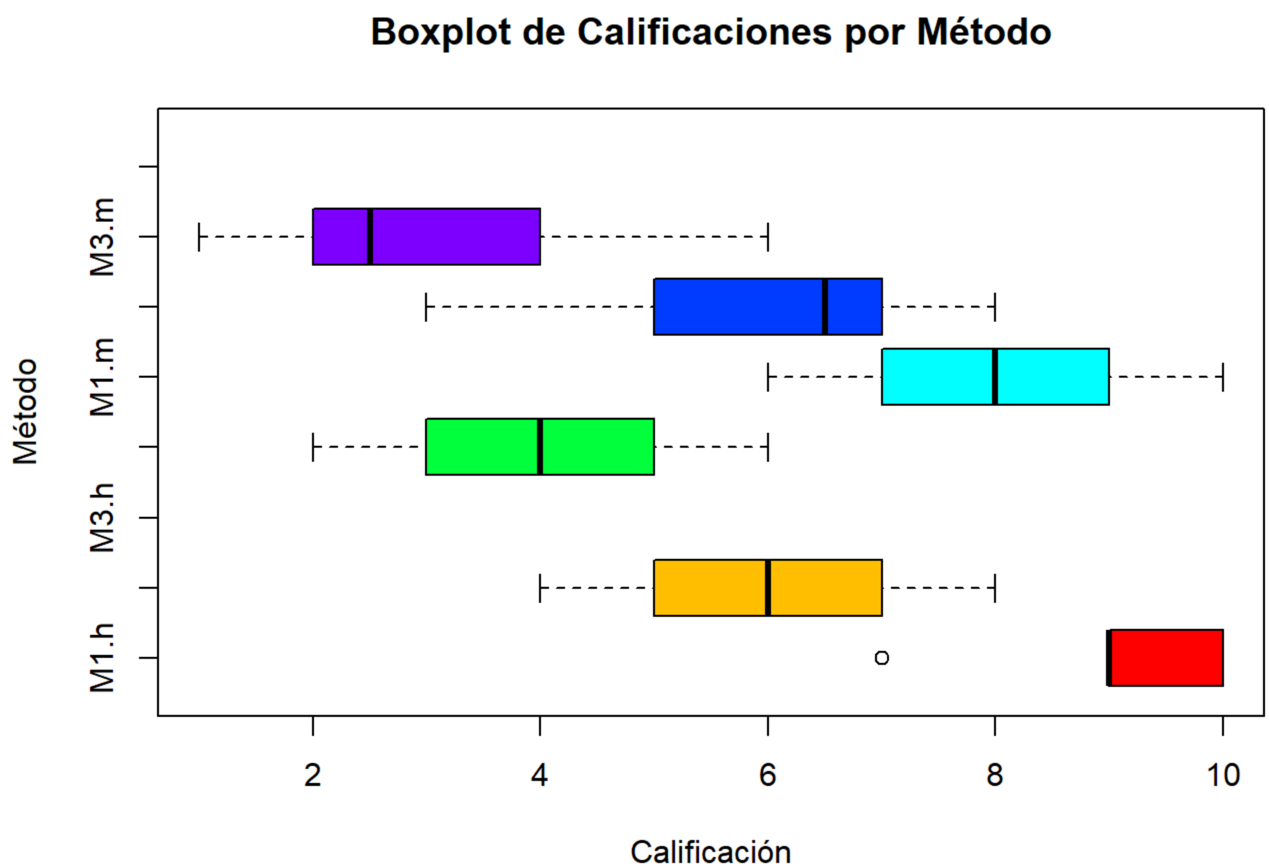
Problema 1: El rendimiento

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:

```
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)
```

- **Análisis exploratorio**
  - Haz el boxplot de la evaluación de los estudiantes por método de enseñanza y sexo.

```
# BoxPlot

boxplot(calificacion ~ interaction(metodo, sexo), col = rainbow(length(levels(interaction(metod
o, sexo)))), main = "Boxplot de Calificaciones por Método", xlab = "Calificación", ylab = "Métod
o", horizontal = TRUE)
```



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

De acuerdo con la gráfica podemos notar que el método más efectivo es el M1, ya que muestra un mejor desempeño en general. De igual manera, que su mediana cae dentro del octavo rango, lo que nos dice que las calificaciones con este método son altas. Asimismo, su IQR no es tan amplio, lo que nos dice que las calificaciones no variaron tanto. Por otro lado, en el caso de M3 podemos ver que el desempeño de los alumnos no fue el mejor ya que se encuentra de lado izquiero de la gráfica, donde las calificaciones son menores. Igualmente, podemos observar que tiene una mediana baja, la cual está al rededor de 4.

- Escribe tus conclusiones parciales.

Habiendo dicho lo anterior y observando la gráfica, nos podemos dar cuenta que el Método 1 es el más efectivo, y que los hombres muestran un mejor desempeño en este. Por otro lado, el método menos efectivo sería el Método 3, y esto se ve mejor reflejado en el caso de las mujeres.

- **Las hipótesis estadísticas**
- **Realiza el ANOVA para dos niveles con interacción**

```
A <- aov(calificacion ~ interaction(metodo, sexo))

cat("Resultados del ANOVA:", "\n")
```

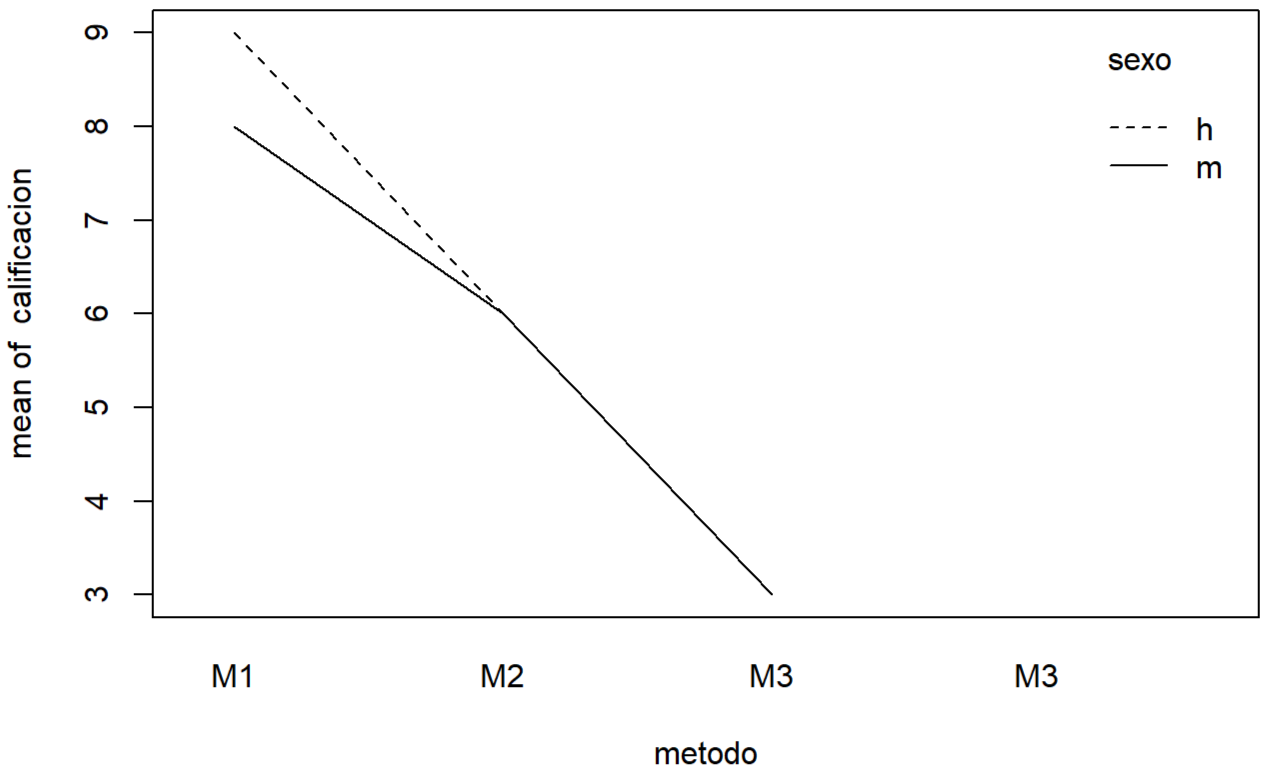
```
## Resultados del ANOVA:
```

```
summary(A)
```

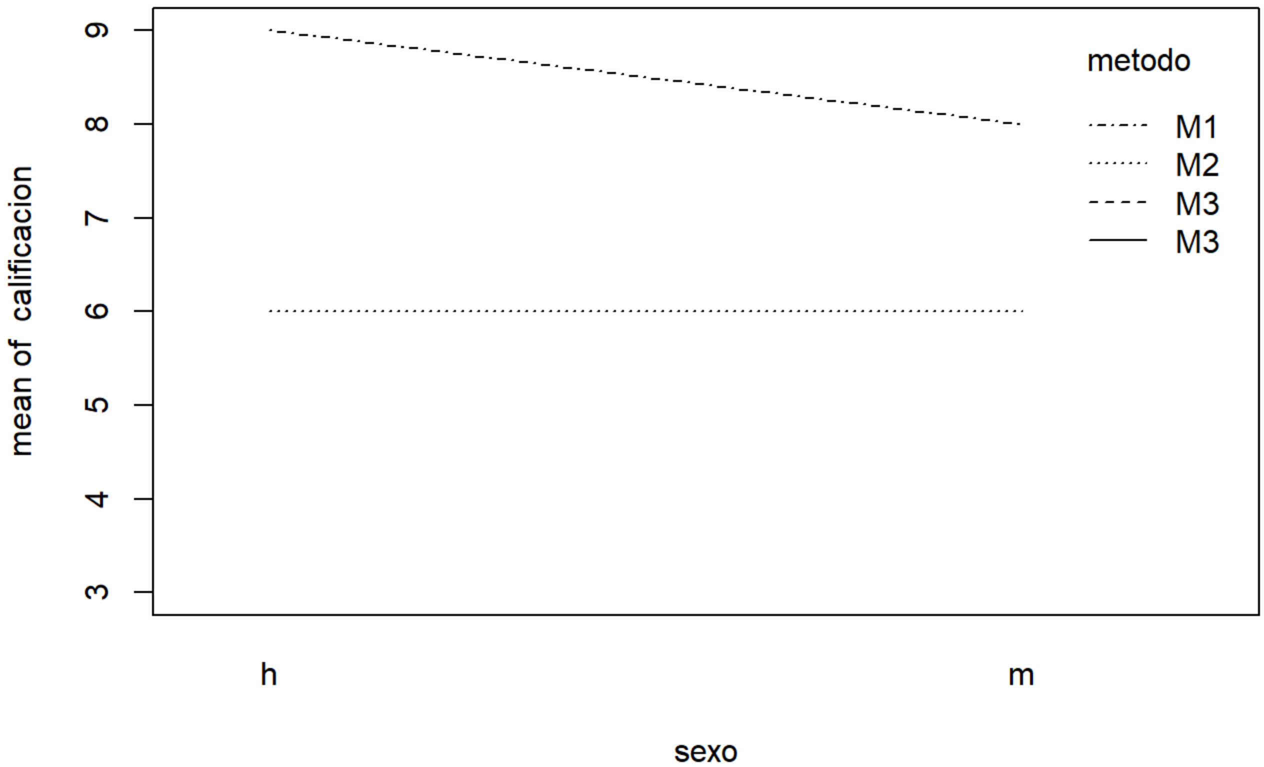
```
##                           Df Sum Sq Mean Sq F value   Pr(>F)    
## interaction(metodo, sexo)  5    156  31.200   13.37 6.82e-07 ***
## Residuals                 30     70   2.333                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Haz la gráfica de interacción de dos factores en ANOVA

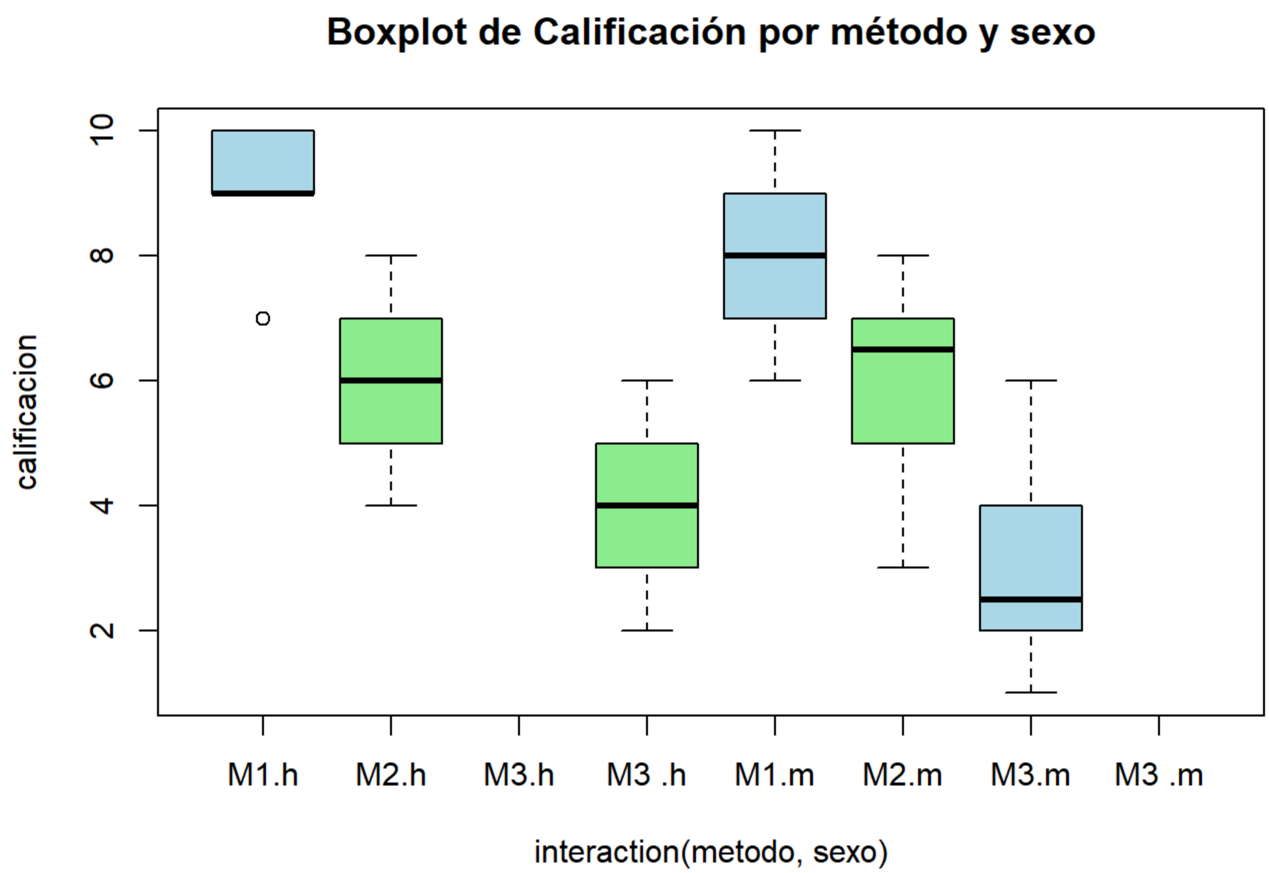
```
interaction.plot(metodo, sexo, calificacion)
```




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




```
boxplot(calificacion ~ interaction(metodo, sexo), col = c("lightblue", "lightgreen"), main = "Bo
xplot de Calificación por método y sexo")
```



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

Así como se dijo anteriormente, podemos notar que el método más efectivo es el M1, ya que muestra un mejor desempeño en general. De igual manera, que su mediana cae dentro del octavo rango, lo que nos dice que las calificaciones con este método son altas. Asimismo, su IQR no es tan amplio, lo que nos dice que las calificaciones no variaron tanto. Por otro lado, en el caso de M3 podemos ver que el desempeño de los alumnos no fue el mejor ya que se encuentra de lado izquiero de la gráfica, donde las calificaciones son menores. Igualmente, podemos observar que tiene una mediana baja, la cual está al rededor de 4. Finalmente, en esta gráfica se ve mejor el desempeño del Método 2, el cual se muestra como un método bueno, el cual tiene una mediana al rededor de 6. Esto nos dice que tal vez no es el mejor método, sin embargo muestra mejores resultados que el M3.

- Escribe tus conclusiones parciales.

Habiendo analizado lo anterior, se concluye que el mejor método es el M1, luego el M2, y finalmente el M3.

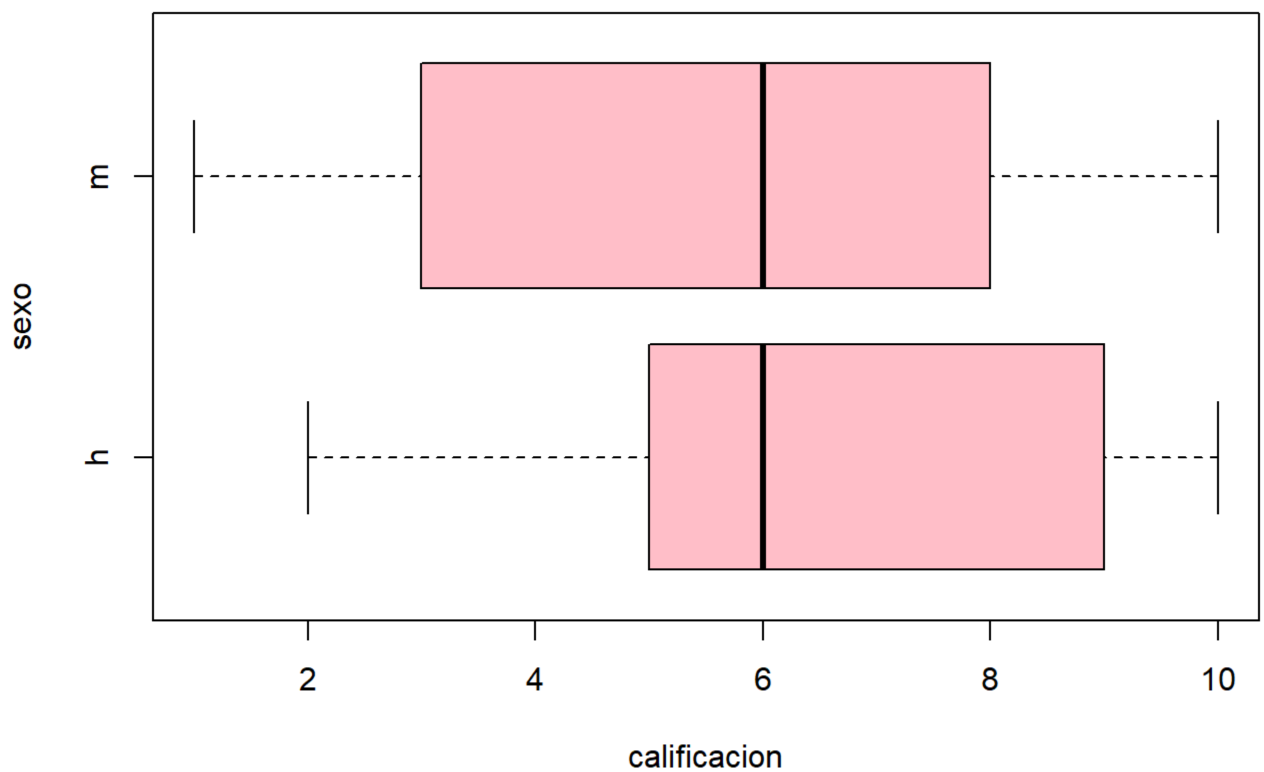
- **Realiza el ANOVA para un efecto principal**

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

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## metodo       3  153.0   51.00   22.11 7.56e-08 ***
## sexo         1    1.5    1.50    0.65    0.426    
## Residuals   31   71.5    2.31                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
# BoxPlot

boxplot(calificacion ~ sexo, col = "pink", horizontal = TRUE)
```




```
# Media

msex = tapply(calificacion, sexo, mean)
mmet = tapply(calificacion, metodo, mean)

cat("La media para la calificación por sexo es: \n")
```

```
## La media para la calificación por sexo es:
```

```
msex
```

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

```
cat("La media para la calificación por método es: \n")
```

```
## La media para la calificación por método es:
```

```
mmet
```

```
##  M1  M2  M3 M3
## 8.5 6.0 3.0 4.0
```

- Haz los intervalos de confianza de rendimiento por sexo. Grafícalos.

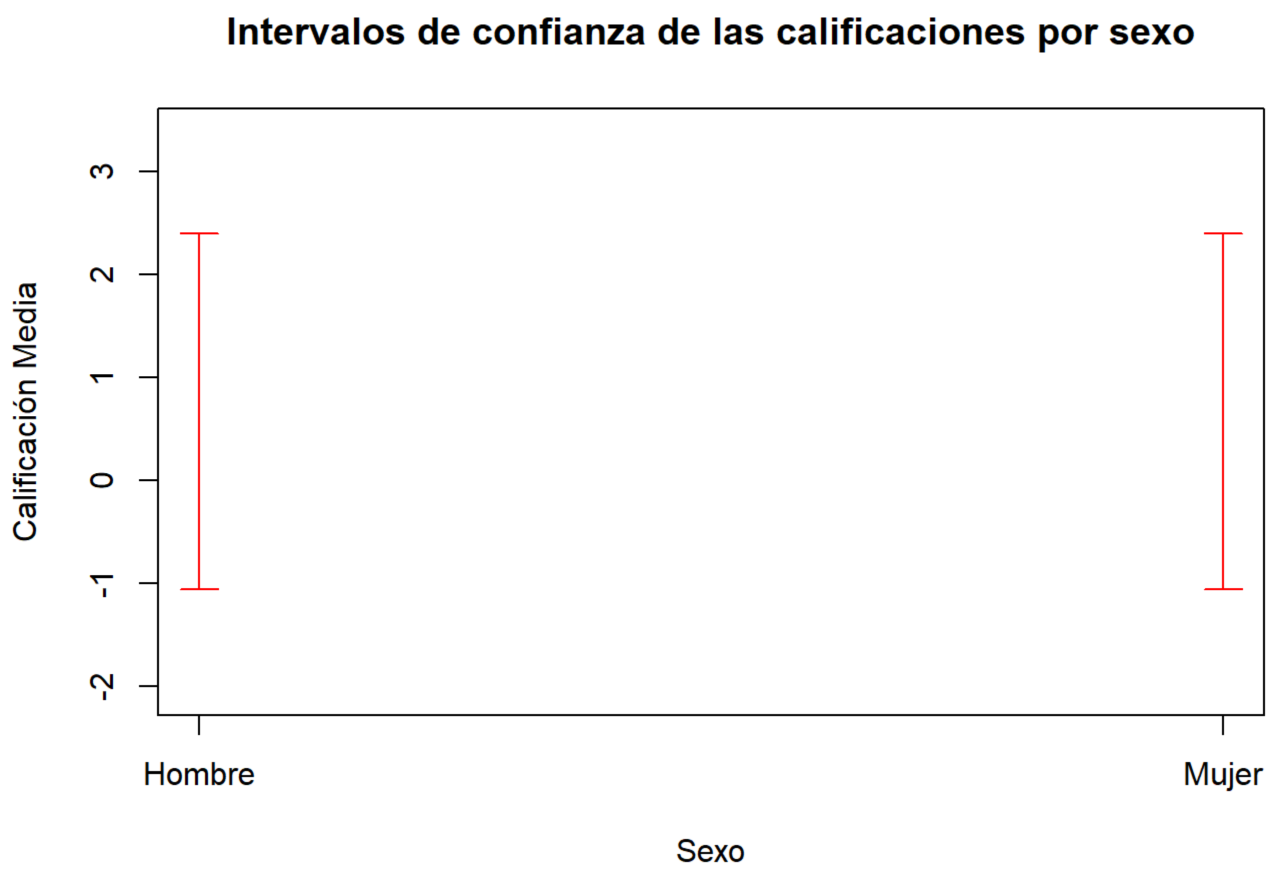
```
int_sex = t.test(calificacion ~ sexo)$conf.int

cat("Los intervalos de confianza de las calificaciones por sexo son: \n")
```

```
## Los intervalos de confianza de las calificaciones por sexo son:
```

```
int_sex
```

```
## [1] -1.064566  2.397899
## attr(,"conf.level")
## [1] 0.95
```

```
media_hombre <- mean(calificacion[sexo == "h"])
media_mujer <- mean(calificacion[sexo == "m"])

medias <- c(media_hombre, media_mujer)
sexos <- c("Hombre", "Mujer")
x_pos <- 1:length(sexos)

plot(x_pos, medias, ylim = c(min(int_sex) - 1, max(int_sex) + 1),
     xaxt = "n", pch = 16, col = "blue", xlab = "Sexo", ylab = "Calificación Media",
     main = "Intervalos de confianza de las calificaciones por sexo")

axis(1, at = x_pos, labels = sexos)

arrows(x0 = x_pos, y0 = rep(int_sex[1], length(x_pos)), y1 = rep(int_sex[2], length(x_pos)),
       angle = 90, code = 3, length = 0.1, col = "red")
```



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

En la gráfica podemos ver que debido a que los intervalos de confianza son muy parecidos, y además muy grandes, se puede decir que no hay una gran diferencia entre las calificaciones medias de hombres y mujeres.

- Escribe tus conclusiones parciales.

Las calificaciones de los hombres no son drásticamente diferentes.

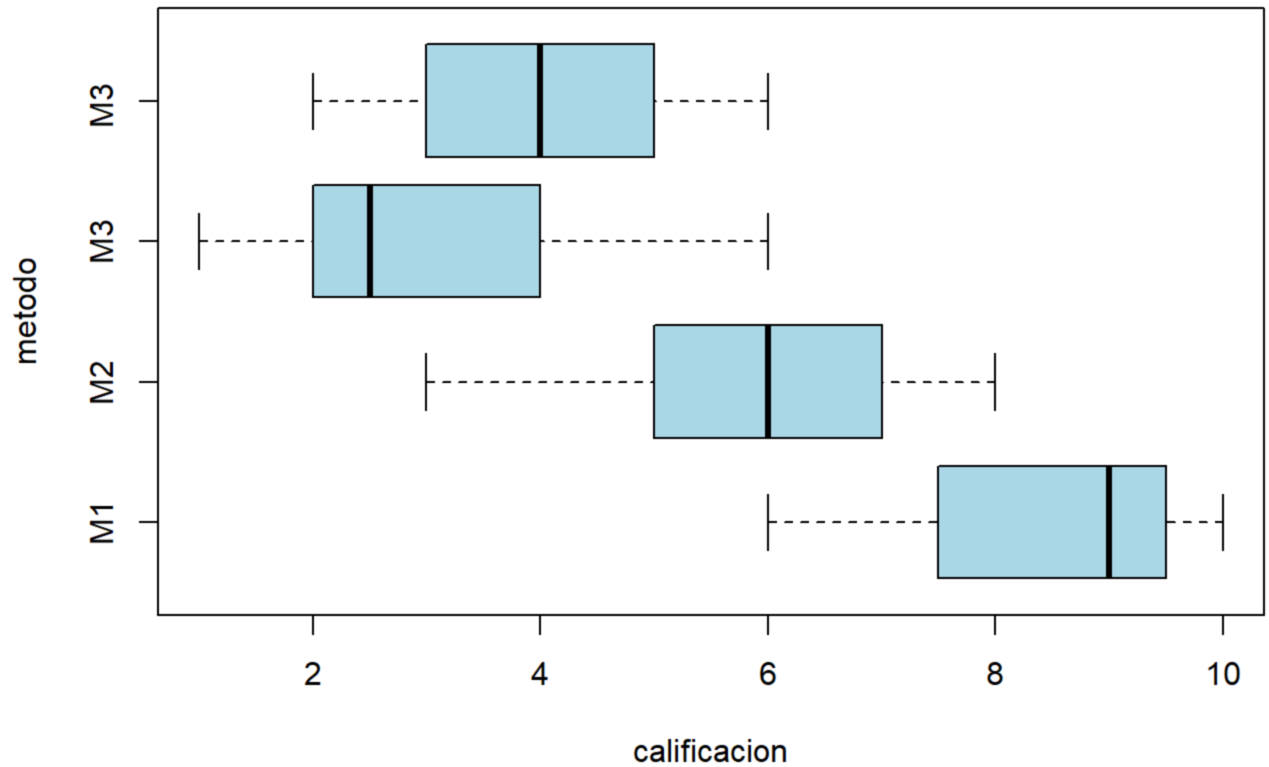
- **Realiza el ANOVA para un efecto principal**

```
C <- aov(calificacion ~ metodo)

summary(C)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## metodo       3    153   51.00   22.36 5.41e-08 ***
## Residuals   32     73    2.28                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Haz el boxplot de rendimiento por método de enseñanza. Calcula la media.

```
# Boxplot

boxplot(calificacion ~ metodo, col = "lightblue", horizontal = TRUE)
```




```
# Media

mmet = tapply(calificacion, metodo, mean)

cat("La media para la calificación por método es: \n")
```

```
## La media para la calificación por método es:
```

```
mmet
```

```
##  M1  M2  M3 M3
## 8.5 6.0 3.0 4.0
```

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

```
n_M1 <- length(calificacion[metodo == "M1"])
mean_M1 <- mean(calificacion[metodo == "M1"])
stderr_M1 <- sd(calificacion[metodo == "M1"]) / sqrt(n_M1)
error_margin_M1 <- qt(0.975, df = n_M1 - 1) * stderr_M1
int_met_M1 <- c(mean_M1 - error_margin_M1, mean_M1 + error_margin_M1)

n_M2 <- length(calificacion[metodo == "M2"])
mean_M2 <- mean(calificacion[metodo == "M2"])
stderr_M2 <- sd(calificacion[metodo == "M2"]) / sqrt(n_M2)
error_margin_M2 <- qt(0.975, df = n_M2 - 1) * stderr_M2
int_met_M2 <- c(mean_M2 - error_margin_M2, mean_M2 + error_margin_M2)

n_M3 <- length(calificacion[metodo == "M3"])
mean_M3 <- mean(calificacion[metodo == "M3"])
stderr_M3 <- sd(calificacion[metodo == "M3"]) / sqrt(n_M3)
error_margin_M3 <- qt(0.975, df = n_M3 - 1) * stderr_M3
int_met_M3 <- c(mean_M3 - error_margin_M3, mean_M3 + error_margin_M3)

cat("Intervalos de confianza para M1: ", int_met_M1, "\n")
```

```
## Intervalos de confianza para M1:  7.664961 9.335039
```

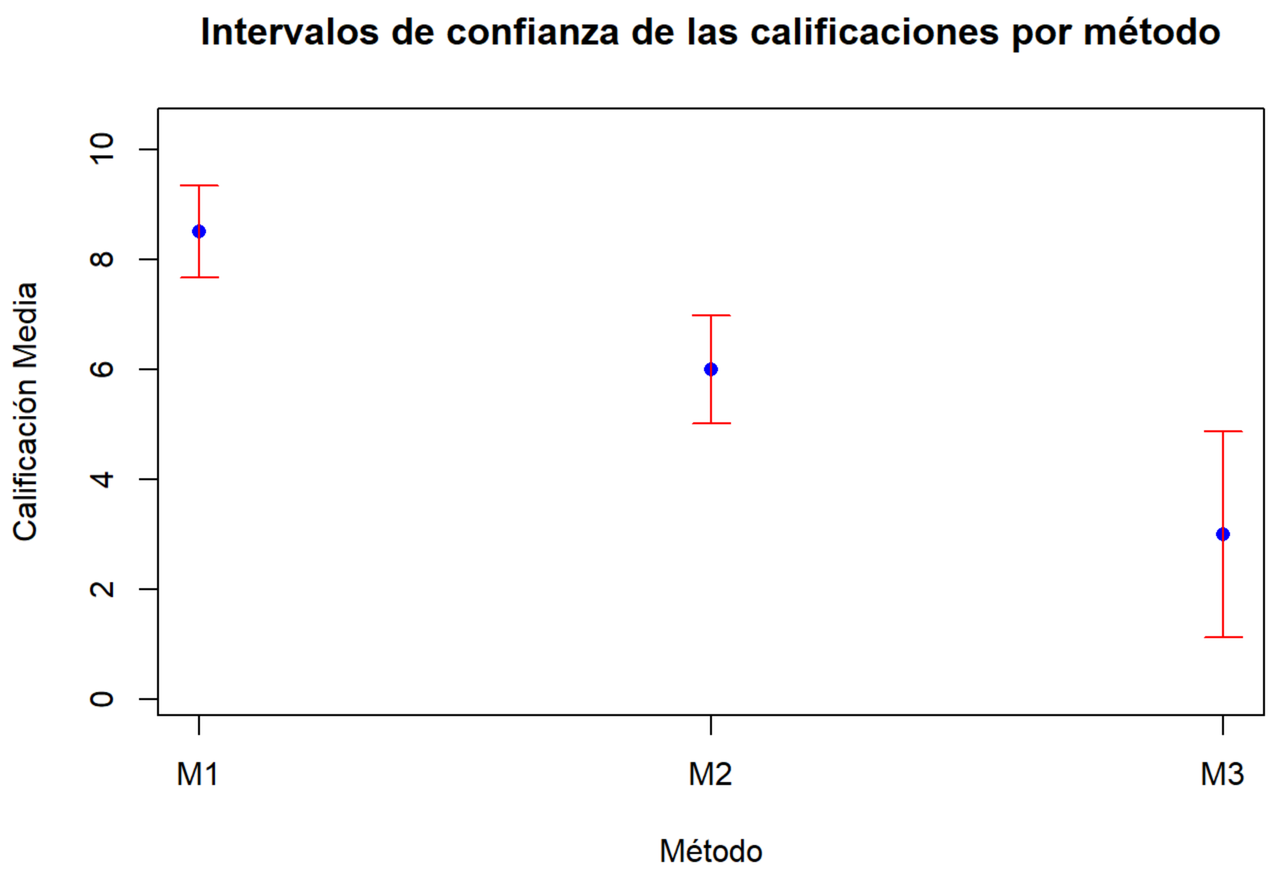
```
cat("Intervalos de confianza para M2: ", int_met_M2, "\n")
```

```
## Intervalos de confianza para M2:  5.023175 6.976825
```

```
cat("Intervalos de confianza para M3: ", int_met_M3, "\n")
```

```
## Intervalos de confianza para M3:  1.122712 4.877288
```

```
medias <- c(mean_M1, mean_M2, mean_M3)

metodos <- c("M1", "M2", "M3")

x_pos <- 1:length(medias)

plot(x_pos, medias, ylim = c(min(c(int_met_M1, int_met_M2, int_met_M3)) - 1,
                             max(c(int_met_M1, int_met_M2, int_met_M3)) + 1),
     xaxt = "n", pch = 16, col = "blue", xlab = "Método", ylab = "Calificación Media",
     main = "Intervalos de confianza de las calificaciones por método")

axis(1, at = x_pos, labels = metodos)

arrows(x0 = x_pos, y0 = c(int_met_M1[1], int_met_M2[1], int_met_M3[1]),
       y1 = c(int_met_M1[2], int_met_M2[2], int_met_M3[2]),
       angle = 90, code = 3, length = 0.1, col = "red")
```



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

```
I = TukeyHSD(C)
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 -4.170620 -0.82937953 0.0016311
## M3-M1  -5.5 -7.546084 -3.45391615 0.0000002
## M3 -M1 -4.5 -6.546084 -2.45391615 0.0000071
## M3-M2  -3.0 -5.046084 -0.95391615 0.0020429
## M3 -M2 -2.0 -4.046084  0.04608385 0.0573005
## M3 -M3  1.0 -1.362614  3.36261412 0.6638727
```

```
plot(I)
```

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

Como se mencionó en las gráficas anteriores, para el M1 podemos ver que su media varía muy poco al redor del número 8, lo que lo vuelve más confiable ya que no hay valores extremos dentro de este. Por otro lado, en el caso de M2 su media va al redor de 5, lo que nos dice que las calificaciones obtenidas por este método son menores que las del M1. Finalmente, para el M3 podemos ver que su media varía al redor de 2, y que además sus intervalos son muy amplios, lo que nos dice que las calificaciones pueden variar mucho al usar este método.

- Escribe tus conclusiones parciales.

En la gráfica podemos observar la efectividad de cada método, siendo el M1 el mejor, luego el M2 y al final el M3.

- **Comprueba la validez del modelo. Comprueba:**
  - Normalidad

```
shapiro.test(resid(C))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(C)
## W = 0.97449, p-value = 0.5603
```

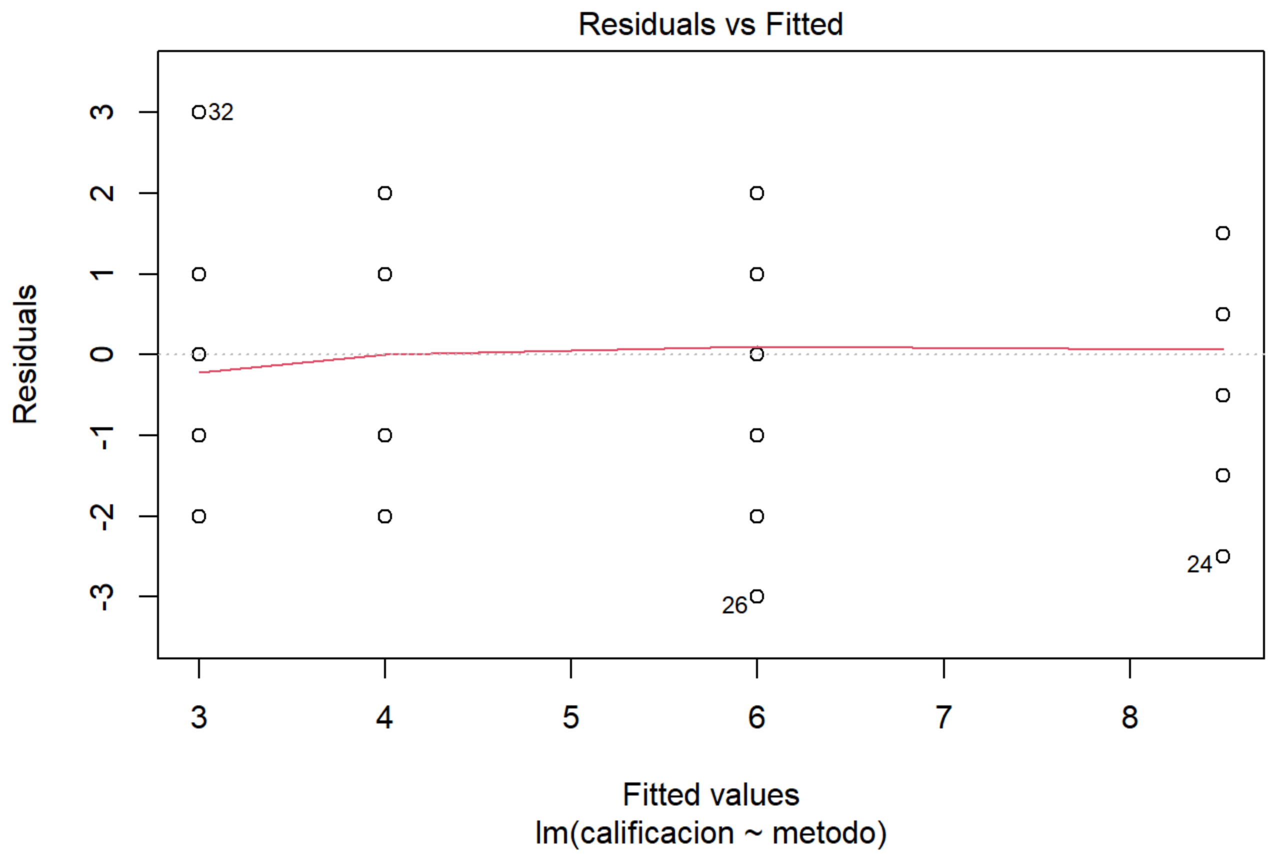
- Homocedasticidad

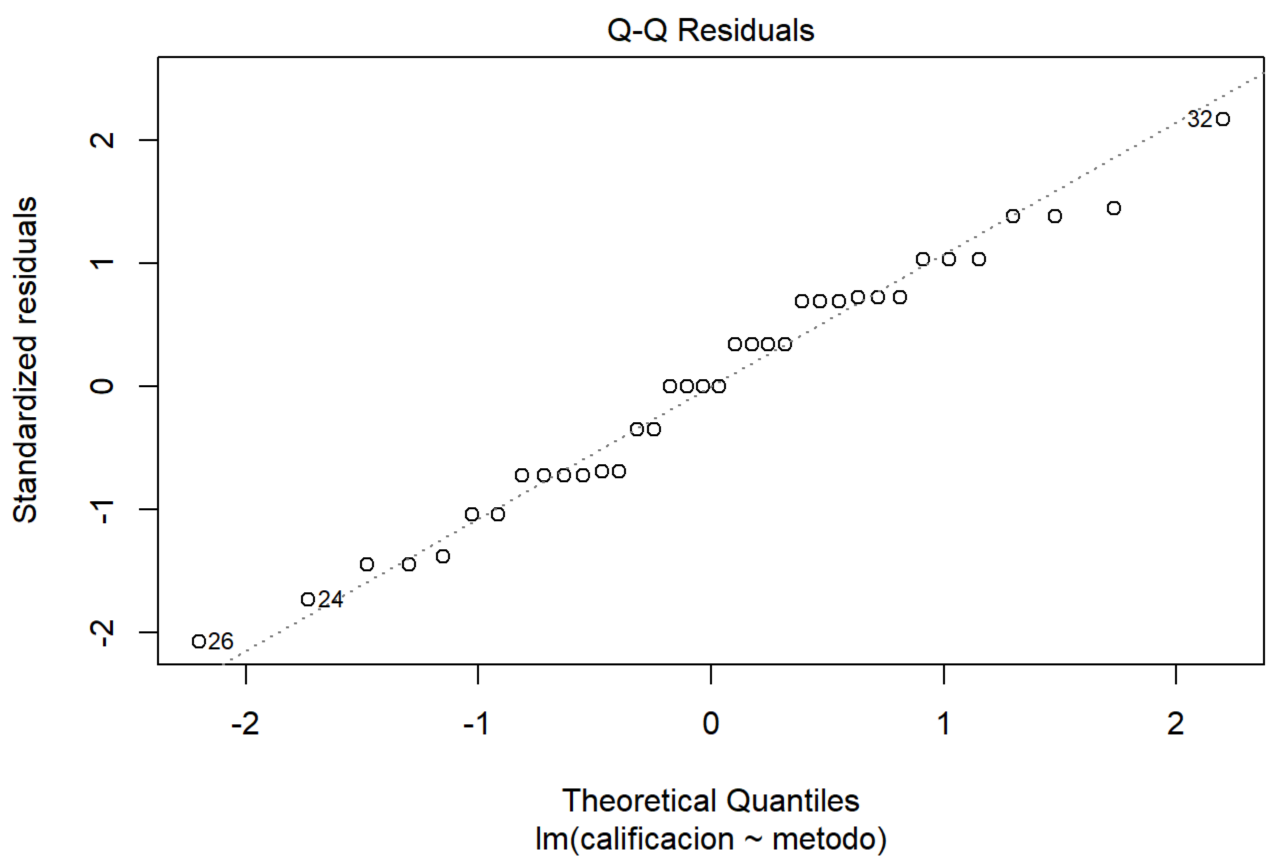
```
bartlett.test(calificacion ~ metodo)
```

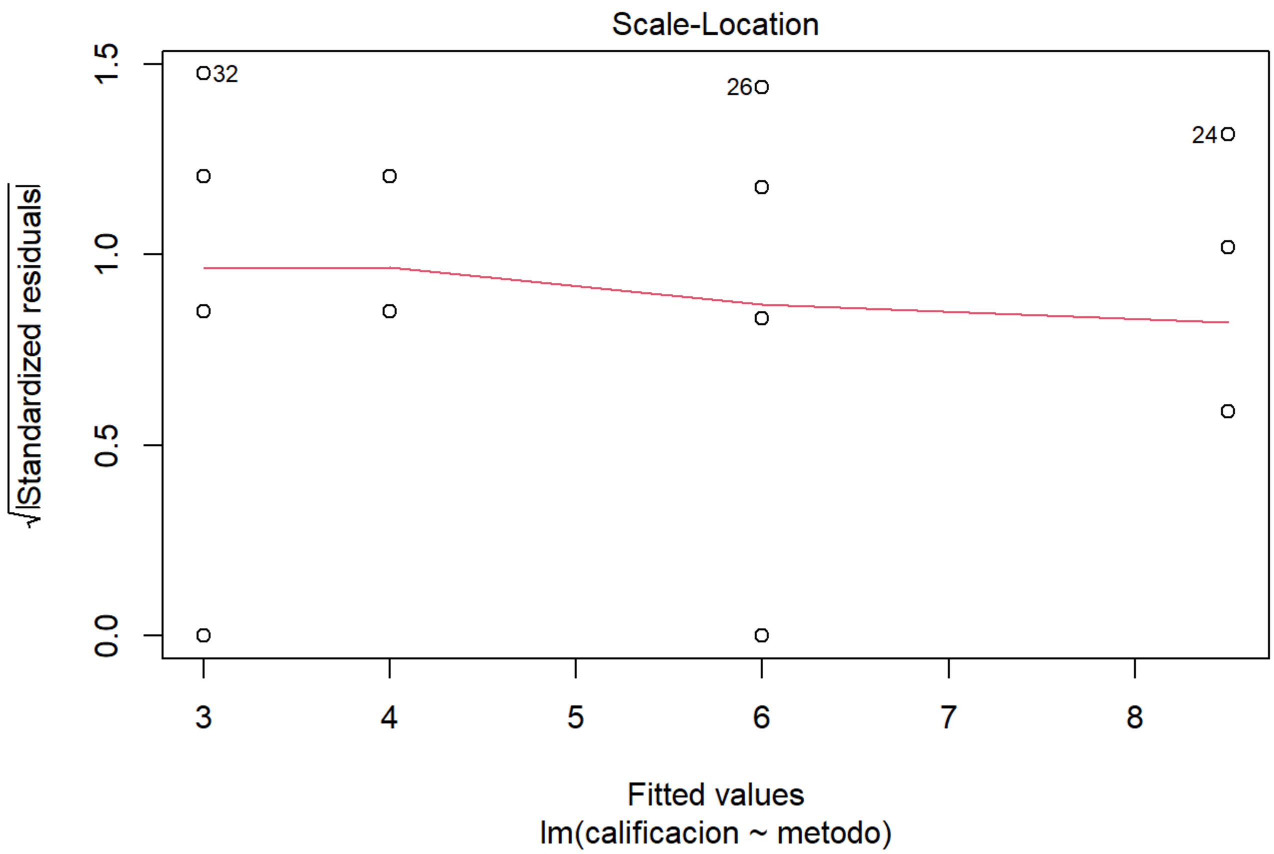
```
##
##  Bartlett test of homogeneity of variances
##
## data:  calificacion by metodo
## Bartlett's K-squared = 0.68213, df = 3, p-value = 0.8774
```

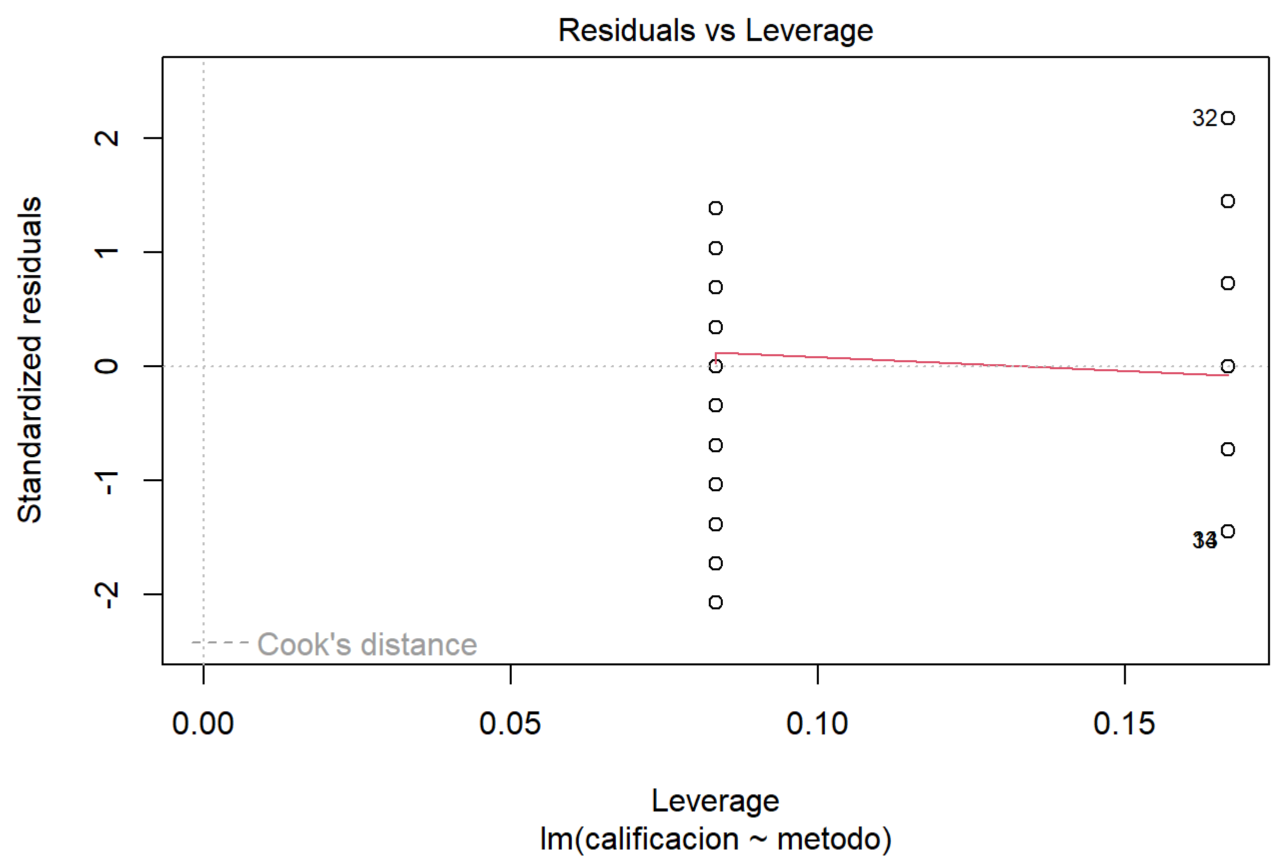
- Relación lineal entre las variables (coeficiente de determinación)

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


Residuals vs Fitted
Residuals
Fitted values
lm(calificacion ~ metodo)
32
26
24


Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(calificacion ~ metodo)
32
24
26


Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
32
26
24
3
4
5
6
7
8
Fitted values
lm(calificacion ~ metodo)


Residuals vs Leverage
Standardized residuals
-2
-1
0
1
2
32
Cook's distance
0.00
0.05
0.10
0.15
Leverage
lm(calificacion ~ metodo)

```
CD = 150/(150 + 76)
```

- **Concluye en el contexto del problema.**

Finalmente, se puede concluir que los hombres tienen mejor desempeño en cualquier método; sin embargo, es más evidente para el Método 1, el que es el mejor método de enseñanza. Usando este método se obtienen mejores calificaciones que con el M2 y el M3. Igualmente, concluyo que el peor de los métodos es el M3 ya que las calificaciones que se obtuvieron con este fueron muy bajas.

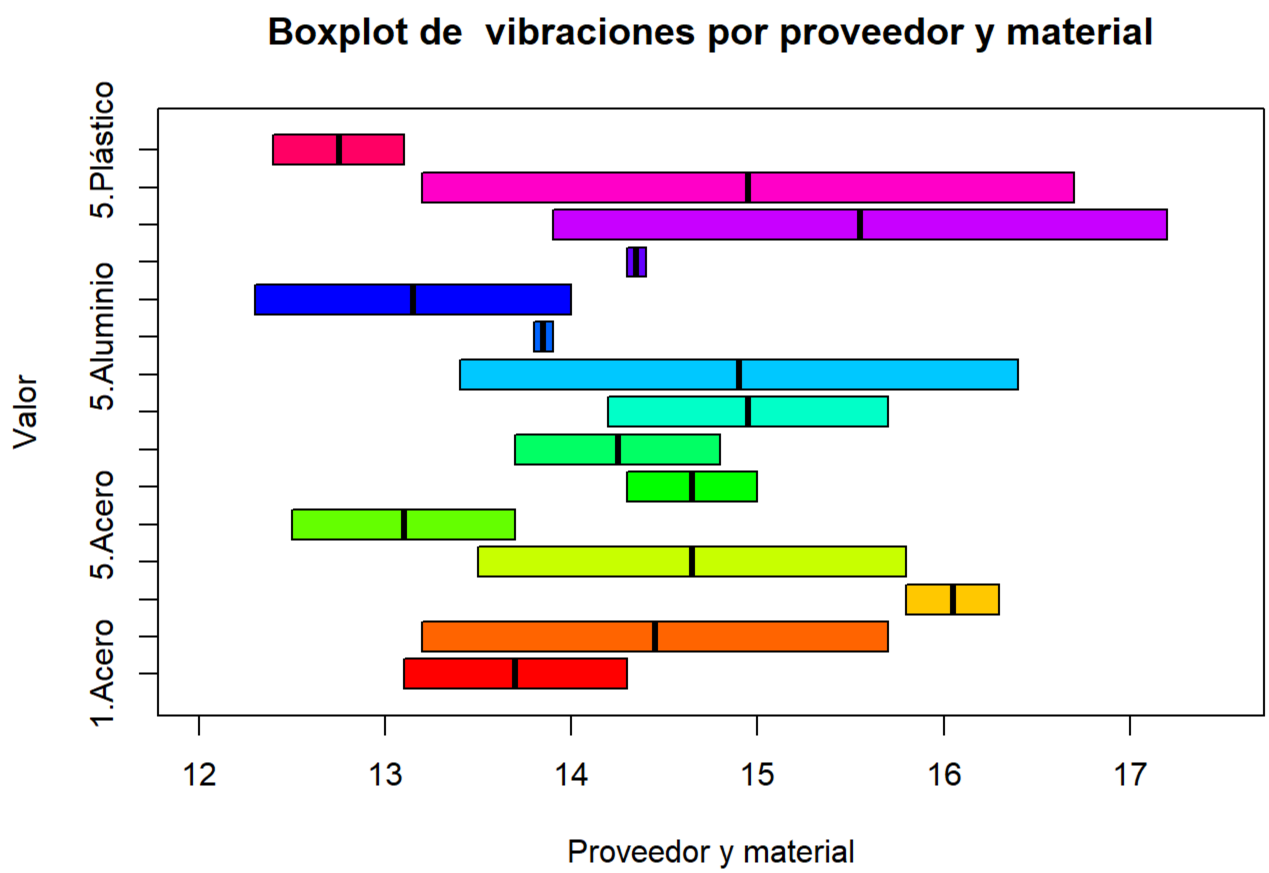
Problema 2: Vibración de motores

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. Los resultados son los siguientes:

```
valores = c(13.1, 13.2, 16.3, 15.8, 13.7, 14.3, 15.7, 15.8, 13.5, 12.5,
            15.0, 14.8, 15.7, 16.4, 13.9, 14.3, 13.7, 14.2, 13.4, 13.8,
            14.0, 14.3, 17.2, 16.7, 12.4, 12.3, 14.4, 13.9, 13.2, 13.1)

material = c(rep("Acero", 10), rep("Aluminio", 10), rep("Plástico", 10))

proveedor = rep(1:5, 6)

material = factor(material)
proveedor = factor(proveedor)
```

- **Análisis exploratorio**
  - Haz el boxplot de los valores por proveedor y material.

```
# BoxPlot

boxplot(valores ~ interaction(proveedor, material), col = rainbow(length(levels(interaction(prov
eedor, material)))), main = "Boxplot de  vibraciones por proveedor y material", xlab = "Proveedo
r y material", ylab = "Valor", ylim = c(12, 17.5), horizontal = TRUE)
```



- Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.
- Escribe tus conclusiones parciales.
- **Las hipótesis estadísticas**
- **Realiza el ANOVA para dos niveles con interacción**

```
A <- aov(valores ~ interaction(material, proveedor))

cat("Resultados del ANOVA:", "\n")
```

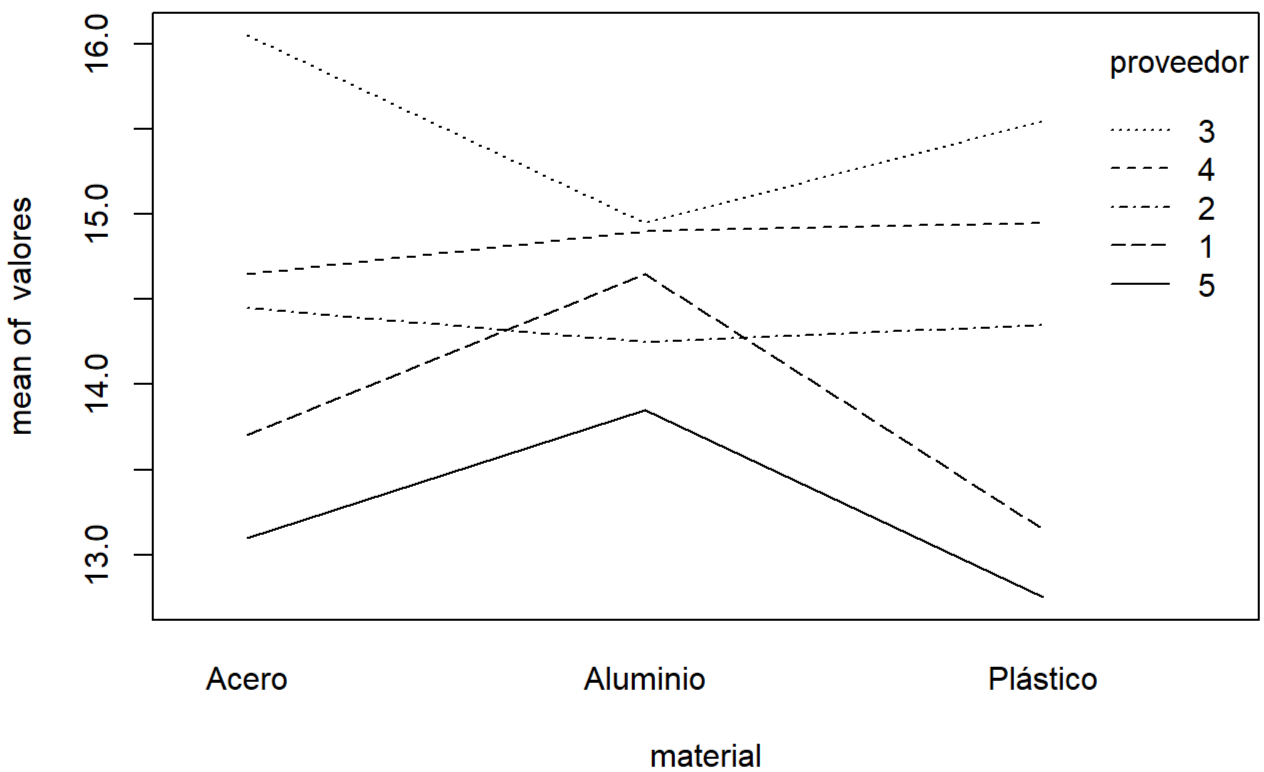
```
## Resultados del ANOVA:
```

```
summary(A)
```

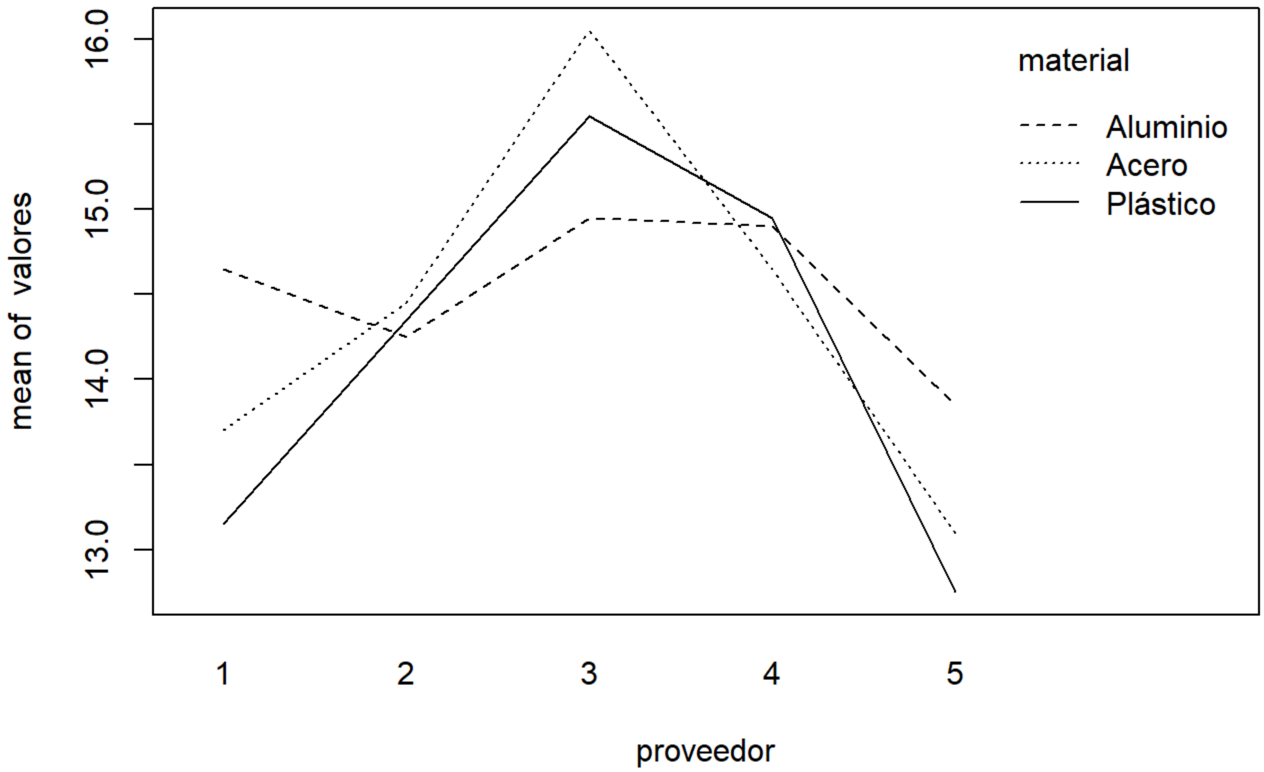
```
##                                  Df Sum Sq Mean Sq F value Pr(>F)
## interaction(material, proveedor) 14  23.57   1.684   0.933  0.549
## Residuals                        15  27.08   1.805
```

- Haz la gráfica de interacción de dos factores en ANOVA

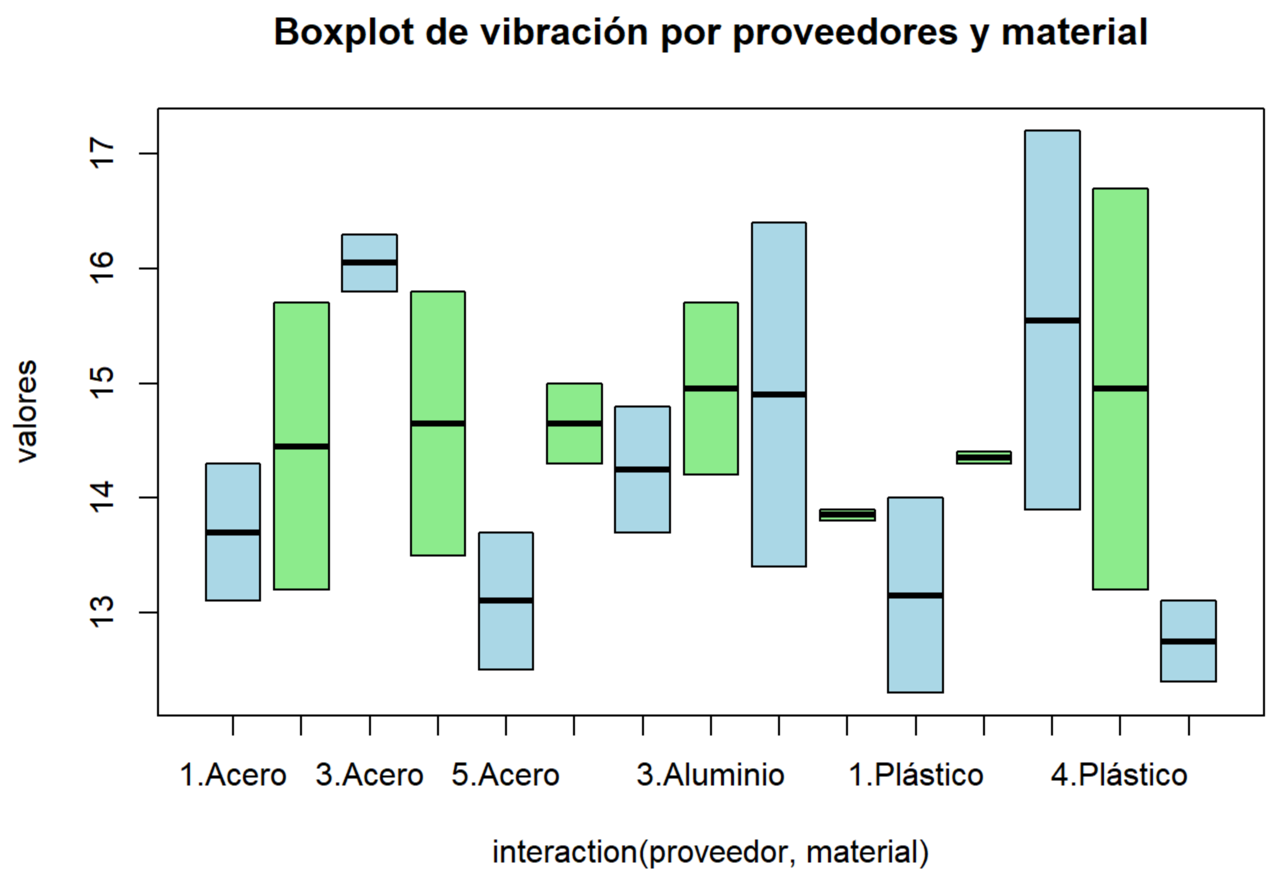
```
interaction.plot(material , proveedor, valores)
```




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




```
boxplot(valores ~ interaction(proveedor, material), col = c("lightblue", "lightgreen"), main =
"Boxplot de vibración por proveedores y material")
```



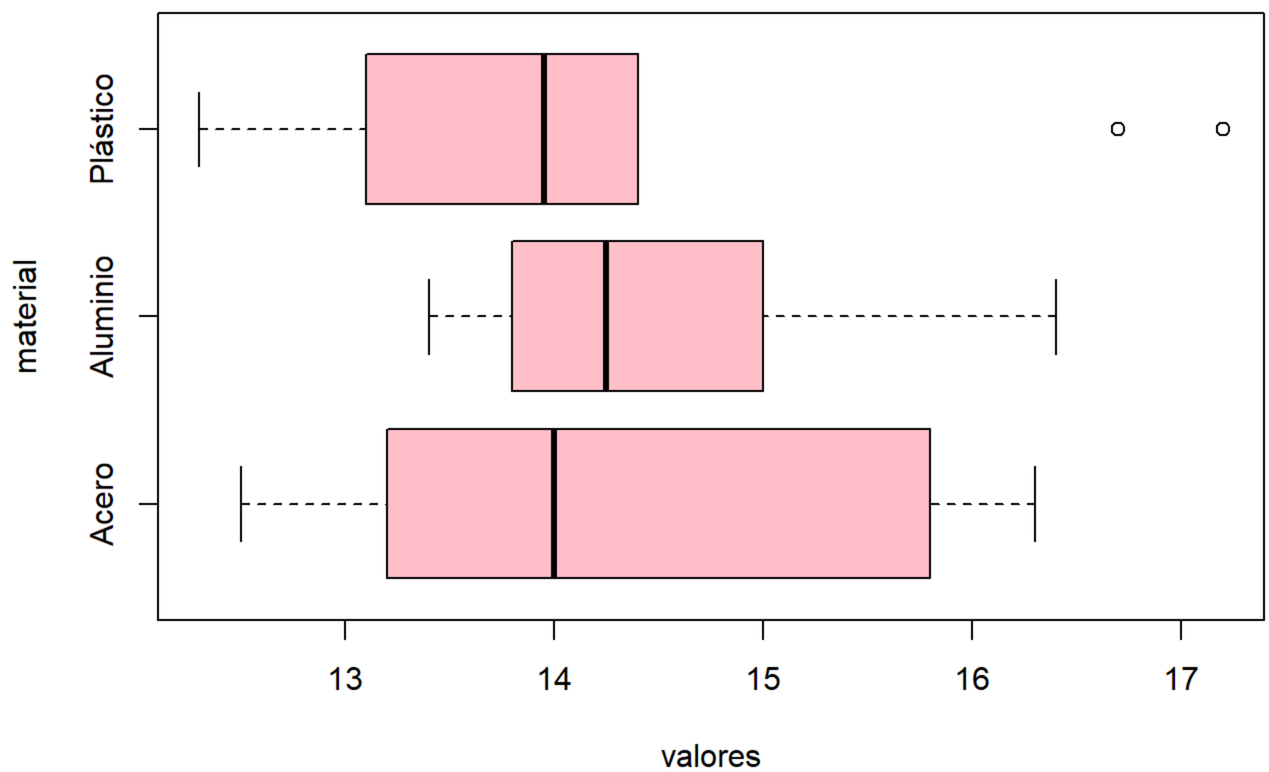
- Interpreta el resultado desde la perspectiva estadística y el contexto del problema.
- Escribe tus conclusiones parciales.
- **Realiza el ANOVA para un efecto principal**

```
B <- aov(valores ~ proveedor + material)
summary(B)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## proveedor    4 18.651   4.663   3.427 0.0245 *
## material     2  0.705   0.352   0.259 0.7741
## Residuals   23 31.299   1.361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Haz el boxplot de vibraciones por material. Calcula la media para las vibraciones por material y proveedor.

```
# BoxPlot
boxplot(valores ~ material, col = "pink", horizontal = TRUE)
```




```
# Media

mmat = tapply(valores, material, mean)
mprov = tapply(valores, proveedor, mean)

cat("La media para la calificación por material es: \n")
```

```
## La media para la calificación por material es:
```

```
mmat
```

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

```
cat("La media para la calificación por proveedor es: \n")
```

```
## La media para la calificación por proveedor es:
```

```
mprov
```

```
##        1        2        3        4        5
## 13.83333 14.35000 15.51667 14.83333 13.23333
```

- Haz los intervalos de confianza de vibración por material. Grafícalos.

```
n_Acero <- length(valores[material == "Acero"])
mean_Acero <- mean(valores[material == "Acero"])
stderr_Acero <- sd(valores[material == "Acero"]) / sqrt(n_Acero)
error_margin_Acero <- qt(0.975, df = n_Acero - 1) * stderr_Acero
int_met_Acero <- c(mean_Acero - error_margin_Acero, mean_Acero + error_margin_Acero)

n_Aluminio <- length(valores[material == "Aluminio"])
mean_Aluminio <- mean(valores[material == "Aluminio"])
stderr_Aluminio <- sd(valores[material == "Aluminio"]) / sqrt(n_Aluminio)
error_margin_Aluminio <- qt(0.975, df = n_Aluminio - 1) * stderr_Aluminio
int_met_Aluminio <- c(mean_Aluminio - error_margin_Aluminio, mean_Aluminio + error_margin_Alumin
io)

n_Plastico <- length(valores[material == "Plástico"])
mean_Plastico <- mean(valores[material == "Plástico"])
stderr_Plastico <- sd(valores[material == "Plástico"]) / sqrt(n_Plastico)
error_margin_Plastico <- qt(0.975, df = n_Plastico - 1) * stderr_Plastico
int_met_Plastico <- c(mean_Plastico - error_margin_Plastico, mean_Plastico + error_margin_Plasti
co)

cat("Intervalos de confianza para Acero: ", int_met_Acero, "\n")
```

```
## Intervalos de confianza para Acero:  13.39909 15.38091
```

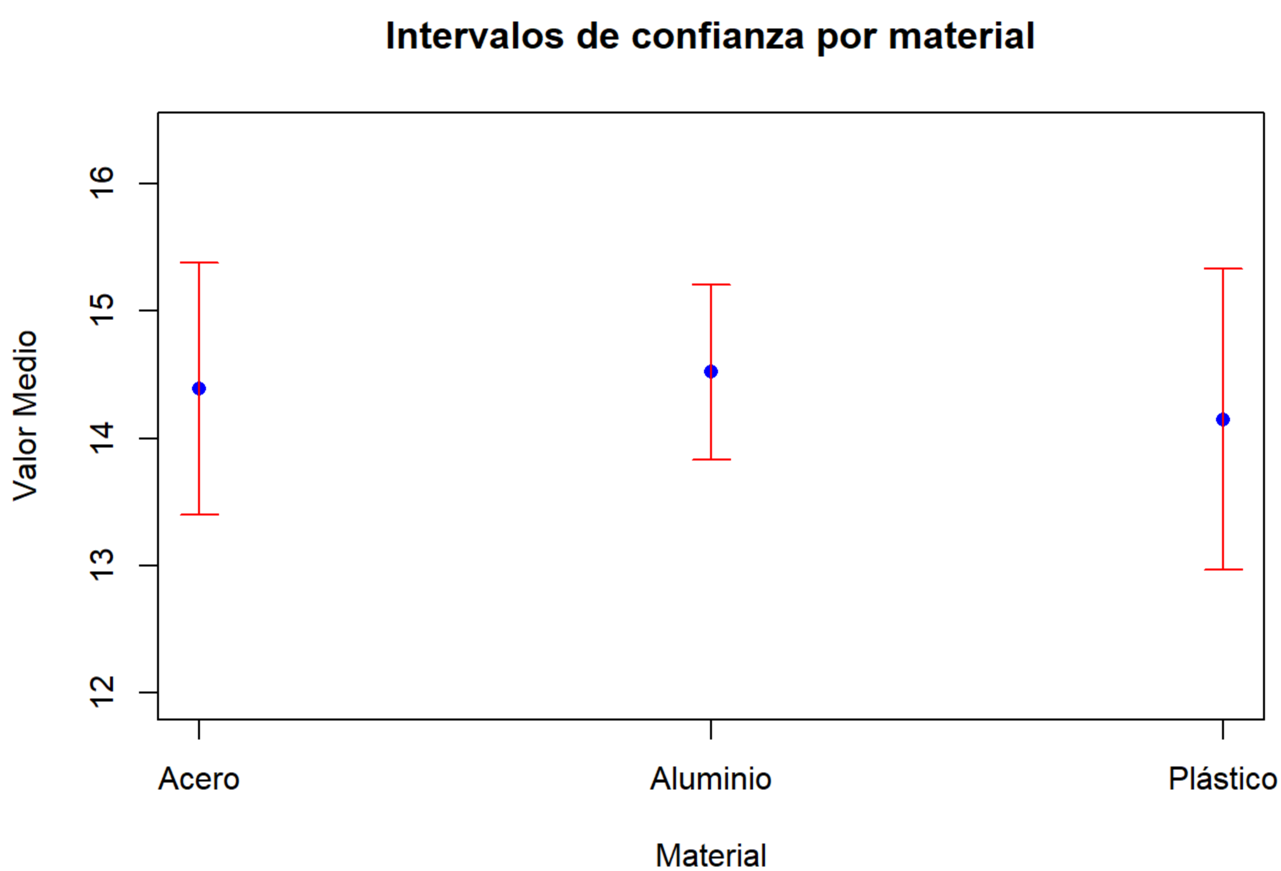
```
cat("Intervalos de confianza para Aluminio: ", int_met_Aluminio, "\n")
```

```
## Intervalos de confianza para Aluminio:  13.83651 15.20349
```

```
cat("Intervalos de confianza para Plástico: ", int_met_Plastico, "\n")
```

```
## Intervalos de confianza para Plástico:  12.97056 15.32944
```

```
medias <- c(mean_Acero, mean_Aluminio, mean_Plastico)

materiales <- c("Acero", "Aluminio", "Plástico")

x_pos <- 1:length(medias)

plot(x_pos, medias, ylim = c(min(c(int_met_Acero, int_met_Aluminio, int_met_Plastico)) - 1,
                             max(c(int_met_Acero, int_met_Aluminio, int_met_Plastico)) + 1),
     xaxt = "n", pch = 16, col = "blue", xlab = "Material", ylab = "Valor Medio",
     main = "Intervalos de confianza por material")

axis(1, at = x_pos, labels = materiales)

arrows(x0 = x_pos, y0 = c(int_met_Acero[1], int_met_Aluminio[1], int_met_Plastico[1]),
       y1 = c(int_met_Acero[2], int_met_Aluminio[2], int_met_Plastico[2]),
       angle = 90, code = 3, length = 0.1, col = "red")
```



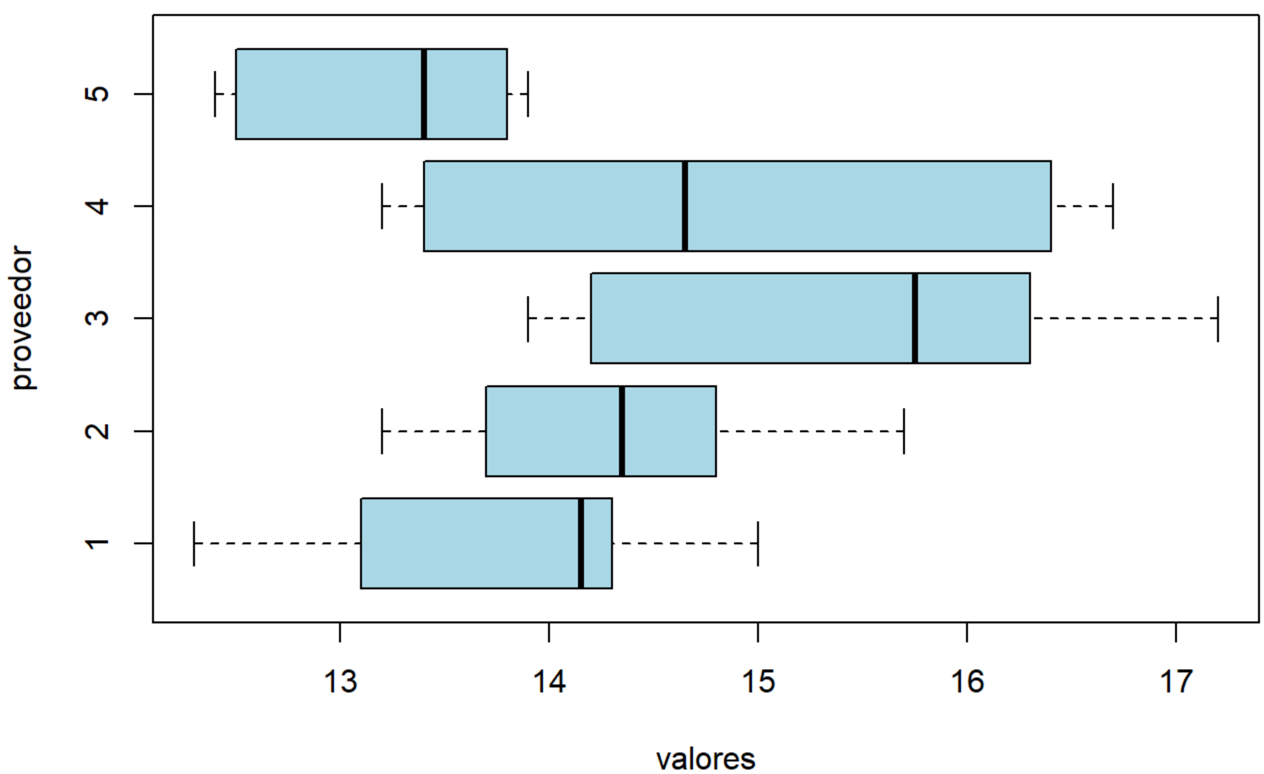
- Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.
- Escribe tus conclusiones parciales.
- **Realiza el ANOVA para un efecto principal**

```
C <- aov(valores ~ proveedor)

summary(C)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## proveedor    4  18.65   4.663   3.642  0.018 *
## Residuals   25  32.00   1.280
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Haz el boxplot de vibración por proveedor. Calcula la media.

```
# Boxplot

boxplot(valores ~ proveedor, col = "lightblue", horizontal = TRUE)
```




```
# Media

mprov = tapply(valores, proveedor, mean)

cat("La media para las vibraciones por proveedor es: \n")
```

```
## La media para las vibraciones por proveedor es:
```

```
mprov
```

```
##        1        2        3        4        5 
## 13.83333 14.35000 15.51667 14.83333 13.23333
```

- Haz los intervalos de confianza de rendimiento por proveedor. Grafícalos.

```
n_Prov1 <- length(valores[proveedor == 1])
mean_Prov1 <- mean(valores[proveedor == 1])
stderr_Prov1 <- sd(valores[proveedor == 1]) / sqrt(n_Prov1)
error_margin_Prov1 <- qt(0.975, df = n_Prov1 - 1) * stderr_Prov1
int_met_Prov1 <- c(mean_Prov1 - error_margin_Prov1, mean_Prov1 + error_margin_Prov1)

n_Prov2 <- length(valores[proveedor == 2])
mean_Prov2 <- mean(valores[proveedor == 2])
stderr_Prov2 <- sd(valores[proveedor == 2]) / sqrt(n_Prov2)
error_margin_Prov2 <- qt(0.975, df = n_Prov2 - 1) * stderr_Prov2
int_met_Prov2 <- c(mean_Prov2 - error_margin_Prov2, mean_Prov2 + error_margin_Prov2)

n_Prov3 <- length(valores[proveedor == 3])
mean_Prov3 <- mean(valores[proveedor == 3])
stderr_Prov3 <- sd(valores[proveedor == 3]) / sqrt(n_Prov3)
error_margin_Prov3 <- qt(0.975, df = n_Prov3 - 1) * stderr_Prov3
int_met_Prov3 <- c(mean_Prov3 - error_margin_Prov3, mean_Prov3 + error_margin_Prov3)

n_Prov4 <- length(valores[proveedor == 4])
mean_Prov4 <- mean(valores[proveedor == 4])
stderr_Prov4 <- sd(valores[proveedor == 4]) / sqrt(n_Prov4)
error_margin_Prov4 <- qt(0.975, df = n_Prov4 - 1) * stderr_Prov4
int_met_Prov4 <- c(mean_Prov4 - error_margin_Prov4, mean_Prov4 + error_margin_Prov4)

n_Prov5 <- length(valores[proveedor == 5])
mean_Prov5 <- mean(valores[proveedor == 5])
stderr_Prov5 <- sd(valores[proveedor == 5]) / sqrt(n_Prov5)
error_margin_Prov5 <- qt(0.975, df = n_Prov5 - 1) * stderr_Prov5
int_met_Prov5 <- c(mean_Prov5 - error_margin_Prov5, mean_Prov5 + error_margin_Prov5)

cat("Intervalos de confianza para Proveedor 1: ", int_met_Prov1, "\n")
```

```
## Intervalos de confianza para Proveedor 1:  12.81443 14.85224
```

```
cat("Intervalos de confianza para Proveedor 2: ", int_met_Prov2, "\n")
```

```
## Intervalos de confianza para Proveedor 2:  13.43814 15.26186
```

```
cat("Intervalos de confianza para Proveedor 3: ", int_met_Prov3, "\n")
```

```
## Intervalos de confianza para Proveedor 3:  14.19685 16.83648
```

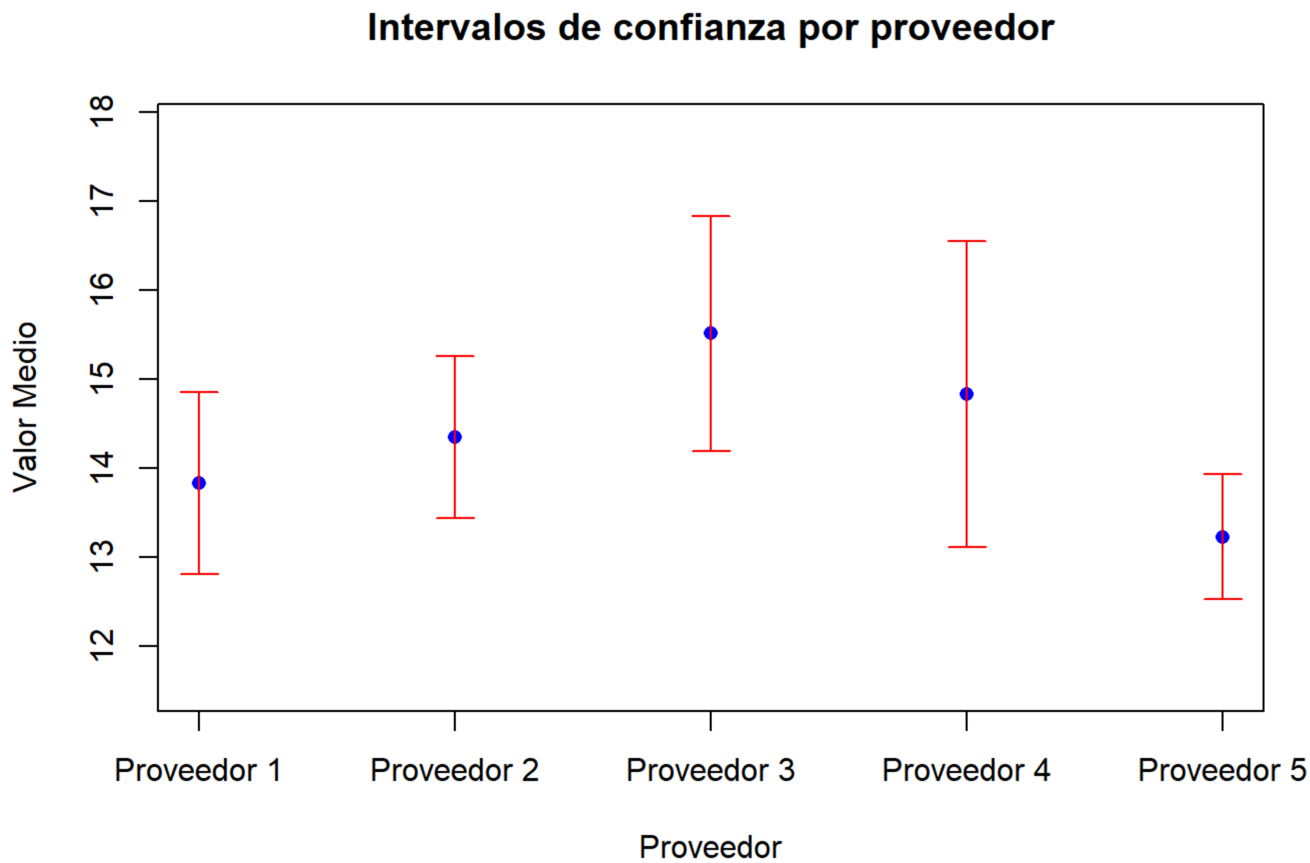
```
cat("Intervalos de confianza para Proveedor 4: ", int_met_Prov4, "\n")
```

```
## Intervalos de confianza para Proveedor 4:  13.11704 16.54962
```

```
cat("Intervalos de confianza para Proveedor 5: ", int_met_Prov5, "\n")
```

```
## Intervalos de confianza para Proveedor 5:  12.53196 13.9347
```

```
medias <- c(mean_Prov1, mean_Prov2, mean_Prov3, mean_Prov4, mean_Prov5)
proveedores <- c("Proveedor 1", "Proveedor 2", "Proveedor 3", "Proveedor 4", "Proveedor 5")
x_pos <- 1:length(medias)

plot(x_pos, medias, ylim = c(min(c(int_met_Prov1, int_met_Prov2, int_met_Prov3, int_met_Prov4, i
nt_met_Prov5)) - 1,
                             max(c(int_met_Prov1, int_met_Prov2, int_met_Prov3, int_met_Prov4, i
nt_met_Prov5)) + 1),
     xaxt = "n", pch = 16, col = "blue", xlab = "Proveedor", ylab = "Valor Medio",
     main = "Intervalos de confianza por proveedor")

axis(1, at = x_pos, labels = proveedores)

arrows(x0 = x_pos, y0 = c(int_met_Prov1[1], int_met_Prov2[1], int_met_Prov3[1], int_met_Prov4
[1], int_met_Prov5[1]),
       y1 = c(int_met_Prov1[2], int_met_Prov2[2], int_met_Prov3[2], int_met_Prov4[2], int_met_Pr
ov5[2]),
       angle = 90, code = 3, length = 0.1, col = "red")
```

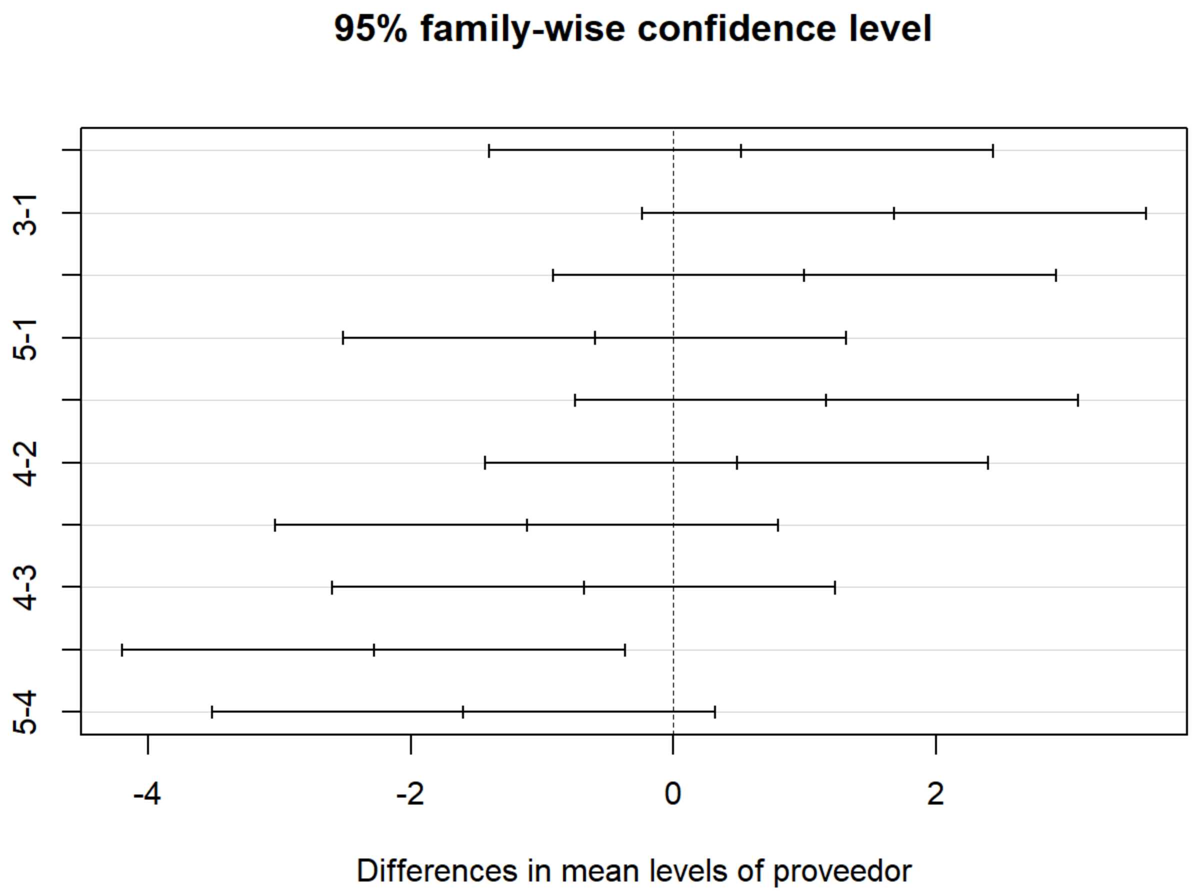


- Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey.

```
I = TukeyHSD(C)
I
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = valores ~ proveedor)
##
## $proveedor
##           diff        lwr        upr     p adj
## 2-1  0.5166667 -1.4017895  2.4351228 0.9308488
## 3-1  1.6833333 -0.2351228  3.6017895 0.1055788
## 4-1  1.0000000 -0.9184562  2.9184562 0.5530518
## 5-1 -0.6000000 -2.5184562  1.3184562 0.8870654
## 3-2  1.1666667 -0.7517895  3.0851228 0.4034099
## 4-2  0.4833333 -1.4351228  2.4017895 0.9449196
## 5-2 -1.1166667 -3.0351228  0.8017895 0.4465140
## 4-3 -0.6833333 -2.6017895  1.2351228 0.8315109
## 5-3 -2.2833333 -4.2017895 -0.3648772 0.0140278
## 5-4 -1.6000000 -3.5184562  0.3184562 0.1352033
```

```
plot(I)
```



- Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.
- Escribe tus conclusiones parciales.
- **Comprueba la validez del modelo. Comprueba:**
    - Normalidad

```
shapiro.test(resid(C))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(C)
## W = 0.94938, p-value = 0.1627
```

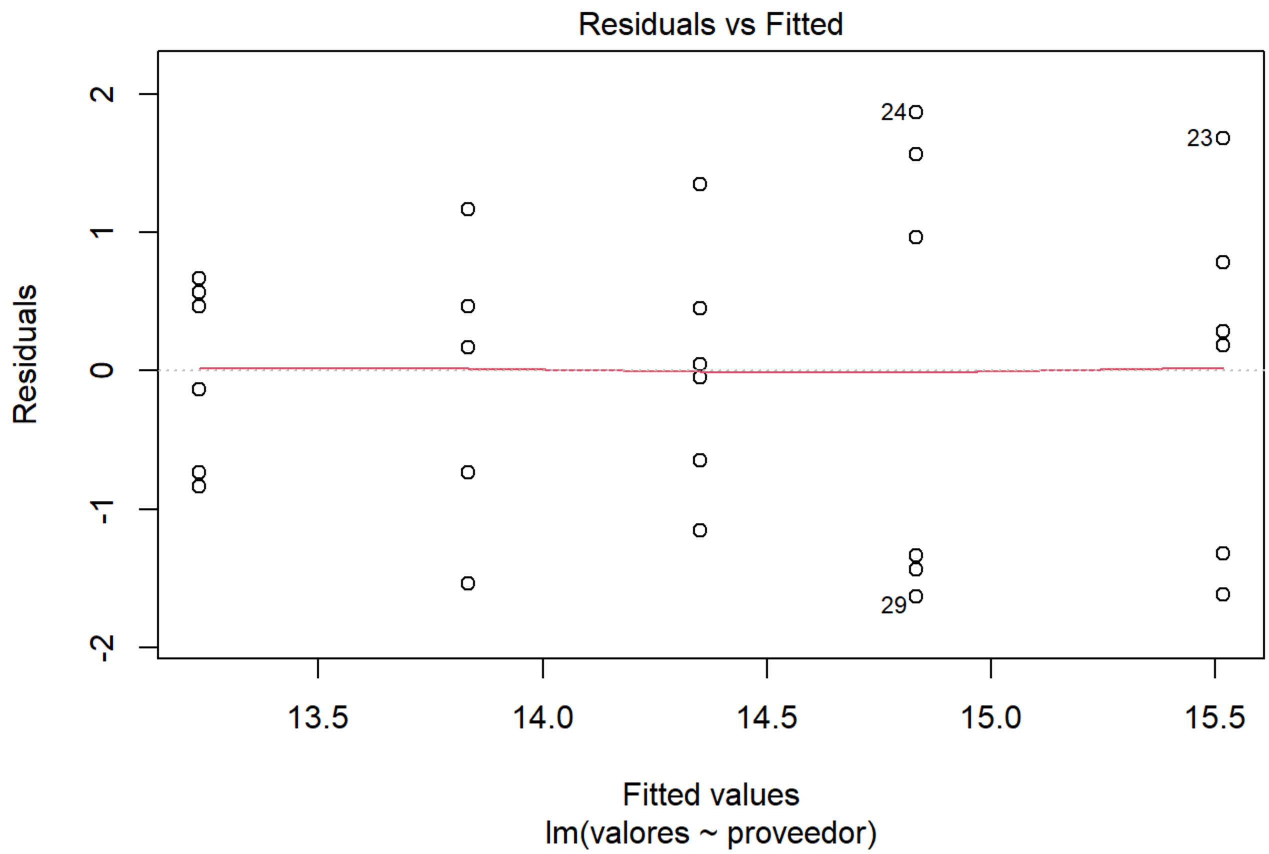
- Homocedasticidad

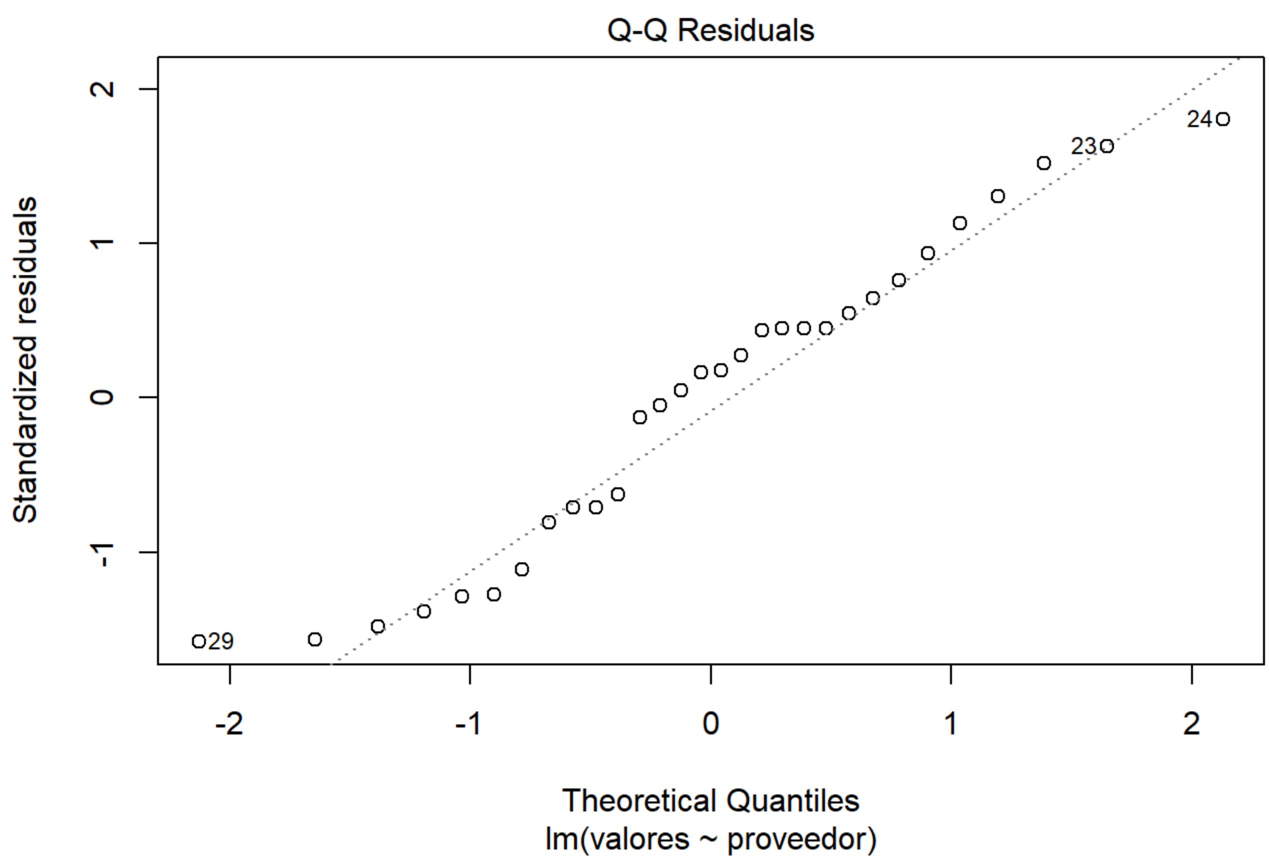
```
bartlett.test(valores ~ proveedor)
```

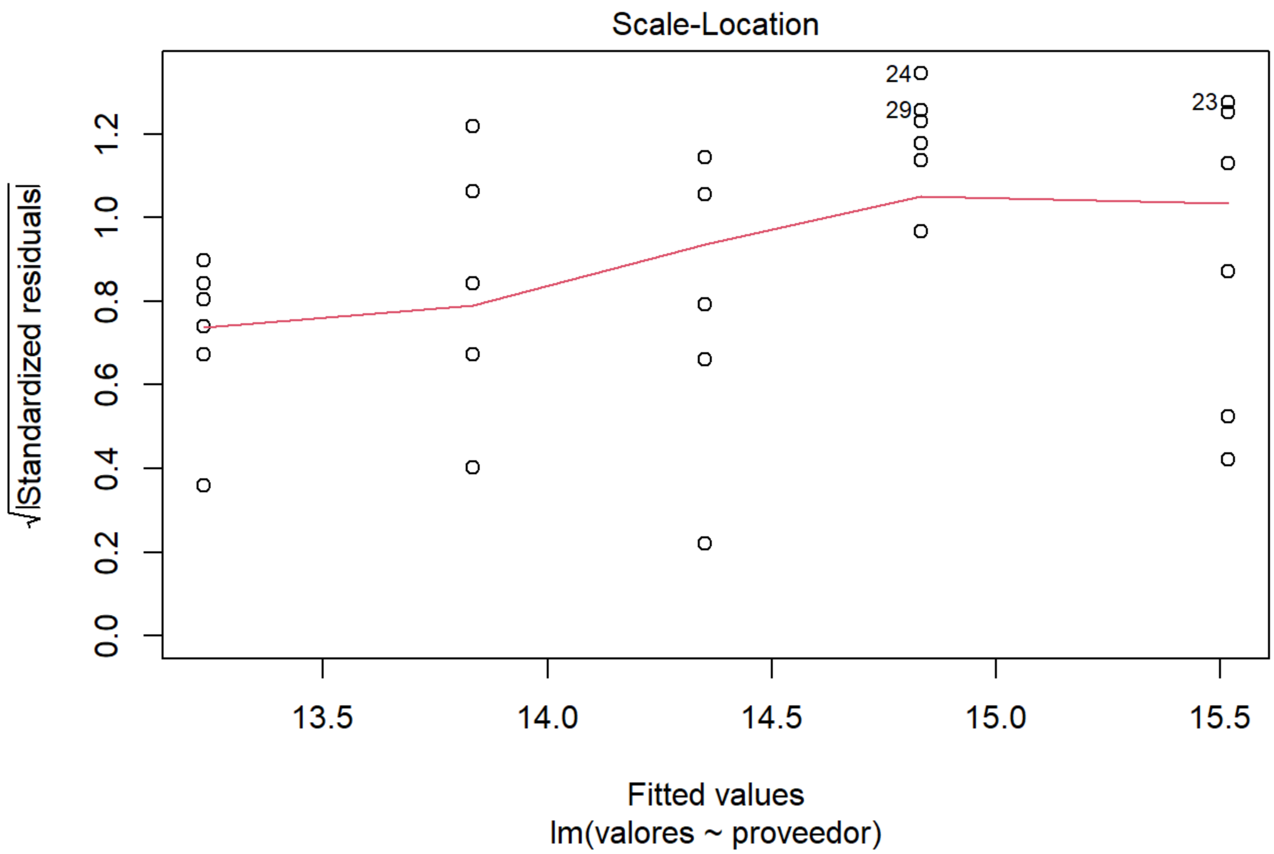
```
##
##  Bartlett test of homogeneity of variances
##
## data:  valores by proveedor
## Bartlett's K-squared = 4.3451, df = 4, p-value = 0.3613
```

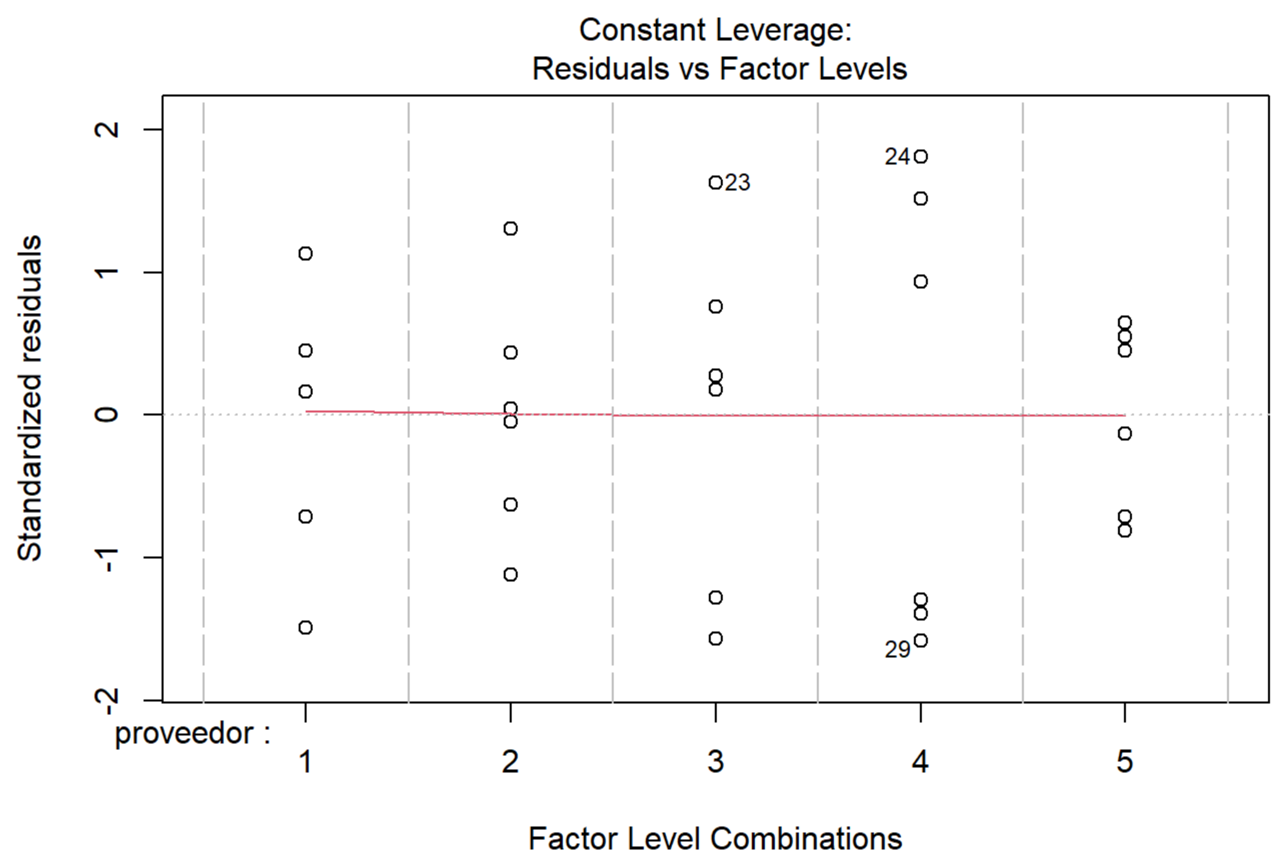
- Relación lineal entre las variables (coeficiente de determinación)

```
plot(lm(valores ~ proveedor))
```


Residuals vs Fitted
Residuals
Fitted values
lm(valores ~ proveedor)
24
23
29
2
1
0
-1
-2
13.5
14.0
14.5
15.0
15.5


Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(valores ~ proveedor)
24
23
29
2
1
0
-1
-2
-1
0
1
2


Scale-Location
√|Standardized residuals|
0.0
0.2
0.4
0.6
0.8
1.0
1.2
24
29
23
13.5
14.0
14.5
15.0
15.5
Fitted values
lm(valores ~ proveedor)


Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
-2
-1
0
1
2
24
23
29
proveedor :
1
2
3
4
5
Factor Level Combinations

```
CD = 150/(150 + 76)
```

- **Concluye en el contexto del problema.**