9. ANOVA

Juan Bernal

2024-08-27

Problema 1. “El rendimiento”

Resuelve las dos partes del problema “El rendimiento”. Se encuentra en los apoyos de clase de “ANOVA”. Para ello se te recomienda que sigas los siguientes pasos.

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:

```
chicos <- data.frame(
  Metodo_1 = c(10, 7, 9, 9, 9, 10),
  Metodo_2 = c(5, 7, 6, 6, 8, 4),
  Metodo_3 = c(2, 6, 3, 5, 5, 3)
)
chicos
```

```
##   Metodo_1 Metodo_2 Metodo_3
## 1       10        5        2
## 2        7        7        6
## 3        9        6        3
## 4        9        6        5
## 5        9        8        5
## 6       10        4        3
```

```
chicas = data.frame(
  Metodo_1 = c(9, 7, 8, 8, 10, 6),
  Metodo_2 = c(8, 3, 5, 6, 7, 7),
  Metodo_3 = c(2, 6, 2, 1, 4, 3)
)
chicas
```

```
##   Metodo_1 Metodo_2 Metodo_3
## 1        9        8        2
## 2        7        3        6
## 3        8        5        2
## 4        8        6        1
## 5       10        7        4
## 6        6        7        3
```

1. Análisis exploratorio. Calcula la media para el rendimiento por

método de enseñanza.

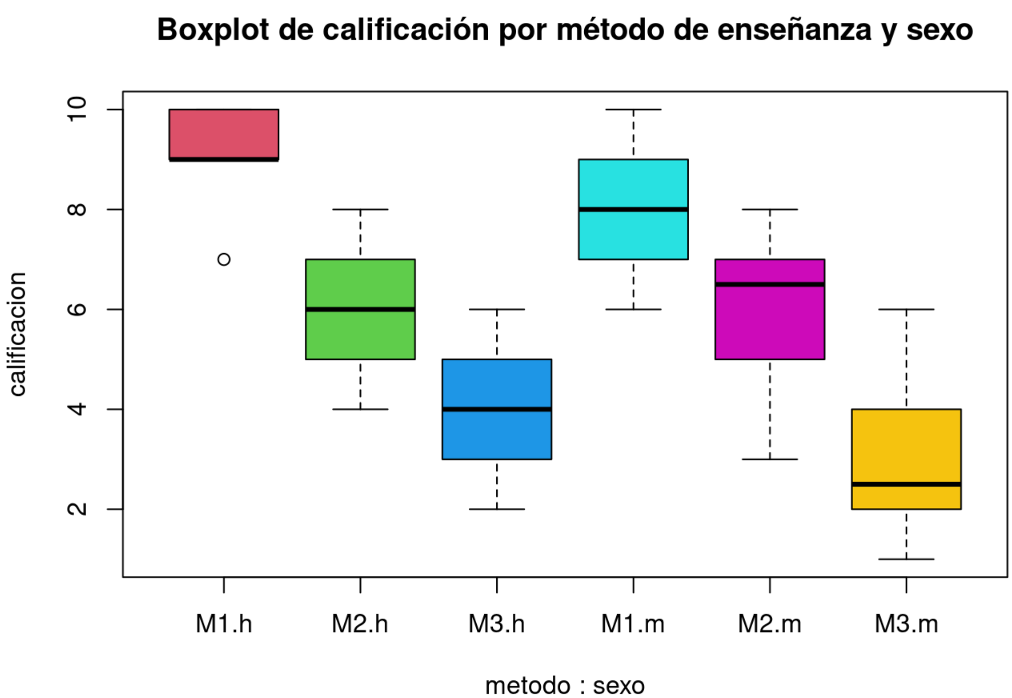
*Consulta el código en R en los apoyos de clase de “ANOVA”

```
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)
data = data.frame(calificacion, metodo,sexo)

tapply(calificacion,metodo,mean)
```

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

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

```
boxplot(calificacion~metodo:sexo, data, col=2:8,
        main = "Boxplot de calificación por método de enseñanza y sexo")
```



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

De las boxplots y medias obtenidas, podemos observar que el método de enseñanza 1 conserva un mejor promedio de rendimiento que los otros métodos de enseñanza, además de que el método 3 se posiciona como el método más bajo.

*Escribe tus conclusiones parciales

Del análisis de interpretación hecho anteriormente, concluimos que el mejor método de enseñanza es el #1, pues conserva un promedio mayor a los otros métodos, y el #3 es el peor método, pues conserva un promedio de calificación reprobatoria.

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

La primera hipótesis es:

*$H_0: \tau_i = 0$. El método de enseñanza influye en la calificación.

*$H_i: _i $. El método de enseñanza no influye en la calificación.

La segunda hipótesis es:

*$H_0: \alpha_j = 0$. El sexo del estudiante influye en la calificación.

*$H_i: _j $. El sexo del estudiante no influye en la calificación.

La tercera hipótesis es:

*$H_0: \tau_i \alpha_j = 0$. Existe una interacción entre el método y el sexo del estudiante.

*$H_i: \exists \tau_i \alpha_j \neq 0$. No existe una interacción entre el método y el sexo del estudiante.

3. Realiza el ANOVA para dos niveles con interacción:

*Consulta el código en R en los apoyos de clase de "ANOVA":

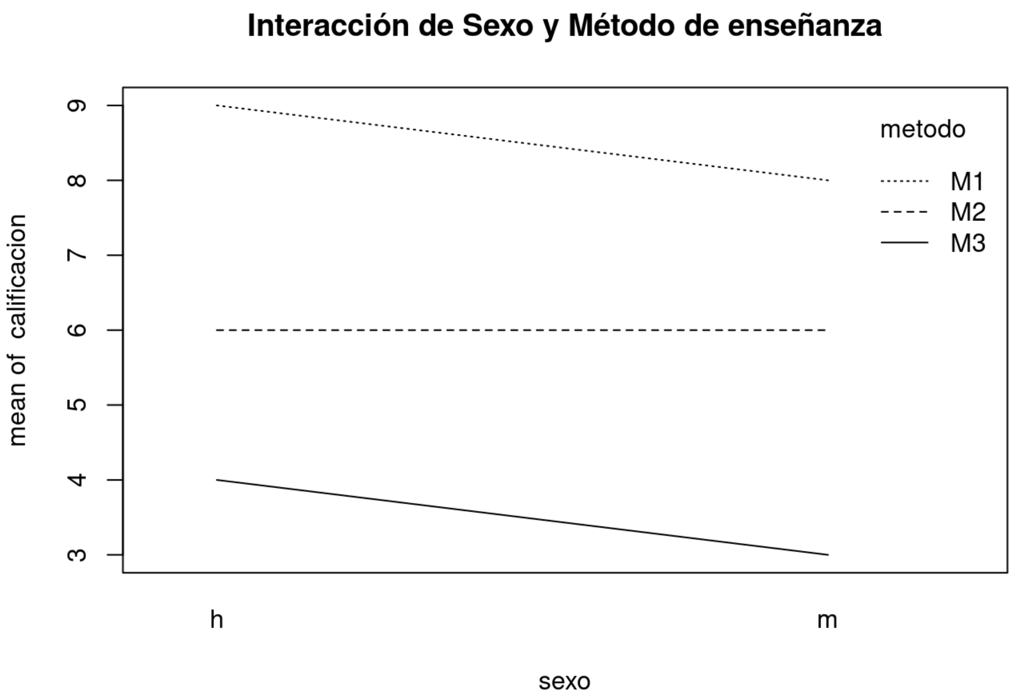
```
A<-aov(calificacion~metodo*sexo)
summary(A)
```

```
##             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
```

*Haz la gráfica de interacción de dos factores en ANOVA

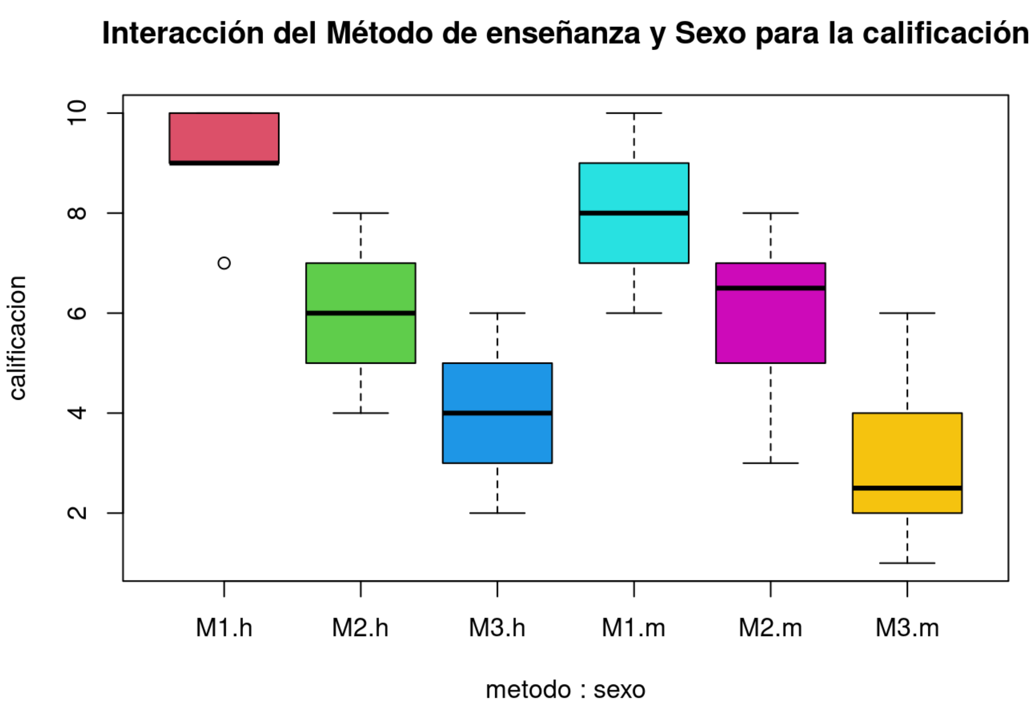
```
interaction.plot(metodo,sexo,calificacion,
                 main = "Interacción de Método de enseñanza y Sexo")
```


```
interaction.plot(sexo, metodo, calificacion,
                 main = "Interacción de Sexo y Método de enseñanza")
```



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

```
boxplot(calificacion~metodo:sexo, data, col=2:8,
        main = "Interacción del Método de enseñanza y Sexo para la calificación")
```



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

Con los resultados del ANOVA se obtuvo que los métodos de enseñanza si influyen en la calificación del alumno, mas no hay interacción entre el método y el sexo, y el sexo del estudiante no influye en su calificación. Todo esto se obtuvo con base en pruebas de hipótesis del 95% de confianza con los valores P obtenidos del análisis de varianza con interacción entre método y sexo. Para reforzar esta idea de que el método influye en la calificación y el sexo no, las gráficas de interacción muestran que no existe una gran diferencia o brecha de calificación entre el sexo de los estudiantes, si no que dicha brecha se da únicamente entre los métodos de enseñanza.

*Escribe tus conclusiones parciales

El factor significativo que influye en la calificación de los estudiantes es el método de enseñanza.

4. Realiza el ANOVA para dos niveles sin interacción.

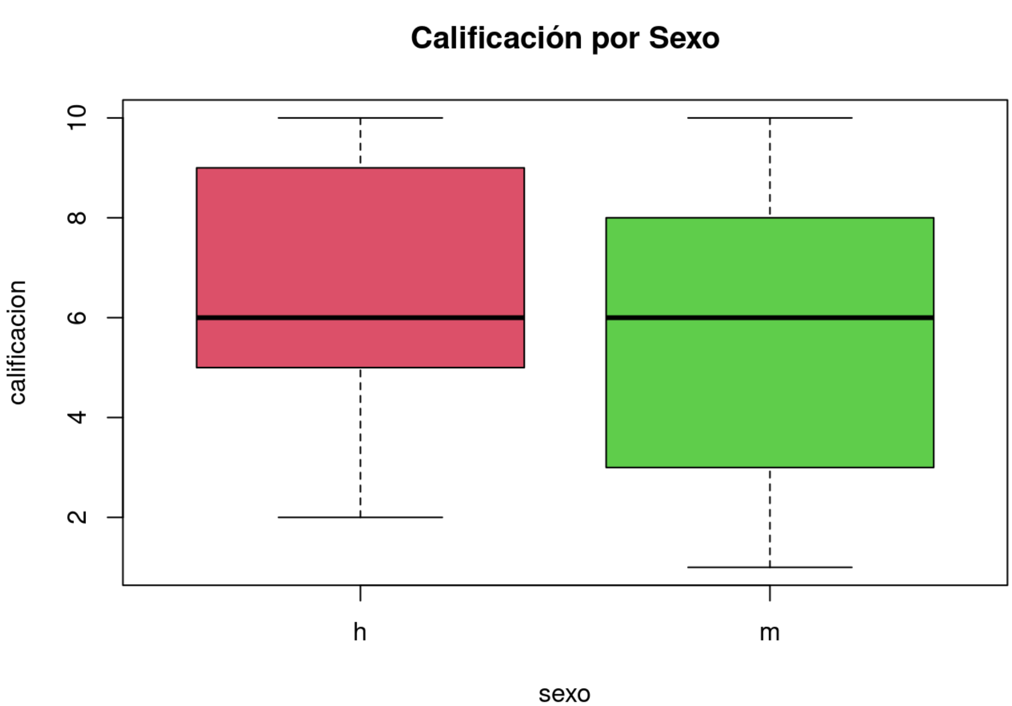
*Consulta el código de R en los apoyos de clase de "ANOVA"

```
A<-aov(calificacion~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
```

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

```
boxplot(calificacion~sexo, data, col=2:8,
        main = "Calificación por Sexo")
```




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

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

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

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

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

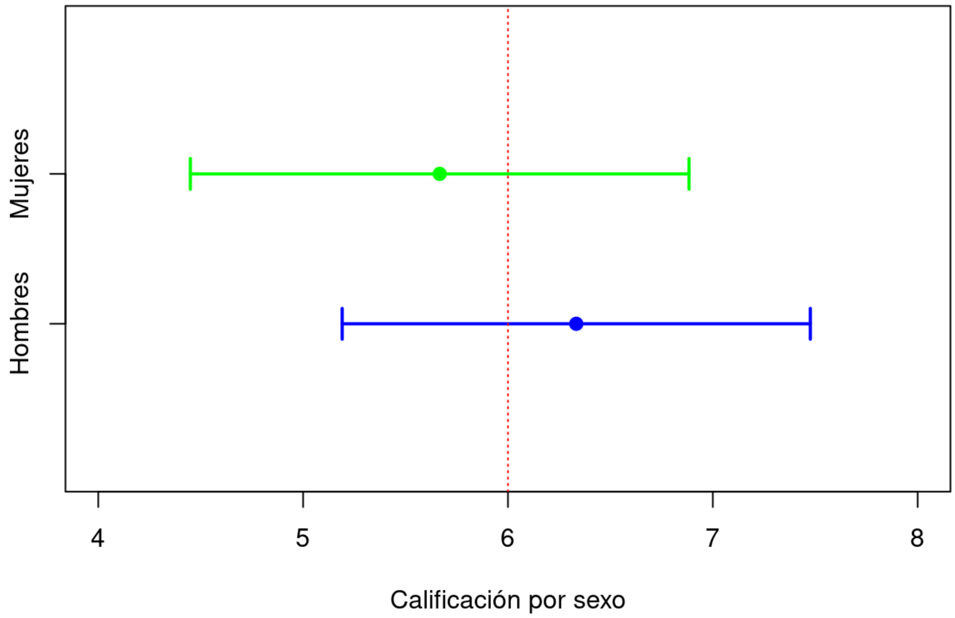
```
## M1:h M1:m M2:h M2:m M3:h M3:m 
##    9    8    6    6    4    3
```

```
cat("La media de la calificación es ",mean(calificacion))
```

```
## La media de la calificación es  6
```

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

```
n = 18 # Número de estudiantes por sexo
a = 0.05 # Nivel de confianza

de_h = sd(calificacion[1:18])
de_m = sd(calificacion[19:36])
med_h = mean(calificacion[1:18])
med_m = mean(calificacion[19:36])

E_h = abs(qnorm(a/2))*de_h/sqrt(n) # Error del 95% confianza
E_m = abs(qnorm(a/2))*de_m/sqrt(n)

A_h = med_h - E_h # Cota inferior
A_m = med_m - E_m # Cota inferior

B_h = med_h + E_h # Cota superior
B_m = med_m + E_m # Cota superior

plot(0, ylim = c(0,2+1), xlim = c(4,8), yaxt = "n", ylab = "", xlab = "Calificación por sexo")
axis(2, at = c(1,2), labels = c("Hombres", "Mujeres"))

arrows(A_h, 1, B_h, 1, angle=90, code=3, length = 0.1, lwd = 2, col = 'blue')
arrows(A_m, 2, B_m, 2, angle=90, code=3, length = 0.1, lwd = 2, col = 'green')

points(med_h, 1, pch = 19, cex = 1.1, col = 'blue')
points(med_m, 2, pch = 19, cex = 1.1, col = 'green')
abline(v = mean(calificacion), lty = 3, col = 'red')
```



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

En el ANOVA sin interacción de las variables método y sexo observamos nuevamente que el método es significativo para la calificación del alumno, mas no el sexo, y las gráficas de boxplot e intervalos de confianza nos ayudan a confirmar esta idea, pues ambos, hombres y mujeres, se encuentran dentro de la media de calificación general y existe apenas 1 punto de diferencia en promedio entre ambos.

*Escribe tus conclusiones parciales

Descartamos la variable “Sexo” como factor de influencia en la calificación de los alumnos.

5. Realiza el ANOVA para un efecto principal

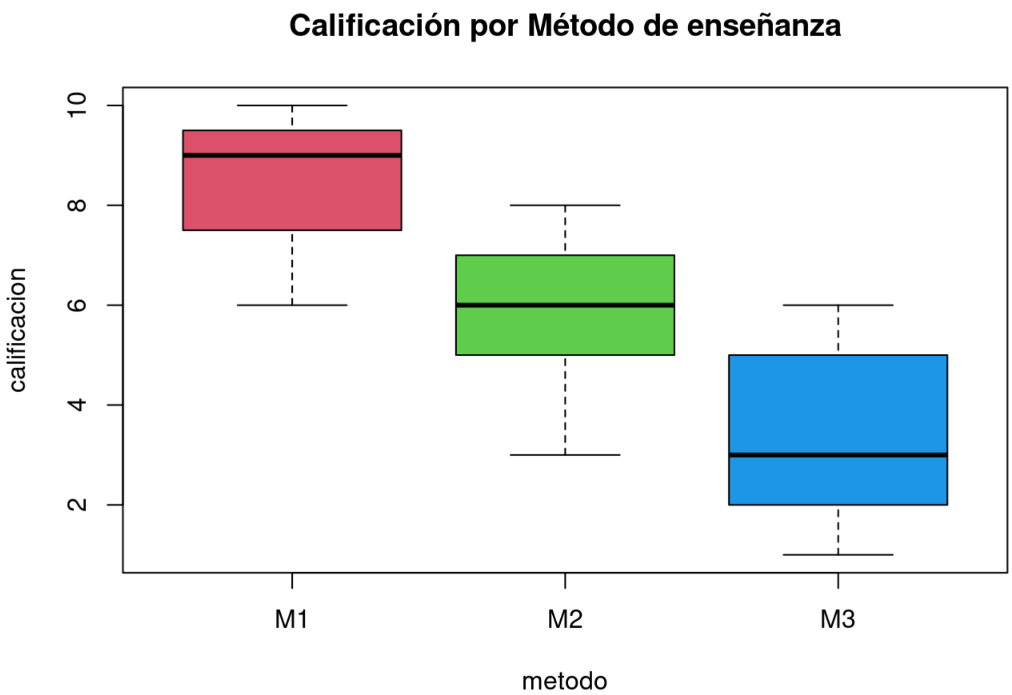
*Consulta el código de R en los apoyos de clase de “ANOVA”

```
A<-aov(calificacion~metodo)
summary(A)
```

```
##             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
```

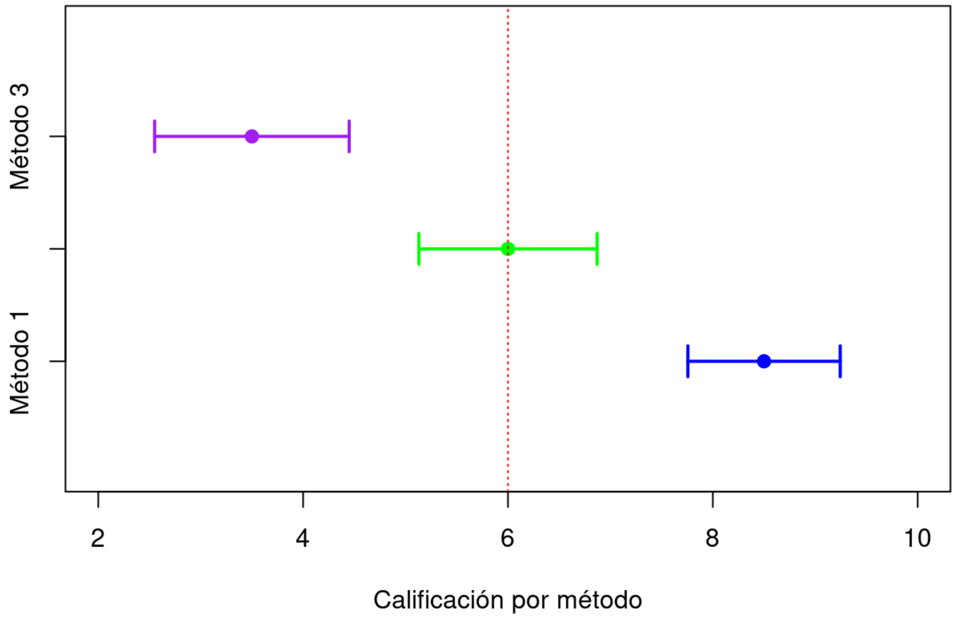
*Haz el boxplot de rendimiento por método de enseñanza. Calcula la media.

```
boxplot(calificacion~metodo, data, col=2:8,
        main = "Calificación por Método de enseñanza")
```



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

```
n = 12 # Número de estudiantes por sexo
a = 0.05 # Nivel de confianza

de1 = sd(calificacion[c(1:6,19:24)])
de2 = sd(calificacion[c(7:12,25:30)])
de3 = sd(calificacion[c(13:18,31:36)])

med1 = mean(calificacion[c(1:6,19:24)])
med2 = mean(calificacion[c(7:12,25:30)])
med3 = mean(calificacion[c(13:18,31:36)])

E1 = abs(qnorm(a/2))*de1/sqrt(n) # Error del 95% confianza
E2 = abs(qnorm(a/2))*de2/sqrt(n)
E3 = abs(qnorm(a/2))*de3/sqrt(n)

A1 = med1 - E1 # Cota inferior
A2 = med2 - E2 # Cota inferior
A3 = med3 - E3 # Cota inferior

B1 = med1 + E1 # Cota superior
B2 = med2 + E2 # Cota superior
B3 = med3 + E3 # Cota superior

plot(0, ylim = c(0,3+1), xlim = c(2,10), yaxt = "n", ylab = "", xlab = "Calificación por método")
axis(2, at = c(1,2,3), labels = c("Método 1", "Método 2",'Método 3'))

arrows(A1, 1, B1, 1, angle=90, code=3, length = 0.1, lwd = 2, col = 'blue')
arrows(A2, 2, B2, 2, angle=90, code=3, length = 0.1, lwd = 2, col = 'green')
arrows(A3, 3, B3, 3, angle=90, code=3, length = 0.1, lwd = 2, col = 'purple')

points(med1, 1, pch = 19, cex = 1.1, col = 'blue')
points(med2, 2, pch = 19, cex = 1.1, col = 'green')
points(med3, 3, pch = 19, cex = 1.1, col = 'purple')
abline(v = mean(calificacion), lty = 3, col = 'red')
```

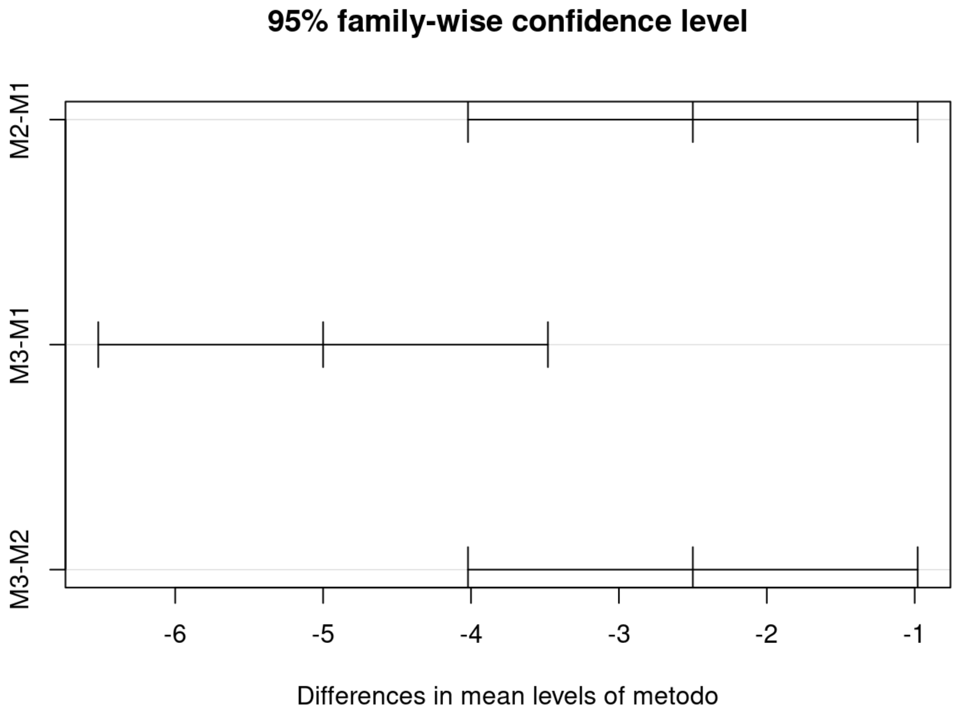


*Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey.

```
Tu=TukeyHSD(A)
Tu
```

```
##   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(TukeyHSD(A))
```



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

De análisis anteriores obtuvimos que la variable “método” era el factor principal de influencia en la calificación de los alumnos. Observando el boxplot y analizando los intervalos de confianza, podemos notar que existe mucha diferencia entre los métodos de enseñanza, tanto así que ni siquiera se interceptan como los sexos. Entre el método 1 y método 3 existen 5 puntos de promedio de diferencia, entre método 1 y método 2 son 2.5 puntos de diferencia y lo mismo entre los métodos 2 y 3.

*Escribe tus conclusiones parciales

Las diferencias antes mencionadas, además de los análisis de varianza ya presentados, dan prueba suficiente de que el método de enseñanza presenta un gran impacto en la calificación de los estudiantes.

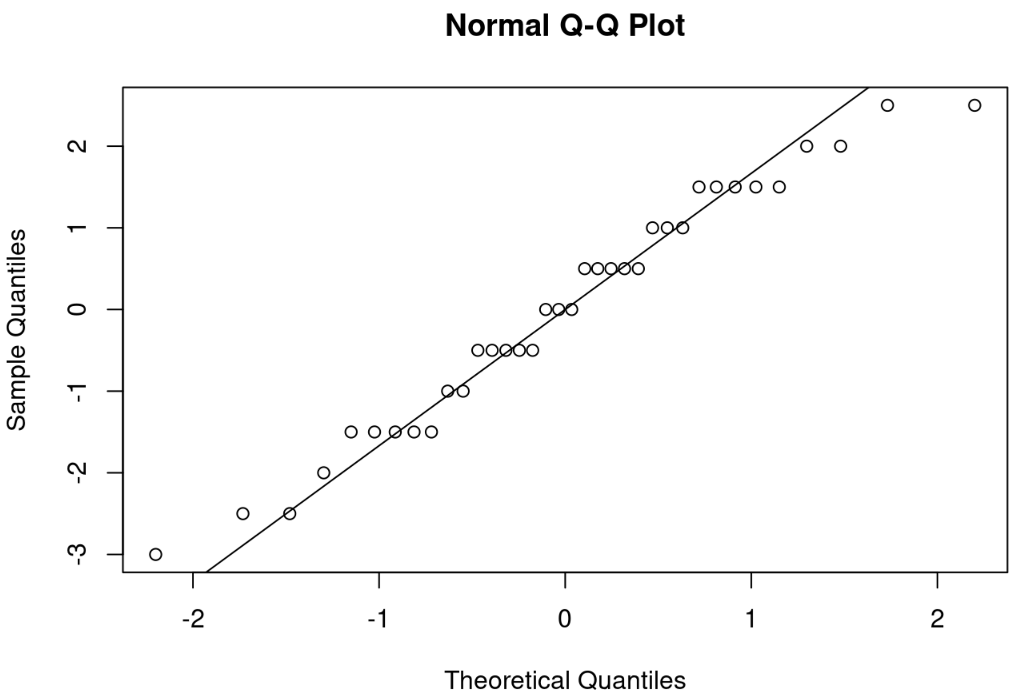
Es así que el modelo finalmente queda como:

$Y_{ijk} = \mu + \tau_i + \varepsilon_{ijk}$

6. Comprueba la validez del modelo. Comprueba:

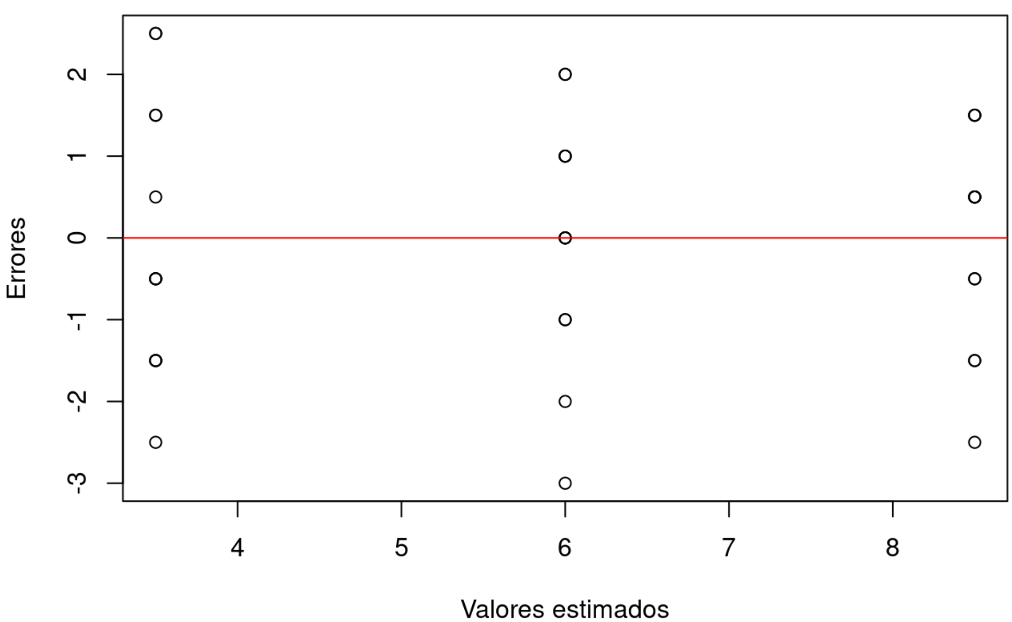
*Normalidad

```
anova = aov(calificacion~metodo)
residuos=anova$residuals
qqnorm(residuos)
qqline(residuos)
```



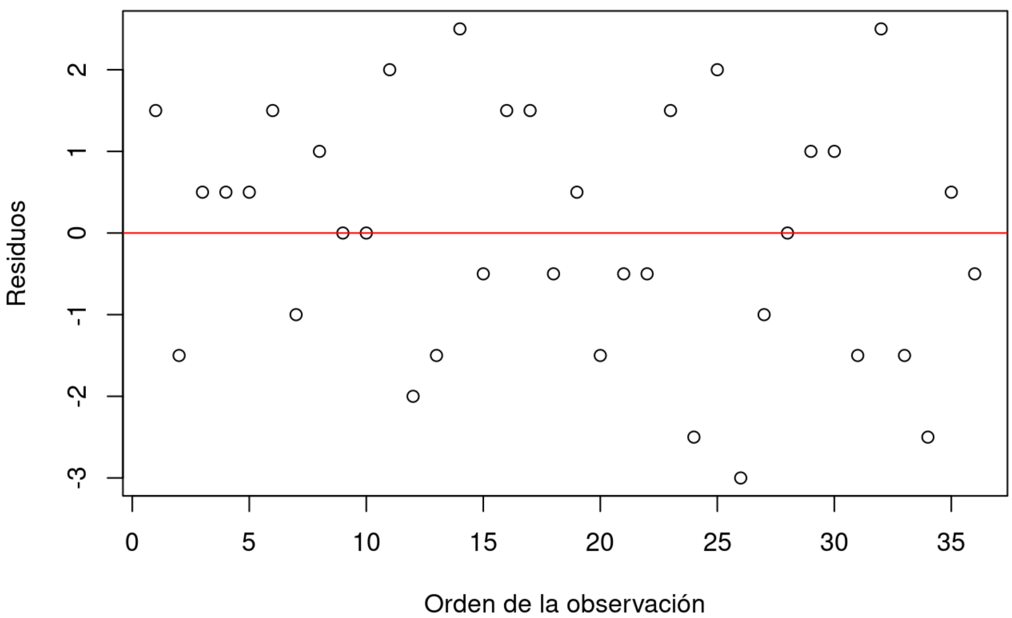
*Homocedasticidad

```
plot(anova$fitted.values,anova$residuals,ylab="Errores",xlab="Valores estimados")
abline(h=0,col="red")
```



*Independencia

```
n = tapply(calificacion, metodo:sexo, length)
plot(c(1:sum(n)),anova$residuals,xlab="Orden de la observación",ylab="Residuos")
abline(h=0,col="red")
```



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

```
CD = 1-(76/(150+76))
CD
```

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

7. Concluye en el contexto del problema.

Se observó que los 3 métodos producen un efecto diferente en el rendimiento de los alumnos. El efecto del Método 3 es un método deficiente, puesto que se disminuye su rendimiento con respecto a la media general, el Método 2 no tiene efecto, es un método que no modifica el rendimiento de los estudiantes, y el Método 1 incrementa su rendimiento con respecto a la media general por lo que resulta ser el mejor método de enseñanza. El modelo explica el 66.37% de la variación. Por lo tanto, el Método de enseñanza es un factor determinante en el rendimiento de los estudiantes (puesto que es el único que fue significativo en el modelo), sin embargo, es posible que haya otros factores que expliquen el resto del porcentaje de variación (32.73%) y que en este modelo se le atribuye a la aleatoriedad (al error).

El número de datos en cada tratamiento fue igual por lo que es un diseño equilibrado que es robusto a heterocedasticidad. De acuerdo al análisis de los gráficos Q-Q y de los residuos vs. el valor esperado (ajustado), los datos aparentemente cumplen con normalidad e independencia. También los errores tienen una media cero y variación constante.

Problema 2. “Vibración de Motores”

Repite los pasos anteriores para el problema “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.

```
vibracion = c(13.1,13.2,15,14.8,14,14.3,16.3,15.8,15.7,16.4,17.2,16.7,13.7,14.3,13.9,14.3,12.4,12.3,15.7,15.8,13.7,14.2,14.
4,13.9,13.5,12.5,13.4,13.8,13.2,13.1)
material = c(rep("Acero",2),rep("Aluminio",2),rep("Plastico",2),rep("Acero",2),rep("Aluminio",2),rep("Plastico",2),rep("Acer
o",2),rep("Aluminio",2),rep("Plastico",2),rep("Acero",2),rep("Aluminio",2),rep("Plastico",2),rep("Acero",2),rep("Aluminio",
2),rep("Plastico",2))
prove = c(rep("1",6), rep("2",6), rep("3",6), rep("4",6), rep("5",6))
material = factor(material)
prove = factor(prove)
data2 = data.frame(vibracion, prove,material)
head(data2)
```

```
##   vibracion prove material
## 1      13.1     1    Acero
## 2      13.2     1    Acero
## 3      15.0     1 Aluminio
## 4      14.8     1 Aluminio
## 5      14.0     1 Plastico
## 6      14.3     1 Plastico
```

1. Análisis exploratorio. Calcula la media para la cantidad por material.

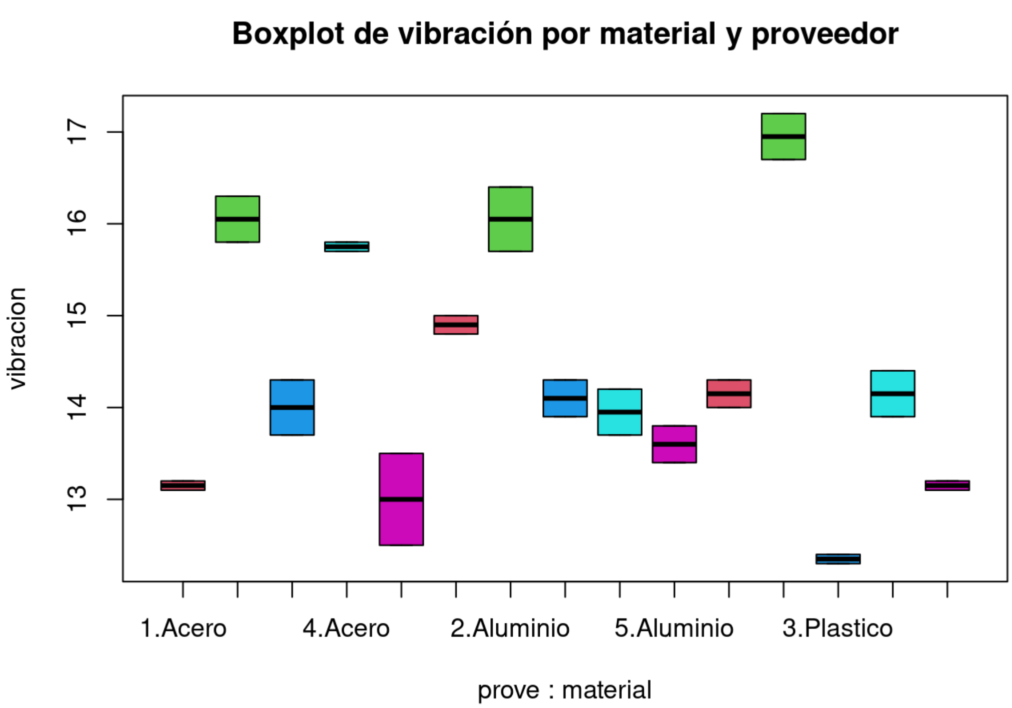
*Consulta el código en R en los apoyos de clase de “ANOVA”

```
tapply(vibracion, material, mean)
```

```
##    Acero Aluminio Plastico
##    14.39    14.52    14.15
```

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

```
boxplot(vibracion~prove:material, data2, col=2:6,
        main = "Boxplot de vibración por material y proveedor")
```



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

Como podemos observar por el promedio de vibraciones por material y el boxplot de las vibraciones por material de cada proveedor, las vibraciones por material son casi iguales sin importar el proveedor.

*Escribe tus conclusiones parciales

Por el momento, tenemos como hipótesis que la media de vibraciones por material es igual para todos los materiales, es decir, el factor "Material" no influye en las vibraciones de los motores.

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

La primera hipótesis es:

*$H_0 : \tau_i = 0$. El proveedor influye en la vibración de los motores.

*$H_i: _i $. El proveedor no influye en la vibración de los motores.

La segunda hipótesis es:

*$H_0 : \alpha_j = 0$. El tipo de material influye en la vibración de los motores.

*$H_i: _j $. El tipo de material no influye en la vibración de los motores.

La tercera hipótesis es:

*$H_0 : \tau_i \alpha_j = 0$. Existe una interacción entre el material y el proveedor en cuanto a las vibraciones de los motores.

*$H_i : \exists \tau_i \alpha_j \neq 0$. No existe una interacción entre el material y el proveedor en cuanto a las vibraciones de los motores.

3. Realiza el ANOVA para dos niveles con interacción:

*Consulta el código en R en los apoyos de clase de "ANOVA":

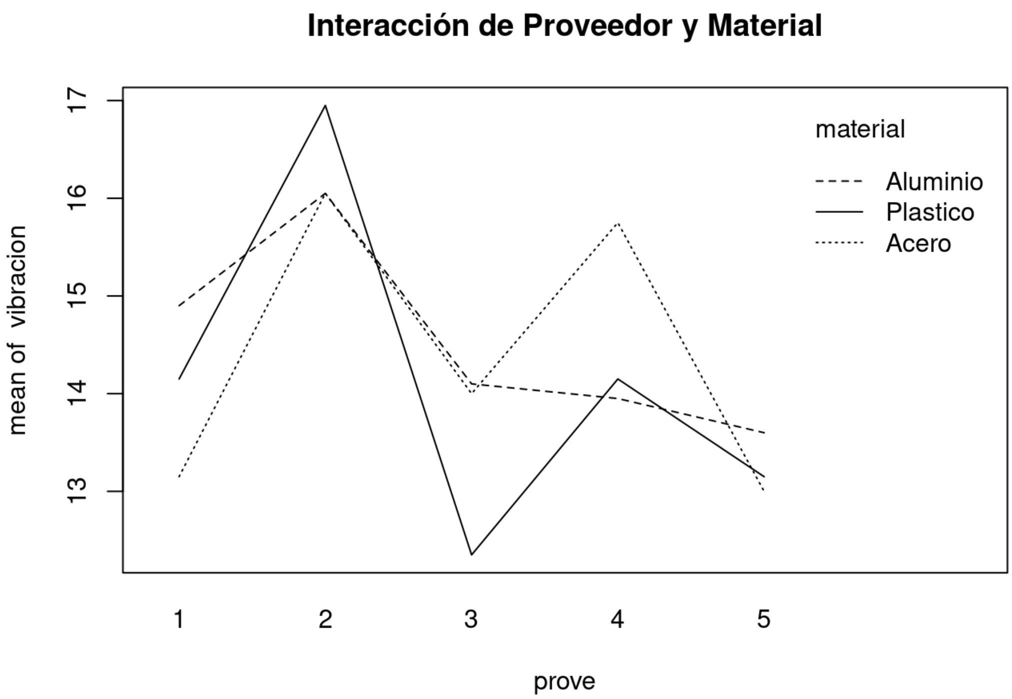
anova = aov(variable~factores) summary(anova)

```
anova = aov(vibracion~prove*material)
summary(anova)
```

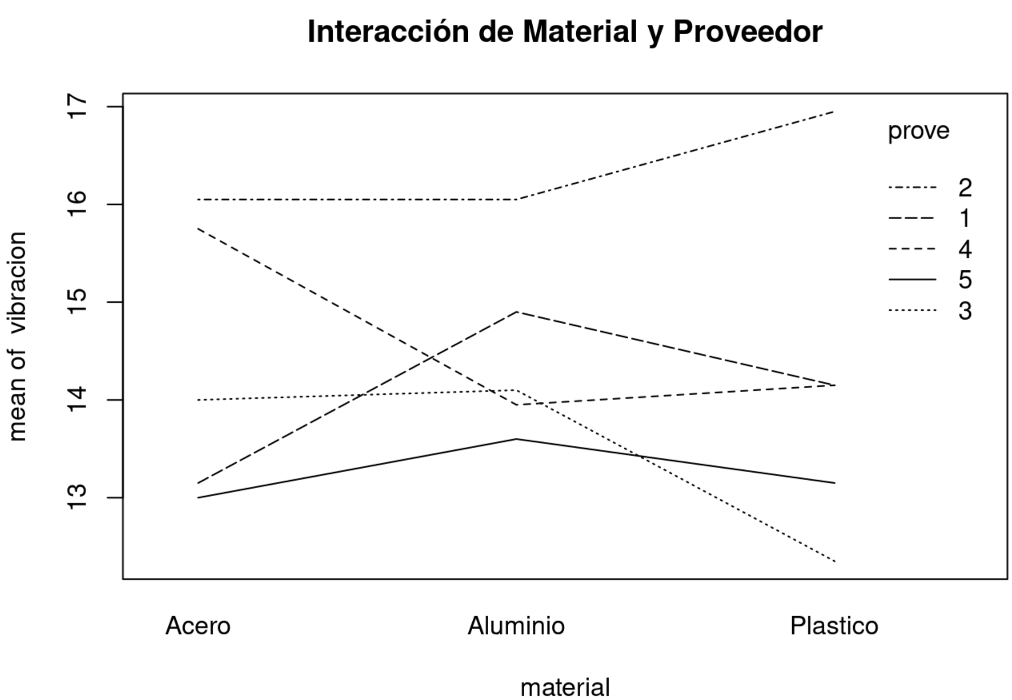
```
##                Df Sum Sq Mean Sq F value   Pr(>F)    
## prove           4  36.67   9.169  82.353 5.07e-10 ***
## material        2   0.70   0.352   3.165   0.0713 .  
## prove:material  8  11.61   1.451  13.030 1.76e-05 ***
## Residuals      15   1.67   0.111                     
## ---
## Signif.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

*Haz la gráfica de interacción de dos factores en ANOVA

```
interaction.plot(prove,material,vibracion,
                 main = "Interacción de Proveedor y Material")
```



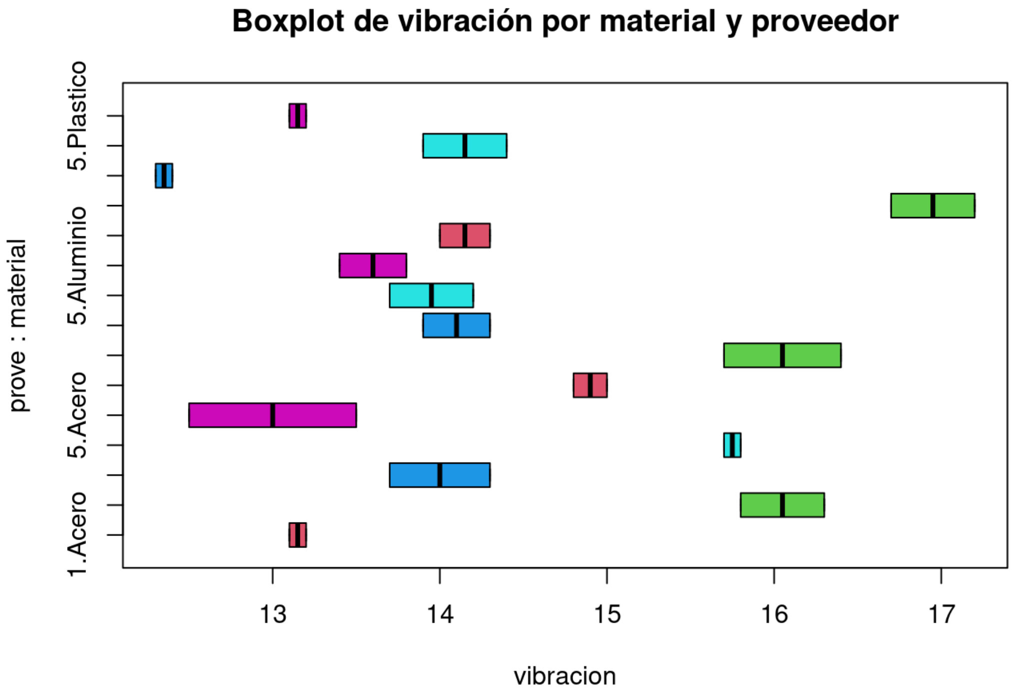

```
interaction.plot(material, prove,vibracion,
                 main = "Interacción de Material y Proveedor")
```



*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(vibracion~prove:material, data2, col=2:6,
        main = "Boxplot de vibración por material y proveedor", horizontal = TRUE)
```



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

De las gráficas y resultados de ANOVA podemos observar que, efectivamente, con un 95% de confianza, el material no tiene ningún efecto sobre las vibraciones de los motores, pero, existe también una interacción entre el proveedor y material que sí afecta a las vibraciones.

*Escribe tus conclusiones parciales

Dado que existe una interacción entre las variables “Material” y “Proveedor”, no necesitamos realizar más análisis de varianza, solo debemos seguir trabajando con el modelo actual, que sería:

$Y_{ijk} = \mu + \tau_i + \alpha_j + \tau_i\alpha_j + \varepsilon_{ijk}$

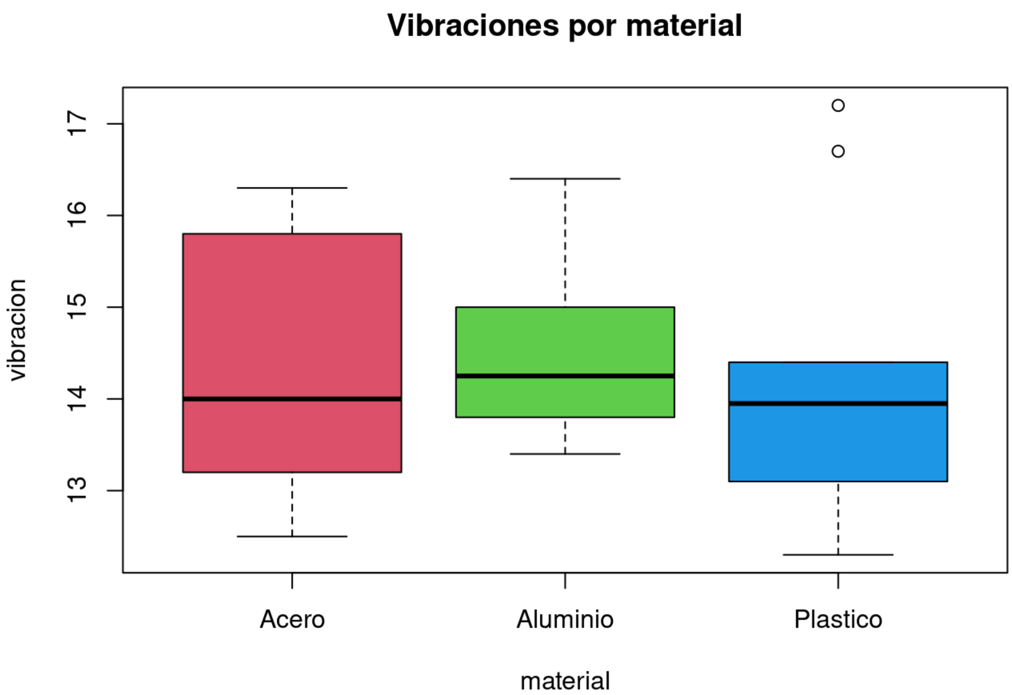
4. Realiza el ANOVA para dos niveles sin interacción.

*Consulta el código de R en los apoyos de clase de “ANOVA”

```
# No es necesario, pues sí existe interacción entre las variables.
```

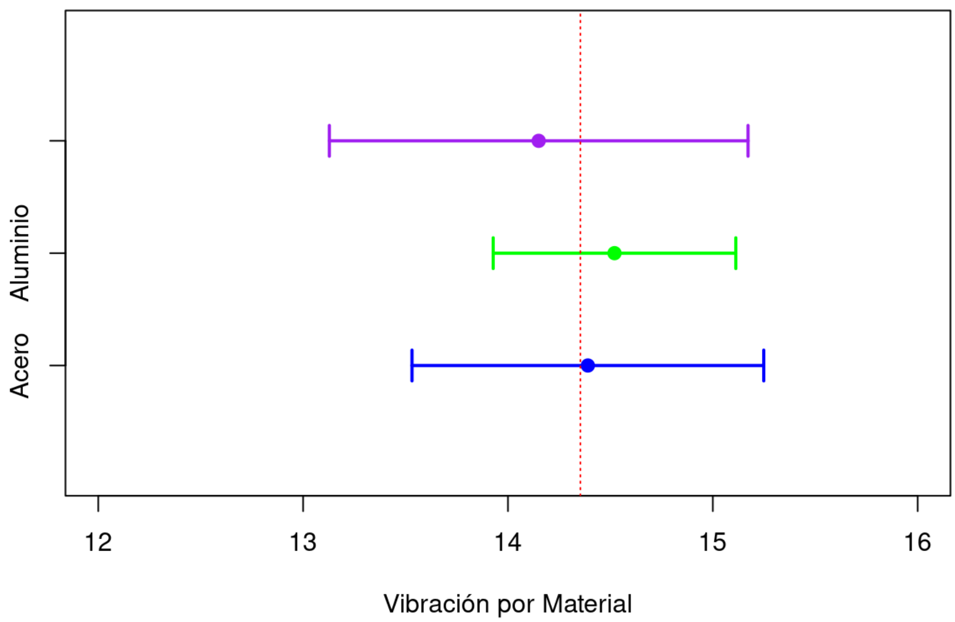
*Haz el boxplot de rendimiento por sexo. Calcula la media para el rendimiento por sexo y método.

```
boxplot(vibracion~material, data2, col=2:8,
        main = "Vibraciones por material")
```



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

```
n = 10 # Número de estudiantes por sexo
a = 0.05 # Nivel de confianza

de1 = sd(vibracion[c(1:2,7:8,13:14,19:20,25:26)])
de2 = sd(vibracion[c(3:4,9:10,15:16,21:22,27:28)])
de3 = sd(vibracion[c(5:6,11:12,17:18,23:24,29:30)])

med1 = mean(vibracion[c(1:2,7:8,13:14,19:20,25:26)])
med2 = mean(vibracion[c(3:4,9:10,15:16,21:22,27:28)])
med3 = mean(vibracion[c(5:6,11:12,17:18,23:24,29:30)])

E1 = abs(qnorm(a/2))*de1/sqrt(n) # Error del 95% confianza
E2 = abs(qnorm(a/2))*de2/sqrt(n)
E3 = abs(qnorm(a/2))*de3/sqrt(n)

A1 = med1 - E1 # Cota inferior
A2 = med2 - E2 # Cota inferior
A3 = med3 - E3 # Cota inferior

B1 = med1 + E1 # Cota superior
B2 = med2 + E2 # Cota superior
B3 = med3 + E3 # Cota superior

plot(0, ylim = c(0,3+1), xlim = c(12,16), yaxt = "n", ylab = "", xlab = "Vibración por Material")
axis(2, at = c(1,2,3), labels = c("Acero", "Aluminio",'Plástico'))

arrows(A1, 1, B1, 1, angle=90, code=3, length = 0.1, lwd = 2, col = 'blue')
arrows(A2, 2, B2, 2, angle=90, code=3, length = 0.1, lwd = 2, col = 'green')
arrows(A3, 3, B3, 3, angle=90, code=3, length = 0.1, lwd = 2, col = 'purple')

points(med1, 1, pch = 19, cex = 1.1, col = 'blue')
points(med2, 2, pch = 19, cex = 1.1, col = 'green')
points(med3, 3, pch = 19, cex = 1.1, col = 'purple')
abline(v = mean(vibracion), lty = 3, col = 'red')
```



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

Tal como el análisis de la variable “Sexo” del problema pasado, la variable “Material” no es un factor principal por su cuenta, pues los intervalos de confianza del acero, aluminio y plástico todos se intersectan y se encuentran dentro de la media general.

*Escribe tus conclusiones parciales

Es de esta manera que realizamos un análisis exhaustivo y podemos decri con un 95% de confianza que el “Material” no influye en las vibraciones de los motores.

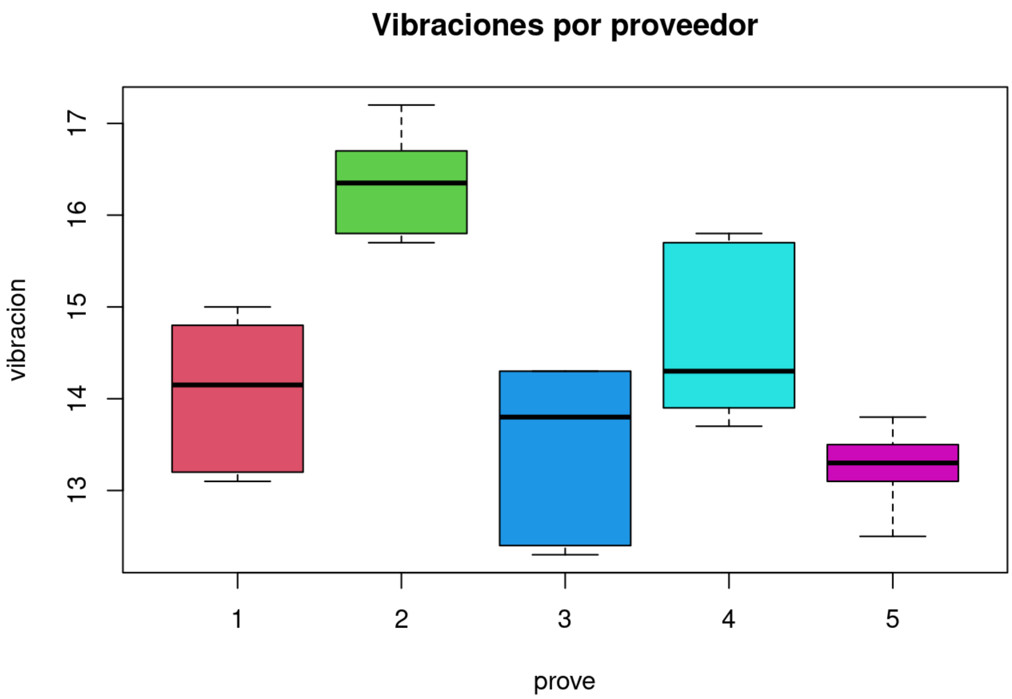
5. Realiza el ANOVA para un efecto principal

*Consulta el código de R en los apoyos de clase de “ANOVA”

```
# No es necesario, pues ya conocemos los factores principales.
```

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

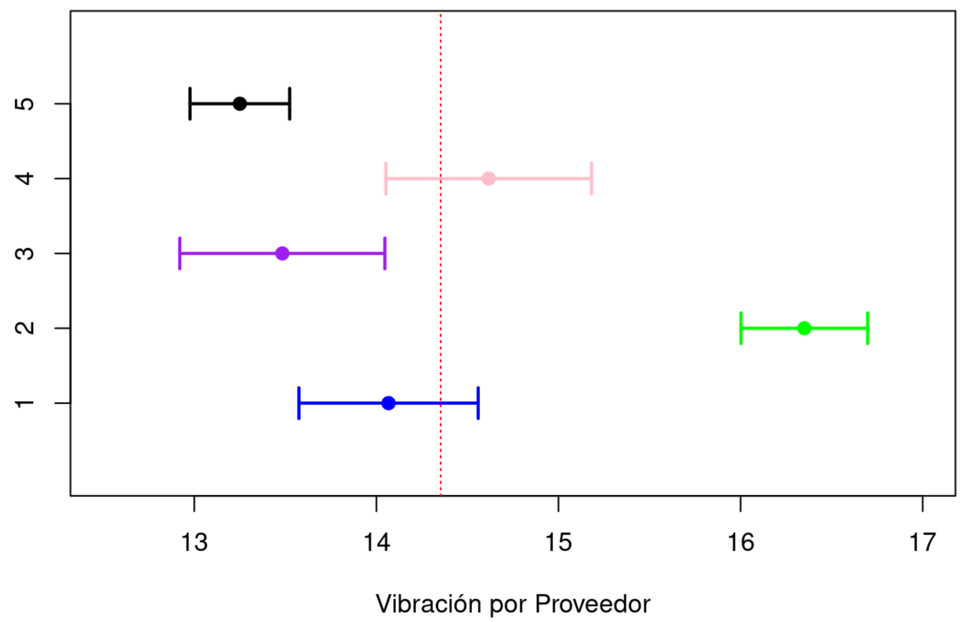
```
boxplot(vibracion~prove, data2, col=2:8,
        main = "Vibraciones por proveedor")
```



Vibraciones por proveedor
vibracion
17
16
15
14
13
1
2
3
4
5
prove

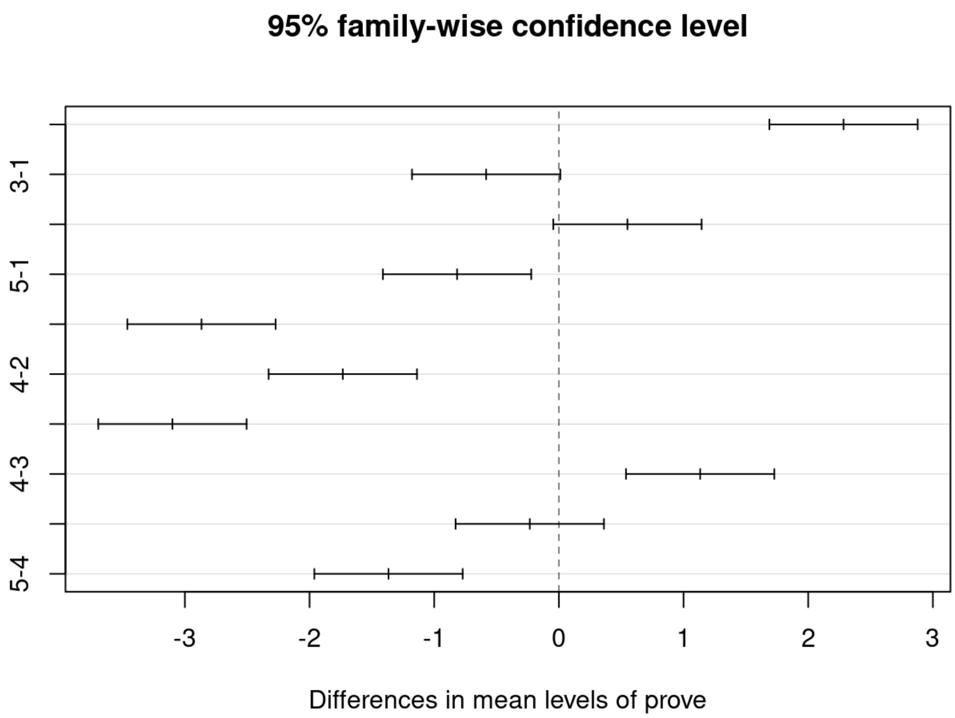
*Haz los intervalos de confianza de rendimiento por método. Grafícalos

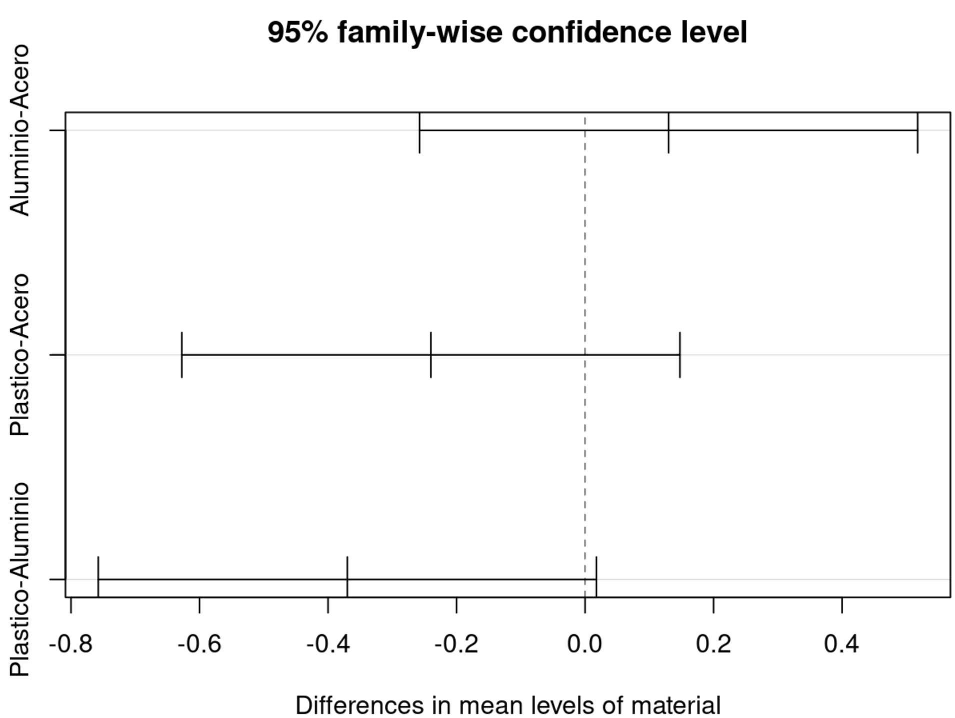
```
n = 10 # Número de estudiantes por sexo
a = 0.05 # Nivel de confianza

de1 = sd(vibracion[1:6])
de2 = sd(vibracion[7:12])
de3 = sd(vibracion[13:18])
de4 = sd(vibracion[19:24])
de5 = sd(vibracion[25:30])

med1 = mean(vibracion[1:6])
med2 = mean(vibracion[7:12])
med3 = mean(vibracion[13:18])
med4 = mean(vibracion[19:24])
med5 = mean(vibracion[25:30])

E1 = abs(qnorm(a/2))*de1/sqrt(n) # Error del 95% confianza
E2 = abs(qnorm(a/2))*de2/sqrt(n)
E3 = abs(qnorm(a/2))*de3/sqrt(n)
E4 = abs(qnorm(a/2))*de4/sqrt(n)
E5 = abs(qnorm(a/2))*de5/sqrt(n)

A1 = med1 - E1 # Cota inferior
A2 = med2 - E2 # Cota inferior
A3 = med3 - E3 # Cota inferior
A4 = med4 - E4 # Cota inferior
A5 = med5 - E5 # Cota inferior

B1 = med1 + E1 # Cota superior
B2 = med2 + E2 # Cota superior
B3 = med3 + E3 # Cota superior
B4 = med4 + E4 # Cota superior
B5 = med5 + E5 # Cota superior

plot(0, ylim = c(0,5+1), xlim = c(12.5,17), yaxt = "n", ylab = "", xlab = "Vibración por Proveedor")
axis(2, at = c(1,2,3,4,5), labels = c("1", "2",'3','4','5'))

arrows(A1, 1, B1, 1, angle=90, code=3, length = 0.1, lwd = 2, col = 'blue')
arrows(A2, 2, B2, 2, angle=90, code=3, length = 0.1, lwd = 2, col = 'green')
arrows(A3, 3, B3, 3, angle=90, code=3, length = 0.1, lwd = 2, col = 'purple')
arrows(A4, 4, B4, 4, angle=90, code=3, length = 0.1, lwd = 2, col = 'pink')
arrows(A5, 5, B5, 5, angle=90, code=3, length = 0.1, lwd = 2, col = 'black')

points(med1, 1, pch = 19, cex = 1.1, col = 'blue')
points(med2, 2, pch = 19, cex = 1.1, col = 'green')
points(med3, 3, pch = 19, cex = 1.1, col = 'purple')
points(med4, 4, pch = 19, cex = 1.1, col = 'pink')
points(med5, 5, pch = 19, cex = 1.1, col = 'black')
abline(v = mean(vibracion), lty = 3, col = 'red')
```

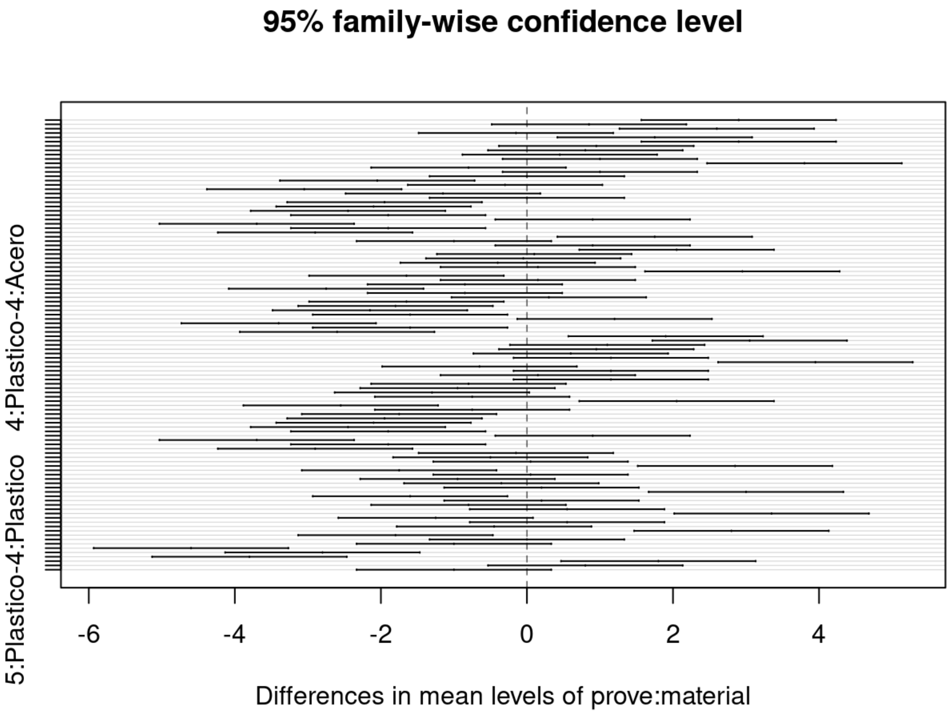


*Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey.

```
Tu=TukeyHSD(anova)
plot(Tu)
```


95% family-wise confidence level
3-1
5-1
4-2
4-3
5-4
-3
-2
-1
0
1
2
3
Differences in mean levels of prove


95% family-wise confidence level
Aluminio-Acero
Plastico-Acero
Plastico-Aluminio
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
Differences in mean levels of material



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

El boxplot muestra que hay variabilidad en las vibraciones entre los diferentes proveedores, con el proveedor 2 mostrando las vibraciones más altas y el proveedor 5 las más bajas.

El gráfico de intervalos de confianza muestra que los proveedores tienen diferentes medias de vibración, con algunos mostrando mayor dispersión (como el proveedor 3 y 4) y otros una dispersión más estrecha (como el proveedor 1 y 2). La línea roja vertical indica la media de vibración considerado importante para la comparación.

Finalmente, el gráfico de Turkey permite identificar qué pares de proveedores tienen diferencias significativas en los niveles medios de vibración. Los intervalos que no incluyen el 0 indican una diferencia significativa, mientras que aquellos que lo cruzan sugieren que no hay una diferencia significativa entre esos pares de proveedores.

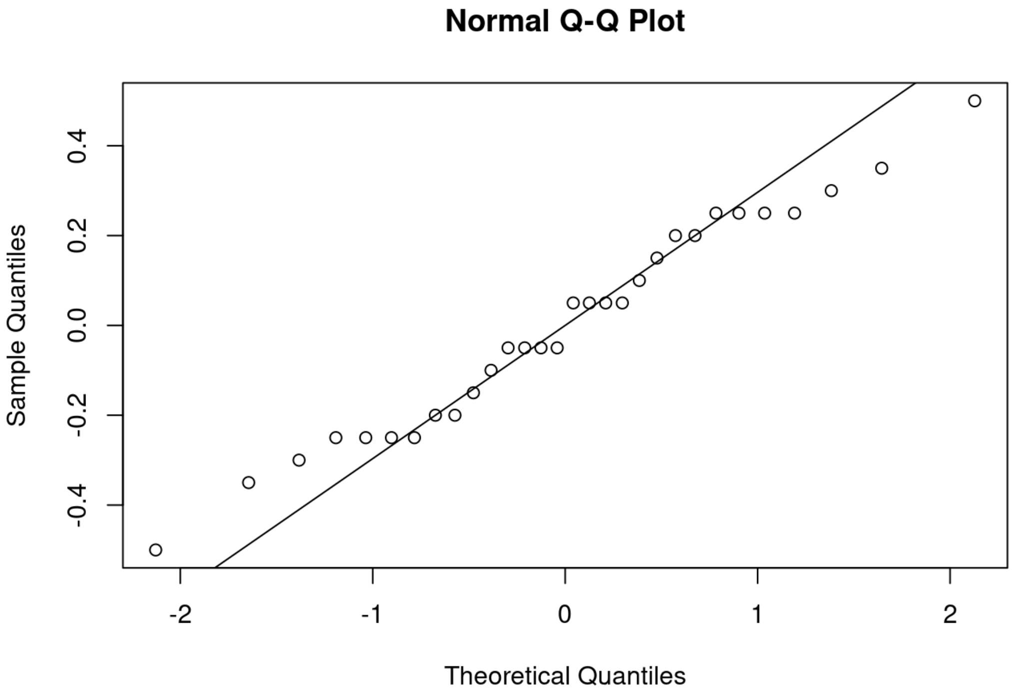
*Escribe tus conclusiones parciales

Estas gráficas demuestran que el "Proveedor" del material es un factor principal para explicar la vibración de los motores. Demostrando también que existe la interacción de "Material" y "Proveedor".

6. Comprueba la validez del modelo. Comprueba:

*Normalidad

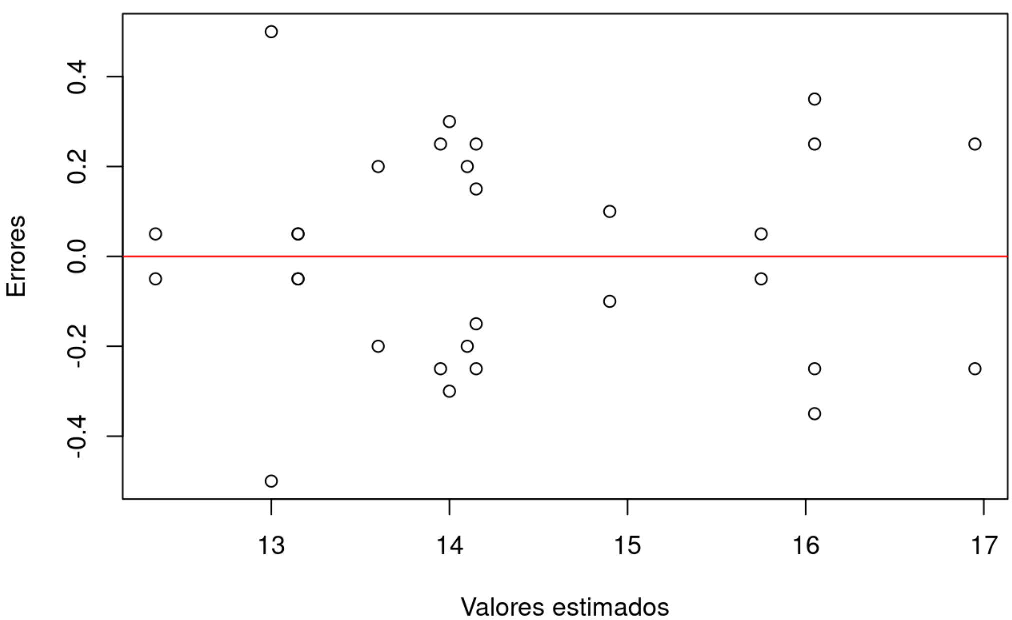
```
anova = aov(vibracion~prove*material)
residuos=anova$residuals
qqnorm(residuos)
qqline(residuos)
```



El modelo se acerca a la normalidad, pero muestra leptocurtosis, pues se alza por encia de la línea, pero luego queda por debajo de manera no tan pronunciada.

*Homocedasticidad

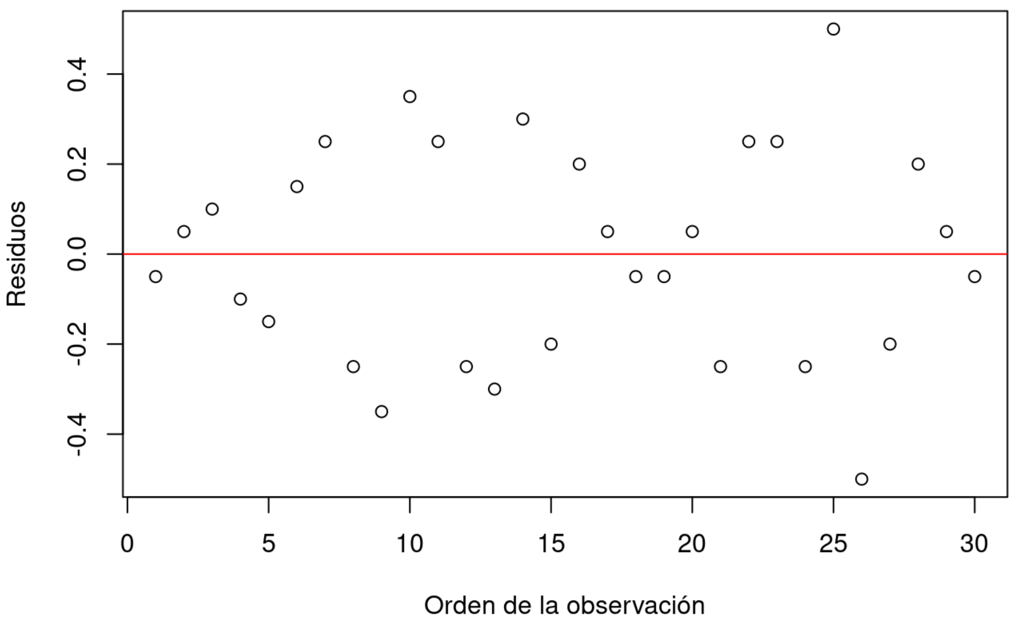
```
plot(anova$fitted.values,anova$residuals,ylab="Errores",xlab="Valores estimados")
abline(h=0,col="red")
```



Se muestra que los residuos tienen una varianza constante.

*Independencia

```
n = tapply(vibracion, prove:material, length)
plot(c(1:sum(n)),anova$residuals,xlab="Orden de la observación",ylab="Residuos")
abline(h=0,col="red")
```



Los datos no aparentan una tendencia.

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

```
(36.67+0.7+11.61)/(36.67+0.7+11.61+1.67)
```

```
## [1] 0.9670286
```

Existe una alta relación entre las variables, se explica un 96% de la varianza con las variables “Proveedor”, “Material” y la interacción entre ambos, siendo las más significativas “Proveedor” y la interacción.

7. Concluye en el contexto del problema.

Se observó que los materiales vendidos por los proveedores producen un efecto diferente en las vibraciones de los motores. El efecto del proveedor 2 es una vibración mucho mayor que los demás, aumenta las vibraciones con respecto a la media general, el proveedor 5, en cambio, tiene el menor promedio de vibraciones en los motores, y los proveedores 1 y 4 mantienen la vibraciones en la media general por lo que resultan ser los proveedores de material con vibraciones más estables, mientras que el proveedor 5 es el que menos vibraciones da en su material. El modelo explica el 96.7% de la variación. Por lo tanto, el proveedor y el material dado por este son factores determinantes en la vibración de los motores, sin embargo, es posible que haya otros factores que expliquen el resto del porcentaje de variación (3.3%) y que en este modelo se le atribuye a la aleatoriedad (al error).

El número de datos en cada tratamiento fue igual por lo que es un diseño equilibrado que es robusto a heterocedasticidad. De acuerdo al análisis de los gráficos Q-Q y de los residuos vs. el valor esperado (ajustado), los datos aparentemente cumplen con normalidad e independencia. También los errores tienen una media cero y variación constante.