

ANOVA

Erika Martínez Meneses

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

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

##Análisis exploratorio.

Calcula la media para el rendimiento por método de enseñanza.

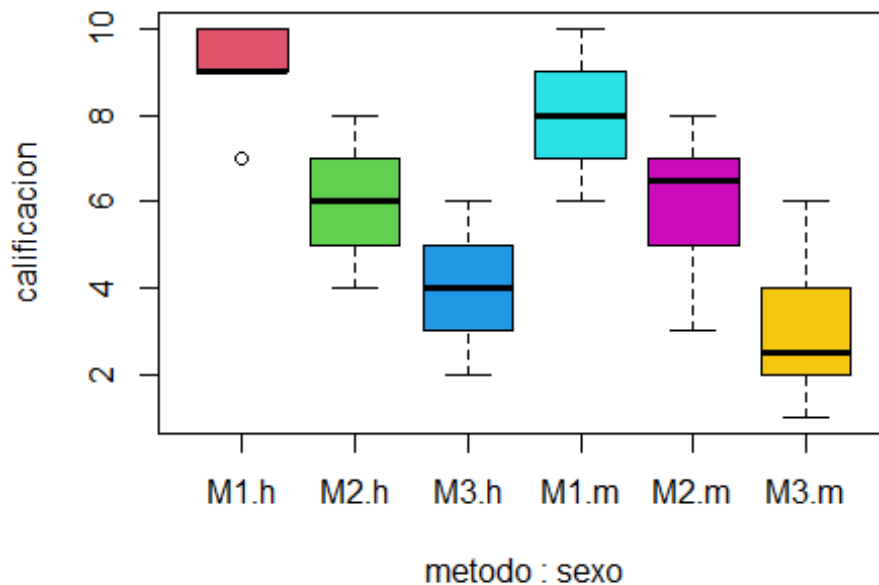
```
media_metodo <- tapply(calificacion, metodo, mean)
media_metodo

##  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, col = 2:8,
main = "Boxplot de Rendimiento por Método de Enseñanza y Sexo")
```

Boxplot de Rendimiento por Método de Enseñanza y



Valor frontera

```
qf(0.95,3,20)
```

```
## [1] 3.098391
```

A partir de que sea 3 veces más grande, voy a rechazar H_0

Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.
Escribe tus conclusiones parciales

Medias: $M1 > M2 > M3$

Podemos observar gracias al boxplot y a las medias que el método 1 tiene mayor rendimiento, en segundo lugar tenemos al método 2 y con menor rendimiento de los métodos el método 3, esto se aprecia tanto para hombres como para mujeres.

2. Las hipótesis.

Establece las hipótesis estadísticas (tienen que ser 3).

Hipótesis 1

- $H_0: \tau_i = 0$ El rendimiento no depende de la metodología de enseñanza.
- $H_1: \exists \tau_i \neq 0$ El rendimiento depende de la metodología de enseñanza.

Hipótesis 2

- $H_0: \alpha_j = 0$ El rendimiento no depende del género de los estudiantes.
- $H_1: \exists \alpha_j \neq 0$ El rendimiento depende del género de los estudiantes. (algún α_j es distinto de cero)

Hipótesis 3

- $H_0: \tau_i * \alpha_j = 0$ No existe interacción entre el género y la metodología de enseñanza en el rendimiento de los estudiantes.
- $H_1: \tau_i * \alpha_j \neq 0$ Existe interacción entre el género y la metodología de enseñanza en el rendimiento de los estudiantes.

3. Realiza el ANOVA para dos niveles con interacción:

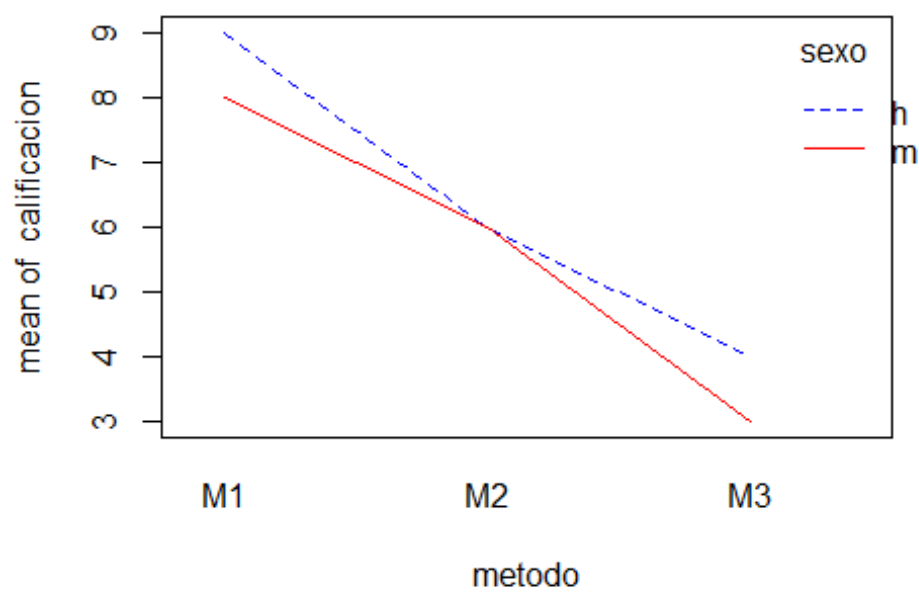
```
anova_interaccio<-aov(calificacion~metodo*sexo)
summary(anova_interaccio)
```

```
##              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
```

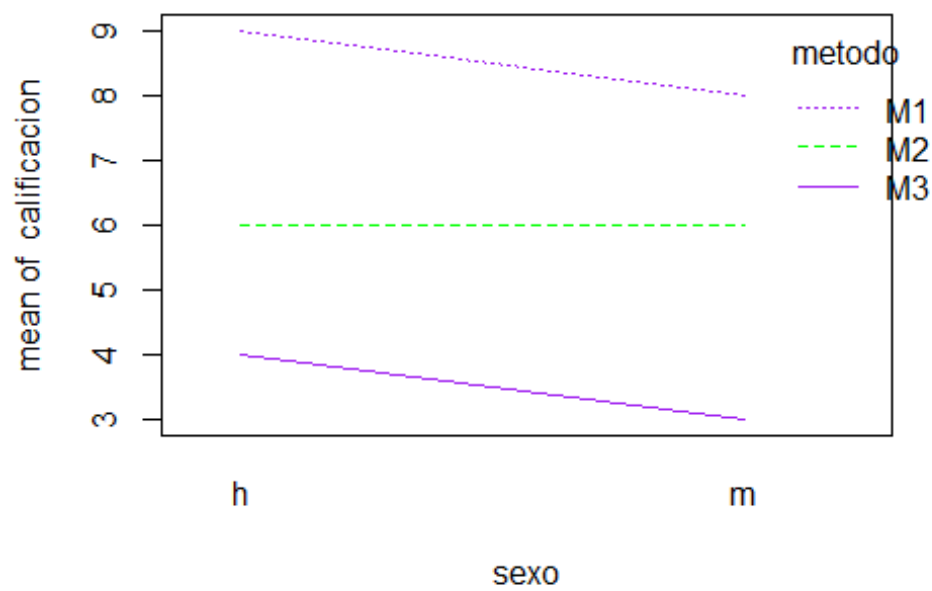
La interacción (Pr(>F)) es alta (.655) y el valor F es bajo (.429), la interacción no es significativa

###Haz la gráfica de interacción de dos factores en ANOVA

```
interaction.plot(metodo,sexo,calificacion, col=c("blue","red"), pch=19)
```



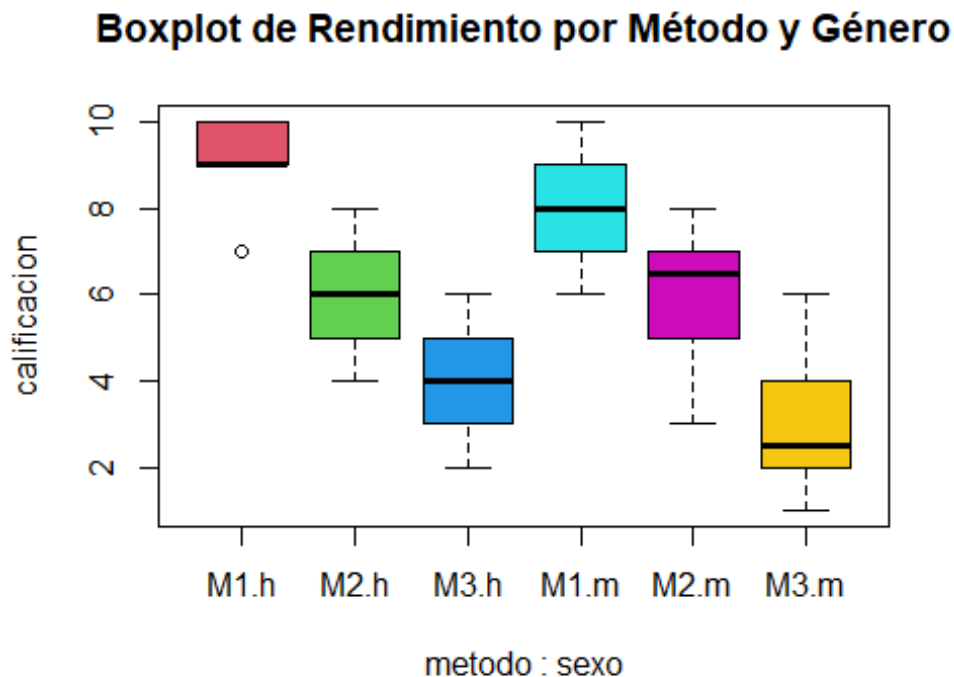
```
interaction.plot(sexo, metodo, calificacion, col=c("purple","green"),
pch=19)
```



Cuando las líneas se cruzan significa que hay efecto de interacción pero como no lo hacen significa que la interacción no es significativa.

Haz el boxplot para visualizar la interacción de los factores:

```
boxplot(calificacion ~ metodo * sexo, col = 2:8,
        main = "Boxplot de Rendimiento por Método y Género")
```



Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Escribe tus conclusiones parciales.

Después de realizar el ANOVA con el modelo completo, encontramos que método*sexo (la interacción de los factores) fue no significativa (no hay efecto de interacción). No se rechaza la tercera hipótesis nula (H_0) y el modelo se reduce.

H_0 No existe interacción entre el género y la metodología de enseñanza en el rendimiento de los estudiantes <- No se rechaza

##4. Realiza el ANOVA para dos niveles sin interacción.

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

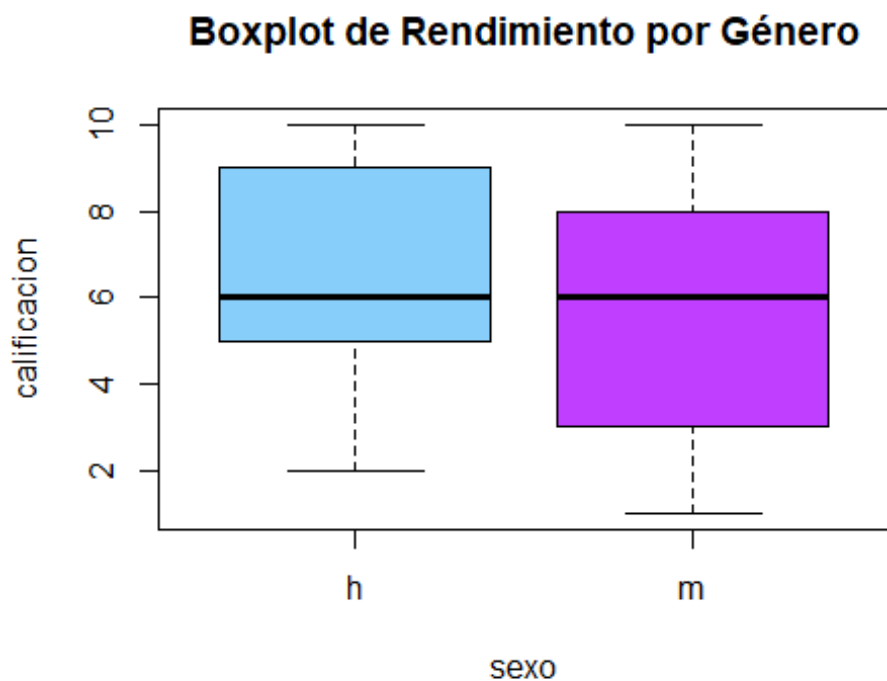
##		Df	Sum Sq	Mean Sq	F value	Pr(>F)
##	sexo	1	4	4.00	1.778	0.192
##	metodo	2	150	75.00	33.333	1.5e-08 ***
##	Residuals	32	72	2.25		

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La variación (F value) es pequeña (1.778) y $\Pr(>F)=0.192$ es mayor que 0.05 lo que significa que no es significativa.

Haz el boxplot de rendimiento por sexo.

```
boxplot(calificacion ~ sexo, col = c("#87CEFA", "#BF3EFF"),
        main = "Boxplot de Rendimiento por Género")
```



Calcula la media para el rendimiento por sexo y método.

Media por sexo

```
media_sexo <- tapply(calificacion, sexo, mean)
media_sexo

##           h           m
## 6.333333  5.666667
```

Media por método

```
media_metodo <- tapply(calificacion, metodo, mean)
media_metodo

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

Media por sexo y método

```
media_sexo_metodo <- tapply(calificacion, list(sexo, metodo), mean)
media_sexo_metodo

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

Haz los intervalos de confianza de rendimiento por sexo.

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

sexo_conf_int <- datos %>%
  group_by(sexo) %>%
  summarise(media = mean(calificacion),
            sd = sd(calificacion),
            n = n()) %>%
  mutate(lower = media - qt(0.975, df=n-1) * sd / sqrt(n),
         upper = media + qt(0.975, df=n-1) * sd / sqrt(n))

print(sexo_conf_int)

## # A tibble: 2 × 6
##   sexo media    sd      n lower upper
##   <fct> <dbl> <dbl> <int> <dbl> <dbl>
## 1 h      6.33  2.47    18  5.10  7.56
## 2 m      5.67  2.63    18  4.36  6.98
```

Grafícalos

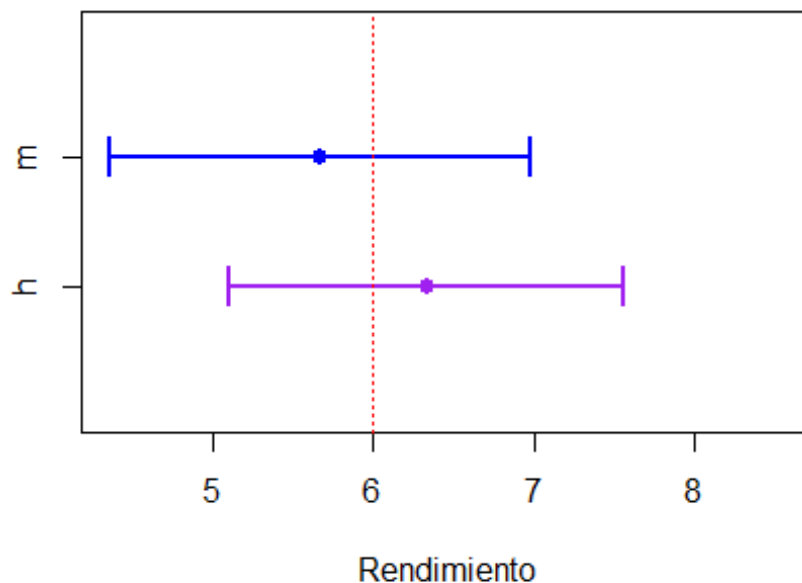
```
plot(0, ylim=c(0, 3), xlim=c(min(sexo_conf_int$lower),
max(sexo_conf_int$upper) + 1),
     yaxt="n", ylab="", xlab = "Rendimiento", main = "Intervalos de
Confianza de Rendimiento por Género")
axis(2, at=c(1,2), labels=sexo_conf_int$sexo)

arrows(sexo_conf_int$lower[1], 1, sexo_conf_int$upper[1], 1, angle=90,
code=3, length=0.1, lwd=2, col="purple")
arrows(sexo_conf_int$lower[2], 2, sexo_conf_int$upper[2], 2, angle=90,
code=3, length=0.1, lwd=2, col="blue")
```

```
points(sexo_conf_int$media[1], 1, pch=19, cex=1.1, col="purple")
points(sexo_conf_int$media[2], 2, pch=19, cex=1.1, col="blue")

abline(v = mean(sexo_conf_int$media), lty = 3, col = "red")
```

Intervalos de Confianza de Rendimiento por Género



Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.
Escribe tus conclusiones parciales.

Después de realizar el ANOVA con el modelo con los efectos principales, encontramos que el sexo fue no significativa (no hay efecto del sexo). No se rechaza la segunda hipótesis nula (H_0) y el modelo se reduce.

H_0 : El rendimiento no depende del género de los estudiantes. <- No se rechaza

5. Realiza el ANOVA para un efecto principal

```
# ANOVA para un efecto principal (método de enseñanza)
anova_metodo <- aov(calificacion ~ metodo)
summary(anova_metodo)
```

```
##              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
```

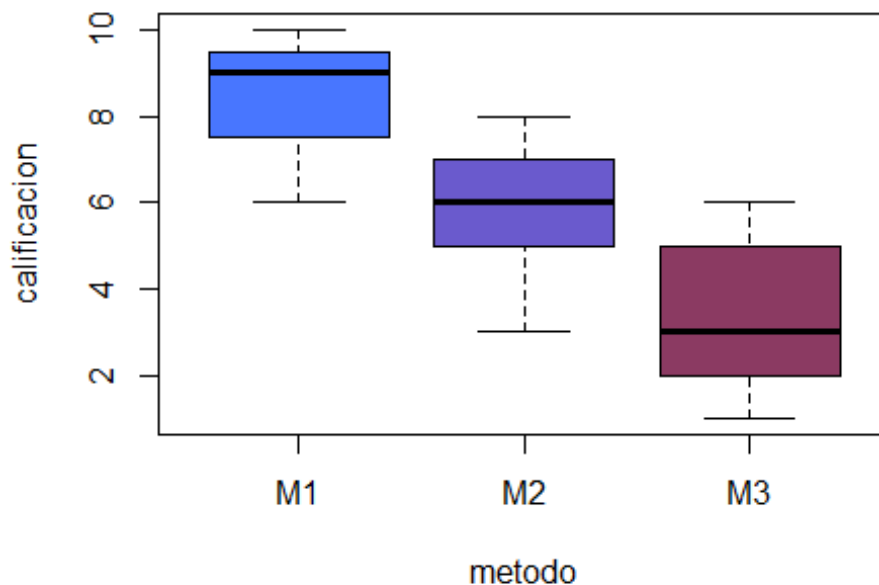
El método sí es significativo, el F value es alto y $\text{Pr}(>F)$ es menor a 0.05.

Nota: Los valores que se utilizan en la Relación lineal entre las variables son el 150 del método y el 76 de Residuals.

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

```
boxplot(calificacion ~ metodo, col = c("#4876FF", "#6A5ACD", "#8B3A62"),
        main = "Boxplot de Rendimiento por Método de Enseñanza")
```

Boxplot de Rendimiento por Método de Enseñanza



Haz los intervalos de confianza de rendimiento por método.

```
metodo_conf_int <- datos %>%
  group_by(metodo) %>%
  summarise(media = mean(calificacion),
            sd = sd(calificacion),
            n = n()) %>%
  mutate(lower = media - qt(0.975, df=n-1) * sd / sqrt(n),
         upper = media + qt(0.975, df=n-1) * sd / sqrt(n))
metodo_conf_int
```

```
## # A tibble: 3 × 6
##   metodo media    sd    n lower upper
##   <fct> <dbl> <dbl> <int> <dbl> <dbl>
## 1 M1      8.5  1.31    12  7.66  9.34
## 2 M2      6    1.54    12  5.02  6.98
## 3 M3      3.5  1.68    12  2.43  4.57
```

Grafícalos

```

plot(0, ylim=c(0, 3), xlim=c(min(metodo_conf_int$lower),
max(metodo_conf_int$upper) + 1),
     yaxt="n", ylab="", xlab = "Rendimiento", main = "Intervalos de
Confianza de Rendimiento por Método")
axis(2, at=c(1,2,3), labels=metodo_conf_int$metodo)

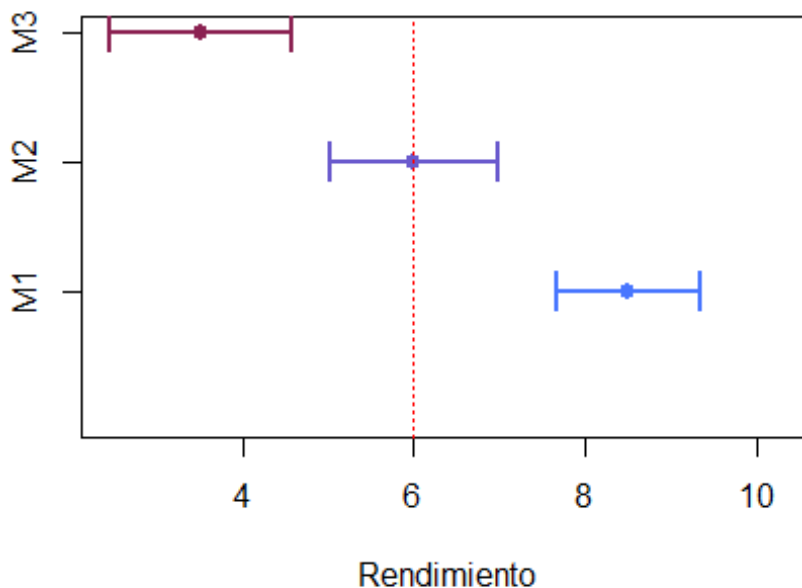
arrows(metodo_conf_int$lower[1], 1, metodo_conf_int$upper[1], 1,
       angle=90, code=3, length=0.1, lwd=2, col="#4876FF")
arrows(metodo_conf_int$lower[2], 2, metodo_conf_int$upper[2], 2,
       angle=90, code=3, length=0.1, lwd=2, col="#6A5ACD")
arrows(metodo_conf_int$lower[3], 3, metodo_conf_int$upper[3], 3,
       angle=90, code=3, length=0.1, lwd=2, col="#8B2252")

points(metodo_conf_int$media[1], 1, pch=19, cex=1.1, col="#4876FF")
points(metodo_conf_int$media[2], 2, pch=19, cex=1.1, col="#6A5ACD")
points(metodo_conf_int$media[3], 3, pch=19, cex=1.1, col="#8B2252")

abline(v = mean(metodo_conf_int$media), lty = 3, col = "red")

```

Intervalos de Confianza de Rendimiento por Método



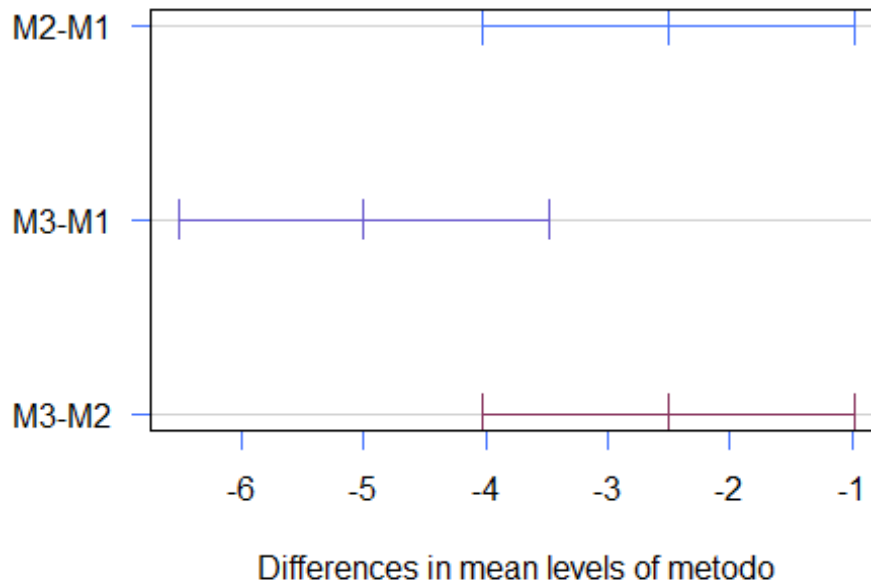
Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey.

```

tukey_test <- TukeyHSD(anova_metodo)
plot(tukey_test, las=1, col = c("#4876FF", "#6A5ACD", "#8B3A62") )

```

95% family-wise confidence level



```
tukey_test
## 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
```

Nota : Si pasa por cero significa que no hay diferencia entre M2-M1

Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.
Escribe tus conclusiones parciales.

Las medias muestrales nos dieron que $\mu_1 = 8.5$, $\mu_2 = 6$, $\mu_3 = 3.5$ y nuestra media poblacional es $\mu = 6$ por lo que los efectos para cada nivel son $M_1 = 2.5$, $M_2 = 0.0$, $M_3 = -2.5$

El método de enseñanza sí es significativo para el rendimiento de los estudiantes por lo que la primera hipótesis nula se rechaza.

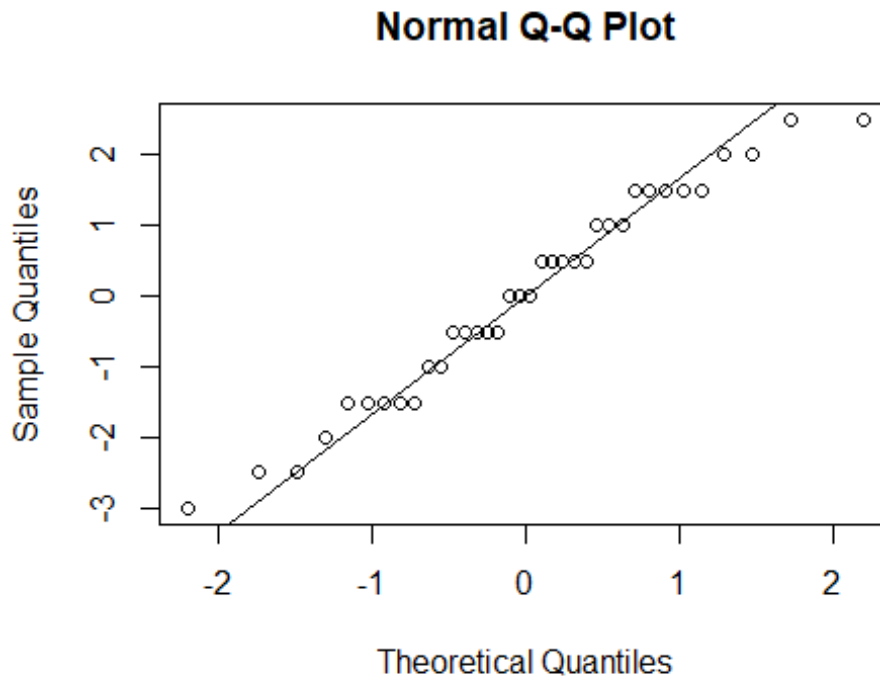
H_0 : El rendimiento no depende de la metodología de enseñanza. <- Se Rechaza

6. Comprueba la validez del modelo.

Comprueba:

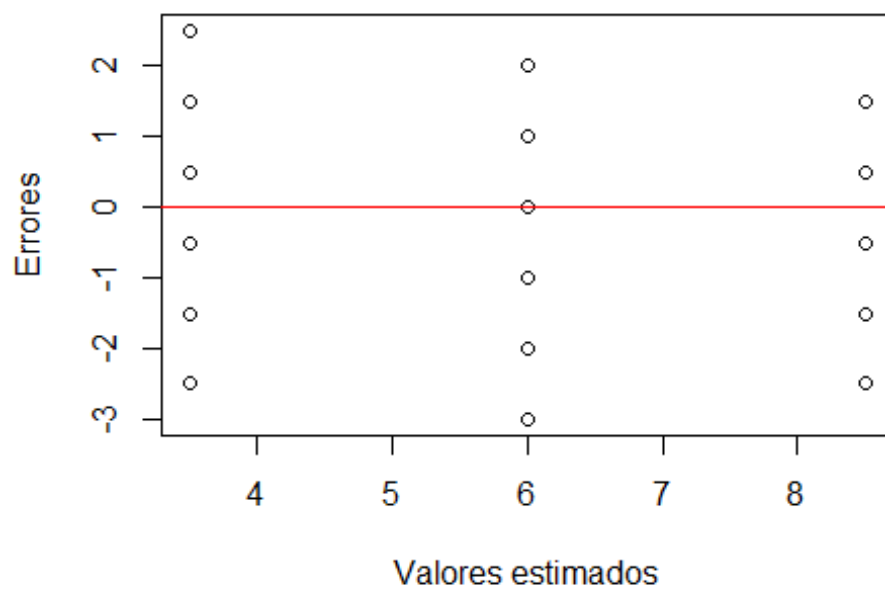
Normalidad

```
residuos = anova_metodo$residuals  
qqnorm(residuos)  
qqline(residuos)
```



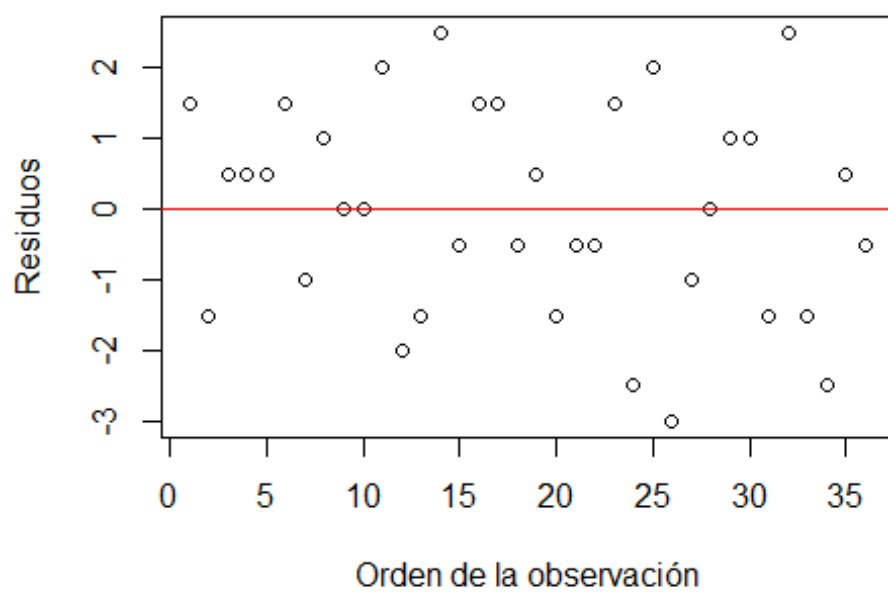
Homocedasticidad

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



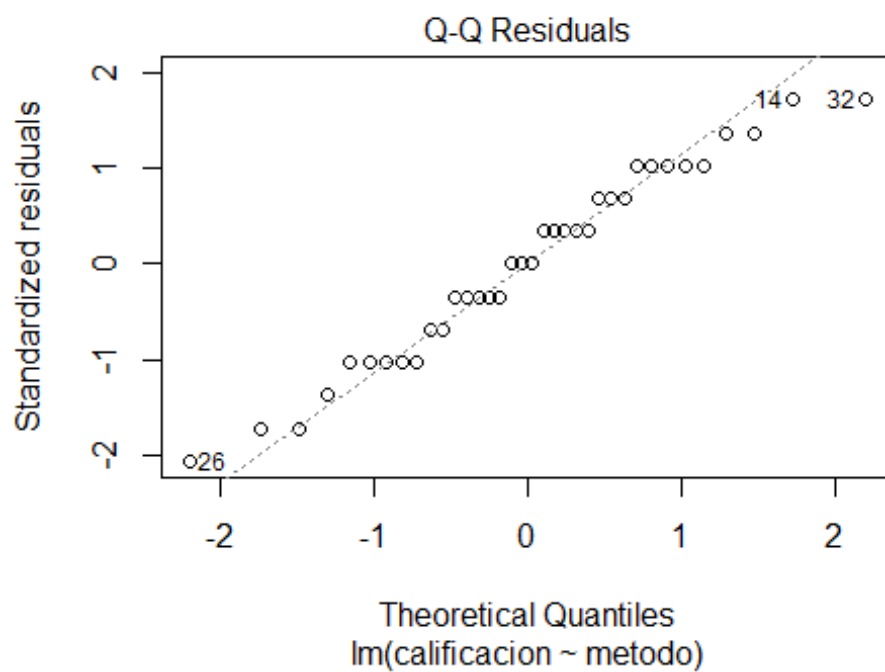
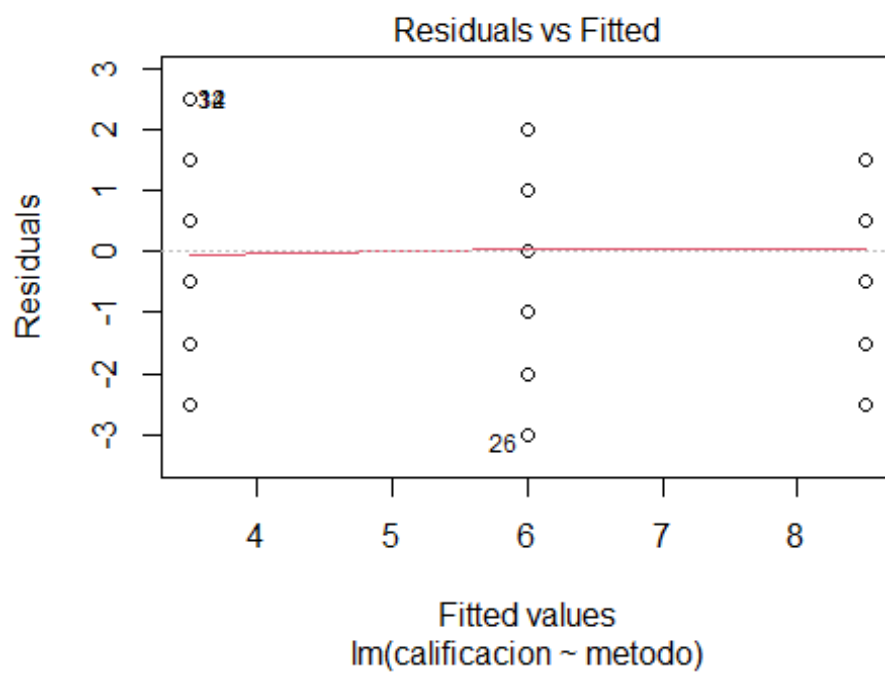
Independencia

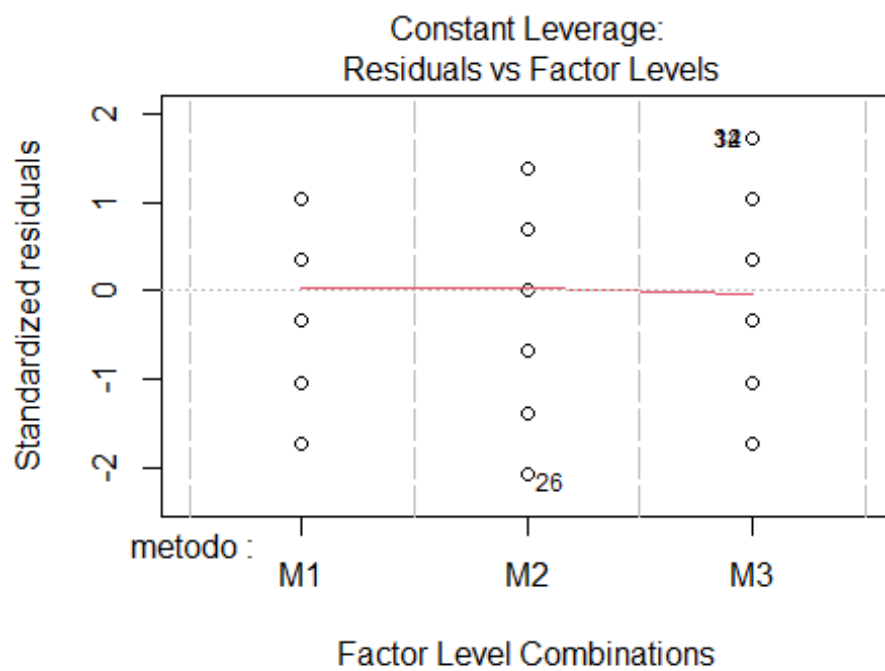
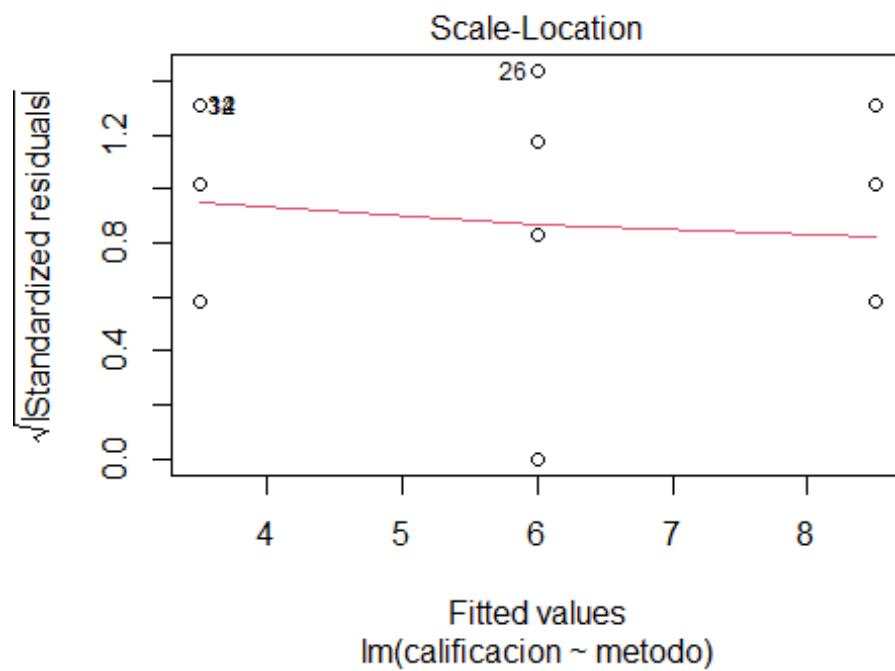
```
n = tapply(calificacion, metodo, length)
plot(c(1:sum(n)), anova_metodo$residuals, xlab="Orden de la
observación", ylab="Residuos")
abline(h=0, col="red")
```



Relación lineal entre las variables (coeficiente de determinación).

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





```
CD= 150/(150+76) #coeficiente de determinación para el modelo
CD # r^2
```

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


Concluye en el contexto del problema.

Varianza Explicada por el modelo: 66.37% ($SC_{Tratamiento}/SC_{totales}$)

Efectos significativos: ¿Hubo efectos significativo? Si así es: ¿todos son diferentes? ¿hay grupos? ¿qué significa en el contexto del problema? Se observó que los 3 métodos producen un efecto diferente en el rendimiento de los niños. 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

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, 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)
proveedor=c(rep("P1",2),rep("P2",2),rep("P3",2),rep("P4",2),rep("P5",2),re
p("P1",2),rep("P2",2),rep("P3",2),rep("P4",2),rep("P5",2),rep("P1",2),rep
("P2",2),rep("P3",2),rep("P4",2),rep("P5",2))
material = c(rep("Acero", 10), rep("Aluminio",10), rep("Plástico",10))
proveedor = factor(proveedor)
material = factor(material)
datos = data.frame(vibracion, proveedor, material)
```

¿Existe evidencia de que el proveedor y el material son causas de la vibración de los motores eléctricos?

##Análisis exploratorio.

Calcula la media para la vibración por proveedor.

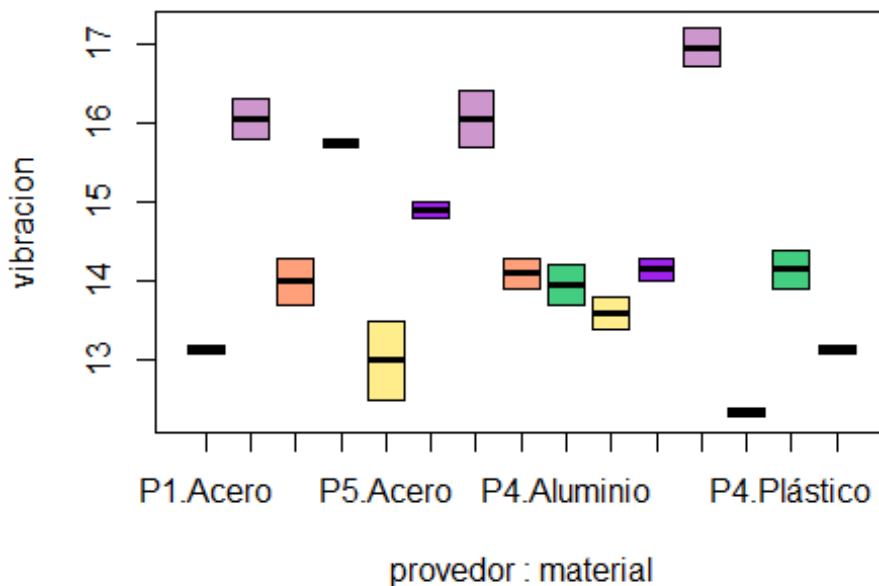
```
media_proveedor <- tapply(vibracion, proveedor, mean)
media_proveedor
```

```
##      P1      P2      P3      P4      P5
## 14.06667 16.35000 13.48333 14.61667 13.25000
```

Haz el boxplot de la cantidad de vibración por proveedor y material.

```
boxplot(vibracion ~ proveedor:material, col=c("purple", "#CD96CD",
"#FFA07A", "#43CD80", "#FFEC8B"),
        main = "Boxplot de Cantidad de vibración por proveedor y
materia.")
```

Boxplot de Cantidad de vibración por proveedor y ma



Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.
Escribe tus conclusiones parciales

Se puede observar que en la media para la vibración por proveedor la más alta es para el proveedor 2 con $\mu = 16.35$ y la menor es para el proveedor 5 con $\mu = 13.25$. En la gráfica de boxplot se pueden observar las medias para cada proveedor con respecto a cada material (los colores están asociados al tipo de proveedor).

2. Las hipótesis.

Establece las hipótesis estadísticas (tienen que ser 3).

Hipótesis 1

- $H_0: \tau_i = 0$ No hay efecto de los proveedores
- $H_1: \exists \tau_i \neq 0$ Existe al menos alguno de los proveedores distinto de 0

Hipótesis 2

- $H_0: \alpha_j = 0$ No hay efecto en los materiales.
- $H_1: \exists \alpha_j \neq 0$ Existe al menos alguno de los materiales distinto de 0

Hipótesis 3

- $H_0: \tau_i * \alpha_j = 0$ No hay interacción entre el material y el proveedor en cuanto a la vibración.
- $H_1: \tau_i * \alpha_j \neq 0$ Existe interacción entre el material y el proveedor en cuanto a la vibración.

3. Realiza el ANOVA para dos niveles con interacción:

```
anova_interaccion<-aov(vibracion~proveedor*material)
summary(anova_interaccion)
```

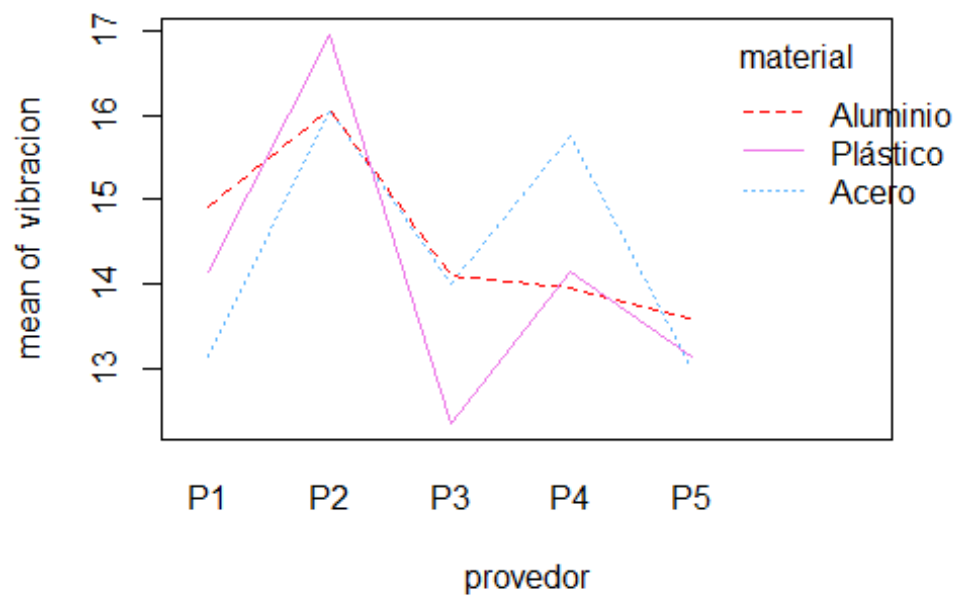
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## proveedor      4  36.67    9.169   82.353 5.07e-10 ***
## material       2   0.70    0.352    3.165  0.0713 .
## proveedor:material  8  11.61    1.451   13.030 1.76e-05 ***
## Residuals     15   1.67    0.111
## ---
## Signif. codes:  0 '***'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

El efecto influye en el interacción por lo cuál se decide no quitar ningún factor.

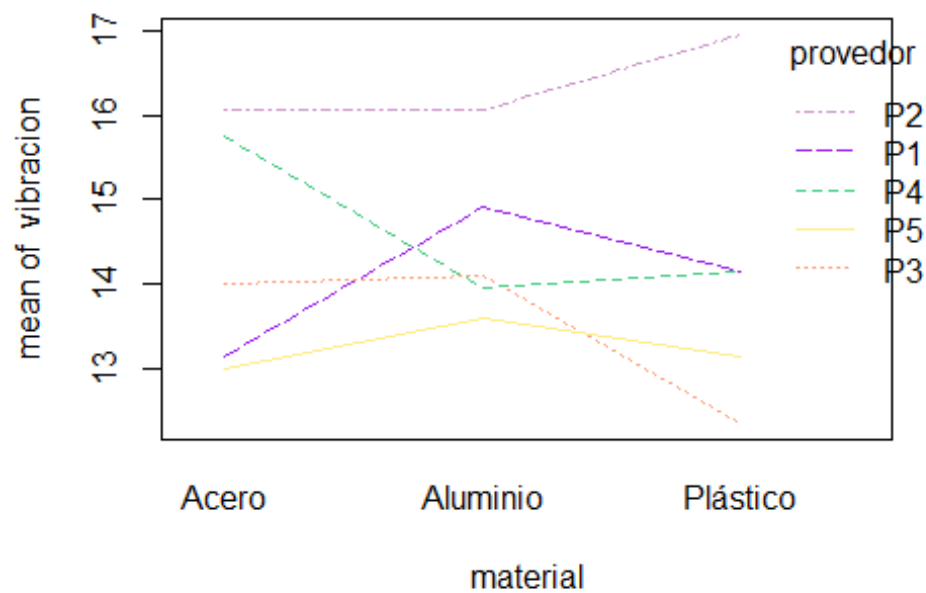
$(36.67+.7+11.61)/(36.67+.7+11.61+1.67)$

###Haz la gráfica de interacción de dos factores en ANOVA

```
interaction.plot(proveedor,material,vibracion, col=c("#63B8FF","red",
"#EE7AE9" ), pch=19)
```



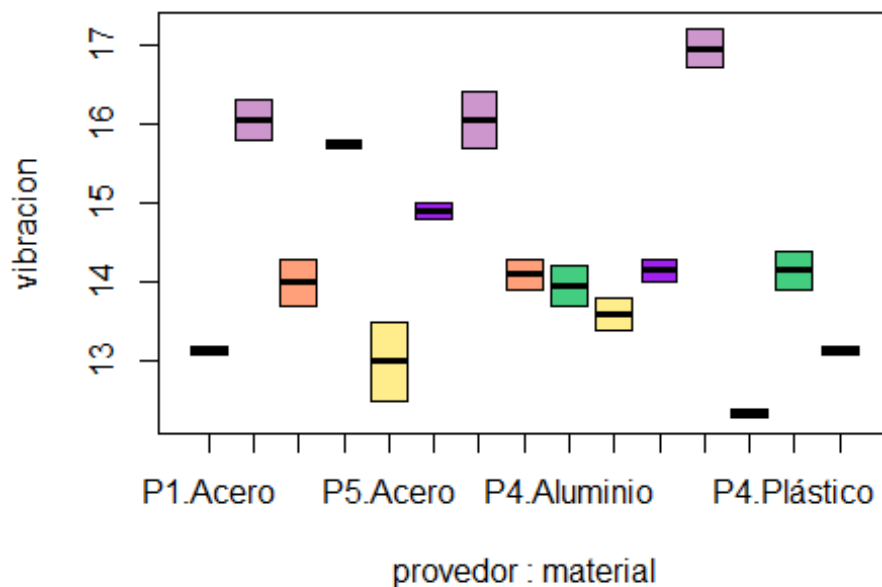
```
interaction.plot(material, proveedor, vibracion, col=c("purple", "#CD96CD",
"#FFA07A", "#43CD80", "#FFEC8B"), pch=19)
```



Haz el boxplot para visualizar la interacción de los factores:

```
boxplot(vibracion ~ proveedor * material, col=c("purple", "#CD96CD",
"#FFA07A", "#43CD80", "#FFEC8B"),
main = "Boxplot de Rendimiento por Proveedor y Material")
```

Boxplot de Rendimiento por Proveedor y Material



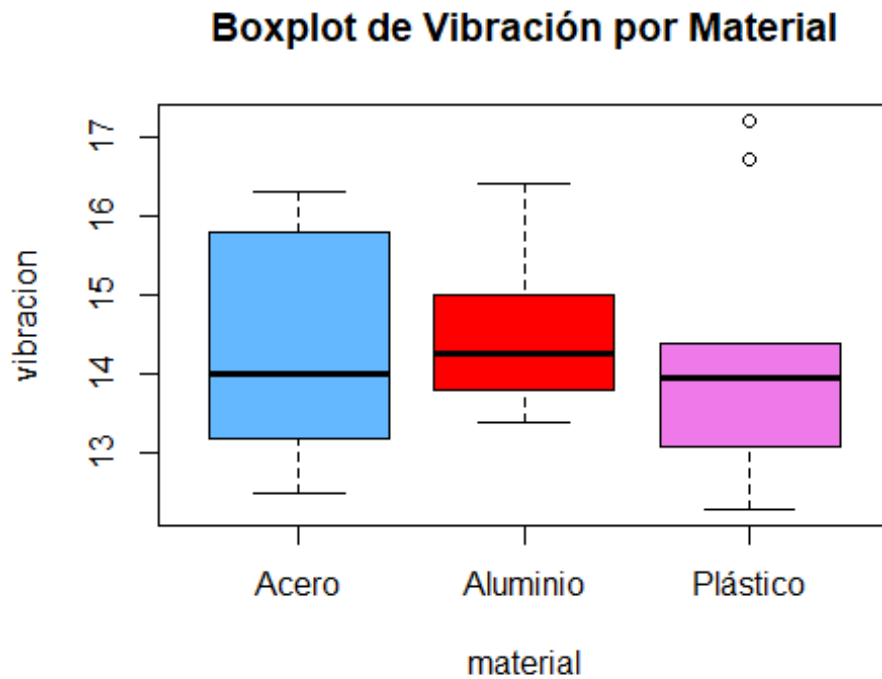
Interpreta el resultado desde la perspectiva estadística y en el contexto del problema

Después de realizar el ANOVA con el modelo completo, encontramos que la interacción entre método y sexo es significativa (con F value = 13.030, $Pr(>F) = 1.76e-05$), es decir, hay un efecto de interacción significativo.

Nota: No quitamos ninguno porque tiene relevancia en la interacción (observamos el anova)

Haz el boxplot de vibración por material

```
boxplot(vibracion ~ material, col=c("#63B8FF", "red", "#EE7AE9" ),
main = "Boxplot de Vibración por Material")
```



Calcula la media para la vibración por material y proveedor.

Media por material.

```
media_material <- tapply(vibracion, material, mean)
media_material

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

Media por proveedor.

```
media_proveedor <- tapply(vibracion, proveedor, mean)
media_proveedor

##      P1      P2      P3      P4      P5
## 14.06667 16.35000 13.48333 14.61667 13.25000
```

Media por material y proveedor.

```
media_material_proveedor <- tapply(vibracion, list(material, proveedor),
mean)
media_material_proveedor

##      P1      P2      P3      P4      P5
## Acero  13.15 16.05 14.00 15.75 13.00
## Aluminio 14.90 16.05 14.10 13.95 13.60
## Plástico 14.15 16.95 12.35 14.15 13.15
```

Haz los intervalos de confianza de vibración por material.

```
library(dplyr)
material_conf_int <- datos %>%
  group_by(material) %>%
  summarise(media = mean(vibracion),
            sd = sd(vibracion),
            n = n()) %>%
  mutate(lower = media - qt(0.975, df=n-1) * sd / sqrt(n),
         upper = media + qt(0.975, df=n-1) * sd / sqrt(n))
```

material_conf_int

```
## # A tibble: 3 × 6
##   material media    sd      n lower upper
##   <fct>    <dbl> <dbl> <int> <dbl> <dbl>
## 1 Acero      14.4  1.39     10  13.4  15.4
## 2 Aluminio   14.5  0.955     10  13.8  15.2
## 3 Plástico   14.2  1.65     10  13.0  15.3
```

Grafícalos

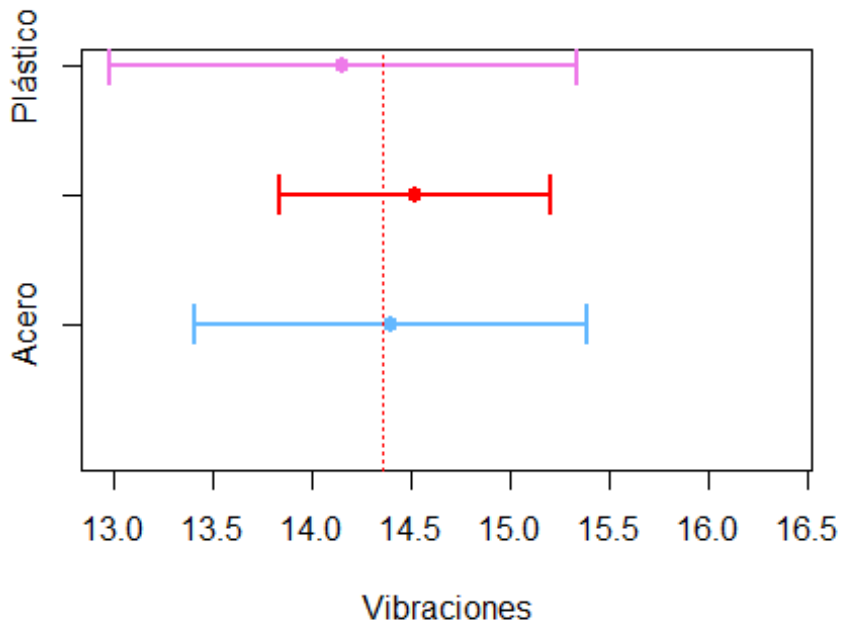
```
plot(0, ylim=c(0, 3), xlim=c(min(material_conf_int$lower),
max(material_conf_int$upper) + 1),
     yaxt="n", ylab="", xlab = "Vibraciones", main = "Intervalos de
Confianza de Vibraciones por Material")
axis(2, at=c(1,2, 3), labels=material_conf_int$material)

arrows(material_conf_int$lower[1], 1, material_conf_int$upper[1], 1,
angle=90, code=3, length=0.1, lwd=2, col="#63B8FF")
arrows(material_conf_int$lower[2], 2, material_conf_int$upper[2], 2,
angle=90, code=3, length=0.1, lwd=2, col="red")
arrows(material_conf_int$lower[3], 3, material_conf_int$upper[3], 3,
angle=90, code=3, length=0.1, lwd=2, col="#EE7AE9")

points(material_conf_int$media[1], 1, pch=19, cex=1.1, col="#63B8FF")
points(material_conf_int$media[2], 2, pch=19, cex=1.1, col="red")
points(material_conf_int$media[3], 3, pch=19, cex=1.1, col="#EE7AE9")

abline(v = mean(material_conf_int$media), lty = 3, col = "red")
```

Intervalos de Confianza de Vibraciones por Materi

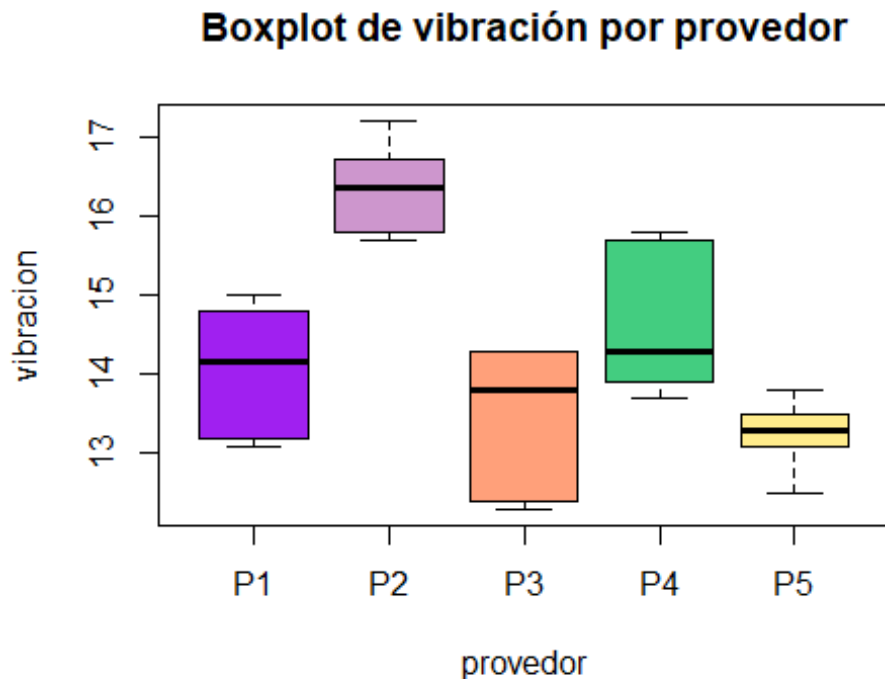


Interpreta el resultado desde la perspectiva estadística y en el contexto del problema.
Escribe tus conclusiones parciales.

Se observó que los 3 materiales producen un efecto diferente en la vibración. El efecto del Plástico es un material que disminuye la cantidad de vibraciones ya que su media es menor con respecto a la media general, el Aluminio aumenta la cantidad de vibraciones con respecto a la media general y el Acero también aumenta sin embargo es mínima la diferencia a la media general.

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

```
boxplot(vibracion ~ proveedor, col=c("purple", "#CD96CD", "#FFA07A",  
"#43CD80", "#FFEC8B"),  
main = "Boxplot de vibración por proveedor")
```

Haz los intervalos de confianza de vibración por proveedor.

```
proveedor_conf_int <- datos %>%
  group_by(proveedor) %>%
  summarise(media = mean(vibracion),
            sd = sd(vibracion),
            n = n()) %>%
  mutate(lower = media - qt(0.975, df=n-1) * sd / sqrt(n),
         upper = media + qt(0.975, df=n-1) * sd / sqrt(n))
proveedor_conf_int
```

```
## # A tibble: 5 × 6
##   proveedor media    sd      n lower upper
##   <fct>      <dbl> <dbl> <int> <dbl> <dbl>
## 1 P1         14.1 0.794     6  13.2  14.9
## 2 P2         16.4 0.561     6  15.8  16.9
## 3 P3         13.5 0.909     6  12.5  14.4
## 4 P4         14.6 0.911     6  13.7  15.6
## 5 P5         13.2 0.442     6  12.8  13.7
```

Grafícalos

```
plot(0, ylim=c(0, 5), xlim=c(min(proveedor_conf_int$lower),
max(proveedor_conf_int$upper) + 1),
     yaxt="n", ylab="", xlab = "Vibración", main = "Intervalos de
Confianza de Vibración por Proveedor")
axis(2, at=c(1,2,3,4,5), labels=proveedor_conf_int$proveedor)
```

```

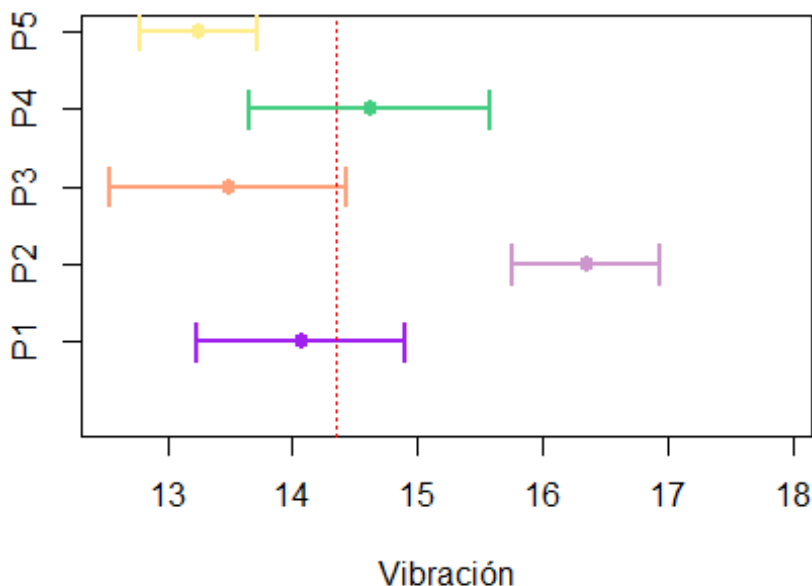
arrows(proveedor_conf_int$lower[1], 1, proveedor_conf_int$upper[1], 1,
angle=90, code=3, length=0.1, lwd=2, col="purple")
arrows(proveedor_conf_int$lower[2], 2, proveedor_conf_int$upper[2], 2,
angle=90, code=3, length=0.1, lwd=2, col="#CD96CD")
arrows(proveedor_conf_int$lower[3], 3, proveedor_conf_int$upper[3], 3,
angle=90, code=3, length=0.1, lwd=2, col="#FFA07A")
arrows(proveedor_conf_int$lower[4], 4, proveedor_conf_int$upper[4], 4,
angle=90, code=3, length=0.1, lwd=2, col="#43CD80")
arrows(proveedor_conf_int$lower[5], 5, proveedor_conf_int$upper[5], 5,
angle=90, code=3, length=0.1, lwd=2, col="#FFEC8B")

points(proveedor_conf_int$media[1], 1, pch=19, cex=1.1, col="purple")
points(proveedor_conf_int$media[2], 2, pch=19, cex=1.1, col="#CD96CD")
points(proveedor_conf_int$media[3], 3, pch=19, cex=1.1, col="#FFA07A")
points(proveedor_conf_int$media[4], 4, pch=19, cex=1.1, col="#43CD80")
points(proveedor_conf_int$media[5], 5, pch=19, cex=1.1, col="#FFEC8B")

abline(v = mean(proveedor_conf_int$media), lty = 3, col = "red")

```

Intervalos de Confianza de Vibración por Proveedor

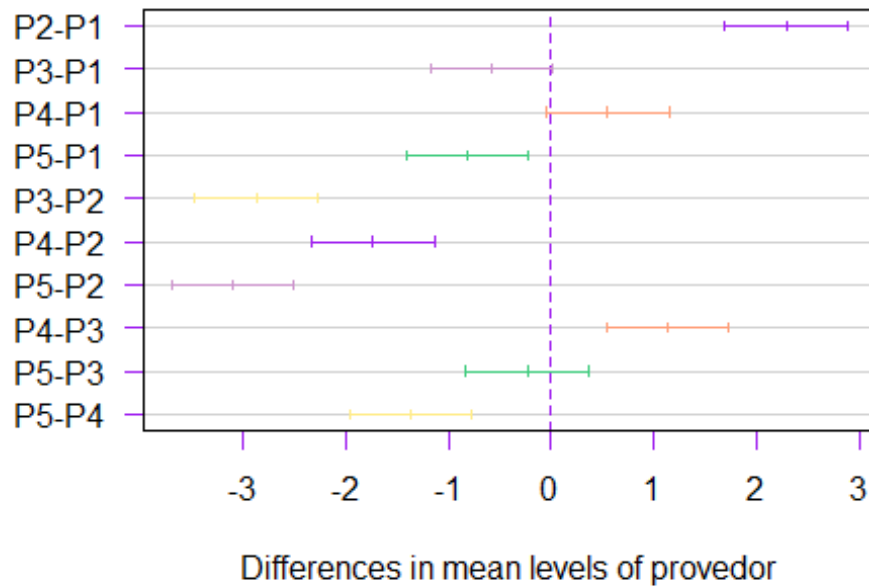


Se observó que los 5 proveedores producen un efecto diferente en la vibración, ninguno permanece en la media general. El proveedor 5 es el que más reduce la cantidad de vibraciones y los proveedores 1 y 3 igual reducen pero en menor proporción. Para el caso de los proveedores 4 y 2, ambos aumentan la cantidad de vibraciones, sin embargo el proveedor 2 es el que más aumenta la cantidad de vibraciones y el que mayor diferencia tiene con respecto a la media general.

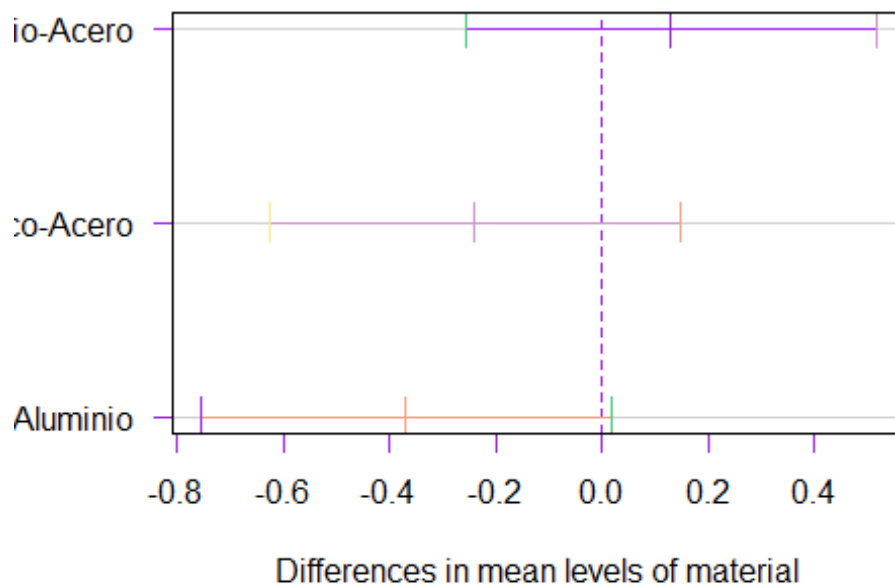
Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey.

```
tukey_test <- TukeyHSD(anova_interaccion)
plot(tukey_test, las=1, col=c("purple", "#CD96CD", "#FFA07A", "#43CD80",
"#FFEC8B"))
```

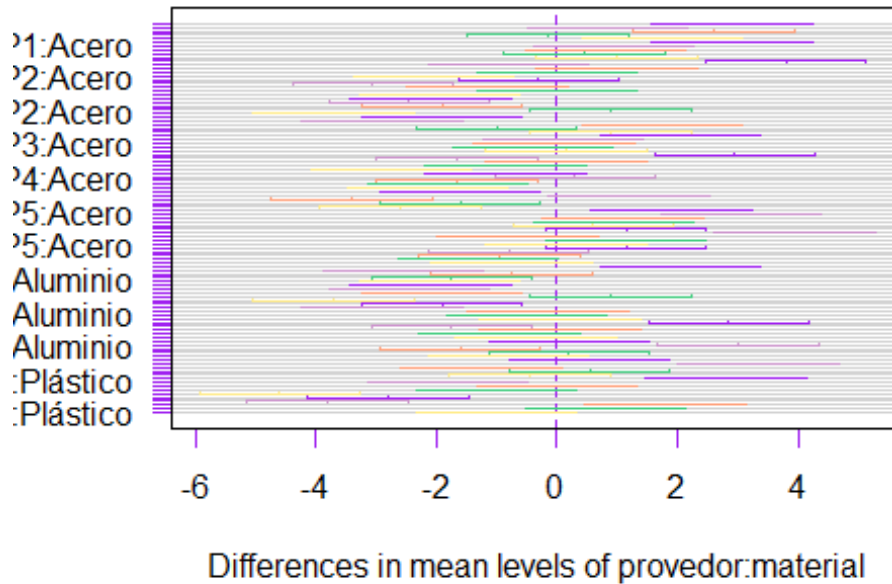
95% family-wise confidence level



95% family-wise confidence level



95% family-wise confidence level



tukey_test

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = vibracion ~ provedor * material)
##
## $provedor
##      diff      lwr      upr    p adj
## P2-P1  2.283333  1.688468  2.878199 0.000000
## P3-P1 -0.583333 -1.178199  0.011532 0.055877
## P4-P1  0.550000 -0.044865  1.144865 0.076690
## P5-P1 -0.816667 -1.411532 -0.221801 0.005471
## P3-P2 -2.866667 -3.461532 -2.271801 0.000000
## P4-P2 -1.733333 -2.328199 -1.138468 0.000017
## P5-P2 -3.100000 -3.694865 -2.505135 0.000000
## P4-P3  1.133333  0.538468  1.728199 0.000250
## P5-P3 -0.233333 -0.828199  0.361532 0.745281
## P5-P4 -1.366667 -1.961532 -0.771801 0.000313
##
## $material
##      diff      lwr      upr    p adj
## Aluminio-Acero  0.13 -0.257595  0.517595 0.665965
## Plástico-Acero -0.24 -0.627595  0.147595 0.272635
## Plástico-Aluminio -0.37 -0.757595  0.017595 0.062370
##
## $`provedor:material`
```

##	diff	lwr	upr	p adj
## P2:Acero-P1:Acero	2.90000e+00	1.5671172	4.23288275	0.0000207
## P3:Acero-P1:Acero	8.50000e-01	-0.4828828	2.18288275	0.4635832
## P4:Acero-P1:Acero	2.60000e+00	1.2671172	3.93288275	0.0000782
## P5:Acero-P1:Acero	-1.50000e-01	-1.4828828	1.18288275	0.9999999
## P1:Aluminio-P1:Acero	1.75000e+00	0.4171172	3.08288275	0.0053857
## P2:Aluminio-P1:Acero	2.90000e+00	1.5671172	4.23288275	0.0000207
## P3:Aluminio-P1:Acero	9.50000e-01	-0.3828828	2.28288275	0.3161401
## P4:Aluminio-P1:Acero	8.00000e-01	-0.5328828	2.13288275	0.5475696
## P5:Aluminio-P1:Acero	4.50000e-01	-0.8828828	1.78288275	0.9809255
## P1:Plástico-P1:Acero	1.00000e+00	-0.3328828	2.33288275	0.2556633
## P2:Plástico-P1:Acero	3.80000e+00	2.4671172	5.13288275	0.0000006
## P3:Plástico-P1:Acero	-8.00000e-01	-2.1328828	0.53288275	0.5475696
## P4:Plástico-P1:Acero	1.00000e+00	-0.3328828	2.33288275	0.2556633
## P5:Plástico-P1:Acero	1.24345e-14	-1.3328828	1.33288275	1.0000000
## P3:Acero-P2:Acero	-2.05000e+00	-3.3828828	-0.71711725	0.0011233
## P4:Acero-P2:Acero	-3.00000e-01	-1.6328828	1.03288275	0.9995788
## P5:Acero-P2:Acero	-3.05000e+00	-4.3828828	-1.71711725	0.0000110
## P1:Aluminio-P2:Acero	-1.15000e+00	-2.4828828	0.18288275	0.1270924
## P2:Aluminio-P2:Acero	0.00000e+00	-1.3328828	1.33288275	1.0000000
## P3:Aluminio-P2:Acero	-1.95000e+00	-3.2828828	-0.61711725	0.0018801
## P4:Aluminio-P2:Acero	-2.10000e+00	-3.4328828	-0.76711725	0.0008711
## P5:Aluminio-P2:Acero	-2.45000e+00	-3.7828828	-1.11711725	0.0001570
## P1:Plástico-P2:Acero	-1.90000e+00	-3.2328828	-0.56711725	0.0024397
## P2:Plástico-P2:Acero	9.00000e-01	-0.4328828	2.23288275	0.3857597
## P3:Plástico-P2:Acero	-3.70000e+00	-5.0328828	-2.36711725	0.0000009
## P4:Plástico-P2:Acero	-1.90000e+00	-3.2328828	-0.56711725	0.0024397
## P5:Plástico-P2:Acero	-2.90000e+00	-4.2328828	-1.56711725	0.0000207
## P4:Acero-P3:Acero	1.75000e+00	0.4171172	3.08288275	0.0053857
## P5:Acero-P3:Acero	-1.00000e+00	-2.3328828	0.33288275	0.2556633
## P1:Aluminio-P3:Acero	9.00000e-01	-0.4328828	2.23288275	0.3857597
## P2:Aluminio-P3:Acero	2.05000e+00	0.7171172	3.38288275	0.0011233
## P3:Aluminio-P3:Acero	1.00000e-01	-1.2328828	1.43288275	1.0000000
## P4:Aluminio-P3:Acero	-5.00000e-02	-1.3828828	1.28288275	1.0000000
## P5:Aluminio-P3:Acero	-4.00000e-01	-1.7328828	0.93288275	0.9929168
## P1:Plástico-P3:Acero	1.50000e-01	-1.1828828	1.48288275	0.9999999
## P2:Plástico-P3:Acero	2.95000e+00	1.6171172	4.28288275	0.0000168
## P3:Plástico-P3:Acero	-1.65000e+00	-2.9828828	-0.31711725	0.0091899
## P4:Plástico-P3:Acero	1.50000e-01	-1.1828828	1.48288275	0.9999999
## P5:Plástico-P3:Acero	-8.50000e-01	-2.1828828	0.48288275	0.4635832
## P5:Acero-P4:Acero	-2.75000e+00	-4.0828828	-1.41711725	0.0000398
## P1:Aluminio-P4:Acero	-8.50000e-01	-2.1828828	0.48288275	0.4635832
## P2:Aluminio-P4:Acero	3.00000e-01	-1.0328828	1.63288275	0.9995788
## P3:Aluminio-P4:Acero	-1.65000e+00	-2.9828828	-0.31711725	0.0091899
## P4:Aluminio-P4:Acero	-1.80000e+00	-3.1328828	-0.46711725	0.0041300
## P5:Aluminio-P4:Acero	-2.15000e+00	-3.4828828	-0.81711725	0.0006770
## P1:Plástico-P4:Acero	-1.60000e+00	-2.9328828	-0.26711725	0.0120178
## P2:Plástico-P4:Acero	1.20000e+00	-0.1328828	2.53288275	0.0991311
## P3:Plástico-P4:Acero	-3.40000e+00	-4.7328828	-2.06711725	0.0000027
## P4:Plástico-P4:Acero	-1.60000e+00	-2.9328828	-0.26711725	0.0120178

## P5:Plástico-P4:Acero	-2.60000e+00	-3.9328828	-1.26711725	0.0000782
## P1:Aluminio-P5:Acero	1.90000e+00	0.5671172	3.23288275	0.0024397
## P2:Aluminio-P5:Acero	3.05000e+00	1.7171172	4.38288275	0.0000110
## P3:Aluminio-P5:Acero	1.10000e+00	-0.2328828	2.43288275	0.1618380
## P4:Aluminio-P5:Acero	9.50000e-01	-0.3828828	2.28288275	0.3161401
## P5:Aluminio-P5:Acero	6.00000e-01	-0.7328828	1.93288275	0.8669272
## P1:Plástico-P5:Acero	1.15000e+00	-0.1828828	2.48288275	0.1270924
## P2:Plástico-P5:Acero	3.95000e+00	2.6171172	5.28288275	0.0000004
## P3:Plástico-P5:Acero	-6.50000e-01	-1.9828828	0.68288275	0.7989869
## P4:Plástico-P5:Acero	1.15000e+00	-0.1828828	2.48288275	0.1270924
## P5:Plástico-P5:Acero	1.50000e-01	-1.1828828	1.48288275	0.9999999
## P2:Aluminio-P1:Aluminio	1.15000e+00	-0.1828828	2.48288275	0.1270924
## P3:Aluminio-P1:Aluminio	-8.00000e-01	-2.1328828	0.53288275	0.5475696
## P4:Aluminio-P1:Aluminio	-9.50000e-01	-2.2828828	0.38288275	0.3161401
## P5:Aluminio-P1:Aluminio	-1.30000e+00	-2.6328828	0.03288275	0.0593800
## P1:Plástico-P1:Aluminio	-7.50000e-01	-2.0828828	0.58288275	0.6344630
## P2:Plástico-P1:Aluminio	2.05000e+00	0.7171172	3.38288275	0.0011233
## P3:Plástico-P1:Aluminio	-2.55000e+00	-3.8828828	-1.21711725	0.0000984
## P4:Plástico-P1:Aluminio	-7.50000e-01	-2.0828828	0.58288275	0.6344630
## P5:Plástico-P1:Aluminio	-1.75000e+00	-3.0828828	-0.41711725	0.0053857
## P3:Aluminio-P2:Aluminio	-1.95000e+00	-3.2828828	-0.61711725	0.0018801
## P4:Aluminio-P2:Aluminio	-2.10000e+00	-3.4328828	-0.76711725	0.0008711
## P5:Aluminio-P2:Aluminio	-2.45000e+00	-3.7828828	-1.11711725	0.0001570
## P1:Plástico-P2:Aluminio	-1.90000e+00	-3.2328828	-0.56711725	0.0024397
## P2:Plástico-P2:Aluminio	9.00000e-01	-0.4328828	2.23288275	0.3857597
## P3:Plástico-P2:Aluminio	-3.70000e+00	-5.0328828	-2.36711725	0.0000009
## P4:Plástico-P2:Aluminio	-1.90000e+00	-3.2328828	-0.56711725	0.0024397
## P5:Plástico-P2:Aluminio	-2.90000e+00	-4.2328828	-1.56711725	0.0000207
## P4:Aluminio-P3:Aluminio	-1.50000e-01	-1.4828828	1.18288275	0.9999999
## P5:Aluminio-P3:Aluminio	-5.00000e-01	-1.8328828	0.83288275	0.9578436
## P1:Plástico-P3:Aluminio	5.00000e-02	-1.2828828	1.38288275	1.0000000
## P2:Plástico-P3:Aluminio	2.85000e+00	1.5171172	4.18288275	0.0000257
## P3:Plástico-P3:Aluminio	-1.75000e+00	-3.0828828	-0.41711725	0.0053857
## P4:Plástico-P3:Aluminio	5.00000e-02	-1.2828828	1.38288275	1.0000000
## P5:Plástico-P3:Aluminio	-9.50000e-01	-2.2828828	0.38288275	0.3161401
## P5:Aluminio-P4:Aluminio	-3.50000e-01	-1.6828828	0.98288275	0.9979595
## P1:Plástico-P4:Aluminio	2.00000e-01	-1.1328828	1.53288275	0.9999962
## P2:Plástico-P4:Aluminio	3.00000e+00	1.6671172	4.33288275	0.0000136
## P3:Plástico-P4:Aluminio	-1.60000e+00	-2.9328828	-0.26711725	0.0120178
## P4:Plástico-P4:Aluminio	2.00000e-01	-1.1328828	1.53288275	0.9999962
## P5:Plástico-P4:Aluminio	-8.00000e-01	-2.1328828	0.53288275	0.5475696
## P1:Plástico-P5:Aluminio	5.50000e-01	-0.7828828	1.88288275	0.9202907
## P2:Plástico-P5:Aluminio	3.35000e+00	2.0171172	4.68288275	0.0000033
## P3:Plástico-P5:Aluminio	-1.25000e+00	-2.5828828	0.08288275	0.0768938
## P4:Plástico-P5:Aluminio	5.50000e-01	-0.7828828	1.88288275	0.9202907
## P5:Plástico-P5:Aluminio	-4.50000e-01	-1.7828828	0.88288275	0.9809255
## P2:Plástico-P1:Plástico	2.80000e+00	1.4671172	4.13288275	0.0000320
## P3:Plástico-P1:Plástico	-1.80000e+00	-3.1328828	-0.46711725	0.0041300
## P4:Plástico-P1:Plástico	0.00000e+00	-1.3328828	1.33288275	1.0000000
## P5:Plástico-P1:Plástico	-1.00000e+00	-2.3328828	0.33288275	0.2556633

```
## P3:Plástico-P2:Plástico -4.60000e+00 -5.9328828 -3.26711725 0.0000000
## P4:Plástico-P2:Plástico -2.80000e+00 -4.1328828 -1.46711725 0.0000320
## P5:Plástico-P2:Plástico -3.80000e+00 -5.1328828 -2.46711725 0.0000006
## P4:Plástico-P3:Plástico 1.80000e+00 0.4671172 3.13288275 0.0041300
## P5:Plástico-P3:Plástico 8.00000e-01 -0.5328828 2.13288275 0.5475696
## P5:Plástico-P4:Plástico -1.00000e+00 -2.3328828 0.33288275 0.2556633
```

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

Escribe tus conclusiones parciales

```
mean(proveedor_conf_int$media)
```

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

Las medias muestrales nos dieron que $\mu_1 = 14.06$, $\mu_2 = 16.35$, $\mu_3 = 13.48$, $\mu_4 = 14.61$, $\mu_5 = 13.25$ y nuestra media poblacional es $\mu = 14.35$ por lo que los efectos para cada nivel son $M_1 = -0.29$, $M_2 = 2$, $M_3 = -0.87$, $M_4 = 0.26$, $M_5 = -1.1$

6. Comprueba la validez del modelo.

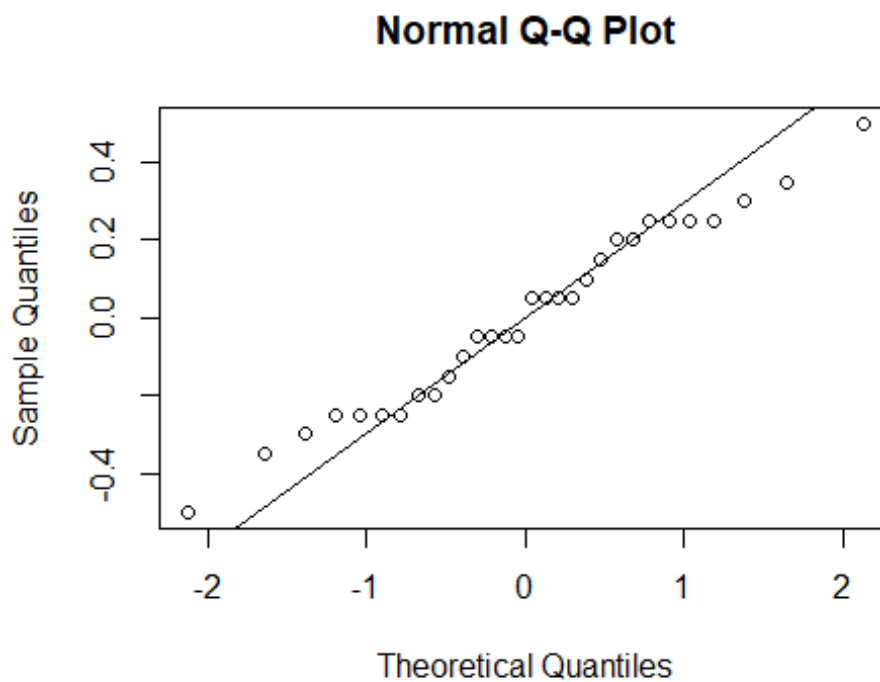
Comprueba:

Normalidad

```
residuos = anova_interaccion$residuals
```

```
qqnorm(residuos)
```

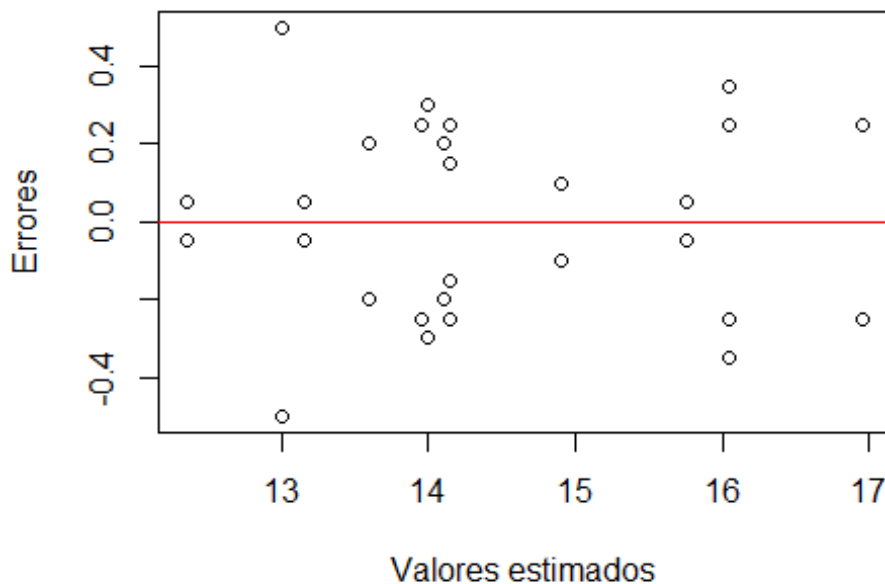
```
qqline(residuos)
```



La gráfica de Normal Q-Q muestra que los residuos siguen aproximadamente la línea diagonal, lo cual sugiere que los residuos se distribuyen de manera cercana a una distribución normal. Sin embargo, algunos puntos se desvían ligeramente en los extremos, lo que indica posibles ligeras desviaciones de la normalidad.

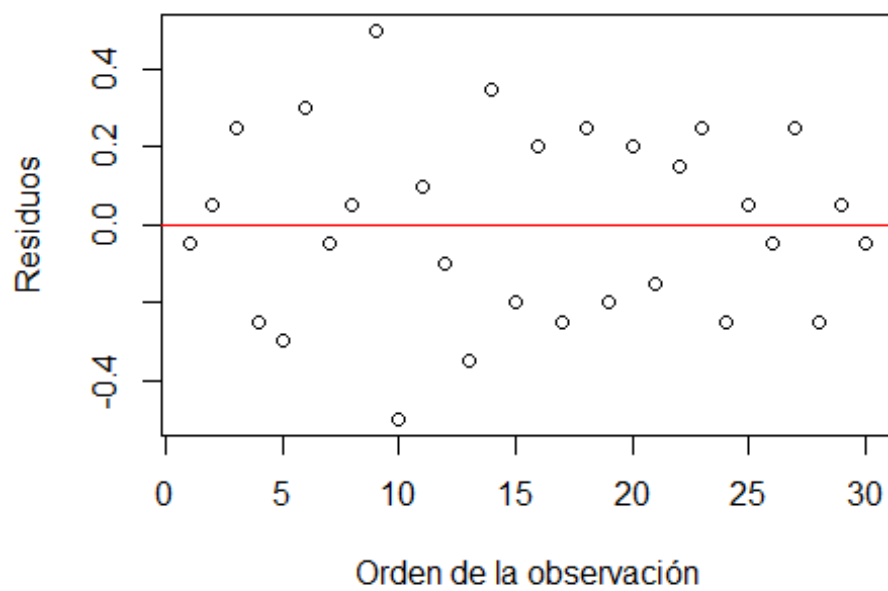
Homocedasticidad

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



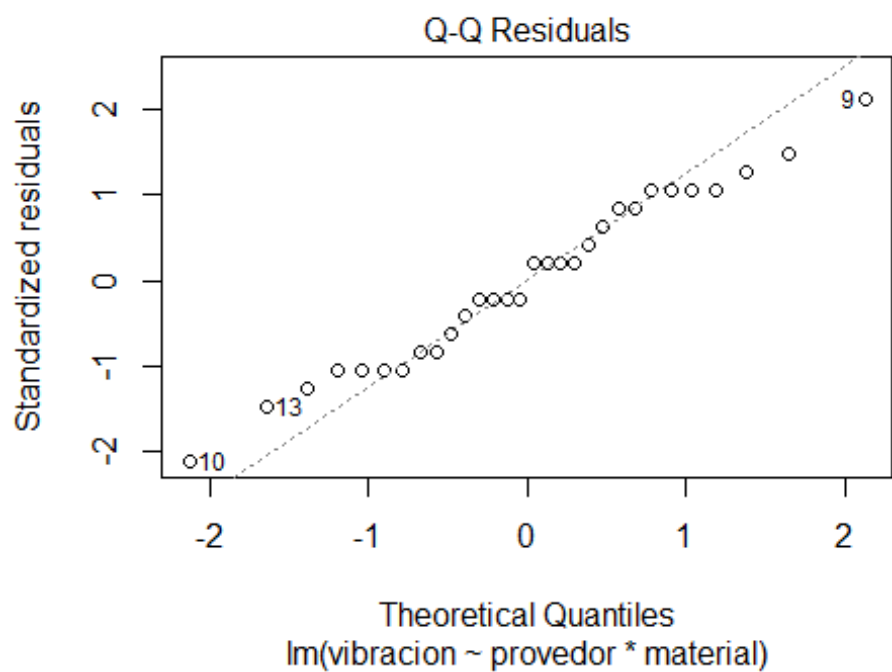
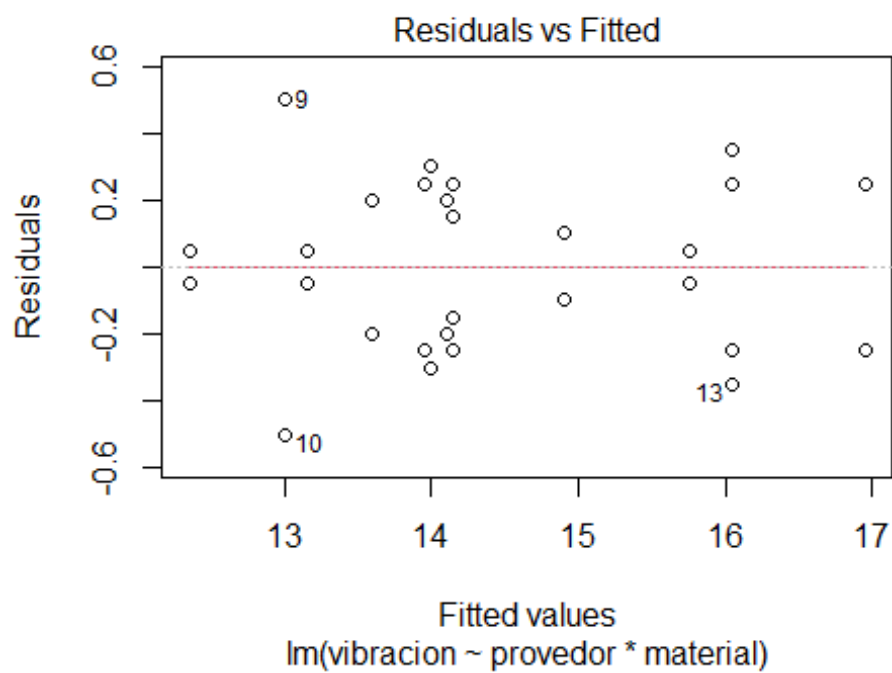
Independencia

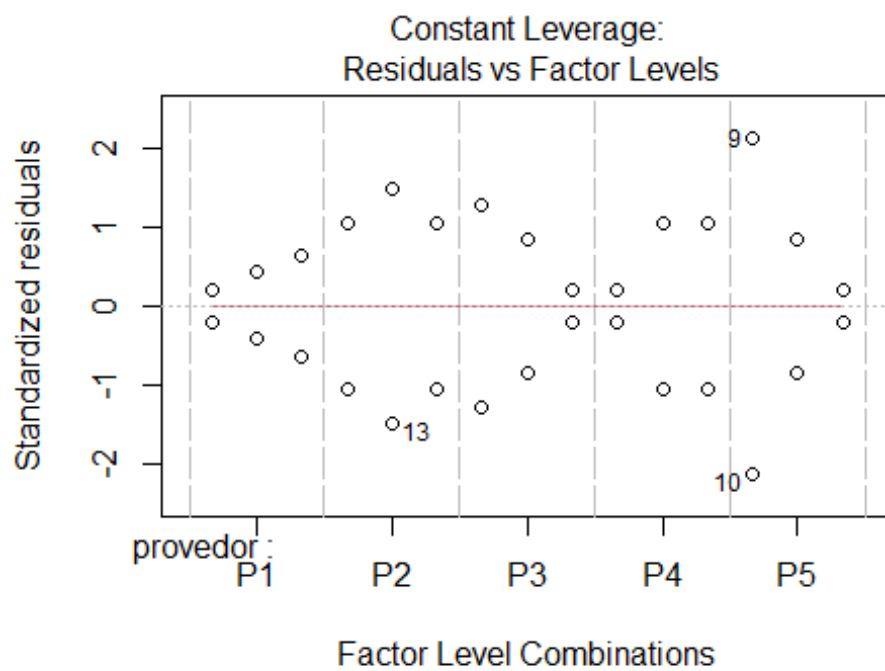
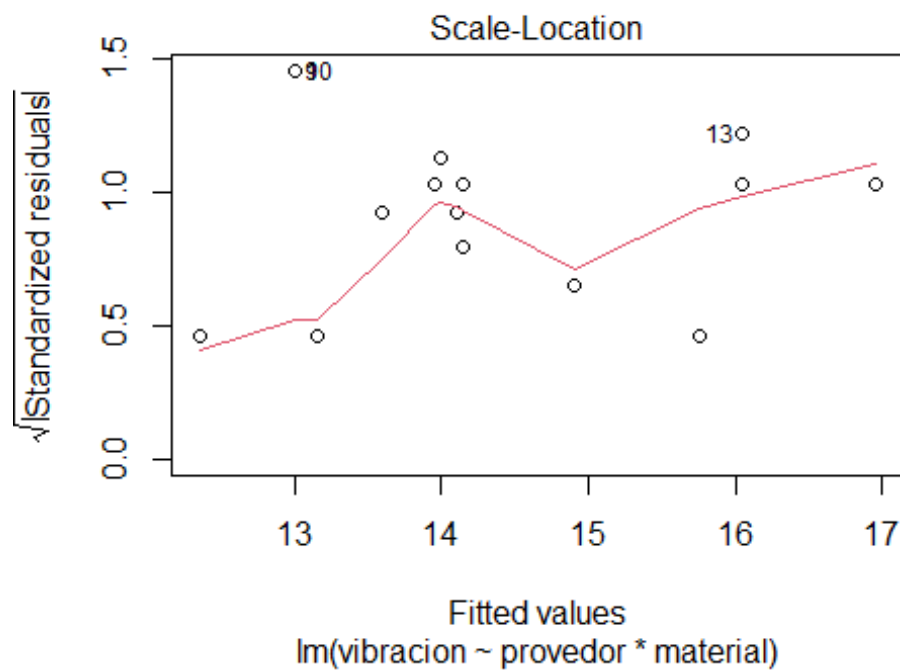
```
n = tapply(vibracion, proveedor:material, length)
plot(c(1:sum(n)), anova_interaccion$residuals, xlab="Orden de la observación", ylab="Residuos")
abline(h=0, col="red")
```



Relación lineal entre las variables (coeficiente de determinación).

```
plot(lm(vibracion~proveedor*material))
```





```
CD= (36.67+.7+11.61)/(36.67+.7+11.61+1.67) #coeficiente de determinación
para el modelo
CD
```

[1] 0.9670286

Concluye en el contexto del problema.

Varianza Explicada por el modelo: 96.7% (SCTratamiento/SCtotales)

Efectos significativos: ¿Hubo efectos significativo? Si así es: ¿todos son diferentes? ¿hay grupos ¿qué significa en el contexto del problema? ¿Existe evidencia de que el proveedor y el material son causas de la vibración de los motores eléctricos? Se observó que los 5 proveedores producen un efecto diferente en la cantidad de vibraciones de los motores eléctricos. El proveedor 2 es el mayor causante de vibraciones en los motores eléctricos, mientras que el proveedor 5 es el que más reduce la cantidad de vibraciones.

Asimismo, se observó que los 3 materiales producen un efecto diferente en la vibración. El plástico disminuye la cantidad de vibraciones, el Acero aumenta ligeramente las vibraciones y el Aluminio aumenta en mayor proporción la cantidad de vibraciones, sin embargo podemos observar que para ninguno de los 3 materiales la diferencia con respecto a la media general es grande, lo que nos indica que no hay mucha diferencia, y nuestro valor p es mayor a 0.05 sin embargo en el anova para dos niveles con interacción el efecto influye en la interacción. Por este motivo es que no tenemos la suficiente información para rechazar o aceptar la hipótesis.

El modelo explica el 96.7% de la variación. Por lo tanto, el Proveedor y el material son factores determinante en las causas de la vibración de los motores eléctricos, 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).

Hipótesis 1

- $H_0: \tau_i = 0$ No hay efecto de los proveedores <- Se Rechaza

Hipótesis 2

- $H_0: \alpha_j = 0$ No hay efecto en los materiales. <- No tenemos suficiente información para rechazar o aceptar la hipótesis, el valor p es mayor a 0.05 pero el efecto influye en la interacción.

Hipótesis 3

- $H_0: \tau_i * \alpha_j = 0$ No hay interacción entre el material y el proveedor en cuanto a la vibración. <- Se rechaza