

ANOVA

Facundo Colasurdo Caldironi

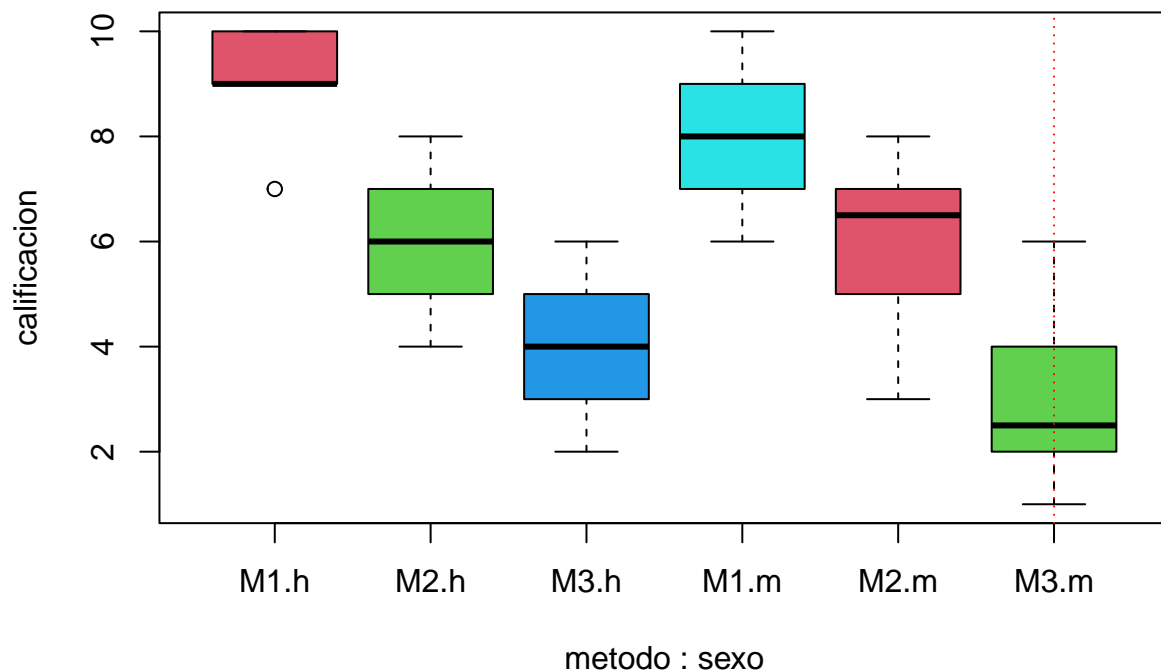
2024-08-27

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

Consulta el código en R en los apoyos de clase de “ANOVA” Haz el boxplot de resistencia a la tensión por concentración de madera dura. Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Escribe tus conclusiones parciales

```
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)
```

```
boxplot(calificacion ~ metodo:sexo, col = 2:5)
abline(v = mean(calificacion), lty = 3, col = "red")
```



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

M1.h y M1.m son los que tienen un incremento positivo, por lo tanto, se puede decir que tienen un mejor resultado, específicamente, el método uno a comparación del método dos y tres, nos da los resultados más a la derecha de la gráfica, justificando la anterior afirmación.

Escribe tus conclusiones parciales El método uno es el mejor de los tres disponibles métodos, ya que es el que da el mejor resultado positivo.

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

Primera hipótesis: $H_0: T_i = 0$ H_1 : algún T_i es distinto a 0

Segunda hipótesis: $H_0: A_j = 0$ H_1 : algún A_j es distinto a 0

Tercera hipótesis: $H_0: T_i A_j = 0$ H_1 : algún $T_i A_j$ es distinto a 0

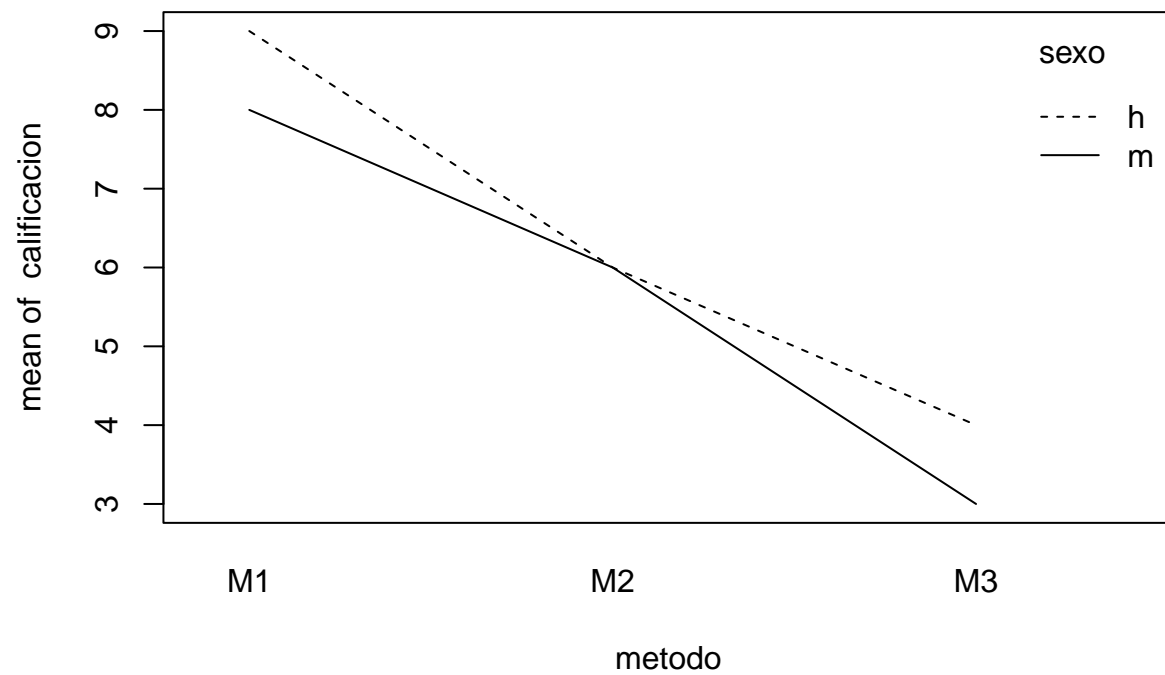
Realiza el ANOVA para dos niveles con interacción:

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

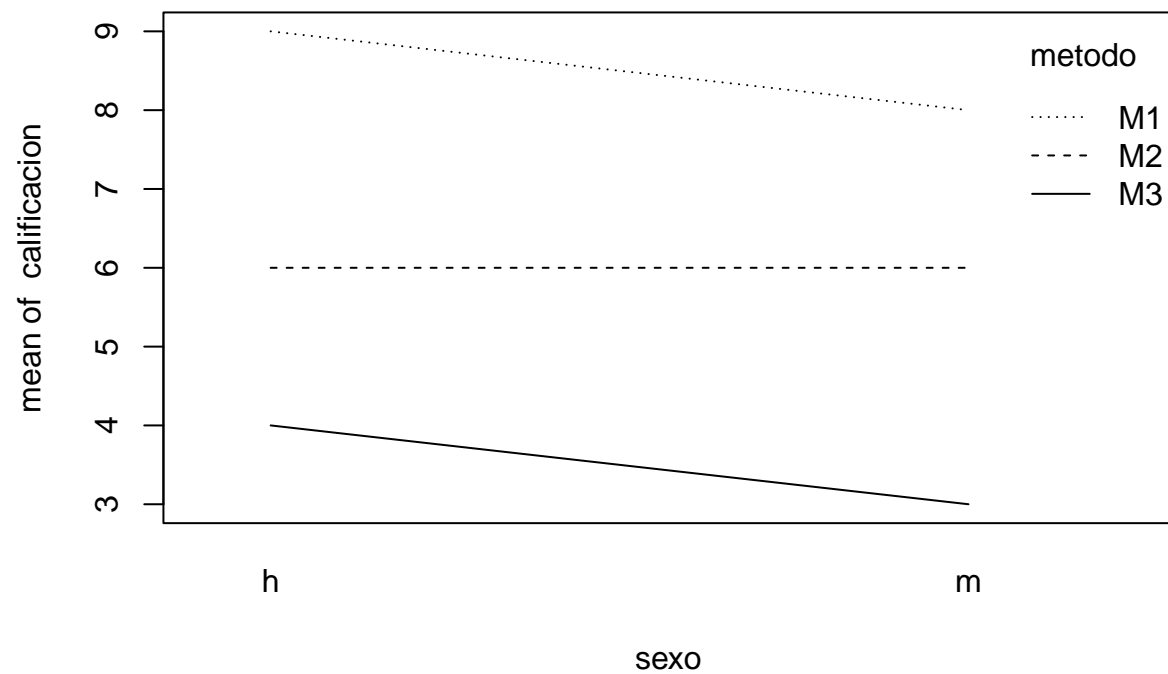
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## metodo	2	150	75.00	32.143	3.47e-08 ***
## sexo	1	4	4.00	1.714	0.200

```
## 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
```

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

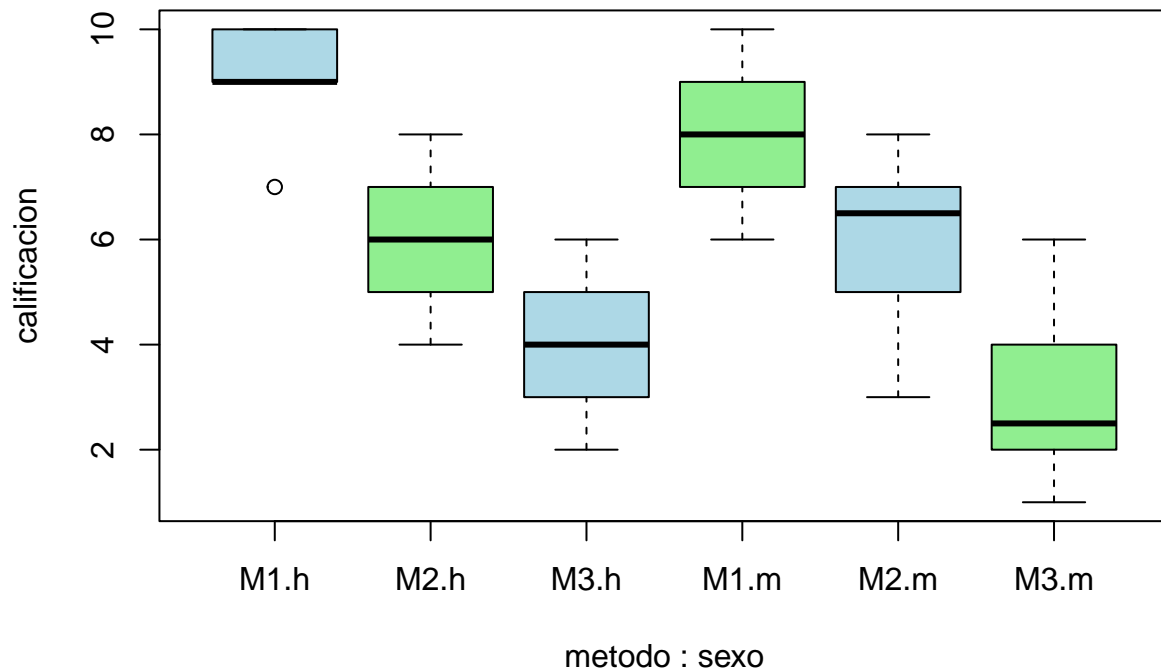


```
interaction.plot(sexo, metodo, calificacion)
```



```
boxplot(calificación ~ metodo * sexo, data = datos, col = c("lightblue", "lightgreen"), main = "Boxplot")
```

Boxplot de calificacion por metodo y sexo



valor frontera

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

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

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

En las graficas de interaccion, es posible ver que los hombres siguen una linea de tendencia mas lineal, mientras que el metodo dos tiene un pico de interaccion, pero realmente no atraviesa para generar un cambio, por otra parte, tambien podemos ver en la otra grafica de interaccion de los modelos, el modelo 1 sigue siendo la mejor opcion de los resultados. Por otra parte en el ANOVA es posible ver que existen diferencia de los valores de los resultados entre los distintos metodos de enseñanza, sin embargo, el sexo no parece influir en las calificaciones, y la efectividad entre hombres y mujeres.

Escribe tus conclusiones parciales Los metodos son los que mayor influencia tienen en los resultados de los alumnos, por otra parte, se reconfirma que el metodo 1 es el mejor metodo de los tres disponibles.

Realiza el ANOVA para dos niveles sin interacción.

Consulta el código de R en los apoyos de clase de “ANOVA” Haz el boxplot de rendimiento por sexo. Calcula la media para el rendimiento por sexo y método. Haz los intervalos de confianza de rendimiento por sexo. Graficalos Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Escribe tus conclusiones parciales

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

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## metodo      2    150   75.00  33.333 1.5e-08 ***
## sexo        1      4    4.00   1.778  0.192
## Residuals   32     72    2.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

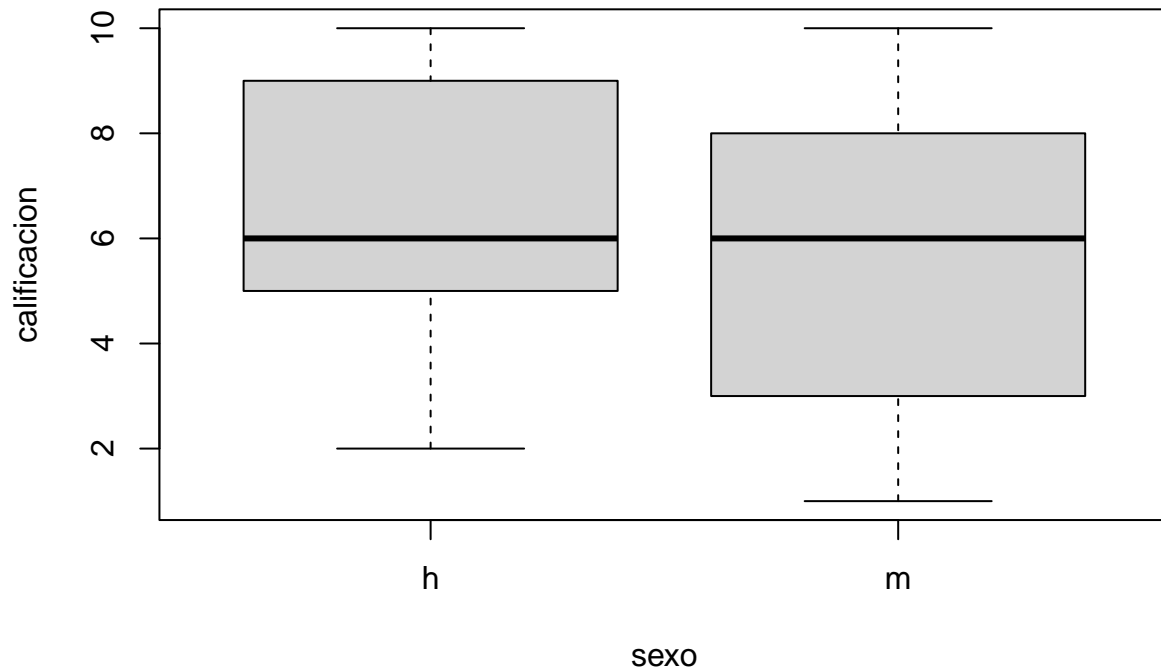
```
tapply(calificacion,sexo,mean)
```

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

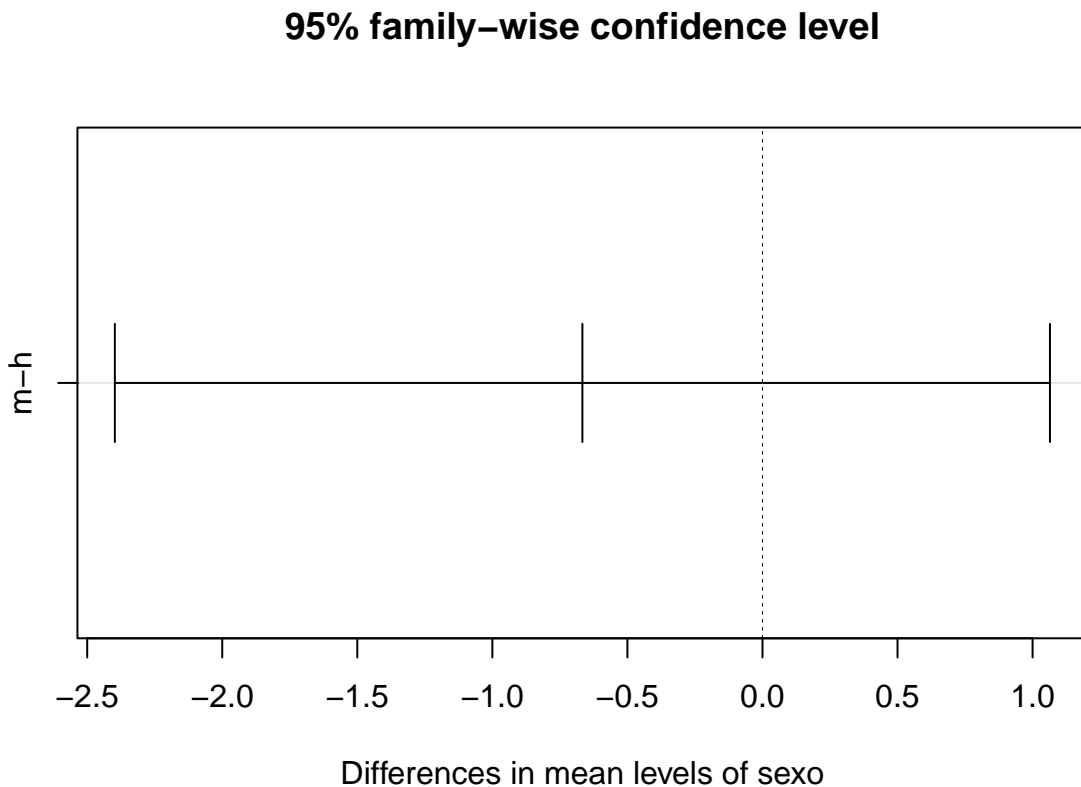
```
tapply(calificacion,metodo,mean)
```

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

```
M=mean(calificacion)
boxplot(calificacion ~ sexo)
```



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



Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Los resultados de los tres metodos nos demuestran que el primer metodo, tiene un mejor resultado, liderando con 8.5, seguido por el metodo dos, por otra parte, se denota en la grafica de boxplot que realmente no hay gran diferencia en los resultados debido al genero, por lo que puede entenderse que el genero no afecta en gran medida.

Escribe tus conclusiones parciales El resultado nos demuestra que los metodos de enseñanza si toman un importante rol en los resultados de los alumnos, mientras que el genero realmente no tiene impacto alguno.

Realiza el ANOVA para un efecto principal

Consulta el código de R en los apoyos de clase de “ANOVA” Haz el boxplot de rendimiento por método de enseñanza. Calcula la media. Haz los intervalos de confianza de rendimiento por método. Gráficalos Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey. Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Escribe tus conclusiones parciales

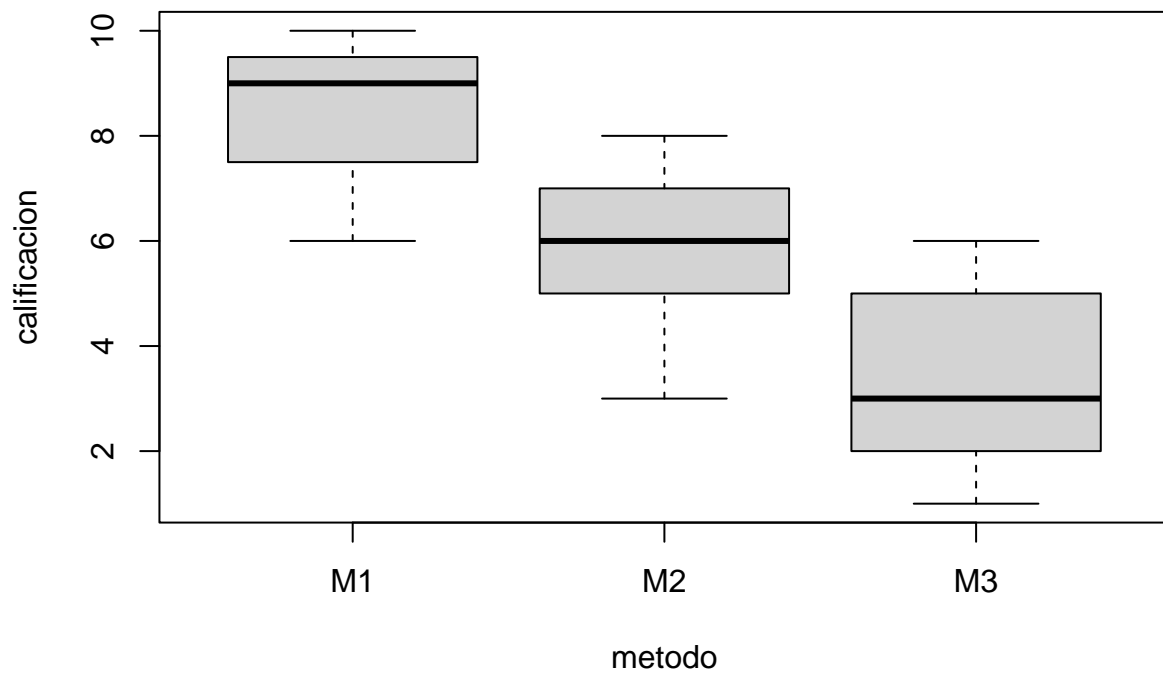
```
B<-aov(calificacion~metodo)
summary(B)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## metodo      2    150    75.0    32.57 1.55e-08 ***
## Residuals   33     76     2.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

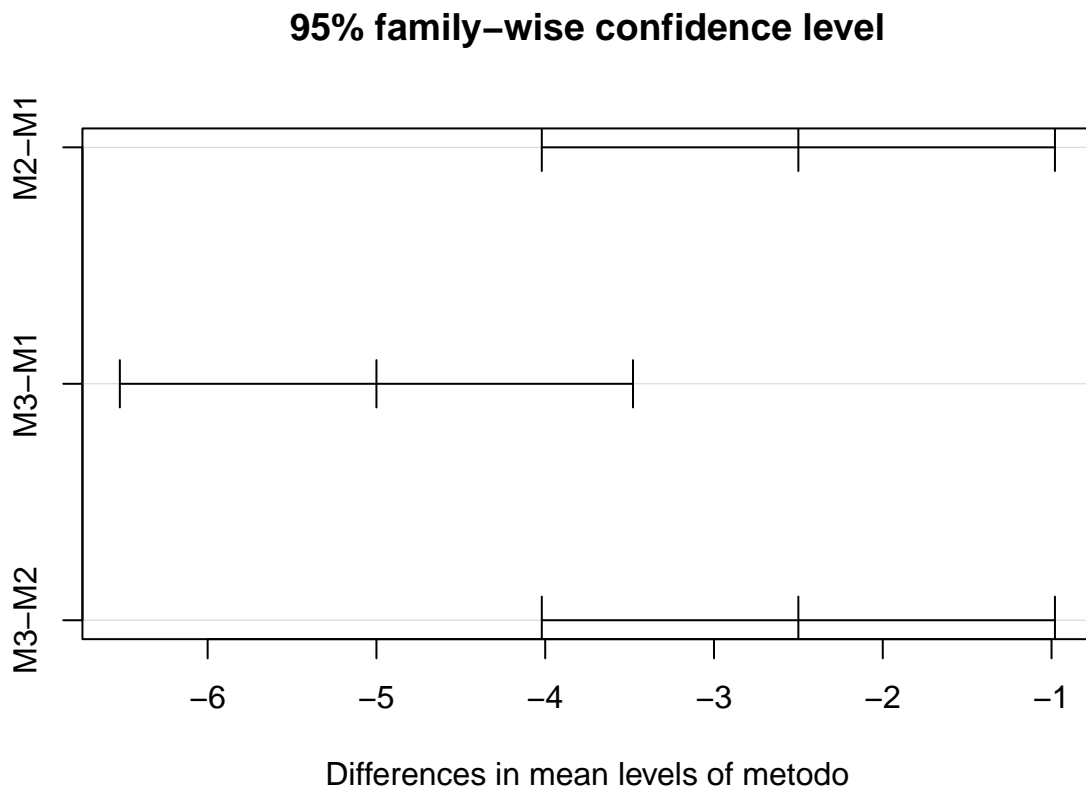
```
tapply(calificacion,metodo,mean)
```

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

```
M=mean(calificacion)  
boxplot(calificacion ~ metodo)
```



```
I = TukeyHSD(aov(calificacion ~ metodo))  
plot(I)
```

Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Se puede denotar la diferencia entre el primer metodo sobre el resto, por lo tanto, es facil decir que es el mas apto para ser usado.

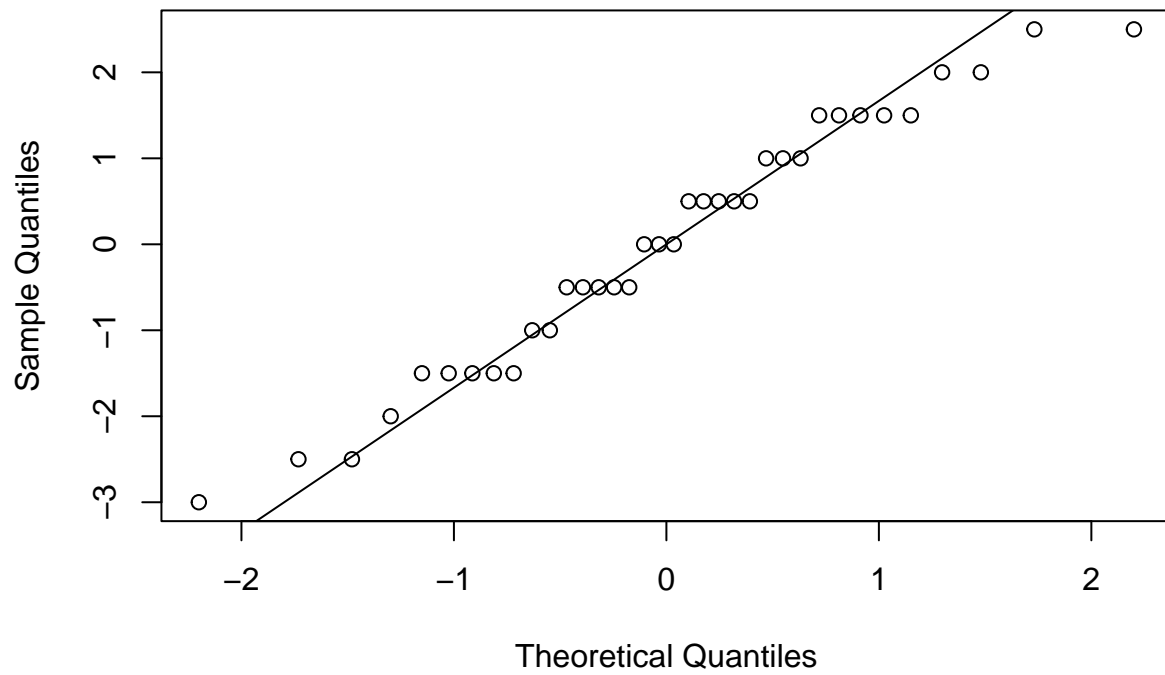
Escribe tus conclusiones parciales El resultado nos demuestra que el metodo uno es el mas efectivo, a comparacion del metodo tres, el cual es el menos efectivo.

Comprueba la validez del modelo. Comprueba:

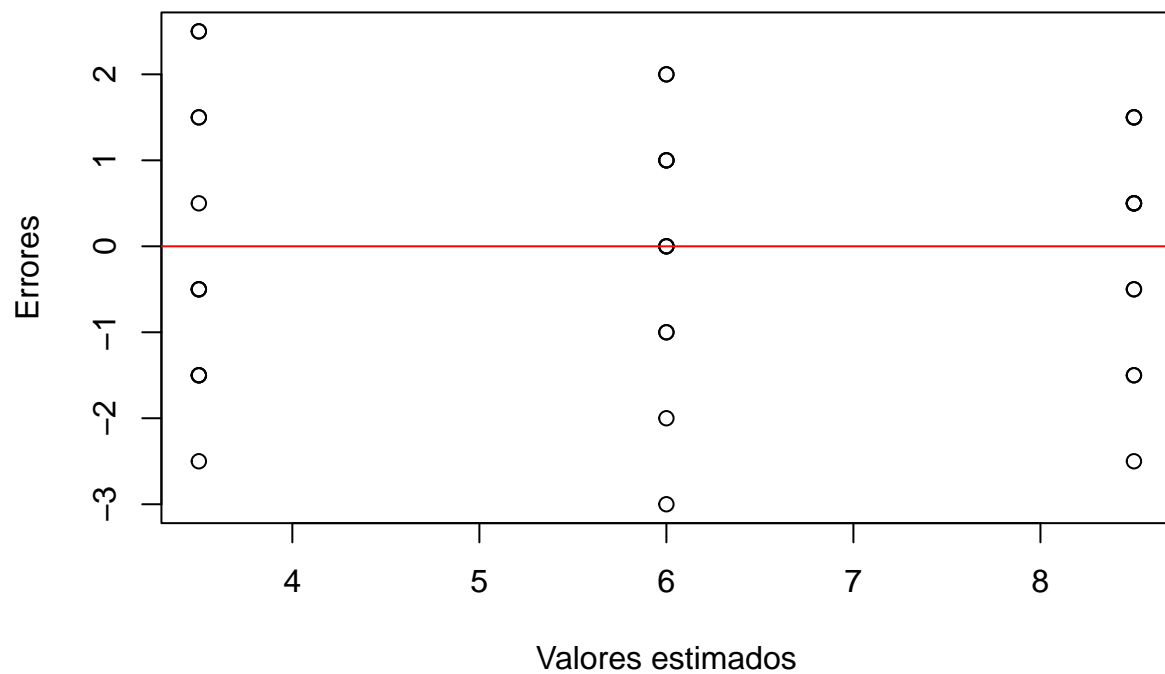
Normalidad Homocedasticidad Independencia Relación lineal entre las variables (coeficiente de determinación).

```
# Normalidad
anova <- aov(calificacion ~ metodo)
residuos <- anova$residuals
qqnorm(residuos)
qqline(residuos)
```

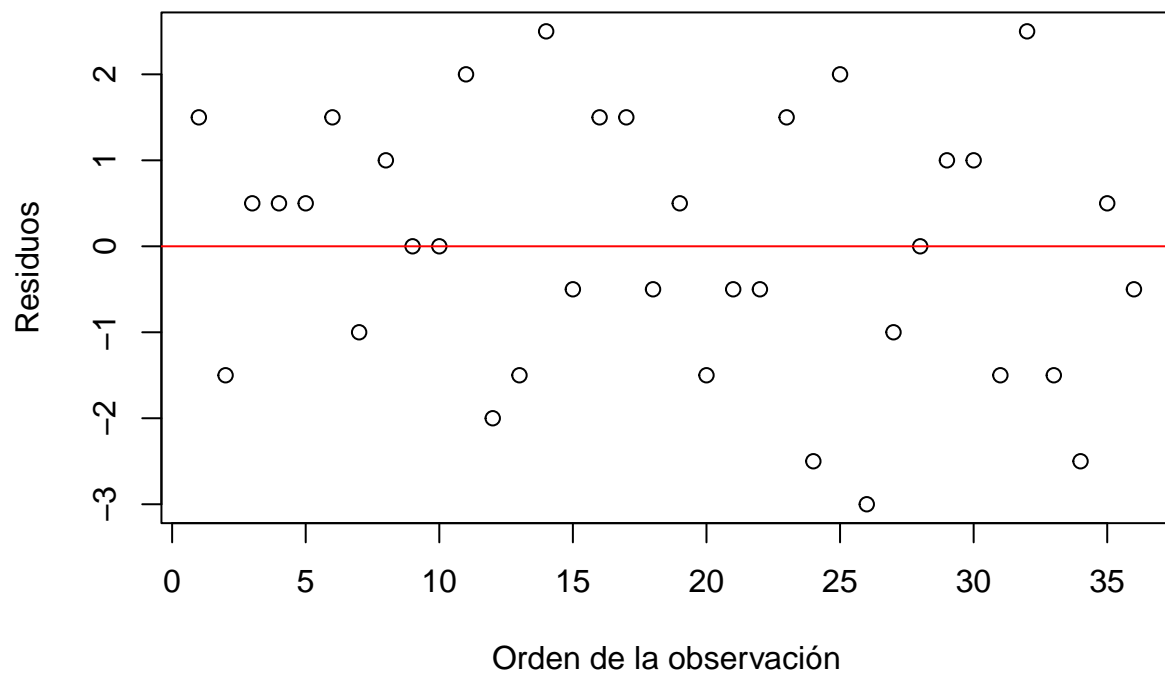
Normal Q-Q Plot



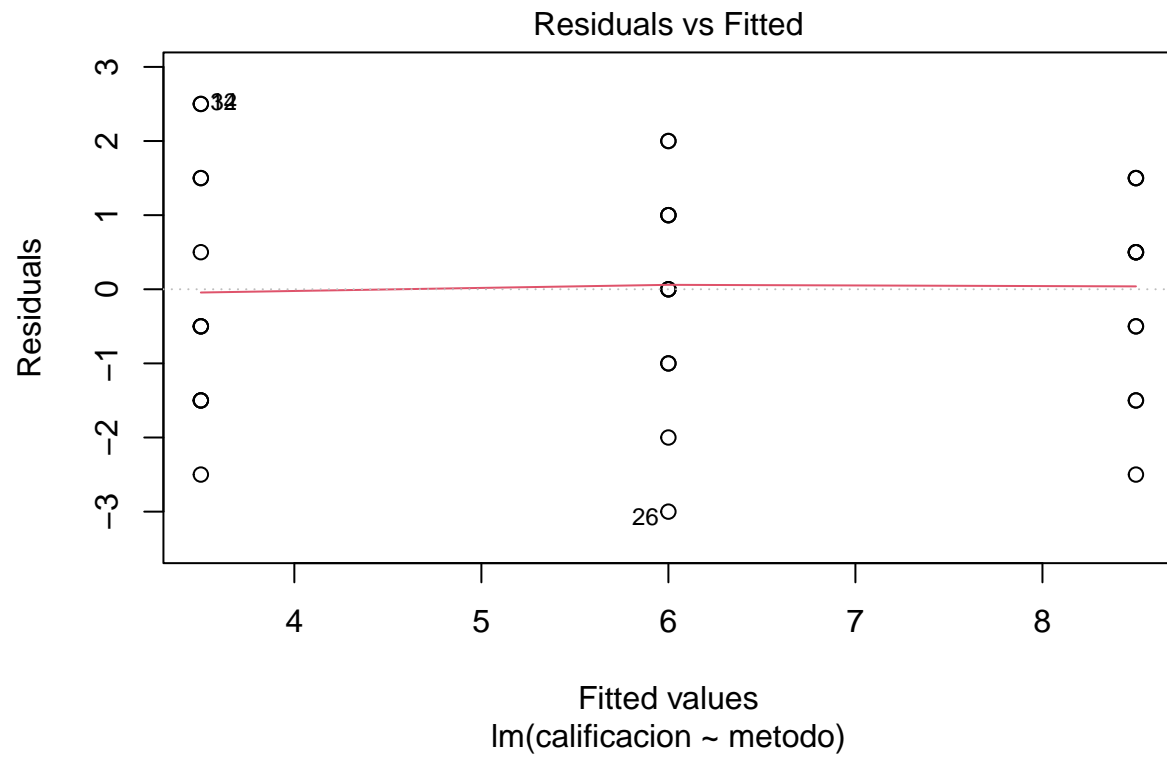
```
# Homocedasticidad  
plot(anova$fitted.values, anova$residuals, ylab="Errores", xlab="Valores estimados")  
abline(h=0, col="red")
```

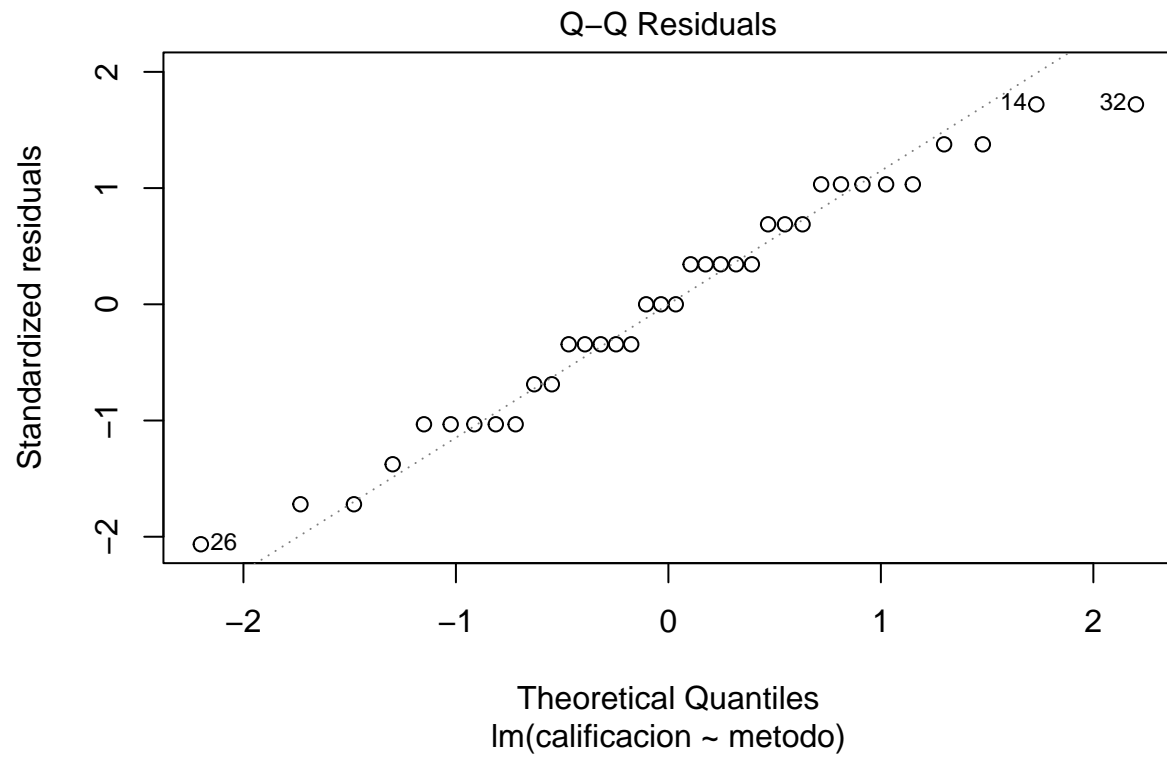


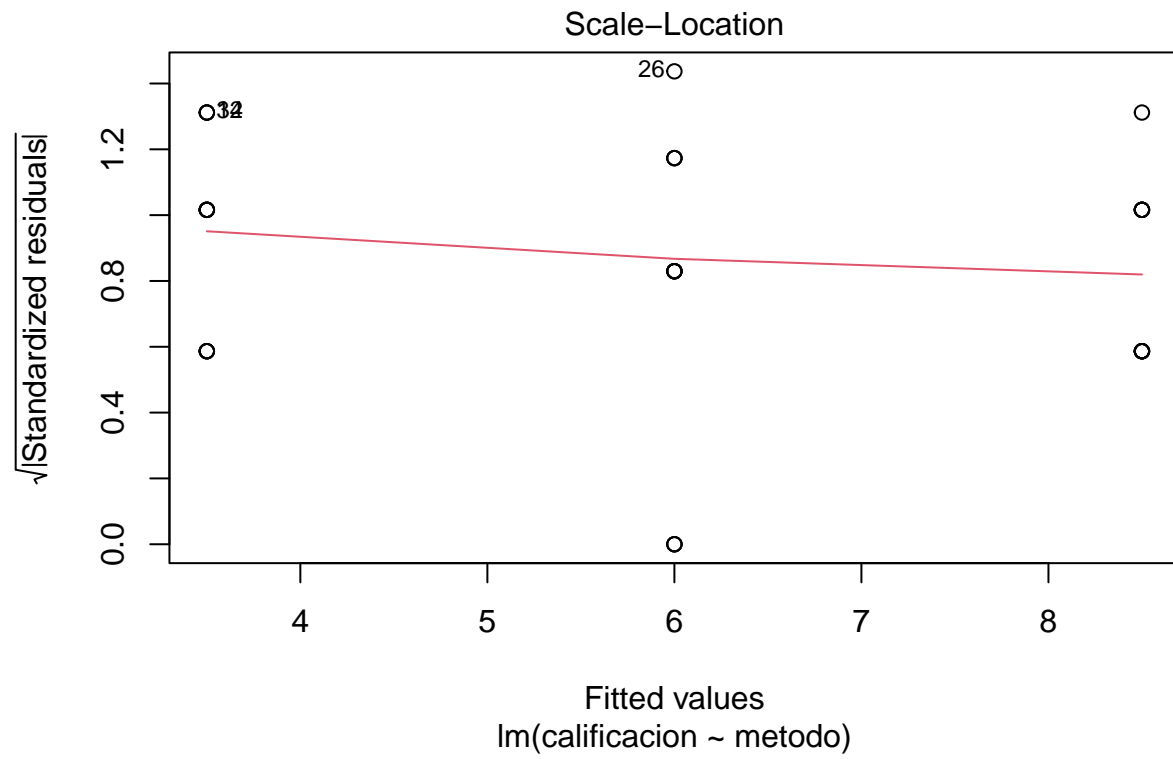
```
# Independencia
n = tapply(calificacion, metodo, length)
plot(c(1:sum(n)), anova$residuals, xlab="Orden de la observación", ylab="Residuos")
abline(h=0, col="red")
```

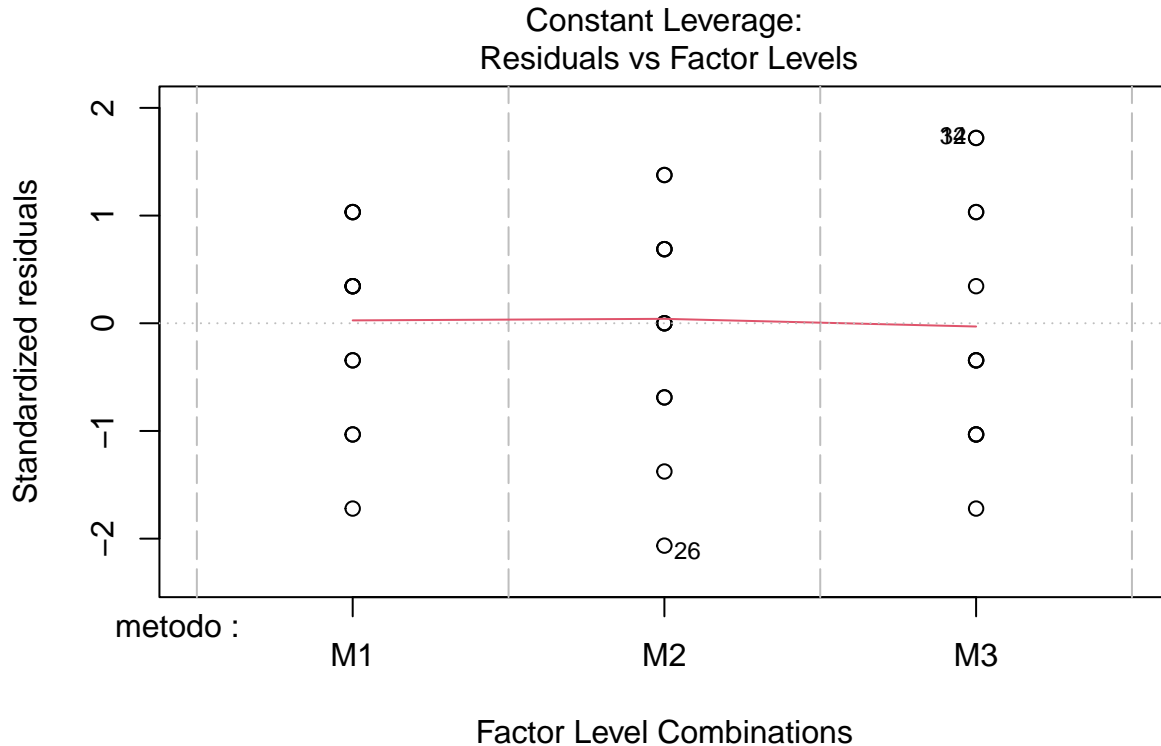


```
plot(lm(calificacion~metodo))
```









```
CD= 150/(150+76) #coeficiente de
summary(lm(calificacion~metodo))$r.squared
```

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

Concluye en el contexto del problema.

Se puede concluir que los tres métodos dan resultados diferentes con los estudiantes, siendo el mejor de los tres el método 1, ya que es el que mejora el rendimiento de los mismos a respecto de la media general, a comparación del método 2 y 1, los cuales no tuvieron un resultado tan favorable con los alumnos. El método explica el 0.6637 de la varianza, es decir, el 66.37% de la variación total, por lo que podemos decir que toma un papel muy importante en los resultados de los alumnos, aunque el porcentaje restante puede ser afectado por otros factores, pero en este caso, se le atribuye a la aleatoriedad de los datos. Al analizar los gráficos q-q y de residuos por valor ajustado, estos datos cumplen con normalidad, al mismo tiempo que los errores cumplen con una media cero.

Problema 2

Repita 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

Consulta el código en R en los apoyos de clase de “ANOVA” Haz el boxplot de resistencia a la tensión por concentración de madera dura. Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Escribe tus conclusiones parciales

```
vibracion = c(13.1, 13.2, 15.0, 14.8, 14.0, 14.3, 16.3, 15.8, 15.7, 16.4, 17.2, 16.7, 13.7, 14.3, 13.9,
proveedor = c(rep("1", 6), rep("2", 6), rep("3", 6), rep("4", 6), rep("5", 6))
material = c(rep("acero", 2), rep("aluminio", 2), rep("plastico", 2))

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

datos = data.frame(vibracion, proveedor, material)

print(datos)
```

```
##      vibracion proveedor 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
## 7         16.3         2      acero
## 8         15.8         2      acero
## 9         15.7         2 aluminio
## 10        16.4         2 aluminio
## 11        17.2         2  plastico
## 12        16.7         2  plastico
## 13        13.7         3      acero
## 14        14.3         3      acero
## 15        13.9         3 aluminio
## 16        14.3         3 aluminio
## 17        12.4         3  plastico
## 18        12.3         3  plastico
## 19        15.7         4      acero
## 20        15.8         4      acero
## 21        13.7         4 aluminio
## 22        14.2         4 aluminio
## 23        14.4         4  plastico
## 24        13.9         4  plastico
## 25        13.5         5      acero
## 26        12.5         5      acero
## 27        13.4         5 aluminio
## 28        13.8         5 aluminio
## 29        13.2         5  plastico
## 30        13.1         5  plastico
```

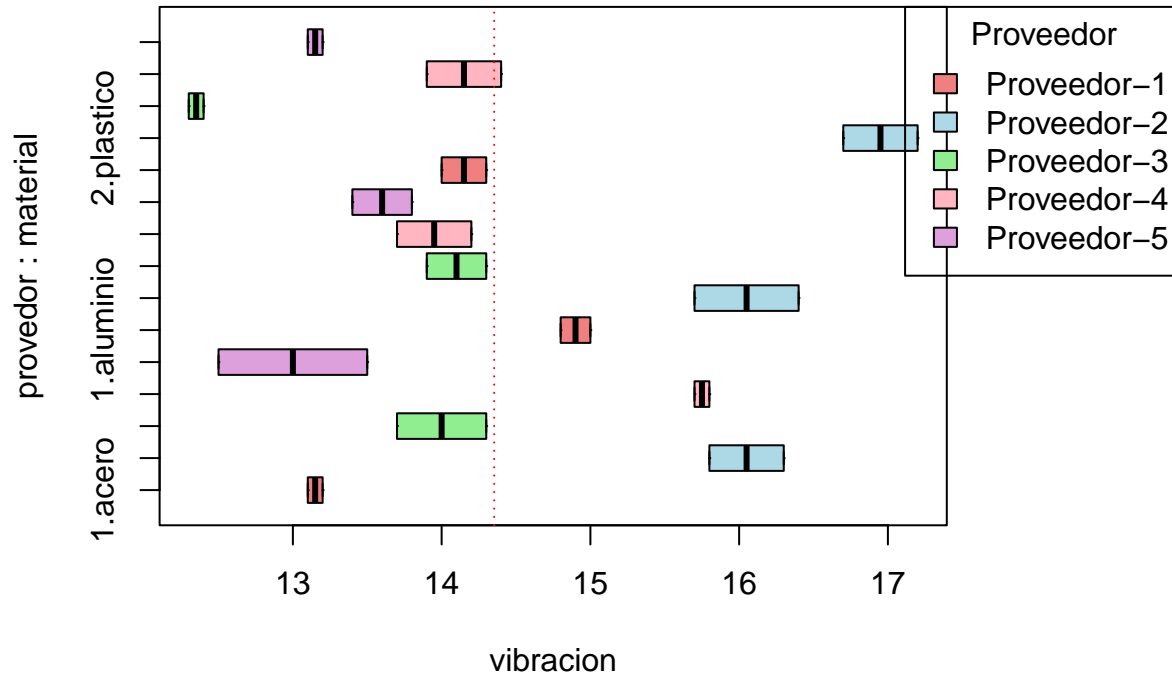
```
par(mar = c(5, 4, 4, 8))

datos = data.frame(vibracion, proveedor, material)
```

```

boxplot(vibracion ~ proveedor : material, datos, col = c("lightcoral", "lightblue", "lightgreen", "lightpink", "plum"),
abline(v = mean(vibracion), lty= 3, col = "red")
legend("topright", legend = c("Proveedor-1", "Proveedor-2", "Proveedor-3", "Proveedor-4", "Proveedor-5"),
      fill = c("lightcoral", "lightblue", "lightgreen", "lightpink", "plum"),
      title = "Proveedor", inset = c(-0.3, 0), xpd = TRUE)

```



Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Podemos ver que el resultado nos dice que realmente los materiales se mantienen por el mismo rango, alrededor de 14.3, mientras que el valor que hace diferencia es el proveedor dos, el cual tiene los mayores materiales con mayor cantidad de vibración, entre 16 a 17,

Escribe tus conclusiones parciales Podemos decir que el proveedor dos es el cual genera los materiales que tienen mayor vibración dentro de los motores eléctricos.

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

Primera hipótesis: $H_0: A = 0$ H_1 : algun A es distinto a 0

Segunda hipótesis: $H_0: B = 0$ H_1 : algun B es distinto a 0

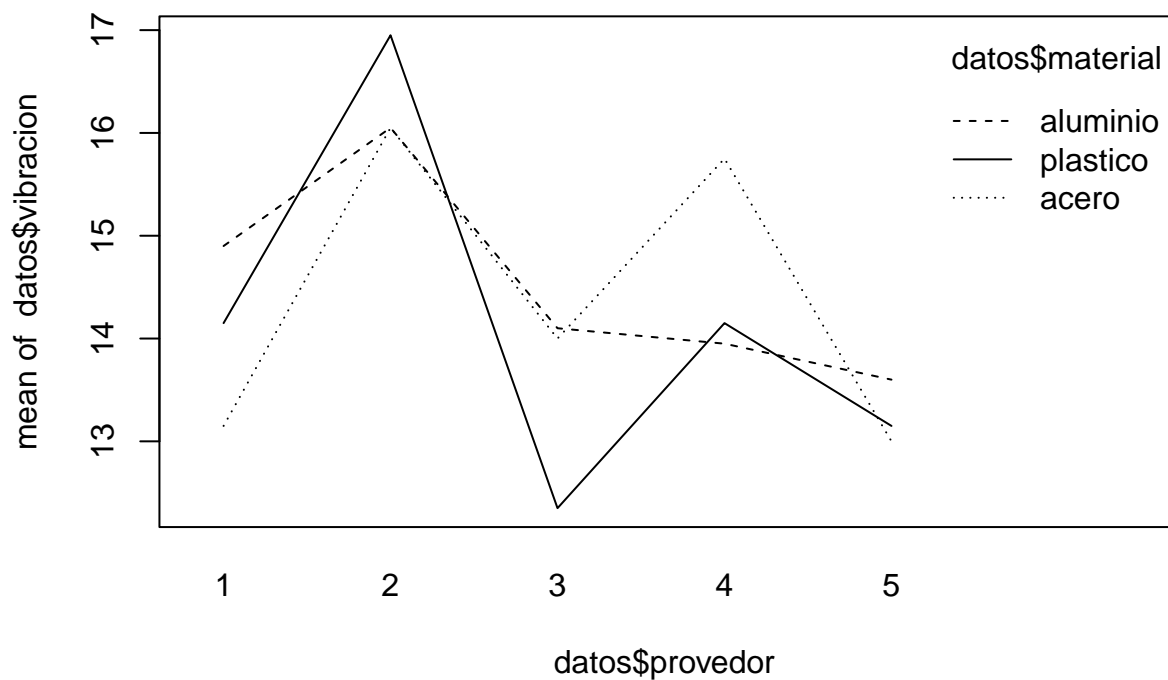
Tercera hipótesis: $H_0: A \times B = 0$ H_1 : algun $A \times B$ es distinto a 0

Realiza el ANOVA para dos niveles con interacción:

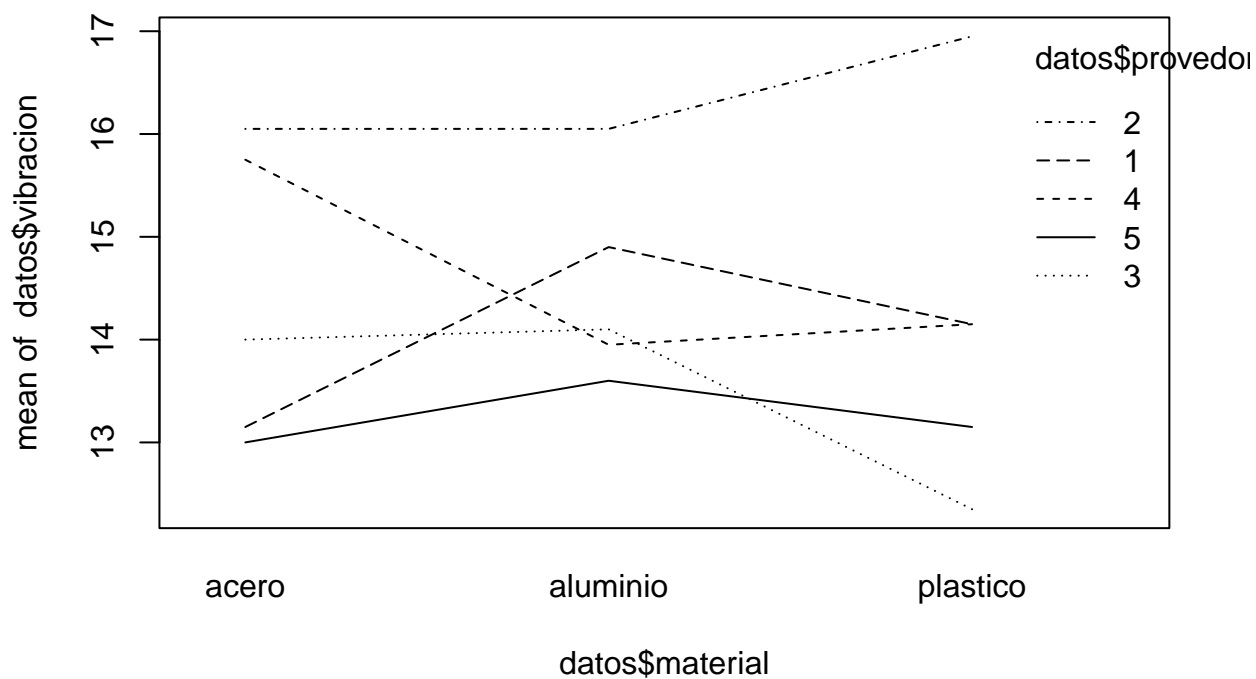
```
C <- aov(vibracion ~ material * proveedor, data = datos)
summary(C)
```

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

```
interaction.plot(datos$proveedor, datos$material, datos$vibracion)
```

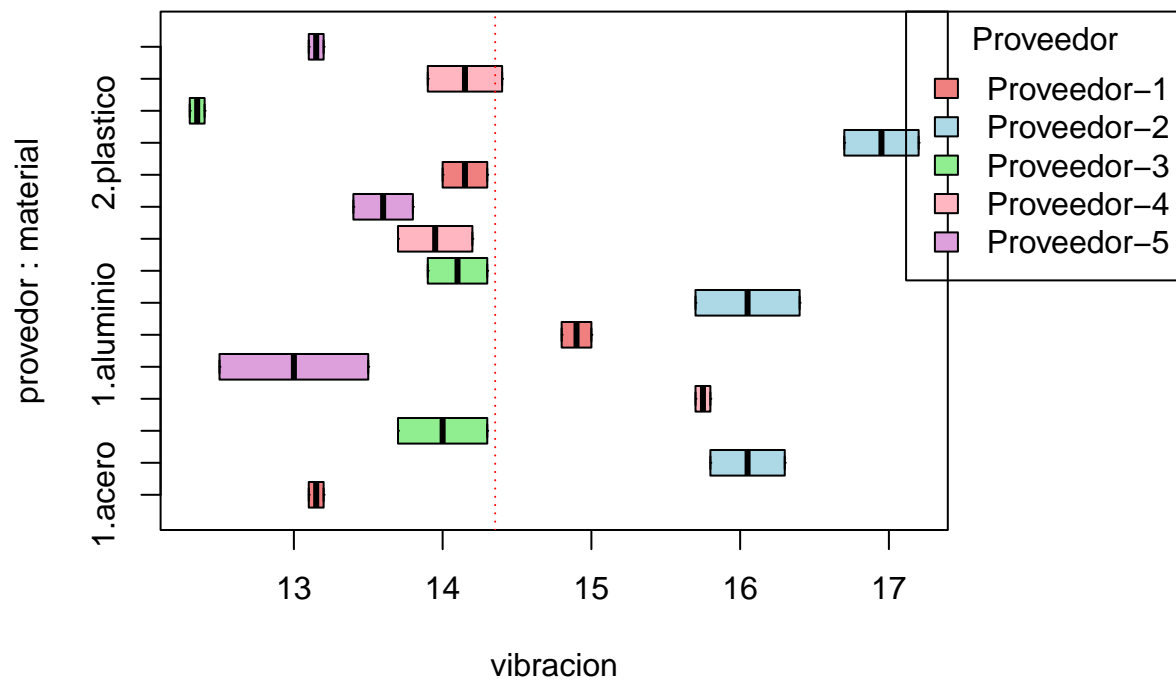


```
# Alternativa: Graficar con los ejes en diferente orden
interaction.plot(datos$material, datos$proveedor, datos$vibracion)
```



```
par(mar = c(5, 4, 4, 8))
```

```
datos = data.frame(vibracion, proveedor, material)
boxplot(vibracion ~ proveedor : material, datos, col = c("lightcoral", "lightblue", "lightgreen", "lightpink", "plum"),
        abline(v = mean(vibracion), lty= 3, col = "red"))
legend("topright", legend = c("Proveedor-1", "Proveedor-2", "Proveedor-3", "Proveedor-4", "Proveedor-5"),
       fill = c("lightcoral", "lightblue", "lightgreen", "lightpink", "plum"),
       title = "Proveedor", inset = c(-0.3, 0), xpd = TRUE)
```



valor frontera

```
qf_material = qf(0.95, df1 = 2, df2 = 15)
qf_material
```

```
## [1] 3.68232
```

```
qf_proveedor = qf(0.95, df1 = 4, df2 = 15)
qf_proveedor
```

```
## [1] 3.055568
```

```
qf_interaction = qf(0.95, df1 = 8, df2 = 15)
qf_interaction
```

```
## [1] 2.640797
```

Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. El proveedor dos es el que tiene los tres elementos con la mayor cantidad de vibraciones, seguidos después por el proveedor cuatro, el proveedor tres, proveedor uno y proveedor cinco, por lo que podemos ver que el problema se encuentra más que nada con el proveedor dos y específicamente el material con mayor vibración es el plástico.

Escribe tus conclusiones parciales Los materiales de ciertos proveedores demuestran una mayor cantidad de vibraciones al compararlos los unos con los otros, por lo que e posible vislumbrar que el proveedor es el aspecto que más nos interesa investigar, ya que en realidad los materiales no tienen muchos cambios, ya que son dependientes del proveedor

Realiza el ANOVA para dos niveles sin interacción.

Consulta el código de R en los apoyos de clase de “ANOVA” Haz el boxplot de rendimiento por sexo. Calcula la media para el rendimiento por sexo y método. Haz los intervalos de confianza de rendimiento por sexo. Gráficalos Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Escribe tus conclusiones parciales

```
B<-aov(datos$vibracion~datos$material+datos$proveedor)
summary(B)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## datos$material  2   0.70   0.352    0.61    0.552
## datos$proveedor  4  36.67   9.169   15.88 2.28e-06 ***
## Residuals      23  13.28   0.577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

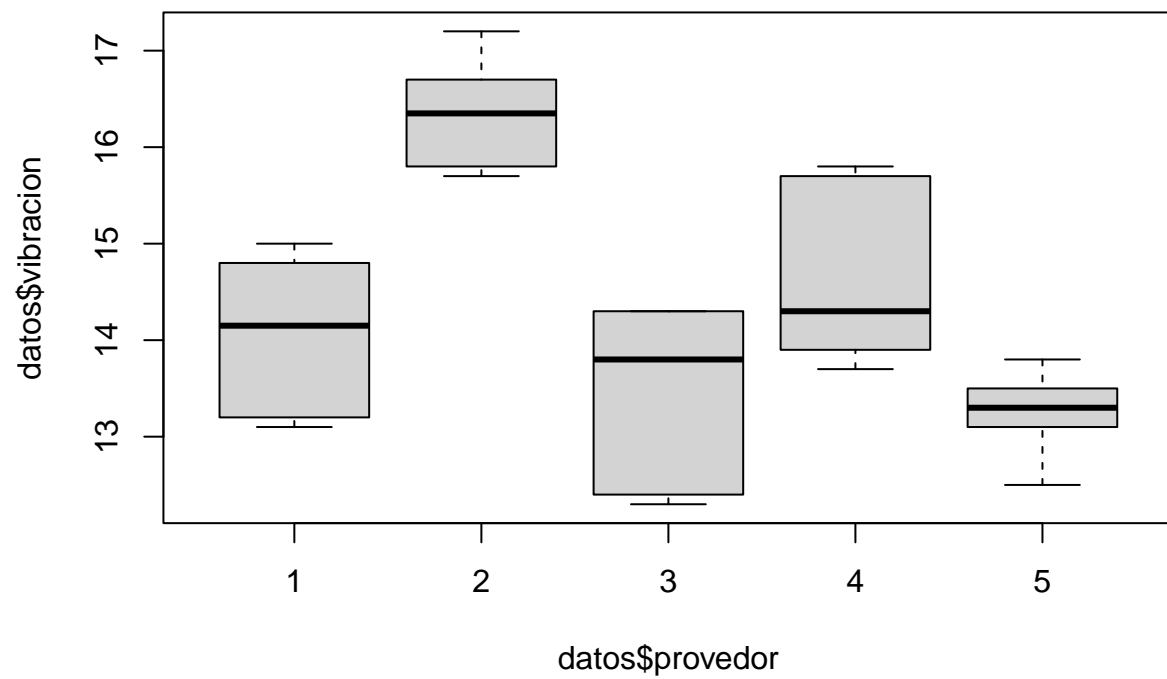
```
tapply(datos$vibracion,datos$proveedor,mean)
```

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

```
tapply(datos$vibracion,datos$proveedor,mean)
```

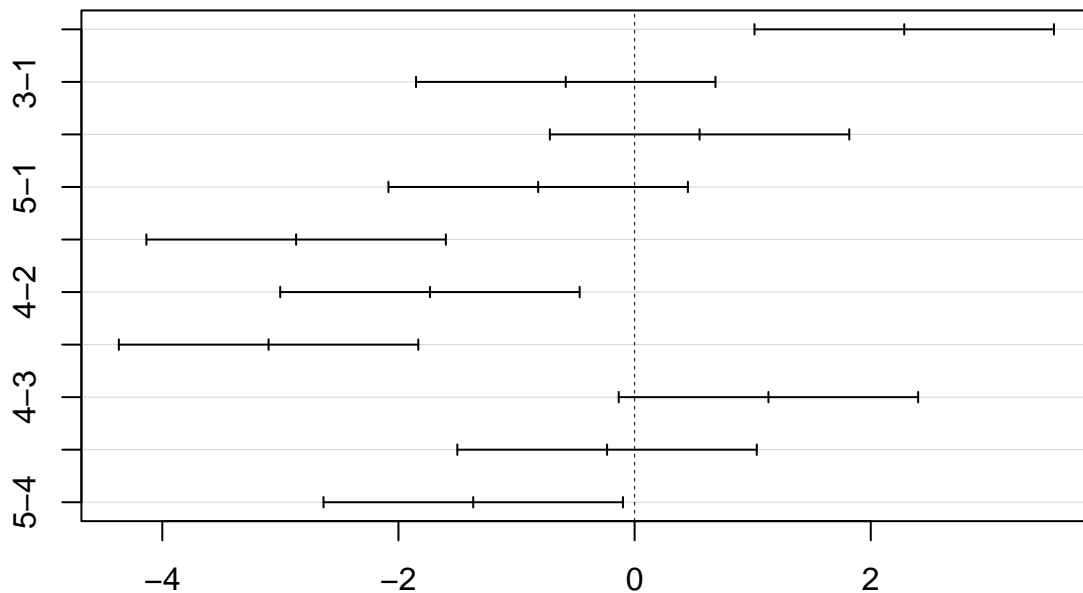
```
##          1          2          3          4          5
## 14.06667 16.35000 13.48333 14.61667 13.25000
```

```
M=mean(datos$vibracion)
boxplot(datos$vibracion ~ datos$proveedor)
```



```
Ix = TukeyHSD(aov(datos$vibracion ~ datos$proveedor))  
plot(Ix)
```

95% family-wise confidence level



Differences in mean levels of datos\$proveedor

Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. El material no tiene un efecto significativo en las vibraciones del mismo, por otra parte, el valor del proveedor es significativo, lo cual nos indica que este tiene un impacto muy significativo en las vibraciones de los materiales, al mismo tiempo se muestra que se tiene diferencias grandes entre cada uno de los distribuidores.

Escribe tus conclusiones parciales Se confirma que los proveedores son quienes causan las vibraciones entre los materiales y no los materiales en sí, al analizar las gráficas, es posible confirmar que el proveedor número dos es el que tiene la mayor cantidad de vibraciones en comparación de los demás proveedores.

Realiza el ANOVA para un efecto principal

Consulta el código de R en los apoyos de clase de “ANOVA” Haz el boxplot de rendimiento por método de enseñanza. Calcula la media. Haz los intervalos de confianza de rendimiento por método. Grafícalos Realiza la prueba de comparaciones múltiples de Tukey. Grafica los intervalos de confianza de Tukey. Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Escribe tus conclusiones parciales

```
Bi<-aov(datos$vibracion~datos$material)
summary(Bi)
```

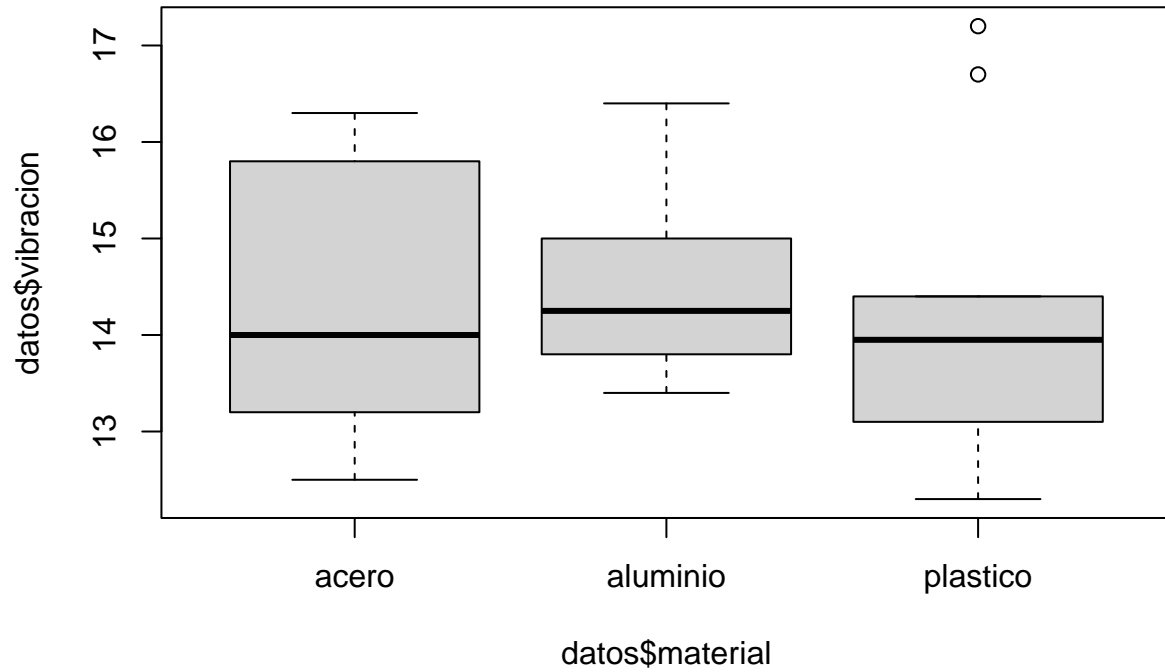
```
##              Df Sum Sq Mean Sq F value Pr(>F)
## datos$material  2   0.70   0.3523    0.19  0.828
## Residuals      27  49.95   1.8500
```

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

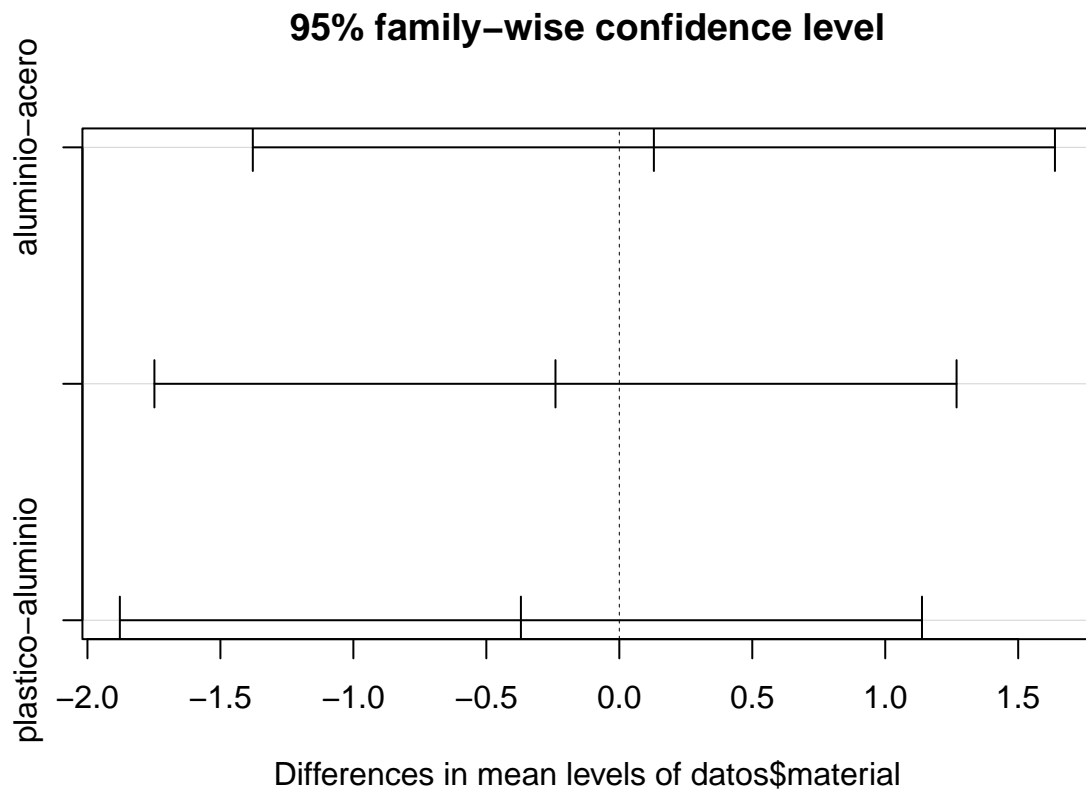


```
##      acero aluminio plastico
##      14.39    14.52    14.15
```

```
M=mean(datos$vibracion)
boxplot(datos$vibracion ~ datos$material)
```



```
Ix = TukeyHSD(aov(datos$vibracion ~ datos$material))
plot(Ix)
```



Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. Los tres materiales tienen resultados muy cercanos en sus vibraciones, lo cual sugiere que no tienen diferencias los unos con los otros.

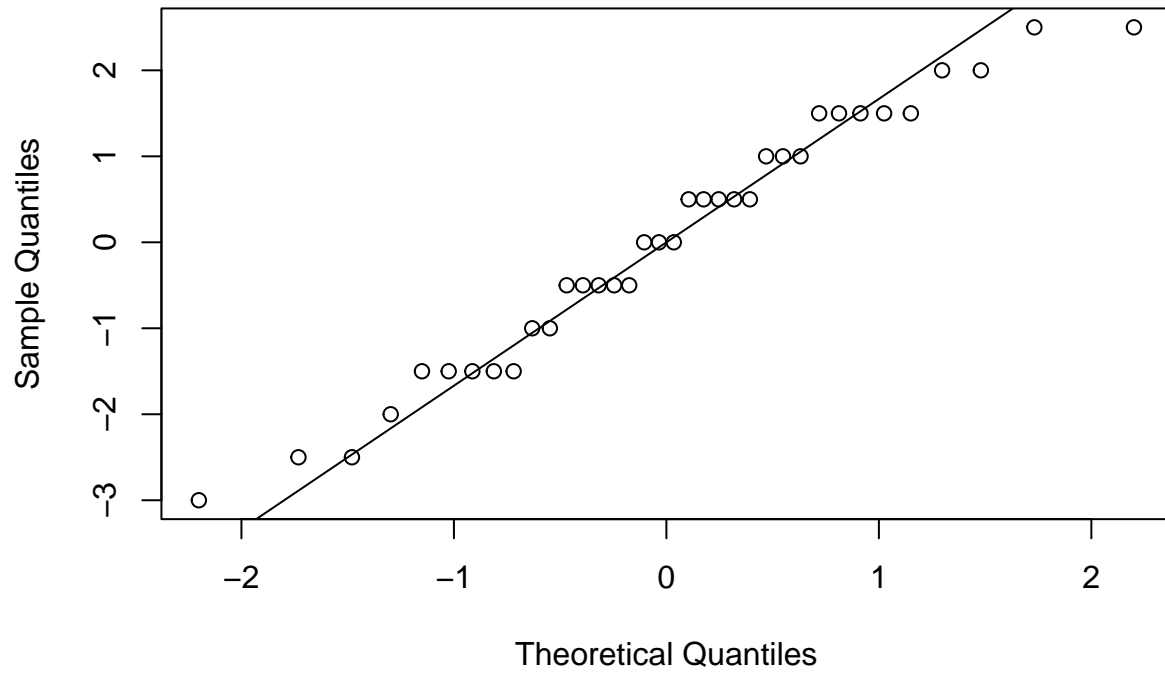
Escribe tus conclusiones parciales Podemos confirmar que el material no es relevante en la diferencia de vibraciones en cada uno de ellos, por lo que podemos dejar en claro que son los proveedores quienes tienen el efecto importante en este aspecto.

Comprueba la validez del modelo. Comprueba:

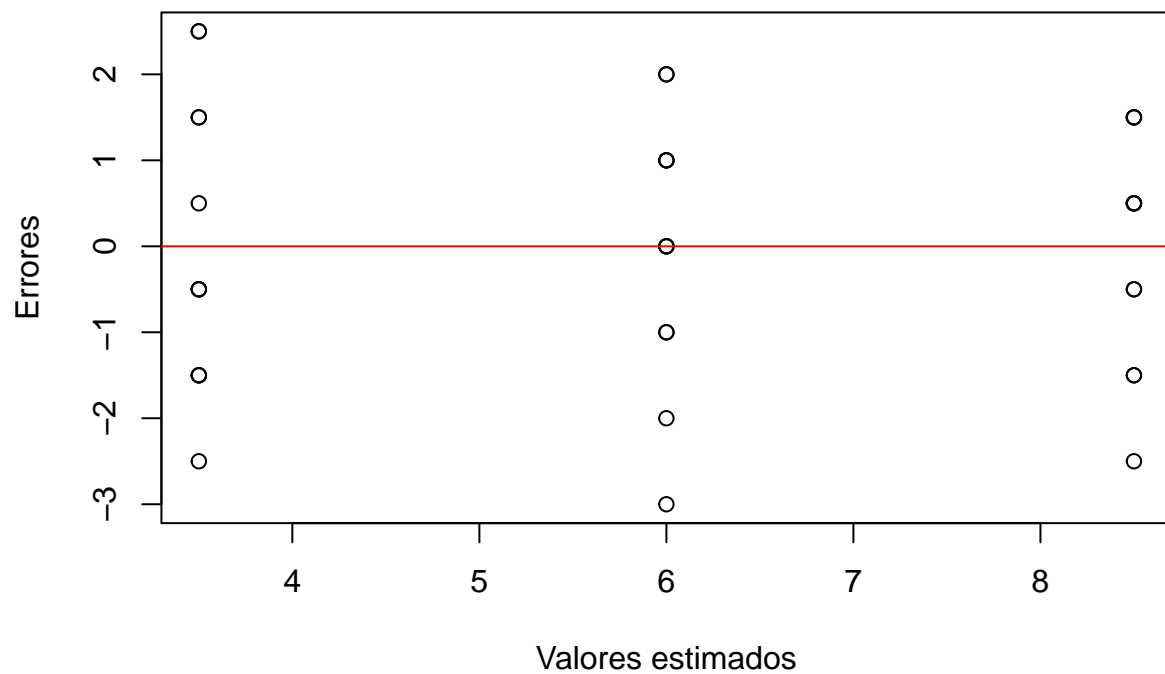
Normalidad Homocedasticidad Independencia Relación lineal entre las variables (coeficiente de determinación).

```
# Normalidad
residuos=anova$residuals
qqnorm(residuos)
qqline(residuos)
```

Normal Q-Q Plot



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



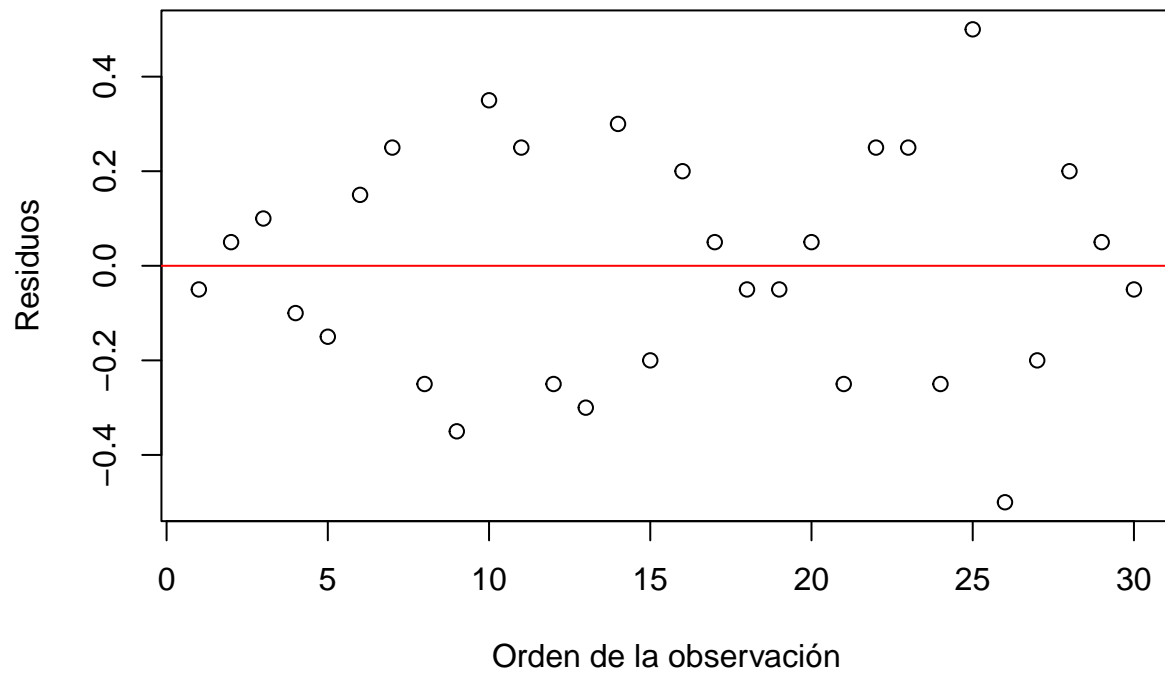
```
# Independencia
C <- aov(datos$vibracion ~datos$material *datos$proveedor , data = datos)

residuos <- residuals(C)

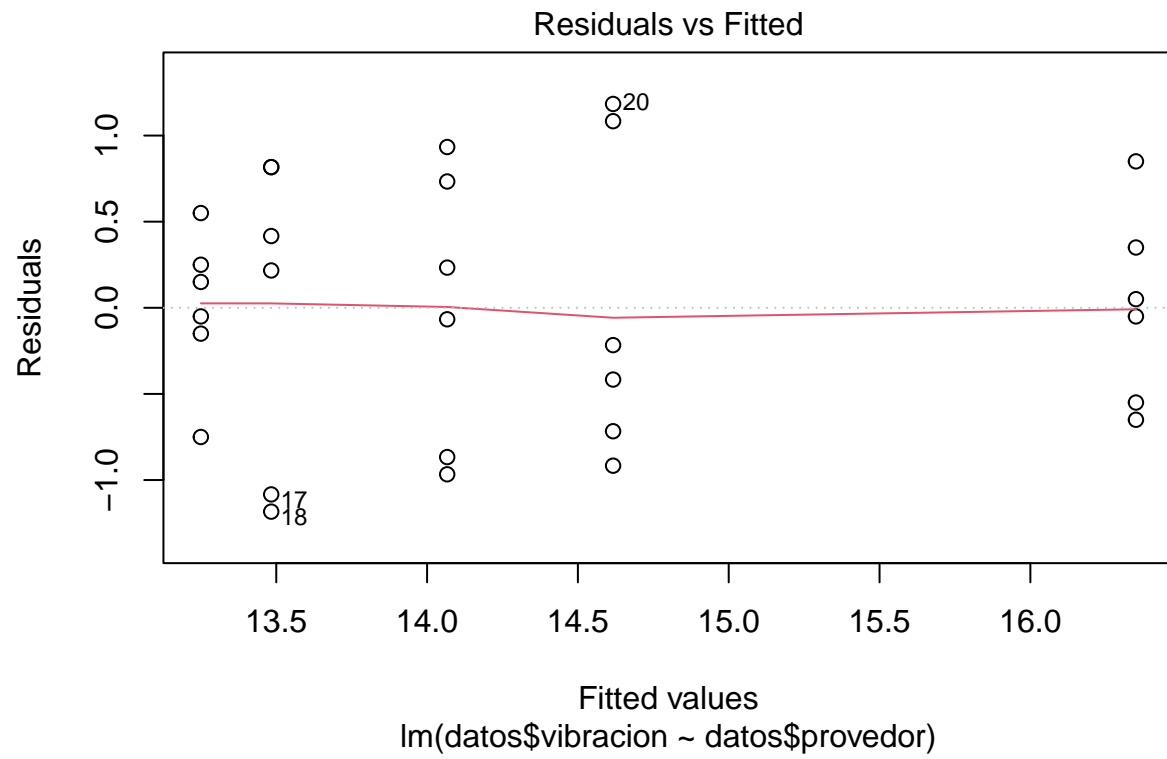
indices <- 1:length(residuos)

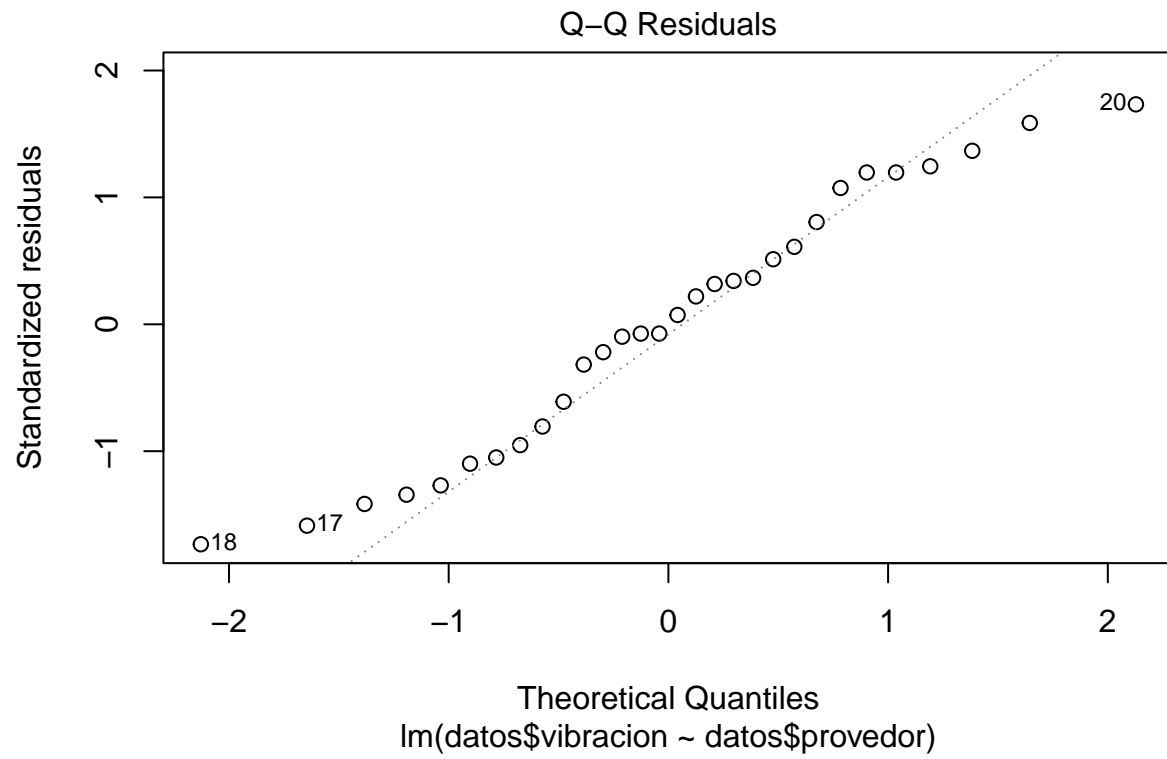
plot(indices, residuos, xlab = "Orden de la observación", ylab = "Residuos", main = "Residuos del Modelo")
abline(h = 0, col = "red")
```

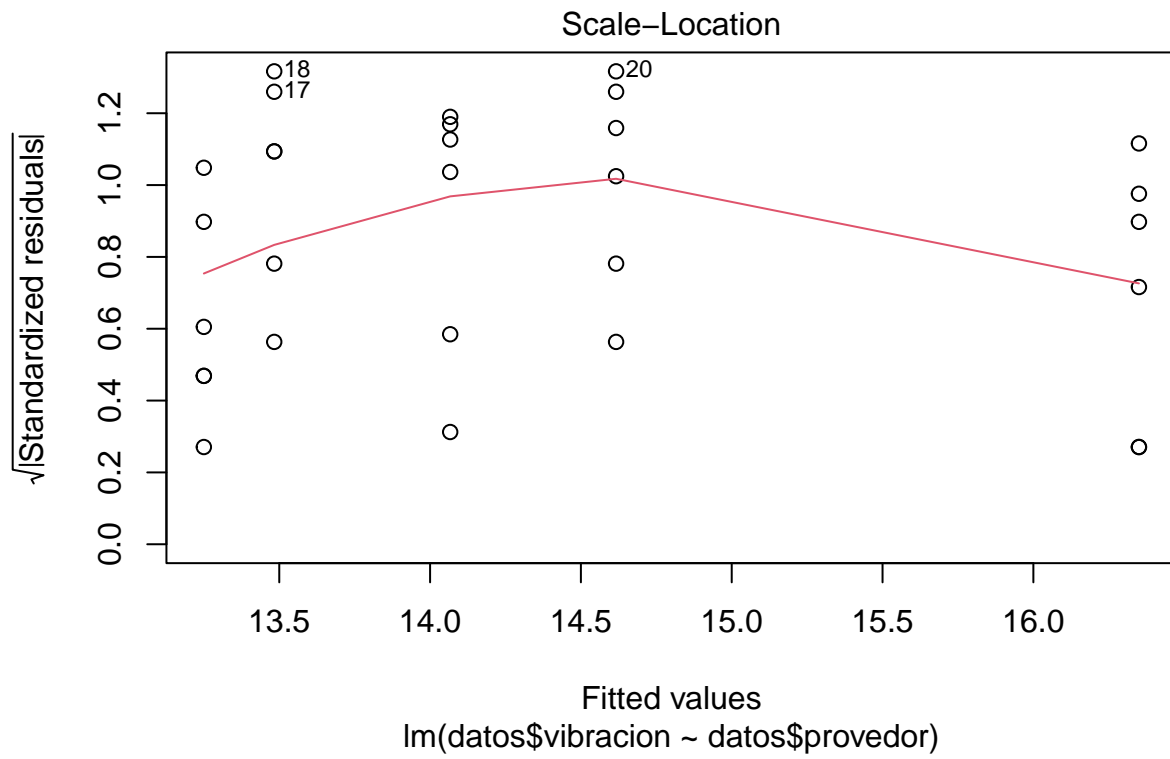
Residuos del Modelo ANOVA

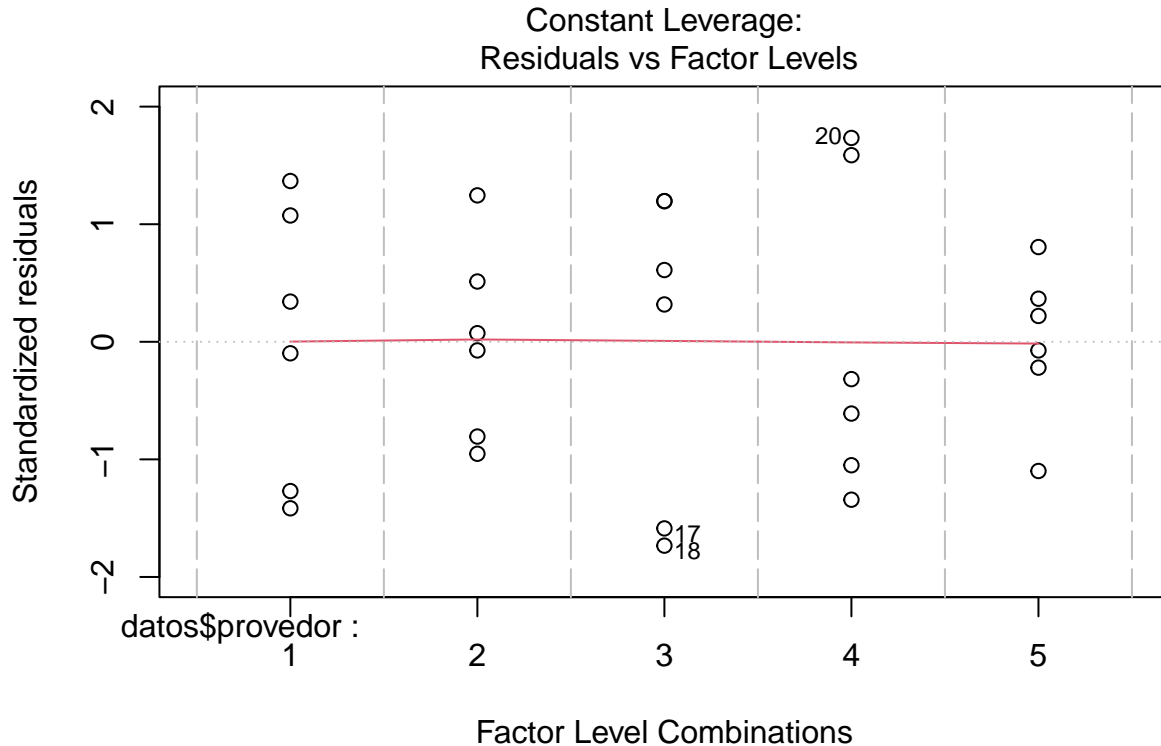


```
plot(lm(datos$vibracion~datos$proveedor))
```









```
CD= 0.70/(0.70+49.95) #coeficiente de
summary(lm(datos$vibracion~datos$proveedor))$r.squared
```

```
## [1] 0.7240136
```

Concluye en el contexto del problema.

Se puede concluir que de los cinco proveedores, el proveedor número dos es el causante de las vibraciones, al compararlo con los otros proveedores, es obvio la diferencia causada por este. El método explica el 0.7240 de la varianza, lo cual se traduce a 72.40% de la variación total, lo cual refuerza que los proveedores son quienes tienen un papel muy importante en los resultados de las vibraciones, mientras que el restante se le atribuye tanto a la aleatoriedad como a los mismos materiales. Al analizar los gráficos q-q y de residuos por valor ajustado, estos datos cumplen con normalidad, al mismo tiempo que los errores cumplen con una media cero.