

## 9. ANOVA

Eliezer Cavazos

2024-08-27

### Problema 1

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

```
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)

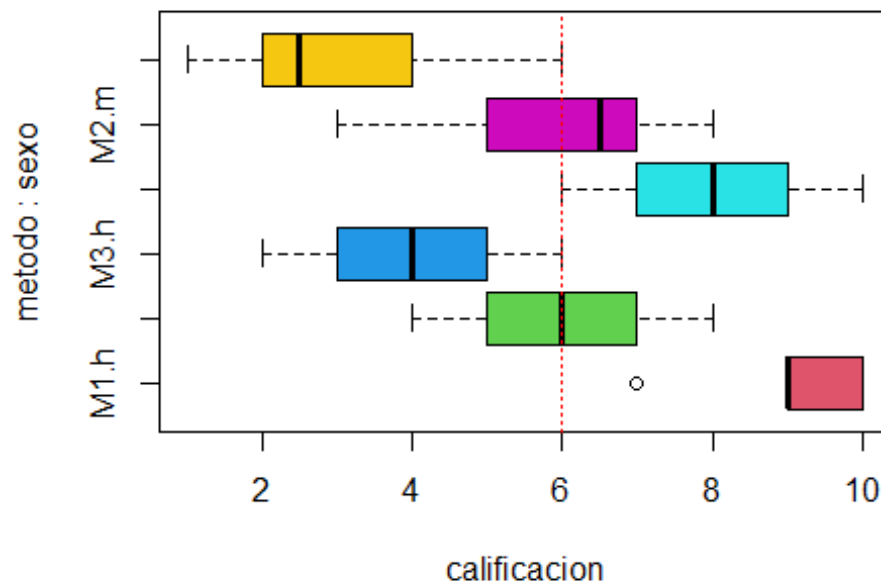
data = data.frame(calificacion, metodo, sexo)
```

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

```
boxplot(calificacion ~ metodo:sexo,data, col = 2:8, horizontal=TRUE)
abline(v = mean(calificacion), lty = 3, col = "red")
```



#### Interpreta el

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

El resultado muestra que el Metodo 1 tiende estar arriba de la media con hombres y mujeres, El Metodo 2 queda en la media y el Metodo 3 tiende a estar abajo de la media de calificacion.

*Escribe tus conclusiones parciales*

Viendo el Boxplot podemos ver que de los 3 metodos el Metodo 1 Incrementa la Calificacion, el Metodo 2 queda en la media y el Metodo 3 es el que más decrece, viendo esto en el contexto del problema lo que nosotros queremos saber es que Metodo es el que más le ayuda a los alumnos y podemos identificar el Metodo 1 es el metodo de los 3

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

F1:Metodo de enseñanza F2: Sexo

$$Y_{ijk} = \mu + T_i + a_j + T_{ia_j} + E_{ijk}$$

Primera Hipótesis:  $H_0: T_i = 0$   $H_1$ : algún  $T_i$  es distinto de 0

Segunda Hipótesis:  $H_0: a_j = 0$   $H_1$ : algún  $a_j$  es distinto de 0

Tercera Hipótesis Interaccion:  $H_0: T_i a_j = 0$   $H_1$ : algún  $T_i a_j$  es distinto de 0

Realiza el ANOVA para dos niveles con interacción:

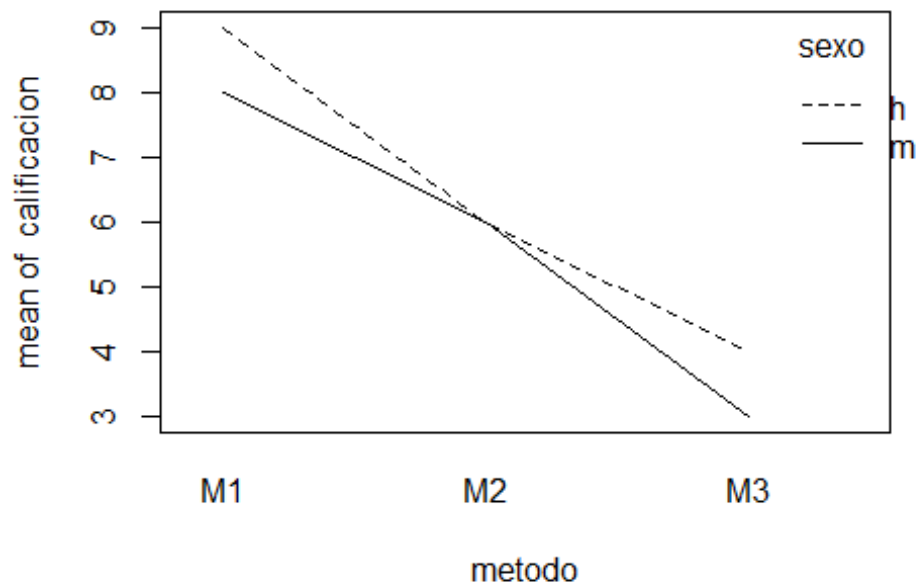
*Haz la gráfica de interacción de dos factores en ANOVA*

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

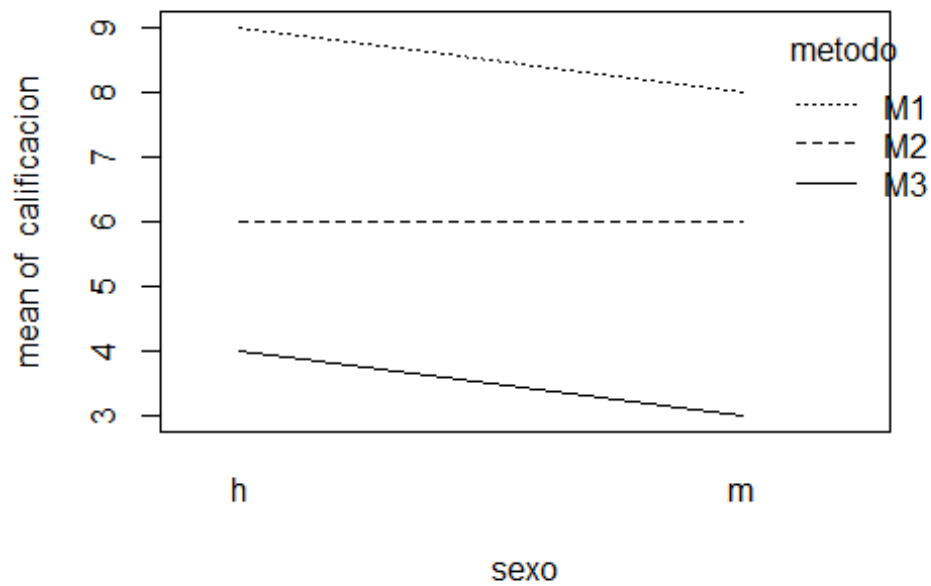
```
summary(A)
```

```
##              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)
```

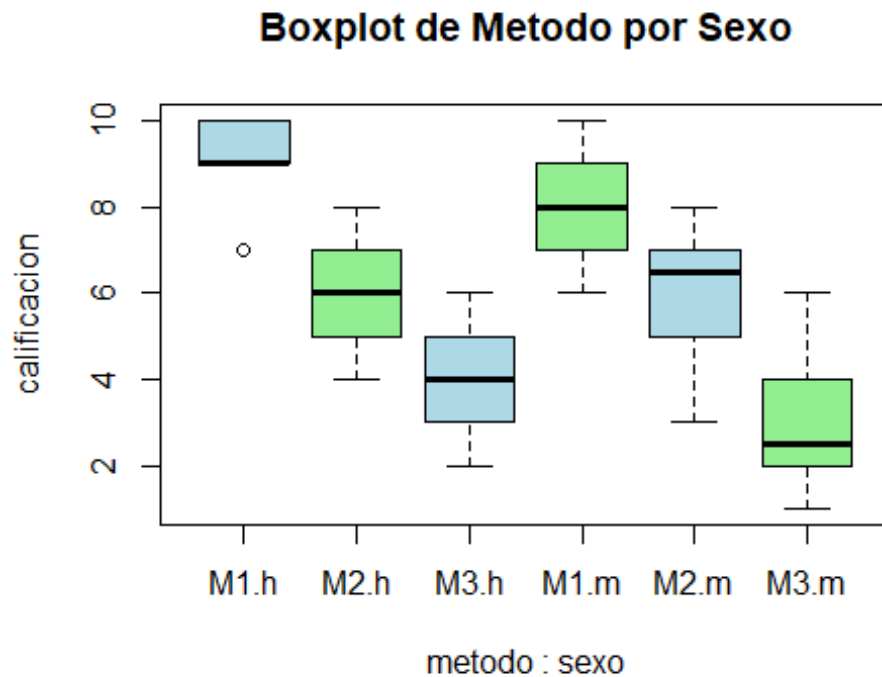


```
interaction.plot(sexo, metodo, calificacion)
```



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

```
boxplot(calificacion ~ metodo*sexo, data = data, col = c("lightblue",  
"lightgreen"), main = "Boxplot de Metodo por Sexo")
```



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

El Resultado muestra que la interaccion tiene poca significacion en los resultados de las calificaciones

*Escribe tus conclusiones parciales*

La interaccion resulto NO SIGNIFICATIVA, ya que el valor F es muy bajo eso significa que no es significativo para las calificaciones, por eso rechazamos tambien la Tercera Hipotesis, ya que no nos va a funcionar la Interaccion de las dos variables

**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.

```
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

print("Media de Calificacion por Sexo y Metodo")
```

```
## [1] "Media de Calificacion por Sexo y Metodo"

tapply(calificacion,sexo,mean)

##           h           m
## 6.333333 5.666667

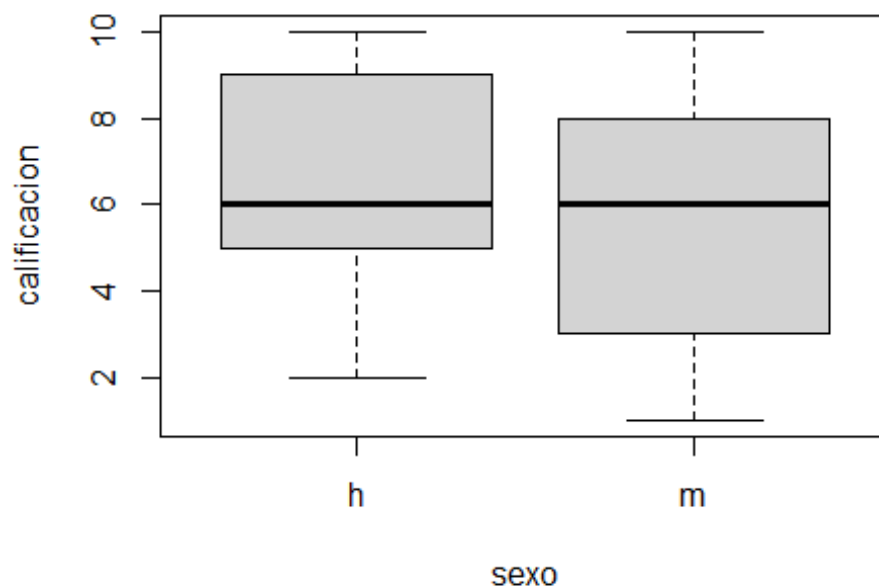
tapply(calificacion,metodo,mean)

##  M1  M2  M3
## 8.5 6.0 3.5

M=mean(calificacion)
M

## [1] 6

boxplot(calificacion ~ sexo)
```



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

```
C<-aov(calificacion~sexo)
summary(C)

##           Df Sum Sq Mean Sq F value Pr(>F)
## sexo         1      4   4.000    0.613   0.439
## Residuals    34     22   6.529

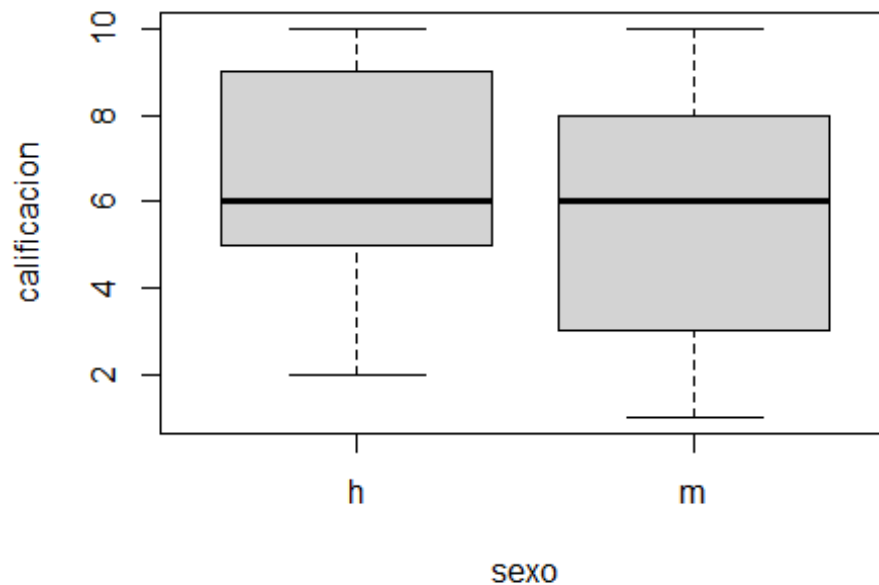
tapply(calificacion,sexo,mean)
```

```
##           h           m
## 6.333333 5.666667

mean(calificacion)

## [1] 6

boxplot(calificacion ~ sexo)
```

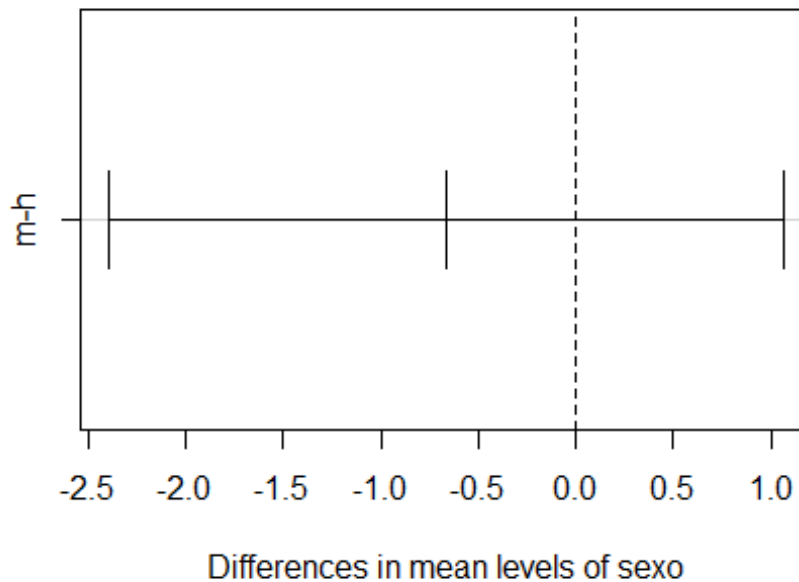


```
I = TukeyHSD(aov(calificacion ~ sexo))
I

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = calificacion ~ sexo)
##
## $sexo
##           diff           lwr           upr           p adj
## m-h -0.6666667 -2.397645 1.064312 0.4392235

plot(I) #Los intervalos de confianza se observan mejor si se grafican
```

### 95% family-wise confidence level



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

El Sexo tiene una baja significacncia en el resultado de las calificaciones

*Escribe tus conclusiones parciales*

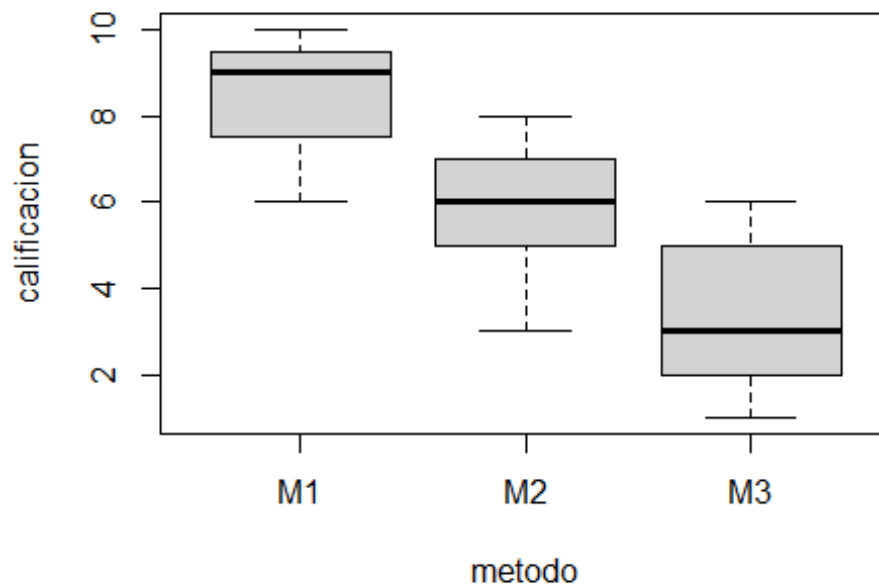
Podemos ver que el Sexo tampoco es tan significativo para las Calificaciones, la diferencia es muy pequeña pero igual se puede identificar que dependiendo del Metodo los dos sexos tendran una mejor calificacion.

**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.

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





```
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

print("Media de Calificacion por Metodo")

## [1] "Media de Calificacion por Metodo"

tapply(calificacion,metodo,mean)

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

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

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

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

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

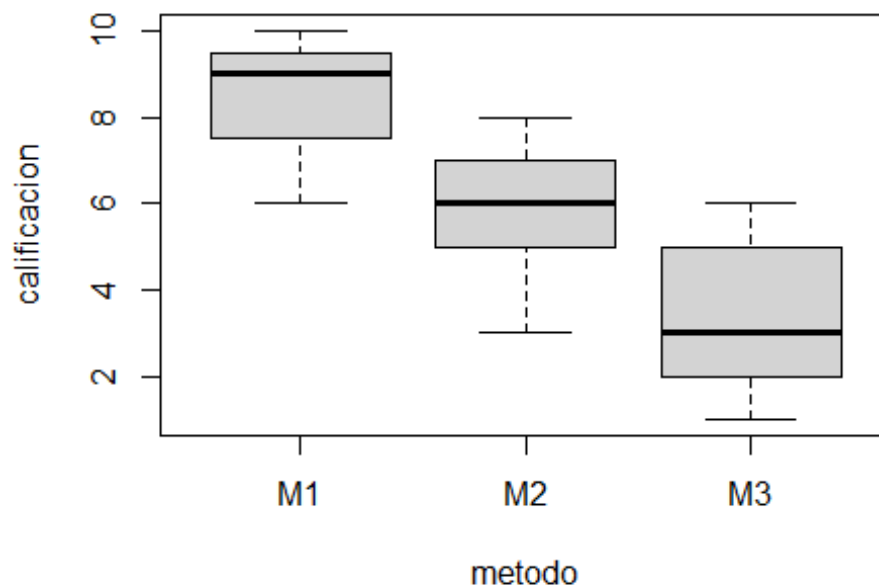
```
##  M1  M2  M3
```

```
## 8.5 6.0 3.5
```

```
mean(calificacion)
```

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

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



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

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

```
I
```

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

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

```
##
```

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

```
##
```

```
## $metodo
```

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

```
## M2-M1 -2.5 -4.020241 -0.9797592 0.0008674
```

```
## M3-M1 -5.0 -6.520241 -3.4797592 0.0000000
```

```
## M3-M2 -2.5 -4.020241 -0.9797592 0.0008674
```

```
plot(I) #Los intervalos de confianza se observan mejor si se grafican
```



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

El efecto principal que afecta la media de las Calificaciones es el Metodo de aprendizaje

*Escribe tus conclusiones parciales*

Podemos identificar que el metodo de aprendizaje SI es significativo para las Calificaciones, donde sin importar el sexo termina teniendo un mejor resultado

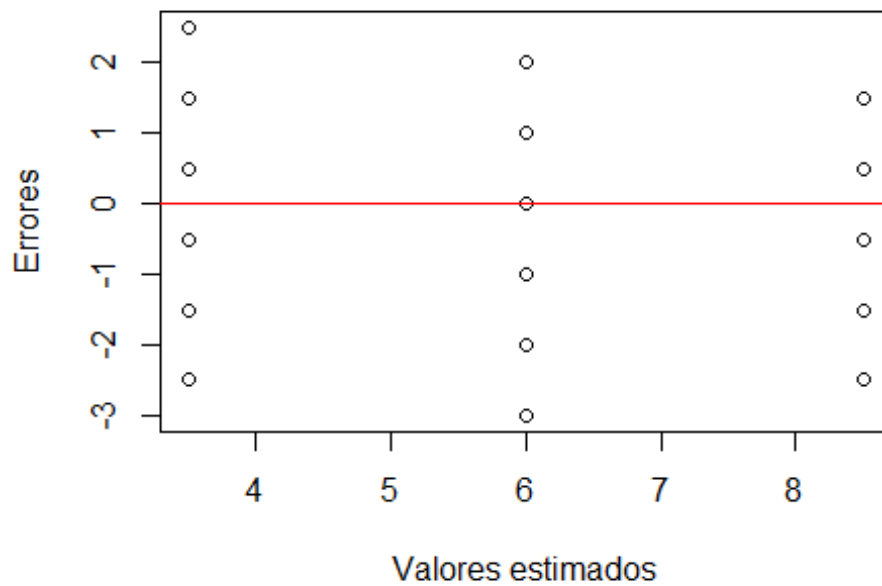
*Comprueba la validez del modelo. Comprueba:*

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

```
anova = aov(calificacion~metodo)
```

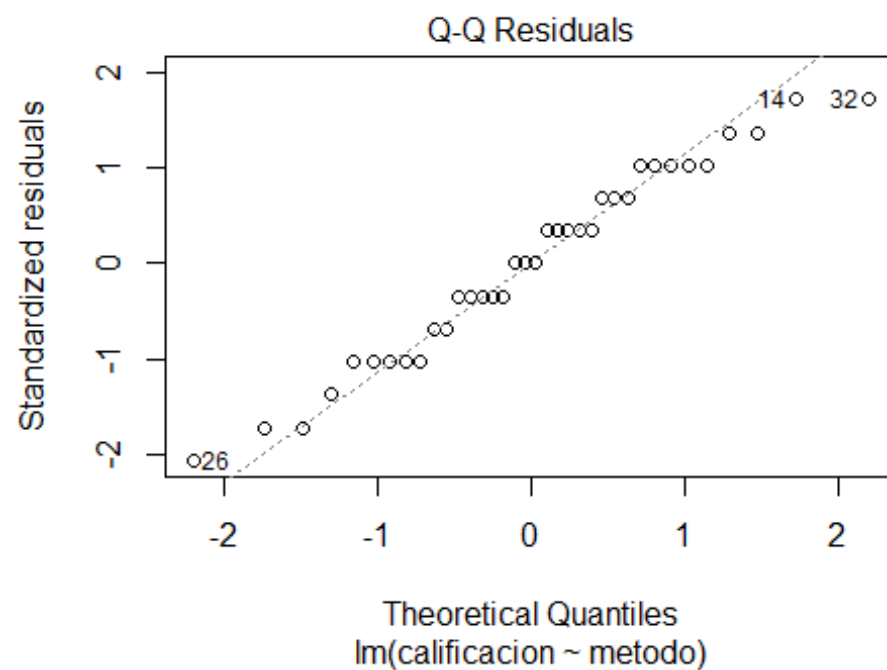
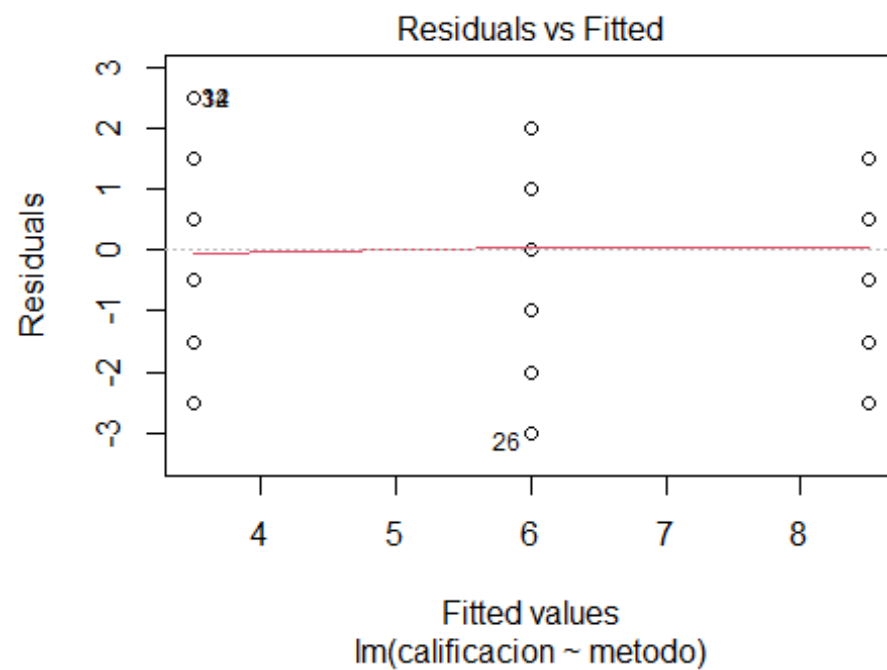
```
residuos= anova$residuals
#qqnorm(residuos)
#qqline(residuos)
```

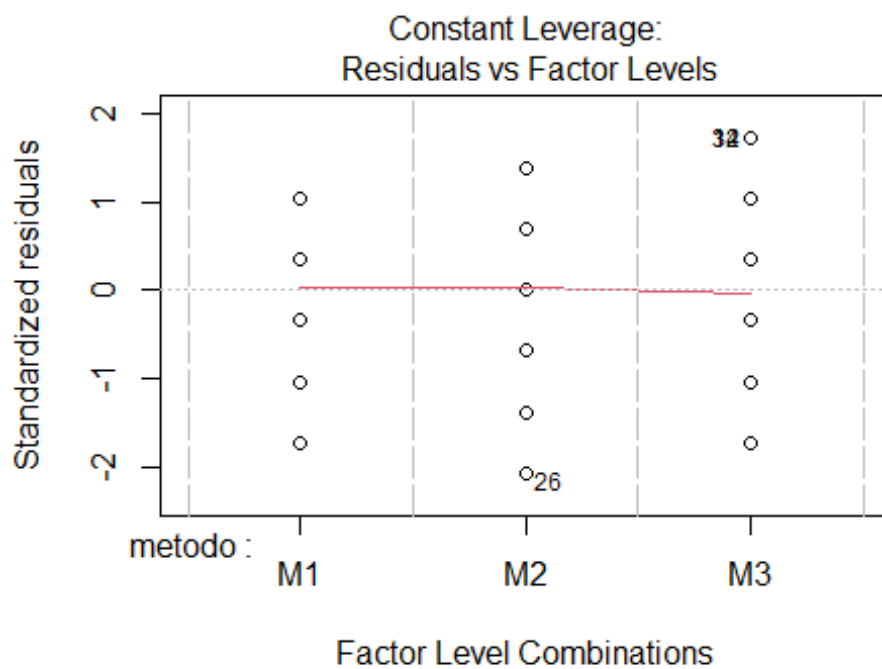
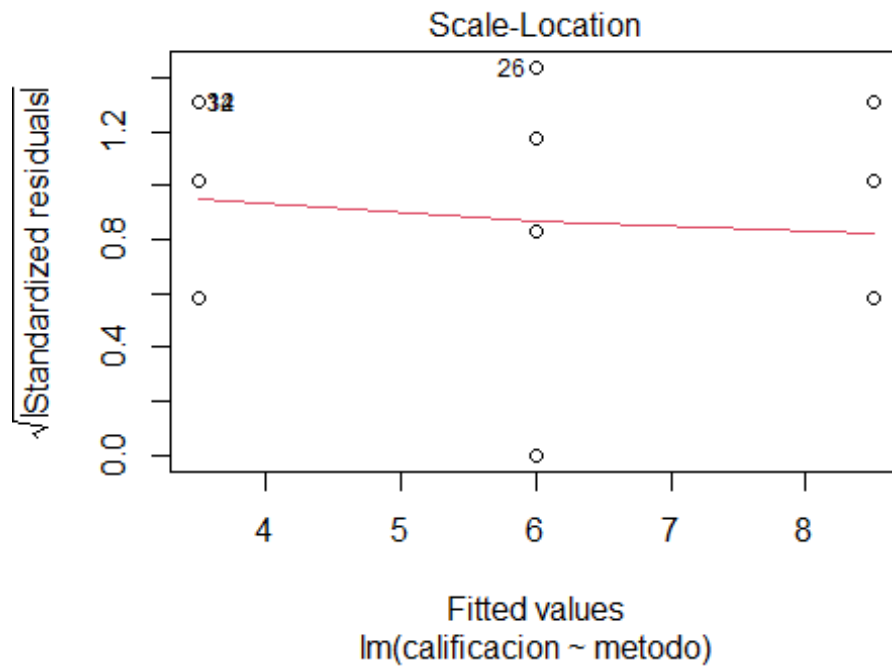
```
plot(anova$fitted.values,anova$residuals,ylab="Errores",xlab="Valores
estimados")
abline(h=0,col="red")
```



```
#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 determinación para el modelo.

```
summary(lm(calificacion~metodo))$r.squared
```

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

### Concluye en el contexto del problema.

Con estos resultados podemos identificar que si existe una influencia de la metodología que afecta el promedio de las calificaciones de los alumnos, el género también tiene un pequeño cambio en las calificaciones pero es más significativo la metodología, además la mejora Metodología si se quiere que los alumnos mejoren sus calificaciones es el Metodo 1

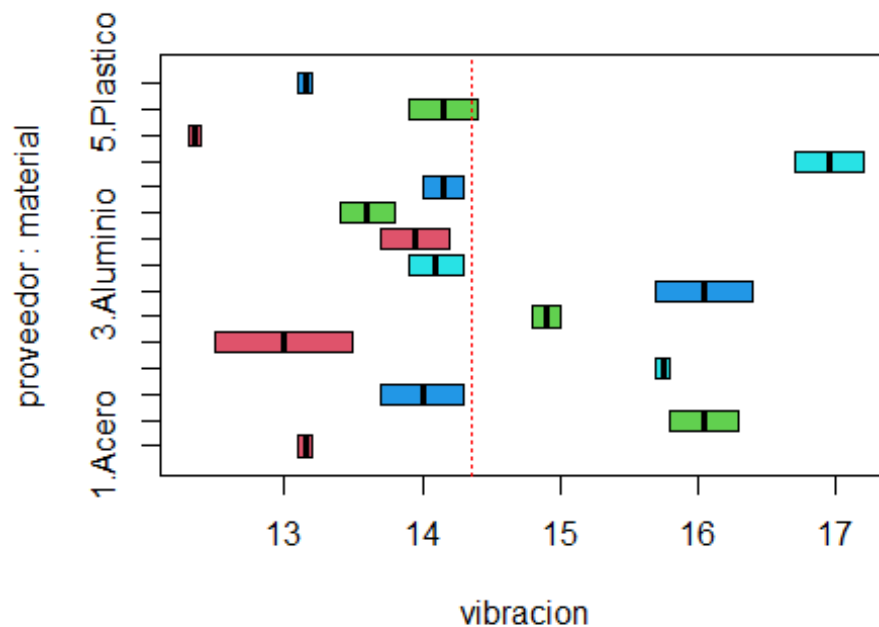
### Problema 2 Vibracion de Motores

```
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)
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))
#material = c(rep("h", 18), rep("m",18))
material = factor(material)
proveedor = factor(proveedor)

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

*Haz el boxplot de la evaluación de los motores por Proveedor y Material.*

```
boxplot(vibracion ~ proveedor:material, data, col = 2:5, horizontal=TRUE)
abline(v = mean(vibracion), lty = 3, col = "red")
```



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

Con el boxplot podemos identificar que la mayoría de datos se encuentran abajo de la media

*Escribe tus conclusiones parciales*

Por el momento no se puede decir mucho más que el proveedor puede que tenga cierto significado sobre las vibraciones pero se ocupa un poco más de información para sacar conclusiones

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

F1: Material F2: Proveedor

$Y_{ijk} = \mu + T_i + a_j + T_{ia_j} + E_{ijk}$

Primera Hipótesis:  $H_0: T_i = 0$   $H_1$ : algún  $T_i$  es distinto de 0

Segunda Hipótesis:  $H_0: a_j = 0$   $H_1$ : algún  $a_j$  es distinto de 0

Tercera Hipótesis Interacción:  $H_0: T_i a_j = 0$   $H_1$ : algún  $T_i a_j$  es distinto de 0

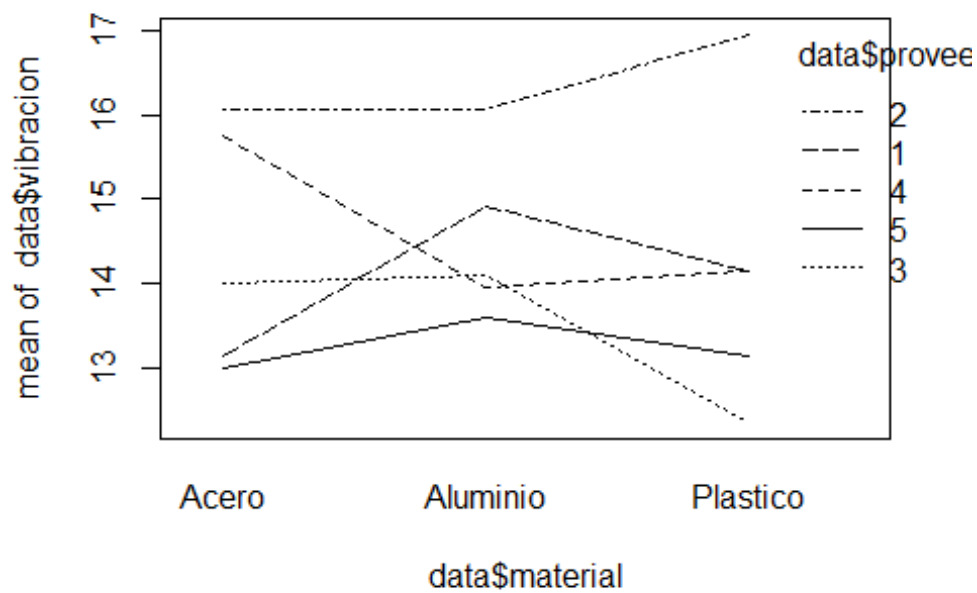
*Haz la gráfica de interacción de dos factores en ANOVA*

```
A<-aov(vibracion~material*proveedor,data)
summary(A)
```

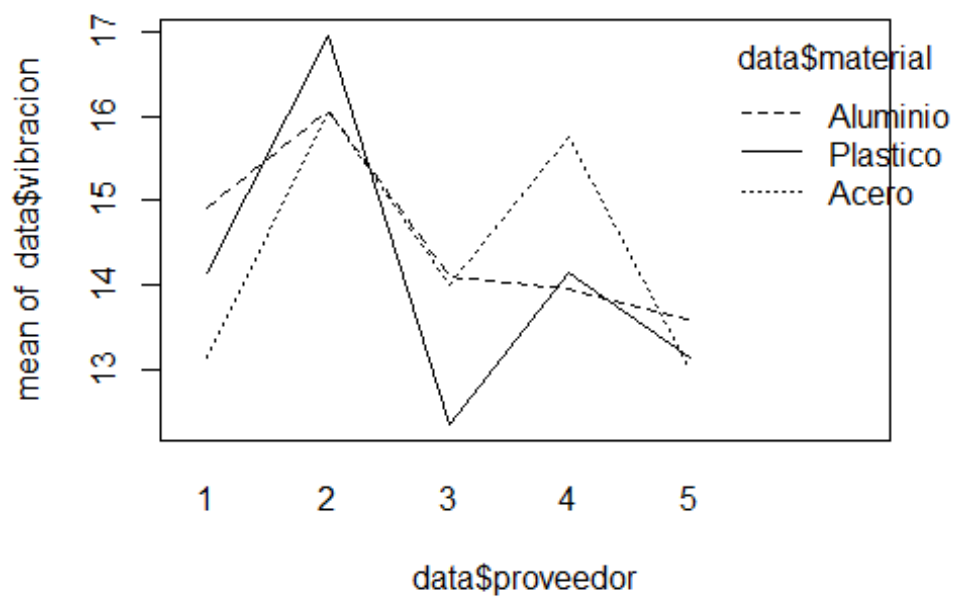
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## material          2    0.70    0.352     3.165    0.0713 .
## proveedor         4   36.67    9.169   82.353 5.07e-10 ***
## material:proveedor  8   11.61    1.451   13.030 1.76e-05 ***
## Residuals        15    1.67    0.111
## ---
## Signif.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

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





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



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

Despues de sacar nuestras hipotesis y sacar la interaccion de las dos variables podemos ver que no es muy significativo los dos valores juntos como usar solo Proveedor

*Escribe tus conclusiones parciales*

Por el momento podemos identificar con el valor de F que el Proveedor individual de la variable de material es mas significativo que la interaccion de ambos.

**Realiza el ANOVA para dos niveles sin interacción.**

*Haz el boxplot de vibracion por proveedor Calcula la media para el vibracion por proveedor y material*

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

##              Df Sum Sq Mean Sq F value    Pr(>F)
## data$material   2   0.70   0.352    0.61   0.552
## data$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

print("Media de Vibracion por Material y Proveedor")

## [1] "Media de Vibracion por Material y Proveedor"

tapply(data$vibracion,data$proveedor,mean)

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

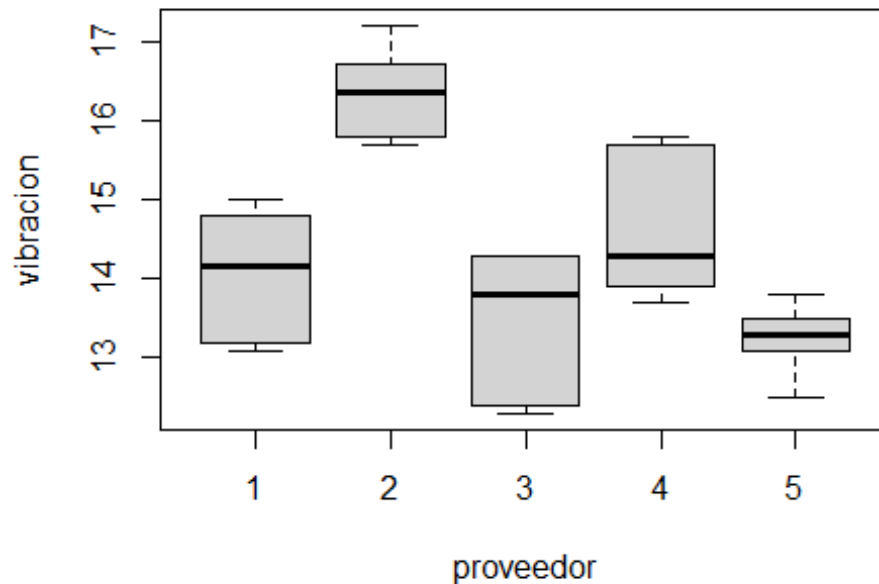
tapply(data$vibracion,data$material,mean)

##   Acero Aluminio Plastico
##   14.39   14.52   14.15

M=mean(data$vibracion)
M

## [1] 14.35333

boxplot(vibracion ~ proveedor)
```



#### Haz los

intervalos de confianza de rendimiento por Proveedor. Grafícalos

```
C<-aov(data$vibracion~data$proveedor)
summary(C)

##              Df Sum Sq Mean Sq F value    Pr(>F)    
## data$proveedor  4  36.67   9.169    16.4 1.03e-06 ***
## Residuals      25  13.98   0.559                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

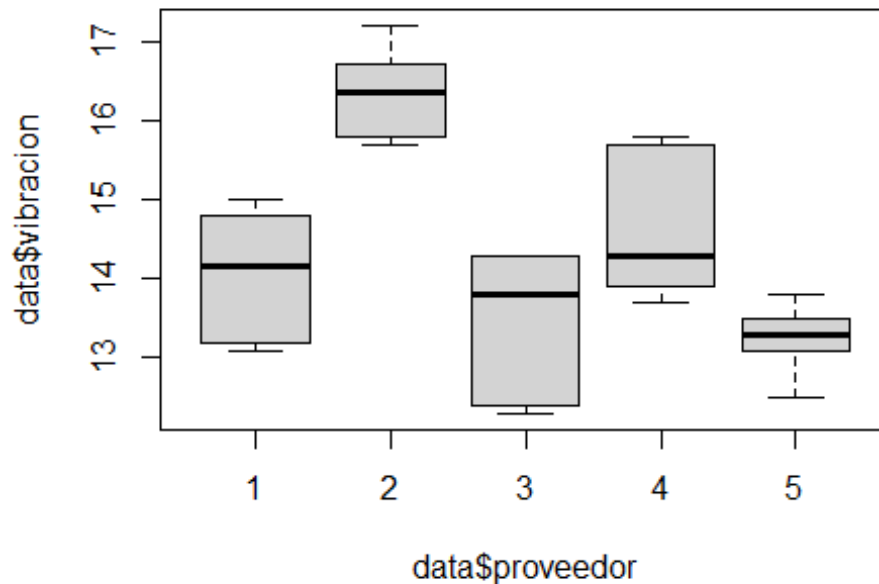
tapply(data$vibracion,data$proveedor,mean)

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

mean(data$vibracion)

## [1] 14.35333

boxplot(data$vibracion ~ data$proveedor)
```

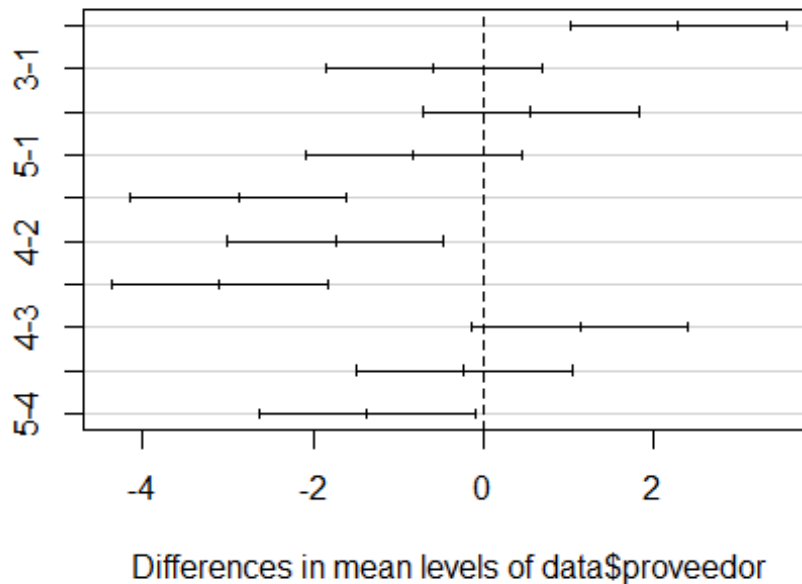


```
I = TukeyHSD(aov(data$vibracion ~ data$proveedor))
I

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = data$vibracion ~ data$proveedor)
##
## $`data$proveedor`
##      diff      lwr      upr    p adj
## 2-1  2.283333  1.015366  3.551300 0.0001595
## 3-1 -0.583333 -1.851300  0.684633 0.6630108
## 4-1  0.550000 -0.717967  1.817967 0.7089904
## 5-1 -0.816667 -2.084633  0.451300 0.3474956
## 3-2 -2.866667 -4.134633 -1.598699 0.0000055
## 4-2 -1.733333 -3.001300 -0.465366 0.0039774
## 5-2 -3.100000 -4.367967 -1.832033 0.0000015
## 4-3  1.133333 -0.134633  2.401300 0.0959316
## 5-3 -0.233333 -1.501300  1.034633 0.9821261
## 5-4 -1.366667 -2.634633 -0.098699 0.0301318
```

`plot(I)` *#Los intervalos de confianza se observan mejor si se grafican*

### 95% family-wise confidence level



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

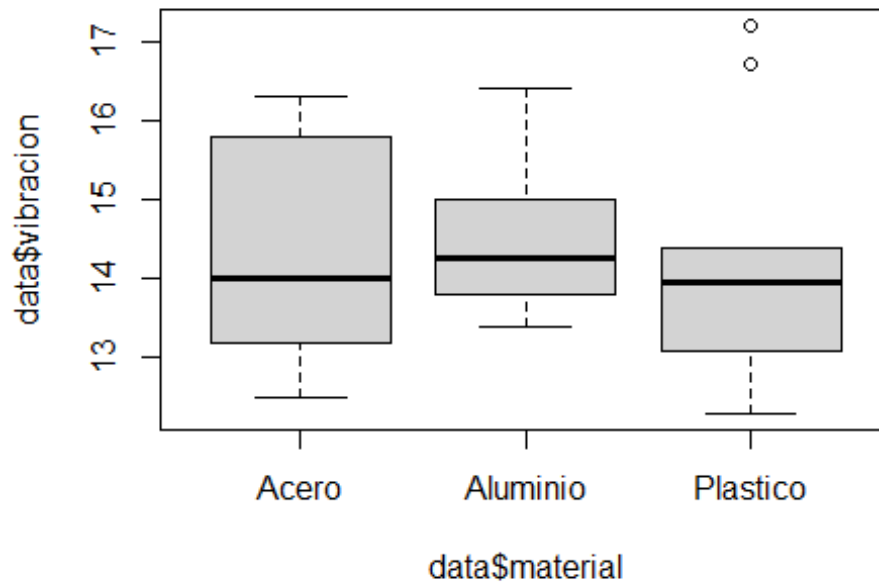
Podemos identificar que el valor de F tiene un valor muy significativo cuando se trata del proveedor, aun falta verificar los datos con Material pero por el momento se puede decir que el Proveedor es la mayor causa de las vibraciones.

*Escribe tus conclusiones parciales*

La mayor causa de las vibraciones es el proveedor hasta el momento ya que la cantidad de vibraciones cambia en base al proveedor principalmente.

*Haz el boxplot de vibracion por material. Calcula la media.*

```
boxplot(data$vibracion ~ data$material)
```



```
B<-aov(data$vibracion~data$material)
summary(B)

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

print("Media de Vibracion por Material")

## [1] "Media de Vibracion por Material"

tapply(data$vibracion,data$material,mean)

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

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

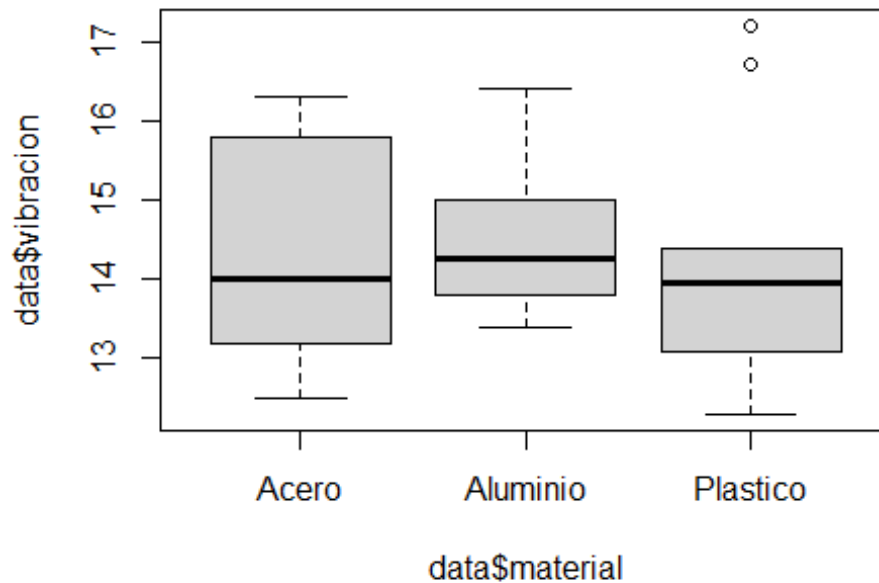
```
C<-aov(data$vibracion~data$material)
summary(C)

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

tapply(data$vibracion,data$material,mean)

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

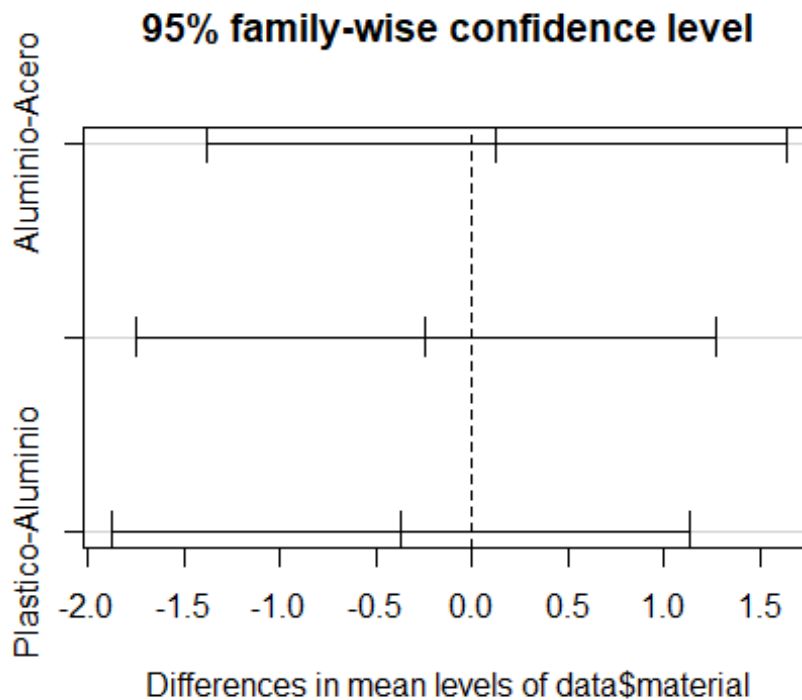
```
mean(data$vibracion)
## [1] 14.35333
boxplot(data$vibracion ~ data$material)
```



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

```
I = TukeyHSD(aov(data$vibracion ~ data$material))
I
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = data$vibracion ~ data$material)
##
## $`data$material`
##              diff          lwr          upr        p adj
## Aluminio-Acero   0.13 -1.378171  1.638171  0.9751575
## Plastico-Acero  -0.24 -1.748171  1.268171  0.9180284
## Plastico-Aluminio -0.37 -1.878171  1.138171  0.8168495

plot(I) #Los intervalos de confianza se observan mejor si se grafican
```



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

Después de estudiar más el uso de los materiales con sus vibraciones podemos identificar que no es muy significativo para ser la causa de las vibraciones y podemos confirmar que la mayor causa de las vibraciones de los motores es el proveedor

*Escribe tus conclusiones parciales*

Los materiales no son muy significativos para ser la causa de las vibraciones de los motores eléctricos.

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

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

```
anova = aov(data$vibracion~data$proveedor)
```

```
residuos= anova$residuals
```

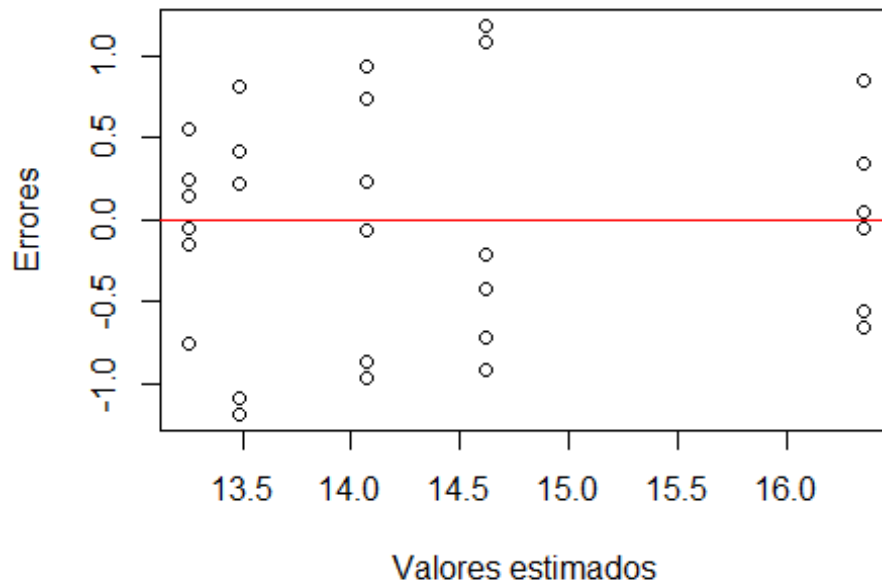
```
#qqnorm(residuos)
```

```
#qqline(residuos)
```

```
plot(anova$fitted.values, anova$residuals, ylab="Errores", xlab="Valores")
```

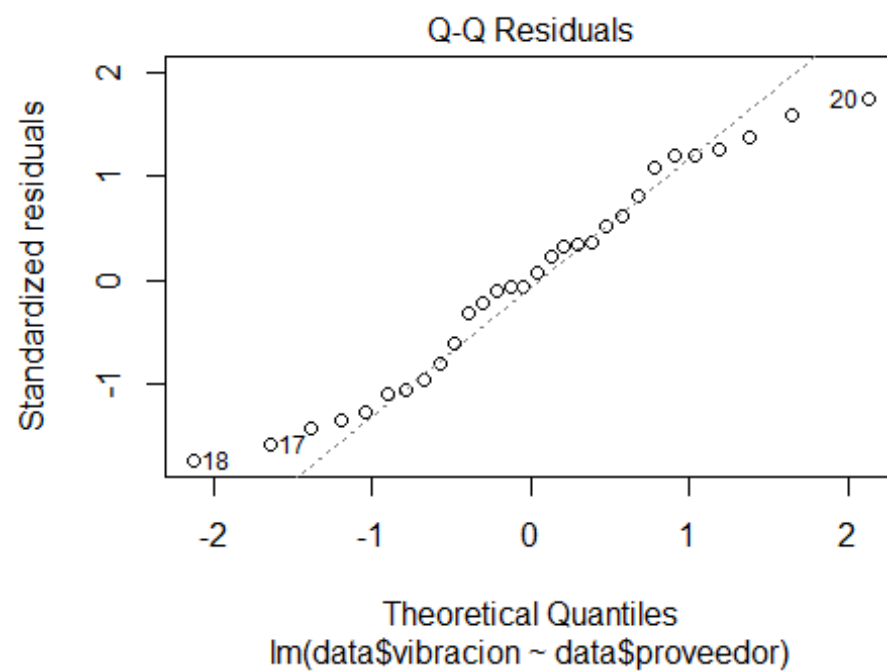
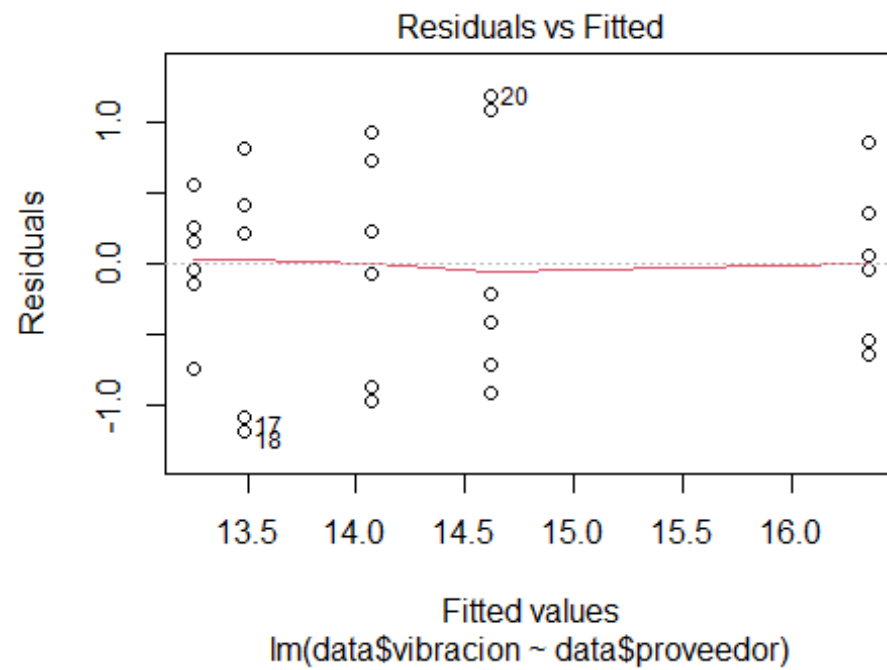


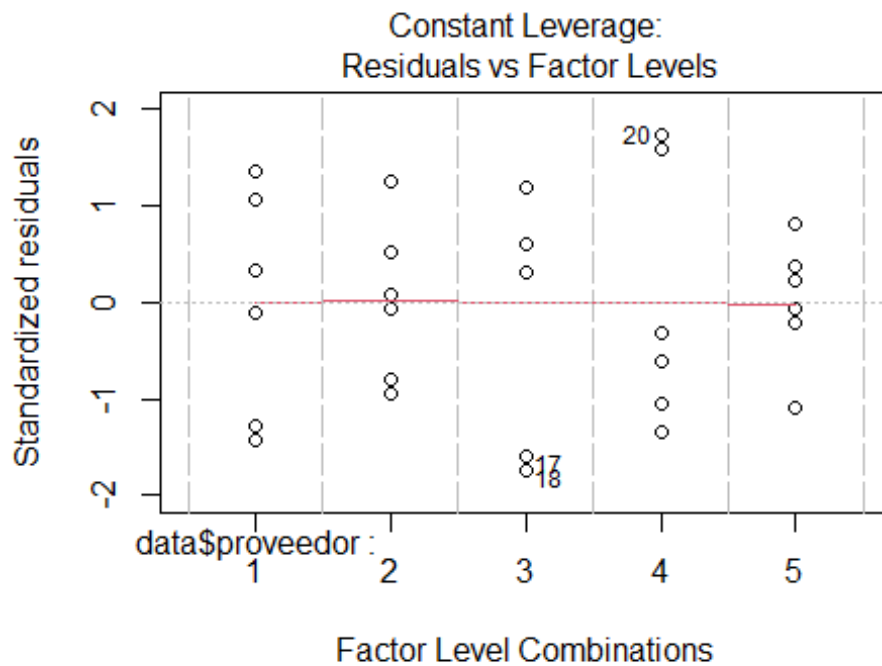
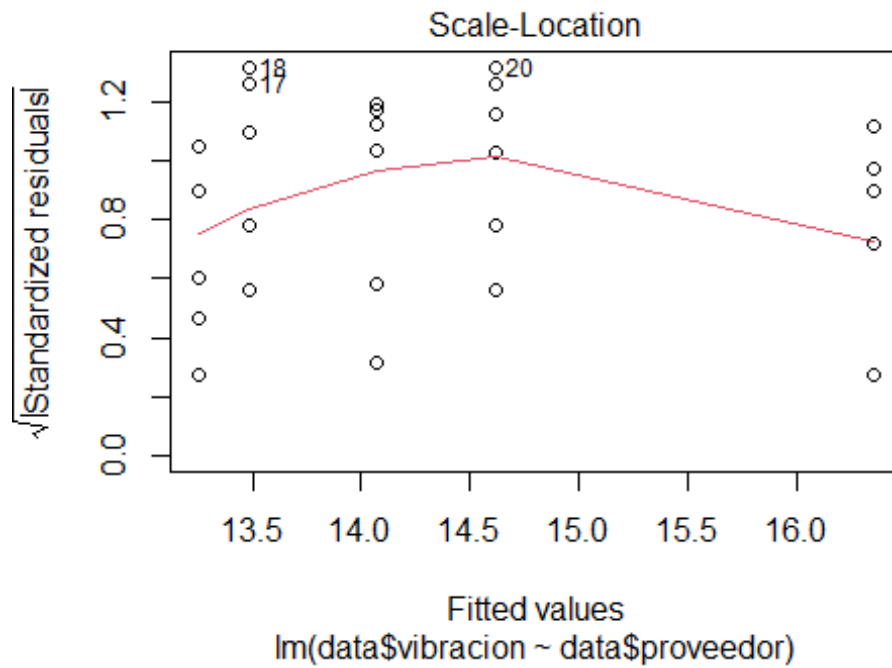
```
estimados")
abline(h=0,col="red")
```



```
#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(data$vibracion~data$proveedor))
```





CD=  $150/(150+76)$  #coeficiente de determinación para el modelo.

### Concluye en el contexto del problema.

En el contexto del problema se puede identificar que la mayor causa de las vibraciones de los motores electricos depende del Proveedor. Especificamente con los productos del Proveedor 2 las vibraciones de los motores electricos aumenta.