

Actividad 9

Saúl Francisco Vázquez del Río

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

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)
```

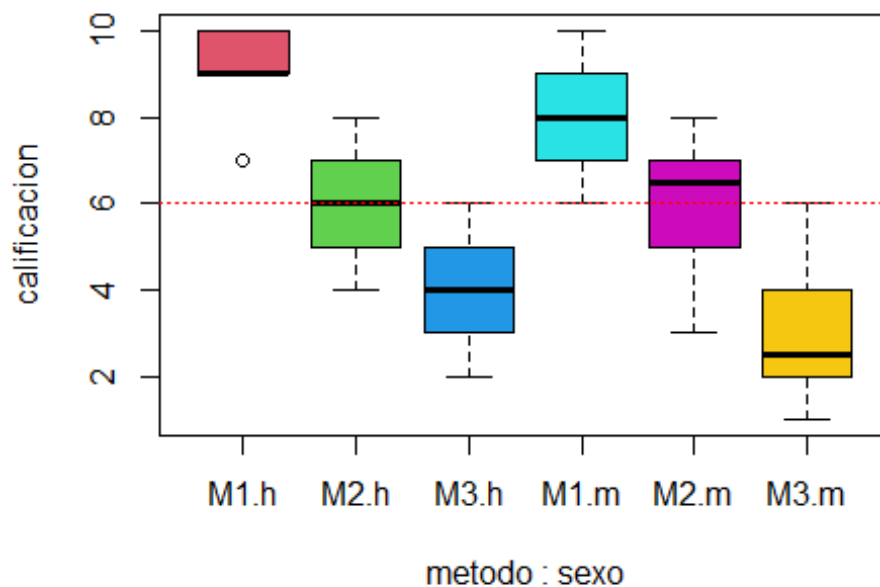
```
datos = data.frame(calificacion, metodo, sexo)
```

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

```
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
```

```
boxplot(calificacion~metodo+sexo, col = 2:8)
abline(h = mean(calificacion), lty = 3, col = "red")
```



Valor

Frontera

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

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

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

Se puede observar el metodo 1 es el que tiene un mejor desempeño tanto en hombre como en mujeres, el metodo 2 no se nota un gran diferencia ya que este pasa por la media y el metodo 3 es el que tiene peor desempeño de los 3 ya que esta a la izquierda de la grafica

Escribe tus conclusiones parciales

Que el mejor metodo de los 3 tanto para hombres y mujeres es el metodo 1 ya que es el que tiene una mayor dierenfia de los 3

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

##Primer hipotesis

$$h_0: T_i = 0$$

h_1: algun T_i es distinto de cero

##Segunda hipotesis

$$h_0: a_j = 0$$

h_1 : algun a_j es distinto de cero

##Tercer hipotesis

$$h_0: T_i a_j = 0$$

h_1 : algun T_{ia_j} es distinto de cero

Realiza el ANOVA para dos niveles con interacción:

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

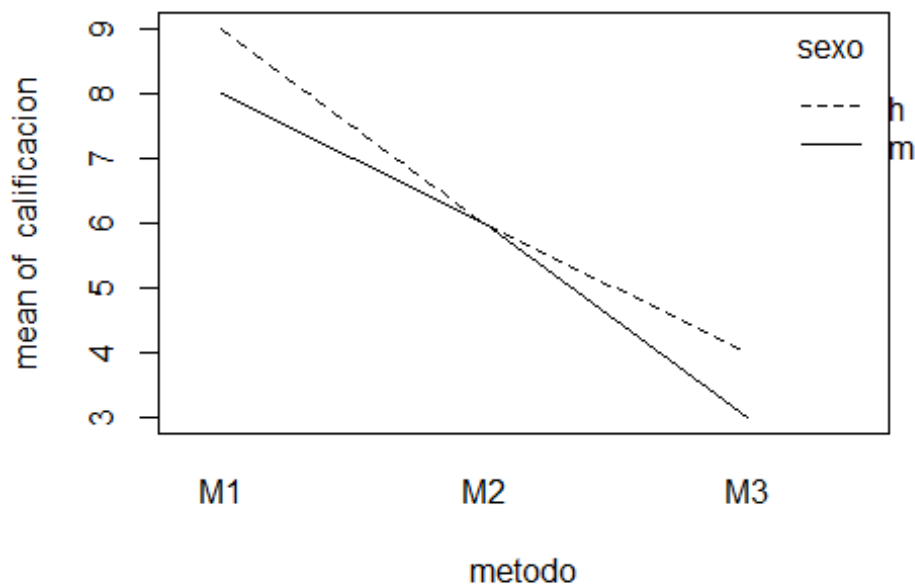
```
anova = aov(calificacion~metodo*sexo,datos)
summary(anova)
```

	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			

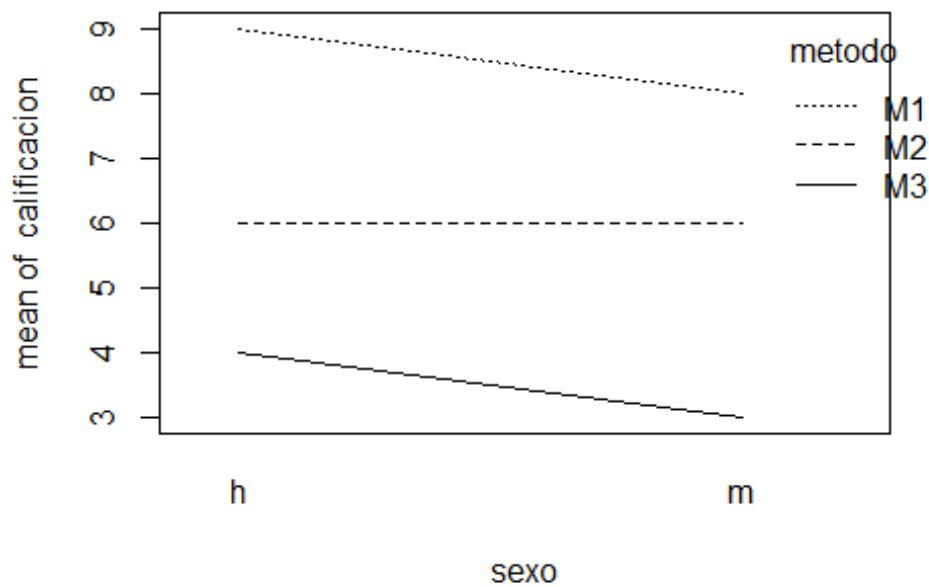
Signif. codes:	0	'***'	0.001	'**'	0.01	'*' 0.05 '.' 0.1 ' ' 1

Haz la gráfica de interacción de dos factores en ANOVA

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



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



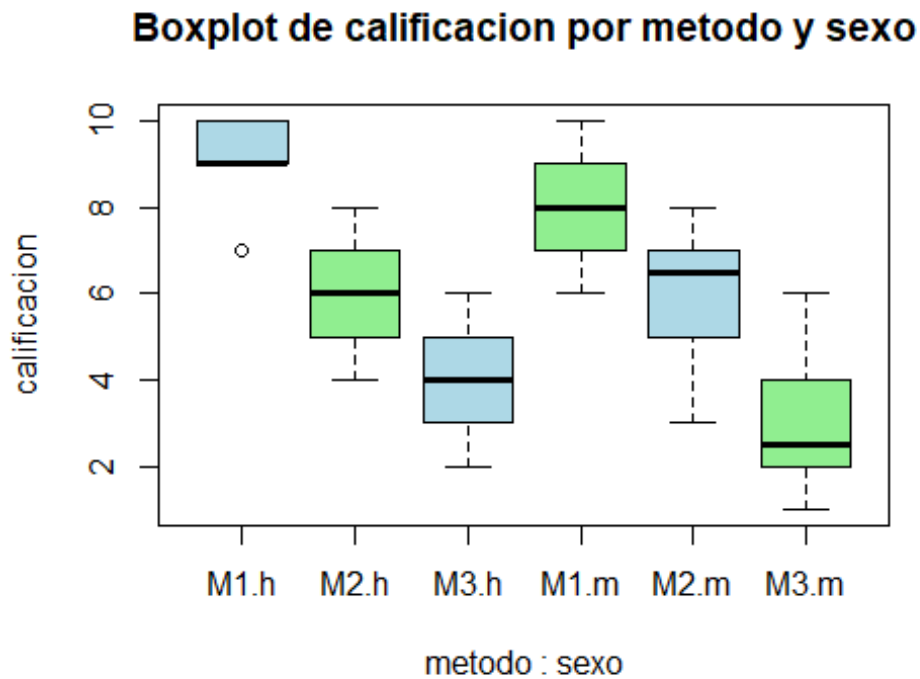
```
A2 = aov(calificacion~metodo+sexo,datos)
```

```
summary(A2)
```

```
##           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 para visualizar la interacción de los factores, por ejemplo, peso por dieta interacción ejercicio:

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



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

En la primera grafica de interaccion se nota que tanto como hombres y mujeres aumentan su desempeño en calificaciones si se usa el metodo 1, haciendo el peor metodo el 3 y dejando al metodo 2 como un desempeño medio por parte de ambos sexos.

En la segunda grafica de interpretacion se observa que el metodo 1 es el que mejor se desempeña en calificaiones, el metodo 2 teniendo un desempeño medio en calificaciones y el metodo 3 es el que peor se desempeña en calificaciones.

En el boxplot se observa el metodo 1 es el que tiene un mejor desempeño tanto en hombre como en mujeres, el metodo 2 no se nota un gran diferencia ya que este pasa por la media y el metodo 3 es el que tiene peor desempeño de los 3 ya que esta a la izquierda de la grafica.

Escribe tus conclusiones parciales

Se puede observar en las graficas interaccion que el metodo 1 es el que mejor se desarrolla de los 3 metodos al igual que en el box plot haciendo que el que tenga un peor desempeño en calificaciones siendo el metodo y dejando al metodo 2 como un desempeño medio en calificaciones.

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. Grafícalos

```
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

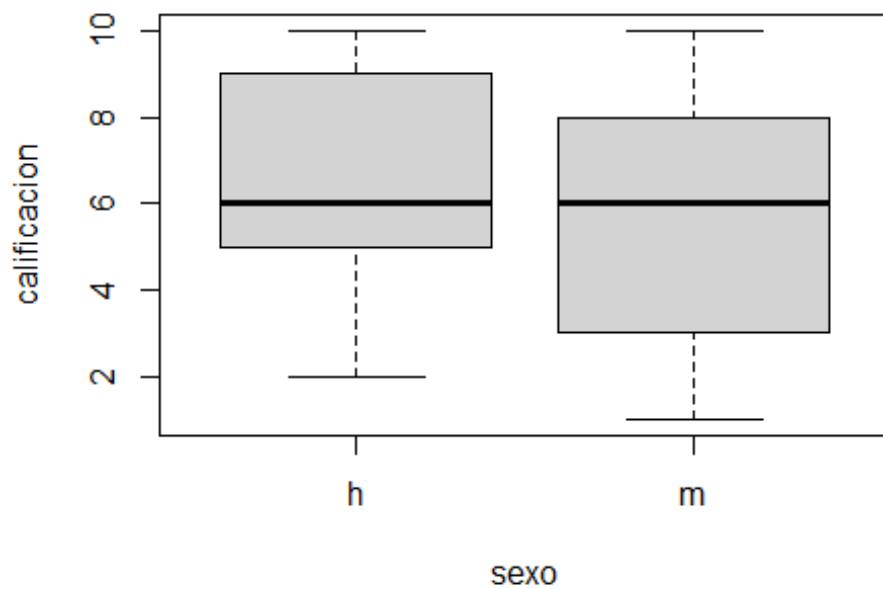
tapply(calificacion,metodo,mean)

##  M1  M2  M3
## 8.5 6.0 3.5

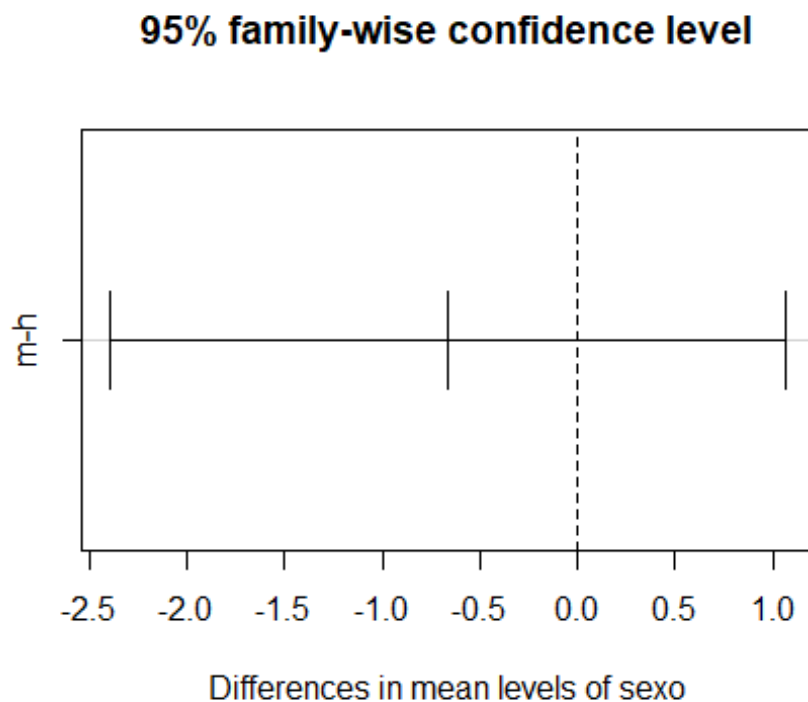
M=mean(calificacion)
M

## [1] 6

boxplot(calificacion ~ sexo)
```



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



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

En la primera grafica se puede observar que los hombres tienen un mejor desempeño en calificaiones que las mujeres, pero este dato no es tan relevante a comparacion de metodo.

Y en la segunda grafica no se nota tanta diferencia entre los hombre y mujeres ya que estan iguales en sus intervalos de confianza.

Escribe tus conclusiones parciales

Gracias a la interpretacion de las graficas podemos determinar que el sexo no es un valor tan importante como el metodo de educacion que se les imparte.

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.

```
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

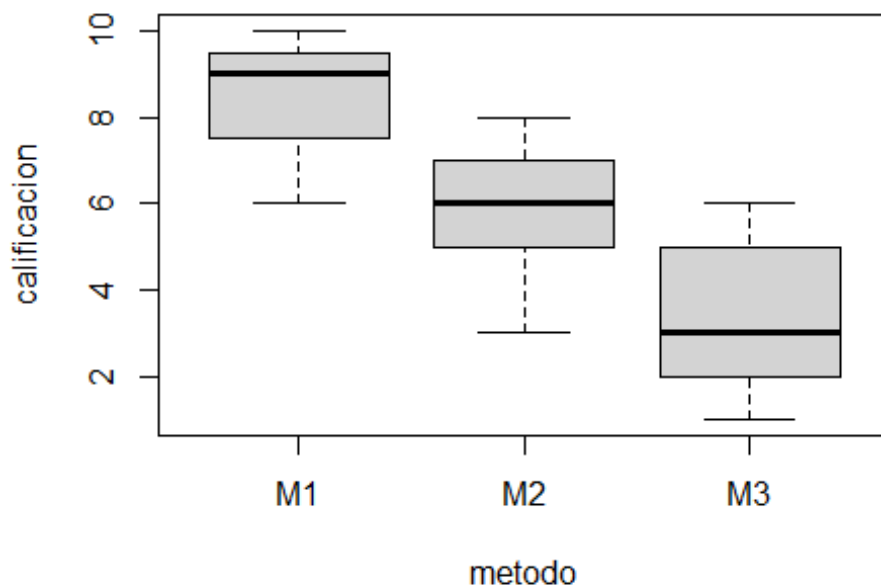
tapply(calificacion,metodo,mean)

##  M1  M2  M3
## 8.5 6.0 3.5

M=mean(calificacion)
M

## [1] 6

boxplot(calificacion ~ metodo)
```



```
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)
```



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

En la grafica del box plot se observa que el metodo 1 es que tiene un mejor desempeño en calificaciones, el metodo 2 teniendo un desempeño medio y el metodo 3 teniendo el peor desempeño en calificaciones.

En la grafica de intervalos de confianza se aprecia que se tiene más confianza hacia el metodo 1 y que el nivel de desconfianza es el metodo 3.

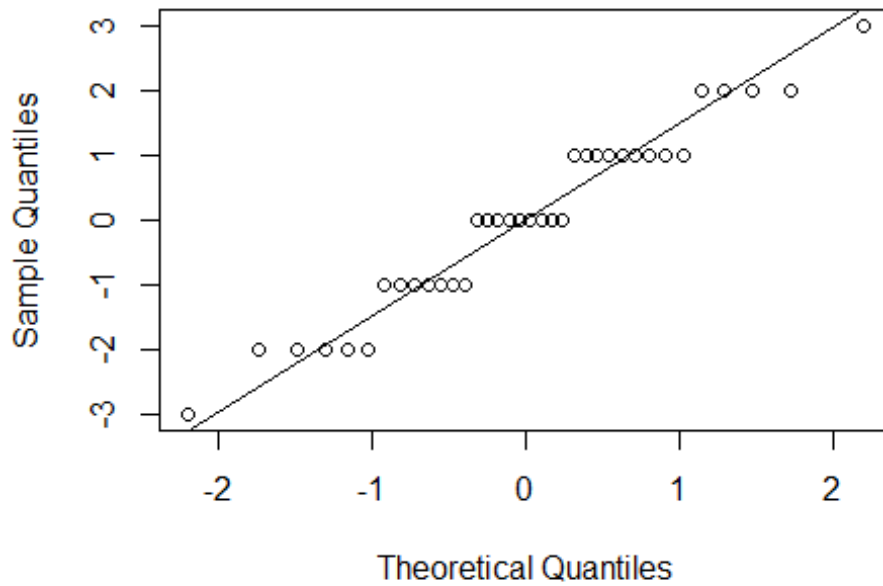
Escribe tus conclusiones parciales

Gracias a la interpretacion de las graficas podemos determinar que el metodo 1 es un valor importante ya que este puede hacer un aumento en calificaciones.

Comprueba la validez del modelo. Comprueba: Normalidad Homocedasticidad Independencia Relación lineal entre las variables (coeficiente de determinación).

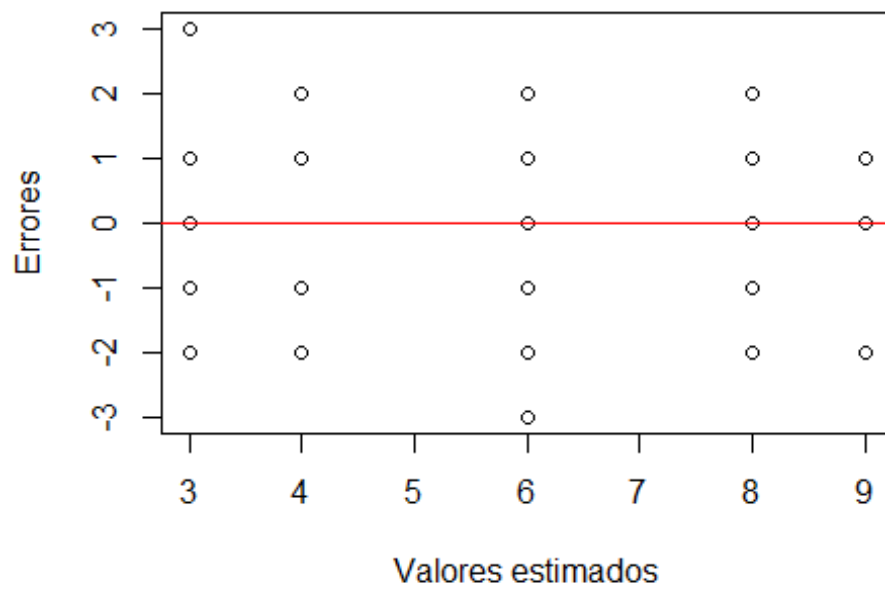
```
##Normalidad
residuos=anova$residuals
qqnorm(residuos)
qqline(residuos)
```

Normal Q-Q Plot



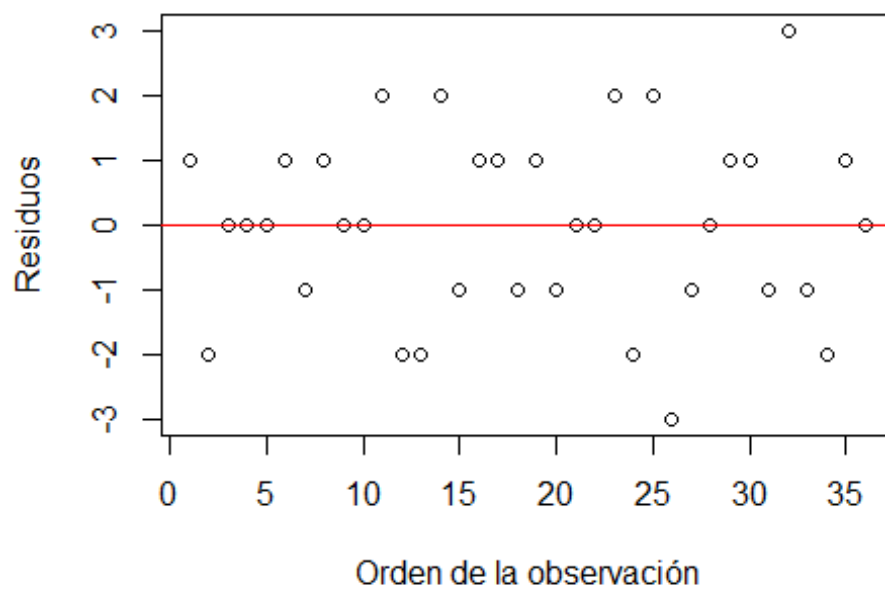
##Homocedasticidad

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

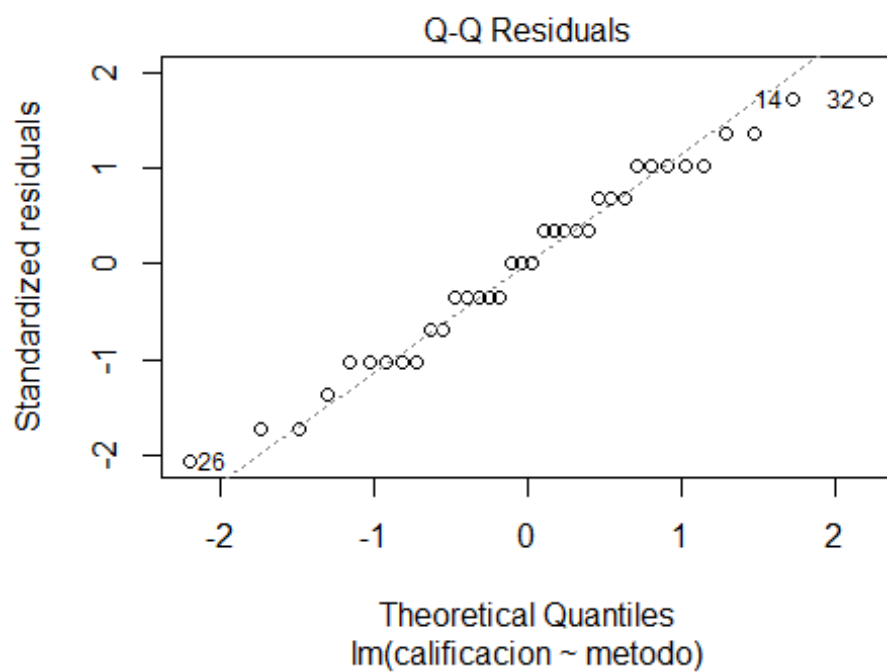
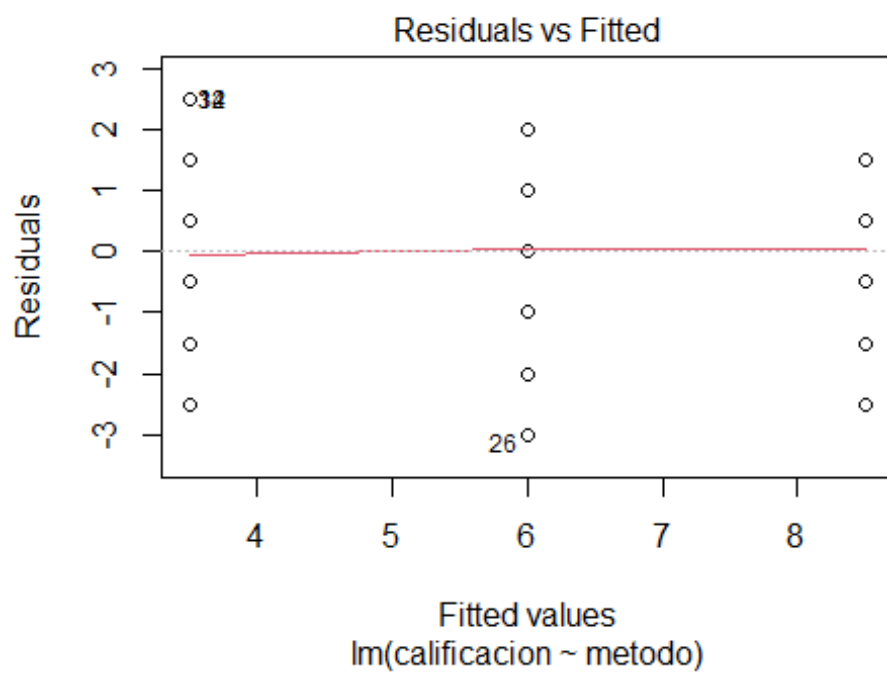


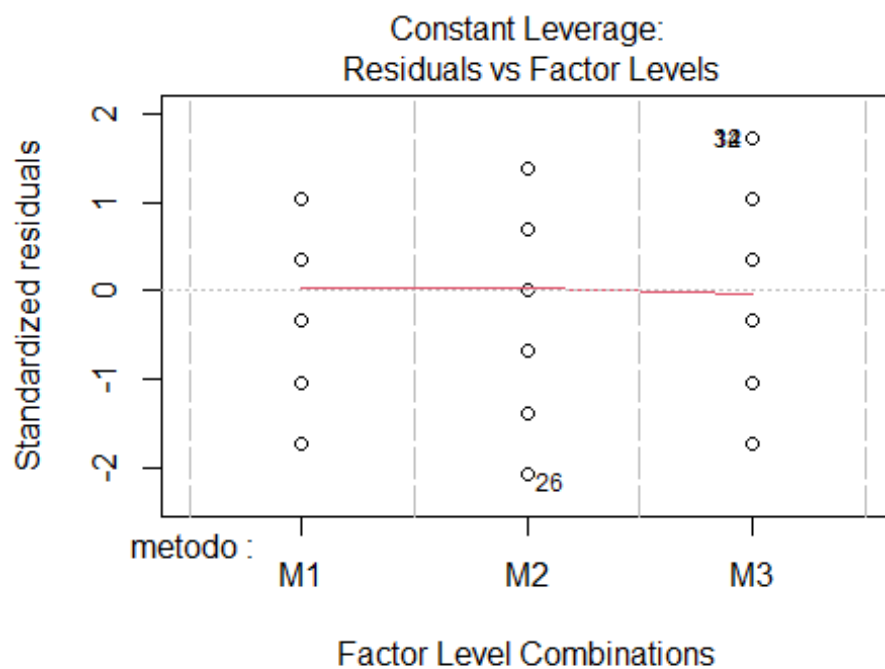
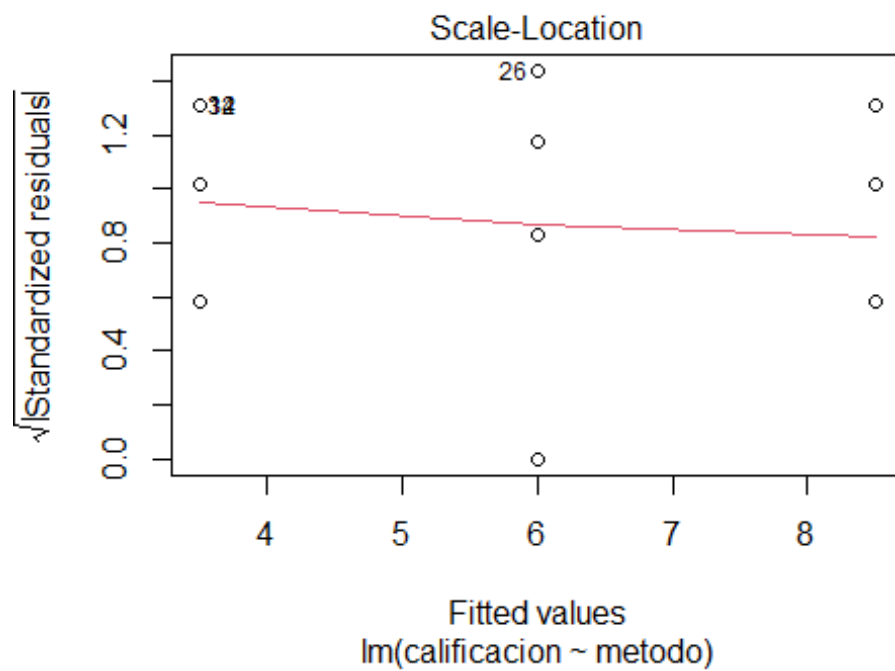
##Independencia

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



```
##Coeficiente de determinacion  
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.

En conclusion el metodo de educacion es el más importante para ver si los alumnos tienen un mejor rendimiento en sus calificaciones, haciendo que el dato del sexo tenga un menor impacto a la hora de las calificaciones, ademas gracias al r cuadrada podemos ver que un metodo de enseñansa tiene el 66% de afectar las calificaciones.

##Problema 2

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.

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

```
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, 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)
```

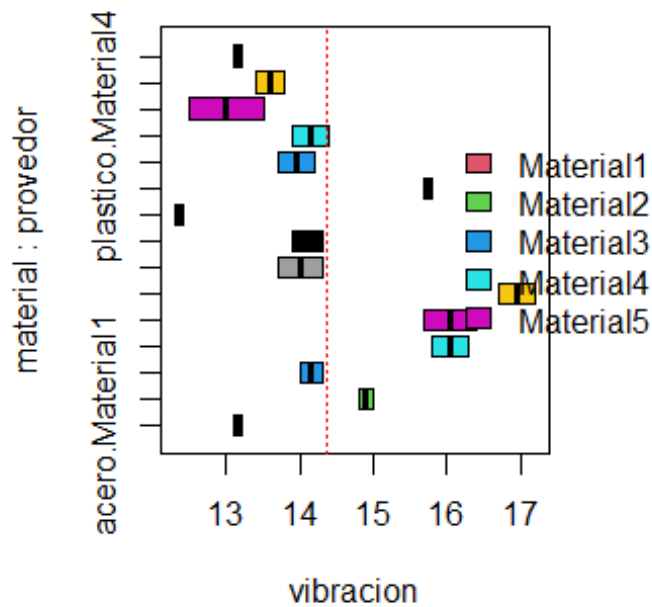
```
proveedor = c(rep("Material1", 6), rep("Material2", 6), rep("Material3",
6), rep("Material4", 6), rep("Material5", 6))
material = c(rep("acero", 2), rep("aluminio", 2), rep("plastico", 2))
```

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

```
datos <- data.frame(vibracion, material, proveedor)
```

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

```
par(mar = c(5, 7, 4, 8))
boxplot(vibracion ~ material : proveedor, datos, col = 2:28, horizontal =
TRUE)
abline(v = mean(vibracion), lty = 3, col = "red")
legend("right", inset = c(-0.3, 0), legend = levels(datos$proveedor), fill
= 2:28, xpd = TRUE, bty = "n")
```

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

Se observa que los materiales que el proveedor 2 trae son los que más vibran y que los materiales de los demás proveedores no vibran tanto.

Escribe tus conclusiones parciales.

Que los materiales del proveedor 2 son los que más vibran comparados con los demás materiales de los proveedores.

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

##Primer hipótesis

$$h_0: T_i = 0$$

h_1 : algún T_i es distinto de cero

##Segunda hipótesis

$$h_0: a_j = 0$$

h_1 : algún a_j es distinto de cero

##Tercer hipótesis

$$h_0: T_i a_j = 0$$

h_1 : algún $T_i a_j$ es distinto de cero

Realiza el ANOVA para dos niveles con interacción:

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

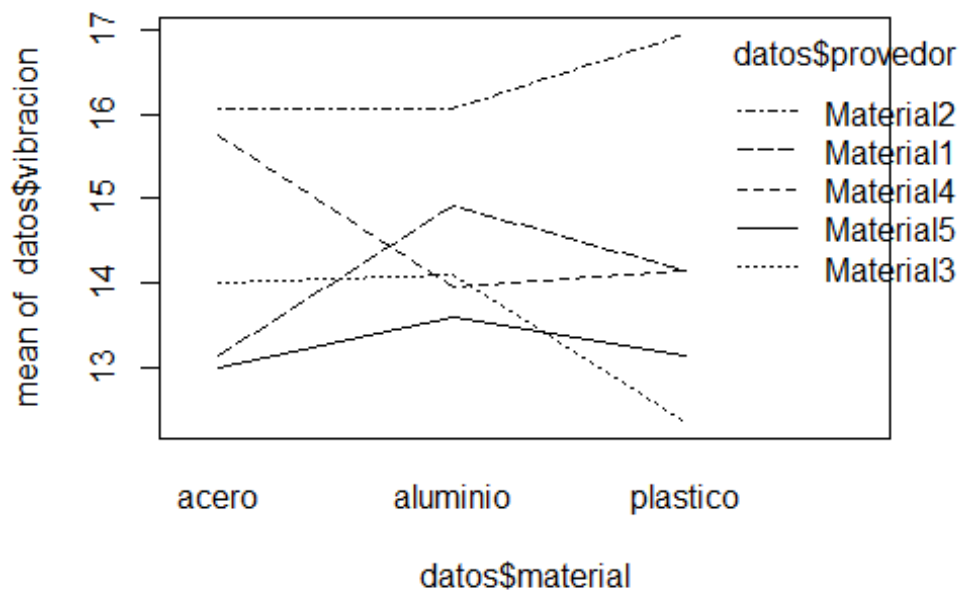
```
anova = aov(vibracion ~ material*proveedor, datos)
summary(anova)
```

	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		

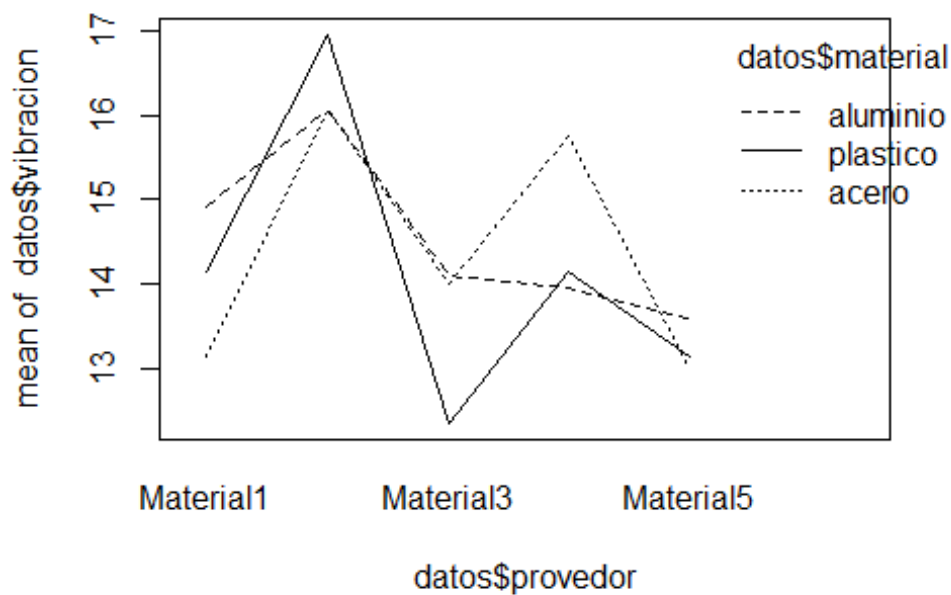
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Haz la gráfica de interacción de dos factores en ANOVA

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



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



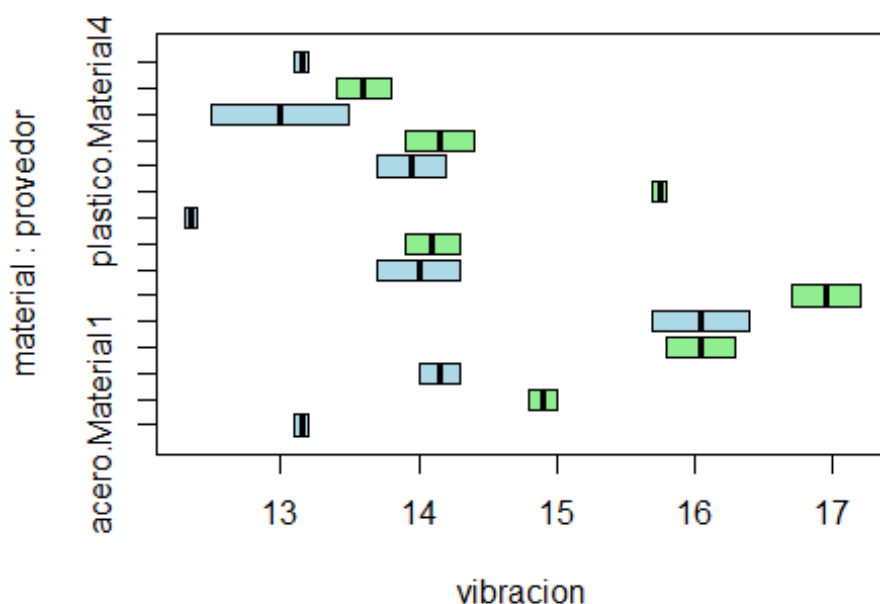
```
A2 = aov(vibracion ~ material+proveedor, datos)
summary(A2)
```

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

boxplot(vibracion~material*proveedor, data = datos, horizontal = TRUE, col
= c("lightblue", "lightgreen"), main = "Boxplot de calificacion por
material y proveedor")
```

Boxplot de calificacion por material y proveedor



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

Gracias a la primera grafica podemos observar que el material que más vibra es plastico y que es del proveedor 2.

En la segunda grafica podemos observar que los materiales del proveedor 2 son los que más vibran.

Escribe tus conclusiones parciales

Gracias a las dos graficas se puede observar que el material que más vibra es el plastico y que los materiales el proveedor 2 son los que más vibran.

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. Grafícalos

```
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(vibracion,provedor,mean)

## Material1 Material2 Material3 Material4 Material5
## 14.06667  16.35000  13.48333  14.61667  13.25000

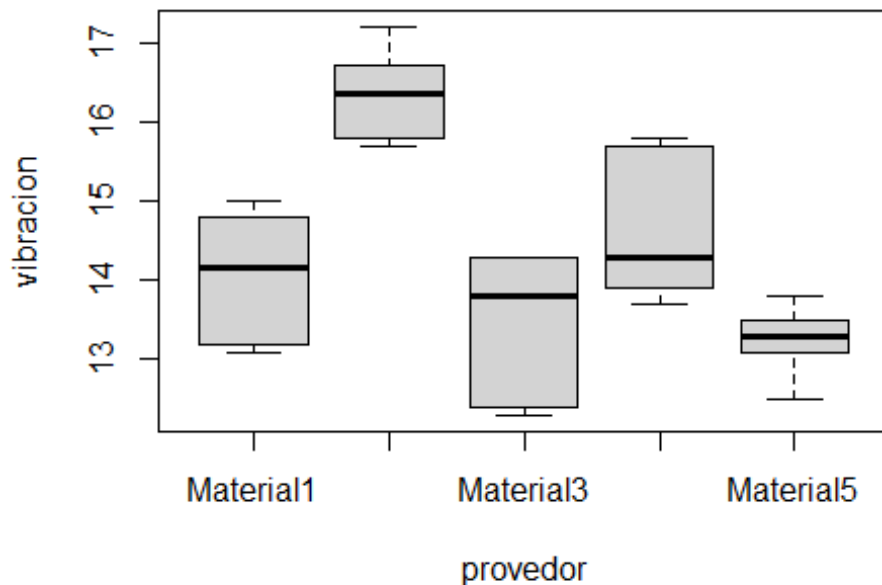
tapply(datos$vibracion,datos$provedor,mean)

## Material1 Material2 Material3 Material4 Material5
## 14.06667  16.35000  13.48333  14.61667  13.25000

M=mean(vibracion)
M

## [1] 14.35333

boxplot(vibracion ~ provedor)
```

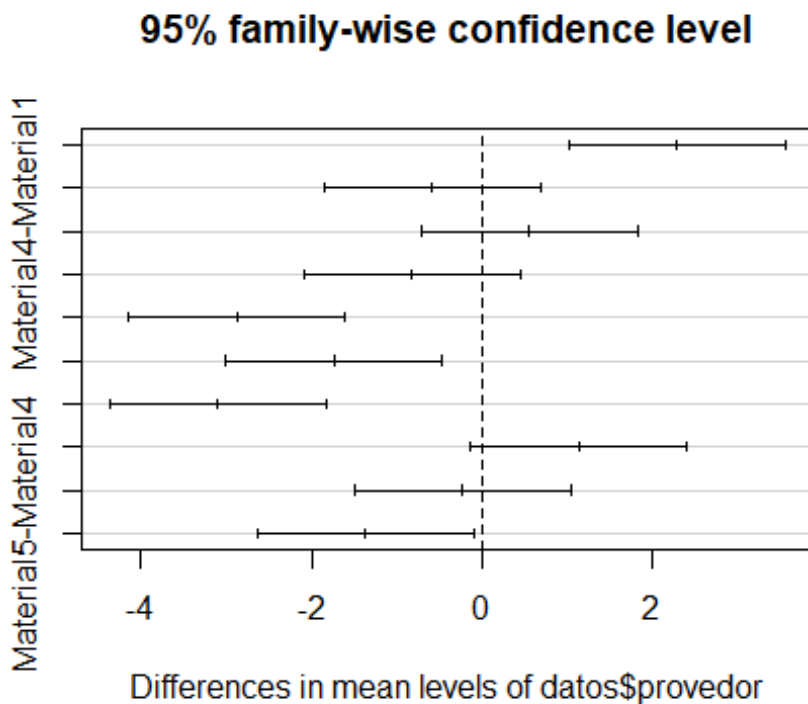


```
I = TukeyHSD(aov(datos$vibracion ~ datos$provedor))
I

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = datos$vibracion ~ datos$provedor)
##
```

```
## `$datos$proveedor`
##
##          diff      lwr      upr    p adj
## Material2-Material1  2.283333  1.0153666  3.55130006 0.0001595
## Material3-Material1 -0.5833333 -1.8513001  0.68463339 0.6630108
## Material4-Material1  0.5500000 -0.7179667  1.81796672 0.7089904
## Material5-Material1 -0.8166667 -2.0846334  0.45130006 0.3474956
## Material3-Material2 -2.8666667 -4.1346334 -1.59869994 0.0000055
## Material4-Material2 -1.7333333 -3.0013001 -0.46536661 0.0039774
## Material5-Material2 -3.1000000 -4.3679667 -1.83203328 0.0000015
## Material4-Material3  1.1333333 -0.1346334  2.40130006 0.0959316
## Material5-Material3 -0.2333333 -1.5013001  1.03463339 0.9821261
## Material5-Material4 -1.3666667 -2.6346334 -0.09869994 0.0301318
```

```
plot(I)
```



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

En el box plot podemos observar que los materiales del proveedor 2 son los más probables a vibrar.

Y en la grafico de intervalos de confianza vemos que los materiales del proveedor tienden más a vibrar.

Escribe tus conclusiones parciales

Gracias a ambas graficas podemos sacar la conclusion que los materiales que más probabilidad de vibrar son los del proveedor 2

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.

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

##              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

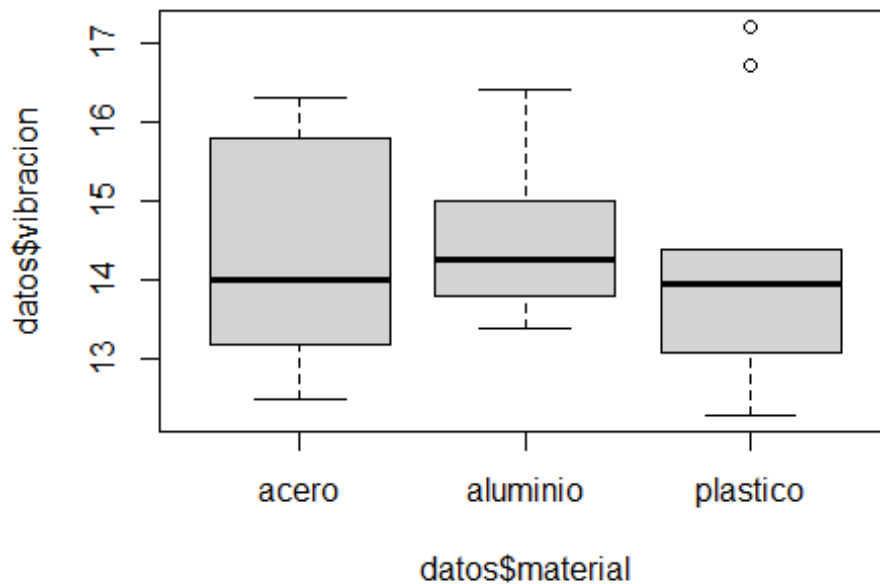
tapply(datos$vibracion,datos$material,mean)

##      acero aluminio plastico
##      14.39    14.52    14.15

M=mean(vibracion)
M

## [1] 14.35333

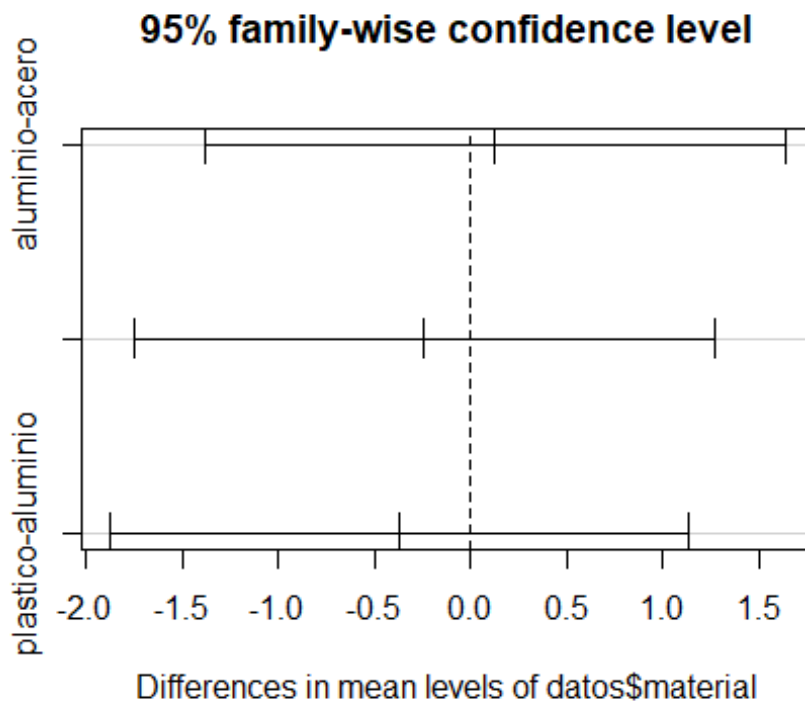
boxplot(datos$vibracion ~ datos$material)
```



```

I = TukeyHSD(aov(datos$vibracion ~ datos$material))
I
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = datos$vibracion ~ datos$material)
##
## $`datos$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)

```



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

En el box plot observamos que no hay mucha diferencia en que material vibra más.

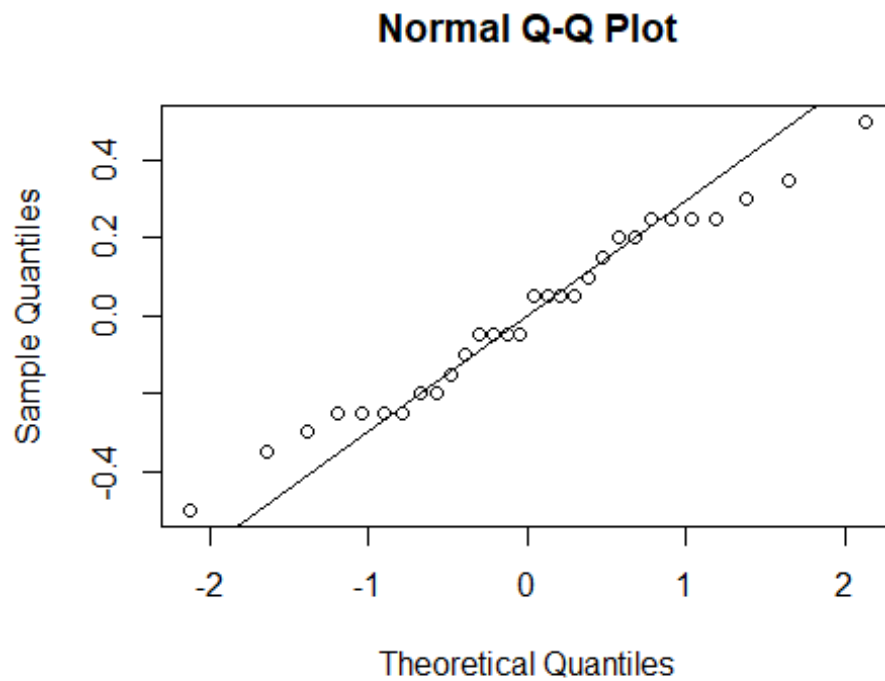
Y en la grafica de intervalo de confianza vemos que no hay mucha diferencia en ver que material vibra más.

Escribe tus conclusiones parciales

Comprueba la validez del modelo. Comprueba: Normalidad Homocedasticidad Independencia Relación lineal entre las variables (coeficiente de determinación).

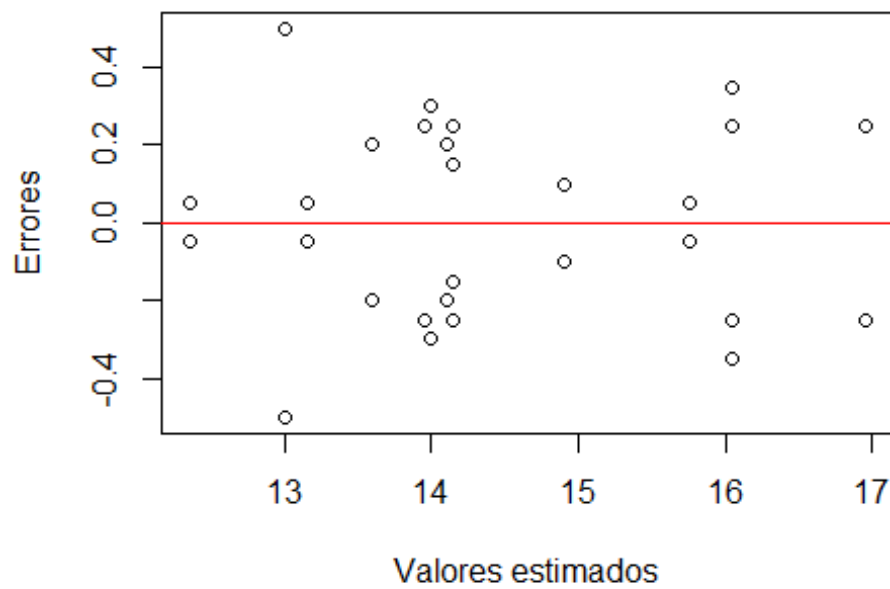
##Normalidad

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



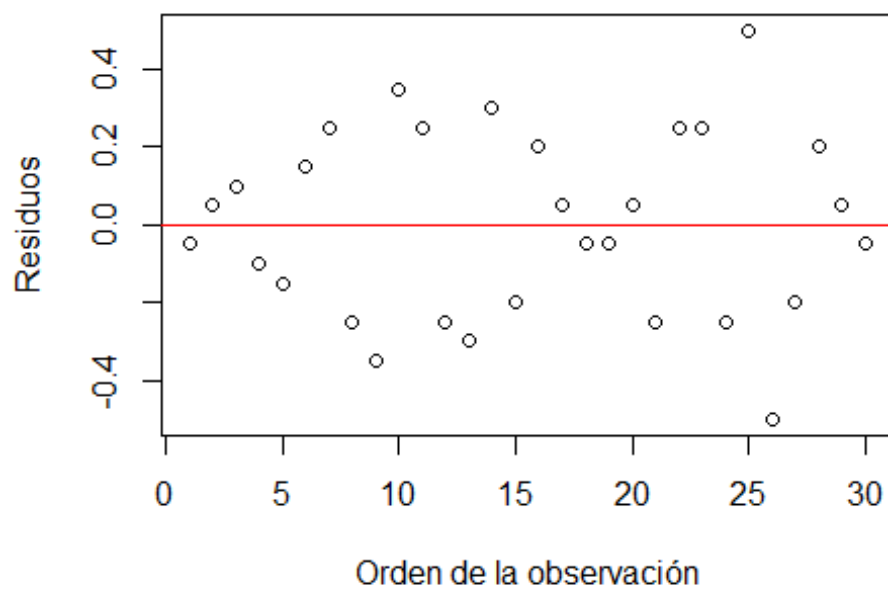
##Homocedasticidad

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

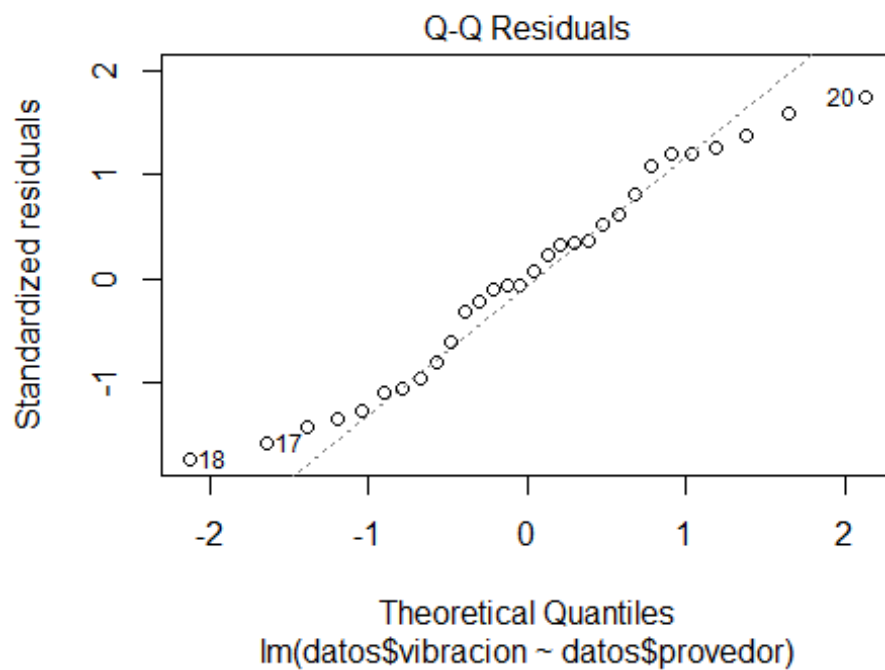
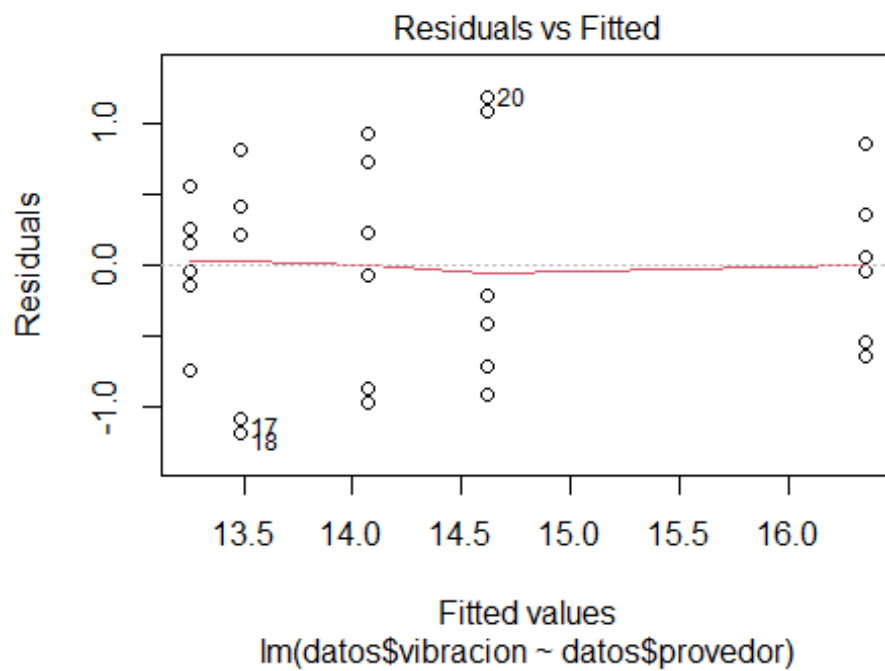


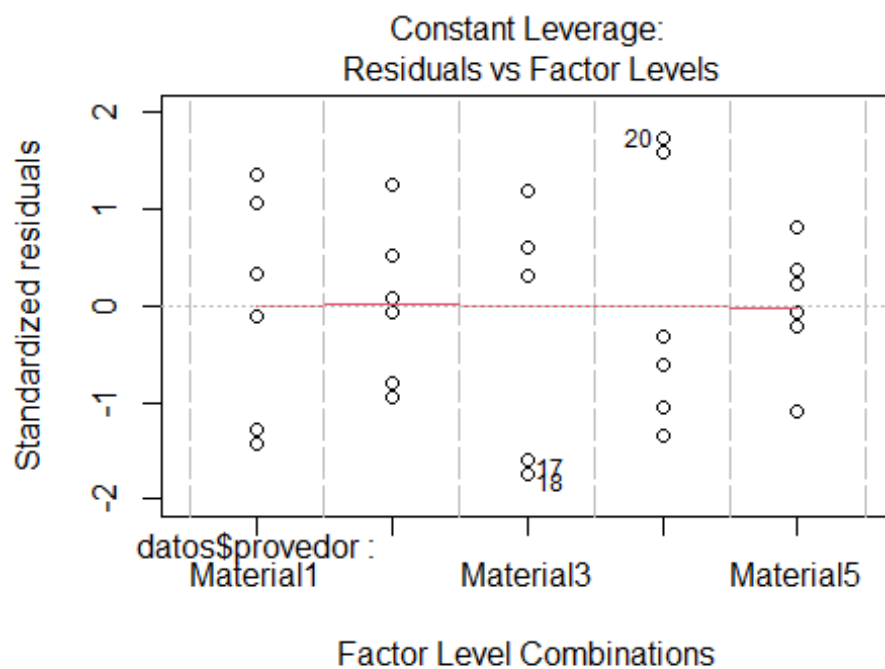
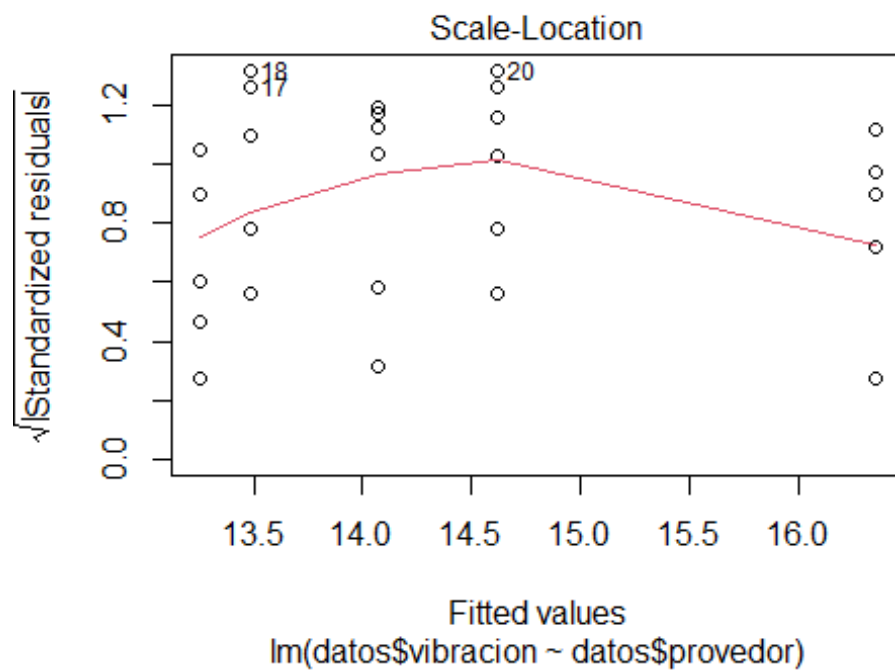
##Independencia

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



```
##Coeficiente de determinacion  
plot(lm(datos$vibracion~datos$proveedor))
```





CD= 0.70/(0.70+49.95) *#coeficiente de determinación para el modelo.*

```
summary(lm(datos$vibracion~datos$proveedor))$r.squared
```

[1] 0.7240136

Concluye en el contexto del problema.

En conclusion se puede ver el material más probable que vibre es el plastico y que de los 5 proveedores que hay el proveedor 2 tiene más probabilidades que su materiales vibren, ademas que gracias a la r cuadrada podemos observar que hay un 72% que los materiales y el proveedor de estos afectan significativamente a la vibracion de los motores.