act_9_A01742161

Rogelio Lizárraga

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.

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

```r
rendimiento <- 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)

m <- tapply(rendimiento, metodo, mean)
print(m)
```

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

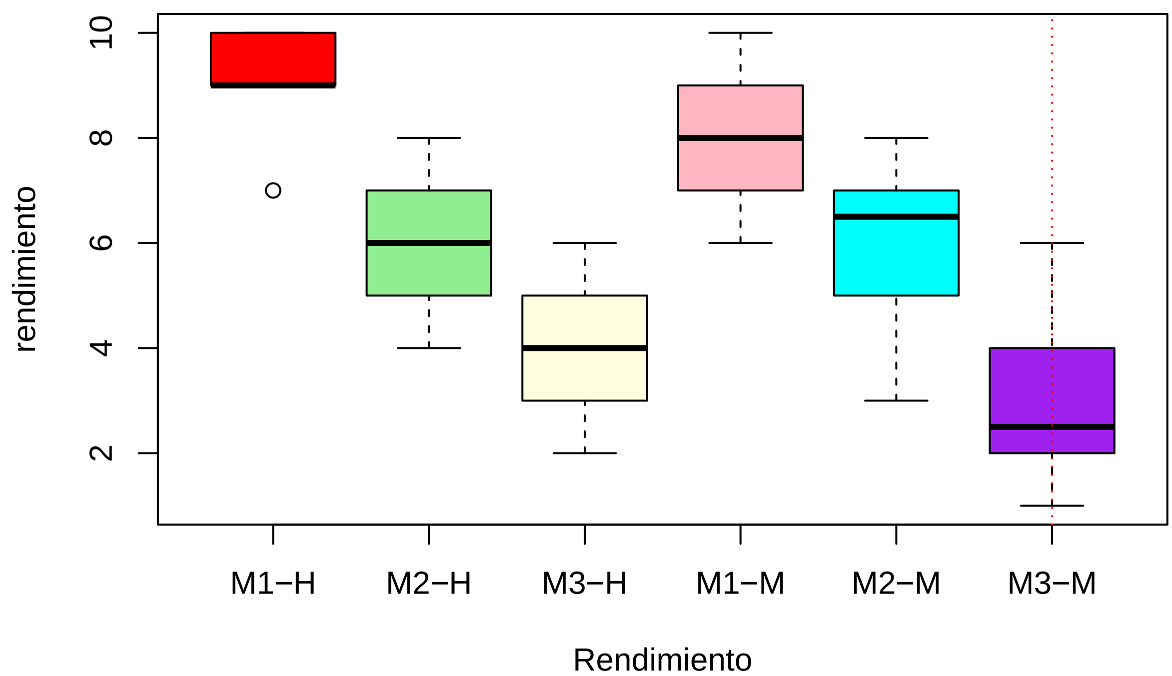
```r
print(mean(rendimiento))
```

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

```r
cat('Efectos:' ,(m- mean(rendimiento)))
```

```
## Efectos: 2.5 0 -2.5
```

```r
boxplot(rendimiento ~ metodo * sexo, col = c("red", "lightgreen", "lightyellow", "lightpink", "cyan", "
        horizontal = FALSE, xlab = "Rendimiento",
        names = c("M1-H", "M2-H", "M3-H", "M1-M", "M2-M", "M3-M"))

abline(v = mean(rendimiento), lty = 3, col = "red")
```



Observamos que el rendimiento varía conforme el método, pues sus medias son diferentes. Además, observamos en las gráficasa de caja de bigotes que el sexo también influye en la diferencia de los métodos.

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

Primera hipótesis

$H_0 : \tau_i = 0$

$H_1 : \exists \tau_i \neq 0$

Segunda hipótesis

$H_0 : \alpha_j = 0$

$H_1 : \exists \alpha_j \neq 0$

Tercera hipótesis

$H_0 : \tau_j \alpha_j = 0$

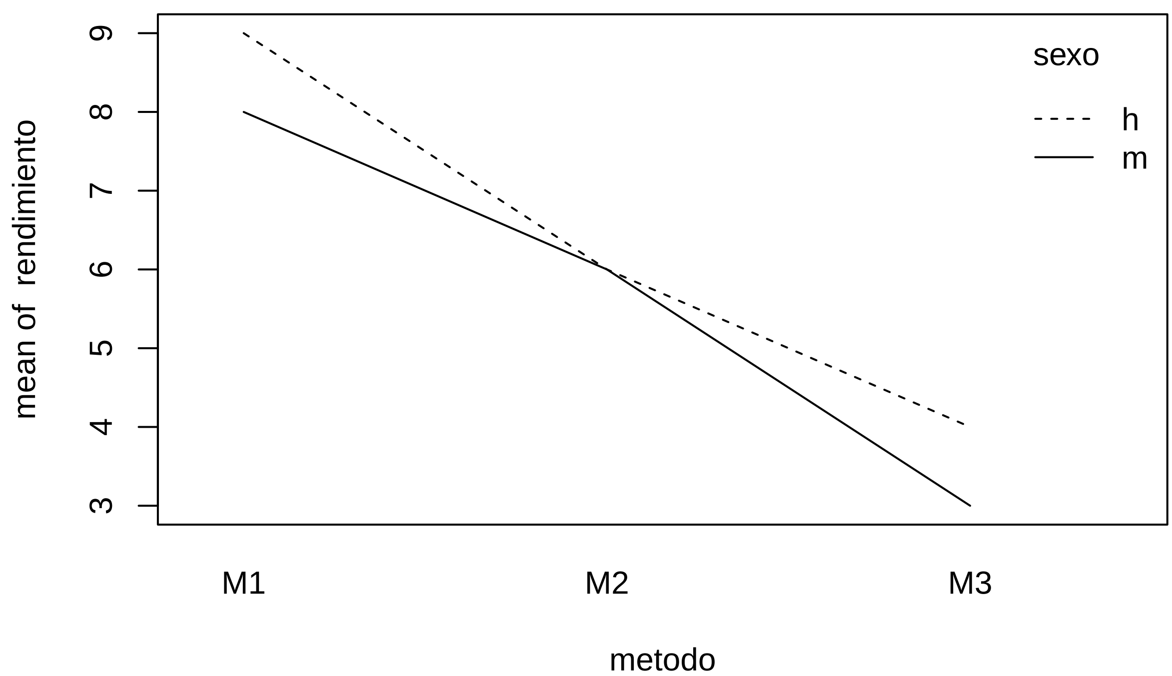
$H_1 : \exists \tau_j \alpha_j \neq 0$

Realiza el ANOVA para dos niveles con interacción:

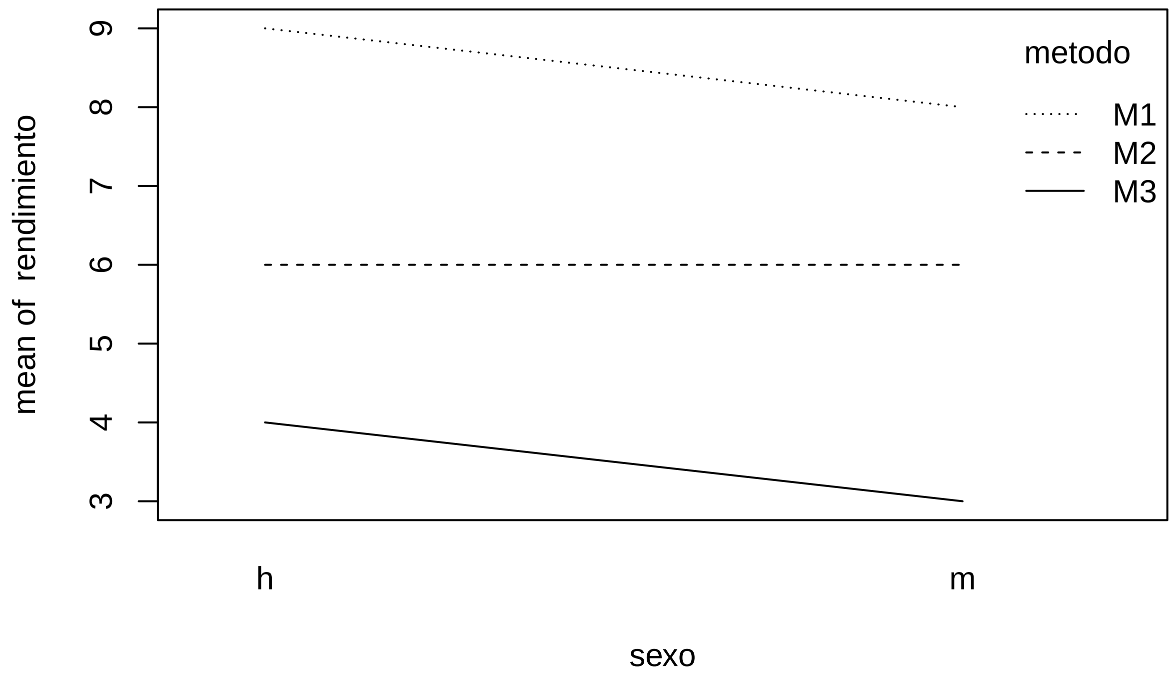
```
A<-aov(rendimiento~metodo*sexo)
summary(A)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## metodo        2    150   75.00  32.143 3.47e-08 ***
## sexo          1      4    4.00   1.714    0.200    
## metodo:sexo   2      2    1.00   0.429    0.655    
## Residuals    30     70    2.33                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

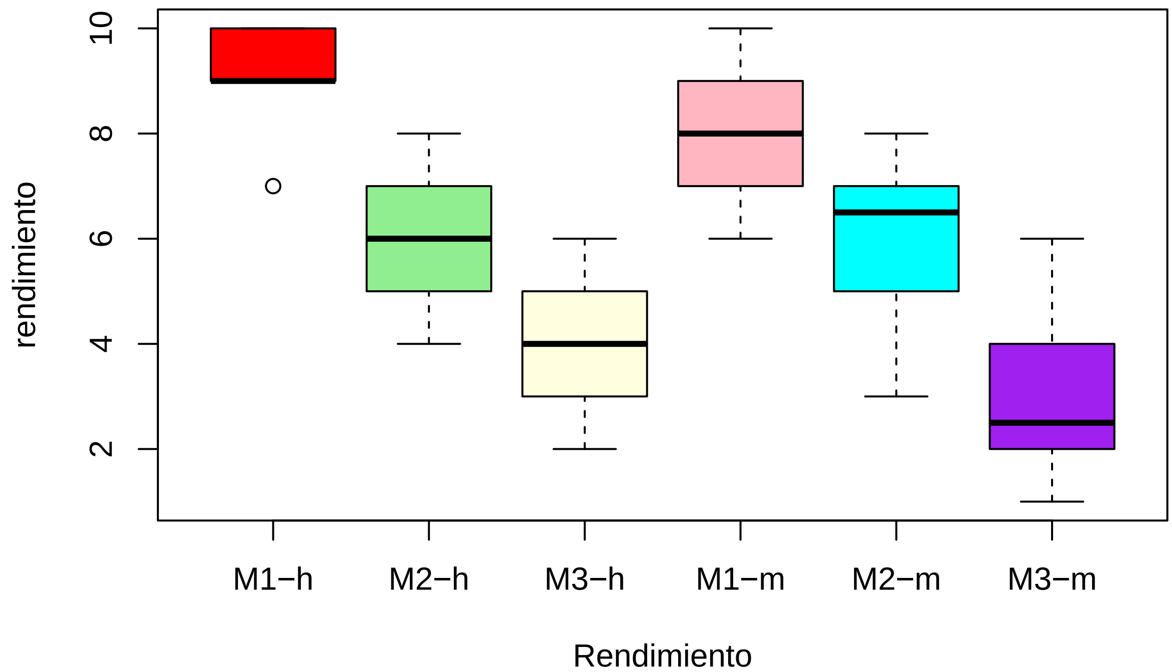
```r
interaction.plot(metodo,sexo, rendimiento)
```




```r
interaction.plot(sexo, metodo, rendimiento)
```




```
boxplot(rendimiento ~ metodo * sexo, col = c("red", "lightgreen", "lightyellow", "lightpink", "cyan", "p
        horizontal = FALSE, xlab = "Rendimiento",
        names = c("M1-h", "M2-h", "M3-h", "M1-m", "M2-m", "M3-m"))
```




```
m <- tapply(rendimiento, metodo, mean)
print(m)
```

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

```
print(mean(rendimiento))
```

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

```
cat('Efectos:' ,(m- mean(rendimiento)))
```

```
## Efectos: 2.5 0 -2.5
```

Observamos que la media de los rendimientos parecen no ser estadísticamente distintas. En cambio, la media por método sí son distintas, pues no se observa ningún cruce entre las medias. Además, se observa que el método 1 tiene el mayor rendimiento, tanto en hombres como mujeres.

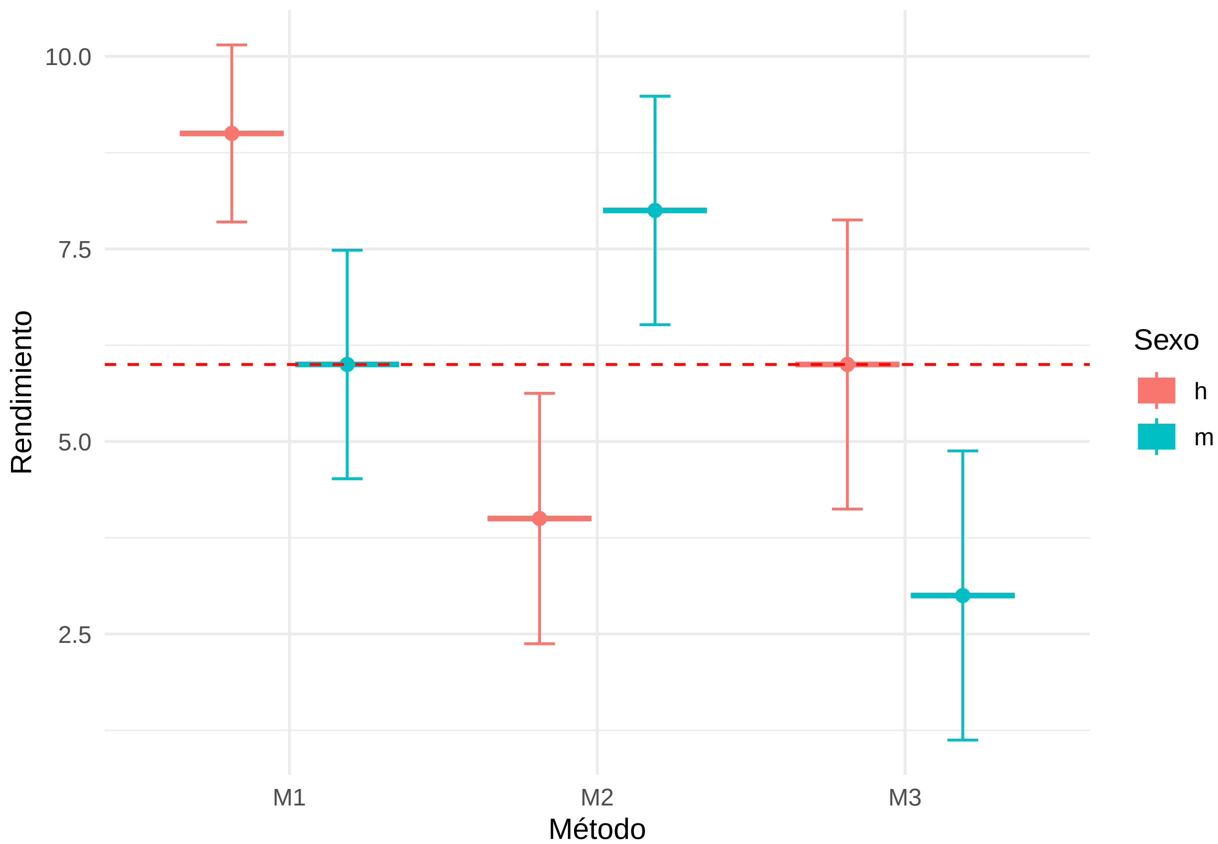
Realiza el ANOVA para dos niveles sin interacción.

Consulta el código de R en los apoyos de clase de "ANOVA". Haz los intervalos de confianza de rendimiento por sexo. Grafícalos Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. scribe tus conclusiones parciales

```r
B<-aov(rendimiento~metodo+sexo)
library(ggplot2)
means <- tapply(rendimiento, list(metodo, sexo), mean)
sds <- tapply(rendimiento, list(metodo, sexo), sd)
n <- tapply(rendimiento, list(metodo, sexo), length)
print(means)
```

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

```r
sem <- sds / sqrt(n)
conf.low <- means - qt(0.975, n - 1) * sem
conf.high <- means + qt(0.975, n - 1) * sem
data <- data.frame(
    Metodo = rep(levels(metodo), each = 2),
    Sexo = rep(levels(sexo), times = 3),
    Rendimiento = c(means),
    Conf.low = c(conf.low),
    Conf.high = c(conf.high)
)
ggplot(data, aes(x = Metodo, y = Rendimiento, fill = Sexo, color =Sexo)) +
    geom_boxplot(aes(ymin = Conf.low, ymax = Conf.high)) +
    geom_point(position = position_dodge(width = 0.75), size = 2) +
    geom_errorbar(aes(ymin = Conf.low, ymax = Conf.high), width = 0.2, position = position_dodge(width =
    geom_hline(yintercept = mean(rendimiento), linetype = "dashed", color = "red") +
    labs(x = "Método", y = "Rendimiento") +
    theme_minimal()
```




```r
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
```

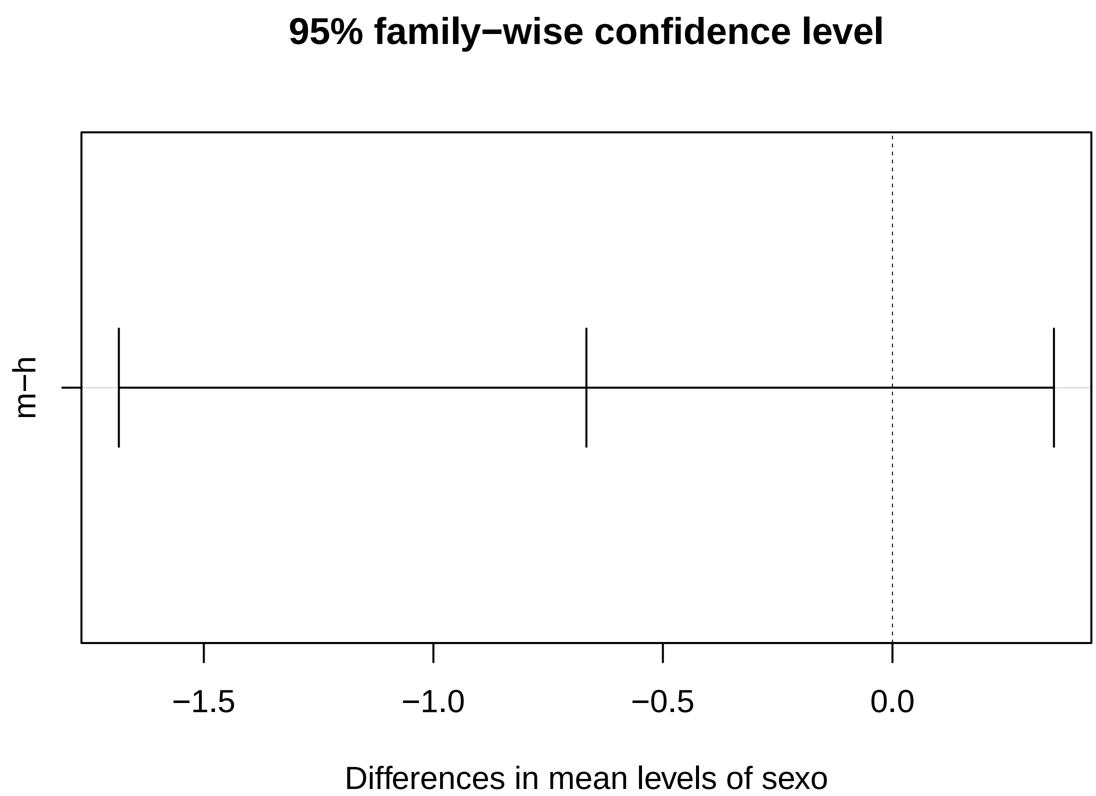
Se observa que las medias difieren por el método y también difieren por sexo, pero las diferencias por sexo parecen no ser estadísticamente significativas. Observamos que el sexo tiene una suma cuadrada muy baja, por lo que parece que este factor no es significativo en el modelo. Por otro lado, observamos que las medias sí difieren conforme el método varía.

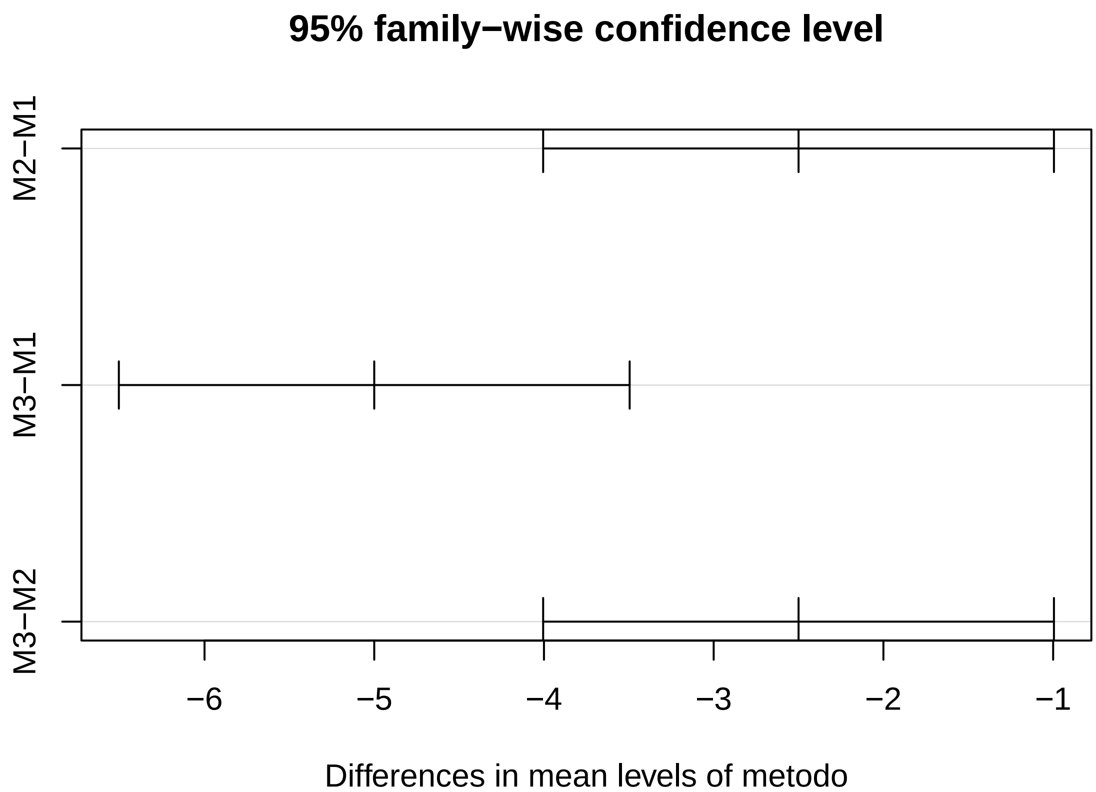
```r
modelo_anova <- aov(rendimiento ~ sexo + metodo)
prueba_tukey <- TukeyHSD(modelo_anova)
print(prueba_tukey)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = rendimiento ~ sexo + metodo)
##
## $sexo
```

```
##            diff       lwr    upr     p adj
## m-h -0.6666667 -1.685133 0.3518 0.1918369
##
## $metodo
##        diff       lwr        upr     p adj
## M2-M1 -2.5 -4.004827 -0.9951734 0.0007912
## M3-M1 -5.0 -6.504827 -3.4951734 0.0000000
## M3-M2 -2.5 -4.004827 -0.9951734 0.0007912
```

```r
plot(prueba_tukey)
```





Observamos que no hay diferencias significativas entre el sexo, pero sí hay diferencias significativas entre todos los métodos, por lo que el método sí es relevante para nosotros # 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

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

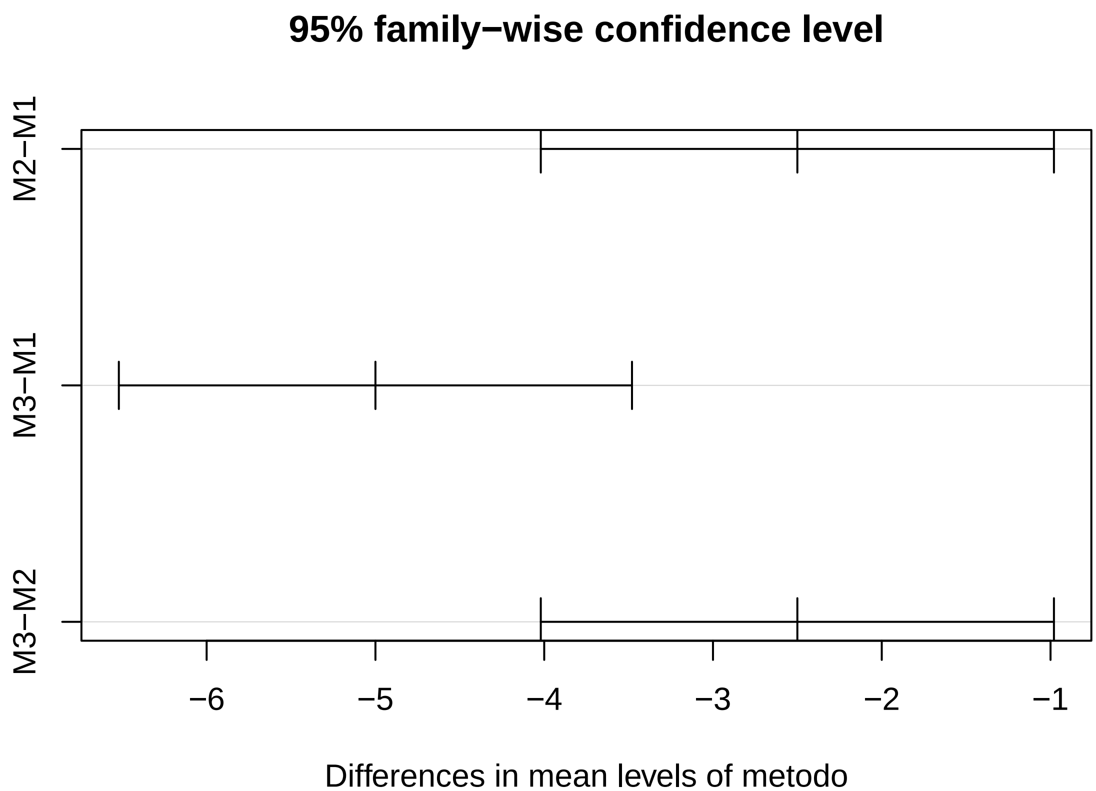
```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## metodo       2    150    75.0   32.57 1.55e-08 ***
## Residuals   33     76     2.3                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
modelo_anova <- aov(rendimiento ~ metodo)
prueba_tukey <- TukeyHSD(modelo_anova)
print(prueba_tukey)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = rendimiento ~ 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
```

```r
plot(prueba_tukey)
```



Se muestra lo que se había observado anteriormente con la prueba de Tukey y los boxplots, pues el método es relevante, al ser las diferencias significativas.

Comprueba la validez del modelo. Comprueba:

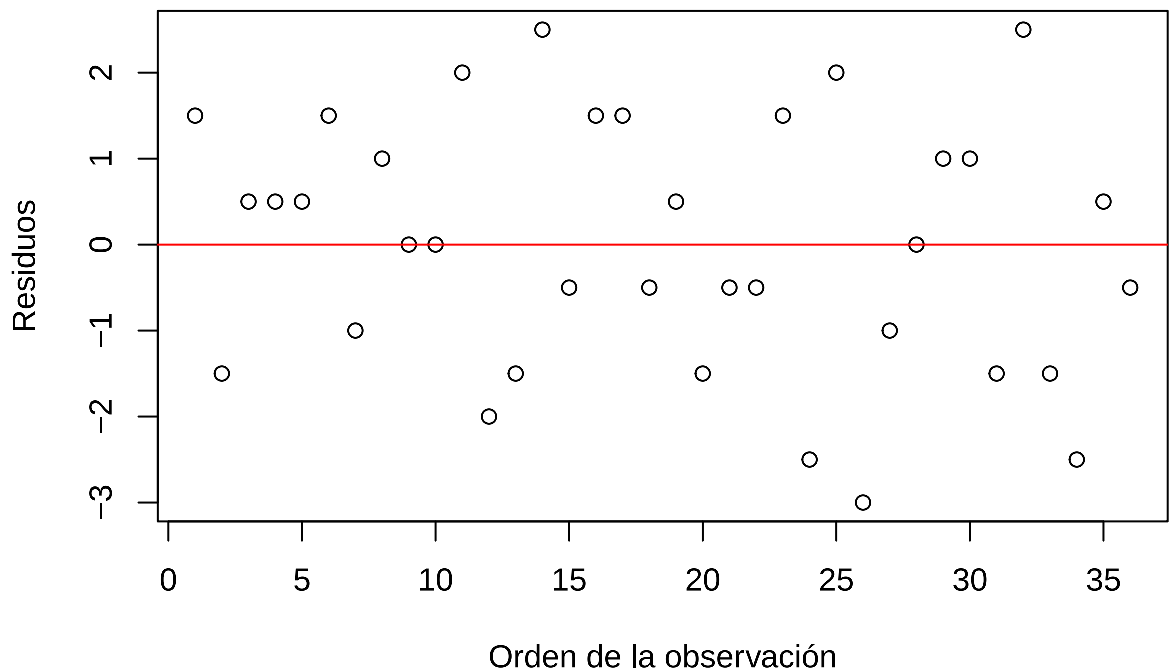
```
Normalidad
Homocedasticidad
Independencia
Relación lineal entre las variables (coeficiente de determinación).
```

```r
residuos=modelo_anova$residuals
shapiro.test(residuos)
```

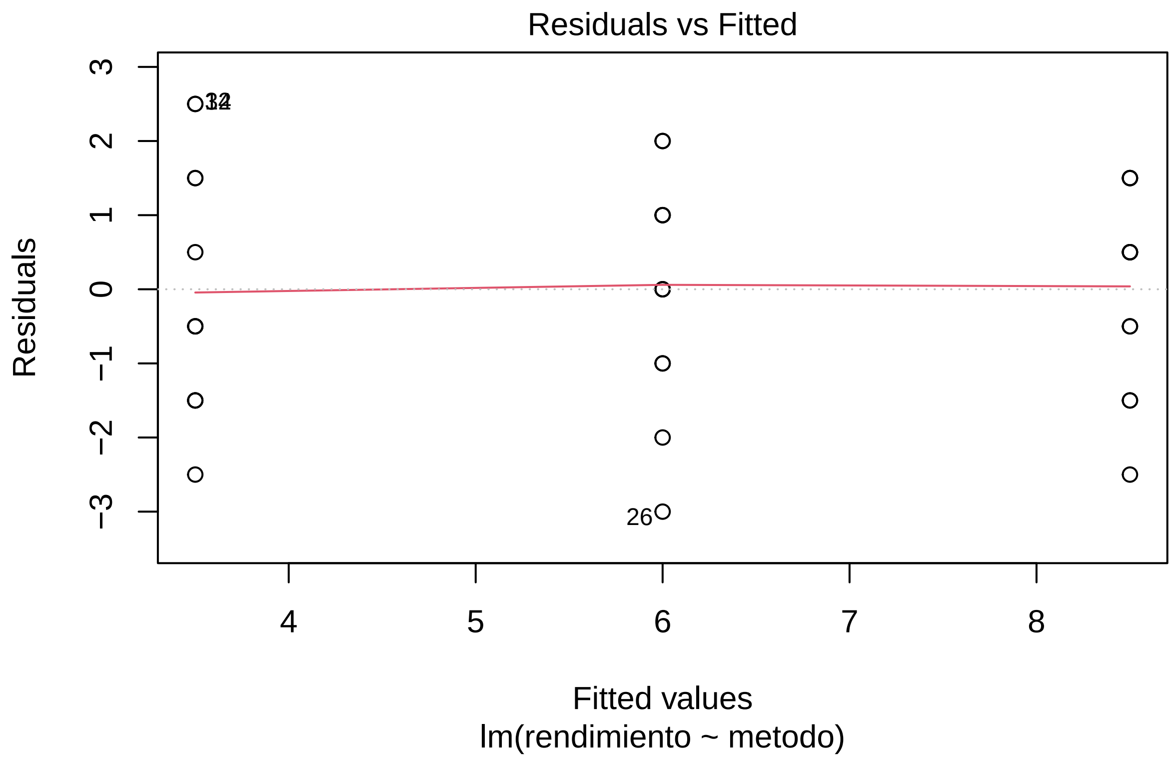
```
##
##  Shapiro-Wilk normality test
```

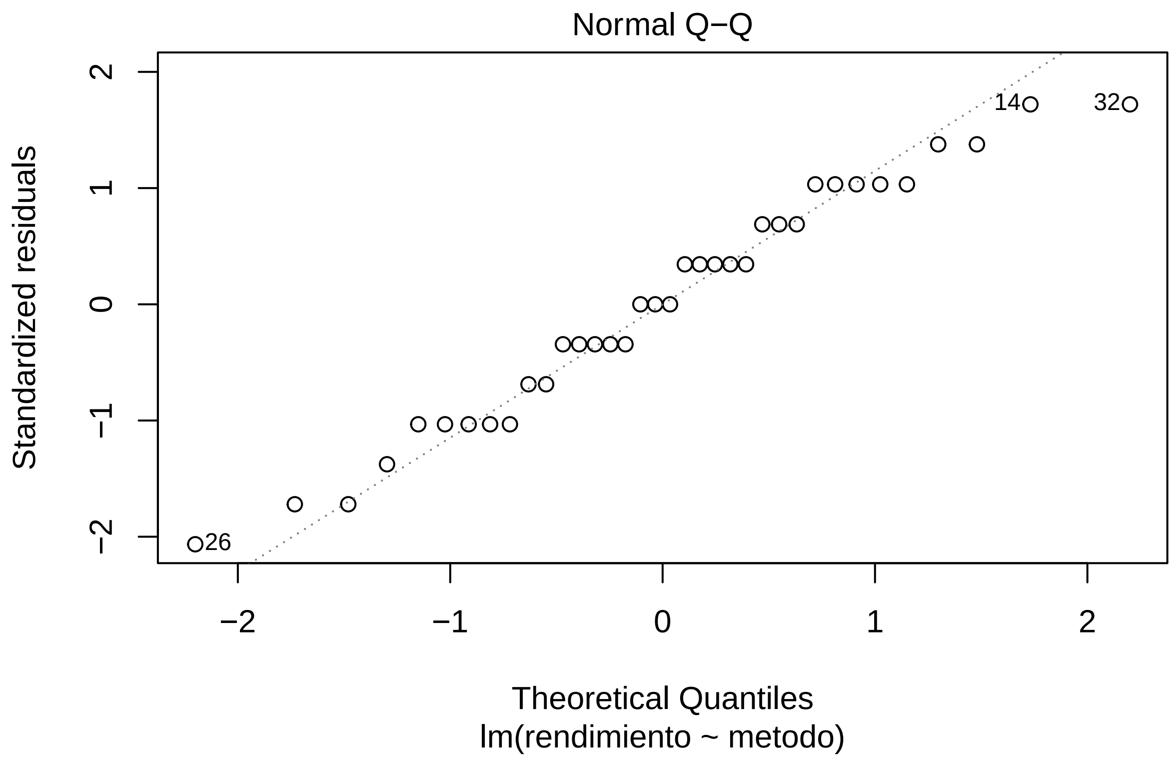
```
##
## data:  residuos
## W = 0.96734, p-value = 0.3573
```

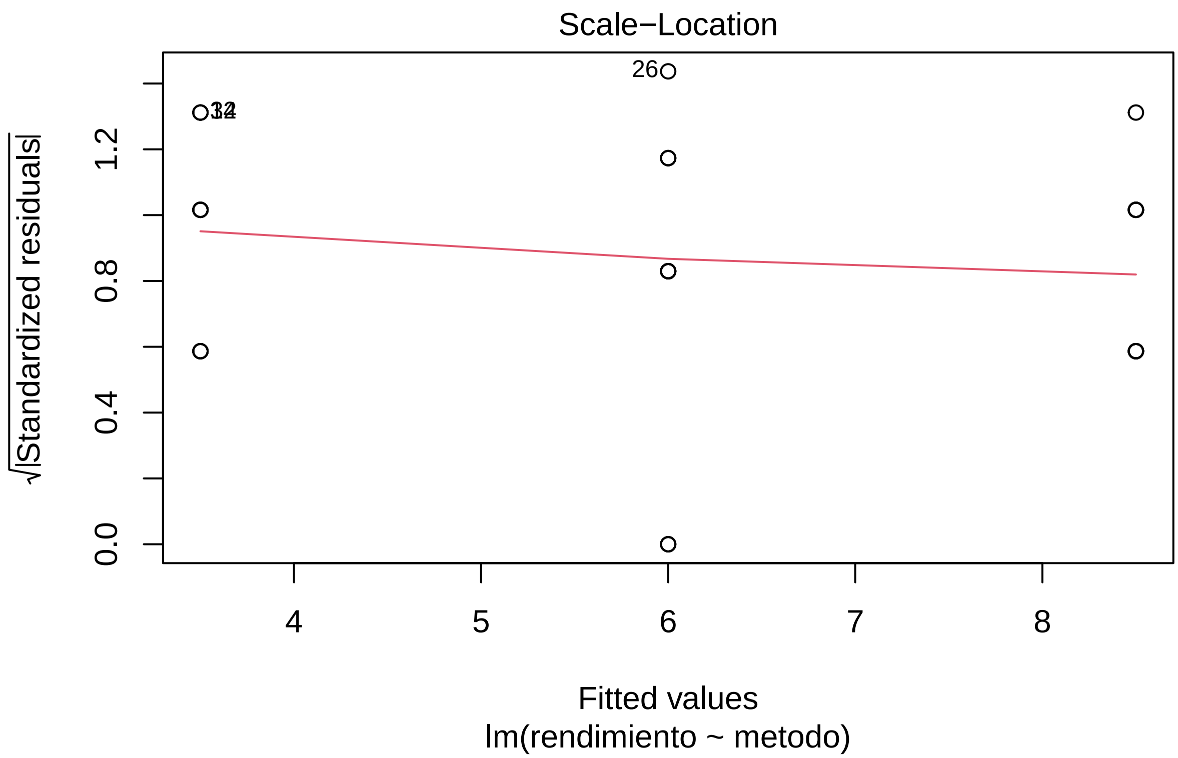
```
n = tapply(rendimiento, metodo, length)
plot(c(1:sum(n)),modelo_anova$residuals,xlab="Orden de la observación",ylab="Residuos")
abline(h=0,col="red")
```



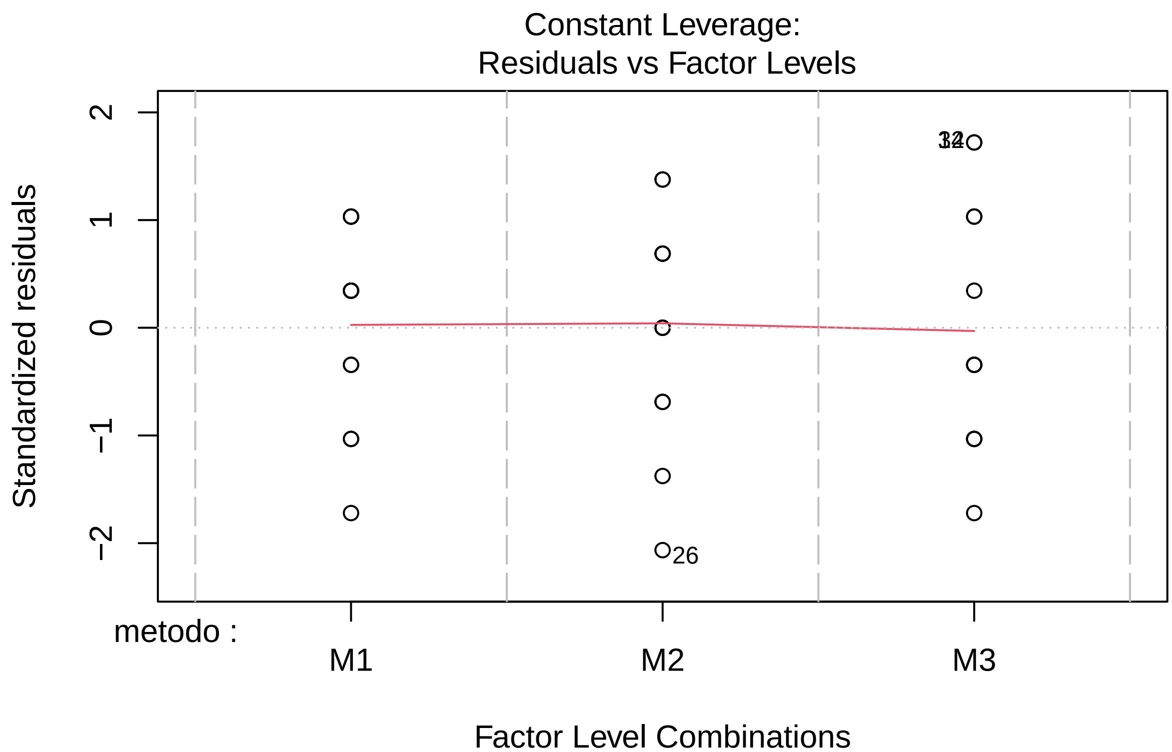

```
plot(lm(rendimiento~metodo))
```


Residuals vs Fitted
Residuals
Fitted values
lm(rendimiento ~ metodo)
3
2
1
0
−1
−2
−3
4
5
6
7
8
26


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(rendimiento ~ metodo)
26
14
32
−2
−1
0
1
2
−2
−1
0
1
2



Scale−Location
26
3
14
√|Standardized residuals|
0.0
0.4
0.8
1.2
4
5
6
7
8
Fitted values
lm(rendimiento ~ metodo)




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

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

Conclusión

El gráfico qq tiene una distribución con colas suaves pequeñas, pero tenemos un valor p de 0.3573 en nuestra prueba de normalidad, por lo que la hipótesis nula no se rechaza y la distribución es normal. Observamos que los errores, aunque estén alineados de cierta manera vertical, no están relacionados, pues no muestran un patrón, por lo que hay homocedasticidad. Observamos que los residuos son independientes, pues están totalmente datos por variabilidad sin seguimiento ni dependencia. Finalmente, tenemos un coeficiente de determinación de 0.6637, por lo que el 66% de la variabilidad por el rendimiento puede explicarse por el modelo lineal basado en el método. Es decir, hay una relación lineal moderadamente fuerte entre las variables, pero aún hay un 34% de la variabilidad que no es explicada por el modelo, lo cual quiere decir que hay otros factores (aparte del sexo) que no se están tomando en cuenta para realizar un modelo ajustado adecuado. Finalmente, vemos que las hipótesis iniciales se rechazan, pues

$$\exists \tau_i \neq 0, \exists \alpha_j \neq 0, \exists \tau_j \alpha_j \neq 0$$

Es decir hay al menos una media y un efecto que $\neq 0$

Problema 2

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.

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

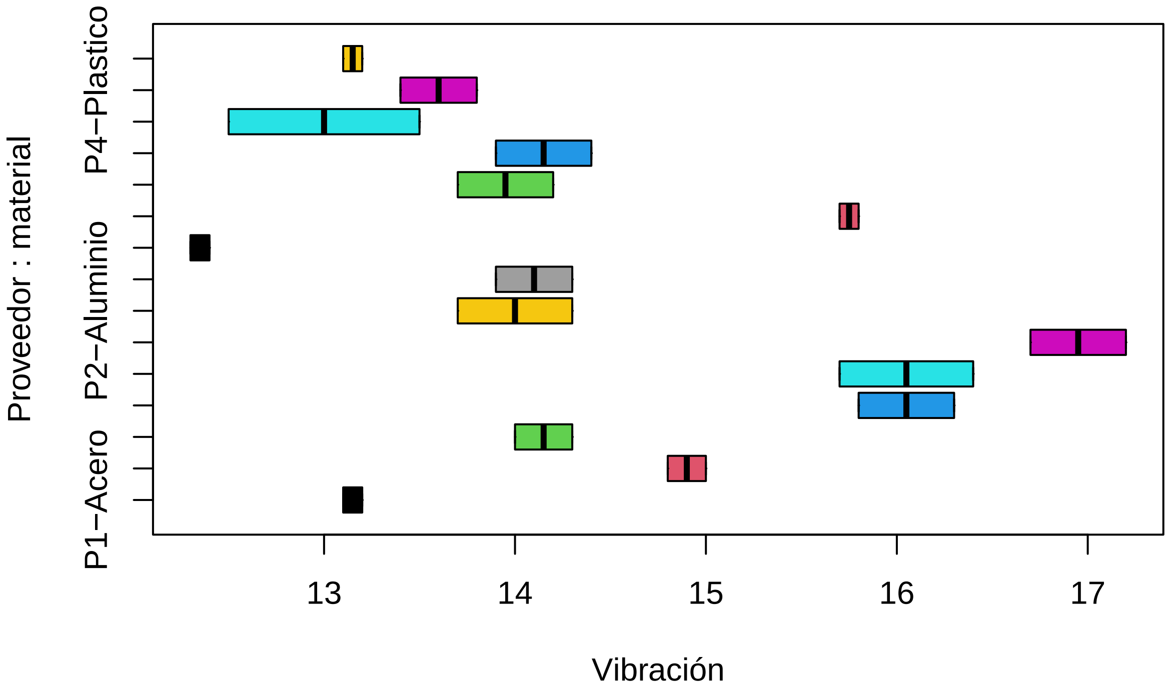
```
vibracion <- c(13.1, 13.2, 15.0, 14.8, 14.0, 14.3, 16.3, 15.8, 15.7, 16.4, 17.2, 16.7, 13.7, 14.3, 13.9
                 14.3, 12.4, 12.3, 15.7, 15.8, 13.7, 14.2, 14.4, 13.9, 13.5, 12.5, 13.4, 13.8, 13.2, 13
proveedor <- c(rep("P1", 2), rep("P2", 2), rep("P3", 2), rep("P4", 2), rep("P5", 2),
            rep("P1", 2), rep("P2", 2), rep("P3", 2), rep("P4", 2), rep("P5", 2),
            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)

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

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

```
print(mean(vibracion))
```

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

```
new_names <- c("P1-Acero", "P2-Acero", "P3-Acero", "P4-Acero", "P5-Acero",
               "P1-Aluminio", "P2-Aluminio", "P3-Aluminio", "P4-Aluminio", "P5-Aluminio",
               "P1-Plastico", "P2-Plastico", "P3-Plastico", "P4-Plastico", "P5-Plastico")
boxplot(vibracion ~ proveedor * material, col = c(1:33),
        horizontal = TRUE, xlab = "Vibración", ylab = 'Proveedor : material',
        names = new_names)
```



Observamos que la vibración varía conforme el proveedor y el material, pues se observa en la gráfica de bigotes que influyen sobre la vibración. Por otro lado, observamos que las diferencias de las medias de los materiales y los proveedores parecen no ser estadísticamente significativas, pero se necesitará un análisis más profundo para confirmar lo anterior.

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

Primera hipótesis

$H_0 : \tau_i = 0$

$H_1 : \exists \tau_i \neq 0$

Segunda hipótesis

$H_0 : \alpha_j = 0$

$H_1 : \exists \alpha_j \neq 0$

Tercera hipótesis

$H_0 : \tau_j \alpha_j = 0$

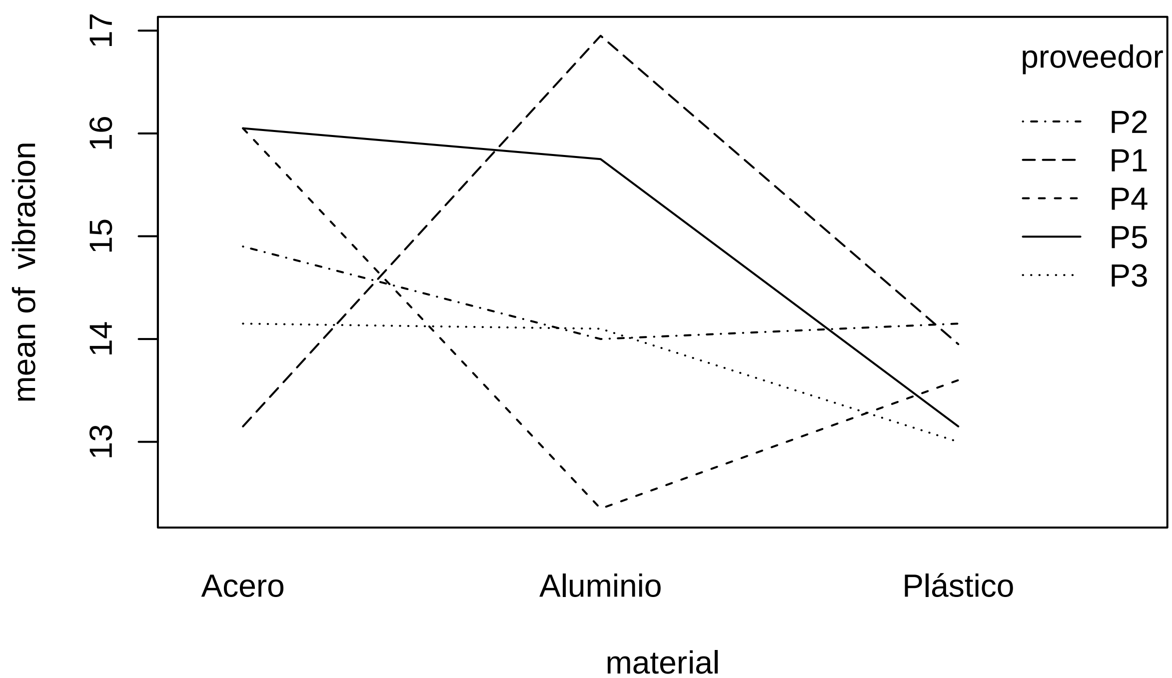
$H_1 : \exists \tau_j \alpha_j \neq 0$

Realiza el ANOVA para dos niveles con interacción:

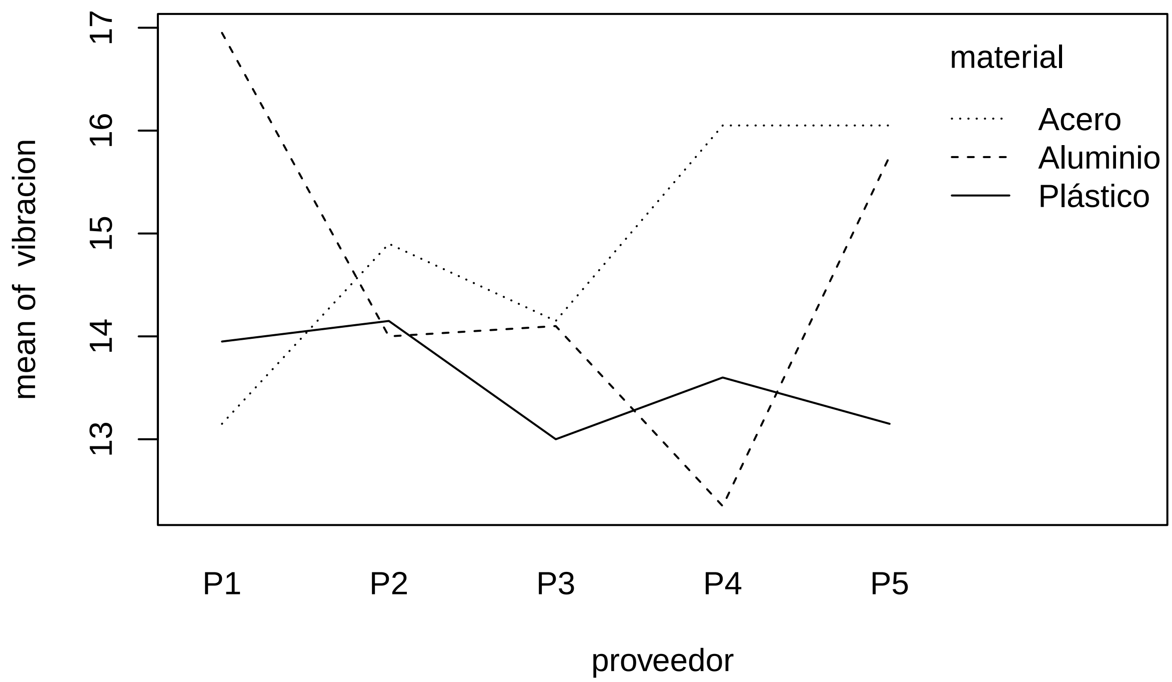
```r
A<-aov(vibracion~material*proveedor)
summary(A)
```

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## material            2   9.47   4.734   42.52 6.59e-07 ***
## proveedor           4   5.97   1.492   13.40 7.66e-05 ***
## material:proveedor  8  33.55   4.194   37.67 1.46e-08 ***
## Residuals          15   1.67   0.111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

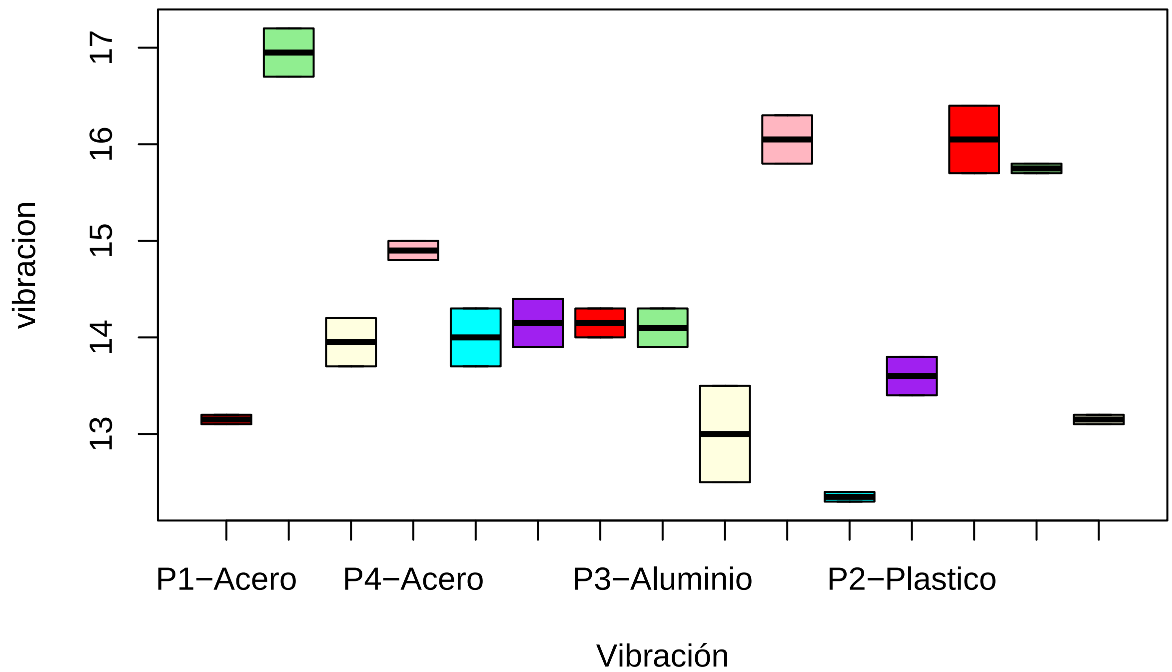
```r
interaction.plot(material, proveedor, vibracion)
```




```r
interaction.plot(proveedor, material, vibracion)
```




```
boxplot(vibracion ~ material * proveedor, col = c("red", "lightgreen", "lightyellow", "lightpink", "cyan
        horizontal = FALSE, xlab = "Vibración",
        names = new_names)
```

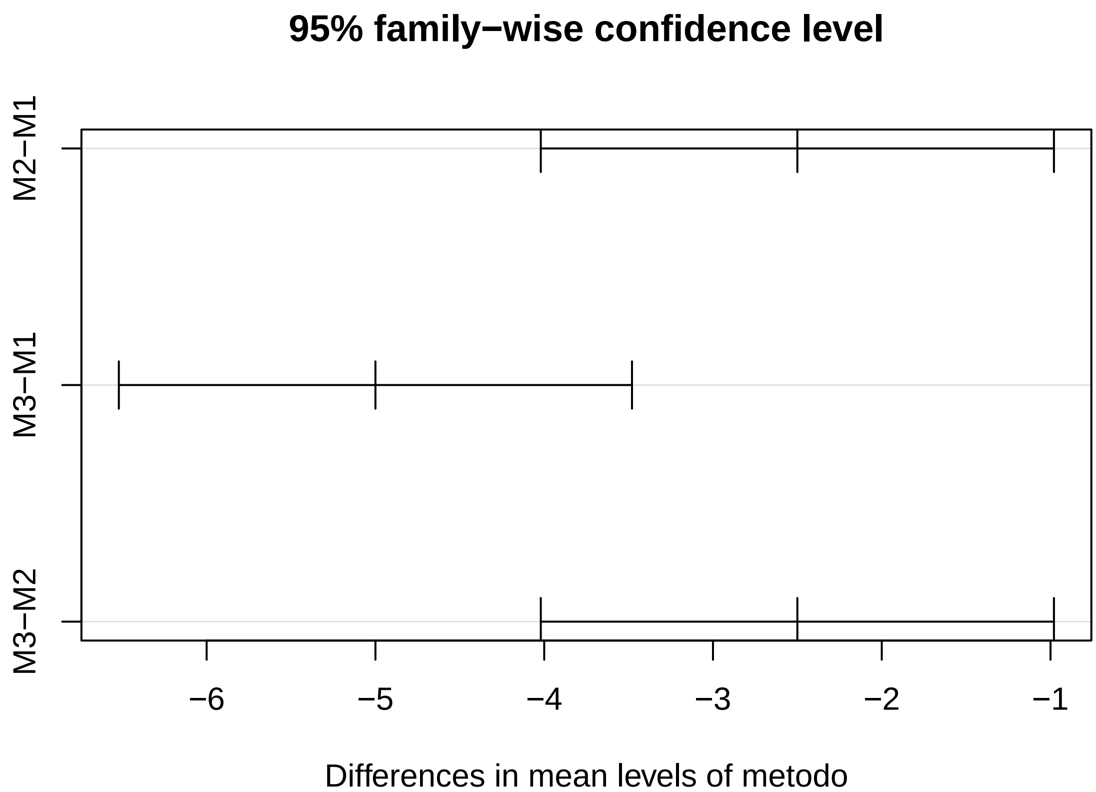


Parece ser que el proveedor 1 tiene la mayor vibración con el material aluminio, pues se observa el pico más alto en ese punto. Observamos en el summary que la suma cuadrada de la interacción entre el material y el proveedor es mucho mayor que los residuos, por lo que estos dos factores con interacción parecen ajustarse al modelo.

```
prueba_tukey <- TukeyHSD(modelo_anova)
print(prueba_tukey)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = rendimiento ~ 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(prueba_tukey)
```



Observamos en los tres gráficos que sí hay diferencias significativas entre los materiales y los proveedores cuando hay interacción entre los factores.

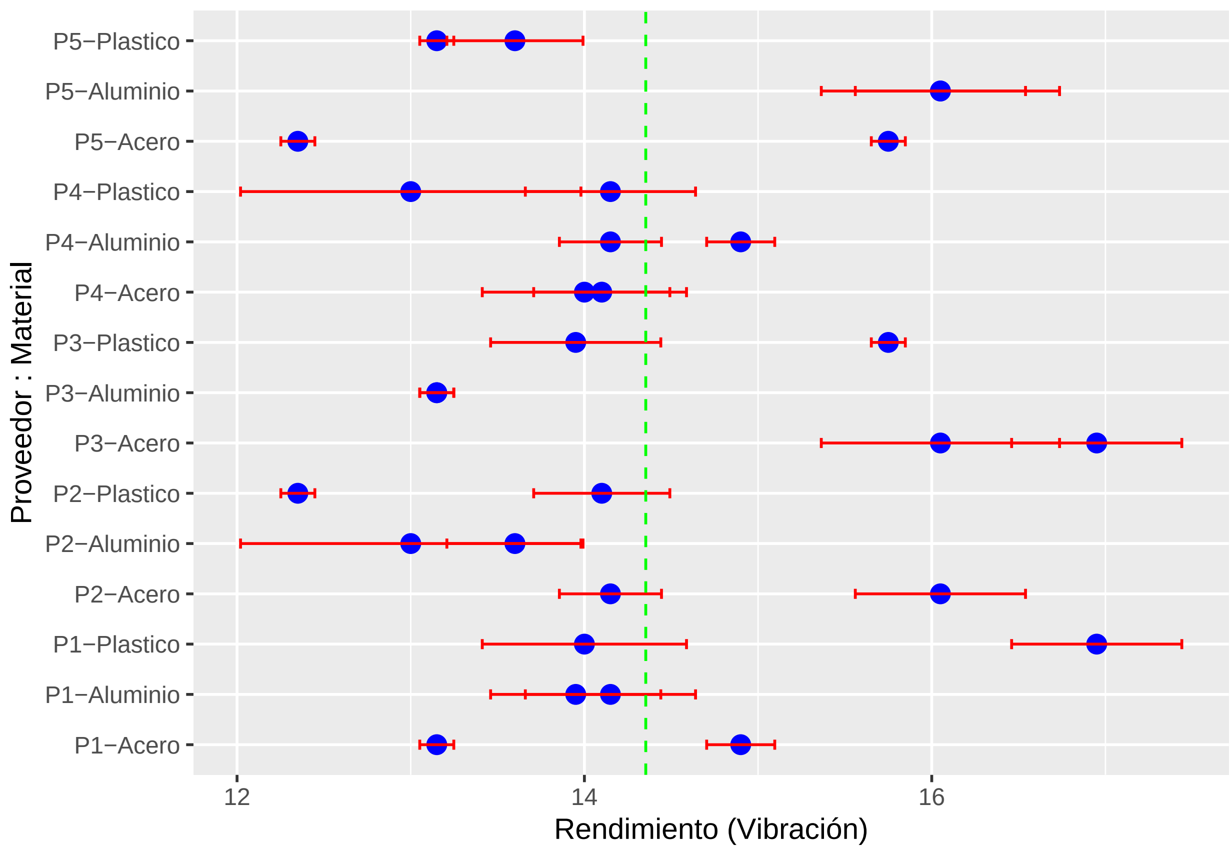
Realiza el ANOVA para dos niveles sin interacción.

Consulta el código de R en los apoyos de clase de "ANOVA". Haz los intervalos de confianza de rendimiento por sexo. Grafícalos Interpreta el resultado desde la perspectiva estadística y en el contexto del problema. scribe tus conclusiones parciales

```
means <- tapply(vibracion, list(proveedor, material), mean)
stderr <- tapply(vibracion, list(proveedor, material), function(x) sd(x) / sqrt(length(x)))
conf_intervals <- tapply(vibracion, list(proveedor, material), function(x) {
  m <- mean(x)
  se <- sd(x) / sqrt(length(x))
  c(m - 1.96 * se, m + 1.96 * se)
})
global_mean <- mean(vibracion)
print(conf_intervals)
```

```
##    Acero     Aluminio  Plástico
## P1 numeric,2 numeric,2 numeric,2
## P2 numeric,2 numeric,2 numeric,2
## P3 numeric,2 numeric,2 numeric,2
## P4 numeric,2 numeric,2 numeric,2
## P5 numeric,2 numeric,2 numeric,2
```

```r
data <- data.frame(
  ProveedorMaterial = rep(new_names, each = 2),
  Media = c(means),
  Inferior = unlist(lapply(conf_intervals, "[", 1)),
  Superior = unlist(lapply(conf_intervals, "[", 2))
)
ggplot(data, aes(x = Media, y = ProveedorMaterial)) +
  geom_point(size = 3, color = "blue") +
  geom_errorbar(aes(xmin = Inferior, xmax = Superior), width = 0.2, color = "red") +
  geom_vline(xintercept = global_mean, linetype = "dashed", color = "green") +
  labs(x = "Rendimiento (Vibración)", y = "Proveedor : Material") +
  theme(axis.text.y = element_text(angle = 0, hjust = 1))
```

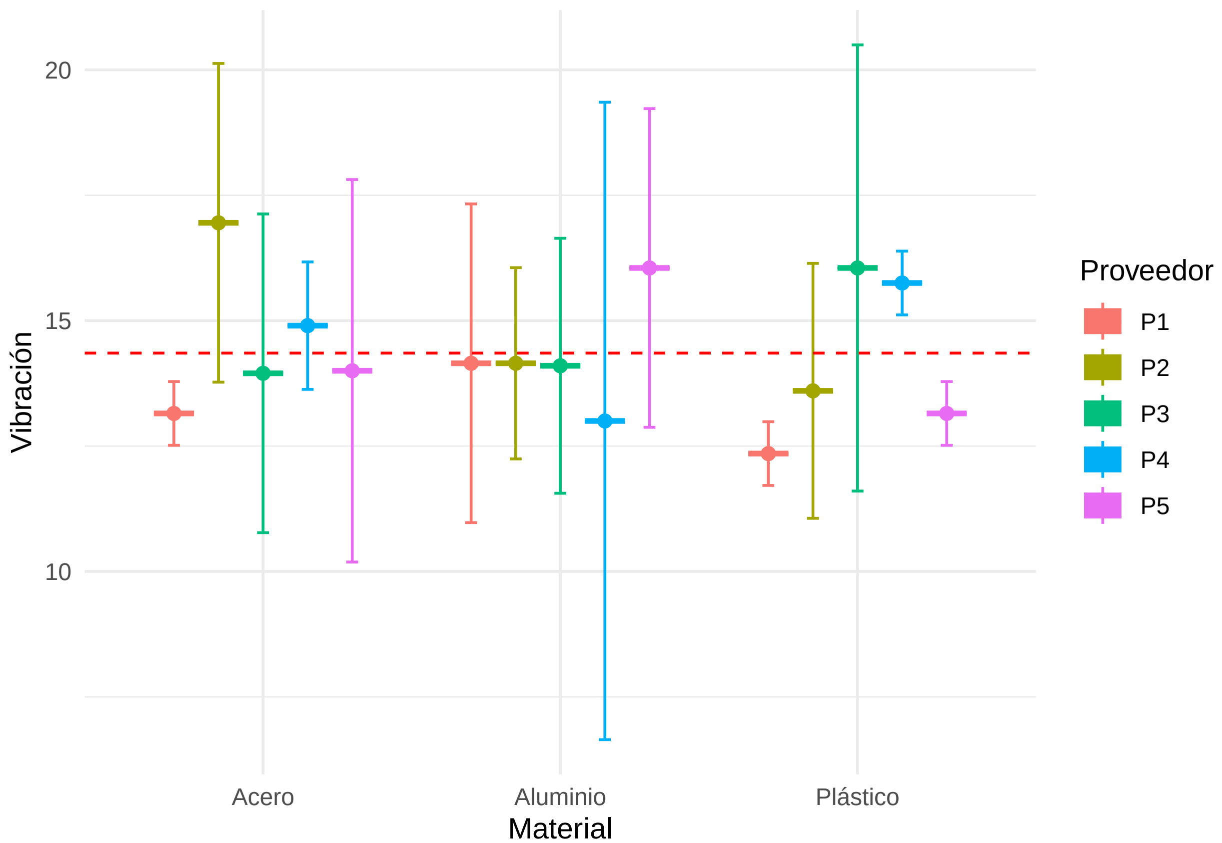


Observamos que los intervalos de confianza varían con respecto al proveedor y el material

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

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## material     2   9.47   4.734   3.092 0.0647 .
## proveedor    4   5.97   1.492   0.974 0.4407
## Residuals   23  35.22   1.531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
means <- tapply(vibracion, list(material, proveedor), mean)
sds <- tapply(vibracion, list(material, proveedor), sd)
n <- tapply(vibracion, list(material, proveedor), length)
sem <- sds / sqrt(n)
conf.low <- means - qt(0.975, n - 1) * sem
conf.high <- means + qt(0.975, n - 1) * sem
data <- data.frame(
    Material = rep(levels(material), each = 5),
    Proveedor = rep(levels(proveedor), times = length(levels(material))),
    Vibracion = c(means),
    Conf.low = c(conf.low),
    Conf.high = c(conf.high)
)
ggplot(data, aes(x = Material, y = Vibracion, fill = Proveedor, color = Proveedor)) +
    geom_boxplot(aes(ymin = Conf.low, ymax = Conf.high)) +
    geom_point(position = position_dodge(width = 0.75), size = 2) +
    geom_errorbar(aes(ymin = Conf.low, ymax = Conf.high), width = 0.2, position = position_dodge(width =
    geom_hline(yintercept = mean(vibracion), linetype = "dashed", color = "red") +
    labs(x = "Material", y = "Vibración") +
    theme_minimal()
```




```
print(means)
```

```
##              P1    P2    P3    P4    P5
## Acero     13.15 14.90 14.15 16.05 16.05
```

```
## Aluminio 16.95 14.00 14.10 12.35 15.75
## Plástico 13.95 14.15 13.00 13.60 13.15
```

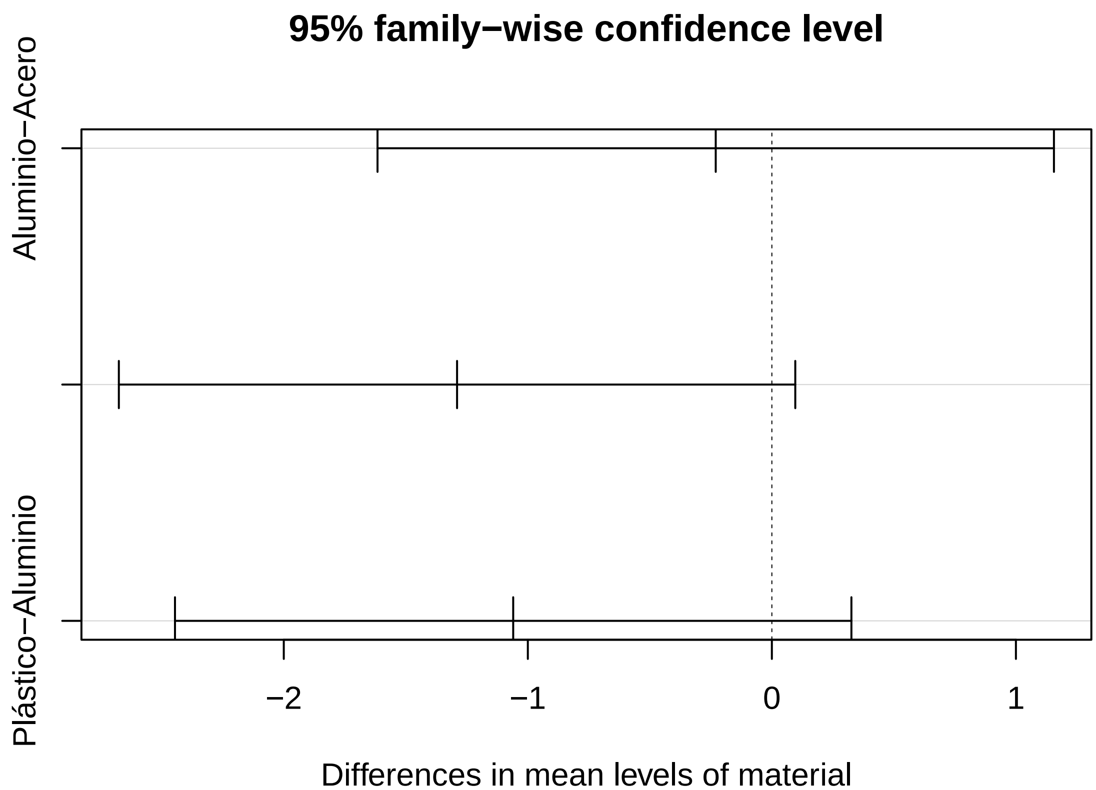
Se observa que las medias difieren por el proveedor, y el material, pero no parecen ser significativas. No sabemos qué tan estadísticamente significativa estas diferencias son, por lo que haremos una prueba de Tukey.

Observamos que ambos factores tienen una suma cuadrada muy baja y mucho menor a los residuos, por lo que parece que estos factores no son relevantes para explicar la variabilidad de la vibración.

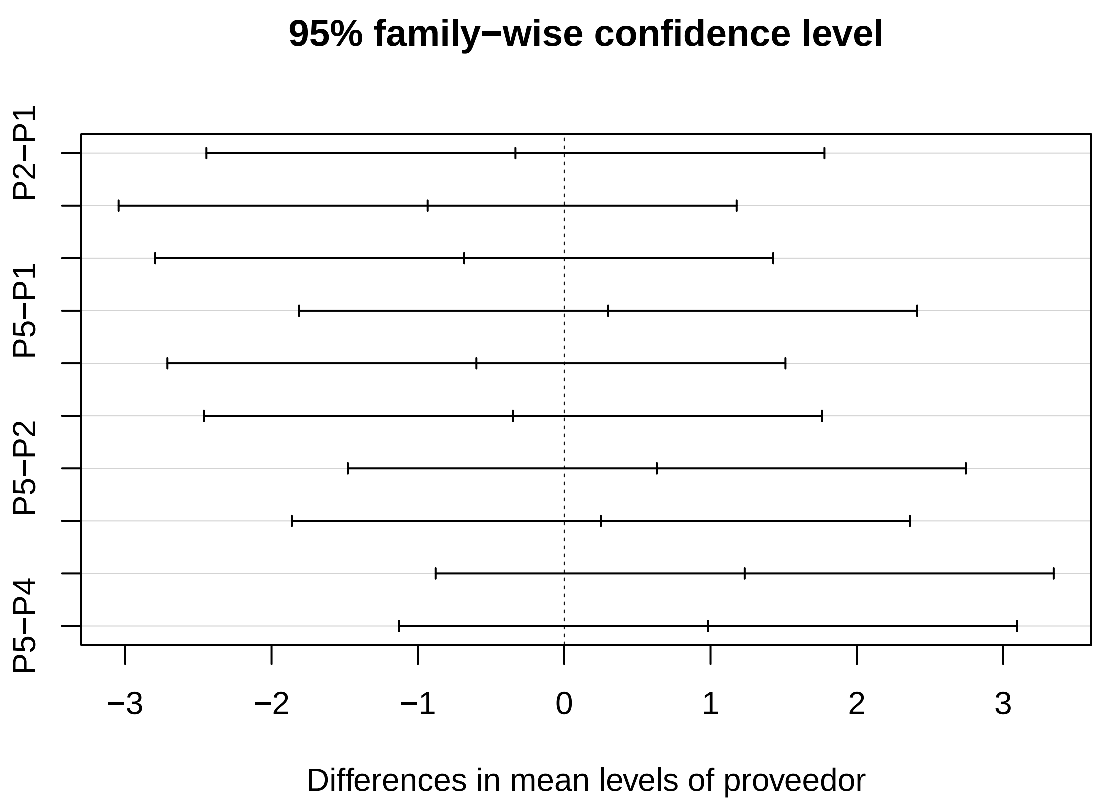
```r
modelo_anova <- aov(vibracion ~ material + proveedor)
prueba_tukey <- TukeyHSD(modelo_anova)
print(prueba_tukey)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = vibracion ~ material + proveedor)
##
## $material
##                    diff       lwr        upr     p adj
## Aluminio-Acero    -0.23 -1.615881 1.15588138 0.9095490
## Plástico-Acero    -1.29 -2.675881 0.09588138 0.0713361
## Plástico-Aluminio -1.06 -2.445881 0.32588138 0.1569892
##
## $proveedor
##             diff        lwr      upr     p adj
## P2-P1 -0.3333333 -2.4451968 1.778530 0.9896240
## P3-P1 -0.9333333 -3.0451968 1.178530 0.6900414
## P4-P1 -0.6833333 -2.7951968 1.428530 0.8714831
## P5-P1  0.3000000 -1.8118635 2.411863 0.9930404
## P3-P2 -0.6000000 -2.7118635 1.511863 0.9152968
## P4-P2 -0.3500000 -2.4618635 1.761863 0.9875366
## P5-P2  0.6333333 -1.4785301 2.745197 0.8989735
## P4-P3  0.2500000 -1.8618635 2.361863 0.9965452
## P5-P3  1.2333333 -0.8785301 3.345197 0.4383929
## P5-P4  0.9833333 -1.1285301 3.095197 0.6480965
```

```r
plot(prueba_tukey)
```



95% family−wise confidence level
Aluminio−Acero
Plástico−Aluminio
−2
−1
0
1
Differences in mean levels of material



Observamos que no hay diferencias significativas entre el tipo de material y el proveedor.

Realiza el ANOVA para un efecto principal

```r
C<-aov(vibracion~proveedor)
summary(C)
```

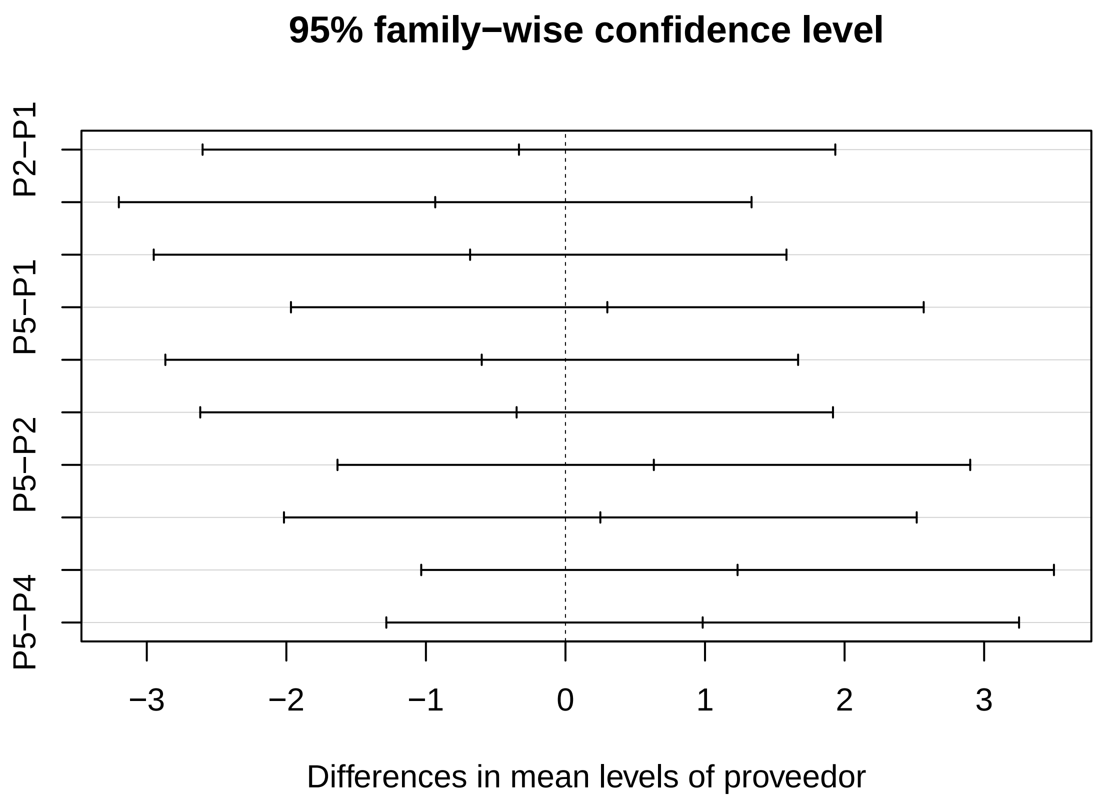
```
##             Df Sum Sq Mean Sq F value Pr(>F)
## proveedor    4   5.97   1.492   0.835  0.516
## Residuals   25  44.69   1.788
```

```r
modelo_anova <- C
prueba_tukey <- TukeyHSD(modelo_anova)
print(prueba_tukey)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = vibracion ~ proveedor)
##
## $proveedor
##             diff       lwr      upr     p adj
## P2-P1 -0.3333333 -2.600290 1.933623 0.9922986
```

```
## P3-P1 -0.9333333 -3.200290 1.333623 0.7462004
## P4-P1 -0.6833333 -2.950290 1.583623 0.8996512
## P5-P1  0.3000000 -1.966957 2.566957 0.9948514
## P3-P2 -0.6000000 -2.866957 1.666957 0.9348000
## P4-P2 -0.3500000 -2.616957 1.916957 0.9907329
## P5-P2  0.6333333 -1.633623 2.900290 0.9217998
## P4-P3  0.2500000 -2.016957 2.516957 0.9974553
## P5-P3  1.2333333 -1.033623 3.500290 0.5123908
## P5-P4  0.9833333 -1.283623 3.250290 0.7089858
```

```
plot(prueba_tukey)
```



Se muestra lo que se había observado anteriormente con la prueba de Tukey, pues el proveedor no es relevante, al no ser las diferencias significativas. De la misma manera, se observa que la suma cuadrada de los residuos es mayor al proveedor, por lo que el modelo no se ajusta correctamente.

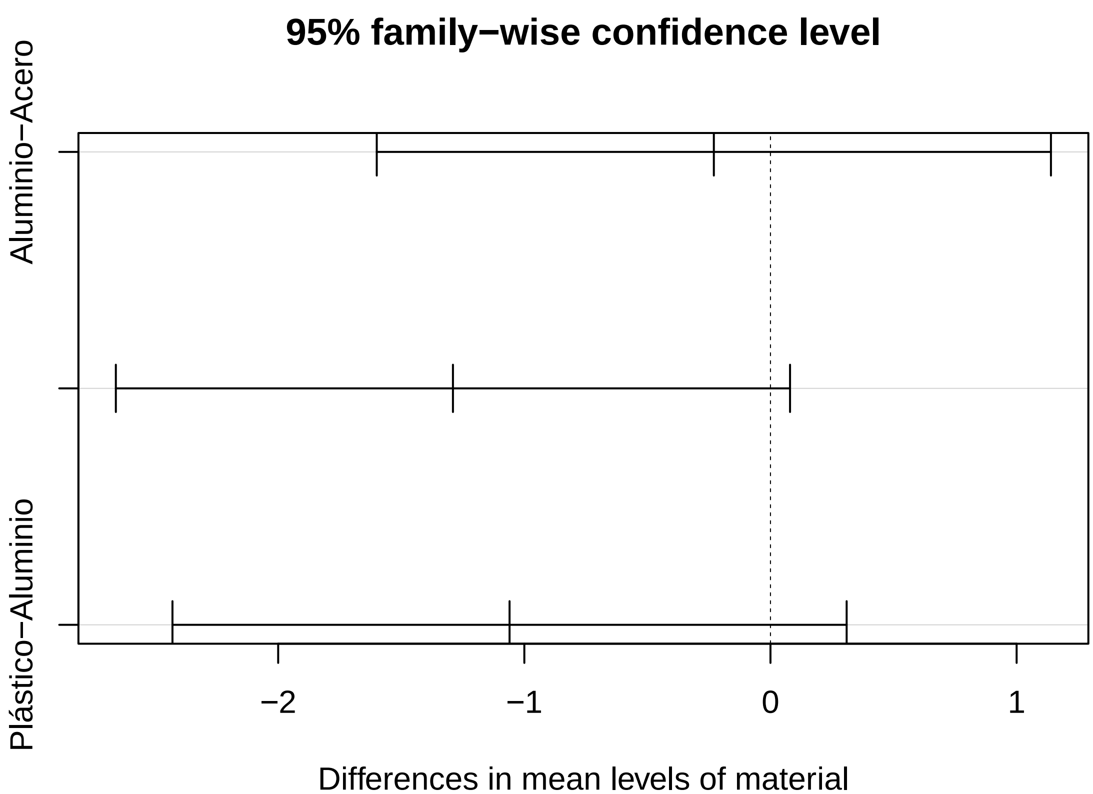
```
C<-aov(vibracion~material)
summary(C)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## material     2   9.47   4.734   3.104 0.0612 .
## Residuals   27  41.19   1.525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
modelo_anova <- C
prueba_tukey <- TukeyHSD(modelo_anova)
print(prueba_tukey)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = vibracion ~ material)
##
## $material
##                     diff       lwr        upr     p adj
## Aluminio-Acero     -0.23 -1.599486 1.13948607 0.9091688
## Plástico-Acero     -1.29 -2.659486 0.07948607 0.0677680
## Plástico-Aluminio -1.06 -2.429486 0.30948607 0.1526917
```

```r
plot(prueba_tukey)
```



Se vuelve a observar lo mismo que se había observado anteriormente con la prueba de Tukey, al material no ser relevante, por lo que no hay diferencias significativas. De la misma manera, se observa que la suma cuadrada de los residuos es mayor al material, lo cual no es bueno.

Por lo anterior, utilizaremos el modelo con interacción entre el proveedor y el material, siempre y cuando cumpla con los supuestos:

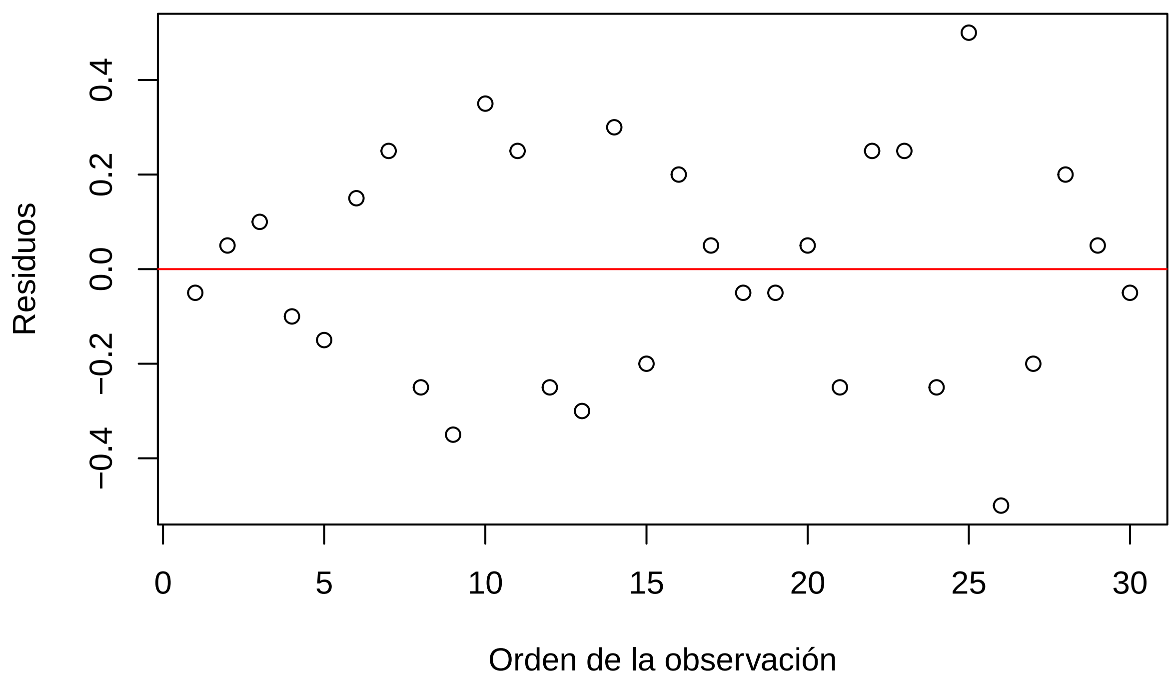
- Normalidad
- Homocedasticidad

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

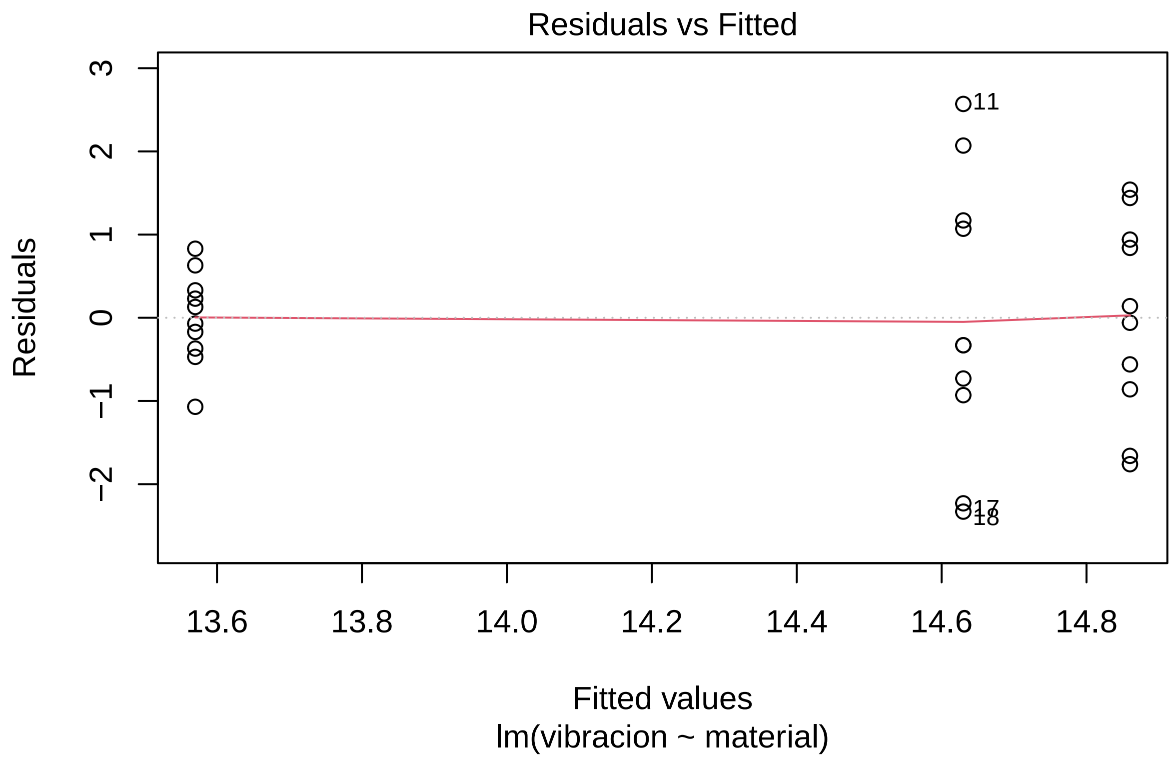
```
library(nortest)
shapiro.test(modelo_anova$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  modelo_anova$residuals
## W = 0.98767, p-value = 0.9736
```

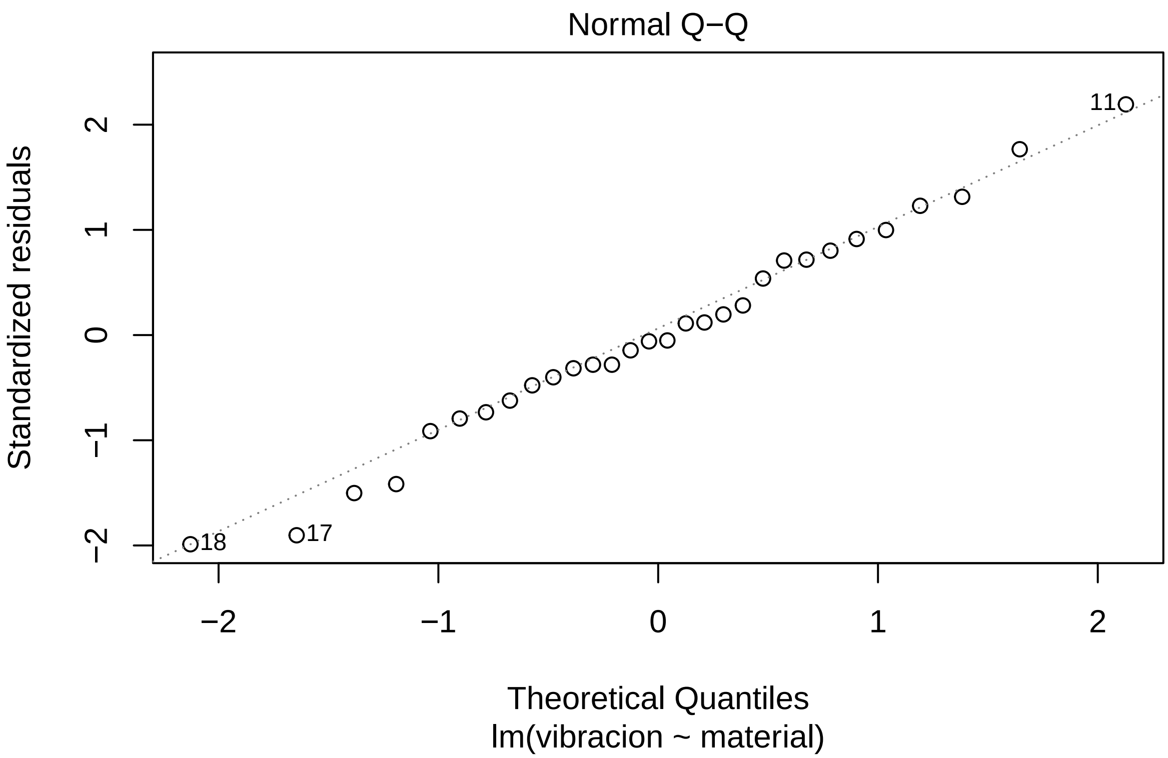
```
modelo_anova <- aov(vibracion ~ material * proveedor)
residuos=modelo_anova$residuals
n = tapply(vibracion, material, length)
plot(c(1:sum(n)),modelo_anova$residuals,xlab="Orden de la observación",ylab="Residuos")
abline(h=0,col="red")
```

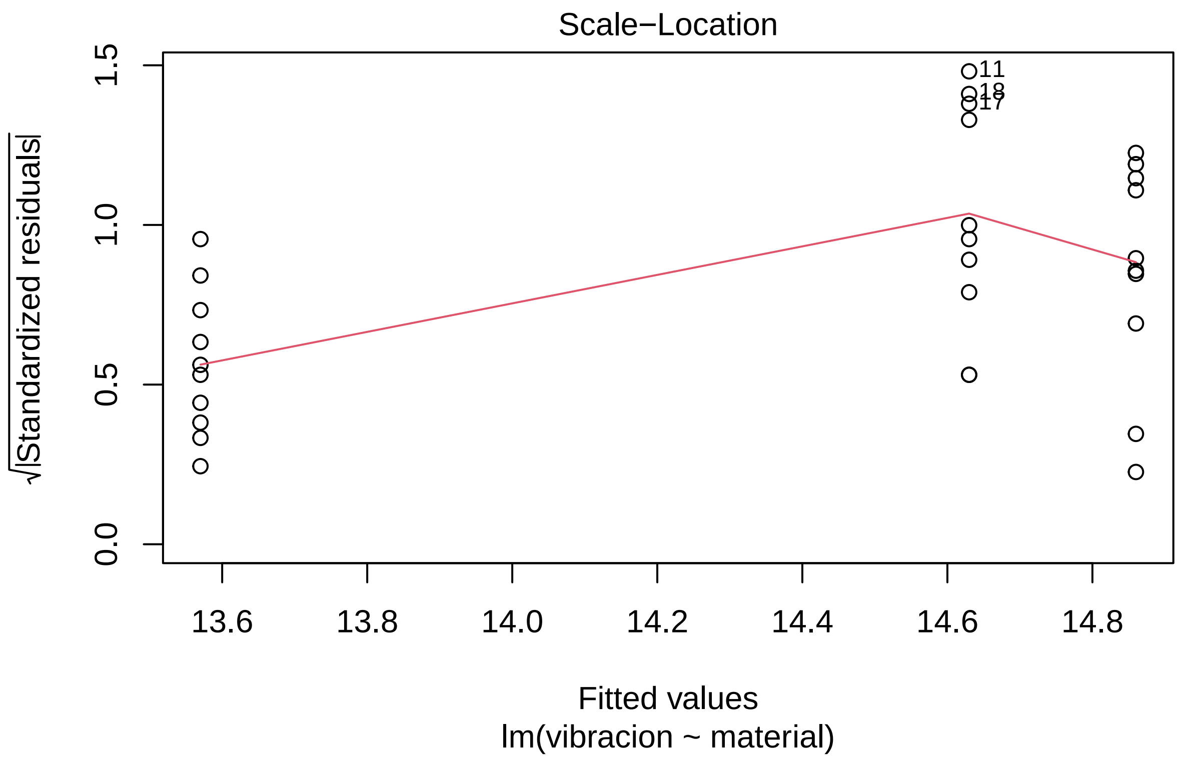


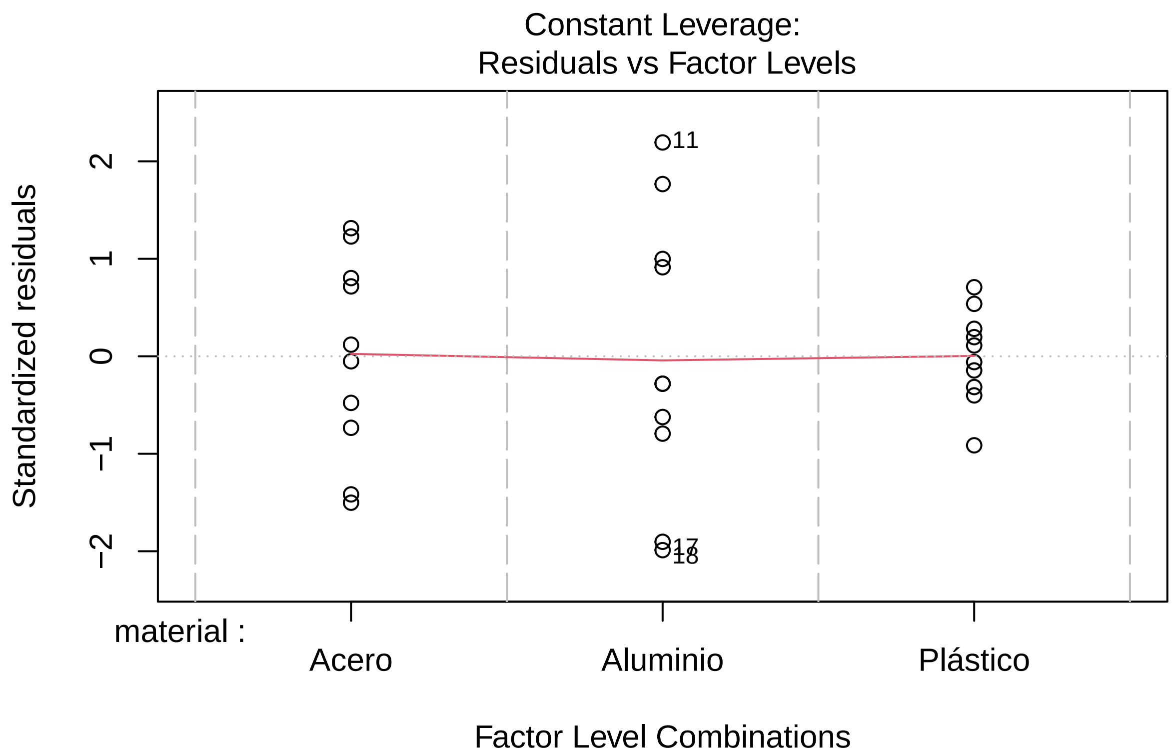

```
plot(lm(vibracion~material))
```



Residuals vs Fitted
Residuals
3
2
1
0
−1
−2
11
17
18
13.6
13.8
14.0
14.2
14.4
14.6
14.8
Fitted values
lm(vibracion ~ material)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(vibracion ~ material)
−2
−1
0
1
2
18
17
11


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
13.6
13.8
14.0
14.2
14.4
14.6
14.8
11
18
17
Fitted values
lm(vibracion ~ material)




```
CD= 1 -sum(modelo_anova$residuals^2)/sum((vibracion - mean(vibracion))^2)
CD
```

```
## [1] 0.9670317
```

Conclusión

El gráfico qq tiene una buena distribución, pues se observa que el modelo se ajusta adecuadamente a los datos. También tenemos un valor p de 0.72 en nuestra prueba de normalidad, por lo que la hipótesis nula no se rechaza y la distribución es normal. Observamos que los errores no están relacionados, pues no muestran un patrón, por lo que hay homocedasticidad. Observamos que los residuos son independientes, pues están totalmente datos por variabilidad sin seguimiento ni dependencia. Finalmente, tenemos un coeficiente de determinación de 0.9670, por lo que casi el 97% de la variabilidad en la vibración puede explicarse por el modelo lineal basado en la interacción entre el proveedor y el material. Es decir, hay una relación lineal muy fuerte entre las variables, con solo un 3% de variabilidad que no es explicada por el modelo, lo cual quiere decir que es muy bueno el modelo y se ajusta adecuadamente.

Finalmente, vemos que las hipótesis iniciales no se rechazan, pues

$$\tau_i = 0, \alpha_j = 0, \tau_j \alpha_j = 0$$

Es decir, las diferencias entre las medias y efectos no son estadísticamente significativas.