

# Anova

Jacobo Hirsch Rodriguez

2024-08-27

## #Problema 1

En un instituto se han matriculado 36 estudiantes. Se desea explicar el rendimiento de ciencias naturales en función de dos variables: género y metodología de enseñanza. La de enseñanza se analiza en tres niveles: explicación oral y realización del experimento (1er nivel) explicación oral e imágenes (2º nivel) y explicación oral (tercer nivel). En los alumnos matriculados había el mismo número de chicos que de chicas, por lo que formamos dos grupos de 18 sujetos; en cada uno de ellos, el mismo profesor aplicará a grupos aleatorios de 6 estudiantes las 3 metodologías de estudio. A fin de curso los alumnos son sometidos a la misma prueba de rendimiento. Los resultados son los siguientes:

```
calificacion=c(10,7,9,9,9,10,5,7,6,6,8,4,2,6,3,5
,5,3,9,7,8,8,10,6,8,3,5,6,7,7,2,6,2,1,4,3)
```

*#el comando*

```
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))
```

*#factor te devuelve las etiquetas, es decir que encuentra cada valor unico y lo convierte en un nuevo el  
#estas indicando que son factores para el indicador del analisis de varianza*

```
metodo = factor(metodo)
```

```
sexo = factor(sexo)
```

*#cheamos que si esten todas las calificaciones*  

```
print(length(calificacion))
```

```
## [1] 36
```

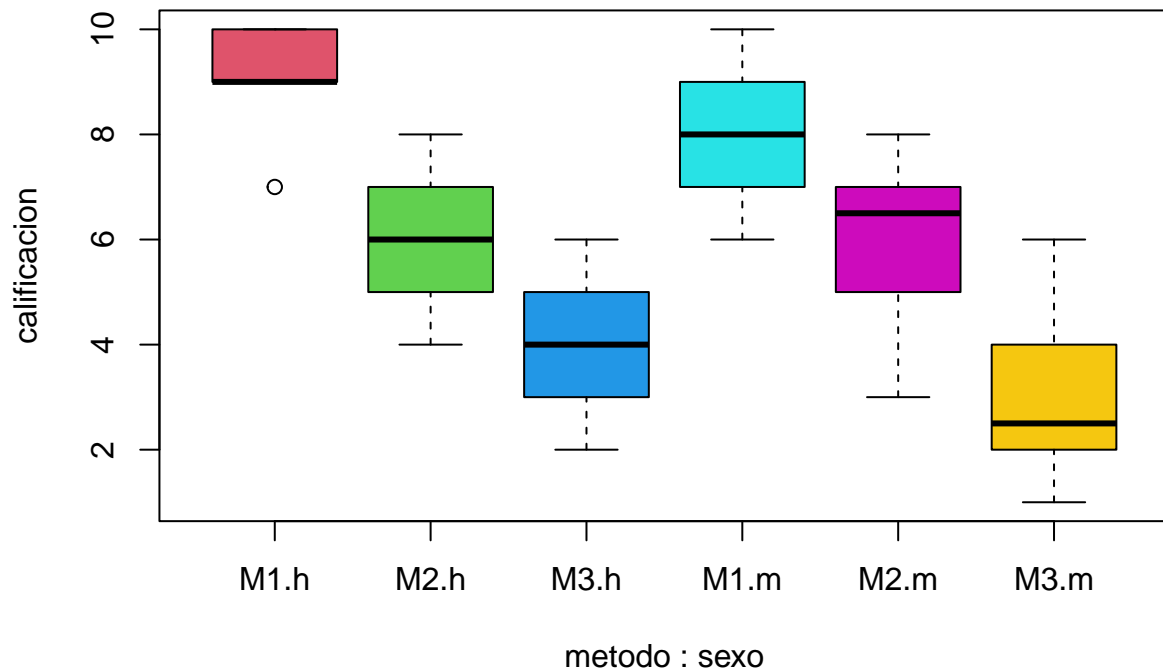
*#boxplot (analisis)*

vamos a hacer un boxplot de las variables

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

*# en R se utiliza para crear un data frame, que es una estructura de datos fundamental en R. Un data fr  
#de cálculo en Excel, donde los datos se organizan en filas y columnas.*

```
boxplot(calificacion~metodo:sexo, datos, col= 2:8)
```



Se puede observar que para hombres tanto como para mujeres el método 1 se comporta de la misma forma para los hombres tanto para las mujeres

#Hipotesis

primeera hipotesis

$h_0: \tau_i = 0$   $h_1$ : algun  $\tau_i$  es distinto de 0

segunda hipotesis

$h_0: \alpha_i = 0$   $h_1$ : algun  $\alpha_i$  es distinto de 0

tercera hipotesis

$h_0: \tau_i \alpha_i = 0$   $h_1$ : algun  $\tau_i \alpha_i$  es distinto de 0

## PUNTO 3

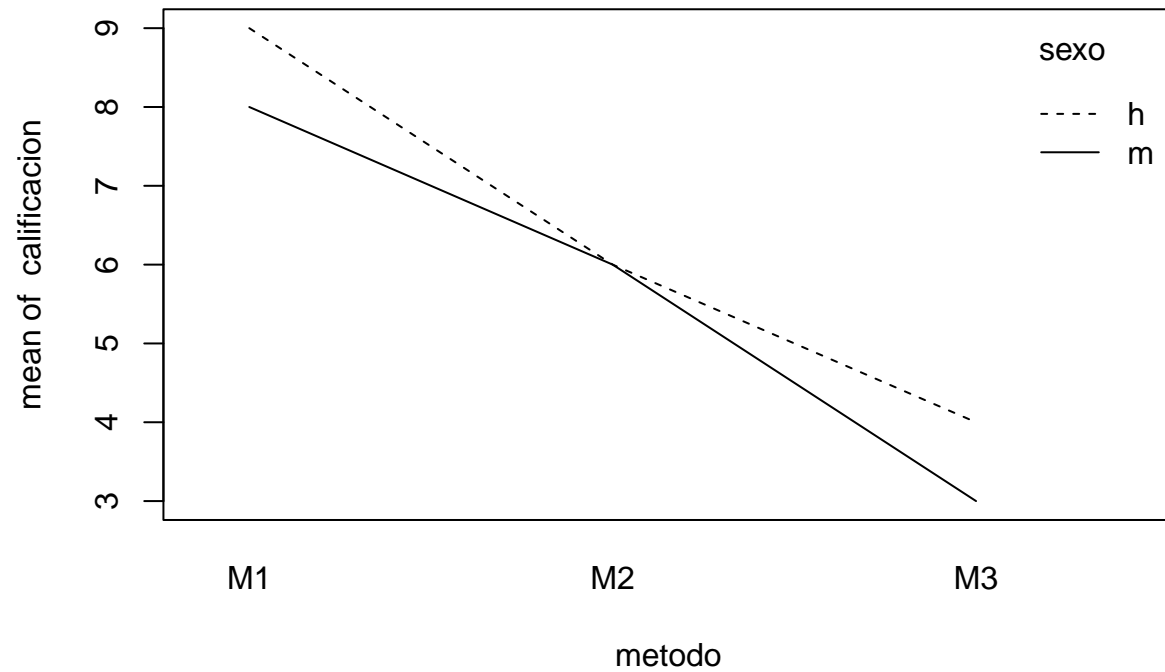
#Anova para dos vias de interaccion

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

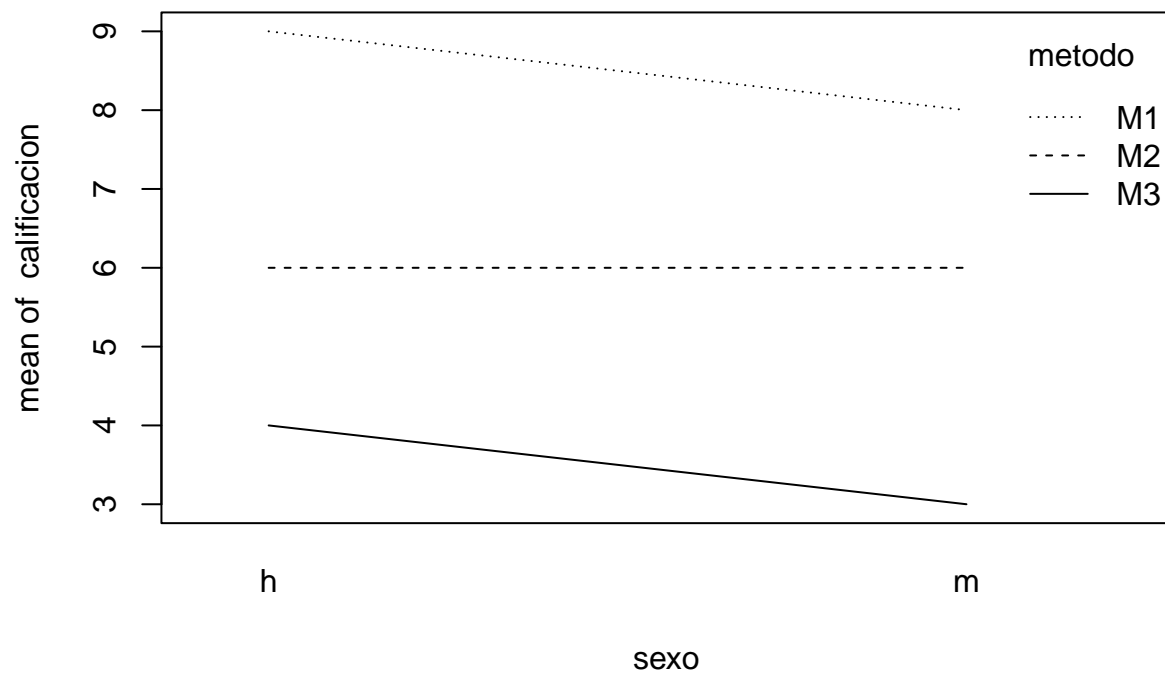
##		Df	Sum Sq	Mean Sq	F value	Pr(>F)
##	metodo	2	150	75.00	32.143	3.47e-08 ***
##	sexo	1	4	4.00	1.714	0.200
##	metodo:sexo	2	2	1.00	0.429	0.655

```
## Residuals    30      70    2.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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



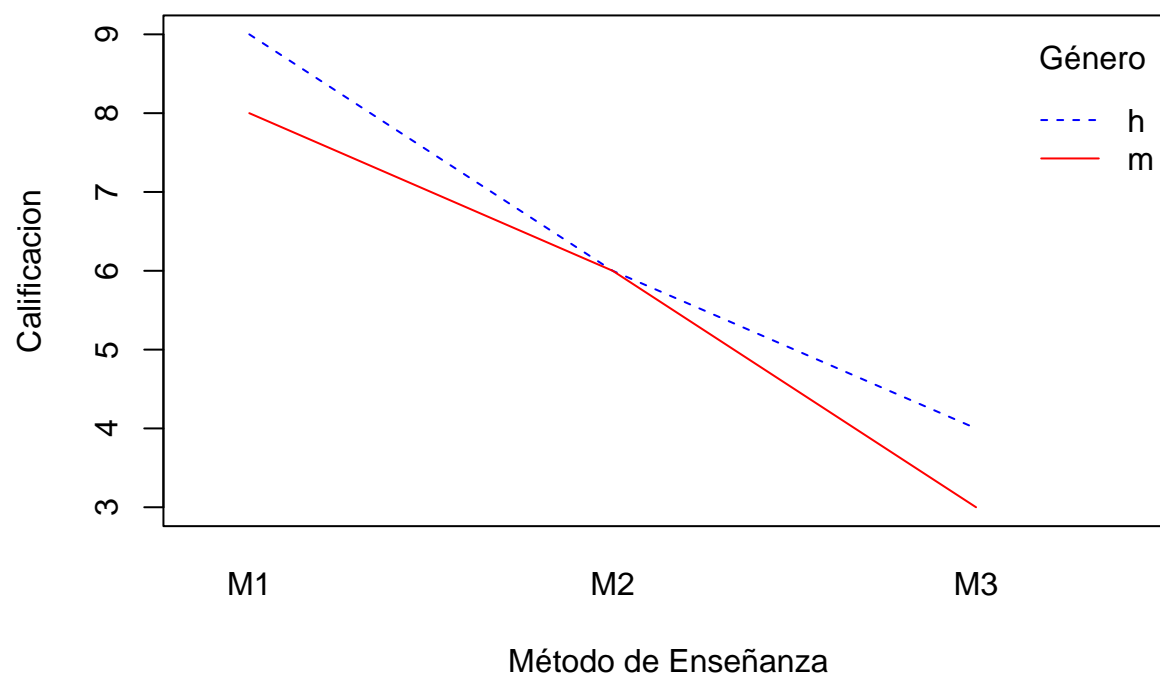
```
interaction.plot(sexo, metodo, calificacion)
```



#grafica de interaccion

```
interaction.plot(x.factor =metodo, trace.factor = sexo,
  response = calificacion,
  col = c("blue", "red"),
  main = "Gráfica de Interacción entre Método y sexo",
  xlab = "Método de Enseñanza",
  ylab = "Calificacion",
  trace.label = "Género")
```

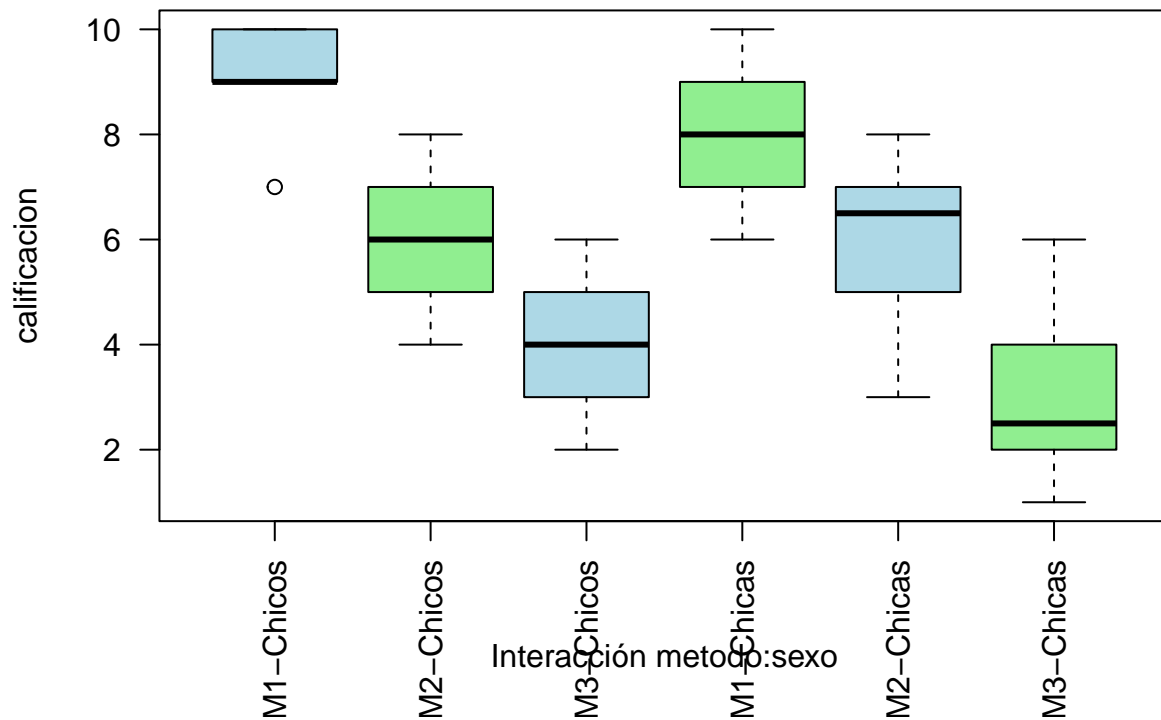
## Gráfica de Interacción entre Método y sexo



#boxplot

```
# Crear el boxplot con interacción
boxplot(calificacion ~ metodo:sexo,
  data = datos,
  col = c("lightblue", "lightgreen"),
  main = "Rendimiento por Interacción entre Método de Enseñanza y sexo",
  xlab = "Interacción metodo:sexo",
  ylab = "calificacion",
  names = c("M1-Chicos", "M2-Chicos", "M3-Chicos",
    "M1-Chicas", "M2-Chicas", "M3-Chicas"),
  las = 2)
```

## Rendimiento por Interacción entre Método de Enseñanza y sexo



#análisis de Anova para dos vías de interacción entre las 3 variables que se analizaron (la de las hipótesis) únicamente fue significativa la del método, con un valor extremadamente pequeño de 3.47e-08

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

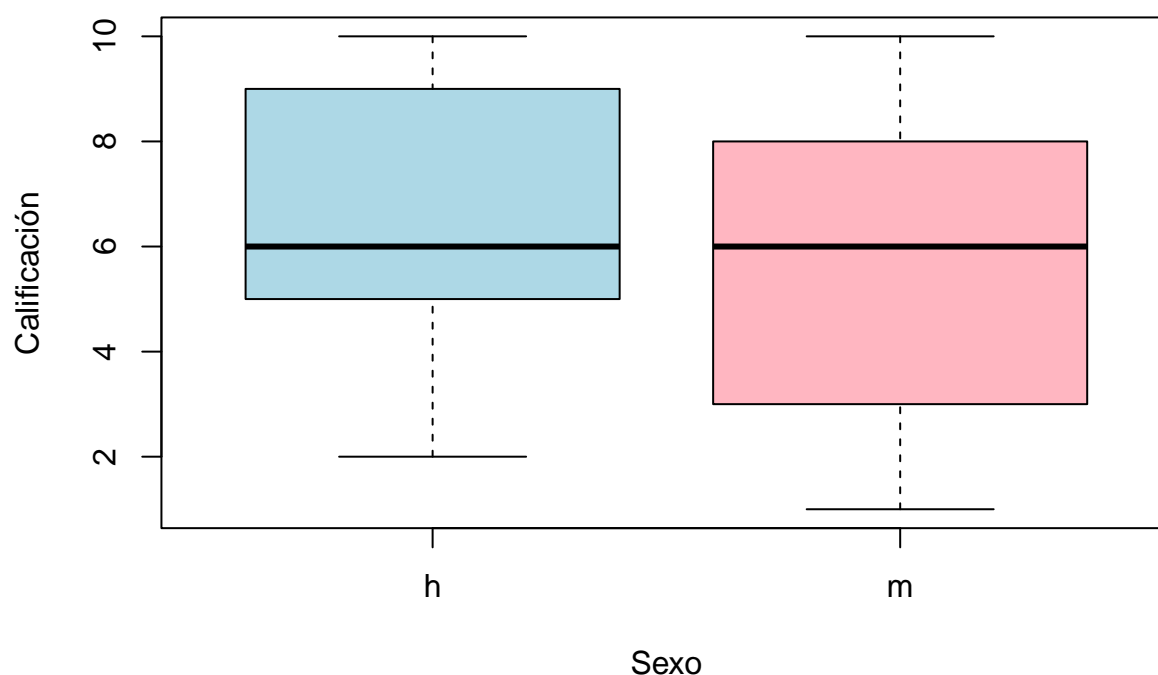
```
anova_sin_interaccion <- aov(calificacion ~ sexo + metodo, data = datos)
summary(anova_sin_interaccion)
```

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

Haz el boxplot de rendimiento por sexo

```
# Boxplot de rendimiento por sexo
boxplot(calificacion ~ sexo, data = datos,
        main = "Boxplot del Rendimiento por Sexo",
        xlab = "Sexo", ylab = "Calificación",
        col = c("lightblue", "lightpink"))
```

## Boxplot del Rendimiento por Sexo



Calcula la media para el rendimiento por sexo y método. para ello vamos a utilizar la función `aggregate` que en R se utiliza para aplicar funciones a subconjuntos de datos agrupados por una o más variables. devolviendo un dataframe con la media de la variable calificación con los pares correspondientes de los subconjuntos

```
media_sexo_metodo <- aggregate(calificacion ~ sexo + metodo, data = datos, mean)
print(media_sexo_metodo)
```

```
##  sexo metodo calificacion
## 1    h     M1            9
## 2    m     M1            8
## 3    h     M2            6
## 4    m     M2            6
## 5    h     M3            4
## 6    m     M3            3
```

```
media_h <- mean(datos$calificacion[datos$sexo == "h"])
media_m <- mean(datos$calificacion[datos$sexo == "m"])
```

Haz los intervalos de confianza de rendimiento por sexo y Gráficalos.vamos a utiliza `t.test` para el calculo de los intervalos de confianza

```
# Calcular intervalos de confianza para chicos
ic_h <- t.test(datos$calificacion[datos$sexo == "h"])$conf.int

# Calcular intervalos de confianza para chicas
```

```
ic_m <- t.test(datos$calificacion[datos$sexo == "m"])$conf.int
```

```
# Mostrar intervalos de confianza
```

```
cat("Intervalo de confianza para chicos: ", ic_h, "\n")
```

```
## Intervalo de confianza para chicos: 5.103347 7.56332
```

```
cat("Intervalo de confianza para chicas: ", ic_m, "\n")
```

```
## Intervalo de confianza para chicas: 4.356505 6.976828
```

# conclusiones parciales Aunque no se encontraron diferencias significativas en las calificaciones basadas en el sexo, sí se encontraron diferencias significativas en función del método de enseñanza. Esto sugiere que lo más relevante para el rendimiento académico de los estudiantes es el método utilizado, más que el sexo de los estudiantes.

# PUNTO 5 realiza el anova para el efecto principal.

```
anova_metodo <- aov(calificacion ~ metodo, data = datos)
summary(anova_metodo)
```

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

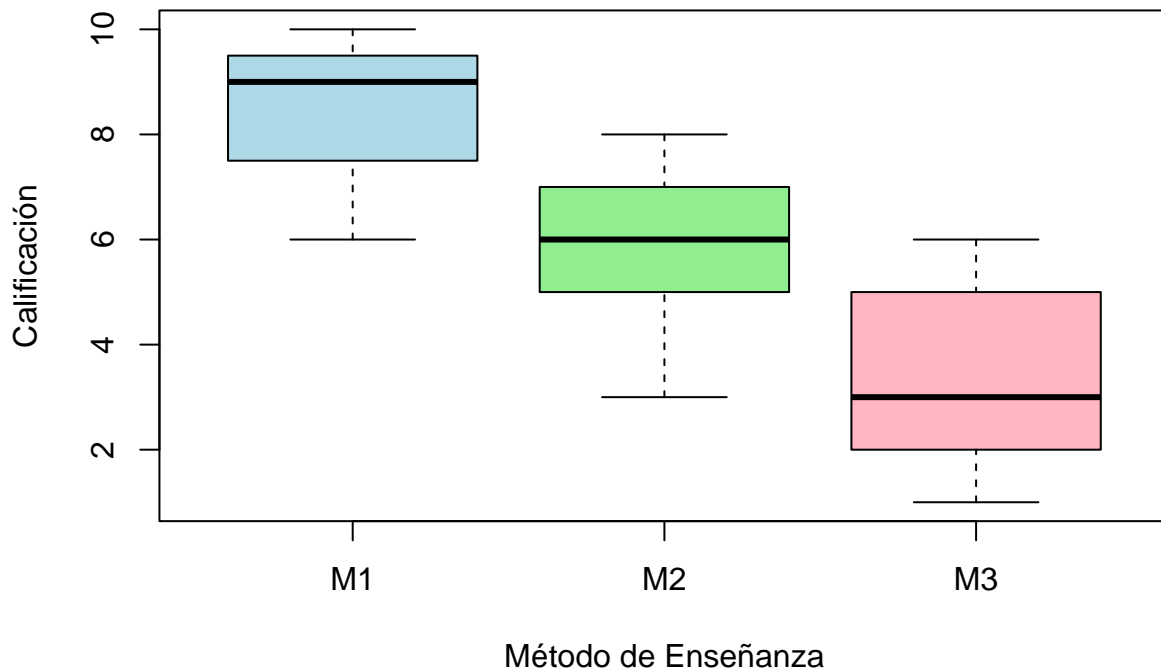
boxplot de rendimiento

```
# Boxplot de rendimiento por método de enseñanza
```

```
boxplot(calificacion ~ metodo, data = datos,
        main = "Boxplot de Rendimiento por Método de Enseñanza",
        xlab = "Método de Enseñanza", ylab = "Calificación",
        col = c("lightblue", "lightgreen", "lightpink"))
```



## Boxplot de Rendimiento por Método de Enseñanza



Calcula la media.

```
media_metodo <- aggregate(calificacion ~ metodo, data = datos, mean)
print(media_metodo)
```

```
##  metodo calificacion
## 1     M1           8.5
## 2     M2           6.0
## 3     M3           3.5
```

Haz los intervalos de confianza de rendimiento por método y Gráficelos

```
ic_m1 <- t.test(datos$calificacion[datos$metodo == "M1"])$conf.int
ic_m2 <- t.test(datos$calificacion[datos$metodo == "M2"])$conf.int
ic_m3 <- t.test(datos$calificacion[datos$metodo == "M3"])$conf.int

cat("Intervalo de confianza para Método 1: ", ic_m1, "\n")
```

```
## Intervalo de confianza para Método 1:  7.664961 9.335039
```

```
cat("Intervalo de confianza para Método 2: ", ic_m2, "\n")
```

```
## Intervalo de confianza para Método 2:  5.023175 6.976825
```

```
cat("Intervalo de confianza para Método 3: ", ic_m3, "\n")
```

```
## Intervalo de confianza para Método 3:  2.433377 4.566623
```

Realiza la prueba de comparaciones múltiples de Tukey.

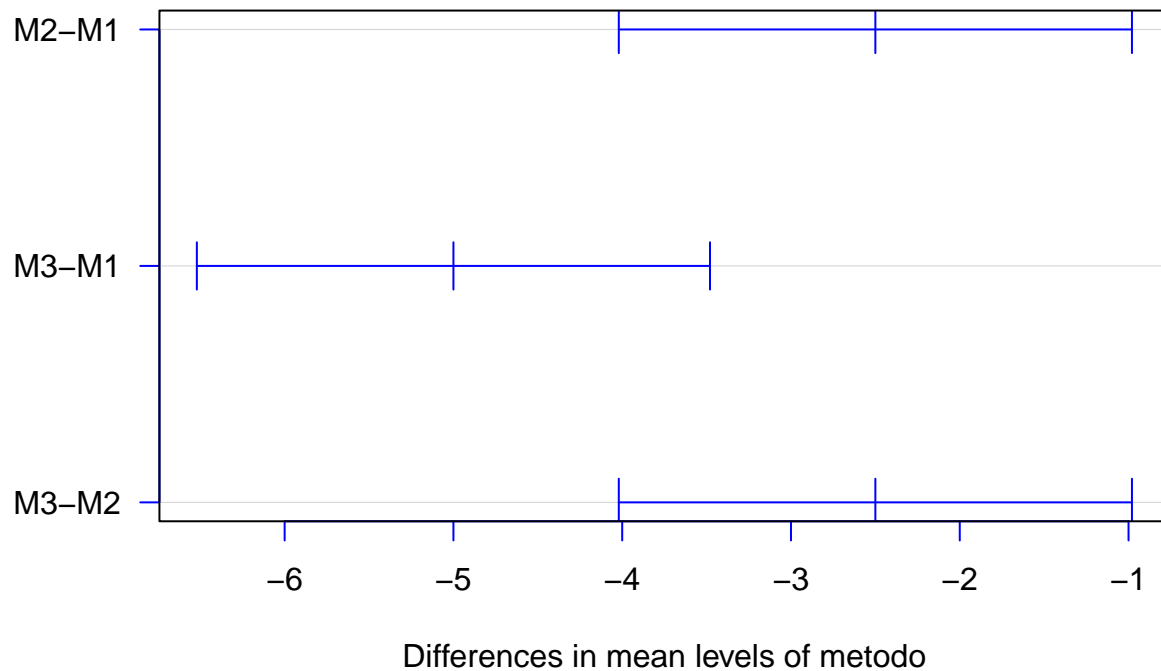
```
# Prueba de Tukey para comparaciones múltiples
tukey_result <- TukeyHSD(anova_metodo)
print(tukey_result)
```

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

Grafica los intervalos de confianza de Tukey.

```
plot(tukey_result, las = 1, col = "blue")
```

### 95% family-wise confidence level



#conclusiones parciales El método de enseñanza tiene un impacto significativo en las calificaciones de

los estudiantes. Específicamente, el método 1 (posiblemente el más completo o efectivo) genera mejores resultados que los métodos 2 y 3. Los métodos tienen un impacto lineal en la calificación, siendo el mejor de los métodos el primero y el peor el tercero.

#PUNTO 6 Validez del modelo principal #Normalidad

```
# Residuos del modelo
residuos <- residuals(anova_metodo)

# Prueba de normalidad de Shapiro-Wilk
shapiro_test <- shapiro.test(residuos)
shapiro_test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuos
## W = 0.96734, p-value = 0.3573
```

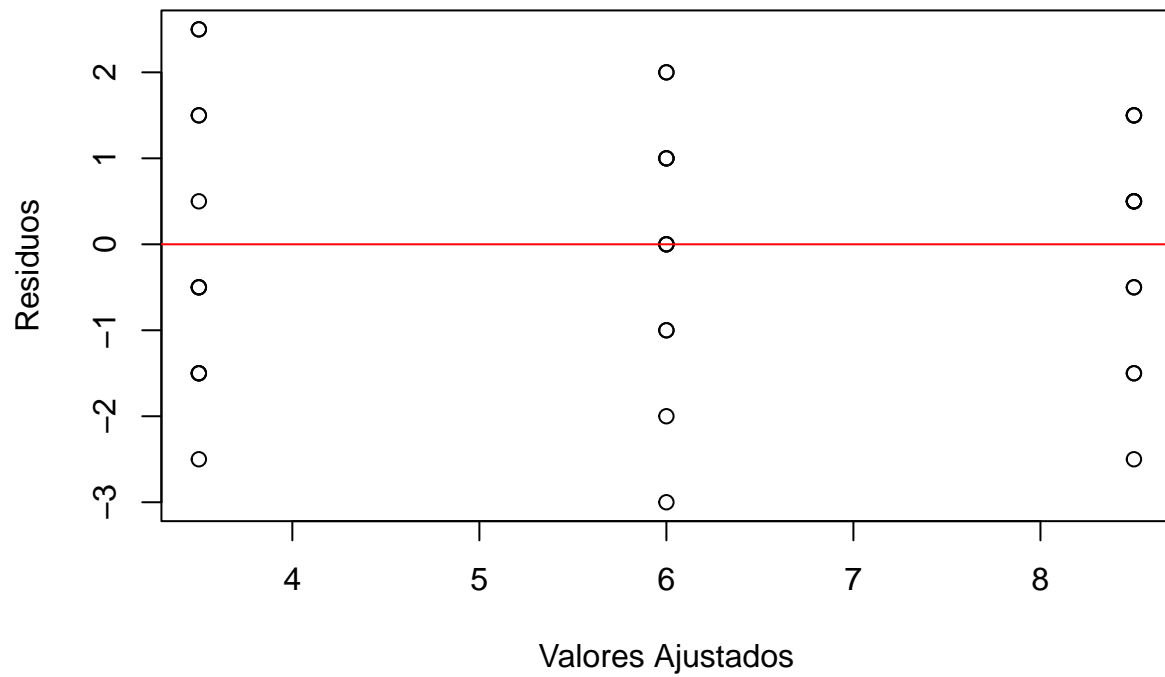
#Homocedasticidad

```
# Prueba de homocedasticidad de Bartlett
bartlett_test <- bartlett.test(calificacion ~ metodo, data = datos)
bartlett_test
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  calificacion by metodo
## Bartlett's K-squared = 0.63268, df = 2, p-value = 0.7288
```

```
# Gráfico de residuos vs. valores ajustados
plot(fitted(anova_metodo), residuos,
     main = "Residuos vs. Valores Ajustados",
     xlab = "Valores Ajustados", ylab = "Residuos")
abline(h = 0, col = "red")
```

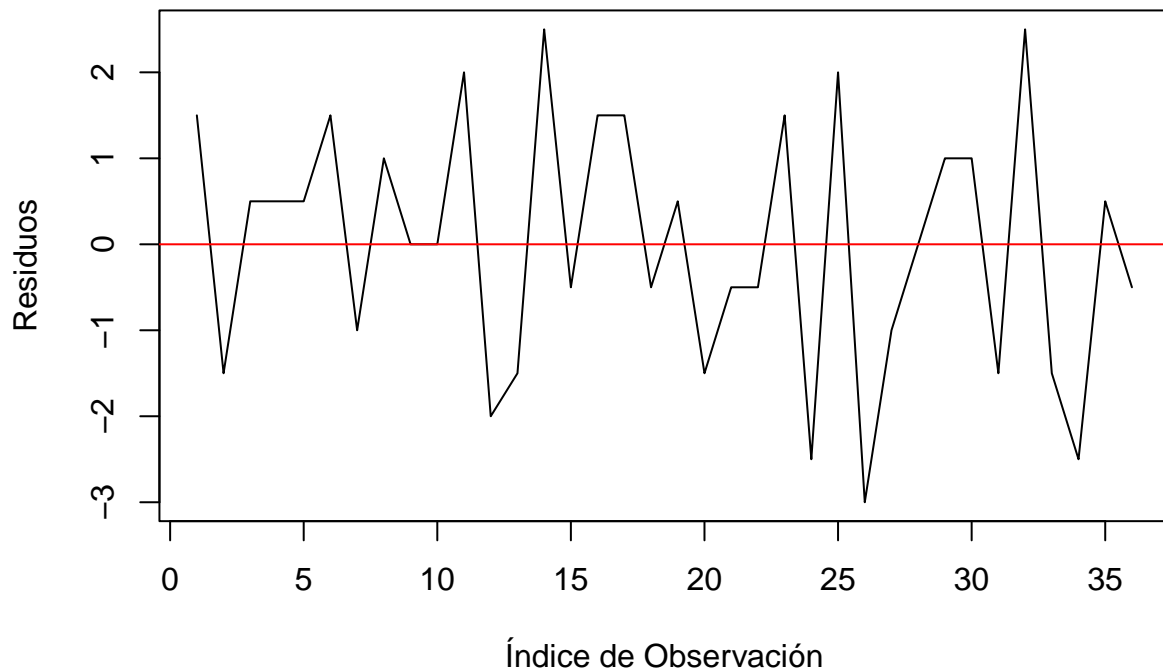
## Residuos vs. Valores Ajustados



#Independencia

```
# Gráfico de residuos en función del orden de los datos
plot(residuos, type = "l",
     main = "Independencia de los Residuos",
     xlab = "Índice de Observación", ylab = "Residuos")
abline(h = 0, col = "red")
```

## Independencia de los Residuos



#Relación lineal entre las variables (coeficiente de determinación).

```
ss_total <- sum((datos$calificacion - mean(datos$calificacion))^2)

# Suma de cuadrados de los residuos (SS_Residual)
ss_residual <- sum(residuals(anova_metodo)^2)

# Calcular R^2
r_squared <- 1 - (ss_residual / ss_total)
r_squared
```

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

#Conclusiones

el análisis realizado demuestra de manera concluyente que el método de enseñanza es un factor determinante en las calificaciones de los estudiantes. Los resultados no solo son estadísticamente significativos, sino que también tienen implicaciones prácticas importantes. La validación del modelo a través de pruebas de normalidad, homocedasticidad, independencia y la relación lineal asegura que las conclusiones derivadas del ANOVA son robustas y confiables. Recomendaría en este escenario implementar el método de enseñanza 1.

#

#

#

```
#-----
#-----
#-----
#-----
#-----
#-----
```

## #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 (factorA) 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

```
# Crear el DataFrame
material <- c(rep("Acero", 2*5), rep("Aluminio", 2*5), rep("Plástico", 2*5))
proveedor <- rep(1:5, times = 3, each = 2)
vibracion <- c(13.1, 13.2, 16.3, 15.8, 13.7, 14.3, 15.7, 15.8, 13.5, 12.5,
               15.0, 14.8, 15.7, 16.4, 13.9, 14.3, 13.7, 14.2, 13.4, 13.8,
               14.0, 14.3, 17.2, 16.7, 12.4, 12.3, 14.4, 13.9, 13.2, 13.1)

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

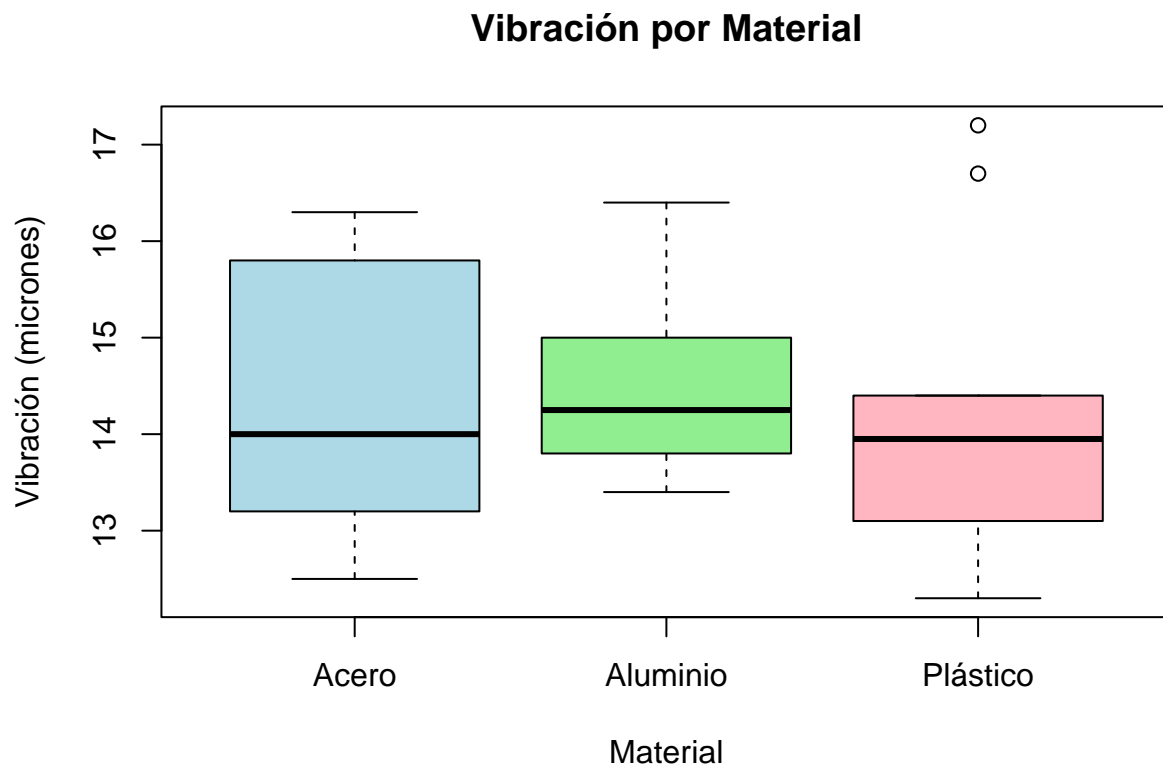
# Ver el DataFrame
print(datos)
```

```
##      material proveedor vibracion
## 1      Acero          1      13.1
## 2      Acero          1      13.2
## 3      Acero          2      16.3
## 4      Acero          2      15.8
## 5      Acero          3      13.7
## 6      Acero          3      14.3
## 7      Acero          4      15.7
## 8      Acero          4      15.8
## 9      Acero          5      13.5
## 10     Acero          5      12.5
## 11 Aluminio          1      15.0
## 12 Aluminio          1      14.8
## 13 Aluminio          2      15.7
## 14 Aluminio          2      16.4
## 15 Aluminio          3      13.9
## 16 Aluminio          3      14.3
## 17 Aluminio          4      13.7
## 18 Aluminio          4      14.2
## 19 Aluminio          5      13.4
## 20 Aluminio          5      13.8
## 21 Plástico          1      14.0
## 22 Plástico          1      14.3
## 23 Plástico          2      17.2
## 24 Plástico          2      16.7
```

```
## 25 Plástico      3      12.4
## 26 Plástico      3      12.3
## 27 Plástico      4      14.4
## 28 Plástico      4      13.9
## 29 Plástico      5      13.2
## 30 Plástico      5      13.1
```

```
#boxplot (analysis)
```

```
boxplot(vibracion ~ material, data = datos,
        main = "Vibración por Material",
        xlab = "Material", ylab = "Vibración (micrones)",
        col = c("lightblue", "lightgreen", "lightpink"))
```



el gráfico sugiere que el material de la carcasa tiene un impacto en la vibración de los motores pareciendo ser el plástico el material más consistente para reducir las vibraciones

#Hipotesis

primera hipótesis

$H_0: \mu = 0$   $H_1: \mu \neq 0$

segunda hipótesis

$H_0: \sigma = 0$   $H_1: \sigma \neq 0$

tercera hipótesis

$H_0: \mu = 0$   $H_1: \mu \neq 0$

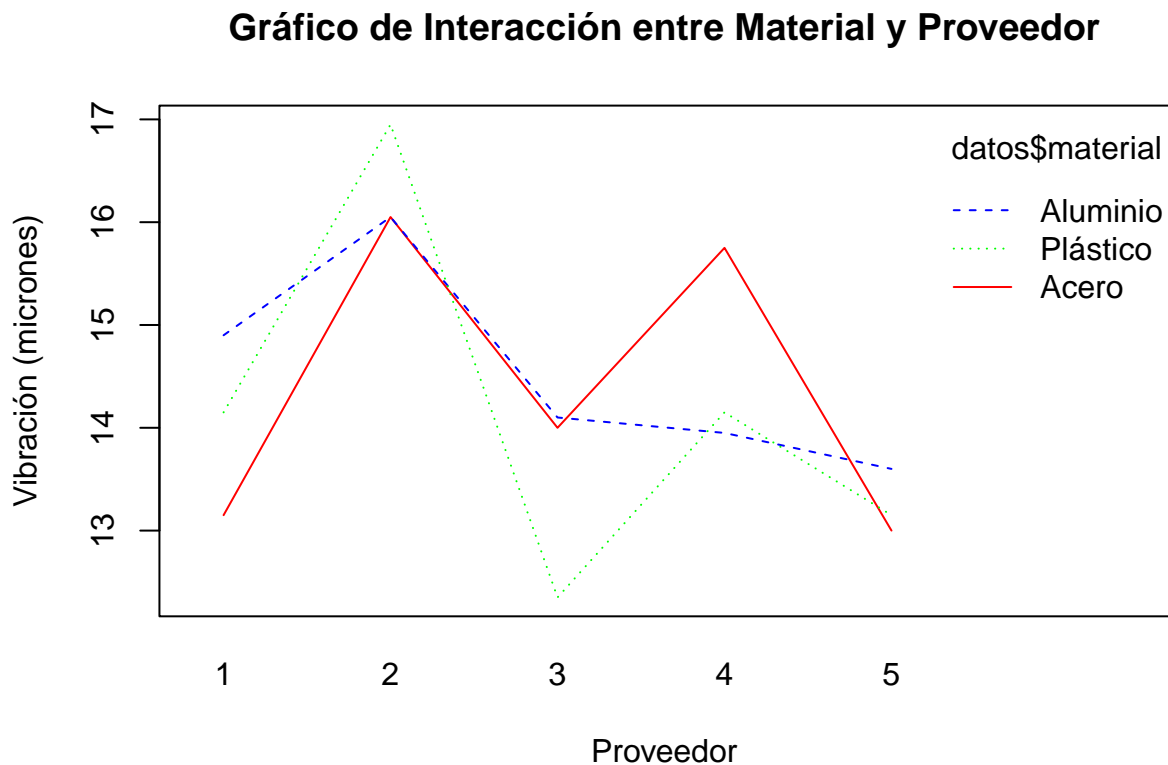
#PUNTO 3 #Anova para dos vias de interaccion

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

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## material      2   0.70   0.352   0.207 0.8145
## proveedor     1   6.80   6.801   3.995 0.0571 .
## material:proveedor 2   2.30   1.149   0.675 0.5186
## Residuals    24  40.85   1.702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#grafica de interaccion

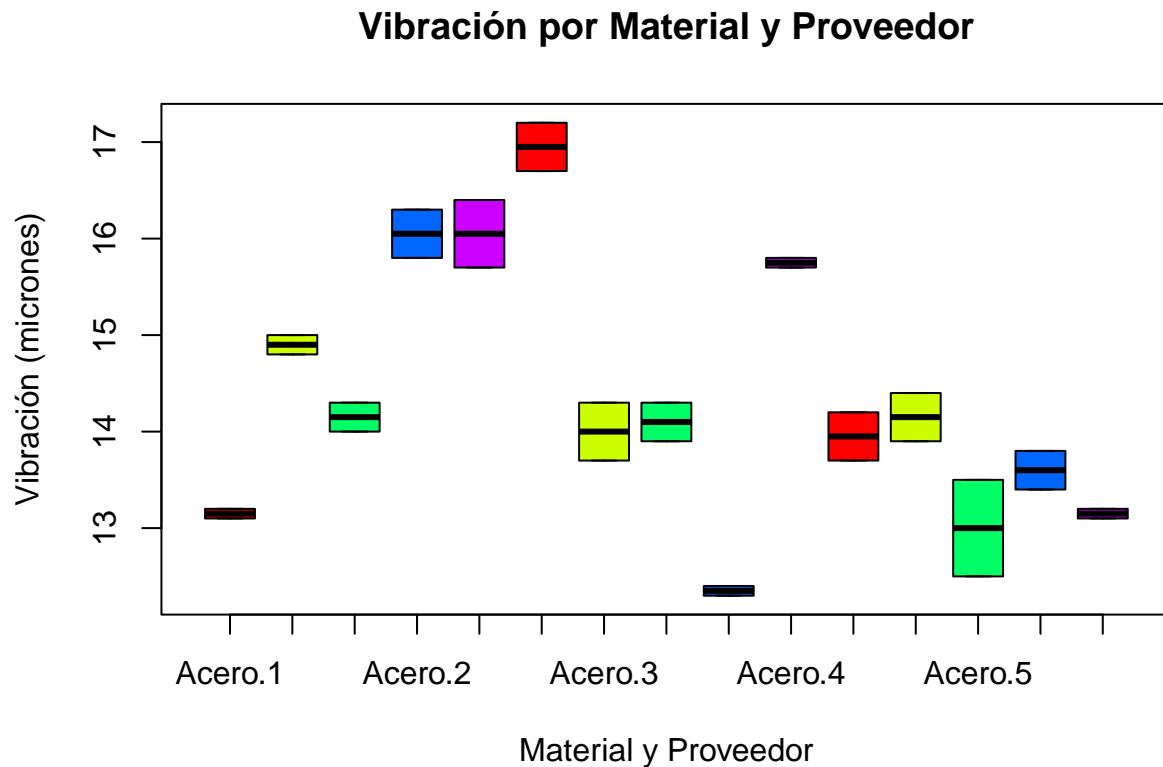
```
# Gráfica de interacción
interaction.plot(datos$proveedor, datos$material, datos$vibracion,
  main = "Gráfico de Interacción entre Material y Proveedor",
  xlab = "Proveedor", ylab = "Vibración (micrones)",
  col = c("red", "blue", "green"), lty = 1:3)
```



#boxplot



```
# Boxplot de vibración por material y proveedor
boxplot(vibracion ~ material:proveedor, data = datos,
        main = "Vibración por Material y Proveedor",
        xlab = "Material y Proveedor", ylab = "Vibración (micrones)",
        col = rainbow(5))
```



# analisis de Anova para dos vias de interaccion

```
# ANOVA de dos vías sin interacción
anova_sin_interaccion_dos <- aov(vibracion ~ material + proveedor, data = datos)
summary(anova_sin_interaccion_dos)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## material    2   0.70   0.352   0.212 0.8101
## proveedor    1   6.80   6.801   4.098 0.0533 .
## Residuals   26  43.15   1.660
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Conclusiones aceleradas

Debido a los resultados obtenidos de la prueba anova con y sin interacción se puede concluir que no hay interacción significativa entre el material y el proveedor, así como se puede concluir que el material no influye en las vibraciones y la influencia del proveedor es insignificante.