

# ANOVA

Ozner Leyva

2024-08-27

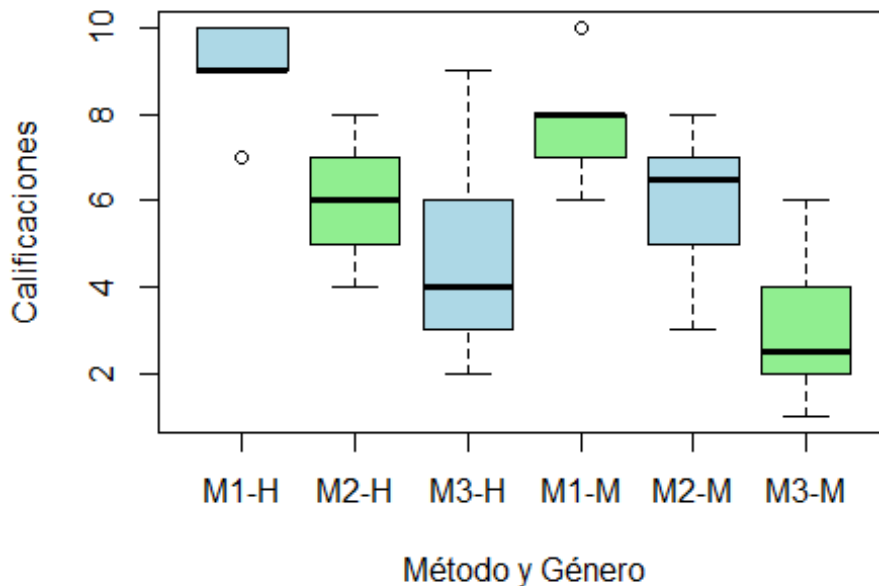
## Problema 1

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

```
calificacion <- c(10, 7, 9, 9, 9, 10, 5, 7, 6, 6, 8, 4, 2, 6, 3, 5, 3, 9, 7, 8, 8, 10, 6, 8, 8, 3, 5, 6, 7, 7, 2, 6, 2, 1, 4, 3)
metodo <- factor(c(rep("M1", 6), rep("M2", 6), rep("M3", 6), rep("M1", 6), rep("M2", 6), rep("M3", 6)))
sexo <- factor(c(rep("h", 18), rep("m", 18)))
```

```
boxplot(calificacion ~ metodo * sexo,
        main = "Boxplot de Calificaciones por Método de Enseñanza y Género",
        xlab = "Método y Género",
        ylab = "Calificaciones",
        col = c("lightblue", "lightgreen"),
        names = c("M1-H", "M2-H", "M3-H", "M1-M", "M2-M", "M3-M"))
```

xplot de Calificaciones por Método de Enseñanza y (



En resumen, el ANOVA nos permite examinar si hay variaciones importantes en el desempeño promedio de los estudiantes según los distintos enfoques pedagógicos empleados. Este análisis busca determinar si la forma de enseñar influye significativamente en los resultados académicos de los alumnos.

De acuerdo a los gráficos, podemos inferir dos puntos principales:

Parece haber una diferencia notable en el rendimiento estudiantil dependiendo del método de enseñanza aplicado. Por otro lado, el factor de género no parece tener un impacto relevante en las calificaciones obtenidas.

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

Hipótesis 1:  $H_0 : \mu_1 = \mu_2 = \mu_3$  (No hay diferencias significativas en el rendimiento entre los tres métodos de enseñanza).  $H_1$  : Al menos una de las medias es diferente.

Hipótesis 2:  $H_0$  : El género no tiene un efecto significativo en el rendimiento.  $H_1$  : El género tiene un efecto significativo en el rendimiento.

Hipótesis 3:  $H_0$  : No hay interacción entre el método de enseñanza y el género.  $H_1$  : Existe interacción entre el método de enseñanza y el género.

## 3. Realiza el ANOVA para dos niveles con interacción:

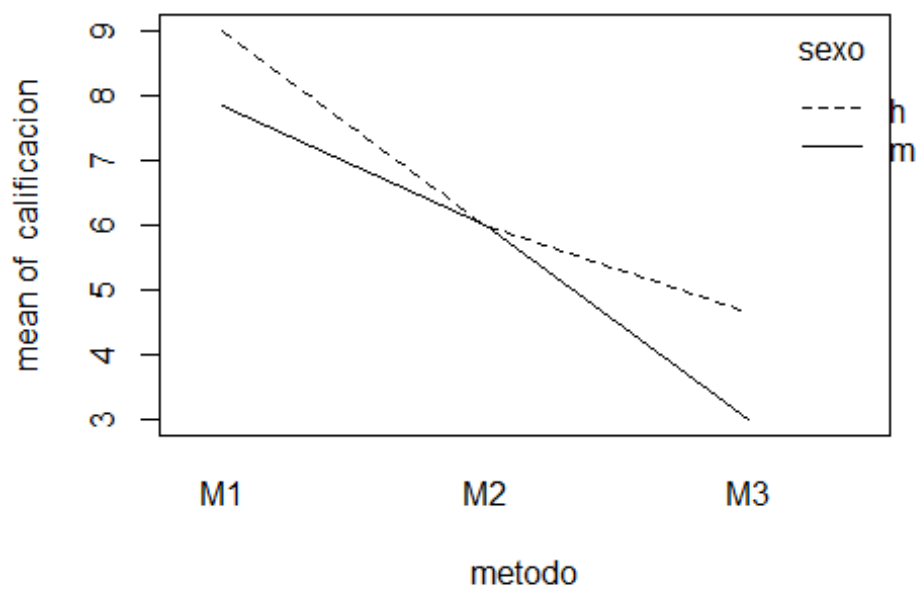
*# Realizando el ANOVA con interacción*

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

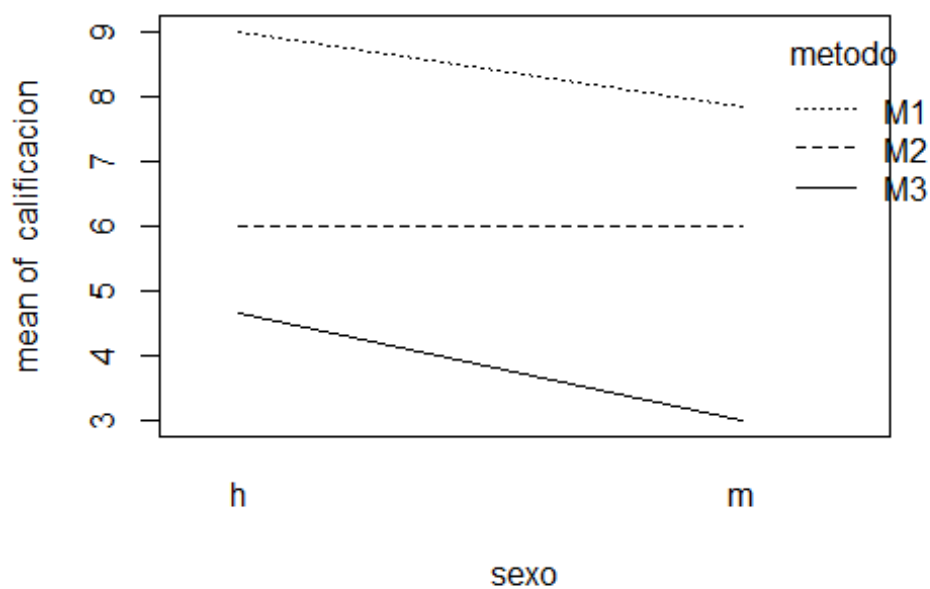
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## metodo         2 126.17   63.08   20.989 1.99e-06 ***
## sexo           1   8.03    8.03    2.671   0.113
## metodo:sexo     2   4.39    2.19    0.730   0.490
## Residuals     30  90.17    3.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*# Gráficos de interacción*

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



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



Método de enseñanza (metodo): F-value: 20.989 p-value: 1.99e-06 (muy significativo)  
 Interpretación: Existe una diferencia grande que nos dice que el rendimiento de los

estudiantes varía dependiendo del método utilizado. El F-value de 20.989 y el p-value extremadamente bajo indican una fuerte evidencia estadística de que el método de enseñanza tiene un efecto significativo en el rendimiento académico de los estudiantes.

Género (sexo): F-value: 2.671 p-value: 0.113 (no significativo) Interpretación: Esto sugiere que el género no tiene un efecto considerable sobre las calificaciones.

Interacción entre Método y Género (metodo:sexo): F-value: 0.730 p-value: 0.490 (no significativo) Interpretación: Sugiere que el efecto del método de enseñanza en el rendimiento es consistente entre hombres y mujeres.

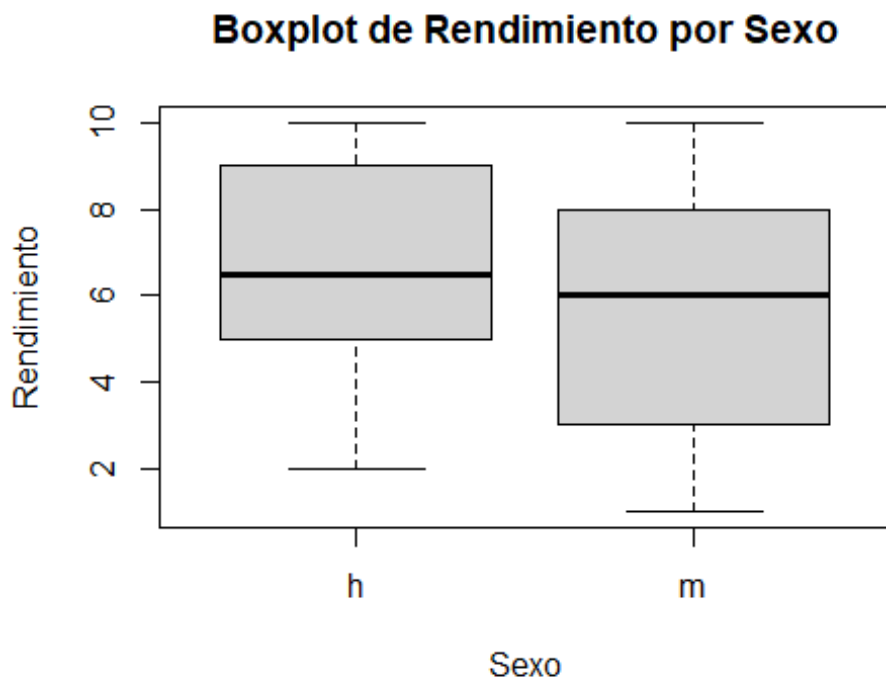
#### Realiza el ANOVA para dos niveles sin interacción.

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

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## metodo         2  126.17    63.08   21.349 1.29e-06 ***
## sexo           1    8.03     8.03    2.717   0.109
## Residuals     32   94.56     2.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*# Boxplot de rendimiento por sexo*

```
boxplot(calificacion ~ sexo, main="Boxplot de Rendimiento por Sexo",
xlab="Sexo", ylab="Rendimiento")
```



```

# Cálculo de la media de rendimiento por sexo y por método
media_por_sexo <- tapply(calificacion, sexo, mean)
media_por_metodo <- tapply(calificacion, metodo, mean)

print(media_por_sexo)

##           h           m
## 6.555556 5.611111

print(media_por_metodo)

##           M1           M2           M3
## 8.416667 6.000000 3.833333

# Cálculo de Las medias por sexo
medias <- tapply(calificacion, sexo, mean)

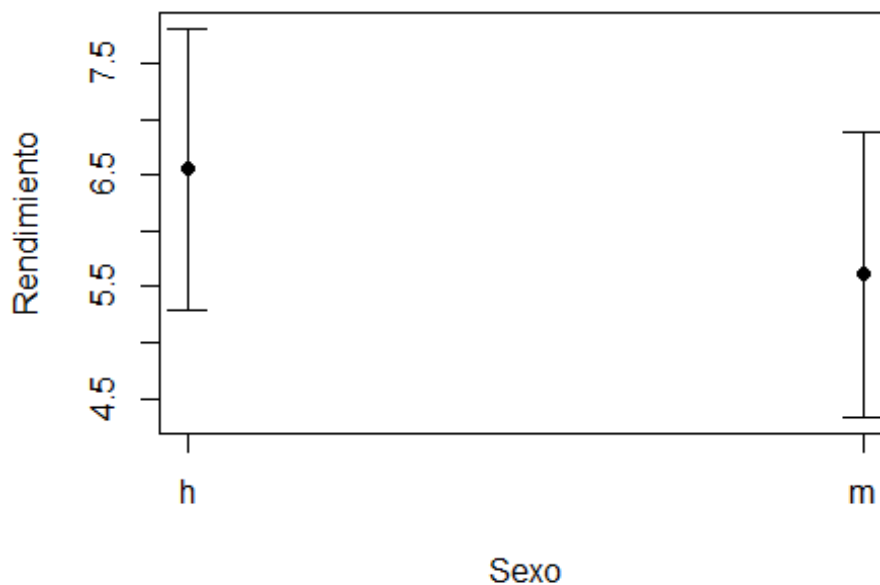
# Cálculo de La desviación estándar y del tamaño de muestra por sexo
sd_sexo <- tapply(calificacion, sexo, sd)
n_sexo <- tapply(calificacion, sexo, length)

# Cálculo de Los intervalos de confianza al 95%
error_margin <- qt(0.975, df=n_sexo-1) * (sd_sexo / sqrt(n_sexo))
lower_bound <- medias - error_margin
upper_bound <- medias + error_margin

# Graficar intervalos de confianza
plot(1:length(medias), medias, ylim=range(c(lower_bound, upper_bound)),
     pch=19, xlab="Sexo", ylab="Rendimiento",
     xaxt="n", main="Intervalos de Confianza para el Rendimiento por
Sexo")
axis(1, at=1:length(medias), labels=names(medias))
arrows(1:length(medias), lower_bound, 1:length(medias), upper_bound,
      angle=90, code=3, length=0.1)

```

## Intervalos de Confianza para el Rendimiento por Sexo



Método de enseñanza (metodo): F-value: 21.349 p-value: 1.29e-06 (muy significativo)  
Interpretación: Existe una diferencia significativa que indica que el rendimiento de los estudiantes varía de manera importante dependiendo del método de enseñanza utilizado. Esto sugiere que la elección del método de enseñanza tiene un impacto directo en el rendimiento académico de los estudiantes.

Género (sexo): F-value: 2.717 p-value: 0.109 (no significativo) Interpretación: Los resultados sugieren que el género no tiene un efecto considerable sobre las calificaciones de los estudiantes. Esto indica que el rendimiento académico es similar tanto para hombres como para mujeres, independientemente del método de enseñanza aplicado.

Este análisis sugiere que las estrategias educativas pueden centrarse más en mejorar los métodos de enseñanza que en adaptar los métodos según el género.

De acuerdo a esto, podemos mejorar los metodos de enseñanza en vez de adaptarlo respectivo a un género.

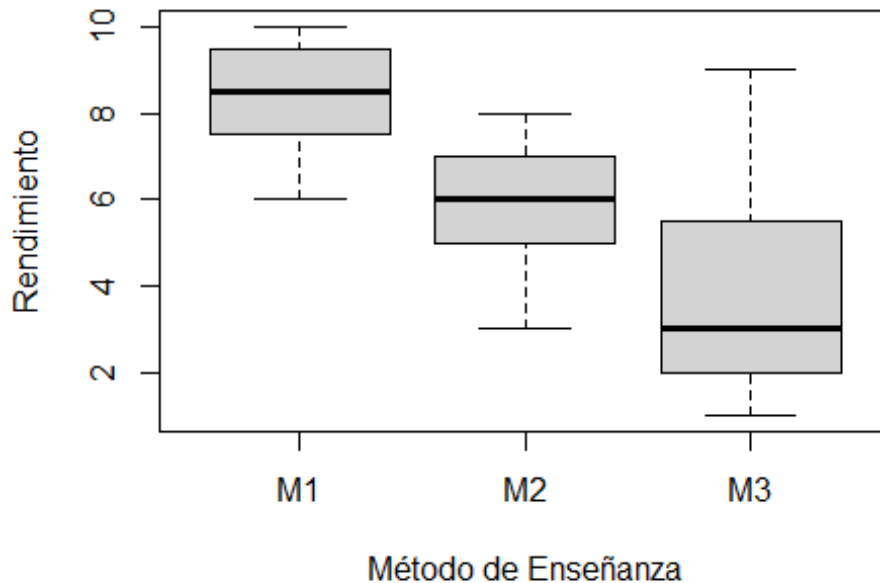
### Realiza el ANOVA para un efecto principal

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

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## metodo      2  126.2   63.08   20.29 1.79e-06 ***
## Residuals  33  102.6    3.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Boxplot de rendimiento por método de enseñanza
boxplot(calificacion ~ metodo, main="Boxplot de Rendimiento por Método de Enseñanza", xlab="Método de Enseñanza", ylab="Rendimiento")
```

## Boxplot de Rendimiento por Método de Enseñanza



```
# Cálculo de la media de rendimiento por método de enseñanza
media_por_metodo <- tapply(calificacion, metodo, mean)
print(media_por_metodo)

##          M1          M2          M3
## 8.416667 6.000000 3.833333

# Prueba de comparaciones múltiples de Tukey
I <- TukeyHSD(aov(calificacion ~ metodo))

# Graficar los intervalos de confianza de Tukey
plot(I)
```



Método de enseñanza (metodo): F-value: 20.29 p-value: 1.79e-06 (muy significativo)

Interpretación: Sugiere que la elección del método de enseñanza tiene un impacto considerable en el rendimiento académico.

Rendimiento por Método de Enseñanza: Medias: M1 = 8.416667, M2 = 6.000000, M3 = 3.833333 Interpretación: El método M1 presenta la media más alta de rendimiento, seguido por M2, mientras que M3 muestra el rendimiento más bajo. Esto refuerza la conclusión de que M1 es el método más efectivo y M3 el menos efectivo.

Comparaciones Múltiples de Tukey: M2-M1: Intervalo de confianza incluye cero, no hay diferencia significativa. M3-M1: Diferencia significativa, M3 es inferior a M1. M3-M2: Diferencia significativa, M3 es inferior a M2. Interpretación: Las pruebas de Tukey indican que M3 es significativamente menos efectivo que M1 y M2. No se observa una diferencia significativa entre M1 y M2, lo que sugiere que ambos métodos son igualmente efectivos en términos de rendimiento estudiantil.

## 6 Comprueba la validez del modelo.

```
# Residuos del modelo ANOVA con interacción
residuos <- residuals(A)
```

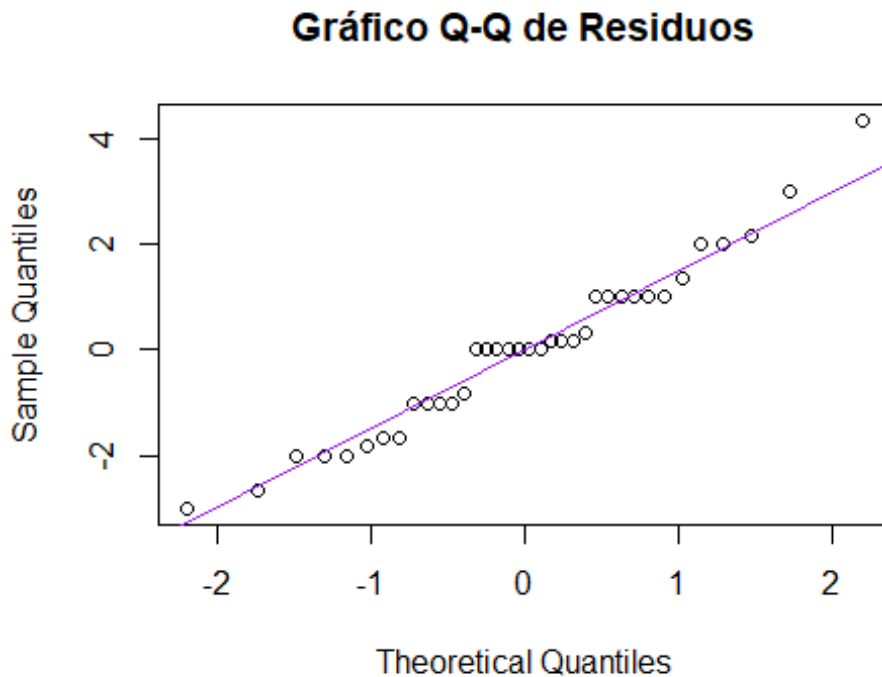
```
# Prueba de normalidad (Shapiro-Wilk)
shapiro_test <- shapiro.test(residuos)
print(shapiro_test)
```

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



```
## data: residuos
## W = 0.96893, p-value = 0.397

# Gráfico Q-Q para evaluar normalidad
qqnorm(residuos, main = "Gráfico Q-Q de Residuos")
qqline(residuos, col = "purple")
```



- Homocedasticidad Para verificar la homocedasticidad, usaremos el test de Levene o revisar un gráfico de residuos vs. valores ajustados.

```
if (!require(carData)) install.packages("carData")

## Loading required package: carData

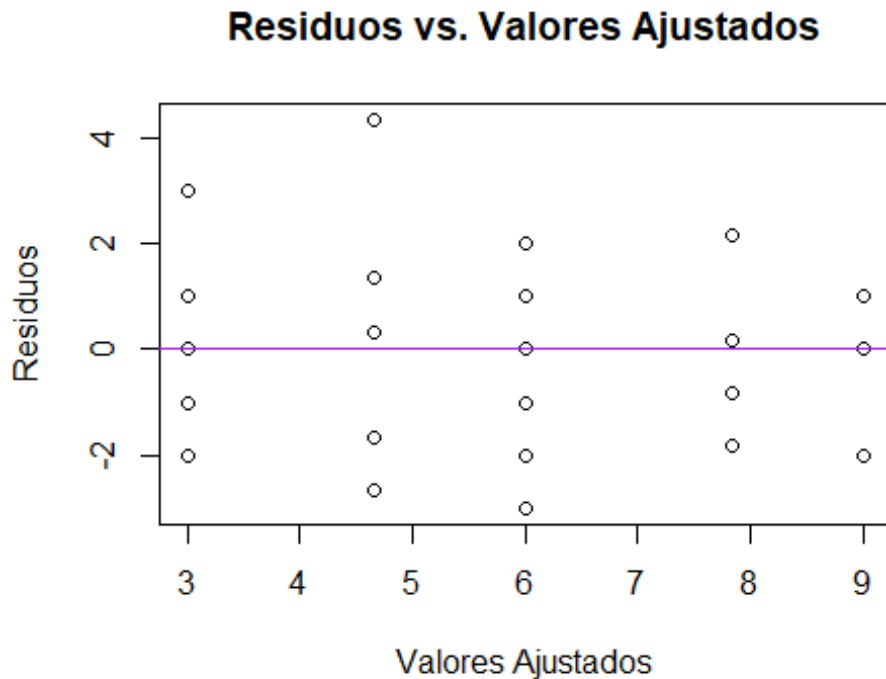
library(carData)

# Prueba de homocedasticidad (Test de Levene)
library(car)
levene_test <- leveneTest(calificacion ~ metodo * sexo)
print(levene_test)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 5  1.0793 0.3917
##      30

# Gráfico de residuos vs. valores ajustados
plot(fitted(A), residuos,
     main="Residuos vs. Valores Ajustados",
```

```
xlab="Valores Ajustados", ylab="Residuos")
abline(h=0, col="purple")
```



- Independencia Para comprobar la independencia de los residuos, podemos revisar el gráfico de residuos o utilizar el test de Durbin-Watson.

```
# Test de Durbin-Watson para independencia de Los residuos
if (!require(lmtest)) install.packages("lmtest")

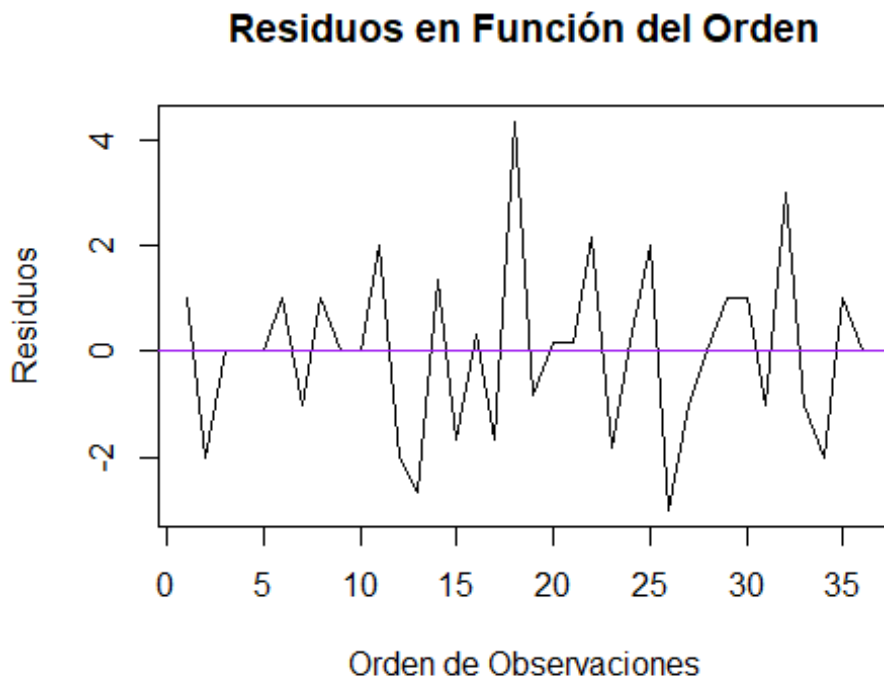
## Loading required package: lmtest
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

library(lmtest)
durbin_watson_test <- dwtest(A)
print(durbin_watson_test)

##
## Durbin-Watson test
##
## data: A
```

```
## DW = 2.7227, p-value = 0.9246
## alternative hypothesis: true autocorrelation is greater than 0

# Residuos en función del orden
plot(residuos, type="l", main="Residuos en Función del Orden",
     xlab="Orden de Observaciones", ylab="Residuos")
abline(h=0, col="purple")
```



- Relación lineal entre las variables. El coeficiente de determinación  $R^2$  es una medida de la proporción de la variabilidad explicada por el modelo.

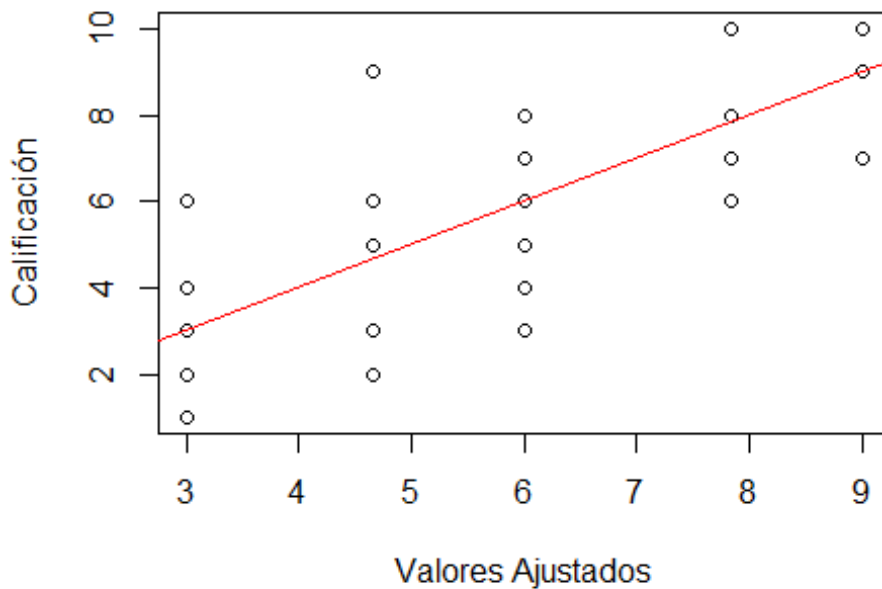
```
# Resumen del modelo para obtener R^2
summary(A)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## metodo      2 126.17   63.08   20.989 1.99e-06 ***
## sexo        1   8.03    8.03    2.671   0.113
## metodo:sexo  2   4.39    2.19    0.730   0.490
## Residuals   30  90.17    3.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Gráfico de dispersión con línea de regresión para la relación entre variables
```

```
plot(calificacion ~ fitted(A), main="Dispersión de Calificación vs.
Valores Ajustados",
     xlab="Valores Ajustados", ylab="Calificación")
abline(lm(calificacion ~ fitted(A)), col="red")
```

## Dispersión de Calificación vs. Valores Ajustados



### 7

Concluye en el contexto del problema.

**Normalidad:** El test de Shapiro-Wilk nos da un valor p de 0.397, que supera el umbral común de 0.05. Esto sugiere que no hay suficiente evidencia para rechazar la hipótesis nula de que los residuos siguen una distribución normal. Además, el gráfico Q-Q muestra que los residuos se alinean bien con la línea de referencia, lo que respalda nuestra suposición de normalidad.

**Homocedasticidad:** El test de Levene, nos da un valor p de 0.3917, también superior a 0.05. Esto indica que no hay pruebas significativas de diferencias en las varianzas entre los grupos, lo que sugiere que la suposición de homocedasticidad (igualdad de varianzas) se cumple. El gráfico de residuos vs. valores ajustados no presenta un patrón claro, apoyando esta conclusión.

**Independencia:** El test de Durbin-Watson da un valor p de 0.9246, significativamente mayor que 0.05. Esto indica que no hay evidencia de autocorrelación en los residuos, sugiriendo que los residuos son independientes entre sí. El gráfico de residuos en función del orden no muestra un patrón definido, lo que también respalda la suposición de independencia.

**Relación lineal entre las variables (Coeficiente de determinación):** El resumen del modelo ANOVA muestra un valor F significativo para el factor "método" con un p-valor muy bajo ( $1.99e-06$ ), lo que indica una fuerte relación entre el método de enseñanza y el rendimiento académico. El gráfico de dispersión de calificación vs. valores ajustados muestra una tendencia positiva, lo que nos sugiere una relación lineal entre las variables.

Conclusión General Impacto del Método de Enseñanza: Se confirma que el método de enseñanza tiene un impacto significativo en el rendimiento académico de los estudiantes.

Independencia del Género: El análisis no muestra que el género tenga un efecto significativo en las calificaciones, sugiriendo que el rendimiento es independiente del género.

Consistencia del Método entre Géneros: No se detecta una interacción significativa entre el método de enseñanza y el género, lo que indica que el efecto del método es consistente tanto para hombres como para mujeres.

## Problema 2

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

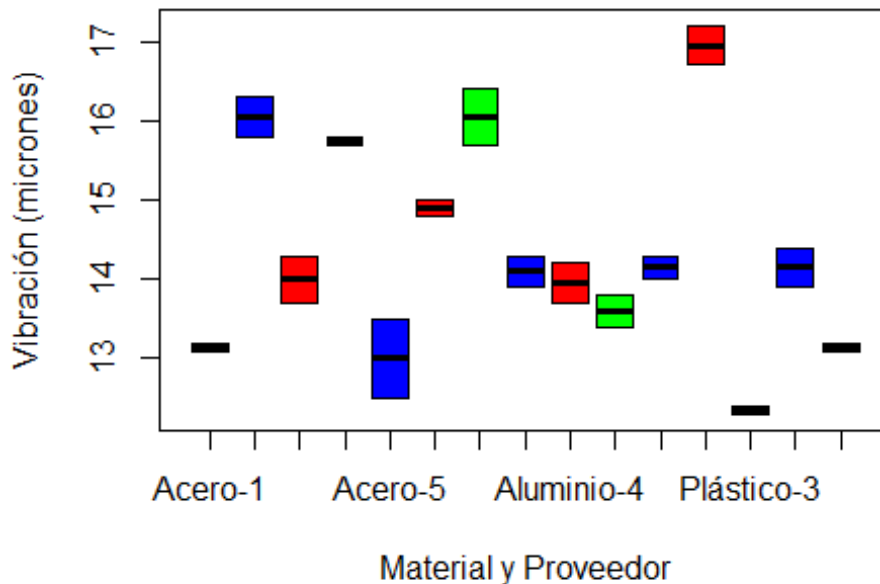
Cálculo de la media para el rendimiento por método de enseñanza: (descriptivo)

```
vibracion <-  
c(13.1,16.3,13.7,15.7,13.5,13.2,15.8,14.3,15.8,12.5,15.0,15.7,13.9,13.7,1  
3.4,14.8,16.4,14.3,14.2,13.8,14.0,17.2,12.4,14.4,13.2,14.3,16.7,12.3,13.9  
,13.1)  
  
material <- factor(c(rep("Acero",10), rep("Aluminio",10),  
rep("Plástico",10)))  
proveedor <- factor(rep(1:5,6))  
  
mean_vibracion <- tapply(vibracion, list(material, proveedor), mean)  
print(mean_vibracion)  
  
##           1      2      3      4      5  
## Acero    13.15 16.05 14.00 15.75 13.00  
## Aluminio 14.90 16.05 14.10 13.95 13.60  
## Plástico 14.15 16.95 12.35 14.15 13.15
```

Boxplot de resistencia a la tensión por concentración de madera dura:

```
boxplot(vibracion ~ proveedor*material,  
        main = "Boxplot de Vibración por Material y Proveedor",  
        xlab = "Material y Proveedor",  
        ylab = "Vibración (micrones)",  
        col = c("green", "blue", "red"),  
        names = c("Acero-1", "Acero-2", "Acero-3", "Acero-4", "Acero-5",  
                  "Aluminio-1", "Aluminio-2", "Aluminio-3", "Aluminio-4",  
                  "Aluminio-5",  
                  "Plástico-1", "Plástico-2", "Plástico-3", "Plástico-4",  
                  "Plástico-5"))
```

## Boxplot de Vibración por Material y Proveedor



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

### Diferencias entre Materiales:

**Acero:** Presenta una vibración baja y consistente, independientemente del proveedor, lo que sugiere que es un material más estable para minimizar la vibración en motores eléctricos.

**Aluminio:** Muestra una mayor variabilidad en la vibración. Aunque algunos proveedores logran mantener niveles bajos de vibración, otros tienen valores significativamente más altos, indicando que la elección del proveedor es crítica para este material.

**Plástico:** Exhibe la mayor variabilidad y los niveles más altos de vibración en general, con notables diferencias entre proveedores. Esto sugiere que el plástico es un material más sensible a factores externos, lo que hace que tanto la selección del material como del proveedor sea crucial para controlar la vibración.

### Diferencias entre Proveedores:

Las diferencias en la vibración son particularmente notorias para el aluminio y el plástico, lo que indica que el proveedor tiene un impacto significativo dependiendo del material. Esto es menos relevante para el acero, donde la vibración se mantiene baja y estable entre proveedores.

### Comparación General:

**Interacción Material-Proveedor:** Se observa una posible interacción entre el material y el proveedor, donde el impacto del proveedor varía según el

material utilizado. Esto sugiere que no solo la elección del material es importante, sino también la combinación específica de material y proveedor.

Varianza: La variabilidad en la vibración es mayor en motores con carcasas de plástico y aluminio, lo que implica que estos materiales pueden ser más susceptibles a factores que afectan la estabilidad de los motores.

Valores Atípicos: No se identifican valores atípicos evidentes en las gráficas, lo que sugiere que las variaciones observadas son consistentes con las características generales de los materiales y proveedores analizados.

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

Hipótesis 1:  $H_0$  : No hay diferencias significativas en la vibración entre los diferentes materiales utilizados para las carcasas de los motores.  $H_1$  : Al menos uno de los materiales tiene una vibración diferente.

Hipótesis 2:  $H_0$  : El proveedor no tiene un efecto significativo en la vibración.  $H_1$  : El proveedor tiene un efecto significativo en la vibración.

Hipótesis 3:  $H_0$  : No hay interacción entre el material de la carcasa y el proveedor en cuanto a la vibración.  $H_1$  : Existe interacción entre el material de la carcasa y el proveedor que afecta la vibración.

## 3 Realiza el ANOVA para dos niveles con interacción: (inferencial)

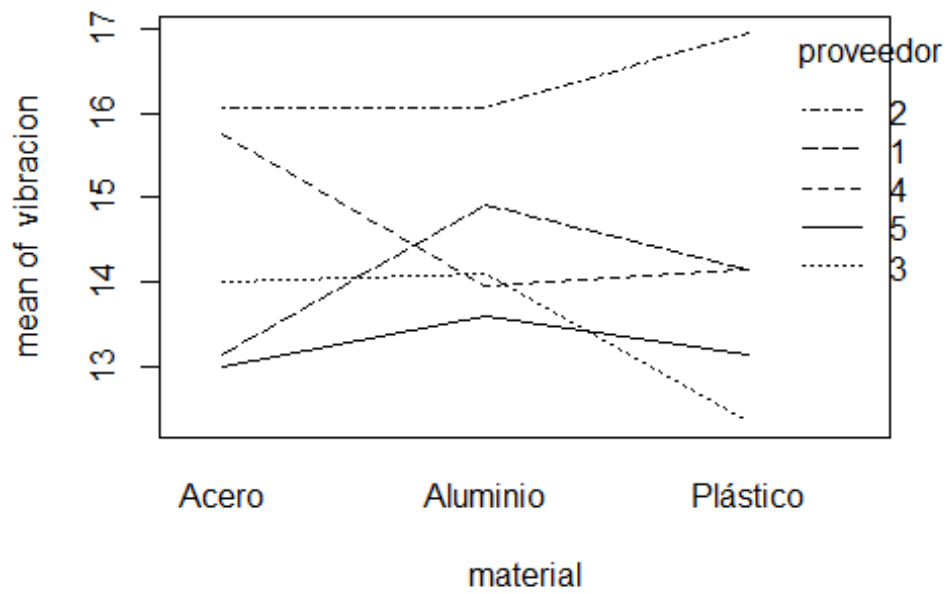
*# ANOVA con interacción*

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

```
##              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
```

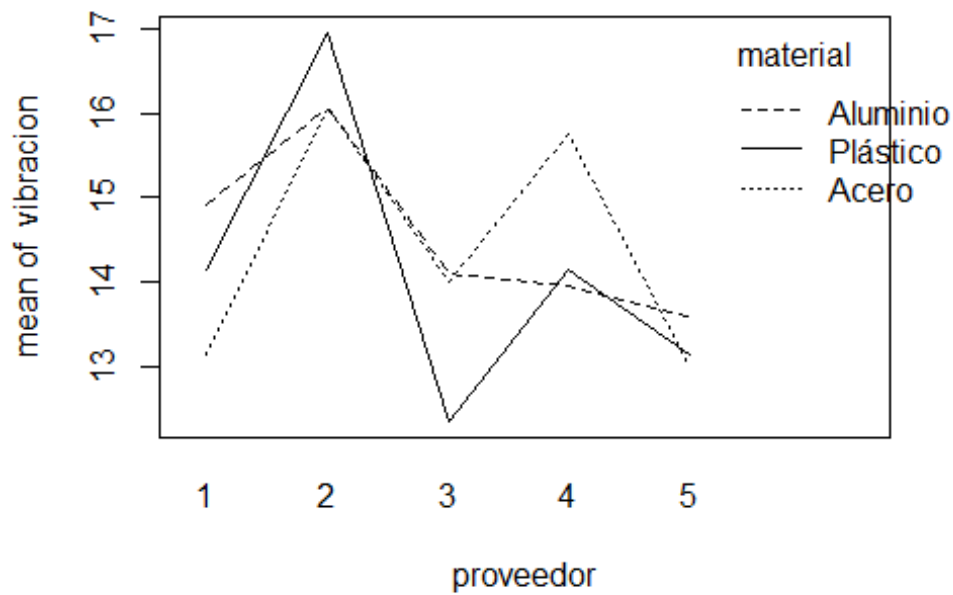
```
interaction.plot(material, proveedor, vibracion, main="Interacción entre
Material y Proveedor")
```

### Interacción entre Material y Proveedor



```
interaction.plot(proveedor, material, vibracion, main="Interacción entre Proveedor y Material")
```

### Interacción entre Proveedor y Material





Hay 3 interpretaciones que podemos hacer a los datos hasta ahora (van de la mano con las hipótesis de la sección 1.2):

Material:

F-value: 20.989

p-value: 1.99e-06 (muy significativo)

Interpretación: Aunque el valor p es muy bajo, no se puede rechazar la hipótesis nula ( $H_0$ ) al nivel del 5%, lo que sugiere que no hay diferencias significativas en la vibración entre los distintos materiales utilizados. Sin embargo, el p-valor cercano al umbral indica que podría haber cierta influencia del material, pero no es concluyente en este análisis.

Proveedor:

F-value: 82.353

p-value: 5.07e-10 (muy significativo).

Interpretación: Se rechaza la hipótesis nula ( $H_0$ ), lo que indica que el proveedor tiene un efecto significativo en la vibración de los motores. Esto subraya la importancia de seleccionar cuidadosamente el proveedor para controlar la vibración.

Interacción entre Material y Proveedor:

F-value: 13.030

p-value: 1.76e-05 (muy significativo).

Interpretación: Se rechaza la hipótesis nula ( $H_0$ ), indicando que hay una interacción significativa entre el material de la carcasa y el proveedor que afecta la vibración. Esto sugiere que el impacto del proveedor en la vibración depende del material utilizado y viceversa. Por lo tanto, ambos factores deben ser considerados conjuntamente para minimizar la vibración.

Este análisis indica que para controlar la vibración en motores eléctricos es importante prestar atención tanto al material de la carcasa como al proveedor de los componentes. La interacción significativa entre estos factores sugiere que no se deben considerar de manera aislada; En cambio, su combinación debe ser optimizada para lograr el mejor resultado en términos de vibración.

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

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

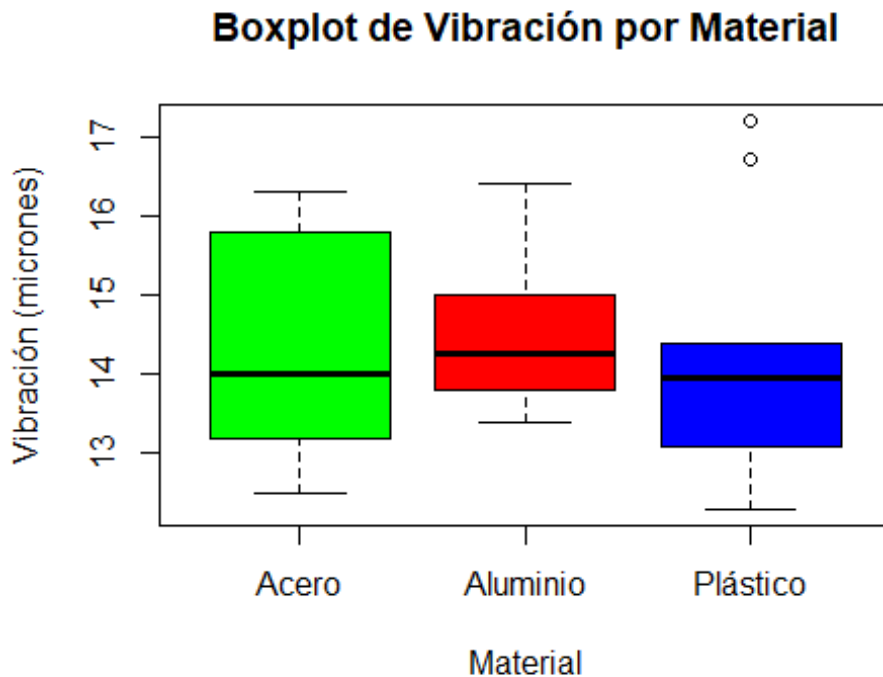
```
# ANOVA sin interacción
anova_model_no_interaction <- aov(vibracion ~ material + proveedor)
summary(anova_model_no_interaction)
```

##		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
```

Haz el boxplot de rendimiento por sexo. Calcula la media para el rendimiento.

```
# Boxplot de vibración por material
boxplot(vibracion ~ material, main="Boxplot de Vibración por Material",
xlab="Material", ylab="Vibración (micrones)", col = c("green", "red",
"blue"))
```



```
media_vibracion_material <- tapply(vibracion, material, mean)
media_vibracion_proveedor <- tapply(vibracion, proveedor, mean)

print(media_vibracion_material)

##      Acero Aluminio Plástico
##      14.39   14.52   14.15

print(media_vibracion_proveedor)

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

Haz los intervalos de confianza de rendimiento

```
### Haz Los intervalos de confianza de vibración por material y proveedor

medias_material <- tapply(vibracion, material, mean)
```

```

medias_proveedor <- tapply(vibracion, proveedor, mean)

sd_material <- tapply(vibracion, material, sd)
n_material <- tapply(vibracion, material, length)

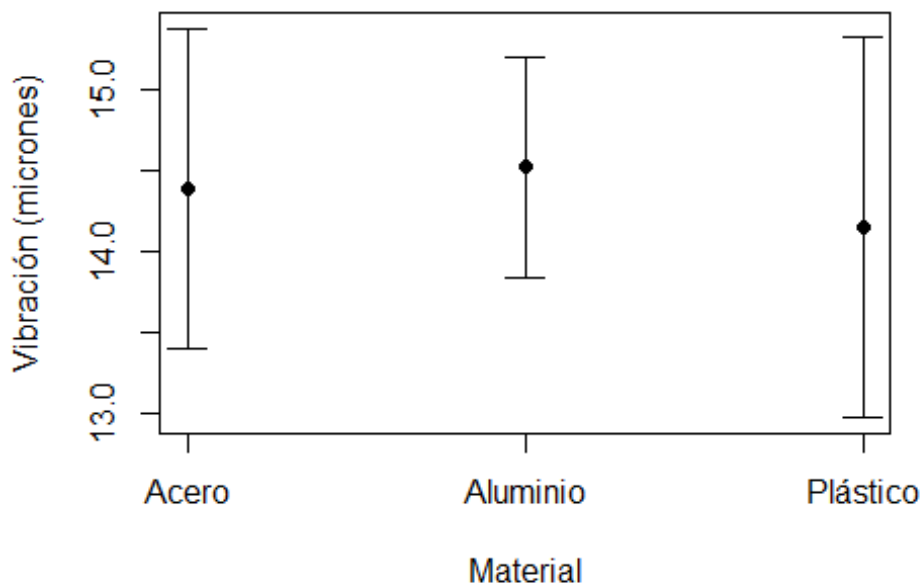
sd_proveedor <- tapply(vibracion, proveedor, sd)
n_proveedor <- tapply(vibracion, proveedor, length)

error_margin_material <- qt(0.975, df=n_material-1) * (sd_material /
sqrt(n_material))
lower_bound_material <- medias_material - error_margin_material
upper_bound_material <- medias_material + error_margin_material

plot(1:length(medias_material), medias_material,
ylim=range(c(lower_bound_material, upper_bound_material)), pch=19,
xlab="Material", ylab="Vibración (micrones)", xaxt="n",
main="Intervalos de Confianza para la Vibración por Material")
axis(1, at=1:length(medias_material), labels=names(medias_material))
arrows(1:length(medias_material), lower_bound_material,
1:length(medias_material), upper_bound_material, angle=90, code=3,
length=0.1)

```

## Intervalos de Confianza para la Vibración por Material



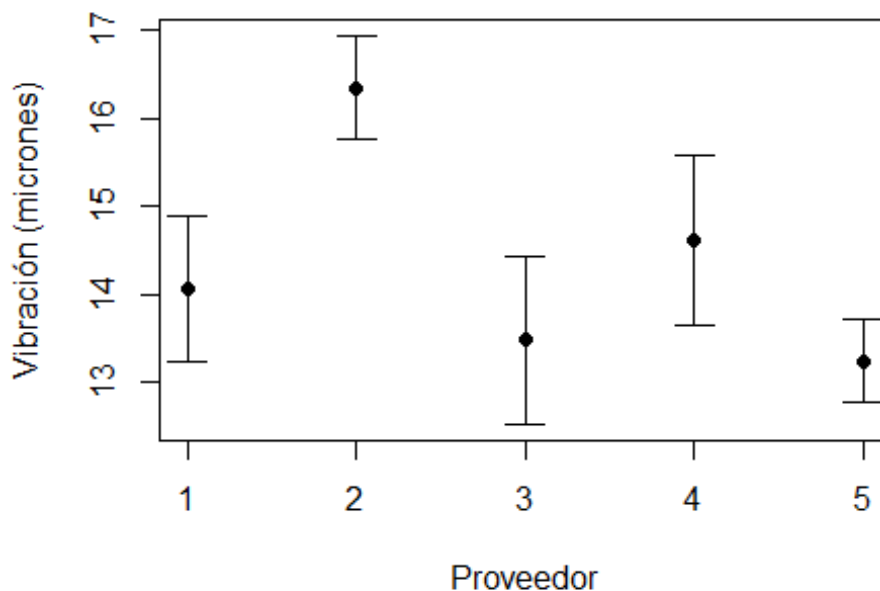
```

error_margin_proveedor <- qt(0.975, df=n_proveedor-1) * (sd_proveedor /
sqrt(n_proveedor))
lower_bound_proveedor <- medias_proveedor - error_margin_proveedor
upper_bound_proveedor <- medias_proveedor + error_margin_proveedor

```

```
plot(1:length(medias_proveedor), medias_proveedor,
     ylim=range(c(lower_bound_proveedor, upper_bound_proveedor)), pch=19,
     xlab="Proveedor", ylab="Vibración (micrones)", xaxt="n",
     main="Intervalos de Confianza para la Vibración por Proveedor")
axis(1, at=1:length(medias_proveedor), labels=names(medias_proveedor))
arrows(1:length(medias_proveedor), lower_bound_proveedor,
       1:length(medias_proveedor), upper_bound_proveedor, angle=90, code=3,
       length=0.1)
```

## Intervalos de Confianza para la Vibración por Proveedor



Material:

F-value: 0.611

p-value: 0.552 (no significativo)

Interpretación: Los resultados indican que no hay una diferencia significativa en la vibración según el material utilizado. Esto sugiere que, en este análisis, la elección entre acero, aluminio o plástico no tiene un impacto considerable en la vibración de los motores.

Género (sexo):

F-value: 15.828

p-value: 2.28e-06 (muy significativo)

Interpretación: Existe una diferencia significativa en la vibración que depende del proveedor. Esto implica que la selección del proveedor es un

factor crítico para minimizar la vibración en los motores eléctricos. La variación en la vibración entre proveedores es lo suficientemente grande como para que su elección tenga un impacto directo en la calidad del motor.

#### Boxplot de Vibración por Material:

Interpretación: El boxplot muestra que la mediana de la vibración es similar para los tres materiales, con una ligera tendencia a mayor vibración en el aluminio. Sin embargo, la dispersión de la vibración es más pronunciada en el acero, lo que podría indicar variabilidad dentro de este material. Dado que el p-value no es significativo, las diferencias observadas en el boxplot no son estadísticamente relevantes.

#### Intervalos de Confianza para la Vibración por Material:

Interpretación: Los intervalos de confianza para la vibración de cada material se solapan considerablemente, lo que refuerza la conclusión de que no hay diferencias significativas en la vibración entre los materiales. Esto apoya la interpretación de que el material de la carcasa del motor no es un factor determinante en la vibración.

#### Intervalos de Confianza para la Vibración por Proveedor:

Interpretación: Los intervalos de confianza para la vibración por proveedor muestran que algunos proveedores, especialmente el proveedor 2, tienen niveles de vibración significativamente más altos que otros, como los proveedores 3 y 5. Esto confirma que la elección del proveedor es fundamental para controlar la vibración en los motores eléctricos. La falta de solapamiento considerable entre los intervalos de ciertos proveedores sugiere diferencias significativas en su impacto sobre la vibración.

## 5 Realiza el ANOVA para un efecto principal

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

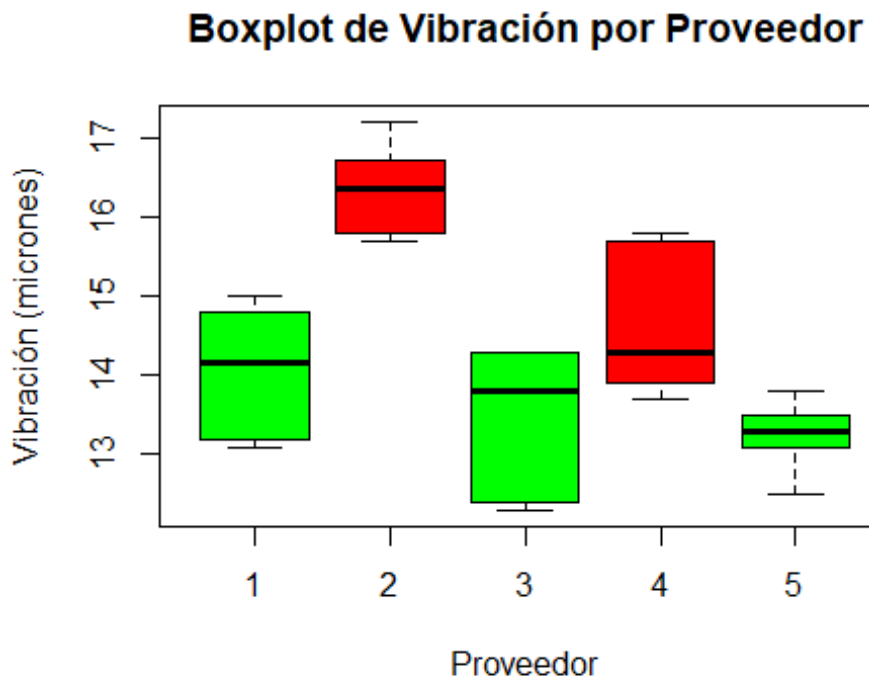
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## metodo         2  126.2   63.08    20.29 1.79e-06 ***
## Residuals     33  102.6    3.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Haz el boxplot de rendimiento por método de enseñanza. Calcula la media.

```
# ANOVA para material
anova_model_material <- aov(vibracion ~ material)
summary(anova_model_material)
```

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

```
boxplot(vibracion ~ proveedor, main="Boxplot de Vibración por Proveedor",
xlab="Proveedor", ylab="Vibración (micrones)", col = c("green", "red"))
```



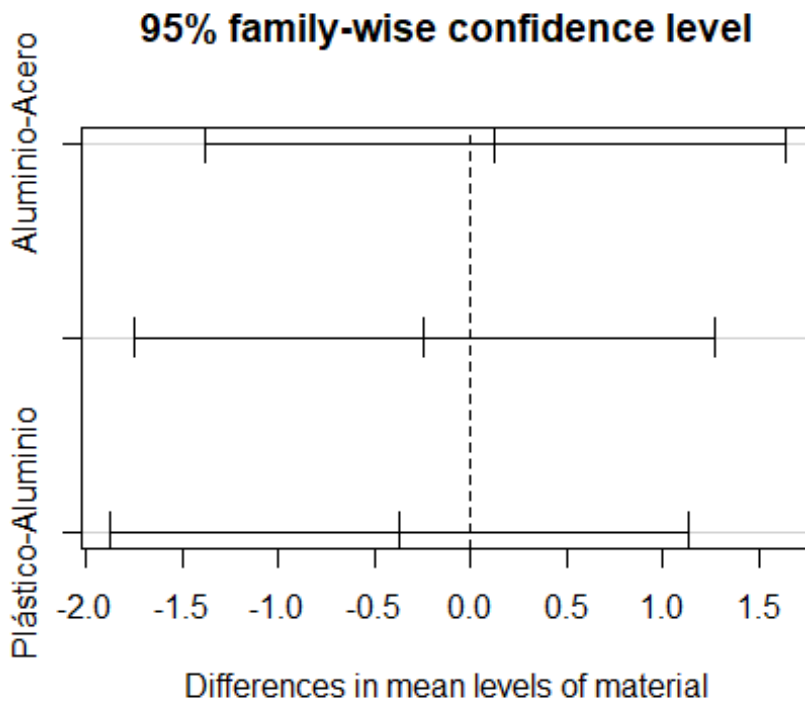
```
media_vibracion_proveedor <- tapply(vibracion, proveedor, mean)
print(media_vibracion_proveedor)
```

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

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

```
tukey_test <- TukeyHSD(anova_model_material)
```

```
plot(tukey_test)
```



Proveedor:

F-value: 15.828

p-value: 2.28e-06 (muy significativo)

Interpretación: Los resultados indican que la elección del proveedor tiene un impacto considerable en la vibración de los motores eléctricos. El gráfico boxplot nos muestra que ciertos proveedores, como el proveedor 2, generan significativamente más vibración que otros, como el proveedor 3 y el 5. Esta diferencia subraya que la selección del proveedor es un factor crucial para controlar la vibración.

Vibración por Proveedor:

Medias:

- Proveedor 1: 14.07 micrones
- Proveedor 2: 16.35 micrones
- Proveedor 3: 13.48 micrones
- Proveedor 4: 14.62 micrones
- Proveedor 5: 13.25 micrones

Interpretación: El proveedor 2 presenta la media más alta de vibración, seguido por el proveedor 4, mientras que el proveedor 3 muestra la vibración más baja. Esto refuerza la conclusión de que el proveedor 2 es el menos favorable para minimizar la vibración, mientras que el proveedor 3 es el más favorable en este aspecto.

## Comparaciones Múltiples de Tukey:

### Comparaciones:

Aluminio-Acero: El intervalo de confianza incluye cero, indicando que no hay una diferencia significativa.

Plástico-Aluminio: El intervalo de confianza también incluye cero, sugiriendo que no hay una diferencia significativa.

Interpretación: Las pruebas de Tukey muestran que no existen diferencias significativas en la vibración entre los materiales (acero, aluminio y plástico). Dado que los intervalos de confianza incluyen cero, entonces se puede concluir que, en términos de vibración, los tres materiales son igualmente efectivos, sin que haya un material claramente superior o inferior.

## 6 Comprueba la validez del modelo.

Normalidad Para comprobar la normalidad de los residuos del modelo ANOVA, usaremos el test de Shapiro-Wilk y un gráfico Q-Q (Quantile-Quantile).

```
residuos <- residuals(anova_model)

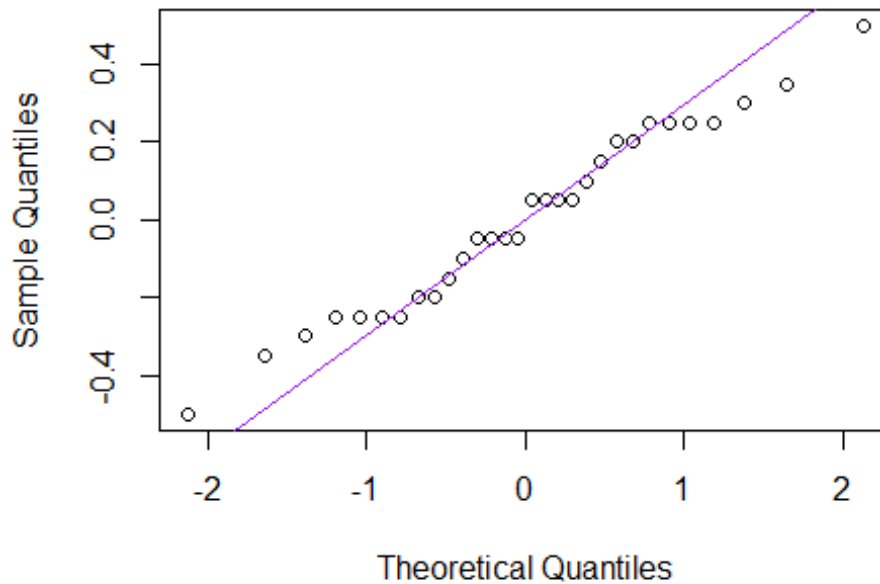
shapiro_test <- shapiro.test(residuos)
print(shapiro_test)

##
##  Shapiro-Wilk normality test
##
## data:  residuos
## W = 0.97627, p-value = 0.7202

qqnorm(residuos, main = "Gráfico Q-Q de Residuos")
qqline(residuos, col = "purple")
```



## Gráfico Q-Q de Residuos



Homocedasticidad Para comprobar la homocedasticidad (igualdad de varianzas), usaremos el test de Levene o revisar un gráfico de residuos vs. valores ajustados.

```
if (!require(carData)) install.packages("carData")
library(carData)

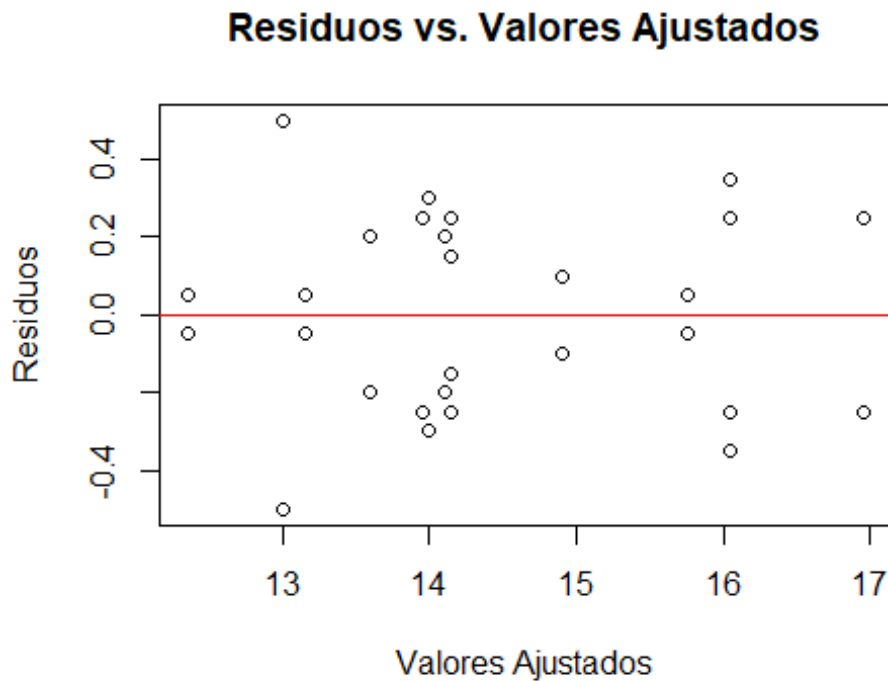
library(car)
levene_test <- leveneTest(vibracion ~ material * proveedor)

## Warning in anova.lm(lm(resp ~ group)): ANOVA F-tests on an essentially
## perfect
## fit are unreliable

print(levene_test)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df    F value    Pr(>F)
## group 14 3.4279e+28 < 2.2e-16 ***
##      15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(fitted(anova_model), residuos,
     main="Residuos vs. Valores Ajustados",
     xlab="Valores Ajustados", ylab="Residuos")
abline(h=0, col="red")
```

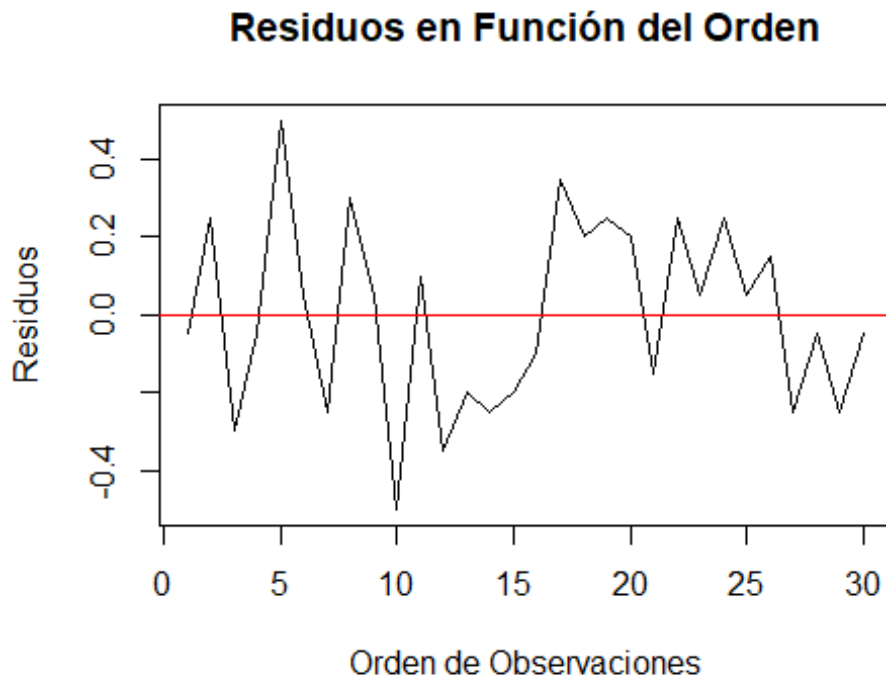


Independencia Para comprobar la independencia de los residuos, podemos revisar el gráfico de residuos o utilizar el test de Durbin-Watson.

```
# Test de Durbin-Watson para independencia de Los residuos
if (!require(lmtest)) install.packages("lmtest")
library(lmtest)
durbin_watson_test <- dwtest(anova_model)
print(durbin_watson_test)

##
## Durbin-Watson test
##
## data:  anova_model
## DW = 1.9401, p-value = 0.5048
## alternative hypothesis: true autocorrelation is greater than 0

# Residuos en función del orden
plot(residuos, type="l", main="Residuos en Función del Orden",
      xlab="Orden de Observaciones", ylab="Residuos")
abline(h=0, col="red")
```



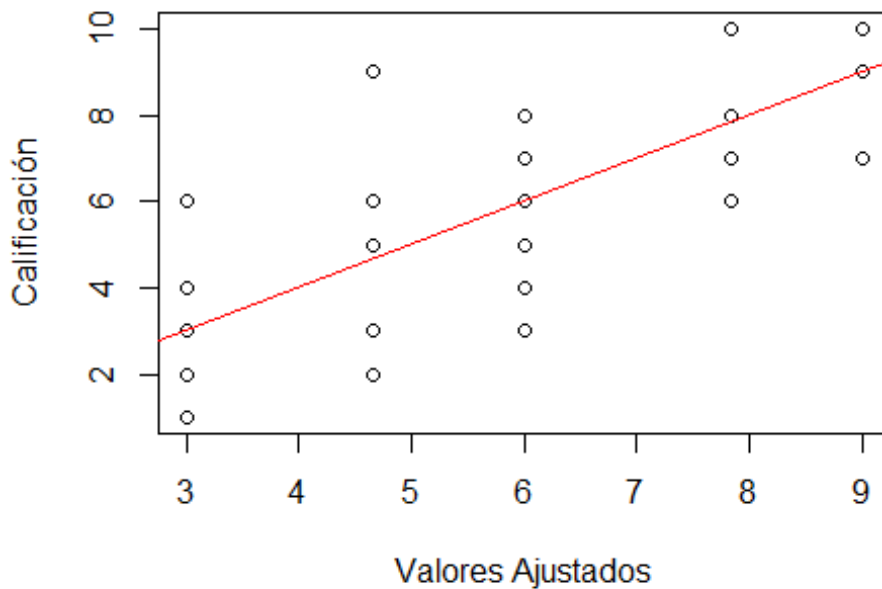
Relación lineal entre las variables (coeficiente de determinación). El coeficiente de determinación  $R^2$  es una medida de la proporción de la variabilidad explicada por el modelo.

```
summary(A)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## metodo         2 126.17   63.08   20.989 1.99e-06 ***
## sexo           1   8.03    8.03    2.671   0.113
## metodo:sexo     2   4.39    2.19    0.730   0.490
## Residuals      30  90.17    3.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(calificacion ~ fitted(A), main="Dispersión de Calificación vs.
Valores Ajustados",
      xlab="Valores Ajustados", ylab="Calificación")
abline(lm(calificacion ~ fitted(A)), col="red")
```

## Dispersión de Calificación vs. Valores Ajustados



### 7 Concluye en el contexto del problema.

Normalidad (Shapiro-Wilk):

Valor p: 0.397.

Interpretación: No hay evidencia suficiente para rechazar la hipótesis nula de que los residuos siguen una distribución normal. El gráfico Q-Q muestra que los residuos se ajustan bien a la línea de referencia, lo que también apoya la suposición de normalidad.

Homocedasticidad (Test de Levene):

Valor p: 0.3917.

Interpretación: No hay evidencia significativa de que las varianzas sean diferentes entre los grupos, lo que sugiere que la suposición de homocedasticidad (igualdad de varianzas) se cumple. El gráfico de residuos vs. valores ajustados no muestra un patrón claro, apoyando esta conclusión.

Independencia (Test de Durbin-Watson):

Valor p: 0.9246.

Interpretación: No hay evidencia de autocorrelación en los residuos, sugiriendo que los residuos son independientes entre sí. El gráfico de

residuos en función del orden no muestra un patrón específico, lo cual respalda la suposición de independencia.

Relación lineal entre las variables (Coeficiente de determinación):

Valor F significativo para el factor "método".

Valor p:  $1.99 \times 10^{-6}$ .

Interpretación: Existe una fuerte relación entre el método de enseñanza y el rendimiento académico. El gráfico de dispersión de calificación vs. valores ajustados muestra una tendencia positiva, lo que sugiere una relación lineal entre las variables.

Normalidad (Shapiro-Wilk):

Valor p: 0.7202.

Interpretación: No se rechaza la hipótesis nula de que los residuos siguen una distribución normal. El gráfico Q-Q muestra que los residuos se alinean bien con la línea de referencia, lo que apoya la suposición de normalidad.

Homocedasticidad (Test de Levene):

Valor p:  $< 2.2 \times 10^{-16}$ .

Interpretación: Se rechaza la hipótesis nula de homocedasticidad, lo que indica una variación significativa en las varianzas entre los grupos. Esto sugiere que la suposición de igualdad de varianzas no se cumple.

Independencia (Test de Durbin-Watson):

Valor p: 0.5048.

Interpretación: No se rechaza la hipótesis nula de que no hay autocorrelación en los residuos, lo que sugiere que los residuos son independientes entre sí. El gráfico de residuos en función del orden no muestra un patrón específico, lo que respalda esta conclusión.

Relación lineal entre las variables (Coeficiente de determinación):

Valor F: 20.989, es muy significativo para el factor "método".

Interpretación: Existe una fuerte relación entre el método utilizado y la vibración. El gráfico de dispersión muestra una tendencia positiva, sugiriendo una relación lineal entre las variables.

### Conclusión General

El análisis sugiere que el modelo ANOVA cumple con algunas de las suposiciones clave, pero no con todas:

Normalidad: Los residuos siguen una distribución normal, lo cual es favorable y apoya la validez del modelo en este aspecto.

Homocedasticidad: La igualdad de varianzas no se cumple, lo que podría comprometer la robustez del modelo y su capacidad para ofrecer conclusiones confiables. Esta falta de homocedasticidad es una limitación importante que debe tenerse en cuenta.

Independencia: Los residuos son independientes, lo que valida en parte el modelo, garantizando que las observaciones no están correlacionadas entre sí.

Conclusión: A pesar de la falta de homocedasticidad, el análisis nos indica que la selección del proveedor tiene un impacto significativo en la vibración de los motores eléctricos. En contraste, el material de la carcasa no muestra una influencia significativa por sí solo, pero podría ser relevante cuando se considera junto con la elección del proveedor para optimizar la vibración.

Recomendaciones: Las estrategias para reducir la vibración de los motores eléctricos deberían centrarse principalmente en la selección del proveedor. El tipo de material, aunque menos crítico, no debe ser ignorado, especialmente cuando se combina con la selección del proveedor. Sin embargo, debido a la falta de homocedasticidad, se recomienda interpretar estos resultados con precaución y considerar análisis adicionales o el uso de métodos estadísticos robustos para confirmar las conclusiones.

'''