

ANOVA

Catherine Rojas

2024-08-27

El rendimiento

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 metodología 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.

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

```
# Definir Los vectores con Los datos proporcionados
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))

# Convertir Las variables 'metodo' y 'sexo' en factores
metodo <- factor(metodo)
sexo <- factor(sexo)

# Crear un data frame con Los datos
data <- data.frame(Calificacion = calificacion, Metodo = metodo, Sexo =
sexo)

# Calcular La media del rendimiento por método de enseñanza
media_por_metodo <- aggregate(Calificacion ~ Metodo, data = data, FUN =
mean)
```

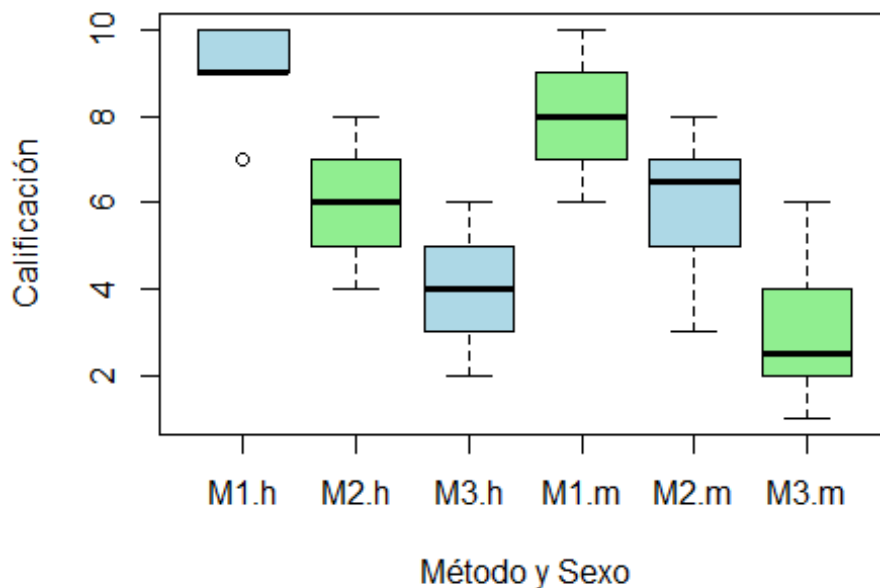
```
# Mostrar Las medias
print(media_por_metodo)

##   Metodo Calificacion
## 1      M1           8.5
## 2      M2           6.0
## 3      M3           3.5
```

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

```
# Crear el boxplot de Calificacion por Metodo y Sexo
boxplot(Calificacion ~ Metodo + Sexo,
        data = data,
        main = "Boxplot de Evaluación de Estudiantes por Método de
Enseñanza y Sexo",
        xlab = "Método y Sexo",
        ylab = "Calificación",
        col = c("lightblue", "lightgreen"))
```

de Evaluación de Estudiantes por Método de Enseñ



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

Las medias pr método de enseñanza, sugieren que el método 1 es el más efectivo, seguido del dos y tres en términos de rendimiento estudiantil.

Método 1 (M1): Tanto para chicos (M1.h) como para chicas (M1.m), las calificaciones tienden a ser altas, pero con un leve sesgo hacia mayores calificaciones en chicos. Además, hay un valor atípico en las calificaciones de los chicos.

Método 2 (M2): Las calificaciones son más homogéneas entre chicos y chicas.

Método 3 (M3): Las calificaciones son las más bajas en general, tanto para chicos (M3.h) como para chicas (M3.m). Las chicas parecen tener una ligera ventaja, pero el rendimiento en este método es bajo en ambos sexos.

4. Escribe tus conclusiones parciales

En términos generales, no parece haber una diferencia muy marcada en el rendimiento entre chicos y chicas para la mayoría de los métodos, aunque se observan algunas ligeras diferencias. Esto sugiere que tanto chicos como chicas responden de manera relativamente similar a los distintos métodos de enseñanza, con algunas variaciones menores.

Podría considerarse la posibilidad de enfocar más recursos y tiempo en el Método 1, ya que parece ser el que mejor apoya el rendimiento de los estudiantes. El Método 3 podría necesitar ser revisado o mejorado, ya que es el que menos impacto positivo tiene en el rendimiento.

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

1. Hipótesis para el primer factor (Método de enseñanza):

- **Hipótesis nula (H_{01}):** No hay diferencia en el rendimiento promedio entre los distintos métodos de enseñanza.

$$H_{01}: \tau_i = 0$$

, donde

$$\tau_i$$

representa el efecto del i-ésimo método de enseñanza.

- **Hipótesis alternativa (H_{11}):** Al menos uno de los métodos de enseñanza tiene un efecto diferente en el rendimiento.

$$H_{11}: \tau_i \neq 0$$

para al menos un i .

2. Hipótesis para el segundo factor (Género):

- **Hipótesis nula (H_{02}):** No hay diferencia en el rendimiento promedio entre chicos y chicas.

$$H_{02}: \alpha_j = 0$$

, donde α_j representa el efecto del j-ésimo nivel del factor género (hombre o mujer).

- **Hipótesis alternativa (H_{12}):** El género tiene un efecto en el rendimiento de los estudiantes.

$$H_{12}: \alpha_j \neq 0$$

para al menos un

j

.

3. Hipótesis para la interacción entre los factores (Método de enseñanza y Género):

- **Hipótesis nula (H_{03}):** No hay interacción entre el método de enseñanza y el género; es decir, el efecto del método de enseñanza es el mismo para chicos y chicas.

$$H_{03}: \tau_i \alpha_j = 0$$

, donde

$\tau_i \alpha_j$

representa la interacción entre el método de enseñanza y el género.

- **Hipótesis alternativa (H_{13}):** Existe una interacción entre el método de enseñanza y el género; es decir, el efecto del método de enseñanza depende del género.

$$H_{13}: \tau_i \alpha_j \neq 0$$

para al menos una combinación de

i

y

j

.

- **Primera hipótesis (Efecto del método de enseñanza):**

$$H_{01}: \tau_i = 0$$

$$H_{11}: \tau_i \neq 0$$

- **Segunda hipótesis (Efecto del género):**

$$H_{02}: \alpha_j = 0$$

$$H_{12}: \alpha_j \neq 0$$

- Tercera hipótesis (Interacción entre método de enseñanza y género):

$$H_{03}: \tau_i \alpha_j = 0$$

$$H_{13}: \tau_i \alpha_j \neq 0$$

6. Realiza el ANOVA para dos niveles con interacción

Realizar el ANOVA de dos factores con interacción

```
anova_A <- aov(Calificacion ~ Metodo * Sexo, data = data)
```

Mostrar el resumen del ANOVA

```
summary(anova_A)
```

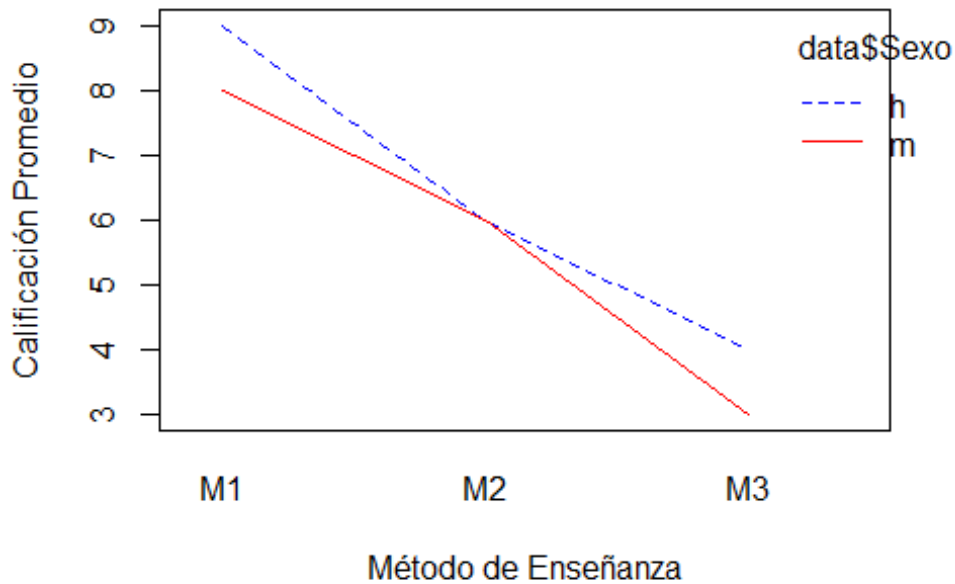
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Metodo         2    150   75.00  32.143 3.47e-08 ***
## Sexo           1     4    4.00   1.714  0.200
## Metodo:Sexo    2     2    1.00   0.429  0.655
## Residuals     30    70    2.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05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

7. Haz la gráfica de interacción de dos factores en ANOVA

Crear la gráfica de interacción

```
interaction.plot(data$Metodo, data$Sexo, data$Calificacion,
                 main = "Gráfica de Interacción: Método de Enseñanza y
Sexo",
                 xlab = "Método de Enseñanza",
                 ylab = "Calificación Promedio",
                 col = c("blue", "red"),
                 legend = TRUE)
```

Gráfica de Interacción: Método de Enseñanza y Se



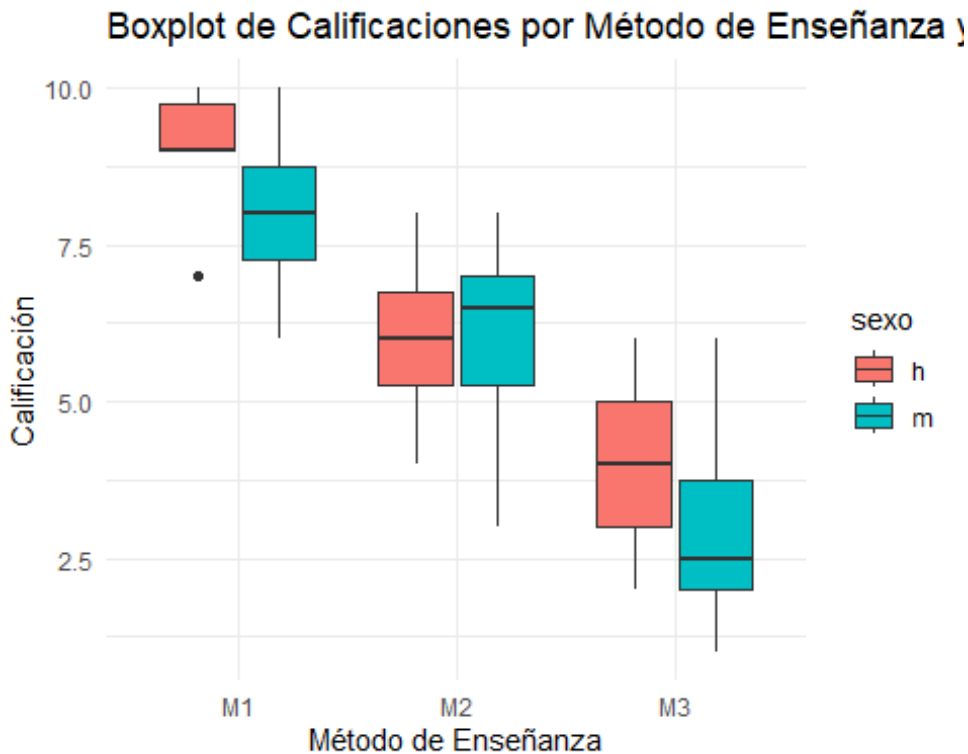
8. Haz el boxplot para visualizar la interacción de los factores

Generar el boxplot

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
ggplot(data, aes(x = metodo, y = calificacion, fill = sexo)) +  
  geom_boxplot() +  
  labs(title = "Boxplot de Calificaciones por Método de Enseñanza y  
Género",  
        x = "Método de Enseñanza", y = "Calificación") + theme_minimal()
```



9.

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

Resultados del ANOVA

Método: El valor p es $3.47e-08$, lo que es extremadamente pequeño y menor al nivel de significancia típico de 0.05. Esto indica que hay una diferencia significativa en las calificaciones promedio entre los diferentes métodos de enseñanza.

Sexo: El valor p es 0.200, lo que es mayor que 0.05, sugiriendo que no hay una diferencia significativa en las calificaciones promedio entre chicos y chicas.

Interacción Método: El valor p es 0.655, lo que también es mayor que 0.05, indicando que no hay evidencia significativa de una interacción entre el método de enseñanza y el sexo en cuanto al efecto sobre las calificaciones.

Gráfica de Interacción: Se observa que las líneas correspondientes a los chicos (línea azul punteada) y a las chicas (línea roja continua) son casi paralelas, lo que respalda el resultado del ANOVA de que no existe una interacción significativa entre el método de enseñanza y el sexo. Sin embargo, ambas líneas muestran una tendencia decreciente desde el Método 1 (M1) hasta el Método 3 (M3), lo que sugiere que ambos géneros tienen un mejor rendimiento con el Método 1 y un peor rendimiento con el Método 3. Este mismo comportamiento se observa en el boxplot.

10. Escribe tus conclusiones parciales

No se encontraron diferencias estadísticamente significativas entre los géneros en términos de calificaciones, lo que sugiere que, en general, tanto chicos como chicas

responden de manera similar a los métodos de enseñanza. y se recomendaría enfocarse en el Método 1 para maximizar el rendimiento de los estudiantes, ya que parece ser el más efectivo.

11. Realiza el ANOVA para dos niveles sin interacción.

```
# Realizar el ANOVA de dos factores sin interacción
anova_B <- aov(Calificacion ~ Metodo + Sexo, data = data)
```

```
# Mostrar el resumen del ANOVA
```

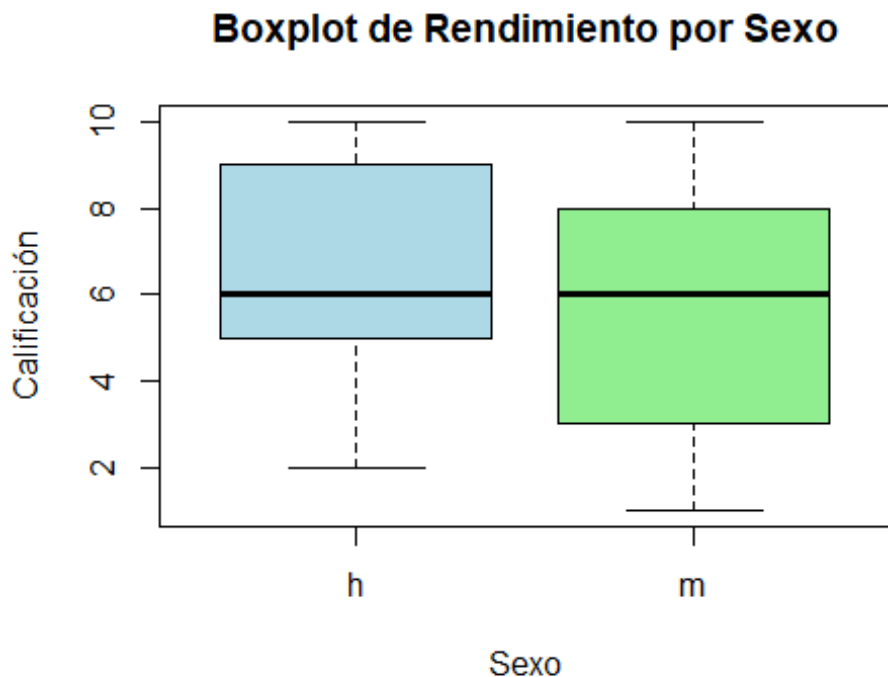
```
summary(anova_B)
```

```
##              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
```

12. Haz el boxplot de rendimiento por sexo.

```
# Crear el boxplot de rendimiento por sexo
```

```
boxplot(Calificacion ~ Sexo,
        data = data,
        main = "Boxplot de Rendimiento por Sexo",
        xlab = "Sexo",
        ylab = "Calificación",
        col = c("lightblue", "lightgreen"))
```



13. Calcula la media para el rendimiento por sexo y método.

```
# Calcular la media del rendimiento por sexo y método de enseñanza
media_por_sexo_metodo <- aggregate(Calificacion ~ Sexo + Metodo, data =
data, FUN = mean)
```

```
# Mostrar las medias
media_por_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
```

14. Haz los intervalos de confianza de rendimiento por sexo. Gráficalo

```
# Calcular la media y el intervalo de confianza por sexo
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##   filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
# Agrupamos por sexo y calculamos la media y el intervalo de confianza
```

```
data_summary <- data %>%
```

```
  group_by(Sexo) %>%
```

```
  summarise(
```

```
    mean_calificacion = mean(Calificacion),
```

```
    n = n(),
```

```
    se = sd(Calificacion) / sqrt(n),
```

```
    lower_ci = mean_calificacion - qt(1 - 0.05 / 2, n - 1) * se,
```

```
    upper_ci = mean_calificacion + qt(1 - 0.05 / 2, n - 1) * se
```

```
  )
```

```
print(data_summary)
```

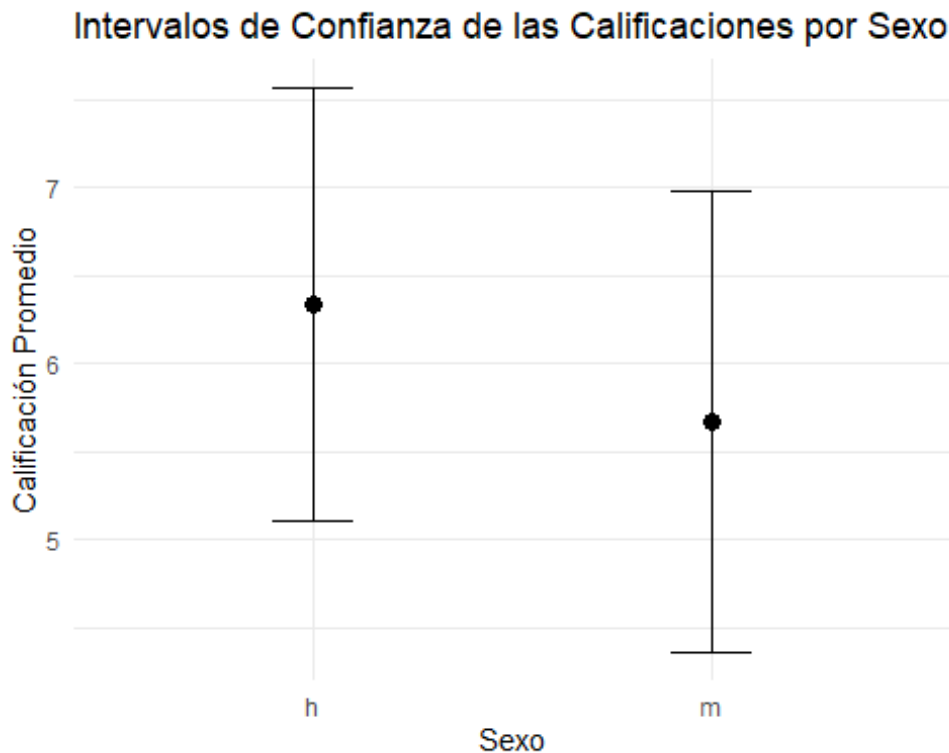
```
## # A tibble: 2 × 6
```

```
##   Sexo mean_calificacion      n      se lower_ci upper_ci
```

```
##   <fct>          <dbl> <int> <dbl>    <dbl>    <dbl>
```

```
## 1 h          6.33    18 0.583    5.10    7.56
## 2 m          5.67    18 0.621    4.36    6.98

# Graficar Los intervalos de confianza
ggplot(data_summary, aes(x = Sexo, y = mean_calificacion)) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = lower_ci, ymax = upper_ci), width = 0.2) +
  labs(title = "Intervalos de Confianza de las Calificaciones por Sexo",
       x = "Sexo", y = "Calificación Promedio") +
  theme_minimal()
```



15.

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

Anova El valor de p para el factor “Método” es extremadamente bajo ($1.5e-08$), lo que indica que hay una diferencia estadísticamente significativa entre los métodos de enseñanza. Es decir, las calificaciones varían significativamente dependiendo del método aplicado.

El valor de p para el factor “Sexo” es 0.192, lo que es mayor que el nivel de significancia comúnmente usado (0.05). Esto sugiere que no hay una diferencia estadísticamente significativa en el rendimiento entre hombres y mujeres en este contexto.

Intervalos Los intervalos de confianza para las calificaciones medias por sexo muestran lo siguiente:

Hombres (h): La calificación media es de 6.33 con un intervalo de confianza que va de 5.10 a 7.56.

Mujeres (m): La calificación media es de 5.67 con un intervalo de confianza que va de 4.36 a 6.98.

Estos intervalos se superponen, lo que refuerza la conclusión del ANOVA de que no hay una diferencia significativa en el rendimiento por sexo.

16. Escribe tus conclusiones parciales

El método de enseñanza tiene un impacto significativo en el rendimiento de ciencias naturales. No se observa una diferencia significativa en el rendimiento entre hombres y mujeres. Esto sugiere que, al menos en este grupo de estudiantes, las variaciones en el rendimiento son más atribuibles al método de enseñanza utilizado que al género del estudiante.

17. Realiza el ANOVA para un efecto principal

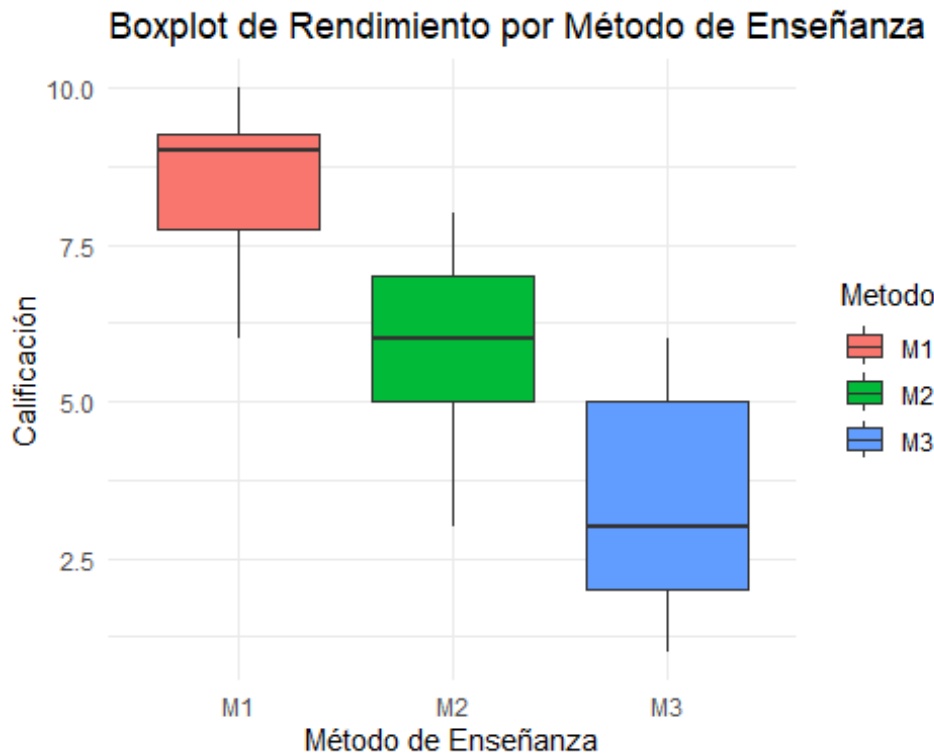
```
# Realizar el ANOVA considerando solo el efecto principal
anova_C <- aov(Calificacion ~ Metodo, data = data)
```

```
# Mostrar el resumen del ANOVA
summary(anova_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
```

18. Haz el boxplot de rendimiento por método de enseñanza.

```
# Graficar el boxplot del rendimiento por método de enseñanza
ggplot(data, aes(x = Metodo, y = Calificacion, fill = Metodo)) +
  geom_boxplot() +
  labs(title = "Boxplot de Rendimiento por Método de Enseñanza",
       x = "Método de Enseñanza",
       y = "Calificación") +
  theme_minimal()
```



19. Calcula la media.

```
# Calcular la media del rendimiento por método de enseñanza
media_por_metodo <- aggregate(Calificacion ~ Metodo, data = data, FUN =
mean)
```

```
# Mostrar las medias
```

```
media_por_metodo
```

```
##  Metodo Calificacion
## 1      M1           8.5
## 2      M2           6.0
## 3      M3           3.5
```

20. Haz los intervalos de confianza de rendimiento por método. Gráficelos

```
# Calcular la media y el intervalo de confianza por sexo
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
# Agrupamos por sexo y calculamos la media y el intervalo de confianza
```

```
data_summary <- data %>%
```

```
  group_by(Metodo) %>%
```

```
  summarise(
```

```
    mean_calificacion = mean(Calificacion),
```

```
    n = n(),
```

```
    se = sd(Calificacion) / sqrt(n),
```

```
    lower_ci = mean_calificacion - qt(1 - 0.05 / 2, n - 1) * se,
```

```

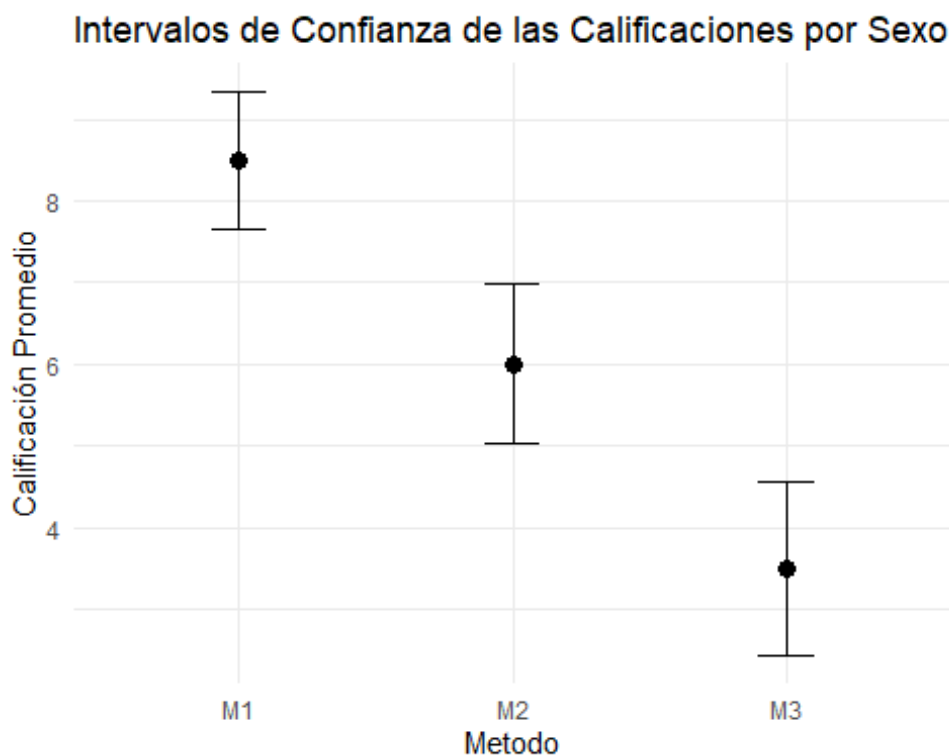
    upper_ci = mean_calificacion + qt(1 - 0.05 / 2, n - 1) * se
  )

print(data_summary)

## # A tibble: 3 × 6
##   Metodo mean_calificacion     n     se lower_ci upper_ci
##   <fct>         <dbl> <int> <dbl>   <dbl>   <dbl>
## 1 M1             8.5    12 0.379     7.66     9.34
## 2 M2             6     12 0.444     5.02     6.98
## 3 M3             3.5    12 0.485     2.43     4.57

# Graficar los intervalos de confianza
ggplot(data_summary, aes(x = Metodo, y = mean_calificacion)) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = lower_ci, ymax = upper_ci), width = 0.2) +
  labs(title = "Intervalos de Confianza de las Calificaciones por Sexo",
       x = "Metodo", y = "Calificación Promedio") +
  theme_minimal()

```



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

```

# Realizar la prueba de comparaciones múltiples de Tukey
tukey_result <- TukeyHSD(anova_C)

```

```
# Mostrar los resultados de Tukey
print(tukey_result)

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

# Graficar los intervalos de confianza de Tukey
plot(tukey_result, las = 1) # las = 1 rota las etiquetas en el eje y
```



22.

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

Boxplot y Turkey El método 1, tiene las calificaciones más altas, mientras que el método 3 tiene las calificaciones más bajas. Esto sugiere que el método que involucra la realización de experimentos (M1) es más efectivo para mejorar el rendimiento de los estudiantes, mientras que solo la explicación oral (M3) es menos efectiva.

23. Escribe tus conclusiones parciales

El método de enseñanza M1 (explicación oral y realización de experimentos) es el más efectivo para mejorar el rendimiento en ciencias naturales, seguido por M2

(explicación oral e imágenes) y, finalmente, M3 (solo explicación oral) que es el menos efectivo.

24. Comprueba la validez del modelo. Comprueba:

Normalidad

- (H_0)= Los datos siguen una distribución normal.
- (H_1)= Los datos no siguen una distribución normal

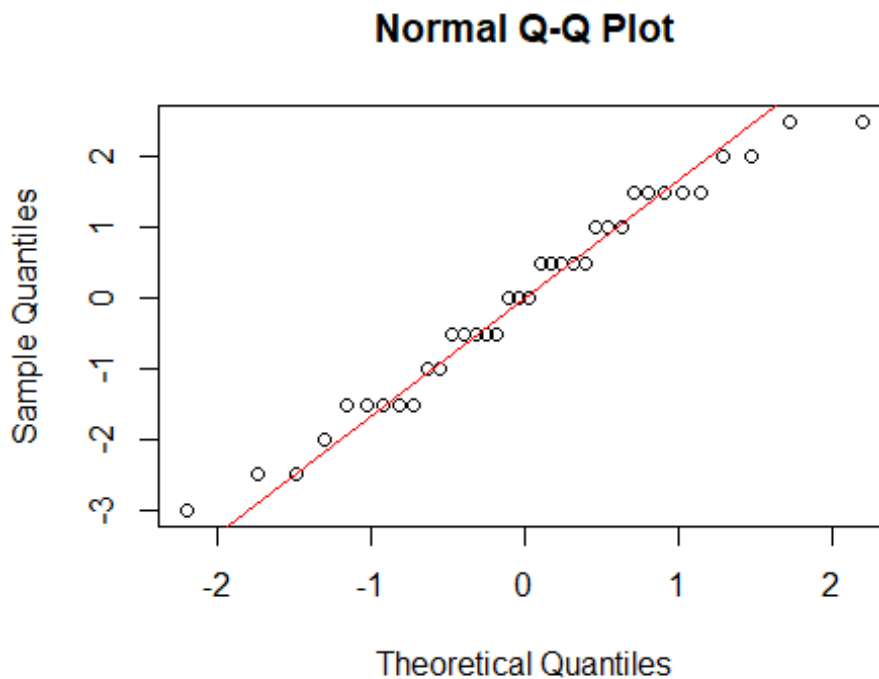
```
# Extraer los residuos del modelo ANOVA
```

```
residuos <- residuals(anova_C)
```

```
# Gráfico Q-Q para verificar la normalidad
```

```
qqnorm(residuos)
```

```
qqline(residuos, col = "red")
```



```
# Prueba de Shapiro-Wilk para normalidad
```

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

```
print(shapiro_test)
```

```
##
```

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

```
##
```

```
## data:  residuos
```

```
## W = 0.96734, p-value = 0.3573
```

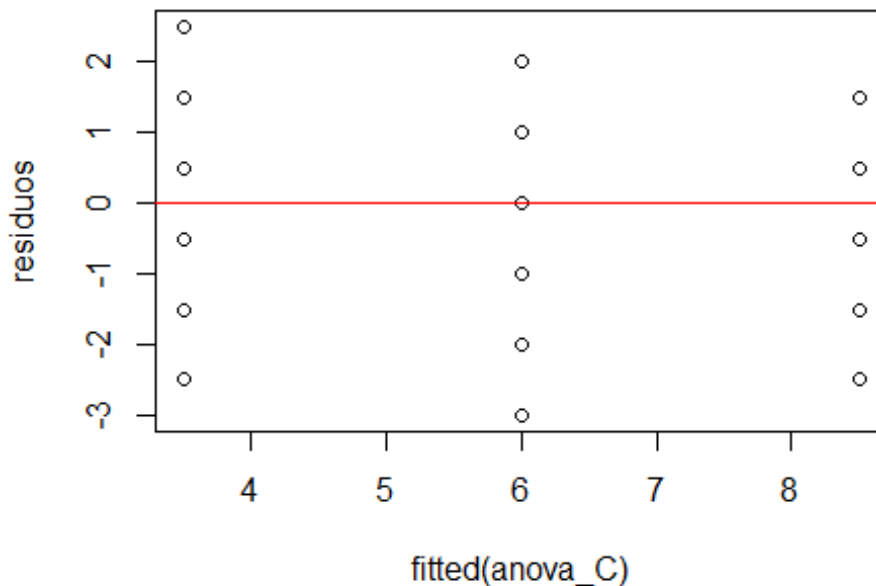
El p-valor es mayor que 0.05, lo que indica que no hay evidencia suficiente para rechazar la hipótesis nula de que los residuos siguen una distribución normal. Esto sugiere que el supuesto de normalidad se cumple para los residuos del modelo ANOVA.

Homocedasticidad

- (H_0)= La varianza de los residuos es constante en todos los niveles del predictor.
- (H_1)= La varianza de los residuos no es constante en todos los niveles del predictor.

Gráfico de residuos vs valores ajustados

```
plot(fitted(anova_C), residuos)  
abline(h = 0, col = "red")
```



Prueba de Bartlett para homocedasticidad

```
bartlett_test <- bartlett.test(Calificacion ~ Metodo, data = data)  
print(bartlett_test)
```

```
##
```

```
## Bartlett test of homogeneity of variances
```

```
##
```

```
## data: Calificacion by Metodo
```

```
## Bartlett's K-squared = 0.63268, df = 2, p-value = 0.7288
```


El p-valor es mayor que 0.05, lo que indica que no hay evidencia suficiente para rechazar la hipótesis nula de homocedasticidad (igualdad de varianzas). Esto sugiere que las varianzas entre los grupos son homogéneas, cumpliendo otro supuesto importante del ANOVA

Independencia

- (H_0) = No hay autocorrelación positiva en los residuos.
- (H_1) = Hay autocorrelación positiva en los residuos.

```
# Instalar y cargar el paquete lmtest si no está instalado
library(lmtest)

## Warning: package 'lmtest' was built under R version 4.3.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.3.3

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

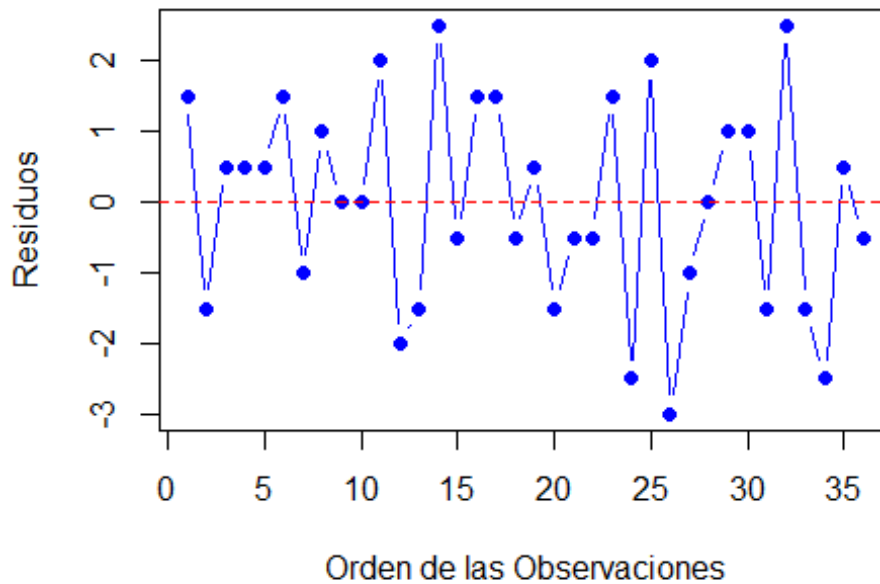
# Prueba de Durbin-Watson para independencia
durbin_watson_test <- dwtest(anova_C)
print(durbin_watson_test)

##
## Durbin-Watson test
##
## data: anova_C
## DW = 2.6974, p-value = 0.9706
## alternative hypothesis: true autocorrelation is greater than 0

# Crear una secuencia que represente el orden de las observaciones
orden <- 1:length(residuos)

# Graficar los residuos en función del orden de las observaciones
plot(orden, residuos, type = "b", pch = 19, col = "blue",
     xlab = "Orden de las Observaciones", ylab = "Residuos",
     main = "Gráfico de Independencia de Residuos")
abline(h = 0, col = "red", lty = 2)
```

Gráfico de Independencia de Residuos



El valor de Durbin-Watson es cercano a 2 y el p-valor es mayor que 0.05, lo que indica que no hay evidencia de autocorrelación en los residuos. Esto sugiere que los residuos son independientes, cumpliendo así el supuesto de independencia del ANOVA.

Relación lineal entre las variables (coeficiente de determinación).

```
# Calcular el R^2
suma_cuadrados_total <- sum((data$Calificacion -
mean(data$Calificacion))^2)
suma_cuadrados_modelo <- sum((fitted(anova_C) -
mean(data$Calificacion))^2)

R2 <- suma_cuadrados_modelo / suma_cuadrados_total

# Mostrar el R^2
print(paste("Coeficiente de Determinación (R^2):", R2))

## [1] "Coeficiente de Determinación (R^2): 0.663716814159292"
```

El coeficiente de determinación R^2 indica que aproximadamente el 66.37% de la variabilidad en las calificaciones se explica por el método de enseñanza utilizado. Esto sugiere que el método de enseñanza es un factor significativo y explica una parte considerable de las diferencias en el rendimiento de los estudiantes.

25. Concluye en el contexto del problema.

El método que incluye explicación oral y realización de experimentos (M1) es el más efectivo, seguido por el método que incluye explicación oral e imágenes (M2), y finalmente, el método menos efectivo es el de solo explicación oral (M3). Estos resultados son robustos y estadísticamente significativos, lo que apoya la recomendación de utilizar métodos de enseñanza más interactivos y prácticos para mejorar el rendimiento académico.

Vibración de motores

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 eleccionados al azar.

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

```
# Materiales de La carcasa (Factor A)
Material <- c(rep("Acero", 10), rep("Aluminio", 10), rep("Plastico", 10))

# Proveedores de cojinetes (Factor B)
Proveedor <- c("Proveedor1", "Proveedor2", "Proveedor3", "Proveedor4",
               "Proveedor5",
               "Proveedor1", "Proveedor2", "Proveedor3", "Proveedor4",
               "Proveedor5",
               "Proveedor1", "Proveedor2", "Proveedor3", "Proveedor4",
               "Proveedor5")

# Datos de vibración en micrones
Vibracion <- c(13.1, 16.3, 13.7, 15.7, 13.5, # Acero
               13.2, 15.8, 14.3, 15.8, 12.5, # Acero
               15.0, 15.7, 13.9, 13.7, 13.4, # Aluminio
               14.8, 16.4, 14.3, 14.2, 13.8, # Aluminio
               14.0, 17.2, 12.4, 14.4, 13.2, # Plástico
               14.3, 16.7, 12.3, 13.9, 13.1)

# Convertir las variables 'material' y 'proveedor' en factores
Material <- factor(Material)
Proveedor <- factor(Proveedor)

# Crear un data frame con Los datos
data <- data.frame(Vibracion = Vibracion, Material = Material, Proveedor
```

```

= Proveedor)
data

##      Vibracion Material  Proveedor
## 1      13.1      Acero Proveedor1
## 2      16.3      Acero Proveedor2
## 3      13.7      Acero Proveedor3
## 4      15.7      Acero Proveedor4
## 5      13.5      Acero Proveedor5
## 6      13.2      Acero Proveedor1
## 7      15.8      Acero Proveedor2
## 8      14.3      Acero Proveedor3
## 9      15.8      Acero Proveedor4
## 10     12.5      Acero Proveedor5
## 11     15.0 Aluminio Proveedor1
## 12     15.7 Aluminio Proveedor2
## 13     13.9 Aluminio Proveedor3
## 14     13.7 Aluminio Proveedor4
## 15     13.4 Aluminio Proveedor5
## 16     14.8 Aluminio Proveedor1
## 17     16.4 Aluminio Proveedor2
## 18     14.3 Aluminio Proveedor3
## 19     14.2 Aluminio Proveedor4
## 20     13.8 Aluminio Proveedor5
## 21     14.0 Plastico Proveedor1
## 22     17.2 Plastico Proveedor2
## 23     12.4 Plastico Proveedor3
## 24     14.4 Plastico Proveedor4
## 25     13.2 Plastico Proveedor5
## 26     14.3 Plastico Proveedor1
## 27     16.7 Plastico Proveedor2
## 28     12.3 Plastico Proveedor3
## 29     13.9 Plastico Proveedor4
## 30     13.1 Plastico Proveedor5

# Calcular La media de vibración por material
media_por_material <- aggregate(Vibracion ~ Material, data = data, mean)
print("Media de Vibración por Material:")

## [1] "Media de Vibración por Material:"

print(media_por_material)

##      Material Vibracion
## 1      Acero      14.39
## 2 Aluminio      14.52
## 3 Plastico      14.15

# Calcular La media de vibración por proveedor
media_por_proveedor <- aggregate(Vibracion ~ Proveedor, data = data,

```

```

mean)
print("Media de Vibración por Proveedor:")

## [1] "Media de Vibración por Proveedor:"

print(media_por_proveedor)

##      Proveedor Vibracion
## 1 Proveedor1   14.06667
## 2 Proveedor2   16.35000
## 3 Proveedor3   13.48333
## 4 Proveedor4   14.61667
## 5 Proveedor5   13.25000

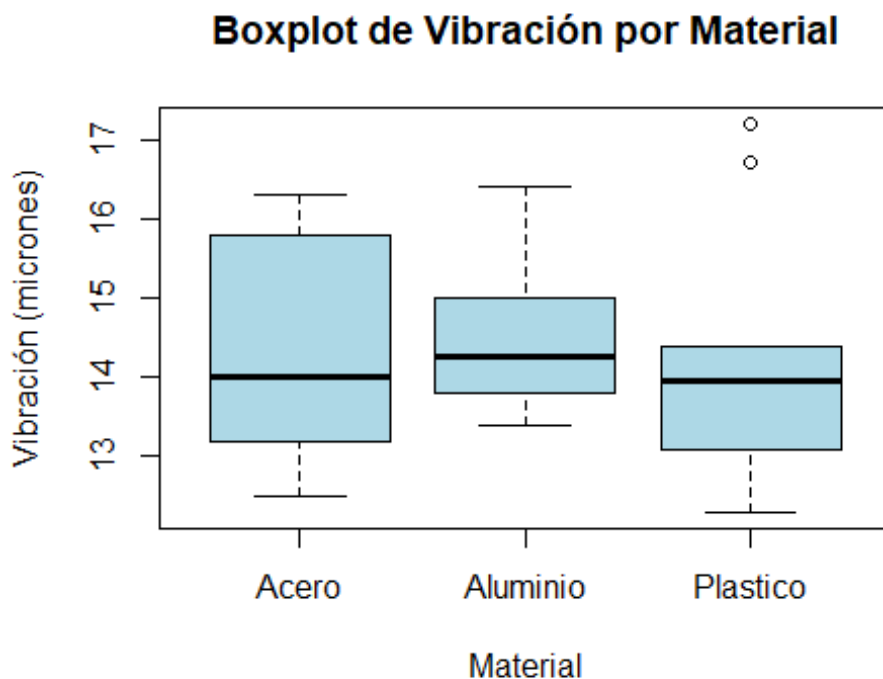
```

2. Haz el boxplot de la evaluación de las vibraciones por material y proveedor.

```

# Realizar el boxplot de las vibraciones por material
boxplot(Vibracion ~ Material, data = data, main = "Boxplot de Vibración
por Material",
        xlab = "Material", ylab = "Vibración (micrones)", col =
"lightblue")

```

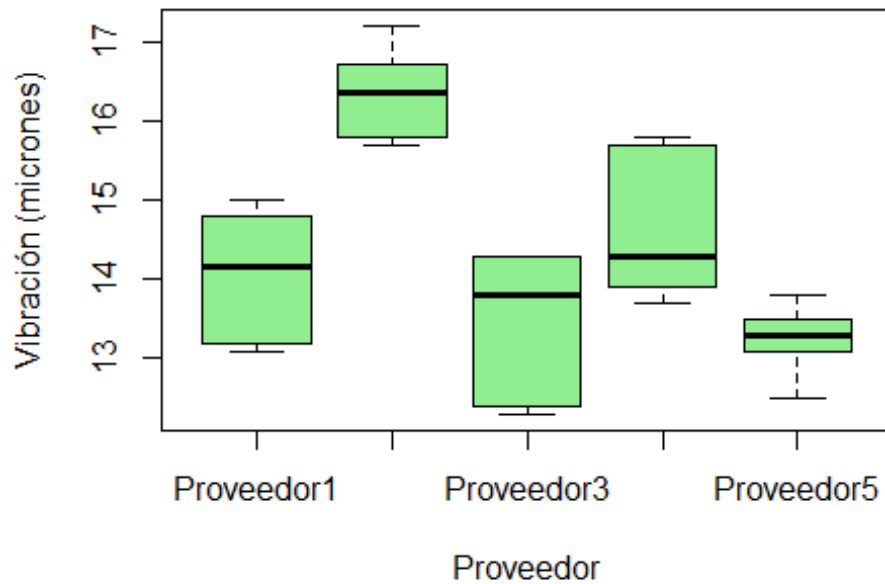


```

# Realizar el boxplot de las vibraciones por proveedor
boxplot(Vibracion ~ Proveedor, data = data, main = "Boxplot de Vibración
por Proveedor",
        xlab = "Proveedor", ylab = "Vibración (micrones)", col =
"lightgreen")

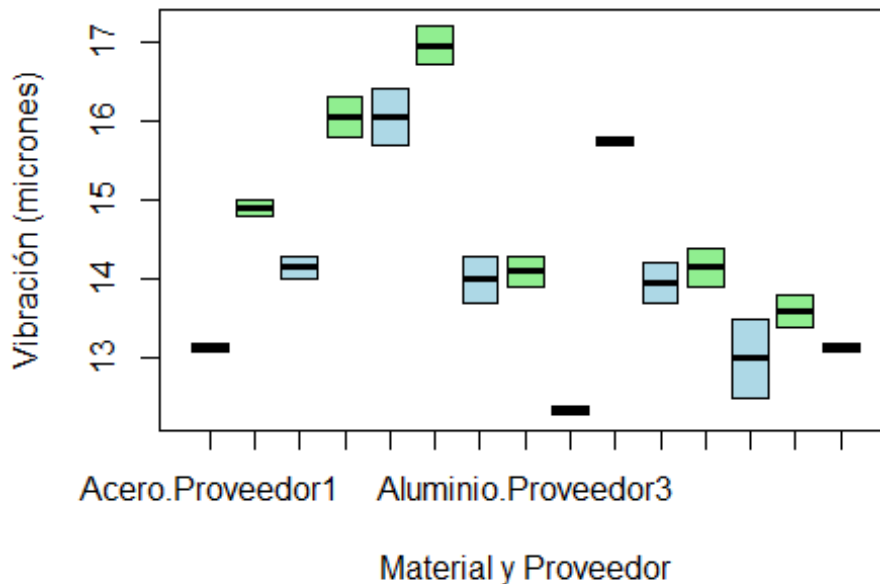
```

Boxplot de Vibración por Proveedor



```
# Realizar el boxplot de las vibraciones por material y proveedor
boxplot(Vibracion ~ Material + Proveedor,
  data = data,
  main = "Boxplot de Vibración por Material y Proveedor",
  xlab = "Material y Proveedor",
  ylab = "Vibración (micrones)",
  col = c("lightblue", "lightgreen"))
```

Boxplot de Vibración por Material y Proveedor



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

Las medias de vibración para los materiales de carcasa (Acero, Aluminio y Plástico) están bastante cerca unas de otras. Esto sugiere que el material de la carcasa tiene un impacto limitado en la vibración, con una variabilidad relativamente pequeña. Por otro lado, hay una mayor variabilidad en las medias de vibración entre los diferentes proveedores. Esto sugiere que el proveedor es un factor importante en la cantidad de vibración experimentada por los motores eléctricos. El Proveedor 2 parece ser una fuente de vibración más alta, lo que podría indicar una calidad inferior o un ajuste menos adecuado en los cojinetes.

En los boxplot podemos observar que existe una mayor variabilidad en las vibraciones con respecto a el proveedor que con respecto a el material. Estas diferencias, podrían indicar que algunos proveedores son mejores para cierto tipo de material.

4. Escribe tus conclusiones parciales

Estos resultados sugieren que el proveedor de los cojinetes tiene un mayor impacto en la vibración que el material de la carcasa. Sería recomendable realizar un análisis más detallado para el Proveedor 2, ya que en este caso, los cojinetes están asociados con niveles de vibración más altos.

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

1. Hipótesis sobre el Efecto del Material (Factor A):

- **Hipótesis Nula (H_{0A}):** El material de la carcasa del motor (Acero, Aluminio, Plástico) no tiene un efecto significativo sobre la vibración de los motores eléctricos. $H_{0A}: \mu_{\text{Acero}} = \mu_{\text{Aluminio}} = \mu_{\text{Plástico}}$
- **Hipótesis Alternativa (H_{1A}):** Al menos uno de los materiales de la carcasa del motor tiene un efecto significativo sobre la vibración de los motores eléctricos. H_{1A} : Al menos uno de los μ_{Material} es diferente.

2. Hipótesis sobre el Efecto del Proveedor (Factor B):

- **Hipótesis Nula (H_{0B}):** El proveedor de los cojinetes (Proveedor 1, Proveedor 2, Proveedor 3, Proveedor 4, Proveedor 5) no tiene un efecto significativo sobre la vibración de los motores eléctricos. $H_{0B}: \mu_{\text{Proveedor 1}} = \mu_{\text{Proveedor 2}} = \mu_{\text{Proveedor 3}} = \mu_{\text{Proveedor 4}} = \mu_{\text{Proveedor 5}}$
- **Hipótesis Alternativa (H_{1B}):** Al menos uno de los proveedores de cojinetes tiene un efecto significativo sobre la vibración de los motores eléctricos. H_{1B} : Al menos uno de los $\mu_{\text{Proveedor}}$ es diferente.

3. Hipótesis sobre la Interacción entre el Material y el Proveedor (Interacción A x B):

- **Hipótesis Nula (H_{0AB}):** No existe una interacción significativa entre el material de la carcasa y el proveedor de cojinetes en relación con la vibración de los motores eléctricos. H_{0AB} : No hay interacción significativa entre el Material y el Proveedor.
- **Hipótesis Alternativa (H_{1AB}):** Existe una interacción significativa entre el material de la carcasa y el proveedor de cojinetes en relación con la vibración de los motores eléctricos. H_{1AB} : Existe interacción significativa entre el Material y el Proveedor.

6. Realiza el ANOVA para dos niveles con interacción

Realizar el ANOVA de dos factores con interacción

```
anova_A <- aov(Vibracion ~ Material * Proveedor, data = data)
```

Mostrar el resumen del ANOVA

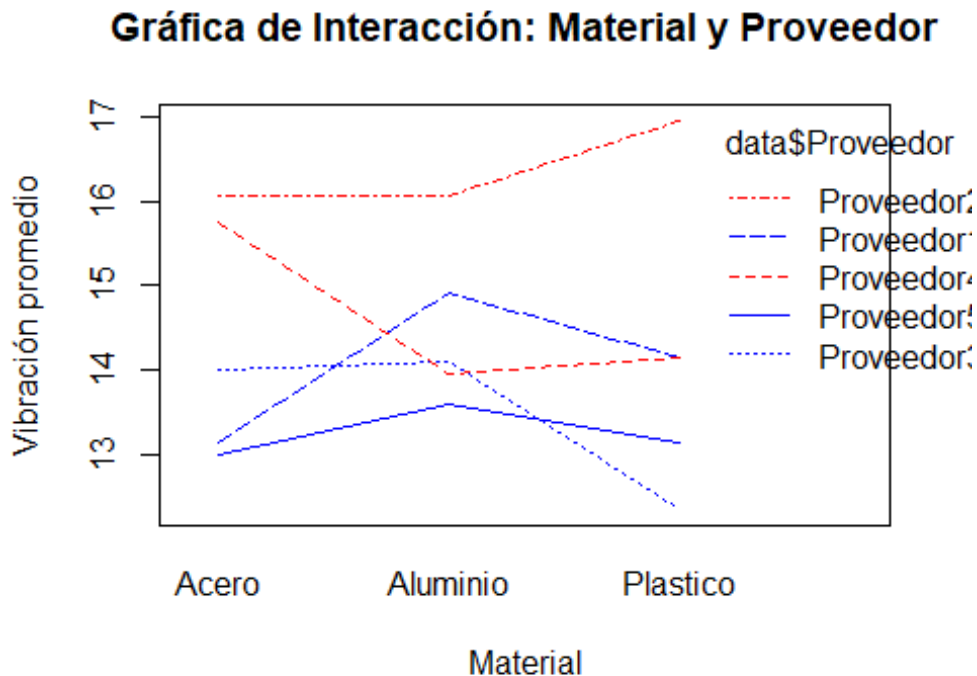
```
summary(anova_A)
```

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


7. Haz la gráfica de interacción de dos factores en ANOVA

Crear la gráfica de interacción

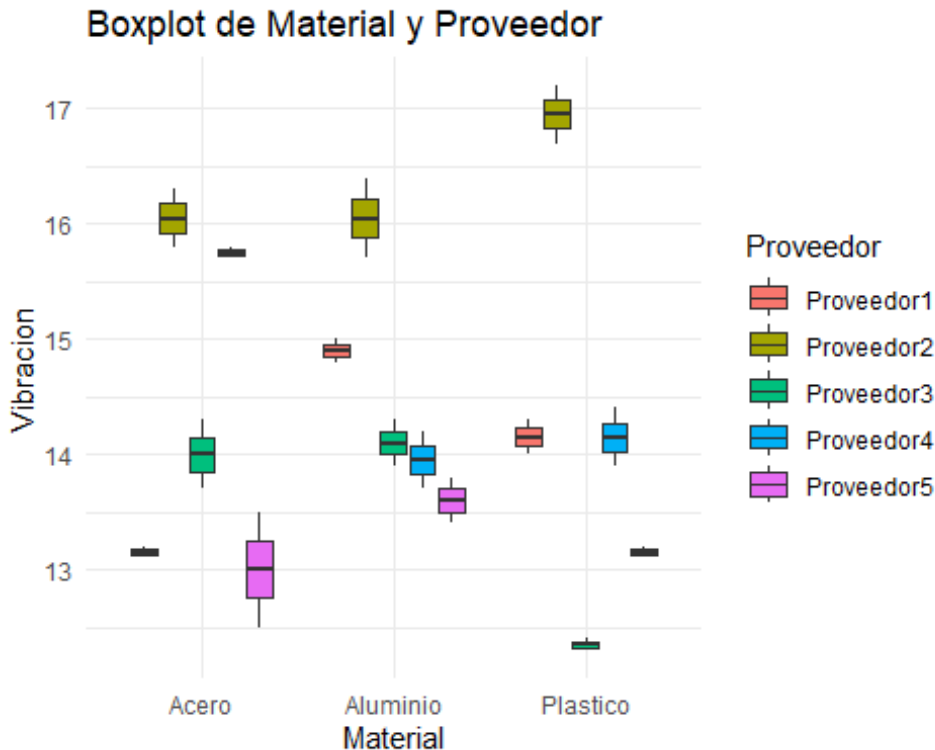
```
interaction.plot(data$Material, data$Proveedor, data$Vibracion,  
  main = "Gráfica de Interacción: Material y Proveedor",  
  xlab = "Material",  
  ylab = "Vibración promedio",  
  col = c("blue", "red"),  
  legend = TRUE)
```



8. Haz el boxplot para visualizar la interacción de los factores

Generar el boxplot

```
ggplot(data, aes(x = Material, y = Vibracion, fill = Proveedor)) +  
  geom_boxplot() +  
  labs(title = "Boxplot de Material y Proveedor",  
    x = "Material", y = "Vibracion") + theme_minimal()
```



9.

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

Resultados del ANOVA

Material El efecto del material de la carcasa es marginalmente no significativo, pues el p-valor es 0.0713. Utilizando el nivel de significancia tradicional de 0.05, no hay evidencia estadística suficiente para afirmar que el material tiene un impacto significativo en la vibración.

Proveedor El efecto del proveedor es altamente significativo con un p-valor $< 2e-10$). Esto indica que los diferentes proveedores de cojinetes tienen un impacto significativo en la vibración de los motores eléctricos. Este resultado coincide con la observación previa de que algunos proveedores parecen generar más vibración que otros.

Interacción Material: La interacción entre el material de la carcasa y el proveedor también es altamente significativa con un p-valor = $1.76e-05$. Esto sugiere que ciertos materiales pueden amplificar o reducir los efectos de ciertos proveedores en la vibración.

Gráfica de Interacción: Se observa que las líneas no son paralelas, lo cual es un indicio visual de una interacción significativa. En algunos casos, la vibración aumenta o disminuye notablemente según el material utilizado con un determinado proveedor. Nuevamente observamos que el Proveedor 2 parece generar más vibración con casi todos los materiales, mientras que otros proveedores tienen un rendimiento más variable dependiendo del material.

Boxplot

Se observa una gran dispersión en los valores de vibración, especialmente para algunas combinaciones de material y proveedor, lo que coincide con la significancia estadística de la interacción detectada por el ANOVA.

10. Escribe tus conclusiones parciales

El proveedor de cojinetes tiene un impacto significativo en la cantidad de vibración que experimentan los motores. Elegir un proveedor diferente podría reducir notablemente la vibración.

El material de la carcasa tiene un impacto menor en comparación con el proveedor.

Existe una interacción significativa entre el material y el proveedor, lo que significa que la combinación específica de ambos factores influye de manera importante en la vibración por lo que el modelo debe mantener ambos factores.

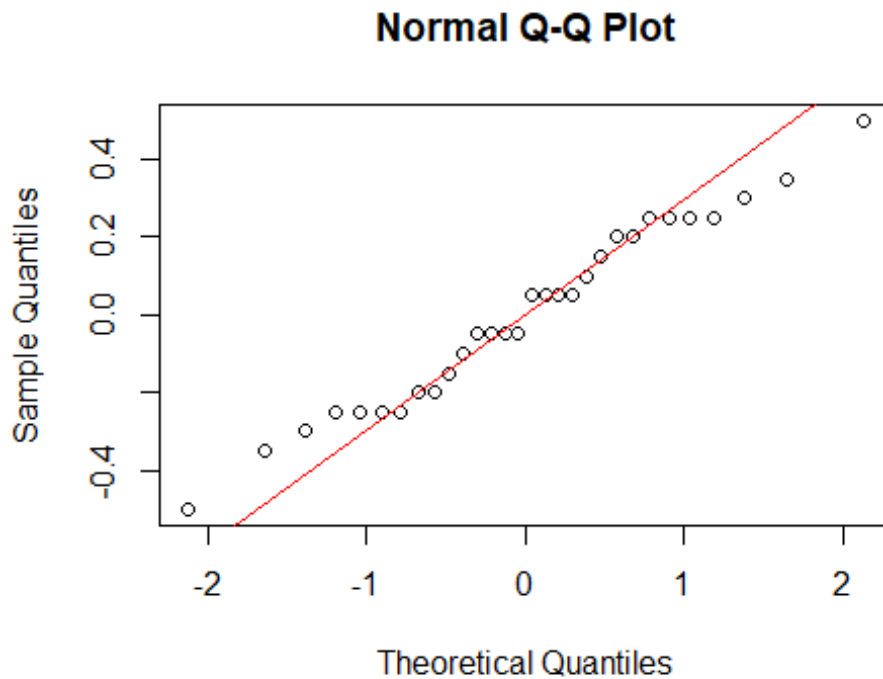
11. Comprueba la validez del modelo.

Normalidad

- (H_0) = Los datos siguen una distribución normal.
- (H_1) = Los datos no siguen una distribución normal

```
# Extraer Los residuos del modelo ANOVA
residuos <- residuals(anova_A)

# Gráfico Q-Q para verificar La normalidad
qqnorm(residuos)
qqline(residuos, col = "red")
```



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

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

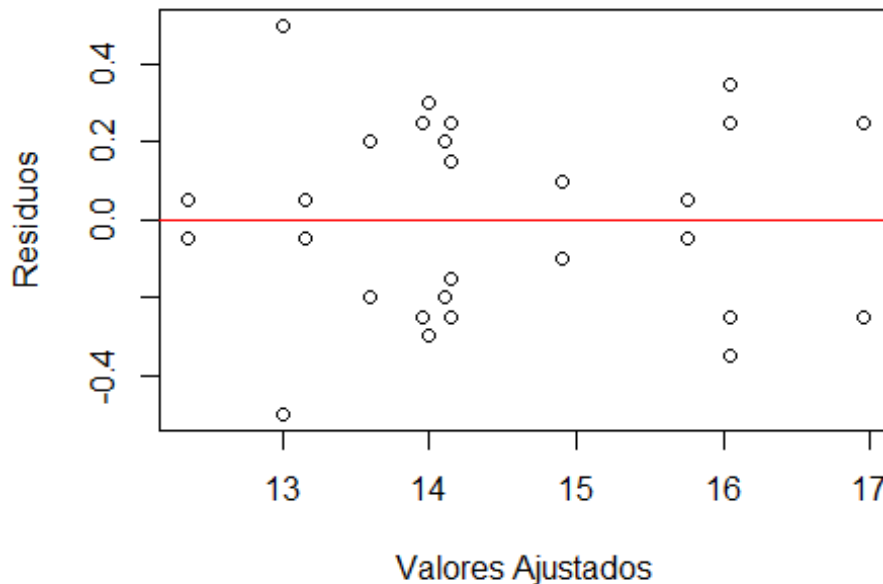
El p-valor es mayor que 0.05, lo que indica que no hay evidencia suficiente para rechazar la hipótesis nula de que los residuos siguen una distribución normal. Esto sugiere que el supuesto de normalidad se cumple para los residuos del modelo ANOVA.

Homocedasticidad

- (H_0)= La varianza de los residuos es constante en todos los niveles del predictor.
- (H_1)= La varianza de los residuos no es constante en todos los niveles del predictor.

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

Gráfico de Residuos vs Valores Ajustados



```
# Prueba de Bartlett para homocedasticidad
bartlett_test <- bartlett.test(residuos ~ interaction(Material,
Proveedor), data = data)
print(bartlett_test)

##
## Bartlett test of homogeneity of variances
##
## data:  residuos by interaction(Material, Proveedor)
## Bartlett's K-squared = 9.193, df = 14, p-value = 0.8185
```

El p-valor es mayor que 0.05, lo que indica que no hay evidencia suficiente para rechazar la hipótesis nula de homocedasticidad (igualdad de varianzas). Esto sugiere que las varianzas entre los grupos son homogéneas, cumpliendo otro supuesto importante del ANOVA

Independencia

- (H_0)= No hay autocorrelación positiva en los residuos.
- (H_1)= Hay autocorrelación positiva en los residuos.

```
# Instalar y cargar el paquete lmtest si no está instalado
library(lmtest)
```

```
# Prueba de Durbin-Watson para independencia
```

```

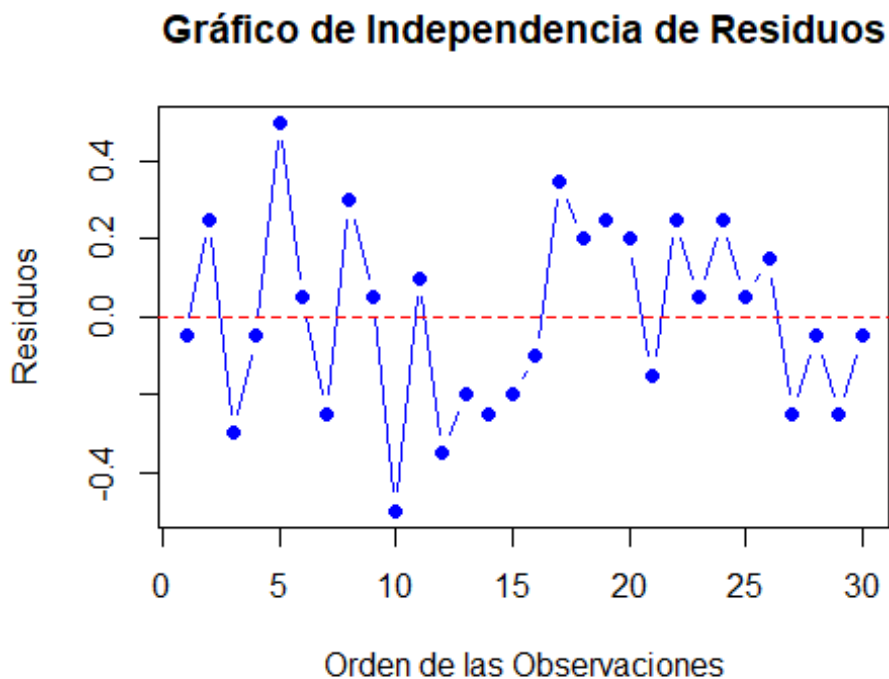
durbin_watson_test <- dwtest(anova_A)
print(durbin_watson_test)

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

# Crear una secuencia que represente el orden de las observaciones
orden <- 1:length(residuos)

# Graficar los residuos en función del orden de las observaciones
plot(orden, residuos, type = "b", pch = 19, col = "blue",
     xlab = "Orden de las Observaciones", ylab = "Residuos",
     main = "Gráfico de Independencia de Residuos")
abline(h = 0, col = "red", lty = 2)

```



El valor de Durbin-Watson es cercano a 2 y el p-valor es mayor que 0.05, lo que indica que no hay evidencia de autocorrelación en los residuos. Esto sugiere que los residuos son independientes, cumpliendo así el supuesto de independencia del ANOVA.

Relación lineal entre las variables (coeficiente de determinación).

```

# Ajustar el modelo ANOVA
anova_A <- aov(Vibracion ~ Material * Proveedor, data = data)

```

```

# Calcular el R^2
suma_cuadrados_total <- sum((data$Vibracion - mean(data$Vibracion))^2)
suma_cuadrados_modelo <- sum((fitted(anova_A) - mean(data$Vibracion))^2)

R2 <- suma_cuadrados_modelo / suma_cuadrados_total

# Mostrar el R^2
print(paste("Coeficiente de Determinación (R^2):", R2))

## [1] "Coeficiente de Determinación (R^2): 0.967031665394436"

```

El coeficiente de determinación R^2 indica que aproximadamente el 96.7% de la variabilidad en las Vibraciones se explica por el material y el proveedor. Esto sugiere que ambos factores son significativos y explica una parte considerable de las diferencias entre las vibraciones por material y proveedor.

25. Concluye en el contexto del problema.

El proveedor tiene un impacto significativo en la vibración de los motores eléctricos. Y la interacción entre material y proveedor también es significativa, lo que implica que la combinación específica de material de carcasa y proveedor de cojinetes influye en la vibración.

Para minimizar la vibración en los motores eléctricos, es crucial seleccionar cuidadosamente tanto el material de la carcasa como el proveedor de los cojinetes y así buscar asegurar la compatibilidad óptima y un rendimiento superior del motor.