Act9. Anova

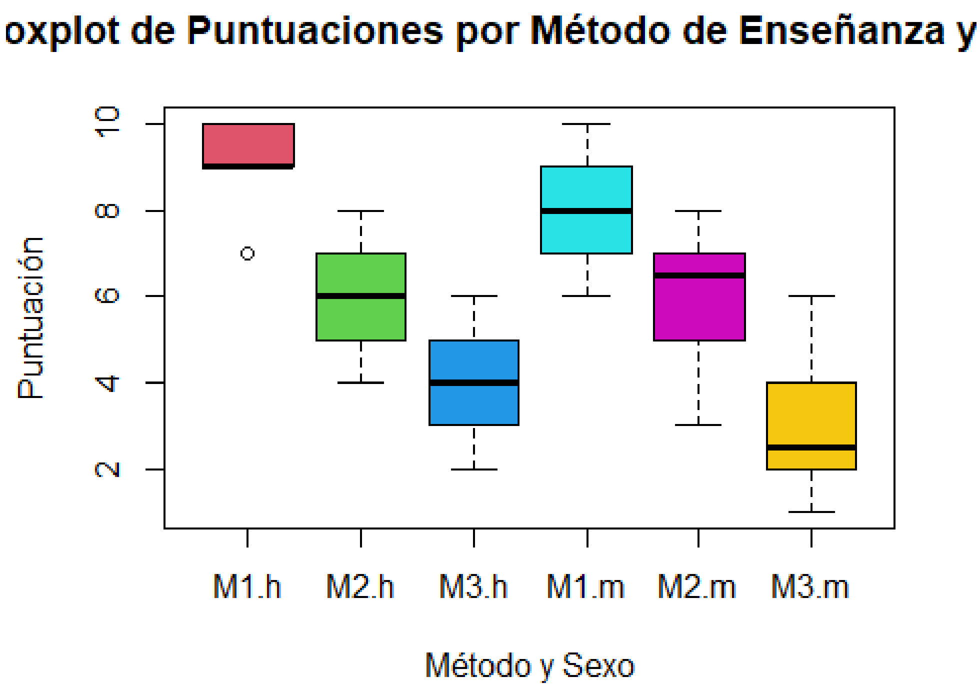
Andrés Villarreal González

2024-08-30

Problema 1 El rendimiento

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

```
Puntuacion=c(10,7,9,9,9,10,5,7,6,6,8,4,2,6,3,5,5,3,9,7,8,8,10,6,8,3,5,6,7
,7,2,6,2,1,4,3)
metodo=c(rep("M1",6),rep("M2",6),rep("M3",6),rep("M1",6),rep("M2",6),rep(
"M3",6))
sexo = c(rep("h", 18), rep("m",18))
Metodo = factor(metodo)
Sexo = factor(sexo)
rendimiento = data.frame(Puntuacion, Metodo, Sexo)

# Calcular la media por método de enseñanza
media_por_metodo <- aggregate(Puntuacion ~ Metodo, data = rendimiento,
mean)
print(media_por_metodo)

##   Metodo Puntuacion
## 1     M1        8.5
## 2     M2        6.0
## 3     M3        3.5

# Boxplot por método de enseñanza y sexo
boxplot(Puntuacion ~ Metodo * Sexo, data = rendimiento,
        col = c(2:7),
        main = "Boxplot de Puntuaciones por Método de Enseñanza y Sexo",
        xlab = "Método y Sexo", ylab = "Puntuación")
```



2. Escribe las hipotesis $H_0: \tau_i = 0$ (No hay efecto significativo del método de enseñanza en el rendimiento de los estudiantes). $H_1: \tau_i \neq 0$ (Hay al menos un método de enseñanza que tiene un efecto significativo en el rendimiento de los estudiantes).

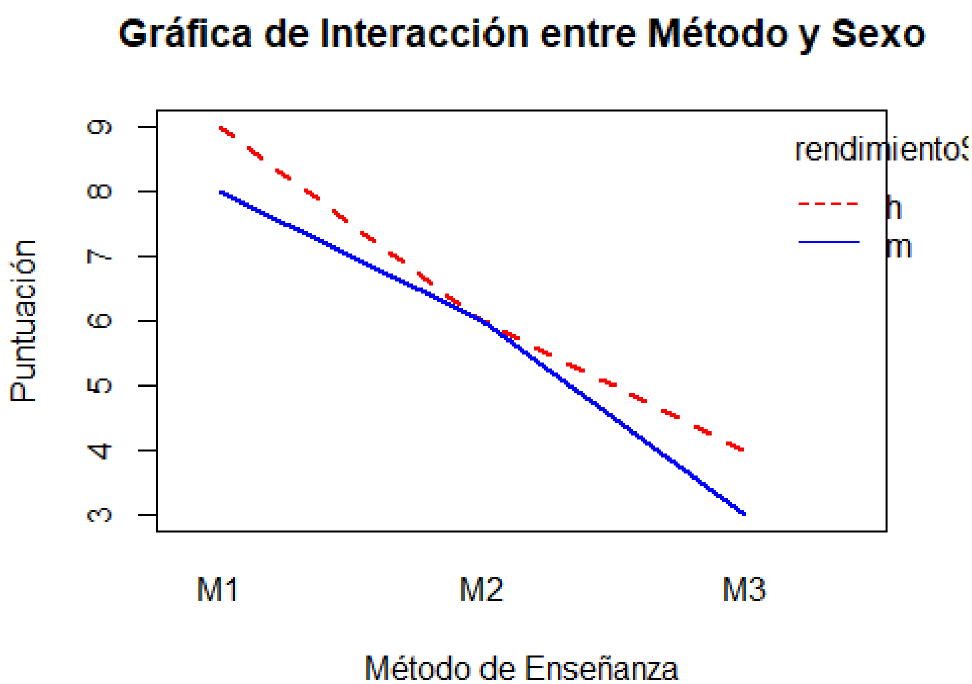
$H_0: \alpha_j = 0$ (No hay efecto significativo del género en el rendimiento de los estudiantes). $H_1: \alpha_j \neq 0$ (El género tiene un efecto significativo en el rendimiento de los estudiantes).

$H_0: \tau_i \alpha_j = 0$ (No hay interacción significativa entre el método de enseñanza y el género en el rendimiento de los estudiantes). $H_1: \tau_i \alpha_j \neq 0$ (Hay una interacción significativa entre el método de enseñanza y el género en el rendimiento de los estudiantes).

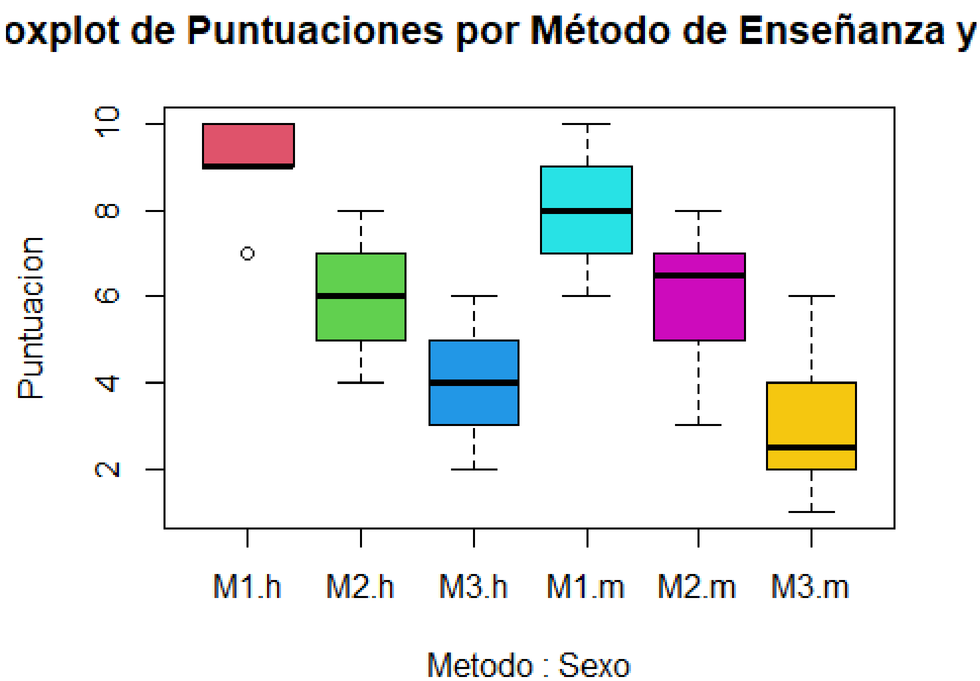
3. Realización del ANOVA para dos niveles con interacción

```r
# Realizar el ANOVA
anova_interaccion <- aov(Puntuacion ~ Metodo * Sexo, data = rendimiento)
summary(anova_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
```

```r
# Gráfica de interacción
interaction.plot(rendimiento$Metodo, rendimiento$Sexo,
rendimiento$Puntuacion,
                 col = c("red", "blue"), lwd = 2,
                 main = "Gráfica de Interacción entre Método y Sexo",
                 xlab = "Método de Enseñanza", ylab = "Puntuación")
```

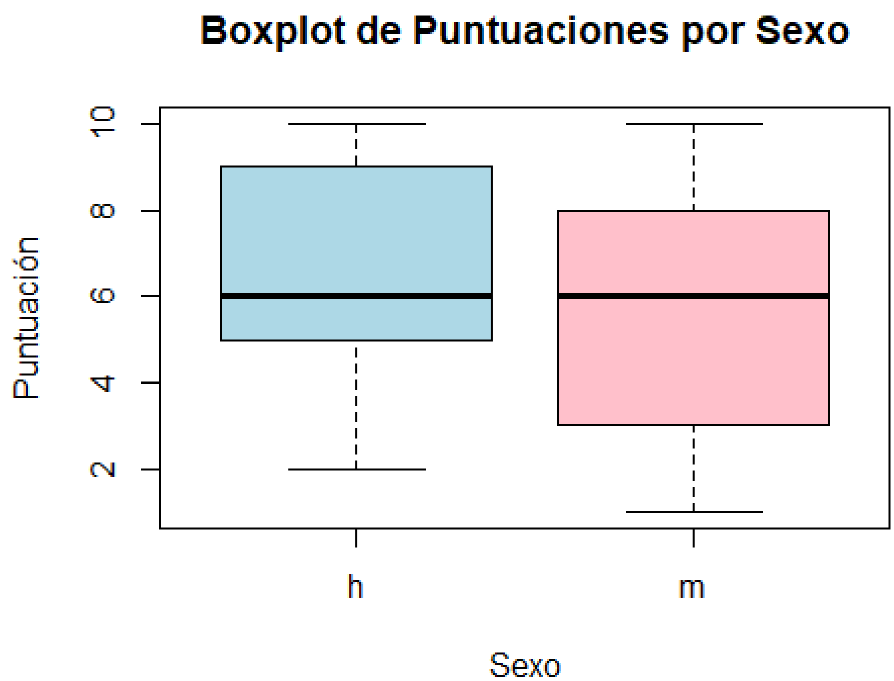



```r
# Boxplot para visualizar la interacción de los factores
boxplot(Puntuacion ~ Metodo * Sexo, data = rendimiento,
        col = c(2:7),
        main = "Boxplot de Puntuaciones por Método de Enseñanza y Sexo")
```



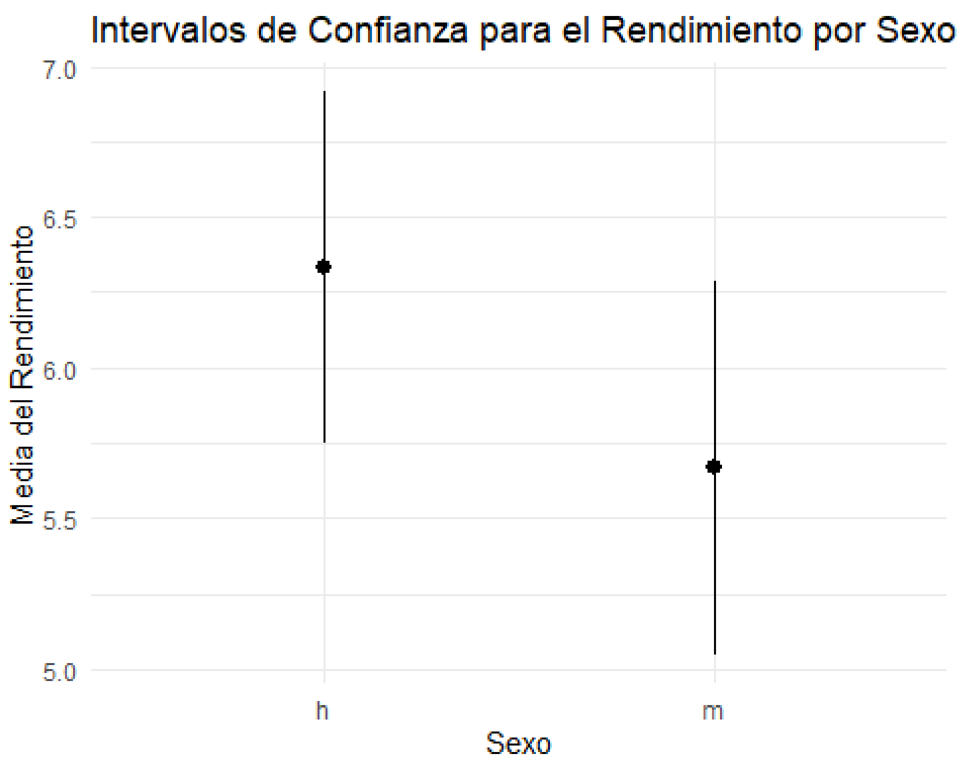
4. Realiza el ANOVA para dos niveles sin interacción.

```
# Realizar el ANOVA sin interacción
anova_sin_interaccion <- aov(Puntuacion ~ Metodo + Sexo, data =
rendimiento)
summary(anova_sin_interaccion)

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

# Boxplot de rendimiento por sexo
boxplot(Puntuacion ~ Sexo, data = rendimiento,
        col = c("lightblue", "pink"),
        main = "Boxplot de Puntuaciones por Sexo",
        xlab = "Sexo", ylab = "Puntuación")
```




```
# Media de rendimiento por sexo y método
media_por_sexo_y_metodo <- aggregate(Puntuacion ~ Sexo + Metodo, data =
rendimiento, mean)
print(media_por_sexo_y_metodo)

##   Sexo Metodo Puntuacion
## 1    h     M1          9
## 2    m     M1          8
## 3    h     M2          6
## 4    m     M2          6
## 5    h     M3          4
## 6    m     M3          3

# Cargar dplyr
library(dplyr)

##
## 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
```

```r
# Intervalos de confianza para el rendimiento por sexo
library(ggplot2)

# Crear un resumen de los datos con la media y el error estándar
summary_sex <- rendimiento %>%
  group_by(Sexo) %>%
  summarise(
    mean = mean(Puntuacion),
    sd = sd(Puntuacion),
    n = n(),
    se = sd/sqrt(n)
  )

# Gráfico de intervalos de confianza
ggplot(summary_sex, aes(x = Sexo, y = mean, ymin = mean-se, ymax =
mean+se)) +
  geom_pointrange() +
  labs(title = "Intervalos de Confianza para el Rendimiento por Sexo",
       x = "Sexo", y = "Media del Rendimiento") +
  theme_minimal()
```

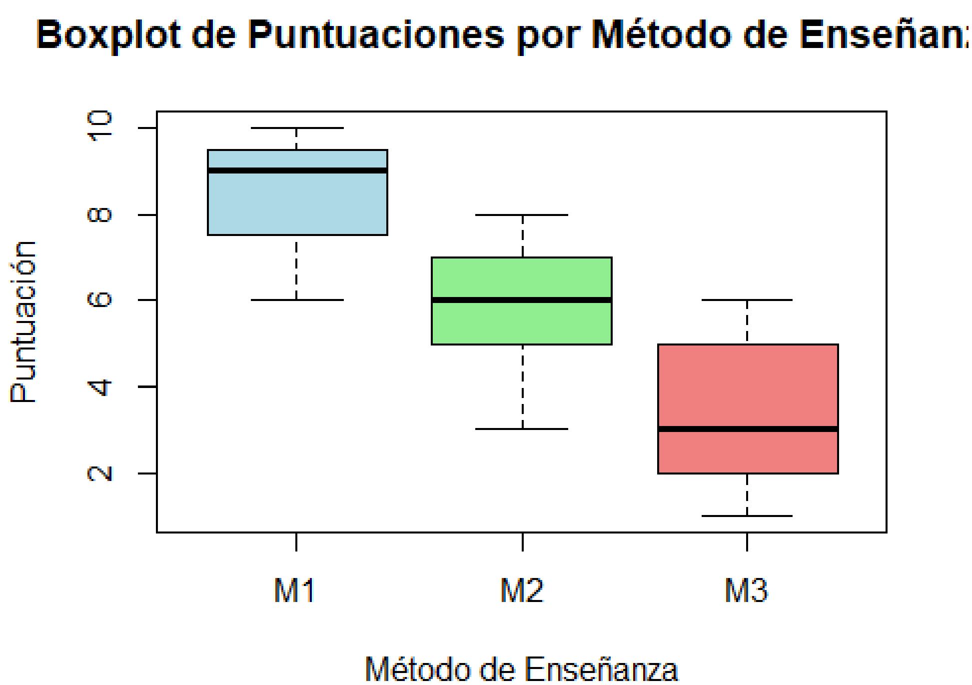


5. Realiza el ANOVA para un efecto principal

```r
# Realizar el ANOVA solo para el método de enseñanza
anova_metodo <- aov(Puntuacion ~ Metodo, data = rendimiento)
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
```

```r
# Boxplot de rendimiento por método de enseñanza
boxplot(Puntuacion ~ Metodo, data = rendimiento,
        col = c("lightblue", "lightgreen", "lightcoral"),
        main = "Boxplot de Puntuaciones por Método de Enseñanza",
        xlab = "Método de Enseñanza", ylab = "Puntuación")
```



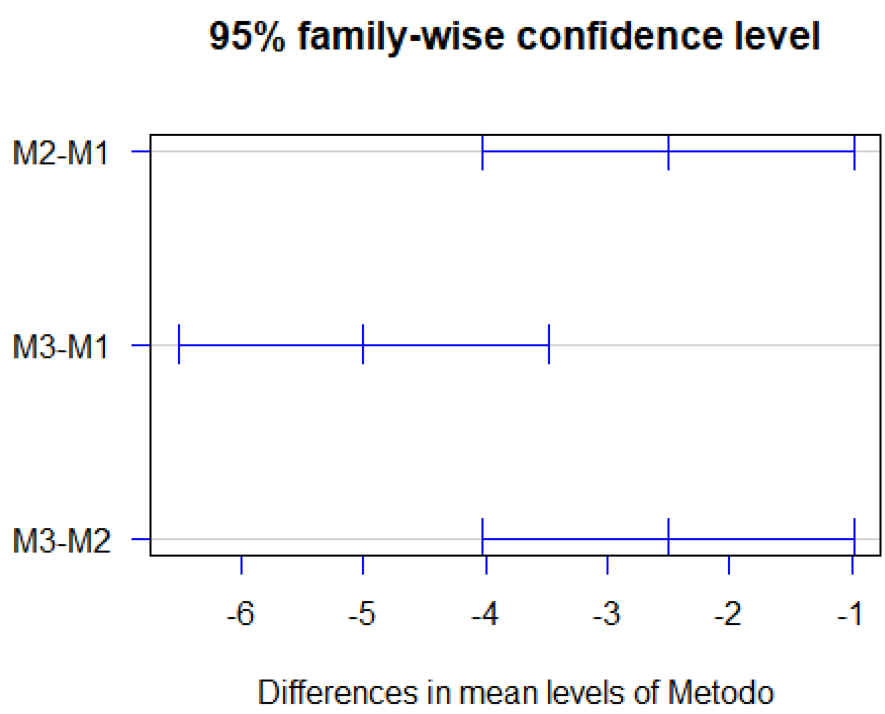

```r
# Calcular la media y los intervalos de confianza por método de enseñanza
summary_metodo <- rendimiento %>%
  group_by(Metodo) %>%
  summarise(
    mean = mean(Puntuacion),
    sd = sd(Puntuacion),
    n = n(),
    se = sd/sqrt(n),
    lim_inf = mean - qt(1 - 0.05 / 2, n - 1) * se,
    lim_sup = mean + qt(1 - 0.05 / 2, n - 1) * se
  )

print(summary_metodo)
```

```
## # A tibble: 3 × 7
##   Metodo  mean    sd     n    se lim_inf lim_sup
##   <fct>  <dbl> <dbl> <int> <dbl>   <dbl>   <dbl>
## 1 M1       8.5  1.31    12 0.379    7.66    9.34
## 2 M2       6    1.54    12 0.444    5.02    6.98
## 3 M3       3.5  1.68    12 0.485    2.43    4.57
```

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

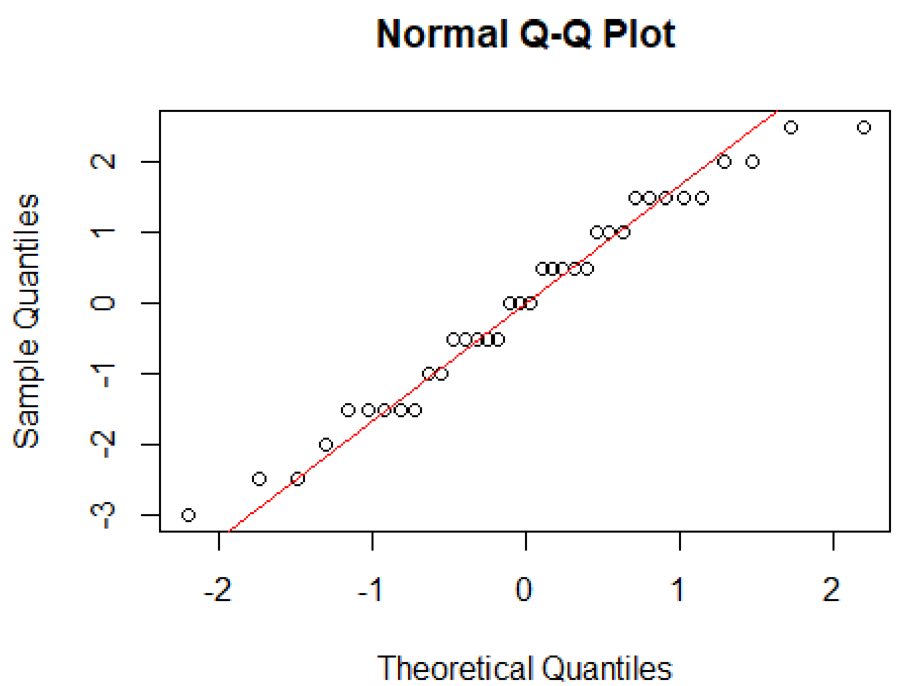
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Puntuacion ~ Metodo, data = rendimiento)
##
## $Metodo
##        diff       lwr        upr     p adj
## M2-M1 -2.5 -4.020241 -0.9797592 0.0008674
## M3-M1 -5.0 -6.520241 -3.4797592 0.0000000
## M3-M2 -2.5 -4.020241 -0.9797592 0.0008674
```

```r
# Graficar los intervalos de confianza de Tukey
plot(tukey_result, las = 1, col = "blue")
```



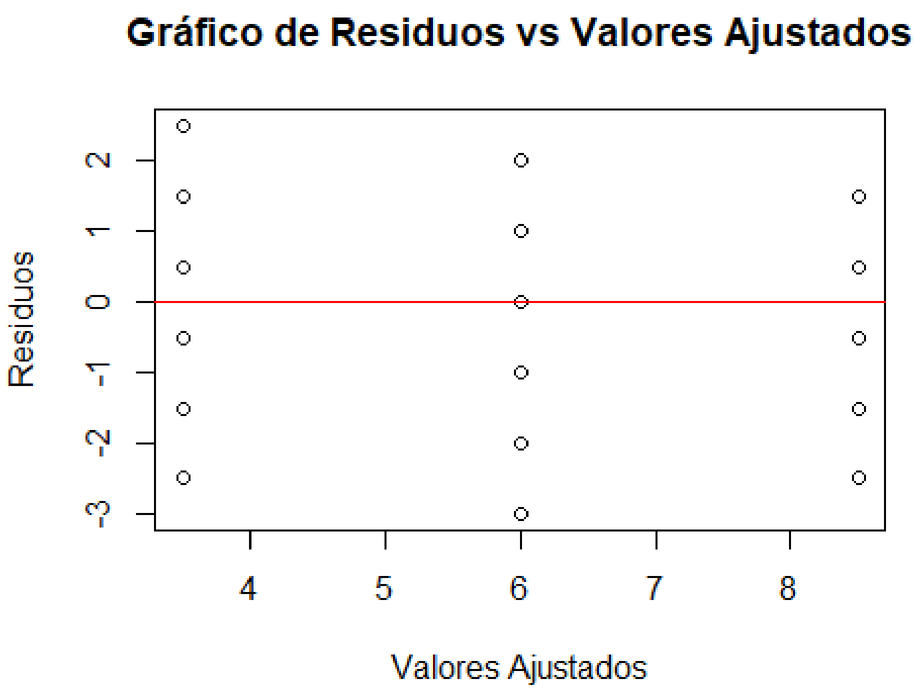
6. Compruebe la validez del modelo

```r
# Normalidad de los residuos
residuos <- residuals(anova_metodo)

# Q-Q plot
qqnorm(residuos)
qqline(residuos, col = "red")
```

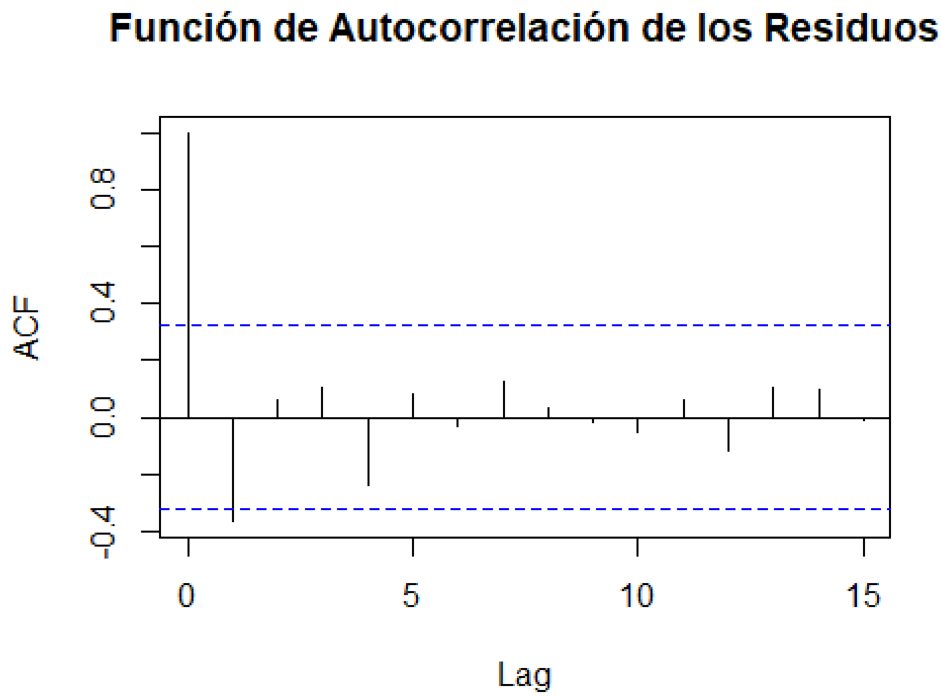



```r
# Prueba de Shapiro-Wilk
shapiro_test <- shapiro.test(residuos)
print(shapiro_test)

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

# Test de Bartlett para homocedasticidad
bartlett_test <- bartlett.test(Puntuacion ~ Metodo, data = rendimiento)
print(bartlett_test)

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

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




```r
# Gráfico de autocorrelación de residuos (si se sospecha de
independencia)
acf(residuos, main = "Función de Autocorrelación de los Residuos")
```




```r
# Coeficiente de determinación (R²)
summary_anova <- summary(anova_metodo)
r_squared <- summary_anova[[1]]["Metodo", "Sum Sq"] /
sum(summary_anova[[1]]["Sum Sq"])
print(paste("R²: ", round(r_squared, 2)))

## [1] "R²:  0.66"
```

7. Conclusiones Generales

Efecto del método de enseñanza: ANOVA con interacción: El análisis de varianza (ANOVA) indicó que el método de enseñanza tiene un efecto significativo en el rendimiento de los estudiantes. Esto sugiere que los diferentes métodos de enseñanza no son igualmente efectivos. Prueba de Tukey: La prueba de comparaciones múltiples de Tukey reveló que el Método 1 es significativamente mejor que los Métodos 2 y 3, y que el Método 2 es significativamente mejor que el Método 3.

Efecto del género: ANOVA sin interacción: No se encontró un efecto significativo del género en el rendimiento de los estudiantes ($p > 0.05$). Esto indica que, en general, no hay diferencias significativas entre el rendimiento de chicos y chicas. Interacción género-método: No se encontró una interacción significativa entre el método de enseñanza y el género ($p > 0.05$), lo que sugiere que el efecto del método de enseñanza es similar tanto para chicos como para chicas.

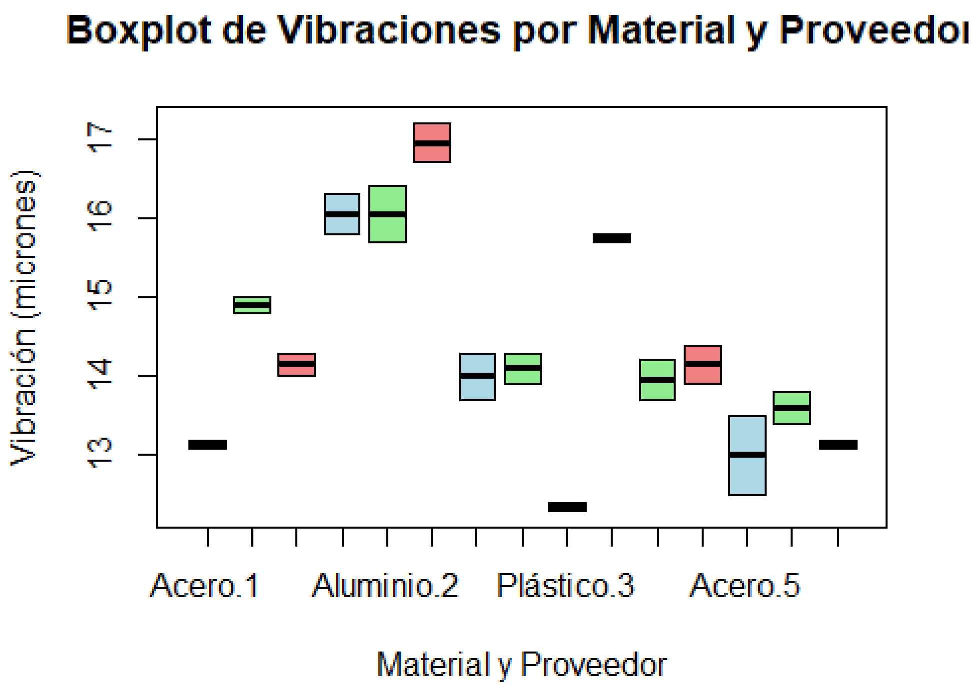
Validez del modelo: Normalidad de los residuos: La prueba de Shapiro-Wilk no rechazó la normalidad de los residuos ($p > 0.05$), lo que indica que el supuesto de normalidad se cumple. Homocedasticidad: La prueba de Bartlett mostró que las varianzas son homogéneas ($p > 0.05$), cumpliendo así el supuesto de homocedasticidad. Independencia de los residuos: El gráfico de autocorrelación de residuos no mostró patrones significativos de autocorrelación, sugiriendo que los residuos son independientes. Relación lineal: El coeficiente de determinación (R^2) fue aproximadamente 0.66, lo que indica que el 66% de la variabilidad en el rendimiento puede ser explicada por el método de enseñanza.

Problema 2 Vibracion de motores

1. Análisis exploratorio

```r
# Datos de vibración
vibracion <- data.frame(
  Material = rep(c("Acero", "Aluminio", "Plástico"), each = 10),
  Proveedor = rep(rep(1:5, each = 2), 3),
  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)
)

# Calcular la media por material y proveedor
media_por_material_proveedor <- aggregate(Vibracion ~ Material + 
Proveedor, data = vibracion, mean)
print(media_por_material_proveedor)

##    Material Proveedor Vibracion
## 1     Acero         1     13.15
## 2  Aluminio         1     14.90
## 3  Plástico         1     14.15
## 4     Acero         2     16.05
## 5  Aluminio         2     16.05
## 6  Plástico         2     16.95
## 7     Acero         3     14.00
## 8  Aluminio         3     14.10
## 9  Plástico         3     12.35
## 10    Acero         4     15.75
## 11 Aluminio         4     13.95
## 12 Plástico         4     14.15
## 13    Acero         5     13.00
## 14 Aluminio         5     13.60
## 15 Plástico         5     13.15

# Boxplot por material y proveedor
boxplot(Vibracion ~ Material * Proveedor, data = vibracion,
        col = c("lightblue", "lightgreen", "lightcoral"),
```

```
        main = "Boxplot de Vibraciones por Material y Proveedor",
        xlab = "Material y Proveedor", ylab = "Vibración (micrones)")
```



2. Las hipotesis

$H_0: \tau_i = 0$ No hay diferencias significativas en la vibración media entre los materiales.
$H_1: \tau_i \neq 0$ Hay diferencias significativas en la vibración media entre los materiales.

$H_0: \alpha_j = 0$ No hay diferencias significativas en la vibración media entre los proveedores. $H_1: \alpha_j \neq 0$ Hay diferencias significativas en la vibración media entre los proveedores.

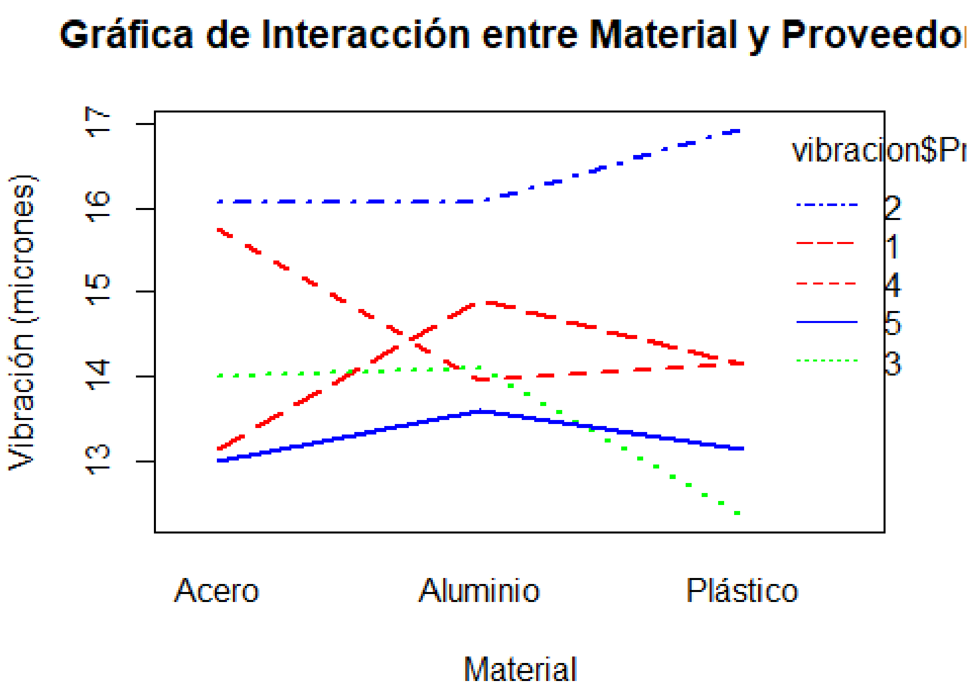
$H_0: \tau_i \alpha_j = 0$ No hay una interacción significativa entre el material y el proveedor en la vibración de los motores. $H_1: \tau_i \alpha_j \neq 0$ Hay una interacción significativa entre el material y el proveedor en la vibración de los motores.

3. Realizar el Anova

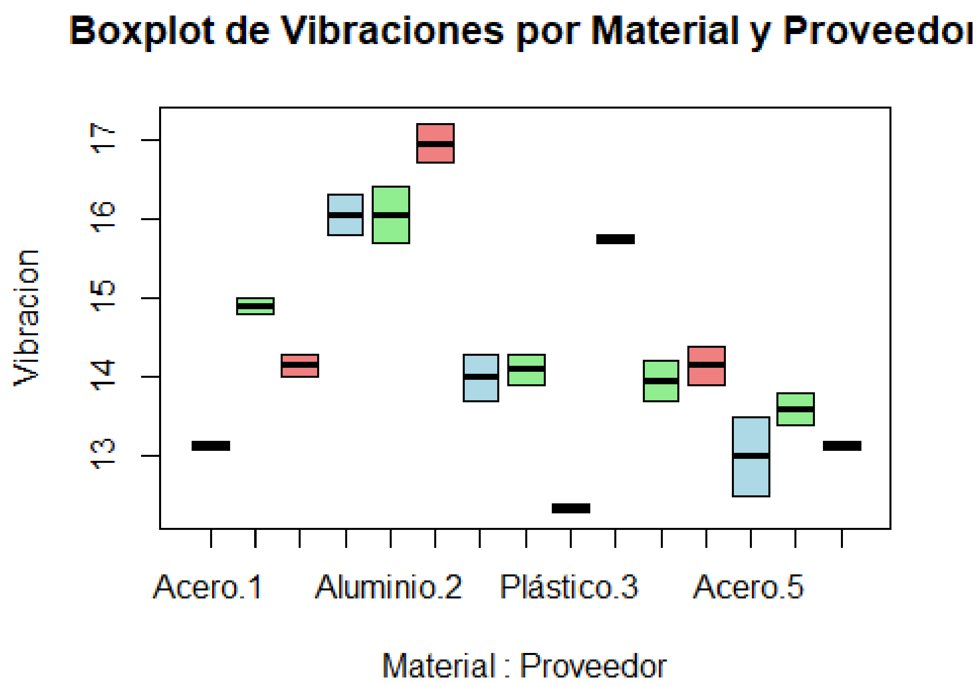
```
# Realizar el ANOVA
anova_interaccion <- aov(Vibracion ~ Material * Proveedor, data =
vibracion)
summary(anova_interaccion)

##                     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
```

```r
# Gráfica de interacción
interaction.plot(vibracion$Material, vibracion$Proveedor,
vibracion$Vibracion,
                 col = c("red", "blue", "green"), lwd = 2,
                 main = "Gráfica de Interacción entre Material y
Proveedor",
                 xlab = "Material", ylab = "Vibración (micrones)")
```

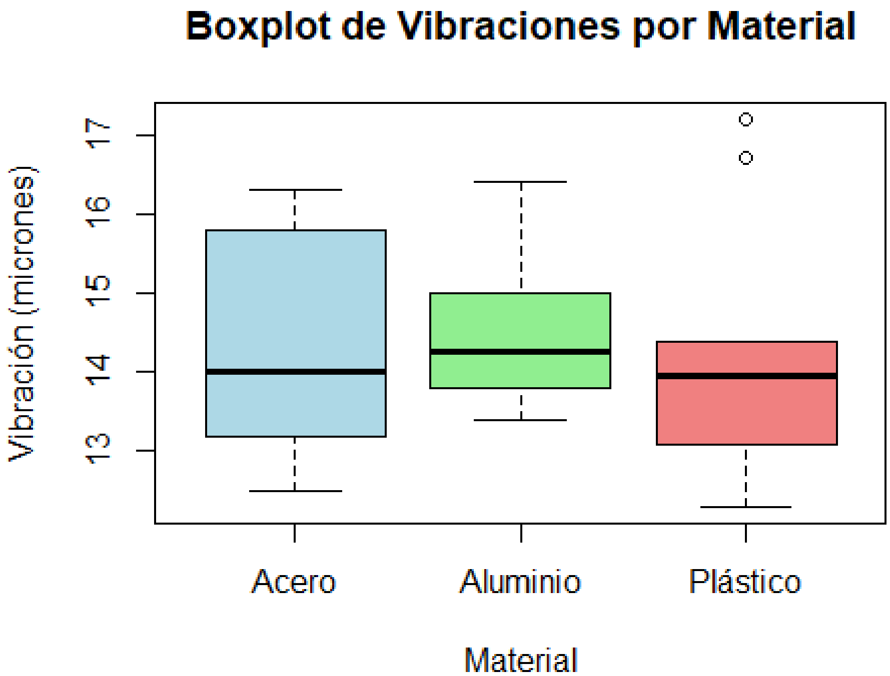



```r
# Boxplot para visualizar la interacción de los factores
boxplot(Vibracion ~ Material * Proveedor, data = vibracion,
        col = c("lightblue", "lightgreen", "lightcoral"),
        main = "Boxplot de Vibraciones por Material y Proveedor")
```

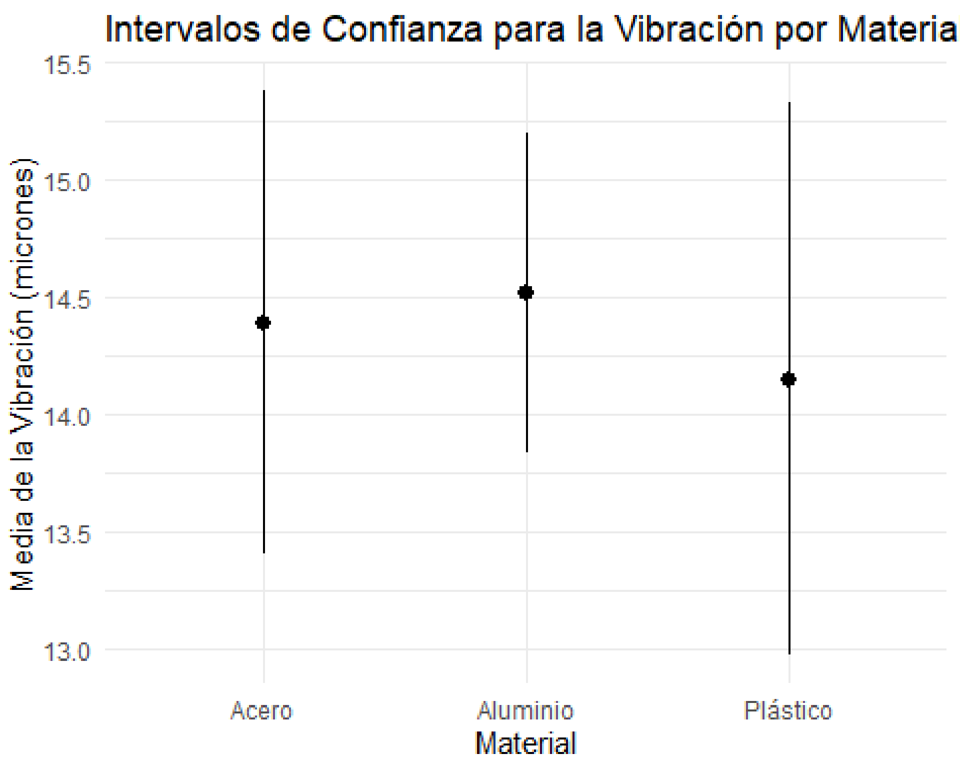


4. Realiza el ANOVA para dos niveles sin interacción

```
# Realizar el ANOVA sin interacción
anova_sin_interaccion <- aov(Vibracion ~ Material + Proveedor, data =
vibracion)
summary(anova_sin_interaccion)

##              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

# Boxplot de vibración por material
boxplot(Vibracion ~ Material, data = vibracion,
        col = c("lightblue", "lightgreen", "lightcoral"),
        main = "Boxplot de Vibraciones por Material",
        xlab = "Material", ylab = "Vibración (micrones)")
```

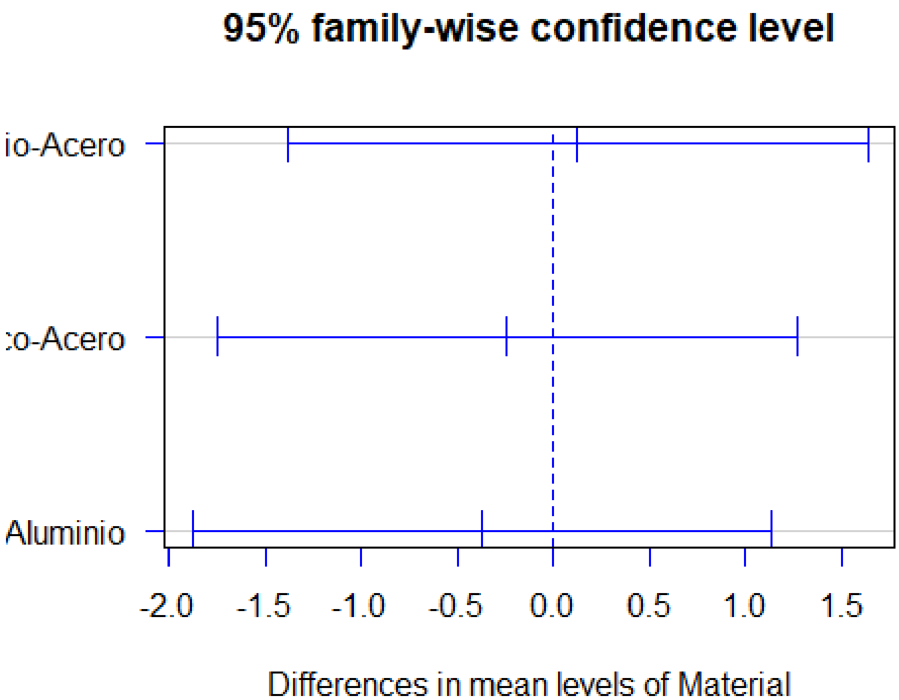



```r
# Intervalos de confianza para la vibración por material
summary_material <- vibracion %>%
  group_by(Material) %>%
  summarise(
    mean = mean(Vibracion),
    sd = sd(Vibracion),
    n = n(),
    se = sd/sqrt(n),
    lim_inf = mean - qt(1 - 0.05 / 2, n - 1) * se,
    lim_sup = mean + qt(1 - 0.05 / 2, n - 1) * se
  )

# Gráfico de intervalos de confianza
ggplot(summary_material, aes(x = Material, y = mean, ymin = lim_inf, ymax
= lim_sup)) +
  geom_pointrange() +
  labs(title = "Intervalos de Confianza para la Vibración por Material",
       x = "Material", y = "Media de la Vibración (micrones)") +
  theme_minimal()
```



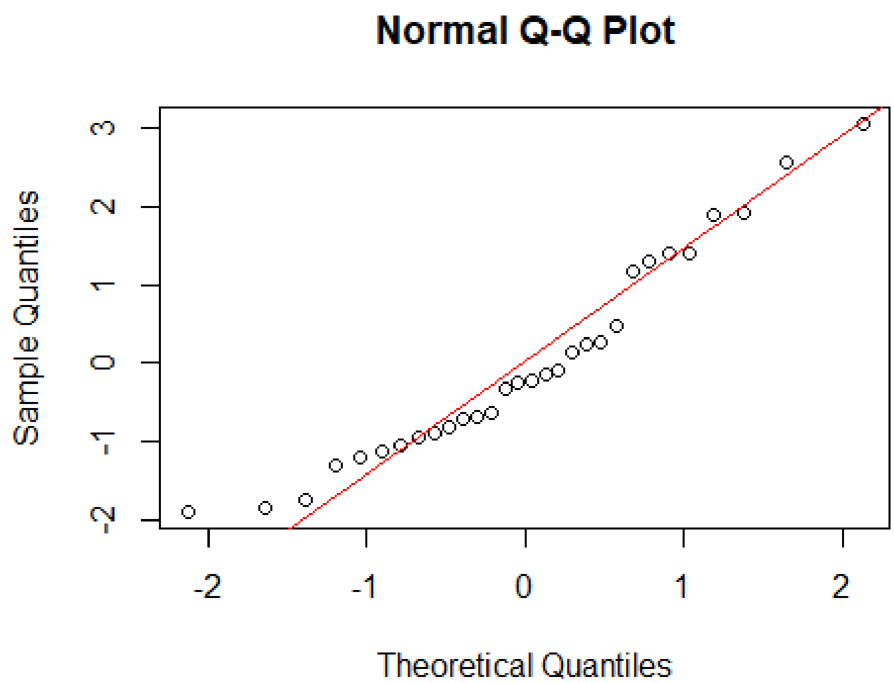
5. Realiza el ANOVA para un efecto principal

```
# Realizar el ANOVA solo para el material
anova_material <- aov(Vibracion ~ Material, data = vibracion)
summary(anova_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

# Prueba de comparaciones múltiples de Tukey
tukey_result <- TukeyHSD(anova_material)
print(tukey_result)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Vibracion ~ Material, data = vibracion)
##
## $Material
##                    diff       lwr      upr     p adj
## Aluminio-Acero     0.13 -1.378171 1.638171 0.9751575
## Plástico-Acero    -0.24 -1.748171 1.268171 0.9180284
## Plástico-Aluminio -0.37 -1.878171 1.138171 0.8168495

# Graficar los intervalos de confianza de Tukey
plot(tukey_result, las = 1, col = "blue")
```

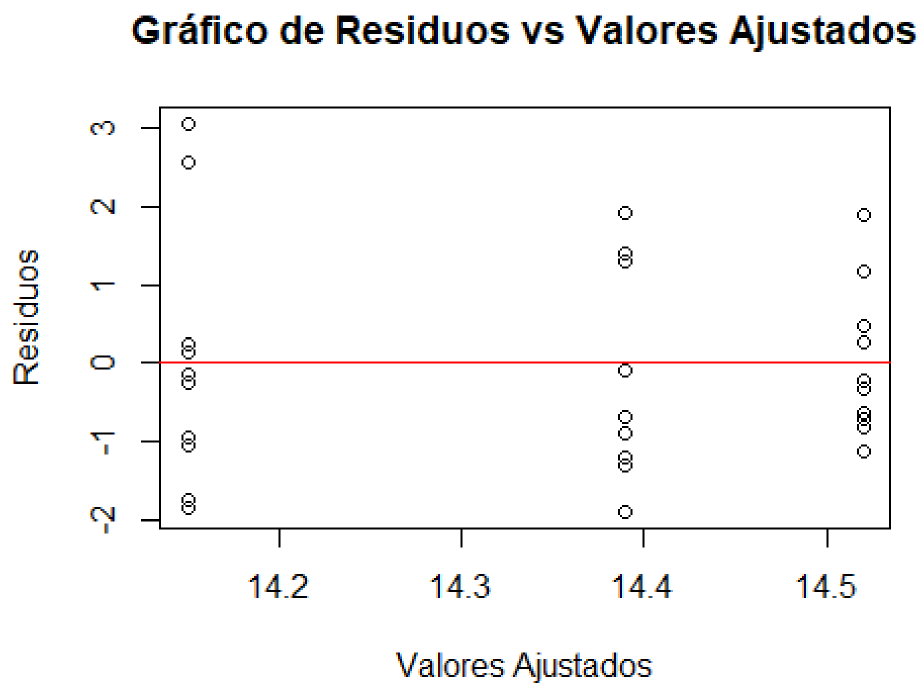


6. Comprueba la validez del modelo

```
# Normalidad de los residuos
residuos <- residuals(anova_material)

# Q-Q plot
qqnorm(residuos)
qqline(residuos, col = "red")
```




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

##
##  Shapiro-Wilk normality test
##
## data:  residuos
## W = 0.94362, p-value = 0.1139

# Test de Bartlett para homocedasticidad
bartlett_test <- bartlett.test(Vibracion ~ Material, data = vibracion)
print(bartlett_test)

##
##  Bartlett test of homogeneity of variances
##
## data:  Vibracion by Material
## Bartlett's K-squared = 2.4442, df = 2, p-value = 0.2946

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



7. Conclusiones Generales:

1. Efecto del Material y del Proveedor en la Vibración El análisis ANOVA mostró que el material no tiene un efecto significativo en la vibración de los motores eléctricos ($p > 0.8$). Esto indica que los tres materiales (acero, aluminio y plástico) tienen un rendimiento similar en términos de vibración. Aunque el proveedor parece tener un efecto más fuerte en comparación con el material, los resultados aún no alcanzan el nivel de significancia estadística típico ($p \approx 0.05$). Esto sugiere que, aunque puede haber alguna variación en la vibración según el proveedor, no es suficientemente fuerte para considerarse significativa bajo el umbral estándar de 0.05. No se observó una interacción significativa entre el material y el proveedor ($p > 0.5$). Esto significa que el impacto del proveedor en la vibración no depende del material utilizado en la carcasa del motor.
2. Análisis de Comparaciones Múltiples La prueba de comparaciones múltiples de Tukey confirmó que no existen diferencias significativas entre los pares de materiales. Esto refuerza la conclusión de que la elección del material no es un factor determinante en la vibración de los motores en este experimento.
3. Validación del Modelo La prueba de Shapiro-Wilk no rechazó la hipótesis nula de normalidad ($p > 0.1$), lo que sugiere que los residuos del modelo siguen una distribución aproximadamente normal. La prueba de Bartlett no encontró evidencia de heterocedasticidad ($p > 0.2$), lo que indica que las varianzas son homogéneas a través de los grupos. El análisis visual de los residuos y el gráfico de autocorrelación indicaron que no hay patrones de autocorrelación significativos, lo que sugiere que los residuos son independientes.