Actividad 9

Adrian Pineda Sanchez

2024-08-27

Rendimientos

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).

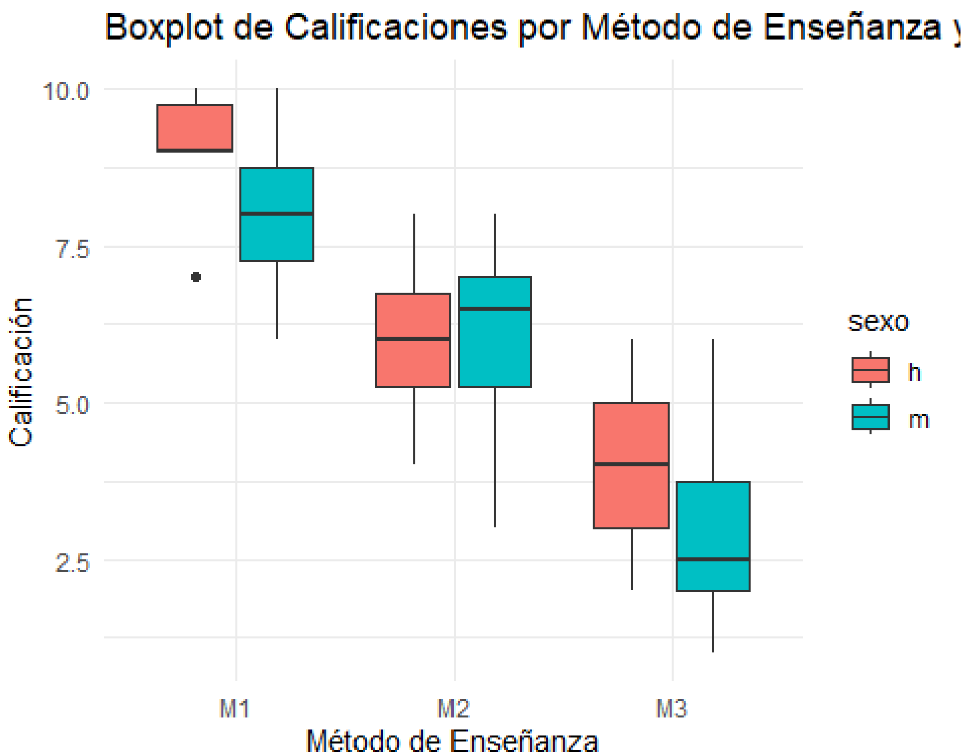
En los alumnos matriculados había el mismo número de chicos que de chicas, por lo que formamos dos grupos de 18 sujetos; en cada uno de ellos, el mismo profesor aplicará a grupos aleatorios de 6 estudiantes las 3 metodologías de estudio. A fin de curso los alumnos son sometidos a la misma prueba de rendimiento. Los resultados son los siguientes:

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

```
# Crear el DataFrame 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 metodo y sexo en factores
metodo <- factor(metodo)
sexo <- factor(sexo)

# Crear el DataFrame
df <- data.frame(calificacion, metodo, sexo)

# Calcular la media por método de enseñanza
media_metodo <- aggregate(calificacion ~ metodo, data = df, FUN = mean)
print(media_metodo)

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

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

```r
# Generar el boxplot
library(ggplot2)
ggplot(df, 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()
```



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

Método 1 (M1): Las calificaciones de los hombres son consistentemente más altas y menos dispersas en comparación con las de las mujeres. La falta de superposición entre las cajas sugiere una diferencia significativa en las medianas entre géneros. Esto podría indicar que este método es particularmente más efectivo para los hombres que para las mujeres.

Método 2 (M2): Existe una superposición entre las calificaciones de hombres y mujeres, lo que sugiere que este método tiene un rendimiento más similar entre géneros, aunque la dispersión es mayor en los hombres.

Método 3 (M3): Al igual que en el Método 1, se observa poca superposición entre las cajas, pero en este caso, ambos géneros tienen calificaciones bajas. Este método parece ser ineficaz para ambos géneros, aunque es ligeramente peor para las mujeres.

Escribe tus conclusiones parciales

Método 1: Aunque es el más efectivo en general, hay una disparidad de género significativa. Este método podría necesitar ajustes para mejorar la equidad de resultados entre hombres y mujeres.

Método 2: Es más equitativo en términos de género, pero con una mayor variabilidad en los resultados para los hombres.

Método 3: Es el menos efectivo y más desigual, especialmente para las mujeres. Este método requiere una revisión completa o incluso ser reemplazado.

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

H_{0_1}: $\sum_{i=1}^{n_\tau} \tau_i = 0$ H_{a_1}:Algún τ_i es distinto de cero

H_{0_2}: $\sum_{j=1}^{n_\alpha} \alpha_j = 0$ H_{a_2}:Algún α_j es distinto de cero

H_{0_3}: $\sum_{i=1}^{n_\tau} \sum_{j=1}^{n_\alpha} \tau_i \alpha_j = 0$ H_{a_3}:Algún $\tau_i\alpha_j$ es distinto de cero

3. Realiza el ANOVA para dos niveles con interacción:

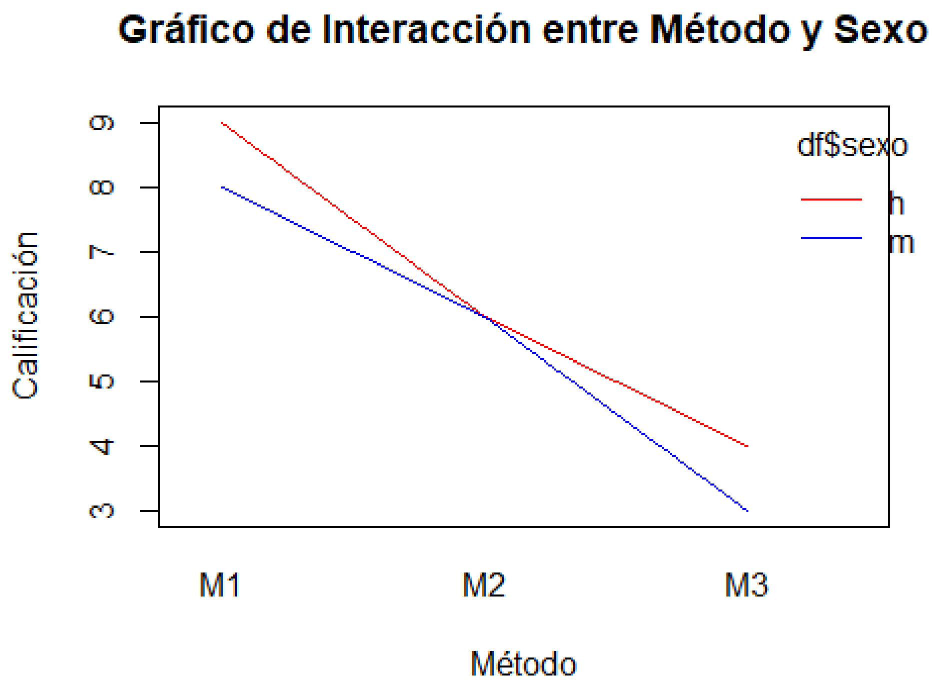
Consulta el código en R en los apoyos de clase de "ANOVA":

```
# Realizar el ANOVA para dos factores con interacción
anova_result <- aov(calificacion ~ metodo * sexo, data = df)

# Mostrar el resumen del ANOVA
summary(anova_result)
```

```
##             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
```

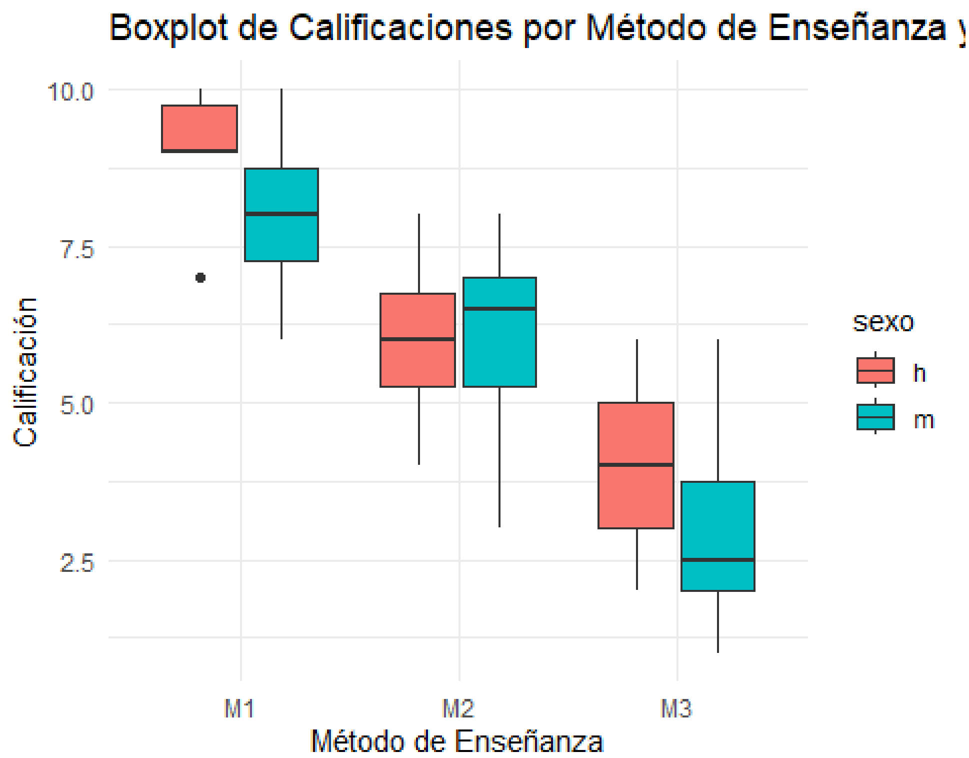
Haz la gráfica de interacción de dos factores en ANOVA

```
# Gráfico de interacción de los dos factores
interaction.plot(df$metodo, df$sexo, df$calificacion,
                 col = c("red", "blue"), lty = 1, pch = c(19, 17),
                 main = "Gráfico de Interacción entre Método y Sexo",
                 xlab = "Método", ylab = "Calificación", legend = TRUE)
```



Haz el boxplot para visualizar la interacción de los factores, por ejemplo, peso por dieta interacción ejercicio: boxplot(peso ~ dieta * ejercicio, data = data, col = c("lightblue", "lightgreen"), main = "Boxplot de Peso por Dieta y Ejercicio")

```
# Generar el boxplot
library(ggplot2)
ggplot(df, 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()
```



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

Método 1 (M1):

Los hombres tienen un rendimiento promedio más alto que las mujeres.

La diferencia entre géneros es bastante notable, lo que sugiere que este método puede estar favoreciendo a los hombres.

Método 2 (M2):

Las calificaciones promedio para ambos géneros disminuyen en comparación con M1, pero las mujeres parecen verse más afectadas, lo que resulta en un rendimiento promedio más bajo que el de los hombres.

Aquí, las líneas se cruzan, lo que indica un cambio en la relación entre género y método, sugiriendo una interacción significativa.

Método 3 (M3):

Las calificaciones continúan disminuyendo para ambos géneros, pero la diferencia entre hombres y mujeres es menor.

Sin embargo, el rendimiento de las mujeres es consistentemente más bajo que el de los hombres a través de todos los métodos.

Escribe tus conclusiones parciales

Interacción Significativa: El cruce de las líneas en el gráfico de interacción entre el Método 2 y el género sugiere que la efectividad del método de enseñanza depende del género del estudiante. Esta interacción indica que los métodos no afectan a hombres y mujeres de la misma manera.

Diferencias de Rendimiento: Los hombres tienden a tener calificaciones más altas en todos los métodos en comparación con las mujeres, pero esta diferencia es más pronunciada en el Método 1.

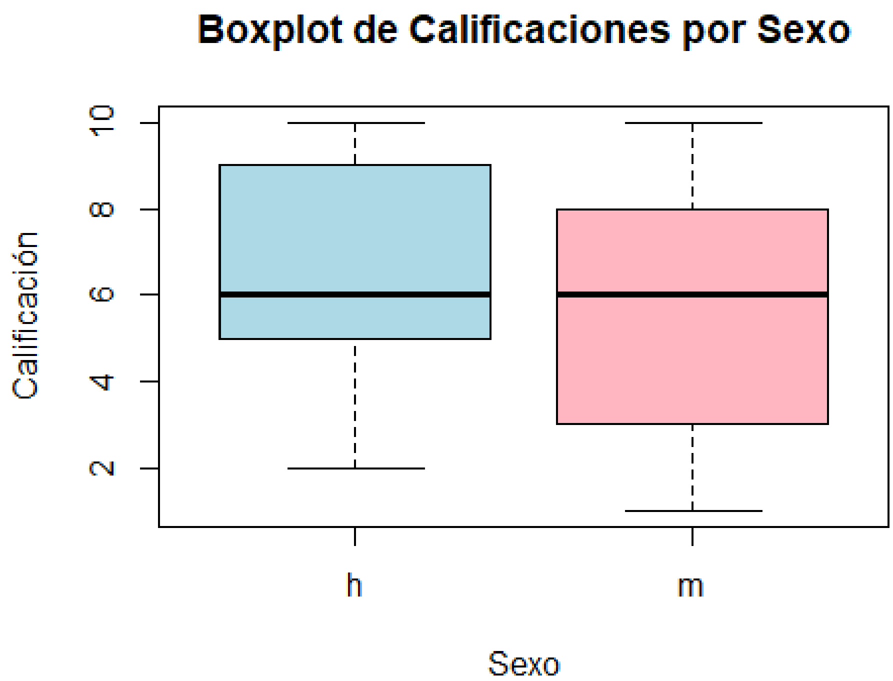
4. Realiza el ANOVA para dos niveles sin interacción.

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

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

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

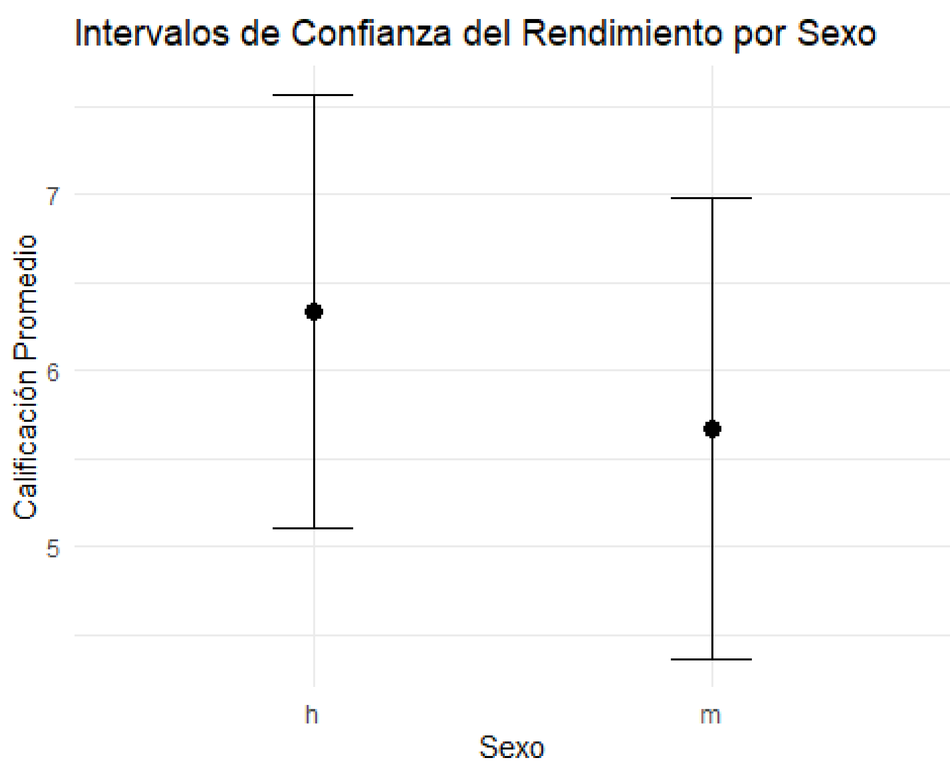



```
# Calcular la media para el rendimiento por sexo y método
media_sexo_metodo <- aggregate(calificacion ~ sexo + metodo, data = df, FUN =
mean)
print(media_sexo_metodo)

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

Haz los intervalos de confianza de rendimiento por sexo. Grafícalos

```
# Intervalos de confianza de rendimiento por sexo
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
media_confianza_sexo <- df %>%
  group_by(sexo) %>%
  summarise(
    media = mean(calificacion),
    sd = sd(calificacion),
    n = n(),
    se = sd / sqrt(n),
    ic_inf = media - qt(1 - 0.05 / 2, n - 1) * se,
    ic_sup = media + qt(1 - 0.05 / 2, n - 1) * se
  )

# Graficar los intervalos de confianza
library(ggplot2)
ggplot(media_confianza_sexo, aes(x = sexo, y = media)) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = ic_inf, ymax = ic_sup), width = 0.2) +
  labs(title = "Intervalos de Confianza del Rendimiento por Sexo",
       x = "Sexo", y = "Calificación Promedio") +
  theme_minimal()
```



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

1. Resultados del ANOVA sin interacción:

Método de enseñanza: El valor p para el método de enseñanza es extremadamente bajo ($p<0.001$), lo que indica que el método de enseñanza tiene un efecto significativo en el rendimiento de los estudiantes.

Sexo: El valor p para el sexo es 0.192, lo que significa que, sin considerar la interacción, no hay evidencia suficiente para concluir que el sexo tenga un efecto significativo en el rendimiento. Sin embargo, esto no descarta que el sexo pueda influir en combinación con otros factores.

2. Boxplot de calificaciones por sexo: El boxplot muestra que las calificaciones de los hombres (h) tienden a ser ligeramente más altas que las de las mujeres (m). Aunque hay cierta variabilidad, no parece haber una gran diferencia entre los sexos en términos de la mediana.

3. Medias de rendimiento por sexo y método: Los datos tabulados muestran las calificaciones medias por método y sexo. Esto refuerza la observación de que los hombres tienden a tener calificaciones ligeramente más altas, pero la diferencia entre sexos no es tan marcada como la diferencia entre métodos.

4. Intervalos de confianza de rendimiento por sexo: Los intervalos de confianza para los dos sexos se superponen, lo que sugiere que las diferencias en las calificaciones medias entre hombres y mujeres no son estadísticamente significativas.

Escribe tus conclusiones parciales

Efecto del Método de Enseñanza:El método de enseñanza tiene un impacto significativo en el rendimiento de los estudiantes, como lo indica el análisis ANOVA. Esto significa que el rendimiento varía significativamente dependiendo del método utilizado.

Impacto del Sexo: Aunque las calificaciones de los hombres son ligeramente más altas en general, el análisis ANOVA sugiere que el sexo, por sí solo, no tiene un efecto significativo en el rendimiento, ya que el valor p es mayor a 0.05.

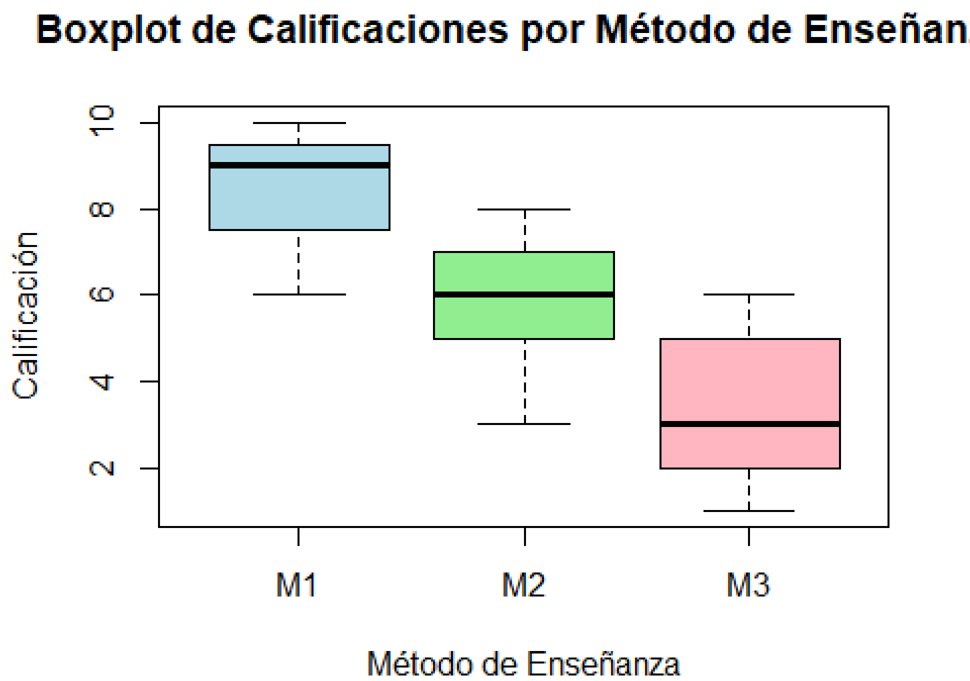
5. Realiza el ANOVA para un efecto principal

```r
# Realizar el ANOVA para un efecto principal (método de enseñanza)
anova_metodo <- aov(calificacion ~ metodo, data = df)
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
```

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

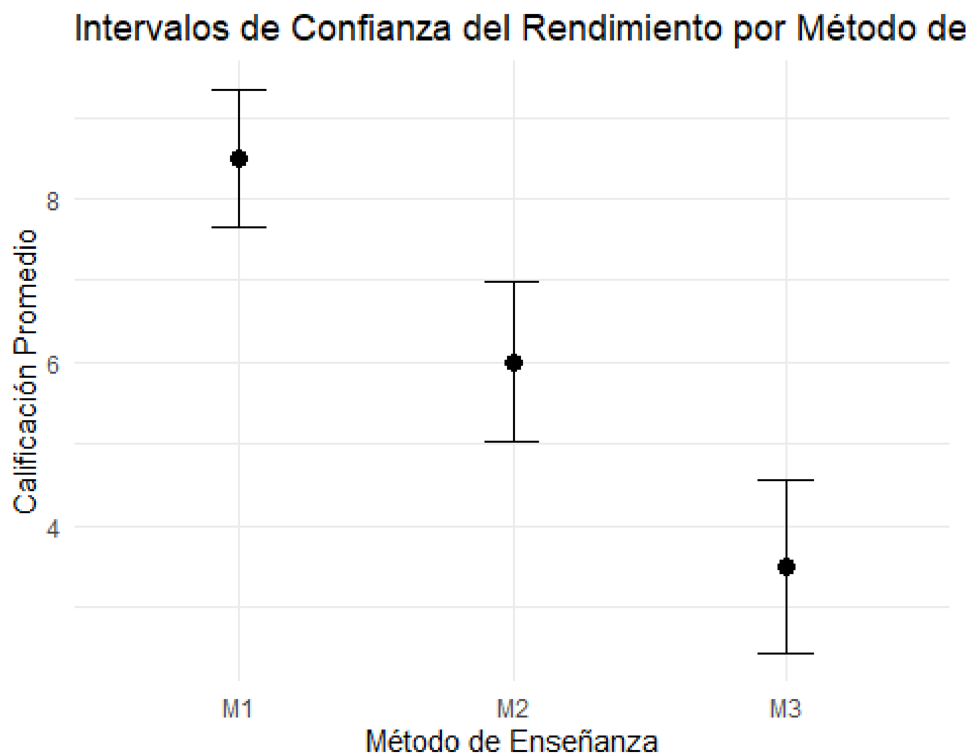
```r
# Boxplot de rendimiento por método de enseñanza
boxplot(calificacion ~ metodo, data = df,
        col = c("lightblue", "lightgreen", "lightpink"),
        main = "Boxplot de Calificaciones por Método de Enseñanza",
        xlab = "Método de Enseñanza", ylab = "Calificación")
```



Haz los intervalos de confianza de rendimiento por método. Grafícalos

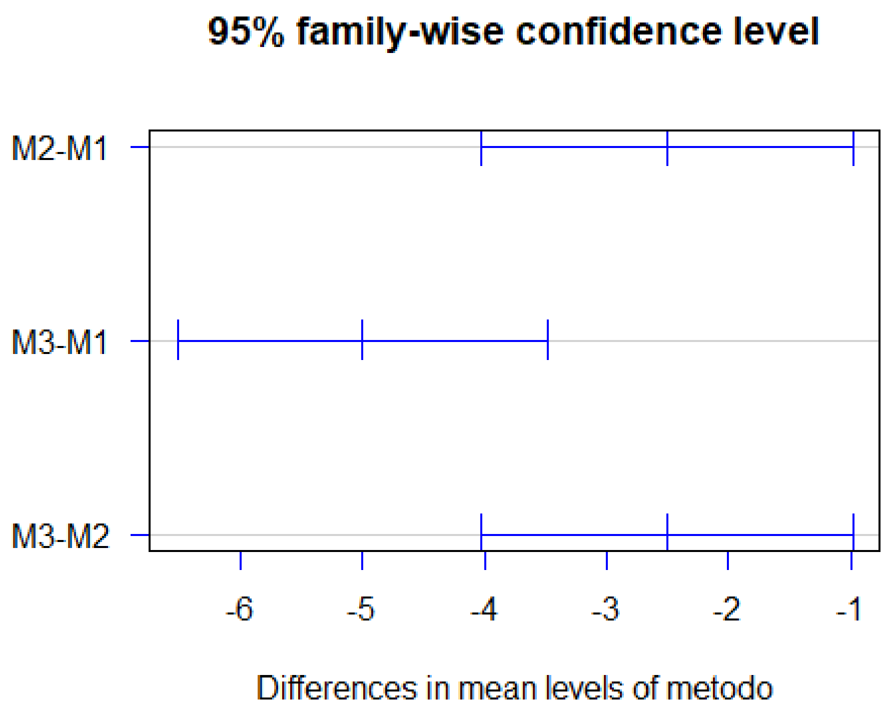
```
# Calcular la media para el rendimiento por método de enseñanza
media_metodo <- aggregate(calificacion ~ metodo, data = df, FUN = mean)
print(media_metodo)

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

# Intervalos de confianza de rendimiento por método
media_confianza_metodo <- df %>%
  group_by(metodo) %>%
  summarise(
    media = mean(calificacion),
    sd = sd(calificacion),
    n = n(),
    se = sd / sqrt(n),
    ic_inf = media - qt(1 - 0.05 / 2, n - 1) * se,
    ic_sup = media + qt(1 - 0.05 / 2, n - 1) * se
  )

# Graficar los intervalos de confianza por método
ggplot(media_confianza_metodo, aes(x = metodo, y = media)) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = ic_inf, ymax = ic_sup), width = 0.2) +
  labs(title = "Intervalos de Confianza del Rendimiento por Método de
```

```r
Enseñanza",
       x = "Método de Enseñanza", y = "Calificación Promedio") +
  theme_minimal()
```



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

```r
# Realizar la prueba de comparaciones múltiples de Tukey
tukey_result <- TukeyHSD(anova_metodo)

# Mostrar los resultados de Tukey
print(tukey_result)

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



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

1. Resultados del ANOVA para un efecto principal (Método de Enseñanza):

El valor p=1.55e−08 es extremadamente bajo, lo que indica que hay diferencias significativas entre los métodos de enseñanza. Esto significa que el método de enseñanza tiene un impacto importante en el rendimiento de los estudiantes.

2. Boxplot de rendimiento por método de enseñanza:

El boxplot muestra que el Método 1 (M1) tiene la calificación promedio más alta y la menor variabilidad, lo que sugiere que es el más efectivo y consistente.

El Método 3 (M3) tiene la calificación promedio más baja, indicando que es el menos efectivo.

El Método 2 (M2) se encuentra entre los otros dos en términos de rendimiento promedio, pero con una variabilidad más notable.

3. Intervalos de confianza del rendimiento por método:

Los intervalos de confianza muestran la precisión de las medias calculadas para cada método.

No hay superposición significativa entre los intervalos de los tres métodos, lo que refuerza la idea de que las diferencias entre ellos son estadísticamente significativas.

4. Prueba de comparaciones múltiples de Tukey:

La prueba de Tukey confirma que todas las diferencias entre los pares de métodos son significativas, con valores p extremadamente bajos.

Esto significa que cada método difiere significativamente de los otros dos en términos de rendimiento.

El gráfico de intervalos de confianza de Tukey aunque se solapabn o se sobreponen muestra que las diferencias entre los métodos no incluyen el cero, lo que confirma su significancia estadística.

Escribe tus conclusiones parciales

Impacto del Método de Enseñanza:

El análisis muestra claramente que el método de enseñanza tiene un efecto significativo en el rendimiento de los estudiantes. Esto se evidencia en los resultados del ANOVA, el boxplot, y la prueba de Tukey.

Efectividad de los Métodos:

Método 1 (M1) es el más efectivo, con la calificación promedio más alta y la menor variabilidad.

Método 3 (M3) es el menos efectivo, con la calificación promedio más baja.

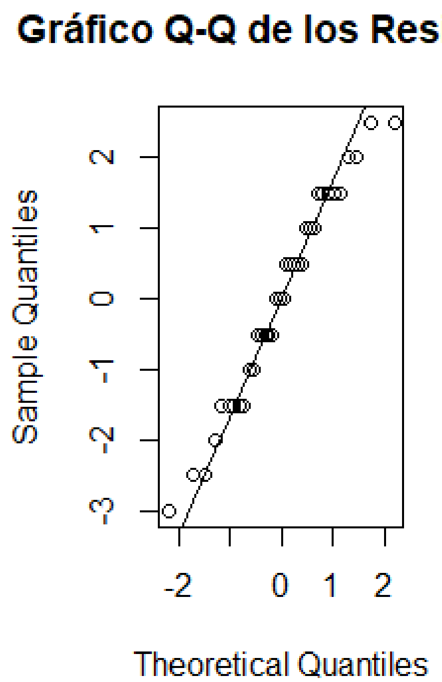
Método 2 (M2), aunque intermedio en rendimiento, muestra una mayor variabilidad, lo que podría indicar que no es igualmente efectivo para todos los estudiantes.

6. Comprueba la validez del modelo. Comprueba:

Normalidad

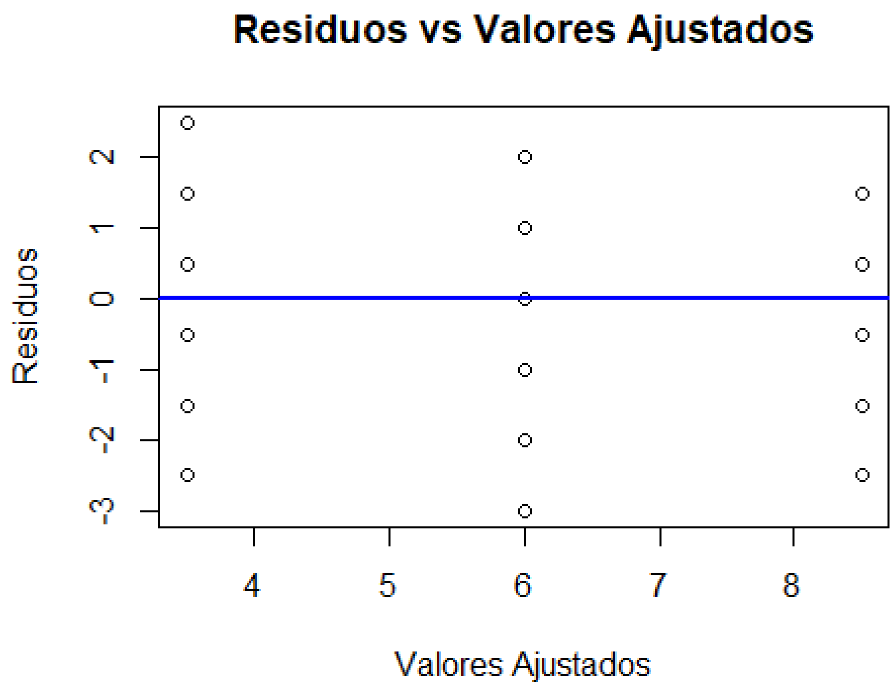
```r
# Normalidad de los residuos
par(mfrow = c(1, 2))
qqnorm(residuals(anova_metodo), main = "Gráfico Q-Q de los Residuos")
qqline(residuals(anova_metodo))

shapiro.test(residuals(anova_metodo))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(anova_metodo)
## W = 0.96734, p-value = 0.3573
```



Homocedasticidad

```r
# Homocedasticidad
plot(fitted(anova_metodo), residuals(anova_metodo),
     main = "Residuos vs Valores Ajustados",
     xlab = "Valores Ajustados", ylab = "Residuos")
abline(h = 0, col = "red", lwd = 2)

# Añadir una línea de regresión para observar la tendencia
coeficiente <- lm(residuals(anova_metodo) ~ fitted(anova_metodo))
abline(coef(coeficiente), col = "blue", lwd = 2)
```

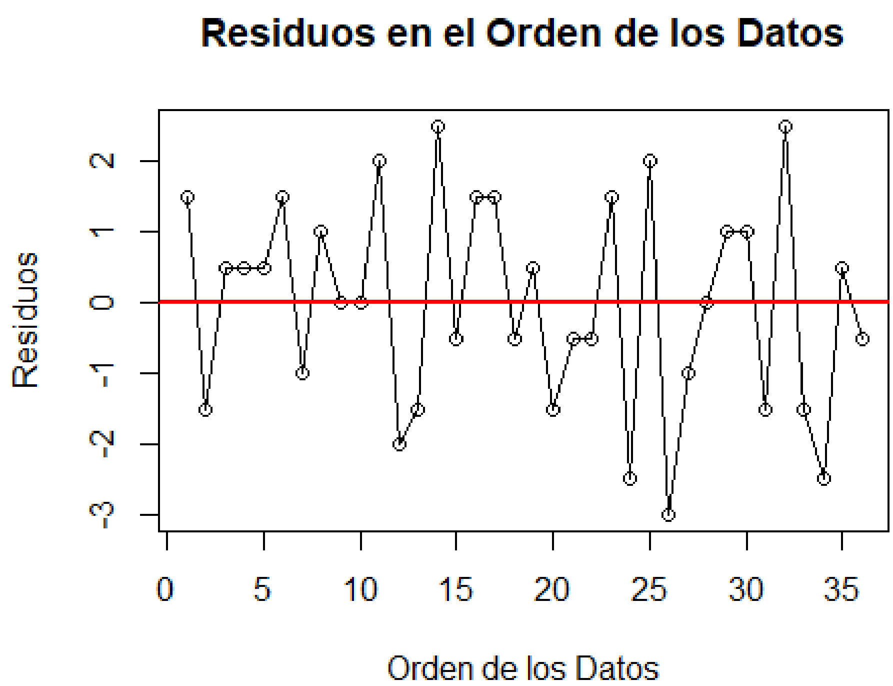



```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

# Realizar la prueba de Breusch-Pagan
bptest_result <- bptest(anova_metodo)

# Imprimir el resultado del test de Breusch-Pagan
print(bptest_result)

##
##  studentized Breusch-Pagan test
##
## data:  anova_metodo
## BP = 1.2041, df = 2, p-value = 0.5477
```

Independencia

```
# Independencia de los residuos
plot(residuals(anova_metodo), type = "o",
     main = "Residuos en el Orden de los Datos",
```

```
     xlab = "Orden de los Datos", ylab = "Residuos")
abline(h = 0, col = "red", lwd = 2)
```




```
# Prueba de Durbin-Watson para independencia de los residuos
dw_test1 <- dwtest(anova_metodo)
print(dw_test1)

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

Relación lineal entre las variables (coeficiente de determinación).

```
# Ajustar el modelo ANOVA
anova_metodo <- aov(calificacion ~ metodo, data = df)

# Calcular el R^2
suma_cuadrados_total <- sum((df$calificacion - mean(df$calificacion))^2)
suma_cuadrados_modelo <- sum((fitted(anova_metodo) -
mean(df$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"
```

7. Concluye en el contexto del problema.

Boxplot: Los boxplots muestran diferencias en las calificaciones por método de enseñanza y por género. Parece haber una variabilidad notable dependiendo del método, y también hay ciertas diferencias entre chicos y chicas en cada método. Esto sugiere que tanto la metodología de enseñanza como el género pueden estar influyendo en el rendimiento. Normalidad de los residuos:

Gráfico Q-Q y Prueba de Shapiro-Wilk: El gráfico Q-Q de los residuos y la prueba de Shapiro-Wilk (p-valor = 0.3573) indican que no podemos negar normalidad , lo cual es un buen indicio para la validez del análisis de varianza (ANOVA).

Homoscedasticidad:

Gráfico de residuos vs valores ajustados y Prueba de Breusch-Pagan: Los gráficos de residuos versus valores ajustados y el test de Breusch-Pagan (p-valor = 0.5477) sugieren que no hay problemas de heteroscedasticidad, es decir, la variabilidad de los residuos es constante a lo largo de las predicciones, lo cual refuerza la validez del modelo.

Ajuste del modelo: Coeficiente de Determinación (R^2): Un R^2 de aproximadamente 0.66 indica que el modelo explica el 66.3% de la variabilidad en las calificaciones. Esto sugiere que aunque el modelo tiene un buen ajuste, todavía hay un 33.7% de la variabilidad que no es explicada por las variables en el modelo.

En resumen, tanto el método de enseñanza como el género parecen influir en el rendimiento académico, y existe evidencia que sugiere una posible interacción entre ambos factores. El modelo estadístico utilizado es válido bajo los supuestos de normalidad, homocedasticidad e independencia de los residuos.

“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 seleccionados al azar.

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

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

# Proveedores de cojinetes (Factor B)
```

```r
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, FUN =
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, FUN =
mean)
print("Media de Vibración por Proveedor:")
```

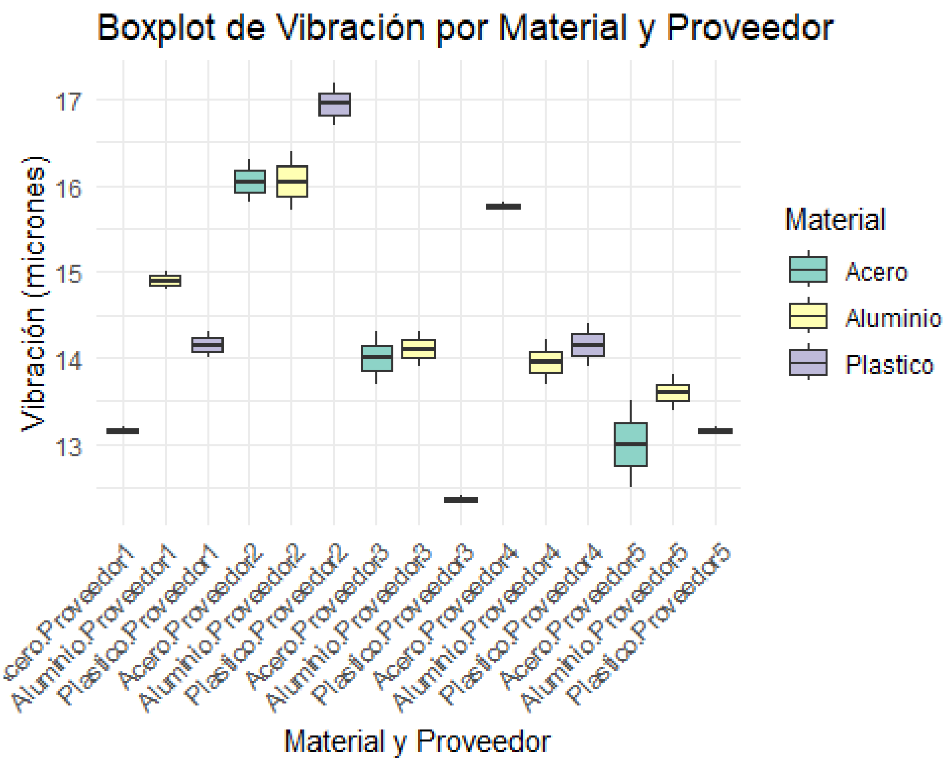
```
## [1] "Media de Vibración por Proveedor:"
```

```
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
```

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

```
# Generar el boxplot con ggplot2
ggplot(data, aes(x = interaction(Material, Proveedor), y = Vibracion, fill =
Material)) +
  geom_boxplot() +
  labs(title = "Boxplot de Vibración por Material y Proveedor",
       x = "Material y Proveedor", y = "Vibración (micrones)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_brewer(palette = "Set3")  # Añadir una paleta de colores para
mejorar la visualización
```



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

Escribe tus conclusiones parciales

Material:

*Boxplot: El boxplot muestra la dispersión de las vibraciones según el material y proveedor. No hay diferencias extremas en la vibración media entre los materiales, aunque el aluminio muestra la media más alta y el plástico la más baja. Sin embargo, la variabilidad entre proveedores parece ser significativa.

*Conclusión: Aunque las diferencias en la media de vibración entre materiales no son muy grandes, el aluminio tiende a generar una mayor vibración en comparación con el plástico y el acero.

Proveedor:

*Boxplot: Los boxplots muestran que la variabilidad en la vibración depende significativamente del proveedor. El Proveedor 2 muestra la vibración más alta en comparación con los otros, lo cual coincide con su media de 16.35 micrones, la más alta entre todos los proveedores.

*Conclusión: El Proveedor 2 parece ser responsable de una mayor vibración en general, lo que sugiere que la selección del proveedor es crucial para controlar los niveles de vibración.

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

Primera Hipótesis (Efecto del Material):

$$H_{0_1}: \tau_i = 0 \quad \text{para todos los } i$$

$$H_{a_1}: \tau_i \neq 0 \quad \text{para algún } i$$

Segunda Hipótesis (Efecto del Proveedor):

$$H_{0_2}: \alpha_j = 0 \quad \text{para todos los } j$$

$$H_{a_2}: \alpha_j \neq 0 \quad \text{para algún } j$$

Tercera Hipótesis (Interacción entre Material y Proveedor):

$$H_{0_3}: (\tau\alpha)_{ij} = 0 \quad \text{para todos los } i, j$$

$$H_{a_3}: (\tau\alpha)_{ij} \neq 0 \quad \text{para algún } i, j$$

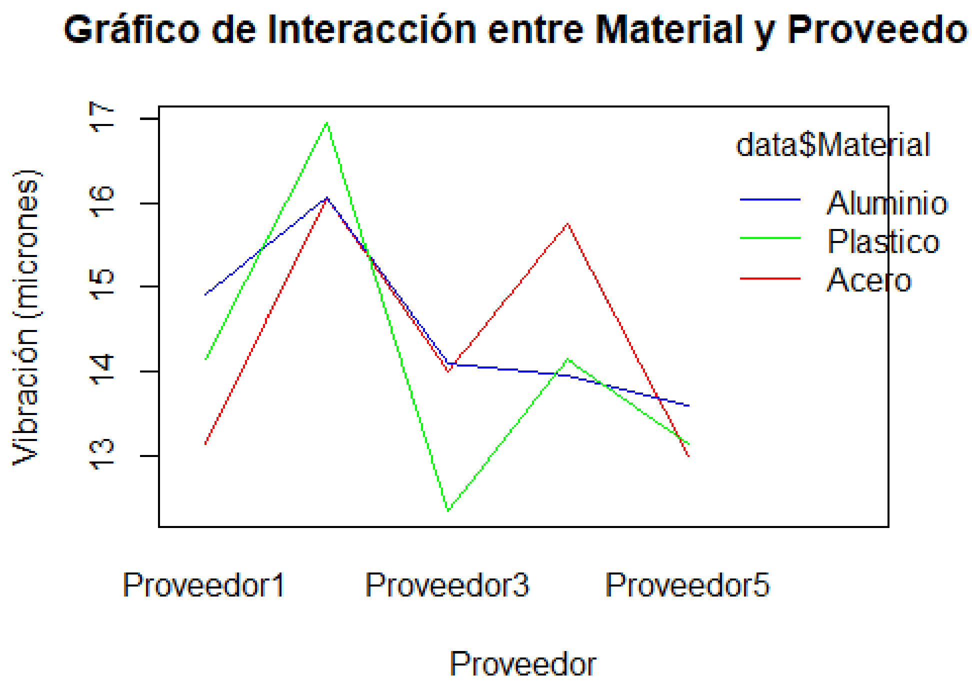
3. Realiza el ANOVA para dos niveles con interacción:

```
# Realizar el ANOVA con interacción
anova_result <- aov(Vibracion ~ Material * Proveedor, data = data)
summary(anova_result)
```

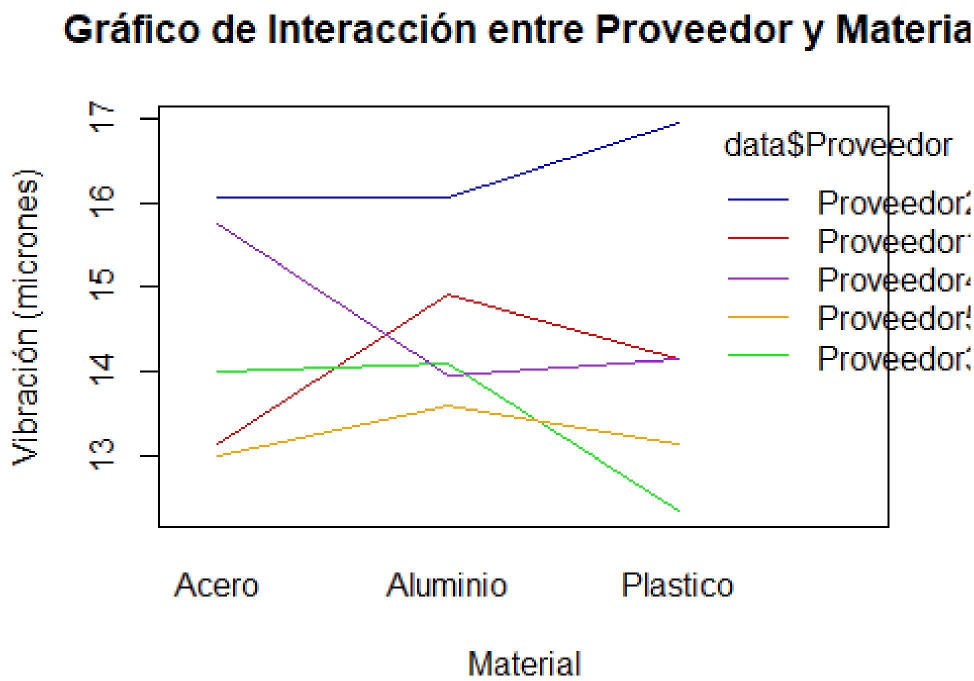
```
##                    Df Sum Sq Mean Sq F value   Pr(>F)    
## Material            2   0.70   0.352   3.165   0.0713 .  
## Proveedor           4  36.67   9.169  82.353 5.07e-10 ***
## Material:Proveedor  8  11.61   1.451  13.030 1.76e-05 ***
## Residuals          15   1.67   0.111             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

Haz la gráfica de interacción de dos factores en ANOVA

```
# Gráfico de interacción entre Material y Proveedor
interaction.plot(data$Proveedor, data$Material, data$Vibracion,
                 col = c("red", "blue", "green"), lty = 1, pch = c(19, 17,
15),
                 main = "Gráfico de Interacción entre Material y Proveedor",
                 xlab = "Proveedor", ylab = "Vibración (micrones)")
```

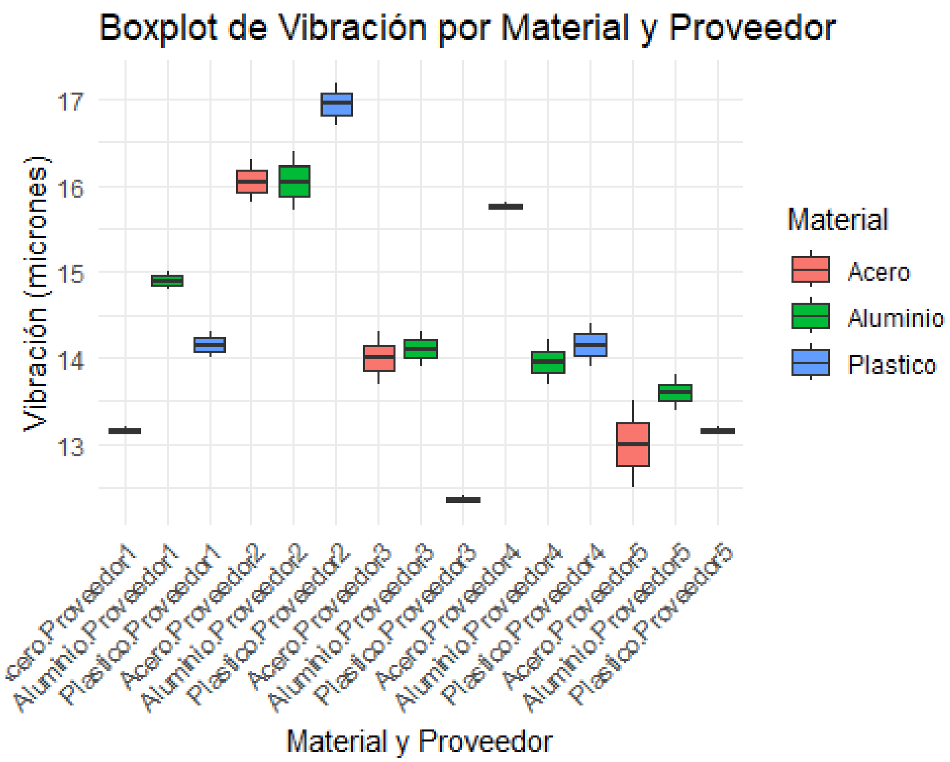



```r
# Gráfico de interacción entre Proveedor y Material (con Material en el eje x)
interaction.plot(data$Material, data$Proveedor, data$Vibracion,
                 col = c("red", "blue", "green", "purple", "orange"), lty = 1, pch = c(19, 17, 15, 13, 11),
                 main = "Gráfico de Interacción entre Proveedor y Material",
                 xlab = "Material", ylab = "Vibración (micrones)")
```



Haz el boxplot para visualizar la interacción de los factores, por ejemplo, peso por dieta interacción ejercicio:

```r
library(ggplot2)

# Generar el boxplot con ggplot2
ggplot(data, aes(x = interaction(Material, Proveedor), y = Vibracion, fill =
Material)) +
  geom_boxplot() +
  labs(title = "Boxplot de Vibración por Material y Proveedor",
       x = "Material y Proveedor", y = "Vibración (micrones)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



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

Escribe tus conclusiones parciales

Estadística del ANOVA

Efecto del Material: p-valor = 0.0713: El efecto del material sobre la vibración tiene un p-valor de 0.0713. Esto es ligeramente mayor que el nivel de significancia común de 0.05, lo que sugiere que no hay suficiente evidencia para rechazar la hipótesis nula. Sin embargo, está cerca del umbral, lo que podría considerarse como marginalmente significativo. Esto indica que el tipo de material podría tener un efecto en la vibración, pero no es concluyente con el 95% de confianza.

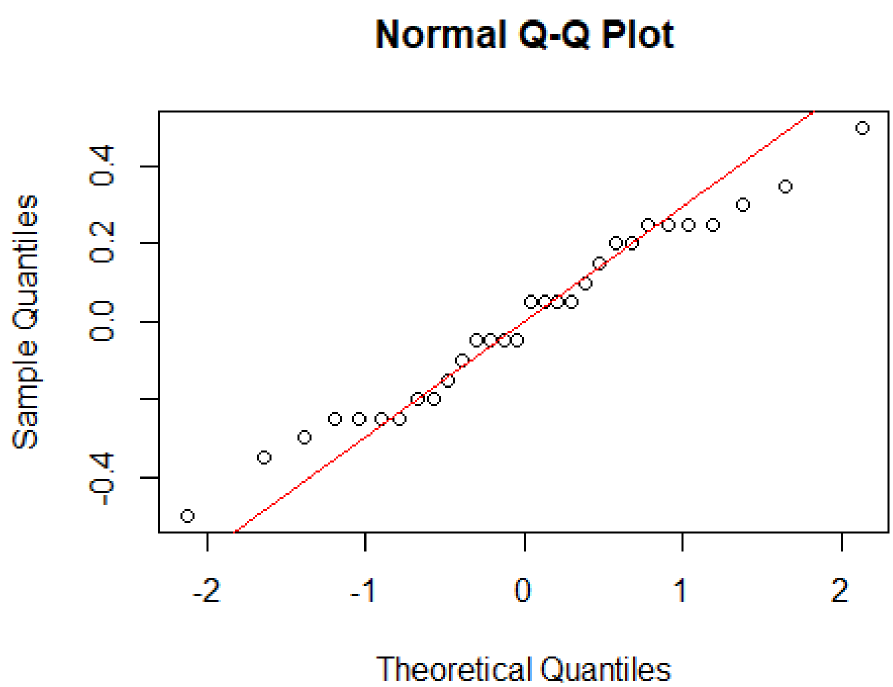
Efecto del Proveedor: p-valor < 0.001: El efecto del proveedor es altamente significativo con un p-valor de 5.07e-10. Esto indica que los diferentes proveedores tienen un impacto significativo en la vibración de los motores. En otras palabras, cambiar de proveedor tiene un efecto claro en la cantidad de vibración.

Interacción Material:Proveedor: p-valor < 0.001: La interacción entre el material y el proveedor también es altamente significativa con un p-valor de 1.76e-05. Esto significa que el efecto del material en la vibración depende del proveedor, y viceversa. Este resultado sugiere que no se puede considerar el efecto de los materiales o de los proveedores de manera independiente, ya que hay una interacción significativa entre ambos factores.

6. Comprueba la validez del modelo. Comprueba:

Normalidad

```r
# Obtener los residuos del modelo ANOVA
residuos <- residuals(anova_result)

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

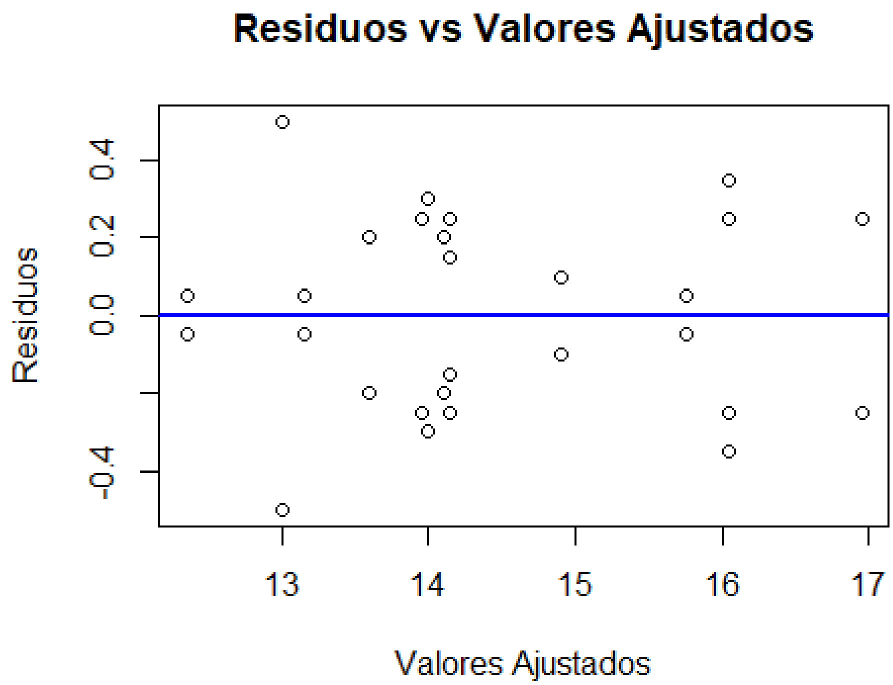



```r
# Prueba de Shapiro-Wilk para normalidad
shapiro_test <- shapiro.test(residuos)
print(shapiro_test)

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

##Homocedasticidad

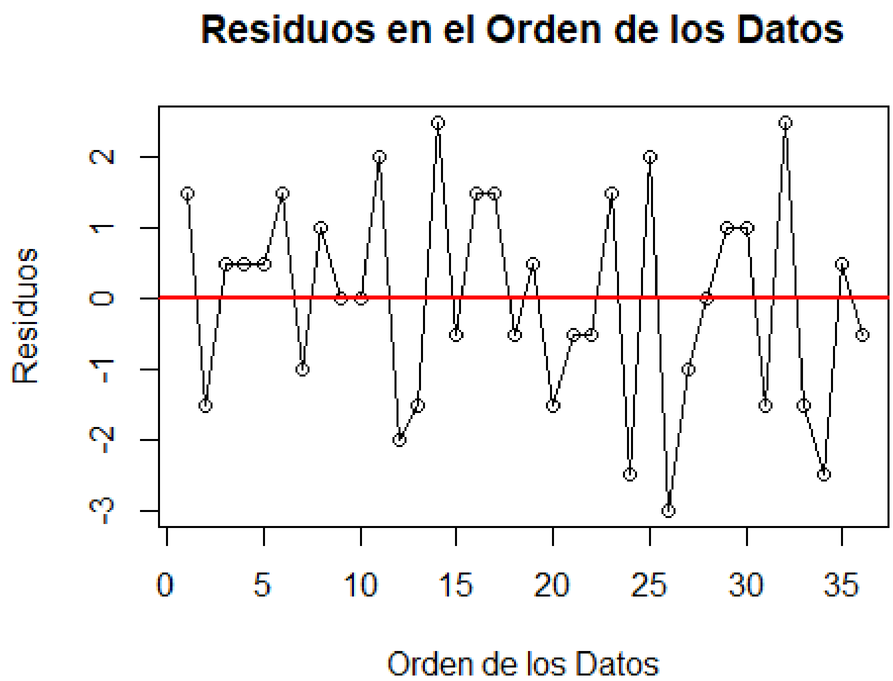
```r
# Gráfico de residuos vs valores ajustados
plot(fitted(anova_result), residuos,
     main = "Residuos vs Valores Ajustados",
     xlab = "Valores Ajustados", ylab = "Residuos")
abline(h = 0, col = "red", lwd = 2)
```

```r
# Añadir una línea de regresión para observar la tendencia
coeficiente <- lm(residuos ~ fitted(anova_result))
abline(coef(coeficiente), col = "blue", lwd = 2)
```



Independencia

```r
# Gráfico de Residuos en el Orden de los Datos
plot(residuals(anova_metodo), type = "o",
     main = "Residuos en el Orden de los Datos",
     xlab = "Orden de los Datos", ylab = "Residuos")
abline(h = 0, col = "red", lwd = 2)
```




```
# Prueba de Durbin-Watson para independencia de los residuos
dw_test <- dwtest(anova_result)
print(dw_test)
```

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

Relación lineal entre las variables (coeficiente de determinación).

```
# Calcular el R^2 manualmente
suma_cuadrados_total <- sum((data$Vibracion - mean(data$Vibracion))^2)
suma_cuadrados_modelo <- sum((fitted(anova_result) - 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"
```

7. Concluye en el contexto del problema.

1. Normalidad: Gráfico Q-Q: El gráfico Q-Q muestra que los residuos se ajustan bien a la línea de referencia, lo que sugiere que los residuos se distribuyen normalmente. Prueba de Shapiro-Wilk: El p-valor obtenido es de 0.7202, que es mayor que 0.05. Esto indica que no hay evidencia suficiente para rechazar la hipótesis nula de normalidad. Por lo tanto, podemos asumir que los residuos del modelo siguen una distribución normal.
2. Homocedasticidad: Gráfico de Residuos vs Valores Ajustados: El gráfico muestra que los residuos están distribuidos aleatoriamente alrededor de cero y no muestran un patrón claro, lo que sugiere homocedasticidad

3. Independencia: Gráfico de Residuos en el Orden de los Datos: El gráfico de los residuos en el orden de los datos muestra que no hay un patrón claro de autocorrelación.Prueba de Durbin-Watson: El p-valor es 0.5048, que es mayor que 0.05, indicando que no hay evidencia de autocorrelación en los residuos. Por lo tanto, podemos asumir que los residuos son independientes.

4. Relación Lineal entre las Variables (Coeficiente de Determinación): Coeficiente de Determinación (R^2): El valor de R^2 es 0.967, lo que indica que el 96.7% de la variabilidad en la vibración se explica por el modelo. Esto sugiere una fuerte relación lineal entre las variables predictoras y la respuesta.

Conclusión General Basada en las Hipótesis:

*Material: Aunque no se encontró un efecto significativo a un nivel del 5%, el material podría tener un efecto marginal sobre la vibración.

*Proveedor: El proveedor tiene un efecto significativo sobre la vibración, lo que sugiere que es un factor crítico en el diseño y selección de componentes para minimizar la vibración.

*Interacción Material-Proveedor: Existe una interacción significativa entre el material y el proveedor