Actividad_ANOVA

Rodolfo Jesús Cruz Rebollar

2024-08-29

Problema 1

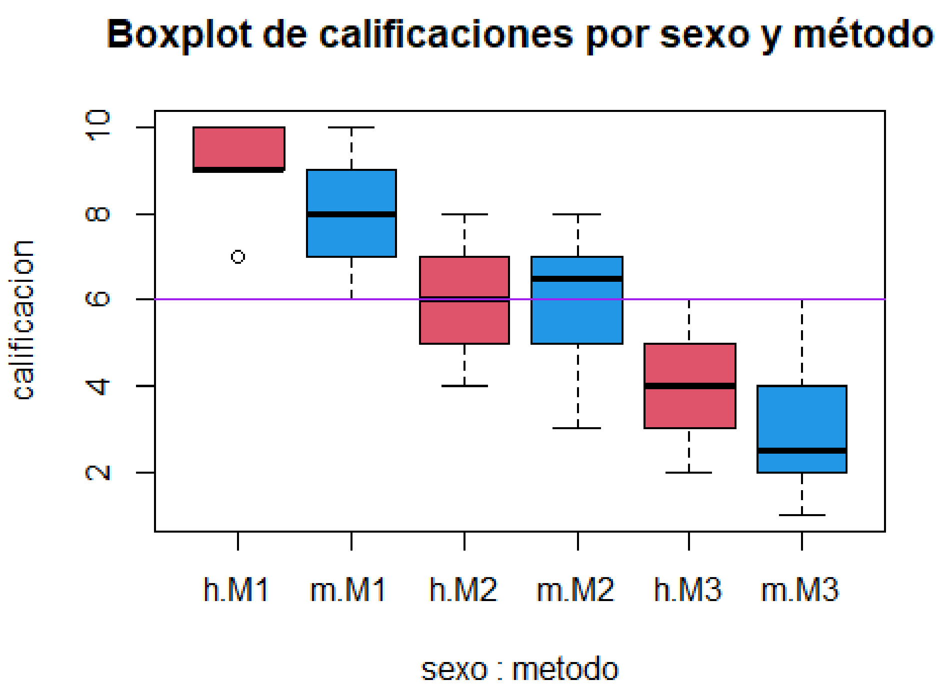
En un instituto se han matriculado 36 estudiantes. Se desea explicar el rendimiento de ciencias naturales en función de dos variables: género y metodología de enseñanza. La metodología de enseñanza se analiza en tres niveles: explicación oral y realización del experimento (1er nivel) explicación oral 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:

```r
# Datos de calificaciones tanto de chicos como de chicas

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)

# Datos correspondientes al método de enseñanza

metodo=c(rep("M1",6),rep("M2",6),rep("M3",6),rep("M1",6),rep("M2",6),rep(
"M3",6))

# Sexo de los estudiantes

sexo = c(rep("h", 18), rep("m",18))

# Método de enseñanza convertido a datos de tipo factor

metodo = factor(metodo)

# Convertir datos de sexo en datos de tipo factor

sexo = factor(sexo)
```

1. Análisis exploratorio

```
# Colocar todos los datos previos en un dataframe

alumnos = data.frame(calificacion, metodo, sexo)

# Calcular la media de rendimiento por método de enseñanza

boxplot(calificacion~sexo:metodo, alumnos, col = c(2,4),
        main = "Boxplot de calificaciones por sexo y método")

# Calcular la media general de las calificaciones independientemente de los valores
# de los factores y trazar una recta horizontal sobre dicho valor

abline(h = mean(alumnos$calificacion), col = "purple")
```




```
# Calcular la media de calificaciones por método y sexo

medias_sexo_metodo = c(mean(alumnos[(alumnos$sexo=="h") &
(alumnos$metodo=="M1"), 1]),
                        mean(alumnos[(alumnos$sexo=="h") &
(alumnos$metodo=="M2"), 1]),
                        mean(alumnos[(alumnos$sexo=="h") &
(alumnos$metodo=="M3"), 1]),
                        mean(alumnos[(alumnos$sexo=="m") &
(alumnos$metodo=="M1"), 1]),
```

```
                         mean(alumnos[(alumnos$sexo=="m") &
(alumnos$metodo=="M2"), 1]),
                         mean(alumnos[(alumnos$sexo=="m") &
(alumnos$metodo=="M3"), 1]))

# Definir etiquetas para cada media calculada

names(medias_sexo_metodo) = c("h.M1", "h.M2", "h.M3", "m.M1", "m.M2",
"m.M3")

# Mostrar la media de calificaciones por grupo (sexo y método)

medias_sexo_metodo

## h.M1 h.M2 h.M3 m.M1 m.M2 m.M3
##    9    6    4    8    6    3
```

En el boxplot anterior se observa que en cuanto a los alumnos varones (boxplots en color rojo), el primer método de enseñanza M1 da como resultado calificaciones considerablemente por encima de la media general de calificaciones, siendo éstas mismas entre 7 y 10, mientras que por otro lado, en cuanto al método 2 de enseñanza para los hombres, se observa que la media de las calificaciones de dicho grupo de alumnos se ubica justamente sobre la media general de calificaciones, es decir que la media de dicho grupo es muy cercana al promedio general (6), mientras que a su vez, también se aprecia que el método de enseñanza 3 para los hombres, da como resultado calificaciones que se encuentran por debajo del promedio general, siendo éstas mismas inferiores a 6, en concreto 4. Por otra parte, en cuanto a las mujeres, se observa que con el método 1 de enseñanza, la media de las calificaciones de dicho grupo de alumnas es superior a la media general, siendo dicha media de 8, mientras que al mismo tiempo, con el método 2 de enseñanza para los mujeres se evidencia que las calificaciones de dichas estudiantes se encuentran exactamente en la media general de calificaciones, no obstante, en cuanto al método de enseñanza 3 para mujeres, se aprecia que la media de las calificaciones de dichas alumnas es inferior a la media general de 6, motivo por el cual, desde la perspectiva estadística y en el contexto del problema, se puede concluir que en base a los boxplots y las medias calculadas para cada grupo de estudiantes, que el método 1 permite que los alumnos varones aprendan ligeramente mejor que las alumnas mujeres, mientras que al mismo tiempo, el método de enseñanza 2 permite a las alumnas mujeres aprender ligeramente mejor que a los hombres, además de que el tercer método de enseñanza en apariencia resulta ser el menos eficiente para que tanto los hombres como las mujeres logren un nivel de aprendizaje adecuado, por lo que hasta este punto, es posible concluir que el sexo de los estudiantes sí parece tener un cierto grado de influencia en la manera de aprender que sea más efectiva para ellos.

2. Las hipótesis estadísticas

- Factor 1: sexo
- Factor 2: método de enseñanza

Hipótesis 1: sexo

H_0: el rendimiento de los estudiantes no se ve afectado por su sexo (las medias de las calificaciones por sexo son aproximadamente iguales entre sí).

H_1: el rendimiento de los estudiantes sí se ve afectado por su sexo (existe al menos 1 1 media de calificaciones por sexo que resulta ser distinta a las demás).

Hipótesis 2: método de enseñanza

H_0: el rendimiento de los alumnos no se ve afectado por el método de enseñanza (las medias de calificaciones por método de enseñanza son todas aproximadamente idénticas entre sí).

H_1: el rendimiento de los alumnos sí se ve afectado por el método de enseñanza (las medias de calificaciones por método de enseñanza no son todas iguales).

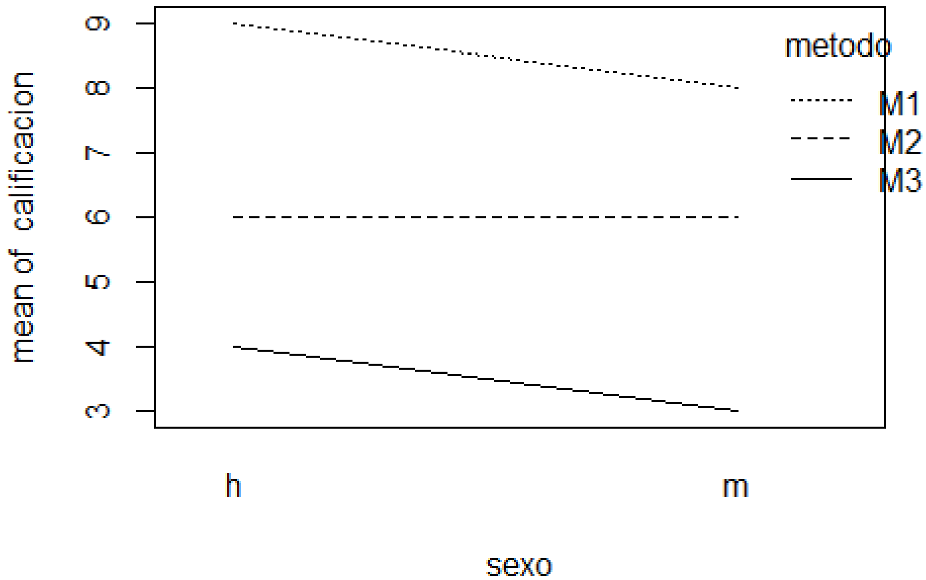
Hipótesis 3: interacción entre sexo y método de enseñanza

H_0: el rendimiento de los alumnos no se ve afectado por la combinación o interacción de su sexo y el método de enseñanza (las medias de calificaciones por sexo y método de enseñanza son todas aproximadamente iguales).

H_1: el rendimiento de los alumnos sí se ve afectado por la combinación o interacción de su sexo y el método de enseñanza (las medias de calificaciones por sexo y método de enseñanza no son todas idénticas).

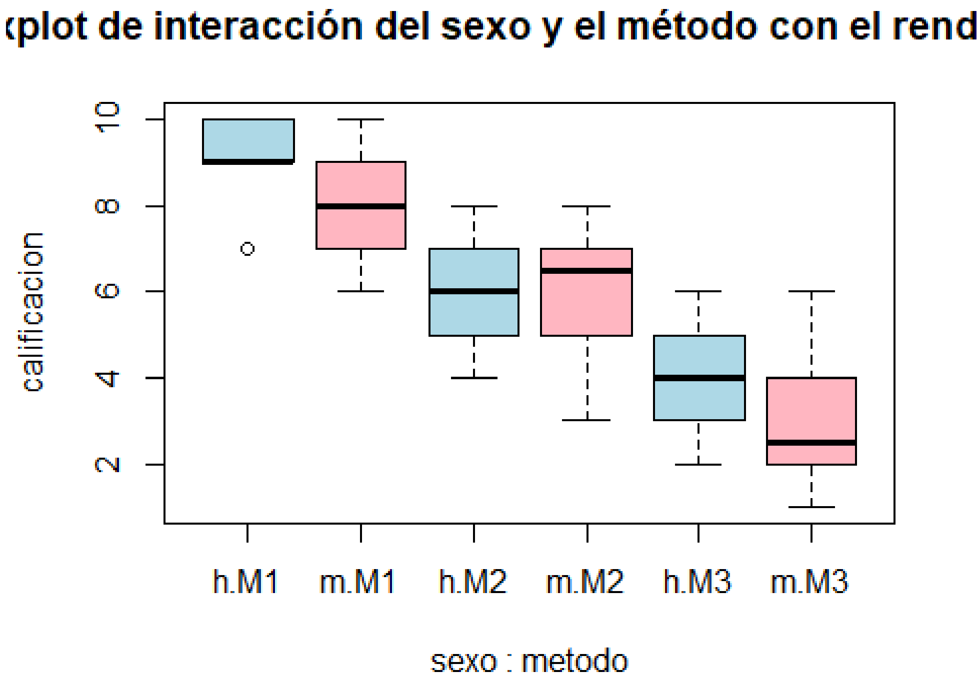
3. ANOVA para 2 niveles con interacción

```
# Realizar el ANOVA para los factores de sexo y método de enseñanza con
interacción

anova = aov(formula = calificacion ~ sexo * alumnos$metodo,
            data = alumnos)

# Resumen del modelo ANOVA

summary(anova)

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

# Gráfica de interacción

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



Boxplot para visualizar la interacción entre factores

```
# Boxplot de la interacción del sexo y el método de enseñanza con el
rendimiento o
# calificaciones de los estudiantes

boxplot(calificacion ~ sexo * metodo, data = alumnos,
        col = c("lightblue", "lightpink"),
        main = "Boxplot de interacción del sexo y el método con el
rendimiento")
```



En cuanto a los boxplots previos y el análisis ANOVA realizado previamente, es posible apreciar que el valor p correspondiente al factor del sexo es igual a 0.2, lo cual al ser superior a 0.05 (nivel de significancia), no se rechaza H_0 de la prueba de hipótesis correspondiente al sexo, por lo cual, se afirma estadísticamente que el factor del sexo no resulta ser significativo en el rendimiento de los estudiantes, sin embargo, también se observa que el segundo factor que corresponde al método de enseñanza, tiene un valor p de 3.47e-08, mismo que al ser menor que 0.05, se rechaza H_0 de la prueba de hipótesis correspondiente al método de enseñanza, por lo cual, se afirma estadísticamente que el método de enseñanza sí es significativo en el rendimiento de los alumnos, mientras que al mismo tiempo, el valor p corespondiente a la interacción del sexo y el método de enseñanza es de 0.655, lo cual al ser mayor a 0.05, no se rechaza H_0 de la prueba de hipótesis correspondiente a la interacción entre sexo y método de enseñanza, por lo tanto, se afirma estadísticamente que la interacción del sexo con el método de enseñanza no es significativa, en otras palabras, ambos factores no se afectan mutuamente, por lo que existe independencia entre ellos, además de que también la interacción entre sexo y método de enseñanza no tiene efecto en el rendimiento de los alumnos.

En conclusión, el rendimiento o calificaciones de los estudiantes no se ve condicionado o afectado por su sexo pero sí por el método de enseñanza al que se someten, además de que el sexo y el método de enseñanza no tienen una relación entre sí, es decir, no influyen uno en el otro de acuerdo con el análisis ANOVA realizado previamente, por lo cual a continuación es necesario analizar con más detenimiento la relación existente entre el método de enseñanza al que se someten los estudiantes y su rendimiento medido en forma de calificaciones.

4. ANOVA para 2 factores sin interacción

```r
# Realizar el análisis ANOVA ahora sin interacción entre los factores

anova_no_interaccion = aov(calificacion ~ metodo + sexo, data = alumnos)

# Resumen del modelo ANOVA

summary(anova_no_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
```

```r
# Calcular la media de calificaciones por sexo

medias_sexo = tapply(calificacion, sexo, mean)

medias_sexo
```

```
##        h        m
## 6.333333 5.666667
```

```r
# Calcular la media de calificaciones por método de enseñanza

medias_metodo = tapply(calificacion, metodo, mean)

medias_metodo
```

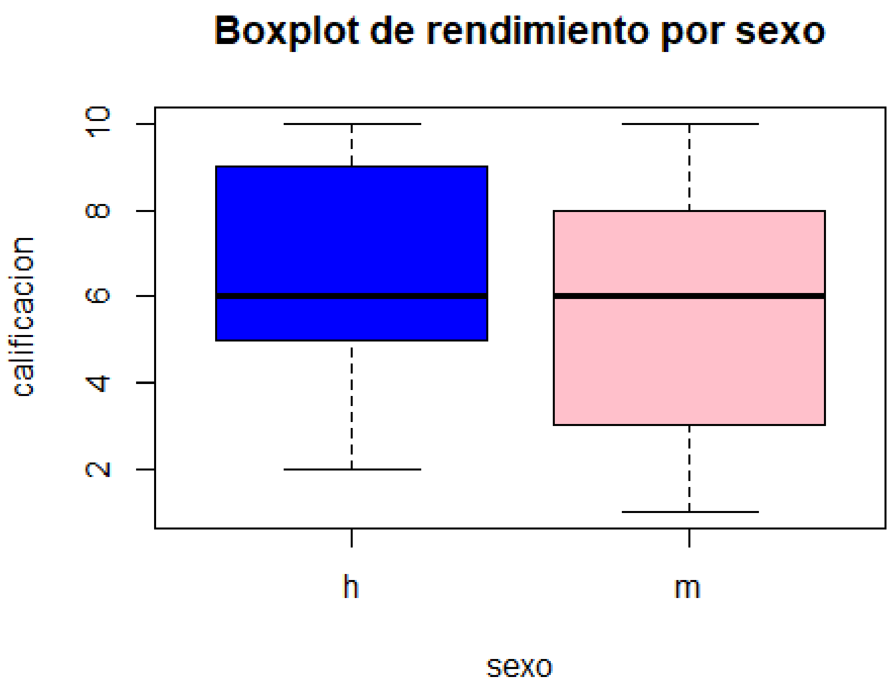
```
##  M1  M2  M3
## 8.5 6.0 3.5
```

```r
# Calcular media general de calificaciones

media_general = mean(alumnos$calificacion)

cat("Media general de calificaciones: ", media_general)
```

```
## Media general de calificaciones:  6
```

Boxplot de rendimiento por sexo

```r
# Graficar el boxplot para las calificaciones por sexo

boxplot(calificacion ~ sexo, data = alumnos, col = c("blue", "pink"),
        main = "Boxplot de rendimiento por sexo")
```



Intervalos de confianza de rendimiento por sexo y gráfico

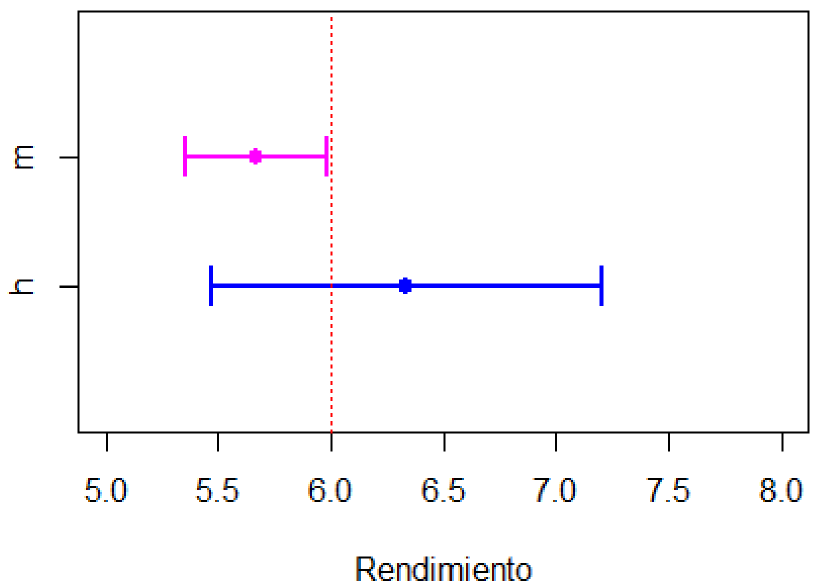
Se necesitará calcular el margen de error E para la media tanto de mujeres como de hombres, mediante la siguiente expresión:

$$E = \frac{z_{\alpha/2}\sigma_x}{\sqrt{n}}$$

```
# Calcular desviación estándar de calificaciones por sexo

desv_sexo = tapply(calificacion, sexo, sd)

desv_sexo

##        h        m
## 2.473388 2.634611

# Cantidad de datos de alumnos hombres

nh = nrow(alumnos[alumnos$sexo == "h", ])

# Cantidad de datos de alumnas mujeres

nm = nrow(alumnos[alumnos$sexo == "m", ])

# Calcular el margen de error E para la media de calificaciones de los
hombres
```

```r
Eh = abs((qnorm(0.05 / 2, medias_sexo["h"], desv_sexo["h"]) *
desv_sexo["h"])/sqrt(nh))

# Calcular margen de error E para la media de calificaciones de las
mujeres

Em = abs((qnorm(0.05 / 2, medias_sexo["m"], desv_sexo["m"]) *
desv_sexo["m"])/sqrt(nm))

# Mostrar intervalo de confianza para hombres

cat("IC de rendimiento para hombres: ", medias_sexo["h"] - Eh, " - ",
    medias_sexo["h"] + Eh, "\n")

## IC de rendimiento para hombres:  5.467264  -  7.199403

# Mostrar intervalo de confianza para mujeres

cat("IC de rendimiento para mujeres: ", medias_sexo["m"] - Em, " - ",
    medias_sexo["m"] + Em)

## IC de rendimiento para mujeres:  5.354359  -  5.978974

# Asignar límites inferiores y superiores al gráfico de ICs en ambos ejes

plot(0, ylim=c(0, 2 + 1), xlim=c(5, 8), yaxt="n", ylab="", xlab =
"Rendimiento")

# Asignar etiquetas al eje vertical del gráfico

axis(2, at=c(1, 2), labels=c("h", "m"))

# Límite inferior y superior del IC para hombres

arrows(medias_sexo["h"] - Eh, 1, medias_sexo["h"] + Eh, 1, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "blue")

# Límite inferior y superior del IC para mujeres

arrows(medias_sexo["m"] - Em, 2, medias_sexo["m"] + Em, 2, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "magenta")

# Grafiar punto sobre la media de hombres

points(medias_sexo["h"], 1, pch = 19, cex = 1.1, col = "blue")

# Graficar punto sobre la media de mujeres
```

```r
points(medias_sexo["m"], 2, pch = 19, cex = 1.1, col = "magenta")

# Trazar una recta vertical sobre el valor de la media general de
calificaciones

abline(v = media_general, lty = 3, col = "red")
```



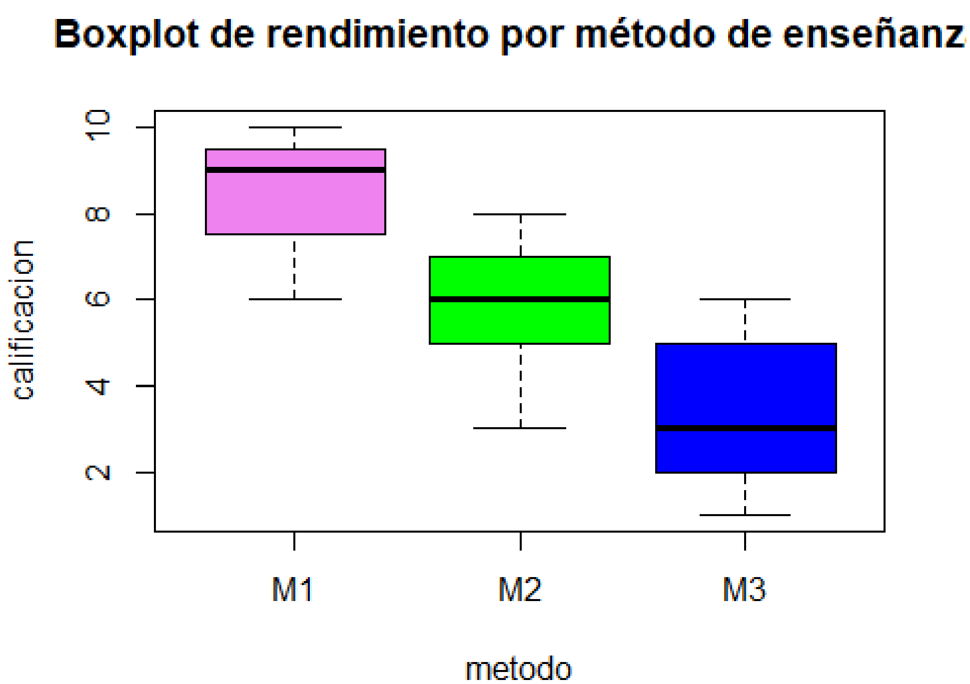
De acuerdo con el boxplot realizado anteriormente, se puede observar que la anchura del intervalo de confianza del rendimiento de las mujeres es significativamente menor que la del intervalo de confianza del rendimiento de los hombres, además de que también la media general de las calificaciones se encuentra contenida únicamente en el intervalo del rendimiento de los hombres, pero no dentro del de las mujeres, motivo por el cual, estadísticamente hablando, la media del rendimiento los hombres fue superior a la de las mujeres, además de que también la calificación mínima y máxima obtenidas por los alumnos hombres fueron ligeramente mayores a las de las mujeres, además de que la mayoría de los hombres obtuvieron calificaciones que fueron superiores a 6.3, mientras que las mujeres tuvieron calificaciones que fueron superiores a 5.6 pero a su vez inferiores a 6.3, lo cual respalda el hecho de que los hombres tuvieron un mejor rendimiento que las mujeres en términos generales.

En conclusión, el rendimiento de los alumnos varones fue considerablemente mejor al de las alumnas mujeres, debido a que el promedio de los hombres resulta ser superior al de las mujeres, además de que también el promedio de los hombres tuvo una mayor variación que el de las mujeres, esto nos indica que las mujeres obtuvieron en realidad calificaciones muy cercanas entre sí, mientras que los hombres por el contrario, obtuvieron resultados considerablemente más alejados entre sí que las mujeres, lo cual sugiere que el método de enseñanza que resultó más efectivo fue en realidad más efectivo para fomentar el aprendizaje de los hombres que el de las mujeres.

5. ANOVA para un efecto principal (método de enseñanza)

Boxplot de rendimiento por método de enseñanza

```r
# Graficar boxplot de calificaciones (rendimiento) por cada método de
enseñanza

boxplot(calificacion ~ metodo, data = alumnos, col = c("violet", "green",
"blue"),
        main = "Boxplot de rendimiento por método de enseñanza")
```




```r
# Media de las calificaciones por método de enseñanza

print("Media de calificaciones por método de enseñanza:")

## [1] "Media de calificaciones por método de enseñanza:"

medias_metodo

##  M1  M2  M3
## 8.5 6.0 3.5

# Realizar el ANOVA pero ahora considerando un solo efecto principal
(método) en vez
```

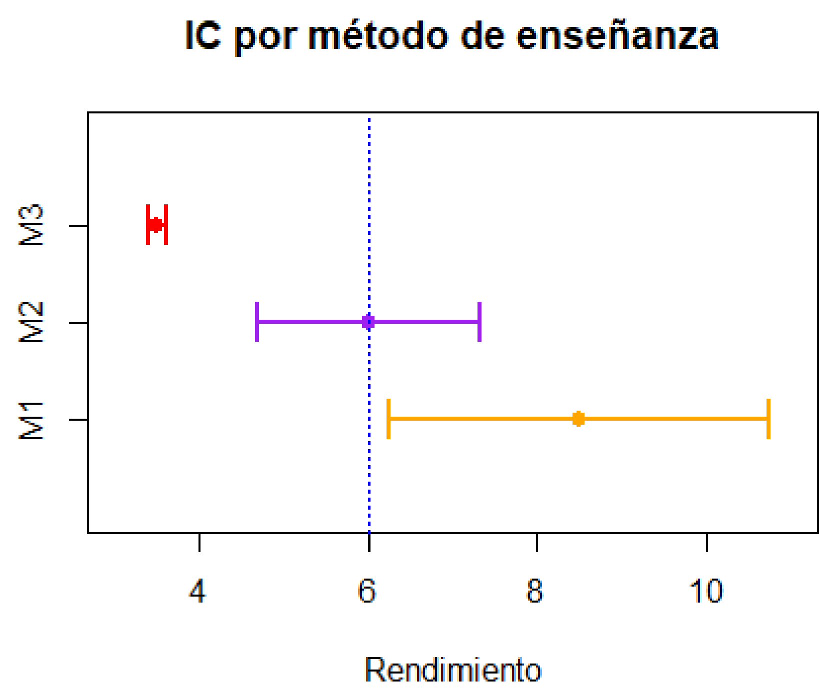
```r
# de ambos (método y sexo)

anova_1_efecto = aov(calificacion ~ metodo, data = alumnos)

# Mostrar resumen del modelo ANOVA

summary(anova_1_efecto)

##             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
```

Intervalos de confianza de rendimiento por método

```r
# Calcular desviación estándar de calificaciones por método

desv_metodo = tapply(calificacion, metodo, sd)

print("Desviación estándar por método de enseñanza:")

## [1] "Desviación estándar por método de enseñanza:"

desv_metodo

##       M1       M2       M3 
## 1.314257 1.537412 1.678744

# Definir cantidad de registros de calificaciones con los 3 métodos

n1 = length(alumnos[alumnos$metodo == "M1",]$calificacion) # metodo 1

n2 = length(alumnos[alumnos$metodo == "M2",]$calificacion) # metodo 2

n3 = length(alumnos[alumnos$metodo == "M3",]$calificacion) # metodo 3

# Calcular el margen de error E para la media de calificaciones del
metodo 1

EM1 = abs((qnorm(0.05 / 2, medias_metodo["M1"],
                 desv_metodo["M1"]) * desv_metodo["M1"])/sqrt(n1))

# Calcular margen de error E para la media de calificaciones del metodo 2

EM2 = abs((qnorm(0.05 / 2, medias_metodo["M2"],
                 desv_metodo["M2"]) * desv_metodo["M2"])/sqrt(n2))

# Calcular margen de error E para la media de calificaciones del metodo 3
```

```r
EM3 = abs((qnorm(0.05 / 2, medias_metodo["M3"],
                 desv_metodo["M3"]) * desv_metodo["M3"])/sqrt(n3))

# Mostrar intervalo de confianza para metodo 1

cat("IC de rendimiento para M1: ", medias_metodo["M1"] - EM1, " - ",
    medias_metodo["M1"] + EM1, "\n")

## IC de rendimiento para M1:  6.252434  -  10.74757

# Mostrar intervalo de confianza para metodo 2

cat("IC de rendimiento para M2: ", medias_metodo["M2"] - EM2, " - ",
    medias_metodo["M2"] + EM2, "\n")

## IC de rendimiento para M2:  4.674453  -  7.325547

# Mostrar intervalo de confianza para metodo 3

cat("IC de rendimiento para M3: ", medias_metodo["M3"] - EM3, " - ",
    medias_metodo["M3"] + EM3, "\n")

## IC de rendimiento para M3:  3.398366  -  3.601634

# Asignar límites inferiores y superiores al gráfico de ICs en ambos ejes

plot(0, ylim=c(0, 3 + 1), xlim=c(3, 11), yaxt="n", ylab="",
     xlab = "Rendimiento", main = "IC por método de enseñanza")

# Asignar etiquetas al eje vertical del gráfico correspondientes a los
métodos
# de enseñanza

axis(2, at=c(1, 2, 3), labels=c("M1", "M2", "M3"))

# Límite inferior y superior del IC para M1

arrows(medias_metodo["M1"] - EM1, 1, medias_metodo["M1"] + EM1, 1, angle
= 90,
       code = 3, length = 0.1, lwd = 2, col = "orange")

# Límite inferior y superior del IC para M2

arrows(medias_metodo["M2"] - EM2, 2, medias_metodo["M2"] + EM2, 2, angle
= 90,
       code = 3, length = 0.1, lwd = 2, col = "purple")

# Límite inferior y superior del IC para M3

arrows(medias_metodo["M3"] - EM3, 3, medias_metodo["M3"] + EM3, 3, angle
```

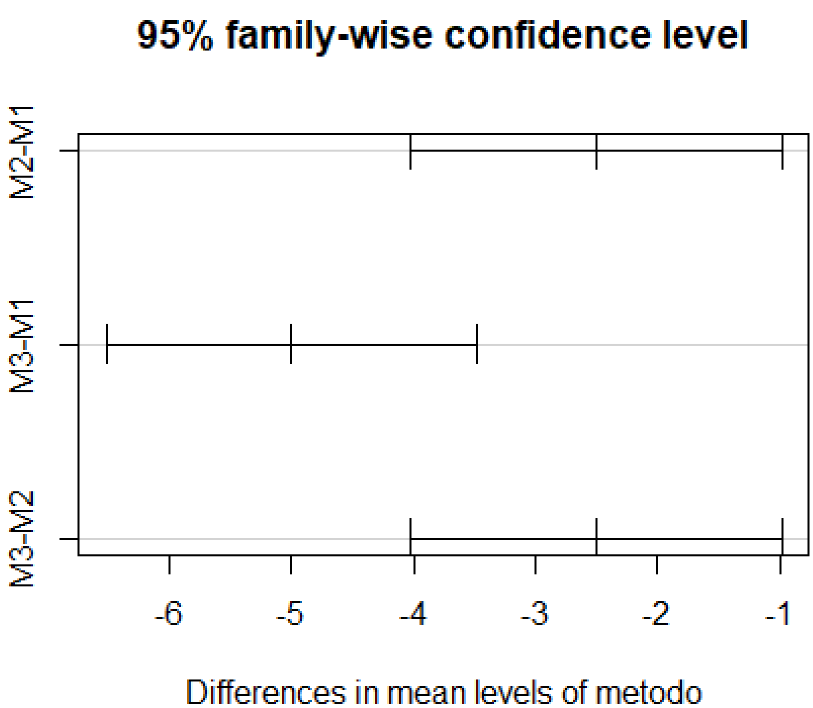
```r
= 90,
       code = 3, length = 0.1, lwd = 2, col = "red")

# Grafiar punto sobre la media de M1

points(medias_metodo["M1"], 1, pch = 19, cex = 1.1, col = "orange")

# Graficar punto sobre la media de M2

points(medias_metodo["M2"], 2, pch = 19, cex = 1.1, col = "purple")

# Graficar punto sobre la media de M3

points(medias_metodo["M3"], 3, pch = 19, cex = 1.1, col = "red")

# Trazar una recta vertical sobre el valor de la media general de
calificaciones

abline(v = media_general, lty = 3, col = "blue")
```



Prueba de comparaciones múltiples de Tukey

```r
# Realizar la prueba de Tukey para comparar las medias de los métodos
(todas contra todas)

tukey = TukeyHSD(anova_1_efecto)
```

```r
# Mostrar resultados de la prueba

tukey

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = calificacion ~ metodo, data = alumnos)
##
## $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 intervalos de confianza resultantes de la prueba

plot(tukey)
```



En relación a los resultados previmente obtenidos, es posible apreciar que tras efectuar la prueba de Tukey para comparar las 3 medias entre sí, las 3 medias comparadas tienen todas un valor p menor a 0.05, motivo por el cual, se cuenta con suficiente evidencia estadística para rechazar H_0 de la prueba de hipótesis correspondiente al factor del método de enseñanza, por lo tanto, se puede afirmar que en el contexto estadístico, las 3 medias de calificaciones correspondientes a cada uno de los métodos de enseñanza son significativas, por lo tanto, en el contexto del problema se puede afirmar que el método de enseñanza efectivamente sí afecta el rendimiento de los estudiantes que se someten al mismo.

Finalmente, a manera de conclusión, se puede afirmar que debido al hecho de que el método de enseñanza al que se someten los alumnos sí influye en su rendimiento escolar, se presentarán variaciones en cuanto al método que mejor le funciona a cada uno de los estudiantes para aprender de manera efectiva, motivo por el cual, en un contexto práctico, será necesario que el docente adapte la explicación de una temática específica de tal forma que se adapte lo mejor posible a algunas de las distintas y diversas formas de aprender que pueden tener los alumnos para que éstos puedan consolidar lo mejor posible los conocimientos adquiridos en clase.

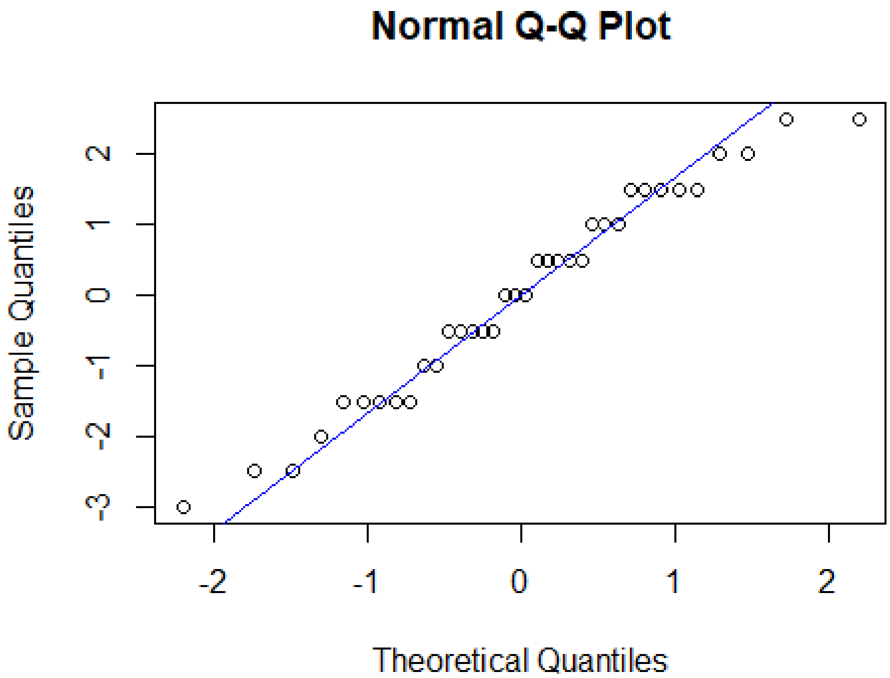
6. Comprobar validez del modelo

Para validar que el modelo ANOVA con el único efecto principal significativo sea estadísticamente válido y confiable, será necesario validar los siguientes supuestos:

- Normalidad de los residuos
- Homocedasticidad de los residuos (los residuos poseen varianza constante)
- Independencia de los residuos (no hay presencia de correlación entre residuos)

Normalidad

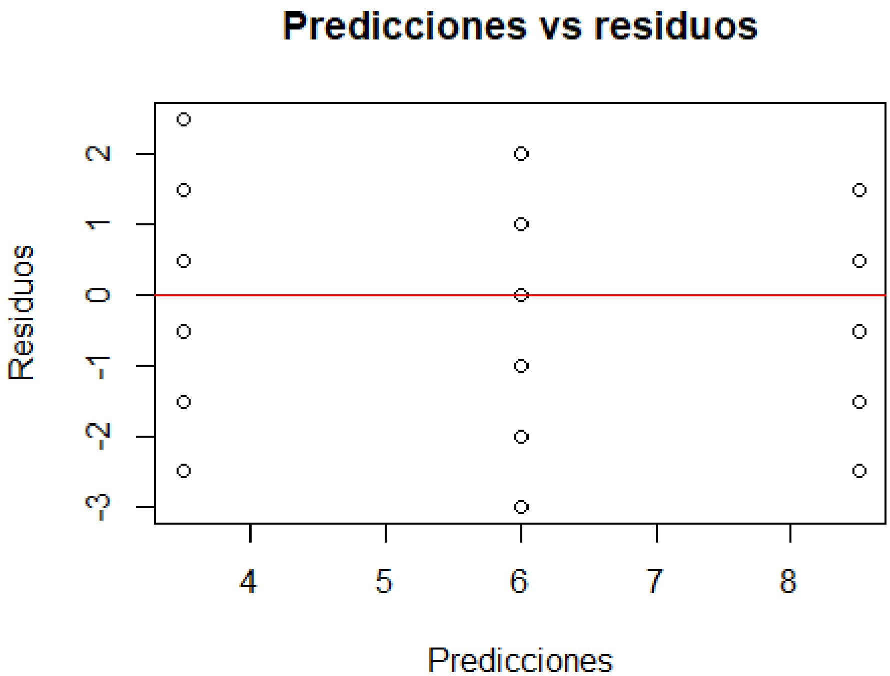
```r
# Para verificar que los residuos sigan una distribución normal, se
realizará
# el QQ-plot de los residuos del modelo ANOVA

# Graficar percentiles de normalidad de los residuos

qqnorm(anova_1_efecto$residuals)

# Graficar la recta correspondiente a la distribución normal

qqline(anova_1_efecto$residuals, col = "blue")
```



En el QQ-plot de los residuos del ANOVA se observa que los residuos del ANOVA se ajustan en su mayoría adecuadamente a la recta en color azul que representa la distribución normal, no obstante también se aprecia que están presentes ciertos datos que se alejan de dicha recta, sin embargo, al ser una cantidad reducida de datos la que se aleja de la recta de normalidad, pero la mayoría de todos los datos (percentiles) sí se ajustan bien a la recta de normalidad, motivo por el cual, se concluye que los residuos del modelo ANOVA generado con el efecto principal significativo (método de enseñanza), siguen en su gran mayoría una distribución normal.

Homocedasticidad

```
# Para verificar si hay presencia de homocedasticidad en los residuos del
ANOVA, se realizará una gráfica de los valores predichos por el modelo y
los residuos del mismo

plot(anova_1_efecto$fitted.values, anova_1_efecto$residuals,
     ylab="Residuos",xlab="Predicciones", main = "Predicciones vs
residuos")

# Graficar una recta horizontal en el valor 0

abline(h = 0, col = "red")
```

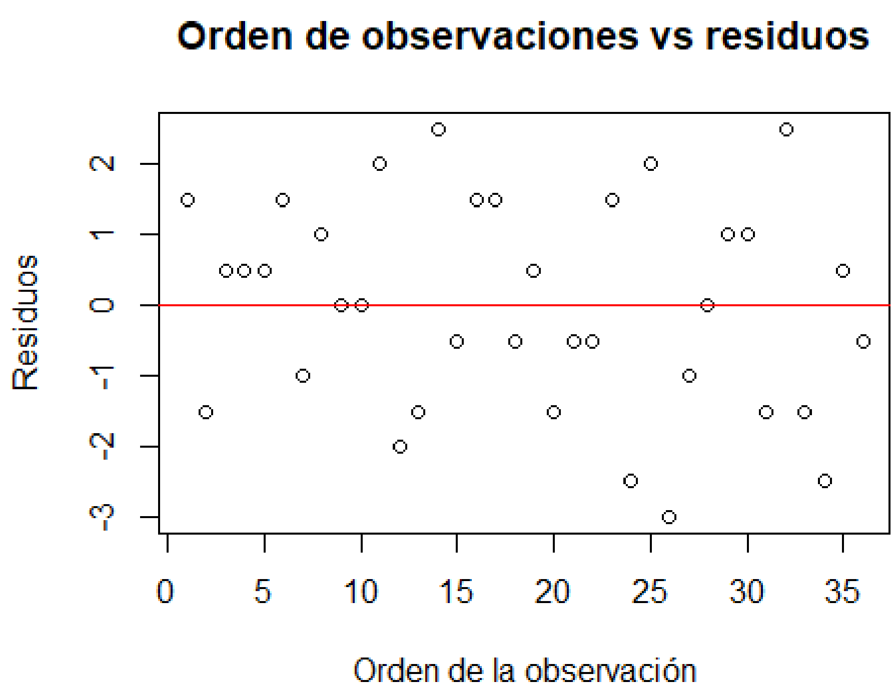


En cuanto a la homocedasticidad, se observa en el gráfico de predicciones contra residuos, que la dispersión de los puntos (residuos) a lo largo del gráfico resulta ser aproximadamente la misma, ya que se aprecia que los residuos se encuentran prácticamente alineados entre sí y cada grupo de puntos se encuentra distribuido de una forma aproximadamente equitativa en el gráfico, motivo por el cual, se puede concluir en base al gráfico anterior que los residuos del modelo ANOVA con el efecto principal significativo tienen una dispersión constante en toda la gráfica, lo cual significa que es posible afirmar que los residuos del ANOVA tienen una varianza constante u homocedasticidad.

Independencia

```
# Para validar si hay presencia de independencia entre los residuos del modelo ANOVA, se
# realizará una gráfica de las observaciones ordenadas conforme fueron registradas contra
# los residuos del modelo

# Definir variable tiempo con valores desde 1 hasta la cantidad de datos del dataset
# alumnos (36 datos)

tiempo = 1:nrow(alumnos)

# Cantidad de datos registrados por método de enseñanza

n = tapply(tiempo, metodo, length)
```

```r
# Graficar las observaciones contra sus cprrespondientes residuos en el orden en el que
# fueron registradas

plot(tiempo, anova_1_efecto$residuals, xlab="Orden de la observación",
     ylab="Residuos", main = "Orden de observaciones vs residuos")

# Línea horizontal en 0

abline(h=0,col="red")
```



En el gráfico de orden de observaciones contra residuos, se aprecia que los puntos del gráfico (residuos), se encuentran todos dispersos a lo largo del mismo, sin seguir alguna tendencia o patrón en particular, por lo cual dicha dispersión se puede atribuir prácticamente al azar y no a alguna dependencia entre las observaciones ordenadas y sus residuos corespondientes, motivo por el cual, se concluye que los residuos del modelo ANOVA muestran independencia entre ellos.

Relación lineal entre variables (coeficiente de determinación)

```r
# Calcular el coeficiente de determinación entre las calificaciones de los alumnos y
# el método de enseñanza al que se someten

r2 = 150 / (150 + 76)
```

```r
cat("Coeficiente de determinación: ", r2)

## Coeficiente de determinación:  0.6637168
```

Al calcular el coeficiente de determinación R^2 se aprecia que éste mismo tiene un valor de 0.6637, lo cual indica que el modelo es capaz de explicar el 66.37% de la variabilidad total de los datos analizados, en otras palabras, la proporción de variación de la variable de calificaciones que es explicada por el método de enseñanza es del 66.37%, lo cual significa que la relación lineal entre ambas variables resulta ser medianamente adecuada ya que aún posee un porcentaje considerable de error (33.63%).

Problema 2

Un ingeniero de procesos ha identificado dos causas potenciales de vibración de los motores eléctricos, el material utilizado para la carcasa del motor (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.

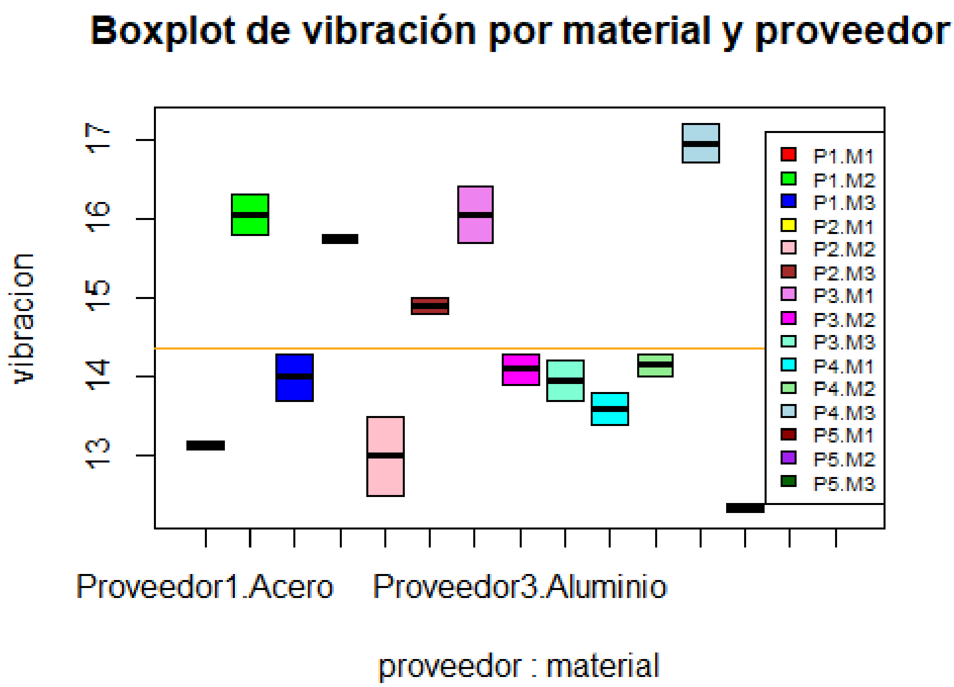
```r
# Factor A: materiales de la carcasa

material = c(rep("Acero", 10), rep("Aluminio", 10), rep("Plastico", 10))

# Factor B: proveedores de cojinetes

proveedor = c(rep(c("Proveedor1", "Proveedor2", "Proveedor3",
                    "Proveedor4", "Proveedor5"), 6))

# 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)
```

```r
# Crear dataframe con los datos anteriores

datos = data.frame(vibracion,material, proveedor)

colnames(datos) = c("vibración", "material", "proveedor")

# Mostrar dataframe con los datos anteriores

head(datos)

##   vibración 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
```

1. Análisis exploratorio

```r
# Calcular la media de vibraciones por material

medias_material = tapply(vibracion, material, mean)

print("Media de vibraciones por material:")

## [1] "Media de vibraciones por material:"

medias_material

##    Acero Aluminio Plastico
##    14.39    14.52    14.15

# Calcular la media de vibraciones por proveedor

medias_proveedor = tapply(datos$vibración, proveedor, mean)

print("Media de vibraciones por proveedor:")

## [1] "Media de vibraciones por proveedor:"

medias_proveedor

## Proveedor1 Proveedor2 Proveedor3 Proveedor4 Proveedor5
##   14.06667   16.35000   13.48333   14.61667   13.25000

# Calcular media de vibraciones por proveedor y material

medias_prov_mat = tapply(datos$vibración, list(proveedor, material),
mean)
```

```
print("Media de vibraciones por proveedor y material:")

## [1] "Media de vibraciones por proveedor y material:"

medias_prov_mat

##            Acero Aluminio Plastico
## Proveedor1 13.15    14.90    14.15
## Proveedor2 16.05    16.05    16.95
## Proveedor3 14.00    14.10    12.35
## Proveedor4 15.75    13.95    14.15
## Proveedor5 13.00    13.60    13.15

boxplot(vibracion ~ proveedor:material, datos,
        main = "Boxplot de vibración por material y proveedor",
        col = c("red", "green", "blue", "yellow", "pink", "brown",
                "violet", "magenta", "aquamarine", "cyan", "lightgreen",
                "lightblue", "darkred", "purple", "darkgreen"))

# Calcular la media general de la vibración independientemente de los
valores
# de los factores y trazar una recta horizontal sobre dicho valor

abline(h = mean(datos$vibración), col = "orange")

# Agregar una leyenda al gráfico para identificar el número de proveedor
y el material

legend(x = "right", y = "top", legend = c("P1.M1", "P1.M2", "P1.M3",
"P2.M1", "P2.M2",
                                  "P2.M3", "P3.M1", "P3.M2", "P3.M3",
"P4.M1",
                                  "P4.M2", "P4.M3", "P5.M1", "P5.M2",
"P5.M3"),
       fill = c("red", "green", "blue", "yellow", "pink", "brown",
                "violet", "magenta", "aquamarine", "cyan", "lightgreen",
                "lightblue", "darkred", "purple", "darkgreen"),
       horiz = F, cex = 0.6, xpd = T)
```



De acuerdo al boxplot anterior, se aprecia que la vibración resulta tener una cantidad de variación mayormente significativa en el material 2 provisto por el proveedor 2, que corresponde al aluminio del proveedor 2, ya que se aprecia que el boxplot tiene una anchura medianamente considerable con respecto a otros boxplots, además, también se aprecia que el boxplot de color violeta correspondiente al material 1 del proveedor 3 (acero del proveedor 3), también tiene una anchura medianamente considerable, motivo por el cual, las vibraciones en dicho material también son más variable que en otros materiales provistos por otros proveedores.

Adicionalmente, también es evidente que algunos boxlots son muy pequeños, lo cual indica que la variación de las vibraciones presentes en los materiales asociados a dichos boxplots son mucha menor magnitud que las vibraciones de los materiales descritos previamente que tenían más variación en cuanto a sus vibraciones, además, entre aquellos materiales que presentan muy escasas vibraciones se encuentran en primer lugar, el acero del proveedor 1 (P1.M1), el acero del proveedor 2 (P2.M1), plástico del proveedor 2 (P2.M3), y el acero del proveedor 5 (P5.M1), por lo cual, desde un punto de vista estadístico, se puede afirmar que el promedio de vibraciones de los materiales: aluminio del proveedor 2 y el acero del proveedor 3, poseen una media de vibraciones superior a la media de vibraciones de los materiales: acero del proveedor 1, acero del proveedor 2, plástico del proveedor 2, acero del proveedor 5, además del hecho de que también se observa que 9 materiales se encuentran por debajo de la media general de vibraciones, los cuales son: el acero y el plástico de proveedor 1, el aluminio del proveedor 2, el aluminio y plástico del proveedor 3, el acero y aluminio del proveedor 4, junto con el acero y aluminio del proveedor 5, mientras que el resto de materiales por el contrario tienen vibraciones superiores a la media, por lo cual, en el contexto del problema es posible concluir a partir de éste análisis que los proveedores que proveen la mayoría de los productos con vibraciones por debajo de la media, son el proveedor 1, en concreto el acero y el plástico, además del proveedor 3, en concreto el aluminio y el plástico, además del proveedor 4, que provee específicamente acero y aluminio con vibraciones por debajo de la media, incluyendo el proveedor 5 que también provee productos cuyas vibraciones se encuentran también por debajo de la media general, específicamente el acero y el aluminio, motivo por el cual, considerando todo lo anterior, es posible concluir que el proveedor de cojinetes así como los materiales provistos por ellos parecen tener cierta influencia en la cantidad de vibraciones de los materiales utilizados para la fabricación de los motores, por lo cual será mejor recibir el material de cada tipo que menos vibraciones genere al momento de fabricar los motores, por lo cual, en conclusión, será más conveniente recibir acero del proveedor 1 y del proveedor 4 y 5, ya que son los que menos vibraciones generan, mientras que a su vez será más conveniente recibir el aluminio del proveedor 2, del proveedor 3, del proveedor 4 y del proveedor 5, mientras que en cuanto al plástico, es mejor recibirlo de los proveedores 1 y 3.

2. Hipótesis estadísticas

Hipótesis para material

H_0: las vibraciones de los motores no se ven afectadas por el material con el que se fabrican (las medias de vibraciones por material son todas aproximadamente iguales entre ellas).

H_1: las vibraciones de los motores sí se ven afectadas por el material con el que se fabrican (las medias de vibraciones por material no son todas iguales).

Hipótesis para proveedor

H_0: las vibraciones de los motores no se ven afectadas por el proveedor del material (las medias de vibraciones por proveedor son todas aproximadamente idénticas).

H_1: las vibraciones de los motores sí se ven afecadas por el proveedor del material (las medias de las vibraciones por proveedor no son todas idénticas).

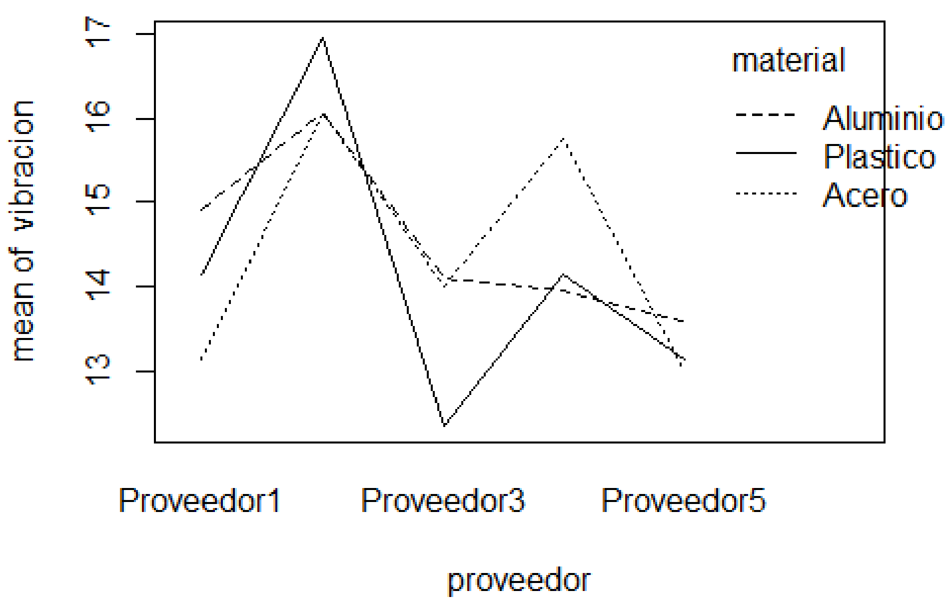
Hipótesis para la interacción entre material y proveedor

H_0: las vibraciones de los motores no se ven afectadas por la interacción o combinación del proveedor y el material (las medias de vibrciones por proveedor y material son todas aproximadamente idénticas).

H_1: las vibraciones de los motores sí se ven afectadas por la interacción o combinación del proveedor y el material (las medias de vibraciones por proveedor y material no son todas iguales).

3. ANOVA para 2 niveles con interacción:

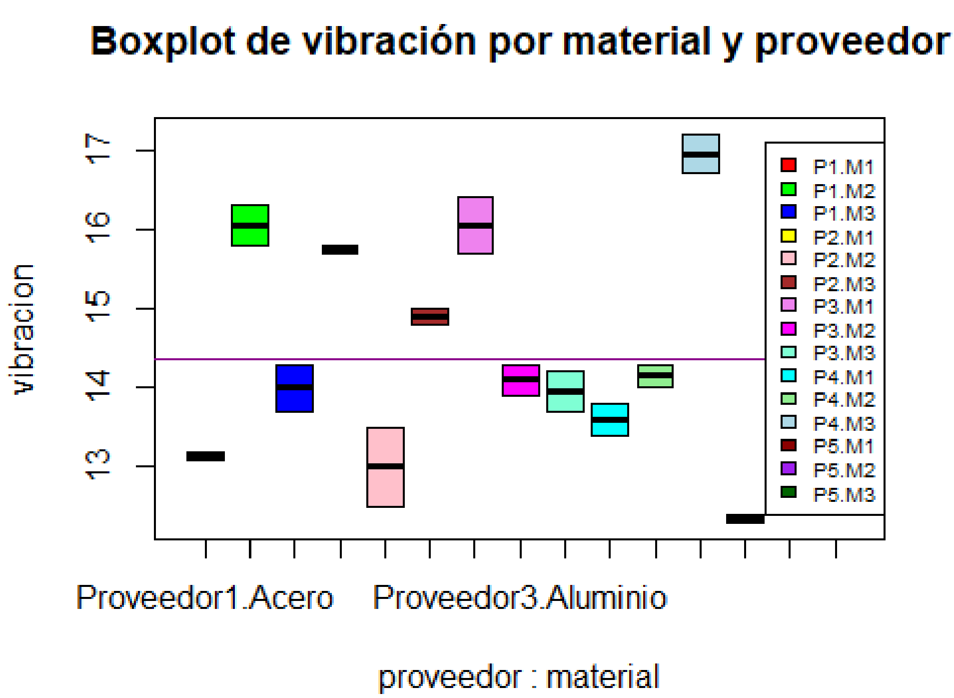
```
# ANOVA con 2 factores que interactúan entre ellos

anova_p2_interaccion = aov(vibracion ~ proveedor * material, data =
datos)

# Resumen del modelo ANOVA generado

summary(anova_p2_interaccion)

##                    Df Sum Sq Mean Sq F value   Pr(>F)
## proveedor           4  36.67   9.169  82.353 5.07e-10 ***
## material            2   0.70   0.352   3.165   0.0713 .
## proveedor:material  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

# Realizar la gráfica de interacción de los factores material y proveedor
(2 factores)

interaction.plot(proveedor, material, vibracion)
```



Boxplot para visualizar la relación de proveedor y material con vibraciones

```
# Boxplot para ver la relación de material y proveedor con las
vibraciones
```

```
boxplot(vibracion ~ proveedor * material, datos,
        main = "Boxplot de vibración por material y proveedor",
        col = c("red", "green", "blue", "yellow", "pink", "brown",
                "violet", "magenta", "aquamarine", "cyan", "lightgreen",
                "lightblue", "darkred", "purple", "darkgreen"))

# Calcular la media general de la vibración independientemente de los
valores
# de los factores y trazar una recta horizontal sobre dicho valor

abline(h = mean(datos$vibración), col = "darkmagenta")

# Agregar una leyenda al gráfico para identificar el número de proveedor
y material

legend(x = "right", y = "top", legend = c("P1.M1", "P1.M2", "P1.M3",
"P2.M1", "P2.M2",
                                  "P2.M3", "P3.M1", "P3.M2", "P3.M3",
"P4.M1",
                                  "P4.M2", "P4.M3", "P5.M1", "P5.M2",
"P5.M3"),
       fill = c("red", "green", "blue", "yellow", "pink", "brown",
                "violet", "magenta", "aquamarine", "cyan", "lightgreen",
                "lightblue", "darkred", "purple", "darkgreen"),
       horiz = F, cex = 0.6, xpd = T)
```



En cuanto a los resultados anteriores, es posible observar que desde la perspectiva estadística, la media de vibraciones correspondiente al aluminio del proveedor 2 es una de las medias más pequeñas de todos los materiales de los distintos proveedores, al igual que la media correspondiente a acero de los proveedores 1, 4 y 5, al igual que la media del aluminio y del acero del proveedor 5, entre otros materiales, motivo por el cual, lo anterior sugiere que los materiales de aquellos proveedores mencionados con anterioridad, posiblemente sean de una mejor calidad que aquellos cuya media de vibraciones sobrepasa el promedio general de vibraciones, por lo cual, dentro del contexto del problema, es posible afirmar que aquel material que tuvo el menor promedio de vibraciones considerando a los 5 proveedores, es el plástico del proveedor 3, mismo que tiene un promedio de 12.35, por lo que resulta ser el tipo material más adecuado para fabricar los motores debido a su cantidad de vibraciones relativamente bajo, sin embargo, en cuanto al material con el máximo promedio de vibraciones se encuentra el plástico pero del proveedor 2, teniendo una media de vibraciones de 16.95, por lo que resulat ser el tipo de material menos indicado para fabricar motores dado su alta cantidad de vibraciones que se producen al utilizarlo, no obstante, también hay otros materiales que no tienen el mayor ni el menor número de vibraciones, es decir que su calidad es intermedia y por tanto también podrían utlizarse para fabricar motores, entre ellos se encuentran el acero del proveedor 2, el acero del proveedor 4 y el aluminio del proveedor 2, teniendo como media de vibraciones: 16.05, 15.75 y 16.05, respectivamente.

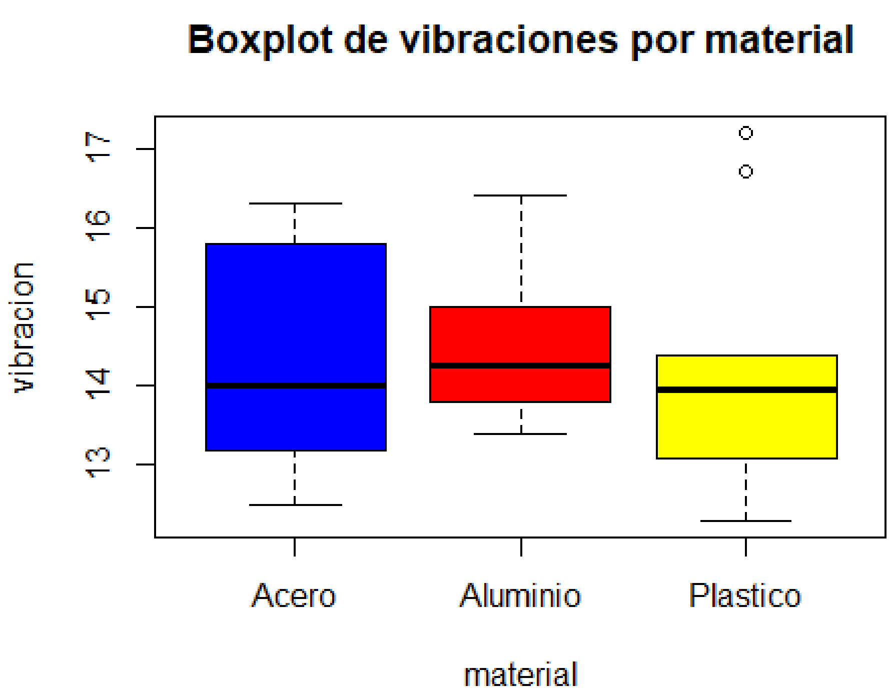
Tomando en cuenta todo lo anterior, es posible concluir que en base al modelo ANOVA generado (con interacción), en el resumen o summary del mismo se observa que el valor p del proveedor es igual a 5.07e-10, lo cual al ser menor que 0.05, se tiene suficiente evidencia estadística para rechazar H_0 de la prueba de hipótesis correspondiente a la influencia del proveedor en el número de vibraciones de los motores, por lo que estadísticamente hablando, se concluye que ls cojinetes provistos por el proveedor sí son significativos en la cantidad de vibraciones de los motores, por lo cual, como consecuencia, el proveedor sí influye en las vibraciones de los motores debido a los materiales que provee. Por otro lado, también se observa que el factor correspondiente a la interacción entre el proveedor y el material tiene un valor p de 1.76e-05, lo cual al ser inferior a 0.05, se tiene evidencia estadística suficiente para rechazar H_0 de la prueba de hipótesis correspondiente al efecto del material en el número de vibraciones de los motores, motivo por el cual, en el contexto del problema eso implica que la interacción del proveedor y el material con el que se fabrican los motores tiene una influencia en el número de vibraciones de los mismos, además, se aprecia que en el resumen del modelo ANOVA, el factor del material tiene un valor p de 0.0713, lo cual es ligeramente superior a 0.05, por lo cual se tiene evidencia estadística suficente para no rechazar H_0 de la prueba de hipótesis asociada al material, lo cual significa en el contexto del problema, que el material no es significativo en el número de vibraciones de los motores, sin embargo, dado que la interacción entre proveedor y material sí es significativa y el factor del material está implicado en dicha interacción, entonces el factor del material se vuelve significativo para explicar el número de vibraciones de los motores, por lo cual el factor del material no se descarta del modelo aunque tenga un valor p mayor a 0.05.

4. ANOVA para 2 niveles sin interacción

```
# Realizar el análisis ANOVA sin interacción entre el proveedor y el
material

anova_p2_NoInt = aov(vibracion ~ proveedor + material, data = datos)

# Resumen del modelo ANOVA

summary(anova_p2_NoInt)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## proveedor    4  36.67   9.169   15.88 2.28e-06 ***
## material     2   0.70   0.352    0.61    0.552
## Residuals   23  13.28   0.577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

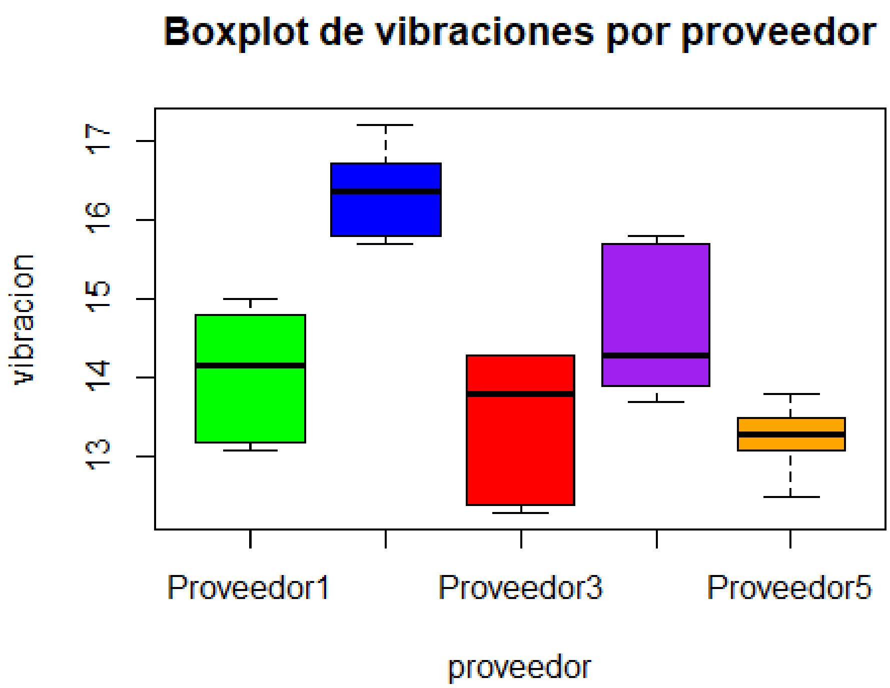
Boxplot de vibraciones por material

```
# Graficar el boxplot para la cantidad de vibraciones por material

boxplot(vibracion ~ material, data = datos, col = c("blue", "red",
"yellow"),
        main = "Boxplot de vibraciones por material")
```



Boxplot de vibraciones por proveedor

```r
# Graficar el boxplot de vibraciones por proveedor

boxplot(vibracion ~ proveedor, data = datos, col = c("green", "blue",
"red", "purple",
                                                     "orange"),
        main = "Boxplot de vibraciones por proveedor")
```



Media para las vibraciones por proveedor y material

```r
# Mostrar la media de ibraciones por proveedor y material

print("Medias por proveedor y material:")

## [1] "Medias por proveedor y material:"

medias_prov_mat

##            Acero Aluminio Plastico
## Proveedor1 13.15    14.90    14.15
## Proveedor2 16.05    16.05    16.95
## Proveedor3 14.00    14.10    12.35
## Proveedor4 15.75    13.95    14.15
## Proveedor5 13.00    13.60    13.15
```

A continuación se procederá a calcular y graficar los intervalos de confianza para el factor asociado con el proveedor, ya que es único factor que resultó significativo en el ANOVA sin interacción, por lo cual, por el contrario, el factor de material se descarta en esta ocasión ya que aparte de no resultar estadísticamente significativo, también no está implicado en ninguna interacción que sí sea significativa.

IC de vibraciones por proveedor

```
# Calcular desviación estándar de vibraciones por material

desv_proveedor = tapply(vibracion, proveedor, sd)

# Cantidad de datos de proveedores

n_prov1 = length(datos[datos$proveedor == "Proveedor1", 1]) # proveedor1
n_prov2 = length(datos[datos$proveedor == "Proveedor2", 1]) # proveedor2
n_prov3 = length(datos[datos$proveedor == "Proveedor3", 1]) # proveedor3
n_prov4 = length(datos[datos$proveedor == "Proveedor4", 1]) # proveedor4
n_prov5 = length(datos[datos$proveedor == "Proveedor5", 1]) # proveedor5

# Calcular el margen de error E para la media de vibraciones del
proveedor1

E_prov1 = abs((qnorm(0.05 / 2, medias_proveedor["Proveedor1"],
                     desv_proveedor["Proveedor1"]) *
                 desv_proveedor["Proveedor1"])/sqrt(n_prov1))

# Calcular margen de error E para la media de vibraciones del proveedor2

E_prov2 = abs((qnorm(0.05 / 2, medias_proveedor["Proveedor2"],
                     desv_proveedor["Proveedor2"]) *
                 desv_proveedor["Proveedor2"])/sqrt(n_prov2))

# Calcular margen de error E para la media de vibraciones del proveedor3

E_prov3 = abs((qnorm(0.05 / 2, medias_proveedor["Proveedor3"],
                     desv_proveedor["Proveedor3"]) *
                 desv_proveedor["Proveedor3"])/sqrt(n_prov3))

# Calcular margen de error E para la media de vibraciones del proveedor4

E_prov4 = abs((qnorm(0.05 / 2, medias_proveedor["Proveedor4"],
                     desv_proveedor["Proveedor4"]) *
                 desv_proveedor["Proveedor4"])/sqrt(n_prov4))

# Calcular margen de error E para la media de vibraciones del proveedor5

E_prov5 = abs((qnorm(0.05 / 2, medias_proveedor["Proveedor5"],
                     desv_proveedor["Proveedor5"]) *
                 desv_proveedor["Proveedor5"])/sqrt(n_prov5))

# Mostrar intervalo de confianza para proveedor1
```

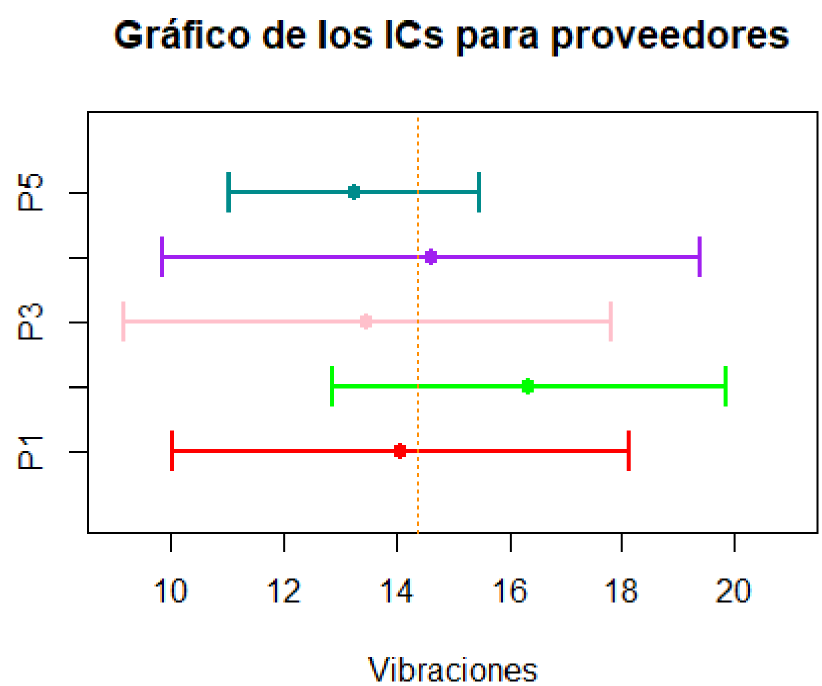
```r
cat("IC de vibraciones para proveedor1: ", medias_proveedor["Proveedor1"]
- E_prov1,
    " - ", medias_proveedor["Proveedor1"] + E_prov1, "\n")

## IC de vibraciones para proveedor1:  10.01076  -  18.12257

# Mostrar intervalo de confianza para proveedor2

cat("IC de vibraciones para proveedor2: ", medias_proveedor["Proveedor2"]
- E_prov2,
    " - ", medias_proveedor["Proveedor2"] + E_prov2, "\n")

## IC de vibraciones para proveedor2:  12.85579  -  19.84421

# Mostrar intervalo de confianza para proveedor3

cat("IC de vibraciones para proveedor3: ", medias_proveedor["Proveedor3"]
- E_prov3,
    " - ", medias_proveedor["Proveedor3"] + E_prov3, "\n")

## IC de vibraciones para proveedor3:  9.142219  -  17.82445

# Mostrar intervalo de confianza para proveedor4

cat("IC de vibraciones para proveedor4: ", medias_proveedor["Proveedor4"]
- E_prov4,
    " - ", medias_proveedor["Proveedor4"] + E_prov4, "\n")

## IC de vibraciones para proveedor4:  9.845213  -  19.38812

# Mostrar intervalo de confianza para proveedor5

cat("IC de vibraciones para proveedor5: ", medias_proveedor["Proveedor5"]
- E_prov5,
    " - ", medias_proveedor["Proveedor5"] + E_prov5)

## IC de vibraciones para proveedor5:  11.01735  -  15.48265
```

Gráfico de los IC para los proveedores

```r
# Asignar límites inferiores y superiores al gráfico de ICs en ambos ejes

plot(0, ylim=c(0, 5 + 1), xlim=c(9, 21), yaxt="n", ylab="", xlab =
"Vibraciones",
     main = "Gráfico de los ICs para proveedores")

# Asignar etiquetas al eje vertical del gráfico

axis(2, at=c(1, 2, 3, 4, 5), labels=c("P1", "P2", "P3", "P4", "P5"))

# Límite inferior y superior del IC para proveedor1 (P1)
```

```r
arrows(medias_proveedor["Proveedor1"] - E_prov1, 1,
       medias_proveedor["Proveedor1"] + E_prov1, 1, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "red")

# Límite inferior y superior del IC para proveedor2 (P2)

arrows(medias_proveedor["Proveedor2"] - E_prov2, 2,
       medias_proveedor["Proveedor2"] + E_prov2, 2, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "green")

# Límite inferior y superior del IC para proveedor3 (P3)

arrows(medias_proveedor["Proveedor3"] - E_prov3, 3,
       medias_proveedor["Proveedor3"] + E_prov3, 3, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "pink")

# Límite inferior y superior del IC para proveedor4 (P4)

arrows(medias_proveedor["Proveedor4"] - E_prov4, 4,
       medias_proveedor["Proveedor4"] + E_prov4, 4, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "purple")

# Límite inferior y superior del IC para proveedor5 (P5)

arrows(medias_proveedor["Proveedor5"] - E_prov5, 5,
       medias_proveedor["Proveedor5"] + E_prov5, 5, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "darkcyan")

# Graficar punto sobre la media de P1

points(medias_proveedor["Proveedor1"], 1, pch = 19, cex = 1.1, col =
"red")

# Graficar punto sobre la media de P2

points(medias_proveedor["Proveedor2"], 2, pch = 19, cex = 1.1, col =
"green")

# Graficar punto sobre la media de P3

points(medias_proveedor["Proveedor3"], 3, pch = 19, cex = 1.1, col =
"pink")

# Graficar punto sobre la media de P4

points(medias_proveedor["Proveedor4"], 4, pch = 19, cex = 1.1, col =
"purple")
```

```r
# Graficar punto sobre la media de P5

points(medias_proveedor["Proveedor5"], 5, pch = 19, cex = 1.1, col =
"darkcyan")

# Trazar una recta vertical sobre el valor de la media general de
vibraciones

abline(v = mean(datos$vibración), lty = 3, col = "darkorange")
```



Desde la perspectiva estadística, se aprecia que la media tanto del proveedor 2 y del proveedor 4 son superiores a la media general de vibraciones, misma que tiene un valor de 14.3533333, mientras que por el contrario, las medias de vibraciones correspondientes a los proveedores 1, 3 y 5, se ubican por debajo de la media general de vibraciones, motivo por el cual eso indica que de las 5 medias e intervalos de confianza graficados, solamente el 40% (2/5) son superiores al promedio general mientras que el 60% restante es inferior al mismo, lo cual, dentro del contexto del problema, nos indica que la mayoría de los proveedores, proveen materiales para fabricación de motores que se encuentran por debajo de la media general de vibraciones, lo cual es benéfico en el sentido de que la fábrica de motores tendrá más opciones de proveedores cuyos materiales sean de buena calidad (que vibren lo menos posible) para que de esa manera, se asegure que los motores producidos en la fábrica igualmente sean de alta calidad y que propicien el adecuado funcionamiento de los automóviles que los utilicen sin ocasionar algún fallo en los automóviles a corto plazo.

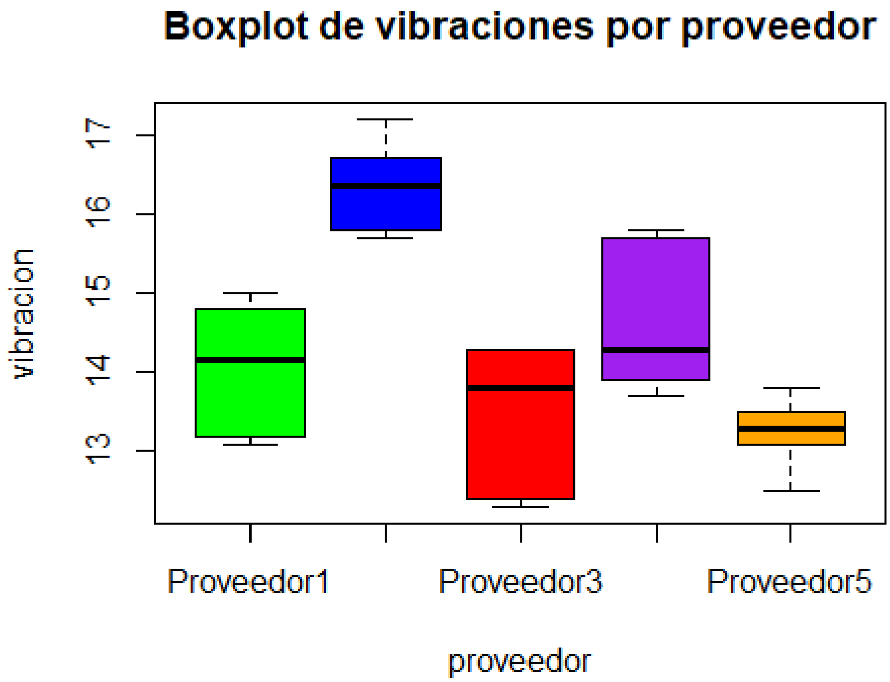
Finalmente, a manera de conclusión de ésta parte del análisis, el efecto principal que resultó significativo al realizar el modelo ANOVA sin interacción fue el proveedor, lo cual en el sentido práctico tiene sentido, ya que la cantidad de vibraciones no se atribuye al tipo de material en si usado para fabricar los motores, sino más bien a los cojinentes que cada proveedor de materiales le suministra a la fábrica, por lo que en función de la calidad de dichos cojinetes, será el desempeño que tendrán los motores fabricados con ellos, por lo que en caso de que los cojinetes provistos por un cierto proveedor sean de alta calidad, los motores producidos a partir de ellos también serán de buena calidad, provocando que presenten una menor cantidad de vibraciones, mientras que si por el contrario, los cojinetes provistos son de una calidad deficiente, entonces los motores fabricados con ellos también serán de calidad baja, conduciendo una mayor cantidad y frecuencia de vibraciones junto con otras posibles fallas durante el tiempo de uso del motor fabricado con ellos.

5. ANOVA para 1 efecto principal significativo (proveedor)

Dado que el efecto que resultó significativo en el ANOVA anterior fue el relacionado a los cojinetes provistos por los proveedores, se procederá a realizar el ANOVA solamente considerando dicho efecto principal, ya sin tomar en cuenta el material y la interacción entre material y proveedor.

Boxplot de vibraciones por proveedor

```
# Graficar el boxplot de vibraciones por proveedor

boxplot(vibracion ~ proveedor, data = datos, col = c("green", "blue",
"red", "purple",
                                                    "orange"),
        main = "Boxplot de vibraciones por proveedor")
```




```
# Media por proveedor (efecto principal significativo)

tapply(vibracion, proveedor, mean)

## Proveedor1 Proveedor2 Proveedor3 Proveedor4 Proveedor5
##   14.06667   16.35000   13.48333   14.61667   13.25000
```

ANOVA con 1 efecto principal (proveedor)

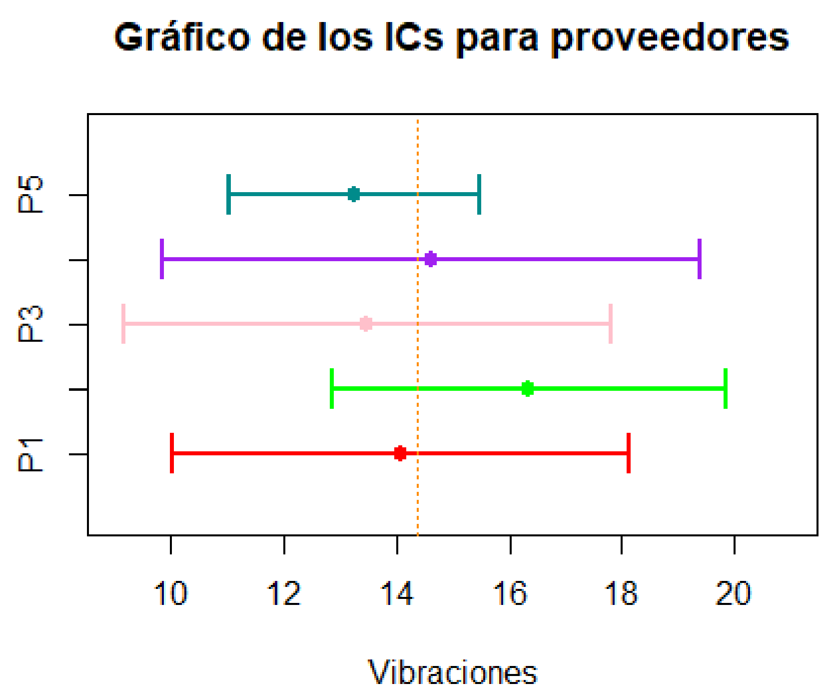
```
# Realizar el ANOVA pero ahora considerando 1 efecto principal
(proveedor)

anova_1_efecto_prov = aov(vibracion ~ proveedor, data = datos)

# Mostrar resumen del modelo ANOVA

summary(anova_1_efecto_prov)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## proveedor    4  36.67   9.169    16.4 1.03e-06 ***
## Residuals   25  13.98   0.559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

IC de vibraciones por proveedor (ya realizado previamente)

```
# Asignar límites inferiores y superiores al gráfico de ICs en ambos ejes
```

```r
plot(0, ylim=c(0, 5 + 1), xlim=c(9, 21), yaxt="n", ylab="", xlab =
"Vibraciones",
     main = "Gráfico de los ICs para proveedores")

# Asignar etiquetas al eje vertical del gráfico

axis(2, at=c(1, 2, 3, 4, 5), labels=c("P1", "P2", "P3", "P4", "P5"))

# Límite inferior y superior del IC para proveedor1 (P1)

arrows(medias_proveedor["Proveedor1"] - E_prov1, 1,
       medias_proveedor["Proveedor1"] + E_prov1, 1, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "red")

# Límite inferior y superior del IC para proveedor2 (P2)

arrows(medias_proveedor["Proveedor2"] - E_prov2, 2,
       medias_proveedor["Proveedor2"] + E_prov2, 2, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "green")

# Límite inferior y superior del IC para proveedor3 (P3)

arrows(medias_proveedor["Proveedor3"] - E_prov3, 3,
       medias_proveedor["Proveedor3"] + E_prov3, 3, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "pink")

# Límite inferior y superior del IC para proveedor4 (P4)

arrows(medias_proveedor["Proveedor4"] - E_prov4, 4,
       medias_proveedor["Proveedor4"] + E_prov4, 4, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "purple")

# Límite inferior y superior del IC para proveedor5 (P5)

arrows(medias_proveedor["Proveedor5"] - E_prov5, 5,
       medias_proveedor["Proveedor5"] + E_prov5, 5, angle = 90,
       code = 3, length = 0.1, lwd = 2, col = "darkcyan")

# Graficar punto sobre la media de P1

points(medias_proveedor["Proveedor1"], 1, pch = 19, cex = 1.1, col =
"red")

# Graficar punto sobre la media de P2

points(medias_proveedor["Proveedor2"], 2, pch = 19, cex = 1.1, col =
"green")
```

```r
# Graficar punto sobre la media de P3

points(medias_proveedor["Proveedor3"], 3, pch = 19, cex = 1.1, col =
"pink")

# Graficar punto sobre la media de P4

points(medias_proveedor["Proveedor4"], 4, pch = 19, cex = 1.1, col =
"purple")

# Graficar punto sobre la media de P5

points(medias_proveedor["Proveedor5"], 5, pch = 19, cex = 1.1, col =
"darkcyan")

# Trazar una recta vertical sobre el valor de la media general de
vibraciones

abline(v = mean(datos$vibración), lty = 3, col = "darkorange")
```



Prueba de comparaciones múltiples de Tukey

```r
# Realizar la prueba de Tukey para comparar las medias de los métodos
(todas contra todas)

tukey_proveedores = TukeyHSD(anova_1_efecto_prov)
```

```r
# Mostrar resultados de la prueba

tukey_proveedores

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = vibracion ~ proveedor, data = datos)
##
## $proveedor
##                              diff        lwr         upr     p adj
## Proveedor2-Proveedor1  2.2833333  1.0153666  3.55130006 0.0001595
## Proveedor3-Proveedor1 -0.5833333 -1.8513001  0.68463339 0.6630108
## Proveedor4-Proveedor1  0.5500000 -0.7179667  1.81796672 0.7089904
## Proveedor5-Proveedor1 -0.8166667 -2.0846334  0.45130006 0.3474956
## Proveedor3-Proveedor2 -2.8666667 -4.1346334 -1.59869994 0.0000055
## Proveedor4-Proveedor2 -1.7333333 -3.0013001 -0.46536661 0.0039774
## Proveedor5-Proveedor2 -3.1000000 -4.3679667 -1.83203328 0.0000015
## Proveedor4-Proveedor3  1.1333333 -0.1346334  2.40130006 0.0959316
## Proveedor5-Proveedor3 -0.2333333 -1.5013001  1.03463339 0.9821261
## Proveedor5-Proveedor4 -1.3666667 -2.6346334 -0.09869994 0.0301318

# Graficar intervalos de confianza resultantes de la prueba

plot(tukey_proveedores)
```

Desde una perspectiva estadística, los resultados anteriores muestran la comparación de múltiples medias entre sí mediante la prueba de Tukey, por lo que ésta misma determina todas las combinaciones diferentes posibles que se pueden hacer considerando los 5 proveedores mencionados en este problema, por lo que después de determinar todas las posibles combinaciones de proveedores, se procede a calcular para cada una de ellas, el intervalo de confianza del 95%, es decir aquel intervalo que posee un 95% de probabilidad de contener a cada una de las medias calculadas, además se calcula el valor p para cada una de las medias, por lo que en base a ésto último, se aprecia que aquellas medias que son estadísticamente significativas, es decir que tienen un valor p por debajo de 0.05 son: Proveedor2-Proveedor1, Proveedor3-Proveedor2, Proveedor4-Proveedor2, Proveedor5-Proveedor2 Proveedor5-Proveedor4, por lo cual en dichos casos, se cuenta con evidencia estadística suficiente para rechazar H_0 correspondiente a la hipótesis de significancia de medias, la cual es la siguiente:

H_0: la media de los proveedores es no es significativa.

H_1: la media de los proveedores sí es significativa.

Por lo tanto, en el caso particular de las medias mencionadas anteriormente, es posible concluir que sí son significativas estadísticamente hablando, lo cual en el contexto del problema implica que la media de vibraciones de los motores elaborados utilizando los cojinentes provistos por los proveedores 2-1, 3-2, 4-2, 5-2 y 5-4 son significativamente diferentes al resto de las medias comparadas mediante la prueba de Tukey, lo cual significa que los cojinentes de dichos pares de proveedores son de una calidad más alta que los cojinentes de los proveedores sobrantes, por lo cual para la fábrica es más conveniente aceptar los cojinetes de los proveedores descritos previamente para garantizar un adecuado proceso de fabricación de los motores y en consecuencia también un óptimo rendimiento de los mismos al momento de emplearse en vehículos.

Por último, a manera de conclusión de ésta parte del análisis, se puede afirmar que el único factor de los 2 factores originalmente involucrados en el análisis que fue realmente significativo en la vibración de los motores son los cojinetes provistos a la fábrica por los proveedores y que se utlizan en la fabricación de los motores, independientemente del material con el que se elaboren éstos mismos, esto debido a que en los 3 modelos ANOVA generados para este problema, el factor de cojinetes del proveedor fue estadísticamente significativo, mientras que el material no lo fue, motivo por el cual, respondiendo a la pregunta de investigación de este problema, existe evidencia de que el proveedor es la causa principal de la vibración de los motores eléctricos, ya que en función de la calidad de los cojinetes que éste mismo le provea a la fábrica, será la calidad que tengan los motores eléctricos resultantes del proceso de fabriación, esto independientemente del tipo de material con el que se fabriquen a su vez los motores.

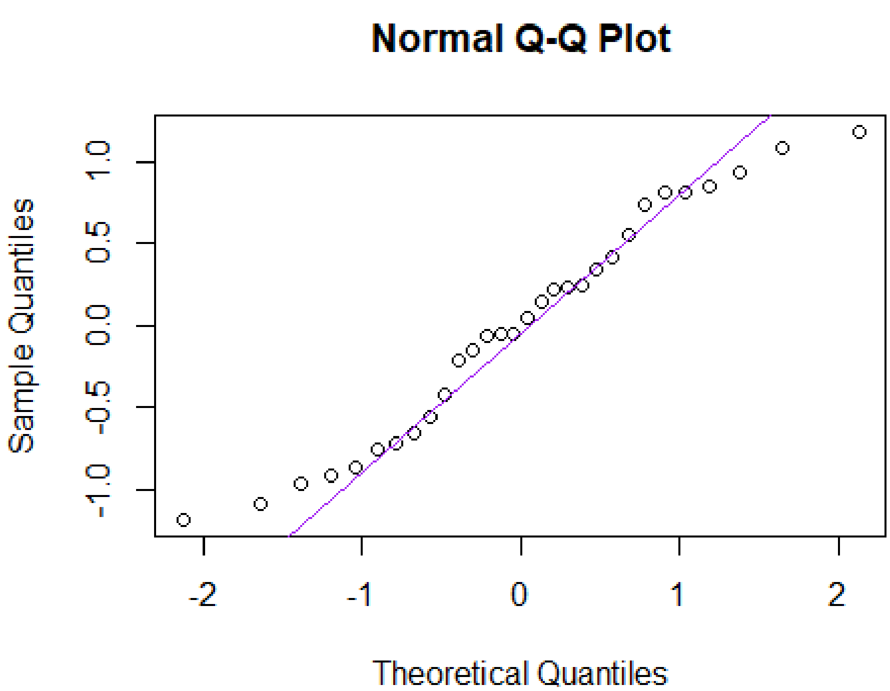
6. Comprobar la validez del modelo

Para comprobar que el modelo ANOVA con el único efecto principal significativo sea estadísticamente adecuado y los resultados derivados del mismo sean confiables, será necesario validar los supuestos a continuación:

- Normalidad de los residuos
- Homocedasticidad de los residuos (los residuos tienen varianza constante)
- Independencia de los residuos (no hay presencia de correlación entre residuos)
- Relación lineal entre variables (coeficiente de determinación R^2)

Normalidad de residuos

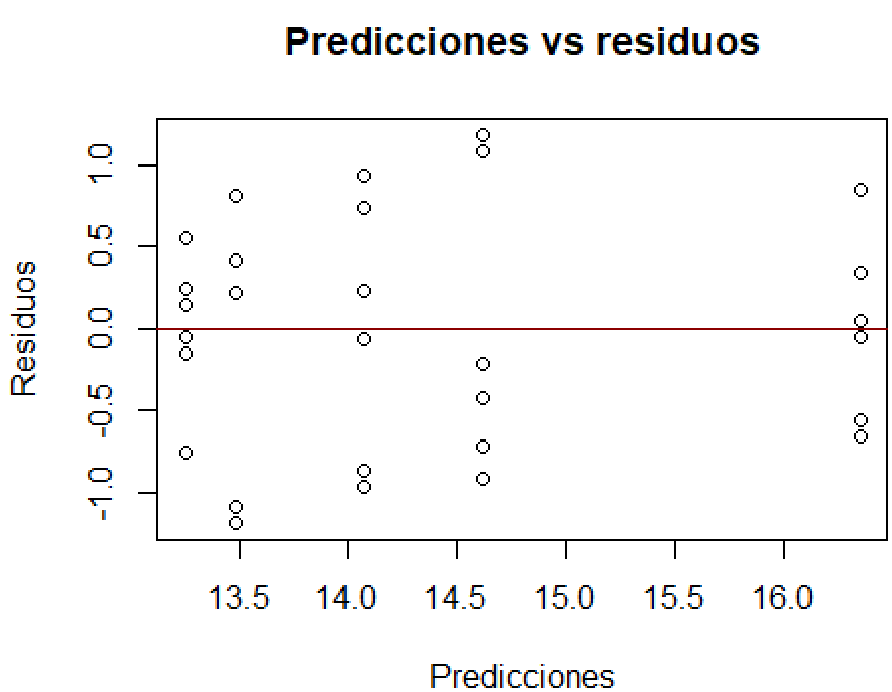
```
# Para verificar que los residuos sigan una distribución normal, se
realizará
# el QQ-plot de los residuos del modelo ANOVA

# Graficar percentiles de normalidad de los residuos

qqnorm(anova_1_efecto_prov$residuals)

# Graficar la recta correspondiente a la distribución normal

qqline(anova_1_efecto_prov$residuals, col = "purple")
```



En el QQ plot correspondiente a los valores de los residuos del modelo ANOVA de 1 efecto principal significativo (proveedor), se puede apreciar que los puntos del gráfico (quantiles de normalidad) de dichos residuos se encuentran en su mayoría sobre y a una distancia muy pequeña alrededor de la recta de normalidad del QQplot, no obstante, también están presentes otros quantiles que se alejan de la recta ideal de normalidad, mismos que se ubican principalmente en los los extremos de la recta de normalidad, motivo por el cual, en base a esto, es posible afirmar que los residuos del modelo ANOVA de 1 efecto principal significativo (proveedor), se alejan parcialmente de la distribución normal, sin embargo los datos que se alejan significativamente de la recta principal son minoría, por lo cual, se concluye que los residuos del ANOVA siguen mayormente una distribución normal.

Homocedasticidad de residuos

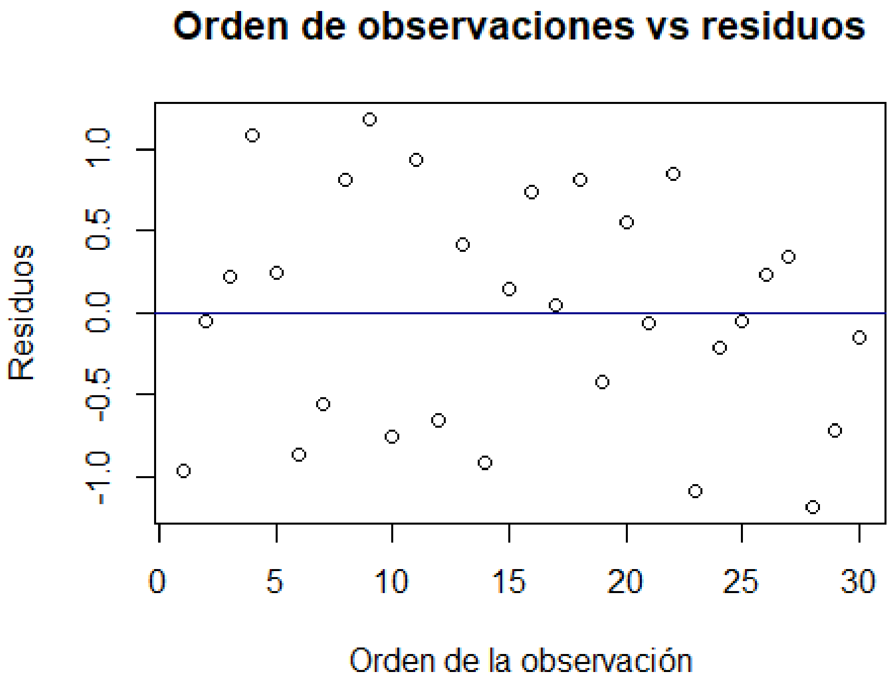
```
# Para verificar si hay presencia de homocedasticidad en los residuos del
ANOVA, se realizará una gráfica de los valores predichos por el modelo y
los residuos del mismo

plot(anova_1_efecto_prov$fitted.values, anova_1_efecto_prov$residuals,
     ylab="Residuos", xlab="Predicciones", main = "Predicciones vs
residuos")

# Graficar una recta horizontal en el valor 0 (media de los residuos del
ANOVA)

abline(h = 0, col = "darkred")
```



En el gráfico de las predicciones derivadas del modelo ANOVA contra los residuos obtenidos del mismo, es posible observar que cada serie de puntos orientada de forma vertical, presenta entre sus datos una dispersión mayormente desigual, dado que cada serie de puntos tienen una separación diferente entre ellos en cada una de las series de puntos graficadas, motivo por el cual, en términos generales, se observa que en el gráfico de las predicciones del ANOVA contra los residuos del mismo, los datos no presentan una dispersión constante a lo largo de todo el gráfico, ya que la separación entre puntos en cada serie es diferente, por lo tanto, tomando en cuenta lo anteriormente mencionado, se concluye que los residuos del modelo ANOVA no presentan varianza constante entre ellos u homocedasticidad.

Independencia de residuos

```
# Para validar si hay presencia de independencia entre los residuos del
modelo ANOVA, se
# realizará una gráfica de las observaciones ordenadas conforme fueron
registradas contra
# los residuos del modelo

# Definir variable tiempo con valores desde 1 hasta la cantidad de datos
del dataset
# de datos (30 registros)

tiempo = 1:nrow(datos)

# Cantidad de datos registrados por proveedor

n = tapply(tiempo, proveedor, length)

# Graficar las observaciones contra sus correspondientes residuos en el
orden en el que
# fueron registradas

plot(tiempo, anova_1_efecto_prov$residuals, xlab="Orden de la
observación",
     ylab="Residuos", main = "Orden de observaciones vs residuos")

# Línea horizontal en 0 (media de los residuos del modelo ANOVA)

abline(h = 0, col = "darkblue")
```



En cuanto al supuesto de independencia de los residuos del modelo ANOVA, en la gráfica previa de las observaciones graficadas en el orden como fueron registradas contra los residuos del modelo ANOVA, se observa que los puntos u observaciones se encuentran completamente dispersas a lo largo del gráfico, sin seguir alguna tendencia o patrón en específico, por lo cual se afirma que dichas observaciones al estar dispersas por todo el gráfico sin evidenciar alguna tendencia o patrón aparente, entonces la dispersión de dichas observaciones se puede atribuir simplemente al azar, motivo por el cual, a partir de esto se concluye que los residuos del modelo ANOVA en cuestión, son independientes entre ellos.

Relación lineal entre las variables (coeficiente de determinación)

```
# Mostrar resumen del modelo ANOVA del factor proveedor

summary(anova_1_efecto_prov)

##             Df Sum Sq Mean Sq F value   Pr(>F)    
## proveedor    4  36.67   9.169    16.4 1.03e-06 ***
## Residuals   25  13.98   0.559                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Para determinar el grado de relación lineal entre las vibraciones de
los motores
# y los cojinetes provistos por un determinado proveedor, se procederá a
dividir el
# valor de la suma de cuadrados del proveedor entre la suma de la suma de
```

```r
cuadrados
# del proveedor más la suma de cuadrados de los residuos (36.67 + 13.98)

# Calcular el coeficiente de determinación entre el proveedor del
material y la
# cantidad de vibraciones de los motores

r2_prov = 36.67 / (36.67 + 13.98)

cat("Coeficiente de determinación: ", r2_prov)

## Coeficiente de determinación:  0.7239882
```

Después de calcular el coeficiente de determinación dividiendo la suma de cuadrados del efecto principal significativo entre la sumatoria de la suma de cuadrados del efecto principal significativo más la suma de cuadrados de los residuos, se obtuvo que el coeficiente de determinación R^2 para el modelo ANOVA en cuestión es igual a 0.7239, lo cual indica que el modelo es capaz de explicar el 72.39% de la variabilidad total de los datos, lo cual es una mayoría de dicha variabilidad, por lo que se puede concluir que los cojinetes provistos por un cierto proveedor tienen una relación mayormente lineal con la cantidad de vibraciones de los motores.

7. Conclusión final en el contexto del problema

Finalmente, a manera de conclusión general, se puede concluir que la cantidad de vibraciones producidad por los motores eléctricos se atribuye al tipo de cojinetes provistos por un proveedor en particular a la fábrica donde éstos son producidos, por lo que en función de la calidad de dichos cojinetes, los motores eléctricos resultantes producirán más o menos vibraciones, esto independientemente del tipo de material con el que hayan sido fabricados los motores, ya sean de acero, aluminio o plástico, además de que también la interacción entre los cojinetes del proveedor y el tipo de material también juega un papel importante al momento de determinar cuántas vibraciones producen los motores eléctricos, esto dado que en el ANOVA con interacción, dicha interacción resultó ser significativa estadísticamente hablando, por lo cual, en resumen, podemos concluir en base a todas las evidencias y análisis previos, que las vibraciones de los motores eléctricos se atribuyen tanto a los cojinetes provistos por el proveedor como a la interacción entre dichos cojinetes y el material seleccionado para fabricar los motores.