

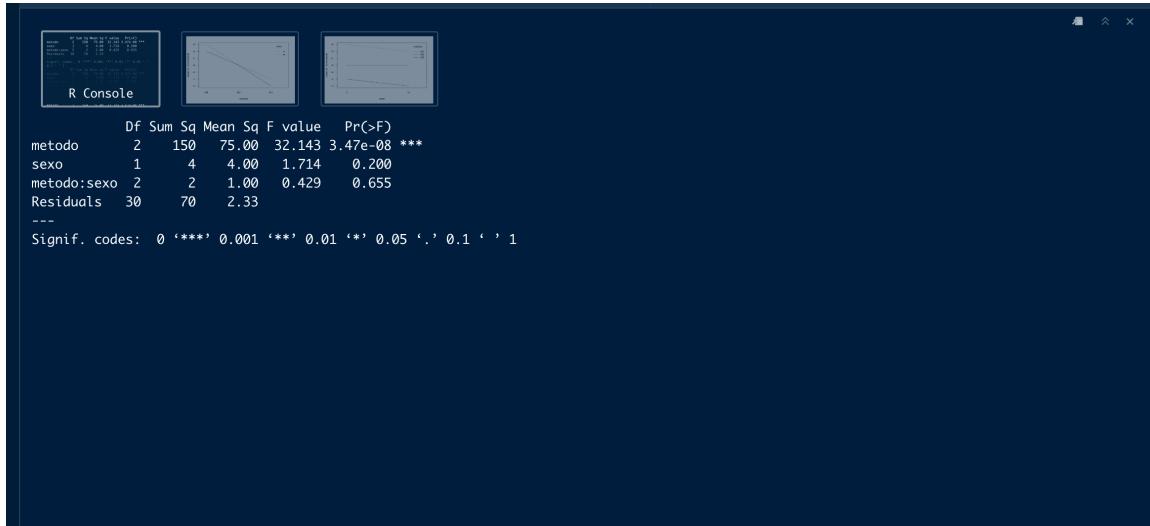
Jorge Javier Sosa Briseño

A01749489

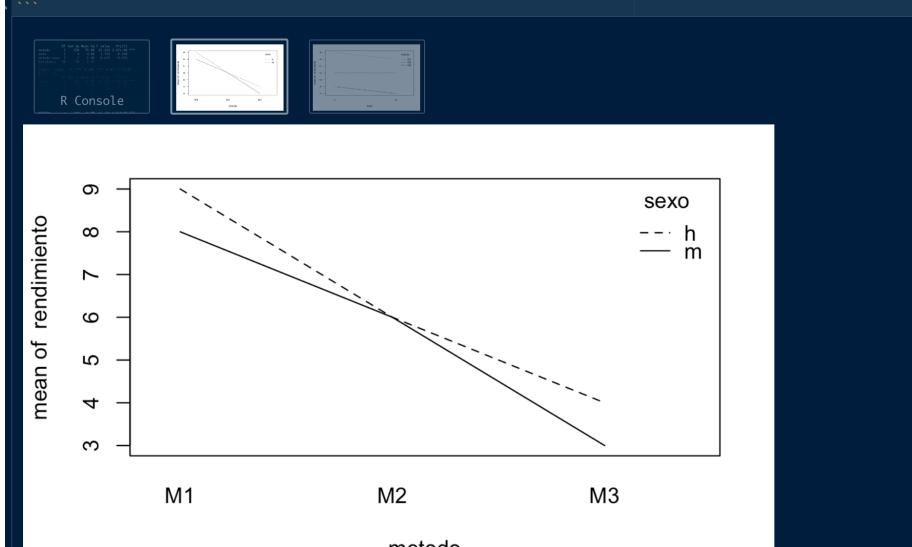
August 30, 2023

ANOVA

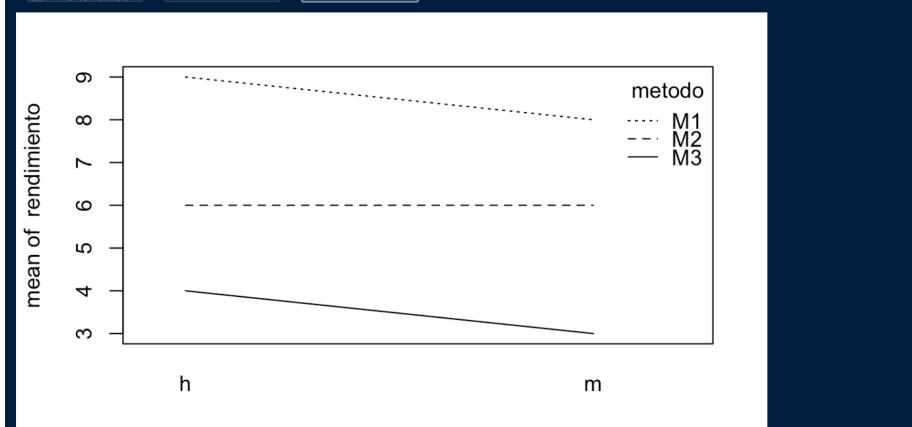
```
1 v ---
2   title: "ANOVA"
3   author: "Jorge Јориј А01749489"
4   date: "8/25/2023"
5   output: html_document
6 ▾ ---
7   En un instituto se han matriculado 36 estudiantes. Se desea explicar el rendimiento de
8   ciencias naturales en función de dos variables: género y metodología de enseñanza. La
9   metodología de enseñanza se analiza en tres niveles: explicación oral y realización del
10  experimento (1er nivel) explicación oral e imágenes (2º nivel) y explicación oral (tercer
11  nivel).
12  En los alumnos matriculados había el mismo número de chicos que de chicas, por lo que
13  formamos dos grupos de 18 sujetos; en cada uno de ellos, el mismo profesor aplicará a
14  grupos aleatorios de 6 estudiantes las 3 metodologías de estudio. A fin de curso los alumnos
15  son sometidos a la misma prueba de rendimiento. Los resultados son los siguientes:
16
17  Chicos Chicas
18  Método 1 Método 2 Método 3 Método 1 Método 2 Método 3
19  10 5 2 9 8 2
20  7 7 6 7 3 6
21  9 6 3 8 5 2
22  9 6 5 8 6 1
23  9 8 5 10 7 4
24  10 4 3 6 7 3
25 ▾ ## Pregunta 1
26  Primero Hip
27 ▾ {r}
28
29  rendimiento=c(10,7,9,9,9,10,5,7,6,6,8,4,2,6,3,
30  5,5,3,9,7,8,8,10,6,8,3,5,6,7,7,2,6,2,1,4,3)
31  metodo=c(rep("M1",6),rep("M2",6),rep("M3",6),rep("M1",6),rep("M2",6),rep("M3",6))
```



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



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



```

#GRAFICA 1
El metodo dos es aprox igual para h y para m, las lineas se tocan pero no se cruzan
##GRAFICA 2
VALORES DEL M1 SON MEJORES QUE EL M2 Y M3
NO HAY UNTERACCION ENTRE EL SEXO Y EL METODO YA QUE NO SE CRUZAN LAS LINEAS, PARA HOMBRES Y PARA MUJERES EL M2 ES APROX IGUAL

Quitamos el efecto de interacion y lo mandamos al error
``{r}
B<-aov(rendimiento~metodo+sexo)
summary(B)
``

Df Sum Sq Mean Sq F value Pr(>F)
metodo 2 150 75.00 33.333 1.5e-08 ***
sexo 1 4 4.00 1.778 0.192
Residuals 32 72 2.25
---

```

Quitamos el efecto de interacion y lo mandamos al error

```

``{r}
B<-aov(rendimiento~metodo+sexo)
summary(B)
``

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

Para observar mejor los efectos de los factores principales, se calcula la media por nivel y se grafica por nivel. También se calcula la media general.
``{r}
tapply(rendimiento,sexo,mean)
tapply(rendimiento,metodo,mean)
M=mean(rendimiento)
M
boxplot(rendimiento ~ sexo)
``
```

Para observar mejor los efectos de los factores principales, se calcula la media por nivel y se grafica por nivel. También se calcula la media general.

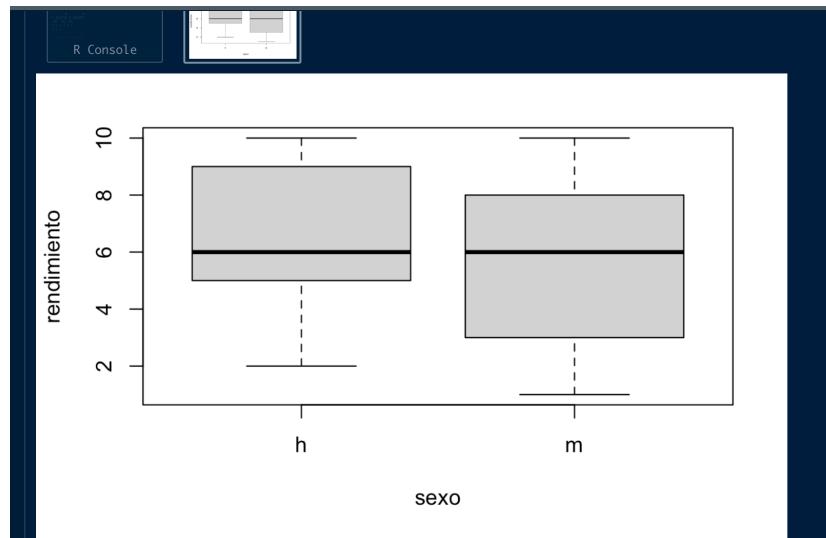
```

``{r}
tapply(rendimiento,sexo,mean)
tapply(rendimiento,metodo,mean)
M=mean(rendimiento)
M
boxplot(rendimiento ~ sexo)
``
```

```

h m
6.333333 5.666667
M1 M2 M3
8.5 6.0 3.5
[1] 6

```



```

```{r}
C<-aov(rendimiento~metodo)
summary(C)
tapply(rendimiento,metodo,mean)
mean(rendimiento)
boxplot(rendimiento ~ metodo)
I = TukeyHSD(aov(rendimiento ~ metodo))
I
plot(I) #Los intervalos de confianza se observan mejor si se grafican
```



Df Sum Sq Mean Sq F value Pr(>F)



	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



M1	M2	M3
8.5	6.0	3.5



[1] 6



Tukey multiple comparisons of means  
95% family-wise confidence level



Fit: aov(formula = rendimiento ~ metodo)

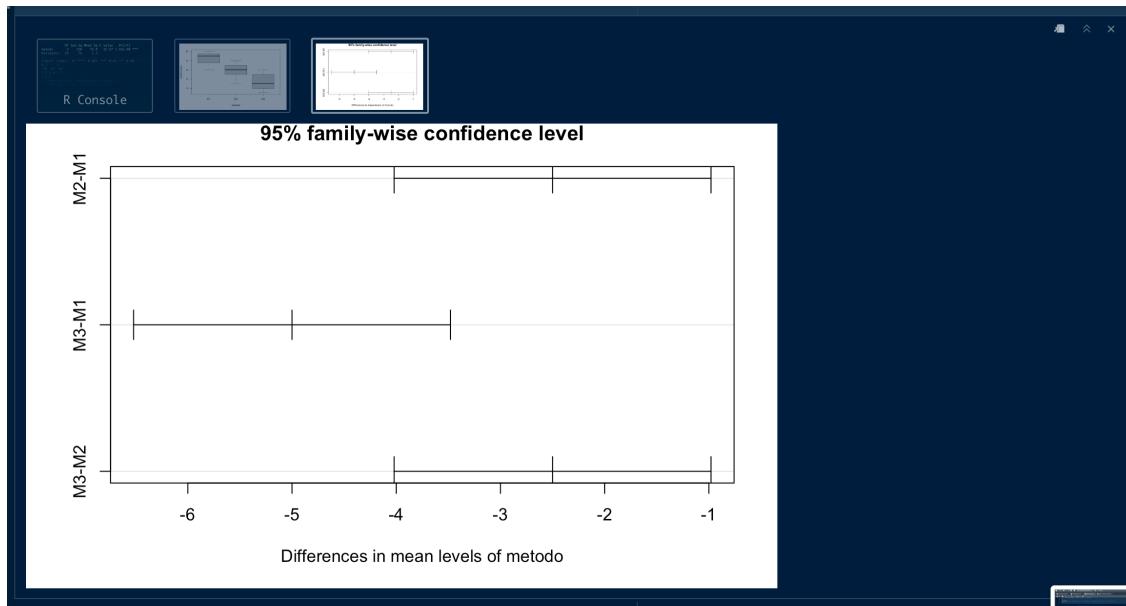
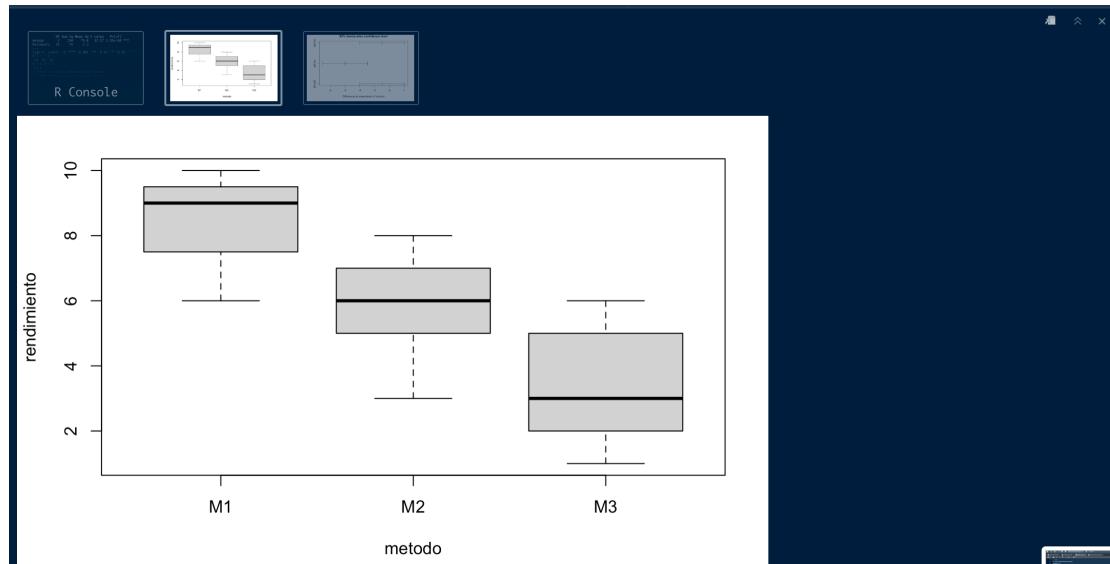


$metodo



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


```



```

+ ## Podemos ver que los valores P despues de pasar por Tukey son menores a 0.05 lo cual se interpreta como que se rechaza la hipótesis nula.

+ ## PREGUNTA 2
Hipótesis Nula ( $H_0$ ):
No hay efectos significativos del factor Material ni del factor Proveedor, ni de su interacción, en la cantidad de vibración de los motores eléctricos.

Hipótesis Alternativa ( $H_1$ ):
Al menos uno de los factores (Material, Proveedor o su interacción) tiene un efecto significativo en la cantidad de vibración de los motores eléctricos.

De manera más específica, para cada factor y su interacción, las hipótesis se formulan de la siguiente manera:

Para el Factor Material:
 $H_0$  (Nula): No hay efecto significativo del factor Material en la cantidad de vibración de los motores eléctricos. (Las medias de vibración son iguales para todas las categorías de Material: Acero, Aluminio y Plástico).
 $H_1$  (Alternativa): Existe un efecto significativo del factor Material en la cantidad de vibración de los motores eléctricos. (Las medias de vibración son diferentes para al menos una pareja de categorías de Material).

Para el Factor Proveedor:
 $H_0$  (Nula): No hay efecto significativo del factor Proveedor en la cantidad de vibración de los motores eléctricos. (Las medias de vibración son iguales para todos los proveedores).
 $H_1$  (Alternativa): Existe un efecto significativo del factor Proveedor en la cantidad de vibración de los motores eléctricos. (Las medias de vibración son diferentes para al menos una pareja de proveedores).

Para la Interacción Material:Proveedor:
 $H_0$  (Nula): No hay efecto significativo de la interacción entre el factor Material y el factor Proveedor en la cantidad de vibración de los motores eléctricos.
 $H_1$  (Alternativa): Existe un efecto significativo de la interacción entre el factor Material y el factor Proveedor en la cantidad de vibración de los motores eléctricos.

```

```

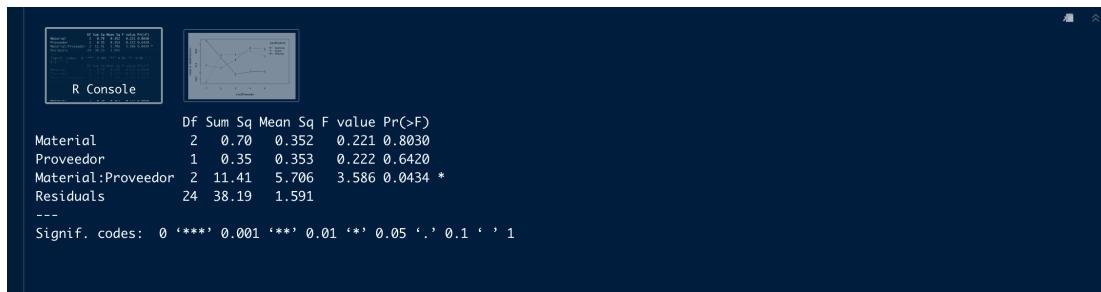
```{r}
Crear un dataframe con los datos
data <- data.frame(
 Material = rep(c("Acero", "Aluminio", "Plástico"), each = 2),
 Proveedor = rep(c(1, 2, 3, 4, 5), times = 3),
 Vibracion = c(
 13.1, 13.2, 15.0, 14.8, 14.0, 14.3,
 16.3, 15.8, 15.7, 16.4, 17.2, 16.7,
 13.7, 14.3, 13.9, 14.3, 12.4, 12.3,
 15.7, 15.8, 13.7, 14.2, 14.4, 13.9,
 13.5, 12.5, 13.4, 13.8, 13.2, 13.1
)
)

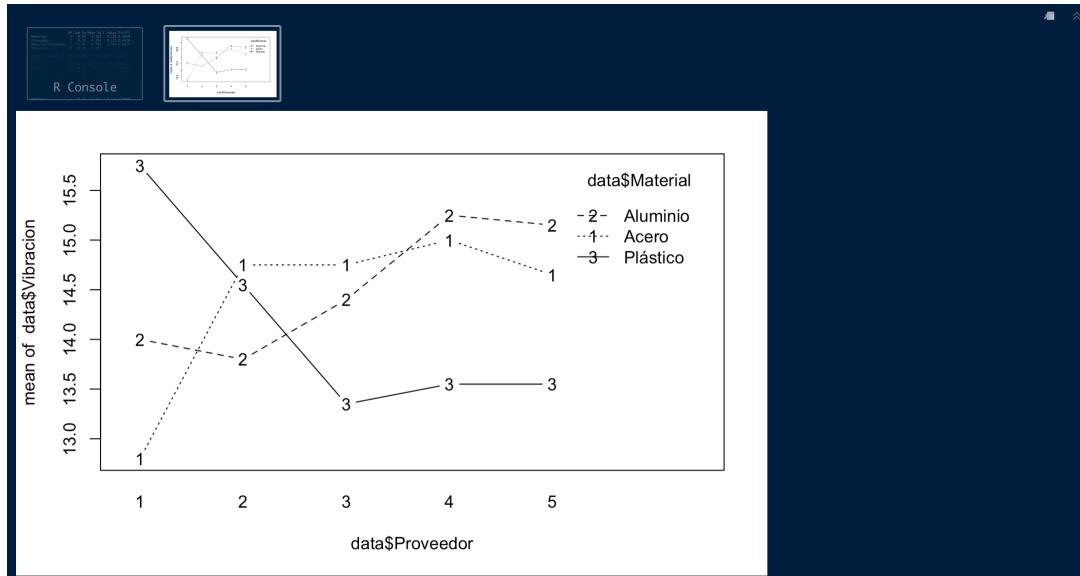
Realizar el análisis de varianza (ANOVA)
modelo_anova <- aov(Vibracion ~ Material + Proveedor + Material:Proveedor, data = data)

Resumen del análisis
summary(modelo_anova)

Gráfico de interacción Material vs. Proveedor
interaction.plot(data$Proveedor, data$Material, data$Vibracion, type = "b", legend = TRUE)
```

```





Material:

```
Df (Degrees of Freedom): 2
Sum Sq (Sum of Squares): 0.70
Mean Sq (Mean of Squares): 0.352
F value: 0.221
Pr(>F): 0.8030
```

Interpretación: El efecto del factor "Material" en la vibración no es estadísticamente significativo. El valor p ($\text{Pr}(>F)$) de 0.8030 es mayor que cualquier nivel de significancia común (como 0.05), lo que indica que no hay suficiente evidencia para rechazar la hipótesis nula de que el material no tiene un efecto significativo en la vibración.

Proveedor:

```
Df: 1
Sum Sq: 0.35
Mean Sq: 0.353
F value: 0.222
Pr(>F): 0.6420
```

Interpretación: El efecto del factor "Proveedor" tampoco es estadísticamente significativo. El valor p de 0.6420 es alto y no permite rechazar la hipótesis nula de que el proveedor no tiene un efecto significativo en la vibración.

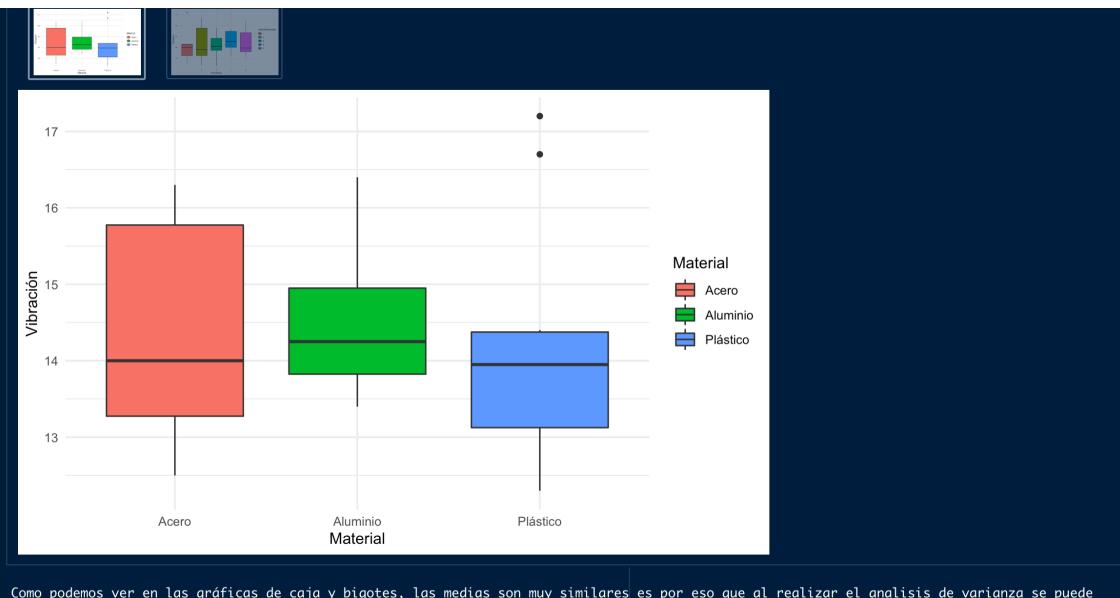
Material:Proveedor:

```
Df: 2
Sum Sq: 11.41
Mean Sq: 5.706
F value: 3.586
Pr(>F): 0.0434 *
```

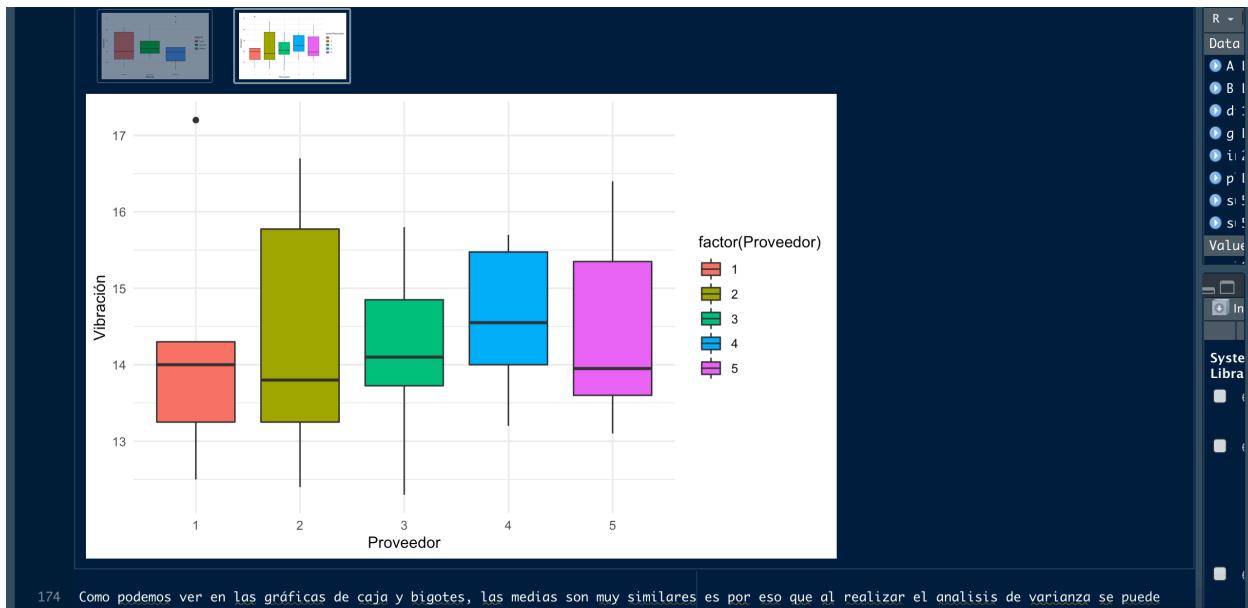
Interpretación: Sin embargo, hay evidencia para sugerir que la interacción entre "Material" y "Proveedor" tiene un efecto significativo en la vibración. El valor p de 0.0434 es menor que 0.05, lo que indica que existe una diferencia significativa en la vibración debido a la combinación de ambos factores.

```
151 Residuals:
152
153 Df: 24
154 Sum Sq: 38.19
155 Mean Sq: 1.591
156 Interpretación: Esta fila representa el error residual, que es la variación no explicada por los factores del modelo.
157
158 La columna "Signif. codes" muestra simbolos que indican el nivel de significancia asociado a los valores p. En este caso, vemos un asterisco (*)
159 a la fila de "Material:Proveedor", lo que indica que la interacción entre "Material" y "Proveedor" es estadísticamente significativa
160 a un nivel de 0.05.
161 En resumen, según los resultados de este ANOVA, la interacción entre los factores "Material" y "Proveedor" parece tener un efecto significativo
162 en la vibración de los motores eléctricos, mientras que los factores individuales "Material" y "Proveedor" no son significativos por sí mismos.
163 # Crear un boxplot por Material
164 ggplot(data, aes(x = Material, y = Vibracion, fill = Material)) +
165   geom_boxplot() +
166   labs(x = "Material", y = "Vibración") +
167   theme_minimal()
168 # Crear un boxplot por Proveedor
169 ggplot(data, aes(x = factor(Proveedor), y = Vibracion, fill = factor(Proveedor))) +
170   geom_boxplot() +
171   labs(x = "Proveedor", y = "Vibración") +
172   theme_minimal()
```





Como podemos ver en las gráficas de caja y bigotes, las medias son muy similares es por eso que al realizar el análisis de varianza se puede analizar si la hipótesis nula se rechaza o no.



174 Como podemos ver en las gráficas de caja y bigotes, las medias son muy similares es por eso que al realizar el análisis de varianza se puede analizar si la hipótesis nula se rechaza o no.

```

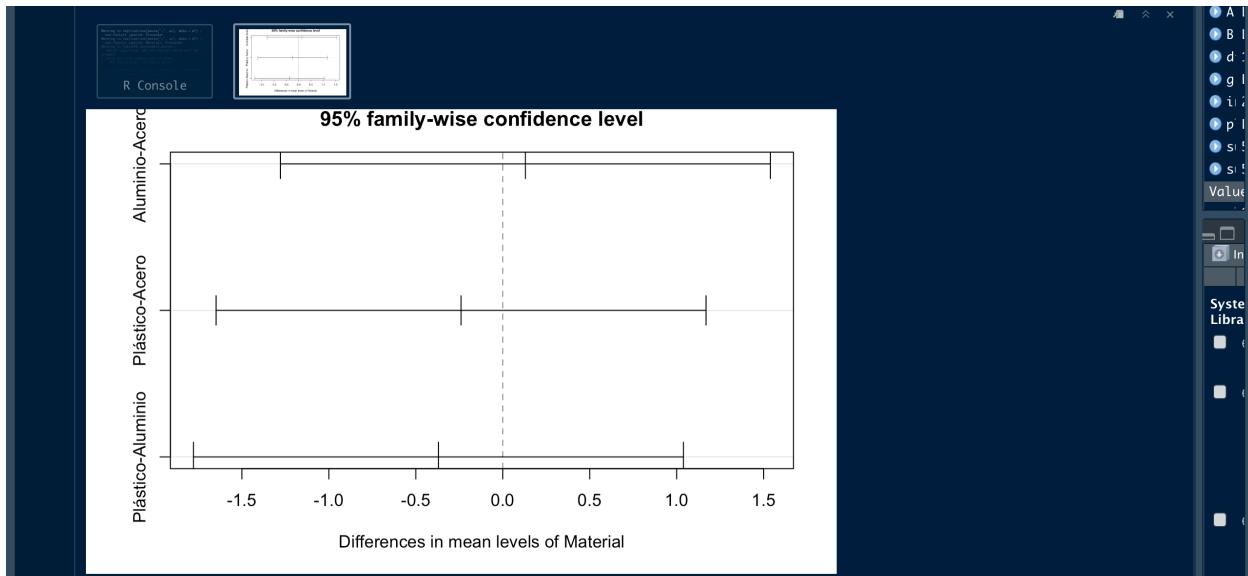
175
176 ````{r}
177 # Realizar la prueba de comparaciones múltiples de Tukey
178 I = TukeyHSD(modelo_anova)
179 I
180 plot(I)
181 ````

Warning in replications(paste("~", xx), data = mf) :
non-factors ignored: Proveedor
Warning in replications(paste("~", xx), data = mf) :
non-factors ignored: Material, Proveedor
Warning in TukeyHSD.aov(modelo_anova) :
'which' specified some non-factors which will be dropped
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = Vibracion ~ Material + Proveedor + Material:Proveedor, data = data)

$Material
            diff      lwr      upr
Aluminio-Acero  0.13 -1.278719 1.538719
Plástico-Acero -0.24 -1.648719 1.168719
Plástico-Aluminio 0.37 -1.778719 1.038719
                p adj
Aluminio-Acero  0.9711816
Plástico-Acero  0.9054306
Plástico-Aluminio 0.7908426

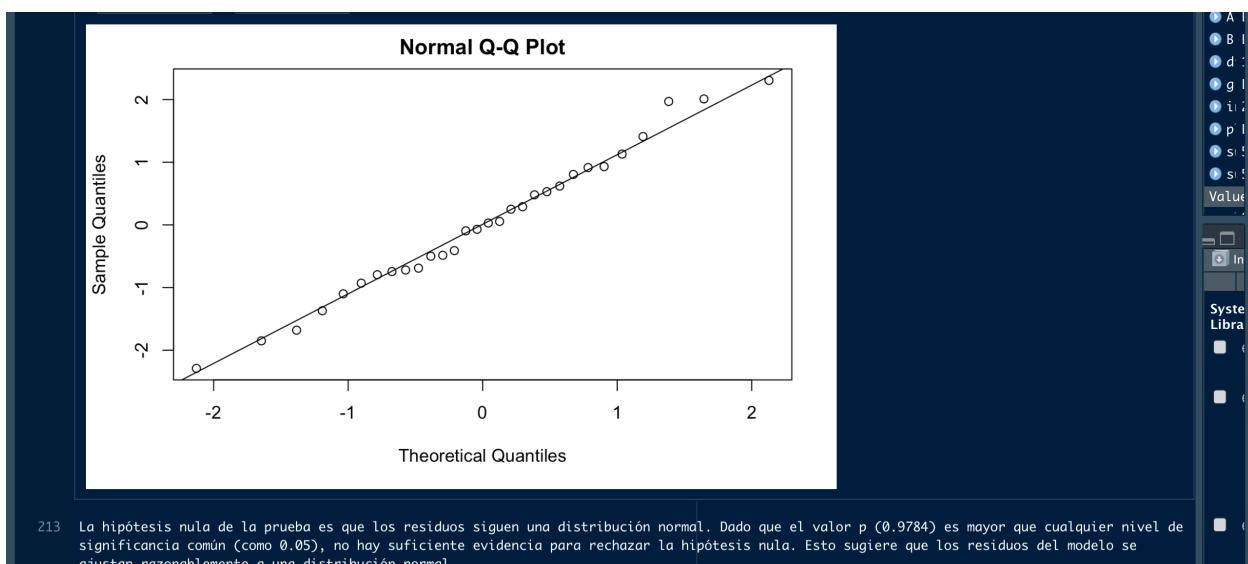
```



```

183 Aluminio vs. Acero:
184
185 La diferencia estimada entre el nivel "Aluminio" y el nivel "Acero" es de 0.13 unidades en la variable de vibración.
186 El intervalo de confianza del 95% para esta diferencia va desde -1.278719 hasta 1.538719.
187 El valor p ajustado para esta comparación es 0.9711816.
188 Interpretación: Dado que el valor p (0.9711816) es mayor que el nivel de significancia común de 0.05, no hay evidencia suficiente para concluir que existe una diferencia significativa en la vibración entre los niveles "Aluminio" y "Acero".
189 Plástico vs. Acero:
190
191 La diferencia estimada entre el nivel "Plástico" y el nivel "Acero" es de -0.24 unidades en la variable de vibración.
192 El intervalo de confianza del 95% para esta diferencia va desde -1.648719 hasta 1.168719.
193 El valor p ajustado para esta comparación es 0.9054306.
194 Interpretación: Dado que el valor p (0.9054306) es mayor que 0.05, no hay evidencia suficiente para afirmar que existe una diferencia significativa en la vibración entre los niveles "Plástico" y "Acero".
195 Plástico vs. Aluminio:
196
197 La diferencia estimada entre el nivel "Plástico" y el nivel "Aluminio" es de -0.37 unidades en la variable de vibración.
198 El intervalo de confianza del 95% para esta diferencia va desde -1.778719 hasta 1.038719.
199 El valor p ajustado para esta comparación es 0.7908426.
200 Interpretación: Dado que el valor p (0.7908426) es mayor que 0.05, no hay evidencia suficiente para concluir que existe una diferencia significativa en la vibración entre los niveles "Plástico" y "Aluminio".
201 En resumen, según los resultados de la prueba de Tukey, no se encontraron diferencias significativas en la vibración entre los diferentes niveles del factor "Material". Los valores p ajustados para todas las comparaciones son mayores que 0.05, lo que indica que las diferencias observadas podrían deberse al azar y no son estadísticamente significativas.
202

```



213 La hipótesis nula de la prueba es que los residuos siguen una distribución normal. Dado que el valor p (0.9784) es mayor que cualquier nivel de significancia común (como 0.05), no hay suficiente evidencia para rechazar la hipótesis nula. Esto sugiere que los residuos del modelo se ajustan razonablemente a una distribución normal.

