Actividad Integradora 2

Adrian Pineda Sanchez

2024-09-06

Grupo 1: Distancia entre los ejes (wheelbase), tipo de gasolina que usa y caballos de fuerza

1. Exploración de la base de datos

```
data = read.csv("precios_autos.csv")

# Seleccionar columnas usando la notación base de R
selected_data <- data[, c("fueltype", "horsepower", "wheelbase")]

# Mostrar los primeros registros del nuevo conjunto de datos seleccionado
head(selected_data)
```

```
##   fueltype horsepower wheelbase
## 1      gas        111      88.6
## 2      gas        111      88.6
## 3      gas        154      94.5
## 4      gas        102      99.8
## 5      gas        115      99.4
## 6      gas        110      99.8
```

Exploración de la base de datos

Calcula medidas estadísticas apropiadas para las variables:

cuantitativas (media, desviación estándar, cuantiles, etc)

```
# Media de Horsepower
cat("La media de Horsepower es:", mean(data$horsepower, na.rm = TRUE), "\n")
```

```
## La media de Horsepower es: 104.1171
```

```
# Media de Wheelbas
cat("La media de Wheelbase es:", mean(data$wheelbase, na.rm = TRUE), "\n")
```

```
## La media de Wheelbase es: 98.75659
```

```
# Desviación estándar de Horsepower
cat("La desviación estándar de Horsepower es:", sd(data$horsepower, na.rm =
TRUE), "\n")
```

```
## La desviación estándar de Horsepower es: 39.54417

# Desviación estándar de Wheelbase
cat("La desviación estándar de Wheelbase es:", sd(data$wheelbase, na.rm =
TRUE), "\n")

## La desviación estándar de Wheelbase es: 6.021776

# Cuantiles de Horsepower (cuartiles)
horsepower_quartiles <- quantile(data$horsepower, probs = c(0.25, 0.5, 0.75),
na.rm = TRUE)
cat("Los cuartiles de Horsepower son:\n")

## Los cuartiles de Horsepower son:

cat("Cuartil 25%:", horsepower_quartiles[1], "\n")

## Cuartil 25%: 70

cat("Mediana (50%):", horsepower_quartiles[2], "\n")

## Mediana (50%): 95

cat("Cuartil 75%:", horsepower_quartiles[3], "\n")

## Cuartil 75%: 116

# Cuantiles de Wheelbase (cuartiles)
wheelbase_quartiles <- quantile(data$wheelbase, probs = c(0.25, 0.5, 0.75),
na.rm = TRUE)
cat("Los cuartiles de Wheelbase son:\n")

## Los cuartiles de Wheelbase son:

cat("Cuartil 25%:", wheelbase_quartiles[1], "\n")

## Cuartil 25%: 94.5

cat("Mediana (50%):", wheelbase_quartiles[2], "\n")

## Mediana (50%): 97

cat("Cuartil 75%:", wheelbase_quartiles[3], "\n")

## Cuartil 75%: 102.4
```

cualitativas: cuantiles, frecuencias (puedes usar el comando table o prop.table)

```
# Frecuencia de Fuel Type
cat("La frecuencia de los tipos de combustible (Fuel Type):\n")

## La frecuencia de los tipos de combustible (Fuel Type):

print(table(data$fueltype))
```

```
##
## diesel    gas
##     20    185
```

```
# Proporciones de Fuel Type
cat("Las proporciones de los tipos de combustible (Fuel Type):\n")
```

```
## Las proporciones de los tipos de combustible (Fuel Type):
```

```
print(prop.table(table(data$fueltype)))
```

```
##
##     diesel        gas
## 0.09756098 0.90243902
```

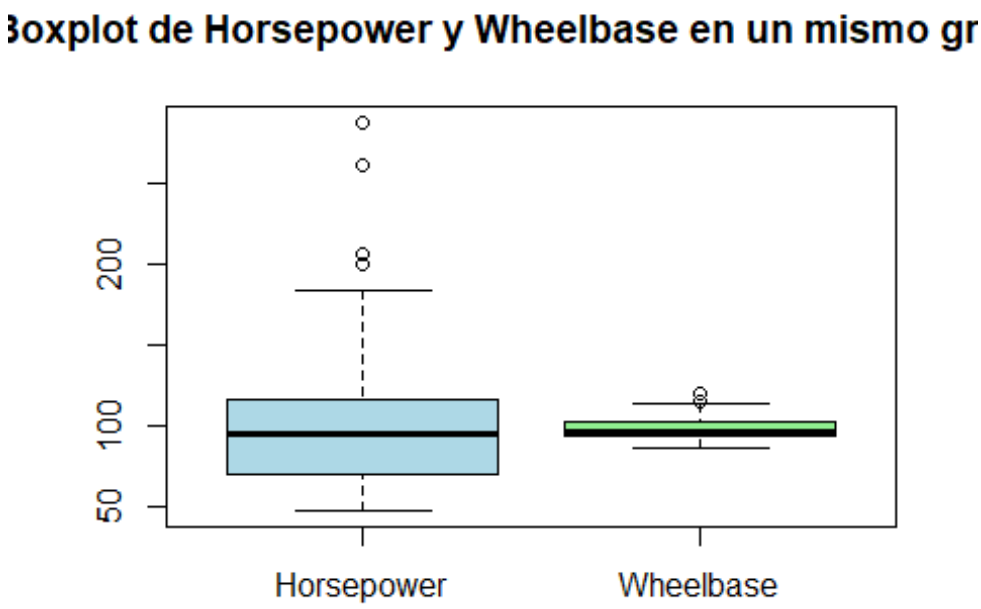
1.2. Analiza la correlación entre las variables (analiza posible colinealidad entre las variables)

```
# Correlación entre Horsepower y Wheelbase
correlation <- cor(data$horsepower, data$wheelbase, use = "complete.obs")
cat("La correlación entre Horsepower y Wheelbase es:", correlation, "\n")
```

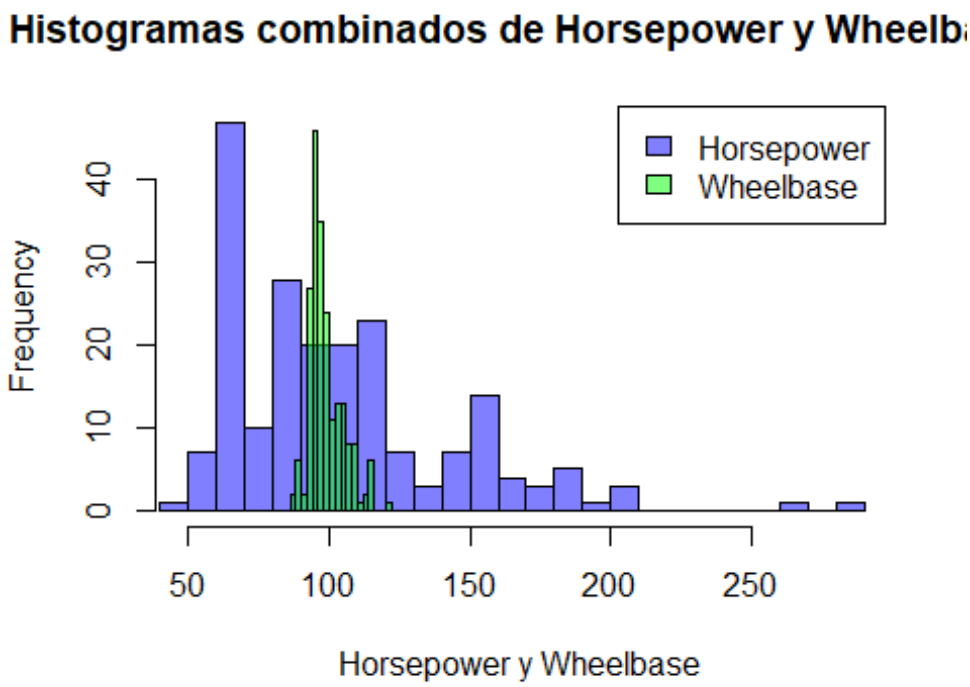
```
## La correlación entre Horsepower y Wheelbase es: 0.3532945
```

1.3. Explora los datos usando herramientas de visualización (si lo consideras necesario):

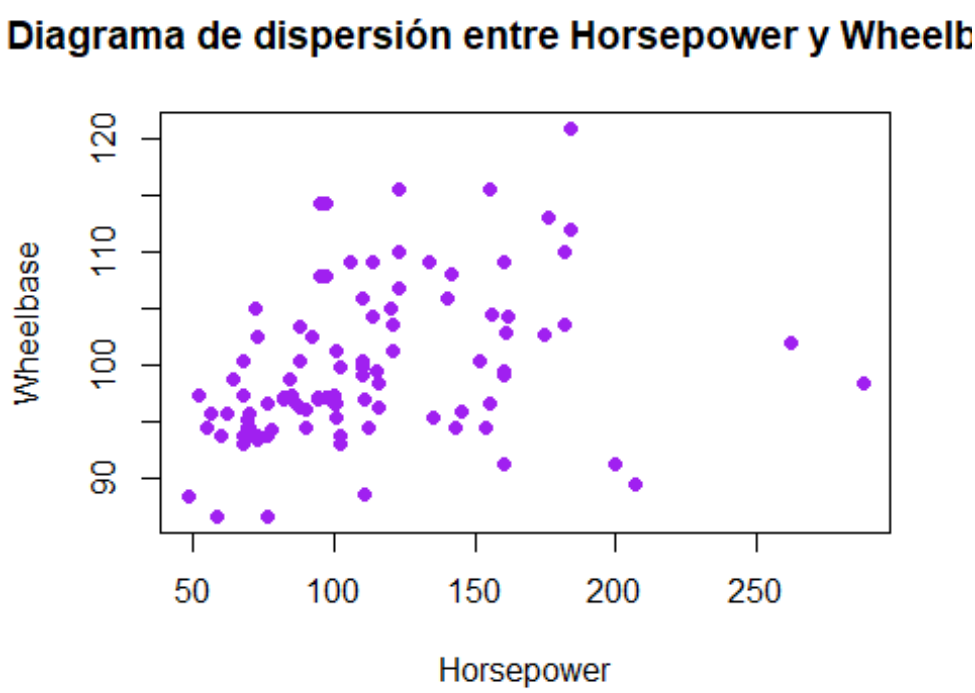
```
boxplot(data$horsepower, data$wheelbase,
        names = c("Horsepower", "Wheelbase"),
        main = "Boxplot de Horsepower y Wheelbase en un mismo gráfico",
        col = c("lightblue", "lightgreen"))
```



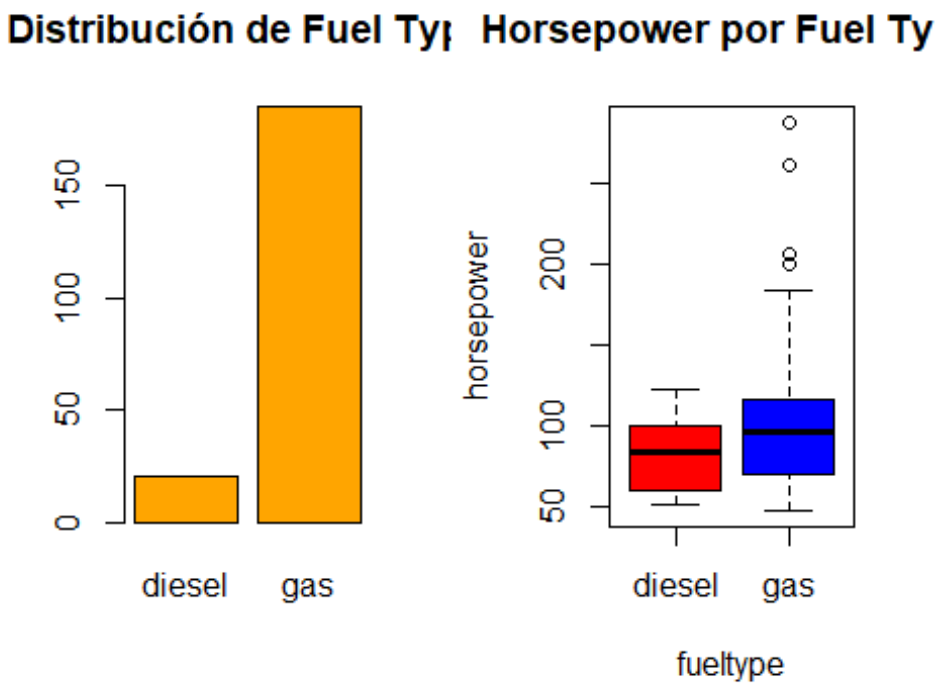

```r
# Generar un único histograma para ambas variables
hist(data$horsepower,
     main = "Histogramas combinados de Horsepower y Wheelbase",
     xlab = "Horsepower y Wheelbase",
     col = rgb(0, 0, 1, 0.5), # Color azul semitransparente para horsepower
     xlim = range(c(data$horsepower, data$wheelbase)),
     breaks = 20)

hist(data$wheelbase,
     col = rgb(0, 1, 0, 0.5), # Color verde semitransparente para wheelbase
     add = TRUE, # Superponer el histograma sobre el anterior
     breaks = 20)

legend("topright", legend = c("Horsepower", "Wheelbase"),
       fill = c(rgb(0, 0, 1, 0.5), rgb(0, 1, 0, 0.5)))
```




```r
# Diagrama de Dispersión entre Horsepower y Wheelbase
plot(data$horsepower, data$wheelbase,
     main = "Diagrama de dispersión entre Horsepower y Wheelbase",
     xlab = "Horsepower", ylab = "Wheelbase",
     pch = 19, col = "purple")
```




```r
par(mfrow = c(1, 2))  # Divide el área gráfica en 1 fila y 2 columnas

# Diagrama de barras para Fuel Type
barplot(table(data$fueltype), main = "Distribución de Fuel Type", col =
"orange")

# Boxplot de Horsepower por Fuel Type
boxplot(horsepower ~ fueltype, data = data, main = "Horsepower por Fuel
Type", col = c("red", "blue"))
```




```r
# Resetear la configuración del layout
par(mfrow = c(1, 1))
```

2. Modelación y verificación del modelo

```r
# Seleccionar las variables relevantes para el análisis
selected_data <- data.frame(data$fueltype, data$horsepower, data$wheelbase,
data$price)
# Renombrar las columnas correctamente
names(selected_data) <- c("fueltype", "horsepower", "wheelbase", "price")

# Calcular la matriz de correlación asegurando que las columnas existen
cor_matrix <- cor(selected_data[, c("horsepower", "wheelbase", "price")], use
= "complete.obs")
print(cor_matrix)
```

```
##             horsepower wheelbase     price
## horsepower   1.0000000 0.3532945 0.8081388
## wheelbase    0.3532945 1.0000000 0.5778156
## price        0.8081388 0.5778156 1.0000000
```

2.1 Encuentra la ecuación de regresión de mejor ajuste. Propón al menos 2 modelos de ajuste para encontrar la mejor forma de ajustar la variable precio.

2.2 Para cada uno de los modelos propuestos:

2.2.1 Realiza la regresión entre las variables involucradas

2.2.2 Analiza la significancia del modelo:

Valida la significancia del modelo con un alfa de 0.04 (incluye las hipótesis que pruebas y el valor frontera) Valida la significancia de β̂i con un alfa de 0.04 (incluye las hipótesis que pruebas y el valor frontera de cada una de ellas) Indica cuál es el porcentaje de variación explicada por el modelo. Dibuja el diagrama de dispersión de los datos por pares y la recta de mejor ajuste. Interpreta en el contexto del problema cada uno de los análisis que hiciste.

Hipotesis

H_0: B_0 y B_1 = 0 H_1: B_0 y B_1 diferente de 0

Modelos

```
# Ajustar el modelo de regresión simple
modelo1 <- lm(price ~ horsepower, data = selected_data)
summary(modelo1)

##
## Call:
## lm(formula = price ~ horsepower, data = selected_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11897.5  -2350.4   -711.1   1644.6  19081.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3721.761    929.849  -4.003 8.78e-05 ***
## horsepower    163.263      8.351  19.549  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4717 on 203 degrees of freedom
## Multiple R-squared:  0.6531, Adjusted R-squared:  0.6514
## F-statistic: 382.2 on 1 and 203 DF,  p-value: < 2.2e-16
```

Paso 1: Validación de la significancia del modelo

Valor p del modelo: < 2.2e-16 Conclusión: Dado que el valor p del modelo es mucho menor que 0.04, rechazamos la hipótesis nula y concluimos que el modelo es altamente significativo. Esto indica que horsepower tiene un efecto significativo sobre el price de los vehículos.

Paso 2: Validación de la significancia de los coeficientes

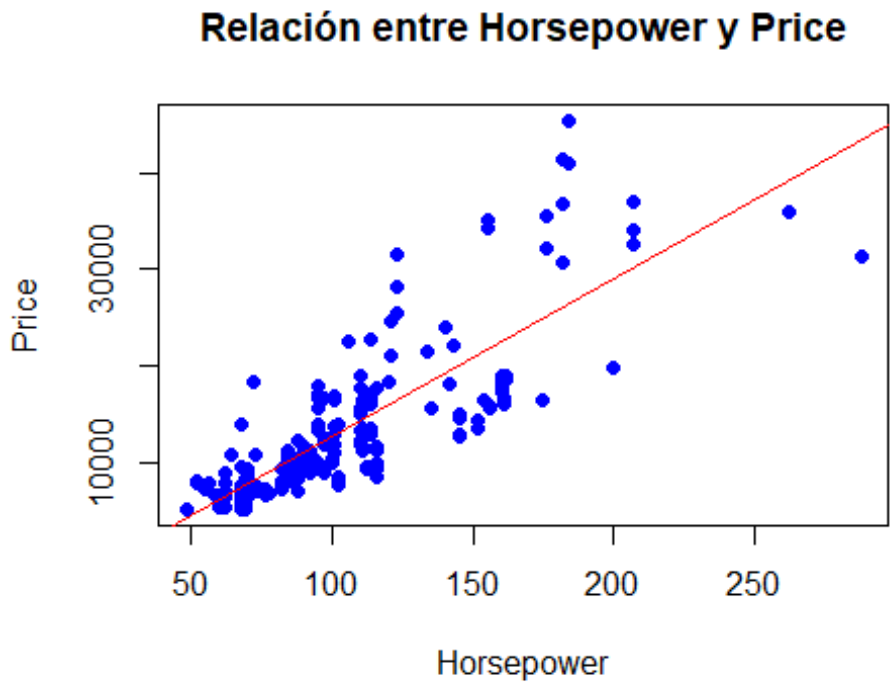
Coeficiente B_0(Intercepto): Valor p: < 8.78e-05 Conclusión: El valor p para B_0es mucho menor que 0.04, por lo que rechazamos la hipótesis nula. Esto significa que el intercepto es significativamente diferente de cero.

Coeficiente B_1(Horsepower): Valor p: < 2e-16 Conclusión: El valor p para B_1 es mucho menor que 0.04, por lo que también rechazamos la hipótesis nula. Esto confirma que existe una relación significativa entre horsepower y price. Por cada unidad adicional de horsepower, el precio promedio aumenta en aproximadamente 163.263 unidades monetarias.

Paso 3: Porcentaje de variación explicada por el modelo

R^2: 0.6531 Interpretación: El modelo explica aproximadamente el 65.31% de la variabilidad en el precio de los vehículos a partir del horsepower. Esto sugiere que horsepower es un buen predictor del price, aunque hay una parte significativa de la variación que no se explica por esta variable.

```
# Graficar la relación entre horsepower y price con la línea de mejor ajuste
plot(selected_data$horsepower, selected_data$price,
     main = "Relación entre Horsepower y Price",
     xlab = "Horsepower", ylab = "Price", pch = 19, col = "blue")
abline(modelo1, col = "red")
```



H_0: B_0, B_1, B_2 = 0 H_1: B_0, B_1, B_2 diferente de 0

```r
# Ajustar el modelo de regresión múltiple sin interaccion
modelo2 <- lm(price ~ horsepower + wheelbase, data = selected_data)
summary(modelo2)
```

```
##
## Call:
## lm(formula = price ~ horsepower + wheelbase, data = selected_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8403.9 -2303.7  -227.6  1608.4 15640.5
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -44998.311   4707.546  -9.559  < 2e-16 ***
## horsepower     139.425      7.586  18.379  < 2e-16 ***
## wheelbase      443.095     49.818   8.894 3.33e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4008 on 202 degrees of freedom
## Multiple R-squared:  0.7507, Adjusted R-squared:  0.7482
## F-statistic: 304.2 on 2 and 202 DF,  p-value: < 2.2e-16
```

Valor p del modelo: < 2.2e-16

Conclusión: Dado que el valor p del modelo es mucho menor que 0.04, rechazamos la hipótesis nula y concluimos que el modelo es altamente significativo. Esto indica que tanto horsepower como wheelbase tienen un efecto significativo sobre el price de los vehículos.

Paso 2: Validación de la significancia de los coeficientes

Coeficiente B_0 (Intercepto): Valor p: < 2e-16 Conclusión: El valor p para B_0 es mucho menor que 0.04, por lo que rechazamos la hipótesis nula. Esto significa que el intercepto es significativamente diferente de cero, lo que indica que hay un punto de inicio importante en el modelo.

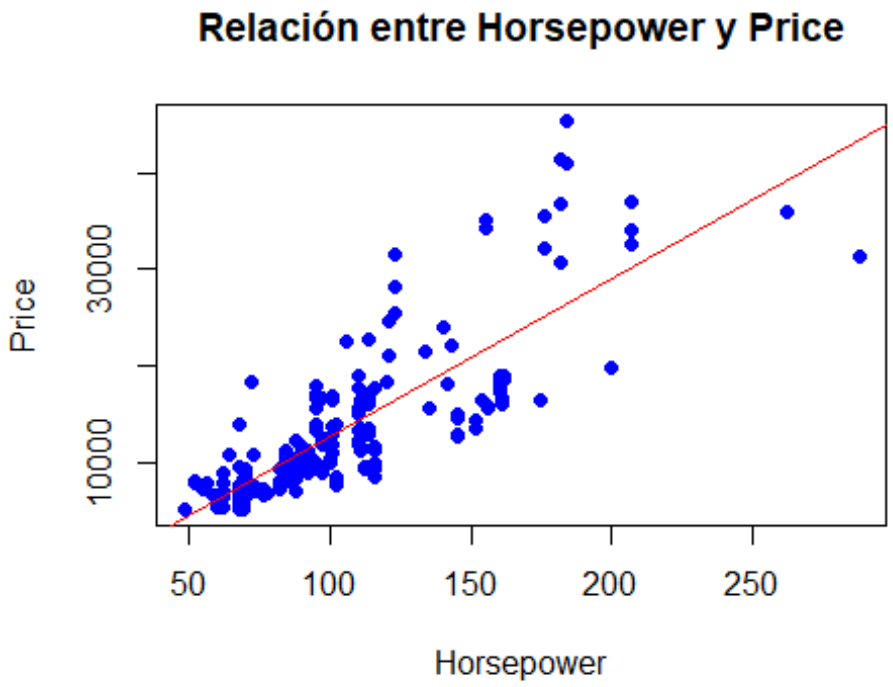
Coeficiente B_1 (Horsepower): Valor p: < 2e-16 Conclusión: El valor p para B_1 es mucho menor que 0.04, por lo que también rechazamos la hipótesis nula. Esto confirma que existe una relación significativa entre horsepower y price. Por cada unidad adicional de horsepower, el precio promedio aumenta en aproximadamente 139.425 unidades monetarias.

Coeficiente B_2 (Wheelbase): Valor p: 3.33e-16 Conclusión: El valor p para B_2también es mucho menor que 0.04, lo que confirma que wheelbase también tiene un efecto significativo sobre el precio. Por cada unidad adicional de wheelbase, el precio promedio aumenta en aproximadamente 443.095 unidades monetarias.

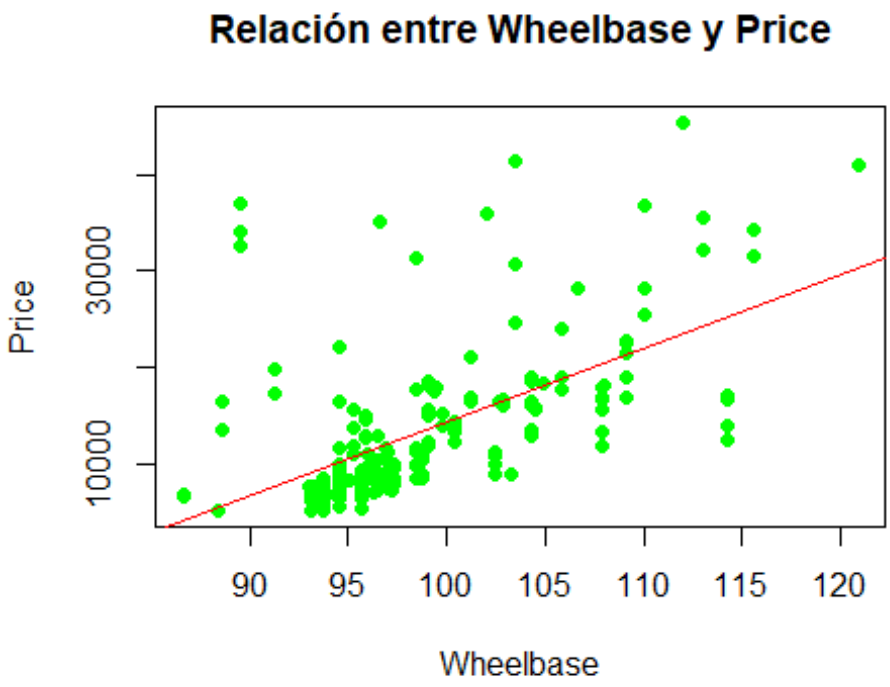
Paso 3: Porcentaje de variación explicada por el modelo

R^2: 0.7482 Interpretación: El modelo explica aproximadamente el 74.82% de la variabilidad en el precio de los vehículos a partir de horsepower y wheelbase. Esto sugiere que ambos son buenos predictores del price, mejorando el ajuste del modelo anterior que solo incluía horsepower.

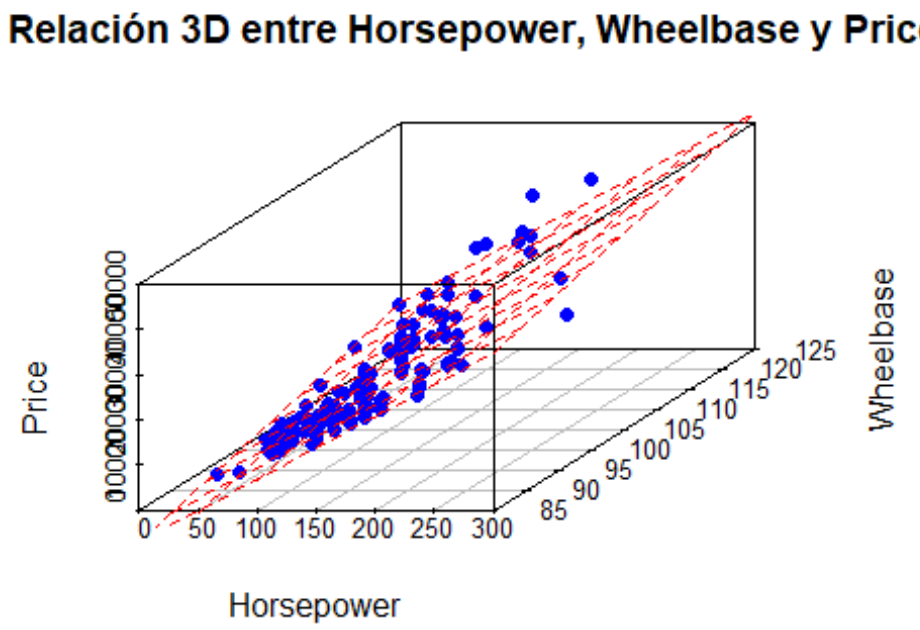
```
# Graficar la relación entre Horsepower y Price
plot(selected_data$horsepower, selected_data$price,
     main = "Relación entre Horsepower y Price",
     xlab = "Horsepower", ylab = "Price", pch = 19, col = "blue")
abline(lm(price ~ horsepower, data = selected_data), col = "red")
```




```
# Graficar la relación entre Wheelbase y Price
plot(selected_data$wheelbase, selected_data$price,
     main = "Relación entre Wheelbase y Price",
     xlab = "Wheelbase", ylab = "Price", pch = 19, col = "green")
abline(lm(price ~ wheelbase, data = selected_data), col = "red")
```




```r
# Graficar la relación combinada entre horsepower, wheelbase y price con el modelo de regresión múltiple
library(scatterplot3d)

# Crear gráfico en 3D
s3d <- scatterplot3d(selected_data$horsepower, selected_data$wheelbase, selected_data$price,
                     main = "Relación 3D entre Horsepower, Wheelbase y Price",
                     xlab = "Horsepower", ylab = "Wheelbase", zlab = "Price", pch = 19, color = "blue")

# Agregar el plano de regresión del modelo2
coef_modelo2 <- coef(modelo2)
s3d$plane3d(coef_modelo2, col = "red")
```



H_0: $B_0, B_1, B_2, B_3 = 0$ H_1: B_0, B_1, B_2, B_3 diferente de 0

```
# Ajustar el modelo de regresión múltiple con interaccion
modelo3 <- lm(price ~ horsepower * wheelbase, data = selected_data)
summary(modelo3)

##
## Call:
## lm(formula = price ~ horsepower * wheelbase, data = selected_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8847  -2050   -177   1350  15889
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -17059.574  14377.287  -1.187   0.2368
## horsepower              -89.721    111.777  -0.803   0.4231
## wheelbase               155.900    148.256   1.052   0.2943
## horsepower:wheelbase      2.342      1.140   2.055   0.0412 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3977 on 201 degrees of freedom
## Multiple R-squared:  0.7558, Adjusted R-squared:  0.7522
## F-statistic: 207.4 on 3 and 201 DF,  p-value: < 2.2e-16
```

Al ajustar un modelo de regresión múltiple con interacción entre las variables horsepower y wheelbase, los resultados son los siguientes:

Fórmula: price ~ horsepower * wheelbase R^2 del modelo: 0.7558 Adjusted R^2: 0.7522 Valor p del modelo: < 2.2e-16 Análisis de los coeficientes:

Intercepto: B_0 Valor p: 0.2368 No significativo (valor p mayor a 0.04), por lo que no podemos rechazar la hipótesis nula de que el intercepto es igual a cero.

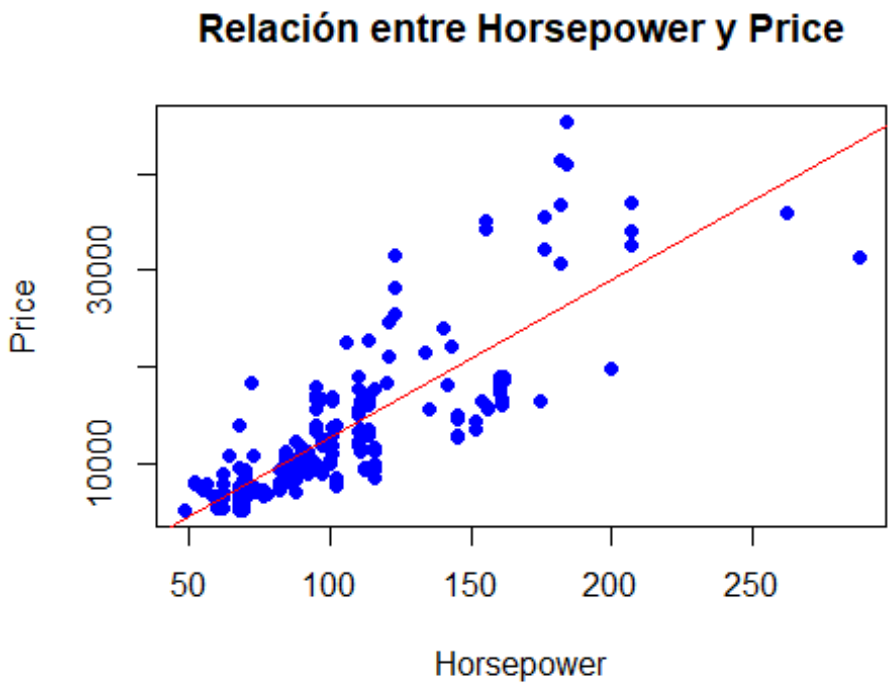
Horsepower: B_1 Valor p: 0.4231 No significativo, lo que indica que al introducir la interacción, la relación directa entre horsepower y price pierde significancia.

Wheelbase:B_2 Valor p: 0.2943 No significativo, lo que sugiere que al incluir la interacción, la relación entre wheelbase y price también deja de ser significativa.

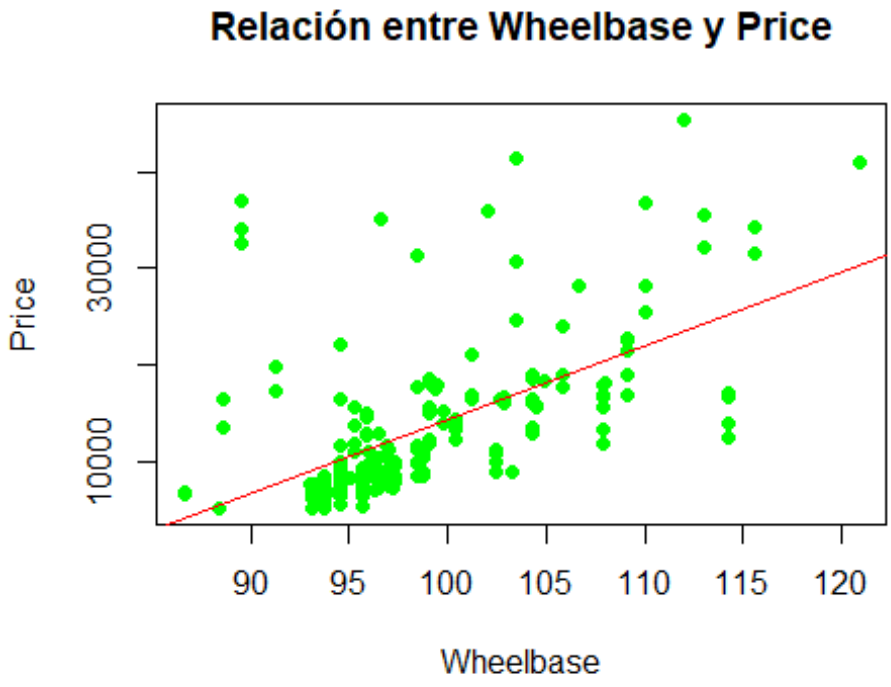
Interacción (Horsepower): B_3 Valor p: 0.0412 Significativa, lo que indica que la interacción entre horsepower y wheelbase afecta de manera conjunta el price.

Conclusión: Aunque el valor de R^2 es mayor (0.7558) comparado con los modelos anteriores, el hecho de que los coeficientes individuales para horsepower y wheelbase no sean significativos implica que este modelo no es el más adecuado.

```r
# Graficar la relación entre Horsepower y Price
plot(selected_data$horsepower, selected_data$price,
     main = "Relación entre Horsepower y Price",
     xlab = "Horsepower", ylab = "Price", pch = 19, col = "blue")
abline(lm(price ~ horsepower, data = selected_data), col = "red")
```

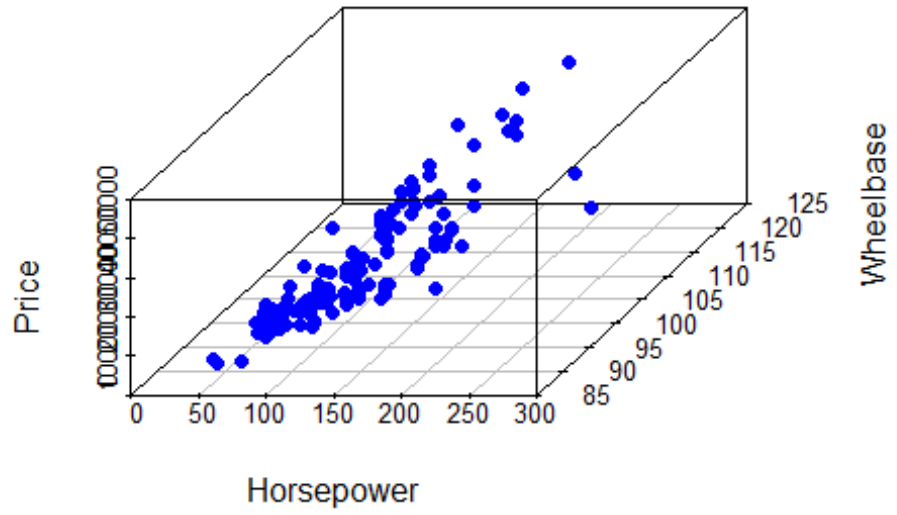


```r
# Graficar la relación entre Wheelbase y Price
plot(selected_data$wheelbase, selected_data$price,
     main = "Relación entre Wheelbase y Price",
     xlab = "Wheelbase", ylab = "Price", pch = 19, col = "green")
abline(lm(price ~ wheelbase, data = selected_data), col = "red")
```




```r
# Graficar la relación con combinada entre horsepower, wheelbase y price con el modelo de regresión múltiple
library(scatterplot3d)

# Crear gráfico en 3D
s3d <- scatterplot3d(selected_data$horsepower, selected_data$wheelbase,
selected_data$price,
                     main = "Relación 3D con interaccion entre Horsepower,
Wheelbase y Price",
                     xlab = "Horsepower", ylab = "Wheelbase", zlab = "Price",
                     pch = 19, color = "blue", grid = TRUE, angle = 55)
```

n 3D con interaccion entre Horsepower, Wheelbase




```
# Agregar el plano de regresión del modelo con interacción
coef_modelo3 <- coef(modelo3)

# Ajustar el modelo de regresión múltiple sin interaccion
modelo4<- lm(price ~ horsepower + wheelbase +fueltype, data = selected_data)
summary(modelo4)
```

```
##
## Call:
## lm(formula = price ~ horsepower + wheelbase + fueltype, data =
selected_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8650  -2191   -197   1606  15816
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -34754.325   5314.194  -6.540 4.99e-10 ***
## horsepower     148.323      7.723  19.205  < 2e-16 ***
## wheelbase      364.657     52.594   6.933 5.48e-11 ***
## fueltypegas  -3794.450   1009.750  -3.758 0.000225 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3884 on 201 degrees of freedom
```

```
## Multiple R-squared:  0.7671, Adjusted R-squared:  0.7636
## F-statistic: 220.7 on 3 and 201 DF,  p-value: < 2.2e-16
```

órmula: price ~ horsepower + wheelbase + fueltype

R^2 del modelo: 0.7671 Adjusted R^2: 0.7636 Valor p del modelo: < 2.2e-16

Análisis de los coeficientes:

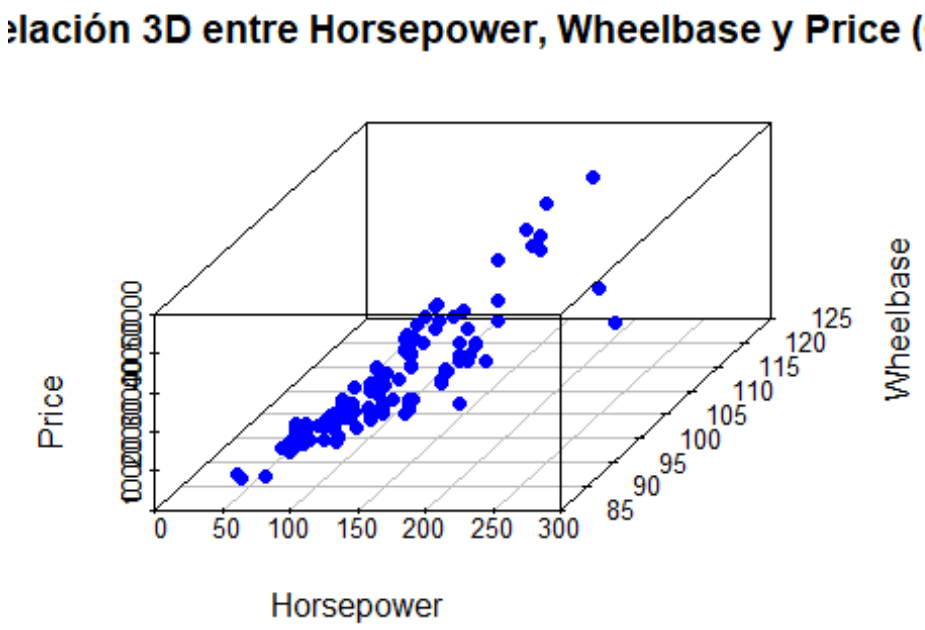
Intercepto: B_0 Valor p: 4.99e-10 Significativo: El valor p es menor que 0.04, lo que indica que el intercepto es significativo, y podemos rechazar la hipótesis nula de que el intercepto es igual a cero.

Horsepower: B_1 Valor p: < 2e-16 Significativo: Dado que el valor p es mucho menor que 0.04, el coeficiente de horsepower es altamente significativo, lo que indica que horsepower tiene un efecto fuerte en el precio.

Wheelbase: B_2 Valor p: 5.48e-11 Significativo: El valor p es menor que 0.04, lo que sugiere que wheelbase es también una variable significativa que afecta el precio de los vehículos.

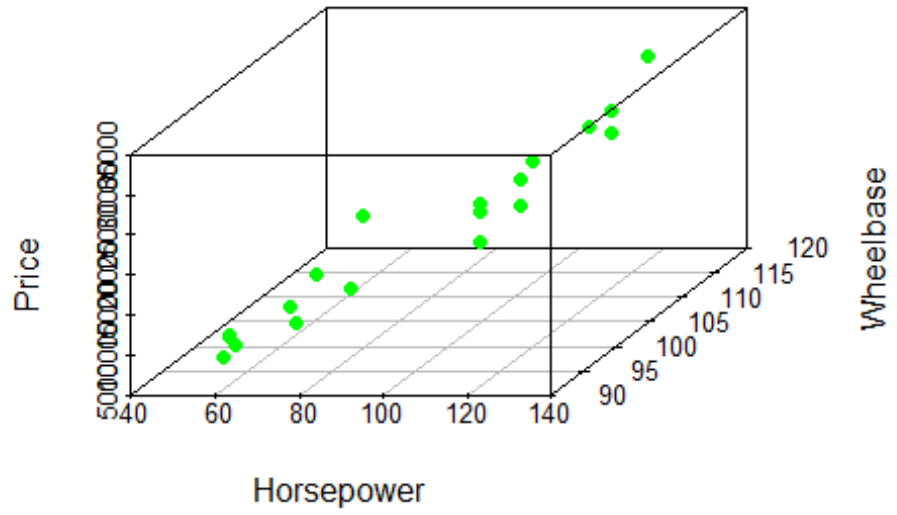
Fueltype (gas): B_3 Valor p: 0.000225 Significativo: Este valor p es menor que 0.04, lo que sugiere que el tipo de combustible (fueltype) es un factor significativo. En este caso, tener un automóvil de tipo gas reduce el precio en comparación con el tipo base.

```
# Filtrar datos por fueltype
selected_gas <- subset(selected_data, fueltype == "gas")
selected_diesel <- subset(selected_data, fueltype == "diesel")

# Graficar la relación combinada entre horsepower, wheelbase y price para gas
y diesel usando scatterplot3d
library(scatterplot3d)

# Gráfico 3D para fueltype = gas
s3d_gas <- scatterplot3d(selected_gas$horsepower, selected_gas$wheelbase,
selected_gas$price,
                         main = "Relación 3D entre Horsepower, Wheelbase y
Price (Gas)",
                         xlab = "Horsepower", ylab = "Wheelbase", zlab =
"Price",
                         pch = 19, color = "blue", grid = TRUE, angle = 55)
```




```r
# Gráfico 3D para fueltype = diesel
s3d_diesel <- scatterplot3d(selected_diesel$horsepower,
selected_diesel$wheelbase, selected_diesel$price,
                            main = "Relación 3D entre Horsepower, Wheelbase y
Price (Diesel)",
                            xlab = "Horsepower", ylab = "Wheelbase", zlab =
"Price",
                            pch = 19, color = "green", grid = TRUE, angle =
55)
```

ación 3D entre Horsepower, Wheelbase y Price (Dies



2.3 Analiza la validez de los modelos propuestos:

Solo seleccionaremos modelo 1,2,4 para las pruebas debido a las conclusiones y resultados anteriores donde descartabamos el 3

Normalidad de los residuos

```
# Cargar la librería nortest
library(nortest)

# Prueba de normalidad para los residuos del Modelo 1
cat("Normalidad de los residuos para el Modelo 1:\n")

## Normalidad de los residuos para el Modelo 1:

ad.test(modelo1$residuals)

##
##  Anderson-Darling normality test
##
## data:  modelo1$residuals
## A = 4.8029, p-value = 6.267e-12

# Prueba de normalidad para los residuos del Modelo 2
cat("Normalidad de los residuos para el Modelo 2:\n")

## Normalidad de los residuos para el Modelo 2:
```

```r
ad.test(modelo2$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  modelo2$residuals
## A = 2.8064, p-value = 4.385e-07
```

```r
# Prueba de normalidad para los residuos del Modelo 4
cat("Normalidad de los residuos para el Modelo 4:\n")
```

```
## Normalidad de los residuos para el Modelo 4:
```

```r
ad.test(modelo4$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  modelo4$residuals
## A = 2.7561, p-value = 5.82e-07
```

$$\alpha = 0.04$$

Modelo 1:

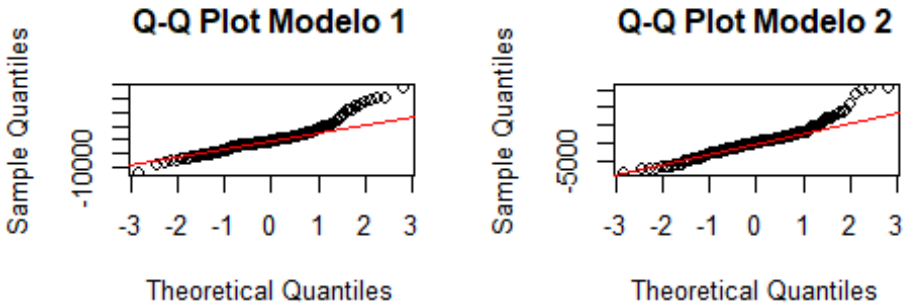
Valor p: 6.267e-12 Conclusión: Dado que el valor p es mucho menor que α=0.04rechazamos la hipótesis nula de normalidad. Esto sugiere que los residuos no son normales para el Modelo 1.

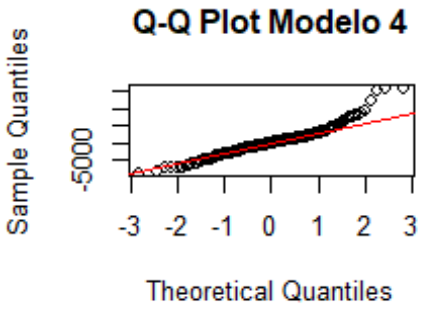
Modelo 2: Valor p: 4.385e-07 Conclusión: Al igual que en el Modelo 1, el valor p es menor que α=0.04 por lo que también rechazamos la hipótesis de normalidad para los residuos del Modelo 2. Los residuos no son normales en el Modelo 2. Aunque son mejores que en el 1

MOdelo 4: Valor p = 5.82e-07 Conclusion: Al igual que los modelos anteriores el valor p es menor que α=0.04 por lo que también rechazamos la hipótesis de normalidad para los residuos del Modelo

```r
# Instalar si no está la librería nortest
# install.packages("nortest")
library(nortest)

# Gráfico Q-Q para visualizar la normalidad de los residuos
par(mfrow = c(2, 2))  # Mostrar cuatro gráficos (dos filas y dos columnas)

# Gráfico Q-Q para el modelo1
qqnorm(modelo1$residuals, main = "Q-Q Plot Modelo 1")
qqline(modelo1$residuals, col = "red")

# Gráfico Q-Q para el modelo2
qqnorm(modelo2$residuals, main = "Q-Q Plot Modelo 2")
```

```r
qqline(modelo2$residuals, col = "red")

# Gráfico Q-Q para el modelo4 (nuevo modelo agregado)
qqnorm(modelo4$residuals, main = "Q-Q Plot Modelo 4")
qqline(modelo4$residuals, col = "red")


# Restaurar la configuración original
par(mfrow = c(1, 1))
```





Podemos ver que aunque su tendencia en ambos es algo mala en las colas debido a su desviacion, parece pronunciarse mas en el primero, mientras que en el segundo es un poco mejor, y no parece difererir tanto en el cuarto modelo del segundo

Media 0

```r
# Prueba t para verificar si la media de los residuos es cero (Modelo 1)
cat("Verificación de media cero para el Modelo 1:\n")

## Verificación de media cero para el Modelo 1:

t.test(modelo1$residuals)

##
##  One Sample t-test
##
## data:  modelo1$residuals
## t = -2.2725e-16, df = 204, p-value = 1
```

```
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -647.9614  647.9614
## sample estimates:
##     mean of x
## -7.468281e-14

# Prueba t para verificar si la media de los residuos es cero (Modelo 2)
cat("Verificación de media cero para el Modelo 2:\n")

## Verificación de media cero para el Modelo 2:

t.test(modelo2$residuals)

##
##  One Sample t-test
##
## data:  modelo2$residuals
## t = 2.4215e-16, df = 204, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -549.2714  549.2714
## sample estimates:
##    mean of x
## 6.745823e-14

cat("Verificación de media cero para el Modelo 4:\n")

## Verificación de media cero para el Modelo 4:

t.test(modelo4$residuals)

##
##  One Sample t-test
##
## data:  modelo4$residuals
## t = -2.6083e-16, df = 204, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -530.9376  530.9376
## sample estimates:
##     mean of x
## -7.023758e-14
```

Modelo 1: Valor p: 1 Conclusión: Dado que el valor p es mayor que 0.04 no rechazamos la hipótesis nula de que la media de los residuos es igual a cero. Esto indica que el modelo no tiene un sesgo sistemático en los residuos.

Modelo 2: Valor p: 1 Conclusión: Al igual que en el Modelo 1, no rechazamos la hipótesis de que la media de los residuos es igual a cero. Esto sugiere que el Modelo 2 también predice sin errores sistemáticos en sus residuos.

Modelo 4: Valor p: 1 Conclusión: Al igual que en el Modelo 1 y 2, no rechazamos la hipótesis de que la media de los residuos es igual a cero. Esto sugiere que el Modelo 4 también predice sin errores sistemáticos en sus residuos.

Homocedasticidad, linealidad e independencia

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

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

```
# Prueba de homocedasticidad para el Modelo 1
cat("Prueba de homocedasticidad para el Modelo 1:\n")
```

```
## Prueba de homocedasticidad para el Modelo 1:
```

```
bptest(modelo1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  modelo1
## BP = 54.573, df = 1, p-value = 1.497e-13
```

```
# Prueba de homocedasticidad para el Modelo 2
cat("Prueba de homocedasticidad para el Modelo 2:\n")
```

```
## Prueba de homocedasticidad para el Modelo 2:
```

```
bptest(modelo2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  modelo2
## BP = 57.388, df = 2, p-value = 3.455e-13
```

```
# Prueba de homocedasticidad para el Modelo 4
cat("Prueba de homocedasticidad para el Modelo 4:\n")
```

```
## Prueba de homocedasticidad para el Modelo 4:
```

```
bptest(modelo4)
```

```
##
##  studentized Breusch-Pagan test
##
```

```
## data:  modelo4
## BP = 62.451, df = 3, p-value = 1.759e-13
```

```
# Prueba de Durbin-Watson para independencia de los residuos (Modelo 1)
cat("Prueba de independencia de los residuos (Modelo 1):\n")
```

```
## Prueba de independencia de los residuos (Modelo 1):
```

```
dwtest(modelo1)
```

```
##
##  Durbin-Watson test
##
## data:  modelo1
## DW = 0.79229, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
# Prueba de Durbin-Watson para independencia de los residuos (Modelo 2)
cat("Prueba de independencia de los residuos (Modelo 2):\n")
```

```
## Prueba de independencia de los residuos (Modelo 2):
```

```
dwtest(modelo2)
```

```
##
##  Durbin-Watson test
##
## data:  modelo2
## DW = 0.98038, p-value = 5.339e-14
## alternative hypothesis: true autocorrelation is greater than 0
```

```
# Prueba de Durbin-Watson para independencia de los residuos (Modelo 4)
cat("Prueba de independencia de los residuos (Modelo 4):\n")
```

```
## Prueba de independencia de los residuos (Modelo 4):
```

```
dwtest(modelo4)
```

```
##
##  Durbin-Watson test
##
## data:  modelo4
## DW = 0.97856, p-value = 4.496e-14
## alternative hypothesis: true autocorrelation is greater than 0
```

Homocedasticidad (Prueba de Breusch-Pagan)

Modelo 1: Valor p: 1.497e-13 Conclusión: El valor p es mucho menor que 0.04 por lo que rechazamos la hipótesis nula de homocedasticidad. Esto indica que existe heterocedasticidad en los residuos del Modelo 1, lo que significa que la varianza de los residuos no es constante.

Modelo 2: Valor p: 3.455e-13 Conclusión: Al igual que en el Modelo 1, el valor p es mucho menor que 0,04 lo que sugiere que también existe heterocedasticidad en los residuos del Modelo 2. La varianza de los residuos no es constante en este modelo.

Modelo 4: Valor p:1.759e-13 Conclusión: Al igual que en el Modelo 1, el valor p es mucho menor que 0,04 lo que sugiere que también existe heterocedasticidad en los residuos del Modelo 2. La varianza de los residuos no es constante en este modelo.

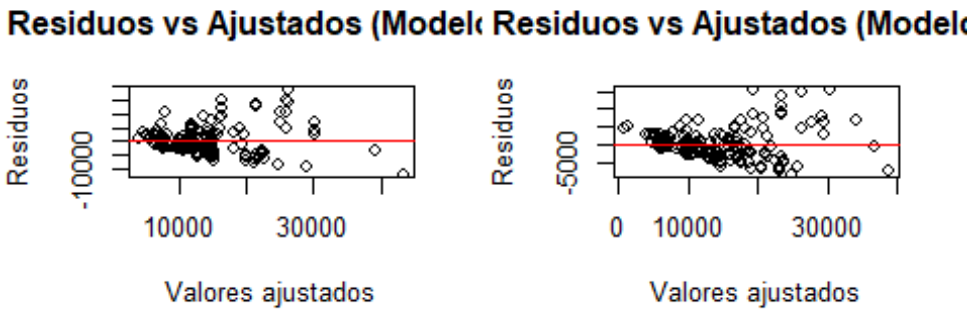
Independencia de los residuos (Prueba de Durbin-Watson)

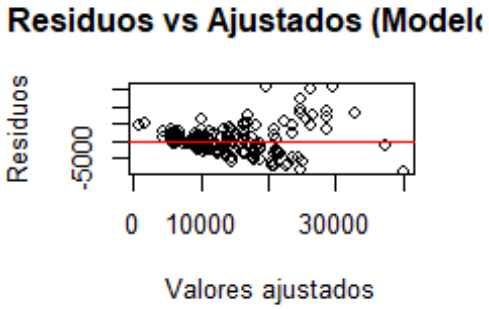
Modelo 1: DW = 0.79229, valor p < 2.2e-16 Conclusión: El valor p es mucho menor que 0.04 lo que indica que rechazamos la hipótesis nula de independencia. Existe autocorrelación en los residuos del Modelo 1, lo que afecta la validez del modelo.

Modelo 2: DW = 0.98038, valor p 5.339e-14 Conclusión: El valor p es menor que 0.04 lo que indica que también existe autocorrelación en los residuos del Modelo 2. Los residuos no son independientes, lo que puede influir en la interpretación del modelo.

Modelo 4: DW = 0.97856, p-value = 4.496e-14 Conclusión: El valor p es menor que 0.04 lo que indica que también existe autocorrelación en los residuos del Modelo 4. Los residuos no son independientes, lo que puede influir en la interpretación del modelo.

```r
# Gráfico de residuos vs valores ajustados para visualizar homocedasticidad (ambos modelos)
par(mfrow = c(2, 2))

plot(modelo1$fitted.values, modelo1$residuals,
     main = "Residuos vs Ajustados (Modelo 1)",
     xlab = "Valores ajustados", ylab = "Residuos")
abline(h = 0, col = "red")

plot(modelo2$fitted.values, modelo2$residuals,
     main = "Residuos vs Ajustados (Modelo 2)",
     xlab = "Valores ajustados", ylab = "Residuos")
abline(h = 0, col = "red")

plot(modelo4$fitted.values, modelo4$residuals,
     main = "Residuos vs Ajustados (Modelo 4)",
     xlab = "Valores ajustados", ylab = "Residuos")
abline(h = 0, col = "red")

par(mfrow = c(1, 1))  # Restaurar configuración original
```





2.4 Emite una conclusión final sobre el mejor modelo de regresión lineal y contesta la pregunta central:

Concluye sobre el mejor modelo que encontraste y argumenta por qué es el mejor ¿Cuáles de las variables asignadas influyen en el precio del auto? ¿de qué manera lo hacen?

El mejor modelo es el 2 y 4, debido a la varianza explica por su r^2 muy superior al 1, y descratamos al 3 por la insignifcancia de sus coeficientes por lo tanto, horsepower y wheelbase ayudan y son mas significantes en la relevancia de este modelo, lo que nos senala que influyen mas en el precio del auto.

sin embargo todos fallan estrepitosamente en las pruebas de Normalidad, lo que sugiere que los datos no se comportan de forma normal. Por lo tanto aunque no es perfectamente normal, el modelo 2 y 4 serian los escogidos, podriamos irnos por el 2 debido a que el ligero incremento en el r^2 en el 4 podria ser debido a la cantidad de variables y no tanto a la influencia de la variable de combustible, pero dado que es significativa podriamos considerarlo el mejor hasta el momento y seleccionado. Por lo tanto el 4

3. Intervalos de predicción y confianza

Con los datos de las variables asignadas construye la gráfica de los intervalos de confianza y predicción para la estimación y predicción del precio para el mejor modelo seleccionado: Calcula los intervalos para la variable Y Selecciona la categoría de la variable cualitativa que, de acuerdo a tu análisis resulte la más importante, y separa la base de datos por esa

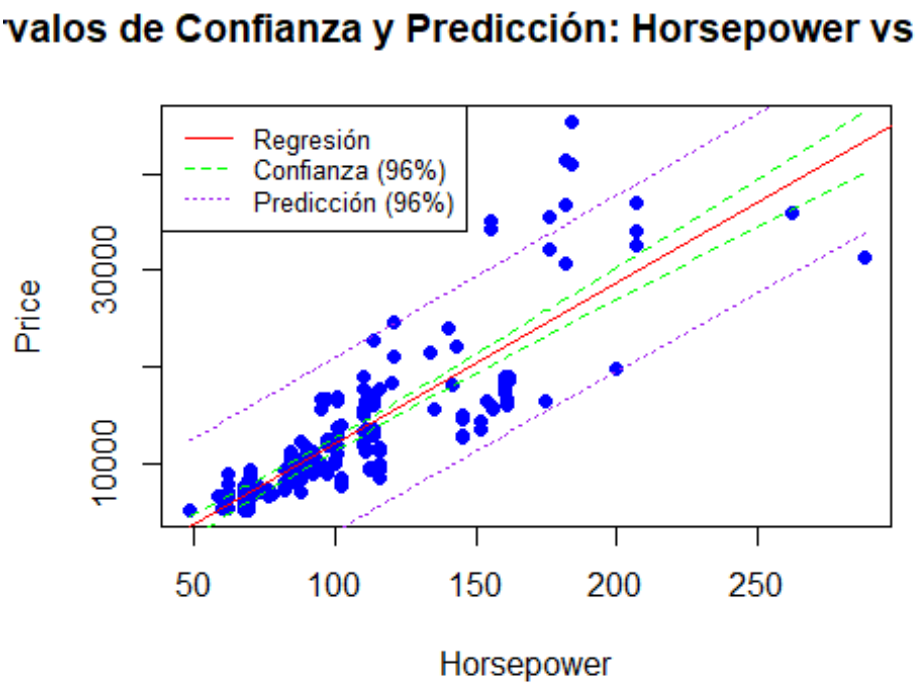
variable categórica. Grafica por pares de variables numéricas Puedes hacer el mismo análisis para otra categoría de la variable cualitativa, pero no es necesario, bastará con que justiques la categoría seleccionada anteriormente. Interpreta en el contexto del problema

Dado que utilizamos el modelo 4 podemos ver que tendremos que hacer 2 graficas por cada categoria, esto debido a las 2 variables numericas que compararemos con price, y en total serian 4 si utilizamos las 2 categorias

Gas

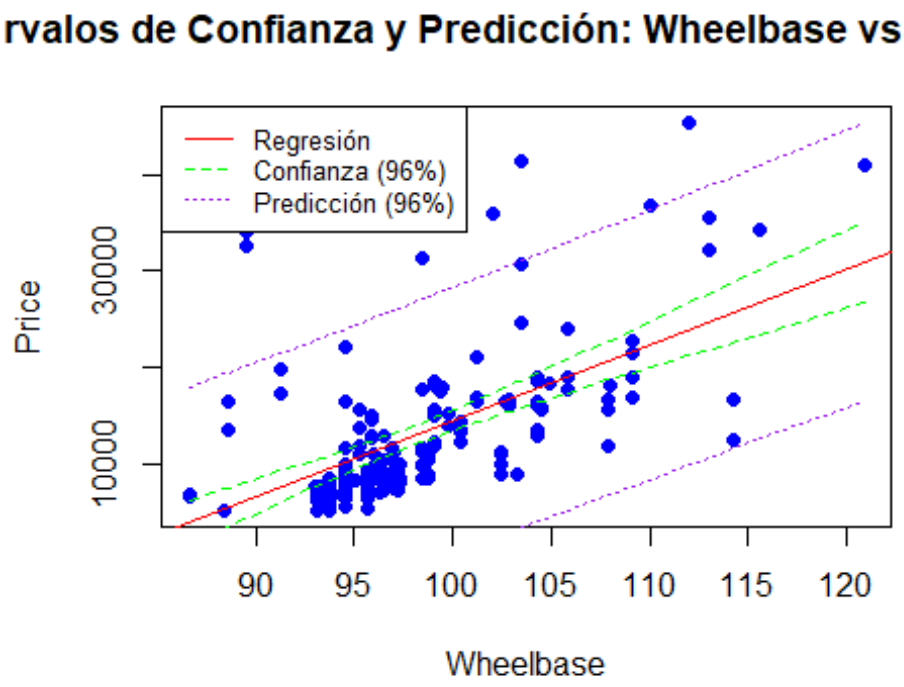
```
# Filtrar los datos para fueltype = gas
selected_gas <- subset(selected_data, fueltype == "gas")

# Calcular los intervalos de confianza y predicción para horsepower
intervalos_confianza_hp <- predict(lm(price ~ horsepower, data =
selected_gas),
                                   interval = "confidence", level = 0.96)
intervalos_prediccion_hp <- predict(lm(price ~ horsepower, data =
selected_gas),
                                    interval = "prediction", level = 0.96)

## Warning in predict.lm(lm(price ~ horsepower, data = selected_gas),
interval = "prediction", : predictions on current data refer to _future_
responses

# Ordenar los valores de horsepower para una mejor visualización
orden_hp <- order(selected_gas$horsepower)

# Graficar los puntos de horsepower vs price
plot(selected_gas$horsepower, selected_gas$price,
     main = "Intervalos de Confianza y Predicción: Horsepower vs Price
(Gas)",
     xlab = "Horsepower", ylab = "Price", pch = 19, col = "blue")

# Agregar línea de regresión
abline(lm(price ~ horsepower, data = selected_gas), col = "red")

# Agregar intervalos de confianza y predicción ordenados
lines(selected_gas$horsepower[orden_hp], intervalos_confianza_hp[orden_hp,
"lwr"], col = "green", lty = 2)
lines(selected_gas$horsepower[orden_hp], intervalos_confianza_hp[orden_hp,
"upr"], col = "green", lty = 2)
lines(selected_gas$horsepower[orden_hp], intervalos_prediccion_hp[orden_hp,
"lwr"], col = "purple", lty = 3)
lines(selected_gas$horsepower[orden_hp], intervalos_prediccion_hp[orden_hp,
"upr"], col = "purple", lty = 3)

# Agregar leyenda
legend("topleft", legend = c("Regresión", "Confianza (96%)", "Predicción
```

```
(96%)"),
       col = c("red", "green", "purple"), lty = c(1, 2, 3), cex = 0.8)
```




```
# Calcular los intervalos de confianza y predicción para wheelbase
intervalos_confianza_wb <- predict(lm(price ~ wheelbase, data =
selected_gas),
                                   interval = "confidence", level = 0.96)
intervalos_prediccion_wb <- predict(lm(price ~ wheelbase, data =
selected_gas),
                                    interval = "prediction", level = 0.96)

## Warning in predict.lm(lm(price ~ wheelbase, data = selected_gas), interval
= "prediction", : predictions on current data refer to _future_ responses

# Ordenar los valores de wheelbase para una mejor visualización
orden_wb <- order(selected_gas$wheelbase)

# Graficar los puntos de wheelbase vs price
plot(selected_gas$wheelbase, selected_gas$price,
     main = "Intervalos de Confianza y Predicción: Wheelbase vs Price (Gas)",
     xlab = "Wheelbase", ylab = "Price", pch = 19, col = "blue")

# Agregar línea de regresión
abline(lm(price ~ wheelbase, data = selected_gas), col = "red")

# Agregar intervalos de confianza y predicción ordenados
lines(selected_gas$wheelbase[orden_wb], intervalos_confianza_wb[orden_wb,
```

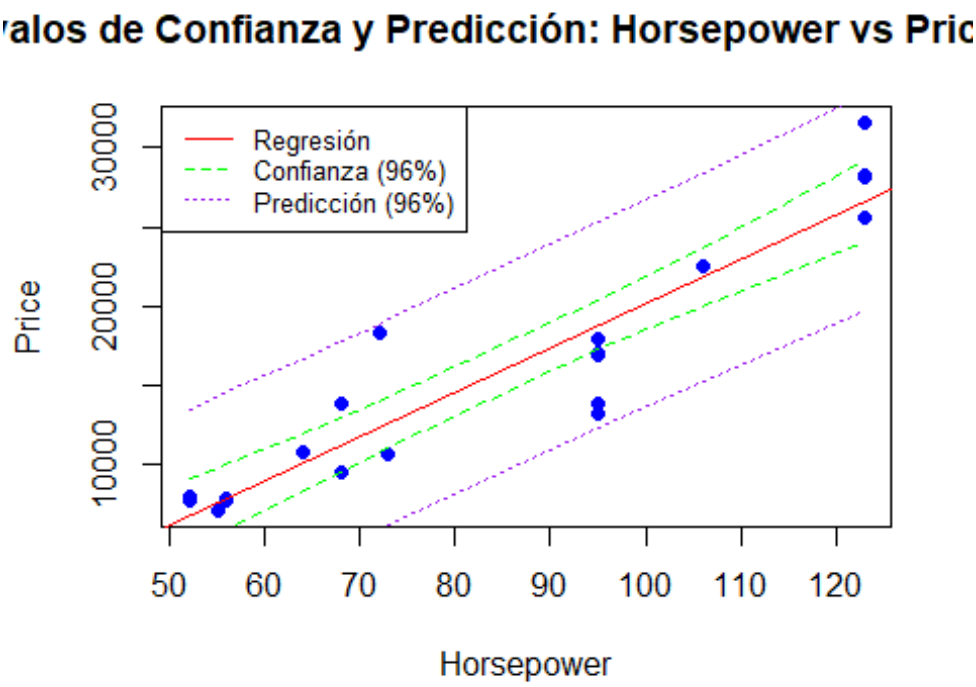
```
"lwr"], col = "green", lty = 2)
lines(selected_gas$wheelbase[orden_wb], intervalos_confianza_wb[orden_wb,
"upr"], col = "green", lty = 2)
lines(selected_gas$wheelbase[orden_wb], intervalos_prediccion_wb[orden_wb,
"lwr"], col = "purple", lty = 3)
lines(selected_gas$wheelbase[orden_wb], intervalos_prediccion_wb[orden_wb,
"upr"], col = "purple", lty = 3)

# Agregar leyenda
legend("topleft", legend = c("Regresión", "Confianza (96%)", "Predicción
(96%)"),
       col = c("red", "green", "purple"), lty = c(1, 2, 3), cex = 0.8)
```



Diesel

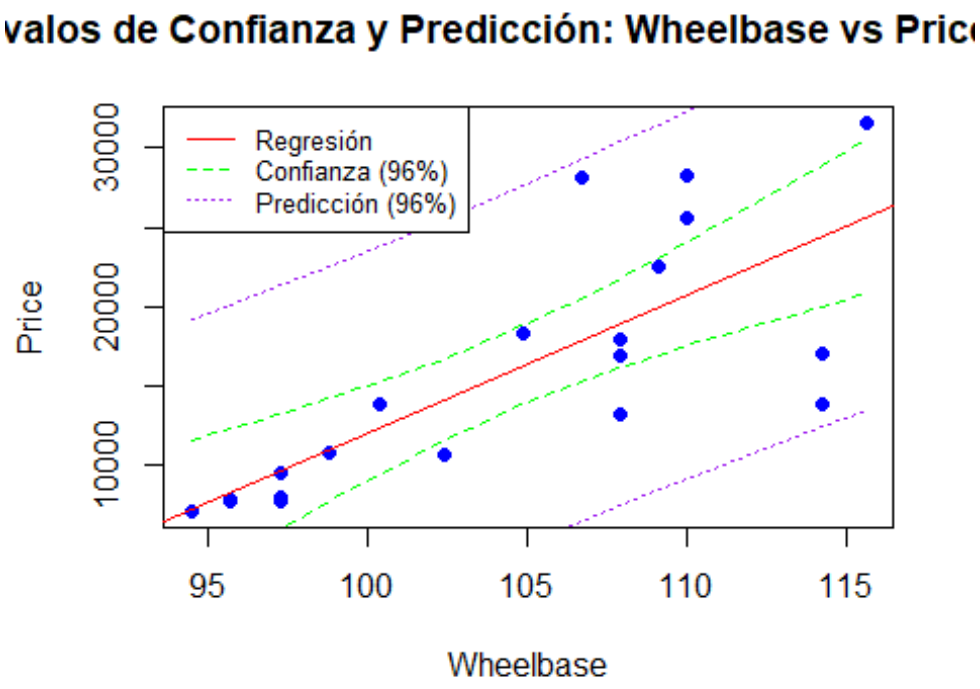
```
# Filtrar los datos para fueltype = diesel
selected_diesel <- subset(selected_data, fueltype == "diesel")

# Calcular los intervalos de confianza y predicción para horsepower
intervalos_confianza_hp_diesel <- predict(lm(price ~ horsepower, data =
selected_diesel),
                                          interval = "confidence", level =
0.96)
intervalos_prediccion_hp_diesel <- predict(lm(price ~ horsepower, data =
selected_diesel),
                                           interval = "prediction", level =
0.96)
```

```
## Warning in predict.lm(lm(price ~ horsepower, data = selected_diesel),
interval = "prediction", : predictions on current data refer to _future_
responses

# Ordenar los valores de horsepower para una mejor visualización
orden_hp_diesel <- order(selected_diesel$horsepower)

# Graficar los puntos de horsepower vs price para diesel
plot(selected_diesel$horsepower, selected_diesel$price,
     main = "Intervalos de Confianza y Predicción: Horsepower vs Price
(Diesel)",
     xlab = "Horsepower", ylab = "Price", pch = 19, col = "blue")

# Agregar línea de regresión para horsepower vs price (diesel)
abline(lm(price ~ horsepower, data = selected_diesel), col = "red")

# Agregar intervalos de confianza y predicción ordenados
lines(selected_diesel$horsepower[orden_hp_diesel],
intervalos_confianza_hp_diesel[orden_hp_diesel, "lwr"], col = "green", lty =
2)
lines(selected_diesel$horsepower[orden_hp_diesel],
intervalos_confianza_hp_diesel[orden_hp_diesel, "upr"], col = "green", lty =
2)
lines(selected_diesel$horsepower[orden_hp_diesel],
intervalos_prediccion_hp_diesel[orden_hp_diesel, "lwr"], col = "purple", lty
= 3)
lines(selected_diesel$horsepower[orden_hp_diesel],
intervalos_prediccion_hp_diesel[orden_hp_diesel, "upr"], col = "purple", lty
= 3)

# Agregar leyenda
legend("topleft", legend = c("Regresión", "Confianza (96%)", "Predicción
(96%)"),
       col = c("red", "green", "purple"), lty = c(1, 2, 3), cex = 0.8)
```




```r
# Calcular los intervalos de confianza y predicción para wheelbase
intervalos_confianza_wb_diesel <- predict(lm(price ~ wheelbase, data =
selected_diesel),
                                          interval = "confidence", level =
0.96)
intervalos_prediccion_wb_diesel <- predict(lm(price ~ wheelbase, data =
selected_diesel),
                                           interval = "prediction", level =
0.96)

## Warning in predict.lm(lm(price ~ wheelbase, data = selected_diesel),
interval = "prediction", : predictions on current data refer to _future_
responses

# Ordenar los valores de wheelbase para una mejor visualización
orden_wb_diesel <- order(selected_diesel$wheelbase)

# Graficar los puntos de wheelbase vs price para diesel
plot(selected_diesel$wheelbase, selected_diesel$price,
     main = "Intervalos de Confianza y Predicción: Wheelbase vs Price
(Diesel)",
     xlab = "Wheelbase", ylab = "Price", pch = 19, col = "blue")

# Agregar línea de regresión para wheelbase vs price (diesel)
abline(lm(price ~ wheelbase, data = selected_diesel), col = "red")

# Agregar intervalos de confianza y predicción ordenados
```

```r
lines(selected_diesel$wheelbase[orden_wb_diesel],
intervalos_confianza_wb_diesel[orden_wb_diesel, "lwr"], col = "green", lty =
2)
lines(selected_diesel$wheelbase[orden_wb_diesel],
intervalos_confianza_wb_diesel[orden_wb_diesel, "upr"], col = "green", lty =
2)
lines(selected_diesel$wheelbase[orden_wb_diesel],
intervalos_prediccion_wb_diesel[orden_wb_diesel, "lwr"], col = "purple", lty
= 3)
lines(selected_diesel$wheelbase[orden_wb_diesel],
intervalos_prediccion_wb_diesel[orden_wb_diesel, "upr"], col = "purple", lty
= 3)

# Agregar Leyenda
legend("topleft", legend = c("Regresión", "Confianza (96%)", "Predicción
(96%)"),
       col = c("red", "green", "purple"), lty = c(1, 2, 3), cex = 0.8)
```



Podemos comparar en estos graficos por intervalos la distribucion y comparativa de las variables numericas de horse y wheel a traves de las diferentes categorias de combustible

1. Gas - Horsepower vs Price: En la gráfica de Horsepower vs Price para vehículos con fueltype = gas, se puede observar una tendencia positiva entre horsepower y price. A medida que aumenta el horsepower, el precio del vehículo también tiende a aumentar.

Línea de Regresión (roja): Representa la relación estimada entre horsepower y price. La inclinación positiva confirma que los vehículos con mayor horsepower tienen un precio más elevado. Intervalos de Confianza (verdes): Los intervalos son bastante ajustados en la parte central de los datos, lo que indica una buena precisión en la estimación del promedio de precios. *Intervalos de Predicción (morados): Son más amplios, lo cual es normal, ya que se trata de predecir un valor individual, y existe más incertidumbre en las predicciones individuales.

2. Gas - Wheelbase vs Price: En la gráfica de Wheelbase vs Price para vehículos de fueltype = gas, la relación es también positiva pero menos pronunciada que con horsepower.

Línea de Regresión (roja): Aunque la pendiente es positiva, la relación entre wheelbase y price es más débil que la observada con horsepower. Intervalos de Confianza (verdes): La dispersión de los puntos muestra que existe cierta variabilidad en los precios, pero la relación es lo suficientemente fuerte como para generar confianza en la predicción. *Intervalos de Predicción (morados): Al igual que con horsepower, los intervalos de predicción son más amplios, indicando mayor incertidumbre en predicciones individuales.

3. Diesel - Horsepower vs Price: En la gráfica de Horsepower vs Price para vehículos con fueltype = diesel, se observa una relación similar a la del caso de gas, pero con menos puntos de datos debido a la menor cantidad de vehículos diesel en la base de datos.

Línea de Regresión (roja): La pendiente sigue siendo positiva, lo que sugiere que el aumento en horsepower conduce a un aumento en el precio de los vehículos diesel. Intervalos de Confianza (verdes): Son relativamente ajustados, aunque menos que en el caso de gas, lo que sugiere una estimación promedio más incierta. *Intervalos de Predicción (morados): Nuevamente, son más amplios debido a la incertidumbre en las predicciones individuales.

4. Diesel - Wheelbase vs Price: En la gráfica de Wheelbase vs Price para vehículos de fueltype = diesel, se puede observar una relación positiva entre wheelbase y price, aunque más débil y con más dispersión que en horsepower.

Línea de Regresión (roja): Indica una relación positiva, pero es menos fuerte comparada con horsepower. Intervalos de Confianza (verdes): Son relativamente más amplios que en las otras gráficas, sugiriendo que hay mayor variabilidad en la relación entre wheelbase y price para los vehículos diesel. *Intervalos de Predicción (morados): Los intervalos de predicción son mucho más amplios, lo que indica que las predicciones individuales del precio para vehículos diesel son más inciertas.

4. Más allá:

Contesta la pregunta referida a la agrupación de variables que propuso la empresa para el análisis: ¿propondrías una nueva agrupación de las variables a la empresa automovilísitica?

Retoma todas las variables y haz un análisis estadístico muy leve (medias y correlación) de cómo crees que se deberían agrupar para analizarlas.

En mi opinion, el conjunto de las variable utilizadas no son las mejores para esta situacion, honestamente siento que el estudio con el tipo de combustible es muy ineficiente ya que no beneficia en demasia al modelo, cambiaria el grupo porque aunque con el modelo 2 y 4 tenemos resultados medianamente decentes, podemos ver en normalidad que fallan la mayoria de las pruebas sugiriendo y viendo que los datos no tienen esa naturaleza de distribucion

```
# Seleccionar solo las columnas numéricas
numerical_columns <- sapply(data, is.numeric)
numerical_data <- data[, numerical_columns]

# Calcular la matriz de correlación
correlation_matrix <- cor(numerical_data)

# Mostrar la matriz de correlación
print(correlation_matrix)
```

```
##                    symboling  wheelbase  carlength   carwidth   carheight
## symboling         1.000000000 -0.5319537 -0.3576115 -0.2329191 -0.54103820
## wheelbase        -0.531953682  1.0000000  0.8745875  0.7951436  0.58943476
## carlength        -0.357611523  0.8745875  1.0000000  0.8411183  0.49102946
## carwidth         -0.232919061  0.7951436  0.8411183  1.0000000  0.27921032
## carheight        -0.541038200  0.5894348  0.4910295  0.2792103  1.00000000
## curbweight       -0.227690588  0.7763863  0.8777285  0.8670325  0.29557173
## enginesize       -0.105789709  0.5693287  0.6833599  0.7354334  0.06714874
## stroke           -0.008735141  0.1609590  0.1295326  0.1829417 -0.05530667
## compressionratio -0.178515084  0.2497858  0.1584137  0.1811286  0.26121423
## horsepower        0.070872724  0.3532945  0.5526230  0.6407321 -0.10880206
## peakrpm           0.273606245 -0.3604687 -0.2872422 -0.2200123 -0.32041072
## citympg          -0.035822628 -0.4704136 -0.6709087 -0.6427043 -0.04863963
## highwaympg        0.034606001 -0.5440819 -0.7046616 -0.6772179 -0.10735763
## price            -0.079978225  0.5778156  0.6829200  0.7593253  0.11933623
##                  curbweight  enginesize       stroke compressionratio
## symboling        -0.2276906 -0.10578971 -0.008735141      -0.17851508
## wheelbase         0.7763863  0.56932868  0.160959047       0.24978585
## carlength         0.8777285  0.68335987  0.129532611       0.15841371
## carwidth          0.8670325  0.73543340  0.182941693       0.18112863
## carheight         0.2955717  0.06714874 -0.055306674       0.26121423
## curbweight        1.0000000  0.85059407  0.168790035       0.15136174
```

```
## enginesize        0.8505941  1.00000000  0.203128588       0.02897136
## stroke            0.1687900  0.20312859  1.000000000       0.18611011
## compressionratio  0.1513617  0.02897136  0.186110110       1.00000000
## horsepower        0.7507393  0.80976865  0.080939536      -0.20432623
## peakrpm          -0.2662432 -0.24465983 -0.067963753      -0.43574051
## citympg          -0.7574138 -0.65365792 -0.042144754       0.32470142
## highwaympg       -0.7974648 -0.67746991 -0.043930930       0.26520139
## price             0.8353049  0.87414480  0.079443084       0.06798351
##                   horsepower     peakrpm     citympg  highwaympg
price
## symboling         0.07087272  0.27360625 -0.03582263  0.03460600 -
0.07997822
## wheelbase         0.35329448 -0.36046875 -0.47041361 -0.54408192
0.57781560
## carlength         0.55262297 -0.28724220 -0.67090866 -0.70466160
0.68292002
## carwidth          0.64073208 -0.22001230 -0.64270434 -0.67721792
0.75932530
## carheight        -0.10880206 -0.32041072 -0.04863963 -0.10735763
0.11933623
## curbweight        0.75073925 -0.26624318 -0.75741378 -0.79746479
0.83530488
## enginesize        0.80976865 -0.24465983 -0.65365792 -0.67746991
0.87414480
## stroke            0.08093954 -0.06796375 -0.04214475 -0.04393093
0.07944308
## compressionratio -0.20432623 -0.43574051  0.32470142  0.26520139
0.06798351
## horsepower        1.00000000  0.13107251 -0.80145618 -0.77054389
0.80813882
## peakrpm           0.13107251  1.00000000 -0.11354438 -0.05427481 -
0.08526715
## citympg          -0.80145618 -0.11354438  1.00000000  0.97133704 -
0.68575134
## highwaympg       -0.77054389 -0.05427481  0.97133704  1.00000000 -
0.69759909
## price             0.80813882 -0.08526715 -0.68575134 -0.69759909
1.00000000
```

1. Wheelbase: Correlación con price: 0.577 (moderada). Correlación con horsepower: 0.353 (baja).

2. Carwidth: Correlación con price: 0.759 (alta). Correlación con horsepower: 0.640 (moderada).

3. Carlength: Correlación con price: 0.682 (alta). Correlación con horsepower: 0.552 (moderada).

Para maximizar la independencia entre las variables y asegurar que no estoy introduciendo multicolinealidad, puedo seleccionar las siguientes variables:

Wheelbase: Tiene una correlación baja con horsepower y una correlación moderada con price, por lo que es una buena opción. Carwidth: Aunque tiene una correlación moderada con horsepower, su alta correlación con price la hace valiosa. Puedo incluirla si el nivel de multicolinealidad es aceptable. *Carlength: Similar a carwidth, es útil, pero debo monitorear la multicolinealidad.