

Actividad Integradora 2

Saúl Francisco Vázquez del Río

2024-09-06

Problema

Una empresa automovilística china aspira a entrar en el mercado estadounidense. Desea establecer allí una unidad de fabricación y producir automóviles localmente para competir con sus contrapartes estadounidenses y europeas. Contrataron una empresa de consultoría de automóviles para identificar los principales factores de los que depende el precio de los automóviles, específicamente, en el mercado estadounidense, ya que pueden ser muy diferentes del mercado chino. Esencialmente, la empresa quiere saber:

Qué variables son significativas para predecir el precio de un automóvil
Qué tan bien describen esas variables el precio de un automóvil
Con base en varias encuestas de mercado, la consultora ha recopilado un gran conjunto de datos de diferentes tipos de automóviles en el mercado estadounidense que presenta en el siguiente archivo Download archivo. Las variables recopiladas vienen descritas en el diccionario de términos Download diccionario de términos. Por un análisis de correlación, la empresa automovilística tiene interés en analizar las variables agrupadas de la siguiente forma para hacer el análisis de variables significativas:

Primer grupo. Distancia entre los ejes (wheelbase), tipo de gasolina que usa y caballos de fuerza
Segundo grupo. Altura del auto, ancho del auto y si es convertible o no.
Tercer grupo. Tamaño del motor (ensinesize), carrera o lanzamiento del pistón (stroke) y localización del motor en el carro
Selecciona uno de los tres grupos analizados (te será asignado por tu profesora) y analiza la significancia de las variables para predecir o influir en la variable precio. ¿propondrías una nueva agrupación a la empresa automovilística?

##Problema 1 Exploración de la base de datos

```
M=read.csv("C:\\Users\\saulv\\OneDrive\\Escritorio\\Septimo
semestre\\precios_autos.csv") #Leer la base de datos
head(M)
```

```
##      symboling      CarName fueltype      carbody drivewheel
## 1           3      alfa-romero giulia      gas convertible      rwd
## 2           3      alfa-romero stelvio      gas convertible      rwd
## 3           1 alfa-romero Quadrifoglio      gas  hatchback      rwd
## 4           2          audi 100 ls      gas      sedan      fwd
## 5           2          audi 100ls      gas      sedan      4wd
## 6           2          audi fox      gas      sedan      fwd
##      enginelocation wheelbase carlength carwidth carheight curbweight
```

```

enginetype
## 1      front      88.6      168.8      64.1      48.8      2548
dohc
## 2      front      88.6      168.8      64.1      48.8      2548
dohc
## 3      front      94.5      171.2      65.5      52.4      2823
ohcv
## 4      front      99.8      176.6      66.2      54.3      2337
ohc
## 5      front      99.4      176.6      66.4      54.3      2824
ohc
## 6      front      99.8      177.3      66.3      53.1      2507
ohc
##      cylindernumber enginesize stroke compressionratio horsepower peakrpm
citympg
## 1      four          130    2.68              9.0          111      5000
21
## 2      four          130    2.68              9.0          111      5000
21
## 3      six           152    3.47              9.0          154      5000
19
## 4      four          109    3.40             10.0          102      5500
24
## 5      five          136    3.40              8.0          115      5500
18
## 6      five          136    3.40              8.5          110      5500
19
##      highwaympg price
## 1          27 13495
## 2          27 16500
## 3          26 16500
## 4          30 13950
## 5          22 17450
## 6          25 15250

```

Calcula medidas estadísticas apropiadas para las variables: cuantitativas (media, desviación estándar, cuantiles, etc) cualitativas: cuantiles, frecuencias (puedes usar el comando table o prop.table)

```

ensinesize = M$enginesize
stroke = M$stroke
engineLocalization = M$engineLocation

summary(ensinesize)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      61.0   97.0   120.0   126.9   141.0   326.0

summary(stroke)

```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.070   3.110   3.290   3.255   3.410   4.170
```

```
table(engineLocalization)
```

```
## engineLocalization
## front  rear
##    202    3
```

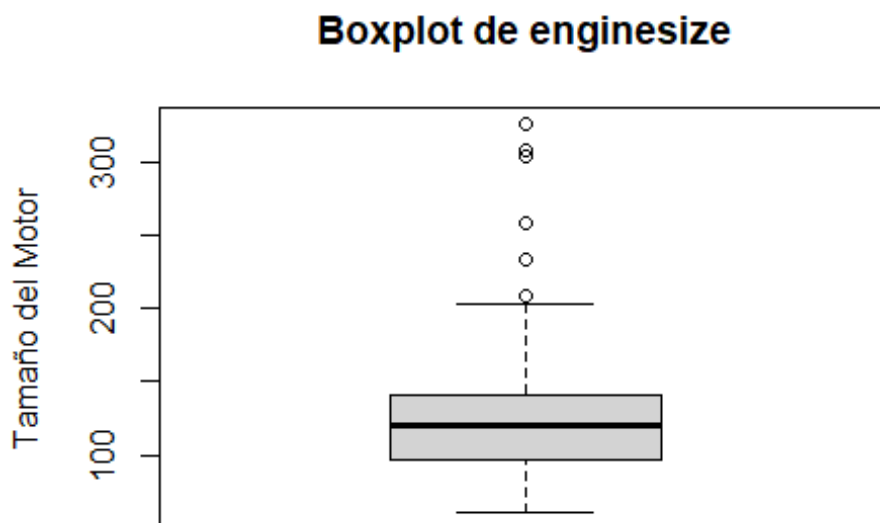
Analiza la correlación entre las variables (analiza posible colinealidad entre las variables)

```
cor(M[, c("enginesize", "stroke", "price")], use = "complete.obs")
```

```
##           enginesize      stroke      price
## enginesize  1.0000000  0.20312859  0.87414480
## stroke      0.2031286  1.00000000  0.07944308
## price      0.8741448  0.07944308  1.00000000
```

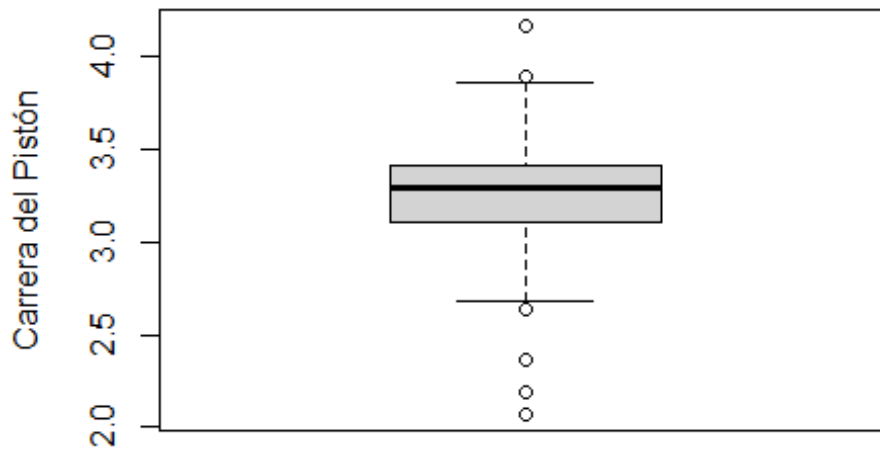
Explora los datos usando herramientas de visualización (si lo consideras necesario):
 Variables cuantitativas: Boxplot (visualización de datos atípicos) Histogramas
 Diagramas de dispersión y correlación por pares

```
# Boxplot de enginesize y stroke
boxplot(M$enginesize, main="Boxplot de enginesize", ylab="Tamaño del Motor")
```



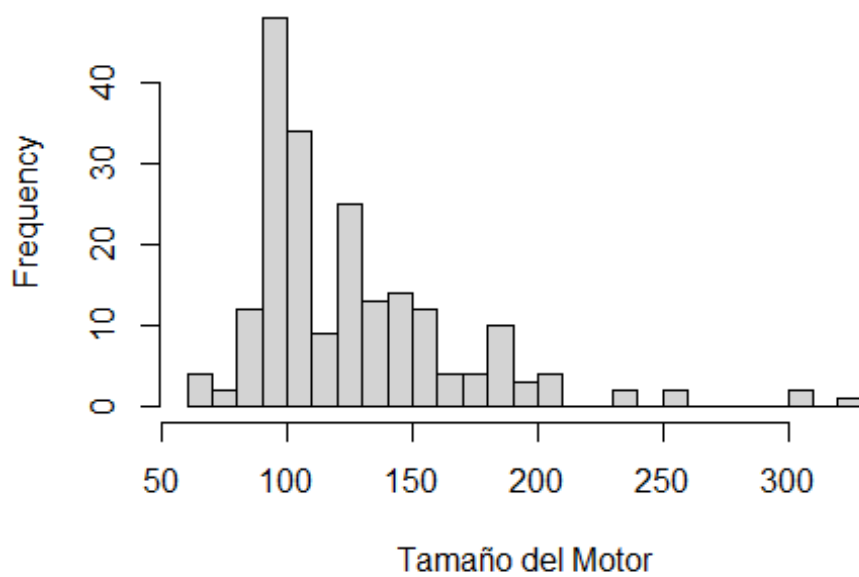
```
boxplot(M$stroke, main="Boxplot de stroke", ylab="Carrera del Pistón")
```

Boxplot de stroke

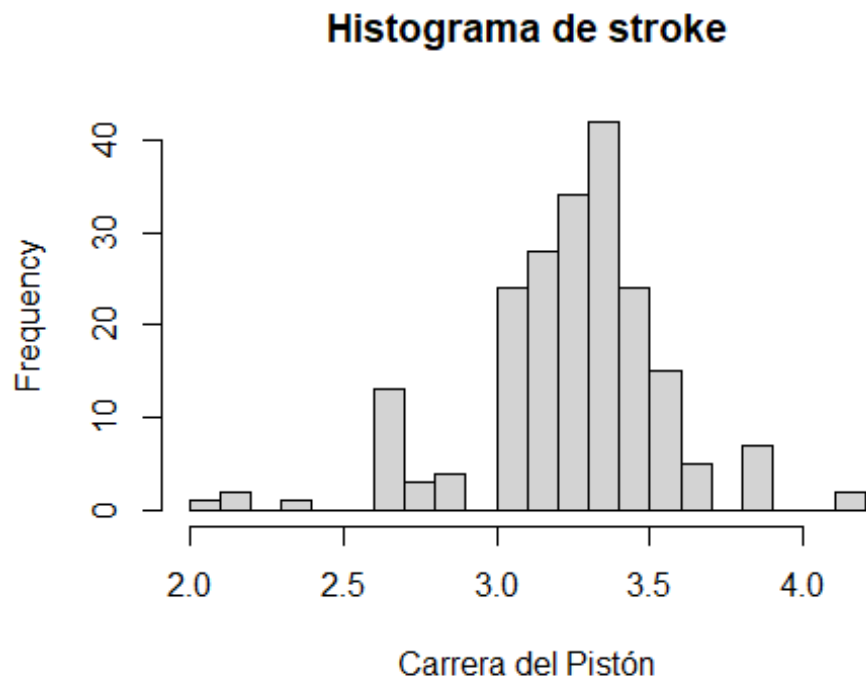


```
# Histogramas para enginesize y stroke  
hist(M$enginesize, main="Histograma de enginesize", xlab="Tamaño del  
Motor", breaks=20)
```

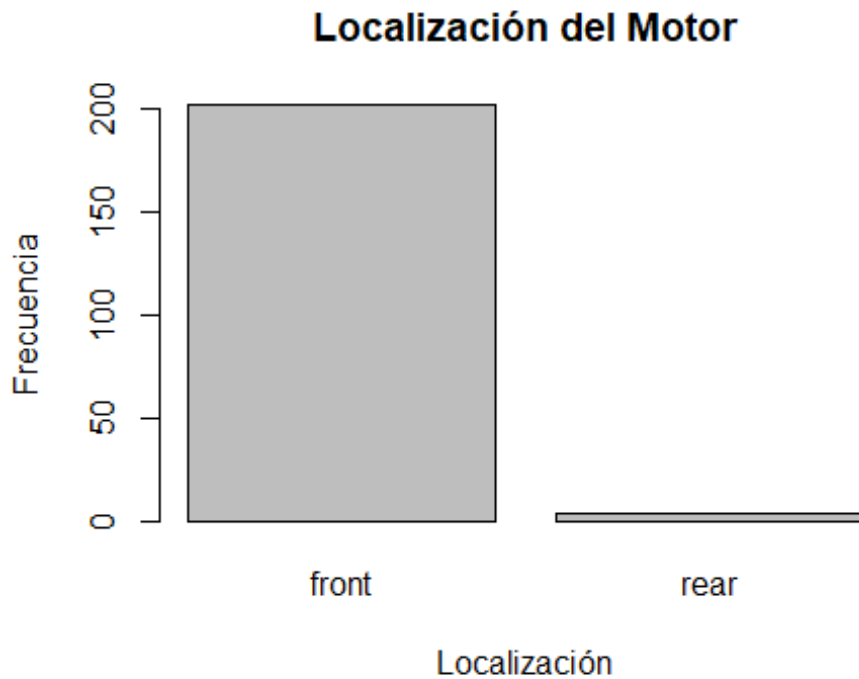
Histograma de enginesize



```
hist(M$stroke, main="Histograma de stroke", xlab="Carrera del Pistón",  
breaks=20)
```



```
# Gráfico de barras para engineLocation  
barplot(table(M$engineLocation), main="Localización del Motor",  
xlab="Localización", ylab="Frecuencia")
```



Problema 2

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.

```
modelo1 <- lm(price ~ enginesize + stroke + enginelocation, data = M)
summary(modelo1)
```

```
##
## Call:
## lm(formula = price ~ enginesize + stroke + enginelocation, data = M)
##
## Residuals:
```

| | Min | 1Q | Median | 3Q | Max |
|--|----------|---------|--------|--------|---------|
| | -10967.9 | -1961.3 | -556.2 | 1576.1 | 14879.5 |

```
##
## Coefficients:
```

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|-----------|------------|---------|--------------|
| (Intercept) | -1575.699 | 2703.795 | -0.583 | 0.5607 |
| enginesize | 165.155 | 6.453 | 25.594 | < 2e-16 *** |
| stroke | -1919.247 | 848.279 | -2.263 | 0.0247 * |
| enginelocationrear | 9629.355 | 2206.969 | 4.363 | 2.05e-05 *** |

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3655 on 201 degrees of freedom
```

```
## Multiple R-squared:  0.7937, Adjusted R-squared:  0.7906
## F-statistic: 257.8 on 3 and 201 DF,  p-value: < 2.2e-16

Modelo2 <- lm(price ~ enginesize * stroke + enginelocation, data = M)
summary(Modelo2)

##
## Call:
## lm(formula = price ~ enginesize * stroke + enginelocation, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8423.1 -2066.1  -467.3   1288.4  14784.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10254.18    8155.59   1.257   0.210
## enginesize         77.54     57.37   1.352   0.178
## stroke        -5537.23    2501.37  -2.214   0.028 *
## enginelocationrear 10327.46    2245.95   4.598 7.55e-06 ***
## enginesize:stroke    26.59     17.30   1.537   0.126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3643 on 200 degrees of freedom
## Multiple R-squared:  0.7961, Adjusted R-squared:  0.792
## F-statistic: 195.2 on 4 and 200 DF,  p-value: < 2.2e-16
```

Para cada uno de los modelos propuestos:

Realiza la regresión entre las variables involucradas 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 β_i con un alfa de 0.04 (incluye las hipótesis que pruebas y el valor frontera de cada una de ellas)

A 0.04 sí es significativo y los modelos quedarían

Modelo 1 $H_0: \beta_1 = 0$ \$H_{1: _1}\$

Modelo 2 $H_0: \beta_1 = 0$ \$H_{1: _1}\$

```
summary(modelo1)$r.squared
## [1] 0.7937094
summary(Modelo2)$r.squared
## [1] 0.7961171
anova(modelo1)
## Analysis of Variance Table
##
```

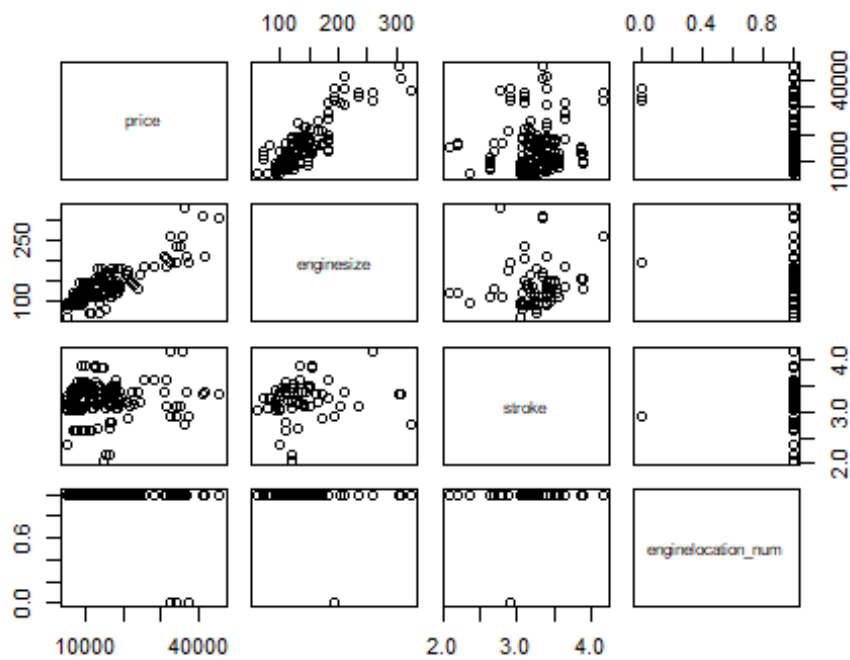
```
## Response: price
##           Df      Sum Sq    Mean Sq  F value    Pr(>F)
## enginesize 1 9948685774 9948685774 744.5320 < 2.2e-16 ***
## stroke     1 130743490 130743490 9.7845 0.002021 **
## enginelocation 1 254380772 254380772 19.0371 2.048e-05 ***
## Residuals 201 2685829326 13362335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Modelo2)
```

```
## Analysis of Variance Table
##
## Response: price
##           Df      Sum Sq    Mean Sq  F value    Pr(>F)
## enginesize 1 9948685774 9948685774 749.5765 < 2.2e-16 ***
## stroke     1 130743490 130743490 9.8508 0.001954 **
## enginelocation 1 254380772 254380772 19.1661 1.93e-05 ***
## enginesize:stroke 1 31347673 31347673 2.3619 0.125914
## Residuals 200 2654481653 13272408
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Valida la significancia del modelo con un alfa de 0.04 (incluye las hipótesis que pruebas y el valor frontera) Valida la significancia de β_i 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.

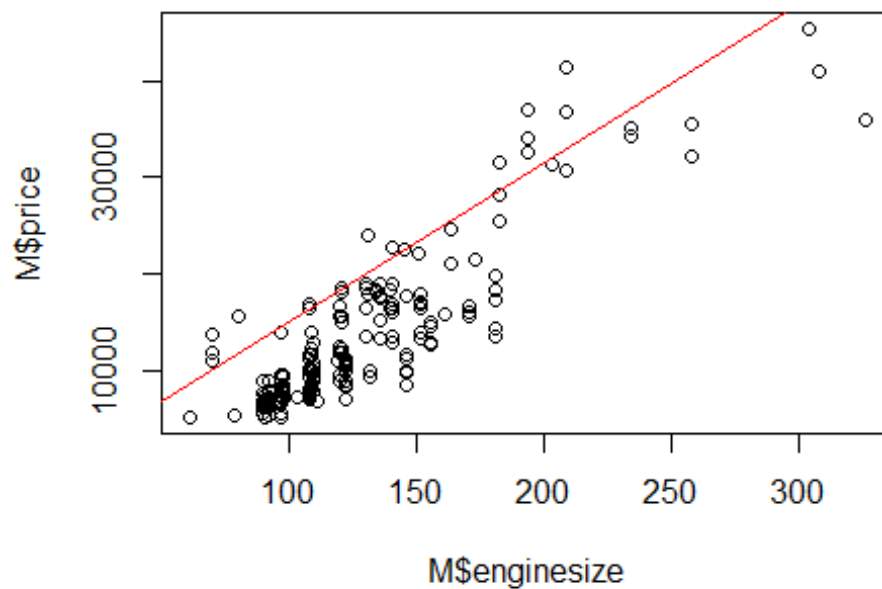
```
# Gráficos de dispersión y recta de ajuste
M$enginelocation_num <- ifelse(M$enginelocation == "front", 1, 0)
pairs(M[,c("price", "enginesize", "stroke", "enginelocation_num")])
```

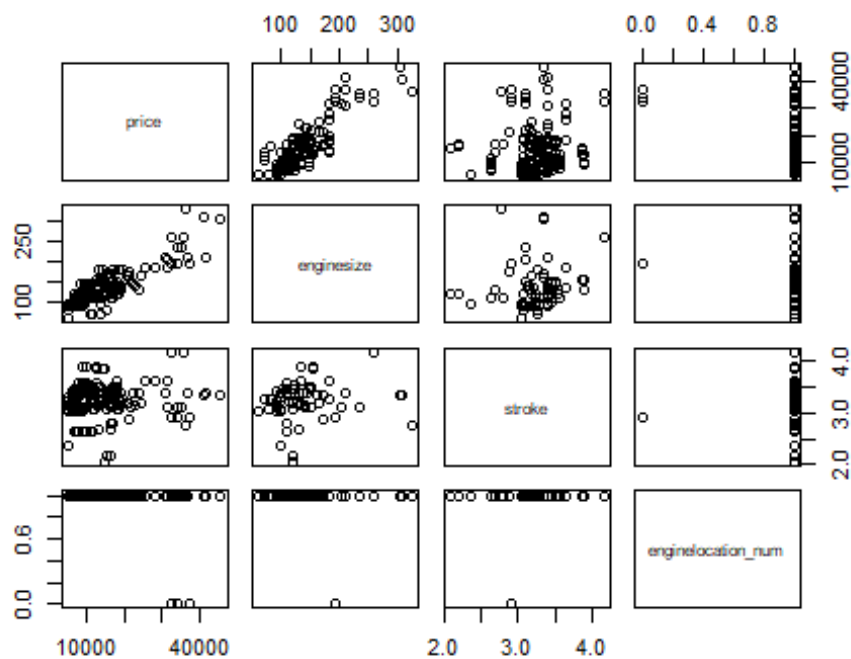
```
# Gráfico para price y enginesize
plot(M$enginesize, M$price, main="Dispersión de Price vs Engine Size
(Modelo 1)")
abline(modelo1, col="red")

## Warning in abline(modelo1, col = "red"): only using the first two of 4
## regression coefficients
```

Dispersión de Price vs Engine Size (Modelo 1)



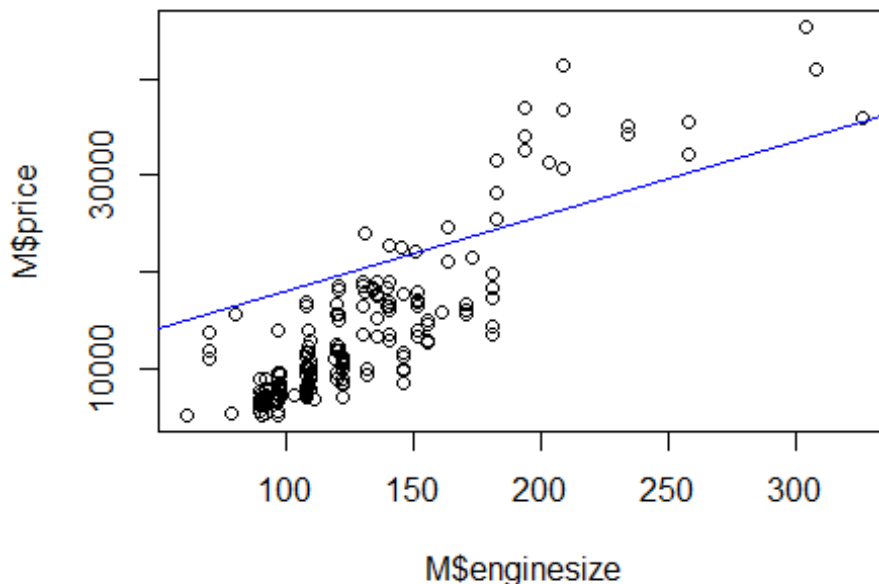
```
# Gráficos de dispersión y recta de ajuste
pairs(M[,c("price", "enginesize", "stroke", "enginelocation_num")])
```



```
# Gráfico para price y enginesize
plot(M$enginesize, M$price, main="Dispersión de Price vs Engine Size
(Modelo 2)")
abline(Modelo2, col="blue")

## Warning in abline(Modelo2, col = "blue"): only using the first two of
5
## regression coefficients
```

Dispersión de Price vs Engine Size (Modelo 2)



Interpreta en el contexto del problema cada uno de los análisis que hiciste.

Observando los dos modelos que se realizaron se tomó la decisión de usar el modelo 2 ya que este tiene un mejor ajuste en la interacción entre enginesize y precio ya que el modelo 1 enginesize y precio depende del stroke y el modelo 2 tiene una mejor comprensión ya que el enginesize y el stroke ambos influyen directamente en price, esto ayudando a tener una mejor validación en el modelo 2.

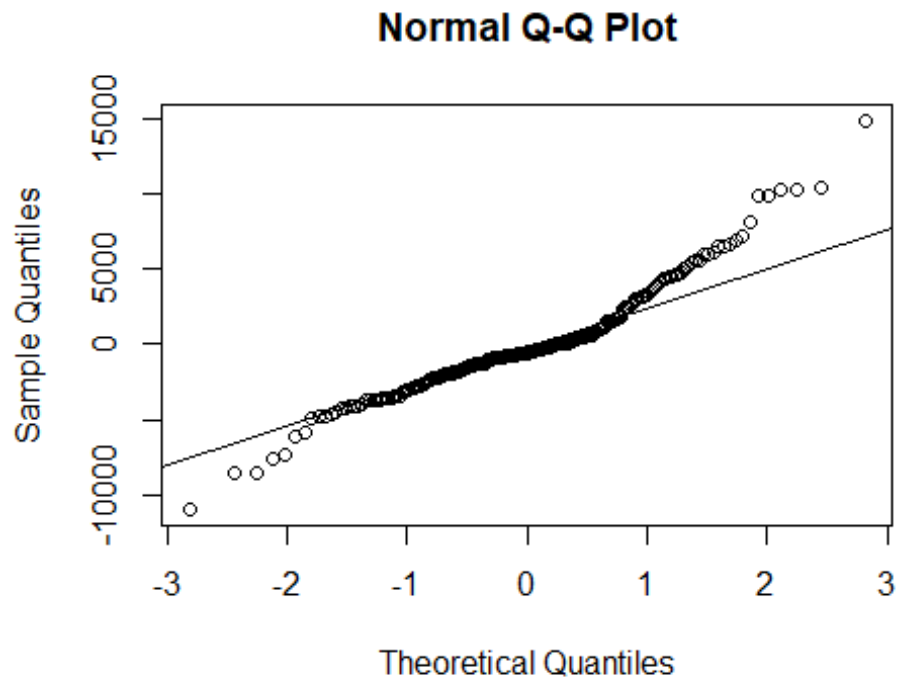
Analiza la validez de los modelos propuestos: Normalidad de los residuos

H0: Los datos provienen de una población normal H1: Los datos no provienen de una población normal

```
library(nortest)

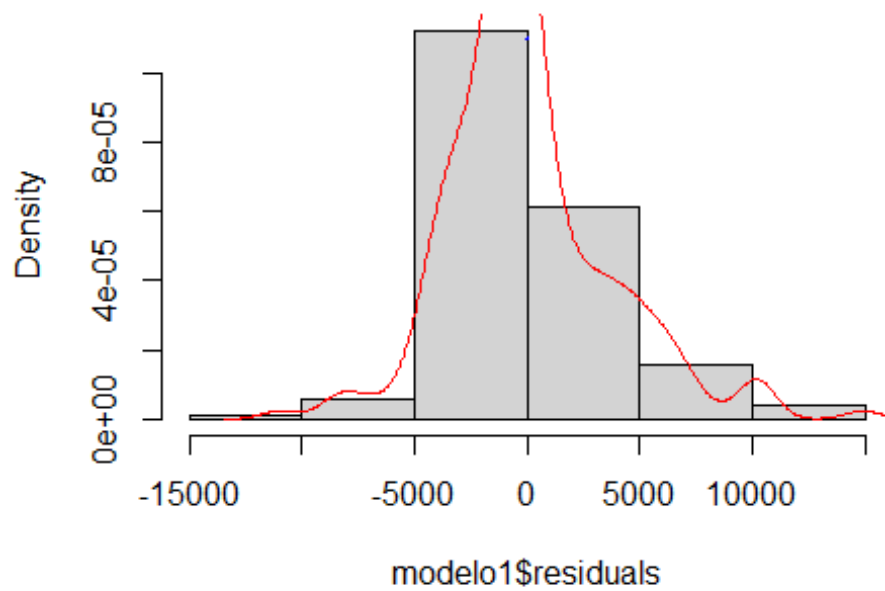
# Modelo 1
ad.test(modelo1$residuals)
```

```
##  
## Anderson-Darling normality test  
##  
## data: modelo1$residuals  
## A = 3.9675, p-value = 6.552e-10  
  
qqnorm(modelo1$residuals)  
qqline(modelo1$residuals)
```



```
hist(modelo1$residuals,freq=FALSE)  
lines(density(modelo1$residual),col="red")  
curve(dnorm(x,mean=mean(modelo1$residuals),sd=sd(modelo1$residuals)),  
from=-  
21, to=21, add=TRUE, col="blue",lwd=2)
```

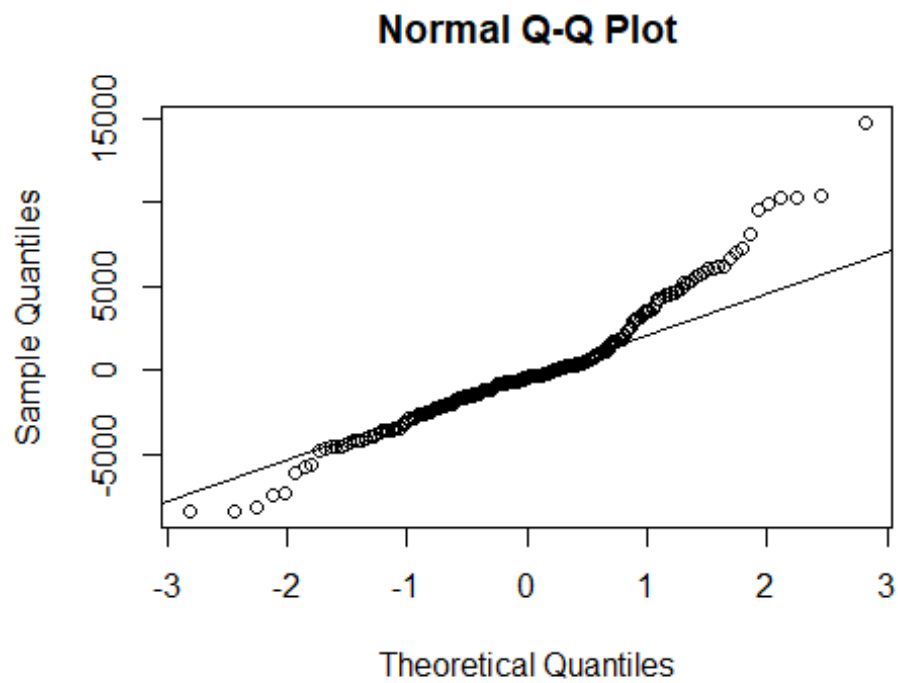
Histogram of modelo1\$residuals



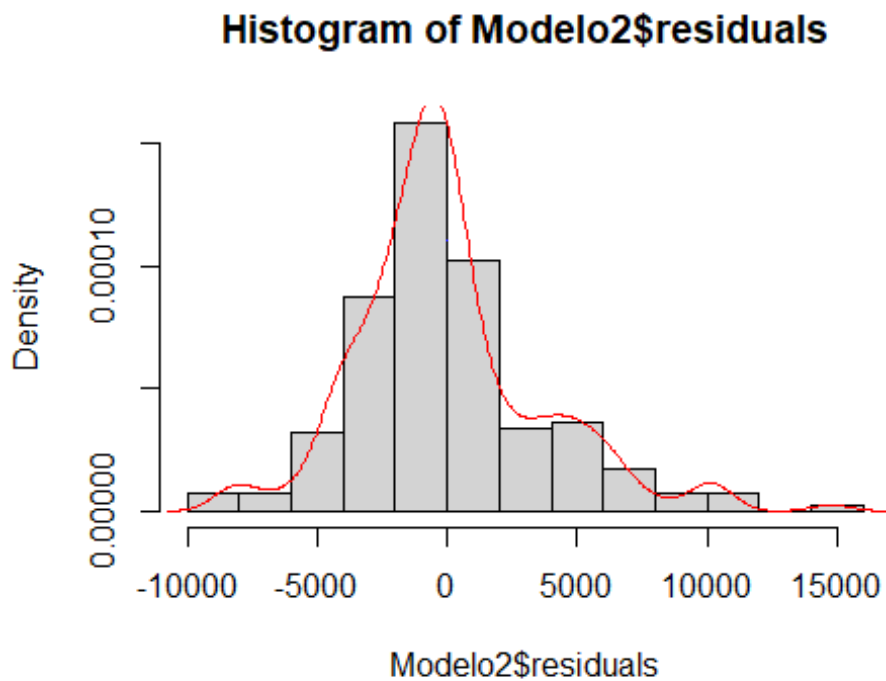
```
# Modelo 2
ad.test(Modelo2$residuals)

##
## Anderson-Darling normality test
##
## data:  Modelo2$residuals
## A = 3.7327, p-value = 2.432e-09

qqnorm(Modelo2$residuals)
qqline(Modelo2$residuals)
```



```
hist(Modelo2$residuals,freq=FALSE)
lines(density(Modelo2$residual),col="red")
curve(dnorm(x,mean=mean(Modelo2$residuals),sd=sd(Modelo2$residuals)),
from=-
21, to=21, add=TRUE, col="blue",lwd=2)
```



Modelo 1 H0

no se rechaza ya que el valorp es mayor que 0.04

Modelo 2 H0 se rechaza ya que el valorp es menor que 0.04

Verificación de media cero

H0: $\mu = 0$ H1: $\mu \neq 0$

```
t.test(modelo1$residuals)

##
## One Sample t-test
##
## data:  modelo1$residuals
## t = -1.6888e-15, df = 204, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -499.6656  499.6656
## sample estimates:
## mean of x
## -4.279807e-13

t.test(Modelo2$residuals)

##
## One Sample t-test
##
## data:  Modelo2$residuals
## t = 6.439e-16, df = 204, p-value = 1
```

```
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -496.7411 496.7411
## sample estimates:
## mean of x
## 1.622247e-13
```

Modelo 1 Se no rechaza H0 porque es mayor a 0.04

Modelo 2 Se no rechaza H0 porque es mayor a 0.04

Homocedasticidad, linealidad e independencia

Homocedasticidad H0: La varianza de los errores es constante (homocedasticidad)

H1: La varianza de los errores no es constante (heterocedasticidad)

Linealidad H0: No hay términos omitidos que indique linealidad H1: Hay una especificación errónea en el modelo que indica no linealidad

Independencia H0: Los errores no están correlacionados H1: Los errores están correlacionados

```
plot(modelo1$fitted.values,modelo1$residuals)
abline(h=0, col="red")
```

```
library(lmtest)
```

```
## Cargando paquete requerido: zoo
```

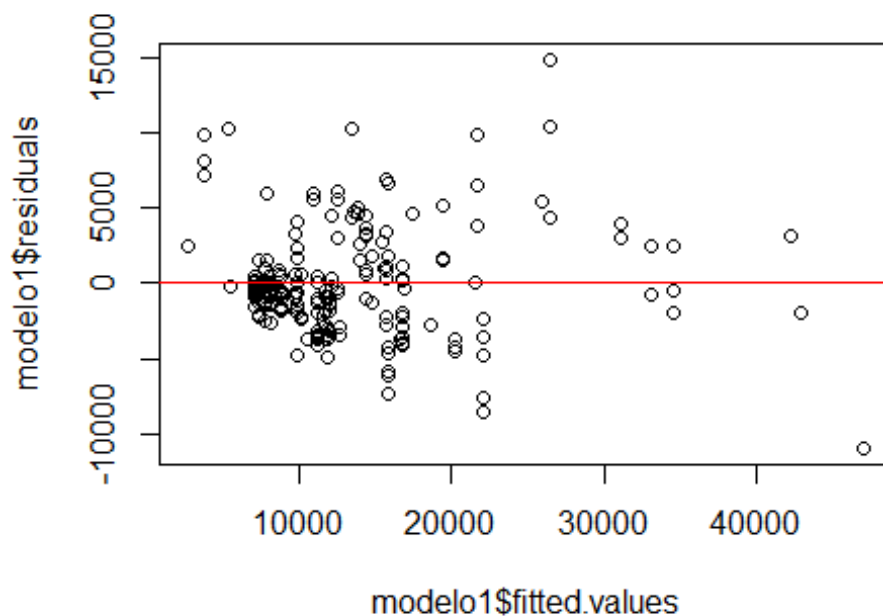
```
##
```

```
## Adjuntando el paquete: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## as.Date, as.Date.numeric
```

```
dwtest(modelo1)

##
##  Durbin-Watson test
##
## data:  modelo1
## DW = 0.82453, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0

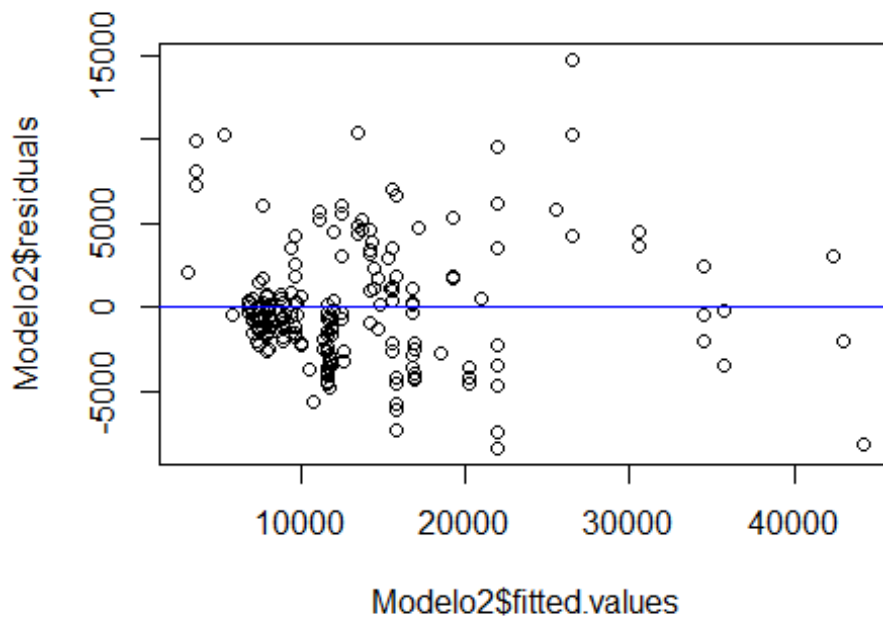
bptest(modelo1)

##
##  studentized Breusch-Pagan test
##
## data:  modelo1
## BP = 24.786, df = 3, p-value = 1.712e-05

resettest(modelo1)

##
##  RESET test
##
## data:  modelo1
## RESET = 14.968, df1 = 2, df2 = 199, p-value = 8.798e-07

plot(Modelo2$fitted.values, Modelo2$residuals)
abline(h=0, col="blue")
```



```
library(lmtest)
dwtest(Modelo2)

##
## Durbin-Watson test
##
## data: Modelo2
## DW = 0.78614, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0

bptest(Modelo2)

##
## studentized Breusch-Pagan test
##
## data: Modelo2
## BP = 20.208, df = 4, p-value = 0.0004543

resettest(Modelo2)

##
## RESET test
##
## data: Modelo2
## RESET = 12.672, df1 = 2, df2 = 198, p-value = 6.629e-06
```

Modelo 1

Homocedasticidad H0 se rechaza porque el valor p es < 0.04

Independencia H0 se rechaza porque el valor p es < 0.04

Linealidad H0 se rechaza porque el valor p es < 0.04

Modelo 2

Homocedasticidad H0 se rechaza porque el valor p es < 0.04

Independencia H0 se rechaza porque el valor p es < 0.04

Linealidad H0 se rechaza porque el valor p es < 0.04

Interpreta cada uno de los análisis que realizaste. Emite una conclusión final sobre el mejor modelo de regresión lineal y contesta la pregunta central: Concluye sobre el mejor modelo que encuentres y argumenta por qué es el mejor.

El modelo2 fue el mejor modelo ya que este tiene una mejor interacción entre el enginesize y stroke haciendo que este modelo sea más preciso con el precio del coche.

¿Cuáles de las variables asignadas influyen en el precio del auto? ¿de qué manera lo hacen? La variable de enginesize afecta en la forma del motor del coche si este es más grande este tiene un mayor precio. La otra variable que afecta el precio es stroke esta influye dependiendo del volumen adentro del motor.

Y la variable que no afecta al precio es la enginelocation esta solo indica donde está ubicado el motor si enfrente o atrás del coche.

Problema 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 justifiques la categoría seleccionada anteriormente.

```
# Calcular intervalos de predicción
Ip = predict(object = Modelo2, interval = "prediction", level = 0.97)

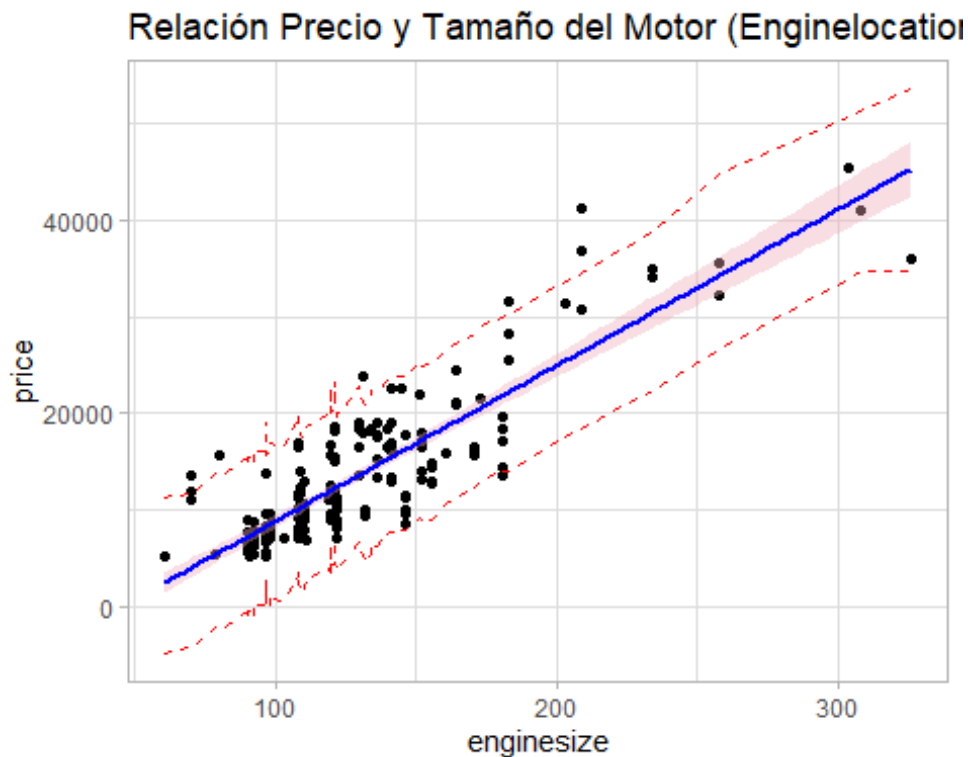
## Warning in predict.lm(object = Modelo2, interval = "prediction", level
= 0.97): predictions on current data refer to _future_ responses

M2 = cbind(M, Ip)

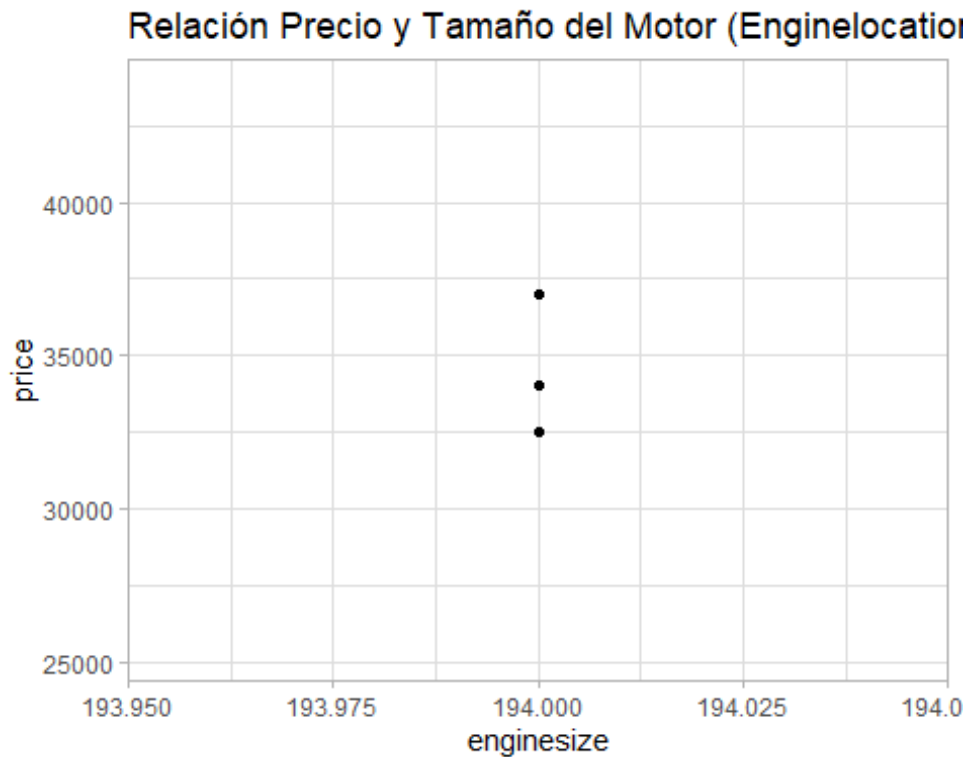
M2_front = subset(M2, enginelocation == "front")
M2_rear = subset(M2, enginelocation == "rear")
```

```
library(ggplot2)

ggplot(M2_front, aes(x = enginesize, y = price)) +
  ggtitle("Relación Precio y Tamaño del Motor (Enginelocation: Front)") +
  geom_point() +
  geom_line(aes(y = lwr), color = "red", linetype = "dashed") +
  geom_line(aes(y = upr), color = "red", linetype = "dashed") +
  geom_smooth(method = lm, formula = y ~ x, se = TRUE, level = 0.97, col
= "blue", fill = "pink2") +
  theme_light()
```



```
ggplot(M2_rear, aes(x = enginesize, y = price)) +
  ggtitle("Relación Precio y Tamaño del Motor (Enginelocation: Rear)") +
  geom_point() +
  geom_line(aes(y = lwr), color = "red", linetype = "dashed") +
  geom_line(aes(y = upr), color = "red", linetype = "dashed") +
  geom_smooth(method = lm, formula = y ~ x, se = TRUE, level = 0.97, col
= "blue", fill = "pink2") +
  theme_light()
```



Interpreta en el contexto del problema

En las graficas se observa que el tamaño del motor si afecta al precio en la primera se nota que dependiendo de la ubicacion esta siendo enfrente del coche dependiendo del tamaño si afecta al precio y en la segunda grafica no se observa mucho ya que la mayoría de los coches tienen el motor enfrenten

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

```
M1=data.frame(M$symboling, M$wheelbase, M$carlength, M$carlength,
M$carwidth, M$carheight, M$curbweight, M$enginesize, M$stroke,
M$compressionratio, M$horsepower, M$peakrpm, M$highwaympg, M$price)
#summary(M1)
cor(M1)
```

| | M.symboling | M.wheelbase | M.carlength | M.carlength.1 |
|------------------|--------------|-------------|-------------|---------------|
| ## | | | | |
| ## M.symboling | 1.000000000 | -0.5319537 | -0.3576115 | -0.3576115 |
| ## M.wheelbase | -0.531953682 | 1.0000000 | 0.8745875 | 0.8745875 |
| ## M.carlength | -0.357611523 | 0.8745875 | 1.0000000 | 1.0000000 |
| ## M.carlength.1 | -0.357611523 | 0.8745875 | 1.0000000 | 1.0000000 |
| ## M.carwidth | -0.232919061 | 0.7951436 | 0.8411183 | 0.8411183 |
| ## M.carheight | -0.541038200 | 0.5894348 | 0.4910295 | 0.4910295 |
| ## M.curbweight | -0.227690588 | 0.7763863 | 0.8777285 | 0.8777285 |

| | | | | |
|-----------------------|--------------|--------------------|--------------|--------------|
| ## M.enginesize | -0.105789709 | 0.5693287 | 0.6833599 | 0.6833599 |
| ## M.stroke | -0.008735141 | 0.1609590 | 0.1295326 | 0.1295326 |
| ## M.compressionratio | -0.178515084 | 0.2497858 | 0.1584137 | 0.1584137 |
| ## M.horsepower | 0.070872724 | 0.3532945 | 0.5526230 | 0.5526230 |
| ## M.peakrpm | 0.273606245 | -0.3604687 | -0.2872422 | -0.2872422 |
| ## M.highwaympg | 0.034606001 | -0.5440819 | -0.7046616 | -0.7046616 |
| ## M.price | -0.079978225 | 0.5778156 | 0.6829200 | 0.6829200 |
| ## | M.carwidth | M.carheight | M.curbweight | M.enginesize |
| ## M.symboling | -0.2329191 | -0.54103820 | -0.2276906 | -0.10578971 |
| ## M.wheelbase | 0.7951436 | 0.58943476 | 0.7763863 | 0.56932868 |
| ## M.carlength | 0.8411183 | 0.49102946 | 0.8777285 | 0.68335987 |
| ## M.carlength.1 | 0.8411183 | 0.49102946 | 0.8777285 | 0.68335987 |
| ## M.carwidth | 1.0000000 | 0.27921032 | 0.8670325 | 0.73543340 |
| ## M.carheight | 0.2792103 | 1.00000000 | 0.2955717 | 0.06714874 |
| ## M.curbweight | 0.8670325 | 0.29557173 | 1.0000000 | 0.85059407 |
| ## M.enginesize | 0.7354334 | 0.06714874 | 0.8505941 | 1.00000000 |
| ## M.stroke | 0.1829417 | -0.05530667 | 0.1687900 | 0.20312859 |
| ## M.compressionratio | 0.1811286 | 0.26121423 | 0.1513617 | 0.02897136 |
| ## M.horsepower | 0.6407321 | -0.10880206 | 0.7507393 | 0.80976865 |
| ## M.peakrpm | -0.2200123 | -0.32041072 | -0.2662432 | -0.24465983 |
| ## M.highwaympg | -0.6772179 | -0.10735763 | -0.7974648 | -0.67746991 |
| ## M.price | 0.7593253 | 0.11933623 | 0.8353049 | 0.87414480 |
| ## | M.stroke | M.compressionratio | M.horsepower | |
| M.peakrpm | | | | |
| ## M.symboling | -0.008735141 | -0.17851508 | 0.07087272 | |
| 0.27360625 | | | | |
| ## M.wheelbase | 0.160959047 | 0.24978585 | 0.35329448 | - |
| 0.36046875 | | | | |
| ## M.carlength | 0.129532611 | 0.15841371 | 0.55262297 | - |
| 0.28724220 | | | | |
| ## M.carlength.1 | 0.129532611 | 0.15841371 | 0.55262297 | - |
| 0.28724220 | | | | |
| ## M.carwidth | 0.182941693 | 0.18112863 | 0.64073208 | - |
| 0.22001230 | | | | |
| ## M.carheight | -0.055306674 | 0.26121423 | -0.10880206 | - |
| 0.32041072 | | | | |
| ## M.curbweight | 0.168790035 | 0.15136174 | 0.75073925 | - |
| 0.26624318 | | | | |
| ## M.enginesize | 0.203128588 | 0.02897136 | 0.80976865 | - |
| 0.24465983 | | | | |
| ## M.stroke | 1.000000000 | 0.18611011 | 0.08093954 | - |
| 0.06796375 | | | | |
| ## M.compressionratio | 0.186110110 | 1.00000000 | -0.20432623 | - |
| 0.43574051 | | | | |
| ## M.horsepower | 0.080939536 | -0.20432623 | 1.00000000 | |
| 0.13107251 | | | | |
| ## M.peakrpm | -0.067963753 | -0.43574051 | 0.13107251 | |
| 1.00000000 | | | | |
| ## M.highwaympg | -0.043930930 | 0.26520139 | -0.77054389 | - |
| 0.05427481 | | | | |

| | | | | |
|-----------------------|--------------|-------------|------------|---|
| ## M.price | 0.079443084 | 0.06798351 | 0.80813882 | - |
| 0.08526715 | | | | |
| ## | M.highwaympg | M.price | | |
| ## M.symboling | 0.03460600 | -0.07997822 | | |
| ## M.wheelbase | -0.54408192 | 0.57781560 | | |
| ## M.carlength | -0.70466160 | 0.68292002 | | |
| ## M.carlength.1 | -0.70466160 | 0.68292002 | | |
| ## M.carwidth | -0.67721792 | 0.75932530 | | |
| ## M.carheight | -0.10735763 | 0.11933623 | | |
| ## M.curbweight | -0.79746479 | 0.83530488 | | |
| ## M.enginesize | -0.67746991 | 0.87414480 | | |
| ## M.stroke | -0.04393093 | 0.07944308 | | |
| ## M.compressionratio | 0.26520139 | 0.06798351 | | |
| ## M.horsepower | -0.77054389 | 0.80813882 | | |
| ## M.peakrpm | -0.05427481 | -0.08526715 | | |
| ## M.highwaympg | 1.00000000 | -0.69759909 | | |
| ## M.price | -0.69759909 | 1.00000000 | | |

Propondria que se usaran las variables que más afectan el precio del coche como el enginesize, curbweight y horsepower