Regresión lineal simple y múltiple

Análisis Exploratorio de Datos. Máster en Ciencia de Datos - UV

Carlos Blom-Dahl Ramos y Daniel Lillo Plaza

14-11-2022

Índice

Índice de figuras

Índice de cuadros

1. Librerías cargadas

Las librerías empleadas para la correcta ejecución del código las encontramos a continuación: `packages = c("kableExtra","tidyverse","knitr", "ggplot2", "car")`

```
 [1] ".GlobalEnv"         "package:car"        "package:carData"    "package:forcats"
 [5] "package:stringr"    "package:dplyr"      "package:purrr"      "package:readr"
 [9] "package:tidyr"      "package:tibble"     "package:ggplot2"    "package:tidyverse"
[13] "package:kableExtra" "package:knitr"      "package:stats"      "package:graphics"
[17] "package:grDevices"  "package:utils"      "package:datasets"   "package:methods"
[21] "Autoloads"          "package:base"
```

2. Introducción.

En la librería MASS puedes encontrar un famoso banco de datos llamado `Cars93` que recoge información sobre 93 coches en venta en los Estados Unidos en 1993. La base contiene 27 variables relativas a 93 coches. Para saber qué información está contenida en las variables puedes escribir: *?Cars93*. En este estudio vamos a usar la base de datos **cars.csv** que encontraréis en el aula virtual (contenido en la carpeta de este proyecto también). Dicha base de datos contiene 15 de las 27 variables del banco de datos `Cars93` indicado anteriormente.

Procedemos a leer el fichero de datos:

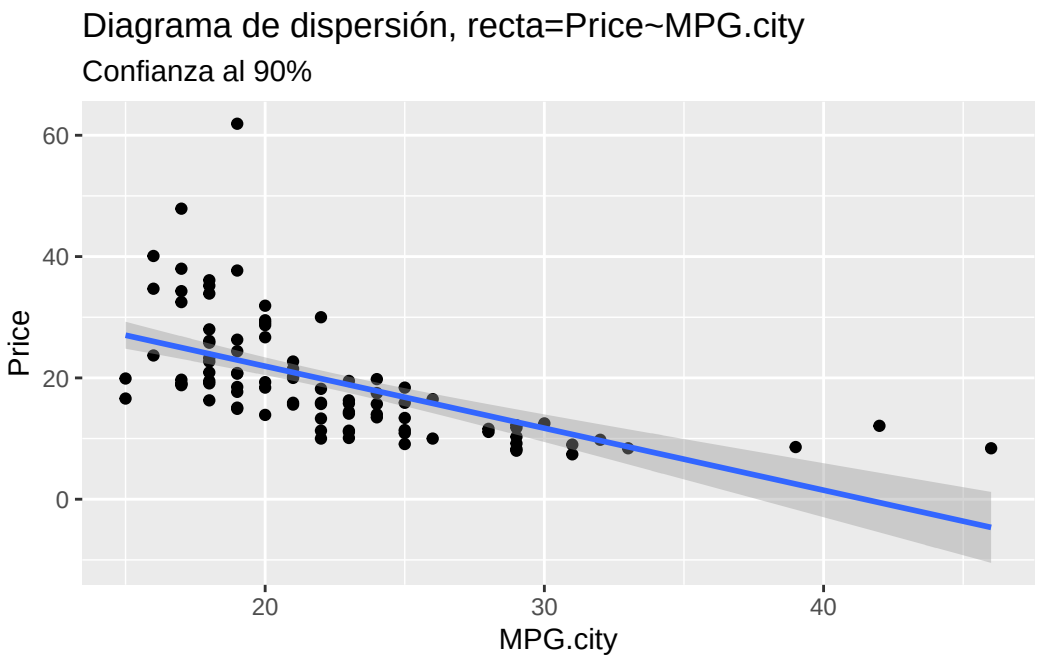
```
cars <- read.csv2("./data/cars.csv")
```

3. Ejercicio 1.

Considera la variable respuesta `Price` relacionándola con el predictor `MPG.city`.

1. Evalúa el efecto de `MPG.city` sobre `Price`.
2. Obtén la recta de mínimos cuadrados. Interpreta los resultados obtenidos.
3. Dibuja el diagrama de dispersión, la recta de regresión y las bandas de confianza al 90 %.
4. ¿Te parece adecuado haber realizado regresión lineal o es preferible otro tipo de regresión?. Ajusta el modelo que te parezca más adecuado y, en dicho caso, dibuja nuevamente las bandas de confianza correspondientes al 90 %.
5. ¿Qué precio mínimo se espera para aquellos coches con un consumo de 12 litros a los 100 km por ciudad? Calcula e interpreta el intervalo de confianza y el de predicción.

```
ggplot(cars, mapping=aes(x=MPG.city, y=Price))+geom_point()+
geom_smooth(method = "lm", se=TRUE, level=0.9)+
ggtitle("Diagrama de dispersión, recta=Price~MPG.city", subtitle="Confianza al 90%")
```

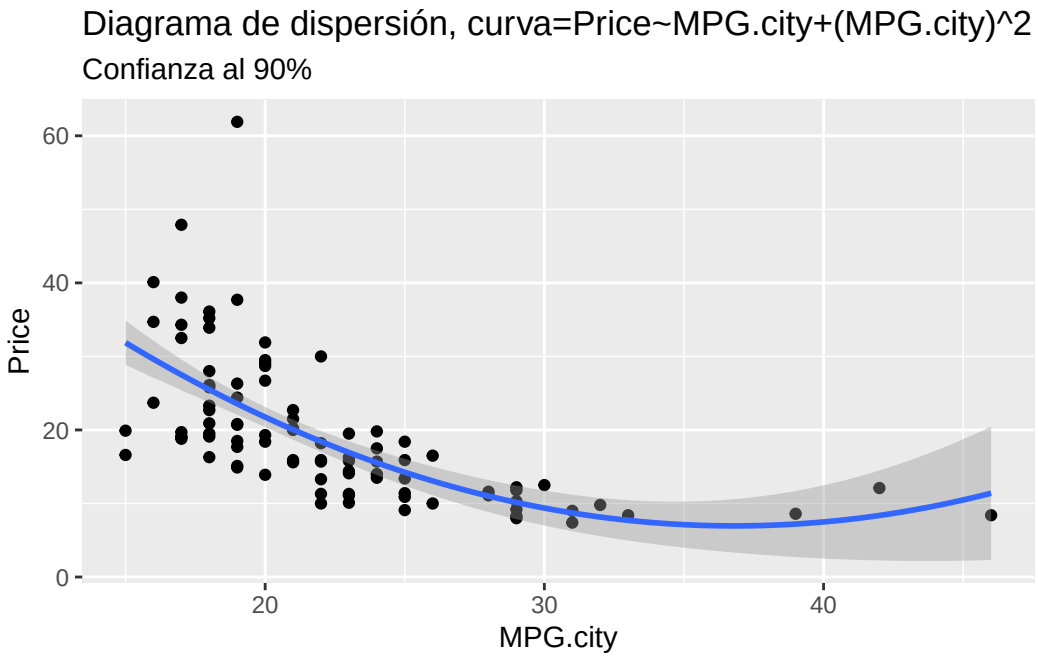



```
cor(cars$MPG.city, cars$Price)
```

```
[1] -0.5945622
```

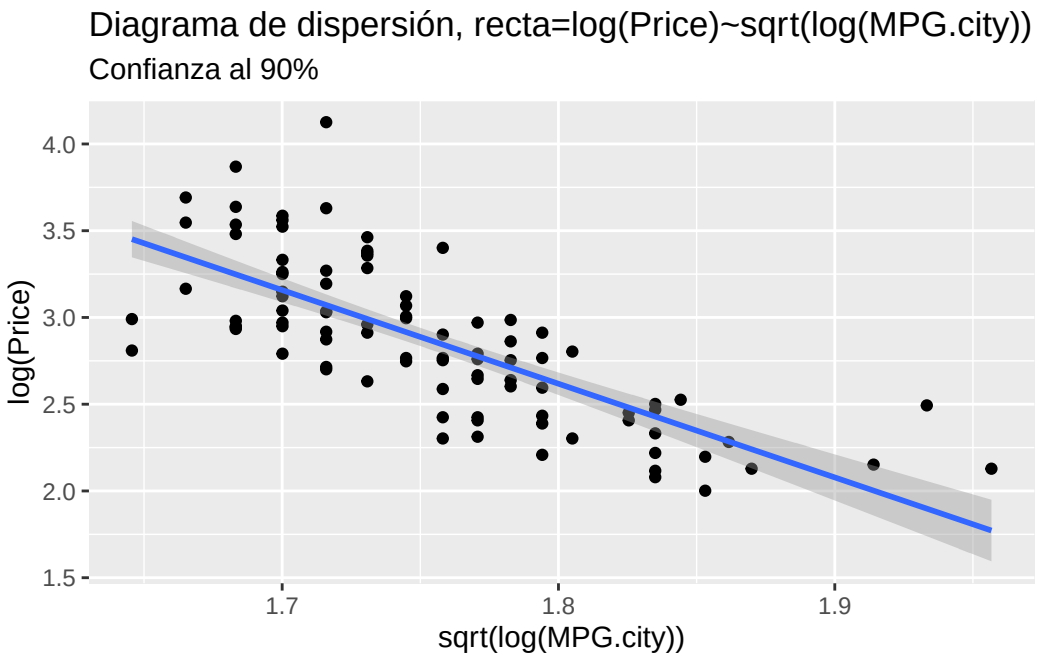
No parece haber mucha relación **LINEAL** entre las variables, pero **sí cuadrática**. Lo comprobamos en el siguiente gráfico.

```
ggplot(cars, mapping=aes(x=MPG.city, y=Price))+geom_point()+
geom_smooth(method = "lm", se=TRUE, level=0.9, formula = y ~ poly(x, 2))+
ggtitle("Diagrama de dispersión, curva=Price~MPG.city+(MPG.city)^2",
subtitle="Confianza al 90%")
```



Para intentar aumentar la relación lineal entre las variables, vamos a realizar una transformación logarítmica de ambas (para no tener problemas de heterocedasticidad) y aplicaremos una transformación de raíz cuadrada para corregir la relación cuadrática expuesta anteriormente.

```
ggplot(cars, mapping=aes(x=sqrt(log(MPG.city)), y=log(Price)))+geom_point()+
geom_smooth(method = "lm", se=TRUE, level=0.9, formula=y~x)+
ggtitle("Diagrama de dispersión, recta=log(Price)~sqrt(log(MPG.city))",
subtitle="Confianza al 90%")
```




```
cor((log(cars$MPG.city))^(1/2), log(cars$Price))
```

```
[1] -0.7469256
```

Vemos como el grado de correlación lineal aumenta (en módulo) sustancialmente: pasamos de -0.5945622 a -0.7469256.

Si quisiéramos mejorar el coeficiente de correlación, ya que sabemos que la relación entre estas dos variables es cuadrática, para corregir más aún la heterocedasticidad podemos estudiar la relación entre *(log(cars$MPG.city))^(1/32)* y *(log(cars$Price))^(1/16)*. Lo importante es mantener la relación cuadrática.

```
cor((log(cars$MPG.city))^(1/32), (log(cars$Price))^(1/16))
```

```
[1] -0.7698761
```

Vemos como aumenta todavía más, y este proceso lo podríamos repetir indefinidamente. Si tomamos un caso extremo, por ejemplo *(log(cars$MPG.city))^(1/128)* y *(log(cars$Price))^(1/64)* obtenemos lo siguiente:

```
cor((log(cars$MPG.city))^(1/128), (log(cars$Price))^(1/64))
```

```
[1] -0.770798
```

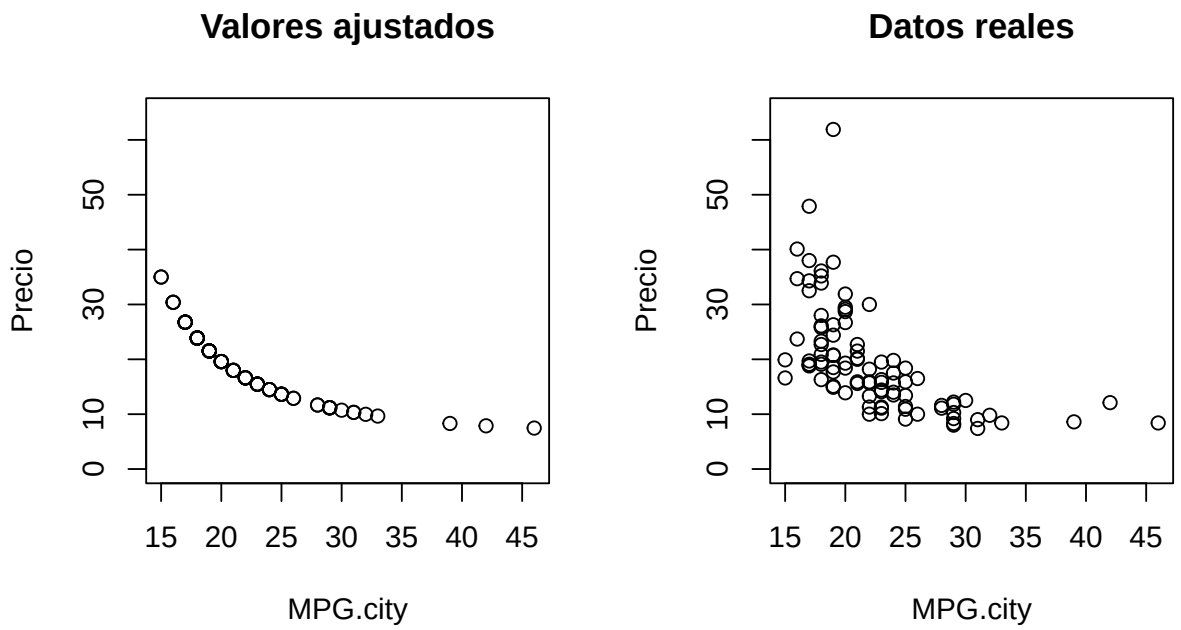
Nuevamente y sin sorpresas, vuelve a aumentar. En cualquiera de los casos, vemos cómo al aumentar `MPG.city` disminuye `Price` lo cual tiene sentido, ya que `MPG.city` es la inversa del consumo de un coche *(1 MPG=235,21 l/100km, 30 MPG=7,84 l/100km)* y los coches más caros, como deportivos o todoterrenos, tienden a consumir más. Además, hemos establecido que guardan una relación cuadrática entre ellas.

Por todo lo expuesto hasta ahora, vamos a crear el siguiente modelo.

```
fit1<-lm(log(Price)~poly(log(MPG.city),2), data=cars)
fit1.s<-summary(fit1)
fit1.s
```

```
Call:
lm(formula = log(Price) ~ poly(log(MPG.city), 2), data = cars)

Residuals:
     Min       1Q   Median       3Q      Max
-0.74595 -0.22300 -0.02439  0.18925  1.05631

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              2.86545    0.03131  91.509   <2e-16 ***
poly(log(MPG.city), 2)1 -3.24687    0.30197 -10.752   <2e-16 ***
poly(log(MPG.city), 2)2  0.52045    0.30197   1.723   0.0882 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.302 on 90 degrees of freedom
Multiple R-squared:  0.5685,    Adjusted R-squared:  0.5589
F-statistic: 59.29 on 2 and 90 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(1,2))
plot(exp(fit1$fitted.values)~exp(log(cars$MPG.city)), ylim=c(0,65),
main="Valores ajustados", ylab="Precio", xlab="MPG.city")
plot(cars$Price~cars$MPG.city, ylim=c(0,65),
main="Datos reales",ylab="Precio", xlab="MPG.city")
```




```
par(mfrow=c(1,1))
```

Como podemos ver, hemos obtenido un R^2 de 0.5685093, lo cual era exactamente lo que esperábamos ya que, en el caso de un modelo de regresión simple, el coeficiente de determinación coincide con el cuadrado del coeficiente de correlación lineal, el cual ya habíamos calculado anteriormente haciendo `cor((log(cars$MPG.city))^(1/2), log(cars$Price))`=-0.7469256. Comprobamos fácilmente que

$$(-0.7469256)^2 = 0.5685093$$

4. Ejercicio 2.

1. Considerando un tope de 10 variables, encuentra el número óptimo de variables a incluir en un modelo predictivo de `Price`, según los criterios R^2, *BIC y AIC*. ¿Qué variables incluye el modelo obtenido?. Interpreta los coeficientes obtenidos, ¿consideras que tiene sentido?.

```
library(leaps)
cars$Type<-as.factor(cars$Type)
cars$Origin<-as.factor(cars$Origin)

fit10<- regsubsets(Price~. , data=cars, nvmax=9)
fit10.s<-summary(fit10)

# Buscamos el mayor R^2
max_r2<-fit10.s$which[which.max(fit10.s$rsq),]
names(max_r2[which(max_r2==TRUE)])
```

```
 [1] "(Intercept)" "TypeMidsize" "TypeSporty"  "TypeVan"     "MPG.highway" "Horsepower"
 [7] "RPM"         "Wheelbase"   "Width"       "OriginUSA"
```

```
# Buscamos el menor BIC
min_bic<-fit10.s$which[which.min(fit10.s$bic),]
names(min_bic[which(min_bic==TRUE)])
```

```
[1] "(Intercept)" "TypeMidsize" "Horsepower"  "RPM"         "Wheelbase"   "Width"
[7] "OriginUSA"
```

```
# Buscamos el menor AIC
min_aic<-fit10.s$which[which.min(fit10.s$cp),]
names(min_aic[which(min_aic==TRUE)])
```

```
[1] "(Intercept)" "TypeMidsize" "Horsepower"  "RPM"         "Wheelbase"   "Width"
[7] "OriginUSA"
```

2. Selecciona el mejor modelo con el método `stepwise`.

```
model0 <- lm(Price ~ . , data=cars)
aj_step0 <- step(model0)
```

```
Start:  AIC=320.75
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Passengers + Length +
    Wheelbase + Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- MPG.city            1      0.02 1944.8 318.75
- Fuel.tank.capacity  1      0.87 1945.6 318.79
- Type                5    177.26 2122.0 318.86
- Passengers          1      4.50 1949.3 318.96
- Length              1      4.91 1949.7 318.98
- MPG.highway         1      7.64 1952.4 319.11
- Weight              1     11.31 1956.1 319.29
- EngineSize          1     16.48 1961.2 319.53
- Rev.per.mile        1     27.77 1972.5 320.07
<none>                           1944.8 320.75
- RPM                 1     67.93 2012.7 321.94
- Origin              1    136.68 2081.4 325.06
- Wheelbase           1    193.56 2138.3 327.57
- Width               1    222.44 2167.2 328.82
- Horsepower          1    456.18 2400.9 338.34

Step:  AIC=318.75
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Fuel.tank.capacity + Passengers + Length +
    Wheelbase + Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- Fuel.tank.capacity  1      0.96 1945.7 316.79
- Type                5    177.92 2122.7 316.89
- Passengers          1      4.48 1949.3 316.96
- Length              1      5.17 1950.0 317.00
- Weight              1     11.42 1956.2 317.29
- EngineSize          1     16.64 1961.4 317.54
- MPG.highway         1     28.19 1973.0 318.09
- Rev.per.mile        1     30.20 1975.0 318.18
<none>                           1944.8 318.75
- RPM                 1     68.76 2013.5 319.98
- Origin              1    136.88 2081.7 323.07
- Wheelbase           1    194.27 2139.1 325.60
- Width               1    222.44 2167.2 326.82
```

```
- Horsepower          1    458.38 2403.2 336.43

Step:  AIC=316.79
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Passengers + Length + Wheelbase + Width +
    Weight + Origin

               Df Sum of Sq    RSS    AIC
- Type          5    177.00 2122.8 314.89
- Passengers    1      4.18 1949.9 314.99
- Length        1      5.14 1950.9 315.04
- Weight        1     10.97 1956.7 315.32
- EngineSize    1     18.25 1964.0 315.66
- Rev.per.mile  1     33.73 1979.5 316.39
- MPG.highway   1     34.18 1979.9 316.41
<none>                      1945.7 316.79
- RPM           1     68.03 2013.8 317.99
- Origin        1    149.89 2095.6 321.70
- Wheelbase     1    193.50 2139.2 323.61
- Width         1    236.95 2182.7 325.48
- Horsepower    1    461.42 2407.2 334.59

Step:  AIC=314.89
Price ~ MPG.highway + EngineSize + Horsepower + RPM + Rev.per.mile +
    Passengers + Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- Weight        1     12.34 2135.1 313.43
- EngineSize    1     15.05 2137.8 313.55
- Passengers    1     37.49 2160.2 314.52
- Rev.per.mile  1     40.83 2163.6 314.66
<none>                      2122.8 314.89
- RPM           1     60.96 2183.7 315.52
- Length        1     66.06 2188.8 315.74
- MPG.highway   1     70.39 2193.1 315.93
- Origin        1    137.29 2260.0 318.72
- Wheelbase     1    218.71 2341.5 322.01
- Width         1    323.61 2446.4 326.09
- Horsepower    1    461.69 2584.4 331.19

Step:  AIC=313.43
Price ~ MPG.highway + EngineSize + Horsepower + RPM + Rev.per.mile +
    Passengers + Length + Wheelbase + Width + Origin

               Df Sum of Sq    RSS    AIC
- EngineSize    1     17.27 2152.4 312.18
- Rev.per.mile  1     42.90 2178.0 313.28
<none>                      2135.1 313.43
- RPM           1     49.63 2184.7 313.57
- Passengers    1     55.63 2190.7 313.82
- Length        1     58.22 2193.3 313.93
- MPG.highway   1     59.09 2194.2 313.97
- Origin        1    126.68 2261.8 316.79
- Wheelbase     1    225.23 2360.3 320.76
```

```
- Width         1    389.74 2524.8 327.02
- Horsepower    1    521.96 2657.1 331.77

Step:  AIC=312.18
Price ~ MPG.highway + Horsepower + RPM + Rev.per.mile + Passengers +
    Length + Wheelbase + Width + Origin

               Df Sum of Sq    RSS    AIC
- Rev.per.mile  1     30.91 2183.3 311.51
<none>                     2152.4 312.18
- Passengers    1     46.91 2199.3 312.19
- MPG.highway   1     46.92 2199.3 312.19
- Length        1     73.51 2225.9 313.30
- Origin        1    123.13 2275.5 315.35
- RPM           1    155.80 2308.2 316.68
- Wheelbase     1    228.58 2380.9 319.57
- Width         1    382.92 2535.3 325.41
- Horsepower    1   1224.31 3376.7 352.06

Step:  AIC=311.51
Price ~ MPG.highway + Horsepower + RPM + Passengers + Length +
    Wheelbase + Width + Origin

              Df Sum of Sq    RSS    AIC
- MPG.highway  1     41.22 2224.5 311.25
- Passengers   1     44.43 2227.7 311.38
<none>                    2183.3 311.51
- Length       1     64.16 2247.4 312.20
- RPM          1    130.52 2313.8 314.91
- Origin       1    157.10 2340.4 315.97
- Wheelbase    1    227.08 2410.3 318.71
- Width        1    428.35 2611.6 326.17
- Horsepower   1   1206.37 3389.6 350.42

Step:  AIC=311.25
Price ~ Horsepower + RPM + Passengers + Length + Wheelbase +
    Width + Origin

             Df Sum of Sq    RSS    AIC
- Passengers  1     19.55 2244.0 310.06
<none>                   2224.5 311.25
- Length      1     56.45 2280.9 311.58
- Origin      1    163.29 2387.8 315.83
- RPM         1    179.88 2404.4 316.48
- Wheelbase   1    224.76 2449.2 318.20
- Width       1    442.28 2666.8 326.11
- Horsepower  1   2101.37 4325.9 371.10

Step:  AIC=310.06
Price ~ Horsepower + RPM + Length + Wheelbase + Width + Origin

             Df Sum of Sq    RSS    AIC
<none>                   2244.0 310.06
- Length      1     66.01 2310.0 310.76
```

```
- Origin      1    165.64 2409.7 314.68
- RPM         1    182.27 2426.3 315.32
- Wheelbase   1    235.98 2480.0 317.36
- Width       1    475.49 2719.5 325.93
- Horsepower  1   2791.71 5035.8 383.23
```

El mejor modelo es el último de los mostrados anteriormente, con un AIC de 310.06.

3. Selecciona el mejor modelo con el método `stepwise` considerando la variable `Passengers` como factor.

```
cars$Passengers<-as.factor(cars$Passengers)
model1 <- lm(Price ~ . , data=cars)
aj_step1 <- step(model1)
```

```
Start:  AIC=323.98
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Passengers + Length +
    Wheelbase + Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- Passengers          4     61.61 1949.3 318.96
- MPG.city            1      0.00 1887.7 321.98
- Fuel.tank.capacity  1      0.84 1888.5 322.02
- Weight              1      1.38 1889.0 322.04
- Length              1      2.43 1890.1 322.10
- MPG.highway         1      6.14 1893.8 322.28
- EngineSize          1     26.75 1914.4 323.28
- Rev.per.mile        1     32.07 1919.7 323.54
- Type                4    160.30 2048.0 323.56
- RPM                 1     40.86 1928.5 323.97
<none>                           1887.7 323.98
- Origin              1    145.15 2032.8 328.87
- Wheelbase           1    212.46 2100.1 331.89
- Width               1    226.74 2114.4 332.52
- Horsepower          1    319.77 2207.4 336.53

Step:  AIC=318.96
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Length + Wheelbase +
    Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- MPG.city            1      0.00 1949.3 316.96
- Fuel.tank.capacity  1      0.61 1949.9 316.99
- Length              1      6.17 1955.4 317.26
- Weight              1      9.24 1958.5 317.40
- MPG.highway         1     10.39 1959.7 317.46
- EngineSize          1     17.38 1966.6 317.79
- Type                5    206.77 2156.0 318.34
- Rev.per.mile        1     29.16 1978.4 318.34
<none>                           1949.3 318.96
- RPM                 1     63.76 2013.0 319.96
- Origin              1    132.20 2081.5 323.07
- Wheelbase           1    191.25 2140.5 325.67
- Width               1    218.13 2167.4 326.83
- Horsepower          1    503.82 2453.1 338.34
```

```
Step:  AIC=316.96
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Fuel.tank.capacity + Length + Wheelbase +
    Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- Fuel.tank.capacity  1      0.65 1949.9 314.99
- Length              1      6.41 1955.7 315.27
- Weight              1      9.31 1958.6 315.41
- EngineSize          1     17.66 1966.9 315.80
- Rev.per.mile        1     32.09 1981.4 316.48
- Type                5    210.78 2160.0 316.51
- MPG.highway         1     38.15 1987.4 316.76
<none>                           1949.3 316.96
- RPM                 1     64.55 2013.8 317.99
- Origin              1    132.42 2081.7 321.08
- Wheelbase           1    191.87 2141.1 323.69
- Width               1    218.14 2167.4 324.83
- Horsepower          1    507.85 2457.1 336.50

Step:  AIC=314.99
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- Length        1      6.35 1956.3 313.30
- Weight        1      9.02 1958.9 313.42
- EngineSize    1     19.06 1969.0 313.90
- Type          5    210.32 2160.2 314.52
- Rev.per.mile  1     35.24 1985.2 314.66
<none>                     1949.9 314.99
- MPG.highway   1     44.50 1994.4 315.09
- RPM           1     64.09 2014.0 316.00
- Origin        1    145.81 2095.7 319.70
- Wheelbase     1    191.32 2141.2 321.70
- Width         1    233.15 2183.1 323.50
- Horsepower    1    515.20 2465.1 334.80

Step:  AIC=313.3
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- Weight        1      5.26 1961.5 311.55
- EngineSize    1     20.99 1977.2 312.29
- Rev.per.mile  1     35.77 1992.0 312.98
<none>                     1956.3 313.30
- MPG.highway   1     43.50 1999.8 313.34
- RPM           1     58.77 2015.0 314.05
- Type          5    287.65 2243.9 316.05
- Origin        1    140.70 2097.0 317.76
- Wheelbase     1    209.33 2165.6 320.75
- Width         1    247.28 2203.6 322.37
```

```
- Horsepower    1    532.93 2489.2 333.70

Step:  AIC=311.55
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Wheelbase + Width + Origin

               Df Sum of Sq    RSS    AIC
- EngineSize    1     19.53 1981.1 310.47
- Rev.per.mile  1     38.15 1999.7 311.34
- MPG.highway   1     38.91 2000.4 311.37
<none>                     1961.5 311.55
- RPM           1     54.11 2015.6 312.08
- Type          5    300.78 2262.3 314.81
- Origin        1    138.82 2100.3 315.91
- Wheelbase     1    233.28 2194.8 320.00
- Width         1    297.31 2258.8 322.67
- Horsepower    1    589.96 2551.5 334.00

Step:  AIC=310.47
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Wheelbase + Width + Origin

               Df Sum of Sq    RSS    AIC
- Rev.per.mile  1     27.16 2008.2 309.73
- MPG.highway   1     36.47 2017.5 310.16
<none>                     1981.1 310.47
- Type          5    301.28 2282.3 313.63
- Origin        1    135.56 2116.6 314.62
- RPM           1    175.78 2156.8 316.37
- Wheelbase     1    242.69 2223.7 319.21
- Width         1    279.66 2260.7 320.75
- Horsepower    1   1459.64 3440.7 359.81

Step:  AIC=309.73
Price ~ Type + MPG.highway + Horsepower + RPM + Wheelbase + Width +
    Origin

              Df Sum of Sq    RSS    AIC
- MPG.highway  1     28.34 2036.6 309.04
<none>                    2008.2 309.73
- Type         5    292.86 2301.1 312.39
- RPM          1    152.74 2161.0 314.55
- Origin       1    161.81 2170.0 314.94
- Wheelbase    1    240.43 2248.7 318.25
- Width        1    319.76 2328.0 321.47
- Horsepower   1   1461.88 3470.1 358.60

Step:  AIC=309.04
Price ~ Type + Horsepower + RPM + Wheelbase + Width + Origin

             Df Sum of Sq    RSS    AIC
<none>                    2036.6 309.04
- Type        5    273.49 2310.0 310.76
- Origin      1    153.02 2189.6 313.77
```

```
- RPM         1    179.86 2216.4 314.91
- Wheelbase   1    243.06 2279.6 317.52
- Width       1    337.90 2374.5 321.31
- Horsepower  1   1988.10 4024.7 370.39
```

```
aj_step1.s<-summary(aj_step1)
```

4. Depura el modelo anterior (*apartado 3*) sólo si te parece oportuno y contesta a las siguientes preguntas: ¿Qué % de la varianza de `Price` explica el modelo? ¿Cuál es la variable menos explicativa?. ¿Cuál es su efecto sobre `Price`?

Respuesta: depuraremos el modelo en función de los grupos que nos presenta la variable "type" y lo haremos de acuerdo a aquellos que tengan medias de precio más similares.

```
cars%>%group_by(Type)%>%summarize(mean(Price))
```

```
# A tibble: 6 x 2
  Type    `mean(Price)`
  <fct>           <dbl>
1 Compact          18.2
2 Large            24.3
3 Midsize          27.2
4 Small            10.2
5 Sporty           19.4
6 Van              19.1
```

```
aj_step1.s
```

```

Call:
lm(formula = Price ~ Type + Horsepower + RPM + Wheelbase + Width +
    Origin, data = cars)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2403  -2.6302  -0.1117   2.0195  22.6970

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 46.649677  27.321819   1.707 0.091529 .
TypeLarge    1.299684   2.914914   0.446 0.656864
TypeMidsize  3.961813   1.914233   2.070 0.041632 *
TypeSmall    0.334407   2.088371   0.160 0.873174
TypeSporty   3.015877   2.299341   1.312 0.193307
TypeVan     -1.849234   2.874896  -0.643 0.521866
Horsepower   0.160788   0.017971   8.947 9.17e-14 ***
RPM         -0.003695   0.001373  -2.691 0.008630 **
Wheelbase    0.663732   0.212168   3.128 0.002434 **
Width       -1.434571   0.388931  -3.688 0.000404 ***
OriginUSA   -3.217185   1.296097  -2.482 0.015098 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.984 on 82 degrees of freedom
Multiple R-squared:  0.7627,    Adjusted R-squared:  0.7338
F-statistic: 26.36 on 10 and 82 DF,  p-value: < 2.2e-16
```

```r
# Vamos a juntar los grupos "Sporty" y "Van"

cars_prueba<-cars
cars_prueba$Type<-factor(cars_prueba$Type, levels=c("Compact", "Large", "Midsize", "Small",
"Sporty", "Van"), labels=c("Compact", "Large", "Midsize", "Small", "Van", "Van"))

model2 <- lm(Price ~ . , data=cars_prueba)
aj_step2 <- step(model2)
```

```
Start:  AIC=323.98
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Passengers + Length +
    Wheelbase + Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- Passengers          5     98.97 1986.6 318.73
- MPG.city            1      0.00 1887.7 321.98
- Fuel.tank.capacity  1      0.84 1888.5 322.02
- Weight              1      1.38 1889.0 322.04
- Length              1      2.43 1890.1 322.10
- MPG.highway         1      6.14 1893.8 322.28
- EngineSize          1     26.75 1914.4 323.28
- Rev.per.mile        1     32.07 1919.7 323.54
- Type                4    160.30 2048.0 323.56
- RPM                 1     40.86 1928.5 323.97
<none>                           1887.7 323.98
- Origin              1    145.15 2032.8 328.87
- Wheelbase           1    212.46 2100.1 331.89
- Width               1    226.74 2114.4 332.52
- Horsepower          1    319.77 2207.4 336.53

Step:  AIC=318.73
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Length + Wheelbase +
    Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- Fuel.tank.capacity  1      0.62 1987.2 316.76
- MPG.city            1      1.77 1988.4 316.81
- MPG.highway         1      4.04 1990.7 316.92
- EngineSize          1     18.03 2004.7 317.57
- Length              1     23.52 2010.1 317.82
- Weight              1     28.17 2014.8 318.04
- Rev.per.mile        1     30.52 2017.1 318.15
- Type                4    169.42 2156.0 318.34
<none>                           1986.6 318.73
- RPM                 1     71.48 2058.1 320.02
- Origin              1    145.39 2132.0 323.30
- Wheelbase           1    153.92 2140.6 323.67
- Width               1    223.49 2210.1 326.64
- Horsepower          1    685.58 2672.2 344.30

Step:  AIC=316.76
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
```

```
    RPM + Rev.per.mile + Length + Wheelbase + Width + Weight +
    Origin

               Df Sum of Sq    RSS    AIC
- MPG.city      1      1.50 1988.7 314.83
- MPG.highway   1      3.94 1991.2 314.94
- EngineSize    1     17.41 2004.7 315.57
- Length        1     25.19 2012.4 315.93
- Rev.per.mile  1     30.24 2017.5 316.16
- Weight        1     30.62 2017.9 316.18
- Type          4    169.48 2156.7 316.37
<none>                      1987.2 316.76
- RPM           1     73.18 2060.4 318.12
- Origin        1    149.27 2136.5 321.49
- Wheelbase     1    153.31 2140.6 321.67
- Width         1    257.18 2244.4 326.08
- Horsepower    1    686.63 2673.9 342.36

Step:  AIC=314.83
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- EngineSize    1     16.45 2005.2 313.59
- Rev.per.mile  1     28.89 2017.6 314.17
- Length        1     29.98 2018.7 314.22
- Weight        1     32.85 2021.6 314.35
- MPG.highway   1     34.64 2023.4 314.43
- Type          4    171.49 2160.2 314.52
<none>                      1988.7 314.83
- RPM           1     76.41 2065.2 316.33
- Origin        1    149.48 2138.2 319.57
- Wheelbase     1    152.58 2141.3 319.70
- Width         1    255.74 2244.5 324.08
- Horsepower    1    718.35 2707.1 341.51

Step:  AIC=313.59
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- Rev.per.mile  1     20.34 2025.5 312.53
- Type          4    162.85 2168.0 312.86
- Weight        1     29.87 2035.1 312.97
- MPG.highway   1     31.65 2036.8 313.05
- Length        1     33.08 2038.3 313.12
<none>                      2005.2 313.59
- Origin        1    145.07 2150.3 318.09
- Wheelbase     1    155.51 2160.7 318.54
- RPM           1    201.00 2206.2 320.48
- Width         1    245.19 2250.4 322.32
- Horsepower    1   1396.39 3401.6 360.74

Step:  AIC=312.53
```

```
Price ~ Type + MPG.highway + Horsepower + RPM + Length + Wheelbase +
    Width + Weight + Origin

              Df Sum of Sq    RSS    AIC
- MPG.highway  1     28.06 2053.6 311.81
- Length       1     31.09 2056.6 311.95
- Type         4    170.25 2195.8 312.04
- Weight       1     33.38 2058.9 312.05
<none>                     2025.5 312.53
- Wheelbase    1    165.85 2191.4 317.85
- Origin       1    172.76 2198.3 318.14
- RPM          1    183.23 2208.8 318.59
- Width        1    263.51 2289.0 321.91
- Horsepower   1   1388.01 3413.5 359.07

Step:  AIC=311.81
Price ~ Type + Horsepower + RPM + Length + Wheelbase + Width +
    Weight + Origin

             Df Sum of Sq    RSS    AIC
- Weight      1     12.00 2065.6 310.35
- Length      1     25.94 2079.5 310.98
<none>                     2053.6 311.81
- Type        4    187.14 2240.7 311.92
- Wheelbase   1    147.75 2201.3 316.27
- Origin      1    150.93 2204.5 316.41
- RPM         1    177.19 2230.8 317.51
- Width       1    303.96 2357.6 322.65
- Horsepower  1   1369.45 3423.1 357.33

Step:  AIC=310.35
Price ~ Type + Horsepower + RPM + Length + Wheelbase + Width +
    Origin

             Df Sum of Sq    RSS    AIC
- Length      1     19.97 2085.6 309.25
- Type        4    178.45 2244.0 310.06
<none>                     2065.6 310.35
- Origin      1    139.02 2204.6 314.41
- RPM         1    169.34 2234.9 315.68
- Wheelbase   1    182.46 2248.1 316.23
- Width       1    320.39 2386.0 321.76
- Horsepower  1   2113.05 4178.6 373.88

Step:  AIC=309.25
Price ~ Type + Horsepower + RPM + Wheelbase + Width + Origin

             Df Sum of Sq    RSS    AIC
<none>                     2085.6 309.25
- Type        4    224.48 2310.0 310.76
- Origin      1    133.79 2219.4 313.03
- RPM         1    162.19 2247.8 314.21
- Wheelbase   1    220.32 2305.9 316.59
- Width       1    329.36 2414.9 320.88
```

```
- Horsepower  1   2093.76 4179.3 371.89
```

```
aj_step2.s<-summary(aj_step2)
aj_step2.s
```

```

Call:
lm(formula = Price ~ Type + Horsepower + RPM + Wheelbase + Width +
    Origin, data = cars_prueba)

Residuals:
     Min       1Q   Median       3Q      Max
-11.0135  -2.7094   0.0789   2.0255  23.1989

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 64.820793  24.206410   2.678 0.008930 **
TypeLarge    3.177957   2.605309   1.220 0.225995
TypeMidsize  4.726142   1.846008   2.560 0.012273 *
TypeSmall   -0.804999   1.935675  -0.416 0.678575
TypeVan      1.187294   1.906482   0.623 0.535145
Horsepower   0.163815   0.017946   9.128 3.62e-14 ***
RPM         -0.003489   0.001373  -2.541 0.012929 *
Wheelbase    0.458758   0.154927   2.961 0.003996 **
Width       -1.415474   0.390966  -3.620 0.000505 ***
OriginUSA   -2.983332   1.292877  -2.308 0.023516 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.013 on 83 degrees of freedom
Multiple R-squared:  0.757, Adjusted R-squared:  0.7307
F-statistic: 28.74 on 9 and 83 DF,  p-value: < 2.2e-16
```

```
# Vamos a juntar los grupos "Sporty", "Van" y "Compact"

cars_prueba2<-cars
cars_prueba2$Type<-factor(cars_prueba2$Type, levels=c("Compact", "Large", "Midsize", "Small",
"Sporty", "Van"), labels=c("Van", "Large", "Midsize", "Small", "Van", "Van"))

model3 <- lm(Price ~ . , data=cars_prueba2)
aj_step3 <- step(model3)
```

```
Start:  AIC=324.45
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Passengers + Length +
    Wheelbase + Width + Weight + Origin

                    Df Sum of Sq    RSS    AIC
- Passengers         5    76.787 2015.3 318.06
- MPG.city           1     0.181 1938.7 322.46
- Length             1     1.145 1939.7 322.50
- Weight             1     2.696 1941.2 322.58
- Fuel.tank.capacity 1     4.862 1943.4 322.68
- MPG.highway        1    13.607 1952.1 323.10
- EngineSize         1    18.875 1957.4 323.35
- Type               3   109.431 2048.0 323.56
```

```
- Rev.per.mile        1    28.378 1966.9 323.80
- RPM                 1    36.345 1974.9 324.18
<none>                             1938.5 324.45
- Origin              1   153.639 2092.2 329.54
- Wheelbase           1   172.452 2111.0 330.37
- Width               1   175.997 2114.5 330.53
- Horsepower          1   314.715 2253.2 336.44

Step:  AIC=318.06
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Length + Wheelbase +
    Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- MPG.city            1      0.87 2016.2 316.10
- Fuel.tank.capacity  1      1.90 2017.2 316.15
- Length              1      7.35 2022.7 316.40
- MPG.highway         1      7.48 2022.8 316.41
- EngineSize          1     13.00 2028.3 316.66
- Weight              1     20.63 2036.0 317.01
- Rev.per.mile        1     30.29 2045.6 317.45
<none>                             2015.3 318.06
- Type                3    140.72 2156.0 318.34
- RPM                 1     69.41 2084.7 319.21
- Wheelbase           1    132.26 2147.6 321.97
- Origin              1    149.30 2164.6 322.71
- Width               1    207.88 2223.2 325.19
- Horsepower          1    668.79 2684.1 342.71

Step:  AIC=316.1
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Fuel.tank.capacity + Length + Wheelbase +
    Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- Fuel.tank.capacity  1      1.54 2017.7 314.17
- Length              1     10.20 2026.4 314.57
- EngineSize          1     12.34 2028.5 314.67
- Weight              1     22.63 2038.8 315.14
- Rev.per.mile        1     30.14 2046.3 315.48
- MPG.highway         1     42.39 2058.6 316.04
<none>                             2016.2 316.10
- Type                3    143.86 2160.0 316.51
- RPM                 1     72.44 2088.6 317.38
- Wheelbase           1    131.83 2148.0 319.99
- Origin              1    148.72 2164.9 320.72
- Width               1    209.32 2225.5 323.29
- Horsepower          1    699.39 2715.6 341.80

Step:  AIC=314.17
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
```

```
- Length        1     10.73 2028.5 312.67
- EngineSize    1     11.23 2029.0 312.69
- Weight        1     24.80 2042.5 313.31
- Rev.per.mile  1     28.60 2046.3 313.48
- MPG.highway   1     41.32 2059.1 314.06
<none>                     2017.7 314.17
- Type          3    142.51 2160.2 314.52
- RPM           1     74.15 2091.9 315.53
- Wheelbase     1    130.30 2148.0 317.99
- Origin        1    149.87 2167.6 318.84
- Width         1    244.25 2262.0 322.80
- Horsepower    1    698.44 2716.2 339.82

Step:  AIC=312.67
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- EngineSize    1     15.16 2043.6 311.36
- Weight        1     20.21 2048.7 311.59
- Rev.per.mile  1     27.82 2056.3 311.93
- MPG.highway   1     35.76 2064.2 312.29
<none>                     2028.5 312.67
- RPM           1     66.78 2095.2 313.68
- Type          3    215.46 2243.9 316.05
- Origin        1    141.79 2170.2 316.95
- Wheelbase     1    142.24 2170.7 316.97
- Width         1    233.55 2262.0 320.80
- Horsepower    1    689.94 2718.4 337.89

Step:  AIC=311.36
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- Weight        1     17.72 2061.3 310.16
- Rev.per.mile  1     19.58 2063.2 310.25
- MPG.highway   1     31.26 2074.9 310.77
<none>                     2043.6 311.36
- Type          3    220.39 2264.0 314.88
- Origin        1    135.95 2179.6 315.35
- Wheelbase     1    151.06 2194.7 315.99
- RPM           1    177.48 2221.1 317.10
- Width         1    226.54 2270.2 319.14
- Horsepower    1   1370.76 3414.4 357.09

Step:  AIC=310.16
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Wheelbase + Width + Origin

               Df Sum of Sq    RSS    AIC
- MPG.highway   1     15.51 2076.8 308.86
- Rev.per.mile  1     22.62 2083.9 309.18
<none>                     2061.3 310.16
```

```
- Origin        1    118.23 2179.6 313.35
- Type          3    221.00 2282.3 313.63
- RPM           1    159.76 2221.1 315.10
- Wheelbase     1    171.33 2232.7 315.59
- Width         1    315.08 2376.4 321.39
- Horsepower    1   1690.60 3751.9 363.86

Step:  AIC=308.86
Price ~ Type + Horsepower + RPM + Rev.per.mile + Wheelbase +
    Width + Origin

               Df Sum of Sq    RSS    AIC
- Rev.per.mile  1     18.48 2095.3 307.68
<none>                     2076.8 308.86
- Type          3    216.05 2292.9 312.06
- Origin        1    120.32 2197.2 312.10
- RPM           1    177.36 2254.2 314.48
- Wheelbase     1    213.45 2290.3 315.96
- Width         1    315.58 2392.4 320.01
- Horsepower    1   1948.47 4025.3 368.40

Step:  AIC=307.68
Price ~ Type + Horsepower + RPM + Wheelbase + Width + Origin

             Df Sum of Sq    RSS    AIC
<none>                    2095.3 307.68
- Type        3    214.73 2310.0 310.76
- Origin      1    143.65 2239.0 311.85
- RPM         1    159.45 2254.8 312.50
- Wheelbase   1    212.05 2307.4 314.65
- Width       1    350.99 2446.3 320.09
- Horsepower  1   2085.56 4180.9 369.93
```

```
aj_step3.s<-summary(aj_step3)
aj_step3.s
```

```
Call:
lm(formula = Price ~ Type + Horsepower + RPM + Wheelbase + Width +
    Origin, data = cars_prueba2)

Residuals:
    Min      1Q  Median      3Q     Max
-11.058  -2.371  -0.095   2.030  23.451

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 60.867714  23.274055   2.615 0.010568 *
TypeLarge    2.266151   2.147125   1.055 0.294252
TypeMidsize  4.020706   1.452291   2.769 0.006927 **
TypeSmall   -1.336328   1.731165  -0.772 0.442326
Horsepower   0.162751   0.017799   9.144 3.05e-14 ***
RPM         -0.003457   0.001367  -2.528 0.013332 *
Wheelbase    0.430909   0.147792   2.916 0.004550 **
Width       -1.306409   0.348272  -3.751 0.000323 ***
OriginUSA   -3.072332   1.280263  -2.400 0.018617 *
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.994 on 84 degrees of freedom
Multiple R-squared:  0.7559,    Adjusted R-squared:  0.7327
F-statistic: 32.52 on 8 and 84 DF,  p-value: < 2.2e-16
```

```r
# Vamos a juntar los grupos "Sporty", "Van" y "Compact" por un lado
# y "Midsize" con "Large" por otro

cars_prueba3<-cars
cars_prueba3$Type<-factor(cars_prueba3$Type, levels=c("Compact", "Large", "Midsize", "Small",
"Sporty", "Van"), labels=c("Van", "Large", "Large", "Small", "Van", "Van"))

model4 <- lm(Price ~ . , data=cars_prueba3)
aj_step4 <- step(model4)
```

```
Start:  AIC=324.68
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Passengers + Length +
    Wheelbase + Width + Weight + Origin

                    Df Sum of Sq    RSS    AIC
- Passengers         5    44.753 2030.3 316.75
- MPG.city           1     0.166 1985.7 322.68
- Weight             1     0.469 1986.0 322.70
- Length             1     2.361 1987.9 322.79
- Fuel.tank.capacity 1     2.887 1988.4 322.81
- EngineSize         1    11.949 1997.5 323.23
- MPG.highway        1    13.086 1998.6 323.29
- Type               2    62.447 2048.0 323.56
- RPM                1    39.330 2024.8 324.50
<none>                           1985.5 324.68
- Rev.per.mile       1    44.471 2030.0 324.74
- Origin             1   131.134 2116.7 328.62
- Wheelbase          1   137.702 2123.2 328.91
- Width              1   218.223 2203.7 332.37
- Horsepower         1   312.140 2297.7 336.26

Step:  AIC=316.75
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Length + Wheelbase +
    Width + Weight + Origin

                    Df Sum of Sq    RSS    AIC
- Fuel.tank.capacity 1      0.71 2031.0 314.78
- MPG.city           1      0.80 2031.1 314.79
- Length             1      6.53 2036.8 315.05
- MPG.highway        1      7.93 2038.2 315.11
- EngineSize         1      8.40 2038.7 315.13
- Weight             1     15.67 2045.9 315.46
- Rev.per.mile       1     37.07 2067.3 316.43
<none>                           2030.3 316.75
- RPM                1     72.03 2102.3 317.99
- Type               2    125.77 2156.0 318.34
```

```
- Wheelbase           1    119.17 2149.4 320.05
- Origin              1    146.40 2176.7 321.22
- Width               1    229.88 2260.1 324.72
- Horsepower          1    679.56 2709.8 341.60

Step:  AIC=314.78
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Length + Wheelbase + Width + Weight +
    Origin

               Df Sum of Sq    RSS    AIC
- MPG.city      1      0.59 2031.6 312.81
- Length        1      7.14 2038.1 313.11
- MPG.highway   1      7.82 2038.8 313.14
- EngineSize    1      7.83 2038.8 313.14
- Weight        1     17.60 2048.6 313.58
- Rev.per.mile  1     37.08 2068.1 314.46
<none>                      2031.0 314.78
- RPM           1     73.64 2104.6 316.09
- Type          2    125.75 2156.7 316.37
- Wheelbase     1    118.47 2149.4 318.05
- Origin        1    149.80 2180.8 319.40
- Width         1    260.80 2291.8 324.02
- Horsepower    1    679.80 2710.8 339.63

Step:  AIC=312.81
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- EngineSize    1      7.45 2039.0 311.15
- Length        1      9.45 2041.0 311.24
- Weight        1     18.88 2050.4 311.67
- Rev.per.mile  1     37.57 2069.1 312.51
<none>                      2031.6 312.81
- MPG.highway   1     44.86 2076.4 312.84
- RPM           1     76.07 2107.6 314.23
- Type          2    128.67 2160.2 314.52
- Wheelbase     1    118.31 2149.9 316.07
- Origin        1    149.93 2181.5 317.43
- Width         1    260.59 2292.2 322.03
- Horsepower    1    709.66 2741.2 338.67

Step:  AIC=311.15
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- Length        1     12.76 2051.8 309.73
- Weight        1     18.52 2057.5 309.99
- Rev.per.mile  1     31.09 2070.1 310.56
- MPG.highway   1     41.82 2080.8 311.04
<none>                      2039.0 311.15
- Type          2    129.03 2168.0 312.86
```

```
- Wheelbase     1    125.87 2164.9 314.72
- Origin        1    146.97 2186.0 315.62
- RPM           1    178.83 2217.8 316.97
- Width         1    256.63 2295.6 320.17
- Horsepower    1   1375.33 3414.3 357.09

Step:  AIC=309.73
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- Weight        1     14.19 2066.0 308.37
- Rev.per.mile  1     27.94 2079.7 308.99
- MPG.highway   1     34.94 2086.7 309.30
<none>                     2051.8 309.73
- Origin        1    137.51 2189.3 313.76
- Wheelbase     1    142.91 2194.7 313.99
- Type          2    212.23 2264.0 314.88
- RPM           1    171.69 2223.5 315.20
- Width         1    243.95 2295.7 318.18
- Horsepower    1   1362.61 3414.4 355.09

Step:  AIC=308.37
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Wheelbase + Width + Origin

               Df Sum of Sq    RSS    AIC
- MPG.highway   1     20.97 2086.9 307.31
- Rev.per.mile  1     29.35 2095.3 307.68
<none>                     2066.0 308.37
- Origin        1    123.72 2189.7 311.78
- RPM           1    157.55 2223.5 313.20
- Wheelbase     1    166.70 2232.7 313.59
- Type          2    216.37 2282.3 313.63
- Width         1    324.19 2390.1 319.93
- Horsepower    1   1694.08 3760.0 362.06

Step:  AIC=307.31
Price ~ Type + Horsepower + RPM + Rev.per.mile + Wheelbase +
    Width + Origin

               Df Sum of Sq    RSS    AIC
- Rev.per.mile  1     26.00 2112.9 306.46
<none>                     2086.9 307.31
- Origin        1    128.87 2215.8 310.88
- Type          2    205.95 2292.9 312.06
- RPM           1    176.15 2263.1 312.85
- Wheelbase     1    205.78 2292.7 314.06
- Width         1    328.53 2415.5 318.91
- Horsepower    1   1991.27 4078.2 367.62

Step:  AIC=306.46
Price ~ Type + Horsepower + RPM + Wheelbase + Width + Origin
```

```
            Df Sum of Sq    RSS    AIC
<none>                    2112.9 306.46
- Type        2    197.11 2310.0 310.76
- RPM         1    152.32 2265.3 310.93
- Origin      1    162.49 2275.4 311.35
- Wheelbase   1    200.76 2313.7 312.90
- Width       1    378.27 2491.2 319.78
- Horsepower  1   2092.28 4205.2 368.47
```

```
aj_step4.s<-summary(aj_step4)
aj_step4.s
```

```
Call:
lm(formula = Price ~ Type + Horsepower + RPM + Wheelbase + Width +
    Origin, data = cars_prueba3)

Residuals:
     Min       1Q   Median       3Q      Max
-10.4306  -2.5747   0.1908   1.8201  23.7847

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 64.600150  22.806957   2.832 0.005766 **
TypeLarge    3.646886   1.380118   2.642 0.009797 **
TypeSmall   -1.640817   1.689907  -0.971 0.334327
Horsepower   0.162991   0.017766   9.174 2.39e-14 ***
RPM         -0.003369   0.001361  -2.475 0.015294 *
Wheelbase    0.416423   0.146530   2.842 0.005613 **
Width       -1.344619   0.344695  -3.901 0.000191 ***
OriginUSA   -3.231565   1.263979  -2.557 0.012345 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.986 on 85 degrees of freedom
Multiple R-squared:  0.7539,    Adjusted R-squared:  0.7336
F-statistic: 37.19 on 7 and 85 DF,  p-value: < 2.2e-16
```

```
# Vamos a juntar los grupos "Sporty", "Van", "Compact" y "Small" por un
# lado y "Midsize" con "Large" por otro.

cars_prueba4<-cars
cars_prueba4$Type<-factor(cars_prueba4$Type, levels=c("Compact", "Large", "Midsize", "Small",
"Sporty", "Van"), labels=c("Van", "Large", "Large", "Van", "Van", "Van"))

model5 <- lm(Price ~ . , data=cars_prueba4)
aj_step5 <- step(model5)
```

```
Start:  AIC=322.9
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Passengers + Length +
    Wheelbase + Width + Weight + Origin

                    Df Sum of Sq    RSS    AIC
- Passengers         5     51.85 2042.1 315.29
- Weight              1      0.23 1990.5 320.91
```

```
- MPG.city            1      0.77 1991.0 320.93
- Fuel.tank.capacity  1      1.84 1992.1 320.98
- Length              1      3.17 1993.4 321.05
- EngineSize          1      9.29 1999.5 321.33
- MPG.highway         1     12.15 2002.4 321.46
<none>                           1990.2 322.90
- Rev.per.mile        1     43.84 2034.1 322.92
- RPM                 1     45.43 2035.7 323.00
- Type                1     57.72 2048.0 323.56
- Origin              1    127.90 2118.1 326.69
- Wheelbase           1    145.04 2135.3 327.44
- Width               1    233.69 2223.9 331.22
- Horsepower          1    321.12 2311.3 334.81

Step:  AIC=315.29
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Fuel.tank.capacity + Length + Wheelbase +
    Width + Weight + Origin

                     Df Sum of Sq    RSS    AIC
- Fuel.tank.capacity  1      0.19 2042.3 313.30
- MPG.city            1      3.26 2045.3 313.44
- EngineSize          1      4.74 2046.8 313.50
- MPG.highway         1      5.39 2047.5 313.53
- Length              1      8.42 2050.5 313.67
- Weight              1     12.99 2055.1 313.88
- Rev.per.mile        1     37.06 2079.1 314.96
<none>                           2042.1 315.29
- RPM                 1     88.65 2130.7 317.24
- Type                1    113.96 2156.0 318.34
- Wheelbase           1    120.41 2162.5 318.62
- Origin              1    141.06 2183.1 319.50
- Width               1    250.02 2292.1 324.03
- Horsepower          1    739.20 2781.3 342.02

Step:  AIC=313.3
Price ~ Type + MPG.city + MPG.highway + EngineSize + Horsepower +
    RPM + Rev.per.mile + Length + Wheelbase + Width + Weight +
    Origin

               Df Sum of Sq    RSS    AIC
- MPG.city      1      3.08 2045.3 311.44
- EngineSize    1      4.57 2046.8 311.50
- MPG.highway   1      5.39 2047.7 311.54
- Length        1      8.79 2051.1 311.70
- Weight        1     14.24 2056.5 311.94
- Rev.per.mile  1     38.53 2080.8 313.04
<none>                     2042.3 313.30
- RPM           1     89.38 2131.7 315.28
- Type          1    114.46 2156.7 316.37
- Wheelbase     1    120.37 2162.6 316.62
- Origin        1    146.23 2188.5 317.73
- Width         1    276.61 2318.9 323.11
- Horsepower    1    739.01 2781.3 340.02
```

```
Step:  AIC=311.44
Price ~ Type + MPG.highway + EngineSize + Horsepower + RPM +
    Rev.per.mile + Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- EngineSize    1      3.43 2048.8 309.59
- Length        1     14.70 2060.0 310.10
- Weight        1     15.88 2061.2 310.16
- Rev.per.mile  1     35.45 2080.8 311.04
<none>                     2045.3 311.44
- MPG.highway   1     55.28 2100.6 311.92
- RPM           1     98.44 2143.8 313.81
- Type          1    114.89 2160.2 314.52
- Wheelbase     1    120.24 2165.6 314.75
- Origin        1    145.93 2191.3 315.85
- Width         1    278.06 2323.4 321.29
- Horsepower    1    810.33 2855.7 340.48

Step:  AIC=309.59
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Length + Wheelbase + Width + Weight + Origin

               Df Sum of Sq    RSS    AIC
- Weight        1     16.05 2064.8 308.32
- Length        1     16.76 2065.5 308.35
- Rev.per.mile  1     32.04 2080.8 309.04
<none>                     2048.8 309.59
- MPG.highway   1     52.46 2101.2 309.94
- Type          1    119.27 2168.0 312.86
- Wheelbase     1    125.65 2174.4 313.13
- Origin        1    144.41 2193.2 313.93
- RPM           1    190.81 2239.6 315.87
- Width         1    274.81 2323.6 319.30
- Horsepower    1   1422.45 3471.2 356.63

Step:  AIC=308.32
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
    Length + Wheelbase + Width + Origin

               Df Sum of Sq    RSS    AIC
- Length        1     11.68 2076.5 306.84
- Rev.per.mile  1     32.81 2097.6 307.78
- MPG.highway   1     36.81 2101.6 307.96
<none>                     2064.8 308.32
- Wheelbase     1    127.71 2192.5 311.90
- Origin        1    128.82 2193.6 311.95
- Type          1    134.45 2199.3 312.19
- RPM           1    175.07 2239.9 313.89
- Width         1    339.66 2404.5 320.48
- Horsepower    1   1820.26 3885.1 365.10

Step:  AIC=306.84
Price ~ Type + MPG.highway + Horsepower + RPM + Rev.per.mile +
```

```
    Wheelbase + Width + Origin

               Df Sum of Sq    RSS    AIC
- Rev.per.mile  1     29.68 2106.2 306.16
- MPG.highway   1     34.56 2111.1 306.38
<none>                     2076.5 306.84
- Origin        1    122.82 2199.3 310.19
- RPM           1    173.69 2250.2 312.31
- Wheelbase     1    184.45 2261.0 312.76
- Type          1    205.83 2282.3 313.63
- Width         1    332.25 2408.8 318.65
- Horsepower    1   1826.75 3903.3 363.54

Step:  AIC=306.16
Price ~ Type + MPG.highway + Horsepower + RPM + Wheelbase + Width +
    Origin

              Df Sum of Sq    RSS    AIC
- MPG.highway  1     30.19 2136.4 305.49
<none>                     2106.2 306.16
- RPM          1    148.05 2254.2 310.48
- Origin       1    157.98 2264.2 310.89
- Wheelbase    1    181.94 2288.1 311.87
- Type         1    194.90 2301.1 312.39
- Width        1    385.77 2492.0 319.80
- Horsepower   1   1869.62 3975.8 363.25

Step:  AIC=305.49
Price ~ Type + Horsepower + RPM + Wheelbase + Width + Origin

             Df Sum of Sq    RSS    AIC
<none>                    2136.4 305.49
- Origin      1    162.35 2298.7 310.30
- Type        1    173.68 2310.0 310.76
- RPM         1    187.38 2323.7 311.31
- Wheelbase   1    251.08 2387.4 313.82
- Width       1    393.17 2529.5 319.20
- Horsepower  1   2585.43 4721.8 377.24
```

```
aj_step5.s<-summary(aj_step5)
aj_step5.s
```

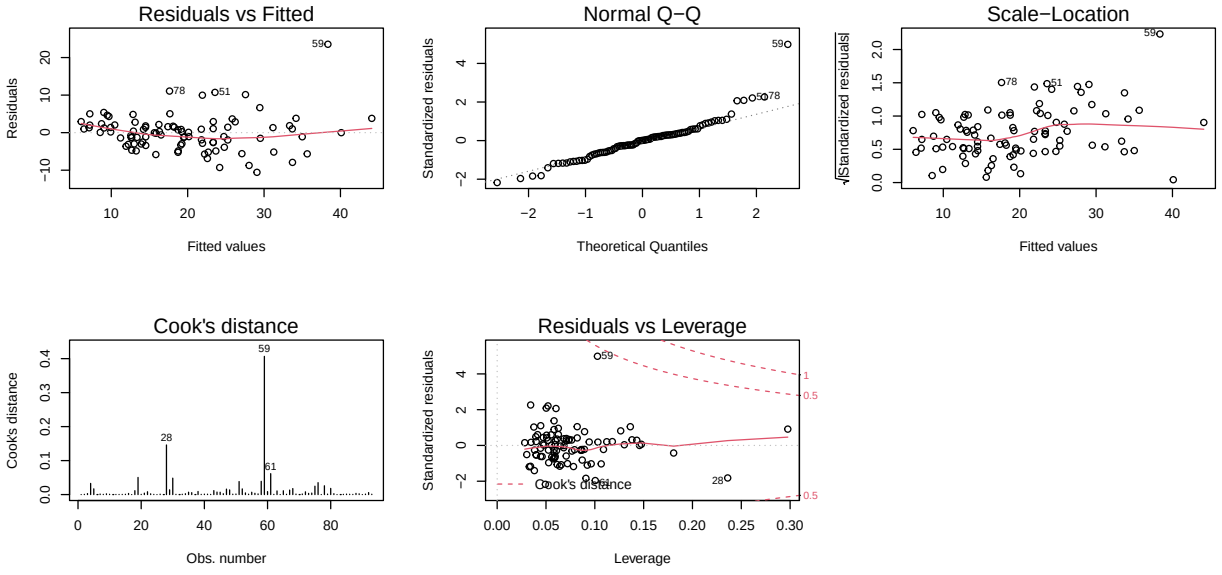
```
Call:
lm(formula = Price ~ Type + Horsepower + RPM + Wheelbase + Width +
    Origin, data = cars_prueba4)

Residuals:
     Min       1Q   Median       3Q      Max
-10.5690  -2.9436   0.0537   1.8573  23.5446

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 62.787472  22.722851   2.763 0.007000 **
TypeLarge    3.647998   1.379658   2.644 0.009733 **
Horsepower   0.169169   0.016582  10.202  < 2e-16 ***
```

```
RPM          -0.003651   0.001329  -2.746 0.007337 **
Wheelbase     0.451397   0.141986   3.179 0.002054 **
Width        -1.367611   0.343766  -3.978 0.000144 ***
OriginUSA    -3.230258   1.263558  -2.556 0.012330 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.984 on 86 degrees of freedom
Multiple R-squared:  0.7511,    Adjusted R-squared:  0.7338
F-statistic: 43.26 on 6 and 86 DF,  p-value: < 2.2e-16
```

La variabilidad que explica nuestro modelo es el coeficiente de determinación , R^2, el cual es 0.7511223. La variable menos explicativa es ***`OriginUSA`****, ya que en el contraste asociado a suponer que su coeficiente es nulo obtenemos el mayor p-valor en comparación con el resto. Suponiendo constante el efecto del resto de variables independientes sobre las variable respuesta, tenemos que un aumento de una unidad en la variable* ***`OriginUSA`****provocará un aumento de 0.12330 unidades en el precio.*

5. Realiza el diagnóstico de tu modelo, sin emprender ninguna acción, e indica los problemas que presenta. *Ayuda: Para ver los valores influyentes podéis utilizar los comandos* ***`influencePlot`*** *e* ***`influence.measures`*** *de la librería* ***`car`****. Para la normalidad usar el qqplot y/o el test de Shapiro-Wilk. Para la linealidad y homocedasticidad evaluar los residuos.*

```
par(mfrow=c(2,3))
for (i in 1:5){
  plot(aj_step5, which=i)
}
par(mfrow=c(1,1))
```




```
shapiro.test(aj_step5.s$residuals)
```

```
    Shapiro-Wilk normality test

data:  aj_step5.s$residuals
W = 0.91389, p-value = 1.331e-05
```

Respuesta: el problema más evidente que presenta el modelo es falta de normalidad de los residuos, ya que no se ajustan correctamente en el QQ-plot y el p-valor obtenido en la prueba de Shapiro-Wilk es muy significativo. Además, no parece haber homocedasticidad ya que en el primer gráfico se aprecia como los residuos se van alejando cada vez más de la línea central de tendencia.

En cuanto a los outliers, consideraremos que un punto es influyente si combina un alto leverage con un alto residuo estandarizado. También lo consideraremos influyente si marca un valor crítico en su distancia de Cook. Atendiendo a esto, el punto correspondiente al residuo 59 tiene una distancia de Cook superior al resto, y aunque podrñiamos considerarlo no influyente, a esto se suma que su residuo estandarizado está por encima del valor 1.7320508. En teoría, los resiudos estandarizados siguen una distribución N(0,1), por lo que a tres desviaciones típicas de la media deberíamos de encontrar aproximadamente el 99 % de los datos. Al hacer la raíz cuadrada, nos quedamos en el intervalo $[0, \sqrt{3}]$*. Por ello consideramos que estar por encima de este valor se corresponde con tener un residuo grande. Además en el QQ-plot el valor 59 es uno de los que más se alejan de la recta.*

La linealidad no parece verse muy afectada, pese a que se dibuja cierta forma de parábola en los residuos.

6. Emprende ahora las acciones que te parezcan oportunas e indica los problemas que has conseguido solucionar o disminuir.

En primer lugar vamos a revisar la observación número 59 de nuestro banco de datos:

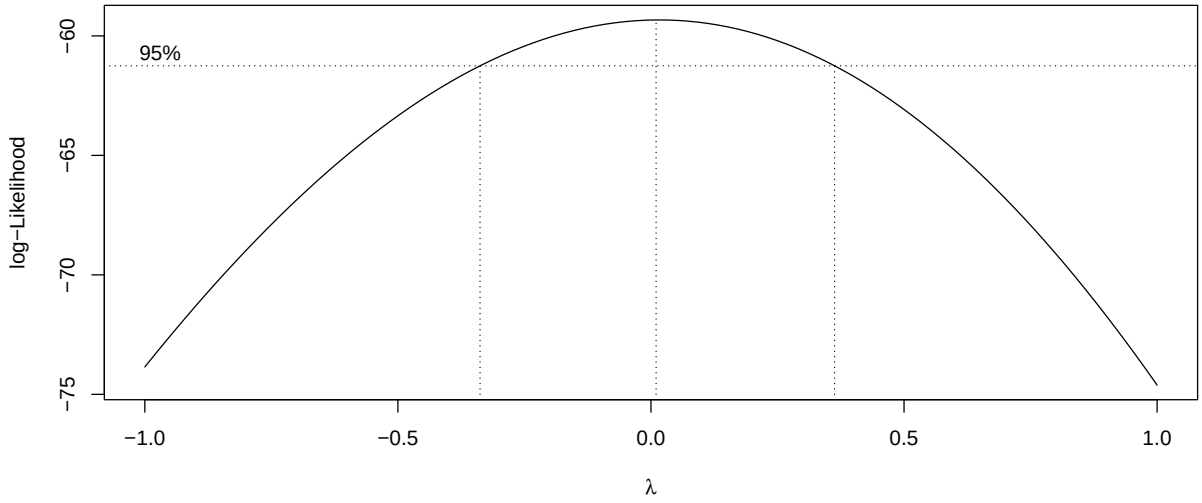
```
cars_prueba4[59,]
```

```
    Type Price MPG.city MPG.highway EngineSize Horsepower  RPM Rev.per.mile Fuel.tank.capacity
59 Large  61.9       19          25        3.2        217 5500         2220               18.5
   Passengers Length Wheelbase Width Weight  Origin
59          5    187       110    69   3525 non-USA
```

Como podemos ver, esta obervación se aleja mucho del resto de las de su grupo. Al comienzo vimos que la media más alta dentro de un grupo era 27.21818, en el grupo Midsize. Es cierto que ahora hay menos grupos, ya que los hemos reorganizado en 2, pero aún así 61.9 se aleja mucho de cualquier media de cualquier grupo. Dado que el objetivo de este estudio es crear un modelo que explique de la mejor forma posible todos los datos, vamos a considerar la eliminación de esta observación.

Probaremos también a realizar una transformación de Box-Cox sobre la variables respuesta seleccionando el λ *óptimo.*

```
library(dplyr)
library(MASS)
cars_prueba5<-cars_prueba4%>%filter(Price<61.9)

aj_step6 <- lm(formula = Price ~ Type + Horsepower + RPM + Wheelbase + Width + Origin,
data = cars_prueba5)
boxcox(aj_step6, lambda = seq(-1,1, length = 10))
```

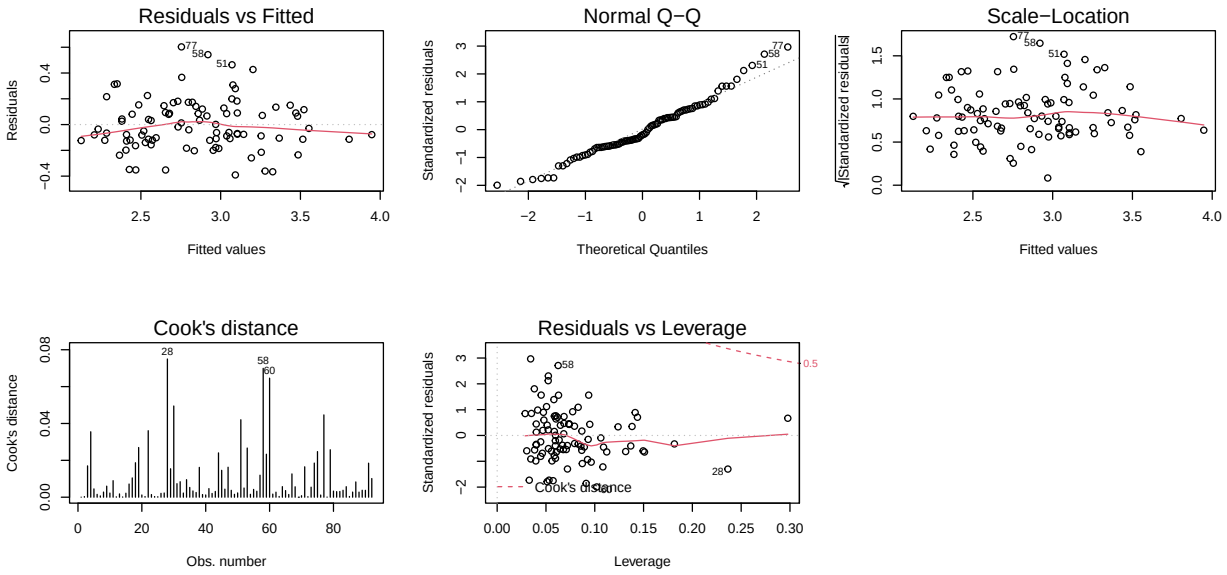



```
# Como el óptimo lo alcanzamos en lambda=0, realizamos una transforma-
# ción logarítmica de la variable respuesta
cars_prueba5$Price<-log(cars_prueba5$Price)
aj_step6<-lm(formula = Price ~ Type + Horsepower + RPM + Wheelbase + Width + Origin,
data = cars_prueba5)
aj_step6.s<-summary(aj_step6)
aj_step6.s
```

```
Call:
lm(formula = Price ~ Type + Horsepower + RPM + Wheelbase + Width +
    Origin, data = cars_prueba5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.39081 -0.12138 -0.03779  0.12237  0.60270

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.392e+00  9.548e-01   2.506  0.01413 *
TypeLarge    1.143e-01  5.742e-02   1.991  0.04968 *
Horsepower   6.729e-03  7.033e-04   9.569 3.82e-15 ***
RPM         -1.171e-04  5.547e-05  -2.111  0.03772 *
Wheelbase    2.131e-02  5.944e-03   3.586  0.00056 ***
Width       -2.967e-02  1.462e-02  -2.030  0.04551 *
OriginUSA   -1.543e-01  5.249e-02  -2.941  0.00422 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2067 on 85 degrees of freedom
Multiple R-squared:  0.7914,    Adjusted R-squared:  0.7767
F-statistic: 53.75 on 6 and 85 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,3))
for (i in 1:5){
```

```
  plot(aj_step6, which=i)
}
par(mfrow=c(1,1))
```




```
shapiro.test(aj_step6.s$residuals)
```

```
    Shapiro-Wilk normality test

data:  aj_step6.s$residuals
W = 0.97163, p-value = 0.04217
```

Se puede apreciar una notable mejora del modelo. Por supuesto todavía no cumple con todas las hipótesis de normalidad y demás, pero sin embargo si considerásemos un nivel de significatividad $\alpha = 0.01$ podríamos asumir mediante el test de Shapir-Wilk la normalidad de los residuos y ya no tenemos ninguno de ellos fuera del intervalo $[0, \sqrt{3}]$ tras normalizarlos.

7. Con el modelo final obtenido, obtén la predicción del precio para un coche en la mediana de los predictores.

Respuesta: para las variables de tipo categórico emplearemos la moda en vez de la mediana.

```
library(modeest)
names(aj_step6$coefficients)
```

```
[1] "(Intercept)" "TypeLarge"   "Horsepower"  "RPM"         "Wheelbase"   "Width"
[7] "OriginUSA"
```

```
data<-data.frame("Type"=mlv(cars_prueba5$Type, method="mvf"), "Horsepower"=
median(cars_prueba5$Horsepower), "RPM"=
median(cars_prueba5$RPM), "Wheelbase"=
median(cars_prueba5$Wheelbase), "Width"=
median(cars_prueba5$Width), "Origin"=
mlv(cars_prueba5$Origin, method="mvf"))

exp(predict(aj_step6, newdata=data, interval="confidence"))
```

```
       fit      lwr      upr
1 15.17195 14.05242 16.38068
```