

Act. Integradora 2

Eliezer Cavazos

2024-09-06

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 . Las variables recopiladas vienen descritas en el diccionario de términos 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:

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?

Primer grupo. Distancia entre los ejes (wheelbase), tipo de gasolina que usa y caballos de fuerza

```
oPreciosAutos =  
read.csv("C:\\Users\\eliez\\OneDrive\\Desktop\\Clases\\precios_autos.csv")  
#Leer la base de datos
```

```
#oRear = subset(oPreciosAutos,oPreciosAutos$engineLocation=="rear")
```

```
M1 = data.frame(oPreciosAutos$wheelbase, oPreciosAutos$horsepower,  
oPreciosAutos$price)#  
#M1$fueltype_num <- ifelse(M1$oPrecioAutos.fueltype == "diesel", 1, 0)  
oDiesel = subset(oPreciosAutos,oPreciosAutos$fueltype=="diesel")  
oGas = subset(oPreciosAutos,oPreciosAutos$fueltype=="gas")  
#M1$fueltype[M1$fueltype == "gas"] = 1  
#M1$fueltype[M1$fueltype == "rear"] = 0
```

Exploración de la base de datos

1. Exploración de la base de datos

1.1 Calcula medidas estadísticas apropiadas para las variables:

1.1.1 cuantitativas (media, desviación estándar, cuantiles, etc)

summary(M1)

```
## oPreciosAutos.wheelbase oPreciosAutos.horsepower oPreciosAutos.price
## Min. : 86.60 Min. : 48.0 Min. : 5118
## 1st Qu.: 94.50 1st Qu.: 70.0 1st Qu.: 7788
## Median : 97.00 Median : 95.0 Median :10295
## Mean : 98.76 Mean :104.1 Mean :13277
## 3rd Qu.:102.40 3rd Qu.:116.0 3rd Qu.:16503
## Max. :120.90 Max. :288.0 Max. :45400
```

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

prop.table(M1)

```
## oPreciosAutos.wheelbase oPreciosAutos.horsepower oPreciosAutos.price
## 1 3.206294e-05 4.016915e-05 0.004883628
## 2 3.206294e-05 4.016915e-05 0.005971090
## 3 3.419806e-05 5.573017e-05 0.005971090
## 4 3.611604e-05 3.691219e-05 0.005048285
## 5 3.597129e-05 4.161668e-05 0.006314880
## 6 3.611604e-05 3.980726e-05 0.005518734
## 7 3.828735e-05 3.980726e-05 0.006408969
## 8 3.828735e-05 3.980726e-05 0.006846849
## 9 3.828735e-05 5.066379e-05 0.008639986
## 10 3.600748e-05 5.790147e-05 0.006462951
## 11 3.662268e-05 3.655031e-05 0.005945758
## 12 3.662268e-05 3.655031e-05 0.006124890
## 13 3.662268e-05 4.378799e-05 0.007588712
## 14 3.662268e-05 4.378799e-05 0.007637566
## 15 3.745502e-05 4.378799e-05 0.008889686
## 16 3.745502e-05 6.586293e-05 0.011131559
## 17 3.745502e-05 6.586293e-05 0.014951246
## 18 3.980726e-05 6.586293e-05 0.013346290
## 19 3.199056e-05 1.737044e-05 0.001864066
## 20 3.419806e-05 2.533190e-05 0.002278061
## 21 3.419806e-05 2.533190e-05 0.002379389
## 22 3.390855e-05 2.460813e-05 0.002016419
## 23 3.390855e-05 2.460813e-05 0.002307736
## 24 3.390855e-05 3.691219e-05 0.002879513
## 25 3.390855e-05 2.460813e-05 0.002254177
## 26 3.390855e-05 2.460813e-05 0.002421729
## 27 3.390855e-05 2.460813e-05 0.002753577
## 28 3.390855e-05 3.691219e-05 0.003097005
## 29 3.738264e-05 3.184581e-05 0.003228369
```

## 30	3.470470e-05	5.247321e-05	0.004691467
## 31	3.133917e-05	2.098928e-05	0.002344648
## 32	3.133917e-05	2.750320e-05	0.002480716
## 33	3.390855e-05	2.171305e-05	0.001953813
## 34	3.390855e-05	2.750320e-05	0.002362742
## 35	3.390855e-05	2.750320e-05	0.002579873
## 36	3.492183e-05	2.750320e-05	0.002639945
## 37	3.492183e-05	2.750320e-05	0.002639945
## 38	3.492183e-05	3.112204e-05	0.002857076
## 39	3.492183e-05	3.112204e-05	0.003291337
## 40	3.492183e-05	3.112204e-05	0.003200866
## 41	3.492183e-05	3.112204e-05	0.003725598
## 42	3.492183e-05	3.655031e-05	0.004684591
## 43	3.492183e-05	3.618842e-05	0.003743692
## 44	3.412568e-05	2.822697e-05	0.002455384
## 45	3.419806e-05	2.533190e-05	0.003226741
## 46	3.419806e-05	2.533190e-05	0.003226741
## 47	3.474088e-05	3.256958e-05	0.003998097
## 48	4.089292e-05	6.369162e-05	0.011670766
## 49	4.089292e-05	6.369162e-05	0.012864984
## 50	3.691219e-05	9.481366e-05	0.013027832
## 51	3.369142e-05	2.460813e-05	0.001879989
## 52	3.369142e-05	2.460813e-05	0.002205684
## 53	3.369142e-05	2.460813e-05	0.002459003
## 54	3.369142e-05	2.460813e-05	0.002422815
## 55	3.369142e-05	2.460813e-05	0.002676134
## 56	3.448757e-05	3.655031e-05	0.003960823
## 57	3.448757e-05	3.655031e-05	0.004286519
## 58	3.448757e-05	3.655031e-05	0.004937910
## 59	3.448757e-05	4.885437e-05	0.005661679
## 60	3.575416e-05	3.039827e-05	0.003200866
## 61	3.575416e-05	3.039827e-05	0.003074206
## 62	3.575416e-05	3.039827e-05	0.003834163
## 63	3.575416e-05	3.039827e-05	0.003707504
## 64	3.575416e-05	2.316059e-05	0.003906540
## 65	3.575416e-05	3.039827e-05	0.004069388
## 66	3.796165e-05	4.342611e-05	0.006615243
## 67	3.796165e-05	2.605566e-05	0.006638404
## 68	3.980726e-05	4.451176e-05	0.009246866
## 69	3.980726e-05	4.451176e-05	0.010222505
## 70	3.861305e-05	4.451176e-05	0.010196450
## 71	4.183382e-05	4.451176e-05	0.011435541
## 72	4.183382e-05	5.609205e-05	0.012370650
## 73	3.495802e-05	5.609205e-05	0.012686213
## 74	4.375180e-05	6.658670e-05	0.014822778
## 75	4.053103e-05	6.658670e-05	0.016429543
## 76	3.716551e-05	6.332974e-05	0.005972175
## 77	3.390855e-05	2.460813e-05	0.001950194
## 78	3.390855e-05	2.460813e-05	0.002239701
## 79	3.390855e-05	2.460813e-05	0.002413406

## 80	3.365523e-05	3.691219e-05	0.002782528
## 81	3.484945e-05	4.197857e-05	0.003604005
## 82	3.484945e-05	3.184581e-05	0.003075654
## 83	3.470470e-05	5.247321e-05	0.004570236
## 84	3.470470e-05	5.247321e-05	0.005380856
## 85	3.470470e-05	5.247321e-05	0.005243340
## 86	3.484945e-05	3.184581e-05	0.002529209
## 87	3.484945e-05	3.184581e-05	0.002963470
## 88	3.484945e-05	4.197857e-05	0.003357924
## 89	3.484945e-05	4.197857e-05	0.003357924
## 90	3.419806e-05	2.497001e-05	0.001990001
## 91	3.419806e-05	1.990363e-05	0.002569016
## 92	3.419806e-05	2.497001e-05	0.002406168
## 93	3.419806e-05	2.497001e-05	0.002478545
## 94	3.419806e-05	2.497001e-05	0.002659487
## 95	3.419806e-05	2.497001e-05	0.002641393
## 96	3.419806e-05	2.497001e-05	0.002822335
## 97	3.419806e-05	2.497001e-05	0.002713770
## 98	3.419806e-05	2.497001e-05	0.002894712
## 99	3.441519e-05	2.497001e-05	0.002985183
## 100	3.517515e-05	3.510277e-05	0.003238502
## 101	3.517515e-05	3.510277e-05	0.003455632
## 102	3.633318e-05	5.500640e-05	0.004885075
## 103	3.633318e-05	5.500640e-05	0.005210771
## 104	3.633318e-05	5.500640e-05	0.004885075
## 105	3.304003e-05	5.790147e-05	0.006224047
## 106	3.304003e-05	7.237684e-05	0.007128757
## 107	3.589891e-05	5.790147e-05	0.006658308
## 108	3.904731e-05	3.510277e-05	0.004306422
## 109	3.904731e-05	3.437900e-05	0.004776872
## 110	4.132718e-05	3.510277e-05	0.004501840
## 111	4.132718e-05	3.437900e-05	0.005015715
## 112	3.904731e-05	3.437900e-05	0.005638156
## 113	3.904731e-05	3.437900e-05	0.006115843
## 114	4.132718e-05	3.437900e-05	0.006041657
## 115	4.132718e-05	3.437900e-05	0.006179173
## 116	3.904731e-05	3.510277e-05	0.006018135
## 117	3.904731e-05	3.437900e-05	0.006495822
## 118	3.908350e-05	5.138756e-05	0.006568199
## 119	3.390855e-05	2.460813e-05	0.002016419
## 120	3.390855e-05	3.691219e-05	0.002879513
## 121	3.390855e-05	2.460813e-05	0.002254177
## 122	3.390855e-05	2.460813e-05	0.002421729
## 123	3.390855e-05	2.460813e-05	0.002753577
## 124	3.738264e-05	3.184581e-05	0.003228369
## 125	3.470470e-05	5.247321e-05	0.004619090
## 126	3.419806e-05	5.174944e-05	0.007967967
## 127	3.238864e-05	7.491003e-05	0.011771370
## 128	3.238864e-05	7.491003e-05	0.012314196
## 129	3.238864e-05	7.491003e-05	0.013399849

## 130	3.560941e-05	1.042227e-04	0.011363345
## 131	3.477707e-05	3.256958e-05	0.003363714
## 132	3.477707e-05	3.256958e-05	0.003580844
## 133	3.586273e-05	3.980726e-05	0.004288328
## 134	3.586273e-05	3.980726e-05	0.004404131
## 135	3.586273e-05	3.980726e-05	0.005442739
## 136	3.586273e-05	3.980726e-05	0.005612824
## 137	3.586273e-05	5.790147e-05	0.006568199
## 138	3.586273e-05	5.790147e-05	0.006738284
## 139	3.390855e-05	2.497001e-05	0.001852123
## 140	3.390855e-05	2.641755e-05	0.002552369
## 141	3.376380e-05	2.641755e-05	0.002751406
## 142	3.517515e-05	2.967451e-05	0.002578787
## 143	3.517515e-05	2.967451e-05	0.002813650
## 144	3.517515e-05	3.401712e-05	0.003604367
## 145	3.510277e-05	2.967451e-05	0.003341277
## 146	3.510277e-05	4.016915e-05	0.004074454
## 147	3.510277e-05	2.967451e-05	0.002700742
## 148	3.510277e-05	3.401712e-05	0.003690495
## 149	3.506658e-05	2.967451e-05	0.002899778
## 150	3.506658e-05	4.016915e-05	0.004231874
## 151	3.463232e-05	2.243682e-05	0.001935357
## 152	3.463232e-05	2.243682e-05	0.002293622
## 153	3.463232e-05	2.243682e-05	0.002347905
## 154	3.463232e-05	2.243682e-05	0.002503515
## 155	3.463232e-05	2.243682e-05	0.002858162
## 156	3.463232e-05	2.243682e-05	0.003176620
## 157	3.463232e-05	2.533190e-05	0.002510753
## 158	3.463232e-05	2.533190e-05	0.002604843
## 159	3.463232e-05	2.026552e-05	0.002858162
## 160	3.463232e-05	2.026552e-05	0.002818354
## 161	3.463232e-05	2.533190e-05	0.002800260
## 162	3.463232e-05	2.533190e-05	0.003024628
## 163	3.463232e-05	2.533190e-05	0.003350324
## 164	3.419806e-05	2.533190e-05	0.002916063
## 165	3.419806e-05	2.533190e-05	0.002981202
## 166	3.419806e-05	4.053103e-05	0.003364799
## 167	3.419806e-05	4.053103e-05	0.003451652
## 168	3.560941e-05	4.197857e-05	0.003057560
## 169	3.560941e-05	4.197857e-05	0.003488202
## 170	3.560941e-05	4.197857e-05	0.003614861
## 171	3.560941e-05	4.197857e-05	0.004052741
## 172	3.560941e-05	4.197857e-05	0.004179401
## 173	3.560941e-05	4.197857e-05	0.006394132
## 174	3.705694e-05	3.329335e-05	0.003238140
## 175	3.705694e-05	2.641755e-05	0.003871437
## 176	3.705694e-05	3.329335e-05	0.003614500
## 177	3.705694e-05	3.329335e-05	0.003943814
## 178	3.705694e-05	3.329335e-05	0.004070474
## 179	3.723789e-05	5.826336e-05	0.005992079

## 180	3.723789e-05	5.826336e-05	0.005789424
## 181	3.781690e-05	5.645394e-05	0.005677963
## 182	3.781690e-05	5.645394e-05	0.005699676
## 183	3.521133e-05	1.881798e-05	0.002813650
## 184	3.521133e-05	3.076016e-05	0.002886027
## 185	3.521133e-05	1.881798e-05	0.002893264
## 186	3.521133e-05	3.076016e-05	0.002965641
## 187	3.521133e-05	3.076016e-05	0.003074206
## 188	3.521133e-05	2.460813e-05	0.003436091
## 189	3.521133e-05	3.618842e-05	0.003617033
## 190	3.419806e-05	3.256958e-05	0.004196047
## 191	3.419806e-05	3.256958e-05	0.003611604
## 192	3.633318e-05	3.980726e-05	0.004811251
## 193	3.633318e-05	2.460813e-05	0.005010287
## 194	3.633318e-05	3.184581e-05	0.004447557
## 195	3.774452e-05	4.125480e-05	0.004682782
## 196	3.774452e-05	4.125480e-05	0.004854677
## 197	3.774452e-05	4.125480e-05	0.005784719
## 198	3.774452e-05	4.125480e-05	0.005976518
## 199	3.774452e-05	5.862524e-05	0.006665907
## 200	3.774452e-05	5.862524e-05	0.006857706
## 201	3.948157e-05	4.125480e-05	0.006095940
## 202	3.948157e-05	5.790147e-05	0.006892085
## 203	3.948157e-05	4.849249e-05	0.007775082
## 204	3.948157e-05	3.835973e-05	0.008131538
## 205	3.948157e-05	4.125480e-05	0.008187630

1.2 Analiza la correlación entre las variables (analiza posible colinealidad entre las variables) cor(M1)

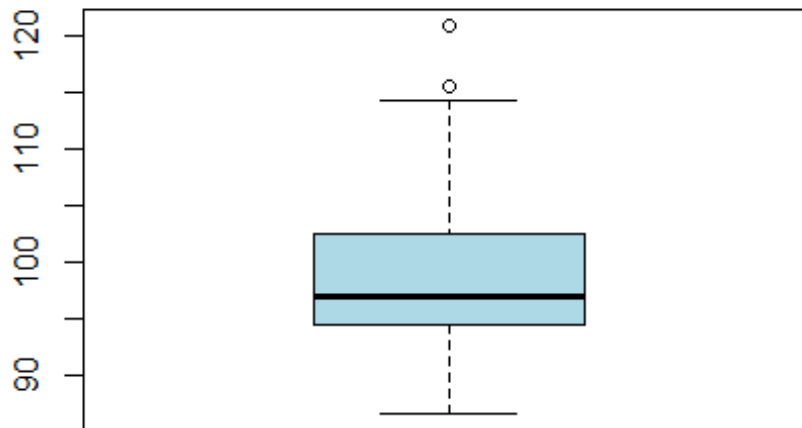
##		oPreciosAutos.wheelbase	oPreciosAutos.horsepower
##	oPreciosAutos.wheelbase	1.0000000	0.3532945
##	oPreciosAutos.horsepower	0.3532945	1.0000000
##	oPreciosAutos.price	0.5778156	0.8081388
##		oPreciosAutos.price	
##	oPreciosAutos.wheelbase	0.5778156	
##	oPreciosAutos.horsepower	0.8081388	
##	oPreciosAutos.price	1.0000000	

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

Variables cuantitativas: - Boxplot (visualización de datos atípicos)

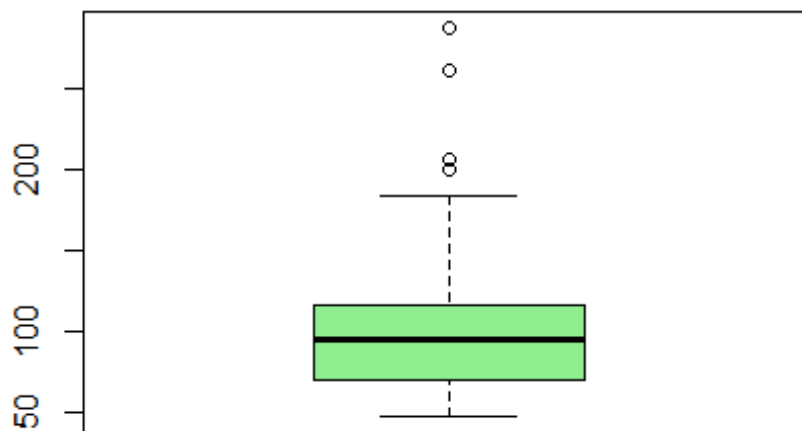
```
boxplot(M1$oPreciosAutos.wheelbase, col="lightblue", main="Wheelbase")
```

Wheelbase



```
boxplot(M1$oPreciosAutos.horsepower, col="lightgreen", main="Horsepower")
```

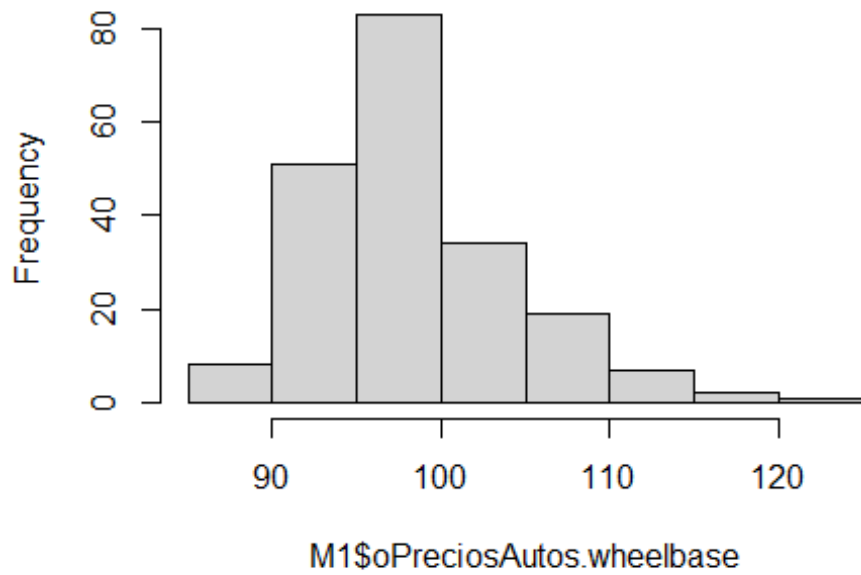
Horsepower



- Histogramas

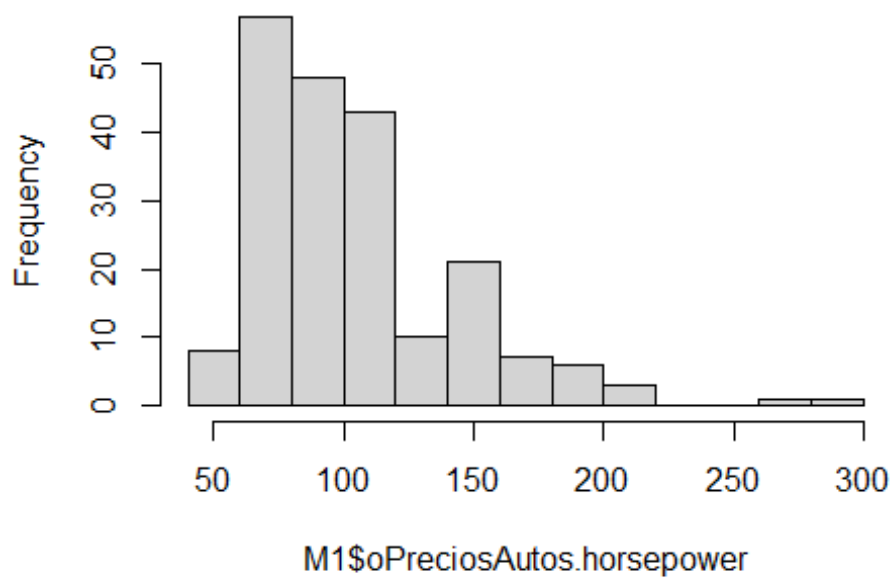
```
hist(M1$oPreciosAutos.wheelbase, main="Grafica Wheelbase")
```

Grafica Wheelbase



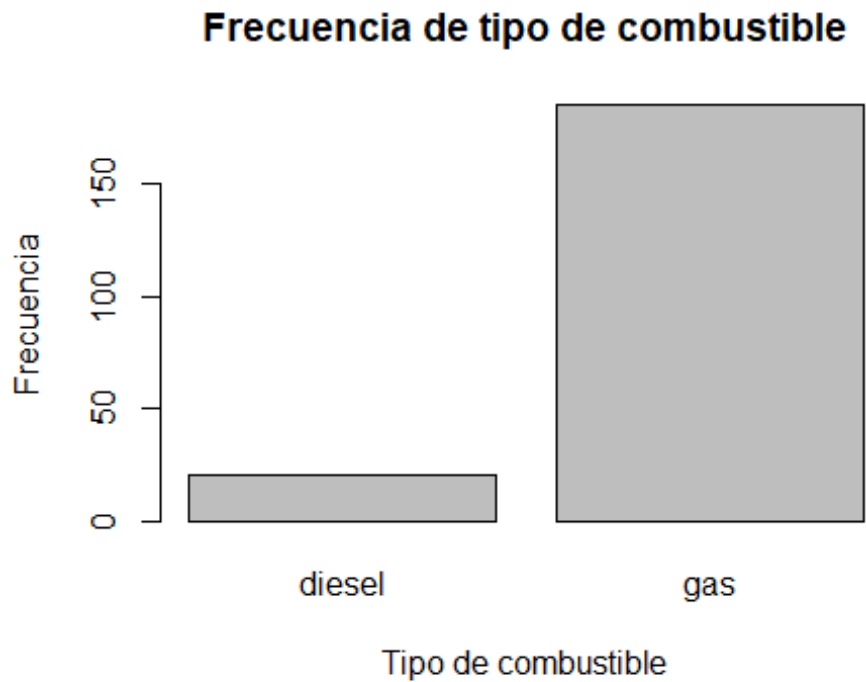
```
hist(M1$oPreciosAutos.horsepower, main="Grafica HorsePower")
```

Grafica HorsePower



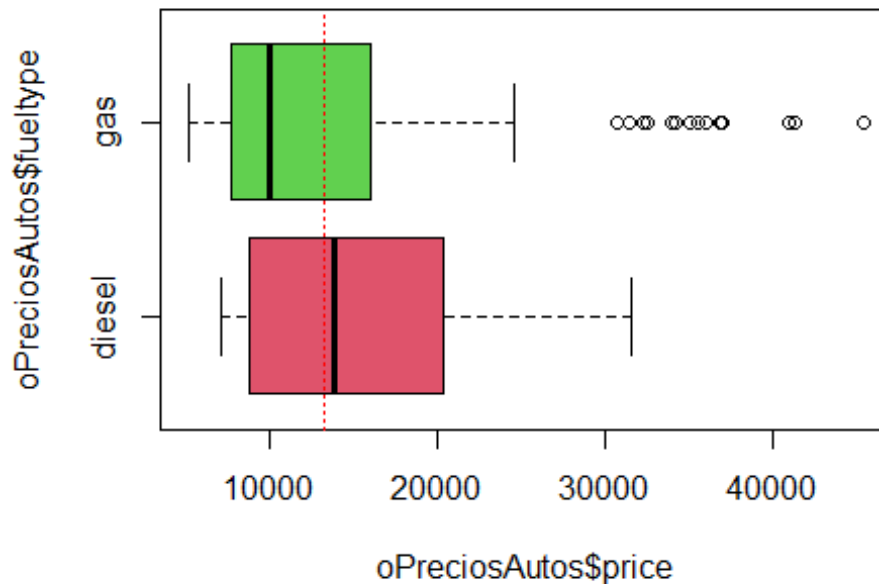
Variables categóricas - Distribución de los datos (diagramas de barras, diagramas de pastel)

```
barplot(table(oPreciosAutos$fueltype), main="Frecuencia de tipo de combustible", xlab="Tipo de combustible", ylab="Frecuencia")
```



- Boxplot por categoría de las variables cuantitativas

```
boxplot(oPreciosAutos$price ~ oPreciosAutos$fueltype, oPreciosAutos, col = 2:5, horizontal=TRUE)  
abline(v = mean(oPreciosAutos$price), lty = 3, col = "red")
```



2. Modelación y verificación del modelo

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.

```
ModeloA = lm(price~wheelbase*fueltype, oPreciosAutos)
```

```
ModeloB = lm(price~horsepower*fueltype, oPreciosAutos)
```

2.2 Para cada uno de los modelos propuestos:

2.2.1 Realiza la regresión entre las variables involucradas

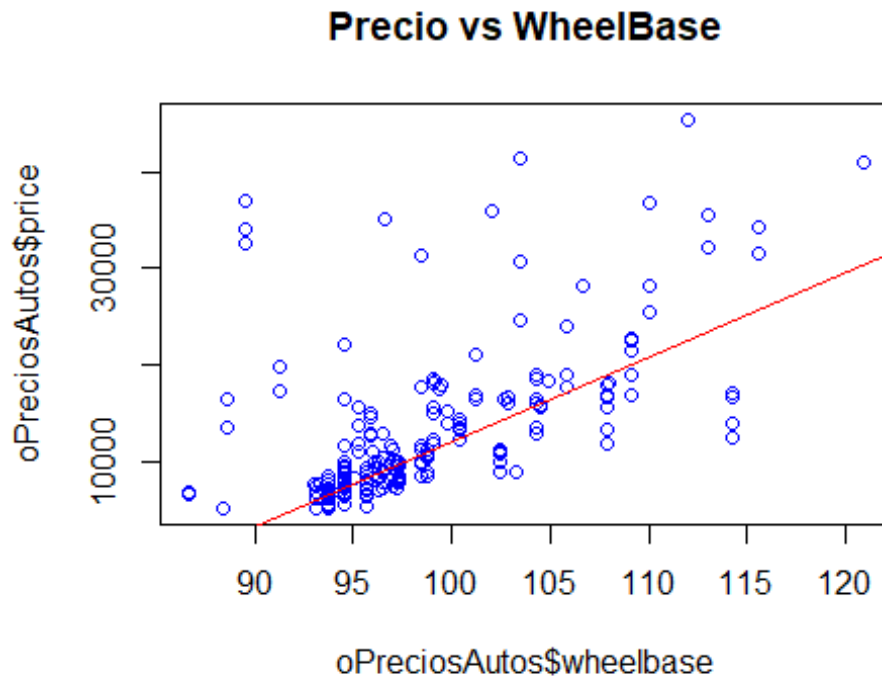
```
summary(ModeloA)
```

```
##
## Call:
## lm(formula = price ~ wheelbase * fueltype, data = oPreciosAutos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13199  -3309  -1889   1106   30837
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -75667.31   22838.89  -3.313  0.00109 **
## wheelbase      876.57    218.34   4.015  8.4e-05 ***
## fueltypegas   11386.91   24348.79   0.468  0.64054
```

```
## wheelbase:fueltypegas    -89.19      234.61   -0.380   0.70424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6538 on 201 degrees of freedom
## Multiple R-squared:  0.3402, Adjusted R-squared:  0.3303
## F-statistic: 34.54 on 3 and 201 DF,  p-value: < 2.2e-16

plot(oPreciosAutos$wheelbase, oPreciosAutos$price, col="blue", main="Precio
vs WheelBase")
abline(ModeloA, col="red")

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



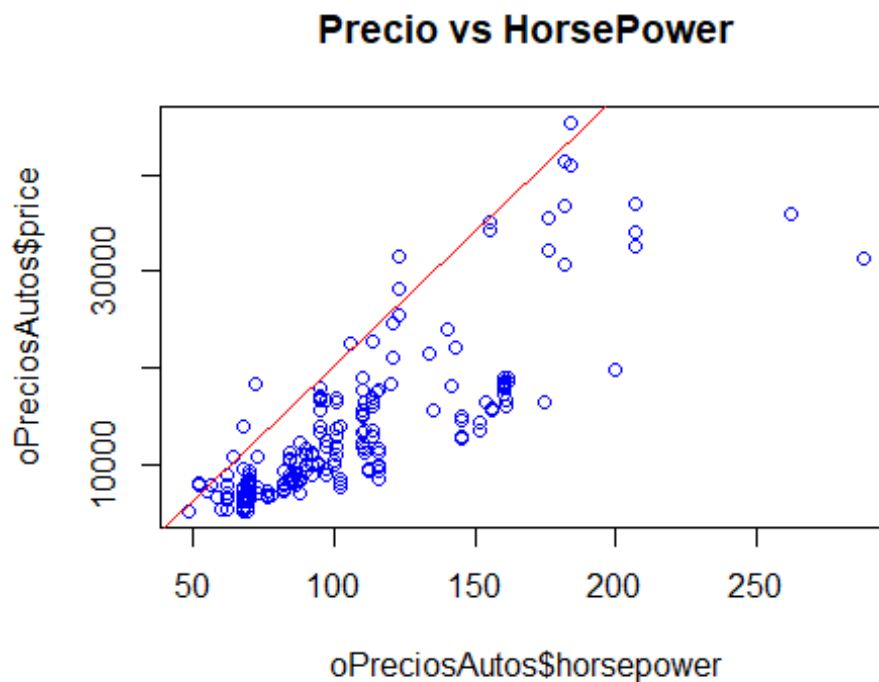
```
summary(ModeloB)

##
## Call:
## lm(formula = price ~ horsepower * fueltype, data = oPreciosAutos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11904.3  -1776.2   -381.8   1458.9  19435.5
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)          -7731.37    3298.65  -2.344  0.02006  *
## horsepower           279.09      37.42   7.459 2.56e-12 ***
## fueltypegas          3016.83    3414.35   0.884  0.37798
## horsepower:fueltypegas -112.36     38.21  -2.940  0.00366 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4234 on 201 degrees of freedom
## Multiple R-squared:  0.7233, Adjusted R-squared:  0.7191
## F-statistic: 175.1 on 3 and 201 DF,  p-value: < 2.2e-16

plot(oPreciosAutos$horsepower,oPreciosAutos$price, col="blue", main="Precio
vs HorsePower")
abline(ModeloB, col="red")

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



2.2.2 Analiza la significancia del modelo:

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

**** Modelo A ****

```

iNModeloA = length(ModeloA)

cat("Valor T limite: ", abs( qt(0.04/2,(iNModeloA-2) ) ))

## Valor T limite:  2.32814

print(" ")

## [1] " "

ModeloA$coefficients

##           (Intercept)           wheelbase           fueltypegas
##          -75667.30912             876.57303             11386.91120
## wheelbase:fueltypegas
##             -89.18525

```

A 0.04 no es significativo y el modelo quedaria:

Precio = -75668.30 + 876.57303WB + 11386.91GAS + -89.18525WB:GAS

**** Modelo B ****

```

iNModeloB = length(ModeloB)

cat("Valor T limite: ", abs( qt(0.04/2,(iNModeloB-2) ) ))

## Valor T limite:  2.32814

print(" ")

## [1] " "

ModeloB$coefficients

##           (Intercept)           horsepower           fueltypegas
##          -7731.3655             279.0943             3016.8270
## horsepower:fueltypegas
##          -112.3606

```

A 0.04 no es significativo y el modelo quedaria:

Precio = -7731.3655 + 279.0943HP + 3016.8270GAS + -112.3606HP:GAS

Al ver que ningun modelo es significativo genero un nuevo modelo solo usando los valores significativos como el wheelbase y horsepower

```

ModeloNuevo = lm(price~wheelbase+horsepower, oPreciosAutos)
summary(ModeloNuevo)

##
## Call:
## lm(formula = price ~ wheelbase + horsepower, data = oPreciosAutos)
##

```

```
## 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 ***
## wheelbase    443.095     49.818   8.894 3.33e-16 ***
## horsepower   139.425      7.586  18.379 < 2e-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

inModeloNuevo = length(ModeloNuevo)

cat("Valor T limite: ", abs( qt(0.04/2,(inModeloNuevo-2) ) ))

## Valor T limite:  2.359315

print(" ")

## [1] " "

ModeloNuevo$coefficients

## (Intercept)  wheelbase  horsepower
## -44998.3110   443.0948   139.4247
```

Los valores quedaron más arriba que el valor limite de t a 0.04 asi que al final este seria el mejor modelo

PrecioHP = -44998.3110 + 139.4247HP PrecioWB = -44998.3110 + 443.0948WB

2.2.2.2 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)

β_0 tiene un valor T absoluto de 9.559 que es mayor al valor limite T que es de 2.35 y tambien tiene un valor P menor a 0.04, eso significa que si es significativo.

β_1 tiene un valor T absoluto de 8.894 que es mayor al valor limite de T que es de 2.35 y tambien tiene un valor P menor a 0.04, eso significa que si es significativo

β_2 tiene un valor T absoluto de 18.379 que es mayor al valor limite de T que es de 2.35 y tambien tiene un valor P menor a 0.04, eso significa que si es significativo

2.2.2.3 Indica cuál es el porcentaje de variación explicada por el modelo.

El porcentaje de variacion es igual a 75.07%

2.2.2.4 Dibuja el diagrama de dispersión de los datos por pares y la recta de mejor ajuste.

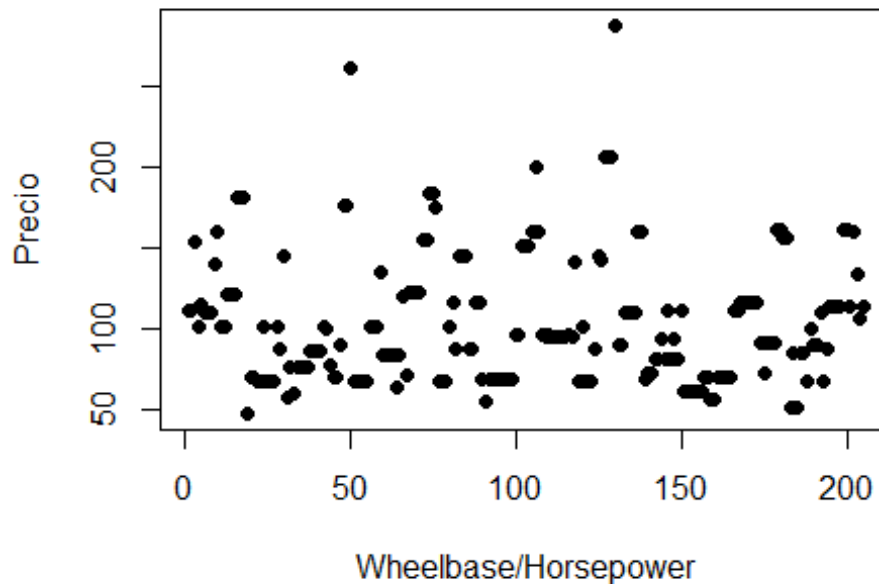
```
b0 = ModeloNuevo$coefficients[1]
b1 = ModeloNuevo$coefficients[2]
b2 = ModeloNuevo$coefficients[3]

Yhp = function(x){b0+b2*x}
Ywb = function(x){b0+b1*x}
colores = c("blue")
plot(oPreciosAutos$horsepower,oPreciosAutos$Precio, data=oPreciosAutos,
pch=19, ylab="Precio", xlab="Wheelbase/Horsepower", main="Relacion Precio vs
Wheelbase y Horsepower")

## Warning in plot.window(...): "data" is not a graphical parameter
## Warning in plot.xy(xy, type, ...): "data" is not a graphical parameter
## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
a
## graphical parameter
## Warning in axis(side = side, at = at, labels = labels, ...): "data" is not
a
## graphical parameter
## Warning in box(...): "data" is not a graphical parameter
## Warning in title(...): "data" is not a graphical parameter

x = seq(0,200,0.01)
lines(x, Yhp(x), col="blue", lwd=2)
```

Relacion Precio vs Wheelbase y Horsepower



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

De los analisis pude identificar cual era el mejor modelo en base a mis vairables y de ahi pude identificar que las variables más significativas son HorsePower y Wheelbase podria haber generado el modelo en base solo a HorsePower pero me parecio más interesante usar las dos variables.

2.3 Analiza la validez de los modelos propuestos:

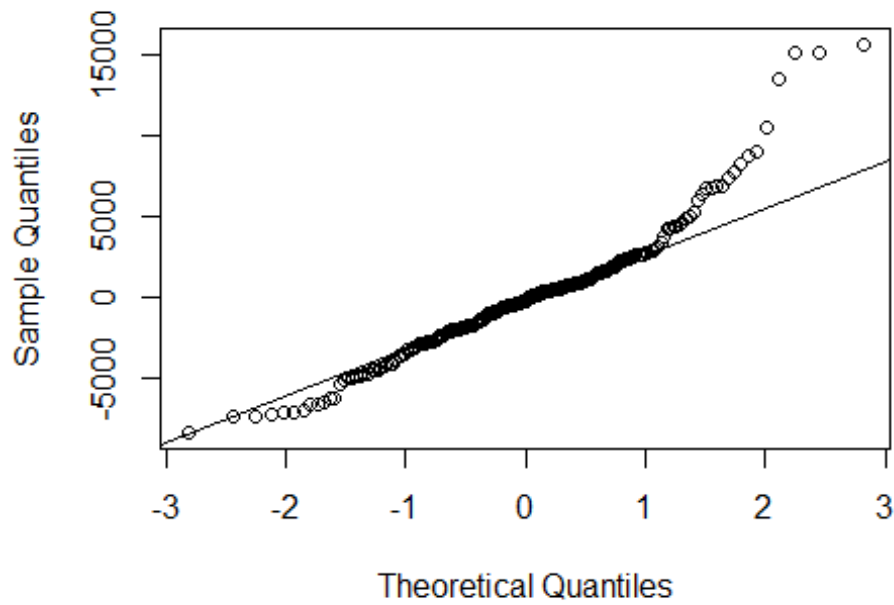
2.3.1 Normalidad de los residuos

```
library(nortest)
ad.test(ModeloNuevo$residuals)

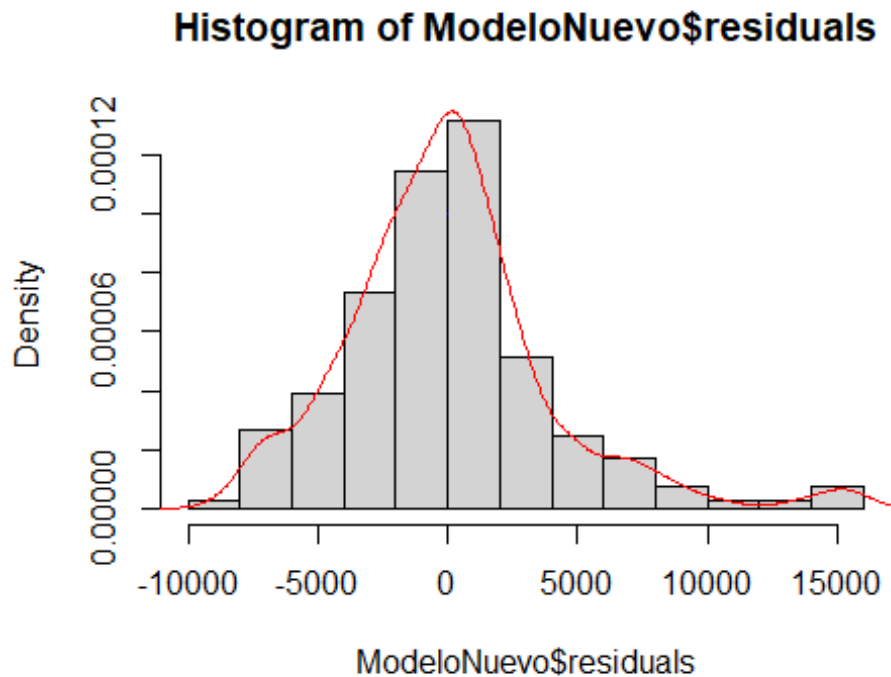
##
## Anderson-Darling normality test
##
## data:  ModeloNuevo$residuals
## A = 2.8064, p-value = 4.385e-07

qqnorm(ModeloNuevo$residuals)
qqline(ModeloNuevo$residuals)
```


Normal Q-Q Plot



```
hist(ModeloNuevo$residuals,freq=FALSE)
lines(density(ModeloNuevo$residual),col="red")
curve(dnorm(x,mean=mean(ModeloNuevo$residuals),sd=sd(ModeloNuevo$residuals)),
from=-
20, to=20, add=TRUE, col="blue",lwd=2)
```



2.3.2 Verificación de media cero

```
t.test(ModeloNuevo$residuals)
```

```
##  
## One Sample t-test  
##  
## data:  ModeloNuevo$residuals  
## t = -6.015e-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  
## -1.675679e-13
```

2.3.3 Homocedasticidad, linealidad e independencia

Homocedasticidad:

```
library(lmtest)  
  
## Loading required package: zoo  
  
##  
## Attaching package: 'zoo'
```

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

```
bptest(ModeloNuevo)
```

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

```
bgtest(ModeloNuevo)
```

```
##  
## Breusch-Godfrey test for serial correlation of order up to 1  
##  
## data: ModeloNuevo  
## LM test = 57.47, df = 1, p-value = 3.432e-14
```

Independencia:

```
dwtest(ModeloNuevo)
```

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

```
gqtest(ModeloNuevo)
```

```
##  
## Goldfeld-Quandt test  
##  
## data: ModeloNuevo  
## GQ = 0.74207, df1 = 100, df2 = 99, p-value = 0.9308  
## alternative hypothesis: variance increases from segment 1 to 2
```

Linealidad:

```
resettest(ModeloNuevo)
```

```
##  
## RESET test  
##  
## data: ModeloNuevo  
## RESET = 12.027, df1 = 2, df2 = 200, p-value = 1.169e-05
```

2.3.4 Interpreta cada uno de los analisis que realizaste

Ninguno de los modelos contiene Homocedasticidad, Independencia y Linealidad, ademas que la normalidad de residuos tiene un valor P menor a 0.04.

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

2.4.1 Concluye sobre el mejor modelo que encontraste y argumenta por qué es el mejor

El mejor modelo es el ModeloNuevo que genere ya que tiene más significancia que los otros modelos donde use fueltype, donde no tenia nada de correlacion, pero al usar las variables de HorsePower y Wheelbase

2.4.2 ¿Cuáles de las variables asignadas influyen en el precio del auto? ¿de qué manera lo hacen

Las variables que influyen son los valores de horsepower y wheelbase donde mientras más horsepower se tenga más aumenta el precio, igual que el wheelbase.

3. Intervalos de predicción y confianza

3.1 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:

3.1.1 Calcula los intervalos para la variable Y

```
Ip=predict(object=ModeloNuevo,interval="prediction",level=0.96)

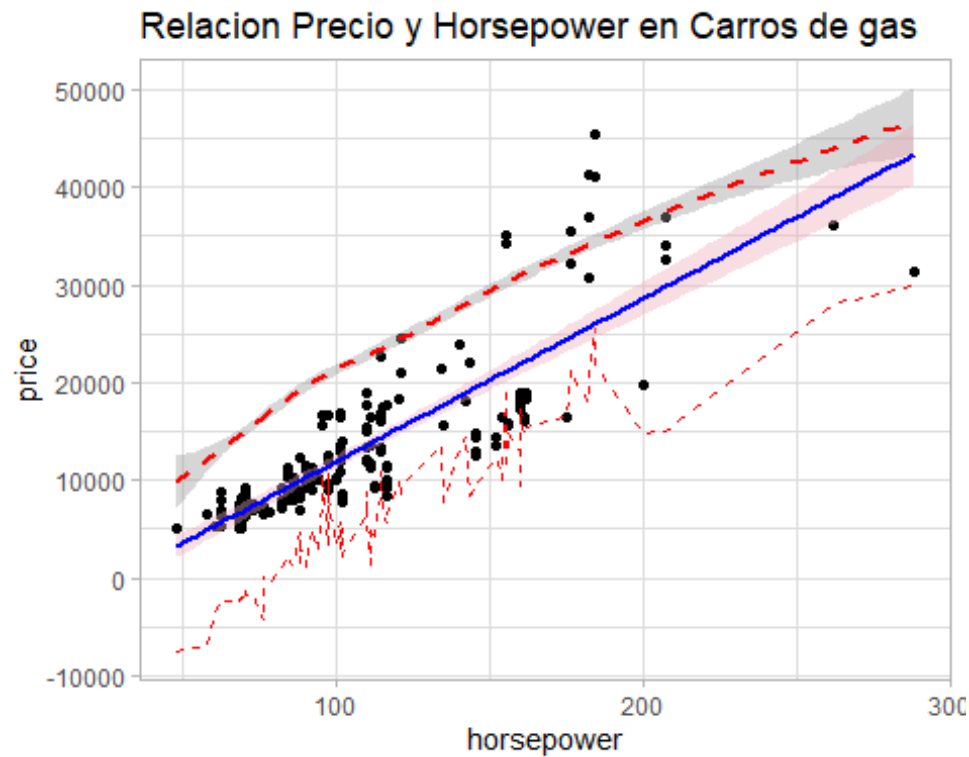
## Warning in predict.lm(object = ModeloNuevo, interval = "prediction", level
= 0.96): predictions on current data refer to _future_ responses

M2=cbind(oPreciosAutos,Ip)
M2g = subset(M2, fueltype == "gas")
M2d = subset(M2, fueltype == "diesel")
```

3.1.2 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.

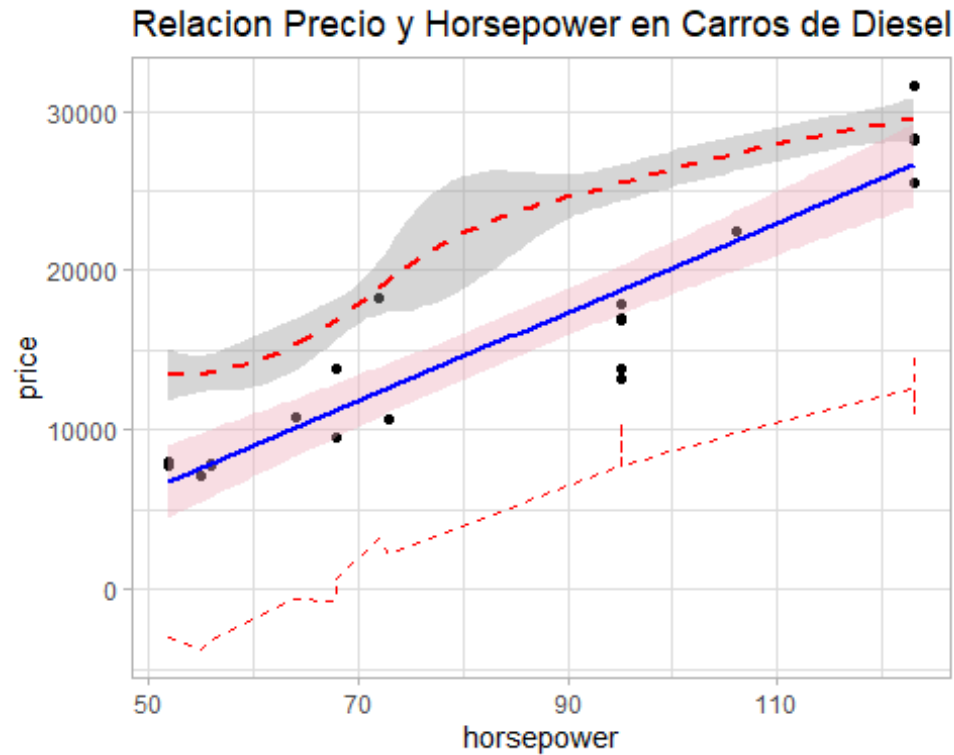
```
library(ggplot2)
ggplot(M2g,aes(x= horsepower ,y= price))+
  ggtitle("Relacion Precio y Horsepower en Carros de gas")+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_smooth(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.96, col="blue",
fill="pink2")+
  theme_light()

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



```
library(ggplot2)
ggplot(M2d,aes(x= horsepower ,y= price))+
  ggtitle("Relacion Precio y Horsepower en Carros de Diesel")+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_smooth(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.96, col="blue",
fill="pink2")+
  theme_light()

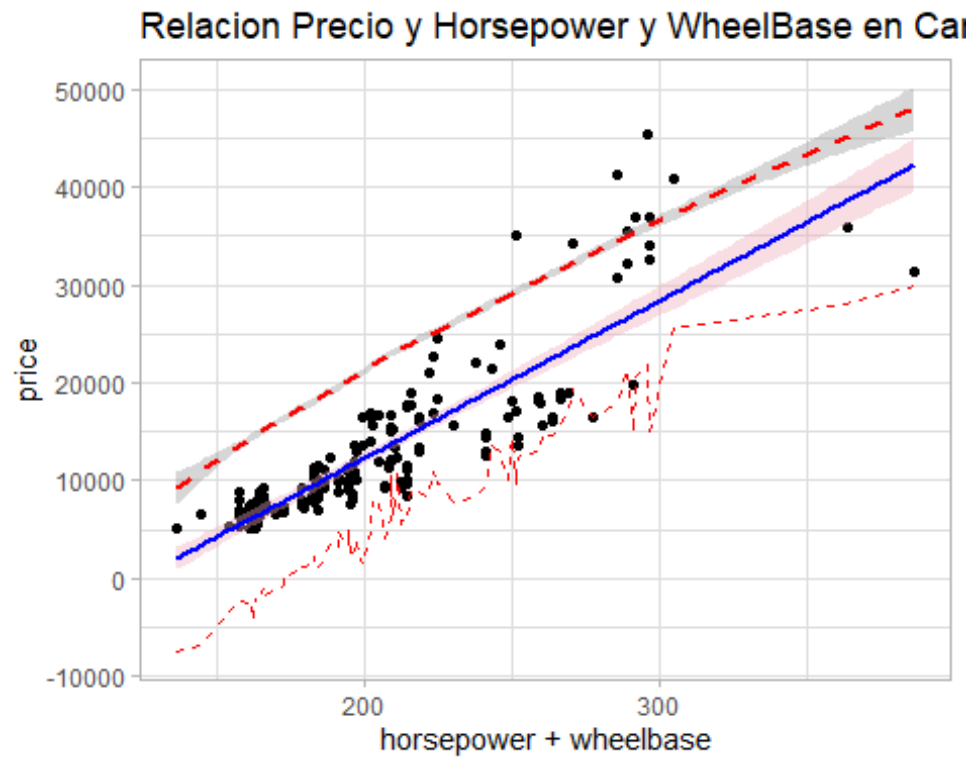
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



3.1.3 Grafica por pares de variables numéricas

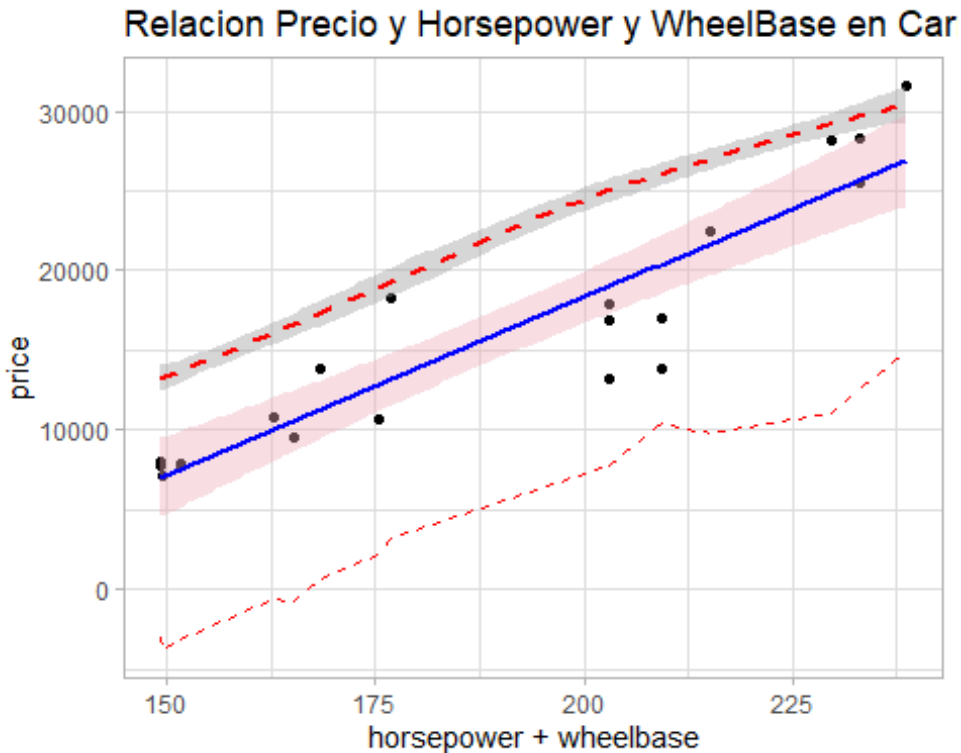
```
library(ggplot2)
ggplot(M2g, aes(x= horsepower+wheelbase ,y= price))+
  ggtitle("Relacion Precio y Horsepower y WheelBase en Carros de gas")+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_smooth(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.96, col="blue",
fill="pink2")+
  theme_light()

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



```
library(ggplot2)
ggplot(M2d,aes(x= horsepower+wheelbase ,y= price))+
  ggtitle("Relacion Precio y Horsepower y WheelBase en Carros de Diesel")+
  geom_point()+
  geom_line(aes(y=lwr), color="red", linetype="dashed")+
  geom_smooth(aes(y=upr), color="red", linetype="dashed")+
  geom_smooth(method=lm, formula=y~x, se=TRUE, level=0.96, col="blue",
fill="pink2")+
  theme_light()

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



3.2 Puedes

hacer el mismo análisis para otra categoría de la variable cualitativa, pero no es necesario, bastará con que justifiques la categoría seleccionada anteriormente. ### 3.3 Interpreta en el contexto del problema

En mis predicciones puedo identificar que la mayoría de datos están dentro de los intervalos de confianza, también se puede identificar la relación de horsepower y wheelbase con el precio donde mientras estos aumenten el precio también lo hace.

4. Más allá:

4.1 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?

Si propondría una agrupación de las variables de Horsepower y Wheelbase ya que tienen significancia con el valor de precio.

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

```
quantitative_data = oPreciosAutos[, c('wheelbase', 'carlength', 'carwidth',
'carheight', 'enginesize',
'stroke', 'compressionratio', 'horsepower',
'peakrpm',
'citympg', 'highwaympg', 'price')]
cor(quantitative_data)
```

	wheelbase	carlength	carwidth	carheight	enginesize
## wheelbase	1.0000000	0.8745875	0.7951436	0.58943476	0.56932868

## carlength	0.8745875	1.0000000	0.8411183	0.49102946	0.68335987
## carwidth	0.7951436	0.8411183	1.0000000	0.27921032	0.73543340
## carheight	0.5894348	0.4910295	0.2792103	1.00000000	0.06714874
## enginesize	0.5693287	0.6833599	0.7354334	0.06714874	1.00000000
## stroke	0.1609590	0.1295326	0.1829417	-0.05530667	0.20312859
## compressionratio	0.2497858	0.1584137	0.1811286	0.26121423	0.02897136
## horsepower	0.3532945	0.5526230	0.6407321	-0.10880206	0.80976865
## peakrpm	-0.3604687	-0.2872422	-0.2200123	-0.32041072	-0.24465983
## citympg	-0.4704136	-0.6709087	-0.6427043	-0.04863963	-0.65365792
## highwaympg	-0.5440819	-0.7046616	-0.6772179	-0.10735763	-0.67746991
## price	0.5778156	0.6829200	0.7593253	0.11933623	0.87414480
##	stroke	compressionratio	horsepower	peakrpm	
## wheelbase	0.16095905	0.24978585	0.35329448	-0.36046875	
## carlength	0.12953261	0.15841371	0.55262297	-0.28724220	
## carwidth	0.18294169	0.18112863	0.64073208	-0.22001230	
## carheight	-0.05530667	0.26121423	-0.10880206	-0.32041072	
## enginesize	0.20312859	0.02897136	0.80976865	-0.24465983	
## stroke	1.00000000	0.18611011	0.08093954	-0.06796375	
## compressionratio	0.18611011	1.00000000	-0.20432623	-0.43574051	
## horsepower	0.08093954	-0.20432623	1.00000000	0.13107251	
## peakrpm	-0.06796375	-0.43574051	0.13107251	1.00000000	
## citympg	-0.04214475	0.32470142	-0.80145618	-0.11354438	
## highwaympg	-0.04393093	0.26520139	-0.77054389	-0.05427481	
## price	0.07944308	0.06798351	0.80813882	-0.08526715	
##	citympg	highwaympg	price		
## wheelbase	-0.47041361	-0.54408192	0.57781560		
## carlength	-0.67090866	-0.70466160	0.68292002		
## carwidth	-0.64270434	-0.67721792	0.75932530		
## carheight	-0.04863963	-0.10735763	0.11933623		
## enginesize	-0.65365792	-0.67746991	0.87414480		
## stroke	-0.04214475	-0.04393093	0.07944308		
## compressionratio	0.32470142	0.26520139	0.06798351		
## horsepower	-0.80145618	-0.77054389	0.80813882		
## peakrpm	-0.11354438	-0.05427481	-0.08526715		
## citympg	1.00000000	0.97133704	-0.68575134		
## highwaympg	0.97133704	1.00000000	-0.69759909		
## price	-0.68575134	-0.69759909	1.00000000		

Las variables que se deberian agrupar son horsepower y enginesize ya que son las variables que más correlacion tienen con el precio.