

# Cluster coches del Jefe

Sergio Casares

23/11/2020

## Contents

<b>1. Introducción al trabajo</b>	<b>2</b>
<b>2. Selección a priori de variables y número de clústers</b>	<b>2</b>
<b>3. Descripción de las variables seleccionadas</b>	<b>3</b>
PVP (Precio de venta al Público) . . . . .	3
CC (Centímetros Cúbicos) . . . . .	4
Velocidad . . . . .	4
Peso . . . . .	5
Cilindros (Número de) . . . . .	6
<b>4. Bibliografía</b>	<b>7</b>

# 1. Introducción al trabajo

Asignar de la forma más eficiente y consistente los distintos vehículos, agrupándolos en virtud de las características que ud. considere oportunas.

El objetivo de este trabajo será el de satisfacer la necesidad de nuestro jefe de ‘clusterizar’ los coches adquiridos en un número determinado de grupos con el fin de que estos puedan ser guardados en hasta un máximo de 10 garajes que tiene nuestro jefe.

## 2. Selección a priori de variables y número de clústers

### Selección de variables:

A priori (y con base en el *análisis de datos*) eliminamos las siguientes variables:

- Aceleración (acelerac) -> Por su alta proporción de valores nulos 46 / 125 (38 %), además de su alta correlación con *potencia* y *velocidad* (por encima de 0.8).
- Tiempo de aceleración (acel2) -> El tiempo de aceleración es una variable que aporta poca información a la hora de poder agrupar los coches (solo 3 coches tienen un valor de 1 y 122 tienen un valor de 2), por lo que optamos por eliminar dicha variable del análisis.
- Potencia (potencia)-> Realizando una matriz de correlaciones, observamos que la correlación entre **velocidad** y **potencia** es de 0.85, lo que indica que la relación lineal entre ambas variables es muy alta y debido a su gran similitud (y que puede generar multicolinealidad), se decide a eliminar la **potencia** (ya que también presenta una correlación alta con el **precio**, **cilindradas**, **centímetros cúbicos** y **consumo urbano**), todas por encima del 0.7 de correlación.

A priori (y con base en el *análisis financiero y de costes*) eliminamos las siguientes variables:

La colección se centra en todo terrenos clásicos (lo que indica que su finalidad es más de coleccionista que de uso personal y diario), desconocemos la finalidad de nuestro jefe (coleccionar para tenerlos almacenados en sus garajes o a lo mejor tiene la visión futura de venderlos a un mayor precio (a modo de inversión a largo plazo)).

- “cons90”, “cons120”, “consurb” -> entiendo que el consumo en gasolina no debe ser un factor discriminante a tener en cuenta a la hora de segmentar ni a la hora de su potencial venta (a nadie le interesa lo que consuma el coche, el que compra este tipo de coches no se fija en lo que le costará la gasolina).
- “plazas” -> por el mismo motivo que el anterior, es un coche de coleccionista, no de uso diario, el coleccionista se fija en el coche, no en el número de plazas que tenga mi todo-terreno clásico.
- “rpm” -> existen variables que se valoran más a la hora de comprar coches de colección, como los cc y el número de cilindros del motor.

### Selección final de variables:

Decidimos seleccionar con su precio de venta al público (pvp), el número de cilindros (cilindro), su velocidad máxima (velocida), su peso (peso) y los centímetros cúbicos del motor (cc).

### Selección del número de clústers:

Aunque el análisis el análisis debe realizarse por similitudes entre observaciones, pero si tenemos en cuenta la estrategia y el enfoque económico a la hora de hacer el reparto de los coches, deberíamos tener las siguientes condiciones en cuenta:

- El peso del vehículo afecta a la hora realizar el transporte (unos coches necesitarán un transporte con más cuidado o con camiones más pesados, lo que encarece el desplazamiento).
- Hay que tener en cuenta el número de coches que puede desplazar el camión (entre 6 y 10), para que cada clúster tenga un número múltiplo de coches y así evitar desplazamientos del camión ineficientes (que no esté lleno).

Por lo tanto, el número de clúster será aquél que minimice el coste de transporte, favoreciendo agrupaciones donde el tamaño de cada grupo pueda hacer que los camiones de desplazamientos vayan lo suficientemente llenos.

### 3. Descripción de las variables seleccionadas

#### PVP (Precio de venta al Público)

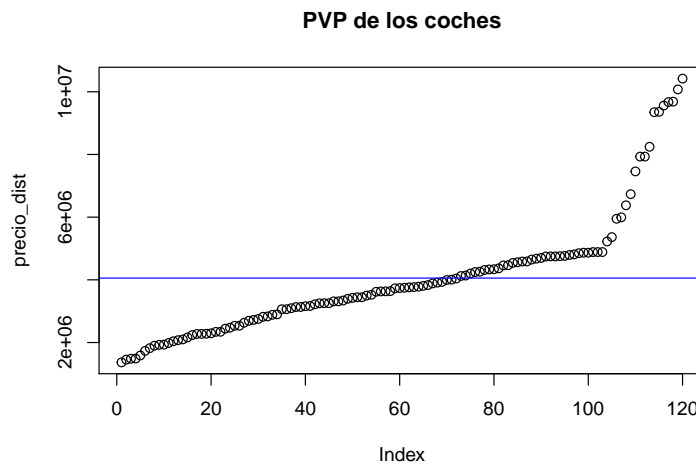
Precio es la cantidad monetaria por el que un comprador estaría dispuesto a adquirir la propiedad del vehículo.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1367000	2803547	3742322	4056164	4710500	10419200

De este 'summary', obtenemos que el coche con menor valoración es de 1.367.000 pesetas (8.200 €) y el más caro de 10.419.200 pesetas (62.620 €).

La media de los coches es de 4.056.164 pesetas (24.435 €), lo que podemos intuir de que la suma total de los coches asciende a más de 3 millones de euros.

A continuación se expondrán los precios de los coches de manera ascendente, además de representar con una línea horizontal azul la media del precio de los coches.



De esta gráfico podemos observar como hay unos 10-15 cuyo precio se aleja en gran medida respecto del resto, lo cual nos hace observar dos grupos claramente diferenciados. Futuras agrupaciones podrían ir encaminadas en función del centil al que pertenezcan.

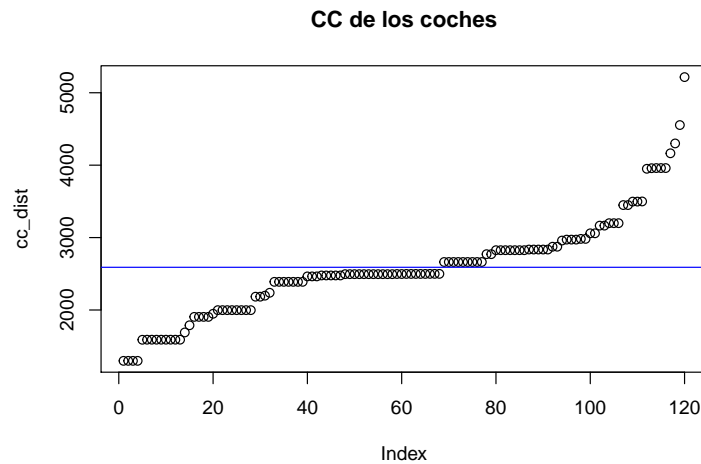
Si quisieramos realizar una agrupación eficiente desde el punto de vista económico, se podrían agrupar en 6 grupos de 20 coches (en función de su precio) con el fin de evitar el menor número de desplazamiento de camiones que transporten esos coches (cada camión puede transportar hasta 10 coches), lo que implicaría hacer dos viajes (llenado el camión) y así evitar un uso eficiente de este servicio de transporte.

## CC (Centímetros Cúbicos)

La cilindrada es el volumen geométrico ocupado por el conjunto de pistones desde el punto muerto inferior (PMI) hasta el más alto (PMS), también llamado punto muerto superior. La cilindrada da una buena medida de la capacidad de trabajo que puede tener un motor.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1298	2194	2498	2590	2835	5216

El valor mínimo es de 1298, el máximo de 5216 y su media es de 2498 cc.



Des esta representación podemos observar una clara diferenciación entre los distintos vehículos, podemos observar diferentes grupos:

- Unos 25 vehículos que tienen menos de 2000 cc
- Unos 40 coches entre 2000 cc y 3000 cc
- Unos 25 coches entre 3000 cc y 4000 cc
- Unos 10 coches por encima de los 4000 cc

Teniendo en cuenta la futura segmentación de los vehículos, podríamos repartir generar alrededor de 5 grupos más o menos homogéneos de entre 20 y 25 vehículos por grupo. Para evitar cierta descompensación, tendría sentido dividir el grupo más grande (coches entre 3000 cc y 4000 cc) en dos para que haya 20 vehículos por grupo.

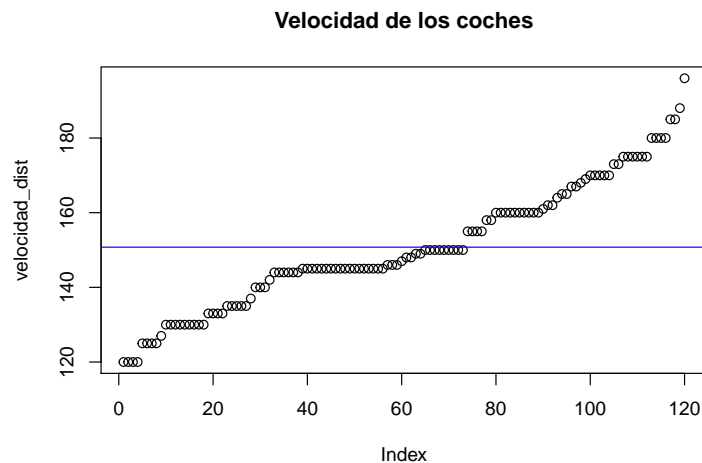
Agrupando los vehículos de esta manera, somos menos eficientes que agrupándolos por precio, pero podría servir como una manera bastante objetiva y equitativa de agrupar los vehículos.

## Velocidad

La velocidad es una magnitud física que expresa la relación entre el espacio recorrido por un objeto, el tiempo empleado para ello y su dirección.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	120.0	140.0	147.5	150.8	161.2	196.0

El valor mínimo es de 120 km/h, el máximo de 196 km/h y su media es de 147,5 km/h.



Se observa una clara agrupación de los coches en el tramo de los 140 y 160 kilómetros/hora. Algunos valores se encuentran por debajo de 130 y otros por encima de 180.

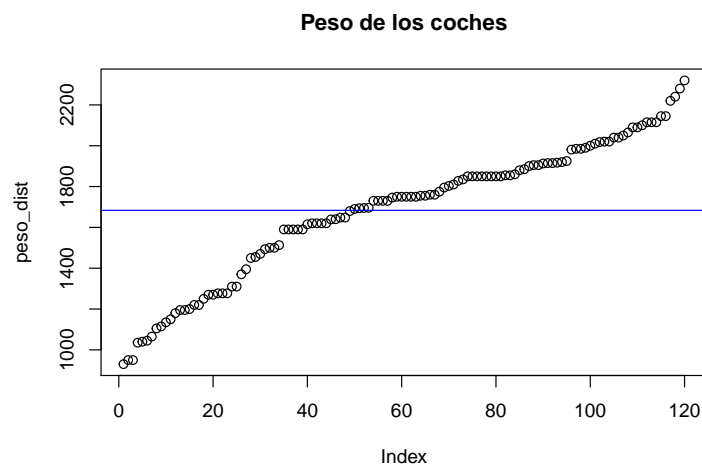
En este caso, su distribución podría resultar más complicada.

## Peso

Masa de los coches, medidos en kilogramos.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	930	1487	1750	1683	1914	2320

El valor mínimo es de 930 kilos, el máximo de 1683 kilos y su media es de 2320 kilos.



Se observa una clara agrupación de los coches en el tramo de los 1600 y 1900 kilos. Algunos valores se encuentran por debajo de 1300 y otros por encima de 2100.

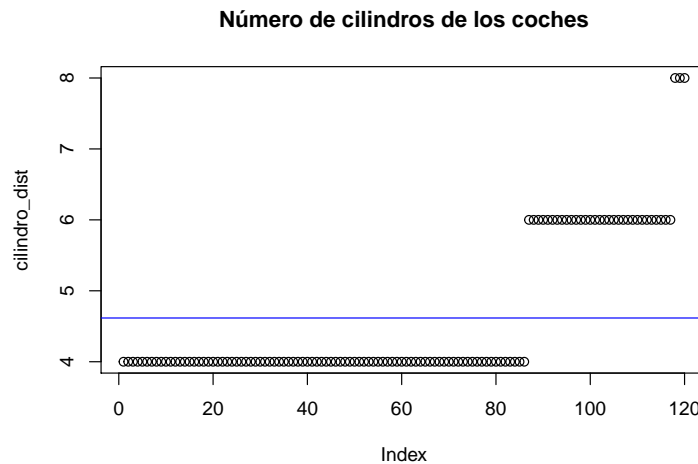
En este caso, se pueden observar a simple vista entre 5 y 6 tramos en los que se podrían agrupar los coches en función de su peso, además, el peso también influye en el precio del transporte de dichos de vehículos, aunque la agrupación pueda ser clara, habría que estudiar el coste económico de transportar los coches según su peso.

## Cilindros (Número de)

Número de cilindros que presenta un coche.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	4.000	4.000	4.000	4.617	6.000	8.000

El valor mínimo es de 4 cilindros, el máximo de 8 mediana es de 4.



Se pueden diferenciar tres grupos en función del número de cilindros (4, 6 y 8), apenas hay vehículos con 8 cilindros, lo que nos indica que quizás deban agruparse con los de 6 por su proximidad.

Tenemos dos grupos bien diferenciados (4 por un lado y 6 y 8 por otro), puede ser una primera aproximación para poder agrupar los vehículos, es decir, haciendo esta división, se podría después utilizar otra variable que pudiera ser capaz de dividir y agrupar cada uno de los dos grupos creados.

## 4. Bibliografía

Tabla de precios de los vehículos realizada por Hacienda:

- <https://www.boe.es/boe/dias/2019/12/31/pdfs/BOE-A-2019-18749.pdf>.

Factores de valoración del mercado de coches clásicos:

- <https://dialnet.unirioja.es/descarga/articulo/2487623.pdf>

Tasación de vehículos normales y clásicos:

- <https://www.motor.es/noticias/tasar-coche-clasico-202067684.html>
- <https://www.motor.es/noticias/claves-tasar-coche-201740935.html>
- <https://www.libertaddigital.com/deportes/motor/2016-01-28/por-que-es-importante-tasar-el-coche-1276566622/>