

Análisis de MTCARS

Obtener datos:

```
> library (datasets)      # necesario sólo si usas Rcmdr
> ? mtcars                 # descripción
> M <- mtcars              # para ahorrar pulsaciones
```

Las variables de este dataframe son:

mpg : millas por galón

cyl : número de cilindros

dis : cubicaje en pulgadas cúbicas (desplazamiento de los cilindros)

hp : potencia en caballos de vapor

drat : razón entre ejes (número de revoluciones de la transmisión por cada vuelta de rueda)

wt : peso en miles de libras

qsec : segundos para recorrer un cuarto de milla

vs : tipo de cilindros (en V; en línea o Serie)

am : tipo de transmisión (automática; manual)

gear : número de marchas hacia adelante

carb : número de carburadores

1. Transformar millas por galón en litros por 100 km

Consideraremos:

- 1 milla $\simeq$ 1,609 km
- 1 galón $\simeq$ 3,785 litros

Cadena de trasformaciones:

```
> km.galon      <-      M$mpg * 1.609      # kilómetros por galón
> km.litro      <-      km.galon / 3.785    # kilómetros por litro
> litros.km     <-      1 / km.litro        # litros por kilómetro
> litros.100km  <-      litros.km * 100     # litros por cada 100 km
```

o bien directamente

```
> M$L100k <- 100 * 3.785 / (1.61 * M$mpg)
```

Siempre conviene comprobar los resultados:

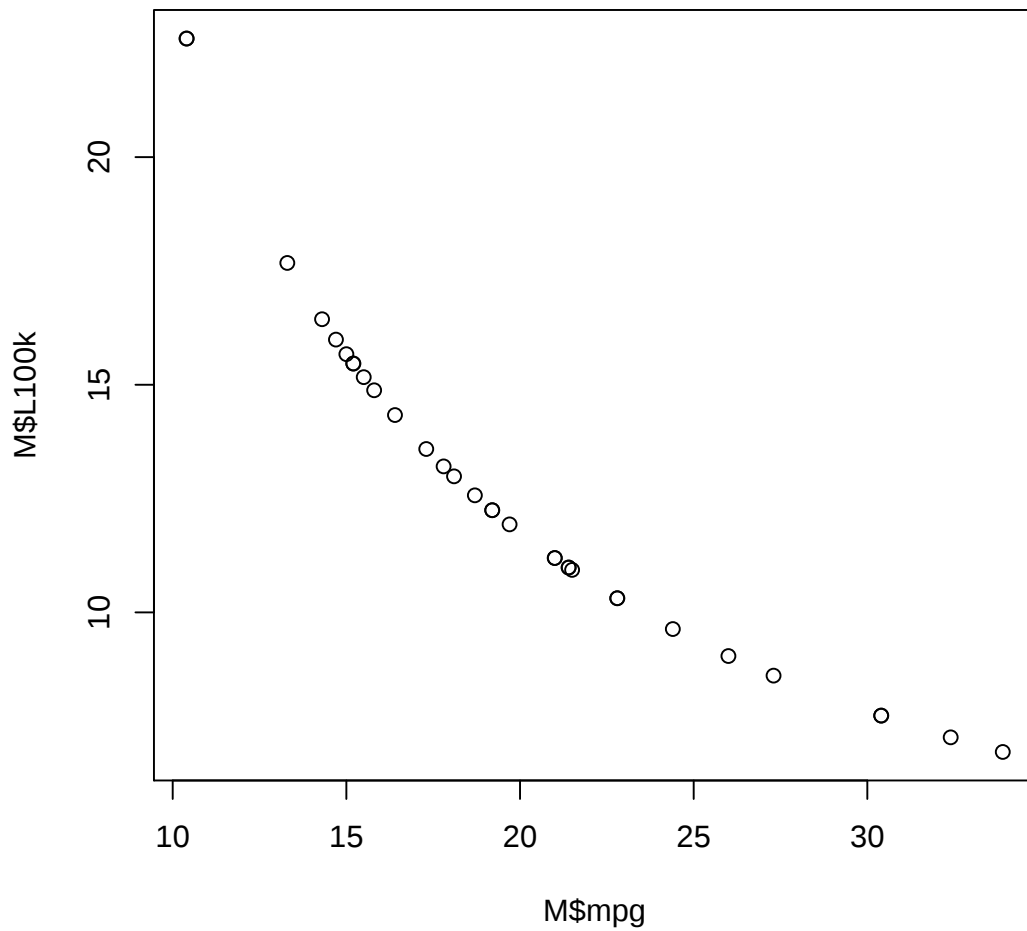
```
> head      (M[, c("mpg", "L100k")])
```

	mpg	L100k
Mazda RX4	21.0	11.2
Mazda RX4 Wag	21.0	11.2
Datsun 710	22.8	10.3
Hornet 4 Drive	21.4	11.0
Hornet Sportabout	18.7	12.6
Valiant	18.1	13.0

```
> summary (M[, c("mpg", "L100k")])
```

mpg	L100k
Min. :10.4	Min. : 6.93
1st Qu.:15.4	1st Qu.:10.31
Median :19.2	Median :12.24
Mean :20.1	Mean :12.75
3rd Qu.:22.8	3rd Qu.:15.24
Max. :33.9	Max. :22.61

```
> plot (M$mpg, M$L100k)
```



2. Transformar “segundos hasta 1/4 de milla” en “segundos de 0 a 100 km/h”

Suponiendo que se parte del reposo y que el movimiento es uniformemente acelerado, las fórmulas que relacionan el espacio recorrido, la velocidad alcanzada y el tiempo empleado son:

$$\begin{aligned}e(t) &= \frac{1}{2} a t^2 \\v(t) &= a t\end{aligned}$$

Aplicaremos estas expresiones a los datos de `mtcars`. Debemos tener cuidado sobre qué unidades usamos en las fórmulas:

1. espacio

$$\begin{aligned}\frac{1}{4} \text{ de milla} &= \frac{1,609}{4} \text{ km} = 0,402 \text{ km} = e(\text{qsec}) = \frac{1}{2} a \text{qsec}^2 \\ \Rightarrow a &= \frac{2 \times 0,402 \text{ km}}{\text{qsec}^2} = \frac{0,804 \text{ km}}{\text{qsec}^2}\end{aligned}$$

Esa aceleración a está expresada en km/s^2 (kilómetros por segundo cada segundo), por lo que vamos a expresar la velocidad en km/s (kilómetros por segundo) en vez de km/h .

2. velocidad

$$\begin{aligned}100 \text{ km/h} &= \frac{100}{3600} \text{ km/s} = \frac{1}{36} \text{ km/s} = \frac{1 \text{ km}}{36 \text{ s}} = v(t) = a t = \frac{0,804 \text{ km}}{\text{qsec}^2} t \\ \Rightarrow t &= \frac{1 \text{ km} \cdot \text{qsec}^2}{0,804 \text{ km} \cdot 36 \text{ s}} = \frac{\text{qsec}^2}{28,944 \text{ s}}\end{aligned}$$

Cálculo y comprobación:

```
> M$T100k <- M$qsec^2 / 28.944      # tiempo en segundos de 0 a 100 km/h
> head (M[, c("qsec", "T100k")])
```

	qsec	T100k
Mazda RX4	16.5	9.36
Mazda RX4 Wag	17.0	10.01
Datsun 710	18.6	11.97
Hornet 4 Drive	19.4	13.06
Hornet Sportabout	17.0	10.01
Valiant	20.2	14.13

```
> summary (M[, c("qsec", "T100k")])
```

	qsec	T100k
Min.	:14.5	Min. : 7.26
1st Qu.:	16.9	1st Qu.: 9.86
Median	:17.7	Median :10.84
Mean	:17.9	Mean :11.11
3rd Qu.:	18.9	3rd Qu.:12.34
Max.	:22.9	Max. :18.12

Un promedio de once segundos en pasar de 0 a 100 es razonable.

3. Transformar otras variables

Ya que hemos hecho la transformación de `mpg` y `qsec` a unidades europeas, haremos lo mismo con

disp (cubicaje): pulgadas cúbicas $\rightarrow$ centímetros cúbicos

wt (peso): miles de libras $\rightarrow$ toneladas

```
> M$cubi <- M$disp * 2.54^3
> M$peso <- M$wt * 0.4536
> ## nos quedamos sólo con variables en unidades europeas:
> M <- M [, c ("L100k", "cyl", "cubi", "hp", "drat", "peso",
+             "T100k", "vs", "am", "gear", "carb")]
> round (head (M), 2)
```

	L100k	cyl	cubi	hp	drat	peso	T100k	vs	am	gear	carb
Mazda RX4	11.2	6	2622	110	3.90	1.19	9.36	0	1	4	4
Mazda RX4 Wag	11.2	6	2622	110	3.90	1.30	10.01	0	1	4	4
Datsun 710	10.3	4	1770	93	3.85	1.05	11.97	1	1	4	1
Hornet 4 Drive	11.0	6	4228	110	3.08	1.46	13.06	1	0	3	1
Hornet Sportabout	12.6	8	5899	175	3.15	1.56	10.01	0	0	3	2
Valiant	13.0	6	3687	105	2.76	1.57	14.13	1	0	3	1

4. Descripción univariante

```
> num.categ <- sapply (M, function (var) length (unique (var)))
> continuas <- num.categ > 10
> pdf ("graf1.pdf")
> for (i in which (continuas))
+   hist (M[,i], main=names(M)[i])
> for (i in which (!continuas))
+   barplot (table (M[,i]), main=names(M)[i])
> dev.off ()
```

null device

1

```
> options (digits = 3)
> RcmdrMisc::numSummary (M)
```

	mean	sd	IQR	0%	25%	50%	75%	100%	n
L100k	12.748	3.861	4.931	6.935	10.31	12.24	15.24	22.61	32
cyl	6.188	1.786	4.000	4.000	4.00	6.00	8.00	8.00	32
cubi	3780.854	2030.991	3362.216	1165.120	1979.97	3216.78	5342.18	7734.69	32
hp	146.688	68.563	83.500	52.000	96.50	123.00	180.00	335.00	32
drat	3.597	0.535	0.840	2.760	3.08	3.70	3.92	4.93	32
peso	1.459	0.444	0.467	0.686	1.17	1.51	1.64	2.46	32
T100k	11.114	2.252	2.482	7.264	9.86	10.84	12.34	18.12	32
vs	0.438	0.504	1.000	0.000	0.00	0.00	1.00	1.00	32
am	0.406	0.499	1.000	0.000	0.00	0.00	1.00	1.00	32
gear	3.688	0.738	1.000	3.000	3.00	4.00	4.00	5.00	32
carb	2.812	1.615	2.000	1.000	2.00	2.00	4.00	8.00	32

5. Relación de L100k con el resto

```
> pdf ("graf2.pdf")
> for (i in which (continuas))
+   plot (M$L100k ~ M[,i], main=names(M)[i])
> for (i in which (!continuas))
+   boxplot (as.formula (paste ("L100k ~", names(M)[i])), M)
> dev.off ()
```

null device

1

```
> cor (M) ["L100k", ]
```

L100k	cyl	cubi	hp	drat	peso	T100k	vs	am	gear	carb
1.000	0.814	0.880	0.763	-0.638	0.890	-0.384	-0.640	-0.540	-0.479	0.526

6. Modelo para predecir L100k

```
> R2 <- sapply (setdiff (names (M), "L100k"),  
+             function (i)  
+                 summary (lm (M$L100k ~ M[[i]])) $ r.squared)  
> mejor <- names (which.max (R2))  
> summary (lm (paste ("L100k ~", mejor), M))
```

Call:

```
lm(formula = paste("L100k ~", mejor), data = M)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.228	-1.041	0.221	1.066	2.741

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.450	1.104	1.31	0.2
peso	7.742	0.725	10.68	9.6e-12

Residual standard error: 1.79 on 30 degrees of freedom

Multiple R-squared: 0.792, Adjusted R-squared: 0.785

F-statistic: 114 on 1 and 30 DF, p-value: 9.57e-12