

# PRÁCTICA 1 ESTADÍSTICA

Análisis de una variable

Alba Calvo Herrero

1º GII

6 de octubre de 2019

## ÍNDICE

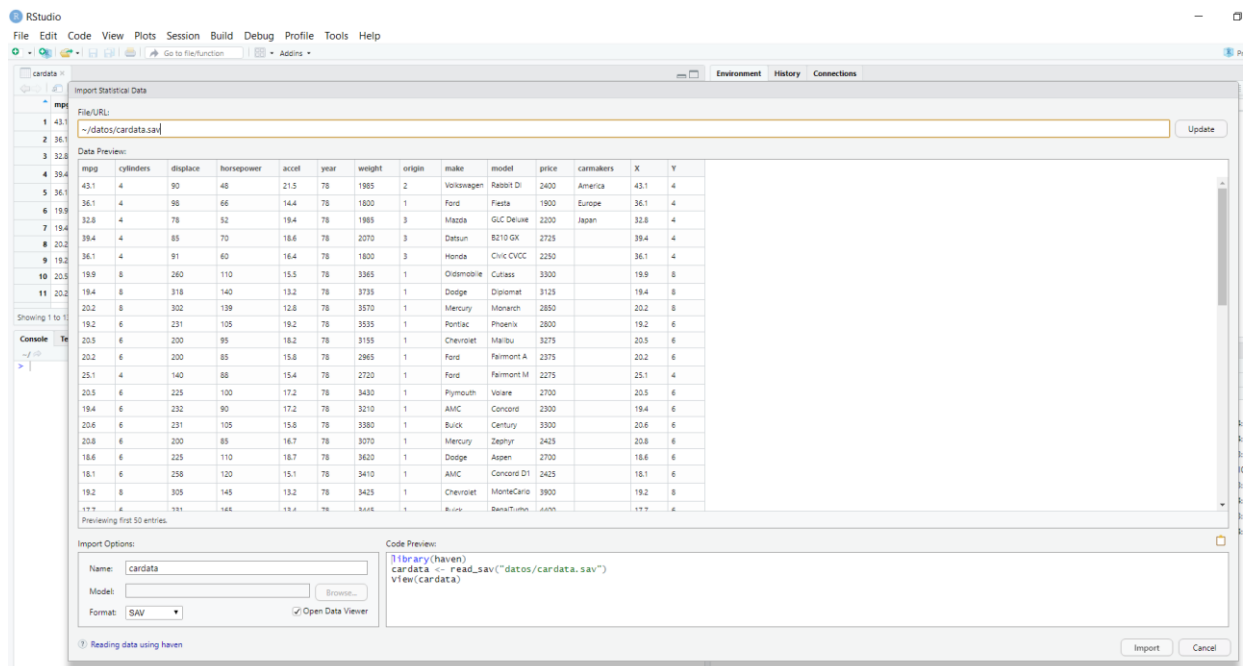
1. Media, mediana, los cuartiles y los percentiles 47, 54 y 82
2. Desviación típica, rango y varianza
3. Histograma de variable Consumo
4. Cuadro de frecuencias para los datos agrupados en 10 clases
5. Cuadro de frecuencias para los datos agrupados en 12 clases

## 1. Media, mediana, los cuartiles y los percentiles 47, 54 y 82

El ejercicio pide abrir el fichero *cardata.sav* y realizar un análisis al consumo de todos los automóviles, es decir, a la primera columna llamada *mpg*. Para ello nos pide calcular la media, la mediana, los cuartiles y los percentiles.

Primero de todo, tenemos que dirigirnos a la ventana inferior derecha a la pestaña "files". Allí buscamos el archivo *cardata* lo clicamos y le damos a importar cambiando *haven* por *foreign*, tal y como aparece en la captura.

```
> library(foreign)
> cardata <- read_sav("datos/cardata.sav")
> view(cardata)
```



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

cardata Rhistory

Filter

	mpg	cylinders	displace	horsepower	accel	year	weight	origin	make	model	price	carmakers	X	Y
1	43.1	4	90	48	21.5	78	1985	2	Volkswagen	Rabbit DI	2400	America	43.1	4
2	36.1	4	98	66	14.4	78	1800	1	Ford	Fiesta	1900	Europe	36.1	4
3	32.8	4	78	52	19.4	78	1985	3	Mazda	GLC Deluxe	2200	Japan	32.8	4
4	39.4	4	85	70	18.6	78	2070	3	Datsun	B210 GX	2725		39.4	4
5	36.1	4	91	60	16.4	78	1800	3	Honda	Civic CVCC	2250		36.1	4
6	19.9	8	260	110	15.5	78	3365	1	Oldsmobile	Cutlass	3300		19.9	8
7	19.4	8	318	140	13.2	78	3735	1	Dodge	Diplomat	3125		19.4	8
8	20.2	8	302	139	12.8	78	3570	1	Mercury	Monarch	2850		20.2	8
9	19.2	6	231	105	19.2	78	3535	1	Pontiac	Phoenix	2800		19.2	6
10	20.5	6	200	95	18.2	78	3155	1	Chevrolet	Malibu	3275		20.5	6
11	20.2	6	200	85	15.8	78	2965	1	Ford	Fairmont A	2375		20.2	6
12	25.1	4	140	88	15.4	78	2720	1	Ford	Fairmont M	2275		25.1	4
13	20.5	6	225	100	17.2	78	3430	1	Plymouth	Volare	2700		20.5	6
14	19.4	6	232	90	17.2	78	3210	1	AMC	Concord	2300		19.4	6
15	20.6	6	231	105	15.8	78	3380	1	Buick	Century	3300		20.6	6
16	20.8	6	200	85	16.7	78	3070	1	Mercury	Zephyr	2425		20.8	6
17	18.6	6	225	110	18.7	78	3620	1	Dodge	Aspen	2700		18.6	6

Showing 1 to 20 of 155 entries, 14 total columns

A continuación, como me interesa trabajar con la columna "mpg" escribo en la esquina inferior izquierda: (`mpg=cardata$mpg`) , simplemente por comodidad y para escribir menos.

Otra cosa a tener en cuenta es que algunos de los valores de este vector son nulos, por lo que, si no los eliminamos nos puede dar error. Para ello, utilizo el comando ***na.omit***.

A partir de aquí, ya podemos hacer los cálculos que se nos piden.

Estos son los comandos para cada uno de los parámetros estadísticos pedidos en este apartado:

- Media: ***mean(mpg)***
- Mediana: ***median(mpg)***
- Cuartiles: ***quantile(mpg)***
- Percentiles: ***quantile(mpg, c(0.47, 0.54, 0.82))***

## 2. Desviación típica, rango y varianza

Para hallar estos parámetros usamos las siguientes funciones:

- Varianza: ***var(mpg)***
- Desviación típica: ***sd(mpg)***
- Rango: ***IQR(mpg)***

Los resultados tanto del apartado uno como del dos se encuentran en las siguientes captura:

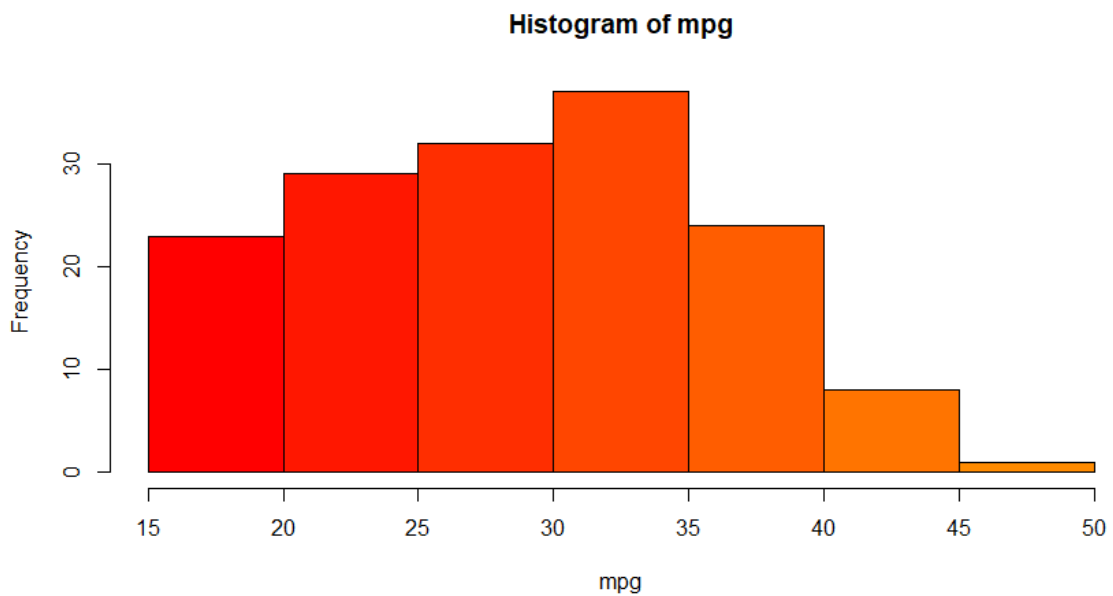
```
[1] 17
> mpg=na.omit(mpg)
> (mediampg=mean(mpg))
[1] 28.79351
> (medianampg=median(mpg))
[1] 28.9
> quantile(mpg)
      0%      25%      50%      75%     100%
15.500 22.550 28.900 34.275 46.600
> quantile(mpg, c(0.47,0.54,0.82))
47% 54% 82%
 28  30  36
```

```
28  30  36
> (varianzampg=var(mpg))
[1] 54.42323
> (desviaciontipmpg=sd(mpg))
[1] 7.37721
> IQR(mpg)
[1] 11.725
> |
```

### 3. Histograma

Utilizamos la función **hist(mpg)**. En este caso he usado además **col=heat.colors(15))** para rellenar las barras de color.

```
[1] 11.725  
> hist(mpg, col=heat.colors(15))  
>
```



#### 4. Cuadro de frecuencias para los datos agrupados en 10 clases

En primer lugar, declaramos  $k$  (el número de intervalos) que en este caso valdrá 10. Después, calculamos la amplitud, que llamaremos  $A$ , dividiendo el rango de los datos entre el número de intervalos, redondeando por exceso.

Calculamos los extremos de los intervalos  $L1$  y  $L$ .

A continuación, determinamos las marcas de clase con su correspondiente fórmula.

Ahora reagrupamos los datos codificándolos en un factor llamado *mpgagrupada* cuyos niveles son las clases que hemos creado y asignando cada dato a su clase. Para ello se utiliza la función *cut*.

Por último, usamos la función de la tabla de frecuencias *fabs=table(mpgagrupada)*.

```
> #TABLA DE FRECUENCIAS
> k=10 #Número de intervalos pedido
> A=diff(range(mpg))/k #cálculo amplitud
> L1=min(mpg)
> L=L1+A*(0:k)
> marcas=(L[0:k]+L[1:k+1])/2
> mpgagrupada=cut(mpg, breaks = L, right = FALSE)
> fabs= table(mpgagrupada)
> fabs
mpgagrupada
[15.5,18.6) [18.6,21.7) [21.7,24.8) [24.8,27.9) [27.9,31.1) [31.1,34.2) [34.2,37.3) [37.3,40.4) [40.4,43.5) [43.5,46.6)
      15         21         15         20         19         24         21         10          5          3
```

## 5. Cuadro de frecuencias para los datos agrupados en 12 clases

En este apartado tenemos que repetir los pasos del ejercicio anterior pero cambiando el valor de k por 12.

Estos son los resultados:

```
> k=12 #Número de intervalos pedido
> A=diff(range(mpg))/k      #Cálculo amplitud
> L1=min(mpg)
> L=L1+A*(0:k)
> marcas=(L[0:k]+L[1:k+1])/2
> mpgagrupada=cut(mpg, breaks = L, right = FALSE)
> fabs= table(mpgagrupada)
> fabs
mpgagrupada
[15.5,18.1) [18.1,20.7) [20.7,23.3) [23.3,25.9) [25.9,28.5) [28.5,31.1) [31.1,33.6) [33.6,36.2) [36.2,38.8) [38.8,41.4)
      10      22      10      14      20      14      18      22      12      5
 [41.4,44) [44,46.6)
       4       2
> |
```