

Práctica 1

Carlos Javier Hellín Asensio

Se pide abrir el fichero y realizar un análisis del consumo de todos los automóviles, calculando para dicha variable:

a) Media, mediana, moda, los cuartiles y los percentiles 47, 54 y 82

```
> mean(cardata$mpg, na.rm = TRUE)
```

```
[1] 28.79351
```

```
> median(cardata$mpg, na.rm = TRUE)
```

```
[1] 28.9
```

```
> table(cardata$mpg)
```

```
15.5 16.2 16.5 16.9 17 17.5 17.6 17.7 18.1 18.2 18.5 18.6 19.1 19.2 19.4
19.8 19.9 20.2 20.3 20.5 20.6
1 1 1 1 2 1 2 1 2 1 1 1 1 3 2
1 1 4 1 2 2
20.8 21.1 21.5 21.6 22 22.3 22.4 23 23.2 23.5 23.6 23.7 23.8 23.9 24
24.2 24.3 25 25.1 25.4 25.8
1 1 1 1 1 1 1 2 1 1 1 1 1 2 1
1 1 1 1 2 1
26 26.4 26.6 26.8 27 27.2 27.4 27.5 27.9 28 28.1 28.4 28.8 29 29.5
29.8 29.9 30 30.4 30.7 30.9
1 1 2 1 4 3 1 1 1 3 1 1 1 1 1
2 1 2 1 1 1
31 31.3 31.5 31.6 31.8 31.9 32 32.1 32.2 32.3 32.4 32.7 32.8 32.9 33
33.5 33.7 33.8 34 34.1 34.2
3 1 1 1 1 1 3 1 1 1 2 1 1 1 1
1 1 1 2 2 1
34.3 34.4 34.5 34.7 35 35.1 35.7 36 36.1 36.4 37 37.2 37.3 37.7 38
38.1 39 39.1 39.4 40.8 40.9
1 1 2 1 1 1 1 5 2 1 3 1 1 1 4
1 1 1 1 1 1
41.5 43.1 43.4 44 44.3 44.6 46.6
1 1 1 1 1 1 1
```

```
> max(table(cardata$mpg))
```

```
[1] 5
```

(la moda es 36)

```
> quantile(cardata$mpg, na.rm = TRUE)
```

```
0% 25% 50% 75% 100%
15.500 22.550 28.900 34.275 46.600
```

```
> quantile(cardata$mpg, c(0.47, 0.54, 0.82), na.rm = TRUE)
```

```
47% 54% 82%
```

```
28 30 36
```

b) Desviación típica, rango y varianza

```
> sd(cardata$mpg, na.rm = TRUE)
```

```
[1] 7.37721
```

```
> range(cardata$mpg, na.rm = TRUE)
```

```
[1] 15.5 46.6
```

```
> var(cardata$mpg, na.rm = TRUE)
```

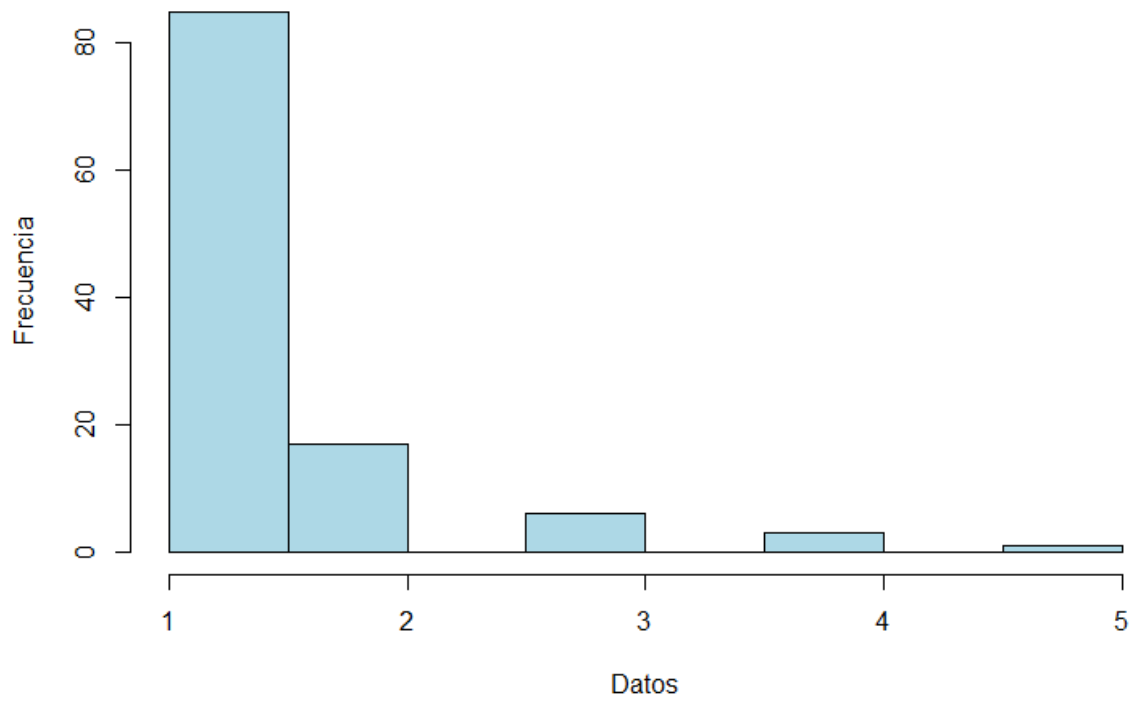
```
[1] 54.42323
```

c) Histograma de frecuencias absolutas y absolutas acumuladas

```
> tablaFrecAbs = table(cardata$mpg)
```

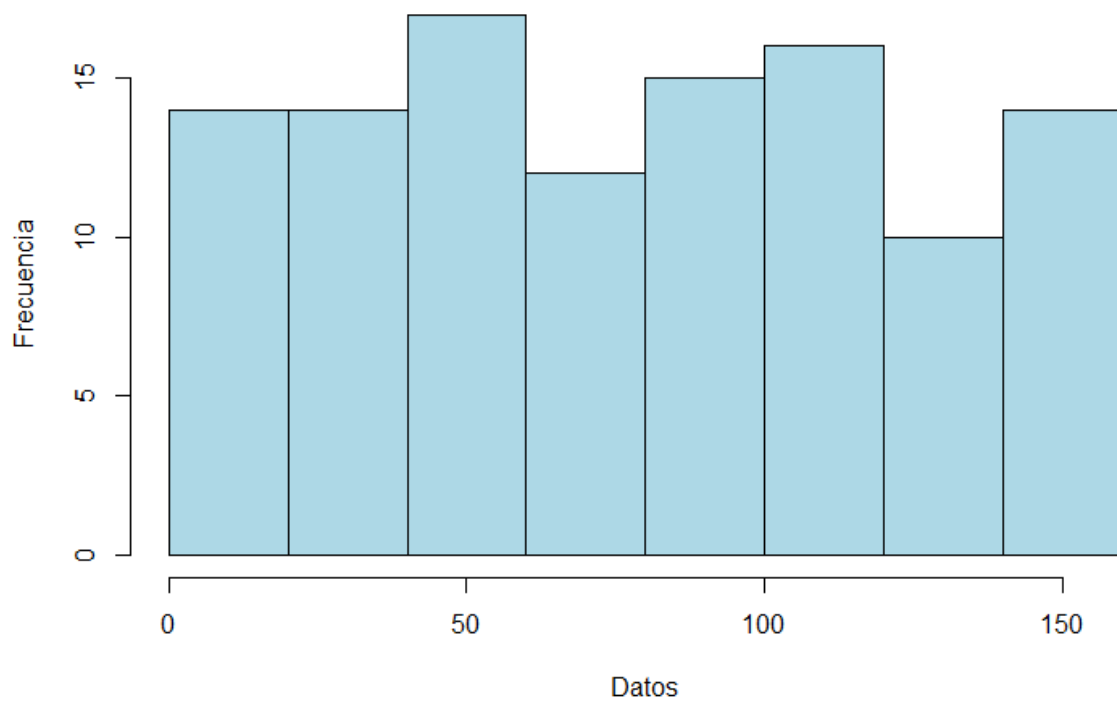
```
> hist(tablaFrecAbs, main = "Frecuencias Absolutas", xlab = "Datos", ylab =
"Frecuencia", col="lightblue")
```

Frecuencias Absolutas



```
> tablaFrecAcu = cumsum(tablaFrecAbs)
> hist(tablaFrecAcu, main = "Frecuencias Absolutas Acumuladas", xlab = "Datos", ylab = "Frecuencia", col="lightblue")
```

Frecuencias Absolutas Acumuladas



d) Cuadro de frecuencias para los datos agrupados en 10 clases

```
> k = 10
> A = diff(range(cardata$mpg, na.rm = TRUE)) / k
> L1 = min(cardata$mpg, na.rm = TRUE)
> L = L1 + A *(0:k)
> marcas = (L[0:k] + L[1:k+1]) / 2
> mpg = cut(cardata$mpg, breaks=L, right=FALSE)
> FrecAbs = table(mpg)
> FrecAcu = cumsum(FrecAbs)
> FrecRel = FrecAbs / n
> data.frame(marcas, FrecAbs, FrecRel, FrecAcu)
```

	marcas	Frec	FrecRel	FrecAcu
[15.5,18.6)	17.055	15	0.09677419	15
[18.6,21.7)	20.165	21	0.13548387	36
[21.7,24.8)	23.275	15	0.09677419	51
[24.8,27.9)	26.385	20	0.12903226	71
[27.9,31.1)	29.495	19	0.12258065	90
[31.1,34.2)	32.605	24	0.15483871	114
[34.2,37.3)	35.715	21	0.13548387	135
[37.3,40.4)	38.825	10	0.06451613	145
[40.4,43.5)	41.935	5	0.03225806	150
[43.5,46.6)	45.045	3	0.01935484	153

e) Cuadro de frecuencias para los datos agrupados en 12 clases

los mismos comandos que d) pero con k = 12

	marcas	Frec	FrecRel	FrecAcu
[15.5,18.1)	16.79583	10	0.06451613	10
[18.1,20.7)	19.38750	22	0.14193548	32
[20.7,23.3)	21.97917	10	0.06451613	42
[23.3,25.9)	24.57083	14	0.09032258	56
[25.9,28.5)	27.16250	20	0.12903226	76
[28.5,31.1)	29.75417	14	0.09032258	90
[31.1,33.6)	32.34583	18	0.11612903	108
[33.6,36.2)	34.93750	22	0.14193548	130
[36.2,38.8)	37.52917	12	0.07741935	142
[38.8,41.4)	40.12083	5	0.03225806	147
[41.4,44)	42.71250	4	0.02580645	151
[44,46.6)	45.30417	2	0.01290323	153

f) Calcular, utilizando las expresiones vistas en teoría, la media, la varianza y desviación típica para los datos agrupados. Comparar los resultados con los obtenidos en los apartados a) y b) ¿Existe mucha diferencia? Razonar la respuesta

Para datos agrupados en 10 clases:

```
> (mediaCalculada = (17.055 * 15 + 20.165 * 21 + 23.275 * 15 + 26.385 * 20
+ 29.495 * 19 + 32.605 * 24 + 35.715 * 21 + 38.825 * 10 + 41.935 * 5 + 45.0
45 * 3) / 153)
[1] 28.64127
```

```
(varianzaCalculada = (((17.055 - mediaCalculada) ^ 2 * 15 + (20.165 - media
Calculada) ^ 2 * 21 + (23.275 - mediaCalculada) ^ 2 * 15 + (26.385 - mediaC
alculada) ^ 2 * 20 + (29.495 - mediaCalculada) ^ 2 * 19 + (32.605 - mediaCa
lculada) ^ 2 * 24 + (35.715 - mediaCalculada) ^ 2 * 21 + (38.825 - mediaCal
culada) ^ 2 * 10 + (41.935 - mediaCalculada) ^ 2 * 5 + (45.045 - mediaCalcu
lada) ^ 2 * 3) / 153))
[1] 53.76364
```

```
> (desviacionCalculada = sqrt(varianzaCalculada))
[1] 7.332369
```

Para datos agrupados en 12 clases:

```
> (mediaCalculada = (16.79583 * 10 + 19.38750 * 22 + 21.97917 * 10 + 24.57083 * 14 + 27.16250 * 20 + 29.75417 * 14 + 32.34583 * 18 + 34.93750 * 22 + 37.52917 * 12 + 40.12083 * 5 + 42.71250 * 5 + 45.30417 * 2) / 153)
[1] 28.91536
```

```
> (varianzaCalculada = ((16.79583 - mediaCalculada) ^ 2 * 10 + (19.38750 - mediaCalculada) ^ 2 * 22 + (21.97917 - mediaCalculada) ^ 2 * 10 + (24.57083 - mediaCalculada) ^ 2 * 14 + (27.16250 - mediaCalculada) ^ 2 * 20 + (29.75417 - mediaCalculada) ^ 2 * 14 + (32.34583 - mediaCalculada) ^ 2 * 18 + (34.93750 - mediaCalculada) ^ 2 * 22 + (37.52917 - mediaCalculada) ^ 2 * 12 + (40.12083 - mediaCalculada) ^ 2 * 5 + (42.71250 - mediaCalculada) ^ 2 * 5 + (45.30417 - mediaCalculada) ^ 2 * 2) / 153)
[1] 52.6933
```

```
> (desviacionCalculada = sqrt(varianzaCalculada))
[1] 7.259015
```

Existen ciertas diferencias con estos resultados y los de a) y b) sobre todo en la varianza y en las pérdidas de decimales. El crear datos agrupados en intervalos parece que determina esta diferencia incluso entre los de 10 clases y 12 clases.