

R-PL1

Gabriel López, Sergio Sanz, Álvaro Zamorano

October 10, 2019

A lo largo de la práctica usaremos una función creada por nosotros para almacenar los resultados obtenidos en un .txt y posteriormente mostrarlos de forma correcta, por ello es necesario ejecutar:

```
> source("toTXT.R")
> toTxt

function(save,name,r) {
  write.table(save, paste("./tmp/",name,sep=""), sep="\t", row.names=r)
}
```

En esta parte de la práctica trabajaremos con el fichero `satelites.txt`.

En primer lugar hay que leer este fichero, para ello usamos la función:

```
> satellites<-read.table("satelites.txt")
```

Para trabajar con la variable radio, y hacer este trabajo más cómodo, la cargamos en una variable:

```
> Radio<-satellites$Radio
```

En el primer análisis de los datos se cuantifica la **frecuencia** de aparición de los mismos.

1. *Frecuencia absoluta:*

```
> frabsradio<-table(Radio)
> toTxt(frabsradio,"frabsradio.txt",F)
```

"Radio"	"Freq"
"13"	1
"15"	1
"16"	1
"20"	2
"22"	1
"27"	1
"29"	1
"30"	1

"33"	1
"34"	1
"42"	1

2. *Frecuencia absoluta acumulada:*

```
> frabsacumradio<-cumsum(table(Radio))
> toTxt(frabsacumradio,"frabsacumradio.txt",T)
```

"x"	
"13"	1
"15"	2
"16"	3
"20"	5
"22"	6
"27"	7
"29"	8
"30"	9
"33"	10
"34"	11
"42"	12

3. *Frecuencia relativa:* En este caso es necesario crear una función para poder calcular este valor. La función es:

```
> frecrel<-function(Radio){table(Radio)/length(Radio)}
> frer<-frecrel(Radio)
> toTxt(frer,"frer.txt",F)
```

"Radio"	"Freq"
"13"	0.0833333333333333
"15"	0.0833333333333333
"16"	0.0833333333333333
"20"	0.1666666666666667
"22"	0.0833333333333333
"27"	0.0833333333333333
"29"	0.0833333333333333
"30"	0.0833333333333333
"33"	0.0833333333333333
"34"	0.0833333333333333
"42"	0.0833333333333333

4. *Frecuencia relativa acumulada:* Haremos uso de la función definida anteriormente:

```
> frecrelacum<-function(Radio){cumsum(table(Radio)/length(Radio))}
> frar<-frecrelacum(Radio)
> toTxt(frar,"frar.txt",T)
```

"x"	
"13"	0.0833333333333333
"15"	0.1666666666666667
"16"	0.25
"20"	0.4166666666666667
"22"	0.5
"27"	0.5833333333333333
"29"	0.6666666666666667
"30"	0.75
"33"	0.8333333333333333
"34"	0.9166666666666667
"42"	1

El segundo análisis de los datos se basa en calcular la **media aritmética**:

```
> mr=mean(Radio)
> mr

[1] 25.08333
```

El tercer análisis de los datos se basa en calcular las **medidas de dispersión**:

1. *Desviación típica*: Para corregir los resultados, se hace el cálculo a través de:

```
> sdr<-sd(Radio)/sqrt(12/11)
> sdr

[1] 8.47996
```

2. *Varianza*: Al igual que en el caso anterior es necesario corregir el resultado por lo que se usa:

```
> varr<-var(Radio)*11/12
> varr

[1] 71.90972
```

El cuarto análisis de los datos se basa en las **medidas de ordenación**, antes de los cálculos es necesario ordenar los datos en función de la variable usada, en este caso el radio.

```
> so<-satelites[order(Radio),]
```

Realmente no sería necesario ordenar los datos, ya que R se encarga de ello en caso de no hacerlo. Se procede a realizar los cálculos:

1. *Mediana:*

```
> mediant<-median(Radio)
> mediant

[1] 24.5
```

2. *Cuartiles:*

```
> cuar1<-quantile(Radio,0.25)
> cuar1

25%
19
```

```
> cuar2<-quantile(Radio,0.5)
> cuar2

50%
24.5
```

```
> cuar3<-quantile(Radio,0.75)
> cuar3

75%
30.75
```

```
> cuar54<-quantile(Radio,0.54)
> cuar54

54%
26.7
```

A continuación pasaremos a trabajar con un fichero generado por SPSS, `cardata.sav`.

En primer lugar hay que leer este fichero pero no disponemos de la librería necesaria para hacerlo, para cargarla usamos:

```
> library(foreign)
```

Esta librería se trata de una librería estándar de R.

Una vez cargada procedemos a su lectura

```
> A<-read.spss("cardata.sav")
```

Para trabajar con la variable `mpg`, y hacer este trabajo más cómodo, la cargamos en una variable:

```
> mpg<-A$mpg
```

La variable mpg contiene valores NA, es decir, valores que no se encuentran disponibles por lo que es imposible realizar cálculos con ella. Para eliminar estos valores usamos:

```
> mpg<-mpg[!is.na(mpg)]
```

En el primer análisis de los datos se cuantifica la **frecuencia** de aparición de los mismos.

1. *Frecuencia absoluta:*

```
> frabsmpg<-table(mpg)
> toTxt(frabsmpg,"frabsmpg.txt",F)
```

"mpg"	"Freq"
"15.5"	1
"16.2"	1
"16.5"	1
"16.9"	1
"17"	2
"17.5"	1
"17.6"	2
"17.7"	1
"18.1"	2
"18.2"	1
"18.5"	1
"18.6"	1
"19.1"	1
"19.2"	3
"19.4"	2
"19.8"	1
"19.9"	1
"20.2"	4
"20.3"	1
"20.5"	2
"20.6"	2
"20.8"	1
"21.1"	1
"21.5"	1
"21.6"	1
"22"	1
"22.3"	1
"22.4"	1
"23"	2
"23.2"	1
"23.5"	1
"23.6"	1
"23.7"	1
"23.8"	1
"23.9"	2
"24"	1
"24.2"	1
"24.3"	1
"25"	1
"25.1"	1
"25.4"	2
"25.8"	1

"26"	1
"26.4"	1
"26.6"	2
"26.8"	1
"27"	4
"27.2"	3
"27.4"	1
"27.5"	1
"27.9"	1
"28"	3
"28.1"	1
"28.4"	1
"28.8"	1
"29"	1
"29.5"	1
"29.8"	2
"29.9"	1
"30"	2
"30.4"	1
"30.7"	1
"30.9"	1
"31"	3
"31.3"	1
"31.5"	1
"31.6"	1
"31.8"	1
"31.9"	1
"32"	3
"32.1"	1
"32.2"	1
"32.3"	1
"32.4"	2
"32.7"	1
"32.8"	1
"32.9"	1
"33"	1
"33.5"	1
"33.7"	1
"33.8"	1
"34"	2
"34.1"	2
"34.2"	1
"34.3"	1
"34.4"	1
"34.5"	2
"34.7"	1
"35"	1
"35.1"	1
"35.7"	1
"36"	5
"36.1"	2
"36.4"	1
"37"	3
"37.2"	1
"37.3"	1
"37.7"	1
"38"	4
"38.1"	1
"39"	1
"39.1"	1
"39.4"	1
"40.8"	1

"40.9"	1
"41.5"	1
"43.1"	1
"43.4"	1
"44"	1
"44.3"	1
"44.6"	1
"46.6"	1

2. *Frecuencia absoluta acumulada:*

```
> frabsacummpg<-cumsum(table(mpg))
> fam<-frabsacummpg
> toTxt(fam,"fam.txt",T)
```

"x"	
"15.5"	1
"16.2"	2
"16.5"	3
"16.9"	4
"17"	6
"17.5"	7
"17.6"	9
"17.7"	10
"18.1"	12
"18.2"	13
"18.5"	14
"18.6"	15
"19.1"	16
"19.2"	19
"19.4"	21
"19.8"	22
"19.9"	23
"20.2"	27
"20.3"	28
"20.5"	30
"20.6"	32
"20.8"	33
"21.1"	34
"21.5"	35
"21.6"	36
"22"	37
"22.3"	38
"22.4"	39
"23"	41
"23.2"	42
"23.5"	43
"23.6"	44
"23.7"	45
"23.8"	46
"23.9"	48
"24"	49
"24.2"	50
"24.3"	51
"25"	52
"25.1"	53
"25.4"	55

"25.8"	56
"26"	57
"26.4"	58
"26.6"	60
"26.8"	61
"27"	65
"27.2"	68
"27.4"	69
"27.5"	70
"27.9"	71
"28"	74
"28.1"	75
"28.4"	76
"28.8"	77
"29"	78
"29.5"	79
"29.8"	81
"29.9"	82
"30"	84
"30.4"	85
"30.7"	86
"30.9"	87
"31"	90
"31.3"	91
"31.5"	92
"31.6"	93
"31.8"	94
"31.9"	95
"32"	98
"32.1"	99
"32.2"	100
"32.3"	101
"32.4"	103
"32.7"	104
"32.8"	105
"32.9"	106
"33"	107
"33.5"	108
"33.7"	109
"33.8"	110
"34"	112
"34.1"	114
"34.2"	115
"34.3"	116
"34.4"	117
"34.5"	119
"34.7"	120
"35"	121
"35.1"	122
"35.7"	123
"36"	128
"36.1"	130
"36.4"	131
"37"	134
"37.2"	135
"37.3"	136
"37.7"	137
"38"	141
"38.1"	142
"39"	143
"39.1"	144
"39.4"	145

"40.8"	146
"40.9"	147
"41.5"	148
"43.1"	149
"43.4"	150
"44"	151
"44.3"	152
"44.6"	153
"46.6"	154

-
3. *Frecuencia relativa*: En este caso es necesario crear una función para poder calcular este valor. La función es:

```
> frecrel<-function(mpg){table(mpg)/length(mpg)}
> frm<-frecrel(mpg)
> toTxt(frm,"frm.txt",F)
```

"mpg"	"Freq"
"15.5"	0.00649350649350649
"16.2"	0.00649350649350649
"16.5"	0.00649350649350649
"16.9"	0.00649350649350649
"17"	0.012987012987013
"17.5"	0.00649350649350649
"17.6"	0.012987012987013
"17.7"	0.00649350649350649
"18.1"	0.012987012987013
"18.2"	0.00649350649350649
"18.5"	0.00649350649350649
"18.6"	0.00649350649350649
"19.1"	0.00649350649350649
"19.2"	0.0194805194805195
"19.4"	0.012987012987013
"19.8"	0.00649350649350649
"19.9"	0.00649350649350649
"20.2"	0.025974025974026
"20.3"	0.00649350649350649
"20.5"	0.012987012987013
"20.6"	0.012987012987013
"20.8"	0.00649350649350649
"21.1"	0.00649350649350649
"21.5"	0.00649350649350649
"21.6"	0.00649350649350649
"22"	0.00649350649350649
"22.3"	0.00649350649350649
"22.4"	0.00649350649350649
"23"	0.012987012987013
"23.2"	0.00649350649350649
"23.5"	0.00649350649350649
"23.6"	0.00649350649350649
"23.7"	0.00649350649350649
"23.8"	0.00649350649350649
"23.9"	0.012987012987013
"24"	0.00649350649350649
"24.2"	0.00649350649350649
"24.3"	0.00649350649350649
"25"	0.00649350649350649

"25.1"	0.00649350649350649
"25.4"	0.012987012987013
"25.8"	0.00649350649350649
"26"	0.00649350649350649
"26.4"	0.00649350649350649
"26.6"	0.012987012987013
"26.8"	0.00649350649350649
"27"	0.025974025974026
"27.2"	0.0194805194805195
"27.4"	0.00649350649350649
"27.5"	0.00649350649350649
"27.9"	0.00649350649350649
"28"	0.0194805194805195
"28.1"	0.00649350649350649
"28.4"	0.00649350649350649
"28.8"	0.00649350649350649
"29"	0.00649350649350649
"29.5"	0.00649350649350649
"29.8"	0.012987012987013
"29.9"	0.00649350649350649
"30"	0.012987012987013
"30.4"	0.00649350649350649
"30.7"	0.00649350649350649
"30.9"	0.00649350649350649
"31"	0.0194805194805195
"31.3"	0.00649350649350649
"31.5"	0.00649350649350649
"31.6"	0.00649350649350649
"31.8"	0.00649350649350649
"31.9"	0.00649350649350649
"32"	0.0194805194805195
"32.1"	0.00649350649350649
"32.2"	0.00649350649350649
"32.3"	0.00649350649350649
"32.4"	0.012987012987013
"32.7"	0.00649350649350649
"32.8"	0.00649350649350649
"32.9"	0.00649350649350649
"33"	0.00649350649350649
"33.5"	0.00649350649350649
"33.7"	0.00649350649350649
"33.8"	0.00649350649350649
"34"	0.012987012987013
"34.1"	0.012987012987013
"34.2"	0.00649350649350649
"34.3"	0.00649350649350649
"34.4"	0.00649350649350649
"34.5"	0.012987012987013
"34.7"	0.00649350649350649
"35"	0.00649350649350649
"35.1"	0.00649350649350649
"35.7"	0.00649350649350649
"36"	0.0324675324675325
"36.1"	0.012987012987013
"36.4"	0.00649350649350649
"37"	0.0194805194805195
"37.2"	0.00649350649350649
"37.3"	0.00649350649350649
"37.7"	0.00649350649350649
"38"	0.025974025974026
"38.1"	0.00649350649350649
"39"	0.00649350649350649

"39.1"	0.00649350649350649
"39.4"	0.00649350649350649
"40.8"	0.00649350649350649
"40.9"	0.00649350649350649
"41.5"	0.00649350649350649
"43.1"	0.00649350649350649
"43.4"	0.00649350649350649
"44"	0.00649350649350649
"44.3"	0.00649350649350649
"44.6"	0.00649350649350649
"46.6"	0.00649350649350649

4. *Frecuencia relativa acumulada*: Haremos uso de la función definida anteriormente:

```
> frecrelacum<-function(mpg){cumsum(table(mpg)/length(mpg))}
> fram<-frecrelacum(mpg)
> toTxt(fram,"fram.txt",T)
```

"x"	
"15.5"	0.00649350649350649
"16.2"	0.012987012987013
"16.5"	0.0194805194805195
"16.9"	0.025974025974026
"17"	0.038961038961039
"17.5"	0.0454545454545455
"17.6"	0.0584415584415584
"17.7"	0.0649350649350649
"18.1"	0.0779220779220779
"18.2"	0.0844155844155844
"18.5"	0.0909090909090909
"18.6"	0.0974025974025974
"19.1"	0.103896103896104
"19.2"	0.123376623376623
"19.4"	0.136363636363636
"19.8"	0.142857142857143
"19.9"	0.149350649350649
"20.2"	0.175324675324675
"20.3"	0.181818181818182
"20.5"	0.194805194805195
"20.6"	0.207792207792208
"20.8"	0.214285714285714
"21.1"	0.220779220779221
"21.5"	0.227272727272727
"21.6"	0.233766233766234
"22"	0.24025974025974
"22.3"	0.246753246753247
"22.4"	0.253246753246753
"23"	0.266233766233766
"23.2"	0.272727272727273
"23.5"	0.279220779220779
"23.6"	0.285714285714286
"23.7"	0.292207792207792
"23.8"	0.298701298701299
"23.9"	0.311688311688312
"24"	0.318181818181818
"24.2"	0.324675324675325

"24.3"	0.331168831168831
"25"	0.337662337662338
"25.1"	0.344155844155844
"25.4"	0.357142857142857
"25.8"	0.363636363636364
"26"	0.37012987012987
"26.4"	0.376623376623377
"26.6"	0.38961038961039
"26.8"	0.396103896103896
"27"	0.422077922077922
"27.2"	0.441558441558442
"27.4"	0.448051948051948
"27.5"	0.454545454545455
"27.9"	0.461038961038961
"28"	0.480519480519481
"28.1"	0.487012987012987
"28.4"	0.493506493506494
"28.8"	0.5
"29"	0.506493506493507
"29.5"	0.512987012987013
"29.8"	0.525974025974026
"29.9"	0.532467532467533
"30"	0.545454545454546
"30.4"	0.551948051948052
"30.7"	0.558441558441559
"30.9"	0.564935064935065
"31"	0.584415584415584
"31.3"	0.590909090909091
"31.5"	0.597402597402597
"31.6"	0.603896103896104
"31.8"	0.61038961038961
"31.9"	0.616883116883117
"32"	0.636363636363636
"32.1"	0.642857142857143
"32.2"	0.649350649350649
"32.3"	0.655844155844156
"32.4"	0.668831168831169
"32.7"	0.675324675324675
"32.8"	0.681818181818182
"32.9"	0.688311688311688
"33"	0.694805194805195
"33.5"	0.701298701298701
"33.7"	0.707792207792208
"33.8"	0.714285714285714
"34"	0.727272727272727
"34.1"	0.74025974025974
"34.2"	0.746753246753247
"34.3"	0.753246753246753
"34.4"	0.75974025974026
"34.5"	0.772727272727273
"34.7"	0.779220779220779
"35"	0.785714285714286
"35.1"	0.792207792207792
"35.7"	0.798701298701299
"36"	0.831168831168831
"36.1"	0.844155844155844
"36.4"	0.850649350649351
"37"	0.87012987012987
"37.2"	0.876623376623377
"37.3"	0.883116883116883
"37.7"	0.88961038961039
"38"	0.915584415584416

"38.1"	0.922077922077922
"39"	0.928571428571429
"39.1"	0.935064935064935
"39.4"	0.941558441558442
"40.8"	0.948051948051948
"40.9"	0.954545454545455
"41.5"	0.961038961038961
"43.1"	0.967532467532468
"43.4"	0.974025974025974
"44"	0.980519480519481
"44.3"	0.987012987012987
"44.6"	0.993506493506494
"46.6"	1

El segundo análisis de los datos se basa en calcular la **media aritmética**:

```
> mm<-mean(mpg)
> mm

[1] 28.79351
```

El tercer análisis de los datos se basa en calcular las **medidas de dispersión**:

1. *Desviación típica*: Para corregir los resultados, se hace el cálculo a través de:

```
> sdm<-sd(mpg)/sqrt(12/11)
> sdm

[1] 7.063141
```

2. *Varianza*: Al igual que en el caso anterior es necesario corregir el resultado por lo que se usa:

```
> varm<-var(mpg)*11/12
> varm

[1] 49.88796
```

El cuarto análisis de los datos se basa en las **medidas de ordenación**,

1. *Mediana*:

```
> mediantm<-median(mpg)
> mediantm

[1] 28.9
```

2. *Cuartiles*:

```
> cuar1m<-quantile(mpg,0.25)
> cuar1m

25%
22.55

> cuar2m<-quantile(mpg,0.5)
> cuar2m

50%
28.9

> cuar3m<-quantile(mpg,0.75)
> cuar3m

75%
34.275

> cuar54m<-quantile(mpg,0.54)
> cuar54m

54%
30
```

En la segunda parte de la práctica vamos a trabajar con una base de datos descargada de *Kaggle* cuyos datos pertenecen a los jugadores de FIFA 19.

El archivo se encuentra en formato *csv*, para proceder a su lectura vamos a usar la librería **readr** por lo que es necesario instalarla mediante:

```
> install.packages("readr")
```

Una vez instala se carga en R haciendo uso de:

```
> library("readr")
```

Por último leemos el archivo.

```
> fifa<-read_csv("fifa19.csv")
```

Trabajaremos con la variable **Age** por lo que vamos a cargarla en una variable local para facilitar el trabajo.

```
> edad<-fifa$Age
```

Con el objetivo de observar las **frecuencias absolutas** de la variable crearemos un histograma con ellas. Hemos delegado la creación del histograma en una función externa la cual almacena este en un archivo *png* para posteriormente poder incluirlo correctamente al documento.

La función es la siguiente:

```
> source("histograma.R")
> histograma

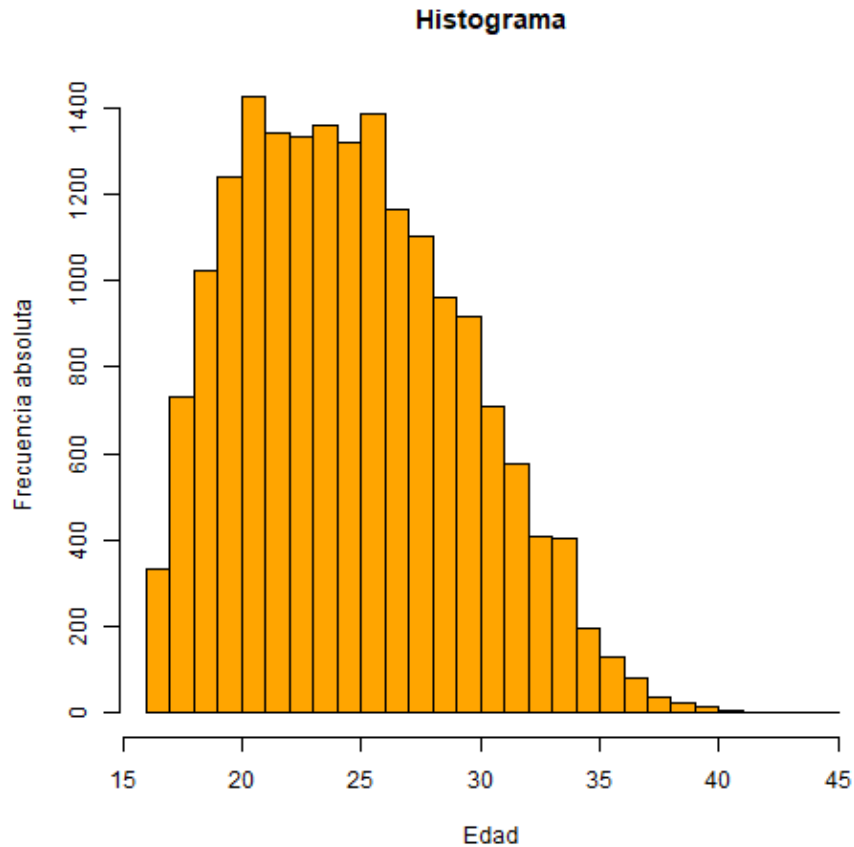
function(var,name,ruta) {
  png(paste("./tmp/",ruta,sep=""))

  h<-hist(var, col='orange', breaks=40, xlab=name,
          ylab="Frecuencia absoluta", main ="Histograma")

  dev.off()
}
```

Procedemos a su realización:

```
> h<-histograma(edad, "Edad", "hist.png")
```



Procedemos a la separación de la variable en **clases de equivalencia** para calcular de ellas los distintos tipos de frecuencia estudiados. Esto se hace gracias al uso de la librería **fdth** y para poder utilizarla, como se ha dicho anteriormente es necesario realizar lo siguiente:

```
> install.packages("fdth")
> library("fdth")
```

Procedemos al cálculo en sí:

```
> dist <- fdt(edad,breaks="Sturges")
> dist
```

Class limits	f	rf	rf(%)	cf	cf(%)
[16,18)	331	0.02	1.82	331	1.82
[18,20)	1756	0.10	9.64	2087	11.46
[20,21)	2663	0.15	14.63	4750	26.09
[21,23)	2672	0.15	14.68	7422	40.76
[23,25)	2677	0.15	14.70	10099	55.47
[25,27)	1387	0.08	7.62	11486	63.09
[27,29)	2263	0.12	12.43	13749	75.51

[29,31)	1876	0.10	10.30	15625	85.82
[31,32)	1281	0.07	7.04	16906	92.85
[32,34)	812	0.04	4.46	17718	97.31
[34,36)	323	0.02	1.77	18041	99.09
[36,38)	119	0.01	0.65	18160	99.74
[38,40)	25	0.00	0.14	18185	99.88
[40,42)	18	0.00	0.10	18203	99.98
[42,44)	1	0.00	0.01	18204	99.98
[44,45)	3	0.00	0.02	18207	100.00

- Frecuencia absoluta -> f
- Frecuencia relativa -> rf
- Frecuencia relativa porcentual -> rf(%)
- Frecuencia abs acumulada -> cf
- Frecuencia rel acumulada porcentual -> cf(%)

Para calcular la **media** de edad hemos construido una función en R la cual es:

```
> source("media.R")
> media

function(var) {
  sum<-0

  for (data in var) {
    sum<-sum+data
  }

  return(sum/length(var))
}
```

Procedemos al cálculo de la media haciendo uso de la misma:

```
> mf<-media(fifa$Age)
> mf

[1] 25.12221
```

En cuanto a los **cuantiles**, para su visualización hemos optado por el uso de un diagrama de bigotes, donde además aparecen representados los límites inferior y superior de la variable de estudio. Para su realización se ha hecho lo mismo que con el histograma de la siguiente manera:

```
> source("bigotes.R")
> bigotes
```

```

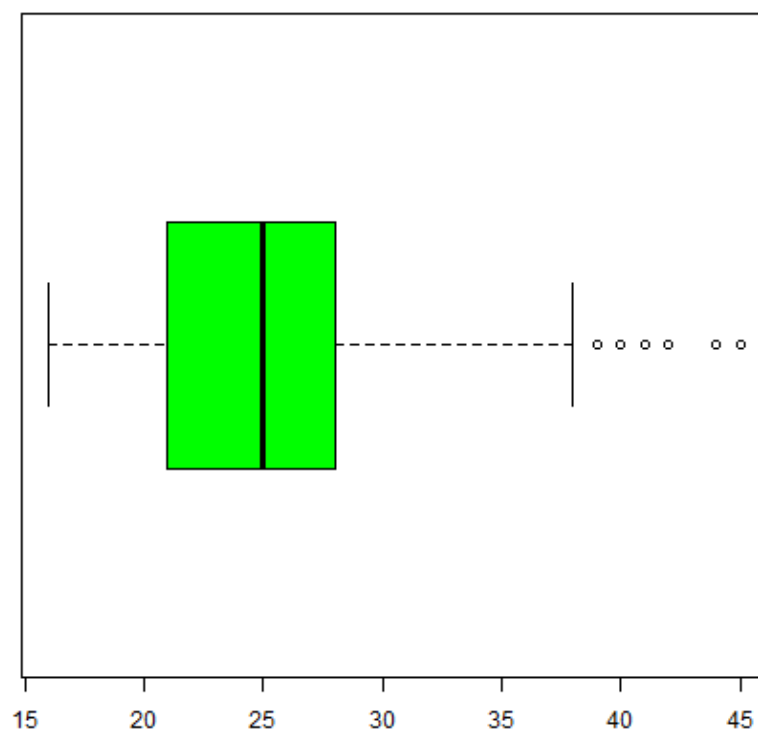
function(var,ruta) {
  png(paste("./tmp/",ruta,sep=""))

  boxplot(var, col='green', horizontal=T)

  dev.off()
}

> b<-bigotes(edad,"bigotes.png")

```



Para calcular la **desviación típica** hemos construido una función de la siguiente forma:

```

> source("desviacion.R")
> desviacion

function(var,media) {
  num<-0
  for (data in var){
    num<-num+((data-media)^2)
  }
}

```

```

    s<-sqrt(num/length(var))

    return(s)
}

```

El cálculo de la misma es:

```

> s<-desviacion(edad,mf)
> s

[1] 4.669814

```

De acuerdo al teorema de *Tchebychev* (realizado en una función aparte) el intervalo en el que se encuentran el 75% de los datos, es decir para $k=2$, es:

```

> source("tche.R")
> t<-tche(s,mf,2)
> toString(t)

[1] "15, 34"

```

Dicha función es:

```

> tche

function(des,media,k) {

    rango<-des*k
    inter<-list(as.integer(media-rango),as.integer(media+rango))

    return(inter)
}

```

Observando el intervalo tan amplio necesario para englobar el 75% de los datos se puede concluir con que la media no es una buena medida representante del conjunto de datos estudiado.

Por último, el cálculo de la **varianza** es sencillo, únicamente es la desviación al cuadrado:

```

> v<-s^2
> v

[1] 21.80717

```