

## 2-Teoria

### Setup para realizar el ejercicio

Primero crearemos una lista con todos los datos de los diferentes tumores dados.

**a) Obténgase una distribución de datos en intervalos de amplitud del 15% de la distribución con los distintos tipos de frecuencias y visualizar el conjunto con un histograma y un diagrama de caja. Razonar sobre los distintos elementos de estos y explicar su contenido**

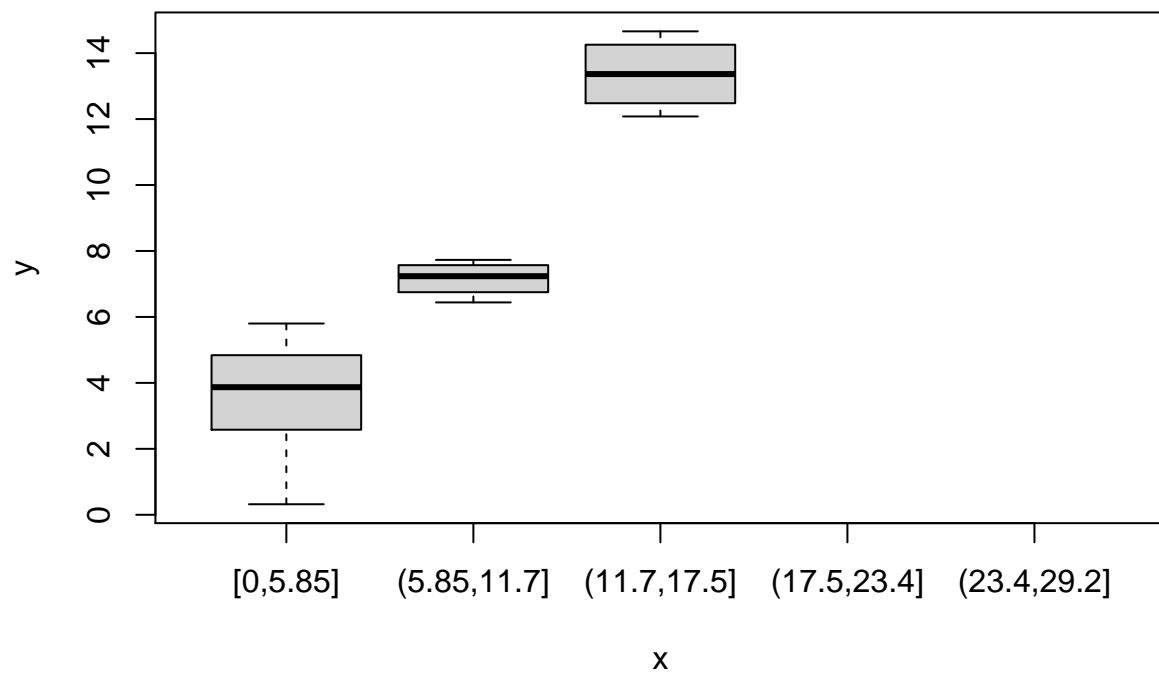
Para la realización de este apartado usaremos el comando cut para dividir los datos en diferentes grupos. Luego lo añadiremos a un data frame donde todos los datos pertenecieran a un intervalo.

Por último mostraremos un boxplot, un histograma y un plot normal de los datos que hemos obtenido. De esta manera podremos analizar mas adelante de una mejor manera los datos.

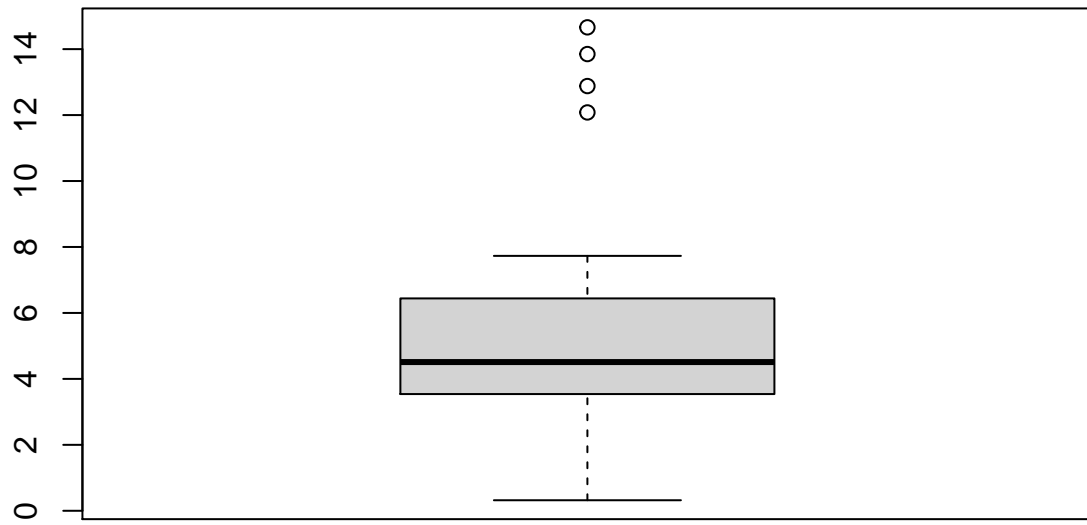
```
data_intervalos <- cut(data, breaks= seq(0,31, 5.85 ), include.lowest = TRUE)

frame <- data.frame(intervalos = data_intervalos, valor = data)

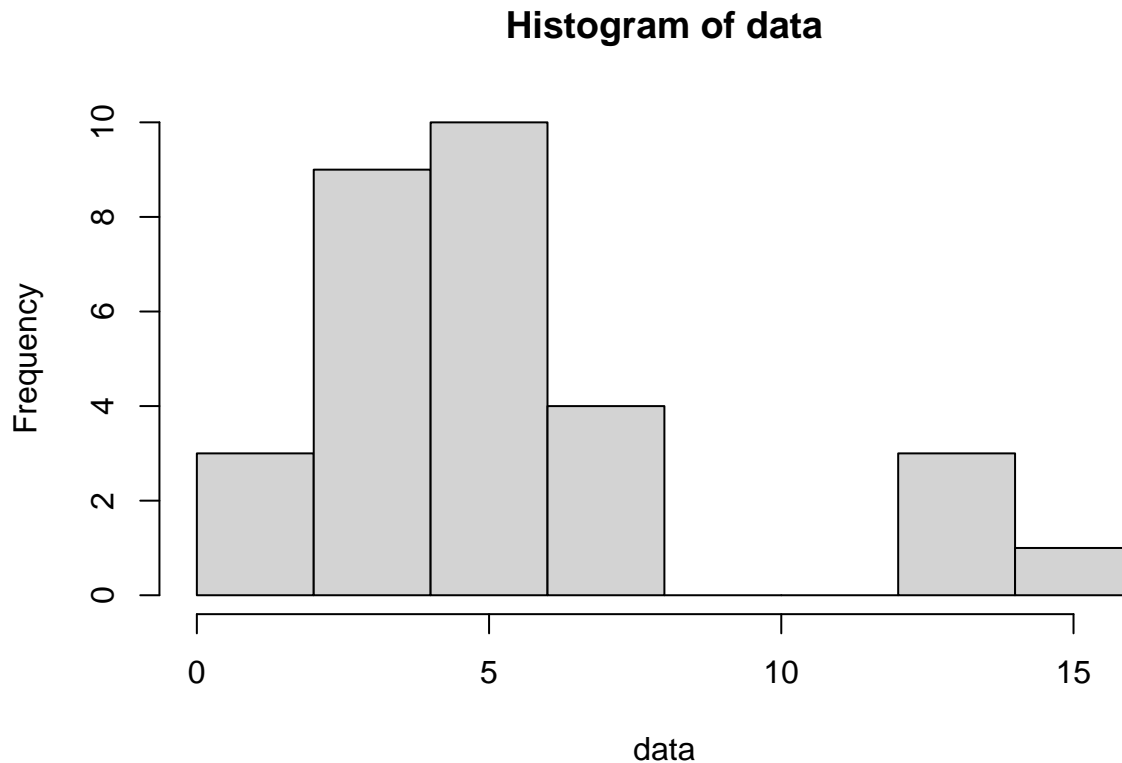
plot(frame$intervalos,frame$valor)
```



```
boxplot(data)
```



```
hist(data)
```



**b) Calcular los tamaños de tumor medio, el más frecuente y su variabilidad.**

Aquí usando la función mean, median y sd calcularemos los datos que necesitamos. Observamos que la media del tumor es aproximadamente 5.3, el tumor medio es del tamaño de 4.51 mm, y el grado de dispersión 3.6mm. A su vez importaremos la librería modeest y usaremos la función mlv para calcular la moda ( el mas frecuente ) el cual ha resultado ser 4.84mm.

```
mean(data)
```

```
## [1] 5.384333
```

```
median(data)
```

```
## [1] 4.51
```

```
sd(data)
```

```
## [1] 3.649017
```

```
library(modeest)
```

```
mlv(data, method="mfv")
```

```
## [1] 4.84
```

**c) ¿Cuántos tiene un grosor mayor de 10 milímetros?**

La realización de este apartado consiste simplemente en usar un filtrado a la hora de recoger los datos entre los corchetes. Y por último calcularemos la longitud de la lista obtenida. Existen 4 casos con tumores mayores a 10mm

```
length(data[data > 10])
```

```
## [1] 4
```

#### d) ¿Cuál es el tamaño de tumor más pequeño que tienen el 25% de pacientes afectados?

Para calcular el tumor mas pequeño que tienen el 25% de los pacientes hace falta calcular primero los cuartiles en concreto el 0.25, y luego calcular el mínimo de este.

```
quantile(data, 0.25)
```

```
## 25%  
## 3.54
```

```
min(data)
```

```
## [1] 0.32
```

#### e) Calcular el tercer cuartil, el percentil 95 y el rango interdecil.

Para calcular el tercer cuartil es tan sencillo como usar la función quantile obteniendo que el tercer cuartil es 6.28 (Percentil 75). Para el percentil 95 sencillamente a la función quantile le agregaremos un parametro, el 0.95 y de esa forma obtendriamos ese percentil. Por último el rango interdecil es la diferencia entre los percentiles 0.90 y 0.10.

```
quantile(data)
```

```
## 0% 25% 50% 75% 100%  
## 0.32 3.54 4.51 6.28 14.66
```

```
quantile(data, 0.95)
```

```
## 95%  
## 13.4135
```

```
rango_interdecil <- quantile(data, 0.90) - quantile(data, 0.10); rango_interdecil
```

```
## 90%  
## 9.892
```

#### f) Estudiar analíticamente la concentración de los tamaños y explicarla.

Para estudiar la concentración ejecutaremos la función gini de la libreria reldist. Esta nos devolvera un número entre 0 y 1 que nos indicara la cantidad de dispersión siendo 0 una dispersión totalmente homogenea y 1 cuando un solo valor se repite a lo largo de todos los datos. Obtenemos que la distribución no es totalmente homogenea pero aún asi es un valor bajo.

```
library(reldist)
```

```
## reldist: Relative Distribution Methods  
## Version 1.6-6 created on 2016-10-07.  
## copyright (c) 2003, Mark S. Handcock, University of California-Los Angeles  
## For citation information, type citation("reldist").  
## Type help(package="reldist") to get started.
```

```
gini(data)
```

```
## [1] 0.3441198
```

### g) Calcular los factores de forma (asimetría y curtosis) de la distribución

Como podemos observar con la función `skewness` tenemos una distribución asimétrica positiva, ya que el valor dado es superior a 0. A su vez vemos como los valores estan bastante concentrados dentro del eje de simetría, al medir su aplastamiento vemos que este esta por encima de la distribución normal por lo tanto nos encontramos con una curva leptocurtica.

```
library(e1071)
```

```
##
```

```
## Attaching package: 'e1071'
```

```
## The following object is masked from 'package:modeest':
```

```
##
```

```
##      skewness
```

```
skewness(data)
```

```
## [1] 1.203033
```

```
kurtosis(data)
```

```
## [1] 0.6359236
```

### h) Sacar conclusiones

Podemos observar tanto en las gráficas mostradas anteriormente como en los datos recogidos que excepto unos valores concretos la mayoría de estos estan por debajo de los 10mm y se concentrán entorno a 5mm siendo la mediana el 4.51.