

Diagramas de Caja

Santiago Pérez Moncada

26/6/2020

El conocido **Diagrama de caja** o **box plot** es un tipo de gráfico que básicamente, remarca 5 valores estadísticos:

- La mediana, representada por la línea gruesa que divide la caja.
- El primer y tercer cuartil, que son los lados inferior y superior, respectivamente. De este modo, la altura de la caja es el rango intercuartílico.
- Los extremos, los valores b_{inf} , b_{sup} , son los **bigotes** (whiskers) del gráfico. Si m y M son el mínimo y el máximo de la variable cuantitativa, entonces los extremos se calculan del siguiente modo.

$$b_{inf} = \max\{m, Q_{0.25} - 1.5(Q_{0.75} - Q_{0.25})\}$$

$$b_{sup} = \min\{M, Q_{0.75} + 1.5(Q_{0.75} - Q_{0.25})\}$$

- Valores atípicos o outliers, que son los valores que están más allá de los bigotes. Se marcan como puntos aislados.

Por su definición, concluimos que los bigotes marcan el mínimo y el máximo de la variable cuantitativa, a no ser que haya datos muy alejados de la caja intercuartílica.

En tal caso, el bigote inferior marcará el valor 1.5 veces el rango intercuartílico por debajo de $Q_{0.25}$, mientras que el superior marcará el valor 1.5 veces el rango intercuartílico por encima de $Q_{0.75}$.

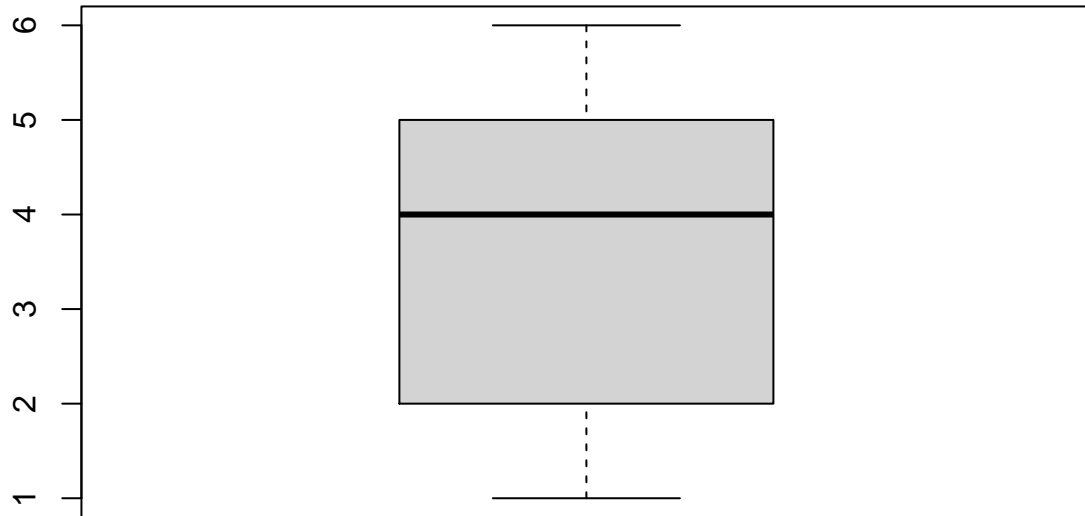
Ejemplo

La instrucción `boxplot()` dibuja diagramas de caja en R.

```
dados1 = sample(1:6, size = 1000, replace = TRUE)
dados2 = sample(1:4, size = 15, replace = TRUE)
dados3 = sample(1:6, size = 100, replace = TRUE)

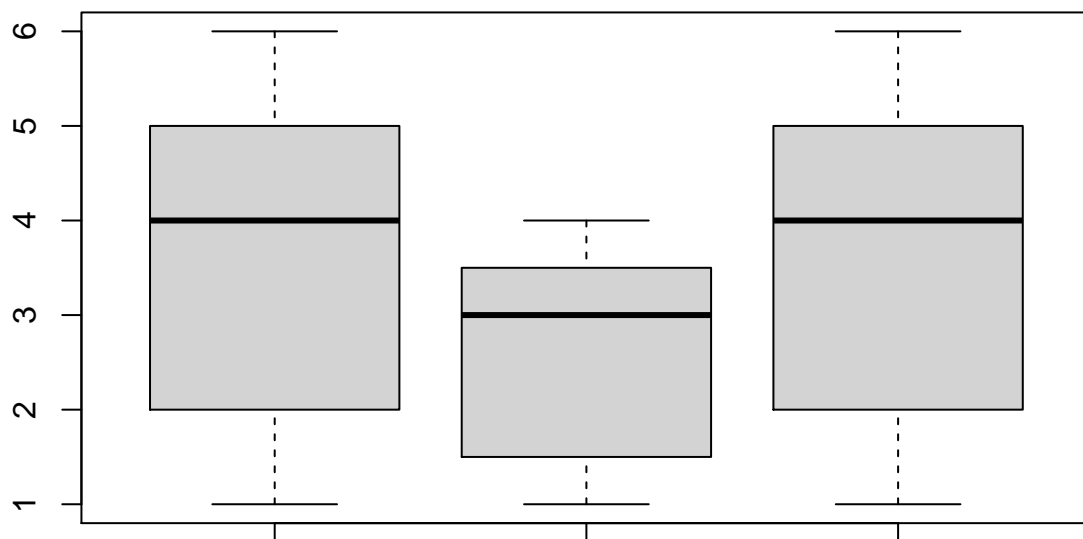
boxplot(dados1, main="Diagrama de caja")
```

Diagrama de caja

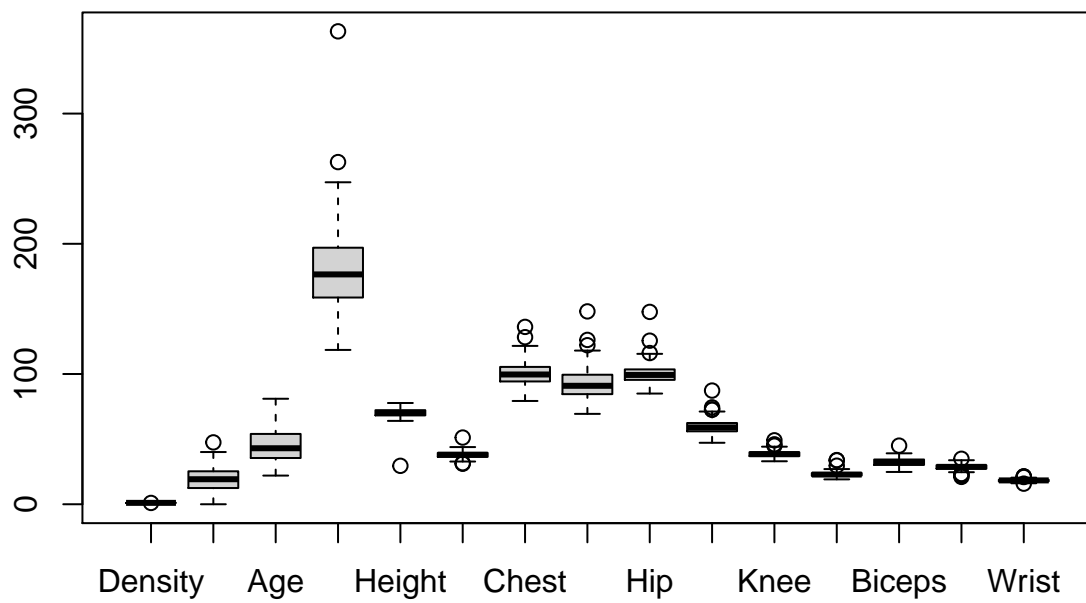


```
boxplot(dados1,dados2,dados3,main="Diagrama de caja")
```

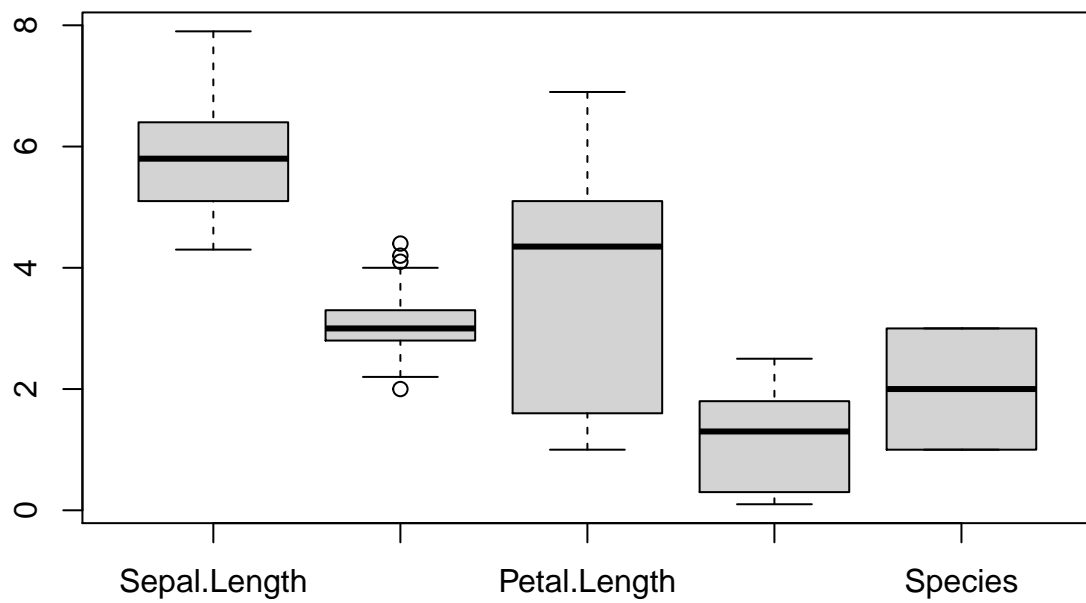
Diagrama de caja



```
body = read.table("../data/bodyfat.txt",header = TRUE)
boxplot(body)
```



```
boxplot(iris)
```



```
boxplot(body[,7:9], names = c("Pecho", "Abdomen", "Cadera"))
```

