Introduction to R Software

Introduction to Statistical Functions
:::
Boxplots, Skewness and Kurtosis

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Summary of observations

In R, quartiles, minimum and maximum values can be easily obtained by the `summary` command

`summary(x)  x: data vector`

It gives information on

❖ minimum,

❖ maximum

❖ first quartile

❖ second quartile (median) and

❖ third quartile.

Summary of observations

Example:

```
> marks <- c(68, 82, 63, 86, 34, 96, 41, 89,
    29, 51, 75, 77, 56, 59, 42)

> summary(marks)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   29.0    46.5    63.0    63.2    79.5    96.0
```

```
R Console

> summary(marks)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   29.0    46.5    63.0    63.2    79.5    96.0
>
```

Summary of observations

Example:

```
> marks1 <- c(628, 812, 613, 186, 34, 986, 41,
   89, 29, 51, 795, 77, 56, 509, 420)


> summary(marks1)
  Min. 1st Qu.  Median    Mean  3rd Qu.   Max.
  29.0    53.5   186.0   355.1   620.5  986.0
```
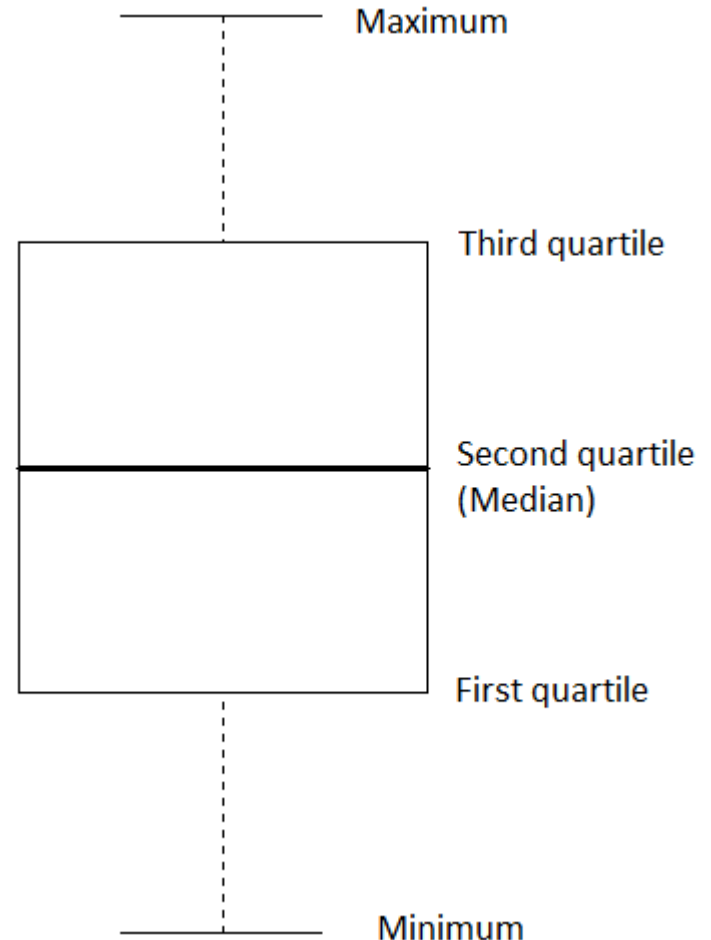
Earlier, we had

```
> summary(marks)
  Min. 1st Qu.  Median    Mean  3rd Qu.   Max.
  29.0    46.5    63.0    63.2    79.5   96.0
```

Boxplot

Box plot is a graph which summarizes the distribution of a variable by using its median, quartiles, minimum and maximum values.
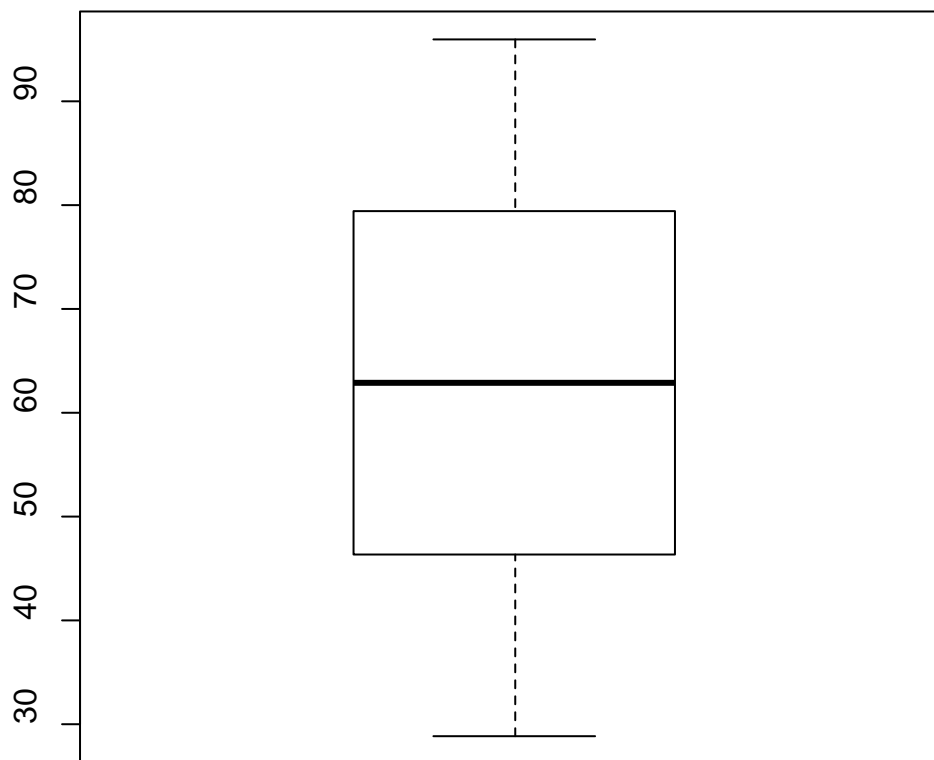
`boxplot()` draws a box plot.

Example:

```
> marks <- c(68, 82, 63, 86, 34, 96, 41, 89,
    29, 51, 75, 77, 56, 59, 42)

> boxplot(marks)
```
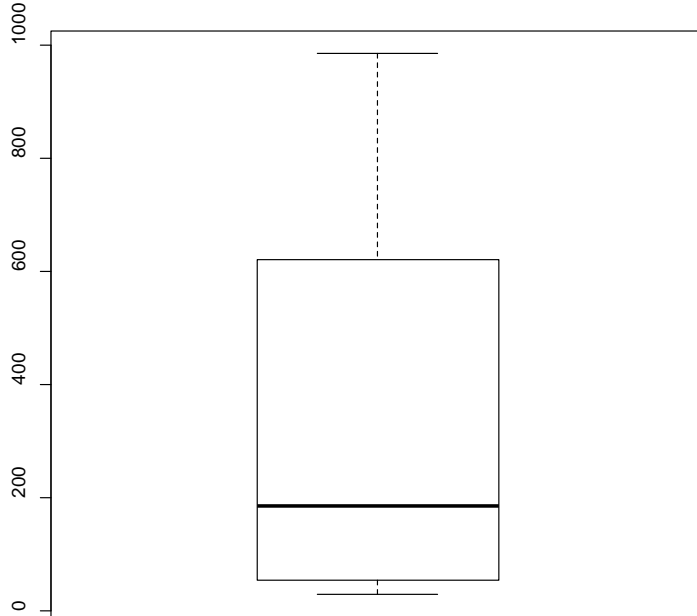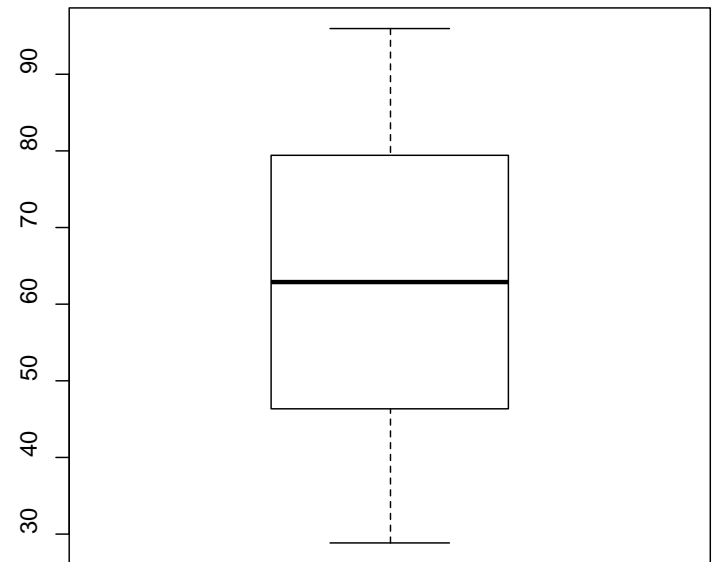
Example:

```
> marks1 <- c(628, 812, 613, 186, 34, 986, 41,
    89, 29, 51, 795, 77, 56, 509, 420)

> boxplot(marks1)
```
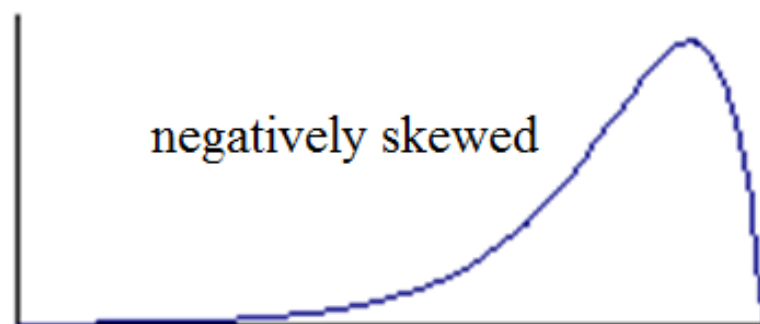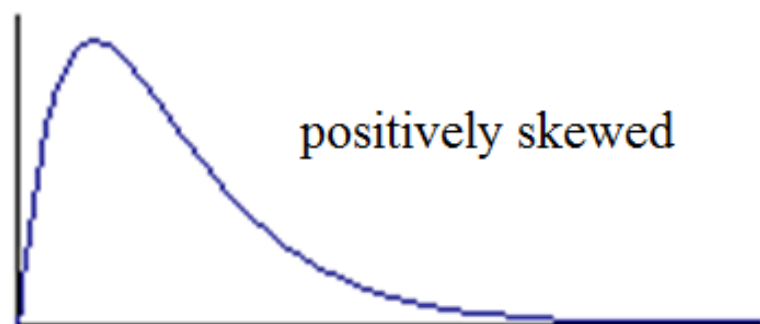

Boxplot(marks1)

Boxplot(marks)

Descriptive statistics:

First hand tools which gives first hand information.

- Structure and shape of data tendency (symmetricity, skewness, kurtosis etc.)

- Relationship study (correlation coefficient, rank correlation, corralation ratio, regression etc.)

Skewness

positively skewed

no skewness

negatively skewed

Kurtosis

platykurtic

mesokurtic

leptokurtic

Skewness

Measures the shift of the hump of frequency curve .

Coefficient of skewness based on values $x_1, x_2,..., x_n$.

$$\gamma_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{3/2}}$$

Mean :

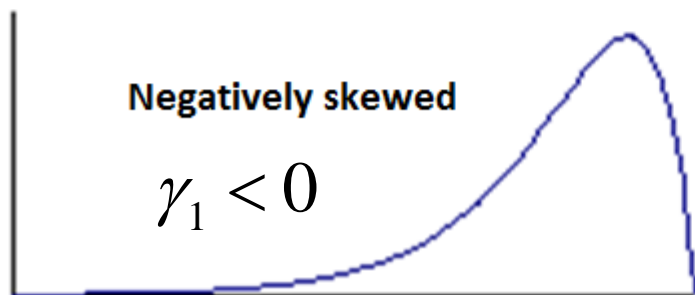$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$$
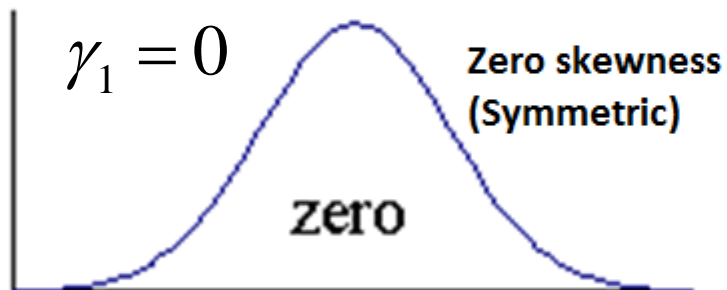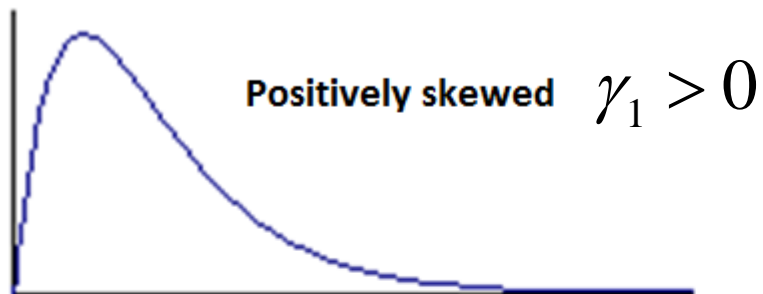
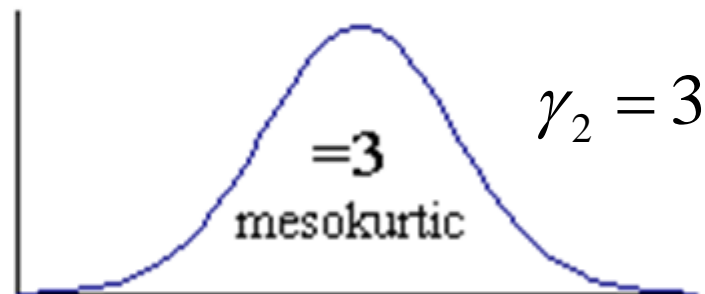# Kurtosis

Measures the peakedness of the frequency curve.

Coefficient of kurtosis based on values x_1, x_2,..., x_n.

$$\gamma_2 = \frac{\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left(\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2} \; , \; -3 < \gamma_2 < 3$$

Skewness

Positively skewed $\gamma_1 > 0$

$\gamma_1 = 0$ Zero skewness (Symmetric)

zero

Negatively skewed $\gamma_1 < 0$

Kurtosis

$\gamma_2 < 3$

<3 platykurtic

$\gamma_2 = 3$

=3 mesokurtic

$\gamma_2 > 3$

>3 leptokurtic

Skewness and kurtosis

First we need to install a package 'moments'

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

> library(moments)

skewness ()  : computes coefficient of skewness

kurtosis ()  : computes coefficient of kurtosis
```
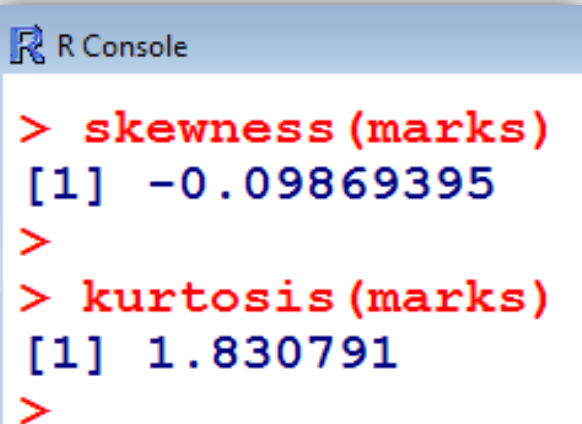
Skewness and kurtosis

Example

```
> marks <- c(68, 82, 63, 86, 34, 96, 41, 89,
29, 51, 75, 77, 56, 59, 42)

> skewness(marks)
[1] -0.09869395

> kurtosis(marks)
[1] 1.830791
```