Correlation Test

- A statistical measure that indicates how strongly two variables are related
- Involves the relationship between multiple variables as well
- For example, when one variable increase and the other increases as well, then these two variables are said to be **positively correlated**.
- The other way around when a variable increase and the other decrease then these two variables are **negatively correlated**.
- In the case of no correlation no pattern will be seen between the two variables.
- For instance, if one is interested to know whether there is a relationship between the heights of fathers and sons, a correlation coefficient can be calculated to answer this question. Generally, it lies between -1 and +1. It is a scaled version of covariance and provides the direction and strength of a relationship.

Types of correlation

There are mainly two types of correlation:

1. Parametric Correlation – **Pearson** correlation(r): It measures a **linear** dependence between two variables (x and y) is known as a parametric correlation test because it depends on the distribution of the data.

2. Non-Parametric Correlation – **Kendall** and **Spearman**: They are **rank-based** correlation coefficients, are known as non-parametric correlation.

Parametric Correlation

$$r = \frac{\Sigma(x - m_x)(y - m_y)}{\sqrt{\Sigma(x - m_x)^2 \Sigma(y - m_y)^2}}$$

Where,

r: pearson correlation coefficient

x and y: two vectors of length n

mx and my: corresponds to the means of x and y, respectively.

Implementation in R

Syntax: cor(x, y, method = “pearson”)
cor.test(x, y, method = “pearson”)

Where:

x, y: numeric vectors with the same length

method: correlation method

Example:1

```
x = c(1, 2, 3, 4, 5, 6, 7)
y = c(1, 3, 6, 2, 7, 4, 5)

result = cor(x, y, method = "pearson")

cat("Pearson correlation coefficient is:", result)
```

OR

```
result = cor.test(x, y, method = "pearson")

print(result)
```

Example:2

Height of Father as x = c(65,66,67,67,68,69,71,73)
Height of Son as y= c(64,65,66,66,67,68,70,71)

```
result = cor(x, y, method = "pearson")

cat("Pearson correlation coefficient is:", result)
```

OR

```
result = cor.test(x, y, method = "pearson")

print(result)
```

To analysis:

Scatter Plot:

```
plot(x,y,xlab='Father H',ylab = "Son H",col='red',xlim=c(60,80),ylim=c(60,80))

abline(lm(y~x),col="blue")
```

Plotting symbols

pch = 0,square
- pch = 1,circle
- pch = 2,triangle point up
- pch = 3,plus
- pch = 4,cross
- pch = 5,diamond
- pch = 6,triangle point down
- pch = 7,square cross
- pch = 8,star
- pch = 9,diamond plus
- pch = 10,circle plus
- pch = 11,triangles up and down
- pch = 12,square plus
- pch = 13,circle cross
- pch = 14,square and triangle down
- pch = 15, filled square
- pch = 16, filled circle
- pch = 17, filled triangle point-up
- pch = 18, filled diamond
- pch = 19, solid circle
- pch = 20,bullet (smaller circle)

pch = 21, filled circle blue
pch = 22, filled square blue
pch = 23, filled diamond blue
pch = 24, filled triangle point-up blue
pch = 25, filled triangle point down blue