Correlation study of the physical development in infants

Lakshmi Arbatti

Saturday, January 31, 2015

Introduction

In this example we study the physical development of an infant by measuring the age and weight of the infant and apply correlation coefficient to determine how the two variables are correlated.

Correlation Study

Correlation study is the study of determining if two variables are related, i.e whether an increase or decrease in one variable has a similar effect on the other variable. The value of the coefficient lies between +1 and -1 inclusive, where 1 is total positive correlation and -1 is total negative correlation. While R provides a simple function "**cor**" which we will be using in this example, it is important that we understand the calculations are performed in the background, to arrive at the result.

Pearson's Correlation Coefficient

Pearson's correlation coefficient defined as the covariance of the two variables divided by the product of their standard deviations. For a population, the Person's coefficient represented by the Greek letter ρ (rho) carries the formula :

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

where, cov is the covariance, σ_X is the standard deviation of X , μ_X is the mean of X , and E is the expectation.

(Source: Wikipedia)

In this example, we will calculate the correlation between age of a child and its weight.

```
age <- c(1,3,5,2,11,9,3,9,12,3)
weight <- c(4.4,5.3,7.2,5.2,8.5,7.3,6.0,10.4,10.2,6.1)
aframe <- data.frame(age,weight)
aframe
```

```
##    age weight
## 1    1    4.4
## 2    3    5.3
## 3    5    7.2
## 4    2    5.2
## 5   11    8.5
## 6    9    7.3
## 7    3    6.0
## 8    9   10.4
## 9   12   10.2
## 10   3    6.1
```

Given the age and weight in the above table, the correlation is “cor(age,weight)” : 0.9075655

```
plot(age,weight)
```