# Correlation Analysis

**MISAB  P.T**

**Ph.D Management**

# Definition of Correlation

- Correlation is the degree of association between two or more variables.

- If two or more quantities vary so that movements in one tend to be accompanied by movements in other, then they are said to be correlated.

- Coefficient of correlation is a numerical measure of the degree of association between two or more variables.

# Meaning

- Correlation is the most popular statistical measure that indicates the relationship between two or more variables.
- It is concerned with finding:
  - Whether or not the relationship exist?
  - Degree of the correlation?
  - Direction of relationship within the variables (Direct or indirect)?
  - Relationship is strong or Weak?

# Examples

- Relationship between income and years of experience
- Relationship between amount of rainfall and yield of rice
- Relationship between price and demand of a commodity
- Relationship between nature of work and motivation to work
- Relationship between height and weight

# Scope of Correlation Analysis

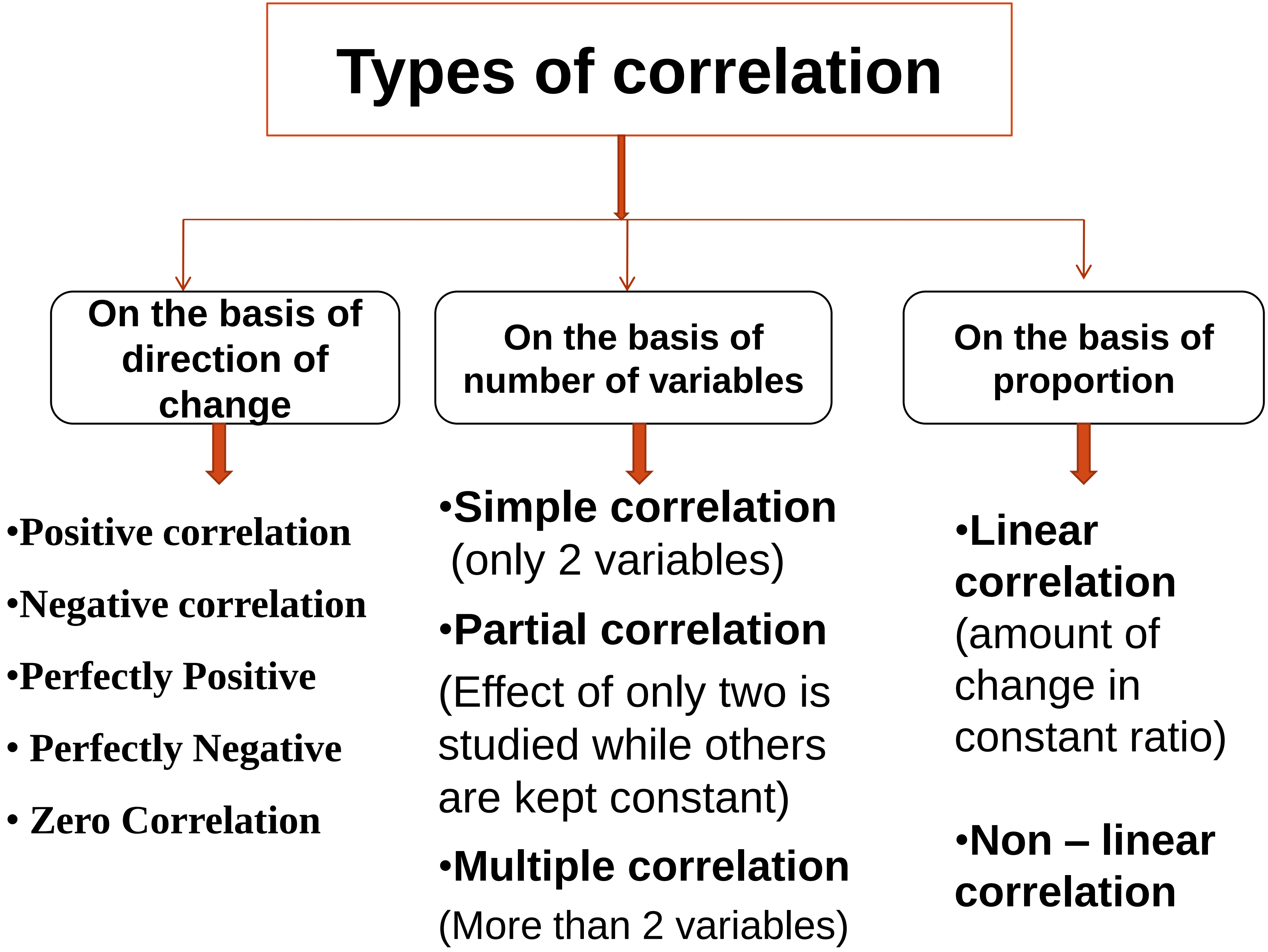
- The existence of correlation between two (or more) variables only implies that these variables:

1. Either tend to increase or decreased together
2. An increase (or decrease) in one is accompanied by the corresponding decrease (or increase) in the other.

- Correlation analysis does not answer the questions like why there is cause and effect between two variables.
- It may be due to following reasons:

# Scope of correlation analysis

- One of the variable may be affecting the other. A correlation calculated from the data on demand and price will only show that degree of association between demand and price is high. It will not show why it happens.
- The two variables may act upon each other. Cause and effect is here also, but it is difficult to find which variable is independent and which is dependent.
- The two variables may be acted upon by the outside influence. Such correlation is called spurious or nonsense correlation.
- A high value of the correlation may be due to sheet coincidence ( or pure chance)



# Types of correlation

## On the basis of direction of change

- **Positive correlation**
- **Negative correlation**
- **Perfectly Positive**
- **Perfectly Negative**
- **Zero Correlation**

## On the basis of number of variables

- **Simple correlation** (only 2 variables)
- **Partial correlation** (Effect of only two is studied while others are kept constant)
- **Multiple correlation** (More than 2 variables)

## On the basis of proportion

- **Linear correlation** (amount of change in constant ratio)
- **Non – linear correlation**

# Types of Correlation

Correlation on the basis of direction of change is as following:

(1) Positive  Correlation
(2) Negative Correlation
(3) Perfectly Positive Correlation
(4) Perfectly Negative Correlation
(5) Zero Correlation

# Positive Correlation

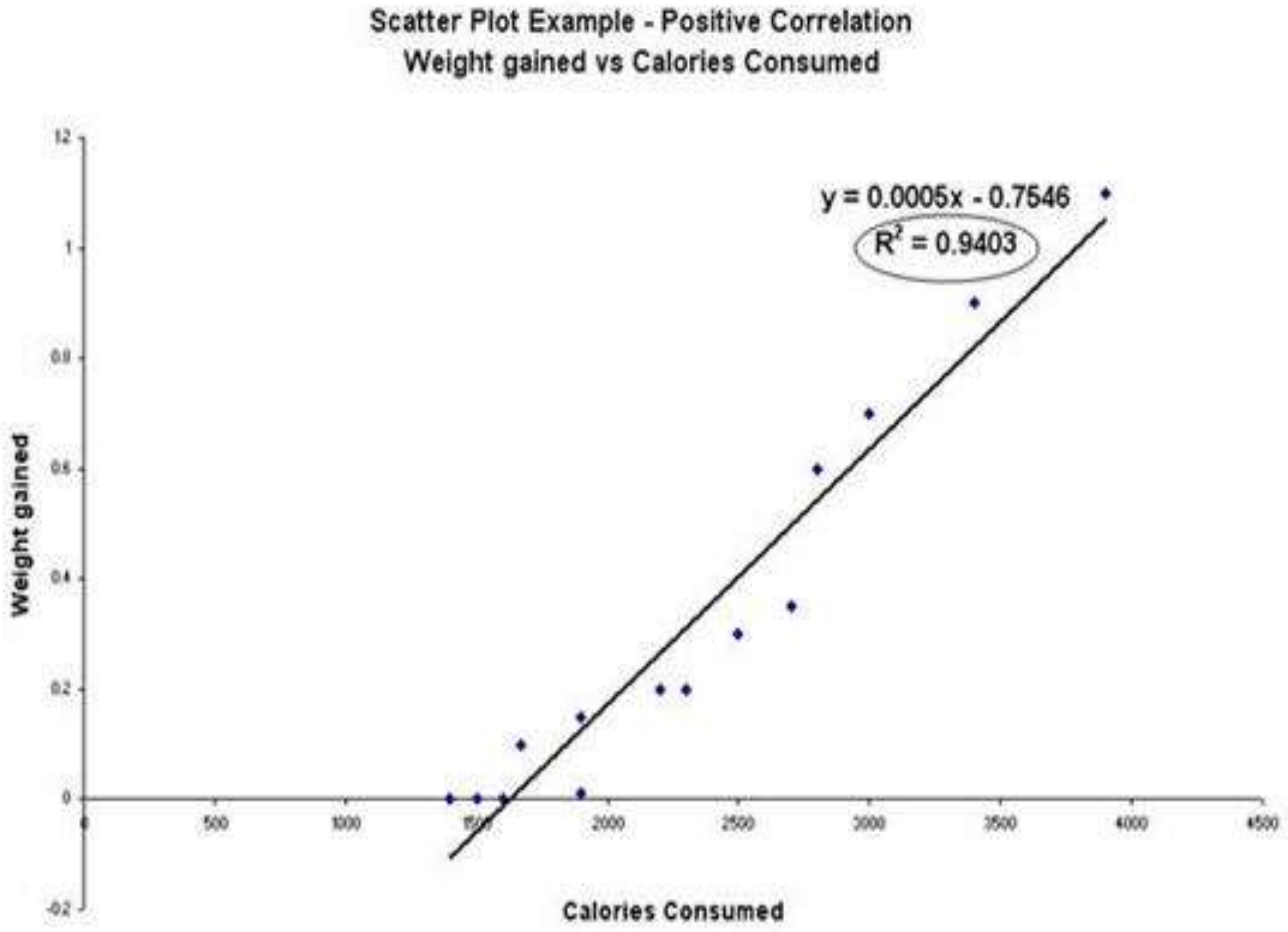
- When two variables move in the same direction then the correlation between these two variables is said to be PositiveCorrelation.

- When the value of one variable increases, the value of other value also increases at the same rate.

**For example** :

| | | | | | |
|---|---|---|---|---|---|
| Training( Rs.) | : | 350 | 360 | 370 | 380 |
| performance( Kg.) | : | 30 | 40 | 50 | 60 |

# Positive Correlation



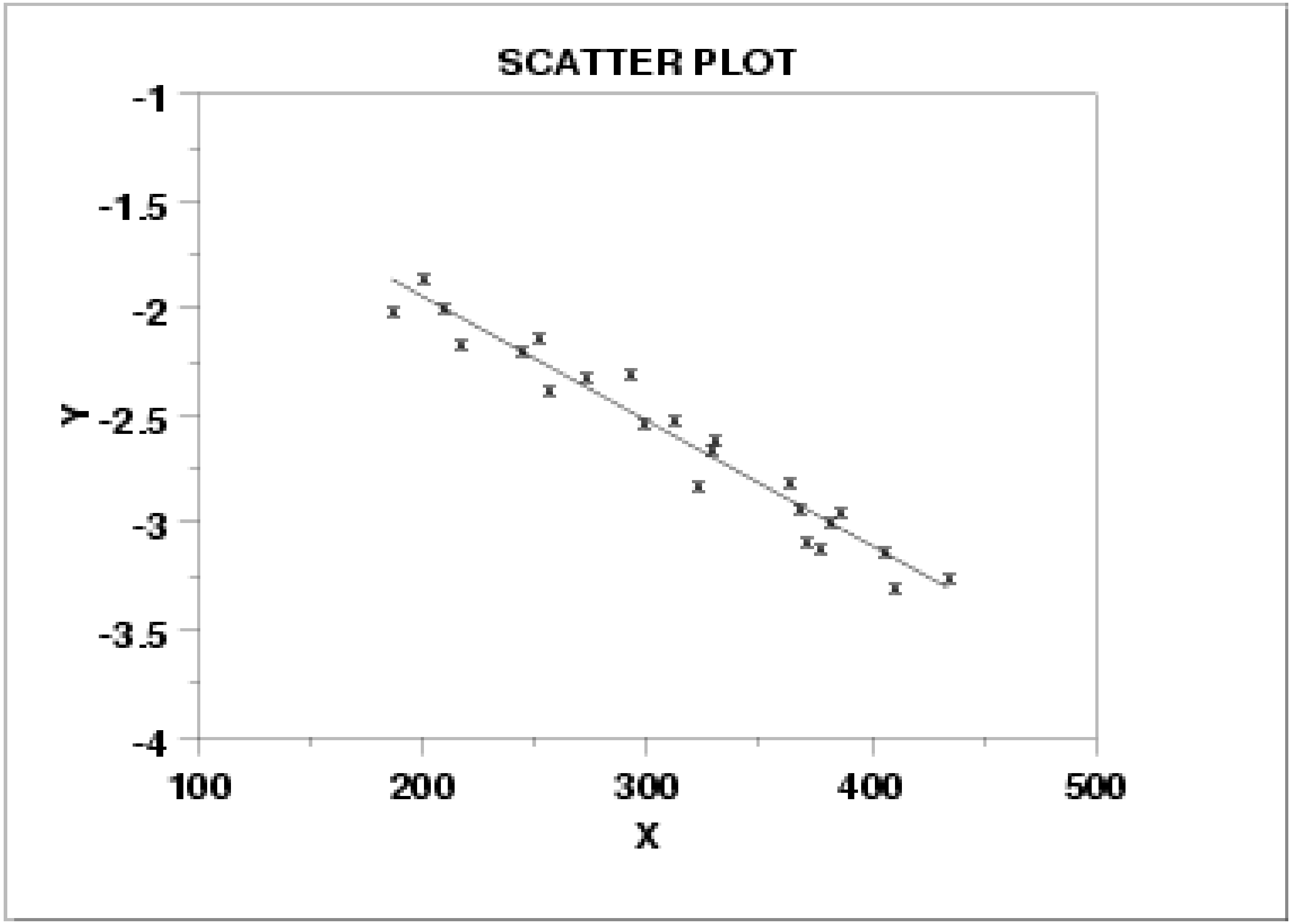
# Negative Correlation

- In this type of correlation, the two variables move in the opposite direction.

- When the value of one variable increases, the value of the other variable decreases.

  For example, the relationship between price and demand.

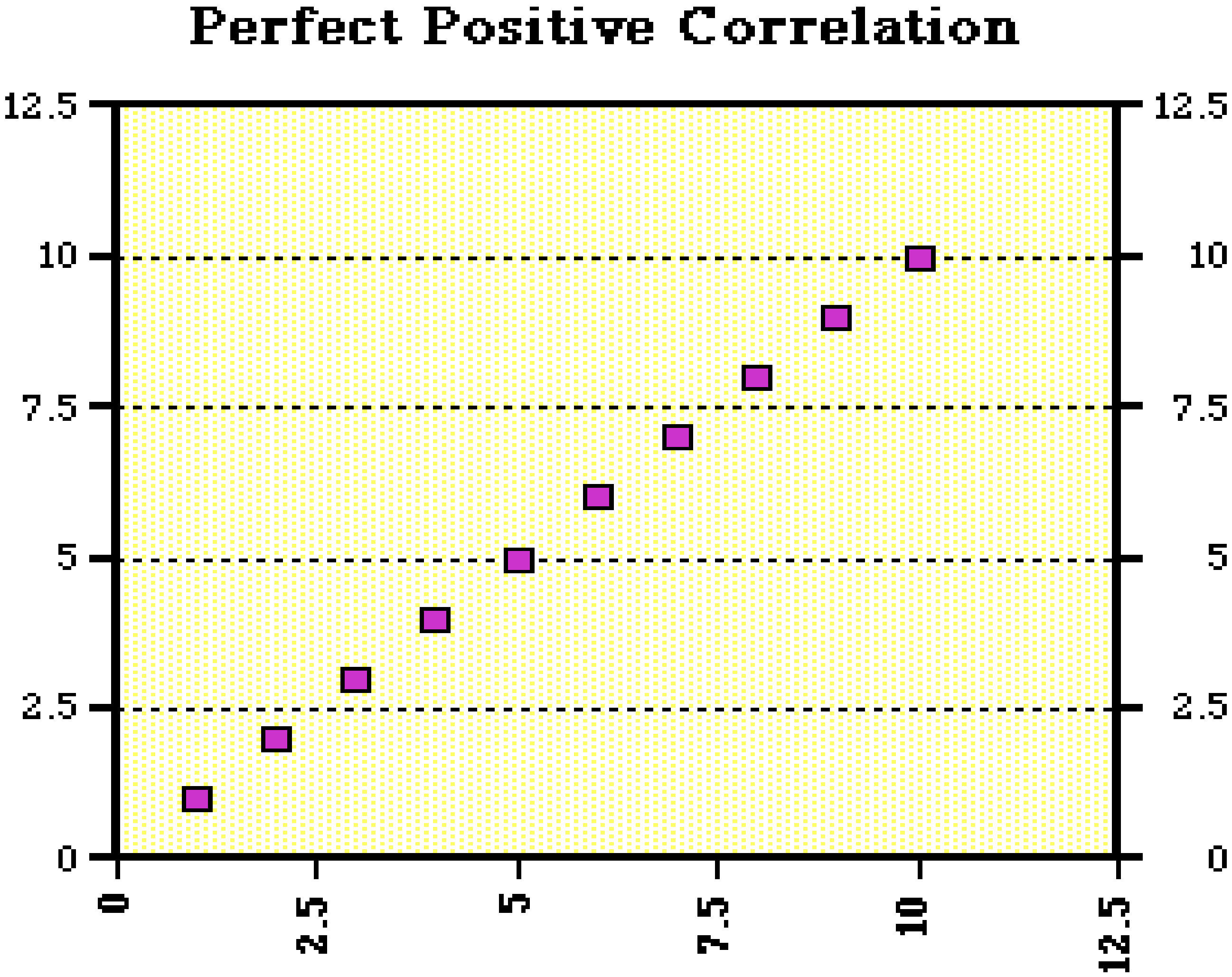
# Negative Correlation



# Perfect Positive Correlation

- When there is a change in one variable X, and if there is equal proportion of change in the other variable say Y in the same direction, then these two variables are said to have a Perfect Positive Correlation.

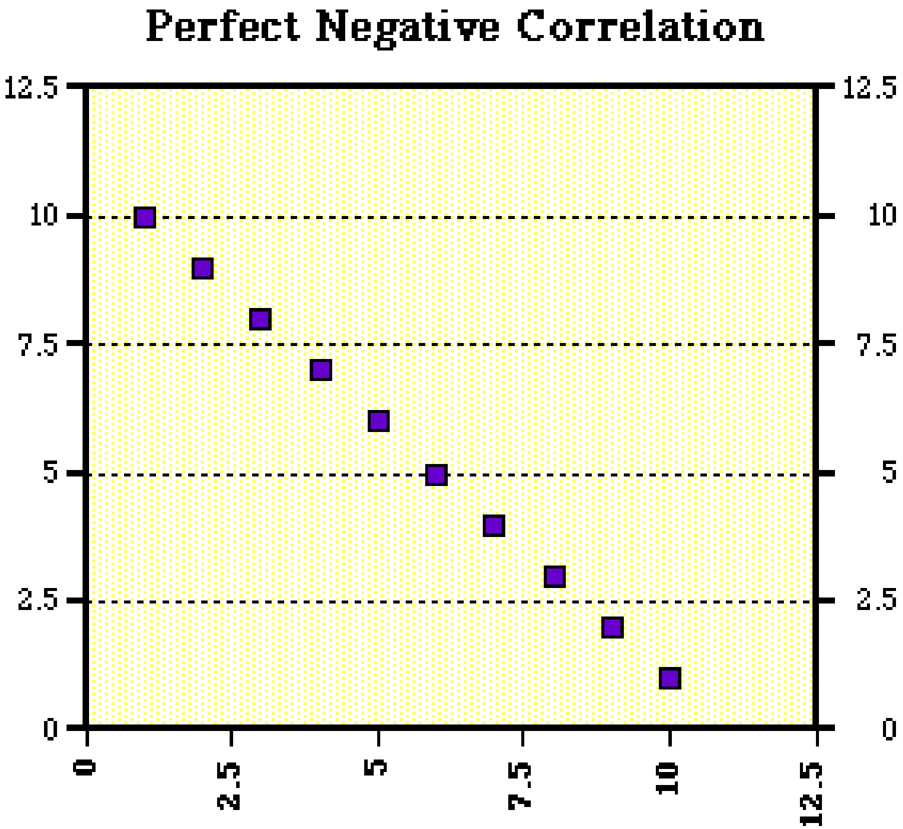
# Perfect Positive Correlation



# Perfectly Negative Correlation

- Between two variables X and Y, if the change in X causes the same amount of change in Y in equal proportion but in opposite direction, then this correlation is called as Perfectly Negative Correlation.

# Perfectly Negative Correlation



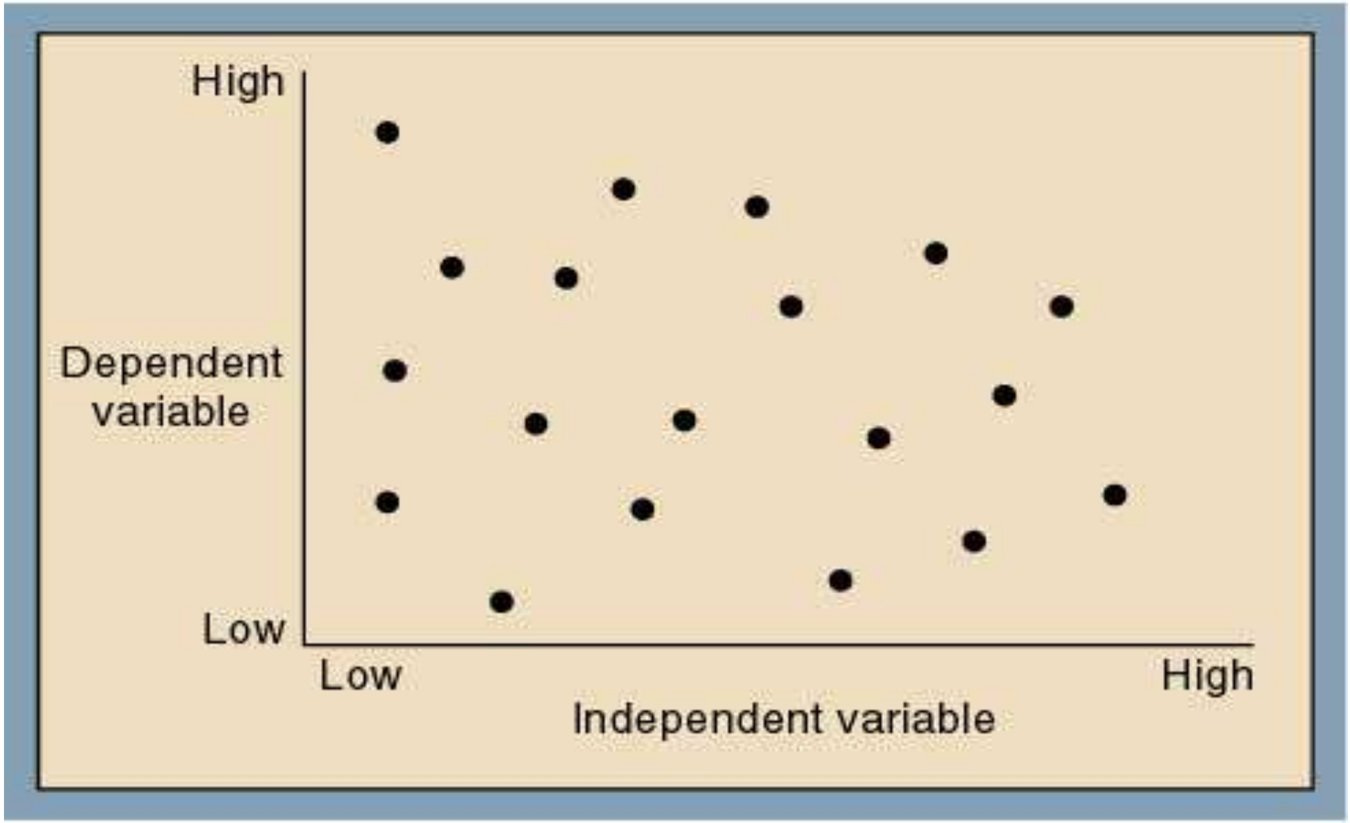
# Zero Correlation

- When the two variables are independent and the change in one variable has no effect in other variable,

  then the correlation between these two variable is known as Zero Correlation.

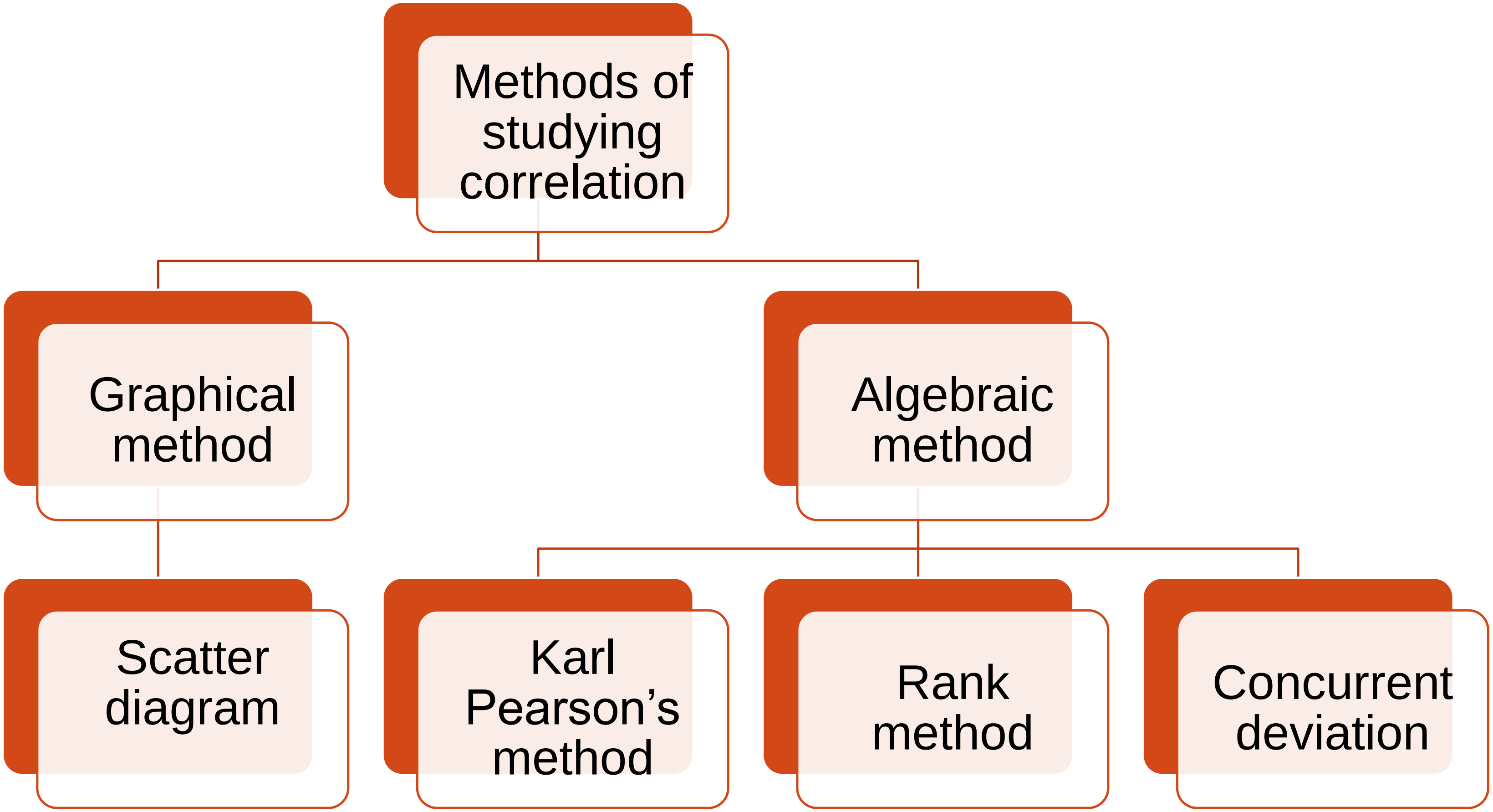
# Zero Correlation

**A Zero Correlation**



A zero correlation indicates that there is no relationship between the independent variable and the dependent variable.

# Methods of studying correlation



Methods of studying correlation

Graphical method

Algebraic method

Scatter diagram

Karl Pearson's method

Rank method

Concurrent deviation

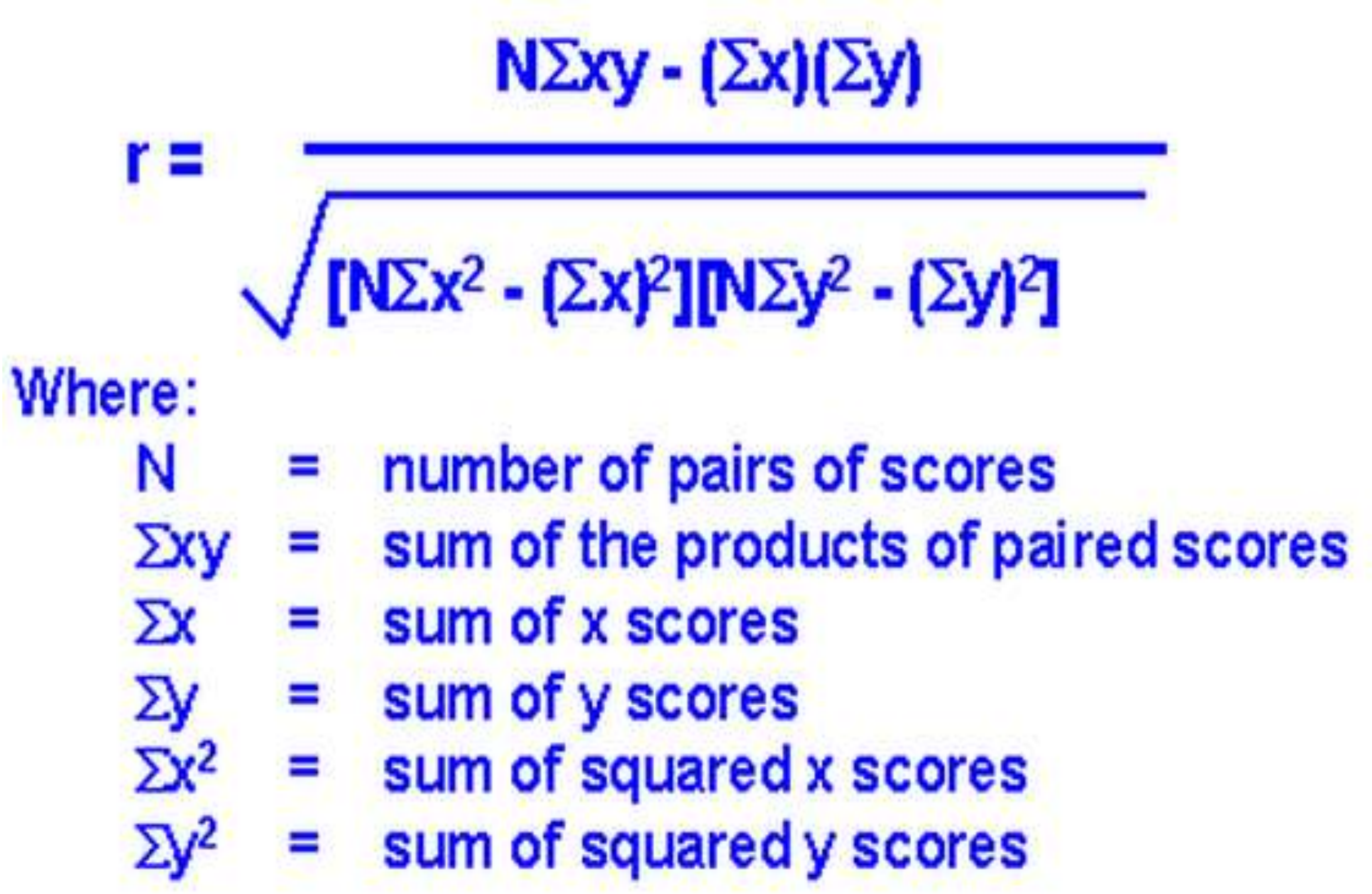
# Karl Pearson's Coefficient of Correlation

- It is the most widely used method of measuring linear relationship between two variables.

- Assumptions of Karl Pearson's Coefficient:

1. There is linear relationship between variables.
2. There is cause and effect relationship

# Calculating the Co-efficient of Correlation by Karl Pearson Method



$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where:

| | | |
|---|---|---|
| $N$ | = | number of pairs of scores |
| $\Sigma xy$ | = | sum of the products of paired scores |
| $\Sigma x$ | = | sum of x scores |
| $\Sigma y$ | = | sum of y scores |
| $\Sigma x^2$ | = | sum of squared x scores |
| $\Sigma y^2$ | = | sum of squared y scores |

# Example

- From the following sets of observations, find the coefficients of correlation:

(a) X : 25 35 (b) X : 8 11

Y : 40 41 Y : 190 100

Ans: (a) r = 1 (b) r = -1

So, in case (a) the variable X and Y are perfectly positive correlated to each other.

And in case (b) the variable X and Y are perfectly negative correlated to each other.

# Features of coefficient of correlation

- Ranges between -1 and 1.
- Closer to -1, stronger the negative relationship
- Closer to 1, stronger the positive relationship
- Closer to 0, weaker the relationship
- If $r=0$ there is no relationship between variable
- If $+0.75 \le r \le +1$ there exist high positive relationship.
- If $-0.75 \ge r \ge -1$ there exist high negative relationship.

# Scatter Diagram

- The first step in determining whether there is a relationship between two variable is to examine the graph of observed data.

- The graph or chart is called scatter diagram.

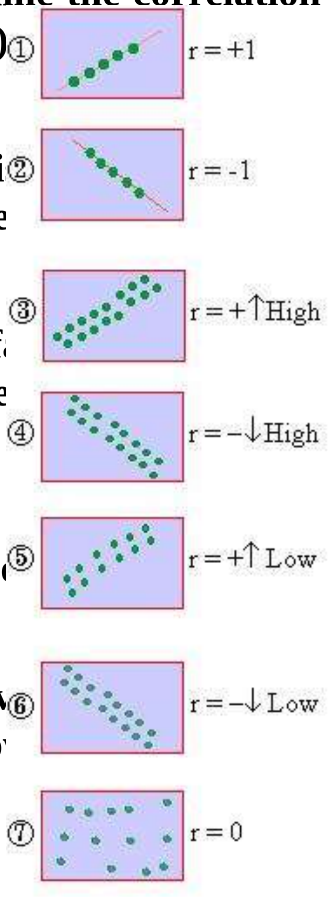
- A scatter diagram gives us information about patterns that indicates that variables are related.

# Scatter Plot ( Scatter diagram or dot diagram )

- In this method the values of the two variables are plotted on a graph paper. One is taken along the horizontal (x-axis) and the other along the vertical (y-axis).
- By plotting the data, we get points (dots) on the graph which are generally scattered and hence the name ‘Scatter Plot’.
- The points plotted on graph may cluster around a straight line or a curve or may not show any tendency of association.

i) If all points lie on a rising straight line the correlation is perfectly positive and r = +1 (see fig.1 )

ii) If all points lie on a falling straight line the correlation is perfectly negative and r = -1 (see fig.2)



**iii) If the points lie in narrow strip, ri**
**correlation is high degree of positive (see**

**iv) If the points lie in a narrow strip, f**
**the correlation is high degree of negative**

**v) If the points are spread widely over a**
**upwards, the correlation is low degree p**

**vi) If the points are spread widely ov**
**falling downward, the correlation is lo**
**(see fig.6)**

**vii) If the points are spread (scatte**
**specific pattern, the correlation is absent. i.e. r = 0. (see fig.7)**

# Scatter diagram continue…

- A scatter diagram of the data helps in having a visual idea about the nature of association between two variables.
- If the point cluster along the straight line the association between variable is linear.
- If the points cluster along the a curve, the association is non-linear or curvilinear.
- If the points neither cluster along a straight line nor along a curve, there is absence of any association between the variables.
- When the low/high value of one variable is associated with low/high value of other variable respectively, the association is called positive.
- In contrast if low/high value of one variable is associated with high/low value of other variable respectively, the association is called negative.

# Example

- Draw a scatter diagram from following data and indicate whether the correlation between the variable is positive or negative.

| **Height (inch)** | **62** | **72** | **70** | **60** | **67** | **70** | **64** | **65** | **60** | **70** |
|---|---|---|---|---|---|---|---|---|---|---|
| Weight ( kgs.) | 50 | 65 | 63 | 52 | 56 | 60 | 59 | 58 | 54 | 65 |

# Standard Error

- Standard error of coefficient of correlation is used to find out probable error of coefficient of correlation.
- Where r = coefficient of correlation
- N = Number of observed pairs
- So S.E. = $1-r^2 / \sqrt{N}$

# Probable Error

- The probable error of coefficient of correlation is an amount which if added to or subtracted from values of r gives upper limit and lower limit within which this coefficient is expected to be.
- Probable error is 0.6745 time of Standard Error
- That means Probable error = 0.6745 (S.E.)

# Use of probable error

- It is used to determine the reliability of coefficient of correlation.
- For ex. If ratio of r and P.E. is greater than 6 then coefficient is reliable, i.e. there is relationship between variable.
- If ratio of r and PE is less than 6 then coefficient is not reliable, i.e. there is no relationship between variable.

# Example

- If r = 0.8 and N = 36, find

(a) Standard Error

(b) Probable Error

(c) Check reliability

Ans. (a) 0.06

(b) 0.04

(c) ratio of r to PE is 20 so coefficient is reliable