HARITZONE

CORRELATION

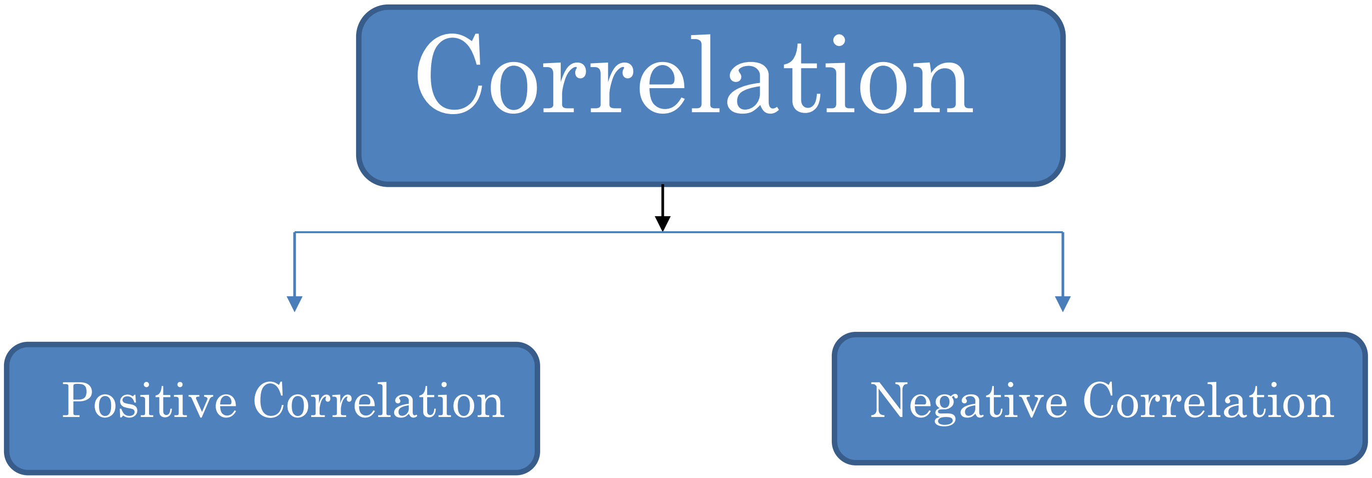
CORRELATION

- Correlation is a statistical tool that helps to measure and analyze the degree of relationship between two variables.
- Correlation analysis deals with the association between two or more variables.

CORRELATION

- The degree of relationship between the variables under consideration is measure through the correlation analysis.
- The measure of correlation called the correlation coefficient .
- The degree of relationship is expressed by coefficient which range from correlation **$(-1 \leq r \geq +1)$**
- The direction of change is indicated by a sign.
- The correlation analysis enable us to have an idea about the degree & direction of the relationship between the two variables under study.

Types of Correlation - Type I



Types of Correlation Type I

- **Positive Correlation:** The correlation is said to be positive correlation if the values of two variables changing with same direction.
  Ex. Pub. Exp. & Sales, Height & Weight.

- **Negative Correlation:** The correlation is said to be negative correlation when the values of variables change with opposite direction.
  Ex. Price & Quantity demanded.

Direction of the Correlation

- **Positive relationship** – Variables change in the same direction.
  - As X is increasing, Y is increasing
  - As X is decreasing, Y is decreasing
  - E.g., As height increases, so does weight.

Indicated by sign; (+) or (-).

- **Negative relationship** – Variables change in opposite directions.
  - As X is increasing, Y is decreasing
  - As X is decreasing, Y is increasing
  - E.g., As TV time increases, grades decrease

Examples

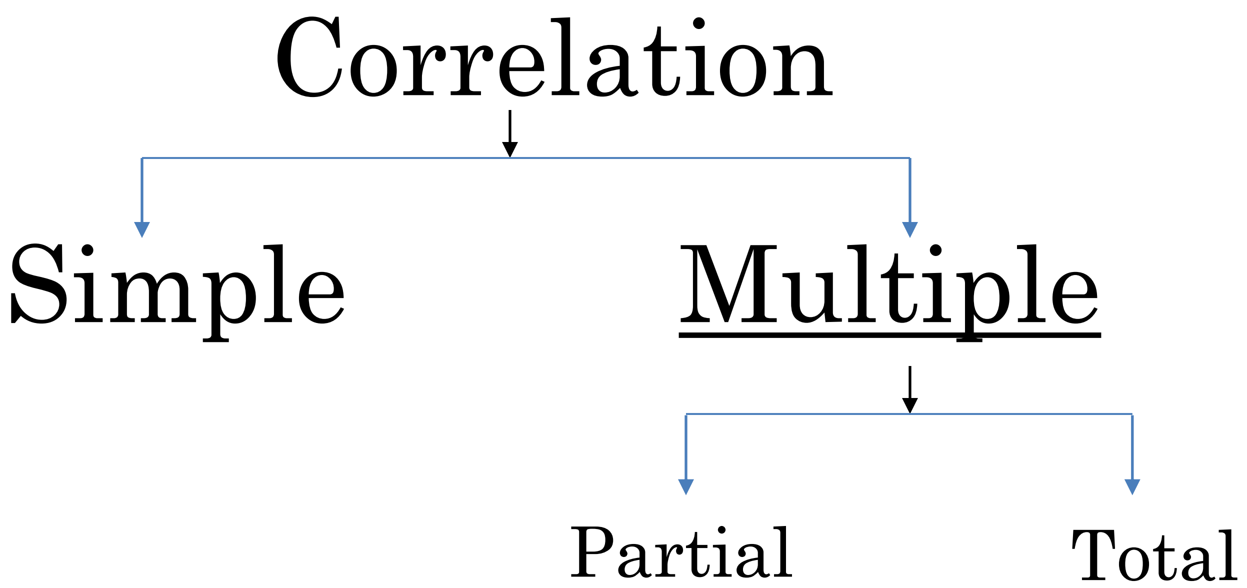
Positive Correlation

- Water consumption and temperature.
- Study time and grades.

Negative Correlation

- Alcohol consumption and driving ability.
- Price & quantity demanded

Types of Correlation Type II



Types of Correlation Type II

- **Simple correlation:** Under simple correlation problem there are only two variables are studied.
- **Multiple Correlation:** Under Multiple Correlation three or more than three variables are studied. Ex. $Q_d = f(P, P_C, P_S, t, y)$
- **Partial correlation:** analysis recognizes more than two variables but considers only two variables keeping the other constant.
- **Total correlation:** is based on all the relevant variables, which is normally not feasible.

Types of Correlation Type III

Correlation

LINEAR

NON LINEAR

Types of Correlation Type III

- **Linear correlation:** Correlation is said to be linear when the amount of change in one variable tends to bear a constant ratio to the amount of change in the other. The graph of the variables having a linear relationship will form a straight line.

  Ex X = 1, 2, 3, 4, 5, 6, 7, 8,

  Y = 5, 7, 9, 11, 13, 15, 17, 19,

  $Y = 3 + 2x$

- **Non Linear correlation:** The correlation would be non linear if the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable.

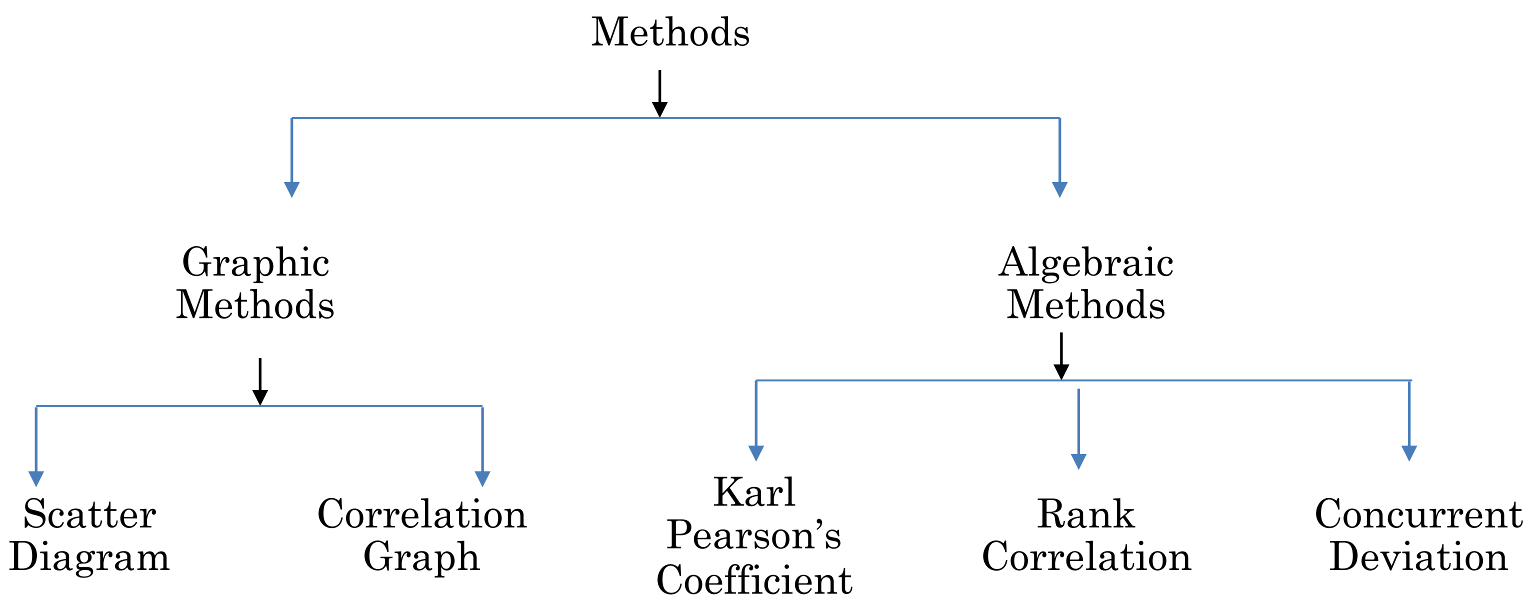
Correlation & Causation

- Causation means cause & effect relation.
- Correlation denotes the interdependency among the variables for correlating two phenomenon, it is essential that the two phenomenon should have cause-effect relationship,& if such relationship does not exist then the two phenomenon can not be correlated.
- If two variables vary in such a way that movement in one are accompanied by movement in other, these variables are called cause and effect relationship.
- Causation always implies correlation but correlation does not necessarily implies causation.

Degree of Correlation

- Perfect Correlation
- High Degree of Correlation
- Moderate Degree of Correlation
- Low Degree of Correlation
- No Correlation

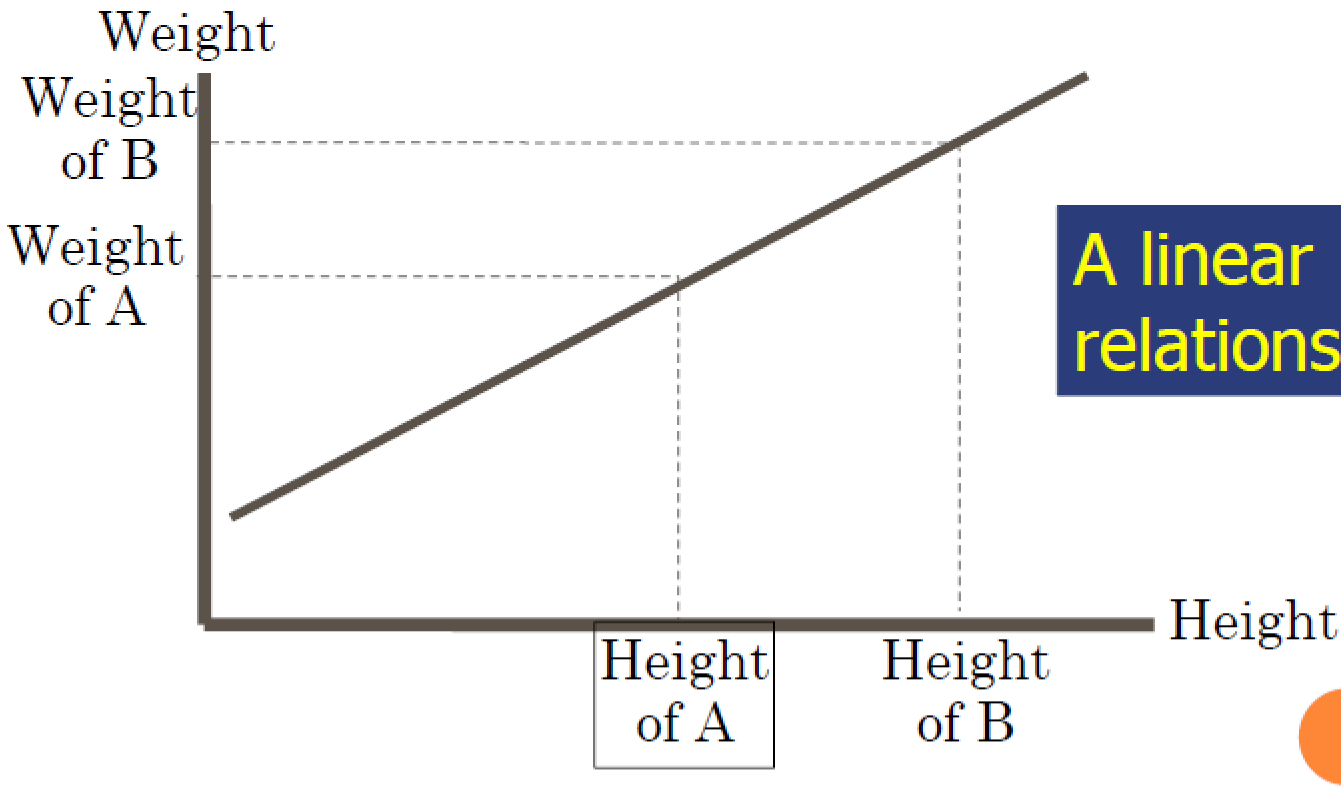
Methods of Studying Correlation



Scatter Diagram Method

- Scatter Diagram is a graph of observed plotted points where each points represents the values of X & Y as a coordinate.
- It portrays the relationship between these two variables graphically.

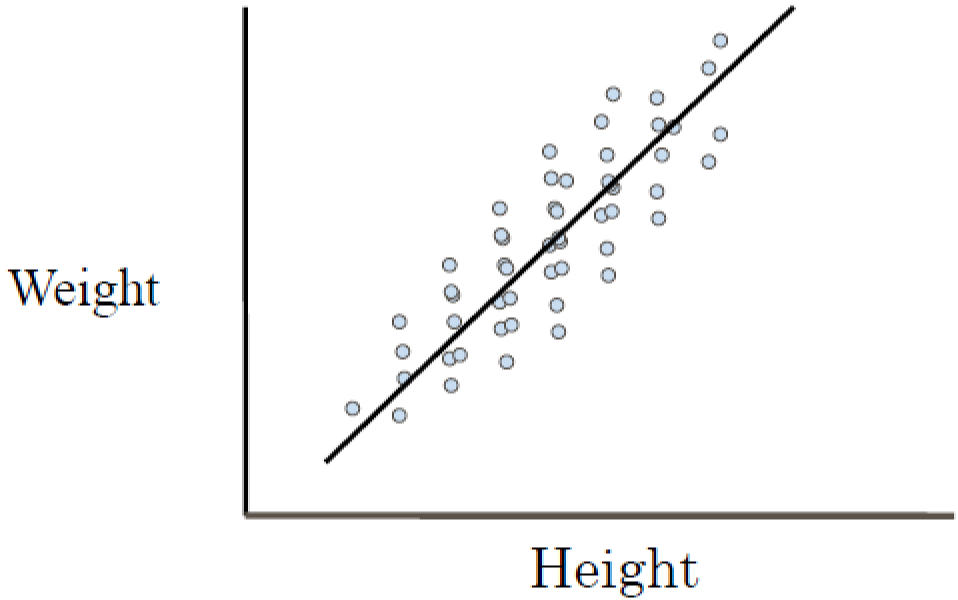
A PERFECT POSITIVE CORRELATION



A linear relationship

HIGH DEGREE OF POSITIVE CORRELATION

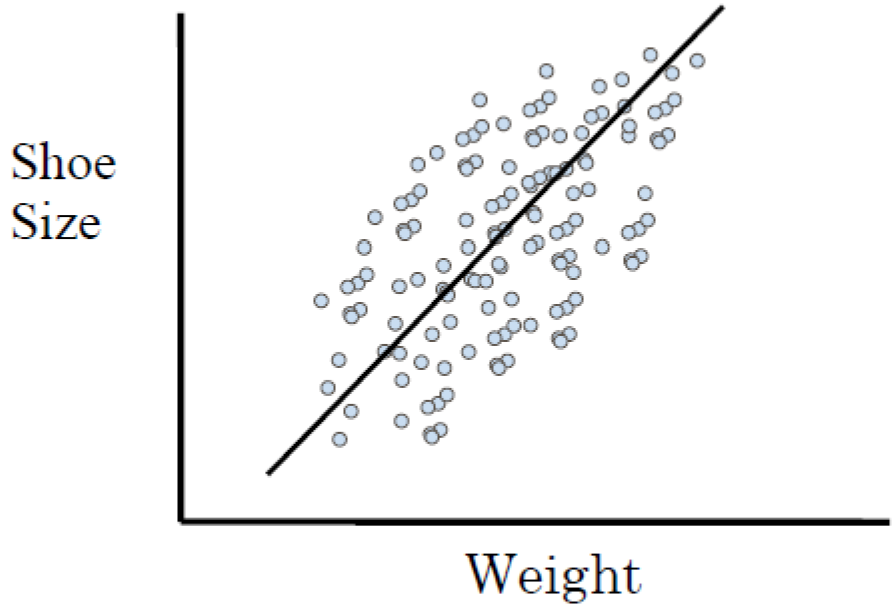
- Positive relationship



r = +.80

Degree of correlation

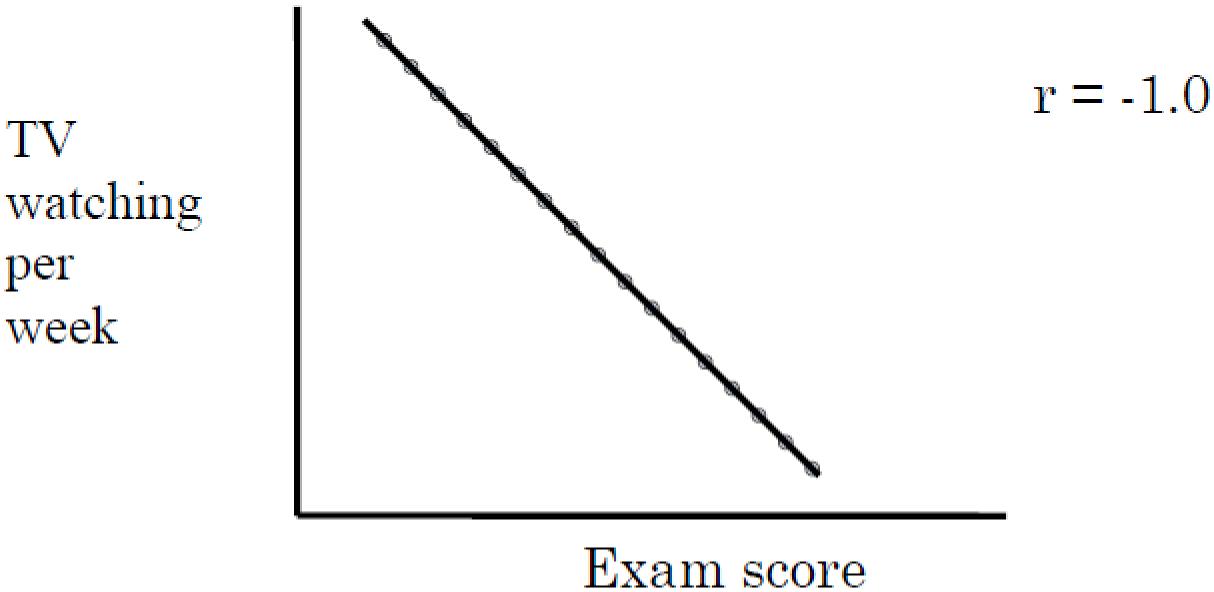
- **Moderate Positive Correlation**



$r = + 0.4$

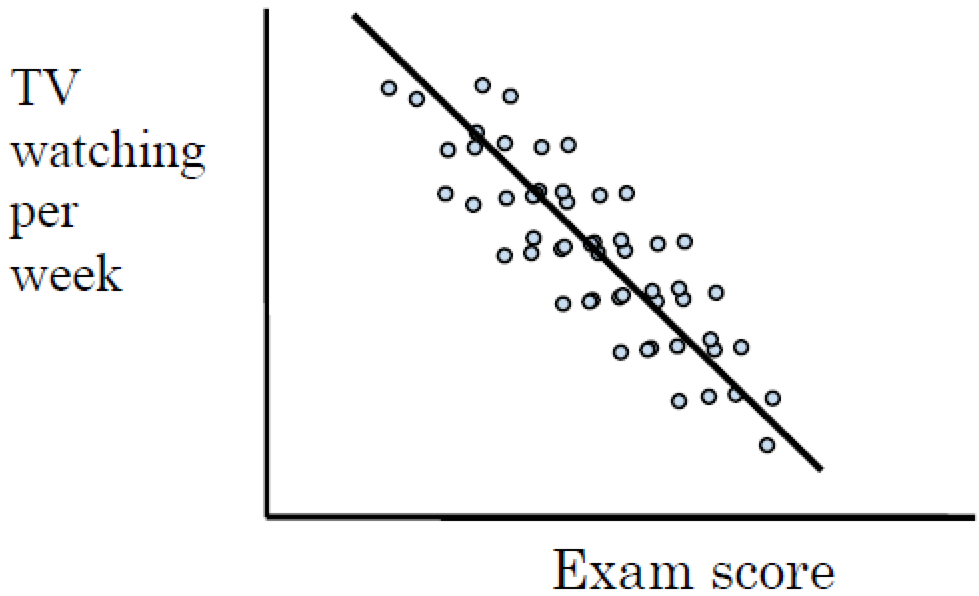
Degree of correlation

- **Perfect Negative Correlation**



Degree of correlation

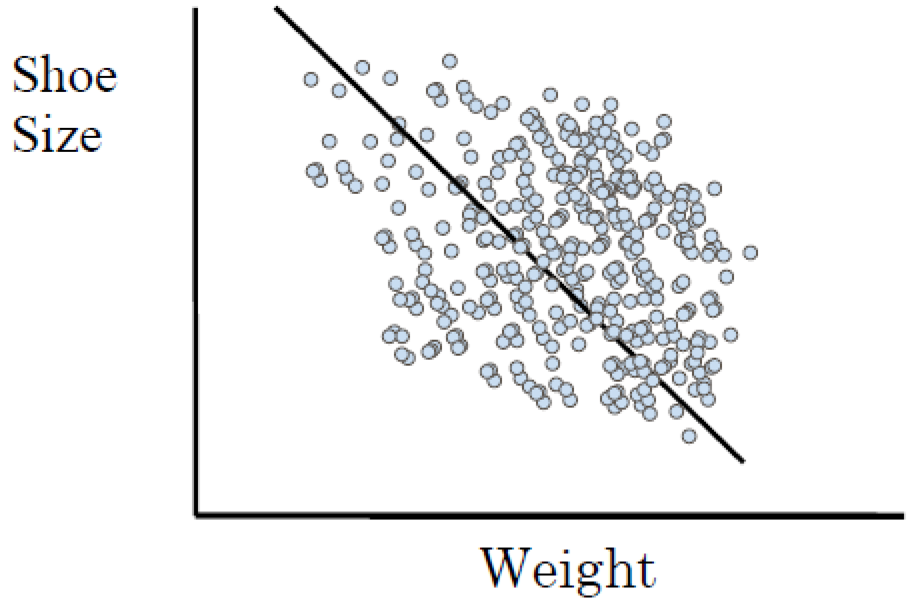
- Moderate Negative Correlation



r = -.80

Degree of correlation

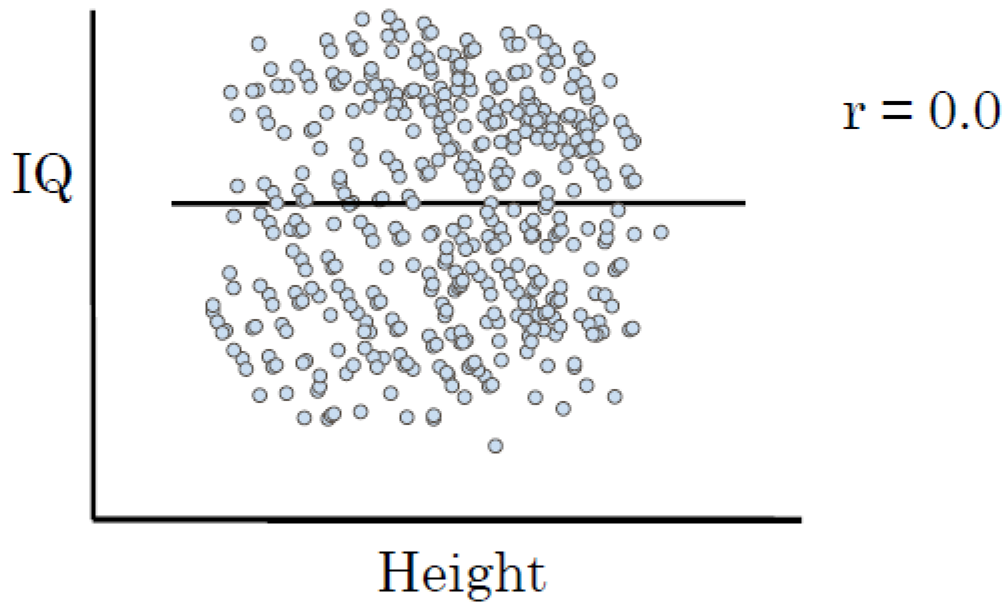
- **Weak negative Correlation**



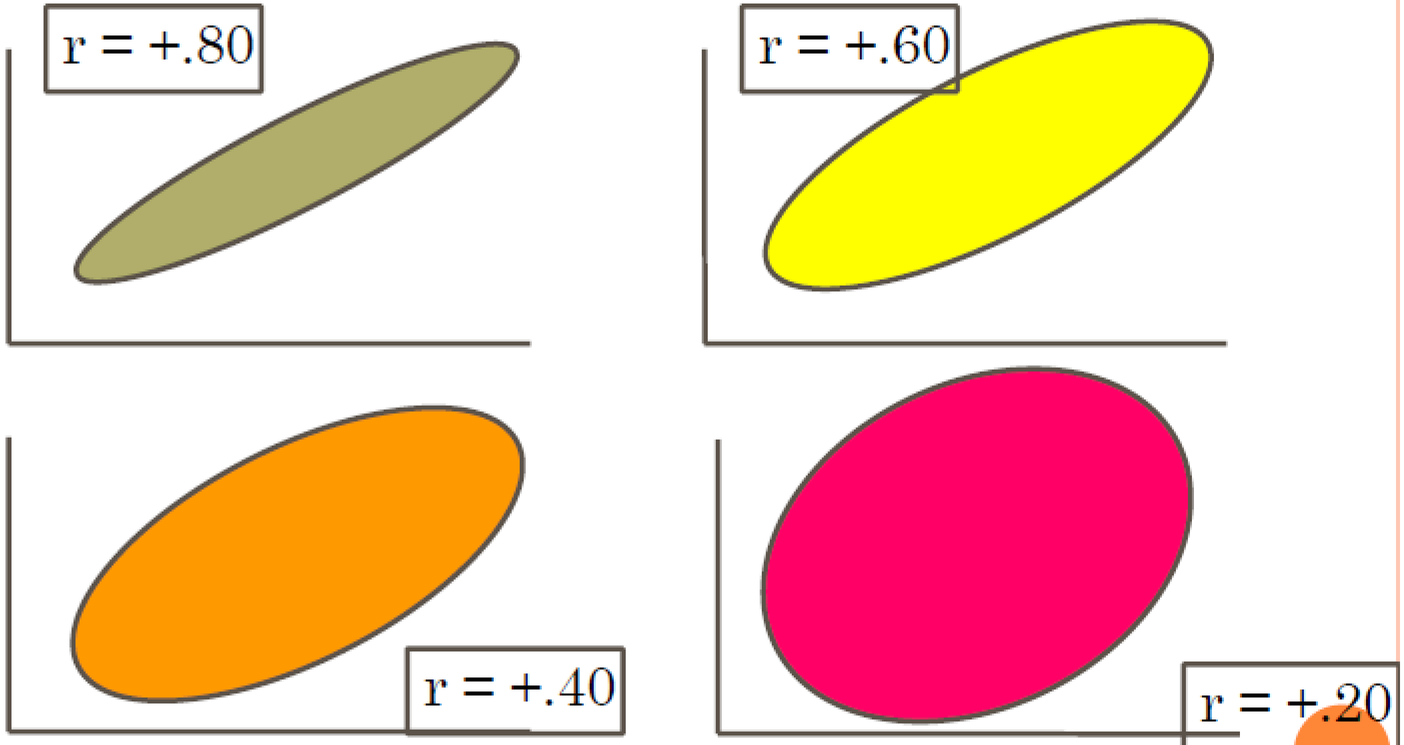
r = - 0.2

Degree of correlation

- No Correlation (horizontal line)



Degree of correlation (r)



Direction of the Relationship

- **Positive relationship** – Variables change in the same direction.
  - As X is increasing, Y is increasing
  - As X is decreasing, Y is decreasing
- E.g., As height increases, so does weight.

Indicated by sign; (+) or (-).

- **Negative relationship** – Variables change in opposite directions.
  - As X is increasing, Y is decreasing
  - As X is decreasing, Y is increasing
- E.g., As TV time increases, grades decrease

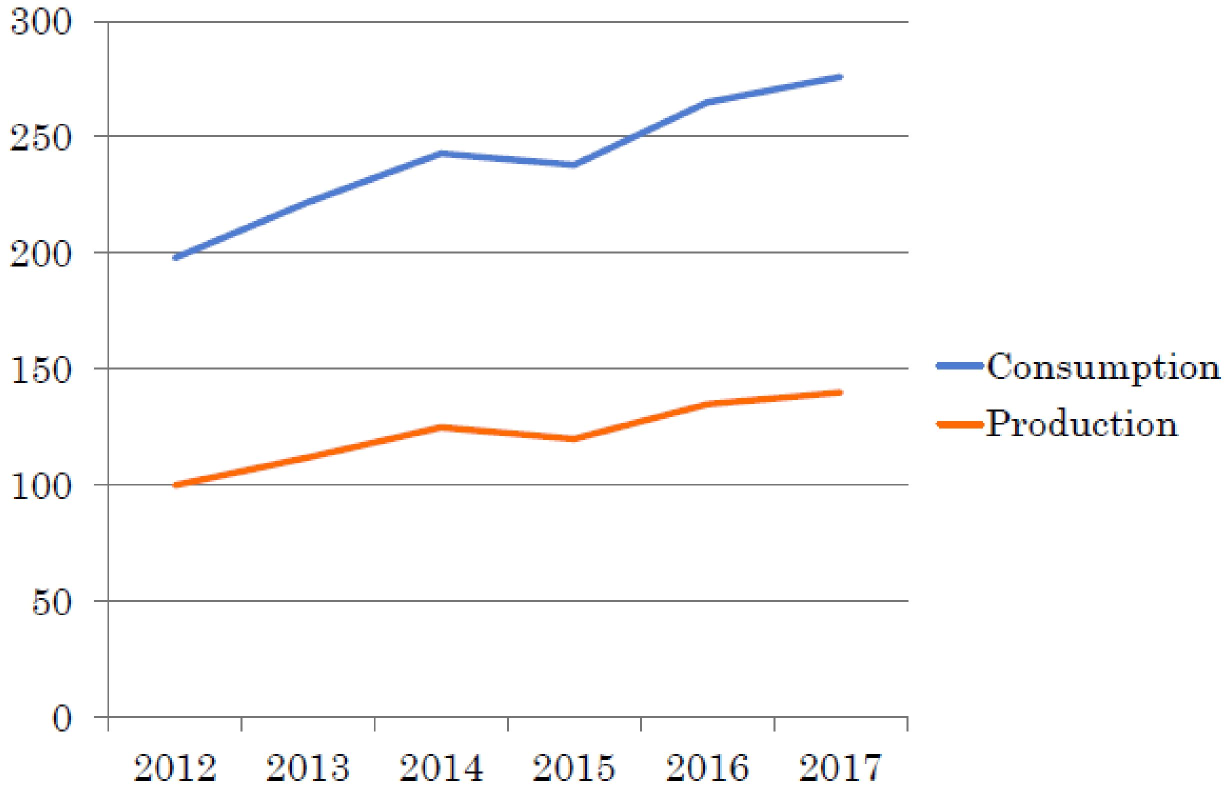
Advantages of Scatter Diagram

- Simple & Non Mathematical method
- Not influenced by the size of extreme item
- First step in investing the relationship between two variables

Disadvantage of scatter diagram

Can not adopt the an exact degree of correlation

Correlation Graph



Karl Pearson's Coefficient of Correlation

- It is quantitative method of measuring correlation
- This method has been given by Karl Pearson
- It's the best method

Calculation of Coefficient of Correlation – Actual Mean Method

- Formula used is:
  - $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$ where $x = X - \bar{X}$; $y = Y - \bar{Y}$

Q1: Find Karl Pearson's coefficient of correlation:

X	2	3	4	5	6	7	8
Y	4	7	8	9	10	14	18

Ans: 0.96

Q2: Find Karl Pearson's coefficient of correlation:

	X- Series	Y-series
No. of items	15	15
AM	25	18
Squares of deviations from mean	136	138

Summation of product of deviations of X & Y series from their respective arithmetic means = 122

Ans: 0.89

PRACTICE PROBLEMS - CORRELATION

Q3: Find Karl Pearson's coefficient of correlation:

X	6	2	10	4	8
Y	9	11	?	8	7

Arithmetic Means of X & Y are 6 & 8 respectively. Ans: – 0.92

Q4: Find the number of items as per the given data:

$r = 0.5, \Sigma xy = 120, \sigma_y = 8, \Sigma x^2 = 90$

where x & y are deviations from arithmetic means

Ans: 10

Q5: Find r:

$\Sigma X = 250, \Sigma Y = 300, \Sigma(X - 25)^2 = 480, \Sigma(Y - 30)^2 = 600$

$\Sigma(X - 25)(Y - 30) = 150$, $N = 10$ Ans: 0.28

Calculation of Coefficient of Correlation – Assumed Mean Method

- Formula used is:
  - $$r = \frac{N.\Sigma dxdy - \Sigma dx.\Sigma dy}{\sqrt{N.\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N.\Sigma dy^2 - (\Sigma dy)^2}}$$

Q6:Find r:

X	10	12	18	16	15	19	18	17
Y	30	35	45	44	42	48	47	46

Ans: 0.98

Q7: Find r, when deviations of two series from assumed mean are as follows: *Ans: 0.895*

Dx	+5	-4	-2	+20	-10	0	+3	0	-15	-5
Dy	+5	-12	-7	+25	-10	-3	0	+2	-9	-15

Calculation of Coefficient of Correlation – Actual Data Method

- Formula used is:

  - $r = \dfrac{N.\Sigma XY - \Sigma X.\Sigma Y}{\sqrt{N.\Sigma X^2 - (\Sigma X)^2}\sqrt{N.\Sigma Y^2 - (\Sigma Y)^2}}$

Q8:Find r:

X	10	12	18	16	15	19	18	17
Y	30	35	45	44	42	48	47	46

Ans: 0.98

Q9: Calculate product moment correlation coefficient from the following data: *Ans: 0.996*

X	-5	-10	-15	-20	-25	-30
Y	50	40	30	20	10	5

Important Typical Problems

Q10: Calculate the coefficient of correlation from the following data and interpret the result: *Ans: 0.76*

$N = 10$, $\Sigma XY = 8425$, $\bar{X} = 28.5$, $\bar{Y} = 28.0$, $\sigma x = 10.5$, $\sigma y = 5.6$

Q11: Following results were obtained from an analysis:

$N = 12$, $\Sigma XY = 334$, $\Sigma X = 30$, $\Sigma Y = 5$, $\Sigma X^2 = 670$, $\Sigma Y^2 = 285$

Later on it was discovered that one pair of values (X = 11, Y = 4) were wrongly copied. The correct value of the pair was (X = 10, Y = 14). Find the correct value of correlation coefficient. Ans: 0.774

Variance – Covariance Method

- This method of determining correlation coefficient is based on covariance.
- $r = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{Cov(X,Y)}{\sigma_x . \sigma_y}$

  where $Cov(X,Y) = \frac{\Sigma xy}{N} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{N} = \frac{\Sigma XY}{N} - \bar{X}\bar{Y}$
- Another Way of calculating $r = \frac{\Sigma xy}{N.\,\sigma_x . \sigma_y .}$

Q12: For two series X & Y, Cov(X,Y) = 15, Var(X)=36, Var (Y)=25. Find r. Ans: 0.5

Q13: Find r when $N = 30$, $\bar{X} = 40$, $\bar{Y} = 50$, $\sigma_x = 6$, $\sigma_y = 7$, $\Sigma xy = 360$

Ans: 0.286

Q14: For two series X & Y, Cov(X,Y) = 25, Var(X)=36, r = 0.6. Find σ_y. Ans: 6.94

Calculation of Correlation Coefficient – Grouped Data

- Formula used is:
  - $$r = \frac{N.\Sigma f dxdy - \Sigma f dx.\Sigma f dy}{\sqrt{N.\Sigma f dx^2 - (\Sigma f dx)^2}\sqrt{N.\Sigma f dy^2 - (\Sigma f dy)^2}}$$

Q15: Calculate Karl Pearson's coefficient of correlation:

X / Y	10-25	25-40	40-55
0-20	10	4	6
20-40	5	40	9
40-60	3	8	15

Ans: 0.33

Properties of Coefficient of Correlation

- Karl Pearson's coefficient of correlation lies between -1 & 1, i.e. $-1 \leq r \leq +1$
- If the scale of a series is changed or the origin is shifted, there is no effect on the value of 'r'.
- 'r' is the geometric mean of the regression coefficients b_{yx} & b_{xy}, i.e. $r = \sqrt{b_{xy} \cdot byx}$
- If X & Y are independent variables, then coefficient of correlation is zero but the converse is not necessarily true.
- 'r' is a pure number and is independent of the units of measurement.
- The coefficient of correlation between the two variables x & y is symmetric. i.e. $r_{yx} = r_{xy}$

Probable Error & Standard Error

- Probable Error is used to test the reliability of Karl Pearson's correlation coefficient.
- Probable Error (P.E.) = 0.6745 x $\frac{1-r^2}{\sqrt{N}}$
- Probable Error is used to interpret the value of the correlation coefficient as per the following:
  - If $|r| > 6$ P.E., then 'r' is significant.
  - If $|r| < 6$ P.E., then 'r' is insignificant. It means that there is no evidence of the existence of correlation in both the series.
- Probable Error also determines the upper & lower limits within which the correlation of randomly selected sample from the same universe will fall.
  - Upper Limit = r + P.E.
  - Lowe Limit = r – P.E.

Practice Problem – Probable Error

Q16: Find Karl Pearson's coefficient of correlation from the following data:

X	9	28	45	60	70	50
Y	100	60	50	40	33	57

Also calculate probable error and check whether it is significant or not. Ans: – 0.94, 0.032

Q17: A student calculates the value of r as 0.7 when N = 5. He concludes that r is highly significant. Comment. Ans: Insignificant

Spearman's Rank Correlation Method

- Given by Prof. Spearman in 1904
- By this method, correlation between qualitative aspects like intelligence, honesty, beauty etc. can be calculated.
- These variables can be assigned ranks but their quantitative measurement is not possible.
- It is denoted by $\mathbf{R = 1 - \frac{6\,\Sigma D^2}{N\,(N^2 - 1)}}$
  - R = Rank correlation coefficient
  - D = Difference between two ranks $(R_1 - R_2)$
  - N = Number of pair of observations
- As in case of r, $-1 \leq R \leq 1$
- *The sum total of Rank Difference is always equal to zero. i.e. $\Sigma D = 0$.*

Three Cases

- Spearman's Rank Correlation Method
  - When ranks are given
  - When ranks are not given
  - When equal or tied ranks exist

Practice Problems – Rank Correlation (When Ranks are Given)

Q18: In a fancy dress competition, two judges accorded the following ranks to eight participants:

Judge X	8	7	6	3	2	1	5	4
Judge Y	7	5	4	1	3	2	6	8

Calculate the coefficient of rank correlation. Ans: .62

Q19: Ten competitors in a beauty contest are ranked by three judges X, Y, Z:

X	1	6	5	10	3	2	4	9	7	8
Y	3	5	8	4	7	10	2	1	6	9
Z	6	4	9	8	1	2	3	10	5	7

Use the rank correlation coefficient to determine which pair of judges has the nearest approach to common tastes in beauty.

Ans: X & Z

Q20: Find out the coefficient of Rank Correlation between X & Y:

X	15	17	14	13	11	12	16	18	10	9
Y	18	12	4	6	7	9	3	10	2	5

Ans: 0.48

Practice Problems – Rank Correlation (When Ranks are Equal or Tied)

- When two or more items have equal values in a series, so common ranks i.e. average of the ranks are assigned to equal values.
- Here **R** = 1 – $\frac{6\left[\Sigma D^2 + \frac{m^3 - m}{12} + \frac{m^3 - m}{12} + \ldots\ldots\ldots\ldots\right]}{N\,(N^2 - 1)}$
  - m = No. of items of equal ranks
  - The correction factor of $\frac{m^3 - m}{12}$ is added to ΣD^2 for such number of times as the cases of equal ranks in the question

Practice Problems – Rank Correlation (When Ranks are Equal or Tied)

Q21: Calculate R:

X	15	10	20	28	12	10	16	18
Y	16	14	10	12	11	15	18	10

Ans: – 0.37

Q22: Calculate Rank Correlation:

X	40	50	60	60	80	50	70	60
Y	80	120	160	170	130	200	210	130

Ans: 0.43

IMPORTANT TYPICAL PROBLEMS – RANK CORRELATION

Q23: Calculate Rank Correlation from the following data:

Ans: 0.64

Serial No.	1	2	3	4	5	6	7	8	9	10
Rank Difference	-2	?	-1	+3	+2	0	-4	+3	+3	-2

Q24: The coefficient of rank correlation of marks obtained by 10 students in English & Math was found to be 0.5. It was later discovered that the difference in the ranks in two subjects was wrongly taken as 3 instead of 7. Find the correct rank correlation.

Ans: 0.26

Q25: The rank correlation coefficient between marks obtained by some students in English & Math is found to be 0.8. If the total of squares of rank differences is 33, find the number of students.

Ans: 10