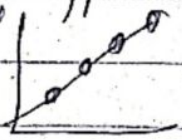
Methods of measuring correlation — Correlation

- Graphical Methods
  - Scatter diagram
  - Correlation Graph
- Algebraic Methods
  - Karl pearson's coefficient of correlation.
  - (2) Spearman's Rank difference Method
  - (3) Concurrent Deviation Method
  - (4) Two-way frequency table Method

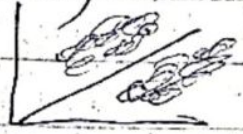
a) Scatter diagram - There are ~~two~~ types.

a) Perfect positive correlation



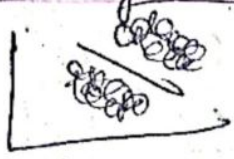
b) Highly positive - If all values are very near to straight line in upward direction, then it is highly positive.

c) Positive correlation - Points are near to the straight line (but not very near) the correlation is positive.



d) Perfect negative - lies in a straight line in downward direction (left top to right bottom), then negative ($r = -1$)

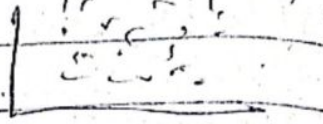
e) High negative - very close to straight line in downward direction,



Correlation -

If Negative - points are close to straight line (not very close) in downward

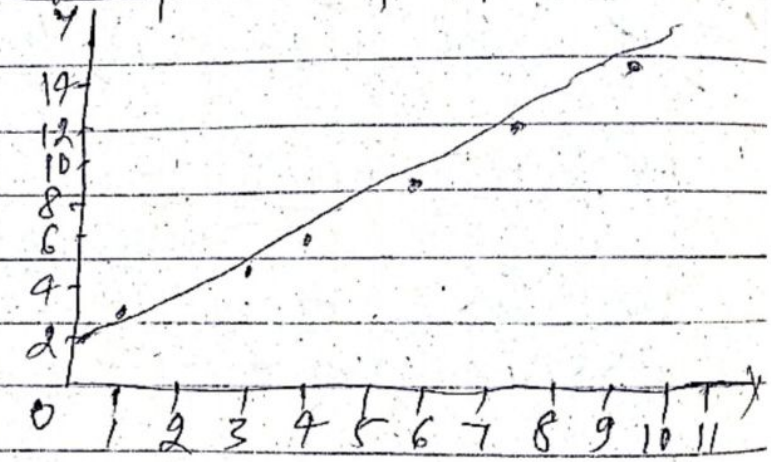
If Zero correction - points are widely scattered in graph.



eg - Draw scatter diagram & find correlation between variables -

X	1	3	4	6	8	9	10
Y	3	4	5	8	9	10	12

Soln -



From above diagram, we can say that the variables X & Y is positively correlation because all the points are near to line

By Correlation Graph -

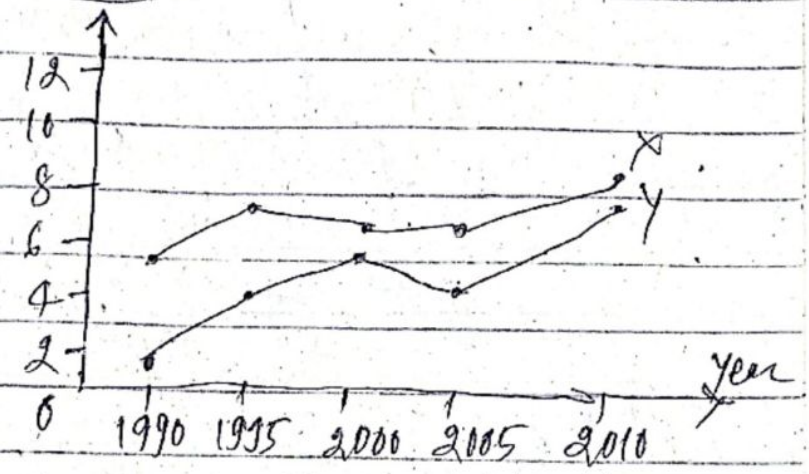
In this method we use individual values of two variables, which are plotted on the graph & we obtain two different curves on graph, by examination of point, we conclude that they will be correlated or not.

not scatter diagram

eg - Draw diagram & examine correlation betⁿ x & y

Year	1990	1995	2000	2005	2010
x	5	7	6	6	8
y	1	4	5	4	7

From diagram, we can say that variables are closely related to each other.



* Merits -
1) It is popular method of measuring the relationship betⁿ two variables.
2) Easy method without involving any mathematical calculations
3) Everyone can easily understood & examine it

* Demerits -
1) we cannot obtain degree of correlation
2) Suitable for only small no. of data.

(3) Karl's Pearson's Coefficient of Correlation.

It is used to measure the degree of linear relationship betn two variables

- It also called moment correlation coefficient.
- Denoted by 'r' & defined as.

$$r = \frac{\Sigma XY}{N\sigma_X \cdot \sigma_Y}$$

where,

$X = x - \bar{x}$ & $Y = y - \bar{y}$.

N = No. of pair of values of variables.

σ = Standard deviation

or.

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

or

$$r = \frac{Cov.(XY)}{S.D(X)\ SD(Y)}$$

where,

$$Cov(XY) = \frac{\Sigma XY}{N}$$

S.D(X) or (σX) = Standard deviation of x series.

SD(Y) or (σY) = Standard deviation of Y series.

KPF - Karl Pearson's formula

Merits -

1. Imp method to give precised quantitative result with meaningful interpretation
2. It also gives a direction (i.e positive or negative) as well as degree of correlation

Demerits -

1. Time consuming
2. limitation of value of correlation is ($-1 \leq r \leq +1$)

eg - Find Karl Pearson's coefficient of correlation

height of father	x	65	66	67	67	69	71
ht of son	y	67	68	64	68	70	69

Sol'n -

$$\text{(KPF)} \quad r = \frac{\Sigma XY}{\sqrt{\Sigma x^2 \, \Sigma y^2}}$$

or

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

where,

$$\bar{x} = \frac{\Sigma x}{n} \quad \& \quad \bar{y} = \frac{\Sigma y}{n}$$

$$\bar{x} = \frac{\Sigma x}{n} = 67.5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{\Sigma y}{6} = 67.6$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
65	67	-2.5	-0.6	1.75	6.25	0.36
66	68	-1.5	0.4	-0.45	2.25	0.16
67	64	-0.5	-3.6	1.85	0.25	12.96
67	68	-0.5	0.4	-0.2	0.25	0.16
69	70	1.5	2.4	3.6	2.25	5.76
71	69	3.5	1.4	4.9	12.25	1.96
Σ				$\Sigma XY = 11.45$	$\Sigma x^2 = 23.50$	$\Sigma y^2 = 21.36$

(☆) not in syllabus - read once.

$$\therefore r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$= \frac{11.45}{\sqrt{(23.50 \times 21.36}}$$

$$= \frac{11.45}{\sqrt{501.96}} \Rightarrow \frac{11.45}{22.40}$$

$$= \underline{0.511}$$

Since the value in betn -1 & +1. So the variables have high degree of correlation betn x & y.

(☆)

(1) Spearman's Rank coefficient of Correlation

It is method of finding the correlation betn two variables by taking their ranks.

- This method of finding useful in dealing with qualitative data.
- We use it if relative position or rank of magnitude are given, but actual magnitude of variables are not given
- Denoted by ρ (rho) & derived as.

$$\rho = 1 - \frac{6\Sigma d^2}{n(n^2-1)} \quad \underline{\text{or}} \quad 1 - \frac{6\Sigma d^2}{(n^3-n)}$$

* Error -

* Probable error (P.E) — Define the interpreted value of the coefficient of correlation.

$$P.E = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$$

where,
r = coefficient of correlation
n = no. of pairs of correlation

* Standard error — It is the measure of extent to which the sample mean deviates from true mean.

$$\text{or } S_E = \frac{s}{\sqrt{n}} \quad \text{or } SE = \frac{1-r^2}{\sqrt{n}} \quad \text{or } \frac{2}{3}(P.E)$$

Imp

* Multiple correlation —

In this, we study relationship betn three or more variable

— Suppose the dependent variable is z & x & y both are independent variables.

$$R_{z,xy} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2 r_{xz} r_{yz} r_{xy}}{1 - r_{xy}^2}}$$

Similarly, $R_{y,zx} = \sqrt{\frac{r_{zy}^2 + r_{xy}^2 - 2 r_{xy} r_{zy} r_{xz}}{1 - r_{xz}^2}}$

& $R_{x,yz} = \sqrt{\frac{r_{xy}^2 + r_{xz}^2 - 2 r_{xz} r_{xy} r_{yz}}{1 - r_{yz}^2}}$

— A coefficient of Multiple correlation lies between 0 & 1. If 1 then perfect & 0 = no