CH-17

DATE : / /

PAGE NO. :

Correlation & Regression

The same point of time, if a variable is changed due to change in another variable then both are Co-related.

★ TYPES :-

i) Positive Correlation :-

If both variables move in same direction or have the relation.
eg: Height and weight / Day temperature and Sale of cold drink.

ii) Negative Correlation :-

If both variable move in opposite direction or have -ve relation.
eg: Price & demand / No of Claim & Profit of Insurance company.

iii) Zero correlation :-

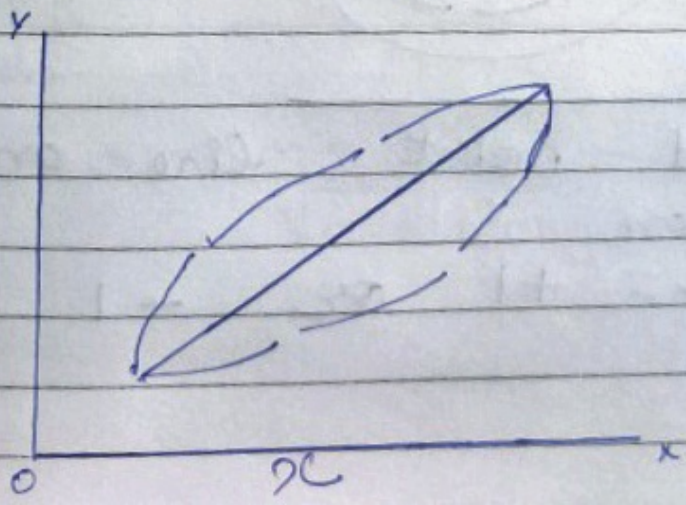
Variable does not move & produce

any difference in the number.
eg: Height & IQ / Rain and space etc.

* Measures of Correlation

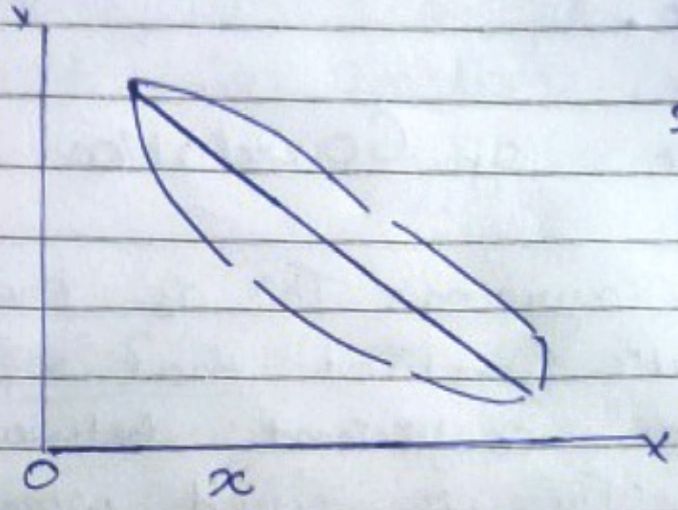
i) Scatter Diagram: This is the simple diagramatic method that represent correlation coefficient between 2 variables. This is used when both variable are in linear or curvilinear relationship. It measures only nature of correlation but fails to find the value of correlation.

- In +ve correlation, plotted points lie from lower left corner to upper right corner.



DATE : / /
PAGE NO. :

- In -ve correlation, plotted points lie from upper left corner to lower right corner.



- In zero correlation, plotted point are equal and evenly distributed.

If plotted point line on single line then,

$r = +1$ or -1

DATE : / /
PAGE NO. :

(2) Karl Pearson Method / Product Movement Method

This is used when variable are in linear relationship so this is shown by

(i) $$r = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$$

σ_x = SD of x
σ_y = SD of y
Cov(x,y) = Covariance

(ii) $$r = \frac{\Sigma dx\,dy}{n\,\sigma_x \times \sigma_y}$$

$\Sigma dx\,dy$ → Sum of product of

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

when table of x & y is given

↓

Huge no.

(iv) $$r = \frac{n\Sigma dx\,dy - \Sigma dx\,\Sigma dy}{\sqrt{n\Sigma d^2n - (\Sigma dx)^2} \times \sqrt{n\Sigma dy - (\Sigma dy)^2}}$$

DATE : / /

PAGE NO. :

(3) Spearman's Rank Correlation Coffic

Is used when qualitative data (attribu are given. This method is alsoused to find the level of aggrement between two judges. eg (singing, dancing, beauty etc)

$$\varrho_R = 1 - \frac{6 \Sigma D^2}{n(n^2-1)}$$

ΣD^2 = Sum of sq of Diffrence in Rank
n = total no. of observation

least imp

$$r = 1 - 6 \left[\frac{\Sigma d^2 + \frac{\Sigma t^3 - t}{12}}{n(n^2-1)} \right]$$

II
In case of tie

t = length of tie

* Properties :

- $-1 \leq \varrho_R \leq 1$
- $\Sigma D = 0$, sum of Diff in Rank is 0

DATE : / /
PAGE NO. :

- If ranks are given by two judges are in reverse order

R_x 1 2 3 4 5

R_y 5 4 3 2 1 $\quad r_R = -1$

④ Coefficient of Concerunt Deviation

It is used when we are not concerned about the magnitude of variables

$$r_c = \pm \sqrt{\pm \frac{(2c - m)}{m}}$$

c = no. of concurrent Deviation

n = total no of observation

* Properties :

- If $(2c - m) > 0$ then we take the sign Inside as well as outside
- If $(2c - m) < 0$ then we take $-ve$ sign

PAGE NO. :

$-1 \leq r_c \leq 1$

This is the quickest Method

② Properties of correlation

- r is unit free measurement
- r always lies b/w −1 to +1 (Both) Inclusive

$-1 \leq r \leq +1$

- r is not affected due to change in origin & scale but affected by sign.

$r_{xy} = .5$ $\quad r_{UV} = ?$

Put like sign can be used

$+ + + + = +$

$+ - + + = -$

$- + + - = +$

$+ - + - = +$

DATE : / /
PAGE NO. :

- Coefficient of determination is equal to r^2
- Coefficient of non-determination is equal to $1 - r^2$
- Percentage of unaccounted variation $(1-r^2) \times 100$
- $\sigma_x \times \sigma_y \geq Cov(x, y)$

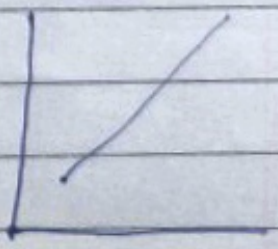
✱ Regression :-

In this concept we estimate or pridict the value of one variable from the given value of another variable. It give exact value / Relation b/w x & y.

It tells what is the linear relation b/w the variable

eg $2x + 3y = 8$

$y = 7 + 2x$



DATE : / /
PAGE NO. :

- There are two regression line :

line of y on x

$y = a + bx$] दिखाने के लिए eq.

a = Intercept

bx = slope or Regg. Coff.

$y - \bar{y} = b_{yx}(x - \bar{x})$

for queo. solving

b_{yx} = coff of regg. of y on x

$= r \cdot \frac{\sigma_y}{\sigma_x}$

line of x on y

$x = a + by$

$x - \bar{x} = b_{xy}(y - \bar{y})$

b_{xy} = coff of regg. of x on y

$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$

DATE : / /
PAGE NO. :

★ Properties of Regression Line

- Intersection point of both regressio line is called mean $(\bar{x}, \bar{y})$
- Least square method is the best method to find regression line

$$\Sigma Y = N a + b \Sigma x$$

$$\Sigma x y = a \Sigma x + b \Sigma x^2$$

- If $r = (+1)$ or (-1) so both regressio line are identical or co-iside
- If $r = 0$ then both line are perpendicular or bisect Right angle
- r is the geometic mean of both regression coefficient

$$r = \sqrt{b_{yx} \times b_{xy}}$$

- r, b_{yx} and b_{xy} always have same sign

DATE : / /

PAGE NO. :

- Product of b_{yx} & b_{xy} is always less then or equal to unity (1).

- Regression coefficient is unaffected due to change in origin but affected by change in scale.

If $u = a_1 + b_1 x$ & $v = a_2 + b_2 y$

then $b_{uv} = \frac{q}{p} \times b_{xy}$ (Regg Coff of uv)

or

Regg Coff of uv / $b_{uv} = \frac{p}{q} \times b_{yx}$

$$p = \frac{\text{Coff of } u}{\text{Coff of } x}$$

$$q = \frac{\text{Coff of } v}{\text{Coff of } y}$$