HANDWRITTEN NOTES

(Quantitative Aptitude)

CORRELATION AND REGRESSION

Correlation & Regression

Correlation

→ Degree of association between two variable.
→ Correlation can be negative, positive or zero
→ $-1 \leq r \leq 1$

- For Direction Relation b/w x & y $x\uparrow y\uparrow$ or $x\downarrow y\downarrow$ r is positive
→ for inverse relation b/w x & y $x\downarrow y\uparrow$ or $x\uparrow y\downarrow$, r is negative.

⇓

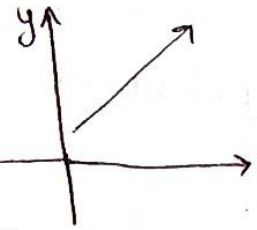
- $\boxed{r=-1}$ ↓ Perfect negative Correlation
- $-1 < r < 0$ ⇓ (Negative Correlation)
- $\boxed{r=0}$ ↓ No correlation (variable are independent)
- $0 < r < 1$ ⇓ (Positive correlation)
- $\boxed{r=1}$ ⇓ Perfect positive Correlation

Methods of Calculating Correlation

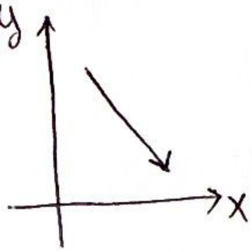
- Graphical Method ⇓ Scattered Diagram
- Non-Graphical Method
  - Karl Pearson method
  - Spearman method
  - Concurrent Deviation

Scattered Diagram Method

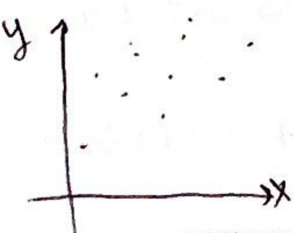
Plot the value of two variable on graph



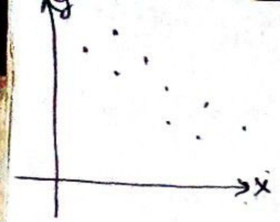
for straight line
(lower left to upper right) $\boxed{r=1}$



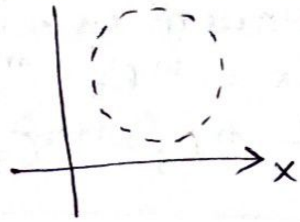
for straight line
(upper left to lower right) $\boxed{r=-1}$



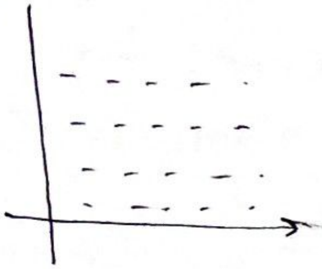
$0 < r < 1$ (No straight line)



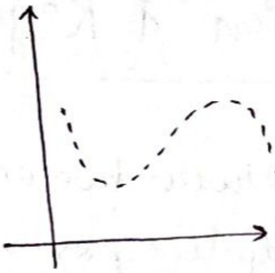
$-1 < r < 0$
No straight line



$r = 0$



$r = 0$



$r = 0$

→ Exact magnitude can't be calculated is scattered Diagram Method

Karl Pearson's coefficient correlation

$$\boxed{r = \frac{\text{cov}(x,y)}{\sigma x \, \sigma y}} \quad \text{or} \quad \boxed{r = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N \, \sigma x \, \sigma y}}$$

$$\boxed{r = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2 \times \Sigma(y_i-\bar{y})^2}}} \quad \text{or} \quad \boxed{r = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{N}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{N}} \sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{N}}}}$$

→ Correlation doesn't change with change of origin
i.e $U_i = x_i - A$ & $V_i = y_i - B$.
then,
$$r = \frac{\Sigma UV - \frac{\Sigma U \times \Sigma V}{N}}{\sqrt{\Sigma U^2 - \frac{(\Sigma U)^2}{N}} \sqrt{\Sigma V^2 - \frac{(\Sigma V)^2}{N}}}$$

Correlation doesn't change with change of scale (Provided Scale is positive)

→ When scale is negative the sign of correlation may change but magnitude doesn't change

Eg:- $r(x,y) = 0.5$; if $U_i = 2x_i$ & $V_i = 4y_i$ then,
$r(U_i, V_i) = (+)(+)0.5 = 0.5$

Eg - $r(x,y) = 0.6$ if $U_i = -2x_i$ & $V_i = 3y_i$ then $r(U,V)$
$r(U,V) = (-)(+)0.6 = -0.6$

Eg - If $r(x,y) = 0.7$ & $U_i = -2x_i$ & $V_i = -4y_i$
then $r(U,V) = (-)(-)0.7 = (+)0.7$

Spearman's Rank Correlation

This method used for qualitative characters & level of agreements & disagreements b/w opinions of Judges.

→ when no numbers repeat ; $r = 1 - \frac{6 \Sigma D^2}{N^3 - N}$

→ when some numbers repeat ;

$$\boxed{r = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 + m_2)\right.}{N^3 - N}}$$

⇒ $\Sigma D = 0$

Concurrent Deviation Method

Concurrent deviation ⇒ when x & y both increase or both decrease

$r = \pm\sqrt{\pm\left(\frac{2C - m}{m}\right)}$; C = Total no. of concurrent deviations

$m = n - 1$

X	10	12	15	14	16	12
y	12	15	14	15	18	20

X: + + − + −

y: + − + + +

Yes (+), No (−), No (−), Yes (+), No (−)

yes (· C = 2

$m = 6 - 1 = 5$

$r = \sqrt{\frac{2(2) - 5}{5}}$

$r = \sqrt{-\frac{1}{5}} = -\sqrt{\frac{1}{5}}$

→ when $\frac{2C - m}{m}$ is negative then (−) sign will be taken out from $\sqrt{\ }$ sign.

For Bivariate Distribution table of "m x n"

$\boxed{\text{Total cells} = m \times n}$ $\boxed{\text{Total marginal distribution} = 2}$

Total concurrent distribution $= m + n$

Coefficient of Determination

$= r^2 = \frac{\text{Explained variance}}{\text{Total Variance}}$

Coefficient of Non-Determination

$= 1 - r^2$

Regressions

- Establishing mathematical relation b/w two variables (Indepent & Dependent)
- Prediction of dependent variable
- For linear regression least square method is used.

There Two linear Regression lines

	When y depends on x	When x depends on y
Standard form	$\boxed{y = a + bx}$	$\boxed{x = a + by}$
How to find	Use formula $\boxed{y - \bar{y} = b_{yx}(x - \bar{x})}$	Use formula $\boxed{x - \bar{x} = b_{xy}(y - \bar{y})}$
Slope of line	$\boxed{\text{slope} = b_{yx}}$	$\boxed{\text{slope} = 1 / b_{xy}}$

b_{yx} & b_{xy} are known as regressions coefficients.

Calculation of Regression coefficients

$$b_{yx} = \frac{\mathrm{COV}(x,y)}{\sigma_x^2}$$

$$b_{yx} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

$$b_{xy} = \frac{\mathrm{COV}(x,y)}{\sigma_y^2}$$

$$b_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(y_i - \bar{y})^2}$$

$$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{N}}{\Sigma y^2 - \frac{(\Sigma y)^2}{N}}$$

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

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

→ 'r' will be positive if both b_{yx} & b_{xy} are positive

→ 'r' will be negative if both b_{yx} & b_{xy} are negative

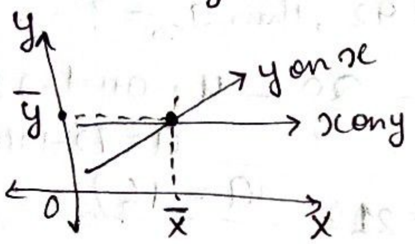
$b_{yx} \times b_{xy} \leq 1$

$r \leq \dfrac{b_{yx} + b_{xy}}{2}$

If regression line y on x is: $ax+by+c=0$ then $b_{yx} = -\dfrac{a}{b}$

If regression line x on y is $Ax+By+C=0$ then $b_{xy} = -\dfrac{B}{A}$

Two regressions lines intersect each other at $(\bar{x}, \bar{y})$



Regression coefficients doesn't change with change of origin

$U_i = x_i - A$ & $V_i = y_i - B$

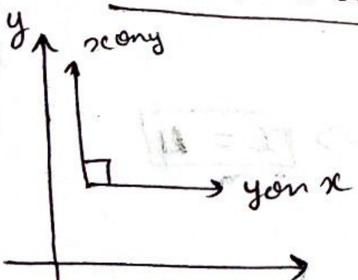
then $b_{yx} = b_{VU}$ & $b_{xy} = b_{UV}$.

Regression coefficients change with change of scale.

$U_i = ax_i + b$ & $V_i = cy_i + d$

So, $b_{VU} = \dfrac{c}{a} \times b_{yx}$ or $b_{VU} = \dfrac{\text{Scale of } y}{\text{Scale of } x} \times b_{yx}$

when two lines are perpendicular



$\boxed{r = 0}$

when two lines are coincident