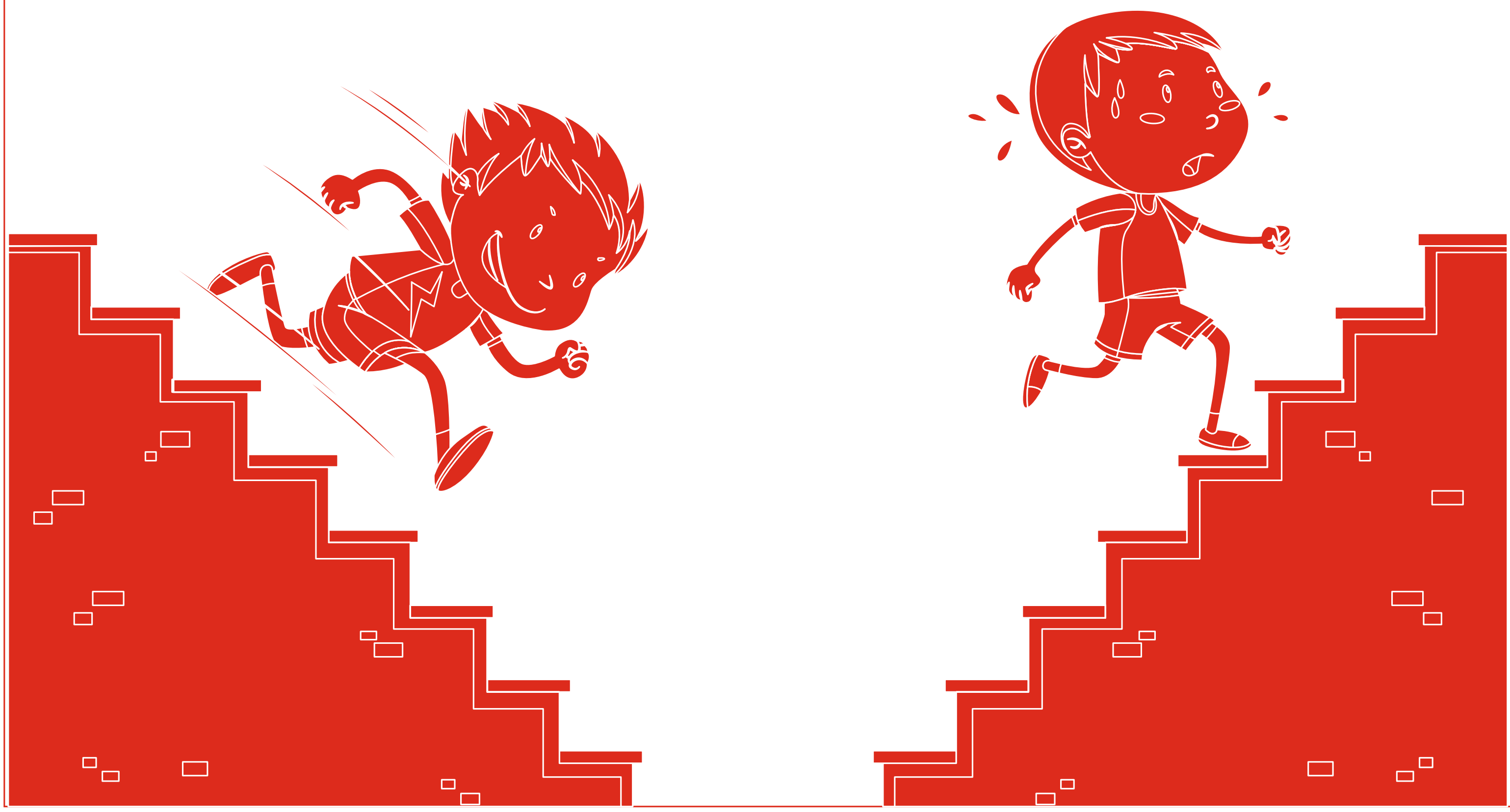
DO or DO NOT
There is no TRY

- CA VINOD REDDY -

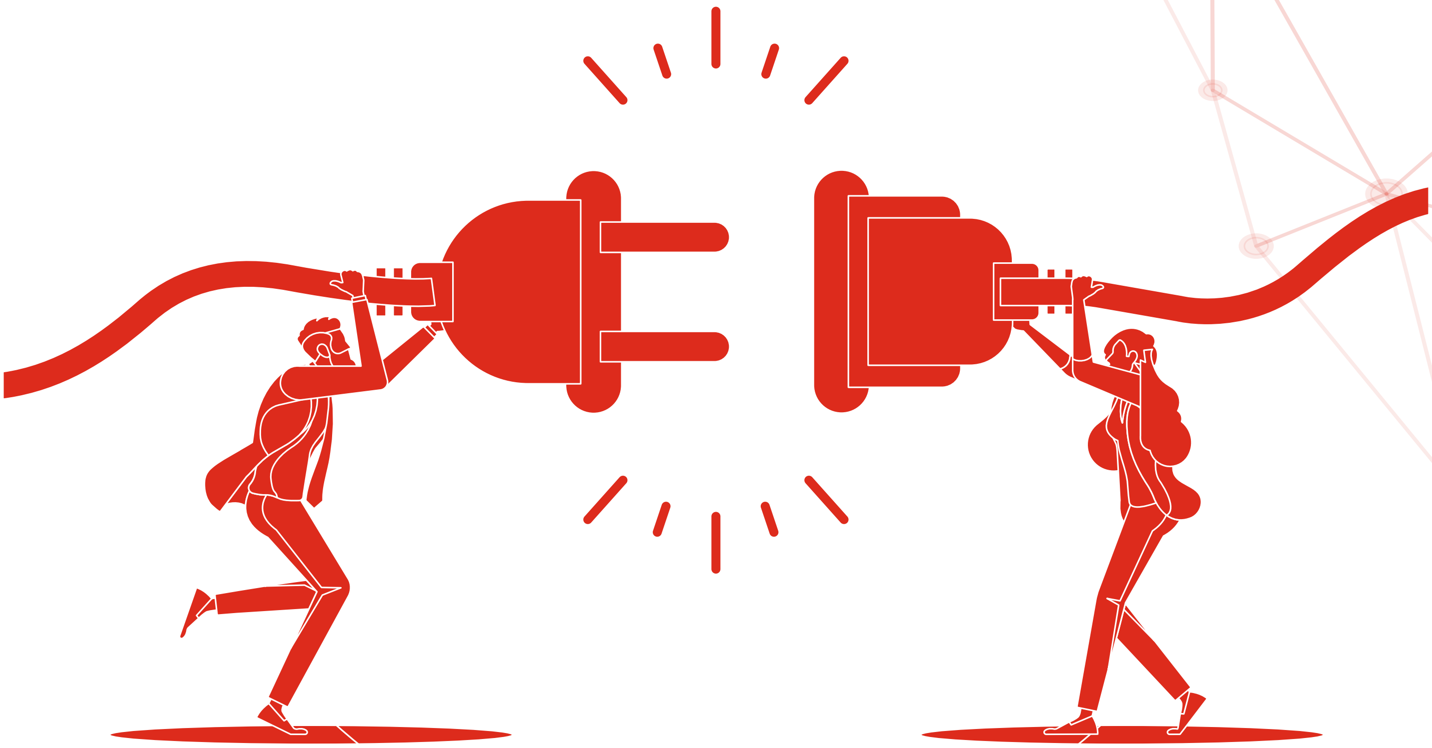
no One is Coming to SAVE You.
This Life is 100% YOUR RESPONSIBILITY

Do not be afraid to give up 'GOOD' to go for the 'GREAT'

- CA VINOD REDDY -



CORRELATION AND REGRESSION ANALYSIS



CA VINOD REDDY

1 What is correlation and what is regression?

⟹ • correlation is the process of establishing relationship / association betn 2 or more variables although they are not in proportion.

• Regression is the process of Determining the value of one variable on the basis of other.

• correlation is the pre-requisite to study Regression.

2 Whether correlation between 2 variables exists or not?

Yes

① **What is the type / Nature of correlation?**

- positive

x	y
↑	↑
↓	↓

- Negative

x	y
↑	↓
↓	↑

No

② **What is the degree of correlation?**

- perfect
- Small OR weak OR poor
- Moderate
- High OR Strong

3 Methods to measure correlation between 2 variables :

a. Scatter Diagram
b. Spearman's rank correlation coefficient (R)
c. coeff. of concurrent deviation (r)
d. Karl pearson's product moment correlation coefficient (r) (Best)

My Notes

• studying association betn 2 variables :
~~Bi-variate correlation~~

• studying association betn 3 or more variables :
~~Multi-variate correlation~~

• examples of positive correlation

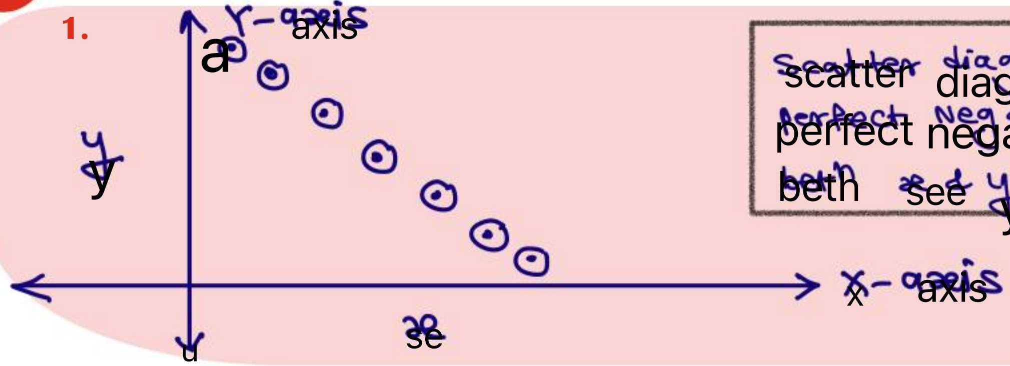
① price & supply
② Temp, sale of cold drinks
③ industry growth rate, Demand for CA's etc

examples of Negative correlation

① Demand & price
② temp, sale of tea/coffee
③ No. of claims, profit of Insu company

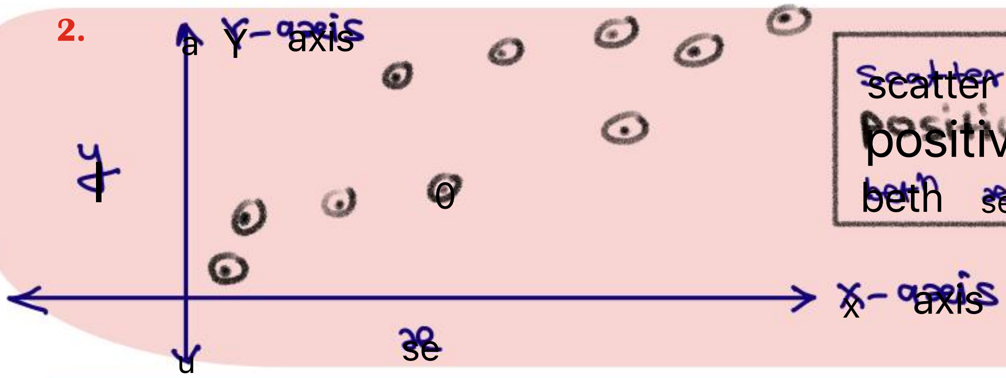
4 Scatter diagram showing

1.



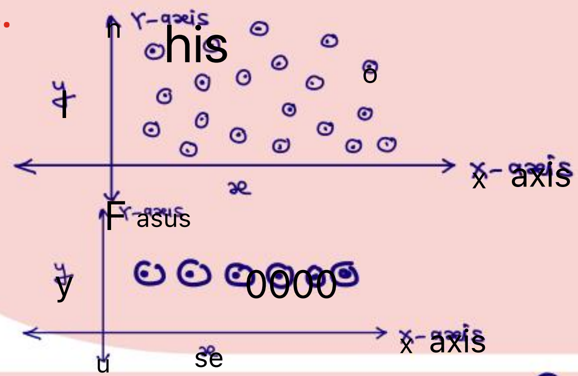
Scatter diagram is showing perfect Negative correlation both x & y

2.



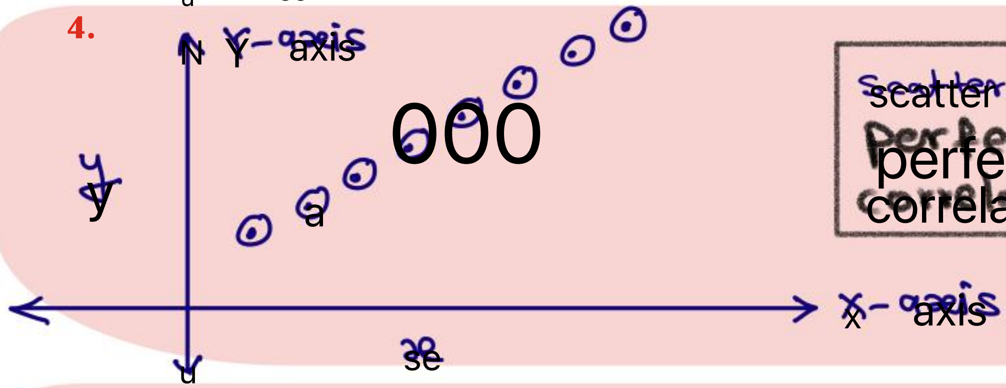
Scatter diagram is showing positive correlation both x & y

3.



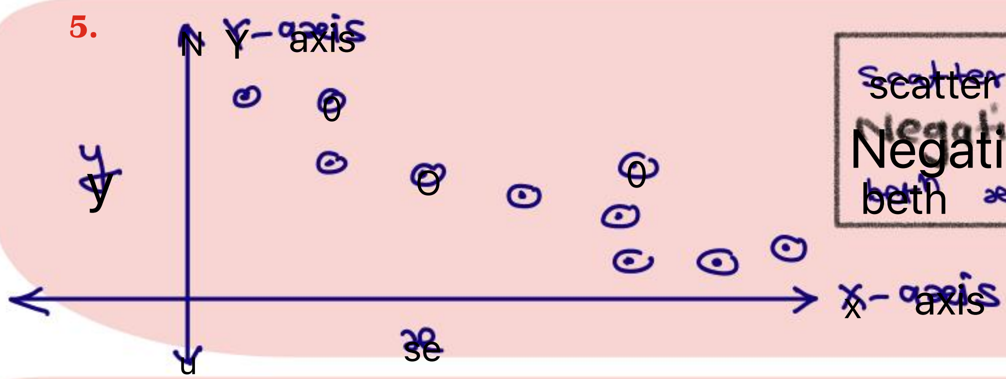
Scatter diagram is showing No correlation both x & y

4.



Scatter diagram is showing Perfect positive correlation both x & y

5.



Scatter diagram is showing Negative correlation both x & y

Scatter diagrams can give an idea about type of correlation but it can't give exact degree of correlation.

To know type as well as degree of correlation we need to obtain correlation coefficient

5 Find Spearman's rank correlation coefficient.

x	30	80	45	63	91	28	222
y	101	111	93	123	86	65	79

x	y	Rank of x	Rank of y	d^2
30	101	6	3	9
80	111	3	2	1
45	93	5	4	1
63	123	4	1	9
91	86	2	5	9
28	65	7	7	0
222	79	1	6	25
			$\Sigma d^2 =$	54

Spearman's Rank correlation coefficient

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

$$= 1 - \left[\frac{6 \times 54}{7(7^2-1)}\right]$$

$$= 1 - \left(\frac{324}{336}\right)$$

$$= 0.0357$$

There is weak / Poor / Low degree of positive correlation.

6 Find Spearman's rank correlation coefficient.

x	58	92	63	63	65	65	63	58
y	20	25	28	25	28	25	30	38

x	y	Rank x	Rank y	d^2
58	20	7.50	8	0.25
92	25	1	6	25
63	28	5	3.50	2.25
63	25	5	6	1
65	28	2.50	3.50	1
65	25	2.50	6	12.25
63	30	5	2	9
58	38	7.50	1	42.25
		Σd^2		93

t = no. of observations involved in a 'tie'

= 2, 3, 2, 2, 3

$$\Sigma\left(\frac{t^3-t}{12}\right) = \left(\frac{2^3-2}{12}\right) + \left(\frac{3^3-3}{12}\right) + \left(\frac{2^3-2}{12}\right) + \left(\frac{2^3-2}{12}\right) + \left(\frac{3^3-3}{12}\right)$$

$$= 0.50 + 2 + 0.50 + 0.50 + 2$$

$$= 5.50$$

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

$$r = 1 - \left[\frac{6(93+5.50)}{8(8^2-1)}\right]$$

$$r = 1 - \left(\frac{591}{504}\right) = -0.172619$$

There is weak degree of negative correlation beth x & y.

6

Find Spearman's correlation coefficient for

x	10	18	26	10
y	36	14	22	22

$\Longrightarrow$

Rank of x :	3.50	2	1	3.50
Rank of y :	1	4	2.50	2.50
d^2 :	6.25	4	2.25	1

$t = 2, 2$ $\qquad \Sigma d^2 = 13.50$

$$\Sigma\frac{t^3-t}{12} = \left(\frac{2^3-2}{12} + \frac{2^3-2}{12}\right)$$

$$= 0.50 + 0.50$$

$$= 1.00$$

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

$$r = 1 - \frac{6(13.50 + 1.00)}{4(4^2-1)}$$

$$r = 1 - \left(\frac{87}{60}\right) = -0.45$$

There is moderate degree of negative correlation.

7

Spearman's Rank Correlation Coefficient.

without tie

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

with tie

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

where d = Diff of Rank
n = no. of pairs of observations
t = no. of observations involved in a tie.

8

Find Coefficient of Concurrent Deviation for -

x	60	90	28	36	51	58	90	95	101	63
y	28	111	93	28	63	78	53	28	99	100

x	y	Deviation of x	Deviation of y	product
60	28			
90	111	+	+	+
28	93	−	−	+
36	28	+	−	−
51	63	+	+	+
58	78	+	+	+
90	53	+	−	−
95	28	+	−	−
101	99	+	+	+
63	100	−	+	−

c = no. of concurrent deviations
= No. of '+' signs in product column
= 5

$m = n - 1 = 10 - 1 = 9$

$$r = \pm\sqrt{\pm\left(\frac{2c-m}{m}\right)}$$

$$r = \sqrt{\frac{2(5)-9}{9}}$$

$$r = \sqrt{\frac{1}{9}}$$

$$r = \frac{1}{3} = 0.3333333$$

8 In the product column : No. of positive signs = x
No. of negative signs = y

x > y	r is positive	→ $0 < r \le 1.00$
x < y	r is negative	→ $-1 \le r < 0$
x = y	r = 0	→ $r = 0$

9 Find Karl Pearson's product moment correlation coefficient

x	8	3	11	9	6
y	5	8	13	20	28

x	y	x^2	y^2	xy
8	5			
3	8			
11	13			
9	20			
6	28			
37	74	311	1442	555

co-variance of x, y

$= cov(x, y)$

$= \frac{\Sigma xy}{n} - \bar{x} \cdot \bar{y}$

$= \frac{555}{5} - (7.40 \times 14.80)$

$= 1.48$

$\bar{x} = 7.40$

$\bar{y} = 14.80$

$SD_x = \sqrt{\frac{311}{5} - 7.40^2} = 2.7276$

$SD_y = \sqrt{\frac{1442}{5} - 14.80^2} = 8.3283$

Karl pearson's corr. coeffi.

$= \left[\frac{cov(x,y)}{\sigma_x \cdot \sigma_y}\right] = \frac{1.48}{2.7276 \times 8.3283}$

$= 0.065151$

10 Covariance of (x,y) = $\left(\frac{\Sigma xy}{n} - \bar{x} \cdot \bar{y}\right) = \left[\frac{\Sigma[(x-\bar{x})(y-\bar{y})]}{n}\right]$

$SD_x = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n}}$

$SD_y = \sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2} = \sqrt{\frac{\Sigma (y-\bar{y})^2}{n}}$

My Notes

Karl pearson's product moment correlation coefficient

$$r = \left[\frac{cov(x,y)}{\sigma_x \cdot \sigma_y}\right] = \left[\frac{\frac{\Sigma xy}{n} - \bar{x} \cdot \bar{y}}{\sqrt{\frac{\Sigma x^2}{n} - (\bar{x})^2} \times \sqrt{\frac{\Sigma y^2}{n} - (\bar{y})^2}}\right]$$

$$r = \left[\frac{\Sigma[(x-\bar{x})(y-\bar{y})]}{\sqrt{\Sigma (x-\bar{x})^2} \times \sqrt{\Sigma (y-\bar{y})^2}}\right]$$

$$-1.00 \le r \le 1.00$$

11

r	Type of Correlation
r = 1.00	perfect positive correlation.
0.30 < r < 0.80	Moderate degree of positive correlation.
0.80 < r < 1.00	strong / High degree of positive correlation.
r = 0	No correlation.
r = -1.00	perfect negative correlation
-1.00 < r < -0.80	strong / High degree of Negative correlation.
-0.80 < r < -0.30	Moderate degree of negative correlation.
0 < r < 0.30	weak / Low degree of positive correlation.
-0.30 < r < 0	weak / Low degree of negative correlation.

12 If $v = 3x+8$; $u = 8y-19$; $r_{xy} = 0.80$

$r_{uv} = r_{xy} = 0.80$

Correlation coefficient is unaffected by change / shift of origin as well as by change in scale.

13 If $u = -3x+53$; $v = -18y+99$; $r_{xy} = 0.70$

$r_{uv} = r_{xy} = 0.70$

14 If $u = -18x+55$; $v = 16y+100$; $r_{xy} = 0.85$

$r_{uv} = -r_{xy} = -0.85$

15 If $u = -8x+19$; $v = -16y-33$; $r_{xy} = -0.56$

$r_{uv} = r_{xy} = -0.56$

16 Find Karl Pearson's Coefficient for -

x	30	60	90	50
y	20	30	40	80

$\Longrightarrow$

$\bar{x} = 57.50$, $\bar{y} = 42.50$

$\sigma_x = \sqrt{\frac{15100}{4} - 57.50^2} = 21.651$

$\sigma_y = \sqrt{\frac{9300}{4} - 42.50^2} = 22.7761$

$cov(x,y) = \frac{10000}{4} - (57.50 \times 42.50)$

$= 56.25$

$r = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{56.25}{21.651 \times 22.7761}$

$r = 0.1141$

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
30	20	−27.50	−22.50	618.75
60	30	2.50	−12.50	−31.25
90	40	32.50	−2.50	−81.25
50	80	−7.50	37.50	−281.25
				225

$\Sigma(x-\bar{x})^2 = 1875$

$\Sigma(y-\bar{y})^2 = 2075$

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

$r = \frac{225}{\sqrt{1875} \times \sqrt{2075}} = 0.1141$

17

x	100	80
y	30	60

Find r by Karl Pearson's Method :

x	y	xy
100	30	3000
80	60	4800

$cov(x,y) = \frac{\Sigma xy}{n} - \bar{x}\bar{y}$

$= \frac{7800}{2} - (90 \times 45)$

$= -150$

$\sigma_x = 10$

$\sigma_y = 15$

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

$r = \frac{-150}{10 \times 15} = -1.00$

x	y
2	30
3	35

$\Rightarrow r = -1.00$

x	y
70	800
80	900

$\Rightarrow r = 1.00$

when n = 2 then

$r = 1.00$ OR $r = -1.00$

x	y
200	300
250	480

$\Rightarrow r = 1.00$

x	y
80	595
70	625

$\Rightarrow r = -1.00$

18

Regression Analysis

After studying correlation between 2 variables, the process of estimating the value of one variable on the basis of other is known as regression analysis

x = Given
y = ?

Regression of y on x

Eqn of Regression line of y on x

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

y = Given
x = ?

Regression of x on y

Eqn of Reg line of x on y

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

r, b_{yx}, b_{xy} all are unit-free

Reg line of y on x is : $(y - \bar{y}) = b_{yx}(x - \bar{x})$

Reg line of x on y is : $(x - \bar{x}) = b_{xy}(y - \bar{y})$

Reg coefficient of y on x is = b_{yx} = $r \cdot \frac{SD_y}{SD_x} = r \cdot \frac{\sigma_y}{\sigma_x}$

Reg coefficient of x on y is = b_{xy} = $r \cdot \frac{SD_x}{SD_y} = r \cdot \frac{\sigma_x}{\sigma_y}$

My Notes

① $y = 182 - 2.5x$ $x = -1.6y + 520$

$r_{xy} = -0.8578$, $r_{yx} = ?$

$\Rightarrow r_{yx} = 0.8578$

19 If $\bar{x} = 30$, $\bar{y} = 90$, $\sigma_x = 5$, $\sigma_y = 8$, $r = 0.80$

Find a. Reg line of x on y

b. Reg line of y on x

c. If x =25, y =?

d. If y =85, x =?

① Eqn of Regression line of x on y

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

$x - \bar{x} = r \cdot \frac{\sigma_x}{\sigma_y}(y - \bar{y})$

$x - 30 = 0.80 \times \frac{5}{8} \times (y - 90)$

$x - 30 = 0.50(y - 90)$

$x - 30 = 0.50y - 45$

$\boxed{x = -15 + 0.50y}$

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

$= 0.80 \times \frac{5}{8}$

$= 0.50$

② Eqn of Regression line of y on x

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

$y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x})$

$y - 90 = 0.80 \times \frac{8}{5} \times (x - 30)$

$y - 90 = 1.28x - 38.40$

$\boxed{y = 51.60 + 1.28x}$

$\boxed{\begin{aligned} b_{yx} &= r \cdot \frac{\sigma_y}{\sigma_x} \\ &= 0.80 \times \frac{8}{5} \\ &= 1.28 \end{aligned}}$

③ $x = 25$, $y = ?$

$y = 51.60 + 1.28x$

$= 51.60 + (1.28 \times 25) = 83.60$

④ $y = 85$, $x = ?$

$x = -15 + 0.50y$

$= -15 + (0.50 \times 85)$

$= 27.50$

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

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

Therefore, $b_{yx} \cdot b_{xy}$

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

$= r^2$

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

$\boxed{r^2 = b_{yx} \cdot b_{xy}}$

Square of correlation coefficient is equal to product of 2 regression coefficients.

Correlation coefficient 'r' is G.M. of 2 regression coefficients $b_{yx} \cdot b_{xy}$

20

$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} \times \frac{\sigma_y}{\sigma_x} = \frac{cov(x,y)}{\text{variance of } x}$

$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} \times \frac{\sigma_x}{\sigma_y} = \frac{cov(x,y)}{\text{variance of } y}$

$b_{yx} \cdot b_{xy} = r^2$

Therefore 'r' is G.M. of b_{yx} & b_{xy}

$r^2 = (b_{yx} \cdot b_{xy})$

$1.00 \geq r \geq -1.00$

$1.00 \geq r^2 \geq 0.00$

r	b_{yx}	b_{xy}
0	0	0
+	+	+
-	-	-

y = ?, x = given } we should use Reg. of y on x

x = ?, y = given } we should use Reg. of x on y

21

If Reg. line of y on x is written in the form of $y = a + bx$ then 'b' represents 'b_{yx}'

If Reg. line of y on x is 3x+5y=83. Find b_{yx}

$3x + 5y = 83$

$5y = 83 - 3x$

$y = \frac{83}{5} - \frac{3x}{5}$

$\boxed{\therefore y = 16.60 - 0.60x}$

$\therefore b_{yx} = -0.60$

22

If Reg. line of x on y is written in the form of $x = a + by$ then 'b' represents 'b_{xy}'

If Reg. line of x on y is 2x-3y=95. Find b_{xy}

$2x - 3y = 95$

$2x = 95 + 3y$

$x = \frac{95}{2} + \frac{3y}{2}$

$x = 47.50 + 1.50y$

$\boxed{\therefore b_{xy} = 1.50}$

$0 \leq b_{yx} \cdot b_{xy} \leq 1.00$

as

$0 \leq r^2 \leq 1.00$

23

On solving 2 regression lines simultaneously. If we get x = 50 and y = 90, then

$\Longrightarrow$ (50,90) is the point of intersection of 2 regression lines $= (\bar{x}, \bar{y})$

$\bar{x} = 50, \quad \bar{y} = 90$

24

Probable Error = $0.674 \times \frac{(1 - r^2)}{\sqrt{N}}$ = [0.674 × standard error]

Standard Error = $\frac{(1 - r^2)}{\sqrt{N}}$

Coefficient of determination = (r^2) = (Explained variance / Total variance)

Coefficient of Non-determination = $(1 - r^2)$ = (unexplained variance / Total variance)

where n = sample size

N = population size

25 2 regression lines become identical i.e. they coincide when r = -1 or r = 1.

i.e. 2 Regression lines coincide when there is perfect positive or perfect negative correlation.

26 $\boxed{\text{If } r = 0\text{; then regression lines are } \perp \text{ to each other.}}$

When there is no correlation between 2 variables then regression lines will be $\perp$ to each other.

27

Particulars	Maths (x)	Stats (y)
AM	88	92
SD	10	12
r	0.75	

Find 1. Reg. line of y on x
2. Reg. line of x on y
3. If x = 95, y = ?
4. If y = 90, x = ?

① Regression line of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$
$$y - 92 = 0.90(x - 88)$$
$$y - 92 = 0.90x - 79.20$$
$$\boxed{y = 12.80 + 0.90x}$$

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} = 0.75 \times \frac{12}{10} = 0.90$$

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} = 0.75 \times \frac{10}{12} = 0.625$$

② Reg. line of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$
$$x - 88 = 0.625(y - 92)$$
$$x - 88 = 0.625y - 57.50$$
$$\boxed{x = 30.50 + 0.625y}$$

③ $y = 12.80 + 0.90(95)$

$y = 98.30$

when $x = 95$ then exp. value of $y = 98.30$

④ when $y = 90$, $x = ?$

$x = 30.50 + 0.625(90) = 86.75$

esti. value of $x = 86.75$ when $y = 90$

$$\left.\begin{aligned} y &= 12.80 + 0.90x \\ x &= 30.50 + 0.625y \end{aligned}\right\}$$ Let's solve these eqns simultaneously

$\rightarrow$ $y = 12.80 + 0.90(30.50 + 0.625y)$

$y = 12.80 + 27.45 + 0.5625y$

$0.4375y = 40.25$ $\quad\therefore x = 30.50 + 0.625(92)$

$\boxed{y = 92}$ $\quad\boxed{x = 88}$

Point of intersection of 2 regression lines $= (88, 92) = (\bar{x}, \bar{y})$

28

$1.00 \geq r \geq -1.00$

$1.00 \geq r^2 \geq 0.00$

$1.00 \geq (b_{yx}.b_{xy}) \geq 0.00$

Minimum value of $r = -1.00$
Maximum value of $r = 1.00$
Minimum value of $r^2 = 0.00$
Maximum value of $r^2 = 1.00$

29

If $b_{yx} > 0$; then $b_{xy} < 0$ ...... This statement is incorrect

sign of b_{yx}, b_{xy} and r will always be same

30

If $b_{yx} = 2.50$, $r = 0.80$, $b_{xy} = ?$

$\Longrightarrow$ $r^2 = b_{yx} \times b_{xy}$

$0.80^2 = 2.50 \times b_{xy}$

$b_{xy} = 0.256$

31

If $b_{xy} = -1.56$, $b_{yx} = -0.20$, $r = ?$

$\Longrightarrow$ $r^2 = b_{yx} \cdot b_{xy} = -0.20 \times -1.56 = 0.312$

$r = \sqrt{0.312} = -0.55856960175$

32

If $b_{xy} = -1.5281$, $b_{yx} = 0.2381$, $r = ?$

$\Longrightarrow$ $r^2 = b_{yx} \cdot b_{xy} = 0.2381 \times -1.5281 = -0.36384061$

As $0 < r^2 < 1.00$ $\therefore$ Given data is wrong.

33

If $b_{yx} = 1.82$, $b_{xy} = 0.90$, $r = ?$

$\Longrightarrow$ $r^2 = b_{yx} \cdot b_{xy}$

$= 1.82 \times 0.90$

$r^2 = 1.638$

As r^2 can not be more than 1.00, This is wrong data.

My Notes

If $r = 0.60$ then

(coefficient of Determination)
$= r^2 = 0.60^2 = 0.36$
$= 36\%$
$=$ Explained variance

(coefficient of non-determination)
$= 1 - r^2 = 1 - 0.60^2 = 1 - 0.36 = 0.64$
$= 64\% =$ unexplained variance

34

If $\bar{x} = 90$, $\bar{y} = 80$, $r = -0.85$, $\sigma_x = 10$, $\sigma_y = 18$

1. If x = 35, y = ?

$$y - 80 = -0.85 \times \frac{18}{10} \times (35 - 90)$$

$$\boxed{y = 164.15}$$

2. If y = 98.70, x = ?

$$x - 90 = -0.85 \times \frac{10}{18} \times (98.70 - 80)$$

$$\boxed{x = 81.1694}$$

35

If r = 0.75. Find coefficient of determination and coefficient of non-determination.

$\Rightarrow$ ① coefficient of Determination $= r^2 = 0.75^2 = 0.5625$

$\therefore$ Explained variance $= 56.25\%$

② coeffi. of Non-determination $= 1 - r^2 = 1 - 0.5625 = 0.4375$

$\therefore$ unexplained variance $= 43.75\%$

36

x	y
35	480
28	410

Find 'r'

$\Rightarrow$ pls remember that for 2 pairs of obsns either $r = 1.00$ or $r = -1.00$

Here both x & y are decreasing

$\therefore r = 1.00$

37

x	y
200	500
180	600

$\Rightarrow r = -1.00$

x	y
200	800
250	703

$\Rightarrow r = -1.00$

x	y
10	30
18	38

$\Rightarrow r = 1.00$

38

If C = 5, m = 11. Find coefficient of concurrent deviation.

$$\text{coeffi. of concurrent deviation} = \pm\sqrt{\pm\left(\frac{2c-m}{m}\right)} = \pm\sqrt{\pm\frac{2(5)-11}{11}} = -\sqrt{-\left(-\frac{1}{11}\right)}$$

$$= -0.3015$$

My Notes

If $\left(\frac{2c-m}{m}\right) = 0$ then $r = 0$

$\left(\frac{2c-m}{m}\right) > 0$ then $0 < r \leq 1.00$

$\left(\frac{2c-m}{m}\right) < 0$ then $-1 \leq r < 0$

39 If cov (x,y) = 0, then r = 0

If cov (x,y) = positive, then $\Longrightarrow$ $1.00 \geq r > 0$

If cov (x,y) = negative, then $\Longrightarrow$ $-1.00 \leq r < 0$

As $r = \dfrac{\text{covariance of } (x,y)}{SD_x \times SD_y} = \left[\dfrac{cov(x,y)}{\sigma_x \cdot \sigma_y}\right]$

40 Karl Pearson's product moment correlation coefficient is the ratio of cov (x,y) to product of standard deviations of x & y

41 Prepare a bi-variate frequency table for the following data relating to marks in stats (x) and maths (y).

(12,18) (2,16) (12,3) (19,12) (5,8) (8,2) (13,14)
(2,6) (13,19) (6,10) (2,12) (14,2) (18,5) (20,1)

x \ y		Marks in Maths (y)		Total
		0-10	10-20	
Marks in Stats (x)	0-10	||| = 3	||| = 3	6
	10-20	|||| = 4	|||| = 4	8
	Total	7	7	14

} Bi-variate Frequency Table

Find Marginal Distribution of x : Distri. of x over all values of y

x	0-10	10-20
f	6	8

Find Marginal Distribution of y : Distri. of y over all values of x

y	0-10	10-20
f	7	7

Find conditional Distribution of x when y is 10-20:

x	0-10	10-20
f	3	4

Find conditional Distribution of y when x is 0-10:

y	0-10	10-20
f	3	3

42

'Marginal Distribution' is the frequency distribution of one variable (x or y) across the other variable's full range of values.

'Conditional Distribution' is the frequency distribution of one variable (x or y) across the particular sub-population of other variable.

43

x \ y	0-10	10-20	20-30	30-40	40-50	Total
0-10	5	20	22	23	25	95
10-20	8	30	26	28	42	134
20-30	9	20	29	38	48	144
30-40	13	50	36	39	56	194
40-50	26	60	28	19	26	159
Total	61	180	141	147	197	726

Bi-variate Frequency Table

Find Marginal Distribution of x :

x	0-10	10-20	20-30	30-40	40-50
f	95	134	144	194	159

Find Marginal Distribution of y :

y	0-10	10-20	20-30	30-40	40-50
f	61	180	141	147	197

Find conditional Distribution of x when y is 30-40:

x	0-10	10-20	20-30	30-40	40-50
f	23	28	38	39	19

Find conditional Distribution of y when x is 20-30:

y	0-10	10-20	20-30	30-40	40-50
f	9	20	29	38	48

My Notes

① bye = boy	ⓐ correct	~~ⓑ~~ incorrect
② γ_{xey} = γ_{yse}	~~ⓐ~~ correct	ⓑ incorrect
③ $b_{ye} \cdot b_{xy}$ = 22	~~ⓐ~~ correct	ⓑ incorrect

44 If 2 variables move in same direction i.e. an increase on the part variable introduces an increase on the part of other variable and Decrease on the part of one variable introduces decrease on the part of other variable also, then 2 variables are known to be positively correlated.

45 If 2 variables move in opposite direction i.e. an increase on the part variable introduces an decrease on the part of other variable and Decrease on the part of one variable results in increase on the part of other variable, then 2 variables are known to be Negatively correlated.

46 2 variables are known to be uncorrelated if movement on the part of one variable does not produce any measureable movement on the part of other variable.

47 1. Correlation coefficient (r) is unit free.
2. Correlation coefficient remains same in value, not necessarily in sign after shift of origin and change in scale.
3. Correlation coefficient lies between -1 and 1, including both limiting values.
4. b_{yx} & b_{xy} are also unit free
5. b_{yx}, b_{xy} and r will always have same sign.

48 For a group of 8 students, the sum of squares of diff. in ranks for maths & stats marks was found to be 50. What is the value of rank correlation coefficient?

$\Rightarrow$ $\Sigma d^2 = 50$, $n = 8$

$$r = 1 - \left[\frac{6\Sigma d^2}{n(n^2-1)}\right] = 1 - \left[\frac{6 \times 50}{8(8^2-1)}\right] = 1 - \frac{300}{504}$$

$$r = 0.40476$$

49 For a number of towns, correlation coefficient between people living below poverty line and increase of population is 0.50. If sum of squares of diff. in rank awarded to these factors are 82.50. Find number of towns.

$\Rightarrow$ $\Sigma d^2 = 82.50$, $r = 0.50$, $n = ?$

$$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)} \qquad 0.50 = 1 - \left[\frac{6 \times 82.50}{n(n^2-1)}\right]$$

$$\left[\frac{495}{n(n^2-1)}\right] = 0.50$$

$$\therefore \; n(n^2-1) = 990$$

$$\boxed{n = 10}$$

No. of towns = 10

My Notes

50 While computing rank correlation coefficient between profit and investments for 10 years of a firm, the diff of rank of one observation was taken as 7 instead of 5 and rank correlation coefficient was 0.80. What is correct value of rank correlation coefficient?

a. 0.95 ✓ b. 0.78 c. -0.80 d. None of these

$0.80 = 1 - \frac{6\Sigma d^2}{10\times 99}$

$\frac{6\Sigma d^2}{990} = 0.20$

wrong $\Sigma d^2 = 33$

correct $\Sigma d^2 = 33 - 7^2 + 5^2$
$= 9$

$r = 1 - \frac{6\times 9}{990}$

$= 1 - \frac{54}{990} = 0.94545$

51 Regression equations are derived from method of least squares.

52 Regression coefficient remain unchanged by shift of origin but affected due to change in scale.

a. If $u = 3 + x$, $v = y - 18$ then $b_{vu} = b_{yx}$ and $b_{uv} = b_{xy}$ → $r_{uv} = r_{xy}$, $\sigma_u = \sigma_x$, $\sigma_v = \sigma_y$ → $b_{vu} = r_{uv}\frac{\sigma_v}{\sigma_u} = r_{xy}\frac{\sigma_y}{\sigma_x}$, $b_{vu} = b_{yx}$, $b_{uv} = b_{xy}$

b. If $u = x + 17$, $v = y + 30$ then $b_{vu} = b_{yx}$, $b_{uv} = b_{xy}$

c. If $u = 3x + 18$, $v = 8y - 19$ then $b_{vu} =$, $b_{uv} =$ ⇒ $b_{vu} = r_{uv}\frac{\sigma_v}{\sigma_u} = r_{xy}\times\frac{8\times\sigma_y}{3\times\sigma_x}$, $b_{vu} = \frac{8}{3}\times r_{xy}\cdot\frac{\sigma_y}{\sigma_x} = \frac{8}{3}\times b_{yx}$, $b_{uv} = \frac{3}{8}\times b_{xy}$

d. If $u = 18x + 17$, $v = 2y - 20$ then $b_{vu} =$, $b_{uv} =$ ⇒ $b_{vu} = \frac{2}{18}\times b_{yx}$, $b_{uv} = \frac{18}{2}\times b_{xy}$

53 Two regression lines i.e. $(y - \bar{y}) = b_{yx}(x - \bar{x})$ and $(x - \bar{x}) = b_{xy}(y - \bar{y})$ intersect at point $(\bar{x}, \bar{y})$

∴ Reg. line will always intersect at point (AM of x, AM of y)

My Notes

① If $u = 3x - 19$, $v = -8y - 63$ then $r_{uv} = -r_{xy}$

② If $u = 8x + 50$, $v = 19y + 83$ then

$b_{uv} = \frac{8}{19}\times b_{xy}$

$b_{vu} = \frac{19}{8}\times b_{yx}$

54

r	b_{yx}	b_{xy}
0.80	5.80	0.11034482758
0.75	0.20	2.8125
-0.60	−0.26471	-1.36
−0.93808	-0.80	-1.10
0.2819	1.23619	0.06428430095

$r^2 = b_{yx} \cdot b_{xy}$

55 There are some cases when we may find a correlation between 2 variables although 2 variables are not casually related. This is due to existence of third variable which is related to both the variables under consideration, such a correlation is known as spurious correlation OR Non-sense correlation

56 Bi-variate data are data collected for :
a. 2 variables.
b. More than 2 variables.
c. 2 variables at same point of time.
d. 2 variables at diff. point of time.

57 If plotted paints in a scatter diagram lie from

Upper left to lower right then → NEGATIVE correlation.

Upper right to lower left then → POSITIVE correlation.

58 If plotted points in a scatter diagram are evenly distributed without depicting any pattern then there is no correlation beth 2 variables.

59 If plotted points in a scatter diagram lie on a single line then correlation is
a. Perfect Positive b. Perfect Negative c. a or b d. None of these

60 The correlation between shoe-size and intelligence is
a. Positive b. Negative c. Zero d. None of these

My Notes

$u = \frac{3x+18}{23}$ & $v = \frac{8y-93}{21}$ then

$= \frac{18}{23} + \frac{3}{23}x$ $v = -\frac{93}{21} + \frac{8}{21}y$

$\Longrightarrow$

$b_{uv} = \frac{3/23}{8/21} \times b_{xy} = \frac{63}{184} \times b_{xy}$

$b_{vu} = \frac{8/21}{3/23} \times b_{yx} = \frac{184}{63} \times b_{yx}$

Value of 'r' helps us to find Type/Nature of correlation as well as Degree of correlation

61 Product moment correlation coefficient is considered for ____________.

a. Finding nature of correlation
b. Finding degree of correlation
c. Both of these
d. None of these

62 If r is positive then points in a scatter diagram tend to cluster :

a. From lower left corner to upper right corner
b. From lower left corner to lower right corner
c. From lower right corner to upper left corner
d. None of these

63 The co-variance between 2 variables is :

a. Strictly positive
b. Strictly negative
c. Always zero
d. Either positive, negative or zero

Similarly SD = zero or positive

Variance = zero or positive

64 To find degree of agreement about beauty between 2 judges in a beauty contest, we use :

a. Scatter Diagram
b. Product moment correlation coefficient
c. Spearman's rank correlation coefficient
d. Coefficient of concurrent deviation

65 The diff. between observed value and estimated value in a regression analysis is known as Error or Residue.

66 What are the limits of 2 regression coefficient ?

a. No limit
b. Both must be positive
c. One positive & other negative
d. Product of 2 regression coefficients must be numerically less than unity.

67 Regression coefficients remain unchanged due to :

a. Shift of origin
b. Change of scale
c. Both a and b
d. Either a or b

My Notes

68 Correlation coefficient between 2 variables is -0.90, then coefficient of determination is :

a. 0.90 b. -0.81 c. 0.19 d. 0.81 ✓

69 Correlation coefficient between 2 variables is 0.70, then % of variation unaccounted for is :

a. 70% b. 49% c. 51% ✓ d. 100%

unaccounted variation $= 1 - 0.7^2 = 1 - 0.49 = 0.51$

70 If cov (x,y) = 15, then $\sigma_x \cdot \sigma_y$

$\Rightarrow \sigma_x \cdot \sigma_y \geq 15$

71 If u + 5x = 6 and 3y - 7v = 20. $(r)_{xy} = 0.58$ then $(r)_{uv}$ = ?

a. 0.58 b. -0.58 ✓ c. 0.84 d. -0.84

$u = 6 - 5x$

$u = -5x + 6$

$3y - 7v = 20$

$3y - 20 = 7v$

$v = -\frac{20}{7} + \frac{3}{7}y$

$r_{uv} = -r_{xy}$

$= -0.58$

72 If sum of squares of diff. in ranks, given by 2 judges A and B of 8 students is 21, what is the value of rank correlation coefficient?

a. 0.70 b. 0.65 c. 0.75 ✓ d. 0.80

$\Rightarrow r = 1 - \left[\frac{6 \times 21}{8(8^2-1)}\right] = 1 - \frac{126}{504} = \frac{378}{504} = 0.75$

73 For 10 pairs of observations, No. of concurrent deviations found to be 4. What is coefficient of concurrent deviation?

a. $\sqrt{0.20}$ b. $-\sqrt{0.20}$ c. 1/3 d. -1/3 ✓

$n = 10, \; m = 9, \; c = 4$

$r = \pm\sqrt{\pm\frac{2c-m}{m}} = -\sqrt{\frac{2(4)-9}{9}} = -\frac{1}{3}$

My Notes

$\frac{1}{3} = \sqrt{\frac{2c-m}{m}}$

$\sqrt{\frac{1}{3}} = \sqrt{\frac{2(6)-m}{m}}$

$\frac{1}{3} = \frac{12-m}{m}$

$m = 36 - 3m$

$4m = 36$

$m = 9$

$\therefore n = 10 = p$

74 The coefficient of concurrent deviation for ‘p’ pairs of observations was found to be $1/\sqrt{3}$ If no. of concurrent deviations was found to be 6. Value of ‘p’ is :

~~a. 10~~ b. 9 c. 8 d. None of these

75 If y = 4 + 3x is regression line of y on x. AM of x = -1; AM of y = ?

~~a. 1~~ b. -1 c. 7 d. None

AM of y $= 4 + 3\,(\text{AM of } x)$

$= 4 + 3\,(-1)$

$= 1$

76 2 regression lines are y = -2x+3 and 8x = -y+3. Find value of r.

a. 0.50 ~~b. - 0.50~~ c. $-1/\sqrt{2}$ d. None of these

$y = -2x + 3$

$y = 3 - 2x$

$b_{yx} = -2$

$8x = 3 - y$

$x = \frac{3}{8} - \frac{1}{8}y$

$b_{xy} = -\frac{1}{8}$

$r^2 = b_{yx} \times b_{xy} = -2 \times -\frac{1}{8} = \frac{1}{4}$

$r = -\frac{1}{2} = -0.50$

77 Given the following equations 2x - 3y = 10 and 3x + 4y = 15, which one is the regression equation of x on y.

a. 3x+4y=15 b. 2x-3y=10 c. Both ~~d. None~~

x on y

$2x - 3y = 10$

$2x = 10 + 3y$

$x = 5 + \frac{3}{2}y$

$b_{xy} = \frac{3}{2} = 1.50$

y on x

$3x + 4y = 15$

$4y = 15 - 3x$

$y = \frac{15}{4} - \frac{3}{4}x$

$b_{yx} = -0.75$

78 2 regression lines are given by : 8x + 10y = 25 and 16x + 5y = 12. & Variance of x = 25, SD of y = ?

a. 16 ~~b. 8~~ c. 64 d. 4 e. None of these

Reg line of y on x

$8x + 10y = 25$

$10y = 25 - 8x$

$y = \frac{25}{10} - \frac{8}{10}x$

$b_{yx} = -0.80$

Reg line of x on y

$16x = 12 - 5y$

$x = \frac{12}{16} - \frac{5}{16}y$

$b_{xy} = -0.3125$

$r^2 = -0.80 \times -0.3125$

$r = -0.50$

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

$-0.80 = -0.50 \times \frac{\sigma_y}{5}$

$\sigma_y = 8$

79

Variables	Nature of Correlation
1. Profit of insurance company and no. of claims	Negative
2. Demand for goods and their prices under normal circumstances	Negative
3. Years of education and Income	Positive
4. Amount of rainfall and Yield of crop	Positive
5. Sale of woollen garments and temperature	Negative

80 For the bivariate data [(20,5), (21,4), (22,3)] the correlation coefficient between x and y is

a. zero b. 1 c. -1 d. 0.50

x	20	21	22
y	5	4	3

81 r = 0.48, cov (x,y) = 36, SD of x = 16, SD of y = ?

a. 18.75 b. -18.75 c. 16.75 d. None of these

$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$ $0.48 = \frac{36}{16 \times \sigma_y}$ $\therefore \sigma_y = 4.6875$

82 r = 0.52, cov (x,y) = 7.80, Variance of x = 16, SD of y = ?

a. 2.85 b. 3.25 c. 1.25 d. 3.75

$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$ $0.52 = \frac{7.80}{4 \times \sigma_y}$ $\therefore \sigma_y = 3.75$

83 If r = 0.40 then coefficient of determination and coefficient of non-determination are resp.

a. 0.16, 0.84 b. 0.36,0.64 c. 0.60,0.40 d. None

Coeffi. of Determination $= r^2 = 0.4^2 = 0.16$

Coeffi. of Non-determination $= 1 - r^2 = 0.84$

84 Simple correlation is known as :

a. Linear correlation b. Non-linear correlation
c. Non-sense correlation d. None of these

85 Slope of regression equation of x on y is :

Slope of the line $ax + by + c = 0$ is $-a/b$

a. b_{xy} b. b_{yx} c. $1/b_{xy}$ d. $1/b_{yx}$

Reg. line of x on y is

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

$x - \bar{x} = b_{xy} \cdot y - b_{xy} \cdot \bar{y}$

$x - b_{xy} \cdot y = \bar{x} - b_{xy} \cdot \bar{y}$

slope $= \frac{-1}{-b_{xy}} = \frac{1}{b_{xy}}$

86 Slope of regression equation of y on x is :

a. b_{xy} b. b_{yx} c. $1/b_{xy}$ d. $1/b_{yx}$

Reg line of y on x

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

$$y - \bar{y} = b_{yx}\cdot x - b_{yx}\cdot \bar{x}$$

$$b_{yx}\cdot x - y = b_{yx}\cdot \bar{x} - \bar{y}$$

87 $(r)_{xy} = (r)_{yx}$

a. correct b. wrong c. can't say d. None of these

slope of the line $= -b_{yx}/-1 = b_{yx}$

88 b_{yx} is always same as b_{xy}

a. correct b. wrong

89 Covariance measures Joint variation between 2 variables.

a. Joint b. Common c. Relative d. None of these

$$\text{variance of } (x, x) = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{\sum x \cdot x}{n} - \bar{x}\cdot\bar{x}$$

$$\text{co-variance of } (x, y) = \frac{\sum x \cdot y}{n} - \bar{x}\cdot\bar{y}$$

90 Karl Pearson's Product Moment Correlation Coefficient $= \left[\dfrac{\text{covariance of } (x, y)}{\sigma_x \cdot \sigma_y}\right]$

$$= \left[\frac{\frac{\sum xy}{n} - \bar{x}\cdot\bar{y}}{\sqrt{\frac{\sum x^2}{n} - \bar{x}^2} \times \sqrt{\frac{\sum y^2}{n} - \bar{y}^2}}\right] = \left[\frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2} \times \sqrt{\sum (y-\bar{y})^2}}\right]$$

Spearman's Rank Correlation Coefficient = without tie : $1 - \left[\dfrac{6\sum d^2}{n(n^2-1)}\right]$

with tie : $1 - \left[\dfrac{6\left(\sum d^2 + \sum \frac{t^3 - t}{12}\right)}{n(n^2-1)}\right]$

Coefficient of Concurrent Deviation $= \pm\sqrt{\pm\left(\dfrac{2c - m}{m}\right)}$

91 $b_{yx} = 1.20$ $b_{xy} = 0.90$; then r = ?

a. 1.039 b. -1.039 c. 1.08 d. Wrong data

$r^2 = b_{yx} \times b_{xy} = 1.20 \times 0.90 = 1.08$

As max. value of $r^2 = 1.00$, This is wrong data.

92 If $\bar{x} = 30$, $\bar{y} = 90$, $\sigma_x = 8$, $\sigma_y = 5$, $r = -0.75$. Find Reg. equation of y on x.

a. Joint b. Common c. Relative d. None of these

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

$y - 90 = -0.75 \times \frac{5}{8} \times (x - 30)$

$y - 90 = -0.46875(x - 30)$

$y - 90 = -0.46875x + 14.0625$

$\boxed{y = 104.0625 - 0.46875x}$

93 If $\Sigma(x-\bar{x})(y-\bar{y}) = 30$, n = 3. Find cov (x,y)

$cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n} = \frac{30}{3} = 10$

94 If cov (x,y) = 36, $\sigma_x = 9$, $\sigma_y = 4$. Find r

a. 1.00 b. -1.00 c. 0 d. None

$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} = \frac{36}{9 \times 4} = 1.00$

95 correlation coefficient is also known as measure of association between 2 variables.

96 Karl Pearson's product moment correlation coefficient is the best method to obtain correlation between 2 variables.

My Notes

97 If Reg line of y on x is 3x + 8y = 13y - 63x + 103. Find b_{yx}

$\Longrightarrow$ $3x + 8y = 13y - 63x + 103$

$-5y = 103 - 66x$

$y = \frac{-103}{5} + \frac{66}{5}x$ $\quad \therefore b_{yx} = \frac{66}{5} = 13.20$

98 If Reg line of x on y is 16x - y = 93x - 21y + 83. Find b_{xy}

$\Longrightarrow$ $16x - y = 93x - 21y + 83$

$-77x = 83 - 20y$

$x = \frac{-83}{77} + \frac{20}{77}y$ $\quad \therefore b_{xy} = 20/77$

$= 0.25974$

99 If r = -0.63812, b_{yx} = -1.36822, b_{xy} = ?

$r^2 = b_{yx} \cdot b_{xy}$

$(-0.63812)^2 = -1.36822 \times b_{xy}$ $\quad \therefore b_{xy} = -0.297611$

100 Correlation between temperature of city and sale of cold drinks is :

a. Positive ✓ b. Negative c. Zero d. Can't say