Correlation & Regression analysis

① Correlation is the process of studying / establishing relationship / association between 2 or more variables although they are not in proportion.

examples :

ⓐ sale of cold-drinks & Temperature has positive correlation

ⓑ No of trees & Rainfall has positive corr.

ⓒ No. of accident & claims, profit of insurance company has negative correlation

ⓓ No. of accidents in pune & USA's GDP has No correlation.

② studying correlation between

- 2 variables → Bi-variate correlation
- 3 or more variables → Multi-variate correlation → Not in CA Foundation syllabus

③ Whether correlation beth 2 variables exists or not ?

Yes

- What is type/Nature of correlation ?
  - Positive correlation
  - Negative correlation
- What is Degree of correlation ?
  1. Low/weak Degree
  2. Moderate degree
  3. High/strong degree
  4. perfect degree

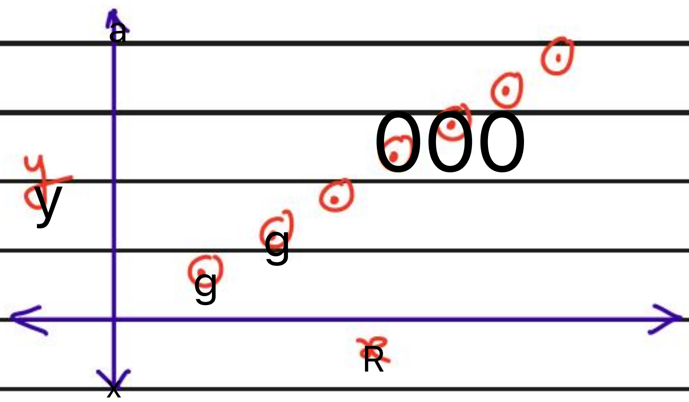
NO

④ There are 4 methods to measure correlation between 2 variables :

1. Scatter Diagram
2. Spearman's rank correlation coefficient
3. Coefficient of concurrent deviation
4. Karl pearson's product moment correlation coefficient (The Best method)

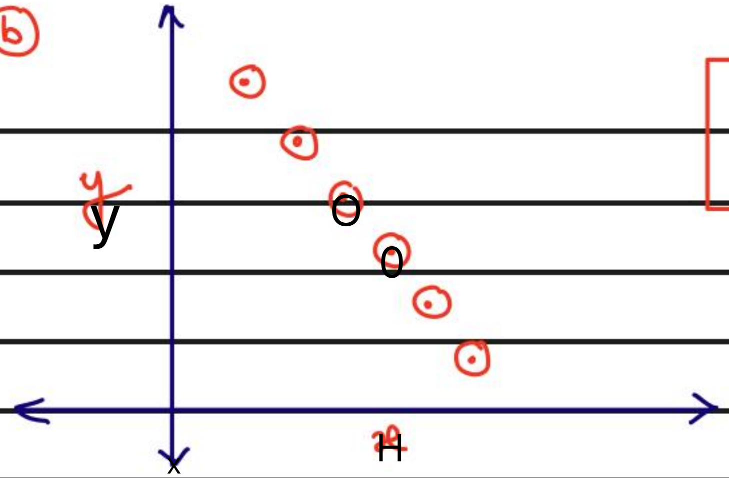
⑤ Scatter diagram

ⓐ



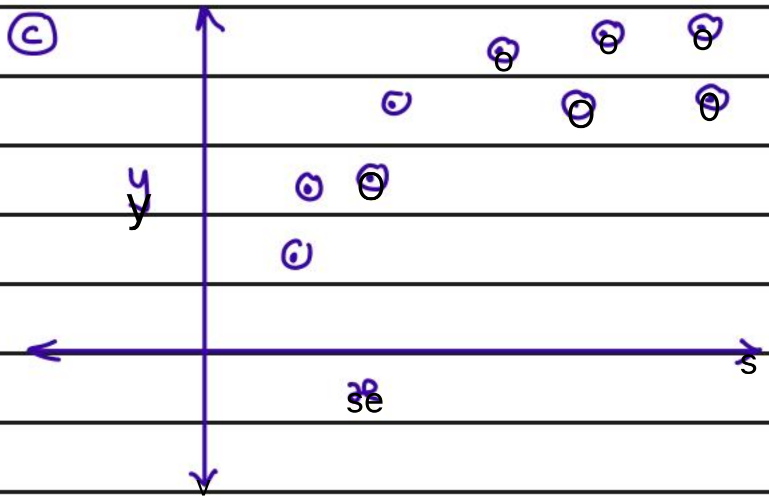
Scatter diagram showing perfect positive correlation

(b)



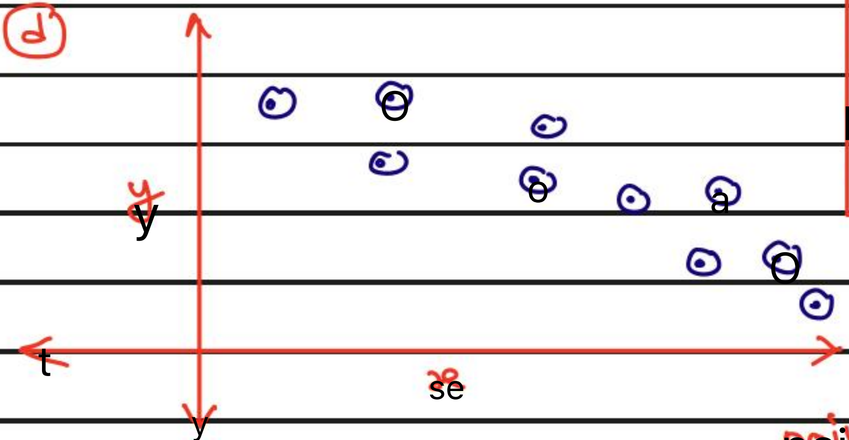
Scatter diagram showing
$\boxed{\text{Perfect Negative correlation}}$

(c)



$\boxed{\text{Scatter diagram showing positive (Relatively) correlation}}$

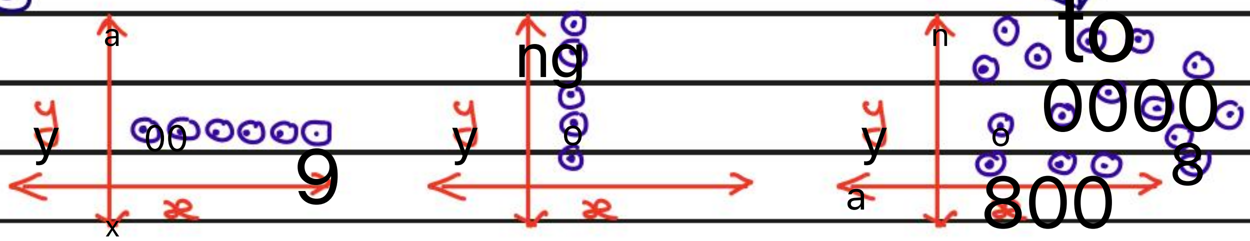
(d)



$\boxed{\text{Scatter Diagram showing Negative (Relatively) correlation}}$

(e)

points are scattered without depicting any pattern



Scatter Diagrams showing <u>No correlation</u>

⑥ Scatter diagrams can give an idea about Type / Nature of correlation but it can not give the exact degree of correlation.

To know Type of correlation as well as degree of correlation we have following 3 methods:

① Spearman's rank correlation coefficient (r)
② Coefficient of concurrent deviation
③ Karl pearson's product moment correlation coeffi.

⑦ Find spearman's rank correlation coefficient for

x	56	59	73	81	79	52	48	88	93	2
y	280	730	831	631	789	123	666	581	983	281

$\Longrightarrow$

x	y	Rank of x	Rank of y	d^2
56	280	7	9	4
59	730	6	4	4
73	831	5	2	9
81	631	3	6	9
79	789	4	3	1
52	123	8	10	4
48	666	9	5	16
88	581	2	7	25
93	983	1	1	0
2	281	10	8	4
			Σd^2	76

$n = 10$

(Spearman's rank correlation coefficient) $= r$

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

$$= 1 - \left[\frac{6 \times 76}{10(10^2-1)}\right]$$

$$r = 1 - \frac{456}{990} = 0.5394$$

There is a moderate degree of positive correlation

8

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

$r = 1.00$	Perfect positive correlation
$0.80 < r < 1.00$	High/strong degree of positive correlation
$0.30 < r < 0.80$	Moderate degree of positive correlation
$0 < r < 0.30$	Weak/Low degree of positive correlation
$r = 0$	No correlation
$-0.30 < r < 0$	Weak/Low degree of negative correlation
$-0.80 < r < -0.30$	Moderate degree of negative correlation
$-1.00 < r < -0.80$	High/strong degree of negative correlation
$r = -1.00$	Perfect negative correlation

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

9 Find spearman's rank correlation coeffi. for

x	123	236	111	886	781	336	893
y	1013	986	993	781	286	583	960

$\Rightarrow$

x	123	236	111	886	781	336	893
y	1013	986	993	781	286	583	960
Rank of x	6	5	7	2	3	4	1
Rank of y	1	3	2	5	7	6	4
d^2	25	4	25	9	16	4	9

$\Sigma d^2 = 92$

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

$$= 1 - \left[\frac{6 \times 92}{7 \times (49-1)}\right] = 1 - \left(\frac{552}{336}\right)$$

$$= -0.642857$$

10) If scatter diagram is showing a straight line from

	correlation is
Lower Left to upper Right	Perfect positive
upper Left to Lower Right	Perfect Negative
Lower Right to upper Left	Perfect negative
upper Right to Lower left	Perfect positive

(11) Find spearman's rank correlation coefficient for

x	11	13	9	14	16	25	31	38
y	46	49	51	89	68	78	33	25

$\Longrightarrow$

x	y	Rank of x	Rank of y	d^2
11	46	7	6	1
13	49	6	5	1
9	51	8	4	16
14	89	5	1	16
16	68	4	3	1
25	78	3	2	1
31	33	2	7	25
38	25	1	8	49

$n = 8$

$\Sigma d^2 = 110$

Spearman rank correlation coefficient $(r) = 1 - \left[\dfrac{6\,\Sigma d^2}{n(n^2-1)}\right]$

$$= 1 - \frac{6 \times 110}{8 \times (64-1)} = 1 - \frac{660}{504}$$

$$= -0.30952$$

There is moderate degree of negative correlation between x & y.

(12) $r = 1.00$: Represents perfect positive correlation

$r = -1.00$: Represents perfect Negative correlation

$r = 0$: Represents $\underline{\text{No correlation}}$

Maximum value of $r = 1.00$

Minimum value of $r = -1.00$

$0 < r \leq 1.00$: Represents positive correlation

$-1 \leq r < 0$: Represents negative correlation

(13) Which of the following is correct

(a) $0 < r \leq 1.00$

(b) $-1.00 \leq r < 0$

(c) $-1.00 < r \leq 1.00$

(d) $-1.00 \leq r \leq 1.00$

Correlation coefficient will always lie between -1.00 and 1.00 including both limits.

(14) Find spearman's rank correlation coeffi for

X	33	381	231	583
Y	1081	2356	5523	1234

x	y	Rank of x	Rank of y	d^2
33	1081	4	4	0
381	2356	2	2	0
231	5523	3	1	4
583	1234	1	3	4

$\Sigma d^2 = 8$

$$r = 1 - \left[\frac{6 \Sigma d^2}{n(n^2-1)}\right] = 1 - \left(\frac{6 \times 8}{4 \times 15}\right)$$

$$= 1 - \frac{48}{60}$$

$$= 1 - 0.80$$

$$= 0.20$$

∴ There is small / weak / Low degree of positive correlation between x & y.

(11) Find Spearman's rank correlation coefficient for

x	11	13	9	14	16	25	31	38
y	46	49	51	89	68	78	33	25

$\Longrightarrow$

x	y	Rank of x	Rank of y	d^2
11	46	7	6	1
13	49	6	5	1
9	51	8	4	16
14	89	5	1	16
16	68	4	3	1
25	78	3	2	1
31	33	2	7	25
38	25	1	8	49
			$\Sigma d^2 =$	110

d = difference of rank

$$r = 1 - \left[\frac{6\Sigma d^2}{n(n^2-1)}\right] = 1 - \frac{6 \times 110}{8 \times 63} = 1 - \frac{660}{504}$$

$$= -0.3095$$

There is moderate degree of negative correlation between x & y.

(12) $r = 1.00$: represents perfect positive correlation

$r = -1.00$: represents perfect negative correlation

$r = 0$: represents No correlation

Maximum value of $r = 1.00$

Minimum value of $r = -1.00$

(13) Which of the following is correct

(a) $0 < r \le 1.00$

(b) $-1.00 \le r \le 0$

(c) $-1.00 < r < 1.00$

~~(d)~~ $-1.00 \le r \le 1.00$

Correlation coefficient will always lie between -1.00 and 1.00 including both limits.

(14) Find Spearman's rank correlation coeffi for

x	33	381	231	583
y	1081	2356	5523	1234

→

x	y	Rank of x	Rank of y	d^2
33	1081	4	4	0
381	2356	2	2	0
231	5523	3	1	4
583	1234	1	3	4
			Σd^2	8

Spearman's rank correlation coefficient $= 1 - \left[\dfrac{6\Sigma d^2}{n(n^2-1)}\right]$

$$= 1 - \frac{6 \times 8}{4(4^2-1)}$$

$$= 1 - \frac{48}{60} = 1 - 0.80 = 0.20$$

There is small / weak / Low degree of positive correlation between x & y.

(15) Find Spearman's rank correlation coefficient for :

(Imp)

x	11	15	16	19	15	13	19	21	15	15	28
y	123	831	583	236	583	781	123	123	281	560	281

$\Rightarrow$

x	y	Rank of x	Rank of y	d^2
11	123	11	10	1
15	831	7.50	1	42.25
16	583	5	3.50	2.25
19	236	3.50	8	20.25
15	583	7.50	3.50	16
13	781	10	2	64
19	123	3.50	10	42.25
21	123	2	10	64
15	281	7.50	6.50	1
15	560	7.50	5	6.25
28	281	1	6.50	30.25

$\Sigma d^2 = 289.50$

t = no. of observations involved in a tie = 2, 4, 2, 2, 3

$$\Sigma\left(\frac{t^3-t}{12}\right) = \left[\frac{2^3-2}{12} + \frac{4^3-4}{12} + \frac{2^3-2}{12} + \frac{2^3-2}{12} + \frac{3^3-3}{12}\right]$$

$$= 0.50 + 5 + 0.50 + 0.50 + 2$$

$$= 8.50$$

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

$$= 1 - \frac{6(289.50 + 8.50)}{11(11^2-1)} = 1 - \frac{1788}{1320}$$

$$= -0.35454545$$

Moderate degree of negative correlation betn x & y.

(16) Spearman's rank correlation coefficient

without tie

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

with tie

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

where
d = Diff of rank
t = No. of obsns involved in a tie
n = No. of pairs of observations

(17) Find spearman's rank correlation coefficient

x	50	80	50	50	45
y	81	93	56	56	28

$\Rightarrow$

x	y	Rank of x	Rank of y	d^2
50	81	3	2	1
80	93	1	1	0
50	56	3	3.50	0.25
50	56	3	3.50	0.25
45	28	5	5	0
				1.50

$t = 3, 2$

$$\sum\left(\frac{t^3-t}{12}\right) = \frac{3^3-3}{12} + \frac{2^3-2}{12} = 2 + 0.50 = 2.50$$

$$r = 1 - \frac{6\left(\sum d^2 + \sum \frac{t^3-t}{12}\right)}{n(n^2-1)} = 1 - \frac{6(1.50+2.50)}{5 \times 24} = 1 - \frac{24}{120} = 1 - 0.20 = 0.80$$

18. Find coeffi. of concurrent deviation for

x	58	63	28	93	58	26	33	38	59	83
y	123	633	833	589	289	333	533	888	231	123

⟹

x	y	Deviation of x	Deviation of y	product
58	123			
63	633	+	+	+
28	833	−	+	−
93	589	+	−	−
58	289	−	−	+
26	333	−	+	−
33	533	+	+	+
38	888	+	+	+
59	231	+	−	−
83	123	+	−	−

$n = 10$

$m = n - 1$

$= 10 - 1 = 9$

$c = \left(\text{No. of concurrent deviations}\right)$

$= \left(\text{No. of '+' signs in product column}\right)$

$= 4$

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

$= \pm\sqrt{\pm\frac{2(4)-9}{9}} = -\sqrt{-\left(-\frac{1}{9}\right)} = -\sqrt{\frac{1}{9}}$

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

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

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

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

(20) If No. of positive signs in product column $= x$, no. of negative sign in product column $= y$

and $x > y$	$1.00 \ge r > 0$
$x < y$	$-1.00 \le r < 0$
$x = y$	$r = 0$

(21) Find coeffi. of concurrent deviation and Spearman's rank correlation coefficient.

x	10	18	10	26	11
y	53	91	98	53	28

⟹

x	y	Rank of x	Rank of y	d^2	Deviation of x	Deviation of y	product
10	53	4.50	3.50	1			
18	91	2	2	0	+	+	+
10	98	4.50	1	12.25	−	+	−
26	53	1	3.50	6.25	+	−	−
11	28	3	5	4	−	−	+

$t = 2, 2$

$\Sigma d^2 = 23.50$

$\Sigma \frac{t^3 - t}{12} = 0.50 + 0.50 = 1.00$

$c = 2$

$m = 4$

Spearman's coefficient $= 1 - \frac{6(23.50 + 1.00)}{5 \times 24}$

$= -0.225$

Coeffi. of concurrent deviation

$= \pm\sqrt{\pm\frac{2(2)-4}{4}}$

$= 0$

(22) ① coeffi. of concurrent deviation method ignores the amount or magnitude of change but it considers only direction of change. ∴ we can say that, it is a casual to obtain correlation coefficient.

② In order to know degree of agreement of 2 Judges in a Beauty contest we can use: Spearman's rank correlation coefficient

23) Find Karl pearson's product moment correlation coefficient for

x	20	30	28	72
y	82	70	38	40

$\Rightarrow$

x	y	x^2	y^2	$x \cdot y$
20	82			
30	70			
28	38			
72	40			
150	230	7268	14668	7684

AM of $x = \bar{x} = \frac{150}{4} = 37.500$

AM of $y = \bar{y} = \frac{230}{4} = 57.500$

SD of $x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{7268}{4} - 37.50^2} = 20.27$

SD of $y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{14668}{4} - 57.50^2} = 18.99$

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

$= \frac{7684}{4} - (37.50 \times 57.50)$

$= -235.25$

Karl Pearson's coefficient $= \frac{cov(x,y)}{SD_x \cdot SD_y} = \frac{-235.25}{20.27 \times 18.99} = -0.6115$

Moderate degree of Negative correlation

(24)

Karl pearson's product moment correlation coefficient

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

24

i) Karl pearson's product moment correlation coefficient

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

$$= \left(\frac{\text{covariance of } (x,y)}{\sigma_x \cdot \sigma_y} \right)$$

ii) coefficient of concurrent deviation

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

iii) spearman's rank correlation coefficient

without tie

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

with tie

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

25. In case of Qualitative data / measurements which of the following is suitable

(a) Scatter Diagram
(b) Spearman's coefficient
(c) Karl Pearson's coefficient
(d) coeffi. of concurrent deviation

26.

x	10	18	12	20
y	30	20	25	15

Find correlation coefficient (Karl pearson's)

$\Longrightarrow$ $\bar{x} = 15$

$\bar{y} = 22.50$

$$SD_x = \sqrt{\frac{968}{4} - 15^2} = 4.1231$$

$$SD_y = \sqrt{\frac{2150}{4} - 22.50^2} = 5.5902$$

$$cov(x,y) = \frac{1260}{4} - (15 \times 22.50) = -22.50$$

$$r = \left[\frac{cov(x,y)}{\sigma_x \sigma_y}\right] = \frac{-22.50}{4.1231 \times 5.5902} = -0.9762$$

very strong / high degree of Negative correlation exists between x & y

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

$$\sigma_x = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - (\bar{y})^2}$$

$$r = \left[\frac{\text{covariance of } x, y}{SD_x \times SD_y}\right]$$

Karl Pearson's method

27) Find Karl pearson's coefficient

x	50	60	48	72	30
y	112	118	138	92	60

⇒ $\bar{x} = 52$

$\bar{y} = 104$

$SD_x = \sqrt{\dfrac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\dfrac{14488}{5} - 52^2} = 13.91402$

$SD_y = \sqrt{\dfrac{\Sigma y^2}{n} - \bar{y}^2} = \sqrt{\dfrac{57576}{5} - 104^2} = 26.4424$

$cov(x,y) = \dfrac{\Sigma xy}{n} - \bar{x}\cdot\bar{y} = \dfrac{27728}{5} - (52 \times 104) = 137.60$

$r = \dfrac{cov(x,y)}{\sigma x \cdot \sigma y} = \left(\dfrac{137.60}{13.91402 \times 26.4424}\right) = 0.374$

Moderate degree of positive correlation

28)

x	11	12	9	18	10
y	15	12	33	8	22

Find correlation coeffi. by all 3 methods

⇒

①

x	y	Rank of x	Rank of y	d^2	Devi. of x	Devi. of y	product	x^2	y^2	xy
11	15	3	3	0						
12	12	2	4	4	+	−	−			
9	33	5	1	16	−	+	−			
18	8	1	5	16	+	−	−			
10	22	4	2	4	−	+	−			
				$\Sigma d^2 = 40$				770	2006	970

i) spearman's rank correlation coefficient $= 1 - \left[\dfrac{6\Sigma d^2}{n(n^2-1)}\right] = 1 - \left(\dfrac{6 \times 40}{5 \times 24}\right)$

$= 1 - \dfrac{240}{120} = 1 - 2 = -1.00$

ii) coefficient of concurrent deviation $= \pm\sqrt{\pm\frac{(2c-m)}{m}} = \pm\sqrt{\pm\frac{(2\times 0-4)}{4}}$

$= -1.00$

iii) Karl pearsons product moment correlation coefficient $= \left[\frac{\frac{\sum xy}{n} - \bar{x}\cdot\bar{y}}{\sqrt{\frac{\sum x^2}{n}-\bar{x}^2}\times\sqrt{\frac{\sum y^2}{n}-\bar{y}^2}}\right]$

$$= \frac{\frac{970}{5} - (12\times 18)}{\sqrt{\frac{770}{5}-12^2}\sqrt{\frac{2006}{5}-18^2}} = \left(\frac{-22}{3.1623\times 8.7864}\right) = -0.7918$$

(29) Find correlation coeffi. by all 3 methods (Home-work)

x	315	833	292	300
y	400	282	188	150

⇒

① Karl pearson's coefficient = 0.1926

② Spearman's rank correlation coefficient = 0.600

③ Coeffi. of concurrent deviation = −0.5773

(30) Find correlation coeffi. by all 3 methods

x	10	30
y	20	100

⇒

x	y	Rank of x	Rank of y	d^2	Devi. of x	Devi. of y	product	xy
10	20	2	2	0				200
30	100	1	1	0	+	+	+	3000
								3200

Spearman's rank correlation coefficient $= 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6\times 0}{2(2^2-1)} = 1 - 0 = 1.00$

Co effi. of concurrent deviation $= \pm\sqrt{\pm\left(\frac{2c-m}{m}\right)} = \pm\sqrt{+\left(\frac{2\times 1-1}{1}\right)} = 1.00$

Karl Pearson's coefficient $= \left[\frac{cov(x,y)}{\sigma_x \cdot \sigma_y}\right] = \frac{\frac{3200}{2} - (40\times 60)}{10\times 40} = \frac{400}{400} = 1.00$

(31) Find correlation co effi. by all 3 methods

x	60	90
y	20	10

$\Longrightarrow$

x	y	Rank x	Rank y	d^2	Devi. of x	Devi. of y	product	xy
60	20	2	1	1				1200
90	10	1	2	1	+	−	−	900

$\Sigma xy = 2100$

Spearman's rank correlation coefficient $= 1 - \left[\frac{6\Sigma d^2}{n(n^2-1)}\right] = 1 - \frac{6\times 2}{2(2^2-1)} = 1 - \frac{12}{6}$

$= 1 - 2 = -1.00$

Co effi. of concurrent deviation $= \pm\sqrt{\pm\left(\frac{2c-m}{m}\right)} = -\sqrt{-\left(\frac{2(0)-1}{1}\right)} = -1.00$

Karl Pearson's coefficient $= \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} = \left[\frac{\frac{2100}{2} - (75\times 15)}{15\times 5}\right]$

$= \frac{-75}{75} = -1.00$

32

When $n = 2$ then $r = 1.00$ OR $r = -1.00$

x	y
10	30
20	80

$r = 1.00$

x	y
50	800
15	600

$r = 1.00$

x	y
100	8500
150	6800

$r = -1.00$

x	y
1000	2800
850	3500

$r = -1.00$

33

x	80	60	70	100
y	35	65	75	25

Find $cov(x,y)$

$\Longrightarrow$

x	y	xy
80	35	
60	65	
70	75	
100	25	
310	200	14450

$\bar{x} = \frac{310}{4} = 77.50$

$\bar{y} = \frac{200}{4} = 50$

covariance of $x, y = \frac{\sum xy}{n} - \bar{x}\cdot\bar{y}$

$= \frac{14450}{4} - (77.50 \times 50) = -262.50$

OR

x	$(x-\bar{x})$	y	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
80	2.50	35	−15	−37.50
60	−17.50	65	15	−262.50
70	−7.50	75	25	−187.50
100	22.50	25	−25	−562.50
				−1050

$cov(x,y) = \left[\frac{\sum(x-\bar{x})(y-\bar{y})}{n}\right]$

$= \frac{-1050}{4}$

$= -262.50$

34) $\sum[(x-\bar{x})(y-\bar{y})] = 1200$

$n = 22$

Find $\text{cov}(x, y)$

$\Longrightarrow$ covariance of $(x, y) = \dfrac{\sum[(x-\bar{x})(y-\bar{y})]}{n}$

$= \dfrac{1200}{22} = 54.54545454$

35) If $\text{cov}(x, y) = 125$, $\sigma_x = 10.50$, $\sigma_y = 13.80$

Find r

~~(a)~~ 0.8626 (b) −0.8626 (c) a or b (d) can't say

$r = \left(125 / 10.50 \times 13.80\right) = 0.8626$

36) If $\text{cov}(x, y) = 138.50$, $\sigma_x = 8.53$, $\sigma_y = 9.80$

Find r

(a) −1.6568 (b) 1.6568 (c) a or b ~~(d)~~ wrong data

$r = \left(\dfrac{138.50}{8.53 \times 9.80}\right) = 1.6568$ as r can't be greater than 1.00

37)

x	y
30	80
40	78

find r

(a) 1.00 ~~(b)~~ −1.00 (c) 0.000 (d) can't find

38) If $\text{cov}(x, y) > 0$ then

~~(a)~~ $1.00 \geq r > 0$

(b) $1.00 \geq r \geq -1$

(c) $0 \geq r > -1$

(d) None of these

39) If $\text{cov}(x, y) \leq 0$ then

$-1 \leq r \leq 0$

(40)

If $cov(x, y) = 0$ then r is also ZERO

$cov(x, y) =$ positive then r is also positive

$cov(x, y) =$ Negative then r is also Negative

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

(41)

Karl pearson's product moment correlation coefficient $= \left[\dfrac{cov(x, y)}{SD_x \times SD_y} \right]$

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

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

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

(42)

$\sum (x-\bar{x})(y-\bar{y}) = 2500$

$\sum (x-\bar{x})^2 = 10{,}000$

$\sum (y-\bar{y})^2 = 8{,}250$

Find 'r'

$$\Longrightarrow r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2} \times \sqrt{\sum (y-\bar{y})^2}} = \left(\frac{2500}{\sqrt{10000} \times \sqrt{8250}} \right) = 0.27524$$

x	10	22	63	35
y	5	3	2	10

Find r_{xy}

$\Rightarrow$

$\bar{x} = 32.50$

$\bar{y} = 5.00$

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

$$= \sqrt{\frac{5778}{4} - 32.50^2}$$

$$= 19.7041$$

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

$$= \sqrt{\frac{138}{4} - 5^2}$$

$$= 3.08221$$

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

$$= \frac{592}{4} - (32.50 \times 5)$$

$$= -14.50$$

$$r_{xy} = \left(\frac{-14.50}{19.7041 \times 3.08221}\right)$$

$$r_{xy} = -0.2388$$

$u = 2x + 10$	30	54	136	80
$v = 5y - 8$	17	7	2	42

Find r_{uv}

$\Rightarrow$

$\bar{u} = 75$

$\bar{v} = 17$

$$SD_u = \sqrt{\frac{\Sigma u^2}{n} - \bar{u}^2}$$

$$= \sqrt{\frac{28712}{4} - 75^2}$$

$$= 39.40812$$

$$SD_v = \sqrt{\frac{\Sigma v^2}{n} - \bar{v}^2}$$

$$= \sqrt{\frac{2106}{4} - 17^2}$$

$$= 15.4110$$

$$cov(u,v) = \frac{\Sigma uv}{n} - \bar{u}\cdot\bar{v}$$

$$= \frac{4520}{4} - (75 \times 17)$$

$$= -145$$

$$r_{uv} = \frac{-145}{39.40812 \times 15.4110}$$

$$r_{uv} = -0.2388$$

'r' is not affected by change of origin as well as by change in scale

covariance is affected only by change in scale not by change of origin

(44) Value of 'r' helps us to know:

(a) Type/Nature of correlation

(b) Degree of correlation

~~(c)~~ Both of these

(d) None of these

(45) Scatter diagram can help use to know:

~~(a)~~ Type/Nature of correlation

(b) Degree of correlation

(c) Both of these

(d) None of these

(46)

x	30	35	39	40	38	60	63
y	40	38	28	39	36	80	98

To Find 'r' by concurrent dev. method

Find c, m. Also Find 'r'

⟹

Devi. of x	+	+	+	−	+	+
Devi. of y	−	−	+	+	−	+
product	−	−	+	−	−	+

$c = 2$, $m = 6$

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

$$r = -\sqrt{1/3} = -0.57735$$

As $\left(\frac{2c-m}{m}\right)$ is negative, r is negative

(47) If $n = 4$, $\Sigma d^2 = 49$ Find Spearman's coefficient.

(a) -3.90 (b) 3.90 (c) a or b ~~(d)~~ wrong data

As $-1.00 \le r \le 1.00$

(48) $SD_x = 3$ If $u = 8x + 20$

$SD_y = 5$ If $v = -3y + 33$

Find SD_u, SD_v

$\Rightarrow$

$$SD_u = |b| \times SD_x = 8 \times 3 = 24$$

$$SD_v = |-3| \times SD_y = 3 \times 5 = 15$$

$$Cov(u, v) = 8 \times -3 \times Cov(x, y)$$

$$Cov(u, v) = -24 \times Cov(x, y)$$

(49) $u = mx + 33$

$v = ky - 88$

then $cov(u, v) = m \times k \times cov(x, y)$

(50) $u = mx - 86$

$v = jy + 2000$

then

$SD_u = |m| \times SD_x$

$SD_v = |j| \times SD_y$

$cov(u, v) = m \times j \times cov(x, y)$

S1

$y = 3x + 18$ $x = 8y - 33$ $r_{xy} = -0.80$ $r_{yx} = -0.80$	$y = -3x + 33$ $x = -18y - 33$ $r_{xy} = -0.57$ $r_{yx} = -0.57$	$y = 8x + 55$ $x = -33y + 21$ $r_{xy} = -0.80$ $r_{yx} = 0.80$
$y = -3x + 21$ $x = 88y - 33$ $r_{xy} = 0.56$ $r_{yx} = -0.56$	$y = \frac{3}{2}x + 55$ $x = -5y + 23$ $r_{xy} = -0.583$ $r_{yx} = 0.583$	$y = 18x + 22$ $x = 33y + 56$ $r_{xy} = 0.81812$ $r_{yx} = 0.81812$
$y = 13x - 212$ $x = -18y + 63$ $r_{yx} = -0.63$ $r_{xy} = 0.63$	$y = -15x - 21$ $x = -18y + 33$ $r_{yx} = 0.2121$ $r_{xy} = 0.2121$	$y = -1.5x + 21$ $x = 81y - 33$ $r_{yx} = -0.8651$ $r_{xy} = 0.8651$

S2

If $\Sigma d^2 = 2$, $n = 4$, find Spearman's coefficient

$\Rightarrow$

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

$$= 1 - \frac{6 \times 2}{4(4^2-1)}$$

$$= 1 - \frac{12}{60} = 1 - 0.20 = 0.80$$

53

Vinod Reddy – C.E.O.

Rishit – Marketing Director.

x = mktg exp for the year

y = Turnover of the year

$\gamma_{xy} = 0.8123$

Question from VR to Rishit

$x = 60$ crores $\qquad y = ?$

$y = 850$ crores $\qquad x = ?$

After studying correlation between 2 variables process of estimating / predicting / Determining value of one variable on the basis of other is known as Regression analysis

54 Regression analysis is used to find

(a) Relation betn 2 variables

~~(b)~~ one variable on the basis of other

(c) Both of these

(d) None of these

55 presense of correlation between 2 variables is the PRE-REQUISITE for studying Regression analysis.

(56) Imp

Determining value of one variable on the basis of other is known as Regression analysis

x = Given
y = ?

Value of y is to be estimated on the basis of given value of x.

↓

Regression of y on x

↓

To Estimate y on the basis of given value of x, we should use Equation of Regression line of y on x.

↓

Eqn of Reg. line of y on x is,

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

where

b_{yx} = Reg. coeffi. of y on x

↓

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

y = Given
x = ?

Value of x is to be estimated on the basis of Value of y

↓

Regression of x on y

↓

To estimate x on the basis of given value of y, we should use Eqn of Reg. line of x on y.

↓

Eqn of Reg line of x on y is

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

where

b_{xy} = Reg. coefficient of x on y

↓

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

where r = corr. coefficient b/w x & y

(57) $r = -0.80$, $\sigma_y = 8$, $\sigma_x = 5$

Find b_{yx}, b_{xy}

$\Longrightarrow$ $$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} = -0.80 \times \frac{8}{5} = -1.28$$

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} = -0.80 \times \frac{5}{8} = -0.50$$

(58) If r is positive then b_{yx} & b_{xy} are also positive & vice versa

If r is Negative then b_{yx} & b_{xy} are also Negative & vice versa

In short b_{yx}, b_{xy}, r either

All these 3 will be positive

OR

All these 3 will be Negative

OR

All these are zero.

(59) $r = -0.8136$, $\sigma_x = +1.2381$, $\sigma_y = +2.8133$

Find b_{yx} & b_{xy}

$\Longrightarrow$ $$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} = -0.8136 \times \frac{2.8133}{1.2381}$$

$$= -1.848721$$

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} = -0.8136 \times \frac{1.2381}{2.8133}$$

$$= -0.358056$$

60) If $\bar{x} = 50$, $\bar{y} = 80$, $\sigma_x = 2$, $\sigma_y = 5$, $r = 0.90$

Find i) Reg. line of y on x

ii) Reg. line of x on y

iii) $x = 58$, $y = ?$

iv) $y = 81.50$, $x = ?$

$\Rightarrow$ i) Reg. line of y on x is

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

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

$$y - 80 = 0.90 \times \frac{5}{2} \times (x - 50)$$

$$y - 80 = 2.25(x - 50)$$

$$y - 80 = 2.25x - 112.50$$

$$\boxed{y = -32.50 + 2.25x}$$

ii) Reg. line of x on y is

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

$$x - 50 = r \cdot \frac{\sigma_x}{\sigma_y} \times (y - 80)$$

$$x - 50 = 0.90 \times \frac{2}{5} \times (y - 80)$$

$$x - 50 = 0.36(y - 80)$$

$$x - 50 = 0.36y - 28.80$$

$$\boxed{x = 21.20 + 0.36y}$$

iii) $x = 58$, $y = ?$

$$y = -32.50 + 2.25(58)$$

$$y = 98$$

iv) $y = 81.50$, $x = ?$

$$x = 21.20 + 0.36(81.50)$$

$$x = 50.54$$

(61) $\bar{x} = 280$, $\bar{y} = 350$, $\sigma_x = 20$, $\sigma_y = 60$

$r = 0.80$

If $x = 295.80$, $y = ?$

$\Longrightarrow$ $y - 350 = 0.80 \times \frac{60}{20} \times (295.80 - 280)$

$y = 387.92$

(62) $\bar{x} = 81.80$, $\bar{y} = 36.52$, $\sigma_x = 5.82$

$\sigma_y = 3.53$, $r = -0.75$

If $y = 50$, $x = ?$

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

$$\boxed{x - 81.80 = -0.75 \times \frac{5.82}{3.53} \times (50 - 36.52)}$$

$x = 65.1314$

(63) $\bar{x} = 135.22$, $\bar{y} = 1083.96$, $\sigma_x = 12.81$

$\sigma_y = 83.56$, $r = -0.8122$

If $x = 148.53$ $y = ?$

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

$y - 1083.96 = -0.8122 \times \frac{83.56}{12.81} \times (148.53 - 135.22)$

$\boxed{y = 1013.4436}$

(64) $b_{yx} = r \cdot \dfrac{\sigma_y}{\sigma_x}$, $b_{xy} = r \cdot \dfrac{\sigma_x}{\sigma_y}$

$b_{yx} \cdot b_{xy} = r \cdot \dfrac{\cancel{\sigma_y}}{\cancel{\sigma_x}} \times r \cdot \dfrac{\cancel{\sigma_x}}{\cancel{\sigma_y}}$

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

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

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

$\therefore$ 'r' is GM of b_{yx} & b_{xy}

$\boxed{\text{correlation coefficient is the Geometric mean of 2 regression coefficients}}$

(65) If $b_{yx} = 1.53$, $b_{xy} = 0.3136$

Find 'r'

$\Rightarrow \quad r^2 = b_{yx} \cdot b_{xy} = 1.53 \times 0.3136$
$= 0.479808$

$r = \sqrt{0.479808} = 0.6926817451$

(66) If $b_{yx} = -5.8188$, $b_{xy} = -0.125833$

Find 'r'

$\Rightarrow \quad r^2 = b_{yx} \cdot b_{xy} = -5.8188 \times -0.125833$

$r^2 = 0.7321970604$

$r = \sqrt{0.7321970604} = -0.8557$

(67) If $y = \dfrac{3x-19}{8}$ & $x = \dfrac{-18y+63}{5}$

then

(a) $r_{xy} = r_{uv}$ ~~(b)~~ $r_{xy} = -r_{uv}$

(c) Both (d) None

$y = \frac{3}{8}x - \frac{19}{8}$ & $x = -\frac{18}{5}y + \frac{63}{5}$

(68)

u	80	90
v	63	61

then Find r_{uv}

$\Longrightarrow$ $r_{uv} = -1.00$

(69) If $\bar{x} = 60$, $\bar{y} = 100$, $\sigma_x = 10$, $\sigma_y = 25$

$r_{xy} = -0.60$

Find ① Reg. line of y on x

② Reg. line of x on y

$\Longrightarrow$

Eqn of Reg line of y on x	Eqn of Reg line of x on y
$y - \bar{y} = b_{yx}(x - \bar{x})$	$x - \bar{x} = b_{xy}(y - \bar{y})$
$y - 100 = -0.60 \times \frac{25}{10}(x - 60)$	$x - 60 = -0.60 \times \frac{10}{25}(y - 100)$
$y - 100 = -1.50(x - 60)$	$x - 60 = -0.24(y - 100)$
$y - 100 = -1.50x + 90$	$x - 60 = -0.24y + 24$
$\boxed{y = 190 - 1.50x}$	$x = 84 - 0.24y$
$1.50x + y = 190$	$x + 0.24y = 84$
$\boxed{3x + 2y = 380}$	$100x + 24y = 8400$
	$25x + 6y = 2100$

(70) If $\bar{x} = 90$, $\bar{y} = 200$, $\sigma_x = 10$, $\sigma_y = 40$

$r = 0.95$

① If $x = 80$, $y = ?$

② If $y = 208$, $x = ?$

$\Rightarrow$ ① $y - 200 = 0.95 \times \frac{40}{10} \times (80 - 90)$

$y = 162$

② $x - 90 = 0.95 \times \frac{10}{40} \times (208 - 200)$

$x = 91.90$

(71) True / False

$r_{xy} = r_{yx}$	True
$b_{yx} = b_{xy}$	False
$r^2 = b_{yx} \cdot b_{xy}$	True
r is GM of b_{yx} & b_{xy}	true
If $r = 0$ then $b_{yx}, b_{xy} = 0$	True
If $\mathrm{cov}(x,y) > 0$ then $r > 0$	True
If $\mathrm{cov}(x,y) = 0$ then $r = 0$	True
$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$	True
$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$	True

72) If $\bar{x} = 50$, $\bar{y} = 90$, $\sigma_x = 3$, $\sigma_y = 6$ and $r = 0.75$

Find i) Reg line of y on x

ii) Reg line of x on y

iii) point of intersection of 2 reg. lines

$\Longrightarrow$ i) Reg. line of y on x is

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

$$y - 90 = 0.75 \times \frac{6}{3} \times (x - 50)$$

$$y - 90 = 1.50x - 75$$

$$\boxed{y = 15 + 1.50x} \quad \text{------ (1)}$$

ii) Reg. line of x on y is

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

$$x - 50 = 0.75 \times \frac{3}{6} \times (y - 90)$$

$$x - 50 = 0.375y - 33.75$$

$$x - 0.375y = 16.25$$

$$\boxed{x = 16.25 + 0.375y} \quad \text{------ (2)}$$

iii) To find point of intersection let's solve these 2 eqns simultaneously

$$y = 15 + 1.50(16.25 + 0.375y)$$

$$y = 15 + 24.375 + 0.5625y$$

$$0.4375y = 39.375$$

$$\therefore \boxed{y = 90}$$

Let's put $y = 90$ in eqn (2)

$$x = 16.25 + (0.375 \times 90) = 50$$

(50, 90) is the point of intersection of 2 reg. lines

$\boxed{\text{Two Regression lines will always intersect at point } (\bar{x}, \bar{y})}$

(73) If $\bar{x} = 85$, $\bar{y} = 20$, $\sigma_x = 10$, $\sigma_y = 2$, $r = -0.80$

Find 2 Reg. lines & point of intersection.

$\Longrightarrow$

① Reg line of y on x : $(y - \bar{y}) = r \cdot \frac{\sigma_y}{\sigma_x} \times (x - \bar{x})$

$y - 20 = -0.80 \times \frac{2}{10} \times (x - 85)$

$y - 20 = -0.16x + 13.60$

$\boxed{y = 33.60 - 0.16x}$ ---------- ①

② Reg. line of x on y : $(x - \bar{x}) = r \cdot \frac{\sigma_x}{\sigma_y} (y - \bar{y})$

$x - 85 = -0.80 \times \frac{10}{2} \times (y - 20)$

$x - 85 = -4y + 80$

$\boxed{x = 165 - 4y}$ ---------- ②

③ Let's solve these eqns simultaneously

$x = 165 - 4(33.60 - 0.16x)$

$x = 165 - 134.40 + 0.64x$

$0.36x = 30.60$

$\boxed{x = 85}$

Let's put $x = 85$ in Eqn ①

$y = 33.60 - 0.16(85) = 20$

$\boxed{y = 20}$

$\therefore$ $(85, 20) = (\bar{x}, \bar{y}) = (\text{AM of } x, \text{AM of } y)$ is the point of intersection of 2 regression lines

$$\boxed{\text{Point of intersection of 2 reg. lines is always } (\bar{x}, \bar{y})}$$

74 If $\bar{x} = 20$, $\bar{y} = 150$, $r = 0.70$, $\sigma_x = 5$, $\sigma_y = 20$. Find b_{yx}, Reg. line of y on x, b_{xy}, Reg. line of x on y.

$\Rightarrow$

(1) $b_{yx} = r \cdot \dfrac{\sigma_y}{\sigma_x} = 0.70 \times \dfrac{20}{5} = 2.80$

(2) Reg. line of y on x

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

$y - 150 = 2.80(x - 20)$

$y - 150 = 2.80x - 56$

$\boxed{y = 94 + 2.80x}$

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

(3) $b_{xy} = r \cdot \dfrac{\sigma_x}{\sigma_y}$

$= 0.70 \times \dfrac{5}{20}$

$= 0.175$

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

$x - 20 = 0.175(y - 150)$

$x - 20 = 0.175y - 26.25$ $\therefore$ $\boxed{x = -6.25 + 0.175y}$

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

75. If Reg line of y on x is $8x + 13y = 28$. Find b_{yx}

$\Rightarrow$ $8x + 13y = 28$

$13y = 28 - 8x$

$y = \frac{28}{13} - \frac{8}{13}x$ $\therefore$ $b_{yx} = -8/13$

$= -0.6154$

76. If Reg. line of x on y is $18x - 55y = 20$
Find b_{xy}

$\Rightarrow$ $18x - 55y = 20$

$18x = 20 + 55y$

$x = \frac{20}{18} + \frac{55}{18}y$

$\therefore$ $b_{xy} = 55/18 = 3.0555555$

77. If Reg. line of y on x is
$3x - 2y - 30 + 55x - 28y = 100x - 20y + 28$
Find b_{yx}

$\Rightarrow$ $3x - 2y - 30 + 55x - 28y = 100x - 20y + 28$

$-2y - 28y + 20y = 28 + 30 + 100x - 55x - 3x$

$-10y = 58 + 42x$

$y = \frac{58}{-10} + \frac{42}{-10}x$

comparing this with $y = a + bx$

$b = b_{yx} = -42/10$

$= -21/5$

$= -4.20$

(78) $b_{yx} = 20$, $b_{xy} = -0.003521$

Find 'r'

(a) 0.26537 (b) -0.26537

(c) 0.07042 ~~(d)~~ wrong data

As b_{yx} is positive & b_{xy} is negative, which is impossible. b_{yx}, b_{xy}, r must be of same sign.

(79) $b_{yx} = -1.80$, $b_{xy} = -0.38$ Find 'r'

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

$r^2 = -1.80 \times -0.38$

$r^2 = 0.684 \quad \therefore r = \pm\sqrt{0.684}$

$r = -0.827043$

As b_{yx} & b_{xy} are Negative, r must be Negative.

(80) $r = -0.8136$, $b_{xy} = -2.8056$, $b_{yx} = ?$

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

$(-0.8136)^2 = b_{yx} \times -2.8056$

$b_{yx} = -0.235937$

(81) $r = 0.228675$, $b_{yx} = 0.583175$, $b_{xy} = ?$

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

$(0.228675)^2 = 0.583175 \times b_{xy}$

$b_{xy} = 0.08966820529$

(82) $\left(\frac{b_{yx}+b_{xy}}{2}\right) \geq r$ ---- True / False

$\Rightarrow$ (AM of b_{yx} & b_{xy}) $\geq$ (GM of b_{yx} & b_{xy})

$\left(\frac{b_{yx}+b_{xy}}{2}\right) \geq r$ $\therefore$ Given statement is True

(83) If $b_{yx} = 1.53$, $b_{xy} = 0.80$ Find r

(a) 1.1063 (b) -1.1063 (c) 1.2240 (d) wrong data

As $-1 \leq r \leq 1.00$ $\therefore$ r^2 must lie between 0 & 1.00 including both limits.

(84) $-1.00 \leq r \leq 1.00$

$0 \leq r^2 \leq 1.00$

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

Product of 2 regression coefficients must lie b/w 0 & 1.00 including both limits

Minimum value of $r = -1.00$

Maximum value of $r = 1.00$

Minimum value of $r^2 = 0.00$

Maximum value of $r^2 = 1.00$

(85) If Reg. line of x on y is

$33x - 21y - 28 = 15x - 101y + 58$. Find b_{xy}

$\Rightarrow 33x - 15x = 58 + 28 - 101y + 21y$

$18x = 86 - 80y$

$x = \frac{86}{18} + \left(\frac{-80}{18}\right) y$

$b_{xy} = -80/18 = -40/9 = -4.44444$

(86) If $b_{yx} > 1$ then b_{xy} must be less than 1. True / False

$\Rightarrow \because (b_{yx} \times b_{xy}) \leq 1.00$

Given statement is : True

(87) If $b_{xy} \leq 1$ then b_{yx} must be greater than 1. True / False

$\Rightarrow$ Given statement is False

(88) If Reg. line of y on x is $y = 88 + 0.20x$
& Reg. line of x on y is $x = 100 + 1.20y$
Find $(\bar{x}, \bar{y})$

$\Rightarrow$ Let's solve these 2 Eqns of regression lines simultaneously

$y = 88 + 0.20(100 + 1.20y)$

$y = 88 + 20 + 0.24y$

$y - 0.24y = 108$

$y(1 - 0.24) = 108$

$0.76y = 108$

$\boxed{y = 142.1053}$

$\therefore x = 100 + 1.20(142.1053)$

$\boxed{x = 270.5263}$

$\therefore (270.5263, 142.1053) = (\bar{x}, \bar{y})$

$\therefore \bar{x} = 270.5263$

$\bar{y} = 142.1053$

Two Regression lines will always intersect at point $(\bar{x}, \bar{y})$

b_{yx} = Reg. coeffi. of y on x

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

$= \dfrac{cov(x,y)}{\cancel{\sigma_x} \cdot \cancel{\sigma_y}} \times \dfrac{\cancel{\sigma_y}}{\sigma_x}$

$= \left[\dfrac{cov(x,y)}{\text{variance of } x}\right]$

$b_{yx} = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})/n}{\Sigma(x-\bar{x})^2/n}$

$b_{yx} = \left[\dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}\right]$

b_{xy} = Reg. coeffi. of x on y

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

$= \dfrac{cov(x,y)}{\cancel{\sigma_x} \cdot \sigma_y} \times \dfrac{\cancel{\sigma_x}}{\sigma_y}$

$= \left[\dfrac{cov(x,y)}{\text{variance of } y}\right]$

$b_{xy} = \left[\dfrac{\Sigma(x-\bar{x})(y-\bar{y})/n}{\Sigma(y-\bar{y})^2/n}\right]$

$b_{xy} = \left[\dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}\right]$

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} = \frac{cov(x,y)}{\text{variance of } x} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y} = \frac{cov(x,y)}{\text{variance of } y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}$$

(90) If $r = 0$ then $cov(x, y) = ?$

(a) 0 (b) 1 (c) Both 0 & 1 (d) can't say

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

$$\boxed{\text{If } cov(x,y) = 0 \text{ then } r = 0}$$

(91) If covariance of $(x, y) = 100$

variance of $x = 8000$

variance of $y = 500$

Find 'r'

$$\Rightarrow \quad r = \left(\frac{\text{covariance of } (x,y)}{\sigma_x \cdot \sigma_y}\right) = \frac{100}{\sqrt{8000} \times \sqrt{500}}$$

$$= \frac{100}{\sqrt{4000000}} = \frac{100}{2000} = 0.05$$

(92) $\sum (x - \bar{x})(y - \bar{y}) = k$

$\sum (x - \bar{x})^2 = m$

$\sum (y - \bar{y})^2 = j$

Find 'r'

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

$$r = \frac{k}{\sqrt{m} \times \sqrt{j}} = \frac{k}{\sqrt{mj}}$$

Karl pearson's coefficient

$$= \frac{\sum X \cdot Y}{\sqrt{\sum X^2 \cdot \sum Y^2}}$$

where

$X = (x - \bar{x})$

$Y = (y - \bar{y})$

(93)

variance of data can be ______	SD of data can be ______	cov(x, y) can be ______
$\Rightarrow$ (a) OR (c)	$\Rightarrow$ (a) OR (c)	$\Rightarrow$ (a) OR (b) OR (c)

(a) zero
(b) negative
(c) positive

(94) If $cov(x, y) = 50$, $\sigma_x = 10$ then

~~(a) $\sigma_y \geq 5$~~
(b) $\sigma_y \leq 5$
(c) $\sigma_y < 5$
(d) can't say

$r = \left(\frac{cov(x,y)}{\sigma_x \cdot \sigma_y} \right) \leq 1$

$r = \left(\frac{50}{10 \times \sigma_y} \right) \leq 1$

$10 \times \sigma_y \geq 50$

$\sigma_y \geq 5$

(95) If scatter diagram is showing a line $\parallel$ to y-axis then there is ______ correlation

(a) positive
(b) negative
(c) spurious
~~(d) No~~

(96) If scatter diagram is showing points without depicting any pattern then there is $\boxed{\text{NO CORRELATION}}$

(97) process of establishing relation / association betn 2 or more variable is known as correlation analysis and process of ascertaining value of one variable on the basis of other is known as Regression analysis

(98) x = Rainfall in city $\quad r_{xy} = 0.60$

y = No. of trees in city

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

coeffi. of Non-determination (unexplained variance) $= 1 - r^2 = 1 - 0.36$
$= 0.64 = 64\%$

(99)

i) Coefficient of Determination $= \left(\dfrac{\text{Explained variance}}{\text{Total variance}}\right) = r^2$

ii) coefficient of Non-determination $= \left(\dfrac{\text{unexplained variance}}{\text{Total variance}}\right) = (1 - r^2)$

$=$

(coeffi. of Determination + coeffi. of Non-determination) $= 1.00$
$= 100\%$

(100) If $r = 0.85$ Find

coeffi. of Determination $= r^2 = 0.85^2 = 0.7225$

coeffi. of Non-determination $= 1 - r^2 = 1 - 0.7225$

$= 0.2775$

∴ Explained variance $= 72.25\%$

unexplained variance $= 27.75\%$

(101) i) 2 Regression lines will coincide when

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

ii) Two Regression lines will coincide when there is perfect positive correlation or perfect Negative correlation

iii) If $r = 0$ then Two regression lines are generally $\perp$ to each other

iv) When 2 regression lines are $\perp$ to each other then there is <u>No correlation.</u>

(102) If $c = 5$, $m = 11$, find coeffi. of concurrent deviation.

$\Rightarrow$ $r = \pm\sqrt{\pm\dfrac{2c - m}{m}}$

$r = -\sqrt{-\left(\dfrac{2(5) - 11}{11}\right)}$

$r = -\sqrt{-\left(-\dfrac{1}{11}\right)}$

∴ $r = -0.3015$

(103) If $c = 8$, $m = 12$. Find coefficient of concurrent deviation.

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

$$= \pm\sqrt{\pm\frac{2(8)-12}{12}}$$

$$= +\sqrt{\frac{4}{12}} = +\sqrt{0.33333333}$$

$$= +0.57735$$

(104)

	Maths (x)	Physics (y)
AM	85	92
SD	8	11
$r = 0.89$		

What expected score of maths if a student scored 90 marks in physics?

$\Rightarrow \quad y = 90, \; x = ?$

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

$$x - 85 = 0.89 \times \frac{8}{11} \times (90 - 92)$$

$$\boxed{x = 83.71}$$

Expected score of maths for a student who scored 90 in physics $= 83.71$

1050 marks of maths, stat of 25 students

(5,16), (18,19), (6,8), (7,13), (2,18)
(12,13), (5,7), (12,8), (19,2), (16,5)
(13,18), (12,15), (6,8), (9,5), (6,12)
(13,15), (7,12), (13,6), (12,12), (13,3)
(8,5), (9,6), (12,5), (3,16), (5,8)

prepare Bi-variate Freq. Distribution Table

→ Bi-variate Freq. Distribution Table

Marks of Maths (x) \ Marks of stat (y)	0 - 5	5 - 10	10 - 15	15 - 20	Total
0 - 5	= 0	| = 1	| = 1	|| = 2	4
5 - 10	= 0	~~||||~~ || = 7	||| = 3	| = 1	11
10 - 15	| = 1	|| = 2	| = 1	||| = 3	7
15 - 20	| = 1	| = 1	= 0	| = 1	3
Total	2	11	5	7	25

i) find Marginal Distribution of x - marks of maths

x	0 - 5	5 - 10	10 - 15	15 - 20
f	4	11	7	3

ii) Find Marginal Distribution of y - Marks of stats

y	0 - 5	5 - 10	10 - 15	15 - 20
f	2	11	5	7

iii) find conditional distri. of y when x is 5 - 10

y	0 - 5	5 - 10	10 - 15	15 - 20
f	0	7	3	1

iv) Find conditional Distri. of x when y is 15 − 20

x	0 − 5	5 − 10	10 − 15	15 − 20
f	2	1	3	1

100

x \ y	10 − 30	30 − 90	90 − 110	Total
10 − 20	53	28	98	179
20 − 30	113	167	283	563
30 − 40	669	813	822	2304
40 − 50	1083	786	555	2424
Total	1918	1794	1758	5470

⟹ i) find Marginal Distribution of x

x	10 − 20	20 − 30	30 − 40	40 − 50
f	179	563	2304	2424

ii) Find Marginal Distribution of y

y	10 − 30	30 − 90	90 − 110
f	1918	1794	1758

iii) Find conditional Distri of x when y is 90 − 110

x	10 − 20	20 − 30	30 − 40	40 − 50
f	98	283	822	555

iv) find conditional Distri. of y when x is 30 − 40

y	10 − 30	30 − 90	90 − 110
f	669	813	822

107 $r = 0.50$, $\Sigma d^2 = 82.50$, $n = ?$

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

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

$$\frac{495}{n(n^2-1)} = 1 - 0.50 = 0.50$$

$$\frac{495}{0.50} = n(n^2-1)$$

$$\therefore n(n^2-1) = 990 = 10 \times 99 = 10 \times (100-1)$$

$$n(n^2-1) = 10(10^2-1) \quad \therefore n = 10$$

---

108 $\boxed{\text{Regression coefficients are affected by change in scale but not affected by change of origin.}}$

109 Theory of Regression has been desired from $\boxed{\text{Method of least squares}}$

110 There are some cases when we find correlation beth 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 type of correlation is known as $\boxed{\text{Spurious or Non-sense correlation}}$

(111) Regression coefficients are unaffected due to ________

~~(a)~~ shift of origin
(b) change in scale
(c) Both
(d) None

(112) lungs damage & cigarette smoking, ________ correlation may exist

~~(a)~~ positive (b) negative (c) zero (d) None

(113) bye and bey are always same. True/False

⟹ False

(114) Yxy and Yyx are always same. True/False

⟹ True

(115) co-variance measures ________ variation between 2 variables.

~~(a)~~ Joint (b) common (c) Relative (d) None

(116) (1) If $u = 3 + 2x$

$v = -8 + 5y$ then

(a) ~~$r_{xy} = r_{uv}$~~ (b) $r_{xy} = -r_{uv}$

(c) $r_{xy} = \frac{1}{r_{uv}}$ (d) $\frac{r_{xy}}{r_{uv}} = 2$

(2) $u = 18 - 3x$

$v = -63 + 55y$

then

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

(117) If $u = 13x + 9$

$v = -18y + 33$ then

Find relation betn $(b_{uv}$ & $b_{xy})$ and $(b_{vu}$ & $b_{yx})$

and $(r_{uv}$ & $r_{xy})$

$\Rightarrow$

(1) $r_{uv} = -r_{xy}$

(2) $b_{uv} = r_{uv} \times \frac{\sigma u}{\sigma v} = -r_{xy} \times \frac{13 \times \sigma x}{18 \times \sigma y}$

$= -\frac{13}{18} \times \left(r_{xy} \cdot \frac{\sigma x}{\sigma y}\right) = -\frac{13}{18} b_{xy}$

(3) $b_{vu} = \frac{-18}{13} \times b_{yx}$

(118) If $u = 15x + 93$

$v = -18 + 61y$

then

$r_{uv} = r_{xy}$

$b_{uv} = b_{xy} \times \frac{15}{61}$

$b_{vu} = b_{yx} \times \frac{61}{15}$

(119) If $u = 59 - 68x$

& $v = -51y + 33$

then

$r_{uv} = r_{xy}$

$b_{uv} = b_{xy} \times \frac{-68}{-51}$

$= \frac{4}{3} \times b_{xy}$

$b_{vu} = \frac{3}{4} \times b_{yx}$

(120) $u = \dfrac{13+5x}{19}$ $\quad v = \dfrac{-16+22y}{101}$

then $u = \dfrac{13}{19} + \dfrac{5}{19}x$, $v = -\dfrac{16}{101} + \dfrac{22}{101}y$

$$b_{uv} = b_{xy} \times \left[\frac{\frac{5}{19}}{\frac{22}{101}}\right] = \frac{505}{418}\, b_{xy}$$

$$b_{vu} = \frac{418}{505}\, b_{yx}$$

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

(121) Spearman's rank correlation coefficient = 0.80 for 10 pairs of obsns. Later on it is observed that one value of 'd' was taken as 7 instead of 6. Find correct 'r'

$\Longrightarrow$

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

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

$$\frac{6\Sigma d^2}{990} = 1 - 0.80$$

wrong $\Sigma d^2 = 33$

correct Σd^2

$= 33 - 7^2 + 6^2$

$= 33 - 49 + 36 = 20$

$\therefore$ correct 'r'

$$= 1 - \frac{6 \times 20}{990}$$

$$= 1 - \frac{120}{990}$$

$= 0.87878787$

(122)

x	y
↑	↑
↓	↓

Type of correlation : Positive

x	y
↓	↑
↑	↓

Type of correlation : Negative

(123) $\left(\text{product of 2 regression coefficients}\right) = \left(\text{correlation coefficient}\right)^2$

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

(124) n = no. of pairs of obs's in given data (sample)

N = population size

As we draw conclusions on the basis of sample & not on the basis of population there can be some errors in our judgement

Probable error

$= 0.6745 \times S.E.$

$= 0.6745 \times \left[\dfrac{1-r^2}{\sqrt{N}}\right]$

Standard error $= \left(\dfrac{1-r^2}{\sqrt{N}}\right)$

125

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

ii) conditional Distribution is the frequency distribution of one variable across the particular sub-population of other variable.

126 Bi-variate data is collected for :

(a) 2 variables

(b) 3 or more variables

(c) ~~2 variables at same point of time~~

(d) 2 variables at diff point of time

127 The diff betn actual value and estimated value is known as Error OR Residue

128 If $r = 0.40$ then

% of known or accounted variation $= r^2 = 0.40^2 = 0.16 = 16\%$

% of unknown or un-accounted variation $= 84\%$

coeffi. of determination $= r^2$

coeffi. of Non-determination $= 1 - (r^2)$

variables	nature of correlation
① No. of claims & profit of insurance company	Negative
② Demand for Giffen goods & price of giffen goods	positive
③ sale of woolen garments & Temp	Negative
④ marketing exp & Turnover	positive
⑤ No. of trees & Rainfall	positive
⑥ cigarette smoking & Lungs damage	positive
⑦ Rainfall & crop yield	positive
⑧ years of education & Income	positive
⑨ Temp & sale of Tea, coffee	Negative
⑩ No. of hrs on social media, marks in exam	Negative

130) For the Bi-variate data

$(20,5), (21,4), (22,3)$ $\quad r = ?$

(a) 1.00 $\quad$ (b) -1.00 $\quad$ (c) can't say $\quad$ (d) $r = 2.00$

x	y
20	5
21	4
22	3

131) Simple correlation is known as $\boxed{\text{Linear correlation}}$

132) Slope of Reg. line of y on x is :

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

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

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

$$(b_{yx})x - 1y = b_{yx} \bar{x} - \bar{y}$$

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

133) Slope of Reg. line of x on y is :

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

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

$$1x - b_{xy} \cdot y = -b_{xy} \cdot \bar{y} + \bar{x}$$

slope of the line $= \dfrac{-1}{-b_{xy}} = \dfrac{1}{b_{xy}}$

(134) The Best method to measure correlation is ____________

$$\boxed{\text{Karl pearson's product moment correlation coefficient}}$$

(135) 2 regression lines

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

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

intersect at point $(\bar{x}, \bar{y}) = (\text{AM of } x, \text{AM of } y)$

(136) 2 Reg. lines are $2x-3y=10$ & $5x-8y=33$. Find b_{xy} & b_{yx} & r

$\Longrightarrow$

x on y	y on x
$2x-3y=10$	$5x-8y=33$
$2x = 10+3y$	$-8y = 33-5x$
$x = \frac{10}{2}+\frac{3}{2}y$	$y = -\frac{33}{8}+\frac{5}{8}x$
$b_{xy} = \frac{3}{2} = 1.50$	$b_{yx} = \frac{5}{8} = 0.625$

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

$= \sqrt{0.625 \times 1.50} = \sqrt{0.9375}$

$= 0.96825$