Correlation & Regression

-6marks

L1

Chapter overview:-

Correlation:-

1.) Defination
2.) Method of computing correlation
3.) Common Properties
4.) Some important points.

Meaning:- it is defines as the measures of Extent Relation b/w 2 variables.

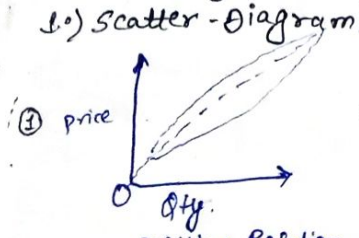
* it is Represented by r.
* it is also called coefficient of correlation.

Important point Related to r:-

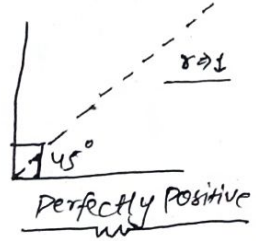
1. it is a unit free measures.
2. $-1 \leq r \leq 1$ - i.e - (-1 to 1)

Method of computing correlation (r):-

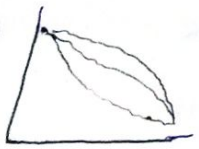
1.) Scatter-Diagram:-



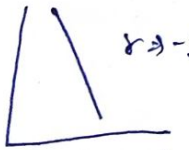
positive Relation.



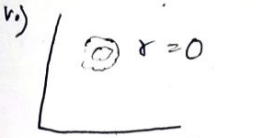
Perfectly Positive



Negative



Perfectly Negative



Zero-Relation

or

Zero Relation

2.) Spearman Rank correlation:- This Method is suitable, when we want to compare the two Quality character.

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

Eg.-

X	y	R_x	R_y	$d \Rightarrow (R_x - R_y)$	d^2
95	100	2	1	1	1
100	80	1	4	-3	9
80	90	4	2	2	4
85	70	3	5	-2	4
60	85	5	3	2	4
					22

$r_R \Rightarrow 1 - \frac{6\Sigma d^2}{n(n^2-1)}$

$\Rightarrow 1 - \frac{6 \times 22}{5(5^2-1)}$

$\Rightarrow 1 - \frac{132}{120}$

$\Rightarrow 1 - 1.1$

$\Rightarrow -0.1$ Ans

Eg.-17*3

X	y	R_x	R_y	$d \Rightarrow (R_x - R_y)$	d^2
90	7	2	2	0	0
85	6	3	3	0	0
68	2	6	7	1	1
75	3	6	6	0	0
82	4	4	5	-1	1
80	5	5	4	1	1
95	8	1	1	0	0
70	1	7	8	-1	1
					4

$r_R \Rightarrow 1 - \frac{6\Sigma 4}{8(8^2-1)}$

$\Rightarrow 1 - \frac{24}{504} \Rightarrow 1 - 0.0476$

$\Rightarrow 0.9524$

Ans

Set-B

Q10

Sol → N = 8, $\Sigma d^2 = 21$

$r_R = 1 - \frac{6\Sigma d^2}{n(n^2-1)} \Rightarrow 1 - \frac{6\times 21}{8(8^2-1)} \Rightarrow 1 - \frac{126}{504}$

$\Rightarrow 1 - 0.25 \Rightarrow 0.75$ — (D) Ans

Q11) $r = 0.6$, $\Sigma d^2 = 66$, $n = ?$ — use option Method

$r_R = 1 - \frac{6\Sigma d^2}{n(n^2-1)} \Rightarrow 1 - \frac{6\times 66}{10(99)}$

$\Rightarrow 1 - \frac{396}{990} \Rightarrow 0.4 \Rightarrow 1 - 0.4 \Rightarrow 0.6$ Ans — (a)

10 ↑ Ans

Q12) N = 6, r = 0.4 (wrong → Right)

Trick to Sol → $\frac{6[N^2 - R^2]}{n(n^2-1)} \Rightarrow \frac{6[3^2 - 4^2]}{6(6^2-1)} \Rightarrow \frac{6[9-16]}{6\times 35} \Rightarrow \frac{9-16}{35} \Rightarrow \frac{-7}{35} \Rightarrow -0.2$

Now at this in old r, i.e $\Rightarrow 0.4 + (-0.2)$

$\Rightarrow 0.4 - 0.2 \Rightarrow 0.2$ — (b) Ans

Eg - 17.14 - N = 10, r = 0.80

Sol - $6\left[\frac{7^2 - 5^2}{10(99)}\right] \Rightarrow \frac{6\times 24}{990} \Rightarrow \frac{144}{990} \Rightarrow 0.14 + r$

$\Rightarrow 0.14 + 0.80 \Rightarrow 0.94$ Ans

Eg. —

X	Y	R_x	R_y	$d \Rightarrow (R_x - R_y)$	d^2
80	90	1	1	0	0
56	75	4	2.5	1.5	2.25
50	75	5.5	2.5	3	9
48	65	7	5	2	4
50	65	5.5	5	0.5	0.25
62	50	2	7	-5	25
60	65	3	5	-2	4
					44.5

$r_R \Rightarrow 1 - \frac{6\Sigma d^2 + \frac{\Sigma t^3 - t}{12}}{n(n^2-1)}$

$1 - 6\left(\frac{44.50 + 3}{7(7^2-1)}\right)$

$\Rightarrow 1 - 6\left(\frac{47.50}{7\times(49-1)}\right)$

$\Rightarrow 1 - \frac{285}{336} \Rightarrow 1 - 0.84$

$\Rightarrow 0.16$ Ans

★ if no. are repeating then see its Rank & add all the ranks of all these Repeating no. & write it. divide it Equally how many times these are Repeating

$\left[\frac{\Sigma t^3 - t}{12}\right] \Rightarrow \frac{(2^3-2) + (2^3-2) + (3^3-3)}{12} \Rightarrow \frac{6+6+24}{12}$

$\Rightarrow \frac{36}{12} \Rightarrow 3$ Ans

{ How many times No. Repeated in R_x }

Concurrent Deviation Method :-

This Method is Suitable when we are not Serious, the magnitude of the variable.

formula :- $r_{CD} \Rightarrow \pm\sqrt{\pm\frac{2C-M}{M}}$

C ⇒ + sign of deviation

M ⇒ N - 1

Ex-17.15

Year	Price	Sign deviation (a)	Demand	Sign devi. (b)	Product of deviation (ab)
1990	25	Nil	35	Nil	Nil
1991	28	+	34	−	−
1992	30	+	35	+	+
1993	23	−	30	−	+
1994	35	+	29	−	−
1995	38	+	28	−	−
1996	39	+	26	−	−
1997	42	+	23	−	−
				C ⇒ 2	

C ⇒ Sum of all Plus Signs

$r_{CO} \Rightarrow \pm\sqrt{\pm\frac{2C-M}{M}}$ ⇒ C ⇒ 2, M ⇒ N−1 ⇒ 8−1

Eg

$\Rightarrow \pm\sqrt{\pm\left(\frac{2\times2-7}{7}\right)} \Rightarrow \pm\sqrt{\pm\left(\frac{-3}{7}\right)} \Rightarrow \sqrt{\frac{-3}{7}} \Rightarrow -0.65$ Ans

Q13-Set-B) N ⇒ 10, C ⇒ 4

$r_{CO} \Rightarrow \pm\sqrt{\pm\frac{2C-M}{M}} \Rightarrow \pm\sqrt{\frac{+2\times4-9}{9}} \Rightarrow -\frac{1}{3}$ Ans (d)

Q14.) n ⇒ P, $r_{CO} \Rightarrow \frac{1}{\sqrt{3}}$, C ⇒ 6, P ⇒ ?

$r_{CO} \Rightarrow \pm\sqrt{\pm\frac{2C-M}{M}}$

$\frac{1}{\sqrt{3}} \Rightarrow \pm\sqrt{\frac{2\times6-(P-1)}{(P-1)}}$

M ⇒ n−1 ⇒ P−1

$\frac{1}{3} \Rightarrow \frac{12-P+1}{P-1} \Rightarrow \frac{1}{3} \times \frac{13-P}{P-1} \Rightarrow 39-3P \Rightarrow P-1$

$40 \Rightarrow 4P$

P ⇒ 10

Karl Pearson correlation coefficient :-

* It is the best Method.
* it is suitable only when Related variable are linear.

X ⇒ y ⌣ | n ⇒ y^2 X

n ⇒ 2y ⌣ | n ⇒ $\sqrt{y}$ X

Formula :-

i) If covariance (x,y) given :-

$r \Rightarrow \frac{cov(x,y)}{\sigma x \times \sigma y}$

σ ⇒ S.D ⇒ Stand deviat

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

Set-B

Q1.) cov(x,y) ⇒ 40, $(\sigma x)^2$ ⇒ 16, $(\sigma y)^2$ ⇒ 256

⇒ σx ⇒ 4, σy ⇒ 16

$r \Rightarrow \frac{cov(x,y)}{\sigma x \times \sigma y} \Rightarrow \frac{40}{4\times16} \Rightarrow \frac{10}{16} \Rightarrow 0.625$ Ans

Eg - if cov(x,y) ⇒ 60, variance of x ⇒ 400, S.D of y ⇒ 50, r ⇒ ?

$r \Rightarrow \frac{cov(x,y)}{\sigma x \times \sigma y} \Rightarrow r \Rightarrow \frac{60}{20\times50} \Rightarrow r \Rightarrow 0.06$

Coefficient of determination $\Rightarrow r^2$

Coefficient of Non-determination $\Rightarrow 1 - r^2$

if cov.(x,y) is not given:-

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

Ex-

x	y	xy	x^2	y^2
10	5	50	100	25
9	6	54	81	36
8	7	56	64	49
7	8	56	49	64
6	9	54	36	81
5	10	50	25	100
45	45	320	355	355

Now, $r \Rightarrow \frac{N\Sigma xy - \Sigma x \times \Sigma y}{\sqrt{N\Sigma x^2 - (\Sigma x)^2} \times \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$

$r \Rightarrow \frac{6 \times 320 - 45 \times 45}{\sqrt{6 \times 355 - (45)^2} \times \sqrt{(6 \times 355) - (45)^2}}$

$r \Rightarrow \frac{1920 - 2025}{\sqrt{2130 - 2025} \times \sqrt{2130 - 2025}}$

$r \Rightarrow \frac{-105}{\sqrt{105} \times \sqrt{105}}$

$r \Rightarrow \frac{-105}{105} \Rightarrow (-1)$ Ans

Eg

x	y
100	700
200	300
400	900
300	1000
0	1100
600	0

★ if any single zero contained in any variable x or y, $r \Rightarrow 0$ that's answer we don't solve further.

Q15 - (c) Ans → (D)

★ The coefficient correlation Remains Invariant under a change or origin &/or scale of the variable under consideration depending on the sign of scale factor.

Ex-17.8 :-

i) $2y + 3x + 4 \Rightarrow 0$ & $4v + 16y + 11 \Rightarrow 0$

$2y + 3x + 4 \Rightarrow 0$

★ $2y \Rightarrow -3x - 4$

$y \Rightarrow \frac{-3x}{2} - \frac{4}{2}$

★ $4v + 16y + 11 \Rightarrow 0$

$4v \Rightarrow -16y - 11$

$v \Rightarrow \frac{-16y}{4} - \frac{11}{4}$

So coefficient will same $\Rightarrow 0.8$ because both variable signs are same.

ii) $2y - 3x + 4 \Rightarrow 0$ & $4v + 16y + 11 = 0$

$2y \Rightarrow 3x - 4$

$y \Rightarrow \frac{3x}{2} - \frac{4}{2}$

$4v \Rightarrow -16 - y - 11$

$v \Rightarrow \frac{-16}{4}y - \frac{11}{4}$

So, in this qtn sign in both variables are different so the answer will be negative -0.8.

iii) $2y - 3x + 4 \Rightarrow 0$ & $4v - 16y + 11 \Rightarrow 0$.

$2y \Rightarrow 3x + 4$

$y \Rightarrow \frac{3x}{2} - \frac{4}{2}$

$4v \Rightarrow 16y - 11$

$v \Rightarrow \frac{16y}{4} - \frac{11}{4}$

0.8 Positive Ans

iv) $2y + 3x + 4 \Rightarrow 0$ & $4v - 16y + 11 \Rightarrow 0$

$2y \Rightarrow -3x - 4$

$y \Rightarrow \frac{-3x}{2} - \frac{4}{2}$

★ $4v \Rightarrow 16y - 11$

$v \Rightarrow \frac{16y}{4} - \frac{11}{4}$

-0.8 Negative Ans

Q6.) $4+5x \Rightarrow 6$ & $3y-7x \Rightarrow 20$. — Coeff of x & y ? $\Rightarrow -0.58$

V.V.I

$4 \Rightarrow \ominus 5x - 6$. $3y \Rightarrow 7x - 20$

$y \Rightarrow \oplus \frac{7x}{3} - \frac{20}{3}$

-0.58 Ans

(b) Negative

★ if only one Eqn. is given & 2nd Eqn missing, then we consider that 2nd Eqn sign is positive.

Eg-Q7 → $3x+4y+7=0$ — Coeff of x & y ? $\Rightarrow -0.6$

$3x \Rightarrow -4y - 7$

$x \Rightarrow \frac{-4y}{3} - \frac{7}{3}$ & 2nd Eqn $\Rightarrow (+) \Rightarrow -0.6$ — (a) Ans

$4y \Rightarrow -3x - 7$

$y \Rightarrow \frac{\ominus 3x}{4} - \frac{7}{4}$ & 2nd Eqn positive $\Rightarrow -$ & -0.6

$\Rightarrow +0.6$ Ans (2)

Q8.) $-3 \quad -2 \quad 0 \quad -1 \quad 2$

$-4 \quad -2 \quad -1 \quad 0 \quad 2$

Add sign-only → $-, -, -, -, \oplus$

$+ \quad - \quad +$

Multiply signs.

last sign if positive then answer Remain same Otherwise diff.

Q9.) $10 \quad 15 \quad 25 \quad 20 \quad 35$

$-24, -36, -42, -48, -60$

$-, -, -, -, -$

$(+) \quad (-) \quad (+) \quad (-)$ ← so negative sign

ans will $-(-0.93)$

(a) Ans

$4x + 3y \Rightarrow 10$, find $\Rightarrow r$?

↳ convert like this $y \Rightarrow a + bx$

$3y \Rightarrow 10 - 4x$

$y \Rightarrow \frac{10}{3} \left(-\frac{4x}{3}\right) \leftarrow b \Rightarrow -\frac{4}{3}$,

So ans will $b < 0, r \Rightarrow -1$ Ans

★★★

Ans V.V.I

if $b > 0$, $r \Rightarrow 1$

if $b < 0$, $r \Rightarrow -1$

Eg— $2x + 4y \Rightarrow 20$, find $r \Rightarrow$?

$4y \Rightarrow 20 - 2x$

$y \Rightarrow \frac{20}{4} \left(-\frac{2x}{4}\right)$ → $y \Rightarrow a + bx$

$b \Rightarrow -\frac{2}{4} < 0$, $r \Rightarrow -1$ Ans

Q4 $y \Rightarrow a + bx$, Coeff. of x & y $\Rightarrow$? (or $r \Rightarrow$?)

Sol- (c) -1 or $r = -1$ according as $b > 0$ or $b < 0$. Ans

— Correlation Ends —

Eg-17.6

$r \Rightarrow 0.4$, $n \Rightarrow 20$

→ AM of $y \Rightarrow 12$

→ Am of $y \Rightarrow 12$

SD of $x \Rightarrow 3$

SD of $y \Rightarrow 4$

$r \Rightarrow \frac{COV(x,y)}{\sigma x \times \sigma y}$

$0.4 \Rightarrow \frac{COV(x,y)}{3 \times 4}$

$COV(x,y) \Rightarrow 4.8$

$COV(x,y) \Rightarrow \frac{\Sigma xy}{n} - \bar{x}\bar{y}$

$4.8 \Rightarrow \frac{\Sigma xy}{20} - 12 \times 15$

$4.8 + 180 \Rightarrow \frac{\Sigma xy}{20}$

$184.8 \Rightarrow \frac{\Sigma xy}{20}$

$\Sigma xy \Rightarrow 3696$

L-4

Regression - Analysis

Topics to be covered :-

1.) Defination
2.) Regression coefficient (b_{yx} & b_{xy})
3.) Properties.
4.) Questions...

Types of Regression Equation :-

i) y depend on x :- [y on x]
ii) x depend on y :- [x on y]

How to calculate Regression Equation (Method) :-

(I) * General form (Normal Equation)

[y on x] ⇒ [$y = a + bx$]

[x on y] ⇒ [$x = a + by$]

Q16 - $5a + 10b = 40$ ×2
$10a + 25b = 95$

$\Rightarrow 10a + 20b = 80$
$\Rightarrow 10a + 25b = 95$
$-5b = -15$

[$b = 3$]

Put (b) in eqn (I)

$5a + 10(3) = 40$
$\Rightarrow 5a + 30 = 40$
$\Rightarrow 5a = 10$

[$a = 2$]

$y = a + bx$
[$y = 2 + 3x$] — (c) Ans

How to calculate Mean:- $\bar{x}, \bar{y}$

Q17 $2x+5y = -1 \times 2$

$5x+6y = -1$

$\Rightarrow 4x+6y = -2$

$5x+6y = -1$

$-x = -1$

$x = 1$

$2(1)+3y = -1$

$3y = -3$

$y = -1$

$(1, -1)$ Ans (a)

Q24 $\rightarrow y = 3x+4$

$y = 3(-1)+4$

$y = -3+4$

$y = 1$ (a) Ans

Point Method:-

- y on $x \Rightarrow y - \bar{y} = b_{yx}(x - \bar{x})$
- x on $y \Rightarrow x - \bar{x} = b_{xy}(y - \bar{y})$
- $\bar{x}, \bar{y} \Rightarrow$ Mean of x & y
- b_{yx} & $b_{xy} \Rightarrow$ Coefficient of Regression.
- $r = \sqrt{b_{yx} + b_{xy}} \Rightarrow$ G.M

How to identify which Eqn. y on x & x on y:-

$4x+3y = 10$ — x on y

$2x+2y = 20$ — y on x

(8) 6

If 1st Value big.

$2x+3y = 10$ — y on x

$3x+2y = 20$ — x on y

4 (9)

- $4x + 3y = 20$ — x on y
- $3x + 4y = 20$ — y on x

(16) 9

always ignore sign Just consider big digit only.

- $5x+4y = 10$
- $2x-3y = 20$

(15) 8

- if Reverse sign came then no any x on y or y on x will come.

Q18 - $3x+y = 13$ — x on y

$2x+5y = 20$ — y on x — 2nd Eqn (b) Ans

15 2

Q19.) $2x-3y = 10$

$3x+4y = 15$

$8 - 9 \Rightarrow$ None of these (d) Ans

How to calculate byx & buy :-

Eg- $4x+4y = 20$ — y on x

$4y = 20-4x$

$y = \frac{20}{4} - \frac{4x}{4}$

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

$3x+4y = 20$ — x on y

$3x = 20-4y$

$x = \frac{20}{3} - \frac{4}{3}y$

$buy = -\frac{4}{3}$

Q21 ☆ $4y-5x = 15$ — y on x

$r = 0.75$ ⇒ bxy ⇒ ?

$4y = 15+5x$

$y = \frac{15}{4} + \frac{5x}{4}$

$byx = \frac{5}{4}$

$r = \sqrt{bxy \times byx}$

$0.75 = \sqrt{bxy \times \frac{5}{4}}$

$0.5625 = bxy \times \frac{5}{4}$

$\frac{0.5625 \times 4}{5}$ ⇒ $bxy = 0.45$

Q22 $y = -2x+3$ — y on x

$byx = -x$

☆ $8x-y+3$ — x on y

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

$bxy = -\frac{1}{8}$

$r = \sqrt{bxy \times byx}$

$r = \sqrt{-2 \times -\frac{1}{8}}$

$r = \sqrt{\frac{1}{4}}$

$r = -0.5$ — (C) Ans

How to calculate buy & byx :-

i) when r is given :-

☆ $byx = r \times \frac{\sigma y}{\sigma x}$

☆ $bxy = r \times \frac{\sigma x}{\sigma y}$

ii) when cov(x,y) given :-

$byx = \frac{cov(x,y)}{(\sigma x)^2}$

$bxy = \frac{cov(x,y)}{\sigma(y)^2}$

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

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

Q23 :- $byx = -\frac{3}{4}$

$r = \frac{\sqrt{3}}{2}$

$(\sigma y)^2$ ⇒ ?

⇒ $byx = r \times \frac{\sigma y}{\sigma x}$

⇒ $-\frac{3}{4} = \frac{\sqrt{3}}{2} \times \frac{2}{\sigma x}$ ⇒

$-\frac{3}{4} = \frac{\sqrt{3}}{\sigma x}$

Squaring both side :-

$\frac{9}{16} = \frac{3}{(\sigma x)^2}$

cross multiply

$(\sigma x)^2 = \frac{16 \times 3}{9}$

$(\sigma x)^2 = \frac{16}{3}$ — (b) Ans

Important Points Related to Correlation & Regression

1. The two lines of regression coincide i.e become identical when $r = \pm 1$ or ± 1. or in other words, there is perfect negative or positive correlation between 2 variable under discussion if $r = 0$, Regression lines are perpendicular to each other.

2. The Ratio between Explained variance of total variance

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

3. Karl Pearson's Product moment correlation coefficient:-

$$r \Rightarrow r_{xy} \Rightarrow \frac{Cov(x,y)}{S_x \times S_y}$$

4. The two lines of Regression Intresect at the point $(\bar{x}, \bar{y})$ where x & y are the variable under consideration.

5. The Regression coefficient Remain Unchanged due to a shift ~~Sto~~ a orign but chage due to a shift of scale.

Q20) $b_{yx} \Rightarrow 2.4$

$b_{YU} \Rightarrow b_{yx} \times \frac{M_Y}{M_X}$

$\Rightarrow 2.4 \times \frac{-3}{2} \Rightarrow -3.6$ - (5)

6.) The Minimization of vertical Distances in the Scatter Diagram.

Normal form is also called 'least square' method.

The difference between the observed value & the Estimated value & ε_i is technically known as Error or residue.

Bivariate Data :- when data are collected on two variables simultaneously, they are known as bivariate Data.

Marginal Distribution In correlation :- 2

Conditional Distribution $\Rightarrow p+2$

cells $\Rightarrow$ Row x Column i.e $p \times q$

cells never be negative it may zero or positive.

Spurious correlation :- where no casual Relation.

Most Quickest Method to find correlation between 2 variables :- Method of concurrent deviation.

Error may positive, negative or zero also.