UNIT-V

①

I Correlation

$$\boxed{r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}} \quad \text{(or)} \quad r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\,\sqrt{\Sigma y^2}}}$$

$x = X - \bar{X}$ $\quad$ $y = Y - \bar{Y}$

BOTH FORMULAE SAME

II Correlation

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{n(\Sigma X^2) - (\Sigma X)^2}\,\sqrt{n(\Sigma Y^2) - (\Sigma Y)^2}}$$

(OR)

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[n(\Sigma X^2) - (\Sigma X)^2\right]\left[n(\Sigma Y)^2 - (\Sigma Y)^2\right]}}$$

III Correlation (Assumed Mean Method)

BOTH FORMULAE SAME

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

OR

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

where $x = X - \bar{X}$

$y = Y - \bar{Y}$

2021/4/20 12:03

N2

(I) (3)

(1) Calculate Karl Pearson's Coefficient of Correlation between the expenditure on advertising and Sales from the data given below

Advertising expenses ('000 RS	39	65	62	90	82	75	25	98	36	78
Sales (Lakh Rs)	47	53	58	86	62	68	60	91	51	84

Let Advertising expenses $= X$

Sales $= Y$

$n = 10$

$\Sigma X = 650$

$\Sigma Y = 660$

$\bar{X} = \frac{\Sigma X}{n} = \frac{650}{10} = 65$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{660}{10} = 66$

X	Y	$x = X - \bar{X}$ $x = X - 65$	$y = Y - \bar{Y}$ $y = Y - 66$	x^2	y^2	xy
39	47	-26	-19	676	361	494
65	53	0	-13	0	169	0
62	58	-3	-8	9	64	24
90	86	25	20	625	400	500
82	62	17	-4	289	16	-68
75	68	10	2	100	4	20
25	60	-40	-6	1600	36	240
98	91	33	25	1089	625	825
36	51	-29	-15	841	225	435
78	84	13	18	169	324	234

$\Sigma X = 650$ $\Sigma Y = 660$ $\Sigma x^2 = 5398$ $\Sigma y^2 = 2224$ $\Sigma xy = 2704$

25

2021/4/20 12:03

$n = 10$

$\Sigma x^2 = 5398$

$\Sigma y^2 = 2224$

$\Sigma xy = 2704$

Correlation Coefficient

$$\boxed{r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \; \Sigma y^2}}}$$

where $x = X - \bar{X}$

$y = Y - \bar{Y}$

$$r = \frac{2704}{\sqrt{5398 \times 2224}}$$

$$r = \frac{2704}{\sqrt{12005152}}$$

$$r = \frac{2704}{3464.8452}$$

$$\boxed{r = 0.7804}$$

MAY 2012

I

2) For the Following data Compute Coefficient of Correlation and interpret the result.

X	100	102	104	107	105	112	103	99
Y	15	12	13	11	12	12	19	26

$n = 8$

$\Sigma X = 832 \qquad \bar{X} = \frac{\Sigma X}{n} = \frac{832}{8} = 104$

$\Sigma Y = 120 \qquad \bar{Y} = \frac{\Sigma Y}{n} = \frac{120}{8} = 15$

X	Y	$x = X - \bar{X}$ $x = X - 104$	$y = Y - \bar{Y}$ $y = Y - 15$	x^2	y^2	xy
100	15	-4	0	16	0	0
102	12	-2	-3	4	9	6
104	13	0	-2	0	4	0
107	11	3	-4	9	16	-12
105	12	1	-3	1	9	-3
112	12	8	-3	64	9	-24
103	19	-1	4	1	16	-4
99	26	-5	11	25	121	-55

$\Sigma x^2 = 120 \quad \Sigma y^2 = 184 \quad \Sigma xy = -92$

$\Sigma X = 832 \quad \Sigma Y = 120$

$n = 8$

$\Sigma x^2 = 120$

$\Sigma y^2 = 184$

$\Sigma xy = -92$

2021/4/20 12:03

$$\boxed{r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}}}$$

$x = X - \bar{X}$
$y = Y - \bar{Y}$

$$r = \frac{-92}{\sqrt{(120)(184)}}$$

$$r = \frac{-92}{\sqrt{22080}}$$

$$r = \frac{-92}{148.5934}$$

$$r = -0.619$$

$$\approx \; r = -0.62$$

There exist fairly negative Correlation between the two variables X & Y.

2021/4/20 12:03

(2) Calculate Correlation Coefficient for the following data

$n = 4$

$\Sigma x = 10 \qquad \Sigma x^2 = 30$

$\Sigma y = 30 \qquad \Sigma y^2 = 354 \qquad \Sigma xy = 100$

Sol

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

$$r = \frac{(4 \times 100) - (10)(30)}{\sqrt{\left[(4 \times 30) - (10)^2\right]\left[(4 \times (354)) - (30)^2\right]}}$$

$$r = \frac{400 - 300}{\sqrt{(120 - 100)(1416 - 900)}}$$

$$r = \frac{400 - 300}{\sqrt{(20)(516)}}$$

$$r = \frac{100}{\sqrt{10320}} = \frac{100}{101.59} = 0.9843$$

$$\boxed{r = 0.9843}$$

$r = 0.9843$???

Note: Not

X	1	2	3	4
Y	1	4	9	16

(1) For the following data compute correlation coefficient.

$n = 25$

$\Sigma X = 125 \qquad \Sigma X^2 = 650$

$\Sigma Y = 100 \qquad \Sigma Y^2 = 436 \qquad \Sigma XY = 520$

Sol:

Correlation Coefficient

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n(\Sigma X^2) - (\Sigma X)^2] \times [n(\Sigma Y^2) - (\Sigma Y)^2]}}$$

$$r = \frac{(25 \times 520) - (125)(100)}{\sqrt{[25(650) - (125)^2] \times [25(436) - (100)^2]}}$$

$$r = \frac{13000 - 12500}{\sqrt{[16255 - 15625][10900 - 10000]}}$$

$$r = \frac{500}{\sqrt{625 \times 900}}$$

$$r = \frac{500}{\sqrt{562500}} = \frac{500}{750} = 0.67$$

$$\boxed{r = 0.67}$$

$n = 8$

working Mean of $X \Rightarrow A = 69$

working Mean of $Y \Rightarrow B = 112$

X	Y	$x = X - A$ $x = X - 69$	$y = Y - B$ $y = Y - 112$	x^2	y^2	xy
78	125	$78 - 69 = 9$	$125 - 112 = 13$	81	169	117
89	137	$89 - 69 = 20$	$137 - 112 = 25$	400	625	500
96	156	$96 - 69 = 27$	$156 - 112 = 44$	729	1936	1188
69	112	$69 - 69 = 0$	$112 - 112 = 0$	0	0	0
59	107	$59 - 69 = -10$	$107 - 112 = -5$	100	25	50
79	136	$79 - 69 = 10$	$136 - 112 = 24$	100	576	240
68	123	$68 - 69 = -1$	$123 - 112 = 11$	1	121	-11
61	108	$61 - 69 = -8$	$108 - 112 = -4$	64	16	32

$\Sigma X = 599$ $\Sigma Y = 1004$ $\Sigma x = 47$ $\Sigma y = 108$ $\Sigma x^2 = 1475$ $\Sigma y^2 = 3468$ $\Sigma xy = 2116$

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

$$r = \frac{(8)(2116) - (47)(108)}{\sqrt{\left((8)(1475) - (47)^2\right)\left(8(3468) - (108)^2\right)}}$$

$$r = \frac{(8)(2116) - (47)(108)}{\sqrt{\left((8)(1475) - (47)^2\right)\left((8)(3468) - (108)^2\right)}}$$

$$r = \frac{16928 - 5076}{\sqrt{(11800 - 2209)(27744 - 11664)}}$$

$$r = \frac{11852}{\sqrt{(9591)(16080)}}$$

$$r = \frac{11852}{\sqrt{154223280}}$$

$$r = \frac{11852}{12418.6666}$$

$$\boxed{r = 0.954}$$

(PTO)

1. Calculate Coefficient of Correlation between X and Y Series from the following data and calculate its probable error also.

X	78	89	96	69	59	79	68	61
Y	125	137	156	112	107	136	123	108

Take 69 as working mean for X and 112 as working mean for Y

Sol: Coefficient of Correlation by Working Mean Method or Assumed Mean Method

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

where $x = X - A$

$y = Y - B$

A = Assumed (working) mean of X's

B = Assumed (working) mean of Y's

n = number of observations

Practice

(3) Find Correlation Coefficient for the following data

$n = 10$

$\Sigma X = 195$ $\quad \Sigma X^2 = 4185$

$\Sigma Y = 149$ $\quad \Sigma Y^2 = 2681$ $\quad \Sigma XY = 3446$

Ans = $\boxed{r = 0.96}$

(4) Find Correlation Coefficient for the followind.

data

$n = 9$ $\quad \Sigma X = 45$ $\quad \Sigma X^2 = 285$ $\quad \Sigma Y = 108$ $\quad \Sigma Y^2 = 1356$ $\quad \Sigma XY = 597$

Ans $\boxed{r = 0.95}$

REGRESSION

(1) For the following data obtain two regression equations

Sales :	91	97	108	121	67	124	51	73	111	57
Purchases :	71	75	69	97	70	91	39	61	80	47

Sol :

let Sales = X

Purchases = Y

X	Y	X^2	Y^2	XY
91	71	8281	5041	6461
97	75	9409	5625	7275
108	69	11664	4761	7452
121	97	14641	9409	11737
67	70	4489	4900	4690
124	91	15376	8281	11284
51	39	2601	1521	1989
73	61	5329	3721	4453
111	80	12321	6400	8880
57	47	3249	2209	2679
$\Sigma X = 900$	$\Sigma Y = 700$	$\Sigma X^2 = 87360$	$\Sigma Y^2 = 51868$	$\Sigma XY = 66900$

$n = 10$ $\qquad \bar{X} = \frac{\Sigma X}{n} = \frac{900}{10} = 90$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{700}{10} = 70$

2021/4/23 11:44

Q. 6

20

$n = 10 \qquad \bar{X} = 90 \qquad \bar{Y} = 70$

$\Sigma X = 900 \qquad \Sigma X^2 = 87360$

$\Sigma Y = 700 \qquad \Sigma Y^2 = 51868 \qquad \Sigma XY = 66900$

Regression Eqn of Y on X

$$\boxed{Y = a + bX}$$

$$\boxed{b = \frac{\Sigma XY - n(\bar{X})(\bar{Y})}{\Sigma X^2 - n(\bar{X})^2}}$$

$$\boxed{a = \bar{Y} - b\bar{X}}$$

$$b = \frac{\Sigma XY - n(\bar{X})(\bar{Y})}{\Sigma X^2 - n(\bar{X})^2}$$

$$b = \frac{66900 - (10)(90)(70)}{87360 - (10)(90)^2}$$

$$b = \frac{66900 - 63000}{87360 - (10)(8100)}$$

$$b = \frac{3900}{87360 - 81000}$$

$$b = \frac{3900}{6360}$$

$$\boxed{b = 0.6132}$$

NOTE: $\boxed{b_{YX} = 0.6132}$

vg.

21

$$\boxed{a = \bar{Y} - b\bar{X}}$$

$$a = 70 - (0.6132)(90)$$

$$a = 70 - 55.188$$

$$\boxed{a = 14.812}$$

Regression eqn of Y on X

$$Y = a + bX$$

$$\boxed{Y = 14.812 + (0.6132)X}$$

$\frac{1}{5}$

Regression eqn of X on Y

$$\boxed{X = a + bY}$$

$\frac{2}{3}$

$$\boxed{b = \frac{\Sigma XY - n(\bar{X})(\bar{Y})}{\Sigma Y^2 - n(\bar{Y})^2}}$$

$$\boxed{a = \bar{X} - b\bar{Y}}$$

$$b = \frac{66900 - (10)(90)(70)}{51868 - (10)(70)^2}$$

$$b = \frac{66900 - 63000}{51868 - (10)(4900)}$$

$$b = \frac{66900 - 63000}{51868 - 49000}$$

2021/4/23

$$b = \frac{3900}{2868}$$

$$\boxed{b = 1.3598}$$

Note $\boxed{b_{XY} = 1.3598}$

$$a = \bar{X} - b\bar{Y}$$

$$a = 90 - (1.3598)(70)$$

$$a = 90 - 95.186$$

$$\boxed{a = -5.186}$$

Regression eqn of X on Y

$$X = a + bY$$

$$\boxed{X = -5.186 + (1.3598)Y}$$

NOTE: Two regression coefficients

$b_{YX} = 0.6132$
$b_{XY} = 1.3598$

Note: To find Correlation $r = \sqrt{b_{YX}\ b_{XY}}$

$$r = \sqrt{(0.6132)(1.3598)}$$

$$r = \sqrt{0.83383}$$

$$r = 0.913$$

2021/4/23 11:44

N8

REGRESSION

Y on x

Regression Equation of Y on X

is $\boxed{Y = a + bX}$ formula

Constant Coeff

$$\boxed{b = \frac{\Sigma XY - n(\bar{X})(\bar{Y})}{\Sigma X^2 - n(\bar{X})^2}}$$

$\bar{X}$ = mean

$$\boxed{a = \bar{Y} - b\bar{X}}$$

X →
Y →

n = no. of observations

$$\bar{X} = \frac{\Sigma X}{n} \qquad \bar{Y} = \frac{\Sigma Y}{n}$$

★ (1) April 2007

For the following data develop the estimating equation the best describes the data.

X	16	6	10	5	12	14	8	15
Y	-4.4	8.0	2.1	8.7	0.1	-2.9	1.9	-3.5

Predict Y for X = 5, 7, and 9.

$n = 8$

$\Sigma X = 86$ $\qquad \bar{X} = \frac{\Sigma X}{n} = \frac{86}{8} = 10.75$

$\Sigma Y = 10$ $\qquad \bar{Y} = \frac{\Sigma Y}{n} = \frac{10}{8} = 1.25$

we have to develop a regression equation of the form $Y = a + bX$

where $b = \frac{\Sigma XY - n(\bar{X})(\bar{Y})}{\Sigma X^2 - n(\bar{X})^2}$ $\qquad a = \bar{Y} - b\bar{X}$

X	Y	X^2	Y^2	XY
16	-4.4	256	19.36	-70.4
6	8.0	36	64	48
10	2.1	100	4.41	21
5	8.7	25	75.69	43.5
12	0.1	144	0.01	1.2
14	-2.9	196	8.41	-40.6
8	1.9	64	3.61	15.2
15	-3.5	225	12.25	-52.5
$\Sigma X = 86$	$\Sigma Y = 10$	$\Sigma X^2 = 1046$	$\Sigma Y^2 = 187.74$	$\Sigma XY = -34.6$

2021/4/23 11:45

$$b = \frac{\Sigma xy - n(\bar{x})(\bar{y})}{\Sigma x^2 - n(\bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{(-34.6) - 8(10.75)(1.25)}{1046 - 8(10.75)^2}$$

$$b = \frac{(-34.6) - 107.5}{1046 - 8(115.5625)}$$

$$b = \frac{-34.6 - 107.5}{1046 - 924.5}$$

$$b = \frac{-142.1}{121.5}$$

$$\boxed{b = -1.1695}$$

$$a = \bar{y} - b\bar{x}$$

$$= (1.25) - (-1.1695)(10.7$$

$$a = 1.25 - (-1.1695)(10.75$$

$$= 1.25 + 12.57$$

$$= 13.82$$

$$\boxed{a = 13.82}$$

$$\boxed{Y = a + bx}$$

$$\boxed{Y = 13.82 - 1.1695x}$$

$+(-1.1695)$

$= -1.695$

This is Regression Equation of Y on x

Now we have to Predict Y when $x = 5, 7, 9$

2021/4/23 11:45

$$\boxed{Y = 13.82 - 1.1695X}$$

when $x = 5$

$$Y = 13.82 - 1.1695(5)$$
$$= 13.82 - 5.8475$$
$$Y = 7.9725$$

when $x = 7$

$$Y = 13.82 - 1.1695(7)$$
$$= 13.82 - 8.1865$$
$$= 5.6335$$

when $x = 9$

$$Y = 13.82 - 1.1695(9)$$
$$= 13.82 - 10.5255$$
$$Y = 3.2945$$