PAGE NO :
DATE :

Unit-2 [Predictive Analytics]

COPA01 ⇒ Collection, Organization, Presentation, Analysis, Interpretation.

Predictive Analytics:- Predictive modeling & Analysis.- Regression Analysis, correlation analysis, Rank correlation coefficient.

BIVARIATE DATA

The above example x, may represent height and y weight of a group of persons such as data (x_i, y_i) $i = 1, 2, \ldots n$ is called BD.

Application of predictive modeling

→ Predictive modeling is the process of using known results to create process, & validate a model that can be used to make future prediction.

→ Two of the most widely used predictive modeling technique are regression & neural network.

→ Companies can use predictive ~~technique~~ modeling to forecaste event, customer behaviour as well as financial ~~income~~ economic & market risk.

→ Other predictive modelings Techniques used by

Correlation

Consider a set of bivariate of ~~dd~~ data (x_i, y_i), $i = 1, 2, \ldots n$.

The variables are said to be correlated, if there exist a change in one variable corresponding, to a change in the other variable.

PAGE NO :
DATE :

Types of Correlation

i] Positive & negative Correlation

Correlation is positive [Direct] if the variables are direct / vary in the same direction. If they increase or decrease together

Correlation is negative [Inverse] if the variables are deviate / vary in the opposite direction. If one variable is increasing, the other variable is decreasing or vise versa.

ii] Simple, Partial & Multiple Correlation

Simple correlation are concerned with two variable only, while partial & multipal corrections are concerned with three or more related variables.

Eg Rice production (x), rain fall (y) & fertility (z) are partially connected.

iii] Linear & Nonlinear [curvilinear] correlation.

If the ~~most~~ amount of change in on variable tends to bear a constant ratio to the amount of change in the other variable, then the correlation is said to be Linear otherwise nonlinear.

Eg. If we double the amount of rainfall, the production of rice or wheat etc; would not necessary be doubled.

Note 1. If the variables do not show related variation, they are said to non-correlated.

Eg. The value of a rupee (x) & atmospheric temperature (y) are non-correlated.

PAGE NO :
DATE :

Measure of Correlation

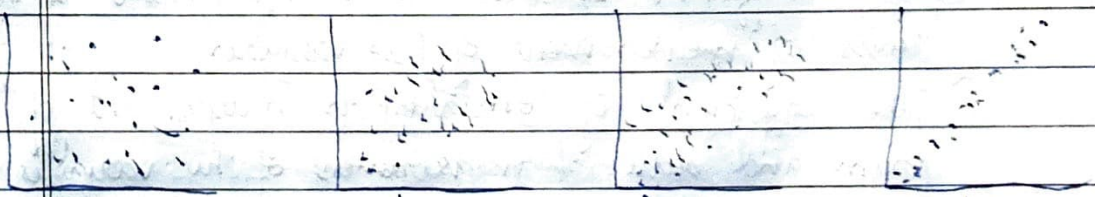
There are various method of ascertaining whether two variables are correlated or not of these methods, the following three are

- Scatter Diagram [Dot diagram]
- Karl Pearson's [Product moments] coefficient of correlation.
- Spearman's coefficient of rank correlation.

Scatter Diagram :-

It is a graphic representation of bivariate data. Here bivariate data with n pairs of values is represented by n points or dots on the xy-plane. The two variables are taken along the two axes and every pair of values in the data is represented by a point of graph.

Degrees of Correlation.



None Low High Perfect

Types of correlation

Karl Pearson's coefficient of correlation

This is a measure of linear relation

PAGE NO :
DATE :

ship b/w the two variables. It indicates the degree of correlations b/w the two variables. It is denoted by 'r'

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

Let $u = x - a$ here 'a' is assume value from 'x'
Let $v = y - b$ here 'b' is assume value from 'y'

$$r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{\left[n\Sigma u^2 - (\Sigma u)^2\right] * \left[n\Sigma v^2 - (\Sigma v)^2\right]}}$$

16/04/25 Thursday

Properties of coefficient of correlation

1. The coefficient of correlation is independent of the units of measurement of the variables
2. The coefficient of correlation is independent of the origin and scale of measurement of the variables
3. The coefficient of correlation is a value b/w -1 & +1.

Problem 1. Calculate the product moments coefficient of correlation b/w the following marks (out of 10) in computer science, mathematics of 5 marks.

Student	1	2	3	4	5
Computer	4	7	8	3	4
Mathematics	5	8	6	3	5

PAGE NO :
DATE :

Sol:- Let x & y be the marks of computer science and mathematics

Here n = 5

x	y	xy	x^2	y^2
4	5	20	16	25
7	8	56	49	64
8	6	48	64	36
3	3	9	9	9
4	5	20	16	25
$\Sigma x = 26$	$\Sigma y = 27$	$\Sigma xy = 153$	$\Sigma x^2 = 154$	$\Sigma y^2 = 158$

$\Sigma x = 26$

$\Sigma y = 27$

$\Sigma xy = 153$

$\Sigma x^2 = 154$

$\Sigma y^2 = 159$

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

$$= \frac{5 \times 153 - 26 \times 27}{\sqrt{5 \times 154 - 26^2 * 5 \times 159 - 27^2}}$$

$$= \frac{765 - 702}{\sqrt{770 - 676 * 795 - 729}}$$

$$= \frac{63}{\sqrt{94 * 66}}$$

$$= \frac{63}{\sqrt{6204}}$$

$$= \frac{63}{78.765} = 0.7998$$

.9598

PAGE NO :
DATE :

(2)

Husband	164	176	178	184	175	167	173	180
Wife	168	174	175	181	173	166	173	179

Sol ~~Let x & y → H. & W~~

Let $u = x - a$

$= x - 170$ (164 − 170)

Let $v = y - b$ (168 − 171)

$= y - 170$

x	y	u	v	uv	u^2	v^2
164	168	-6	-2	12	36	4
176	174	6	4	24	36	16
178	175	8	5	40	64	25
184	181	14	11	154	196	121
175	173	5	3	15	25	9
167	166	-3	-4	12	9	16
173	173	3	3	9	9	9
180	179	10	9	90	100	81
Σ		$\Sigma u = 37$	$\Sigma v = 29$	$\Sigma uv = 356$	$\Sigma u^2 = 475$	$\Sigma v^2 = 281$

$\Sigma u = 37$
$\Sigma v = 29$
$\Sigma uv = 356$
$\Sigma u^2 = 475$
$\Sigma v^2 = 281$
$n = 8$

$$r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{\left[n\Sigma u^2 - (\Sigma u)^2\right] * \left[n\Sigma v^2 - (\Sigma v)^2\right]}}$$

$$= \frac{8 \times 356 - 37 \times 29}{\sqrt{\left[8 \times 475 - 37^2\right] * \left[8 \times 281 - 29^2\right]}}$$

$$= \frac{2,848 - 1073}{\sqrt{3,800 - 1369 * 2248 - 841}}$$

$= \frac{1775}{\sqrt{2431 * 1407}}$

PAGE NO :
DATE :

$= \frac{1775}{\sqrt{3420417}} = \frac{1775}{1849.4369} = 0.9597$

(3)

x	218	220	236	225	220	227	228
y	12.3	12.7	12.0	12.2	12.7	12.1	12.0

Sol Let $u = x - a$
$= x - 220$

$v = y - b$
$= y - 12$

x	y	u	v	uv	u^2	v^2
218	12.3	-2	0.3	-0.6	4	0.09
220	12.7	0	0.7	0	0	0.49
236	12.0	16	0	0	256	0
225	12.2	5	0.2	1	25	0.04
220	12.7	0	0.7	0	0	0.49
227	12.1	7	0.1	0.7	49	0.01
228	12.0	8	0	0	64	0
		$\Sigma u = 34$	$\Sigma v = 2$	$\Sigma uv = 1.1$	$\Sigma u^2 = 398$	$\Sigma v^2 = 1.12$

$r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{\left[n\Sigma u^2 - (\Sigma u)^2\right] * \left[n\Sigma v^2 - (\Sigma v)^2\right]}}$

$= \frac{7 \times 1.1 - 34 \times 2}{\sqrt{7 \times 398 - 34^2 * 7 \times 1.12 - 2^2}}$

$= \frac{7.7 - 68}{\sqrt{2786 - 1156 * 7.84 - 4}} = \frac{-60.3}{\sqrt{1630 * 3.84}} = \frac{-60.3}{\sqrt{6.259.2}}$

PAGE NO :
DATE :

Spearman's coefficient of ranks of correlation

Suppose that a group of n individuals are arranged in the order of merit or efficency with respect to some characteristics.

In a bivariate data, if the values of the variables are ranked in the decreasing (or increasing) order, the correlation b/w these rank is rank correlation. The coefficient of correlation computed for these ranks is Spearman's coefficient of rank correlation. Denoted by ρ. (Read as Rho).

If R_1 and R_2 are the ranks in the two characteristics and $d = R_1 - R_2$ is the d/f b/w the ranks, coefficient of correlation is

$$\rho = 1 - \left(\frac{6\Sigma d^2}{n^3 - n}\right)$$

Problem ①

Marks in Statistics	25	43	27	35	54	61	37	45
Marks in maths	35	47	20	37	63	54	28	40

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
25	35	8	6	8-6 = 2	4
43	47	4	3	4-3 = 1	1
27	20	7	8	7-8 = -1	1
35	37	6	5	6-5 = 1	1
54	63	2	1	2-1 = 1	1
61	54	1	2	1-2 = -1	1
37	28	5	7	5-7 = -2	4
45	40	3	4	3-4 = -1	1
					$\Sigma d^2 = 14$

$$p = 1 - \left(\frac{6 \cdot \Sigma d^2}{n^3 - n}\right)$$

0.8333 sol

PAGE NO :
DATE :

$$= 1 - \frac{6(14)}{8^3 - 8} = 1 - \frac{84}{512 - 8} = 1 - 0.1666$$

$$= 0.8333$$

H/W Karl pearson's problem

①

x	65	66	67	67	68	69	70	73
y	67	68	65	68	72	72	69	71

Sol: Let $u = x - a$

$= x - 68$

$v = y - b$

$= y - 69$

x	y	u	v	uv	u^2	v^2
65	67	-3	-2	6	9	4
66	68	-2	-1	2	4	1
67	65	-1	-4	4	1	16
67	68	-1	-1	1	1	1
68	72	0	3	0	0	9
69	72	1	3	3	1	9
70	69	2	0	0	4	0
73	71	4	2	8	16	4
		$\Sigma u = 0$	$\Sigma v = 0$	$\Sigma uv = 24$	$\Sigma u^2 = 36$	$\Sigma v^2 = 44$

$$r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{\left[n\Sigma u^2 - (\Sigma u)^2\right] * \left[n\Sigma v^2 - (\Sigma v)^2\right]}}$$

$$= \frac{8 \times 24 - 0 \times 0}{\sqrt{8 \times 36 - 0^2 * 8 \times 44 - 0^2}}$$

$= 0.6030$

$$= \frac{192 - 0}{\sqrt{288 * 352}} = \frac{192}{\sqrt{101376}} = \frac{192}{318.3959}$$

17/01/25
Friday

PAGE NO :
DATE :

Spearman's coefficient of ranks correlation

Problem ①

mathematics	38	50	42	61	43	55	67	46	72
Statistics	41	64	70	75	44	55	62	56	60

Sol:- $n = 9$

different data

$$\rho = 1 - \left(\frac{6.\Sigma d^2}{n^3 - n}\right)$$

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
38	41	9	9	0	0
50	64	5	3	2	4
42	70	8	2	6	36
61	75	3	1	2	4
43	44	7	8	-1	1
55	55	4	6.7	-3	9
67	62	2	4	-2	4
46	56	6	6	0	0
72	60	1	5	-4	16
					$\Sigma d^2 = 74$

$$\rho = 1 - \frac{6(74)}{9^3 - 9}$$

$$\rho = 1 - \frac{444}{729 - 9}$$

$$\rho = 1 - \frac{444}{720}$$

$$= 0.3833$$

H/W ②

$3 \quad 3 \quad 3$
36×6
216

96

PAGE NO :
DATE :

Height	165	167	166	170	169	172
Weight	61	60	63.5	63	61.5	64

Sol:- $n = 6$

X	Y	R_1	R_2	$d = R_1 - R_2$	d^2
165	61	6	5	1	1
167	60	64	6	-2	4
166	63.5	45	2	3	49
170	63	2	3	-1	1
169	61.5	3	4	-1	1
172	64	1	1	0	0
					$\Sigma d^2 = 16$

$$P = 1 - \frac{6(16)}{6^3 - 6}$$

$$= 1 - \frac{96}{216-6} = 1 - \frac{96}{210} = 0.5428$$

Repeated Rank [Data with Ties]

If one rank repeat m_1 times, another rank repeat m_2 times and third rank repeat m times and so that the correlation factor [C.F] is

$$C.F = \frac{m_1^3 - m_1}{12} + \frac{m_2^3 - m_2}{12} + \frac{m_3^3 - m_3}{12} + \dots$$

The coefficient of rank correlation

$$p = 1 - \frac{6[\Sigma d^2 + C.F]}{n^3 - n}$$

PAGE NO :
DATE :

Problem ① The following table using repeated ranks

x	43	96	74	38	35	43	22	56	35	80
y	30	94	84	13	30	18	30	41	48	95

Sol:- $n = 10$

There are similar data rank so we use

$$\rho = 1 - \frac{6[\Sigma d^2 + C.F]}{n^3 - n}$$

Correlation Factor $= C.F = \frac{m_1^3 - m_1}{12} + \frac{m_2^3 - m_2}{12} + \frac{m_3^3 - m_3}{12} + \dots$

$\frac{5+6}{2} = 5.5$ $\quad \frac{8+9}{2} = \frac{17}{2} = 8.5$

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
43	30	5.5	7	−1.5	2.25
96	94	1	2	−1	1
74	84	3	3	0	0
38	13	7	10	−3	9
35	30	8.5	7	1.5	2.25
43	18	5.5	9	−3.5	12.25
22	30	10	7	3	9
56	41	4	5	−1	1
35	48	8.5	4	4.5	20.25
80	95	2	1	1	1

$\Sigma d^2 = 58$

$\frac{6+7+8}{3} = \frac{21}{3} = 7$

43 is repeated 2 times so

$m_1 = 2$

35 is repeated 2 times so

$m_2 = 2$

30 is repeated 3 times so

$m_3 = 3$

$\frac{}{12} + \frac{}{12} + \frac{}{12}$

PAGE NO :
DATE :

$= \frac{8-2}{12} + \frac{8-2}{12} + \frac{27-3}{12}$

$= \frac{6+6+24}{12} = \frac{36}{12} = 3$

$p = 1 - \frac{6(58)+3)}{10^3 - 10}$

$= 1 - \frac{366}{990} \Rightarrow 0.6303$

20/01/25 Monday

Introduction Regression:-

The statistical tool with the help of which it is possible to estimate [or predict] the unknown values of one variable from the known values of other variable [when the variable are correlated] is called Regression.

The Regression theory was develop by Sir Fanics galton.

y on x-axis

$(y-\bar{y}) = byx(x-\bar{x})$

where $\bar{x} = a + \frac{\Sigma u}{n}$, $\bar{y} = b + \frac{\Sigma v}{n}$

byx is coefficient of y on x-axis

$byx = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma u^2 - (\Sigma u)^2}$

x on y-axis

$(x-\bar{x}) = bxy(y-\bar{y})$

where $\bar{x} = a + \frac{\Sigma u}{n}$, $\bar{y} = b + \frac{\Sigma v}{n}$

b_{xy} is coefficient of x on y-axis

$$b_{xy} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma v^2 - (\Sigma v)^2}$$

PAGE NO :
DATE :

$$= \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} = r \cdot \frac{\sigma_x}{\sigma_y}$$

Properties of Regression line

→ The regression line intersect at $(\bar{x}, \bar{y})$.

→ If there is perfect correlation, the regression line coincide [there will be only on regression line]

→ The geometric mean of the regression coefficient is equal to coefficient of correlation [numerically]. That is

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

→ The regression coefficients cannot be of opposite sign.

* If r is positive, both the regression coefficient will be positive
* If r is negative, both the regression coefficient will be negative.
* If r is zero, both the regression coefficient will be zero.

Problem ① Obtain two regression equations [height - x] [weight - y]

x	153	157	168	160	170	163
y	48	50	50	49	54	59

i) Height of a person whoes weight is 60 kg
ii) Weight of a person whoes height is 165 kg

Sol. y on x-axis

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

where

$$\bar{x} = a + \frac{\Sigma u}{n} \quad \& \quad \bar{y} = b + \frac{\Sigma v}{n}$$

$$b_{yx} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma u^2 - (\Sigma u)^2}$$

x on y axis

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

PAGE NO :
DATE :

where $\bar{x} = a + \frac{\Sigma u}{n}$ & $\bar{y} = b + \frac{\Sigma v}{n}$

$b_{xy} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma v^2 - (\Sigma v)^2}$

21/6/25 Tuesday

Let $u = x - a$
$= x - 160$

Let $v = y - b$
$= y - 50$

$a = 160$
$b = 50$

x	y	u	v	u^2	v^2	uv
153	48	-7	-2	49	4	+14
157	50	-3	0	9	0	0
168	50	8	0	64	0	0
160	49	0	-1	0	1	0
170	54	10	4	100	16	40
163	53	3	3	259	9	9
x_2	y_2	$\Sigma u = 11$	$\Sigma v = 4$	$\Sigma u^2 = 231$	$\Sigma v^2 = 30$	$\Sigma uv = 63$

$\bar{x} = a + \frac{\Sigma u}{n}$

$= 160 + \frac{11}{6}$

$= 161.83$

$\bar{y} = b + \frac{\Sigma v}{n}$

$= 50 + \frac{4}{6}$

$= 50.66$

$b_{yx} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma u^2 - (\Sigma u)^2}$

$= \frac{6(63) - 11 \times 4}{6(231) - (11)^2}$

$= \frac{378 - 44}{1386 - 121}$

$= \frac{334}{1266}$

$= 0.264$

$b_{xy} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma v^2 - (\Sigma v)^2}$

$= \frac{334}{6(30) - 4^2}$

$= \frac{334}{180 - 16}$

$= \frac{334}{164}$

$= 2.03$

y on x-axis

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

PAGE NO :
DATE :

$(y - 50.66) = 0.264(x - 161.83)$

$y - 50.66 = 0.264x - 42.72$

$y = 0.264x - 42.72 + 50.66$

$y = 0.264x + 7.94$ — (1)

x on y-axis

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

$(x - 161.83) = 2.03(y - 50.66)$

$x = 2.03y - 102.83 + 161.83$

$x = 2.03y + 59$ — (2)

i] To find weight of a person where height is 165

$y = ?$, $x = 165$

$y = 0.264(165) + 7.94$

$= 43.56 + 7.94$

$= 51.5$

ii] To find height of a person where weight is 60

$x = ?$, $y = 60$

$x = 2.03(60) + 59$

$= 121.8 + 59$

$= 180.8$

22/01/25
Wednesday

② Find the co-efficient of correlation & equation of the lines of regression

x	5	2	1	3	4

$a = 3$

$b = 5$

Coefficient of correlation $= \sqrt{b_{xy} * b_{yx}}$

PAGE NO :
DATE :

80.

x	y	u	v	u^2	v^2	uv
6	4	2	2	4	4	4
2	5	-1	0	1	0	0
1	2	-2	-3	4	9	6
3	3	0	-2	0	4	0
4	8	1	3	1	9	3
		$\Sigma u = 0$	$\Sigma v = 0$	$\Sigma u^2 = 10$	$\Sigma v^2 = 26$	$\Sigma uv = 13$

$\bar{x} = a + \frac{\Sigma u}{n}$ $\qquad$ $\bar{y} = b + \frac{\Sigma v}{n}$

$= 3 + \frac{0}{5}$ $\qquad$ $= 5 + \frac{0}{5}$

$= 3$ $\qquad$ $= 5$

$b_{yx} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma u^2 - (\Sigma u)^2}$ $\qquad$ $b_{xy} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma v^2 - (\Sigma v)^2}$

$= \frac{5(13) - 0 \times 0}{5(10) - (0)^2}$ $\qquad$ $= \frac{65}{5(26) - 0^2}$

$= \frac{65 - 0}{50 - 0}$ $\qquad$ $= \frac{65}{130 - 0}$

$= \frac{65}{50}$ $\qquad$ $= \frac{65}{130}$

$= 1.08$ $\qquad$ $= 0.5$

1.3

y on x-axis

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

$y - 5 = 1.3(x - 3)$

$y = 1.3x - 3.9 + 5$

$y = 1.3x + 1.1$ — (1)

x on y-axis

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

$x - 3 = 0.5(y - 5)$

$x = 0.5y - 2.5 + 3$

$x = 0.5y + 0.5$ — (2)

Coefficient of correlation

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

PAGE NO :
DATE :

$= \sqrt{0.5 \times 1.3}$

$= \sqrt{0.65}$ ∴ The coefficient of correlation is highly +ve.

$= 0.8$

(3) Calculate the coefficient of correlation & line of regression. Also estimate y corresponding to x = 6.2

x	1	2	3	4	5	6	7	8
y	9	8	10	12	11	13	14	16

Sol: Find y on x-axis? $\frac{4+5}{2} = \frac{9}{2} = 4.5 \rightarrow a$

x	y	$u = x - 4.5$	$v = y - 11$	u^2	v^2	uv
1	9	-3.5	-2	12.25	4	7
2	8	-2.5	-3	6.25	9	7.5
3	10	-1.5	-1	2.25	1	1.5
4	12	-0.5	1	0.25	1	-0.5
5	11	0.5	0	0.25	0	0
6	13	1.5	2	2.25	4	3
7	14	2.5	3	6.25	9	7.5
8	16	3.5	5	12.25	25	17.5
		$\Sigma u = 0$	$\Sigma v = 5$	$\Sigma u^2 = 42$	$\Sigma v^2 = 53$	$\Sigma uv = 43.5$

$\bar{x} = a + \frac{\Sigma u}{n}$ $\bar{y} = b + \frac{\Sigma v}{n}$

$= 4.5 + \frac{0}{8}$ $= 11 + \frac{5}{8}$

$= 4.5$ $= 11.625$

$b_{yx} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma u^2 - (\Sigma u)^2}$ $b_{xy} = \frac{n\Sigma uv - \Sigma u \Sigma v}{n\Sigma v^2 - (\Sigma v)^2}$

$= \frac{8(43.5) - 0 \times 5}{8(42) - 0^2}$ $= \frac{348}{8(53) - 5^2}$ 0.8721

$= \frac{348}{336} = 1.0357$ $= \frac{348}{424 - 25} = \frac{348}{399} = 0.87$

y on x-axis

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

$y - 11.625 = 1.0357(x-4.5)$

PAGE NO :
DATE :

$y = 1.0357x - 4.6606 + 11.625$ ⑥

$y = 1.0357x + 6.9644$ — ①

Coefficient of correlation

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

$= \sqrt{0.8721 \times 1.0357}$

$= \sqrt{0.9032}$

$= 0.9503$

$=$

∴ The coefficient of correlation is highly +ve.

∴ $x = 6.2$

$y = 1.0357x + 6.9644$

$= 1.0357(6.2) + 6.9644$

$= 6.4213 + 6.9644$

$= 13.3857$

$=$

⇒ In a bivariate data, the regression lines are

$4x - 5y + 33 = 0$ &

$20x - 9y = 107$

find the mean of x & y coefficient of correlation [or find $\bar{x}$, $\bar{y}$ & r]

Sol: Given $4x - 5y = -33$ — ①

$20x - 9y = 107$ — ②

from ① & ② 'y' is a coefficient compared [20 & 4]

∴ 20x is greater so

x on y-axis

$20x - 9y = 107$

$20x = 9y + 107$

$x = \frac{9}{20}y + \frac{107}{20}$

∴ The coefficient of x on y-axis

$b_{xy} = \frac{9}{20}$

y on axis

$4x - 5y = -33$

$-5y = -4x - 33$

$x = \frac{5}{4}y + \frac{33}{4}$

Reverse ⇄ make +ve

∴ The coefficient of y on x-axis

$b_{yx} = \frac{5}{4}$

The coefficient of correlation

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

$= \sqrt{\frac{4}{5} \times \frac{9}{20}}$

$= \sqrt{\frac{9}{25}}$

$= \frac{9}{5}$

$= 0.6$

∴ The coefficient of correlation is highly +ve.

To find $\bar{x}$ & $\bar{y}$, WKT the two regression lines are intersect at $(\bar{x}, \bar{y})$

Here $4x - 5y = -33$ — ①

$20x - 9y = 107$ — ②

① mult by 20 ⇒ $80x - 100y = -660$

② mult by 4 ⇒ $80x - 36y = -428$

$-64y = -1088$

$y = \frac{-1088}{-64}$ 17 $\quad \boxed{y = 17}$

Substitute y with equation ①

$4x - 5y = -33$

$4x - 5(17) = -33$

$4x - 85 = -33$

$4x = -33 + 85$

$x = \frac{53}{4}$

$\boxed{x = 13}$

The mean of x & y is

$\bar{x} = x = 13$

$\bar{y} = y = 17$

4/5 ②

Find the regression lines are

$3x + 2y = 26$ &

$6x + y = 31$

find $\bar{x}$, $\bar{y}$ & r

PAGE NO :
DATE :

Sol:

Given $3x + 2y = 26$ — (1)

$6x + y = 31$ — (2)

∴ 6 is greater than equation (1) [3x] so

x on y-axis y on x-axis

$6x + y \geq 31$

$6x \geq$

PAGE NO :
DATE :

23/01/25
Thursday

Introduction - Curve fit, Least square, Goodness of fit.

One of the fields of elementary mathematics / statistics where the digital computer has a good amount of contribution is that of the curve.

The base problem in curve-fitting can be described as follows:

Let $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ be a set of n observations which can be represented on xy-plane. The problem of finding a functional relation of the form $y = f(x)$ which is satisfied by the given set of points is called curve fitting & the curve $y = f(x)$ is called the curve of the best fit.

In fitting a curve to given data points, there are two possible approaches.

i] To have a graph of the approximating functions that passes exactly through the given data points.

ii] To have an approximating function that has a smooth curve, and serves as the best approximation to the actual curve. The second curve is generally employed for curve fitting and is referred as the method of least squares and this is discussed in this chapter.

1] A Straight line $y = a + bx$

whose normall equations are

$\Sigma \rightarrow \Sigma y = na + b\Sigma x$

$\Sigma xy \rightarrow \Sigma xy = a\Sigma x + b\Sigma x^2$

PAGE NO :
DATE :

from this two equation 'a' & 'b' are determined. Substituting these values of 'a' & 'b' in $y = a + bx$ we get the equation of the (Straight) line of best fit for the given data.

Fitting a straight line problem

(1) Fit in a straight line for the following data

x	6	7	7	8	8	8	9	9	10
y	5	5	4	5	4	3	4	3	3

Sol:- $y = a + bx$

Let $X = x -$ center value

$\therefore$ Straight line becomes $y = a + bX$

Normal Equation

$\Sigma \rightarrow \Sigma y = na + b\Sigma X$

$\Sigma X \rightarrow \Sigma Xy = a\Sigma X + b\Sigma X^2$

$\therefore$ Let $X = x - 8$

$\therefore y = a + bX$

29/01/25 Friday

X	y	$X = x - 8$	Xy	X^2
6	5	-2	-10	4
7	5	-1	-5	1
7	4	-1	-4	1
8	5	0	0	0
8	4	0	0	0
8	3	0	0	0
9	4	1	4	1
9	3	1	3	1
10	3	2	6	4
	$\Sigma y = 36$	$\Sigma X = 0$	$\Sigma Xy = -6$	$\Sigma X^2 = 12$

PAGE NO :
DATE :

$\Sigma y = na + b\Sigma x$

$36 = 9a + b(0)$

$9a = 36$

$a = \frac{36^4}{9_1} = 4$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$-6 = a(0) + b(12)$

$-6 = 12b$

$12b = -6$

$b = \frac{-6}{12} = -0.5$

$y = a + bx$

$= 4 + (-0.5)(x - 8)$

$= 4 - 0.5x + 4$

$y = 0.5x + 8$

$=_1$

(2) Find a striaght line → mid-term

x	1911	1921	1931	1941	1951
y	15	23	28	32	39

Sol:- Let $X = x - 1931$

$\therefore y = a + bX$

x	y	$X = x - \frac{1931}{8}$	Xy	X^2
1911	15	-20	-300	400
1921	23	-10	-230	100
1931	28	0	0	0
1941	32	10	320	100
1951	39	20	780	400

$\Sigma y = 137$ $\Sigma X = 0$ $\Sigma Xy = 570$ $\Sigma X^2 = 1000$

$\Sigma y = na + b\Sigma X$

$137 = 5a + b(0)$

$137 = 5a$

PAGE NO :
DATE :

$a = \frac{137}{5} = 27.4$

$\Sigma Xy = a\Sigma X + b\Sigma X^2$

$570 = a(0) + b(1000)$

$570 = 1000b$

$b = \frac{570}{1000} = 0.57$

$y = a + bX$

$= 27.4 + 0.57(x - 1931)$

$= 27.4 + 0.57x - 1,100.67$

$y = 0.57x - 1,073.27$

(3) Find a straight line → mid term

x	1	2	3	4	5	6	7	8	9
y	3	4	6	5	10	9	10	12	11

Sol.

Let $X = x - 5$

$\therefore y = a + bX$

x	y	$X = x-5$	Xy	X^2
1	3	-4	-12	16
2	4	-3	-12	9
3	6	-2	-12	4
4	5	-1	-5	1
5	10	0	0	0
6	9	1	9	1
7	10	2	20	4
8	12	3	236	9
9	11	4	44	16
	$\Sigma y = 70$	$\Sigma X = 0$	$\Sigma Xy = 68$	$\Sigma X^2 = 60$

$\Sigma y = na + b\Sigma x$

$70 = 9a + b(0)$

$9a = 70$

PAGE NO :

DATE :

$a = \frac{70}{9} = 7.77$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$68 = a(0) + b(60)$

$68 = 60b$

$b = \frac{68}{60} = 1.13$

$y = a + bx$

$= 7.77 + 1.13(x - 5)$

$= 7.77 + 1.13x - 5.65$

$= 1.13x + 2.12$

(4) Find the straight line

x	20	5	10	15	10
y	24	15	17	22	12