B. Sc (Hons) CS (BHCS – 202) UNIT – 4
+
MCA – IInd SEM (MCA – 211) UNIT – 4

I Regression Analysis :–

Regression

- Linear
- Non-linear
- Multiple

Regression Lines :–

① Line of Regression of y on x :–

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

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

where r is the coefficient of correlation, σ_x & σ_y are the standard deviations of x and y series respt.

(2) Line of regression of x on y:

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

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

Example 1: calculate linear regression coefficients from the following data:

x:	1	2	3	4	5	6	7	8
y:	3	7	10	12	14	17	20	24

Solution: The linear regression coefficients are given by

$$b_{yx} = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma x^2 - (\Sigma x)^2} \quad \text{--- (1)}$$

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

Now, make table according to the requirement in formula.

x	y	x^2	y^2	xy
1	3	1	9	3
2	7	4	49	14
3	10	9	100	30
4	12	16	144	48
5	14	25	196	70
6	17	36	289	102
7	20	49	400	140
8	24	64	576	192
$\Sigma x = 36$	$\Sigma y = 107$	$\Sigma x^2 = 204$	$\Sigma y^2 = 1763$	$\Sigma xy = 599$

Here n = number of items given

i.e $n = 8$

$\Sigma x = 36$, $\Sigma y = 107$, $\Sigma x^2 = 204$, $\Sigma y^2 = 1763$

$\Sigma xy = 599$.

Now Putting these values in (1), we get.

Regression line y on x

$$b_{yx} = \frac{8(599)-(36)(107)}{8(204)-(36)^2}$$

On solving these we get

$$\frac{940}{336} = 2.79762$$

$$b_{xy} = \frac{8(599)-(36)(107)}{8(1763)-(107)^2}$$

Regression line x on y

$$= \frac{940}{2655} = 0.35405$$

Q2 Obtain the line of regression of y on x for the following data:

x:	1.53	1.78	2.60	2.95	3.42
y:	33.50	36.30	40.00	45.80	53.50

Solve it yourself & show me the result.

[II] Non-Linear Regression Analysis
(Second degree parabolic curve)

$$\boxed{y = a + bx + cx^2} \text{ of } y \text{ on } x.$$

$$\Sigma y = an + b\Sigma x + c\Sigma x^2 \quad -(1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad -(2)$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad -(3)$$

Equations (1), (2), & (3) are called normal equations. On solving these 3 equations we get the best values of a, b and c. On putting these values of a, b, and c in parabolic equations we therefore obtain a best fit of second degree parabolic curve of regression of y on x.

Example 1: Fit a second degree curve of regression of y on x to the following data:

x:	1	2	3	4
y:	6	11	18	27

Solution:

Let $y = a + bx + cx^2$ — (1) be a second degree curve of regression of y on x which is to be fitted to the given data.

The normal equations for (1) are

$$\Sigma y = an + b\Sigma x + c\Sigma x^2 \quad — (2)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad — (3)$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad — (4)$$

The table is given below:

x	y	x^2	x^3	x^4	xy	x^2y
1	6	1	1	1	6	6
2	11	4	8	16	22	44
3	18	9	27	81	54	152
4	27	16	64	256	108	432
$\Sigma x =$ 10	62	30	100	354	190	644

Here $n = 4$, $\Sigma x = 10$, $\Sigma y = 62$, $\Sigma x^2 = 30$, $\Sigma x^3 = 100$, $\Sigma xy = 190$, $\Sigma x^2 y = 644$

Putting these values in (2), (3) and (4) we get

$$62 = 4a + 10b + 30c \quad -(5)$$
$$190 = 10a + 30b + 100c \quad -(6)$$
$$644 = 30a + 100b + 354c \quad -(7)$$

Solving (5) (6) and (7) for a, b, c values and we get

$a = 3$, $b = 2$, $c = 1$

Hence the required Second degree curve of regression of y on x is

$$\boxed{y = 3 + 2x + x^2}$$

Example 2: Employ the method of least Square to fit a parabola $y = a + bx + cx^2$ in the following data:

x:	-1	0	0	1
y:	2	0	1	2

Solve it yourself & Show me the result.

[III] Multiple Linear Regression Analysis :-

$$\boxed{y = a + bx + cz} \quad \text{--- (1)}$$

$$\Sigma y = an + b\Sigma x + c\Sigma z \quad \text{--- (2)}$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma xz \quad \text{--- (3)}$$

$$\Sigma zy = a\Sigma z + b\Sigma zx + c\Sigma z^2 \quad \text{--- (4)}$$

Solving (2), (3) and (4) are called normal equations. By solving these equations we get the values of a, b and c

Q Obtain a regression plane by using multiple linear given below:

x:	1	2	3	4
y:	12	18	24	30
z:	0	1	2	3

Solution: Let $y = a + bx + cz$ — (1)
be a regression plane which is to be fitted to the given data:

The normal equations for (1) are

$$\Sigma y = an + b\Sigma x + c\Sigma z \quad \text{— (2)}$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma zx \quad \text{— (3)}$$
$$\Sigma zy = a\Sigma z + b\Sigma zx + c\Sigma z^2 \quad \text{— (4)}$$

The table is given below:

	x	y	z	x^2	z^2	xy	zy	zx
	1	12	0	1	0	12	0	0
	2	18	1	4	1	36	18	2
	3	24	2	9	4	72	48	6
	4	30	3	16	9	120	90	12
$\Sigma x =$	10	6	84	30	14	240	156	20

Putting these values in (2), (3) and (4) we get

$$84 = 4a + 10b + 6c \quad \text{---} \quad (5)$$
$$240 = 10a + 30b + 20c \quad - \quad (6)$$
$$156 = 6a + 20b + 14c \quad - \quad (7)$$

Solving (5), (6) and (7), we get

$a = 6, \quad b = 6, \quad c = 0$

Hence the required plane of regression of best fit is

$$\boxed{y = \cancel{6x+} \; 6 + 6x + z}$$

Q Fit a second degree parabola in the following data.