24/7/23

MODULE - 4

CURVE FITTING AND STATISTICAL METHOD

① Fitting of a Straight line :

$$y = ax + b \qquad (\text{or} \quad y = a + bx)$$

Normal equations are :

- $\sum y = a\sum x + nb$

(* x)

- $\sum xy = a\sum x^2 + b\sum x$.

② Fitting of a Second degree parabola :

$$y = ax^2 + bx + c$$
$$= a + bx + cx^2 .$$

Normal equations are ;

- $\sum y = a\sum x^2 + b\sum x + nc$

(* x)

- $\sum xy = a\sum x^3 + b\sum x^2 + c\sum x$

(* x)

- $\sum x^2 y = a\sum x^4 + b\sum x^3 + c\sum x^2$.

③ Fitting of a curve of the form

$$y = ax^b$$

$\Rightarrow \log y = \log(ax^b)$

$\log y = \log a + \log x^b$

$\log y = \log a + b\log x$

put $\log y = Y$, $\log a = A$, $\log x = X$

$Y = A + bX$

Normal equations :

$\sum Y = nA + b\sum X$

(* X)

$\sum XY = A\sum X + b\sum X^2$

Curve fitting method is also called as method of least squares.

1. Find the equation of the best fitting straight line for the following data and hence estimate the value of the dependent variable corresponding to the value 30 of the independent variable,

x	5	10	15	20	25
y	16	19	23	26	30

Ans: We have,

straight line equation as,

$$y = ax + b$$

Normal equations:

$$\Sigma y = a\Sigma x + nb \longrightarrow (1)$$

$(\times x)$

$$\Sigma xy = a\Sigma x^2 + b\Sigma x \longrightarrow (2)$$

x	y	xy	x^2
5	16	80	25
10	19	190	100
15	23	345	225
20	26	520	400
25	30	750	625

$\Rightarrow \Sigma y = \underline{114}$, $\Sigma x = \underline{75}$

$\Sigma xy = \underline{1885}$

$\Sigma x^2 = \underline{1375}$

$\Rightarrow (1) \Rightarrow 114 = 75a + 5b$

$(2) \longrightarrow 1885 = 1375a + 75b$

$\Longrightarrow a = \underline{0.7}$ $b = \underline{12.3}$

$\therefore y = ax + b$

$y = \underline{\underline{0.7x + 12.3}}$

Here, $y \rightarrow$ dependent variable &

$x \rightarrow$ independent "

Given is independent, $X = 30$.

$\Rightarrow y = 0.7(30) + 12.3$

$= 21 + 12.3$

$y = \underline{\underline{33.3}}$

2\. A simple supported beam carries a concentrated load 'P' at its mid point. Corresponding to various values of P the maximum deflection is 'y' is given in the following table.

P	100	120	140	160	180	200
y	0.45	0.55	0.60	0.70	0.80	0.85

Find the of the form,

$Y = a + bP$ and hence estimate Y when P is 150.

Ans: Given $y = a + bP$

Normal equations: $\Sigma y = na + b\Sigma P$ —— (1)

$\Sigma yP = a\Sigma P + b\Sigma P^2$ —— (2)

P	y	yP	P^2
100	0.45	45	10000
120	0.55	66	14400
140	0.60	84	19600
160	0.70	112	25600
180	0.80	144	32400
200	0.85	170	40000
$\Sigma P = \underline{\underline{900}}$	$\Sigma y = \underline{\underline{3.95}}$	$\Sigma yP = \underline{\underline{621}}$	$\Sigma P^2 = \underline{\underline{142000}}$

① ⇒ $3.95 = 6a + b(900)$

② ⇒ $621 = 900a + b(142000)$

$\Rightarrow a = \underline{0.04761}$ $\quad b = 4.071 \times 10^{-3}$

$b = \underline{0.00407}$

$= \underline{0.0041}$

$\therefore \; y = a + bP$

$y = 0.0476 + 0.0041 P$

then, when $P = 150$,

$y = 0.0476 + 0.0041(150)$

$y = \underline{0.6626}$

25/07/23

3. Fit a straight line to the following data:

year	1961	1971	1981	1991	2001
production (in tons)(y)	8	10	12	10	16

Also find the expected production in the year 2006.

Ans:- ($\because$ values are large in 'year')

So $X = x - \text{mid no.}$

$X = x - 1981$

⇒

X	-20	-10	0	10	20
y	8	10	12	10	16

Now take $y = aX + b$

Normal eqns: $\Sigma y = a\Sigma X + nb$ — ① $\quad \Sigma Xy = a\Sigma X^2 + b\Sigma X$

X	y	X^2	Xy
-20	8	400	-160
-10	10	100	-100
0	12	0	0
10	10	100	100
20	16	400	320

$\Sigma X = 0$

$\Sigma y = 56$

$\Sigma X^2 = 1000$

$\Sigma Xy = 160.$

$(1) \Rightarrow \quad 56 = 0 + 5b \Rightarrow \quad b = \frac{56}{5} = \underline{11.2}$.

$(2) \Rightarrow \quad 160 = 1000a + 0$

$a = \underline{0.16}$

$y = aX + b$

$\Rightarrow \quad y = 0.16X + 11.2$

$\Rightarrow \quad y = \underline{0.16(x - 1981)} + 11.2$

$\Rightarrow \quad y = 0.16x - 305.76$.

To find for 2006 $\Rightarrow x = 2006$

$\therefore$ y (production) $= 0.16(2006) - 305.76$

$= \underline{15.2}$ ~~tons~~.

$\therefore$ The production in 2006 is ~~2~~ $\underline{15.2}$ tons.

4. An experiment of lifetime t of cutting tools at different cutting speeds v (units) are given below:

speed (v)	350	400	500	600
life (t)	61	26	7	2.6

Fit a relation of the form,

$v = at^b$.

Ans Given $v = at^b$ (take log $\because$ b is in power).

$\log v = \log(at^b)$

$= \log a + \log t^b$

$\log V = \log a + b \log t$

put $\log V = V$, $\log a = A$, $\log t = T$.

$\because V \rightarrow$ variable this also; A constant; T variable

$\Rightarrow \quad V = A + bT$

Normal equations: $\sum V = nA + b\sum T$

$\sum VT = A\sum T + b\sum T^2$

$V = \log x$	$T = \log T$	VT	T^2
5.8579	4.1109	24.0812	16.8995
5.9914	3.2582	19.5188	10.6146
6.2146	1.9459	12.08624	3.7869
6.3969	0.9555	6.10822	0.9128

$\Sigma V = 24.4608$ $\Sigma T = 10.2705$ $\Sigma VT = 61.8091$ $\Sigma T^2 = 32.2798$

(1) $\Rightarrow$ $24.4609 = 4A + 10.2704b$

$61.8091 = 10.2704A + 32.2142b$

$A = 6.5531$

$b = -0.1705$

$V = A + bT$

$V = 6.5531 ~~A~~ - 0.1705T$

we have $\log a = A$

$~~\log~~ a = e^A$

$a = e^{6.5531}$

$= 701.4151$

$\therefore$ $x = at^b$

$x = (701.4151)t^{-0.1705}$

$x = (701.4151)t^{-0.1706}$

5. Fit a least square geometric curve, $y = ax^b$ for the following data

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

Ans: Given $y = ax^b$

$\Rightarrow$ $\log y = \log a + b \log x$

Put $\log y = Y$, $\log a = A$, $\log x = X$

$y = A + bX$.

Normal equations: $\sum_i y = nA + b\sum_i x$ —— (1)

$\sum_i xy = A\sum_i x + b\sum_i x^2$ —— (2)

$X = \log x$	$Y = \log y$	X^2	XY
0	-0.6931	0	0
0.6931	0.6931	0.4803	0.4803
1.0986	1.5040	1.2069	1.6522
1.3862	2.0794	1.9215	2.8824
1.6094	2.5257	2.5901	4.0648
$\sum_i X =$ 4.7873	$\sum_i Y =$ 6.1091	$\sum_i X^2 =$ 6.1988	$\sum_i XY =$ 9.0797

(1) ⇒ $6.1091 = 5A + b(4.7873)$

(2) ⇒ $9.0797 = 4.7873A + 6.1988b$.

$A = -0.6932$

$b = 2.0001$

we $\log a = A$

$a = e^A = e^{-0.6932}$

$= 0.4999$

∴ $y = ax^b$

⇒ $y = (0.4999)x^{(2.0001)}$

6. Fit a 2nd degree parabola in the least square sense for the following data, and hence estimate y at $x = 6$.

x	1	2	3	4	5
y	10	12	13	16	19

Ans: We have, 2nd degree parabolic equation,

$$y = ax^2 + bx + c$$

Normal equations:

$$\Sigma y = a\Sigma x^2 + b\Sigma x + nc \quad \text{--- (1)}$$

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

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

x	y	xy	x^2y	x^2	x^3	x^4
1	10	10	10	1	1	1
2	12	24	48	4	8	16
3	13	39	117	9	27	81
4	16	64	256	16	64	256
5	19	95	475	25	125	625
$\Sigma x = 15$	$\Sigma y = 70$	$\Sigma xy = 232$	Σx^2y 906	$\Sigma x^2 = 55$	$\Sigma x^3 = 225$	$\Sigma x^4 = 979$

(1) $\Rightarrow$ $70 = 55a + 15b + 5c$

(2) $\Rightarrow$ $232 = 225a + 55b + 15c$

(3) $\Rightarrow$ $906 = 979a + 225b + 55c$

$\Rightarrow a = 0.2857$

$b = 0.4857$

$c = 9.4$

$\therefore$ Eqn:

$$y = 0.2857x^2 + 0.4857x + 9.4$$

At $x = 6$,

$$y = 0.2857(6)^2 + 0.4857(6) + 9.4$$
$$y = \underline{22.5994}$$

26/07/23

7. Fit a parabola $y = a + bx + cx^2$ for the data,

x :	0	1	2	3	4
y :	1	1.8	1.3	2.5	6.3

Ans: Given $y = a + bx + cx^2$.

Normal equations:

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad \text{—— (1)}$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad \text{—— (2)}$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad \text{—— (2)}$$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	1.8	1	1	1	1	1
2	1.3	4	8	16	~~3.2~~ 2.6	5.2
3	2.5	9	27	81	7.5	22.5
4	6.3	16	64	256	25.2	100.8
$\Sigma x =$ 10	$\Sigma y =$ 12.9	$\Sigma x^2 =$ 30	$\Sigma x^3 =$ 100	$\Sigma x^4 =$ 354	$\Sigma xy =$ 37.1	$\Sigma x^2y =$ 130.3

(1) $\Rightarrow$ $12.9 = 5a + 10b + 30c$

(2) $\Rightarrow$ $37.1 = 10a + 30b + 100c$

(3) $\Rightarrow$ $130.3 = 30a + 100b + 354c$.

$\therefore$ $a = \underline{1.42}$

$b = \underline{-1.07}$

$c = \underline{0.55}$.

$\therefore$ $y = a + bx + cx^2$

$\Rightarrow$ $y = \underline{1.42 - 1.07x + 0.55x^2}$

8. Fit a curve of the form $y = ax^b$ for the data

x	1	2	3	4	5	6
y	2.98	4.26	5.21	6.1	6.8	7.5

Ans

Given $y = ax^b$

$\log y = \log(ax^b)$

$\log y = \log a + b \log x$

put $\log y = Y$, $\log a = A$, $\log x = X$

$\Rightarrow Y = A + bX$

$\therefore$ Normal equations:

$\sum Y = nA + b\sum X$ ——— (1)

$\sum XY = A\sum X + b\sum X^2$ ——— (2)

$X = \log x$	$Y = \log y$	XY	X^2
0	1.0919	0	0
0.6931	1.4492	1.0044	0.4803
1.0986	1.6505	1.8132	1.2069
1.3862	1.8082	2.5065	1.9215
1.6094	1.9169	3.0850	2.5901
1.7917	2.0149	3.6100	3.2101
$\sum X =$ ~~4.5696~~ 6.579	$\sum Y =$ 9.9316	$\sum XY =$ 12.0191	$\sum X^2 =$ 9.4089

(1) $\Rightarrow$ $9.9316 = 6A + 6.579b$

(2) $\rightarrow$ $12.0191 = 6.579A + 9.4089b$

$\Rightarrow A = 1.0912$

$b = 0.5143$

$\log a = A$

$a = e^A = e^{1.0912}$

$a = 2.9778$

$\therefore\ y = a x^b$

$y = \underline{(2.9778)\, x^{(0.5143)}}$

28/7/23

Correlation and Regression :

The numerical measure of correlation between 2 variables 'x' and 'y' is known as Karl Pearson's coefficient of correlation denoted by 'r'

① Correlation coefficient :

(1.) $\boxed{r = \dfrac{\Sigma XY}{n\,\sigma_x \sigma_y}}$ where

$X = x - \bar{x}$

$Y = y - \bar{y}$.

$\sigma_x^2 = \dfrac{\Sigma x_i^2}{n} - \bar{x}^2$ $|||^{ly}$, $\sigma_y^2 = \dfrac{\Sigma_i y_i^2}{n} - \bar{y}^2$

$\underset{\text{(mean)}}{\bar{x}} = \dfrac{\Sigma x}{n}$ or $\bar{x} = \dfrac{\Sigma fx}{\Sigma f}$. $|||^{ly}$ $\bar{y} = \dfrac{\Sigma y}{n}$

(2) $r = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}}$

(3.) $r = \dfrac{\sigma_x^2 + \sigma_y^2 - \sigma_{(x-y)}^2}{2\sigma_x \sigma_y}$.

Equation of regression lines :

y on x :

$y - \bar{y} = r\left(\dfrac{\sigma_y}{\sigma_x}\right)(x - \bar{x})$ or $Y = \dfrac{\Sigma_i XY}{\Sigma_i X^2} X$

x on y :

$x - \bar{x} = r\left(\dfrac{\sigma_x}{\sigma_y}\right)(y - \bar{y})$ or $X = \dfrac{\Sigma_i XY}{\Sigma Y^2}(\cancel{X})$

9. Compute the coefficient of correlation and the equations of lines of regression for the data 1, 2, 3, 4, 5, 6, 7 and y values, → 9, 8, 10, 12, 11, 13, 14.

Ans:

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

(first coefficient then regression lines)

we have correlation,

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{(x-y)}^2}{2\sigma_x \sigma_y} \quad \text{——— (1)}$$

$$\sigma_x^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 \quad \to (2) \qquad \sigma_y^2 = \frac{\sum y_i^2}{n} - \bar{y}^2 \quad \to (3)$$

$$\sigma_{(x-y)}^2 = \frac{\sum (x_i - y_i)^2}{n} - (\overline{x-y})^2 \quad \text{——— (4)}$$

$$\bar{x} = \frac{\sum x}{n} \qquad \bar{y} = \frac{\sum y}{n} \qquad (\overline{x-y}) = \frac{\sum (x-y)}{n}.$$

x	y	x^2	y^2	$(x-y)$	$(x-y)^2$
1	9	1	81	-8	64
2	8	4	64	-6	36
3	10	9	100	-7	49
4	12	16	144	-8	64
5	11	25	121	-6	36
6	13	36	169	-7	49
7	14	49	196	-7	49
$\sum x =$ 28	$\sum y =$ 77	$\sum x^2 =$ 140	$\sum y^2$ = 875	$\sum (x-y) =$ -49	$\sum (x-y)^2 =$ 347

$\therefore \bar{x} = \frac{\Sigma x}{n} = \frac{28}{7} = \underline{4}$

12, $\bar{y} = \frac{\Sigma y}{n} = \frac{77}{7} = \underline{11}$.

$\overline{x-y} = \frac{\Sigma (x-y)}{n} = \frac{-49}{7} = \underline{-7}$.

$(2) \Rightarrow \sigma_x^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2$

$= \frac{140}{7} - (4)^2 = 20 - (16) = \underline{4}$

$(3) \Rightarrow \sigma_y^2 = \frac{875}{7} - (11)^2 = \cancel{4}\ 125 - 121$

$= \underline{4}$.

$(4) \Rightarrow \sigma^2_{(x-y)} = \frac{347}{7} - (-7)^2 = 49.5714 - 49$

$= \underline{0.5714}$

$\Rightarrow \sigma_x = \sqrt{4} = 2, \quad \sigma_y = \sqrt{4} = 2$

$\therefore (1) \Rightarrow$

$r = \frac{4 + 4 - 0.5714}{2 \times 2 \times 2}$

$= \underline{0.9285}$

Regression lines:

y on x:

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

$y - 11 = 0.9285\left(\frac{2}{2}\right)(x - 4)$

$y - 11 = 0.9285x - 3.7143$

$$y = 0.9285x + 7.2857$$

x on y:

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

$$x - 4 = 0.9285\left(\frac{\cancel{2}}{\cancel{2}}\right)(y - 11)$$

$$x - 4 = 0.9285y - 10.2135$$

$$x = 0.9285y - 6.2135$$

10. Obtain the line of regression and hence find the coefficient of correlation for the data,

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

Ans: line of regression:

y on x: $\quad Y = \frac{\sum XY}{\sum X^2}(X)$ —— (1)

x on y: $\quad X = \frac{\sum XY}{\sum Y^2}(Y)$ —— (2)

$Y = y - \bar{y} \qquad X = x - \bar{x} \qquad \bar{x} = \frac{\sum x}{n} \quad \bar{y} = \frac{\sum y}{n}$.

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	X^2	Y^2	XY
1	9	-3	-2	9	4	6
2	8	-2	-3	4	9	6
3	10	-1	-1	1	1	1
4	12	0	1	0	1	0
5	11	1	0	1	0	0
6	13	2	2	4	4	4
7	14	3	3	9	9	9
$\sum x =$ 28	$\sum y =$ 77			$\sum X^2 =$ 28	$\sum Y^2 =$ 28	$\sum XY =$ 26

/ /

$\bar{x} = \frac{28}{7} = \underline{\underline{4}}$ $\bar{y} = \frac{77}{7} = \underline{\underline{11}}$

y on x : ① ⇒ $y = \frac{26}{28}(x) = 0.9285(x)$

$\Rightarrow y - \bar{y} = 0.9285(x - \bar{x})$

$y - 11 = 0.9285(x-4) = 0.9285x - 3.7142$

$y = \underline{\underline{0.9285x + 7.2857}}$

x on y :

② $x = \frac{26}{28}(y) = 0.9285(y)$

$(x - \bar{x}) = 0.9285(y - \bar{y})$

$(x-4) = 0.9285(y-11) = 0.9285y - 10.2135$

$x = \underline{\underline{0.9285 - 6.2135}}$

Correlation coefficient :

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{Y^2}}$$

$$= \frac{26}{\sqrt{28}\sqrt{28}}$$

$$= \frac{26}{28}$$

$r = \underline{\underline{0.9285}}$

11. Obtain the line of regression & hence find the coefficient of correlation for data

x	1	3	4	2	5	8	9	10	13	15
y	8	6	10	8	12	16	16	10	32	32

Ans

Regression line :

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

y on x :

here use, $Y = \frac{\Sigma XY}{\Sigma X^2}(X)$ ——— ①

x on y :

$X = \frac{\Sigma XY}{\Sigma Y^2}(Y)$ ——— ②

$Y = (y - \bar{y})$; $X = (x - \bar{x})$

$\bar{x} = \frac{\Sigma x}{n}$ $\bar{y} = \frac{\Sigma y}{n}$

$\bar{x} = \frac{70}{n} = \frac{70}{10} = 7$

$\bar{y} = \frac{150}{10} = 15$

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	XY	X^2	Y^2
1	8	-6	-7	42	36	49
3	6	-4	-9	36	16	81
4	10	-3	-5	15	9	25
2	8	-5	-7	35	25	49
5	12	-2	-3	6	4	9
8	16	1	1	1	1	1
9	16	2	1	2	4	1
10	10	3	-5	-15	9	25
13	32	6	17	102	36	289
15	32	8	17	136	64	289
$\Sigma x = 70$	$\Sigma y = 150$			$\Sigma XY = 360$	$\Sigma X^2 = 204$	$\Sigma Y^2 = 818$

$$\therefore \bar{x} = \frac{70}{n} = \frac{70}{10} = 7$$

$$\bar{y} = \frac{150}{10} = 15$$

y on x :

$$① \Rightarrow \quad Y = \frac{360}{204}(X) = \frac{30}{17}X$$

$$Y = 1.7647X \quad —— ③$$

x on y :

$$② \Rightarrow \quad X = \frac{360}{818}(X) = \frac{360}{818}(\cancel{Y})$$

$$X = 0.44\cancel{X} \quad —— ④$$

y on x :

$$③ \Rightarrow \quad y-\bar{y} = 1.7647(x-\bar{x})$$

$$y - 15 = 1.7647(x-7) = 1.7647x - 12.3529$$

$$y = 1.7647x + 2.6471$$

x on y :

$$④ \Rightarrow \quad x-\bar{x} = 0.44(y-\bar{y})$$

$$x - 7 = 0.44(y-15) = 0.44y - 6.6$$

$$x = 0.44y + 0.4$$

Coefficient of correlation:

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

$$= \frac{360}{\sqrt{204}\sqrt{818}} =$$

$$r = 0.88127$$

12. Find the correlation, coefficient and the equation of regression for the following values of x & y.

x	1	2	3	4	5
y	2	5	3	8	7

Ans correlation coefficient,

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma^2_{(x-y)}}{2\sigma_x \sigma_y} \quad \text{---} (1)$$

$$\sigma_x^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2 \quad \text{---} (2) \qquad \sigma_y^2 = \frac{\sum y_i^2}{n} - (\bar{y})^2 \quad \text{---} (3)$$

$$\sigma^2_{(x-y)} = \frac{\sum (x_i - y_i)^2}{n} - (\overline{x-y})^2 \quad \text{---} (4)$$

x	y	x^2	y^2	$(x-y)$	$(x-y)^2$
1	2	1	4	-1	1
2	5	4	25	-3	9
3	3	9	9	0	0
4	8	16	64	-4	16
5	7	25	49	-2	4
$\sum x =$ 15	$\sum y =$ 25	$\sum x^2 =$ 55	$\sum y^2 =$ 151	$\sum (x-y) =$ -10	$\sum (x-y)^2 =$ 30

$\bar{x} = \frac{\sum x}{n} = \frac{15}{5}$ $\qquad$ $\sum y = \frac{\sum y}{n} = \frac{25}{5}$

$= \underline{\underline{3}}$ $\qquad$ $= \underline{\underline{5}}$

$(\overline{x-y}) = \frac{\sum (x-y)}{n} = \frac{-10}{5} = \underline{\underline{-2}}$

(2) $\Rightarrow \sigma_x^2 = \frac{55}{5} - (3)^2 = 11 - 9 = \underline{\underline{2}}$

(3) $\Rightarrow \sigma_y^2 = \frac{151}{5} - (5)^2 = \underline{\underline{5.2}}$.

(4) $\Rightarrow \sigma^2_{(x-y)} = \frac{30}{5} - (-2)^2 = \underline{\underline{2}}$.

$\sigma_x = \sqrt{2} = \underline{1.4142}$ $\qquad$ $\sigma_y = \sqrt{5.2} = \underline{\underline{2.2803}}$

$r = \frac{\cancel{2} + 5.2 - \cancel{2}}{2(1.4142)(2.2803)}$

$r = \underline{\underline{0.8062}}$

Regression line:

y on x:

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

$(y - 5) = 0.8062\left(\frac{2.2803}{1.4142}\right)(x - 3)$

$= \cancel{0.8062\,(0.62018)\,(x-3)}$

$\cancel{y - 5 = 0.4999\,(x-3)} = 0.4999x - 1.4997$

$\underline{y = 0.4999x - 3.5003}$

$y = \underline{1.2999 - 1.1003}$

x on y:

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

$$(x-3) = 0.8062\left(\frac{1.4142}{2.2803}\right)(y-5)$$

$$x-3 = 0.4999y - 2.4999$$

$$x = \underline{0.4999y + 0.50004}$$

$$(x-3) = 0.8062\left(\frac{1.4142}{2.2803}\right)(y-5)$$

$$x-3 = 0.4999y - 2.4999$$

$$x = \underline{0.4999y + 0.50004}$$

1/7/23

13. The following data gives the age of husbands (x) and the age of wives (y) in years. Form the 2 regression lines, and calculate the age of husband corresponding to 16 years age of wife.

x	36	23	27	28	28	29	30	31	33	35
y	29	18	20	22	27	21	29	27	29	28

Ans

Regression lines:

y on x: $$Y = \frac{\Sigma XY}{\Sigma X^2}(X) \quad \text{---} \quad ①$$

x on y: $$X = \frac{\Sigma XY}{\Sigma Y^2}(Y) \quad \text{---} \quad ②$$

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

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	X^2	Y^2	XY
36	29	6	4	36	16	24
23	18	-7	-7	49	49	49
27	20	-3	-5	9	25	15
28	22	-2	-3	4	9	6
28	27	-2	2	4	4	-4
29	21	-1	-4	1	16	4
30	29	0	4	0	16	0
31	27	1	2	1	4	2
33	29	3	4	9	16	12
35	28	5	3	25	9	15
$\Sigma x =$ 300	$\Sigma y =$ 250			$\Sigma X^2 =$ 138	ΣY^2 = 164	ΣXY = 123

$\bar{x} = \frac{\Sigma x}{n} = \frac{300}{10} = \underline{\underline{30}}$ $\bar{y} = \frac{\Sigma y}{n} = \frac{250}{n} = \underline{\underline{25}}$.

regression lines:

y on x: $y = \frac{123}{138}(x) = \frac{41}{46}(x)$

$$y = 0.8913(x)$$
$$y - \bar{y} = 0.8913(x - \bar{x})$$
$$y - 25 = 0.8913(x - 30)$$
$$y - 25 = 0.8913x - 26.7391$$
$$y = \underline{\underline{0.8913x}} - 1.73913$$

x on y: $x = \frac{123}{164}(y) = \frac{3}{4}(y)$

$$x = 0.75(y)$$
$$x - \bar{x} = 0.75(y - \bar{y})$$
$$x - 30 = 0.75(y - 25)$$
$$x - 30 = 0.75y - 18.75$$
$$x = \underline{\underline{0.75y}} + 11.25$$

Also;

given $y = 16$ years.

$$x = 0.75(16) + 11.25$$
$$x = \underline{23.25} \text{ years}$$
$$\approx \underline{\underline{23}} \text{ years}$$

∴ When age of wife is 16, the husband's age is 23 years.

Thus husband's age is 23 years corresponding to wife's age of 16 years.

14. Find the correlation coefficient and the equation of lines of regression for the following values of x and y.

x	1	2	3	4	5.
y	2	5	3	8	7

14. Find the correlation coefficient for the following 2 groups.

a	92	89	87	86	83	77	71	63	53	50
b	86	83	91	77	68	85	52	82	37	57

Ans: Correlation coefficient

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

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

$\bar{x} = \frac{\Sigma x}{n} = \frac{751}{10} = 75.1$ $\qquad$ $\bar{y} = \frac{\Sigma y}{n} = \frac{718}{10} = 71.8$

a=x	b=y	$X = x - \bar{x}$	$y = y - \bar{y}$	X^2	y^2	xy
92	86	16.9	14.2	285.61	201.64	239.98
89	83	13.9	11.2	193.21	125.44	155.68
87	91	11.9	19.2	141.61	368.64	228.48
86	77	10.9	5.2	118.81	27.04	56.68
83	68	7.9	-3.8	62.41	14.64	-30.02
77	85	1.9	13.2	3.61	174.24	25.08
71	52	-4.1	-19.8	16.81	392.04	81.18
63	82	-12.1	10.2	146.41	104.04	-123.42
53	37	-22.1	-34.8	488.41	1211.04	769.08
50	57	-25.1	-14.8	630.01	219.04	371.48
Σx = 751	Σy = 718			ΣX^2 = 2086.9	Σy^2 = 2837.6	Σxy = 1774.2

$$\therefore \; r = \frac{1774.2}{\sqrt{2086.9}\sqrt{2837.6}} = 0.7290.$$

$r = 0.7290$

$r = \pm\sqrt{(\text{coef } x)(\text{coef } y)}$

• +ve - ans is +ve • -ve ans ⇒ +ve

Rank Correlation:

The coefficient of correlation in respects of the ranks of some 2 characteristics of ~~an~~ individuals of an observation is called Rank Correlation Coefficient, denoted by ρ

i.) $$\boxed{\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}}$$

where $d = x - y$

ii.) Repeated Rank

$$\boxed{\rho = 1 - \frac{6\left[\sum d^2 + \frac{m(m^2-1)}{12} + \dots\right]}{n(n^2-1)}}$$

2/8/23

15. 10 competitors in music competition are ranked by 3 judges A, B, C in the following order. Use the rank correlation coefficient to decide which pair of judges have the nearest approach to a common taste of music

A	1	6	5	10	3	2	4	9	7	8
B	3	5	8	4	7	10	2	1	6	9
C	6	4	9	8	1	2	3	10	5	7

Ans

Since rank is not repeating, the rank correlation coefficient:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

where $d = x - y$

and $n = 10$

(10 participants)

$A\ \rho\ B = \rho_{AB} \rightarrow (A-B)^2$

$B\ \rho\ C = \rho_{BC} \rightarrow (B-C)^2$

$C\ \rho\ A = \rho_{CA} \rightarrow (C-A)^2$

A	B	C	$(A-B)^2$	$(B-C)^2$	$(A-C)^2$
1	3	6	4	9	25
6	5	4	1	1	4
5	8	9	9	1	16
10	4	8	36	16	4
3	7	1	16	36	4
2	10	2	64	64	0
4	2	3	4	1	1
9	1	10	64	81	1
7	6	5	1	1	4
8	9	7	1	4	1
			$\Sigma(A-B)^2 = 200$	$\Sigma(B-C)^2 = 214$	$\Sigma(A-C)^2 = 60$

$$\therefore \rho_{AB} = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6\ \Sigma(A-B)^2}{10(10^2-1)}$$

$$= 1 - \frac{6(200)}{10 \times 99}$$

$$= +0.2121$$

$$\rho_{BC} = 1 - \frac{6\Sigma (B-C)^2}{n(n^2-1)} = 1 - \frac{6\ (214)}{10(10^2-1)}$$

$$= 1 - \frac{6(214)}{10(99)}$$

$$= -0.2969$$

$$\rho_{CA} = 1 - \frac{6\ \Sigma(C-A)^2}{n(n^2-1)} = 1 - \frac{6 \times 60}{10 \times 99}$$

$$= 0.6363$$

It may be observed that row AB and row BC are negative, which means that their tastes (A and B ; B and C) are opposite. But row CA is positive and is near to 1, thus we conclude that judges C and A have the nearest approach to a common taste of music.

16. Compute the Rank correlation coefficient for the following

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

Ans: (no.s are high ⇒ then are marks; hence convert to rank)

x	Rank for x (X)	y	Rank for y (Y)	$d^2 = (X - Y)^2$
68	4	62	5	1
64	6	58	7	1
75	2.5	68	3.5	1
50	9	45	10	1
64	6	81	1	25
80	1	60	6	25
75	2.5	68	3.5	1
40	10	48	9	1
55	8	50	8	0
64	6	70	2	16
				$\Sigma (X-Y)^2 = 72$

rank rank

$\frac{2+3}{2} = 2.5$ $\quad \frac{3+4}{2} = \frac{7}{2} = 3.5$

$\frac{5+6+7}{3} = \frac{18}{3} = 6$

- In x, 75 repeats 2 times (m_1);

$\therefore$ Rank $= \frac{2+3}{2} = 2.5$

64 repeats 3 times (m_2);

Rank $= \frac{5+6+7}{3} = 6$

- In y, 68 repeats 2 times (m_3)

Rank $= \frac{3+4}{2} = 3.5$

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

We have $m = 2, 3, 2$.

$$\rho = 1 - \frac{6\left[72 + \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12}\right]}{10(10^2-1)}$$

$$= 1 - \frac{6\left[72 + \frac{1}{2} + 2 + \frac{1}{2}\right]}{10(99)}$$

$$= 1 - \frac{5}{11}$$

$$= \underline{0.5454}.$$

$\therefore$ The rank correlation coefficient is 0.5454

4/8/

17. 10 students got the following % of marks in 2 subjects x and y, Compute their rank correlation coefficient.

Marks in x	78	36	98	25	75	82	90	62	65	39
Marks in y	84	51	91	60	68	62	86	58	53	47

Ans

x	Rank of x (X)	y	Rank of y (Y)	$d^2 = (X-Y)^2$
78	4	84	3	1
36	9	51	9	0
98	1	91	1	0
25	10	60	6	16
75	5	68	4	1
82	3	62	5	4
90	2	86	2	0
62	7	58	7	0
65	6	53	8	4
39	8	47	10	4

$\Sigma d^2 = 30$

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

$= 1 - \frac{6(30)}{10(99)}$

$= \underline{0.81818}$

4/8/20

Linear Programming Problems (LPP)

The optimization (of maximization or minimization) of the objective function of Z, subject to the constraints, $AX \le B$ or $AX \ge B$ is the mathematical formulation of a LPP

A set of real values $X = (x_1, x_2, x_3 \ldots\ldots x_n)$, which satisfies the constraints $AX \le B$ or $AX \ge B$ is called solution.

A set of real values 'x_i' which satisfies the constraints and also satisfies non negative constraints $x_i \ge 0$ is called feasible solution.

A set of real values 'x_i' which satisfies the constraints along with non negative restrictions and optimizes the objective function is called Optimal solution.

Method 1:

Graphical method of solving an LPP

(using graph.

18 Maximize $z = x + 1.5y$, given $x \geq 0$, $y \geq 0$. subject to the constraints $x + 2y \leq 160$, $3x + 2y \leq 240$ by graphical method.

Ans: Consider the constraints

$x + 2y = 160$

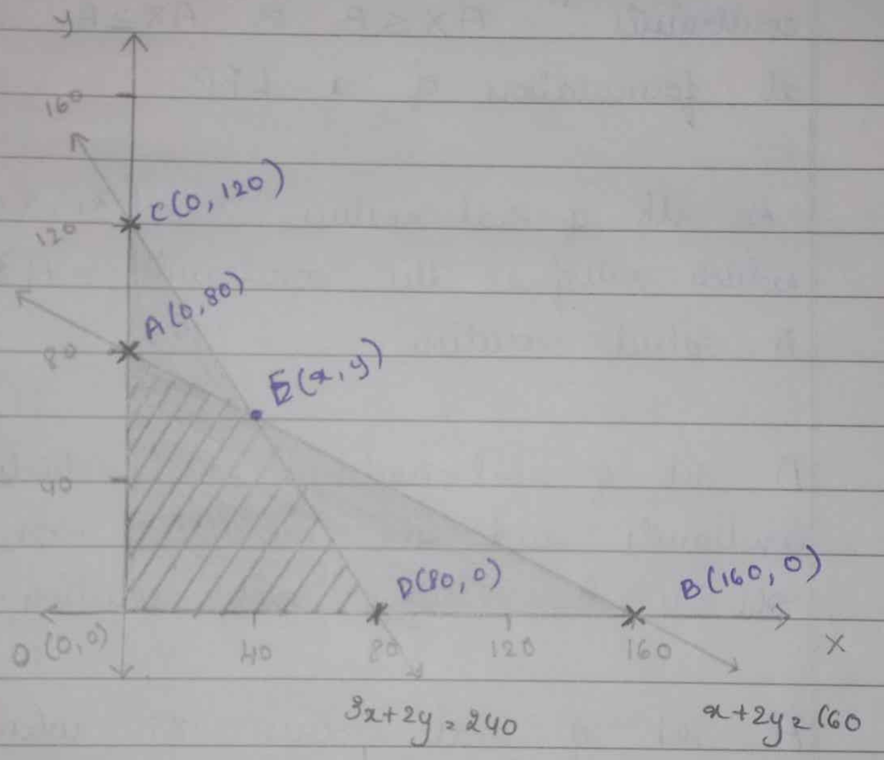
x	0	160
y	80	0

Now we have the points A (0, 80), B (160, 0)

$3x + 2y = 240$

x	0	80.
y	120	0

Now we have the points C (0, 120), D (80, 0)



for $x = 0$, $y = 0$,
$0 + 0 \leq 160$ ✓ → true
Consider region below

for $x = 0$, $y = 0$
$0 + 0 \leq 240$ ✓ → true
Consider the region below.

The corners of feasible region are,
A (0, 80), O (0, 0), D (80, 0), E(x, y)
find E (x, y)
E is intersection of 2 lines.
$3x + 2y = 240$ & $x + 2y = 160$

$x = 40$, $y = 60$.

$\therefore$ $E(x, y) = (40, 60)$.

	$z = x + 1.5y$
A (0, 80)	120
O (0, 0)	0
D (80, 0)	80
E (40, 60)	$\boxed{130}$

Thus, z has maximum value at $x = 40$, and $y = 60$ and the maximum value is 130.

19. Minimize $z = 5x + 4y$, subject to the constraints $x + 2y \geq 10$ and $x + y \geq 8$, $2x + y \geq 12$, $x \geq 0$, $y \geq 0$.

Ans. Consider the constraints.

$x + 2y \geq 10$

x	0	10
y	5	0

Now we have points A (0, 5), B (10, 0)

$x + y \geq 8$

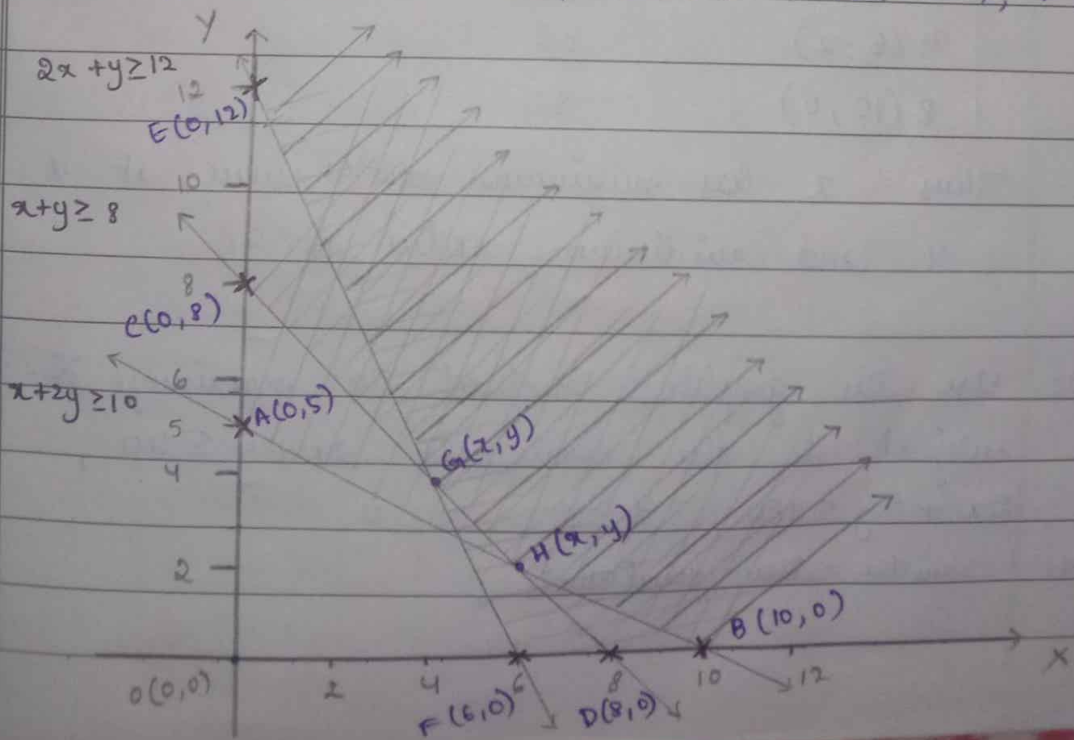
x	0	8
y	8	0

Now we have points C (0, 8), D (8, 0)

$2x + y \geq 12$

x	0	6
y	12	0

Now we have points. E (0, 12), F (6, 0).



for $x + 2y \geq 10$,

$x = 0, y = 0 \Rightarrow 0 \geq 10 \rightarrow$ false

for $x + y \geq 8$,

$x = 0, y = 0 \Rightarrow 0 \geq 8 \Rightarrow$ false

for $2x + y \geq 12$

$x = 0, y = 0 \Rightarrow 0 \geq 12 \rightarrow$ false.

$\therefore$ The corner points of feasible region are

E (0, 12), G (x, y), H (x, y), B (10, 0)

G(x, y) is intersection of lines.

$2x + y = 12$, $x + y = 8$

$\Rightarrow x = 4, y = 4$

$G(x, y) = G(4, 4)$

H(x, y) is intersection of 2 lines

$x + y = 8$, $x + 2y = 10$

$\Rightarrow x = 6, y = 2$

$H(x, y) = H(6, 2)$.

	$z = 5x + 4y$.
E (0, 12)	48
G (4, 4)	$\boxed{36}$
H (6, 2)	38
B (10, 0)	50

Thus z has minimum ~~value~~ value at $x = 4$ and $y = 4$ and minimum value is 36.

20. Use the graphical method to maximize $z = 3x + 4y$ subject to the constraints $2x + y \leq 40$, $2x + 5y \leq 180$, $x \geq 0$, $y \geq 0$.

Ans Consider the constraints;

$2x+y=40$

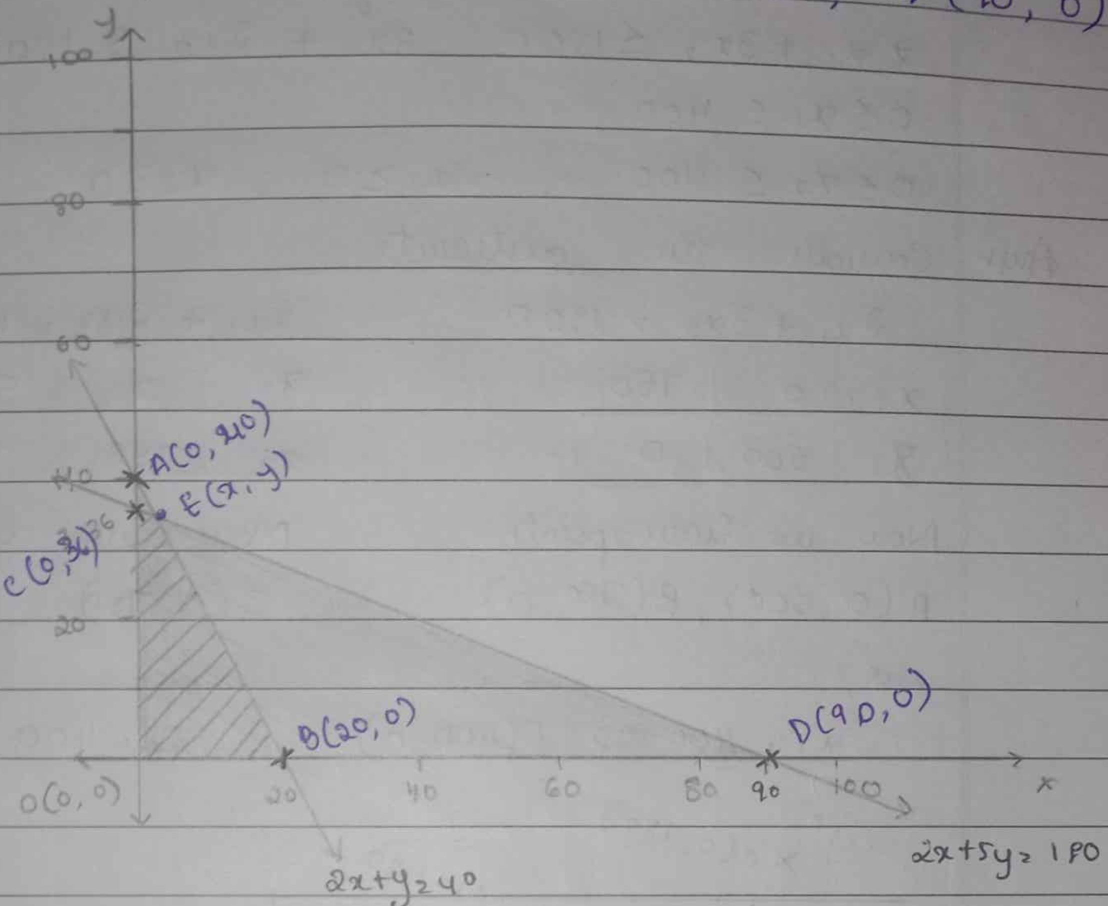
x	0	20
y	40	0

Now we have points
A (0, 40) , B (20, 0)

$2x+5y=180$.

x	0	90
y	36	0

Now we have points.
C (0, 36) , D (90, 0)



for $2x+y \leq 40$

$x=0, y=0$

$0 \leq 40$ ✓ true

for $2x+5y \leq 180$

$x=0, y=0$

$0 \leq 180$ ✓ true.

∴ The corner points of feasible region are
C (0, 36), E (x, y), B (20, 0), O (0, 0)

E(x, y) is intersection of lines :

$2x+y=40$ $\quad$ $2x+5y=180$

$x=2.5$, $y=35$ $\Rightarrow$ $E(x,y) = E(2.5, 35)$

	$z=8x+4y$
C (0, 36)	144
E (2.5, 35)	$\boxed{147.5}$
B (20, 0)	60
O (0, 0)	0

Thus z has a maximum value at $x = 2.5, y = 35$ and maximum value is $\underline{147.5}$

21 Solve the following LPP graphically: maximize $z = 50x_1 + 60x_2$, subject to the constraints
$2x_1 + 3x_2 \le 1500$, $3x_1 + 2x_2 \le 1500$,
$0 \le x_1 \le 400$
$0 \le x_2 \le 400$, $x_1 \ge 0$, $x_2 \ge 0$.

Ans Consider the constraints.

$2x_1 + 3x_2 = 1500$

x_1	0	750
x_2	500	0

Now we have points
A(0, 500), B(750, 0)

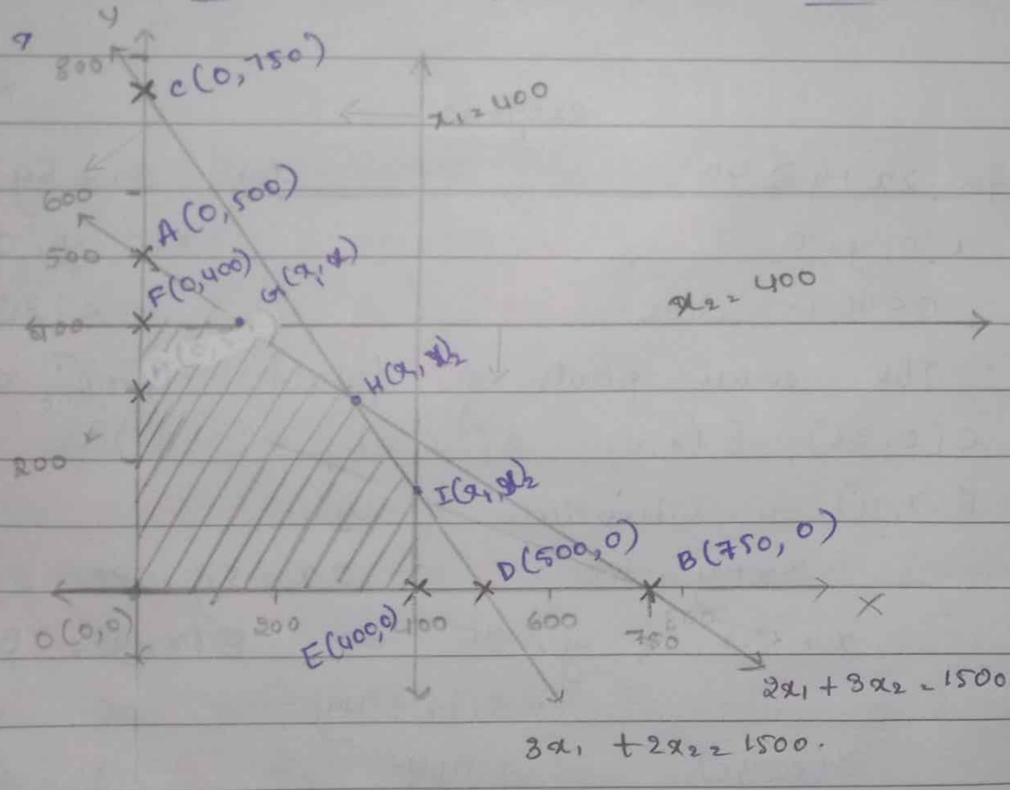
$3x_1 + 2x_2 = 1500$

x_1	0	500
x_2	750	0

Now we have points.
C(0, 750), D(500, 0)

also,

$x_1 = \underline{400} \Rightarrow E(400, 0)$ $x_2 = \underline{400} \Rightarrow F(0, 400)$



- for $2x_1 + 3x_1 \le 1500$
  $0 \le 1500$ ✓ → true
- for $3x_1 + 2x_2 \le 1500$
  $0 \le 1500$ ✓ → true.

∴ corner points of the feasible region is $O(0,0)$, $F(0,400)$, $G(x_1, x_2)$, $H(x_1, x_2)$, $I(x_1, x_2)$, $E(400,0)$.

$G(x, x_2)$ is point of intersection of 2 lines,

$2x_1 + 3x_2 = 1500$ and $x_2 = 400$

$\Rightarrow \quad 2x_1 = 300$

$x_1 = 150$

$G(x_1, x_2) = G(150, 400)$

$H(x_1, x_2)$ is point of intersection of 2 lines.

$2x_1 + 3x_2 = 1500$ & $3x_1 + 2x_2 = 1500$

$x_1 = 300$

$x_2 = 300$

$H(x_1, x_2) = H(300, 300)$

$I(x_1, x_2)$ is intersection of 2 lines

$3x_1 + 2x_2 = 1500$ & $x_1 = 400$

$\Rightarrow \quad 2x_2 = 300$

$x_2 = 150$.

$I(x_1, x_2) = I(400, 150)$

Points	$z = 50x_1 + 60x_2$
$O(0,0)$	0
$F(0,400)$	24000
$G(150,400)$	31500
$H(300,300)$	$\boxed{33000}$
$I(400,150)$	29000
$E(400,0)$	20000

Thus z has a maximum value at $x = 300$, $y = 300$, and maximum value is 33000.

> → subtract something
< → add "

Δ → Indicator (take -ve of original z)

8/8/23

Method 2:

Simplex Method:

22. Maximize $z = x + 1.5y$, given $x \geq 0, y \geq 0$ subject to constraints $x + 2y \leq 160$, $3x + 2y \leq 240$. by Simplex method.

Ans: Given: $z = x + 1.5y$

Consider the constraints:

by adding Slack elements S_1 and S_2

$$\Rightarrow \quad x + 2y + 1S_1 + 0S_2 = 160$$
$$3x + 2y + 0S_1 + 1S_2 = 240$$
$$z = x + 1.5y + 0S_1 + 0S_2$$

Step 1:

	NZV	x	y	S_1	S_2	Qt	Ratio Qt/y
R1	S1	1	2	1	0	160	80
R2	S2	3	2	0	1	240	120
R3	Indicator Δ	-1	-1.5	0	0	0	

pivotal element

pivotal row

pivotal column (least Δ value)

(Now we need to make pivotal element as 1 and the pivotal col as 0)

$\Rightarrow R_1 \to \cancel{R_1 - R_2}\ R_1/2$

$R_2 \to R_2 - R_1$

$R_3 \to R_3 + 3/4\, R_1$.

Step 2:

NZV	x	y	S_1	S_2	Qt	Ratio Qt/x
y	$1/2$	1	$1/2$	0	80	160
S_2	2	0	-1	1	80	40 pivot row
Δ	$-1/4$	0	$3/4$	0	120	

pivot column.

$R_2 \leftrightarrow R_2/2$

$R_1 \rightarrow R_1 - 1/4 R_2$

$R_3 \rightarrow R_3 + 1/8 R_2$.

Step 3:

NZV	x	y	S_1	S_2	Qt	P
y	0	1	$3/4$	$-1/4$	60	
x	1	0	$-1/2$	$1/2$	40	
Δ	0	0	$5/8$	$1/8$	130	

(all indicators should be +ve, then stop process)

Thus, z has maximum value 130 at $x = 40$ and $y = 60$.

23. Use simplex method to maximize $z = 2x + 4y$ subject to constraints $3x + y \leq 22$, $2x + 3y \leq 24$, $x \geq 0$, $y \geq 0$.

Ans: Consider the constraints, by adding the slack elements,

$$3x + y + 1S_1 + 0S_2 = 22$$

$$2x + 3y + 0S_1 + 1S_2 = 24$$

$$z = 2x + 4y + 0S_1 + 0S_2.$$

Step 1

NZV	x	y	S_1	S_2	Qt	Ratio = Qt
S_1	3	1	1	0	22	$22/1 = 22$
S_2	2	3	0	1	24	$24/3 = 8$
Δ	-2	-4	0	0	0	

least = pivot col.

$R_1 \to 3R_1 - R_2$

$R_2 \to R_2 / 3$

$R_3 \to 3R_3 + 4R_2$

Step 2:

NZV	x	y	S_1	S_2	Qt	~~Ratio = Qt/~~
S_1	7	0	3	-1	42	
y	$2/3$	1	0	$1/3$	8	
Δ	2	0	0	4	96	

(all Δ are +ve $\Rightarrow$ Stop.)

$\Rightarrow y = 8, \; x = 0.$

Thus z is maximum at $x = 0$ and $y = 8$ and the maximum value is ~~96~~ 32.

$\because z = 2(0) + 4(8)$

$= \underline{\underline{32}}$