Seg-5 (Curve Fitting)

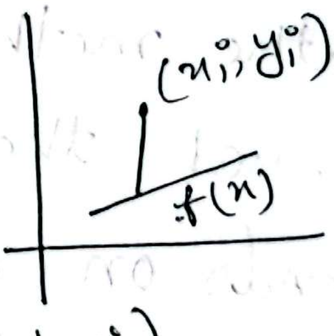
Curve fitting: The exact mathematical relationship between the two variables is given by simple algebraic expression called curve fitting.

Write the importance of curve fitting.

1. Data analysis and prediction
2. Interpolation and Extrapolation
3. Decision making
4. Parameter Estimation

Straight line:

$$y = f(x) = a + bx \quad \text{——} (1)$$



$$\text{error} = q_i = y_i - f(x_i)$$

$$= y_i - a - bx_i$$

1) Minimize the $\Sigma q_i = \Sigma (y_i - a - bx_i)$

2) " " $\Sigma |q_i| = \Sigma |y_i - a - bx_i|$

3) " " $\Sigma q_i^2 = \Sigma (y_i - a - bx_i)^2$

↳ Least square method.

SP-22/SP-23

2 b) Describe the least square method to fit a straight line.

Solve: Let the sum of squares of individual errors be expressed as

$$Q = \sum_{i=1}^{n} q_i^2$$

$$= \sum_{i=1}^{n} [y_i - f(x_i)]^2$$

$$= \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

In the method of (LSM); we choose a and b such that Q is minimum. Since Q depends on a and b, a necessary condition for Q to be minimum is

$$\frac{\partial Q}{\partial a} = 0 \quad \text{and} \quad \frac{\partial Q}{\partial b} = 0$$

Then, $\frac{\partial Q}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i) = 0$

$$\Rightarrow \sum_{i=1}^{n} y_i - a - bx_i = 0$$

$$\Rightarrow -\sum y_i - na + b\sum x_i = 0$$

$$\therefore a = \frac{\sum y_i - b\sum x_i}{n} \quad \text{— (I)}$$

$$\frac{\partial Q}{\partial b} = 0$$

$$\Rightarrow -2\sum_{i=1}^{n}(y_i - a - bx_i)(-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(-x_i y_i + ax_i + bx_i^2) = 0$$

$$\Rightarrow -\sum x_i y_i + a\sum x_i + b\sum x_i^2 = 0$$

$$\Rightarrow \sum x_i y_i = a\sum x_i + b\sum x_i^2 \quad \text{— (II)}$$

In eqn (II),

$$\sum x_i y_i = \frac{\sum y_i - b\sum x_i}{n}\sum x_i + b\sum x_i^2$$

$$\Rightarrow n\Sigma x_i y_i = \Sigma x_i y_i - nb\,\Sigma(x_i)^2 + bn\Sigma x_i^2$$

$$n\Sigma x_i y_i = \Sigma x_i y_i = b\left[n\Sigma x_i^2 - \Sigma(x_i)^2\right]$$

$$b = \frac{n\Sigma x_i y_i - \Sigma x_i y_i}{\quad}$$

$$\Rightarrow n\Sigma x_i y_i = \Sigma x_i \Sigma y_i - b\Sigma(x_i)^2 + bn\Sigma x_i^2$$

$$\Rightarrow n\Sigma x_i y_i - \Sigma x_i \Sigma y_i = b\left[n\Sigma x_i^2 - \Sigma(x_i)^2\right]$$

$$\therefore b = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - \Sigma(x_i)^2}$$

Exc-10.1 (Balagurusamy)

Fit a straight line to the following set of data.

x	1	2	3	4	5
y	3	4	5	6	8

Solve:

x_i	y_i	x_i^2	$x_i y_i$
1	3	1	3
2	4	4	8
3	5	9	15
4	6	16	24
5	8	25	40
$\Sigma x_i = 15$	$\Sigma y_i = 26$	$\Sigma x_i^2 = 55$	$\Sigma x_i y_i = 90$

$$b = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - \Sigma (x_i)^2}$$

$$= \frac{5 \times 90 - 15 \times 26}{5 \times 55 - 15^2}$$

$$= 1.20$$

$$a = \frac{\sum y_i - b \sum x_i}{n}$$

$$= \frac{26 - 1.20 \times 15}{5}$$

$$= 1.60$$

The linear equation is

$$y = a + bx$$

$$\therefore y = 1.6 + 1.2x$$

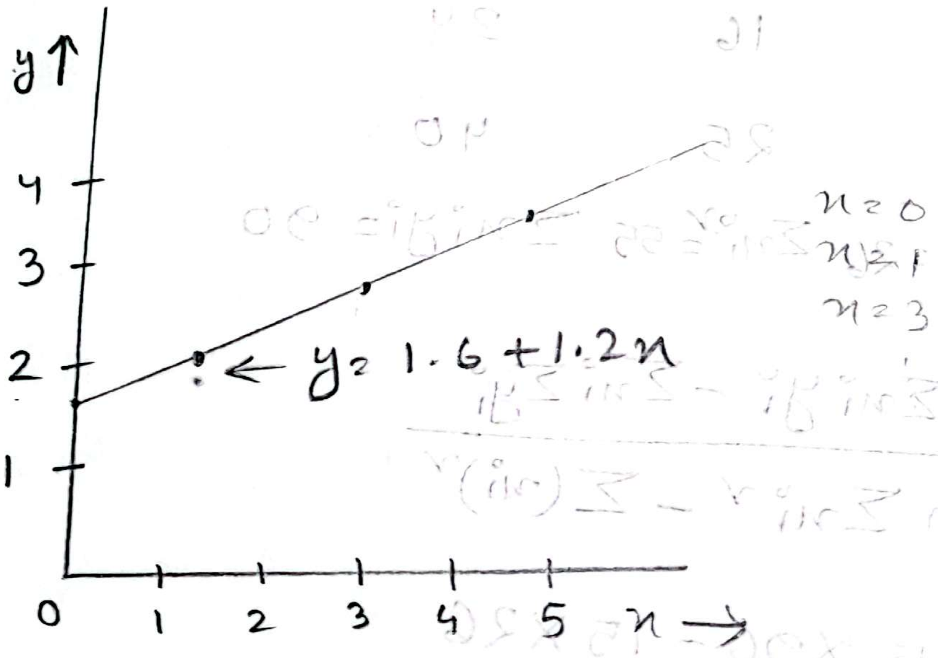


fig: Plot of the data and regression line

SP-23

2(c) The observations from an experiment are given below:

x	10	20	40	50
y	600	500	300	200

It is known that a relation of type $y = a + bx$ exists. Find the best possible values of a and b.

x_i	y_i	x_i^2	$x_i y_i$
10	600	100	6000
20	500	400	10000
40	300	1600	64000
50	200	2500	10000
$\Sigma x_i = 120$	$\Sigma y_i = 1600$	$\Sigma x_i^2 = 4600$	$\Sigma x_i y_i = 38000$

$$b = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

$$= \frac{4 \times 38000 - (120 \times 1600)}{(4 \times 4600) - (120)^2} = -10$$

$$\therefore a = \frac{\Sigma y_i - b\Sigma x_i}{n}$$

$$= \frac{1600 + (10 \times 120)}{4}$$

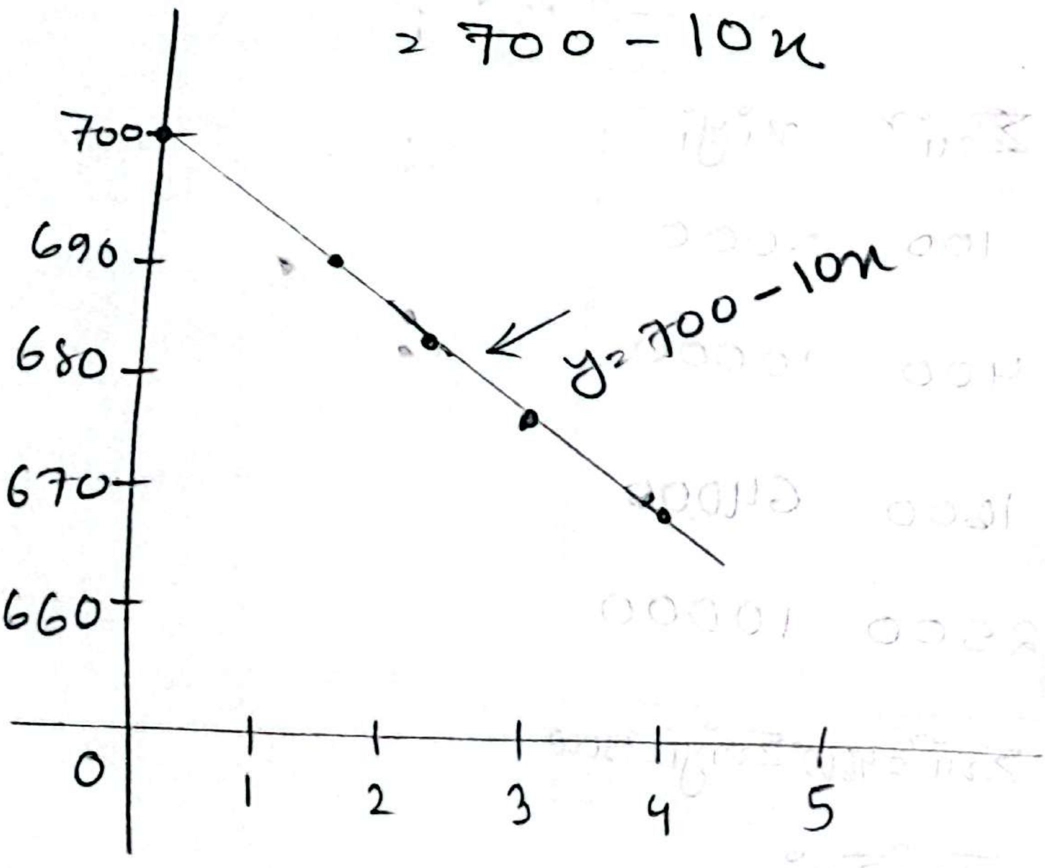
$$= 700$$

The linear eqn is,

$$y = a + bx$$

$$= 700 + (-10)x$$

$$= 700 - 10x$$



$X = 0, Y = 700$

$X = 1, y = 690$

$X = 2, y = 680$

$X = 3, y = 670$

$X = 4, y = 660$

Fig: Plot of the data and regression line

SP-23 / Au-22

(b) Describe the least square method to fit a parabola. / Derive an equation to " "

" "

Solve:

Let the parabolic curve be represented by the eqn,

$$y = f(x) = a_0 + a_1 x + a_2 x^2$$

$$Q = \sum_{i=1}^{n} q_i^2$$

$$= \sum_{i=1}^{n} [y_i - f(x_i)]^2$$

$$= \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$$

According to the method of least square, we have to choose a_0, a_1, a_2 such that Q is minimum. Equating the first derivatives to zero and simplifying, we get the following normal eqn:

$$na_0 + a_1 \sum x_i + a_2 \sum x_i^2 = \sum y_i$$

$$a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 = \sum x_i y_i$$

$$a_0 \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 = \sum x_i^2 y_i$$

If the eqn be a polynomial of nth degree,

$$na_0 + a_1 \sum x_i + a_2 \sum x_i^2 + \dots + a_n \sum x_i^n = \sum y_i$$

$$a_0 \sum x_i + a_1 \sum x_i^2 + a_3 \sum x_i^3 + \dots + a_n \sum x_i^{n+1} = \sum x_i y_i$$

- - - - - - - - - - - - - - - - - -

$$a_0 \sum x_i^n + a_1 \sum x_i^{n+1} + a_2 \sum x_i^{n+2} + \dots + a_n \sum x_i^{2n} = \sum x_i^n \sum y_i$$

Ex-10.4 Similar to SP-24

2C) Fit a second order polynomial to the data in the table below:

x	2	4	6	8
y	1.4	2.0	2.4	2.6

Solve:

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
2	1.4	4	8	16	2.8	5.6
4	2.0	16	64	256	8	32
6	2.4	36	216	1296	14.4	86.4
8	2.6	64	512	4096	20.8	166.4
$\Sigma x_i = 20$	$\Sigma y_i = 8.4$	$\Sigma x_i^2 = 120$	$\Sigma x_i^3 = 800$	$\Sigma x_i^4 = 5664$	$\Sigma x_i y_i = 46$	$\Sigma x_i^2 y_i = 290.4$

We know,

Fitting a parabola,

$$n a_0 + a_1 \Sigma x_i + a_2 \Sigma x_i^2 = \Sigma y_i \quad \text{—(i)}$$

$$a_0 \Sigma x_i + a_1 \Sigma x_i^2 + a_2 \Sigma x_i^3 = \Sigma x_i y_i \quad \text{—(ii)}$$

$$a_0 \Sigma x_i^2 + a_1 \Sigma x_i^3 + a_2 \Sigma x_i^4 = \Sigma x_i^2 y_i \quad \text{—(iii)}$$

Here, $n = 4$

From eqn (i) (ii) and (iii)

$$4a_0 + 20a_1 + 120a_2 = 8.4$$

$$20a_0 + 120a_1 + 800a_3 = 46$$

$$120a_0 + 800a_1 + 5664a_2 = 290.4$$

Solving these eqn,

$$a_0 = 0.6$$

$$a_1 = 0.45$$

$$a_2 = -0.025$$

$\therefore$ The parabola equation is

$$y = 0.6 + 0.45x - 0.025x^2$$

Fitting a power function

$$y = ax^b$$

$$\ln y = \ln a + b \ln x \quad \text{—(i)}$$

$$n \ln a = \sum \ln y_i - b \sum \ln x_i$$

$$\ln a = \frac{\sum \ln y_i - b \sum \ln x_i}{n}$$

$$\ln a = R \qquad \left[R = \frac{\sum \ln y_i - b \sum \ln x_i}{n} \right]$$

$$a = e^R$$

$$\therefore a = e^{\left(\frac{\sum \ln y_i - b \sum \ln x_i}{n}\right)}$$

$$\therefore \cancel{b = n \sum \ln x_i y_i -}$$

$$b = \frac{n \sum \ln x_i \ln y_i - \sum \ln x_i \sum \ln y_i}{n \sum (\ln x_i)^2 - \left(\sum \ln x_i\right)^2}$$

Au-22

Fit the power $y = ax^b$ to the data in table below:

x	2	4	6	8
y	1.4	2.0	2.4	2.6

Solve:

x_i	y_i	$\ln x_i$	$\ln y_i$	$(\ln x_i)^2$	$\ln x_i \cdot \ln y_i$
2	1.4	0.69	0.33	0.48	0.23
4	2.0	1.38	0.69	1.92	0.95
6	2.4	1.79	0.87	3.21	1.53
8	2.6	2.07	0.95	4.32	1.96
$\Sigma x_i = 20$	$\Sigma y_i = 8.4$	$\Sigma \ln x_i = 5.93$	$\Sigma \ln y_i = 2.84$	$\Sigma (\ln x_i)^2 = 9.93$	$\Sigma \ln x_i \Sigma \ln y_i = 4.67$

$$b = \frac{n \Sigma \ln x_i \cdot \Sigma \ln y_i - \Sigma \ln x_i \cdot \Sigma \ln y_i}{n \Sigma (\ln x_i)^2 - (\Sigma \ln x_i)^2}$$

$$= \frac{4 \times 5.93 \times 2.84 - 4.67}{4 \times (5.93)^2 - (5.93)^2}$$

$$= \frac{62.69}{4.55} = 13.79$$

$$a = e^{\frac{\sum \ln y_i - b \sum \ln x_i}{n}}$$

$$= e^{\frac{2.84 - 13.79 \times 5.93}{4}}$$

$$= 7.38 \times 10^{-3}$$

The power function eqn is

$$y = ax^b$$

$$= 7.38 \times 10^{-3} x^{13.79}$$

SP-22

2(a) Compare least square method with interpolation.

LSM	Interpolation
① Minimizes errors in curve fitting	① Estimates values between data point precisely.
② It is suitable for noisy data.	② It is not suitable for noisy data.
③ It can fit curves of any shape.	③ The curve must pass exactly through all given data points.
④ Usage – Regression, Curve fitting etc.	④ Usage – Signal processing, image processing.

Here's a concise comparison between **curve fitting** and **interpolation**:

Aspect	Curve Fitting	Interpolation
Definition	Fits a smooth curve to approximate data points, allowing for deviations.	Passes exactly through all given data points.
Purpose	To model the trend or pattern in data, even if not all points are fitted exactly.	To estimate values within the range of known data points.
Use Case	Used when data has noise or uncertainty.	Used when data is precise and error-free.
Behavior Outside Data	May extrapolate trends beyond the data range.	Rarely used for extrapolation; focuses within the data range.
Examples	Polynomial regression, exponential fitting.	Linear, spline, or Lagrange interpolation.