Unit-5

Curve-fitting and Approximation

Curve-fitting :⇒ Let there be two variables x and y which give us a set of n pairs of numerical values $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. In order to have an approximate idea about the relationship of these two variables, we plot these n paired points on a graph thus, we get a diagram called scatter or dot diagram. From scatter diagram, we get only an approx. non-mathematical rel^n b/w two variables.

Curve-fitting means an exact relationship b/w two variables by algebraic eq^n, infact, this relationship is the eq^n of the curve.

* Curve-fitting means to form an eq^n of the curve from the given-data.
* Theoretically, it is useful in the study of correlation and regression. It enables us to represent the relationship b/w two variables by simple algebraic expressions.
* It is also used to estimate the values of one variable corresponding to the specified values of the other variable.
* The constants occuring in the eq^n of approximate curve can be found by following methods:-
  (i) Graphical method
  (ii) Method of group averages
  (iii) Method of least-square
  (iv) Method of moments.

Method of Least-Square

This method provides a unique set of values to the constants and hence suggests a curve of best-fit to the given data.

Suppose we have n paired of observations values $(x_1, y_1), (x_2, y_2), ---, (x_n, y_n)$ of two variables x & y.

It is required to fit a polynomial of degree n of the type $y = a + bx + cx^2 + ---- + kx^n$ to these values.

* Aim: To determine the constants $a, b, c, \sim k$. such that it represents a best fit to the given data. (reverse side)

(1) Fitting a straight line

$y = a + bx$ be the straight line to be fitted to the given data (x_i, y_i) $i = 1 -- n$

The residual at $x = x_i$ is

$$E_i = y_i - \{a + b x_i\}$$

Introduce a new quantity U s.t.

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

The constants a & b are chosen in such a way that the sum of squares of residuals is minimum.

i.e. by principal of least-square, U is minimum

ie $\frac{\partial U}{\partial a} = 0$ and $\frac{\partial U}{\partial b} = 0$

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

$$\Sigma y - na - b\Sigma x = 0 \qquad \Sigma xy - a\Sigma x - b\Sigma x^2 = 0$$

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

Since (x_i, y_i) are known, solⁿ (1) & (2) for a & b. and put in $y = a + bx$.

Let $y = f(x)$ | $f(x)$ may have any form (r-t)/

$E_i = \underbrace{y_i}_{\text{observed values}} - \underbrace{f(x_i)}_{\text{expected values.}}$

E_i ↓ Residual

Introduce a new quantity U st

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

The constants a, b, c — are chosen in such a way that the sum of squares of residual is minimum.

Hence

$$\frac{\partial U}{\partial a} = 0, \ \frac{\partial U}{\partial b} = 0 \ — \ \frac{\partial U}{\partial k} = 0$$

On simplifying we get

$$\Sigma y = na + b\Sigma x + \cdots + k\Sigma x^n$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + \cdots + k\Sigma x^{n+1}$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + \cdots + k\Sigma x^{n+2}$$

$$\vdots$$

$$\Sigma x^n y = a\Sigma x^n + b\Sigma x^{n+1} + \cdots + k\Sigma x^{2n}$$

These are called Normal eqn.

Q1. By the method of least-square, find the straight line that best fits the following data. 3

x	1	2	3	4	5
y	14	27	40	55	68

Sol 1. Let the eqⁿ of straight line of best fit is

$$y = a + bx$$

then Normal eqⁿ are

$$\Sigma y = na + b\Sigma x \qquad n = 5$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

x	y	x^2	xy
1	14	1	14
2	27	4	$\cancel{8}$ 54
3	40	9	$\cancel{27}$ 120
4	55	16	$\cancel{64}$ 220
5	68	25	$\cancel{125}$ 340
$\Sigma x = 15$	$\Sigma y = 204$	$\Sigma x^2 = 55$	$\Sigma xy = 748$

$$204 = 5a + 15b$$
$$748 = 15a + 55b$$

$$\Rightarrow a = 0,\ b = 13.6$$

∴ line $y = 13.6x$

Q2. Show that the line of fit to the following data is given by $y = 0.7x + 11.285$

$n = 6$

$$y = a + bx$$
$$\left.\begin{aligned}\Sigma y &= 6a + b\Sigma x\\ \Sigma xy &= a\Sigma x + b\Sigma x^2\end{aligned}\right\}$$

x	y	x^2	xy
0	12	0	0
5	15	25	75
10	17	100	170
15	22	225	330
20	24	400	480
25	30	625	750
75	120	1375	1805

$$120 = 6a + 75b$$
$$1805 = 75a + 1375b$$

$$\boxed{\begin{aligned}a &= 11.2857\\ b &= 0.6971\end{aligned}}$$

(2) Fitting of the curve of type

$$y = ax + bx^2$$

By principle of Least-squares,

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

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

$$\sum(-2x_i)(y_i - ax_i - bx_i^2) = 0$$

$$-\sum xy + a\sum x^2 + b\sum x^3 = 0$$

$$\sum xy = a\sum x^2 + b\sum x^3 \quad -(1)$$

$$\sum(-2x_i^2)(y_i - ax_i - bx_i^2) = 0$$

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

Q→ find best fit curve

x	y	x^2	x^3	x^4	xy	x^2y
1	1.8	1	1	1	1.8	1.8
2	5.1	4	8	16	10.2	20.4
3	8.9	9	27	81	26.7	80.1
4	14.1	16	64	256	56.4	225.6
5	19.8	25	125	625	99	495
$\sum x = \cancel{x}$	$\cancel{y}$	$\sum x^2 = 55$	$\sum x^3 = 225$	$\sum x^4 = 979$	$\sum xy = 194.1$	$\sum x^2y = 822.9$

Substitute in (1) & (2)

$$194.1 = 55a + 225b$$

$$822.9 = 225a + 979b$$

$$a = 1.52, \quad b = 0.49$$

(reverse side) →

$$\underbrace{y = a + bx + cx^2}$$ {second degree parabola}

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\}$$

<u>Q1</u> fit a second degree parabola to

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	4	1	1	1	4	4
2	10	4	8	16	20	40
3	17	9	27	81	51	153
4	30	16	64	256	120	480
10	62	30	100	354	195	677

$$62 = 5a + 10b + 30c$$
$$195 = 10a + 30b + 100c$$
$$677 = 30a + 100b + 354c$$

$a = 1.2,\ b = 1.1,\ c = 1.5$

$$\boxed{y = 1.2 + 1.1x + 1.5x^2}$$

(3) fitting of the curve $y = ax^2 + \frac{b}{x}$

$$U = \sum_{i=1}^{n}\left(y_i - ax_i^2 - \frac{b}{x_i}\right)^2$$

$$\frac{\partial U}{\partial a} = 0 \quad , \quad \frac{\partial U}{\partial b} = 0$$

$\cancel{\sum y - a\sum x_i^2 + \frac{b}{x_i}}$

$$\sum(-2x_i^2)\left(y_i - ax_i^2 - \frac{b}{x_i}\right) = 0$$

$$-\sum x^2 y + a\sum x^4 + b\sum x = 0$$

$$\boxed{\sum x^2 y = a\sum x^4 + b\sum x}$$

$$\sum\left(-\frac{2}{x_i}\right)\left(y_i - ax_i^2 - \frac{b}{x_i}\right) = 0$$

$$-\sum\frac{y}{x} + a\sum x - b\sum\frac{1}{x^2} = 0$$

$$\boxed{\sum\frac{y}{x} = a\sum x + b\sum\frac{1}{x^2}}$$

(4) fitting of the curve $y = a + \frac{b}{x} + \frac{c}{x^2}$

$$U = \sum\left(y_i - a_i - \frac{b}{x_i} - \frac{c}{x_i^2}\right)^2$$

$$\frac{\partial U}{\partial a} = 0 \quad , \quad \frac{\partial U}{\partial b} = 0 \quad , \quad \frac{\partial U}{\partial c} = 0$$

$$(-2)\sum\left(y_i - a - \frac{b}{x_i} - \frac{c}{x_i^2}\right) = 0$$

$$\boxed{\sum y = na + b\sum\frac{1}{x} + c\sum\frac{1}{x^2}}$$

$$\sum\frac{2}{x_i}\left(y_i - a - \frac{b}{x_i} - \frac{c}{x_i^2}\right) = 0$$

$$\boxed{\sum\frac{y}{x} = a\sum\frac{1}{x} + b\sum\frac{1}{x^2} + c\sum\frac{1}{x^3}}$$

$$\boxed{\sum\frac{y}{x^2} = a\sum\frac{1}{x^2} + b\sum\frac{1}{x^3} + c\sum\frac{1}{x^4}}$$

(5) fitting of $xy = b + ax$

$$xy = b + ax$$

$$y = \frac{b}{x} + a$$

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

$$\frac{\partial U}{\partial a} = 0 \qquad \frac{\partial U}{\partial b} = 0$$

$$\boxed{\sum y = na + b\sum\frac{1}{x}}$$

$$\boxed{\sum\frac{y}{x} = a\sum\frac{1}{x} + b\sum\frac{1}{x^2}}$$

6 Fitting of an exponential curve

$$y = a e^{bx}$$

Take log of both sides.

$$\log_{10} y = \log_{10} a + bx \log_{10} e$$

$$\boxed{Y = A + Bx}$$

where $Y = \log_{10} y$, $A = \log_{10} a$, $B = b \log_{10} e$

$$U = \sum (Y_i - A - Bx_i)^2$$

$$\frac{\partial U}{\partial A} = 0, \quad \frac{\partial U}{\partial B} = 0$$

$\log_{10} e = 0.4343$

$$\sum Y = nA + B\sum x, \quad \sum xY = A\sum x + B\sum x^2$$

solve & get A, B.

$$A = \log_{10} a \Rightarrow a = \text{antilog } A$$

$$B = b \log_{10} e \Rightarrow b = \frac{B}{\log_{10} e} = \frac{B}{0.4343}$$

Q→ Find the curve of best fit of the type $y = ae^{bx}$ to the following data by method of least-squares

x	y	$Y = \log_{10} y$	x^2	xY
1	10	1	1	1
5	15	1.1761	25	5.8805
7	12	1.0792	49	7.5544
9	15	1.1761	81	10.5849
12	21	1.3222	144	15.8664
$\sum x = 34$		$\sum Y = 5.7536$	$\sum x^2 = 300$	$\sum xY = 40.8862$

$$5.7536 = 5A + 34B$$

$$40.8862 = 34A + 300B$$

$$A = 0.9766, \quad B = 0.02561$$

$$a = \text{antilog } A = 9.4754 \qquad b = \frac{B}{0.4343} = 0.059$$

$$y = ae^{bx}$$

$$y = 9.4754\, e^{0.0597}$$

Q→ ① x : 2 4 6

y : 4.077 11.084 30.128

8 10 $y = 0.5580\, e^{1.0631x}$

81.897 222.62

h)

② x : 1 2 3 4

y : 1.6 4.5 13.8 40.2

5 6

125 300

$$y = 1.49989\, e^{0.50001x}$$

7 fitting of the curve $y = ab^x$

$$y = ab^x$$
$$\log y = \log a + x \log b$$
$$Y = A + xB.$$

Normal eq^n
$$\left.\begin{aligned} \Sigma Y &= nA + B\Sigma x \\ \Sigma xY &= A\Sigma x + B\Sigma x^2. \end{aligned}\right\} \text{solve find } A \,\&\, B.$$

$\Rightarrow a = \text{Antilog } A$

$b = \text{Antilog } B.$

8 $y = \frac{c_0}{x} + c_1\sqrt{x}$

$$U = \Sigma\left(y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i}\right)^2$$

$$\frac{\partial U}{\partial c_0} = 0 \quad , \quad \frac{\partial U}{\partial c_1} = 0$$

$$\boxed{\Sigma \frac{y}{x} = c_0 \Sigma \frac{1}{x^2} + c_1 \Sigma \frac{1}{\Sigma x_1}} \qquad \boxed{\Sigma y\sqrt{x} = c_0 \Sigma \frac{1}{\sqrt{x}} + c_1 \Sigma (x)}$$

9 $y = ae^{-3x} + be^{-2x}$

$$U = \Sigma(y - ae^{-3x} - be^{-2x})^2$$

$$\frac{\partial U}{\partial a} = 0 \qquad \frac{\partial U}{\partial b} = 0$$

$$\Sigma ye^{-3x} = a\Sigma e^{-6x} + b\Sigma e^{-5x}$$
$$\Sigma ye^{-2x} = a\Sigma e^{-5x} + b\Sigma e^{-4x}$$

Ex→ obtain the least square fit for $y = ae^{-3x} + be^{-2x}$

x	$f(x)/y$	e^{-4x}	e^{-5x}	e^{-6x}	ye^{-2x}	ye^{-3x}
0.1	0.76	0.6703	0.6065	0.5488	0.6222	0.5630
0.2	0.58	0.4493	0.3679	0.3012	0.3888	0.3183
0.3	0.44	0.3012	0.2231	0.1653	0.2415	0.1789
0.4	0.35	0.2019	0.1353	0.0907	0.1573	0.1054
		1.6227	1.3328	1.106	1.4098	1.1656

$$\left.\begin{aligned} 1.106a + 1.3328b &= 1.1656 \\ 1.3328a + 1.6227b &= 1.4098 \end{aligned}\right\} \Rightarrow a = 0.6778, \; b = 0.3121$$

$$y = 0.6778e^{-3x} + 0.3121e^{-2x}$$

Multiple linear Regression

Consider such a linear func. as

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

The sum of the squares of residual is

$$U = \sum (y_i - a - bx_i - cz_i)^2$$

diff^n partially w.r.t. a, b & c.

$$\left.\begin{aligned} \frac{\partial U}{\partial a} = 0 &\Rightarrow \sum y = na + b\sum x + c\sum z \\ \frac{\partial U}{\partial b} = 0 &\Rightarrow \sum xy = a\sum x + b\sum x^2 + c\sum xz \\ \frac{\partial U}{\partial c} = 0 &\Rightarrow \sum yz = a\sum z + b\sum xz + c\sum z^2 \end{aligned}\right\} \text{—— (2)}$$

Solve normal eq^n and find a, b, c. and put in (1) which is called regression plane.

Ex→ obtain a regression plane by

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

substitute in above normal eq^n (2)

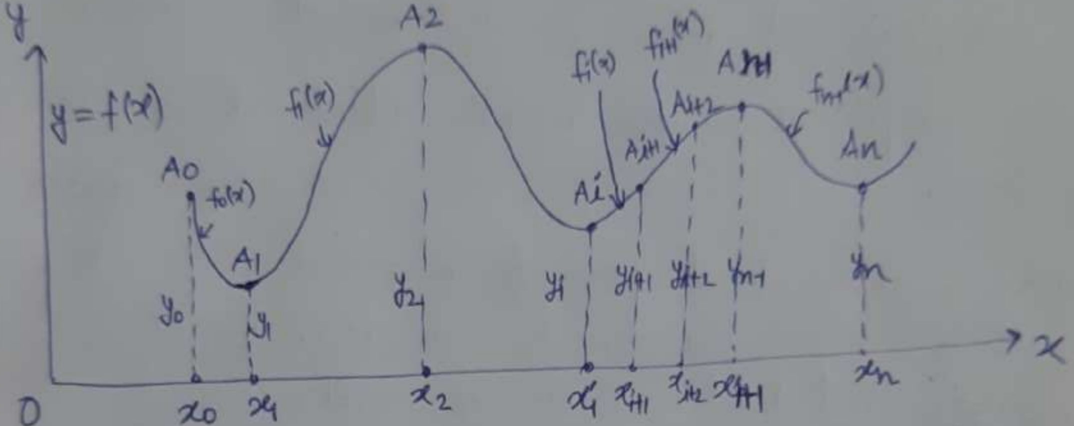
$$84 = 4a + 10b + 6c$$
$$240 = 10a + 30b + 20c$$
$$156 = 6a + 20b + 14c$$

$$\Rightarrow a = 10$$
$$b = 2$$
$$c = 4$$

Regression plane $\Rightarrow y = 10 + 2x + 4z$.

Spline Interpolation

* So far, a single polynomial has been fitted to the tabulated pts.
* This method is not always accurate.
* It is also called piecewise interpolation, a fit for every sub-interval.
* We fit a curve b/w A_i and A_{i+1} & another curve b/w A_{i+1} and A_{i+2} > such that the slopes of two curve matches at A_{i+1}.



* These curve are commonly used, cubic polynomials.
* Higher order polynomials can also be used but resulting graph will not be better.

$$f(x) = \frac{(x_{i+1}-x)^3}{6h} M_i + \frac{(x-x_i)^3}{6h} M_{i+1} + \frac{(x_{i+1}-x)}{h}\left(y_i - \frac{h^2}{6} M_i\right) + \frac{(x-x_i)}{h}\left(y_{i+1} - \frac{h^2}{6} M_{i+1}\right)$$

Cubic-spline

Consider a problem of interpolating b/w the following data points using spline fitting.

x	x_0	x_1	x_2	$\cdots\cdots$	x_n
y	y_0	y_1	y_2	$\cdots\cdots$	y_n

* Assumption of Cubic spline $f(x)$:
  - → $f(x)$ is a linear polynomial outside the interval (x_0, x_n).
  - → $f(x)$ is a cubic polynomial in each of the subintervals.
  - → $f'(x)$ and $f''(x)$ are continuous at each pt.

* Now since $f(x)$ is a cubic in each subinterval so $f''(x)$ shall be linear.

$x_i \quad x_{i+1}$

* Taking equally spaced values of x so that $x_{i+1} - x_i = h$, so according to Lagrange's interpolation, we can write

$$f''(x) = \frac{(x - x_{i+1})}{(x_i - x_{i+1})} f''(x_i) + \frac{(x - x_i)}{(x_{i+1} - x_i)} f''(x_{i+1})$$

$$= \frac{(x - x_{i+1})}{-h} f''(x_i) + \frac{(x - x_i)}{h} f''(x_{i+1})$$

$$f''(x) = \frac{1}{h}\left[(x_{i+1} - x) f''(x_i) + (x - x_i) f''(x_{i+1})\right]$$

Integrating this eqn twice

$$f(x) = \frac{1}{h}\left[\frac{(x_{i+1} - x)^3}{3!} f''(x_i) + \frac{(x - x_i)^3}{3!} f''(x_{i+1})\right] + a_i (x_{i+1} - x) + b_i (x - x_i) \quad \text{——} (1)$$

$$\left.\begin{aligned} a_i &= \frac{1}{h}\left[y_i - \frac{h^2}{3!} f''(x_i)\right] \\ b_i &= \frac{1}{h}\left[y_{i+1} - \frac{h^2}{3!} f''(x_{i+1})\right] \end{aligned}\right\} \text{——} (2)$$

After substituting the values of a_i, b_i and writing $f''(x_i) = M_i$ with the condition of continuity as well, we get

$$M_{i-1} + 4M_i + M_{i+1} = \frac{6}{h^2}(y_{i-1} - 2y_i + y_{i+1}) \; ; \quad i = 1 \text{ to } n-1. \qquad (3)$$

Since graph is linear for $x < x_0$ & $x > x_n$, we have

$$M_0 = 0 \text{ and } M_n = 0 \qquad (4)$$

(3) & (4) give $(n+1)$ equations in $(n+1)$ unknowns M_i ($i = 0 - n$) which can be solved. Substituting the value of M_i in (1) or (2), we get the cubic spline.

Q1. Obtain the cubic spline for

x	0 (x_0)	1 (x_1)	2 (x_2)	3 (x_3)
y	2 (y_0)	-6 (y_1)	-8 (y_2)	2 (y_3)

$h = 1$, $n = 3$

Sol→ The cubic spline can be determined from

$$M_{i-1} + 4M_i + M_{i+1} = \frac{6}{(1)^2}(y_{i-1} - 2y_i + y_{i+1}), \; i = 1 \text{ to } 2.$$

$$M_0 + 4M_1 + M_2 = 6(y_0 - 2y_1 + y_2)$$

$$M_1 + 4M_2 + M_3 = 6(y_1 - 2y_2 + y_3)$$

Now $M_0 = 0$, $M_3 = 0$

$$\begin{cases} 4M_1 + M_2 = 6(2 + 12 - 8) = 36 \\ M_1 + 4M_2 = 6(-6 + 16 + 2) = 72 \end{cases}$$

Solve $\Rightarrow$ $M_1 = 4.8$ and $M_2 = 16.8$

Now the cubic spline in $(x_i \le x \le x_{i+1})$ is

$$f(x) = \frac{1}{6}(x_{i+1} - x)^3 M_i + \frac{1}{6}(x - x_i)^3 M_{i+1} + \frac{(x_{i+1} - x)}{1}\left(y_i - \frac{1}{6}M_i\right) + \frac{(x - x_i)}{1}\left(y_{i+1} - \frac{1}{6}M_{i+1}\right) \qquad (1)$$

There are 3 subintervals so there will be three polynomials corresponding to $i = 0, 1$ & 2.

for 1^{st} subinterval :- $i=0$, the cubic spline in $(0 \le x \le 1)$

$\cancel{f(x) = \frac{1}{6}(1-x)^3(0)}$ from 1

$$f(x) = \frac{1}{6}(x_1 - x)^3 M_0 + \frac{1}{6}(x - x_0)^3 M_1 + (x_1 - x)\left(y_0 - \frac{1}{6}M_0\right) + (x - x_0)\left(y_1 - \frac{1}{6}M_1\right)$$

$$= \frac{1}{6}(1-x)^3(0) + \frac{1}{6}(x-0)^3(4.8) + (1-x)\left(2 - \frac{1}{6}(0)\right) + (x-0)\left(-6 - \frac{1}{6}(4.8)\right)$$

$$f(x) = 0.8x^3 - 8.8x + 2 \quad \text{in } (0 \le x \le 1)$$

for 2^{nd} subinterval :- $i=1$ the cubic spline in $1 \le x \le 2$,
from 1

$$f(x) = 2x^3 - 5.84x^2 - 1.68x + 0.8$$

for 3^{rd} subinterval :- $i=2$, the cubic spline in $2 \le x \le 3$
from 1.

$$f(x) = -0.8x^3 + 2.64x^2 + 9.68x - 14.8$$

Hence;

$$f(x) = \begin{cases} 0.8x^3 - 8.8x + 2 & 0 \le x \le 1 \\ 2x^3 - 5.84x^2 - 1.68x + 0.8 & 1 \le x \le 2 \\ -0.8x^3 + 2.64x^2 + 9.68x - 14.8 & 2 \le x \le 3 \end{cases}$$

Q2. $x: \; 1 \quad 2 \quad 3 \quad 4$

$y: \; 1 \quad 2 \quad 5 \quad 11$

find cubic spline and evaluate $y(1.5)$ & $y'(3)$.

Sol→ $h = 1, \; n = 3$

cubic spline are obtained by

$$M_{i-1} + 4M_i + M_{i+1} = 6(y_{i-1} - 2M_i + y_{i+1}) \; ; \; i = 1, 2$$

$$\begin{cases} M_0 + 4M_1 + M_2 = 6(y_0 - 2y_1 + y_2) \\ M_1 + 4M_2 + M_3 = 6(y_1 - 2y_2 + y_3) \\ M_0 = 0, \; M_3 = 0 \end{cases}$$

$$\Rightarrow M_1 = 2, \; M_2 = 4.$$

The cubic spline in $x_i \le x \le x_{i+1}$ is

$$f(x) = \frac{1}{6}\left[(x_{i+1} - x)^3 M_i\right] + \frac{1}{6}(x - x_i)^3 M_{i+1} + (x_i - x)\left(y_i - \frac{1}{6}M_i\right) + (x - x_i)\left(y_{i+1} - \frac{1}{6}M_{i+1}\right)$$

put $i = 0, \; i = 1, \; i = 2$, the cubic splines are

$$f(x) = \begin{cases} \frac{1}{3}(x^3 - 3x^2 + 5x) & 1 \le x \le 2 \\ \frac{1}{3}(x^3 - 3x^2 + 5x) & 2 \le x \le 3 \\ \frac{1}{3}(-2x^3 + 24x^2 - 76x + 81) & 3 \le x \le 4 \end{cases}$$

$f'(x) = \frac{1}{3}(3x^2 - 6x + 5)$ $\qquad f(1.5) = 11/8$

$f'(1.5) = 11/8$ $\qquad f'(3) = 14/3.$

Regression Analysis

The term 'regression' stands for some sort of functional relationship b/w two or more related variables.

* The fundamental difference b/w problems of curve-fitting and regression is that in regression, any of the variables may be considered as independent or dependent which is curve-fitting, one variable cannot be dependent.
* Regression measures the nature and extent of correlation.
* Regression is the estimation or prediction of unknown values of one variable from known values of another variable.

Curve of Regression and Regression Equation

If two variates x & y are correlated, then the scatter diagram will be more or less concentrated round a curve. This curve is called the curve of regression.

The mathematical eqn of the regression curve is called regression equation.

Linear Regression

when the pts. of the scatter diagram concentrate round a straight line, the regression is called linear and this st line is known as the line of regression.

otherwise (non-linear Regression).

Lines of Regression

A line of regression is the straight line which gives the best-fit in the least-square sense to the given frequency.

In case of n pairs (x_i, y_i); $i = 1, 2, \dots n$. We may choose any one of the variable as independent and another as dependent variable.

Either of the two may be estimated for the given values of the other.

* Thus if we wish to estimate y for given values of x, we shall have the regression eqn of the form $y = a + bx$, called the regression line of y on x.
* If we wish to estimate x for given values of y, we shall have the regression line of the form $x = A + By$, called the regression line of x on y.

$\Rightarrow$ Thus, in general, we always have two lines of regression.

* Regression line of y on x is given by

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

where $\bar{x}$ & $\bar{y}$ are mean values while

Regression -coefficient $b_{yx} = \dfrac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$

or

$b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$ where r = correlation coefficient

* Regression line of x on y is given by

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

Regression coefficient $b_{xy} = \dfrac{n\,\Sigma xy - \Sigma x\,\Sigma y}{n\,\Sigma y^2 - (\Sigma y)^2}$

or

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

Note :- (1) If $r=0$ then two lines of regression become $y=\bar{y}$ and $x=\bar{x}$ which are two straight line parelled to x and y axes respectively and passing through their means $\bar{y}$ and $\bar{x}$. They are mutually perpendicular.

If $r=\pm 1$, the two lines of regression will coincide.

(2) $$r^2 = b_{xy} \times b_{yx}.$$

(3) The correlation co-efficient and the two regression coefficients have same sign.

b_{xy}, b_{yx} and r have same sign

(4) Angle b/w two Regression lines :-

$$\tan\theta = \left(\frac{1-r^2}{r}\right)\frac{\sigma_x\,\sigma_y}{\sigma_x^2+\sigma_y^2}.$$

Ex 1. Calculate linear regression coefficients from

$x \rightarrow$	1	2	3	4	5	6	7	8
y	3	7	10	12	14	17	20	24.

Sol $\rightarrow$

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
$\Sigma x = 36$	$\Sigma y = 107$	$\Sigma x^2 = 204$	$\Sigma y^2 = 1763$	$\Sigma xy = 599$

$\bar{y} = \frac{107}{8}$

$\bar{x} = \frac{\Sigma x}{8} = \frac{36}{8}$

Here $n = 8$

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

$$b_{yx} = \frac{940}{336} = 2.7976 \qquad b_{xy} = 0.3540$$

Regression lines :- y on x

$$y - \bar{y} = b_{yx}(x - \bar{x}) \Rightarrow y - \frac{107}{8} = 2.7976\left(x - \frac{36}{8}\right)$$

x on y

$$x - \bar{x} = b_{xy}(y - \bar{y}) \Rightarrow x - \frac{36}{8} = 0.3540\left(y - \frac{107}{8}\right).$$

Correlation coeff :- $r^2 = b_{xy} \times b_{yx}$

$$r = \sqrt{2.7976 \times 0.3540}$$

estimate y when $x = 3.5$.

use regression line y on x :-

$$y - \frac{107}{8} = 2.7976\left(x - \frac{36}{8}\right)$$

put $x = 3.5$ and

Calculate y.

Q2. variance of $x = 9$

Regression eqn $8x - 10y + 66 = 0$ ← y on x

$40x - 18y = 214$ ← x on y

what are the (a) mean values of x & y (b) standard deviation of y and (c) correlation coefficient

Sol → (a) Since both the lines of regression pass through the pts. $(\bar{x}, \bar{y})$ ⇒

$$\left.\begin{array}{l} 8\bar{x} - 10\bar{y} + 66 = 0 \\ 40\bar{x} - 18\bar{y} = 214 \end{array}\right\} \text{solve } \begin{array}{l}\bar{x} = 13 \\ \bar{y} = 17\end{array}$$

(b) $\sigma_x^2 = 9 \Rightarrow \sigma_x = 3.$

$10y = 6.6 + 8x$	$40x = 18y + 214$
$y = \underbrace{0.8}_{b_{yx}}x + 6.6$	$x = \underbrace{0.45}_{b_{xy}}y + 5.35$

Now $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x} \Rightarrow r\frac{\sigma_y}{\sigma_x} = 0.8$

$\cancel{0.8 = r\frac{\sigma_y}{3}}$

$\cancel{\Rightarrow \sigma_y = 1}$

& $r\frac{\sigma_x}{\sigma_y} = 0.45$ } Multiply

$r^2 = 0.8 \times 0.45$

$r^2 = 0.36$

$\boxed{r = 0.6}$

then $\sigma_y = \frac{0.8\,\sigma_x}{r} = \frac{0.8 \times 3}{0.6}$

$\sigma_y = \frac{2.4}{0.6}$

$\boxed{\sigma_y = 4}$

Q→ The following results were obtained from marks in Applied Mechanics and Eng. Mathematics in an examination:-

	Applied Mech (x)	Eng. Mathem (y)
Mean	47.5	39.5
S.D.	16.8	10.8

and $r = 0.95$. find both regression eqn. Also estimate y for $x = 30$.

Sol:
$$\bar{x} = 47.5, \quad \sigma_x = 16.8$$
$$\bar{y} = 39.5, \quad \sigma_y = 10.8$$

$$b_{yx} = r \frac{\sigma_y}{\sigma_x}$$
$$= 0.95 \times \frac{10.8}{16.8} = 0.6107.$$

$$b_{xy} = r \frac{\sigma_x}{\sigma_y}$$
$$= 0.95 \times \frac{16.8}{10.8} = 1.477$$

Reg. line of y on x is
$$y - \bar{y} = b_{yx}(x - \bar{x})$$
$$y = 0.6107x + 10.49 \quad \text{—— (1)}$$

Reg. line of x on y is
$$x - \bar{x} = b_{xy}(y - \bar{y})$$
$$x = 1.477y - 10.8415. \quad \text{—— (2)}$$

put $x = 30$ in (1);
$$y = 28.81.$$

Q-13. The eqⁿ of two regression lines, obtained in a correlation analysis of 60 observations are

$$5x = 6y + 24 \quad \text{and} \quad 1000y = 768x - 3608.$$

What is the correlation coeff.? show that the ratio of coeff. of variability of x to that of y is $\frac{5}{24}$. What is the ratio of variances of x & y?

Sol → Reg. line of x on y is $5x = 6y + 24$

$$x = \frac{6}{5}y + \frac{24}{5}$$

$$b_{xy} = 6/5 \quad \text{—(1)}$$

Reg. line of y on x is $1000y = 768x - 3608$

$$y = 0.768x - 3.608$$

$$b_{yx} = 0.768 \quad \text{—(2)}$$

from (1) & (2)

$$r\frac{\sigma_x}{\sigma_y} = \frac{6}{5} \quad (3), \quad r\frac{\sigma_y}{\sigma_x} = 0.768 \quad \text{—(4)}$$

$$\Rightarrow r^2 = 0.9216$$

$$r = 0.96$$

Now from (3) & (4)

$$\frac{\sigma_x^2}{\sigma_y^2} = \frac{6}{5 \times 0.768} = 1.5625$$

$$\frac{\sigma_x}{\sigma_y} = 1.25 = \frac{5}{4}$$

∴ Regression lines pass through the pts $(\bar{x}, \bar{y})$, we have

$$\left.\begin{aligned} 5\bar{x} &= 6\bar{y} + 24 \\ 1000\bar{y} &= 768\bar{x} - 3608 \end{aligned}\right\} \Rightarrow \bar{x} = 6, \bar{y} = 1.$$

coeff. of variability of $x = \frac{\sigma_x}{\bar{x}} \Rightarrow$

" " " $y = \frac{\sigma_y}{\bar{y}} \Rightarrow$

Required ratio $= \frac{\sigma_x}{\bar{x}} \times \frac{\bar{y}}{\sigma_y} \Rightarrow \frac{\sigma_x}{\sigma_y}\frac{\bar{y}}{\bar{x}} \Rightarrow \frac{1}{6} \times \frac{5}{4} = \frac{5}{24}$ Ans