

4/4/26]

Unit - 6Curve Fitting and Regression.Curve Fitting

→ Method of least square:

This method is useful to fit straight line, second degree polynomial and exponent curves.

Fitting straight line

Let $y = a_0 + a_1 x$ be a straight line where a_0, a_1 are constants

The normal ~~eqns~~ equations are

$$\sum y_i = n a_0 + a_1 \sum x_i$$

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

$n = \text{no. of paired data}$

second degree polynomial.

Let $y = a_0 + a_1 x + a_2 x^2$

The normal equations are

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

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

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

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

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

let $y = a_0 + a_1 x \rightarrow (1)$

Normal equations

$\sum y_i = na_0 + a_1 \sum x_i \rightarrow (2)$

$\sum x_i y_i = a_0 \sum x_i + a_1 \sum x_i^2 \rightarrow (3)$

x_i	y_i	$x_i y_i$	x_i^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25

$\sum x_i = 15$

$\sum y_i = 204$

$\sum x_i y_i = 748$

$\sum x_i^2 = 55$

$(2) \Rightarrow 204 = 5a_0 + 15a_1 \times 3$

$(3) \Rightarrow 748 = 15a_0 + 55a_1$

$612 = 15a_0 + 45a_1$

$136 = 3a_0 + 10a_1$

$a_1 = 13.6$

$a_0 = 10$

$(4) \Rightarrow y = a_0 + a_1 x$

$y = 10 + 13.6x$

$y = 13.6x$

$\sum x_i^2 = 55, \sum x_i y_i = 748, \sum y_i = 204, \sum x_i = 15, n = 5$

8/4/26

Q2 Fit a 2nd degree polynomial to the foll. data.

x 0 1 2 3 4

y 1 1.8 1.3 2.5 6.3

$$y = a_0 + a_1x + a_2x^2$$

Normal equations

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

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

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

x_i	y_i	$x_i y_i$	x_i^2	x_i^3	x_i^4	$x_i^2 y_i$
0	1	0	0	0	0	0
1	1.8	1.8	1	1	1	1.8
2	1.3	2.6	4	8	16	5.2
3	2.5	7.5	9	27	81	22.5
4	6.3	25.2	16	64	256	100.8

$$\sum x_i = 10 \quad \sum y_i = 12.9 \quad \sum x_i y_i = 37.1$$

$$\sum x_i^2 = 30 \quad \sum x_i^3 = 100 \quad \sum x_i^4 = 354 \quad \sum x_i^2 y_i = 130.3$$

$$12.9 = 5a_0 + 10a_1 + 30a_2$$

$$37.1 = 10a_0 + 30a_1 + 100a_2$$

$$130.3 = 30a_0 + 100a_1 + 354a_2$$

$$a_0 = \frac{71}{50} = 1.42$$

$$a_1 = \frac{-107}{100} = -1.07$$

$$a_2 = \frac{11}{20} = 0.55$$

$$y = 1.42 - 1.07x + 0.55x^2$$

Q3 By the method of least square, fit a straight line of the form $a + bx = y$

x	1	2	3	4	5
y	12	25	40	50	65

$$y = a + bx$$

Normal eqns.

$$\sum y_i = na + b \sum x_i$$

$$\sum x_i y_i = a \sum x_i + b \sum x_i^2$$

x_i	y_i	$x_i y_i$	x_i^2
1	12	12	1
2	25	50	4
3	40	120	9
4	50	200	16
5	65	325	25

$$\sum x_i = 15 \quad \sum y_i = 192$$

$$\sum x_i y_i = 707 \quad \sum x_i^2 = 55$$

$$192 = 5a + 15b$$

$$707 = 15a + 55b$$

$$a = -\frac{9}{10} = -0.9$$

$$b = \frac{131}{10} = 13.1$$

$$y = -0.9 + 13.1x$$

Fitting Exponential functions

(i) Fitting $y = ae^{bx}$

$\log_e$ on b.s.

$$\log y = \log(ae^{bx}) \Rightarrow \log y = \log a + \log e^{bx}$$

$$\log y = \log a + bx$$

$$\boxed{Y = A + BX} \rightarrow \textcircled{1}$$

$$Y = \log y \quad A = \log a \quad B = b \quad X = x.$$

(ii) Fitting $y = ab^x$

$\log_{10}$ on b.s.

$$\log_{10} y = \log_{10} ab^x \Rightarrow \log_{10} y = \log_{10} a + x \log_{10} b$$

$$\boxed{Y = A + BX} \rightarrow \textcircled{1}$$

$$Y = \log_{10} y \quad A = \log_{10} a \quad B = \log_{10} b \quad X = x.$$

Q1 Fit a curve of the form $y = ae^{bx}$ to the foll. data.

x	0	1	2	3
y	1.05	2.10	3.85	8.30

Let $\boxed{y = ae^{bx}}$ $\times \log_e$ on b.s.

$$\log_e y = \log_e (ae^{bx})$$

$$\log(y) = \log a + bx.$$

$$Y = A + BX$$

Normal eqn.

$$\sum Y_i = nA + B \sum X_i$$

$$\sum X_i Y_i = A \sum X_i + B \sum X_i^2$$

X_i	Y_i	X_i	$Y_i = \log y$	$X_i Y_i$	X_i^2
0	1.05	0	0.0489	0	0
1	2.10	1	0.7419	0.7419	1
2	3.85	2	1.3480	2.696	4
3	8.30	3	2.1162	6.3486	9

$$\sum X_i = 6 \quad \sum Y_i = 4.2548 \quad \sum X_i Y_i = 9.7865 \quad \sum X_i^2 = 14$$

$$4.2548 = 4A + 6B$$

$$9.7865 = 6A + 14B$$

$$A = 0.04241$$

$$B = 0.68086$$

$$Y = 0.04241 + 0.68086X$$

$$b = B = 0.68086$$

$$a = e^A = e^{0.04241} = 1.0433$$

$$y = 1.0433 e^{0.68086X}$$

9/4/20

Q If a curve of the form $y = ab^x$

x	0	1	2	3	4	5	6	7
y	10	21	35	59	92	200	400	610

$y = ab^x$ $\times \log_{10}$ on b.s.

$\log_{10} y = \log_{10}(ab^x)$

$\log_{10} y = \log_{10} a + x \log_{10} b$

$Y = A + XB$

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

Normal eqns:

$\sum Y_i = nA + B \sum X_i$

$\sum X_i Y_i = A \sum X_i + B \sum X_i^2$

x	y	X=x	Y=log ₁₀ y	X _i Y _i	X _i ²
0	10	0	1	0	0
1	21	1	1.3222	1.3222	1
2	35	2	1.5440	3.0880	4
3	59	3	1.7708	5.3124	9
4	92	4	1.9637	7.8548	16
5	200	5	2.3010	11.5050	25
6	400	6	2.6020	15.6120	36
7	610	7	2.7853	19.4971	49

$\sum X_i = 28$ $\sum Y_i = 15.289$ $\sum X_i Y_i = 64.1915$ $\sum X_i^2 = 140$

$$15.289 = 8A + 28B$$

$$64.1915 = 28A + 140B$$

$$A = 2.91710$$

$$B = 0.254$$

$$A = 1.0211$$

$$B = 0.254$$

$$a = 10^A = 10^{1.0211} = 10.498$$

$$b = 10^B = 10^{0.254} = 1.7947$$

$$\Rightarrow y = (10.498)(1.7947^x)$$

10/4

Regression

Types

① Linear regression

A regression line is a straight line fitted to the data by the method of least squares and it gives relation b/w 2 variables.

(i) Deviations taken from arithmetic mean of x on y

(ii) Deviations taken from y on x

~~The regression line x on y is~~

→ Regression x on y

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

where b_{xy} = regression coefficient x on y

$$b_{xy} = r \frac{\sigma_x}{\sigma_y} = \frac{\sum xy}{\sum y^2}$$

→ Regression y on x

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

$$b_{yx} = r \frac{\sigma_y}{\sigma_x} = \frac{\sum xy}{\sum x^2}$$

Note:

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

Q1 Calculate regression coeff. y on x .
from the data taking deviations from actual means of x and y and estimate value of y when $x = 20$

x	10	12	13	12	16	15
y	40	38	43	45	37	43

$$\bar{x} = \frac{10+12+14+15}{6} = \frac{48}{6} = 8$$

$$\bar{y} = \frac{40+38+43}{6} = 41$$

$$b_{yx} = \frac{\sum xy}{\sum x^2}$$

$$x = x_i - \bar{x}$$

$$y = y_i - \bar{y}$$

		x	y	xy	x^2
x	y				
10	40	-2	-1	2	4
12	38	-4	-3	12	16
13	43	-1	2	-2	1
12	45	-4	4	-16	16
16	37	8	-4	-32	64
15	43	7	2	14	49

$$\sum xy = -6 \quad \sum x^2 = 24$$

$$b_{yx} = \frac{-6}{24} = -0.25$$

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

$$y - 41 = -\frac{6}{24} (20 - 8)$$

$$y - 41 = -\frac{1}{4} (12)$$

$$y - 41 = -3$$

$$y - 41 = -3 \Rightarrow y = 38$$

$$y = 39.25$$

$$b_{yx} = -0.25$$

$$ZF \cdot 0$$

$$\frac{15}{85}$$

$$\frac{y \cdot x}{x^2}$$

$$(\sum y - Fy) \frac{15}{85} = 85 - y \quad (F - x) \cdot y = \bar{y} - \bar{x}$$

$$2 \cdot 22 = y$$

$$2 \cdot y = 85 - y$$

Q8 A panel of 2 judges P, Q graded performances independantly awarding marks as follows

perform.	1	2	3	4	5	6	7
P - $m_1(x)$	46	42	44	40	43	41	45
Q - $m_2(y)$	40	38	36	35	39	37	41

The eighth performance which Q could not attend, was awarded 37 marks by judge P.

If Q had been present, how many marks would be expected to have awarded by him to the 8th performance.

$$\bar{x} = \frac{46 + 42 + \dots}{7} = 43$$

$$\bar{y} = \frac{40 + 38 + \dots}{7} = 38$$

x	y	x	y	xy	x ²
46	40	3	42	+6	9
42	38	-1	0	-2	1
44	36	1	-2	-2	1
40	35	-3	-3	9	9
43	39	0	1	0	0
41	37	-2	-1	2	4
45	41	2	3	6	4
				$\sum xy = 21$	$\sum x^2 = 28$

$$b_{yx} = \frac{\sum xy}{\sum x^2} = \frac{21}{28} = 0.75$$

$$y - \bar{y} = b_{yx} (x - \bar{x}) \Rightarrow y - 38 = \frac{21}{28} (37 - 43)$$

$$y - 38 = -4.5 \Rightarrow y = 33.5$$

Q3 If $b_{xy} = 0.85$ $b_{yx} = 0.89$ $\sigma_x = 3$

(i) Find r

(ii) σ_y

$$(i) r = \sqrt{b_{xy} b_{yx}} = \sqrt{0.85 \times 0.89}$$

$$= 0.8697$$

$$(ii) b_{xy} = r \frac{\sigma_x}{\sigma_y}$$

$$0.85 = 0.8697 \times \frac{3}{\sigma_y}$$

$$\sigma_y = \frac{0.8697 \times 3}{0.85} = 3.069$$

Q4 Find the most likely production corresponding to rainfall 40 from the foll. table

Rainfall (x) prod (y)

Arg

30

500 kgs

SD

5

1000 kgs

r

0.8

$$\sigma_x = 5 \quad \sigma_y = 100$$

$$\bar{x} = 30$$

$$\bar{y} = 500$$

$$b_{yx} = r \frac{\sigma_y}{\sigma_x} = 0.8 \times \frac{100}{5} = 16$$

$$y - \bar{y} = b_{yx} (x - \bar{x}) \Rightarrow y - 500 = 16(40 - 30)$$

$$y = 660 \text{ kgs}$$

→ Regression equation x on y

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

$$b_{xy} = r \frac{\sigma_x}{\sigma_y} = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sum y^2 - \frac{(\sum y)^2}{N}}$$

" " y on x

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

$$b_{yx} = r \frac{\sigma_y}{\sigma_x} = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sum x^2 - \frac{(\sum x)^2}{N}}$$

→ Angle b/w two regression lines.

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

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

$$\tan \theta = \frac{m_1 - m_2}{1 + m_1 m_2}$$

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

Q1 Price of cotton & wool, eqns for line of regression.

	78	77	85	88	87	82	81	77	76	83	97	93
x												
y	84	82	82	85	89	90	88	92	83	89	96	99

Q2 SD of y is twice SD of x , find $\tan \theta$, $r = 0.25$.

Multiple Regression

We know that prod. of rice depends not only on the amt. of rainfall (x_1) but also on the fertilizers (x_2), quality of seeds (x_3) and quality of soil (x_4)

In multiple regression, the independent variable in a function of more than 1 independent variable

→ Let $y = a_0 + a_1 x_1 + a_2 x_2$ ~~be~~ ^{be} the regression line

where a_0, a_1, a_2 are constants.

Normal eqns are given by

$$\sum y = n a_0 + a_1 \sum x_1 + a_2 \sum x_2$$

$$\sum x_1 y = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2$$

$$\sum x_2 y = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2$$

Q Find the least squares regression eqn of x_1 & x_2 from the foll. data.

y	3	5	6	8	12	14
x_1	16	10	7	4	3	2
x_2	90	72	54	42	30	12

let the regression line.

$$y = a_0 + a_1 x_1 + a_2 x_2$$

y	x ₁	x ₂	x ₁ ²	x ₂ ²	x ₁ y	x ₂ y	x ₁ x ₂
3	16	90	256	8100	48	270	1440
5	10	72	100	5184	50	360	720
6	7	54	49	2916	42	324	378
8	4	42	16	1764	32	336	168
12	3	30	9	900	36	360	90
14	2	12	4	144	28	168	24

$$\sum y = 48 \quad \sum x_1 = 42 \quad \sum x_2 = 300$$

$$\sum x_1^2 = 434 \quad \sum x_2^2 = 19008$$

$$\sum x_1 y = 216 \quad \sum x_2 y = 1818$$

$$\sum x_1 x_2 = 2820$$

$$\sum y = na_0 + a_1 \sum x_1 + a_2 \sum x_2$$

$$\sum x_1 y = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2$$

$$\sum x_2 y = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2$$

$$48 = 6a_0 + 42a_1 + 300a_2$$

$$216 = 42a_0 + 434a_1 + 2820a_2$$

$$1818 = 300a_0 + 2820a_1 + 19008a_2$$

$$a_0 = 16.106 \quad a_1 = 0.426 \quad a_2 = -0.221$$

$$y = 16.106 + 0.426x_1 - 0.221x_2$$