20/11/24

Module -4

Regression + Curve fitting

* Curve fitting. (method of least squares)

* Line of best fit

$$\boxed{Y = ax+b}$$

$a\sum_i x_i + nb = \sum_i y_i$

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

$$\boxed{y = a+bx}$$

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

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

* Parabolic curve (Normal)

$Y = ax^2+bx+c$

$a\sum_i x_i^2 + b\sum_i x_i + nc = \sum_i y_i$

$a\sum_i x_i^3 + b\sum_i x_i^2 + c\sum_i x_i = \sum_i x_i y_i$

$a\sum_i x_i^4 + b\sum_i x_i^3 + c\sum_i x_i^2 = \sum_i x_i^2 y$

* exponential curve

$Y = ae^{bx}$

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

$\log y = \log a + \log e^{bx}$

$\log y = \log a + bx$

$Y = A + bx$

$Y = \log y$

$A = \log a$

$a = e^A$

Q Fit a straight line for the following data by the method of least squares. $Y = ax+b$

x:	1	2	3	4	5
y:	7	14	28	56	112

→ $Y = ax+b$ $\quad n = 5$

x	y	x^2	xy
1	7	1	7
2	14	4	28
3	28	9	84
4	56	16	224
5	112	25	560
15	217	55	903

$15a + 5b = 217 \times 3$

$55a + 15b = 903$

$45a + \cancel{15b} = 651$

$55a + \cancel{15b} = 903$

$-10a = -252$

$a = 25.2 = \frac{126}{5}$

$\frac{\cancel{55}^{11} \times 126}{\cancel{5}} + 15b = 903$

$1386 + 15b = 903$

$15b = 483$

$b = \frac{161}{5}$

$b = 32.2$

$a = 25.2$

$b = 32.2$

$y = 25.2x - 32.2$

Q Find the line of best fit $y = a + bx$ by the method of least squares for the following data.

x :	5	10	15	20	25	30
y :	35	75	95	115	145	170

→ $y = a + bx$ $n = 6$

x	y	x^2	xy
5	35	25	175
10	75	100	750
15	95	225	1425
20	115	400	2300
25	145	625	3625
30	170	900	5100
105	635	2275	13375

Eqⁿ $6a + 105b = 635$

$105a + 2275b = 13375$

$a = 15.33$

$b = 5.1714$

$$y = 15.33 + 5.1714x.$$

Q Fit a parabola $y = ax^2 + bx + c$ for the following data using method of least squares

x ;	1	2	3	4	5
y :	1	4	9	16	25

→ $y = ax^2 + bx + c$ $n = 5$

x	y	x^2	x^3	x^4	xy	x^2y
1	1	1	1	1	1	1
2	4	4	8	16	8	16
3	9	9	27	81	27	81
4	16	16	64	256	64	256
5	25	25	125	625	125	625
15	55	55	225	979	225	979

$55a + 15b + 5c = 55$

$225a + 55b + 15c = 225$

~~$225a$~~

$979a + 225b + 55c = 979.$

$a = 1$

$b = 0$

$c = 0$

$y = x^2$

Q. Fit a parabola $y = a + bx + cx^2$ for the following data ~~using method of least squares~~.

$x = 1 \quad 2 \quad 3 \quad 4 \quad 5$

$y = 1.2 \quad 4.1 \quad 9 \quad 16.2 \quad 25.3$

x	y	x^2	x^3	x^4	xy	x^2y
1	1.2	1	1	1	1.2	1.2
2	4.1	4	8	16	8.2	16.4
3	9	9	27	81	27	81
4	16.2	16	64	256	64.8	259.2
5	25.3	25	125	625	126.5	632.5
15	55.8	55	225	979	227.7	990.3

$5a + 15b + 55c = 55.8$

$15a + 55b + 225c = 227.7$

$55a + 225b + 979c = 990.3$

$a = 0.42$

$b = -0.27$

$c = 1.05.$

$y = 0.42 - 0.27x + 1.05x^2$

Q Predict a mean radiation of at an altitude of 3000 feet by LSM exponential curve for the following data

Altitude	x	50	450	780	1200	4400	4800	5300
dose of radiation	y	28	30	32	36	51	58	69

$n = 7$

$y = ae^{bx}$

$\log y = \log a + bx$

$Y = A + bx$

where $Y = \log y$ & $A = \log a$

x	y	$Y = \log y$	x^2	xY
50	28	3.33	2500	166.5
450	30	3.40	202500	1530
780	32	3.46	608400	2698.8
1200	36	3.58	1440000	4296
4400	51	3.93	19360000	17292
4800	58	4.06	23040000	19488
5300	69	4.23	28090000	22419
16980		25.99	72743400	67890.3

$$nA + b\sum_i x_i = \sum_i Y_i$$

$$A\sum_i x_i + b\sum_i x_i^2 = \sum_i x_i Y_i$$

$$7A + 16980b = 25.99$$

$$16980A + 72743400b = 67890.3$$

$A = 3.34$ $\quad b = 0.00015$

$a = e^{3.34} = 28.21$

$$y = 28.21\, e^{0.00015x}$$

Q A Voltage V across a capacitor at time t secs is given by following table using the method of least square to fit a curve of form $V = ae^{bt}$

x t	0	2	4	6	8
y V	150	63	28	12	5.6

→ $y = a e^{bt}$

$\log y = \log a + bt$ $n = 5$

$Y = A + bt$

when $Y = \log y$ & $A = \log a$

t	y	$Y = \log y$	x^2	xY
0	150	5.010	0	0
2	63	4.143	4	8.286
4	28	3.33	16	13.32
6	12	2.484	36	14.904
8	5.6	1.722	64	13.776
20		16.689	120	50.286

$nA + b\sum t_i = \sum_i y_i$

$A\sum_i t_i + b\sum_i t_i^2 = \sum_i t_i y_i$

$5A + 20b = 16.689$

$20A + 120b = 50.286$

$A = 4.98$

$b = -0.411$

$a = e^{4.98} = 145.47$

$V = 145.4\, e^{-0.411t}$

* Correlation and regression

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

Q Find the ~~correlatio~~ co-efficient of correlation for following data:

x	10	14	18	22	26	30
y	18	12	24	6	30	36

In the probl also find regression lines.

→

x	y	x^2	y^2	xy
10	18	100	324	180
14	12	196	144	168
18	24	324	576	432
22	6	484	36	132
26	30	676	900	780
30	36	900	1290	1080
120	126	2680	3270	2772

$$r = \frac{6(2772) - (120)(126)}{\sqrt{2680 \times 6 - (120)^2}\sqrt{3276 \times 6 - (126)^2}}$$

$$r = \frac{16632 - 15120}{\sqrt{16080 - 14400}\sqrt{}}$$

$$r = \frac{1512}{\sqrt{1680}\sqrt{3780}} = \frac{1512}{(40.98)(61.48)}$$

$$r = \frac{1512}{2519.5}$$

$$r = 0.6001$$

Y on X $\quad y-\bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$ $\quad y = ax+b$

X on Y $\quad x-\bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$ $\quad x = cy+d$

or

$\bar{x} = \frac{\Sigma x}{n}$ $\quad \sigma_x = \sqrt{\frac{1}{n}\Sigma(x^2)-(\bar{x})^2}$

$\bar{y} = \frac{\Sigma y}{n}$ $\quad \sigma_y = \sqrt{\frac{1}{n}\Sigma y^2-(\bar{y})^2}$

Regression on Y

$$y - \frac{126}{6} = 0.6\left(\frac{10.25}{6.83}\right)\left(x-\frac{120}{6}\right)$$

$$Y-21 = 0.9(x-20)$$

$$Y-21 = 0.9x - 18$$

$$Y = 0.9x - 18+21$$

$$Y = 0.9x+3$$

$\sigma_x = \sqrt{\frac{1}{6}(2680)-(20)^2}$

$\sigma_x = 6.83$

$\sigma_y = \sqrt{\frac{1}{6}(3276)-441}$

$\sigma_y = 10.25$

Regression X on Y

$$X-20 = 0.6\left(\frac{6.83}{10.25}\right)(y-21)$$

$$X-20 = 0.399y - 8.396$$

$$X = 0.399y - 8.396+20$$

$$X = 0.399y + 11.6$$

$$X = 0.4y+11.6$$

$$Y-0.9x = 3$$

$$X-0.4y = 11.6$$

$Y = 21$

$X = 20$

Q Find coefficient of Correlation regression lines of following data

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

x	y	x^2	y^2	xy
1	2	1	4	2
2	5	4	25	10
3	3	9	9	9
4	8	16	64	32
5	7	25	49	35
15	25	55	151	88

$$r = \frac{5(88) - (15)(25)}{\sqrt{5(55) - (225)}\ \sqrt{5(151) - (25)^2}}$$

$$r = \frac{440 - 375}{\sqrt{50}\ \sqrt{130}}$$

$$r = \frac{65}{(7)(11.4)} = \frac{65}{80.610} = \underline{0.806}$$

* $\bar{x} = \frac{\Sigma x}{n} \rightarrow \sigma_x = \sqrt{\frac{1}{5}(55) - (3)^2} = 1.414$

* $\bar{x} = \frac{15}{5} = 3$

* $\bar{Y} = \frac{\Sigma y}{n} = \frac{25}{5} = 5 \rightarrow \sigma_y = \sqrt{\frac{1}{5}(151) - 25} = \sqrt{5.2} = 2.280$

Regression on y

$\Rightarrow y - 5 = 0.806 \frac{2.280}{1.414}(x - 3)$

$y - 5 = 1.29x - 3.89$

$y = 1.29x + 1.10$

$y - 1.29x = 1.10$

Regression x on Y

$$x - 3 = 0.806 \frac{1.414}{2.280} (y-5)$$

$$x - 3 = 0.49y - 2.49$$

$$x = 0.49y + 0.5$$

$$x - 0.49y = +0.5$$

Q In Partially destroyed laboratory record of an analysis of correlation data following results are only

$$8x - 10y + 66 = 0$$

$$40x - 18y - 214 = 0$$

Find Correlation co-efficient mean of x and Y

→ $8x - 10y = 66 \rightarrow x = \frac{10}{8}y - \frac{66}{8}$, $b_{xy} = \frac{10}{8}$

$40x - 18y = 214 \rightarrow y = \frac{40x}{18} - \frac{214}{18}$, $b_{yx} = \frac{40}{18}$.

$\bar{x} = 13$ $\bar{y} = 17$

$b_{xy}\, b_{yx} = \sqrt{\frac{400}{144}} = \frac{20}{12} = 1.66 > 1$

eqn Not possible

change eqn

$y = \frac{8}{10}x + \frac{66}{100}$, $b_{yx} = \frac{8}{10}$

$x = \frac{18}{40}y + \frac{214}{40}$, $b_{xy} = \frac{18}{40}$

$$r = \sqrt{\frac{144}{400}} = 0.6 < 1$$

eqn is possible

$x = -0.103$

$y = 0.517$.

Q $7x - 16y + 9 = 0 \rightarrow y = \frac{7x}{16} + \frac{9}{16}$, $b_{yx} = \frac{7}{16}$

$5y - 4x - 3 = 0 \rightarrow x = \frac{5y}{4} - \frac{3}{4}$, $b_{xy} = \frac{5}{4}$

$$\sqrt{b_{xy}\, b_{yx}} = \sqrt{\frac{7}{16} \times \frac{5}{4}} = \sqrt{\frac{35}{64}} = \sqrt{0.54} = 0.73$$

Q. $3x + 2y = 26 \rightarrow x = \frac{26}{3} - \frac{2y}{3}$, $b_{xy} = -\frac{2}{3}$

$6x + y = 31 \rightarrow y = 31 - 6x$, $b_{yx} = -6$

$\bar{x} = 4 \quad \bar{y} = 7$

$$\sqrt{b_{xy}\, b_{yx}} = \sqrt{4} = \underline{2 > 1}$$

eqⁿ is not possible

$3x + 2y = 26 \rightarrow y = \frac{26}{2} - \frac{3x}{2}$, $b_{yx} = \frac{-3}{2}$

$6x + y = 31 \rightarrow x = +\frac{31}{6} - \frac{y}{6}$, $b_{xy} = -\frac{1}{6}$

$$\sqrt{b_{xy}\, b_{yx}} = \sqrt{\frac{-3}{2} \times -\frac{1}{6}} = \sqrt{\frac{3}{12}} = 0.5 < 1$$

<u>eqⁿ is possible</u>