|| Jai Sri Gurudev ||

S. J. C institute of technology, Chickballapur

Department of Mathematics

Lecture notes

Complex Analysis, Probability and statistical Methods [18MAT41]

Prepared by:

Purushotham. P

Assistant Professor

SJC institute of technology

Email id: ppurushothamp48@gmail.com

phone: 8197481658.

MODULE - 04

CURVE FITTING

It is a mathematical procedure for finding the best fitted curve, to the given set of points x and y, by minimizing the sum of squares of the offsets of the points from the curve.

Fitting of a straight line $y = a + bx$

1. Consider the straight line to be fitted $y = a + bx \rightarrow (1)$
2. Write the normal equation for the given straight line as $\Sigma y = na + b\Sigma x \rightarrow (2)$, $\Sigma xy = a\Sigma x + b\Sigma x^2 \rightarrow (3)$
3. Find out $\Sigma x, \Sigma y, \Sigma x^2, \Sigma xy$ from the given table of x and y of values and apply them in the given normal equation.
4. Solve the normal equation and opting a, b as $\hat{a}, \hat{b}$ and substitute the same in the required straight line.
5. The fitted straight line becomes $y = \hat{a} + \hat{b}x$.

1. Fit a straight line of the form $y = a + bx$ for the data

x	-2	-1	0	1	2
y	1	2	3	3	4.

Let the straight line to be fitted $y = a + bx \rightarrow (1)$

$\therefore$ The normal equations for the curve are

$\Sigma y = na + b\Sigma x \rightarrow (2)$ $\quad \Sigma xy = a\Sigma x + b\Sigma x^2 \rightarrow (3)$

x	y	xy	x^2
-2	1	-2	4
-1	2	-2	1
0	3	0	0
1	3	3	1
2	4	8	4
$\Sigma x = 0$	$\Sigma y = 13$	$\Sigma xy = 7$	$\Sigma x^2 = 10$

$n = 5$

$(2) \Rightarrow 13 = 5a + b(0)$ $\qquad (3) \Rightarrow 7 = a(0) + b(10)$

$a = 13/5$ $\qquad 7/10 = b$

$a = 2.6$ $\qquad b = 0.7$

$y = 2.6 + (0.7)x$

2) Fit a straight line $y = a + bx$ for the data

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

Let the straight line to be fitted $y = a + bx \rightarrow (1)$

The normal equation for the curve are

$\Sigma y = na + b\Sigma x \rightarrow (2)$, $\Sigma xy = a\Sigma x + b\Sigma x^2 \rightarrow (3)$

x	y	xy	x^2
1	2	2	1
2	5	10	4
3	3	9	9
4	8	32	16
5	7	35	25
$\Sigma x = 15$	$\Sigma y = 25$	$\Sigma xy = 88$	$\Sigma x^2 = 55$

$(2) \Rightarrow 5a + 15b = 25$

$a + 3b = 5 \rightarrow (1)$

$88 = 15a + 55b \rightarrow (2)$

$a = 1.1$, $b = 1.3$

$y = 1.1 + 1.3x$

3) Fit a straight of the form $y=ax+b$ for the data

x	12	15	21	25
y	50	70	100	120.

Let the straight line to be fitted $= ax+b \rightarrow (1)$

The normal equations of the curve are $\Sigma y = a\Sigma x + nb \rightarrow (2)$

$\Sigma xy = a\Sigma x^2 + b\Sigma x \rightarrow (3)$

x	y	xy	x^2
12	50	600	144
15	70	1050	225
21	100	2100	441
25	120	3000	625
$\Sigma x = 73$	340	6750	1435

$n = 4.$

(2) $\Rightarrow 73a + 4b = 340$ (3) $\Rightarrow 1435a + 73b = 6750$

$a = 5.304$ $b = -11.800$

$y = 5.304x - 11.8.$

4). Fit a straight line of the form $y=ax+b$ for the data

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

Let the straight line to be fitted $= ax+b \rightarrow (1)$

The normal equation of the curve are $\Sigma y = a\Sigma x + nb \rightarrow (2)$

$\Sigma xy = a\Sigma x^2 + b\Sigma x \rightarrow (3)$

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

$n = 5$

(2) $\Rightarrow 15a + 5b = 204$ (3) $\Rightarrow 55a + 15b = 748$

$a = 13.6$ $b = 0$

$y = 13.6x$

5) Fit a straight line of the form $y = a + bp$ for the data

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

hence estimate $P = 150$.

$$Y = a + bp \rightarrow (1)$$

The normal equation of the curve are

$$\Sigma y = na + b\Sigma P \rightarrow (2)$$

$$\Sigma Py = a\Sigma P + b\Sigma P^2 \rightarrow (3)$$

P	Y	PY	P^2
100	0.45	45	10000
120	0.65	66	14400
140	0.55	77	19600
160	0.70	112	25600
180	0.80	144	32400
200	0.85	170	40000
$\Sigma P = 900$	3.9	614	142000

(2) $\Rightarrow 6a + 900b = 3.9$ (3) $\Rightarrow 900a + 142000b = 614$

$a = 0.0285$ $b = 0.004142$

$$Y = 0.0285 + 0.004142\,P$$

when $P = 150$

$$Y = 0.0285 + 0.004142(150)$$

$$Y = 0.6498$$

6) Fit a straight line of the form $P = m\omega + c$ for the data

P	12	15	21	25
ω	50	70	100	120.

Let the straight line $P = m\omega + c \rightarrow (1)$

The normal equation of the curve are $\Sigma P = m\Sigma\omega + cn \rightarrow (2)$

$\Sigma P\omega = m\Sigma\omega^2 + c\,\Sigma\omega \rightarrow (3)$

P	ω	Pω	ω^2
12	50	600	2500
15	70	1050	4900
21	100	2100	10000
25	120	3000	14400
$\Sigma P = 73$	$\Sigma\omega = 340$	6750	$\Sigma\omega^2 = 31800$

(2) $\Rightarrow 340m + 4C = 73$ (3) $\Rightarrow 31800m + 340C = 6750$

$a = 0.1879$ $b = 2.2758$

$$P = 0.1879\omega + 2.2758$$

Fitting of second degree parabola of the form $y = ax^2+bx+c$ or $y = a+bx+cx^2$.

1. Let the curve to be fitted $y = a+bx+cx^2 \rightarrow (1)$

2. Let the normal equation of the curve by principle of least square is defined as.

$\Sigma y = na + b\Sigma x + c\Sigma x^2 \rightarrow (2)$

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \rightarrow (3)$

$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \rightarrow (4)$

3. Solve the above linear equation by using the given set of x and y values.

$\therefore$ The fitted second degree parabola $y = \hat{a} + \hat{b}x + \hat{c}x^2$.

1) Fit a best fitting of $y = a+bx+cx^2$ for the following data

x	-2	-1	0	1	2
y	-3.15	-1.390	0.620	2.880	5.378

Let the curve to be fitted is $y = a+bx+cx^2 \rightarrow (1)$

where a, b, c are constants

$\therefore$ The normal equation of the curve is defined as

$\Sigma y = na + b\Sigma x + c\Sigma x^2 \rightarrow (2)$

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \rightarrow (3)$

$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \rightarrow (4)$

x	y	xy	x^2	x^2y	x^3	x^4
-2	-3.15	6.3	4	-12.6	-8	16
-1	-1.39	1.39	1	-1.39	-1	1
0	0.620	0	0	0	0	0
1	2.880	2.880	1	2.880	1	1
2	5.378	10.756	4	21.512	8	16
$\Sigma x = 0$	$\Sigma y = 4.338$	21.326	10	10.402	0	34

(2) $\Rightarrow 5a + 0b + 10c = 4.338$

(3) $\Rightarrow 0a + 10b + 0c = 21.326$

(4) $\Rightarrow 10a + 0b + 34c = 10.402$

$a = 0.6210 \qquad b = 2.1326 \qquad c = 0.12328$

$$y = 0.6210 + (2.1326)x + (0.1232)x^2$$

2) Fit a second degree parabola of the form ax^2+bx+c for the data given below

x	1	2	3	4	5
y	10	12	13	16	9

Let the curve to be fitted is $y=ax^2+bx+c \rightarrow ①$

The normal equation of the curve are defined as

$\Sigma y = a\Sigma x^2 + b\Sigma x + nc \rightarrow ②$

$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x \rightarrow ③$

$\Sigma x^2y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 \rightarrow ④$

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

② $\Rightarrow 55a + 15b + 5c = 70$

③ $\Rightarrow 225a + 55b + 15c = 232$

④ $\Rightarrow 979a + 225b + 55c = 906$

$a = 0.2857$ $\quad b = 0.4857$ $\quad c = 9.4$

① $\Rightarrow y = (0.2857)x^2 + (0.4857)x + 9.4.$

3) Fit a second degree parabola of the form $y=ax^2+bx+c$ for the given data

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

Let the curve to be fitted $y = ax^2+bx+c \rightarrow ①$

The normal equation of the curve.

$\Sigma y = a\Sigma x^2 + b\Sigma x + nc \rightarrow ②$

$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x \rightarrow ③$

$\Sigma x^2y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 \rightarrow ④$

x	y	xy	x^2	x^3	x^4	x^2y
1	2	2	1	1	1	2
2	6	12	4	8	16	24
3	7	21	9	27	81	63
4	8	32	16	64	256	128
5	10	50	25	125	625	250
6	11	66	36	216	1296	396
7	11	77	49	343	2401	539
8	10	80	64	512	4096	640
9	9	81	81	729	6561	729
45	74	421	285	2025	15333	2771

(2) $\Rightarrow 285a + 45b + 9c = 74$

(3) $\Rightarrow 2025a + 285b + 45c = 421$

(4) $\Rightarrow 15333a + 2025b + 285c = 2771$

$a = -0.267316$, $b = 3.52316$ $c = -0.928571$

(1) $\Rightarrow y = (-0.267316)x^2 + (3.52316)x - 0.928571.$

4) Following are the measurements of air velocity (x) and evaporation co-efficient (y) of burning fuel droplets in an impulse engine. Find a best fitting parabola to the above data and hence estimate y when $x = 99$

x	20	60	100	140	180	220
y	0.18	0.37	0.35	0.78	0.56	0.75

Let the curve to be fitted $y = ax^2 + bx + c \rightarrow$ (1)

The normal equations of the curve are

$\Sigma y = a\Sigma x^2 + b\Sigma x + nc \rightarrow$ (2)

$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x \rightarrow$ (3)

$\Sigma x^2 y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 \rightarrow$ (4)

x	y	xy	x^2	x^3	x^4	x^2y
20	0.18	3.6	400	8000	160000	72
60	0.37	22.2	3600	216000	12960000	1332
100	0.35	35	10000	1000000	100000000	3500
140	0.78	109.2	19600	2744000	384160000	15288
180	0.56	100.8	32400	5832000	1049760000	18144
220	0.75	165	48400	10648000	2342560000	36300
720	2.99	435.8	114400	20448000	3889600000	74636

② $\Rightarrow 114400a + 720b + 6c = 2.99$

③ $\Rightarrow 20448000a + 114400b + 720c = 435.8$

④ $\Rightarrow 3889600000a + 20448000b + 114400c = 74636$

$a = -0.00000892857$, $b = 0.0048928571142$, $c = 0.0814285714 28$

$$y = -0.0000089 2857x^2 + 0.0048928571142x + 0.0814285714 28$$

5) Fit a best fit a second degree parabola for the data.

x	-3	-2	-1	0	1	2	3
y	4.63	2.11	0.67	0.09	0.63	2.15	4.58

Let the curve to be fitted $y = ax^2 + bx + c \rightarrow$ ①

The normal equations for the curve are

$\Sigma y = a\Sigma x^2 + b\Sigma x + nc \rightarrow$ ②

$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x \rightarrow$ ③

$\Sigma x^2y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 \rightarrow$ ④

x	y	xy	x^2	x^3	x^4	x^2y
-3	4.63	-13.89	9	-27	81	41.67
-2	2.11	-4.22	4	-8	16	8.44
-1	0.67	-0.67	1	-1	1	0.67
0	0.09	0	0	0	0	0
1	0.63	0.63	1	1	1	0.63
2	2.15	4.3	4	8	16	8.60
3	4.58	13.74	9	27	81	41.22
0	14.86	-0.08	28	0	196	101.23

② $\Rightarrow 28a + 0b + 7c = 14.86$

③ $\Rightarrow 0a + 28b + 0c = -0.08$

④ $\Rightarrow 196a + 0b + 28c = 101.23$

By solving $\Rightarrow a = 0.4975, b = -0.00285, c = 0.132$

$$\Rightarrow y = 0.4975x^2 - 0.00285x - 0.132.$$

Fitting of the Power curve of the form $y = a \cdot x^b$

1. Let $y = a \cdot x^b \rightarrow ①$ be the power curve where a and b are constants.
2. Take log on both sides

$$\log y = \log (a \cdot x^b)$$
$$\log y = \log a + \log x^b$$
$$\log y = \log a + b \log x$$
$$Y = A + BX \rightarrow ②$$

where $Y = \log_e y$, $X = \log_e x$, $A = \log_e a \Rightarrow a = e^A$, $B = b$

3. Estimate A and B by the principles of least square and fit the power curve $y = \hat{a} \cdot x^{\hat{b}}$.

1) Fit a curve of the form $y = a \cdot x^b$ for the data

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

Let the curve to be fitted $y = a \cdot x^b \rightarrow ①$ where a and b are constants.

By taking log on both sides we get $Y = A + BX \rightarrow ②$ where $Y = \log_e y$, $X = \log_e x$, $A = \log_e a$, $a = e^A$, $B = b$.

The normal eqⁿ of the curve are $\Sigma Y = nA + B\Sigma X \rightarrow ③$
$\Sigma XY = A\Sigma X + B\Sigma X^2 \rightarrow ④$

x	y	$X = \log_e x$	$Y = \log_e y$	XY	X^2
1	0.5	0	-0.6931	0	0
2	2	0.6931	0.6931	0.4803	0.4803
3	4.5	1.0986	1.5040	1.6522	1.2069
4	8	1.3862	2.0794	2.8824	1.9215
5	12.5	1.6094	2.5257	4.0648	2.5901
15	27.5	4.7873	6.1091	9.0797	6.1988

③ $\Rightarrow 5A + 4.7873B = 6.1091$

④ $\Rightarrow 4.7873A + 6.1988B = 9.0797$

$A = -0.6932$ $\quad B = 2.000$

$a = e^{A}$

$= e^{-0.6932}$

$= 0.4999$

$= 0.5$

① $\Rightarrow$ $y = (0.5)x^{2}$

2) 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.

Let the curve to be fitted $y = a \cdot x^{b} \rightarrow$ ① where a and b are constants.

By taking log on both sides with base e $\quad Y = A + BX \rightarrow$ ②

where $Y = \log_e y$, $X = \log_e x$, $A = \log_e a$, $a = e^{A}$, $B = b$.

The normal equation of the curve are

$\Sigma Y = nA + B\Sigma x \rightarrow$ ③

$\Sigma XY = A\Sigma X + B\Sigma X^{2} \rightarrow$ ④

x	y	$X = \log_e x$	$Y = \log_e y$	XY	X^2
1	2.98	0	1.0919	0	0
2	4.26	0.6931	1.4492	1.0044	0.4803
3	5.21	1.0986	1.6505	1.8132	1.2069
4	6.1	1.3862	1.8082	2.5065	1.9215
5	6.8	1.6094	1.9169	3.0850	2.5901
6	7.5	1.7917	2.0149	3.6100	3.2101
		6.5790	9.93162	12.0197	9.4089

③ $\Rightarrow$ $6A + 6.5790B = 9.93162$

④ $\Rightarrow$ $6.5790A + 9.4089B = 12.0197$

$A = 1.0909$ $\quad B = b = 0.5149$

$a = e^{A}$

$= e^{1.0909}$

$= 2.976$

① $\Rightarrow$ $Y = (2.976)x^{0.514}$.

3) Fit a curve of the form $y = a \cdot x^b$ for the data

x	2	3	4	5
y	27.8	62.1	110	161.

Let the curve to be fitted $y = a \cdot x^b \rightarrow (1)$ where a and b are constants.

By taking log on both sides $Y = A + BX \rightarrow (2)$

where $Y = \log_e y$, $X = \log_e x$, $A = \log_e a$, $a = e^A$, $B = b$

The normal equation for the curve are

$\Sigma Y = nA + B\Sigma X \rightarrow (3)$ $\Sigma XY = A\Sigma X + B\Sigma X^2 \rightarrow (4)$

x	y	$X = \log_e x$	$Y = \log_e y$	XY	X^2
2	27.8	0.6981	3.3250	2.3211	0.4873
3	62.1	1.0986	4.1287	4.5357	1.2069
4	110	1.3862	4.7004	6.5156	1.9215
5	161	1.6094	5.0814	8.1780	2.5901
		4.7923	17.2355	21.5504	6.2058

(3) $\Rightarrow 4A + 4.7923B = 17.2355$

(4) $\Rightarrow 4.7923A + 6.2058B = 21.5504$.

$A = 1.9838$ $B = b = 1.9406$

$a = e^{1.9838}$.

$a = 7.27 \simeq 7.3$

(1) $\Rightarrow y = (7.3)x^{1.9406}$

4) Find the best values of a and b by fitting the law $V = at^b$, using the method of least squares.

V	350	400	500	600
t	61	26	7	26.

Let the curve to be fitted $V = at^b \rightarrow (1)$

By taking log on both sides we get (1) $\Rightarrow V = A + BT \rightarrow (2)$

The normal equations for the curve are.

$\Sigma V = nA + B\Sigma T \rightarrow (3)$

$\Sigma VT = A\Sigma T + B\Sigma T^2 \rightarrow (4)$

where $T = \log_e t$, $V = \log_e v$, $A = \log_e a \Rightarrow a = e^A$, $B = b$.

t	v	$T = \log_e t$	$V = \log_e v$	TV	T^2
61	350	4.1108	5.8579	24.0806	16.8986
26	400	3.2580	5.9914	19.5199	10.6145
7	500	1.9459	6.2146	12.0929	3.7865
26	600	3.2580	6.3969	20.8411	10.6145
		12.5727	24.4608	76.5345	41.9141

③ $\Rightarrow 4A + 12.5727B = 24.4608$

④ $\Rightarrow 12.5727B + 41.9141B = 76.5345$

By solving $\Rightarrow$ $A = 6.5744$ $\qquad B = b = -0.1461$

$a = e^{6.5744}$

$a = 716.5155$

① $\Rightarrow$ $V = 716.5155\, t^{-0.1461}$

Correlation and Regression

The co-variant between the two individuals or magnitudes is called the correlation

Karl Pearson's coefficient of correlation:

Suppose x and y be the two individuals, then the Karl Pearson's coefficient of correlation between x and y is denoted by r is defined as.

$$r = \frac{Cov(x,y)}{n\sigma_x \sigma_y}$$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

where $\sigma_x = \sqrt{\frac{1}{n}\Sigma(x-\bar{x})^2}$ $\qquad \sigma_y = \sqrt{\frac{1}{n}\Sigma(y-\bar{y})^2}$

Note:

1) The correlation coefficient always should be in between -1 and $+1$, i.e., $-1 \le r \le +1$.

2) If $r > 0$, we can say that the relationship between the individuals is good or strong.

3) If $r<0$, we can say that the relationship between the individuals is poor or weak.

Regression:

The regression lines are two types they are i) y on x ii) x on y

Regression line of y on x:

The equation of the form $(y-\bar{y}) = b_{yx}(x-\bar{x}) \Rightarrow (y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$ is called the regression line of y on x.

where $\bar{x}, \bar{y}$ are the means of x and y respectively and the regression coefficient of y on x is $b_{yx} = r\frac{\sigma_y}{\sigma_x}$.

Regression line of x on y:

The equation of the form $(x-\bar{x}) = b_{xy}(y-\bar{y}) \Rightarrow (x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$ is called the regression line of x on y. where $\bar{x}, \bar{y}$ are the means of x and y respectively and the regression coefficient of x on y is $b_{xy} = r\frac{\sigma_x}{\sigma_y}$

we can also calculate the correlation coefficient with the regression lines by using $r = \pm\sqrt{b_{xy} \times b_{yx}}$

Problems:

1) a) In a bivariate distribution, it is found $\sigma_x = \sigma_y$ that and the acute angle between the lines of regression is $\tan^{-1}(3)$. Find the correlation coefficient.

b) If is the acute angle between the lines of regression, then show that $\tan\theta = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left[\frac{1-r^2}{r}\right]$. Explain the significance when $r=0$ and $r=\pm 1$.

Soln Case-1: W.K.T the regression line of y on x is

$$y-\bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \rightarrow (1)$$

eqn (1) is of the form $y-y_0 = m(x-x_0)$

$\therefore$ The slope of the eqn (1) is $m_1 = r\frac{\sigma_y}{\sigma_x}$

W.K.T the regression line of x on y is

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

$$y - \bar{y} = \frac{\sigma_y}{r\sigma_x}(x - \bar{x}) \rightarrow (2)$$

eqⁿ (2) is of the form $y - y_0 = m(x - x_0)$

$\therefore$ The slope of the eqⁿ (1) is $m_2 = \frac{\sigma_y}{r\sigma_x}$

W.K.T

$$\tan\theta = \left|\frac{m_1 - m_2}{1 + m_1 m_2}\right| \Rightarrow \tan\theta = \left|\frac{r\frac{\sigma_y}{\sigma_x} - \frac{\sigma_y}{r\cdot\sigma_x}}{1 + r\frac{\sigma_y}{\sigma_x} \times \frac{\sigma_y}{r\sigma_x}}\right|$$

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

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

$$\tan\theta = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\left[\frac{r^2 - 1}{r}\right] \rightarrow (3)$$

i) when $r = 0$

(3) $\Rightarrow \tan\theta = \infty \Rightarrow \theta = \tan^{-1}(\infty) = \pi/2$

when $r = \pm 1$

$\tan\theta = 0 \Rightarrow \theta = \tan^{-1}(0) = 0$

when $\theta = \tan^{-1}(3) = \sigma_x = \sigma_y$

(3) $\Rightarrow \tan(\tan^{-1} 3) = \left|\frac{r^2 - 1}{r}\right| \frac{\sigma^2 x}{2\sigma^2 x}$

$$3 = \frac{r^2 - 1}{2r}$$

$r^2 - 1 = 6r$

$r^2 - 6r - 1 = 0$

$$r = \frac{6 \pm \sqrt{36 - 4(1)(-1)}}{2(1)} = \frac{6 \pm \sqrt{40}}{2(1)}$$

$$r = \frac{6 \pm 2\sqrt{10}}{2}$$

$r = 3 \pm \sqrt{10}$

$r = 3.1622 \qquad r = -0.1622.$

2) The following table gives the heights of fathers (x) and sons (y). Calculate the coeffecient of correlation and lines of regression.

x	65	66	67	67	68	69	70	72
y	67	68	65	68	72	72	69	71.

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
65	67	-3	-2	6	9	4
66	68	-2	-1	2	4	1
67	65	-1	-4	4	1	16
67	68	-1	-1	1	1	1
68	72	0	3	0	0	9
69	72	1	3	3	1	9
70	69	2	0	0	4	0
72	71	4	2	8	16	4
544	552			24	36	44

$\bar{x} = \frac{544}{8} = 68, \qquad \bar{y} = \frac{552}{8} = 69$

W.K.T

$$r = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2}\sqrt{\Sigma (y-\bar{y})^2}}$$

$$r = \frac{24}{\sqrt{36}\sqrt{44}} = \frac{24}{6\sqrt{44}} = \frac{4}{\sqrt{44}}$$

$r = 0.6030$

i) The regression line y on x

$\sigma_x = \sqrt{\frac{1}{n}\Sigma (x-\bar{x})^2} = \sqrt{\frac{36}{8}} = 2.1213 \qquad \sigma_y = \sqrt{\frac{1}{n}\Sigma (y-\bar{y})^2} = \sqrt{\frac{44}{8}} = 2.3452$

① $\Rightarrow y-\bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$

$$y - 69 = \frac{(0.6030)(2.3452)}{2.1213}(x-68)$$

$$y = (0.6666)x + 23.6680$$

$$(0.6666)x - y + 23.6680 = 0$$

ii) regression line x on y is

$$(x-\bar{x}) = r\frac{\sigma x}{\sigma y}(y-\bar{y})$$

$$x-68 = \frac{(0.6030)(2.1213)}{2.3452}(y-69)$$

$$x-0.5454y-30.3652=0.$$

3) Find the coefficient of correlation and the eqn of regression lines for the data

x	1	2	3	4	5	6	7	8	9
y	12	11	13	15	14	17	16	19	18.

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	12	-4	-3	12	16	9
2	11	-3	-4	12	9	16
3	13	-2	-2	4	4	4
4	15	-1	0	0	1	0
5	14	0	-1	0	0	1
6	17	1	2	2	1	4
7	16	2	1	2	4	1
8	19	3	4	12	9	16
9	18	4	3	12	16	9
45	135			56	60	60

$$\bar{x} = \frac{\sigma x}{n} = \frac{45}{9} = 5 \qquad \bar{y} = \frac{\sigma y}{n} = \frac{135}{9} = 15$$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} = \frac{56}{\sqrt{60}\sqrt{60}} = \frac{56}{4\sqrt{15}\sqrt{15}} = \frac{14}{\sqrt{15}\sqrt{15}} = 0.9333$$

$$\sigma x = \sqrt{\frac{1}{n}\Sigma(x-\bar{x})^2} = 2.5819 \qquad \sigma y = \sqrt{\frac{1}{n}\Sigma(y-\bar{y})^2} = 2.5819$$

i) $(y-\bar{y}) = r\frac{\sigma y}{\sigma x}(x-\bar{x})$

$$y-60 = \frac{(0.9333)(2.5819)}{2.5819}(x-60)$$

$$y-60 = 0.9333x-56 \Rightarrow 0.933x-y+4=0.$$

ii) $(x-\bar{x}) = r\dfrac{\sigma x}{\sigma y}(y-\bar{y})$

$$x - 60 = \frac{0.9333(2.5819)}{2.5819}(y-60)$$

$$x - 60 = 0.93534 - 56$$

$$x - 0.93534 - 4 = 0.$$

4) Find the Co-efficient of correlation and line of regression for following data.

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

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
10	18	-10	-3	30	100	9
14	12	-6	-9	54	36	81
18	24	-2	3	-6	4	9
22	6	2	-15	-30	4	225
26	30	6	9	54	36	81
30	36	10	15	150	100	225
120	126			252	280	630

$\bar{x} = \dfrac{120}{6} = 20$ $\quad \bar{y} = \dfrac{126}{6} = 21$

W.K.T $\dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} = \dfrac{252}{\sqrt{280}\sqrt{680}} = 0.6$

The regression line of y on x

$$y-\bar{y} = r\frac{\sigma y}{\sigma x}(x-\bar{x})$$

$$\sigma x = \sqrt{\frac{1}{n}\Sigma(x-\bar{x})^2} = \sqrt{\frac{280}{6}} = 6.8313$$

$$\sigma y = \sqrt{\frac{1}{n}\Sigma(y-\bar{y})^2} = \sqrt{\frac{630}{6}} = 10.2469$$

y on x

$$y-\bar{y} = r\frac{\sigma y}{\sigma x}(x-\bar{x})$$

$$y - 21 = \frac{(0.6)(10.2469}{6.8313}(x-20)$$

$$y - 21 = 0.9x - 18$$
$$y = 0.9x + 3$$

x on y

$$x - \bar{x} = r\frac{\sigma x}{\sigma y}(y - \bar{y})$$

$$x - 20 = \frac{(0.6)(6.8313)}{10.2469}(y - 21)$$

$$x - 20 = 0.4y - 8.4$$
$$x = 0.4y + 11.6.$$

5) Find the Co-efficient of Correlation and line of regression for the data below

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

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	2	-2	-3	6	4	9
2	5	-1	0	0	1	0
3	3	0	-2	0	0	4
4	8	1	3	3	1	9
5	7	2	2	4	4	4
15	25			13	10	26

$$\bar{x} = \frac{15}{5} = 3 \qquad \bar{y} = \frac{25}{5} = 5$$

W.K.T $$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$$r = \frac{13}{\sqrt{10 \times 26}} = 0.806$$

$$\sigma x = \sqrt{\frac{1}{n}\Sigma(x-\bar{x})^2} = \sqrt{\frac{10}{5}} = 1.414, \quad \sigma y = \sqrt{\frac{1}{n}\Sigma(y-\bar{y})^2} = \sqrt{\frac{26}{5}} = 2.280$$

$$y - \bar{y} = r\frac{\sigma y}{\sigma x}(x - \bar{x})$$

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

$$y = 1.3x + 1.1$$

W.K.T the regression line of x on y is

$$(x-\bar{x}) = r\frac{\sigma x}{\sigma y}(y-\bar{y})$$

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

$$x = 0.5y+0.5.$$

6) Find the Co-efficient of correlation between the industrial production and export by the following table.

Production (in lakh ton)	55	56	58	59	60	60	60
Export (in lakh ton)	35	38	38	39	44	43	45

Let x be the production in lakh ton and 'y' be the export in lakh ton., then we have.

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
55	35	-3.2857	-5.2857	17.3672	10.7948	27.9386
56	38	-2.2857	-2.2857	5.2244	5.2244	5.2244
58	38	-0.2857	-2.2857	0.6530	0.0816	5.2244
59	39	0.7143	-1.2857	-0.9183	0.5102	1.6530
60	44	1.7143	3.7143	6.3674	2.9388	13.7960
60	43	1.7143	2.7143	4.6531	2.9388	7.3674
60	45	1.7143	4.7143	8.0817	2.9388	22.2246
408	282			41.4285	25.4285	83.4285

$$\bar{x} = \frac{408}{7} = 58.2857 \qquad \bar{y} = \frac{282}{7} = 40.2857$$

$$r = \frac{41.4285}{\sqrt{25.4285}\sqrt{83.4285}}$$

$$r = 0.9.$$

7) with usual notation, Compute $\bar{x}, \bar{y}$ and r from the following lines of regression: $y = 0.516x + 33.73$ and $x - 32.52 = 0.512y + 32.52$.

Sol^n Given regression line of y on x is

$$y = 0.516x + 33.73 \longrightarrow (1) \text{ where } b_{yx} = 0.516$$

And the regression line of x on y is

$x - 32.52 = 0.512y + 32.52$

$x = 0.512y + 65.04 \rightarrow ②$ where $b_{xy} = 0.512$

The given two regression lines are passing through the means of x and y, i.e. $(\bar{x}, \bar{y})$

① $\Rightarrow$ $\bar{y} = 0.516\bar{x} + 33.73 \rightarrow ③$

② $\Rightarrow$ $\bar{x} = 0.512\bar{y} + 65.04 \rightarrow ④$

By solving $\Rightarrow$ $\bar{x} = 111.8631$, $\bar{y} = 91.45135$

The Correlation Coeffecient is $r = \sqrt{b_{yx} \times b_{xy}}$

$= \sqrt{0.516 \times 0.512}$

$= 0.5140$

8) With usual notation, Compute $\bar{x}, \bar{y}$ and r from the following lines of regression $2x + 3y + 1 = 0$, $x + 6y - 4 = 0$.

Soln Given lines of regression are $2x + 3y + 1 = 0 \rightarrow ①$

$x + 6y - 4 = 0 \rightarrow ②$

Let eqn ① be the regression line of x on y

① $\Rightarrow$ $2x + 3y + 1 = 0$

$\Rightarrow x = -1.5y - 0.5$ where $b_{xy} = -1.5$

Let eqn ② be the regression line of y on x

② $\Rightarrow$ $x + 6y - 4 = 0$

$\Rightarrow y = -0.16666x + 0.6666$, where $b_{yx} = -0.1666$

The given two regression lines are passing through the means of x and y, i.e. $(\bar{x}, \bar{y})$

① $\Rightarrow$ $2\bar{x} + 3\bar{y} + 1 = 0 \rightarrow ③$

② $\Rightarrow$ $\bar{x} + 6\bar{y} - 4 = 0 \rightarrow ④$

By solving above equations, we get

$\bar{x} = -2$

$\bar{y} = 1$

The Correlation coefficient is $r = \sqrt{b_{xy} \times b_{yx}}$

$= \sqrt{(-0.1666) \times (-1.5)}$

$= 0.2499.$

9) With usual notation, compute $\bar{x}$, $\bar{y}$ and r from the following lines of regression $4x-5y+33=0$, $20x-9y=107$.

Sol^n Given lines of regression are

$$4x-5y+33 \Rightarrow 4x-5y=-33 \rightarrow ①$$

$$20x-9y=107 \rightarrow ②$$

Let eq^n ② be the regression line of x on y

$$② \Rightarrow 20x-9y=107$$

$$x=\frac{9}{20}y+\frac{107}{20} \quad \text{where } b_{xy}=\frac{9}{20}$$

Let eq^n ① be the regression line of y on x

$$① \Rightarrow 4x-5y=-33$$

$$y=\frac{4}{5}x+\frac{33}{5}, \quad \text{where } b_{yx}=\frac{4}{5}$$

The given two regression lines are passing through the means of x and y. i.e., $(\bar{x}, \bar{y})$

$$① \Rightarrow 4\bar{x}-5\bar{y}=-33 \rightarrow ③$$

$$② \Rightarrow 20\bar{x}-9\bar{y}=107 \rightarrow ④$$

By solving above equations, we get $\bar{x}=13$, $\bar{y}=17$

The correlation coefficient $r=\sqrt{b_{yx} \times b_{xy}}$

$$=\sqrt{\frac{9}{20} \times \frac{4}{5}}$$

$$=0.6.$$

Rank Correlation :

Spearman's rank Correlation :

The spearman Correlation coefficient is defined as the Pearson correlation coefficient between the rank variables and it will be denoted by the symbol ρ (row).

Suppose x and y are the marks scored in the two subjects, let R_x and R_y are the ranks, $d_i=R_x-R_y$ be the differences of the ranks then the Spearman's rank correlation will be calculated by following formula

$$\rho = 1-\frac{6\sum_{i=1}^{n} d_i^2}{n^3-n}$$

1) The scores for 9 students in Physics and mathematics are as follows.

Physics	35	23	47	17	10	43	9	6	28
Mathematics	30	33	45	23	8	49	12	4	31

Compute the rank of students in the two subjects and also compute the spearman's rank correlation.

Phy(x)	Mat(y)	Rx	Ry	di = Rx - Ry	di^2
35	30	3	5	-2	4
23	33	5	3	2	4
47	45	1	2	-1	1
17	23	6	6	0	0
10	8	7	8	-1	1
43	49	2	1	1	1
9	12	8	7	1	1
6	4	9	9	0	0
28	31	4	4	0	0
					$\Sigma di^2 = 12$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} di^2}{n^3 - n}$$

$$\rho = 1 - \frac{6(12)}{9^3 - 9}$$

$$\rho = 1 - \frac{72}{720}$$

$$\rho = 1 - 0.1$$

$$\rho = 0.9$$

2) The score of 9 students in english and mathematics are as follows

English	56	75	45	71	62	64	58	80	76	61
Mathematics	66	70	40	60	65	56	59	77	67	63

Compute the ranks of students in the two subjects and also compute the spearman's rank correlation.

Eng(x)	Mat(y)	Rx	Ry	$d_i = R_x - R_y$	d_i^2
56	66	9	4	5	25
75	70	3	2	1	1
45	40	10	10	0	0
71	60	4	7	-3	9
62	65	6	5	1	1
64	56	5	9	-4	16
58	59	8	8	0	0
80	77	1	1	0	0
76	67	2	3	1	1
61	63	7	6	1	1
					$\Sigma d_i^2 = 54$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$= 1 - \frac{6(54)}{10^3 - 10}$$

$$= 1 - \frac{324}{990}$$

$$= 1 - 0.33$$

$$\rho = 0.67$$

3) A random sample of recent repair jobs was selected and estimated cost and actual cost were recorded

Estimated Cost (x)	300	450	800	250	500	975	475	400
Actual Cost (y)	273	486	734	297	631	872	396	457

Compute the ranks and the Spearman's rank correlation.

Estimated Cost (x)	Actual Cost (y)	Rx	Ry	$d_i = R_x - R_y$	d_i^2
300	273	7	8	-1	1
450	486	5	4	1	1
800	734	2	2	0	0
250	297	8	7	1	1
500	631	3	3	0	0
975	872	1	1	0	0
475	396	4	6	-2	4
400	457	6	5	1	1
					$\Sigma d_i^2 = 8$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$\rho = 1 - \frac{6(8)}{8^3 - 8}$$

$$\rho = 1 - \frac{48}{504}$$

$$\rho = 1 - 0.095$$

$$\rho = 0.9047$$

4) The rank of 10 students of same batch in two subjects A and B are given below. Calculate the rank correlation Coefficient

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

Given, the ranks of no of students $n = 10$

R_A	R_B	$d_i = R_A - R_B$	d_i^2
1	6	-5	25
2	7	-5	25
3	5	-2	4
4	10	-6	36
5	3	2	4
6	9	-3	9
7	4	3	9
8	1	7	49
9	8	1	1
10	2	8	64
			$\Sigma d_i^2 = 226$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$= 1 - \frac{6(226)}{10^3 - 10}$$

$$= 1 - \frac{1356}{990}$$

$$= 1 - 1.37$$

$$\rho = -0.37$$

5) The participants in a Contest are ranked by two judges as follows

x	1	6	5	10	3	2	4	9	7	8
y	6	4	9	8	1	2	3	10	5	7

Compute the Spearman's rank correlation

Given

the ranks of no. of students $n = 10$

R_x	R_y	$d_i = R_x - R_y$	d_i^2
1	6	-5	25
6	4	2	4
5	9	-4	16
10	8	2	4
3	1	2	4
2	2	0	0
4	3	1	1
9	10	-1	1
7	5	2	4
8	7	1	1
			$\Sigma d_i^2 = 60$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$= 1 - \frac{6(60)}{10^3 - 10}$$

$$= 1 - \frac{360}{990}$$

$$= 1 - 0.3636$$

$$\rho = 0.6363$$

6) The participants in a Contest are ranked by two judges as follows

x	6	4	3	1	2	7	9	8	10	5
y	4	1	6	7	5	8	10	9	3	2

Compute the Spearman's rank Correlation

Sol^n

Given, The ranks of no of students $n = 10$.

R_x	R_y	$d_i = R_x - R_y$	d_i^2
6	4	2	4
4	1	3	9
3	6	-3	9
1	7	-6	36
2	5	-3	9
7	8	-1	1
9	10	-1	1
8	9	-1	1
10	3	7	49
5	2	3	9
			$\Sigma d_i^2 = 128$

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$\rho = 1 - \frac{6(128)}{10^3 - 10}$$

$$= 1 - \frac{768}{990}$$

$$= 1 - 0.7758$$

$$\rho = 0.2242$$

7) Ten competitors in a beauty contest are ranked by two judges A and B in the following order.

I.D. No. of Competitors	1	2	3	4	5	6	7	8	9	10
Judge A	1	6	5	3	10	2	4	9	7	8
Judge B	6	4	9	8	1	2	3	10	5	7

Calculate the rank coefficient.

Sol^n Given, The ranks of no of students $n = 10$.

R_A	R_B	$d_i = R_A - R_B$	d_i^2
1	6	-5	25
6	4	2	4
5	9	-4	16
3	8	-5	25
10	1	9	81
2	2	0	0
4	3	1	1
9	10	-1	1
7	5	2	4
8	7	1	1
			$\Sigma d_i^2 = 158$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$= 1 - \frac{6(158)}{10^3 - 10}$$

$$= 1 - \frac{948}{990}$$

$$= 1 - 0.9575$$

$$\rho = 0.04242$$

8) The rank of 10 students of same batch in two subjects A and B are given below. Calculate the rank correlation coefficient

Roll.No of Student	1	2	3	4	5	6	7	8	9	10
Rank of A	3	5	8	4	7	10	2	1	6	9
Rank of B	6	4	9	8	1	2	3	10	5	7

Sol^n Given, the ranks of no of students $n = 10$

R_A	R_B	$d_i = R_A - R_B$	d_i^2
3	6	-3	9
5	4	1	1
8	9	-1	1
4	8	-4	16
7	1	6	36
10	2	8	64
2	3	-1	1

RA	RB	$d_i = R_A - R_B$	d_i^2
1	10	-9	81
6	5	1	1
9	7	2	4
			$\Sigma d_i^2 = 214$

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$= 1 - \frac{6(214)}{10^3 - 10}$$

$$= 1 - \frac{1284}{990}$$

$$= 1 - 1.297$$

$$= -0.2970$$

9) The students get the following % of marks in 2 subjects A and B.

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

Calculate Spearman's rank Correlation Coefficient

Soln

Marks in A	Marks in B	RA	RB	$d_i = R_A - R_B$	d_i^2
78	84	4	3	1	1
36	51	9	9	0	0
98	91	1	1	0	0
25	60	10	6	4	16
75	68	5	4	-1	1
82	62	3	5	-2	4
90	86	2	2	0	0
62	58	7	7	0	0
65	53	6	8	-2	4
39	47	8	10	-2	4
					$\Sigma d_i^2 = 30$

W.K.T, the Spearman's Correlation Coefficient

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} = 1 - \frac{6(30)}{10^3 - 10}$$

$$= 1 - \frac{180}{990}$$

$$= 1 - 0.1818$$

$$= 0.8181$$