Mat 41

Module - 4

Statistical Methods

Mean (Arithmetic mean) :-

If $x_1, x_2, \dots, x_n$ be a set of 'n' values of a variate 'x', the mean denoted by $\bar{x}$ is defined as follows.

$$\bar{x} = \frac{\sum x}{n} \quad \text{(or)} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For a grouped data in the form of a frequency distribution,

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \quad \text{(or)} \quad \bar{x} = \frac{\sum f x}{\sum f}$$

where f_i's are the frequency of the classes having corresponding midpoint 'x_i'.

Variance (V) and Standard deviation (SD) :-

If a variate 'x' take values $x_1, x_2, \dots, x_n$ the variance (V) is defined as follows.

$$V = \frac{\sum (x-\bar{x})^2}{n} \quad \text{(or)} \quad V = \frac{\sum_{i=1}^{n} (x_i-\bar{x})^2}{n}$$

Also for a grouped data

$$V = \frac{\sum f (x-\bar{x})^2}{\sum f} \quad \text{(or)} \quad V = \frac{\sum_{i=1}^{n} f_i (x_i-\bar{x})^2}{\sum_{i=1}^{n} f_i}$$

Standard deviation (SD), $\sigma = \sqrt{V}$ (or) $\sigma^2 = V$

Alternative Expression for σ^2 :-

consider, $\sigma^2 = \frac{1}{n}\sum (x-\bar{x})^2$

$$\sigma^2 = \frac{1}{n}\sum [x^2 - 2x\bar{x} + (\bar{x})^2]$$

$$= \frac{\sum x^2}{n} - 2(\bar{x})\cdot\bar{x} + \frac{n(\bar{x})^2}{n}$$

Here, $\frac{\sum x}{n} = \bar{x}$ & $(\bar{x})^2$ being a constant added 'n' times gives $n(\bar{x})^2$.

i.e $\sigma^2 = \frac{\sum x^2}{n} - 2(\bar{x})^2 + (\bar{x})^2$

$$\sigma^2 = \frac{\sum x^2}{n} - (\bar{x})^2$$

For a grouped data the expression will be of the form

$$\sigma^2 = \frac{\sum f x^2}{\sum f} - (\bar{x})^2$$

<u>Example</u> :-

① we shall find the mean and standard deviation of a set of observations. 6, 8, 7, 5, 4, 9, 3

<u>Soln</u> :- $\bar{x} = \frac{\sum x}{n}$,

$$\bar{x} = \frac{6+8+7+5+4+9+3}{7} = \frac{42}{7} = 6$$

Thus mean $(\bar{x}) = 6$

$$V \text{ or } \sigma^2 = \frac{\sum(x-\bar{x})^2}{n} = \frac{1}{7}\{(6-6)^2+(8-6)^2+(7-6)^2+(5-6)^2+(4-6)^2+(9-6)^2+(3-6)^2\}$$

$$V = \frac{28}{7} = 4$$

$$\therefore \sigma = \sqrt{4} = 2$$

<u>Alternate</u> : $\sigma^2 = \frac{\sum x^2}{n} - (\bar{x})^2$

$$= \frac{6^2+8^2+7^2+5^2+4^2+9^2+3^2}{7} - (6)^2$$

$$= \frac{280}{7} - 36$$

$$= 4$$

Thus $SD = \sigma = 2$.

② Let us find the mean & SD for the following grouped data

class	1-10	11-20	21-30	31-40	41-50	51-60
Frequency	3	16	26	31	16	8

Class	f	x	fx	$(x-\bar{x})$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
1-10	3	5.5	16.5		702.25	2106.75
11-20	16	15.5	248.0		272.25	4356.00
21-30	26	25.5	663.0		42.25	1098.50
31-40	31	35.5	1100.5		12.25	379.75
41-50	16	45.5	728.0		182.25	2916.00
51-60	8	55.5	444.0		552.25	4418.00
Totals	100		3200			15275

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{3200}{100} = 32$$

$$\sigma^2 = \frac{\sum f(x-\bar{x})^2}{\sum f} = \frac{15275}{100} = 152.75$$

$$\therefore \sigma = \sqrt{152.75} = 12.36$$

Curve Fitting:-

* Fitting of a straight line : $y = ax+b$

Consider a set of 'n' given values (x, y) for fitting the straight line $y = ax+b$ where 'a' & 'b' are parameters to be determined. The residual $R = y-(ax+b)$ is the difference between the observed and estimated values of 'y'. By the method of least squares we find parameters 'a' & 'b' such that the sum of squares of the residuals is minimum.

$$a\sum x + nb = \sum y$$

$$a\sum x^2 + b\sum x = \sum xy$$

Problems:-

① Fit a straight line $y = ax+b$ for the following data.

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

The normal eqns for fitting the straight line $y = ax+b$ are $\Sigma y = a\Sigma x + nb$

$\Sigma xy = a\Sigma x^2 + b\Sigma x$ $(n=8)$

x	y	xy	x^2
1	1	1	1
3	2	6	9
4	4	16	16
6	4	24	36
8	5	40	64
9	7	63	81
11	8	88	121
14	9	126	196
$\Sigma x = 56$	$\Sigma y = 40$	$\Sigma xy = 364$	$\Sigma x^2 = 524$

The normal eqns become

$$56a + 8b = 40.$$
$$524a + 56b = 364$$

$\therefore a = 0.63 \approx 0.64, \; b = 0.54 \approx 0.55$

$\therefore$ Thus by substituting these values in $y = ax+b$ we obtain the eqn.

$$\underline{y = 0.64x + 0.55}$$

② Find the eqn of the best fitting straight line for the following data & hence estimate the value of the dependent variable corresponding to the value 30 of the independent variable.

x	5	10	15	20	25
y	16	19	23	26	30

$\therefore \; y = ax+b$

$\Sigma y = a\Sigma x + nb$

$\Sigma xy = a\Sigma x^2 + b\Sigma x$ $(n=5)$

x	y	xy	x^2
5	16	80	25
10	19	190	100
15	23	345	225
20	26	520	400
25	30	750	625

$\Sigma x = 75$ $\Sigma y = 114$ $\Sigma xy = 1885$ $\Sigma x^2 = 1375$

$\therefore$ $75a + 5b = 114$

$1375a + 75b = 1885$

$\therefore$ $a = 0.7$, $b = 12.3$.

$y = ax + b$

$y = 0.7x + 12.3$

when $x = 30$, we obtain $y = 0.7(30) + 12.3 = 33.3$

(3) A simply supported beam carries a concentrated load P at its mid point. corresponding to various values of 'p' the maximum deflection 'y' is measured & is given in the following table

P	100	120	140	160	180	200
Y	0.45	0.55	0.60	0.70	0.80	0.85

find a law of the form $y = a + bp$ & hence estimate 'y' when 'p' is 150.

Soln: The normal equations associated with $y = a + bp$ are as follows

$\Sigma y = na + b\Sigma p$ $(n=6)$

$\Sigma py = a\Sigma p + b\Sigma p^2$

p	y	py	p^2
100	0.45	45	10000
120	0.55	66	14400
140	0.60	84	19600
160	0.70	112	25600
180	0.80	144	32400
200	0.85	170	40000
$\Sigma p = 900$	$\Sigma y = 3.95$	$\Sigma py = 621$	$\Sigma p^2 = 142000$

$\therefore$ Eqns $\Rightarrow$

$$6a + 900b = 3.95$$

$$900a + 142000b = 621$$

$\therefore a = 0.0476$, $b = 0.0041$

Thus the required Law is $y = 0.0476 + 0.0041p$

Also when $p = 150$, $y = 0.6626 \approx 0.66$

(4) Fit a straight line to the following data.

Year	1961	1971	1981	1991	2001
production (in tons)	8	10	12	10	16

Also find the expected production in the year 2006.

Soln :- let $X = x - 1981$ & the line of fit with be $y = a + bX$

$$\Sigma y = na + b\Sigma X \qquad (n = 5)$$

$$\Sigma Xy = a\Sigma X + b\Sigma X^2$$

x	y	xy	x^2
-20	8	-160	400
-10	10	-100	100
0	12	0	0
10	10	100	100
20	16	320	400

$\Sigma x = 0 \quad \Sigma y = 56 \quad \Sigma xy = 160 \quad \Sigma x^2 = 1000$

The normal equations become,

$5a = 56$ & $1000b = 160$

$a = 11.2$ & $b = 0.16$

Hence $y = a + bX$, with $X = x - 1981$ becomes

$y = 11.2 + 0.16(x - 1981)$

Thus $y = -305.76 + 0.16x$ is the required line of fit.

Also when $x = 2006$, $y = -305.76 + 0.16(2006)$

$y = 15.2$

Expected production in the year 2006 is 15.2 tons.

(5) Find the eqn of the best fitting straight line for the following data.

i)

x	1	2	3	4	5
y	14	13	9	5	2

ii)

x	0	1	2	3	4	5
y	9	8	24	28	26	20

iii)

x	62	64	65	69	70	71	72
y	65.7	66.8	67.2	69.3	69.8	70.5	70.9

iv)

x	1	2	3	4	5	6	7
y	80	90	92	83	94	99	92

(6) Year (x) 1911, 1921, 1931, 1941, 1951

Year (x)	1911	1921	1931	1941	1951
production(y) (in thousand tons)	8	10	12	10	6

Soln:- let $X = x - 1931$ & the line of fit will be $y = a + bX$

The normal eqns associated with $y = a + bX$ are as follows.

$$\Sigma y = na + b\Sigma X \quad (n=5)$$
$$\Sigma Xy = a\Sigma X + b\Sigma X^2$$

X	y	Xy	X^2
-20	8	-160	400
-10	10	-100	100
0	12	0	0
10	10	100	100
20	6	120	400

$\Sigma X = 0$ $\Sigma y = 46$ $\Sigma Xy = -40$ $\Sigma X^2 = 1000$

The normal eqn become,

$$46 = 5a \rightarrow (1)$$
$$-40 = 1000b \rightarrow (2)$$

Eqn (1) $a = 9.2$,

Eqn (2) $b = -0.04$

Hence $y = a + bX$, with $X = x - 1931$

$$y = 9.2 + (-0.04)(x - 1931)$$
$$= 9.2 - 0.04x + 77.24$$

Thus, $y = 86.44 - 0.04x$ is the required line of fit.

* Fitting of a second degree parabola $y = ax^2+bx+c$:-

consider a set of 'n' given values (x,y) for fitting the curve $y = ax^2+bx+c$. The residual $R = y-(ax^2+bx+c)$ is the difference b/w the observed & estimated values of y. We have to find parameters a, b, c such that the sum of the squares of the residuals is the least.

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

*

① Fit a best fitting parabola $y = ax^2+bx+c$ for the following data:

x	1	2	3	4	5
y	10	12	13	16	19

and hence estimate at $x=6$.

Soln:- The normal eqns associated with $y = ax^2+bx+c \rightarrow (*)$ are as follows.

$$\Sigma y = a\Sigma x^2 + b\Sigma x + nc \longrightarrow ①$$
$$\Sigma xy = a\Sigma x^3 + b\Sigma x^2 + c\Sigma x \longrightarrow ②$$
$$\Sigma x^2y = a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 \longrightarrow ③$$

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

$\Sigma x = 15$ $\Sigma y = 70$ $\Sigma xy = 232$ $\Sigma x^2 = 55$ $\Sigma x^2y = 906$ $\Sigma x^3 = 225$ $\Sigma x^4 = 979$

Eqn ① $\Rightarrow$ $70 = 55a + 15b + 5c$

eqn ② $\Rightarrow$ $232 = 225a + 55b + 15c$

Eqn ③ $\Rightarrow$ $906 = 979a + 225b + 55c$

$\therefore a = 0.2857 \approx 0.29$, $b = 0.4857 \approx 0.49$, $c = 9.4$.

Thus the required second degree parabola is

$y = 0.29x^2 + 0.49x + 9.4$ Also at $x = 6$

$y = 22.78$

★ (2) Fit a parabola $y = a + bx + cx^2$ for the data.

x	0	1	2	3	4
y	1	1.8	1.3	2.5	2.3

6.3

Soln:- The normal eqn associated $y = a + bx + cx^2 \rightarrow (*)$ are

$$\Sigma y = an + b\Sigma x + c\Sigma x^2 \rightarrow (1)$$

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

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

x	y	xy	x^2	x^2y	x^3	x^4
0	1	0	0	0	0	0
1	1.8	1.8	1	1.8	1	1
2	1.3	2.6	4	5.2	8	16
3	2.5	7.5	9	22.5	27	81
4	2.3	9.2	16	36.8	64	256
$\Sigma x = 10$	$\Sigma y = 8.9$	$\Sigma xy = 21.1$	$\Sigma x^2 = 30$	$\Sigma x^2y = 66.3$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$

Eqn (1) $\Rightarrow$ $8.9 = 5a + 10b + 30c$

Eqn (2) $\Rightarrow$ $21.1 = 10a + 30b + 100c$

eqn (3) $\Rightarrow$ $66.3 = 30a + 100b + 354c$

$a = 1.0771$, $b = 0.4157$, $c = -0.0214$

Thus the parabola of fit is

$$y = 1.0771 + 0.4157x - 0.0214x^2$$

★ (3) Fit a second degree parabola to the following data.

x	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y	1.1	1.3	1.6	2.0	2.7	3.4	4.1

★ (4) Fit a second degree parabola to the following data:

x	0	1	2	3	4
y	1	1.8	1.3	2.5	6.3

* **Fitting of a curve of the form $y = ae^{bx}$** :-

consider, $y = ae^{bx}$. Taking logarithms (to the base e) on both sides we get.

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

$$= \log_e a + \log_e e^{bx}$$

$$= \log_e a + bx \log_e e$$

| $\log mn = \log m + \log n$
| $\log m^n = n \log_e m$
| $\log_e e = 1$

$$\log_e y = \log_e a + bx \quad \text{(or)}$$

$$Y = A + BX \longrightarrow (1)$$

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

It is evident that eq^n (1) is the eq^n of a straight line & the associated normal eq^{ns} are as follows.

$$\Sigma Y = nA + B\Sigma X \longrightarrow (2)$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2 \longrightarrow (3)$$

Solving eq^n (2) & (3) we obtain A & B. But $\log_e a = A \Rightarrow a = e^A$. Also $b = B$.

Substituting these values in $y = ae^{bx}$ we get the curve of best fit, in the required form.

* (1) Fit a curve of the form $y = ae^{bx}$ to the following data.

x	77	100	185	239	285
y	2.4	3.4	7.0	11.1	19.6

Sol^n :- consider, $y = ae^{bx} \longrightarrow (*)$

$$Y = A + bx$$

The normal eq^{ns} are as follows.

$$\Sigma Y = nA \cancel{\Sigma} + b\Sigma x \longrightarrow (1)$$

$$\Sigma xY = A\Sigma x + b\Sigma x^2 \longrightarrow (2)$$

where $Y = \log_e y$, $A = \log_e a$

x	y	$Y = \log_e y$	xY	x^2
77	2.4	0.8754	67.4058	5929
100	3.4	1.2237	122.37	10000
185	7.0	1.9459	359.9915	34225
239	11.1	2.4069	575.2491	57121
285	19.6	2.9755	848.0175	81225
$\Sigma x = 886$	$\Sigma y = 43.5$	$\Sigma Y = 9.4274$	$\Sigma xY = 1973.0339$	$\Sigma x^2 = 188500$

Eqn ① $\Rightarrow$ $9.4274 = 5A + 886b$

Eqn ② $\Rightarrow$ $1973.0339 = 886A + 188500b$

$A = 0.1838$ $\qquad$ $b = 9.6028 \times 10^{-3}$

$A = \log_e a = 0.1838$

$a = e^{A}$

$= e^{0.1838}$

$a = 1.2017$, $\qquad b = 9.6088 \times 10^{-3}$

The curve of fit is $y = ae^{bx}$

Thus $y = (1.2017)e^{9.6028 \times 10^{-3} x}$ is the curve of fit.

② Fit an exponential curve of the form $y = ae^{bx}$ by the method of least squares for the following data.

No of petals	5	6	7	8	9	10
No of flowers	133	55	23	7	2	2

x	y	$Y=\log_e y$	xY	x^2
5	133	4.8903	24.4515	25
6	55	4.0073	24.0438	36
7	23	3.1355	21.9485	49
8	7	1.9459	15.5672	64
9	2	0.6931	6.2379	81
10	2	0.6931	6.9310	100
$\Sigma x=45$		$\Sigma Y=15.3652$	$\Sigma xY=99.1799$	$\Sigma x^2=355$

The normal eqn becomes

$$6A+45b=15.3652$$

$$45A+355b=99.1799$$

$A=9.4433$ & $b=-0.9177$

Thus the required curve of fit is

$$y=(12623.3)\,e^{-0.9177x}$$

③ 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

④ Find the eqn of the best fitting curve in the form $y=ae^{bx}$ for the data

x	0	2	4
y	5.02	10	31.62

Correlation and correlation co-efficient :-

Co-variation of two independent magnitudes is known as correlation. If two variables 'x' & 'y' are related in such a way that ↑ or ↓ in one of them corresponds to ↑ or ↓ in the other, we say that the variables are +vely correlated. Also if increase or decrease in one of them corresponds to decrease or ↑ in the other, the variables are said to be -vely correlated.

The numerical measure of correlation b/w two variables 'x' & 'y' is known as Pearson's coefficient of correlation usually denoted by 'r' & is defined as follows.

$$\boxed{r = \frac{\sum (x-\bar{x})(y-\bar{y})}{n \sigma_x \sigma_y}} \longrightarrow (1)$$

This can be put in an alternative form as follows. If $X = x-\bar{x}$, $Y = y-\bar{y}$ we can write.

$$\sigma_x^2 = \frac{\sum (x-\bar{x})^2}{n} = \frac{\sum X^2}{n},$$

$$\sigma_y^2 = \frac{\sum (y-\bar{y})^2}{n} = \frac{\sum Y^2}{n}$$

$$\therefore \sigma_x \sigma_y = \frac{\sqrt{\sum X^2}\sqrt{\sum Y^2}}{\sqrt{n^2}} = \frac{\sqrt{\sum X^2}\sqrt{\sum Y^2}}{n}$$

$$\therefore n \sigma_x \cdot \sigma_y = \sqrt{\sum X^2} \cdot \sqrt{\sum Y^2}$$

Thus eqn (1) becomes

$$\boxed{r = \frac{\sum XY}{\sqrt{\sum X^2}\sqrt{\sum Y^2}}}$$

Note :- *. The coefficient of correlation numerically does not exceed unity. i.e $-1 \le r \le +1$

* If $r = \pm 1$ we say that 'x' & 'y' are perfectly correlated & if $r = 0$ we say that 'x' & 'y' are non correlated.

Alternative formula for the correlation co-efficient r :-

$$\boxed{r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}}$$

proof :- let $z = x - y$

$$\therefore \frac{\Sigma z}{n} = \frac{\Sigma x}{n} - \frac{\Sigma y}{n} \quad (\text{or}) \quad \bar{z} = \bar{x} - \bar{y}$$

Hence, $(z - \bar{z}) = (x - y) - (\bar{x} - \bar{y})$

i,e $(z - \bar{z}) = (x - \bar{x}) - (y - \bar{y})$

Squaring both sides, taking summation & dividing by n we have,

$$\frac{\Sigma (z - \bar{z})^2}{n} = \frac{\Sigma [(x - \bar{x}) - (y - \bar{y})]^2}{n}$$

$$= \frac{\Sigma (x - \bar{x})^2}{n} + \frac{\Sigma (y - \bar{y})^2}{n} - 2\frac{\Sigma (x - \bar{x})(y - \bar{y})}{n}$$

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

i,e $\sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x \sigma_y$

i,e $\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x \sigma_y$

Thus $r = \dfrac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$

Note :- In general if $z = ax + by$ we can obtain as before

$$\sigma_z^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2rab\sigma_x \sigma_y$$

i,e $\sigma_{ax+by}^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2rab\sigma_x \sigma_y$

Regression :-

Regression is an estimation of one independent variable in terms of the other. If 'x' & 'y' are correlated, the best fitting straight line in the least square sense gives reasonably a good relation b/w 'x' & 'y'.

The best fitting straight line of the form $y = ax+b$ ('x' being the independent variable) is called the regression line of 'y' on 'x' & $x = ay+b$ ('y' being the independent variable) is called the regression line of 'x' on 'y'. $y-\bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$, $x-\bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$

Note :- 1) The lines of regression (3) a

$$Y = \frac{\Sigma XY}{\Sigma X^2}(X) \quad \text{and} \quad X = \frac{\Sigma XY}{\Sigma Y^2}(Y)$$

where $X = x-\bar{x}$ & $Y = y-\bar{y}$.

This form will be useful to find out the coefficient of correlation by first obtaining the lines of regression as we have deduced that

$$\boxed{r = \pm\sqrt{(\text{coeff. of } x)(\text{coeff. of } y)}}$$

2) To compute the coefficient of correlation we prefer to use the formula.

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$$

where SDs can be found by applying the formula.

$$\sigma^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2$$

If $\bar{x}$ & $\bar{y}$ are integers computation of 'r' by the formula.

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}}$$ is convenient where $X = x-\bar{x}$, $Y = y-\bar{y}$.

problems:-

1. compute the coefficient of correlation and the equation of the lines of regression for the data.

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

Solⁿ:- we have $r = \dfrac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y} \rightarrow (1)$

x	y	$z = x-y$	x^2	y^2	z^2
1	9	-8	1	81	64
2	8	-6	4	64	36
3	10	-7	9	100	49
4	12	-8	16	144	64
5	11	-6	25	121	36
6	13	-7	36	169	49
7	14	-7	49	196	49
$\Sigma x=28$	$\Sigma y=77$	$\Sigma z=-49$	$\Sigma x^2=140$	$\Sigma y^2=875$	$\Sigma z^2=347$

$\bar{x} = \dfrac{\Sigma x}{n}$, $\bar{y} = \dfrac{\Sigma y}{n}$, $\bar{z} = \dfrac{\Sigma z}{n}$

$\bar{x} = \dfrac{28}{7} = 4$, $\bar{y} = \dfrac{77}{7} = 11$, $\bar{z} = \dfrac{-49}{7} = -7$

$\sigma_x^2 = \dfrac{\Sigma x^2}{n} - (\bar{x})^2$, $\sigma_y^2 = \dfrac{\Sigma y^2}{n} - (\bar{y})^2$, $\sigma_z^2 = \dfrac{\Sigma z^2}{n} - (\bar{z})^2$

$= \dfrac{140}{7} - (4)^2 = 4$ $\quad = \dfrac{875}{7} - (11)^2 = 4$ $\quad = \dfrac{347}{7} - (-7)^2 = 0.57$

we have, $\sigma_x^2 = 4$, $\sigma_y^2 = 4$, $\sigma_z^2 = \sigma_{x-y}^2 = 0.57$

Eqⁿ (1) $\Rightarrow$ $r = \dfrac{4+4-0.57}{2\sqrt{4}\sqrt{4}} = 0.92875 \approx 0.93$.

Thus $r = 0.93$

The lines of regression are given by

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

$y - 11 = \dfrac{(0.93)(2)}{2}(x-4)$, $x - 4 = \dfrac{(0.93)\,2}{2}(y-11)$

$y - 11 = 0.93(x-4)$, $x - 4 = 0.93(y-11)$

Thus $y = 0.93x + 7.28$ & $x = 0.93y - 6.23$ are the lines of regression.

(2) obtain the lines of regression & hence find the coefficient of correlation for the data.

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

Soln :- Here $\bar{x} = 4$, $\bar{y} = 11$

$\therefore X = x - \bar{x}$, $Y = y - \bar{y}$

$= x - 4$, $= y - 11$

x	y	X	Y	XY	X^2	Y^2
1	9	-3	-2	6	9	4
2	8	-2	-3	6	4	9
3	10	-1	-1	1	1	1
4	12	0	1	0	0	1
5	11	1	0	0	1	0
6	13	2	2	4	4	4
7	14	3	3	9	9	9
				$\Sigma XY = 26$	$\Sigma X^2 = 28$	$\Sigma Y^2 = 28$

we shall consider regression lines in the form.

$$Y = \frac{\Sigma XY}{\Sigma X^2} \cdot X \quad \text{and} \quad X = \frac{\Sigma XY}{\Sigma Y^2} \cdot Y$$

i,e $y - 11 = \frac{26}{28}(x-4)$, $x - 4 = \frac{26}{28}(y-11)$

$y - 11 = 0.93(x-4)$, $x - 4 = 0.93(y-11)$

$y = 0.93x + 7.28$, $x = 0.93y - 6.23$

These are the regression lines and we compute 'r' as the geometric mean of the regression coefficients.

i,e $r = \sqrt{(\text{coeff of } x)(\text{coeff of } y)} = \sqrt{(0.93)(0.93)}$

$r = 0.93$

The sign of 'r' must be +ve since both the regression coefficients are +ve.

Thus $r = 0.93$.

③ Find the correlation coefficient & the eqn of the lines of regression for the following values of 'x' & 'y'.

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

Soln:- $n = 5$

x	y	$z = x - y$	x^2	y^2	z^2
1	2	-1	1	4	1
2	5	-3	4	25	9
3	3	0	9	9	0
4	8	-4	16	64	16
5	7	-2	25	49	4
$\Sigma x = 15$	$\Sigma y = 25$	$\Sigma z = -10$	$\Sigma x^2 = 55$	$\Sigma y^2 = 151$	$\Sigma z^2 = 30$

$\bar{x} = \frac{\Sigma x}{n}$, $\bar{y} = \frac{\Sigma y}{n}$, $\bar{z} = \frac{\Sigma z}{n}$

$\bar{x} = \frac{15}{5} = 3$, $\bar{y} = \frac{25}{5} = 5$, $\bar{z} = \frac{-10}{5} = -2$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{55}{5} - (3)^2 = 2$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{151}{5} - (5)^2 = 5.2$$

$$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = \frac{30}{5} - (-2)^2 = 2$$

we have, $\sigma_x^2 = 2$, $\sigma_y^2 = 5.2$, $\sigma_z^2 = 2$

Now, $r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$

$$r = \frac{2 + 5.2 - 2}{2\sqrt{2}\sqrt{5.2}} = 0.8062 \approx 0.81$$

Thus $\boxed{r = 0.81}$

The eqn of the regression lines are as follows.

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

$y - 5 = (0.81) \cdot \frac{\sqrt{5.2}}{\sqrt{2}}(x - 3)$, $\quad x - 3 = 0.81 \cdot \frac{\sqrt{2}}{\sqrt{5.2}}(y - 5)$

$$y - 5 = 1.306(x-3), \quad x-3 = 0.502(y-5)$$

Thus $y = 1.306x + 1.082$ & $x = 0.502y + 0.49$

These are the lines of regression.

④ Find the correlation coefficient b/w 'x' & 'y' for the following data. Also obtain the regression lines.

x	1	2	3	4	5	6	7	8	9	10
y	10	12	16	28	25	36	41	49	40	50

Soln:- Here $n = 10$

x	y	$z = x - y$	x^2	y^2	z^2
1	10	-9	1	100	81
2	12	-10	4	144	100
3	16	-13	9	256	169
4	28	-24	16	784	576
5	25	-20	25	625	400
6	36	-30	36	1296	900
7	41	-34	49	1681	1156
8	49	-41	64	2401	1681
9	40	-31	81	1600	961
10	50	-40	100	2500	1600
$\Sigma x = 55$	$\Sigma y = 307$	$\Sigma z = -252$	$\Sigma x^2 = 385$	$\Sigma y^2 = 11387$	$\Sigma z^2 = 7624$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{55}{10} = 5.5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{307}{10} = 30.7$$

$$\bar{z} = \frac{\Sigma z}{n} = \frac{-252}{10} = -25.2$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{385}{10} - (5.5)^2 = 8.25, \quad \sigma_x = 2.87$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = 1138.7 - (30.7)^2 = 196.21, \quad \sigma_y = 14.01$$

$$\sigma_z^2 = \sigma_{x-y}^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = 762.4 - (-25.2)^2 = 127.36$$

we have, $r = \dfrac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$

$$= \frac{8.25 + 196.21 - 127.36}{2 \times 2.87 \times 14.01} = 0.96$$

Thus $\boxed{r = 0.96}$

Equation of the lines of regression are

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

on substituting & simplifying we get,

$$y - 30.7 = 0.96 \times \frac{14.01}{2.87}(x - 5.5)$$

$$\boxed{y = 4.686x + 4.927}$$

$$x - 5.5 = 0.96 \times \frac{2.87}{14.01}(y - 30.7)$$

$$\boxed{x = 0.197y - 0.548}$$

These are the lines of regression.

* (5) Find the coefficient of correlation for the following data.

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

Soln:- we have $\bar{x} = \dfrac{\Sigma x}{n} = \dfrac{120}{6} = 20$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{126}{6} = 21$$

Let $X = x - \bar{x}$, & $Y = y - \bar{y}$

$X = x - 20$, $Y = y - 21$ &

x	y	X	Y	X^2	Y^2	XY
10	18	-10	-3	100	9	30
14	12	-6	-9	36	81	54
18	24	-2	3	4	9	-6
22	6	2	-15	4	225	-30
26	30	6	9	36	81	54
30	36	10	15	100	225	150
				$\Sigma X^2 = 280$	$\Sigma Y^2 = 630$	$\Sigma XY = 252$

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}} = \frac{252}{\sqrt{280} \times \sqrt{630}} = 0.6$$

★

⑥ Find the correlation coefficient and the eq^n of the line of regression for the following.

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

Sol^n:- Let $n = 5$

x	y	$z = x - y$	x^2	y^2	z^2
1	2	-1	1	4	1
2	5	-3	4	25	9
3	3	0	9	9	0
4	8	-4	16	64	16
5	7	-2	25	49	4
$\Sigma x = 15$	$\Sigma y = 25$	$\Sigma z = -10$	$\Sigma x^2 = 55$	$\Sigma y^2 = 151$	$\Sigma z^2 = 30$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5} = 3, \quad \bar{y} = \frac{\Sigma y}{n} = \frac{25}{5} = 5, \quad \bar{z} = \frac{\Sigma z}{n} = \frac{-10}{5} = -2$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{55}{5} - (3)^2 = 2$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{151}{5} - (5)^2 = 5.2$$

$$\sigma_z^2 = \sigma_{x-y}^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = \frac{30}{5} - (-2)^2 = 2$$

$$\text{Now} = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$$

$$r = \frac{2+5.2-2}{2\sqrt{2}\sqrt{5.2}} = 0.8062 \approx 0.81$$

Thus $r = 0.81$

The eqn of the regression lines are as follows.

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

$$y-5 = (0.81)\frac{\sqrt{5.2}}{\sqrt{2}}(x-3) \, , \quad x-3 = (0.81)\frac{\sqrt{2}}{\sqrt{5.2}}(y-5)$$

$$y-5 = 1.306\,(x-3) \quad , \quad x-3 = 0.502\,(y-5)$$

Thus $y = 1.306x + 1.082$ & $x = 0.502y + 0.49$

There are the lines of regression.

⑦ Find the correlation coefficient b/w 'x' & 'y' for the following data. Also obtain the regression lines.

x	1	2	3	4	5	6	7	8	9	10
y	10	12	16	28	25	36	41	49	40	50

Soln:- Here $n = 10$

x	y	$z = x-y$	x^2	y^2	z^2
1	10	-9	1	100	81
2	12	-10	4	144	100
3	16	-13	9	256	169
4	28	-24	16	784	576
5	25	-20	25	625	400
6	36	-30	36	1296	900
7	41	-34	49	1681	1156
8	49	-41	64	2401	1681
9	40	-31	81	1600	961
10	50	-40	100	2500	1600
$\Sigma x = 55$	$\Sigma y = 307$	$\Sigma z = -252$	$\Sigma x^2 = 385$	$\Sigma y^2 = 11387$	$\Sigma z^2 = 7624$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{55}{10} = 5.5 \, , \quad \bar{y} = \frac{\Sigma y}{n} = \frac{307}{10} = 30.7 \, , \quad \bar{z} = \frac{\Sigma z}{n} = \frac{-252}{10} = -25.2$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{385}{10} - (5.5)^2 = 8.25, \quad \sigma_x = 2.87$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{11387}{10} - (30.7)^2 = 196.21, \quad \sigma_y = 14.01$$

$$\sigma_z^2 = \sigma_{x-y}^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = \frac{7624}{10} - (-25.2)^2 = 127.36$$

we have $r = \dfrac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}$

$$r = \frac{8.25 + 196.21 - 127.36}{2 \times 2.87 \times 14.01} = 0.96$$

Thus $r = 0.96$

Eqn of the line of regression are

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

$$y - 30.7 = 0.96 \times \frac{14.01}{2.87}(x - 5.5), \quad (x - 5.5) = 0.96 \times \frac{2.87}{14.01}(y - 30.7)$$

$y = 4.686x + 4.927$ and $x = 0.197y - 0.548$

There are lines of regression.

(8) Find the regression line of 'y' on 'x' for the following data.

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

Soln :- $\bar{x} = \frac{\Sigma x}{n} = \frac{56}{8} = 7, \quad \bar{y} = \frac{\Sigma y}{8} = \frac{40}{8} = 5$

we denote $X = x - \bar{x}$ and $Y = y - \bar{y}$

$X = x - 7, \quad Y = y - 5$

we have lines of regression in the form

$$Y = \frac{\Sigma XY}{\Sigma X^2} X \qquad X = \frac{\Sigma XY}{\Sigma Y^2} \cdot Y.$$

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	X^2	Y^2	XY
1	1	-6	-4	36	16	24
3	2	-4	-3	16	9	12
4	4	-2	-1	4	1	2
6	4	-1	-1	1	1	1
8	5	1	0	1	0	0
9	7	2	2	4	4	4
11	8	4	3	16	9	12
14	9	7	4	49	16	28

$\Sigma x = 56 \quad \Sigma y = 40 \qquad \Sigma X^2 = 132 \quad \Sigma Y^2 = 56 \quad \Sigma XY = 84$

i.e $y - \bar{y} = \dfrac{\Sigma XY}{\Sigma X^2}(x-\bar{x})$, $x - \bar{x} = \dfrac{\Sigma XY}{\Sigma Y^2}(y-\bar{y})$

$y - 5 = \dfrac{84}{132}(x-7) \qquad x-7 = \dfrac{84}{56}(y-5)$

$y-5 = 0.63(x-7) \qquad x-7 = 1.\cancel{4285}\,5\,(y-5)$

$y-5 = 0.63x - 4.41 \qquad x-7 = 1.\cancel{4285}\,5\,y - 7.\cancel{1425}\,5$

$y = 0.63x - 4.41 + 5 \qquad x = 1.\cancel{4285}\,5\,y - 7.\cancel{1425}\,5 + 7$

$y = 0.66x + 0.59 \qquad x = 1.\cancel{4285}\,5\,y - 0.\cancel{1425}\,5$

These are the lines of regression.

* (9) Calculate the Karl pearson's coefficient of correlation for 10 students who have obtained the following % of marks in mathematics & Electronics.

Roll No	1	2	3	4	5	6	7	8	9	10
Marks in mathematics	78	36	98	25	75	82	90	62	65	39
Marks in Electronics	84	51	91	60	68	62	86	58	53	47

Sol^n:- we have $\bar{x} = \dfrac{\Sigma x}{n} = \dfrac{650}{10} = 65$

$\bar{y} = \dfrac{\Sigma y}{n} = \dfrac{660}{10} = 66$

Let $X = x - \bar{x}$ & $Y = y - \bar{y}$

$X = x - 65 \qquad Y = y - 66$

x	y	$X=x-\bar{x}$	$Y=y-\bar{y}$	X^2	Y^2	XY
78	84	13	18	169	324	234
36	51	-29	-15	841	225	435
98	91	33	25	1089	625	825
25	60	-40	-6	1600	36	240
75	68	10	2	100	4	20
82	62	17	-4	289	16	68
90	86	25	20	625	400	500
62	58	-3	-8	9	64	24
65	53	0	-13	0	169	0
39	47	-26	-19	676	361	494
				$\Sigma X^2=5398$	$\Sigma Y^2=2224$	$\Sigma XY=2840$

$$r=\frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}}=\frac{2840}{\sqrt{5398}\sqrt{2224}}=\frac{2840}{73.47\times 47.15}=\frac{2840}{3464.1105}$$

$$r=0.819\approx 0.8$$

★ (10) In a partially destroyed lab record, only the lines of regression of y on x & x on y are available as $4x-5y+33=0$ & $20x-9y=107$ respectively. Calculate $\bar{x}$, $\bar{y}$ and coefficient of correlation b/w 'x' & 'y'.

Soln: we know that regression lines passes through $\bar{x}$ & $\bar{y}$.

$$4\bar{x}-5\bar{y}=-33$$
$$20\bar{x}-9\bar{y}=107$$
$$\bar{x}=13,\quad \bar{y}=17$$

we shall now rewrite the eqn of the regression lines to find the regression coefficients.

$$5y=4x+33 \quad (\text{or}) \quad y=0.8x+6.6 \rightarrow (1)$$
$$20x=9y+107 \quad (\text{or}) \quad x=0.45y+5.35 \rightarrow (2)$$

From ① & ②

$r \cdot \frac{\sigma_y}{\sigma_x} = 0.8$, $r \frac{\sigma_x}{\sigma_y} = 0.45$

correlation coefficient $r = \sqrt{0.8 \times 0.45} = \pm 0.6$ *

Thus $r = 0.6$

★ Show that if 'θ' is the angle b/w the lines of regression, then

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

Soln:- W.K.T if θ is a acute, the angle b/w the lines $y = m_1 x + c_1$ and $y = m_2 x + c_2$ is given by.

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

we have the lines of regression,

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

we write the second of the eqn as

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

Slopes of ① & ② are respectively given by

$$m_1 = \frac{r\sigma_y}{\sigma_x} \quad \& \quad m_2 = \frac{\sigma_y}{r\sigma_x}$$

Substituting these in the formula for $\tan\theta$ we have,

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

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

RANK CORRELATION
18MAT41

MODULE-IV

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

Sol.

Phy(x)	**Mat(** y)	R_x	R_y	$d_i = R_x - R_y$	${d_i}^2$
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**
					$\sum {d_i}^2 =$ **12**

and $n = 9$

We know that The Spearman correlation coefficient is

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

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

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

$$\Rightarrow \rho = 1 - 0.1$$

$$\Rightarrow \rho = 0.9$$

2. The scores for 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 the students in the two subjects and also compute the Spearman's rank correlation.

Sol.

Eng(x)	**Mat(y)**	R_x	R_y	$d_i = R_x - R_y$	$d_i^{\,2}$
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**
					$\sum d_i^{\,2} =$ **54**

and $n = 10$

We know that The Spearman correlation coefficient is

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

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

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

$$\Rightarrow \rho = 1 - 0.33$$

$$\Rightarrow \rho = 0.67$$

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.

Sol.

Estimated Cost(x)	**Actual Cost(y)**	R_x	R_y	$d_i = R_x - R_y$	$d_i^{\,2}$
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
					$\sum d_i^{\,2} = 8$

and $n = 8$

We know that The Spearman correlation coefficient is

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

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

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

$$\Rightarrow \rho = 1 - 0.095$$

$$\Rightarrow \rho = 0.9047$$

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

Sol.

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**
			$\sum d_i^{\,2} =$ **226**

We know that The Spearman correlation coefficient is

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

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

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

$$\Rightarrow \rho = 1 - 1.37$$

$$\Rightarrow \rho = -0.37$$

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.

Sol.

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**
			$\sum d_i^2 =$ **60**

We know that The Spearman correlation coefficient is

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

$$\Rightarrow \rho = 1 - \frac{6(60)}{10^3 - 10}$$
$$\Rightarrow \rho = 1 - \frac{360}{990}$$
$$\Rightarrow \rho = 1 - 0.3636$$
$$\Rightarrow \rho = 0.6363$$

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.

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**
			$\sum d_i^{\,2} = \mathbf{128}$

We know that The Spearman correlation coefficient is

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

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

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

$$\Rightarrow \rho = 1 - 0.7758$$

$$\Rightarrow \rho = 0.2242$$

***