Unit - 5

Curve fitting :

Let x be an independent variable and y an variable depending on x ; i.e $y = f(x)$. If $f(x)$ is a known function, then for any n allowable values $x_1, x_2 \dots x_n$ of x, we can find the corresponding values $y_1, y_2 \dots y_n$ and thereby determine the pairs $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$. These pairs of values of x and y give us points on the curve $y = f(x)$.

Suppose we consider the converse problem. That is, suppose we are given n values $x_1, x_2 \dots x_n$ of an independent variable x & the corresponding values $y_1, y_2 \dots y_n$ of a variable depending on x. Then the point $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ give us n points in the xy plane. Generally, it is not possible to find the actual curve $y = f(x)$ that passes through these points. Hence we try to find a curve that serves as best approximation to the curve $y = f(x)$. Such a curve is referred to as the curve of best fit. The process of determining a curve of best fit is called curve fitting. The method generally employed for curve fitting is known as the method of least squares.

Fitting of a straight line :-

Suppose a straight line that serves as best approximation to the actual curve $y = f(x)$ passing through n given points $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$. The line will be referred to as the line of best fit and we take its eqn as $y = a + bx$ — (1), where a & b are parameter to be determined. Let Y_i be the value of y corresponding to the value x_i of x as determined by eqn (1). Also, let $S = \Sigma (y_i - Y_i)^2$.

Using (1), this becomes $S = \Sigma (y_i - a - bx_i)^2$ — (2)
We determine a & b so that S is minimum (least).
Two necessary conditions for this are

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

AUGUST						2008
S	M	T	W	T	F	S
31					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

Using (2), These conditions yield the following equation:-

$\Sigma(y_i - a - bx_i) = 0 \Rightarrow \boxed{\Sigma y_i = na + b\Sigma x_i}$ —— (3)

& $\Sigma(y_i - a - bx_i)x_i = 0 \Rightarrow \boxed{\Sigma x_i y_i = a\Sigma x_i + b\Sigma x_i^2}$ —— (4)

These two eqn (3) & (4), called normal eqn, serve as two simultaneous eqn for determining a & b. Putting these values of a & b so determined in (1), We get the eqn of the line of best fit for the given data.

1) Fit a straight line to the following data:-

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

Estimate the value of y when $x = 3.5$.

Sol. Given $n = 5$ & the Table is.

x_i	y_i	x_i^2	$x_i y_i$
1	14	1	14
2	13	4	26
3	9	9	27
4	5	16	20
5	2	25	10
$\Sigma x_i = 15$	$\Sigma y_i = 43$	$\Sigma x_i^2 = 55$	$\Sigma x_i y_i = 97$

$\therefore \Sigma y_i = na + b\Sigma x_i$

$43 = 5a + 15b$ —— (1)

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

$97 = 15a + 55b$ —— (2)

On solving (2) & (1), we get $a = 18.2$ & $b = -3.2$. Hence, for the given data, the line of best fit is $y = 18.2 - 3.2x$.

When $x = 3.5$, the estimated value of y (found from the line of best fit) is $y = 18.2 - (3.2)(3.5)$

$y = 7$

2) Fit a straight line (by the method of least squares) to the following data:

x	1	2	3	4	5
y	35	68	100	138	170

$\Rightarrow$ Given $n = 5$ & the following table is

x_i	y_i	x_i^2	$x_i y_i$
1	35	1	35
2	68	4	136
3	100	9	300
4	138	16	552
5	170	25	850
$\Sigma x_i = 15$	$\Sigma y_i = 511$	$\Sigma x_i^2 = 55$	$\Sigma x_i y_i = 1873$

$\therefore \Sigma y_i = na + b\Sigma x_i$

$511 = 5a + 15b$ —— (1)

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

$1873 = 15a + 55b$ —— (2)

On solving (1) & (2), we get $a = 0.2$ & $b = 34$. Hence for the given data, the line of best fit is $y = 0.2 + 34x$.

Exercise

① Fit a straight line by the method of least squares to each of the following data:-

a)

x	0	1	2	3	4
y	1	1.8	2.3	4.5	6.3

$\Rightarrow$ Given $n = 5$ & the following table is.

x_i	y_i	x_i^2	$x_i y_i$
0	1	0	0
1	1.8	1	1.8
2	2.3	4	4.6
3	4.5	9	13.5
4	6.3	16	25.2
$\Sigma x_i = 10$	$\Sigma y_i = 15.9$	$\Sigma x_i^2 = 30$	$\Sigma x_i y_i = 45.1$

$\therefore \Sigma y_i = na + b\Sigma x_i$

$15.9 = 5a + 10b$ —— (1)

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

$45.1 = 10a + 30b$ —— (2)

On solving (1) & (2), we get $a = 0.52$ & $b = 1.33$. The line of best fit is $y = 0.52 + 1.33x$

AUGUST						2008
S	M	T	W	T	F	S
31					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

b)

x	1	2	3	4	6	8
y	2.4	3	3.6	4	5	6

⇒ Given $n = 6$ & Table is

x_i	y_i	x_i^2	x_iy_i
1	2.4	1	2.4
2	3	4	6
3	3.6	9	10.8
4	4	16	16
6	5	36	30
8	6	64	48
$\Sigma x_i =$ 24	$\Sigma y_i =$ 24	$\Sigma x_i^2 =$ 130	$\Sigma x_iy_i =$ 113.2

$\therefore \Sigma y_i = na + b\Sigma x_i$

$24 = 6a + 24b$ —— (1)

$\Sigma x_iy_i = a\Sigma x_i + b\Sigma x_i^2$

$113.2 = 24a + 130b$ —— (2)

On solving (1) & (2), we get $a = 2$ and $b = 0.5$. Hence for the given data, The line of best fit is $y = 2 + 0.5x$

c)

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

⇒ Given $n=7$ and table is

x_i	y_i	x_i^2	x_iy_i
1	80	1	80
2	90	4	180
3	92	9	276
4	83	16	332
5	94	25	470
6	99	36	594
7	92	49	644
$\Sigma x_i =$ 28	$\Sigma y_i =$ 630	$\Sigma x_i^2 =$ 140	$\Sigma x_iy_i =$ 2576

$\therefore \Sigma y_i = na + b\Sigma x_i$

$630 = 7a + 28b$ —— (1)

$\Sigma x_iy_i = a\Sigma x_i + b\Sigma x_i^2$

$2576 = 28a + 140b$ —— (2)

On solving (1) & (2), we get $a = 82$ and $b = 2$. Hence for the given data, The line of best fit is $y = 82 + 2x$

(2) Using the method of least squares, fit a linear relation of the form $P = a + bW$ to the following data when P is the pull required to lift a weight W: Estimate P when W is 150 kg.

W (kg) :	50	70	100	120
P (kg) :	12	15	21	25

⇒ Given $n = 4$ & Table is

W(x_i)	P(y_i)	x_i^2	x_iy_i
50	12	2500	600
70	15	4900	1050
100	21	10,000	2100
120	25	14,400	3000
$\Sigma x_i =$ 340	$\Sigma y_i =$ 73	$\Sigma x_i^2 =$ 31,800	$\Sigma x_iy_i =$ 6750

$\therefore \Sigma y_i = na + b\Sigma x_i$

$73 = 4a + 340b$ —— (1)

$\Sigma x_iy_i = a\Sigma x_i + b\Sigma x_i^2$

$6750 = 340a + 31800b$ —— (2)

On solving (2) & (1), we get $a = 2.95$ and $b = 0.18$. Hence, for the given data, the line of best fit is $y = 2.95 + 0.18x$.

When $x = 150$, the estimated value of P is

$P = 2.95 + 0.18W = 2.95 + 0.18(150) = 2.95 + 27$

$P = \underline{29.95}$

(3) The weights of a calf taken at weekly intervals are given below. Fit a straight line using the method of least squares & estimate the weight when the age is 11 years:

Age in Years (x)	1	2	3	4	5	6	7	8	9	10
Weight in Kgs (y)	52.5	58.7	65.0	70.2	75.4	81.1	87.2	95.3	102.2	108.4

→ Given $n = 10$ & table is

x_i	y_i	x_i^2	$x_i y_i$
1	52.5	1	52.5
2	58.7	4	117.4
3	65.0	9	195
4	70.2	16	280.8
5	75.4	25	377
6	81.1	36	486.6
7	87.2	49	610.4
8	95.3	64	762.4
9	102.2	81	919.8
10	108.4	100	1084
$\Sigma x_i =$ 55	$\Sigma y_i =$ 796	$\Sigma x_i^2 =$ 385	$\Sigma x_i y_i =$ 4885.9

$\therefore \Sigma y_i = na + b\Sigma x_i$

$796 = 10a + 55b$ —— (1)

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

$4885.9 = 55a + 385b$ —— (2)

On solving (1) & (2), we get $a = 45.77$ and $b = 6.15$. Hence, for the given data, the line of best fit is $y = 45.77 + 6.15x$.

When $x = 11$, the estimated weight is

$y = 45.77 + 6.15(11)$

$y = 45.77 + 67.65$

$y = \underline{113.42}$

Fitting of a curve of the form $y = ab^x$

Suppose we wish to have a curve whose eqn is of the form $y = ab^x$ —— (1) as the curve of best fit for a data consisting of the pairs (x_i, y_i), where $i = 1, 2, 3, \ldots n$

Taking log on both sides of (1), we get

$\log y = \log a + x\log b$

$U = A + xB$ —— (2)

where $U = \log y$, $A = \log a$, $B = \log b$

eqn (2) is a linear eqn of the type $y = a + bx$.

Therefore the normal eqn that yield A & B are

$\Sigma U_i = nA + B\Sigma x_i$ —— (3)

$\Sigma x_i U_i = A\Sigma x_i + B\Sigma x_i^2$ —— (4), where $U_i = \log y_i$

Solving these equations, we obtain A & B from which $a = \text{antilog } A$ & $b = \text{antilog } B$ can be found. Substituting these values of a & b in (1), we obtain the curve of best fit for the given data.

1) Fit a curve of the form $y = ab^x$ for the following data:-

x	1	2	3	4	5	6	7	8
y	1.0	1.2	1.8	2.5	3.6	4.7	6.6	9.1

→ Given $n = 8$ & the table is $U_i = \log y_i$

x_i	y_i	$U_i = \log y_i$	x_i^2	$x_i U_i$
1	1.0	0	1	0
2	1.2	0.1823	4	0.3646
3	1.8	0.5878	9	1.7634
4	2.5	0.9163	16	3.6652
5	3.6	1.2809	25	6.4045
6	4.7	1.5476	36	9.2856
7	6.6	1.8871	49	13.2097
8	9.1	2.2083	64	17.6664
$\Sigma x_i =$ 36	$\Sigma y_i =$ 30.5	$\Sigma U_i =$ 8.6103	$\Sigma x_i^2 =$ 204	$\Sigma x_i U_i =$ 52.3594

$U_i = nA + xB$ — (1)

$U_i x_i = xA + x_i^2 B$

The normal eqn that determine the curve of best fit are $8.6103 = 8A + 36B$ (1) & $52.3594 = 36A + 204B$ (2)

On solving (1) & (2), we get $A = -0.3822$ & $B = 0.3241$.
Therefore, $a = \text{antilog } A = 0.6824$ &
$b = \text{antilog } B = 1.3828$

Hence, for the given data, the curve of best fit of the required is $y = (0.6824)(1.3828)^x$.

2) Find the best values of a & b in the formula $y = ab^x$ to fix the following data by the method of least squares.

x	0	2	4	5	7	10
y	100	120	256	390	710	1600

⇒ Given $n = 6$ & table is

x_i	y_i	$U_i = \log_e y_i$	x_i^2	$x_i U_i$
0	100	4.6052	0	0
2	120	4.7875	4	9.575
4	256	5.5452	16	22.1808
5	390	5.9661	25	29.8305
7	710	6.5653	49	45.9571
10	1600	7.3778	100	73.778
$\Sigma x_i = 28$	$\Sigma y_i = 3176$	$\Sigma U_i = 34.8471$	$\Sigma x_i^2 = 194$	$\Sigma x_i U_i = 181.3214$

The normal eqn that determine the curve of the best fit are $34.8471 = 6A + 28B$ &
$181.3214 = 28A + 194B$

On solving A & B, we get $A = 4.4298$ & $B = 0.2953$.

Therefore, $a = \text{antilog } A = 83.9146$
$b = \text{Antilog } B = 1.3435$.

Hence, for the given data, the curve of best fit of the required form is $y = (83.9146)(1.3435)^x$.

Fitting of a Parabola :-

Suppose we wish to have a parabola (second-degree curve) as the curve of best fit for a data consisting of n given pairs (x_i, y_i), $1 \le i \le n$.
Let us take the eqn of this parabola called parabola of best fit, in the form $y = a + bx + cx^2$ —(1) where a, b & c are parameters to be determined.
Let Y_i be the value of y corresponding to the value of x_i of x as determined by eqn (1), & $S = \Sigma(y_i - Y_i)^2$.
Using (1), this becomes $S = \Sigma(y_i - a - bx_i - (x_i^2))^2$ — (2)
We determine a, b, c so that S is least (minimum).
Three necessary conditions for this are

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

Using (2), these conditions yield the following eqns :-

$$\Sigma y_i = na + b\Sigma x_i + c\,\Sigma x_i^2 \quad \text{—— (3)}$$

$$\Sigma x_i y_i = a\Sigma x_i + b\Sigma x_i^2 + c\,\Sigma x_i^3 \quad \text{—— (4)}$$

$$\Sigma x_i^2 y_i = a\Sigma x_i^2 + b\Sigma x_i^3 + c\,\Sigma x_i^4 \quad \text{—— (5)}$$

These three eqn, called the normal eqn, serve as three simultaneous eqn for determining a, b & c. Putting these values of a, b, c so determined in (1), we get the eqn of the parabolas of best fit for the given data :.

1) Fit a parabola to the following data :

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

Estimate y when $x = 4.5$.

⇒ Given $n = 9$ & the table is.

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

The normal eqn that determine the parabola of best fit are

$$74 = 9a + 45b + 285c$$
$$421 = 45a + 285b + 2025c$$
$$2771 = 285a + 2025b + 15333c$$

On solving, we obtain $a = -0.9282$, $b = 3.523$ & $c = -0.2673$.
Hence the parabola of best fit for the given data is

$$y = -0.9282 + 3.523x - 0.2673x^2$$

For $x = 4.5$, the parabola of best fit gives the estimated value of y as

$$y = -0.9282 + (3.523)(4.5) - (0.2673)(4.5)^2$$
$$y = 9.512$$

SEPTEMBER 2008

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

Mode → eqn → 1 → a_1 = → b_1 = c_1 =

September Read → Shift → mode → 3 → 2times =

10
Wednesday
Wk-37 • 254-112

2) Fit a parabola $y = a + bx + cx^2$ by the method of least squares to the following data.

x	1	2	3	4	5	6	7
y	2.3	5.2	9.7	16.5	29.4	35.5	54.4

⇒ Given $n = 7$ & table is

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
1	2.3	1	1	1	2.3	2.3
2	5.2	4	8	16	10.4	20.8
3	9.7	9	27	81	29.1	87.3
4	16.5	16	64	256	66.0	264.0
5	29.4	25	125	625	147.0	735.0
6	35.5	36	216	1296	213.0	1278.0
7	54.4	49	343	2401	380.8	2665.6
$\Sigma x_i =$ 28	$\Sigma y_i =$ 153.	$\Sigma x_i^2 =$ 140	$\Sigma x_i^3 =$ 784	$\Sigma x_i^4 =$ 4676	$\Sigma x_i y_i =$ 848.6	$\Sigma x_i^2 y_i =$ 5053

The normal equation that determine the parabola of best fit are

$$153 = 7a + 28b + 140c$$
$$848.6 = 28a + 140b + 784c$$
$$5053 = 140a + 784b + 4676c$$

Solving these eqn, we obtain $a = 2.373$, $b = -1.094$, $c = 1.193$. Hence the parabola of best fit for the given data is

$$y = 2.3731 - 1.094x + 1.193x^2$$

11
Thursday
Wk-37 • 255-111

Exercise

1) Fit a parabola to each of the following data:

a)

x	-1	0	0	1
y	2	0	1	2

⇒ Given $n = 4$ & table is

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
-1	2	1	-1	1	-2	2
0	0	0	0	0	0	0
0	1	0	0	0	0	0
1	2	1	1	1	2	2
$\Sigma x_i =$ 0	$\Sigma y_i =$ 5	$\Sigma x_i^2 =$ 2	$\Sigma x_i^3 =$ 0	$\Sigma x_i^4 =$ 2	$\Sigma x_i y_i =$ 0	$\Sigma x_i^2 y_i =$ 4

The normal equation that determine the parabola of best fit are

$$5 = 4a + 0b + 2c$$
$$0 = 0a + 2b + 0c$$
$$4 = 2a + 0b + 2c$$

Solving three eqn, we obtain $a = 0.5$, $b = 0$, $c = 1.5$. Hence the parabola of best fit for the given data is

$$y = 0.5 + 0x + 1.5x^2$$

$$y = 0.5 + 1.5x^2$$

SEPTEMBER						2008
S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

b)

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

$\Rightarrow$ Given $n = 7$ & the table is

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
1.0	1.1	1	1	1	1.1	1.1
1.5	1.3	2.25	3.375	5.0625	1.95	2.925
2.0	1.6	4	8	16	3.2	6.4
2.5	2.0	6.25	15.625	39.0625	5	12.5
3.0	2.7	9	27	81	8.1	24.3
3.5	3.4	12.25	42.875	150.0625	11.9	41.65
4.0	4.1	16	64	256	16.4	65.6
$\Sigma x_i =$ 17.5	$\Sigma y_i =$ 16.2	$\Sigma x_i^2 =$ 50.75	$\Sigma x_i^3 =$ 161.875	$\Sigma x_i^4 =$ 548.1875	$\Sigma x_i y_i =$ 47.65	$\Sigma x_i^2 y_i =$ 154.475

The normal equation that determine the parabola of best fit are

$$16.2 = 7a + 17.5b + 50.75c$$
$$47.65 = 17.5a + 50.75b + 161.875c$$
$$154.475 = 50.75a + 161.875b + 548.1875c$$

Solving three eqn, we obtain $a = 1.035$, $b = -0.192$, $c = 0.24$. Hence the parabola of best fit for the given data is

$$y = 1.035 - 0.192x + 0.24x^2$$

② The following table gives the result of ~~measureable~~ measurement of train resistance :-

V	20	40	60	80	100	120
R	5.5	9.1	14.9	22.8	33.3	46.0

Here V is the velocity in km/hr and R is the resistance in Kg/quintal. Fit a relation of the form $R = a + bv + cv^2$ to the data. Estimate R when $V = 90$.

$\Rightarrow$ Given $n = 6$ & the table is

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
20	5.5	400	8000	160000	110	2200
40	9.1	1600	64,000	2560,000	364	14560
60	14.9	3600	216,000	12960,000	894	53640
80	22.8	6400	512,000	40960,000	1824	145,920
100	33.3	10,000	1000000	100000000	3330	3,33000
120	46.0	14,400	1728000	207360,000	5520	662400
$\Sigma x_i =$ 420	$\Sigma y_i =$ 131.6	$\Sigma x_i^2 =$ 36400	$\Sigma x_i^3 =$ 3528000	$\Sigma x_i^4 =$ 364000,000	$\Sigma x_i y_i =$ 12042	$\Sigma x_i^2 y_i =$ 1211720

The normal equation that determine the parabola of best fit are

$$131.6 = 6a + 420b + 36400c$$
$$12042 = 420a + 36400b + 3528000c$$
$$1211720 = 36400a + 3528000b + 364000000c$$

Solving three eqn, we obtain $a = 4.35$, $b = 2.4 \times 10^{-3}$, $c = 2.8 \times 10^{-3}$

Hence the parabola of best fit for the given data is.

$$y = 4.35 + 0.0024x + 0.0028x^2$$
$$y = 4.35 + 0.0024(90) + 0.0028(90)^2$$
$$y = 4.35 + 0.216 + 22.68$$
$$R = 27.246$$

Correlation & Regression

Co-variation of two independent magnitudes is known as correlation. If two variables x & y are related in such a way that increase or decrease in one of them corresponds to increase or decrease in the other, then we say that it is positive correlated.

If increase or decrease in one of them corresponds to decrease or increase in the other, the variable are said to be negatively correlated.

The numerical measure of correlation between the variation x & y is known as Pearson's co-efficient of correlation usually denoted by r & is defined as follows:-

$$r = \frac{\sum_{i}^{n}(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} \quad \text{——— (1)}$$

This can be put in an alternative form as follows:-

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

Standard deviation $\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma X^2}{n}}$

$$\Rightarrow \sigma_x^2 = \frac{\Sigma X^2}{n}$$

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

$\therefore \sigma_x\sigma_y = \frac{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}}{n} \Rightarrow n\,\sigma_x\sigma_y = \sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}$

Thus (1) becomes $r = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}}$ ——— (2)

Note :-

1) The co-efficient of correlation numerically does not exceed unity i.e, $-1 \le r \le 1$.
2) If $r = \pm 1$, we say that x & y are perfectly correlated.
3) If $r = 0$, we say that x & y are non correlated.
4) $\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2$, $\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2$, $\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2$

Alternative formula for r:-

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

proof: Let $z = x - y$

$$\Rightarrow \frac{\Sigma z}{n} = \frac{\Sigma x}{n} - \frac{\Sigma y}{n}$$

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

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

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

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

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

$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$ from (1)

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

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

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

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

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

Derivation of the eqn of the regressive line :

Let $y = ax + b$ be the equation of the regression line of y on x for a given set of n values (x, y). The associated normal eqn are

$nb + a\Sigma x = \Sigma y$ ——(1)

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

(1) ÷ by $n \Rightarrow b + a\frac{\Sigma x}{n} = \frac{\Sigma y}{n}$

$\Rightarrow b + a\bar{x} = \bar{y}$ or $\bar{y} = a\bar{x} + b$

This shows that the regression line passes through $(\bar{x}, \bar{y})$. We also know that the eqn of a straight line passing through (x_1, y_1) having slope is $(y - y_1) = m(x - x_1)$ and hence if $(x_1, y_1) = (\bar{x}, \bar{y})$ we get $(y-\bar{y}) = m(x-\bar{x})$ ——— (3)

SEPTEMBER						2008
S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

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

$\Rightarrow Y = mX$

Multiply by X & taking summation on both sides

$$\Sigma XY = \Sigma mX^2$$

$$\Rightarrow m = \frac{\Sigma XY}{\Sigma X^2}$$

Consider, $r = \frac{\Sigma XY}{n\sigma_x\sigma_y} \Rightarrow \Sigma XY = nr\sigma_x\sigma_y$

& also $\sigma_x^2 = \frac{\Sigma X^2}{n} \Rightarrow \Sigma X^2 = n\sigma_x^2$

$$\therefore m = \frac{nr\sigma_x\sigma_y}{n\sigma_x^2} = \frac{r\sigma_y}{\sigma_x}$$

$\therefore$ (3) becomes $(y - \bar{y}) = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$ — (4)

This is the eqn of regression line of y on x. Similarly assuming the eqn in the form $x = ay + b$ & proceeding on the same lines as before we obtain

$$x - \bar{x} = r \cdot \frac{\sigma_x}{\sigma_y}(y - \bar{y}) \quad \text{— (5)}$$

This is the regression line of x on y. The coefficient of x in (4) is $r\frac{\sigma_y}{\sigma_x}$ & co-efficient of y in (5) is $r\frac{\sigma_x}{\sigma_y}$ are known as regression co-efficient. Their product is equal to r^2.

Alternate formula :-

$$Y = \frac{\Sigma XY}{\Sigma X^2}X \quad \& \quad X = \frac{\Sigma XY}{\Sigma Y^2}Y$$

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

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

Note :-

1) $r = \pm\sqrt{(\text{Co-effi of } x)(\text{Co-effi of } y)}$

2) If $\bar{x}$ & $\bar{y}$ are integers, computation of r by the formula

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

3) Rank Correlation

$$\rho = 1 - \frac{6\Sigma(x-y)^2}{n(n^2-1)} \quad \text{or} \quad \rho = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

a) If the rankings of x, y are entirely in the same order like for example.

x : 1, 2, 3, 4, 5 ; y : 1, 2, 3, 4, 5 then

$\Sigma d^2 = \Sigma(x-y)^2 = 0$ This will give us $\rho = +1$ & is called perfect direct correlation.

b) If the rankings of x & y are entirely in the opposite order like for eg: x : 1, 2, 3, 4, 5 & y : 5, 4, 3, 2, 1 then

$$\Sigma d^2 = \Sigma(x-y)^2$$
$$= (1-5)^2 + (2-4)^2 + (3-3)^2 + (4-2)^2 + (5-1)^2 = 40$$

This will give us $\rho = -1$ & it is called perfect inverse correlation.

4) Standard deviation :-

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

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

$\frac{\bar{x}^2}{n}\Sigma 1$

(or)

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

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

$\Sigma 1 \quad n$

$$= \frac{1}{n}\Sigma x^2 - 2\bar{x}^2 + \bar{x}^2$$

$$\sigma_x^2 = \frac{1}{n}\Sigma x^2 - \bar{x}^2$$

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

$$\sigma_z^2 = \frac{1}{n}\Sigma z^2 - (\bar{z})^2$$

SEPTEMBER 2008

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

1) Find the correlation for the following data & also obtain line of regression.

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

Find the best estimate for y when $x = 3.5$ & also for x when $y = 3.5$.

⇒ Here $n = 5$ values of (x, y) are given

The mean (average) of the y-series is

$$\bar{y} = \frac{\Sigma y}{n} = \frac{2+5+3+8+7}{5} = 5$$

The mean (average) of the x-series is

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1+2+3+4+5}{5} = 3$$

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	X^2	Y^2	XY
1	2	-2	-3	4	9	6
2	5	-1	0	1	0	0
3	3	0	-2	0	4	0
4	8	1	3	1	9	3
5	7	2	2	4	4	4
				$\Sigma X^2 = 10$	$\Sigma Y^2 = 26$	$\Sigma XY = 13$

Then correlation coefficient between x & y is

$$r = \frac{\Sigma XY}{\sqrt{(\Sigma x)^2 (\Sigma y)^2}} = \frac{13}{\sqrt{10 \times 26}} = \frac{13}{\sqrt{260}} = \frac{13}{16.1245} = 0.8062.$$

The lines of regression are

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

$$y - 5 = (0.8062)\frac{\sigma_y}{\sigma_x}(x - 3)$$

$$y - 5 = (0.8062)\frac{(2.2803)}{1.4142}(x - 3)$$

$$\sigma_x^2 = \frac{\Sigma X^2}{n} = \frac{10}{5} = 2 \Rightarrow \sigma_x = 1.4142$$

$$\sigma_y^2 = \frac{\Sigma Y^2}{n} = \frac{26}{5} = 5.2 \Rightarrow \sigma_y = 2.2803$$

$$y - 5 = \frac{1.8383}{1.4142}(x - 3)$$

$$y - 5 = 1.2998(x - 3)$$

$$y - 5 = 1.2998x - 3.8994$$

$$y = 1.2998x + 1.1006$$

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

$$(x - 3) = 0.8062 \frac{(1.4142)}{2.2803}(y - 5)$$

$$(x - 3) = 0.4999(y - 5)$$

$$(x - 3) = 0.4999y - 2.4995$$

$$x = 0.4999y - 2.4995 + 3$$

$$x = 0.4999y + 0.5005$$

When $x = 3.5$

$$y = 1.2998(3.5) + 1.1006$$

$$y = 4.5493 + 1.1006$$

$$y = 5.6499$$

when $y = 3.5$

$$x = 0.4999(3.5) + 0.5005$$

$$= 1.74965 + 0.5005$$

$$x = 2.25015$$

2) Obtain the lines of regression and hence find the co-efficient of correlation for the data.

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

⇒ Here $n = 7$ values of (x, y) are given

$$\bar{x} = \frac{\Sigma x}{n} = 4, \quad \bar{y} = \frac{\Sigma y}{n} = 11$$

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

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \, \Sigma Y^2}} = \frac{26}{\sqrt{28 \times 28}} = \frac{26}{28} = 0.9285$$

SEPTEMBER 2008

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

22 Monday Wk-39 • 266-100

$\sigma_x^2 = \frac{\Sigma x^2}{n} = \frac{28}{7} = 4 \Rightarrow \sigma_x = 2$

$\sigma_y^2 = \frac{\Sigma y^2}{n} = \frac{28}{7} = 4 \Rightarrow \sigma_y = 2$

The line of regression are

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

$(y-11) = (0.9285)\frac{2}{2}(x-4)$

$y = 0.9285x + 7.286$ —(1)

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

$(x-4) = \frac{0.9285(2)}{2}(y-11)$

$x = 0.9285y - 10.2135 + 4$

$x = 0.9285y - 6.2135$ —(2)

(1) & (2) are the line of regression.

3) Find the correlation coefficient for the two groups.

A	92	89	87	86	83	77	71	63	53	50
B	86	83	91	77	68	85	52	82	37	57

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

⇒ Here $n=10$, we shall denote $z = x-y$

x	y	$z = x-y$	x^2	y^2	z^2
92	86	6	8464	7396	36
89	83	6	7921	6889	36
87	91	-4	7569	8281	16
86	77	9	7396	5929	81
83	68	15	6889	4624	225
77	85	-8	5929	7225	64
71	52	19	5041	2704	361
63	82	-19	3969	6724	361
53	37	16	2809	1369	256
50	57	-7	2500	3249	49
$\Sigma x = 751$	$\Sigma y = 718$	$\Sigma z = 33$	Σx^2 58487	54390	1485

23 Tuesday Wk-39 • 267-099

$\bar{x} = \frac{\Sigma x}{n} = 75.1$, $\bar{y} = \frac{\Sigma y}{n} = 71.8$, $\bar{z} = \frac{\Sigma z}{n} = 3.3$

$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{58487}{10} - (75.1)^2 = 208.69 \Rightarrow \sigma_x = 14.446$

$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = 283.76 \Rightarrow \sigma_y = 16.845$

$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = 137.61$

$\therefore r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y} = 0.73$ is the required coefficient.

4) If r is the correlation coefficient between x and y and $z = ax + by$, Then show that $r = \frac{\sigma_z^2 - (a^2\sigma_x^2 + b^2\sigma_y^2)}{2ab\sigma_x\sigma_y}$

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

⇒ Let $z = ax + by$ we have $\bar{z} = a\bar{x} + b\bar{y}$ where $\bar{x}, \bar{y}$ & $\bar{z}$ are the mean of x, y & z series.

Let $z - \bar{z} = (ax + by) - (a\bar{x} + b\bar{y})$

$z - \bar{z} = a(x-\bar{x}) + b(y-\bar{y})$

$(z-\bar{z})^2 = a^2(x-\bar{x})^2 + b^2(y-\bar{y})^2 + 2ab(x-\bar{x})(y-\bar{y})$

Summing up to n term

$\Sigma(z-\bar{z})^2 = a^2\Sigma(x-\bar{x})^2 + b^2\Sigma(y-\bar{y})^2 + 2ab\Sigma(x-\bar{x})(y-\bar{y})$

But we have $\sigma_x^2 = \frac{\Sigma(x-\bar{x})^2}{n}$, $\sigma_y^2 = \frac{\Sigma(y-\bar{y})^2}{n}$

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

$\therefore n\sigma_z^2 = a^2n\sigma_x^2 + b^2n\sigma_y^2 + 2ab\,r\,n\sigma_x\sigma_y$

$\div$ by $n \Rightarrow \sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_x\sigma_y r$

$$\therefore \boxed{r = \frac{\sigma_z^2 - (a^2\sigma_x^2 + b^2\sigma_y^2)}{2ab\sigma_x\sigma_y}}$$

SEPTEMBER 2008

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

24

Wednesday

In particular case

$a=1,\ b=1$

$$r = \frac{\sigma_{x+y}^2 - (\sigma_x^2+\sigma_y^2)}{2\sigma_x\sigma_y} \quad \text{—(1)}$$

and $a=1,\ b=-1$

$$r = \frac{\sigma_{x-y}^2 - (\sigma_x^2 - \sigma_y^2)}{-2\sigma_x\sigma_y} = \frac{\sigma_x^2+\sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y} \quad \text{—(2)}$$

Adding (1) & (2)

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

5) If the variables x & y are such that
a) $x+y$ has variance 15 b) $x-y$ has variable 11
c) $2x+y$ has variance 29. Find σ_x, σ_y & the coefficient of correlation between x & y.

⇒ If $z = ax+by$, we have
$\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2abr\sigma_x\sigma_y$ given

a) $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2r\sigma_x\sigma_y$ when $a=1,\ b=1$

b) $\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$ when $a=1,\ b=-1$

c) $\sigma_{2x+y}^2 = 4\sigma_x^2 + \sigma_y^2 + 4r\sigma_x\sigma_y$ when $a=2,\ b=1$.

It is given that $\sigma_{x+y}^2 = 15$, $\sigma_{x-y}^2 = 11$, $\sigma_{2x+y}^2 = 29$

$$\therefore \sigma_x^2 + \sigma_y^2 + 2r\sigma_x\sigma_y = 15 \quad \text{—(1)}$$

$$\sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y = 11 \quad \text{—(2)}$$

$$4\sigma_x^2 + \sigma_y^2 + 4r\sigma_x\sigma_y = 29 \quad \text{—(3)}$$

Solving 1, 2, 3 we get $\sigma_x^2 = 4$, $\sigma_y^2 = 9 \Rightarrow \sigma_x = 2$, & $\sigma_y = 3$

Now eq (2) gives the coefficient of correlation as.

$$r = \frac{\sigma_x^2 + \sigma_y^2 - 11}{2\sigma_x\sigma_y} = \frac{1}{6}$$

25

Thursday

6) Show that if θ is the angle between the lines of regression, then $\tan\theta = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left[\frac{1-r^2}{r}\right]$. Explain the significance when $r=0$ & hence find $\tan^{-1}(3)$, $\sigma_x = \sigma_y$. Find the correlation coefficient.

⇒ WKT if θ is acute, the angle b/w the lines $y = m_1x + c_1$ & $y = m_2x + c_2$ is given by $\tan\theta = \frac{m_2 - m_1}{1+m_1m_2}$

we have the lines of regression

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \text{ —(1)} \quad \& \quad (x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \text{ —(2)}$$

we write (2) as $y - \bar{y} = \frac{\sigma_y}{r\sigma_x}(x-\bar{x})$ —— (3)

Slopes of (1) & (3) are respectively given by
$m_1 = r\frac{\sigma_y}{\sigma_x}$ & $m_2 = \frac{\sigma_y}{r\sigma_x}$

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

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

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

By data $\tan^{-1}(3) = \theta \Rightarrow \tan\theta = 3$ & $\sigma_x = \sigma_y$

$$\therefore 3 = \frac{\sigma_x^2}{2\sigma_x^2}\left(\frac{1-r^2}{r}\right)$$

$$6r = 1 - r^2 \Rightarrow r^2 + 6r - 1 = 0$$

$$r = \frac{-6 \pm \sqrt{40}}{2} = -\frac{6 \pm 2\sqrt{10}}{2} = -3 \pm \sqrt{10}$$

$r = +0.1623$ or $r = -6.1623$

Since $r = \pm 1 \Rightarrow r = 0.1623$.

SEPTEMBER						2008
S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
28	29	30				

Note: • If $r = \pm 1$, $\tan\theta = 0$. $\therefore \theta = 0$ which implies that 2 regression lines coincide & hence variables are perfectly correlated.

• If $r = 0$, $\tan\theta = \infty$ or $\theta = \pi/2$. This implies that lines are $\perp^r$ & variables are uncorrelated.

26
Friday
Wk-39 • 270-096

7) If the co-efficient of correlation between two variables x & y is 0.5 & the acute angle b/w their lines of regression is $\tan^{-1}(3/5)$. Find the ratio of the standard deviation of x & y.

⇒ We have $\tan\theta = \dfrac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left(\dfrac{1-r^2}{r}\right)$ is the formula for acute angle between the lines of regression.

Here, $\theta = \tan^{-1}(3/5) \Rightarrow \tan\theta = \dfrac{3}{5}$ & $r = 0.5 = \dfrac{1}{2}$

$$\therefore \frac{3}{5} = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left[\frac{1-(1/4)}{1/2}\right]$$

$$\frac{3}{5} = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\left(\frac{3}{2}\right)$$

$$\frac{\cancel{3}}{5}\times\frac{2}{\cancel{3}} = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

$$\frac{2}{5} = \frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

$$2(\sigma_x^2+\sigma_y^2) = 5(\sigma_x\sigma_y)$$

$$2(\sigma_x^2+\sigma_y^2) - 5\sigma_x\sigma_y = 0$$

$$2\sigma_x^2 + 2\sigma_y^2 - 5\sigma_x\sigma_y = 0$$

$$2\sigma_x^2 - 4\sigma_x\sigma_y - \sigma_x\sigma_y + 2\sigma_y^2 = 0$$

$$\cancel{(2\sigma_x - \sigma_y)(\sigma_x - 2\sigma_y) = 0}$$

$$2\sigma_x(\sigma_x - 2\sigma_y) - \sigma_y(\sigma_x - 2\sigma_y) = 0$$

$$(2\sigma_x - \sigma_y)(\sigma_x - 2\sigma_y) = 0$$

Hence $2\sigma_x = \sigma_y$ or $\sigma_x = 2\sigma_y$

So that $\dfrac{\sigma_x}{\sigma_y} = \dfrac{1}{2}$ (or) $\dfrac{\sigma_y}{\sigma_x} = 2$

27
Saturday
Wk-39 • 271-095

8) $8x - 10y + 66 = 0$ & $40x - 18y = 214$ are two regression lines. Find the mean of x's, y's & the correlation coefficient. Find σ_y if $\sigma_x = 3$.

⇒ WKT regression line passes through $\bar{x}$ & $\bar{y}$.

$\therefore 8\bar{x} - 10\bar{y} + 66 = 0$ & $40\bar{x} - 18\bar{y} - 214 = 0$

On solving we get $\bar{x} = 13$, $\bar{y} = 17$.

We shall now rewrite the eqn of the regression lines to find the regression co-efficients.

$10y = 8x + 66$ (or) $y = 0.8x + 6.6$ —(1)

$40x = 18y + 214$ (or) $x = 0.45y + 5.35$ —(2)

from (1), Co-eff of $x = 0.8$

from (2), Co-eff of $y = 0.45$

Correlation co-efficient $r = \sqrt{(0.8)(0.45)} = \pm 0.6$

Then $r = 0.6$, since both the regression co-efficient are +ve.

Also, $\sigma_x = 3$ but we have $r\dfrac{\sigma_y}{\sigma_x} = 0.8$

$0.6\,\sigma_y = (0.8)(3) \Rightarrow 0.6\sigma_y = 2.4 \Rightarrow \sigma_y = 4$

28
Sunday
Wk-39 • 272-094

9) Given the eqn of the regression lines $x = 19.13 - 0.87y$, $y = 11.64 - 0.5x$. Compute mean of x's, mean of y's & the co-efficiently of correlation.

⇒ $\bar{x} = 19.13 - 0.87\bar{y}$ & $\bar{y} = 11.64 - 0.5\bar{x}$

$\bar{x} = 15.9$
$\bar{y} = 3.67$

$x = 19.13 - 0.87y$ —(1)

$y = 11.64 - 0.5x$ —(2)

from (1), co-eff of $y = -0.87$

from (2), co-eff of $x = -0.5$

Correlation co-efficient $r = \sqrt{(-0.87)(-0.5)} = \pm 0.435$

SEPTEMBER 2008

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

10) Compute $\bar{x}$, $\bar{y}$ & r from the following eqn of the regression lines $2x+3y+1=0$ & $x+6y-4=0$

$\Rightarrow$ $2\bar{x}+3\bar{y}=-1$ & $\bar{x}+6\bar{y}=4$

On solving, we get $\bar{x}=-2$, $\bar{y}=1$.

We shall now rewrite the eqn of the regression lines to find the regression coefficient..

$2x=-3y\mp 1$ or $x=-1.5y\mp 0.5$ —(1)

$6y=\cancel{4}-x+4$ or $y=-x+4$ —(2)

~~Correlation~~ from (1), Co-eff of $y=-1.5$
from (2), Co-eff of $x=-1$

Correlation co-efficient $r=\sqrt{(-1.5)(-1)}=\pm 1.22$

11) Given

	x-series	y-series
Mean	18	100
S.D	14	20

and $r=0.8$, Write down the eqn of the line of regression & hence find the most probable value of y when $x=70$.

$\Rightarrow$ By data $\bar{x}=18$, $\bar{y}=100$

$\sigma_x=14$, $\sigma_y=20$

We have eqn of the regression line

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

$y-100=\dfrac{(0.8)(20)}{14}(x-18)$ $\quad$ $x-18=\dfrac{(0.8)(14)}{20}(y-100)$

$y=1.14x+79.48$ —(1) $\quad$ $x=0.56y-38$ —(2)

are the lines of regression.

When $x=70$, we obtain the first eqn

$y=1.14(70)+79.48\}=159.28$

Probability.

The study of probability provides a mathematical frame work for such assertion and is essential in every decision making process.

- Event :- Present happening is called event.
- Exhaustive event :- A set of event is said to be exhaustive, if it includes all the possible events. (or) An event consisting of all the various possibilities is called exhaustive events.

  eg: In tossing a coin there are 2 exhaustive cases either a head or a tail & there is no third possibility.

- Mutually ~~exhaustive~~ Exclusive event :- If the happening of one event prevents the simultaneous happening of the others.

  eg: Just as in tossing a coin either head comes up or the tail & but can't happen both at the same time. i.e., These are two mutually exclusive cases.

- Independent Event :- Two or more events are said to be independent if the happening or non-happening of one event does not prevent the happening or non happening of the others.

  eg. When two coins are tossed, the event of getting head is an independent event as both the coins can turn out heads.

Definition of Probability :-

If there are n exhaustive, mutually exclusive & equally likely (equally possible) cases of which m are favourable to an event A, then probability of A is

$$P(A)=\frac{m}{n}=\frac{\text{no. of favourable cases}}{\text{no. of possible cases}}=\frac{O(m)}{O(n)}.$$

eg: $n = \{HH, TT, TH, HT\}$ = order of n (no. of cases) = 4
$m = \{HH, TH, HT\}$ = order of m (no. of possible) = 3
$\therefore P(A) = \frac{O(m)}{O(n)} = \frac{3}{4} = 75\%$

Since m cases are favourable to the event, it follows that $(n-m)$ cases are not favourable to the event. This set of unfavourable event is denoted by A' or $\bar{A}$.
$\therefore$ Probability of the non happening of the event usually denoted by q & is given by.

$q = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(A)$

i.e., $P(A') = 1 - \frac{m}{n} = 1 - P(A)$

$\Rightarrow P(A') = 1 - P(A) \Rightarrow P(A) + P(A') = 1$

i.e., if an event is certain to happen then its probability is unity, while if it is certain not to happen its probability is zero. In otherwords, if $P(A) = 1$; A is called sure event, if $P(A) = 0$, then A is called impossible event.

Set Theory: A well defined collection of objects are called set & it is denoted as A, B, C, ...
A is subset of B is denoted as $A \subset B$.
if x belong to A is denoted as $x \in B$,
Otherwise $x \notin A$.
if $A \subset B$ & $B \subset A \Rightarrow A = B$.

Set Operations:

Union: The set of all elements belongs to either A or B or both is called union of 2 sets denoted as $A \cup B$.
eg: $A = \{2, 3, 4\}$ & $B = \{2, 5, 6\} \Rightarrow A \cup B = \{2, 3, 4, 5, 6\}$
$A = \{a, b, c\}$ & $B = \{a, m, k\} \Rightarrow A \cup B = \{a, b, c, m, k\}$.

Intersection: The set of all elements belongs to both A & B is called intersection of 2 sets developed as $A \cap B$.
eg: $A = \{1, 2, 3, 4, 5\}$ & $B = \{3, 4, 6, 7\} \Rightarrow A \cap B = \{3, 4\}$

Difference: The set of all elements, which belongs to A but does not belong to B is called difference of A & B denoted by $A - B$.
eg: $A = \{1, 2, 3, 4, 5\}$ & $B = \{3, 4, 6, 7\}$
$\Rightarrow A - B = \{1, 2, 5\}$ & $B - A = \{6, 7\}$

Complements: The set of elements of U which does not belong to A is called compliment of A denoted by A'. i.e., $A' = U - A$.
eg: $U = \{1, 2, 3, 4, \ldots\}$, $A = \{2, 3, 5, 6\}$, $B = \{3, 4, 6\}$
$\Rightarrow \bar{A}$ or $A' = \{1, 4, 7, 8 \ldots\}$ & $\bar{B}$ or $B' = \{1, 2, 5, 7, 8, \ldots\}$

Disjoints: Two sets are said to be disjoint if $A \cap B = \phi$.
eg: $A = \{1, 2, 3\}$ & $B = \{4, 5, 6\}$
$\Rightarrow A \cap B = \phi$.

OCTOBER						2008
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Laws of Set Operation

1) $A \cup B = B \cup A$, $A \cap B = B \cap A$ (Commutative law).
2) $A \cup (B \cup C) = (A \cup B) \cup C$ (Associative law)
3). $\left.\begin{array}{l} A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \end{array}\right\}$ Distributive law.
4) $\left.\begin{array}{l} \overline{A \cup B} = \bar{A} \cap \bar{B} \\ \overline{A \cap B} = \bar{A} \cup \bar{B} \end{array}\right\}$ De Morgan's law.

5) $A - B = A \cap \bar{B}$
$B - A = B \cap \bar{A}$
6) $\overline{(\bar{A})} = A$

Combination : Combination is the selection of r term out of n it is denoted by $^nC_r = \dfrac{n!}{(n-r)!\,r!}$

Random Experiments : The experiments when performed repeatedly giving different results (outcomes) are called random experiments (trials).

Sample Space : A set S consisting of all possible outcomes of a random experiment is called a sample space which corresponds to the universal set.
If a sample space has finite no. of elements, then it is called a finite sample space ; otherwise it is called an infinite sample space.
eg : $S = \{1, 2, 3, 4, 5, 6\}$
Let $E_1 = \{1, 3, 5\}$ & $E_2 = \{2, 4, 6\}$. Then E_1 & E_2 are subsets of S.

Event : An event E is a subset of the sample space S. A particular outcome that is an element S is called a sample.
eg : Suppose a coin is tossed twice & E is the event of getting atleast one head, then we have
$S = \{HH, HT, TH, TT\}$, $E = \{HH, HT, TH\}$.

1. State the axioms of probability. Prove that
a) $P(\bar{A}) = 1 - P(A)$

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ where A and B are any 2 events.

Axioms of Probability (Axiomatic def of probability)
If S is the sample space & E is the set of all events then to each event A in E. We associate a unique real no. $P = P(A)$ known as the probability of the event A, if the following axioms are satisfied. These are known as the axioms of probability.

1) $P(S) = 1$
2) For every event A in E $0 \le P(A) \le 1$
3) If $A_1, A_2, A_3, \ldots\ldots A_n$ are mutually exclusive events of E, then
$P(A_1 \cup A_2 \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

OCTOBER 2008

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Addition Rule (or) Addition law of probability (or) Theorem of total probability

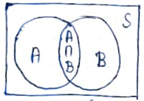
"If A & B are any two arbitrary events, prove that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$"

(or)

If A & B are any two events of S which are not mutually exclusive, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof :- If A & B are two events, then $A \cup B$ is an event consisting of outcomes that are in A or B or both, & $A \cap B$ is an event consisting of outcomes that are common to A & B, so that

$$O(A \cup B) = O(A) + O(B) - O(A \cap B)$$



$$\therefore P(A \cup B) = \frac{O(A \cup B)}{O(S)}$$

$$= \frac{O(A) + O(B) - O(A \cap B)}{O(S)} \qquad \left[\because P(A) = \frac{O(A)}{O(S)} \ldots\ldots\right]$$

$$= \frac{O(A)}{O(S)} + \frac{O(B)}{O(S)} - \frac{O(A \cap B)}{O(S)}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A & B are independent events, then $P(B/A) = P(B)$.

$\therefore P(A \cap B) = P(A) \cdot P(B)$.

Mutually exclusive events

$P(A \cap B) = 0$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= P(A) + P(B)$

Complementary events : $P(A) + P(B) = 1$

Exhaustive events : $P(A \cup B) = 1$

Independent events : $P(A \cap B) = P(A) \cdot P(B)$

Note : $\overline{A \cup B} = \bar{A} \cap \bar{B}$

1.) A student A can solve 75 % of the problems given in the book & a student B can solve 70 %. What is the probability that either A or B can solve a problem choosen at random?

⇒ The student A can solve 75 % of problems so that the probability of his not solving a problem is.

$P(\bar{A}) = 1 - \frac{75}{100} = \frac{1}{4}$ $\quad$ ||ly $\quad P(\bar{B}) = 1 - \frac{70}{100} = \frac{3}{10}$

As the event A & B are mutually independent events, The probability of both not solving a problem is

$$P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B}) = \frac{3}{40}$$

∴ Probability that either A or B solve a problem is

$$P(A \cup B) = 1 - P(\overline{A \cup B}) = 1 - P(\bar{A} \cap \bar{B})$$

$$= 1 - \frac{3}{40}$$

$$= \frac{37}{40}$$

$\frac{1}{4} \cdot \frac{7}{10} = \frac{7}{40}$

$\frac{1}{4} + \frac{7}{10} - \frac{7}{40} = 10 + 28$.

OCTOBER 2008

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

2) A problem is given to three students A, B, C whose chances of solving it are $\frac{1}{2}, \frac{1}{4}, \frac{1}{5}$ respectively. Find the probability that the problem is solved.

⇒ The probability that A fails to solve the problem is

$$P(\bar{A}) = 1 - \frac{1}{2} = \frac{1}{2}$$

III^ly $P(\bar{B}) = 1 - \frac{1}{4} = 3/4$ & $P(\bar{C}) = 1 - 1/5 = 4/5$

Since the events are independent, the probability that all the three students fail to solve the problem is

$$P(\bar{A}).P(\bar{B}).P(\bar{C}) = \frac{1}{2} \times \frac{3}{4} \times \frac{4}{5} = \frac{3}{10}$$

∴ The probability that the problem is solved is

$$1 - \frac{3}{10} = \frac{7}{10}$$

Bayer's Theorem :

Let S be a finite outcome set & $A_1, A_2, \ldots A_n$ be events such that i) every two of these events are mutually exclusive (disjoint) & ii) The union of all these events is equal to S. Also, let E be any event.

Then for any $i = 1, 2, \ldots n$, The multiplication theorem for conditional probability gives

$$P(A_i/E) = \frac{P(A_i \cap E)}{P(E)} = \frac{P(E \cap A_i)}{P(E)}$$

$$= \frac{P(E/A_i)P(A_i)}{P(E)} \quad \text{—— (1)}$$

Consider, $E = E \cap S$

$$E = E \cap (A_1 \cup A_2 \cup \ldots \cup A_n)$$

$$E = (E \cap A_1) \cup (E \cap A_2) \cup \ldots \cup (E \cap A_n)$$

Since $E \cap A_1, E \cap A_2, \ldots, E \cap A_n$ are mutually disjoints, the addition theorem gives

$$P(E) = P(E \cap A_1) + P(E \cap A_2) + \ldots + P(E \cap A_n).$$

Using the multiplication theorem, this becomes

$$P(E) = P(E/A_1)P(A_1) + P(E/A_2)P(A_2) + \ldots + P(E/A_n)P(A_n) \quad \text{—— (2)}$$

Substituting this in (1), we obtain

$$P(A_i/E) = \frac{P(E/A_i)P(A_i)}{P(E/A_1)P(A_1) + P(E/A_2)P(A_2) + \ldots + P(E/A_n)P(A_n)}.$$

$$P(A_i/E) = \frac{P(E/A_i)P(A_i)}{\sum P(E/A_i)P(A_i)} \quad \text{—— (3)}$$

This is known as the Bayer's Theorem (formula).

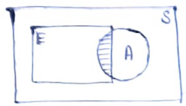
Conditional Probability : The happening of one event depends on the happening of the other.

Def : Let A & B be the two events. Probability of the happening of the events B. When the event A has already happened is called the conditional probability denoted by $P(B/A)$.

i.e., to say that $P(B/A)$ is the probability of B gives A

$$P(B/A) = \frac{\text{Probability of the occurance of both B \& A}}{\text{Probability of the occurance of the given event A}}$$

i.e., $P(B/A) = \frac{P(A \cap B)}{P(A)}$ —— (1)



also $P(A/B) = \frac{P(A \cap B)}{P(B)}$ —— (2)

OCTOBER 2008

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Multiplication Rule (or) Multiplication law of probability (or) Theorem of compound probability

The probability of simultaneous occurance of two events A & B are given by and we have from (1)

$$P(A \cap B) = P(A) \cdot P(B/A) \quad \text{--- (3)}$$

where $P(A) > 0$. This is called the multiplication rule of probability.

1) The probability of conducting an examination on time is 0.95 if there is no strike by students & 0.25 if there is a strike. If the probability that there will be a strike is 0.65, find the probability of holding the examination on time.

⇒ Let A & B respectively denote the events that there is strike by students & there is no strike. Let C denote the event that the examination is held on time. Then we are required to find $P(C)$.

Here, $P(A) = 0.65$, $P(B) = 1 - P(A) = 0.35$

$P\left(\frac{C}{A}\right) = 0.25$ & $P\left(\frac{C}{B}\right) = 0.95$

$\therefore P(C) = P\left(\frac{C}{A}\right) P(A) + P\left(\frac{C}{B}\right) P(B)$

$P(C) = (0.25)(0.65) + (0.95)(0.35)$

$P(C) = 0.495$

2) Three machines A, B, C produce 50 %, 30 %, 20 % of the items in a factory. The percentage of defective outputs of these machines are 3, 4, 5 respectively. If an items is selected at random, what is the probability that it is defective? If a selected item are defective, what is the probability that it is from machine A?

⇒ Let $P(A)$, $P(B)$ & $P(C)$ denote the probability of choosing an item produced by machines A, B, C respectively. Also, let $P(X)$ denote the probability of choosing a defective item from the whole output & $P(X/A)$, $P(X/B)$, $P(X/C)$ denote the probabilities of choosing a defective item from the output of A, B, C respectively. Then from what is given, we have

$P(A) = \frac{50}{100} = 0.5$, $P(B) = \frac{30}{100} = 0.3$, $P(C) = \frac{20}{100} = 0.2$

$P(X/A) = \frac{3}{100} = 0.03$, $P\left(\frac{X}{B}\right) = \frac{4}{100} = 0.04$, $P\left(\frac{X}{C}\right) = \frac{5}{100} = 0.05$

$P(X) = P(A)\,P(X/A) + P(B)\,P(X/B) + P(C)\,P(X/C)$

$P(X) = (0.5)(0.03) + (0.3)(0.04) + (0.2)(0.05)$

$P(X) = 0.037$

Thus, the probability of selecting a defective item is 0.037. Next, we have to find the probability that the item selected is from A, given that it is defective. This probability is

$$P(A/X) = \frac{P(X/A)\,P(A)}{P(X)} = \frac{(0.03)(0.5)}{(0.037)} = 0.4054$$

$$P(A/X) = \frac{P(X/A)\,P(A)}{P(X)} = \frac{(0.03)(0.5)}{(0.037)} = 0.4054$$

By Bayer's theorem.

OCTOBER						2008
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

A variable whose value is the numerical outcome of an experiment or random phenomenon.

Random Variable

In a random experiment, if a real variable is associated with every outcome then it is called a random variable or stochastic variable or variate or chance variable.
eg i) While tossing a coin . 1 → head (outcome)
0 → tail

$S = \{H, T\}$, here x is the random variable.
$x(H) = 1$, $x(T) = 0$, Range of $x = \{0, 1\}$.

ii) If we consider a random experiment of tossing a coin twice, Then the outcomes are HH, HT, TH, TT. Let x_i be the random experiment of getting no. of heads, Then x_i can be zero heads, one heads, 2 heads. This x_i is called random variable. If we consider their corresponding probability are
$P(x=0)$ = Probability of 0 head = 1/4
$P(x=1)$ = Probability of 1 head = 2/4
$P(x=2)$ = Probability of 2 head = 1/4

$$P(x=0) + P(x=1) + P(x=2) = \frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1.$$

It is to note that sum of the probability for random variable $x = x_i$ is always one. i.e., $\sum_{i=0}^{n} P(x=x_i) = 1$.

If we take random variable $x = x_i$ & their corresponding probability $P(x) = x_i$. In the distribution table then it is called random distribution. It is shown as below.

$x = x_i$	0	1	2 ...... n
$P(x = x_i)$	$P(x=0)$	$P(x=1)$	$P(x=2)$

$f : X \to \mathbb{R}$
st $\{ w \in X, \ X(w) = x\}$ $\forall x \in \mathbb{R}$

for any random variable their exists three statistical quantities namely.
1) Arithmetic mean (μ)
2) Variance (V)
3) Standard deviation (σ)

1) Arithmetic mean : If the random variable $x = x_i$ & probability function $P(x) = x_i$. Then the sum of product of random variable with their corresponding probability is called A.M.

$$\mu = \sum_{i=0}^{n} x_i P(x_i)$$

2) Variance : From statistics, it is defined as

$$V = \sum_{i=0}^{n} (x_i - \mu)^2 P(x_i)$$

It is also in the form $V = \sum_{i=0}^{n} x_i^2 P(x_i) - [\sum x_i P(x_i)]^2$

(or) $V = \sum_{i=0}^{n} x_i^2 P(x_i) - \mu^2$

3) Standard deviation : It is a positive square root of Variance & denoted by σ.

$$\sigma = \sqrt{\sum_{i=0}^{n} x_i^2 P(x_i) - \mu^2}$$

1) When a coin is tossed thrice then find mean, variance & SD of random variable of no. of heads.

⇒ When a coin is tossed thrice, then

$S = \{HHH, HTH, HHT, THH, HTT, THT, TTH, TTT\}$

Range of $x = \{0, 1, 2, 3\}$

$P(x=0)$ = Probability of 0 heads = $1/8$
$P(x=1)$ = " " 1 heads = $3/8$
$P(x=2)$ = " " 2 heads = $3/8$
$P(x=3)$ = " " 3 heads = $1/8$

$X = x_i$	0	1	2	3
$P(X=x_i)$	$1/8$	$3/8$	$3/8$	$1/8$

Mean $\mu = \sum_{i=0}^{3} x_i P(x_i) = x_0 P(x_0) + x_1 P(x_1) + x_2 P(x_2) + x_3 P(x_3)$

$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8}$

$= \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8}$

$\mu = 3/2$

Variance $V = \sum_{i=0}^{3} x_i^2 P(x_i) - \mu^2 = [x_0^2 P(x_0) + x_1^2 P(x_1) + x_2^2 P(x_2) + x_3^2 P(x_3)] - \mu^2$

$= (0)^2 \frac{1}{8} + (1)^2 \frac{3}{8} + (2)^2 \frac{3}{8} + (3)^2 \frac{1}{8} - \frac{9}{4}$

$= \frac{3}{8} + \frac{12}{8} + \frac{9}{8} - \frac{9}{4} = \frac{24}{8} - \frac{9}{4} = \frac{6}{8}$

$V = \frac{3}{4}$

Standard Deviation $= \sqrt{V} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$

Classification of Random Variables:

Discrete Random Variable: If a random variable takes finite number of values, Then it is called a discrete random variable.

Continuous random variable: If a random variable takes non countable or infinite number of values, Then it is called Continuous random variable.

Discrete probability distribution: If for each value x_i of a discrete random variable X. We assign a real number $P(x_i)$ such that i) $P(x_i) \geq 0$ ii) $\sum_i P(x_i) = 1$. Then the function $P(x)$ is called a probability function.

The set of values $[x_i, P(x_i)]$ is called a discrete (finite) probability distribution of the discrete random variable x. The function $P(x)$ is called the probability density function (p.d.f) or probability mean function (p.m.f). The mean & variance of the discrete probability is defined as follows.

Mean $\mu = \sum_i x_i P(x_i)$

Variance $V = \sum_i (x_i - \mu)^2 P(x_i)$ (or) $\sum_i x_i^2 P(x_i) - \mu^2$ (or)

$= \sum_i x_i^2 P(x_i) - [\sum_i x_i P(x_i)]^2$

Standard deviation $\sigma = \sqrt{V}$

The distribution function $f(x)$ defined by $f(x) = P(X \leq x) = \sum_{i=1}^{x} P(x_i) = t$, x being an integer is called the cummulative distribution function (cdf) or distribution function.

OCTOBER						2008
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

1) A coin is tossed twice. A random variable X represent the no. of heads turning up. Find the discrete probability distribution for X. Also find its mean, variance & standard deviation.

$\Rightarrow$ $S = \{HH, HT, TH, TT\}$, the associated of the elements of S to the random variable X are respectively 2, 1, 1, 0.

Now $P(HH) = \frac{1}{4}$, $P(HT) = \frac{1}{4}$, $P(TH) = \frac{1}{4}$, $P(TT) = \frac{1}{4}$

$P(X=0$, i.e., no head$) = P(TT) = 1/4$

$P(X=1$, i.e one head$) = P(HT \cup TH) = P(HT) + P(TH)$
$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$P(X=2$, i.e two head$) = P(HH) = \frac{1}{4}$

The discrete probability distribution for X is as follows

$X = x_i$	0	1	2
$P(x_i)$	$1/4$	$1/2$	$1/4$

We observe $P(x_i) > 0$ & $\Sigma P(x_i) = 1$

Mean $\mu = \Sigma x_i P(x_i) = 0 + \frac{1}{2} + \frac{1}{2} = 1$

Variance $V = \Sigma (x_i - \mu)^2 P(x_i)$
$= (0-1)^2 \frac{1}{4} + (1-1)^2 \frac{1}{2} + (2-1)^2 \frac{1}{4} = \frac{1}{4} + \frac{1}{4} = \frac{2}{4}$
$= \frac{1}{2}$

Standard Deviation $= \sqrt{V} = \sqrt{\frac{1}{2}}$

2) The Pdf of a variate x is given by the following data.

x	0	1	2	3	4	5	6
$P(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

For what value of k, this represents a valid probability distribution? Also find $P(x \geq 5)$ & $P(3 < x \leq 6)$.

$\Rightarrow$ The probability distribution valid if $P(x_i) \geq 0$ & $\Sigma P(x_i) = 1$.
Hence we must have $k \geq 0$ &
$k + 3k + 5k + 7k + 9k + 11k + 13k = 1$
$49k = 1$
$k = \frac{1}{49}$

Also, $P(x \geq 5) = P(5) + P(6) = 11k + 13k = 24k = \frac{24}{29}$.

$P(3 < x \leq 6) = P(4) + P(5) + P(6) = 33k = \frac{33}{49}$

3) A die is tossed thrice, a success is getting 1 or 6 a toss. Find the mean & variance of the number of success.

${}^nC_x = \frac{n!}{(n-x)!\,x!}$

$\Rightarrow$ Probability of a success $= \frac{2}{6} = \frac{1}{3}$
Probability of a failures $= 1 - \frac{1}{3} = \frac{2}{3}$

$\therefore$ Probability of no success = Probability of all failures
$= {}^3C_0 \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{8}{27}$

Prob of 1 success & 2 failures $= {}^3C_1 \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$

Prob of 2 success & 1 failure $= {}^3C_2 \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$

Prob of 3 success $= {}^3C_3 \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27}$

OCTOBER 2008

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

x_i	0	1	2	3
$P(x_i)$	$8/27$	$4/9$	$2/9$	$1/27$

$\therefore$ Mean $\mu = \sum x_i P(x_i)$

$= 0 + \frac{4}{9} + \frac{4}{9} + \frac{3}{27} = 1$

Variance $V = \frac{2}{3}$ $\sum x_i^2 P(x_i) - \mu^2$

$= 0 + \frac{16}{81} + \frac{16}{81} + \frac{9}{729} - 1 = \frac{2}{3}$

S.D $\sigma = \sqrt{V} = \sqrt{\frac{2}{3}}$

Joint Probability

The distribution associated with two random variables is referred as joint distribution.

If X & Y are two discrete random variable, we define the joint probability function of X & Y by suppose $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$ then.

$P(X = x_i, Y = y_j) = f(x_i, y_j) = J_{ij}$

where $f(x, y)$ satisfy the conditions 1) $f(x, y) \geq 0$ &

2) $\sum_x \sum_y f(x, y) = 1$ i.e, $\sum_x \sum_y J_{ij} = 1$ [The second condition mean that the sum over all the values of x & y is equal to one].

The values are presented in the form of two way table called the joint probability table.

X \ Y	y_1	y_2	....	y_n	Sum
x_1	J_{11}	J_{12}	....	J_{1n}	$f(x_1)$
x_2	J_{21}	J_{22}	....	J_{2n}	$f(x_2)$
|	|	|	|	|	|
x_m	J_{m1}	J_{m2}	....	J_{mn}	$f(x_m)$
Sum	$g(y_1)$	$g(y_2)$	...	$g(y_n)$	1

x - column
y - row

Here $f(x_1) = J_{11} + J_{12} + \dots + J_{1n}$

$f(x_2) = J_{21} + J_{22} + \dots + J_{2n}$ etc.,

$f(x_m) = J_{m1} + J_{m2} + \dots + J_{mn}$

IIly $g(y_1) = J_{11} + J_{12} + \dots + J_{m1}$

$g(y_2) = J_{12} + J_{22} + \dots + J_{m2}$ etc.,

$g(y_n) = J_{1n} + J_{2n} + \dots + J_{mn}$.

$\{f(x_1), f(x_2), \dots f(x_m)\}$ & $\{g(y_1), g(y_2) \dots g(y_n)\}$ are called marginal probability distribution of X & Y (or) individual distribution.

Note : The distribution is obtained by adding them all the respective row entries & also respective column entries, i.e.,

$f(x_1) + f(x_2) + \dots + f(x_m) = 1$ &

$g(y_1) + g(y_2) + \dots + g(y_n) = 1$.

Expectation, Variance & Co-variance :-

If x is a discrete random variable taking values $x_1, x_2, x_3, \ldots x_n$ having probability function $f(x)$, then expectation of x, denoted by $E(X)$ or μ_x is defined by the relation.

$$\mu_x = E(X) = \sum_{i=1}^{n} x_i f(x_i) \text{ or } \sum x f(x) \text{ \&}$$

$$E(x^2) = \sum_i x_i^2 f(x_i)$$

Variance of X denoted by $V(X)$ is defined as

$$V(X) = \sigma_X^2 = E(X^2) - [E(X)]^2$$

$|||^{ly}$ $V(Y) = \sigma_Y^2 = E(Y^2) - [E(Y)]^2$

$\sigma_X = \sqrt{V(X)}$ is called the standard deviation (S.D) of x.

If X and Y are two discrete random variables having the joint probability function $f(x, y)$, then the expectations of X & Y are defined as follows

$$\mu_x = E(X) = \sum_x \sum_y x f(x, y) = \sum_i x_i f(x_i)$$

$$\mu_y = E(Y) = \sum_x \sum_y y f(x, y) = \sum_j y_j g(y_j)$$

If $z = \phi(x, y)$, then the expectation of z is

$$E(z) = \sum_x \sum_y \phi(x_i, y_j) J_{ij}.$$

Covariance: If X & Y are random variables having mean μ_x & μ_y respectively, then the co-variance of X & Y denoted by $\text{cov}(X, Y)$ are defined by the relation $\text{cov}(X, Y)$

$\text{Co}(X, Y) = E(XY) - E(X)E(Y)$

Further, the correlation of X & Y denoted by $\rho(X, Y)$ & is defined by the relation

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad \text{where } \sigma_X = \sqrt{V(X)}, \ \sigma_Y = \sqrt{V(Y)}$$

Independent random variable

The discrete random variable X & Y are said to be independent random variable if $P(X=x, Y=y) = P(X=x).P(Y=y)$.

This is equivalent to $f(x_i) g(y_j) = J_{ij}$ i.e., to say that X & Y are independent if each entry J_{ij} in the table is equal to the product of its marginal entries; otherwise X & Y are said to be dependent.

Note : a) $E(XY) = E(X).E(Y)$

b) $\text{Cov}(X, Y) = 0$ & hence $\rho(X, Y) = 0$

c) $\text{Cov}(X, X) = E[(X - \mu_x)^2]$
$= V(X) = \sigma_X^2$

d) $V(X) = E(X^2 - 2X\mu_x + \mu_x^2)$
$= E(X^2) - 2E(X)\mu_x + \mu_x^2 E(1)$
$= E(X^2) - 2E(X)E(X) + [E(X)]^2 \cdot 1$

$\sigma_X^2 = E(X^2) - [E(X)]^2$ (or) $\sigma_X^2 = E(X^2) - \mu_X^2$

i) The joint distribution of two random variables X & Y is as follows

X\Y	-4	2	7
1	1/8	1/4	1/8
5	1/4	1/8	1/8

Determine i) The marginal distribution of X & Y
ii) $E(X)$, $E(Y)$ & $E(XY)$, $\text{Cov}(X, Y)$, $\rho(X, Y)$
iii) Are X & Y independent random variable.

$\Rightarrow$ i) The marginal distribution of X & Y are [The distribution is obtained by adding all the respective row entries & also respective column entries].

x_i	1	5
$f(x_i)$	½	½

Distribution of x

y_i	-4	2	7
$g(y_i)$	3/8	3/8	¼

Distribution of Y

ii) $E(x) = \Sigma x_i f(x_i) = (1)(\frac{1}{2}) + 5(\frac{1}{2}) = \frac{6}{2} = 3$

$E(y) = \Sigma y_j\, g(y_j) = (-4)(\frac{3}{8}) + (2)(\frac{3}{8}) + (7)(\frac{1}{4}) = 1$

Thus $\mu_x = E(x) = 3$ & $\mu_y = E(y) = 1$

iii) If X & Y are independent random variables we must have $f(x_i)\, g(y_j) = J_{ij}$

$f(x_1)\, g(y_1) = \frac{1}{2} \times \frac{3}{8} = \frac{3}{16} \neq \frac{1}{8}$ $\quad \therefore f(x_1)\, g(y_1) \neq J_{11}$

III^ly for other also the condition is not satisfied. Hence X & Y are dependent random variables.

2) The joint probability distribution of two random variable X & Y is given below

X \ Y	-3	2	4
1	0.1	0.2	0.2
3	0.3	0.1	0.1

Find (i) Marginal distribution of X & Y
(ii) Covariance of X & Y
(iii) Correlation co-efficient ρ of (X, Y).

⇒ (i) The marginal distribution of X & Y are

Distribution of X

x_i	1	3
$f(x_i)$	0.5	0.5

Distribution of Y

y_j	-3	2	4
$g(y_j)$	0.4	0.3	0.3

ii) $\text{cov}(XY) = E(xy) - E(x)E(y)$

$E(x) = \sum_i x_i f(x_i) = 1(0.5) + 3(0.5) = 2$

$E(y) = \sum_i y_j\, g(y_j) = -3(0.4) + 2(0.3) + 4(0.3) = 0.6$

$E(xy) = \Sigma x_i y_j J_{ij}$

$= 1 \times -3 \times 0.1 + 1 \times 2 \times 0.2 + 1 \times 4 \times 0.2 + 3 \times 0.3 \times -3 + 3 \times 0.1 \times 2 + 3 \times 0.1 \times 4$

$E(xy) = -3 + 3 = 0$

$\text{cov}(xy) = 0 - 2(0.6) = -1.2$

3) The joint probability distribution of two discrete random variable X & Y is given by the table. Determine the marginal distribution of X & Y. Also find whether X & Y are independent.

X \ Y	1	3	6
1	1/9	1/6	1/18
3	1/6	1/4	1/12
6	1/18	1/12	1/36

⇒ Distribution of X

x_i	1	3	6
$f(x_i)$	3/4	½	5/36

Distribution of Y

y_j	1	3	6
$g(y_j)$	3/4	½	5/36

If X & Y are independent variable we must have $f(x_i)\, g(y_j) = J_{ij}$

$f(x_1)\, g(y_1) = (\frac{3}{4})(\frac{3}{4}) = \frac{9}{16} \neq \frac{1}{9}$ $\quad \therefore f(x_1)\, g(y_1) \neq J_{11}$

X & Y are dependent variable.

OCTOBER 2008

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

27
Monday

2nd problem continuity

$$= 1\times -3\times 0.1 + 1\times 2\times 0.2 + 1\times 4\times 0.2 + 3\times -3\times 0.3 + 3\times 2\times 0.1 + 3\times 4\times 0.3$$

$$= -0.3 + 0.4 + 0.8 - 2.7 + 0.6 + 3.6$$

$$E(xy) = 2.4$$

$$Cov(x,y) = E(xy) - E(x)\cdot E(y)$$
$$= 2.4 - 3(1) = -0.6$$

$$\rho(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{-0.6}{\sigma_x \sigma_y}$$

$$\sigma_x = \sqrt{V(x)} = \sqrt{E(x^2) - [E(x)]^2} = \sqrt{13 - 3^2} = \sqrt{4} = 2$$

$$E(x^2) = \sum x_i^2 f(x_i) = x_1^2 f(x_1) + x_2^2 f(x_2)$$
$$= 1\times \tfrac{1}{2} + (5)^2 \tfrac{1}{2} = 13$$

$$E(y^2) = \sum y_j^2 g(x_j)$$
$$= y_1^2 g(x_1) + y_2^2 g(x_2) + y_3^2 g(x_3)$$
$$= (-4)^2 \frac{3}{8} + 2^2 \left(\frac{3}{8}\right) + 7^2 \left(\frac{1}{4}\right) = 6 + \frac{3}{2} + \frac{49}{4} = 19.75$$

$$\sigma_y = \sqrt{V(y)} = \sqrt{E(y^2) - [E(y)]^2} = \sqrt{19.75 - 1}$$

$$\sigma_y = 4.33$$

$$\rho(x,y) = \frac{-0.6}{2\times 4.33} = \frac{-0.6}{8.66} = \underline{-0.0692}$$

28
Tuesday

Unit - 2

Z - transform

The Z-transform is an important tool in the analysis of signals & linear time-invarient (LTI) system. It has many properties which are akin to those of laplace transforms. Infact, it plays the same role in the analysis of discrete-time signals ~~same role in the~~ & LTI system as laplace transform does in the analysis of continuous-time signals & LTI systems.

However, the main difference between the two is that Z-transform operates on sequence of discrete integer-valued arguments $n = 0, \pm 1, \pm 2, \ldots$ where as laplace transform operates on functions of continuous arguments. Also for every operational rule, and applications pertaining to laplace transforms, there is a corresponding operational rule & application for Z-transforms.

Definition : Let $\{U_n\}$ be a sequence defined for discrete values of $n = 0, 1, 2, \ldots$, then the Z-transform of U_n is defined as $Z(U_n) = U(z) = \sum_{n=-\infty}^{\infty} U_n Z^{-n}$

The above function is also known as two-sided Z-transform of a discrete-time sequence $\{U_n\}$. If $Z(U_n)$ converges for all values of z, then it is refered as bilateral Z-transform.

The one sided or unilateral Z-transform is defined as

$$Z(U_n) = U(z) = \sum_{n=0}^{\infty} U_n z^{-n}$$

OCTOBER 2008

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

2^{nd} problem continuity

$= 1\times -3\times 0.1 + 1\times 2\times 0.2 + 1\times 4\times 0.2 + 3\times -3\times 0.3 + 3\times 2\times 0.1 + 3\times 4\times 0.3$

$= -0.3 + 0.4 + 0.8 - 2.7 + 0.6 + 3.6$

$E(xy) = 2.4$

$Cov(X, Y) = E(xy) - E(X).E(Y)$

$= 2.4 - 3(1) = -0.6$

$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y} = \frac{-0.6}{\sigma_x \sigma_y}$

$\sigma_x = \sqrt{V(X)} = \sqrt{E(X^2) - [E(X)]^2} = \sqrt{13 - 3^2} = \sqrt{4} = 2$

$E(X^2) = \sum x_i^2 f(x_i) = x_1^2 f(x_1) + x_2^2 f(x_2)$

$= 1\times \frac{1}{2} + (5)^2 \frac{1}{2} = 13$

$E(Y^2) = \sum y_j^2 g(x_j)$

$= y_1^2 g(x_1) + y_2^2 g(x_2) + y_3^2 g(x_3)$

$= (-4)^2 \frac{3}{8} + 2^2 \left(\frac{3}{8}\right) + 7^2 \left(\frac{1}{4}\right) = 6 + \frac{3}{2} + \frac{49}{4} = 19.75$

$\sigma_y = \sqrt{V(Y)} = \sqrt{E(Y^2) - [E(Y)]^2} = \sqrt{19.75 - 1}$

$\sigma_y = 4.33$

$\rho(X, Y) = \frac{-0.6}{2\times 4.33} = \frac{-0.6}{8.66} = \underline{-0.0692}$