12/4/22

Unit-2

Correlation & Regression

&

Discrete probability Distributions

Correlation: correlation refers the relationship of two or more variables we know that their exist relationship between heights and weights of students in a class, wage and prize index The study of relation is called correlation. It measures the closeness of the relationship between the variables

Def: Correlation is a statistical analysis which measures and analyzes how the two variables related with respect to each other. The correlation express the relationship are independence of two set of variables upon each other.

13/4/22

Coefficient of correlation

Karl Spearson, a British biometrician suggested a method for measuring the magnitude of linear relationship between 2 variables. This is known as Pearson's coefficient of correlation and it is represented by 'r' and defined as

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$ where $X = \Sigma x_i - \bar{x}$

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

$\bar{x}$ = mean of series x

$\bar{y}$ = mean of series y

Properties of correlation coefficient

1) Correlation coefficient lies between -1 to 1

i.e., $-1 \leq r \leq 1$

2) If $r = 1$, there is a perfect positive correlation

If $r = -1$, there is perfect negative correlation

If $r = 0$, there is no correlation

3) The coefficient of correlation is independent of the change of origin and scale of measurement

1. Find if there is any significant correlation between heights and weights given below.

Heights in inches (x)	57	59	62	63	64	65	55	58	57
weight in lbs (y)	113	117	126	126	130	129	111	116	112

We know that coefficient of correlation

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$ where $X = \Sigma x_i - \bar{x}$

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

Here $\bar{x} = \frac{57+59+62+63+64+65+55+58+57}{9} = \frac{540}{9}$

$= 60$

$$\bar{y} = \frac{113+117+126+126+130+129+111+116+112}{9} = 120$$

x	y	$X = x_i - \bar{x}$	$Y = y_i - \bar{y}$	XY	X^2	Y^2
57	113	-3	-7	21	9	49
59	117	-1	-3	3	1	9
62	126	2	6	12	4	36
63	126	3	6	18	9	36
64	130	4	10	40	16	100
65	129	5	9	45	25	81
55	111	-5	-9	45	25	81
58	116	-2	-4	8	4	16
57	112	-3	-8	24	9	64
				216	102	472

$$\therefore r = \frac{216}{\sqrt{(102)(472)}} = \frac{216}{\sqrt{48144}} = \frac{216}{219.41} = 0.98$$

i.e. there is more positive correlation b/w heights and weights

2. find Karl Pearson. coefficient of correlation from the following data

wages(x)	100	101	102	102	100	99	97	98	96	95
cost of living (y)	98	99	99	97	95	92	95	94	90	91

we know that coefficient of correlation,

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$ where $X = \Sigma x_i - \bar{x}$

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

Here $\bar{x} = \frac{100+101+102+102+100+99+97+98+96+95}{10}$

$= \frac{990}{10} = 99$

$\bar{y} = \frac{98+99+99+97+95+92+95+94+90+91}{10}$

$= \frac{950}{10} = 95$

x	y	$X = x_i - \bar{x}$	$Y = y_i - \bar{y}$	XY	X^2	Y^2
100	98	1	3	3	1	9
101	99	2	4	8	4	16
102	99	3	4	12	9	16
102	97	3	2	6	9	4
100	95	1	0	0	1	0
99	92	0	-3	0	0	9
97	95	-2	0	0	4	0
98	94	-1	-1	1	1	1
96	90	-3	-5	15	9	25
95	91	-4	-4	16	16	16
				61	54	96

$\therefore r = \frac{61}{\sqrt{(54)(96)}} = \frac{61}{\sqrt{5184}} = \frac{61}{72} = 0.84$

i.e., there is more positive correlation b/w wages and cost of living

Q. Calculate Karl Pearson coefficient of following data.

x	38	45	46	38	35	38	46	32	36	38
y	28	34	38	34	36	26	28	29	25	36

we know that

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

here $\bar{x} = \frac{38+45+46+38+35+38+46+32+36+38}{10}$

$= \frac{392}{10} = 39.2$

$\bar{y} = \frac{28+34+38+34+36+26+28+29+25+36}{10}$

$= \frac{314}{10} = 31.4$

x	y	$x = x_1 - \bar{x}$	$y = y_1 - \bar{y}$	xy	x^2	y^2
38	28	-1.2	-3.4	4.08	1.44	11.56
45	34	5.8	2.6	15.08	33.64	6.76
46	38	6.8	6.6	44.88	46.24	43.56
38	34	-1.2	2.6	-3.12	1.44	6.76
35	36	-4.2	4.6	-19.32	17.64	21.16
38	26	-1.2	-5.4	6.48	1.44	29.16
46	28	6.8	-3.4	-23.12	46.24	11.56
32	29	-7.2	-2.4	17.28	51.84	5.76
36	25	-3.2	-6.4	20.48	10.24	40.96
38	36	-1.2	4.6	5.52	1.44	21.16
				68.26	211.6	198.4

$$r = \frac{68.26}{\sqrt{(211.6)(198.4)}}$$

$$= \frac{68.26}{\sqrt{41981.44}} \Rightarrow \frac{68.26}{204.8937} = 0.333$$

14/2/2

Rank correlation coefficient:

A British psychologist charles Adward spearman introduced a method to find the ~~rank~~ of given observations using ranks correlation coefficients of given observations using rank. This is known as Rank of correlation coefficient.

It is obtained by $\rho' = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$

where ρ = Rank correlation coefficient

D^2 = Sum of squares of the ~~ranks~~ difference of given ~~observations~~ two ranks

n = no. of given observations

Properties of Rank correlation coefficient

→ The value of Rank correlation Coefficient lies between 1, -1. i.e., $-1 \le \rho \le 1$

→ If ~~value is~~ $\rho = 1$ there is a perfect Positive correlation

→ If $\rho = -1$ there is a perfect Negative correlation b/w given observations/variables

→ If $\rho = 0$ there is no correlation, in such case the given variables are independent to each other.

*Following are the Ranks obtained by 10 students in two subjects statistics, Mathematics. To what extent the knowledge of the students in 2 subjects is related.

Statistics	1	2	3	4	5	6	7	8	9	10
Mathematics	2	4	1	5	3	9	7	10	6	8

we know that

spearman's Rank correlation coefficient is

$$\rho = 1 - \frac{6 \Sigma D^2}{n(n^2-1)}$$

Here $n=10$

Rank in S R_1	Rank in M R_2	$D = R_1 - R_2$	D^2
1	2	-1	1
2	4	-2	4
3	1	2	4
4	5	-1	1
5	3	2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	3	9
10	8	2	4
			40

$$\rho = 1 - \frac{6(40)}{10(100-1)}$$

$$\rho = 1 - \frac{240}{10(99)}$$

$$\rho = 1 - \frac{240}{990}$$

$$\rho = 0.7575$$

$$\rho = 0.7$$

$\therefore$ There is more positive correlation b/w Statistics and Mathematics

2. A random sample of 5 college students are selected and their marks in Mathematics & Statistics are found to be

Mathematics	85	60	73	40	90
Statistics	93	75	65	50	80

calculate spearman's correlation coefficient

we know that spearman's Rank Correlation coefficient $\rho = 1 - \frac{6 \Sigma D^2}{n(n^2-1)}$

Here $n=5$

Marks in 'M'	Marks in 'S'	Rank in M R_1	Rank in S R_2	$D = R_1 - R_2$	D^2
85	93	2	1	1	1
60	75	4	3	1	1
73	65	3	4	-1	1
40	50	5	5	0	0
90	80	1	2	-1	1
		15	15	0	4

$$\rho = 1 - \frac{6(4)}{5(24)}$$

$$\rho = 1 - \frac{24}{5(24)}$$

$$\rho = 1 - \frac{1}{5}$$

$$\rho = \frac{4}{5}$$

$$\rho = 0.8$$

$\therefore$ There is more positive correlation b/w 'S' and 'M'.

3. 10 competitors in a musical test were ranked by the 3 judges 'A', 'B' & 'C' in the following order.

Ranks by 'A'	1	6	5	10	3	2	4	9	7	8
Ranks by 'B'	3	5	8	4	7	10	2	1	6	9
Ranks by 'C'	6	4	9	8	1	2	3	10	5	7

using rank correlation method discuss which pair of judges has the nearest approach to common likings in music.

(b) We know that spearmans rank correlation coefficient $\rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$

Here $n = 10$

Ranks by 'A' Ranks by 'B' Ranks by 'C' $D_1 = A-B$ $D_2 = B-C$ $D_3 = C-A$

Rank A	Rank B	Rank C	$D_1=A-B$	$D_2=B-C$	$D_3=C-A$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-3	5	4	9	25
6	5	4	1	1	-2	1	1	4
5	8	9	-3	-1	4	9	1	16
10	4	8	6	-5	-2	36	25	4
3	7	1	-4	6	-2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	-1	4	1	1
9	1	10	8	-9	1	64	81	1
7	6	5	1	1	-2	1	1	4
8	9	7	-1	2	-1	1	4	1
						200	214	60

$$\rho_{(A,B)} = 1 - \frac{6\Sigma D_1^2}{n(n^2-1)} \Rightarrow 1 - \frac{6(200)}{990}$$

$$= 1 - 1.21$$

$$= -0.21$$

$\rho(B,C) = 1 - \frac{6(214)}{990}$

$= -0.29$

$\rho(C,A) = 1 - \frac{6(60)}{990}$

$= 0.63$

[Judges A, C have nearest approach common in liking music]

Since $\rho(A,C)$ is maximum $\Rightarrow$ Judges A, C have nearest approach to common liking in music

16/04/22

Rank correlation for equal and repeated ranks

If two or more ranks are repeated in any observation then Spearman's rank correlation will slightly change. We use,

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1) + \cdots\right]}{n(n^2-1)}$$

where

ρ = rank correlation coefficient

ΣD^2 = sum of the squares of the difference between ranks

n = no. of given observations

m = no. of items whose ranks are common/equal.

NOTE:

If there is a tie for Rank 4, 5, 6 then the common rank $\frac{4+5+6}{3} = 5$ is assigned to all the equal values. Next rank is given as 7.

1. From the following data calculate the rank correlation coefficient after making adjustments for tied ranks.

x	48	33	40	9	16	16	65	24	16	57
y	13	18	24	6	15	4	20	9	6	19

We know that rank correlation coefficient for equal and repeated ranks is

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12} m(m^2-1) + \frac{1}{12} m(m^2-1) + \cdots\right]}{n(n^2-1)}$$

Here $m = 3, 2$

if $m = 3 \Rightarrow \frac{1}{12} m(m^2-1) = \frac{1}{12} 3(9-1)$

$= \frac{8}{4}$

$= 2$

ii) $m = 2 \Rightarrow \frac{1}{12} m(m^2-1) \Rightarrow \frac{1}{12} 2(4-1)$

$= \frac{6}{12} = 0.5$

x	y	R_1	R_2	$D = R_1 - R_2$	D^2
48	13	3	6	-3	9
33	18	5	4	1	1
40	24	4	1	3	9
9	6	10	8.5	1.5	2.25
16	15	8	5	3	9
16	4	8	10	-2	4
65	20	1	2	-1	1
24	9	6	7	-1	1
16	6	8	8.5	-0.5	0.25
57	19	2	3	-1	1
					37.50

$\frac{7+8+9}{3} = \frac{24}{3} = 8$

$$\rho = 1 - \frac{6[37.50 + 2 + 0.5]}{10(99)}$$

$$= 1 - \frac{6[40]}{990}$$

$$= 1 - \frac{240}{990}$$

$$= 1 - 0.24$$

$$= 0.76$$

i.e., There is more positive correlation b/w 'x' & 'y'

2. A sample of 12 fathers and their elder sons give the following data about their elder sons. calculate rank correlation coefficient

Fathers(x)	65	63	67	64	68	62	70	66	68	67	69	7
Sons(y)	68	66	68	65	69	66	68	65	71	67	68	7

We know that rank correlation coefficient for equal and repeated rank is

$$\rho = 1 - \frac{\left[6\,\Sigma D^2 + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1) + \cdots\right]}{n(n^2-1)}$$

Here $m = 2, 2, 4, 2, 2$

if $m = 2 \Rightarrow \frac{1}{12}m(m^2-1) = \frac{1}{12}2(4-1)$

$$= \frac{1}{\cancel{12}_{\cancel{6}\,2}}\cancel{2}(\cancel{3})$$

$$= 0.5$$

if $m = 4 \Rightarrow \frac{1}{12}m(m^2-1) = \frac{1}{12}4(16-1)$

$$= \frac{1}{\cancel{12}_{\cancel{3}}}\cancel{4}(\cancel{15}^{5})$$

$$= 5$$

Fathers (x)	Sons (y)	R_1	R_2	$D=R_1-R_2$	D^2
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	-1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	-3.5	12.25
66	65	8	11.5	-3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	-1.5	2.25
69	68	3	5.5	-2.5	6.25
71	70	1	2	-1	1
					72.5

$$\rho = 1 - \frac{6[72.5+0.5+0.5+0.5+0.5+5]}{12(144-1)}$$

$$= 1 - \frac{477}{1716}$$

$$= 1 - 0.27$$

$$= 0.73$$

i.e., There is more positive correlation b/w fathers and sons

18/04/2022

Regression: A statistical measure which is used to estimate the unknown value of one variable from the known value of another related variable is known as Regression.

Regression line: A line which is used to describe the average relationship b/w two variables is known as regression line. There are two types of regression lines

1. Regression line of 'x' on 'y'
2. Regression line of 'y' on 'x'

Regression line of x on y:

It is obtained by using $(x-\bar{x}) = b_{xy}(y-\bar{y})$

where $\bar{x}$ = mean of 'x'

$\bar{y}$ = mean of 'y'

$b_{xy} = r\frac{\sigma x}{\sigma y}$ (regression coefficient of 'x' on 'y')

$= \frac{\Sigma xy}{\Sigma y^2}$ (if $\bar{x}, \bar{y}$ are whole numbers)

where $x = \Sigma x - \bar{x}$

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

$= \cancel{\frac{\Sigma x \cdot \Sigma y}{n}}$

$= \frac{\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}$ (if $\bar{x}, \bar{y}$ are fractions)

Regression line of 'y' on 'x':

It is obtained by $(y-\bar{y}) = b_{yx}(x-\bar{x})$

where $b_{yx} = r\frac{\sigma_y}{\sigma_x}$

$= \frac{\Sigma XY}{\Sigma X^2}$ (if $\bar{x}, \bar{y}$ are whole number)

$= \frac{\Sigma XY - \frac{\Sigma X \cdot \Sigma Y}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}$ (if $\bar{x}, \bar{y}$ are fraction)

NOTE:

→ If b_{xy} is regression coefficient of x on y ~~under~~ and b_{yx} is regression coefficient of y on x then correlation coefficient 'r' is $r = \sqrt{b_{xy} \times b_{yx}}$

→ The values of b_{xy}, b_{yx} are either positive(+ve) or negative.

→ Both the regression line will pass through the point $(\bar{x}, \bar{y})$

Uses of regression:

→ It is used to estimate the relation b/w two economic variables like income and expenditure.

→ It is a highly valuable tool used in economics and Business.

→ It is widely used in prediction purpose

→ It is useful in statistical estimation of demand curves, supply curves.

1. find the most likely production corresponding to a rainfall 40 from the following data

	Rainfall(x)	Production(y)
Avg	30	50
SD	5	100
coefficient of correlation	0.8	

Given that $\bar{x} = 30$ $\bar{y} = 50$

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

$r = 0.8$

We know that regression line of y on x is $y - \bar{y} = b_{yx}(x - \bar{x})$

Here $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$

$$y - 50 = 0.8\frac{(100)}{(5)}(x - 30)$$

$$y - 50 = 16(x - 30)$$

$$y - 50 = 16x - 480$$

$$\boxed{y = 16x - 430}$$

if $x = 40$

$$\Rightarrow y = 16(40) - 430$$

$$y = 640 - 430$$

$$y = 210$$

i.e., the production is 210 corresponding to rainfall 40

20/4/22

2. find the regression line of x on y and y on x to the data given below

x	10	12	13	16	17	20	25
y	10	22	24	27	29	33	37

Sol: we know that

$$\bar{x} = \frac{10+12+13+16+17+20+25}{7} \Rightarrow \bar{x} = 16.14 \cong 16$$

$$\bar{y} = \frac{10+22+24+27+29+33+37}{7} \Rightarrow \bar{y} = 180.26$$

We know that

(i) regression line of x on y is

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

here $b_{xy} = \dfrac{\Sigma XY - \frac{\Sigma X \Sigma Y}{n}}{\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}}$

$X = \Sigma(x_i - \bar{x})$

$Y = \Sigma(y_i - \bar{y})$

x	y	$X = x-16$	$Y = y-26$	X^2	Y^2	XY
10	10	-6	-16	36	256	96
12	22	-4	-4	16	16	16
13	24	-3	-2	9	4	6
16	27	0	1	0	1	0
17	29	1	3	1	9	3
20	33	4	7	16	49	28
25	37	9	11	81	121	99
		1	0	159	456	248

$$(x-16) = \frac{248 - \frac{2(0)}{7}}{1156 - \frac{(0)^2}{7}} (y-26)$$

$$(x-16) = \frac{248}{1156}(y-26)$$

$$(x-16) = 0.54(y-26)$$

$$x - 16 = 0.54y - 14.04$$

$$x = 0.54y + 16 - 14.04$$

$$x = 0.54y + 1.96$$

(iii) similarly regression of y on x is

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

where $b_{yx} = \dfrac{\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$

$$(y-26) = \frac{248 - \frac{1(0)}{7}}{159 - \frac{(1)^2}{7}}(x-16)$$

$$y - 26 = \frac{248}{158.86}(x-16)$$

$$y - 26 = 1.56(x-16)$$

$$y - 26 = 1.56x - 24.96$$

$$y = 1.56x + 26 - 24.96$$

$$y = 1.56x + 1.04$$

3. Price indices of cotton and wool are given below for 12 months of a year obtain the equation of lines of regression between the indices.

Price index of cotton(x)	78	77	85	88	87	82	81	77	76	83	97	9[illegible]
wool(y)	84	82	82	85	89	90	88	92	83	89	98	9[illegible]

we know that

$$\bar{x} = \frac{78+77+85+88+87+82+81+77+76+83+97+93}{12}$$

$$= \frac{1004}{12}$$

$$\simeq 83.66$$

$$\cong 84$$

$$\bar{y} = \frac{84+82+82+85+89+90+88+92+83+89+98+9\ldots}{12}$$

$$= \frac{1061}{12}$$

$$= 88.41$$

$$\cong 88$$

(i) Regression line of 'x' on 'y' is

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

where $b_{xy} = \dfrac{\Sigma xy - \dfrac{\Sigma x \cdot \Sigma y}{n}}{\Sigma y^2 - \dfrac{(\Sigma y)^2}{n}}$

x	y	$X = x - 84$	$Y = y - 88$	X^2	Y^2	XY
78	84	-6	-4	36	16	24
77	82	-7	-6	49	36	42
85	82	1	-6	1	36	-6
88	85	4	-3	16	9	+12
87	89	3	1	9	1	3
82	90	-2	2	4	4	-4
81	88	-3	0	9	0	0
77	92	-7	4	49	16	-28
76	83	-8	-5	64	25	+45
83	89	-1	1	1	1	1
97	98	13	10	169	100	130
93	99	9	11	81	121	99
		-4	5	488	365	287

$$(x-84) = \frac{287 - \frac{(-4)(5)}{12}}{365 - \frac{(5)^2}{12}}(y-88)$$

$$(x-84) = \frac{287 + \frac{20}{12}}{365 - \frac{25}{12}}(y-88)$$

$$(x-84) = \frac{287 + 1.66}{365 - 2.08}(y-88)$$

$$(x-84) = \frac{288.66}{362.92}(y-88)$$

$$(x-84) = 0.79(y-88)$$

$x - 84 = 0.79y - 69.99$

$x = 0.79y + 84 - 69.99$

$x = 0.79y + 14.01$

(ii) Regression line of 'y' on 'x' is.

$$(y - \bar{y}) = byx(x - \bar{x})$$

$$byx = \frac{\Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$(y - 88) = \frac{287 - \frac{(-4)(5)}{12}}{488 - \frac{(-4)^2}{12}}(x - 84)$$

$$(y - 88) = \frac{287 + \frac{20}{12}}{488 - \frac{16}{12}}(x - 84)$$

$$(y - 88) = \frac{288.66}{486.67}(x - 84)$$

$(y - 88) = 0.59(x - 84)$

$y - 88 = 0.59x - 49.82$

$y = 0.59x + 38.18$

4. Calculate the regression line of 'y' on 'x' from the data given below then estimate the likely demand when the price is Rs 20

Price (x)	10	12	13	12	16	15
Demand (y)	40	38	43	45	37	43

We know that

$$\bar{x} = \frac{10+12+13+12+16+15}{6}$$

$$= \frac{78}{6}$$

$$= 13$$

$$\bar{y} = \frac{40+38+43+45+37+43}{6}$$

$$= \frac{246}{6}$$

$$= 41$$

∴ Regression line of 'y' on 'x'

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

$$byx = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

x	y	X = x − 13	Y = y − 41	X^2	Y^2	XY
10	40	−3	−1	9	1	3
12	38	−1	−3	1	9	3
13	43	0	2	0	4	0
12	45	−1	4	1	16	−4
16	37	3	−4	9	16	−12
15	43	2	2	4	4	4
		0	0	24	50	−6

$$(y-41) = \frac{-6 - \frac{(0)(0)}{6}}{24 - \frac{(0)^2}{6}}(x-13)$$

$$(y-41) = \frac{-6}{24}(x-13)$$

$$(y-41) = -0.25(x-13)$$

$$y-41 = -0.25x + 3.25$$

$$y = -0.25x + 41 + 3.25$$

$$y = -0.25x + \cancel{37.75}\ 44.25$$

$$y = -0.25x + 44.25$$

Given $x = 20$

$$y = -0.25(20) + 44.25$$

$$y = -5 + 44.25$$

$$y = 39.25$$

23/4/22

5 In the following table set of x, y values are given then construct regression line of y on x. hence find the value of y if $x=50$

x	0	20	40	60	80
y	54	65	75	85	96

we know that $\bar{x} = \dfrac{0+20+40+60+80}{5}$

$= 40$

$\bar{y} = \dfrac{54+65+75+85+96}{5}$

$= 75$

We know that regression line of y on x is $(y-\bar{y}) = b_{yx}(x-\bar{x})$ where

$$b_{yx} = \frac{\Sigma XY}{\Sigma X^2}$$

Here $X = \Sigma x_i - \bar{x}$, $Y = \Sigma(y_i - \bar{y})$

x	y	$X = x-40$	$Y = y-75$	XY	X^2
0	54	-40	-21	840	1600
20	65	-20	-10	200	400
40	75	0	0	0	0
60	85	20	10	200	400
80	96	40	21	840	1600
		0	0	2080	4000

$$y-75 = \frac{2080}{4000}(x-40)$$

$$y-75 = 0.52(x-40)$$

$$y-75 = 0.52x - 20.8$$

$$y = 0.52x + 75 - 20.8$$

$y = 0.52x + 54.2$ which is required regression line.

given $x = 50$

$$y = 0.52(50) + 54.2$$

$$y = 26 + 54.2$$

$$y = 80.2$$

Note:

→ If θ is the angle b/w two regression lines then $\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$

1. If $\sigma_x = \sigma_y = \sigma$ and angle b/w regression lines is $\tan^{-1}\left(\frac{4}{3}\right)$ then find r

Given that $\sigma_x = \sigma_y = \sigma$

$$\tan\theta = \frac{4}{3}$$

we know that angle b/w two regression lines $\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$

$$\frac{4}{3} = \frac{1-r^2}{r}\left(\frac{\sigma \cdot \sigma}{\sigma^2 + \sigma^2}\right)$$

$$\frac{4}{3} = \frac{1-r^2}{r}\left(\frac{\sigma^2}{2\sigma^2}\right)$$

$$\frac{4}{3} = \frac{1-r^2}{r}\left(\frac{1}{2}\right)$$

$$\frac{1-r^2}{r} = \frac{8}{3}$$

$$3 - 3r^2 = 8r$$

$$3r^2 + 8r - 3 = 0$$

$$3r^2 + 9r - r - 3 = 0$$

$$3r(r+3) - 1(r+3) = 0$$

$$(r+3)(3r-1) = 0$$

$$r = -3, \frac{1}{3}$$

Here $r = \frac{1}{3}$ $(\because -1 \le r \le 1)$

2. If the regression line of x on y is $3y - 2x - 10 = 0$, regression line of y on x is $2y - x - 50 = 0$ then find $\bar{x}, \bar{y}$ and coefficient of correlation.

Given both the regression lines are:

$$3y - 2x - 10 = 0 \rightarrow (1)$$

$$2y - x - 50 = 0 \rightarrow (2)$$

we know that eq (1), (2) will pass through the point $(\bar{x}, \bar{y})$

$$(3\bar{y} - 2\bar{x} - 10 = 0) \times 2$$

$$2\bar{y} - \bar{x} - 50 = 0 \times 3$$

$$\cancel{3\bar{y} - 2\bar{x} - 10 = 0}$$

$$\cancel{4\bar{y} - 2\bar{x} - 100 = 0}$$

$$\cancel{7\bar{y} + 90 = 0}$$

$$6\bar{y} - 4\bar{x} - 20 = 0$$

$$6\bar{y} - 3\bar{x} - 150 = 0$$

$$-\bar{x} + 130 = 0$$

$$\boxed{\bar{x} = 130}$$

$$\cancel{-14\bar{y}} = \cancel{-910}$$

$$\cancel{\bar{y}} = \cancel{-18.85}$$

$$2\bar{y} = 130 + 50$$

$$\cancel{2}\bar{y} = \overset{90}{\cancel{180}}$$

$$\boxed{\bar{y} = 90}$$

$$① \Rightarrow 2x = 3y - 10$$

$$x = \frac{3}{2}y - 5$$

$$b_{xy} = \frac{3}{2}$$

IIIy $② \Rightarrow 2y = x + 50$

$$y = \frac{x}{2} + 25$$

$$b_{yx} = \frac{1}{2}$$

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

$$= \sqrt{\frac{3}{2} \cdot \frac{1}{2}}$$

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

$$r^2 = \frac{3}{4}$$

$$r = 0.866$$

3. For 20 army personal the regression weight of kidneys (x) on weight of heart(y) is given as $x = 1.212y + 2.461$ and $y = 0.399x + 6.394$ then find the correlation coefficient and $\bar{x}, \bar{y}$

Given both regression lines are

$$x = 1.212y + 2.461 \rightarrow ①$$

$$y = 0.399x + 6.394 \rightarrow ②$$

$$① \Rightarrow b_{xy} = 1.212$$

$$② \Rightarrow b_{yx} = 0.399$$

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

$$= \sqrt{(1.212)(0.399)}$$

$$= \sqrt{0.483588}$$

$$r = 0.69$$

$$① \Rightarrow \bar{x} - 1.212\bar{y} - 2.461 = 0$$

$$② \Rightarrow -0.399\bar{x} + \bar{y} - 6.394 = 0$$

solving ① & ②

$$-0.399\bar{x} + 0.483588\bar{y} + 0.981939 = 0$$

$$\underset{(+)}{-0.399\bar{x}} \underset{(-)}{+} \bar{y} \quad \underset{(+)}{-6.394} = 0$$

$$-0.516412\bar{y} + 7.375939 = 0$$

$$0.516412\bar{y} = 7.375939$$

$$\bar{y} = 14.28$$

$\bar{x} - 1.212(14.28) - 2.461 = 0$

$\bar{x} - 17.30736 - 2.461 = 0$

$\bar{x} = 19.76$

8/04/2022

Least square curve fitting:

Curve fitting is a systematic procedure to fit an unique curve through the given data points. It's application is wide in ~~tropical~~ Pratical competitions.

(i) fitting a straight line:

Let $y = a_0 + a_1 x \rightarrow (1)$ with the equation of the straight line to be construct to the given data points by the method of least squares where a_0, a_1 are constants.

We may find a_0, a_1 by solving the two normal equations

$$n a_0 + a_1 \Sigma x = \Sigma y \rightarrow (2)$$

$$a_0 \Sigma x + a_1 \Sigma x^2 = \Sigma xy \rightarrow (3)$$

By solving eq (2), (3) we get a_0, a_1 then substitute a_0, a_1 in (1) to get require straight line

1. By the method of least squares find the straight line that best fits to the following data

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

Let $y = a_0 + a_1 x \rightarrow (1)$ be the equation of straight line to be construct to given data points where a_0, a_1 are constants.

We may get a_0, a_1 by solving the normal equations

$$n a_0 + a_1 \Sigma x = \Sigma y$$
$$a_0 \Sigma x + a_1 \Sigma x^2 = \Sigma xy$$

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

$\therefore$ Normal eqn's are

$$5a_0 + a_1(15) = 204 \rightarrow (2)$$
$$15a_0 + 55a_1 = 748 \rightarrow (3)$$

solving ② & ③

$$15a_0 + 45a_1 = 612$$
$$15a_0 + 55a_1 = 748$$

$$-10a_1 = -136$$

$$a_1 = \frac{136}{10}$$

$$a_1 = 13.6$$

$$5a_0 + 15(13.6) = 204$$

$$5a_0 + 204 = 204$$

$$5a_0 = 0$$

$$a_0 = 0$$

Sub a_0 & a_1 in ①

$$y = 0 + 13.6x$$

$$y = 13.6x$$

2. Fit a straight line in the following data $y = a+bx$ to the following data by the method least squares.

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Let $y = a+bx \rightarrow ①$ be the equation of straight line to be construct to given data points where a, b are constants.

we may get a, b by solving the normal equations

$$na + b\Sigma x = \Sigma y$$

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

x	y	x^2	xy
0	12	0	0
5	15	25	75
10	17	100	170
15	22	225	330
20	24	400	480
25	30	625	750
75	120	1375	1805

∴ Normal eqns are

$$6a + b(75) = 120 \rightarrow (2)$$
$$75a + b(1375) = 1805 \rightarrow (3)$$

$$a = 11.28 \quad b = 0.6971$$

Sub a & b in (1)

$y = 11.28 + 0.6971x$ which is required straight line equation.

8. Fit a straight line to the following data by the method of least squares

x	0	1	2	3	4
y	1	1.8	3.3	4.5	6.3

let $y = a_0 + a_1 x$ → ① be the equation of straight line to be construct to given data points where a_0, a_1 are constants.

we may get a_0, a_1 by solving the normal equations

$$n a_0 + a_1 \Sigma x = \Sigma y$$

$$a_0 \Sigma x + a_1 \Sigma x^2 = \Sigma xy$$

x	y	x^2	xy
0	1	0	0
1	1.8	1	1.8
2	3.3	4	6.6
3	4.5	9	13.5
4	6.3	16	25.2
10	16.9	30	47.1

∴ Normal eqn's are

$$5a_0 + 10a_1 = 16.9 \rightarrow ②$$

$$10a_0 + 30a_1 = 47.1 \rightarrow ③$$

$$a_0 = 0.72 \quad a_1 = 1.33$$

Sub a_0 & a_1 in ①

$y = 0.72 + 1.33x$ which is required straight line equation

4/5/22

Second degree polynomial (Parabola):

Let $y = a_0 + a_1 x + a_2 x^2 \rightarrow (1)$ be the equation of the second degree polynomial (Parabola) to be construct to the given data points by the method of least squares.

We may get a_0, a_1, a_2 by solving following three normal equations

$$na_0 + a_1 \Sigma x + a_2 \Sigma x^2 = \Sigma y \rightarrow (2)$$

$$a_0 \Sigma x + a_1 \Sigma x^2 + a_2 \Sigma x^3 = \Sigma xy \rightarrow (3)$$

$$a_0 \Sigma x^2 + a_1 \Sigma x^3 + a_2 \Sigma x^4 = \Sigma x^2 y \rightarrow (4)$$

Substitute a_0, a_1, a_2 in (1) to get second degree polynomial (or) parabola.

1. Fit a second degree polynomial (parabola) to the following data by the method of least squares

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

Let $y = a + bx + cx^2 \rightarrow (1)$ is the eqn of the Parabola to be construct to the given data points.

We may get the constants a, b, c by solving the following 3 normal equations.

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

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

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

x	y	xy	x^2	x^3	x^4	x^2y
0	1	0	0	0	0	0
1	1.8	1.8	1	1	1	1.8
2	1.3	2.6	4	8	16	5.2
3	2.5	7.5	9	27	81	22.5
4	6.3	25.2	16	64	256	100.8
10	12.9	37.1	30	100	354	130.3

$\therefore$ Normal eqns are

$$5a + 10b + 30c = 12.9 \rightarrow (2)$$
$$10a + 30b + 100c = 37.1 \rightarrow (3)$$
$$30a + 100b + 354c = 130.3 \rightarrow (4)$$

$a = 1.42$
$b = -1.07$
$c = 0.55$

Sub a, b, c in eq (1)

$y = 1.42 - 1.07x^2 + 0.55x^2$ is the required parabola.

2. Fit a polynomial of second degree to the data points given in the following table.

x	0	1	2
y	1	6	17

Let $y = a + bx + cx^2 \rightarrow (1)$ is the equation of parabola to be construct to given data points.

We may get constants a, b, c by solving following 3 normal equations

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

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

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

x	y	xy	x^2	x^3	x^4	x^2y
0	1	0	0	0	0	0
1	6	6	1	1	1	6
2	17	34	4	8	16	68
3	24	40	5	9	17	74

∴ Normal eqns are

$3a + 3b + 5c = 24 \rightarrow (2)$

$3a + 5b + 9c = 40 \rightarrow (3)$

$5a + 9b + 17c = 74 \rightarrow (4)$

By solving (2), (3) & (4)

$a = 1, b = 2, c = 3$

sub a, b, c in (1)

$y = 1 + 2x^2 + 3x^3$ is required parabola.

Exponential Function

Working procedure

let $y = ae^{bx} \rightarrow (1)$ be the equation of the exponential function.

Apply log on both sides.

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

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

$\log_e^y = \log_e^a + bx \longrightarrow (2)$

eq (2) is of the form $Y = a_0 + a_1 X \longrightarrow (3)$

where a_0, a_1 are constants

Here $Y = \log_e^y$

$a_0 = \log_e^a \Rightarrow a = e^{a_0}$

$a_1 = b$

$X = x$

a_0, a_1 may be obtained by solving the normal equations

$na_0 + a_1 \Sigma X = \Sigma Y$

$a_0 \Sigma X + a_1 \Sigma X^2 = \Sigma XY$

5/5/2022

1. Find the curve which is of the form $y = ae^{bx}$ (Exponential function) to the following data by the method of least squares

x	0	1	2	3	4	5	6	7	8
y	20	30	52	77	135	211	326	550	1052

Let $y = ae^{bx} \longrightarrow (1)$ is an exponential function to be construct to the given data by the method of least squares

Apply log on both sides

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

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

$$\log_e^y = \log_e^a + bx \rightarrow (2)$$

eq (2) is of the form

$$Y = a_0 + a_1 X \rightarrow (3)$$

where $Y = \log_e^y$

$a_0 = \log_e^a$

$a_1 = b$

$X = x$

we know that eq (3) is a straight line

we may get a_0, a_1 by solving 2 normal equations

$$n a_0 + a_1 \Sigma X = \Sigma Y$$

$$a_0 \Sigma X + a_1 \Sigma X^2 = \Sigma XY$$

$x = X$	$Y = \log_e^y$	XY	X^2
0	2.996	0	0
1	3.401	3.401	1
2	3.951	7.902	4
3	4.344	13.032	9
4	4.905	19.620	16
5	5.352	26.760	25
6	5.767	34.722	36
7	6.310	44.170	49
8	6.958	55.664	64
36	44.004	205.271	204

∴ The normal equations are

$9a_0 + 36a_1 = 44.001 \rightarrow (4)$

$36a_0 + 201a_1 = 205.271 \rightarrow (5)$.

By solving (4) & (5)

$a_0 = 2.939$

$a_1 = 0.487$

But we have $a_0 = \log_e a$

$a = e^{a_0}$

$a = e^{2.939}$

$\boxed{a = 18.897}$

$b = a_1$

$\boxed{b = 0.487}$

sub a,b in (1)

$y = ae^{bx}$

$y = 18.897\, e^{0.487x}$ which is required exponential function.

3. Fit an exponential function which is of the form $y = ae^{bx}$ to the following data by the method of least square

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

let $y = ae^{bx}$ → ① is an exponential function to be construct to the given data by the method of least squares

Apply log on both sides

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

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

$$\log_e y = \log_e a + bx \rightarrow ②$$

eq② is of the form

$$Y = a_0 + a_1 X \rightarrow ③$$

where $Y = \log_e y$

$a_0 = \log_e a$

$a_1 = b$

$X = x$

we know that eq③ is a straight line we may get a_0, a_1 by solving 2 normal equations

$$na_0 + a_1 \Sigma X = \Sigma Y$$

$$a_0 \Sigma X + a_1 \Sigma X^2 = \Sigma XY$$

$x = x$	$y = \log_e y$	xy	x^2
77	0.875	67.375	5929
100	1.224	1224	10,000
185	1.946	360.01	34225
239	2.407	575.273	57121
285	2.975	847.875	81225
886	~~883.552~~ 9.427	1972.933	188500

$\therefore$ The normal equations are

$$5a_0 + 886a_1 = \cancel{883.552}\ 9.427 \rightarrow (4)$$

$$886a_0 + 188500a_1 = 1972.933 \rightarrow (5)$$

By solving (4) & (5)

$$a_0 = \cancel{1046.332}\ 0.184$$

$$a_1 = \cancel{-4.904}\ 0.0096$$

But we have $a_0 = \log_e a$

$$a = e^{a_0}$$

$$\cancel{a = e^{1046.330}}$$

$$a = e^{0.184}$$

$$\boxed{a = 1.202}$$

$$b = 0.0096$$

sub a, b in (1)

$$y = 1.202\,e^{0.0096x}$$

7/5/22

Power function:

Let $y = ab^x \rightarrow (1)$ be the equation of the power function to be construct to the given data points where a, b are constants

Apply log on both sides

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

$$\log_{10}^{y} = \log_{10}^{a} + x\log_{10}^{b} \rightarrow (2)$$

we may write eq (2) as

$$Y = a_0 + a_1 X \rightarrow (3)$$

where $Y = \log_{10}^{y}$

$a_0 = \log_{10}^{a}$

$a_1 = \log_{10}^{b}$

$X = x$

we know that eq (3) is a straight line we may get a_0, a_1 by solving the two normal equations

$$na_0 + a_1\Sigma X = \Sigma Y \rightarrow (4)$$

$$a_0\Sigma X + a_1\Sigma X^2 = \Sigma XY \rightarrow (5)$$

we have $a_0 = \log_{10}^{a}$

$a = 10^{a_0}$

ll^{ly} $a_1 = \log_{10}^{b}$

$b = 10^{a_1}$

sub a, b in ① to get a power function.

1. Fit a curve of the form $y = ab^x$ (Power function) to the following data by the method of least squares

x	2	3	4	5	6
y	8.3	15.4	33.1	65.2	127.4

Let $y = ab^x \rightarrow ①$ is the equation of the power function to be construct.

Apply log on both sides

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

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

$$\log_{10}^{y} = \log_{10}^{a} + x\log_{10}^{b} \rightarrow ②$$

eq ② is of the form

$$Y = a_0 + a_1 X \rightarrow ③$$

where $Y = \log_{10}^{y}$

$a_0 = \log_{10}^{a}$, $a_1 = \log_{10}^{b}$, $X = x$

we know that eq ③ is a straight line.
we may also get a_0, a_1 by solving the two normal equations

$$na_0 + a_1 \Sigma X = \Sigma Y$$
$$a_0 \Sigma X + a_1 \Sigma X^2 = \Sigma XY$$

X	$Y = \log_{10}^{y}$	XY	X^2
2	0.919	1.838	4
3	1.188	3.564	9
4	1.520	6.080	16
5	1.814	9.070	25
6	2.105	12.630	36
20	7.546	33.182	90

$\therefore$ The normal equations are

$$5a_0 + 20a_1 = 7.546 \rightarrow ④$$
$$20a_0 + 90a_1 = 33.182 \rightarrow ⑤$$

By solving ④ & ⑤

$$a_0 = 0.310$$
$$a_1 = 0.300$$

But we have $a_0 = \log_{10}^{a}$

$$a = 10^{a_0}$$
$$a = 10^{0.310}$$
$$\boxed{a = 2.0.42}$$

$a_1 = \log_{10} b$

$b = 10^{a_1}$

$b = 10^{0.300}$

$\boxed{b = 1.995}$

sub a,b in ①

$y = (2.042)(1.995)^x$ which is required power function.

2. Fit a curve which is of the form $y = ab^x$ (Power function) to the following data by the method of least squares

x	2	3	4	5	6
y	144	172.8	207.4	248.8	298.5

Let $y = ab^x \longrightarrow$ ① is the equation of the power function to be construct.

Apply log on both sides

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

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

$\log_{10} y = \log_{10} a + x \log_{10} b \longrightarrow$ ②

eq ② is of the form

$Y = a_0 + a_1 X \longrightarrow$ ③

where $Y = \log_{10} y$

$a_0 = \log_{10} a$, $a_1 = \log_{10} b$; $X = x$

we know that eq (3) is a straight line
we may get a_0, a_1 by solving the two normal equations

$$na_0 + a_1\Sigma X = \Sigma Y$$
$$a_0\Sigma X + a_1\Sigma X^2 = \Sigma XY$$

X	$Y = \log_{10} y$	XY	X^2
2	2.158	4.316	4
3	2.238	6.714	9
4	2.317	9.268	16
5	2.396	11.980	25
6	2.475	14.850	36
20	11.584	47.128	90

∴ The normal equations are

$$5a_0 + 20a_1 = 11.584 \longrightarrow (4)$$
$$20a_0 + 90a_1 = 47.128 \longrightarrow (5)$$

By solving (4) & (5)

$$a_0 = 2, \quad a_1 = 0.079$$

we have $a_0 = \log_{10} a$

$$a = 10^{a_0}$$
$$a = 10^2$$
$$\boxed{a = 100}$$

$a_1 = \log_{10}$

$b = 10^{a_1}$

$b = 10^{0.079}$

$\boxed{b = 1.199}$

sub a,b in (1)

$y = (100)(1.199)^x$ which is required power function

6/6/22

Random variables: Random variable is a function which is defined from sample space (s) to Real number set (R).

i.e., $x : S \to R$ is a Random variable

We represent random variables by capital letters of English alphabets and the values assigned to the random variables are represented with small letters or numericals

Random variables are classified into 2 types

1. Discrete Random variable
2. continuous Random variable

Discrete Random variable: If the random variable x takes only finite no.of values then it is known as a discrete random variable.

Ex: In the random experiment of throwing a Die then the random variable x defined as no. of even numbers in that experiment

Continuous Random variable: If the random variable 'x' takes infinite no. of values then it is known as continuous Random variables.

Ex: 1. The height of student between 5 feet to 6 feet.

2. The weight of the students between 50kgs to 60kgs

Cummulative distributive function: The cummulative distributive function of a discrete random variable 'x' is represented by $F(x)$ and it is defined as $F(x) = P(X \leq x)$

Ex: $F(3) = P(X \leq 3)$

$= P(0) + P(1) + P(2) + P(3)$

Discrete Probability distribution: If the random variable 'x' takes the value $x_1, x_2, x_3, \ldots x_n$ and the corresponding probabilities are $P_1, P_2, P_3, \ldots P_n$ then the set of x_i, P_i values is known as discrete probability distribution of discrete random variable.

x_i	$P(x_i)$
x_1	$P(1)$
x_2	$P(2)$
$\vdots$	
x_n	$P(n)$

Probability function: The probability function of a discrete random variable 'x' is defined as $P(x_i) = f(x_i) = P(X=x)$

NOTE: If $F(x)$ is cummulative distributive function ~~tunch~~ and $f(x)$ is probability function

(i) $\sum_{i=1}^{n} f(x_i) = 1$

(ii) $f(x_i) \geq 0$

(iii) $\frac{d}{dx} F(x) = f(x)$

Mean (or) expectation of a discrete Random variable

If the discrete Random variable 'x' takes the values $x_1, x_2, x_3, \ldots x_n$ and the corresponding probabilities are $P_1, P_2, P_3, \ldots P_n$ then the mean or ~~are~~ expectation of the discrete random variable is defined as

$$\mu = E(x) = \sum_{i=1}^{n} P_i x_i$$

$$\Rightarrow \mu = E(X) = P_1 x_1 + P_2 x_2 + P_3 x_3 + \cdots + P_n x_n$$

Variance: The variance of a discrete random variable 'x' is defined as $\sigma^2 = \sum_{i=1}^{n} P_i x_i^2 - \mu^2$

$$\sigma^2 = E(X^2) - \{E(X)\}^2$$

Standard Deviation: The positive square root of a variance is known as standard Deviation of Discrete random variable

i.e., $\sigma = \sqrt{E(X^2) - \{E(X)\}^2}$ (or) $\sigma = \sqrt{\sum_{i=1}^{n} P_i x_i^2 - \mu^2}$

1. If a random variable 'x' has the following probability function

x	0	1	3	4	5	6	7
$P(x)$	0	K	$2K$	$2K$	$3K$	K^2	$7K^2+K$

then find

(i) value of K

(ii) evaluate $(P(x<6))$, $P(x\geq 6)$

(iii) $P(0<x<5)$

Sol (i) we know that $\sum_{i=1}^{n} P(x_i)=1$

$$\Rightarrow P(0)+P(1)+ \cdots +P(7)=1$$

$$0+K+2K+3K+K^2+7K^2+K=1$$

$$8K^2+9K=1$$

$$8K^2+9K-1=0$$

$$\cancel{8K^2+9K}$$

$$K=\frac{-9\pm\sqrt{113}}{16}$$

$$K=\frac{-9+10.63}{16}, \quad K=\frac{-9-10.63}{16}$$

$$K=0.102, -1.227$$

Here $K=0.1$

(ii) $P(x<6) = P(0)+P(1)+\cancel{P(2)}+P(3)+P(4)+P(5)$

$$=0+K+2K+2K+3K$$

$$=8K$$

$$=8(0.1)$$

$$=0.8$$

$P(x \geq 6) = P(6) + P(7)$

$= K^2 + 7K^2 + K$

$= 8K^2 + K$

$= 8(0.1)^2 + 0.1$

$= 0.18$

(iii) $P(0 < x < 5) = P(1) + P(3) + P(4)$

$= K + 2K + 2K$

$= 5K$

$= 5(0.1)$

$= 0.5$

2. A random variable 'x' has the following probability distribution i.e.,

x_i	0	1	2	3	4	5	6	7
$p(x_i)$	0	K	2K	2K	3K	K^2	$2K^2$	$7K^2+K$

then determine

(i) value of K

(ii) evaluate $P(x<6)$, $P(x \geq 6)$ & $P(0<x<5)$

(iii) if $P(x \leq K) > \frac{1}{2}$ then find min value of K

(iv) Mean (v) variance (vi) S.D (vii) F(x)

Sol: (i) we know that $\sum_{i=1}^{n} p(x_i) = 1$

$P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7) = 1$

$0 + K + 2K + 2K + 3K + K^2 + 2K^2 + 7K^2 + K = 1$

$10K^2 + 9K = 1$

$10K^2 + 9K - 1 = 0$

$10K^2 + 10K - K - 1 = 0$

$10K(K+1) - 1(K+1) = 0$

$(10K-1)(K+1) = 0$

$K = 0.1, -1$

$\therefore K = 0.1$

(ii) $P(x<6) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5)$

$= 0 + K + 2K + 2K + 3K + K^2$

$= K^2 + 8K$

$= (0.1)^2 + 8(0.1)$

$= (0.1)^2 + 0.8$

$= 0.01 + 0.8$

$= 0.81$

$P(x \geq 6) = P(6) + P(7)$

$= 2K^2 + 7K^2 + K$

$= 9K^2 + K$

$= 9(0.01) + K$

$= 0.09 + 0.1$

$= 0.19$

$P(0<x<5) = P(1) + P(2) + P(3) + P(4)$

$= K + 2K + 2K + 3K$

$= 8K$

$= 8(0.1)$

$= 0.8$

(iii) if $K = 1 \Rightarrow P(x \leq 1) = P(0) + P(1)$

$= 0 + K$

$= 0.1 \neq \frac{1}{2}$

if $K = 2 \Rightarrow P(x \leq 2) = P(0) + P(1) + P(2)$

$= 0 + K + 2K$

$= 3K$

$= 3(0.1)$

$= 0.3 \ngtr \frac{1}{2}$

if $K = 3 \Rightarrow P(X \leq 3) = P(0) + P(1) + P(2) + P(3)$

$= 0 + K + 2K + 2K$

$= 5K$

$= 5(0.1)$

$= 0.5 \ngtr \frac{1}{2}$

if $K = 4 \Rightarrow P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4)$

$= 0 + K + 2K + 2K + 3K$

$= 8K$

$= 8(0.1)$

$= 0.8 > \frac{1}{2}$

i.e. minimum value of 'K' such that $P(X \leq K) > \frac{1}{2}$ is '4'

(iv) we know that mean of a discrete random variable $\mu = \sum_{i=1}^{n} p_i x_i$

$\Rightarrow 0(0) + 1(K) + 2(2K) + 3(2K) + 4(3K) + 5(K^2) + 6(2K^2) + 7(7K^2 + K)$

$\Rightarrow 0 + K + 4K + 6K + 12K + 5K^2 + 12K^2 + 49K^2 + 7K$

$\Rightarrow 66K^2 + 30K$

$\Rightarrow 66(0.01) + 30(0.1)$

$\Rightarrow \mu = 3.66$

(v) we know that variance of the discrete random variable

$$\sigma^2 = \sum_{i=1}^{n} p_i x_i^2 - \mu^2$$

$$= 0(0^2) + K(1^2) + 2K(4) + 2K(9) + 3K(16) + K^2(25) + 2K^2(36) + 7K^2 + K(49) - (3.66)$$

$$= K + 8K + 18K + 48K + 25K^2 + 72K^2 + 343K^2 + 49K - 13.39$$

$$\sigma^2 = 3.404$$

(vi) we know that positive square root of variance is SD

$$\sigma = \sqrt{3.404}$$

$$\sigma = 1.84$$

(vii) we know that cummulative distributive function

x_i	$F(x_i) = P(X \leq x_i)$
0	0
1	$K = 0.1$
2	$3K = 0.3$
3	$5K = 0.5$
4	$8K = 0.8$
5	$K^2 + 8K = 0.81$
6	$3K^2 + 8K = 0.83$
7	$10K^2 + 9K = 1$

7/6/22

Probability density function: If $f(x)$ is a continuous function then the probability of the variable $x \in (a,b)$ is defined as $P(a<x<b) = \int_a^b f(x)dx$ here $f(x)$ is known as probability density function.

Cummulative distributive function: Cummulative distributive function of a continuous random variable is defined as

$F(x) = P(X \le x)$

NOTE: If $F(x)$ is a probability distribution function and $f(x)$ is a probability density function then i) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

ii) $0 \le f(x) \le 1$

iii) $\frac{d}{dx} F(x) = f(x)$

Mean (or) Expectation of continuous random variable: If x is a continuous random variable then mean or expectation is defined as $\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$

Variance of a continuous random variable: Variance of a continuous random variable is defined as $\sigma^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - \mu^2$

$$\boxed{\sigma^2 = E(X^2) - (E(X))^2}$$

Standard deviation: The positive square root of variance is known as standard deviation. $\sigma = \sqrt{E(x^2) - \{E(x)\}^2}$

$$\Rightarrow \sigma = \sqrt{\int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2}$$

Median: Median is the value of 'x' which divides the total distribution into two equal parts if $x \in (a,b)$ and 'm' is the median then it is obtained by solving $\int_a^m f(x)\,dx = \int_m^b f(x)\,dx = \frac{1}{2}$

Mode of a continuous random variable:

Mode is the value of 'x' for which f(x) is maximum. It is obtained by taking $f'(x)=0$, where $f''(x)<0$.

1. If a random variable x has the following probability density function $f(x) = 2e^{-2x}$ if $x>0$ and $f(x)=0$ if $x \le 0$ then find the probability that it will take on a value

(i) between 1 & 3

(ii) ~~less~~ greater than >0.5

Sol: (i) We know that $P(a<x<b) = \int_a^b f(x)\,dx$

$$P(1<x<3) = \int_1^3 f(x)\,dx$$

$$= \int_1^3 2e^{-2x}\,dx$$

$$= -\left(\cancel{2}\frac{e^{-2x}}{\cancel{2}}\right)_1^3$$

$$= -(e^{-6} - e^{-2})$$

$$= e^{-2} - e^{-6}$$

(ii) $P(x>0.5) = \int_{0.5}^{\infty} f(x)dx$

$$= \int_{0.5}^{\infty} 2e^{-2x} dx$$

$$= -\left(\cancel{2}\frac{e^{-2x}}{\cancel{2}}\right)_{0.5}^{\infty}$$

$$= -(e^{-\infty} - e^{-1})$$

$$= -(0 - e^{-1})$$

$$= \frac{1}{e}$$

2. The probability density function of a random variable 'x' is given by

$f(x) = K(1-x^2)$ where $0<x<1$

$= 0$ otherwise

then find (i) value of K

(ii) Probability of the variable lies between 0.1 & 0.2

(iii) >0.5

Sol: (i) WKT $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{\infty} f(x)dx = 1$$

$$0 + \int_{0}^{1} K(1-x^2)dx + 0 = 1$$

$$= -\left(\cancel{2}\frac{e^{-2x}}{\cancel{2}}\right)_1^3$$

$$= -(e^{-6} - e^{-2})$$

$$= e^{-2} - e^{-6}$$

(ii) $P(x > 0.5) = \int_{0.5}^{\infty} f(x)\,dx$

$$= \int_{0.5}^{\infty} 2e^{-2x}\,dx$$

$$= -\left(\cancel{2}\frac{e^{-2x}}{\cancel{2}}\right)_{0.5}^{\infty}$$

$$= -(e^{-\infty} - e^{-1})$$

$$= -(0 - e^{-1})$$

$$= \frac{1}{e}$$

2. The probability density function of a random variable 'x' is given by

$f(x) = K(1-x^2)$ where $0 < x < 1$

$= 0$ otherwise

then find (i) value of K

(ii) Probability of the variable lies between 0.1 & 0.2

(iii) > 0.5

Sol: (i) WKT $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx = 1$$

$$0 + \int_{0}^{1} K(1-x^2)\,dx + 0 = 1$$

$$K\left(x-\frac{x^3}{3}\right)_0^1=1$$

$$K\left(1-\frac{1}{3}-0\right)=1$$

$$K\left(\frac{2}{3}\right)=1$$

$$K=\frac{3}{2}$$

$\therefore f(x)=\frac{3}{2}(1-x^2)$ if $0<x<1$

$=0$ otherwise

(ii) $P(0.1<x<0.2)=\int_{0.1}^{0.2} f(x)\,dx$

$$=\int_{0.1}^{0.2}\frac{3}{2}(1-x^2)\,dx$$

$$=\frac{3}{2}\left(x-\frac{x^3}{3}\right)_{0.1}^{0.2}$$

$$=\frac{3}{2}\left[0.2-\frac{(0.2)^3}{3}-\left(0.1-\frac{(0.1)^3}{3}\right)\right]$$

$=0.14$

(iii) $P(x>0.5)=\int_{0.5}^{\infty} f(x)\,dx$

$$=\int_{0.5}^{1} f(x)\,dx+\int_{1}^{\infty} f(x)\,dx$$

$$=\int_{0.5}^{1}\frac{3}{2}(1-x^2)\,dx+0$$

$$=\frac{3}{2}\left(x-\frac{x^3}{3}\right)_{0.5}^{1}$$

$$=\frac{3}{2}\left(1-\frac{1}{3}-\left(0.5-\frac{(0.5)^3}{3}\right)\right)=0.34$$

3. If a probability density function of a continuous random variable is given by $f(x) = Ke^{-|x|}$ where $-\infty < x < \infty$ then find

(i) value of K

(ii) mean

(iii) variance

(iv) the probability of the variable which lies between 0 & 4

(i) wkt $|x| = x$ if $x > 0$

$= -x$ if $x < 0$

wkt $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{0} f(x)dx + \int_{0}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{0} Ke^{x}dx + \int_{0}^{\infty} Ke^{-x}dx = 1$$

$$K(e^{x})_{-\infty}^{0} - K(e^{-x})_{0}^{\infty} = 1$$

$$K(1 - e^{-\infty}) - K(e^{-\infty} - 1) = 1$$

$$K(1-0) - K(0-1) = 1$$

$$K + K = 1$$

$$K = \frac{1}{2}$$

$\therefore f(x) = \frac{1}{2}e^{-|x|}$ if $-\infty < x < \infty$

(ii) wkt $\mu = \int_{-\infty}^{\infty} x \cdot f(x)dx$

$$= \int_{-\infty}^{0} x \cdot f(x)dx + \int_{0}^{\infty} xf(x)dx$$

$$\int_{-\infty}^{0} x\cdot\frac{1}{2}e^{x}dx + \int_{0}^{\infty} x\cdot\frac{1}{2}e^{-x}dx$$

$$\frac{1}{2}\int_{-\infty}^{0} xe^{x}dx + \frac{1}{2}\int_{0}^{\infty} x\cdot e^{-x}dx$$

$$\boxed{\because \int uv\,dx = u\int v\,dx - \int\left[u'\int v\,dx\right]dx}$$

$$\frac{1}{2}\left[xe^{x} - \int e^{x}dx\right]_{-\infty}^{0} + \frac{1}{2}\left[-xe^{-x} + \int e^{-x}dx\right]_{0}^{\infty}$$

$$\frac{1}{2}\left[xe^{x} - e^{x}\right]_{-\infty}^{0} + \frac{1}{2}\left[-xe^{-x} - e^{-x}\right]_{0}^{\infty}$$

$$\frac{1}{2}[-1-0] + \frac{1}{2}[0-0+1]$$

$$\cancel{-\frac{1}{2}} + \cancel{\frac{1}{2}}$$

$\mu = 0$

(iii) WKT $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$

$$\Rightarrow \int_{-\infty}^{0} x^2 f(x)dx + \int_{0}^{\infty} x^2 f(x)dx - 0$$

$$\Rightarrow \int_{-\infty}^{0} x^2\cdot\frac{1}{2}e^{x}dx + \int_{0}^{\infty} x^2\cdot\frac{1}{2}e^{-x}dx$$

$$= \frac{1}{2}\int_{-\infty}^{0} x^2 e^{x}dx + \frac{1}{2}\int_{0}^{\infty} x^2 e^{-x}dx$$

$$= \frac{1}{2}\left[x^2\int e^{x}dx - \int\left[2x\int e^{x}dx\right]dx\right]_{-\infty}^{0} +$$

$$\frac{1}{2}\left[x^2\int e^{-x}dx - \int\left[2x\int e^{-x}dx\right]dx\right]_{0}^{\infty}$$

$= \frac{1}{2}\left[x^2 e^{x} - 2\int (x e^{x}\, dx)\, dx\right]_{-\infty}^{0} +$

$$\frac{1}{2}\left[-x^2 e^{-x} - \int -2x e^{-x}\, dx\right]_0^{\infty}$$

$= \frac{1}{2}\left[x^2 e^{x}\right]_{-\infty}^{0} - \cancel{2}\left[-\frac{1}{\cancel{2}}\right] + \frac{1}{2}\left[-x^2 e^{-x}\right]_0^{\infty} + \cancel{2}\left(\frac{1}{\cancel{2}}\right)$

$= 0 + 1 + 0 + 1$

$= 2$

(ii) $P(0 < x < 4) = \int_0^4 f(x)\, dx$

$\int_0^4 \frac{1}{2} e^{-x}\, dx = \frac{1}{2}\left(e^{-x}\right)_0^4$

$-\frac{1}{2}\left(e^{-4} - 1\right) = \frac{1}{2}\left(1 - e^{-4}\right)$

4. The probability density function of a random variable is given by $f(x) = \frac{1}{2}\sin x$ where $0 \le x \le \pi$ and $f(x) = 0$ in other cases then find

(i) Mean (ii) Median (iii) Mode (iv) probability of the variable which lies b/w 0 & $\frac{\pi}{2}$.

(i) wkt mean of random variable

$$\mu = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

$$= \int_{-\infty}^{0} x \cdot f(x) dx + \int_{0}^{\pi} x \cdot f(x) dx + \int_{\pi}^{\infty} x f(x) dx$$

$$= 0 + \int_{0}^{\pi} x \cdot \frac{1}{2} \sin x dx + 0$$

$$= \frac{1}{2} \int_{0}^{\pi} x \cdot \sin x dx$$

$$\boxed{\because \int uv dx = u \int v dx - \int (u' \int v dx) dx}$$

$$= \frac{1}{2}\left[-x\cos x + \int_0^{\pi} 1 \cos x dx\right]_0^{\pi}$$

$$= \frac{1}{2}\left[-x\cos x + \sin x\right]_0^{\pi}$$

$$= \frac{1}{2}[\pi - 0] = \frac{\pi}{2}$$

(ii) wkt median is the point which divides the total distribution into 2 equal parts

If 'm' is the median then we have

$$\int_0^m f(x)\,dx = \int_m^\pi f(x)\,dx = \frac{1}{2}$$

$$\int_0^m f(x)\,dx = \frac{1}{2}$$

$$\int_0^m \frac{1}{2}\sin x = \frac{1}{2}$$

$$\cancel{\frac{1}{2}}\int_0^m \sin x = \frac{1}{\cancel{2}}$$

$$-[\cos x]_0^m = 1$$

$$-[\cos m - 1] = 1$$

$$1 + \cos m = 1$$

$$\cos m = 0$$

$$m = \frac{\pi}{2}$$

(vii) wkt mode is the value of x for which $f(x)$ is maximum.

It is obtained by taking $f'(x) = 0$ where $f''(x) < 0$.

wkt $f(x) = \frac{1}{2}\sin x \Rightarrow f'(x) = 0$

$$\Rightarrow \frac{1}{2}\cos x = 0$$

$$\cos x = 0$$

$$\cos x = \cos\frac{\pi}{2}$$

$$x = \frac{\pi}{2}$$

here $f''(x) = -\frac{1}{2}\sin x$

$$f''\left(\frac{\pi}{2}\right) = -\frac{1}{2}\sin\left(\frac{\pi}{2}\right)$$

$$= \frac{-1}{2} < 0$$

i.e. mode $= \frac{\pi}{2}$

(iv) $P\left(0 < x < \frac{\pi}{2}\right) = \int_0^{\pi/2} f(x)\,dx$

$$= \int_0^{\pi/2} \frac{1}{2}\sin x\,dx$$

$$= \frac{1}{2}\left(-\cos x\right)_0^{\pi/2}$$

$$= -\frac{1}{2}\left[\cos\frac{\pi}{2} - \cos(0)\right]$$

$$= -\frac{1}{2}\left[0 - 1\right]$$

$$= -\frac{1}{2}(-1)$$

$$= \frac{1}{2}$$

Probability Distributions: There are 2 types of probability distributions
1. Discrete probability distribution
2. Continuous Probability distribution

→ Binomial distribution, Possion distribution are discrete probability distributions in which random variable will take finite no. of values.

→ Normal distribution is continuous Probability distribution in which random variable will take infinite no. of values

Binomial distribution: A random variable X is said to have a binomial distribution if it takes only non-negative values & its probability distribution is given by

$p(X=r) = p(r) = {}^nC_r\, p^r \cdot q^{n-r} \quad \forall\, r = 0,1,2,3, \ldots n.$

Here n is no. of independent trials,

p is probability of success &

q is probability of failure

Mean of Binomial distribution:

Proof: We know that mean, $\mu = \sum_{r=0}^{n} r \cdot p(r)$

$$\Rightarrow \sum_{r=0}^{n} r \cdot {}^nC_r\, p^r q^{n-r}$$

$$\Rightarrow 0 + {}^nC_1\, p^1 q^{n-1} + 2 \cdot {}^nC_2\, p^2 q^{n-2} + \cdots + n \cdot {}^nC_n\, p^n$$

$$\Rightarrow n\, p\, q^{n-1} + 2 \cdot \frac{n(n-1)}{2!}\, p^2 \cdot q^{n-2} + \cdots + n p^n$$

$$\Rightarrow np\left[{}^{n-1}C_0\, q^{n-1} + {}^{n-1}C_1\, p^1 q^{(n-1)-1} + \cdots + {}^{n-1}C_{n-1}\, p^{n-1}\right]$$

$$\Rightarrow np[q+p]^{n-1}$$

$$\boxed{\mu = np}$$

Variance of the binomial distribution:

Proof: we know that variance

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

$$= \sum_{r=0}^{n} (r(r-1)+r) P(r) - \mu^2$$

$$= \sum_{r=0}^{n} r(r-1)p(r) + \sum_{r=0}^{n} r.P(r) - \mu^2$$

$$= \sum_{r=0}^{n} r(r-1) nC_r p^r q^{n-r} + \mu - \mu^2$$

$$= 2.1.nC_2 p^2 q^{n-2} + 3.2\, nC_3 p^3 q^{n-3} + \cdots + n(n-1) nC_n P^n + \mu - \mu^2$$

$$= 2\frac{n(n-1)}{2} p^2 q^{n-2} + 6.\frac{n(n-1)(n-2)}{6} p^3 q^{n-3} + \cdots + n(n-1)p^n + np - n^2p^2$$

$$= n(n-1)p^2 \left[{}^{n-2}C_0 q^{n-2} + {}^{n-2}C_1 p^1 q^{(n-2)-1} + \cdots + {}^{n-2}C_{n-2} p^{n-2}\right] + np - n^2p^2$$

$$= n(n-1)P^2 [q+p]^{n-2} + np - n^2p^2$$

$$\sigma^2 = n(n-1)p^2 + np - n^2p^2$$

$$= \cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2}$$

$$= np(1-p)$$

$$= np(q)$$

$$\sigma^2 = npq$$

standard deviation of Binomial distribution:

we know that positive square root of variance is SD.

$\therefore \sigma = \sqrt{npq}$

Mode of the Binomial distribution:

case-(i): if $(n+1)p$ is not an integer then mode of the Binomial distribution is Integral part of $[(n+1)p]$

case-(ii): if $(n+1)p$ is an integer then mode of the Binomial distribution is

$(n+1)p, (n+1)p-1$

Binomial frequency distribution: If n independent trials performs one experiment such that p is the probability of success, q is the probability of failure and this ~~experiment is repeated N times then Binomial frequency distribution~~ the Binomial frequency distribution is defined as $N[q+p]^n$ here $N = \Sigma f$

1. If fair coin is tossed 6 times then find the probability of getting 4 heads

Sol: Given that $n=6$, $r=4$

let p be the probability of getting head $\Rightarrow p = \frac{1}{2}$

$q = 1-p$

$q = \frac{1}{2}$

WKT $P(X=r) = P(r) = nC_r\, p^r . q^{n-r}$

$P(4) = 6C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2$

$= \frac{6\times5\times4!}{2!\,4!} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2$

$= 15 \left(\frac{1}{2}\right)^6$

$= 15 \left(\frac{1}{64}\right)$

2. A die is thrown 6 times & getting an even number is success then find the probability of getting

(i) atleast one success

(ii) less than or equal to 3 successes

(iii) 4 successes

Given that $\boxed{n=6}$

let p is the probability of getting an even number $\Rightarrow \frac{3}{6}$

$\Rightarrow \frac{1}{2}$

$\boxed{p=\frac{1}{2}}$

$q = 1-p$

$\boxed{q=\frac{1}{2}}$

(i) $P(r\geq1) = 1-P(r<1)$

$= 1-P(0)$

$= 1 - \left[6C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6 \right]$

$= 1 - \left[\left(\frac{1}{2}\right)^6 \right]$

$= 1 - \frac{1}{64}$

$= \frac{63}{64}$

(ii) $P(r \leq 3) = P(0) + P(1) + P(2) + P(3)$

$= 6C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6 + 6C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^5 + 6C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 +$
$6C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^3$

$= \left(\frac{1}{2}\right)^6 \left[6C_0 + 6C_1 + 6C_2 + 6C_3 \right]$

$= \left(\frac{1}{2}\right)^6 \left[1 + 6 + 15 + 20 \right]$

$= \frac{1}{64} [42]$

$= \frac{21}{32}$

(iii) $P(r=4) = 6C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2$

$= \frac{6!}{2!\,4!} \left(\frac{1}{2}\right)^6$

$= \frac{\overset{3}{\cancel{6}} \times 5 \times \cancel{4!}}{\cancel{2!} \times \cancel{4!}} \left(\frac{1}{2}\right)^6$

$= \frac{15}{64}$

3. Determine mode of the Binomial distribution for which mean is 4 and variance is 3

Sol: Given that $\mu = np = 4$

$$\sigma^2 = npq = 3$$

$$\frac{npq}{np} = \frac{3}{4}$$

$$\boxed{q = \frac{3}{4}}$$

$$p = 1 - q$$

$$p = 1 - \frac{3}{4}$$

$$\boxed{p = \frac{1}{4}}$$

We know that $np = 4$

$$n\left(\frac{1}{4}\right) = 4$$

$$\boxed{n = 16}$$

Here $(n+1)p = (17)\left(\frac{1}{4}\right) = 4.25$

$\therefore$ Mode = Integral part of (4.25)

$= 4$

Q. The mean and variance of a binomial distribution is given by 16 and 8 respectively then find probability (i) $P(r \geq 1)$ (ii) $P(r > 2)$

Sol:

$\mu = 16$	$npq = 8$
$np = 16$	$\boxed{q = \frac{1}{2}}$

$P = 1 - q$

$P = 1 - \frac{1}{2}$

$\boxed{P = \frac{1}{2}}$

$nP = 16$

$n\left(\frac{1}{2}\right) = 16$

$\boxed{n = 32}$

(i) $P(r \geq 1) = 1 - P(r < 1)$

$= 1 - P(0)$

$= 1 - \left[{}^{32}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{32}\right]$

$= 1 - \left(\frac{1}{2}\right)^{32}$

(ii) $P(r > 2) = 1 - P(r \leq 2)$

$= 1 - [P(0) + P(1) + P(2)]$

$= 1 - \left[{}^{32}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{32} + {}^{32}C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{31} + {}^{32}C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{30}\right]$

$= 1 - \left(\frac{1}{2}\right)^{32} \left[{}^{32}C_0 + {}^{32}C_1 + {}^{32}C_2\right]$

$= 1 - \left(\frac{1}{2}\right)^{32} \left[\cancel{32}\ 1 + 32 + 496\right]$

$= 1 - \left(\frac{1}{2}\right)^{32} [529]$

5. Fit a Binomial frequency distribution to the following data

x	0	1	2	3	4	5
f	2	14	20	34	22	8

We know that binomial frequency distribution is $N(q+p)^n$

Given that $\boxed{n=5}$

$$\boxed{N=\Sigma f=100}$$

wkt $\mu = np = \frac{\Sigma f x}{\Sigma f}$

$$np = \frac{0+14+40+102+88+40}{100}$$

$$np = \frac{284}{100}$$

$$np = 2.84$$

$$5p = 2.84$$

$$p = \frac{2.84}{5}$$

$$\boxed{p = 0.57}$$

$$q = 1-p$$

$$= 1-0.5$$

$$\boxed{q = 0.43}$$

∴ The Binomial frequency distribution is

$$= 100(0.43+0.57)^5$$

$$100(1.00)^5$$

$$= 100\left[{}^5C_0(0.43)^5 + {}^5C_1(0.43)^4(0.57)^1 + {}^5C_2(0.43)^3(0.57)^2 + {}^5C_3(0.43)^2(0.57)^3 + {}^5C_4(0.43)^1(0.57)^4 + {}^5C_5(0.57)^5\right]$$

∴ The expected frequencies are

$$= 100[0.016 + 0.105 + 0.267 + 0.339 + 0.216 + 0.055]$$

$$= 1.6,\ 10.5,\ 26.7,\ 33.9,\ 21.6,\ 5.5$$

x	f	Expected frequency
0	2	2
1	14	10
2	20	27
3	34	34
4	22	22
5	8	5