Unit-3

Correlation coefficient : In a bivariant if the change in one variable affects the change in other variable, the variables are said to be correlated.

Note -
1. Correlation analysis deals with association b/w 2 or more variables.
2. The degree of relⁿ ship b/w the variable under consideration is measured through the correlation analysis.
3. The measure of correlation called the correlation coefficient or correlation index summarises in one figure that is direction & degree of correlation.

$r = -1$	$r = 0$	$r = 1$
perfect negative correlation coefficient	no relⁿ	perfect positive correlation coefficient

Types of correlation

- Positive correlation (direct correlation)
  Both variables vary in the same direction.
  $r = 0$ to 1
- Negative correlation (inverse correlation)
  Both variables vary in opposite direction
  $r = -1$ to 0

Methods of finding correlation coefficient -

1. Karl Pearson's coefficient of correlation (Product moment correlation coefficient)
2. Spear man Rank correlation coefficient
3. Scatter or dot diagram method
4. Coefficient of concurrent deviation.

(1) Karl Pearson's Method -

$r = \dfrac{cov(x,y)}{\sigma_x \sigma_y}$; $cov(x,y) = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$,

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

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

Q. find correlation coefficient of the following data -

x	y	x^2	y^2	xy
9	15	81	225	135
8	16	64	256	128
7	14	49	196	98
6	13	36	169	96
5	11	25	121	55
4	12	16	144	48
3	10	9	100	30
2	8	4	64	16
1	9	1	81	1
$\Sigma x = 45$	108	285	1356	597

$$r = \frac{9 \times 597 - 45 \times 108}{\sqrt{9 \times 285 - 2025}\ \sqrt{9 \times 1356 - (108)^2}} = \underline{0.95}$$ (strongly +ve)

Q. find corr. coeff. of the given table -

x	y	x^2	y^2	xy
10	18	100	324	180
14	12	196	144	168
18	24	324	576	432
22	6	484	36	132
26	30	676	900	780
30	36	900	1296	1080
120	126	2680	3276	2772

$r = \underline{0.6}$

Positive Correlation

Q: find corr. coeff. of given data –

x	y	x^2	y^2	xy
100	30	10000	900	3000
200	50	40000	2500	10000
300	60	90000	3600	18000
400	80	160000	6400	32000
500	100	250000	10000	50000
600	110	360000	12100	66000
700	130	490000	16900	91000
2800	560	1400000	52400	270000

$$r = \frac{7 \times 270000 - 2800 \times 560}{\sqrt{7 \times 1400000 - (2800)^2}\sqrt{7 \times 52400 - (560)^2}} = 0.9$$

Q. find corr. coeff. b/w size group & defect in quality from the given data –

size group	15-16	16-17	17-18	18-19	19-20	20-21
No. of items	200	270	340	360	400	300
No. of defective items	150	162	170	180	100	114
% of defective items	75	60	50	50	45	38

$\frac{150}{200} \times 100$

x	y	x^2	y^2	xy
15.5	75			
16.5	60			
17.5	50			
18.5	50			
19.5	45			
20.5	38			
108	318	1961.5	17694	5609

$r = -0.95$

Regression Analysis

- It is the technique for measuring or estimating the relationship among variables.
- Regression analysis ~~de~~ provides estimate values of the dependent variable from the values of independent variables.
- The device used to accomplish this estimation procedure is the regression line.
- The regression line describes the average relⁿ ship existing b/w x & y.

Difference b/w regression analysis & correlation analysis -

Correlation analysis:

1. Corr. coefficient is the measure of degree of covariability b/w x & y.
2. In correlation analysis, we can't say that one variable is the cause & other is the effect.
3. In correlation analysis both r_{xy} & r_{yx} are symmetric ($r_{yx} = r_{xy}$)
4. There may be nonsense correlation which has no practical relevance such as increase in income & increase in weight.
5. Correlation coefficient is independent of change of origin & change of scale.

Regression analysis -

1. Objective of regression analysis is to study the nature of relationship b/w variables.
2. In regression analysis, it is possible to study cause & effect relationship.
3. In regression analysis, b_{xy} & b_{yx} are not symmetric ($b_{xy} \neq b_{yx}$)
4. There is nothing like non sense regression.
5. Regression coefficients are independent of change of origin but not of scale.

Similarity - Correlation coefficient(r) & regression coefficient always have same sign.

Regression lines -

① Regression line of y on x - It is used to calculate y for given x.

$$(y-\bar{y}) = b_{yx}(x-\bar{x}) \quad \text{where, } \bar{x} = \frac{\Sigma x_i}{n}, \ \bar{y} = \frac{\Sigma y_i}{n}$$

b_{yx} = regression coefficient of y on x

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

where x is independent ~~of~~ variable, y is dependent variable and slope of line (coefficient of x) = b_{yx}.

② Regression line of x on y -

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

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

slope of line (coeff. of x) $= \dfrac{1}{b_{xy}}$

Q. From the given table obtain - (i) both regression lines (ii) Calculate corr. coeff. (iii) estimate value of y for x = 6.2

x	y	x^2	y^2	xy
1	9			
2	8			
3	10			
4	12			
5	11			
6	13			
7	14			
8	16			
9	15			
45	108	285	1356	597

$$b_{xy} = \frac{9 \times 597 - 45 \times 108}{9 \times 1356 - (108)^2} = 0.95$$

$b_{yx} = 0.95$

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

$$y = 13.14$$

Correlation coefficient

$$= \sqrt{b_{xy} \times b_{yx}}$$

$$= \sqrt{0.95 \times 0.95}$$

$$= 0.95$$

Q.

x	y	x^2	y^2	xy
1	1			
3	2			
4	4			
6	4			
8	5			
9	7			
11	8			
14	9			
56	40	524	256	364

$\bar{x} = 7$ $\quad \bar{y} = 5$

$$b_{xy} = \frac{8 \times 364 - 56 \times 40}{8 \times 256 - (40)^2} = 1.5$$

$$b_{yx} = 0.636$$

$$r = \sqrt{b_{xy}\, b_{yx}} = 0.976$$

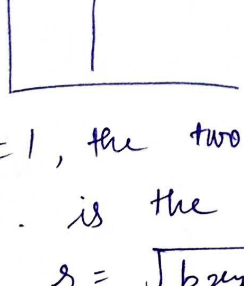
(2 marks)
Note -

① $b_{yx} = \dfrac{r \sigma_y}{\sigma_x}$

② $b_{xy} = \dfrac{r \sigma_x}{\sigma_y}$

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

③ If $r = 0$, then the two lines of regression become $y = \bar{y}$ & $x = \bar{x}$ which are two straight lines parallel to x axis & y axis respectively & passing through ~~this~~ their means $\bar{y}$ and $\bar{x}$ and they are perpendicular to each other.



④ If $r = \pm 1$, the two lines of regression will coincide.

⑤ Corr. coeff. is the geometric mean b/w the regression coeff.

$$r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{\frac{r\sigma_x}{\sigma_y} \times \frac{r\sigma_y}{\sigma_x}} = \sqrt{r^2} = r$$

⑥ If one of the regression coefficient is greater than unity, then the other must be less than unity.

V. Imp Q. If θ is the acute angle b/w the two regression lines in case of two variables x & y. Show that

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

Explain the significance of the formula when $r=0$ & $r=\pm 1$

=) Regression line y on x –

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

$$m_1 = b_{yx} = \frac{r\sigma_y}{\sigma_x}$$

Regression line x on y –

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

$$m_2 = \frac{1}{b_{yx}} = \frac{\sigma_y}{r\sigma_x}$$

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

$$= \pm \frac{\frac{\sigma_y}{r\sigma_x} - \frac{r\sigma_y}{\sigma_x}}{1+\frac{\sigma_y}{r\sigma_x} \times \frac{r\sigma_y}{\sigma_x}}$$

$$= \pm \frac{(1-r^2)}{r} \frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}$$

Since, $r^2 \le 1$ and σ_x and σ_y are +ve and θ is acute angle

=) $\tan\theta$ = +ve

Case 1 - when $r=0$, $\tan\theta = \infty$ =) $\boxed{\theta = \frac{\pi}{2}}$

The lines of regression are $\perp$ to each other.

Case 2 - when $r=\pm 1$, $\tan\theta = 0$ =) $\boxed{\theta = 0, \pi}$

The lines of regression coincide & there is perfect correlation b/w x & y.

Spare Man Rank Correlation Method :-

$$r = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}$$

Q. The marks scored by some students in Maths & Physics are given below :-

Marks in Maths (x)	Marks in Physics (y)	R_x	R_y	$d = R_x - R_y$	d^2
10	30	9	9	0	0
15	42	5	3	2	4
12	45	8	2	6	36
17	46	3	1	2	4
13	33	7	8	-1	1
16	34	4	7	-3	9
24	40	1	4	-3	9
14	35	6	6	0	0
22	39	2	5	-3	9

$\Sigma d^2 = 72$

$$r = 1 - \frac{6 \times 72}{9 \times 80} = 0.4$$

Q. The competitors in a beauty contest got marks of three judges in the given table -

I judge	II	III	d_{12}	d_{23}	d_{13}	d_{12}^2	d_{23}^2	d_{13}^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	-4	2	36	16	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
						208	~~60~~ 214	~~214~~ 60

$$r_{12} = 1 - \frac{6 \Sigma d_{12}^2}{n(n^2-1)} = 1 - \frac{208 \times 6}{10(10^2-1)} = -0.26$$

$$r_{23} = \frac{1 - 6\Sigma d_{23}^2}{n(n^2-1)} = \frac{1 - 6 \times 214}{10(10^2-1)} = -0.29$$

$$r_{31} = \frac{1 - 6\Sigma d_{31}^2}{n(n^2-1)} = \frac{1 - 6 \times 60}{10(10^2-1)} \doteq 0.64$$

Tied Rank – (Ranks are repeated)

$$\boxed{r = 1 - \frac{6(\Sigma d^2 + F)}{n(n^2-1)} \qquad F = \frac{m(m^2-1)}{12}}$$

X	Y	Rx	Ry	$d_{xy} = R_x - R_y$	d_{xy}^2
(4) 68	62 (5)	4	5	−1	1
(5) 64	58 (7)	6	7	−1	1
(2) 75	68 (3)	2.5	3.5	−1	1
(9) 50	45 (10)	9	10	−1	1
(6) 64	81 (1)	6	1	5	25
(1) 80	60 (6)	1	6	−5	25
(3) 75	68 (4)	2.5	3.5	−1	1
(10) 40	48 (9)	10	9	1	1
(8) 55	50 (8)	8	8	0	0
(7) 64	70 (2)	6	2	4	16
					$\Sigma d_{xy}^2 = 72$

$$F = f_{75} + f_{64} + f_{68}$$

$$= \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12}$$

$$= \frac{2 \times 3}{12} + \frac{3 \times 8}{12} + \frac{2 \times 3}{12}$$

$$F = 3$$

$$r = 1 - \frac{6(72+3)}{10(10^2-1)} = \frac{1 - 6 \times 75}{10 \times 99} = \underline{\underline{0.585}}$$

In a partially destroyed laboratory record of an analysis of a correlation data the following results are :-

variance of $x = 9$ $(\sigma_x^2 = 9)$

Regression eq^n : $8x - 10y + 66 = 0$, $40x - 18y - 214 = 0$

what were : (a) the means values of x & y.
(b) the standard deviation of y & the coefficient of correlation b/w x & y.

=) $\sigma_y = ?$, $r = ?$

Since both the lines of regression pass through the point $(\bar{x}, \bar{y})$

$$8\bar{x} - 10\bar{y} + 66 = 0$$
$$40\bar{x} - 18\bar{y} - 214 = 0$$
$$\bar{x} = +13 \quad , \bar{y} = +17$$

$$8x - 10y + 66 = 0$$
$$y = 0.8x + 6.6 \qquad \Rightarrow b_{yx} = 0.8$$

$$40x - 18y - 214 = 0$$
$$x = \frac{18}{40}x + \frac{214}{40}$$

$$b_{xy} = 0.45$$

$$r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{0.45 \times 0.8} = 0.6$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$\underline{\underline{\sigma_y = 4}}$$

Curve fitting -

By the method of least square -

① Fitting ~~in~~ a straight line -

$$y = a + bx \quad \text{——— ①}$$

$$\Sigma y = \Sigma a + \Sigma bx$$

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

on multiplying both sides of eq. 1 by Σx -

$$\boxed{\Sigma xy = a\Sigma x + b\Sigma x^2}$$

② fitting a second degree parabola -

$$y = a + bx + cx^2 \quad \text{—— ②}$$

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

mul. both sides of eq. 2 by Σx -

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

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

③ fitting of an exponential curve -

$$y = ae^{bx}$$

Taking log on both sides -

$$\log_{10} y = \log_{10} a + \underbrace{b \log_{10} e}\, . x$$

$$Y = A + BX$$

$$\Sigma Y = nA + B\Sigma X$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

④ fitting of curve $y = ax^b$

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

$$Y = A + B \log_{10} x$$

$$\Sigma Y = nA + B\Sigma X$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

⑤ fitting of curve $y =$

Q. Fit a curve $y = a + bx + cx^2$ for following data using least square method

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

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

$$12.9 = 5a + b \times 10 + c \times 30$$

$$12.9 = 5a + 10b + 30c \quad \text{——— (1)}$$

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

$$37.1 = 10a + 30b + 100c \quad \text{——— (2)}$$

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

$$130.3 = 30a + 100b + 354 \quad \text{——— (3)}$$

$$a = 1.42, \quad b = -1.07, \quad c = 0.55$$

Q. Determine the constants a and b by the method of least squares such that $y = ae^{bx}$. fix the following data -

x	y	$Y = \log_{10} y$	XY	X^2
2	4.077	0.610	1.2	
4	11.084	1.044	4.176	
6	30.128	1.478	8.86	
8	81.897	1.913	15.30	
10	222.62	2.347	23.47	
30		7.39	53.038	220

$$\Sigma Y = nA + B\,\Sigma X$$

$$7.39 = 5A + 30B \quad \text{——— (1)}$$

$$\Sigma XY = A\,\Sigma X + B\,\Sigma X^2$$

$$53.038 = 30A + 220B \quad \text{——— (2)}$$

$$A = 0.1755, \quad B = 0.217$$

Q. ⑤ fitting of the curve $y = \cancel{ax^b}\ ab^x$

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

$$Y = A + BX$$

$$Y = \log_{10} y$$
$$A = \log_{10} a$$
$$B = \log_{10} b$$
$$X = x$$

$$\Sigma Y = nA + B\Sigma X$$
$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

⑥ fitting of the curve $PV^\gamma = k$

$$\log_{10} P + \gamma \log V = \log k$$

$$\log_{10} P = -\gamma \log V + \log k$$

$$V^\gamma = \frac{k}{P} \Rightarrow V = \left(\frac{k}{P}\right)^{1/\gamma}$$

$$V = \frac{k^{1/\gamma}}{P^{1/\gamma}}$$

$$\log V = \log k^{1/\gamma} - \log P^{1/\gamma}$$

$$\log V = \frac{1}{\gamma}\log k - \frac{1}{\gamma}\log P$$

$$Y = A + BX$$
$$Y = \log V$$
$$A = \frac{1}{\gamma}\log k$$
$$B = -\frac{1}{\gamma}, \quad X = \log P$$

Q. fit the curve $PV^\gamma = k$ to the following data —

P	V	Y	X	X^2	XY
0.5	1620	3.209	-0.301		
1	1000	3	0		
1.5	750	2.875	0.176		
2	620	2.792	0.301		
2.5	520	2.716	0.397		
3	460	2.662	0.417		
		17.254	1.05	0.597	2.728

$\Sigma Y = nA + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$A = 3.0013$ or $\underline{2.999}$

$B = -0.716$ or $\underline{-0.705}$

$B = -\frac{1}{\gamma}$

$\cancel{\gamma = \frac{-1}{-0.716} = 1.396}$

$\gamma = \frac{-1}{-0.705} = \underline{1.418}$

$2.999 = 1.418 \log k$

$\cancel{\log k = \frac{2.999}{1.418} = 2.1149}$

$\log k = 4.252$

$k = 17864.87$

Q. independent / dependent

V	P	Y = log P	X = log V
50	64.7	1.8109	
60	51.3	1.710	
70	40.5	1.607	
90	25.9	1.413	
100	78	1.892	

$PV^{\gamma} = k$

$PV^{\gamma} = k$

$P = kV^{-\gamma}$

$\log P = \log k - \gamma \log V$

$Y = \log P$

$A = \log k$

$B = -\gamma$

$X = \log V$

$Y = A + BX$

$\Sigma Y = nA + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

Multiple Linear Regression -

The dependent variable is a function of 2 or more linear or non-linear independent variables. Consider such a linear $f x^n$ as -

$$y = a + bx + cz$$

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

$$\Sigma yx = a\Sigma x + b\Sigma x^2 + c\Sigma zx$$

$$\Sigma yz = a\Sigma z + b\Sigma xz + c\Sigma z^2$$

Q. Obtain a regression plane by using multiple linear regression to fit the data given below :-

x	z	y	x^2	z^2	yx	yz	zx
1	0	12	1	0	12	0	0
2	1	18	4	1	36	18	2
3	2	24	9	4	72	48	6
4	3	30	16	9	120	90	12
10	6	84	30	14	240	156	20

$$84 = 4a + 10b + 6c$$

$$240 = 10a + 30b + 20c$$

$$156 = 6a + 20b + 14c$$

Math error

Q. Find multiple linear regression of x_1 on x_2 and x_3 from the data relating to 3 variables

$$x_1 = a + bx_2 + cx_3$$

$$\Sigma x_1 = na + b\Sigma x_2 + c\Sigma x_3$$

$$\Sigma x_1 x_2 = a\Sigma x_2 + b\Sigma x_2^2 + c\Sigma x_3 x_2$$

$$\Sigma x_1 x_3 = a\Sigma x_3 + b\Sigma x_2 x_3 + c\Sigma x_3^2$$

x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	$x_2 x_3$	$x_1 x_3$
4	15	30	16	225	900	450	
6	12	24	36	144	576	288	
7	8	20	49	64	400	160	
9	6	14	81	36	196	84	
13	4	10	169	16	100	40	
15	3	4	225	9	16	12	
54	48	102		494	2188	1034	720

$a = 16.477$, $b = 0.3899$, $c = -0.623344$

Moment -

moments are the statistical tool to -

- To describe the characteristic of distribution.
- describe the distribution of frequency.
- gives an idea about the shape of distribution.

~~Utility the~~

Utility of Moments -

- With the help of moments, we can measure the central tendency, variability (variance), skewness, asymmetry & Kurtosis, height of the peak of curve.

Moments about mean -

If $x_1, x_2, x_3, ---, x_n$ are the values of the variables under consideration, the rth moment μ_r about mean $\bar{x}$ is defined as -

$$\mu_r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^r \quad ; \quad r = 0, 1, 2, ----$$

$r = 0$: $\mu_0 = \frac{1}{n}$

For a frequency distribution –
If $x_1, x_2, \dots x_n$ are the values of variable x with the corresponding freq. $f_1, f_2, \dots f_n$ respectively, then rth moment μ_r about the mean $\bar{x}$ is defined as –

$$\mu_r = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^r \quad ; \quad r = 0, 1, 2, \dots$$

$$N = \sum_{i=1}^{n} f_i$$

$r = 0$: $\mu_0 = 1$

~~$r=1$; $\mu_1 =$~~

<u>Moments about Mean (central moment)</u>

~~For an individual series –~~
~~If $x_1, x_2, \dots, x_n$ are the values of~~

$r = 1$: $$\mu_1 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^1$$

$$= \frac{1}{N} \sum_{i=1}^{n} f_1 x_i - \frac{\bar{x}}{N} \sum_{i=1}^{n} f_i$$

$$= \bar{x} - \frac{\bar{x} N}{N} = 0$$

$r = 2$: $$\mu_2 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2$$

$$= \text{variance} = (S.D)^2$$

$r = 3$: $$\mu_3 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^3$$

<u>Moments about arbitrary number (Assumed Mean 'A')</u> –

For frequency series –

$$\mu'_r = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - A)^r \quad ; \quad N = \sum_{i=1}^{n} f_i$$

$r = 0$: $\mu'_0 = \frac{1}{N} \times N = 1$

$r = 1$: $\mu_1' = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - A)^1$

$= \frac{1}{N} \sum_{i=1}^{n} f_i x_i - \frac{1}{N} \sum_{i=1}^{n} f_i A$

$= \bar{x} - \frac{1}{N} \times A \times N$

$$\boxed{\mu_1' = \bar{x} - A}$$

$r = 2$: $\mu_2' = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - A)^2$

$\cancel{= \frac{N}{N} \sum_{i=1}^{n} f_i x_i - \frac{1}{N}}$

$r = 3$: $\mu_3' = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - A)^3$

Moments about the origin – (V_r)

$$V_r = \frac{1}{N} \sum_{i=1}^{n} f_i x_i^r$$

$r = 0$: $V_0 = \frac{1}{N} \sum_{i=1}^{n} f_i x_i' = \frac{1}{N} \times N = 1$

$r = 1$: $V_1 = \frac{1}{N} \sum_{i=1}^{n} f_i x_i = \bar{x}$

$r = 2$: $V_2 = \frac{1}{N} \sum_{i=1}^{n} f_i x_i^2$

Relation b/w μ_r & μ_r' –

$\mu_r = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2$, $\mu_r' = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - A)^r$

$\mu_r = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - A + A - \bar{x})^r$

$\mu_r = \frac{1}{N} \sum_{i=1}^{n} f_i ((x_i - A) - (\bar{x} - A))^r$

$$= \frac{1}{N}\sum_{i=1} f_i\left[(x_i - A) - (\bar{x} - A)\right]^r$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \left[{}^rC_0 (x_i - A)^r (\bar{x}-A)^0 - {}^rC_1 (x_i - A)^{r-1}(\bar{x}-A)^1 + {}^rC_2 (x_i-A)^{r-2}(\bar{x}-A)^2 - \cdots \; {}^rC_r (x_i - A)^0 (\bar{x}-A)^r \right]$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \left[(x_i - A)^r - r(x_i-A)^{r-1}(\bar{x}-A) + \frac{r(r-1)}{2}(x_i-A)^{r-2}(\bar{x}-A)^2 - \cdots + (\bar{x}-A)^r \right]$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^r - \frac{1}{N}\sum_{i=1}^{n} f_i\, r (x_i - A)^{r-1}(\bar{x}-A) + \cdots$$

$$\mu_r = \mu_r' - {}^rC_1 \mu_{r-1}' \mu_1' + {}^rC_2 \mu_{r-2}' \mu_1'^2 - \cdots + (-1)^r \mu_1'^r$$

$r=2$: $\mu_2 = \mu_2' - 2\mu_1'\mu_1' + \mu_0' \mu_1'^2$

$$= \mu_2' - 2(\mu_1')^2 + \mu_1'^2$$

$$\boxed{\mu_2 = \mu_2' - (\mu_1')^2}$$

$r=3$: $\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 3\mu_1'^3 - \mu_1'^3$

$$\boxed{\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3}$$

$r=4$: $\boxed{\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4}$

Relation b/w v_r & μ_r –

$v_1 = \bar{x}$

$v_2 = \mu_2 + \bar{x}^2$

$v_3 = \mu_3 + 3\mu_2\bar{x} + \bar{x}^3$

$v_4 = \mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4$

Karl Pearson's coefficients

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\gamma_1 = \sqrt{\beta_1}$$

$$\gamma_2 = \beta_2 - 3$$

Q. The first four moments of a distribution about $x = 2$ are 1, 2.5, 5.5 & 16. Calculate the first four moments about the mean & about the origin.

$A = 2$

$\mu_1' = 1, \quad \mu_2' = 2.5, \quad \mu_3' = 5.5, \quad \mu_4' = 16$

$\mu_1' = \bar{x} - A$

$\bar{x} = 1 + 2 = 3$

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1')^2 = 2.5 - 1^2 = \cancel{2.25}\ 1.5$

$\mu_3 = 0$

$\mu_4 = 6$

$\nu_1 = \bar{x} = 3$

$\nu_2 = 1.5 + 9 = 10.5$

$\nu_3 = 0 + 3 \times 1.5 \times 3 + 27 = 13.5 + 27$

$= 40.5$

$\nu_4 = 6 + 4 \times 0 + 6 \times 1.5 \times 9 + 81$

$= 6 + 81 + 81$

$= 168$

$A = 4$

$\mu_1' = \bar{x} - A =$

$\beta_1 = 0, \quad \beta_2 = \frac{6}{2.25} = 2.66$

Q. The first four moments of the distribution about a value 4 are $-1.5, 17, -30$ & 108. find moments about mean, about origin β_1 & β_2. Also find moments about a pt. $x = 2$.

$A = 4$

$\mu_1' = -1.5, \quad \mu_2' = 17, \quad \mu_3' = -30, \quad \mu_4' = 108$

$\bar{x} = -1.5 + 4 = 2.5$

$\mu_1 = 0$

$\mu_2 =$

Q. In a certain distribution the first four moments about pt $x = 4$ are -1.5, 17, -30 & 308
Also calculate β_1 & β_2

Q. Calcute $\mu_1, \mu_2, \mu_3, \mu_4$ for the given freq. distributi

	Marks	No. of students	$(x_i - \bar{x})^2$	$f_i(x_i-\bar{x})^2$
10	5-15	10	576	5760
20	15-25	20	196	3920
30	25-35	25	16	4000
40	35-45	20	36	720
50	45-55	15	256	3840
60	55-65	10	576	5760
			1656	24000

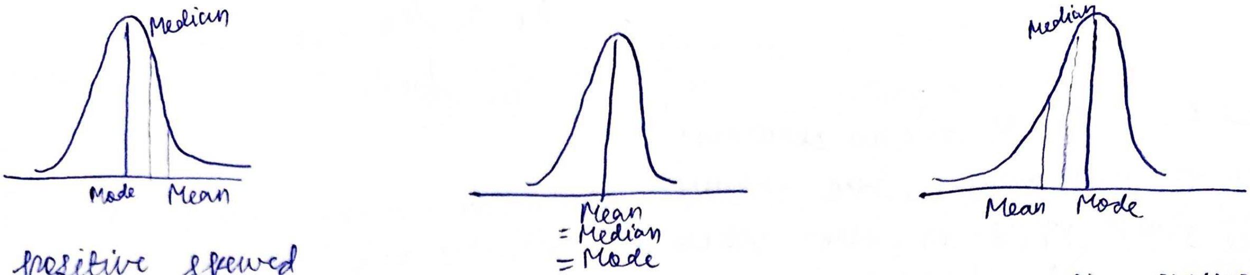
$\bar{x} = 34$

$\mu_1 = 0$

$$\mu_2 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2$$

$$\mu_3 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^3$$

Skewness



positive skewed
Mean > Median > Mode

Symmetric curve
no skewness

negative skewed.
Mode > Median > Mean

Methods of measuring skewness -

① Karl Pearson's Method -

Pearson's coefficient of skewness $(S_{KP}) = \dfrac{\text{Mean} - \text{Mode}}{\text{S.D.}}$

$$\boxed{\text{empirical mode} = 3\ \text{Median} - 2\ \text{mean}}$$

$$S_{KP} = \frac{\text{Mean} - (3\,\text{Median} - 2\,\text{mean})}{\text{S.D.}}$$

$$\boxed{S_{KP} = \frac{3(\text{mean} - \text{median})}{\text{S.D.}}}$$

Case 1: If $S_{KP} = 0$, then mean = mode, which implies no skewness.

Case 2: If $S_{KP} > 0$, then mean > mode, which implies positively skewed distribution.

Case 3: If $S_{KP} < 0$, then mean < mode, which implies negatively skewed distribution.

② Method of Moments - (coefficient of ~~mo~~ skewness based upon moments)

Moment coefficient of skewness gives the magnitude as well as direction of skewness present in the distribution.

Moment coefficient of skewness $\gamma_1 = \sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}}$

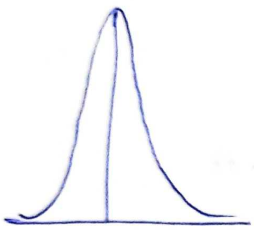
$$\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}}$$

Case 1 : $\gamma_1 = 0 \Rightarrow$ no skewness

Case 2 : $\gamma_1 > 0 \Rightarrow$ +ve skewed.

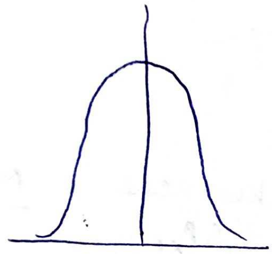
Case 3 : $\gamma_1 < 0 \Rightarrow$ -ve skewed.

Kurtosis



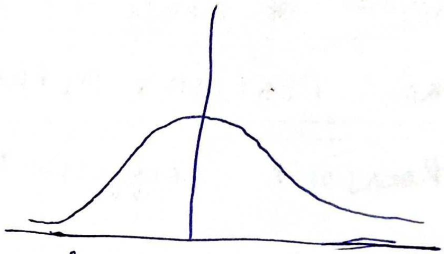
lepto kurtic (more sharpley peaked than normal)

$\beta_2 > 3$



meso kurtic (normal curve)

$\beta_2 = 3$



platy kurtic (flatter than normal)

$\beta_2 < 3$

$$\gamma_2 = \beta_2 - 3$$

Case 1 : $\gamma_2 = 0 \Rightarrow \beta_2 = 3$

Case 2 : $\gamma_2 > 0 \Rightarrow \beta_2 > 3$

Case 3 : $\gamma_2 < 0 \Rightarrow \beta_2 < 3$

Q. The first four moments of a distribution about the value 4 are -1.5, +17, -30 & 108. State whether the distribution is lacto kurtic or platy kurtic

$\mu_1' = -1.5, \quad \mu_2' = -17, \quad \mu_3' = -30, \quad \mu_4' = 108$

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1)^2 = +17 - 2.25 = +14.25$

$\mu_3 = 39.75$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

$= 108 + 120 \times 1.5 + 6 \times (-17) \times 2.25 - 3 \times (-1.5)^4$

$\mu_4 = 142.3125$

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, $\beta_2 = \frac{\mu_4}{\mu_2^2}$, $\beta_1 = 0.492$

$\beta_2 = \frac{142.3125}{\cancel{(11)^2}(14.95)^2} = \cancel{0.492}$ $0.654 < 3$

platy kurtic

Q. The first four moments about the working mean 28.5 of a distribution are 0.294, 7.144, 42.409, 454.98. Calculate the moment about the mean. Also evaluate β_1, β_2 and comment upon the skewness & kurtosis of the distribution.

$\mu_1' = 0.294$, $\mu_2' = 7.144$, $\mu_3' = 42.409$, $\mu_4' = 454.98$

$A = 28.5$

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1')^2 = 7.144 - (0.294)^2 = 7.057$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$

$= 36.158$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = 408.789$

~~= 454.98 − 4 × 3~~

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(36.158)^2}{(7.057)^3} = 3.\cancel{0}72$, $\gamma_1 = 1.92$

$\beta_2 = \frac{408.789}{(7.057)^2} = 8.208$, $\gamma_2 = 5.208$

$\Rightarrow$ +ve skewness , lapto kurtic

Q. Calculate the moment coefficient of skewness & coefficient of kurtosis

Classes
2-5-?5

	Classes	f	$f_i x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^3$
5	2·5 − 7·5	8	40	−14·916	222·30		
10	7·5 − 12·5	15	150	−9·916	98·20		
15	12·5 − 17·5	20	300	−4·916	24·108		
20	17·5 − 22·5	32	640	0·084	0·008		
25	22·5 − 27·5	23	575	5·084	25·90		
30	27·5 − 32·5	17	510	10·084	101·80		
35	32·5 − 37·5	5	175	15·084	227·70		
		120	2390			0·8	7198·456

$$\bar{x} = \frac{2390}{120} = 19.916$$

Note - for an individual series :-

① $\mu_r' = \frac{1}{n} \sum_{i=1}^{n} (x_i - A)^r \quad ; r = 0, 1, 2, \ldots$

② $\mu_r' = \frac{1}{N} \left[\sum_{i=1}^{n} f_i u_i^r \right] h^r \quad ; r = 0, 1, 2, \ldots$

where $u = \frac{x - A}{h}$

Q. Calculate the variance and third central moment (μ_3) from the given data -

x	f	u	fu	fu^2	fu^3
0	1	−4	−4	16	−64
1	9	−3	−27	81	−243
2	26	−2	−52	104	−208
3	59	−1	−59	59	−59
A 4	72	0	0	0	0
5	52	1	52	52	52
6	29	2	58	116	232
7	7	3	21	63	189
8	1	4	4	16	64
			−7	507	−37

$\mu_1' = \frac{1}{N} \left[\sum_{i=1}^{9} f_i u_i \right] h = -0.027$

$\mu_2' = \frac{1}{N} \left[\sum_{i=1}^{9} f_i u_i^2 \right] h^2 = 1.98$

$\mu_3' = \frac{1}{N} \left[\sum_{i=1}^{9} f_i u_i^3 \right] h^3 = -0.144$

$\mu_1 = 0$ (always)

$\mu_2 = \mu_2' - {\mu_1'}^2 = 1.979$

$\mu_3 = 0.0163$

Variance $= \mu_2 = 1.979$

Q. The first four moments about $x = 4$ are 1, 4, 10, 45 of the distribution. Obtain the various characteristics of the distribution on the basis of given information. Comment upon the nature of distribution.

$$\mu_1' = 1, \quad \mu_2' = 4, \quad \mu_3' = 10, \quad \mu_4' = 45$$

$$A = 4$$

$$\mu_1' = \bar{x} - A$$

$$\bar{x} = 1 + 4 = 5$$

$$\mu_1 = 0$$

$$\mu_2 = 3$$

$$\mu_3 = 0$$

$$\mu_4 = 26$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{26}{9} = 2.88 \Rightarrow \text{platykurtic}$$

$\gamma_1 = 0$ No skewness $\Rightarrow$ distribution is symmetric

Q. The first four moments of a distribution about the value 0 are 0.20, 1.76, −2.36, 10.88. Find the moments about mean & measure the kurtosis.

$\Rightarrow$ $A = 0$, $\mu_1' = 0.20$, $\mu_2' = 1.76$, $\mu_3' = -2.36$, $\mu_4' = 10.88$

$$\mu_1 = 0, \quad \mu_2 = \mu_2' - \mu_1'^2 = 1.76 - (0.20)^2 = 1.72$$

$$\mu_3 = -3.4, \quad \mu_4 = 13.1856$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 4.45 > 3 \quad \text{laptokurtic}$$

Q. The following table represents the height of a batch of 100 students. Calculate kurtosis.

height	No. of students
59	0
61	2
63	6
65	20
67	40
69	20
71	8
73	2
75	2

Q. find all 4 central moments & discuss skewness & kurtosis

x	f
2-4	38
4-6	292
6-8	389
8-10	212
10-12	69

Q. A bag contains 7 white, 6 red & 5 black balls. Two balls are drawn at random. Find the probability that they will both be white.

$$P(E) = \frac{{}^7C_2}{{}^{18}C_2} = \frac{7! \times 16! \, 2!}{5! \, 2! \times 18!}$$

$$= \frac{7 \times 6}{18 \times 17} = \frac{7}{51}$$

Random experiment -

Occurrences which can be repeated a no. of times essentially under the same conditions and whose result can not be predicted before hand are known as random experiment.

Sample space - Out of the several possible outcomes of a random experiment one & only one can take place in a trial. The set of all these possible outcomes is called sample space.

Axioms -

(i) With each event E is associated a real no. b/w 0 & 1, called the probability of that event & is denoted by

$$\boxed{0 \leq P(E) \leq 1}$$

(ii) The sum of the probabilities of all ~~poss~~ simple events constituting the sample space is 1.

$P(S) = 1$

(iii) The probability of a compound event is the sum of the probabilities of the simple events comprising the compound event.

* Probability of the impossible event is 0. $\boxed{P(\phi) = 0}$

* Probability of the complementary event is given by -

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

* for any 2 events A & B ~~probability of~~

$$\boxed{P(\bar{A} \cap B) = P(B) - P(A \cap B)}$$

* If $B \subset A$, then -

(i) $P(A \cap \bar{B}) = P(A) - P(B)$

(ii) $P(B) \leq P(A)$

* Addition theorem of probability = (theorem of total probability)

If A & B are 2 events then -

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

Note - If 2 events are mutually exclusive -

$P(A \cap B) = 0$

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

or $$\boxed{P(A + B) = P(A) + P(B)}$$

* If A, B & C are any 3 events -

$$\boxed{P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)}$$

If A, B, C are mutually exclusive -

$$\boxed{P(A \cup B \cup C) = P(A) + P(B) + P(C)}$$

Conditional Probability -

The probability of the happening of the event E_1 when another event E_2 is known to have already happened is called conditional probability & is denoted by $P\left(\frac{E_1}{E_2}\right)$

Note - An event E_1 is said to be independent of an event E_2, then -

$$\boxed{P\left(\frac{E_1}{E_2}\right) = P(E_1)}$$

Multiplicative law of Probability (Theorem of Compound probability) -

The probability of simultaneous occurrence of 2 events is equal to the prob. of one of the events is multiplied by the conditional prob. of the other.

for ~~an~~ two events A & B -

$$\boxed{P(A \cap B) = P(A) \cdot P\left(\frac{B}{A}\right)}$$

or $\boxed{P(AB) = P(A)\, P(B/A)}$

- When two events A & B are ~~mutually exclusive or~~ independent then,

$$\boxed{P(A \cap B) = P(A)P(B)}$$

- If P is the chance that an event will happen in one trial then the chance that it will happen in a succession of r trials is -

$$p \cdot p \cdot p \cdots p = p^r$$

- If $p_1, p_2, \ldots p_n$ are the prob. that certain events happen then the prob. of their non-happening -

$$(1-p_1)(1-p_2)(1-p_3) \cdots (1-p_n)$$

Therefore the prob. of all of these failing is

$$(1-p_1)\cdot(1-p_2)\cdot(1-p_3) \cdots \cdot(1-p_n)$$

Hence the chance in which atleast one of these events ~~must~~ happen is -

$$1 - \left[(1-p_1)(1-p_2) \cdots (1-p_n)\right]$$

Q. A problem in mathematical is given to 3 Students A, B & C, whose chances to solving it are $\frac{1}{2}$, $\frac{1}{3}$ & $\frac{1}{4}$ respectively what is the prob. that the problem will be solved by A, B & C.

$p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{4}$

$(1-p_1) = \frac{1}{2}$, $(1-p_2) = \frac{2}{3}$, $(1-p_3) = \frac{3}{4}$

P(not solving the problem) $= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} = \frac{1}{4}$

P(solving the problem) $= \frac{3}{4}$

Q. A can hit a target 4 times in 5 shorts, B 3 times in 4 shorts, C 2 times in 3 shorts. what is the prob. that atleast 2 shorts hit.

$\Rightarrow$ $P(A) = \frac{4}{5}$, $P(B) = \frac{3}{4}$, $P(C) = \frac{2}{3}$

for atleast 2 hits, we may have then

- when A, B, C all hit the target,

$P(A, B, C) = \frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} = \frac{2}{5}$

- when A & B hit the target, but C miss it

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

- When A & C hit, but B miss –

$P(A, \bar{B}, C) = \frac{4}{5} \times \frac{1}{4} \times \frac{2}{3} = \frac{2}{15}$

- When A miss, B & C hit –

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

P(at least 2 shorts hit) $= \frac{2}{5} + \frac{1}{5} + \frac{2}{15} + \frac{1}{10}$

$= \frac{12+6+4+3}{30} = \frac{25}{30} = \frac{5}{6}$

Q. A bag contains 10 white & 3 black balls. Another bag contains 3 white & 5 black balls. 2 balls are drawn from first bag & put into the II bag & then a ball is drawn from the latter. What is the prob. that it is a white ball.

=> Bag 1:

$P(W) = \frac{10}{13}$, $P(B) = \frac{3}{13}$

Bag 2:

$P(W) = \frac{3}{8}$, $P(B) = \frac{5}{8}$

$P(\text{having 2 white balls}) = \frac{\cancel{10}^{5}}{13} \times \frac{\cancel{9}^{3}}{\cancel{12}_{\cancel{6}\,2}} = \frac{15}{26}$

$P(W, B) = \frac{10}{13} \times \frac{2}{12}$

$P(B, B) = \frac{3}{13} \times \frac{2}{12} = \frac{3}{78} = \frac{1}{26}$

Baye's theorem –

$E_1, E_2, E_3 \ldots E_n$ are mutually exclusive and exhaustive events with $P(E_i) \neq 0$ when $i = 1$ to n of a random experiment, then for any arbitrary event, then for arbitrary event A of sample space of above exp. with $P(A) > 0$ then we have

$$\boxed{P\left(\frac{E_i}{A}\right) = \frac{P(A) \cdot P\left(\frac{A}{E_i}\right)}{\sum_{i=1}^{n} P(E_i)\, P(A/E_i)}}$$

Proof-

Let S be ~~a the~~ the sample of the random experiment & $E_1, E_2, E_3 \ldots E_n$ are mutually exclusive (given)

$$S = E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n$$

$$\therefore A \subset S$$

$$A = A \cap S$$

$$A = A \cap [E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n]$$

By distribution law -

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \ldots \cup (A \cap E_n) \quad \text{—— (1)}$$

we know that -

$$P\left(\frac{E_1}{E_2}\right) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

$$\Rightarrow P(E_1 \cap E_2) = P(E_2)\, P\left(\frac{E_1}{E_2}\right)$$

from eq. 1 -

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \ldots + P(A \cap E_n)$$

$$P(A) = P(E_1)\, P\left(\frac{A}{E_1}\right) + P(E_2)\, P\left(\frac{A}{E_2}\right) + \ldots + P(E_n)\, P\left(\frac{A}{E_n}\right)$$

$$P(A) = \sum_{i=1}^{n} [P(E_i)\, P(A/E_i)]$$

$$P(E_i/A) = \frac{P(E_i \cap A)}{P(A)}, \quad P(A/E_i) = \frac{P(A \cap E_i)}{P(E_i)}$$

$$\Rightarrow P(E_i \cap A) = P(A)\, P(E_i/A), \quad \Rightarrow P(A \cap E_i) = P(E_i)\, P(A/E_i)$$

$$\Rightarrow P(A)\, P(E_i/A) = P(E_i)\, P(A/E_i)$$

$$\Rightarrow P(E_i/A) = \frac{P(E_i)\, P(A/E_i)}{P(A)}$$

$$\Rightarrow \boxed{P(E_i/A) = \frac{P(E_i)\, P(A/E_i)}{\sum_{i=1}^{n} [P(E_i)\, P(A/E_i)]}}$$

Hence Proved

Q. A bag x contains 2 white and 3 red balls & a bag y contains 4 white, 5 red balls. One ball is drawn at random from one of the bags and is found to be red. Find the probability that it was drawn from the bag y.

⇒ E_1 : a ball drawn from Bag x.

E_2 : a ball drawn from Bag y.

A : ball is red.

$$P\left(\frac{E_2}{A}\right) = \frac{P(E_2)\,P(A/E_2)}{P(E_2)P(A/E_2) + P(E_1)P(A/E_1)}$$

$$= \frac{\frac{1}{2} \times \frac{3}{5}}{\frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{5}{9}} = \frac{3/10}{3/10 + 5/18}$$

$$= \frac{3/10}{\frac{54+50}{180}} = \frac{3}{10} \times \frac{180}{104} = \frac{25}{52}$$

Q. The content of 3 bags are :-

Bag 1 : 1 white, 2 black, 3 red balls

Bag 2 : 2 white, 1 black, 1 red balls

Bag 3 : 4 white, 5 black, 3 red balls

One bag is chosen at random & 2 balls drawn they happened to be white & red. So what is the probability that they come from the bag 1, 2 or 3.

⇒ E_1 : 2 balls drawn from Bag 1

E_2 : " " " " " 2

E_3 : " " " " " 3

A : one ball red & one white

$$P\left(\frac{E_1}{A}\right) = \frac{P(E_1)\,P(A/E_1)}{P(E_1)P(A/E_1) + P(E_2)P(A/E_2) + P(E_3)P(A/E_3)}$$

$$= \frac{\frac{1}{3}\times\frac{1}{6_2}\times\frac{3}{5}}{\frac{1}{3}\times\frac{1}{6}\times\frac{3}{5} + \frac{1}{3}\times\frac{2}{4}\times\frac{1}{3} + \frac{1}{3}\times\frac{4}{12}\times\frac{3}{11}}$$

$$= \frac{\frac{1}{10}}{\frac{1}{10}+\frac{1}{6}+\frac{1}{11}} = \frac{\frac{1}{10}}{\frac{66+110+60}{660}}$$

110
66
60
236

$$= \frac{1}{\cancel{10}} \times \frac{66\cancel{0}}{236} = \frac{33}{118}$$

$$P\left(\frac{E_2}{A}\right) = \frac{1}{6} \times \frac{660}{236} = \frac{110}{236} = \frac{55}{118}$$

$$P\left(\frac{E_3}{A}\right) = \frac{1}{11} \times \frac{660}{236} = \frac{60}{236} = \frac{30}{118} = \frac{15}{59}$$

Q. In a bolt factory, machine A, B & C manufacture respectively 25%, 35% & 40% of total of their output 5%, 4% & 2% are defective bolts. A bolt is drawn at random & is found to be defective. What is the prob. that it was manufactured by machine B.

ans = .4057

=> E_1 :

Random Variable :-

Random variable is a real valued fx^n which assign a real no. to each sample point in sample place. It is also called chance variable or stochastic variable.

~~Random variable~~
example - tossing a coin 2 times.

$S = \{HH, HT, TH, TT\}$

No. of heads $X(HH) = 2$

$X(HT) = X(TH) = 1$

$X(TT) = 0$

$X = \{0, 1, 2\}$

X (no. of heads)	0	1	2
P	$\frac{1}{4}$	$\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$	$\frac{1}{4}$

Types of Random Variable -

① **Discrete Random variable** - A discrete Random variable is one which can assume only isolated values. ex - tossing a coin 2 times.

② **Continuous Random variable** - A continuous random variable is one which can assume any value within an interval.
ex - the weight of group of individuals.

Probability mass function -

let small x be a discrete random variable such that probability

$$P[X = x] = p_i$$

is said to be prob. mass fx^n (PMF) of random

Variable if it satisfies the following conditions -

(i) $p(x_i) \geq 0$

(ii) $\Sigma p(x_i) = 1$

ex - tossing a coin 3 times.

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

X = No. of heads

X	0	1	2	3
P	$1/8$	$3/8$	$3/8$	$1/8$

Mean of discrete Random variable -

Mean or expected value (μ or $E x$) $= \Sigma p_i x_i$

Variance of Random variable -

$$\sigma^2 = \Sigma p_i (x_i - \mu)^2$$

* If μ is not whole no., then

$$\sigma^2 = \Sigma p_i x_i^2 - \mu^2$$

Probability density function - (PDF)

A fx^n $f(x)$ is PDF if it satisfies the following conditions -

(i) $f(x) \geq 0$, $-\infty < x < \infty$

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

Q. If X is a continuous random variable with the following PDF

$$f(x) = \begin{cases} \alpha(2x - x^2) & ; \ 0 < x < 2 \\ 0 & ; \ \text{otherwise} \end{cases}$$

- find the value of α
- $P(X > 1)$

$\Rightarrow$ By definition of PDF —

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

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

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

$$\alpha \left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_0^2 = 1$$

$$\alpha \left[4 - \frac{8}{3}\right] = 1$$

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

$$P(X > 1) = \int_{1}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx$$

$$= \int_{1}^{2} \frac{3}{4}(2x - x^2)\,dx = \frac{3}{4}\left[x^2 - \frac{x^3}{3}\right]_1^2$$

$$= \frac{3}{4}\left[4 - \frac{8}{3} - 1 + \frac{1}{3}\right]$$

$$= \frac{3}{4}\left[\frac{2}{3}\right] = \frac{1}{2}$$

$3 - \frac{7}{3}$

$\frac{2}{3}$

Q. A Random variable X has density $f x^n$

$$f(x) = \begin{cases} kx^2 & ; -3 < x < 3 \\ 0 & ; \text{otherwise} \end{cases}$$

(i) find the value of k

(ii) $p(1 < X < 2)$

(iii) $p(X \leq 2)$ (iv) $p(X > 0)$

$\Rightarrow$

(i) $$\int_{-\infty}^{-3} f(x)\,dx + \int_{-3}^{3} f(x)\,dx + \int_{3}^{\infty} f(x)\,dx = 1$$

$$\int_{-3}^{3} k x^2 \cdot dx = 1$$

$$k\left[\frac{x^3}{3}\right]_{-3}^{3} = 1$$

$$k\left(\frac{27}{3} + \frac{27}{3}\right) = 1$$

$$k = \frac{3}{54} = \frac{1}{18}$$

(ii) $p(1 < X < 2)$

$$\int_{1}^{2} f(x)\,dx = \int_{1}^{2} \frac{1}{18} x^2 \cdot dx = \frac{1}{18}\left[\frac{8}{3} - \frac{1}{3}\right] = \frac{1}{18} \times \frac{7}{3} = \frac{7}{54}$$

(iii) $p(X \leq 2)$

$$\int_{-3}^{2} \frac{1}{18} x^2 \cdot dx = \frac{1}{18}\left[\frac{8}{3} + \frac{27}{3}\right] = \frac{1}{18} \times \frac{35}{3} = \frac{35}{54}$$

(iv) $p(X > 0)$

$$\int_{0}^{3} \frac{1}{18} x^2 \cdot dx = \frac{1}{18} \times \frac{27}{3} = \frac{9}{27}\ \frac{1}{2}$$

Q. A Random variable has the following prob. $f x^n$

X	0	1	2	3	4	5	6	7
p	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

(i) find k

(ii) $P(X<6)$ (iii) $p(x \geq 6)$ (iv) $p(3<x<6)$

(v) find min. value of x so that $p(X \leq x) \geq \frac{1}{2}$

$\Rightarrow$ (i) $10k^2 + 9k = 1$

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

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

$k = -1, \quad k = \frac{1}{10}$

(ii) $P(X<6) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5) = \frac{81}{100}$

(iii) $P(X \geq 6) = 2k^2 + 7k^2 + k$

$= 9k^2 + k$

$= \frac{9}{100} + \frac{1}{10} = \frac{19}{100}$

$\cancel{3k + k^2 = \frac{3}{10} + \frac{1}{100}}$

$\cancel{= 0.3 + 0.01}$

$\cancel{= 0.31}$

(iv) $P(3 < x \leq 6) = \frac{33}{100}$

(v) $P(X \leq 1) = P(X=0) + P(X=1) = \frac{1}{10} < \frac{1}{2}$

$P(X \leq 2) = \frac{3}{10} < \frac{1}{2}$

$P(X \leq 3) = \frac{1}{2} = \frac{1}{2}$

$P(X \leq 4) = \frac{8}{10} > \frac{1}{2}$

Mean for continuous random variable -

$$E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Variance -

$$\sigma^2 = E(x^2) - [E(x)]^2$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) \cdot dx - \left[\int_{-\infty}^{\infty} x\, f(x) \cdot dx\right]^2$$

Q. A dice is ~~#~~ tossed thrice. A success is getting 1 or 6 on a toss. Find the mean & variance of the no. of success.

$\Rightarrow$ $n = 3$

Distribution

- Discrete
  - Binomial distribution
  - Poission distribution
- continuous
  - Normal distribution

Binomial distribution – • It is a discrete prob. dist.

• It is used in those scenerio where the outcomes of an experiment are in the form of success & failure i.e. the set of only 2 alternatives. Hence the name is binomial.

ex - (i) a new drug is introduced & an experiment is conducted to identify whether the drug will cure the disease or it doesn't.

(ii) a coin tossed 10 times & identify the prob. of getting heads 3 times.

* Probability of r success is –

$$\boxed{P(r) = {}^{n}C_{r}\, p^{r} q^{n-r}}$$

where n is no. of repetitive trials,

p = prob. of a success.

q = prob. of a failure

$p + q = 1$

* Mean $(\mu) = np$
* Variance $(\sigma^2) = npq$
* Standard deviation $(\sigma) = \sqrt{npq}$
* mean > variance

Note – When the experiment be repeated N times the frequency of r success is –

$$\boxed{P(r) = N\, {}^{n}C_{r}\, p^{r} q^{n-r}}$$

(ii) $n = 10$, $x = 3$, $p = \frac{1}{2}$, $q = \frac{1}{2}$,

$$P(r) = {}^{10}C_3 \left(\tfrac{1}{2}\right)^3 \left(\tfrac{1}{2}\right)^7$$

$$= \frac{10!}{7!\,3!} \times \left(\frac{1}{2}\right)^{10} = \frac{15}{128}$$

Where to apply binomial distribution -

- finite no. of trials.
- 2 possible outcomes which are mutually exclusive.
- each trial is independent.
- the probability of success is same in each trial.

ex - 40% of people who purchased luxury cars are women. If 10 luxury car owners are randomly selected find the prob. that exactly 6 are women.

$$\Rightarrow \quad P(x) = {}^{10}C_6 (0.4)^6 (0.6)^4$$

$$= \frac{10!}{6!\,4!} (0.4)^6 (0.6)^4$$

$$= \frac{10 \times \overset{3}{\cancel{9}} \times \cancel{8} \times 7}{\cancel{24}_{3}} \times (0.4)^6 (0.6)^4$$

$$= \underline{\underline{0.114}}$$

Properties of Binomial distribution -

- discrete prob. distribution
- there are 2 main parameters n & p & so is called biparametric distribution. ~~which are mean~~
- ~~constants of binomial dist.~~
~~& variance~~.

Poisson distribution -

- It is also a discrete prob. distribution.
- It is used in those scenerios where the prob. of happening of an event is very small i.e. the chances of happenning of event is rare.

This means that the prob. of success is very small & the value of n is very large.

- Probabilities are calculated for a certain time period.

ex - prob. of defective items in a manufacturing company for a month.

* Probability of r success is -

$$\boxed{P(r) = \frac{e^{-\lambda}\lambda^{r}}{r!}} \quad , \text{ where } \lambda = np$$

* Mean = Variance = np

* Poisson distribution is a limiting case of binomial distribution.

$$P(x=r) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

$$\Rightarrow \lambda = np \Rightarrow p = \frac{\lambda}{n}$$

$$P(x=r) = \frac{n!}{(n-r)!\, r!} \left(\frac{\lambda}{n}\right)^{r} (1-p)^{n-r} \qquad [p+q=1]$$

$$= \frac{n!}{(n-r)!\, r!} \left(\frac{\lambda}{n}\right)^{r} \left(1-\frac{\lambda}{n}\right)^{n-r}$$

$$= \frac{n(n-1)(n-2)\ldots(n-(r-1))\cancel{(n-r)!}}{\cancel{(n-r)!}\, r!\, n^{r} \left(1-\frac{\lambda}{n}\right)^{r}} \lambda^{r} \left(1-\frac{\lambda}{n}\right)^{n}$$

$$= \frac{n^{r}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{(r-1)}{n}\right) \lambda^{r} \left(1-\frac{\lambda}{n}\right)^{n}}{r!\, n^{r} \left(1-\frac{\lambda}{n}\right)^{r}}$$

$$\lim_{n\to\infty} P(x=r) = \lim_{n\to\infty} \frac{\lambda^{r}\left(1-\frac{\lambda}{n}\right)^{n}}{r!} = \frac{\lambda^{r}}{r!} \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{n}$$

$$= \frac{\lambda^r}{r!} e^{-\lambda}$$

$$\boxed{P(x=r) = \frac{e^{-\lambda}\lambda^r}{r!}}$$

Q. Comment on the following statement for a binomial distribution mean is 6 & variance is 9

$$\text{Mean} = np = 6$$
$$\sigma^2 = npq = 9$$

$$\frac{1}{q} = 2/3$$
$$q = 1.5$$

~~Q. A dice is tossed thrice~~. A success of getting 1 ~~or 6 on toss, find me.~~

Q. If 10% of bolts produced by a machine are defective. Determine the prob. that out of 10 bolts chosen at random will be defected,
(i) one (ii) none (iii) at most 2 bolts

$\Rightarrow$ $n = 10$

$$p = \frac{10}{100} = \frac{1}{10}$$

$$q = 1 - \frac{1}{10} = \frac{9}{10}, \quad r = 1$$

$$P(X=1) = {}^nC_r\, p^r q^{n-r}$$
$$= {}^{10}C_1 \left(\frac{1}{10}\right)^1 \left(\frac{9}{10}\right)^{10-1} = 0.3874$$

$$P(X=0) = {}^{10}C_0 \left(\frac{1}{10}\right)^0 \left(\frac{9}{10}\right)^{10-0} = 0.3486$$

$$P(X \le 2) = P(x=0) + P(x=1) + P(X=2) = 0.3874 + 0.3486 + 0.1932$$
$$= 0.9297$$

Q. Out of 800 families with 4 children each how many families would be expected to have

(i) 2 boys & 2 girls

(ii) at least 1 boy

(iii) no girl

(iv) at most 2 girls

Assume equal prob. for boys & girls.

$\Rightarrow$ Total families = 800 = N

No. of children = 4 = n

$\cancel{800 C_4}$ $P(X=r) = {}^nC_r\, p^r q^{n-r} \cdot N$

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

(i) $P(X=r) = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2}$

$= \frac{4!}{2!2!} \times \frac{1}{16} = \frac{3}{8} \times 800 = \underline{\underline{300}}$

(ii) $P(X \geq 1) = {}^4C_1 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{4-1} + {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 + {}^4C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^1 + {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0$

$= \frac{4!}{3!1!} \left(\frac{1}{2}\right)^4 + \frac{4!}{2!2!} \left(\frac{1}{2}\right)^4 + \frac{4!}{3!1!} \left(\frac{1}{2}\right)^4 + \frac{4!}{4!0!} \frac{1}{16}$

$= \frac{1}{4} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16}$

$= \frac{4+6+4+1}{16} = \frac{15}{16} = \cancel{\frac{7}{8}}$

$= \frac{15}{16} \times 800$

$= \underline{\underline{750}}$

(iii) $P(X=0) = {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{4-4} = \frac{1}{16} \times 800 = \underline{\underline{50}}$

(iv) $P(X>2) = \frac{11}{16} \times 800 = \underline{\underline{550}}$

Q. ~~Let~~ A binomial variable X satisfies the relⁿ $9P(X=4) = P(X=2)$ when $n=6$. Find the value of the parameter p and $P(X=1)$.

$$9\,{}^nC_4\, p^4 q^{n-4} = {}^nC_2\, p^2 q^{n-2}$$

$$\frac{9\,{}^nC_4\, p^4 q^{n-4}}{{}^nC_2\, p^2 q^{n-2}} = 1$$

$$9\,\frac{\cancel{n!} \times (n-2)!\,2!\; p^2 q^{n-4-n+2}}{(n-4)!\,4! \times \cancel{n!}} = 1$$

$$9\,\frac{(n-2)(n-3)}{12} \times \frac{p^2}{q^2} = 1$$

$$9\,\frac{(6-2)(6-3)}{12}\,\frac{p^2}{q^2} = 1$$

~~$p^2 = 9q^2$~~

~~$p = 3q$~~

$$\Rightarrow p = \frac{q}{3}, \quad \cancel{q = 1 - \frac{q}{3} \Rightarrow q = \frac{2q}{3}}$$

~~$P(X=1) = {}^nC_1 (p)^1 (q)^{n-1}$~~

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

$$q + \frac{q}{3} = 1$$

$$\frac{4q}{3} = 1 \quad \Rightarrow q = \frac{3}{4}$$

$$\Rightarrow p = \frac{3}{4} \times \frac{1}{3} = \frac{1}{4}$$

$$\Rightarrow \boxed{p = \frac{1}{4}} \;\&\; \boxed{q = \frac{3}{4}}$$

$$P(X=1) = {}^nC_1 (p)^1 (q)^{n-1}$$

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

$$= 0.3559$$

Q. A bag contains 5 white, 7 red & 8 black balls
If 4 balls are drawn one by one with replacement
What is the probability that -
(i) none is white.
(ii) all are white
(iii) at least one is white
(iv) only 2 are white

=)

Q.1 If the mean of a binomial distribution is 3 and variance is 3/2, find the prob of obtaining at least 4 success.

Q.2 The sum & product of mean & variance of a binomial distribution are $\frac{25}{3}$ and $\frac{50}{3}$ respectively. Find the distribution.

Q.3 Fit a binomial distribution for the following data & compare the theoretical frequencies with the actual ones.

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

Q.4 - In 100 sets of 10 tosses of an unbiased coin. In how many cases do you expect to get
(i) 7 heads & 3 tails.
(ii) at least 7 heads

1.

$$\mu = 3, \quad \sigma^2 = 3/2$$

$$np = 3 \qquad npq = 3/2$$

$$\frac{1}{q} = \cancel{\phi}\, 2$$

$$q = 1/2, \quad p = 1/2$$

$$n \times \frac{1}{2} = 3$$

$$n = 6$$

$$P(X \geq 4) = P(X=4) + P(X=5) + P(X=6)$$

$$= {}^6C_4 \left(\tfrac{1}{2}\right)^4 \left(\tfrac{1}{2}\right)^2 + {}^6C_5 \left(\tfrac{1}{2}\right)^5 \left(\tfrac{1}{2}\right)^1 + {}^6C_6 \left(\tfrac{1}{2}\right)^6 \left(\tfrac{1}{2}\right)^0$$

$$= \frac{6!}{4!\,2!} \left(\tfrac{1}{2}\right)^6 + \frac{6!}{5!\,1!} \left(\tfrac{1}{2}\right)^6 + \frac{6!}{6!\,0!} \left(\tfrac{1}{2}\right)^6$$

$$= \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{22}{64} = 11/32$$

2.

$$np + npq = \frac{25}{3} \Rightarrow np(1+q) = \frac{25}{3}$$

$$np \times npq = \frac{50}{3}$$

Q. If the variance of a poisson distribution is 2. find the prob. for $r = 1, 2, 3, 4$. Also find ~~p where~~ $P(r \geq 4)$

$\Rightarrow$ $$P(X=r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

$$P(1) = \frac{e^{-\lambda}\lambda}{1!} = \cancel{e^{-2}\lambda}\ 2/e^2 = 0.27$$

$$P(2) = \frac{e^{-\lambda}\lambda^2}{2!} = \frac{e^{-2}\times 4}{2} = \frac{2}{e^2} = 0.27$$

$$P(3) = \frac{e^{-\lambda}\lambda^3}{3!} = \frac{e^{-2}\times 8}{6} = \frac{4}{3e^2} = 0.180$$

$$P(4) = \frac{e^{-\lambda}\lambda^4}{4!} = \frac{16}{24e^2} = \frac{2}{3e^2} = 0.0902$$

$$P(r \geq 4) = 1 - P(r < 4)$$

$$P(r \geq 4) = 1 - (0.27 + 0.27 + 0.180) \cancel{- 0.0902)}$$

$$= 0.145$$

Q. Find the prob. that the ace of spades will be drawn from a pack of well shuffled cards at least once in 104 consecutive trials

$\Rightarrow$ $$p = \frac{1}{52}, \quad n = 104$$

$$\lambda = np = \frac{1}{52} \times 104 = 2$$

$$P(X \geq 1) = 1 - P(X=0)$$

$$= 1 - \cancel{{}^nC_0\, p^0 q^{n-0}}\ \frac{e^{-\lambda}\lambda^r}{r!}$$

$$= 0.865$$

Q. In a certain factory turning out razor blades there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10. Calculate the approximate no. of packets containing no defective, one defective & two defective blades in a consignment of 10000 packets

⇒ ~~N = 10000~~ $n = 10$

$p = 0.002$, $\lambda = np = 0.02$

$$P(0) = \frac{e^{-\lambda}\lambda^{r}}{r!} = \frac{e^{-(0.02)}(0.02)^{0}}{0!} = \cancel{0.9802 \times 10000} = \cancel{9802}$$

$$P(1) = \frac{e^{-(0.02)}(0.02)^{1}}{1} = 0.0196 = \cancel{0.0196 \times 10000} = \cancel{196}$$

$$P(2) = \frac{e^{-(0.02)}(0.02)^{2}}{2!} = 0.000196 =$$

$$= (0.9802 + 0.0196 + 0.000196) \times 10000$$

$$= 999.996$$

Q.

death	0	1	2	3	4
f	122	60	15	2	1