Topic

Statistical Methods

Correlation, Lines of Regression, rank correlation

Prepared by:

Dr. Sunil

NIT Hamirpur (HP)

Last updated on 30-10-2007

CORRELATION:

There are many phenomenae where the changes in one variable are related to the changes in the other variable. For instance, the production of a crop varies with the amount of rainfall, the price of a commodity increases with the reduction in its supply and so on. Such a simultaneous variation, i. e. when the changes in one variable are associated or followed by changes in the other, is called correlation. Such a data connecting two variables is called bivariate population.

In a bivariate distribution, if the change in one variable affects a change in the other variable, the variables are said to be correlated.

If the two variables deviate in the same direction i. e., if the increase (or decrease) in one results in a corresponding increase (or decrease) in the other, correlation is said to be direct or positive. e. g., the correlation between income and expenditure is positive.

If the two variables deviate in opposite direction i. e., if the increase (or decrease) in one results in a corresponding decrease (or increase) in the other, correlation is said to be inverse or negative. e. g., the correlation between volume and the pressure of a perfect gas or the correlation between price and demand is negative.

Correlation is said to be perfect if the deviation in one variable is followed by a corresponding proportional deviation in the other.

If there is no relationship indicated between the variables, they are said to be independent or uncorrelated.

SCATTER OR DOT DIAGRAM:

To obtain a measure of relationship between the two variables, we plot their corresponding values on the graph taking one of the variables along the x-axis and the other along the y-axis. The resulting diagram showing a collection of dots is called a scatter diagram.

Let the origin be shifted to $(\bar{x}, \bar{y})$, where $\bar{x}$, $\bar{y}$ are the means of x's and y's that the new co-ordinates are given by

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

Now the points (X, Y) are so distributed over the four quadrants of XY-plane that the product XY is positive in the first and third quadrants but negative in the second and fourth quadrants. The algebraic sum of the products can be taken as describing the trend of the dots in all the quadrants.

$\therefore$ (i) If $\sum XY$ is positive, the trend of the dots is through the first and third quadrants.

(ii) If $\sum XY$ is negative, the trend of the dots is in the second and fourth quadrants, and

(iii) If $\sum XY$ is zero, the points indicates no trend i. e. the points are evenly distributed over the four quadrants.

The $\sum XY$ or better still $\frac{1}{n}\sum XY$, i. e the average of n products may be taken as a measure of correlation. If we put X and Y in their units, i. e. taking σ_x as the unit for x and σ_y for y, then

$$\frac{1}{n}\sum \frac{X}{\sigma_x}.\frac{Y}{\sigma_y}, \quad \text{i. e. } \frac{\sum XY}{n\sigma_x\sigma_y} \text{ is the measure of correlation.}$$

CO-EFFICIENT OF CORRELATION:

Correlation co-efficient between two variables x and y, usually denoted by r(x, y) or r_{xy} is a numerical measure of linear relationship between them. In other words, the numerical measure of correlation is called the co-efficient of correlation and is defined by the relation

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 . \sum (y_i - \bar{y})^2}} = \frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2 . \frac{1}{n}\sum (y_i - \bar{y})^2}}$$

$$= \frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \Rightarrow r = r_{xy} = \frac{\sum XY}{n\sigma_x \sigma_y}$$

where X = deviation from the mean $\bar{x} = x_i - \bar{x}$, Y = deviation from the mean $\bar{y} = y_i - \bar{y}$, σ_x = S. D. of x-series, σ_y = S. D. of y-series, and n = number of values of the two variables.

Remarks: Correlation co-efficient is independent of change of origin and scale.

Let us define two new variables u and v as

$u = \frac{x-a}{h}$, $v = \frac{y-b}{k}$, where a, b, h, k are constants, then $r_{xy} = r_{uv}$.

METHODS OF CALCULATION:

(a). Direct method:

Substituting the values of σ_x and σ_y in the formula $\frac{\sum XY}{n\sigma_x \sigma_y}$, we get

$$r = \frac{\sum XY}{\sqrt{\left(\sum X^2 . \sum Y^2\right)}} \qquad \text{(i)}$$

Another form of the formula (i) which is quite handy for calculation is

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[\left\{n\sum x^2 - \left(\sum x\right)^2\right\} \times \left\{n\sum y^2 - \left(\sum y\right)^2\right\}\right]}} \qquad \text{(ii)}$$

(b). Step-deviation method:

The direct method becomes very lengthy and tedious if the means of the two series are not integers. In such cases, use is made of assumed means. If d_x and d_y are step-deviations from the assumed means, then

$$r = \frac{n\sum d_x d_y - \sum d_x \sum d_y}{\sqrt{\left[\left\{n\sum d_x^2 - \left(\sum d_x\right)^2\right\} \times \left\{n\sum d_y^2 - \left(\sum d_y\right)^2\right\}\right]}} \qquad \text{(iii)}$$

where $d_x = \frac{(x-a)}{h}$ and $d_y = \frac{(y-b)}{k}$.

Remarks: The change of origin and units do not alter the value of correlation co-efficient since r is a pure number.

(c). Coefficient of correlation for grouped data:

When x and y series are both given as frequency distributions, these can be represented by a two-way table known as the correlation-table. It is double-entry table with one series along the horizontal and the other along the vertical. The co-efficient of correlation for such a bivariate frequency distribution is calculated by the formula

$$r = \frac{n\left(\sum f d_x d_y\right) - \left(\sum f d_x\right)\left(\sum f d_y\right)}{\sqrt{\left[\left\{n\sum f d_x^2 - \left(\sum f d_x\right)^2\right\} \times \left\{n\sum f d_y^2 - \left(\sum f d_y\right)^2\right\}\right]}},$$

where d_x = deviation of the central values from the assumed mean of x-series,

d_y = deviation of the central values from the assumed mean of y-series,

f is the frequency corresponding to the pair (x, y)

and $n\left(=\sum f\right)$ is the total number of frequencies.

RANK CORRELATION:
[Spearman's formula for rank correlation]

A group of n individuals may be arranged in order of merit with respect to some characteristics. The same group would give different orders for different characteristics. Considering the orders corresponding to two characteristics A and B, the correlation between these n pairs of ranks is called the rank correlations in the characteristics A and B for that group of individuals.

Let x_i, y_i be the ranks of the i^{th} individuals in A and B respectively. Assuming that no two individuals are bracketed equal in either case, each of the variables taking the values 1, 2, 3, ...,n, we have

$$\bar{x} = \bar{y} = \frac{1+2+3+......+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

If X, Y be the deviations of x, y from their means, then

$$\sum X_i^2 = \sum (x-\bar{x})^2 = \sum x_i^2 + n(\bar{x})^2 - 2\bar{x}\sum x_i = \sum n^2 + \frac{n(n+1)^2}{4} - 2\frac{n+1}{2}.\sum n$$

$$= \frac{n(n+1)(2n+1)}{6} + \frac{n(n+1)^2}{4} - \frac{n(n+1)^2}{2} = \frac{1}{12}\left(n^3-n\right).$$

Similarly, $\sum Y_i^2 = \frac{1}{12}\left(n^3-n\right)$.

If d_i denotes the difference in ranks of the i^{th} individual, then

$d_i = x_i - y_i$ so that $d_i = (x_i-\bar{x})-(y_i-\bar{y}) = X_i - Y_i$ $\qquad [\because \bar{x}=\bar{y}]$

$$\therefore \sum d_i^2 = \sum X_i^2 + \sum Y_i^2 - 2\sum X_i Y_i$$

$$\Rightarrow \sum X_i Y_i = \frac{1}{2}\left(\sum X_i^2 + \sum Y_i^2 - \sum d_i^2\right) = \frac{1}{12}\left(n^3-n\right) - \frac{1}{2}\sum d_i^2 .$$

Hence the correlation co-efficient between these variables is

$$r = \frac{\sum X_i Y_i}{\sqrt{\left(\sum X_i^2 \sum Y_i^2\right)}} = \frac{\frac{1}{12}\left(n^3-n\right) - \frac{1}{2}\sum d_i^2}{\frac{1}{12}\left(n^3-n\right)} = 1 - \frac{6\sum d_i^2}{n^3-n}.$$

This is called the rank correlation co-efficient and is denoted by ρ.

REGRESSION:

Regression is the estimation or prediction of unknown values of one variable from known values of another variable.

After establishing the fact of correlation between two variables, it is natural curiosity to known the extent to which one variable varies in response to a given variation in the other variable i. e., one is interested to know the nature of relationship between the two variables.

Regression measures the nature and extent of correlation.

LINEAR REGRESSION:

If two variates x and y are correlated i. e., there exist an association or relationship between them, then the scatter diagram will be more or less concentrated round a curve. This curve is called the curve of regression and relationship is said to be expressed by means of curvilinear regression. In the particular case, when the curve is straight line, it is called a line of regression and regression is said to be linear.

A line of regression is the straight line which gives the best fit in the least square sense to the given frequency.

If the line of regression is so chosen that the sum of squares of deviation parallel to the axis of y is minimized, it is called the line of regression of y on x and it gives the best estimate of y for given value of x.

If the line of regression is so chosen that the sum of squares of deviation parallel to the axis of x is minimized, it is called the line of regression of x on y and it gives the best estimate of x for given value of y.

LINES OF REGRESSION:

Let the equation of line of regression of y on x be

$y = a + bx$ (i)

Then $\bar{y} = a + b\bar{x}$ (ii)

Subtracting (ii) from (i), we have $y - \bar{y} = b(x - \bar{x})$ (iii)

The normal equations are $\sum y = na + b\sum x \Rightarrow \sum xy = a\sum x + b\sum x^2$ (iv)

Shifting the origin to $(\bar{x}, \bar{y})$, (iv) becomes

$$\sum(x-\bar{x})(y-\bar{y}) = a\sum(x-\bar{x}) + b\sum(x-\bar{x})^2 \qquad \text{(v)}$$

Since $\dfrac{\sum(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} = r$, $\sum(x-\bar{x}) = 0$ and $\dfrac{1}{n}\sum(x-\bar{x})^2 = \sigma_x^2$.

$\therefore$ From (v), $n\,r\,\sigma_x\sigma_y = a.0 + b.n\sigma_x^2 \Rightarrow b = \dfrac{r\sigma_y}{\sigma_x}$

Hence, from (iii), the line regression of y on x is $y - \bar{y} = r\dfrac{\sigma_y}{\sigma_x}(x - \bar{x})$

Similarly, the line regression of x on y is $x - \bar{x} = r\dfrac{\sigma_x}{\sigma_y}(y - \bar{y})$

Here $\dfrac{r\sigma_y}{\sigma_x}$ is called the regression co-efficient of y on x and is denoted by b_{yx}

and $\frac{r\sigma_x}{\sigma_y}$ is called the regression co-efficient of x on y and is denoted by b_{xy}.

Remarks: If r = 0, the two lines of regression become $y=\bar{y}$ and $x=\bar{x}$ which are two straight lines parallel to X and Y axes respectively and passing through their means $\bar{y}$ and $\bar{x}$. They are mutually perpendicular.

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

PROPERTIES OF REGRESSION CO-EFFICIENTS:

Property I: Correlation co-efficient is the geometric mean between the regression co-efficients.

Proof: The co-efficients of regression are $\frac{r\sigma_y}{\sigma_x}$ and $\frac{r\sigma_x}{\sigma_y}$.

G. M. between them $\sqrt{\frac{r\sigma_y}{\sigma_x}\times\frac{r\sigma_x}{\sigma_y}}=\sqrt{r^2}=r$ = co-efficient of correlation.

Property II: If one of the regression co-efficients is greater than unity, the other must be less than unity.

Proof: The two regression co-efficients are $b_{yx}=\frac{r\sigma_y}{\sigma_x}$ and $b_{xy}=\frac{r\sigma_x}{\sigma_y}$.

Let $b_{yx}>1$, then $\frac{1}{b_{yx}}<1$ (i)

Since $b_{yx}.b_{xy}=r^2\le 1$ $[\because -1\le r\le 1]$

$b_{xy}\le\frac{1}{b_{yx}}<1$. [using (i)]

Similarly, if $b_{xy}>1$, then $b_{xy}<1$.

Property III: Arithmetic mean of regression co-efficients is greater than the correlation co-efficient.

Proof: We have to prove that $\frac{b_{yx}+b_{xy}}{2}>r\Rightarrow\frac{\frac{r\sigma_y}{\sigma_x}+\frac{r\sigma_x}{\sigma_y}}{2}>r$

$\Rightarrow\sigma_y^2+\sigma_x^2>2\sigma_x\sigma_y\Rightarrow(\sigma_x-\sigma_y)^2>0$, which is true.

Property IV: Regression co-efficients are independent of the origin but not of scale.

Proof: Let $u = \frac{x-a}{h}$, $v = \frac{y-v}{k}$, where a, b, h and k are constants.

$$b_{yx} = \frac{r\sigma_y}{\sigma_x} = r.\frac{k\sigma_v}{h\sigma_u} = \frac{k}{h}\left(\frac{r\sigma_v}{\sigma_u}\right) = \frac{k}{h}b_{vu}.$$

Similarly, $b_{xy} = \frac{h}{k}b_{uv}$.

Thus b_{yx} and b_{xy} are both independent of a and b but not of h and k.

Property V: The correlation co-efficient and the two regression co-efficients have same sign.

Proof: Regression co-efficient of y on x $= b_{xy} = r\frac{\sigma_y}{\sigma_x}$.

Regression co-efficient of x on y $= b_{xy} = r\frac{\sigma_x}{\sigma_y}$.

Since σ_x and σ_y are both positive $\Rightarrow b_{yx}$, b_{xy} and r have same signs.

STANDARD ERROR OF ESTIMATE:

Result: To show that the coefficient of correlation lies between ± 1.

Proof: The sum of the squares of the deviations of the points from the line of regression of y on x is $\sum(y-a-bx)^2 = \sum(Y-bX)^2$, where $X = x-\bar{x}$, $Y = y-\bar{y}$

$$= \sum\left(Y - r\frac{\sigma_y}{\sigma_x}X\right)^2 = \sum Y^2 - 2r\left(\frac{\sigma_y}{\sigma_x}\right)\sum XY + r^2\left(\frac{\sigma_y^2}{\sigma_x^2}\right)\sum X^2$$

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

Denoting this sum of squares by nS_y^2, we have $S_y = \sigma_y\sqrt{\left(1-r^2\right)}$ (i)

Since S_y is the root mean square deviation of the points from the regression line of y on x, it is called the standard error of estimate of y.

Similarly, the standard error of estimate of x is given by

$$S_x = \sigma_x\sqrt{\left(1-r^2\right)}.$$

Since S_y is the sum of the squares of real quantities, it is never negative, hence it follows that $1-r^2 \geq 0 \Rightarrow r^2 \leq 1 \Rightarrow -1 \leq r \leq 1$.

If r = 1 or -1, the sum of the squares of the deviations from either line of regression is zero. Consequently each deviation is zero and all the points lie on both the lines of regression. These two lines coincide and we say that the correlation between the variables is perfect. The nearer r^2 is to unity the closer are the points to the lines of regression. Thus the departure of r^2 from unity is a measure of departure from linearity of the relationship between the variables.

ANGLE BETWEEN TWO LINES OF REGRESSION:

If θ is the acute angle between the two regression lines in the case of two variables x and y, show that $\tan\theta = \frac{1-r^2}{r}.\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$, where r, σ_x, σ_y have their usual meanings.

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

Proof: The equations to the lines of regression of y on x and x on y are

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

$\therefore$ Their slopes are $m_1 = \frac{r\,\sigma_y}{\sigma_x}$ and $m_2 = \frac{\sigma_y}{r\,\sigma_x}$.

$$\text{Thus } \tan\theta = \pm\frac{m_2-m_1}{1+m_1m_2} = \pm\frac{\frac{\sigma_y}{r\,\sigma_x}-\frac{r\,\sigma_y}{\sigma_x}}{1+\frac{\sigma_y^2}{\sigma_x^2}} = \pm\frac{1-r^2}{r}.\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}.$$

Since $r^2 \leq 1$ and σ_x, σ_y are positive .

$\therefore$ Positive sign gives the acute angle between the lines.

$$\text{Hence } \tan\theta = \frac{1-r^2}{r}.\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

This completes the proof of the first part.

When $r = 0,\ \tan\theta \to \infty \ \Rightarrow \theta = \frac{\pi}{2}$. The two lines of regression are perpendicular to each other. Hence the estimate value of y is same for all values of x and vice-versa.

When $r = \pm 1,\ \tan\theta = 0 \Rightarrow \theta = 0$ or π. Thus the lines of regression coincide and there is perfect correlation between the two variables x and y.

Now let us solve some problems:

Q.No.1.: Psychological tests of intelligence and of engineering ability were applied to 10 students. Here is a record of ungrouped data showing intelligence ratio (I. R.) and engineering ratio (E. R.). Calculate the co-efficient of correlation.

Student	A	B	C	D	E	F	G	H	I	J
I. R.	105	104	102	101	100	99	98	96	93	92
E. R.	101	103	100	98	95	96	104	92	97	94

Sol.: We consist the following table:

Student	Intelligence ratio		Engineering ratio		X^2	Y^2	XY
	x	$x - \bar{x} = X$	y	$y - \bar{y} = Y$			
A	105	6	101	3	36	9	18
B	104	5	103	5	25	25	25
C	102	3	100	2	9	4	6
D	101	2	98	0	4	0	0
E	100	1	95	-3	1	9	-3
F	99	0	96	-2	0	4	0
G	98	-1	104	6	1	36	-6
H	96	-3	92	-6	9	36	18
I	93	-6	97	-1	36	1	6
J	92	-7	94	-4	49	16	28
Total	990	0	980	0	170	140	92

With the help of above table, we have

mean of x, i. e. $\bar{x} = \frac{\sum x}{n} = \frac{990}{10} = 99$ and mean of y, i. e. $\bar{y} = \frac{\sum y}{n} = \frac{980}{10} = 98$.

$\sum X^2 = 170$, $\sum Y^2 = 140$ and $\sum XY = 92$.

Substitute these values in the formula $r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$, we get

$$r = \frac{\sum XY}{\sqrt{\left(\sum X^2 \sum Y^2\right)}} = \frac{92}{\sqrt{(170 \times 140)}} = \frac{92}{154.27} = 0.596 \text{. Ans.}$$

Q.No.2.: Find the correlation coefficient between x and y from the given data**:**

x :	78	89	97	69	59	79	68	57
y :	125	137	156	112	107	138	123	108

Sol.: Mean $\bar{x} = \frac{\sum x}{8} = \frac{596}{8} = 74.5$, $\bar{y} = \frac{\sum y}{8} = \frac{1006}{8} = 125.75$.

x	$x - \bar{x} = X$ $\Rightarrow x - 74.5 = X$	y	$y - \bar{y} = Y$ $\Rightarrow y - 125.75 = Y$	X^2	Y^2	XY
78	3.5	125	-0.75	12.25	0.5625	-2.625
89	14.5	137	11.25	210.25	126.5625	163.125
97	22.5	156	30.25	506.25	915.0625	680.625
69	-5.5	112	-13.75	30.25	189.625	75.625
59	-15.5	107	-18.75	240.25	351.5625	290.625
79	4.5	138	12.25	20.25	150.0625	55.125
68	-6.5	123	-2.75	42.25	7.6525	17.875
57	-17.5	108	-17.75	306.25	315.0625	310.625
$\sum x = 596$	Total $\sum X = 0$	1006	$\sum Y = 0$	$\sum X^2 = 1368$	$\sum Y^2 = 2055.5$	$\sum XY = 1591$

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} = \frac{1591}{\sqrt{1368 \times 2055.5}} = \frac{1591}{1676.88} = 0.949 = 0.95 \text{. Ans.}$$

Q.No.3.: Find the co-efficient of correlation between industrial production and export using the following data and comment on the result.

Production (in crore tons)	55	56	58	59	60	60	62
Exports (in crore tons)	35	38	38	39	44	43	45

Sol.: Here mean $\bar{x} = \frac{\sum x}{7} = 58.57$ and $\bar{y} = \frac{\sum y}{7} = 40.3$.

Production x	Deviation of x from mean (=58.57) $X = x - \bar{x}$	Export y	Deviation of y from mean (=40.3) $Y = y - \bar{y}$	X^2	Y^2	XY
55	-3.57	35	-5.3	12.75	28.09	18.921
56	-2.57	38	-2.3	6.6	5.29	5.911
58	-0.57	38	-2.3	0.325	5.29	1.311
59	0.43	39	-1.3	0.185	1.69	-0.559
60	1.43	44	3.7	2.045	13.69	5.291
60	1.43	43	2.7	2.045	7.29	3.861
62	3.43	45	4.7	11.765	22.69	16.121
$\sum x = 410$	$\sum X = 0.01$	282	$\sum Y = 0.1$	$\sum X^2 = 35.715$	$\sum Y^2 = 83.43$	$\sum XY = 50.857$

$$\therefore \text{Coefficient of correlation, } r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}} = \frac{50.857}{54.587} = 0.93. \text{ Ans.}$$

Q.No.4.: Ten people of various heights as under, were requested to read the letters on a car at 25 yard distance. The number of letters correctly read is given below:

Height (in feet)	5.1	5.3	5.6	5.7	5.8	5.9	5.10	5.11	6.0	6.1
No. of letters	11	17	19	14	8	15	20	6	18	12

Is there any correlation between height and visual power?

Sol.: Let the assumed mean of x is 5.8 and assumed mean of y is 14.

x	$X = x - \bar{x}$	y	$Y = y - \bar{y}$	X^2	Y^2	XY
5.1	-0.7	11	-3	0.49	9	2.1

5.3	-0.5	17	3	0.25	9	-1.5
5.6	-0.2	19	5	0.04	25	-1
5.7	-0.1	14	0	0.01	0	0
5.8	0	8	-6	0	36	0
5.9	0.1	15	1	0.01	1	0.1
5.10	0.2	20	6	0.04	36	1.2
5.11	0.3	6	-8	0.09	64	-2.4
6.0	0.4	18	4	0.16	16	1.6
6.1	0.4	12	-2	0.25	4	-1
		$\sum y = 140$		$\sum X^2 = 1.34$	$\sum Y^2 = 200$	$\sum XY = -0.9$

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} = \frac{-0.9}{\sqrt{1.34 \times 200}} = -0.055. \text{ Ans.}$$

Q.No.5.: In the two regression equations of the variables x and y are $x = 19.13 - 0.87y$ and $y = 11.64 - 0.50x$. Find **(i)** mean of x's **(ii)** mean of y's and **(iii)** the correlation co-efficient between x and y.

Sol.: Since the mean of x's and the mean of y's lie on the two regression lines, we have

$\bar{x} = 19.13 - 0.87\bar{y}$ (i)

$\bar{y} = 11.64 - 0.50\bar{x}$ (ii)

(i). Multiplying (ii) by 0.87 and subtracting from (i), we have

$$[1 - (0.87)(0.50)]\bar{x} = 19.13 - (11.64)(0.87) \Rightarrow 0.565\bar{x} = 9.00 \Rightarrow \bar{x} = 15.93. \text{ Ans.}$$

(ii). $\bar{y} = 11.64 - (0.50)(15.93) = 3.675$. Ans

(iii). Now regression co-efficient of y on x is -0.50 and that of x on y is -0.87.

Now since the co-efficient of correlation is the geometric mean between the two regression co-efficients.

$$\therefore r = \sqrt{[(-0.50)(-0.87)]} = \sqrt{(0.435)} = -0.6595 = -0.66. \text{ Ans.}$$

[Negative sign is taken since both the regression co-efficients are negative.]

Q.No.6.: Two random variables have the regression lines with equations $3x + 2y = 26$ and $6x + y = 31$. Find the mean values and the correlation coefficient between x and y.

Sol.: Given regression lines are: $3x + 2y = 26$ (i) $6x + y = 31$ (ii)

Now from (i), we get $2y = 26 - 3x \Rightarrow y = -1.5x + 13$ (iii) $\therefore r\dfrac{\sigma_y}{\sigma_x} = -1.5$.

From (ii), we get $x = -0.167y + 5.167$ (iv) $\therefore r\dfrac{\sigma_x}{\sigma_y} = -0.167$.

Since mean of x's and mean of y's lie on two regression lines, we have

$\overline{y} = 13 - 1.5\overline{x}$ (v) and $\overline{x} = 5.167 - 0.167\overline{y}$ (vi)

From (v) and (vi), we get

$\overline{y} = 13 - 1.5\overline{x} \Rightarrow 7.7505 - 0.2505\overline{y} = 1.5\overline{x} \Rightarrow 0.7495\overline{y} = 5.2495$

$\therefore \overline{y} == 7$ and $\overline{x} = 4$. Ans.

$\therefore$ Regression coefficient of y on x is -1.5 and that of x on y is -0.167.

Now since the coefficient of correlation is the geometric mean between the two regression coefficients:

$r^2 = (-1.5)(-0.167) = 0.2505$, $\therefore r = -0.5005 = -0.5$. Ans.

[Here we take negative sign because both the regression co-efficients are negative.]

Q.No.7.: The regression equations of two variables x and y are $x = 0.7y + 5.2$, $y = 0.3x + 2.8$. Find the means of the variables and the coefficient of correlation between them.

Sol.: Given regression lines are: $y = 0.3x + 2.8$ and $x = 0.7y + 5.2$

$\therefore r\dfrac{\sigma_y}{\sigma_x} = 0.3$ and $r\dfrac{\sigma_x}{\sigma_y} = 0.7$

Since mean of x's and mean of y's lie on two regression lines, we have

$\overline{y} = 0.3\overline{x} + 2.8$ (i) $\overline{x} = 0.7\overline{y} + 5.2$ (ii)

Multiplying (i) with 0.7, we get $0.21\overline{x} + 1.96 = 0.7\overline{y}$ (iii)

Solving (ii) and (iii), we get

$0.79\overline{x} = 7.16$. $\therefore \overline{x} = 9.06$ and $\overline{y} = 5.52$. Ans.

Also regression coefficient of y on x is 0.3 and that of x on y is 0.7.

Now since the coefficient of correlation is the geometric mean between the two regression coefficients:

$r^2 = 0.3 \times 0.7 = 0.4582$. $\therefore r = 0.46$. Ans.

[Here we take positive sign because both the regression co-efficients are positive.]

Q.No.8.: In a particular destroyed laboratory data, only the equations giving the two lines of regression of y on x and x on y are available and are respectively, $7x - 16y + 9 = 0$, $5y - 4x - 3 = 0$. Calculate the coefficient of correlation, $\bar{x}$ and $\bar{y}$.

Sol.: Given regression lines are: $7x - 16y + 9 = 0$ and $5y - 4x - 3 = 0$

$\Rightarrow y = 0.4375x + 0.5625$ (i) and $x = 1.25y - 0.75$ (ii)

$\therefore r\dfrac{\sigma_y}{\sigma_x} = 0.4375 =$ regression co-efficient of y on x.

and $r\dfrac{\sigma_x}{\sigma_y} = 1.25 =$ regression co-efficient of x on y.

Since mean of x's and mean of y's lie on two regression lines, we have

$\bar{y} = 0.4375\bar{x} + 0.5625 \Rightarrow 0.4375\bar{x} - \bar{y} + 0.5625 = 0$ (iii)

$\bar{x} = 1.25\bar{y} - 0.75 \Rightarrow \bar{x} - 1.25\bar{y} + 0.75 = 0$ (iv)

Multiplying (iii) by 1.25 and subtracting, we get

$-0.453125\bar{x} - 0.046875 = 0 \Rightarrow \bar{x} = -0.103448 = -0.1034$. Ans.

$\therefore$ From (iv), we get

$-0.103448 - 1.25\bar{y} + 0.75 = 0 \Rightarrow \bar{y} = 0.51724$. Ans.

Now since the coefficient of correlation is the geometric mean between the two regression coefficients:

$\therefore r^2 = 1.25 \times 0.4375 = 0.546875 \Rightarrow r = 0.7395$. Ans.

Q.No.9.: In a partially destroyed laboratory record , only the lines of regression of y on x and x on y are available as $4x - 5y + 33 = 0$ and $20x - 9y = 107$ respectively. Calculate $\bar{x}$, $\bar{y}$ and co-efficient of correlation between x and y.

Sol.: The lines of regression of y on x $4x - 5y + 33 = 0$ (i)

and the lines of regression of x on y is $20x - 9y = 107$. (ii)

Since the regression lines pass through $(\overline{x}, \overline{y})$, therefore,

$4\overline{x} - 5\overline{y} + 33 = 0$, (iii) $20\overline{x} - 9\overline{y} = 107$ (iv)

Multiplying (iii) by and (iv) by 5 and subtracting , we get

$36\overline{x} - 45\overline{y} + 297 - 100\overline{x} + 45\overline{y} + 535 = 0 \Rightarrow -64\overline{x} + 832 = 0 \Rightarrow \overline{x} = 13$. Ans.

Substituting the value of $\overline{x}$ in (iii), we get

$52 - 5\overline{y} + 33 = 0 \Rightarrow 5\overline{y} = 85 \Rightarrow \overline{y} = 17$. Ans.

Also (i) $\Rightarrow y = \frac{4}{5}x + \frac{33}{5}$, we get $b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{4}{5}$. (v)

And (ii) $\Rightarrow x = \frac{9}{20}y + \frac{107}{9}$, we get $b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{9}{20}$. (v)

Since $r^2 = r\frac{\sigma_y}{\sigma_x} \times r\frac{\sigma_x}{\sigma_y} \Rightarrow r^2 = \frac{4}{5} \times \frac{9}{20} = 0.36 \Rightarrow r = \pm 0.6$.

Hence r = 0.6. Ans.

(Here we take the positive sign because b_{yx} and b_{xy} both are positive.)

Q.No.10.: In the following table are recorded data showing the test scores made by salesmen on an intelligence test and their weekly sales.

Salesmen	1	2	3	4	5	6	7	8	9	10
Test scores	40	70	50	60	80	50	90	40	60	60
Sales (000)	2.5	6.0	4.5	5.0	4.5	2.0	5.5	3.0	4.5	3.0

Calculate the regression line of sales on test scores and estimate the most probable weekly sales volume if a salesman makes a score of 70.

Sol.:

Test Scores x	Sales y	Deviation of x from assumed mean (=60) d_x	Deviation of y from assumed average (=4.5) d_y	$d_x \times d_y$	d_x^2	d_y^2
40	2.5	−20	−2	40	400	4
70	6.0	10	1.5	15	100	2.25

50	4.5	-10	0	0	100	0
60	5.0	0	0.5	0	0	2.25
80	4.5	20	0	0	400	0
50	2.0	-10	-2.5	25	100	6.25
90	5.5	30	1	30	900	1.00
40	3.0	-20	-1.5	30	400	2.25
60	4.5	0	0	0	0	0
60	3.0	0	-1.5	0	0	2.25
		$\sum d_x = 0$	$\sum d_y = -4.5$	$\sum d_x d_y = 140$	$\sum d_x^2 = 2400$	$\sum d_y^2 = 18.25$

With the help of the table, we have

$$\bar{x} = \text{mean of (test scores)} = 60 + \frac{\sum d_x}{10} = 60 + \frac{0}{10} = 60.$$

$$\bar{y} = \text{mean of y (sales)} = 4.5 + \frac{\sum d_y}{10} = 4.5 + \frac{(-4.5)}{10} = 4.05.$$

Regression line of sales (y) on score (x) is given by $y - \bar{y} = r\left(\frac{\sigma_y}{\sigma_x}\right)(x - \bar{x})$,

$$\text{where } r\frac{\sigma_y}{\sigma_x} = \frac{\sum XY}{\sigma_x \sigma_y} \times \frac{\sigma_y}{\sigma_x} = \frac{\sum XY}{(\sigma_x)^2} = \frac{\left[\sum d_x d_y - \frac{\sum d_x \sum d_y}{n}\right]}{\left[\sum d_x^2 - \frac{\left(\sum d_x\right)^2}{n}\right]}$$

$$= \frac{140 - \frac{0 \times (-4.5)}{10}}{2400 - \frac{0^2}{10}} = \frac{140}{2400} = 0.06.$$

$\therefore$ The required regression line is

$y - 4.05 = 0.06(x - 60) \Rightarrow y = 0.06x + 0.45$. Ans.

For $x = 70$, $y = 0.06 \times 70 + 0.45 = 4.65$. Ans.

Thus the most probable weekly sales volume for a score of 70 is 4.65.

Q. No.11.: Find the correlation coefficient between x and y for the given values. Find also the regression lines.

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

Sol.: Assume mean for x = 6, and for y = 36.

x	y	d_x	d_y	$d_x d_y$	d_x^2	d_y^2
1	10	-5	-26	130	25	676
2	12	-4	-24	96	16	576
3	16	-3	-20	60	9	400
4	28	-2	-8	16	4	64
5	25	-1	-11	11	1	121
6	36	0	0	0	0	0
7	41	1	5	5	1	25
8	49	2	13	26	4	169
9	40	3	4	12	9	16
10	50	4	14	56	16	196
$\sum$		$\sum d_x = -5$	$\sum d_y = -53$	$\sum d_x.d_y = 412$	$\sum d_x^2 = 85$	$\sum d_y^2 = 2243$

$$\therefore \ \overline{x} = x + \frac{\sum d_x}{10} = 6 + \left(\frac{-5}{10}\right) = 5.5 .$$

and $\overline{y} = y + \dfrac{\sum d_y}{10} = 36 + \dfrac{-53}{10} = 36 - 5.3 = 30.7$.

Also $r\dfrac{\sigma_y}{\sigma_x} = \dfrac{\sum(d_x d_y) - \dfrac{\sum d_x \sum d_y}{n}}{\sum d_x^2 - \dfrac{\left(\sum d_x\right)^2}{n}} = \dfrac{412 - \dfrac{(-5)\times(-53)}{10}}{85 - \dfrac{25}{10}} = \dfrac{385.5}{82.5} = 4.67$.

Hence the line of regression of y on x is $y - \overline{y} = r\dfrac{\sigma_y}{\sigma_x}(x - \overline{x})$

$\Rightarrow y - 30.7 = 4.67(x - 5.5) \Rightarrow y = 4.67x + 4.96$. Ans.

Now

$$r\frac{\sigma_x}{\sigma_y}=\frac{\left[\sum d_x d_y-\frac{\sum d_x \sum d_y}{n}\right]}{\left[\sum d_x^2-\frac{\left(\sum d_y\right)^2}{n}\right]}=\frac{\left[412-\frac{(-5)(-53)}{10}\right]}{\left[2243-\frac{(-53)^2}{10}\right]}=\frac{412-26.5}{2243-280.9}=\frac{385.5}{1962.1}=0.2.$$

Hence the line regression of y on x is $x-\bar{x}=(y-\bar{y})r\frac{\sigma_x}{\sigma_y}\Rightarrow x-5.5=0.2(y-30.7)$

$\Rightarrow x=0.2y-0.64$. Ans.

Now $r^2=r\frac{\sigma_y}{\sigma_x}\times r\frac{\sigma_x}{\sigma_y}=4.67\times 0.2=0.934\Rightarrow r=0.966$. Ans.

Q.No.12.: Following table gives the data on rainfall and discharge in a certain river.

Obtain the line of regression of y on x.

Rain fall x (inches)	15.3	1.78	2.60	2.95	3.42
Discharge y (1000cc)	33.5	36.3	40.0	45.8	53.5

Sol.: Let assumed mean of x is 2.6 and assumed mean of y is 40.

x	$d_x=x-\bar{x}$	y	$d_y=y-\bar{y}$	$d_x\times d_y$	d_x^2	d_y^2
1.53	-1.07	33.5	-6.5	6.955	1.1449	42.25
1.78	-0.82	36.3	$-6.5-3.7$	3.034	0.6724	13.69
2.60	0.0	40.0	0.0	0	0	0
2.95	0.35	45.8	5.8	2.03	0.1125	33.64
3.42	0.82	53.5	13.5	11.07	0.6724	182.25
$\sum x=12.28$	$\sum d_x=-0.72$	$\sum y=209.1$	$\sum d_y=9.1$	$\sum d_x d_y=23.089$	$\sum d_x^2=2.6122$	$\sum d_y^2=271.83$

$\bar{x}=$ Mean of x $=\frac{\sum x}{5}=2.456$ and $\bar{y}=$ Mean of y $=\frac{\sum y}{5}=41.82$.

$$\text{Now } r\frac{\sigma_x}{\sigma_y}=\frac{\sum d_x d_y-\frac{\sum d_x\sum d_y}{n}}{\sum d_x^2-\frac{\left(\sum d_x\right)^2}{n}}=\frac{23.0-\frac{(-0.72)\times(9.1)}{5}}{2.61122-\frac{(0.72)^2}{5}}=9.7266.$$

$\therefore$ The line regression of y on x is $y-\bar{y}=r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \Rightarrow y-41.82=9.7266(x-2.456)$

$\Rightarrow y=9.7266x+17.93$. Ans.

Q.No.13.: The following results were obtained from records of age (x) and blood pressure (y) of a group of 10 men:

$$\left.\begin{array}{lcc} & x & y \\ \text{Mean} & 53 & 142 \\ \text{Variance} & 130 & 165 \end{array}\right\} \text{ and } \sum (x-\bar{x})(y-\bar{y})=1220.$$

Find the appropriate regression equation and use it to estimate the blood pressure of a man whose age is 45.

Sol.: Given mean of age $\bar{x}=53$ and mean of blood pressure $\bar{y}=142$.

Also given variance of age $\sigma_x^2=130 \Rightarrow \sigma_x=11.401754$,

and variance of blood pressure $\sigma_y^2=165 \Rightarrow \sigma_y=12.845233$.

Now since $\frac{\sum (x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}=r \Rightarrow \frac{1220}{10(11.401754)(12.845233)}=r$

$\Rightarrow r=\frac{1220}{1464.5819}=0.8330022$.

The line of regression of y on x is $y-\bar{y}=r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \Rightarrow y-142-0.9384615(x-53)$

$\Rightarrow y=0.9384615x+92.26154$,

which is the required regression equation.

When $x=45$, i. e. age of a man is 45, then

Blood pressure $y=42.230768+92.26154=134.49231 \Rightarrow y=134.5$.

Thus the estimate blood pressure of a man whose age is 45 is 134.5. Ans.

Q.No.14.: Establish the formula $r=\frac{\sigma_x^2+\sigma_y^2-\sigma_{x-y}^2}{2\sigma_x\sigma_y}$.

Hence calculate r from the following data:

x :	21	23	30	54	57	58	72	78	87	90
y :	60	71	72	83	110	84	100	92	113	135

Sol.: (a). Let $z = x - y$ so that $\bar{z} = \bar{x} - \bar{y}$.

$$\therefore z - \bar{z} = (x - \bar{x}) - (y - \bar{y}) \Rightarrow (z - \bar{z})^2 = (x - \bar{x})^2 + (y - \bar{y}) - 2(x - \bar{x})(y - \bar{y})$$

Summing up for n terms, we have

$$\sum (z - \bar{z})^2 = \sum (x - \bar{x})^2 + \sum (y - \bar{y})^2 - 2\sum (x - \bar{x})(y - \bar{y})$$

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

$$\Rightarrow \sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y. \qquad \left[\therefore r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n\sigma_x\sigma_y}\right]$$

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

which is the required result.

(b). To find r, we have to calculate σ_x, σ_y and σ_{x-y}. We make the following table.

x	$X = x - 54$	X^2	y	$Y = y - 100$	Y^2	$y - x$	$(y - x)^2$
21	-33	1089	60	-40	1600	39	1521
23	-31	961	71	-29	841	48	2304
30	-24	576	72	-28	784	42	1764
54	0	0	83	-17	289	29	841
57	3	9	110	10	100	53	2809
58	4	16	84	-16	256	26	676
72	18	324	100	0	0	28	784
78	24	576	92	-8	64	14	196
87	33	1089	113	13	169	26	676
90	36	1296	135	35	1225	45	2025
Total	30	5936		-80	5328	350	13596

$$\therefore \sigma_x^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = \frac{5936}{10} - \left(\frac{30}{10}\right)^2 = 593.6 - 9 = 584.6$$

$$\sigma_y^2 = \frac{\sum Y^2}{N} - \left(\frac{\sum Y}{N}\right)^2 = \frac{5328}{10} - \left(\frac{-80}{10}\right)^2 = 532.8 - 64 = 468.8$$

$$\sigma_{x-y}^2 = \frac{\sum (x-y)^2}{N} - \left(\frac{\sum (x-y)}{N}\right)^2 = 1359.6 - 1225 = 134.6$$

$$\therefore r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y} = \frac{584.6 + 468.8 - 134.6}{2 \times 24.18 \times 21.65} = \frac{918.8}{1046.994} = 0.878. \text{ Ans.}$$

Q.No.15.: Using the formula $r = \dfrac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}$, find r from the following data:

x :	92	89	87	86	83	77	71	63	53	50
y :	86	88	91	77	68	85	52	82	37	57

Sol.: Let $z = x - y$, so that $\bar{z} = \bar{x} - \bar{y} \Rightarrow z - \bar{z} = (x - \bar{x}) - (y - \bar{y})$

$$\Rightarrow \sum (z - \bar{z})^2 = [(x - \bar{x}) - (y - \bar{y})]^2$$

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

$$\text{As } r = \frac{\sum (x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}$$

$$\Rightarrow \sigma_z^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y r \Rightarrow 2\sigma_x\sigma_r = \sigma_x^2 + \sigma_y^2 - \sigma_z^2$$

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

Mean of x is 75.1 and mean of y is 72.3.

x	$x - \bar{x} = X$	X^2	y	$y - \bar{y} = Y$	Y^2	$X - Y = Z$	Z^2
92	16.9	285.61	86	13.7	187.69	3.2	10.24
89	13.9	193.21	88	15.7	246.49	−1.8	3.24
87	11.9	141.61	91	18.7	349.69	−6.8	46.24
86	10.9	118.81	77	4.7	22.09	6.2	38.44
83	7.9	62.41	68	−4.3	18.49	12.2	148.84
77	1.9	3.61	85	12.7	161.29	−10.8	116.64

71	-4.1	16.81	52	-20.3	412.09	16.2	262.44
63	-12.1	146.41	82	9.7	94.09	-21.8	475.24
53	-22.1	488.41	37	-35.3	1246.09	13.2	174.24
50	-25.1	630.01	57	-15.3	234.09	-9.8	96.04
$\sum x = 751$		$\sum X^2 = 2086.9$	$\sum y = 723$		$\sum Y^2 = 2972.1$		$\sum Z^2 = 1371.6$

$$\sigma_x^2 = \frac{\sum X^2}{n} = \frac{2086.9}{10} = 208.69 \quad \therefore \sigma_x = 14.45.$$

$$\sigma_y^2 = \frac{\sum Y^2}{n} = \frac{2972.1}{10} = 297.21 \quad \therefore \sigma_y = 17.24.$$

$$\sigma_z^2 = \frac{\sum Z^2}{n} = \frac{1371.6}{10} = 137.16.$$

$$\therefore r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y} \frac{208.69 + 297.21 - 137.16}{2 \times (17.24) \times (14.45)} = 0.74. \text{ Ans.}$$

Q.No.16.: Compute the standard error of estimate S_x for the respective heights of the following 12 couples:

Height x of husband (inches)	68	66	68	65	69	66	68	65	71	67	68	70
Height y of wife (inches)	65	63	67	64	68	62	70	66	68	67	69	71

Sol.: Here $\bar{x} = \frac{\sum x}{12} = \frac{811}{12} = 67.5833$ and $\bar{y} = \frac{\sum y}{12} = \frac{800}{12} = 66.6667$.

x	$x - \bar{x} = X$	X^2	y	$y - \bar{y} = Y$	Y^2	$X - Y = Z$	Z^2
68	0.4167	0.1736	65	–1.6667	2.7778	2.0833	4.3403
66	– 1.5833	2.5069	63	–3.6667	13.4444	2.0833	4.3403
68	0.4167	0.1736	67	0.3333	0.1111	0.0833	0.0069
65	–2.5833	6.6736	64	–2.667	7.1111	0.0833	0.0069
69	1.4167	2.0069	68	1.3333	1.7778	0.0833	0.0069
66	–1.5833	2.5069	62	–4.6667	21.7778	3.08333	9.5069
68	0.4167	0.1736	70	3.3333	11.1111	–2.9167	8.5069
65	–2.5833	6.6736	66	–0.6667	0.4444	–1.9167	3.6736

71	3.4167	11.6736	68	1.3333	1.7778	2.0833	4.3403
67	–0.5833	0.3403	67	0.3333	0.1111	–0.9167	0.8403
68	0.4167	0.1736	69	2.3333	5.4444	–1.9167	3.6736
70	2.4167	5.8403	71	4.3333	18.7778	–1.9167	3.6736
$\sum x$ =811		$\sum X^2 = 38.9167$	$\sum y$ =800		$\sum Y^2 = 84.6667$		$\sum Z^2 = 42.9167$

$$\therefore \sigma_x^2 = \frac{\sum X^2}{n} = \frac{38.9169}{12} = 3.2431 \qquad \therefore \sigma_x = 1.8008.$$

$$\sigma_y^2 = \frac{\sum Y^2}{n} = \frac{84.6667}{12} = 7.0556 \qquad \therefore \sigma_y = 2.6562$$

$$\sigma_z^2 = \frac{\sum Z^2}{n} = \frac{42.9167}{12} = 3.5764 .$$

$$\text{Now } r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x\sigma_y} = \frac{3.2431 + 7.0556 - 3.5764}{2(1.8008)(2.6562)} = 0.7027 .$$

Hence standard error of estimate $S_x = \sigma_x\sqrt{1-r^2} = 1.8008\sqrt{1-(0.7027)^2}$

$= 1.28$. Inch. Ans.

Q.No.17.: While calculating correlation coefficient between two variables x and y from 25 pairs of observations, the following results were obtained: n = 25, $\sum x = 125$, $\sum x^2 = 650$, $\sum y = 100$, $\sum y^2 = 460$, $\sum xy = 508$. Later it was discovered at the time of checking that the pairs of values x : 8, 6. y: 12, 8 were copied down as x : 6, 8. y : 14, 6. Obtain the correct value of correlation coefficient.

Sol.: To get the correct results, we subtract the incorrect values and add the corresponding correct values.

$\therefore$ The correct result would be

$$\sum n = 25,\ \sum x = 125 - 6 - 8 + 8 + 6 = 125,\ \sum x^2 = 650 - 6^2 - 8^2 + 8^2 + 6^2 = 650$$

$$\sum y = 100 - 14 - 6 + 12 + 8 = 100,\ \sum y^2 = 460 - 14^2 - 6^2 + 12^2 + 8^2 = 436$$

$$\sum xy = 508 - 6\times14 - 8\times6 + 8\times12 + 6\times8 = 520.$$

$$\therefore r=\frac{n\sum xy-\left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[\left(n\sum x^2-\left(\sum x\right)^2\right)\left(n\sum y^2-\left(\sum y\right)^2\right)\right]}}=\frac{25\times 520-125\times 100}{\sqrt{\left[\left\{25\times 650-(125)^2\right\}\left\{25\times 436-(100)^2\right\}\right]}}$$

$$=\frac{500}{\sqrt{(625)(900)}}=\frac{500}{25\times 30}=\frac{2}{3}=0.667\text{ . Ans.}$$

Q.No.18.: Find the correlation between x (marks in Mathematics) and y (marks in Engineering Drawing) given in the following data.

y \ x	10-40	40-70	70-100	Total
0-30	5	20	-----	25
30-60	-----	28	2	30
60-90	-----	32	13	45
Total	5	80	15	100

Sol.:

Marks in E. D.			Marks in Maths (X)				Suppose $dx=(x-45)/10$ $dy=(y-55)/10$		
			10-40	40-70	70-100	Total f			
Marks group	Mid Pt. y	Mid pt. x	25	55	85		fd_y	fd_y^2	fd_xd_y
		dy \ dx	-3	0	3				
0-30	15	-3	**45** 5	**0** 20	-	**25**	-75	225	**45**
30-60	45	0	----	**0** 28	**0** 2	30	0	0	**0**
60-90	75	3	---	**0** 32	**117** 13	45	135	405	**117**
Total f			5	80	15	100 = n	60	630	**162**
fd_x			-15	0	45	30	Thick figures in small Sqs. Stand for fd_xd_y.		
fd_x^2			45	0	135	180			
fd_xd_y			**45**	**0**	**117**	**162**			

With the help of the above correlation table, we have

$$r = \frac{n\left(\sum fd_x d_y\right) - \left(\sum fd_x\right)\left(\sum fd_y\right)}{\sqrt{\left[\left\{n\sum fd_x^2 - \left(\sum fd_x\right)^2\right\} \times \left\{n\sum fd_y^2 - \left(\sum fd_y\right)^2\right\}\right]}}$$

$$= \frac{(100\times 162) - (30\times 60)}{\sqrt{(100\times 180 - 900)\times(100\times 630 - 3600)}} = \frac{14400}{31870.676} = 0.4518. \text{ Ans.}$$

Q.No.19.: Ten participants in a contest are ranked by two judges as follows :

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

Calculate the rank correlation coefficient ρ.

Sol.: If $d_i = x_i - y_i$, then $d_i = -5, 2, -4, 2, 2, 0, 1, -1, 2, 1$, $n = 10$

$\therefore \sum d_i^2 = 25+4+16+4+4+0+1+1+4+1 = 60$.

Hence $\rho = 1 - \dfrac{6\sum d_i^2}{n^3 - n} = 1 - \dfrac{6\times 60}{990} = 0.636$ (nearly). Ans.

Q.No.20.: Calculate the rank correlation co-efficient from the following data showing ranks of 10 students in two subjects:

Maths (x_i)	3	8	9	2	7	10	4	6	1	5
Physics (y_i)	5	9	10	1	8	7	3	4	2	6

Sol.: Here n = 10, $d_i = x_i - y_i$, then $d_i = -2, -1 -1, 1, -1, 3, 1, 2, -1, -1$

$\Rightarrow \sum d_i^2 = 4+1+1+1+1+9+1+4+1+1 \Rightarrow \sum d_i^2 = 24$.

Now since rank-correlation co-efficient $= 1 - \dfrac{6\sum d_i^2}{n^3 - n}$

$$= 1 - \frac{6\times 24}{990} = 1 - \frac{144}{990} = 1 - 0.1455 = 0.8545$$

Hence rank-correlation co-efficient = 0.8545. Ans.

*** *** *** *** ***

*** *** ***

***

Topic

Probability

Probability, Permutations and Combinations, Theorem of total probability, Conditional probability, Theorem of compound probability

Prepared by:

Dr. Sunil

NIT Hamirpur (HP)

Last updated on 30-10-2007

Probability:

Exhaustive: A set of events is said to be exhaustive, if it includes all the possible events.

For example in tossing a coin there are two exhaustive cases either a head or a tail and there is no third possibility.

Mutually exclusive: A set of events is said to be mutually exhaustive, if the occurrence of one of the events precludes (stop) the occurrence of all others.

For example, as in tossing a coin either head comes up or a tail and both can not happen at the same time. These are two mutually exhaustive cases.

Equally likely: A set of events is said to be equally likely, if one of the events cannot be expected to happen in preference to another.

For example, when a cubical die is thrown, the turnings up of the six different faces of the die are exhaustive, mutually exhaustive and equally likely.

Definition: If there are n exhaustive, mutually exclusive and equally likely outcomes of a random experiment and m of them are favourable to the happening of an event A, then the probability (P) of the happening of A is $P(A) = \frac{m}{n}$.

Since there are $n-m$ cases in which A will not happen, therefore, the chance of A not happening is q or $P(A')$ so that

$$q=\frac{n-m}{n}=1-\frac{m}{n}=1-p.$$

i. e. $P(A')=1-P(A)$ so that $P(A)+P(A')=1$, i. e. if an event is certain to happen then its probability is unity. While it is certain not to happen, its probability is zero.

Permutations and Combinations:

Definition: A permutation of a number of objects is their arrangement in some definite order. Given three letters a, b, c, we can permute them two at a time as follows:

bc, cb; ca, ac; ab, ba yielding 6 permutations. The combinations or selections or groupings are only 3 i. e. bc, ca, ab. Here the order is immetrial.

The number of **permutations** of n different things taken r at a time is $n(n-1)(n-2)............(n-r+1)$, which is denoted by ${}^{n}P_{r}$.

Thus ${}^{n}P_{r}=n(n-r)(n-2).........(n-r+1)=\dfrac{n!}{(n-r)!}$. (i)

The number of **combinations** of n different objects taken r at a time is denoted by ${}^{n}C_{r}$. If we take any one of the combinations, its r objects can be arranged in r ! ways. So the total number of arrangements which can be obtained from all the combinations is ${}^{n}P_{r}={}^{n}C_{r}.r!$.

Thus ${}^{n}C_{r}=\dfrac{n!}{r!(n-r)!}=\dfrac{n(n-r)(n-r)..........(n-r+1)}{r!}$ (ii)

Also ${}^{n}C_{n-r}={}^{n}C_{r}$ (iii)

e. g. ${}^{25}P_{4}=25\times24\times23\times22$; ${}^{25}C_{21}={}^{25}C_{4}=\dfrac{25\times24\times23\times22}{4\times3\times2\times1}$.

Remarks: It is often found convenient to use the following notations:

1. $P(A+B)$ or $P(A\cup B)$ stands for the probability of happening of **at least one** the events A and B.
2. $P(AB)$ or $P(A\cap B)$ stands for the probability of happening of **both** the events A and B.

Addition law of probability or Theorem of total probability:

Statement:

If the probability of an event happening as a result of a trial is P(A) and the probability of a mutually exclusive event B happening is P(B). then the probability of either of the events happening as a result of the trial is P(A+B) or $P(A \cup B) = P(A) + P(B)$.

Proof: Let n be the total number of equally likely cases and let m_1 be favourable to the event A and m_2 be favourable to the event B. Then the number of cases favourable to A or B is $m_1 + m_2$.

Hence the probability of A or B happening as a result of the trial

$$= \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B).$$

If the events A and B are not mutually exclusive, then, there are some outcomes which favour both A and B. If m_3 be their number, then these are included in both m_1 and m_2. Hence the total number of outcomes favouring either A or B or both is $m_1 + m_2 - m_3$.

Thus the probability P(A+B) or $P(A \cup B)$ of occurrence of A and B or both

$$= \frac{m_1 + m_2 - m_3}{n} = \frac{m_1}{n} + \frac{m_2}{n} - \frac{m_3}{n}$$

Hence $P(A + B) = P(A) + P(B) - P(AB)$

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

When A and B are mutually exclusive, then P(AB) or $P(A \cap B) = 0$ and we get

P(A+B) or $P(A \cup B) = P(A) + P(B)$

In general, for a number of mutually exclusive events $A_1, A_2, ... A_n$, we have

$$P(A_1 + A_2 + ... + A_n) \text{ or } P(A_1 \cup A_2 \cup .... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n).$$

This completes the proof.

Independent Events:

Definition: Two events are said to be independent, if happening or failure of one does not affect the happening or failure of other. Otherwise the events are said to be dependent.

Conditional probability:

For the dependent events A and B, the symbol $P\left(\frac{B}{A}\right)$ denotes the probability of occurrence of B, when A has already occurred. It is known as the conditional probability and is read as a 'probability of B given A'.

Multiplication law of probability or Theorem of compound probability:

Statement:

If the probability of an event A happening as a result of trial is P(A) and after A has happened the probability of an event B happening as a result of another trial (i. e. conditional probability of B given A) is $P\left(\frac{B}{A}\right)$, then the probability of both the events A and B happening as a result of two trials is P(AB) or $P(A \cap B) = P(A).P\left(\frac{B}{A}\right)$.

Proof: Let n be the total number of outcomes in the first trial and m be favourable to the event A so that $P(A) = \frac{m}{n}$.

Let n_1 be the total number of outcomes in the second trial of which m_1 are favourable to the event B so that $P\left(\frac{B}{A}\right) = \frac{m_1}{n_1}$.

Now each of the n outcomes can be associated with each of the n_1 outcomes. So the total number of outcomes in the combined trial is nn_1. Of these mm_1 are favourable to both the events A and B.

Hence $P(AB)$ or $P(A \cap B) = \frac{mm_1}{nn_1} = P(A).P\left(\frac{B}{A}\right)$.

Similarly, the conditional probability of A given B is $P\left(\frac{A}{B}\right)$

$\therefore P(AB)$ or $P(A \cap B) = P(B).P\left(\frac{A}{B}\right)$.

Thus $P)A \cap B) = P(A).P\left(\frac{B}{A}\right) = P(B).P\left(\frac{A}{B}\right)$.

If the events A and B are independent, i. e. if the happening of B does not depend on whether A has happened or not, then $P\left(\frac{B}{A}\right) = P(B)$ and $P\left(\frac{A}{B}\right) = P(A)$.

$\therefore P(AB)$ or $P(A \cap B) = P(A).P(B)$.

In general, $P(A_1\, A_2 \ldots\ldots\ldots A_n)$ or $P(A_1 \cap A_2 \cap \ldots\ldots\ldots \cap A_n) = P(A_1).P(A_2)\ldots\ldots\ldots .P(A_n)$.

Remarks: If p_1, p_2 be the probabilities of happening of two independent events, then

(i) the probability that the first event happens and the second fails is $p_1(1-p_2)$.

(ii) the probability that both events fail to happen is $(1-p_1)(1-p_2)$.

(iii) the probability that at least one of the events happens is $1-(1-p_1)(1-p_2)$. This is commonly known as their **cumulative probability**.

In general, if $p_1, p_2, p_3, \ldots\ldots\ldots p_n$ be the chances of happening of n independent events, then their cumulative probability (i. e. the chance that atleast one of the events will happen) is $1-(1-p_1)(1-p_2)(1-p_3)\ldots\ldots\ldots(1-p_n)$.

Now let us solve some problems:

Q.No.1.: Find the chance of throwing **(a)** four , **(b)** an even number with an ordinary six faced dice.

Sol.: (a): There are six possible ways in which the dice can fall and of these there is only one way of throwing 4. Thus the required chance $= \frac{1}{6}$. Ans.

(b): There are six possible ways in which the dice can fall. Of these there are only 3 ways of getting 2, 4 or 6. Thus the required chance $= \frac{3}{6} = \frac{1}{2}$. Ans.

Q.No.2.: What is the chance that a leap year selected at random will contain 53 Sundays?

Sol.: A leap year consists of 366 days, so that there are 52 full weeks (and hence 52 Sundays) and two extra days. These two days can be (i) Monday, Tuesday (ii) Tuesday, Wednesday (iii) Wednesday, Thursday (iv) Thursday, Friday (v) Friday, Saturday (vi) Saturday, Sunday (vii) Sunday, Monday.

Of these 7 cases, the last two are favourable and hence the required probability $= \frac{2}{7}$. Ans.

Q.No.3.: A five figure number is formed by the digits 0, 1, 2, 3, 4 without repetition. Find the probability that the number formed is divisible by 4.

Sol.: The five digits can be arranged in 5 ! ways, out of which 4 !will begin with zero.

$\therefore$ Total number of five-figure numbers formed $5! - 4! = 96$.

Those numbers formed will be divisible by 4 which will have two extreme right digits divisible by 4, i. e. numbers ending in 04, 12, 20. 24, 32, 40.

Now numbers ending in 04 = 3 ! = 6, numbers ending in $12 = 3! - 2! = 4$,

numbers ending in 20 = 3 ! = 6, numbers ending in $24 = 3! - 2! = 4$,

numbers ending in 32 = 3 ! = 6, numbers ending in $40 = 3! = 6$.

[The numbers having 12, 24, 32, in the extreme right are $(3! - 2!)$ since the numbers having zero on the extreme left are to excluded.]

$\therefore$ Total number of favourable ways = 6 + 4 + 6 + 4 + 4 + 6 =30.

Hence, the required probability $= \frac{30}{96} = \frac{5}{16} = 0.3125$. Ans.

Q.No.4.: A bag contains 40 tickets numbered 1, 2, 3,.............40, of which four are drawn at random and arranged in ascending order $(t_1 < t_2 < t_3 < t_4)$. Find the probability of t_3 being 25.

Sol.: Here exhaustive number of cases $= {}^{40}C_4$.

If $t_3 = 25$, then the tickets t_1 and t_2 must come out of 24 tickets numbered 1 to 24. This can be done in ${}^{24}C_2$ ways.

Then t_4 must come out of the 15 tickets (numbering 25 to 40) which can be done in ${}^{15}C_1$ ways.

$\therefore$ Favourable number of cases $= {}^{24}C_2 \times {}^{15}C_1$.

Hence, the probability of t_3 being 25 $= \frac{{}^{24}C_2 \times {}^{15}C_1}{{}^{40}C_4} = \frac{414}{9139} = 0.0453$. Ans.

Q.No.5.: An urn contains 5 red and 10 black balls. Eight of them are placed in another urn. What is the chance that the latter then contains 2 red and 6 black balls?

Sol.: The number of ways in which 8 balls can be drawn out of 15 is ${}^{15}C_8$.

The number of ways of drawing 2 red balls is ${}^{5}C_2$ and corresponding to each of these ${}^{5}C_2$ ways of drawing a red ball, there are ${}^{10}C_6$ ways of drawing 6 black balls.

$\therefore$ The total number of ways in which 2 red and 6 black balls can be drawn is ${}^{5}C_2 \times {}^{10}C_6$.

$\therefore$ The required probability $= \dfrac{{}^{5}C_2 \times {}^{10}C_6}{{}^{15}C_8} = \dfrac{2100}{6435} = 0.3263$. Ans.

Q.No.6.: A committee consist of 9 students two of which are from 1st year, three from 2nd year and four from 3rd year. Three student are removed at random. What is the chance that (i) the three students belong to different classes, (ii) two belong to the same class and third to the different class, (iii) the three belong to the same class?

Sol.: (i). The total number of ways of choosing 3 students out of 9 is

$${}^{9}C_3 = \frac{9\times 8\times 7}{3\times 2\times 1} = 84.$$

A student can be removed from 1st year students in two ways, from 2nd year in 3 ways and from 3rd year in 4 ways.

$\therefore$ Total ways of removing three students, one from each group $= 2\times 3\times 4 = 24$.

Hence, the required chance $= \dfrac{24}{84} = \dfrac{2}{7} = 0.2857$. Ans.

(ii). The number of ways of removing two from 1st year students and one from others

$= {}^{2}C_2 \times {}^{7}C_1 = 7$.

The number of ways of removing two from 2nd year students and one from others

$= {}^{3}C_2 \times {}^{6}C_1 = \dfrac{3\times 2}{2\times 1} \times \dfrac{6}{1} = 18$.

The number of ways of removing two from 3rd year students and one from others

$= {}^{4}C_2 \times {}^{5}C_1 = \dfrac{4\times 3}{2\times 1} \times \dfrac{5}{1} = 30$.

$\therefore$ The total number of ways in which two students of the same class and third from the others may be removed $= 7 + 18 + 30 = 55$.

Hence, the required chance $=\frac{55}{84}=0.6548$. Ans.

(ii). Three students can be removed from 2nd year group in ${}^3C_3 = 1$ way and from 3rd year group in ${}^4C_3 = 4$ ways.

$\therefore$ The total number of ways in which three students belong to the same class = 1 + 4 = 5.

Hence, the required chance $=\frac{5}{84}=0.0595$. Ans.

Q.No.7.: A has one share in a lottery in which there is 1 prize and 2 blanks; B has three shares in a lottery in which there are 3 prizes and 6 blanks; compare the probability of A's success to that of B's success.

Sol.: A can draw a ticket in ${}^3C_1 = 3$ ways.

The number of cases in which A can get a prize is clearly 1.

$\therefore$ The probability of A's success $=\frac{1}{3}$.

Again B can draw a ticket in ${}^9C_3 = \frac{9.8.7}{3.2.1} = 84$ ways.

The number of ways in which B gets all blanks $= {}^6C_3 = \frac{6.5.4}{3.2.1} = 20$.

$\therefore$ The number of ways of getting a prize $= 84 - 20 = 64$.

Thus, the probability of B's success $=\frac{64}{84}=\frac{16}{21}$.

Hence, A's probability of success : B's probability of success $=\frac{1}{3}:\frac{16}{21}=7:16$. Ans.

Q.No.8.: In a race, the odds in favour of the four horses H_1, H_2, H_3, H_4 are 1 : 4, 1 : 5, 1 : 6, 1 : 7 respectively. Assuming that a dead heat is not possible, find the chance that one of them wins the race.

Sol.: Since it is not possible for all the horses to cover the same distance in the same time (a dead heat), the events are mutually exclusive.

If p_1, p_2, p_3, p_4 be the probabilities of winning of the horses H_1, H_2, H_3, H_4 respectively, then

$p_1 = \frac{1}{1+4} = \frac{1}{5}$ [$\therefore$ Odds in favour of H_1 are 1 : 4]

$p_2 = \frac{1}{6}$, $p_3 = \frac{1}{7}$, $p_4 = \frac{1}{8}$.

Hence, the chance that one of them wins $= p_1 + p_2 + p_3 + p_4$

$$= \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} = \frac{533}{840} = 0.635 \text{. Ans.}$$

Q.No.9.: A bag contains 8 white balls and 6 red balls. Find the probability of drawing two balls of the same colour.

Sol.: Two balls out of 14 can be drawn in ${}^{14}C_2 = \frac{14\times13}{2\times1} = 91$ ways which is the total number of outcomes.

Two white balls out of 8 can be drawn in ${}^{8}C_2 = \frac{8\times7}{2\times1} = 28$ ways.

Thus the probability of drawing 2 white balls $= \frac{28}{91} = 0.3077$.

Similarly, 2 red balls out of 6 can be drawn in ${}^{6}C_2 = \frac{6\times5}{2\times1} = 15$ ways. Thus the probability of drawing 2 red balls $= \frac{15}{91} = 0.165$.

Hence the probability of drawing 2 balls of the same colour (either both white or both red) is $= \frac{28}{91} + \frac{15}{91} = \frac{43}{91} = 0.473$. Ans.

Q.No.10.: Find the probability of drawing an ace or a spade or both from a deck of cards?

Sol.: The probability of drawing an ace from a deck of 52 cards $= \frac{4}{52}$.

Similarly, the probability of drawing a card of spades $= \frac{13}{52}$, and the probability of drawing an ace of spades $= \frac{1}{52}$.

Since the two events (i. e. a card being an ace and a card being of spades) are not mutually exclusive, therefore, the probability of drawing an ace or a spade

$=\frac{4}{52}+\frac{13}{52}-\frac{1}{52}=\frac{4}{13}=0.308$. Ans.

Q.No.11.: Two cards are drawn in succession from a pack of 52 cards. Find the chance that first is a king and the second is a queen if the first card is **(i)** replaced, **(ii)** not replaced.

Sol.: (i). The probability of drawing a king $=\frac{4}{52}=\frac{1}{13}$.

If the card is replaced, the pack will again have 52 cards so that the probability of drawing a queen is $\frac{1}{13}$.

The two events being independent, the probability of drawing both cards in succession

$=\frac{1}{13}\times\frac{1}{13}=\frac{1}{169}=0.0059$. Ans.

(ii). The probability of drawing a king $=\frac{1}{13}$.

If the card is not replaced, the pack will have 51 cards so that the chance of drawing a queen is $\frac{4}{51}$.

Hence the probability of drawing both cards $=\frac{1}{13}\times\frac{4}{51}=\frac{4}{663}=0.006$. Ans.

Q.No.12.: A pair of dice is tossed twice. Find the probability of scoring 7 points **(a)** once**, (b)** atleast once, **(c)** twice.

Sol.: In a single toss of two dice, the sum 7 can be obtained as (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1) i. e. in six ways, so that the probability of getting 7 $=\frac{6}{36}=\frac{1}{6}$.

Also the probability of not getting 7 $=1-\frac{1}{6}=\frac{5}{6}$.

(a). The probability of getting 7 in the first toss and not getting 7 in the second toss

$=\frac{1}{6}\times\frac{5}{6}=\frac{5}{36}$.

Similarly, the probability of not getting 7 in the first toss and getting 7 in the second toss

$=\frac{5}{6}\times\frac{1}{6}=\frac{5}{36}$.

Since these are mutually exclusive events, addition law of probability applies.

$\therefore$ The required probability $=\frac{5}{36}+\frac{5}{36}=\frac{5}{18}=0.278$. Ans.

(b). The probability of not getting 7 in either toss $=\frac{5}{6}\times\frac{5}{6}=\frac{25}{36}$.

$\therefore$ The probability of getting 7 atleast once $=1-\frac{25}{36}=\frac{11}{36}=0.306$. Ans.

(c). The probability of getting 7 twice $=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36}=0.028$. Ans.

Q.No.13.: A box A contains 2 white and 4 black balls. Another box B contains 5 white and 7 black balls. A ball is transferred from the box A to the box B. Then a ball is drawn from the box B. Find the probability that it is white.

Sol.: The probability of drawing a white ball from box B will depend on weather the transferred ball is white or black.

If black ball is transferred, its probability is $\frac{4}{6}$. There are now 5 white and 8 black balls in the box B.

Then the probability of drawing white ball from box B is $\frac{5}{13}$.

Thus, the probability of drawing a white ball from urn B, if the transferred ball is black $=\frac{4}{6}\times\frac{5}{13}=\frac{10}{39}$.

Similarly, the probability of drawing a white ball from urn B, if the transferred ball is white $=\frac{2}{6}\times\frac{6}{13}=\frac{2}{13}$.

Hence, the required probability $=\frac{10}{39}+\frac{2}{13}=\frac{16}{39}=0.410$. Ans.

Q.No.14.: The odds that a book will be reviewed favourably by three independent critics are 5 to 2, 4 to 3 and 3 to 4. What is the probability that of the three reviews, a majority will be favourable.

Sol.: The probability that the book shall be reviewed favourably by first critic is $\frac{5}{7}$, by second $\frac{4}{7}$ and by third $\frac{3}{7}$.

A majority of the three reviews will be favourable when two or three are favourable.

$\therefore$ Probability that the first two are favourable and the third unfavourable

$$=\frac{5}{7}\times\frac{4}{7}\times\left(1-\frac{3}{7}\right)=\frac{80}{343}\ .$$

Probability that the first and third are favourable and the second unfavourable

$$=\frac{5}{7}\times\frac{3}{7}\times\left(1-\frac{4}{7}\right)=\frac{45}{343}\ .$$

Probability that the second and third are favourable and the first unfavourable

$$=\frac{4}{7}\times\frac{3}{7}\times\left(1-\frac{5}{7}\right)=\frac{24}{343}\ .$$

Finally, probability that the all the three are favourable

$$=\frac{5}{7}\times\frac{4}{7}\times\frac{3}{7}=\frac{60}{343}\ .$$

Since they are mutually exclusive events, the required probability

$$=\frac{80}{343}+\frac{45}{343}+\frac{24}{343}+\frac{60}{343}=\frac{209}{343}=0.609\text{. Ans.}$$

Q.No.15.: A can hit a target 3 times in 5 shots, B 2 times in 5 shots and C 3 times in 4 shots. They fire a volley. What is the probability that **(i)** two shots hits, **(ii)** At least two shots hit?

Sol.: Probability of A hitting the target $=\frac{3}{5}$, Probability of B hitting the target $=\frac{2}{5}$, Probability of C hitting the target $=\frac{3}{4}$.

(i). In order that two shots hit the target, the following cases must be considered:

$p_1 =$ Chance that A and B hit and C fails to hit $=\frac{3}{5}\times\frac{2}{5}\times\left(1-\frac{3}{4}\right)=\frac{6}{100}$.

$p_2 =$ Chance that B and C hit and A fails to hit $=\frac{2}{5}\times\frac{3}{4}\times\left(1-\frac{3}{5}\right)=\frac{12}{100}$.

$p_3 =$ Chance that C and A hit and B fails to hit $= \frac{3}{4} \times \frac{3}{5} \times \left(1 - \frac{2}{5}\right) = \frac{27}{100}$.

Since these are mutually exclusive events, the probability that any 2 shots hit

$= p_1 + p_2 + p_3 = \frac{6}{100} + \frac{12}{100} + \frac{27}{100} = 0.45$. Ans.

(ii). In order that atleast two shots may hit the target, we must also consider the case of all A, B, C hitting the target [in addition to three cases of (i)] for which

$p_4 =$ Chance that A, B, C all hit $= \frac{3}{5} \times \frac{2}{5} \times \frac{3}{4} = \frac{18}{100}$.

Since all these are mutually exclusive events, the probability of atleast 2 shots hit

$= p_1 + p_2 + p_3 + p_4 = \frac{6}{100} + \frac{12}{100} + \frac{27}{100} + \frac{18}{100} = 0.63$. Ans.

Q.No.16.: A problem in mechanics is given to three students A, B and C whose chances of solving it are $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ respectively. What is the probability that the problem will be solved.

Sol.: The probability that A can solve the problem is $\frac{1}{2}$.

The probability that A can not solve the problem is $1 - \frac{1}{2}$.

Similarly, the probabilities that B and C can not solve the problem are $1 - \frac{1}{3}$ and $1 - \frac{1}{4}$.

$\therefore$ The probability that A, B and C can not solve the problem is $\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{4}\right)$.

Hence, the probability that the problem will be solved, i. e. at least one student will solve it $= 1 - \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{4}\right) = \frac{3}{4} = 0.75$. Ans.

Q.No.17.: **(Huyghen's problem)** A and B throw alternately with a pair of dice. A wins if he throws 6 before B throws 7 and B wins if he throws 7 before A throws 6. If A begins, find his chance of winning.

Sol.: The sum 6 can be obtained as follows : (1, 5), (2, 4), (3, 3), (4, 2), (5, 1), i. e. in 5 ways.

The probability of A's throwing 6 with 2 dice is $\frac{5}{36}$.

$\therefore$ The probability of A's not throwing 6 is $1-\frac{5}{36}=\frac{31}{36}$.

Similarly, the probability of B's throwing 7 is $\frac{6}{36}=\frac{1}{6}$.

$\therefore$ The probability of B's not throwing 7 is $1-\frac{1}{6}=\frac{5}{6}$.

Now A can win if he throws 6 in the first, third , fifth, seventh etc. throws.

$\therefore$ The chance of A's winning

$$=\frac{5}{36}+\left(\frac{31}{36}\times\frac{5}{6}\times\frac{5}{36}\right)+\left(\frac{31}{36}\times\frac{5}{6}\times\frac{31}{36}\times\frac{5}{6}\times\frac{5}{36}\right)+\ldots\ldots\ldots$$

$$=\frac{5}{36}\left[1+\left(\frac{31}{36}\times\frac{5}{6}\right)+\left(\frac{31}{36}\times\frac{5}{6}\right)^2+\left(\frac{31}{36}\times\frac{5}{6}\right)^3+\ldots\ldots\right]$$

$$=\frac{5}{36}.\frac{1}{1-\left(\frac{31}{36}\right)\times\left(\frac{5}{6}\right)}=\frac{5}{36}\times\frac{36\times 6}{(216-155)}=\frac{30}{61}=0.492\text{. Ans.}$$

Q.No.18: Let A and B be two events with $P(A)=\frac{1}{2},\ P(B)=\frac{1}{3}$ and $P(A\cap B)=\frac{1}{4}$. Find

$$P\left(\frac{A}{B}\right),\ P(A\cup B),\ P\left(\frac{A'}{B'}\right).$$

Sol.: Given $P(A)=\frac{1}{2},\ P(B)=\frac{1}{3}$ and $P(A\cap B)=\frac{1}{4}$.

(i) Here A and B are two dependent events. Then $P\left(\frac{A}{B}\right)$ denotes the probability of occurrence of A, when B has already occurred. It is known as the conditional probability and is read as a "probability of A given B".

By using multiplication law of probability i.e. $P(AB)$ or $P(A\cap B)=P(B).P\left(\frac{A}{B}\right)$

Then $P\left(\frac{A}{B}\right)=\frac{P(A\cap B)}{P(B)}=\frac{1/4}{1/3}=\frac{3}{4}=0.75$. Ans.

(ii) Here A and B are not mutually exclusive, then

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

$$= \frac{1}{2} + \frac{1}{3} - \frac{1}{4} = \frac{6+4-3}{12} = \frac{7}{12} = 0.5833. \text{ Ans.}$$

(iii) Also $P\left(\frac{A'}{B'}\right) = \frac{P(A \cap B)}{P(B')} = \frac{P(A \cap B)}{1-P(B)} = \frac{\frac{1}{4}}{1-\frac{1}{3}} = \frac{3}{8} = 0.375$. Ans.

Q.No.19: In a single throw with two dice, what is the chance of throwing

(a) two aces ? **(b)** 7 ? Is this probability the same as that for getting 7 in two throws of a single die ?

Sol.: Total events are $6^2 = 36$.

(a) Number of events of getting two aces = 1 [i.e. (1, 1)]

Therefore, the probability of throwing two aces = $\frac{1}{36} = 0.0278$. Ans.

(b) Number of events of getting 7 in two throw = 6 [i.e. (1, 6), (2, 5), (3,4), (4, 3), (5, 2), (6, 1)]

Therefore, the probability of throwing 7 = $\frac{6}{36} = \frac{1}{6} = 0.1667$. Ans.

(c) Yes, because in that case number of events of getting 7 in two throws of a single die = 6 [i.e. (1, 6), (2, 5), (3,4), (4, 3), (5, 2), (6, 1)], which is same as number of events of getting 7 in two throw.

Q.No.20: Compare the chances of throwing 4 with one die, 8 with two dice 12 with three dice.

Sol.: With one die, number of events of getting 4 is 1.

Therefore, chances of getting 4 = $\frac{1}{6}$.

With two dice, total events are $6^2 = 36$ and number of events of getting 8 are 5 [i.e. (2, 6), (3, 5), (4, 4), (5, 3), (6, 2)]

Therefore, chances of getting 8 = $\frac{5}{36}$.

With three dice, total events are $6^3 = 216$ and number of events of getting 12 are 25 [i.e. (1,5,6), (1,6,5), (2,4,6), (2,6,4), (2,5,5), (3,4,5), (3,5,4), (3,6,3), (3,3,6), (4,4,4), (4,6,2), (4,2,6), (4,3,5), (4,5,3), (5,1,6), (5,6,1), (5,2,5), (5,5,2), (5,3,4), (5,4,3), (6,1,5), (6,5,1), (6,2,4), (6,4,2), (6,3,3)]

Therefore, chances of getting 12 $=\dfrac{25}{216}$.

Therefore, ratio of probabilities = $\dfrac{1}{6}:\dfrac{5}{36}:\dfrac{25}{216} \Rightarrow 36:30:25$. Ans.

Q.No.21: Find the probability that a non-leap year should have 53 Saturdays ?

Sol.: A non-leap year has 365 days.

Number of Saturdays in 365 days = $\dfrac{365}{7} = 52$ Saturdays + one day.

This day may be one of the following:

{Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday}

Hence, probability of Saturday = $\dfrac{1}{7}$. Ans.

Q.No.22: When a coin is tossed 4 times, find the probability of getting

(i) exactly one head,

(ii) at most three heads,

(iii) at least two heads.

Sol.: Total events are $2^4 = 16$.

(i). Exactly one head ={H,T,T,T}, total favorable events = $\dfrac{4!}{3!} = 4$.

Therefore, the probability of getting exactly one head $=\dfrac{4}{16}=\dfrac{1}{4}=0.25$. Ans.

(ii). At most three heads =[{H,T,T,T}, {H,H,T,T}, {H,H,H,T}].

Here total favorable events = $\dfrac{4!}{3!}+\dfrac{4!}{2!2!}+\dfrac{4!}{3!}=4+6+4=14$.

Therefore, the probability of getting at most three heads $=\dfrac{14}{16}=\dfrac{7}{8}=0.875$. Ans.

(iii). At least two heads =[{H,H,T,T}, {H,H,H,T}, {H,H,H,H}].

Here total favorable events = $\dfrac{4!}{2!2!}+\dfrac{4!}{3!1!}+\dfrac{4!}{4!}=6+4+1=11$.

Therefore, the probability of getting at least three heads $= \frac{11}{16} = 0.6875$. Ans.

Q.No.23: If all the letters of word "ENGINEER" be written at random, what is the probability that all letters 'E' are found together.

Sol.: Total possible arrangements $= \frac{8!}{3!2!} = \frac{40320}{126} = 3360$.

(EEE)NGINR

Total favorable events = $\frac{6!}{2!} = \frac{720}{2} = 360$.

Therefore, the probability that all letters 'E' are found together = $\frac{360}{3360} = \frac{9}{84} = \frac{3}{28}$. Ans.

Q.No.24: A ten digit number is formed using the digits from zero to nine, every digit being used only once. Find the probability that the number is divisible by 4.

Sol.: The ten digits can be arranged in 10 ! ways, out of which 9 ! will begin with zero.

$\therefore$ Total number of 10-figure numbers formed = 10 ! – 9 ! = 9. 9!

Thus, the total number of events = 9. 9!

Those numbers formed will be divisible by 4, which will have two extreme right digits divisible by 4, i.e. numbers ending in

(04, 08, 20, 40, 60, 80), (12, 16, 24, 28, 32, 36, 48, 52, 56, 64, 68, 72, 76, 84, 92, 96)

Thus total favorable events = (numbers of number where zero is in last two digits)
+ (numbers of number where zero is not in last two digits)
= 6. 8! + 16.7. 7! = 20.8!

Thus the probability that the number is divisible by 4 = $\frac{20\times 8!}{9\times 9!} = \frac{20}{81} = 0.2469$. Ans.

Q.No.25: From a pack of 52 cards three are drawn at random. Find the chance that they are king, a queen and a knave.

Sol.: Total number of events = ${}^{52}C_3 = \frac{52\times 51\times 50}{3\times 2\times 1} = 22100$.

Total favorable events = ${}^4C_1 \times {}^4C_1 \times {}^4C_1 = 4\times 4\times 4 = 64$.

Thus, the probability of this event $= \frac{64}{22100} = \frac{16}{5525} = 0.0029$. Ans.

Q.No.26: Four cards are drawn from a pack of 52 cards. What is the chance that

(i) no two cards are of equal value ?

(ii) each belong to a different suit ?

Sol.: Total number of events = ${}^{52}P_4 = 52\times51\times50\times49 = 6497400$.

(i). Total number of ways that no two cards are of equal value

$= {}^{52}C_1 + {}^{48}C_1 + {}^{44}C_1 + {}^{40}C_1 = 52\times48\times44\times40 = 4392960$.

Thus, the probability of this event $= \dfrac{52\times48\times44\times40}{52\times51\times50\times49} = \dfrac{4392960}{6497400} = \dfrac{2816}{4165}$

$= 0.6761104$. Ans.

(ii). Total number of ways that each belong to a different suit $= {}^{52}C_1 + {}^{39}C_1 + {}^{26}C_1 + {}^{13}C_1$

$= 52\times39\times26\times13 = 685464$.

Thus, the probability of this event $= \dfrac{52\times39\times26\times13}{52\times51\times50\times49} = \dfrac{2197}{20825} = 0.1054982$. Ans.

Q.No.27: Out of 50 rare books, 3 of which are especially valuable, 5 are stolen at random by a thief. What is the probability that

(a) none of the 3 is included ?

(b) 2 of the 3 are included ?

Sol.: Total number of events = ${}^{50}C_5 = \dfrac{50\times49\times48\times47\times46}{5\times4\times3\times2\times1} = 2118760$.

(a). Total number of ways that none of the 3 is included

$= {}^{47}C_5 = \dfrac{47\times46\times45\times44\times43}{5\times4\times3\times2\times1} = 1533939$.

Thus, the probability of this event = $\dfrac{{}^{47}C_5}{{}^{50}C_5} = \dfrac{1533939}{2118760} = 0.72398$. Ans.

(b). Total number of ways that2 of the 3 are included

$= {}^{3}C_2 \times {}^{47}C_3 = \dfrac{3\times2\times1}{2\times1}\times\dfrac{47\times46\times45}{3\times2\times1} = 48645$.

Thus, the probability of this event = $\dfrac{{}^{3}C_2\times{}^{47}C_3}{{}^{50}C_5} = \dfrac{48645}{2118760} = 0.0229591$. Ans.

Q.No.28: Five men in a company of twenty are graduates. If 3 men are picked out of 20 at random, what is the probability that

(a) they are all graduates ?

(b) at least one is graduate ?

Sol.: The number of ways in which 3 men are picked out of 20 at random is

$$^{20}C_3 = \frac{20\times19\times18}{3\times2\times1} = \frac{6840}{6} = 1140.$$

(a). Since five men in a company of twenty are graduates, therefore, the number of ways in which all 3 are graduates $= {}^5C_3 = \frac{5\times4\times3}{3\times2\times1} = \frac{60}{6} = 10$.

Thus, the probability that all 3 men are graduates $= \frac{10}{1140} = \frac{1}{114} = 0.00877$. Ans.

(b). Again, since five men in a company of twenty are graduates, therefore, the number of ways in which at least one is graduate $= {}^{20}C_3 - {}^{15}C_3 = 1140 - 455 = 685$.

Thus, the probability that at least one is graduate $= \frac{685}{1140} = \frac{137}{228} = 0.6009$. Ans.

Q.No.29: From 20 tickets marked from 1 to 20, one ticket is drawn at random. Find the probability that it is marked with a multiple of 3 or 5.

Sol.: Ticket marked with a multiple of 3 are 3, 6, 9, 12, 15, 18. (6 numbers)

Ticket marked with a multiple of 5 are 5, 10, 15, 20. (4 numbers)

Ticket marked with a multiple of both i.e. 3 and 5 are 1. (1 number)

Thus $P(3) = \frac{6}{20}$, $P(5) = \frac{4}{20}$, $P(3\cap5) = \frac{1}{20}$.

Therefore, probability that one ticket is marked with a multiple of 3 or 5

$$= P(3) + P(5) - P(3\cap5) = \frac{6}{20} + \frac{4}{20} - \frac{1}{20} = \frac{9}{20} = 0.45. \text{ Ans.}$$

Q.No.30: Five balls are drawn from a bag containing 6 white and 4 black balls what is the chance that 3 white and 2 black balls are drawn.

Sol.: The number of ways in which 5 balls are drawn from a bag containing 10 balls is

$$^{10}C_5 = \frac{10\times9\times8\times7\times6}{5\times4\times3\times2\times1} = \frac{30240}{120} = 252.$$

The number of ways in which 3 white balls are drawn from a bag containing 6 white balls is $^6C_3 = \frac{6\times5\times4}{3\times2\times1} = \frac{120}{6} = 20$.

The number of ways in which 2 black balls are drawn from a bag containing 4 black balls is ${}^4C_2 = \frac{4\times 3}{2\times 1} = 6$.

Therefore, the probability that 3 white and 2 black balls are drawn from a bag containing 6 white and 4 black balls $= \frac{20\times 6}{252} = \frac{10}{21} = 0.4762$. Ans.

Q.No.31: The probability of n independent events are $p_1, p_2, p_3, \ldots\ldots\ldots\ldots, p_n$. Find the probability that at least one of the events will happen. Use this result to find out the chance of getting at least one six in a throw of 4 dice.

Sol.: Given the probability of n independent events are $p_1, p_2, p_3, \ldots\ldots\ldots\ldots, p_n$.

Then the probability of not happening the events are $(1-p_1), (1-p_2), \ldots\ldots\ldots\ldots\ldots, (1 - p_n)$.

Thus the probability that no event would happen = $1-(1-p_1)(1-p_2)\ldots\ldots\ldots\ldots..(1 - p_n)$.

Second part:

Putting $p_1 = p_2 = p_3 = p_4 = \frac{1}{6}$, we get, the probability of getting at least one six in a throw of 4 dice $= 1-\left(1-\frac{1}{6}\right)\left(1-\frac{1}{6}\right)\left(1-\frac{1}{6}\right)\left(1-\frac{1}{6}\right) = 1-\left(1-\frac{1}{6}\right)^4 = 1-\left(\frac{5}{6}\right)^4 = 0.518$. Ans.

Q.No.32: Find the probability of drawing 4 white balls and 2 black balls without replacement from a bag containing 1 red, 4 black and 6 white balls.

Sol.: The number of ways in which 6 balls are drawn from a bag containing 11 balls is

$${}^{11}C_6 = \frac{11\times 10\times 9\times 8\times 7\times 6}{6\times 5\times 4\times 3\times 2\times 1} = \frac{332640}{720} = 462.$$

The number of ways in which 4 white balls are drawn from a bag containing 6 white balls is ${}^6C_4 = \frac{6\times 5\times 4\times 3}{4\times 3\times 2\times 1} = \frac{360}{24} = 15$.

The number of ways in which 2 black balls are drawn from a bag containing 4 black balls is ${}^4C_2 = \frac{4\times 3}{2\times 1} = \frac{12}{2} = 6$.

Therefore, the probability that 4 white and 2 black balls are drawn from a bag containing 1 red, 4 black and 6 white balls $= \frac{15\times 6}{462} = \frac{15}{77} = 0.1948$. Ans.

Q.No.33: A bag contains 10 white and 15 black balls. Two balls are drawn in succession. What id the probability that one of them is black and the other white ?

Sol.: The number of ways in which 2 balls are drawn from a bag containing 25 balls is

$^{25}C_2 = \frac{25 \times 24}{2 \times 1} = \frac{600}{2} = 300$.

The number of ways in which 1 white ball are drawn from a bag containing 10 white balls is $^{10}C_1 = 10$.

The number of ways in which 1 black ball are drawn from a bag containing 15 black balls is $^{15}C_1 = 15$.

Therefore, the probability that 1 white and 1 black balls are drawn from a bag containing 10 white and 15 black balls $= \frac{10 \times 15}{300} = \frac{1}{2} = 0.5$. Ans.

Q.No.34: A purse contains 2 silver and 4 copper coins and a second purse contains 4 silver and 4 copper coins. If a coin is selected at random from one of the two purses, what is the probability that it is silver coin ?

Sol.: Total number of ways of selecting purse = $^{2}C_1 = 2$.

Thus the probability of drawing silver coin from first purse = $\frac{^{2}C_1}{^{6}C_1} = \frac{2}{6} = \frac{1}{3}$.

Thus the probability of drawing silver coin from second purse = $\frac{^{4}C_1}{^{8}C_1} = \frac{4}{8} = \frac{1}{3}$.

As both the probabilities are mutually exclusive,

therefore, the required probability = $\left(\frac{1}{3} + \frac{1}{2}\right) \times \frac{1}{2} = \frac{5}{12} = 0.4167$. Ans.

Q.No.35: An urn contains 2 white and 2 black balls and a second urn contains 2 white and 4 black balls.

(a) If one ball is drawn from each urn what is the probability that they will be the same colour ?

(b) If an urn is selected at random and one ball is drawn from it, what is the probability that it will be white ?

Sol.: (a). Probability of choosing white ball from first urn = $\frac{{}^{2}C_1}{{}^{4}C_1} = \frac{2}{4} = \frac{1}{2}$.

Probability of choosing black ball from first urn = $\frac{{}^{2}C_1}{{}^{4}C_1} = \frac{2}{4} = \frac{1}{2}$.

Probability of choosing white ball from second urn = $\frac{{}^{2}C_1}{{}^{6}C_1} = \frac{2}{6} = \frac{1}{3}$.

Probability of choosing black ball from second urn = $\frac{{}^{4}C_1}{{}^{6}C_1} = \frac{4}{6} = \frac{2}{3}$.

Thus, the probability that all of same colour = $\left(\frac{1}{2}\times\frac{1}{3}\right)+\left(\frac{1}{2}\times\frac{1}{3}\right) = \frac{1}{6}+\frac{1}{3} = \frac{1}{2} = 0.5$. Ans.

(b). Total number of ways of selecting urn = ${}^{2}C_1 = 2$.

Thus, the probability of choosing white ball from first urn = $\frac{{}^{2}C_1}{{}^{4}C_1} = \frac{2}{4} = \frac{1}{2}$.

Thus, the probability of choosing white ball from second urn = $\frac{{}^{2}C_1}{{}^{6}C_1} = \frac{2}{6} = \frac{1}{3}$.

As both the probabilities are mutually exclusive,

therefore, the required probability = $\left(\frac{1}{2}+\frac{1}{3}\right)\times\frac{1}{2} = \frac{5}{12} = 0.4167$. Ans.

Q.No.36: A party of n persons take their seats at random at a round table; find the probability that two specified persons do not sit together.

Sol.: Let the 2 persons sit together so they form a group. Then the rest (n – 2) persons can sit in (n – 2) ! ways.

Thus 2 persons can sit in 2 ! ways and rest in (n – 2) ! ways.

Therefore, number of ways that two persons can sit together in round table = $2!\times(n-2)!$

Now in a round table n persons can sit in (n – 1) ! ways, because number of circular permutations of n things = (n – 1) !.

$\therefore$ The probability that two persons can sit together = $\frac{2!\times(n-2)!}{(n-1)!} = \frac{2}{n-1}$.

Thus, the probability that two persons can not sit together = $1-\frac{2}{n-1} = \frac{n-3}{n-1}$. Ans.

Q.No.37: A speaks the truth in 75% cases, and B in 80% of the cases. In what percentage of cases, are they likely to contradict each other in stating the same fact ?

Sol.: Let p(A) and p(B) be the probabilities of speaking the truth of persons A and B.

$\therefore \quad p(A)=\frac{3}{4}$, $q(A)=1-\frac{3}{4}=\frac{1}{4}$, $p(B)=\frac{4}{5}$, $q(B)=1-\frac{4}{5}=\frac{1}{5}$.

Thus, the probability that they contradict each other in stating the same fact

= Probability that A speaks truth and B fate & lie and vice-versa

= p(A).q(B) + p(B).q(A)

$$=\left(\frac{3}{4}\times\frac{1}{5}\right)+\left(\frac{4}{5}\times\frac{1}{4}\right)=\frac{3}{20}+\frac{4}{20}=\frac{7}{20}=0.35\text{ . Ans.}$$

Q.No.38: The probability that Sushil will solve a problem is $\frac{1}{4}$ and the probability that Ram will solve it is $\frac{2}{3}$. If Sushil and Ram work independently, what is the probability that the problem will solved by

(a) both of them,

(b) at least one of them ?

Sol.: Let p(S) and p(R) be the probabilities that the problem will be solved by Sushil and Ram, respectively.

Here $p(S)=\frac{1}{4},\ \ p(R)=\frac{2}{3}$.

(a). Since both the events are independent.

$\therefore$ The probability that the problem will be solved by both of them

$$=P(S\cap R)=P(S)\times P(R)=\frac{1}{4}\times\frac{2}{3}=\frac{1}{6}=0.1667\text{ . Ans.}$$

(b). Since both the events are not mutually exclusive.

$\therefore$ The probability that the problem will be solved by at least one of them

$$=P(S\cup R)=P(S)+P(R)-P(S\cap R)=\frac{1}{4}+\frac{2}{3}-\frac{1}{6}=\frac{9}{12}=\frac{3}{4}=0.75\text{ . Ans.}$$

Q.No.39: A student takes his examination in four subjects P, Q, R, S. He estimates his chances of passing in P as $\frac{4}{5}$, in Q as $\frac{3}{4}$, in R as $\frac{5}{6}$ and in S as $\frac{2}{3}$. To qualify, he must pass in P and at least two other subjects. What is the probability that he qualify ?

Sol.: Given $P(P) = \frac{4}{5},\ P(Q) = \frac{3}{4},\ P(R) = \frac{5}{6},\ P(S) = \frac{2}{3}$.

To qualify, he must pass in P and at least two other subjects. Here four cases arises

(i) When he passes in all four, then the probability $= \frac{4}{5}\times\frac{3}{4}\times\frac{5}{6}\times\frac{2}{3} = \frac{1}{3}$.

(ii) When he fails in Q, then the probability $= \frac{4}{5}\times\frac{1}{4}\times\frac{5}{6}\times\frac{2}{3} = \frac{1}{9}$.

(iii) When he fails in R, then the probability $= \frac{4}{5}\times\frac{3}{4}\times\frac{1}{6}\times\frac{2}{3} = \frac{1}{15}$.

(iv) When he fails in S, then the probability $= \frac{4}{5}\times\frac{3}{4}\times\frac{5}{6}\times\frac{1}{3} = \frac{1}{6}$.

$\therefore$ The required probability $= \frac{1}{3} + \frac{1}{15} + \frac{1}{9} + \frac{1}{6} = \frac{30+6+10+15}{90} = \frac{61}{90} = 0.6778$. Ans.

Q.No.40: The probability that a 50 year old man will be alive at 60 is 0.83 and the probability that a 45 year old woman will be alive at 55 is 0.87. What is the probability that a man who is 50 and his wife is 45 will both be alive 10 years hence ?

Sol.: The probability of aliving be P(A) and P(B) for 10 years.

Here $P(A) = 0.83,\ P(B) = 0.87$.

As both events are mutually independent,

$\therefore$ The probability that both A and B alive for 10 years

$= P(A \cap B) = P(A) \times P(B) = 0.83 \times 0.87 = 0.7221$. Ans.

Q.No.41: If on an average one birth in 80 is a case of twins, what is the probability that there will be at least one case of twins in a maternity hospital on a day when 20 births occur ?

Sol.: Probability that the twin occurs $= \frac{1}{80} = 0.0125$.

$\therefore$ Probability not having twins in one birth $= 1 - 0.0125 = 0.9875$.

In case of 20 births occurs in a day, then probability of occurring

$= (0.9875)^{20} = 0.7775746$.

$\therefore$ Probability of occurring at least one twin = $1 - 0.7775746 = 0.2224253 = 0.2224$. Ans.

Q.No.42: Two persons A and B fire at a target independently and have a probability 0.6 and 0.7 respectively of hitting the target. Find the probability that the target is destroyed.

Sol.: Let P(A) and P(B) be the probabilities of A and B for hitting the target.

Given P(A) = 0.6, P(B) = 0.7. As both events are not mutually independent,

$\therefore$ The probability that the target is destroyed

$= P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A).P(B) = 0.6 + 0.7 - (0.6 \times 0.7)$

=1.3 – 0.42 = 0.88. Ans.

Q.No.43: A manufacturer supplies cheap quarter horse-power motors in lots of 25. A buyer, before taking a lot, tests a random sample of 5 motors and accepts the lot if they buy are all good; otherwise he rejects the lot. Find the probability

(a) he will accept a lot containing 5 defective motors;

(b) he will reject a lot containing only one defective motor.

Sol.: Number of ways of selecting 5 motors out of 25 motors

$$= {}^{25}C_5 = \frac{25 \times 24 \times 23 \times 22 \times 21}{5 \times 4 \times 3 \times 2 \times 1} = 53130.$$

(a). Number of ways of selecting 5 motors out of 20 motors

$$= {}^{20}C_5 = \frac{20 \times 19 \times 18 \times 17 \times 16}{5 \times 4 \times 3 \times 2 \times 1} = 15504.$$

$\therefore$ Probability that of accepting 5 defecting motors

$$= \frac{{}^{20}C_5}{{}^{25}C_5} = \frac{15504}{53130} = 0.2918125 = 0.292. \text{ Ans.}$$

(b). Number of ways of selecting 5 motors out of 24 good motors

$$= {}^{24}C_5 = \frac{24 \times 23 \times 22 \times 21 \times 20}{5 \times 4 \times 3 \times 2 \times 1} = 42504.$$

$$\therefore \text{ Probability that of selecting} = \frac{{}^{24}C_5}{{}^{25}C_5} = \frac{42504}{53130} = 0.8.$$

$\therefore$ Probability that of rejecting $= 1 - 0.8 = 0.2$. Ans.

Q.No.44: A and B throw alternately with a pair of dice. The one who throws 9 first wins. Show that the chances of their winning are 9 : 8.

Sol.: Total number of ways, when A and B throw alternately with a pair of dice

$= 6\times 6 = 36$.

The probability that A throws 9 first = $\frac{4}{36} = \frac{1}{9}$. {(3, 6), (4, 5), (5, 4), (6, 3)}

The probability that A does not throws 9 first = $1 - \frac{1}{9} = \frac{8}{9}$.

Similarly, the probability that B does not throws 9 first = $\frac{8}{9}$.

If A stares tossing, then the probability of A winning is

$$= \frac{1}{9} + \left(\frac{8}{9}\times\frac{8}{9}\times\frac{1}{9}\right) + \left(\frac{8}{9}\times\frac{8}{9}\times\frac{8}{9}\times\frac{8}{9}\times\frac{1}{9}\right) + \ldots\ldots.$$

$$= \frac{1}{9}\left[1 + \left(\frac{8}{9}\right)^2 + \left(\frac{8}{9}\right)^4 + \ldots\ldots\right] = \frac{1}{9}\left[\frac{1}{1-\left(\frac{8}{9}\right)^2}\right] = \frac{1}{9}\left[\frac{81}{81-64}\right] = \frac{9}{17}.$$

The probability of winning B = The probability of losing A = $1 - \frac{9}{17} = \frac{8}{17}$.

Then the ratio is $\frac{\left(\frac{9}{17}\right)}{\left(\frac{8}{17}\right)} = 9:8$.

Hence the chances of winning A and B are 9 : 8.

Hence prove.

*** *** *** *** ***

*** *** ***

***

Topic

Binomial Distribution

Binomial probability distribution or Bernoulli's distribution, Probability of r successes in n trials, Recurrence or Recursion formula, Mean and Variance

Prepared by:

Dr. Sunil

NIT Hamirpur (HP)

Last updated on 30-10-2007

Binomial Distribution

This distribution was discovered by a Swiss mathematician Jacob Bernoulli in 1713. It is concerned with trials of a repetitive nature in which only the occurrence or non-occurrence, success or failure, acceptance or rejection, yes or no of a particular event is of interest. This distribution is applied to problems concerning:

- **Number of defectives in a sample from production line,**
- **Estimation of reliability of system,**
- **Number of rounds fired from a gun hitting a target,**
- **Radar detection.**

Binomial probability distribution or Bernoulli's distribution:

Let there be n independent trials in an experiment. Let a random variable X denote the number of successes in these n trials. Let p be the probability of a success and q that of a failure in a single trial so that $p + q = 1$. Let the trial be independent and q be constant for every trial.

Probability of r successes in n trials:

Since r success can be obtained in n trials in ${}^nC_r$ ways.

$$\therefore P(X=r) = {}^nC_r P\,(\underbrace{S\,S\,S.......S}_{r\ \text{times}}\ \ \underbrace{F\,F\,F.........F}_{(n-r)\,\text{times}})$$

$$= {}^nC_r \underbrace{P(S)\,P(S)\,P(S).......P(S)}_{r\ \text{factors}}\ \ \underbrace{P(F)\ P(F).........P(F)}_{(n-r)\,\text{factors}}$$

$$= {}^nC_r \underbrace{P\,P\,P.......P}_{n\ \text{factors}}\ \ \underbrace{q\,q\,q..........q}_{(n-r)\,\text{factors}}$$

$$= {}^nC_r p^r q^{n-r} \qquad \text{(i)}$$

Hence $P(X=r) = {}^nC_r q^{n-r} P^r$, where $p+q=1$ and $r = 0, 1, 2, ...,n$.

The distribution (i) is called the binomial probability distribution and X is called the binomial variate.

Note 1. P(X=r) is usually written as P(r).

2. The successive probability P(r) in (i) for r = 0, 1, 2, 3,.......,n are

$${}^nC_0 q^n,\ {}^nC_1 q^{n-1} p,\ {}^nC_2 q^{n-2} p^2, ... {}^nC_n p^n,$$

which are the successive terms of the binomial expansion of $(q+p)^n$. That is why this distribution is called "binomial" distribution.

3. n and p occurring in the binomial distribution are called the parameters of the distribution.

4. In a binomial distribution:

(i) n, the number of trials is finite.

(ii) each trial has only two possible outcomes usually called success and failure.

(iii) all the trials are independent.

(iv) p (and hence q) is constant for all the trials.

Recurrence or Recursion formula for the binomial distribution:

In a binomial distribution,

$$P(r) = {}^nC_r q^{n-r} p^r = \frac{n!}{(n-r)!\,r!} q^{n-r} p^r$$

$$P(r+1) = {}^nC_{r+1} q^{n-r-1} p^{r+1} = \frac{n!}{(n-r-1)!\,(r+1)!} q^{n-r-1} p^{r+1}$$

$$\therefore \frac{P(r+1)}{P(r)} = \frac{(n-r)!}{(n-r-1)!} \times \frac{r!}{(r+1)!} \times \frac{p}{q} = \frac{(n-r)\times(n-r-1)!}{(n-r-1)!} \times \frac{r!}{(r+1)\times r!} \times \frac{p}{q} = \frac{n-r}{r+1}.\frac{p}{q}$$

$$\Rightarrow P(r+1) = \frac{n-r}{r+1}.\frac{p}{q}P(r),$$

which is the required recurrence formula. Applying this formula successively, we can find P(1), P(2), P(3),…, if P(0) is known.

Constants of the binomial distribution:

Mean and Variance of the binomial distribution:

For the binomial distribution, $P(r)={}^{n}C_{r}q^{n-r}p^{r}$.

$$\text{Mean } \mu = \sum_{r=0}^{n} rP(r) = \sum_{r=0}^{n} r.{}^{n}C_{r}q^{n-r}p^{r}$$

$$= 0 + 1.{}^{n}C_{1}q^{n-1}p + 2.{}^{n}C_{2}q^{n-2}p^{2} + 3.{}^{n}C_{3}q^{n-3}p^{3} + \ldots\ldots\ldots\ldots\ldots + {}^{n}C_{n}p^{n}$$

$$= nq^{n-1}p + 2.\frac{n(n-1)}{2.1}q^{n-2}p^{2} + 3.\frac{n(n-1)(n-2)}{3.2.1}q^{n-3}p^{3} + \ldots\ldots\ldots + n.p^{n}$$

$$= nq^{n-1}p + n(n-1)q^{n-2}p^{2} + \frac{n(n-1)(n-2)}{2.1}q^{n-3}p^{3} + \ldots\ldots + np^{n}$$

$$= np\left[q^{n-1} + (n-1)q^{n-2}p + \frac{(n-1)(n-2)}{2.1}q^{n-3}p^{2} + \ldots\ldots\ldots p^{n-1}\right]$$

$$= np\left[{}^{n-1}C_{0}q^{n-1} + {}^{n-1}C_{1}q^{n-2}p + {}^{n-1}C_{2}q^{n-3}.p^{2} + \ldots\ldots + {}^{n-1}C_{n-1}p^{n-1}\right]$$

$$= np(p+p)^{n-1} = np. \qquad [\because p+q=1]$$

Hence the mean of the binomial distribution is np.

$$\text{Variance } \sigma^{2} = \sum_{r=0}^{n} r^{2}P(r) - \mu^{2} = \sum_{r=0}^{n} [r + r(r-1)]P(r) - \mu^{2}$$

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

(since the contribution due to r = 0 and r = 1 is zero)

$$= \mu + \left[2.1.{}^{n}C_{2}q^{n-2}p^{2} + 3.2.{}^{n}C_{3}q^{n-3}p^{3} + \ldots\ldots\ldots + n(n-1){}^{n}C_{n}p^{n}\right] - \mu^{2}$$

$$= \mu + \left[2.1.\frac{n(n-1)}{2.1}q^{n-2}p^{2} + 3.2.\frac{n(n-1)(n-2)}{3.2.1}q^{n-3}p^{3} + \ldots\ldots + n(n-1)p^{n}\right]$$

$$= \mu + \left[n(n-1)q^{n-2}p^2 + n(n-1)(n-2)q^{n-3}p^3 + \ldots\ldots\ldots\ldots + n(n-1)p^n\right] - \mu^2$$

$$= \mu + n(n-1)p^2\left[q^{n-2} + (n-2)q^{n-3}p + \ldots\ldots. + p^{n-2}\right] - \mu^2$$

$$= \mu + n(n-1)p^2\left[{}^{n-2}C_0 q^{n-2} + {}^{n-2}C_1 q^{n-3}p + \ldots\ldots\ldots + {}^{n-2}C_{n-2}p^{n-2}\right] - \mu^2$$

$$= \mu + n(n-1)p^2(q+p)^{n-2} - \mu^2 = \mu + n(n-1)p^2 - \mu^2 \qquad [\because q + p = 1]$$

$$= np + n(n-1)p^2 - n^2p^2 \qquad [\because \mu = np]$$

$$= np[1 + (n-1)p - np] = np[1-p] = npq\,.$$

Hence the variance of the binomial distribution is npq.

Standard deviation of the binomial distribution is $\sqrt{npq}$.

The moment generating function about the origin is

$$M_0(t) = E\left(e^{tx}\right) = \sum {}^{n}C_x p^x q^{n-x} e^{tx} = \sum {}^{n}C_x \left(pe^t\right)^x q^{n-x} = \left(q + pe^t\right)^n.$$

Differentiating w. r. t. to t and putting t = 0, we get the mean $\mu_1' = np$.

Since $M_a(t) = e^{-at}M_0(t)$, the m. g. f. of the binomial distribution about its mean $(m) = np$, is given by

$$M_m(t) = e^{-npt}\left(q + pe^t\right)^n = \left(qe^{-pt} + pe^{qt}\right)^n$$

$$= \left(1 + pq\frac{t^2}{2!} + pq\left(q^2 - p^2\right)\frac{t^3}{3!} + pq\left(q^3 - p^3\right)\frac{t^4}{4!} + \ldots\ldots\ldots\ldots\right)^n$$

$$\Rightarrow 1 + \mu_1 t + \mu_2\frac{t^2}{2!} + \mu_3\frac{t^3}{3!} + \mu_4\frac{t^4}{4!} + \ldots\ldots\ldots\ldots\ldots\ldots$$

$$= 1 + npq\frac{t^2}{2!} + npq(q-p)\frac{t^3}{3!} + npq[1 + 3(n-2)pq]\frac{t^4}{4!} + \ldots\ldots\ldots$$

Equating the coefficient of like powers of t on either side, we have

$\mu_2 = npq,\ \mu_3 = npq(q-p),\ \mu_4 = npq[1 + 3(n-2)pq]$.

Also $\beta_1 = \dfrac{\mu_2^3}{\mu_2^3} = \dfrac{(q-p)^2}{npq} = \dfrac{(1-2p)^2}{npq}$ and $\beta_2 = \dfrac{\mu_2^3}{\mu_2^3} = 3 + \dfrac{1-6pt}{npq}$.

Thus mean = np, standard deviation = $\sqrt{(npq)}$

Skewness $= \dfrac{(1-2p)}{\sqrt{(npq)}}$, kurtosis $= \beta_2$.

Remarks: The skewness is positive for $p < \dfrac{1}{2}$ and negative for $p > \dfrac{1}{2}$. When $p = \dfrac{1}{2}$, the skewness is zero, i. e. the probability curve of the binomial distribution will be symmetrical.

As the number of the trials increase indefinitely, $\beta_1 \to 0$ and $\beta_3 \to 3$.

Now let us solve some problems for better illustration of the Binomial distribution:

Q.No.1.: The probability of that a pen manufactured by a company will be defective is $\dfrac{1}{10}$. If 12 such pens are manufactured , find the probability that

(a) exactly two will be defective.

(b) at least two will be defective.

(c) none will be defective.

Sol.: The probability of a defective pen is $\dfrac{1}{10} = 0.1$

$\therefore$ The probability of a non-defective pen is $1 - 0.1 = 0.9$

(a) The probability that exactly two will be defective

$$= {}^{12}C_2 (0.1)^2 (0.9)^{10} = \frac{12 \times 11}{2 \times 1}(0.01)(0.3487) = 0.2301 \text{. Ans.}$$

(b) The probability that at least two will be defective

$$= 1 - \left[{}^{12}C_0 (0.1)^0 (0.9)^{12} + {}^{12}C_1 (0.1)(0.9)^{11}\right] = 1 - \left[0.28243 + (12 \times 0.1 \times 0.31381)\right].$$

$$= 1 - [0.28243 + 0.37657] = 1 - 0.659 = 0.341 \text{. Ans.}$$

(c) The probability that none will be defective $= {}^{12}C_{12} (0.9)^{12} = 0.28243$. Ans.

Q.No.2.: In sampling a large number of parts manufactured by a machine, the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples, how many would be expected to contain at least 3 defective parts.

Sol.: Mean number of defectives = 2 = np = 20p.

$\therefore$ The probability of defective part is $p = \dfrac{2}{20} = 0.1$,

and the probability of a non-defective part = 0.9.

$\therefore$ The probability of at least three defectives in a sample of 20

$$= 1 - \left[{}^{20}C_0 (0.9)^{20} + {}^{20}C_1 (0.1)(0.9)^{19} + {}^{20}C_2 (0.1)^2 (0.9)^{18} \right]$$

$$= 1 - [0.121577 + 0.27017 + 0.28518] = 1 - 0.676927 = 0.323073$$

Thus the number of samples having at least three defective parts out of 1000 samples

$= 1000 \times 0.323 = 323$. Ans.

Q.No.3.: The following data are the number of seeds germinating out of 10 on damp filter paper for 80 sets of seeds. Fit a binomial distribution to these data .

x:	0	1	2	3	4	5	6	7	8	9	10
f:	6	20	28	12	8	6		0	0	0	0

Sol.: Here n = 10 and $N = \sum f_i = 80$

$$\therefore \text{Mean } = \frac{\sum f_i x_i}{\sum f_i} = \frac{20 + 56 + 36 + 32 + 30}{80} = \frac{174}{80} = 2.175$$

Now the mean of a binomial distribution = np

i. e., $np = 10p = 2.175$ $\therefore p = 0.2175,$ $q = 1 - p = 0.7825$

Hence the binomial distribution to be fitted is

$$N(q + p)^n = 80(0.7825 + 0.2175)^{10}$$

$$= 80. \begin{bmatrix} {}^{10}C_0 (.7825)^{10} + {}^{10}C_1 (.7825)^9 (.2175)^1 + {}^{10}C_2 (.7825)^8 (.2175)^2 + {}^{10}C_3 (.7825)^7 (.2175)^3 \\ + {}^{10}C_4 (.7825)^6 (.2175)^4 + {}^{10}C_5 (.7825)^5 (.2175)^5 + {}^{10}C_6 (.7825)^4 (.2175)^6 \\ {}^{10}C_7 (.7825)^3 (.2175)^7 + {}^{10}C_8 (.7825)^2 (.2175)^8 + {}^{10}C_9 (.7825)^1 (.2175)^9 + {}^{10}C_{10} (.2175)^{10} \end{bmatrix}$$

$$= 6.885 + 19.139 + 23.94 + 17.74 + 8.63 + 2.88 + 0.67 + 0.11 + 0.011 + 0.0007 + 0.00002$$

$\therefore$ The successive terms in the expansion give the expected or theoretical frequencies which are;

x:	0	1	2	3	4	5	6	7	8	9	10
f:	6.9	19.1	24.0	17.8	8.6	2.9	.7	0.1	0	0	0

Q.No.4.: An ordinary six-faced die is thrown four times. What are the probabilities of obtaining 4, 3, 2, 1 and 0 faces, having same number ?

Sol.: There are six possible ways in which the die can fall and of these there is only one way of throwing any number .

Thus the probability of occurrence of a particular number = $p = \frac{1}{6}$, then $q = 1 - \frac{1}{6} = \frac{5}{6}$.

Since this problem is concerned with trials of a repetitive nature in which only the occurrence or non-occurrence of a particular event is of interest, so we use Binomial distribution and in this distribution the probability of r successes in a series of 4 trials is given by

$$ {}^{n}C_{r}p^{r}q^{n-r} = {}^{4}C_{r}\left(\frac{1}{6}\right)^{r}\left(\frac{5}{6}\right)^{4-r}, \text{ where } r = 0, 1, 2, 3, 4. $$

Q.No.5.: If the chance that one of the ten telephone lines is busy at an instant is 0.2.

(a) What is the chance that 5 of the lines are busy ?

(b) What is the most probable number of busy lines and what is the probability of this number ?

(c) What is the probability that all the lines are busy ?

Sol.: In this problem,

the chance that one of the ten telephone lines is busy at an instant is $p = 0.2$

$\therefore q = 1 - 0.2 = 0.8$.

Since this problem is concerned with trials of a repetitive nature in which only the occurrence or non-occurrence of a particular event is of interest, so we use Binomial distribution and in this distribution the probability of r successes in a series of n trials is given by ${}^{n}C_{r}p^{r}q^{n-r}$.

(a). Here n = 10, r = 5, then

the chance that 5 of the lines are busy $= {}^{n}C_{r}p^{r}q^{n-r} = {}^{10}C_{5}(0.2)^{5}(0.8)^{10-5}$

$$ = \frac{10\times 9\times 8\times 7\times 6}{5\times 4\times 3\times 2\times 1}\times(0.0032)(0.32768) = 0.026424. \text{ Ans.} $$

(b). 0.04571. Ans.

(c). Here n = 10, r = 10, then

the probability that all the lines i.e. 10 are busy $= {}^{n}C_{r}p^{r}q^{n-r} = {}^{10}C_{10}(0.2)^{10}(0.8)^{10-10}$

$$= 1 \times (1.024) \times 10^{-7} = 1.024 \times 10^{-7}. \text{ Ans.}$$

Q.No.6.: If the probability that a new-born child is a male is 0.6, find the probability that in a family of 5 children there are exactly 3 boys.

Sol.: Given the probability that a new-born child is a male = p = 0.6

$\therefore q = 1 - 0.6 = 0.4$.

From the Binomial distribution the probability of r successes in a series of n trials is given by ${}^{n}C_{r}p^{r}q^{n-r}$.

Thus the probability that in a family of 5 children (i.e. n = 5) there are exactly 3 boys (i.e. r = 3)

$$= {}^{n}C_{r}p^{r}q^{n-r} = {}^{5}C_{3}(0.6)^{3}(0.4)^{5-3} = \frac{5 \times 4}{2 \times 1} \times (0.216) \times (0.16) = 0.3456. \text{ Ans.}$$

Q.No.7.: If on an average 1 vessel in every 10 is wrecked, find the probability that out of 5 vessels expected to arrive, at least 4 will arrive safely.

Sol.: The probability of a wrecked vessel is $p = \frac{1}{10} = 0.1$.

$\therefore$ The probability of a non wrecked vessel is $q = 1 - \frac{1}{10} = \frac{9}{10} = 0.9$.

$\therefore$ The probability of atleast 4 will arrive safely

= The probability of 4 will arrive safely + the probability of 5 will arrive safely

$$= {}^{5}C_{4}(0.9)^{4}(0.1)^{1} + {}^{5}C_{5}(0.9)^{5}(0.1)^{0}$$

$$= (0.9)^{4}[0.5 + 0.9] = (0.9)^{4}.(1.4) = (0.6561).(1.4) = 0.91854. \text{ Ans.}$$

Q.No.8.: The probability that a bomb dropped from a plane will strike the target is 1/5. If six bombs are dropped, find the probability that

(i) exactly two will strike the target,

(ii) at least two will strike the target.

Sol.: Given that the probability that a bomb dropped from a plane will strike the target is 1/5. i.e. $p = 0.2$.

Therefore, the probability that a bomb not to strike the target i.e. q = 1-0.2 = 0.8

(i). Thus the probability that exactly two (i.e. r = 2) will strike the target, when six (i.e. n = 6) bombs are dropped

$$= {}^{n}C_{r}p^{r}q^{n-r} = {}^{6}C_{2}(0.2)^{2}(0.8)^{6-2} = \frac{6\times5}{2\times1}\times(0.04)\times(0.4096) = 0.24576 \text{. Ans.}$$

(ii). The probability that at least two (i.e. r = 2) will strike the target, when six (i.e. n = 6) bombs are dropped

= 1- (Probability that non or one will strike the target)

$$= 1-\left[{}^{6}C_{0}(0.2)^{0}(0.8)^{6}+{}^{6}C_{1}(0.2)^{1}(0.8)^{6-1}\right] = 1-[0.262144+6\times(0.2)\times(0.32768)]$$
$$= 1-0.65536 = 0.34464 \text{. Ans.}$$

Q.No.9.: A sortie of 20 aeroplane is sent on an operational flight. The chances that an aeroplane fails to return is 5%. Find the probability that

(i) one plane does not return,

(ii) at the most five planes do not return, and

(iii) what is the most probable number of returns ?

Sol.: The probability that an aeroplane does not return $= 5\% = \frac{5}{100} = \frac{1}{20} = 0.05$

So the probability of return $= 1-0.05 = 0.95$

(i) The probability that one plane does not return $= {}^{20}C_{1}(0.05)^{1}(0.95)^{19} = 0.377$. Ans.

(ii) The probability that at the most five planes do not return

$$= \sum_{r=0}^{r=5} {}^{20}C_{r}(0.05)^{r}(0.95)^{20-r} \text{. Ans.}$$

(iii) The most probable number of returns $= 20-1 = 19$. Ans.

Q.No.10.: The probability that an entering student will graduate is 0.4. Determine the probability that out of 5 students

(a) none

(b) one and

(c) at least one will graduate.

Sol.: The probability that an entering student will graduate = 0.4

Therefore, the probability that an entering student will not graduate $= 1-0.4 = 0.6$

(a) The probability that out of 5 students none will graduate

$$= {}^{5}C_{0}(0.4)^{0}(0.6)^{5} = 0.07776 = 0.08 \text{. Ans.}$$

(b) The probability that out of 5 students one will be graduate

$= {}^5C_1(0.4)^1(0.6)^4 = 0.2592 = 0.26$. Ans.

(c) The probability that out of 5 students at least one will be graduate

= 1 – (probability that out of 5 students none will graduate)

$= 1 - 0.07776 = 0.92224 = 0.92$. Ans.

Q.No.11.: Out of 800 families with 5 children each, how many would you expect to have

(a) 3 boys,

(b) 5 girls,

(c) either 2 or 3 boys ? Assume equal probabilities for boys and girls.

Sol.: Since the probabilities for boys and girls are equal. Therefore

p = probability of having a boy = $\frac{1}{2} = 0.5$, and

q = probability of having a girl = $\frac{1}{2} = 0.5$.

Here n = 5 and N = 800.

(a) The expected number of families having 3 boys = $800\ {}^5C_3(0.5)^3(0.5)^2 = 250$. Ans.

(b) The expected number of families having 5 girls = $800\ {}^5C_5(0.5)^0(0.5)^5 = 25$. Ans.

(c) The expected number of families having either 2 or 3 boys

$= 800\left[{}^5C_2(0.5)^2(0.5)^3 + {}^5C_3(0.5)^3(0.5)^2\right] = 800[0.3125 + 0.3125] = 500$. Ans.

Q.No.12.: If 10 per cent of the rivets produced by a machine are defective, find the probability that out of 5 rivets chosen at random

(i) none will be defective,

(ii) one will be defective, and

(iii) at least two will be defective.

Sol.: No of defective rivets produced by a machine out of 100 = 10 %.

$\therefore$ Probability of defective rivets out of 100 rivets $p = \frac{10}{100} = \frac{1}{10}$.

Then probability of non-defective rivets $q = 1 - \frac{1}{10} = \frac{9}{10}$.

Given n = 5.

(i). Probability of none defective rivets $= {}^5C_0\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^5 = \left(\frac{9}{10}\right)^5 = 0.59049$. Ans.

(ii). Probability of one defective rivets $= {}^5C_1\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^{5-1} = {}^5C_1\left(\frac{1}{10}\right)\left(\frac{9}{10}\right)^4$

$$= \frac{5\times 9^4}{10^5} = 0.32805 \text{. Ans.}$$

(iii). Probability of at least two defective rivets

= 1 − [Probability of none defective rivets + Probability of one defective rivets]

$$= 1 - \left[{}^5C_0\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^5 + {}^5C_1\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^4\right]$$

$$= 1 - [0.59049 + 0.32805] = 0.08146 \text{. Ans.}$$

Q.No.13.: In a bombing action there is 50% chance that any bomb will strike the target. Two direct hits are needed to destroy the target completely. How many bombs are required to be dropped to give a 99% chance or better of completely destroying the target.

Sol.: Since there is 50% chance that any bomb will strike the target.

$\therefore$ Probability that a target gets hit by a bomb is $p = \frac{1}{2}$.

And probability that a target not getting hit $q = 1 - \frac{1}{2}$.

We know that only two direct hits of bombs can destroy the target completely.

$\therefore$ Probability of target being destroyed

$\Rightarrow$ 1 − [Probability of zero bomb hitting target + Probability of one bomb hitting target]

$$= 1 - \left[{}^nC_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^n + {}^nC_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{n-1}\right] = 1 - \left[\left(\frac{1}{2}\right)^n + \frac{n}{2^n}\right].$$

Probability of target destroyed $= 1 - \left[\frac{n+1}{2^n}\right]$.

Since 99%probability or more of completely destroying the target.

$$\Rightarrow 1 - \frac{n+1}{2^n} \geq \frac{99}{100} \Rightarrow \frac{1}{100} \geq \frac{n+1}{2^n}.$$

Taking log on both sides and solving, we get $n = 11$. Ans.

Q.No.14.: A product is 0.5% defective and is packed in cartons of 100. What percentage contains not more than 3 defectives ?

Sol.: Since the %age of defective product is 0.5%.

$\therefore$ Probability of the product being defective $p = \frac{0.5}{100} = 0.005$.

Probability of the product not being defective q= 0.995.

Here n = 100.

$\therefore$ Probability of not more than 3 defectives

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

$$= {}^{100}C_0(0.005)^0(0.995)^{100} + {}^{100}C_1(0.005)(0.995)^{99} + {}^{100}C_2(0.005)^2(0.995)^{98} + {}^{100}C_3(0.005)^3(0.995)^{97}$$

$$= (0.995)^{97}\left[{}^{100}C_0(0.995)^3 + {}^{100}C_1(0.005)(0.995)^2 + {}^{100}C_2(0.005)^2(0.995)^1 + {}^{100}C_3(0.005)^3\right]$$

$$= (0.995)^{97}[0.98505 + 0.4950125 + 0.12313125 + 0.0202125]$$

$$= (0.995)^{97}[1.62342625] = 0.998323739.$$

Hence the percentage = 99.83% Ans.

Q.No.15.: If in a lot of 500 solenoids 25 are defective, find the probability of 0, 1, 2, 3 defective solenoids in a random sample of 20 solenoids.

Sol.: Here n = 500.

The mean number of defectives = 25 = np = 500 p.

Therefore the probability of a defective part is $p = \frac{25}{500} = 0.05$, and

the probability of a non-defective part = 1 – 0.05 = 0.95.

Therefore, the probability of defective solenoids in a random sample of 20 solenoids are as follows:

(i) The probability of 0 defective solenoids in a random sample of 20 solenoids

$$= {}^{20}C_0(0.05)^0(0.95)^{20} = 0.358486 = 0.3585. \text{ Ans.}$$

(ii) The probability of 1 defective solenoids in a random sample of 20 solenoids

$$= {}^{20}C_1(0.05)^1(0.95)^{19} = 0.3773. \text{ Ans.}$$

(iii) The probability of 2 defective solenoids in a random sample of 20 solenoids

$= {}^{20}C_2 (0.05)^2 (0.95)^{18} = 0.1887$. Ans.

(iv) The probability of 3 defective solenoids in a random sample of 20 solenoids

$= {}^{20}C_3 (0.05)^3 (0.95)^{17} = 0.05958 = 0.0596$. Ans.

Q.No.16.: Fit a Binomial distribution for the following data and compare the theoretical frequencies with the actual ones

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

Sol.: Here $n = 5$ and $N = \sum f_i = 100$.

$$\therefore \text{ Mean } = \frac{\sum f_i x_i}{\sum f_i} = \frac{0+14+40+102+88+40}{100} = \frac{284}{100} = 2.84.$$

Now mean of a Binomial distribution = np

i.e. $np = 5p = 2.84 \Rightarrow p = \dfrac{2.84}{5} = 0.568$, and $\therefore\ q = 1 - p = 1 - 0.568 = 0.432$

Hence the Binomial distribution to be fitted is

$$\begin{aligned} N(q+p)^n &= 100(0.432+0.568)^5 \\ &= 100.{}^5C_0(0.432)^5(0.568)^0 + 100.{}^5C_1(0.432)^4(0.568)^1 + 100.{}^5C_2(0.432)^3(0.568)^2 \\ &\quad + 100.{}^5C_3(0.432)^2(0.568)^3 + 100.{}^5C_4(0.432)^1(0.568)^4 + 100.{}^5C_5(0.432)^0(0.568)^5 \\ &= 1.5045 + 9.89 + 26.0104 + 34.198 + 22.4826 + 5.91 \end{aligned}$$

$\therefore$ The successive terms in the expansion give the expected or theoretical frequencies which are

	x :	0	1	2	3	4	5
Actual	f_i :	2	14	20	34	22	8
Theoretical	f_i :	2	10	26	34	22	6

Q.No.17.: Fit a Binomial distribution for the following frequency distribution

x :	0	1	2	3	4	5	6
f_i :	13	25	52	58	32	16	4

Sol.: Here $n = 6$ and $N = \sum f_i = 200$.

$$\therefore \text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0+25+104+174+128+80+24}{200} = \frac{535}{200} = 2.675.$$

Now mean of a Binomial distribution = np i.e. $np = 6p = 2.8675 \Rightarrow p = \dfrac{2.675}{6} = 0.4458$, and $\therefore\ q = 1 - p = 1 - 0.4458 = 0.5542$

Hence the Binomial distribution to be fitted is

$$\begin{aligned}
N(q+p)^n &= 200(0.5542+0.4458)^6 \\
&= 200.{}^6C_0(0.5542)^6(0.4458)^0 + 200.{}^6C_1(0.5542)^5(0.4458)^1 \\
&\quad + 200.{}^6C_2(0.5542)^4(0.4458)^2 + 200.{}^6C_3(0.5542)^3(0.4458)^3 \\
&\quad + 200.{}^6C_4(0.5542)^2(0.4458)^4 + 200.{}^6C_5(0.5542)^1(0.4458)^5 \\
&\quad + 200.{}^6C_6(0.5542)^0(0.4458)^5 \\
&= 5.79 + 27.97 + 56.24 + 60.32 + 36.39 + 11.71 + 1.5699
\end{aligned}$$

$\therefore$ The successive terms in the expansion give the expected or theoretical frequencies which are

x :	0	1	2	3	4	5	6
Actual f_i :	13	25	52	58	32	16	4
Theoretical f_i :	06	28	56	60	36	12	2

*** *** *** *** ***

*** *** ***

***

Topic

Poisson distribution

Poisson distribution as a limiting case of Binomial distribution, Recurrence Formula, Mean and Variance

(Discrete Probability Distribution)

Prepared by:

Dr. Sunil

NIT Hamirpur (HP)

Last updated on 17-11-2007

Poisson distribution

This distribution was discovered by French mathematician S. D. Poisson in 1837. It is a distribution related to the probabilities of events which are extremely rare, but which have a large number of independent opportunities for occurrence. The following are some of the physical situations illustrating the Poisson distribution:

- **The number of person born blind per year in a large city.**
- **The number of deaths from a disease such as heart attaches.**
- **The number of deaths by horse kicks in an army corps.**
- **Number of printing mistakes per page of one of the early proofs of a book.**
- **Arrival pattern of 'defective vehicles in a workshop', 'patient in hospital', or 'telephone calls'.**
- **Demand pattern for certain spare parts.**
- **Number of fragments from a shell hitting a target.**
- **Spatial distribution of bomb hits.**
- **Number of cars passing a crossing per minute.**

Poisson distribution as a limiting case of Binomial distribution:

If the parameters n and p of a binomial distribution are known, then we can find the distribution. But in the situations where n is very large and p is very small, then the application of binomial distribution is very laborious. However, if we assume that as $n \to \infty$ and $p \to 0$ such that np always remains finite, say m, then we get the Poisson approximation to the binomial distribution.

The probability of r successes in a binomial-distribution is

$$P(r) = {}^nC_r p^r q^{n-r} = \frac{n(n-1)(n-2)........(n-r+1)}{r!} p^r q^{n-r}$$

$$= \frac{np(np-p)(np-2p).........(np-\overline{r-1}\,p)}{r!}(1-p)^{n-r}$$

Since $np = m$, we get

$$P(r) = \frac{m(m-p)(m-2p).........(m-\overline{r-1}\,p)}{r!}\left(1-\frac{m}{n}\right)^{n-r}.$$

As $n \to \infty,\ p \to 0$, we have

$$P(r) = \frac{m^r}{r!}\lim_{n\to\infty}\frac{\left(1-\frac{m}{n}\right)^n}{\left(1-\frac{m}{n}\right)^r} = \frac{m^r}{r!}\lim_{n\to\infty}\left(1-\frac{m}{n}\right)^n, \qquad \left[\because \lim_{n\to\infty}\left(1-\frac{m}{n}\right)^r = 1^r = 1\right]$$

$$= \frac{m^r}{r!}\lim_{n\to\infty}\left[\left(1-\frac{m}{n}\right)^{-\frac{n}{m}}\right]^{-m} = \frac{m^r}{r!}\lim_{n\to\infty}\left[\left(1-\frac{1}{\frac{n}{m}}\right)^{-\frac{n}{m}}\right]^{-m}.$$

Since $\lim_{x\to\pm\infty}\left(1+\frac{1}{x}\right)^x = e$, the Naperian base. $\therefore \lim_{n\to\infty}\left[\left(1-\frac{1}{\frac{n}{m}}\right)^{-\frac{n}{m}}\right]^{-m} \to e^{-m}$

$$\therefore P(r) = \frac{m^r}{r!}e^{-m} \qquad \text{(i)}$$

so that the probabilities of 0, 1, 2,................,r,...... successes in a Poisson distribution are given by

$$e^{-m},\ me^{-m},\ \frac{m^2e^{-m}}{2!},........,\frac{m^re^{-m}}{r!},..........$$

(i) represents a probability distribution, which is called the Poisson probability distribution.

Note: m is called the parameter of this distribution.

Remark: The sum of the probabilities P(r) for r = 0, 1, 2, 3, 4,.......... is 1.

Proof: Since $P(0)+P(1)+P(2)+P(3)+.........\ = e^{-m}+\frac{me^{-m}}{1!}+\frac{m^2e^{-m}}{2!}+\frac{m^3e^{-m}}{3!}+......$

$$= e^{-m}\left[1+\frac{m}{1!}+\frac{m^2}{2!}+\frac{m^3}{3!}+...\right]$$

$$= e^{-m}.e^{m} = 1.$$

This completes the proof.

Recurrence Formula for the Poisson distribution:

For Poisson distribution, $P(r)=\frac{m^re^{-m}}{r!}$ and $P(r+1)=\frac{m^{r+1}e^{-m}}{(r+1)!}$

$$\therefore \frac{P(r+1)}{P(r)}=\frac{mr!}{(r+1)!}=\frac{m}{r+1} \Rightarrow P(r+1)=\frac{m}{r+1}P(r),\quad r=0,\ 1,\ 2,\ 3,\ ....$$

This is called the recurrence formula for the Poisson distribution.

Mean and Variance of the Poisson distribution:

For the Poisson distribution, $P(r)=\frac{m^re^{-m}}{r!}$

$$\text{Mean } \mu=\sum_{r=0}^{\infty} rP(r)=\sum_{r=0}^{\infty} r.\frac{m^re^{-m}}{r!}=e^{-m}\sum_{r=0}^{\infty}\frac{m^r}{(r-1)!}=e^{-m}\left(m+\frac{m^2}{1!}+\frac{m^3}{2!}+.......\right)$$

$$= me^{-m}\left(1+\frac{m}{1!}+\frac{m^2}{2!}+.........\right)= me^{-m}.e^{m}=m.$$

Thus, the mean of the Poisson distribution is equal to the parameter m.

$$\text{Variance } \sigma^2=\sum_{r=0}^{\infty} r^2P(r)-\mu^2=\sum_{r=0}^{\infty} r^2.\frac{m^re^{-m}}{r!}-m^2=e^{-m}\sum_{r=1}^{\infty}\frac{r^2m^r}{r!}-m^2$$

$$= e^{-m}\left[\frac{1^2.m}{1!}+\frac{2^2.m}{2!}+\frac{3^2.m}{3!}+\frac{4^2.m}{4!}.............\right]-m^2$$

$$= me^{-m}\left[1+\frac{2m^2}{1!}+\frac{3m^2}{2!}+\frac{4m^2}{3!}+..........\right]-m^2$$

$$= me^{-m}\left[1+\frac{(1+1)m}{1!}+\frac{(1+2)m^2}{2!}+\frac{(1+3)m^2}{3!}+........\right]-m^3$$

$$= me^{m}\left[\left(1+\frac{m}{1!}+\frac{m^2}{2!}+\frac{m^3}{3!}+..........\right)+\left(\frac{m}{1!}+\frac{2m^2}{2!}+\frac{3m^3}{3!}+..........\right)\right]-m^2$$

$$= me^{m}\left[e^m+m\left(1+\frac{m}{1!}+\frac{m^2}{2!}+...........\right)\right]-m^2$$

$$= me^{-m}\left[e^m+me^m\right]-m^2 = me^{-m}.e^m(1+m)-m^2 = m(1+m)-m^2 = m\,.$$

Hence, the variance of the Poisson distribution is also m.

Thus, the mean and variance of the Poisson distribution are equal to the parameter m.

Note: The mean and the variance of the Poisson distribution can also be derived from those of the binomial distribution in the limiting case $n\to\infty$ and $p\to 0$ and $np = m$.

Mean of Poisson Distribution is np.

$\therefore$ Mean of Poisson distribution $= \underset{n\to\infty}{\mathrm{Lt}}\ np = \underset{n\to\infty}{\mathrm{Lt}}\ m = m$.

Variance of binomial distribution $npq = np(1-p)$.

$\therefore$ Variance of Poisson distribution $= \underset{n\to\infty}{\mathrm{Lt}}\ np(1-p) = \underset{n\to\infty}{\mathrm{Lt}}\ m\left(1-\frac{m}{n}\right) = m$.

Remarks: $\mu_2 = \mathrm{Lt}(npq) = m\,\mathrm{Lt}(q) = m$.

$\therefore$ Standard deviation $= \sqrt{m}$

Also $\mu_3 = m,\ \mu_4 = m+3m^2$

$\therefore$ Skewness $\left(=\sqrt{\beta_1}\right) = \frac{1}{m}$, Kurtosis $3+\frac{1}{m}$.

Since μ_3 is positive, Poisson distribution is positively skewed and since $\beta_2 > 3$, it is Leptokurtic.

Now let us solve some problems for better illustration of the Poisson distribution:

Q.No.1.: If the probability of a bad reaction from a certain injection is 0.001, determine the chance that out of 2,000 individuals more than two will get a bad reaction.

Sol.: Since the probability of occurrence is very-very small, so by Poisson distribution

Mean m = np =2000(0.001) = 2

Probability that more than 2 will get a bad reaction

$= 1 - [$prob. that no one gets a bad reaction+ prob. that one gets a bad reaction + prob. that two get bad reaction$]$

$$= 1 - \left[e^{-m} + \frac{m^1 e^{-m}}{1!} + \frac{m^2 e^{-m}}{2!} \right] = 1 - \left[\frac{1}{e^2} + \frac{2}{e^2} + \frac{2}{e^2} \right]$$

$$= 1 - \frac{5}{e^2} = 0.32. \text{ Ans.}$$

Q.No.2.: 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 a packet of 10, use Poisson distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10,000 packets.

Sol.: In a Poisson distribution, we know that $m = np = 10 \times 0.002 = 0.02$

$$\therefore e^{-0.02} = 1 - 0.02 + \frac{(0.02)^2}{2!} - .... = 0.9802 \text{ approximately.}$$

Probability of no defective blade is $e^{-m} = e^{-0.02} = 0.9802$

(i) $\therefore$ No. of packets containing no defective blade is $10{,}000 \times 0.9802 = 9802$. Ans.

(ii) Similarly the number of packets containing one defective blade $= 10{,}000 \times me^{-m}$

$= 10{,}000 \times (0.02) \times 0.9802 = 196$. Ans.

(iii) Finally the number of packets containing two defective blades

$$= 10.000 \times \frac{m^2 e^{-m}}{2!} = 10{,}000 \times \frac{(0.02)^2}{2!} \times 0.9802 = 2 \text{ approximately. Ans.}$$

Q.No.3.: If the variance of the Poisson distribution is 2, find the probabilities for $r = 1, 2, 3, 4$ from the recurrence relation of the Poisson distribution.

Sol.: The variance of Poisson distribution= σ^2 =m = 2.

Recurrence relation for the Poisson distribution is $P(r+1) = \frac{m}{r+1}P(r) = \frac{2}{r+1}P(r)$

Now $P(r) = \frac{m^r e^{-m}}{r!} \Rightarrow P(0) = \frac{e^{-2}}{0!}e^{-2} = 0.1353$.

Putting $r = 0, 1, 2, 3$, for evaluating P(1), P(2), P(3) and P(4), we get

$P(1) = 2P(0) = 2 \times 0.1353 = 0.2706$. Ans.

$P(2) = \frac{2}{2}P(1) = 0.2706$. Ans.

$P(3) = \frac{2}{3}P(2) = \frac{2}{3} \times .2706 = 0.1806$. Ans.

$P(4) = \frac{2}{4}P(3) = \frac{2}{3} \times .1806 = 0.0902$. Ans.

Q.No.4.: Six coins are tossed 6400 times. Using the Poisson distribution, determine the approximate probability of getting six times heads x times.

Sol.: Probability of getting one head with one coin $= \frac{1}{2}$.

$\therefore$ Probability of getting six heads with six coins $= \left(\frac{1}{2}\right)^6 = \frac{1}{64}$.

$\therefore$ Average number of six heads with six coins in 6400 throws $= np = 6400 \times \frac{1}{64} = 100$.

$\therefore$ The mean of the Poisson distribution = 100.

Approximate probability of getting six heads x times when the distribution is Poisson

$= \frac{m^x e^{-m}}{x!} = \frac{(100)^x . e^{-100}}{(100)!}$. Ans.

Q.No.5.: If a random variable has a Poisson distribution such that P(1) = P(2), find (i) mean of the distribution, (ii) P(4).

Sol.: Given P(1) = P(2).

(i) Since $P(r) = \frac{m^r}{r!}e^{-m}$. Therefore $P(1) = P(2)$.

$\Rightarrow \frac{m^1}{1!}e^{-m} = \frac{m^2}{2!}e^{-m} \Rightarrow m = 2 = \text{mean}$.

Hence, mean of the distribution is 2.

(ii) Since $P(r) = \frac{m^r}{r!}e^{-m}$. Therefore $P(4) = \frac{m^r}{r!}e^{-m} = \frac{2^4}{4!}e^{-2} = \frac{16}{24}e^{-2} = \frac{2}{3}e^{-2}$

$= (0.6667).(0.1353) = 0.0902$. Ans.

Q.No.6.: X is a Poisson variable and it is found that the probability that X = 2 is two-thirds of the probability that X = 1. Find the probability that X = 0 and the probability that X = 3. What is the probability that X exceeds 3 ?

Sol.: Given $P(2) = \frac{2}{3}P(1) \Rightarrow \left(\frac{m^2}{2}e^{-m}\right) = \frac{2}{3}\left(\frac{m^1}{1}e^{-m}\right) \Rightarrow m = \frac{4}{3}$.

Since $P(r) = \frac{m^r}{r!}e^{-m}$, therefore $P(0) = \frac{m^r}{r!}e^{-m} = \frac{m^0}{0!}e^{--m} = e^{-\frac{4}{3}} = 0.2636$. Ans.

And $P(3) = \frac{\left(\frac{4}{3}\right)^3}{3!}e^{-m} = \frac{64}{162}e^{-\frac{4}{3}} = \frac{32}{81}e^{-\frac{4}{3}} = 0.1041$. Ans.

Also $P(1) = \frac{m^1}{1!}e^{-m} = \frac{4}{3}e^{-\frac{4}{3}} = 0.35146$ and $P(2) = \frac{m^2}{2!}e^{-m} = \frac{8}{9}e^{-\frac{4}{3}} = 0.234308$

$P(X > 3) = 1 - [P(0) + P(1) + P(2) + P(3)] = [1 - (0.2636 + 0.35146 + 0.234308 + 0.1041)]$

$= 1 - 0.849363 = 0.150632$. Ans.

Q.No.7.: A certain screw making machine produces on average of 2 defective screws out of 100, and packs them in boxes of 500. Find the probability that a box contains 15 defective screws.

Sol.: The probability of 2 defective screws out of 100 = $\frac{2}{100} = 0.02$.

Now probability of occurrence is very small, so by Poisson distribution

mean $(m) = np = 500 \times 0.02 = 10$.

$\therefore$ Probability that a box contains15 defective screws

$$=\frac{m^r}{r\,!}e^{-m}=\frac{10^{15}}{15\,!}e^{-10}=0.0347\text{ . Ans.}$$

Q.No.8.: A manufacturer knows that the condensers he makes contain on the average 1% defectives. He packs them in boxes of 100. What is the probability that a box picked at random will contain 3 or more faulty condensers ?

Sol.: The probability of defective condensers = 1% = 0.01

Now probability of occurrence is very small, so by Poisson distribution

Mean (m) = np = $100\times 0.01=1$

Since $P(r)=\frac{m^r}{r\,!}e^{-m}$.

$$\therefore P(0)=\frac{1^0}{0\,!}e^{-1}=\frac{1}{e},\quad P(1)=\frac{1^1}{1!}e^{-1}=\frac{1}{e},\quad P(2)=\frac{1^2}{2\,!}e^{-1}=\frac{1}{2e}.$$

Therefore, probability that a box picked at random will contain 3 or more faulty condensers is

$$P(r\ge 3)=1-[P(0)+P(1)+P(2)]=1-\left[\frac{1}{e}+\frac{1}{e}+\frac{1}{2e}\right]$$

$$=1-\left[\frac{2+2+1}{2e}\right]=1-\frac{5}{2e}=\frac{2e-5}{2e}=0.08030.\text{ Ans.}$$

Q.No.9.: A car-hire firm has two cars, which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson distribution with mean 1.5. Calculate the proportion of days

(i) on which there is no demand,

(ii) on which demand is refused. $\left(e^{-1.5}=0.2231\right)$.

Sol.: Since the number of demands for a car is distributed as a Poisson distribution with mean 1.5. So mean (m) = 1.5 and $e^{-m}=e^{-1.5}=0.2231$. Since $P(r)=\frac{m^r}{r\,!}e^{-m}$.

(i) The proportion of days on which neither car is used

= Probability of there being no demand for the car

$$=P(0)=\frac{m^0}{0\,!}e^{-m}=e^{-1.5}=0.2231\text{ . Ans.}$$

(ii) The proportion of days on which some demand is refused

= Probability for the number of demands to be more than two

$=1-P(X \le 2)$

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

Since $P(1) = \dfrac{m^1}{1!}e^{-m} = (1.5)\times e^{-1.5} = 0.33465$.

$$P(2) = \frac{m^2}{2!}e^{-m} = \frac{(1.5)^2 \times e^{-1.5}}{2} = 0.2509875.$$

Therefore, the proportion of days on which demand is refused is

$$P = 1-[P(0)+P(1)+P(2)]$$
$$= 1-[0.2231+0.33465+0.2509875] = 1-0.8087375 = 0.1912625. \text{ Ans.}$$

Q.No.10.: The incidence of occupational disease in an industry is such that the workmen have a 10% chance of suffering from it. What is probability that in a group of 7, five or more will suffer from it ?

Sol.: The probability of suffering from occupational disease in an industry =10% = 0.1

Now probability of occurrence is very small, so by Poisson distribution

mean $(m) = np = 7\times 0.1 = 0.7$. [Here given n = 7]

Since $P(r) = \dfrac{m^r}{r!}e^{-m}$, therefore,

probability that in a group of 7, five or more will suffer from it

$$= 1-[P(0)+P(1)+P(2)+P(3)+P(4)]$$
$$= 1-\left[\frac{m^0}{0!}e^{-m} + \frac{m^1}{1!}e^{-m} + \frac{m^2}{2!}e^{-m} + \frac{m^3}{3!}e^{-m} + \frac{m^4}{4!}e^{-m}\right]$$
$$= 1-e^{-m}\left[1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\frac{m^4}{4!}\right] = 1-e^{-0.7}\left[1+0.7+\frac{(0.7)^2}{2!}+\frac{(0.7)^3}{3!}+\frac{(0.7)^4}{4!}\right]$$
$$= 1-\frac{(1+0.7+0.245+0.05717+0.01)}{2.0137} = 1-0.9992 = 0.0008. \text{ Ans.}$$

Q.No.11.: The frequency of accidents per shift in a factory is as shown in the following table:

Accidents per shift	0	1	2	3	4
Frequency	180	92	24	3	1

Calculate the mean number of accidents per shift and the corresponding Poisson distribution and compare with actual observations.

Sol.: The actual frequency of accidents per shift in a factory is given in the following table

Accidents per shift (x_i)	0	1	2	3	4
Actual Frequency (f_i)	180	92	24	3	1

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0+92+48+9+4}{300} = \frac{153}{300} = 0.51.$$

$\therefore$ Mean of the Poisson distribution i.e. m = 0.51.

Hence the theoretical frequency of r successes is $P(r) = \frac{Nm^r}{r!}e^{-m} = \frac{300\times(0.51)^r.e^{-0.51}}{r!}$,

where r = 0, 1, 2, 3, 4.

Then the corresponding Poisson distributions are

$$f(0) = \frac{300\times(0.51)^0.e^{-0.51}}{0!} = 300\times e^{-0.51} = 180.14867.$$

$$f(1) = \frac{300\times(0.51)^1.e^{-0.51}}{1!} = 300\times(0.51)\times e^{-0.51} = 91.875824.$$

$$f(2) = \frac{300\times(0.51)^2.e^{-0.51}}{2!} = \frac{300\times(0.2601)\times e^{-0.51}}{2} = 23.428335.$$

$$f(3) = \frac{300\times(0.51)^3.e^{-0.51}}{3!} = \frac{300\times(0.132651)\times e^{-0.51}}{6} = 3.982817.$$

$$f(4) = \frac{300\times(0.51)^4.e^{-0.51}}{4!} = \frac{300\times(0.067652)\times e^{-0.51}}{24} = 0.5078091.$$

$\therefore$ The theoretical frequencies are

Accidents per shift (x_i)	0	1	2	3	4
Actual Frequency (f_i)	180	92	24	3	1
Theoretical frequency (f_i)	180.1	91.9	23.4	4	0.5

Q.No.12.: Fit a Poisson distribution to the following :

x :	0	1	2	3	4
f :	46	38	22	9	1

Sol.: Mean $= \dfrac{\sum f_i x_i}{\sum f_i} = \dfrac{0+38+44+27+4}{116} = \dfrac{113}{116} = 0.974$

$\therefore$ Mean of the Poisson distribution i.e. m = 0.974.

Hence the theoretical frequency of r successes is

$$f(r) = \frac{Nm^r}{r!}e^{-m} = \frac{116 \times (0.974)^r . e^{-0.974}}{r!}, \text{ where } r = 0, 1, 2, 3, 4.$$

$$f(0) = \frac{116 \times (0.974)^0 . e^{-0.974}}{0!} = 116 \times e^{-0.974} = 43.798089.$$

$$f(1) = \frac{116 \times (0.974)^1 . e^{-0.974}}{1!} = 116 \times (0.974) \times e^{-0.974} = 42.659339.$$

$$f(2) = \frac{116 \times (0.974)^2 . e^{-0.974}}{2!} = \frac{116 \times (0.948676) \times e^{-0.974}}{2} = 20.775098.$$

$$f(3) = \frac{116 \times (0.974)^3 . e^{-0.974}}{3!} = \frac{116 \times (0.9240104) \times e^{-0.974}}{6} = 6.7449818.$$

$$f(4) = \frac{116 \times (0.974)^4 . e^{-0.974}}{4!} = \frac{116 \times (0.8999861) \times e^{-0.974}}{24} = 1.6424.$$

$\therefore$ The theoretical frequencies are

x :	0	1	2	3	4
f :	44	43	21	7	2

Q.No.13.: Fit a Poisson distribution to the following data given the number of yeast cells per square for 400 squares :

No. of cells per sq.	: 0	1	2	3	4	5	6	7	8	9	10
No. of squares	: 103	143	98	42	8	4	2	0	0	0	0

Sol.: The number of yeast cells per square for 400 squares is given in the following table

No. of cells per sq.	: 0	1	2	3	4	5	6	7	8	9	10
No. of squares	: 103	143	98	42	8	4	2	0	0	0	0

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0+143+196+126+32+20+12+0+0+0+0}{400} = \frac{529}{400} = 1.3225.$$

$\therefore$ Mean of the Poisson distribution i.e. m = 1.3225.

Hence the theoretical frequency of r successes is

$$f(r)=\frac{Nm^r}{r!}e^{-m}=\frac{400\times(1.3225)^r.e^{-1.3225}}{r!}\text{, where } r = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.$$

$$f(0)=\frac{400\times(1.3225)^0.e^{-1.3225}}{0!}=400\times e^{-1.3225}=106.64\,.$$

$$f(1)=\frac{400\times(1.3225)^1.e^{-1.3225}}{1!}=400\times(1.3225)\times e^{-1.3225}=141.03\,.$$

$$f(2)=\frac{400\times(1.3225)^2.e^{-1.3225}}{2!}=\frac{400\times(1.749)\times e^{-1.3225}}{2}=93.257561\,.$$

$$f(3)=\frac{400\times(1.3225)^3.e^{-1.3225}}{3!}=\frac{400\times(2.313)\times e^{-1.3225}}{6}=41.111\,.$$

$$f(4)=\frac{400\times(1.3225)^4.e^{-1.3225}}{4!}=\frac{400\times(3.059)\times e^{-1.3225}}{24}=13.59\,.$$

$$f(5)=\frac{400\times(1.3225)^5.e^{-1.3225}}{5!}=\frac{400\times(4.0455)\times e^{-1.3225}}{120}=3.595\,.$$

$$f(6)=\frac{400\times(1.3225)^6.e^{-1.3225}}{6!}=\frac{400\times(5.35)\times e^{-1.3225}}{720}=0.7924\,.$$

$$f(7)=\frac{400\times(1.3225)^7.e^{-1.3225}}{7!}=\frac{400\times(7.0754)\times e^{-1.3225}}{5070}=0.1497\,.$$

$$f(8)=\frac{400\times(1.3225)^8.e^{-1.3225}}{8!}=\frac{400\times(9.3572)\times e^{-1.3225}}{40560}=0.025\,.$$

$$f(9)=\frac{400\times(1.3225)^9.e^{-1.3225}}{9!}=\frac{400\times(12.375)\times e^{-1.3225}}{365040}=0.00367\,.$$

$$f(10)=\frac{400\times(1.3225)^{10}.e^{-1.3225}}{10!}=\frac{400\times(16.366)\times e^{-1.3225}}{3650400}=0.000485\,.$$

$\therefore$ The theoretical frequencies are

No. of cells per sq.	: 0	1	2	3	4	5	6	7	8	9	10
No. of squares	: 107	141	93	41	14	4	1	0	0	0	0

Q.No.14.: Fit a Poisson distribution to the set of observations :

x :	0	1	2	3	4
f :	122	60	15	2	1

Sol.: Mean $= \dfrac{\sum f_i x_i}{\sum f_i} = \dfrac{0+60+36+6+4}{200} = \dfrac{106}{200} = 0.53$

$\therefore$ Mean of the Poisson distribution i.e. m = 0.53

Hence the theoretical frequency of r successes is

$$f(r) = \frac{Nm^r}{r!}e^{-m} = \frac{200\times(0.53)^r.e^{-0.53}}{r!}, \text{ where } r = 0, 1, 2, 3, 4.$$

$$f(0) = \frac{200\times(0.53)^0.e^{-0.53}}{0!} = 200\times e^{-0.53} = 117.7\,.$$

$$f(1) = \frac{200\times(0.53)^1.e^{-0.53}}{1!} = 62.38\,.$$

$$f(2) = \frac{200\times(0.53)^2.e^{-0.53}}{2!} = 16.53\,.$$

$$f(3) = \frac{200\times(0.53)^3.e^{-0.53}}{3!} = 2.92\,.$$

$$f(4) = \frac{200\times(0.53)^4.e^{-0.53}}{4!} = 0.39\,.$$

$\therefore$ The theoretical frequencies are

x :	0	1	2	3	4
f :	118	62	17	3	0

*** *** *** *** ***

*** *** ***

***

Topic

Normal distribution

Standard form of the Normal Distribution, Properties of normal distribution , Mean deviation from the mean μ, Moments about the mean, Probability of x lying between x_1 and x_2, Probable error, Normal frequency distribution

Prepared by:

Dr. Sunil

NIT Hamirpur (HP)

Last updated on 30-10-2007

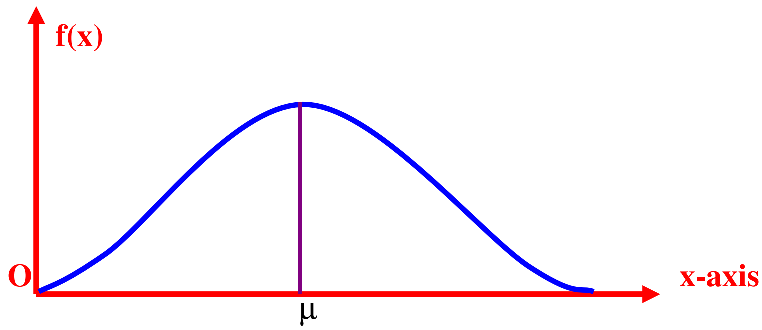
Normal distribution:

The normal distribution is a continuous distribution. It can be derived from the binomial distribution in the limiting case when n, the number of trials is very large and p, the probability of a success, is close to $\frac{1}{2}$. The general equation of the normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

where the variable x can assume all values from $-\infty$ to $+\infty$. μ and σ, called the parameters of the distribution, are respectively the mean and standard deviation of the distribution and $-\infty < \mu < \infty$, $\sigma > 0$, x is called the normal variate and f(x) is called probability density function of the normal distribution.

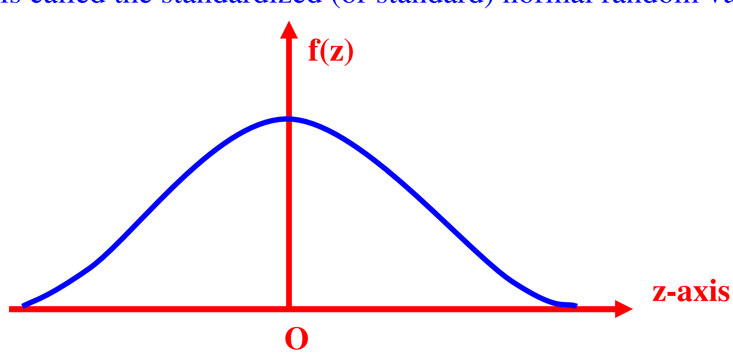
If a variable x has the normal distribution with **Mean** μ and **standard deviation** σ, we briefly write $x : N(\mu, \sigma^2)$.



The graph of the normal distribution is called the normal curve. It is bell-shaped and symmetrical about the mean μ. The two tails of the curve extend to $+\infty$ to $-\infty$ towards the positive and negative directions of the x-axis respectively and gradually approach the x-axis without ever meeting it. The curve is unimodal and the mode of the normal distribution coincides with its mean μ. The line $x = \mu$ divides the area under the normal curve above x-axis into two equal parts. Thus, the median of the distribution also coincides with its mean and mode. The area under the normal curve between any two given ordinates $x = x_1$ and $x = x_2$ represents the probability of values falling into the given interval. The total area under the normal curve above the x-axis is 1.

Standard form of the Normal Distribution:

If x is a normal variable with mean μ and standard deviation σ, then the random variable $z = \dfrac{x - \mu}{\sigma}$ has the normal distribution with mean 0 and standard deviation 1. The random variable z is called the standardized (or standard) normal random variable.



The probability density function for the normal distribution in standard form is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad (-\infty < z < \infty)$$

It is free from any parameter. This helps us to compute area under the normal probability curve by making use of standard tables.

Remarks (i). If f(z) is the probability density function for the normal distribution, then

$$P(z_1 \le Z \le z_2) = \int_{z_1}^{z_2} f(z)dz = F(z_2) - F(z_1) \quad \text{where } F(z) = \int_{-\infty}^{z} f(z)dz = P(Z \le z)$$

The function F(z) defined above is called the distribution function for the normal distribution.

2. The probabilities $P(z_1 \le Z \le z_2)$, $P(z_1 < Z \le z_2)$, $P(z_1 \le Z < z_2)$, and $P(z_1 < Z < z_2)$ are all regarded to be the same.

3. $F(-z_1) = 1 - F(z_1)$.

Applications of the normal distribution:

This distribution is applied to problems concerning:

(i) Calculation of errors made by chance in experiment measurements.

(ii) Computation of hit probability of a shot.

(iii) Statistical inference in almost every branch of science.

Properties of normal distribution:

(1). Basic Properties:

The probability density function of the normal distribution is given

$$\text{by} \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

then

(i) $f(x) \ge 0$.

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$, i.e. the total area under the normal curve above the x-axis is 1.

(iii) The normal curve is bell-shaped and is symmetrical about its mean.

(iv) It is unimodal distribution with ordinates decreasing rapidly on both sides of the mean. The maximum ordinate is $\frac{1}{\sigma\sqrt{(2\pi)}}$, found by putting $x = \mu$

(v) As it is symmetrical, its mean, median and mode are the same. i.e. the mean, mode and medium of this distribution coincide.

(vi) Its points of inflexion (found by putting $\frac{d^2y}{dx^2} = 0$ and verifying that at these points $\frac{d^3y}{dx^3} \neq 0$) are given by $x = \mu \pm \sigma$, i. e. these points are equidistant from the mean on either side.

(2). Mean deviation from the mean μ:

$$= \int_{-\infty}^{\infty} |x-\mu| . \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} dx , \qquad [\text{Put } z = \frac{(x-\mu)}{\sigma}]$$

$$= \frac{\sigma}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} |z| e^{-\frac{z^2}{2}} dz = \frac{\sigma}{\sqrt{(2\pi)}} \left[\int_{-\infty}^{0} ze^{-\frac{z^2}{2}} dz + \int_{0}^{\infty} ze^{-\frac{z^2}{2}} dz \right] = \frac{2\sigma}{\sqrt{(2\pi)}} \int_{0}^{\infty} ze^{-\frac{z^2}{2}} dz$$

$$= \frac{\sigma}{\sqrt{(2\pi)}} \left| -e^{-\frac{z^2}{2}} \right|_{0}^{\infty} = -\sqrt{\left(\frac{2}{\pi}\right)} \sigma(0-1) = 0.7979\sigma = \left(\frac{4}{5}\right)\sigma .$$

(3). Moments about the mean:

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{\sigma^{2n+1}}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} z^{2n+1} e^{-z^2/2} dz , \qquad [\text{where } z = \frac{(x-\mu)}{\sigma}]$$

$= 0$, since the integral is an odd function.

Thus all odd order moments about the mean vanish.

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} z^{2n-1} e^{-z^2/2}.zdz \qquad \text{[Integrate by parts]}$$

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} \left[\left| -z^{2n-1} e^{-z^2/2} \right|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} (2n-1).z^{2n-2} e^{-z^2/2} dz \right]$$

$$= \frac{\sigma^{2n}}{\sqrt{(2\pi)}} (0-0) + (2n-1)\sigma^2 \mu_{2n-2}.$$

Repeated application of this reduction formula, gives

$\mu_{2n} = (2n-1)(2n-3)..........3.1\sigma^{2n}$.

In particular, $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$.

Hence $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$ and $\beta_2 = \frac{\mu_4}{\mu_2^2} = 3.$

i. e. the coefficient of skewness is zero (i. e. the curve is symmetrical) and the Kurtosis is 3. This is the basis for the choice of the value 3 in the definitions of platykurtic and leptokurtic.

(4). Probability of x lying between x_1 and x_2:

The probability of x lying between x_1 and x_2 is given by the area under the normal curve from x_1 to x_2, i. e. $(x_1 \le x \le x_2)$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{x_1}^{x_2} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \int_{z_1}^{z_2} e^{-z^2/2} dz, \text{ where } z = \frac{(x-\mu)}{\sigma},\ dz = \frac{dx}{\sigma} \text{ and } z_1 = \frac{(x_1-\mu)}{\sigma},\ z_2 = \frac{(x_2-\mu)}{\sigma}.$$

$$= \frac{1}{\sigma\sqrt{(2\pi)}} \left[\int_0^{z_2} e^{-z^2/2} dz - \int_0^{z_1} e^{-z^2/2} dz \right] = P_2(z) - P_1(z).$$

The values of each of the above integrals can be found from the table of area under the normal curve from 0 to z, which gives the values of $P(z) = \frac{1}{\sqrt{(2\pi)}} \int_0^z e^{-z^2/2} dz$, for various values of z. This integral is called the probability integral or the error function due to its use in the theory of sampling and the theory of errors.

Using the table, we see that the area under the curve from z = 0 to z = 1, i. e. from $x = \mu$ to $\mu + \sigma = 0.3413$.

(i). The area under the normal curve between the ordinates $x = \mu - \sigma$ and $x = \mu + \sigma$ is 0.6826, $\sim 68\%$ nearly. Thus approximately $\frac{2}{3}$ of the values lie within these limits.

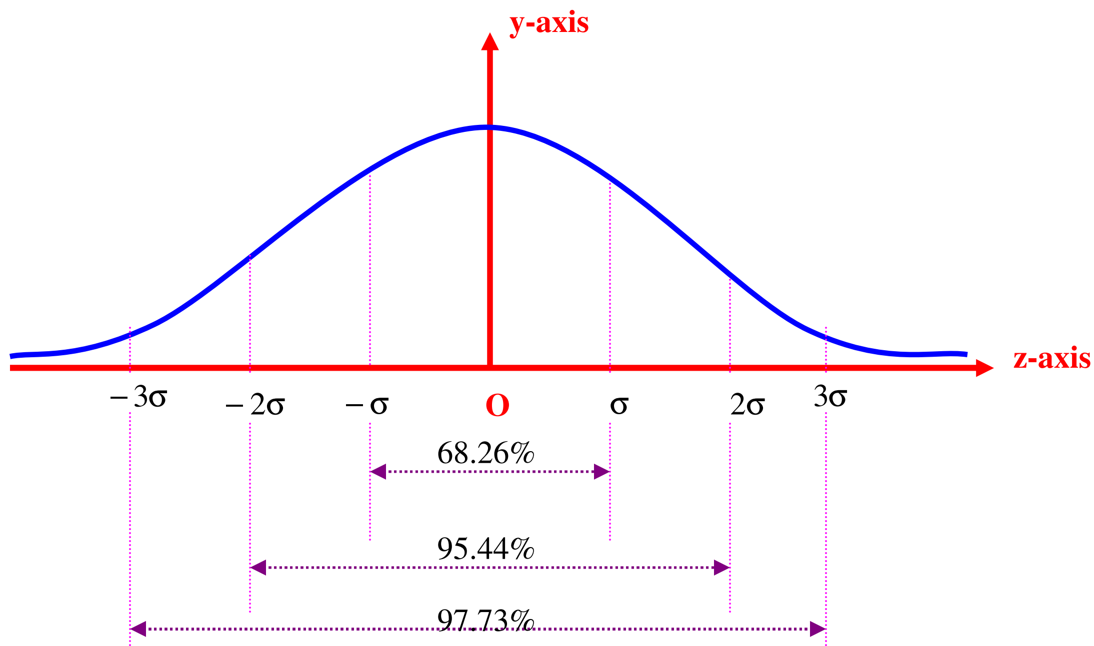
(ii). The area under the normal curve between the ordinates $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$ is 0.9544 -95%, which implies that about $4\frac{1}{2}\%$ of the values lie outside these limits.

(iii). 99.73% of the values lie between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$, i. e. only a quarter % of the whole lies outside these limits.

(iv). 95% of the values lie between $x = \mu - 1.96\sigma$ and $x = \mu + 1.96\sigma$ i. e. only 5% of the values lie outside these limits.

(v). 99% of the values lie between $x = \mu - 2.58\sigma$ and $x = \mu + 2.58\sigma$ i. e. only 1% of the values lie outside the limits.

(vi). 99.9% of the values lie between $x = \mu - 3.29\sigma$ and $x = \mu + 3.29\sigma$.



In other words, a value that deviates more than σ from μ occurs about once in 3 trials. A value that deviates more than 2σ or 3σ occurs about once in 20 or 400 trials. Almost all values lie within 3σ of the mean.

The shape of standardized normal curve is

$$y = \frac{1}{\sigma\sqrt{(2\pi)}}\int_0^z e^{-z^2/2}dz, \text{ where } z = \frac{(x-\mu)}{\sigma},$$

and the respective areas are shown in fig. 'z' is called a normal variate.

Probable error:

Any lot of articles manufactured to certain specifications is subject to small errors. In fact, measurement of any physical quantity shows slight error. In general, these errors of manufacture or experiment are of random nature and therefore, follow a normal distribution. While quoting a specification of an experimental result, we usually mention the probable error (λ). It is such that the probability of an error falling within the limits $\mu-\lambda$ and $\mu+\lambda$ is exactly equal to the chance of an error falling outside these limits,

i. e. the chance of an error lying with in $\mu-\lambda$ and $\mu+\lambda$ is $\frac{1}{2}$.

$$\therefore \frac{1}{\sigma\sqrt{(2\pi)}}\int_{\mu-\lambda}^{\mu+\lambda} e^{-(x-\mu)^2/2\sigma^2}dx = \frac{1}{2} \Rightarrow \frac{1}{\sqrt{(2\pi)}}\int_0^{\lambda/\mu} e^{-z^2/2}dz = \frac{1}{4}. \qquad \left[z = \frac{x-\mu}{\sigma}\right]$$

The table of area under the normal curve from 0 to z gives $\frac{\lambda}{\sigma} = 0.6745$

Hence the probable error $\lambda = 0.6745\sigma \sim \frac{2}{3}\sigma$.

Remarks: Quartile deviation $= \frac{1}{2}(Q_3 - Q_1) \sim \frac{2}{3}\sigma$. Mean deviation $= \frac{4}{5}\sigma$

$\therefore$ Q. D. : M. D. : S. D. $= \frac{2}{3}:\frac{4}{5}:1$.

Normal frequency distribution:

We can fit a normal curve to any distribution. If N be the total frequency, μ the mean and σ the standard deviation of the given distribution, then the curve is

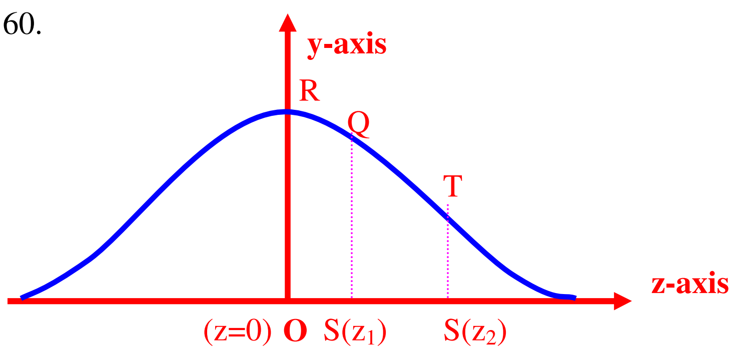
$y = \frac{N}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2}$ will fit the given distribution as best as the data will permit. The frequency of the variate between x_1 and x_2 as given by the fitted curve, will be the area under $z = \frac{x - np}{\sqrt{npq}}$ from x_1 to x_2.

Now let us solve some problems for better illustration of the Normal distribution:

Q..No.1.: A certain number of articles manufactured in one batch were classified into three categories according to a particular characteristic, being less than 50, between 50 and 60, and greater than 60. If this characteristic is known to be normally distributed, determine the mean and standard deviation for this batch if 60%, 35% and 5% were found in this categories.

Sol.: Let μ be the mean (at $z = 0$) and σ the standard deviation of the normal curve.

Now 60% of the articles have the characteristic below 50, 35% between 50 and 60 and only 5% greater than 60.



Let the area to the left of the ordinate PQ be 60% and that between the ordinates PQ and ST be 35% so that the areas to the left of PQ $(z = z_1)$ and ST $(z = z_2)$ are 0.6 and 0.95 respectively, i. e. the area OPQR $= 0.6 - 0.5 = 0.1$ and the area OSTR $= 0.45$.

$\therefore$ Area corresponding to $z_1\left(= \frac{50 - \mu}{\sigma}\right) = 0.1$,

and that corresponding to $z_2\left(= \frac{60 - \mu}{\sigma}\right) = 0.45$.

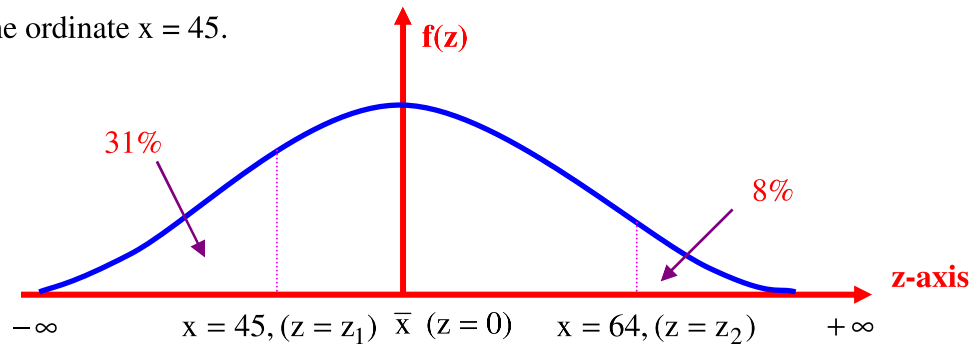
With the help of table of area under the normal curve from 0 to z, we have

$\frac{(50-\mu)}{\sigma} = 0.2533$ and $\frac{(60-\mu)}{\sigma} = 1.645$

$\Rightarrow \sigma = 7.543$. Ans. and $\mu = 48.092$. Ans.

Q.No.2.: In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and S. D. of the distribution.

Sol.: Let $\overline{x}$ be the mean and σ the S. D. 31% of the items are under 45 means area to the left of the ordinate x = 45.



When x = 45, let $z = z_1$ so that $z_1 = \frac{45-\overline{x}}{\sigma}$ (i)

$$\therefore \int_{-\infty}^{z_1} \phi(z)dz = 0.31 \Rightarrow \int_{-\infty}^{0} \phi(z)dz - \int_{z_1}^{0} \phi(z)dz = 0.31$$

$$\text{Hence } \int_{z_1}^{0} \phi(z)dz = \int_{-\infty}^{0} \phi(z)dz - 0.31 = 0.5 - 0.31 = 0.19$$

From table, $z_1 = -0.5$ (ii)

When x = 64, let $z = z_2$ so that $z_2 = \frac{64-\overline{x}}{\sigma}$ (iii)

$$\therefore \int_{z_2}^{\infty} \phi(z)dz = 0.08 \Rightarrow \int_{0}^{\infty} \phi(z)dz - \int_{0}^{z_2} \phi(z)dz = 0.08$$

$$\Rightarrow \int_{0}^{z_2} \phi(z)dz = \int_{0}^{\infty} \phi(z)dz - 0.08 = 0.5 - 0.08 = 0.42.$$

From table , $z_2 = 1.4$. (iv)

From (i) and (ii), $45 - \overline{x} = -0.5\sigma$. Ans.

From (iii) and (iv), $64 - \overline{x} = +1.4\sigma$. Ans.

Solving the equations, we get $\overline{x} = 50$ and $\sigma = 10$.

Q.No3.: In a test on 2000 electric bulbs, it was found that the life of a particular make, was normally distributed with an average life of 2040 hours and S. D. of 60 hours. Estimate the number of bulbs likely to burn for

(a) more than 2150 hours,

(b) less than 1960 hours and

(c) more than 1920 hours but less than 2160 hours.

Sol.: Here $\mu = 2040$ hours and $\sigma = 60$ hours

(a). For x = 2150, $z = \dfrac{x-\mu}{\sigma} = 1.833$.

$\therefore$ Area against z = 1.83 in the table = 0.4664.

We, however, require the area to the right of the ordinate at z = 1.83. This area $= 0.5 - 0.4664 = 0.0336$.

Thus the number of bulbs expected to burn for more than 2150 hours

$= 0.0336 \times 2000 = 67$ approximately.

(b). For x = 1960, $z = \dfrac{x-\mu}{\sigma} = -1.33$.

The area require in this case is to the left of $z = -1.33$

i. e. $= 0.5 - 0.4082$ (table value for z = 1.33) = 0.0918.

$\therefore$ The number of bulbs expected to burn for less than 1950 hours

$= 0.0918 \times 2000 = 184$ approximately.

(c). When x = 1920, $z = \dfrac{1920-2040}{60} = -2$

When x = 2160, $z = \dfrac{2160-2040}{60} = 2$.

The number of bulbs expected to burn for more than 1920 hours but less than 2160 hours will be represented by the area between $z = -2$ and z = 2. This is twice the area from the table for z = 2, i. e. $2 \times 0.4772 = 0.9544$.

Thus the required number of bulbs $= 0.9544 \times 2000 = 1909$ approximately.

Q.No.4.: If the probability of committing an error of magnitude x is given by

$y = \dfrac{h}{\sqrt{\pi}} e^{-h^2x^2}$; compute the probable error from the following data:

$m_1 = 1.305$; $m_2 = 1.301$; $m_3 = 1.295$; $m_4 = 1.286$; $m_5 = 1.318$;

$m_6 = 1.321$; $m_7 = 1.283$; $m_8 = 1.289$; $m_9 = 1.300$; $m_{10} = 1.286$.

Sol.: From the given data which is normally distributed, we have

mean $= \frac{1}{10}\sum m_i = \frac{12.984}{10} = 1.2984$ and

$$\sigma^2 = \frac{1}{10}\sum (m_i - \text{mean})^2$$

$$= \frac{1}{10}\left[\begin{array}{c}(0.007)^2 + (0.003)^2 + (0.003)^2 + (0.012)^2 + (0.02)^2 + (0.023)^2 \\ + (0.015)^2 + (0.09)^2 + (0.002)^2 + (0.012)^2\end{array}\right]$$

$= 0001595 \Rightarrow \sigma = 0.0126$.

$\therefore$ Probable error $= \frac{2}{3}\sigma = 0.0084$ approximately.

Q.No.5.: Fit a normal curve to the following distribution.

x :	2	4	6	8	10
f :	1	4	6	4	1

Sol.: Mean $\frac{\sum fx}{\sum f} = \frac{2+16+36+32+10}{16} = 6$.

S. D. $= \sqrt{\left[\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2\right]} = \sqrt{(40-36)} = 2$

Taking $\mu = 6$, $\sigma = 2$ and $N = 16$, the equation of the normal curve is

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \Rightarrow \frac{1}{2\sqrt{2\pi}} e^{-(x-6)^2/8} \qquad \text{(i)}$$

Area under (i) in (x_1, x_2) or (z_1, z_2)

$$= \frac{1}{\sqrt{2\pi}}\int_0^{z_2} e^{-z^2/2}dz - \frac{1}{\sqrt{2\pi}}\int_0^{z_1} e^{-z^2/2}dz \quad \text{whence } z = \frac{x-6}{2}$$

To evaluate these integrals, we refer to table of the area under the normal curve from 0 to z.

Calculations:

Mid x	(x_1, x_2)	(z_1, z_2)	Area under (i) in (x_1, x_2)	Expected frequency
2	(1, 3)	$(-2.5, -1.5)$	$0.4938-0.4332$	$16\times 0.606 = 0.97$
4	(3, 5)	$(-1.5, 0.5)$	$0.4332-0.1915$	$16\times 0.2417 = 3.9$
6	(5, 7)	$(-0.5, 0.5)$	$0.1915+0.1915$	$16\times 0.383 = 6.1$
8	(7, 9)	$(0.5, 1.5)$	$0.4332-0.1915$	$16\times 0.2417 = 3.9$
10	(9,11)	$(1.5, 2.5)$	$0.4938-0.4332$	$16\times 0.606 = 0.97$

Hence the expected (theoretical) frequencies corrected to nearest integer are 1, 4, 6, 4, 1 which agree with the observed frequencies. This shows that the normal curve (i) is proper fit to the given distribution.

Q.No.6.: Show that the standard deviation for a normal distribution is approximately 25% more than the mean deviation.

Sol.: Since the mean deviation from the mean μ

$$= \int_{-\infty}^{\infty} |x-\mu| . \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-\mu)^2/2\sigma^2} dx \;, \qquad [\text{Put } z = \frac{(x-\mu)}{\sigma}]$$

$$= \frac{\sigma}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} |z| e^{-\frac{z^2}{2}} dz = \frac{\sigma}{\sqrt{(2\pi)}} \left[\int_{-\infty}^{0} ze^{-\frac{z^2}{2}} dz + \int_{0}^{\infty} ze^{-\frac{z^2}{2}} dz \right] = \frac{2\sigma}{\sqrt{(2\pi)}} \int_{0}^{\infty} ze^{-\frac{z^2}{2}} dz$$

$$= \frac{\sigma}{\sqrt{(2\pi)}} \left| -e^{-\frac{z^2}{2}} \right|_0^{\infty} = -\sqrt{\left(\frac{2}{\pi}\right)} \sigma(0-1) = 0.7979\sigma = \left(\frac{4}{5}\right)\sigma .$$

Thus mean deviation = $\left(\frac{4}{5}\right)$ times standard deviation.

$\Rightarrow$ standard deviation = 1.25 times mean deviation

Hence the standard deviation for a normal distribution is approximately 25% more than the mean deviation.

Q.No.7.: For a normally distributed variate with mean 1 and S. D. 3, find the probabilities that

(i) $3.43 \le x \le 6.19$, **(ii)** $-1.43 \le x \le 6.19$

Sol.: Given mean , $\mu = 1$, and standard deviation (S. D), $\sigma = 3$.

(i). To find the probability that $3.43 \le x \le 6.19$.

i. e $P(x_1 \le x \le x_2) = P\ (3.43 \le x \le 6.19)$.

Now when $x_1 = 3.43$, $z_1 = \dfrac{x_1 - \mu}{\sigma} = \dfrac{3.43 - 1}{3} = 0.81$,

and when $x_2 = 6.19$, $z_2 = \dfrac{x_2 - \mu}{\sigma} = \dfrac{6.19 - 1}{3} = 1.73$.

Since we know that the probability of x lying between x_1 and x_2 is given by the area under the normal curve from x_1 to x_2.

i. e. $P(x_1 \le x \le x_2) = \dfrac{1}{\sqrt{2\pi}}\left[\int_0^{z_2} e^{-z^2/2}dz - \int_0^{z_1} e^{-z^2/2}dz\right]$.

The values of each of the above integral can be found from the table of area under the normal curve from 0 to z.

Thus $P(3.43 \le x \le 6.19) = P(0.81 \le z \le 1.73) = P(0 < z \le 1.73) - P(0 < z < 0.81)$

$$= 0.4582 - 0.2910 = 0.1672\text{. Ans.}$$

(ii). To find the probability that $-1.43 \le x \le 6.19$. i. e

$P(x_1 \le x \le x_2) = P(-1.43 \le x \le 6.19)$.

Now when $x_1 = -1.43$, $z_1 = \dfrac{x_1 - \mu}{\sigma} = \dfrac{-1.43 - 1}{3} = -0.81$,

and when $x_2 = 6.19$, $z_2 = \dfrac{x_2 - \mu}{\sigma} = \dfrac{6.19 - 1}{3} = 1.73$.

Thus $P(-1.43 \le x \le 6.19) = P(-0.81 \le z \le 1.73) = P(0 \le z \le 1.73) + P(-0.81 \le z \le 0)$

$$= 0.4582 + 0.2910 = 0.7492\text{. Ans.}$$

Q.No.8.: If z is normally distributed with mean 0 and variance 1, find

(i) $P_r\{z \le -1.64\}$, **(ii)** z_1 if $P_r\{z \ge z_1\} = 0.84$.

Sol.: (i). To find : $P_r\{z \le -1.64\}$.

Now $P_r\{z \le -1.64\} = P_r(z \ge 1.64) = P_r(0 < z < \infty) - (0 < z < 1.64)$

$$= 0.5 - 0.4495 = 0.0505\text{. Ans.}$$

(ii) To find : z_1 if $P_r\{z \ge z_1\} = 0.84$.

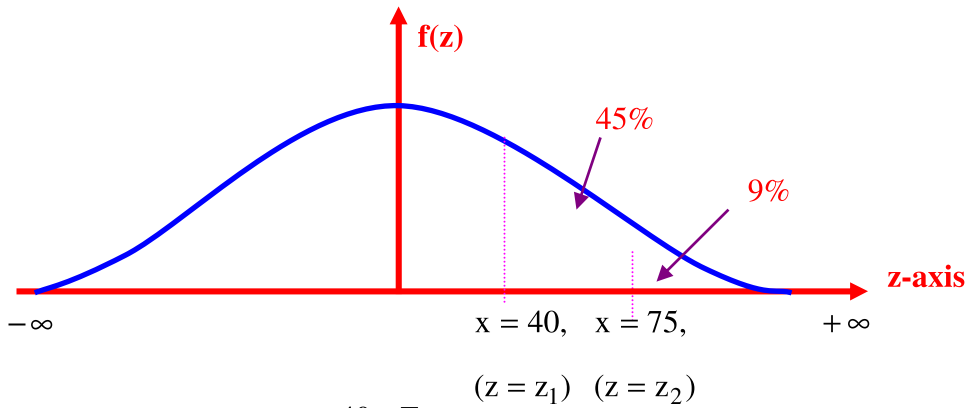
Now $P_r(z \geq z_1) = 0.34 + 0.5 = P(-0.995 \leq z \leq 0) + P_r(z > 0)$

$z_1 = -0.995$. Ans.

Q.No.9.: In a certain examination, the percentage of candidates passing and getting distinctions were 45 and 9 respectively. Estimate the average marks obtained by the candidates, the minimum pass and distinction marks being 40 and 75 respectively. (Assume the distribution of marks to be normal).

Sol.: Let $\bar{x}$ be the mean and σ the S. D.

Now 45% of the students are above 40 marks means area to the right of the ordinate x = 40 and 09% of the students are above 75 marks means area to the right of the ordinate x = 75



When x = 40, let $z = z_1$ so that $z_1 = \dfrac{40 - \bar{x}}{\sigma} = 0.5 - 0.45 = 0.05$ (i)

When x = 75, let $z = z_2$ so that $z_2 = \dfrac{75 - \bar{x}}{\sigma} = 0.5 - 0.09 = 0.41$ (ii)

From (i) and (ii), we get

$40 - \bar{x} = 0.05\sigma$.

$75 - \bar{x} = 0.41\sigma$.

Solving these equations, we get $\bar{x} = 35.13$. Ans.

Q.No.10.: A manufacturer of air-mail envelopes knows from experience that the weight of the envelopes is normally distributed with mean 1.95 gm and standard deviation 0.05 gm. About how many envelopes weighing

(i) 2 gm or more;

(ii) 2.05 gm or more can be expected in a given packet of 100 envelopes.

Sol.: Given mean , $\mu = 1.95$, and standard deviation (S. D), $\sigma = 0.05$.

(i). To find : Total number of envelopes weighing 2 gm or more.

Now when $x = 2$, $z = \dfrac{x-\mu}{\sigma} = \dfrac{2-1.95}{0.05} = 1$,

From the table of area under the normal curve from 0 to z.

Area against $z = 1$ equal to 0.3413.

Now $P\{x > 2\} = P(z \geq 1) = P(0 < z < \infty) - (0 < z < 1)$

$= 0.5 - 0.3413 = 0.1587$.

$\therefore$ Total number of envelopes weighing 2 gm or more

$= 100 \times 0.1587 = 15.87 = 16$. (nearly) Ans.

(i). To find : Total number of envelopes weighing 2.05 gm or more can be expected in a given packet of 100 envelopes.

Now when $x = 2.05$, $z = \dfrac{x-\mu}{\sigma} = \dfrac{2.05-1.95}{0.05} = 2$,

From the table of area under the normal curve from 0 to z.

Area against $z = 2$ equal to 0.4772.

Now $P\{x > 2.05\} = P(z \geq 2) = P(0 < z < \infty) - (0 < z < 2)$

$= 0.5 - 0.4772 = 0.0228$.

$\therefore$ Total number of envelopes weighing 2.05 gm or more

$= 100 \times 0.0228 = 2.28 = 2$. (nearly) Ans.

Q.No.11.: The mean height of 500 students is 151cm. and the standard deviation is 15 cm. Assuming that the heights are normally distributed, find how many students' heights lie between 120 and 155 cm.

Sol.: Given total number of student = 500.

Mean height, $\mu = 151$cm. Standard deviation, $\sigma = 15$ cm.

For x = 120, $z = \dfrac{x-\mu}{\sigma} = \dfrac{120-151}{15} = \dfrac{-31}{15} = -2.066 \sim -2.07$.

For x = 155, $z = \dfrac{155-151}{15} = \dfrac{4}{15} = 0.266 \approx 0.27$.

From the table of area under the normal curve from 0 to z.

Area against $z = -2.07$ equal to 0.4808.

Area against z = 0.27 is equal to 0.1064.

$\therefore$ Total area = 0.4808 + 0.1064 = 0.5872.

$\therefore$ The required number of students whose height lie between 120 and 155 cm. are

$0.5872 \times 500 = 293.6 = 294$ (approximately). Ans.

Q.No.12.: The mean and standard deviation of the marks obtained by 1000 students in an examination are respectively 34.4 and 16.5.Assuming the normality of the distribution, find the approximate number of students expected to obtain marks between 30 and 60.

Sol.: Given the total number of students = 1000, mean of marks, $\mu = 34.4$ and standard deviation of marks, $\sigma = 16.5$.

For x = 30, $z = \dfrac{x - \mu}{\sigma} = \dfrac{30 - 34.4}{16.5} = \dfrac{-4.4}{16.5} = -0.266 \approx -0.27$

The area required in this case is to the right of $z = -0.27$

$\therefore$ Area against $z = -0.27$ is equal to 0.1064.

For x = 60, $z = \dfrac{x - \mu}{\sigma} = \dfrac{60 - 34.4}{16.5} = \dfrac{25.4}{16.5} = 1.55$.

The area required in this case is to the left of z = 1.55.

$\therefore$ Area against z = 1.55 is equal to 0.4394.

$\therefore$ Total area = 0.1064+0.4394 = 0.5458.

The approximate number of students expected to obtain marks between 30 and 60 are

$= 0.5458 \times 1000 = 545.8 = 546$ (approximately). Ans.

Q.No.13.: In an examination taken by 500 candidates, the average and the standard deviation of marks obtained (normally distributed) are 40% and 10%. Find approximately

(i) how many will pass, if 50% is fixed as a minimum?

(ii) what should be the minimum if 350 candidates are to pass?

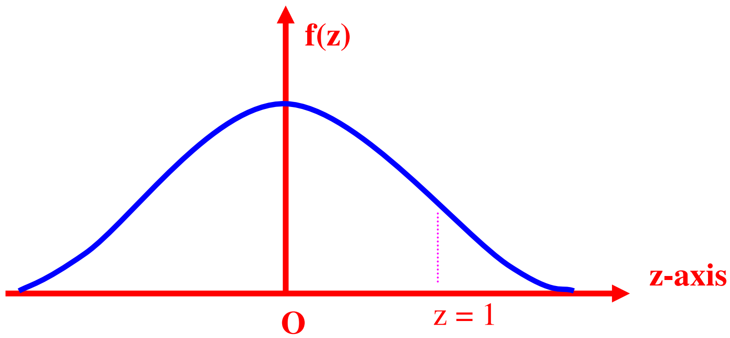
(iii) how many have scored marks above 60%?

Sol.: (i). Given total number of students = 500.

Mean of marks, $\mu = 40$, standard deviation of marks, $\sigma = 10$.

For x = 50, $z = \dfrac{x - \mu}{\sigma} = \dfrac{50 - 40}{10} = 1$.

The area required in this case is to the right of z = 1.



$\therefore$ Area against z = 1 is equal to 0.3413.

$\therefore P(z > 1) = P(0 < z < \infty) - P(0 < z < 1) = 0.5 - 0.3413 = 0.1587$.

$\therefore$ Number of students passing $= 0.1587 \times 500 = 79$. Ans.

(ii). Given number of students N = 350.

$P \times 500 = 350 \Rightarrow P(z > z_1) = 0.7 \Rightarrow P(z > z_1) = 0.2 + 0.5$

In order to satisfy the given condition for passing 350 students,

the value of $z = -0.525 \Rightarrow -0.525 = \dfrac{x - 40}{10} \Rightarrow x = 34.75$. Ans.

$\therefore$ Minimum 34.8% marks must be obtained if 350 candidates are to pass.

(iii). If $x = 60\%$, then $z = \dfrac{x - \mu}{\sigma} = \dfrac{60 - 40}{10} = 2$.

Area against z =2 is equal to 0.4772.

The area required in this case is to the right of z = 0.4772.

$\therefore P(z > 2) = P(0 < z < \infty) - P(0 < z < 0.4772)$

$= 0.5 - 0.4772 = 0.0228$.

$\therefore$ Total number of students scored marks above 60% $= 0.0228 \times 500 = 11.4 = 11$ students (approximately). Ans.

Q.No.14.: The mean inside diameter of a sample of 200 washers produced by a machine is 5.02 mm and the standard deviation is 0.05 mm. The purpose for which these washers are intended allows a maximum tolerance in the diameter of 4.96 to 5.08 mm, otherwise the washers are considered defective. Determine the percentage of defective washers produced by the machine, assuming the diameters are normally distributed.

Sol.: Given mean $\mu = 5.02$ mm. and S. D. $\sigma = 0.05$ mm.

For x = 4.96 mm, $z = \frac{x-\mu}{\sigma} = \frac{4.96-5.02}{0.05} = \frac{-0.06}{0.05} = -1.2$.

For x = 5.08 mm, $z = \frac{x-\mu}{\sigma} = \frac{5.08-5.02}{0.05} = \frac{-0.06}{0.05} = 1.2$.

The area between $z = -1.2$ to $z = 1.2$

= 2(area between z = 0 and z = 1.2)

$= 2 \times 0.3849 = 0.7698$ or 77%. (nearly).

Thus the proportion of non-defective washers = 77%. nearly

$\therefore$ The percentage of defective washers produced by the machine = $100 - 77 = 23\%$.

Q.No.15.: Assuming that the diameters of 1000 brass plugs taken consecutively from a machine, form a normal distribution with mean 0.7515 cm. and standard deviation 0.0020cm., how many of the plugs are likely to be rejected if the approved diameter is 0.752 ± 0.004 cm.?

Sol.: Given total number of brass plugs = 1000, S. D., $\sigma = 0.002$, mean, $\mu = 0.7515$.

$$\therefore z_1 = \frac{x_1-\mu}{\sigma} = \frac{0.756-0.7515}{0.002} = 2.25$$

$$z_2 = \frac{x_2-\mu}{\sigma} = \frac{0.748-0.7515}{0.002} = -1.75.$$

From table of area under the normal curve from 0 to z, we get

Area against z = 2.25 is equal to 0.4878 and against $z = -1.75$ is equal to 0.4599.

$\therefore$ Total area $= P(-1.75 < z < 2.25) = 0.4878 + 0.4599 = 0.9477$ i. e. 94.77%

$\therefore$ Total number of unrejected plugs = $94.77\% \times 1000 = 948$ plugs.

Hence rejected plugs $= 1000 - 948 = 52$ plugs. Ans.

Q.No.16.: It is given that the age of thermostats of a particular make follow the normal law with mean 5 years and S. D. 2 years. 1000 units are sold out every month. How many of them have to be replaced at the end of the second year.

Sol.: In this problem mean μ = 5, standard deviation σ = 2, N = 1000, x = 2.

$$\therefore z = \frac{x-\mu}{\sigma} = \frac{2-5}{2} = -1.5.$$

The area against $z = -1.5$ in the table is 0.4332.

We, however, require the area to the left of the ordinate $z = -1.5$.

This area = $0.5 - 0.4332 = 0.0668$.

Thus the number of thermostats of a particular make which have to be replaced at the end of the second year = $1000 \times 0.0668 = 66.7 = 67$ (nearly). Ans.

Q.No.17.: Find the equation of the best fitting normal curve to the following distribution:

x :	0	1	2	3	4	5
f :	13	23	34	15	11	4

Sol.: Mean $= \mu = \dfrac{\sum fx}{\sum f} = \dfrac{0+23+68+45+44+20}{100} = \dfrac{200}{100} = 2$.

$$\text{S. D.} = \sqrt{\left[\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2\right]} = \sqrt{[5.7-4]} = 1.3038$$

Taking $\mu = 2$, $\sigma = 1.3038$ and $N = 100$, the equation of the normal curve is

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \Rightarrow y = \frac{100}{\sqrt{3.4\pi}} e^{-(x-2)^2/3.4},$$

which is the required equation of the best fitting normal curve.

Q.No.18.: Obtain the equation of the normal probability curve that may be fitted to the following data:

Variable x:	4	6	8	10	12	14	16	18	20	22	24
Frequency f:	1	7	15	22	35	43	38	20	13	5	1

Sol.: Mean $\mu = \dfrac{\sum fx}{\sum f} = \dfrac{4+42+120+220+420+602+608+360+260+110++24}{200}$

$$= \frac{2770}{200} = 13.85. \text{ Ans.}$$

$$\text{S. D.} = \sqrt{\left[\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2\right]}$$

Now $\dfrac{\sum fx.x}{\sum f} = \dfrac{16+252+960+2200+5040+8428+9728+6480+5200+2420+576}{200}$

$$= \frac{41300}{200} = 206.5$$

$$\left(\frac{\sum fx}{\sum f}\right)^2 = (13.85)^2 = 191.8225$$

S. D. $= \sqrt{206.5 - 191.8225} = \sqrt{14.6775} = 3.83$. Ans

Taking $\mu = 13.85,\ \sigma = 3.83$ and $N = 200$, the equation of the normal probability curve is

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \Rightarrow y = \frac{200}{\sqrt{29.34\pi}} e^{-(x-13.85)^2/29.34}, \qquad \text{(i)}$$

which is the required equation of the best fitting normal probability curve.

By substituting $x = x_1,\ x_2,\ x_3, \ldots\ldots\ldots\ldots$, we get probability equation for different values. (i) gives the main or general equation for normal probability curve, having $\mu = 13.85$. $\sigma = 3.83$.

Q.No.19.: A factory turns out an article by mass production and it is found that 10% of the product is rejected. Find the S. D. of the number of rejects and the equation to the normal curve to represent the number of rejects.

Sol.: Given data shows that if n = 100, then

p is the success having probability $= \frac{10}{100} = 0.1$,

and q is that of failure $= 1 - \frac{10}{100} = 0.9$,

$\therefore$ Binomial distribution of rejects gives mean, $\mu = np = 100 \times 0.1 = 10$.

and S. D. $= \sigma = \sqrt{npq} = \sqrt{100 \times 0.1 \times 0.9} = \sqrt{9} = 3$. Ans.

Now if this binomial distribution is approximated by a normal distribution, then the equation to the normal curve is

$$y = \frac{n}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \Rightarrow y = \frac{100}{3\sqrt{2\pi}} e^{\frac{-(x-10)^2}{2.(3)^2}} \Rightarrow y = \frac{100}{\sqrt{18\pi}} e^{\frac{-(x-10)^2}{18}}.$$

This equation represent to the normal curve to represent the number of rejects.

Q.No.20.: Given that the probability of committing an error of magnitude x is

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}, \text{ show that the probable error is } \frac{0.4769}{h}.$$

Sol.: Given probability of committing an error of magnitude x is $y = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2}$

Comparing this equation with $y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, we get

$$h = \frac{1}{\sigma\sqrt{2}} \qquad \text{(i)}$$

$$h^2 x^2 = \frac{(x-\mu)^2}{2\sigma^2} \Rightarrow hx = \frac{x-\mu}{\sigma\sqrt{2}} \qquad \text{(ii)}$$

Put $h = \frac{1}{\sigma\sqrt{2}}$, we get

$$\frac{x}{\sqrt{2}\,\sigma} = \frac{x-\mu}{\sqrt{2}\,\sigma} \Rightarrow x = x-\mu \qquad \text{(iii)}$$

This is only possible when $\mu = 0$

Put in (ii), we get $hx = \frac{x}{\sigma\sqrt{2}} \Rightarrow \sigma = \frac{1}{h\sqrt{2}}$.

Now we know that probable error $= \frac{2}{3}\sigma = \frac{2}{3}\frac{1}{h\sqrt{2}} = \frac{\sqrt{2}}{3}h = \frac{0.4714}{h}$,

Hence this proves the result.

*** *** *** *** ***

*** *** ***

***

Topic

Probability

Baye's Theorem

Prepared by:

Dr. Sunil

NIT Hamirpur (HP)

(Last updated on 18-11-2007)

Baye's Theorem: (Theorem of inverse probability)

(Formula for the probability of "causes")

Statement: Let the sample space S be divided into n subsets $B_1, B_2, \ldots\ldots\ldots, B_n$, with $P(B_i) \neq 0$ for i = 1,2,3,….,n. For any event arbitrary event A in S corresponds to a number of mutually exclusive and exhaustive events $B_1, B_2, \ldots\ldots\ldots, B_n$ with $P(A) \neq 0$, then

$$P\left(\frac{B_i}{A}\right) = \frac{P(B_i)P\left(\frac{A}{B_i}\right)}{\sum P(B_i)P\left(\frac{A}{B_i}\right)}.$$

Proof: By multiplication law of probability, i.e.

Multiplication law of probability or Theorem of compound probability:

If the probability of an event A happening as a result of trial is P(A) and after A has happened the probability of an event B happening as a result of another trial (i. e. conditional probability of B given A) is $P\left(\frac{B}{A}\right)$, then the probability of both the events A and B happening as a result of two trials is P(AB) or $P(A \cap B) = P(A).P\left(\frac{B}{A}\right)$.

we have $P(AB_i) = P(A)P\left(\frac{B_i}{A}\right) = P(B_i)P\left(\frac{A}{B_i}\right)$. (i)

$$\therefore P\left(\frac{B_i}{A}\right) = \frac{P(B_i)P\left(\frac{A}{B_i}\right)}{P(A)} \quad \text{(ii)}$$

Since the event A corresponds to $B_1, B_2, \ldots\ldots\ldots, B_n$, we have by the additional law of probability or theorem of total probability, i.e.

Addition law of probability or Theorem of total probability:

If the probability of an event happening as a result of a trial is P(A) and the probability of a mutually exclusive event B happening is P(B), then the probability of either of the events happening as a result of the trial is P(A+B) or $P(A \cup B) = P(A) + P(B)$.

we have

$$P(A) = P(AB_i) + P(AB_2) + \ldots\ldots\ldots\ldots P(AB_n) = \sum P(AB_i) = \sum P(B_i)P\left(\frac{A}{B_i}\right). \qquad \text{[by (i)]}$$

Hence, from (ii), we have

$$P\left(\frac{B_i}{A}\right) = \frac{P(B_i)P\left(\frac{A}{B_i}\right)}{\sum P(B_i)P\left(\frac{A}{B_i}\right)},$$

which is also known as the **theorem of inverse probability**.

Baye's theorem is also known as **formula for the probability of "causes",** i.e., probability of particular (cause) B_i, given that event A has happened already.

Remarks:

1. The probabilities $P(B_i)$, i = 1, 2, ……., n are called **apriori probabilities** because these exist before we get any information from the experiment.
2. The probabilities $P(A/B_i)$, i = 1, 2, ……., n are called **postapriori probabilities,** because these are found after the experimental results are known.

Now let us solve some problems, where 'Baye's theorem' will be used for evaluating the probability.

Q.No.1.: There are three bags: first containing 1 white, 2 red, 3green balls: second 2 white, 3 red, 1 green balls and third 3 white, 1 red , 2 green balls. Two balls are drawn from a bag chosen at random. These are found to be **one white** and **one red.** Find the probability that the balls so drawn came from the second bag.

Sol.: Let B_1, B_2, B_3 pertain to first, second, third bags chosen and A: the two balls are white and red.

Now $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3} = 0.333$.

$$P\left(\frac{A}{B_1}\right) = P\ \text{(a white and a red ball are drawn from first bag)} = \frac{\left({}^1C_1 \times {}^2C_1\right)}{{}^6C_2} = \frac{2}{15} = 0.1333.$$

Similarly, $P\left(\frac{A}{B_2}\right)=\frac{\left({}^2C_1 \times {}^3C_1\right)}{{}^6C_2}=\frac{2}{5}=0.4$, $P\left(\frac{A}{B_3}\right)=\frac{\left({}^3C_1 \times {}^1C_1\right)}{{}^6C_2}=\frac{1}{5}=0.2$.

By Baye's theorem, we have $P\left(\frac{B_2}{A}\right)=\frac{P(B_2)P\left(\frac{A}{B_2}\right)}{P(B_1)P\left(\frac{A}{B_1}\right)+P(B_2)P\left(\frac{A}{B_2}\right)+P(B_3)P\left(\frac{A}{B_3}\right)}$

$$=\frac{\frac{1}{3}\times\frac{2}{5}}{\frac{1}{3}\times\frac{2}{15}\times\frac{1}{3}\times\frac{2}{5}+\frac{1}{3}\times\frac{1}{5}}=\frac{6}{11}=0.5455\text{. Ans.}$$

Q.No.2.: Three machines M_1, M_2 and M_3 produce identical items. Of their respective output 5%, 4% and 3% of items are faulty. On a certain day, M_1 has produced 25% of the total output, M_2 has produced 30% and M_3 the remainder. An item selected at random is found to be faulty. What are the chances that it was produced by the machine with the highest output?

Sol.: Let the event of drawing a faulty item from any of the machine be A, and the event that an item drawn at random was produced by M_i be B_i.

We have to find $P(B_i/A)$ for which we proceed as follow:

	M_1	M_2	M_3	Remarks
$P(B_i)$	0.25	0.30	0.45	$\because$ Sum = 1
$P(A/B_i)$	0.05	0.04	0.03	
$P(B_i)P(A/B_i)$	0.0125	0.12	0.0135	Sum = 0.38
$P(B_i/A)$	$\frac{0.0125}{0.038}$	$\frac{0.012}{0.038}$	$\frac{0.0135}{0.038}$	By Baye's theorem

The highest output being from M_3, the required probability = $\frac{0.0135}{0.038}$=0.355. Ans.

Q.No.3.: In a certain college 25% of boys and 10% of girls are studying mathematics. The girls continue 60% of the students body.

(a) What is the probability that Mathematics is being studies?

(b) If a student is selected at random and is found to be studying Mathematics, find the probability that the student is a girl?

(c) If a student is selected at random and is found to be studying Mathematics, find the probability that the student is a boy?

Sol.: Given that P(Boy) = P(B) $=\frac{40}{100}=\frac{2}{5}$;

$$P(\text{Girl}) = P(G) = \frac{60}{100} = \frac{3}{5},$$

Probability that Maths is studied given that the student is a boy $= P\left(\frac{M}{B}\right) = \frac{25}{100} = \frac{1}{4}$.

Similarly, $P\left(\frac{M}{G}\right) = \frac{10}{100} = \frac{1}{10} = 0.1$.

(a) Probability that the Maths is studied $= P(M) = P(G)P\left(\frac{M}{G}\right) + P(B)P\left(\frac{M}{B}\right)$.

By theorem on total probability, we have

$$P(M) = \frac{3}{5}.\frac{1}{10} + \frac{2}{5}.\frac{1}{4} = \frac{4}{25} = 0.16. \text{ Ans.}$$

(b) By Baye's theorem, we have

Probability that a Maths student is a girl $= P\left(\frac{G}{M}\right) = \dfrac{P(G)P\left(\frac{M}{G}\right)}{P(M)} = \dfrac{\frac{3}{5}.\frac{1}{10}}{\frac{4}{25}} = \frac{3}{8} = 0.375$. Ans.

(c) Probability that a Maths student is a boy $= P\left(\frac{B}{M}\right) = \dfrac{P(B)P\left(\frac{M}{B}\right)}{P(M)} = \dfrac{\frac{2}{5}.\frac{1}{4}}{\frac{4}{25}} = \frac{5}{8} = 0.625$. Ans.

Q.No.4.: A businessman goes to hotels X, Y, Z 20%, 50%, 30% of the time, respectively. It is known that 5%, 4% and 8% of rooms in X, Y, Z hotels has faulty plumbing.

(a) Determine the probability that the businessman goes to hotel with faulty plumbing.

(b) What is the probability that businessman's room having faulty plumbing is assigned to hotel Z?

Sol.: A: event of faulty plumbing

$B_1 = X, \ B_2 = Y, \ B_3 = Z.$

(a). By theorem on total probability, we have

$$P(\text{Faulty plumbing}) = P(A) \sum_{i=1}^{3} P(B_i)P\left(\frac{A}{B_i}\right) = P(X)P\left(\frac{A}{X}\right) + P(Y)P\left(\frac{A}{Y}\right) + P(Z)P\left(\frac{A}{Z}\right).$$

It is known (given) that $P(B_1) = P(X) = \frac{20}{100} = 0.2$,

$$P(B_2) = P(Y) = \frac{50}{100} = 0.5,$$

$$P(B_3) = P(Z) = 0.3,$$

$$P\left(\frac{A}{X}\right)=\frac{5}{100}=0.05,$$

$$P\left(\frac{A}{Y}\right)=\frac{4}{100}=0.04,$$

$$P\left(\frac{A}{Z}\right)=\frac{8}{100}=0.08.$$

Thus, P(Faulty plumbing) = P(A) = (0.2)(0.05) + (0.5)(0.04) + (0.3)(0.08) = 0.054. Ans.

(b). P(Assigned to hotel Z given that room has faulty plumbing) $=P\left(\frac{Z}{A}\right)$.

By Baye's theorem, we have $P\left(\frac{Z}{A}\right)=\frac{P(Z)P\left(\frac{A}{Z}\right)}{P(A)}=\frac{(0.3)(0.08)}{0.054}=\frac{4}{9}=0.444$. Ans.

Q.No.5.: A box X contains 2 white and 3 red balls and a bag Y contains 4 white and 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 bag Y.

Sol.: Let E_1: the event that the ball is drawn from bag X;

E_2: the event that the ball is drawn from bag Y;

and A : the event that the ball is red.

We have to find $P\left(\frac{E_2}{A}\right)$:

By Baye's theorem, we have $$P\left(\frac{E_2}{A}\right)=\frac{P(E_2)P\left(\frac{A}{E_2}\right)}{P(E_1)P\left(\frac{A}{E_1}\right)+P(E_2)P\left(\frac{A}{E_2}\right)}. \qquad \text{(i)}$$

Since the two bags are equally likely to be selected, $P(E_1)=P(E_2)=\frac{1}{2}$.

Also $P\left(\frac{A}{E_1}\right)$ = P(a red ball is drawn from bag X) = $\frac{3}{5}$.

$P\left(\frac{A}{E_2}\right)$ = P(a red ball is drawn from bag Y) = $\frac{5}{9}$.

From (i), we get $P\left(\frac{E_2}{A}\right)=\frac{\frac{1}{2}\times\frac{5}{9}}{\frac{1}{2}\times\frac{3}{5}+\frac{1}{2}\times\frac{5}{9}}=\frac{25}{52}=0.48077$. Ans.

Q.No.6.: In a bolt factory, machines A, B and C manufacture respectively 25%, 35% and 40% of the total. Of their output 5, 4, and 2 percent are defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the probability that it was manufactured by machine B.

Sol.: Let E_1, E_2 and E_3 denote the events that a bolt selected at random is manufacture by machines A, B, and C respectively and let H denote the event of its being defective. Then

$P(E_1) = 0.25$, $P(E_2) = 0.35$, $P(E_3) = 0.40$.

The probability of drawing a defective bolt manufactured by machine A is $P\left(\frac{H}{E_1}\right) = 0.05$.

Similarly, $P\left(\frac{H}{E_2}\right) = 0.04$ and $P\left(\frac{H}{E_3}\right) = 0.02$.

By Baye's theorem, we have

$$P\left(\frac{E_2}{H}\right) = \frac{P(E_2)P\left(\frac{H}{E_2}\right)}{P(E_1)P\left(\frac{H}{E_1}\right) + P(E_2)P\left(\frac{H}{E_2}\right) + P(E_3)P\left(\frac{H}{E_3}\right)}.$$

$$= \frac{0.35\times 0.04}{0.25\times 0.05 + 0.35\times 0.04 + 0.40\times 0.02} = \frac{0.0140}{0.0345} = 0.41. \text{ Ans.}$$

Q.No.7.: The contents of urns I, II, III are as follows;

1 white, 2 black and 3 red balls,

2 white, 1 black and 1 red balls and

4 white, 5 black and 3 red balls.

One urn is chosen at random and two balls drawn. They happen to be white and red. What is the probability that they come from urns I, II or III?

Sol.: Let E_1 : the event that urn I chosen,

E_2 : the event that urn II chosen,

E_3 : the event that urn III chosen, and

A : the event that the two balls are white and red.

We have to find $P\left(\frac{E_1}{A}\right)$, $P\left(\frac{E_2}{A}\right)$ and $P\left(\frac{E_3}{A}\right)$.

Now $P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$.

$P\left(\frac{A}{E_1}\right)$ = P (a white and a red ball are drawn from urn I) $= \frac{{}^1C_1 \times {}^3C_1}{{}^6C_2} = \frac{1}{5}$.

$$P\left(\frac{A}{E_2}\right)=\frac{{}^2C_1\times{}^1C_1}{{}^4C_2}=\frac{1}{3}, \qquad P\left(\frac{A}{E_3}\right)=\frac{{}^4C_1\times{}^3C_1}{{}^{12}C_2}=\frac{2}{11}.$$

By Baye's theorem, we have

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

$$=\frac{\frac{1}{3}\times\frac{1}{5}}{\frac{1}{3}\times\frac{1}{5}+\frac{1}{3}\times\frac{1}{3}+\frac{1}{3}\times\frac{2}{11}}=\frac{\frac{1}{15}}{\frac{1}{15}+\frac{1}{9}+\frac{2}{33}}$$

$$=\frac{0.0667}{0.0667+0.111+0.06061}=\frac{0.0667}{0.23831}=0.279661.\text{Ans.}$$

Similarly,

$$P\left(\frac{E_2}{A}\right)=\frac{P(E_2)P\left(\frac{A}{E_2}\right)}{P(E_1)P\left(\frac{A}{E_1}\right)+P(E_2)P\left(\frac{A}{E_2}\right)+P(E_3)P\left(\frac{A}{E_3}\right)}=\frac{0.111}{0.23831}=0.466\text{. Ans.}$$

$$P\left(\frac{E_3}{A}\right)=\frac{P(E_3)P\left(\frac{A}{E_3}\right)}{P(E_1)P\left(\frac{A}{E_1}\right)+P(E_2)P\left(\frac{A}{E_2}\right)+P(E_3)P\left(\frac{A}{E_3}\right)}=\frac{0.0606}{0.23831}=0.454\text{. Ans.}$$

*** *** *** *** ***

*** *** ***

***

Home Assignments

Q.No.1.: In a certain college, 4% of the boys and 1% girls are taller than 1.8 m. Further more 60% of the students are girls. If a student is selected at random and is found to be taller than 1.8 m., what is the probability that the student is a girl.

Ans.: 3/11, i.e., 0.2727.

Q.No.2.: In a bolt factory, machine A, B and C manufacture 25%, 35% and 40% of the total. Of their output 5%, 4% and 2% are defective bolts. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by machines A, B or C?

Ans.: 25/69, i.e.,0.3623, 28/69, i.e.,0.4058, 16/69, i.e.,0.2319.

Q.No.3.: In a bolt factory, there are four machine A, B, C and D manufacturing 20%, 15%, 25% and 40% of the total output, respectively. Of their output 5%, 4%, 3% and 2% in the same order are defective bolts. A bolt is chosen at random from the factory production and is found defective. What is the probability that it was manufactured by machine A or machine D?

Ans.: 0.3175, 0.254.

Q.No.4.: The contents of three urns are : 1 white, 2 red, 3 green balls; 2 white, 1 red, 1 green balls and 4 white, 5 red, 3 green balls. Two balls are drawn from an urn chosen at random. These found to be one white and one green.

Find the probability that the balls so drawn came from third urn.

Ans.: 15/59, i.e.,0.2542.

Q.No.5.: Two urns contain 4 white, 6 blue and 4 white, 5 blue balls, respectively. One of the urns is selected at random and a ball is drawn from it. If the ball drawn is white, find the probability that it is drawn from the **(i)** first urn, **(ii)** second urn.

Ans. (i) $\frac{9}{19}$, i.e., 0.4737 **(ii)** $\frac{10}{19}$, i.e., 0.5263**.**

Q.No.6.: Three urns contain 6 red 4 black; 4 red 6 black; 5 red 5 black balls, respectively. One of the urns is selected at random and a ball is drawn from it. If the ball drawn is red, find the probability that it is drawn from the first urn.

Ans. $\frac{2}{5}$, i.e.,0.4**.**

Q.No.7.: A factory has two machines A and B. Past record shows that machine A produced 60% of the items of output and machine B produced 40% of the items. Further, 2% items produced by machine A were defective and 1% produced by machine B were defective. If a defective item is drawn at random, what is the probability that it was produced by machine A.

Ans. $\frac{3}{4}$, i.e.,0.75**.**

Q.No.8.: An insurance company insured 2000 scooter drivers, 4000 car drivers and 6000 truck drivers. The probability of accident is 0.01, 0.03 and 0.15, respectively. One of the insured persons meets an accident. What is the probability that he is scooter driver.

Ans. $\frac{1}{52}$, i.e.,0.0192**.**

Q.No.9.: A company has two plants to manufacture scooters. Plant I manufactures 70% of scooters and plant II manufactures 30%. At plant I, 80% scooters are standard quality and at plant II, 90% of scooters are rated standard quality. A scooter is chosen at random and is found to be of standard quality. What is the chance that it has come from plant II?

Ans. $\frac{27}{83}$, i.e.,0.3253**.**

Q.No.10.: Companies B_1, B_2 and B_3 produced 30%, 45% and 25% of the cars respectively, It is known that 2%, 3% and 2% of cars produced from B_1, B_2 and B_3, are defective.

(a) What is the probability that a car purchased is defective?

(b) If a car purchased is found to be defective, what is the probability that this car is produced by company B_3.

Ans.: **(a).** P(Defective) = (0.3)(0.02) + (0.45)(0.03) + (0.25)(0.02) = 0.0245

(b). $P\left(\frac{B_3}{D}\right) = \frac{(0.25)(0.02)}{P(D)} = \frac{10}{49} = 0.2041$**.**

Q.No.11.: Of the three men, the chances that a politician, a businessman and an academician will be appointed as a vice chancellor of a university are 0.50, 0.30 and 0.20, respectively. Probability that research is promoted by these people if they are appointed as V.C. are 0.3, 0.7 and 0.8, respectively.

(a) Determine the probability that research is promoted in the university.

(b) If research is promoted in the university, what is the probability that VC is an academician?

(c) If research is promoted in the university, what is the probability that VC is a businessman?.

Ans.: (a). P(Research) = (0.5)(0.3) + (0.3)(0.7) + (0.2)(0.8) = 0.52.

(b). $P(\text{Academician/Research}) = \frac{(0.2)(0.8)}{0.52} = 0.30769$.

(c). $P(\text{Businessman/Research}) = \frac{(0.3)(0.7)}{0.52} = 0.4038$.

Q.No.12.: Suppose three companies X, Y, Z produce TV's. X produce twice as many as Y while Y and Z produce the same number. It is known that 2% of X, 2% of Y and 4% of Z are defective. All the TV's produced are put into one shop and then one TV is chosen at random.

(a). What is the probability that the TV is defective?

(b). Suppose a TV chosen is defective, what is the probability that this TV is produced by company X?

Ans.: (a). P(D) = P(X) P(D/X) + P(Y) P(D/Y) + P(Z) P(D/Z)

$$= \frac{1}{2}(0.02) + \frac{1}{4}(0.02) + \frac{1}{4}(0.04) = 0.025.$$

(b). $$P\left(\frac{X}{D}\right) = \frac{(0.02)\left(\frac{1}{2}\right)}{0.025} = 0.40.$$

Q.No.13.: For a certain binary communication channel, the probability that a transmitted '0' is received as a '0' is 0.95 and the probability that a transmitted '1' is received as '1' is 0.90. If the probability that a '0' is transmitted is 0.4 , find the probability that

(i). a '1' is received

(ii) a'1' was transmitted given that '1' was received..

Ans.: **(i).** 0.56. **(ii).** 27/28,i.e., 0.9643.

Hint: A: event of transmitting '1', $\overline{A}$: event of transmitting '0',

B: event of receiving '1', $\overline{B}$: event of receiving '0'.

(i) $$P(B) = P(A)P\left(\frac{B}{A}\right) + P(\overline{A})P\left(\frac{B}{\overline{A}}\right) = (0.6)(0.9) + (0.4)(0.05) = 0.56.$$

(ii) $$P\left(\frac{A}{B}\right) = \frac{P(A)P\left(\frac{B}{A}\right)}{P(B)} = \frac{(0.6)(0.9)}{0.56} = 0.9643.$$

Q.No.14.: A student has to answer a multiple-choice question with 5 alternatives. What is the probability that the student knew the answer given that he answered it correctly?

Ans.: $\frac{5p}{(4p+1)}$, where p = probability that he knew the correct answer.

Hint: B_1: knew right answer, B_2: guesses right answer, A : gets the right answer.

$$P\left(\frac{B_1}{A}\right) = \frac{(p)1}{(p)1 + (1-p)\left(\frac{1}{5}\right)} = \frac{5p}{(4p+1)}, \text{ since } P\left(\frac{A}{B_2}\right) = \frac{1}{5}.$$

*** *** *** *** ***

*** *** ***

***