10

Statistics

10.1 Statistics is a branch of science dealing with the collection of data, organising, summarising, presenting and analysing data and drawing valid conclusions and thereafter making reasonable decisions on the basis of such analysis.

10.2 Frequency distribution is the arranged data, summarised by distributing it into classes or categories with their frequencies.

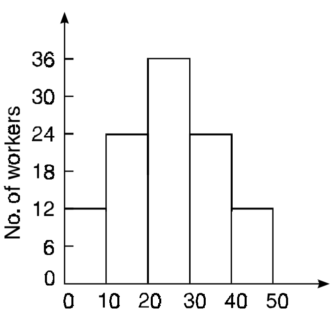
Wages of 100 workers

Wages in Rs.	0-10	10-20	20-30	30-40	40-50
Numbers of workers	12	23	35	20	10

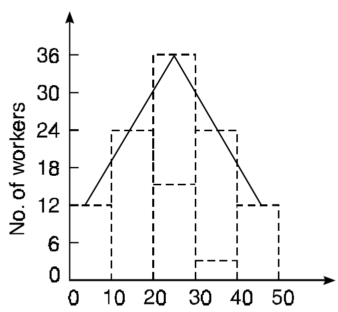
10.3 Graphical representation. It is often useful to represent frequency distribution by means of a diagram. The different types of diagrams are

1. Histogram
2. Frequency polygon
3. Frequency curve
4. Cumulative frequency curve or Ogive
5. Bar chart
6. Circles or Pie diagrams.

1. **Histogram** consists of a set of rectangles having their heights proportional to the class-frequencies, for equal class-intervals. For unequal class-interval, the areas of rectangles are proportional to the frequencies.



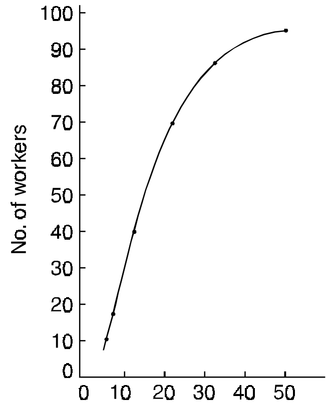
Histogram



Frequency polygon

2. **Frequency Polygon** is a line graph of class-frequency plotted against class-mark. It can be obtained by connecting mid-points on the tops of the rectangles in the histogram.

3. **Cumulative Frequency curve or the Ogive.** If the various points are plotted according to the upper limit of the class as x co-ordinate and the cumulative frequency as y co-ordinate and these points are joined by a free hand smooth curve, the curve obtained is known as cumulative frequency curve or the Ogive.



Frequency polygon

10.4 AVERAGE OR MEASURES OF CENTRAL TENDENCY

An average is a value which is representative of a set of data. Average value may also be termed as measures of central tendency. There are five types of averages in common.

(*i*) Arithmetic average or mean (*ii*) Median (*iii*) Mode

(*iv*) Geometric Mean (*v*) Harmonic Mean

10.5 ARITHMETIC MEAN If $x_1, x_2, x_3, ... x_n$ are n numbers, then their arithmetic mean (A.M.) is defined by

$$A.M. = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\Sigma x}{n}$$

If the number x_1 occurs f_1 times, x_2 occurs f_2 times and so on, then

$$A.M. = \frac{f_1 x_1 + f_2 x_2 + ... + f_n x_n}{f_1 + f_2 + ... + f_n} = \frac{\Sigma f x}{\Sigma f}$$

This is known as direct method.

Example 1. *Find the mean of 20, 22, 25, 28, 30.*

Solution. $$A.M. = \frac{20 + 22 + 25 + 28 + 30}{5} = \frac{125}{5} = 25$$ **Ans.**

Example 2. *Find the mean of the following :*

Numbers	*8*	*10*	*15*	*20*
Frequency	*5*	*8*	*8*	*4*

Solution. $\Sigma f x = 8 \times 5 + 10 \times 8 + 15 \times 8 + 20 \times 4 = 40 + 80 + 120 + 80 = 320$

$$\Sigma f = 5 + 8 + 8 + 4 = 25$$

$$A.M. = \frac{\Sigma f x}{\Sigma f} = \frac{320}{25} = 12.8.$$ **Ans.**

(*b*) **Short cut method**

Let a be the assumed mean, d the deviation of the variate x from a. Then

$$\frac{\Sigma f d}{\Sigma f} = \frac{\Sigma f (x - a)}{\Sigma f} = \frac{\Sigma f x}{\Sigma f} - \frac{\Sigma f a}{\Sigma f} = \text{A.M.} - \frac{a \Sigma f}{\Sigma f} = A.M. - a$$

$$\therefore \quad A.M. = a + \frac{\Sigma f d}{\Sigma f}$$

Example 3. *Find the arithmetic mean for the following distribution :*

Class	*0–10*	*10–20*	*20–30*	*30–40*	*40–50*
Frequency	*7*	*8*	*20*	*10*	*5*

Solution. Let assumed mean (a) = 25.

Class	*Mid-value* x	*Frequency* f	$x - 25 = d$	$f.d$
0–10	5	7	– 20	– 140
10–20	15	8	– 10	– 80
20–30	25	20	0	0
30–40	35	10	+ 10	+ 100
40–50	45	5	+ 20	+ 100
Total		50		– 20

$$A.M. = a + \frac{\Sigma f d}{\Sigma f} = 25 + \frac{-20}{50} = 24.6$$ **Ans.**

(*c*) **Step deviation method**

Let a be the assumed mean, i the width of the class interval and

$$D = \frac{x-a}{i}, \quad A.M. = a + \frac{\Sigma f D}{\Sigma f}\, i$$

Example 4. *Find the arithmetic mean of the data given in example 3 by step deviation method.*

Solution. $a = 25$

Class	*Mid-value* x	*frequency* f	$D = \frac{x-a}{i}$	$f \, . \, D$
0–10	5	7	– 2	– 14
10–20	15	8	– 1	– 8
20–30	25	20	0	0
30–40	35	10	+ 1	+ 10
40–50	45	5	+ 2	+ 10
Total		50		– 2

$$A.M. = a + \frac{\Sigma f D}{\Sigma f} \,.\, i = 25 + \frac{-2}{50} \times 10 = 24.6$$ **Ans.**

10.6 MEDIAN

Median is defined as the measure of the central item when they are arranged in ascending or descending order of magnitude.

When the total number of the items is odd and equal to say n, then the value of $\frac{1}{2}(n + 1)$th item gives the median.

When the total number of the frequencies is even, say n, then there are two middle items, and so the mean of the values of $\frac{1}{2}n$th and $\left(\frac{1}{2}\, n+1\right)$ th items is the median.

Example 5. *Find the median of 6, 8, 9, 10, 11, 12, 13.*

Solution. Total number of items = 7

$$\text{The middle item} = \frac{1}{2}(7+1)^{th} = 4th$$

$$\text{Median} = \text{Value of the 4th item} = 10$$ **Ans.**

For grouped data, Median $= l + \dfrac{\frac{1}{2}N - F}{f} \cdot i$

where l is the lower limit of the median class, f is the frequency of the class, i is the width of the class-interval, F is the total of all the preceeding frequencies of the median-class and N is total frequency of the data.

Example 6. *Find the value of Median from the following data:*

No. of days for which absent (less than)	5	10	15	20	25	30	35	40	45
No. of students	29	224	465	582	634	644	650	653	655

Solution. The given cumulative frequency distribution will first be converted into ordinary frequency as under

Class-Interval	*Cumulative frequency*	*Ordinary frequency*
0–5	29	29 = 29
5–10	$\boxed{224}$	224 – 29 = 195
$\boxed{10\text{–}15}$	465	465 – 224 = $\boxed{241}$
15–20	582	582 – 465 = 117
20–25	634	634 – 582 = 52
25–30	644	644 –634 = 10
30–35	650	650 – 644 = 6
35–40	653	653 – 650 = 3
40–45	655	655 – 653 = 2

$$\text{Median} = \text{size of } \frac{655}{2} \quad \text{or} \quad 327.5\text{th item}$$

327.5th item lies in 10–15 which is the median class.

$$M = l + \frac{\frac{N}{2} - C}{f} i$$

where l stands for lower limit of median class,

N stands for the total frequency,

C stands for the cumulative frequency just preceeding the median class,

i stands for class interval

f stands for frequency for the median class.

$$\text{Median} = 10 + \frac{\frac{655}{2} - 224}{241} \times 5$$

$$= 10 + \frac{103.5 \times 5}{241} = 10 + 2.15 = 12.15$$ **Ans.**

10.7 MODE

Mode is defined to be the size of the variable which occurs most frequently.

Example 7. *Find the mode of the following items :*

0, 1, 6, 7, 2, 3, 7, 6, 6, 2, 6, 0, 5, 6, 0.

Solution. 6 occurs 5 times and no other item occurs 5 or more than 5 times, hence the mode is 6. **Ans.**

For grouped data, $$\text{Mode} = l + \frac{f - f_{-1}}{2f - f_{-1} - f_1} \cdot i$$

where l is the lower limit of the modal class, f is the frequency of the modal class, i is the width of the class, f_{-1} is the frequency before the modal class and f_1 is the frequency after the modal class.

Emperical formula

Mean – Mode = 3 [Mean – Median]

Example 8. *Find the mode from the following data*:

Age	0–6	6–12	12–18	18–24	24–30	30–36	36–42
Frequency	6	11	25	35	18	12	6

Solution.

Age	*Frequency*	*Cumulative frequency*
0–6	6	6
6–12	11	17
12–18	$25 = f_{-1}$	42
$\boxed{18–24}$	$35 = f$	77
24–30	$\boxed{18} = f_1$	95
30–36	12	107
36–42	6	113

$$\text{Mode} = l + \frac{f - f_{-1}}{2f - f_{-1} - f_1} \times i$$

$$= 18 + \frac{35 - 25}{70 - 25 - 18} \times 6$$

$$= 18 + \frac{60}{27} = 18 + 2.22 = 20.22$$ **Ans.**

10.8 GEOMETRIC MEAN

If $x_1, x_2, x_3, \ldots.., x_n$ be n values of variates x, then the geometric mean

$$G = (x_1 \times x_2 \times x_3 \times x_4 \times \ldots \times x_n)^{1/n}$$

Example 9. *Find the geometric mean of 4, 8, 16.*

Solution. $$G.M. = (4 \times 8 \times 16)^{1/3} = 8.$$ **Ans.**

10.9 HARMONIC MEAN

Harmonic mean of a series of values is defined as the reciprocal of the arithmetic mean of their reciprocals. Thus if H be the harmonic mean, then

$$\frac{1}{H} = \frac{1}{n}\left[\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots + \frac{1}{x_n}\right]$$

Example 10. *Calculate the harmonic mean of 4, 8, 16.*

Solution.
$$\frac{1}{H} = \frac{1}{3}\left[\frac{1}{4}+\frac{1}{8}+\frac{1}{16}\right] = \frac{7}{48}$$
$$H = \frac{48}{7} = 6.853$$ **Ans.**

10.10 AVERAGE DEVIATION OR MEAN DEVIATION

It is the mean of the absolute values of the deviations of a given set of numbers from their arithmetic mean.

If $x_1, x_2, x_3, ..., x_n$ be a set of numbers with frequencies $f_1, f_2, ...f_n$ respectively. Let x be the arithmetic mean of the numbers $x_1, x_2, ..., x_n$, then

$$\text{Mean deviation} = \frac{\Sigma f_i \mid x_i - \overline{x} \mid}{\Sigma f_i}$$

Example 11. *Find the mean deviation of the following frequency distribution.*

Class	0-6	6-12	12-18	18-24	24-30
Frequency	8	10	12	9	5

Solution. $a = 15$

Class	*Mid-value* x	*Frequency* f	$d = x - a$	fd	$\lvert x - 14\rvert$	$f\lvert x - 14\rvert$
0–6	3	8	– 12	– 96	11	88
6–12	9	10	– 6	– 60	5	50
12–18	15	12	0	0	1	12
18–24	21	9	+ 6	54	7	63
24–30	27	5	+ 12	60	13	65
Total		44		– 42		278

$$\text{Mean} = a + \frac{\Sigma fd}{\Sigma f} = 15 - \frac{42}{44} = 14 \text{ nearly}$$

$$\text{Average deviation} = \frac{\Sigma f\lvert x - \overline{x}\rvert}{\Sigma f} = \frac{278}{44} = 6.3$$ **Ans.**

10.11 STANDARD DEVIATION

Standard deviation is defined as the square root of the mean of the square of the deviation from the arithmetic mean.

$$S.D. = \sigma = \sqrt{\frac{\Sigma f (x - \overline{x})^2}{\Sigma f}}$$

Note. 1. The square of the standard deviation σ^2 is called variance.

2. σ^2 is called the second moment about the mean and is denoted by μ_2.

10.12 SHORTEST METHOD FOR CALCULATING STANDARD DEVIATION

We know that
$$\sigma^2 = \frac{1}{N}\,\Sigma f (x - \overline{x})^2 = \frac{1}{N}\,\Sigma f (x - a - \overline{x - a})^2$$
$$= \frac{1}{N}\,\Sigma f (d - \overline{x - a})^2 \qquad \text{where } x - a = d$$

$$= \frac{1}{N}\,\Sigma f d^2 - 2\,(\bar{x}-a)\,\frac{1}{N}\,\Sigma f d + (\bar{x}-a)^2\,\frac{1}{N}\Sigma f \quad \Sigma f = N$$

$$= \frac{1}{N}\,\Sigma f d^2 - 2\,(\bar{x}-a)\,\frac{1}{N}\,\Sigma f d + (\bar{x}-a)^2$$

$$\bar{x} = a + \frac{\Sigma f d}{N} \quad \text{or} \quad \bar{x} - a = \frac{\Sigma f d}{N}$$

$$\sigma^2 = \frac{1}{N}\,\Sigma f d^2 - 2\left(\frac{\Sigma f d}{N}\right)\left(\frac{1}{N}\,\Sigma f d\right) + \left(\frac{\Sigma f d}{N}\right)^2$$

$$= \frac{1}{N}\,\Sigma f d^2 - \left(\frac{\Sigma f d}{N}\right)^2$$

$$S.D. = \sigma = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}$$

Note. Coefficient of variation $= \dfrac{\sigma}{\bar{x}} \times 100$

Example 12. *Calculate the mean and standard deviation for the following data :*

Size of item	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4

Solution. Assumed mean = 9 (*A.M.I.E., Winter 2001*)

x	f	$d = x - a$	$f\,.\,d$	$f\,.\,d^2$
6	3	− 3	− 9	27
7	6	− 2	− 12	24
8	9	− 1	− 9	9
9	13	0	0	0
10	8	+ 1	8	8
11	5	+ 2	10	20
12	4	+ 3	12	36
	$\Sigma f = 48$		$\Sigma f d = 0$	$\Sigma f d^2 = 124$

$$\text{Mean} = a + \frac{\Sigma f d}{\Sigma f} = 9 + 0 = 9$$

$$S.D. = \sqrt{\frac{\Sigma f\,(x-\bar{x})^2}{\Sigma f}}$$

$$= \sqrt{\frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2} = \sqrt{\frac{124}{48}} = 1.6 \qquad \textbf{Ans.}$$

Example 13. *From the following frequency distribution, compute the standard deviation of 100 students :*

Mass in kg	60–62	63–65	66–68	69–71	72–74
Number of students	5	18	42	27	8

Solution. Assumed mean = 67

Mass in kg	*Number of students f*	*x*	*d = x – 67*	*f . d*	*f . d²*
60-62	5	61	– 6	– 30	180
63-65	18	64	– 3	– 54	162
66-68	42	67	0	0	0
69-71	27	70	3	81	243
72-74	8	73	6	48	288
	$\Sigma f = 100$			$\Sigma f d = 45$	$\Sigma f d^2 = 873$

$$S.D. = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} = \sqrt{\frac{873}{100} - \left(\frac{45}{100}\right)^2}$$

$$= \sqrt{8.73 - 0.2025} = \sqrt{8.5275} = 2.9202 \qquad \textbf{Ans.}$$

Example 14. *Compute the standard deviation for the following frequency distribution:*

Class interval	0–4	4–8	8–12	12–16
Frequency	4	8	2	1

Solution. Assumed mean = 6

Class interval	*f*	*x*	*d = x – 6*	*f d*	*f d²*
0-4	4	2	– 4	– 16	64
4-8	8	6	0	0	0
8-12	2	10	+ 4	8	32
12-16	1	14	+ 8	8	64
	$\Sigma f = 15$			$\Sigma f d = 0$	$\Sigma f d^2 = 160$

$$S.D. = \sqrt{\frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2} = \sqrt{\frac{160}{15} - 0} = 3.266 \qquad \textbf{Ans.}$$

10.13 MOMENTS

The rth moment of a variable x about the mean $\overline{x}$ is usually denoted by μ_r is given by

$$\mu_r = \frac{1}{N} \Sigma f_i (x_i - \overline{x})^r, \qquad \Sigma f_i = N$$

The rth moment of a variable x about any point a is defined by

$$\mu'_r = \frac{1}{N} f_i (x_i - a)^r$$

In particular

$$\mu_0 = \frac{1}{N} \Sigma f_i (x - \overline{x})^\circ = \frac{1}{N} \Sigma f_i = \frac{N}{N} = 1$$

$$\mu'_0 = \frac{1}{N} \Sigma f_i (x - a)^\circ = \frac{1}{N} \Sigma f_i = \frac{N}{N} = 1$$

$$\mu_1 = \frac{1}{N}\Sigma f_i (x - \overline{x}) = 0, \quad \mu'_1 = \frac{1}{N}\Sigma f_i (x - a) = \overline{x} - a$$

$$\mu_2 = \frac{1}{N}\Sigma f_i (x - \overline{x})^2 = \sigma^2 .$$

Relation between moments about mean and moment about any point.

$$\mu_r = \frac{1}{N}\Sigma f_i (x - \overline{x})^r = \frac{1}{N}\Sigma f_i [(x - a) - (\overline{x} - a)]^r$$

$$= \frac{1}{N}\Sigma f_i (X_i - d)^r \quad \text{where} \quad X_i = x - a \text{ and } d = \overline{x} - a$$

$$= \frac{1}{N}[\, \Sigma f_i X_i^r - {}^rC_1 (\Sigma f_i X_i^{r-1})\, d + {}^rC_2 (\Sigma f_i x_i^{r-2})\, d^2$$

$$- {}^rC_3 (\Sigma f_i x_i^{r-3})\, d^3 + \ldots\ldots]$$

$$= \mu'_r - {}^rC_1\, d\, \mu'_{r-1} + {}^rC_2\, d^2\, \mu'_{r-2} - {}^rC_3\, d^3\, \mu'_{r-3} + \ldots\ldots$$

In particular

$$\mu_2 = \mu'_2 - \mu'^2_1$$

$$\mu_3 = \mu'_3 - 3\,\mu'_2\,\mu'_1 + 2\,\mu'^3_1$$

$$\mu_4 = \mu'_4 - 4\,\mu'_3\,\mu'_1 + 6\,\mu'_2\,\mu'^2_1 - 3\,\mu^4_1$$

Note. 1. The sum of the coefficients of the various terms on the right-hand side is zero.

2. The dimension of each term on right-hand side is the same as that of terms on the left.

10. 14 MOMENT GENERATING FUNCTION

The moment generating function of the variate x about x = a is defined as the expected value of $e^{t(x-a)}$ and is denoted $M_a(t)$.

$$M_a(t) = \Sigma P_i\, e^{t(x_i - a)}$$

$$= \Sigma p_i + t\,\Sigma p_i (x_i - a) + \frac{t^2}{2\,!}\Sigma p_i (x_i - a)^2 + \ldots \frac{t^r}{r\,!}\Sigma p_i (x_i - a)^r + \ldots$$

$$= 1 + t\,\mu_1' + \frac{t^2}{2\,!}\mu_2' + \ldots. + \frac{t^r}{r\,!}\mu_r' + \ldots.$$

where μ_r' is the moment of order r about a

Hence $\mu_r' = \text{coefficient of } \frac{t^r}{r'!}$ or $\mu_r' = \left[\frac{d^r}{dt^r} M_a(t)\right]_{t=0}$

again $M_a(t)$

$$= \Sigma p_i\, e^{t(x_i - a)}$$

$$= e^{-at}\,\Sigma p_i\, e^{txi}$$

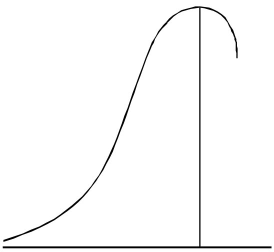
$$= e^{-at}\, M_0(t)$$

Thus the moment generating function about the point $a = e^{-at}$ moment generating function about the origin

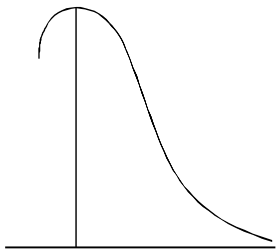
10.15 (1) SKEWNESS

Skewness denotes the opposite of symmetry. It is lack of symmetry. In a symmetrical series, the mode, the median, and the arithmetic average are identical.

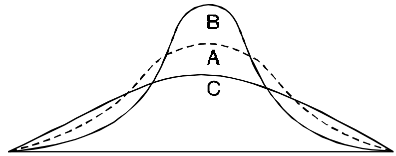
$$\text{Coefficient of skewness} = \frac{\text{Mean} - \text{Mode}}{\text{standard deviation}}$$



Negetive skewness



Positive skewness



A Mesokurtic B Leptokurtic C Platykurtic

(2) KURTOSIS. It measures the degree of peakedness of a distribution and is given by

Measure of kurtosis

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \quad \mu_2 = \frac{\Sigma (x-\bar{x})^2}{N}, \mu_4 = \frac{\Sigma (x-\bar{x})^4}{N}$$

If $\beta_2 = 3$, the curve is normal or mesokurtic.

If $\beta_2 > 3$, the curve is peaked or leptokurtic.

If $\beta_2 < 3$, the curve is flat topped or platykurtic.

Exercise 10.1

1. Marks obtained by 9 students in statistics are given below :

52, 57, 40, 70, 43, 40, 65, 35, 48.

Calculate the arithmetic mean. **Ans.** 50.

2. Calculate the mean of the following :

Height in cm	65	66	67	68	69	70	71	72	73
Number of plants	1	4	5	7	11	10	6	4	2

Ans. 69.18.

3. Find the mean for the following distribution :

Marks	*No. of students*	*Marks*	*No. of students*
0–10	3	50–60	15
10–20	5	60–70	12
20–30	7	70–80	6
30–40	10	80–90	2
40–50	12	90–100	8

Ans. 51.75

4. Determine the mode from the following figures :

25, 15, 23, 40, 27, 25, 23, 25, 20. **Ans.** 25.

5. Find the median of the following :

20, 18, 22, 27, 25, 12, 15. **Ans.** 20.

6. The Mean of 200 items was 50. Later on it was discovered that two items were misread as 92 and 8 instead of 192 and 88. Find the corrected mean. **Ans.** 53.6

7. Calculate the mean and standard deviation of the following frequency distribution.

Weekly wages in Rs.	*No. of persons*	*Weekly wages in Rs.*	*No of persons*
4.5–12.5	4	44.5–52.5	3
12.5–20.5	24	52.5–60.5	5
20.5–28.5	21	60.5–68.5	8
28.5–36.5	18	68.5–76.5	2
36.5–44.5	5		

Ans. Mean = 31.34, S.D. = 16.67

8. Compute the standard deviation from the following distribution of marks obtained by 90 students:

Marks	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–99
No. of students	5	12	15	20	18	10	6	4

Ans. 17.65.

9. The following table shows the Marks obtained by 100 candidates in an examination. Calculate the mean, median and standard deviation.

Marks	1–10	11–20	21–30	31–40	41–50	51–60
No. of candidates	3	16	26	31	16	8

Ans. Mean = 32, S.D. = 12.36, Median = 32.11

10. Fill in the blanks :

(*a*) Average value may be termed as measure of .... (*b*) $\beta_2 =$ ....

(*c*) $\mu_2 =$ .... (*d*) The curve is normal if $\beta_2 =$ ....

(*e*) The value of $\Sigma f(x - \bar{x}) =$ ....

(*f*) The measure of central item is called as ....

(*g*) The size of the variable which occurs most frequently is known as ....

(*h*) Coefficient of skewness = ...

(*i*) The ratio of the standard deviation to the mean is known as ....

(*j*) The square of standard deviation is known as the ....

Ans. (*a*) Central tendency, (*b*) $\frac{\mu_4}{\mu_2^2}$, (*c*) $\frac{\Sigma (x - \bar{x})^2}{N}$, (*d*) 3, (*e*) 0, (*f*) median,

(*g*) mode, (*h*) $\frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}}$, (*i*) Coefficient of standard deviation, (*j*) Variance

11. The expected value of a random variable X is 2 and its variance is 1, then variance of $3X + 4$ is

(*a*) 9 (*b*) 7 (*c*) 3 (*d*) 13 (*A.M.I.E.T.E., Dec. 2004*) **Ans.** (*a*)

12. The expected value of a random variable X is 3 and its variance is 2. Then the variance of $2X + 5$ is

(*a*) 8 (*b*) 9 (*c*) 10 (*d*) 11 (*A.M.I.E.T.E., June 2006*) **Ans.** (*a*)

10.16 CORRELATION

Whenever two variables x and y are so related that an increase in the one is accompanied by an increase or decrease in the other, then the variables are said to be correlated.

For example, the yield of crop varies with the amount of rainfall.

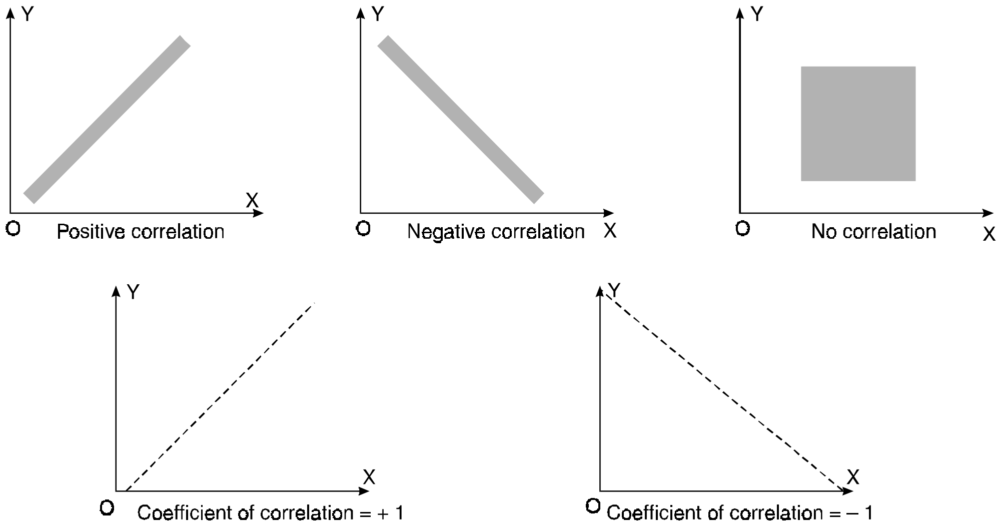
If an increase in one variable corresponds to an increase in the other, the correlation is said to be positive. If increase in one corresponds to the decrease in the other the correlation is said to be negative. If there is no relationship between the two variables, they are said to be independent.

Perfect Correlation: If two variables vary in such a way that their ratio is always constant, then the correlation is said to be prefect.

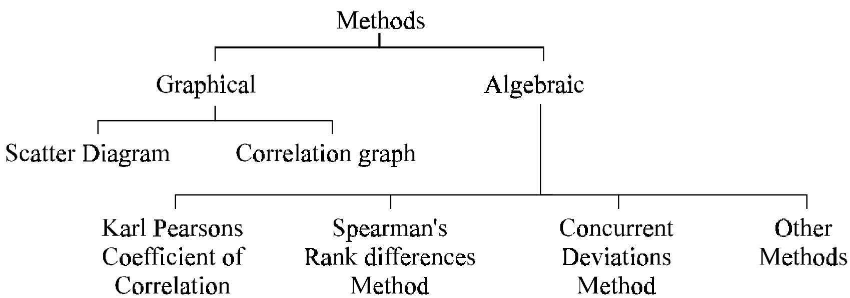
10.17 SCATTER OR DOT-DIAGRAM

When we plot the corresponding values of two variables, taking one on x-axis and the other along y-axis, it shows a collection of dots.

This collection of dots is called a dot diagram or a scatter diagram



Methods of Determining Simple Correlation



10.18 KARL PEARSON'S COEFFICIENT OF CORRELATION

r between two variables x and y is defined by the relation

$$r = \frac{\Sigma XY}{\sqrt{(\Sigma X^2)(\Sigma Y^2)}} = \frac{P}{\sigma_x \sigma_y} = \frac{Covariance\,(x, y)}{\sqrt{variance\ x}\,\sqrt{variance\ y}},$$

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

i.e. X, Y are the deviations measured from their respective means,

$$P = \left(\frac{\Sigma XY}{n}\right) = \text{co-variance.}$$

and $\sigma_x,\ \sigma_y$ being the standard deviations of these series.

Example 15. *Ten students got the following percentage of marks in Economics and Statistics.*

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

Calculate the coefficient of correlation.

Solution. Let the marks of two subjects be denoted by x and y respectively.

Then the mean for x marks $= \frac{650}{10} = 65$ and the mean of y marks $= \frac{660}{10} = 66$

If X and Y are deviations of x's and y's from their respective means, then the data may be arranged in the following form :

x	y	$X = x - 65$	$Y = y - 66$	X^2	Y^2	XY
78	84	13	18	169	324	234
36	51	– 29	– 15	841	225	435
98	91	33	25	1089	625	825
25	60	– 40	– 6	1600	36	240
75	68	10	2	100	4	20
82	62	17	– 4	289	16	– 68
90	86	25	20	625	400	500
62	58	– 3	– 8	9	64	24
65	53	0	– 13	0	169	0
39	47	– 26	– 19	676	361	494
650	660	0	0	5398	2224	2704

Here $\Sigma X^2 = 5398,\ \Sigma Y^2 = 2224,\ \Sigma XY = 2704$

$$\therefore \qquad r = \frac{\Sigma XY}{\sqrt{(\Sigma X^2)(\Sigma Y^2)}} = \frac{2704}{\sqrt{5398 \times 2224}}$$

$$= \frac{2704}{73.4 \times 47.1} = \frac{2704}{3457} = 0.78 \qquad \textbf{Ans.}$$

Example 16. *Find the coefficient of correlation between the age and the sum assured from the following table:*

	Sum assured (in Rs.)				
Age group	*10,000*	*20,000*	*30,000*	*40,000*	*50,000*
20–30	4	6	3	7	1
30–40	2	8	15	7	1
40–50	3	9	12	6	2
50–60	8	4	2	—	—

Solution. Let the sum assured denote by x and the age group by y.

$$x' = \frac{x - 30{,}000}{10{,}000},\ y' = \frac{y - 45}{10}$$

x			10,000		20,000		30,000		40,000		50,000					
y		y' \ x'	–2		–1		0		1		2		Σf (Rows)	$f \cdot y'$	$f y'^2$	$fx'y'$
			f	$fx'y'$	f	$fx'y'$	f	$fx'y'$	f	$fx'y'$	f	$fx'y'$				
20–30	25	–2	4	16	6	12	3	0	7	–14	1	–4	21	–42	84	+10
30–40	35	–1	2	4	8	8	15	0	7	–7	1	–2	33	–33	33	+3
40–50	45	0	3	0	9	0	12	0	6	0	2	0	32	0	0	0
50–60	55	1	8	–16	4	–4	2	0	–	0	–	0	14	14	14	–20
		Σf column	17		27		32		20		4		N = 100	$\Sigma f y'$ = –61	$\Sigma f y'^2$ = 131	$\Sigma fx'y'$ = –7
		fx'	–34		–27		0		20		8		$\Sigma fx'$ = –33			
		$f' x^2$	68		27		0		20		16		$\Sigma f x'^2$ = 131			
		$f x' y'$		4		16		0		–21		–6	$\Sigma fx'y'$ = –7			

$$r = \frac{N \Sigma fx'y' - \Sigma f \cdot x' \Sigma f \cdot y'}{\sqrt{N \Sigma f \cdot x'^2 - (\Sigma f \cdot x')^2}\ \sqrt{N \Sigma f \cdot x'^2 - (\Sigma f \cdot y')^2}}$$

$$= \frac{100\,(-7) - (-33)\,(-61)}{\sqrt{100\,(131) - (-33)^2}\ \sqrt{100\,(131) - (-61)^2}} = \frac{-700 - 2013}{\sqrt{13100 - 1089}\ \sqrt{13100 - 3721}}$$

$$= \frac{-2713}{\sqrt{12011}\ \sqrt{9379}} = \frac{-2713}{109.59 \times 96.85} = \frac{-2713}{10613 \cdot 7915} = -0.2556$$

Hence, the age and sum assured are negatively correlated, *i.e.*, as age goes up the sum assured comes down. **Ans.**

10.19 SHORT-CUT METHOD

$$r = \frac{\dfrac{\Sigma X' Y'}{N} - \left(\dfrac{\Sigma X'}{N}\right)\left(\dfrac{\Sigma Y'}{N}\right)}{\sqrt{\left\{\dfrac{\Sigma X'^2}{N} - \left(\dfrac{\Sigma X'}{N}\right)^2\right\}}\ \sqrt{\left\{\dfrac{\Sigma Y'^2}{N} - \left(\dfrac{\Sigma Y'}{N}\right)^2\right\}}}$$

where r is the coefficient of correlation.

X' = deviation from *assumed* mean of $x = x - a$

Y' = deviation from *assumed* mean of $y = y - b$

N = Total number of items.

Example 17. *Calculate the coefficient of correlation for the following table :*

x–age / marks	0–4	4–8	8–12	12–16
0–5	7	—	—	—
5–10	6	8	—	—
10–15	—	5	3	—
15–20	—	7	2	—
20-25	—	—	—	9

Solution. Replace the class-interval for x and y by their mid-points and then let

$$X' = \frac{x - 10}{4} \text{ and } Y' = \frac{y - 12.5}{5}$$

	x / y	X' / Y'	2		6		10		14		Σf (row)	fY'	fY'^2	$fX'Y'$
			–2		–1		0		1					
			f	$fX'Y'$	f	$fX'Y'$	f	$fX'Y'$	f	$fX'Y'$				
0–5	2.5	–2	7	28							7	– 14	– 28	28
5–10	7.5	–1	6	12	8	8					14	–14	– 14	20
10–15	12.5	0			5	0	3	0			8	0	0	0
15–20	17.5	1			7 –7		2	0			9	9	9	–7
20–25	22.5	2							9	18	9	18	36	18
	Σf		13		20		5		9		47	$\Sigma fY' = -1$	$\Sigma fY' = 87$	$\Sigma fX'Y' = 59$
	fX'		–26		– 20		0		9		$\Sigma fX' =$ –37			
	fX'^2		52		20		0		9		$\Sigma fX'^2 =$ 81			
	$fX'Y'$			40		1		0		18	$\Sigma fX'Y' =$ 59			

Here, $\quad \Sigma f X' = -37,\ \Sigma f X'^2 = 81,\ \Sigma f Y' = -1,\ \Sigma f Y'^2 = 87, \Sigma fX'Y' = 59$

$$r = \frac{\dfrac{\Sigma f X'Y'}{N} - \left(\dfrac{\Sigma fX'}{N}\right)\left(\dfrac{\Sigma fY'}{N}\right)}{\sqrt{\dfrac{\Sigma fX'^2}{N} - \left(\dfrac{\Sigma fX'}{N}\right)^2}\sqrt{\dfrac{\Sigma fY'^2}{N} - \left(\dfrac{\Sigma fY'}{N}\right)^2}}$$

$$= \frac{\dfrac{59}{47} - \left(\dfrac{-37}{47}\right)\left(\dfrac{-1}{47}\right)}{\sqrt{\left\{\dfrac{81}{47} - \left(\dfrac{-37}{47}\right)^2\right\}}\sqrt{\left\{\dfrac{87}{47} - \left(\dfrac{-1}{47}\right)^2\right\}}} = \frac{1.255 - 0.017}{\sqrt{1.723 - 0.620}\sqrt{1.851 - 0.0005}}$$

$$= \frac{1.238}{\sqrt{1.103}\sqrt{1.8505}} = \frac{1.238}{1.05 \times 1.36} = \frac{1.238}{1.428} = 0.87 \quad \textbf{Ans.}$$

10. 20 SPEARMAN'S RANK CORRELATION

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

Solution. Let (x_1, y_1), (x_2, y_2) … (x_n, y_n) be the ranks of n individuals corresponding to two characteristics.

Assuming nor two individuals are equal in either classification, each individual takes the values 1, 2, 3, ... n and hence their arithmetic means are, each

$$= \frac{\Sigma\,n}{n} = \frac{1}{n}\,\frac{n\,(n+1)}{2} = \frac{n+1}{2}$$

Let x_1, x_2, x_3, ... x_n be the values of variable X and y_1, y_2, y_3, ... y_n those of Y.

Then $$d = X - Y = \left(x - \frac{n+1}{2}\right) - \left(y - \frac{n+1}{2}\right) = x - y$$

where X and Y are deviations from the mean.

$$\Sigma\,X^2 = \Sigma\left(x - \frac{n+1}{2}\right)^2 = \Sigma\,x^2 - (n+1)\,\Sigma\,x + \Sigma\left(\frac{n+1}{2}\right)^2$$

$$= \frac{n\,(n+1)\,(2n+1)}{6} - \frac{(n+1)\,n\,(n+1)}{2} + n\left(\frac{n+1}{2}\right)^2$$

$$= \frac{n\,(n^2 - 1)}{12}$$

Clearly, $$\Sigma\,X = \Sigma\,Y \quad \text{and} \quad \Sigma\,X^2 = \Sigma\,Y^2$$

$\therefore$ $$\Sigma\,Y^2 = \frac{n\,(n^2-1)}{12}$$

Hence $$\Sigma\,d^2 = \Sigma\,(x - y)^2 = \Sigma\,x^2 + \Sigma\,y^2 - 2\,\Sigma\,xy$$

$\therefore$ $$\Sigma\,XY = \frac{1}{2}\left[\frac{n\,(n^2-1)}{6} - \Sigma\,d^2\right]$$

$$= \frac{1}{12}\,n\,(n^2 - 1) - \frac{1}{2}\,\Sigma\,d^2$$

Putting these values in $$r = \frac{\Sigma\,XY}{\sqrt{\Sigma\,X^2}\,\sqrt{\Sigma\,Y^2}}$$

$$= \frac{\frac{1}{12}\,n\,(n^2-1) - \frac{1}{2}\,\Sigma\,d^2}{\frac{n\,(n^2-1)}{12}}$$

$$= 1 - \frac{6\,\Sigma\,d^2}{n\,(n^2-1)}$$ **Ans.**

10.21 SPEARMAN'S RANK CORRELATION COEFFICIENT

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

where r denotes rank coefficient of correlation and d refers to the difference of ranks between paired items in two series.

Example 18. *Compute Spearman's rank correlation coefficient r for the following data:*

Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Rank in statistics	9	10	6	5	7	2	4	8	1	3
Rank in income	1	2	3	4	5	6	7	8	9	10

Solution.

Person	*Rank in statistics*	*Rank in income*	$d = R_1 - R_2$	d^2
A	9	1	8	64
B	10	2	8	64
C	6	3	3	9
D	5	4	1	1
E	7	5	2	4
F	2	6	– 4	16
G	4	7	– 3	9
H	8	8	0	0
I	1	9	– 8	64
J	3	10	– 7	49
				$\Sigma d^2 = 280$

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

$$r = 1 - \frac{6 \times 280}{10\,(100 - 1)} = 1 - 1.697 = -0.697 \qquad \textbf{Ans.}$$

Example 19. *Establish the formula*

$$\sigma^2_{x-y} = \sigma^2_x + \sigma^2_y - 2r\,\sigma_x\,\sigma_y$$

where r is the correlation coefficient between x and y.

Solution. We know that $\sigma^2_x = \dfrac{\Sigma\,(x - \overline{x})^2}{n}$

$$\therefore \qquad \sigma^2_{x-y} = \frac{\Sigma\,[(x - y) - (\overline{x - y})]^2}{n}$$

$$\overline{x - y} = \text{mean of } (x - y) \text{ series.} = \text{mean of } x - \text{mean of } y = \overline{x} - \overline{y}$$

$$\sigma^2_{x-y} = \frac{\Sigma\,[\,(x - y) - (\overline{x} - \overline{y})\,]^2}{n} = \frac{\Sigma[\,(x - \overline{x}) - (y - \overline{y})\,]^2}{n}$$

$$= \frac{\Sigma\,[\,(x - \overline{x})^2 + (y - \overline{y})^2 - 2\,(x - \overline{x})\,(y - \overline{y})\,]}{n}$$

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

$$= \sigma^2_x + \sigma^2_y - \frac{2\,\Sigma\,(x - \overline{x})\,(y - \overline{y})}{n} \qquad ...(1)$$

We know that $r = \dfrac{\Sigma\,(x - \overline{x})\,(y - \overline{y})}{n\,\sigma_x\,\sigma_y}$ or $\dfrac{\Sigma\,(x - \overline{x})\,(y - \overline{y})}{n} = r\,\sigma_x\,\sigma_y$

Putting this value in (1) we get,

$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2r\,\sigma_x\,\sigma_y \qquad \textbf{Proved}$$

Example 20. *If X and Y are uncorrelated random variables, find the coefficient of correlation between X + Y and X − Y.*

Solution.

Let $u = X + Y$ and $v = X - Y$

Then $r = \dfrac{\Sigma (u-\bar{u})(v-\bar{v})}{n\,\sigma_u\,\sigma_v}$

Now $u = X + Y, \quad \bar{u} = \bar{X} + \bar{Y}$

Similarly $\bar{v} = \bar{X} - \bar{Y}$

Now $\Sigma (u-\bar{u})(v-\bar{v})$

$$= \Sigma (X - \bar{X} + Y - \bar{Y})\,[(X-\bar{X}) - (Y-\bar{Y})]$$
$$= \Sigma (x+y)(x-y)$$
$$= \Sigma x^2 - \Sigma y^2$$
$$= n\,\sigma_x^2 - n\,\sigma_y^2$$

Also $\sigma_u^2 = \dfrac{\Sigma (u-\bar{u})^2}{n} = \dfrac{1}{n}\Sigma [(X-\bar{X}) + (Y-\bar{Y})]^2$

$$= \frac{1}{n}\Sigma (x+y)^2$$
$$= \frac{1}{n}(\Sigma x^2 + \Sigma y^2 + 2\,\Sigma xy)$$
$$= \sigma_x^2 + \sigma_y^2 \quad \text{(As } X \text{ and } Y \text{ are not correlated, we have } \Sigma xy = 0)$$

Similarly $\sigma_v^2 = \sigma_x^2 + \sigma_y^2$

$\therefore$ $r = \dfrac{\Sigma (u-\bar{u})(v-\bar{v})}{n\,\sigma_u\,\sigma_v}$

$$= \frac{n(\sigma_x^2 - \sigma_y^2)}{\sqrt{n(\sigma_x^2+\sigma_y^2)}\,\sqrt{n(\sigma_x^2+\sigma_y^2)}}$$
$$= \frac{\sigma_x^2 - \sigma_y^2}{\sigma_x^2 + \sigma_y^2} \qquad \textbf{Ans.}$$

10.22 REGRESSION

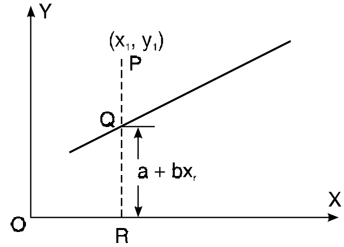
If the scatter diagram indicates some relationship between two variables x and y, then the dots of the scatter diagram will be concentrated round a curve. This curve is called the *curve of regression.*

Regression analysis is the method used for estimating the unknown values of one variable corresponding to the known value of another variable.

10.23 LINE OF REGRESSION

When the curve is a straight line, it is called a line of regression. A line of regression is the straight line which gives the best fit in the least square sense to the given frequency.

10.24 EQUATIONS TO THE LINES OF REGRESSION



Let $$y = a + bx \quad ...(1)$$

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

Let (x_r, y_r) be any point of dot.

From the figure

$$PR = y_r$$

$$QR = a + bx_r$$

$$PQ = PR - QR = y_r - a - bx_r$$

Let S be the sum of the squares of such distances, then

$$S = \Sigma (y - a - bx)^2$$

According to the principle of least squares, we have to choose a and b so that S is minimum. The method of least square gives the condition for minimum value of S.

$$\frac{\partial S}{\partial a} = -2\,\Sigma (y - a - bx), \quad \frac{\partial S}{\partial b} = -\;2\,\Sigma (y - a - bx)\,x$$

$$\frac{\partial S}{\partial a} = 0, \; \frac{\partial S}{\partial b} = 0, \quad \text{for } S \text{ minimum}$$

i.e. $$\Sigma (y - a - bx) = 0 \quad \text{or} \quad \Sigma y - na - b\,\Sigma x = 0$$

or $$\Sigma y = na + b\,\Sigma x \quad ...(2)$$

and $$\Sigma (xy - ax - bx^2) = 0 \quad or \quad \Sigma xy - a\,\Sigma x - b\,\Sigma x^2 = 0$$

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 \quad ...(3)$$

Dividing (2) by n we get

$$\frac{\Sigma y}{n} = a + b\,\frac{\Sigma x}{n} \qquad \left(\bar{y} = \frac{\Sigma y}{n},\; \bar{x} = \frac{\Sigma x}{n}\right)$$

$$\bar{y} = a + b\,\bar{x}$$

where $\bar{x}$ and $\bar{y}$ are the means of x series and y series.

This shows that $(\bar{x}, \bar{y})$ lie on the line of regression (1), shifting the origin to $(\bar{x}, \bar{y})$, the equation (3) becomes

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

But $$\Sigma (x - \bar{x}) = 0 \quad i.e. \quad \Sigma (x - \bar{x})(y - \bar{y}) = b\,\Sigma (x - \bar{x})^2$$

or $$b = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2} = \frac{\Sigma XY}{\Sigma X^2} \quad ...(4)$$

We know $$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\,\sqrt{\Sigma Y^2}} = \frac{\Sigma XY}{n\sqrt{\dfrac{\Sigma X^2}{n}}\,\sqrt{\dfrac{\Sigma Y^2}{n}}} = \frac{\Sigma XY}{n\,\sigma_x\,\sigma_y}$$

or $$\Sigma XY = nr\,\sigma_x\,\sigma_y$$

Putting the value of ΣXY in (4) we get

$$b = \frac{nr\,\sigma_x\sigma_y}{\Sigma X^2} = \frac{r\,\sigma_x\,\sigma_y}{\dfrac{\Sigma X^2}{n}} = \frac{r\,\sigma_x\,\sigma_y}{\sigma_x^2} = \frac{r\,\sigma_y}{\sigma_x}$$

i.e. slope of the line of regression = $b = r\dfrac{\sigma_y}{\sigma_x}$

The line of regression passes through $(\bar{x}, \bar{y})$.

Hence the equation to the line of regression is

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

Similarly the regression line of x on y is

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

Note. $b_{yx} = r\,\dfrac{\sigma_y}{\sigma_x}$ and $b_{xy} = r\,\dfrac{\sigma_x}{\sigma_y}$ are known as the coefficients of regression.

$$b_{yx} \,.\, b_{xy} = \left(r\,\frac{\sigma_y}{\sigma_x}\right)\left(r\,\frac{\sigma_x}{\sigma_y}\right) = r^2$$

Example 21. *If θ be 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 where $r = 0$ and $r = \pm 1$. (*A.M.I.E., Winter 2001*)

Solution. Lines of regression are

$$y - \bar{y} = r\,\frac{\sigma_y}{\sigma_x}\,(x - \bar{x}) \qquad ...(1) \qquad \therefore\; m_1 = r\,\frac{\sigma_y}{\sigma_x}$$

and

$$x - \bar{x} = r\,\frac{\sigma_x}{\sigma_y}\,(y - \bar{y}) \qquad ...(2) \qquad \therefore\; m_2 = \frac{1}{r}\,\frac{\sigma_y}{\sigma_x}$$

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

$$= \frac{\dfrac{1}{r}\,\dfrac{\sigma_y}{\sigma_x} - r\,\dfrac{\sigma_y}{\sigma_x}}{1 + r\,\dfrac{\sigma_y}{\sigma_x} \times \dfrac{1}{r}\,\dfrac{\sigma_y}{\sigma_x}} = \frac{\left(\dfrac{1}{r} - r\right)\dfrac{\sigma_y}{\sigma_x}}{1 + \dfrac{\sigma_y^2}{\sigma_x^2}}$$

$$= \frac{1 - r^2}{r}\,.\,\frac{\left(\dfrac{\sigma_y}{\sigma_x}\right)\sigma_x^2}{\sigma_x^2 + \sigma_y^2} = \frac{1 - r^2}{r}\,\frac{\sigma_x\,\sigma_y}{\sigma_x^2 + \sigma_y^2} \qquad ...(3) \quad \textbf{Proved}$$

(*a*) If $r = 0$, then there is no relationship between the two variables and they are independent.

On putting the value of $r = 0$ in (3) we get $\tan\theta = \infty$, $\theta = \dfrac{\pi}{2}$. So the lines (1) and (2) are perpendicular. (*A.M.I.E., Summer 1998*)

(*b*) If $r = 1$ or -1

On putting these values of r in (3) we get, $\tan\theta = 0$ or $\theta = 0$

i.e. lines (1) and (2) coincide.

The correlation between the variables is perfect. **Ans.**

Example 22. *Find the correlation coefficient between x and y, when the lines of regression are:*

$$2x - 9y + 6 = 0, \qquad x - 2y + 1 = 0$$

Solution. Let the line of regression of x on y be $2x - 9y + 6 = 0$

Then, the line of regression of y on x is $x - 2y + 1 = 0$

$$\therefore \quad 2x - 9y + 6 = 0 \Rightarrow x = \frac{9}{2}y - 3 \Rightarrow b_{xy} = \frac{9}{2}$$

and
$$x - 2y + 1 = 0 \Rightarrow y = \frac{1}{2}x + \frac{1}{2} \Rightarrow b_{yx} = \frac{1}{2}$$

$$r = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{\frac{9}{2} \times \frac{1}{2}} = \frac{3}{2} > 1 \quad \text{which is not possible.}$$

So our choice of regression line is incorrect.

$\therefore$ The regression line of x on y is $x - 2y + 1 = 0$

And, the regression line of y on x is $2x - 9y + 6 = 0$

$$\therefore \quad x - 2y + 1 = 0 \Rightarrow x = 2y - 1 \Rightarrow b_{xy} = 2$$

And
$$2x - 9y + 6 = 0 \Rightarrow y = \frac{2}{9}x + \frac{2}{3} \Rightarrow b_{yx} = \frac{2}{9}$$

$$r = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{2 \times \frac{2}{9}} = \frac{2}{3}$$

Hence the correlation coefficient between x and y is $\frac{2}{3}$.

Example 23. *The following regression equations were obtained from a correlation table:*

$$y = 0.516\,x + 33.73, \quad x = 0.512\,y + 32.52$$

Find the value of (a) the correlation coefficient, (b) the mean of x's and (c) the mean of y's.

Solution.
$$y = 0.516\,x + 33.73 \qquad \text{...(1)}$$
$$x = 0.512\,y + 32.52 \qquad \text{...(2)}$$

(*a*) From (1),
$$r\frac{\sigma_y}{\sigma_x} = 0.516 \qquad \text{...(3)}$$

From (2),
$$r\frac{\sigma_x}{\sigma_y} = 0.512 \qquad \text{...(4)}$$

From (3) and (4)

$$\left(r\frac{\sigma_y}{\sigma_x}\right)\left(r\frac{\sigma_x}{\sigma_y}\right) = (0.516)\,(0.512)$$

$$r^2 = 0.516 \times 0.512 \quad \text{or} \quad r = 0.514$$

Coefficient of correlation $= 0.514$. **Ans.**

(*b*) (1) and (2) pass through the point $(\overline{x}, \overline{y})$.

$$\therefore \quad \overline{y} = 0.516\,\overline{x} + 33.73 \qquad \text{...(5)}$$
$$\overline{x} = 0.512\,\overline{y} + 32.52 \qquad \text{...(6)}$$

On solving (5) and (6), we get

$$\bar{x} = 67.6,\ \bar{y} = 68.61$$ **Ans.**

Example 24. *The two regression equations of the variables x and y are*

$$x = 19.13 - 0.87\,y \text{ and } y = 11.64 - 0.50\,x.$$

Find (i) Mean of x's; (ii) Mean of y's ; (iii) The correlation coefficient between x and y.

(*A.M.I.E., Summer 1997, 1996*)

Solution.

$$x = 19.13 - 0.87\,y \qquad ...(1)$$

$$y = 11.64 - 0.50\,x \qquad ...(2)$$

As (1) and (2) pass through $(\bar{x}, \bar{y})$:

$$\bar{x} = 19.13 - 0.87\,\bar{y} \qquad ...(3)$$

$$\bar{y} = 11.64 - 0.50\,\bar{x} \qquad ...(4)$$

On solving (3) and (4) we get

$$\bar{x} = 15.935,\ \bar{y} = 3.67$$

From (1) $$r\,\frac{\sigma_x}{\sigma_y} = -0.87 \qquad ...(5)$$

From (2) $$r\,\frac{\sigma_y}{\sigma_x} = -0.50 \qquad ...(6)$$

As σ_x and σ_y are always positive, so r is negative.

Multiplying (5) and (6) we get

$$r\,\frac{\sigma_x}{\sigma_y} \cdot r\,\frac{\sigma_y}{\sigma_x} = -0.87 \times (-0.50)$$

$$r^2 = 0.435 \quad \text{or} \quad r = -0.66$$ **Ans.**

Example 25. *The regression equations calculated from a given set of observations for two random variables are*

$$x = -0.4y + 6.4 \quad \textit{and} \quad y = -0.6x + 4.6$$

Calculate $\bar{x}$, $\bar{y}$ *and r.* (*A.M.I.E., Winter 1997*)

Solution. The regression equations are

$$x = -0.4y + 6.4 \qquad ...\ (1)$$

$$y = -0.6x + 4.6 \qquad ...\ (2)$$

From (1) coefficient of regression of x on $y = r\,\frac{\sigma_x}{\sigma_y} = -0.4$ ... (3)

From (2) coefficient of regression of y on $x = r\,\frac{\sigma_y}{\sigma_x} = -0.6$ ... (4)

From (3) and (4)

$$\left(r\,\frac{\sigma_x}{\sigma_y}\right)\left(r\,\frac{\sigma_y}{\sigma_x}\right) = (-0.4)\,(-0.6)$$

or $$r^2 = 0.24$$

$$r = \pm 0.49$$

In (3) and (4), σ_x and σ_y are (always) positive so r is negative

$$r = -0.49$$

To find $\bar{x}$ and $\bar{y}$, we solve the equations (1) and (2) simultaneously. Their point of intersection is $(\bar{x}, \bar{y})$.

$$\bar{x} = 6, \quad \bar{y} = 1$$ **Ans.**

Example 26. *Show that the geometric mean of the coefficients of regression is the coefficient of correlation.*

Solution. The coefficients of regressions are $r\,\dfrac{\sigma_y}{\sigma_x}$ and $r\,\dfrac{\sigma_x}{\sigma_y}$

i.e. $$G.M. = \sqrt{r\,\frac{\sigma_y}{\sigma_x} \cdot r\,\frac{\sigma_x}{\sigma_y}} = r$$

$=$ coefficient of correlation. **Proved.**

Example 27. *Prove that arithmetic mean of the coefficients of regression is greater than the coefficient of correlation.* (*A.M.I.E., Summer 2000*)

Solution. Coefficients of regression are $r\,\dfrac{\sigma_y}{\sigma_x}$, $r\,\dfrac{\sigma_x}{\sigma_y}$

We have to prove that $A.M. > r$

or $$\frac{1}{2}\left[r\,\frac{\sigma_y}{\sigma_x} + r\,\frac{\sigma_x}{\sigma_y}\right] > r \quad \text{or} \quad \frac{1}{2}\left[\frac{\sigma_y}{\sigma_x} + \frac{\sigma_x}{\sigma_y}\right] > 1$$

or $$\frac{\sigma_y}{\sigma_x} + \frac{\sigma_x}{\sigma_y} - 2 > 0 \quad \text{or} \quad \frac{1}{\sigma_x\,\sigma_y}\,[\sigma_x^2 + \sigma_y^2 - 2\,\sigma_x\,\sigma_y] > 0$$

or $$\frac{1}{\sigma_x\,\sigma_y}\,[\sigma_x - \sigma_y]^2 > 0 \quad \text{which is true.}$$ **Proved**

Example 28. *Find the regression line of y on x for the following data :*

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

Estimate the value of y, when x = 10.

Solution.

S. No.	*x*	*y*	*xy*	x^2
1	1	1	1	1
2	3	2	6	9
3	4	4	16	16
4	6	4	24	36
5	8	5	40	64
6	9	7	63	81
7	11	8	88	121
8	14	9	126	196
Total	56	40	364	524

Let $y = a + bx$ be the line of regression of y on x, where a and b are given by the following equations :

$$\Sigma y = na + b\Sigma x \quad \text{or} \quad 40 = 8a + 56b \qquad ...(1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{or} \quad 364 = 56a + 524b \qquad ...(2)$$

On solving (1) and (2) we get,

$$a = \frac{6}{11} \quad \text{and} \quad b = \frac{7}{11}$$

The equation of the required line is

$$y = \frac{6}{11} + \frac{7}{11}x \quad \text{or} \quad 7x - 11y + 6 = 0 \qquad \textbf{Ans.}$$

If $$x = 10,\ y = \frac{6}{11} + \frac{7}{11}(10) = \frac{76}{11} = 6\frac{10}{11} \qquad \textbf{Ans.}$$

Example 29. *In a study between the amount of rainfall and the quantity of air pollution removed the following data were collected.*

Daily Rainfall in 0.01 cm	*4.3*	*4.5*	*5.9*	*5.6*	*6.1*	*5.2*	*3.8*	*2.1*
Pollution Removed (mg/m^3)	*12.6*	*12.1*	*11.6*	*11.8*	*11.4*	*11.8*	*13.2*	*14.1*

Find the regression line of y on x. (*A.M.I.E., Summer 2000*)

Solution.

S.N.	*x (metre)*	*y*	*xy*	*x^2*
1	4.3	12.6	54.18	18.49
2	4.5	12.1	54.45	20.25
3	5.9	11.6	68.44	34.81
4	5.6	11.8	66.08	31.36
5	6.1	11.4	69.54	37.21
6	5.2	11.8	61.36	27.04
7	3.8	13.2	50.16	14.44
8	2.1	14.1	29.61	4.41
	37.5	98.6	453.82	188.01

Let $y = a + bx$ be the equation of the line of regression of y on x, where a and b are given by the following equations.

$$\Sigma y = na + b\Sigma x \quad \text{or} \quad 98.6 = 8a + 37.5b \qquad ...\ (1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{or} \quad 453.82 = 37.5a + 188.01b \qquad ...\ (2)$$

On solving (1) and (2), we get $a = 15.49$ and $b = -\ 0.675$.

The equation of the line of regression is $y = 15.49 - 0.675x$ **Ans.**

Example 30. *The following data regarding the heights (y) and the weights (x) of 100 college students are given :*

$$\Sigma x = 15000 \qquad \Sigma x^2 = 2272500$$

$$\Sigma y = 6800 \qquad \Sigma y^2 = 46.3025$$

$$\Sigma xy = 1022250$$

Find the correlation coefficient between height and weight and state the equation of regression of height on weight.

Solution.
$$\bar{x} = \frac{\Sigma x}{n} = \frac{15000}{100} = 150, \quad \bar{y} = \frac{\Sigma y}{n} = \frac{6800}{100} = 68$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{2272500}{100} - \left(\frac{15000}{100}\right)^2}$$

$$\sigma_x = \sqrt{22725 - 22500} = \sqrt{225} = 15$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2} = \sqrt{\frac{463025}{100} - \left(\frac{6800}{100}\right)^2}$$

$$= \sqrt{4630.25 - 4624} = \sqrt{6.25} = 2.5$$

$$r = \frac{\frac{\Sigma xy}{n} - (\bar{x})(\bar{y})}{(\sigma_x)(\sigma_y)} = \frac{\frac{1022250}{100} - (150)(68)}{15 \times 2.5}$$

$$= \frac{10222.5 - 10200}{15 \times 2.5} = \frac{22.5}{15 \times 2.5} = \frac{1.5}{2.5} = 0.6$$

Regression equation of y on x we have

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x}), \quad y - 68 = 0.6\left(\frac{2.5}{15}\right)(x - 150)$$

$$y - 68 = \frac{1}{10}(x - 150) \quad \text{or} \quad 10y - 680 = x - 150$$

$$10y = x + 530$$ **Ans.**

10.25 ERROR OF PREDICTION

The deviation of the predicted value from the observed value is known as the standard error of prediction. It is given by

$$E_{yx} = \sqrt{\frac{\Sigma (y - y_r)^2}{n}}$$

where y is the actual value and y_r the predicted value.

Example 31. *Prove that*

(*i*) $E_{yx} = \sigma_y \cdot \sqrt{1 - r^2}$ (*ii*) $E_{xy} = \sigma_x \sqrt{1 - r^2}$

Solution. The equation of the line of regression of y on x is

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

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

So,
$$E_{yx} = \sqrt{\frac{\Sigma (y - y_r)^2}{n}} = \left[\frac{1}{n}\Sigma\left\{y - \bar{y} - r\frac{\sigma_y}{\sigma_x}(x - \bar{x})\right\}^2\right]^{1/2}$$

$$= \left[\frac{1}{n}\cdot\Sigma\left\{(y - \bar{y})^2 + \frac{r^2\sigma_y^2}{\sigma_x^2}(x - \bar{x})^2 - \frac{2r\sigma_y}{\sigma_x}(x - \bar{x})(y - \bar{y})\right\}\right]^{1/2}$$

$$= \left[\Sigma \frac{(y-\overline{y})^2}{n} + r^2 \frac{\sigma_y^2}{\sigma_x^2} \Sigma \frac{(x-\overline{x})^2}{n} - 2r \frac{\sigma_y}{\sigma_x} \Sigma \frac{(x-\overline{x})(y-\overline{y})}{n}\right]^{1/2}$$

$$= \left[\sigma_y^2 + r^2 \frac{\sigma_y^2}{\sigma_x^2} \cdot \sigma_x^2 - 2r \frac{\sigma_y}{\sigma_x} r \cdot \sigma_x \cdot \sigma_y\right]^{1/2}$$

$$= \left[\sigma_y^2 + r^2 \sigma_y^2 - 2r^2 \sigma_y^2\right]^{1/2} = \left[\sigma_y^2 - r^2 \sigma_y^2\right]^{1/2}$$

$$= \sigma_y \sqrt{1-r^2}$$ **Proved.**

(*ii*) Similarly (*ii*) may be proved.

Example 32. *Find the standard error of estimate of y on x for the data given below:*

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

Solution. The equation of the line of regression of y on x is

$$y = \frac{7}{11}x + \frac{6}{11}. \quad \text{So} \quad y_r = \frac{7x}{11} + \frac{6}{11}$$ (*See Example 28 on page 757*)

S. No.	x	y	y_r	$(y - y_r)$	$(y - y_r)^2$
1	1	1	$\frac{13}{11}$	$-\frac{2}{11}$	$\frac{4}{121}$
2	3	2	$\frac{27}{11}$	$-\frac{5}{11}$	$\frac{25}{121}$
3	4	4	$\frac{34}{11}$	$\frac{10}{11}$	$\frac{100}{121}$
4	6	4	$\frac{48}{11}$	$-\frac{4}{11}$	$\frac{16}{121}$
5	8	5	$\frac{62}{11}$	$-\frac{7}{11}$	$\frac{49}{121}$
6	9	7	$\frac{69}{11}$	$\frac{8}{11}$	$\frac{64}{121}$
7	11	8	$\frac{83}{11}$	$\frac{5}{11}$	$\frac{25}{121}$
8	14	9	$\frac{104}{11}$	$-\frac{5}{11}$	$\frac{25}{121}$
					$\Sigma(y-y_r)^2 = \frac{308}{121}$

$$E_{yx} = \sqrt{\frac{\Sigma (y-y_r)^2}{n}} = \sqrt{\frac{308}{121 \times 8}} = \sqrt{\frac{7}{22}} = 0.564$$ **Ans.**

Exercise 10.2

1. Find the coefficient of correlation between x and y from the table of their values :

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

Ans. 0.977.

2. Find the coefficient of correlation of the following data taking new origin of x at 70 and for y at 67.

x	67	68	64	68	72	70	69	70
y	65	66	67	67	68	69	71	73

(*AMIE winter 2002*) **Ans.** 0.472

3. x and y are two random variables with the same standard deviation and correlation coefficient r. Show that the coefficient of correlation between x and $x + y$ is $\sqrt{\frac{1+r}{2}}$.

4. Find the regression line of y on x for the data :

x	1	4	2	3	5
y	3	1	2	5	4

Ans. $y = 2.7 + 0.1x$

5. Find the correlation coefficient and the equations of regression lines from the following data :

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

Ans. $r = 0.81$, $x = 0.5y + 0.5$, $y = 1.3x + 1.1$

6. Find the regression line of y on x if

x	40	70	50	60	80	50	90	40	60	60
y	2.5	6.0	4.5	5.0	4.5	2.0	5.5	3.0	4.5	3.0

Ans. $y = 0.55 + 0.0583\ x$

7. The following marks have been obtained by a class of students in statistics.

Paper I	80	45	55	56	58	60	65	68	70	75	85
Paper II	81	56	50	48	60	62	64	65	70	74	90

Compute the coefficient of correlation for the above data. Find the lines of regression.

Ans. $r = .918$, $y - 65.45 = 0.981\ (x - 65.18)$

$x - 65.18 = 0.859\ (y - 65.45)$

8. Find the equations to the lines of regression and the coefficient of correlation for the following data:

x	2	4	5	6	8	11
y	18	12	10	8	7	5

Ans. $y - 10 = -\ 1.34\ (x - 6)$, $x - 6 = -\ 0.632\ (y - 10)$, $r = -\ 0.92$

9. Obtain normal equations for fitting a curve of the form $y = ax + \frac{b}{x}$

for n points $(x_r,\ y_r)$, $r = 1, 2, \ldots n$.

Ans. $\Sigma xy = nb + a\,\Sigma x^2$, $\Sigma \frac{y}{x} = na + b\,\Sigma \frac{1}{x^2}$

10. The following results were obtained from lineups in Applied Mechanics and Engineering Mathematics in an examination :

	Applied Mechanics *(x)*	*Engg. Maths.* *(y)*
Mean	47.5	10.5
Standard deviation	16.8	10.8

$r = 0.95$

Find both the regression equations. Also estimate the value of y for $x = 30$.

Ans. $y = 0.611x + 10.5$, $x = 1.478\,y - 1.143$, $y = 28.83$

11. The following results were obtained from records of age (x) and systolic blood pressure (y) of a group of 10 men :

	x	y
Mean	53	142
Variance	130	165

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.

Ans. $y = 0.94\, x + 92.26$, Blood pressure = 134.56

12. The regression equation are : $7x - 16y + 9 = 0,\ 5y - 4x - 3 = 0$ find $\bar{x},\ \bar{y}$ and r

(*AMIE, Winter 2003*) **Ans.** $\bar{x} = -\dfrac{3}{29},\ \bar{y} = \dfrac{15}{29},\ r = \dfrac{3}{4}$

13. If two regression coefficients are 0.8 and 0.2, what would be the value of coefficient of correlation?

Ans. $r = 0.4$

14. In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible :

Variance of $x = 9$

Regression equations : $8x - 10y + 66 = 0$, $40x - 18y - 214 = 0$.

What were (*a*) the mean values of x and y, (*b*) the standard deviation of y, and (*c*) the coefficient of correlation between x and y. *(A.M.I.E., Summer 2001, 2000)*

Ans. $\bar{x} = 13,\ \bar{y} = 17,\ y = .8x + 6.6\ ,\ x = .45y - 5.35,\ r = 0.6,\ \sigma_y = 4.$

15. The following regression equations and variances are obtained from a correlation table :

$20x - 9y - 107 = 0$, $4x - 5y + 33 = 0$, variance of $x = 9$.

Find (*i*) the mean values of x and y, (*ii*) the standard deviation of y. **Ans.** $\bar{x} = 13,\ \bar{y} = 17,\ \sigma_y = 4.$

(*A.M.I.E., Winter 2000*)

16. Two random variables have the least square regression lines with equations $3x + 2y = 26$ and $6x + y = 31$. Find mean values and correlation coefficient between x and y.

(*A.M.I.E., Summer 1998*) **Ans.** $\bar{x} = 4,\ \bar{y} = 7,\ r = -0.5$

17. Find the standard error of estimate of y on x for the data given below

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

Ans. 1.349

18. Fill in the blanks :

(*a*) The correlation coefficient is the .... mean between the regression coefficients.

(*b*) The lines of regression always pass through a point ....

(*c*) Arithmetic mean of the coefficients of regressions is .... than the coefficient of correlation.

(*A.M.I.E., Summer 2000*)

(*d*) The value of coefficient of correlation lies between .... and ....

(*e*) If the two regression lines are perpendicular to each other, then the coefficient of correlation is equal to ....

(*f*) If two regression coefficients are, – 0.1 and – 0.9, the value of r is ....

(*g*) The normal equations for fitting a curve of the form $y = a + bx + cx^2$ are ...., ... and ....

(*h*) If r_1 and r_2 are two regression coefficients, then signs of $r_1,\ r_2$ depend on ....

(*i*) If coefficient of correlation $r = 0$, the two lines of regression are ....

(*j*) If two regression lines concide then the coefficient of correlation is .... (*A.M.I.E., Winter 2000*)

Ans. (*a*) geometric, (*b*) $(\bar{x},\ \bar{y})$, (*c*) greater, (*d*) – 1 and 1, (*e*) 0, (*f*) – 0.3,

(*g*) $\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2, \Sigma xy = a\,\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3$ and $\Sigma x^2 y = a\,\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4$,

(*h*) Coefficient of regression (*i*) Perpendicular. (j) ± 1

11

Probability

11.1 PROBABILITY

Probability is a concept which numerically measure the degree of uncertainty and therefore, of certainity of the occurrence of events.

If an event A can happen in m ways, and fail in n ways, all these ways being equally likely to occur, then the probability of the happening of A is

$$= \frac{\text{Number of favourable cases}}{\text{Total number of mutually exclusive and equally likely cases}} = \frac{m}{m+n}$$

and that of its failing is defined as $\frac{n}{m+n}$.

If the probability of the happening $= p$
and the probability of not happening $= q$

then $$p+q = \frac{m}{m+n} + \frac{n}{m+n} = \frac{m+n}{m+n} = 1 \quad \text{or} \quad p+q = 1.$$

For instance, on tossing a coin, the probability of getting a head is $\frac{1}{2}$.

11.2 DEFINITIONS

1. **Die :** It is a small cube. Dots are . .. ... :: :·: ::: marked on its faces. Plural of the die is dice. On throwing a die, the outcome is the number of dots on its upper face.

2. **Cards :** A pack of cards consists of four suits *i.e.* Spades, Hearts, Diamonds and Clubs. Each suit consists of 13 cards, nine cards numbered 2, 3, 4, ..., 10, an Ace, a King, a Queen and a Jack or Knave. Colour of Spades and Clubs is black and that of Hearts and Diamonds is red. Aces, Kings, Queens, and Jacks are known as *face* cards.

3. **Exhaustive Events or Sample Space :** The set of all possible outcomes of a single performance of an experiment is exhaustive events or sample space. Each outcome is called a sample point. In case of tossing a coin once, $S = (H, T)$ is the *sample space.* Two outcomes – Head and Tail – constitute an exhaustive event because no other outcome is possible.

4. **Random Experiment :** There are experiments, in which results may be altogether different, even though they are performed under identical conditions. They are known as random experiments. Tossing a coin or throwing a die is random experiment.

5. **Trial and Event :** Performing a random experiment is called a *trial* and outcome is termed as *event.* Tossing of a coin is a trial and the turning up of head or tail is an event.

6. **Equally likely events :** Two events are said to be *'equally likely'*, if one of them cannot be expected in preference to the other. For instance, if we draw a card from well-shuffled pack, we may get any card, then the 52 different cases are equally likely.

7. **Independent events :** Two events may be *independent*, when the actual happening of one does not influence in any way the probability of the happening of the other.

Example. The event of getting head on first coin and the event of getting tail on the second coin in a simultaneous throw of two coins are independent.

8. **Mutually Exclusive events :** Two events are known as *mutually exclusive*, when the occurrence of one of them excludes the occurrence of the other. For example, on tossing of a coin, either we get head or tail, but not both.

9. **Compound Event :** When two or more events occur in composition with each other, the simultaneous occurrence is called a compound event. When a die is thrown, getting a 5 or 6 is a compound event.

10. **Favourable Events :** The events, which ensure the required happening, are said to be favourable events. For example, in throwing a die, to have the even numbers, 2, 4 and 6 are favourable cases.

11. **Conditional Probability :** The probability of happening an event A, such that event B has already happened, is called the conditional probability of happening of A on the condition that B has already happened. It is usually denoted by $P\ (A/B)$.

12. **Odds in favour of an event and odds against an event**

If number of favourable ways = m, number of not favourable events = n

(i) Odds in favour of the event = $\frac{m}{n}$, Odds against the event = $\frac{n}{m}$.

13. **Classical Definition of Probability.** If there are N equally likely, mutually, exclusive and exhaustive of events of an experiment and m of these are favourable, then the probability of the happening of the event is defined as $\frac{m}{N}$..

14. **Expected value.** If p_1, p_2, p_3 ... p_n of the probabilities of the events x_1, x_2, x_3 ... x_n respectively then expected value

$$E\,(x) = p_1\,x_1 + p_2\,x_2 + p_3\,x_3 + \ldots + p_n\,x_n = \sum_{r=1}^{n} p_r\,x_r$$

Example 1. *Find the probability of throwing (a) 5, (b) an even number with an ordinary six faced die.*

Solution. (a) There are 6 possible ways in which the die can fall and there is only one way of throwing 5.

$$\text{Probability} = \frac{\text{Number of favourable ways}}{\text{Total number of equally likely ways}} = \frac{1}{6} \qquad \textbf{Ans.}$$

(b) Total number of ways of throwing a die = 6

Number of ways falling 2, 4, 6 = 3

$$\text{The required probability} = \frac{3}{6} = \frac{1}{2} \qquad \textbf{Ans.}$$

Example 2. *Find the probability of throwing 9 with two dice.*

Solution. Total number of possible ways of throwing two dice

$$= 6 \times 6 = 36.$$

Number of ways getting 9 *i.e.*, (3 + 6), (4 + 5), (5 + 4), (6 + 3) = 4.

$$\therefore \qquad \text{The required probability} = \frac{4}{36} = \frac{1}{9}. \qquad \textbf{Ans.}$$

Example 3. *From a pack of 52 cards, one is drawn at random. Find the probability of getting a king.*

Solution. A king can be chosen in 4 ways. But a card can be drawn in 52 ways.

$\therefore$ The required probability $= \dfrac{4}{52} = \dfrac{1}{13}$. **Ans.**

Exercise 11.1

1. In a class of 12 students, 5 are boys and the rest are girls. Find the probability that a student selected will be a girl. **Ans.** $\dfrac{7}{12}$

2. A bag contains 7 red and 8 black balls. Find the probability of drawing a red ball. **Ans.** $\dfrac{7}{15}$

3. Three of the six vertices of a regular hexagon are choosen at random. Find the probability that the triangle with three vertices is equilateral. (*A.M.I.E.T.E., Summer 1999*) **Ans.** $\dfrac{1}{10}$

4. What is the probability that a leap year, selected at random, will contain 53 Sundays. (*A.M.I.E.T.E., Summer 2002, A.M.I.E., Summer 2001, Winter 1998*) **Ans.** $\dfrac{2}{7}$

5. Choose the correct answer :

(a) In solving any problem, odds against A are 4 to 3 and odds in favour of B in solving the same problem are 7 to 5. The probability that the problem will be solved is

(*i*) $\dfrac{5}{21}$ (*ii*) $\dfrac{16}{21}$ (*iii*) $\dfrac{15}{84}$ (*iv*) $\dfrac{69}{84}$ (*A.M.I.E.T.E., Winter 2003*) **Ans.** (*ii*)

(*b*) In a given race, the odds in favour of horses *A, B, C, D* are 1 : 3, 1 : 4, 1 : 5, 1 : 6 respectively. The probability that horse *C* wins the race is

(*i*) $\dfrac{1}{4}$ (*ii*) $\dfrac{1}{5}$ (*iii*) $\dfrac{1}{6}$ (*iv*) $\dfrac{1}{7}$ **Ans.** (*iii*)

(*c*) In tossing a fair die, the probability of getting on odd number or a number less than 4 is

(*i*) 2 (*ii*) 1/2 (*iii*) 2/3 (*iv*) 3/4 (*A.M.I.E.T.E., Summer 1997*) **Ans.** (*iii*)

(*d*) An unbiased coin is tossed 3 times. The probability of obtaining two heads is

(*i*) $\dfrac{1}{2}$ (*ii*) $\dfrac{3}{8}$ (*iii*) 1 (*iv*) $\dfrac{1}{8}$ (*A.M.I.E.T.E., Winter 2002*) **Ans.** (*ii*)

6. Fill in the blanks with appropriate correct answer

(*a*) Chance of throwing 6 at least once in four throws with single dice is ....

(*A.M.I.E., Summer 2000*) **Ans.** $\dfrac{671}{1296}$

(*b*) A pair of fair dice is thrown and one die shows a four. The probability that the other die shows 5 is .... (*A.M.I.E., Summer 2000*) **Ans.** $\dfrac{1}{36}$

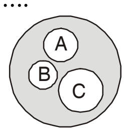
11.3 ADDITION LAW OF PROBABILITY

If $p_1, p_2, ..., p_n$ be separate probabilities of mutually exclusive events, then the probability P, that any of these events will happen is given by $P = p_1 + p_2 + p_3 + ... + p_n$

Proof. Let *A, B, C,......* be the events, where probabilities are respectively $p_1, p_2, ... p_n$.

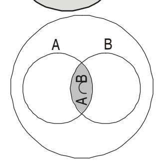
Let n be the total number of favourable cases to either A or B or C or......

$$= m_1 + m_2 + m_3 + ... + m_n$$



Hence $P\,(A + B + C...)$ $= \dfrac{m_1 + m_2 + m_3 + ... + m_n}{n}$

$$= \frac{m_1}{n} + \frac{m_2}{n} + \frac{m_3}{n} + ... + \frac{m_n}{n}$$



$$= P\,(A) + P\,(B) + P\,(C) + ...$$

$$P = p_1 + p_2 + p_3 + ... + p_n \qquad \textbf{Proved}$$

NOT MUTUALLY EXCLUSIVE EVENTS

Consider the case where two events *A* and *B* are not mutually exclusive. The probability of the event that either *A* or *B* or both occur is given as

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

Example 4. *An urn contains 10 black and 10 white balls. Find the probability of drawing two balls of the same colour.*

Solution. Probability of drawing two black balls = $\dfrac{^{10}C_2}{^{20}C_2}$

$\therefore$ Probability of drawing two red balls = $\dfrac{^{10}C_2}{^{20}C_2}$

$\therefore$ Probability of drawing two balls of the same colour

$$= \frac{^{10}C_2}{^{20}C_2} + \frac{^{10}C_2}{^{20}C_2} = 2 \cdot \frac{^{10}C_2}{^{20}C_2} = 2 \cdot \frac{\dfrac{10 \times 9}{2 \times 1}}{\dfrac{20 \times 19}{2 \times 1}}$$

$$= \frac{9}{19}$$ **Ans.**

Example 5. *A bag contains four white and two black balls and a second bag contains three of each colour. A bag is selected at random, and a ball is then drawn at random from the bag chosen. What is the probability that the ball drawn is white ?*

Solution. There are two mutually exclusive cases,

(*i*) when the first bag is chosen, (*ii*) when the second bag is chosen.

Now the chance of choosing the first bag is $\frac{1}{2}$ and if this bag is chosen, the probability of drawing a white ball is 4/6. Hence the probability of drawing a white ball from first bag is

$$\frac{1}{2} \times \frac{4}{6} = \frac{1}{3}$$

Similarly the probability of drawing a white ball from second bag is

$$\frac{1}{2} \times \frac{3}{6} = \frac{1}{4}$$

Since the events are mutually exclusive the required probability

$$= \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$ **Ans.**

Example 6. *Three machines I, II and III manufacture respectively 0.4, 0.5 and 0.1 of the total production. The percentage of defective items produced by I, II and III is 2, 4 and 1 per cent respectively. For an item chosen at random, what is the probability it is defective ?*

Solution. The defective item produced by machine I = $\dfrac{0.4 \times 2}{100} = \dfrac{0.8}{100}$

The defective item produced by machine II = $\dfrac{0.5 \times 4}{100} = \dfrac{2}{100}$

The defective item produced by machine III = $\dfrac{0.1 \times 1}{100} = \dfrac{0.1}{100}$

The total defective items produced by machines I, II, III

$$= \frac{0.8}{100} + \frac{2}{100} + \frac{0.1}{100} = \frac{2.9}{100} = 0.029$$

The required probability $= \dfrac{0.029}{1} = 0.029$ **Ans.**

11.4 MULTIPLICATION LAW OF PROBABILITY

If there are two independent events the respective probabilities of which are known, then the probability that both will happen is the product of the probabilities of their happening respectively.

$$P(AB) = P(A) \times P(B)$$

Proof. Suppose A and B are two independent events. Let A happen in m_1 ways and fail in n_1 ways.

$$\therefore \quad P(A) = \frac{m_1}{m_1 + n_1}$$

Also let B happen in m_2 ways and fail in n_2 ways.

$$\therefore \quad P(B) = \frac{m_2}{m_2 + n_2}$$

Now there are four possibilities

A and B both may happen, then the number of ways $= m_1 \,.\, m_2$.

A may happen and B may fail, then the number of ways $= m_1 \,.\, n_2$

A may fail and B may happen, then the number of ways $= n_1 \,.\, m_2$

A and B both may fail, then the number of ways $= n_1 \,.\, n_2$

Thus, the total number of ways $= m_1m_2 + m_1n_2 + n_1m_2 + n_1n_2$

$$= (m_1 + n_1)(m_2 + n_2)$$

Hence the probabilities of the happening of both A and B

$$P(AB) = \frac{m_1m_2}{(m_1 + n_1)(m_2 + n_2)} = \frac{m_1}{m_1 + n_1} \cdot \frac{m_2}{m_2 + n_2}$$

$$= P(A) \,.\, P(B)$$ **Proved.**

Example 7. *An article manufactured by a company consists of two parts A and B. In the process of manufacture of part A, 9 out of 100 are likely to be defective. Similarly, 5 out of 100 are likely to be defective in the manufacture of part B. Calculate the probability that the assembled article will not be defective (assuming that the events of finding the part A non-defective and that of B are independent).*

Solution. Probability that part A will be defective $= \dfrac{9}{100}$

Probability that part A will not be defective $= \left(1 - \dfrac{9}{100}\right) = \dfrac{91}{100}$

Probability that part B will be defective $= \dfrac{5}{100}$

Probability that part B will not be defective $= \left(1 - \dfrac{5}{100}\right) = \dfrac{95}{100}$

Probability that the assembled article will not be defective = (Probability that part A will not be defective) × (Probability that part B will not be defective)

$$= \left(\frac{91}{100}\right) \times \left(\frac{95}{100}\right) = 0.8645$$ **Ans.**

Example 8. *The probability that machine A will be performing an usual function in 5 years' time is $\frac{1}{4}$, while the probability that machine B will still be operating usefully at the end of the same period, is $\frac{1}{3}$.*

Find the probability in the following cases that in 5 years time:

(i) Both machines will be performing an usual function.

(ii) Neither will be operating.

(iii) Only machine B will be operating.

(iv) At least one of the machines will be operating.

Solution. P (A operating usefully) $= \frac{1}{4}$, $\quad q(A) = 1 - \frac{1}{4} = \frac{3}{4}$

P (B operating usefully) $= \frac{1}{3}$ so $q(B) = 1 - \frac{1}{3} = \frac{2}{3}$

(i) P (Both A and B will operate usefully) $= P(A) \,.\, P(B) = \left(\frac{1}{4}\right) \times \left(\frac{1}{3}\right) = \frac{1}{12}$

(ii) P (Neither will be operating) $= q(A) \,.\, q(B) = \left(\frac{3}{4}\right)\left(\frac{2}{3}\right) = \frac{1}{2}$

(iii) P (Only B will be operating) $= p(B) \times q(A) = \left(\frac{1}{3}\right) \times \left(\frac{3}{4}\right) = \frac{1}{4}$

(iv) P (At least one of the machines will be operating)

$$= 1 - P = 1 - P \qquad \text{(none of them operates)}$$

$$= 1 - \frac{1}{2} = \frac{1}{2}$$ **Ans.**

Example 9. *There are two groups of subjects one of which consists of 5 science and 3 engineering subjects and the other consists of 3 science and 5 engineering subjects. An unbiased die is cast. If number 3 or number 5 turns up, a subject is selected at random from the first group, otherwise the subject is selected at random from the second group. Find the probability that an engineering subject is selected ultimately.* (*A.M.I.E.T.E., Summer 2000*)

Solution. Probability of turning up 3 or 5 $= \frac{2}{6} = \frac{1}{3}$

Probability of selecting engineering subject from first group $= \frac{3}{8}$

Now the probability of selecting engineering subject from first group on turning up 3 or 5

$$= \left(\frac{1}{3}\right) \times \left(\frac{3}{8}\right) = \frac{1}{8} \qquad \text{... (1)}$$

Probability of not turning up 3 or 5 $= 1 - \frac{1}{3} = \frac{2}{3}$

Probability of selecting engineering subject from second group $= \frac{5}{8}$

$$= \frac{2}{3} \times \frac{5}{8} = \frac{5}{12} \quad \text{... (2)}$$

Probability of the selection of engineering subject $= \frac{1}{8} + \frac{5}{12}$ [From (1) and (2)]

$$= \frac{13}{24} \quad \textbf{Ans.}$$

Example 10. *An urn contains nine balls, two of which are red, three blue and four black. Three balls are drawn from the urn at random. What is the probability that*

(*i*) *the three balls are of different colours?*

(*ii*) *the three balls are of the same colour?* (*A.M.I.E., Summer 2000*)

Solution.

Urn contains 2 Red balls, 3 Blue balls and 4 Black balls.

(*i*) Three balls will be of different colours if one ball is red, one blue and one black ball are drawn.

$$\text{Required probability} = \frac{{}^2C_1 \times {}^3C_1 \times {}^4C_1}{{}^9C_3} = \frac{2 \times 3 \times 4}{84} = \frac{2}{7} \quad \textbf{Ans.}$$

(*ii*) Three balls will be of the same colour if either 3 blue balls or 3 black balls are drawn.

P (3 Blue balls or 3 Black balls) = P (3 Blue balls) + P (3 Black balls)

$$= \frac{{}^3C_3}{{}^9C_3} + \frac{{}^4C_3}{{}^9C_3} = \frac{1+4}{84} = \frac{5}{84} \quad \textbf{Ans.}$$

Example 11. *An urn A contains 2 white and 4 black balls. Another urn B contains 5 white and 7 black balls. A ball is transferred from the urn A to the urn B, then a ball is drawn from urn B. Find the probability that it is white.*

Solution. Urn A contains 2 white and 4 black balls.

Urn B contains 5 white and 7 black balls.

Now there are two cases of transferring a ball from A to B.

Case I. When a white ball is transferred from A to B

$$P\text{ (Transfer of a white ball)} = \frac{2}{2+4} = \frac{1}{3}$$

After transfer of a white ball, urn B contains 6 white balls and 7 black balls.

P (Drawing a white ball from urn B after transfer)

$$= P\text{ (Transfer of a white ball)} \times P\text{ (Drawing of a white ball)}$$

$$= \left(\frac{1}{3}\right)\left(\frac{6}{6+7}\right) = \frac{1}{3} \times \frac{6}{13} = \frac{2}{13}$$

Case II. When a black ball is transferred from A to B.

$$P\text{ (Transfer of a black ball)} = \frac{4}{2+4} = \frac{2}{3}$$

After transfer of a black ball, urn B contains 5 white and 8 black balls.

P (Drawing a white ball from urn B after transfer)

$$= P\text{ (Transfer of a black ball)} \times P\text{ (Drawing of a white ball)}$$

$$\text{Required probability} = \frac{2}{13} + \frac{10}{39} = \frac{16}{39}$$ **Ans.**

Example 12. *A bag contains 10 white and 15 black balls. Two balls are drawn in succession. What is the probability that first is white and second is black ?*

Solution. Probability of drawing one white ball = $\frac{10}{25}$

Probability of drawing one black ball without replacement = 15/24

Required probability of drawing first white ball and second black ball = $\frac{10}{25} \times \frac{15}{24} = \frac{1}{4}$

Example 13. *A committee is to be formed by choosing two boys and four girls out of a group of five boys and six girls. What is the probability that a particular boy named A and a particular girl named B are selected in the committee?* (*A.M.I.E., Summer 1997*)

Solution. Two boys are to be selected out of 5 boys. A particular boy A is to be included in the committee. It means that only 1 boy is to be selected out of 4 boys.

Number of ways of selection = ${}^4C_1$

Similarly a girl B is to be included in the committee.

Then only 3 girls are to be selected out of 5 girls.

Number of ways of selection = ${}^5C_3$

$$\text{Required probability} = \frac{{}^4C_1 \times {}^5C_3}{{}^5C_2 \times {}^6C_4} = \frac{4 \times 10}{10 \times 15} = \frac{4}{15}$$ **Ans.**

Example 14. *Three groups of children contain respectively 3 girls and 1 boy; 2 girls and 2 boys; 1 girl and 3 boys. One child is selected at random from each group. Find the chance of selecting 1 girl and 2 boys.*

Solution. There are three ways of selecting 1 girl and two boys.

I way : Girl is selected from first group, boy from second group and second boy from third group.

Probability of the selection of (Girl + Boy + Boy) = $\frac{3}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{18}{64}$

II way : Boy is selected from first group, girl from second group and second boy from third group.

Probability of the selection of (Boy + Girl + Boy) = $\frac{1}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{6}{64}$

III way : Boy is selected from first group, second boy from second group and the girl from the third group.

Probability of selection of (Boy + Boy + Girl) = $\frac{1}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{2}{64}$

$$\text{Total probability} = \frac{18}{64} + \frac{6}{64} + \frac{2}{64} = \frac{26}{64} = \frac{13}{32}$$ **Ans.**

Example 15. *The number of children in a family in a region are either 0, 1 or 2 with probability 0.2, 0.3 and 0.5 respectively. The probability of each child being a boy or girl 0.5. Find the probability that a family has no boy.* (*A.M.I.E.T.E., Winter 1998*)

Solution. Here there are three types of families

(*i*) Probability of zero child (boys) = 0.2

(*ii*)

Boy	*Girl*
0	1
1	0

Probability of zero boy in case II

$$= 0.3 \times 0.5 = 0.15$$

(*iii*)

Boy	*Girl*
0	2
1	1
2	0

In this case probability of zero boy $= 0.5 \times \frac{1}{3} = 0.167$

Considering all the three cases, the probability of zero boy

$$= 0.2 + 0.15 + 0.167 = 0.517$$ **Ans.**

Example 16. *A husband and wife appear in an interview for two vacancies in the same post. The probability of husband's selection is* $\frac{1}{7}$ *and that of wife's selection is* $\frac{1}{5}$*. What is the probability that*

(i) both of them will be selected. *(ii) only one of them will be selected, and*

(iii) none of them will be selected ?

Solution. P (husband's selection) $= \frac{1}{7}$, P (wife's selection) $= \frac{1}{5}$

(*i*) P (both selected) $= \frac{1}{7} \times \frac{1}{5} = \frac{1}{35}$

(*ii*) P (only one selected) $= P$ (only husband's selection) $+ P$ (only wife's selection)

$$= \frac{1}{7} \times \frac{4}{5} + \frac{1}{5} \times \frac{6}{7} = \frac{10}{35} = \frac{2}{7}$$

(*iii*) P (none of them will be selected) $= \frac{6}{7} \times \frac{4}{5} = \frac{24}{35}$ **Ans.**

Example 17. *A problem of statistics is given to three students A, B and C whose chances of solving it are* $\frac{1}{2}$, $\frac{3}{4}$ *and* $\frac{1}{4}$ *respectively. What is the probability that the problem will be solved ?* (*A.M.I.E., Winter 2001*)

Solution. The probability that A can solve the problem $= \frac{1}{2}$

The probability that A cannot solve the problem $= 1 - \frac{1}{2}$.

Similarly the probability that B and C cannot solve the problem are $\left(1 - \frac{3}{4}\right)$ and $\left(1 - \frac{1}{4}\right)$.

$\therefore$ The probability that A, B, C cannot solve the problem

$$= \left(1 - \frac{1}{2}\right) \times \left(1 - \frac{3}{4}\right) \times \left(1 - \frac{1}{4}\right) = \frac{1}{2} \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{32}.$$

Hence the probability that the problem can be solved

$$= 1 - \frac{3}{32} = \frac{29}{32} \qquad \textbf{Ans.}$$

Example 18. *A student takes his examination in four subjects* $\alpha, \beta, \gamma, \delta$. *He estimates his chances of passing in* α *as* $\frac{4}{5}$, *in* β *as* $\frac{3}{4}$, *in* γ *as* $\frac{5}{6}$ *and in* δ *as* $\frac{2}{3}$. *To qualify, he must pass in* α *and at least two other subjects. What is the probability that he qualifies ?*

Solution. $P(\alpha) = \frac{4}{5}$, $P(\beta) = \frac{3}{4}$, $P(\gamma) = \frac{5}{6}$, $P(\delta) = \frac{2}{3}$

There are four possibilities of passing at least two subjects

(*i*) Probability of passing β, γ and failing δ

$$= \frac{3}{4} \times \frac{5}{6} \times \left(1 - \frac{2}{3}\right) = \frac{3}{4} \times \frac{5}{6} \times \frac{1}{3} = \frac{5}{24}$$

(*ii*) Probability of passing γ, δ and failing β

$$= \frac{5}{6} \times \frac{2}{3} \times \left(1 - \frac{3}{4}\right) = \frac{5}{6} \times \frac{2}{3} \times \frac{1}{4} = \frac{5}{36}$$

(*iii*) Probability of passing δ, β and failing γ

$$= \frac{2}{3} \times \frac{3}{4} \times \left(1 - \frac{5}{6}\right) = \frac{2}{3} \times \frac{3}{4} \times \frac{1}{6} = \frac{1}{12}$$

(*iv*) Probability of passing β, γ, δ

$$= \frac{3}{4} \times \frac{5}{6} \times \frac{2}{3} = \frac{5}{12}.$$

Probability of passing at least two subjects

$$= \frac{5}{24} + \frac{5}{36} + \frac{1}{12} + \frac{5}{12} = \frac{61}{72}$$

Probability of passing α and at least two subjects

$$= \frac{4}{5} \times \frac{61}{72} = \frac{61}{90} \qquad \textbf{Ans.}$$

Example 19. *There are 6 positive and 8 negative numbers. Four numbers are chosen at random, without replacement, and multiplied. What is the probability that the product is a positive number ?*

Solution. To get from the product of four numbers, a positive number, the possible combinations are as follows :

S. No.	*Out of 6 Positive Numbers*	*Out of 8 Negative Numbers*	*Positive Numbers*
1.	4	0	${}^6C_4 \times {}^8C_0 = \frac{6 \times 5}{1 \times 2} \times 1 = 15$
2.	2	2	${}^6C_2 \times {}^8C_2 = \frac{6 \times 5}{1 \times 2} \times \frac{8 \times 7}{1 \times 2} = 420$
3.	0	4	${}^6C_0 \times {}^8C_4 = 1 \times \frac{8 \times 7 \times 6 \times 5}{1 \times 2 \times 3 \times 4} = 70$
			Total = 505

$$\text{Probability} = \frac{{}^6C_4 \times {}^8C_0 + {}^6C_2 \times {}^8C_2 + {}^6C_0 \times {}^8C_4}{14\, C_4}$$

$$= \frac{15 + 420 + 70}{\dfrac{14 \times 13 \times 12 \times 11}{1 \times 2 \times 3 \times 4}} = \frac{505 \times 4 \times 3 \times 2 \times 1}{14 \times 13 \times 12 \times 11} = \frac{505}{1001}$$ **Ans.**

Example 20. *A six-faced die is so biased that, when thrown, it is twice as likely to show an even number than an odd number. If it is thrown twice, what is the probability that the sum of two numbers thrown is odd.*

Solution. A biased die, when thrown, shows even number twice than an odd number.

Probability of showing even number $= \dfrac{2}{2+1} = \dfrac{2}{3}$

Probability of showing odd number $= \dfrac{1}{1+2} = \dfrac{1}{3}$

Sum of two numbers is odd if the first is even and the second is odd or vice versa.

Probability of sum to be odd = Probability of an even number × Probability of an odd number + Probability of an odd number × Probability of an even number.

$$= \frac{2}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} = \frac{2}{9} + \frac{2}{9} = \frac{4}{9}$$ **Ans.**

Example 21. *A can hit a target 3 times in 5 shots, B 2 times in 5 shots and C three times in 4 shots. All of them fire one shot each simultaneously at the target. What is the probability that* (*i*) *2 shots hit* (*ii*) *At least two shots hit*? (*A.M.I.E.T.E., Summer 2003*)

Solution. Probability of A hitting the target $= \dfrac{3}{5}$

Probability of B hitting the target $= \dfrac{2}{5}$

Probability of C hitting the target $= \dfrac{3}{4}$

Probability that 2 shots hit the target

$$= P(A)\ P(B)\ q(C) + P(A)\ P(C)\ q(B) + P(B)\ P(C)\ q(A)$$

$$= \frac{3}{5} \times \frac{2}{5} \times \left(1 - \frac{3}{4}\right) + \frac{3}{5} \times \frac{3}{4} \times \left(1 - \frac{2}{5}\right) + \frac{2}{5} \times \frac{3}{4} \times \left(1 - \frac{3}{5}\right)$$

$$= \frac{6}{25} \times \frac{1}{4} + \frac{9}{20} \times \frac{3}{5} + \frac{6}{20} \times \frac{2}{5}$$

$$= \frac{6 + 27 + 12}{100} = \frac{45}{100} = \frac{9}{20}$$ **Ans.**

(*ii*) Probability of at least two shots hitting the target

= Probability of 2 shots + probability of 3 shots hitting the target

$$= \frac{9}{20} + P(A)\ P(B)\ P(C) = \frac{9}{20} + \frac{3}{5} \times \frac{2}{5} \times \frac{3}{4} = \frac{63}{100}$$ **Ans.**

Example 22. *A and B take turns in throwing two dice, the first to throw 10 being awarded the prize. Show that if A has the first throw, their chances of winning are in the ratio* 12:11.

Solution. The combinations of throwing 10 from two dice can be (6 + 4), (4 + 6), (5 + 5).

The number of combinations is 3.

Total combinations from two dice $= 6 \times 6 = 36$.

$\therefore$ The probability of throwing 10 $= p = \dfrac{3}{36} = \dfrac{1}{12}$

The probability of not getting 10 $= q = 1 - \left(\dfrac{1}{12}\right) = \dfrac{11}{12}$

If A is to win, he should throw 10 in either the first, the third, the fifth, ... throws.

Their respective probabilities are $= p, q^2 p, q^4 p, \ldots\ldots = \dfrac{1}{12}, \left(\dfrac{11}{12}\right)^2 \dfrac{1}{12}, \left(\dfrac{11}{12}\right)^4 \dfrac{1}{12} \ldots$

A's total probability of winning $= \dfrac{1}{12} + \left(\dfrac{11}{12}\right)^2 \cdot \dfrac{1}{12} + \left(\dfrac{11}{12}\right)^4 \cdot \dfrac{1}{12} + \ldots$

$$= \frac{\frac{1}{12}}{1 - \left(\frac{11}{12}\right)^2} = \frac{12}{23} \quad \left[\text{This is infinite G.P. Its sum} = \frac{a}{1-r}\right]$$

B can win in either 2nd, 4th, 6th ... throws.

So B's total chance of winning $= q\,p + q^3\,p + q^5 p + \ldots\ldots$

$$= \left(\frac{11}{12}\right)\left(\frac{1}{12}\right) + \left(\frac{11}{12}\right)^3 \left(\frac{1}{12}\right) + \left(\frac{11}{12}\right)^5 \left(\frac{1}{12}\right) + \ldots = \frac{\left(\frac{11}{12}\right)\left(\frac{1}{12}\right)}{1 - \left(\frac{11}{12}\right)^2} = \frac{11}{23}$$

Hence A's chance to B's chance $= \dfrac{12}{23} : \dfrac{11}{23} = 12 : 11.$ **Proved.**

Example 24. *A and B throw alternatively 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. Find their respective chances of winning, if A begins.* (*A.M.I.E.T.E., Summer 2002*)

Solution. Number of ways of throwing 6

i.e. $(1 + 5), (2 + 4), (3 + 3), (4 + 2), (5 + 1) = 5.$

Probability of throwing 6 $= \dfrac{5}{36} = p_1, \quad q_1 = \dfrac{31}{36}$

Number of ways of throwing 7

i.e. $(1 + 6), (2 + 5), (3 + 4), (4 + 3), (5 + 2), (6 + 1) = 6$

Probability of throwing $7 = \dfrac{6}{36} = \dfrac{1}{6} = P_2, \quad q_2 = \dfrac{5}{6}$

$$P(A) = p_1 + q_1\,q_2\,p_1 + q_1^2\,q_2^2\,p_1 + \ldots\ldots$$

$$P(B) = q_1\,p_2 + q_1^2\,q_2\,p_2 + q_1^3\,q_2^2\,p_2 + \ldots\ldots.$$

Probability of A's winning $= p_1 + q_1\,q_2\,p_1 + q_1^2\,q_2^2\;p_1 + \ldots\ldots.$

$$= \frac{p_1}{1 - q_1\,q_2} = \frac{\frac{5}{36}}{1 - \frac{31}{36} \times \frac{5}{6}} = \frac{5}{36} \times \frac{36 \times 6}{61} = \frac{30}{61}$$

Probability of B's winning $= q_1\,p_2 + q_1^2\,q_2\,p_2 + q_1^3\,q_2^2\,p_2 + \ldots$

$$= \frac{q_1 p_2}{1 - q_1 q_2} = \frac{\frac{31}{36} \times \frac{1}{6}}{1 - \left(\frac{31}{36}\right)\left(\frac{5}{6}\right)} = \frac{31}{36 \times 6} \times \frac{36 \times 6}{61} = \frac{31}{61}$$ **Ans.**

Exercise 11.2

1. The probability that Nirmal will solve a problem is $\frac{2}{3}$ and the probability that Satyajit will solve it is $\frac{3}{4}$. What is the probability that (*a*) *the problem will be solved* (*b*) *neither can solve it.*
(*A.M.I.E., Summer 1999*) **Ans.** (*a*) $\frac{11}{12}$, (*b*) $\frac{1}{12}$

2. Two persons A and B toss an unbiased coin alternately on the understanding that the first who gets the head wins. If A starts the game, then what are their respective chances of winning ?
(*A.M.I.E.T.E. Summer 2004*) **Ans.** 4 : 1

3. Four persons are chosen at random from a group containing 3 men, 2 women, and 4 children. Show that the probability that exactly two of them will be children is $\frac{10}{21}$.

4. A five digit number is formed by using the digits 0, 1, 2, 3, 4 and 5 without repetition. Find the probability that the number is divisible by 6. (*A.M.I.E.T.E., Summer 1999*) **Ans.** $\frac{4}{25}$

5. The chances that doctor A will diagnose a disease X correctly is 60%. The chances that a patient will die by his treatment after correct diagnosis is 40% and the chances of death by wrong diagnosis is 70%. A patient of doctor A, who had disease X, died, what is the chance that his disease was diagnosed correctly. **Ans.** $\frac{6}{13}$

6. An anti-aircraft gun can take a maximum of four shots on enemy's plane moving from it. The probabilities of hitting the plane at first, second, third and fourth shots are 0.4, 0.3, 0.2 and 0.1 respectively. Find the probability that the gun hits the plane. **Ans.** 0.6976.

7. An electronic component consists of three parts. Each part has probability 0.99 of performing satisfactorily. The component fails if two or more parts do not perform satisfactorily. Assuming that the parts perform independently, determine the probability that the component does not perform satisfactorily. **Ans.** 0.000298

8. The face cards are removed from a full pack. Out of the remaining 40 cards, 4 are drawn at random. What is the probability that they belong to different suits ? **Ans.** $\frac{1000}{9139}$

9. Of the cigarette smoking population, 70% are men and 30% women, 10% of these men and 20% of these women smoke 'WILLS.' What is the probability that a person seen smoking a 'WILLS' will be a man. **Ans.** $\frac{7}{13}$.

10. A machine contains a component C that is vital to its operation. The reliability of component C is 80%. To improve the reliability of a machine, a similar component is used in parallel to form a system S. The machine will work provided that one of these components functions correctly. Calculate the reliability of the system S. **Ans.** 96%

11. In a bolt factory, machines A, B and C manufacture 25%, 35% and 40% of the total output respectively. Of their outputs, 5%, 4% and 2% are defective bolts. A bolt is chosen at random and found to be defective. What is the probability that the bolt came from machine A ? B ? C ?

Ans. $\frac{25}{69}$, $\frac{28}{69}$, $\frac{16}{69}$

12. One bag contains four white and two black beads and another contains three of each colour. A bead is drawn from each bag. What is the probability that one is white and one is black ? **Ans.** $\frac{1}{2}$.

13. The odds that a book will be favourably reviewed by three independent critics are 5 to 2, 4 to 3, 3 to 4 respectively. What is the probability that of the three reviews, a majority will be favourable ?

(A.M.I.E., Summer 2004) **Ans.** $\frac{209}{343}$

14. Let E and F be independent events. The probability that both E and F happen is $\frac{1}{12}$ and the probability that neither E nor F happen is $\frac{1}{2}$. Then find $P(E)$ and $P(F)$.

Ans. $P(E) = \frac{1}{3}$, $P(F) = \frac{1}{4}$

15. Given a random variable whose range set is (1, 2) and whose probability is $f(1) = \frac{1}{4}$ and $f(2) = \frac{3}{4}$. Find the mean and variance of the distribution. **Ans.** Mean $= \frac{7}{4}$, Var $= \frac{3}{16}$

16. A man takes a step forward with probability 0.4 and backward with probability 0.6. Find the probability that at the end of 11 steps, he is just one step away from the starting point. **Ans.** 0.210677186

17. What would be the expectation of the number of failures preceding the first success in an infinite series of independent trials with the constant probability of success p ?

Solution. The probabilities of success in 1st, 2nd, 3rd trials respectively are $p, qp, q^2p, q^3p, \ldots\ldots$

The expected number of failures preceding the first success

$$E(x) = (0 \,.\, p) + (1 \,.\, qp) + (2 \,.\, q^2p) + \ldots\ldots \infty$$
$$= qp\,[1 + 2q + 3q^2 + \ldots\ldots \infty] \quad \text{where} \quad q < 1.$$
$$= \frac{qp}{(1-q)^2} = \frac{qp}{p^2} = \frac{q}{p}$$ **Ans.**

18. The probability of an airplane engine failure (independent of other engines) when the aircraft is in flight is (1-P). For a successful flight at least 50% of the airplane engines should remain operational. For which values of P would you prefer a four engine airplane to a two engine one

(A.M.I.E.T.E., Dec. 2004)

19. A person plays m independent games. The probability of his winning any game is $\frac{a}{a+b}$ (a, b are positive number). Show that probability that the person wins an odd number of games is $\frac{1}{2}\,[(b+a)^m - (b-a)^m\,]/(b+a)^m$.

20. Fill in the blanks :

(*a*) If the probabilities of n independent events are $p_1, p_2, p_3, \ldots, p_n$, then the probability that at least one of the event will happen is ....

(*b*) For a biased die, the probabilities for the different faces to turn up are given below :

Face	1	2	3	4	5	6
Prob.	0.1	0.32	0.21	0.15	0.05	0.17

The die is tossed and you are told that either face 1 or face 2 has turned up. Then the probability that it is face 1, is ....

(*c*) The probability of getting a ticket of number of multiple of 5 in a random draw from a bag containing tickets of even numbers from 1 to 100, is ....

(*d*) A town has two doctors X and Y operating independently. If the prob. the doctor X is available, is 0.9 and that for Y is 0.8, then the prob. that at least one doctor is available, when needed is ....

(*e*) From a pack of well shuffled cards, one card is drawn randomly. A gambler bets it as a diamond or a king. The odds in favour of his winning the bet are ....

(*f*) From a pack of cards, 2 cards are drawn, the first being replaced before the second is drawn. The probability that the first is a diamond and the second is a king will be ....

(*g*) From an urn containing 12 white and 8 black balls two balls are drawn at random. The probability that both the balls will turn to be black is .... (*A.M.I.E., Winter 1999*)

(*h*) A ball is taken out of a pot containing 6 white and 12 red balls. The probability that the ball is white is .... (*A.M.I.E., Summer 1997*)

(*i*) A speaks truth in 75% and B in 80% of the cases. The percentage of cases in which they likely to contradict each other narrating the same incident is ....

Ans. (*a*) $1-(1-p_1)(1-p_2)\ .....\ (1-p_n)$, (*b*) $\frac{5}{21}$, (*c*) $\frac{1}{5}$, (*d*) 0.98, (*e*) 4 : 9,

(*f*) $\frac{1}{52}$, (*g*) $\frac{14}{95}$, (*h*) $\frac{1}{3}$, (*i*) 35%

20. Tick √ the correct answer :

(*i*) The probability that at least one of the events A and B occurs is 0.8 and the probability that both the events occur simultaneously is 0.25. The probability $P(A) + P(B)$ is (*A.M.I.E.T.E. Summer 1997)*

(*i*) 0.65 (*ii*) 0.75 (*iii*) 0.85 (*iv*) 0.95

(*ii*) A, B, C are independent events such that $P(A) = P(B)$ and probability that at least one of them happens is ½. The probability that A or B happens given that at least one of A, B, or C happens is $\frac{2}{9}$. Find *P (A)* and *P (C)*. (*A.M.I.E.T.E., Winter 1996*) **Ans.** $P(A) = 1 - \frac{\sqrt{7}}{3}$, $P(i) = \frac{5}{14}$.

(*iii*) An unbiased coin is tossed five times. Given that heads were obtained in two of the tosses, the probability that these were obtained in the first two tosses is

(*a*) 1/10 (*b*) 1/4 (*c*) 1/32 (*d*) None of these. (*A.M.I.E.T.E., Summer 1996*)

(*iv*) Groups are formed of 4 persons out of 12 persons. The probability that one particular person is never included is

(*a*) 2/3 (*b*) 1/3 (*c*) 1/4 (*d*) none of these

(*v*) 50 tickets are serially numbered 1 to 50. One ticket is drawn from these at random. The probability of its being a multiple of 3 or 4 is

(*a*) 12/25 (*b*) 14/25 (*c*) 2/5 (*d*) none of these

(*vi*) The probabilities of occurring of two events *E, F* are 0.25 and 0.5 respectively and of occurring both simultaneously is 0.14. Then the probability of the occurrence of the neither event is

(*a*) 0.61 (*b*) 0.39 (*c*) 0.89 (*d*) none of these

(*vii*) A bag contains 5 black and 4 white balls. Two balls are drawn at random. The probability that they match, is

(*a*) 7/12 (*b*) 5/8 (*c*) 5/9 (*d*) 4/9

(*viii*) A, B, C in order toss a coin. The first to throw a head wins. Assuming the game continues indefinitely their respective chances of winning the games are

(*a*) $\frac{4}{7}, \frac{2}{7}, \frac{1}{7}$ (*b*) $\frac{1}{7}, \frac{4}{7}, \frac{2}{7}$ (*c*) $\frac{2}{7}, \frac{4}{7}, \frac{1}{7}$ (*d*) None of these

(*A.M.I.E.T.E., Winter 2000*)

(*ix*) A purse contains 4 copper coins, 3 silver coins, the second purse contains 6 copper coins and 2 silver coins. A coin is taken out of any purse, the probability that it is a copper coin is

(*a*) 4/7 (*b*) 3/4 (*c*) 3/7 (*d*) 37/56

(*x*) In rolling two fair dice, the probability of getting equal numbers or numbers with an even product is

(*a*) 6/36 (*b*) 30/36 (*c*) 27/36 (*d*) 3/36 (*A.M.I.E.T.E., Summer 1998*)

(*xi*) One of the two events must occur. If the chance of one is 2/3 of the other, then odds in favour of the other are

(*a*) 1 : 3 (*b*) 2 : 3 (*c*) 3 : 1 (*d*) none of these (*xii*) The probability that a certain beginner at golf gets a good shot if he uses the correct club is 1/3, and the probability of a good shot with an incorrect club is 1/4. In his bag are 5 different clubs, only one of which is correct for the shot in question. If he chooses a club at random and takes a stroke, the probability that he gets a good shot is

(*a*) $\frac{1}{3}$, (*b*) $\frac{1}{12}$, (*c*) $\frac{4}{15}$, (*d*) $\frac{7}{12}$.

(*xiii*) India plays two matches each with West Indies and Australia. In any match, the probabilities of India getting points 0, 1 and 2, are 0.45, 0.05 and 0.50 respectively. Assuming that the outcomes are independent, the probability of India getting at least 7 points is

(*a*) 0.8750 (*b*) 0.0875 (*c*) 0.625 (*d*) 0.0250. (*A.M.I.E.T.E., Summer 2001*)

(*xiv*) A bag contains 10 bolts, 3 of which are defective. Two bolts are drawn without replacement. The probability that both the bolts drawn are not defective is

(*a*) $\frac{49}{100}$ (*b*) $\frac{7}{15}$ (*c*) $\frac{4}{9}$ (*d*) $\frac{3}{10}$

(*xv*) The probability that a family has k children is $(0.5)^{k+1}$, $k = 0, 1, 2, ...$ If four families are chosen at random, the probability that each family has at least one child is

(*a*) 1/16 (*b*) 1/256 (*c*) 3/16 (*d*) 3/256

(*xvi*) The random variable X has $N(1, 4)$ distribution, then

(*a*) $P(x > 3) > P(x > 1)$ (*b*) $P(x > 3) < P(x < 1)$

(*c*) $P(x < 3) < P(x > 1)$ (*d*) $P(x < 3) < P(x < 1)$

(*xvii*) Two distinguishable dice are tossed simultaneously. The probability that multiple of 2 does not occur on the first die or multiple of 3 does not occur on the second die is

(*a*) $\frac{5}{36}$ (*b*) $\frac{10}{36}$ (*c*) $\frac{20}{36}$ (*d*) $\frac{30}{36}$

(*xviii*) An unbiased die with faces marked 1, 2, 3, 4, 5, 6 is rolled 4 times, out of four face values obtained, the probability that the minimum face value is not less than 2 and the maximum face value is not greater than 5 is then

(*a*) $\frac{16}{81}$ (*b*) $\frac{2}{9}$ (*c*) $\frac{80}{81}$ (*d*) $\frac{8}{9}$ (*A.M.I.E.T.E., Summer 2000*)

(*xix*) There are q persons sitting in a row. Two of them are selected at random, the probability that the two selected persons are not together is

(*a*) $\frac{2}{q}$ (*b*) $1 - \frac{2}{q}$ (*c*) $\frac{q(q-1)}{(q+1)(q+2)}$ (*d*) None of these

Ans. *(i) (b), (ii) (b), (iii) (a), (iv) (a), (v) (a), (vi) (b), (vii) (d), (viii) (a), (ix) (d), (x) (b), (xi) (d), (xii) (c), (xiii) (b) (xiv) (b), (xv) (a) . (xvi) (b), (xvii) (d), (xviii) (a), (xix) (b)*

11.4 (*a*) CONDITIONAL PROBABILITY

Let A and B be two events of a sample space S and let $P(B) \neq 0$. Then conditional probability of the event A, given B, denoted by $P(A/B)$, is defined by

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad \text{... (1)}$$

Theorem. If the events A and B defined on a sample space S of a random experiment are independent, then

$$P(A/B) = P(A) \text{ and } P(B/A) = P(B)$$

Proof. A and B are given to be independent events,

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

$$\Rightarrow \qquad P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

$$\Rightarrow \qquad P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B) \cdot P(A)}{P(A)} = P(B)$$

11.4 (*b*) BAYES THEOREM

If $B_1, B_2, B_3, \ldots B_n$ are mutually exclusive events with $P(B_i) \neq 0$, $(i = 1, 2, \ldots n)$ of a random experiment then for any arbitrary event A of the sample space of the above experiment with $P(A) > 0$, we have

$$P(B_i/A) = \frac{P(B_i)\, P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\, P(A/B_i)} \qquad \text{(for } n = 3)$$

$$P(B_2/A) = \frac{P(B_2)\, P(A/B_2)}{P(B_1)\, P(A/B_1) + P(B_2)\, P(A/B_2) + P(B_3)\, P(A/B_3)}$$

Proof. Let S be the sample space of the random experiment.

The events $B_1, B_2, \ldots, B_n$ being exhaustive

$$S = B_1 \cup B_2 \cup \ldots\ldots \cup B_n$$

$$\therefore \qquad A = A \cap S \qquad [\because\ A \subset S]$$

$$= A \cap (B_1 \cup B_2 \cup \ldots\ldots \cup B_n)$$

$$= (A \cap B_1) \cup (A \cap B_2) \cup \ldots\ldots \cup (A \cap B_n)$$

[Distributive Law]

$$\Rightarrow \qquad P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots\ldots + P(A \cap B_n)$$

$$= P(B_1)\, P(A/B_1) + P(B_2)\, P(A/B_2) + \ldots\ldots + P(B_n)\, P(A/B_n)$$

$$= \sum_{i=1}^{n} P(B_i)\, P(A/B_i) \qquad \ldots (1)$$

Now $\qquad P(A \cap B_i) = P(A)\, P(B_i/A)$

$$\Rightarrow \qquad P(B_i/A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(B_i)\, P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\, P(A/B_i)} \qquad \text{[Using (1)]}$$

Note. $P(B)$ is the probability of occurrence B. If we are told that the event A has already occurred. On knowing about the event A, $P(B)$ is changed to $P(B/A)$. With the help of Baye's theorem we can calculate $P(B/A)$.

Example (A) *An urn I contains 3 white and 4 red balls and an urn II contains 5 white and 6 red balls. One ball is drawn at random from one of the urns and is found to be white. Find the probability that it was drawn from urn I.*

Solution. Let $\quad U_1$: the ball is drawn from urn I

U_2 : the ball is drawn from urn II

W : the ball is white.

We have to find $P(U_1/W)$

By Baye's Theorem

$$P(U_1/W) = \frac{P(U_1)\,P(W/U_1)}{P(U_1)\,P(W/U_1) + P(U_2)\,P(W/U_2)} \quad \text{... (1)}$$

Since two urns are equally likely to be selected, $P(U_1) = P(U_2) = \frac{1}{2}$

$P(W/U_1) = P$ (a white ball is drawn from urn I) $= \frac{3}{7}$

$P(W/U_2) = P$ (a white ball is drawn from urn II) $= \frac{5}{11}$

$\therefore$ From (1) $\quad P(U_1/W) = \dfrac{\frac{1}{2}\times\frac{3}{7}}{\frac{1}{2}\times\frac{3}{7}+\frac{1}{2}\times\frac{5}{11}} = \dfrac{33}{68}$ **Ans.**

Example (B) *Three urns contains 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.*

Solution. Let U_1 : the ball is drawn from U_1.

U_2 : the ball is drawn from U_2.

U_3 : the ball is drawn from U_3.

R : the ball is red.

We have to find $P(U_1/R)$.

By Baye's Theorem,

$$P(U_1/R) = \frac{P(U_1)\,P(R/U_1)}{P(U_1)\,P(R/U_1) + P(U_2)\,P(R/U_2) + P(U_3)\,P(R/U_3)} \quad \text{... (1)}$$

Since the three urns are equally likely to be selected $P(U_1) = P(U_2) = P(U_3) = \frac{1}{3}$

Also $P(R/U_1) = P$ (a red ball is drawn from urn I) $= \frac{6}{10}$

$P(R/U_2) = P$ (a red ball is drawn from urn II) $= \frac{4}{10}$

$P(R/U_3) = P$ (a red ball is drawn from urn III) $= \frac{5}{10}$

$\therefore$ From (1), we have $P(U_1/R) = \dfrac{\frac{1}{3}\times\frac{6}{10}}{\frac{1}{3}\times\frac{6}{10}+\frac{1}{3}\times\frac{4}{10}+\frac{1}{3}\times\frac{5}{10}} = \dfrac{2}{5}$ **Ans.**

Example (C) *In a bolt factory, machines A, B and C manufacture respectively 25%, 35% and 40% of the total. If their output 5, 4 and 2 per cent 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?*

Solution. A : bolt is manufactured by machine A.

B : bolt is manufactured by machine B.

C : bolt is manufactured by machine C.

$$P(A) = 0.25, \quad P(B) = 0.35, \quad P(C) = 0.40$$

The probability of drawing a defective bolt manufactures by machine A is $P(D/A) = 0.05$

Similarly, $P(D/B) = 0.04$ and $P(D/C) = 0.02$

By Baye's theorem

$$P(B/D) = \frac{P(B)\,P(D/B)}{P(A)\,P(D/A) + P(B)\,P(D/B) + P(C)\,P(D/C)}$$

$$= \frac{0.35 \times 0.04}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02} = 0.41 \quad \textbf{Ans.}$$

BINOMIAL DISTRIBUTION

11.5 (a) DISCRETE PROBABILITY DISTRIBUTION

$P(X = x_i) = p\,i$ or $p(x_i)$ for $i = 1, 2, \ldots$ where (i) $p(x_i) \geq 0$ for all values of i,

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

The set of values x_i with their probabilities pi constitute a discreate probability distribution of the distrete variate X.

11.5 (b) BINOMIAL DISTRIBUTION P (r) = $^nC_r\, p^r . q^{n-r}$

To find the probability of the happening of an event once, twice, thrice, ... r times ...... exactly in n trials.

Let the probability of the happening of an event A in one trial be p and its probability of not happening be $1 - p = q$.

We assume that there are n trials and the happening of the event A is r times and its not happening is $n - r$ times.

This may be shown as follows

$$\underset{r \text{ times}}{A\,A \ldots\ldots A} \qquad \underset{n - r \text{ times}}{\overline{A}\,\overline{A} \ldots\ldots \overline{A}} \qquad \ldots(1)$$

A indicates its happening, $\overline{A}$ its failure and $P(A) = p$ and $P(\overline{A}) = q$.

We see that (1) has the probability

$$\underset{r \text{ times}}{pp \ldots\ldots p} \quad \underset{n - r \text{ times}}{q.q \ldots\ldots q} = p^r \,.\, q^{n-r} \qquad \ldots(2)$$

Clearly (1) is merely one order of arranging r A's.

The probability of (1) $= p^r q^{n-r} \times$ Number of different arrangements of r A's and $(n - r)\,\overline{A}$'s.

The number of different arrangements of r A's and $(n - r)\,\overline{A}$'s $= {^nC_r}$

$\therefore$ Probability of the happening of an event r times $= {^nC_r}\, p^r\, q^{n-r}$.

$= (r + 1)$th term of $(q + p)^n$ $\quad (r = 0, 1, 2, \ldots\ldots, n)$.

If $r = 0$, probability of happening of an event 0 times $= {^nC_0}\, q^n\, p^0 = q^n$

If $r = 1$, probability of happening of an event 1 time $= {^nC_1}\, q^{n-1}\, p$

If $r = 2$, probability of happening of an event 2 times $= {^nC_2}\, q^{n-2}\, p^2$

If $r = 3$, probability of happening of an event 3 times $= {^nC_3}\, q^{n-3}\, p^3$ and so on.

These terms are clearly the successive terms in the expansion of $(q + p)^n$.

Hence it is called Binomial Distribution.

Example 24. *Find the probability of getting 4 heads in 6 tosses of a fair coin.*

Solution. $p = \frac{1}{2},\ q = \frac{1}{2},\ n = 6,\ r = 4.$

We know that $P(r) = {^nC_r}\, q^{n-r}\, p^r \Rightarrow P(4) = {^6C_4}\, q^{6-4}\, p^4$

$$= \frac{6 \times 5}{1 \times 2}\left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 = 15 \times \left(\frac{1}{2}\right)^6 = \frac{15}{64} \quad \textbf{Ans.}$$

Example 25. *If on an average one ship in every ten is wrecked, find the probability that out of 5 ships expected to arrive, 4 at least will arrive safely.*

Solution. Out of 10 ships, one ship is wrecked.

i.e., Nine ships out of ten ships are safe. P (safety) $= \dfrac{9}{10}$

P (At least 4 ships out of 5 are safe) $= P(4 \text{ or } 5) = P(4) + P(5)$

$$= {}^5C_4\, p^4 q^{5-4} + {}^5C_5\, p^5 q^0 = 5\left(\frac{9}{10}\right)^4\left(\frac{1}{10}\right) + \left(\frac{9}{10}\right)^5 = \left(\frac{9}{10}\right)^4\left(\frac{5}{10}+\frac{9}{10}\right) = \frac{7}{5}\left(\frac{9}{10}\right)^4 \textbf{ Ans.}$$

Example 26. *The overall percentage of failures in a certain examination is 20. If six candidates appear in the examination, what is the probability that at least five pass the examination?*

Solution. Probability of failures $= 20\% = \dfrac{20}{100} = \dfrac{1}{5}$

Probability of pass $(P) = 1 - \dfrac{1}{5} = \dfrac{4}{5}$

Probability of at least five pass $= P(5 \text{ or } 6)$

$$= P(5) + P(6) = {}^6C_5\, p^5 q + {}^6C_6\, p^6 q^0$$

$$= 6\left(\frac{4}{5}\right)^5\left(\frac{1}{5}\right) + \left(\frac{4}{5}\right)^6 = \left(\frac{4}{5}\right)^5\left[\frac{6}{5}+\frac{4}{5}\right] = 2\left(\frac{4}{5}\right)^5 = \frac{2048}{3125} = 0.65536 \qquad \textbf{Ans.}$$

Example 27. *Ten percent of screws produced in a certain factory turn out to be defective. Find the probability that in a sample of 10 screws chosen at random, exactly two will be defective.*

Solution. $p = \dfrac{1}{10},\ q = \dfrac{9}{10},\ n = 10,\ r = 2\ P(r) = {}^nC_r\, p^r q^{n-r}$

$$P(2) = {}^{10}C_2\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^{10-2} = \frac{10\times 9}{1\times 2}\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^8 = \frac{1}{2}\cdot\left(\frac{9}{10}\right)^9 = 0.1937 \textbf{ Ans.}$$

Example 28. *The probability that a man aged 60 will live to be 70 is 0.65. What is the probability that out of 10 men, now 60, at least 7 will live to be 70 ?*

Solution. The probability that a man aged 60 will live to be 70

$$= p = 0.65$$

$$q = 1 - p = 1 - 0.65 = 0.35$$

Number of men $= n = 10$

Probability that at least 7 men will live to 70 = (7 or 8 or 9 or 10)

$$= P(7) + P(8) + P(9) + P(10) = {}^{10}C_7\, q^3 p^7 + {}^{10}C_8\, q^2 p^8 + {}^{10}C_9\, q\, p^9 + p^{10}$$

$$= \frac{10\times 9\times 8}{1\times 2\times 3}(.35)^3\,(.65)^7 + \frac{10\times 9}{1\times 2}(.35)^2\,(.65)^8 + 10\,(.35)\,(.65)^9 + (.65)^{10}$$

$$= (.65)^7\,[120\,(.35)^3 + 45\,(.35)^2\,(.65) + 10\,(.35)\,(.65)^2 + (.65)^3]$$

$$= (.65)^7 \times 125\,[120\times(.07)^3 + 45\times(.07)^2\,(.13) + 10\,(.07)\,(.13)^2 + (.13)^3]$$

$$= 0.04901 \times 125\,[0.04116 + 0.028665 + 0.011830 + .002197]$$

$$= 6.12625 \times 0.083852 = 0.5137 \qquad \textbf{Ans.}$$

Example 29. *If 10% of bolts produced by a machine are defective. Determine the probability that out of 10 bolts, chosen at random (i) 1 (ii) none (iii) at most 2 bolts will be defective.*

Solution. Probability of defective bolts = p = 10% = 0.1

Probability of not defective bolts = $q = 1 - p = 1 - 0.1 = 0.9$

Total number of bolts = n = 10

(i) Probability of 1 defective bolt = $^{10}C_1\,(0.1)^1\,(0.9)^9 = 0.3874$

(ii) Probability that none is defective = Probability of 0 defective bolt

$$= P(0) = {}^{10}C_0\,(0.1)^0\,(0.9)^{10} = .3487$$

(iii) Probability of 2 defective = $^{10}C_2\,(0.1)^2\,(0.9)^8 = 0.1937$

Probability of at most 2 defective = P (0 or 1 or 2)

$$= P(0) + P(1) + P(2) = 0.3487 + 0.3874 + 0.1937$$

$$= 0.9298$$ **Ans.**

Example 30. *A die is thrown 8 times and it is required to find the probability that 3 will show (i) Exactly 2 times (ii) At least seven times (iii) At least once.*

Solution. The probability of throwing 3 in a single trial = $p = \frac{1}{6}$

The probability of not throwing 3 in a single trial = $q = \frac{5}{6}$

(i) P (getting 3, exactly 2 times)

$$= {}^8C_2\,q^6p^2 = 28\left(\frac{5}{6}\right)^6\left(\frac{1}{6}\right)^2 = \frac{28\times 5^6}{6^8}$$

(ii) P (getting 3, at least seven times) = P (getting 3, at 7 or 8 times)

$$= P(7) + P(8) = {}^8C_7q^1p^7 + {}^8C_8\,q^0p^8 = 8\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)^7 + \left(\frac{1}{6}\right)^8 = \frac{41}{6^8}$$

(iii) P (getting 3 at least once)

$= P$ (getting 3, at 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 times)

$= P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7) + P(8)$

$= 1 - P(0) = 1 - {}^8C_0\,q^8\,p^0$

$= 1 - \left(\frac{5}{6}\right)^8$ **Ans.**

Example 31. *An underground mine has 5 pumps installed for pumping out storm water, the probability of any one of the pumps failing during the storm is* $\frac{1}{8}$*. What is the probability that* (i) *at least 2 pumps will be working;* (ii) *all the pumps will be working during a particular storm?*

Solution. (i) Probability of pump failing = $\frac{1}{8}$

Probability of pump working = $1 - \frac{1}{8} = \frac{7}{8}$, $p = \frac{7}{8}$, $q = \frac{1}{8}$, $n = 5$

(i) P (At least 2 pumps working) = P (2 or 3 or 4 or 5 pumps working)

$$= P(2) + P(3) + P(4) + P(5)$$

$$= 5_{C_2}\,p^2\,q^3 + 5_{C_3}\,p^3q^2 + 5_{C_4}\,p^4q + 5_{C_5}\,p^5q^0$$

$$= 10\left(\frac{7}{8}\right)^2\left(\frac{1}{8}\right)^3 + 10\left(\frac{7}{8}\right)^3\left(\frac{1}{8}\right)^2 + 5\left(\frac{7}{8}\right)^4\left(\frac{1}{8}\right) + \left(\frac{7}{8}\right)^5$$

$$= \frac{1}{8^5}[10\times 49 + 10\times 343 + 5\times 2401 + 16807]$$

$$= \frac{1}{8^5}[490 + 3430 + 12005 + 16807] = \frac{32732}{8^5} = \frac{8183}{8192}$$

(*ii*) P (All the 5 pumps working) $= P(5) = {}^5C_5\, p^5 q^0 = \left(\frac{7}{8}\right)^5 = \frac{16807}{32768}$

Ans. (*i*) $\frac{8183}{8192}$ (*ii*) $\frac{16807}{32768}$

Example 32. *Assuming that 20% of the population of a city are literate, so that the chance of an individual being literate is* $\frac{1}{5}$ *and assuming that 100 investigators each take 10 individuals to see whether they are literate, how many investigators would you expect to report 3 or less were literate.* (*A.M.I.E.T.E., Summer 2000*)

Solution. $p = \frac{1}{5},\quad n = 10$

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

$$= {}^{10}C_0\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{10} + {}^{10}C_1\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^9 + {}^{10}C_2\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^8 + {}^{10}C_3\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^7$$

$$= \left(\frac{4}{5}\right)^{10} + \frac{10}{5}\left(\frac{4}{5}\right)^9 + \frac{45}{25}\left(\frac{4}{5}\right)^8 + \frac{120}{125}\left(\frac{4}{5}\right)^7$$

$$= \left(\frac{4}{5}\right)^7 [(0.8)^3 + 2\,(0.8)^2 + 1.8\,(0.8) + 0.96]$$

$$= 0.2097152\,[0.512 + 1.28 + 1.44 + 0.96] = 0.2097152 \times 4.192 = 0.879126118$$

Required number of investigators $= 0.879126118 \times 100 = 87.9126118$

$= 88$ approximate **Ans.**

Example 33. *Assuming half the population of a town consumes chocolates and that 100 investigators each take 10 individuals to see whether they are consumers, how many investigators would you expect to report that three people or less were consumers?* (*A.M.I.E., Winter 1999*)

Solution. The chance for an individual to be consumer is $P = \frac{1}{2}$

The chance of not being a consumer $= q = 1 - \frac{1}{2} = \frac{1}{2}$.

Here we have to find the probabilities of 0, 1, 2 and 3 successes.

$P(r \le 3) = P(0) + P(1) + P(2) + P(3) = q^\circ + {}^{10}C_1\, q^9 p^1 + {}^{10}C_2\, q^8 p^2 + {}^{10}C_3\, q^7 p^3$

$$= \left(\frac{1}{2}\right)^{10} + 10\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right) + 45\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + 120\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3$$

$$= \left(\frac{1}{2}\right)^{10}[1 + 10 + 45 + 120] = \frac{176}{1024}$$

The number of investigators to report that three or less people were consumers of chocolates is given by

$$\frac{176}{1024} \times 100 = 17.2$$

Hence, 17 investigators would report that 3 or less people are consumers. **Ans.**

Example 34. *For special security in a certain protected area, it was decided to put three lighting bulbs on each pole. If each bulb has a probability p of burning out in the first 100 hours of service, calculate the probability that at least one of them is still good after 100 hours.*

If p = 0.3, how many bulbs would be needed on each pole to ensure 99% safety so that at least one is good after 100 hours ?

Solution. Probability of burning out in the first 100 hours of service $= p = 0.3$

Probability of working good in the first 100 hours

$$q = 1 - p = 1 - .3 = .7$$

(*i*) Probability that at least one of them is still good after 100 hours

$$= 3C_1\, q\, p^2 + 3C_2\, q^2\, p^1 + 3C_3\, q^3 p^0$$

$$= \left[3C_0\, q^0\, p^3 + 3C_1\, q\, p^2 + 3C_2\, q^2\, p + {}^3C_3\, q^3 p^0 \right] - 3C_0\, q^0\, p^3$$

$$= 1 - p^3 = 1 - (.3)^3 = 1 - 0.027 = 0.973 \qquad \textbf{Ans.}$$

Let the number of bulbs required be n.

$$P \text{ (At least one bulb is good)} = 1 - p^n$$

$$0.99 = 1 - (0.3)^n \text{ or } (0.3)^n = 1 - 0.99$$

$$(0.3)^n = 0.01 \text{ or } \log (0.3)^n = \log 0.01$$

$$n \log 0.3 = \log 0.01 \text{ or } n = \frac{\log 0.01}{\log 0.3}$$

$$n = \frac{\bar{2}.000}{\bar{1}.477} = \frac{-2.000}{-0.523} = 3.8, \simeq 4 \text{ Bulbs} \qquad \textbf{Ans.}$$

Exercise 11.3

1. If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 bolts chosen at random

(*a*) 1 (*b*) 0 (*c*) At most 2

bolts will be defective. **Ans.** (*a*) 0.4096, (*b*) 0.4096, (*c*) 0.9728.

2. Six dice are thrown 729 times. How many times do you expect at least three dice to show a five or a six ? **Ans.** 233

3. If the chance that any one of the 10 telephone lines is busy at any instant is 0.2, what is the chance that 5 of the lines are busy ? What is the probability that all the lines are busy ?

Ans. $10C_5\,(0.2)^5\,(0.8)^5, (0.2)^{10}$

4. An insurance salesman sells policies to 5 men, all of identical age in good health. According to the actuarial tables the probability that a man of this particular age will be alive 30 years hence is $\frac{2}{3}$.

Find the probability that in 30 years.

(*a*) All 5 men (*b*) At least 3 men (*c*) Only 2 men (*d*) At least 1 man

will be alive. **Ans.** (*a*) $\frac{32}{243}$ (*b*) $\frac{192}{243}$ (*c*) $\frac{40}{243}$ (*d*) $\frac{242}{243}$

5. Assuming a Binomial distribution, find the probability of obtaining at least two "six" in rolling a fair die 4 times. **Ans.** $\frac{171}{1296}$

6. If successive trials are independent and the probability of success on any trial is p, show that the probability that the first success occurs on the nth trial is

$$p\,(1 - p)^{n-1}, \qquad n = 1, 2, 3, \ldots$$

7. Consider an urn in which 4 balls have been placed by the following scheme : A fair coin is tossed; if the coin falls head, a white ball is placed in the urn, and if the coin falls tail, a red ball is placed in urn. (*i*) What is the probability that the urn will contain exactly 3 white balls ? (*ii*) What is the probability that the urn will contain exactly 3 red balls, given that the first ball placed was red?

Ans. (*i*) $\frac{1}{8}$, (*ii*) $\frac{3}{8}$.

8. A box contains 10 screws, 3 of which are defective. Two screws are drawn at random without replacement. Find the probability that none of the two screws is defective. **Ans.** $\frac{7}{15}$

9. Out of 800 families with four children each, how many families would be expected to have :

(*i*) 2 boys and 2 girls; (*ii*) at least one boy; (*iii*) no girl; (*iv*) at most two girls ?

Assume equal probabilities for boys and girls. **Ans.** (*i*) 300, (*ii*) 750, (*iii*) 50, (*iv*) 550.

10. In a hurdle race, a player has to cross 10 hurdles. The probability that he will clear each hurdle is 5/6. What is the probability that he will knock down less than 2 hurdles ? **Ans.** $\frac{8}{3}\left(\frac{5}{6}\right)^9$

11. An electronic component consists of three parts. Each part has probability 0.99 of performing satisfactorily. The component fails if 2 or more parts do not perform satisfactorily. Assuming that the parts perform independently, determine the probability that the component does not perform satisfactorily. **Ans.** 0.000298

12. Find the binomial distribution whose mean is 5 and variance is 10/3. **Ans.** ${}^{15}C_r\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{15-r}$

13. The probability that on, joining Engineering College, a student will successfully complete the course of studies is $\frac{3}{5}$. Determine the probability that out of 5 students joining the College (*i*) none and (*ii*) at least two will successfully complete the course. **Ans.** (*i*) $\frac{32}{3125}$ (*ii*) $\frac{2853}{3125}$

14. A carton contains 20 fuses, 5 of which are defective. Three fuses are chosen at random and inspected. What is the probability that at most one defective fuse is found. **Ans.** $\frac{27}{32}$

15. A bag contains three coins, one of which is coined with two heads, while the other two coins are normal and not biased. A coin is thrown at random from the bag and tossed three times in succession. If heads turn up each time, what is the probability that this is the two-headed coin? **Ans.** $\frac{4}{5}$

16. 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 1,000 such samples, how many would be expected to contain at least 3 defective parts? **Ans.** 324

17. The incidence of occupational disease in an industry is such that the workers have 20% chance of suffering from it. What is the probability that out of 6 workers 4 or more will catch the disease ?

Ans. $\frac{53}{3125}$

18. If the probability of hitting a target is 10% and 10 shots are fired independently, what is the probability that the target will be hit at least once ? **Ans.** 1 – (0.9)10 = .65 nearly

19. Among 10,000 random digits, find the probability p that the digit 3 appears at most 950 times.

(*A.M.I.E., Summer 2003*) **Ans.** $e^{-100}\sum_{r=0}^{r=950}\left(\frac{(10000)\,r}{r\,!}\right)$

20. A fair coin is tossed 400 times. Using normal approximation to the binomial, find the probability that a head will occur (*a*) more than 180 times and (*b*) less than 195 times. (*A.M.I.E. Winter 2004*)

21. Four coins were tossed 200 times. The number of tosses showing 0, 1, 2, 3 and 4 heads were found to be as under. Fit a binomial distribution to these observed results. Find the expected frequencies.

No. of heads:	0	1	2	3	4
No. of tosses:	15	35	90	40	20

22. A firm plans to bid Rs. 300 per tonne for a contract to supply 1000 tonnes of a metal. It has two competitors A and B and it assumes that the probability that A will bid less than 300/- per tonne is 0.3 and that B will bid less than Rs. 300 per tonne is 0.7. If the lowest bidder gets all the business and the firms bid independently, what is the expected value of business in rupees to the firm.

(*A.M.I.E.T.E., Dec. 2006*)

11.6 MEAN OF BINOMIAL DISTRIBUTION (*A.M.I.E.T.E., Dec. 2006, Summer 2000*)

$$(q+p)^n = q^n + {}^nC_1\, q^{n-1} p^1 + {}^nC_2\, q^{n-2} p^2 + {}^nC_3 q^{n-3} p^3 + \ldots\ldots + {}^nC_r\, q^{n-r} p^r + \ldots\ldots + p^n$$

Successes r	*Frequency* f	*Product* rf
0	q^n	0
1	$n\,q^{n-1}\,p$	$n\,q^{n-1}\,p$
2	$\frac{n\,(n-1)}{2}\; q^{n-2}\,p^2$	
3	$\frac{n\,(n-1)\,(n-2)}{6}\; q^{n-3}\,p^3$	
....	....	....
n	p^n	np^n

$$\Sigma fr = n\,q^{n-1}\; p + n\,(n-1)\,q^{n-2}\,p^2 + \frac{n\,(n-1)\,(n-2)}{2}\; q^{n-3}\,p^3 + \ldots\ldots + np^n$$

$$= np\,[q^{n-1} + \frac{(n-1)}{1\,!}\; q^{n-2}\,p + \frac{(n-1)\,(n-2)}{2}\; q^{n-3}\,p^2 + \ldots + p^{n-1}]$$

$$= np\,(q+p)^{n-1} = n\,p \qquad \text{(since } q+p=1\text{)}$$

$$\Sigma f = q^n + nq^{n-1}\,p + \frac{n\,(n-1)}{2}\; q^{n-2}\,p^2 + \ldots\ldots + p^n$$

$$= (q+p)^n = 1 \qquad \text{since } q+p=1$$

Hence $\qquad \text{Mean} = \frac{\Sigma f r}{\Sigma f} = np \qquad$ **Ans.**

11.7 STANDARD DEVIATION OF BINOMIAL DISTRIBUTION (*A.M.I.E.T.E., Dec.2006*)

Successes r	*Frequency* f	*Product* $r^2 f$
0	q^n	0
1	$n\,q^{n-1}\,p$	$n\,q^{n-1}\,p$
2	$\frac{n\,(n-1)}{2}\; q^{n-2}\,p^2$	
3	$\frac{n\,(n-1)\,(n-2)}{6}\; q^{n-3}\,p^3$	
....	....	....
n	p^n	n^2p^n

We know that $\qquad \sigma^2 = \frac{\Sigma f r^2}{\Sigma f} - \left(\frac{\Sigma fr}{\Sigma f}\right)^2 \qquad$ ...(1)

r is the deviation of items (successes) from 0.

$$\Sigma f = 1,\; \Sigma f r = np$$

$$\Sigma f r^2 = 0 + nq^{n-1}\,p + 2n\,(n-1)\,q^{n-2}\,p^2 + \frac{3n\,(n-1)\,(n-2)}{2}\; q^{n-3}\,p^3 + \ldots\ldots + n^2p^n$$

$$= np\,[\,q^{n-1} + \frac{2\,(n-1)}{1\,!} q^{n-2}\,p + \frac{3\,(n-1)\,(n-2)}{2\,!}\; q^{n-3}\,p^2 + \ldots\ldots + n\,p^{n-1}\,]$$

$$= np\Big[q^{n-1} + \frac{(n-1)\,q^{n-2}\,p}{1\,!} + \frac{(n-1)(n-2)}{2\,!}\,q^{n-3}\,p^2 + \ldots.. + p^{n-1}$$

$$+ \frac{(n-1)\,q^{n-2}\,p}{1\,!} + \frac{2\,(n-1)(n-2)}{2\,!}\,q^{n-3}\,p^2 + \ldots.. + (n-1)\,p^{n-1}\Big]$$

$$= np\Big[q^{n-1} + (n-1)\,q^{n-2}\,p + \frac{(n-1)(n-2)}{2\,!}\,q^{n-3}\,p^2 + \ldots... + p^{n-1}$$

$$+ (n-1)\,p\left\{q^{n-2} + (n-2)\,q^{n-3}\,p + \frac{(n-2)(n-3)}{2\,!}\,q^{n-4}\,p^2 + \ldots + p^{n-2}\right\}\Big]$$

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

$$= np\,[np + (1-p)] = np\,[np + q] = n^2p^2 + npq$$

Putting these values in (1), we have

$$\text{Variance} = \sigma^2 = \frac{n^2p^2 + n\,pq}{1} - \left(\frac{np}{1}\right)^2 = npq,$$

$$S.D. = \sigma = \sqrt{n\,p\,q}$$

Hence, for the binomial distribution,

$$\text{Mean} = np,\quad \mu_2 = \sigma^2 = n\,p\,q$$

Example 35. *Find the first four moments of the binomial distribution.*

(*A.M.I.E., Summer 2000*)

Solution. First moment about the origin

$$\mu_1' = \sum_{r=0}^{n} {}^nC_r\,p^r\,q^{n-r}\cdot r = \sum_{r=0}^{n} r\cdot\frac{n\,(n-1)(n-2)\ldots(n-r+1)}{r\,!}\,p^r\,q^{n-r}$$

$$= n\sum_{r=1}^{n}\frac{(n-1)(n-2)\ldots(n-r+1)}{(n-r)!}\,p^r\,q^{n-r}$$

$$= np\sum_{r=1}^{n} {}^{n-1}C_{r-1}\,p^{r-1}\,q^{n-r} = np\,(q+p)^{n-1} = np$$

Thus, the mean of the Binomial distribution is np.

Second moment about the origin

$$\mu_2' = \sum_{r=0}^{n} {}^nC_r\,p^r\,q^{n-r}\cdot r^2 \qquad [r^2 = r\,(r-1) + r]$$

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

$$= \sum_{r=0}^{n}\frac{r\,(r-1)\,n\,(n-1)(n-2)\ldots(n-r+1)}{r\,!}\,p^r\,q^{n-r}$$

$$\sum_{r=0}^{n}\frac{rn\,(n-1)(n-2)\ldots(n-r+1)}{r\,!}\,p^r\,q^{n-r}$$

$$= n\,(n-1)\,p^2\sum_{r=2}^{n}\frac{(n-2)(n-3)\ldots(n-r+1)}{(r-2)\,!}\,p^{r-2}\,q^{n-r}$$

$$+ np\,\sum\frac{(n-1)(n-2)\ldots(n-r+1)}{(r-1)\,!}\,p^{r-1}\,q^{n-r}$$

$$= n\,(n-1)\,p^2\,(q+p)^{n-2} + np\,(q+p)^{n-1} = n\,(n-1)\,p^2 + np$$

Third moment about the origin

$$\mu_3' = \sum_{r=0}^{n} {}^nC_r\,p^r\,q^{n-r}\cdot r^3$$

[Let $r^3 = A\,r(r-1)(r-2) + Br\,(r-1) + Cr$

By putting r = 1,2, 3, we get A = 1 B = 3 C = 1]

$$\mu_3' = \sum_{r=0}^{n} \left\{r(r-1)(r-2) + 3r(r-1) + r\right\} {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} r(r-1)(r-2)\, {}^nC_r\, p^r q^{n-r} + 3\sum_{r=0}^{n} r(r-1)\cdot {}^nC_r\, p^r q^{n-r} + \sum_{r=0}^{n} r\cdot {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} \frac{r(r-1)(r-2)\cdot n(n-1)\ldots(n-r+1)}{r!}\, p^r q^{n-r}$$

$$+ 3\sum_{r=0}^{n} \frac{r(r-1)\, n\cdot(n-1)\ldots(n-r+1)}{r!}\, p^r q^{n-r} + \sum_{r=0}^{n} r\frac{n(n-1)\ldots(n-r+1)}{r!}\, p^r q^{n-r}$$

$$= \sum_{r=3}^{n} \frac{n(n-1)(n-2)(n-3)\ldots(n-r+1)}{(r-3)!}\, p^r q^{n-r}$$

$$+ 3\sum_{r=2}^{n} \frac{n(n-1)(n-2)(n-3)\ldots(n-r+1)}{(r-2)!}\, p^r q^{n-r}$$

$$+ \sum_{r=1}^{n} \frac{n(n-1)(n-2)\ldots(n-r+1)}{(r-1)!}\, p^r q^{n-r}$$

$$= n(n-1)(n-2)\, p^3 \sum_{r=3}^{n} {}^{n-3}C_{r-3}\, p^{r-3} q^{n-3} + 3n(n-1)\, p^2 \sum_{r=2}^{n} {}^{n-2}C_{r-2}\, p^{r-2} q^{n-2}$$

$$+ np \sum_{r=1}^{n} {}^{(n-1)}C_{r-1}\, p^{r-1} q^{n-1}$$

$$= n(n-1)(n-2)\, p^3 (q+p)^{n-3} + 3n(n-1)\, p^2 (q+p)^{n-2} + np\,(q+p)^{n-1}$$

$$= n(n-1)(n-2)\, p^3 + 3n(n-1)\, p^2 + np$$

Fourth Moment

$$\mu_4' = \sum_{r=0}^{n} {}^nC_r\, p^r q^{n-r}\cdot r^4$$

[Let $r^4 = A\,r(r-1)(r-2)(r-3) + B\,r(r-1)(r-2) + Cr(r-1) + Dr$

By putting r = 1, 2, 3, 4 we get A = 1, B = 6, C = 7, D = 1]

$$\mu_4' = \sum_{r=0}^{n} r(r-1)(r-2)(r-3)\cdot {}^nC_r\, p^r q^{n-r} + \sum_{r=0}^{n} 6r(r-1)(r-2)\cdot {}^nC_r\, p^r q^{n-r}$$

$$+ \sum_{r=0}^{n} 7r(r-1)\cdot {}^nC_r\, p^r q^{n-r} + \sum_{r=0}^{n} r\cdot {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} \frac{r(r-1)(r-2)(r-3)\cdot n(n-1)\ldots(n-r+1)}{r!}\, p^r q^{n-r}$$

$$+ 6\sum_{r=0}^{n} \frac{r(r-1)(r-2)\cdot n(n-1)\ldots(n-r+1)}{r!}\, p^r q^{n-r}$$

$$+ 7\sum_{r=0}^{n} \frac{r(r-1)\cdot n(n-1)\ldots(n-r+1)}{r!}\, p^r q^{n-r} + \sum_{r=0}^{n} \frac{r\cdot n(n-1)\ldots(n-r+1)}{r!}\, p^r q^{n-r}$$

$$= \sum_{r=4}^{n} \frac{n(n-1)(n-2)(n-3)(n-4)\ldots(n-r+1)}{(r-4)!} p^r q^{n-r}$$

$$+ 6\sum_{r=3}^{n} \frac{n(n-1)(n-2)(n-3)\ldots(n-r+1)}{(r-3)!} p^r q^{n-r}$$

$$+ 7\sum_{r=2}^{n} \frac{n(n-1)(n-2)\ldots(n-r+1)}{(r-2)!} p^r q^{n-r} + \sum_{r=1}^{n} \frac{n(n-1)\ldots(n-r+1)}{(r-1)!} p^r q^{n-r}$$

$$= n(n-1)(n-2)(n-3) \sum_{r=4}^{n} {}^{n-4}C_{r-4}\, p^r q^{n-r} + 6n(n-1)(n-2) \sum_{r=3}^{n} {}^{n-3}C_{r-3}\, p^r q^{n-r}$$

$$+ 7n(n-1) \sum_{r=2}^{n} {}^{n-2}C_{r-2}\, p^r q^{n-r} + n \sum_{r=1}^{n} {}^{n-1}C_{r-1} \cdot p^r q^{n-r}$$

$$= n(n-1)(n-2)(n-3)\, p^r (q+p)^{n-4} + 6n(n-1)(n-2)\, p^3 (q+r)^{n-3}$$

$$+ 7n(n-1)\, p^2 (q+r)^{n-2} + np\,(q+r)^{n-1}$$

$$= n(n-1)(n-2)(n-3)\, p^4 + 6n(n-1)(n-2)\, p^3 + 7n(n-1)\, p^2 + np$$

11.8 CENTRAL MOMENTS : (Moments about the mean)

Now, the first four central moments are obtained as follows:

Second Central Moment

$\mu_2 = \mu_2' - \mu_1' = [n(n-1)p^2 + np] - n^2 p^2 = np[(n-1)p + 1 - np] = np(1-p) = npq$

Variance of Binomial distribution is npq

Third Central Moment

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

$$= \{n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np\} - 3\{(n^2p^2 + npq)\,np\} + 2n^3p^3$$

$$= np[-3np^2 + 3np + 2p^2 - 3p + 1 - 3npq]$$

$$= np[3np(1-p) + 2p^2 - 3p + 1 - 3npq]$$

$$= np[3npq + 2p^2 - 3p + 1 - 3npq] = np[2p^2 - 3p + 1] = np[2p^2 - 2p + q]$$

$$= np[-2p(1-p) + q] = np(-2pq + q) = npq(1-2p) = npq(q-p)$$

Fourth Central Moment

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

$$= n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2$$
$$+ np - 4[n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np]\,np$$
$$+ 6[n(n-1)p^2 + np]\,n^2p^2 - 3n^4p^4$$

$$= np\Big[(n-1)(n-2)(n-3)p^3 + 6(n-1)(n-2)p^2 + 7(n-1)p$$
$$+ 1 - 4\{n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np\}$$
$$+ 6\{n(n-1)p^2 + np\}\,np - 3n^3p^3\Big]$$

$$= np\{(n^3 - 6n^2 + 11n - 6)p^3 + (6n^2 - 18n + 12)p^2 + 7np - 7p + 1\}$$
$$+ \{(-4n^3 + 12n^2 - 8n)p^3 - 4(3n^2 - 3n)p^2 - 4np\}$$
$$+ \{(6n^3 - 6n^2)p^3 + 6n^2p^2\} - 3n^3p^3]$$

$$= np\Big[(n^3 - 6n^2 + 11n - 6 - 4n^3 + 12n^2 - 8n + 6n^3 - 6n^2 - 3n^3)p^3$$
$$+ (6n^2 - 18n + 12 - 12n^2 + 12n + 6n^2)p^2 + (7n - 7 - 4n)p + 1\Big]$$

$$= np\Big[(3n-6)p^3 + (-6n+12)p^2 + (3n-7)p + 1\Big]$$

$$= np[3np^3 - 6p^3 - 6np^2 + 12p^2 + 3np - 7p + 1]$$

$$= np\,[3np^3 - 3np^2 - 6p^3 + 6p^2 - 3np^2 + 3np + 6p^2 - 6p - p + 1]$$
$$= np\,[-3np^2(1-p) + 6p^2(1-p) + 3np(1-p) - 6p(1-p) + (1-p)]$$
$$= np[-3np^2q + 6p^2q + 3npq - 6pq + q] = npq\,[-3np^2 + 6p^2 + 3np - 6p + 1]$$
$$= npq\,[3np(1-p) - 6p(1-p) + 1] = npq\,[3npq - 6pq + 1]$$
$$= npq\,[1 + 3(n-2)pq] \qquad \textbf{Ans.}$$

11.9 MOMENT GENERATING FUNCTIONS OF BINOMIAL DISTRIBUTION ABOUT ORIGIN

$$M_0(t) = E(e^{tx}) = \Sigma\; {}^nC_x\, p^x q^{n-x} \cdot e^{tx}$$
$$= \Sigma\; {}^nC_x\,(p\,e^t)^x q^{n-x} = (q + p\,e^t)^n$$

Differentiating w.r.t. 't', we get $M_0'(t) = n\,(q + p\,e^t)^{n-1}\,pe^t$

On putting $t = 0$, we get

$$\mu_1' = n\,(q+p)^{n-1}\;p$$
$$\mu_1' = np$$

Since, $\quad M_a(t) = e^{-at} M_0(t)$

Moment generating function of the Binomial distribution about its mean $(m) = np$ is given by

$$M_m(t) = e^{-npt} M_0(t)$$

$$M_m(t) = e^{-npt}(q + p\,e^t)^n = (q\,e^{e^{-pt}} + p\,e^{-pt+t})^n = (q\,e^{-pt} + pe^{(1-p)t]} = (qe^{-pt} + pe^{qt})$$

$$= \left[q\left(1 - pt + \frac{p^2t^2}{2!} - \frac{p^3t^3}{3!} + \frac{p^4t^4}{4!} + \ldots\right) + p\left(1 + qt + \frac{q^2t^2}{2!} + \frac{q^3t^3}{3\,!} + \frac{q^4t^4}{4!} + \ldots\right)\right]^n$$

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

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

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

$$1 + \mu_1 t + \mu_2\frac{t^2}{2!} + \mu_3\frac{t^3}{3!} + \mu_4\frac{t^4}{4!} + \ldots$$
$$= 1 + npq\frac{t^2}{2!} + npq\,(q-p)\frac{t^3}{3!} + npq\,[1 + 3(n-2)pq]\frac{t^4}{4!} + \ldots$$

Equating the coefficients of like powers of t on both sides, we get

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

Hence the moment coefficient of skewness is

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{[\,npq\,(q-p)\,]^2}{(npq)^3} = \frac{(q-p)^2}{npq}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}}$$

Coefficient of Kurtosis is given by

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{npq\,[1 + 3\,pq\,(n-2)]}{(npq)^2} = 3 + \frac{1 - 6\,pq}{npq}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1 - 6\,pq}{npq}$$

Example 36. *If the probability of a defective bolt is 0.1, find*
(a) the mean (b) the standard deviation for the distribution bolts in a total of 400.

Solution. $n = 400,\ p = 0.1,$ Mean $= np = 400 \times 0.1 = 40$

Standard deviation $= \sqrt{npq} = \sqrt{400 \times 0.1 \times (1 - 0.1)}$

$$= \sqrt{400 \times 0.1 \times 0.9} = 20 \times .3 = 6 \qquad \textbf{Ans.}$$

Example 37. *A die is tossed thrice. A success is getting 1 or 6 on a toss. Find the mean and variance of the number of successes.*

Solution. $n = 3,\ p = \dfrac{1}{3},\ q = \dfrac{2}{3}$

$$\text{Mean} = np = 3 \times \frac{1}{3} = 1$$

$$\text{Variance} = npq = 3 \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{3} \qquad \textbf{Ans.}$$

11.10 RECURRENCE RELATION FOR THE BINOMIAL DISTRIBUTION

By Binomial distribution, $P(r) = {}^nC_r\, p^r q^{n-r}$ ...(1)

$$P(r+1) = {}^nC_{r+1}\, p^{r+1} q^{n-r-1} \qquad \text{...(2)}$$

On dividing (2) by (1), we get

$$\frac{P(r+1)}{P(r)} = \frac{{}^nC_{r+1}}{{}^nC_r}\,\frac{p^{r+1} q^{n-r-1}}{p^r q^{n-r}} = \frac{n(n-1)(n-2)\ldots(n-r)}{\lfloor r+1} \cdot \frac{\lfloor r}{n(n-1)(n-2)\ldots(n-r+1)}\,\frac{p}{q}$$

$$\frac{P(r+1)}{P(r)} = \frac{n-r}{r+1}\,\frac{p}{q} \quad \text{or} \quad P(r+1) = \frac{n-r}{r+1}\,\frac{p}{q} \cdot P(r) \qquad \textbf{Ans.}$$

Exercise 11.4

1. Fit a binomial distribution to the following frequency data :

x	0	1	3	4
f	28	62	10	4

(*UP III Sem. Dec. 2004*) **Ans.** $P(r) = {}^{104}C_r\,(0.00999)^r\,(0.99111)^{104-r}$

2. Fill in the blanks :

(*a*) A coin is biased so that a head is twice as likely to occur as a tail. If the coin is tossed 3 times, the prob. of getting exactly 2 tails, is ....

(*b*) The probability of getting number 5 exactly two times in five throws of an unbiased die is ....

(*c*) A die is thrown 6 times. The probability to get greater than 4 appears at least once is ....

(*d*) For what, one should be?

(*i*) Obtaining 6 at least once in 4 throws of a die.

or (*ii*) obtaining a double-six at least once in 24 throws with two dice.

(*e*) The probability of producing a defective bolt is 0.1. The probability that out of 5 bolts one will be defective is ....

(*f*) If the probability of hitting a target is 5% and 5 shots are fired independently, the probability that the target will be hit at least once is ....

(*g*) If n and p are the parameters of a binomial distribution the standard deviation is ....

(*h*) The mean, standard deviation and skewness of Binomial distribution are ____ and ___. (*A.M.I.E., Summer 2001*)

(*i*) If three persons selected at random are stopped on a street, then the probability that all of them were born on Sunday is ____. (*A.M.I.E., Winter 2001*)

Ans. (*a*) $\frac{2}{9}$, (*b*) $10.\frac{5^3}{6^5}$, (*c*) $\frac{665}{729}$, (*d*) (*i*), (*e*) $\frac{1}{2}\left(\frac{9}{10}\right)^4$, (*f*) $1-(.95)^5$, (*g*) $\sqrt{npq}$ (*h*) $np, \sqrt{npq}, 0$ (*i*) $\frac{1}{343}$

3. Tick √ the correct answer :

(a) If a coin is tossed 6 times in succession, the probability of getting at least one head is
(i) 1/64 (ii) 3/32 (iii) 63/64 (iv) 1/2

(b) A coin is tossed until a tail appears or at the most five times. Given that the tail does not appear on the first two tosses, the probability that the coin will be tossed 5 times, is
(i) 1/2 (ii) 3/5 (iii) 1/3 (iv) 1/4

(c) In a certain manufacturing process it is known that on an average, 1 in every 100 items is defective. What is the probability that 5 items are inspected before a defective item is found?
(i) .0096 (ii) .96 (iii) .096 (iv) none of these

(d) The probability that a marksman will hit a target is given as $\frac{1}{5}$. Then his probability of at least one hit in 10 shots is
(i) $1-\left(\frac{4}{5}\right)^{10}$ (ii) $\frac{1}{5^{10}}$ (iii) $1-\frac{1}{5^{10}}$ (iv) None of these.

(e) The probability of having at least one tail in 4 throws with a coin is
(i) $\frac{15}{16}$, (ii) $\frac{1}{16}$, (iii) $\frac{1}{4}$, (iv) 1.

(f) A coin is tossed 3 times. The probability of obtaining two heads will be
(i) $\frac{3}{8}$, (ii) $\frac{1}{2}$, (iii) 1, (iv) 2.

(g) 8 coins are tossed simultaneously. The probability of getting at least 6 heads is
(i) $\frac{57}{64}$, (ii) $\frac{229}{256}$, (iii) $\frac{7}{64}$, (iv) $\frac{37}{256}$.

(h) Three unbiased coins are tossed simultaneously. This is repeated four times. The probability of getting at least one head each time is
(i) $\left(\frac{3}{4}\right)^4$ (ii) $\left(\frac{7}{8}\right)^4$ (iii) $\left(\frac{1}{8}\right)^4$ (iv) $\left(\frac{1}{4}\right)^4$

(i) In rolling two fair dice, the probability of getting equal numbers or numbers with an even product
(i) $\frac{6}{36}$ (ii) $\frac{30}{36}$ (iii) $\frac{27}{36}$ (iv) $\frac{3}{36}$.

(j) In a binomial distribution the sum and the product of the mean and variance are $\frac{25}{3}$ and $\frac{50}{3}$ respectively. The distribution is
(i) $\left(\frac{4}{5}+\frac{1}{5}\right)^{15}$ (ii) $\left(\frac{2}{3}+\frac{1}{3}\right)^{15}$ (iii) $\left(\frac{3}{4}+\frac{1}{4}\right)^{15}$ (iv) None of these.

(k) A room has three lamp sockets. From a collection of 10 light bulbs of which only 6 are good. A person selects 3 at random and puts them in a socket. What is the probability that room will have light.
(i) 29/120 (ii) 39/60 (iii) 19/30 (iv) 29/30 (*A.M.I.E.T.E. Dec 2005*)

(l) The inequality between mean and variance of Binomial distribution which is true is
(a) Mean < Variance (b) Mean = Variance
(c) Mean > Variance (d) Mean × Variance = 1 (*A.M.I.E.T.E., Dec. 2006*)

Ans. *(a) (iii), (b) (iv), (c) (i), (d) (i), (e) (i), (f) (i), (g) (iv), (h) (ii), (i) (ii), (j) (ii) (l) (c)*

4. Find the Binomial distribution whose mean is 5 and variance is $\frac{10}{3}$. **Ans.** ${}^{15}C_r\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{15-r}$.

5. (a) The mean, standard deriation and skewness of Binomial distribution are ....,........... and ....
(*AMIE Summer 2001*) **Ans.** np, $\sqrt{npq}$, 0.

(b) If three persons selected at random are stopped on a street, then the probability that all of them were born on Sunday is .... (*AMIE, Winter 2001*) **Ans.** $\frac{1}{343}$

POISSON DISTRIBUTION

11.11 POISSON DISTRIBUTION

Poisson distribution is a particular limiting form of the Binomial distribution when p (or q) is very small and n is large enough.

Poisson distribution is

$$\boldsymbol{P(r) = \frac{m^r e^{-m}}{r!}}$$ (*A.M.I.E.T.E., Summer 1996*)

where m is the mean of the distribution.

Proof. In Binomial distribution.

$$P(r) = {}^nC_r\, q^{n-r}\, p^r = {}^nC_r\,(1-p)^{n-r}\, p^r$$

$$\left(\text{since mean} = m = np,\ p = \frac{m}{n}\right)$$

$$= {}^nC_r\left(1-\frac{m}{n}\right)^{n-r}\left(\frac{m}{n}\right)^r \qquad (m \text{ is constant})$$

$$= \frac{n(n-1)(n-2)\ldots(n-\overline{r-1})}{r!}\left(\frac{m}{n}\right)^r\left(1-\frac{m}{n}\right)^{n-r}$$

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

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

Taking limits, when n tends to infinity

$$\lim_{n\to\infty}\left(1-\frac{m}{n}\right)^n = \lim_{n\to\infty}\left[\left(1-\frac{m}{n}\right)^{-n/m}\right]^{-m} = e^{-m}$$

$$P(r) = \frac{m^r}{r!}e^{-m}$$

$$\boldsymbol{P(r) = \frac{e^{-m}\cdot m^r}{r!}}$$

11.12 MEAN OF POISSON DISTRIBUTION

$$P(r) = \frac{e^{-m}\,m^r}{r!}$$ (*A.M.I.E.T.E., Summer 2004, 1996*)

Successes r	*Frequency f*	*f . r*
0	$\dfrac{e^{-m}\,m^0}{0\,!}$	0
1	$\dfrac{e^{-m}\,m^1}{1\,!}$	$e^{-m}\,.\,m$
2	$\dfrac{e^{-m}\,m^2}{2\,!}$	$e^{-m}\,.\,m^2$
3	$\dfrac{e^{-m}\,m^3}{3\,!}$	$\dfrac{e^{-m}\,.\,m^3}{2\,!}$
...	...	...
r	$\dfrac{e^{-m}\,m^r}{r\,!}$	$\dfrac{e^{-m}\,.\,m^r}{(r-1)\,!}$
...	...	...

$$\Sigma fr = 0 + e^{-m}\,.\,m + e^{-m}\,.\,m^2 + e^{-m}\,.\,\frac{m^3}{2!} + \ldots + e^{-m}\,\frac{m^r}{(r-1)\,!} + \ldots$$

$$= e^{-m}\,.\,m\left[1 + \frac{m}{1\,!} + \frac{m^2}{2\,!} + \ldots + \frac{m^{r-1}}{(r-1)\,!} + \ldots\right]$$

$$= m\,.\,e^{-m}\,.\,[\,e^m\,] = m$$

$$\text{Mean} = \frac{\Sigma f r}{\Sigma f} = \frac{m}{1},\quad \textbf{Mean} = \boldsymbol{m}.\quad \textbf{Ans.}$$

11.13 STANDARD DEVIATION OF POISSON DISTRIBUTION

$$P\,(r) = \frac{e^{-m}\,m^r}{r\,!}.$$

(A.M.I.E.T.E., Summer 1996)

Successes r	*Frequency* f	*Product* rf	*Product* r^2f
0	$\dfrac{e^{-m}\,m^0}{0\,!}$	0	0
1	$\dfrac{e^{-m}\,m^1}{1\,!}$	$e^{-m}\,.\,m$	$e^{-m}\,.\,m$
2	$\dfrac{e^{-m}\,m^2}{2\,!}$	$e^{-m}\,.\,m^2$	$2\,e^{-m}\,.\,m^2$
3.	$\dfrac{e^{-m}\,m^3}{3\,!}$	$e^{-m}\,.\,\dfrac{m^3}{2\,!}$	$3\,e^{-m}\,.\,\dfrac{m^3}{2\,!}$
...	...	...	...
r	$\dfrac{e^{-m}\,m^r}{r\,!}$	$\dfrac{e^{-m}\,m^r}{(r-1)\,!}$	$\dfrac{r\,e^{-m}\,.\,m^r}{(r-1)\,!}$
...	...	...	...

$$\Sigma f = 1 \quad \Sigma f r = m$$

$$
\begin{aligned}
\Sigma f r^2 &= 0 + e^{-m} \,.\, m + 2\, e^{-m} \,.\, m^2 + 3 \,.\, e^{-m} \,.\, \frac{m^3}{2} + \ldots + \frac{r e^{-m} \,.\, m^r}{(r-1)\,!} + \ldots \\
&= m \,.\, e^{-m} \left[1 + 2m + \frac{3m^2}{2!} + \frac{4m^3}{3!} + \ldots + \frac{r \,.\, m^{r-1}}{(r-1)\,!} + \ldots \right] \\
&= m \,.\, e^{-m} \left[1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots + \frac{m^{r-1}}{(r-1)\,!} + \ldots \right. \\
&\qquad \left. + \; m + \frac{2m^2}{2\,!} + \frac{3m^3}{3\,!} + \ldots + \frac{(r-1)\, m^{r-1}}{(r-1)\,!} + \ldots \right] \\
&= m \,.\, e^{-m} \left[\left\{ 1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \ldots + \frac{m^{r-1}}{(r-1)\,!} + \ldots \right\} \right. \\
&\qquad \left. + m \left\{ 1 + \frac{m}{1!} + \frac{m^2}{2!} + \ldots + \frac{m^{r-2}}{(r-2)\,!} + \ldots \right\} \right] \\
&= m \,.\, e^{-m} \left[e^m + m\, e^m \right] = m + m^2 \\
\sigma^2 &= \frac{\Sigma f r^2}{\Sigma f} - \left(\frac{\Sigma f r}{\Sigma f} \right)^2 = \frac{m + m^2}{1} - (m)^2 = m \quad \text{or} \quad \sigma = \sqrt{m}
\end{aligned}
$$

Hence mean and variance of a Poisson distribution are each equal to m. Similarly we can obtain,

$$
\begin{aligned}
\mu_3 &= m\,, & \mu_4 &= 3\, m^2 + m \\
\beta_1 &= \frac{1}{m}\,, & \beta_2 &= 3 + \frac{1}{m} \\
\gamma_1 &= \frac{1}{\sqrt{m}}\,, & \gamma_2 &= \frac{1}{m}
\end{aligned}
$$

11.14 MEAN DEVIATION

Show that in a Poisson distribution with unit mean, and the mean deviation about the mean is $\left(\frac{2}{e} \right)$ times the standard deviation. (*A.M.I.E.T.E., Dec. 2005*)

Solution. $P(r) = \dfrac{m^r}{r!}\, e^{-m}$ But mean = 1 *i.e.* $m = 1$ and S.D. = $\sqrt{m} = 1$

Hence $P(r) = \dfrac{e^{-1}}{r\,!} = \dfrac{1}{e} \cdot \dfrac{1}{r\,!}$

r	$P(r)$	$\lvert r - 1 \rvert$	$P(r) \lvert r - 1 \rvert$
0	$\frac{1}{e}$	1	$\frac{1}{e}$
1	$\frac{1}{e}$	0	0
2	$\frac{1}{e}\,\frac{1}{2!}$	1	$\frac{1}{e}\,\frac{1}{2!}$
3	$\frac{1}{e}\,\frac{1}{3!}$	2	$\frac{1}{e}\,\frac{2}{3!}$
4	$\frac{1}{e}\,\frac{1}{4!}$	3	$\frac{1}{e}\,\frac{3}{4!}$
r	$\frac{1}{e}\,\frac{1}{r\,!}$	$r - 1$	$\frac{1}{e}\,\frac{r-1}{r\,!}$

$$\Sigma P(r)\,|r-1| = \frac{1}{e}+0+\frac{1}{e}\,\frac{1}{2!}+\frac{1}{e}\,\frac{2}{3!}+\frac{1}{e}\,\frac{3}{4!}+\ldots+\frac{1}{e}\,\frac{r-1}{r!}+\ldots$$

$$= \frac{1}{e}\left[1+0+\frac{1}{2!}+\frac{2}{3\,!}+\frac{3}{4\,!}+\ldots\ldots+\frac{r-1}{r\,!}+\ldots\right]$$

$$= \frac{1}{e}\left[1+\left(\frac{1}{1\,!}-\frac{1}{1\,!}\right)+\left(\frac{+2}{2\,!}-\frac{1}{2\,!}\right)+\left(\frac{3}{3\,!}-\frac{1}{3\,!}\right)+\left(\frac{4}{4\,!}-\frac{1}{4\,!}\right)+\ldots+\left(\frac{r}{r\,!}-\frac{1}{r\,!}\right)+\ldots\right]$$

$$= \frac{1}{e}\left[1+\frac{1}{1\,!}+\frac{2}{.\,2!}+\frac{3}{3\,!}+\frac{4}{4\,!}+\ldots+\frac{r}{r\,!}+\ldots-\frac{1}{1\,!}-\frac{1}{2\,!}-\frac{1}{3\,!}-\frac{1}{4\,!}\ldots-\frac{1}{r\,!}-\ldots\right]$$

$$= \frac{1}{e}\left[1+\left\{1+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\ldots+\frac{1}{(r-1)\,!}+\ldots\right\}\right.$$

$$\left.-\left\{1+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\frac{1}{4!}+\ldots+\frac{1}{r!}\;\ldots\right\}+1\right]$$

$$= \frac{1}{e}\;[1+e-e+1] = \frac{2}{e}$$

$$= \frac{2}{e}\;(1) = \frac{2}{e}\text{ S.D.}$$

Proved

11.15 MOMENT GENERATING FUNCTION OF POISSON DISTRIBUTION

(*A.M.I.E., Summer 2000*)

Solution. $$P(r) = \frac{e^{-m}\,m^r}{\lfloor r}$$

Let $M_x(t)$ be the moment generating function, then

$$M_x(t) = \sum_{r=0}^{\infty} e^{tr}\cdot\frac{e^{-m}\cdot m^r}{\lfloor r} = \sum_{r=0}^{\infty} e^{-m}\cdot\frac{(me^t)^r}{\lfloor r}$$

$$= e^{-m}\left[1+me^t+\frac{(me^t)^2}{\lfloor 2}+\frac{(me^t)^3}{\lfloor 3}+\ldots\right] = e^{-m}\cdot e^{me^t} = e^{m(e^t-1)}$$

11.16 CUMULANTS

The cumulant generating function $K_x(t)$ is given by

$$K_x(t) = \log_e M_x(t) = \log_e e^{m(e^t-1)} = m(e^t-1)\log_e e$$

$$= m(e^t-1) = m\left[1+t+\frac{t^2}{\lfloor 2}+\frac{t^3}{\lfloor 3}+\ldots+\frac{t^r}{\lfloor r}+\ldots-1\right]$$

$$= m\left[t+\frac{t^2}{\lfloor 2}+\frac{t^3}{\lfloor 3}+\ldots+\frac{t^r}{\lfloor r}+\ldots\right]$$

Now $$k_r = r\text{th cumulant}$$

$$= \text{coefficient of }\frac{t^2}{\lfloor r}\text{ in }K(t) = m$$

i.e., $$k_r = m,\; r = 1, 2, 3, \ldots$$

Hence, all the cumulants of the Poisson distribution are equal. In particular, we have

Mean $= K_1 = m,\quad \mu_2 = K_2 = m,\quad \mu_3 = K_3 = m$

$$\mu_4 = K_4 + 3\,K_2^2 = m + 3\,m^2$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{m^2}{m^3} = \frac{1}{m}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{m + 3\,m^2}{m^2} = \frac{1}{m} + 3$$

11.17 RECURRENCE FORMULA FOR POISSON DISTRIBUTION

Solution. By Poisson distribution

$$P(r) = \frac{e^{-m} \,.\, m^r}{r\,!} \quad ...(1) \qquad P(r+1) = \frac{e^{-m}\, m^{r+1}}{(r+1)\,!} \quad ...(2)$$

By dividing (2) by (1), we get

$$\frac{P(r+1)}{P(r)} = \frac{e^{-m}\, m^{r+1}}{(r+1)\,!}\; \frac{sub\; r\,!}{e^{-m} \,.\, m^r} = \frac{m}{r+1}$$

$$P(r+1) = \frac{m}{r+1}\, P(r) \qquad \textbf{Ans.}$$

Example 38. *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. Also find P(r ≥ 4).*

Solution. Variance = m = 2; Mean = 2

$$P(r+1) = \frac{m}{r+1} P(r) \qquad \text{[Recurrence relation]}$$

Now $$P(r+1) = \frac{2}{r+1} P(r) \qquad \left[\begin{array}{l} P(r) = \dfrac{e^{-m} \cdot m^r}{r\,!} \\ P(0) = e^{-m} = e^{-2} = 0.1353 \end{array}\right]$$

If $r = 0$, $P(1) = \dfrac{2}{0+1} P(0) = \dfrac{2}{0+1}(0.1353) = 0.2706$

If $r = 1$, $P(2) = \dfrac{2}{1+1} P(1) = \dfrac{2}{2}(0.2706) = 0.2706$

If $r = 2$, $P(3) = \dfrac{2}{2+1} P(2) = \dfrac{2}{3}(0.2706) = 0.1804$

If $r = 3$, $P(4) = \dfrac{2}{3+1} P(3) = \dfrac{1}{3}(0.1804) = 0.0902$

$$\begin{aligned} P(r \geq 4) &= P(4) + P(5) + P(6) + ... \\ &= 1 - [P(0) + P(1) + P(2) + P(3)] \\ &= 1 - [0.1353 + 0.2706 + 0.2706 + 0.1804] \\ &= 1 - 0.8569 = 0.1431 \qquad \textbf{Ans.} \end{aligned}$$

Example 39. *Assume that the probability of an individual coal miner being killed in a mine accident during a year is $\frac{1}{2400}$. Use appropriate statistical distribution to calculate the probability that in a mine employing 200 miners, there will be at least one fatal accident in a year.*

Solution. $$P = \frac{1}{2400}, \quad n = 200$$

$$m = np = \frac{200}{2400} = \frac{1}{12}$$

$$\begin{aligned} \text{P (At least one)} &= P(1 \text{ or } 2 \text{ or } 3 \text{ or ... or } 200) \\ &= P(1) + P(2) + P(3) + \ldots + P(200) \end{aligned}$$

$$= 1 - P(0) = 1 - \frac{e^{-m} . m^0}{0!}$$

$$= 1 - e^{-\frac{1}{12}} = 1 - 0.92 = 0.08 \qquad \textbf{Ans.}$$

Example 40. *Suppose 3% of bolts made by a machine are defective, the defects occurring at random during production. If bolts are packaged 50 per box, find (a) exact probability and (b) Poisson approximation to it, that a given box will contain 5 defectives.*

Solution. $p = \frac{3}{100} = 0.03$

(a) $q = 1 - p = 1 - 0.03 = 0.97$

Hence the probability for 5 defective bolts in a lot of 50

$$= {}^{50}C_5 (0.03)^5 (0.97)^{45} = 0.013074$$

(b) To get Poisson approximation $m = n\,p = 50 \times \frac{3}{100} = \frac{3}{2} = 1.5$

Required Poisson approximation $= \frac{m^r e^{-m}}{r!} = \frac{(1.5)^5 e^{-1.5}}{5!} = 0.01412$ **Ans.**

Example 41. *The number of arrivals of customers during any day folllows Poisson distibution with a mean of 5. What is the probability that the total number of customers on two days selected at random is less than 2?*

Solution. $m = 5$

$$P(r) = \frac{e^{-m} m^r}{r!}, \quad P(r) = \frac{e^{-5} (5)^r}{r!}$$

If the number of customers on two days $< 2 = 1$ or 0

No. of customers

First day	*Second day*	*Total*
0	0	0
0	1	1
1	0	1

Required probability $= P(0)\,P(0) + P(0)\,P(1) + P(1)\,P(0)$

$$= \frac{e^{-5}(5)^0}{0!} \cdot \frac{e^{-5}(5)^0}{0!} + \frac{e^{-5}(5)^0}{0!} \cdot \frac{e^{-5}(5)^1}{1!} + \frac{e^{-5}(5)^1}{1!} \frac{e^{-5}(5)^0}{0!}$$

$$= e^{-5} \cdot e^{-5} + e^{-5} \cdot e^{-5} \cdot 5 \cdot + e^{-5} \cdot 5 \cdot e^{-5}$$

$$= e^{-10} [1 + 5 + 5] = 11\, e^{-10} = 11 \times 4.54 \times 10^{-5}$$

$$= 4.994 \times 10^{-4} \qquad \textbf{Ans.}$$

Example 42. *Using Poisson distribution, find the probability that the ace of spades will be drawn from a pack of well-shuffled cards at least once in 104 consecutive trials.*

Solution. Probability of the ace of spades $= P = \frac{1}{52}$, $n = 104$

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

$$P(r) = e^{-m} . \frac{m^r}{r!} = e^{-2} . \frac{2^r}{r!} = \frac{1}{e^2} \frac{2^r}{r!}$$

P (At least once) = P (1) + P (2) + P (3) + ... + P (104) = 1– P (0)

$$= 1 - \frac{1}{e^2} \times \frac{2^o}{0!} = 1 - \frac{1}{e^2} = 1 - 0.135 = 0.865 \qquad \text{Ans.}$$

Example 43. *In a certain factory producing cycle tyres, there is a small chance of 1 in 500 tyres to be defective. The tyres are supplied in lots of 10. Using Poisson distribution, calculate the approximate number of lots containing no defective, one defective and two defective tyres, respectively, in a consignment of 10,000 lots.* (*A.M.I.E.T.E. Winter 2003*)

Solution.
$$p = \frac{1}{500}, \quad n = 10$$
$$m = np = 10 \times \frac{1}{500} = \frac{1}{50} = 0.02$$
$$P(r) = \frac{e^{-m} \,.\, m^r}{r!}$$

(*i*) Probability of no defective = P (0) $= \dfrac{e^{-0.02}\ (0.02)^0}{0!} = e^{-0.02} = 0.9802$

Number of lots containing no defective = $10{,}000 \times 0.9802 = 9802$ lots

(*ii*) Probability of one defective = P (1) = $\dfrac{e^{-0.02}\,(0.02)^1}{1\,!}$

$$= 0.9802 \times 0.02 = 0.019604$$

Number of lots containing 1 defective

$$= 10{,}000 \times 0.019604 = 196 \text{ lots}$$

(*iii*) Probability of two defectives = P (2)

$$= \frac{e^{-0.02}\,(0.02)^2}{2\,!}$$
$$= 0.9802 \times 0.0002 = 0.00019604$$

Number of lots containing 2 defectives = 10,000 × 0.000196 = 2 lots **Ans.**

Example 44. *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 number of days in a year on which (i) car is not used (ii) the number of days in a year on which some demand is refused.* (*A.M.I.E Summer 2004., Winter 2001*)

Solution. $m = 1{\cdot}5$

(*i*) If the car is not used, then demand (r) = 0.

$$P(r) = \frac{e^{-m} \,.\, m^r}{r!}$$
$$P(0) = \frac{e^{-1.5}\,(1.5)^0}{0!} = e^{-1.5} = 0.2231$$

Number of days in a year when there is no demand of car = $365 \times 0.2231 = 81.4315$

Ans. 81 days

(*ii*) Some demand is refused if the number of demands is more than two *i.e.* $r > 2$.

$$P(r > 2) = P(3) + P(4) + \ldots\ldots = 1 - [P(0) + P(1) + P(2)]$$
$$= 1 - \left[\frac{e^{-1.5}\,(1.5)^o}{0\,!} + \frac{e^{-1.5}\ (1.5)^1}{1\,!} + \frac{e^{-1.5}\ (1.5)^2}{2\,!}\right]$$
$$= 1 - [e^{-1.5} + e^{-1.5} \times 1.5 + e^{-1.5} \times 1.125]$$

$$= 1 - e^{-1.5}[1 + 1.5 + 1.125] = 1 - e^{-1.5} \times 3.625$$
$$= 1 - 0.2231 \times 3.625 = 1 - 0.8087375$$
$$= 0.1912625$$

Number of days in a year when some demand is refused $= 365 \times 0.1912625 = 69.81$

Ans. 70 days

Example 45. *If the probability that an individual suffers a bad reaction from a certain injection is 0.001, determine the probability that out of 2000 individuals*

(a) exactly 3 (b) more than 2 individuals (c) None (d) More than one individual will suffer a bad reaction. (*A.M.I.E.T.E., Winter 2000*)

Solution. $P = 0.001, \; n = 2000$
$$m = np = 2000 \times 0.001 = 2$$

$\therefore$ $$P(r) = \frac{e^{-m} m^r}{r!} = e^{-2}\frac{2^r}{r!} = \frac{1}{e^2} \times \frac{2^r}{r!}$$

(*a*) $$P(3) = \frac{1}{e^2} \times \frac{2^3}{3!} = \frac{1}{(2.718)^2} \times \frac{8}{6} = (.135) \times \frac{4}{3}$$
$$= 0.18$$

(*b*) P (more than 2) $= P(3) + P(4) + P(5) + \ldots + P(2000)$
$$= 1 - [P(0) + P(1) + P(2)]$$
$$= 1 - \left[\frac{e^{-2}(2)^0}{0!} + \frac{e^{-2}(2)^1}{1!} + \frac{e^{-2}(2)^2}{2!}\right]$$
$$= 1 - e^{-2}[1 + 2 + 2] = 1 - \frac{5}{e^2}$$
$$= 1 - 5 \times 0.135 = 1 - 0.675 = 0.325$$ **Ans.**

(*c*) $$P(\text{none}) = P(0) = \frac{e^{-2}(2)^0}{0!} = 0.135$$

(*d*) P (more than 1) $= P(2) + P(3) + P(4) + \ldots + P(2000)$
$$= 1 - [P(0) + P(1)]$$
$$= 1 - \left[\frac{e^{-2}(2)^0}{0!} + \frac{e^{-2} \,.\, 2^1}{1!}\right] = 1 - 3e^{-2} = 1 - 3 \times 0.135 = 1 - 0.405$$
$$= 0.595$$ **Ans.**

Example 46. *A manufacturer knows that the razor blades he makes contain on an average 0.5% of defectives. He packs them in packets of 5. What is the probability that a packet picked at random will contain 3 or more faulty blades ?*

Solution. $p = 0.5\% = 0.005, \quad n = 5$
$$m = np = 5 \times 0.005 = 0.025$$
$$P(r) = \frac{e^{-m} \,.\, m^r}{r!} = \frac{e^{-0.025}(.025)^r}{r!}$$
$$P(3 \text{ or more}) = P(3) + P(4) + P(5) = \frac{e^{-0.025}(.025)^3}{3!} + \frac{e^{-0.025}(.025)^4}{4!} + \frac{e^{-0.025}(.025)^5}{5!}$$
$$= \frac{e^{-.025}(.025)^3}{5!}[20 + 5(.025) + (.025)^2]$$

$$= \frac{0.975 \times .000015625 \times 20.125625}{120}$$

$$= 0.000002555.$$ **Ans.**

Example 47. *Suppose that a book of 600 pages contains 40 printing mistakes. Assume that these errors are randomly distributed throughout the book and x, the number of errors per page has a Poisson distribution. What is the probability that 10 pages selected at random will be free of errors ?*

Solution. $p = \frac{40}{600} = \frac{1}{15}, \quad n = 10, \quad m = np = 10 \times \frac{1}{15} = \frac{2}{3}$

$$P(r) = \frac{e^{-m} \cdot m^r}{r\,!} = \frac{e^{-2/3}\left(\frac{2}{3}\right)^r}{r\,!}$$

$$P(0) = \frac{e^{-2/3}\left(\frac{2}{3}\right)^0}{0\,!} = e^{-2/3} = 0.51$$ **Ans.**

Example 48. *A manufacturer knows that the condensers he makes contain on an average 1% of defectives. He packs them in boxes of 100. What is the probability that a box picked at random will contain 4 or more faulty condensers ?*

Solution. $p = 1\% = 0.01$, n = 100, $m = np = 100 \times 0.01 = 1$

$$P(r) = \frac{e^{-m} \cdot (m)^r}{r!} = \frac{e^{-1}(1)^r}{r\,!} = \frac{e^{-1}}{r\,!}$$

P (4 or more faulty condensers) $= P(4) + P(5) + \ldots + P(100)$

$$= 1 - [P(0) + P(1) + P(2) + P(3)] = 1 - \left[\frac{e^{-1}}{0!} + \frac{e^{-1}}{1!} + \frac{e^{-1}}{2!} + \frac{e^{-1}}{3!}\right]$$

$$= 1 - e^{-1}\ [1 + 1 + \frac{1}{2} + \frac{1}{6}]$$

$$= 1 - \frac{8}{3e} = 1 - 0.981 = 0.019$$ **Ans.**

Example 49. *An insurance company found that only 0.01% of the population is involved in a certain type of accident each year. If its 1000 policy holders were randomly selected from the population, what is the probability that not more than two of its clients are involved in such an accident next year ? (given that $e^{-0.1} = 0.9048$)*

Solution.
$$p = 0.01\% = \frac{1}{100} \times \frac{1}{100} = \frac{1}{10000}, \quad n = 1000$$

$$m = np = (1000) \times \frac{1}{10000} = \frac{1}{10} = 0.1$$

$$P(r) = \frac{e^{-m}\, m^r}{r!}$$

P (not more than 2) $= P$ (0, 1, and 2) $= P(0) + P(1) + P(2)$

$$= \frac{e^{-0.1}(0.1)^0}{0!} + \frac{e^{-0.1}(0.1)^1}{1!} + \frac{e^{-0.1}(0.1)^2}{2!}$$

$$= e^{-0.1}\left(1 + 0.1 + \frac{0.01}{2}\right)$$

$$= 0.9048 \times 1.105 = 0.9998$$ **Ans.**

Example 50. *A skilled typist, on routine work, kept a record of mistakes made per day during 300 working days.*

Mistakes per day	*0*	*1*	*2*	*3*	*4*	*5*	*6*
No. of days	*143*	*90*	*42*	*12*	*9*	*3*	*1*

Fit a Poisson distribution to the above data and hence calculate the theoretical frequencies.

Solution. The mean number of mistakes

$$= \frac{1}{300}(143 \times 0 + 90 \times 1 + 42 \times 2 + 12 \times 3 + 9 \times 4 + 3 \times 5 + 1 \times 6)$$

$$= \frac{1}{300}(90 + 84 + 36 + 36 + 15 + 6) = \frac{267}{300} = 0.89$$

Number of mistakes	*Probability* $P(r) = \frac{e^{-0.89} \times (0.89)^r}{r\,!}$	*Theoreitical frequency*
0	$\frac{e^{-0.89} \times (0.89)^0}{0\,!} = 0.411$	$0.411 \times 300 = 123.3 = 123$ (say)
1	$\frac{e^{-0.89} \times (0.89)^1}{1\,!} = 0.365$	$0.365 \times 300 = 109.5 = 110$ (say)
2	$\frac{e^{-0.89} \times (0.89)^2}{2\,!} = 0.163$	$0.163 \times 300 = 48.9 = 49$ (say)
3	$\frac{e^{-0.89} \times (0.89)^3}{3\,!} = 0.048$	$0.048 \times 300 = 14.4 = 14$ (say)
4	$\frac{e^{-0.89} \times (0.89)^4}{4\,!} = 0.011$	$0.011 \times 300 = 3.3 = 3$ (say)
5	$\frac{e^{-0.89} \times (0.89)^5}{5\,!} = 0.002$	$0.002 \times 300 = 0.6 = 1$ (say)
6	$\frac{e^{-0.89} \times (0.89)^6}{6\,!} = 0.0003$	$0.0003 \times 300 = 0.09 = 0$ (say)

Example 51. Fit a Poisson distribution to the following data which gives the number of yeast cells per square for 400 squares.

No. of cells per square (x)	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Total*
No. of squares (f)	*103*	*143*	*98*	*42*	*8*	*4*	*2*	*0*	*0*	*0*	*0*	*400*

It is given that $e^{-1.32} = 0.2674$ *(A.M.I.E., Summer 2000)*

Solution.

x	0	1	2	3	4	5	6	7	8	9	10	Total
f	103	143	98	42	8	4	2	0	0	0	0	400
$f \cdot x$	0	143	196	126	32	20	12	0	0	0	0	529

$$m = \text{Mean} = \frac{\Sigma f \cdot x}{\Sigma f} = \frac{529}{400} = 1.32$$

But Poisson distribution is $P(x) = \dfrac{e^{-m} \cdot m^x}{r\,!} = \dfrac{e^{-1.32}\,(1.32)^x}{x\,!}$ or $P(r) = \dfrac{0.2674\,(1.32)^x}{x\,!}$

No. of cells	*Probability* $P(x) = \dfrac{0.2674\,(1.32)^x}{x\,!}$	*Theoretical frequency*
0	$\dfrac{0.2674\,(1.32)^0}{0\,!} = 0.2674$	$0.2674 \times 400 = 107$
1	$\dfrac{0.2674\,(1.32)^1}{1\,!} = 0.353$	$0.353 \times 400 = 141$
2	$\dfrac{0.2674\,(1.32)^2}{2\,!} = 0.233$	$0.233 \times 400 = 93.2$
3	$\dfrac{0.2674\,(1.32)^3}{3\,!} = 0.1025$	$0.1025 \times 400 = 41$
4	$\dfrac{0.2674\,(1.32)^4}{4\,!} = 0.0338$	$0.0338 \times 400 = 13.52$ i.e., 14
5	$\dfrac{0.2674\,(1.32)^5}{5\,!} = 0.00893$	$0.00893 \times 400 = 3.57$ i.e., 4
6	$\dfrac{0.2674\,(1.32)^6}{6\,!} = 0.00196$	$0.00196 \times 400 = 0.784$ i.e., 1
7	$\dfrac{0.2674\,(1.32)^7}{7\,!} = 0.00037$	$0.00037 \times 400 = 0.148$ i.e., 0
8	$\dfrac{0.2674\,(1.32)^8}{8\,!} = 0.00006$	$0.00006 \times 400 = 0.24$ i.e., 0
9	$\dfrac{0.2674\,(1.32)^9}{9\,!} = 0.00000897$	$0.00000897 \times 400 = 0.003588$ i.e., 0
10	$\dfrac{0.2674\,(1.32)^{10}}{10\,!} = 0.00000118$	$0.00000118 \times 400 = 0.000472$ i.e., 0

Exercise 11.5

1. Find the probability that at most 5 defective fuses will be found in a box of 200 fuses if experience shows that 2 per cent of such fuses are defective. **Ans.** 0.785

2. The number of accidents during a year in a factory has the Poisson distribution with mean 1.5. The accidents during different years are assumed independent. Find the probability that only 2 accidents take place during 2 years time. **Ans.** 0.224

3. A manufacturer of cotter pins knows that 5% of his product is defective. If he sells cotter pins in boxes of 100 and guarantee that not more than 10 pins will be defective, what is the approximate probability that a box will fail to meet the guaranteed quality. $[e^{-5} = 0.006738]$

Ans. 0.0136875

4. Suppose the number of telephone calls on an operator received from 9.00 to 9.05 follow a Poisson distribution with mean 3. Find the probability that

(*i*) the operator will receive no calls in that time interval tomorrow,

(*ii*) in the next three days the operator will receive a total of 1 call in that time interval. $[e^{-3} = 0.04978]$

Ans. (*i*) e^{-3} (*ii*) $3 \times (e^{-3})^2\ (e^{-3} . 3)$

5.

On the basis of past record it has been found that there is a 70% chance of power-cut in a city on any particular day. What is the probability that from the first to the 10th day of the month, there are 5 or more days without power cut. (*A.M.I.E.T.E., Summer 2001*)

Ans. $\left(\frac{3^5}{5\,!} + \frac{3^6}{6\,!} + \frac{3^7}{7\,!} + \frac{3^8}{8\,!} + \frac{3^9}{9\,!} + \frac{3^{10}}{10\,!}\right) e^{-3}$

6. The distribution of typing mistakes committed by a typist is given below. Assuming a Poisson model, find out the expected frequencies:

Mistakes per pages	0	1	2	3	4	5
No. of pages	142	156	69	27	5	1

Ans. 147, 147, 74, 25, 6, 1 pages.

7. Let x be the number of cars per minute passing a certain crossing of roads between 5.00 P.M. and 7.00 P.M. on a holiday. Assume x has a Poisson distribution with mean 4. Find the probability of observing atmost 3 cars during any given minute between 5.00 P.M. and 7 P.M. (given $e^{-4} = 0.0183$)

Ans. 0.4331

8. Let x be the number of cars, passing a certain point, per minute at a particular time. Assuming that x has a poisson distribution with mean 0.5, find the probability of observing 3 or fewer cars during any given minute. **Ans.** 0.998

9. Number of customers arriving at a service counter during a day has a Poisson distribution with mean 100. Find the probability that at least one customer will arrive on each day during a period of five days. Also find the probability that exactly 3 customers will arrive during two days.

Ans. $(1 - e^{-100})^5,\ e^{-200} \times \frac{4\,(100)^3}{3}$

10. The random variable X has a Poisson distribution. If

$P\,(X = 1) = 0.01487, P\,(X = 2) = 0.04461$. Then find $P\,(X = 3)$. **Ans.** 0.08922

11. A source of water is known to contain bacteria with mean number of bacteria per cc equal to 2. Five 1 cc test tubes were filled with water. Assuming that Poisson distribution is applicable, calculate the probability that exactly 2 test tubes contain at least 1 bacterium each. **Ans.** $\frac{2}{5}\,(1 - e^{-2}) = 0.3459$

12. In a normal summer, a truck driver gets on an average one puncture in 1000 km. Applying Poisson distribution, find the probability that he will have

(*i*) no puncture, (*ii*) two punctures in a journey of 3000 kms. **Ans.** (*i*) e^{-3} (*ii*) 4.5 e^{-3}

13. Wireless sets are manufactured with 25 soldered joints each. On the average, 1 joint in 500 is defective. How many sets can be expected to be free from defective joints in a consignment of 10000 sets ?

Ans. 9512

14. In a certain factory turning out razor blades, there is small chance $\frac{1}{500}$ for any blade to be defective. The blades are supplied in packets of 10. Using Poisson's distribution, calculate the approximate number of packets containing (i) no defective (ii) one defective and (iii) two defective blades respectively in a consignment of 10,000 packets. ($e^{-0.02} = 0.9802$).

Ans. (*i*) 9802 (*ii*) 196 (*iii*) 2

15. If m and μ_r denote by the mean and central rth moment of a Poisson distribution, then prove that

$$\mu_{r+1} = r\,m\,\mu_{r-1} + m\frac{d\mu_r}{dm}.$$

$$\left[\textbf{Hint. } \mu_r = \sum_{n=0}^{\infty} (x - m)^r \frac{e^{-m} m^x}{x\,!},\ \text{find } \frac{d\,\mu r}{dm}\right]$$

16. The random variable x has a Poisson distribution. If $P(x=3)=\frac{1}{6}$, $P(x=2)=\frac{1}{3}$, then $P(x=0)$ is

(i) exp (–3/2) (ii) exp (3/2) (iii) exp (–3) (iv) exp (–1/2) **Ans.** (i)

17. Suppose that on an average 1 house in 1000 houses gets fire in a year in a district. If there are 2000 houses in that district find the probability that exactly 5 houses will have fire during the year. Also find approximate probability using Poisson distribution. (*A.M.I.E.T.E., Dec. 2006*)

18. Assuming that the probability of a fatal accident in a factory during the year is $\frac{1}{1200}$. Calculate the probability that in a factory employing 300 workers, there will be at least two fatal accidents in a year ($e^{-0.25}=-0.770$)

19. An insurance company found that only 0.01% of the population is involved in a certain type of accident each year. If 1000 policy holders were randomly selected from the population, what is the probability that not more than 2 of its clients are involved in such an accident next year.
(*A.M.I.E.T.E., Summer 2001*) **Ans.** 0.9998

20. Fill in the blanks :

(a) If a random variable x follows Poisson distribution such that $P(x=1)=P(x=2)$, then the mean of the distribution is ....

(b) Mean and variance of a Poisson distribution are ....

(c) If the probability of a defective fuse is .05, the variance for the distribution of defective fuses in a total of 40 is ....

(d) The probability of the king of hearts drawn from a pack of cards once in 52 trials is ....

(e) If the standard deviation of the Poisson distribution is $\sqrt{2}$, the probability for $r=2$ is ....

(f) If x has a modified Poisson distribution

$P_k = P_r(x=k) = \frac{(e^m-1)^{-1}\, m^k}{k\,!}$, $(k=1, 2, 3, ...)$, then the expectation of x is ....

(g) If x has a poisson distribution such that $P(x=k)=P(x=k+1)$ for some positive integer k then mean of x is .... (*A.M.I.E., Summer 2000*)

Ans. (a) 2, (b) equal, (c) 2, (d) $\frac{1}{e}$, (e) $\frac{2}{e^2}$, (f) $\frac{m\, e^m}{e^m-1}$ (g) $k+1$.

21. Choose the correct answer:

(a) Let X be a Poisson random variable, such that $2\,P(X=0)=P(X=2)$. Then standard deviation of x is

(i) 4. (ii) 2. (iii) $-\sqrt{2}$ (iv) $\sqrt{2}$.

(b) A card is drawn from a well shuffled pack of cards. A sequence of 156 consecutive trials are made. Using Poisson distribution, the probability that the Queen of clubs will be drawn at least once is obtained as

(i) e^{-3}; (ii) $1-e^{-3}$ (iii) $e^{-\frac{1}{3}}$; (iv) $1-e^{-\frac{1}{3}}$ **Ans.** (a) (iv), (b) (ii)

NORMAL DISTRIBUTION

11.18 CONTINUOUS DISTRIBUTION

So far we have dealt with discrete distributions where the variate takes only the integral values. But the variates like temperature, heights and weights can take all values in a given interval. Such variables are called continuous variables.

Distribution function.

If $F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\, dx$, then $f(x)$ is defined as the Distribution Function.

Let $f(x)$ be a continuous function, then Mean $= \int_{-\infty}^{+\infty} x.f(x)\,dx$

$$\text{Variance} = \int_{-\infty}^{+\infty} (x-\bar{x})^2 . f(x)\,dx. \qquad (\bar{x} = \text{mean})$$

Note. $f(x)$ is called probability density function if

(1) $f(x) \geq 0$ for every value of x. (2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (3) $\int_{a}^{b} f(x)\,dx = P, \; (a < x < b)$

(*A.M.I.E., Summer 1995*)

Example 52. *A function $f(x)$ is defined as follows*

$$f(x) = \begin{cases} 0, & x < 2 \\ \frac{1}{18}(2x+3), & 2 \leq x \leq 4 \\ 0, & x > 4. \end{cases}$$

Show that it is a probability density function.

Solution. $$f(x) = \begin{cases} 0, & x < 2 \\ \frac{1}{18}(2x+3), & 2 \leq x \leq 4 \\ 0, & x > 4. \end{cases}$$

If $f(x)$ is a probability density function, then

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

Here $$\int_2^4 \frac{1}{18}(2x+3)\,dx = \frac{1}{18}[x^2+3x]_2^4 = \frac{1}{18}(16+12-4-6) = 1$$

(*ii*) $f(x) > 0$ for $2 \leq x \leq 4$

Hence the given function is a probability density function. **Proved.**

Example 53. *The diameter of an electric cable is assumed to be continuous random variate with probability density function:*

$$f(x) = 6x(1-x), \quad 0 \leq x \leq 1$$

(*i*) *verify that above is a p.d.f.*
(*ii*) *find the mean and variance.*

Solution. (*i*) $$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 6x(1-x)\,dx = \int_0^1 (6x-6x^2)\,dx$$

$$= (3x^2 - 2x^3)_0^1 = 3 - 2 = 1$$

Secondly $f(x) > 0$ for $0 \leq x \leq 1$.
Hence the given function is a probability density function.

(*ii*) $$\text{Mean} = \int_{-\infty}^{\infty} x.f(x)\,dx = \int_0^1 x.6x(1-x)\,dx = \int_0^1 (6x^2-6x^3)\,dx$$

$$= \left(2x^3 - \frac{3}{2}x^4\right)_0^1 = 2 - \frac{3}{2} = \frac{1}{2} \qquad \textbf{Ans.}$$

$$\text{Variance} = \int_{-\infty}^{\infty} (x-\bar{x})^2 . f(x)\,dx = \int_0^1 \left(x - \frac{1}{2}\right)^2 . 6x(1-x)\,dx$$

$$= \int_0^1 \left(x^2 - x + \frac{1}{4} \right)(6x - 6x^2)\, dx \quad = \int_0^1 \left(12x^3 - 6x^4 - \frac{15}{2}x^2 + \frac{3}{2}x \right) dx$$

$$= \left(3x^4 - \frac{6}{5}x^5 - \frac{5}{2}x^3 + \frac{3x^2}{4} \right)_0^1 = \left(3 - \frac{6}{5} - \frac{5}{2} + \frac{3}{4} \right) = \frac{1}{20}$$ **Ans.**

Example 54. *If the probability density function of a random variable x is*

$$f(x) = \begin{cases} kx^{\alpha-1}(1-x)^{\beta-1}, & \text{for } 0 < x < 1,\ \alpha > 0,\ \beta > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find k and mean of x.

Solution. If $f(x)$ is a probability density function,

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

Here $$\int_0^1 k\, x^{\alpha-1}(1-x)^{\beta-1}\, dx = 1$$ [$f(x)$ is a beta function.]

or $$k\,\frac{\Gamma\alpha\,\Gamma\beta}{\Gamma(\alpha+\beta)} = 1 \quad \text{or} \quad k = \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\,\Gamma\beta}$$ **Ans.**

$$\text{Mean} = \int_{-\infty}^{\infty} x\,.\,f(x)\, dx = \int_0^1 x\,.\,k\,x^{\alpha-1}(1-x)^{\beta-1}\, dx = k\int_0^1 x^{\alpha+1-1}(1-x)^{\beta-1}\, dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\,\Gamma\beta}\cdot\frac{\Gamma(\alpha+1)\,\Gamma\beta}{\Gamma(\alpha+\beta+1)} = \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\,\Gamma\beta}\,\frac{\alpha\,\Gamma\alpha\,\Gamma\beta}{(\alpha+\beta)\,\Gamma(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta}$$ **Ans.**

Exercise 11.6

1. The two equal sides of an isosceles triangle are of length a each and the angle θ between them has a probability density function proportional to $\theta(\pi - 0)$ in the range $\left(0, \frac{\pi}{2}\right)$ and zero otherwise. Find the mean value and variance of area of triangle. (*AMIETE Dec. 2005*)
2. Suppose that certain bolts have length L = 400 + X mm, where X is a random variable with probability distribution function.
   $f(x) = \frac{3}{4}(1 - x^2),\ -1 \geq x \leq 1$ and 0, otherwise
   (*i*) Determine C so that with probability $\frac{11}{16}$, a bolt will have length between 400–C and 400+C
   (*ii*) Find the mean and variance of bolt length L. Also find mean and variance of $(2L + 5)$.
3. Let $f(x)$ be a function defined as $f(x) = e^{-x}$, for $x \geq 0$ and $f(x) = 0$ for $x < 0$, then the value of probability distribution function $x = 2$.
   (*a*) $1 + e^{-2}$ (*b*) $1 - e^{-2}$ (*c*) $1 + e^{2}$ (*d*) $1 + e^{-2.5}$ (*A.M.I.E.T.E., Dec. 2006*) **Ans.** (*b*)

11.19 MOMENT GENERATING FUNCTION OF THE CONTINUOUS PROBABILITY DISTRIBUTION ABOUT $x = a$ is given by

$$M_a(t) = \int_{-\infty}^{\infty} e^{t(x-a)} f(x)\, dx \qquad \text{where } f(x) \text{ is p.d.f.}$$

Example 55. *Find the moment generating function of the exponential distribution*

$$f(x) = \frac{1}{c}e^{-x/c} \qquad 0 \leq x \leq \infty,\ c > 0$$

Hence find its mean and S.D.

Solution. The moment generating function about origin is

$$M_0(t) = \int_0^{\infty} e^{tx}\frac{1}{c}e^{-x/c}\, dx = \frac{1}{c}\int_0^{\infty} e^{(t-1/c)\,x} dx$$

$$= \frac{1}{c}\left[\frac{e^{(t-1/c)x}}{t-\frac{1}{c}}\right]_0^\infty = \frac{1}{c}\left[-\frac{1}{t-\frac{1}{c}}\right] = \frac{1}{1-ct} = (1-ct)^{-1}$$

$$= 1 + ct + c^2t^2 + c^3t^3 + c^4t^4 + \ldots.$$

$$\mu_1' = \frac{d}{dt}[M_0(t)]_{t=0} = [c + 2c^2t + 3c^3t^2 + 4c^4t^3 + \ldots]_{t=0} = c$$

$$\mu_2' = \frac{d^2}{dt^2}[M_0(t)]_{t=0} = [2c^2 + 6c^3t + 12c^4t^2 + \ldots]_{t=0} = 2c^2$$

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

$$\text{S.D.} = c$$

Ans. Mean = c, S.D. = c

11.20 NORMAL DISTRIBUTION

Normal distribution is a continuous distribution. It is derived as the limiting form of the Binomial distribution for large values of n and p and q are not very small.

The normal distribution is given by the equation

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \ldots(1)$$

where μ = mean, σ = standard deviation, $\pi = 3.14159 \ldots$, $e = 2.71828 \ldots$

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

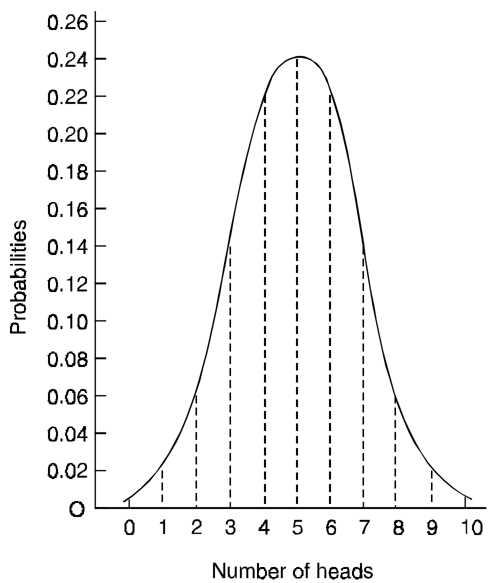
On substitution $z = \frac{x-\mu}{\sigma}$ in (1), we get $f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}$ ...(2)

Here mean = 0, standard deviation = 1.

(2) is known as standard form of normal distribution.

11.21 NORMAL CURVE

(*A.M.I.E., Summer 1995*)



Let us show binomial distribution graphically. The probabilities of heads in 1 tosses are ${}^{10}C_0\, q^{10}\, p^0$, ${}^{10}C_1\, q^9 p^1$, ${}^{10}C_2\, q^8 p^2$, ${}^{10}C_3\, q^7\, p^3$, ${}^{10}C_4\, q^6 p^4$, ${}^{10}C_5\, q^5 p^5$, ${}^{10}C_6\, q^4 p^6$, ${}^{10}C_7\, q^3 p^7$, ${}^{10}C_8\, q^2 p^8$, ${}^{10}C_9\, q^1 p^9$, ${}^{10}C_{10}\, q^0\, p^{10}$.

$p = \frac{1}{2}, q = \frac{1}{2}$. It is shown in the given figure.

If the variates (heads here) are treated as if they were continuous, the required probability curve will be a *normal curve* as shown in the above figure by dotted line.

Properties of the normal curve. $y = y_0\, e^{-\frac{x^2}{2\sigma^2}}$

1. The curve is symmetrical about the y-axis. The mean, median and mode coincide at the origin.

2. The curve is drawn, if mean (origin of x) and standard deviation are given. The value of y_0 can be calculated from the fact that the area of the curve must be equal to the total number of observations.

3. y decreases rapidly as x increases numerically. The curve extends to infinity on either side of the origin.

4. (a) $P\,(\mu - \sigma < x < \mu + \sigma) = 68\%$

(b) $P\,(\mu - 2\sigma < x < \mu + 2\sigma) = 95.5\%$

(c) $P\,(\mu - 3\sigma < x < \mu + 3\sigma) = 99.7\%$

Hence (a) About $\frac{2}{3}$ of the values will lie between $(\mu - \sigma)$ and $(\mu + \sigma)$.

(b) About 95% of the values will lie between $(\mu - 2\sigma)$ and $(\mu + 2\sigma)$.

(c) About 99.7 % of the values will be between $(\mu - 3\sigma)$ and $(\mu + 3\sigma)$.

11.22 MEAN FOR NORMAL DISTRIBUTION

$$\text{Mean} = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2\sigma^2}} .\, x\, dx \qquad \left[\text{Putting}\, \frac{x}{\sigma} = t\right]$$

$$= \frac{1}{\sigma\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\, (t\sigma)\,(\sigma\, dt)$$

$$= \frac{\sigma}{\sqrt{2\,\pi}} \int_{-\infty}^{+\infty} t\; e^{-\frac{t^2}{2}}\, dt = \frac{\sigma}{\sqrt{2\,\pi}} \left[e^{-\frac{t^2}{2}}\right]_{-\infty}^{+\infty}$$

$$= \frac{\sigma}{\sqrt{2\,\pi}}\, [0] = 0$$

11.23 STANDARD DEVIATION FOR NORMAL DISTRIBUTION

$$\mu_2' = \int x^2 .\, f(x)\, dx \quad \text{or} \quad \mu_2' = \int_{-\infty}^{+\infty} x^2 \,.\, \frac{1}{\sigma\sqrt{2\,\pi}}\, e^{-\frac{x^2}{2\sigma^2}}\, dx$$

$$\text{Put} \qquad \frac{x^2}{2\sigma^2} = t \quad \text{or} \quad x = \sqrt{2}\sigma\, t^{1/2} .\, dx = \frac{\sqrt{2}\,\sigma\, dt}{2t^{1/2}}$$

$$\mu_2' = \int_{-\infty}^{+\infty} (2\sigma^2\, t)\, \frac{1}{\sigma\sqrt{2\,\pi}}\, e^{-t} \left(\frac{\sqrt{2}\sigma\, dt}{2t^{1/2}}\right)$$

$$= \frac{2\,\sigma^2}{\sigma\sqrt{(2\,\pi)}}\, \frac{\sqrt{2}\sigma}{2} \int_{-\infty}^{+\infty} t^{\frac{3}{2}-1}\; e^{-t}\, dt, \left[\int_0^{\infty} x^{n-1}\; e^{-x}\, dx = \Gamma n\right]$$

$$= \frac{\sigma^2}{\sqrt{\pi}} \cdot 2\,\Gamma\frac{3}{2} = 2\,\frac{\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2}\,\Gamma\frac{1}{2} = \frac{\sigma^2}{\sqrt{\pi}}\sqrt{\pi} = \sigma^2$$

$$\mu_2 = \mu'_2 - (\mu_1)^2 = \sigma^2 - (0)^2 = \sigma^2$$

$$S.D. = \sigma$$ **Ans.**

11.24 MEDIAN OF THE NORMAL DISTRIBUTION

If a is the median, then it divides the total area into two equal halves so that

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

where $$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Suppose $a >$ mean, μ then

$$\int_{-\infty}^{\mu} f(x)\,dx + \int_{\mu}^{a} f(x)\,dx = \frac{1}{2} \qquad \left[\text{But} \int_{-\infty}^{\mu} f(x)\,dx = \frac{1}{2}\right]$$

$$\frac{1}{2} + \int_{\mu}^{a} f(x)\,dx = \frac{1}{2} \qquad (\mu = \text{mean})$$

$$\int_{\mu}^{a} f(x)\,dx = 0$$

Thus . $a = \mu$

Similarly, when $a <$ mean, we have $a = \mu$.

Thus, median = mean = μ.

Q. Let X be a random variable having a normal distribution.

If $P(X < 0) = P(X > 2) = 0.4$, then mean value of X is.

(*a*) 0 (*b*)1 (*c*) 1.5 (*d*) 2 (*A.M.I.E.T.E., Dec. 2004*) **Ans.** (*b*)

11.25 MEAN DEVIATION ABOUT THE MEAN μ

$$\text{Mean deviation} = E\,|x-\mu|$$

$$= \int_{-\infty}^{\infty} |x-\mu|\,\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

$$= \sigma\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} |z|\, e^{-\frac{z^2}{2}}\,dz \qquad \text{where } z = \frac{x-\mu}{\sigma}$$

$$= \sigma\,\frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{0} -z\,e^{-\frac{z^2}{2}}\,dz + \int_{0}^{\infty} z e^{-\frac{z^2}{2}}\,dz\right]$$

$$= \frac{2\sigma}{\sqrt{2\pi}}\int_0^{\infty} z e^{-\frac{z^2}{2}}\,dz \qquad \text{(as the function is even)}$$

$$= \sigma\sqrt{\frac{2}{\pi}} = \frac{4}{5}\sigma \quad \text{approximately.}$$

11.26 MODE OF THE NORMAL DISTRIBUTION

We know that mode is the value of the variate x for which $f(x)$ is maximum. Thus, by differential calculus $f(x)$ is maximum if $f'(x) = 0$ and $f''(x) < 0$

where $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Clearly $f(x)$ will be maximum when the exponent will be maximum which will be the case when $x = \mu$.

Thus mode is μ, and modal ordinate $= \dfrac{1}{\sigma\sqrt{2\pi}}$

11.27 MOMENT OF NORMAL DISTRIBUTION

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} f(x)\, dx \qquad \textit{(A.M.I.E., Winter 2001)}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2n+1}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1}\, e^{-\frac{z^2}{2}}\, dz \qquad \left[z = \frac{x-\mu}{\sigma}\right]$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1}\, e^{-\frac{z^2}{2}}\, dz$$

$$= 0 \qquad \text{(since } z^{2n+1}\, e^{-\frac{z^2}{2}} \text{ is an odd function)}$$

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x)\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n}\, e^{-\frac{z^2}{2}}\, dz = \frac{\sigma^{2n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n}\, e^{-\frac{z^2}{2}}\, dz$$

$$= \frac{2\sigma^{2n}}{\sqrt{2\pi}} \int_{0}^{\infty} z^{2n}\, e^{-\frac{z^2}{2}}\, dz \qquad \left[z^{2n} \cdot e^{-\frac{z^2}{2}} \text{ is an even function.}\right]$$

$$= \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_{0}^{\infty} e^{-t}\, t^{\left(n-\frac{1}{2}\right)}\, dt \qquad \left[\frac{z^2}{2} = t\right]$$

$$= \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \left\lceil n+\frac{1}{2} \right.$$

Changing n to $(n-1)$, we get

$$\mu_{2n-2} = \frac{2^{n-1}\sigma^{2n-2}}{\sqrt{\pi}} \left\lceil n-\frac{1}{2} \right.$$

On dividing, we get

$$\frac{\mu_{2n}}{\mu_{2n-2}} = 2\sigma^2 \frac{\left\lceil n+\frac{1}{2}\right.}{\left\lceil n-\frac{1}{2}\right.} = \frac{2\sigma^2\left(n-\frac{1}{2}\right)\left\lceil n-\frac{1}{2}\right.}{\left\lceil n-\frac{1}{2}\right.} = 2\sigma^2\left(n-\frac{1}{2}\right)$$

$$\mu_{2n} = \sigma^2 (2n-1)\, \mu_{2n-2}$$

which gives the recurrence relation for the moments of normal distribution

$$\mu_{2n} = [(2n-1)\,\sigma^2]\,[(2n-3)\,\sigma^2]\,\mu_{2n-4}$$
$$= [(2n-1)\,\sigma^2]\,[(2n-3)\,\sigma^2]\,[(2n-5)\,\sigma^2]\,\mu_{2n-6}$$
$$= [(2n-1)\,\sigma^2]\,[(2n-3)\,\sigma^2]\,[(2n-5)\,\sigma^2] - (3\,\sigma^2)\,(1\cdot\sigma^2)\,\mu_0$$
$$= (2n-1)\,(2n-3)\,(2n-5) - ....1\cdot\sigma^{2n}$$
$$= 1.3.5.7\,......\,(2n-5)\,(2n-3)\,(2n-1)\,\sigma^{2n}$$

Example 56. *Fit a normal curve to the following data :*

Length of line (in cm.)	8.60	8.59	8.58	8.57	8.56	8.55	8.54	8.53	8.52
Frequency	2	3	4	9	10	8	4	1	1

Solution. $a = 8.56$

x	f	$d = (x-a)$	fd	fd^2
8.60	2	0.04	0.08	0.0032
8.59	3	0.03	0.09	0.0027
8.58	4	0.02	0.08	0.0016
8.57	9	0.01	0.09	0.0009
8.56	10	0	0	0
8.55	8	– 0.01	– 0.08	0.0008
8.54	4	– 0.02	– 0.08	0.0016
8.53	1	– 0.03	– 0.03	0.0009
8.52	1	– 0.04	– 0.04	0.0016
	$\Sigma f = 42$		$\Sigma fd = 0.11$	$\Sigma fd^2 = 0.0133$

$$\text{Mean } (\mu) = a + \frac{\Sigma fd}{\Sigma f} = 8.56 + \frac{0.11}{42} = 8.56262$$

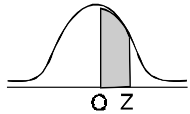
$$S.D.\ (\sigma) = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$$

$$\sigma = \sqrt{\frac{0.0133}{42} - \left(\frac{0.11}{42}\right)^2} = \sqrt{0.00031666 - 0.00000686}$$
$$= 0.0176$$

Hence the equation of the normal curve fitted to the given data is

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}}\,e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty \le x \le \infty$$

where $\mu = 8.56262$, $\sigma = 0.0176$ **Ans.**



Table

Area under standard normal curve from 0 to $\dfrac{x-\mu}{\sigma}$

$\dfrac{x-\mu}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0159	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1084	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2611	.2642	.2671	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4232	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4430	.4441
1.6	.4452	.4463	.4474	.4485	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4762	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4990
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4980	.4980	.4981
2.9	.4981	.4982	.4983	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.49865	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.49903	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993

11.28 AREA UNDER THE NORMAL CURVE

By taking $z = \dfrac{x - \overline{x}}{\sigma}$, the standard normal curve is formed.

The total area under this curve is 1. The area under the curve is divided into two equal parts by $z = 0$. Left hand side area and right hand side area to $z = 0$ is 0.5. The area between the ordinate $z = 0$ and any other ordinate can be noted from the table given on the next page.

Example 57. *On a final examination in mathematics, the mean was 72, and the standard deviation was 15. Determine the standard scores of students receiving grades.*

(*a*) *60* (b) *93* (c) *72.*

Solution.

(*a*) $z = \dfrac{x - \overline{x}}{\sigma} = \dfrac{60 - 72}{15} = -0.8$ (*b*) $z = \dfrac{93 - 72}{15} = +1.4$ (*c*) $z = \dfrac{72 - 72}{15} = 0$ **Ans.**

Example 58. *Find the area under the normal curve in each of the cases*

(a) $z = 0$ *and* $z = 1.2$; *(b)* $z = -0.68$ *and* $z = 0$;

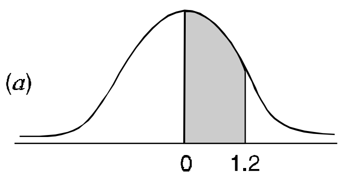
(c) $z = -0.46$ *and* $z = 2.21$; *(d)* $z = 0.81$ *and* $z = 1.94$;

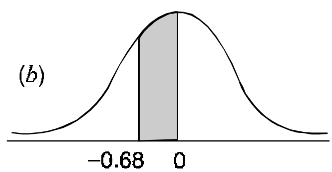
(e) To the left of $z = -0.6$; *(f) Right of* $z = -1.28$.

Solution.

(*a*) Area between $z = 0$ and $z = 1.2$ is .3849

(*b*) Area between $z = 0$ and $z = 0.68$ is 0.2518

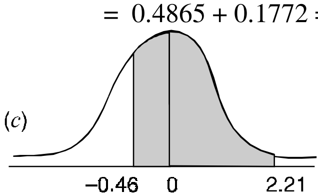


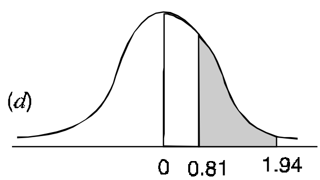


(*c*) Required area = (Area between $z = 0$ and $z = 2.21$) + (Area between $z = 0$ and $z = -0.46$)

= (Area between $z = 0$ and $z = 2.21$) + (Area between $z = 0$ and $z = 0.46$)

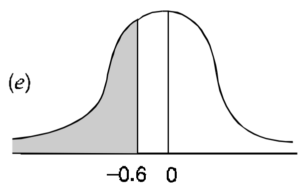
= 0.4865 + 0.1772 = 0.6637.





(*d*) Required area = (Area between $z = 0$ and $z = 1.94$) − (Area between $z = 0$ and $z = 0.81$)

= 0.4738 − 0.2910 = 0.1828



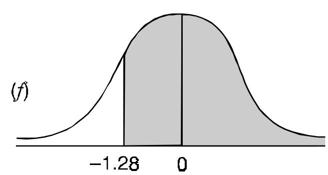
(*e*) Required area = 0.5 − (Area between $z = 0$ and $z = -0.6$)

= 0.5 − 0.2257 = 0.2743

(*f*) Required area = (Area between $z = 0$ and $z = -1.28$) + 0.5

= 0.3997 + 0.5

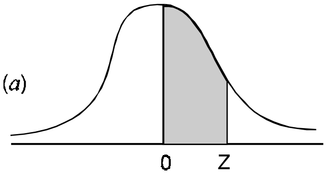
= 0.8997.

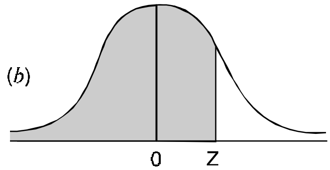


Example 59. *Find the value of z in each of the cases*

(*a*) Area between 0 and z is 0.3770

(*b*) Area to the left of z is 0.8621





Solution.

(*a*) $z = \pm\ 1.16$

(*b*) Since the area is greater than 0.5.

Area between 0 and z.

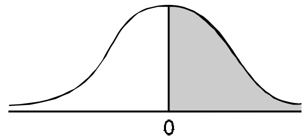
$= 0.8621 - 0.5 = 0.3621$

from which $z = 1.09$ **Ans.**

Example 60. *Students of a class were given an aptitude test. Their marks were found to be normally distributed with mean 60 and standard deviation 5. What percentage of students scored more than 60 marks?*

Solution. $x = 60, \quad \bar{x} = 60, \quad \sigma = 5$

$$z = \frac{x - \bar{x}}{\sigma} = \frac{60 - 60}{5} = 0$$



if $x > 60$ then $z > 0$

Area lying to the right of $z = 0$ is 0.5.

The percentage of students getting more than 60 marks = 50 % **Ans.**

Example 61. *In a sample of 1000 cases, the mean of a certain test is 14 and standard deviation is 2.5. Assuming the distribution to be normal, find*

(*i*) *how many students score between 12 and 15 ?*

(*ii*) *how many score above 18 ?*

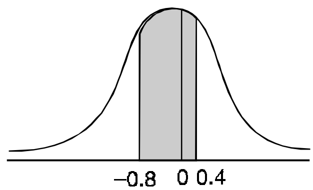
(*iii*) *how many score below 8 ?*

(*iv*) *how many score 16 ?*

Solution. $n = 1000, \ \bar{x} = 14, \ \sigma = 2.5$

(*i*) $$z_1 = \frac{x - \bar{x}}{\sigma} = \frac{12 - 14}{2.5} = -0.8$$

$$z_2 = \frac{15 - 14}{2.5} = \frac{1}{2.5} = 0.4$$



The area lying between – 0.8 to 0.4 = Area from 0 to –0.8 + area from 0 to 0.4

$= 0.2881 + 0.1554 = 0.4435$

The required number of students $= 1000 \times 0.4435 = 443.5 = 444$ (say)

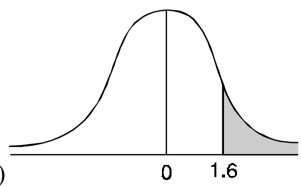
$$z_1 = \frac{18 - 14}{2.5} = \frac{4}{2.5} = 1.6$$

(*ii*) Area right to 1.6 $= 0.5 -$ Area between 0 and 1.6

$= 0.5 - 0.4452 = 0.0548$

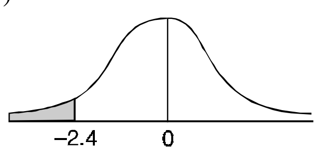
The required number of students

$= 1000 \times 0.0548 = 54.8 = 55$ (say)



(*iii*) $$z = \frac{8 - 14}{2.5} = -\frac{6}{2.5} = -2.4$$

Area left to -2.4 $= 0.5 -$ area between 0 and -2.4

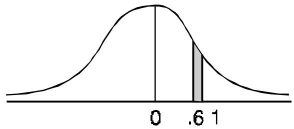


$$= 0.5 - 0.4918 = 0.0082$$

The required number of students $= 1000 \times 0.0082 = 8.2 = 8$ (say)

(*iv*) Area between 15.5 and 16.5

$$z_1 = \frac{15.5 - 14}{2.5} = 0.6$$



and $$z_2 = \frac{16.5 - 14}{2.5} = 1$$

Area between 0.6 and 1 $= 0.3413 - 0.2257 = 0.1156$

The required number of students $= 0.1156 \times 1000$

$= 115.6 = 116$ say **Ans.**

Example 62. *Five thousand candidates appeared in a certain examination carrying a maximum of 100 marks. It was found that the marks were normally distributed with mean 39.5 and with standard deviation 12.5. Determine approximately the number of candidates who secured a first class for which a minimum of 60 marks is necessary. You may see the table given below (x denotes the deviation from the mean).*

The proportion A of the whole area of the normal curve lying to the left of the ordinate at the deviation $\frac{x}{\sigma}$ *is :*

$\frac{x}{\sigma}$	*1.5*	*1.6*	*1.7*	*1.8*
A	*.93319*	*.94520*	*.95543*	*.96407*

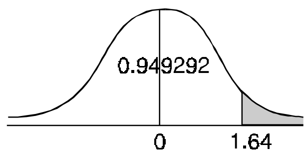
Solution. Mean $= \bar{x} = 39.5$

Standard deviation $= \sigma = 12.5$

$$\frac{x}{\sigma} = \frac{60 - 39.5}{12.5} = \frac{20.5}{12.5} = \frac{41}{25} = 1.64$$

We have to find out area for $\frac{x}{\sigma} = 1.64$.

Area for $\frac{x}{\sigma} = 1.6$ is .94520.



Area for $\frac{x}{\sigma} = 1.7$ is .95543.

Difference for .1 = .01023

Difference for .04 = 0.004092

$\therefore$ Area for $\frac{x}{\sigma} = 1.64$ is $0.94520 + .004092 = 0.949292$.

Area right to ordinate $1.64 = 1 - 0.949292 = 0.050708$

Number of candidates who secured marks 60 or more $= 5000 \times 0.050708 = 253.54$

Candidates securing first division = 254**Ans.**

Example 63. *The mean inside diameter of a sample of 200 washers produced by a machine is 0·502 cm and the standard deviation is 0.005 cm. The purpose for which these washers are intended allows a maximum tolerance in the diameter of 0.496 to 0.508 cm, otherwise the washers are considered defective. Determine the percentage of defective washers produced by the machine, assuming the diameters are normally distributed.*

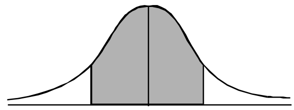
(*A.M.I.E., Summer 2001*)

Solution. $z_1 = \dfrac{x-\overline{x}}{\sigma} = \dfrac{0.496-0.502}{0.005} = -1.2$

$$z_2 = \frac{x-\overline{x}}{\sigma} = \frac{0.508-0.502}{0.005} = +1.2$$

Area for non-defective washers = Area *b*etween $z = -1.2$ and $z = +1.2$

= 2 Area between $z = 0$ and $z = 1.2$.



$= 2 \times (0.3849) = 0.7698 = 76.98\%$

Percentage of defective washers = 100 – 76.98

= 23.02% **Ans.**

Example 64. *A manufacturer of envelopes knows that the weight of the envelopes is normally distributed with mean 1.9 gm and variance 0.01 gm. Find how many envelopes weighing (i) 2 gm or more, (ii) 2.1 gm or more, can be expected in a given packet of 1000 envelopes. [Given : if t is the normal variable, then* $\phi\,(o \le t \le 1) = 0.3413$ *and* $\phi\,(0 \le t \le 2) = 0.4772$*].*

(*A.M.I.E.T.E., Winter 1995*)

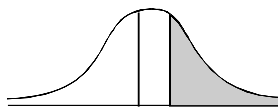
Solution. $\mu = 1.9$ gm, Variance = 0.01 gm

(*i*) $x = 2$ gms or more

$$z = \frac{x-\mu}{\sigma} = \frac{2-1.9}{0.1} = \frac{0.1}{0.1} = 1$$

$P\,(z > 1)$ = Area right to $z = 1$

$= 0.5 - 0.3413 = 0.1587$



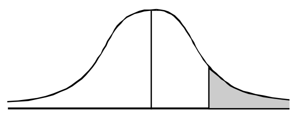
Number of envelopes heavier than 2 gm in a lot of 1000

$= 1000 \times 0.1587 = 158.7 = 159$ (app)

(*ii*) $z = \dfrac{2.1-1.9}{0.1} = \dfrac{0.2}{0.1} = 2$

$P\,(z > 2)$ = Area right to $z = 2$

$= 0.5 - 0.4772 = 0.0228$



Number of envelopes heavier than 2.1 gm in a lot of 1000

$= 1000 \times 0.0228 = 22.8 = 23$ (app.) **Ans.** (*i*) 159 (*ii*) 23

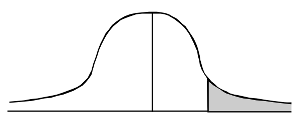
Example 65. *The life of army shoes is 'normally' distributed with mean 8 months and standard deviation 2 months. If 5000 pairs are issued how many pairs would be expected to need replacement after 12 months?* [*Given that* $P\,(z \ge 2) = 0.0228$ *and* $z = (x-\mu)/\sigma$]

Solution. Mean $(\mu) = 8$,

Standard deviation $(\sigma) = 2$

Number of pairs of shoes = 5000

Total months $(x) = 12$



When $z = \dfrac{x-\mu}{\sigma} = \dfrac{12-8}{2} = 2$

Area $(z \ge 2) = 0.0228$

Number of pairs whose life is more than 12 months $(z > 2)$

$$= 5000 \times 0.0228 = 114$$

Replacement after 12 months = 5000 – 114 = 4886 pairs of shoes **Ans.**

Example 66. *In a male population of 1000, the mean height is 68.16 inches and standard deviation is 3.2 inches. How many men may be more than 6 feet (72 inches)?*

[ϕ (1.15) = *0.8749,* ϕ (1.2) = *0.8849,* ϕ (1.25) = *0.8944*]

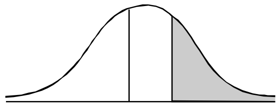
where the argument is the standard normal variable. (*A.M.I.E.T.E., Winter 1997*)

Solution. Male population = 1000

Mean height = 68.16 inches

Standard deviation = 3.2 inches

Men more than 7.2 inches = ?



$$\phi(1.15) = 0.8749, \phi(1.2) = 0.8849$$

$$\phi(1.25) = 0.8944$$

$$z = \frac{x - \bar{x}}{\sigma} = \frac{72 - 68.16}{3.2} = 1.2$$

$$\phi(1.2) = 0.8849$$

ϕ for more than 1.2 = 1 – 0.8849 = 0.1151

Men more than 72 inches = 1000 × 0.1151 = 115.1

= 115 (say) **Ans.**

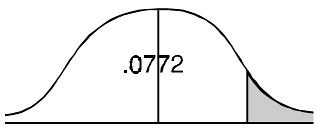
Example 67. *Pipes for tobacco are being packed in fancy plastic boxes. The length of the pipes is normally distributed with μ = 5″* and *σ = 0.1″. The internal length of the boxes is 5.2". What is the probability that the box would be small for the pipe?*

[*given that* ϕ (*1.8*) = *0.9641,* ϕ (2) = *0.9772,* ϕ (2.5) = *0.9938*]

Solution. $\mu = 5''$, $\sigma = 0.1''$, $x = 5.2''$

$$\phi(1.8) = 0.9641, \phi(2) = 0.9722, \phi(2.5) = 0.9938$$

$$z = \frac{x - \mu}{\sigma} = \frac{5.2 - 5}{0.1} = 2$$



$$\phi(2) = 0.9772$$

$$\phi(z > 2) = 1 - 0.9772 = 0.0228$$

The box will be small if the length of the pipe is more than 5.2" (z = 2).

Hence the probability is 0.0228 **Ans.**

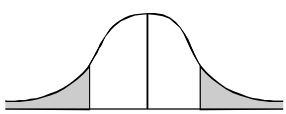
Example 68. *Assuming that the diameters of 1,000 brass plugs taken consecutively from a machine form a normal distribution with mean 0.7515 cm and standard deviation 0.0020 cm, how many of the plugs are likely to be rejected if the approved diameter is 0.752 ± .004 cm ?*

Solution. Tolerance limits of the diameter of non-defective plugs are

$$0.752 - 0.004 = 0.748 \text{ cm and}$$

$$0.752 + 0.004 = 0.756 \text{ cm}$$

$$z = \frac{x - \mu}{\sigma}$$



If x_1 = 0.748, $$z_1 = \frac{0.748 - 0.7515}{0.002} = -1.75$$

If $x_2 = 0.756$, $\qquad z_2 = \dfrac{0.756 - 0.7515}{0.002} = 2.25$

Area under $\qquad z_1 = -1.75$ to $z_2 = 2.25$

$$= (\text{Area from } z = 0 \text{ to } z_1 = -1.75) + (\text{Area from } z = 0 \text{ to } z_2 = 2.25)$$

$$= 0.4599 + 0.4878 = 0.9477$$

Number of plugs likely to be rejected

$$= 1000\,(1 - 0.9477)$$

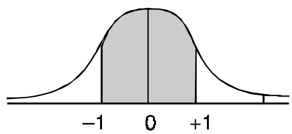
$$= 1000 \times .0523 = 52.3$$

Approximately 52 plugs are likely to be rejected. **Ans.**

Example 69. *A manufacturer knows from experience that the resistance of resistors he produces is normal with mean $\mu = 100$ ohms and standard deviation $\sigma = 2$ ohms. What percentage of resistors will have resistance between 98 ohms and 102 ohms ?*

(*A.M.I.E.T.E., Winter 2003*)

Solution. $\mu = 100$ ohms, $\sigma = 2$ ohms



$$x_1 = 98,\ x_2 = 102$$

$$z = \frac{x - \mu}{\sigma},\ z_1 = \frac{98 - 100}{2} = -1$$

$$z_2 = \frac{102 - 100}{2} = +1$$

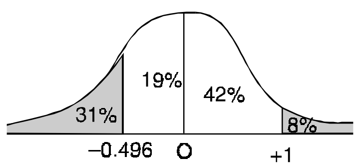
Area between $z_1 = -1$ and $z_2 = +1$

$$= (\text{Area betweem } z = 0 \text{ and } z = -1)$$
$$+ (\text{Area between z = 0 and z = + 1})$$
$$= 2\,(\text{Area between } z = 0 \text{ and } z = +1) = 2 \times 0.3413 = 0.6826$$

Percentage of resistors having resistance between 98 ohms and 102 ohms = 68.26 **Ans.**

Example 70. *In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.* (*A.M.I.E., Winter 1998*)

Solution. Let μ be the mean and σ the S.D.



If $x = 45$, $\qquad z = \dfrac{45 - \mu}{\sigma}$

If $x = 64$, $\qquad z = \dfrac{64 - \mu}{\sigma}$

Area between 0 and $\dfrac{45 - \mu}{\sigma} = 0.50 - .31 = 0.19$

[From the table, for the area 0.19, $z = 0.496$]

$$\frac{45 - \mu}{\sigma} = -0.496 \qquad \text{...(1)}$$

Area between $z = 0$ and $z = \dfrac{64 - \mu}{\sigma} = 0.5 - 0.08 = 0.42.$

(From the table, for area 0.42, $z = 1.405$)

$$\frac{64 - \mu}{\sigma} = 1.405 \qquad \text{...(2)}$$

Solving (1) and (2) we get $\mu = 50, \sigma = 10.$ **Ans.**

Exercise 11.7

1. In a regiment of 1000, the mean height of the soldiers is 68.12 units and the standard deviation is 3.374 units. Assuming a normal distribution, how many soldiers could be expected to be more than 72 units? It is given that

   $P(z = 1.00) = 0.3413$, $P(z = 1.15) = 0.3749$ and

   $P(z = 1.25) = 0.3944$, where z is the standard normal variable. **Ans.** 125

2. If the height of 300 students are normally distributed with mean 64.5 inches and standard deviation 3.3 inches, find the height below which 99% of the students lie. **Ans.** 68.7295 inches

3. The lifetime of radio tubes manufactured in a factory is known to have an average value of 10 years. Find the probability that the lifetime of a tube taken randomly (*i*) exceeds 15 years, (*ii*) is less than 5 years, assuming that the exponential probability law is followed. **Ans.** (*i*) 0.2231, (*ii*) 0.3935.

4. Analysis of past data shows that hub thickness of a particular type of gear is normally distributed about a mean thickness of 2.00 cm with a standard deviation of 0.04 cm.

   (i) What is the probability that a gear chosen at random will have a thickness greater than 2.06 cm?

   (*ii*) How many gears will have a thickness between 1.89 and 1.95 cm?

   Given $\phi(1.5) = 0.4332$, $\phi(2.75) = 6.4970$, $\phi(1.25) = 0.3944$.

   (*A.M.I.E., Summer 2001*) **Ans.** (*i*) 0.068 (*ii*) 62

5. The breaking strength X of a cotton fabric is normally distributed with $E(x) = 16$ and $\sigma(x) = 1$. The fabric is said to be good if $X \geq 14$. What is the probability that a fabric chosen at random is good. Given that $\phi(2) = 0.9772$ **Ans.** 0.9772

6. A manufacturer knows from experience that the resistance of resistors he produces is normal with mean $\mu = 140\ \Omega$ and standard deviation $\sigma = 5\ \Omega$. Find the percentage of the resistors that will have resistance between 138 Ω and 142 Ω . (given $\phi(0.4) = 0.6554$, where z is the standard normal variate). **Ans.** 31.08%

7. A manufacturing company packs pencils in fancy plastic boxes. The length of the pencils is normally distributed with $\mu = 6"$ and $\sigma = 0.2"$. The internal length of the boxes is 6.4". What is the probability that the box would be too small for the pencils (Given that a value of the standardized normal distribution function is $\phi(2) = 0.9772$). **Ans.** 0.0228.

8. A manufacturer produces airmail envelopes, whose weight is normal with mean $\mu = 1.95$ gm and standard deviation $\sigma = 0.05$ gm. The envelopes are sold in lots of 1000. How many envelopes in a lot will be heavier than 2 gm? Use the fact that

   $$\frac{1}{\sqrt{2\pi}} \int_0^1 \exp(-x^2/2)\, dx = 0.3413$$ **Ans.** 159

9. A sample of 100 dry battery cells tested to find the length of life produced the following results. $\bar{x} = 12$ hours, $\sigma = 3$ hours. Assuming the data to be normally distributed, what percentage of battery cells expected to have life?

   (*i*) more than 15 hours. (*ii*) less than 6 hours.

   [Given $P(0 < z < 1) = 0.3413$ and $P(0 < z < 2) = 0.4772$.] **Ans.** (*i*) 15.87% (*ii*) 2.28%

10. Find the mean and variance of the density function $f(x) = \lambda e^{-\lambda x}$. **Ans.** $\frac{1}{\lambda}, \frac{1}{\lambda^2}$.

11. Fill in the blanks :

    (*a*) The mean of the marks obtained by the students is 50 and the variance is 25. If a student gets 60 marks, his standard score is ....

    (*b*) If $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, then its mean is .... and standard deviation is ....

    (*c*) In the standard normal curve the area between z = – 1 and z = 1 is nearly ....

    (*d*) If $\sigma = 2$, $\bar{x} = 5$, the equation of normal distribution is ....

(*e*) The marks obtained were found normally distributed with mean 75 and variance 100 . The percentage of students who scored more than 75 marks is ....

(*f*) The mean, median and mode of a normal distribution are .... (*A.M.I.E., Summer 2000*)

(*g*) Exponential distribution $f(x)$ is defined by $f(x) = a\,e^{-2x}$, $0 < x < \infty$, then a = ...

(*h*) The probability density function of beta distribution with $\alpha = 1, \beta = 4$ is $f(x)$ = ...

(*i*) For a standard normal variate z $P(-0.72 \le z \le 0)$ = ... (*A.M.I.E., Winter 1997*) **Ans.** 0.2642

Ans. (*a*) 2 , (*b*) 0.1 , (*c*) 68% , (*d*) $f(x) = \dfrac{1}{2\sqrt{2\pi}} e^{-\frac{(x-5)^2}{8}}$, (*e*) 50% , (*f*) zero, (*g*) 2, (*h*) $4(1-x)^3$

12. The probability density function $f(x)$ of a continuous random variable x is defined by

$$f(x) = \frac{A}{x^3}, \quad 5 \le x \le 10$$
$$= 0, \quad \text{elsewhere}$$

The value of A is

(*i*) 50, (*ii*) 1, (*iii*) – 200, (*iv*) $\dfrac{200}{3}$. **Ans.** (*iv*)

13. The cumulative distribution function F of a continuous variate X is such that $F(a) = 0.5$ and $F(b) = 0.7$. then value of $P(a \le X \le b)$ is given as

(*a*) 0 (*b*) 0.5 (*c*) 0.2 (*d*) 0.7 (*A.M.I.E.T.E. Dec, 2005*)

14. A discrete random variate X has probability mass function f which is positive at x = – 1, 0, 1 and is zero elsewhere. If $f(0) = \dfrac{1}{2}$, then the value of $E(x^2)$ is

(*a*) 1 (*b*) 0 (*c*) $= \dfrac{1}{2}$ (*d*) $= -\dfrac{1}{2}$ (*A.M.I.E.T.E. Dec. 2005*)

15. If x is normally distributed with mean 1 and variance 4,
(*i*) Find $Pr(-3 \le x \le 3)$; (*ii*) Obtain k if $Pr(x \le k) = 0.90$ **Ans.** (*i*) .8185, (ii) 3.56.

16. A normal variable × has mean 1 and variance 4. Find the probability that $x \ge 3$. (Given: z is the standard normal variable and $\phi(0) = 0.5$, $\phi(0.5) = 0.6915$, $\phi(1) = 0.8413$, $\phi(1.5) = 0.9332$) **Ans.** 0.1587

17. (*a*) If x is normally distributed with mean 4 and variance 9; find
(*i*) $Pr(2.5 \le x \le 5.5)$. (*ii*) Obtain k if $Pr(x \le k) = 0.9$.
Use $Pr(z \le .5) = 0.691$ and $Pr(z \le 1.3) = 0.90$. **Ans.** (*i*) 0.382 (*ii*) 7.9.

(*b*) If $\log_e x$ is normally distributed with mean 1 and variance 4, find $P\left(\dfrac{1}{2} < x < 2\right)$, given that $\log_e 2 = 0.693$.

Ans. 0.24.

(*c*) For a standard normal variate z $P(-0.72 \le z \le 0)$ = ... **Ans.** 0.2642

18. The random variable x is normally distributed with $E(x) = 2$ and variance $V(x) = 4$. Find a number p (approximately), such that $P(x > p) = 2\,P(x \le p)$. [The values of the standard normal distribution are $\phi(-0.43) = 0.3336$, and $\phi(-0.44) = 0.3300$]. (*A.M.I.E.T.E., Summer 1995*) **Ans.** 1.13834

19. The continuous random variable x is normally distributed with $E(x) = \mu$ and $V(x) = \sigma^2$. If $Y = cx + d$, then find *V(Y).* **Ans.** $c^2\sigma^2$

20. The pdf of X is given by $f(X) = \lambda e^{-\lambda X}$, $x \ge 0, \lambda > 0$.
Calculate $Pr[X > E(X)]$. If $X \sim N(75, 25)$, find $Pr[X > 80/X > 77]$.

If $X \sim N(10, 4)$ find $Pr[|X| \ge 5]$. **Ans.** $\dfrac{1}{e}$, $\dfrac{1}{5\sqrt{2}\,\pi} e^{-\frac{(x-75)^2}{2(0.5)}}$, 0.062

21. If the resistance X of certain wires in an electrical networks have a normal distribution with mean of 0.01 ohm and a standard deviation of 0.001 ohm, and specification requires that the wires should have resistance between 0.009 ohm and 0.011 ohms, then find the expected number of wires in a sample of 1000 that are within the specification. Also, find the expected number among 1000 wires that cross the upper specification.

(You may use normal table values $\Phi(.5) = .6915$, $\Phi(1) = .8413$, $\Phi(1.5) = .9332$, $\Phi(2) = .9772$)

(*A.M.I.E.T.E., Dec. 2004*)

22. A random variable x has a standard normal distribution ϕ. Prove : $\Pr(|x| > k) = 2[1 - \phi(k)]$

23. The random variable x has the probability density function

$$f(x) = kx \text{ if } 0 \le x \le 2$$

Find k. Find x such that

(*i*) $Pr(X \le x) = 0.1$ (*ii*) $Pr(X \le x) = 0.95$ **Ans.** $k = \frac{1}{2}$, (*i*) $x = 0.632$ (*ii*) $x = 1.949$

24. For a normal curve, show that $\mu_{2n+1} = 0$ and $\mu_{2n} = (2n-1)\sigma^2 \mu_{2n-2}$.

25. In a normal distribution, 7% of the items are under 35 and 89% are under 63. Determine mean and variance of distribution. [Area of z for 0.43 = 1.48. Area of z for 0.39 = 1.23]

(*A.M.I.E.T.E., Winter 2001*) **Ans.** $\mu = 50.29$, $\sigma^2 = 106.73$

26. The length of an item manufactured on an automatic machine tool is a normally distributed randam variable with parameters $m(\overline{x}) = 10$, and $\sigma^2 = \frac{1}{200}$. Find the probability of defective production of the tolerance is 10 ± 0.05. (*A.M.I.E.T.E., Winter 2001*) **Ans.** 0.04798

27. In a mathematics examination, the average grade was 82 and the standard deviation was 5. All the students with grades from 88 to 94 received a grade B. If the grades are normally distributed and 8 students received a B grade, find how many students took the examination.

Given:

x/6	1.20	2.00	2.40	2.45
A	0.3849	0.4772	0.4918	0.4929

(*A.M.I.E., Winter 2001*) **Ans.** 75 students

28. The income of a group of 10,000 persons was found to be normally distributed with mean Rs. 750 p.m. and standard deviation of Rs. 50. Show that, of this group, about 95% had income exceeding Rs. 668 and only 5% had income exceeding Rs. 832. Also find the lowest income among the richest 100. (*U.P. III Semester, Dec. 2004*) **Ans.** Rs. 866

29. A continuous type random variable X has probability density $f(x)$ which is proportional to x^2 and X takes values in the interval [0, 2]. Find the distribution function of the random variable use this to find $P(X > 1.2)$ and conditional probability $P(X > 1.2/X > 1)$. (*A.M.I.E.T.E., Dec. 2006*)

11.29 OTHER DISTRIBUTIONS

(1) Uniform (or Rectangular) Distribution

Here $p(x) = \frac{1}{n}$ $\Sigma p(x) = n\left(\frac{1}{n}\right) = 1$

The value of probability for all variates $x_1, x_2 \ldots\ldots x_n$ is the same $\frac{1}{n}$.

(2) Geometric Distribution

Let r be the number of failures preceding the first success

$$\begin{aligned} p(r) &= q^r p \quad \text{where } r = 0,1,2,3\ldots\ldots,\ q = 1-p \\ \Sigma p(r) &= \Sigma q^r p \\ &= p(1 + q + q^2 + \ldots\ldots + q^r \ldots\ldots) \quad \text{(Geometric series)} \\ &= p\frac{1}{1-q} = \frac{p}{p} = 1 \end{aligned}$$

$$\text{Mean} = \frac{q}{p}, \quad \text{Variance} = \frac{q}{p^2}$$

(3) Negative Binomial Distribution

The probability of the event that occurs for the kth time on the rth trial

$$p(k,r) = {}^{r-1}C_{k-1}\, p^k q^{r-k}$$

For $k = 1$, the negative Binomial distribution becomes geometric distribution.

(4) Hypergeometric Distribution

Let the number y white balls be m and n black balls in a bag. If r balls are drawn at a time with replacement

$$p\,(k \text{ white}) = {}^{m}C_k \,\frac{{}^{m}C_{r-k}}{{}^{m+n}C_r} \quad \text{where } k = 0,1,2,\ldots.\ r,\ r \le m,\ r \le n$$

$$\sum_{k=0}^{r} p\,(k) = 1 \text{ since } \Sigma\ {}^{m}C_k\,{}^{n}C_{r-k} = (m+n)\,C_r$$

(5) Exponential Distribution

Let $f(x)$ be a continuous distribution

$$f(x) = e^{-cx} \qquad \text{from } x > 0$$

Hence mean $= \dfrac{1}{C}$ = standard deviation $= \dfrac{1}{C}$

(6) Weibull Distribution is given by $f(x) = \dfrac{\alpha}{C} x^{\alpha-1} e^{-\frac{x^{\alpha}}{C}},\ x > 0,\ c > 0$

where C is a scale parameter and α is a shape parameter.

This distribution is used for

(1) variation in the fatigue resistance of steel and its elastic limits.

(2) variation of length of service of radio service equipment.

Exercise 11.8

Find the mean and variance for the following distributions

1. Rectangular distribution **Ans.** $\dfrac{1}{2}, \dfrac{1}{12}$
2. Uniform distribution $f(x) = \dfrac{1}{n},\ x = 1, 2, \ldots n$ **Ans.** $\dfrac{1}{2}(n+1), \dfrac{1}{12}(n^2-1)$
3. Geometric distribution $p\,(r) = 2^r,\ r = 1,2,3 \ldots$
4. Exponential distribution $p\,(x) = \lambda\, e^{-\lambda x}$ **Ans.** $\dfrac{1}{\lambda}, \dfrac{1}{\lambda}$

SAMPLING OF VARIABLES

11.30 POPULATION (Universe)

Before giving the notion of sampling, we will first define *population*. The group of individuals under study is called *population* or *universe*. It may be finite or infinite.

11.31 SAMPLING

A part selected from the population is called *a sample*. The process of selection of a sample is called sampling. A *Random sample* is one in which each member of population has an equal chance of being included in it. There are ${}^{N}C_n$ different samples of size n that can be picked up from a population of size N.

11.32 PARAMETERS AND STATISTICS

The statistical constants of the population such as mean (μ), standard deviation (σ) are called parameters. Parameters are denoted by Greek letters.

The mean $(\bar{x})$, standard deviation $|\,S\,|$ of a sample are known as statistics. Statistics are denoted by Roman letters.

Symbols for Population and Samples

Characteristic	Population	Sample
	Parameter	Statistic
Symbols	population size $= N$ population mean $= \mu$ population standard deviation $= \sigma$ population proportion $= p$	sample size $= n$ sample mean $= \bar{x}$ sample standard deviation $= s$ sample proportion $= \tilde{p}$

11.33 AIMS OF A SAMPLE

The population parameters are not known generally. Then the sample characteristics are utilised to approximately determine or estimate of the population. Thus, static is an estimate of the parameter. To what extent can we depend on the sample estimates?

The estimate of mean and standard deviation of the population is a primary purpose of all scientific experimentation. The logic of the sampling theory is the logic of *induction*. In induction, we pass from a particular (sample) to general (population). This type of generalization here is known as *statistical inference*. The conclusion in the sampling studies are based not on certainties but on probabilities.

11.34 TYPES OF SAMPLING

Following types of sampling are common:

(1) Purposive sampling (2) Random sampling

(3) Stratified sampling (4) Systematic sampling

11.35 SAMPLING DISTRIBUTION

From a population a number of samples are drawn of equal size n. Find out the mean of each sample. The means of samples are not equal. The means with their respective frequencies are grouped. The frequency distribution so formed is known as *sampling distribution of the mean*. Similarly, sampling distribution of standard deviation we can have.

11.36. STANDARD ERROR (S.E.) is the standard deviation of the sampling distribution. For assessing the difference between the expected value and observed value, standard error is used. Reciprocal of standard error is known as *precision*.

11.37 SAMPLING DISTRIBUTION OF MEANS FROM INFINITE POPULATION

Let the population be infinitely large and having a population mean of μ and a population variance of σ^2. If x is a random variable denoting the measurement of the characteristic, then

Expected value of x, $E(x) = \mu$

Variance of x, Var $(x) = \sigma^2$

The sample mean $\overline{x}$ is the sum of n random variables, *viz*., $x_1, x_2, ..., x_n$, each being divided by n. Here, $x_1, x_2, ..., x_n$ are independent random variables from the infinitely large population.

$$\therefore \qquad E(x_1) = \mu \quad \text{and} \quad Var(x_1) = \sigma^2$$

$$E(x_2) = \mu \quad \text{and} \quad Var(x_2) = \sigma^2 \text{ and so on}$$

Finally

$$\begin{aligned} E(\overline{x}) &= E\left[\frac{x_1 + x_2 + \dots + x_n}{n}\right] \\ &= \frac{1}{n} E\, x_1 + \frac{1}{n} E(x_2) + \dots + \frac{1}{n} E(x_n) \\ &= \frac{1}{n}\mu + \frac{1}{n}\mu + \dots + \frac{1}{n}\mu \\ &= \mu \end{aligned}$$

and

$$\begin{aligned} Var(\overline{x}) &= Var\left[\frac{x_1 + x_2 + \dots + x_n}{n}\right] \\ &= Var\left(\frac{x_1}{n}\right) + Var\left(\frac{x_2}{n}\right) + \dots + Var\left(\frac{x_n}{n}\right) \\ &= \frac{1}{n^2} Var(x_1) + \frac{1}{n^2} Var(x_2) + \dots + \frac{1}{n^2} Var(x_n) \end{aligned}$$

$$= \frac{1}{n^2} \cdot \sigma^2 + \frac{1}{n^2} \cdot \sigma^2 + \ldots + \frac{1}{n^2} \cdot \sigma^2$$

$$= \frac{n\,\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

The expected value of the sample mean is the same as population mean. The variance of the sample mean is the variance of the population divided by the sample size.

The average value of the sample tends to true population mean. If sample size (n) is increased then variance of $\bar{x}$, $\left(\frac{\sigma^2}{n}\right)$ gets reduced, by taking large value of n, the variance $\left(\frac{\sigma^2}{n}\right)$ of $\bar{x}$ can be made as small as desired. The standard deviation $\left(\frac{\sigma}{\sqrt{n}}\right)$ of $\bar{x}$ is also called **standard error of the mean**. It is denoted by $\sigma_{\bar{x}}$.

Sampling with Replacement

When the sampling is done with replacement, so that the population is back to the same form before the next sample member is picked up. We have

$$E(\bar{x}) = \mu$$

$$Var(\bar{x}) = \frac{\sigma^2}{n} \quad \text{or} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Sampling without replacement from Finite population

When a sample is picked up without replacement from a finite population, the probability distribution of second random variable depends on the outcome of the first pick up. n sample members do not remain independent. Now we have

$$E(\bar{x}) = \mu$$

and $$Var(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \quad \text{or} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \frac{\sigma}{\sqrt{n}} \text{ app.} \qquad \left(\text{if } \frac{n}{N} \text{ is very small}\right)$$

Sampling from Normal Population

If $x \sim N(\mu, \sigma^2)$ then it follows that $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

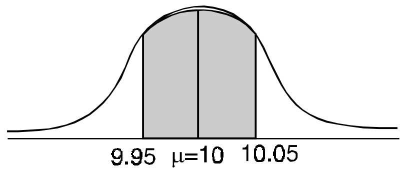
Example 71. *The diameter of a component produced on a semi-automatic machine is known to be distributed normally with a mean of 10 mm and a standard deviation of 0.1 mm. If we pick up a random sample of size 5, what is the probability that the same mean will be between 9.95 and 10.05 mm?*

Solution. Let x be a random variable representing the diameter of one component picked up at random.

Here $x \sim N(10, 0.01)$, Therefore, $\bar{x} \sim N\left(10, \frac{0.01}{5}\right)$ $\qquad \left[\bar{x} = N\left(\bar{x}, \frac{\sigma^2}{n}\right)\right]$

$$Pr\{9.95 \le \bar{x} \le 10.05\} = 2 \times Pr\{10 \le \bar{x} \le 10.05\} \qquad \left\{z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right\}$$

$$= 2 \times Pr\left\{\frac{10-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{10.05-\mu}{\frac{\sigma}{\sqrt{n}}}\right\}$$



$$= 2 \times Pr\left\{0 \le z \le \frac{10.05-10}{\frac{0.1}{\sqrt{5}}}\right\}$$

$$= 2 \times Pr\{0 \le z \le 1.12\}$$

$$= 2 \times 0.3686$$

$$= 0.7372$$ **Ans.**

Similar Question

A sample of size 25 is picked up at random from a population which is normally distributed with a mean 100 and a variance of 36. Calculate (*a*) $Pr\{\bar{x} \le 99\}$, (*b*) $Pr\{98 \le \bar{x} \le 100\}$

Ans. (*a*) 0.2023 (*b*) 0.4522

11.38 SAMPLING DISTRIBUTION OF THE VARIANCE

We use a sample statistic called the sample variance to estimate the population variance. The sample variance is usually denoted by s^2

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

11.39 TESTING A HYPOTHESIS

On the basis of sample information, we make certain decisions about the population. In taking such decisions we make certain assumptions. These assumptions are known as *statistical hypothesis*. These hypothesis are tested. Assuming the hypothesis correct we calculate the probability of getting the observed sample. If this probability is less than a certain assigned value, the hypothesis is to be rejected.

11.40 NULL HYPOTHESIS (H_0)

Null hypothesis is based for analysing the problem. Null hypothesis is the *hypothesis of no difference*. Thus, we shall persume that there is no significant difference between the observed value and expected value. Then, we shall test whether this hypothesis is satisfied by the data or not. If the hypothesis is not approved the difference is considered to be significant. If hypothesis is approved then the difference would be described as due to sampling fluctuation. Null hypothesis is denoted by H_0.

11.41 ERRORS

In sampling theory to draw valid inferences about the population parameter on the basis of the sample results.

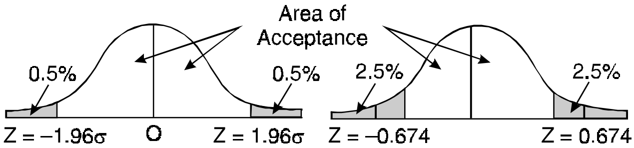
We decide to accept or to reject the lot after examining a sample from it. As such, we are liable to commit the following two types of errors.

Type 1 Error. If H_0 is rejected while it should have been accepted.

Type II Error. If H_0 is accepted while it should have been rejected.

11.42 LEVEL OF SIGNIFICANCE

There are two critical regions which cover 5% and 1% areas of the normal curve. The shaded portions are the critical regions.



1% level of significance 5% level of significance

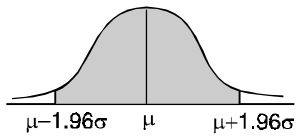
Thus, the probability of the value of the variate falling in the critical region is the level of significance. If the variate falls in the critical area, the hypothesis is to be rejected.

11.43 TEST OF SIGNIFICANCE

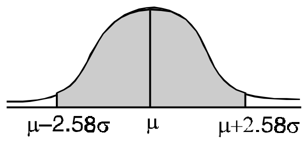
The tests which enables us to decide whether to accept or to reject the null hypothesis is called the tests of significance. If the difference between the sample values and the population values are so large (lies in critical area), it is to be rejected.

11.44 CONFIDENCE LIMITS

$\mu - 1.96\,\sigma$, $\mu + 1.96\,\sigma$ are 95% confidence limits as the area between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 95%. If a sample statistics lies in the interval $\mu - 1.96\,\sigma$, $\mu + 1.96\sigma$, we call 95% confidence interval.



Similarly, $\mu - 2.58\,\sigma$, $\mu + 2.58\,\sigma$ is 99% confidence limits as the area between $\mu - 2.58\,\sigma$ and $\mu + 2.58\,\sigma$ is 99%. The numbers 1.96, 2.58 are called confidence coefficients.



11.45 TEST OF SIGNIFICANCE OF LARGE SAMPLES ($N > 30$)

Normal distribution is the limiting case of Binomial distribution when n is large enough. For normal distribution 5% of the items lie outside $\mu \pm 1.96\,\sigma$ while only 1% of the items lie outside $\mu \pm 2.586\,\sigma$.

$$z = \frac{x - \mu}{\sigma}$$

where z is the standard normal variate and x is the observed number of successes.

First we find the value of z. Test of significance depends upon the value of z.

(*i*) (*a*) If $|z| < 1.96$, difference between the observed and expected number of successes is not significant at the 5% level of significance.

(*b*) If $|z| > 1.96$, difference is significant at 5% level of significance.

(*ii*) (*a*) If $|z| < 2.58$, difference between the observed and expected number of successes is not significant at1% level of significance.

(*b*) If $|z| > 2.58$, difference is significant at 1% level of significance.

Example 72. *A cubical die was thrown 9,000 times and 1 or 6 was obtained 3120 times. Can the deviation from expected value lie due to fluctuations of sampling?*

Solution. Let us consider the hypothesis that the die is an unbiased one and hence the probability of obtaining 1 or 6 $= \frac{2}{6} = \frac{1}{3}$ *i.e.*, $P = \frac{1}{3}$, $q = \frac{2}{3}$

The expected value of the number of successes $= np = 9000 \times \frac{1}{3} = 3000$

Also $\sigma = \text{S.D.} = \sqrt{npq} = \sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}} = \sqrt{2000} = 44.72$

$$3\,\sigma = 3 \times 44.72 = 134.16$$

Actual number of successes = 3120

Difference between the actual number of successes and expected number of successes

$$= 3120 - 3000 = 120 \text{ which is } < 3\,\sigma$$

Hence, the hypothesis is correct and the deviation is due to fluctuations of sampling due to random causes. **Ans.**

11.46 SAMPLING DISTRIBUTION OF THE PROPORTION

A simple sample of n items is drawn from the population. It is same as a series of n independent trials with the probability p of success. The probabilities of 0, 1, 2, ..., n success are the terms in the binomial expansion of $(q+p)^n$.

Here mean = np and standard deviation = $\sqrt{npq}$.

Let us consider the proportion of successes, then

(a) Mean proportion of successes = $\dfrac{np}{n} = p$

(b) Standard deviation (standard error) of proportion of successes = $\dfrac{\sqrt{npq}}{n} = \sqrt{\dfrac{pq}{n}}$

(c) Precision of the proportion of success = $\dfrac{1}{\text{S.E.}} = \sqrt{\dfrac{n}{pq}}$.

Example 73. *A group of scientific mens reported 1705 sons and 1527 daughters. Do these figures conform to the hypothesis that the sex ratio is* $\frac{1}{2}$.

Solution. The total number of observations = 1705 + 1527 = 3232

The number of sons = 1705

Therefore, the observed male ratio = $\dfrac{1705}{3232} = 0.5175$

On the given hypothesis the male ratio = 0.5000

Thus, the difference between the observed ratio and theoretical ratio

$$= 0.5275 - 0.5000$$
$$= 0.0275$$

The standard deviation of the proportion $= s = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{\frac{1}{2} \times \frac{1}{2}}{3232}} = 0.0088$

The difference is more than 3 times of standard deviation.

Hence, it can be definitely said that the figures given do not conform to the given hypothesis.

11.47 ESTIMATION OF THE PARAMETERS OF THE POPULATION

The mean, standard deviation etc. of the population are known as parameters. They are denoted by μ and σ. Their estimates are based on the sample values. The mean and standard deviation of a sample are denoted by $\bar{x}$ and s respectively. Thus, a static is an estimate of the parameter. There are two types of estimates.

(i) *Point estimation*: An estimate of a population parameter given by a single number is called a point estimation of the parameter. For example,

$$(\text{S.D.})^2 = \frac{\Sigma\,(x - \bar{x})^2}{n - 1}$$

(*ii*) *Interval estimation*: An interval in which population parameter may be expected to lie with a given degree of confidence. The intervals are

(*i*) $\bar{x} - \sigma_s$ to $\bar{x} + \sigma_s$ (68.27% confidence level)

(*ii*) $\bar{x} - 2\,\sigma_s$ to $\bar{x} + 2\,\sigma_s$ (95.45% confidence level)

(*iii*) $\bar{x} - 3\,\sigma_s$ to $\bar{x} + 3\,\sigma_s$ (99.73% confidence level)

$\bar{x}$ and σ_s are the mean and S.D. of the sample.

Similarly, $\bar{x} \pm 1.96\,\sigma_s$, $\bar{x} \pm 2.58\,\sigma_s$ are 95% and 99% confidence of limits for μ.

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ and $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$ are also the intervals as $\sigma_s = \frac{\sigma}{\sqrt{n}}$.

11.48 COMPARISON OF LARGE SAMPLES

Let two large samples of size n_1, n_2 be drawn from two populations of proportions of attributes A's as P_1, P_2 respectively.

(*i*) *Hypothesis*: As regards the attribute A, the two populations are similar. On combining the two samples

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

where p is the common proportion of attributes.

Let e_1, e_2 be the standard errors in the two samples, then

$$e_1^2 = \frac{pq}{n_1} \text{ and } e_2^2 = \frac{pq}{n_2}$$

If e be the standard error of the combined samples, then

$$e = P_1^2 + P_2^2 = \frac{pq}{n_1} + \frac{pq}{n_2} = pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

$$z = \frac{P_1 - P_2}{e}$$

1. If $z > 3$, the difference between P_1 and P_2 is significant.
2. If $z < 2$, the difference may be due to fluctuations of sampling.
3. If $2 < z < 3$, the difference is significant at 5% level of significance.

(*ii*) *Hypothesis*. In the two populations, the proportions of attribute A are not the same, then standard error e of the difference $p_1 - p_2$ is

$$e^2 = p_1 + p_2$$

$$= \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}, \quad z = \frac{P_1 - P_2}{e} < 3,$$

difference is due to fluctuations of samples.

Example 74. *In a sample of 600 men from a certain city, 450 are found smokers. In another sample of 900 men from another city, 450 are smokers. Do the data indicate that the cities are significantly different with respect to the habit of smoking among men.*

Solution.

$$n_1 = 600 \text{ men, Number of smokers} = 450, \; P_1 = \frac{450}{600} = 0.75$$

$$n_2 = 900 \text{ men, Number of smokers} = 450, \; P_2 = \frac{450}{900} = 0.50$$

$$p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{600 \times 0.75 + 900 \times 0.5}{600 + 900} = \frac{900}{1500} = 0.60$$

$$q = 1 - p = 1 - 0.6 = 0.4$$

$$e^2 = P_1^2 + P_2^2 = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

$$e^2 = 0.6 \times 0.4\left(\frac{1}{600} + \frac{1}{900}\right) = 0.000667$$

$$e = 0.02582$$

$$z = \frac{P_1 - P_2}{e} = \frac{0.75 - 0.50}{0.02582} = 9.682$$

$z > 3$ so that the difference is significant. **Ans.**

Example 75. *One type of aircraft is found to develop engine trouble in 5 flights out of a total of 100 and another type in 7 flights out of a total of 200 flights. Is there a significant difference in the two types of aircrafts so far as engine defects are concerned.*

Solution. $n_1 = 100$ flights, Number of troubled flights = 5, $p_1 = \frac{5}{100} = \frac{1}{20}$

$n_2 = 200$ flights, Number of troubled flights = 7, $p_2 = \frac{7}{200}$

$$e^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = \frac{0.05 \times 0.95}{100} + \frac{0.035 \times 0.965}{200}$$

$$= 0.000475 + 0.0001689 = 0.0006439$$

$$e = 0.0254$$

$$z = \frac{0.05 - 0.035}{0.0254} = 0.59$$

$z < 1$, Difference is not significant. **Ans.**

11.49 THE t-DISTRIBUTION (For small sample)

The students distribution is used to test the significance of

(*i*) The mean of a small sample.

(*ii*) The difference between the means of two small samples or to compare two small samples.

(*iii*) The correlation coefficient.

Let $x_1, x_2, x_3, \ldots, x_n$ be the members of random sample drawn from a normal population with mean μ. If $\bar{x}$ be the mean of the sample then

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{where} \quad s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1}$$

Example 76. *A machine which produces mica insulating washers for use in electric device to turn out washers having a thickness of 10 mm. A sample of 10 washers has an average thickness 9.52 mm with a standard deviation of 0.6 mm. Find out t.*

Solution. $\bar{x} = 9.52$, $M = 10$, $S' = 0.6$, $n = 10$

$$t = \frac{\bar{x} - M}{\frac{s}{\sqrt{n}}} = \frac{9.52 - 10}{\frac{0.6}{\sqrt{10}}} = \frac{0.48\sqrt{10}}{0.6} = -\frac{4}{5}\sqrt{10}$$

$$= -0.8 \times 3.16 = -2.528$$ **Ans.**

Example 77. *Compute the students t for the following values in a sample of eight: – 4, –2, –2, 0, 2, 2, 3, 3 taking the mean of universe to be zero.* (*A.M.I.E., Summer 1995*)

Solution. $\mu = 0$

S.No.	x	$x - \bar{x} = \left(x - \frac{1}{4}\right)$	$(x - \bar{x})^2 = \left(x - \frac{1}{4}\right)^2$
1	–4	$-\frac{17}{4}$	$\frac{289}{16}$
2	–2	$-\frac{9}{4}$	$\frac{81}{16}$
3	–2	$-\frac{9}{4}$	$\frac{81}{16}$
4	0	$-\frac{1}{4}$	$\frac{1}{16}$
5	2	$\frac{7}{4}$	$\frac{49}{16}$
6	2	$\frac{7}{4}$	$\frac{49}{16}$
7	3	$\frac{11}{4}$	$\frac{121}{16}$
8	3	$\frac{11}{4}$	$\frac{121}{16}$
$n = 8$	$\Sigma x = 2$		$\Sigma (x - \bar{x})^2 = \frac{792}{16}$

$$\bar{x} = \frac{2}{8} = \frac{1}{4}$$

$$\text{S.D.} = s = \sqrt{\frac{(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{792}{16 \times 7}} = \sqrt{7.07} = 2.66$$

$$t = \frac{\bar{x} - \mu}{\frac{S'}{\sqrt{n}}} = \frac{\frac{1}{4} - 0}{\frac{2.66}{\sqrt{8}}} = \frac{\sqrt{8}}{4\,(2.66)} = \frac{2.83}{10.64} = 0.266$$ **Ans.**

11.50 WORKING RULE

To calculate significance of sample mean at 5% level.

Calculate $t = \frac{\bar{x} - \mu}{s}\sqrt{n}$ and compare it to the value of t with $(n - 1)$ degrees of freedom at 5% level, obtained from the table. Let this tabulated value of t be t_1.

If $t < t_1$, then we accept the hypothesis *i.e.*, we say that the sample is drawn from the population.

If $t > t_1$, we compare it with the tabulated value of t at 1% level of significance for $(n-1)$ degrees of freedom. Denote it by t_2. If $t_1 < t < t_2$ then we say that the value of t is significant.

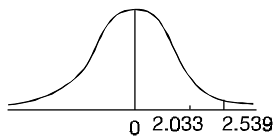
If $t > t_1$, we reject the hypothesis and the sample is not drawn from the population.

Example 78. *A manufacturer intends that his electric bulbs have a life of 1000 hours. He tests a sample of 20 bulbs, drawn at random from a batch and discovers that the mean life of the sample bulbs is 990 hours with a s.d. of 22 hours. Does this signify that the batch is not up to the standard?*

[**Given:** *The table value of t at 1% level is significance with 19 degrees of freedom is 2.539*]

Solution. $\overline{x} = 990,\ \sigma = 22,\ x = 1000$

$$t = \frac{\overline{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{1000-990}{\frac{22}{\sqrt{20}}}$$



$$= \frac{10\sqrt{20}}{22} = \frac{22.36}{11} = 2.033$$

Since the calculated value of t (2.032) is less than the value of t (2.539) from the table. Hence, it is not correct to say that this batch is not upto this standard. **Ans.**

Example 79. *Ten individuals are chosen at random from a population and their heights are found to be in inches 63, 63, 64, 65, 66, 69, 70, 70, 71. Discuss the suggestion that the Mean height of universe is 65.*

For 9 degree of freedom t at 5% level of significance = 2.262.

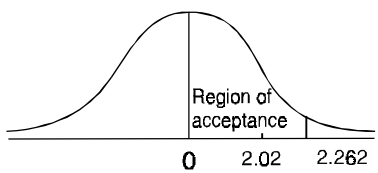
Solution.

x	$x - 67$	$(x-67)^2$
63	-4	16
63	-4	16
64	-3	9
65	-2	4
66	-1	1
69	$+2$	4
69	$+2$	4
70	$+3$	9
70	$+3$	9
71	$+4$	16
$\Sigma x = 670$		$\Sigma(x-\overline{x})^2 = 88$

$$\overline{x} = \frac{\Sigma x}{n} = \frac{670}{10} = 67,\quad s = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n-1}} = \sqrt{\frac{88}{9}} = 3.13$$

$$t = \frac{\overline{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{67-65}{\frac{3.13}{\sqrt{10}}}$$

$$= \frac{2\sqrt{10}}{3.13} = 2.02$$



$$2.02 < 2.262$$

Calculated value of t (2.02) is less than the table value of t (2.262). The hypothesis is accepted the mean height of universe is 65 inches. **Ans.**

Example 80. *The mean life time of sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. The company claims that the average life of the bulbs produced by it is 1600 hours. Using the level of significance of 0.05, is the claim acceptable?*

Solution. $\bar{x} = 1570,\ S = 120,\ n = 100,\ \mu = 1600$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{1570-1600}{\frac{120}{\sqrt{100}}} = \frac{1570-1600}{12} = 2.5$$

At 0.05 the level of significance, $t = 1.96$.

Calculated value of t > Table value of t.

$$2.5 > 1.96$$

Hence the claim is to be rejected. **Ans.**

Example 81. *A sample of 6 persons in an office revealed an average daily smoking of 10, 12, 8, 9, 16, 5 cigarettes. The average level of smoking in the whole office has to be estimated at 90% level of confidence.*

t = 2.015 for 5 degree of freedom.

Solution.

x	$x-10$	$(x-10)^2$
10	0	0
12	2	4
8	-2	4
9	-1	1
16	$+6$	36
5	-5	25
Total	0	$\Sigma(x-10)^2 = 70$

$$\text{Mean} = a + \frac{\Sigma fd}{\Sigma f} = 10 + \frac{0}{6} = 10$$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{70}{5}} = 3.74$$

At 90% level of confidence, $t = \pm 2.015$.

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \quad \text{or} \quad \pm 2.015 = \frac{10-\mu}{\frac{3.74}{\sqrt{6}}}$$

or $$\mu = 2.015 \times \frac{3.74}{\sqrt{6}} + 10 = 6.92,\ 13.08$$ **Ans.**

Example 82. *A certain stimulus administered to each of 12 patients resulted in the following increase in the blood pressures 5, 2, 8, –1, 3, 0, 6, –2, 1, 5, 0, 4. Can it be calculated that stimulus is accompanied by an increase in blood pressure given that for 11 degrees of freedom the value of $t_{0.5}$ is 2.201?*

Solution.

$$\bar{x} = \frac{5+2+8-1+3+0+6-2+1+5+0+4}{12}$$

$$= \frac{31}{12} = 2.583 = 2.6 \text{ approx.}$$

x	$x - 2.6$	$(x - 2.6)^2$
5	2.4	5.76
2	–0.6	0.36
8	5.4	29.16
–1	–3.6	12.96
3	0.4	0.16
0	–2.6	6.76
6	3.4	11.56
–2	–4.6	21.16
1	–1.6	2.56
5	2.4	5.76
0	–2.6	6.76
4	1.4	1.96
$\Sigma x = 12$	$\Sigma (x - 2.6)$	$\Sigma (x - 2.6)^2 = 104.92$

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n-1} = \frac{104.92}{12-1} = 9.54$$

$$s = 3.08$$

Assuming that the stimulus will not be accompanied by increase in blood pressure, *i.e.*, the mean of increase in blood pressure for the population is zero, we have

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{2.6 - 0}{3.08}\sqrt{12} = \frac{2.6}{3.08} \times 3.464 = 2.924$$

As the computed value of t, *i.e.*, 2.924 is greater than $t_{0.05}$, *i.e.*, 2.201 we find that our assumption is wrong and we conclude that as a result of the stimulus blood pressure will increase. **Ans.**

Example 83. *A fertiliser mixing machine is set to give 12 kg of nitrate for quintal bag of fertiliser. Ten 100 kg bags are examined. The percentages of nitrate per bag are as follows:*

11, 14, 13, 12, 13, 12, 13, 14, 11, 12

Is there any reason to believe that the machine is defective? Value of t for 9 degrees of freedom is 2.262. (*A.M.I.E., Winter 1997*)

Solution.

The calculation of $\bar{x}$ and s is given in the following table:

x	$d = x - 12$	d^2
11	–1	1
14	2	4
13	1	1
12	0	0
13	1	1

12	0	0
13	1	1
14	2	4
11	–1	1
12	0	0
$\Sigma x = 125$	$\Sigma d = 5$	$\Sigma d^2 = 13$

$$\mu = 12 \quad \text{kg}, \quad n = 10, \bar{x} = \frac{\Sigma x}{n} = \frac{125}{10} = 12.5$$

$$s^2 = \frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2 = \frac{13}{10} - \left(\frac{5}{10}\right)^2 = \frac{13}{10} - \frac{1}{4} = \frac{21}{20} = \frac{105}{100}$$

$$s = 1.024$$

Value of t for 9 degrees of freedom = 2.262

Also $$t = \frac{\bar{x} - \mu}{s}\sqrt{n}$$

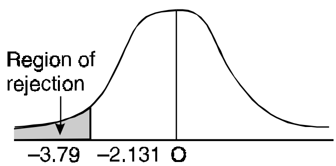
$$= \frac{12.5 - 12}{1.024}\sqrt{10} = 1.54$$

Since the value of t is less than 2.262, there in no reason to believe that machine is defective. **Ans.**

Example 84. *A random sample of size 16 values from a normal population showed a mean of 53 and a sum of squares of deviation from the mean equals to 150. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% and 99% confidence limits of the mean of the population.*

$$\gamma = 15, \quad \alpha = 0.05, \; t = 2.131$$

$$\alpha = 0.01, \; t = 2.947$$



Solution. $\mu = 56, \; n = 16, \; \bar{x} = 53, \; \Sigma(x - \bar{x})^2 = 150$

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{150}{15} = 10$$

$$s = \sqrt{10}$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{53 - 56}{\frac{\sqrt{10}}{\sqrt{16}}} = \frac{-3 \times 4}{\sqrt{10}} = -3.79$$

$$t = 3.79$$

3.79 > 2.131 and also 3.79 > 2.947.

Thus, the sample cannot be regarded as taken from the population. **Ans.**

11.51 TESTING FOR DIFFERENCE BETWEEN MEANS OF TWO SMALL SAMPLES

Let the mean and variance of the first population be μ_1 and σ_1^2 and $\mu_2 \cdot \sigma_2^2$ be the mean and variance of the second population.

Let $\bar{x}_1$ be the mean of small sample of size n_1 from first population and $\bar{x}_2$ the mean of a sample of size n_2 from second population.

We know that

$$E(\bar{x}_1) = \mu_1 \text{ and } Var(\bar{x}_1) = \frac{\sigma_1^2}{n_1}$$

$$E(\bar{x}_2) = \mu_2 \text{ and } Var(\bar{x}_2) = \frac{\sigma_2^2}{n_2}$$

If the samples are independent, then $\bar{x}_1$ and $\bar{x}_2$ are also independent.

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2 \text{ and } Var(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\bar{x}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ and } \bar{x}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

then
$$(\bar{x}_1 - \bar{x}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If the population is the same then

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (\mu_1 - \mu_2 = \mu_1 - \mu_1 = 0)$$

Example 85. *Two independent samples of 8 and 7 items respectively had the following values of the variable (weight in ounces):*

Sample 1: 9 11 13 11 15 9 12 14

Sample 2: 10 12 10 14 9 8 10

Is the difference between the means of the sample significant?

[Given for V = 13, $t_{0.05}$ = 2.16] (*A.M.I.E., Winter 1996*)

Solution.

Assumed mean of $x = 12$, Assumed mean of $y = 10$

x	$(x-12)$	$(x-12)^2$	y	$(y-10)$	$(y-10)^2$
9	–3	9	10	0	0
11	–1	1	12	2	4
13	1	1	10	0	0
11	–1	1	14	4	16
15	3	9	9	–1	1
9	–3	9	8	–2	4
12	0	0	10	0	0
14	2	4	–	–	–
94	–2	34	73	3	25

$$\bar{x} = \frac{\Sigma x}{n} = \frac{94}{8} = 11.75$$

$$\sigma_x^2 = \frac{\Sigma (x-12)^2}{n} - \left(\Sigma \frac{(x-12)}{n}\right)^2 = \frac{34}{8} - \left(\frac{-2}{8}\right)^2 = 4.1875$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{73}{7} = 10.43$$

$$\sigma_y^2 = \frac{\Sigma (y-10)^2}{n} - \left[\frac{\Sigma (y-10)}{n}\right]^2 = \frac{25}{7} - \left(\frac{3}{7}\right)^2 = 3.438$$

$$s = \sqrt{\frac{(x-\bar{x})^2 + \Sigma (y-\bar{y})^2}{n_1+n_2-2}} = \sqrt{\frac{34+25}{8+7-2}} = \sqrt{\frac{59}{13}} = \sqrt{4.54} = 2.13$$

$$t = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{11.75-10.43}{2.13\sqrt{\frac{1}{8}+\frac{1}{7}}} = \frac{1.32}{2.13\sqrt{0.268}} = \frac{1.32}{2.13\times 0.518}$$

$$= \frac{1.32}{1.103} = 1.12$$

The 5% value of t for 13 degree of freedom is given to be 2.16. Since calculated value of t is 1.12 is less than 2.16, the difference between the means of samples is not significant. **Ans.**

Exercise 11.9

1. A random sample of six steel beams has mean compressive strength of 58.392 psi (pounds per square inch) with a standard deviation of $s = 648$ psi. Test the null hypothesis $H_0 = \mu = 58{,}000$ psi against the alternative hypothesis $H_1 : \mu > 58{,}000$ psi at 5% level of significance (value for t at 5 degree of freedom and 5% significance level is 2.0157). Here μ denotes the population mean.
(*A.M.I.E., Summer 2000*)

2. A certain cubical die was thrown 96 times and shows 2 upwards 184 times. Is the die biased?
Ans. die is biased.

3. In a sample of 100 residents of a colony 60 are found to be wheat eaters and 40 rice eaters. Can we assume that both food articles are equally popular?

4. Out of 400 children, 150 are found to be under weight. Assuming the conditions of simple sampling, estimate the percentage of children who are underweight in, and assign limits within which the percentage probably lies. **Ans.** 37.5% approx. Limits = 37.5 ± 3 (2.4)

5. 500 eggs are taken at random from a large consignment, and 50 are found to be bad. Estimate the percentage of bad eggs in the consignment and assign limits within which the percentage probably lies. **Ans.** 10%, 10 ± 3.9

6. A machine puts out 16 imprefect articles in a sample of 500. After the machine is repaired, puts out 3 imprefect articles in a batch of 100. Has the machine been improved?
Ans. The machine has not been improved.

7. In a city A, 20% of a random sample of 900 school boys had a certain slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference between the proportions significant? **Ans.** $z = 0.37$, Difference between proportions is significant.

8. In two large populations there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations?
Ans. $z = 2.5$, not hidden at 5% level of significance.

9. One thousand articles from a factory are examined and found to be three percent defective. Fifteen hundred similar articles from a second factory are found to be only 2 percent defective. Can it reasonably be concluded that the product of the first factory is inferior to the second?
Ans. It cannot be reasonable concluded that the product of the first factory is inferior to that of the second.

10. A manufacturing company claims 90% assurance that the capacitors manufactured by them will show a tolerance of better than 5%. The capacitors are packaged and sold in lots of 10. Show that about 26% of his customers ought to complain that capacitors do not reach the specified standard.

11. An experiment was conducted on nine individuals. The experiment showed that due to smoking, the pulse rate increased in the following order:

$$5,\ 3,\ 4,\ -1,\ 2,\ -3,\ 4,\ 3,\ 1.$$

Can you maintain that smoking leads to an increase in the pulse rate?
(t for 8 d.f. at 5% level of significance = 2.31). **Ans.** Yes.

12. Nine patients to whom a certain drink was administered registered the following in blood pressure:

$$7, 3, -1, 4, -3, 5, 6, -4, 1.$$

Show that the data do not indicate that the drink was responsible for these increments.

13. A machine has produced washers having a thickness of 0.50 mm. To determine whether the machine is in proper working order, a sample of 10 washers is chosen for which the mean thickness is 0.53 mm. and the standard deviation is 0.03 mm. Test the hypothesis that the machine is in proper working order using a level of significance (a) 0.05 (b) 0.01.

Ans. (a) The machine is not in proper working order at 0.05 level of significance.
(b) The machine is in proper working order at 0.01 level of significance.

14. Eleven school boys were given a test in drawing. They were given a months further tuition and a second test of equal difficulty was held at the end of it. Do the marks give evidence that the students have benefitted by extra coaching.

Boys	1	2	3	4	5	6	7	8	9	10	11
Marks I Test	23	20	19	21	18	20	18	17	23	16	19
Marks II Test	24	19	22	18	20	22	20	20	23	20	17

Ans. t = 1.48, The value of t is not significant at 5% level of significance. (*i.e.*, the test, i.e., the students) no evidence that the students have benefitted by extra coaching.

15. Two horses A and B were tested according to the time (in seconds) to run a particular race with the following results:

Horse A	28	30	32	33	33	29	39
Horse B	29	30	30	24	27	29	

Test whether you can discriminate between two horses? **Ans.** Yes with 75% confidence.

11.52 THE CHI-SQUARE DISTRIBUTION

Chi-square is a measure of actual divergence of the observed and expected frequencies. If f_0 is the observed frequency and f_e the expected frequency of a class interval, then x^2 is defined as

$$x^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

11.53 DEGREE OF FREEDOM (df)

The degree of freedom refers to the number of "independent constrants" in a set of data. We shall illustrate this concept with example. There is a 2×2 association table and the actual frequencies as under:

Let the two attributes A and B be independent.

Expected frequency of $(AB) = \frac{30 \times 60}{100} = 18$

After finding the frequency of (AB), the expected frequencies of the remaining three classes are automatically fixed.

Expected frequency of $(\alpha B) = 60 - 18 = 42$

Expected frequency of $(A \beta) = 30 - 18 = 12$

Expected frequency of $(\alpha \beta) = 70 - 42 = 28$

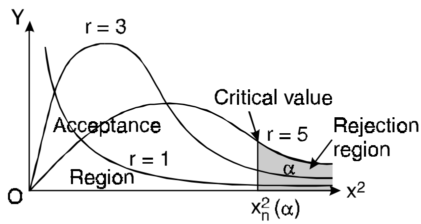
It means that only one choice is fixing of frequency of AB is independent choice. Frequencies of the remaining three classes depend on the frequency of (AB). It means, we have only one degree of freedom.

Degree of freedom $= (r - 1)(c - 1)$

where r is the number of rows and c is the number of columns.

11.54 x^2 – CURVE

Let $x_1, x_2, ..., x_n$ be n standard variates with mean zero and S.D. unity. Then x^2-distribution has the $x_1^2, x_2^2, x_3^2, ..., x_n^2$ random variates.



Equation of the x^2-curve is

$$y = y_0 \, e^{-\frac{x^2}{2}} \, (x^2)^{\frac{r-1}{2}}, \quad r = n - 1$$

where r is the degree of freedom.

Since this equation does not have any parameter. So it can be used for every problem of chi-square. $x_n^2(\alpha)$ denote the value of chi-square for n degree of freedom such that the area to the right of this point is α.

11.55 GOODNESS OF FIT

The value of x^2 is used to find the divergence of the observed frequency from the expected frequency.

If the value of P is high the fit is said to be good. It means that there no significant divergence between observed and expected data.

If the curve of the expected frequency is super imposed on the curve of observed frequencies there would not be much divergence between the two. The fit would be good. If the value of P is small, the fit is said to be poor.

11.56 STEPS FOR TESTING

(*i*) First calculate the value of x^2.

(*ii*) From the table read the value of x^2 for a given degree of freedom.

(*iii*) Find out the probability P corresponding to the calculated values of x^2.

(*iv*) If $P > 0.05$, the value is not significant and it is a good fit.

(*v*) If $P < 0.05$, the deviations are significant.

Example 86. *The following table is given*

		Eye colour in sons		
		not light	*light*	
Eye colour in fathers	*Not light*	*230*	*148*	*378*
	light	*251*	*471*	*622*
		381	*619*	*1000*

Test whether the colour of the son's eyes is associated with that of the fathers.

Given: *value of x^2 is 3.84 for 1 degree of freedom.*

Solution.

Hypothesis: Let the eye colour of sons and the eye colour of fathers independent.

		Eye colour in sons	
		not light	*light*
Eye colour in fathers	Not light	$\frac{378 \times 381}{1000} = 144$	$\frac{378 \times 619}{1000} = 234$
	light	$\frac{622 \times 381}{1000} = 237$	$\frac{622 \times 619}{1000} = 385$

$$X^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

$$X^2 = \frac{(230-144)^2}{144} + \frac{(148-234)^2}{234} + \frac{(151-237)^2}{237} + \frac{(471-385)^2}{385}$$

$$= (86)^2 \left[\frac{1}{144} + \frac{1}{234} + \frac{1}{237} + \frac{1}{385}\right] = 133.37$$

The degree of freedom $= (c-1)(r-1) = (2-1)(2-1) = 1$

The value of x^2 at 5% level of significance for 1 degree of freedom is 3.841 and the calculated value is 133.37

$$133.37 > 3.841$$

This leads to the conclusion that the hypothesis is wrong and the colour of son's eyes is associated with that of the fathers to a great extent. **Ans.**

Example 87. *From the following table, showing the number of plants having certain characters, test the hypothesis that the flower colour is independent of flatness of leaf.*

	Flat leaves	*Curled leaves*	*Total*
White Flowers	99	36	135
Red Flowers	20	5	25
Total	119	41	160

Solution. Null Hypothesis: The flower colour is dependent of flatness of leaf.

The following table shows the theoretical frequencies.

	Flat leaves	*Curled leaves*	*Total*
White flowers	$\frac{135 \times 119}{160} = 100$	$\frac{135 \times 41}{160} = 35$	135
Red flowers	$\frac{25 \times 119}{160} = 19$	$\frac{25 \times 41}{160} = 6$	25
Total	119	41	160

$$x^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

$$x^2 = \frac{(99-100)^2}{100} + \frac{(36-35)^2}{35} + \frac{(20-19)^2}{19} + \frac{(5-6)^2}{6}$$

$$x^2 = \frac{1}{100} + \frac{1}{35} + \frac{1}{19} + \frac{1}{6} = 0.2579$$

Degree of freedom $= (r-1)(c-1) = (2-1)(2-1) = 1$

We have $x^2 = 0.0158$ at 0.1 level of significance.

$$0.2579 > 0.0158$$

This leads to the conclusion that the hypothesis is wrong and the flower colour is independent of flatness of leaf at the 0.1 level of significance. **Ans.**

Example 88. *The following table gives the number of aircraft accidents that occurs during various days of the week. Find whether the accidents are uniformly distributed over the week.*

Days	*Sun.*	*Mon.*	*Tues.*	*Wed.*	*Thurs.*	*Fri.*	*Sat.*
No. of accidents	*14*	*16*	*8*	*12*	*11*	*9*	*14*

Given: *The values of chi-square significant at 5, 6, 7, d.f. are respectively 11.07, 12.59, 14.07 at the 5% level of significance.*

Solution. Null Hypothesis: The accidents are uniformly distributed over the week.

Expected frequencies of the accidents are given below:

Days	Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Total
No. of accidents	12	12	12	12	12	12	12	84

$$x^2 = \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(11-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-12)^2}{12}$$

$$= \frac{1}{12}[4+16+16+0+1+9+4] = \frac{50}{12} = 4.17$$

The number of degrees of freedom = Number of observations – Number of independent constants = 7 – 1 = 6.

The tabulated $x^2_{0.05}$ for 6 d.f. = 12.59

Since the calculated x^2 is much less than the tabulated value, we accept the null hypothesis. Hence, the accidents are uniformly distributed over the week. **Ans.**

Example 89. *A set of five similar coins is tossed 320 times and the result is*

No. of heads	*0*	*1*	*2*	*3*	*4*	*5*
Frequency	*6*	*27*	*72*	*112*	*71*	*32*

Test the hypothesis that the data follow a binomial distribution. (*Warangal 1995*)

Solution. P (Head) $= \frac{1}{2},\ q = 1 - \frac{1}{2} = \frac{1}{2}$

Theoretical frequencies are

$$P(0H) = q^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}, \text{ Frequency of 0 head} = \frac{320}{32} = 10$$

$$P(1H) = {}^5C_1\, pq^4 = {}^5C_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{5}{32}, \text{ Frequency of 1 head} = \frac{5}{32}\times 320 = 50$$

$$P(2H) = {}^5C_2\, p^2 q^3 = 10\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32}, \text{ Frequency of 2 heads} = \frac{10}{32}\times 320 = 100$$

$$P(3H) = {}^5C_3\, p^3 q^2 = 10\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}, \text{ Frequency of 3 heads} = \frac{10}{32}\times 320 = 100$$

$$P(4H) = {}^5C_4\ p^4\ q = 5\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right) = \frac{5}{32},\ \text{Frequency of 4 heads} = \frac{5}{32}\times 320 = 50$$

$$P(5H) = {}^5C_5\ p^5\ q^0 = \left(\frac{1}{2}\right)^5 = \frac{1}{32},\ \text{Frequency of 5 heads} = \frac{1}{32}\times 320 = 10$$

$$x^2 = \frac{(6-10)^2}{10}+\frac{(27-50)^2}{50}+\frac{(72-100)^2}{100}+\frac{(112-100)^2}{100}+\frac{(71-50)^2}{50}+\frac{(32-10)^2}{10}$$

$$= \frac{1}{100}[160+1058+784+144+882+4840] = \frac{7868}{100} = 78.68$$

Degree of freedom = 6 – 1 = 5

For 5 degree of freedom, x^2 = 11.07

Since the calculated value of x^2 is much greater than that of x^2 at 5% level of significance, the hypothesis that the data follow the binomial law is rejected. **Ans.**

Example 90. *Fit a Poisson distribution to the following data and test the goodness of fit.*

x	0	1	2	3	4	5	6
f	275	72	30	7	5	2	1

Solution.

Mean, $$m = \frac{\Sigma fx}{\Sigma f} = \frac{0+72+60+21+20+10+6}{392} = \frac{189}{392} = 0.482$$

Poissondistribution

$$P(r) = \frac{e^{-m}\, m^r}{r\,!} \quad \text{or} \quad P(r) = \frac{e^{-0.482}\ (0.482)^r}{r\,!}$$

$$P(0) = e^{-0.482} = 0.6175, \quad f(0) = 392\times 0.6175 = 242.1$$

$$P(1) = \frac{e^{-0.482}(0.482)^1}{1\,!} = 0.2976, \quad f(1) = 392\times 0.2976 = 116.7$$

$$P(2) = \frac{e^{-0.482}(0.482)^2}{2\,!} = 0.0717, \quad f(2) = 392\times 0.0717 = 28.1$$

$$\left.\begin{aligned} P(3) &= \frac{e^{-0.482}(0.482)^3}{3\,!} = 0.0115, \quad f(3) = 392\times 0.0115 = 4.5\\ P(4) &= \frac{e^{-0.482}(0.482)^4}{4\,!} = 0.00139, \quad f(4) = 392\times 0.00139 = 0.5\\ P(5) &= \frac{e^{-0.482}(0.482)^5}{5\,!} = 0.0001, \quad f(5) = 392\times 0.0001 = 0.1\\ P(6) &= \frac{e^{-0.482}(0.482)^6}{6\,!} = 0.00001, \quad f(6) = 392\times 0.00001 = 0 \end{aligned}\right\} \quad (5.1)$$

$$x^2 = \frac{(275-242.1)^2}{242.1}+\frac{(72-116.7)^2}{116.7}+\frac{(30-28.1)^2}{28.1}+\frac{[(7+5+2+1)-(4.5+0.5+0.1)]^2}{4.5+0.5+0.1}$$

or $$x^2 = 4.471+17.122+0.128+19.217 = 40.938$$

Degree of freedom = 7 – 1 – 1 – 3 = 2

[One d.f. being lost because $\Sigma\, 0 = \Sigma\, E$; 1 d.f. is lost because the parameter m has been estimated; 3 d.f. are lost because of pooling the last four expected cell frequencies which are less than 5]

Tabulated value of x^2 for 2 d.f. at 5% level of significance = 5.99.

Since, the calculated value of x^2 (40.938) is much greater than 5.99, it is highly significant. Hence, Poisson distribution is not good fit. **Ans.**

Exercise 11.10

1. The following information is obtained concerning an investigation of 50 ordinary shops of small size.

	Shops		Total
	In Town	In Villages,	
Run by men	17	18	35
Run by women	3	12	15
Total	20	30	50

Can it be inferred that shops run by women are relatively more in villages than in towns? Use x^2 test. **Ans.** $x^2 = 3.57$, Hypothesis is wrong.

2. Of a group of patients who complained they did not sleep well, some were given sleeping pills while others were given sugar pills (although they all thought they were getting sleeping pills). They were later asked whether the pills helped them or not. The result of their responses are shown in the table given below. Assuming that all patients told the truth, test the hypothesis that there is no difference between sleeping pills and sugar pills at a significance level of 0.05.

	Slept well	*Did not sleep well*
Took sleeping pills	44	10
Took sugar pills	81	35

Ans. The hypothesis cannot be rejected at the 0.05 level.

3. In an experiment on immunization of cattle from tuberculosis the following results were obtained

	Died	*Unaffected*
Inoculated	12	26
Not inoculated	16	6

Examine the effect of vaccine in controlling susceptibility to tuberculosis.

Ans. $x^2 = 9.367$, vaccine is effective.

4. Genetic theory states that children having one parent of blood type M and other blood type N will always be one of three types M, MN, N and that the proportions of these types will on average be 1 : 2 : 1. A report states that out of 300 children having one M parent and one N parent, 30% were found to be of type M, 45% of type MN and remainder of type N. Test the hypothesis by x^2 test.

Ans. Hypothesis is correct.

5. In an experiment on pea-breeding, Mendal obtained the following frequencies of seeds; 315 round and yellow, 101 wrinkled and yellow; 108 round and green, 32 wrinkled and green. Total 556. Theory predicts that the frequencies should be in the proportion 9 : 3 : 3 : 1 respectively. Set up proper hypothesis and test it at 10% level of significance.

Ans. $x^2 = 0.51$. There seems to be good correspondence between theory and experiment.

6. On a particular proposal of national importance, political party A and party B cast votes as given in the table. At a level of significance of (a) 0.01 and (b) 0.05, test the hypothesis that there is no difference between the two parties in so far as this proposal is concerned.

	In favour	*Opposed*	*Undecided*
Party *A*	85	78	37
Party *B*	118	61	25

Ans. The hypothesis can be rejected at both levels.

7. The table shows the relation between the performance in mathematics and electronics, using a (*a*) 0.05 (*b*) 0.01 significance level.

Mathematics	*Electronics*		
	High marks	*Medium marks*	*Low marks*
High marks	56	71	12
Medium marks	47	163	38
Low marks	14	42	85

Ans. The hypothesis can be rejected at both levels.

8. The results of a survey made to determine whether the age of a driver 21 years of age and older has any effect on the number of automobile accidents in which he is involved (including minor accidents) are given in the table below. At a level of significance of (*a*) 0.05 and (*b*) 0.01, test the hypothesis that number of accidents is independent of the age of the driver.

		Age of the driver				
		21–30	*31–40*	*41–50*	*51–60*	*61–70*
Number of accidents	0	748	821	786	720	672
	1	74	60	51	66	50
	2	31	25	22	16	15
	more than 2	9	10	6	5	7

Ans. The hypothesis cannot be rejected at either level.

9. A die is thrown 60 times with the following results.

Face	1	2	3	4	5	6
Frequency	8	7	12	8	14	11

Test at 5% level of significance if the die is honest, assuming that $P(x^2 > 11.1) = 0.05$ with 5 d.f.

10. Fit a Binomial Distribution to the data

x	0	1	2	3	4	5
f	38	144	342	287	164	25

and test for goodness of fit at the level of significance 0.05.

Ans. $x^2 = 7.97$, Binomial distribution gives a good fit at 5% level.

11. Fit a Poisson distribution to the following data and test for its goodness of fit at level of significance 0.05.

x	0	1	2	3	4
f	419	352	154	56	19

Ans. Poisson distribution can be fitted to the data.

12. A bird watching sitting in a park has spotted a number of birds belonging to 6 categories. The exact classification is given below:

Category	1	2	3	4	5	6
Frequency	6	7	13	17	6	5

Test at 5% level of significance whether or not the data is compatible with the assumption that this particular park is visited by birds belonging to these six categories in the proportion

$= 1 : 1 : 2 : 3 : 1 : 1.$

Given $P\ (x^2 = 1.07\ ) = 0.05$ for 5 degree of freedom.

13. Two hundred digits were chosen at random from a set of tables. The frequencies of the digits were as follows:

Digits	0	1	2	3	4	5	6	7	8	9
Frequency	18	19	23	21	16	25	22	20	21	15

Use x^2 test to assess the corrections of hypothesis that the digits were distributed in equal numbers in the table from which they were chosen.

Given that the values of x^2 are respectively 16.9, 18.3., 19.7 for 9, 10 and 11 degrees of freedom at 5% level of significance. **Ans.** $x^2 = 4.3$, the hypothesis seems reasonable correct.

14. A survey of 320 families with 5 children each revealed the following distribution:

No. of boys	5	4	3	2	1	0
No. of girls	0	1	2	3	4	5
No. of families	14	56	110	88	40	12

Is this result consistent with the hypothesis that male and female births are equally probable?

Ans. $x^2 = 7.16$, Equal probability for male and female births may be accepted.

11.57 F-DISTRIBUTION HAS THE FOLLOWING APPLICATION

F-Test for Equality of Population Variances

Suppose we want to test

(*i*) Whether two independent samples $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$ have been drawn from the normal population with the same variance σ^2.

(*ii*) Whether the two independent estimates of the population variance are homogeneous or not.

Under the null hypothesis (H_0)

(*i*) $\sigma_x^2 = \sigma_y^2 = \sigma^2$, population variances are equal.

(*ii*) Two independent estimates of population variance are homogeneous, F is given by

$$F = \frac{S_x^2}{S_y^2}$$

where $S_x^2 = \dfrac{\sum\limits_{n=1}^{n_2} (x - \overline{x})^2}{n_1 - 1}, \quad S_y^2 = \dfrac{\sum\limits_{n=1}^{n_2} (y - \overline{y})^2}{n - 1}$

and $\overline{x}, \overline{y}$ are sample means, S_1^2, S_2^2 are unbiased estimates of two samples from two normal populations with $(\sigma_1 = \sigma_2)$.

The distributions of variance ratio F with r_1 and r_2 is

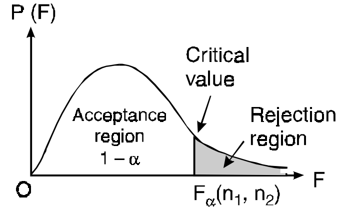
$$y = y_0 \cdot F^{\frac{r_1 - r_2}{2}} \left[1 + \frac{r_1}{r_2} F\right]^{-\frac{(r_1 + r_2)}{2}}$$

where $r_1 = n_1 - 1$, $r_2 = n_2 - 1$, r_1 and r_2 are d.f. of two samples.

Note. (*i*) Greater of the two variances S_x^2, S_y^2 is to be taken in the numerator and n_1 corresponds to the greater variance. Calculated value of F is compared with tabulated value of F at certain level of significance, H_0 is either rejected or accepted.

(*ii*) *Significative Test*

If the calculated value of F is higher than this table value of F for the given degree of freedom at 5% level of significance as such the difference is significant. It means that the variance between the samples is significantly greater than variance within the samples. In other words, the samples are not picked up from the same population or the mean value of various samples are significantly different from each other.



11.58 FISHER'S Z-DISTRIBUTION

On substitution $z = \frac{1}{2}\log_e F$ or $F = e^{2z}$ in the F-distribution, we have the Fisher's z-distribution. It is of the form

$$y = y_0 \, e^{r_1 z} \, (r_1 \, e^{2z} + r_2)$$

The curve is more symmetrical than F-distribution curve.

Example 91. *The I.Q.'s of 25 students from one college showed a variance of 16 and those of an equal number from the other college had a variance of 8. Discuss whether there is any significant difference in variability of intelligence.*

Given: F (5%) = 1.98, F (1%) = 2.62

Solution. $\sigma_1^2 = 16, \ \sigma_2^2 = 8$

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{16}{8} = 2$$

Tabulated value of F at 5% level of significance = 1.98

Calculated value of F (2) Tabulated value of F (1.98)

Hence, variability of intelligence is just significant at 5% level of significant.

Tabulated value of F at 1% level of significance = 2.62

Calculated value F (2) < Tabulated value of F (2.62)

Hence, variability of intelligence is not significant at 1% level of significance.**Ans.**

Example 92. *Two random samples are as are :*

Sample	*Size*	*Sum of squares of deviations from the mean*
1	*10*	*90*
2	*12*	*108*

Test whether the samples come from the same normal population at 5% level of significance.

[**Given:** $F_{0.05}$ *(9, 11) = 2.90,* $F_{0.05}$ *(11, 9) = 3.10*]

Solution. Null Hypothesis: The two samples have been drawn from the same normal population.

$$n_1 = 10, \ \Sigma (x - \bar{x})^2 = 90$$

$$n_2 = 12, \ \Sigma (y - \bar{y})^2 = 108$$

Here, $$S_1^2 = \frac{1}{n_1 - 1} \Sigma (x - \overline{x})^2 = \frac{90}{9} = 10$$

$$S_2^2 = \frac{1}{n_2 - 1} \Sigma (y - \overline{y})^2 = \frac{108}{11} = 9.82$$

$$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.82} = 1.018$$

Tabulated $F_{0.05}$ (9, 11) = 2.90

Since the calculated F is less than tabulated F, it is not significant. Hence, null hypothesis of equality of population variances may be accepted.

Example 93. *Two random samples from two normal populations are given below:*

Sample *I*	16	26	27	23	24	22
Sample *II*	33	42	35	32	28	31

Do the estimates of population variances differ significantly?

Degree of Freedom	*(5, 5)*	*(5, 6)*	*(6, 5)*
5% value of F	*5.05*	*4.39*	*4.95*

Solution.

Sample I x	$x - \overline{x}$ $x - 23$	$(x - \overline{x})^2$	*Sample II*	$y - \overline{y}$ $y - 33.5,$	$(y - \overline{y})^2$
16	–7	49	31	–0.5	0.25
26	3	9	42	8.5	72.25
27	4	16	35	1.5	2.25
23	0	0	32	–1.5	2.25
24	1	1	28	–5.5	30.25
22	–1	1	31	–2.5	6.25
138		76	201	σ_{76}	113.50

$$\overline{x} = \frac{\Sigma x}{n} = \frac{138}{6} = 23, \quad \overline{y} = \frac{\Sigma y}{n} = \frac{201}{6} = 33.5$$

$$S_1^2 = \frac{\Sigma (x - \overline{x})^2}{n - 1} = \frac{76}{5} = 15.2$$

$$S_2^2 = \frac{\Sigma (y - \overline{y})^2}{n - 1} = \frac{113.50}{5} = 22.7$$

$$F = \frac{S_2^2}{S_1^2} = \frac{22.7}{15.2} = 1.4934$$

Tabulated value of $F_{0.05}$ = 5.05

Since the calculated value (1.4934) of F is less than the tabulated value of F (5.05). Hence, the difference is not significant. **Ans.**

EXERCISE 11.11

1. The diameters of two random samples, each of size 10, of bulbs produced by two machines have standard deviations $S_1 = 0.01$ and $S_2 = 0.015$. Assuming that the diameters have independent distributions, test the hypothesis that, the two machines are equally good by testing.

Ans. F = 1.5, yes hypothesis is correct.

2. The mean diameter of rivets produced by two firms A and B are practically the same but their standard deviations are different. For 16 rivets manufactured by firm A, the S.D. is 3.8 mm while for 22 rivets manufactured by firm B is 2.9 mm. Do you think products from firm A are better quality than those of firm B.

Ans. Yes

3. Mango-trees were grown under two experimental conditions. Two random samples of 11 and 9 mango-trees show the samples standard deviations of their weights as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test the hypothesis that the true variances are equal, against the alternative that they are not, at the 10% level.

[Assume that $P(F_{10, 8} > 3.35) = 0.05$ and $P(F_{8, 10} > 3.07) = 0.05$]

Ans. $F = 2.5$, Not significant, hence null hypothesis of equality of population variances may be accepted at level of significance $\alpha = 1.0$.

4. Two random samples drawn from two normal populations are:

Sample *I*	20	16	26	27	23	22	18	29	25	19		
Sample *II*	27	33	42	35	32	34	38	28	41	23	30	37

Obtain the estimates of the variances of the populations and test whether the population have the same variance.

(**Given:** $F_{0.05} = 3.11$ for 11 and 9 d.f.)

Ans. $F = 2.368$, The hypothesis seems to be correct at the 0.05 level of significance.