PROBABILITY DISTRIBUTIONS

There are two types of probability distributions or theoretical distributions.

1. Discrete probability distributions.

(i) Binomial distribution

(ii) Poisson distribution

2. Continuous probability distributions

(i) Normal distribution

(ii) Exponential distribution.

* Bernoulli's Distribution

A random variable 'x' which takes two values '0' & '1' with probability 'q' & 'p' respectively ie, $P(X=0) = q$ & $P(X=1) = p$ where $q = 1-p$ is known as a Bernoulli's discrete random variable & is said to have a Bernoulli's distribution.

The probability function of Bernoulli's distribution can be written as

$$P(x) = p^x q^{1-x} = p^x (1-p)^{1-x} \quad ; \; x = 0, 1$$

Bernoulli's theorem

If the probability of the occurence of an event (success) in a single trial is p, then the probability that it will occur exactly 'r' times out of 'n' independent Bernoullian trials is $^nC_r\, p^r q^{n-r}$ where $p+q=1$, & $r=0,1$

* Binomial Distribution

Definition: A random variable 'x' has a binomial distribution if it assumes only non-negative values and its probability function is given by

$$P(X=r) = P(r) = \begin{cases} ^nC_r\, p^r q^{n-r} & ; r=0,1,2,\ldots n ; q=1-p \\ 0 & ; \text{otherwise.} \end{cases}$$

→ Here 'n' & 'p' are known as parameters of Binomial distribution.

→ 'n' is also known as degree of the distribution

1. Mean of Binomial distribution

$$\text{Mean } (\mu) = E(X) = \sum_{x=0}^{n} x\, p(x)$$

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

2. Variance of Binomial Distribution:-

$$V(x) = E(x^2) - [E(x)]^2$$

$$\Rightarrow V(x) = npq$$

$$S.D\ (\sigma) = \sqrt{\text{Variance}} = \sqrt{npq}$$

Note:-

The conditions for the applicability of B.D as follows:-

1. There are 'n' independent trails
2. Each trial has only 2 possible outcomes
3. The probability of 2 outcomes remains constant.

* Binomial frequency distribution:-

If 'n' independent trials constitute one Experiment & this Experiment is repeated 'N' times then the frequency of 'r' successes is $N \cdot {}^nC_r\, p^r q^{n-r}$ (or) $N \cdot P(r)$.

In 'N' sets of 'n' trials the theoretical frequencies of $0, 1, 2, \ldots, r, \ldots n$ successes will be given by the terms of expansion of $N(q+p)^n$.

The possible no. of successes & their frequencies is known as ~~pr~~ Binomial frequency Distribution.

* Problems :-

1. A fair coin is tossed six times. Find the probability of getting 4 heads.

Sol: Given that :-

A fair coin is tossed six times.

$\therefore$ No. of trials $(n) = 6$

$r = 4$

Let, $p =$ probability of getting head $= 1/2$

$q =$ probability of not getting head $= 1-p = 1/2$

By using Binomial distribution,

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

$$\therefore P(4) = {}^{6}C_{4} \left(\tfrac{1}{2}\right)^{4} \left(\tfrac{1}{2}\right)^{6-4} = {}^{6}C_{4} \left(\tfrac{1}{2}\right)^{4} \left(\tfrac{1}{2}\right)^{2}$$

$$= 0.2344$$

2. The probability that John hits the target is 1/2. If he fires 6 times, find the probability that he hits the target (i) exactly 2 times (ii) more than 4 times (iii) atleast once.

Sol: Here, $n = 6$

$p = 1/2$; $q = 1-p = 1/2$

(i) Exactly 2 times

$$P(r=2) = {}^{6}C_{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{4} = 0.23$$

(ii) More than 4 times.

$$\text{probability} \Rightarrow P(r>4) = P(5)+P(6)$$

$$= {}^{6}C_{5}\left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)^{1} + {}^{6}C_{6}\left(\frac{1}{2}\right)^{6}$$

$$= 0.109$$

(iii) atleast once

$$P(r\geq 1) = P(1)+P(2)+P(3)+P(4)+P(5)+P(6)$$

$$= {}^{6}C_{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{5} + {}^{6}C_{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{4} + {}^{6}C_{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{3}$$

$$+ {}^{6}C_{4}\left(\frac{1}{2}\right)^{4}\left(\frac{1}{2}\right)^{2} + {}^{6}C_{5}\left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)^{1} + {}^{6}C_{6}\left(\frac{1}{2}\right)^{6}$$

$$= 0.98$$

3. 20% of items are produced from a factory are defective. Find the probability that in a sample of 5 choosen at random (i) None is defective (ii) one is defective (iii) $P(1<x<4)$.

Soln Here $n=5$

$P = 20\% = 0.2$; $q = 0.8$

(ii) None is defective

$$P(r=0) = {}^5C_0 (0.2)^0 (0.8)^5 = 0.327$$

(ii) one is defective

$$P(r=1) = {}^5C_1 (0.2)^1 (0.8)^4 = 0.4096$$

(iii) $P(1 < X < 4) = P(2) + P(3)$

$$= {}^5C_2 (0.2)^2 (0.8)^3 + {}^5C_3 (0.2)^3 (0.8)^2$$

$$= 0.25.$$

4. The probability that life of a bulb in 100 days is 0.05. Find the probability that out of 6 bulbs (i) atleast one (ii) greater than 4 (iii) None, will be having a life of 100 days.

Soln Here, $n = 6$

$p = 0.05$; $q = 0.95$

(i) Atleast once

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

$$= 1 - {}^6C_0 (0.05)^0 (0.95)^6$$

$$= 0.244$$

(ii) greater than 4

$$P(r > 4) = P(5) + P(6)$$

$= {}^6C_5 (0.05)^5 (0.95)^1 + {}^6C_6 (0.05)^6$

$= 0.000002$

(iii) None

$P(r=0) = {}^6C_0 (0.05)^0 (0.95)^6$

$= 0.735$

5. Out of 800 families with 5 children each, how many would you expect to have (i) 3 boys (ii) 5 girls (iii) either 2 or 3 boys (iv) atleast one boy? Assume equal probabilities for boys and girls.

Sol Given that

$n = 5$

There are 800 families with 5 children each.

Let 'r' be the no. of boys.

Here, $p = 1/2$; $q = 1/2$

(i) 3 boys

$P(r=3) = {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = 0.3$

(ii) 5 girls

$P(r=0) = {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = 0.036$

(iii) Either 2 or 3 boys

$P(r=2) + P(r=3) = {}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 + {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2$

$= 0.62$

(iv) Atleast one boy

$P(r \geq 1) = P(1) + P(2) + P(3) + P(4) + P(5)$

(or)

$= 1 - P(r=0) = 1 - 0.031 = 0.969$

$\therefore$ No. of families having 3 boys $= 0.31 \times 800 = 248$

No. of families having 5 girls $= 0.036 \times 800 = 25$

No. of families having either 2 or 3 boys

$= 0.62 \times 800 = 496$

No. of families having atleast one boy

$= 0.969 \times 800 = 768$

6. Determine the Binomial distribution for which mean is 4 & variance is 3.

Sol: Given that

mean $= 4$

$\Rightarrow np = 4 \rightarrow (1)$

variance $= 3$

$npq = 3 \rightarrow (2)$

Sub. (1) in (2),

$4q = 3 \Rightarrow q = 3/4 = 0.75$

$\therefore p = 1 - q = 1 - 0.75 = 0.25$

From (1),

$n(0.25) = 4 \Rightarrow n = \frac{4}{0.25} = 16.$

$\therefore$ The parameters of given binomial distribution are $n = 16$ & $p = 0.25 = 1/4$

7. Seven coins are tossed and the no. of heads are noted. The experiment is repeated 128 times and the following distribution is obtained.

No. of heads	0	1	2	3	4	5	6	7
frequency	7	6	19	35	30	23	7	1

Fit a binomial distribution

Soln Given that

$n = 7$; $N = \Sigma f = 7+6+19+35+30+23+7+1$
$= 128$

Here, $p = 1/2$; $q = 1/2$

By binominal distribution

$P(x) = {}^{n}C_{x}\, p^{x} q^{n-x}$ where $x = 0,1,2,3,4,5,6,7$

No. of heads	observed freq	probability $P(x)$	Expected or theoretical freq $f(x) = N \cdot P(x)$
0	7	$P(0) = {}^{7}C_{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{7} = \frac{1}{2^7}$	$f(0) = 128\left(\frac{1}{2^7}\right) = 1$
1	6	$P(1) = 7/2^7$	$f(1) = 128\left[\frac{7}{2^7}\right] = 7$
2	19	$P(2) = 21/2^7$	$f(2) = 128\left(\frac{21}{2^7}\right) = 21$
3	35	$P(3) = 35/2^7$	$f(3) = 128\left(\frac{35}{128}\right) = 35$
4	30	$P(4) = 35/128$	$f(4) = 128\left(\frac{35}{128}\right) = 35$
5	23	$P(5) = 21/128$	$f(5) = 128\left(\frac{21}{128}\right) = 21$
6	7	$P(6) = 7/128$	$f(6) = 128\left(\frac{7}{128}\right) = 7$
7	1	$P(7) = 1/128$	$f(7) = 128\left(\frac{1}{128}\right) = 1$
	128		128

$\therefore$ The required Binomial distribution is

x	0	1	2	3	4	5	6	7
freq	7	6	10	35	30	23	7	1
Expected freq	1	7	21	35	35	21	7	1

8. Fit a Binomial distribution for the following data:

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

Soln Given that

$n = 5$; $N = \Sigma f = 2+14+20+34+22+8 = 100$

$\therefore$ Mean $= \dfrac{\Sigma fx}{\Sigma f} = \dfrac{0+14+40+102+88+40}{100} = \dfrac{284}{100}$

$= 2.84$

ie, $np = 2.84$

$5p = 2.84 \Rightarrow p = \dfrac{2.84}{5} = 0.568$

$q = 1-p = 0.432$

By Binomial distribution,

$P(x) = {}^nC_x \, p^x \, q^{n-x}$; $x = 0, 1, 2, 3, 4, 5$

9. Show that mean of Binomial distribution is np.

Sol By using B.D,

$$P(r) = {}^nC_r\, p^r q^{n-r} \quad ; \quad r = 0, 1, 2, \ldots, n$$

$$\text{Mean } (\mu) = \sum_{r=0}^{n} r\, P(r)$$

$$= \sum_{r=0}^{n} r\left[{}^nC_r\, p^r q^{n-r}\right]$$

$$= {}^nC_1\, p q^{n-1} + 2\, {}^nC_2\, p^2 q^{n-2} + 3\, {}^nC_3\, p^3 q^{n-3} + \ldots + n\, {}^nC_n\, p^n$$

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

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

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

$$= np\,[1]^{n-1} = np(1) = np$$

10. Show that variance of Binomial distribution is npq.

Sol we know that:-

$$P(r) = {}^nC_r\, p^r q^{n-r} \quad ; \quad r = 0, 1, 2, \ldots, n.$$

$$\text{Variance, } V(x) = E(x^2) - [E(x)]^2$$

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

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

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

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

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

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

$$= \left[2 \frac{n(n-1)}{2!} p^2 q^{n-2} + 6 \frac{n(n-1)(n-2)}{3!} p^3 q^{n-3} + \cdots + n(n-1) p^n\right] + \mu - \mu^2$$

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

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

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

$$= \cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2} = np[1-p] = \underline{\underline{npq}}$$

* **Poisson Distribution :-**

A random variable 'x' is said to follow a poisson distribution if it assumes only non-negative values & its probability density function is given by

$$P(x) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^{x}}{x!} & ; x = 0, 1, 2, \ldots \\ 0 & ; \text{otherwise} \end{cases}$$

Here, $\lambda > 0$ is known as parameter of poisson distribution.

Note :- poisson distribution is suitable for rare events for which the probability of occurence 'p' is very small & the no. of trials 'n' is very large.

* → Mean of poisson distribution $(\mu) = \lambda$
  → Variance of poisson distribution $V(x) = \lambda$.

1. Show that mean of P.D is 'λ'.

Sol:- By using P.D,

$$P(r) = \frac{e^{-\lambda} \lambda^{r}}{r!} \quad ; r = 0, 1, 2, \ldots$$

$$\therefore \text{Mean } (\mu) = \sum_{r=0}^{\infty} r\, P(r)$$

$$= \sum_{r=0}^{\infty} r \left[\frac{e^{-\lambda} \lambda^{r}}{r!} \right]$$

$$= e^{-\lambda}\left[\sum_{r=0}^{\infty} \frac{\lambda^r}{(r-1)!}\right]$$

$$= e^{-\lambda}\left[\sum_{r=1}^{\infty} \frac{\lambda^r}{(r-1)!}\right]$$

$$= e^{-\lambda}\left[\sum_{y=0}^{\infty} \frac{\lambda^{y+1}}{y!}\right] \quad | \because \cancel{r+1} = y$$

$$= \lambda e^{-\lambda}\left[\sum_{y=0}^{\infty} \frac{\lambda^y}{y!}\right]$$

$$= \lambda e^{-\lambda} . e^{\lambda} = \underline{\underline{\lambda}}$$

2. Show that variance of P.D is 'λ'.

__Soln__ we know that,

$$P(r) = \frac{e^{-\lambda}\lambda^r}{r!} \; ; r = 0, 1, 2, \text{---}$$

$\therefore$ Variance, $V(X) = E(X^2) - [E(X)]^2$

$$= \sum_{r=0}^{\infty} r^2 p(r) - \mu^2$$

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

$$= \sum_{r=0}^{\infty} [r(r-1) + r]\, P(r) - \mu^2$$

$$= \sum_{r=0}^{\infty} [r(r-1)\, p(r)] + \sum_{r=0}^{\infty} r\, P(r) - \mu^2$$

$$= \sum_{r=0}^{\infty} r(r-1) \frac{e^{-\lambda}\lambda^{r}}{r!} + \mu - \mu^2$$

$$= e^{-\lambda}\left[\sum_{r=0}^{\infty} \frac{\lambda^{r}}{(r-2)!}\right] + \mu - \mu^2$$

$$= e^{-\lambda}\left[\sum_{r=2}^{\infty} \frac{\lambda^{r}}{(r-2)!}\right] + \mu - \mu^2 \quad \Big| \because r-2=y$$

$$= e^{-\lambda}\left[\sum_{y=0}^{\infty} \frac{\lambda^{y+2}}{y!}\right] + \mu - \mu^2$$

$$= e^{-\lambda}.\lambda^2\left[\sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!}\right] + \mu - \mu^2$$

$$= e^{-\lambda}.\lambda^2 e^{\lambda} + \mu - \mu^2 = \cancel{\lambda^2} + \lambda - \cancel{\lambda^2}$$

$$= \lambda$$

Note: Mean & variance of poisson distribution are equal.

3. If a bank received on the average 6 bad cheques per day, find the probability that it will receive 4 bad cheques on any given day.

Soln: Given that

A bank receives 6 bad cheques per day

ie, $\lambda = 6$

we have to find the probability that it will receive 4 bad cheques.

ie, $P(r=4)$

By poisson distribution,

$$P(r=4) = \frac{e^{-\lambda}\lambda^{r}}{r!} = \frac{e^{-6}(6)^4}{4!} = 0.133$$

4. 2% of the items of a factory are defective. The items are packed in boxes. What is the probability that there will be (i) 2 defective items (ii) atleast three defective items in a box of 100 items?

Soln Given that

$$n = 100$$

$$p = 2\% = 0.02$$

$$\lambda = \text{mean} = np = 100(0.02) = 2$$

By P.D,

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

2 defective items

$$P(r=2) = \frac{e^{-2}\cdot(2)^2}{2!} = 0.2706$$

Atleast 3 def. items

$$P(r\geq 3) = P(r=3) + P(r=4) + \text{—} + P(r=100)$$

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

$$= 1 - \frac{e^{-2}(2)^0}{0!} - \frac{e^{-2}(2)^1}{1!} - \frac{e^{-2}(2)^2}{2!} = 0.3233$$

5. If 2% of light bulbs are defective. Find
(i) Atleast one is defective
(ii) Exactly 7 are defective (iii) $P(1 < x < 8)$
in a sample of 100.

6. Average number of accidents on any day on a national highway is 1.8. Determine the probability that the number of accidents are
(i) atleast one (ii) atmost one.

Soln Given that

$$\lambda = 1.8$$

we know that

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

(i) Atleast one

$$P(r \geq 1) = 1 - P(r=0) = 1 - \frac{e^{-\lambda} \lambda^0}{0!} = 1 - e^{-1.8}$$

$$= 0.8347$$

(ii) Atmost one

$$P(r \leq 1) = P(r=0) + P(r=1)$$

$$= \frac{e^{-1.8} (1.8)^0}{0!} + \frac{e^{-1.8} (1.8)^1}{1!} = 0.4628$$

7. A manufacturer of cotter pins knows that 5% of his product is defective. Pins are sold in boxes of 100. He guarantees that not

more than 10 pins will be defective. What is the approximate probability that a box will fail to meet the guaranteed quality?

Sol: Given that

probability of cotter pins to be defective

$$p = 5\% = 0.05$$

Total no. of pins $(n) = 100$

$\therefore$ mean $(\lambda) = np = 100(0.05) = 5$

By P.D, we have

$$P(r) = \frac{e^{-\lambda}\lambda^{r}}{r!} = \frac{e^{-5}.5^{r}}{r!}$$

$\therefore$ The prob. that a box will fail to meet the guarantee $= P(X>10) = 1-P(X\leq 10)$

$$= 1-[P(X=0)+P(r=1)+ \text{——} + P(r=10)]$$

$$= 1-0.9863 = 0.0137$$

8. A car-hire firm has two cars which it hires out day by day. The no. of demands for a car on each day is distributed as a poisson distribution with mean 1.5. Calculate the proportion of days (i) on which there is no demand

(ii) on which demand is refused.

Sol Given that

$\lambda = 1.5$

(i) No demand

$$P(r=0) = \frac{e^{-1.5}(1.5)^0}{0!} = 0.223$$

(ii) demand is refused

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

$$= 1 - \left[0.223 + \frac{e^{-1.5}(1.5)^1}{1!} + \frac{e^{-1.5}(1.5)^2}{2!}\right]$$

$$= 1 - [0.223 + 0.334 + 0.251]$$

$$= 1 - 0.808 = 0.192$$

9. If 'x' is a poisson variate such that $P(x=1) = 24\,P(x=3)$. find μ.

Sol Given that

$$P(x=1) = 24 \cdot P(x=3)$$

$$\frac{e^{-\lambda}\cdot\lambda^1}{1!} = 24\,\frac{e^{-\lambda}\cdot\lambda^3}{3!}$$

$$\Rightarrow \lambda = \frac{\cancel{24}^{4}}{\cancel{6}}\lambda^3 \Rightarrow 4\lambda^3 - \lambda = 0$$

$$\Rightarrow \lambda[4\lambda^2 - 1] = 0$$

$$\Rightarrow \lambda = 0 \quad | \quad 4\lambda^2 - 1 = 0$$

$$4\lambda^2 = 1$$

$$\lambda^2 = 1/4$$

$$\lambda = \pm 1/2$$

$\therefore \boxed{\lambda = 1/2}$ $\quad [\because \lambda > 0$

10. If 'x' is a poisson variate such that $3P(x=4) = \frac{1}{2}P(x=2) + P(x=0)$. find (i) mean (ii) $P(x \le 2)$.

Sol: Given that

$$3P(x=4) = \frac{1}{2}P(x=2) + P(x=0)$$

$$3\left[\frac{e^{-\lambda}\lambda^4}{4!}\right] = \frac{1}{2}\left[\frac{e^{-\lambda}\lambda^2}{2!}\right] + \left[\frac{e^{-\lambda}\lambda^0}{0!}\right]$$

$$\Rightarrow \frac{3}{24}\cancel{e^{-\lambda}}\lambda^4 = \cancel{e^{-\lambda}}\left[\frac{\lambda^2}{4} + \frac{1}{1}\right]$$

$$\Rightarrow \frac{\cancel{3}\lambda^4}{\cancel{24}_{\,8\,2}} = \frac{\lambda^2 + 4}{\cancel{4}} \Rightarrow \lambda^4 = 2\lambda^2 + 8$$

$$\Rightarrow \lambda^4 - 2\lambda^2 - 8 = 0$$

$$\Rightarrow (\lambda^2 - 4)(\lambda^2 + 2) = 0$$

$$\Rightarrow \lambda = 2, -2 \qquad | \because \lambda > 0$$

$$\Rightarrow \boxed{\lambda = 2}$$

$\therefore$ mean $(\mu) = \lambda = 2$

$$P(x \le 2) = P(x=0) + P(x=1) + P(x=2)$$

$$= e^{-2} + \frac{2}{1}e^{-2} + \frac{e^{-2}.2^2}{2!}$$

$$= 3e^{-2} + 2e^{-2} = 5e^{-2} =$$

11. Fit a poisson distribution for the following data

x	0	1	2	3	4	5
f	142	156	69	27	5	1

Sol: Here, $N = \Sigma f = 142 + 156 + 69 + 27 + 5 + 1 = 400$

Now, mean $(\mu) = \frac{\Sigma f x}{N} = \frac{0 + 156 + 138 + 81 + 20 + 5}{400}$

$= \frac{400}{400} = 1$

$\therefore \boxed{\lambda = 1}$

we know that

$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \; ; \; x = 0, \rightarrow 5$

x	observed freq (f)	$P(x)$	Expected or theoritical freq $f(x) = N \cdot P(x)$
0	142	$P(0) = e^{-1}$	$f(0) = 400(e^{-1}) = 147.15$
1	156	$P(1) = e^{-1}$	$f(1) = 400(e^{-1}) = 147.15$
2	69	$P(2) = \frac{e^{-1}}{2}$	$f(2) = 400\left[\frac{e^{-1}}{2}\right] = 73.58$
3	27	$P(3) = \frac{e^{-1}}{6}$	$f(3) = 400\left[\frac{e^{-1}}{6}\right] = 24.53$
4	5	$P(4) = \frac{e^{-1}}{24}$	$f(4) = 400\left[\frac{e^{-1}}{24}\right] = 6.13$
5	1	$P(5) = \frac{e^{-1}}{120}$	$f(5) = 400\left[\frac{e^{-1}}{120}\right] = 1.23$

$\therefore$ The probability distribution is

x	0	1	2	3	4	5	Total
$f(x)$	142	156	69	27	5	1	400
Theoretical freq $f(x)$	147	147	74	25	6	1	400

11. The distribution of typing mistakes committed by a typist is given below assuming the distribution to be poisson, find the expected frequencies.

x	0	1	2	3	4	5
$f(x)$	125	95	49	20	8	3

* Normal Distribution :-

A random variable 'x' is said to have a normal distribution if its density function (or) probability distribution is given by,

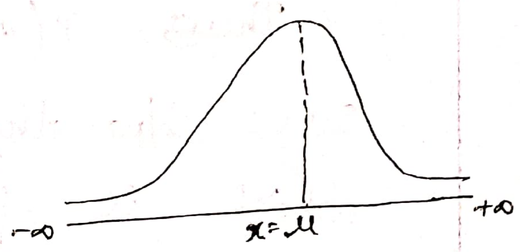
$$f(x, \mu, \sigma) = \frac{1}{\sigma . \sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \quad ; \quad -\infty < x < \infty, \; -\infty < \mu < \infty \;\&\; \sigma > 0$$

where 'μ' is the mean & 'σ' is the standard deviation of 'x'.

→ Here 'μ' & 'σ' are known as parameters of the normal distribution.

Note:-

→ The curve has maximum value at $x=\mu$ & tapers off on either side but never touches the horizontal line.



→ The curve on the left side goes upto $-\infty$, & on right side it goes upto $+\infty$.

→ However as much as 99.73% of the area under the curve lies between $(\mu - 3\sigma)$ & $(\mu + 3\sigma)$ & only 0.27% of the area lies beyond these points.

→ The random variable 'x' is then said to be normal random variable (or) normal variate & the curve representing normal distribution is known as normal curve.

→ The total area bounded by the curve & the x-axis is 1.

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

→ The area under the curve between the ordinates $x=a$ & $x=b$ where $a<b$ represents the probability that 'x' lies b/w 'a' & 'b'

ie, $P(a<x<b)$.

→ Thus, $P(a<x<b)$ = area under the normal curve b/w the vertical lines $x=a$ & $x=b$ which is $\int_a^b f(x)\, dx$.

→ A random variable 'x' with mean 'μ' & variance σ^2 then the probability law of normal distribution is expressed by

$x \sim N(\mu, \sigma^2)$.

* **Normal distribution as a limiting form of Binomial distribution :-**

Normal distribution is a limiting case of the Binomial distribution under the following conditions.

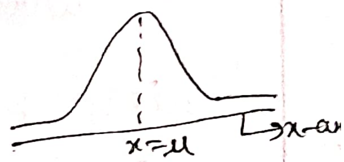
(i) n, the no. of trials is indefinitely large ie, $n \to \infty$.

(ii) neither 'p' nor 'q' is very small.

* **Chief characteristics of Normal distribution :-**

→ The graph of normal distribution $y = f(x)$ in the xy-plane is known as normal curve. ~~$y=f(x)$~~.

→ The curve is bell shaped and symmetrical about the line $x = \mu$.



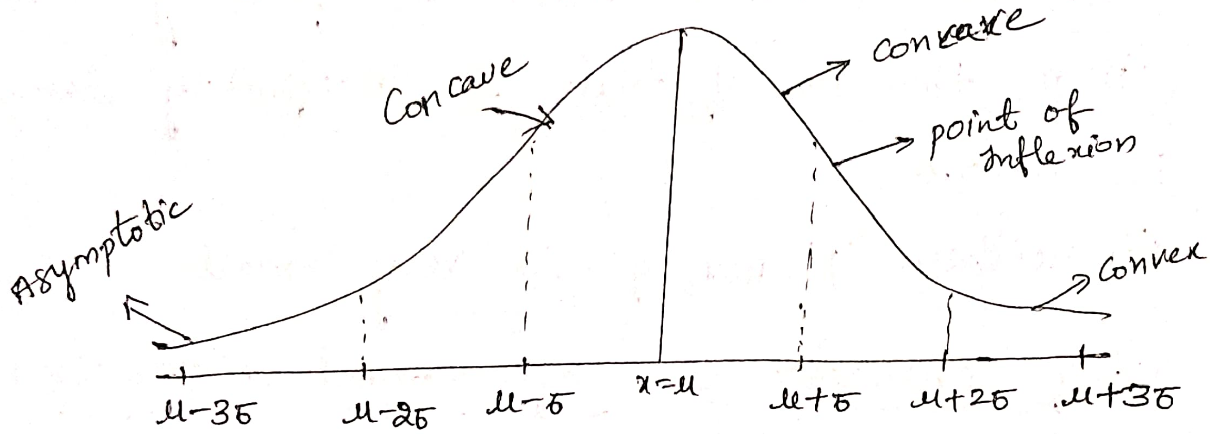
→ The area under the normal curve represents the total probability (population).

→ Mean, median & mode of the distribution coincide so the curve is unimodel. ie, it has only one maximum point.

→ x-axis is an asymptote to the curve.

→ The linear combination of independent normal variates is also a normal variate.

→ The points of inflexion of the curve are at $x = \mu \pm \sigma$ & the curve changes from concave to convex at $x = \mu + \sigma$ & to $x = \mu - \sigma$.



→ The probability that the normal variate 'x' with mean 'μ' & standard deviation σ lies b/w x_1 & x_2 which is given by $P[x_1 \leq X \leq x_2]$.

$$\text{ie, } P[x_1 \leq X \leq x_2] = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \cdot dx \quad \text{①}$$

<u>Note</u>: Since ① depends on two parameters 'μ' & 'σ' we get different normal curves for different values of 'μ' & 'σ' & it is difficult task to plot all such normal curves.

Instead, by substituting $z = \frac{x-\mu}{\sigma}$, the R.H.S of ① becomes independent of 'μ' & 'σ'.

Here 'z' is known as the <u>standard variable</u>.

* Standard normal Distribution :-

The normal distribution with mean $\mu = 0$ & standard deviation $\sigma = 1$ is known as standard normal distribution.

Here the random variable that follows this distribution is 'z' & is given by $z = \frac{x-\mu}{\sigma}$

$$\therefore \; P(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2}\, dz \; ; \quad -\infty < z < \infty .$$

* How to find probability Density of normal curve :-

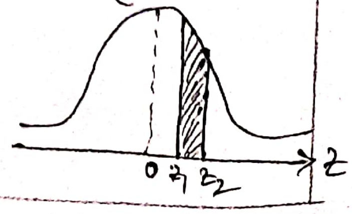
The probability that the normal variate 'X' with mean 'μ' & standard deviation 'σ' lies b/w two specific values x_1 & x_2 with $x_1 \leq x_2$ can be obtained using area under the standard normal curve as follows.

step-1 :- perform the change of scale $z = \frac{x-\mu}{\sigma}$ & find z_1 & z_2 corresponding to the values of x_1 & x_2 respectively.

step-2 :- To find $P[x_1 \leq X \leq x_2] = P[z_1 \leq z \leq z_2]$

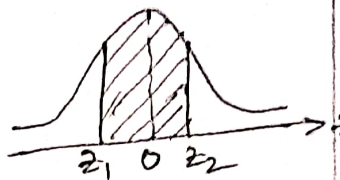
Case-(i) :- If both z_1 & z_2 are positive (both negative) then

$$P[x_1 \leq X \leq x_2] = |A(z_2) - A(z_1)|$$



Case-(iii):- If $z_1 < 0$ & $z_2 > 0$ then

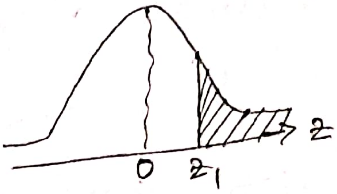
$$P[x_1 \le X \le x_2] = A(z_2) + A(z_1)$$



Step-3:- To find $P[z > z_1]$

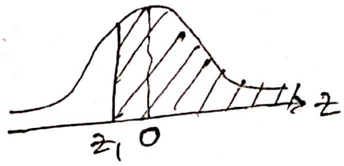
Case-(i):- If $z_1 > 0$ then

$$P[z > z_1] = 0.5 - A(z_1)$$

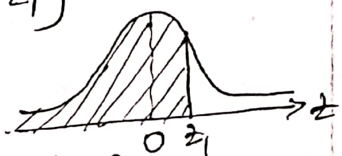


Case-(ii):- If $z_1 < 0$ then

$$P[z > z_1] = 0.5 + A(z_1)$$



Step-4:- To find $P[z < z_1] = 1 - P[z > z_1]$



Case-(i):- If $z_1 > 0$ then

$$P[z < z_1] = 1 - P[z > z_1] = 0.5 + A(z_1)$$

Case-(ii):- If $z_1 \le 0$ then

$$P[z < z_1] = 0.5 - A(z_1)$$

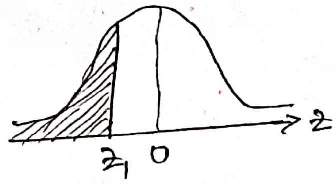
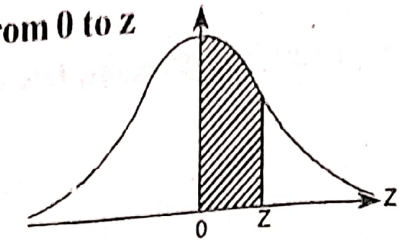


Table – 3

Areas under the Standard Normal Curve from 0 to z



$$Z = \frac{X - \mu}{\sigma}$$

0.81

z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1256	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1916	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2649
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.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	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4654	.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	.4761	.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	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
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	.4979	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

* **Estimation:-**

* **Parameter:-**

Quantities appearing in the distributions such as 'p' in binomial distribution & 'μ' & 'σ' in normal distribution are known as parameters.

* **Estimate:-**

An estimate is a statement made to find an unknown population parameter.

* **Estimator:-**

The procedure or rule to determine an unknown population parameter is known as an estimator.

* **Estimation:-**

Estimation is the process to use the statistics obtained from the sample as estimate of the unknown parameter of the population from which the sample is drawn.

Note:- A parameter can have one (or) two (or) many estimators.

* **Types of estimation :-**

There are two kinds of estimates to determine the statistic of a population parameter namely,

(i) point estimation (ii) Interval estimation.

1. **Point estimation :-** If an estimate of the population parameter given by a single value then the estimate is known as point estimation of the parameter.

Ex :- If height of a student is measured as 162 cm then the measurement gives a point estimation.

2. **Interval Estimation :-** If an estimate of a population parameter is given by two different values b/w which the parameter may be considered to lie then the estimate is known as interval estimation of the parameter.

Ex :- If the height is given as (163 ± 3.5) cm then the height lies b/w 159.9 & 166.5 cm & the measurement gives an interval estimation.

Note :-

→ A point estimate of a parameter 'θ' is a Single numerical value which is computed from a given sample & serves as an approximation of the unknown Exact value of the parameter.

→ An estimate / estimator is denoted by $\hat{\theta}$ & parameter is denoted by θ.

Properties of estimation :-

A good estimator is one which is as close to the true value of the parameter as possible.

* The important properties of a good estimator are

1. Consistency
2. unbiasedness
3. Efficiency
4. Sufficiency

→ An estimator $\hat{\theta}_n$ of a parameter 'θ' is Consistant if it converges to 'θ' as $n \to \infty$.

→ A statistic $\hat{\theta}$ is said to be an unbiased estimate of parameter 'θ' if $E(\hat{\theta}) = \theta$.

→ A Statistic $\hat{\theta}_1$ is said to be more efficient unbiased estimator of the parameter 'θ' than the Statistic $\hat{\theta}_2$ if

(i) $\hat{\theta}_1$ & $\hat{\theta}_2$ are both unbiased estimators of 'θ'.

(ii) $V(\hat{\theta}_1) < V(\hat{\theta}_2)$

Then $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$.

→ An estimator is said to be <u>sufficient</u> for a parameter 'θ' if it contains all the information with the sample regarding the parameter.

* <u>Statistical Inference</u> :-

It is defined as "the process by which we draw a conclusion about some measure of a population based on a sample value. The measure might be a variable, such as the mean, S.D etc. The purpose of sampling is to estimate some characteristics for the population from which the sample is selected."

→ Basically, there are two types of problems under statistical Inference.

1. <u>Hypothesis testing</u> :- to test some hypothesis about parent population from which the sample is drawn.

2. <u>Estimation</u> :- to use the statistics obtained from the sample as estimate of the unknown parameter of the population from which the sample is drawn.

Note: An important problem of the statistical inference is the estimation of population parameters (ie, population mean, population S.D etc) from the corresponding sample statistics (ie, sample mean, sample S.D etc.).

* Confidence Interval:

Confidence Interval is a range in which a certain percentage of normally distributed values lie. (or) Confidence Interval is a probability that a single sample result will fall in certain range of normally distributed values.

Formulae:

→ If $\bar{x}$ is a mean of random sample of size 'n' from population with known variance 'σ^2', $(1-\alpha)100\%$ confidence interval for 'μ' is given by

$$\bar{x} - z_{\alpha/2}\left[\frac{\sigma}{\sqrt{n}}\right] < \mu < \bar{x} + z_{\alpha/2}\left[\frac{\sigma}{\sqrt{n}}\right] \quad \text{(for large sample)}$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Note: Standard error $(S.E) = \frac{\sigma}{\sqrt{n}}$

→ The maximum error of estimate with $1-\alpha$ probability is given by

$$\text{Max. Error } (E) = z_{\alpha/2} \cdot \left[\frac{\sigma}{\sqrt{n}}\right]$$ (large sample)

→ When α, E, σ are known then the Sample size 'n' is given by

$$n = \left[\frac{z_{\alpha/2} \cdot \sigma}{E}\right]^2$$

→ When σ is unknown, in this case 'σ' is replaced by 's', the S.D of sample to determine maximum error with $(1-\alpha)$ probability.

$$\text{Max. error } (E) = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$ (small samples)

→ If $\bar{x}$ & 's' are the mean & standard deviation of a random sample from a normal population with unknown variance σ^2, $(1-\alpha)100\%$ confidence interval for 'μ' is given by

$$\bar{x} - t_{\alpha/2}\left[\frac{s}{\sqrt{n}}\right] < \mu < \bar{x} + t_{\alpha/2}\left[\frac{s}{\sqrt{n}}\right].$$ (Small samples)

where $t_{\alpha/2}$ is a t-value with $\nu = n-1$ degrees of freedom leaving an area of $\alpha/2$ to the right.

→ Maximum error of estimate of true proportion is given by,

$$E = z_{\alpha/2}\sqrt{\frac{PQ}{n}} = z_{\alpha/2}\sqrt{\frac{pq}{n}}$$ | ∵ If 'P' is not known take $P = p$

→ Confidence interval is given by,

$$p \pm z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

where, p = sample proportion & P = population proportion.

→ Maximum error of estimate for true proportion is given by,

$$E = z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$; If 'P' is known.

→ Confidence interval is given by

$$P \pm z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

→ The sample size 'n' of unknown proportion 'P' is given by,

$$n = \left[\frac{z_{\alpha/2}}{E}\right]^2 .(P.Q)$$ | If 'P' is not given Consider $P = 1/2$

where E = max. error

Note: The fraction $(1-\alpha)$ is known as the confidence coefficient (or) degree of confidence.

→ 95% confidence → $z_{\alpha/2} = 1.96$

98% confidence → $z_{\alpha/2} = 2.33$

99% confidence → $z_{\alpha/2} = 2.58$

90% confidence → $z_{\alpha/2} = 1.645$

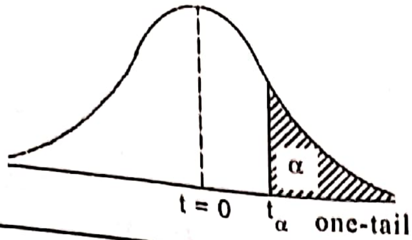
→ A sample of size > 30 is known as large sample.

→ A sample of size ≤ 30 is known as small sample.

→ t - distribution: It is used for testing of hypothesis when the sample size is small & population S.D 'σ' is not known.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$ where $\nu = n-1$ degrees of freedom

t_α - Critical Values of the t-Distribution with ν Degrees of Freedom Table – 4



ν	α = 0.40	0.30	0.20	0.15	0.10	0.05	0.025	0.01	0.005
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Note : The above table gives the values of t for one-tail test (either left-tail or right-tail test). If we have to find the value of t for a two-tail test at a level, we take the value of $\alpha / 2$ for α. For example, the value of t at 5% level with 9 d.f. is $t_{0.025} = 2.262$ and the value of t at 1% level with 11 d.f. is $t_{0.005} = 3.106$