Percentiles: Percentile are the points which divide the entire scale of measurement into 100 equal parts. Usually, percentile point are represented by $P_0, P_1, P_2, \ldots P_{100}$.

First percentile may be defined as that point in a frequency distribution below which lie one percent of the total scores. Similarly, P_{20} or 20th percentile may be defined as that point in a frequency distribution below which 20% percent of the measure of score fall. On the similar pattern we define other percentile points. It should be noted that P_{25} is also known as first Quartile, P_{50} is known as median & P_{75} is also third quartile. As shown above, P_0 lies at the [redacted] beginning of the distribution and P_{100} at the end of the distribution. In fact, percentile points provide us the information about the relative position of scores in a distribution. The percentile points are calculated with help of below mentioned formula

$$P_p = L + \frac{[pN - cf_b]}{f_p} \times i$$

Where
P_p : stands for percentage of the distribution wanted.
L : Stands for lower limit of the class interval in which desired percentile falls
pN : Part of N to be counted off in order to reach desired point
cf_b : Sum of all scores upto interval below L.
f_p : Number of scores [redacted] within the interval upon which percentile falls and
i : Length / Size of the class Interval

②

Q. Compute the value of P_{10} & P_{30} on the below mentioned data.

C.I	f
60-69	4
50-59	6
40-49	10
30-39	5
20-29	3
10-19	2

Solution:

C.I	f	cf	
60-69	4	30	
50-59	6	26	
40-49	10	20	
30-39	5	10	← P_{30}
20-29	3	5	← P_{10}
10-19	2	2	
	N=30		

We know that

$$P_p = L + \left[\frac{PN - cf_b}{f_p}\right] \times i$$

Now, P_{10}

$PN = 10\%$ of $N = \frac{10}{100} \times 30 = 3$

$\therefore \boxed{PN = 3}$

Here $cf_b = 2$, $f_p = 3$, $L = 19.5$ & $i = 10$

$$\therefore P_{10} = 19.5 + \left[\frac{3-2}{3}\right] \times 10$$

$$= 19.5 + \frac{10}{3}$$

$$= 19.5 + 3.33$$

$$= 22.83$$

i.e; $\boxed{P_{10} = 22.83}$

Now, P_{30}

$PN = 30\%$ of $N = \frac{30}{100} \times 30 = 9$

i.e; $\boxed{PN = 9}$

Here $cf_b = 5$, $f_p = 5$, $L = 29.5$ & $i = 10$

$$\therefore P_{30} = 29.5 + \left[\frac{9-5}{5}\right] \times 10$$

$$= 29.5 + 8$$

$$= 37.5$$

i.e; $\boxed{P_{30} = 37.5}$

Hence $P_{10} = 22.83$ & $P_{30} = 37.5$

#

Scanned by CamScanner

Questions for Practice (3)

1 Compute the value of P_{50} and P_{75} on the below mentioned data.

C.I	f
65-69	2
60-64	4
55-59	10
50-54	7
45-49	4
40-44	3

2. Calculate the value P_{10}, P_{90} and P_{30} on the below mentioned data

Scores	f
70-79	5
60-69	15
50-59	10
40-49	10
30-39	5
20-29	3
10-19	2

Percentile Rank: It is the point in the distribution below which a given percentage of scores fall. In other words, we can say that percentile rank is the position or rank of an individual obtained on the basis of percent of cases falling below the score attained by an individual in the distribution. It is generally denoted by PR & it indicates percent of all the scores in a frequency distribution that fall below a given raw score.

The procedure followed in calculating percentile rank is reverse to the process of percentiles. The formula for computing the value of PR is given below

$$PR = \frac{100}{N}\left[cf_b + \frac{X-L}{i} \times f_p\right]$$

OR

$$PR = \frac{f_p(X-L) + i\,Cf_b}{Ni} \times 100$$

where

PR : stands for Percentile Rank

N : Total number of cases/scores in the distribution

L : Lower limit of the class interval in which PR falls.

Cf_b : Number of cummulative frequencies below the class interval in which PR falls.

f_p : Actual frequencies present against the class interval in which PR falls.

i : Size of the class interval and

X : Raw score for which percentile rank is to be found out.

Q. Compute the value of PR of 36 on the below mentioned data.

C.I	f
50-54	2
45-49	4
40-44	6
35-39	12
30-34	7
25-29	5
20-24	4

Solution:

C.I	f	cf
50-54	2	40
45-49	4	38
40-44	6	34
35-39	12 ← fp	28
30-34	7	16 ← cfb
25-29	5	9
20-24	4	4

We know that $PR = \frac{f_p(X-L)+i\,cf_b}{N.i} \times 100$

Here $X = 36$, $cf_b = 16$, $f_p = 12$, $L = 34.5$, $N = 40$ & $i = 5$

$\therefore$ PR of 36 $= \frac{12(36-34.5)+5\times16}{40\times5} \times 100$

$= \frac{12(1.5)+80}{200} \times 100$

$= \frac{18+80}{200} \times 100$

$= \frac{98}{200} \times 100$

i.e; $\boxed{\text{PR of } 36 = 49}$: Hence, the PR of 36 is 49. It means that there are 49% of cases which fall below the score of 36 #

Alternate

$$PR = \frac{100}{N}\left[cf_b + \frac{X-L}{i} \times f_p\right]$$

But here $X = 36$, $cf_b = 16$, $f_p = 12$, $N = 40$, $L = 34.5$ & $i = 5$

$$\therefore PR \text{ of } 36 = \frac{100}{40}\left[16 + \frac{36-34.5}{5} \times 12\right]$$

$$= \frac{100}{40}\left[16 + \frac{1.5}{5} \times 12\right]$$

$$= \frac{100}{40}\left[16 + \frac{18}{5}\right]$$

$$= \frac{100}{40}\left[16 + 3.6\right]$$

$$= \frac{100}{40} \times 19.6$$

$$= \frac{1960}{40}$$

$$= 49$$

i.e; $\boxed{PR \text{ of } 36 = 49}$

Hence, the PR of 36 is 49. It means that there are 49% of the cases which fall below the score of 36.

#

Questions for Practice

Q.1 Compute the value of PR of 28 on the below mentioned data.

C.I	f
30-34	1
25-29	8
20-24	9
15-19	10
10-14	14
5-9	10

Q.2 Calculate the value of PR of 127 + PR of 138 on the below mentioned data.

C.I	f
140-144	10
135-139	11
130-134	12
125-129	15
120-124	14
115-119	13
110-114	5

Q.3 Compute the value of P_{50} for the above mentioned data in Q No 1 and Q.No 2.

#

Questions for Practice (Additional)

Q1 Find the value of mean, median and mode

C.I	f
27-31	5
22-26	6
17-21	9
12-16	10
7-11	6
2-6	4

Q.2. Determine the value of Correlation by Product Moment Method and Rank difference Method.

X	Y
10	12
20	12
15	10
05	12
10	15
30	10
25	05

Q.3. Compute the value of P_{30} and PR of 45 on the below mentioned data.

Scores	f
80-89	0
70-79	2
60-69	8
50-59	15
40-49	20
30-39	10
20-29	3
10-19	2

#

The concept of normal curve was first developed by a French mathematician, De Moivre (1733) on the basis of mathematical equation. Gauss and Laplace, German mathematicians further developed the concept of the curve and probability. The curve and probability are two concepts which are intimately connected.

The probability of a given event is defined as expected frequency of occurrence of this event among events of a like sort. It can be best explained either through coin-tossing or dice-throwing etc. The unbiased coin-tossing helps to understand how the law of chance operates in determining probability of the occurrence of an event. The probability is determined by examining the number of possible outcomes. For example, if ten coins are tossed into air a large number of times, the most common combination of results would be 5 heads and 5 tails. Somewhat less often 4 heads and 6 tails or 6 tails and 4 heads would come up. It would be indeed rare to obtain 10 heads or no tails or 10 tails and no heads. If the frequency with which each combination of heads and tails were depicted on a graph, a rough approximation of normal distribution or limiting form of binomial distribution would be obtained. The binomial distribution for the number of heads appearing on throw of 'n' coins correspond to $\left(\frac{1}{2}+\frac{1}{2}\right)^n$ and as n grows larger this distribution approaches normal probability. The binomial for n things can be expanded as such :

$$(P+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{(1)(2)}p^{n-2}q^2 \frac{n(n-1)(n-2)}{(1)(2)(3)}p^{n-3}q^3 + \_\_ q^n$$

In other words, normal distribution is a special kind of binomial distribution with p = ·5 and as n grows infinitely larger, the Normal and binomial probabilities become identical.

It is observed that test scores always tend to be distributed around the averages. There is a tendency for many students to get average or near average score than the high or low scores on many tests. If a test is neither too easy nor too difficult for the group, scores will usually be distributed in a manner that approximates the normal curve (Lord, 1955). Even behavioural scientists and educators believe that many of the traits when studied such as intellectual, physical etc. generate sets of scores in an approximately normal fashion. It is believed that in total population, many human abilities and skills tend to be distributed in a manner similar to the normal curve. In actual practice, the form of distribution may or may not be normal but the characteristics measured may be normal in general population. The reason being that the distribution of scores is dependent not only on the characteristics being measured but also on the manner in which measured, the individuals being measured and a number of chance factors etc. So all variables dealt within testing are considered to be normally distributed and present an idealized distribution of scores for any group on some tests. It is assumed that large samples are often required to obtain a curve which reflects normality. When the distribution on a small sample is plotted, the shape will be irregular or uneven.

The normal distribution is a special kind of distribution whose relationship of height to width at every score is mathematically specified. The normal distribution is not a distribution of actual scores but a theoretical distribution plotted from a mathematical equation as described below :

$$y = \frac{N}{\sigma\sqrt{\pi}} e^{1/2} \left(\frac{\pi}{\sigma}\right)^2 \qquad ...(9)$$

In which,

y = Height of curve for a particular value of X

π = 3·1416 a constant

e = 2·7183 base of Napierian Logarithms

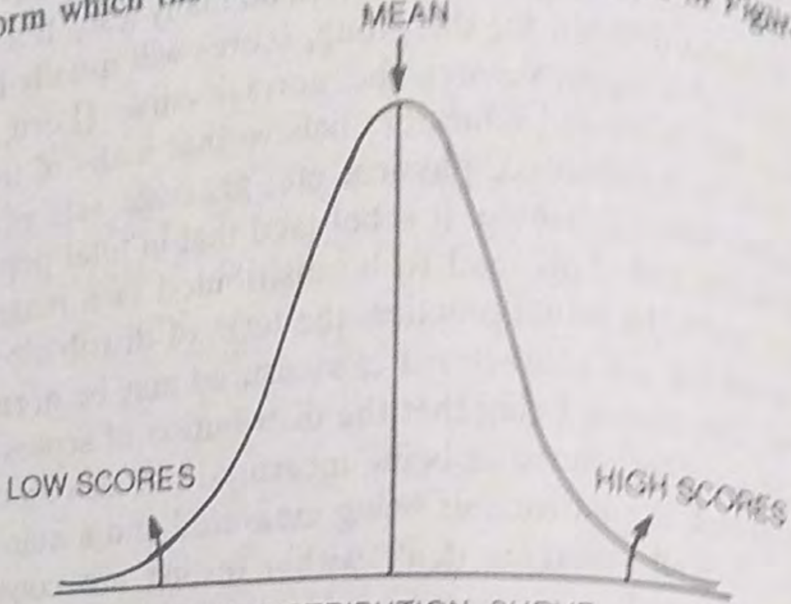
x = Scores expressed as deviation from the mean

σ = Standard deviation of the distribution

N = Number of cases in the sample

If for example, abilities of large number of individuals, were graphically depicted by a curve, most of the people will be located around the averages which fall at the centre of the distribution. In addition there will be lesser number of individuals whose scores would fall at the two extremes of the curve. The averages *i.e.* mean, median and mode are located in the same central position bisecting area under the curve into two equal halves. The random samples from a population usually conform to the ideal bilaterally symmetrical curve. The obtained scores have to be converted into Z scores. The Z score (Sigma score) is used as a yardstick to mark off distances around the mean on the baseline of the curve.

The form which the curve usually takes is shown in Figure 9.1.



NORMAL DISTRIBUTION CURVE

Fig. 9.1.

This curve is unimodal and symmetrical about the central point which is the mean. The mesokurtic or bell-shaped curve represents this tendency and is used as a standard for a normal curve model. The normal curve is usually written in standard score (Z) form. Standard scores have a mean of 0 and standard deviation of 1. The area under the curve is taken as unity, that is, N = 1. It is also referred to as unit normal curve in standard deviation units because the value of the area under the curve is set at unity to interpret the meaning of scores in the distribution. In case when values of N and σ are different from one, then scores are converted into Z scores. In order to convert raw scores into Z scores, the below given Formula is used. The score then gets presented in terms of the unit of standard deviation.

$$Z = \frac{X - M}{\sigma} \qquad \text{......(6)}$$

$\because \quad X - M = x$

$\therefore$ Formula (6) can also be written as :

$$Z = \frac{x}{\sigma}$$

Here, Z is a standard score on X and is equal to $(X-M)/\sigma$. The score Z is a deviation in standard deviation units measured along the base line of the curve from a mean of 0. Deviations to the right of the mean being positive and those to the left negative. The curve has a unit area and a unit standard deviation. Here we find the ordinates and other values of normal curve for various standard deviation units from the mean.

In actual practice, the students are not required to calculate different values of Z or of different values of Y or ordinate i.e. vertical lines drawn from curve to the base line to demarcate different points for

Characteristics or properties of Normal curve :

The following is a summary of the characteristics of normal curve :

1. The curve is bell shaped.
2. The curve is symmetrical.
3. In the curve, the mean, median and mode fall at the same place.
4. The mean, median and mode have the same numerical value.
5. It is a unimodal distribution.

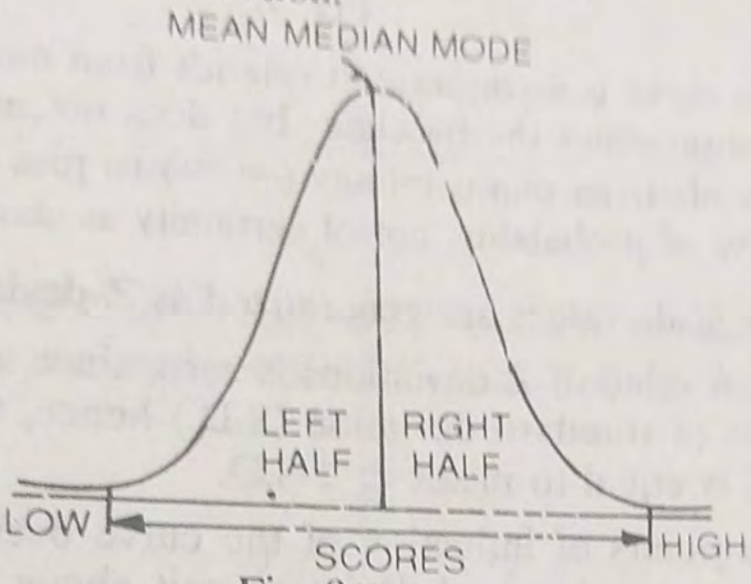


Fig. 9.2

6. The area under the curve represents the total frequency (N) of the distribution.
7. It is bilateral *i.e.* it is divided into two equal halves *i.e.* right and left half from the highest central points.
8. The maximum ordinate of the curve occurs at the mean, that is, where Z = 0 and the value of the highest ordinate is 0·3989. (This value of *y* is obtained when we select mean as the first point on the base line. For the mean, the deviation value, *x*, is equal to 0. Since *x* is equal to O, the exponent of *e* is 0 and any quantity raised to 0 power is equal to 1. Then Equation (9) becomes :

$$y = \frac{1}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\,(3\cdot 1416)}} = \frac{1}{\sqrt{6\cdot 2832}} = \frac{1}{2\cdot 50}$$

$$= 0\cdot 3989$$

9. The first and third quartiles are at equal distance from the median.

10. Q is equal to the probable error (PE) which is 0·6745 of the standard deviation.

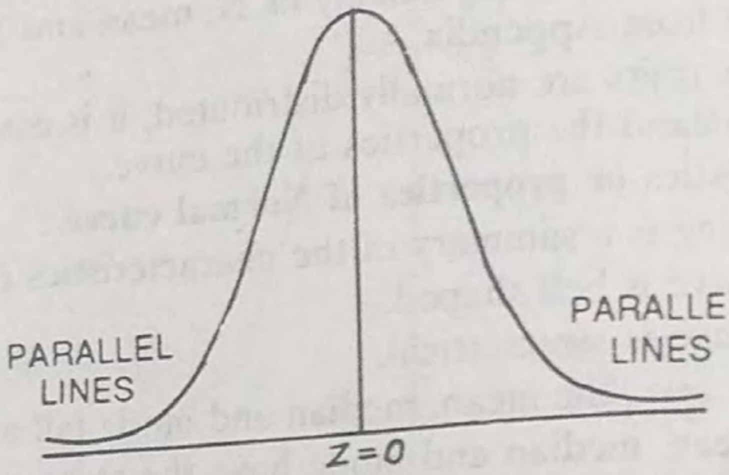
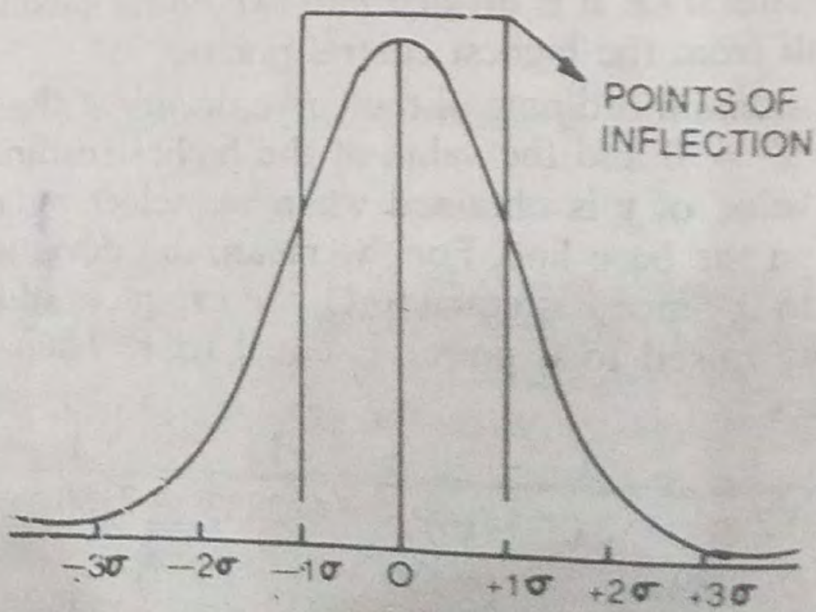


Fig. 9.3

11. The curve is asymptotic. It extends from mean in either direction. It approaches the baseline but does not meet the base line and extends from minus infinity ($-\infty$) to plus infinity ($+\infty$). It is a curve of probability, not of certainity as shown in Fig. 9.3.
12. The scale values are generalized as Z-deviations, $\left(\frac{X-M}{\sigma}\right)$. The mean value of Z deviations is zero, since all the Z values are in units of standard deviation (S.D.) hence, the range of mean $\pm$ 1 Z is equal to mean $\pm$ 1 S.D.
13. The points of inflection of the curve occur at points plus and minus one standard deviation unit above and below the mean. It is at these points that the tails of the curve start decreasing. Thus, the curve changes from convex to concave in relation to horizontal axis at these points.



14. Mean deviation is 0·7979 of standard deviation.

15. For practical purposes, the base line of the curve is divided into 6 sigma distances ranging from -3σ to $+3\sigma$ (Fig 9.5). The unit of distance is used as a standard to divide the base line into equal segments. Each segment is one standard unit of distance or one standard deviation from the mean. Very few scores deviate by more than 3 standard deviations above or below the mean.

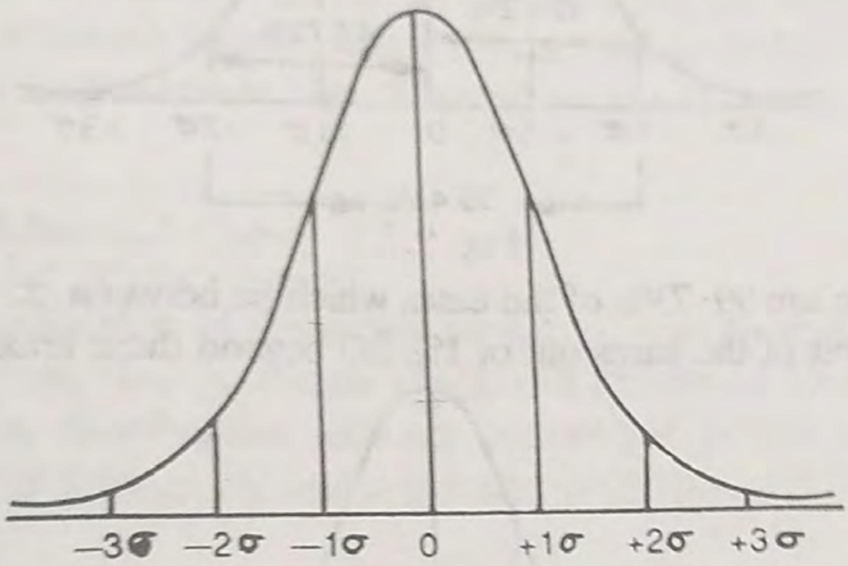


Fig. 9.5

16. In normal curve, 1σ on each side of the mean includes 68·26 per cent of the curve or roughly two-third of the cases as shown in Fig. 9.6.

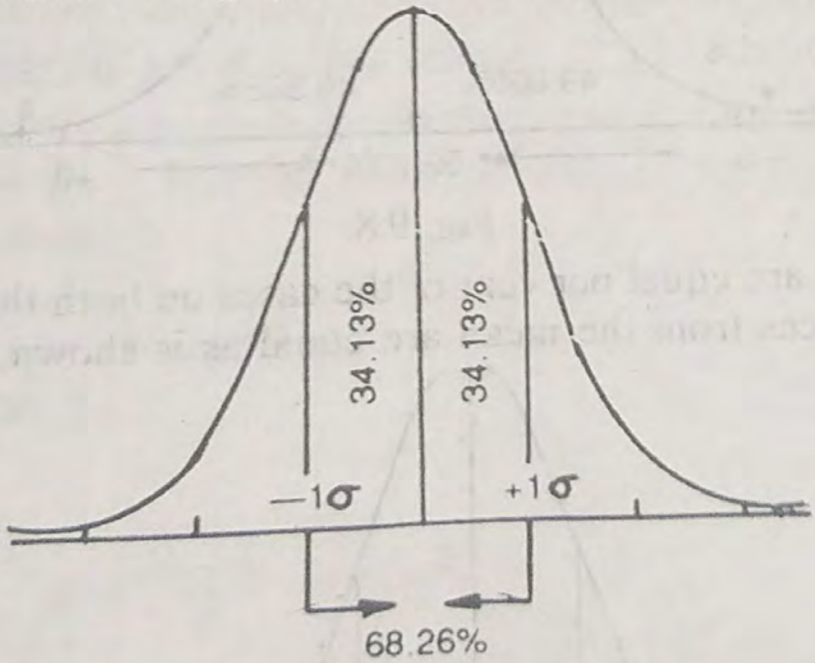


Fig. 9.6.

17. 95% of the cases fall between 2σ on both sides of the mean. One third of the scores lie by more than one standard deviation from the mean. Consequently, one-sixth of the total scores fall above $+1\ \sigma$ from the mean and one sixth fall below $-1\ \sigma$ from the mean. (See Fig. 9.7)

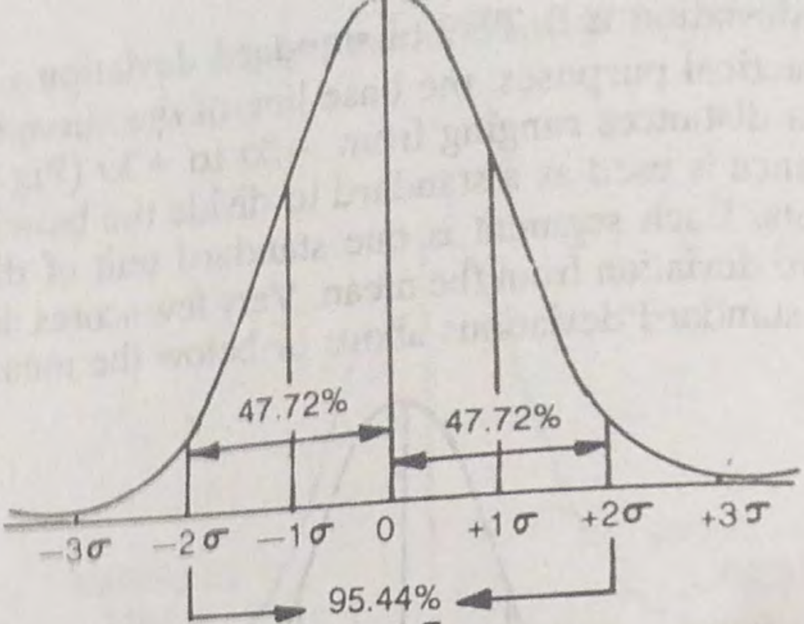


Fig. 9.7.

18. There are 99·73% of the cases which lie between ± 3σ distance and the rest of the cases out of 1% fall beyond these limits. (See Fig. 9.8)

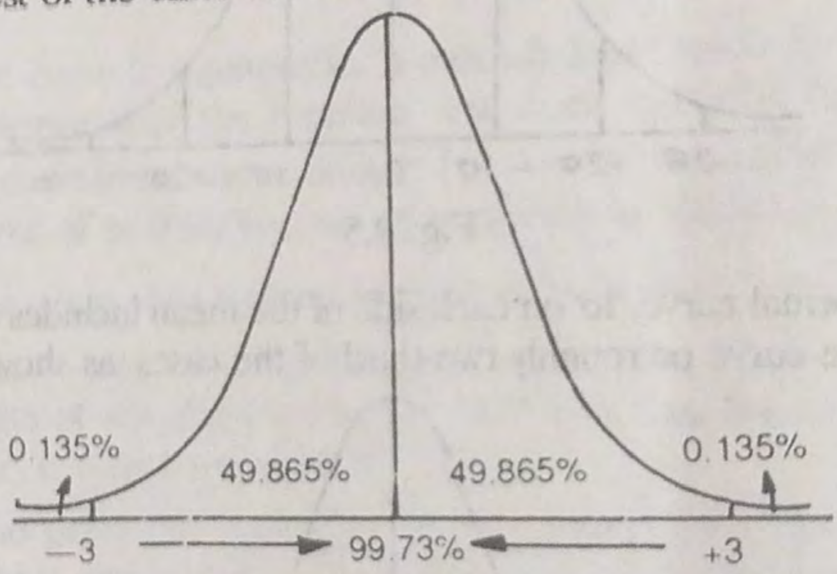


Fig. 9.8.

19. There are equal per cent of the cases on both the sides, if sigma distances from the mean are equal as is shown in Fig. 9.9.