UNIT 1 CHARACTERISTICS OF NORMAL DISTRIBUTION

Structure

1.0 INTRODUCTION

So far you have learnt in descriptive statistics, how to organise a distribution of scores and how to describe its shape, central value and variation. You have used histogram and frequency polygon to illustrate the shape of a frequency distribution, measures of central tendency to describe the central value and measures of variability to indicate its variation. All these descriptions have gone a long way in providing information about a set of scores, but we also need procedures to describe individual scores or cutting point scores to categorize the entire group of individuals on the basis of their ability or the nature of test paper, which a psychometerician or teacher has used to assess the outcomes of the individual on a certain ability test. For example, suppose a teacher has administered a test designed to appraise the level of achievement and a student has got some score on the test. What did that score mean? The obtained score has some meaning only with respect to other scores either the teacher may be interested to know how many students lie within the certain range of scores? Or how many students are above and below certain referenced score? Or how many students may be assign A, B, C, D etc. grades according to their ability?

To have an answer to such problems, the curve of Bell shape, which is known as Normal curve, and the related distribution of scores, through which the bell shaped curve is obtained, generally known as Normal Distribution, is much helpful.

Thus the present unit presents the concept, characteristics and use of Normal Distributions and Normal Curve, by suitable illustrations and explanations.

1.1 OBJECTIVES

After reading this unit, you will be able to:

- Explain the concept of normal distribution and normal probability curve;
- Draw the normal probability curve on the basis of given normal distribution;
- Explain the theoretical basis of the normal probability curve;
- Elucidate the Characteristics of the normal probability curve and normal distribution;
- Analyse the normal curve obtained on the basis of large number of observations;
- Describe the importance of normal distribution curve in mental and educational measurements;
- Explain the applications of normal curve in mental measurement and educational evaluation;
- Read the table of area under normal probability curve;
- Compare the Non-Normal with normal Distribution and express the causes of divergence from normalcy; and
- Explain the significance of skewness and kurtosis in the mental measurement and educational evaluation.

1.2 NORMAL DISTRIBUTION/ NORMAL PROBABILITY CURVE

1.2.1 Concept of Normal Distribution

Carefully look at the following hypothetical frequency distribution, which a teacher has obtained after examining 150 students of class IX on a Mathematics achievement test.

Table 1.2.1: Frequency distribution of the Mathematics achievement test scores

Class Intervals	Tallies	Frequency
85 – 89	I	1
80 – 84	II	2
75 – 79	IIII	4
70 – 74	~~IIII~~ II	7
65 – 69	~~IIII~~ ~~IIII~~	10
60 – 64	~~IIII~~ ~~IIII~~ ~~IIII~~ I	16
55 – 59	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	20
50 – 54	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	30
45 – 49	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	20
40 – 44	~~IIII~~ ~~IIII~~ ~~IIII~~ I	16
35 – 39	~~IIII~~ ~~IIII~~	10
30 – 34	~~IIII~~ II	7
25 – 29	IIII	4
20 – 24	II	2
15 – 19	I	1
	Total	**150**

Are you able to find some special trend in the frequencies shown in the column 3 of the above table? Probably yes! The concentration of maximum frequencies (f = 30) lies near a central value of distribution and frequencies gradually tapper off symmetrically on both the sides of this value.

1.2.2 Concept of Normal Curve

Now, suppose if we draw a frequency polygone with the help of above distribution, we will have a curve as shown in the fig. 1.2.1

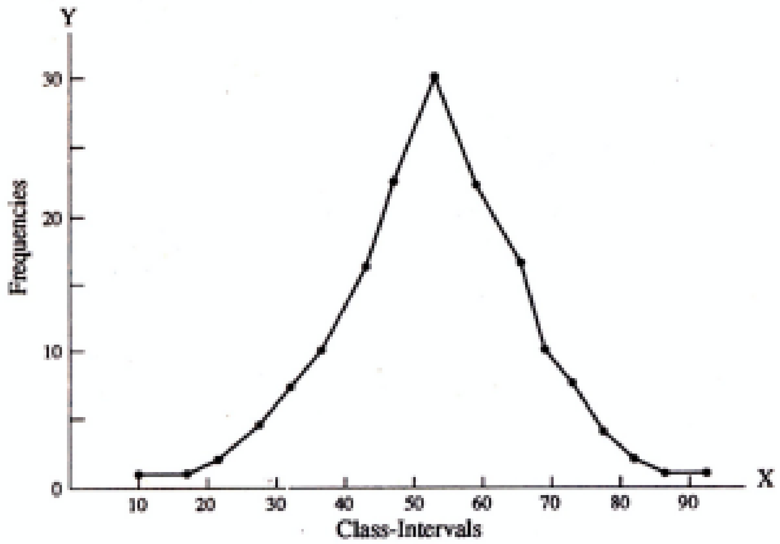

Fig. 1.2.1: Frequency Polygon of the data given in Table 1.2.1

The shape of the curve in Fig. 1.2.1 is just like a 'Bell' and is symmetrical on both the sides.

If you compute the values of Mean, Median and Mode, you will find that these three are approximately the same (M = 52; Md = 52 and Mo = 52)

This Bell shaped curve technically known as Normal Probability Curve or simply Normal Curve and the corresponding frequency distribution of scores, having just the same values of all three measures of central tendency (Mean, Median and Mode) is known as Normal Distribution.

Many variables in the physical (e.g. height, weight, temperature etc.) biological (e.g. age, longevity, blood sugar level and behavioural (e.g. Intelligence; Achievement; Adjustment; Anxiety; Socio-Economic-Status etc.) sciences are normally distributed in the nature. This normal curve has a great significance in mental measurement. Hence to measure such behavioural aspects, the Normal Probability Curve in simple terms Normal Curve worked as reference curve and the unit of measurement is described as σ (Sigma).

1.2.3 Theoretical Base of the Normal Probability Curve

The normal probability curve is based upon the law of Probability (the various games of chance) discovered by French Mathematician Abraham Demoiver (1667-1754). In the eighteenth century, he developed its mathematical equation and graphical representation also.

The law of probability and the normal curve that illust-rates it are based upon the law of chance or the probable occurrence of certain events. When any body of observations conforms to this mathematical form, it can be represented by a bell shaped curve with definite characteristics.

1.2.4 Characteristics or Properties of Normal Probability Curve (NPC)

The characteristics of the normal probability curve are:

1) **The Normal Curve is Symmetrical:** The normal probability curve is symmetrical around it's vertical axis called ordinate. The symmetry about the ordinate at the central point of the curve implies that the size, shape and slope of the curve on one side of the curve is identical to that of the other. In other words the left and right halves to the middle central point are mirror images, as shown in the figure given here.

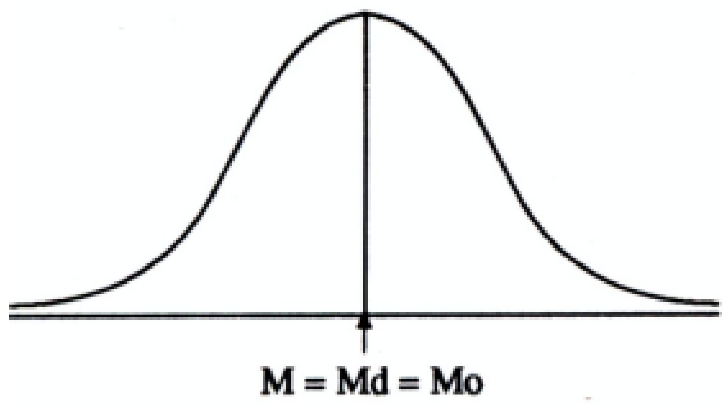

Fig. 1.2.2

2) **The Normal Curve is Unimodel:** Since there is only one maximum point in the curve, thus the normal probability curve is unimodel, i.e. it has only one mode.

3) **The Maximum Ordinate occurs at the Center:** The maximum height of the ordinate always occur at the central point of the curve, that is the mid-point. In the unit normal curve it is equal to 0.3989.

4) **The Normal Curve is Asymptotic to the X Axis:** The normal probability curve approaches the horizontal axis asymptotically; i.e. the curve continues to decrease in height on both ends away from the middle point (the maximum ordinate point); but it never touches the horizontal axis. Therefore its ends extend from minus infinity (- ∞) to plus infinity (+ ∞).

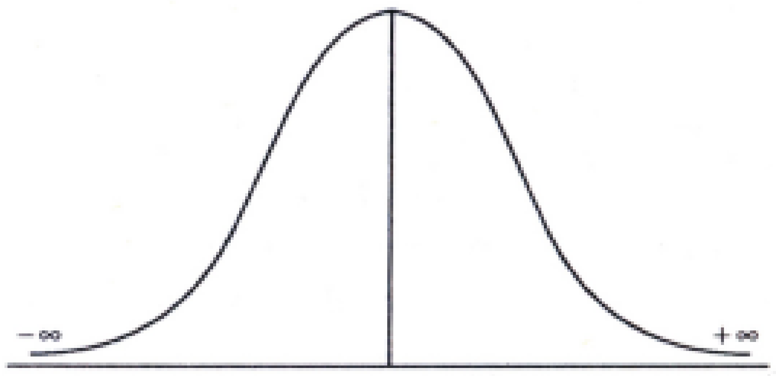

Fig. 1.2.3

5) **The Height of the Curve declines Symmetrically:** In the normal probability curve the height declines symmetrically in either direction from the maximum point.

6) **The Points of Influx occur at point ±1 Standard Deviation (± 1 σ):** The normal curve changes its direction from convex to concave at a point recognised as point of influx. If we draw the perpendiculars from these two points of influx of the curve to the horizontal X axis; touch at a distance one standard deviation unit from above and below the mean (the central point).

7) **The Total Percentage of Area of the Normal Curve within Two Points of Influxation is fixed:** Approximately 68.26% area of the curve lies within the limits of ± 1 standard deviation (± 1 σ) unit from the mean.

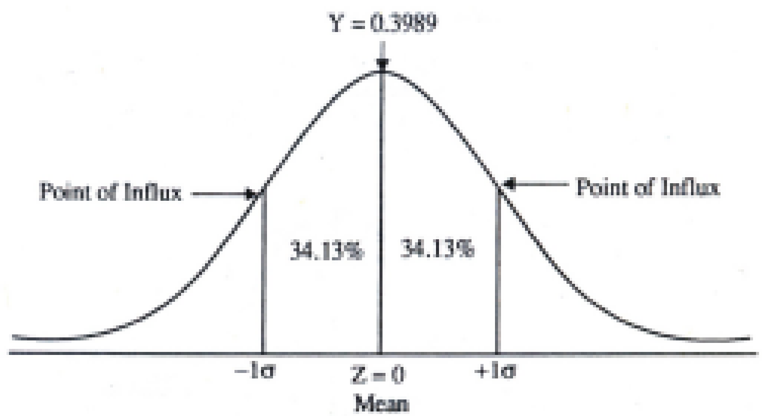

Fig. 1.2.4

8) **The Total Area under Normal Curve may be also considered 100 Percent Probability:** The total area under the normal curve may be considered to approach 100 percent probability; interpreted in terms of standard deviations. The specified area under each unit of standard deviation are shown in this figure.

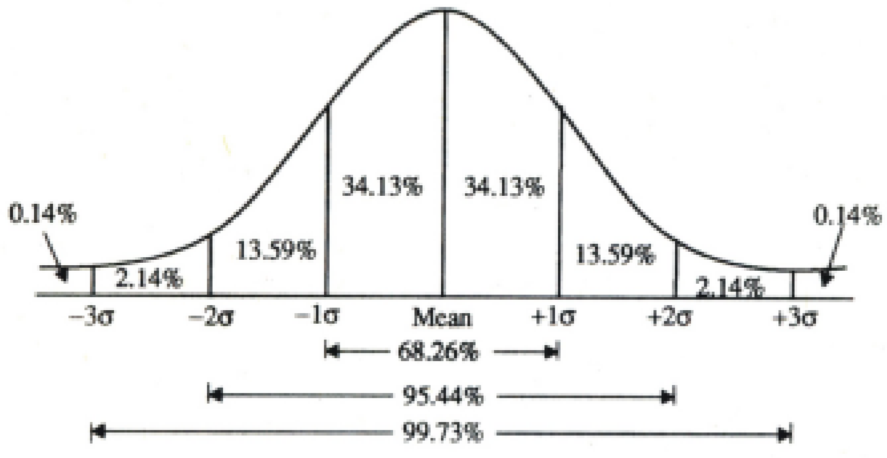

Fig. 1.2.5: The Percentage of the Cases Failing Between Successive Standard Deviation in Normal Distribution

9) **The Normal Curve is Bilateral:** The 50% area of the curve lies to the left side of the maximum central ordinate and 50% of the area lies to the right side. Hence the curve is bilateral.

10) **The Normal Curve is a mathematical model in behavioural Sciences Specially in Mental Measurement:** This curve is used as a measurement scale. The measurement unit of this scale is ± 1σ (the unit standard deviation).

Self Assessment Questions

1) Define a Normal Probability Curve.

....

....

....

....

2) Write the properties of Normal Distribution.

....

....

....

....

3) Mention the conditions under which the frequency distribution can be approximated to the normal distribution.

....

....

....

....

4) In a distribution what percentage of frequencies are lie in between

(a) -1 σ to + 1 σ

(b) -2 σ to + 2 σ

(c) -3 σ to + 3 σ

....

....

....

....

5) Practically, why are the two ends of normal curve considered closed at the points ±3 σ of the base.

....

....

....

....

1.3 INTERPRETATION OF NORMAL CURVE/ NORMAL DISTRIBUTION

What do normal curve/ normal distribution indicate? Normal curve has great significance in the mental measurement and educational evaluation. It gives important information about the trait being measured.

If the frequency polygon of observations or measurements of certain trait is a normal curve, it is a indication that

1) The measured trait is normally distributed in the universe.
2) Most of the cases i.e. individuals are average in the measured trait and their percentage in the total population is about 68.26%.
3) Approximately 15.87% (50-34.13%) of cases are high in the trait measured.
4) Similarly 15.87% of cases are low in the trait measured.
5) The test which is used to measure the trait is good.
6) The test which is used to measure the trait has good discrimination power as it differentiates between poor, average and high ability group individuals.
7) The items of the test used are fairly distributed in terms of difficulty level.

1.4 IMPORTANCE OF NORMAL DISTRIBUTION

The Normal distribution is by far the most used distribution in inferential statistics because of the following reasons:

1) Number of evidences are accumulated to show that normal distribution provides a good fit or describe the frequencies of occurrence of many variable facts in biological statistics, e.g. sex ratio in births, in a country over a number of years. The anthropometrical data, e.g. height, weight, etc. The social and economic data e.g. rate of births, marriages and deaths. In psychological measurements

e.g. Intelligence, perception span, reaction time, adjustment, anxiety etc. In errors of observation in physics, chemistry, astronomy and other physical sciences.

2) The normal distribution is of great value in educational evaluation and educational research, when we make use of mental measurement. It may be noted that normal distribution is not an actual distribution of scores on any test of ability or academic achievement, but is, instead, a mathematical model. The distributions of test scores approach the theoretical normal distribution as a limit, but the fit is rarely ideal and perfect.

1.5 APPLICATIONS/USES OF NORMAL DISTRIBUTION CURVE

There are number of applications of normal curve in the field of psychology as well as educational measurement and evaluation. These are:

i) To determine the percentage of cases (in a normal distribution) within given limits or scores.

ii) To determine the percentage of cases that are above or below a given score or reference point.

iii) To determine the limits of scores which include a given percentage of cases to determine the percentile rank of an individual or a student in his own group.

v) To find out the percentile value of an individual on the basis of his percentile rank.

vi) Dividing a group into sub-groups according to certain ability and assigning the grades.

vii) To compare the two distributions in terms of overlapping.

viii) To determine the relative difficulty of test items.

1.6 TABLE OF AREAS UNDER THE NORMAL PROBABILITY CURVE

How do we use all the above applications of normal curve in mental as well as in educational measurement and evaluation? It is essential first to know about the Table of areas under the normal curve.

The Table 1.6.1 gives the fractional parts of the total area under the normal curve found between the mean and ordinates erected at various σ (sigma) distances from the mean.

The normal probability curve table is generally limited to the areas under unit normal curve with N = 1, σ = 1. In case, when the values of N and σ are different from these, the measurements or scores should be converted into sigma scores (also referred to as standard scores or z scores). The process is as follows :

$$z = \frac{X - M}{\sigma} \quad \text{or} \quad z = \frac{x}{\sigma}$$

In which: z = Standard Score   X = Raw Score

M = Mean of X Scores   σ = Standard Deviation of X Scores

The table of areas of normal probability curve are then referred to find out the proportion of area between the mean and the z value.

Though the total area under the N.P.C. is 1, but for convenience, the total area under the curve is taken to be 10,000 because of the greater ease with which fractional parts of the total area, may be then calculated.

The first column of the table, x/σ gives distance in tenths of σ measured off on the base line for the normal curve from the mean as origin. In the row, the x/σ distance are given to the second place of the decimal.

To find the number of cases in the normal distribution between the mean, and the ordinate erected at a distance of lσ unit from the mean, we go down the x/σ column until 1.0 is reached and in the next column under .00 we take the entry opposite 1.0, namely 3413. This figure means that 3413 cases in 10,000; or 34.13 percent of the entire area of the curve lies between the mean and lσ. Similarly, if we have to find the percentage of the distribution between the mean and 1 .56 σ, say, we go down the x/σ column to 1.5, then across horizontally to the column headed by .06, and note the entry 44.06. This is the percentage of the total area that lies between the mean and 1.56σ.

Table 1.6.1: Fractional parts of the total area (taken as 10,000) under the normal probability curve, corresponding to distance on the baseline between the mean and successive points laid off from the mean in units of standard deviation.

x/σ	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0160	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	1064	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	2457	2389	2422	2454	2486	2517	2549
0.7	2580	2611	2642	2673	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	3290	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	3889	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	4383	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	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	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	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4988	4984	4984	4985	4985	4986	4986
3.0	4986.5	4986.9	4987.4	4987.8	4988.2	4988.6	4988.9	4989.3	4989.7	4990.0
3.1	4990.3	4990.6	4991.0	4991.3	4991.6	4991.8	4992.1	4992.4	4992.6	4992.9
3.2	4993.129									

3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.922									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

Example: Between the mean and a point 1.38 σ $\left(\frac{x}{\sigma} = 1.38\right)$ are found 41.62% of the entire area under the curve.

We have so far considered only σ distances measured in the positive direction from the mean. For this we have taken into account only the right half of the normal curve. Since the curve is symmetrical about the mean, the entries in Table apply to distances measured in the negative direction (to the left) as well as to those measured in the positive direction. If we have to find the percentage of the distribution between mean and -1.28σ, for instance, we take entry 3997 in the column .08, opposite 1.2 in the x/σ column. This entry means that 39.97 percent of the cases in the normal distribution fall between the mean and -1.28σ.

For practical purposes we take the curve to end at points -3σ and +3σ distant from the mean as the normal curve does not actually meet the base line. Table of area under normal probability curve shows that 4986.5 cases lie between mean and ordinate at +3σ. Thus 99.73 percent of the entire distribution, would lie within the limits -3σ and +3σ. The rest 0.27 percent of the distribution beyond ±3σ is considered too small or negligible except where N is very large.

1.7 POINTS TO BE KEPT IN MIND WHILE CONSULTING TABLE OF AREA UNDER NORMAL PROBABILITY CURVE

The following points are to be kept in mind to avoid errors, while consulting the N.P.C. Table.

1) Every given score or observation must be converted into standard measure i.e. Z score, by using the following formula:

$$z = \frac{X - M}{\sigma}$$

2) The mean of the curve is always the reference point, and all the values of areas are given in terms of distances from mean which is zero.

3) The area in terms of proportion can be converted into percentage, and

4) While consulting the table, absolute values of z should be taken. However, a negative value of z shows that the scores and the area lie below the mean and this fact should be kept in mind while doing further calculation on the area. A positive value of z shows that the score lies above the mean *i.e.* right side.

Self Assessment Questions

i) What formula is to use to convert raw score X into standard score i.e. z score.

....

....

ii) What is the reference point on the normal probability curve.

iii) Mean value of the z scores is ____________

iv) The value of standard deviation of z scores is ___________

v) The total area under the N.P.C. is always ___________

vi) The negative value of z scores shows that ___________

vii) The positive value of z scores shows that _________

1.8 PRACTICAL PROBLEMS RELATED TO APPLICATION OF THE NORMAL PROBABILITY CURVE

Under the caption 1.5 you have studied the Application of normal Distribution/ Normal Curve in mental and educational measurements. Now how the practical problems related to these application are solved you go through the following examples carefully and thoughroly.

1) **To determine the percentage of cases in a normal distribution within given limits of scores.**

Often a psychometrician or psychology teacher is interested to know the number of cases or individuals that lie in between two points or two limits. For example, a teacher may be interested as to how many students of his class got marks in between 60% and 70% in the annual examination, or he may be interested in how many students of his got marks above 80%.

Example 1

An adjustment test was administered on a sample of 500 students of class VIII. The mean of the adjustment scores of the total sample obtained was 40 and standard deviation obtained was 8, what percentage of cases lie between the score 36 and 48, if the distribution of adjustment scores is normal in the universe.

Solution:

In the problem it is given that

N = 500

M = 40

$\sigma = 8$

We have to find out the total % of the students who obtained score in between 36 and 48 on the adjustment test.

To find the required percentage of cases, first we have to find out the z scores for the raw scores (X) 36 and 48, by using the formula.

$$z = \frac{X - M}{\sigma}$$

∴ z score for raw score 36 is

$$z_1 = \frac{36-40}{8} =$$

or z_1 = -0.5 σ

Similarly z score for raw score 40 is

$$z_2 = \frac{48-40}{8} =$$

or z_2 = +1 σ

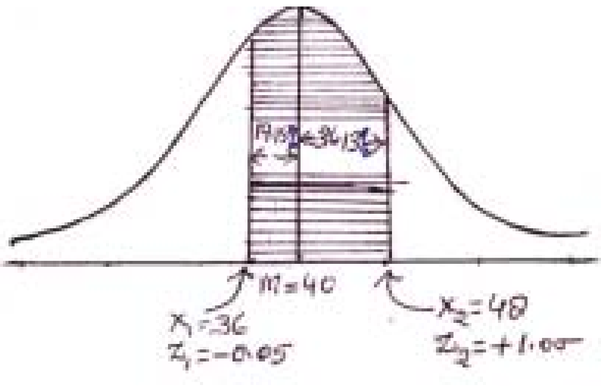

Fig. 1.8.1

According to table of area under Normal Probability curve (N.P.C.) *i.e.* Table No. 1.6.1 the total area of the curve lie in between M to +1σ is 34.13 and in between M to -0.56 is 19.15.

∴ The total area of the curve in between -0.5 σ to +1 σ is 19.15 + 34.13 = 53.28

Thus the total percentage of students who got scores in between 36 and 48 on the adjustment test is 53.28 (Ans.)

Example 2

A reading ability test was administered on the sample of 200 cases studying in IX class. The mean and standard deviation of the reading ability test score was obtained 60 and 10 respectively. Find how many cases lie in between the scores 40 and 70. Assume that reading ability scores are normally distributed.

Solution:

Given N = 200

M = 60

σ = 10

X_1 = 40 and

X_2 = 70

To find out: The total no. of cases in between the two scores 40 and 70.

To find the required no. of cases, first we have to find out the total percentage of cases lie in between Mean and 40 and mean and 70. See the Fig. 1.8.2 For the purpose, first the given raw scores (40 & 70) should be converted into z scores by using the formula

$$z = \frac{X-M}{\sigma}$$

$$\therefore \quad z_1 = \frac{40-60}{10} =$$

or z_1 = - 2σ

Similarly $z_2 = \frac{70-60}{10} =$

or $z_2 = +1\sigma$

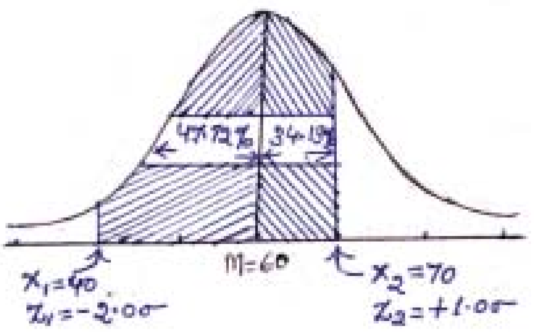

Fig. 1.8.2

According to Table 1.6.1 the area of the curve in between M and -2σ is 47.72% and in between M and +1σ is 34.13%.

∴ The total area of the curve in between -2σ to +1σ is = 47.72 + 34.13 = 81.85%

Therefore, the total no. of cases in between the two scores 40 and 70 are =

$$\frac{81.85 \times 200}{100} = 163.7 \text{ or } 164$$

Thus total no. of cases who got scores in between 40 and 70 are = 164. (Ans.)

2) **To determine the percentage of cases lie above or below a given score or reference point.**

Example 3

An intelligence test was administered on a group of 500 cases of class V. The mean I.Q. of the students was found 100 and the S.D. of the I.Q. scores was 16. Find how many students of class V having the I.Q. below 80 and above 120.

Solution:

Given M = 100, σ = 16, X_1 = 80 and X_2 = 120

To find out : (i) The total no. of cases below 80

(ii) The total no. of cases above 120

To find the required no. of cases first we have to find z scores of the raw scores X_1 = 80 and X_2 = 120 by using the formula

$$z = \frac{X-M}{\sigma}$$

$$z_1 = \frac{80-100}{16} = -\frac{20}{16}$$

or $z_1 = -1.25\ \sigma$

Similarly,

$$z_2 = \frac{120-100}{16} = +\frac{20}{16}$$

or $z_2 = +1.25\ \sigma$

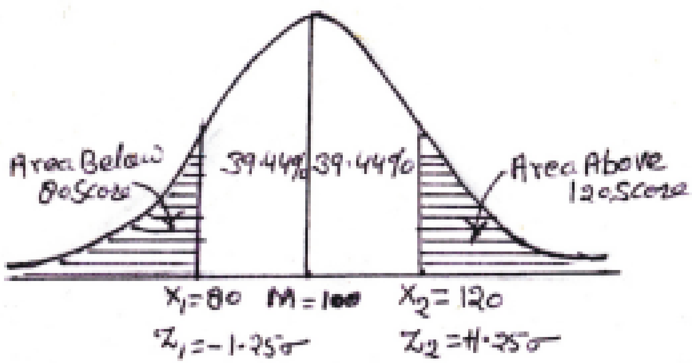

Fig. 1.8.3

According to NPC table (Table 1.6.1) the total percentage of area of the curve lie in between Mean to 1.25 σ is = 39.44.

According to the properties of N.P.C. the 50% area lies below to the mean *i.e.* in left side and 50% area lie above to the mean i.e. in right side.

Thus the total area of NPC curve below M = (100) is = 50 – 39.44 = 10.56

Similarly the total area of NPC curve above M = (100) is = 50 – 39.44 = 10.56

Therefore total cases below to the I.Q. 80 = = 52.8 = 53 Appox.

Similarly Total cases above to the I.Q. 120 = = 52.8 = 53 Appox.

Thus in the group of 500 students of V class there are total 53 students having I.Q. below 80. Similarly there are 53 students who have I.Q. above 120. (Ans.)

3) **To determine the limits of scores which includes a given percentage of cases**

Some time a psychometerician or a teacher is interested to know the limits of the scores in which a specified group of individuals lies. To understand, read the following example-4 and its solution.

Example 4

An achievement test of mathematics was administered on a group of 75 students of class VIII. The value of mean and standard deviation was found 50 and 10 respectively. Find limits of the scores in which middle 60% students lies.

Solution:

Given that, N = 75, M = 50, σ = 10

To find out: Value of the limits of middle 60% cases i.e. X_1 and X_2

As per given condition (middle 60% cases), 30%-30% cases lie left and right to the mean value of the group (see the fig. 1.8.4.)

According to the formula

$$z = \frac{X - M}{\sigma}$$

If the value of M, σ and z is known, the value of X can be find out. In the given problem the value of M and σ are given. We can find out the value of z with the help of the NPC Table No. 1.6.1 as the area of the curve situated right and left to the mean (30%-30% respectively) is also given.

According to the table (1.6.1) the value of z_1 and z_2 of the 30% area is ± 0.84σ

Therefore by using formula

$$z_1 = \frac{X_1 - M}{\sigma}$$

$$-0.84 = \frac{X_1 - 50}{10}$$

or $X_1 = 50 - 0.84 \times 10$

$= 41.60$ or 42

Similarly,

$$z_2 = \frac{X_2 - M}{\sigma}$$

$$-0.84 = \frac{X_2 - 50}{10}$$

or $X_2 = 50 + 0.84 \times 10$

$= 58.4$ or 58

Thus $X_1 = 42$

$X_2 = 58$

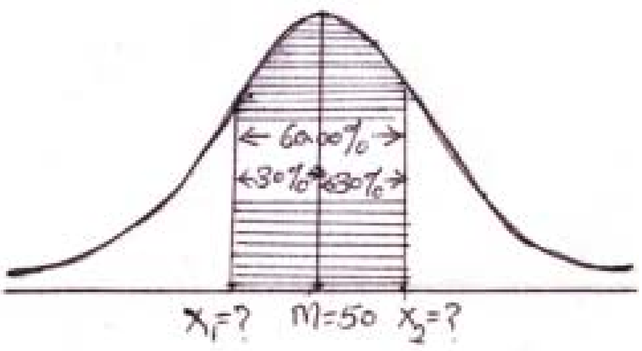

Fig. 1.8.4

Therefore, the middle 60% cases of the entire group (75, students) got marks on achievement test of mathematics in between 42 – 58. Ans.

Self Assessment Questions

The observation given in the example 4, i.e. M = 50 and S.D. (σ) = 10

1) Find the limits of the scores middle 30% cases

....

....

2) Find the limits of the scores middle 75% cases

....

....

3) Find the limits of the scores middle 50% cases

....

....

4) **To determine the percentage rank of the individual in his group.**

The percentile rank is defined as the percentage of cases lie below to a certain score (X) or a point.

Some time a psychologist or a teacher is interested to know the position of an individual or a student in his own group on the bases of the trait is measured (for more clarification go through the following example carefully)

Example 5

In a group of 60 students of class X, Sumit got 75% marks in board examination. If the mean of whole class marks is 50 and S.D. is 10. Find the percentile rank of the Sumit in the class.

Solution:

See the fig. 1.8.5. and pay the attention to the definition of percentile given above carefully.

It is clear from the fig. that we have to find out the total percentage of cases (i.e. the area of N.P.C.) lie below to the point X = 75 (See Fig. 1.8.5.)

To find the total required area (shaded part) of the curve, it is essential first to know the area of the curve lie in between the points 50 and 75.

This area can be determined very easily, by taking up the help of N.P.C. Table, i.e. Table No. 1.6.1., if we know the value of z of score 75.

According to the formula

$$z = \frac{X - M}{\sigma}$$

$$z = \frac{75 - 50}{10} = \frac{25}{10}$$

or $= +2.50\ \sigma$

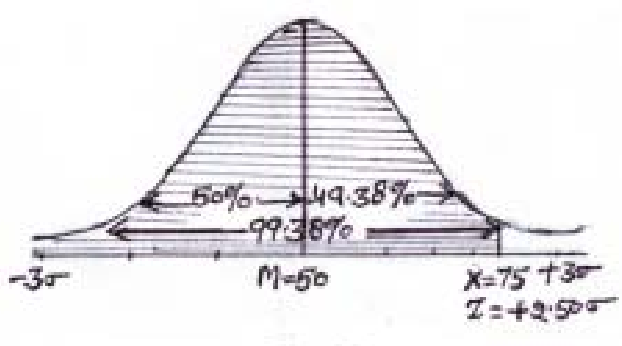

Fig. 1.8.5

According to NPC Table (Table No. 1.6.1) the area of the curve lies M and +2.50 σ is 49.387.

In the present problem we have determined 49.38% area lies right to the mean and 50% area lies to the left of the Mean. (According to the properties of NPC see caption 1.2.4 property no. 9)

Thus according to the definition of percentile the total area of the curve lies below to the point X = 75 is

= 50 + 49.38%

= 99.38% or 99% Approx.

Therefore the percentile rank of the Sumit in the class is 99th. In other words Sumit is the topper student in the class, remaining 99% students lie below to him. (Ans.)

Self Assessment Questions

In a test of 200 items, each correct item has 1 mark.

If M = 100, σ = 10

1) Find the position of Rohit in the group who secured 85 marks on the test.

   ....

   ....

2) Find the percentile rank of Sunita she got 130 marks on the test.

....

....

5) **To find out the percentile value of an individual's percentile rank.**

Some time we are interested to know that the person or an individual having a specific percentile rank in the group, than what is the percentage of score he got on the test paper. To understand, go through the following example and its solution –

Example 6

An intelligence test was administered on a large group of student of class VIII. The mean and standard deviation of the scores was obtained 65 and 15 respectively. On the basis intelligence test if the Ramesh's percentile rank in the class is 80, find what is the score of the Ramesh, he got on the test?

Solution:

Given : M = 65, σ = 15, and PR = 80

To find out : The value of P_{80}

Look at the Fig. No. 1.8.6., as per definition of percentile rank, the 30% area of the curve lie from mean to the point P_{80} and 50% are lie to the left side of the mean.

The z value of the 30% area of the curve lie in between M and P_{80} is = +0.85 σ

(Table No. 1.16)

We know that $z = \dfrac{X - M}{\sigma}$

or $+ 0.85 = \dfrac{X - 65}{15}$

or $X = 65 + 15 \times 0.85$

$= 65 + 12.75$

$= 77.75$ or 78 Approx.

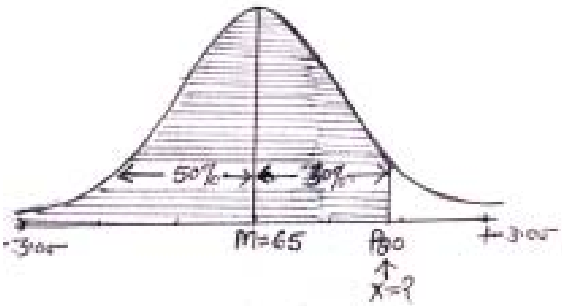

Fig. 1.8.6

Thus Ramesh's intelligence score on the test is = 78 (Ans.)

Self Assessment Questions

1) If M = 100, σ = 10

Find the values of

i) P_{75} = ____________________

ii) P_{10} = ____________________

iii) P_{50} = ____________________

iv) P_{80} = ____________________

6) **Dividing a group of individuals into sub-group according to the level of ability or a certain trait. If the trait or ability is normally distributed in the universe.**

Some time we are making qualitative evaluation of the person or an individual on the basis of trait or ability, and assign them grades like A, B, C, D, E etc. or 1st grade 2nd grade, 3rd grade etc. or High, Average or Low. For example a company evaluate their salesman as A grade, B grade and C grade salesman. A teacher provides A, B, C etc. grades to his students on the basis of their performance in the examination. A psychologist may classify a group of person on the basis of their adjustment as highly adjusted, Average and poorly adjusted. In such conditions, always there is a question that how many persons or individuals, we have to provide A, B, C, D and E etc. grades to the individuals and categorize them in different groups.

For further clarification go through the following examples:

Example 7

A company wants to classify the group of salesman into four categories as Excellent, Good, Average and Poor on the basis of the sale of a product of the company, to provide incentive to them. If the number of salesman in the company is 100, their average sale of the product per week is 10,00,000 Rs. and standard deviation is Rs. 500/-. Find the number of salesman to place as Excellent, Good, Average and Poor.

Solution:

As per property of the N.P.C. we know that total area of the curve is 6σ over a range of -3σ to +3σ.

According to the problem, the total area of the curve is divided into four categories.

Therefore area of each category is 6σ/4 = ± 1.5σ. It means the distance of each category from the mean on the curve is 1.5σ respectively.

The distance of each category is shown in the figure 1.8.7

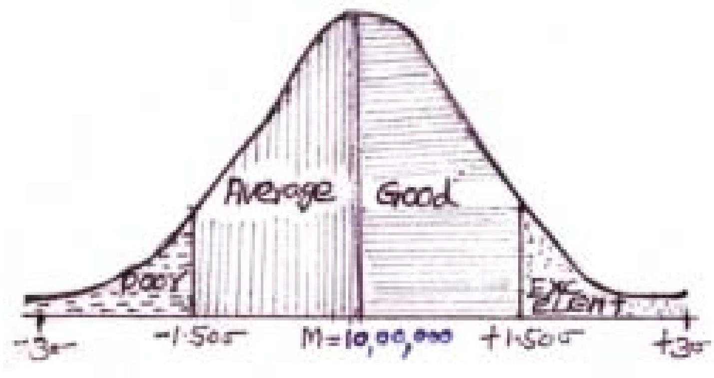

Fig. 1.8.7

i) Total % of salesman in "Good" category

According to N.P.C. Table (Table No. 1.6.1), the

Total area of the curve lies in between M an +1.5σ is = 43.32%

∴ The total % of salesman in "Good" category is 43.32%

ii) Total % of salesman in "Average" category

Total area of the curve lies in between Mean and – 1.5σ is also = 43.32%

∴ The total % of salesman in Average category is = 43.32

iii) Total % of salesman in "Excellent" category

The total area of the curve from M to +3σ and above is

= 50% .... (As per properties of Normal Curve)

∴ The total % of salesman in the category Excellent is = 50 – 43.32 = 6.68%

iv) Total % of salesman in "Poor" category

The total area of the curve from M to -3σ and below is

= 50% .... (As per properties of Normal Curve)

∴ The total % of the salesman in the poor category is = 50 – 43.32 = 6.68%

Thus,

i) The number of salesman should place in "Excellent" category

$$= \frac{6.68 \times 100}{100} = 6.68 \text{ or } 7$$

ii) The number of salesman should place in "Good" category

$$= \frac{43.32 \times 100}{100} = 43.32 \text{ or } 43$$

iii) The number of salesman should place in "Average" category

$$= \frac{43.32 \times 100}{100} = 43.32 \text{ or } 43$$

iv) The number of salesman should place in "Poor" category

$$= \frac{6.68 \times 100}{100} = 6.68 \text{ or } 7$$

Total = 100 (Ans.)

Self Assessment Questions

In the above example no. 7 if the salesman are categorised into six categories as excellent, v. good, good average, poor and v. poor. Find the number of salesman in each category as per their sales ability.

....

....

Example 8

A group of 1000 applicant's who wishes to take admission in a psychology course. The selection committee decided to classify the entire group into five sub-categories A, B, C, D and E according to their academic ability of last qualifying examination.

If the range of ability being equal in each sub category, calculate the number of applicants that can be placed in groups ABCD and E.

Solution:

Given: N = 1000

To find out: The 1000 cases to be categorised into five categories A, B, C, D, and E.

We know that the base line of a normal distribution curve is considered extend from -3σ to +3σ that is range of 6σ.

Dividing this range by 5 (the five subgroups) to obtain σ distance of each category, i.e. the z value of the cutting point of each category (see the fig. given below)

$$\therefore\ z = \frac{6\sigma}{5} = \pm\ 1.20\ \sigma$$

(It is to be noted here that the entire group of 1000 cases is divided into five categories. The number of subgroups is odd number. In such condition the middle group or middle category (c) will lie equally to the centre i.e. M of the distribution of scores. In other words the number of cases of "c" category or middle category remain half to the left area of the curve from the point of mean and half of the right area of the curve from the mean.

$$\therefore \quad \text{the limits of "c" category is} = \frac{1.2\sigma}{2} = \pm\ 0.60\ \sigma$$

i.e. the "c" category will remain on NPC curve in between the two limits -0.6 σ to +0.6 σ

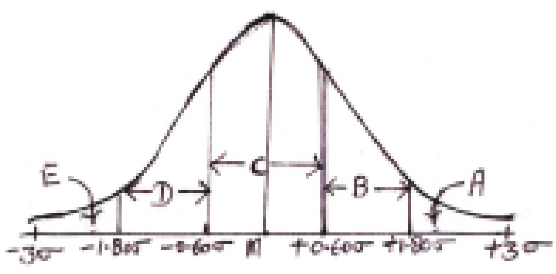

Fig. 1.8.8

Now,

The limits of B category

Lower limit = +0.6 σ

and Upper limit = 0.60 σ + 1.20 σ

or = +1.80 σ

The limits of A category

Lower limit = + 1.8 σ

and Upper limit = + 3 σ and above

Similarly, the limits of D category

Upper limit = - 0.6 σ

Lower limit = (- 0.60 σ) + (-1.20 σ)

or = -1.80 σ

The limits of E category

Upper limit = -1.8 σ

Lower limit = -3σ and below

(For limits of each category see the fig. 1.8.8 carefully)

i) The total % area of the NPC for A category

According to NPC Table (1.6.1) the total % of area in between

Mean to +1.80 σ is = 46.41

∴ The total % of area of the NPC for A category is = 50 – 46.41 = 3.59

ii) The total % Area of the NPC for B category –

According to NPC Table (1.6.1) the total % of Area in between

Mean and + 0.60 σ is = 22.57

∴ The total % area of NPC for B category is = 46.41 – 22.57 = 23.84

iii) The total % area of the NPC for C category –

According to NPC table the total % area of NPC in between

M and + 0.06 σ is = 22.57

Similarly the total % area of NPC in between

M and – 0.06 σ is also = 22.57

∴ The total % area of NPC for C category is = 22.57 + 22.57 = 45.14

iv) In similar way the total % area of NPC for D category is = 23.84

v) The total % area of NPC for E category is = 3.59

Thus the total number of applicants (N = 1000) in –

A category is = $\frac{3.59 \times 1000}{100}$ = 35.9 = 36

B category is = $\frac{23.84 \times 1000}{100}$ = 238.4 = 238

C category is = $\frac{45.14 \times 1000}{100}$ = 451.4 = 452

D category is = $\frac{23.84 \times 1000}{100}$ = 238.4 = 238

E category is = $\frac{3.59 \times 1000}{100}$ = 35.9 = 36

Total= 1000 (Ans.)

Self Assessment Questions

1) In the example 8 if the total applicants are categorised into three categories. Find how many applicants will be the categories A, B and C?

....

....

7) **To compare the two distributions in terms of overlapping.**

Example 9

A numerical ability test was administered on 300 graduate boys and 200 graduate girls. The boys Mean score is 26 with S.D. (σ) of 4. The girls' mean. Mean score is 28 with a σ 8. Find the total number of boys who exceed the mean of the girls and total number of girls who got score below to the mean of boys.

Solution:

Given: For Boys, N = 300, M = 26 and σ = 6

For Girls, N = 200, M = 28 and σ = 8

To find: 1- Number of boys who exceed the mean of girls

2- Number of girls who scored below to the mean of boys

As per given conditions, first we have to find the number of cases above the point 28

(The mean of the numerical ability scores of girls) by considering M=26 and σ=6

Second, we to find no. of cases below to the point 26 (The mean score of the boys), by considering M = 28 and 5 – 8 (see the fig. 1.8.9 given below carefully)

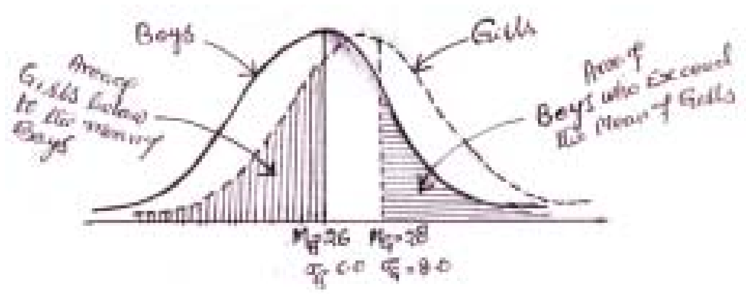

Fig. 1.8.9

1) The z score of X (28) is = $\frac{28-26}{6} = \frac{2}{6}$

or = + 0.33 σ

According to NPC Table (1.6.1) the total % of area of the NPC from M = 26 to + 0.33 σ is = 12.93

∴ The total % of cases above to the point 28 is = 50 – 12.93 = 37.07

Thus the total number of boys above to the point 28 (mean of the girls) is

$= \frac{37.07 \times 300}{100} = 111.21 = 111$

2) The z score of X = 26 is = $\frac{26-28}{8} = \frac{2}{8}$ = - 0.25 σ

According to the NPC table the total % of area of the curve in between M = 28 and -0.25 σ is = 9.87

∴ Total % of cases below to the point 26 is = 50 – 9.87 = 40.13

Thus the total number of girls below to the point 26 (mean of the boys) is

$$= \frac{40.13 \times 200}{100} = 80.26 = 80$$

Therefore,

1) The total number of boys who exceed the mean of the girls in numerical ability is = 111

2) The total number of girls who are below to the mean of the boys is = 88 (Ans.)

Self Assessment Questions

1) In the example given above (Example 9) find.

   i) Number of boys between the two means 26 and 28 __________

   ii) Number of girls between the two means 26 and 28 __________

   iii) Number of boys below to the mean of girls __________

   iv) Number of girls above to the mean of boys __________

   v) Number of boys above to the Md of girls which is 28.20 __________

   vi) Number of girls exceed to the Md of the boys which is 26.20 __________

8) **To determine the relative difficulty of a test items:**

Example 10

In a mathematics achievement test ment for 10th standard class, Q.No. 1, 2 and 3 are solved by the students 60%, 30% and 10% respectively find the relative difficulty level of each Q. Assume that solving capacity of the students is normally distributed in the universe.

Given: The percentage of the students who are solving the test items (Qs) of a question paper correctly.

To Find: The relative difficulty level of each item of the test paper given.

Solution:

First of all we shall mark the relative position of test items on the basis of percentage of students solving the items successfully on the NPC scale.

Q.No.3 of the test paper is correctly solved by the 10% students only. It means 90% students unable to attend the Q.No. 3. On the NPC scale, these 10% cases lies extreme to the right side of the mean (see the fig. given below). Similarly 30% students who are solving Q.No. 2 correctly also lying to the right side of the curve. While the 60% students who are solving Q.No. 1 correctly are lying left side of the N.P.C. curve.

Now, we have to find out the z value of the cut point of the each item (Q.No.) on the NPC base line

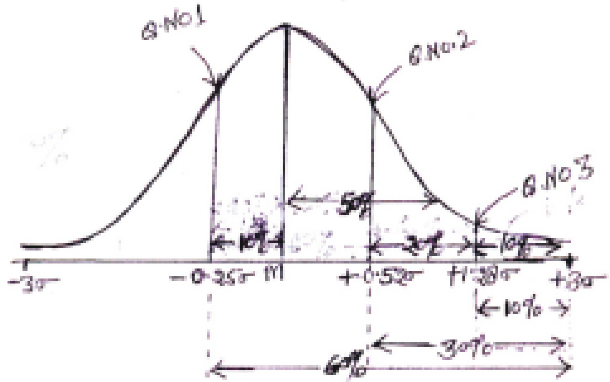

Fig. 1.8.10

i) The z value of the cut point of Q.No. 3

The total percentage of cases lie in between mean and cut point of Q.No. 3 is = (50% - 10%) in right half of NPC

∴ The z value of the right 40% of area of the NPC is = 11.28 σ

ii) The z value of the cut point of Q.No.2

The total percentage of cases lie between the mean and cut point of Q.No. 2 is = 20% (50% - 30%) in right half of NPC

∴ The z value of the right 20% area of the NPC is = + 0.52 σ

iii) The z value of the cut point of Q.No. 3

The total percentage of cases lie between the mean and cut point of Q.No. 3 is = (60% - 50%) in left half of NPC

∴ The z value of the left of 10% of area = - 0.25 σ

Therefore corresponding z value of each item (Q) passed by the students is

Item (Q.No.)	Passed By	z value	Z difference
3	10%	+ 1.28 σ	-
2	30%	+ 0.52 σ	0.76 σ
1	60%	- 0.25 σ	0.77 σ

We may now compare the three questions of the mathematics achievement test, Q.No. 1 has a difficulty value of 0.76 σ higher than the Q.No. 2. Similarly the Q.No. 2 has a difficulty value of 0.77 σ higher than the Q.No. 3. Thus the Q.No. 1, 2 and 3 of the mathematics achievement test are the good items having equal level of difficulty and are quite discriminative. (Ans.)

Self Assessment Question

1) The three test items 1, 2 and 3 of an ability test are solved by 10%, 20% and 30% respectively. What are the relative difficulty values of these items?

....

....

....

1.9 DIVERGENCE IN NORMALITY (THE NON-NORMAL DISTRIBUTION)

In a frequency polygon or histogram of test scores, usually the first thing that strikes one is the symmetry or lack of it in the shape of the curve. In the normal curve model, the mean, the median and the mode all coincide and there is perfect balance between the right and left halves of the curve. Generally two types of divergence occur in the normal curve.

1) Skewness

2) Kurtosis

1) **Skewnes:** A distribution is said to be "skewed" when the mean and median fall at different points in the distribution and the balance *i.e.* the point of center of gravity is shifted to one side or the other to left or right. In a normal distribution the mean equals, the median exactly and the skewness is of course zero ($S_K = 0$).

   There are two types of skewness which appear in the normal curve.

   a) **Negative Skewness :** Distribution said to be skewed negatively or to the left when scores are massed at the high end of the scale, *i.e.* the right side of the curve are spread out more gradually toward the low end *i.e.* the left side of the curve. In negatively skewed distribution the value of median will be higher than that of the value of the mean.

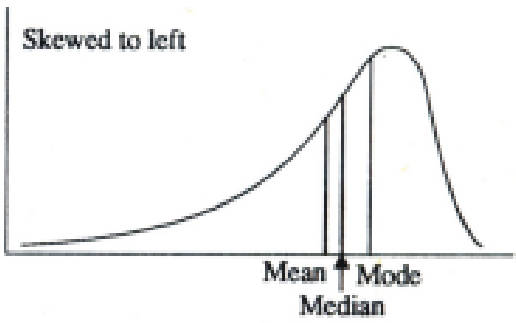

Fig. 1.9.1: Negative Skewness

   b) **Positive Skewness:** Distributions are skewed positively or to the right, when scores are massed at the low; *i.e.* the left end of the scale, and are spread out gradually toward the high or right end as shown in the fig.

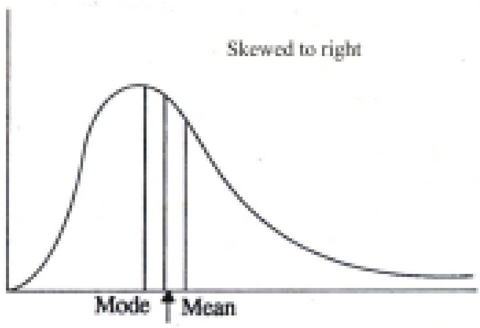

Fig. 1.9.2: Negative Skewness

2) **Kurtosis:** The term kurtosis refers to (the divergence) in the height of the curve, specially in the peakness. There are two types of divergence in the peakness of the curve

   a) **Leptokurtosis:** Suppose you have a normal curve which is made up of a steel wire. If you push both the ends of the wire curve together. What would happen in the shape of the curve? Probably your answer may be that by pressing both the ends of the wire curve, the curve become more peeked *i.e.* its top become more narrow than the normal curve and scatterdness in the scores or area of the curve shrink towards the center.

Thus in a Leptokurtic distribution, the frequency distribution curve is more peaked than to the normal distribution curve.

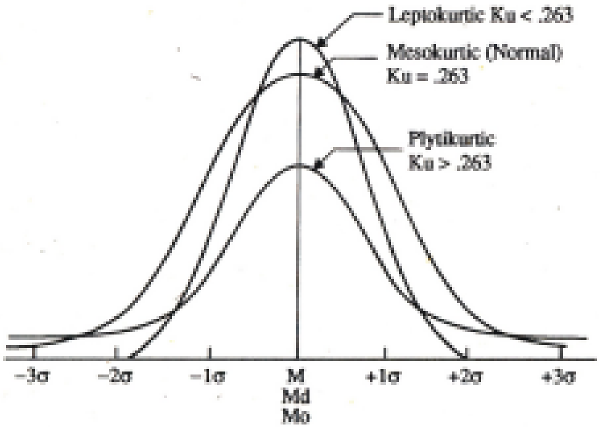

Fig. 1.9.3: Kurtosis in the Normal Curve

   b) **Platykurtosis:** Now suppose we put a heavy pressure on the top of the wire made normal curve. What would be the change in the, shape of the curve? Probably you may say that the top of the curve become more flat than to the normal.

Thus a distribution of flatter Peak than to the normal is known Platykurtosis distribution.

When the distribution and related curve is normal, the vain of kurtosis is 0.263 (KU = 0.263). If the value of the KU is greater than 0.263, the distribution and related curve obtained will be platykurtic. When the value of KU is less than 0.263, the distribution and related curve obtained will be Leptokurtic.

1.10 FACTORS CAUSING DIVERGENCE IN THE NORMAL DISTRIBUTION /NORMAL CURVE

The reasons on why distribution exhibit skewness and kurtosis are numerous and often complex, but a careful analysis of the data will often permit the common causes of asymmetry. Some of common causes are –

1) **Selection of the Sample:** Selection of the subjects (individuals) produce skewness and kurtosis in the distribution. If the sample size is small or sample is biased one, skewness is possible in the distribution of scores obtained on the basis of selected sample or group of individuals.

If the scores made by small and homogeneous group are likely to yield narrow and leptokurtic distribution. Scores from small and highly hetrogeneous groups yield platykurtic distribution.

2) **Unsuitable or Poorly Made Tests:** If the measuring tool or test is unappropriate, or poorly made, the asymmetry is possible in the distribution of scores. If a test is too easy, scores will pile up at the high end of the scale, whereas the test is too hard, scores will pile up at the low end of the scale.

3) **The Trait being Measured is Non-Normal:** Skewness or Kurtosis or both will appear when there is a real lack of normality in the trait being measured, e.g. interest, attitude, suggestibility, deaths in a old age or early childhood due to certain degenerative diseases etc.

4) **Errors in the Construction and Administration of Tests:** The unstandardised with poor item-analysis test may cause asymmetry in the distribution of the scores. Similarly, while administrating the test, the unclear instructions – Error in timings, Errors in the scoring, practice and motivation to complete the test all the these factors may cause skewness in the distribution.

Self Assessment Questions

1) Define the following:

a) Skewness

....

....

b) Negative and Positive Skewness

....

....

c) Kurtosis

....

....

d) Platykurtosis

....

....

e) Leptokurtosis

....

....

2) In case of normal distribution what should be the value of skewness.

....

....

....

....

3) In case of normal distribution what should be the value of Kurtosis.

....

....

....

....

4) What is the significance of the knowledge of skewness and kurtosis to a school teacher?

....

....

....

....

1.11 MEASURING DIVERGENCE IN THE NORMAL DISTRIBUTION / NORMAL CURVE

In psychology and education the divergence in normal distribution/normal curve have a significant role in construction of the ability and mental tests and to test the representativeness of a sample taken from a large population. Further the divergence in the distribution of scores or measurements obtained of a certain population reflects some important information about the trait of population measured. Thus there is a need to measure the two divergence i.e. skewness and kurtosis of the distribution of the scores.

1.11.1 Measuring Skewness

There are two methods to study the skewness in a distribution.

i) Observation Method

ii) Statistical Method

i) **Observation Method:** There is a simple method of detecting the directions of skewness by the inspection of frequency polygon prepared on the basis of the scores obtained regarding a trait of the population or a sample drawn from a population.

Looking at the tails of the frequency polygon of the distribution obtained, if longer tail of the curve is towards the higher value or upper side or right side to the centre or mean, the skewness is positive. If the longer tail is towards the lower values or lower side or left to the mean, the skewness is negative.

ii) **Statistical Method:** To know the skewness in the distribution we may also use the statistical method. For the purpose we use measures of central tendency, specifically mean and median values and use the following formula.

$$S_K = \frac{3(Mean - Median)}{\sigma}$$

Another measure of skewness based on percentile values is as under

$$S_K = \frac{(P_{90} - P_{10})}{2} - P_{50}$$

Here, it is to be kept in mind that the above two measures are not mathematically equivalent. A normal curve has the value of $S_K = 0$. Deviations from normality can be negative and positively direction leading to negatively skewed and positively skewed distributions respectively.

1.11.2 Measuring Kurtosis

For judging whether a distribution lacks normal symmetry or peakedness; it may detected by inspection of the frequency polygon obtained. If a peak of curve is thin and sides are narrow to the centre, the distribution is leptokurtic and if the peak of the frequency distribution is too flat and sides of the curve are deviating from the centre towards $\pm 4\sigma$ or $\pm 5\sigma$ than the distribution is platikurtic.

Kurtosis can be measured by following formula using percentile values.

$$K_U = \frac{Q}{P_{90} - P_{10}}$$

where Q = quartile deviation i.e.

P_{10} = 10th percentile

P_{90} = 90th percentile

A normal distribution has $K_U = 0.263$. If the value of K_U is less than 0.263 ($K_U <$ 0.263), the distribution is leptokurtic and if K_U is greater than 0.263 ($K_U > 0.263$), the distribution is platykurtic.

Self Assessment Questions

1) How we can instantly study the skewness in a distribution.

....

....

....

....

2) What is the formula to measure skewness in a distribution?

....

....

....

....

3) What indicates the kurtosis of a distribution?

....

....

....

....

4) What formula is used to calculate the value of kurtosis in a distribution?

....

....

....

....

5) How we decide that a distribution is leptokurtic or platykurtic?

....

....

....

....

1.12 LET US SUM UP

The normal distribution is a very important concept in the behavioural sciences because many variables used in behavioural research are assumed to be normally distributed.

In behavioural science each variable has a specific mean and standard deviation, there is a family of normal distribution rather than just a single distribution. However, if we know the mean and standard deviation for any normal distribution we can transform it into the standard normal distribution. The standard normal distribution is the normal distribution in standard score (z) form with mean equal to 0 and standard deviation equal to 1.

Normal curve is much helpful in psychological and educational measurement and educational evaluation. It provides relative positioning of the individual in a group. It can be also used as a scale of measurement in behavioural sciences.

The normal distribution is a significant tool in the hands of teacher and researcher of psychology and education. Through which he can decide the nature of the distribution of the scores obtained on the basis of measured variable. Also he can decide about his own scoring process which is very lenient or hard; he can Judge the difficulty level of the test items in the question paper and finally he may know about his class, whether it is homogeneous to the ability measured or it is hetrogeneous one.

1.13 UNIT END QUESTIONS

1) Take some frequency distributions and prepare the frequency polygons. Study the normalcy in the distribution. If you will obtained non-normal distribution, determine the type of skewness and kurtosis. Also list down the probable causes associated to the non-normal distribution.

2) Collect the annual examination marks of various subjects of any class and study the nature of distribution of scores of each subject. Also determine the difficulty level of the question papers of each subject.

3) Determine which variables related to cognitive and affective domain of behaviour are normally distributed.

4) As a psychological test constructor or teacher, what precautions are to be considered, while preparing a question paper or test paper.

1.14 SUGGESTED READINGS

Aggarwal, Y.P.: *"Statistical Methods-Concepts, Applications and Computation"*. New Delhi: Sterling Publishers Pvt. Ltd.

Ferguson, G.A.: *"Statistical Analysis in Psychology and Education"*. New York: McGraw Hill Co.

Garrett, H.E. & Woodworth, R.S.: *"Statistics in Psychology and Education"*. Bombay : Vakils, Feffer & Simons Pvt. Ltd.

Guilford, J.P. & Benjamin, F.: *"Fundamental Statistics in Psychology and Education"*. New York: McGraw Hill Co.

Srivastava, A.B.L. & Sharma, K.K.: *"Elementary Statistics in Psychology and Education"*, New Delhi: Sterling Publishers Pvt. Ltd.