UNIT 17 NATURE OF DISTRIBUTION AND ITS INTERPRETATION

Structure

17.1 INTRODUCTION

By now you are well aware of the distribution of scores. In the previous units, you have also learnt to illustrate the shape of a frequency distribution through different kinds of graphs like histogram and frequency polygons. In histogram or frequency polygon all the scores of a distribution have some place or space. The central tendency helps you describe the central value of this distribution and the measure of variability helps you indicate the extent of variation in the distribution. These distributions or graphs help you in getting information about the set of scores or position of scores in the set. These graphs are varied from distribution to distribution.

But many a times, we need to categorise the group of individuals on certain measure or test or the trait. We may like to know the procedures for describing individual's position within a group or a set of scores on that trait. It may be distribution of the students on the basis of ability, achievement, and intelligence through a set of frequency distribution.

For Example: Suppose you had administered an achievement test in your subject to ascertain the level of achievement of your students and a student had got some marks (score) on that test. What does this score mean? Any obtained score has a meaning only with respect to other scores of that distribution or the scores of some other distribution on similar criteria. If you see any distribution that is occurring or happening without any effect from outside, it will have a definite trend or shape in graphical term. If we take any variable or any event of the nature whether educational, psychological or otherwise, it is occurring on a definite pattern and follows a *Natural Probability Theory* in its occurrence. This pattern in graphical terms is known as Normal Probability Curve, which is bell shaped. This curve helps us understand the meaning of any score with respect to other scores or the position of an individual or even helps us in doing categorisation of a group.

The present unit presents the concept and use of Normal Distribution in relation to the educational variables through suitable illustrations and explanations.

17.2 OBJECTIVES

After going through this unit, you will be able to:

- explain the concept of Normal Distribution and Normal Probability Curve;
- recall the theoretical base of the Normal Probability Curve;
- explain the need of a Normal Distribution;
- write the properties of Normal Probability Curve;
- draw a curve of any kind of distribution;
- recognize the various divergence in the Normal Curve;
- recall the definitions of various divergence of Normal Probability Curve;
- justify the significance of Skewness and Kurtosis in the educational measurement and evaluation;
- interpret the Normal Curve obtained on the basis of large number of observations;
- discuss the importance of Normal Curve in educational measurement;
- recall the various applications of Normal Curve in educational measurement and evaluation;
- find the number of cases in any one sector of a normal distribution; and
- apply the knowledge of Normal Probability Curve in solving various practical problems related to educational measurement and evaluation.

17.3 NORMAL DISTRIBUTION/NORMAL PROBABILITY CURVE

A normal distribution is also called as a normal probability curve. Let us try to understand first the concept of a normal distribution.

17.3.1 The Concept of Normal Distribution

Although it is not possible to predict the exact nature of occurrence of any distribution, still mathematics and statistics, with the help of well established theories, have applied certain methods of calculating the numerical values and predict the kind of distribution that is expected in general. This is made possible with the help of theory of probability.

If we measure any variable for a large number of times, we will find that they observe a definite pattern or symmetry. On plotting a graph of these measurements we will find the distribution to be symmetrical. Such distributions which are perfectly symmetrical and follow the laws of nature are called Normal Distributions i.e. in which the left half is a mirror image of the right half, as well as, mean, median and (if the distribution is uni-modal) the mode have the same value. Such normal distributions are widely occurring in nature. However, equality of mean and median does not guarantee that the distribution is symmetrical, although it is not likely to depart very far from that condition. On the other hand, if mean and median are different or have different values, the distribution cannot be normal.

Normal distribution is important because many of the educational and psychological variables are distributed close to normality. Measures of intelligence, reading ability, introversion and extroversion, job satisfaction, other personality traits or memory are among some of the psychological variables distributed approximately normally. Although these distributions are only approximately normal, they are quite close to normality. Many kinds of statistical tests can be derived for normal distributions. These tests work very well even if the distribution is approximately normally distributed. If the normality of the distribution is known, it is easy to convert back and forth from raw scores to the distribution and distribution to the raw scores.

Carefully look at the following hypothetical frequency distribution, which a teacher has obtained after examining 150 students of class X on a mathematics achievement test (out of 150) (see Table 17.1).

Table 17.1: Frequency Distribution of the Mathematics Achievement Test Scores

Class Intervals	Tallies	Frequency
115–119	I	1
110–114	II	2
105–109	IIII	4
100–104	~~IIII~~ II	7
95–99	~~IIII~~ ~~IIII~~	10
90–94	~~IIII~~ ~~IIII~~ ~~IIII~~ I	16
85–89	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	20
80–84	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	30
75–79	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	20
70–74	~~IIII~~ ~~IIII~~ ~~IIII~~ I	16

65–69	~~IIII~~ ~~IIII~~	10
60–64	~~IIII~~ II	7
55–59	IIII	4
50–54	II	2
45–49	I	1
	Total	**150**

Can you see some special trend in the frequencies shown in the column 3 of the above table? Probably Yes! The concentration of maximum frequency (*f*=30) is at the central value of distribution and frequencies gradually taper off symmetrically on both the sides of this value.

If we draw a frequency polygon with the help of the above distribution, we will have a curve as shown in the Fig. 17.1

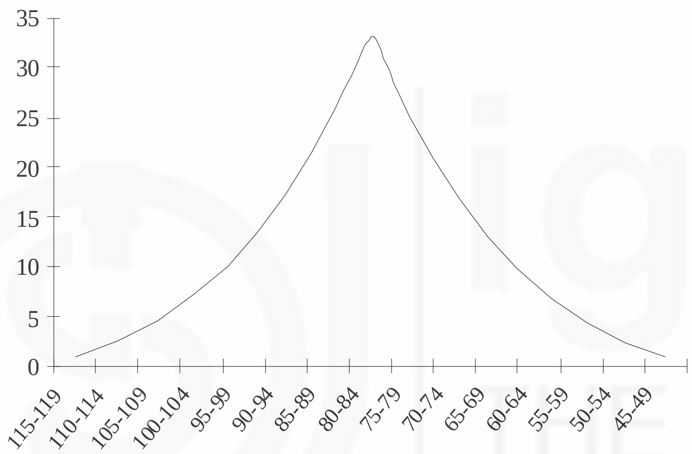

Figure 17.1: Frequency Polygon of the data given in Table 17.1

The shape of the curve in Fig. 17.1 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 = Md = Mo = 82).

This 'Bell' shaped curve is technically known as Normal Probability Curve or simply Normal Curve and the corresponding frequency distribution of scores, having equal values of all three measures of central tendency, is known as Normal Distribution. This normal curve has great significance in cognitive and educational measurement. In measurement of behavioural aspects, the normal probability curve has often been used as reference curve.

17.3.2 The Normal Probability Curve: Its Theoretical Base

The normal probability curve is based upon the law of probability (the games of chance) discovered by French Mathematician Abraham Demoivre (1667-1754) in the eighteenth century. He developed its mathematical equation and graphical representation also. You must have studied Binomial Distribution in mathematics in your earlier classes at the school or college level. Normal distribution is a limiting form of the Binomial distribution in which neither of the two variables (chances) p and q is very small and number of trials is very large to make it random.

17.3.3 Properties of Normal Probability Curve

By now you must have understood the meaning of normal distribution from the above discussion that it is a peculiar distribution and has its own specific properties, which make it so important. It is more important for you to understand its important properties as they will be useful for you in times to come. In this section, we will be discussing these important properties or characteristics which make it so distinct. Some of the properties of a normal distribution are:

1. ***The Normal Curve is Symmetrical:*** The normal distribution curve is a symmetrical curve around the mean i.e., if we draw a curve with mean as a variant, then the number of cases above the mean value and the number of cases below the mean value will be equal. It is not only equal number of cases but even their distribution on either side i.e., below and above the mean value will be equal. The curve will be symmetrical around the vertical axis. If we draw a curve for a number of trials or various cases of a distribution means as the vertical axis, it will look alike on both sides of the axis.

It 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 values to the middle central point are mirror images, as shown in Figure 17.2 below.

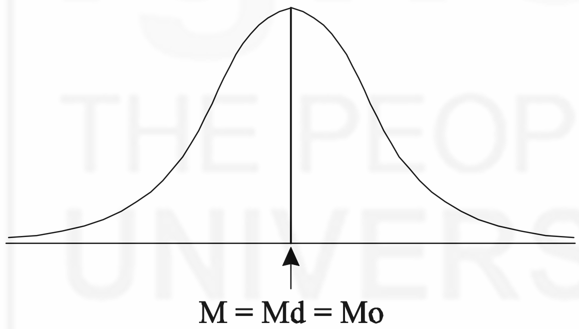

Figure 17.2: Normal Distribution Curve

2. When we draw a normal probability curve for a distribution, it is always a bell shaped curve irrespective of the nature of the variable for which it is being drawn. We may take any variable and see its occurrence in large number of cases; we will get the same kind of distribution. Hence the curve drawn for any variable will be a normal distribution curve or a bell shaped curve. e.g. if we take a coin which is symmetrical or unbiased and toss it a large number of times and we represent these observations on a graph; it shows a bell shaped curve. Similarly, if we throw an unbiased dice showing a three – sport and get a large number of observations. On plotting these scatter diagram points on a graph, we will get the same bell shaped curve because the distribution in this case also tends to be normal.

   Such a bell shaped curve is observed whenever we plot any normal distribution graphically for any variable. Even by seeing a graph of a distribution as bell shaped it could easily be said that the distribution is a normal distribution.

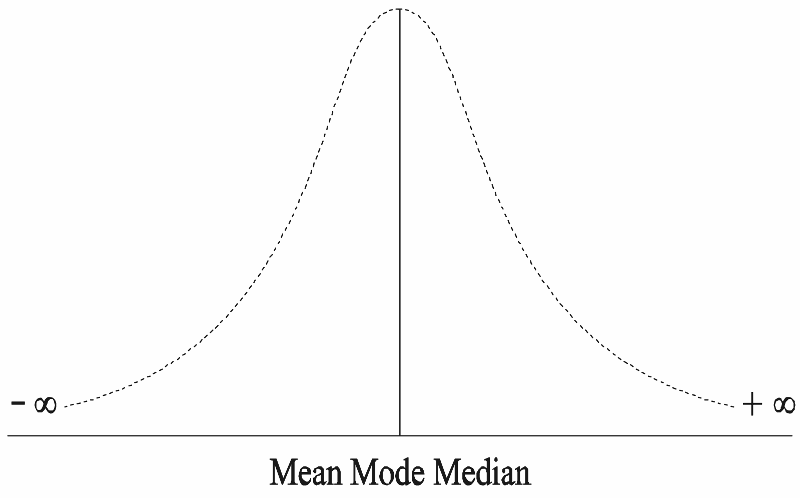

Figure 17.3: Normal Distribution Curve

3. Size, shape, and slope of the curve on one side of the curve is identical to the size, shape or the slope on the other side of the curve. This can be seen in the above said figure or any other diagram of a normal distribution curve.

4. The values of the mean, median and mode (in uni-modal distributions) computed for a normal distribution or those following the normal distribution curve always coincide at the same point and have the same value. The above diagram has shows it clearly.

   Mean = Mode = Median

5. The height of the normal curve is maximum at the mid point or the mean value. We can say that the height of the vertical axis drawn from the peak (called ordinate) is maximum at mean and in the unit normal curve, its value is equal to 0.3989.

6. The first and the third quartiles (i.e., Q_1 and Q_3) are equidistant from the mean or median in a normal distribution.

   We can say that

   $Q_3 -$ Median = Median $- Q_1$

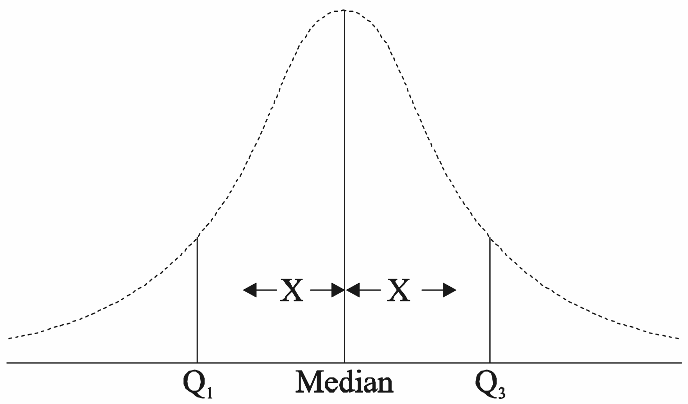

Figure 17.4: Quartiles showing in Normal Distribution Curve

If x is the distance between the third quartile and the median, then the distance between median and the first quartile will also be the x.

7. The distribution is continuous and is never ending on either side. Hence the normal distribution curve has no boundaries in either direction and the curve never ends or touches the base line, no matter how far it is extended. Such a curve which is never ending is called *symptotic* curve and it extends from minus infinity to plus infinity. Normal distribution is continuous for all values of x between plus infinity to minus infinity so that each conceivable interval of real numbers has a probability other than zero. You can see the above diagram for your understanding. The curve is not touching the base line at the ends on left side or the right side. It continues to approach the base line but never touches it. Another point to note in these diagrams is that even no portion of the curve lies below the base line in a normal distribution curve.

8. Since there is only one point where the frequency is maximum in a normal distribution (i.e., Mean), so the normal distribution will always be a uni – modal distribution. Normal distribution has only one mode.

9. The mean deviation in a normal distribution, if calculated, is four fifth of its Standard deviation

   The mean deviation = 4/5 standard deviation = 4/5 σ

   If Standard deviation of a normal distribution is 5 then its mean deviation will be 4. However the distribution has to be perfectly normal.

10. Semi – inter quartile range in a normal distribution also has a fixed value which is also called the probable error and its value is 0.6745 of the standard deviation.

    Semi – inter quartile range = Probable Error = 0.6745 standard deviation = 0.6745 σ

11. If we want to find the points in the same normal curve at which the curve changes its path or direction, we will observe that these points are also symmetrical. They are at a distance of one standard deviation from the mean on either side. These points are called points of inflexion.

12. The various ordinates at different points or different distances (distance in terms of standard deviation) from the mean ordinate in a normal distribution stand in a fixed proportion to the height of the mean ordinate. Thus the height of the ordinate at one standard deviation distance on either side of the mean ordinate is 60.653% of the height of the mean ordinate and so on.

13. The data cluster around the mean: The percentage of distribution area under a normal curve is given by the percentages given below in different cases. The percentages of area around mean are:

    i. Between mean and one Standard deviation or

       Between Mean to 1 S.D. = 34.13%

       Mean to 1σ = 34.13%

Mean to – 1 S.D. = 34.13%

Mean to – 1 σ = 34.13%

+ 1 σ to – 1 σ = 68.26%

ii. Between one Standard deviation to two standard deviation or

Between 1 S.D. to 2 S.D. = 13.59%

1 σ to 2 σ = 13.59%

Between – 1 S.D. to –2 S.D. = 13.59%

iii. Between mean and two Standard deviation or

Between Mean to 2 S.D. = 47.72%

Mean to 2 σ = 47.72%

Mean to –2 S.D. = 47.72%

Mean to –2 σ = 47.72%

+ 2 σ to –2 σ = 95.44%

iv. Between mean and 1.96 Standard Deviation or

Mean to 1.96 S.D. = 47.5%

Mean to 1.96 σ = 47.5%

Mean to –1.96 S.D.= 47.5%

Mean to –1.96 σ = 47.5%

+1.96 σ to –1.96 σ = 95%

v. Between mean and 2.58 Standard Deviation or

Mean to 2.58 S.D. = 49.5%

Mean to 2.58 σ = 49.5%

Mean to – 2.58 S.D. = 49.5%

Mean to – 2.58 σ = 49.5%

+ 2.58σ to –2.58σ = 99%

vi. Between mean and three Standard deviation or

Mean to 3 S.D. = 49.86%

Mean to 3 σ = 49.86%

Mean to –3 S.D. = 49.86%

Mean to –3 σ = 49.86%

+ 3 σ to –3 σ = 99.72%

v. Between mean and 2.58 Standard deviation or

Mean to 2.58 S.D. = 49.5%

Mean to 2.58 σ = 49.5%

Mean to –2.58 S.D.= 49.5%

Mean to -2.58σ = 49.5%

+ 2.58 σ to –2.58 σ = 99%

vi. Between mean and three Standard deviation or

Mean to 3 S.D. = 49.86%

Mean to 3 σ = 49.86%

Mean to –3 S.D.= 49.86%

Mean to –3 σ = 49.86%

+ 3 σ to –3 σ = 99.72%

vii. Between two and three Standard deviation or

2 S.D. to 3 S.D. = 2.15%

2 σ to 3 σ = 2.15%

–2 S.D. to – 3 S.D.= 2.15%

–2 σ to –3 σ = 2.15%

All these figures are approximated to two decimal places.

Let us try to understand this analysis with the help of normal distribution curve. In the following curve we have shown the areas under different conditions.

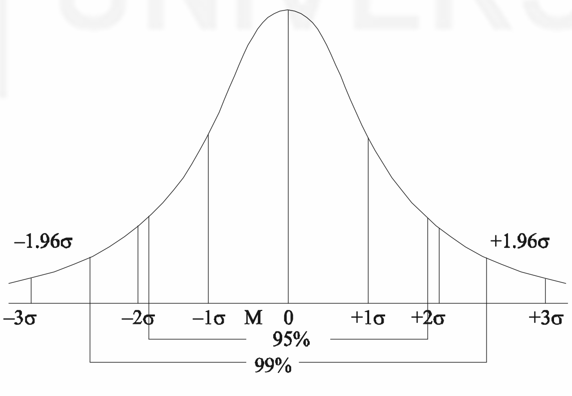

Figure 17.5 : Areas under Normal Probability Curve

The percentage of cases in a given normal curve are greatest around the centre i.e. around mean and go on decreasing as you move away from the Mean. Secondly, more than two third cases, i.e., 68.26% of the total area of the curve fall between ±1 standard deviation. If we find the number of cases between ±1.96 standard deviation, they come out to be 95% and between ±2.58 standard deviation are 99% of the cases. These observations are very important in applying the significance limits.

14. The total area under the normal curve in terms of probability is taken as 1 (or 100%).

15. The probability that a normal random variable X equals particular value is 0

16. It links frequency distribution to probability distribution.

17. Normal distribution has the same shape as Standard Normal Distribution.

18. Normal distribution is actually a family of distributions and has two important parameters as mean and standard deviation. These two parameters determine the shape of the distribution. If we have different means and different standard deviations then yield will be different density curves. We will have different normal distributions in this case. All these distributions have the same general shape. They may differ in their spread but they bear similar bell shaped curve.

Check Your Progress 1

Note: a) Write your answers in the space given below.

b) Compare your answers with those given at the end of the Unit.

1. Define a Normal Probability Curve.

....

....

....

2. Mention five properties of Normal Curve.

....

....

....

3. Why is Normal Distribution so important in education?

....

....

....

4. In a normal distribution, what percentage of frequencies is :

(a) between -1σ to $+1\sigma$

(b) between -2σ to $+2\sigma$

(c) between -3σ to $+3\sigma$

....

....

....

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

....

....

....

6. Fill in the blanks

   i. The shape of a normal distribution curve is …………….

   ii. In a standard normal distribution, mean = ……………. and standard deviation = …………….

   iii. Most of the statistical tests are ……………. distributed.

   iv. Percentage of cases between Mean and 1 σ = …………….

   v. Percentage of cases between Mean and 2 σ = …………….

   vi. 95% cases lie between Mean ± ……………. σ

   vii. 99% cases lie between Mean ± ……………. σ

17.3.4 Interpretation of Normal Probability Curve

Normal Curve has great significance in the cognitive measurement and educational evaluation. It gives important information about the trait being measured. If the frequency polygon of observations of measurements of a certain trait is a normal curve, it indicates that:

i. The measured trait is normally distributed in the Universe;

ii. Most of the cases are average in the measured trait and their percentage in the total population is about 68.26%;

iii. Approximately 15.87% of (50–34.13%) cases are high in the trait measured;

iv. Similarly, 15.87% cases approximately are low in the trait measured;

v. The test which is used to measure the trait is good;

vi. The test has good discrimination power as it differentiates between poor, average and high ability group individuals; and

vii. The items of the test used are fairly distributed in term of difficulty level.

17.3.5 Importance of Normal Probability Curve

The Normal Distribution is by far the most used distribution for drawing inferences from statistical data 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 and facts in (i) *biological statistics* e.g. sex ratio in births in a country over a number of years, (ii) *the anthropometrical data* e.g. height, weight, (iii) *wages and output of large numbers of workers* in the same occupation under comparable conditions, (iv) *psychological measurements* e.g. intelligence, reaction time, adjustment, anxiety and (v) *errors of observations* in Physics, Chemistry and other Physical Sciences.

2. The Normal distribution is of great value in educational evaluation and educational research, when we make use of cognitive measurements. 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 distribution of test scores approach the theoretical normal distribution as a limit, but the fit is rarely ideal and perfect.

17.3.6 Applications of Normal Probability Curve

Some of applications of normal curve in the field of educational measurement and evaluation 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;

(iv) hypothesis testing;

(v) determining standard error of mean;

(vi) to determine the percentile rank of a student in his own group;

(vii) to find out the percentile value of a student's percentile rank;

(viii) to compare the two distributions in terms of overlapping;

(ix) to determine the relative difficulty or test items; and

(x) dividing a group into sub-groups according to certain ability and assigning the grades.

17.3.7 Table of Areas Under the Normal Probability Curve

In order to use above applications of normal curve in educational measurement and evaluation, we need to know the Table of areas under the normal curve.

The Table 17.6 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. 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:

$$Z = \frac{X - M}{\sigma} \quad = \quad Z = \frac{x}{\sigma}$$

Where:

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 then be calculated.

The first column of the Table 17.6, I.E., x/σ gives distance in tenths ofσ measured off on the base line for the normal curve from the mean as origin. In the row, the x/ σ distance is 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 1 ? unit from the mean, we go down the x/ σcolumn 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 per cent of the entire area of the curve lie between the mean and 1σ. 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, 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 17.2 : *Fractional parts of the total area (taken as 10,000) under the normal probability curve, corresponding to distances on the baseline between the mean and successive points laid off from the mean in units of standard deviation*

Example : Between the mean and a point 1.38σ (x/ σ = 1.38) are found 41.62% of the entire area under the curve:

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	2357	2389	2422	2454	2486	2517	2549
0.7	2580	2611	2642	2673	2704	2734	276	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	2749	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	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								

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 17.2 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 +3 σ distance 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 +3 σ. The rest 0.27 percent of the distribution beyond ±3σ is considered too small or negligible except where N is very large.

17.3.8 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 NPC. 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.

17.3.9 Problems Related to Application of the Normal Probability Curve

(a) To determine the percentage of cases in a Normal Distribution within given limits or scores

Example 1

In a class mean score in an Urdu test is 26.1 and the standard deviation is 6.45. How many cases lie between mean and score 35. Now let us find Z score

We know Z score = $\frac{\text{Mean deviation}}{\text{Standard deviation}}$

$$Z = \frac{\text{Score} - \text{Mean}}{\sigma}$$

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

In this $X = 35 \quad \overline{X} = 26.1$ and $\sigma = 6.45$

Thus,

$$Z = \frac{35 - 26.1}{6.45}$$

$$\frac{8}{6.45}$$

$$= 1.38$$

Now from the table of standard score, we can find area. Since 35 is 1.38 standard deviation on right side of the curve, the area between mean and 1.38 σ from table comes to be 41.62%. Hence 41.62% students get marks between mean and 35 marks.

Example 2

If you want to find the percentage of students getting marks between Mean and 20 then

$$Z = \frac{20 - 26.1}{6.45}$$

$$= \frac{-6.1}{6.45}$$

$$= -0.946$$

Thus score 20 in a normal probability curve will lie 0.946σ on left side of mean as it is negative .The percentage of cases between mean and 0.946 σ from table is 32.76%. Hence 32.76% students score marks between 20 and 26.1.

Example 3

If we want to know the students getting marks between 20 and 35 then we can add the two percentages

Percentage of students scoring marks between 20–35 = 41.62 + 32.76 = 74.38%.

(b) To determine the percentile rank of a student in his own group

The percentile rank is defined as the percentage of scores below a given score.

Example 4

The raw score of a student of class X on an achievement test is 60. The mean of the whole class is 50 with standard deviation 5. Find the percentile rank of the student.

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

$$Z = \frac{60 - 50}{5} = \frac{10}{5}$$

$$Z = +2.00\ \sigma$$

According to the table of area under N.P.C., the area of the curve that lie between M and +2 σ is 47.72%.

The total percentage of cases below the score 60 is 50 + 47.22 = 97.72% or 98%

Thus, the percentile rank of a student who secured 60 marks in an achievement test in the class is 98.

(c) To determine the percentile value(score) of a student whose percentile rank is known

Example 5

In a class Rohit's percentile rank in the mathematics class is 75. The mean of the class in mathematics is 60 with standard deviation 10. Find out Rohit's marks in Mathematics achievement test. According to definition of percentile rank the position of Rohit on the N.P.C. scale is 25% scores above the Mean.

According to the N.P.C. Table the σ score of 25% cases from the Mean is +.67 σ.

Using the formula

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

$$\text{or} \quad +.67 = \frac{X - 60}{10}$$

$$\text{or} \quad X - 60 = 10 \times .67$$

$$\text{or} \quad X = 60 + 6.7$$

$$\text{or} \quad X = 66.7 \text{ (Say 67)}$$

Rohit's marks in mathematics are 67.

(d) Dividing a group into sub-groups according to the level of ability

Example 6

Given a group of 500 college students who have been administered a general mental ability test. The teacher wishes to classify the group in five categories and assign them grades A, B, C, D, E according to ability. Assuming that general mental ability is normally distributed in the population; calculate the number of students that can be placed in groups A, B, C, D and E.

We know that total area of the Normal Curve extends from – 3σ to +3 σ that is over a range of 6 σ.

Dividing this range by 5, we get the σ distance of each category = 6 σ /5 = 1.2 σ. Thus, each category is spread over a distance of 1.2σ. The category

C will lie in the middle. Half of its area will be below the mean, and the other half above the mean.

Theσ distance of each category is shown in the figure.

According to N.P.C. table the total percentage of cases from mean to .6 σ is 22.57.

The total cases in between –.6 σ to +.6 σ are 22.57 + 22.57 = 45.14%

Hence, in category C, the total percentage of students is = 45.14.

Similarly, according to N.P.C. Table the total percentage of cases from Mean to 1.8 σ is 46.41.

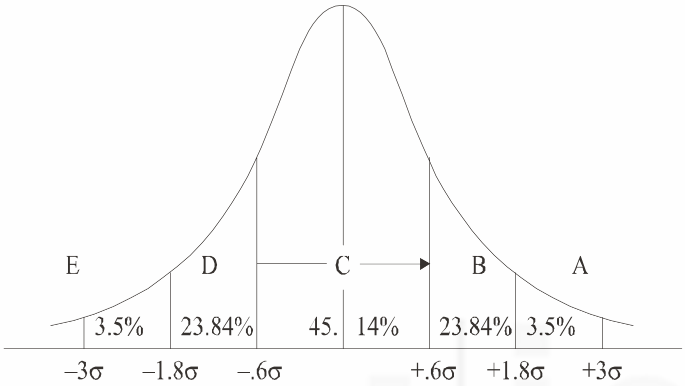

Figure 17.6: Percentages of Areas under Normal Probability Curve

The total percentage of cases in category B is 46.41 – 22.57 = 23.84%

In category A the total percentage of the cases will be 50 – 46.41 = 3.59%

Similarly, in category D and E the total percentages of the students will be 23.84% and 3.59% respectively. Thus :

In category A the total percentage of students is 3.59%

In category B the total percentage of students is 23.84%

In category C the total percentage of students is 45.14%

In category D the total percentage of students is 23.84%

In category E the total percentage of student is 3.59%

Exact numbers out of 500 are:

Category 'A' = 3.59 × 5 = 17.95 = 18 (Approx.)

Category 'B' = 23.84 × 5 = 119.2 = 119 (Approx.)

Category 'C' = 45.14 × 5 = 225.7 = 226 (Approx.)

Category 'D' = 119 (Approx.)

Category 'E' = 18 (Approx.)

17.4 DIVERGENCE FROM NORMALITY

The important property of the normal curve is its symmetry about the mean and bell shaped curve. However, there may be distributions which may be different, then normal may be asymmetrical. For studying this it becomes essential to understand the nature of asymmetry. The asymmetry of the distribution is studied though the measure of skewness. In such cases the measures of central tendency and the dispersion are not sufficient to describe a distribution completely. In a normal distribution also there is a divergence in shape and size or form. The peak may also differ in different distributions and even in a number of normal distributions. It is also possible to have frequency distributions which differ widely in their nature and compositions but they have same measures of central tendency and dispersion. In uni-variate data, we need some more measures to supplement the measures of central tendency and dispersion so as to make it more complete in its description. Two such measures are:

(i) Skewness,

(ii) Kurtosis.

(i) Skewness

The skewness will help you in understanding the nature and kind of asymmetry in a distribution. It will be helping you in understanding the distributions which are not normal although they may be symmetrical. In histogram you got a general idea of the shape of a distribution but what about the amount and direction of asymmetry. How much was the deviation from symmetry and to which side is the distribution departing from a normal distribution can be understood by studying the skewness of the distribution.

The skewness can be defined as a degree of asymmetry of a distribution around its mean. It is due to the lack of symmetry. Skewness quantifies how symmetrical the distribution is. A symmetrical distribution has a skewness value of '0'. Two distributions have the same means as 20 and the same standard deviation of 5. But the two distributions may not be alike in nature. One may be symmetrical distribution but the other may not be symmetrical distribution. The second may be asymmetrical or skewed. Measure of skewness will help you to distinguish between different types of distributions. In such cases, the median falls at a point other than the mean.

There are two types of skewness which appear in the Normal Curve.

(a) Negative Skewness

(b) Positive Skewness

(a) Negative Skewness

Distribution is 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, and are spread out gradually towards the low end i.e. the left side of the curve. In a negatively skewed distribution the value of median will be higher than that of the value of the mean. Can you think why would it be so?

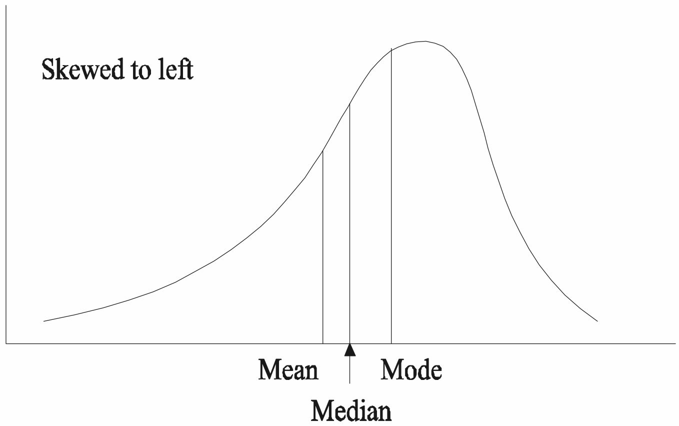

Figure 17.7: Negative Skewness

(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 towards the high or right end as shown in the figure below.

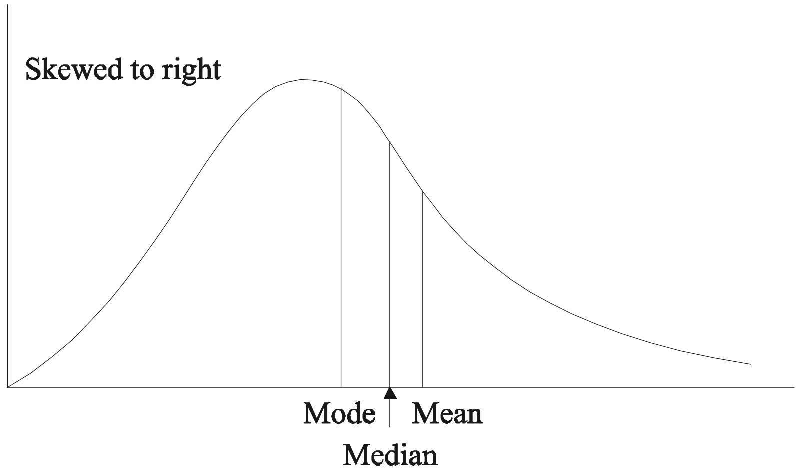

Figure 17.8: Positive Skewness

ii) Kurtosis

Another parameter of divergence is the vertical divergence in symmetry. Tallness or sharpness of the peak of a distribution comes under this category of divergence. Tallness and shapes of the central peak can be measured through kurtosis. Kurtosis will help you to get an idea of the height of the peak, sharpness of the peak or the broadness of the central peak of a distribution. Even if the distribution happens to be normal or symmetrical; the height, breadth or the sharpness of the central peak in such a distribution may vary. If we try to measure the height, breadth and the sharpness of the central peak with respect to the rest of the data, a new term or number comes and this number is known as kurtosis. Kurtosis is actually the measure of the shape of distribution.

There are two types of kurtosis which appear in the Normal Curve.

(a) Lepto kurtosis

(b) Platy kurtosis

(c) Mesokurtic

(a) Leptokurtosis

If you have a normal curve which is made up of steel wire and you push both the ends of the wire curve together, the curve become more peaked i.e. its top becomes narrower than the normal curve and scatterness in the scores or area of the curve shrink towards the centre.

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

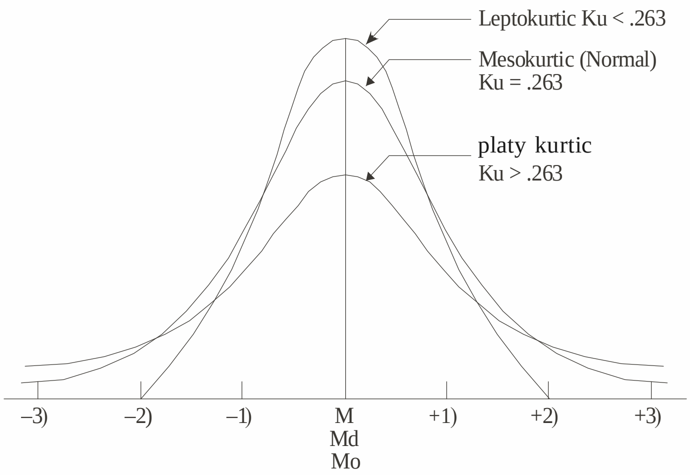

Figure 17.9: Kurtosis in the Normal Curve

(b) Platy kurtosis

If we put a heavy pressure on the top normal curve made from the steel wire. The top of the curve would become more flat than that of the normal.

Thus a distribution of flatter peak than of the normal distribution is known as platykurtic distribution.

When the distribution and related curve is normal, the value of kurtosis is .263 (Ku = .263). If the value of the Ku is greater than .263, the distribution and related curve obtained will be Platykurtic. When the value of Ku is less than .263, the distribution and related curve obtained will be Leptokurtic.

(C) Mesokurtic

Mesokurtic is called as the normal probability curve. Which is moderately peaked and not highly flattened. This curve is also not skewed to any direction. It follows all the characteristics of a normal distribution.

17.4.1 Factors Causing Divergence from Normal Probability Curve

The reasons why distributions exhibit skewness and kurtosis are numerous and often complex, but a careful analysis of the data can often throw some light on the symmetry. Some of the common causes are:

1. Selection of the Sample

Selection of the subjects (individuals) can 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.

The scores made by small and homogeneous group are likely to yield narrow and leptokurtic distribution. Scores from small and highly heterogeneous group yield platykurtic distribution.

2. Unsuitable or Poorly Made Tests

If the measuring tool of test is inappropriate for the group on which it has been administered, or poorly made, the asymmetry is likely to occur in the distribution of scores. If a test is too easy, scores will pile up at the high end of the scale, whereas when the test is too difficult, scores will pile up at the low end of the scale.

3. ***The Trait being measured is Non-Normal***

Skewness or Kurtosis will appear when there is a real lack of normality in the trait being measured. e.g. interests or attitudes.

4. ***Errors in the Construction and Administration of Tests***

A poorly constructed test may cause asymmetry in the distribution of the scores. Similarly, while administrating the test, unclear instructions, error in timings, errors in the scoring practice and lack of motivation to complete the test may cause skewness in the distribution.

Check Your Progress 2

Note: a) Write your answers in the space given below.

b) Compare your answers with those given at the end of the Unit.

7. Define the following:

(a) Skewness

....

....

....

(b) Negative and Positive Skewness

....

....

....

(c) Kurtosis

....

....

....

(d) Platykurtosis

....

....

....

(e) Leptokurtosis

....

....

....

(8) In case of normal distribution what should be the value of kurtosis?

....

....

....

....

(9) What is the significance of the knowledge of skewness and kurtosis for a school teacher?

....

....

....

(10) Fill in the blanks.

1. Skewness is the degree of .... of a distribution.
2. Negatively skewed distribution has .... in a distribution curve.
3. Positively skewed distribution has .... in a distribution curve.

17.5 LET US SUM UP

The normal distribution helps us to study variables used in behavioural research because they tend to be normally distributed.

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

The normal distribution is a significant tool in the hands of teacher, through which one can decide the nature of the distribution of the scores obtained on the basis of measured variable. He/she can judge the difficulty level of the test items in the question paper and finally he may know about his/her class, whether it is homogeneous to the ability measured or it is heterogeneous one. The normal distribution also helps any individual to classify the distribution in different categories as per needed criteria.

17.6 REFERENCES AND SUGGESTED READINGS

Aggarwal, Y.P. (1990), *'Statistical Methods – Concepts, Applications and Computation'*, Sterling Publishers Pvt. Ltd., New Delhi.

Anderson, Richard L. (1952), *'Statistical Theory in Research'*, McGraw hills Book Co., New York.

Best John W and James V. Kohn (2004), *'Research in Education'*, Prentice Hall of India Pvt. Ltd., New Delhi.

Downie, N.M. (1983), *'Basic Statistical Methods'*, HarperCollins College Div., New York.

Ferguson, G.A. (1974), *'Statistical Analysis in Psychology and Education'*, McGraw Hill B.K. Co., New York.

Garrett. H.E. & Woodworth, R.S.(1969), *'Statistics in Psychology and Education'*. Vakils, Feffer & Simons Pvt. Ltd., Bombay.

Guilford, J.P. & Benjamin, F.(1973), *'Fundamental Statistics in Psychology and Education'*, McGraw Hill B.K. Co., New York.

Kerlinger, F.N.(1993), *'Foundations of Behavioural Research: Educational and Psychological Inquiry'*. Holt, Rinchart & Winston, New York.

Loether, H.L. & McTavish, D.G. (1988), *'Descriptive and Inferential Statistics: An Introduction'*, 3rd ed. Boston: Allyn & Bacon.

Sonedecor, G.W. & Cochran, W.G.(1967), *'Statistical Methods'* (6th ed.), Lowa State University Press, Ames, Lowa.

17.7 ANSWERS TO CHECK YOUR PROGRESS

1. Normal Probability Curve is a bell shaped curve obtained for a distribution having maximum frequency near the central value of distribution and the frequency gradually tapers off symmetrically on both the sides.

2. The Normal Curve is symmetrical about the ordinate at the central point of the curve.

   - It is unimodal; the mode is always at the central point of the curve.
   - It is asymptotic to the x-axis.
   - The points of influx occur at ±1.
   - The area of the curve between the points of influxation is fixed.

3. Normal Distribution is important in education for use in biological statistics, psychological measurement, errors in observations, ability testing, estimation of scores, categorisation, etc.

4. (a) Between –1 σ and + 1 σ, there are 68.26% of the frequencies

   (b) Between –2 σ and + 2 σ, there are 95.44% of the frequencies

   (c) Between –3 σ and + 3 σ, there are 99.73% of the frequencies

5. The two ends of the normal probability curve are considered closed at the points ± 3 σ, as almost all the cases (99.73% of the cases, to be exact) lie between these two points and there is a rare probability of a case going beyond these two limits.

6. i. Bell shaped

   ii. Mean = 0 and Standard Deviation = One

   iii. Normally

iv. 34.13 percent

v. 47.72 percent

vi. 2 σ

vii. 2.58 σ

7. (a) A distribution is said to be skewed, if the point of centre of gravity is located on one side of the distribution i.e. away from the centre of the scale of measurement.

   (b) A distribution is said to be negatively skewed if the scores are concentrated at the higher end of the measurement scale and it is said to be positively skewed it the scores are concentrated at the lower end of the measurement scale.

   (c) Kurtosis refers to the divergence in the height of the curve or the peakedness of the curve.

   (d) A distribution of flatter peak than the normal one is known as platy – kurtosis distribution.

   (e) A distribution which is more peaked than the normal one is known as lepto-Kurtosis distribution.

8. In case of normal distribution the value of Kurtosis is 0.263.

9. If the distribution of scores obtained by a school teacher is not normal, he/she will try to find out the reasons of skewness and kurtosis of the distribution. One of the reasons may be that the group of individuals are different from the normal one, or it may be due to the nature of the trait itself. Teacher should know that ultimately the goal of education is to have the distribution of scores of individuals as negatively skewed.

10. i. Asymmetry

    ii. Skewness on left

    iii. Skewness on right