UNIT 14 ANALYSIS OF QUANTITATIVE DATA (DESCRIPTIVE STATISTICAL MEASURES : SELECTION AND APPLICATION)

Structure

14.1 INTRODUCTION

In the Block 2, you were introduced to various approaches to research: historical, philosophical, descriptive, experimental, etc. You were also explained the techniques of sampling, alongwith the development and use of tools in the collection of data. The data are of two types: Qualitative and Quantitative. Descriptive narrations, observation notes, responses to questions, etc. are qualitative data; whereas quantitative data comprise numerical figures. The quantitative data are either parametric or non-parametric in nature. Various descriptive statistical measures are used in the analysis of parametric and non-parametric data. These measures include: (i) measures of central tendency; (ii) measures of variability; (iii) measures of relative position; and (iv) measures of relationship. Descriptive statistical analysis using these measures limits generalisation to the particular group of individuals observed. No conclusions are extended beyond this group, and any similarity to those outside the group cannot be assumed.

In this unit you will study the nature of qualitative and quantitative data. The techniques which produce data of quantitative nature will also be explained. You will also be introduced to various descriptive statistical measures which are used in the analysis of quantitative data. The computation, selection and application of these descriptive statistical measures will also be explained with the help of illustrations. You will also understand the nature and characteristics of a normal probability curve, and its applications in educational research.

14.2 OBJECTIVES

After studying this unit, you will be able to:

- name and describe the nature of various types of data;
- describe the techniques/tools which generate data of quantitative type;
- describe the procedures of classifying data and their graphical representation;
- name and describe various statistical measures which are used in the analysis of quantitative data;
- define, Mean, Median and Mode as the measures of central tendency; their computation, selection and applications;
- define Range, Average Deviation, Quartile Deviation and Standard Deviation as the measures of variability; their computation, selection and application;
- discuss the nature, characteristics and applications of Normal Probability Curve;
- define Sigma Scores, T-Scores and Percentile Rank as the measures of comparing individuals; their computation, selection and application; and
- define Product moment correlation and Rank difference correlation as the measures of relationship; their computation, selection and application.

14.3 TYPES OF DATA

The data collected from various sources through the use of different tools and techniques generally consists of numerical figures, ratings, narrations, responses to open-ended questions comprising a questionnaire or an interview schedule, quotations, field notes etc. In educational research usually, the use of two types of data are recognised. These are qualitative data and quantitative data.

14.3.1 Qualitative Data

Qualitative data are verbal or other symbolic materials. The detailed descriptions of observed behaviours, people, situations and events, are some examples of qualitative data. These data are gathered by a variety of methods and techniques such as: (i) observation; (ii) interviews; (iii) questionnaires, opinionnaries and inventories; and (iv) recorded data from newspapers and other documents maintained by educational institutions, courts, clinics, government and non-government agencies etc. These data are narrations and indicate what people have said in their own words about their experiences and interactions in natural settings. The data gathered through observation may consist of patterns of action and verbal and non-verbal interaction between members of the group in their natural settings, and are qualitative in nature. The responses to open ended questions in a questionnaire are neither systematic nor standardized. However, such responses are detailed and comprehensive which provide data of qualitative nature in the form of emotions of respondents, their thoughts and experiences about the phenomena under study.

Further details about the nature of qualitative data, their analysis and interpretation will be explained to you in the units 17 and 18.

14.3.2 Quantitative Data

Quantitative data are obtained by using various tools and tests based on scales of measurement: nominal, ordinal, interval or ratio. In educational research, we generally come across nominal, ordinal or interval scale of measurement. Ratio scale measurements are almost non-existent in the educational studies. The experiences of people are fitted into standard responses to which numerical values are attached. These data are close-ended and hardly provide depth and detail which is possible only in the case of qualitative data.

The quantitative data are either **parametric** or **non-parametric**

Parametric data are measured on interval or ratio scale measurements. Ratio scale occupies the highest level in the measurement, because it permits all the four operations of addition, subtraction, multiplication and division and hence all statistical techniques are permissible with such scales. The measurement scales have all the characteristics of the interval scale, with additional advantage of a true zero point. For example, the zero point on a centimeter scale indicates complete absense of length and height. It may be noted that ratio scales are almost non-existent in psychological and educational measurements except in the area of pscho-physical judgement where we measure the reaction time with the help of customary time unit like second and fraction of the seconds. Since in educational research we would be mostly dealing with interval scale measurement data, some examples about this scale are provided for clarifying the concept. If we actually measure the weight of the three students, A, B and C by using a weighing machine and find their weights as 41, 49 and 53 kilograms respectively, we have a measurement on a scale of equal units. This scale of measurement is called an interval scale. By this measurement we can say that C is 4 kilograms heavier than B and 12 kilograms heavier than A. We can also infer that the difference in weights between C and A is thrice than between C and B. The marks obtained by students in Hindi in a certain class provide data on interval scale.

In interval scale, the difference between consecutive points on the scale are equal over the entire scale but there is no zero point on it. The zero point (point of reference) of the scale is chosen conventionally or arbitrarily. The scores on an intelligence test or an attitude scale are also based on interval scales. They have no real zero point. To illustrate this concept, suppose a student gets "zero" score in a test of mathematics, this does not mean that the student has no knowledge of mathematics.

The operations of addition and subtraction can be performed on interval scales; the statistical techniques based on these operations are permissible. However, since operations of multiplication and division assume the existence of an exact zero point, these operations cannot be used with interval scales.

Nominal as well as ordinal scales of measurement provide non-parametric data. These data are either counted or ranked. Nominal scales are used when a set of objects among two or more categories are to be differentiated on the basis of some similar defined characteristics. Usually symbols or numericals are chosen to represent all objects in a given category. We assign students to such categories as locality (rural and urban), gender (females and males) etc. Nominal scales are non-orderable and only arithmetical operation applicable to such scales is Counting, the mere enumeration of individuals in a particular category. Statistical analysis based on counting is permissible in this type of measurement.

The measurement based on ranks is ordinal scale measurement. In this scale, objects or individuals are ordered (ranked) on some continuum in a series from lowest to highest according to measured characteristics. The ranking of students in a class for weight, height or academic achievement are the examples of ordinal scale. Suppose we rank three students X, Y and Z in order of their height, X as the tallest, and assign the number 1, 2, 3 respectively as their ranks.

In this way we have information about the serial arrangement. We cannot infer any information as how much X is taller than Y or Z, even though the three numbers assigned to them are equally spaced on the scales of measurement. The statistical analysis based rank is permissible in this measurement.

14.4 GRAPHIC REPRESENTATION OF QUANTITATIVE DATA

Graphic representation of quantitative data is helpful in reading and interpreting data. The following four methods of graphic representation are in general use.

- Histogram or Column Diagram
- Frequency Polygon
- Cumulative Frequency Curve
- Cumulative Percentage Curve or Ogive

14.4.1 The Histogram or Column Diagram

Consider the following distribution of scores. The distribution has been obtained on the basis of the scores by a group of 40 students in a test of chemistry in which the maximum score achieved by a student was 65 and minimum score was 10.

Scores (Class Interval)	Exact Units of Class Intervals	f (Frequency)
60 - 69	59.5 - 69.5	1
50 - 59	49.5 - 59.5	4
40 - 49	39.5 - 49.5	10
30 - 39	29.5 - 39.5	15
20 - 29	19.5 - 29.5	8
10 - 19	9.5 - 19.5	2
		N = 40

Let us construct a histogram or column diagram representing the above distribution. Histogram is a graph in which the exact class intervals are represented along the horizontal axis called x-axis and their corresponding frequencies are represented by areas in the form of rectangular basis drawn on the intervals as shown in the figure (Fig. 14.1).

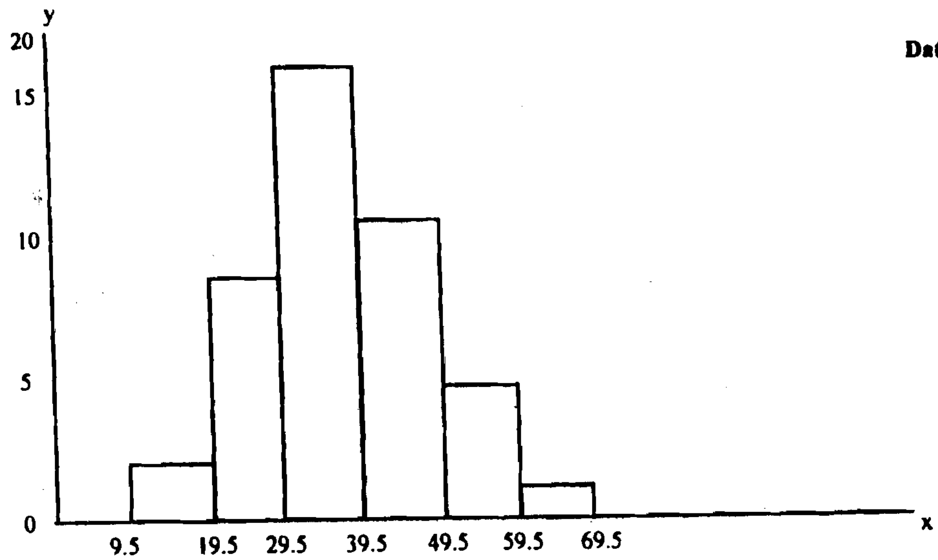

Fig. 14.1 : The Histogram or Column

14.4.2 Frequency Polygon

Another method of representing a frequency distribution graphically is 'Frequency Polygon'. It is obtained by identifying the point which represent the mid-point of each class interval and then joining all these points by straight lines as shown in the figure (Fig. 14.2).

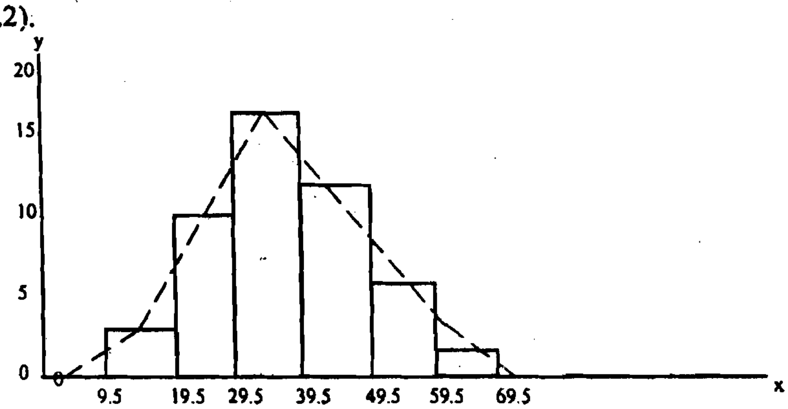

Fig. 14.2 : Frequency Polygon

14.4.3 Cumulative Frequency Curve

Cumulative frequency curve differs from frequency polygon in two ways. First, instead of plotting points corresponding to frequencies, we plot points corresponding to cumulative frequencies. Second, instead of plotting mid points of each class, we plot our points above the exact upper limit of the class interval. This is done because in this graph we wish to represent the number of cases falling above or below the particular values. Let us consider the distribution as given in 14.4.1.

Exact Units of Class Intervals	f (Frequency)	F (Cumulative Frequency)
59.5 - 69.5	1	40 = (39+1)
49.5 - 59.5	4	39 = (35+4)
39.5 - 49.5	10	35 = (25+10)
29.5 - 39.5	15	25 = (10+15)
19.5 - 29.5	8	10 = (2+8)
9.5 - 19.5	2	2
	N = 40	

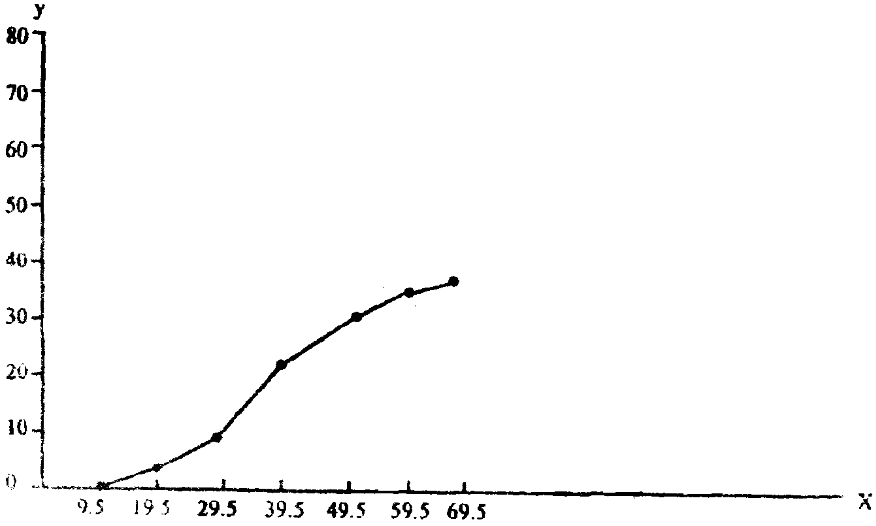

Fig. 14.3 : Cumulative Frequency Curve

For drawing the cumulative frequency curve of this distribution, each cumulative frequency is plotted at the exact upper limit of the class interval upon which it falls. For the given distribution first value of cumulative frequency corresponding to 19.5 is 2; the second value of cumulative frequency corresponding to 29.5 is 10 and so on to the last point for which value of cumulative frequency corresponding to 69.5 is 40. By joining the plotted points by lines, we get 'frequency curve' as illustrated in the Figure (Fig. 14.3). It may be noted that in order to have the curve begin on the horizontal axis, we start at the exact lower limit of the lower class (i.e. 9.5 in the present case), the cumulative frequency of which is taken to be '0'.

14.4.4 Cumulative Percentage Curve or Ogive

Cumulative percentage curve or Ogive differs from cumulative frequency curve in that frequencies are expressed as cumulative percents of N on the vertical axis instead of as cumulative frequencies, as illustrated in the given example.

Exact Units of Class Intervals	f	F (Cumulative Frequency)	Percentage (Cumulative Frequency)
59.5 - 69.5	1	40	100
49.5 - 59.5	4	39	97.5
39.5 - 49.5	10	35	87.5
29.5 - 39.5	15	25	62.5
19.5 - 29.5	8	10	25.0
9.5 - 19.5	2	2	5.0
	N = 4[illegible]		

In the above distribution, the percentage cumulative frequency are expressed as percentages of N in the last column. After finding cumulative percentage frequencies we plot these frequencies corresponding to upper exact limits of class intervals, as shown in the figure (Fig. 14.4). The curve joining the points thus drawn is called cumulative percentage curve or Ogive.

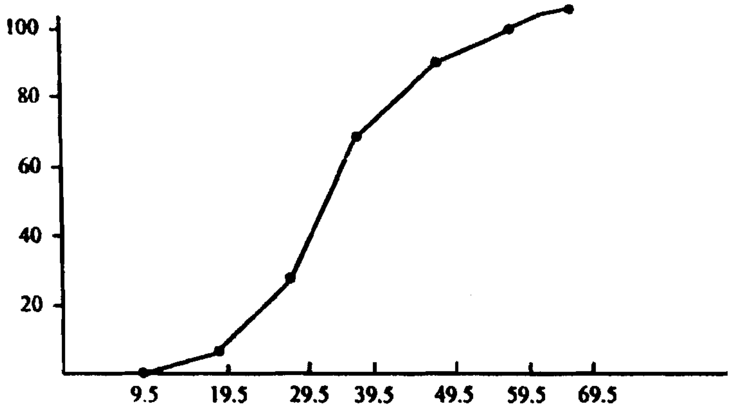

Fig. 14.4 : Cumulative Percentage Curve or Ogive

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

1. i) What are the various types of quantitative data.

....

....

ii) Name various techniques of presenting data graphically.

....

....

14.5 DESCRIPTIVE STATISTICAL MEASURES

Statistical methods are extensively used in analyzing quantitative data in educational research. Generally, these methods are classified as 'descriptive statistics' and 'inferential statistics'. Descriptive statistics are computed to describe the characteristics (attributes) of a sample or population in totality and thus limit generalization to the particular group (sample). No conclusions are extended beyond this group whereas inferential statistical methods are used to draw generalizations beyond the sample with a known degree of accuracy.

In this unit we shall confine our discussion to the understanding of four measures of descriptive statistics commonly used in educational research. These include measures of: (i) central tendency or averages; (ii) variability or dispersion ; (iii) relative position; and (iv) relationship.

14.5.1 Measures of Central Tendency or Averages

We use averages to describe the characteristics of samples (groups) in a general way. Economic status of groups indicated by 'average income', performance of students in a class is judged by 'grade point averages', or the climate of a city is reported by 'average temperature'. But in statistical language, use of the term 'average' does not convey any specific meaning, because there are a number of types of averages, only one of which may be appropriate to use in describing given characteristics of a sample/group. Of the many averages that may be used, three are mostly used in the analysis of educational data. These include: Mean, Median and Mode.

The Mean

The Mean of a distribution is the arithmetic average. It is perhaps the most familiar, most frequently used and well understood average. It is computed by dividing the sum of all the observations by the total number of observation. If $x_1, x_2, x_3, \ldots\ldots\ldots x_n$ are the N observations, the formula for computing the Mean (X) is given

$$\text{Mean} = M = \frac{x_1 + x_2 + x_3 + \ldots\ldots\ldots\ldots\ldots\ldots x_n}{N}$$

$$= \frac{\sum x}{N}$$

Suppose 10, 12, 7, 6, 4, 10, 15 are the marks obtained by 7 students in a test of Hindi, the Mean is

$$\text{Mean} = M = \frac{10+12+7+6+4+10+5}{7} = \frac{64}{7} = 9.14$$

In case of large data, the observations are arranged in a frequency distribution. Consider the following distribution which was used in 14.4.1.

Scores Class Interval	Exact Units of Class Interval	Mid Point x	f	x'	fx'
60 - 69	59.5 - 69.5	64.5	1	3	3
50 - 59	49.5 - 59.5	54.5	4	2	8
40 - 49	39.5 - 49.5	44.5	10	1	10
30 - 39	29.5 - 39.5	$\boxed{34.5 \text{ A.M.}}$	15	0	0
20 - 29	19.5 - 29.5	24.5	8	–1	–8
10 - 19	9.5 - 19.5	14.5	2	–2	–4
			N = 40		$\sum fx' = 9$

These data are grouped in a frequency distribution. The following formula is used for computing the Mean in such cases:

$$\text{Mean} = M = A.M. + \frac{\sum fx'}{N} \times i$$

in which

A.M. = Assumed Mean

f = Frequency of the Class Interval

x' = Deviation of the score from the Assumed Mean Divided by the length of the class interval

i = Width of the class interval

N = Total number of observations

In the column Mid point (x) of the example, we have computed the mid-point of each class interval and taken the assumed mean (A.M.) at the interval which has the maximum frequency.

The values for the column (x') have been calculated by taking deviation of each mid point (x) from the assumed mean dividing it by the length of the class interval i.e. 10 in the present case. The values fx' are obtained by multiplying each f with the corresponding x'.

Using the formula:

$$\text{Mean} = M = A.M. + \frac{\sum fx'}{N} \times i$$

$$= 34.5 + \frac{9}{10} \times 10$$

$$= 34.5 + 2.25$$

$$= 36.75$$

Selection and Application of Mean

Generally, Mean is a more useful measure of central tendency than its other measures.

Mean is an appropriate statistic when

i) the observations or measures are distributed symmetrically about the centre of the distribution;

ii) the most stable measure of central tendency is desired;

iii) additional statistics are to be computed later.

The Median

The Median is a point in an ordered arrangement of observations, above and below which one-half of the observations fall. It is a measure of a position rather than one of magnitude.

If the number of observations are ungrouped and small, we arrange them in order of magnitude and identify the middle observation by counting up half the value of N (the total number of observations). When the number of observations (N) is odd, the mid-observation is the Median.

For example, 45 is the median of the marks 30, 31, 45, 48, 52 obtained by 5 students in a test of Mathematics.

When the number of observations is even, the Median is the mid-point between two middle observations.

For example, $\frac{35+37}{2}$ = 36 is the Median of the observations:

29, 31, 35, 37, 40, 44.

In case of data, grouped in frequency distribution we use the following formula:

$$\text{Median} = \text{Mdn} = l + \frac{(N/2 - F)}{f} \times i$$

in which

l = Exact lower limit of the class interval upon which the median lies.

i = Width of class interval in which the median falls.

f = Frequency within the class interval upon which the median lies.

F = Sum of all the frequencies below l.

$\frac{N}{2}$ = One-half of the total number of observations.

To illustrate the use of this formula, consider the data which was used in 14.4.1 for computing the Mean.

Here $\frac{N}{2}=\frac{40}{2}=20$; $l=29.5$; $F=8$; $f=15$ and $i=10$

Using the formula:

$$\text{Median} = (\text{Mdn.}) = l + \frac{(N/2-F)}{f} \times i$$

$$= 29.5 + \frac{(20-8)}{15} \times 10$$

$$= 29.5 + \frac{12}{15} \times 10$$

$$= 29.5 + 8$$

$$= 35.5$$

Selection and Application of Median

We should compute Median when

i) the distribution of the observation measures is markedly skewed;

ii) there is not sufficient time to compute a Mean;

iii) an incomplete distribution is given;

iv) we are interested in whether cases fall within the upper or lower halves of distribution and not particularly in how far they are from the central point.

The Mode

The mode is defined as the most frequently occurring observation in a distribution. It is located by inspection rather than by computation. If there is only one value which occurs a maximum number of times, then the distribution is said to have one mode or to be unimodal. In some distributions there may be more than one mode. A two mode distribution is bimodal; more than two, multimodal.

In a simple ungrouped series of measures of observations, the crude or empirical mode is that single measure which occurs most frequently. For example, in the series 19, 20, 21, 26, 28, 28, 29 and 31, the most often recurring measure, namely, 28, is the crude or empirical mode.

In grouped data distributions, the mode is assumed to be the mid score of the interval in which the greatest frequency occurs. For example, in the grouped distribution under 14.4.1 which was used for computing Mean, 34.5 is the Mode. It is the mid score of the interval 29.5-39.5 with maximum frequency of 15.

Selection and Application of Mode

The mode is the most appropriate statistics when

i) the quickest estimate of central tendency is wanted;

ii) a very rough estimate of central tendency will suffice;

iii) the purpose is to know the most typical case.

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

2. i) Name the most stable measure of central tendency.

....

ii) Compute mean and mode for the following distribution:

10, 12, 5, 7, 8, 9, 15, 5

....

....

iii) Compute median for the following distribution.

Scores	f
120 - 139	50
100 - 119	150
80 - 99	500
60 - 79	250
40 - 59	50
	N = 1000

14.5.2 Measures of Variability or Dispersion

Measures of central tendency describe location along an ordered scale. Although these statistics are very useful in describing the nature of a distribution of measures, these will not help the researcher how the observation data tend to be distributed. For this another kind of statistics, the measure of variability, is used. It is also called the measure of dispersion or spread.

There are several measures of variability, but in this unit we shall confine our discussion to the: (i) range, (ii) average deviation and (iii) standard deviation.

The Range

The range is the simplest measure of variability. It is the difference between the highest and lowest observation or score. For example, suppose the marks obtained by a group of 10 students in a test of mathematics are:

90, 80, 72, 71, 70, 70, 69, 68, 60, 50

The range for this distribution will be (90 -50) + 1 = 41

Selection and Application of Range

Although the range has the advantage of being easily calculated, it has some limitations. First, the value of range is based on only two extreme scores in the total distribution and thus it does not provide us any information of the variation of many other scores of the distribution. Second, it is not a stable statistics as its value can differ from sample to sample drawn from the same population.

The range is used when:

i) the data are too scant or too scattered;

ii) only an idea of extreme scores or of total spread is wanted.

Average Deviation (AD)

The average deviation or AD is the mean of the deviations of all of the separate observations or scores in a series taken from their mean. However, in averaging deviations to find average deviation (AD), the signs (+ and –) are not taken into consideration i.e. all the deviations, whether plus or minus, are treated as positive.

The formula for computing average deviation of an ungrouped data is:

$$\text{Average Deviation} = \text{AD} = \frac{\sum|x|}{N}$$

in which

$\sum|x|$ = Sum of deviations of scores from their mean disregarding plus and minus signs.

N = Total number of scores.

To illustrate the computation of average deviation, consider the following ungrouped distribution of marks obtained by a group of ten students in a test of spellings.

18, 19, 21, 19, 27, 31, 22, 25, 28, 20

Let us find the mean of the distribution:

X	x	$\lvert x \rvert$
18	– 5	5
19	– 4	4
21	– 2	2
19	– 4	4
27	+ 4	4
31	+ 8	8
22	– 1	1
25	+ 2	2
28	+ 5	5
20	– 3	3
$\sum x = 230$	$\sum x' = 0$	$\sum\lvert x'\rvert = 23$

$$\text{Mean} = \frac{\sum x}{N} = \frac{230}{10} = 23$$

The deviations of the scores from the mean, neglecting the negative signs, are given in the column |X|.

Using the formula:

$$\text{Average Deviation} = \frac{\sum|x|}{N} = \frac{23}{10} = 2.3$$

It may be noted that the sum of the deviations of the observations (scores) from their actual mean is zero. In the present case = $\sum x = 0$

For a grouped data distribution, the formula for computing average deviation is:

$$\text{Average Deviation (AD)} = \frac{\sum|fx|}{N}$$

in which $\sum|fx|$ = sum of the product of f and x, ignoring the minus sign.

Using the data of 14.4.1, the Average Deviation of the distribution will be:

$$\text{Average Deviation} = \frac{\sum|fx|}{N} = \frac{33}{40} = 0.825$$

Selection and Application of Average Deviation (AD)

Average deviation is used when:

i) It is desired to consider all deviations from the mean according to their size;

ii) Extreme deviations would affect standard deviation unduly.

Quartile Deviation (Q)

The quartile deviation or Q is one-half the scale distance between the 75th and 25th percentiles in a frequency distribution. The 25th percentile or Q_1 is the first quartile on the score scale, the point below which lie 25 per cent of the scores. The 75th percentile or Q_3 is the third quartile on the score scale, the point below which lie 75 per cent of the score. Mathematically,

in which

$$Q = \frac{Q_3 - Q_1}{2}$$

$$Q_1 = l + i\frac{\left[\frac{N}{4} - F\right]}{f}$$

$$Q_3 = l + i\frac{\left[\frac{3N}{4} - F\right]}{f}$$

where

l = The exact lower limit of the interval in which the quartile falls

i = The length of the interval

F = Cumulative frequency upto the interval which contains the quartile

f = The frequency on the interval containing the quartile.

To illustrate the use of the formula, consider the following distribution:

Scores	Exact Units of Class Interval	f	F
52 - 55	51.5 - 55.5	1	65
48 - 51	47.5 - 51.5	0	64
44 - 47	43.5 - 47.5	5	64
40 - 43	39.-5 - 43.5	10	59
36 - 39	35.5 - 39.5	20	49
32 - 35	31.5 - 35.5	12	29
28 - 31	27.5 - 31.5	8	17
24 - 27	23.5 - 27.5	2	9
20 - 23	19.5 - 23.5	3	7
16 - 19	15.5 - 19.5	4	4
		N = 65	

Here $\frac{N}{4} = \frac{65}{4} = 16.25$ and $\frac{3N}{4} = 48.75$

$$Q_1 = l + \frac{\left[\frac{N}{4} - F\right]}{f} = 27.5 + 4\,\frac{(16.25-9)}{8}$$

$$= 27.5 + 4\,\frac{(7.25)}{8}$$

$$= 27.5 + \frac{7.25}{2}$$

$$= 27.5 + 3.63 = 31.13$$

$$Q_3 = l + i\,\frac{\left[\frac{3N}{4} - F\right]}{f} = 35.5 + 4\,\frac{(48.75-29)}{20}$$

$$= 35.5 + 4\,\frac{(19.75)}{20}$$

$$= 35.5 + \frac{19.75}{5}$$

$$= 35.5 + 3.95 = 39.45$$

$$Q = \frac{Q_3 - Q_1}{2} = \frac{39.45 - 31.13}{2} = 4.16$$

Selection and Application of Quartile Deviation (Q)

Quartile deviation is when:

i) the mean is the measure of central tendency;

ii) there are scattered or extreme scores which would influence the standard deviation disproportionately;

iii) the concentration around the median, the middle 50 per cent of cases, is of primary interest.

Standard Deviation (SD)

The standard deviation (SD) is the most general and stable measure of variability. It is the positive square root of variance. The average of the squared deviations of the measures or scores from their mean is known as variance which is generally denoted as σ^2. The standard deviation is also denoted as σ .

The standard deviation for the ungrouped is computed by using the formula:

$$SD = \sigma = \sqrt{\frac{\sum x^2}{N}}$$

in which

x = Deviation of the raw score or measure from the Mean.

N = Number of scores or measures.

To illustrate the use of this formula, let us consider the data:

15, 10, 15, 20, 8, 10, 25, 9

The mean or average score is:

$$M = \frac{\sum x}{N} = \frac{112}{8} = 14$$

Now we subtract the mean from each of the raw scores in the distribution and proceed as under:

Score (x)	X – M or x	x^2
15	1	1
10	– 4	16
15	1	1
20	6	36
8	– 6	36
10	– 4	16
25	11	121
9	– 5	25
		$\sum x^2 = 252$

Using the formula:

Standard Deviation = $\sigma = \sqrt{\frac{\sum x^2}{N}} = \frac{252}{8}$

$$= \sqrt{31.8} \quad = 5.64$$

We can also compute standard deviation directly from the raw scores without using the deviation with the help of the formula:

Standard Deviation = $\sigma = \frac{\sqrt{N\sum x^2 - (\sum x)^2}}{N}$

in which

x = Raw score

N = The number of scores in the distribution.

x	x^2
15	225
10	100
15	225
20	400
8	64
10	100
25	625
9	81
$\sum x = 112$	$\sum x^2 = 1820$

Standard Deviation = $\sigma = \dfrac{\sqrt{N\sum x^2 - \left(\sum x\right)^2}}{N}$

$$= \frac{\sqrt{(8)(1820) - (112)^2}}{8}$$

$$= \frac{\sqrt{14560 - 12544}}{8}$$

$$= \frac{\sqrt{2016}}{8}$$

$$= \frac{44.90}{8}$$

$$= 5.612$$

The formula for computing standard deviation for the grouped data is:

$$S.D. = \sigma = \frac{i}{N}\sqrt{N\sum fx'^2 - \left(\sum fx'\right)^2}$$

in which

i = length of the class interval

N = total number of measures or scores

x = deviation of the raw score from the assumed mean divided by the length of the class interval.

To illustrate the use of this formula let us consider the data used for calculating S.D.

Scores	f	Mid Point (x)	x'	fx'	fx²
52 - 55	1	53.5	4	4	16
48 - 51	0	49.5	3	0	0
44 - 47	5	45.5	2	10	20
40 - 43	10	41.5	1	10	10
36 - 39	20	AM $\boxed{37.5}$	0	0	0
32 - 35	12	33.5	–1	–12	12
28 - 31	8	29.5	–2	–16	32
24 - 27	2	25.5	–3	–6	18
20 - 23	3	21.5	–4	–12	48
16 - 19	4	17.5	–5	–20	100
	N = 65			$\sum fx = -42$	$\sum fx^2 = 256$

Using the formula:

$$S.D. = \sigma = \frac{i}{N}\sqrt{\sum fx'^2 - \left(\sum fx'\right)^2}$$

$$= \frac{4}{65}\sqrt{(65)(256)-(-42)^2}$$

$$= \frac{4}{65}\sqrt{16640-1764}$$

$$= \frac{4}{65}\sqrt{14876}$$

$$= 7.51$$

Selection and Application of Standard Deviation (σ)

Standard deviation is used when:

i) the statistics having greatest stability is sought;

ii) extreme deviations exercise a proportionally greater effect upon the variability;

iii) co-efficient of correlation and other statistics are subsequently computed.

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

3. i) List the situations when we compute Range and Average Deviation.

....

....

....

ii) Mention the uses of quartile deviation.

....

....

....

iii) Compute Standard Deviation for the following data:

18, 25, 21, 19, 27, 31, 22, 25, 28, 20

....

....

....

14.5.3 Normal Probability Curve

The normal probability curve is based upon the law of probability or probable occurrence of certain events The 'probability' of a given event is defined as the expected frequency of occu ıce of this event among events of a like sort. It is expressed as a 'ratio'. For example, the probability of an unbiased coin falling heads is ½ and that of a dice showing a three-sport is 1/6. When any set of observations confirms to this mathematical form, it can be represented by a bell-shaped curve with the following characteristics.

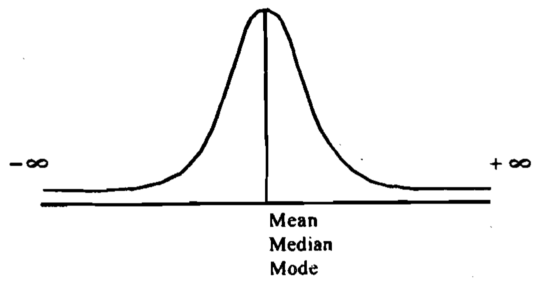

Fig. 14.5 : Normal Probability Curve

1. The curve is symmetrical around its vertical axis. It implies that the size, shape and slope of the curve on one side of the curve is identical to that on the other side of the curve.

2. The values of mean, mode and median computed for a distribution following this curve always coincide and have the same value.

3. The height of the vertical line called ordinate is maximum at the mean and in the unit normal curve is equal to 0.3989.

4. The curve has no boundaries in either direction, for the curve never touches the base line, no matter how far it is extended i.e. the curve is asymptotic and extends from $-\infty$ (minus infinity) to $+\infty$ (plus infinity).

5. The data 'cluster' around the mean. The percentages in a given standard deviation are greatest around the Mean and decrease as one moves away from the Mean.

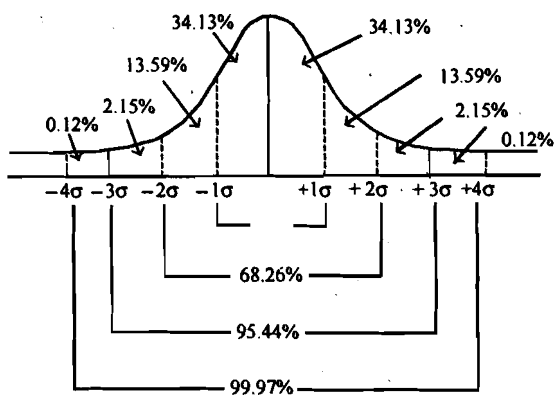

Fig. 14.6

The percentage area around the Mean are:

i) Mean $\pm$ 1.00 σ 34.13%

ii) $\pm$ 1.00 to $\pm$ 2.00 σ 13.59 %

iii) + 2.00 to $\pm$ 3.00 σ 2.15%

This implies that 68.26% of the total area of the curve falls between the limits Mean + 1σ and Mean –1; 95.44 per cent of the total area of the curve falls between limits Mean + 2 σ and Mean – 2 σ and 99.73 per cent of the total area of the curve falls between Mean + 3 σ and Mean – 3 σ.

The equation of the normal curve is

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

in which

x = scores, expressed as deviations from the mean

y = height of the curve above the horizontal axis i.e. the frequency of a given x value

N = number of cases in the sample

π = 3.1416

e = 2.7183

When N and σ are known, it is possible from the equation to compute: (i) the frequency (y) of a given value x; and (ii) the number or percentage between two points, or above or below a given value point in the distribution.

However, these calculations are not needed as Normal Probability Tables are available from which these values can be readily obtained.

Applications of the Normal Curve

There are a number of applications of the normal curve in the field of educational research.

1. When the measures or scores are normally or nearly normally distributed, the Normal Probability Table is useful: (i) to find the percentage of cases, or the number when N is known, that fall between the mean and a given σ distance from the mean; (ii) the percentage of the total area included between the mean and a given σ distance from the mean.

2. The normal curve is used to convert a raw score into standard score. Suppose the student A gets a score of 50 in an achievement test in mathematics which are administered to random sample of 300 students. The mean and standard deviation of the test scores, assumed to be normally distributed, for the sample are 60 and 6 respectively. The standard score or sigma score of the student A is :

$$Z = \frac{X-M}{\sigma} = \frac{50-60}{6} = \frac{-10}{6} = -1.67$$

indicating that the score of 50 is 1.67 standard deviation below the Mean.

3. Normal curve is useful in calculating the percentile rank of scores in a normal distribution.

4. For normalising a given frequency distribution, the normal curve is used.

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

4. i) What percentage of cases in a normal probability curve lie between M + 2σ and M − 2σ ?

....

ii) State two characteristics of a normal probability curve.

....

....

14.5.4 Measures of Relative Positions

When we administer a test on a sample of students drawn randomly from a population, the scores obtained are raw. A raw score, taken by itself, has no meaning. It gets meaning only by comparison with some reference group or groups. We generally use the following measures for making comparisons.

1. Sigma Score (Z)
2. T Score
3. Percentile Rank

Sigma Score (Z)

In describing a score in a distribution, its deviation from the mean is more meaningful than the score itself. The unit of measurement is the standard deviation. Mathematically, it is expressed as;

$$Z = \frac{X - X}{\sigma} \text{ or } \frac{x}{\sigma}$$

in which

X = raw score

X = mean

σ = standard deviation

x = (X – X) score deviation from the mean.

To illustrate the concept, let us consider the following data:

Test 1	Test 2
X = 77	X = 64
X = 84	X = 58
$\sigma = 4$	$\sigma = 5$
$Z = \frac{77-84}{4} = \frac{-7}{4} = -1.75$	$Z = \frac{64-58}{5} = \frac{6}{5} = +1.20$

The raw score of 77 in Test 1. may be expressed as a sigma score of –1.75, indicating that 77 is 1.75 standard deviation below the mean. The raw score of 64 in Test 2 may be expressed as a sigma score of + 1.20, indicating that 64 is 1.20 standard deviation above the mean.

On sigma scale, the mean of any distribution is converted to zero and the standard deviation is equal to 1. Thus a sigma score is useful in making realistic comparison of scores and thus provides a basis for equal weighting of the scores.

Suppose a teacher wants to determine a students' equally weighted average (mean) achievement on a physics test and on a chemistry test on the basis of the following data:

Subject	Test Score	Mean	Highest Possible Score	Standard Deviation
Physics	50	60	80	6
Chemistry	65	92	90	15

It is evident that the mean of the two raw scores obtained by the student in Physics and Chemistry tests would not provide a valid picture of the student's performance because the mean would be weighted more in favour of the chemistry test score. The conversion of each test score to a sigma score makes them equally weighted and comparable as both the raw test scores have been expressed on a scale with a mean of zero and standard deviation of one.

$$Z = \frac{Z-O}{1} = \frac{x-x}{\sigma}$$

$$\text{Physics Z Score} = \frac{50-60}{6} = \frac{-10}{6} = -1.67$$

$$\text{Chemistry Z Score} = \frac{65-92}{15} = \frac{-27}{15} = -1.80$$

The teacher, on an equally weighted basis, may conclude that the performance of the student was fairly consistent, i.e., 1.67 standard deviation below the mean in physics and 1.80 standard deviation below the mean in chemistry.

In the application of normal probability curve, it was explained that the curve describes the percentage of area lying between the mean and successive deviation units. The sigma or Z score can be used in hypothesis testing and determination of percentile ranks.

T Score (T)

The value sigma (Z) score in most of the cases is expressed as scores with decimals or negatives. In order to avoid this type of situation, another version of a standard score, the T score, has been devised. It is expressed as:

$$T = 50 + 10\frac{x-x}{\sigma} \text{ or } 50 + 10Z$$

Using the scores in the previous example, we can compute the physics and chemistry T scores as:

Physics T = 50 + 10 (– 1.67) = 33.30

Chemistry T = 50 + 10 (– 1.80) = 32.00

T scores are always rounded to the nearest whole number. Thus the T scores of the student in Physics and Chemistry are 33 and 32 respectively.

Percentile Rank

The percentile rank is the point below which a given percentage of scores fall. For example, if the 80th percentile is a score of 60, 80 per cent of the scores fall below 60. The median is the 50th percentile rank, for 50 per cent of the scores fall below it.

The formula for converting a given rank into percentile ranks is:

$$\text{Percentile rank} = 100 - \left[\frac{100\,RK - 50}{N}\right]$$

in which RK = rank from the top.

For example, a student A ranks 6th in a class of 150 students. This means 5 students rank above the student A, 144 below him. The percentile rank of the student A is:

$$= 100 - \left[\frac{100 \times 6 - 50}{150}\right]$$

$= 100 - (3.67)$

$= 100 - 4$

$= 96$

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

5. i) Define sigma score and T score.

....

....

ii) Shyam ranks twenty seventh in his class of 139 students. Find his percentile rank.

....

iii) Find the sigma score of a student whose raw score in a test of Hindi is 76. The mean of the Hindi scores of the students in his class is 82 and standard deviation 4. What will be his T score.

....

....

14.5.5 Measures of Relationship

We have so far discussed the statistical description of a single variable. Now we shall study the problem of describing the degree of simultaneous variation of two variables. The data in which we obtain measures of two variables for each individual is called a bivariate data. For example, we get bivariate data if we have scores of the tests of Mathematics and Hindi for a group of school children. The essential feature of the bivariate data is that one measure can be paired with another measure for each member of the group.

When we study bivariate data we may like to know the degree of relationship between variables of such data. This degree of relationship between variables is known as 'correlation' which is represented by the co-efficient of correlation. This co-efficient may be identified by either the r, the Greek symbol rho (r), or other symbols depending upon the data distributions and the way the co-efficient has been calculated.

Students who obtain high scores on an intelligence test tend to get high scores in an achievement test in mathematics; whereas those with low scores on an intelligence test tend to score low in mathematics. When this type of relationship is obtained the two variables are said to be positively related.

When the increase in one variable is associated with decrease in the other variable, the variables are said to be negatively related.

When the relationship between two sets of variables is a pure chance relationship, we say there is no correlation.

The intensity or degree of linear relationship is represented quantitatively by co-efficient of correlation. Its value ranges from –1.00 to +1.00. A value of –1.00 indicates a perfect relationship between the two variables and +1.00 describes perfect positive relationship. A zero value indicates complete lack of correlation between the two variables. The sign (– or +) indicates the **direction** of the relationship and the numerical value its **magnitude** or strength.

Methods of Correlation

There are various methods of correlation. Their use is relative to the situation and type of data. We may have data in scores. There are many situations in which the researcher does not have scores and has to deal with data in which differences in a given attribute or characteristic can be expressed only by ranks or classifying an individual into one of several descriptive categories. In this unit we will discuss two methods of correlation, namely, product moment correlation and rank order correlation.

Product Moment Correlation (r)

In some situations the data for two variables x and y are expressed in interval or ratio level of measurement and the distribution of these variables have a linear relationship. Moreover, the distributions of the variables are uni-modal and their variances are approximately equal. In such situations we make use of product moment method of correlation. It is also called Pearson's r.

1. Calculation of Pearson's r from Ungrouped Data

When the 'N' (size of the sample or group) is small or raw scores are small numbers, there is no need of grouping the data. The formula for computing 'r' is:

$$rxy = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}}$$

in which

x = deviation of x measures from the assumed mean

y = deviation of y measures from the assumed mean

To illustrate the use of this formula, let us compute product-moment r form the following data in the two variables X and Y for 8 students.

X: 15, 18, 22, 17, 19, 20, 16, 21

Y: 40, 42, 50, 45, 43, 46, 41, 41

For computing the values of $\sum x, \sum y, \sum x^2, \sum y^2$ and $\sum xy$, we proceed as under:

X	Y	x	y	x^2	y^2	xy
15	40	−2	−5	4	25	10
18	42	+1	−3	1	9	−3
22	50	+5	+5	25	25	25
$\boxed{17\ AM}$	$\boxed{45\ AM}$	0	0	0	0	0
19	43	+2	−2	4	4	−4
20	46	+3	+1	9	1	3
16	41	−1	−4	1	16	4
21	41	+4	−4	16	16	−16
		$\sum x = +12$	$\sum y = -12$	$\sum x^2 = 60$	$\sum y^2 = 96$	$\sum xy = 19$

$$rxy = \frac{N\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[N\sum x^2 - \left(\sum x\right)^2\right]\left[N\sum y^2 - \left(\sum y\right)^2\right]}}$$

$$= \frac{(8)(19)-(12)(-12)}{\left[(8)(60)-(12)^2\right]\left[(8)(96)-(-12)^2\right]}$$

$$= \frac{152+144}{\sqrt{[480-144][768-144]}} = \frac{296}{\sqrt{(336)(624)}}$$

$$= \frac{296}{\sqrt{209664}} = \frac{296}{457.891} = 0.646$$

$$= 0.65$$

2. Calculation of Pearson's r from Grouped Data

When N is large or even moderate in size, the best procedure is to group data in both variables X and Y in the form of a scattergram. The Pearson's r is computed by using the following formula:

$$r = \frac{N\sum fxy - \left(\sum fx\right)\left(\sum fy\right)}{\sqrt{\left[N\sum fx^2 - \left(\sum fx\right)^2\right]\left[N\sum fy^2 - \left(\sum fy\right)^2\right]}}$$

To illustrate the use of this formula, let us consider the data in the following scattergram:

		Mathematics Test Scores (X)											
		12-13	14-15	16-17	18-19	20-21	22-23	24-25	f	y	fy	fy²	fxy
Language Test Scores (Y)	35-37					0 1 0		6 1 6	2	+3	6	18	6
	32-34					0 6 0	2 3 6		9	+2	18	36	6
	29-31		–3 1 –3	–2 2 –4	–1 6 –6	0 8 0	1 1 1		18	+1	18	18	–12
	26-28		0 4 0	0 4 0	6 0	0 11 0	0 4 0	0 1 0	30	0	0	0	0
	23-25	4 2 8	3 1 3	2 6 12	1 5 5	0 4 0	–1 1 –1		19	–1	–19	19	27
	20-22	6 3 24	43 1 4	2 1 2					7	–2	–14	28	42
	f	5	8	13	18	30	9	2	85		9	119	69
	x	–4	–3	–2	–1	0	+1	+2					
	fx	–20	–24	–26	–18	0	9	4	–75				
	fx²	80	72	52	18	0	9	8	239				
	fxy	32	12	12	1	0	6	6	62				

Using the formula:

$$r = \frac{N\sum fxy - \left(\sum fx\right)\left(\sum fy\right)}{\sqrt{\left[N\sum fx^2 - \left(\sum fx\right)^2\right]\left[N\sum fy^2 - \left(\sum fy\right)^2\right]}}$$

$$= \frac{(85)(69)-(-75)(9)}{\sqrt{\left[(85)(239)-(-75)^2\right]\left[(85)(119)-\left(-9^2\right)\right]}}$$

$= 0.54$

The computation for the values of Σfx, Σfx^2, Σfy, Σfy^2, and Σfxy, is done using the following steps:

Step 1

The distribution of language test scores for 85 students is found in the f column at the right of the scattergram. Assume a mean for the distribution of scores for a language test (the mid point of that interval which contains the largest frequency), and draw bold lines to mark off the row in which the assumed mean falls. Here the mean for the language test scores has been taken at 27 (mid-point of interval 26-28) and y's (deviations from the assumed mean) have been taken from this point. Fill in the values of fy and fy^2 columns.

Step 2

The distribution of the mathematics test scores of 85 students is found in the f row at the bottom of the scattergram. Assume a mean for this distribution, and draw bold lines to designate the column under the assumed mean. The mean for the mathematics test scores is taken at 20.5 (mid-point of interval 20-21) and x's (deviations from the assumed mean) are taken from this point. Fill in the fx and fx^2 rows.

Step 3

The fxy for a cell is computed by multiplying the frequency given in the particular cell with the corresponding x and y.

Rank Order Correlation (ρ)

Rank order correlation is also known as the Spearman rank order co-efficient of correlation and is denoted by rho (ρ). This method is used when the data are available only in ordinal form of measurement (ranked) rather than in interval or ratio form, or if the number of observations or measures is small.

For computing the rank order correlation (r), the following formula is used:

$$\rho = 1 - \frac{6\sum D^2}{N\left(N^2-1\right)}$$

in which

D = difference between paired ranks

N = number of paired ranks

To illustrate the use of this formula, let us consider the following data:

X: 15, 18, 22, 17, 19, 20, 16, 14, 17, 22

Y: 41, 40, 42, 50, 45, 38, 46, 41, 40, 39

First, let us convert the data to ranks by assigning rank 1 to highest measure and so on. If a measure is repeated, the ranks are averaged. For example, in the variable X, 22 is the highest measure and is repeated. The ranks assigned are 1 and 2.

The average is $\frac{1+2}{2}=\frac{3}{2}=1.5$. The next measure is given the rank 3 and so on.

X	Y	Rx	Ry	D	D^2
15	41	9	6.5	2.5	6.25
18	40	5	8	–3	9.00
22	42	1.5	5	–3.5	12.25
17	50	6.5	1	5.5	30.25
19	45	4	4	0	0
20	38	3	10	–7	49.00
16	46	8	3	5	25.00
14	41	10	6.5	3.5	12.25
17	49	6.5	2	4.5	20.25
22	39	1.5	9	–7.5	56.25
					$\sum D^2 = 220.50$

Using the formula,

$$\rho = 1 - \frac{6\sum D^2}{N\left(N^2-1\right)}$$

$$= 1 - \frac{(6)(220.50)}{10(100-1)}$$

$$= 1 - \frac{1323}{10 \times 99}$$

$$= 1 - \frac{1323}{990}$$

$$= 1 - 1.336$$

$$= 0.336$$

$$= 0.34$$

Check Your Progress

Notes : a) Space is given below for your answer.

b) Compare your answer with the one given at the end of this unit.

6. i) List the situations in which product moment correlation is used.

....

....

....

ii) Compute product moment correlation for the following data:

X : 50, 54, 56, 59, 60, 62, 61, 65, 67, 71, 71, 74

Y : 22, 25, 34, 28, 26, 30, 32, 30, 28, 34, 36, 40

....

iii) Compute rank difference correlation for the data given under (ii). Compare the result with the one obtained using product moment correlation.

....

14.6 LET US SUM UP

1. In this unit, we listed the tools and techniques which generate data of a quantitative nature. We also discussed that quantitative data are either parametric or non-parametric.
2. The methods of presenting the data graphically using histogram, frequency polygon, cumulative frequency curve and Ogive were also explained.
3. The four measures of descriptive statistics, viz., measures of central tendency, variability, relative position and relationship were discussed.
4. The use, application and computation of mean, median and mode as the measures of central tendency were explained with the help of examples.
5. The measures of variability, namely, range, average deviation, quartile deviation and standard deviation were discussed. Their application, use and computation were explained with the help of examples.
6. The characteristics of normal probability curve alongwith its use in the educational research were discussed.
7. The raw test score, taken by itself, has no meaning. It is meaningful only by comparison with some reference group or groups. For which we make use of sigma score (Z), T score and percentile ranks. The method for computing sigma score, T score and percentile ranks have been explained in this unit.
8. The direction and magnitude of a relationship between the two variables is measured with the help of co-efficient of correlation. The product moment correlation co-efficient is computed when the variables (X and Y) are expressed in interval or ratio scales of measurement and the distribution of the variables have a linear relationship. In case of ordinal (ranked) data we compute rank difference correlation. The procedure for computing these co-efficients has been discussed in this unit.

14.7 UNIT-END ACTIVITIES

1. List the situations in which we obtain data of a quantitative nature.
2. What are the types of quantitative data? Illustrate with the help of examples.
3. Compute Median and Standard Deviation for the following data::

Scores	20-21	18-19	16-17	14-15	12-13	10-11	8-9
f	2	2	4	0	4	0	4

4. Compute Mean and Quartile Deviation for the following distribution:

Class Interval	f
195 - 199	1
190 - 194	2
185 - 189	4
180 - 184	5
175 - 179	8
170 - 174	10
165 - 169	6
160 - 164	4
155 - 159	4
150 - 154	2
145 - 159	3
140 - 144	1

5. Compute Mode, Median and Standard Deviation for the following scores:
   15, 20, 17, 18, 10, 11, 12, 17, 15, 9
6. Compute Product Moment Correlation (r) and Rank Difference Correlation (ρ) between the following sets of scores:

   X : 50, 42, 51, 26, 35, 42, 60, 41, 70, 55, 62, 38

   Y : 62, 40, 61, 35, 30, 52, 68, 51, 84, 63, 72, 50

14.8 POINTS FOR DISCUSSION

1. Discuss the uses of various measures of (i) central tendency and (ii) variability.
2. What are the characteristics and uses of a normal distribution curve?
3. What are the uses of sigma score, T score and percentile rank?

14.9 SUGGESTED READINGS

Best, John and James V. Kahn (1992): *Research in Education.* New Delhi: Prentice Hall of India Pvt. Ltd.

Garrett, H.E. (1962): *Statistics in Psychology and Education.* Bombay: Allied Pacific Pvt. Ltd.

Guilford, J.P. (1965): *Fundamental Statistics in Psychology and Education.* New York : McGraw Hill Book Company.

Koul, Lokesh (1997): *Methodology of Educational Research.* Third Revised Edition, New Delhi: Vikas Publishing House Pvt. Ltd.

14.10 ANSWERS TO CHECK YOUR PROGRESS

1. i) There are two types of quantitative data : parametric and non-parametric.

   ii) The data can be represented graphically using Histogram, Frequency Polygon, Frequency Cumulative Curve, and Ogive.

2. i) Mean

   ii) Mean = 8.88

   Mode = 5

   iii) Median = 87.5

3. i) Range is used: (a) when the data are too scant or too scattered; and (b) a knowledge of extreme scores or of total spread is wanted.

   We use average deviation when:

   a) it is desired to consider all deviations from the mean according to their size; and (b) extreme deviations would influence the standard deviation unduly.

   ii) Quartile deviation is used: (a) when the mean is the measure of central tendency; and (b) when the data are scattered or there are extreme score which would influence the standard deviation.

   iii) SD = 4.05.

4. i) 95.44%

   ii) a) The normal curve is bell shaped and symmetrical around its vertical axis called ordinate; and

   b) the curve is highest at the mean — the mean, median, and mode have the same value.

5. i) Sigma score is defined as = Z $\frac{X - X}{\sigma}$ or $\frac{X}{\sigma}$ in which x is the mean and σ is the standard deviation of the distribution.

   T score is defined as:

   $$T = 50 + 10\ \frac{x - x}{\sigma}$$
   $$= 50 + 10\ Z.$$

ii) percentile rank = 81

iii) sigma score = – 1.50

T score = 35

6. i) Product moment correlation is used when:

a) The data for two variables (X and Y say) are expressed in interval or ratio scale of measurement.

b) The distributions of the variables have a linear relationship.

c) The distributions of the variables are unimodal and their variances are approximately equal.

ii) $r = 0.78$

iii) $\rho = 0.78$