UNIT 7 STATISTICS

Structure

7.1 INTRODUCTION

The word 'statistics' appears to have been derived from the Latin word 'status' meaning a (political) state. In its origin, statistics was simply the study of "The political arrangement of the modern states of the known world". The description of states was at first verbal but later the increasing proportion of numerical data in the description gradually gave the word 'statistics'. The scope of statistics now includes collection of numerical data pertaining to almost every field; for this reason it is very useful in economics, sociology, business, education, agriculture, psychology, biology and related fields. It may be defined as a science which enables us to draw representative samples, analyse the data collected, interpret and make inferences.

Objectives

After studying this unit, you should be able to

- construct the frequency tables given a numerical data,
- define the measures of location namely mean and the median and analyse the information it conveys, and
- define the measures of dispersion namely the standard deviation, and the mean deviation and analyse the information it conveys.

7.2 STATISTICAL DATA AND VARIABLES AND UNITS OF OBSERVATIONS

7.2.1 Raw Data

Statistics is a collection of information in numerical terms. For example, marks obtained by the students of a class; monthly wages of workers in a factory; numbers indicating births, deaths and marriages in different states etc. are called

Statistical Data or Numerical Data and statistics is the science which deals with the collection, analysis and interpretation of statistical data.

The numerical data or information is collected in two ways. When information is collected in respect of every individual person or item, then the numerical data has been carried out by means of complete Enumeration or Census. But if information is collected only from a selected portion (or sample) of a given population, the procedure is called Sample Survey.

For example, during the census operation, the population of a country is enumerated and all citizens are included in such operation but while verifying the accuracy of entries in Books of Accounts, we check only a portion of entries. This is an example of Sample Survey.

In addition to these two methods, we also come across regular collection and recording of information in a routine manner for example the Railways keep a daily record of movement of passengers and goods, the income earned through fares and freights etc.

The information collected through censuses and surveys or in a routine manner is called raw data.

7.2.2 Variables of Observation

In a census, suppose for each person we have recorded the age and sex and whether the person belongs to a rural or urban area, we say that we have taken observations on three variables : age, sex and place of residence. The term variable stands for what is being observed. A variable is completely described by its descriptive name and the description of all the values it can possibly take. In the above example, the variable called 'place of residence' has two possible values as 'rural' or 'urban'. The value of this variable could also have been recorded as the name of the state, district, city, village etc. Thus, the variable 'place of residence' has the same name but the set of values may be different for them. We regard two variables to be different if their set of possible values are different even if they have the same name.

Qualitative and Quantitative Variables

Variables of observations with numbers as possible values are called quantitative variables and variables of observation with name of things, places etc. are called qualitative variables. For example in the example of census, the value of age variable were numbers so the age variable is a quantitative variable whereas the variable called 'place of residence' is a qualitative variable. A word of caution is necessary here. Suppose in the recording of the variable 'place of residence' we mark '1' if urban and '2' if rural but this does not make the variable a quantitative variable. A quantitative variable has not only its values recorded as numbers but they are really number on which arithmetic operations can be carried out. For example, age is a quantitative variable as sum, product, difference of age has a sense whereas the sum of rural or urban 'place of residence' has no sense.

Age, height, income of a worker etc. are examples of quantitative variables whereas variables such as sex, religion, caste etc. are qualitative variables.

7.2.3 Unit of Observation

Suppose in the census example we have recorded age, sex and place of residence of all the persons alive at the time of the census and in another example we have recorded the results of a particular examination, i.e. we have recorded the name of the students who appeared in that examination and marks obtained by each student.

The term unit of observation will be used to describe what the values of a variable are attached to. In the census example, the unit of observations are person alive at the time of census and to each unit of observation we recorded the value of three variables : age, sex and place of residence. In the example of results of a particular examination, the unit of observations would be the students who have appeared in the said examination and the variable of observation will be marks obtained by a particular student. Thus different variables of observation may be associated with the same unit of observation.

7.3 CONSTRUCTION OF FREQUENCY TABLES (OR FREQUENCY DISTRIBUTIONS) FROM RAW DATA

If a sample of a population contains a 'large' number of observations, the investigator has to devise methods to condense it and present it in the form of tables and charts to bring out its main characteristics. This is called **data presentation**.

Let us consider, for example, the marks obtained by 30 students of Class XI in a class test (out of 80 marks) in mathematics.

60, 49, 53, 57, 73, 62, 40, 39, 68, 55, 36, 61, 40, 31, 43, 41, 47, 52, 67, 44, 54, 52, 24, 38, 48, 46, 72, 46, 47, 49.

These observations constitute **raw data** or **ungrouped data**. What do these 30 numbers convey to us? Not much. We, therefore, would like to bring out certain features of this data. For instance, we could arrange numbers in ascending or descending order of magnitude. But, this method would involve difficulties when the number of observations is very large. So, what we generally do is, to condense the data into **classes** (or **groups**) as follows :

We find the difference between maximum and minimum observations. This difference is called **range** of the raw data. Then, we decide about the number of classes into which the raw data is to be grouped. Care should be taken that the classes cover the entire range and there is a class to include the least observation and also a class to include greatest observation. In general, we make sure that we have **not less than 5 or more than 15 classes**.

In the above example, the range is 73 – 24 = 49. So, it is convenient to have 10 classes, each of width (or size) 5. In general, the width of each class (or class-interval) is a convenient whole number immediately greater than the quotient obtained by dividing the range by the number of classes to be made. While setting up the class limits (i.e. the maximum and minimum numbers which can be put in the class), the following rules may be observed :

(i) Classes should be non-overlapping.

(ii) The classes should be continuous without any gap.

(iii) As far as possible, the classes should be of the same size.

(iv) Classes such as, less than 3 or more than 8, i.e. open-ended classes should be avoided.

(v) The boundaries of each class should be determined in such a way that there is no ambiguity as to which class a particular observation of the data belongs.

In our example, we choose the classes as

24 – 29, 29 – 34, 34 – 39, . . . , 69 – 74.

Note that the class 24 – 29 will contain all observations which are greater than or equal to 24 but less than 29. The observation '29' will be put in the next class and so on.

24 is the **lower limit** of the class 24 – 29 and 29 is the **upper limit** of the class. The arithmetic average of the lower limit and the upper limit of a class is called the **class-mark** of that class. The class mark of the class 24 – 29 is $26.5\left(=\frac{24+29}{2}\right)$.

To prepare the frequency distribution table, we take each observation from the data, one at a time and put a tally mark (a right-handed dash) opposite the class in which the observation lies. For sake of convenience and symmetry, we record tally marks in bunches of five, the fifth one crossing the other four diagonally. The count of tally marks in a particular class is called **frequency** of that class and is recorded opposite the class next to the tally marks. It may be noted that the sum of all the frequencies is equal to the total number of observations in the raw data.

Table 7.1

Class-interval (Marks out of 80)	Tally Marks	Frequency
24 – 29	/	1
29 – 34	/	1
34 – 39	//	2
39 – 44	////	5
44 – 49	//// /	6
49 – 54	////	5
54 – 59	///	3
59 – 64	///	3
64 – 69	//	2
69 – 74	//	2
Total		30

Using the above steps, the **frequency distribution table** of the marks obtained by 30 students of class XI in a class test (out of a maximum of 80) in mathematics is as shown in Table 7.1.

The data in the above form are called **grouped data**. We have condensed 30 observations into ten classes and we observe from this data that

(i) There are 4 students (out of 30) who have secured less than 39 (of course, ≥ 24) marks.

(ii) Nearly half the students (16) have secured marks between 39 and 54.

(iii) Only two students have secured 69 or more than 69 (of course, ≤ 74) marks.

Often, we shall be interested in knowing the number of observations less than a particular number. For this purpose we add another column in the above frequency table. Opposite to each class we write in this column, the sum of frequencies of all the previous classes and that particular class. The new number we get is called the **cumulative frequency** of the class and the modified table is called **cumulative frequency table**.

Table 7.2

Class-interval (Marks out of 80)	Tally Marks	Frequency	Cumulative Frequency
24 – 29	/	1	1
29 – 34	/	1	2
34 – 39	//	2	4
39 – 44	~~////~~	5	9
44 – 49	//// /	6	15
49 – 54	////	5	20
54 – 59	///	3	23
59 – 64	///	3	26
64 – 69	//	2	28
69 - 74	//	2	30
Total		30	

Surely, the cumulative frequency of the last class is same as the total number of observations in the data. The cumulative frequency table in case of our example is the Table 7.2.

From this table we can say, at a glance, that 15 (our of 30) students have secured less than 49 marks and only 4 students have secured less than 39 marks.

In fact, there are two types of cumulative frequencies *upward* or *downwards*. A cumulative frequency distribution may be made on a '*less than*' basis or on a '*more than*' basis. Table 7.2 has been constructed on a 'less than' basis. For comparison, we give below, both the tables, one constructed on less than basis and the other on more than basis.

Table 7.3 (i)
Cumulative Frequency Table Constructed on Less than Basis

Score	Number of Students Scoring Less than the Indicated Score
24	0
29	1
34	2
39	4
44	9
49	15
54	20
59	23
64	26
69	28
Total	30

Table 7.3 (ii)
Cumulative Frequency Table Constructed on More than Basis

Score	Number of Students Scoring More than or Equal to the Indicated Score
24	30
29	29
34	28
39	26
44	21
49	15
54	10
59	7
64	4
69	2
Total	0

In practical situations, it is often desired to compare class frequencies in two or more distributions based upon a very different number of total items. This becomes very easy, if we transform the absolute frequencies into relative frequencies. The class frequencies expressed *relative* to the total frequency (total number of items) are called **percentage frequencies**. We arrive at percentage frequencies by dividing the frequencies in each class by the total number of items in the distribution and express the resulting fraction as percents. The following table illustrates the whole process.

Table 7.4

Class	Frequency	Cumulative Frequency
24 – 29	1	$\frac{1}{30} = 0.0333 = 3.33\%$
29 – 34	1	$\frac{1}{30} = 0.0333 = 3.33\%$
34 – 39	2	$\frac{2}{30} = 0.0667 = 6.67\%$
39 – 44	5	$\frac{5}{30} = 0.1667 = 16.67\%$
44 – 49	6	$\frac{6}{30} = 0.2000 = 20.00\%$
49 – 54	5	$\frac{5}{30} = 0.1667 = 16.67\%$
54 – 59	3	$\frac{3}{30} = 0.1000 = 10.00\%$
59 – 64	3	$\frac{3}{30} = 0.1000 = 10.00\%$
64 – 69	2	$\frac{2}{30} = 0.0667 = 6.67\%$
69 – 74	2	$\frac{2}{30} = 0.0667 = 6.67\%$
Total	30	

Sometimes, the classes are not continuous, i.e. the upper class limit of a class is not equal to the lower class-limit of the next class; then we make the classes continuous by decreasing lower limit of each class by 0.5 and increasing upper limit by 0.5. Consider the example of distribution of ages (in years) of primary school teachers in a Tehsil :

Table 7.5(a)

Age (in Years)	**Number of Primary Teachers**
21 – 25	20
26 – 30	26
31 – 35	34
36 – 40	47
41 – 45	15
46 – 50	7
51 – 55	3
Total	152

We shall modify the above frequency distribution table as follows :

Table 7.5(b)

Age (in Years)	**Number of Primary Teachers**
20.5 – 25.5	20
25.5 – 30.5	26
30.5 – 35.5	34
35.5 – 40.5	47
40.5 – 45.5	15
45.5 – 50.5	7
50.5 – 55.5	3
Total	152

Example 7.1

Given the following distribution of weekly wage rates of a selected group of junior and senior typists in a private factory; compare the two distributions by constructing relative frequency distributions :

Weekly Wage Rate (in Rupees)	**Number of Junior Typists**	**Number of Senior Typists**
100-200	12	
200-300	32	
300-400	38	16
400-500	30	19
500-600	13	10
600-700		3
700-800		2

Solution

We construct the table of relative frequencies as follows :

Table 7.6

Weekly Wage Rate (in Rupees)	Number of Typists		Percent of Total Number of Typists	
	Junior	Senior	Junior	Senior
100-200	12		$\frac{1200}{125} = 9.6$	
200-300	32		$\frac{3200}{125} = 25.6$	
300-400	38	16	$\frac{3800}{125} = 30.4$	$\frac{1600}{50} = 32.0$
400-500	30	19	$\frac{3000}{125} = 24.0$	$\frac{1900}{50} = 38.0$
500-600	13	10	$\frac{1300}{125} = 10.4$	$\frac{1000}{50} = 20.0$
600-700		3		$\frac{300}{50} = 6.0$
700-800		2		$\frac{200}{50} = 4.0$
Total	125	50	100	100

7.4 GRAPHICAL PRESENTATION OF FREQUENCY DISTRIBUTIONS

The main features of frequency distribution are conveniently communicated by representing the frequency distribution in the term of a diagram, since a diagram is more easily and more quickly understood than a collection of numbers. Diagrammatic presentation is particularly useful when the number of classes or the class frequency distribution is large.

There are various methods of graphical presentation of frequency distribution which are in use. We shall discuss only two of them namely the bar diagram and the pie diagram.

The Bar Diagram

To draw the bar diagram of a frequency distribution, we mark **equal lengths** on horizontal axis for representing difference classes. These lengths must be equal even if the classes are of unequal size. On each of these lengths (on the horizontal axis), we erect a *rectangle* whose height is proportional to the frequency of the class represented by its base. This means that the heights of a rectangle represents the relative frequency of the class represented by its base. Thus, we shall get '*bars*' and hence the name bar diagram. The bar diagram of frequency distribution of Table 7.1 (using the Table 7.4) is shown in Figure 7.1.

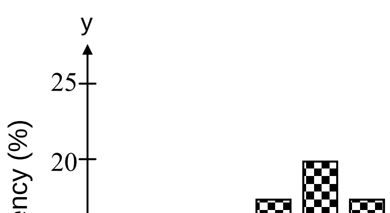

Figure 7.1

The bar diagram (or bar chart) is particularly useful when it is desired to compare two different frequency distributions. This is achieved by drawing two bars (for the same class) adjacent to each other, one for the first distribution and the other for the second, their heights representing the relative frequencies in their respective distribution. The bar diagram for studying the comparison of two distribution of Table 7.6 (Example 7.1) is shown in Figure 7.2.

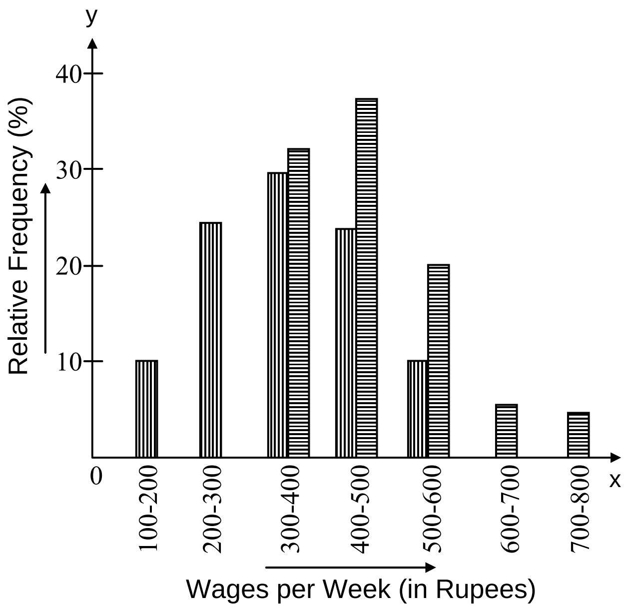

Figure 7.2

The Pie Diagram

The pie diagram (or pie chart) is used to represent relative frequencies only. The relative frequencies of different classes are represented by sectors of a circle. The angle of each sector is proportional to the relative frequency of the particular class represented by the sector. The angle of a sector representing a particular class is calculated by multiplying 360^o with the relative frequency of that class. The angles of sectors representing the frequency distribution of Table 7.1 is calculated in the following table using Table 7.4.

The pie diagram of the distribution (Table 7.7) is shown in Figure 7.3.

Table 7.7

Class	Percentage Frequency	Angle of the Sector Representing the Class
24 – 29	3.33	$\frac{3.33 \times 360^o}{100} = 12^o$
29 – 34	3.33	$\frac{3.33 \times 360^o}{100} = 12^o$
34 – 39	6.67	$\frac{6.67 \times 360^o}{100} = 24^o$
39 – 44	16.67	$\frac{16.67 \times 360^o}{100} = 60^o$
44 – 49	20.00	$\frac{20 \times 360^o}{100} = 72^o$
49 – 54	16.67	$\frac{16.67 \times 360^o}{100} = 60^o$
54 – 59	10.00	$\frac{10 \times 360^o}{100} = 36^o$
59 – 64	10.00	$\frac{10 \times 360^o}{100} = 36^o$
64 – 69	6.67	$\frac{6.67 \times 360^o}{100} = 24^o$
69 – 74	6.67	$\frac{6.67 \times 360^o}{100} = 24^o$
Total	100	360^o

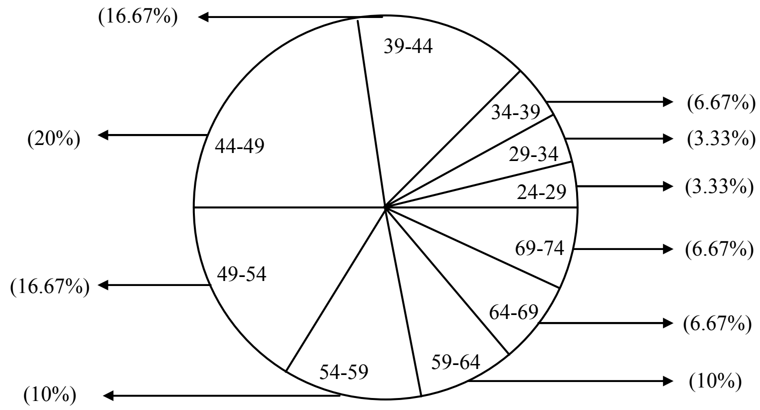

Figure 7.3 : Pie Diagram Showing the Frequency Distribution of Table 7.7

7.5 MEASURES OF LOCATION AND DISPERSION

So far we have discussed the presentation of raw data in a form suitable for communicating the information contained in it and have studied the use of frequency tables. In case of quantitative variables the information contained in the raw data can be summarized by means of a few numerical values. Such a summary is partly provided by what are called measures of location and measures of dispersion.

7.5.1 Measures of Location

Definition 1

The Arithmetic Mean of the values $x_1, x_2, \ldots, x_n$ of a variable recorded for n units of observation is defined as $\dfrac{x_1 + x_2 + \ldots + x_n}{n}$ *and is denoted by* $\bar{x}$.

$$\therefore \qquad \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

From the definition of $\bar{x}$, we have

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

i.e. $\quad (x_1 - \bar{x}) + (x_2 - \bar{x}) + \ldots + (x_n - \bar{x}) = 0$

Thus some values of $x_i - \bar{x}$ must be positive and some negative so that the sum is zero. If we add all the positive $x_i - \bar{x}$ and all the negative $x_i - \bar{x}$, these two sums will have the same value but opposite in sign so that their algebraic sum is zero.

Hence we say that $\bar{x}$ lies at the centre of all the observations. Unless otherwise stated, the word mean denotes the arithmetic mean.

For calculating the mean of the group data, suppose the observed values in the different classes of the frequency table are $y_1, y_2, \ldots, y_k$ and the frequencies in each class are $f_1, f_2, \ldots, f_k$, then mean will be given by

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i\, y_i}{\sum_{i=1}^{k} f_i}$$

In some cases the classes are no longer defined by single values, each class consists of many values of the variable. In such a case the method is to replace all the observed values belonging to a class by the mid value of that class and then use the mid value to determine the mean. In this method one of the class marks (preferably near the middle) is designated as 'a' (called the assumed mean) and the deviation $d_i = y_i - a$ are calculated for each class.

The arithmetic mean is then

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i\, y_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i\,(d_i + a)}{\sum_{i=1}^{k} f_i}$$

$$\bar{x} = \frac{\sum_{i=1}^{k} (f_i\, d_i + f_i\, a)}{\sum_{i=1}^{k} f_i} = a + \frac{1}{n}\sum_{i=1}^{k} f_i\, d_i$$

where *k* is the number of classes and *n*, the number of observations, i.e. $$n = \sum_{i=1}^{k} f_i$$

i.e. $$\bar{x} = a + \frac{1}{n}\sum_{i=1}^{k} f_i\, d_i$$

Since in most of the problems, the width of all the classes is same, we can further simplify the calculations of the mean of the grouped data by calculating the mean of μ_i's (denoted by Mu), where $\mu_i = \frac{y_i - a}{c}$, c is the width of each class and a is an assumed mean.

Now, $$\bar{x} = a + \frac{1}{n}\sum_{i=1}^{k} f_i\, d_i$$

$$= a + \frac{1}{n}\sum_{i=1}^{k} c f_i \,.\, \frac{d_i}{c}$$

$$= a + \frac{c}{n}\sum_{i=1}^{k} f_i \,.\, \frac{d_i}{c}$$

$$= a + c\left[\frac{1}{n}\sum_{i=1}^{k} f_i\, \mu_i\right]$$

$$= \left(\because \frac{d_i}{c} = \frac{y_i - a}{c} = \mu_i\right)$$

$$= a + c\,\mu$$

This method is known as the step deviation method.

Remark

The step deviation method works equally well if the classes are of unequal width. The only care to be taken is that the number c (which may not be the class size) should be a divisor of each class size.

Example 7.2

The marks obtained by 20 students in a test were 13, 17, 11, 5, 18, 16, 11, 14, 13, 12, 18, 11, 9, 6, 8, 17, 21, 22, 7, 6.

Find (i) The mean marks per student.

(ii) The mean marks per student when marks of each student are increased by 5.

(iii) The mean marks per student when the marks of each student are doubled.

Solution

The sum of marks of all the students = 13 + 17 + 11 + 5 + 18 + 16 + 11 + 14 + 13 + 12 + 18 + 11 + 9 + 6 + 8 + 17 + 21 + 22 + 7 + 6 = 255.

(i) Mean = $\frac{\sum x_i}{\text{Number of students}} = \frac{255}{20} = 12.75$

(ii) When marks of each student is increased by 5, then sum of their marks is increased by 20 × 5 = 100, i.e. sum of marks = 255 + 100 = 355.

$$\text{Mean} = \frac{355}{20} = 17.75$$

Thus we see that mean is also increased by 5.

(iii) When marks of each student is doubled, the sum of their marks is also doubled, i.e. the sum of the marks = 255 × 2 = 510.

$$\text{Mean} = \frac{510}{20} = 25.5 = 2 \times 12.75$$

i.e. the mean has also doubled.

Example 7.3

The following table shows the gain in weight by 25 children in a year

Gain in Weight (in kg)	2	2.5	3	3.5	4	4.5	5	5.5	6
No. of Children	2	3	4	2	5	1	4	3	1

Find the mean of gain in weight.

Solution

For calculation of the mean, we construct the table

y_i	f_i	$f_i y_i$
2.0	2	4.0
2.5	3	7.5
3.0	4	12.0
3.5	2	7.0
4.0	5	20.0
4.5	1	4.5
5.0	4	20.0
5.5	3	16.5
6.0	1	6.0
Total	25	97.5

$$\text{Mean} = \frac{\sum f_i\, y_i}{\sum f_i} = \frac{97.5}{25} = 3.9$$

So mean of gain in weight = 3.9.

Example 7.4

The weekly observations on cost of living index in a certain city for a particular year are

Cost of Living Index	140-150	150-160	160-170	170-180	180-190	190-200

Number of Weeks	5	10	20	9	6	2

Compute the average weekly cost of Living Index.

Solution

We shall use the deviation method by taking $a = 165$ the assumed mean.

Class	Class Mark (y_i)	Frequency (f_i)	$d_i = y_i - a$	$f_i\, d_i$
140 – 150	145	5	– 20	– 100
150 – 160	155	10	– 10	– 100
160 – 170	165	20	0	+ 0
170 – 180	175	9	10	+ 90
180 – 190	185	6	20	120
190 – 200	195	2	30	60
Total		52		70

$$\therefore \quad \text{Mean } = a + \frac{\Sigma f_i\, d_i}{\Sigma f_i} = 165 + \frac{70}{52}$$

$$= 165 + 1.35 = 166.35$$

Example 7.5

The ages of all the male inhabitants of a village were received and the following frequencies distribution was obtained.

Age (years)	0-5	5-10	10-20	20-30	30-40	40-50	50-60	60-80
Number of Persons	12	18	16	19	14	11	4	3

Obtain the mean age per male inhabitant.

Solution

We construct the following table, taking assuming mean $a = 25$, $c = 2.5$.

Class	Class Mark (y_i)	Frequency ($f_i i$)	$\mu_i = \frac{y_i - a}{c}$	$f_i\, \mu_i$
0 – 5	2.5	12	– 9	– 108
5 – 10	7.5	18	– 7	– 126
10 – 20	15	16	– 4	– 64
20 – 30	25	19	0	0
30 – 40	35	14	4	56
40 – 50	45	11	8	88
40 – 60	55	4	12	48
60 – 80	70	3	18	54
Total		97		– 52

$$\therefore \quad \text{Mean} = a + c \times \frac{\Sigma f_i\, \mu_i}{\Sigma f_i}$$

$$\therefore \qquad \text{Mean} = 25 + 2.5 \times \frac{-52}{97}$$

$$= 25 - \frac{130}{97} = 25 - 1.34$$

$$= 23.66 \text{ nearly.}$$

Definition 2 : Median

Median is defined as the central value of a set of observations. It divides the whole series of observations into two parts. If there are n observations $x_1, x_2, x_3, \ldots, x_n$*, then*

$$Median = \begin{cases} \dfrac{n+1}{2} th\ observation, & if\ n\ is\ odd \\ \dfrac{\dfrac{n}{2} th \text{ observation} + \left(\dfrac{n}{2}+1\right) th\ observation}{2}, & if\ n\ is\ even \end{cases}$$

where $x_1, x_2, x_3, \ldots, x_n$ *are either in ascending or in descending order.*

For example, the median of 1, 2, 4, 8, 9, 10, 12 is $\frac{7+1}{2}$ *th* term

$$= 4\ th \text{ term} = 8$$

and the median of 3, 5, 8, 9, 12, 15, 16, 18, 19, 23

$$= \frac{1}{2}\left[10th \text{ term} + \left(\frac{10}{2}+1\right) th \text{ term}\right]$$

$$= \frac{1}{2}(5th \text{ term} + 6th \text{ term})$$

$$= \frac{1}{2}(12 + 15) = 13.5 .$$

Alternatively, the median of a set of n observations is a number M which satisfies the conditions :

(i) *(Number of observations* $\geq M$*)* $\geq \frac{n}{2}$.

(ii) *(Number of observations* $\leq M$*)* $\geq \frac{n}{2}$.

Consider the set 8, 9, 5, 3, 12, 18, 15, 16, 23, 19 of 10 numbers. The numbers arranged in ascending order are 3, 5, 8, 9, 12, 15, 16, 18, 19, 23.

If M is the median, then there should be atleast $5\left(=\frac{10}{2}\right)$ numbers greater than or equal to M; this suggests that $M \leq 15$. Also, there should be atleast 5 numbers less than or equal to M; this suggests that $M \geq 12$. This means that any number between 12 and 15 can be taken as the median. Conventionally, we take the mean of 12 and 15 as the median.

We observe from the above definition of median that unlike the arithmetic mean, median of a set of observations may not be unique. However, the methods given here are conventional and these determine the median without ambiguity.

In case of *grouped data*, the median is calculated by formula :

$$\text{Median} = l + \frac{\frac{n}{2} - C_{f-1}}{f_m} \times c_m,$$

where l = lower limit of the median class,

f_m = frequency of the median class,

C_{f-1} = cumulative frequency of the class preceding to the median class,

c_m = width of the median class, and

n = sum of all the frequencies, i.e. total number of observations.

The **median class** being the class which contains the $\frac{n}{2}th$ observation.

Remark

The above method of finding the median in case of grouped data works well even if the classes are of unequal widths. Of course, we assume that the **frequency of the median class is uniformly distributed over the whole class** and the classes are without gaps and they have been arranged according to the ascending order of the variable.

Example 7.6

The number of students absent in a school was recorded every day for 147 days and the new data was presented in the form of the frequency table given below :

Number of Students Absent	5	6	7	8	9	10	11	12	13	15	18	20
Number of Days	1	5	11	14	16	13	10	70	4	1	1	1

Obtain the median and describe the information conveyed by it.

Solution

Here, n = 147, and odd number, therefore the median $= \frac{n+1}{2}th$, i.e.

$\frac{147+1}{2}th$ observations, i.e. the median is 74 *th* observation. To find it, we construct the cumulative frequency table.

x_i	5	6	7	8	9	10	11	12	13	15	18	20
y_i	1	5	11	14	16	13	10	70	4	1	1	1
Cunulative Frequency	1	6	17	31	47	60	70	140	144	145	146	147

We notice that 74 *th* observation is 12.

($\because$ all observations from 71st upto 140 are equal, each being 12.)

This value of the median suggests that for half the number of days 12 or more than 12 students remained absent and on the other days 12 or less than 12 students remained absent.

Example 7.7

Calculate the mean and median of the following data :

Number of Workers	12	30	65	107	157	202	222	230
Wages per Week up to (Rs.)	15	30	45	60	75	90	105	120

Solution

In this case, we are given the cumulative frequencies. We construct the following table for mean and the median. Here, width of each class = 15. We shall use step deviation method taking a = 67.5.

Class	**Frequency** y_i	**Cumulative Frequency**	f_i	$\mu_i = \frac{y_i - a}{15}$	$f_i\,\mu_i$
0 – 15	7.5	12	12	– 4	– 48
15 – 30	22.5	30	18	– 3	– 54
30 – 45	37.5	65	35	– 2	– 70
45 – 60	52.5	107	42	– 1	– 42
60 – 74	67.5	157	50	0	0
75 – 90	82.5	202	45	1	45
90 – 105	97.5	222	20	2	40
105 – 120	112.5	230	8	3	24
Total			230		– 105

$\therefore$ The mean $= a + c \times \left(\frac{\Sigma f_i\,\mu_i}{\Sigma f_i}\right) = 67.5 + 15 \times \frac{-105}{230}$

$= 67.5 - 6.85 = 60.65$ nearly.

Here $\frac{n}{2} = \frac{230}{2} = 115$

$\therefore$ Median class is 60 – 75.

$$\therefore \quad \text{Median } = l + \frac{\frac{n}{2} - C_{f-1}}{f_m} \times c_m$$

$$= 60 + \frac{115 - 107}{50} \times 15$$

$$= 60 + 2.4 = 62.4$$

7.5.2 Measures of Dispersion

Suppose a cricket team to represent India was to be selected and all the members of the team have been selected except one. Two players X and Y are available and the last member to be selected has to be one of them. The managers look at the runs scored by these two batsmans in the last 5 matches which are as follows :

X	38, 70	48, 34	42, 55	63, 46	54, 44
Y	5, 11	8, 29	83, 104	20, 28	81, 123

The average score per inning is nearly 50 for both the players. We observe that the runs made by player X do not change much from inning to inning whereas Y's scores show great variation with very high scores in one inning and very low in another. We use the word dispersion and say that the runs scored by Y show a higher dispersion than the runs made by X. The mean of runs made by X is 49.4 and his scores are close to the mean score whereas the mean score of Y is 49.2 and his scores in different innings are not close to the mean score.

The measure of location, the mean and the median give us a central value around which the values of the variables are located but gives us no idea of how far these values are from the central value. The measure of dispersion which we are going to study now provides us this information.

The commonly used measure of dispersion are Standard Deviation (SD) and the Mean Deviation (MD). Standard deviation is the measure of dispersion about the mean and mean deviation is the measure of dispersion about the median.

Definition 3 : Mean Deviation

The mean deviation is the mean of the absolute differences of the values from the mean or median. Thus mean deviation (MD)

$$= \frac{1}{n} \Sigma f_i |x_i - A|$$

where A is either the mean or median. As the positive and negative differences leave equal effects, only the absolute value of differences is taken into account.

Now let us consider the mean deviation from the median.

Mean Deviation from the Median

Mean deviation from the median

$$\text{MD} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{\sum_{i=1}^{n} |d_i|}{n},$$

n being the total number of observations and $\bar{x}$ being the median. In case of a grouped data, the mean deviation about median

$$\text{MD} = \frac{\sum_{i=1}^{k} f_i |y_i - \bar{y}|}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i |d_i|}{\sum_{i=1}^{k} f_i}$$

where $d_i = y_i - \bar{y}$

Definition 4 : Standard Deviation

Let us consider a set of n observations $x_1, x_2, x_3, \ldots, x_n$. *We compute the sum S of squares of deviations of these observations from an arbitrary number a.*

So, $S = (x_1 - a)^2 + (x_2 - a)^2 + (x_3 - a)^2 + \ldots + (x_n - a)^2$

$$= \sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} [(x_i - \bar{x}) + (\bar{x} - a)]^2$$

where $\bar{x}$ is the mean of n observations in reference

$$= \sum_{i=1}^{n} [(x_i - \bar{x})^2 + (\bar{x} - a)^2 + 2(x_i - \bar{x})(\bar{x} - a)]$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - a)^2 + 2(\bar{x} - a) \sum_{i=1}^{n} (x_i - \bar{x})$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - a)^2$$

$$\left[\because \sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = 0\right]$$

Clearly, S is minimum when $\bar{x} - a = 0$, i.e. when $\bar{x} = a$, i.e. when the deviations are considered from the arithmetic mean.

In view of the above idea, *Karl Pearson introduced the concept of standard deviation*. It is most popular measure of dispersion. It is denoted by σ and is defined as

(i) $$\sigma = \left[\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}\right]^{\frac{1}{2}} \quad \ldots$$

(i)

in case of ungrouped data, and

(ii) $$\sigma = \left[\frac{\sum_{i=1}^{k} f_i (y_i - \bar{y})^2}{\sum_{i=1}^{k} f_i}\right]^{\frac{1}{2}}$$

in case of grouped data, it being assumed that frequency of a class is centred at the class mark.

Calculation for standard deviation can be more simplified if we take deviations of the variates (or class marks) from an assumed mean 'a'.

Let $d_i = x_i - a$, then $x_i = d_i + a$ . . . (ii)

From Eq. (i),

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i^2 + \bar{x}^2 - 2x_i \bar{x})$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 + \frac{1}{n}(n\bar{x}^2) - \frac{2\bar{x}}{n} \sum_{i=1}^{n} x_i$$

$$\left(\because \sum_{i=1}^{n} \bar{x} = \bar{x} + \bar{x} + \bar{x} + \ldots \text{upto } n \text{ terms} = n\bar{x}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 + \bar{x}^2 - 2\frac{\bar{x}}{n}(n\bar{x}) \left(\because \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

or $$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \quad \text{. . . (iii)}$$

$$= \frac{1}{n}\sum_{i=1}^{n} (d_i + a)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} (d_i + a)\right)^2 \quad \text{(using (ii))}$$

$$= \frac{1}{n}\sum_{i=1}^{n} (d_i^2 + a^2 + 2d_i\, a) - \frac{1}{n^2}\left(\sum_{i=1}^{n} d_i + na\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} d_i^2 + \frac{1}{n}(na^2) + \frac{2a}{n}\sum_{i=1}^{n} d_i - \frac{1}{n^2}\left[\left(\sum_{i=1}^{n} d_i\right)^2 + n^2a^2 + 2na\sum_{i=1}^{n} d_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)^2$$

$\therefore$ $$\sigma = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n} - \left(\frac{\sum_{i=1}^{n} d_i}{n}\right)^2} \quad \text{. . . (iv)}$$

In case of grouped data, this formula takes the shape

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i\, d_i^2}{n} - \left(\frac{\sum_{i=1}^{n} f_i\, d_i}{n}\right)^2}$$

where $$n = \sum_{i=1}^{n} f_i \quad \text{. . . (v)}$$

We can still modify this formula by defining $u_i = \frac{y_i - a}{c}$, c being the class size.

We are assuming that all classes are of equal width and the frequency of each class is centred at its class mark. In this method

$$\sigma = c\sqrt{\frac{\sum f_i\, \mu_i^2}{n} - \left(\frac{\sum_{i=1}^{n} f_i\, u_i}{n}\right)^2} \quad \text{. . . (vi)}$$

Standard deviation is usually abbreviated as SD.

Example 7.8

Find the SD of the first n natural numbers.

Solution

The series is 1, 2, 3, . . . , n.

We know that $$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left[\frac{\sum_{i=1}^{n} x_i}{n}\right]^2 \quad \ldots$$

(i)

$$\therefore \quad \sigma = \sqrt{\frac{\sum n^2}{n} - \left(\frac{\sum n}{n}\right)^2} \qquad (\because\ x_i = i \text{ in this case})$$

$$= \sqrt{\frac{n\,(n+1)\,(2n+1)}{6n} - \frac{n^2\,(n+1)^2}{4n^2}}$$

$$= \sqrt{(n+1)\left(\frac{2n+1}{6} - \frac{n+1}{4}\right)}$$

$$= \sqrt{(n+1)\left(\frac{4n+2-3n-3}{12}\right)} = \sqrt{\frac{n^2-1}{12}}$$

Example 7.9

The scores of a batsman in 10 different matches were 38, 70, 48, 34, 42, 55, 63, 46, 54, 44. Find the MD and SD of these scores.

Solution

The scores arranged in ascending order are 34, 38, 42, 44, 46, 48, 54, 55, 63, 70.

Number of observations = 10

$$\therefore \quad \text{Median} = \text{Mean of } \frac{10}{2}th \text{ and } \left(\frac{10}{2}+1\right)^{\text{th}} \text{ observations}$$

= Mean of 5^{th} and 6^{th} observations

$$= \frac{46+48}{2} = 47\,.$$

x_i	34	38	42	44	46	48	54	55	63	70	Total
$\lvert x_i - 47 \rvert$	13	9	5	3	1	1	7	8	16	23	86

Hence $$\text{MD} = \frac{\sum_{i=1}^{10} |x_i - \text{median}|}{10} = \frac{13+9+5+3+1+1+7+8+16+23}{10}$$

$$= \frac{86}{10} = 8.6$$

To find the SD, let $a = 48$, consider the following table

x_i	$d_i = x_i - a$	d_i^2
34	– 14	196
38	– 10	100
42	– 6	36
44	– 4	16
46	– 2	4
48	0	0
54	6	36
55	7	49
63	15	225
70	22	484
Total	14	1146

$$\therefore \qquad \sigma^2 = \frac{\sum d_i^2}{n} - \left(\frac{\sum d_i}{n}\right)^2$$

$$= \frac{1146}{10} - \left(\frac{14}{10}\right)^2 = 114.6 - 1.96 = 112.64$$

$$\therefore \qquad \sigma = \sqrt{112.64} = 10.61 \text{ nearly.}$$

Example 7.10

In a survey of 950 families in a village, the following distribution of number of children was obtained.

No. of Children	0 – 2	2 – 4	4 – 6	6 – 8	8 – 10	10 – 12
No. of Families	272	328	205	120	15	10

Find the mean, median and the standard deviation.

Solution

Let us take $a = 5$ the assumed mean. We construct the following table

Class-interval	y_i	f_i	**Comm. f_i**	$u_i = \frac{y_i - 5}{2}$	u_i^2	$f_i u_i$	$f_i u_i^2$
0 – 2	1	272	272	– 2	4	– 544	1088
2 – 4	3	328	600	– 1	1	– 328	328
4 – 6	5	205	805	0	0	0	0
6 – 8	7	120	925	1	1	120	120
8 – 10	9	15	940	2	4	30	60
10 – 12	11	10	950	3	9	30	90
Total		950				– 692	1686

$$\text{Mean } = a + \frac{\sum f_i u_i}{\sum f_i} \times c = 5 + \frac{-692}{950} \times 2 = 5 - \frac{692}{475}$$

$$= 5 - 1.457 = 3.543$$

$$\text{SD } = c \times \sqrt{\frac{\sum f_i u_i^2}{\sum f_i} - \left(\frac{\sum f_i u_i}{\sum f_i}\right)^2}$$

$$= 2 \times \sqrt{\frac{1686}{950} - \left(\frac{-692}{950}\right)^2}$$

$$= 2 \times \sqrt{1.7747 - 0.5300} = 2 \times \sqrt{1.2442}$$

$$= 2 \times 1.115 = 2.23 \text{ nearly.}$$

To find median, we note $\frac{n}{2} = \frac{950}{2} = 475$, and the median class is 2 – 4.

$$\text{Hence median } = 2 + \frac{475 - 272}{328} \times 2 \quad \left(l + \frac{\frac{n}{2} - c_{f-1}}{f_m} c_m \right)$$

$$= 2 + 1.238 = 3.238 \text{ nearly.}$$

Example 7.11

Calculate the mean deviation for the following data :

Marks	**No. of Children**
0 – 10	5
10 – 20	8
20 – 30	15
30 – 40	16
40 – 50	6

Solution

Construct the table (last two columns to be completed after the calculation of median).

Class	y_i	f_i	**Comm. f_i**	$\lvert y_i - 28\rvert$	$f_i \lvert y_i - 28\rvert$
0 – 10	5	5	5	23	115
10 – 20	15	8	13	13	104
20 – 30	25	15	28	3	45
30 – 40	35	16	44	7	112
40 – 50	45	6	50	17	102
Total		50			478

$$n = 50 \Rightarrow \frac{n}{2} = 25 .$$

The class of median is 20 – 30.

$$\text{Hence median } = 20 + \frac{25-13}{15} \times 10 \quad \left| \; l + \frac{\frac{n}{2} - c_{f-1}}{f_m} \times c_m \right.$$

$$= 20 + 8 = 28$$

$$\text{M. D } = \frac{478}{50} = 9.56 \quad \left| \; \frac{\Sigma f_i \,|y_i - 28|}{\Sigma f_i} \right.$$

Example 7.12

The score of 48 children in a test are shown in the following frequency table :

Score	71	76	79	83	86	89	92	97	101	103	107	110	114
Frequency	4	3	4	5	6	5	4	4	3	3	3	2	2

Find σ^2.

Solution

Let $a = 90$, the assumed mean. Construct the following table

x_i	f_i	$d_i = x_i - a$	d_i^2	$f_i d_i$	$f_i d_i^2$
71	4	– 19	361	– 76	1444
76	3	– 14	196	– 42	588
79	4	– 11	121	– 44	484
83	5	– 7	49	– 35	245
86	6	– 4	16	– 24	96
89	5	– 1	1	– 5	5
92	4	2	4	8	16
97	4	7	49	28	196
101	3	11	121	33	363
103	3	13	169	39	507
107	3	17	289	51	867
110	2	20	400	40	800
114	2	24	576	48	1152
Total	48			21	6763

$$\sigma^2 = \frac{6763}{48} - \left(\frac{21}{48}\right)^2 \quad \left| \quad \frac{\Sigma f_i \, d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i \, d_i}{\Sigma f_i}\right)^2 \right.$$

$$= 140.896 - 0.191 = 140.705 \text{ nearly.}$$

SAQ 1

(a) The postal expenses on the letters dispatched from an office on a given day is given in the following frequency distribution :

Postage (P)	15	30	35	60	70
Number of Letters	47	33	56	41	25

Find the mean postage per letter.

(b) The mean age (in years) per student and the number of students in each of the four classes of two primary schools are given below :

	School A		**School B**	
	No.	**Mean Age**	**No.**	**Mean Age**
Class I	6	6.2	25	7.1
Class II	10	7.5	32	8.4
Class III	28	8.6	12	9.2
Class IV	30	10.0	4	10.7

Obtain the mean age per student for the two schools.

(c) The measurements (in mm) of the diameters of the heads of 107 screws gave the following frequencies distribution.

Diameter	33 – 35	36 – 38	39 – 41	42 – 44	45 – 47
Frequency	17	19	23	21	27

Find the mean head diameter per screw.

(d) The marks obtained out of 50 by 102 students in a test were recorded and were according to the following frequency table :

Marks	20	22	23	24	26	31	38	43
Number of Students	8	15	28	27	20	2	1	1

Obtain the median and describe what information it conveys.

SAQ 2

(a) The following table gives the frequency distribution of married women by age at marriage :

Age (in years)	15 – 19	20 - 24	25 - 29	30 - 34	35 - 39
Frequency	53	140	98	32	12
Age (in years)	40 - 44	45 – 49	50 – 54	55 – 59	60 and above
Frequency	9	5	3	3	2

Calculate the median.

(b) The following table gives the weekly consumption of electricity of 50 families. Find the mean and median weekly consumption :

Weekly Consumption	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Number of Families	6	12	18	3	1

(c) The following data is about the number of days patients stayed in a hospital after an operation. Calculate the SD.

Hospital Stay (in days)	1 – 4	4 – 7	7 – 10	10 – 13	13 – 16	16 – 19	19 – 22
Number of Patients	32	108	67	28	14	7	3

(d) Calculate the SD for the following data :

Wages per Week up to (Rs.)	15	30	45	60	75	90	105	120
Number of Workers	12	30	65	107	157	202	220	225

SAQ 3

(a) Find the SD of the following table :

x_i	140	145	150	155	160	165	170	175
f_i	4	6	15	30	36	24	8	2

(b) The length (in cm) of 10 small pieces of cloth were :

5, 3, 9, 12, 3, 10, 12, 21, 18, 12

Find the mean deviation and the standard deviation.

(c) Find the mean deviation for the following observations

3, 3, 5, 9, 10, 12, 12, 12, 18, 21, 21.

7.6 SUMMARY

- The arithmetic of *n* individual observations $x_1, x_2, x_3, \ldots, x_n$ is given by (denoted by $\bar{x}$)

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- The arithmetic mean $\bar{y}$ of a grouped data classified into *k* classes with class marks $y_1, y_2, \ldots, y_k$ and frequencies $f_1, f_2, \ldots, f_k$ is given by

$$\bar{y} = \frac{\sum_{i=1}^{k} f_i\, y_i}{\sum_{i=1}^{k} f_i} = a + \frac{\sum_{i=1}^{k} f_i\, d_i}{\sum_{i=1}^{k} f_i}$$

$$d_i = y_i - a$$

or $$\bar{y} = a + c\,\frac{\sum_{i=1}^{k} f_i\, u_i}{\sum_{i=1}^{k} f_i}$$

where $u_i = \dfrac{y_i - a}{c}$, *c* is the width of the class.

- The median *M* of *n* individual observations $x_1, x_2, \ldots, x_n$ is

$$M = \begin{cases} \left\{\dfrac{n+1}{2}\right\}^{\text{th}} \text{observation}, & \text{if } n \text{ is odd} \\[2ex] \dfrac{\dfrac{n}{2}\text{th observation} + \left(\dfrac{n}{2}+1\right)\text{th observation}}{2}, & \text{if } n \text{ is even} \end{cases}$$

- The median *M* of *k* observations $x_1, x_2, \ldots, x_k$ with frequencies $f_1, f_2, \ldots, f_k$ is given by

$$M = l + \frac{\frac{n}{2} - c_{f-1}}{f_m} \times c_m$$

where l = lower limit of the median class,

f_m = frequency of the median class,

c_{f-1} = commulative frequency of the class preceding the median class,

n = sum of all the frequencies, i.e. all the observations, and

c_m = width of the median class.

The median class being the class which contains the $\frac{n}{2}$th observations.

- The mean deviation denoted by MD about median is

$$MD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{1}{n} \sum_{i=1}^{k} |d_i|,$$

n being the total number of individual observations, x_1, x_2, …,x_n and $\bar{x}$ being the median.

- In case of grouped data

$$\text{MD} = \frac{\sum_{i=1}^{k} f_i |(y_i - \bar{y})|}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i |d_i|}{\sum_{i=1}^{n} f_i}$$

where $d_i = y_i - \bar{y}$.

- The standard deviation denoted by SD or σ is given by

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

In case of grouped data

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (y_i - \bar{y})^2}{\sum_{i=1}^{k} f_i}$$

- If $d_i = x_i - a$, where a is an assumed mean then

$$\sigma^2 = \frac{\sum_{i=1}^{n} d_i^2}{n} - \left(\frac{\sum_{i=1}^{n} d_i}{n} \right)^2$$

and in case of grouped data

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i\, d_i^2}{\sum_{i=1}^{k} f_i} - \left(\frac{\sum_{i=1}^{k} f_i\, d_i}{\sum_{i=1}^{k} f_i}\right)^2$$

If we define $u_i = \frac{y_i - a}{c}$, c being the class size, then

$$\sigma^2 = c^2 \left[\frac{\sum_{i=1}^{k} f_i\, u_i^2}{\sum_{i=1}^{k} f_i} - \left(\frac{\sum_{i=1}^{k} f_i\, u_i}{\sum_{i=1}^{k} f_i}\right)^2\right].$$

7.7 ANSWERS TO SAQs

SAQ 1

(a) 38.9 paise.

(b) 8.8 years, 8.2 years.

(c) 40.617 mm.

(d) Median = 23.5, 50% of the student obtain less than 23.5 marks out of 50.

SAQ 2

(a) Median = 24.5.

(b) 17.2 units, 17.5 units.

(c) 3.75.

(d) 636.1.

SAQ 3

(a) SD = 7.26 nearly.

(b) 4.5, 31.85.

(c) $\frac{54}{11}$

FURTHER READINGS

Shanti Narayan, *Integral Calculus*, by S. Chand & Co.

Georgy Thomas and Ross L. Finney, *Calculus and Analytic Geometry*, Narosa Publishing House.

M. K. Singal and Asha Rani Singal, *Complete Course in I. S. C. Mathematics Part-I and Part-II*, Pitambar Publishing House.