# FREQUENCY TABLES

**Statistics:** Statistics is the science which deals with the method of collecting and presenting numerical data to throw light on any sphere of enquiry and enables us to interpret it and draw interferences from it.

**Characteristics of Statistics:**

1. Statistics are the aggregate of facts.
2. Statistics are numerically expressed.
3. Statistics should be placed in relation to each other.
4. Statistics should be collected for a pre-determined purpose.
5. Statistics are affected to a marked extent by multiplicity of causes and not by a single cause.
6. Statistics should be collected in a systematic manner.
7. The reasonable standard of accuracy should be maintained in statistics.

**Importance and Usefulness of Statistics:**

1. Statistics help in presenting large quantity of data in a simple and classified form.
2. It helps in finding the conditions of relationship between the variables.
3. It gives the methods of comparison of data and it weighs and judges them in the right perspective.
4. It tries to give material for businessman as well as to the administrators so as to serve as a guide in planning and in shaping future policies and programmes.
5. It enlarges individual mind.
6. It when considered as the logic of figures, assists in arriving at, correct views based of facts.
7. It proves useful in a number of fields like Railways, Banks, Army etc.

**Raw Data:**

The information which we get through censuses, suerveys or in a routine matter is called raw data.

The word 'data' means *information*(or given facts).

The word raw attached to data indicates that this information, thus collected and recorded, cannot be put into use and requires processing.

**Collection of Data:** The collection of data is the primary need for any statistical investigation. There are two types of data:

**Primary Data:** It is the data collected by a particular person or organisation for his own use from the primary source.

**Mehod of Collecting Primary data:**

1. Direct personal observations.
2. Indirect oral investigations.
3. Estimates from local sources and correspondence.
4. Data through questionnaires.
5. Investigation through enumerators.

**Secondary Data:** It is the data collected by some other person or organisation for their own use but the investigator also gets it for his use.

**Method of Collecting the Secondary Data:**

1. Information collected through newspaper and periodicals.
2. Information obtained from the research papers published by University department or research bureaus of U.G.C.
3. Information obtained from the official publications of the central, state and the local government dealing with crop statistics, industrial statistics, trade and transport statistics.

4. Informatin obtained from the publications of trade association.
5. Information obtained from the official publications of the foreign government for international organisation.

**Classification of Data:** There are four bases of classification of data:

1. Quantitative e.g., a class of students split up into groups according to their heights or ages.
2. Qualitative e.g., rich and poor, educated and uneducated persons, intelligent and dull students.
3. Spatial or Geographical (basis of classification is according to difference in geographical location or space) e.g., birth rate of India is divided statewise.
4. Temporal (classification is according to difference in time).

**Variables of Observation:** Suppose that in a census we record the age, and place of residence (rural or urban) of each person. Here we take observations on three variables: age, sex, and place of residence. Thus, the term variable stands for what is being observed. This is variable because the results are different for different persons. A person may have age 10, another 32, and so on.

Thus 10, 32, .... are the variables. Similarly the values of sex variable are 'male' and 'female'. The age variable can take any value 1, 2, 3, .... but we may not observe all as there may be no person with age 2 at a given time. In view of this we distinguish between 'observed values' and 'possible values'.

**Qualitative and Quantitative Variables:** In the above example of census, the values of age-variable are numbers, those of sex-variable are names ('male or female') and those variable of place of residence are also name ('rural' or 'urban').

*The variables of observation with numbers as possible values are called quantitative variables.*

**Example:** Age, height, income; etc. are quantitative variables. The variables of

observation with names of things, place, attributes; etc. as possible values are called quantitative variables.

**Example:** Sex, religion, caste; etc. quantitative variables.

**Units of Observation:** The term, unit of observation, is used to describe what the values of a variable are attached to. In the above census example, the units of observation are the persons alive at the time of census and to each unit of observation. We associate, the value of three variables; age, sex and place of residence.

**Types of Variable:** There are tow types of variable:

(*i*) **Continuous Variable:** A quantity which can take any numerical value within a certain range, is called continuous variable.
**Example:** Measurement such as height, pressure, temperature; etc. are all continuous variables.

(*ii*) **Discontinuous (Discrete) Variable:** A quantity which is incapable of taking all possible values is called discontinuous variable.

This is also called discrete variable.

**Examples:**

(I) Number of members in a family can't be $4\frac{1}{2}$.

(II) Marks obtained by a student can't be $47\frac{1}{2}$ etc.

**Construction of Frequency Table:** If the number of observations in the raw data is large, the investigator has to devise way to condense the data and present them in tables and charts in order to bring out their main features. This is known as data presentation.

**Example:** The marks obtained by 30 students of XI class in a test (out of 25 marks) in Mathematics are:

17, 5, 18, 17, 18, 2, 16, 13, 8, 17, 8, 18, 2, 13, 17, 8, 16, 18, 8, 5, 13, 8, 18, 16, 8, 13, 18, 2, 5, 18.

These are called ungrouped (or raw) data which do not give any useful information. We can classify them in two ways:

We could, for instance arrange them in ascending or descending order. (This is said to be an array).

But this can not be taken as a rule of the number if observations are large.
We can condense the data into classes (or groups) as below:

(*a*) **Determine the range of raw data** [Range of raw data is the difference between the maximum and the minimum number occurring in the data]

(*b*) **Decide upon the number of classes into which raw data are to be grouped:** [**Rule:** Have no fewer than 5 or more than 12 classes].

The accuracy is lost if we take less number maximum numbers occurring in the data are included in the same class.

**Rules:**

1. Classes should not be overlapping.
2. There should be no gaps between classes.
3. Classes should be of the same size.

(*e*) Take each number from the data (one at a time) and place a tally mark (/) opposite to the class to which it belongs.

The tally marks can be recorded in bunches of 5.

After the occurrence four times, the fifth ocurrrence is recorded by a cross tally (卌) on the first four tallies.

[The counting of tally marks in a particular class is called the frequency of that class].

The table showing how frequencies are distributed is called frequency table.

If the number of observation is less than a particular value, we find cumulative frequency of each class.

(*f*) The table showing the manner in which cumulative frequencies are distributed is called Cumulative Frequency Table.

**Relative Frequency Table:** Sometimes the frequency is expressed as a fraction of the total frequency and this fraction (usually expressed in percentage) is called the relative frequency.

Consider the frequency inhabitants of U.P. by age (1971 census)

| Age | Frequency | % |
|---|---|---|
| 0–9 | 26,105,403 | 29.5 |
| 10–14 | 10,859,933 | 12.3 |
| 15–19 | 7,184,548 | 8.1 |
| 20–24 | 6,531,468 | 7.3 |
| 25–29 | 6,474,870 | 7.4 |
| 30–34 | 5,910,572 | 6.7 |
| 35–39 | 5,174,221 | 5.9 |
| 40–44 | 4,737,880 | 5.4 |
| 45–49 | 3,676,085 | 4.2 |
| 50–54 | 3,598,058 | 4.1 |
| 55–59 | 2,103,947 | 2.4 |
| 60–64 | 2,636,013 | 3.0 |
| 65–69 | 1,241,876 | 1.4 |
| 70 and above | 2,099,009 | 2.4 |

The relative frequencies make it easy to understand when the class frequencies are very large.

**Types of Classes:** There are two types of classes viz. Exclusive and Inclusive.

(*a*) Consider the classes

5–10, 10–15, 15–20, .....

Here, in the class 5–10, 5 is the lower limit and 10 is the upper limit.

Similar is the case with other classes. Here, the upper limit of a class is the lower limit of the next class in each case.

Thus, there are no gaps.

Size of each class = Its upper limit – Its lower limit

= 10–5 = 15–10 = 20–15= ... = 5.

The common point of two classes is included in the higher class.

**Example:** $10 \in (10–15)$ and $10 \notin (5–10)$

Thus, $x \in (10–15) \Rightarrow 10 \leq x < 15$

[$x$ can't = 15]

Thus, the upper limit is excluded.

Such classes are called exclusive classes.

(*b*) Consider the classes

10–19, 20–29, 30–39, .....

$\therefore$ $20 \in (20–29)$ and $29 \in (20–29)$.

Here $x \in (20–29) \Rightarrow 20 \leq x \leq 29$

Thus, the upper limit is included.

Such classes are called inclusive classes.

**Sum up:**

The lower limit is always inclusive.

The class is inclusive or exclusive according as the upper limit is included or excluded.

Method to convert inclusive classes into exclusive classes?

We call

$$\left.\begin{array}{l}\textbf{adjustment}\\ \textbf{factor}\end{array}\right\} = \frac{\begin{array}{c}\text{Lower limit}\\ \text{of 2nd class}\end{array} - \begin{array}{c}\text{Upper limit}\\ \text{of 1st class}\end{array}}{2}$$

Subtract this adjustment factor from the lower limits and add to the upper limits so as to get the required exclusive classes.

**Example:** The inclusive classes are

10–19, 20–29, 30–39, 40–49, .....

$$\textbf{Adjustment factor} = \frac{20-19}{2} = \frac{1}{2} = .5$$

Subtracting .5 from the lower limits and adding to upper limits, we get

9.5–19.5, 19.5–29.5, 29.5–39.5, 39.5–49.5,....

These are exclusive classes.

The mid-point of a class-interval is called its Class-mark or Mid-value.

**Example:** class-mark of 9.5–19.5 is

$$\frac{9.5+19.5}{2} = \frac{29}{2} = 14.5$$

**Remember:**

$$\text{Class - mark} = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

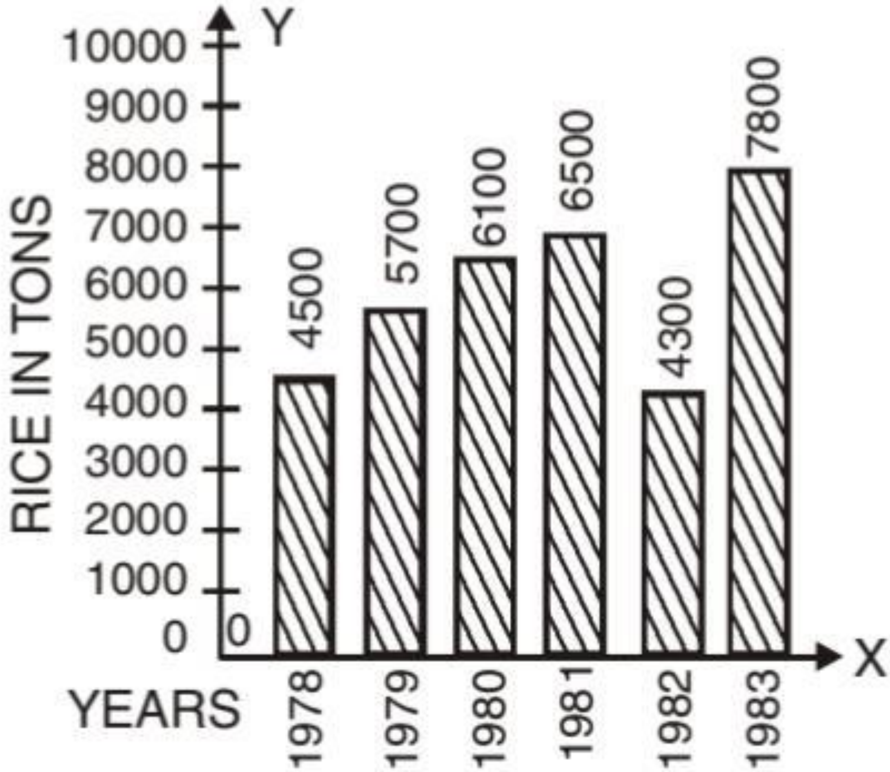
**Graphical Presentation of Frequency Distribution:** The graphical representation of the frequency distribution is convenient communication method.

(*i*) **Simple Bar Diagram:** It is used to compare two or more items related to variable. In this case the data are presented with the help of bars. These bars are usually arrranged according to relative magnitude of bars. The length of a bar is determined by the value or the amount of the variable. A limitation of Simple Bar Diagram is that only one variable can be represented on it.

**Illustration:** Draw a Bar chart of the procurement of rice (in tons) in an Indian state:

| **Year:** | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 |
|---|---|---|---|---|---|---|
| **Rice (in tons):** | 4500 | 5700 | 6100 | 6500 | 4300 | 7800 |

**Solution:** The given data is represented by the Simple Bar Diagram as only one variable is to be presented.




Here we represent years on the $x$-axis and procurement of rice on the $y$-axis.

(*ii*) **Multiple or Grouped Bar Diagram:** A multiple or grouped bar diagram is used when a number of items are to be compared in respect of two, three or more values. In this case the numerical values of major categories are arranged in ascending or descending order so that the categories can be readily distinguished. Different shades or colours are used for each category.

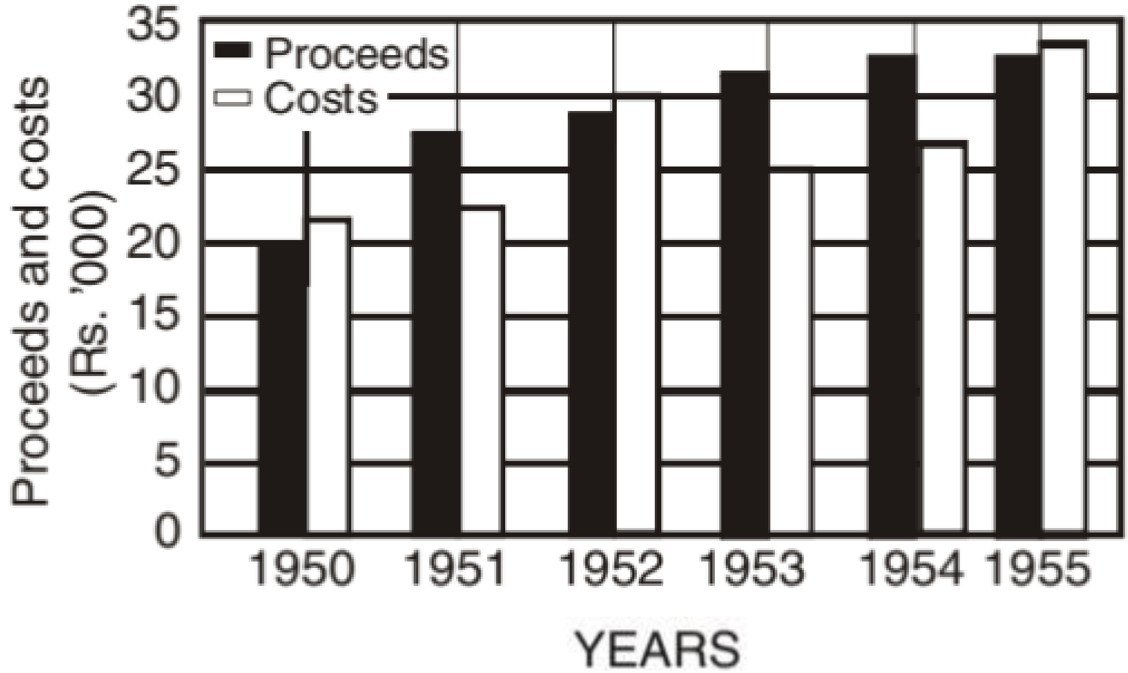
**Illustration:** Represent the following data by a suitable diagram showing the difference between proceeds and costs:

*Proceeds and costs of a firm (in thousand of rupees)*

| Year | Total proceeds | Total costs |
|---|---|---|
| 1950 | 22.0 | 19.5 |
| 1951 | 27.3 | 21.7 |
| 1952 | 28.2 | 30.0 |
| 1953 | 30.3 | 25.6 |
| 1954 | 32.7 | 26.1 |
| 1955 | 33.3 | 34.2 |

**Solution:** In this case the two types of information, proceeds and costs, are shown on a diagram indicating the difference between them. The two types of information for each year are placed in such a way that a comparison may be made between them.

So, a grouped bar diagram is drawn in order to represent the given data.



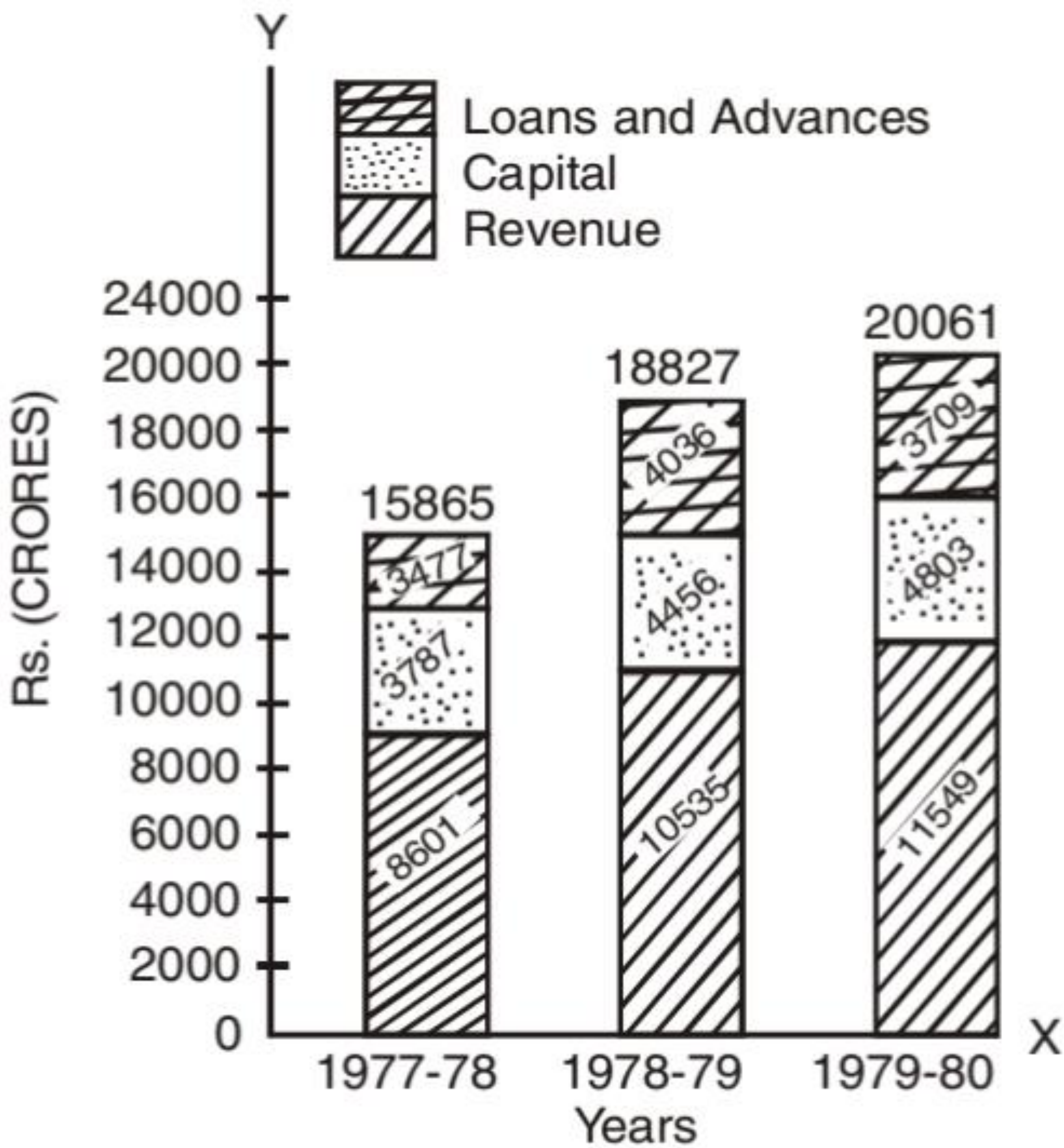

(*iii*) **Subdivided or Component Bar Diagram:** A component bar diagram is one which is formed by dividing a single bar into several component parts. A simple bar represents the aggregate value whereas the component parts represent the component values of the aggregate value. It shows the relationship among the different parts and also between the different parts and the main bar. The design and procedure of constructing it is similar to simple bar diagrams except that in this form of presentation each bar is subdivided into its components. Different shades or colours or design of different types of cross hatchings are used to distinguish the various components.

**Illustration:** Represent the following data of the development expenditure of Central and the State Governments in India during 1977–78, 1978–79, 1979–80 by bar diagram.

| Year | Loans & Advances | Capital | Revenue | Total |
|---|---|---|---|---|
| 1977–78 | 8,601 | 3,787 | 3,477 | 15,865 |
| 1978–79 | 10,535 | 4,456 | 4,036 | 18,827 |
| 1979–80 | 11,549 | 4,803 | 3,709 | 20,061 |

**Solution:** In this case we use subdivided bar diagram. Here the data for different years is represented by various parts of the single bar for the years. Different parts are indicating loans and advances; Capital and revenues for each year as shown in the figure.

(*iv*) **Percentage Sub-divided Bar Diagrams:** It consists of one or more than one bars where each bar totals 100%. Its construction is similar to the subdivided bar diagram with the only difference that whereas in the subdivided bar diagram segments are used in absolute quantities, in the percentage




bar diagram the quantities are transformed into percentages. Its construction is based on the following steps:

**Step I.** Convert the quantities in each case into percentage of the whole

**Step II.** Take the cumulative percentages.

**Step III.** Represent each category of items in different shades or colours or different type of cross-hatchings.

**Step IV.** In case two diagrams showing different periods are given, then show each category of item with the same shade or colour or cross hatching.

**Illustration:** Following are the heads of income of Railways during 1949 and 1950.

| | **1949** | **1950** |
|---|---|---|
| | **(in crores of rupees)** | **(in crores of rupees)** |
| Coaching | 26 | 31 |
| Goods | 40 | 39 |
| Others | 4.50 | 3.50 |

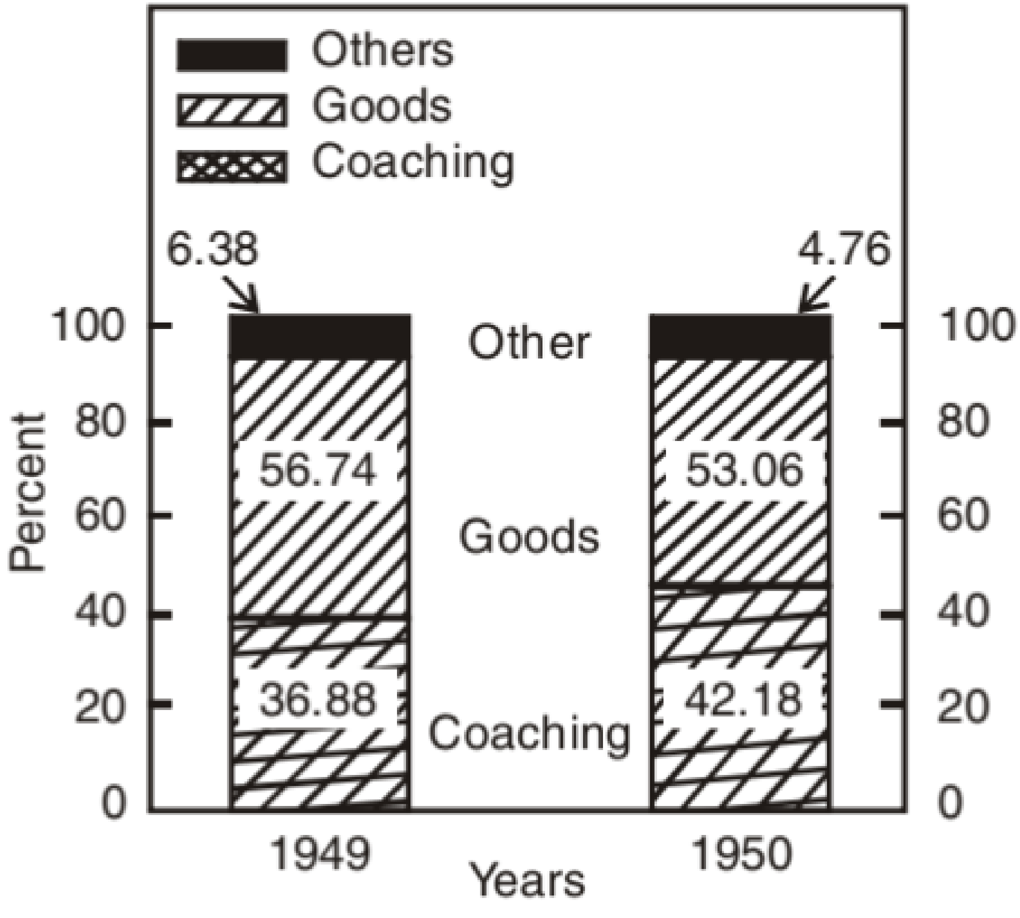
*Represent the above data by a bar chart.*

**Solution:** We are given three types of information–Coaching, goods and others for two years. In order to facilitate comparison among them and also between the two years, a component bar chart is drawn to represent the given data. The graph has been drawn on percentage figures. Percentage of each item has been expressed on the total income of respective years by formula

$$\text{Percentage} = \frac{\text{Income of the item}}{\text{Total income of respective year}} \times 100$$

**Calculation of percentage.**

| Items | 1949 | | 1950 | |
|---|---|---|---|---|
| | Income | Percentage | Income | Percentage |
| Coaching | 26 | 36.88 | 31 | 42.18 |
| Goods | 40 | 56.74 | 39 | 53.06 |
| Other | 4.50 | 6.38 | 3.50 | 4.76 |
| | 70.50 | 100 | 73.50 | 100 |




(*v*) **Pie Chart or Sector Chart:** A pie chart is a circular graph which represents the total value with its components. The area of a circle represents the total value and the different sectors of the circle represent the different part. The circle is divided into sectors by radii and the areas of the sectors are proportional to the angles at the centre. It is generally used for comparing the relation between various components of a value and between components and the total value. In pie chart, the data is expressed as percentage. Each component is expressed as percentage of the total value. A pie chart is also known as Circular Chart or Sector Chart.

A pie diagram is used for representing relative frequency distributions, where different sectors of the circles represent relative frequencies.

**Construction of Pie Diagram**

**Step I.** Plot a circle of appropriate size with protractor, pencil and compass. The angles of a circle total 360°.

**Step II.** Convert the given value of the components of an item in percentage of the total value of the item.

**Step III.** In laying out the sector for a pie chart it is logical to adopt the common procedure to arrange sectors according to size with the largest at the top and the others in sequence running clockwise.

**Step IV.** Transpose the various component values corresponding to the degree on the circle. Since 100% is represented by 360° angle at the centre of the circle, therefore 1% value is represented by

$$\frac{360^\circ}{100} = 3.6^\circ.$$

If 8 be the percentage of a certain component, the angle which represents the percentage of such component is (3.6 × 8) degrees.

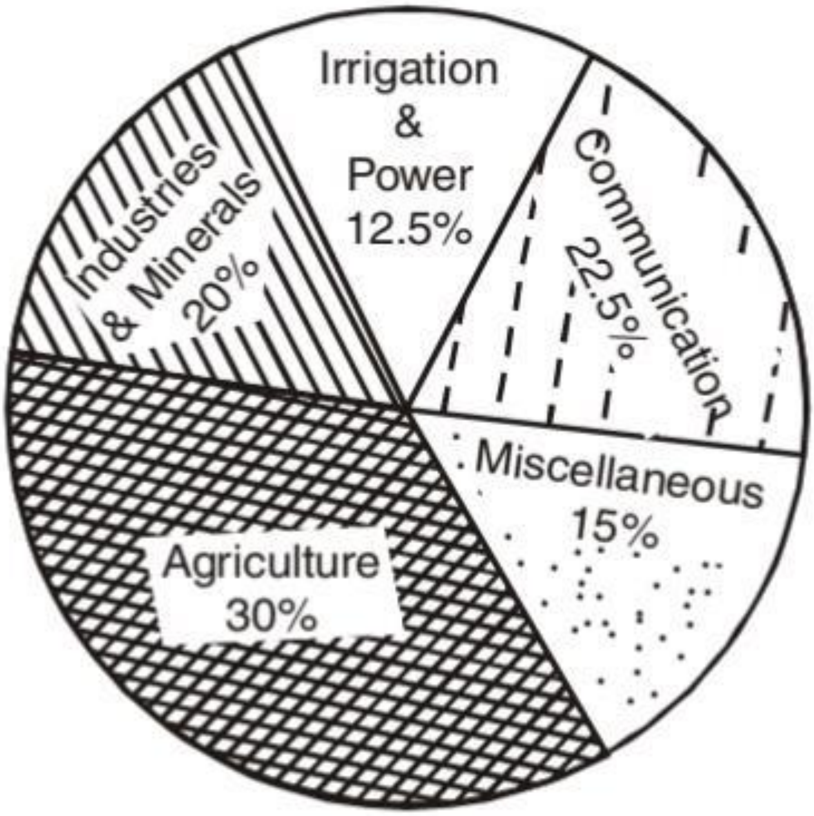
**Setp V.** Measure with protractor the points on a circle representing the size of each sector.

**Step VI** Label each sector for identification.

**Illustration:** Draw a pie chart to represent the following data on the proposed out lay during the Fourth Five Year Plan.

| Items | Agriculture | Industries and Minerals | Irrigation and Power | Commu-nication | Miscella-neous |
|---|---|---|---|---|---|
| Rs. in Crores | 6000 | 4000 | 2500 | 4500 | 3000 |

**Solution:** In constructing a pie chart, it is necessary to convert the percentages into angles of different degrees.




**Calculation for Pie Chart**

| Items | Amount in (Rs.) Crores | Percentage on total (%) | Angle for each percentage (360° ÷ 100) | Angle for each item at the centre of the Pie-chart |
|---|---|---|---|---|
| Agriculture | 6000 | $\frac{6000 \times 100}{20000} = 30$ | 3.6° | 30 ×3.6° = 108° |
| Industries and Minerals | 4000 | $\frac{4000 \times 100}{20000} = 20$ | 3.6° | 20 ×3.6° = 72° |
| Irrigation and Power | 2500 | $\frac{2500 \times 100}{20000} = 12.5$ | 3.6° | 12.5 × 3.6° = 45° |
| Communication | 4500 | $\frac{4500 \times 100}{20000} = 22.5$ | 3.6° | 22.5 × 3.6° = 81° |

| | | | | |
|---|---|---|---|---|
| Miscellaneous | 3000 | $\frac{3000 \times 100}{20000} = 15$ | 3.6° | 15 × 3.6° = 54° |
| Total | 20000 | 100 | | 360° |

## Arithmetic Mean

(*i*) **Mean of Raw Data:** We know that the average of $n$-numbers is obtained by finding their sum (by adding) and then dividing it by $n$. Let $x_1, x_2, x_3, \ldots, x_n$ be $n$-numbers, then their average is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 \ldots + x_n}{n}$$

**Illustration:** Find the arithmetic mean of the marks obtained by 10 students of class $X$ in Mathematics in a certain examination. The marks obtained are: 25, 30, 21, 55, 47, 10, 15, 17, 45, 35.

**Solution:** Let $\bar{x}$ be the average marks

∴ Sum of all the observations

= 25 + 30 + 21 + 55 + 47 + 10 +15 + 17+ 45 +35 =300

Number of students = 10

$$\therefore \text{Arithmetic mean} = \frac{300}{10} = 30.$$

(*ii*) **Mean of Grouped Data:**

Let $x_1, x_2, x_3, \ldots x_n$ be the variates and let $f_1, f_2, f_3, \ldots f_n$ be their corresponding frequencies, then their mean $\bar{x}$ is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots\ldots f_nx_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum f_i} = \frac{\sum_{i=1}^{n} f_ix_i}{N}$$

where $N = f_1 + f_2 + f_3 + \ldots. + f_n$

**Illustration:** Find the Arithmetic Mean from the frequency table

| **Marks** | 52 | 58 | 60 | 65 | 68 | 70 | 75 |
|---|---|---|---|---|---|---|---|
| **No. of Students** | 7 | 5 | 4 | 6 | 3 | 3 | 2 |

**Solution:** Let $x$ be the marks and $f$ be the frequency so that we have the following table:

| $x$ | $f$ | $fx$ |
|---|---|---|
| 52 | 7 | 364 |
| 58 | 5 | 290 |
| 60 | 4 | 240 |
| 65 | 6 | 390 |
| 68 | 3 | 204 |
| 70 | 3 | 210 |
| 75 | 2 | 150 |
| Total | 30 | 1848 |

Here $N = \Sigma f = 30$ and $\Sigma fx = 1848$

$$\therefore \text{Mean} = \bar{x} = \frac{1848}{30} = \frac{616}{10} = 61.6$$

**Short-cut Method**

($i$) In the case of ungrouped data

$$\bar{x} = a + \frac{\Sigma d}{n}, \text{ where}$$

$a$ = assumed mean, $n$ = number of items,
$d = x - a$ = deviations of any variate from $a$.

**Working Rule for Short Cut Method for Ungrouped Data**

**Step I.** Denote the variable of the discrete series by $x$ or $X$.

**Step II.** Take any item of series, preferable the middle one, and denote it by $a$. This number $a$ is called the assumed mean or provision mean.

**Step III.** Take the difference $x - a$ and denote it by $d$ or $dx$ or $d' = x - \text{a}$, where $d'$ is the deviation of variate from '$a$'.

**Step IV.** Find the sum $\Sigma dx$ or $\Sigma d$.

**Step V.** Use the following formula to calculate the arithmetic mean.

$$\bar{x} = a + \frac{\Sigma d}{n}$$

**Illustration:** Find, by short-cut method, the mean height of the following 8 students whose heights in centimetre are

59, 65, 71, 67, 61, 63, 69, 73.

**Solution:** Let us take 65 as assumed mean i.e., $a = 65$. Let us prepare the following table.

| $x$ | $d = x - 65$ |
|---|---|
| 59 | –6 |
| 61 | –4 |
| 63 | –2 |
| 65 | 0 |
| 67 | +2 |
| 69 | +4 |
| 71 | +6 |
| 73 | +8 |
| | $\Sigma d = 8$ |

Total deviation = $\Sigma d = 8$

Here $a = 65$, $n = 8$, $\Sigma d = 8$

$$\therefore \quad \bar{x} = a + \frac{\Sigma d}{n} = 65 + \frac{8}{8} = 66 \text{ cm.}$$

(*ii*) In the case of grouped data $\bar{x} = a + \dfrac{\Sigma fd}{n}$

where $fd$ = product of the frequency and the corresponding deviation.

N = Σf = the sum of all the frequencies.

**Working Rule for short-cut Method for Grouped Data**

**Step I.** In the case of discrete series, denote the variable by $x$ or $X$ and the corresponding frequency by $f$. (But in the case of continuous series $x$ is the mid-value of the interval and $f$, the frequency corresponding to that interval.)

**Step II.** Take any item of $x$ series, preferable the middle one and denote it by '$a$'. This number '$a$' is called assumed mean or provisional mean.

**Step III.** Take the difference $x - a$ and denote it by $d$ or $dx$ or $d' = x - a$ = deviation of any variate $x$ from $a$, the assumed mean.

**Step IV.** Multiply the respective $f$ and $d$ and denote the product under the column $fd$.

**Step V.** Find $\Sigma fd$.

**Step VI.** Use the following formula to calculate the arithmetic mean.

$$\bar{x} = a + \frac{\Sigma fd}{\Sigma f}$$

**Illustration.** Ten coins were tossed together and the number of tails resulting from them were observed. Calculate the A.M.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $f$: | 2 | 8 | 43 | 133 | 207 | 260 | 213 | 120 | 54 | 9 | 1 |

**Solution:** Let 5 be the assumed mean i.e., $a = 5$.

| $x$ | $f$ | $d = x - 5$ | $fd$ |
|---|---|---|---|
| 0 | 2 | – 5 | – 10 |

| | | | |
|---|---|---|---|
| 1 | 8 | – 4 | – 32 |
| 2 | 43 | – 3 | – 129 |
| 3 | 133 | – 2 | – 266 |
| 4 | 207 | – 1 | – 207 |
| 5 | 260 | 0 | 0 |
| 6 | 213 | 1 | + 213 |
| 7 | 120 | 2 | + 240 |
| 8 | 54 | 3 | + 162 |
| 9 | 9 | 4 | + 36 |
| 10 | 1 | 5 | + 5 |
| | $\Sigma f = 1050$ | | $\Sigma fd = 12$ |

$$\therefore \text{ A.M.} = \bar{x} = a + \frac{\Sigma fd}{\Sigma f} = 5 + \frac{12}{1050} = 5 + 0.0114$$

$$= 5.0114$$

**Illustration:** For the following frequency table, find the mean class:

| 100–120 | 120–140 | 140–160 | 160–180 | 180–200 | 200–220 | 220–240 |
|---|---|---|---|---|---|---|

**Frequency:**

| | | | | | | |
|---|---|---|---|---|---|---|
| 10 | 8 | 4 | 4 | 3 | 1 | 2 |

**Solution:** Let $a = 170$, then we have the following table.

| Class | Frequency $f$ | Mid-value $x$ | $d = x - a$ $= x - 170$ | $fd$ |
|---|---|---|---|---|
| 100–120 | 10 | 110 | – 60 | – 600 |

| | | | | |
|---|---|---|---|---|
| 120–140 | 8 | 130 | – 40 | – 320 |
| 140–160 | 4 | 150 | – 20 | – 80 |
| 160–180 | 4 | 170 | 0 | 0 |
| 180–200 | 3 | 190 | 20 | 60 |
| 200–220 | 1 | 210 | 40 | 40 |
| 220–240 | 2 | 230 | 60 | 120 |
| | $\Sigma f = 32$ | | | $\Sigma fd = -780$ |

$$\text{Mean } = \bar{x} = a + \frac{\Sigma fd}{\Sigma f} = 170 - \frac{780}{32}$$

$$= 170 - 24.3 = 145.7$$

**Step Deviation Method**

When the class intervals in a grouped data are equal, then the calculations can be simplified further by taking out the common factor from the deviations.

$$\text{A.M.} = \bar{x} = a + \frac{\Sigma fd}{N} \times i$$

where $a$ = assumed mean; $d = \frac{x - a}{i}$ = the deviation of any variate from $a$; $i$ = the width of the class-interval, $N$ = Number of observations.

**Illustration:** Find the Arithmetic mean of the followig data:

| 10–20 | 20–30 | 30–40 | 40–50 | 50–60 | 60–70 | 70–80 | 80–90 | 90–100 |
|---|---|---|---|---|---|---|---|---|

**Frequency:**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | 7 | 17 | 29 | 29 | 10 | 3 | 2 | 1 |

**Solution:** Let the assumed mean $a = 45$

| Class Interval | Mid.value $x$ | $d = \frac{x-45}{10}$ | $f$ | $fd$ |
|---|---|---|---|---|
| 10–20 | 15 | – 3 | 2 | – 6 |
| 20–30 | 25 | – 2 | 7 | – 14 |
| 30–40 | 35 | – 1 | 17 | – 17 |
| 40–50 | 45 | 0 | 29 | 0 |
| 50–60 | 55 | 1 | 29 | 29 |
| 60–70 | 65 | 2 | 10 | 20 |
| 70–80 | 75 | 3 | 3 | 9 |
| 80–90 | 85 | 4 | 2 | 8 |
| 90–100 | 95 | 5 | 1 | 5 |
| | | | $N$=100 | $\Sigma fd$=34 |

$a = 45$, N = 100, $\Sigma fd = 34$ and $i = 10$

$$\bar{x} = a + \frac{\Sigma fd}{\Sigma f} \times i$$

$$= 45 + \frac{34}{100} \times 10 = 45 + 3.4 = 48.4$$

## Weighted Arithmetic Mean

If $w_1, w_2, w_3, \ldots, w_n$ are the weights assigned to the values $x_1, x_2, x_3, \ldots, x_n$ respectively, then the weighted average is defined as:

Weighted Arithmetic Mean

$$= \frac{w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n}{w_1 + w_2 + w_3 + \ldots + w_n}$$

**Median:** Median is defined as the middle most or the central value of the variables in a set of observations, when the observations are arranged either in ascending or descending order of their magnitudes. It is denoted by $M$.

**Calculation of Median**

(*a*) **When the data is ungrouped:** Arrange the $n$ values of the variable in ascending (or descending) order of magnitudes.

**Case I. When *n* is odd:** In this case $\frac{n+1}{2}$th value is the median

*i.e.,* $M = \frac{n+1}{2}$th term.

**Illustration:** The number of runs scored by 11 players of a cricked team of a school are 5, 19, 42, 11, 50, 30, 21, 0, 52, 36, 27. Find the median.

**Solution:** Let us arrange the values in ascending order

0, 5, 11, 19, 21, 27, 30, 36, 42, 50, 52 ...(1)

$$\therefore \text{Median, } M = \left(\frac{n+1}{2}\right)\text{th value}$$

$$= \left(\frac{11+1}{2}\right)\text{th value} = 6\text{th value.}$$

Now the 6th value in the data (1) is 27,

$\therefore$ Median = 27 runs.

**Case II. When *n* is even:** In this case there are two middle terms $\frac{n}{2}$th and $\left(\frac{n}{2}+1\right)$th. The median is the average of these two terms *i.e.,*.

$$M = \frac{\frac{n}{2}+\left(\frac{n}{2}+1\right)}{2}$$

**Illustration:** Find the median of the following item: 6, 10, 4, 3, 9, 11, 22, 18.

**Solution:** Let us arrange the items in ascending order, 3, 4, 6, 9, 10, 11, 18, 22

In this data the number of items is $n = 8$, which is even.

$\therefore$ Median $= M =$ average of $\left(\frac{n}{2}\right)$th and $\left(\frac{n}{2}+1\right)$th terms

$$= \text{Average of } \left(\frac{8}{2}\right)\text{th and } \left(\frac{8}{2}+1\right)\text{th terms}$$

=Average of 4th and 5th terms

$$= \frac{9+10}{2} = \frac{19}{2} = 9.5$$

(*b*) **When the data is grouped:**

(i) **When the series is discrete:** In this case the values of the variable are arranged in ascending or descending order of magnitudes. A table is prepared showing the corresponding frequencies and cumulative frequencies. Then the median is calculated as:

$$M = \left(\frac{n+1}{2}\right)\text{th value}$$

where n = Σf = total frequencies.

**Illustration:** Calculate median for the following data:

| | | | | | | |
|---|---|---|---|---|---|---|
| **No. of Students:** | 6 | 4 | 16 | 7 | 8 | 2 |
| **Marks** : | 20 | 9 | 25 | 50 | 40 | 80 |

**Solution:** Arranging the marks in ascending order and preparing the following table.

| Marks | Frequency | Cumulative Frequency |
|---|---|---|
| 9 | 4 | 4 |
| 20 | 6 | 10 |
| 25 | 16 | 26 |
| 40 | 8 | 34 |
| 50 | 7 | 41 |
| 80 | 2 | 43 |
| | $n = \Sigma f = 43$ | |

Here, $n = 43$

$$\therefore \text{Median } = M = \left(\frac{n+1}{2}\right)\text{th value.}$$

$$= \left(\frac{43+1}{2}\right)\text{th value} = 22\text{nd value}$$

The above table shows that all items from 11 to 26 have their values 25. Since 22nd item lies in this interval, therfore its value is 25.

Hence, median = 25.

(*c*) **When the series is continuous.** In this case the data is given in the form of a frequency table with class-interval etc., and the following formula is used to calculate the median.

$$M = L + \frac{\frac{n}{2} - C}{f} \times i, \text{ where}$$

$L$ = lower limit of the class in which the median lies

$n$ = total number of frequencies, i.e., $n = \Sigma f$

$f$ = frequency of the class in which the median lies

$C$ = cumulative frequency of the class proceeding the median class

$i$ = width of the class-interval of the class in which the median lies.

**Illustration:** The following table gives the marks obtained by students in Economics. Find the median.

| Marks | No. of Students |
|---|---|
| 10–14 | 4 |
| 15–19 | 6 |
| 20–24 | 10 |
| 25–29 | 5 |
| 30–34 | 7 |
| 35–39 | 3 |
| 40–44 | 9 |
| 45–49 | 6 |

**Solution:** Let us prepare the following table showing the frequencies and cumulative frequencies.

| Marks | Frequency | Cumulative Frequencies |
|---|---|---|
| 10–14 | 4 | 4 |
| 15–19 | 6 | 10 |
| 20–24 | 10 | 20 |
| 25–29 | 5 | 25 |
| 30–34 | 7 | 32 |
| 35–39 | 3 | 35 |
| 40–44 | 9 | 44 |
| 45–49 | 6 | 50 |

Here $n = 50$, $\therefore \frac{n}{2} = 25.$

It is evident that 25th item is the median and in the table the cumulative frequency which contains the 25th item lies in the interval 25 –29.

Median class = 25 – 29

Now the given series is inclusive one. Let us convert it to exclusive series and hence the value of L will be $\frac{25+24}{2} = 24.5$

Also, $\frac{n}{2} = 25$

$\therefore$ L = lower limit of the median class = 24.5

C = cumulative frequency of the class (20–24) preceding the median class = 20

$f$ = frequency of the median class = 5

$i$ = class-interval of the median class = 5

$$\therefore \text{Median} = 24.5 + \frac{25-20}{5} \times 5$$

$$= 24.5 + 5 = 29.5$$

**Measures of Dispersion.** Dispersion means scatterness. The degree to which numerical data tend to spread about an average value is called the dispersion of the data. The term dispersion is used in two senses. The first relates to the limits within which the data fall and the second takes into account the amount, absolute or relative, by which the value of item differ from an average.

**Mean Deviation:** It is the average of the modulus of the deviations of the observations in a series taken from mean or median.

**Methods for Calculation of Mean Deviations**

(*a*) **For ungrouped data:** In this case the mean deviation is given by the formula

$$\text{Mean deviation} = \frac{\Sigma\,|x - A|}{n} = \frac{\Sigma\,|d|}{n}$$

where '*d*' stands for the deviation from the mean or median and | d | is always positive whether *d* itself is positive or negative and $n$ is the total number of items.

**Illustration:** Find the mean deviation from (*i*) mean and (*ii*) median for the following data:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Marks** | : 20 | 18 | 16 | 14 | 12 | 10 | 8 | 6 |
| **No. of Students** | : 2 | 4 | 9 | 18 | 27 | 25 | 14 | 1 |

**Solution:** (*i*) **Mean deviation from mean:** Let us calculate the mean of the given data by forming the following table:

| Marks ($x$) | No. of Students ($f$) | $f \times x$ |
|---|---|---|
| 6 | 1 | 6 |
| 8 | 14 | 112 |
| 10 | 25 | 250 |
| 12 | 27 | 324 |
| 14 | 18 | 252 |
| 16 | 9 | 144 |
| 18 | 4 | 72 |
| 20 | 2 | 40 |
| | $\Sigma f = 100$ | $\Sigma f \times x = 1200$ |

$$\therefore \text{ Arithmetic mean } = \frac{1200}{100} = 12$$

We shall now proceed to calculate the mean deviation. Let us prepare the following table for it.

| $(x)$ | $(f)$ | $\lvert d \rvert = \lvert x - 12 \rvert$ | $f \lvert d \rvert$ |
|---|---|---|---|
| 6 | 1 | 6 | 6 |
| 8 | 14 | 4 | 56 |
| 10 | 25 | 2 | 50 |
| 12 | 27 | 0 | 0 |
| 14 | 18 | 2 | 36 |
| 16 | 9 | 4 | 36 |
| 18 | 4 | 6 | 24 |
| 20 | 2 | 8 | 16 |
| | $\Sigma f = 100$ | | $\Sigma f \lvert d \rvert = 224$ |

∴ Mean deviation (about mean)

$$= \frac{\Sigma f \lvert d \rvert}{n} = \frac{224}{100} = 2.24$$

(*ii*) **Mean deviation about median:** Let us prepare the following table in order to calculate the median:

| Marks $(x)$ | Frequency | Cumulative frequency |
|---|---|---|
| 6 | 1 | 1 |
| 8 | 14 | 15 |
| 10 | 25 | 40 |

| | | |
|---|---|---|
| 12 | 27 | 67 |
| 14 | 18 | 85 |
| 16 | 9 | 94 |
| 18 | 4 | 98 |
| 20 | 2 | 100 |

Now Median = Average of $\frac{n}{2}$th and $\left(\frac{n}{2}+1\right)$th item.

= Average of 50th and 51st item = 12

Let us now prepare the following table in order to calculate the mean deviation from median:

| $x$ | $f$ | $\lvert d \rvert = \lvert x - 12 \rvert$ | $f \times \lvert d \rvert$ |
|---|---|---|---|
| 6 | 1 | 6 | 6 |
| 8 | 14 | 4 | 56 |
| 10 | 25 | 2 | 50 |
| 12 | 27 | 0 | 0 |
| 14 | 18 | 2 | 36 |
| 16 | 9 | 4 | 36 |
| 18 | 4 | 6 | 24 |
| 20 | 2 | 8 | 16 |
| | $\Sigma f = 100$ | | $\Sigma f \times \lvert d \rvert = 224$ |

$\therefore$ Mean deviation (about median) = $\frac{224}{100} = 2.24$

(*b*) **Mean Deviation for grouped data:** Let $x_1, x_2, x_3, \ldots\ldots x_n$ occur with frequency $f_1, f_2, f_3, \ldots\ldots f_n$ respectively and let $\Sigma f = n$ and $M$ can be either Mean or Median then the mean deviation is given by the formula.

**Mean Deviation =** $\dfrac{\Sigma f|x - M|}{\Sigma f} = \dfrac{\Sigma f|d|}{n}$

where $d = |x - M|$ and $\Sigma f = n$.

**Illustration:** Calculate the mean deviation from the mean for the following data:

| | | | | | |
|---|---|---|---|---|---|
| Class-interval : | 0–4 | 4–8 | 8–12 | 12–16 | 16–20 |
| Frequency : | 4 | 6 | 8 | 5 | 2 |

**Solution:** Let us prepare the following table by assuming that the frequencies in each class are centred at its mid-value.

| **Class-interval** | **Mid-value** | ***f*** | ***fx*** | $\lvert d\rvert = \lvert x-9.2\rvert$ | $f\lvert d\rvert$ |
|---|---|---|---|---|---|
| 0–4 | 2 | 4 | 8 | 7.2 | 28.8 |
| 4–8 | 6 | 6 | 36 | 3.2 | 19.2 |
| 8–12 | 10 | 8 | 80 | 0.8 | 6.4 |
| 12–16 | 14 | 5 | 70 | 4.8 | 24.0 |
| 16–20 | 18 | 2 | 36 | 8.8 | 17.6 |
| | $\Sigma f = 25$ | | $\Sigma fx = 230$ | $\Sigma f\lvert d\rvert = 96.0$ | |

Now A.M. $= \frac{\Sigma fx}{\Sigma f}$, $\therefore$ A.M. $= \frac{230}{25} = 9.2$

Also, Mean Deviation (about mean) $= \frac{\Sigma f|d|}{\Sigma f}$

$\therefore$ Mean deviation $= \frac{96}{25} = 3.84$

**Merits and Demerits of Mean Deviation**

**Merits:**

1. It is easy to understand and compute.
2. Mean deviation is less affected by the extreme values as compared to standard deviation.
3. Mean deviation about an arbitrary point is least when the point is median.

**Demerits:**

1. In mean deviation the signs of all deviations are taken as positive and therefore, it is not suitable for further algebraic treatment.
2. It is rarely used in social sciences.
3. It does not give accurate results because the mean deviation from the median is least but median itself is not considered a satisfactory average when the variation in the series is large.

**Standard Deviation:** The positive square root of the average of squared deviations of all observations taken from their mean is called standard deviation. It is generally denoted by the Greek alphabet $\sigma$ or $s$.

**Variance:** The square of the standard deviation is called variance and is denoted by $\sigma^2$.

**Coefficient of Standard Deviation:** It is the ratio of the standard deviation to its A.M. *i.e.*,

$$\text{Coefficient of standard deviation} = \frac{\sigma}{\bar{x}}$$

**Coefficient of Variance:** It is the product of coefficient of standard deviation × 100.

$$\text{Coefficient of variance} = \frac{\sigma}{\bar{x}} \times 100.$$

**Computation of Standard Deviation:** The methods of calculating the standard deviation depend upon the nature of data and also on the number of observations for ungrouped data.

(*a*) **Standard Deviation for Ungrouped Data:**

**Direct method:** In case of simple series, the standard deviation can be obtained by the formula:

$$\sigma = \sqrt{\frac{(x-\overline{x})^2}{n}} = \sqrt{\frac{\Sigma d^2}{n}}, \text{ where } d = x - \overline{x}$$

and $x$ = value of the variable or observation, $\overline{x}$ = arithmetic mean, $n$ = total number of observations.

**Illustration:** Find the standard deviation of 16, 13, 17, 22.

**Solution:** Here

$$\text{A.M.} = \overline{x} = \frac{16+13+17+22}{4} = \frac{68}{4} = 17.$$

Let us prepare the following table in order to calculate the standard deviation.

| $(x)$ | $d = x - \overline{x} = x - 17$ | $(x-\overline{x})^2$ |
|---|---|---|
| 16 | – 1 | 1 |
| 13 | – 4 | 16 |
| 17 | 0 | 0 |
| 22 | 5 | 25 |
| | | $\Sigma d^2 = 42$ |

Now $\sigma = \sqrt{\frac{\Sigma(x-\overline{x})^2}{n}} = \sqrt{\frac{42}{4}} = 3.2$

**Short-cut Method:** This method is applied to calculate standard deviation, when the mean of the data comes out to be a fraction. In that case it is very difficult and tedious to find the deviations of all observations from the mean by the earlier method. The formula used is

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2},$$

where $d = x - A$, $A$ = assumed mean, $n$ = total number of observations.

**Illustration:** Find the standard deviation of the following data:

48, 43, 65, 57, 31, 60, 37, 48, 59, 78.

**Solution:** Let us prepare the following table in order to calculate the value of S.D:

| **Value ($x$)** | $d = x - A, (A = 50)$ | $d^2$ |
|---|---|---|
| 48 | – 2 | 4 |
| 43 | – 7 | 49 |
| 65 | 15 | 225 |
| 57 | 7 | 49 |
| 31 | – 19 | 361 |
| 60 | 10 | 100 |

| | | |
|---|---|---|
| 37 | – 13 | 169 |
| 48 | – 2 | 4 |
| 59 | 9 | 81 |
| 78 | 28 | 784 |
| $n = 10$ | $\Sigma d = 26$ | $\Sigma d^2 = 1826$ |

Here, $\bar{x} = A + \frac{\Sigma d}{n} = 50 + \frac{26}{10} = 52.6$, which is fraction. Let us apply the short-cut formula in order to calculate S.D.

$$\therefore \text{ S.D.} = \sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{1826}{10} - \left(\frac{26}{10}\right)^2}$$

$$= \sqrt{182.60 - 6.76} = \sqrt{175.84} = 13.26$$

(*b*) **Standard Deviation for Grouped Data or Discrete Series.**

**Direct Method:** The standard deviation for the discrete series is given by the formula.

$$\sigma = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{n}},$$

where $\bar{x}$ is A.M., $x$ is the size of the item, and $f$ is the corresponding frequency in the case of discrete series. But when the mean has a fractional value, then the following formula is applied to calculate S.D.

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2}$$

where $d = x - A$, $A$ = assumed mean, $n = \Sigma f$ = Total frequency.

**Illustration:** Find the standard deviation from the following data:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Size of the item* : | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| *Frequency* : | 2 | 7 | 11 | 15 | 10 | 4 | 1 |

Also, find the coefficient of variation.

**Solution:** Let us prepare the following table:

| ***Size of the (x) item*** | ***Frequency f*** | ***d = x − A, A = 13*** | ***fd*** | ***fd²*** |
|---|---|---|---|---|
| 10 | 2 | − 3 | − 6 | 18 |
| 11 | 7 | − 2 | − 14 | 28 |
| 12 | 11 | − 1 | − 11 | 11 |
| 13 | 15 | 0 | 0 | 0 |
| 14 | 10 | 1 | 10 | 10 |
| 15 | 4 | 2 | 8 | 16 |
| 16 | 1 | 3 | 3 | 9 |
| Total | $n = \Sigma f = 50$ | | $\Sigma fd = -10$ | $\Sigma fd^2 = 92$ |

Now A.M. = $\bar{x} = A + \frac{\Sigma fd}{n} = 13 + \frac{(-10)}{50} = 12.8$

Here $\bar{x}$ = 12.8 is a fraction,

$$\therefore \text{ S.D.} = \sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} = \sqrt{\frac{92}{50} - \left(-\frac{10}{50}\right)^2}$$

$$= \sqrt{1.84 - 0.04} = \sqrt{1.80} = 1.342$$

$\therefore$ Now the coefficient of variation

$$= \frac{\sigma}{x} \times 100 = \frac{1.342}{12.8} \times 100 = 10.4$$

**(c) Standard Deviation in Continuous Series:**

**Direct Method:** The standard deviation in the case of continuous series is obtained by the following formula

$$\sigma = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{n}},$$

where $x$ = mid-value, $\bar{x}$ = A.M., $f$ = frequency, $n$ = total frequency.

**Illustration:** Calculate the standard deviation for the following frequency distribution.

| **Class Interval :** | 0–4 | 4–8 | 8–12 | 12–16 |
|---|---|---|---|---|
| **Frequency :** | 4 | 8 | 2 | 1 |

**Solution:** Assuming that frequency at each class is centred at its mid-value, let us prepare the following table:

### Table for Arithmetic Mean

| Class Interval | Mid-value ($x$) | Frequency $f$ | Product $f \times x$ |
|---|---|---|---|
| 0–4 | 2 | 4 | 8 |
| 4–8 | 6 | 8 | 48 |
| 8–12 | 10 | 2 | 20 |
| 12–16 | 14 | 1 | 14 |
| | | $\Sigma f = 15$ | $\Sigma fx = 90$ |

$$\text{A.M.} = \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{90}{15} = 6$$

**Table for Standard Deviation**

| Mid-value $x$ | Frequency $f$ | $x - \bar{x}$ $= x - 6$ | $(x - \bar{x})^2$ $= (x - 6)^2$ | $f(x - \bar{x})^2$ $= f(x - 6)^2$ |
|---|---|---|---|---|
| 2 | 4 | – 4 | 16 | 64 |
| 6 | 8 | 0 | 0 | 0 |
| 10 | 2 | 4 | 16 | 32 |
| 14 | 1 | 8 | 64 | 64 |
| | $\Sigma f = 15$ | | | $\Sigma f(x - \bar{x})^2 = 160$ |

$$\therefore \text{SD} = \sqrt{\frac{\Sigma f(x - x)^2}{n}} = \sqrt{\frac{160}{15}} = \sqrt{10.67} = 3.27$$

**Short Method:** The formula for short method to find the standard deviation of continuous series is

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} \times i \text{ ,where } d = \frac{x - A}{i},$$

A = Assumed mean, $n$ = total frequency,
$i$ = class width.

**Illustration.** Find the standard deviation for the following distribution.

| Marks | 10–20 | 20–30 | 30–40 | 40–50 | 50–60 | 60–70 | 70–80 |
|---|---|---|---|---|---|---|---|
| No. of Students | 5 | 12 | 15 | 20 | 10 | 4 | 2 |

**Solution:** Let us prepare the following table in order to calculate the standard deviation.

| Mark (Class interval) | No. of students ($f$) | Mid-value ($x$) | $d = \frac{x-45}{10}$ | $fd$ | $fd^2$ |
|---|---|---|---|---|---|
| 10–20 | 5 | 15 | –3 | –15 | 45 |
| 20–30 | 12 | 25 | –2 | –24 | 48 |
| 30–40 | 15 | 35 | –1 | –15 | 15 |
| 40–50 | 20 | 45 | 0 | 0 | 0 |
| 50–60 | 10 | 55 | 1 | 10 | 10 |
| 60–70 | 4 | 65 | 2 | 8 | 16 |
| 70–80 | 2 | 75 | 3 | 6 | 18 |
| Total | $\Sigma f = n$ = 68 | | | $\Sigma fd$ = –30 | $\Sigma fd^2$ = 152 |

$$\therefore \sigma = i \times \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} = 10 \times \sqrt{\frac{152}{68} - \left(\frac{-30}{68}\right)^2}$$

$= 14.3$ Approx.

## Merits and Demerits of Standard Deviation

**Merits:**

1. It is based on all the observations.
2. It is rigidly defined.
3. It lend itself to further algebraic treatment.
4. It is less affected by fluctuations of sampling as compared to other measures of dispersion.
5. It is extremely useful in correlation.
6. Like mean deviation, there is no artificiality in it.

**Demerits:**

1. It is difficult to compute unlike other measures of dispersion.
2. It is not simple to understand.
3. It gives more weightage to extreme values.