

# Statistics

## Some Basic Definitions

### i. Variable

- A quantity which takes different values at stated condition is called *variable* and the different values taken by the variable are called *variate values*. The height of the men in a city, the wages of the workers of a factory and the price of the goods in a shop are some examples of variables. The variables are usually denoted by  $X, Y, Z, x, y, z$  etc. The variate values of the variable  $X$  are  $x_1, x_2, x_3, \dots, x_n$ . A variable is of two types:

- Continuous variable and
- Discrete variable

A variable which can theoretically assume any value (integral or fractional) between two given values is called a *continuous variable*, otherwise it is called a *discrete variable*. For example, the height ( $H$ ) in cm of persons in a town is a continuous variable because it can take all possible values in a certain range. The height may be 140 cm, 148.2 cm, 159.33 cm and so on. The number ( $N$ ) of children in a family can take certain values like 0, 1, 2, 3, and 4. But it cannot be a fractional number like 1.4.  $N$  is thus a discrete variable. The number of goals scored in a football match is also a discrete variable.

### ii. Raw Data and Array

The data obtained in original form are called *raw data* or *ungrouped data*. Suppose 40 students of a class obtained the following marks out of 10:

4, 3, 5, 1, 0, 3, 8, 2, 6, 5, 4, 6, 5, 4, 5, 6, 7, 6, 9, 6,  
5, 10, 3, 6, 5, 7, 6, 4, 4, 2, 8, 1, 3, 8, 7, 9, 4, 5, 7, 6

The data are original in form and not organised in any way. In this form they are called *raw data*. In the raw form the data are confusing and they cannot give any useful information. If we rearrange the data in ascending or descending order, the new arrangement is called an *array*. An arrangement of the above data in the form of array is given below:

0, 1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5,  
5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 9, 9, 10

It is a better arrangement, but it is also not of much use. Such data will be useful and comprehensive if they are presented in the form of frequency distribution.

### iii. Frequency

The number of times an observation (i.e., a variate value) occurs is called the *frequency*. It is denoted by  $f$ . For example,

Consider the following set of observations

5, 2, 3, 3, 1, 2, 2, 4, 4, 3, 5, 1, 2, 3, 1.

Here, 1 appears 3 times, so frequency of 1 is 3. Similarly, 2 appears 4 times, so frequency of 2 is 4 and so on. The frequency table may be presented as under.

Variable	1	2	3	4	5
Frequency	3	4	4	2	2



#### iv. Frequency Distribution

Frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups, with their corresponding frequencies side by side. In other words, a tabular arrangement of data showing the frequency of each observation is called a *frequency distribution*. Thus, a frequency distribution has two parts, on its left there are sizes or magnitudes of values and on its right the number of times a value or a group of values has repeated. There are two types of frequency distribution namely *univariate* and *bivariate frequency distribution*.

All the frequency distributions showing only one variable are called a *univariate frequency distribution*. A univariate frequency may be, further, divided into the following three types.

- Individual series or Arrayed series.
- Discrete series or ungrouped frequency distribution.
- Continuous series or grouped frequency distribution.

(a) **Individual series or arrayed series:** A series where items are listed individually rather than keeping in groups is called *individual series*. In other words, the set of individual data (unorganised or organised) is called *individual series*. For example, the temperatures on different months form individual series.

Month	Baisakh	Jestha	Ashadh	Shrawan	Bhadra
Temp (°C)	33	31	35	36	34

(b) **Discrete series or ungrouped frequency distribution:** A discrete frequency distribution is the arrangement of data in a table showing the frequency with which each successive value of a discrete variable occurs. For example,

No. of accident	0	1	2	3	4	5
No. of days	10	15	20	8	5	2

Here, 10 days occur zero accident; 15 days occur 1 accident and so on. Since the variable (accident) can take some specific values, the data are discrete data. Hence, a series formed by a discrete variable is called a *discrete series*. In this form of discrete frequency distribution, the frequency refers to a given discrete value.

(c) **Continuous series or grouped frequency distribution:** In this form of distribution, the frequencies refer to groups of values where values of variable can take any fractional value within limits (interval) and an exact measurement is not possible. Such variable is called *continuous variable*. Hence, the series formed by a continuous variable is called *continuous series*. It is also called continuous frequency distribution. For example,

Marks obtained	0-20	20-40	40-60	60-80	80-100
No. of students	3	6	8	5	2

#### v. Open end classes

If the lower limit of the first class or the upper limit of the last class or both are not specified and such classes in which one of the limits is missing are called open end classes. In this case the value at one end is not mentioned and classes are written as "Below 10" or "Above 90". They are known as *open end classes*. The mid-value of such a class cannot be determined and further mathematical computation becomes a problem. It will also be difficult to present the data graphically. Open end classes are therefore avoided as far as possible. For examples,

(a) Salary (in Rs.) :	Below 200	200-300	300-400	Above 400
No. of workers :	5	7	8	3

(b) Temp. (in °C) :	Below 5	5-10	10-15	15-20
No. of days :	3	6	8	2

#### vi. Range

The difference between the maximum and minimum observations is called *range*.

Approved by Curriculum Development Centre (CDC), Nepal

#### vii. Mid-

The v  
sum o

Thu

#### viii. Class

The en  
lower  
lower

### 5.2 Types of

There are  
inclusive

(i) Exclu  
class in

No. of stu

In the above

second class

Thus, such

immediate

ensures con

(ii) Includi

then the

M

No. of stud

In the above

second class

lower and th

included in t

When the cl

lower limit o

To ensure co

order to chan

adjustments a

(i) Find the

$$C_f = \frac{\text{low}}{\text{low}}$$

(ii) Subtract this  
required exclu

Real lowe

and real uppe

Here, the real

In the above ex



**vii. Mid-value of the class**

The value exactly at the middle of a class-interval is called the mid-value. It is obtained by dividing the sum of the lower and upper limits of a class by two. It is also called its *class mark* or *mid-point*.

$$\text{Thus, Mid-value} = \frac{\text{upper limit} + \text{lower limit}}{2}$$

$$\text{e.g., in the class (20-30), mid-value} = \frac{30+20}{2} = 25.$$

**viii. Class-limits**

The end points of a class interval are called *class limits*. The smallest observation of the interval is called *lower limit* and the largest observation of the interval is called *upper limit*. In the class interval 20-25, the lower limit is 20 and the upper limit is 25.

**5.2 Types of Class Intervals**

There are two methods of classifying the data according to class-intervals, namely exclusive type classes and inclusive type classes.

- (i) **Exclusive type classes:** When the upper limit of each class is the lower limit of the next class, then the class intervals are called *exclusive type classes*. For example,

Marks	0-20	20-40	40-60	60-80	80-100
No. of students	3	6	8	5	2

In the above table, the first class (0 - 20) is of those students who secured marks from 0 to 19. Similarly, the second class (20 - 40) is of those students who secured marks from 20 to 39 and so on.

Thus, such classes in which upper limits are excluded from the respective classes and are included in the immediate next class are called as *exclusive classes*. As the classes are formed without any gap, this method ensures continuity of data.

- (ii) **Inclusive type classes:** When the upper limit of none of the classes is the lower limit of the next class, then the class intervals are called *inclusive type classes*. For example,

Marks	20-29	30-39	40-49	50-59	60-69
No. of students	3	6	8	5	2

In the above table, first class (20 - 29) is of those students who secured marks 20 and up to 29. Similarly, the second class (30 - 39) is of those students who secured marks 30 and up to 39 and so on. In this method, both lower and the upper limits are included in a particular class. Since the upper limit of each class interval is included in that interval, such types of class intervals are called *inclusive classes*.

When the class intervals are inclusive type classes, there is a gap between the upper limit of a class and the lower limit of the next class.

To ensure continuity and to get correct class interval, we should adopt 'exclusive' method of classification. In order to change some inclusive classes to exclusive ones and to obtain continuous distribution, the following adjustments are needed.

- (i) Find the correction factor which is denoted by  $C_f$  and it is given by

$$C_f = \frac{\text{lower limit of the 2}^{\text{nd}} \text{ class} - \text{upper limit of the 1}^{\text{st}} \text{ class}}{2}$$

- (ii) Subtract this correction factor from the lower limit and add it to the upper limit of each class to get the required exclusive classes. Thus,

$$\text{Real lower limit} = \text{lower limit} - C_f$$

$$\text{and real upper limit} = \text{upper limit} + C_f$$

Here, the real upper and lower limits are said to be the *class boundaries* of new exclusive type classes.

$$\text{In the above example, the correction factor } (C_f) = \frac{30-29}{2} = \frac{40-39}{2} = 0.5$$



Subtracting 0.5 from the lower limits of each class and adding it to the upper limits of each class, the exclusive type classes will be as follows.

Marks	Frequency	Exclusive class
20 – 29	3	19.5 – 29.5
30 – 39	6	29.5 – 39.5
40 – 49	8	39.5 – 49.5
50 – 59	5	49.5 – 59.5
60 – 69	2	59.5 – 69.5
Total	24	

(iii) Sometimes classes are defined by their mid values. We can find the class boundaries in the following way:

- (a) Find the class interval by taking the difference between any two consecutive mid values. Consider the following example.

Marks (Mid values)	10	20	30	40	50
No. of students	3	6	8	5	2

Here, class interval =  $20 - 10 = 30 - 20 = \dots = 10$

- (b) Divide the class interval by 2. Thus, we have  $\frac{10}{2} = 5$ .
- (c) Add this result 5 to each mid value and subtract from each mid value to get the class boundaries. Thus, the class boundaries become

$$(10 - 5) - (10 + 5) = (5 - 15)$$

$$(20 - 5) - (20 + 5) = (15 - 25) \text{ and so on.}$$

Thus, the above continuous frequency distribution of exclusive type is as follows

Marks (Mid-values)	Frequency	Class
10	3	5 – 15
20	6	15 – 25
30	8	25 – 35
40	5	35 – 45
50	2	45 – 55
Total	24	

### 5.3 Cumulative Frequency Distribution

A frequency distribution in which frequencies are cumulated either from top to bottom or bottom to top is called *cumulative frequency distribution*. In other words, a frequency distribution showing the cumulative frequencies against values of the variable systematically arranged in increasing (or decreasing) order is known as *cumulative frequency distribution*. In a cumulative frequency distribution, the cumulative frequencies (c.f.) are obtained by the cumulating (successively adding) of the frequencies of the successive individual class intervals. The cumulative frequency of a given class interval thus represents the total of all the previous class frequencies including the class against which it is written. Sometimes, we have to change an ordinary frequency table into a cumulative frequency table. There are two types of cumulative frequency distribution namely less than cumulative frequency distribution and more than cumulative frequency distribution.

- (i) **Less than cumulative frequency distribution:** When the classes (or values) are arranged in ascending order of magnitude and frequencies are cumulated from the top to bottom, frequencies thus cumulated are called *less than cumulative frequency* and the distribution having less than cumulative frequency is called *less than cumulative frequency distribution*. When we cumulate the frequencies from the top, we take the upper limits of the class intervals and write the word "below or less than or under or up to" before them. Let us consider the following ordinary frequency distribution.

Approved by Curriculum Development Centre (CDC), Nepal



Marks obtained	0-20	20-40	40-60	60-80	80-100
No. of students	3	6	8	5	2

Here, 3 students get (0 – 20) marks, 6 students get (20 – 40) marks and so on. As each class has individual frequency, this is the simple frequency.

Again, the total number of students with marks less than 40 is  $3+6=9$ .

Similarly, the total number of students with marks less than 60 is the sum of all the previous frequencies upto and including the class  $(40-60) = 3+6+8=17$  and so on.

The final distribution is given below.

**Less than cumulative frequency distribution of marks**

Marks	Frequency ( $f$ )	Less than cumulative frequency (c.f.)
0 – 20	3	3
20 – 40	6	$3+6=9$
40 – 60	8	$9+8=17$
60 – 80	5	$17+5=22$
80 – 100	2	$22+2=24$

The above 'less than' cumulative frequency distribution can also be written as follows

Marks	Cumulative frequency
Less than 20	3
Less than 40	9
Less than 60	17
Less than 80	22
Less than 100	24

- (ii) **More than cumulative frequency distribution:** When classes (or variate values) are arranged in ascending order of magnitude and the frequencies are cumulated from bottom to top, the frequencies thus cumulated are called *more than cumulative frequency* and the distribution having more than cumulative frequency is called *more than cumulative frequency distribution*.

When we cumulate the frequencies from the bottom to top, we take the lower limits of the class intervals and write the word "Above or more than or over" before them. Thus, in above example, the total no. of students getting marks more than 80 is  $2+5=7$ .

Similarly, the total no. of students getting marks more than 60 is  $2+5+8=15$  and so on.

The final distribution is given below.

**More than cumulative frequency distribution of marks**

Marks	Frequency ( $f$ )	More than cumulative frequency (c.f.)
0 – 20	3	$21+3=24$
20 – 40	6	$15+6=21$
40 – 60	8	$7+8=15$
60 – 80	5	$2+5=7$
80 – 100	2	2

The above more than c.f. distribution can also be written as follows

Marks	Cumulative frequency
More than 0	24
More than 20	21
More than 40	15
More than 60	7
More than 80	2



**Remarks:**

- (i) In less than cumulative frequency distribution, frequencies are cumulated from top to bottom and only the upper limit of each class is written. In more than cumulative frequency distribution, frequencies are cumulated from bottom to top and only the lower limit of each class is written.
- (ii) In fact 'less than' and 'more than' words also include the equality sign i.e. 'less than a given value' means 'less than or equal to that value' and 'more than a given value' means 'more than or equal to that value.'

## 5.4 Construction of Ordinary Frequency Distribution from Cumulative Frequency Distribution

As stated above the less than or more than cumulative frequency is obtained from an ordinary frequency, likewise an ordinary frequency (simple frequency) may also be obtained from given cumulative frequency distribution. In the case of less than cumulative frequency distribution, the frequencies of individual class intervals can always be obtained by deducting the totals of the previous class or classes from the cumulative frequency of the given class. It may be noted that the cumulative frequency of the first class interval is the frequency itself. However, in the case of more than cumulative frequency distribution, the frequencies of individual class intervals can be obtained by subtracting the totals of the succeeding class or classes from the preceding frequencies. In this case, the cumulative frequency of the last class interval is the frequency itself.

Let us consider the following less than cumulative frequency table.

Marks obtained	No. of students
Below 20	5
Below 40	15
Below 60	30
Below 80	40
Below 100	45

In the above table, cumulative frequencies are given, 5 students get marks below 20. As the class intervals except the first are 20, the first class interval should be (0 – 20). So, the number of students who get (0 – 20) marks is 5.

Again, there are 5 students below 40 marks and 5 students below 20 marks. So, there should be  $15 - 5 = 10$  students who get marks (20 – 40).

Similarly, the number of students who get (40 – 60) marks =  $30 - 15 = 15$  and so on.

In this way, the simple frequencies are obtained by deducting the first from the second, second from third and so on. Thus, the simple frequency table will be as follows:

Marks	No. of students
0 – 20	5
20 – 40	$15 - 5 = 10$
40 – 60	$30 - 15 = 15$
60 – 80	$40 - 30 = 10$
80 – 100	$45 - 40 = 5$

Again, let us consider the following more than cumulative frequency table.

Marks obtained	No. of students
Above 0	45
Above 20	40
Above 40	30
Above 60	15
Above 80	5

In the above table, the class sizes of all class intervals except last are 20. So, the last class interval should be (80 – 100). Hence, the class intervals will be (0 – 20), (20 – 40), (40 – 60), (60 – 80) and (80 – 100). The corresponding frequencies will be obtained by subtracting the second from the first, the third from the second

Approved by Curriculum Development Centre (CDC), Nepal

and so on.  
distribution

## 5.5 Measures

After the data objectives of huge and unw Thus, the ave the entire dat i.e., the large part of the d called measu

## 5.6 Types of

The follow  
a. Arithmet  
b. Median  
But in curr

## 5.7 Arithme

The most p  
'the mean'.  
(1) Simple  
(1) Simple  
arithm  
4, 6 and

Calculati  
Individu  
(i) Direct

where  $\Sigma$   
(ii) Short  
defin

wher



and so on. Thus, the simple frequency distribution obtained from the more than cumulative frequency distribution is as follows.

Simple frequency table	
Marks	No. of students
0 – 20	45 – 40 = 5
20 – 40	40 – 30 = 10
40 – 60	30 – 15 = 15
60 – 80	15 – 5 = 10
80 – 100	5

## 5.5 Measures of Central Tendency

After the data have been classified and tabulated, the next step is to analyse it. One of the most important objectives of statistical analysis is to get one single value that describes the characteristic of the entire mass of huge and unwieldy data. Such a value is called the *central value* or an *average* of the entire distribution.

Thus, the averages are the measures which condense a huge unwieldy data into a single value which represents the entire data. Since an average represents the entire data, its value lies between the two extreme observations, i.e., the largest and the smallest items. This tendency of the items or values of the items to cluster in the central part of the distribution is known as the *Measures of Central Tendency*. Measures of central tendency are also called *measures of location* since they enable us to locate the position or place of the distribution.

## 5.6 Types of Averages

The following are the important types of averages:

- Arithmetic mean: (i) Simple and (ii) Weighted
- Median
- Mode
- Geometric mean
- Harmonic mean

But in curriculum, arithmetic mean, median and mode are included only.

## 5.7 Arithmetic Mean

The most popular and widely used measure of central tendency is the arithmetic mean. It is also called simply 'the mean'. Arithmetic mean may either be

- Simple arithmetic mean, or
- Weighted arithmetic mean

(1) **Simple arithmetic mean:** The sum of all the observations divided by the number of observations is called arithmetic mean. In such cases all the items are equally important. For example, the arithmetic mean of 2, 4, 6 and 8 is

$$\frac{2 + 4 + 6 + 8}{4} = \frac{20}{4} = 5$$

### Calculation of Arithmetic Mean

#### Individual Series

(i) **Direct method:** The arithmetic mean, usually denoted by  $\bar{X}$  is defined by the following formula.

$$\bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + \dots + X_n}{n} = \frac{\sum X}{n} \quad \dots (1)$$

where  $\sum X$  = the sum of observations,  $n$  = no. of observations

(ii) **Short-cut method or assumed mean method:** The formula for calculating the mean by this method is defined by

$$\bar{X} = A + \frac{\sum d}{n} \quad \dots (2)$$

where  $A$  = assumed mean.

$d = X - A$  = deviations of the items from the assumed mean

$\sum d$  = sum of the deviations of the items.

$n$  = no. of observations.



**Example 1:** The following are the monthly incomes of 5 families in a certain locality.

Family	1	2	3	4	5
Income (Rs)	300	500	900	1500	800

Calculate the arithmetic mean by (i) direct method (ii) short-cut method.

**Solution:**

Let  $X$  be the incomes. Then calculation of A.M.

Family	Incomes ( $X$ )	$d = X - A = X - 900$
1	300	-600
2	500	-400
3	900	0
4	1500	600
5	800	-100
Total	$\Sigma X = 4000$	$\Sigma d = -500$

**For direct method:**

Here,  $\Sigma X = 4000$ ,  $n = 5$ ,  $\bar{X} = ?$

$$\text{Now, } \bar{X} = \frac{\Sigma X}{n} = \frac{4000}{5} = 800$$

Hence, the average income is Rs. 800.

**For short-cut method:**

Here,  $A = 900$ ,  $\Sigma d = -500$ ,  $n = 5$ ,  $\bar{X} = ?$

$$\begin{aligned} \text{Now, } \bar{X} &= A + \frac{\Sigma d}{n} = 900 + \frac{-500}{5} \\ &= 900 - 100 = 800 \end{aligned}$$

Hence, the average income is Rs. 800

**Remark:** Any number can serve the purpose of the arbitrary number 'A', but generally the value of items  $X$  corresponding to the middle part of the distribution will be more convenient. In fact, 'A' need not necessarily be one of the values of the items  $X$ .

### Discrete Series

(i) **Direct method:** The arithmetic mean is defined by following formula:

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + f_3 X_3 + \dots + f_n X_n}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\Sigma fX}{\Sigma f} = \frac{\Sigma fX}{N} \quad \dots (3),$$

where  $N = \Sigma f$  = total frequency

(ii) **Short-cut method or assumed mean method:** The formula for calculating arithmetic mean  $\bar{X}$  by this method is defined by

$$\bar{X} = A + \frac{\Sigma fd}{N} \quad \dots (4),$$

where  $A$  = Assumed mean

$d = X - A$  = deviation of the items from the assumed mean

$N$  = total frequency.

(iii) **Step-deviation method or coding method:** When common factor is taken from all the items of observations, then step-deviation method is used to calculate the arithmetic mean. Thus, the formula for calculating the arithmetic mean is defined by

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times h \quad \dots (5),$$

where  $h$  = common factor

$A$  = Assumed mean

$$d' = \frac{X - A}{h}$$

**Example 2:** Calculate the arithmetic mean from the following data by (i) direct method (ii) short-cut method (iii) step-deviation method.

**Solution:**

Let  $X$  denote

Calcul

Marks ( $X$ )
5
10
15
20
25

(i) For dir

Here,  $N = \Sigma$

Now,

(iii) For Ste

Here,  $N$

Now,  $\Sigma$

Hence, the a

**Remark:** In the calcula distrib

**Continuou**

(i) Direct m

$\bar{X} =$

$N =$

(ii) Short-c method

$\bar{X}$

(iii) Step de

$\bar{X}$

In the case of necessary to

**Example 3:** From

(i) di



Marks obtained	5	10	15	20	25
No. of students	4	6	8	5	2

**Solution:**

Let  $X$  denotes the marks obtained by the students.

**Calculation of A.M. by direct method, Short-cut method and step-deviation method**

Marks (X)	No. of students (f)	fX	d = X - A = X - 15	fd	$d' = \frac{X - 15}{5}$	fd'
5	4	20	-10	-40	-2	-8
10	6	60	-5	-30	-1	-6
15	8	120	0	0	0	0
20	5	100	5	25	1	5
25	2	50	10	20	2	4
	$N = \sum f = 25$	$\sum fX = 350$		$\sum fd = -25$		$\sum fd' = -5$

(i) For direct method:

Here,  $N = \sum f = 25$ ,  $\sum fX = 350$ ,  $\bar{X} = ?$

$$\text{Now, } \bar{X} = \frac{\sum fX}{N} = \frac{350}{25} = 14$$

(ii) For Short-cut method:

Here,  $N = 25$ ,  $\sum fd = -25$ ,  $A = 15$ ,  $\bar{X} = ?$

$$\text{Now, } \bar{X} = A + \frac{\sum fd}{N} = 15 + \frac{-25}{25} = 15 - 1 = 14$$

(iii) For Step-deviation method:

Here,  $N = 25$ ,  $h = 5$  (common factor),  $A = 15$ ,  $\sum fd' = -5$ ,  $\bar{X} = ?$

$$\begin{aligned} \text{Now, } \bar{X} &= A + \frac{\sum fd'}{N} \times h \\ &= 15 + \frac{-5}{25} \times 5 = 15 - 1 = 14 \end{aligned}$$

Hence, the average mark is 14.

**Remark:** In the case of discrete or ungrouped frequency distribution, the step-deviation method can be used to calculate the arithmetic mean only when common factor is taken from all the items of the given distribution.

**Continuous Series**

(i) **Direct method:** The arithmetic mean is defined by following formula:

$$\bar{X} = \frac{\sum fX}{N}, \text{ where } X = \text{mid-value of corresponding class}$$

$N$  = total frequency

(ii) **Short-cut method or assumed mean method:** The formula for calculating the arithmetic mean by this method is defined by

$$\bar{X} = A + \frac{\sum fd}{N}, \text{ where, } A = \text{Assumed mean}$$

$$d = X - A$$

$X$  = mid-value of corresponding class

(iii) **Step deviation method:** The formula for calculating the arithmetic mean by this method is defined by

$$\bar{X} = A + \frac{\sum fd'}{N} \times h, \text{ where } A = \text{Assumed mean}$$

$$d' = \frac{X - A}{h}, \quad h = \text{class size or common factor}$$

In the case of unequal class intervals,  $h$  is taken as common factor from the mid-value of the classes. It is not necessary to make unequal class intervals into equal class intervals for calculating arithmetic mean.

**Example 3:** From the following data of the marks obtained by 25 students of a class, calculate the average mark by (i) direct method (ii) short-cut method and (iii) step-deviation method:



Marks obtained	0-20	20-40	40-60	60-80	80-100
No. of students	4	6	8	5	2

**Solution:**

Let  $X$  denotes the mid. value of corresponding classes.

**Calculation of average marks by direct method, short-cut method and step-deviation method**

Marks	$f$	Mid. value ( $X$ )	$fX$	$d = X - 50$	$fd$	$d' = \frac{X-50}{20}$	$fd'$
0-20	4	10	40	-40	-160	-2	-8
20-40	6	30	180	-20	-120	-1	-6
40-60	8	50	400	0	0	0	0
60-80	5	70	350	20	100	1	5
80-100	2	90	180	40	80	2	4
Total	$N = 25$		$\sum fX = 1150$		$\sum fd = -100$		$\sum fd' = -5$

(i) For direct method:

Here,  $N = 25$ ,  $\sum fX = 1150$ ,  $\bar{X} = ?$

$$\text{Now, } \bar{X} = \frac{\sum fX}{N} = \frac{1150}{25} = 46$$

(ii) For short-cut method:

Here,  $N = 25$ ,  $\sum fd = -100$ ,  $A = 50$ ,  $\bar{X} = ?$

$$\text{Now, } \bar{X} = A + \frac{\sum fd}{N} = 50 + \frac{-100}{25} = 50 - 4 = 46$$

(iii) For step-deviation method:

Here,  $A = 50$ ,  $h = 20$  (class size),  $\sum fd' = -5$ ,  $\bar{X} = ?$

$$\text{Now, } \bar{X} = A + \frac{\sum fd'}{N} \times h$$

$$= 50 + \frac{-5}{25} \times 20$$

$$= 50 - 4 = 46$$

Hence, the average mark is 46.

**Note:** If we are dealing with quantitatively measures such as average bonus, average pension, average return, average sales, average revenue, average weight, average income, etc., in that case, mean is more suitable average than others. It is also applied when the given classes have equal size. As arithmetic mean depends on all the observations, it is also called a mathematical average.

(2) **Weighted arithmetic mean:** The formulae discussed above for calculating the simple arithmetic mean are based on the assumption that all the items are of equal importance in the distribution. But, there are cases where the relative importance of the different items in the distribution is not the same. When this is so, we compute weighted arithmetic mean. In such cases, proper weightage is to be given to the various items. For example, commodities like wheat, rice, pulses, housing, fuel and lighting are more important than cigarettes, tea, confectionery, cosmetics, etc. In such cases we use weighted arithmetic mean to calculate the prices of commodities consumed by a certain group of people. The formula for computing weighted arithmetic mean is defined by

$$\begin{aligned} \bar{X}_w &= \frac{W_1X_1 + W_2X_2 + \dots + W_nX_n}{W_1 + W_2 + \dots + W_n} \\ &= \frac{\sum WX}{\sum W}, \text{ where, } \bar{X}_w \text{ denotes the weighted arithmetic mean.} \end{aligned}$$

**Example 4:** A candidate obtained the following percentages in an examination: English 50; Accountancy 80; Statistics 90; Economics 60; Marketing 55. Find the candidate's weighted arithmetic mean, if weights 1,3,3,2 and 1 respectively are allotted to the subjects.

**Solution:**

Let th

Subj
Engl
Acco
Stat
Econ
Mar
Tota

Here,

Now,

Hence

**Example 5**

**Solution:**

In ord  
marks

S

Statist
Account
Economi
Business

For st

For st

Since

### 5.7.1 Math

A few

(i) Th  
i.e.

(ii) Th  
ar



**Solution:**

Let the variable  $X$  represents the percentage of marks in the examination.

**Calculation of weighted A.M.**

Subject	Marks (%) $X$	Weights ( $W$ )	$WX$
English	50	1	50
Accountancy	80	3	240
Statistics	90	3	270
Economics	60	2	120
Marketing	55	1	55
Total		$\Sigma W = 10$	$\Sigma WX = 735$

Here,  $\Sigma W = 10$ ,  $\Sigma WX = 735$ ,  $\bar{X}_w = ?$

$$\text{Now, } \bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{735}{10} = 73.5$$

Hence, weighted arithmetic mean is 73.5

**Example 5:** An examination was held to decide about the award of a scholarship in Kathmandu University. The weights of various subjects were different. The marks obtained by 4 candidates (out of 100 in each subject) are given below:

Subject	Weight	Marks			
		A	B	C	D
Statistics	4	70	80	85	92
Accountancy	3	90	75	75	40
Economics	1	50	60	45	70
Business Law	2	60	45	65	50

If the candidate getting the highest marks is to be awarded the scholarship who should get it?

**Solution:**

In order to decide to whom the scholarship should be given, compare the weighted arithmetic means of the marks obtained by these candidates.

**Calculation of weighted mean**

Subjects	Weight ( $W$ )	A		B		C		D	
		$X$	$WX$	$X$	$WX$	$X$	$WX$	$X$	$WX$
Statistics	4	70	280	80	320	85	340	92	368
Accountancy	3	90	270	75	225	75	225	40	120
Economics	1	50	50	60	60	45	45	70	70
Business Law	2	60	120	45	90	65	130	50	100
	$\Sigma W = 10$		$\Sigma WX = 720$		$\Sigma WX = 695$		$\Sigma WX = 740$		$\Sigma WX = 658$

$$\text{For student A, } \bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{720}{10} = 72 \quad \text{For student B, } \bar{X}_w = \frac{695}{10} = 69.5$$

$$\text{For student C, } \bar{X}_w = \frac{740}{10} = 74 \quad \text{For student D, } \bar{X}_w = \frac{658}{10} = 65.8$$

Since the weighted mean is the highest in case of student C, he should be awarded the scholarship.

**5.7.1 Mathematical Properties of A.M.**

A few important mathematical properties of the arithmetic mean are as follows:

- (i) The algebraic sum of the deviations of the given set of observations from their arithmetic mean is zero. i.e.,

$$(i) \Sigma (X - \bar{X}) = 0 \quad (ii) \Sigma f(X - \bar{X}) = 0$$

- (ii) The sum of the squares of deviations of the given set of observations is minimum when taken from the arithmetic mean.

$$\Sigma (X - \bar{X})^2 < \Sigma (X - A)^2, \text{ where } A \text{ is the assumed mean.}$$

Approved by Curriculum Development Centre (CDC), Nepal



(iii) Since  $\bar{X} = \frac{\sum X}{n}$ , it is clear that any two given values will enable us to compute the third value. Thus,

$$\sum X = n \bar{X} \text{ and } n = \frac{\sum X}{\bar{X}}$$

(iv) *Combined mean of two series.* If  $n_1$  and  $n_2$  be the sizes, and  $\bar{X}_1$ ,  $\bar{X}_2$  be the respective means of two series, then the combined mean  $\bar{X}_{12}$  of two series of size  $(n_1 + n_2)$  is given by

$$\bar{X}_{12} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

**Example 6:** The average wage of the workers of a factory is Rs 150. There are 15 workers working in the factory. Find the total wage bill.

**Solution:**

Here,  $\bar{X} = 150$ ,  $n = 15$ , the total wage bill ( $\sum X$ ) = ?

We know,

$$\bar{X} = \frac{\sum X}{n} \text{ or } \sum X = n \bar{X} = 15 \times 150 = \text{Rs } 2250$$

Hence, the total wage bill is Rs. 2250.

**Example 7:** The average wage of the male workers working in a factory is Rs. 80 and the number of workers is 100. The average wage of the female workers is Rs. 70 and their number is 50. Find the average wage of all the workers taken together.

**Solution:**

Here,  $n_1 = 100$ ,  $\bar{X}_1 = 80$ ,  $n_2 = 50$ ,  $\bar{X}_2 = 70$ ,  $\bar{X}_{12} = ?$

$$\begin{aligned} \therefore \bar{X}_{12} &= \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2} \\ &= \frac{100 \times 80 + 50 \times 70}{100 + 50} = \frac{8000 + 3500}{150} = \frac{11500}{150} = \text{Rs. } 76.67 \end{aligned}$$

Hence, the average wage of all the workers is Rs. 76.67.

**Example 8:** The mean of marks in Statistics of 100 students in a class was 72. The mean of marks of boys was 75, while their number was 70. Find out the mean marks of girls in the class.

**Solution:**

Here,  $n_1 = 70$ ,  $\bar{X}_1 = 75$ ,  $n_1 + n_2 = 100$ ,  $\bar{X}_{12} = 72$ ,  $\bar{X}_2 = ?$

Now,  $n_2 = 100 - 70 = 30$

$$\therefore \bar{X}_{12} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

$$\text{or } 72 = \frac{70 \times 75 + 30 \times \bar{X}_2}{70 + 30}$$

$$\text{or } 72 = \frac{5250 + 30\bar{X}_2}{100}$$

$$\text{or } 5250 + 30\bar{X}_2 = 7200$$

$$\text{or } 30\bar{X}_2 = 7200 - 5250 = 1950$$

$$\text{or } \bar{X}_2 = \frac{1950}{30} = 65$$

Hence, the mean mark of girls in the class is 65.

**Example 9:** The average monthly wage of all workers in a factory is Rs. 444. If the average wages paid to male and female workers are Rs. 480 and Rs. 360 respectively, find the percentage of male and female workers employed by the factory.



**Solution:**

Let  $n_1$  and  $n_2$  denote the number of male and female workers and  $\bar{X}_1$ ,  $\bar{X}_2$  be respectively their average salary in the factory. Then

$$\bar{X}_1 = 480, \bar{X}_2 = 360, \bar{X}_{12} = 444, \% n_1 = ?, \% n_2 = ?$$

$$\text{Let, } n_1 + n_2 = 100\%$$

...(1)

$$\text{We know, } \bar{X}_{12} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

$$\text{or } 444 = \frac{n_1 \times 480 + n_2 \times 360}{100}$$

$$\text{or } 480n_1 + 360n_2 = 44400$$

$$\text{or } 4n_1 + 3n_2 = 370$$

...(2)

Now, multiplying (1) by 3 and then subtracting from (2), we get

$$4n_1 + 3n_2 = 370$$

$$3n_1 + 3n_2 = 300$$

$$\hline$$

$$n_1 = 70$$

Also, from (1),  $n_2 = 30$

Hence, percentage no. of male and female workers employed in the factory are 70 and 30 respectively.

**Example 10:** Calculate the mean wages (Rs.) from the following data

Wages Above	50	100	150	200	250	300	350
No. of workers	100	90	75	50	30	15	5

**Solution:**

As given frequency distribution is more than cumulative frequency distribution, we first need to find the simple frequencies from the more than cumulative frequencies. Again, the class width between two successive terms is 50. So, the last class should be (350 – 400).

#### Calculation of A.M.

Wages (Rs.)	$f$	Mid. value (X)	$d' = \frac{X-225}{50}$	$fd'$
50 – 100	10	75	-3	-30
100 – 150	15	125	-2	-30
150 – 200	25	175	-1	-25
200 – 250	20	225	0	0
250 – 300	15	275	1	15
300 – 350	10	325	2	20
350 – 400	5	375	3	15
Total	$N = 100$			$\sum fd' = -35$

Here,  $N = 100$ ,  $\sum fd' = -35$ ,  $A = 225$ ,  $h = 50$ ,  $\bar{X} = ?$

$$\text{Now, } \bar{X} = A + \frac{\sum fd'}{N} \times h = 225 + \frac{-35}{100} \times 50 = 225 - 17.5 = 207.50$$

Hence, the average wage is Rs. 207.50

**Example 11:** The following are the monthly salaries in Rupees of 30 employees of a firm:

91, 139, 126, 110, 100, 87, 65, 77, 99, 95, 108, 127, 86, 148, 116,

76, 69, 88, 112, 118, 89, 116, 97, 105, 95, 80, 86, 106, 93, 135.

The firm gave bonuses of Rs. 10, 15, 20, 25, 30, 35, 40, 45 and 50 to employees in the respective salary groups: exceeding 60 but not exceeding 70, exceeding 70 but not exceeding 80 and so on up to exceeding 140 but not exceeding 150. Construct a frequency distribution and find out the total bonus paid per employee.



Solution:

Calculation of average bonus per employee

Salary (Rs.)	Tally Marks	Frequency (f)	Bonus (Rs) (X)	$d' = \frac{X-30}{5}$	$fd'$
61 - 70		2	10	-4	-8
71 - 80		3	15	-3	-9
81 - 90		5	20	-2	-10
91 - 100		7	25	-1	-7
101 - 110		4	30	0	0
111 - 120		4	35	1	4
121 - 130		2	40	2	4
131 - 140		2	45	3	6
141 - 150		1	50	4	4
		N = 30			$\sum fd' = -16$

Here,  $A = 30$ ,  $h = 5$ ,  $\sum fd' = -16$ ,  $N = 30$ ,  $\bar{X} = ?$

$$\therefore \bar{X} = A + \frac{\sum fd'}{N} \times h = 30 + \frac{-16}{30} \times 5$$

$$= 30 - 2.67 = 27.33$$

Hence, the average bonus is Rs. 27.33.

**Example 12:** For a certain frequency table which has only been partly reproduced here, the mean was found to be 1.46.

No. of accidents	0	1	2	3	4	5	Total
No. of days	46	?	?	25	10	5	200

Calculate the missing frequencies.

Solution:

Let  $X$  denote the number of accidents and let the missing frequencies corresponding to  $X = 1$  and  $X = 2$  be  $f_1$  and  $f_2$  respectively.

Calculation of A.M.		
No. of accidents (X)	Frequency (f)	fX
0	46	0
1	$f_1$	$f_1$
2	$f_2$	$2f_2$
3	25	75
4	10	40
5	5	25
	$N = 86 + f_1 + f_2$	$\sum fX = 140 + f_1 + 2f_2$

$$\text{Here, } N = 200 = 86 + f_1 + f_2$$

$$\text{or } f_1 + f_2 = 200 - 86$$

$$\text{or } f_1 + f_2 = 114$$

$$\text{Also, } \bar{X} = 1.46, \sum fX = 140 + f_1 + 2f_2$$

$$\therefore \bar{X} = \frac{\sum fX}{N}$$

$$\text{or } 1.46 = \frac{140 + f_1 + 2f_2}{200}$$

$$\text{or } 140 + f_1 + 2f_2 = 1.46 \times 200 = 292$$

$$\text{or } f_1 + 2f_2 = 292 - 140 = 152$$

Subtracting (1) from (2), we get

$$\text{Substituting this in (1), we get}$$

$$f_2 = 152 - 114 = 38$$

$$f_1 = 114 - 38 = 76$$

Hence, the missing frequencies are 38 and 76.

Approved by Curriculum Development Centre (CDC), Nepal

Example 13:

Solution:

Let the v  
wrong fi

Now,  $\bar{X}$

or  $\sum X$

$\therefore$  Cor

$\therefore$  Cor

Example 14:

Solution:

Let  $\bar{X}_1$  den  
mark obtai

Also let  $\bar{X}_2$

Then,

$$\bar{X}_{12} = \frac{\bar{X}_1 + \bar{X}_2}{2}$$

Marks
0 - 20
20 - 40
40 - 60

Here,  $\sum fX =$

$$\therefore \bar{X}_1 = \frac{\sum fX}{N}$$

Again,  $\bar{X}_{12} =$

$$\text{or } 50 = \frac{70}{2}$$

$$\text{or } 2380 + 3$$

$$\text{or } 30\bar{X}_2 = 5$$

$$\text{or } \bar{X}_2 = \frac{262}{30}$$

Hence, the ave

## 5.8. Median

The variate value  
other words, med  
total number of o  
the items lie below



**Example 13:** The mean salary paid to 1,000 employees of an establishment was found to be Rs. 180.40. Later on, after disbursement of salary, it was discovered that the salary of two employees was wrongly entered as Rs. 297 and 165. Their correct salaries were Rs. 197 and 185. Find the correct arithmetic mean.

**Solution:**

Let the variable  $X$  denote the salary (in rupees) of an employee. Then,  $n = 1000$ ,  $\bar{X} = 180.40$ , wrong figures = 297 and 165, correct figures = 197 and 185, correct mean  $\bar{X} = ?$

$$\text{Now, } \bar{X} = \frac{\sum X}{n}$$

$$\text{or } \sum X = n \bar{X} = 1000 \times 180.40 = 180400$$

$$\therefore \text{Corrected } \sum X = 180400 - (297 + 165) + (197 + 185) \\ = 180400 - 462 + 382 = 180320$$

$$\therefore \text{Corrected mean salary } (\bar{X}) = \frac{\text{corrected } \sum X}{n} = \frac{180320}{1000} = \text{Rs. } 180.32$$

**Example 14:** 100 students took a test. The result of those who secured less than 60% mark is given below:

Mark	0 - 20	20 - 40	40 - 60
No. of Students	16	24	30

If the average marks of all students were 50, find out the average marks of those who secured more than 60% marks.

**Solution:**

Let  $\bar{X}_1$  denote the average mark obtained by students who secured less than 60% marks and  $\bar{X}_2$  the average mark obtained by students who secured more than 60% marks.

Also let  $\bar{X}_{12}$  be average marks of all students.

Then,

$$\bar{X}_{12} = 50, N = 100, N_1 = 70, N_2 = 100 - 70 = 30, \bar{X}_2 = ?$$

Calculation of average mark ( $\bar{X}_1$ )

Marks	No. of Students (f)	Mid value (X)	fX
0 - 20	16	10	160
20 - 40	24	30	720
40 - 60	30	50	1500
	$N = 70$		$\sum fX = 2380$

Here,  $\sum fX = 2380$ ,  $N_1 = 70$ ,  $\bar{X}_1 = ?$

$$\therefore \bar{X}_1 = \frac{\sum fX}{N_1} = \frac{2380}{70} = 34$$

$$\text{Again, } \bar{X}_{12} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2}$$

$$\text{or } 50 = \frac{70 \times 34 + 30 \bar{X}_2}{100}$$

$$\text{or } 2380 + 30 \bar{X}_2 = 5000$$

$$\text{or } 30 \bar{X}_2 = 5000 - 2380 = 2620$$

$$\text{or } \bar{X}_2 = \frac{2620}{30} = 87.33$$

Hence, the average mark obtained by students who secured more than 60% marks is 87.33.

## 5.8. Median

The variate value which divides the total number of observations into two equal parts is called the *median*. In other words, median is the *middle most* or most central item in the set of observations. The median divides the total number of observations into two equal parts such that 50% of the items lie above median value and 50% of the items lie below median value. As distinct from the arithmetic mean which is calculated from the value of all



the items of the distribution, the median depends only on the positional value. So, it is also called *positional average*. It is denoted by  $M_d$ .

### Calculation of Median

(A) **Individual series:** The median is calculated by the following formula.

$$M_d = \text{Size of } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item, where } n \text{ is the no. of observations.}$$

Here, following two cases arise.

**Case (i):** If the number of observations is odd, then the middle value gives the median. For example, consider the following 7 observations.

8, 6, 2, 15, 12, 4, 10

Arranging the given set of observations in ascending order, we get  
2, 4, 6, 8, 10, 12, 15

Here,  $n = \text{no. of observations} = 7$

$$\begin{aligned} \text{Now, } M_d &= \text{Size of } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } \left(\frac{7+1}{2}\right)^{\text{th}} \text{ item} = \text{Size of } 4^{\text{th}} \text{ item} = 8 \end{aligned}$$

Hence, the median is 8.

**Case (ii):** If the number of observations is even, then there are two middle values. In such cases, the median is obtained by taking arithmetic mean of the two middle values. Consider the following 6 observations.

3, 5, 2, 12, 10, 18

Arranging the given data in ascending order, we get, 2, 3, 5, 10, 12, 18

Here,  $n = \text{no. of observations} = 6$

$$\begin{aligned} \text{Now, } M_d &= \text{Size of } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } \left(\frac{6+1}{2}\right)^{\text{th}} \text{ item} = \text{Size of } 3.5^{\text{th}} \text{ item} \\ &= \frac{3^{\text{rd}} \text{ item} + 4^{\text{th}} \text{ item}}{2} = \frac{5 + 10}{2} = 7.5 \end{aligned}$$

**Note:** It should be noted that the answer would remain the same whether the data are arranged in ascending or descending order. Students are to show only one arrangement not both.

(B) **Discrete series or ungrouped frequency distribution:** The necessary steps for determining the value of median in a discrete series are as follows:

- Arrange the data in ascending order. When data is arranged, the order of frequencies would also change accordingly.
- Obtain the less than cumulative frequency distribution.
- Use the formula,  $M_d = \text{size of } \left(\frac{N+1}{2}\right)^{\text{th}} \text{ item}$ , where  $N = \sum f = \text{total frequency}$ .
- Now see the c.f. just greater than or equal to  $\left(\frac{N+1}{2}\right)$  at cumulative frequency column, the corresponding value of the variable gives the median.

**Example 15:** From the following data of the weight of 46 persons, determine the median weight.

Weight in lbs.	150	100	140	130	120	160
No. of persons	2	3	16	14	10	1

**Solution:**

Here, we first arrange the data in ascending order and then order of frequencies would also change accordingly.

Weight

Here,  $N =$   
Now,  $\frac{N+1}{2} =$   
The cumulative  
 $\therefore M_d =$   
Hence,  $M_d =$

(C) **Continuous series:**  
median in a  
(i) We find  
(ii) Find  $N$   
(iii) Now see  
class.

where  $l =$   
 $f =$   
 $h =$   
c.f. =  
 $N =$

The above  
order, it shows

**Example 16:** Calculate the median

**Solution:**

Marks obtained
0 - 10
10 - 20
20 - 30
30 - 40
40 - 50

Now,  $\frac{N}{2} =$

Here,  $l = 20$ ,

$\therefore M_d = l +$

Hence, the median



Calculation of Median		
Weights in ascending order	No. of persons (f)	c.f.
100	3	3
120	10	13
130	14	27
140	16	43
150	2	45
160	1	46
	N = 46	

Here,  $N = 46$

$$\text{Now, } \frac{N+1}{2} = \frac{46+1}{2} = \frac{47}{2} = 23.5.$$

The cumulative frequency just greater than 23.5 is 27. So, corresponding median value is 130 lbs.

$$\therefore M_d = 130 \text{ lbs.}$$

Hence, Median weight of persons is 130 lbs.

**(C) Continuous series or grouped frequency distribution:** The necessary steps for determining the value of median in a continuous frequency distribution are as follows:

- We first obtain 'less than' cumulative frequency (c.f.) distribution.
- Find  $N/2$ .
- Now see cumulative frequency just greater than or equal to  $N/2$ . So, the corresponding class is median class. Then value of median is obtained by the following formula:

$$M_d = l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h,$$

where  $l$  = lower limit of the corresponding median class

$f$  = frequency of the corresponding median class.

$h$  = class interval or width of the corresponding median class.

c.f. = cumulative frequency of the class preceding the median class.

$N = \sum f$  = total frequency.

The above formula is applied, when the series is in ascending order. In case, the series is given in descending order, it should be arranged into ascending order before applying the above formula.

**Example 16:** Calculate the median from the following data

Marks obtained	0-10	10-20	20-30	30-40	40-50
No. of students	3	6	10	5	2

**Solution:**

Calculation of median		
Marks obtained	$f$	Cumulative frequency (c.f.)
0-10	3	3
10-20	6	9
20-30	10	19
30-40	5	24
40-50	2	26
	N = 26	

Now,  $\frac{N}{2} = \frac{26}{2} = 13$ . The c.f. just greater than 13 is 19. So, corresponding median class is 20-30.

Here,  $l = 20$ ,  $h = 10$ ,  $f = 10$ , c.f. = 9,  $M_d = ?$

$$\therefore M_d = l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h = 20 + \frac{13 - 9}{10} \times 10 = 20 + 4 = 24$$

Hence, the median mark is 24.



**Note:** For open end classes, incomplete data and unequal classes, median is more suitable measure of average (central tendency) than others. For qualitative characteristics such as intelligence, beauty, honesty, talent, habit, knowledge, defective or non-defective etc., it is more suitable average than others. The distribution of the variable under consideration is continuous with exclusive type classes without any gaps. If the classes are inclusive classes, then it must be converted into exclusive classes (i.e. continuous series) before using the median formula.

**Example 17:** Calculate the suitable measure of average giving reason for your choice.

Income (Rs)	Below 100	100-200	200-500	500-1000	Above 1000
No. of persons	4	6	10	7	3

**Solution:**

Since given classes are open ended, median is more suitable measure of average than others.

Calculation of Median		
Income (Rs)	f	c.f.
Below 100	4	4
100 - 200	6	10
200 - 500	10	20
500 - 1000	7	27
Above 1000	3	30
	N = 30	

Now,  $\frac{N}{2} = \frac{30}{2} = 15$ . The cumulative frequency just greater than 15 is 20.

So, corresponding median class is 200 - 500.

Here,  $l = 200$ ,  $h = 500 - 200 = 300$ ,  $f = 10$ ,  $c.f. = 10$ .

$$\begin{aligned}\therefore M_d &= l + \frac{\frac{N}{2} - c.f.}{f} \times h = 200 + \frac{15 - 10}{10} \times 300 \\ &= 200 + \frac{5}{10} \times 300 = \text{Rs. } 350\end{aligned}$$

Hence, the median income is Rs. 350.

**Example 18** In a class of 15 students, 3 failed in a test. The marks of 12 students who passed were 9, 6, 7, 8, 8, 9, 6, 8, 7, 5, 4 and 7. What was the median of all the 15 students?

**Solution:**

Let  $x_1, x_2$  and  $x_3$  be the marks obtained by failed 3 students respectively such that  $x_1 < x_2 < x_3$ . Now, arranging the given data in ascending order, we get

$x_1, x_2, x_3, 4, 5, 6, 6, 7, 7, 7, 8, 8, 8, 9, 9$

Here,  $n = 15$

$$\begin{aligned}\text{Now, } M_d &= \text{Value of } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item} \\ &= \text{Value of } \left(\frac{15+1}{2}\right)^{\text{th}} \text{ item} \\ &= \text{Value of } 8^{\text{th}} \text{ item} = 7\end{aligned}$$

### 5.8.1 Calculation of Median when Class Intervals are Unequal

When the class intervals are unequal it is not necessary to make class intervals equal for determining median. If, it makes equal, at least three classes should be formed for calculating median.

**Example 19:** Calculate median from the following data:

Marks	0-10	10-30	30-60	60-80	80-90
No. of students	5	15	30	8	2

**Solution:**

Now,  $\frac{N}{2}$

Here,  $l =$

$\therefore M_d =$

**Alternative me**

If we mak

Now,  $\frac{N}{2} =$

Here,  $l = 30$

$\therefore M_d = l$

**Example 20:** Cal

**Solution:**

Now,  $\frac{N}{2} = \frac{50}{2}$

Here,  $l = 40$ ,  $h$

$\therefore M_d = l +$

Hence, the me



**Solution:****Calculation of median when class intervals are unequal**

Marks	f	c.f.
0 - 10	5	5
10 - 30	15	20
30 - 60	30	50
60 - 80	8	58
80 - 90	2	60
	N = 60	

Now,  $\frac{N}{2} = \frac{60}{2} = 30$ . The c.f. just greater than 30 is 50. So, corresponding median class is 30-60.

Here,  $l = 30$ ,  $h = 30$ , c.f. = 20,  $f = 30$

$$\therefore M_d = l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h = 30 + \frac{30 - 20}{30} \times 30 = 30 + 10 = 40$$

**Alternative method**

If we make the class intervals equal, the same answer will be obtained.

**Calculation of median by making class intervals equal**

Marks	f	c.f.
0 - 30	20	20
30 - 60	30	50
60 - 90	10	60
	N = 60	

Now,  $\frac{N}{2} = \frac{60}{2} = 30$ . The c.f. just greater than or equal to 30 is 30. So, corresponding median class is 30 - 40.

Here,  $l = 30$ ,  $h = 10$ , c.f. = 20,  $f = 10$

$$\begin{aligned} \therefore M_d &= l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h \\ &= 30 + \frac{30 - 20}{10} \times 10 \\ &= 30 + 10 = 40. \end{aligned}$$

**Example 20:** Calculate the median wages (Rs.) from the following data

Wages (more than)	0	20	40	60	80	100
No. of workers	50	40	25	15	7	2

**Solution:****Calculation of median**

Wages (Rs.)	f	c.f.
0 - 20	10	10
20 - 40	15	25
40 - 60	10	35
60 - 80	8	43
80 - 100	5	48
100 and above	2	50
	N = 50	

Now,  $\frac{N}{2} = \frac{50}{2} = 25$ . The c.f. just greater than 25 is 35. So, corresponding median class is (40-60).

Here,  $l = 40$ ,  $h = 20$ ,  $f = 10$ , c.f. = 25,  $M_d = ?$

$$\therefore M_d = l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h = 40 + \frac{25 - 25}{10} \times 20 = 40$$

Hence, the median wage is Rs. 40



**Example 21:** Find the median from the following table:

Daily income (Rs.)	10-19	10-29	30-39	40-49	50-59	60-69
No. of persons	4	6	20	10	7	3

**Solution:**

Since, we are given inclusive class intervals; we should convert it to the exclusive class intervals before calculating median.

$$\text{Correction factor (c.f.)} = \frac{20-19}{2} = 0.5$$

**Calculation of median**

Daily income (Rs.)	f	c.f.
9.5 - 19.5	4	4
19.5 - 29.5	6	10
29.5 - 39.5	20	30
39.5 - 49.5	10	40
49.5 - 59.5	7	47
59.5 - 69.5	3	50
	N = 50	

Now,  $\frac{N}{2} = \frac{50}{2} = 25$ . The c.f. just greater than 25 is 30. So, corresponding median class is (29.5 - 39.5).

Here,  $l = 29.5$ ,  $h = 10$ ,  $f = 20$ ,  $c.f. = 10$ ,  $M_d = ?$

$$\begin{aligned} \therefore M_d &= l + \frac{\frac{N}{2} - c.f.}{f} \times h = 29.5 + \frac{25 - 10}{20} \times 10 \\ &= 29.5 + \frac{15}{2} = 29.5 + 7.5 = 37 \end{aligned}$$

**Note:** The distribution of the variable under consideration is continuous with exclusive type classes without any gaps. If the classes are inclusive classes, then it must be converted into exclusive classes (i.e. continuous series) before using the median formula.

**Example 22:** Find the missing frequency from the following distribution of sales of shops, given that the median sale of shops is Rs. 2,400.

Sale in hundred Rs.	0-10	10-20	20-30	30-40	40-50
No. of shops	5	25	?	18	7

**Solution:**

Let the missing frequency be  $f_1$ .

**Calculation of missing frequency**

Sales ('00) Rs.	f	c.f.
0 - 10	5	5
10 - 20	25	30
20 - 30	$f_1$	$30 + f_1$
30 - 40	18	$48 + f_1$
40 - 50	7	$55 + f_1$
	N = 55 + $f_1$	

Since median is Rs. 2400 (Rs. 24 hundred), median class is Rs. 20 - 30.

Here,  $l = 20$ ,  $h = 10$ ,  $f = f_1$ ,  $c.f. = 30$ ,  $M_d = \text{Rs. 24 hundred}$ ,  $N = 55 + f_1$

$$\begin{aligned} \therefore M_d &= l + \frac{\frac{N}{2} - c.f.}{f} \times h \\ \text{or } 24 &= 20 + \frac{\frac{55 + f_1}{2} - 30}{f_1} \times 10 \end{aligned}$$



$$\text{or } 4 = \frac{55 + f_1 - 60}{2f_1} \times 10$$

$$\text{or } 4f_1 = (f_1 - 5)5$$

$$\text{or } 4f_1 = 5f_1 - 25$$

$$\text{or } -f_1 = -25$$

$$\therefore f_1 = 25.$$

Hence, the missing frequency is 25.

**Example 23:** In the frequency distribution of 100 families given below, the number of families corresponding to expenditure groups 20 - 40 and 60 - 80 are missing from the table. However, the median is known to be 50. Find the missing frequencies.

Expenditure	0-20	20-40	40-60	60-80	80-100
No. of families	14	?	27	?	15

**Solution:**

Let  $f_1$  and  $f_2$  be the missing frequencies for the classes 20 - 40 and 60 - 80 respectively.

#### Calculation of missing frequencies

Expenditure (Rs.)	f	c.f.
0 - 20	14	14
20 - 40	$f_1$	$14 + f_1$
40 - 60	27	$41 + f_1$
60 - 80	$f_2$	$41 + f_1 + f_2$
80 - 100	15	$56 + f_1 + f_2$
	$N = 56 + f_1 + f_2 = 100$	

$$\text{Here, } N = 100 = 56 + f_1 + f_2$$

$$\text{or } f_1 + f_2 = 44$$

Since, median is 50; median class is 40 - 60

Here,  $l = 40$ ,  $h = 20$ ,  $f = 27$ , c.f. =  $14 + f_1$ ,  $M_d = 50$

$$\therefore M_d = l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h$$

$$\text{or } 50 = 40 + \frac{\frac{100}{2} - (14 + f_1)}{27} \times 20$$

$$\text{or } 10 = \frac{50 - 14 - f_1}{27} \times 20$$

$$\text{or } 27 = (36 - f_1) \times 2$$

$$\text{or } 27 = 72 - 2f_1$$

$$\text{or } 2f_1 = 72 - 27$$

$$\text{or } 2f_1 = 45$$

$$\text{or } f_1 = \frac{45}{2} = 22.5 \approx 23. \text{ Since frequency can't be fractional value.}$$

Putting the value of  $f_1$  in (1), we get

$$f_2 = 44 - f_1 = 44 - 23 = 21.$$

Hence, the missing frequencies corresponding to classes 20 - 40 and 60 - 80 are 23 and 21 respectively.

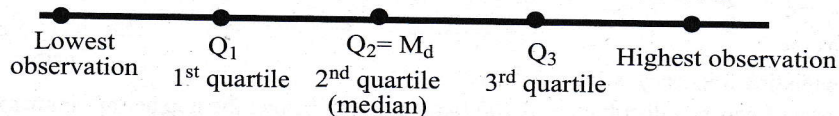
### 5.8.2 Partition Values

The variate values which divide the total number of observations into a number of equal parts are called the **Partition Values**.

**Quartiles:** The variate values which divide the total number of observations into four equal parts are called Quartiles. In such cases there are three points  $Q_1$ ,  $Q_2$  and  $Q_3$  such that  $Q_1 < Q_2 < Q_3$ .  $Q_1$  is called lower or first



quartile which has 25% of the items of the distribution below it and consequently 75% of the items are greater than it.  $Q_2$  is called 2nd quartile which coincides with median and has 50% of the items of the distribution below it and 50% of the items of the distribution above it.  $Q_3$ , is called upper or third quartile which has 75% of the observations below it and 25% of the observations above it.



**Deciles:** The variate values which divide the total number of observations into ten equal parts are called deciles. Then there are 9 points  $D_1, D_2, \dots, D_9$  such that  $D_1 < D_2 < D_3 < \dots < D_9$ .

**Percentiles:** The variate values which divide the total number of observations into 100 equal parts are called percentiles. Then there will be 99 points  $P_1, P_2, \dots, P_{99}$  such that  $P_1 < P_2 < \dots < P_{99}$ .

#### Calculation of Quartiles, Deciles and Percentiles

**Individual Series:** After the given data are arranged in ascending order, the quartiles, deciles and percentiles can be obtained by following formulae:

$$Q_1 = \text{first quartile} = \text{Size of } \left(\frac{n+1}{4}\right)^{\text{th}} \text{ item}$$

$$Q_2 = 2^{\text{nd}} \text{ quartile (or Median)} = \text{Size of } 2\left(\frac{n+1}{4}\right)^{\text{th}} \text{ item}$$

$$Q_3 = 3^{\text{rd}} \text{ quartile} = \text{Size of } 3\left(\frac{n+1}{4}\right)^{\text{th}} \text{ item}$$

$$P_1 = \text{first percentile} = \text{Size of } \left(\frac{n+1}{100}\right)^{\text{th}} \text{ item}$$

$$P_2 = 2^{\text{nd}} \text{ percentile} = \text{Size of } 2\left(\frac{n+1}{100}\right)^{\text{th}} \text{ item}$$

$$P_3 = 3^{\text{rd}} \text{ percentile} = \text{Size of } 3\left(\frac{n+1}{100}\right)^{\text{th}} \text{ item}$$

$$Q_1 = \text{First Quartile or lower quartile}; \quad Q_2 = \text{Second Quartile or median}$$

$$Q_3 = \text{Third quartile or upper quartile}; \quad D_1 = \text{First decile}$$

$$D_3 = \text{Third decile}; \quad D_9 = 9^{\text{th}} \text{ decile}$$

$$P_1 = \text{First percentile}; \quad P_{80} = 80^{\text{th}} \text{ percentile}$$

$n$  = Number of observations

**Discrete series:** The formulae for calculating quartiles, deciles and percentiles are defined as follows:

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{\text{th}} \text{ item, where } N = \sum f = \text{total frequency}$$

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)^{\text{th}} \text{ item}$$

$$D_1 = \text{Size of } \left(\frac{N+1}{10}\right)^{\text{th}} \text{ item}$$

$$D_2 = \text{Size of } 2\left(\frac{N+1}{10}\right)^{\text{th}} \text{ item}$$

$$D_1 = \text{first decile} = \text{Size of } \left(\frac{n+1}{10}\right)^{\text{th}} \text{ item}$$

$$D_2 = 2^{\text{nd}} \text{ decile} = \text{Size of } 2\left(\frac{n+1}{10}\right)^{\text{th}} \text{ item}$$

$$D_3 = 3^{\text{rd}} \text{ decile} = \text{Size of } 3\left(\frac{n+1}{10}\right)^{\text{th}} \text{ item}$$

$$D_9 = 9^{\text{th}} \text{ decile} = \text{Size of } 9\left(\frac{n+1}{10}\right)^{\text{th}} \text{ item}$$

$$P_{99} = 99^{\text{th}} \text{ percentile} = \text{Size of } 99\left(\frac{n+1}{100}\right)^{\text{th}} \text{ item,}$$

where  $n$  = no. of items.



**Continuous series:** The formulae for calculating quartiles, deciles and percentiles are defined as follows:

$$Q_1 = l + \frac{\frac{N}{4} - \text{c.f.}}{f} \times h$$

$$Q_2 = l + \frac{\frac{2N}{4} - \text{c.f.}}{f} \times h$$

$$Q_3 = l + \frac{\frac{3N}{4} - \text{c.f.}}{f} \times h$$

$$D_1 = l + \frac{\frac{N}{10} - \text{c.f.}}{f} \times h$$

$$D_2 = l + \frac{\frac{2N}{10} - \text{c.f.}}{f} \times h$$

$$D_3 = l + \frac{\frac{3N}{10} - \text{c.f.}}{f} \times h$$

$$D_9 = l + \frac{\frac{9N}{10} - \text{c.f.}}{f} \times h$$

$$P_1 = l + \frac{\frac{N}{100} - \text{c.f.}}{f} \times h$$

$$P_2 = l + \frac{\frac{2N}{100} - \text{c.f.}}{f} \times h$$

$$P_{60} = l + \frac{\frac{60N}{100} - \text{c.f.}}{f} \times h$$

$$P_{99} = l + \frac{\frac{99N}{100} - \text{c.f.}}{f} \times h,$$

where,

$l$  = lower limit of corresponding quartile class (or decile class or percentile class)

$f$  = frequency of corresponding class

$h$  = class size of corresponding class

$\text{c.f.}$  = cumulative frequency of preceding quartile class (or decile or percentile class)

**Remarks:** 1. In particular, we shall have

$$Q_1 = P_{25}, Q_3 = P_{75}, M_d = Q_2 = D_5 = P_{50}, D_1 = P_{10}, D_2 = P_{20}, D_3 = P_{30}, D_4 = P_{40} \dots, D_8 = P_{80}, D_9 = P_{90}$$

2. The procedure for calculating quartiles, deciles and percentiles is same as that for median.

**Example 24:** Calculate median, quartiles, 8<sup>th</sup> decile and sixty-first percentile from the following data of the marks obtained by 10 students in an examination.

22, 26, 14, 30, 18, 11, 35, 41, 12, 32

**Solution:** Arranging the marks in ascending order, we get

11, 12, 14, 18, 22, 26, 30, 32, 35, 41

Here,  $n = 10$

$$\begin{aligned} M_d &= \text{Size of } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } \left(\frac{10+1}{2}\right)^{\text{th}} \text{ item} = \text{Size of } 5.5^{\text{th}} \text{ item} \\ &= \frac{5^{\text{th}} \text{ item} + 6^{\text{th}} \text{ item}}{2} = \frac{22 + 26}{2} = \frac{48}{2} = 24 \end{aligned}$$

$$\begin{aligned} Q_1 &= \text{Size of } \left(\frac{n+1}{4}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } \left(\frac{10+1}{4}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } 2.75^{\text{th}} \text{ item.} \\ &= 2^{\text{nd}} \text{ item} + 0.75 (\text{size of } 3^{\text{rd}} \text{ item} - \text{size of } 2^{\text{nd}} \text{ item}) \\ &= 12 + 0.75 (14 - 12) \\ &= 12 + 1.5 = 13.5 \text{ marks.} \end{aligned}$$

$$\begin{aligned} D_8 &= \text{Size of } 8\left(\frac{n+1}{10}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } 8\left(\frac{10+1}{10}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } 8.8^{\text{th}} \text{ item} \\ &= 8^{\text{th}} \text{ item} + 0.8 (9^{\text{th}} \text{ item} - 8^{\text{th}} \text{ item}) \\ &= 32 + .8 (35 - 32) \\ &= 32 + .8 \times 3 = 32 + 2.4 \\ &= 34.4 \text{ marks} \end{aligned}$$

#### Alternative method

$$\begin{aligned} M_d &= \text{size of } 5.5^{\text{th}} \text{ item} \\ &= 5^{\text{th}} \text{ item} + 0.5 (6^{\text{th}} \text{ item} - 5^{\text{th}} \text{ item}) \\ &= 22 + 0.5 (26 - 22) \\ &= 22 + 0.5 \times 4 = 24 \end{aligned}$$

Hence median marks = 24

$$\begin{aligned} Q_3 &= \text{Size of } 3\left(\frac{n+1}{4}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } 3\left(\frac{10+1}{4}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } 8.25^{\text{th}} \text{ item} \\ &= 8^{\text{th}} \text{ item} + 0.25 (9^{\text{th}} \text{ item} - 8^{\text{th}} \text{ item}) \\ &= 32 + 0.25 (35 - 32) = 32 + 0.25 \times 3 = 32.75 \text{ marks.} \end{aligned}$$

$$\begin{aligned} P_{61} &= \text{Size of } 61\left(\frac{n+1}{100}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } 61\left(\frac{10+1}{100}\right)^{\text{th}} \text{ item} \\ &= \text{Size of } 6.71^{\text{th}} \text{ item} \\ &= 6^{\text{th}} \text{ item} + 0.71 (7^{\text{th}} \text{ item} - 6^{\text{th}} \text{ item}) \\ &= 26 + 0.71 (30 - 26) \\ &= 26 + 0.71 \times 4 = 26 + 2.84 \\ &= 28.84 \text{ marks.} \end{aligned}$$



**Example 25:** Calculate lower and upper quartiles,  $D_4$  and  $P_{80}$  from the following data:

Marks	5	10	15	20	25
No. of students	2	8	10	9	6

**Solution:**

**Calculation of  $Q_1$ ,  $Q_3$ ,  $D_4$  and  $P_{80}$**

Marks	$f$	c.f.
5	2	2
10	8	10
15	10	20
20	9	29
25	6	35
	$N = 35$	

Here,  $N = 35$

For  $Q_1$ ,  $\frac{N+1}{4} = \frac{35+1}{4} = 9$ . The c.f. just greater than 9 is 10. So, corresponding quartile value is 10.

$\therefore Q_1 = 10$

For  $Q_3$ ,  $3\left(\frac{N+1}{4}\right) = 3\left(\frac{35+1}{4}\right) = 27$ . The c.f. just greater than 27 is 29. So, corresponding quartile value is 20.

$\therefore Q_3 = 20$

For  $D_4$ ,  $4\left(\frac{N+1}{10}\right) = 4\left(\frac{35+1}{10}\right) = 12.4$ . The c.f. just greater than 12.4 is 20. So, corresponding value is 15.

$\therefore D_4 = 15$

For  $P_{80}$ ,  $80\left(\frac{N+1}{100}\right) = 80\left(\frac{35+1}{100}\right) = 28.80$ . The c.f. just greater than 28.80 is 29. So, corresponding percentile value is 20.

$\therefore P_{80} = 20$

**Example 26:** Find the median, quartiles, 4<sup>th</sup> decile and 20<sup>th</sup> percentile from the following data

Incomes (less than Rs.)	500	400	350	275	200	150	100
No. of persons	100	90	75	50	30	15	5

**Solution:**

Since given classes are in descending order, we first arrange it ascending order. Then, find median and its partition values.

**Calculation of  $M_d$ ,  $Q_1$ ,  $Q_2$ ,  $D_4$  and  $P_{20}$**

Incomes (Rs.)	$f$	c.f.
Below 100	5	5
100 – 150	10	15
150 – 200	15	30
200 – 275	20	50
275 – 350	25	75
350 – 400	15	90
400 – 500	10	100
	$N = 100$	

For  $M_d$ ,  $\frac{N}{2} = \frac{100}{2} = 50$ . The c.f. just greater than or equal to 50 is 50.

So, corresponding median class is 200 – 275.

Her,  $l = 200$ ,  $h = 75$ ,  $f = 20$ , c.f. = 30,  $M_d = ?$

$$\begin{aligned} \therefore M_d &= l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h \\ &= 200 + \frac{50 - 30}{20} \times 75 = 275 \end{aligned}$$

Hence, median income is Rs. 275.

**Alternative method**

For  $M_d$ ,  $\frac{N}{2} = \frac{100}{2} = 50$ . The c.f. just greater than 50 is 75.

So, corresponding median class is 275 – 350.

Her,  $l = 275$ ,  $h = 75$ ,  $f = 25$ , c.f. = 50,  $M_d = ?$

$$\begin{aligned} \therefore M_d &= l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h \\ &= 275 + \frac{50 - 50}{25} \times 75 = 275 \end{aligned}$$

Hence, median income is Rs. 275.



Again, for  $Q_1 = \frac{N}{4} = \frac{100}{4} = 25$ . The c.f. just greater than 25 is 30. So, corresponding class is 150 – 200.

Here,  $l = 150$ ,  $h = 50$ ,  $f = 15$ , c.f. = 15,  $Q_1 = ?$

$$\begin{aligned}\therefore Q_1 &= l + \frac{\frac{N}{4} - \text{c.f.}}{f} \times h \\ &= 150 + \frac{25 - 15}{15} \times 50 \\ &= 150 + \frac{10}{15} \times 50 = 150 + \frac{500}{15} \\ &= 150 + 33.33 = 183.33\end{aligned}$$

For  $Q_3$ ,  $\frac{3N}{4} = \frac{3 \times 100}{4} = 75$ . The c.f. just greater than or equal to 75 is 75. So, corresponding quartile class is 275 – 350.

Hence,  $l = 275$ ,  $h = 75$ ,  $f = 25$ , c.f. = 50,  $Q_3 = ?$

$$\begin{aligned}\therefore Q_3 &= l + \frac{\frac{3N}{4} - \text{c.f.}}{f} \times h \\ &= 275 + \frac{75 - 50}{25} \times 75 = 350\end{aligned}$$

#### Alternative method

For  $Q_3$ ,  $\frac{3N}{4} = \frac{3 \times 100}{4} = 75$ . The c.f. just greater than 75 is 90. So, corresponding quartile class is 350 – 400.

Hence,  $l = 350$ ,  $h = 50$ ,  $f = 15$ , c.f. = 75,  $Q_3 = ?$

$$\begin{aligned}\therefore Q_3 &= l + \frac{\frac{3N}{4} - \text{c.f.}}{f} \times h \\ &= 350 + \frac{75 - 75}{15} \times 50 = 350\end{aligned}$$

For  $D_4$ ,  $\frac{4N}{10} = \frac{4 \times 100}{10} = 40$ . The c.f. just greater than 40 is 50. So, corresponding decile class is 200 – 275.

Here,  $l = 200$ ,  $h = 75$ ,  $f = 20$ , c.f. = 30,  $D_4 = ?$

$$\begin{aligned}\therefore D_4 &= l + \frac{\frac{4N}{10} - \text{c.f.}}{f} \times h = 200 + \frac{40 - 30}{20} \times 75 \\ &= 200 + \frac{75}{2} = 200 + 37.5 = 237.5\end{aligned}$$

For  $P_{20}$ ,  $\frac{20N}{100} = \frac{20 \times 100}{100} = 20$ . The c.f. just greater than 20 is 30. So, corresponding class is 150 – 200.

Here,  $l = 150$ ,  $h = 50$ ,  $f = 15$ , c.f. = 15,  $P_{20} = ?$

$$\begin{aligned}\therefore P_{20} &= l + \frac{\frac{20N}{100} - \text{c.f.}}{f} \times h = 150 + \frac{20 - 15}{15} \times 50 \\ &= 150 + \frac{50}{3} = 150 + 16.67 = 166.67\end{aligned}$$

**Note:** If the quartile, decile or percentile lies in the first class, then the cumulative frequency of the preceding class shall be zero.

**Example 27:** Monthly incomes (in Rs. hundred) of 400 employees in an organization are given below:

Monthly incomes (Below Rs. '00')	80	120	160	200	240	280
No. of employees	10	35	165	310	380	400

Obtain the limits and range of income of central 60% of observed employees.

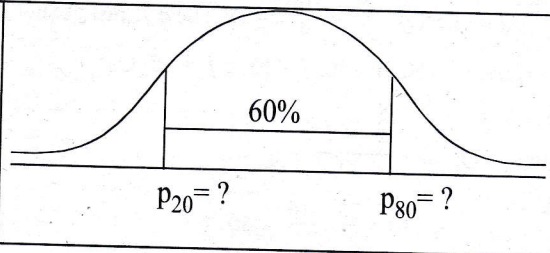
#### Solution:

To find the limits of income of central (i.e. middle) 60%, we have to find  $P_{20}$  and  $P_{80}$ .



Calculation of  $P_{20}$  and  $P_{80}$ 

Monthly Incomes (Rs)	f	c.f.
Below 80	10	10
80 - 120	25	35
120 - 160	130	165
160 - 200	145	310
200 - 240	70	380
240 - 280	20	400
	N = 400	



For  $P_{20}$ ,  $\frac{20N}{100} = \frac{20 \times 400}{100} = 80$ . The c.f. just greater than 80 is 165. So, corresponding class is 120 - 160.

Here,  $l = 120$ ,  $h = 40$ , c.f. = 35,  $f = 130$

$$\begin{aligned} \therefore P_{20} &= l + \frac{\frac{20N}{100} - \text{c.f.}}{f} \times h \\ &= 120 + \frac{80 - 35}{130} \times 40 = 120 + \frac{45}{130} \times 40 \\ &= 120 + 13.85 = \text{Rs. } 133.85 \text{ hundred.} \end{aligned}$$

For  $P_{80}$ ,  $\frac{80N}{100} = \frac{80 \times 400}{100} = 320$ . The c.f. just greater than 320 is 380.

So, corresponding percentile class is 200 - 240.

Here,  $l = 200$ ,  $h = 40$ ,  $f = 70$ , c.f. = 310

$$\begin{aligned} \therefore P_{80} &= l + \frac{\frac{80N}{100} - \text{c.f.}}{f} \times h \\ &= 200 + \frac{320 - 310}{70} \times 40 \\ &= 200 + \frac{10}{7} \times 4 = 200 + \frac{40}{7} = 200 + 5.71 = \text{Rs. } 205.71 \text{ hundred.} \end{aligned}$$

Hence, the limits of income of central 60% of observed employees are Rs. 133.85 and Rs. 205.11.

Range of income of central 60% of observed employees =  $P_{80} - P_{20} = 205.71 - 133.85 = \text{Rs. } 71.86$  hundred.

## 5.9. Mode

The mode or the modal value is that value in a distribution which occurs most frequently. In other words, the *most repeated value* in a distribution is called mode. It is an actual value which has the highest concentration of items in and around it. Actually the word 'mode' has been derived from the French word 'La Mode' which signifies fashion. Thus, mode is the value occurring most frequently in a set of observations and around which other items of the set cluster most densely. The value of the variable at which the curve reaches a maximum is called the mode.

### Calculation of Mode

**Individual and discrete series:** In the case of individual and discrete series, mode is often found by mere inspection.

**Continuous series:** In the case of continuous series, the class corresponding to the maximum frequency is called the modal class and the mode is defined by following formula.

$$M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h, \text{ where}$$

$l$  = lower limit of the modal class

$$\Delta_1 = f_1 - f_0, \Delta_2 = f_1 - f_2$$

Approved by Curriculum Development Centre (CDC), Nepal

$f_1 =$   
 $f_0 =$   
 $f_2 =$   
 $h =$   
 $M_0 =$

### Remark:

- The class in which the frequency is maximum is called the modal class. They are used in the calculation of mode.
- The frequency of the modal class is given in the table.

### Example 28: Find the mode of the following data.

### Solution:

Here, maximum frequency is 100.

Now,  $l = 20$

$$\Delta_1 = f_1 - f_0$$

$$\Delta_2 = f_1 - f_2$$

$$\therefore M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

### Example 29: Find the mode of the following data.

### Solution:

Weight
93
98
103
108
113
118
123
128

Here, maximum frequency is 100.

Now,  $l = 100$

$$\Delta_1 = f_1 - f_0$$

$$\Delta_2 = f_1 - f_2$$

$$\therefore M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$



$f_1$  = maximum frequency

$f_0$  = frequency of the class preceding the modal class

$f_2$  = frequency of the class succeeding the modal class

$h$  = class size of modal class

$M_0$  = mode

**Remark:**

- (i) The class intervals must be uniform throughout i.e., the width of all the class intervals must be the same. If they are unequal, they should first be made equal on the assumption that the frequencies are equally distributed throughout the classes, otherwise the value of mode calculated from formula will give misleading results. It is noted that at least five classes should be formed to make unequal classes into equal classes for calculating mode.
- (ii) The frequency distribution must be continuous with exclusive type classes without any gaps. If the data are given in the form of inclusive classes, it must first be changed into exclusive classes (i.e., continuous classes) before calculating mode.

**Example 28:** Find the mode from the following frequency distribution.

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	3	6	8	5	2

**Solution:**

Here, maximum frequency is 8, so modal class is 20 - 30.

Now,  $l = 20$ ,  $h = 10$ ,  $f_1 = 8$ ,  $f_0 = 6$ ,  $f_2 = 5$ .

$$\Delta_1 = f_1 - f_0 = 8 - 6 = 2$$

$$\Delta_2 = f_1 - f_2 = 8 - 5 = 3$$

$$\therefore M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h = 20 + \frac{2}{2+3} \times 10$$

$$= 20 + \frac{2}{5} \times 10 = 20 + 4 = 24$$

**Example 29:** Find the value of mode from the data given below:

Weight (kg)	No. of students	Weight(kg)	No. of students
93 - 97	2	113 - 117	14
98 - 102	5	118 - 122	6
103 - 107	12	123 - 127	3
108 - 112	17	128 - 131	1

**Solution:**

**Calculation of mode**

Weight (kg)	Exclusive class	f
93 - 97	92.5 - 97.5	2
98 - 102	97.5 - 102.5	5
103 - 107	102.5 - 107.5	12
108 - 112	107.5 - 112.5	17
113 - 117	112.5 - 117.5	14
118 - 122	117.5 - 122.5	6
123 - 127	122.5 - 127.5	3
128 - 131	127.5 - 131.5	1

Here, maximum frequency is 17, so modal class is 107.5 - 112.5

Now,  $l = 107.5$ ,  $h = 5$ ,  $f_1 = 17$ ,  $f_0 = 12$ ,  $f_2 = 14$

$$\Delta_1 = f_1 - f_0 = 17 - 12 = 5$$

$$\Delta_2 = f_1 - f_2 = 17 - 14 = 3$$

$$\therefore M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h = 107.5 + \frac{5}{5+3} \times 5$$

$$= 107.5 + \frac{25}{8} = 107.5 + 3.125 = 110.625$$



**Example 30:** Calculate the mode, median and arithmetic average from the following data.

Class	f	Class	f
0 - 2	8	25 - 30	45
2 - 4	12	30 - 40	60
4 - 10	20	40 - 50	20
10 - 15	10	50 - 60	13
15 - 20	16	60 - 80	15
20 - 25	25	80 - 100	4

**Solution:**

Since the class-intervals are unequal, class intervals should be made equal before computing the mode.

**Calculation of mode, median and mean**

Class	F	c.f.	mid value (X)	$d' = \frac{X-50}{20}$	fd'
0 - 20	66	66	10	- 2	- 132
20 - 40	130	196	30	- 1	- 130
40 - 60	33	229	50	0	0
60 - 80	15	244	70	1	15
80 - 100	4	248	90	2	8
	N = 248				$\sum fd' = -239$

Here, maximum frequency is 130, so modal class is 20 - 40.

Now,  $l = 20$ ,  $h = 20$ ,  $f_1 = 130$ ,  $f_0 = 66$ ,  $f_2 = 33$

$$\Delta_1 = f_1 - f_0 = 130 - 66 = 64$$

$$\Delta_2 = f_1 - f_2 = 130 - 33 = 97$$

$$\begin{aligned} \therefore M_0 &= l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h = 20 + \frac{64}{64 + 97} \times 20 \\ &= 20 + \frac{64 \times 20}{161} = 20 + 7.95 = 27.95 \end{aligned}$$

For  $M_d$ ,  $\frac{N}{2} = \frac{248}{2} = 124$ . The c.f. just greater than 124 is 196. So, corresponding median class is 20 - 40.

$l = 20$ ,  $h = 20$ , c.f. = 66,  $f = 130$

$$\begin{aligned} \therefore M_d &= l + \frac{\frac{N}{2} - \text{c.f.}}{f} \times h = 20 + \frac{124 - 66}{130} \times 20 \\ &= 20 + \frac{58 \times 20}{130} = 20 + 8.92 = 28.92 \end{aligned}$$

$$\begin{aligned} \text{For mean, } \bar{X} &= A + \frac{\sum fd'}{N} \times h = 50 + \frac{-239}{248} \times 20 \\ &= 50 - 19.27 = 30.73. \end{aligned}$$

### 5.9.1 Empirical Relation between Mean, Median and Mode

If mean, median and mode of a distribution are equal, and then frequency distribution is a symmetrical distribution. However, in case of moderately asymmetrical (skewed) distribution, mean, median and mode are not equal. In such cases, mode can be obtained by following empirical relationship, given by *Prof. Karl Pearson*.

$$\text{Mean} - \text{Mode} = 3 (\text{Mean} - \text{Median})$$

$$\text{or } \text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

*Approved by Curriculum Development Centre (CDC), Nepal*

### Calculati

A distribu  
follows:

X	5
f	3

If a distr  
example

X	5
f	3

If the dist  
example

X	5
f	3

In case o  
following

**Example 31:** (a)

**Solution:**

(a) Here, M

$$M_d = 68$$

$$\therefore M$$

$$\text{or } 64$$

$$\text{or } 27$$

$$\text{or } 27$$

$$\text{or } X$$

**Example 32:** H

**Solution:**

Let x be th

Since  $M_0$

Here,  $l = 2$

$$\Delta_1 =$$

$$\Delta_2 =$$

$$\therefore M_0 =$$

$$24 =$$

$$\text{or } 4 =$$

$$\text{or } 1 =$$

$$\text{or } 31 =$$

$$\text{or } x =$$

Hence, m



**Calculation of Mode by using the Empirical Formula**

A distribution having only one maximum frequency is called *unimodal*. e.g., the unimodal distribution is as follows:

X	5	10	15	20	25
f	3	6	8	5	2

If a distribution has two equal maximum frequencies, then distribution is called *bimodal*. The following example is a bimodal distribution.

X	5	10	15	20	25	30
f	3	6	8	8	5	2

If the distribution has three or more equal maximum frequencies, then distribution is called *multimodal*. The example of multimodal distribution is as follows:

X	5	10	15	20	25	30	35
f	3	6	8	8	8	5	2

In case of bimodal or multimodal distributions, mode is ill-defined. In such cases, mode is obtained by the following empirical formula.

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

- Example 31:** (a) The values of mode and median for a moderately skewed distribution are 64.2 and 68.6 respectively. Find the value of the mean.  
 (b) In a moderately asymmetrical distribution the values of mode and mean are 32.1 and 35.4 respectively. Find the median value.  
 (c) If the mean and median of a moderately asymmetrical series are 26.8 and 27.9 respectively, what would be its most probable mode?

**Solution:**

(a) Here,  $M_0 = 64.2$ ,  
 $M_d = 68.6$ , Mean = ?

$$\begin{aligned} \therefore M_0 &= 3M_d - 2\bar{X} \\ \text{or } 64.2 &= 3 \times 68.6 - 2\bar{X} \\ \text{or } 2\bar{X} &= 205.8 - 64.2 \\ \text{or } 2\bar{X} &= 141.6 \\ \text{or } \bar{X} &= 70.8 \end{aligned}$$

(b) Here,  $M_0 = 32.1$ ,  $\bar{X} = 35.4$ ,  $M_d = ?$

$$\begin{aligned} \therefore M_0 &= 3M_d - 2\bar{X} \\ \text{or } 32.1 &= 3M_d - 2 \times 35.4 \\ \text{or } 3M_d &= 32.1 + 70.8 = 102.9 \\ \text{or } M_d &= 34.3 \end{aligned}$$

(c) Here,  $\bar{X} = 26.8$ ,

$$\begin{aligned} M_d &= 27.9, M_0 = ? \\ \therefore M_0 &= 3M_d - 2\bar{X} \\ &= 3 \times 27.9 - 2 \times 26.8 \\ &= 83.7 - 53.6 = 30.1 \end{aligned}$$

**Example 32:** If the mode for the following distribution is Rs. 24, find the missing frequency corresponding to the class 30 - 40.

Expenditure (Rs.)	0-10	10-20	20-30	30-40	40-50
No. of families	14	23	27	?	15

**Solution:**

Let  $x$  be the missing frequency.

Since  $M_0 = 24$ , the modal class is 20 - 30.

Here,  $l = 20$ ,  $h = 10$ ,  $f_1 = 27$ ,  $f_0 = 23$ ,  $f_2 = x$

$$\Delta_1 = f_1 - f_0 = 27 - 23 = 4$$

$$\Delta_2 = f_1 - f_2 = 27 - x$$

$$\therefore M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

$$24 = 20 + \frac{4}{4 + 27 - x} \times 10$$

$$\text{or } 4 = \frac{40}{31 - x}$$

$$\text{or } 1 = \frac{10}{31 - x}$$

$$\text{or } 31 - x = 10$$

$$\text{or } x = 31 - 10 = 21$$

Hence, missing frequency is 21.



### 5.10 Choice of an Average

An average in Statistics is a measure of central tendency; it represents the entire series by a single value which can be substituted for each and every value in the series without causing any change in the total magnitude of the series. As discussed earlier, there are many averages, each of which has its own characteristics and usefulness. It is obvious that no single average is suitable for all circumstances. The choice of a particular average will depend upon:

- the nature and availability of the data,
- the nature of the variable involved,
- the purpose of the enquiry and
- the system of classification adopted.

The following points should be kept in mind while making a choice of an average for use:

**Arithmetic mean:** If all the values in a series are to be given equal importance, arithmetic mean will be a suitable average. When we talk of average cost of production, average price, average exports-imports etc., we always use arithmetic mean. However in the following cases, arithmetic mean should not be used.

- In highly skewed distributions,
- In distributions with open-end classes,
- Where there are very large and very small items,
- When average ratios and rates of change are required. In such cases, the geometric mean is more suitable.

**Median:** Since median is positional average, it is especially used for open-end classes and incomplete data. Also for unequal classes, median is more suitable average than others. For qualitative characteristics such as intelligence, beauty, honesty etc., it is used.

**Mode:** Like median, mode is also used for open-end classes. To determine the most stylish or most frequently occurring item, mode should be computed. For example, when we talk of most common wage, most common income, most common height, most common size of shoe or ready-made garments, we have to use mode and not the arithmetic mean and median.

### 5.11 Limitations of Averages

- An average in itself is simply a single value, so it can yield wrong interpretations too.
- A wrong choice of average may lead to fallacious and funny conclusions.
- Averages in themselves fail to convey the nature of the series, because two quite different series may have the same average.

### 5.12 Data Analysis using Ms-Excel

**Example 1.** Calculate mean, median, first quartile, third quartile, 6<sup>th</sup> decile, 80<sup>th</sup> percentile and mode from the following data of the heights (in inches) of a group of 15 students.

61	62	62	63	61	63	64	64	60	65	63	64	65	66	64
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

**Solution:**

The excel results are as follows:



**Exercise 5.1****A. Practical Problems**

1. The monthly income of ten families (in rupees) in a certain locality are given below:

Family	1	2	3	4	5	6	7	8	9	10
Income	30	70	10	75	500	8	42	250	40	36

Calculate the arithmetic average by (a) Direct method, and (b) short-cut method.

2. Compute mean from the following data by (a) direct method (b) short cut method (c) step-deviation method:

Marks	20	30	40	50	60	70
No. of students	8	12	20	10	6	4

3. From the following data of marks obtained by 100 students of a class, calculate the arithmetic mean of marks by (i) direct method (ii) short-cut method and (iii) step-deviation method.

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	8	12	20	40	12	8

4. (a) Calculate the average daily wages for the workers of two factories

	Factory A	Factory B
No. of wage earners	80	20
Average daily wages	Rs. 1500	Rs. 1260

- (b) From the information given below find:

(i) Which factory pays larger amount as daily wages?

(ii) What is the average daily wage for the workers of the two factories taken together?

	Factory A	Factory B
No. of wage earners	250	200
Average daily wages	Rs. 45.0	Rs. 47.5

5. (a) The mean weight of 100 students in a certain class is 59 kg. The mean weight of boys in the class is 65 kg and that of girls is 50 kg. Find the number of boys and girls in the class.  
 (b) The mean weight of 150 students in a certain class is 60 kg. The mean weight of boys in the class is 70 kg and that of girls is 55 kg. Find the number of boys and the number of girls in the class.

6. (a) In a class of 50 students, 10 have failed and their average of marks is 2.5. The total marks secured by the entire class were 281. Find the average marks of those who have passed.

- (b) The pass result of 40 students who took up a class test is given below:

Marks	4	5	6	7	8	9
No. of students	8	10	9	6	4	3

If the average marks for all the fifty students were 5.16, find out the average marks of the students who failed.

7. (a) From the following data, find the missing frequency when mean is 15.38

Size	10	12	14	16	18	20
Frequency	3	7	?	20	8	5

- (b) From the following frequency distribution of 51 households, calculate missing frequencies if arithmetic mean is 5.

Size	2	3	4	5	6	7
No. of households	2	3	?	21	?	5

8. (a) The average income for a group of 50 persons working in a factory was calculated to be Rs 169. It was later discovered that one figure was misread as 134 instead of the correct value 143. Calculate the correct average income.

- (b) The mean of 200 items was 50. Later on it was discovered that two items were wrongly read as 92 and 8 instead of 192 and 88. Find out the correct mean.



9. (a) Calculate the simple and weighted average from the following data:

Price per ton (Rs)	45	50	40
Tons purchased	130	40	30

- (b) An examination was held to decide the award of a scholarship. The weight given to the various subjects were different. The marks obtained by 3 candidates (out of 100 in each subject) are given below:

Subjects	Weight	Marks		
		A	B	C
Statistics	4	63	60	65
Accountancy	3	65	64	70
Economics	2	58	56	63
Business Law	1	70	80	52

If the candidate getting the highest marks is to be awarded the scholarship, who should get it?

10. (a) Determine the arithmetic mean in the following series:

Marks below	10	20	30	40	50	60	70	80
Frequency	25	40	60	75	95	125	190	240

- (b) Calculate the average weight from the following data:

Weight (in lbs) Above	100	110	120	130	140	150
No. of persons	400	304	170	100	80	50

11. Find the median of the following data:

(i) 15, 8, 12, 7, 18      (ii) 20, 15, 25, 22, 17, 28

12. (a) In a batch of 15 students, 3 students failed. The marks of passed 12 students were

9, 6, 7, 8, 4, 5, 8, 4, 9, 7, 5, 7. What was the median of all 15 students?

- (b) Only 12 students took a test, of which four students have failed with less than 5 marks. What was the median of all the 12 students when the marks of passed students were 5, 7, 8, 6, 9, 10, 8, 7.

13. Find the median from the following data

Marks	10	20	30	40	50	60	70
No. of students	5	10	15	20	12	8	5

14. Find mean and median from the data given below:

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	12	18	27	20	17	6

15. (a) Calculate an appropriate measure of suitable average for the following distribution:

Value	less than 10	10-20	20-30	30-40	40 and above
Frequency	5	10	20	10	5

- (b) You are given below a certain statistical distribution:

Value	less than 100	100-200	200-300	300-400	400 and above
Frequency	40	89	148	64	39

Calculate an appropriate measure of central tendency giving reasons for your choice.

16. (a) In the frequency distribution of 100 families given below, frequency corresponding to two groups are missing from the table. However median is known to be 49.5. Find the missing frequencies.

Expenditure (Rs.)	0-19	20-39	40-59	60-79	80-99
No. of families	14	?	26	?	16

- (b) If the median income of 150 workers of a factory is Rs. 7000, find the missing frequencies and compute mean income of the workers.

Income in '000 Rs.'	0-2	2-4	4-6	6-8	8-10	10-12	12-14
No. of workers	10	?	25	50	?	15	10

17. Find the median,  $Q_1$ ,  $Q_3$ ,  $D_6$  and  $P_{40}$  from following data:

(i) 25, 15, 10, 18, 28, 30, 35      (ii) 55, 47, 42, 59, 65, 75, 85, 90



18. Find the median, upper and lower quartiles from the following frequency distribution.

Expenditure (Rs)	15	30	45	60	75	90	105
No. of persons	3	7	10	5	3	2	5

19. (a) Calculate the median and the quartiles for the following data:

Wages (Rs.) per day	6-8	8-10	10-12	12-14
No. of workers	85	65	59	50

- (b) Calculate the lower and upper quartiles, third decile and 20
- <sup>th</sup>
- percentile from the following data:

Central value	2.5	7.5	12.5	17.5	22.5
Frequency	7	18	25	30	20

20. (a) Calculate the Median and quartiles from the following data:

Marks less than	80	70	60	50	40	30	20	10
No. of students	100	90	80	60	32	20	13	5

- (b) Calculate median, 7
- <sup>th</sup>
- decile and 85
- <sup>th</sup>
- percentile from the following data:

Deposit balance (less than Rs.)	1000	900	800	600	550	500	400	250	100
No. of deposits	500	498	480	475	440	374	300	125	25

21. Compute mean, median and mode from following marks obtained by 9 students:

Marks: 55, 65, 80, 60, 50, 60, 65, 45, 60.

22. The number of telephone calls received in 245 successive one minute intervals at an exchange is shown in the following frequency distribution:

No. of calls	0	1	2	3	4	5	6	7
Frequency	14	21	25	43	51	40	39	12

Calculate the median and mode.

23. (a) Calculate the mode from the following data :

Class intervals	0-20	20-40	40-60	60-80	80-100
Frequency	5	8	10	7	5

- (b) Find the mode of the distribution for the following data:

Value (mid-point)	5	10	15	20	25	30
Frequency	3	7	12	16	13	3

24. A random sample survey was conducted by ABC Shoe Company to determine the size of the shoe that company should produce in order to fit majority of the population. Following information was obtained from the survey.

Size of the foot (in inches)	4-5	5-6	6-7	7-8	8-9	9 and above
No. of persons	10	35	70	35	12	6

What size of the shoe should the company produce to fulfill its objective?

25. (a) Find out the mean, the median and the mode from the following table:

Income (Rs.)	100-200	100-300	100-400	100-500	100-600
No. of persons	15	33	63	83	100

- (b) Calculate the most repeated value after regrouping the frequency distribution given below:

Value	0-10	10-20	20-40	40-50	50-60	60-80	80-100
Frequency	10	15	50	20	10	20	5

26. (a) The following table represents the marks of 100 students.

Marks	0-20	20-40	40-60	60-80	80-100
No. of students	14	?	27	?	15

If the mode value is 48, find the missing frequencies and the mean marks of all 100 students.

- (b) Modal marks for a group of 94 students are 54. 10 students got mark between 0 - 20, 30 students between 40 - 60 and 14 students between 80-100. Find the number of students getting marks between 20 - 40 and 60 - 80, if the maximum mark of the test were 100.

27. (a) In a mode the mean.

- (b) If the mo compute

**B. Data Analysis u**

1. Calculate mea group of stude

59, 61, 61, 62

2. Compute mea

34 62 42

65 80 57

3. From the follo

Size of items

Frequency

4. Calculate mea

Ma

No. of Stude

5. (a) The average workers is F

- (b) The mean v distribution

Weight (kg)

Frequency:

**Answers****A. Practical Pro**

1. Rs. 106.1

5. (a) 60, 40 (b)

7. (a) 12 (b) 9, 11

9. (a) Rs. 45; Rs. 4

10. (a) 49.58 (b)

13. 40

15. (a) 25 (b) Ope

17. (i)
- $M_d = 25$
- ,
- $Q_1 =$

- 18.
- $M_d = 45$
- ,
- $Q_1 =$

19. (a)
- $M_d = 9.37$

20. (a)
- $M_d = 46.43$

- 21.
- $\bar{X} = 60$
- ,
- $M_d = 6$

- 22.
- $M_d = M_0 = 4$

25. (a)
- $\bar{X} = 356$
- ,
- $M_d =$

26. (a) 23, 21, 50

**B. Data Analysis**

- 1.
- $\bar{X} = 62$
- ,
- $M_d =$

- 2.
- $\bar{X} = 55.46$
- ,
- $Q_1 =$

- 3.
- $\bar{X} = 7.25$
- ,
- $M_d =$

- 4.
- $\bar{X} = 27$
- ,
- $M_d =$

5. (a) Rs. 20400



27. (a) In a moderately asymmetrical distribution, the value of median is 42.8 and the value of mode is 40. Find the mean.
- (b) If the mode and mean of a moderately skewed distribution are respectively 16 inches and 20.2 inches, compute the most probable median.

### B. Data Analysis using Ms-Excel

1. Calculate mean, median,  $Q_1$ ,  $Q_3$ ,  $D_6$ ,  $P_{20}$  and mode from the following data of the heights (in inches) of a group of students:

59, 61, 61, 62, 59, 62, 63, 63, 59, 64, 62, 63, 64, 65, 63

2. Compute mean, first quartile, 6<sup>th</sup> decile, 80<sup>th</sup> percentile and mode for the following data:

34	62	42	56	61	42	46	53	67	61	57	45	52
65	80	57	42	61	71	46	49	58	61	51	70	53

3. From the following data, calculate mean, median, lower quartile, upper quartile, 4<sup>th</sup> decile, 90<sup>th</sup> percentile and mode.

Size of items	2	4	6	8	10	12
Frequency	3	5	10	12	6	4

4. Calculate mean, median, lower quartile, upper quartile, 7<sup>th</sup> decile, 10<sup>th</sup> percentile and mode for the following data

Marks	0-10	10-20	20-30	30-40	40-50
No. of Students	5	8	15	16	6

5. (a) The average wage of the male 30 workers in a factory is Rs. 22000 and the average wage of the female 20 workers is Rs. 18000. Find the average wage of all the works taken together.

- (b) The mean weight of 150 students is 56 kg. The weight of 100 students of these is given in the frequency distribution below. Find the arithmetic mean weight of the remaining 50 students.

Weight (kg):	30-40	40-50	50-60	60-70	70-80
Frequency:	10	30	40	10	10

## Answers

### A. Practical Problems

1. Rs. 106.1      2. 41      3. 31      4. (a) Rs. 1500 (b) (i) A (ii) Rs. 46.11
5. (a) 60, 40 (b) 50, 100      6. (a) 6.4 (b) 2.1
7. (a) 12 (b) 9, 11      8. (a) Rs. 169.18 (b) 50.9
9. (a) Rs. 45; Rs. 45.25      (b)  $\bar{X}_w(A) = 63.3$ ,  $\bar{X}_w(B) = 62.4$ ,  $\bar{X}_w(C) = 64.8$ ; C
10. (a) 49.58 (b) 122.6 lbs      11. (i) 12 (ii) 21      12. (a) 6 (b) 6.5
13. 40      14.  $\bar{X} = 28$ ,  $M_d = 27.41$
15. (a) 25 (b) Open end classes,  $M_d = 241.22$       16. (a) 23, 21 (b) 15, 25, Rs. 7000
17. (i)  $M_d = 25$ ,  $Q_1 = 15$ ,  $Q_3 = 30$ ,  $D_6 = 27.4$ ,  $P_{40} = 19.4$       (ii)  $M_d = 62$ ,  $Q_1 = 49$ ,  $Q_3 = 82.5$ ,  $D_6 = 69$ ,  $P_{40} = 57.4$
18.  $M_d = 45$ ,  $Q_1 = 30$ ,  $Q_3 = 75$
19. (a)  $M_d = 9.37$ ,  $Q_1 = 7.52$ ,  $Q_3 = 11.50$  (b)  $Q_1 = 10$ ,  $Q_3 = 19.17$ ,  $D_3 = 11$  &  $P_{20} = 8.61$
20. (a)  $M_d = 46.43$ ,  $Q_1 = 34.17$ ,  $Q_3 = 57.5$       (b)  $M_d = 357.14$ ,  $D_7 = 467.57$ ,  $P_{85} = 538.64$
21.  $\bar{X} = 60$ ,  $M_d = 60$ ,  $M_0 = 60$
22.  $M_d = M_0 = 4$       23. (a) 48      (b) 20.36      24.  $M_0 = 6.5$  inches
25. (a)  $\bar{X} = 356$ ,  $M_d = 356.67$ ,  $M_0 = 354.54$       (b)  $M_0 = 31.11$
26. (a) 23, 21, 50 (b) 16, 24      27. (a) 44.2      (b) 18.8 inches

### B. Data Analysis using Ms-Excel

1.  $\bar{X} = 62$ ,  $M_d = 62$ ,  $Q_1 = 61$ ,  $Q_3 = 63$ ,  $D_6 = 63$ ,  $P_{20} = 60.6$ ,  $M_0 = 63$ .
2.  $\bar{X} = 55.46$ ,  $Q_1 = 46.75$ ,  $D_6 = 58$ ,  $P_{80} = 61.4$ ,  $M_0 = 61$
3.  $\bar{X} = 7.25$ ,  $M_d = 8$ ,  $Q_1 = 6$ ,  $Q_3 = 10$ ,  $D_4 = 6$ ,  $P_{90} = 12$ ,  $M_0 = 8$
4.  $\bar{X} = 27$ ,  $M_d = 28$ ,  $Q_1 = 19.375$ ,  $Q_3 = 35.9375$ ,  $D_7 = 34.375$ ,  $P_{10} = 10$ ,  $M_0 = 30.91$
5. (a) Rs. 20400 (b) 62 kg