3 MEASURES OF LOCATION

3.1 Introduction In the previous chapter, we have seen that it is difficult to learn anything by looking at the data which have not been properly arranged. When the data have been arranged into a frequency distribution, the information contained in the data is easily understood. We have also seen that important features of the data become clear at a glance when the frequency distribution is represented by means of a graph. We can still go further and find a *single value* which will represent all the values of the distribution in some definite way. A value which is used in this way to represent the distribution is called an *average*. Since the averages tend to lie in the centre of a distribution they are called *measures of central tendency*. They are also called *measures of location* because they locate the centre of a distribution.

3.2 Types of Averages The most commonly used averages are (i) the arithmetic mean, (ii) the geometric mean, (iii) the harmonic mean, (iv) the median, and (v) the mode.

3.3 The Arithmetic Mean The arithmetic mean is the most commonly used average. In view of its common use, it is usually referred to as *the average* or simply *the mean.*

The arithmetic mean or simply the mean is defined as the value obtained by dividing the sum of the values by their number. Thus the mean of the values X_1, X_2, ..., X_n denoted by $\bar{X}$ (read as X-bar) is

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum X}{n} \qquad (3.1)$$

Example 3.1(a) Total annual incomes of eight families are Rs.3200, Rs.4000, Rs.3500, Rs.4500, Rs.3800, Rs.4200, Rs.3600 and Rs.53200. Their arithmetic mean is obtained as

$$\bar{X} = \frac{3200 + 4000 + 3500 + 4500 + 3800 + 4200 + 3600 + 53200}{8} = \frac{80000}{8}$$

$= \text{Rs.}10000.$

Example 3.1(b) The mean wage of 5 employees is Rs.1000. If the wages of four employees are Rs.800, Rs.1200, Rs.1300 and Rs.900, find the wage of the fifth employee.

Solution Here $n = 5$ and $\bar{X} = 1000$. We have $\bar{X} = \frac{\sum X}{n}$ or $\sum X = n\,\bar{X}$.

Sum of wages of 5 employees: $\sum X = n\,\bar{X} = 5(1000) = \text{Rs.}5000$

Sum of wages of 4 employees $= 800 + 1200 + 1300 + 900 = \text{Rs.}4200$

Wage of the fifth employee $= 5000 - 4200 = \text{Rs.}800$.

Example 3.1(c) For a set of 15 observations, the mean came out to be 18.2. Later, checking it was discovered that an observation 19.7 was incorrectly recorded where the correct value was 17.9. Calculate the correct mean from this information.

Solution Here $n = 15$ and mean $\bar{X} = 18.2$. We have $\bar{X} = \sum X/n$
or $\sum X = n\bar{X} = 15(18.2) = 273.0$

$$\begin{aligned}\sum X\,(\text{corrected}) &= \sum X + \text{correct value} - \text{incorrect value} = 273.0 + 17.9 - 19.7\\ &= 271.2\end{aligned}$$

$$\bar{X}\,(\text{corrected}) = \frac{\sum X\,(\text{corrected})}{n} = \frac{271.2}{15} = 18.08.$$

Example 3.1(d) Arithmetic mean of 20 values is 25. By adding 3 more values the mean becomes 28. Find the three values if the ratio between these values is 1 : 2 : 3.

Solution Since mean of 20 values is 25, their sum is

$\sum X = n\bar{X} = 20(25) = 500$

On adding 3 more values, $n = 23$ and $\bar{X} = 28$

Sum of the 23 values: $\sum X = 23(28) = 644$

Sum of the newly added values $= 644 - 500 = 144$

Since the values are in the ratio 1 : 2 : 3, and sum of ratios is 6, the values are

$$\frac{144}{6} \times 1 = 24; \quad \frac{144}{6} \times 2 = 48; \text{ and } \frac{144}{6} \times 3 = 72$$

Hence the newly added values are 24, 48 and 72.

3.3.1 Arithmetic Mean from Grouped Data Formula (3.1) is used when the number of values is small. If the number of values is large, they are grouped into a frequency distribution. When the data have been grouped into a frequency distribution, all the values falling in a class are assumed to be equal to the class mark or midpoint of that class. If $X_1, X_2, \ldots, X_k$ are the class marks with $f_1, f_2, \ldots, f_k$ as the corresponding class frequencies, the sum of the values in the first class would be f_1X_1, in the second class f_2X_2 and so on the sum of the values in the kth class would be f_kX_k. Thus the sum of the values in all the k classes would be

$$f_1X_1 + f_2X_2 + \cdots + f_kX_k = \sum_{i=1}^{k} f_iX_i = \sum fX$$

The total number of values is the sum of the class frequencies, namely,

$$f_1 + f_2 + \cdots + f_k = \sum_{i=1}^{k} f_i = \sum f$$

Since the mean is a value obtained by dividing the sum of the values by their number, the mean for grouped data is given by

$$\bar{X} = \frac{\sum_{i=1}^{k} f_iX_i}{\sum_{i=1}^{k} f_i} = \frac{\sum fX}{\sum f} \tag{3.2}$$

Example 3.2(a) The following frequency distribution shows the hourly income of 100 households in a locality.

Income	35–39	40–44	45–49	50–54	55–59	60–64	65–69
f	13	15	28	17	12	10	5

Calculate the arithmetic mean.

Solution

Income	f	X	fX
35 – 39	13	37	481
40 – 44	15	42	630
45 – 49	28	47	1316
50 – 54	17	52	884
55 – 59	12	57	684
60 – 64	10	62	620
65 – 69	5	67	335
	$\Sigma f = 100$		$\Sigma fX = 4950$

$$X = \frac{\Sigma fX}{\Sigma f} = \frac{4950}{100} = 49.50$$

Example 3.2 (b) Find the arithmetic mean for the following distribution showing marks obtained by 50 students in English at a certain examination.

Marks	20–24	25–29	30–34	35–39	40–44	45–49	50–54
Frequency	1	4	8	11	15	9	2

Solution Computation of the arithmetic mean is outlined in Table 3.2.

Table 3.2

Marks	Frequency (f)	Class Mark (X)	fX
20 – 24	1	22	22
25 – 29	4	27	108
30 – 34	8	32	256
35 – 39	11	37	407
40 – 44	15	42	630
45 – 49	9	47	423
50 – 54	2	52	104
	$n = \Sigma f = 50$		$\Sigma fX = 1950$

Here $n = 50$ and $\Sigma fX = 1950$. Using Formula (3.2), we get

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{1950}{50} = 39 \text{ marks.}$$

3.3.2 Short Formulae for Computing Mean The computation of mean using Formula (3.2) is easy provided the values X_i and f_i are not large. If the values X_i and f_i are large, considerable time can be saved by taking deviations from an *assumed* or *guessed mean*. If A is an assumed or guessed mean (which may be any number) and D_i denotes the deviations of X_i from A, i.e. $D_i = X_i - A$, then $X_i = A + D_i$. Substituting this value of X_i, Formulae (3.1) and (3.2) become respectively

$$\bar{X} = A + \frac{\sum_{i=1}^{n} D_i}{n} = A + \frac{\Sigma D}{n} \tag{3.3}$$

$$\bar{X} = A + \frac{\sum_{i=1}^{k} f_i D_i}{\sum_{i=1}^{k} f_i} = A + \frac{\Sigma f D}{\Sigma f} \tag{3.4}$$

Computations using Formulae (3.1) and (3.2), and Formulae (3.3) and (3.4) are sometimes called *long* and *short* methods respectively.

Example 3.3 (a) Find the arithmetic mean from the following values using Formula (3.3).

184 191 172 168 187 189 196 189 193 195

Solution Taking deviations from the assumed mean $A = 180$, $D = X - A = X - 180$. We get $D = 4, 11, -8, -12, 7, 9, 16, 9, 13, 15$ so that $\Sigma D = 64$. Using Formula (3.3), we obtain

$$\bar{X} = A + \frac{\Sigma D}{n} = 180 + \frac{64}{10} = 186.4.$$

Example 3.3 (b) Deviations from 10.5 of ten items are:

$-1.3,\ 2.0,\ 2.9,\ 7.5,\ -4.6,\ -3.4,\ 8.2,\ 9.3,\ -7.4,\ 5.6$.

Calculate the arithmetic mean.

(B.I.S.E., Lahore 1987)

Solution Here $A = 10.5$, $\Sigma D = 18.8$ and $n = 10$. Using Formula (3.3), we have

$$\bar{X} = A + \frac{\Sigma D}{n} = 10.5 + \frac{18.8}{10} = 12.38.$$

Example 3.4(a) Use formula (3.4) to find the mean income of 100 households in a locality.

Solution

$D = X - A,\ X = 47$

Income (Rs.)	Frequency	X	D = X - 47	fD
35 — 39	13	37	-10	-130
40 — 44	15	42	-5	-75
45 — 49	28	A 47	0	0
50 — 54	17	52	5	85
55 — 59	12	57	10	120
60 — 64	10	62	15	150
65 — 69	5	67	20	100
	$\Sigma f = 100$			$\Sigma fD = 250$

$$\bar{X} = A + \frac{\Sigma fD}{\Sigma f} = 47 + \frac{250}{100} = 47 + 2.5 = 49.5$$

Example 3.4 (b) Use Formula (3.4) to find the mean for the frequency distribution of waist measurements of 40 children given in the first two columns of Table 3.4 below.

Solution The computation of the mean is outlined in Table 3.4.

Table 3.4

Waist Measurement (cm)	No. of Children (f)	Class Mark (X)	D = X − 33	fD
20 — 22	1	21	−12	−12
23 — 25	2	24	−9	−18
26 — 28	4	27	−6	−24
29 — 31	7	30	−3	−21
32 — 34	9	A → 33	0	0
35 — 37	10	36	3	30
38 — 40	4	39	6	24
41 — 43	2	42	9	18
44 — 46	1	45	12	12
	$n = \Sigma f = 40$		$\Sigma fD = 28 - 25 = 3$	

Here $A = 17.95$, $n = 40$ and $\Sigma fD = 9$. Using Formula (3.4), we get

$$\bar{X} = A + \frac{\Sigma fD}{n} = 17.95 + \frac{9}{40} = 17.95 + 0.225 = 18.175 \text{ inches.}$$

From the above examples, it is clear that Formula (3.4) is easy to use when size of the class interval is 1.

If all the class intervals are of equal size, say h, the computation of the can be further simplified using a *coding variable* u_i, defined by $u_i = \frac{X_i - A}{h} = \frac{D_i}{h}$.

From this we have $hu_i = X_i - A$ or $X_i = A + hu_i$

Substituting the value of X_i in Formulae (3.1) and (3.2), we get

$$\left.\begin{aligned} \bar{X} &= A + \left(\frac{\sum_{i=1}^{n} u_i}{n}\right) h = A + \left(\frac{\Sigma u}{n}\right) h \\ \text{and} \quad \bar{X} &= A + \left(\frac{\sum_{i=1}^{k} f_i u_i}{\sum_{i=1}^{k} f_i}\right) h = A + \left(\frac{\Sigma fu}{n}\right) h \end{aligned}\right\} \tag{3.5}$$

This is called the *coding method* for computing the mean. It is a very method and should always be used for grouped data where class interval size equal. In the coding method, the values X_i are transformed into the values u_i t deviations of X_i from A and dividing the resulting deviations by the class interva h. The values of u_i will be positive or negative integers or zero, i.e. ..., −3, −2, −1, 2, 3, ... as shown in Tables 3.6 and 3.7.

Example 3.5 The following table shows yields of grains (in tons) for 50 small Use the coding method to find the mean yield per plot.

Yield (ton)	10	14	18	22	26	30	34
No. of Plots	2	4	10	18	9	5	2

Solution The computation of the mean is outlined in Table 3.6. Here $A = 22$, h $n = 50$ and $\Sigma fu = 1$. Using Formula (3.7), we have

$$\bar{X} = A + h\left(\frac{\Sigma fu}{n}\right) = 22 + 4\left(\frac{1}{50}\right) = 22 + 0.08 = 20.08 \text{ tons.}$$

Table 3.6

Yield (ton) X	No. of Plots (f)	$u = \frac{X-A}{h}$	fu
10	2	−3	−6
14	4	−2	−8
18	10	−1	−10
A → 22	18	0	0
26	9	1	9
30	5	2	10
34	2	3	6
	$n = \Sigma f = 50$		$\Sigma fu = 25 - 24 = 1$

Example 3.6 (a) Find the arithmetic mean, given

(i) $D = X - 14$, $\Sigma D = 40$ and $n = 5$

(ii) $u = \dfrac{X-20}{2}$, $\Sigma u = 20$ and $n = 10$

(iii) $X = 50 + 2u$, $\Sigma u = 40$ and $n = 20$

(iv) $D = X - 140$, $\Sigma fD = 400$ and $n = 100$

(v) $u = \dfrac{X-120}{10}$, $\Sigma fu = 200$ and $n = 50$

(vi) $X = 200 + 5u$, $\Sigma fu = 100$ and $n = 100$

Solution (i) Here $A = 14$. Hence $\bar{X} = A + \dfrac{\Sigma D}{n} = 14 + \dfrac{40}{5} = 22$.

(ii) Here $A = 20$ and $h = 2$. Hence $\bar{X} = A + \dfrac{\Sigma u}{n} \times h = 20 + \dfrac{20}{10} \times 2 = 24$.

(iii) Here $A = 50$ and $h = 2$. Hence $\bar{X} = A + \dfrac{\Sigma u}{n} \times h = 50 + \dfrac{40}{20} \times 2 = 54$.

(iv) Here $A = 140$. Hence $\bar{X} = A + \dfrac{\Sigma fD}{n} = 140 + \dfrac{400}{100} = 144$.

(v) Here $A = 120$ and $h = 10$. Hence $\bar{X} = A + \dfrac{\Sigma fu}{\Sigma f} \times h = 120 + \dfrac{200}{50} \times 10 = 160$.

(vi) Here $A = 200$ and $h = 5$. Hence $\bar{X} = A + \dfrac{\Sigma fu}{\Sigma f} \times h = 200 + \dfrac{100}{100} \times 5 = 205$.

Example 3.6. (b) The following data have been obtained from a frequency distribution of a continuous variable X after making the substitution $u = (X - 130)/5$.

u	−3	−2	−1	0	1	2	3	Total
f	5	12	26	32	13	8	4	100

Find the mean using (i) coding method (ii) direct method. (B.I.S.E., Lahore 2014)

Solution (i) Computation of the mean using coding method is outlined in the following table. Last two columns have been added for the computation of mean by direct method for (ii).

u	f	fu	X	fX
−3	5	−15	115	575
−2	12	−24	120	1440
−1	26	−26	125	3250
0	32	0	130	4160
1	13	13	135	1755
2	8	16	140	1120
3	4	12	145	580
	$n = \Sigma f = 100$	$\Sigma fu = 41 - 65 = -24$		$\Sigma fX = 12880$

Here $A = 130$ and $h = 5$. Hence $\bar{X} = A + \left(\dfrac{\Sigma fu}{\Sigma f}\right) h = 130 + \dfrac{(-24)}{100} \times 5 = 128.8$.

(ii) Since $u = \frac{X-130}{5}$, $X = 130 + 5u$. The values of X are obtained putting u = −2, −1, 0, 1, 2, 3. For example, for $u = -3$, $X = 130 + 5(-3) = 115$. The remai values of X are shown in the above table.

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{12880}{100} = 128.8.$$

Example 3.6(c) If μ = X − 112 calculate A.M. for the following data

μ	−2	−1	0	1	2
f	24	30	45	65	72

(B.I.S.E., Gujranwala 20

Solution

μ	f	fu
−2	24	−48
−1	30	−30
0	45	0
1	65	65
2	72	144
	236	131

$$\text{A.M.} = A + \frac{\Sigma fu}{\Sigma f} \times h$$
$$= 112 + \frac{131}{236} \times 5$$
$$= 112 + 2.775$$
$$= 114.78$$

3.3.3 Weighted Arithmetic Mean Sometimes we want to find the average certain values which are not of equal importance. When the values are not of equ importance, we assign them certain numerical values to express their relati importance. These numerical values are called *weights*. If $X_1, X_2, \ldots, X_k$ ha weights $W_1, W_2, \ldots, W_k$, then the *weighted arithmetic mean* or the *weighted me* (denoted by $\bar{X}_w$) is defined as

$$\bar{X}_w = \frac{W_1X_1 + W_2X_2 + \cdots + W_kX_k}{W_1 + W_2 + \cdots + W_k} = \frac{\sum_{i=1}^{k} W_iX_i}{\sum_{i=1}^{k} W_i} = \frac{\Sigma WX}{\Sigma W} \qquad (3.6)$$

Formula (3.6) becomes Formula (3.2) if we replace the weights W_i by cla frequencies f_i. Thus the mean of grouped data may be regarded as the weighted me of the values $X_1, X_2, \ldots, X_k$ whose weights are the respective class frequencies f_1, …, f_k.

Example 3.7(a) The marks obtained by a student in English, Urdu and Statisti were 70, 76 and 82 respectively. Find the appropriate average if weights of 5, 4 and are assigned to these subjects.

Solution We use the weighted mean, the weights attached to the marks being 5, and 3. Thus

$$\bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{5(70) + 4(76) + 3(82)}{5 + 4 + 3} = \frac{350 + 304 + 246}{12} = \frac{900}{12} = 75 \text{ marks.}$$

Example 3.7(b) Three teachers of statistics reported mean examination grades of 79, 74 and 82 in their classes, which consisted of 32, 25 and 17 students respectively. Determine the mean grade for all the classes.

Solution We use a weighted mean. The weights assigned to each grade are taken as the number of students in each class. Thus

$$\bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{(32)\,(79) + (25)\,(74) + (17)\,(82)}{32 + 25 + 17} = \frac{5772}{74} = 78.$$

3.3.4 Properties of the Arithmetic Mean Some important properties of the arithmetic mean are given below:

(i) The sum of deviations of values from their mean is zero. Symbolically $\Sigma(X_i - \bar{X}) = 0$ or $\Sigma f_i\,(X_i - \bar{X}) = 0$.

Example 3.8 The mean of the values 3, 10, 8, 5 and 4 is 6. The deviations of these values from the mean 6 are $-3, 4, 2, -1$ and -2 with sum equal to zero.

(ii) If n_1 values have mean $\bar{X}_1$, n_2 values have mean $\bar{X}_2, \ldots, n_k$ values have mean $\bar{X}_k$, the mean of all the values is

$$\bar{\bar{X}} = \frac{n_1\bar{X}_1 + n_2\,\bar{X}_2 + \cdots + n_k\,\bar{X}_k}{n_1 + n_2 + \cdots + n_k} = \frac{\sum_{i=1}^{k} n_i\bar{X}_i}{\sum_{i=1}^{k} n_i} = \frac{\Sigma n\bar{X}}{\Sigma n} \qquad (3.7)$$

Example 3.9(a) The mean weights of four groups of students consisting of 15, 20, 10 and 18 students were 162, 148, 153 and 140 pounds respectively. Find the mean weight of all the students.

Solution $\bar{\bar{X}} = \dfrac{\Sigma n\bar{X}}{\Sigma n} = \dfrac{15(162) + 20(148) + 10(153) + 18(140)}{15 + 20 + 10 + 18} = \dfrac{9440}{63}$

$= 149.84$ pounds.

(iii) The sum of squares of the deviations of the values X_i from any value a is minimum if and only if $a = \bar{X}$. Symbolically

$$\Sigma(X_i - a)^2 \text{ is a minimum if and only if } a = \bar{X}. \qquad (3.8)$$

In Example 3.8 above, the sum of squares of deviations from mean is $(-3)^2 + (4)^2 + (2)^2 + (-1)^2 + (-2)^2 = 34$. Now let us take deviations from 5 (a value other than mean). The deviations from 5 are $-2, 5, 3, 0$ and -1. The sum of square of these deviations is $(-2)^2 + (5)^2 + (3)^2 + (0)^2 + (-1)^2 = 39$, which is more than 34.

(iv) The arithmetic mean is affected by change of origin and scale. By this we mean that if we add or subtract a constant from all the values or multiply or divide all the values by a constant, the mean is affected by these changes. Symbolically

(a) If $X = a$ (a constant), then $\bar{X} = a$

(b) If $Y = X \pm a$, then $\bar{Y} = \bar{X} \pm a$

(c) If $Y = bX$, then $\bar{Y} = b\,\bar{X}$.

(d) If $Y = \dfrac{X}{a}$, then $\bar{Y} = \dfrac{\bar{X}}{a}$ or $\bar{Y} = \dfrac{1}{a}(\bar{X})$

Combining the above, if $Y = bX \pm a$, then $\bar{Y} = b\,\bar{X} \pm a$

and if $Y = \dfrac{X}{b} \pm a$, then $\bar{Y} = \dfrac{1}{b}(\bar{X}) \pm a$. (3.9)

Example 3.9(b) The mean weight of 150 students in a class is 60 kg. The mean weight of boys in the class is 70 kg and that of the girls is 55 kg. Find the number of boys and girls in the class.

Solution Let the number of boys $= n_1$ and number of girls $= n_2$

We are given $n_1 + n_2 = 150$, $\bar{x}_1 = 70$ $\bar{x}_2 = 55$, $\bar{x}_c = 60$

$$n_1 = 150 - n_2 \qquad (1)$$

Now $\bar{x}_c = \dfrac{n_1\bar{x}_1 + n_2\bar{x}}{n_1 + n_2}$

$$\Rightarrow \quad 60 = \frac{n_1 \times 70 + n_2 \times 55}{150}$$

$$60 \times 150 = 70n_1 + 55n_2$$

$$9000 = 70n_1 + 55n_2 \qquad (2)$$

Putting (1) in (2) we get

$$9000 = 70(150 - n_2) + 55n_2$$

$$9000 = 10500 - 70n_2 + 55n_2$$

$$15n_2 = 10500 - 9000$$

$$n_2 = \frac{1500}{15} = 100$$

The number of girls in the class is 100, putting $n_2 = 100$ in (1) $n_1 = 150 - 100 = 50$ and the number of boys in the class is 50.

Example 3.10 (a) Compute the mean of the observations X_i : 4, 8, 16, 20 and 12. Using this mean, compute the mean in the following cases after, (i) adding 5 to each observation, (ii) subtracting 3 from each observation, (iii) multiplying each observation by 2, (iv) dividing each value by 4, (v) multiplying each value by 2 and then adding 5, (vi) dividing each value by 4 and subtracting 3. Verify the results in (i) – (vi) by using the properties of the mean.

Solution $\bar{X} = \dfrac{4 + 8 + 16 + 20 + 12}{5} = \dfrac{60}{5} = 12.$

(i) Adding 5 the observations become Y_i : 9, 13, 21, 25 and 17.

$$\bar{Y} = \frac{9 + 13 + 21 + 25 + 17}{5} = \frac{85}{5} = 17.$$

(ii) Subtracting 3 the observations become Y_i: 1, 5, 13, 17 and 9.

$$\bar{Y} = \frac{1 + 5 + 13 + 17 + 9}{5} = \frac{45}{5} = 9.$$

(iii) Multiplying by 2, the observations become Y_i : 8, 16, 32, 40 and 24.

$$\bar{Y} = \frac{8 + 16 + 32 + 40 + 24}{5} = \frac{120}{5} = 24.$$

(iv) Dividing by 4, the observations become Y_i: 1, 2, 4, 5 and 3.

$$\bar{Y} = \frac{1+2+4+5+3}{5} = \frac{15}{5} = 3.$$

(v) Multiplying by 2 and adding 5, the observations become Y_i: 13, 21, 37, 45, 29.

$$\bar{Y} = \frac{13+21+37+45+29}{5} = \frac{145}{5} = 29.$$

(vi) Dividing by 4 and subtracting 3, the observations become
Y_i: −2, −1, 1, 2 and 0.

$$\bar{Y} = \frac{-2-1+1+2+0}{5} = 0.$$

$\bar{X} = 12$. The result (i) – (vi) are verified below:

(i) $Y = X + 5$, $\bar{Y} = \bar{X} + 5 = 12 + 5 = 17$.

(ii) $Y = X - 3$, $\bar{Y} = \bar{X} - 3 = 12 - 3 = 9$.

(iii) $Y = 2X$, $\bar{Y} = 2\bar{X} = 2(12) = 24$.

(iv) $Y = \frac{X}{4}$, $\bar{Y} = \frac{\bar{X}}{4} = \frac{12}{4} = 3$.

(v) $Y = 2X + 5$, $\bar{Y} = 2\bar{X} + 5 = 2(12) + 5 = 29$.

(vi) $Y = \frac{X}{4} - 3$, $\bar{Y} = \frac{\bar{X}}{4} - 3 = \frac{12}{4} - 3 = 3 - 3 = 0$.

Example 3.10 (b) By multiplying each of the values 2, 4, 6, 3 and 5 by 3 and then adding 5, we obtain 11, 17, 23, 14 and 20. What is the relation between the means of the two sets.

Solution. Let the first set be denoted by X and the second set by Y so that $Y = 5 + 3X$.

Mean of first set, $\bar{X} = \frac{\Sigma X}{n} = \frac{2+4+6+3+5}{5} = \frac{20}{5} = 4.$

Mean of second set, $\bar{Y} = \frac{\Sigma Y}{n} = \frac{11+17+23+14+20}{5} = \frac{85}{5} = 17.$

The relation is $\bar{Y} = 5 + 3\bar{X} = 5 + 3(4) = 17$.

3.4 The Geometric Mean The geometric mean, G, of a set of n positive values $X_1, X_2, \ldots, X_n$ is the nth root of the product of the values. Thus

$$G = \sqrt[n]{X_1 \cdot X_2 \cdot \ldots \cdot X_n} = (X_1 \cdot X_2 \cdot \ldots \cdot X_n)^{1/n} \qquad (3.10)$$

Example 3.11 The geometric mean of the values 2, 4 and 8 is

(B.I.S.E., Gujranwala 2010)

$$G = \sqrt[3]{2 \times 4 \times 8} = \sqrt[3]{64} = 4.$$

In practice, it is difficult to extract higher roots. The geometric mean is, therefore, computed using logarithms. It is given by

$$\text{Log } G = \frac{[\log X_1 + \log X_2 + \cdots + \log X_n]}{n} = \frac{1}{n}\left(\sum_{i=1}^{n} \log X_i\right) = \frac{\Sigma \log X}{n} \qquad (3.11)$$

Here we assume that all the values are positive, otherwise the logarithms are not defined.

Example 3.12 Find the geometric mean of the values

(i) 3, 5, 6, 6, 7, 10, 12.

(ii) 7.96, 13.82, 22.95, 35.34.

Solution

$$\text{(i)}\ \log G = \frac{(\log 3 + \log 5 + \log 6 + \log 6 + \log 7 + \log 10 + \log 12)}{7}$$

$$= \frac{0.47712 + 0.69897 + 0.77815 + 0.77815 + 0.84510 + 1.00000 + 1.07918}{7}$$

$$= 5.65667/7 = 0.8081$$

$$G = \text{antilog}\ (0.8081) = 6.42836.$$

The arithmetic mean of these values is

$$\bar{X} = \frac{3 + 5 + 6 + 6 + 7 + 10 + 12}{7} = \frac{49}{7} = 7.$$

This illustrates that the geometric mean of a set of values (not all equal) is less than their arithmetic mean. Moreover, if any one of the original values is zero, their geometric mean is zero.

$$\text{(ii)}\ \log G = \frac{\log 7.96 + \log 13.82 + \log 22.95 + \log 35.34}{4}$$

$$= \frac{0.90091 + 1.14051 + 1.36078 + 1.54827}{4} = 4.95047/4 = 1.23762$$

$$G = \text{antilog}\ (1.23762) = 17.28.$$

3.4.1 The Geometric Mean for Grouped Data When the data have been arranged into a frequency distribution, each of the original observations in a class is assumed to have a value equal to its class mark. Suppose $X_1, X_2, \ldots, X_k$ represent the class marks in a frequency distribution with $f_1, f_2, \ldots, f_k$ as the corresponding class frequencies (where $f_1 + f_2 + \cdots + f_k = \Sigma f = n$). Since X_1 occurs f_1 times, X_2 occurs f_2 times, ..., X_k occurs f_k times, then the product of original values may be written as

$$\underbrace{X_1 \cdot X_1 \cdot \ldots \cdot X_1}_{f_1\ \text{times}} \quad \underbrace{X_2 \cdot X_2 \cdot \ldots \cdot X_2}_{f_2\ \text{times}} \ldots \underbrace{X_k \cdot X_k \cdot \ldots \cdot X_k}_{f_k\ \text{times}}$$

or $X_1^{f_1}\ X_2^{f_2} \ldots X_k^{f_k}$

and the geometric mean is

$$G = \sqrt[n]{X_1^{f_1}\ X_2^{f_2} \ldots X_k^{f_k}} = \left(X_1^{f_1}\ X_2^{f_2} \ldots X_k^{f_k}\right)^{1/n} \tag{3.12}$$

where $n = \Sigma f$. This is sometimes called the *weighted geometric mean* with weights $f_1, f_2, \ldots, f_k$.

Taking logarithm of both sides of (3.12), we have

$$\log G = \frac{[f_1 \log X_1 + f_2 \log X_2 + \cdots + f_k \log X_k]}{n} = \frac{1}{n}\sum_{i=1}^{k} f_i \log X_i = \frac{\Sigma f \log X}{n} \quad (3.13)$$

Note that the logarithm of the geometric mean of a set of n positive values is the arithmetic mean of the logarithms of the values.

Example 3.13 Find geometric mean for the following frequency distribution using (i) the basic definition (ii) logarithms.

X	1	2	3	4	Total
f	2	3	4	1	10

Solution (i) Using the basic definition:

$$G = \left[X_1^{f_1} \cdot X_2^{f_2} \cdot \ldots \cdot X_k^{f_k}\right]^{1/\Sigma f} = [(1)^2 \cdot (2)^3 \cdot (3)^4 \cdot (4)^1]^{1/10}$$

$$= [(1) \cdot (8) \cdot (81) \cdot (4)]^{1/10} = (2592)^{1/10} = 2.1946.$$

(ii)

X	f	$\log X$	$f \log X$
1	2	0.0000	0.0000
2	3	0.3010	0.9030
3	4	0.4771	1.9084
4	1	0.6021	0.6021
	$\Sigma f = 10$	$\Sigma f \log X = 3.4135$	

$$\log G = \frac{\Sigma f \log X}{\Sigma f} = \frac{3.4135}{10} = 0.34135$$

$G = \text{antilog}\ (0.34135) = 2.1946.$

Example 3.14 The logarithm of 10 values of given below. 1.5315, 1.7559, 1.0414, 1.6812, 1.7782, 1.4314, 1.5798, 1.6532, 1.2304, 1.8062. Find $\bar{X}$.

Solution We are given the column logX. To make the column X, we have to take the antilog of the column logX.

log X	X-Antilog (logX)
1.5315	34
1.7559	57
1.0414	11
1.6872	48
1.7782	60
1.4314	27
1.5798	38
1.6532	45
1.2304	17
1.8062	64
Total	401

$$\bar{X} = \frac{\Sigma X}{n} = \frac{401}{10} = 40.1$$

Example 3.15 (a) Find the average rate of increase of income if the income of person increased by 25% during the first year and 40% during the second year.

Solution During the first year, income increased by 25%. Thus the income at the end of first year was 125% of the income at the beginning of the first year and the income at the end of second year was 140% of the income at the end of first year.

$$G = \sqrt{125 \times 140} = 132.3\% \text{ of the income in the first year.}$$

Thus average rate of increase = 132.3 – 100 = 32.3%.

We can also find the average rate of increase by *compound interest formula*. Suppose the capital P_0 is invested at the rate of interest (r). After n years the capital will be P_n which may be determined by the formula:

$$P_n = P_0\,(1 + r)^n \tag{3.14}$$

In the above example,

P_0 = Income in the first year = 100

P_1 = Income in the second year $= 100 \times \frac{125}{100} = 125$

P_2 = Income in the third year $= 125 \times \frac{140}{100} = 175$

Here $n = 2$. Using Formula (3.14), we get

$$P_2 = P_0(1 + r)^2 \text{ or } 175 = 100(1 + r)^2$$

$$\text{or } (1 + r)^2 = \frac{175}{100} \text{ or } (1 + r) = \sqrt{\frac{175}{100}} \text{ or}$$

$$r = \sqrt{\frac{175}{100}} - 1 = \sqrt{1.75} - 1 = 1.323 - 1 = 0.323 \text{ or } 32.3\%.$$

It is important to note that if we find the arithmetic mean of the incomes 125 and 175, it is 150 which means an increase of 150 – 100 = 50%. If income increases by an average rate of 50%, then income in the second year would be 150 and in the third year 225. Thus the arithmetic mean is not the appropriate average here.

Example 3.15 (b) A man gets a rise of 10% in salary at the end of his first year of service and further rises of 20% and 25% at the end of second and third years respectively; the rise in each year being calculated on his salary at the beginning of the year. To what average annual percentage increase is this equivalent?

Solution. Salary at the end of first year is 110% of the salary at the beginning of the first year of service; at the end of the second year it is 120% of the salary at the end of first year and at the end of the third year it is 125% of the salary at the end of the second year.

$$G = (110 \times 120 \times 125)^{1/3}$$

$$\log G = \frac{1}{3}[\log 110 + \log 120 + \log 125] = \frac{1}{3}(2.04139 + 2.07918 + 2.09691)$$

$$= \frac{6.21748}{3} = 2.07249$$

$G = \text{antilog}\ (2.07249) = 118.165\%$

Thus average increase per year = 118.17 − 100 = 18.17%.

3.5 The Harmonic Mean The harmonic mean, H, of a set of n values $X_1, X_2, \ldots, X_n$ is the reciprocal of the arithmetic mean of the reciprocals of the values. The mean of the reciprocals

$$\frac{1}{X_1}, \frac{1}{X_2}, \ldots, \frac{1}{X_n} \text{ is } \left(\frac{\frac{1}{X_1} + \frac{1}{X_2} + \ldots + \frac{1}{X_n}}{n}\right).$$

Hence the harmonic mean, H, is given by

$$H = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \ldots + \frac{1}{X_n}} = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{X_i}\right)} = \frac{n}{\Sigma\left(\frac{1}{X}\right)} \qquad (3.14)$$

In practice it is easier to remember

$$\frac{1}{H} = \frac{\sum_{i=1}^{n}\left(\frac{1}{X_i}\right)}{n} = \frac{1}{n}\Sigma\left(\frac{1}{X}\right) \qquad (3.15)$$

Example 3.16. (a) The harmonic mean of the values 2, 4 and 8 is

$$H = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}} = \frac{3}{7/8} = 3 \times \frac{8}{7} = \frac{24}{7} = 3.43.$$

The geometric mean and the arithmetic mean of these values are 4 and 4.67 respectively. This shows that the harmonic mean of values (not all equal) is less than their geometric mean which in turn is less than their arithmetic mean.

Example 3.16(b) The harmonic mean of the values 3, 5, 6, 6, 7, 10 and 12 is

$$H = \frac{7}{\frac{1}{3} + \frac{1}{5} + \frac{1}{6} + \frac{1}{6} + \frac{1}{7} + \frac{1}{10} + \frac{1}{12}}$$

$$= \frac{7}{0.3333 + 0.2000 + 0.1667 + 0.1667 + 0.1429 + 0.1000 + 0.0833} = \frac{7}{1.1929}$$

$= 5.87.$

Example 3.15(c) The daily wages for a group of 200 persons have been obtained from a frequency distribution of a continuous variable x after making the substitution $u = \frac{X - 130}{20}$

u	−2	−1	0	1	2
No. of Persons	7	50	50	40	3

(B.I.S.E., Gujranwala 2018)

Sol.

u	f	X = 20u + 30	log X	f log X
–2	7	90	1.9542	13.6794
–1	50	110	2.0414	102.07
0	50	130	2.1139	105.695
1	40	150	2.1761	87.044
2	3	170	2.2304	6.6912
	150			318.1796

$$\log GM = \frac{\Sigma f \log X}{\Sigma f} = \frac{318.1796}{150} = 2.12119$$

$$G.M. = \text{Antilog}\ (2.12119)$$

$$= 132.1896$$

Example 3.16(d) Reciprocal of 'X' are given below.
0.0267, 0.0235, 0.0211, 0.0191, 0.0174, 0.0160, 0.0148.

(B.I.S.E., Gujranwala 2015; Lahore 2018)

Solution

$$H.M. = \frac{n}{\Sigma(1/X)} = \frac{7}{0.0267 + 0.0235 + 0.0211 + 0.0191 + 0.0174 + 0.0148 + 0.0160}$$

$$= \frac{7}{0.1386} = 50.51$$

3.5.1 The Harmonic Mean for Grouped Data Suppose $X_1, X_2, \ldots, X_k$ represent the class marks and $f_1, f_2, \ldots, f_k$ as the corresponding class frequencies (where $f_1 + f_2 + \cdots + f_k = \Sigma f_i = n$). Then the reciprocals of the class marks will be $\frac{1}{X_1}, \frac{1}{X_2}, \ldots, \frac{1}{X_k}$. Since the reciprocals occur with frequencies $f_1, f_2, \ldots, f_k$, the total value of the reciprocals in the first class is $\frac{f_1}{X_1}$, in the second class is $\frac{f_2}{X_2}, \ldots$, in the kth class is $\frac{f_k}{X_k}$. The sum of the reciprocals in all the k classes would be

$$\frac{f_1}{X_1} + \frac{f_2}{X_2} + \cdots + \frac{f_k}{X_k} = \sum_{i=1}^{k} \frac{f_i}{X_i} = \Sigma f\left(\frac{1}{X}\right).$$

Since the harmonic mean is defined as the reciprocal of the arithmetic mean of reciprocals of the values, it is given by

$$H = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} f_i\left(\frac{1}{X_i}\right)} = \frac{\Sigma f}{\Sigma f\left(\frac{1}{X}\right)} = \frac{n}{\Sigma f\left(\frac{1}{X}\right)} \tag{3.16}$$

where $n = \Sigma f$. This is sometimes called the *weighted harmonic mean* with weights $f_1, f_2, \ldots, f_k$.

Example 3.17(a) Calculate H.M. of the data given below. (B.I.S.E. Lahore, 2012)

Weight	40 — 44	45 — 49	50 — 54	55 — 59
f	20	30	40	10

Solution

Weight	f	X	$1/X$	f/X	C.B.	C.F.
40 – 44	20	42	0.0238	0.476	39.5 – 44.5	20
45 – 49	30 f_1	47	0.0213	0.639	44.5 – 49.5	50
50 – 54	40 f_m	52	0.0192	0.768	49.5 – 54.5	90
55 – 59	10 f_2	57	0.0175	0.175	54.5 – 59.5	100
	100			2.058		

$$H.M. = \frac{\Sigma f}{\Sigma(f/X)} = \frac{100}{2.058} = 48.59$$

Example 3.17(b) Calculate mode and P_{40} of the data given above.

Solution

$$\text{Mode} = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h = 49.5 + \frac{(40-30)}{(40-30)+(40-10)} \times 5$$

$$\text{Mode} = 49.5 + \frac{10}{10+30} \times 5 = 49.5 + \frac{50}{40} = 49.5 + 1.25 = 50.75$$

$$P_{40} = l + \frac{h}{f}\left(\frac{40\,n}{100} - C.F.\right)$$

$$\frac{40\,n}{100} = \frac{40(100)}{400} = 40$$

$$P_{40} = 44.5 + \frac{5}{30}(40-20) = 44.5 + \frac{(5)\,(20)}{30} = 44.5 + 3.33 = 47.83$$

Example 3.18 (a) A man travels from A to B at an average speed of 30 km per hour and returns from B to A along the same route at an average speed of 60 km per hour. Find the average speed for the entire trip.

Solution The average speed for the entire trip is given by the harmonic mean of 30 and 60 as

$$H = \frac{2}{\frac{1}{30} + \frac{1}{60}} = \frac{2}{3/60} = 2 \times \frac{60}{3} = 40 \text{ km per hour.}$$

Now suppose that distance from A to B is 60 km (although any distance can be assumed). Then it will take 2 hours to travel from A to B at the speed of 30 km per hour. On return, it will take 1 hour because now the person travels at the speed of 60 km per hour.

Total distance = 60 + 60 = 120 km (both ways).

Total time = 2 + 1 = 3 hours.

$$\text{Average speed} = \frac{\text{Total distance}}{\text{Total time}} = \frac{120}{3} = 40 \text{ km per hour.}$$

Thus we see that harmonic mean is the appropriate average here.

Example 3.18 (b) Daska is situated at a distance of 15 miles each from Gujranwala and Sialkot. A bus travels from Gujranwala to Daska at an average speed of 25 miles per hour, from Daska to Sialkot at an average speed of 40 miles per hour and return to Daska at an average speed of 30 miles per hour. Find the average speed for the entire trip.

Solution The average speed for the entire trip is given by the harmonic mean of 25, 40 and 30 as

$$H = \frac{3}{\frac{1}{25} + \frac{1}{40} + \frac{1}{30}} = \frac{3}{0.0400 + 0.0250 + 0.0333} = \frac{3}{0.0983} = 30.52 \text{ miles per hour.}$$

3.5.2 Relation between the Arithmetic Mean, the Geometric Mean and the Harmonic Mean We observe from Examples 3.2(a), 3.14 and 3.17 that the arithmetic mean (156.7) is greater than the geometric mean (155.046) which is in turn greater than the harmonic mean (153.99). The general relation between $\bar{X}$, G and H is

$$\bar{X} \geq G \geq H \quad \text{or} \quad H \leq G \leq \bar{X}$$

The equality sign holds when all the values are equal.

Example 3.19(a) The arithmetic mean, geometric mean and harmonic mean of the distribution of a continuous variable were calculated and recorded as 26.37, 29.84 and 25.83 by some person. Identify the mean, geometric mean and harmonic mean.

Solution Since $\bar{X} > G > H$, we have

$$\bar{X} = 29.84, \ G = 26.37 \text{ and } H = 25.83.$$

Example 3.19(b) If a and b are any two positive numbers and A, G, H are their respective A.M., G.M. and H.M. then shown that $G = \sqrt{A \times H}$

Solution

X	1/X
a	1/a
b	1/b
a + b	$\frac{1}{a} + \frac{1}{b}$

$$\text{A.M.} = \frac{a+b}{2} \qquad \text{G.M.} = \sqrt{a \times b}$$

$$H = \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2}{\frac{b+a}{ab}} = \frac{2ab}{a+b}$$

Now $\sqrt{A \times H} = \left(\frac{a+b}{2} \times \frac{2ab}{a+b}\right) = \sqrt{ab}$

Hence it is proved.

$$G = \sqrt{A \times H}$$

$$\sqrt{ab} = \sqrt{ab}$$

3.6 The Median The median of a set of values arranged in ascending or descending order of magnitude is defined as the middle value if the number of values is odd and the mean of the two middle values if the number of values is even. The median divides a distribution into two halves and the number of values greater than the median is equal to the number of values smaller than the median.

Example 3.20 (a) The median of the values 4, 5, 6, 8, 10, 11 and 12 is 8.

(b) The median of the values, 4, 6, 7, 9, 11 and 13 is $\frac{7+9}{2} = 8$.

When the number of values is *odd*, the median is the middle value and when the number of values is *even*, the median is the mean of the two middle values. In both cases, the median is the value of $\left(\frac{n+1}{2}\right)$th item from either end in the array. In Example 3.20(a), $n = 7$ and the $\left(\frac{7+1}{2}\right)$th item or the 4th item from either end is 8. In Example 3.20(b), $n = 6$ and the median, which is the value of $\left(\frac{6+1}{2}\right)$th or 3.5th item (i.e. half way between the 3rd and the 4th items) from either end in the array, is $\left(\frac{7+9}{2}\right) = 8$.

It should be remembered that the expression $\left(\frac{n+1}{2}\right)$ only tells us the position of the median; it is not the median itself.

The median and the mean do not necessarily have the same value. If we say that median weight of 100 students is 149 pounds, it means that as many students are heavier than 149 pounds as are lighter than 149 pounds. This may not be true of the mean as we can see from the following example. The mean of the value 4, 5, 7, 9 and 25 is 50/5 = 10. But 10 is exceeded by only one of the items; four items are smaller than 10. The median is the value of $\left(\frac{5+1}{2}\right)$th or 3rd item in the array, which is 7.

3.6.1 The Median for Grouped Data We have seen that the median is the value of $\left(\frac{n+1}{2}\right)$th item. In case of a frequency distribution, the median is the value of $\left(\frac{n}{2}\right)$th item from either end. Thus if we have 100 items in a frequency distribution, the median will be the value of the 50th item. To find the median from a frequency distribution, we form a cumulative frequency distribution. The median lies in the class which corresponds to the cumulative frequency in which $\left(\frac{n}{2}\right)$ lies. It is given the formula (obtained by interpolation)

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - F\right) \tag{3.17}$$

l = lower class boundary of the median class, i.e. the class corresponding to the cumulative frequency in which $(n/2)$ lies

h = class interval size of the median class

f = frequency of the median class

n = number of values or the total frequency

F = cumulative frequency of the class preceding the median class

Geometrically median is the value of X (abscissa) corresponding to the vertical line which divides a histogram into two parts having equal areas. This value of X is sometimes denoted by $\tilde{X}$ (X-childa).

Example 3.21(a) Find median form the given data. (B.I.S.E., 2012)

$C-I$	F	$C.B.$	f
0 — 9	2	–0.5 — 9.5	2
10 — 19	3	9.5 — 19.5	1
20 — 29	11	19.5 — 29.5	8
30 — 39	24	29.5 — 39.5	13
40 — 49	32	39.5 — 49.5	8
50 — 59	40	49.5 — 59.5	8

$$\text{Median} = l + \frac{i}{f}\left(\frac{n}{2} - C.F.\right)$$

Here $n = 40$, $i = 10$

First of all we find

$$\frac{n}{2} = \frac{40}{2} = 20$$

$$f = 13, \quad l = 29.5$$

$$\text{So } \tilde{X} = 29.5 + \frac{10}{13}\left(\frac{40}{2} - 11\right) = 29.5 + \frac{10}{13}(20 - 11) = 29.5 + \frac{10}{13}(9) = 29.5 + \frac{90}{13}$$

$$= 29.5 + 6.92 = 36.42$$

Example 3.21(b) The class mark for the ages of sales clerks employed in a departmental store are: 18.5, 28.5, 38.5, 48.5, 58.5 and 68.5. Find the class-boundaries at this distribution and compute median if the class frequencies are 7, 12, 23, 35, 25, 8 respectively.

(B.I.S.E. Gujranwala 2015)

Sol.

Class Marks X	f	C.B.	C.F.
18.5	7	13.5 — 23.5	7
28.5	12	23.5 — 33.5	19
38.5	23	33.5 — 43.5	42
48.5	35	43.5 — 53.5	77
58.5	25	53.5 — 63.5	102
68.5	8	63.5 — 73.5	110

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

First of all we find

$$\frac{n}{2} = \frac{110}{2} = 55$$

$l = 43.5$, $h = 10$

$$\text{Median} = 43.5 + \frac{10}{35}(55 - 42) = 43.5 + \frac{10}{35}(13) = 43.5 + \frac{130}{35} = 43.5 + 3.7143$$

$$= 47.2143$$

Example 3.21(c) Find median (B.I.S.E., Multan Practical 2014)

u	f	$u = \dfrac{X\ 130}{20}$	C.B.	C.F.
–2	07	90	80 – 100	7
–1	50	110	100 – 120	57
0	80	130	120 – 140	137
1	60	150	140 – 160	197
2	03	170	160 – 180	200

$$\tilde{X} = 1 + \frac{1}{f}\left(\frac{n}{2} - C\right) = 120 + \frac{20}{80}(100 - 57) = 120 + 0.25(43) = 130.75$$

3.6.2 Median from Discrete Data In case of discrete data grouped by sizes, the median is the size of the $\left(\frac{n+1}{2}\right)$th item as in the case of ungrouped data. To find the median from discrete data, we form a cumulative frequency distribution. The median is the value corresponding to the cumulative frequency in which $(n + 1)/2$ lies.

Table below gives the child distribution for 123 families of a locality. The cumulative frequencies have been shown in the last column. Since $(n + 1)/2 =$ $(123 + 1)/2 = 62$ lies in the cumulative frequency corresponding to 2, the median is 2 as indicated by the arrow.

Number of Children	Number of Families	Cumulative Frequency
0	4	4
1	25	29
2 ←	53	82
3	18	100
4	14	114
5	7	121
6	2	123
	$n = 123$	

3.6.3 Graphic Location of Median The approximate value of the median can be located from an ogive (a cumulative frequency polygon). In an ogive, the median is the value of X corresponding to $n/2$. Thus to locate median, we mark $n/2$ along the Y-axis and draw a perpendicular from this point of the Y-axis and extend it so as to intersect the ogive. Then we drop a perpendicular on the X-axis from the point of intersection. The point at which the perpendicular intersects the X-axis is the value of the median. Fig. 3.1 shows the graphic location of median for the frequency distribution of weights in Table 2.3

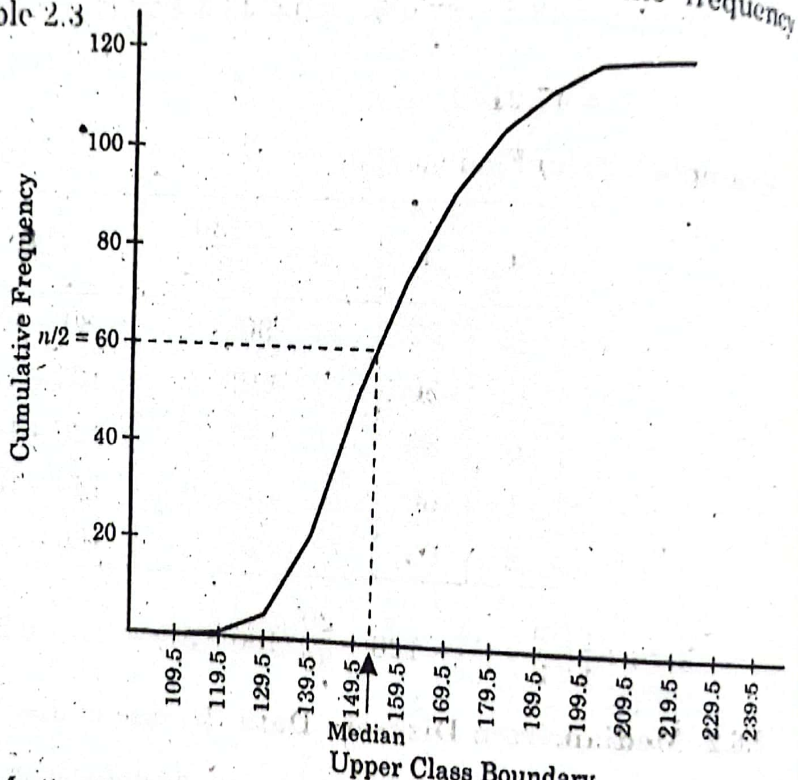

Fig. 3.1 Graphic Location of Median for the Frequency Distribution of Weights of 120 students

3.6.4 Quartiles, Deciles and Percentiles We know that the median of an array is the middle value (or the mean of the two middle values). It divides a set of data into two equal parts. There are also certain other values which divide a set of data into four, ten or hundred parts. The values which divide an array into four equal parts are called the *first, second and third quartiles* and are denoted by Q_1, Q_2 and Q_3 respectively. The first and third quartiles are also called the *lower* and *upper* *quartiles* respectively. The second quartile is the median. The quartiles are given by the formulae

$$\left.\begin{aligned} Q_1 &= \text{Value of } \frac{(n+1)}{4} \text{th item} \\ Q_2 &= \text{Value of } \frac{2(n+1)}{4} \text{th item or } \frac{(n+1)}{2} \text{th item} \\ Q_3 &= \text{Value of } \frac{3(n+1)}{4} \text{th item} \end{aligned}\right\} \quad (3.18)$$

The values which divide an array into ten equal parts are called *deciles*. The first, second, ..., ninth deciles are denoted by $D_1, D_2, \ldots, D_9$ respectively. The deciles are given by the formulae

$$\left.\begin{aligned} D_1 &= \text{Value of } \frac{(n+1)}{10} \text{th item} \\ D_2 &= \text{Value of } \frac{2(n+1)}{10} \text{th item} \\ &\vdots \\ D_9 &= \text{Value of } \frac{9(n+1)}{10} \text{th item} \end{aligned}\right\} \quad (3.19)$$

The fifth decile (D_5) corresponds to the median.

Finally the values which divide an array into one hundred equal parts are called *percentiles*. The first, second, ..., ninety-ninth percentiles are denoted by $P_1, P_2, \ldots, P_{99}$ respectively. The percentiles are given by the formulae

$$\left.\begin{aligned} P_1 &= \text{Value of } \frac{(n+1)}{100} \text{th item} \\ P_2 &= \text{Value of } \frac{2(n+1)}{100} \text{th item} \\ &\vdots \\ P_{99} &= \text{Value of } \frac{99(n+1)}{100} \text{th item} \end{aligned}\right\} \quad (3.20)$$

The 50th percentile (P_{50}) corresponds to the median. The 25th percentile (P_{25}) and 75th percentile (P_{75}) correspond to the first and third quartiles respectively.

Collectively the quartiles, deciles, percentiles and other values obtained by equal sub-division of the data are called *quantiles*. The second, fourth, sixth and eighth deciles which divide the data into five equal parts are also called *quintiles*.

Example 3.22 . Find all the quartiles, second, third and seventh deciles and the fifteenth, thirty-seventh and sixty-fourth percentiles from the following marks obtained by 20 students on a test in statistics.

53	74	82	42	39	20	81	68	58	28	67	54	93
70	30	55	36	37	29	61						

Solution The marks of $n = 20$ students are arranged in ascending order as follows:

20	28	29	30	36	37	39	42	53	54	55	58	61
67	68	70	74	81	82	93						

Q_1 = value of $\frac{(n+1)}{4}$th or $\frac{(20+1)}{4}$th or 5.25th item from below. The value of the 5th item is 36 and that of the 6th item is 37. Thus the first quartile is a value 0.25th of the way between 36 and 37, which is 36.25. Hence $Q_1 = 36.25$.

Q_2 (median) = value of $\frac{2(n+1)}{4}$th or $\frac{2(20+1)}{4}$th or 10.5th item from below. The value of 10th item is 54 and that of the 11th item is 55. Thus the second quartile is a value 0.5th of the way between 54 and 55, which is 54.5. Hence $Q_2 = 54.5$.

Q_3 = value of $\frac{3(n+1)}{4}$th or $\frac{3(20+1)}{4}$th or 15.75th item from below. The value of the 15th item is 68 and that of the 16th item is 70. Thus the third quartile is a value 0.75th of the way between 68 and 70, which is 68 + 2(0.75) = 69.5. Hence $Q_3 = 69.5$.

D_2 = value of $\frac{2(n+1)}{10}$th or $\frac{2(20+1)}{10}$th or 4.2th item from below. The value of the 4th item is 30 and that of the 5th item is 36. Thus the second decile is a value 0.2th of the way between 30 and 36, which is 30 + 6(0.2) = 31.2. Hence $D_2 = 31.2$.

D_3 = value of $\frac{3(n+1)}{10}$th or $\frac{3(20+1)}{10}$th or 6.3th item from below. The value of the 6th item is 37 and that of the 7th is 39. Thus the third decile is 0.3th of the way between 37 and 39, which is 37 + 2(0.3) = 37.6. Hence $D_3 = 37.6$.

D_7 = value of $\frac{7(n+1)}{10}$th or $\frac{7(20+1)}{10}$th or 14.7th item from below. The value of the 14th item is 67 and that of the 15th item is 68. Thus the 7th decile is 0.7th of the way between 67 and 68, which is 67 + 0.7(1) = 67.7. Hence $D_7 = 67.7$.

P_{15} = value of $\frac{15(n+1)}{100}$th or $\frac{15(20+1)}{100}$th or 3.15th item from below. The value of the 3rd item is 29 and that of the 4th item is 30. Thus the 15th percentile is 0.15th of the way between 29 and 30, which is 29 + 0.15(1) = 29.15. Hence $P_{15} = 29.15$.

P_{37} = value of $\frac{37(n+1)}{100}$th or $\frac{37(20+1)}{100}$th or 7.77th item from below. The value of the 7th item is 39 and that of the 8th item is 42. Thus the 37th percentile is 0.77th of the way between 39 and 42, which is 39 + 3(0.77) = 41.31. Hence $P_{37} = 41.31$.

P_{64} = value of $\frac{64(n+1)}{100}$th or $\frac{64(20+1)}{100}$th or 13.44th item from below. The value of the 13th item is 61 and that of the 14th item is 67. Thus the 64th percentile is 0.44th of the way between 61 and 67, which is 61 + 6(0.44) = 63.64. Hence P_{64} = 63.64.

3.6.5 Quantiles from Grouped Data The quartiles, deciles and percentiles may be determined from grouped data in the same way as the median except that in place of $n/2$ we will use $n/4$, $n/10$ and $n/100$. For example, to find Q_1 and Q_3 we will form cumulative frequencies and find the classes which contain $\frac{n}{4}$ th and $\frac{3n}{4}$ th items. Thus Q_1 and Q_3 are given by the formulae

$$\left.\begin{aligned} Q_1 &= l + \frac{h}{f}\left(\frac{n}{4} - F\right) \\ Q_3 &= l + \frac{h}{f}\left(\frac{3n}{4} - F\right) \end{aligned}\right\} \qquad (3.21)$$

where l, h, f and F are defined in their respective quartiles. Thus for Q_1 (or Q_3), we have

l = lower class boundary of the class containing Q_1 (or Q_3), i.e. the class corresponding to the cumulative frequency in which $n/4$ (or $3n/4$) lies.

h = class interval size of the class containing Q_1 (or Q_3).

f = frequency of the class containing Q_1 (or Q_3).

F = cumulative frequency of the class preceding the class containing Q_1 (or Q_3).

The deciles and the percentiles from grouped data are given by

$$\left.\begin{aligned} D_1 &= l + \frac{h}{f}\left(\frac{n}{10} - F\right) \\ D_2 &= l + \frac{h}{f}\left(\frac{2n}{10} - F\right) \\ &\vdots \\ D_9 &= l + \frac{h}{f}\left(\frac{9n}{10} - F\right) \end{aligned}\right\} \qquad (3.22)$$

$$\left.\begin{aligned} P_1 &= l + \frac{h}{f}\left(\frac{n}{100} - F\right) \\ P_2 &= l + \frac{h}{f}\left(\frac{2n}{100} - F\right) \\ &\vdots \\ P_{51} &= l + \frac{h}{f}\left(\frac{51n}{100} - F\right) \\ &\vdots \\ P_{99} &= l + \frac{h}{f}\left(\frac{99n}{100} - F\right) \end{aligned}\right\} \qquad (3.23)$$

From D_1, D_4, D_7 and D_9, we conclude that 10% of the students have weights 133.62 pounds or less, 40% of the students have weights 148.8 pounds or less, 70% of the students have weights 164.5 pounds or less and 90% of the students have weights 182.83 pounds or less.

79.22   71.17   79.79   72.24   79.56   72.87   75.48

Identify the median and the quantiles.

Solution Arranging the values in ascending order we have

71.17   72.24   72.87   75.48   79.22   79.56   79.79

The median and quantiles are identified below:

Q_1	< D_3	< P_{37}	< Median	< D_7	< P_{72}	< Q_3
(25%)	(30%)	(37%)	(50%)	(70%)	(72%)	(75%)
71.17	72.24	72.87	75.48	79.22	79.56	79.79

3.7 The Mode The *mode* is defined as that value in the data which occurs the greatest number of times provided such a value exists.

Example 3.26(a) (i) The mode of the values 2, 5, 7, 8, 9, 9, 9 and 10 is 9.
(ii) The mode of the values 11, 12, 12, 14, 15, 15, 15, 17, 17 and 19 is 15.

If each value occurs the same number of times, then there is *no mode*. If two or more values occur the same number of times but more frequently than any of the other values, then there is more than one mode. In this respect the mode differs from the mean and the median because there is only one mean and only one median.

Example 3.26(b) (i) The set of values 2, 3, 4, 4, 4, 6, 8, 9, 9 and 9 has two modes 4 and 9.

(ii) The set of values 1, 2, 5, 6, 12, 13 and 14 has no mode.

A distribution having only one mode is called *uni-modal distribution*, a distribution having two modes is called *bi-modal distribution* and a distribution having more than two modes is called a *multi-modal distribution*.

3.7.1 The Mode from Grouped Data In case of a unimodal frequency distribution where a frequency curve has been constructed for the data, the mode is defined as that value of X which corresponds to the highest point on the curve. This value of X is sometimes denoted by $\hat{X}$ (read as X-caret).

In a frequency distribution with equal class interval sizes, the class with the highest frequency is called the modal class. The mode is given by the formula.

$$\text{Mode} = l + \frac{(f_m - f_{m-1})}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \times h \tag{3.24}$$

where l = lower class boundary of the modal class, i.e. the class with the highest frequency

f_m = frequency of the modal class

f_{m-1} = frequency of the class preceding the modal class

f_{m+1} = frequency of the class following the modal class

h = class interval size of the modal class

Example 3.28 Given below is the height distribution of 60 college students

Height	145–149.9	150–154.9	155–159.9	160–164.9
Number of students	2	5	9	15
Height	165–169.9	170–174.9	175–179.9	180–184.9
Number of students	16	7	5	1

Draw the histogram for this distribution and find the modal height, then to check this result by using the algebraic formula.

Solution

Check by algebraic formula.

Classes	f	Class boundaries
145 – 149.9	2	144.95 – 149.95
150 – 154.9	5	149.95 – 154.95
155 – 159.9	9	154.95 – 159.95
160 – 164.9	15	159.95 – 164.95
165 – 169.9	16	164.95 – 169.95
170 – 174.9	7	169.95 – 174.95
175 – 179.9	5	174.95 – 179.95
180 – 184.9	1	179.95 – 184.95

Model class is 164.95 — 169.95 which corresponds to maximum frequency.

$$\hat{X} = \text{Modal} = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h = 164.95 + \frac{16 - 15}{(16 - 15) + (16 - 7)} \times 5$$

$$= 164.95 + \frac{1}{1 + 9} \times 5 = 164.95 + \frac{1}{10} \times 5 = 165.45$$

3.7.2 Mode from Discrete Data In case of discrete data, the mode may be found by inspection. It is the most common value i.e. the value with greatest frequency. The following table gives the number of times various black cards appeared in 36 deals of 5 playing cards. The data are discrete because we may get 4 or 5 black cards but never 4.5 or 4.33 black cards. The most common number of black cards is 2 (corresponding to the greatest frequency) which is, therefore, the mode.

Number of Black Cards	0	1	2	3	4	5
Frequency	2	7	12	9	5	1

Example 3.29 Find the median and the mode for each of the following distributions.

(a)

Marks	No. of Students
Less than 40	0
Less than 45	5
Less than 50	25
Less than 55	56
Less than 60	78
Less than 65	95
Less than 70	100

(b)

Daily Wage (Rs)	No. of Workers
More than 55	5
More than 50	17
More than 45	28
More than 40	38
More than 35	47
More than 30	49
More than 25	50

Solution (a) This is a "*less than*" cumulative frequency distribution. The frequency distribution is given below.

Marks	Frequency	Cumulative Frequency
40 – 45	5	5
45 – 50	20	25
50 – 55	31	56
55 – 60	22	78
60 – 65	17	95
65 – 75	5	100
	$n = \sum f = 100$	

Here $n/2 = 100/2 = 50$ which lies in the class 50 – 55; this is the median class. Thus

$$\text{Median } = l + \frac{h}{f}\left(\frac{n}{2} - F\right) = 50 + \frac{5}{31}(50 - 25) = 50 + 4.032$$

$$= 54.032 \text{ marks or } 54 \text{ marks.}$$

The maximum frequency lies in the class 50 – 55. This is the modal class. Thus

$$\text{Mode } = l + \frac{(f_m - f_{m-1})}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \times h = 50 + \frac{(31 - 20)}{(31 - 20) + (31 - 22)} \times 5$$

$$= 50 + 2.75 = 52.75 \text{ marks or } 53 \text{ marks.}$$

(b) This is a "*more than*" cumulative or a *decumulative* frequency distribution. The frequency distribution is given below.

Since $n/2 = 50/2 = 25$ lies in the class 45 – 50, this is the median class. Thus

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - F\right) = 45 + \frac{5}{11}(25 - 22) = 45 + 1.364 = \text{Rs.}46.36.$$

Daily Wage (Rs)	Frequency	Commulative Frequency
25 – 30	1	1
30 – 35	2	3
35 – 40	9	12
40 – 45	10	22
45 – 50	11	33
50 – 55	12	45
55 – 60	5	50
	$n = \sum f = 50$	

The maximum frequency lies in the class 50 – 55; this is the modal class. Thus

$$\text{Mode } = l + \frac{(f_m - f_{m-1})}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \times h = 50 + \frac{(12 - 11)}{(12 - 11) + (12 - 5)} \times 5$$

$$= 50 + 0.625 = \text{Rs.}50.62.$$

Example 3.30(a) Find the median and the mode for the frequency distribution of Example 3.6(b).

Solution Since $u = \dfrac{X - 130}{5}$ or $X = 130 + 5u$, the values of X are obtained from this relation. Since $h = 5$, The class boundaries are obtained by subtracting and adding 2.5 (half of $h/2$). These values are given in the following table.

u	f	$X = 130+5u$	Class Boundaries	Cumulative Frequency (F)
–3	5	115	112.5 – 117.5.	5
–2	12	120	117.5 – 122.5	17
–1	26	125	122.5 – 127.5	43
0	32	130	127.5 – 132.5	75
1	13	135	132.5 – 137.5	88
2	8	140	137.5 – 142.5	96
3	4	145	142.5 – 147.5	100
$n = \sum f = 100$				

Since $n/2 = 50$ lies in the class 127.5 – 132.5, this is the median class. Thus

$$\text{Median } = l + \frac{h}{f}\left(\frac{n}{2} - F\right) = 127.5 + \frac{5}{32}(50 - 43) = 127.5 + 1.09 = 128.59.$$

The maximum frequency lies in the class 127.5 – 132.5; this is the modal class. Thus

$$\text{Mode } = l + \frac{(f_m - f_{m-1})}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \times h = 127.5 + \frac{(32 - 26)}{(32 - 26) + (32 - 13)} \times 5$$

$$= 127.5 + 1.2 = 128.7.$$

3.5 The Empirical Relation between the Mean, Median and Mode For unimodal frequency curves which are moderately skewed (asymmetrical), we have the following empirical relation

$$\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median}) \qquad (3.25)$$

Figures 3.3(a) and 3.3(b) show the relative positions of the mean, median and mode for distributions which are skewed to the right and left respectively. As we can see from Figs. 3.3(a) and 3.3(b), the median lies between the mean and the mode and it is twice as far from the mode as from the mean. It is interesting to note that the mean, median and mode occur in the same order as they occur in the dictionary in distributions skewed to the left and in reverse order in distributions skewed to the right.

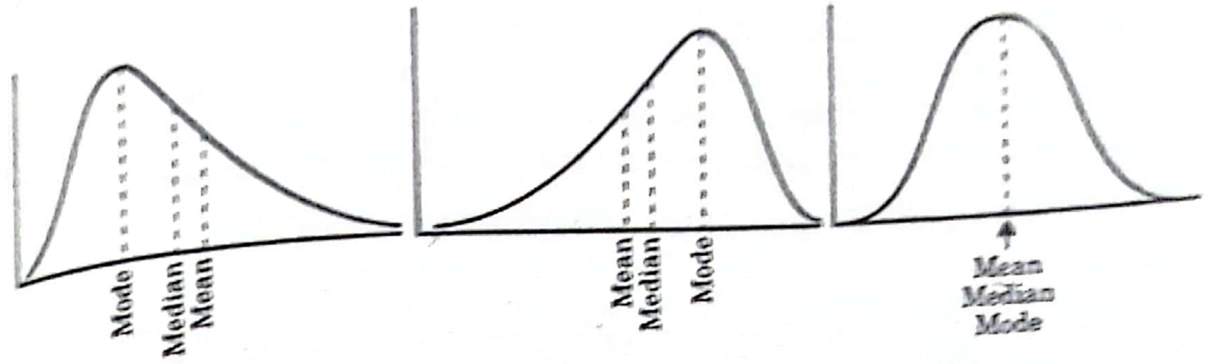

Fig. 3.3(a) Fig. 3.3(b) Fig. 3.3(c)

In symmetrical distributions, the mean, median and mode all coincide. This is shown in Fig. 3.3(c).

The relation (3.25) can be used to find the mode when we have computed the values of mean and median in moderately skewed distributions. This gives

$$\text{Mode} = 3\text{Median} - 2\text{Mean} \qquad (3.26)$$

From Examples 3.2 (a) and 3.21, the mean and median for the distribution of weights in Table 3.1 are 156.17 pounds and 153.5 pounds respectively. Substituting these values in Formula (3.26), we find

$$\text{Mode} = 3\text{ Median} - 2\text{ Mean} = 3(153.5) - 2(156.17) = 460.5 - 312.34$$

$$= 148.16 \text{ pounds}$$

which is not the same as found in Example 3.28.

Example 3.31 (a) In a symmetrical distribution mean is 60. What is median and mode?

Solution In a symmetrical distribution the mean, median and mode coincide. Hence the median and the mode are also 60.

Example 3.31 (b) In a moderately skewed distribution, mean is 82 and median is 78. What is the approximate value of mode.

Solution Mode = 3 Median − 2 Mean = 3(78) − 2(82) = 70.