Harmonic mean : Harmonic mean of a set of observation is defined as the reciprocal of the arithmetic mean of the reciprocals of the observation

~~$H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(1/x_i)}$~~

Therefore $x_1, x_2, \ldots, x_n$ are 'n' observations

Then their

$$\text{harmonic mean } (H) = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(1/x_i)}$$

In the case of a freaquncy distribution

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}(f_i/x_i)}, \quad N = \sum_{i=1}^{n} f_i$$

Here $x_i \neq 0$

Find the H M of the numbers

x 1, 2, 3

$$H = \frac{1}{\frac{1}{3}\left(\frac{6}{6}+\frac{3}{6}+\frac{2}{6}\right)} = \frac{1}{\frac{1}{3}\left(\frac{11}{6}\right)} = \frac{1}{\frac{11}{18}} = \frac{18}{11} = 1.72\ldots$$

1.7
11) 18
11
80
77

2·9. Harmonic Mean. Harmonic mean of a number of observations is the reciprocal of the arithmetic mean of the reciprocals of the given values. Thus, harmonic mean H, of n observations x_i, $i = 1, 2, ..., n$ is

$$H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(1/x_i)} \qquad ...(2·12)$$

In case of frequency distribution $x_i | f_i$, $(i = 1, 2, ..., n)$,

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}(f_i/x_i)}, \quad \left[N = \sum_{i=1}^{n} f_i\right] \qquad ...(2·12a)$$

2·9·1. Merits and Demerits of Harmonic Mean

Merits. Harmonic mean is rigidly defined, based upon all the observations and is suitable for further mathematical treatment. Like geometric mean, it is not affected much by fluctuations of sampling. It gives greater importance to small items and is useful only when small items have to be given a greater weightage.

Demerits. Harmonic mean is not easily understood and is difficult to compute.

Example 2·13. *A cyclist pedals from his house to his college at a speed of 10 m.p.h. and back from the college to his house at 15 m.p.h. Find the average speed.*

Solution. Let the distance from the house to the college be x miles. In going from house to college, the distance (x miles) is covered in $\frac{x}{10}$ hours, while in coming from college to house, the distance is covered in $\frac{x}{15}$ hours. Thus a total distance of $2x$ miles is covered in $\left(\frac{x}{10} + \frac{x}{15}\right)$ hours.

$$\text{Hence average speed} = \frac{\text{Total distance travelled}}{\text{Total time taken}} = \frac{2x}{\left(\frac{x}{10} + \frac{x}{15}\right)}$$

$$= \frac{2}{\left(\frac{1}{10} + \frac{1}{15}\right)} = 12 \text{ m.p.h.}$$

Remark. 1. In this case the average speed is given by the harmonic mean of 10 and 15 and not by the airthmetic mean.

Rather, we have the following general result :

If equal distances are covered (travelled) per unit of time with speeds equal to $V_1, V_2, ..., V_n$, say, then the average speed is given by the harmonic mean of $V_1, V_2, ..., V_n$, *i.e.*,

$$\text{Average speed} = \frac{n}{\left(\frac{1}{V_1} + \frac{1}{V_2} + ... + \frac{1}{V_n}\right)} = \frac{n}{\Sigma\left(\frac{1}{V}\right)}$$

Proof is left as an exercise to the reader.

Hint. $\text{Speed} = \dfrac{\text{Distance}}{\text{Time}} \Rightarrow \text{Time} = \dfrac{\text{Distance}}{\text{Speed}}$

$$\text{Average Speed} = \frac{\text{Total distance travelled}}{\text{Total time taken}}$$

2. Weighted Harmonic Mean. Instead of fixed (constant) distance being travelled with varying speed, let us now suppose that different distances, say, $S_1, S_2, ..., S_n$, are travelled with different speeds, say, $V_1, V_2, ..., V_n$ respectively. In that case, the average speed is given by the weighted harmonic mean of the speeds, the weights being the corresponding distances travelled, *i.e.*,

$$\text{Average speed} = \frac{S_1 + S_2 + \ldots + S_n}{\left(\frac{S_1}{V_1} + \frac{S_2}{V_2} + \ldots + \frac{S_n}{V_n}\right)} = \frac{\Sigma S}{\Sigma\left(\frac{S}{V}\right)}$$

Example 2·14. *You can take a trip which entails travelling 900 km. by train at an average speed of 60 km. per hour, 3000 km. by boat at an average of 25 km. p.h., 400 km. by plane at 350 km. per hour and finally 15 km. by taxi at 25 km. per hour. What is your average speed for the entire distance ?*

Solution. Since different distances are covered with varying speeds, the required average speed for the entire distance is given by the weighted harmonic mean of the speeds (in km.p.h.), the weights being the corresponding distances covered (in kms.).

COMPUTATION OF WEIGHTED H. M.

Speed (*km. / hr.*) X	*Distance* (*in km.*) W	W/X
60	900	15·00
25	3000	120·00
350	400	1·43
25	15	0·60
Total	$\Sigma W = 4315$	$\Sigma (W/X) = 137·03$

Average speed

$$= \frac{\Sigma W}{\Sigma (W/X)} = \frac{4315}{137·03} = 31·489 \text{ km.p.h.}$$

2·10. Selection of an Average. From the preceding discussion it is evident that no single average is suitable for all practical purposes. Each one of the average has its own merits and demerits and thus its own particular field of importance and utility. We cannot use the averages indiscriminately. A judicious selection of the average depending on the nature of the data and the purpose of the enquiry is essential for sound statistical analysis. Since arithmetic mean satisfies all the properties of an ideal average as laid down by Prof. Yule, is familiar to a layman and further has wide applications in statistical theory at large, it may be regarded as the best of all the averages.

2·11. Partition Values. These are the values which divide the series into a number of equal parts.

The three points which divide the series into four equal parts are called *quartiles*. The first, second and third points are known as the first, second and third quartiles respectively. The first quartile, Q_1, is the value which exceed 25% of the observations and is exceeded by 75% of the observations. The second quartile, Q_2, coincides with median. The third quartile, Q_3, is the point which has 75% observations before it and 25% observations after it.

The nine points which divide the series into ten equal parts are called *deciles* whereas *percentiles* are the ninety-nine points which divide the series into hundred equal parts. For example, D_7, the seventh decile, has 70% observations before it and P_{47}, the forty-seventh percentile, is the point which exceed 47% of the observations. The methods of computing the partition values are the same as those of locating the median in the case of both discrete and continuous distributions.

Example 2·15. *Eight coins were tossed together and the number of heads resulting was noted. The operation was repeated 256 times and the frequencies (f) that were obtained for different values of x, the number of heads, are shown in the following table. Calculate median, quartiles, 4th decile and 27th precentile.*

x :	0	1	2	3	4	5	6	7	8
f :	1	9	26	59	72	52	29	7	1

Solution.

x :	0	1	2	3	4	5	6	7	8
f :	1	9	26	59	72	52	29	7	1
c.f. :	1	10	36	95	167	219	248	255	256

Median : Here $N/2 = 256/2 = 128$. Cumulative frequency (*c.f.*) just greater than 128 is 167. Thus, median = 4.

Q_1 : Here $N/4 = 64$. *c.f.* just greater than 64 is 95. Hence, $Q_1 = 3$.

Q_3 : Here $3N/4 = 192$ and *c.f.* just greater than 192 is 219. Thus $Q_3 = 5$.

D_4 : $\dfrac{4N}{10} = 4 \times 25{\cdot}6 = 102{\cdot}4$ and *c.f.* just greater than 102·4 is 167. Hence $D_4 = 4$.

P_{27} : $\dfrac{27N}{100} = 27 \times 2{\cdot}56 = 69{\cdot}12$ and *c.f.* just greater than 69·12 is 95. Hence $P_{27} = 3$.

2·11·1. Graphical Location of the Partition Values. The partition values, *viz.*, quartiles, deciles and percentiles, can be conveniently located with the help of a curve called the 'cumulative frequency curve' or 'Ogive'. The procedure is illustrated below.

First form the cumulative frequency table. Take the class intervals (or the variate values) along the *x*–axis and plot the corresponding cumulative frequencies along the *y*–axis against the *upper limit* of the class interval (or against the variate value in the case of discrete frequency distribution). The curve obtained on joining the points so obtained by means of free hand drawing is called the *cumulative frequency curve* or *ogive*. The graphical location of partition values from this curve is explained below by means of an example.

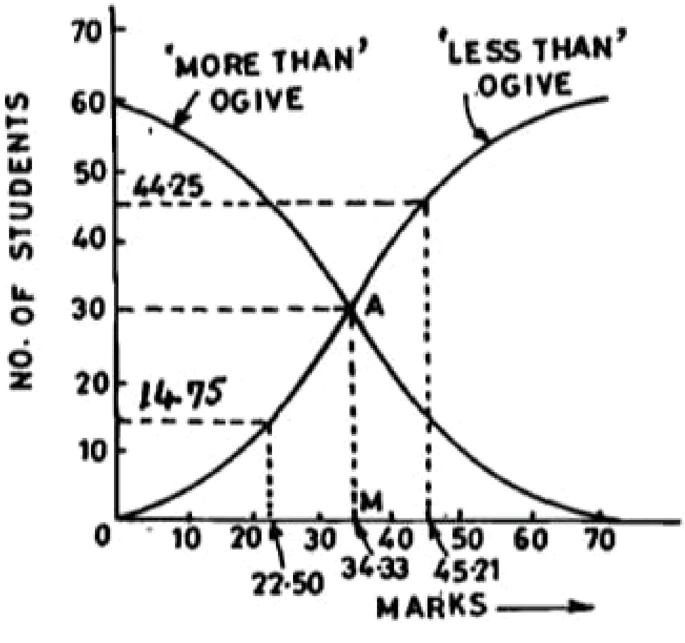
Example 2·16. *Draw the cumulative frequency curve for the following distribution showing the number of marks of 59 students in Statistics.*

Marks-group :	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of Students :	4	8	11	15	12	6	3

Solution.

Marks-group	*No. of Students*	*Less than c.f.*	*More than c.f.*
0—10	4	4	59
10—20	8	12	55
20—30	11	23	47
30—40	15	38	36
40—50	12	50	21
50—60	6	56	9
60—70	3	59	3

Taking the marks–group along x-axis and *c.f.* along y-axis, we plot the cumulative frequencies, *viz.*, 4, 12, 23, ..., 59 against the upper limits of the corresponding classes, *viz.*, 10, 20, ..., 70 respectively. The smooth curve obtained on joining these points is called *ogive* or more particularly *'less than' ogive*.



If we plot the 'more than' cumulative frequencies, *viz.*, 59, 55,..., 3 against the lower limits of the corresponding classes, *viz.*, 0, 10, ..., 60 and join the points by a smooth curve, we get cumulative frequency curve which is also known as *ogive* or more particularly *'more than' ogive*.

To locate graphically the value of median, mark a point corresponding to $N/2$ along y-axis. At this point draw a line parallel to x-axis meeting the ogive at the point 'A' (say). From 'A' draw a line prependicular to x-axis meeting it in 'M' (say). Then abscissa of 'M' gives the value of median.

To locate the values of Q_1 (or Q_3), we mark the points along y-axis corresponding to $N/4$ (or $3N/4$) and proceed exactly similarly.

In the above example, we get from ogive

Median= 34·33, Q_1 = 22·50, and Q_3 = 45·21.

Remarks. 1. The median can also be located as follows :

From the point of intersection of 'less than' ogive and 'more than' ogive, draw perpendicular to OX. The abscissa of the point so obtained gives median.

2. Other partition values, *viz.*, deciles and percentiles, can be similarly located from 'ogive'.

33. Find the minimum value of :

(i) $f(x) = (x-6)^2 + (x+3)^2 + (x-8)^2 + (x+4)^2 + (x-3)^2$

(ii) $g(x) = |x-6| + |x+3| + |x-8| + |x+4| + |x-3|$.

[Delhi Univ. B.Sc.(Stat. Hons.), 1991]

Hint. The sum of squares of deviations is minimum when taken from arithmetic mean and the sum of absolute deviations is minimum when taken from median.