3 Measures Of Location

$0 \leq P(A) \leq 1$

3.1 Introduction

The diagrammatic representation of a set of data can give us some impressions about its distribution. Even then, there remains a need for a single quantitative measure which could be used to indicate the centre of the distribution. The measures commonly used for this purpose are **mean**, **median** and **mode**. **Geometric mean** and **harmonic mean** are also sometimes used. These measures are single values, which represent the given data and are known as averages or measures of location or measures of central tendency. The name measures of location arises as these measures give an indication where to locate a distribution; The decision as to which measure is to be used depends upon the particular situation under consideration.

Properties of a good average are: -

i) It is well defined.
ii) It is easy to calculate.
iii) It is easy to understand.
iv) It is based on all the values.
v) It is capable of mathematical treatment.

The important types of averages are:

i) Arithmetic mean and Weighted mean.
ii) Geometric mean
iii) Harmonic mean
iv) Median
v) Mode

3.2 Arithmetic Mean And Weighted Mean

Arithmetic Mean is calculated by adding up all the observations and dividing the sum by the total number of observations. The Greek letter μ (meu) is used as a symbol for the population mean.

The population mean of N observations $Y_1, Y_2, \ldots, Y_N$ is defined as:

$$\mu = \frac{1}{N}(Y_1 + Y_2 + \ldots + Y_N)$$

$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N} Y_i$$

$$= \frac{\sum Y}{N} \qquad (3.1)$$

The μ is a parameter, a fixed value and is usually unknown in practice. It is also called as arithmetic mean, abbreviated as A.M. The estimate of population mean μ is sample mean and is denoted by $\overline{Y}$. $\overline{Y}$ is a statistic and its value varies from sample to sample drawn from the population. The sample mean $\overline{Y}$ of n observations $y_1, y_2, \ldots, y_n$ from a population is defined as:

$$\overline{Y} = \frac{1}{n}(y_1 + y_2 + \ldots y_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$= \frac{\sum y}{n} \qquad (3.2)$$

The sample mean is a good estimate of the population mean as it is an unbiased statistic. Unbiased simply means that if means are calculated for all possible samples drawn from the population, the mean of these sample means would be equal to the population mean. The arithmetic mean has the same units as the original observation i.e., if the original observations are in centimeters, the unit of mean would be in centimeters.

Example 3.1: Arithmetic mean for ungrouped data.

Following are the data on students heights (cms).

87, 91, 89, 88, 89, 91, 87, 92, 90, 98.

There are ten observations and their sum is 902.

											Total
Y_i	89	91	89	88	89	91	87	92	90	98	902

$$\text{Arithmetic mean } (\overline{Y}) = \frac{\sum y_i}{n}$$

$$= \frac{902}{10}$$

$$= 90.20\ cm.$$

3.2.1 Updating or correcting the mean

It may happen that an observation was overlooked while calculating the mean. In such situations, the observation may be incorporated.

Suppose that in the above example 3.1, it was found later on that the last observation is 94 instead of 98. To correct the mean, we proceed as follows:

The mean of n observation is given by:

$$\bar{Y} = \frac{\sum y_i}{n} \Rightarrow \sum y = n\bar{Y}$$

$$= 10(902) = 9020$$

The corrected total can be obtained by subtracting 98 and adding 94 from the total 9020 so the corrected sum is given by

Corrected $\Sigma Y = 9020 - 98 + 94 = 9016$

So corrected mean is given by

$$\bar{Y} = \frac{\Sigma y_i}{n}$$

$$= 9016/10$$

$$= 90.16$$

A.M for grouped data

When data is lengthy, it is usually grouped into different classes and $\bar{Y}$ is calculated by the following formula:

$$\bar{Y} = \frac{f_1 y_1 + f_2 y_2 + ... + f_k y_k}{f_1 + f_2 + ... + f_k}$$

$$= \frac{\sum_{i=1}^{k} f_i y_i}{\sum_{i=1}^{k} f_i} = \frac{\sum f_i y_i}{\sum f_i} \qquad (3.3)$$

where k is the number of classes. y_i is the mid point of the ith class and f_i is the corresponding frequency. Here, it is assumed that each of the observations is equal to the mid point of the class in which it occurs. This causes the value of arithmetic mean a bit different for grouped data when the same is calculated from ungrouped data. This difference is called grouping error.

4/4

Example 3.2: Find arithmetic mean consider the grouped student height data as in the table 2.1.

Solution: The first two columns indicate frequency distribution and the columns 3 and 4 are useful to calculate mean according to the definition.

C.1	f	**Mid point** y_i	$f_i y_i$
86-90	6	88	528
91-95	4	93	372
96-100	10	98	980
101-105	6	103	618
106-110	3	108	324
111-115	1	113	113
Total	30	---	2935

$$\sum f_i y_i = 2935$$

$$\sum f_i = 30$$

$$\bar{y} = \text{Arithmetic mean} = \frac{\sum f_i y_i}{\sum f_i}$$

$$= \frac{2935}{30}$$

$$= 97.8333\text{cm}$$

3.2.2 Properties of arithmetic mean

i) The algebraic sum of the deviations of the observations from their mean is zero. i.e.,

$$\Sigma(y_i - \bar{y}) = 0 \qquad \ldots. \ (3.4)$$

It can be proved as:

$$\sum_{i=1}^{n} (Y_i - \bar{Y}) = \sum_{i=1}^{n} Y_i - n\bar{Y}$$

$$= \sum y_i - n\bar{Y}$$

$$= \sum y_i - n \frac{\sum y_i}{n} = 0 \quad \therefore \ \bar{Y} \text{ is constant for any specific values of } Y_i$$

for grouped data

$$\sum f_i\ (y_i - \bar{y}) = \sum_{i=1}^{n} f_i y_i - n\frac{\sum_{i=1}^{n} f_i y_i}{n} \qquad as\ \bar{y} = \frac{\sum_{i=1}^{n} f_i y_i}{n}$$

$$= \sum_{i=1}^{n}\ f_i y_i - \sum_{i=1}^{n}\ f_i y_i$$

$$= 0$$

Numerically, it can be easily seen from the following sample

y_i	2	4	5	6	7	6	Total 30
$y_i - \bar{y}$	3	−1	0	1	2	1	0

$$\bar{Y} = \frac{(2+4+5+6+7+6)}{6}$$

$$= \frac{30}{6} = 5$$

We see that $\sum(y_i - \bar{y}) = 0$

ii) Sometimes, it is desirable to calculate the combined mean of two or more sample means using the individual sample means and their sample sizes. It would be denoted by $\bar{Y}c$. The combined mean is calculated as:

$$\bar{Y}c = \frac{n_1\bar{Y}_1 + n_2\bar{Y}_2 + ... + n_k\bar{Y}_k}{n_1 + n_1 + ... + n_k} \qquad (3.5)$$

$$= \frac{\sum_{i=1}^{k} n_i\bar{Y}_i}{\sum_{i=1}^{k} n_i}$$

e.g., the combined mean for two groups is given by

$$\bar{Y}c = \frac{n_1\bar{Y}_1 + n_2\bar{Y}_2}{n_1 + n_2}$$

If $\bar{Y}_1 = 3$ with $n_1 = 3$ and $\bar{Y}_2 = 4$ with $n_2 = 2$, then $\bar{Y}c$ is given by

$$\bar{Y}_c = \frac{3(3)+2(4)}{3+2}$$

$$= \frac{17}{5}$$

$$= 3.40$$

Similarly, $\bar{Y}_c$ can be calculated for more than two groups.

(iii) The sum of squares of the deviations of the observations f rom their mea minimum i.e.,

$$\sum_{i=1}^{n} (Y - \bar{Y})^2 \text{ is minimum} \qquad (3.6)$$

forungrouped data

$\sum f_i(Y_i - \bar{Y})^2$ is minimum for frequency distribution.

It means that when we take the sum of squares of the deviations from an value a other than $\bar{Y}$, then

$$\sum_{i=1}^{n} (Y_i - a)^2 > \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

It can be proved as:

$$\sum_{i=1}^{n} (Y_i - a)^2 = \sum_{i=1}^{n} (Y_i - \bar{Y} + \bar{Y} - a)^2.$$

$$= \sum_{i=1}^{n} \left[(Y_i - \bar{Y}) + (\bar{Y} - a)\right]^2$$

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

$$= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 + n(\bar{Y} - a)^2 + 0 \qquad \text{as } \sum_{i=1}^{n} (Y_i - \bar{Y}) = 0$$

Now $\sum_{i=1}^{n} (Y_i - a)^2$ is the sum of two terms which are both positive, i. $\sum_{i=1}^{n} (Y_i - \bar{Y})^2$ and $n(Y - a)^2$ are positive being squares. So, $\sum_{i=1}^{n} (Y_i - a)^2$ greater than single term $\sum_{i=1}^{n} (Y_i - \bar{Y})^2$ i.e.,

$$\sum_{i=1}^{n} (Y_i - a)^2 > \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

iv) The mean is affected by the change of origin. If a constant A is added to each of the observations $Y_1, Y_2, \ldots, Y_n$ having mean $\overline{Y}$, then the mean increases by that constant.By adding a to all the observations we have $a+Y_1, a+Y_2, \ldots, a+Y_n$ and the mean would be a+ $\overline{Y}$. For $a = 10$, mean woubd be 10 + $\overline{Y}$, and similarly, if a is subtracted from each Y_i then the mean is $\overline{Y} - a$.

v) The man is affected by the change of scale. If $Y_1, Y_2, \ldots, Y_n$ have mean $\overline{Y}$ then the mean after multiplying each observation by a constant a, is the mean multiplied by that constant. The mean of $aY_i, aY_2, \ldots, aY_n$. i.e.,

$$\overline{Y} = \frac{aY_1 + aY_2 + \ldots + aY_n}{n}$$

$$= \frac{a(Y_1 + Y_2 + \ldots + Y_n)}{n}$$

$$= a\frac{\sum_{i=1}^{n} Y_i}{n}$$

$$= a\overline{Y}$$

If $Y_1, Y_2, \ldots, Y_n$ are multiplied by 10 then the resulting mean would be 10 $\overline{Y}$.

3.2.3 Calculation of A.M. by coding / short-cut method

The arithmetic mean may be calculated by the following formula:.

$$\overline{Y} = a + \frac{\sum_{i=1}^{n} D_i}{n} \qquad (3.7)$$

Where $D_i = (Y_i - a)$ and a is an arbitrary value called provisional mean. The relation (3.7) can be derived as follows:-

$$\overline{Y} = a + \frac{\sum_{i=1}^{n} Y_i}{n}$$

$$= \frac{\sum_{i=1}^{n} (Y - a + a)}{n}$$

$$= a + \frac{\sum_{i=1}^{n} (Y - a)}{n} = a + \frac{\sum_{i=1}^{n} D}{n} \quad \text{where, } Y_i - a = D_i$$

This is for ungrouped data, and similarly, for grouped data

$$\overline{Y} = a + \frac{\sum_{i=1}^{k} f_i D_i}{\sum_{i=1}^{k} f_i} \quad (3.8)$$

If the class intervals are equal then the arithmetic mean may also be calculated as:

$$\overline{Y} = a + \frac{\sum_{i=1}^{k} f_i u_i}{\sum_{i=1}^{k} f_i} \times h \quad (3.9)$$

where $u_i = \frac{Y_i - a}{h}$ and h = common width of the classes.

Example 3.3: Find A.M. for the data of the examples 3.1 and 3.2 by short cut method.

Solution: Taking 90 as an arbitrary origin.

Y_i	87	91	89	88	89	91	87	92	90	98	Total
$D_i = Y_i$-90	-3	1	-1	-2	-1	1	-3	2	0	8	2

$$\overline{Y} = a + \frac{\Sigma D_i}{n}$$

$$= 90 + \frac{2}{10}$$

$$= 90.20 \text{ cm}$$

Arithmetic mean for grouped data of the example 3.2 taking $a = 98$

C.I	f_i	Y_i	$D_i=Y_i$-98	$f_i D_i$
86 – 90	6	88	–10	–60
91 – 95	4	93	–5	–20
96 – 100	10	98	0	0
101 – 105	6	103	5	30
106 – 110	3	108	10	30
111 – 115	1	113	15	15
Total	**30**	--	--	**–5**

$$\overline{Y} = a + \frac{\sum_{i=1}^{k} f_i D_i}{\sum_{i=1}^{k} f_i}$$

$$= 98 + \left(\frac{-5}{30}\right)$$

$$= 98 - 0.1667$$

$$= 97.8333 \text{ cm}$$

3.2.4 Merits of arithmetic mean

i) It is rigidly defined by mathematical formula.
ii) It is easy to calculate.
iii) It is easy to understand.
iv) It is based upon all the values.
v) It is stable statistic in repeated sampling experiments.
vi) Sum of the observations can be found if mean and number of observations are known.

3.2.5 Demerits of arithmetic mean

i) It is very sensitive to any marked departure from the bell shaped distribution and hence is not suitable for skewed distributions.
ii) It gives fallacious and misleading conclusions when there is too much variation in data.
iii) It is greatly affected by extreme values.
iv) It can not be calculated for open-end classes without assuming open ends.

3.2.6 Weighted Mean

Arithmetic mean is used when all the observations are given equal importance but there are certain situations in which the different observations get different weights. In these situations, weighted mean denoted by $\bar{Y}_w$ is preferred. The weighted mean of $Y_1, Y_2, \ldots, Y_n$ with corresponding weights $w_1, w_2, \ldots, w_n$ is calculated as:

$$\bar{Y}_w = \frac{w_1Y_1 + w_2Y_2 + \ldots + w_nY_n}{w_1 + w_2 + \ldots + w_n}$$

$$= \frac{\sum_{i=1}^{n} w_iY_i}{\sum_{i=1}^{n} w_i} = \frac{\sum w_iY_i}{\sum w_i} \qquad (3.10)$$

Example 3.4: The following data is about the percentage kill (Y_i) and the number of insects (w_i) used in a study, the interest is to calculate the mean of the percentage kill.

Y_i	88	85.7	52.1	33.3	12.0
w_i	44	42	24	16	6

Solution: Weighted Mean $= \bar{Y}_w = \dfrac{\sum w_i Y_i}{\sum w_i}$

Now $\qquad \Sigma w_i Y_i = 44(88) + 42(85.7) + 24(52.1) + 16(33.3) + 6(12)$

$= 9326.6$

and $\qquad \Sigma w_i = 44 + 42 + 24 + 16 + 6$

$= 132$

therefore $\qquad \bar{Y}_w = \dfrac{9326.6}{132}$

$= 70.65606\%$

3.3 Geometric Mean

Geometric mean is useful measures of central tendency for positive values. It is appropriate for averaging rates and ratios. It may be appropriately calculated only for ratio scale data.

The geometric mean is defined as the nth root of the product of n positive numbers. If we have n positive values $Y_1, Y_2, \ldots, Y_n$ then geometric mean, denoted by G.M is defined by

$$\text{G.M} = \sqrt[n]{Y_1 \times Y_2 \ldots \times Y_n} \qquad (3.11)$$

$$= (Y_1 \times Y_2 \ldots \times Y_n)^{1/n}$$

Taking log of both sides, we get.

$$\text{Log G.M} = \frac{1}{n}(\log Y_1 + \log Y_2 + \ldots + \log Y_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \log Y_i$$

or $$\text{G.M} = \text{antilog}\left[\frac{1}{n}\sum_{i=1}^{n} \log Y_i\right] \qquad (3.12)$$

This measure is useful when dealing with relative values such as to find the average of percentage changes, independent ratios and index numbers. The formula for grouped data is:

$$G.M = \sqrt[n]{(Y_1)^{f1}(Y_2)^{f2} + ... + (Y_3)^{fk}} \qquad (3.13)$$

Where $n = \sum_{i=1}^{k} f_i$

$$G.M. = [(Y_1)^{f_1} \times (Y_2)^{f_2} \times .... \times (Y_k)^{f_k}]^{\frac{1}{n}}$$

Taking log, we get

$$\log(G.M) = \frac{1}{n}[f_1 \log(Y_1) + f_2 \log(Y_2) + ... + f_k \log(Y_k)]$$

$$= \frac{1}{n}\sum_{i=1}^{k} f_i \log(Y_i)$$

or $$G.M = \text{antilog}\left(\frac{1}{n}\sum_{i=1}^{k} f_i \log(Y_i)\right) \qquad (3.14)$$

Example 3.5: Calculate geometric mean for the following ungrouped data of the percentage changes in the weight of eight animals.

45, 30, 35, 40, 44, 32, 42, 37

Solution: we know that

$$G.M = \text{antilog}\left[\frac{1}{n}\sum_{i=1}^{n} \log Y_i\right]$$

$$\text{Log}(G.M) = \frac{1}{n}\sum_{i=1}^{n} \log Y_i$$

$$= \frac{1}{8}[\log 45 + \log 30 + \log 35 + \log 40 + \log 44 + \log 32 + \log 42 + \log 37].$$

$$= \frac{1}{8}[1.6532 + 1.4771 + 1.5441 + 1.6021 + 1.6434 + 1.5051 + 1.6232 + 1.5682].$$

$$= \frac{12.6164}{8}$$

$$= 1.57705$$

$$\Rightarrow \quad G.M = \text{antilog}(1.57705)$$

$$= 37.7616$$

Example 3.6: Compute G.M by using the basic definition for the observations:
0.5, 10.0, 2.7, 3.48, 4.7

Solution: Geometric mean of five observations is given by

$$\text{G.M.} = (Y_1 \times Y_2 \times Y_3 \times Y_4 \times Y_5)^{\frac{1}{5}}$$

$$= [(0.5)\ (1.0)\ (2.7)\ (3.48)\ (4.71)\ ]^{\frac{1}{5}}$$

$$= 1.8577$$

Example 3.7: The grouped data available on insect growth population for age and corresponding frequencies are given.

Class boundaries	f_i	Y_i	Log Y_i	f_i log Y_i
0 – 4	2	2	0.3010	0.6021
4 – 8	5	6	0.7782	3.8908
8 – 12	7	10	1.0000	7.0000
12 – 16	8	14	1.1461	9.1690
16 – 20	7	18	1.2553	8.7869
20 – 24	4	22	1.3424	5.3697
24 – 28	1	26	1.4150	1.4150
Total	34	--	--	36.2334

Find geometric mean for the above data

Solution: $\therefore \text{G.M} = \text{antilog}\left[\frac{1}{n}\sum_{i=1}^{k} f_i \log(Y_i)\right],$

Where $n = \sum f_i$

To compute G.M. we calculate column 3, 4 & 5 of table

$$\sum f_i \log Y_i = 36.2334$$

$$\sum f_i = 34$$

$$\text{thus G.M.} = \text{antilog}\left[\frac{36.2334}{34}\right]$$

$$= \text{antilog}\ (1.0657)$$

$$= 11.6329$$

Example 3.8: A man gets a rise of 10% in salary at the end of his first year of service and further rise of 20% and 25% at the end of the second and third years respectively.

The rise in each case being calculated on his salary at the beginning of the year. To what annual percentage increase is this equivalent?

Solution: Suppose the initial salary of the man = 100

Increase after first year = 10%

Salary at the end of the year = 100 + 10 = 110

Salary at the end of the second year = 100 + 20 = 120

Salary at the end of the third year = 100 + 25 = 125

$$G.M. = (110 \times 120 \times 25)^{\frac{1}{3}} = 118.16$$

Annual percentage increase = 118.16 – 100 = 18.16%

Example 3.9: The frequency distribution given below has been derived from the use of working origin. If $D = Y - 18$, find Arithmetic Mean and Geometric Mean.

D	-12	-8	-4	0	4	8	12	16
f	2	5	8	18	22	13	8	4

Solution: Here, $D = Y - 18$ or $Y = D + 18$

D	f	Y	fY	$\log Y$	$f \log Y$
-12	2	6	12	0.776815	1.55630
-8	5	10	50	1.00000	5.00000
-4	8	14	112	1.14613	9.16904
0	18	18	324	2.25527	22.59490
4	22	22	484	1.34242	29.53324
8	13	26	338	1.41497	18.9461
12	8	30	240	1.47712	11.81696
16	4	34	136	1.53148	6.12592
Total	80	--	1696	--	104.19097

$$\text{Arithmetic Mean} = \overline{Y} = \frac{\sum fY}{\sum f}$$

$$= \frac{1696}{80}$$

$$= 21.2$$

$$\text{and G.M} = \text{Antilog}\left(\frac{\sum f \log Y}{\sum f}\right)$$

$$= \text{Antilog}\left(\frac{104.19097}{80}\right)$$

$$= 20.06$$

3.3.1 Properties of geometric mean

i) If there are k sets, each with observations $n_1, n_2, \ldots, n_k$ and G_1, G_2, G_k as their geometric means. Then the combined geometric mean G_{com} of total observations is given by

$$G_{com} = \frac{\sum_{i=1}^{k} n_i \log G_i}{\sum_{i=1}^{k} n_i} \qquad (3.15)$$

ii) If there are two sets each consisting of n positive observations $Y_1, Y_2, \ldots, Y_n$ with geometric mean G_1 and $X_1, X_2, \ldots, X_n$, with geometric mean G_2, then the geometric mean G of the ratio $Z=Y/X$ is the ratio of their geometric means i.e.,

$$G = \frac{G_1}{G_2} \qquad (3.16)$$

This can easily be proved as under:

$$G = \left[\frac{Y_1}{X_1} \times \frac{Y_2}{X_2} \times \ldots \times \frac{Y_n}{X_n}\right]^{\frac{1}{n}}$$

$$= \left[\frac{Y_1}{X_1} \times \frac{Y_2}{X_2} \times \ldots \times \frac{Y_n}{X_n}\right]^{\frac{1}{n}}$$

$$= \frac{(Y_1 \times Y_2 \times \ldots \times Y_n)^{\frac{1}{n}}}{(X_1 \times .X_2 \times \ldots \times X_n)^{\frac{1}{n}}}$$

$$= \frac{G_1}{G_2}$$

3.3.2 Merits of geometric mean

i) It is rigidly defined by a mathematical formula.

ii) It is based on all the observations.

iii) It is capable of mathematical development.

iv) It is less affected by the extreme values as compared with the mean.

3.3.3 Demerits of geometric mean

i) It becomes zero if any of the observations is zero.

ii) It is sensible only for positive values and it becomes imaginary for negative values.

3.4 Harmonic mean

The harmonic mean is particularly useful when dealing with the averages of certain types of rates and ratios. The harmonic mean of n values $Y_1, Y_2, \ldots Y_n$ is defined as reciprocal of the arithmetic mean of the reciprocals of the values. The harmonic mean is denoted by H.M and is defined by:

$$\text{H.M} = \text{Reciprocal of } \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right) \quad (3.17)$$

Where $Y_i \neq 0$

$$= \frac{n}{\sum_{i=1}^{n}\frac{1}{Y_i}} \quad (3.18)$$

Harmonic mean for the grouped data is given by

$$\text{H.M} = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k}\frac{f_i}{Y_i}} \quad (3.19)$$

Harmonic mean deals with the rates independent on each other.

Example 3.10: A tractor is running at the rate of 10 Km / hr. during the first 60 Km; at 20 Km/ hr. during second 60 Km; 30 Km/hr. during the third 60 Km; 40 Km/ hr.

during the fourth 60 Km and 50 Km/hr. during the (last) fifth 60 Km. What would be the average speed?

Solution: Harmonic mean of the values shall give the average speed.

$$\text{H.M} = \text{Reciprocal of } \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right)$$

$$= \text{Reciprocal of } \left[\frac{\frac{1}{10}+\frac{1}{20}+\frac{1}{30}+\frac{1}{40}+\frac{1}{50}}{5}\right]$$

$$= \left[\frac{5}{\frac{1}{10}+\frac{1}{20}+\frac{1}{30}+\frac{1}{40}+\frac{1}{50}}\right]$$

$$= \frac{5}{0.22833}$$

$$= 21.89813 \text{ Km/hr}$$

Example 3.11: Find Harmonic mean for grouped data of the Example 3.7

Solution: We know that $\text{H.M} = \dfrac{\Sigma f_i}{\Sigma f_i / Y_i}$

The data is given by

Class boundaries	f_i	Y_i	$\frac{1}{Y_i}$	$f_i \frac{1}{Y_i}$
0 – 4	2	2	0.5000	1.0000
4 – 8	5	6	0.1667	0.8333
8 – 12	7	10	0.1000	0.7000
12 – 16	8	14	0.0714	0.5714
16 – 20	7	18	0.0556	0.3889
20 – 24	4	22	0.0454	0.1816
24 – 28	1	26	0.0385	0.0385
Total	34	--	00	3.7137

$$\sum f_i = 34$$

$$\sum \frac{f_i}{Y_i} = 3.7137$$

$$\therefore \text{H.M} = \frac{34}{3.7137} = 9.1553$$

Example 3.12: Calculate Harmonic Mean and Geometric Mean from the following data:

3, 8, 11, 0, 10

Solution: It is not possible to calculate Geometric Mean and Harmonic Mean as the data involves the value 0 because while calculating Geometric Mean, the multiplication with 0 makes product of the given values zero (0) and while calculating Harmonic Mean, division by zero (0) is undefined.

3.4.1 Properties of Harmonic Mean

i) If there are k sets each with observations $n_1, n_2, ..., n_k$ and $k_1, k_2, ..., k_k$ as their harmonic means. Then the combined harmonic mean H.M_{comb} of all the observations is given by:

$$\text{H.M}_{\text{comb}} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{H_i}} \qquad (3.20)$$

3.4.2 Merits of harmonic mean

i) It is defined by a mathematical formula.

ii) It is based on all the observations.

iii) It is capable of future mathematical development.

3.4.3 Demerits of Harmonic Mean

i) It can not be calculated if any of the observations is zero.

ii) It is not simple to calculate as compared to the arithmetic mean.

iii) It gives less weight to large values and more weight to small values.

3.4.4 General relationship between A.M. , G.M. and H.M.

If $Y_1, Y_2, ..., Y_n$ are n positive observations, then the arithmetic mean, geometric mean and harmonic mean satisfy the following relation.

$$\text{A.M} \geq \text{G.M.} \geq \text{H.M}$$

The three means are equal only if all the observations are identical.

3.5 Median

It is the value which divides an arranged data in order of magnitude into two equal parts. In case of odd number of observations, median is the value of $\left(\frac{n+1}{2}\right)$ th item and in case of even number of observations, median is the mean value of $(n/2)$ th and $\left(\frac{n+2}{2}\right)$ th items of a set of values arranged in ascending or descending order of magnitude i.e., 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.

Example 3.13: Following are the heights (cms) of 5 students measured at the time of registration. Also find median for the data of the example 3.1.

Y_i: 88.03, 94.50, 94.90, 95.05, 84.60

Solution: The ordered observations are:

84.60, 88.03, 94.50, 94.90, 95.05

Here $n = 5$, so

$$\text{Median} = \text{value of } \left(\frac{5+1}{2}\right) \text{th observation}$$

$$= \text{value of } 3^{rd} \text{ observation}$$

$$= 94.50$$

or

$$\text{Median} = y_{\left(\frac{n+1}{2}\right)}$$

$$= y_{\left(\frac{5+1}{2}\right)}$$

$= Y_{(3)}$, the third value in the ordered observations.

$= 94.50$ cms

The data of the example 3.1 is used to calculate median for even n. The ordered observations are:

87, 87, 88, 89, 89, 90, 91, 91, 92, 98.

or Median = value of $\left(\frac{n+1}{2}\right)^{th}$ obs.

$= \text{value of } \left(\frac{10+1}{2}\right)^{th}$ obs.

$= \text{value of } (5.5)^{th}$ obs.

$= 5^{th} \text{ obs} + 0.5\ (6^{th}\ obs - 5^{th}\ obs)$

$= 89 + 0.5\ (90 - 89)$

$= 89 + 0.5\ (1) = 89.5$

or Median $= \frac{1}{2}\left[y_{\left(\frac{n}{2}\right)} + y_{\left(\frac{n}{2}+1\right)}\right]$

$y_{\left(\frac{n}{2}\right)} = y_{\left(\frac{10}{2}\right)} = y_{(5)} = 89$

$y_{\left(\frac{n+2}{2}\right)} = y_{(6)} = 90$

So, median is the mean of $Y_{(5)}$ and $Y_{(6)}$

$$\text{Median} = \frac{89+90}{2} = 89.5\text{cm}$$

For the grouped data (given in ascending order) median is calculated by the relation:

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - c\right) \qquad (2.21)$$

Where l is the lower class boundary of the class containing the median.

h is width of the class containing median.

f is the frequency of the class containing median.

$\frac{n}{2}$ is used to locate the median class i.e., where the $\left(\frac{n}{2}\right)$th observation falls and this is done by looking at the class corresponding to the cumulative frequency in which $\left(\frac{n}{2}\right)$th observation lies.

c is the cumulative frequency of the class preceding to the median class.

Example 3.14: Find the median for the following student height grouped data.

Class boundaries	y	f_i	c.f
85.5 – 90.5	87	6	6
90.5 – 95.5	93	4	10
95.5 – 100.5	98	10	20
100.5 – 105.5	103	6	26
105.5 – 110.5	108	3	29
110.5 – 115.5	113	1	30

Solution: To find median class $\left(\frac{n}{2}\right)$th observation falls is $(\frac{30}{2})$ th observation.

i.e $\frac{n}{2}$th observation $= \frac{30}{2}$th observation

$= 15$ th observation

The 15th observation falls in the class 95.5 – 100.5

So, Median group = 95.5 – 100.5

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

$$= 95.5 + \frac{5}{10}(15 - 10)$$

$=98.0$ cm

In case, when data is discrete but grouped, the median is calculated by using the formal definition of median.

Example 3.15: Discrete grouped data of 26 plants of cotton are taken and the number of bolls per plant observed, the data is grouped as follows:

Number of bolls	0	1	2	4	5	6	7
Number of plants	5	6	3	6	3	2	1

Solution: As n is even so the median is the mean of $Y_{(\frac{n}{2})}$ and $Y_{(\frac{n}{2}+1)}$ *obs.*

$Y_{(\frac{n}{2})} = Y_{(13)}$; $Y_{(\frac{n}{2}+1)} = Y_{(14)}$.

To locate $Y_{(13)}$ and $Y_{(14)}$ we need to make cumulative frequency column:

Number of bolls	Number of plants	c.f
0	5	5
1	6	11
2	3	14
4	6	20
5	3	23
6	2	25
7	1	26

From the c.f column it is clear that $Y_{(13)}$ and $Y_{(14)}$ are the plants which have number of bolls equal to 2. Thus 2 is the median.

3.5.1 Properties of median

i) If a constant a is added to each of the n observations $Y_1, Y_2, \ldots, Y_n$ having median M, then the median of $a + Y_1, a + Y_2, \ldots, a + Y_n$ would be $a + M$. If a is multiplied to each of the n observations, then median of $aY_1, aY_2, \ldots, aY_n$ would be aM.

ii) The sum of absolute deviations of the observations from their median is minimum i.e.,

$$\sum |Y - \text{median}| \text{ is minimum} \quad (3.22)$$

The bars indicate absolute value meaning thereby we are taking all the deviations positive.

iii) For a symmetrical distribution median is equidistant from the first and third quartiles i.e.,

$$Q_3 - \text{Median} = \text{Median} - Q_1 \quad (3.23)$$

Where, Q_1 and Q_3 are first and third quartiles respectively.

3.5.2 Merits of median

i) It is quick to find.

ii) It is not much affected by exceptionally large or small values in the data.

iii) It is suitable for skewed distributions.

3.5.3 Demerits of median

i) It is not rigidly defined.

ii) It is not readily suitable for algebraic development.

iii) It is less stable in repeated sampling experiments than the mean.

iv) It is not based on all the observations.

3.6 Quantiles

Sometimes, our interest is to know the position of an observation relative to the others in a data set. For example, in the grouped student height data of the example 2.1, we may be interested to know the percentage of students having height less than some specified value. The measures used for this purpose are called quantiles or fractiles. These are usually calculated under the following headings:

i) Quartiles and Deciles ii) Percentiles

3.6.1 Quartiles

Quartiles are the values in the order statistic that divide the data into four equal parts. These are the first quartile Q_1, second quartile Q_2 (median) and third quartile Q_3. The first quartile, also known as lower quartile, is the value of order statistic that exceeds ¼ of the observations and less than the remaining ¾ observations. The second quartile is the median and the third quartile, known as upper quartile, is the value in the order statistic that exceeds ¾ of the observations and is less than remaining ¼ observations.

In case of ungrouped data, the quartiles are calculated by splitting the order statistic at the median and calculating the median of the two halves. If n is odd, the median can be included in both halves.

Example 3.16: Find Quartiles for ungrouped data of the example 3.13

Solution: We know that median of data is the mid value of the order statistic. For finding quartiles, we split the order statistic at the median and calculate the median of two halves. Since n is odd, we can include the median in both halves.

The orders statistic is

84.60, 88.03, 94.50, 94.90, 95.05

$$Q_2 = \text{Median} = Y_{\left(\frac{n+1}{2}\right)}$$

$= Y_{(3)}$, the third observation

$= 94.50$

$$Q_1 = \text{Median of the first three value} = Y_{\left(\frac{3+1}{2}\right)}$$

$= Y_{(2)}$, the second observation

$= 88.03$

$$Q_3 = \text{Median of the last three values} = Y_{\left(\frac{3+5}{2}\right)}$$

$= Y_{(4)}$, the fourth observation

$= 94.90$

For the grouped data (in ascending order) the quartiles are calculated as:

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right) \tag{3.24}$$

$$Q_2 = l + \frac{h}{f}\left(\frac{2n}{4} - c\right) \tag{3.25}$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right) \tag{3.26}$$

Where l is the lower class boundry of the class containing the Q1, Q2 or Q3.
h is the width of the class containing the Q1, Q2 or Q3.
f is the frequency of the class containing Q1, Q2 or Q3.
c is the cumulative frequency of the class immediately preceeding to the class containing Q_1, Q_2 or Q_3, $\left[\frac{n}{4}, \frac{2n}{4} \text{ or } \frac{3n}{4}\right]$ are used to locate Q_1, Q_2 or Q_3 group.

Example 3.17: Find quartiles for data of the example 3.2.

Solution:

Class boundaries	Y_i	f_i	c.f
85.5 – 90.5	87	6	6
90.5 – 95.5	93	4	10
95.5 – 100.5	98	10	20
100.5 – 105.5	103	6	26
105.5 – 110.5	108	3	29
110.5 – 115.5	113	1	30
Total	-	**30**	-

To locate the class containing Q_1,

$\frac{n}{4}$ th observation $= \frac{30}{4}$ th observation

$= 7.5$ th observation

7.5 th observation falls in the group 90.5 – 95.5.

So, Q_1 group = 90.5 – 95.5

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

$$= 90.5 + \frac{5}{4}(7.5-6)$$

$$= 92.3750 \text{ cm}$$

for Q_2,

$$\frac{2n}{4} \text{th observation} = \frac{2\times30}{4} \text{th observation}$$

= 15 th observation falls in the group 95.5 – 100.5

So, $$Q_2 = l + \frac{h}{f}\left(2\frac{n}{4} - c\right)$$

$$= 95.5 + \frac{5}{10}(15 - 10)$$

$$= 98 \text{ cm}$$

for Q_3,

$$\frac{3n}{4} \text{ th observation} = \frac{3\times30}{4} \text{th observation}$$

= 22.5 th observation

So, Q_3 group = 100.5 – 105.5

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

$$= 100.5 + \frac{5}{6}(22.5 - 20)$$

$$= 100.5 + 2.0833$$

$$= 102.5833 \text{ cm}$$

3.6.2 Deciles

Deciles are the values in the order statistic that divide the data into ten equal parts. These are denoted by $D_1, D_2, D_3, \ldots, D_9$. D_1 is the value of order statistic that exceeds 1/10 of the observations and less than the remaining 9/10. The fifth decile is the median and D_9, the ninth decile is the value in the order statistic that exceeds 9/10 of the observations and is less than 1/10 remaining observations.

Deciles for Ungrouped Data

To calculate deciles for the ungrouped data the following procedure may be followed.

i) Order the observations.

ii) For the *m*th decile, determine the product $\frac{m.n}{10}$. If $\frac{m.n}{10}$ is not an integer, round it up and find the corresponding ordered value but if $\frac{m.n}{10}$ is an integer, say *k*, then calculate the mean of *k*th and (k+1) th ordered observations.

Example 3.18: The data of the example 2.1 is used to explain the procedure and D_7 and D_3 have been calculated.

i) The ordered observations are
87, 87, 88, 89, 89, 90, 91, 91, 92, 95, 96, 96, 97, 98, 98, 98, 99,99, 100, 100, 101, 101, 102, 103, 105, 105, 106, 107, 107, 112.
Here $n = 30$

ii) To calculate D_7, the (7) (30) / 10 = 21, so, we calculate the mean of 21st and 22nd observations i.e., $D_7 = (101 + 101)/2 = 101$.
To calculate D_3, the (3) (30) / 10 = 9, so, we calculate the mean of 9th and 10th observation i.e., $D_3 = (92 + 95)/2 = 93.5$.

Deciles for grouped data

The *m*th decile for grouped data (in ascending order) is

$$D_m = l + \frac{h}{f}\left(\frac{m.n}{10} - c\right) \qquad (3.27)$$

Like the median, $\frac{m.n}{10}$ is used to locate the *m*th decile group.

l is the lower class boundary of the class containing *m*th decile.
h is the width of the class containing D_m.
f is the frequency of the class containing D_m.
n is the total number of frequencies.
c is the cumulative frequency of the class immediately preceding to the class containing D_m.

Example 3.19: Data of the example 3.17 is used to explain the procedure for grouped data. D_1 and D_7 are calculated

Calculation for D_1

$\frac{1.n}{10}$th observation = $\frac{1\times 30}{10}$th observation

= 3^{rd} observation

So, D_1 group = 85.5 – 90.5

$$\therefore \quad D_1 = l + \frac{h}{f}\left(\frac{1\times n}{10} - c\right)$$

$$= 85.5 + \frac{5}{6}(3 - 0)$$

$$= 88.00 \text{ cm}$$

Calculation for D_7

$\frac{7\times n}{10}$th observation $= \frac{7\times 30}{10} = 21^{st}$ observation

$$D_7 = l + \frac{h}{f}\left(\frac{7\times n}{10} - c\right)$$

$$= 100.5 + \frac{5}{6}(21 - 20)$$

$$= 101.3333 \text{ cm}$$

3.6.3 Percentiles

These are the measures of relative standing of an observation within a data. The *pth* percentile is the value $Y_{(p)}$ in the order statistic such that p percent of the values are less than the value $Y_{(p)}$ and (100-p) percent of the values are greater than $Y_{(p)}$. The 5th percentile is denoted by P_5, the 10th by P_{10} and 95th by P_{95}.

Percentiles for the ungrouped data

To calculate percentiles for the ungrouped data, the following procedure is adopted:

i) Order the observations.

The procedure is explained on the data of the example 2.1 and P_{10} and P_5 have been calculated.

ii) For the *m*th percentile, determine the product $\frac{m.n.}{100}$. If $\frac{m.n.}{100}$ is not an integer, round it up and find the corresponding ordered value and if $\frac{m.n.}{100}$ is an integer, say k, then calculate the mean of the *Kth* and (k+1) *th* ordered observations.

The ordered observations of the example 2.1 are:

87, 87, 88, 89, 89, 90, 91, 91, 92, 95, 96, 96, 97, 98, 98, 98, 99,99, 100, 100, 101, 101, 102, 103, 105, 105, 106, 107, 107, 112.

To calculate P_{10}, the (10) (30) / 100 = 3, so, we calculate the mean of 3rd and 4th observations i.e., $P_{10} = (88 + 89)/2 = 88.5$.

To calculate P_{95}, the (95) (30) / 100 = 28.5, so, 29th observation is our 95th percentile i.e., $P_{95} = 107$.

Percentiles for the grouped data

The m_{th} percentile for grouped data (given in ascending order) is

$$P_m = l + \frac{h}{f}\left(\frac{m.n}{100} - c\right) \qquad (3.28)$$

Like the median, $\frac{m.n}{100}$ is used to locate the mth percentile group.

l is the lower class boundary of the class containing the mth percentile.

h is the width of the class containing P_m.

f is the frequency of the class containing P_m.

n is the total number of frequencies.

c is the the cumulative frequency of the class immediately preceding to the class containing P_m.

The 50th percentile is the median by definition as half of the values in the data are smaller than the median and half are larger than the median.

The 25th and 75th percentiles are the lower and upper quartiles respectively. The quartiles, deciles and percentiles are also called quantiles or fractiles.

Example 3.20: Find P_{10}, P_{25}, P_{50} and P_{95} of grouped data for the example 3.17.

Solution: $\frac{10n}{100}$ th observation = $\frac{10\times 30}{100}$ th observation

= 3rd observation

So, P_{10} group= 85.5 – 90.5

$$P_{10} = l + \frac{h}{f}\left(\frac{10n}{100} - c\right)$$

$$= 85.5 + \frac{5}{6}(3 - 0)$$

$$= 85.5 + 2.5 = 88.00 \text{ cm}$$

$$P_{25} = l + \frac{h}{f}\left(\frac{25n}{100} - c\right)$$

$$= l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

$$= Q_1$$

Similarly, $P_{50} = Q_2$ = Median

and $P_{75} = Q_3$, already calculated under the example 3.17

$\frac{95n}{100}$ th observation $= \frac{95 \times 30}{100}$ th observation

= 28.5 th observation

So, P_{95} group= 105.5 – 110.5

$$P_{95} = l + \frac{h}{f}\left(\frac{95n}{100} - c\right)$$

$$= 105.5 + \frac{5}{3}\ (28.5 - 26)$$

$$= 105.5 + 4.1667$$

= 109.6667 cm

The percentiles and quartiles may be read directly from the graphs of the cumulative frequency function as in chapter 2 where, Q_1 is indicated. The Q_3 may be read corresponding to a relative cumulative frequency of 0.75.

3.7 Mode

Mode is defined as the most frequent value in a data set. In case of ungrouped data, the mode can be found by inspection of the order statistic. For example, five plants having heights in cms. 87, 82, 87, 90, 89. The order statistic for this data would be 82, 87, 87, 89, 90.

87 is the value that comes twice while others are only once. So, by definition 87 is the mode of this data. If data has only one mode, then it is called unimodal. The data may have more than one mode. It may be bimodal (having two modes) or multimodal (having more than two modes). The data is said to have no mode, if every value of the data equal number of times.

The mode for the grouped data (given in ascending order) is calculated by

$$\text{Mode} = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h \qquad (3.29)$$

l is the lower class boundary of the modal class.

f_m is the frequency of the modal class.

f_1 is the frequency associated with the class preceding the modal class.

f_2 is the frequency associated with the class following the modal class.

h is the width of the model class.

model class is the class in which maximum frequency lies.

Example 3.21: Find mode for the data of the example 3.17.

Solution: The maximum frequency is 10 for the class 95.5 – 100.5, so, it is a model class.

$$\therefore \text{ Mode } (\hat{X}) = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

Here $l = 95.5,\ h = 5,\ f_m = 10,\ f_1 = 4,\ f_2 = 6$

$$\text{Mode} = 95.5 + \frac{10 - 4}{(10-4) + (10-6)} \times 5$$

$$= 95.5 + 3.0$$

$$= 98.5 \text{ cm}$$

Example 3.22: The table shows the distribution of the maximum loads is shot tons supported by certain cables produced by a company.

Maximum loads	No. of Cables
9.3 – 9.7	2
9.8 – 10.2	5
10.3 – 10.7	12
10.8 – 11.2	17
10.3 – 11.7	14
11.8 – 12.2	6
12.3 – 12.7	3
12.8 – 13.2	1

Determine its mode.

Solution:

Maximum loads	f	C.B.
9.3 – 9.7	2	9.25 – 9.75
9.8 – 10.2	5	9.75 – 10.25
10.3 – 10.7	12	10.25 – 10.75
10.8 – 11.2	17	10.75 – 11.25
11.3– 11.7	14	11.25 – 11.75
11.8 – 12.2	6	11.75 – 12.25
12.3 – 12.7	3	12.25 – 12.75
12.8 – 13.2	1	12.75 – 13.25

$$\text{Mode } (\hat{X}) = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

$$= 10.75 + \frac{17-12}{(17-12)+(17+14)} \times 5$$

$$= 11.06$$

Example 3.23: Find mode for the data of the example 2.1.

y	87	88	89	90	91	92	95	96	97	98	99	100	101	102	103	105	106	107	112
f	2	1	2	1	2	1	1	2	1	3	2	2	2	1	1	2	1	2	1

Mode = The value which occurs most frequently in the data.

$\therefore$ Mode = 98 cm.

3.7.1 Properties of mode

i) If a constant a is added to each of the n observations $Y_1, Y_2, \ldots, Y_n$ having mode m, then the mode of $a+Y_1, a+Y_2, \ldots, a+Y_n$ would be $a+m$.

ii) If a is multiplied with each of the n observations $Y_1, Y_2, \ldots, Y_n$ having mode m then the mode of $aY_1, aY_2, \ldots, aY_n$ would be am.

3.7.2 Merits of mode

i) It is very quick to find.

ii) It is not affected by extreme values.

3.7.3 Demerits of mode

i) It is not rigidly defined.

ii) It is not capable of further mathematical development easily.

iii) It uses only a few members of the population, so can be misleading in small data sets.

iv) It is an unstable measure like median.

v) There may be more than one values of the mode in a data set.

vi) It may not exist in many cases.

3.7.4 Empirical Relationship Between Mean, Median And Mode

The empirical relationship depends upon the shape of the distribution of the data. The distribution of a data set is called symmetrical if the frequency curve for the data is such that the left of curve to its mean is the mirror image of the portion to the right of mean. Otherwise, the distribution is called skewed. The skew may be to the right or to the left depending upon the shape of curve. The empirical relationship is described as follows:

a) In a single peaked symmetrical distributions mean, median and mode are equal i.e.,

$$\text{Mean} = \text{Median} = \text{Mode} \qquad (3.30)$$

It is indicated in figure 3.1

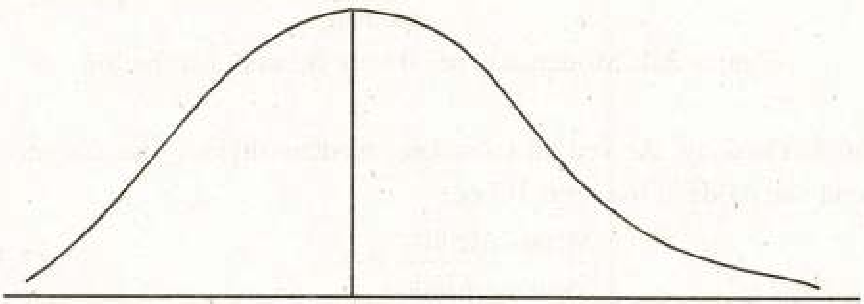

Figure 3.1: single peaked symmetrical distribution.

b) For moderately positively skewed distributions, the following empirical relation holds.

$$\text{Mean} > \text{Median} > \text{Mode} \qquad (3.31)$$

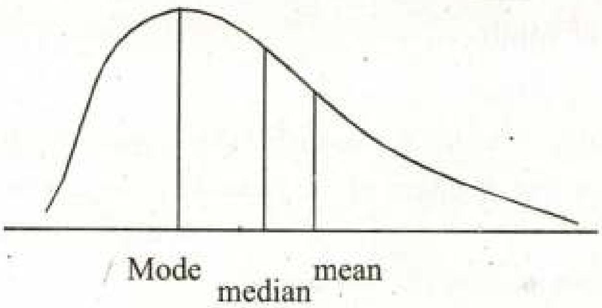

Figure 3.2: Moderately Positively skewed distribution

c) For moderately negatively skewed distributions, the following empirical relation holds.

$$\text{Mean} < \text{Median} < \text{Mode} \qquad (3.32)$$

It is indicated in figure 3.3.

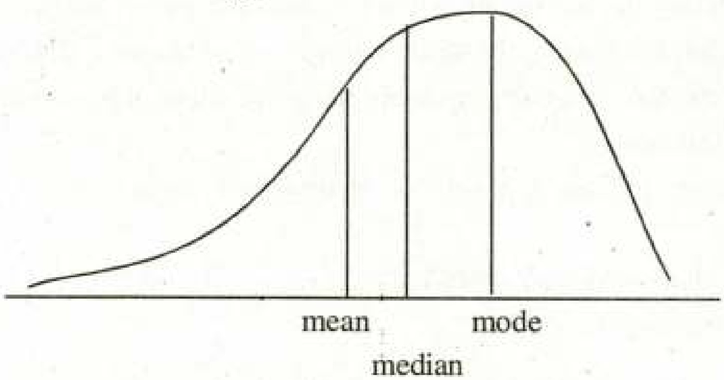

Figure 3.3: Moderately negatively skewed distribution

d) For moderately skewed distributions median divides the distance between mean and mode in the ratio 1:2 i.e.,

$$\frac{\text{Mean - Median}}{\text{Median - Mode}} = \frac{1}{2} \qquad (3.33)$$

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

Example 3.24: If mode = 15 and Median = 12, find mean.

Solution: If Mode = 15, Median = 12, Mean = ?

We know that

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

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

$$= \frac{3(12) - 15}{2} = \frac{36 - 15}{2} = 10.5$$

Example 3.25: Mean and Median of a frequency distribution are 45 and 30 respectively. Find mode of the distribution.

Solution: Mean = 45, Median = 30, Mode = ?

We know that

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

$$= 3(30) - 2(45) = 90 - 90 = 0$$

3.8 Selecting a Suitable Measure of Central Tendency

To select an appropriate measure for a situation, certain factors are taken into account. This includes the type of variable, the purpose of the statistic for which it would be used and the type of distribution.

For the quantitative variables, arithmetic mean is usually appropriate. For the categorical variables, the median and mode are appropriate depending upon the type of categories. For example, if we consider the eye colour, then mode is appropriate but if the categories are income groups, then the appropriate measure is median.

The type of the distribution is an important aspect to evaluate which statistic is appropriate. If the distribution is symmetrical, all the measures i.e., mean, median and mode being equal are equally good. In the skewed distributions median is preferred as it is not affected by the extreme observations. Medians are also preferred to the means when the sample constitutes only small part of the population. Geometric and harmonic means are useful for averaging rates and ratios.

Exercise 3

Ans on Page 250

3.1 i) Define arithmetic mean, geometric mean and harmonic mean. Explain the situations when each of them is used perfectly?

ii) The relation between arithmetic mean (A.M), geometric mean (G.M) and harmonic mean (H.M) is

$$A.M \geq G.M. \geq H.M.$$

Under what situation these are equal.

3.2 Find the geometric mean of 50, 67, 39, 40, 36, 60, 54, 43.

3.3 A man traveling 100 kilometers has 5 stages at equal intervals. The speed of the man in the various stages was observed to be 10, 16, 20, 14, 15 kilometers per hour.

Find the average speed at which the man travels.

3.4 Calculate mean, median and mode for table 2.5

3.5 Calculate mean, median and mode for the grouped data of table 2.7

3.6 Calculate the following for the data of exercise 2.8 (b)

i) Calculate an estimate for the mean leaf length.

ii) Construct a cumulative frequency table and use it to estimate the sample median.

3.7 What do you understand by weighted mean? In what circumstances is it preferred to ordinary mean and why?

3.8 Define the mode of a frequency distribution. How does it compare with other types of averages?

3.9 i) Write down the empirical relation between mean, median and mode for unimodal distribution of moderate asymmetry. Illustrate graphically the relative positions of the mean, median and mode for frequency curves which are skewed to the right and to the left.

ii) For a certain frequency distribution, with the mean and median 45 and 36 respectively, find the mode approximately using the empirical relation between the three.

3.10 Bilal gets a rise of 10% in salary at the end of his first year of service and further rise of 20% and 25% at the end of the second and third year respectively. The rise in each case being calculated on his salary at the beginning of the year. To what annual percentage increase in this equivalent.

3.11 Find the Mean for the following distribution.

Classes	0–10	10–40	40–90	90–100	100–105	105–120	120–140
f	40	110	150	200	120	30	20

3.12 The frequency distribution given below has been derived from the use of working origin. If $D = X - 18$, find arithmetic mean and Geometric mean.

D	−12	−8	−4	0	4	8	12	16
f	2	5	8	18	22	13	8	4

3.13 The reciprocals of 11 values of x are given below:

0.0500, 0.0454, 0.0400, 0.0333, 0.0285, 0.0232,
0.0213, 0.02000, 0.0182, 0.0151, 0.0143.

Calculate Harmonic and Arithmetic Mean of the data.

3.14 Reciprocals of x are given below:

0.0267, 0.0235, 0.0211, 0.0191, 0.0174, 0.0160, 0.0148

Calculate Harmonic Mean of the data.

3.15 Three cities A, B, C are equidistant from each other. Fatima travels from A to B at the speed of 30 miles per hour by car. From B to C at speed of 50 miles per hour. Determine her average speed for the entire trip.

3.16 Harmonic Mean and Geometric Mean of two numbers are 3.2 and 4 respectively. Find their Arithmetic Mean and both the numbers as well.

3.17 The Arithmetic Mean and Geometric Mean of three numbers are 34 and 18 respectively. Find all the three numbers, when the Geometric Mean of the first two numbers is 9.

3.18 Find out

i) The average rate of motion in the case of a person who rides the first mile at the rate of 10 miles per hour the next mile at the rate of 8 miles per hour ar the third at the rate of 6 miles per hour.

ii) Increase in population which in the first decade has increased 20% in the next 25% and in the third 4%.

3.19 The given table shows the distribution of the maximum loads in short tons supported by certain cables produced by a company. Determine Mean, Median and Mode.

Maximums Loads	No. of cables
9.3 – 9.7	2
9.8 –10.2	5
10.3 – 10.7	12
10.8 – 11.2	17
11.3 – 11.7	14
11.8 – 12.2	6
12.3 – 12.7	3
12.8 – 13.2	1

3.20 Compute Mean, Median, Mode, 6^{th} Decile, and 74^{th} percentile for the data given in the table:

Classes	Frequency
0.7312 – 0.7313	10
0.7314 – 0.7315	15
0.7316 – 0.7317	20
0.7318 – 0.7319	25
0.7320 – 0.7321	30
0.7322 – 0.7323	8
0.7324 – 0.7325	2

3.21 Find the value Q_3, D_5, P_5 and mode for the following data:

Groups	Frequency	Groups	Frequency
0 – 4.9	3	25–29.9	13
5 – 9.9	4	30–34.9	13
10 – 14.9	9	35–39.9	5
15 – 19.9	11	40–44.9	2
20 – 24.9	15	45–49.9	2

3.22 If for any frequency distribution the Mean is 45 and the Median is 30. Find Mode approximately, using formula connecting the three.

3.23 A bus traveling 200 miles has ten stages at equal intervals. The speed of the bus in the various stages was observed to be 10, 15, 20, 75, 20, 30, 40, 50, 30, 40 miles per hour. Find the average speed at which the bus has traveled.

3.24 The following data has been obtained from a frequency distribution of a continuous variable x after making the substitution $u = \frac{x-136.5}{6}$.

U	−4	−3	−2	−1	0	1	2	3
f	2	5	8	18	22	13	8	4

Find Harmonic Mean.

3.25 Salman obtained the following marks in a certain examination. Find the weighted mean if weights 4, 3, 3, 2 and 2 respectively are allotted to the subjects.

English	Urdu	Math	Stat	Physics
73	82	80	57	62

3.26 Calculate weighted mean for the following items:

Items	Expenditures	Weights
Food	290	7.5
Rent	54	2
Clothing	98	1.5
Fuel and light	75	1.0
Other Items	75	0.5

3.27 For a certain distribution, if $\Sigma (x - 15) = 5$, $\Sigma (x - 18) = 0$, $\Sigma(x - 211) = -21$ What is the value of A.M and why?

3.28 Arithmetic Mean of 15 values is 20 and by adding 3 more values, the mean remains 20. Find the new three values if ratio is $a:b:c$ ::3:2:1.

3.29 State the following as true or false.

i) In a symmetrical distribution, mean, median and mode are equal.

ii) The algebraic sum of the deviations for a set of observations from their mean is always zero.

iii) Median is affected by extreme observations.

iv) Mode is not affected by extreme observations.

v) Frequency polygon is an increasing function.

vi) For highly skewed distributions, median is preferred over mean.

vii) Mean of a data set remains unchanged if a constant is added in each observation.

3.30 What type of averages would you prefer to average the following:

i) Marks obtained in an examination.

ii) Growth rate of population of different cities.

iii) Height of students.

iv) Size of agricultural holdings.

v) Increase in salaries.

vi) IQ level of students in a class.

3.31 Fill in the blanks

i) An average obtained to represent a data is called __________.

ii) A good average should not be effected by __________ values.

iii) Sum of deviations from mean is always __________.

iv) Median is a value that divides an ordered data into __________ parts.

v) For estimating an average rate of change of population __________ is a better average.

vi) The mean and median of two values is always ______.

vii) In qualitative data, the most suitable average is __________.

viii) A distribution having two modes is called __________ distribution.

ix) In symmetrical distribution, the three averages mean, median and mode are ______.

x) If extreme large or small values are changed, values of __________ are not effected.

3.32 Write T for true and F for false against each statement.

i) The sum of deviations of the values from mean is minimum.

ii) Geometric mean is possible only for negative values.

iii) The median divides the data into two halves.

iv) The third quartile is the median.

v) A distribution having only one mode is called uni-modal distribution.

vi) Arithmetic mean depends on all the values of the data.

vii) Mean, Median and Mode in a symmetrical distribution are not equal.

viii) Harmonic Mean can be calculated if any value is zero.

ix) Median is not affected by extreme values.

x) Geometric Mean cannot be calculated if any value is negative.