

UNIT 3: Measures of central tendency/Average *

Definition and purpose of averages

Definition: Statistics provides tools to reduce each group of values in to a single number or figure representing each group. These representative values are called averages (the measures of central tendency). In other words, they are measures, which condense a huge and mass of numerical data in to a single value. Its value always lies between the minimum and maximum values or it has a tendency to be somewhere at the center. In general, the measures of central tendency are divided in to two:

1. Mathematical Measures of Central Tendency
2. Positional Measures of Central Tendency

Purpose of computing averages

The main objectives of calculating average are:

1. It provides one single value. An average reduces the bulk mass of data in to a single representative value, which enables to grasp the salient feature of data.
2. It facilitates comparison. Comparison of two sets of huge raw data can be possible by working out their averages.
3. It facilitates statistical data inference. An inference, which is derived on the values calculated from the sample, is called statistical inference. It is helpful in knowing the parameters of population.
4. It aids in decision-making. The management is often interested in knowing normal output of a plant, representative sales, overall productivity index, price index, etc and takes decision on future course of action. All these are the connotations of an average.

To sum up, the averages are very much useful in:

- Describing the distribution in concise manner.
- Comparative study of different distributions.
- Computing various other statistical measures such as dispersion, skew ness and other basic characteristics of mass of data.

Note: There are different ways of computing averages depending on the characteristic of the quantities, depending on the significance/ weight or importance of the quantities being averaged. The sort of average chosen as a representative of the set of data depends on the particular characteristic we want to describe.

Requisites of a good average

An average should be:

- ◆ Vigorously (rigidly) defined.
- ◆ Easy to understand and calculate.
- ◆ Based on all values of the given data.
- ◆ Suitable for further mathematical treatment.
- ◆ Affected as little as possible by fluctuations of sampling.
- ◆ Not be affected much by extreme observations. The extreme values should not pull up or pull down the value of the average.

Summation notation and its properties

The symbol “ $\sum_1^n x_i$ ” is read as the sum of x sub i where i vary from 1 to n written as =

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n \text{ and } i \text{ and } n \text{ are lower and upper limit of summation and}$$

x_i the i^{th} value of the variable x .

Properties of summation

1. Summation of a constant

Case 1: If lower limit equal to 1, $\sum_{i=1}^n k = nk$

Example: Suppose $k = 6$ and upper limit $n = 4$, then $\sum_{i=1}^n k = \sum_{i=1}^4 6 = 4 \times 6 = 24$

Case 2: If lower limit is different from 1 or lower limit is greater than 1,

$$\sum_{i=m}^n k = [(n - m) + 1]k \text{ for } m < n$$

Example: Suppose $k = 8$ and upper limit $n = 6$ & lower limit $k = 4$ then

$$\sum_{i=m}^n k = [(n-m) + 1]k$$

$$\sum_{i=4}^6 8 = [(6-4) + 1]8$$

$$8 + 8 + 8 = 3(8) \text{ which gives } 24 = 24$$

2. The summation of sums and of differences'

$$\text{i. } \sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i \quad \text{ii. } \sum_{i=1}^n (x_i - y_i) = \sum_{i=1}^n x_i - \sum_{i=1}^n y_i$$

Example: Suppose $x_1 = 1, x_2 = 3, x_3 = 4, y_1 = 2, y_2 = 5, y_3 = 3$

$$\text{Then } \sum_{i=1}^3 (x_i + y_i) = \sum_{i=1}^3 x_i + \sum_{i=1}^3 y_i$$

$$(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) = (x_1 + x_2 + x_3) + (y_1 + y_2 + y_3)$$

$$(1 + 2) + (3 + 5) + (4 + 3) = (1 + 3 + 4) + (2 + 5 + 3)$$

$$3 + 8 + 7 = 8 + 10$$

$$18 = 18$$

To show $\sum_{i=1}^3 (x_i - y_i) = \sum_{i=1}^3 x_i - \sum_{i=1}^3 y_i$ is left for the student

3. The sum of constant multiple of summand

$$\sum_{i=1}^n kx_i = k \sum_{i=1}^n x_i$$

Example: Suppose $k = 7$ and $x_1 = 2, x_2 = 5, x_3 = 4$ Then $\sum_{i=1}^3 kx_i = k \sum_{i=1}^3 x_i$

$$kx_1 + kx_2 + kx_3 = k(x_1 + x_2 + x_3)$$

$$7(2) + 7(5) + 7(4) = 7(2 + 5 + 4)$$

$$14 + 35 + 28 = 7(11)$$

$$77 = 77$$

4. Sum of summations

$$\sum_{i=1}^k x_i + \sum_{i=k+1}^n x_i = \sum_{i=1}^n x_i \text{ for any } k < n$$

Example: Suppose $x_1 = 2$, $x_2 = 4$, $x_3 = 6$, $x_4 = 3$, $x_5 = 2$, $x_6 = 4$ let $k = 3$ then

$$\sum_{i=1}^k x_i + \sum_{i=k+1}^n x_i = \sum_{i=1}^n x_i$$

$$\sum_{i=1}^3 x_i + \sum_{i=4}^6 x_i = \sum_{i=1}^6 x_i$$

$$(x_1 + x_2 + x_3) + (x_4 + x_5 + x_6) = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

$$(2 + 4 + 6) + (3 + 2 + 4) = (2 + 4 + 6 + 3 + 2 + 4)$$

$$21 = 21$$

Exercise: Let $\sum_{i=3}^6 x_i = 10$, $\sum_{i=3}^6 x_i^2 = 148$, $x_1 = 3$, $x_2 = 2$ then Find **i.** $\sum_{i=1}^6 x_i$

$$\text{ii. } \sum_{i=1}^6 x_i^2 \quad \text{iii. } \sum_{i=1}^6 x_i(x_i - 2) \quad \text{iv. } \sum_{i=1}^6 (2x_i - 3)^2 \quad \text{v. } \sum_{i=1}^2 (ix_i + 4)$$

a. Mathematical Measures of Central Tendency

i. The Arithmetic Mean (AM)

Definition: The arithmetic mean is the sum of the values in a group divided by the number of items in that group. Let $x_1, x_2, \dots, x_n$ be the n values of a variable x , then

their arithmetic mean is defined by: $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum x}{n}$ Where

$\sum x$: sum of all observations and n : total number of observations.

Computation of Arithmetic Mean for Ungrouped and Grouped Data

For ungrouped data: Direct method: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ and Short cut method:

$\bar{x} = A + \frac{\sum d}{n}$ Where n – number of items, A = assumed mean , $\sum d$ = sum of deviations i.e. $(x_i - A)$.

For Discrete Series: Direct method: $\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{n} = \frac{\sum f_x}{n}$, Short cut method:

$\bar{x} = A \pm \frac{\sum fd}{n}$, Where f - frequency, d - deviation of items from assumed mean $(x_i - A)$ assumed mean, n - number of observations

For continuous series: Direct method $\bar{x} = \frac{\sum_{i=1}^k f_i cm_i}{n}$

Step deviation method $\bar{x} = A + \frac{\sum_{i=1}^k f_i d_i}{n}$. Where, f_i – frequency

n – number of observation cm_i – class mark A – assumed mean

c – class width d_i – derivation of class marks from assumed mean $(cm_i - A)$

Example 1: Find the arithmetic mean of the numbers 18, 12, 15, 13, 19, 10, 17, 16, 14, and 24 choosing a guessed mean 17.

Sol: Using direct Method: - The deviations of the numbers from the guessed 17 are 1, -5, -2, -4, 2, -7, 0, -1, -3, 7 and the sum of the deviations is -12. Then

$$\bar{x} = A \pm \frac{\sum fd}{n} = 17 + -12/10 = \underline{15.8}$$

Example 2: Suppose the frequency distribution of the test scores of 20 students in a class is given below. Find the mean test score choosing a guessed mean 58.

Test score(x)	34	51	58	63	77	95
f	3	2	5	4	4	2

Sol: The deviations of the numbers from the guessed mean 58 are -24, -7, 0, 5, 19 and 37

with frequencies 3,2,5,4,4 and 2 resp. then $\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{n} = \frac{\sum f_x}{n}$

$$= 58 + (3(-24) + 2(-7) + 5(0) + 4(5) + 4(19) + 2(37))/20 = 58 + 4.2 = \underline{\underline{62.2}}$$

Example3: Suppose the weight in kg os 50 students in a class .is given by the distribution below. Find the mean weight of the students choosing the guessed mean 58.

Wt. in kg	46---50	51---55	56---60	61---65	66---70	71---75
f	4	11	6	14	8	7

Sol: Direct method $\bar{x} = \frac{\sum_{i=1}^k f_i c_{m_i}}{n} = \frac{4x48+11x53+6x58+14x63+8x68+7x73}{50} = \underline{\underline{61.2kg}}$

Step deviation method $\bar{x} = A + \frac{\sum_{i=1}^k f_i d_i}{n}$

$$= 58 + \frac{48x(-10)+53x(-5)+58(0)+63x5+68x10+73x15}{50}$$

$$= \underline{\underline{61.2kg}}$$

Example4: Find the arithmetic mean for the following data by

- i. direct method ii. Short cut method

23.4, 15.6, 22.1, 20.0, 26.7, 31.4, 18.9, 22.3

Sol.: i. $\sum_{i=1}^8 x_i = 180.4$, $n = 8$ $\bar{x} = \frac{\sum_{i=1}^8 x_i}{n} = \frac{180.4}{8} = \underline{\underline{22.55}}$

- ii. Let $A = 22$ then d_i : 1.4, -6.4, 0.1, -2, 4.7, 9.4, -3.1, 0.3

$$\sum_{i=1}^8 d_i = 4.4, n = 8 \quad \bar{x} = A + \frac{\sum_{i=1}^8 d_i}{n} = 22 + \frac{4.4}{8} = 22 + 0.55 = \underline{\underline{22.55}}$$

Example5: Given data of a test score of 50 students in a class. Calculate the arithmetic mean by i. direct method ii. short cut method.

No. of Students	20	30	40	50	60	70
Marks	8	10	16	8	5	3

Sol. :

Marks x_i	f_i	$f x$	$D_i = (x - 40)$ where $A = 40$	$f d$
20	8	160	-20	-160
30	10	300	-10	-100
40	16	640	0	0
50	8	400	10	80
60	5	300	20	100
70	3	210	30	90
	50	2010		10

i. $\bar{x} = \frac{\sum fx}{n} = \frac{2010}{50} = \underline{\underline{40.20}}$

ii. $\bar{x} = A + \frac{\sum fd}{n} = 40 + \frac{10}{50} = \underline{\underline{40.20}}$

—

Example 6: In a survey, the number of persons at different ages is found as follows:

Age in Year	5 - 15	15 - 25	25 - 35	35 - 45	45 - 55	55 - 65
No. of Persons	8	10	14	20	16	12

Sol.:

Classes	f	cmi	fcm	d = cm – A where A = 30	df	d1 = d/c	fd1
5 - 15	8	10	80	-20	-160	-2	-16
15 - 25	10	20	200	-10	-100	-1	-10
25 - 35	14	30	420	0	0	0	0
35 - 45	20	40	800	10	200	1	20
45 - 55	16	50	800	20	320	2	32
55 - 65	12	60	720	30	360	3	36
	80		3020		620		62

$$\text{i. } \bar{x} = \frac{\sum fcm}{n} = \frac{3020}{80} = \underline{\underline{37.75}}$$

$$\text{ii. } \bar{x} = A + \frac{\sum fd}{n} = 30 + \frac{620}{80} = 30 + 7.75 = \underline{\underline{37.75}}$$

$$\text{iii. } \bar{x} = A + \left(\frac{\sum fd^1}{n} \right) \times c = 30 + \left(\frac{62}{80} \right) 10 = 30 + (0.775) 10 = \underline{\underline{37.75}}$$

Exercise: Find the arithmetic mean of the following data.

i. 53, 54, 52, 32, 30, 60, 47, 46, 35, 29

ii.

Height (inches)	64	65	66	67	68	69	70	71	72	73
No of Students	4	9	12	18	20	12	10	9	4	2

iii.

Classes	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35
Frequencies	5	6	7	7	5

Properties of the Arithmetic Mean

The mathematical properties of an arithmetic mean are discussed below:

1. The arithmetic mean can be computed for any finite set of numerical data, so it always exist and is unique (ie the set of numerical data has exactly one mean).
2. The product of arithmetic mean $(\bar{x})$ and the number of observations (n) on which the mean based is equal to the sum of all given values i.e. $\sum_1^n x_i = n\bar{x}$
3. If each of the observation $x_1, x_2, x_3, \dots, x_n$ with mean $(\bar{x})$ is changed by a constant c (adding or subtracting c) then the new mean is changed by that constant. Ie. Suppose the mean of the values $x_1, x_2, \dots, x_n$ be $\bar{x}_0$. Then
i.if a constant k is added to each x_i , then the new mean $\bar{x}_n$ will be $\bar{x}_0 + k$.

Proof: Arithmetic mean of $x_1 + k, x_2 + k, \dots, x_n + k$ is

$$A.M = \frac{\sum_{i=1}^n (x_i + k)}{n} = \frac{(x_1 + k) + (x_2 + k) + \dots + (x_n + k)}{n}$$

$$A.M = \frac{(x_1 + x_2 + \dots + x_n) + (k + k + \dots + k)}{n}$$

$$A.M = \frac{x_1 + x_2 + \dots + x_n}{n} + \frac{nk}{n}$$

$$A.M = \underline{\underline{\bar{x}_0 + k}}$$

4. If each of the observation $x_1, x_2, x_3, \dots, x_n$ with mean $(\bar{x})$ is multiplied by a constant c then the mean of $c x_1, c x_2, \dots, c x_n$ is given by $c\bar{x}$.
5. The algebraic sum of the deviations of the given values $x_1, x_2, x_3, \dots, x_n$ from the arithmetic mean $\bar{x}$ is equal to zero. Mathematically: $\sum (x_i - \bar{x}) = 0$ for ungrouped data, $\sum f_i (x_i - \bar{x}) = 0$... For grouped data.
Because of this property, the arithmetic mean may be characterized as a point of balance.
6. The sum of squares of deviations is minimum when deviations are taken from the arithmetic mean. i.e. $\sum (x_i - \bar{x})^2 \leq \sum (x_i - A)^2$ for ungrouped data, and

$\sum f_i (x_i - \bar{x})^2 \leq \sum f_i (x_i - A)^2$ for grouped data. Where, A is different from mean.

Example: Given data 12, 10, 8, 6, 16, 7, 11. If each item is multiplied by $\frac{4}{5}$ and 8 is added, what will be the new mean?

$$\bar{x}_0 = \frac{\sum_{i=1}^7 x_i}{7} = \frac{70}{7} = \underline{\underline{10}} \quad \text{New mean}$$

$$\bar{x}_n = \frac{4}{5} \bar{x}_0 + 8$$

$$\bar{x}_n = \frac{4}{5} (10) + 8 = \underline{\underline{16}}$$

Exercise: Given data 3, 8, 9, 4, 7, 5, 10, 11, 6 if each item is multiplied by 2 and 6 is added, then

- i. The new mean will be _____
- ii. $\sum (x_i - \bar{x}) =$ _____

Combined Arithmetic Mean

Let $\bar{x}_1$ and $\bar{x}_2$ be the AM of two groups, having N_1 and N_2 observations respectively. If we combine the two groups in to a single group, then the arithmetic mean of the combined group will be $\bar{x}_c = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N_1 + N_2}$

For n number of groups, $\bar{x}_c = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2 + \dots + N_n \bar{x}_n}{N_1 + N_2 + \dots + N_n} = \frac{\sum_{i=1}^n N_i \bar{x}_i}{\sum_{i=1}^n N_i}$

Example: The mean height of 25 male and 20 female is 161.0cm and 155.6cm. What will be the combined mean height?

$$\bar{x}_m = 161.0\text{cm}, \quad \bar{x}_f = 155.6\text{cm}, \quad N_M = 25, \quad N_F = 20$$

$$\bar{x}_c = \frac{\bar{x}_m N_m + \bar{x}_F N_F}{N_m + N_F}$$

$$\bar{x}_c = \frac{161.0(25) + 155.6(20)}{25 + 20} = \frac{7137}{45} = \underline{\underline{158.60cm}}$$

Exercise: In a factory, 120 workers get an average wage of birr 30 a day, 160 workers get Birr 50 a day, 80 workers get Birr 60 a day and 40 workers get birr 80 a day. Find
i. the average of averages. ii. the general average.

Correcting the Arithmetic Mean

Suppose the mean of n numbers $x_1, x_2, x_3, \dots, x_n$ is wrongly computed due to some reasons (such as miss reading) one or more of the numbers then the wrong mean can be

corrected as follows: $\bar{x}_c = \frac{n\bar{x}_w - \sum x_w + \sum x_c}{n}$ ie

$$\text{Correct Mean} = \frac{\text{Wrong Sum} - \text{Wrong Entry} + \text{Correct Entry}}{\text{Total Number of Observations}}$$

Examples: 1. The mean score of 50 students in an exam. Was found to be 74. But later on it was discovered that the scores of two students was misread 64 and 32 as 94 and 82. What was the correct mean.

$$\begin{aligned} \text{Sol: } \bar{x}_c &= \frac{n\bar{x}_w - \sum x_w + \sum x_c}{n} & \text{Given: } \bar{x}_{wr} &= 71, \sum x_{wr} = 94 + 82 = 176 \\ &= 71 + \frac{96-176}{50} \\ &= 71 + \frac{-80}{50} & \sum x_c &= 64 + 72 = 96, n = 50 \\ &= \underline{\underline{69.4}} \end{aligned}$$

2. The average mark of 100 students was found to be 40 but later it is discovered that a score of 33 was misread as 83. Find the correct average corresponding to the correct sum.

$$\text{Sol: } \left. \begin{array}{l} \bar{x} = 40 \\ N = 100 \end{array} \right\} \rightarrow \sum x_i = \bar{x} N = 40(100) = 4000 \text{ wrong sum}$$

Wrong Entry = 83, Correct Entry = 33

$$\text{Correct Mean} = \frac{4000 - 83 + 33}{100} = \frac{3950}{100} = \underline{\underline{39.5}}$$

3. The average of age of a class having 35 pupils is 14 years. When the age of the class teacher is added to the sum of the ages of the pupils, the average rises by 0.5 year. What must be the age of the teacher?

Sol:

$$\left. \begin{array}{l} \bar{x} = 14 \\ N = 35 \end{array} \right\} \rightarrow \sum x_i = 14(35) = 490 \quad \dots \text{Sum of ages of the pupils}$$

$$\left. \begin{array}{l} \bar{x} = 14.5 \\ N = 36 \end{array} \right\} \rightarrow \sum x_i = 14.5(36) = 522 \quad \dots \text{Sum of ages of the pupils and the teacher}$$

$$\therefore \text{Age of the teacher is } 522 - 490 = \underline{\underline{32 \text{ years}}}.$$

4. Goals scored by a football team in successive matches are 5, 2, 4, 3, 6, 0, 4 and 6. What is the number of goals the team must score in the next match in order that the average comes to 4 goals per match?

Sol:

$$\text{Total goal scored in 8 matches} = 5 + 2 + 4 + 3 + 6 + 0 + 4 + 6 = 30$$

$$\text{Total goal scored in 9 matches} = \bar{x} \cdot N = 4 \times 9 = 36$$

$$\text{Hence the goals required in the 9}^{\text{th}} \text{ match to bring the average 4} = 36 - 30 = \underline{\underline{6}}$$

Exercise: 1. The mean of 200 items is 50. Later on it is discovered that two items were wrongly taken as 92 and 8 instead of 192 and 88. Find out the correct mean.

2. The average rainfall for a week, excluding Sundays, was 10cm. Due to heavy rainfall on Sunday, the average rainfall for the week rose to 15cm. How much rain fall was there on Sunday?

ii. Weighted Arithmetic Mean

An item or value may be relatively more important or less important than other items. This relative importance is technically known as weight. In case where the relative importance of the different items is not the same we compute weighted arithmetic mean.

If $w_1, w_2, \dots, w_n$ are weights attached to the values $x_1, x_2, \dots, x_n$ respectively, then the

6. weighted AM is defined as
$$\bar{x}_w = \frac{x_1 w_1 + x_2 w_2 + \dots + x_n w_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum wx}{\sum w}$$

Note: If the weights are equal then the quantities reduces to that of the ordinary Arithmetic Mean

Example: If a student took five courses whose credit hours are 4, 4, 3, 2, 2 and at the end of the semester the student achieves letter grades A, A, B, C, D respectively in the exam. What is the GPA of the student?

Sol. Points for the letter grades: 4 for A, 3 for B, 2 for C, 1 for D, 0 for F and the quantities to be averaged are $x_1 = 4, x_2 = 4, x_3 = 3, x_4 = 2$ and $x_5 = 1$. Therefore the GPA is the weighted mean given as GPA =

$$\bar{x}_w = \frac{x_1 w_1 + x_2 w_2 + \dots + x_n w_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum wx}{\sum w}$$

$$= 47/15 = \underline{\underline{3.13}}$$

Exercise : What is the GPA of a student if his grades were D, C, B, A, A in the courses
==
respectively?

Example: An auto ride costs Birr 5 for the first km, Birr 4 for the next 3kms and Birr 9 for each of the subsequent kms. Find the average cost per km for 10 kms.

Rate (Birr)	Distance (km) w	xw
5.00	1	5.00
4.00	3	12.00
9.00	6	54.00
	10	71.00

$$\bar{x}_w = \frac{\sum xw}{\sum w} = \frac{71.00}{10} = \underline{\underline{7.10\text{Birr}}}$$

Exercise: Given data

Designation	Class I Officer	Class II Officer	Subordinate Staff	Clerical Staff	Lower Staff
Monthly Salary (in birr) x	1500	800	500	250	100
Strength of Cadre. w	10	20	70	100	150

7.

Calculate i. the simple arithmetic mean.

ii. the weighted arithmetic mean.

iii. Geometric mean (GM)

Definition: Geometric Mean is defined as the n^{th} root of the product of n items or values of a series. If there are two items, we take square root; if there are three items, the cube root and so on. Symbolically, let $x_1, x_2, \dots, x_n$ be the n values of a variable x , then their

G.M is defined as $G.M = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$

If the number of observation is more than three or more, the computation of the n^{th} root is very tedious. To simplify computation, the logarithms are used. In terms of log.

$$\begin{aligned}
 \text{Log } G.M &= \text{Log } \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n} \\
 &= \text{Log } (x_1 \cdot x_2 \dots x_n)^{\frac{1}{n}} \\
 &= \frac{1}{n} \cdot \text{Log } (x_1 \cdot x_2 \dots x_n) \\
 &= \frac{1}{n} [\text{Log } x_1 + \text{Log } x_2 + \dots + \text{Log } x_n] \\
 &= \frac{1}{n} \cdot \sum_{i=1}^n \text{Log } x_i
 \end{aligned}$$

$$\text{Anti log } (\text{Log } GM) = \text{Anti log } \left[\frac{1}{n} \cdot \sum_{i=1}^n \text{Log } x_i \right]$$

$$GM = \text{Anti log } \left[\frac{1}{n} \sum_{i=1}^n \text{Log } x_i \right]$$

Computation GM for Ungrouped and Grouped Data

For ungrouped data: $G.M = \text{Anti log } \left[\frac{1}{n} \sum_{i=1}^n \text{Log } x_i \right]$

For grouped data: $G.M = \text{Anti log } \left[\frac{1}{n} \sum_{i=1}^n f_i \cdot \text{Log } x_i \right]$

Example: Compute GM of the following data.

i. **x :** 23 27 54 35 50

ii. **x:** 10 16 22 28 34

f: 5 4 3 6 2

iii. **Classes:** 30 – 40 40 – 50 50 – 60 60 – 70

fi 5 8 4 3

Sol.

i.

x:	23	27	54	35	50
Log x	1.3617	1.4314	1.7324	1.5441	1.6990

$$\sum_{i=1}^5 \text{Log } x_i = 9.4021 \quad n = 5$$

$$G.M = \text{Anti log} \left[\frac{1}{5} (9.4021) \right] = \text{Anti log} (1.5670) = \underline{\underline{36.9}}$$

ii.

x_i	10	16	22	28	34
f_i	5	4	3	6	2
Log x_i	1	1.2041	1.3424	1.4472	1.5315
fi log x_i	5	4.8164	4.0272	8.6532	3.0630

$$\sum_{i=1}^{20} f_i \text{Log } x_i = 25.5598$$

$$G.M = \text{Anti log} \left[\frac{1}{20} (25.5598) \right] = \text{Anti log} (1.2780) = \underline{\underline{18.6}}$$

iii.

Classes:	30 – 40	40 – 50	50 – 60	60 – 70
fi :	5	8	4	3
CMi :	35	45	55	65
Log CMi :	1.5441	1.6532	1.7401	1.8129
fi Log CMi :	7.7200	13.2256	6.9612	5.4387

$$G.M = \text{Anti log} \left[\frac{1}{20} (33.3455) \right] = \text{Anti log} (1.6673) = \underline{\underline{45.81}}$$

Exercise: Calculate GM for the following data.

i. x: 8 40 175 1209 2000

ii. x: 2 3 4 5 6

f: 5 7 8 3 2

iii. Classes: 0 – 10 10 – 20 20 – 30 30 – 40 40 – 50 50 – 60

fi : 2 5 6 18 13 6

iv. Harmonic mean (H.M)

Definition: Harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the given observations. Symbolically, if $x_1, x_2, \dots, x_n$ are n variables of x then their HM is

given by:
$$H.M = \frac{1}{\frac{1}{n} \left[\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right]} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Computations of Harmonic Mean for Ungrouped and Grouped Data

For ungrouped data: $H.M = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$ For grouped data: $H.M = \frac{n}{\sum_{i=1}^n \frac{f_i}{x_i}}$

Example: Find the harmonic mean of the following data

i. x : 20 30

ii.

x	2	3	4	5	6
fi	5	7	8	3	2

iii.

Classes	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54
fi	11	18	32	37	21	47	13

Sol:

i. 20, 30
$$H.M = \frac{2}{\frac{1}{20} + \frac{1}{30}} = \frac{2}{\frac{5}{60}} = \frac{120}{5} = \underline{\underline{24}}$$

ii.

x	2	3	4	5	6
f	5	7	8	3	2
f/x	2.5	2.33	2	0.6	0.33

$$\sum_{i=1}^{25} \frac{f_i}{x_i} = 7.76 \quad H.M = \frac{25}{7.76} = \underline{\underline{3.22}}$$

iii.

Classes	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54
fi	11	18	32	37	21	47	13
Cm_i	22	27	32	37	42	47	52
f_i/Cm_i	0.5	0.67	1	1	0.5	1	0.25

$$\sum_{i=1}^{179} \frac{f_i}{CM_i} = 4.92 \quad H.M = \frac{179}{4.92} = \underline{\underline{36.38}}$$

Example: Find the H.M of the following data

i. **x:** 10 25 12 8 5

ii.

Marks	40	50	60	70
No. of Students	20	30	50	10

iii.

Classes	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
fi	4	6	10	12	5	3

Advantages and application areas of the three means

Advantages:

All are i. rigidly defined.

ii. based on all the observations.

iii. suitable for further mathematical tea.

AM: i. easy to calculate and understand.

ii. is least affected by fluctuations of sampling compared to other averages.

GM: i. it gives highest weightage to smaller values and smaller weightage to large values.

ii. it is a proper average to measure the relative change (like percentage increase in Population, sales over a period of time, etc.

HM: i. is not affected very much by fluctuation of sampling.

ii. Particularly useful in averaging speed, special types of rates and ratios where time factor is involved.

iii. Since the reciprocals of the variables are involved, it gives greater weightage to smaller values.

Remarks: Relationship of A.M, G.M. and H.M.

i. $AM = GM = HM$ if all the values are equal in a series.

ii. $HM < GM < AM$ if the values are different in a series.

proof

i. Suppose there are two items x and y in the series, If $x = y = 7$, then

$$AM = \frac{x + y}{2}, \quad GM = \sqrt{xy}, \quad HM = \frac{2}{\frac{1}{x} + \frac{1}{y}}$$

$$AM = \frac{7 + 7}{2} = 7, \quad GM = \sqrt{7 \times 7} = 7, \quad HM = \frac{2}{\frac{1}{7} + \frac{1}{7}} = \frac{2 \times 7}{2} = 7$$

Therefore, $AM = GM = HM$

ii. Suppose there are two items x and y in the series

$$\text{Then } AM = \frac{x + y}{2}, \quad GM = \sqrt{xy}$$

If $x \neq y$, then $x - y > 0$

$$\sqrt{x} - \sqrt{y} > 0$$

$$(\sqrt{x} - \sqrt{y})^2 > 0$$

$$x + y - 2\sqrt{xy} > 0$$

$$x + y > 2\sqrt{xy}$$

$$\frac{x + y}{2} > \sqrt{xy}$$

$$\underline{\underline{AM > GM}}$$

Consider $\frac{x + y}{2} > \sqrt{xy}$ This is proved above.

$$\frac{\sqrt{xy}}{2} > \frac{xy}{x + y} \quad \text{by multiplying both sides by } \frac{\sqrt{xy}}{x + y} \quad \text{we get}$$

$$\sqrt{xy} > \frac{2xy}{x + y} = 2 \cdot \frac{xy}{x + y} = 2 \cdot \frac{1}{\frac{x + y}{xy}} = \frac{2}{\frac{x}{xy} + \frac{y}{xy}} = \frac{2}{\frac{1}{y} + \frac{1}{x}}$$

$$\therefore \sqrt{xy} > \frac{2}{\frac{1}{x} + \frac{1}{y}} \Rightarrow \underline{\underline{GM > HM}}$$

Therefore, $HM < GM < AM$

Note:- We can have the following relationship between the three means.

$$AM \cdot HM = \frac{x + y}{2} \left(\frac{2xy}{x + y} \right) = xy$$

To equalize AM . HM to GM, we put AM . HM under square root

$$GM = \sqrt{xy} = \sqrt{\frac{x + y}{2} \cdot \left(\frac{2xy}{x + y} \right)} = \sqrt{AM \cdot HM}$$

$\therefore GM = \sqrt{AM \cdot HM}$ If there are only two positive observations in the series.

$$\text{Therefore H.M.} = \frac{G.M^2}{A.M.}$$

Some more examples

1. The price of a commodity increased by 5% from 1979 to 1980, by 9% from 1980 to 1981 and by 73% from 1981 to 1982. The average increase from 1979 to 1982 is quoted as 25.6% and not 29%. Verify.

Sol.

Year	Price at the end of the year taking proceeding as 100%. (X)	Log X
1980	100 + 5 = 105	2.0212
1981	100 + 9 = 109	2.0374
1982	100 + 73 = 173	2.2380
		6.2966

$$AM = \frac{5+9+73}{3} = \frac{87}{3} = \underline{\underline{29}}$$

$$GM = \text{Antilog } [1/3(6.2966)] = \text{Antilog } (2.0989) = 125.6$$

Therefore, Rise in price is 125.6 - 100 = 25.6%

Verification:

<u>Year</u>	<u>Rise</u>	<u>Price would be</u>	<u>Growth 25.6%</u>	<u>Growth 29%</u>
1979		100	100	100
1980	5%	105	125.6	129
1981	9%	114.45	157.75	166.41
1982	73%	198	198	214.67

Thus GM is the best average to give us the true rise in price.

2. World Population has increased from 5 billion to 6 billion within 12 years.
Calculate the average increment per year.

Sol.

The average annual increase is computed by applying the formula

$$P_n = P_0(1 + r)^n \quad \text{or} \quad r = \sqrt[n]{P_n / P_0} - 1 \quad \text{Where } P_n \text{ - the amount at the end of the } n^{\text{th}} \text{ period}$$

P_0 - the amount at the beginning of the period n - time (years)

r - rate of change $P_n = 6$, $P_0 = 5$, $n = 12$ $r = ?$

$$r = \sqrt[12]{6/5} - 1 = 1.01 - 1 = 0.01 \quad \therefore \text{The average increment per annum} = 1\%$$

Therefore, GM is used in determination of average percentage of change in amount.

3. A machine depreciates by 40% in the first year, by 25% in the second year and by 10% per annum for the next three years. Each percentage being calculated on the diminishing value, what is the average percentage of depreciation for the entire period?

Sol.

<u>Depreciation (%)</u>	<u>After depreciation (%) = X</u>	<u>Log X</u>
40%	60%	1.7782
25%	75%	1.8751
10%	90%	1.9542
10%	90%	1.9542
10%	90%	1.9542
		9.5159

$$GM = \text{Antilog} [1/5(9.5159)] = \text{Antilog} (1.9032) \approx 81$$

$\therefore$ Rate of depreciation per annum is $100 - 81 = 19\%$

4. A cyclist pedals from his house to his college at a speed of 8 kmph and back from the college to home at 12 kmph. Find the average speed.

Sol.

Let the distance between the house and the college be x kms. Then the distance from house to college is covered in $x/8$ hrs and from college to house in $x/12$ hrs.

And the total distance = $2x$ (house to college and back) is covered in $(x/8 + x/12)$ hrs.

$\therefore$ Average Speed = $\frac{\text{Total distance traveled}}{\text{Time taken}}$

$$= \frac{2x}{x/8 + x/12} = \frac{2x}{5x/24} = \frac{48x}{5x} = 9.60 \text{ kmph}$$

Exercise:

1. If the arithmetic mean and the geometric mean of two items is 12.5 and 10 respectively, then
 - i. find the HM of the two items.
 - ii. find the value of the two items.
2. A motorist travels at a uniform speed of 20 kmph, 60 kmph and 30 kmph from A to B, B to C and C to D respectively. Find the average speed.
3. In a factory, a unit of work is completed by A in 5 minutes, by B in 7 minutes, by C in 4 minutes, by D in 8 minutes and by E in 6 minutes.
 - i. What is their average rate of work?
 - ii. What is the average number of units of work completed per minute?
 - iii. At this rate, how many units of work will they complete in six hours a day?
5. Find the average rate of increase in Population which in the first decade had increased by 20%, in the next by 30% and in the third by 40%.

Summary:

Arithmetic mean is mostly used in practice of all areas because its characteristics value being represented to all items in the variable.

Geometric mean is widely used in averaging ratios and percentages and in computing average rates of increase or decrease.

Harmonic mean is useful in comparing the values of a variable with constant quantity of another variable. i.e.time, rate, distance covered, quantities purchased or sold per unit etc.

Exercise : i. Given data:

Income ('000)	Below 10	10-30	30-60	60-100	Above 100
# of Students	5	10	15	8	2

Calculate the mean income.

ii. Calculate the number of shops corresponding to class interval 30 - 40 of the following distribution if the mean profit is 28.

Profit per Shop	0-10	10-20	20-30	30-40	40-50	50-60
# of Shops	12	18	27	f_4	17	6

iii. Find the class intervals if the AM of the following distribution is 30.1 and assumed mean is 31.5.

Step deviations (d')	-3	-2	-1	0	1	2
Frequency (f_i)	5	10	25	30	20	10

iv. The mean weight of 150 students of a class is 60kgs. The mean weight of boys is 70kgs and that of girls is 55kgs. Find the number of boys and girls in the class.

- v. The price of a commodity increased by 20% in 1989, decreased by 12% in 1990 and increased by 15% in 1991. Calculate the average annual change in price.
- vi. If the price of a commodity triples in a period of 6 years, what is the average percentage increase per year?
- vii. A train runs the first 40 kms at a speed of 60 kmph, the next 60 kms at a speed of 80 kmph and the last 80 kms at a speed of 100 kmph. What is the average speed of the train for the whole journey?
- viii. If the GM of two positive observations is $\frac{2}{3}$ of their AM and the sum of the two observations is 18, then
- their HM is _____
 - the two observations are _____ and _____

b. Positional measures of central tendency

Introduction

The mode and median are called positional measures of central tendency. The term position refers to the place of a value in the series. The values being divided by a number of equal parts are called partition values. Besides median, which divides a series in to equal parts, the quartiles, deciles and percentiles are important measures.

The Mode : A value, which occurs most frequently in a series of observations, is called Mode. So by looking the observations mode can be identified.

It is the value, which has the greatest frequency density in its immediate neighborhood.

Importance:

1. Mode can be used as a central location for qualitative as well as quantitative data, like the median. Example, if a beauty measurement turns in to three impressions or scores, which we rate 'very beautiful', 'beautiful' and 'not beautiful', then the modal value is beautiful.
2. Like the mean, the mode is not affected by extreme values.
3. Mode can be used when one or more of the classes are open-ended.

Computation of Mode for Ungrouped and Grouped Data

For ungrouped data: Mode $(\hat{x})$ = that value in the data set, which occurs most often.

For grouped data: Discrete Series: Mode $(\hat{x})$ = the value of the variable corresponding to the maximum frequency.

Continuous Series: The class corresponding to the maximum frequency is called the modal class. The value of mode is obtained by the following interpolation formula.

$$\text{Mode}(\hat{x}) = l + \left[\frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \right] \times c$$

or

$$\text{Mode} \hat{x} = l + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c$$

Where l – LCB of the modal class

f_2 – frequency succeeding f_1

f_1 – maximum frequency

C – magnitude of the class

f_0 – frequency preceding f_1

$\Delta_1 = f_1 - f_0$

f_2 – frequency succeeding f_1

$\Delta_2 = f_1 - f_2$

Example: Find the value of mode for the following data

i. 25, 15, 23, 40, 27, 25, 23, 25, 20, 19, 22, 24, 25

ii.

x	10	20	30	40	50	60
f	4	9	16	25	22	15

iii.

Classes	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 - 79
fi	328	350	720	664	598	524	378	244

Sol.

- i. Mode = value which occurs most often , Mode = 25
- ii. Mode = Value of the variable with maximum frequency so Mode = 40
- iii. Modal Class = 19.5 - 29.5

$$l = 19.5 \quad f_0 = 350 \quad f_1 = 720 \quad f_2 = 664 \quad c = 10$$

$$\begin{aligned} \text{Mode}(\hat{x}) &= l + \left[\frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \right] \times c \\ &= 19.5 + \left[\frac{720 - 350}{(720 - 350) + (720 - 664)} \right] \times 10 = 19.5 + \frac{3700}{426} = \underline{\underline{28.1854}} \end{aligned}$$

Exercise: Find the modal value of the following data.

- i. 27, 33, 42, 25, 23, 27, 25, 33, 27, 28, 16, 18

ii.

Height (in inches)	58	60	61	62	63	65	68	70
No. of Persons	4	6	10	8	20	24	9	5

iii.

Classes	0 - 400	400 - 800	800 - 1200	1200 - 1600	1600 - 2000
fi	4	12	40	41	27

ii. The midrange: The midrange of a set of numbers is the average, obtained by adding the highest score and the lowest score. Ie add the highest and the lowest and divide the result by 2.

iii.. The Median

The median is that value of the variable, which divides the group in to two equal parts, one part comprising all the values greater and the other all the values less than median.

Or median can be defined as the middle value of a set of data values when they are arranged in ascending or descending order.

Importance:

- ◆ In dealing with qualitative data, median is more suitable average
- ◆ Median is recommended if the distribution has unequal classes, since it is simple to compute than the mean.
- ◆ Median is especially useful in case of open-ended classes since it is only positional and not calculated average.
- ◆ The magnitudes of extreme deviations do not influence the median.

Commutation of Median for Ungrouped and Grouped Data

For ungrouped data:

First, rearrange the values in the order of magnitude. Then apply the following formula.

$$\begin{aligned} \text{Median}(\tilde{x}) &= \text{value of the } \left(\frac{N+1}{2}\right)^{\text{th}} \text{ item (where n is odd)} \\ &= \frac{x_{n+1}}{2} \\ \text{Median}(\tilde{x}) &= \frac{1}{2} \left[\text{Value of } \left(\frac{N}{2}\right)^{\text{th}} \text{ item} + \text{Value of } \left(\frac{N}{2} + 1\right)^{\text{th}} \text{ item} \right] \text{ (Where n is even)} \\ &= \frac{1}{2} \left[x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right] \end{aligned}$$

For grouped data:

Discrete Series:

1. Compute the $< cf_i$ column.
2. Search for the value of $\frac{N+1}{2}$, if not available, consider the value just greater than it, in the column of $< cf_i$.

Continuous Series:

1. Compute the $< cf_i$ column.
2. Search for the value of $\frac{N}{2}$ if not available, consider the value just greater than it, in the column of $< cf_i$.
3. Then the following interpolation formula is used to calculate the median.

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

Where l - LCB of the median class

c - class interval of the median class

f - frequency of the median class

$c.f$ - cumulative frequency just less than $\frac{N}{2}$

Example: Find the median of the following data.

i. a) 27, 33, 42, 25, 23, 25, 33, 28, 27, 16, 18, 12

b) 8, 5, 2, 6, 15, 10, 25

ii.

x	4	6	8	10	12	14	16
f	2	4	5	3	2	4	1

iii.

x	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100	100 - 110
fi	20	21	50	40	53	16

Sol.

i. a. Rearranging:

12 16 18 23 25 25 27 27 28 33 33 42

$n = 12 \dots$ even

$$\tilde{x} = \frac{1}{2} \left[x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right] = \frac{1}{2} [x_6 + x_7] = \frac{1}{2} [25 + 27] = \underline{\underline{26}}$$

b. Rearranging: 2 5 6 8 10 15 25 n = 7 ... odd

$$\tilde{x} = \frac{x_{n+1}}{2} = x_4 = \underline{\underline{8}}$$

ii.

x	4	6	8	10	12	14	16
f	2	4	5	3	2	4	1
<cfi	2	6	11	14	16	20	21

$$\begin{aligned} n &= 21 & \text{Median} &= \text{The value of } \left(\frac{N+1}{2} \right)^{\text{th}} \text{ item} \\ & & &= \left(\frac{21+1}{2} \right)^{\text{th}} \text{ item} \\ & & &= \text{The value of the } 11^{\text{th}} \text{ item} \\ & & &= \underline{\underline{8}} \end{aligned}$$

iii.

x	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100	100 - 110
fi	20	21	50	40	53	16
<cfi	20	41	91	131	184	200

$$\text{Median class} = \text{Value of } \left(\frac{n}{2} \right)^{\text{th}} \text{ item} = 100^{\text{th}} \text{ item} = 80 - 90$$

$$l = 80, \quad c = 10, \quad f = 40, \quad c.f = 91$$

$$\text{Median } (\tilde{x}) = l + \frac{c}{f} \left(\frac{n}{2} - c.f \right) = 80 + \frac{10}{40} (100 - 91) = 80 + \frac{9}{4} = \underline{\underline{82.25}}$$

Exercise: Find the median of the following data.

i. a. 20, 15, 21, 13, 22, 24, 22, 25, 26, 27, 25

b. 120, 125, 112, 137, 129, 127

ii.

x	28	30	32	34	36	38	40	42
f	14	15	16	24	16	10	6	4

iii.

x	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59
f_i	5	10	15	20	6	4

Remark: The value of median can be determined graphically as follows:

Draw both the more than and less than o-gives on the same graph. From the point of intersection of these two curves, draw a perpendicular line to the x – axis. The foot of the perpendicular line is the value of the median.

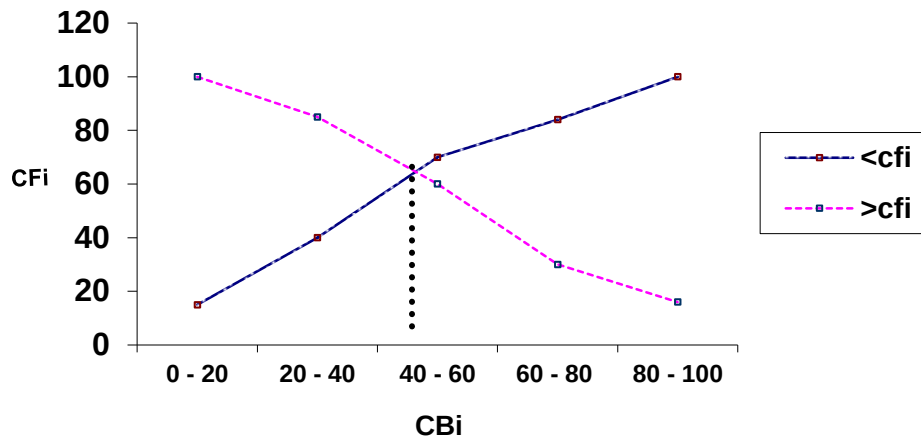
Examples: Find the median of the following frequency distribution graphically.

Classes	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
f_i	15	25	30	14	16

Sol.:

Classes	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
f_i	15	25	30	14	16
<c_fi	15	40	70	84	100
>c_fi	100	85	60	30	16

The < & > Ogives



The perpendicular line drawn from the intersection point meets the x-axis approximately at 46. Therefore, the Median of the distribution is 46.

iv. Quartiles, Deciles and percentiles

Definitions:

Quartiles: are the three values, which divide the given data in to four equal parts. They are denoted by Q_1 , Q_2 and Q_3 .

Q_1 - The lower or first quartile. It covers 25% of the distribution.

Q_2 - The middle or second quartile. It covers 50% of the distribution.

Q_3 - The upper or third quartile. It covers 75% of the distribution.

Deciles: are the nine values, which divide the series in to ten equal parts. They are denoted by $D_1, D_2, \dots, D_9$.

D_1 = Covers 10% of the distribution

D_2 = Covers 20% of the distribution

.

.

.

D_9 = Covers 90% of the distribution

Percentiles: are the 99 values, which divide the series in to 100 equal parts. They are denoted by $P_1, P_2, \dots, P_{99}$.

Note that: i. $Q_1 = P_{25}$ ii. $Q_2 = D_5 = P_{50} = \text{median}$ $Q_3 = P_{75}$ and
iii. $D_1 = P_{10}, D_2 = P_{20}, D_3 = P_{30}, \dots, D_9 = P_{90}$.

Importance: The quartiles are more widely used in Economics and Business while the deciles and percentiles are important in Psychology and Educational Statistics concerning grades, rates, ranks, etc. The working principle for computing the partition value is basically the same as that of computing the median.

Computation of Quartiles, Deciles and Percentiles for Ungrouped and Grouped Data

For ungrouped data and discrete series:

First, for ungrouped data, rearrange the values in the order of magnitude and for discrete series, compute the $<Cf_i$ column. Then apply the following formula.

$$Q_i = \text{value of } \frac{i(N + 1)^{th}}{4} \text{ item}$$

$$D_i = \text{value of } \frac{i(N + 1)^{th}}{10} \text{ item}$$

$$P_i = \text{value of } \frac{i(N + 1)^{th}}{100} \text{ item}$$

For continuous series:

1. Compute the $<cf_i$ column.
2. Determine the quartile, decile or percentile class.
3. Apply the following interpolation formula.

$$Q_i = l + \frac{c}{f} \left(\frac{iN}{4} - c.f \right)$$

$$D_i = l + \frac{c}{f} \left(\frac{iN}{10} - c.f \right)$$

$$P_i = l + \frac{c}{f} \left(\frac{iN}{100} - c.f \right)$$

Example: For the data given below, compute the value of Quartiles, D_3 , D_7 , P_{15} and P_{88} and interpret.

Marks	Below 10	10 - 20	20 - 40	40 - 60	60 - 80	Above 80
No. of Students	10	15	25	30	14	6
<cf _i	10	25	50	80	94	100

Sol.

Q_1 – size of $\frac{N}{4}$ item = 25th item 10 – 20 quartile class

$$l = 10, \quad c = 10, \quad f = 15, \quad c.f = 10$$

$$Q_1 = l + \frac{c}{f} \left(\frac{n}{4} - c.f \right) = 10 + \frac{10}{15} (25 - 10) = \underline{\underline{20}}$$

Mark of 25% of students is less than 20.

Q_2 – size of $\frac{2N}{4}$ item = 50th item 20 – 40 quartile class

$$l = 20, \quad c = 20, \quad f = 25, \quad c.f = 25$$

$$Q_2 = l + \frac{c}{f} \left(\frac{n}{2} - c.f \right) = 20 + \frac{20}{25} (50 - 25) = \underline{\underline{40}}$$

Mark of half of students is below 40.

Q_3 – size of $\frac{3N}{4}$ item = 75th item 40 – 60 quartile class

$$l = 40, \quad c = 20, \quad f = 30, \quad c.f = 50$$

$$Q_3 = l + \frac{c}{f} \left(\frac{3n}{4} - c.f \right) = 40 + \frac{20}{30} (75 - 50) = \underline{\underline{73.33}}$$

Mark of $\frac{3}{4}^{\text{th}}$ of students is below 73.33.

D₃ – size of $\frac{3N}{10}^{\text{th}}$ item = 30th item 20 – 40 decile class

$$L = 20, \quad c = 20, \quad f = 25, \quad c.f = 25$$

$$D_3 = l + \frac{c}{f} \left(\frac{3n}{10} - c.f \right) = 20 + \frac{20}{25} (30 - 25) = \underline{\underline{24}}$$

Mark of 30% of students is below 24.

D₇ – size of $\frac{7N}{10}^{\text{th}}$ item = 70th item 40 – 60 decile class

$$L = 40, \quad c = 20, \quad f = 30, \quad c.f = 50$$

$$D_7 = l + \frac{c}{f} \left(\frac{7n}{10} - c.f \right) = 40 + \frac{20}{30} (70 - 50) = \underline{\underline{53.33}}$$

Mark of 70% of students is below 53.33.

P₁₅ – size of $\frac{15N}{100}^{\text{th}}$ item = 15th item 10 – 20 percentile class

$$L = 10, \quad c = 10, \quad f = 15, \quad c.f = 10$$

$$P_{15} = l + \frac{c}{f} \left(\frac{15n}{100} - c.f \right) = 10 + \frac{10}{15} (15 - 10) = \underline{\underline{13.3}}$$

Mark of 15% of students is below 13.3.

P₈₈ – size of $\frac{88N}{100}^{\text{th}}$ item = 88th item 60 – 80 percentile class

$$L = 60, \quad c = 20, \quad f = 14, \quad c.f = 80$$

$$P_{88} = l + \frac{c}{f} \left(\frac{88n}{100} - c.f \right) = 60 + \frac{20}{14} (88 - 80) = \underline{\underline{71.43}}$$

Mark of 88% of students is below 71.43.

Exercise: Compute the value of Quartiles, D₄, P₆₉ and interpret for the data given below.

- i. 46, 35, 28, 52, 54, 43, 35, 49, 46, 50, 41

ii.

Daily Wages	40	45	50	55	60	65	70
No. of Workers	9	22	26	18	13	8	5

iii.

Rent in Birr	150-250	250-350	350-450	450-550	550-650	650-750	750-850	850-950
No. of Houses	8	10	15	25	40	20	15	7

Summary:

The arithmetic mean and median satisfy the conditions of definition and stability. Median has a distinct merit over mean insofar as easy calculations. Mode can be located just by inspection. In case, every value occurs the same number of times mode is useless measure. It is observed that the median, quartiles, deciles and percentiles have good relation.

Exercise:

1. In a class of 15 students, three failed in a test. The marks of 12 students who passed were 9, 6, 7, 8, 8, 9, 6, 8, 7, 5, 4 and 7. What was the median of all the 15 students?
2. Calculate the mode and median for the distribution of the weights in kgs of 150 people from the data given below.

Weight in Kgs	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
No. of People	18	37	45	27	15	8

3. For the data given below, find the missing frequencies if median are 37 and mode is 43 million birr.

Fund raised in millions of birr	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
No. of NGO's	3	F ₂	16	20	F ₂	16

4. The following distribution shows the marks of 60 students in Economics.
Calculate Q_3 , D_5 , P_{57} and the median.

Marks	31 - 39	41 - 49	51 - 59	61 - 69	71 - 79	81 - 89	91 - 99
No. of Students	12	10	12	9	6	7	4

5. For the following data Q_1 is found to be 41. Find the missing frequency.

Classes	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59
f_i	8	10	f_3	20	12	25

Properties of average

The different averages namely the mean, the median, the mode and the midrange are characterized by their different advantages and disadvantages, some are listed below.

1. The mean

Advantages:

- it always exist and is a unique number for any finite set of scores which can be used for further statistical analysis.
- It is the most familiar of the four averages and take every score of the set of data into account, can be located graphically.
- it is used to compare two or more sets of data.
- It is the only measure of central tendency/ average when the sum of the deviation of each value from the mean will always be zero.

Disadvantages:

- The mean can drastically be affected by exceptionally extreme high or low values.
- It cannot be used for non-numerical data unlike the median and the mode,
- It cannot be computed for an open ended frequency distribution.

2. The median

Advantages:

- The median always exist as a unique number or value, relatively easy to compute and can be computed graphically.

- ii. The median exhibits the advantage of not being affected by some exceptional extreme values unlike the mean and midrange. Therefore it is valuable average when such values do occur.
- iii. It can be used for non-numerical data which can be arranged in some order such as order of significance/importance, order of quality, order of difficulty and soon.
- iv. It can be computed for an open ended frequency distribution if the median does not lie in an open ended class.

Disadvantages: the median is not sensitive to the value of every score.

3. The midrange:

Advantage: it always exists as a unique number for any finite set of observation and is easy to compute.

Dis advantage: i) it is too strongly affected by extreme values.

ii) it does not take every score in to account.

4. The mode

Advantage: i) it is not affected by extreme value and is the easiest average to compute whenever it exists, can be dictated graphically.

ii) it is applicable to qualitative (non-numerical) data with no order such as eye color, hair color, body style, opinions and soon.

Dis advantage :i) it may not exist for a given set of data and even if it exists, it may not be unique.

ii) it is poor choice of average in many cases.

Remarks: Though there are no objective criteria which determine the most (best) representative average for a given set of data, we usually consider the list of advantages and disadvantages in the selection of a particular average. The following indicate which average is best or appropriate:

1. If we have a numerical data with extreme high or low score(exceptionally) in the given set of data, the median may be the best average, but if we have a numerical data with no exceptionally high or low scores/ values we use the mean as the best average.

For instance: For the set of numbers,1,2,3,4,5,9,12,13,15,16, the mean and the median are 8 and 7 respectively. But if the largest 16 is replaced by 336, the mean is 40 and the

median is still 7, which remains the same which is close to all except the extreme value (336). So the median is the best representative of the given data set.

2. In an open ended GFD the median is the best average.
3. For a numerical data with no extreme high or low scores the mid-range is relatively better average than the median and the mode.
4. For non-numerical data which can be arranged in some order such as the order of quality, the order of difficulty, the order of importance, the order of significance and soon the median may be the best measure of central tendency.
5. For non-numerical data with no order the mode is the best measure of central tendency. Moreover if a researcher quotes opinions of people, he refers the most frequently expressed opinion of people which is called modal opinion.

Empirical relationship of mean median and mode

For a frequency distribution if $\bar{x} = \tilde{x} = \hat{x}$ then the curve of a frequency distribution is bell shaped and is symmetrical called normal distribution and the distribution which is not symmetrical is called asymmetrical or skewed.

Skewness of a distribution is a tendency of a distribution to depart from symmetry. If extremely high or low values are added to a distribution then the mean tends to shift towards those values and the curve becomes skewed.

If the frequency curve of a distribution has longer tail in the left than of the right from the central maximum then the distribution is said to be skewed to the left (negatively skewed)

If the reverse is true, it said to be skewed to the right (positively skewed). That is

If in a distribution the greatest number of values is at the right of the curve and few scattered extreme small values are at the left end, then the curve is called negatively skewed or left skewed and if the reverse is true, it is called positively or skewed right.

For a symmetrical distribution (normal distribution) the mean, median and mode are identical while for a unimodal frequency curve which are moderately skewed distribution (not very high peaked and not flat topped

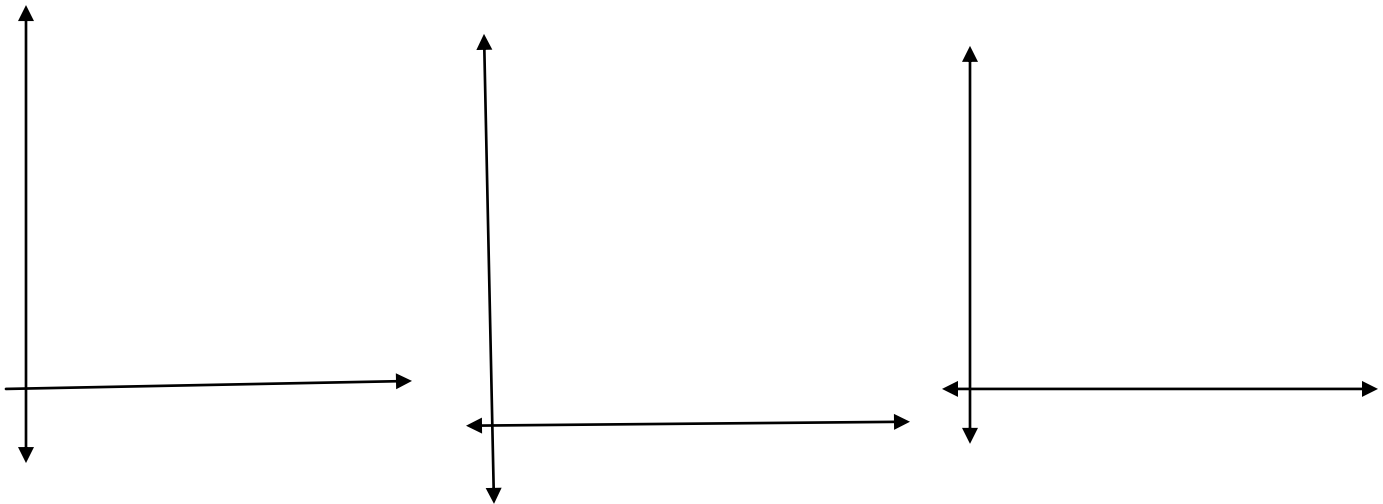
We have the empirical relationship of mean, median and mode as

$$\text{mean} - \text{mode} = 3(\text{mean} - \text{median}) \text{ ie if } \bar{x} - \hat{x} = 3(\bar{x} - \tilde{x})$$

which is an approximate relation that characterize many but not all skewed distribution.

Note: 1. The empirical relationship of mean, median and mode is used to estimate one of the three averages when the value of the other two are known in a moderately skewed distribution.

2. For a unimodal frequency curve which is moderately skewed distribution the relative positions of the mean, median and mode are as shown in the fig below.



Normal symmetrical dist.

$$\bar{x} = \hat{x} = \tilde{x}$$

Left skewed dist.

$$\bar{x} < \tilde{x} < \hat{x}$$

Right skewed dist

$$\bar{x} > \tilde{x} > \hat{x}$$

Sketch the curve for each condition.

Example: If the mode and the mean of a moderately skewed distribution are 63.4 and 61.2 respectively then a. estimate the median b. determine whether the frequency curve is right or left skewed.

Sol.: Given a. $\bar{x} = 61.2$, $\hat{x} = 63.4$, $\tilde{x} = ?$

$$\bar{x} - \hat{x} = 3(\bar{x} - \tilde{x}) \rightarrow 61.2 - 63.4 = 3(61.2 - \tilde{x}) \rightarrow \tilde{x} = 61.92.$$

b. Since $\bar{x} < \tilde{x} < \hat{x}$, the curve of the distribution is left skewed.

Exercise: 1. If the mode and the median of a moderately skewed distribution are 53.8 and 56.8 respectively then a. estimate the mean b. determine whether the frequency curve is right or left skewed.

2. If the median score of the distribution given below is 64.5 and the total number of observation (total frequency) is 90, then find m and n . Find also the a. mean b. mode c. quartile 1 (Q_1) d. quartile 1 (Q_2) e. quartile 1 (Q_1) f. 5th decile

score	40---49	50---59	60---69	70---79	80---89	90---99
f	16	22	m	15	11	n

3. Suppose the distribution given below is the exam score of 75 students with modal score 55.5 then a. determine f_1 and f_2 b. find the mean c. find the median d. find the 10th, 25th, 50th, 75th, 90th percentiles, and interpret.

score	21---30	31---40	41---50	51---60	61---70	71---80	81---90
f	7	11	f_1	15	f_2	11	7