MEASURES OF CENTRAL TENDENCY

INTRODUCTION:

In the study of a population with respect to one in which we are interested, we may get a large number of observations. It is not possible to grasp any idea about the characteristic when we look at all the observations. So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic. Such representative number can be a central value for all these observations. This central value is called a measure of central tendency or an average or a measure of locations. There are five averages. Among them mean, median
and mode are called simple averages and the other two averages geometric mean and harmonic mean are called special averages.

The meaning of average is nicely given in the following definitions.

“A measure of central tendency is a typical value around which other figures congregate.”

“An average stands for the whole group of which it forms a part yet represents the whole.”

“One of the most widely used set of summary figures is known as measures of location.”

CHARACTERISTICS FOR A GOOD OR AN IDEAL AVERAGE:

The following properties should possess for an ideal average.

1. It should be rigidly defined.

2. It should be easy to understand and compute.

3. It should be based on all items in the data.

4. Its definition shall be in the form of a mathematical formula.

5. It should be capable of further algebraic treatment.

6. It should have sampling stability.

7. It should be capable of being used in further statistical computations or processing.

ARITHMETIC MEAN OR MEAN:

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. If the variable x assumes n values x_1, x_2 … x_n then the mean, x, is given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

This formula is for the ungrouped or raw data.

Grouped Data: (Discrete & Continuous Case)

The mean for grouped data is obtained from the following formula:

$$\bar{x} = \frac{\sum fx}{N}$$

where *x* = the mid-point of individual class

f = the frequency of individual class

N = the sum of the frequencies or total frequencies.

MERITS AND DEMERITS OF ARITHMETIC MEAN:

Merits:

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. If the number of items is sufficiently large, it is more accurate and more reliable.
4. It is a calculated value and is not based on its position in the series.
5. It is possible to calculate even if some of the details of the data are lacking.
6. Of all averages, it is affected least by fluctuations of sampling.
7. It provides a good basis for comparison.

Demerits:

1. It cannot be obtained by inspection nor located through a frequency graph.
2. It cannot be in the study of qualitative phenomena not capable of numerical measurement i.e. Intelligence, beauty, honesty etc.,
3. It can ignore any single item only at the risk of losing its accuracy.
4. It is affected very much by extreme values.
5. It cannot be calculated for open-end classes.

6. It may lead to fallacious conclusions, if the details of the data from which it is computed are not given.

calculate the arithmetic mean for the following

1600, 1590, 1560, 1610, 1640, 10

$$\text{Arithmetic mean, } \bar{x} = \frac{1600 + 1590 + 1560 + 1610 + 1640 + 10}{6}$$

$$= \frac{8010}{6} = 1335$$

1. Calculate Arithmetic mean

S.No.	1	2	3	4	5	6	7	8	9	10
Sales in 1000's(x)	34	55	45	62	48	57	28	57	62	78

$$\text{Arithmetic mean, } \bar{x} = \frac{34+ 55 + 45 + 62 + 48 +57+28+57+62+78}{10}$$

$$= \frac{526}{10} = 52.6 \text{ (average sales)}$$

Discrete data

Let x_i, x_2, $x3$…. x_n be the n values of the variable x with corresponding frequency f_i, f_2, f_3…. f_n. then

$$\text{the arithmetic mean } \bar{x} = \frac{x_1.f_1 + x_2.\, f_2 + x_3.f_3 + \ldots + x_n \,.\, f_n}{f_1 + f_2 + f_3 + \ldots + f_n}$$

$$= \frac{\Sigma\, x_i\, f_i}{\Sigma\, f_i}$$

2. Calculate the arithmetic mean

x	f	xf
2	4	2x4= 8
4	6	24
6	10	60
8	12	96
10	8	80
12	7	84
14	3	42
	$\Sigma f_i =$ 50	$\Sigma x_i f_i$ = 394

,

$$\text{Arithmetic mean, X bar} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{394}{50} = 78.8$$

Continuous Series

Type I: Exclusive Class Intervals

3. Calculate the Arithmetic Mean for the following data

Marks	20-30	30-40	40-50	50-60	60-70	70-80
No of Students	5	8	12	15	6	4

Solution

Marks	f	m	fm
20-30	5	25	125
30-40	8	35	280
40-50	12	45	540
50-60	15	55	825
60-70	6	65	390
70-80	4	75	300
	$\sum f = 50$		$\sum fm = 2460$

$$\overline{X} = \frac{\sum fm}{\sum f} = 2460/50 = 49.20$$

4. From the following data, compute Arithmetic mean

Marks Obtained	0-10	10-20	20-30	30-40	40-50	50-60
No of Students	5	10	25	30	20	10

Solution

Marks Obtained	f	m	fm
0-10	5	5	25
10-20	10	15	150
20-30	25	25	625
30-40	30	35	1050
40-50	20	45	900
50-60	10	55	550
	$\sum f = 100$		$\sum fm = 3300$

$$\overline{X} = \frac{\sum fm}{\sum f} = 3300/100 = 33$$

MEDIAN

Definition

Median is the value of the middle most items when all the items are in the order of magnitude. It is denoted by M.

Merits of Median:

1. Median is not influenced by extreme values because it is a positional average.

2. Median can be calculated in case of distribution with open end intervals.

3. Median can be located even if the data are incomplete.

4. Median can be located even for qualitative factors such as ability, honesty etc.

Demerits of Median:

1. A slight change in the series may bring drastic change in median value.

2. In case of even number of items or continuous series, median is an estimated value other than any value in the series.

3. It is not suitable for further mathematical treatment except its use in mean deviation.

4. It is not taken into account all the observations.

Formulae:

Individual Series: Position of M = $\left(\frac{N+1}{2}\right)^{th}$ term

Discrete Series: Position of M = $\left(\frac{\sum f+1}{2}\right)^{th}$ term

Continuous Series: M = $L+\left[\frac{\sum f/2 - p.c.f}{f}\right]i$

Where L = Lower limit of the Median class

f = Frequency of the median class

p.c.f = Cumulative frequency preceding the median class

i = Length of the class interval

Individual Series:

1. Find the Median for the following data

6, 9, 21, 5, 7, -2, 0, 32, 9

Solution:

N = 9 (Odd number)

Ascending Order: -2, 0, 5, 6, 7, 9, 9, 21, 32,

Position of Median = $\left(\frac{N+1}{2}\right)^{th}$ term

$$= \left(\frac{9+1}{2}\right)^{th} \text{ term}$$

$$= \left(\frac{10}{2}\right)^{th} \text{ term}$$

$= 5^{th}$ term

$\therefore$ M = 7

2. Find the Median

57, 58, 61, 42, 38, 65, 72, 66

Solution:

N = 8 (Even number)

Ascending Order: 38, 42, 57, 58, 61, 65, 66, 72,

Position of Median = $\left(\frac{N+1}{2}\right)^{th}$ term

$$= \left(\frac{8+1}{2}\right)^{th} \text{ term}$$

$$= \left(\frac{9}{2}\right)^{th} \text{ term}$$

$= 4.5^{th}$ term

$\therefore$ M = average of 4^{th} and 5^{th} number

= 58+61/2

= 119/2

$\therefore$ M = 59.5

Discrete Series:

3. The marks (out of a maximum of 10) scored by the students of a class are given below. Find the Median mark

Mark	3	4	5	6	7	8	9	10	Total
No of Students	1	5	6	7	10	15	10	5	59

Solution:

Marks	No of Students	Cumulative frequency

x	f	cf
3	1	1
4	5	5+1 = 6
5	6	6+6 = 12
6	7	19
7	10	29
8	15	44
9	10	54
10	5	59
	$\sum f = 59$	

Solution:

N = 59(Odd Number)

Position of M = $\left(\frac{\sum f + 1}{2}\right)^{th}$ term

$= \left(\frac{59+1}{2}\right)^{th}$ term

$= \left(\frac{60}{2}\right)^{th}$ term

$= 30^{th}$ term

$\therefore$ M = 8

4. Find the Median from the following data

Wages	50	75	100	150	250	Total
No of Labourers	8	14	10	5	3	40

Solution:

Wages x	No of Labourers f	Cumulative frequency cf
50	8	8
75	14	22
100	10	32

150	5	37
250	3	40
	$\sum f$ = 40	

N = 40(Even Number)

Position of M = $\left(\frac{\sum f + 1}{2}\right)^{th}$ term

$= \left(\frac{40+1}{2}\right)^{th}$ term

$= \left(\frac{41}{2}\right)^{th}$ term

$= 20.5^{th}$ term

$\therefore$ M = Average of 20^{th} term and the 21^{st} term

= 75+75/2

= 150/2

$\therefore$ M = 75

5. Find the Median from the following data

No of cars sold in a day	10	15	17	18	21	Total
No of Days	4	16	12	5	3	40

Solution:

Wages x	No of Labourers f	Cumulative frequency cf
10	4	4
15	16	20
17	12	32
18	5	37
21	3	40
	$\sum f$ = 40	

Solution:

N = 40(Even Number)

Position of M = $\left(\frac{\sum f+1}{2}\right)^{th}$ term

$= \left(\frac{40+1}{2}\right)^{th}$ term

$= \left(\frac{41}{2}\right)^{th}$ term = 20.5^{th} term

∴ M = Average of 20^{th} term and the 21^{st} term

= 15+17/2 = 32/2

∴ M = 16

Continuous Series:

Type I: Exclusive Class Interval

6. Calculate the Median Height.

Height (cms.)	145-150	150-155	155-160	160-165	165-170	170-175
No of Students	2	5	10	8	4	1

Solution:

Height (cms.)	No of Students f	Cumulative frequency cf
145-150	2	2
150-155	5	7
155-160	10	17
160-165	8	25
165-170	4	29
170-175	1	30
	$\sum f$ = 30	

Solution:

$\frac{\sum f}{2} = \frac{30}{2} = 15$

Median Class = 155-160

L = 155

p.c.f = 7

f= 10

i= 5

M = $L + \left[\frac{\sum f/2 - p.c.f}{f} \right] i$

$$= 155 + \left[\frac{15-7}{10} \right] 5 = 155 + \left[\frac{8}{10} \right] 5 = 155+4 = 159 \text{ cms.}$$

$\therefore$ M = 159 cms.

7. Calculate the Median from the following data:

Marks	10-25	25-40	40-55	55-70	70-85	85-100
Frequency	6	20	44	26	3	1

Solution:

Marks	frequency f	Cumulative frequency cf
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100
	$\sum f$ = 100	

$$\frac{\sum f}{2} = \frac{100}{2} = 50$$

Median Class = 40-55

L = 40

p.c.f = 26

f=44

i= 15

M = $L + \left[\frac{\sum f/2 - p.c.f}{f} \right] i$

$$= 40 + \left[\frac{50-26}{44}\right]15 = 40 + \left[\frac{24}{44}\right]15 = 40+8.18$$

$\therefore$ M = 48.18

MODE

Definition:

Mode is the value which has the greatest frequency density. It is denoted by 'Z'.

Formulae:

Individual Series: Most repeated numbers.

Discrete Series: The value straight to the greatest frequency.

Continuous Series: $Z = L + \left[\frac{D_1}{D_1 + D_2}\right]i$

Where L = Lower limit of the modal class.

$D_1 = f_1 - f_0$

= Frequency of the modal class – frequency of the class preceding the modal class.

$D_2 = f_1 – f_2$

= Frequency of the modal class – frequency of the class succeeding the modal class.

i = Size of the Modal class.

Individual Series:

1.Determine the Mode.

(i) 320, 395, 342, 444, 551, 395, 425, 417, 395, 401, 390, 400.

Mode = 395

(ii) 3,6,7,5,8,4,9.

Mode = No mode

(iii) 25,32,24,27,32,27,25,32,24,27,25,24.

Mode = No Mode

(iv) 0,2,5,6,9,5,6,14,6,15,5,6,5.

Mode = 5 and 6

Discrete Series:

2.Determine the mode:

Size of the dress (x)	18	20	22	24
No of sets produced(f)	55	120	108	45

Solution: Greatest frequency = 120

Mode = 20

3.Determine the modal size:

Size of Shoes (x)	3	4	5	6	7	8	9
No of pairs sold (f)	10	25	32	38	61	47	34

Solution: Greatest frequency = 61

Mode = 7

Continuous Series:

Type I: Exclusive Class Interval

4.Calculate the Mode:

Daily wages in Rs.	50-60	60-70	70-80	80-90	90-100
No of Labourers	40	62	75	100	65

Solution:

Daily wages in Rs.	No of Labourers
50-60	40
60-70	62
70-80	$75 = f_0$
80-90	$100 = f_1$
90-100	$65 = f_2$

Greatest Frequency = 100

Modal class = 80-90

$L = 80$

$D_1 = f_1 - f_0 = 100-75 = 25$

$D_2 = f_1 - f_2 = 100-65 = 35$

$i = 10$

$$Z = L + \left[\frac{D_1}{D_1 + D_2}\right] i = 80 + \left[\frac{25}{25+35}\right]10 = 80 + \left[\frac{25}{60}\right]10 = 80+4.17$$

Z = Rs. 84.17

Merits of Mode:

1. It is easy to calculate and in some cases it can be located mere inspection
2. Mode is not at all affected by extreme values.
3. It can be calculated for open-end classes.
4. It is usually an actual value of an important part of the series.
   In some circumstances it is the best representative of data.

Demerits of mode:

1. It is not based on all observations.
2. It is not capable of further mathematical treatment.
3. The Mode is ill-defined, generally, it is not possible to find mode in some cases.
4. As compared with the mean, mode is affected to a great extent, by sampling fluctuations.
5. It is unsuitable in cases where the relative importance of items has to be considered.

Geometric Mean

Definition:

Geometric mean of N values is the N^{th} root of the product of the N values. It is abbreviated as G.M.

Formulae:

Individual Series: G.M = Antilog$\left(\frac{\sum \log X}{N}\right)$

Discrete Series: G.M = Antilog$\left(\frac{\sum f \log X}{\sum f}\right)$

Continuous Series: G.M = Antilog$\left(\frac{\sum f \log m}{\sum f}\right)$

1.Find the Geometric Mean of 3, 6, 24, 48

Solution:

X	Log X
3	0.4771
6	0.7782
24	1.3802
48	1.6812
	∑log X = 4.3167

$$G.M = \text{Antilog}\left(\frac{\sum \log X}{N}\right)$$

$$= \text{Antilog}\left(\frac{4.3167}{4}\right) = \text{Antilog}(1.0792) = 12$$

2.Calculate the Geometric Mean for the following data:

X	f	Log X	f log X
10	4	1.0000	4.0000
15	6	1.1761	7.0566
25	10	1.3979	13.9790
40	7	1.6021	11.2147
50	3	1.6990	5.0970
	$\sum f = 30$		∑f log X = 41.3473

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

$$= \text{Antilog}\left(\frac{41.3473}{30}\right) = \text{Antilog}(1.3782) = 23.89$$

3. Compute the Geometric Mean for the following series:

Marks	No of Students f	m	Log m	f log m
0-10	5	5	0.6990	3.4950
10-20	7	15	1.1761	8.2327
20-30	15	25	1.3979	20.9685
30-40	25	35	1.5441	38.6025
40-50	8	45	1.6532	13.2256
	$\sum f = 60$			$\sum f \log m = 84.5243$

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

$$= \text{Antilog}\left(\frac{84.5243}{60}\right) = \text{Antilog}(1.4087) = 25.63.$$

Merits of GM

- It is based on all the items of the data.
- It is rigidly defined. It means different investigators will find the same result from the given set of data.
- It is a relative measure and given less importance to large items and more to small ones unlike the arithmetic mean.
- Geometric mean is useful in ratios and percentages and in determining rates of increase or decrease.
- It is capable of algebraic treatment. It mean we can find out the combined geometric mean of two or more series.

Demerits or Limitations of GM

- It is not easily understood and therefore is not widely used.
- It is difficult to compute as it involves the knowledge of ratios, roots, logs and antilog.
- It becomes indeterminate in case any value in the given series happen to be zero or negative.
- With open-end class intervals of the data, geometric mean cannot be calculated.
- Geometric mean may not correspond to any value of the given data.

HARMONIC MEAN

Definition:

Harmonic mean is the reciprocal of the mean of the reciprocals of the values. It is abbreviated as H.M.

Formulae:

Individual Series: H.M = $\dfrac{N}{\sum \frac{1}{X}}$

Discrete Series: H.M = $\dfrac{\sum f}{\sum \frac{f}{X}}$

Continuous Series: H.M = $\dfrac{\sum f}{\sum \frac{f}{m}}$

1. Find the Harmonic Mean for the following individual data:

6, 15, 35, 40, 900, 520, 300, 400, 1800, 2000

X	1/X
6	0.1667
15	0.0667
35	0.0286
40	0.0250
900	0.0011
520	0.0019
300	0.0033
400	0.0025
1800	0.0006
2000	0.0005
	$\sum \frac{1}{X} = 0.2969$

$$H.M = \frac{N}{\sum \frac{1}{X}} = \frac{10}{0.2969} = 33.68$$

2. Calculate the Harmonic mean from the following data:

X	f	f/X
10	5	0.5000
12	18	1.5000
14	20	1.4286
16	10	0.6250
18	6	0.3333
20	1	0.0500
	$\sum f = 60$	$\sum \frac{f}{X} = 4.4369$

$$\text{H.M} = \frac{\sum f}{\sum \frac{f}{X}} = \frac{60}{4.4369} = 13.52$$

3..Calculate the Harmonic mean for the following data:

Value	frequency	m	f/m
0-10	8	5	1.6000
10-20	12	15	0.8000
20-30	20	25	0.8000
30-40	6	35	0.1714
40-50	4	45	0.0889
	$\sum f = 50$		$\sum \frac{f}{m} = 3.4603$

$$\text{H.M} = \frac{\sum f}{\sum \frac{f}{m}} = \frac{50}{3.4603} = 14.45$$

The harmonic mean has the following merits.

1. It is rigidly defined.
2. It is based on all the observations of a series i.e. it cannot be calculated ignoring any item of a series.
3. It is capable of further algebraic treatment.
4. It gives better result when the ends to be achieved are the same for the different means adopted.
5. It gives the greatest weight to the smallest item of a series.

6. It can be calculated even when a series contains any negative value.
7. It makes a skewed distribution a normal one.
8. It gives a curve straighter than that of the arithmetic and geometric mean.

Demerits

However, the harmonic mean suffers from the following demerits.

1. It is not easy to understand by a man of ordinary prudence.
2. Its calculation is cumbersome as it involves finding out of the reciprocals of the numbers.
3. It does not give better and accurate results when the means adopted are the same for the different ends achieved.
4. Its algebraic treatment is very much limited and not far and wide as that of the arithmetic mean.
5. It is greatly affected by the values of the extreme items.
6. It can not be calculated, if any, of the items is zero.