Module 2

MEASURES OF CENTRAL TENDENCY

A measure of central tendency helps to get a single representative value for a set of usually unequal values. This single value is the point of location around which the individual values of the set cluster. Hence the averages are known also as *measures of location*.

The important measures of central tendencies or statistical averages are the following.

- Arithmetic Mean
- Geometric Mean
- Harmonic Mean
- Median
- Mode

Weighted averages, positional values, viz., quartiles, deciles and percentiles, also are considered in this chapter.

Criteria or Desirable Properties of an Average

It should be rigidly defined: That is, it should have a formula and procedure such that different persons who calculate it for a set of values get the same answer.

It should have sampling stability: A number of samples can be drawn from a population. The average of one sample is likely to be different from that of another. It is desired that the average of any sample is not much different from that of any other.

1. Arithmetic Mean

The arithmetic mean (AM) or simply mean is the most popular and widely used average. It is the value obtained by dividing sum of all given observations by the number of observations. AM is denoted by $\bar{x}$ (x bar).

Definition for a raw data

For a raw data or ungrouped data if $x_1, x_2, x_3, \ldots, x_n$ are n observations,

then $$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

ie., $\bar{x} = \frac{\sum x}{n}$ where the symbol $\sum$ (sigma) denotes summation.

Example 1

Calculate the AM of 12, 18, 14, 15, 16

Solution

$$\bar{x} = \frac{\sum x}{n} = \frac{12 + 18 + 14 + 15 + 16}{5} = \frac{75}{5} = \mathbf{15}$$

Definition for a frequency data

For a frequency data if $x_1, x_2, x_3, \ldots, x_n$ are 'n' observations or middle values of 'n' classes with the corresponding frequencies

$f_1, f_2, \ldots, f_n$ then AM is given by

$$\bar{x} = \frac{f_1 \times x_1 + f_2 \times x_2 + \ldots + f_n \times x_n}{f_1 + f_2 + \ldots. + f_n} = \frac{\sum fx}{\sum f}$$

ie., $\bar{x} = \frac{\sum fx}{N}$ where $N = \sum f$ = Total frequency

Example 2

The following data indicate daily earnings (in rupees) of 40 workers in a factory.

Daily earnings in ₹	:	5	6	7	8	9
No of workers	:	3	8	12	10	7

Calculate the average income per worker.

X d = x 320

Solution 305 15

Daily Earnings in ₹320(x)	No. 0of workers (f)	fx
5 332	12 3	15
6 350	30 8	48
7	12	84
8	10	80
9	7	63
Total	40	290

$$\bar{x} = \frac{\sum fx}{N} = \frac{290}{40} = \mathbf{7.25}$$

Average income per worker is ₹ **7.25**

Example 3

Calculate the AM of the following data

Class	:	04	4-8	8-12	12-16
Frequency	:	1	4	3	2

Solution

Class	f	Mid values (x)	fx
0-4	1	2	2
4-8	4	6	24
8-12	3	10	30
12-16	2	14	28
Total	10		84

$$\bar{x} = \frac{\sum fx}{N} = \frac{84}{10} = \mathbf{8.4}$$

Shortcut Method: Raw data

Suppose the values of a variable under study are large, choose any value in between them. Preferably a value that lies more or less in the middle, called arbitrary origin or assumed mean, denoted by A. Take deviations of every value from the assumed mean A.

Let d = x A, Taking summation of both sides and dividing by n, we get

$$\bar{x} = A + \frac{\sum d}{n}$$

Example 4

Calculate the AM of 305, 320, 332, 350

Solution

X	d = x 320
305	15
320	0
332	12
350	30
	27

$$\bar{x} = A + \frac{\sum d}{n}$$

$$320 + \frac{27}{4}$$

320+6.75

326.75

Shortcut Method: Frequency Data

When the frequencies and the values of the variable x are large the calculation of AM is tedious. So a simpler method is adopted. The deviations of the mid values of the classes are taken from a convenient origin. Usually the mid value of the class with the maximum frequency is chosen as the arbitrary origin or assumed mean. Thus change x values to ‘d’ values by the rule,

$$d = \frac{x - A}{c}$$

where A-assumed mean, c-class interval, x-mid values. Then the formula for calculating AM is given by

$$\bar{x} = A + \frac{\sum fd}{N} \times c$$

Example 5

Calculate AM from the following data

Weekly wages	:	0-10	10-20	20-30	30-40	40-50
Frequency	:	3	12	20	10	5

Solution

Weekly wages	f	Mid value x	$d = \frac{x-25}{10}$	fd	
0-10	3	5	2	6	18
10-20	12	15	1	12	
20-30	20	25	0	0	
30-40	10	35	1	10	20
40-50	5	45	2	10	
Total	50			2	

$$\bar{x} = A + \frac{\sum fd}{N} \times c = 25 + \frac{2}{50} \times 10 = 25 + 0.4 = \mathbf{25.4}$$

Properties

The AM is preserved under a linear transformation of scale.

That is, if x_i is changed to y_i by the rule

$y_i = a + b\,x_i$, then $\bar{y} = a + b\,\bar{x}$, which is also linear.

The mean of a sum of variables is equal to the sum of the means of the variables.

Algebraic sum of the deviations of every observation from the A.M. zero.

If n_1 observations have an A.M $\bar{x}_1$ and n_2 observations have an AM $\bar{x}_2$ then the AM of the combined group of $n_1 + n_2$ observations is given by $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$.

Example 6

Let the average mark of 40 students of class A be 38; the average mark of 60 students of another class B is 42. What is the average mark of the combined group of 100 students?

Here $n_1 = 40$, $\bar{x}_1 = 38$, $n_2 = 60$, $\bar{x}_2 = 42$

Here $\bar{x} = \dfrac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n_1 + n_2} = \dfrac{(40 \times 38) + (60 \times 42)}{40 + 60}$

$= \dfrac{1520 + 2520}{100} = \dfrac{4040}{100}$ = **40.4**

Note

The above property can be extended as follows. When there are three groups, the combined mean is given by

The algebraic sum of the squares of the observations from AM is always minimum. ie., is always minimum.

Merits and Demerits

Merits

The most widely used arithmetic mean has the following merits.

It is rigidly defined. Clear cut mathematical formulae are available.

It is based on all the items. The magnitudes of all the items are considered for its computation.

It lends itself for algebraic manipulations. Total of a set, Combined Mean etc., could be calculated.

It is simple to understand and is not difficult to calculate. Because of its practical use, provisions are made in calculators to find it.

It has sampling stability. It does not vary very much when samples are repeatedly taken from one and the same population.

It is very much useful in day-to-day activities, later chapters in Statistics and many disciplines of knowledge.

Many forms of the formula are available. The form appropriate and easy for the data on hand can be used.

Demerits

It is unduly affected by extreme items. One greatest item may pull up the mean of the set to such an extent that its representative character is questioned. For example, the mean mark is 35 for the 3 students whose individual marks are 0, 5 and 100.

Theoretically, it cannot be calculated for open-end data.

It cannot be found graphically.

It is not defined to deal with qualities.

Weighted Arithmetic Mean

In calculating simple arithmetic mean it was assumed that all items are of equal importance. This may not be true always. When items vary in importance they must be assigned weights in proportion to their relative importance. Thus, a weighted mean is the mean of weighted items. The weighted arithmetic mean is sum of the product of the values and their respective weights divided by the sum of the weights.

Symbolically, if $x_1, x_2, x_3, \ldots x_n$ are the values of items and $w_1, w_2, \ldots w_n$ are their respective weights, then

$$\text{WAM} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots + w_n x_n}{w_1 + w_2 + w_3 + \ldots + w_n} = \frac{\sum wx}{\sum w}$$

Weighted AM is preferred in computing the average of percentages, ratios or rates relating to different classes of a group of observations. Also WAM is invariably applied in the computation of birth and death rates and index numbers.

Example 7

A student obtains 60 marks in Statistics, 48 marks in Economics, 55 marks in law, 72 marks in Commerce and 45 marks in taxation in an examination. The weights of marks respectively are 2, 1, 3, 4, 2. Calculate the simple AM and weighted AM of the marks.

Solution

$$\text{Simple AM} = \frac{\sum x}{n} = \frac{60 + 48 + 55 + 72 + 45}{5} = \frac{280}{5} = \mathbf{56}$$

Marks (x)	Weights (w)	wx
60	2	120
48	1	48
55	3	165
72	4	288
45	2	90
	12	711

$$\text{WAM} = \frac{\sum wx}{\sum w} = \frac{711}{12} = \mathbf{59.25}$$

Geometric Mean

Geometric mean (GM) is the appropriate root (corresponding to the number of observations) of the product of observations. If there are n observations GM is the n-th root of the product of n observations.

Definition for a raw data

If $x_1, x_2, x_3, \ldots, x_n$ are n observations;

$$\text{GM} = \sqrt[n]{x_1, x_2, \ldots\ldots x_n}$$

Using logarithms, we can calculate GM using the formula,

$$\text{GM} = \text{Anti}\log\left(\frac{\sum \log x}{n}\right)$$

Definition for a frequency distribution

For a frequency distribution if $x_1, x_2, x_3, \ldots, x_n$ are n observations with the corresponding frequencies $f_1, f_2, \ldots, f_n$

$$\text{GM} = \sqrt[N]{x_1^{f_1}, x_2^{f_2}, \ldots\ldots x_n^{f_n}}$$

using logarithm,

$$\textbf{GM} = \text{Ant ilog}\left(\frac{\sum \mathbf{f \log x}}{\mathbf{N}}\right) \quad \text{where } N = \sum f.$$

Note

GM is the appropriate average for calculating index number and average rates of change.

GM can be calculated only for non zero and non negative values.

3. Weighted GM = $\text{Anti}\log\left(\frac{\sum w \log x}{\sum w}\right)$

where w's are the weights assigned.

Example 8

Calculate GM of 2, 4, 8

Solution

$$GM = \sqrt[n]{x_1, x_2, ...x_n} = \sqrt[3]{2 \times 4 \times 8} = \sqrt[3]{64} = \mathbf{4}$$

Example 9

Calculate GM of 4, 6, 9, 1 1 and 15

Solution

x	logx
4	0.6021
6	0.7782
9	0.9542
11	1.0414
15	1.1761
	4.5520

$$GM = Anti\log\left(\frac{\sum \log x}{n}\right)$$

$$= Anti\log\left(\frac{4.5520}{5}\right)$$

= Antilog0.9104

= **8.136**

Example 10

Calculate GM of the following data

Classes	:	1-3	4-6	7-9	10-12
Frequency	:	8	16	15	3

Solution

Classes	f	X	logx	f.logx
1-3	8	2	0.3010	2.4080
4-6	16	5	0.6990	11.1840
7-9	15	8	0.9031	13.5465
10-12	3	11	1.0414	3.1242
Total	42			30.2627

Merits and Demerits

Merits

It is rigidly defined. It has clear cut mathematical formula.

It is based on all the items. The magnitude of every item is considered for its computation.

It is not as unduly affected by extreme items as A.M. because it gives less weight to large items and more weight to small items.

It can be algebraically manipulated. The G.M. of the combined set can be calculated from the GMs and sizes of the sets.

It is useful in averaging ratios and percentages. It is suitable to find the average rate (not amount) of increase or decrease and to compute index numbers.

Demerits

It is neither simple to understand nor easy to calculate. Usage of logarithm makes the computation easy.

It has less sampling stability than the A.M.

It cannot be calculated for open-end data.

It cannot be found graphically.

It is not defined for qualities. Further, when one item is zero, it is zero and thereby loses its representative character. It cannot be calculated even if one value or one mid value is negative.