Unit - 2 B. Bouvick

Measures of Central Tendency (or) location.

2M

(A figure which is used to represent a whole series, possibly in the centre, where most of the items of the series cluster, is called Measures of central Tendency (or) Averages.)

It is also called Measures of Location."

Averages are statistical constants which help us to put in a single effort the significance of the whole.

There are five types of Averages.

1. 8th Arithmetic Mean
2. Median
3. Mode
4. Geometric Mean.
5. Harmonic Mean.

Objectives (or) Functions of an Average.

5th 1). Averages provide a (quick - understanding of Complex data. Complex data are reduced to a single understandable number by averages.

2. Average facilitates Sampling Techniques. Sampling Techniques are useful because of the role played by averages. The average of a sample tells what is the average of the population likely tobe.

3). <u>Averages enable comparison.</u>
Two or more sets of Values canbe Compared on the basis of their averages.

4) <u>Average gives the way for furthur Statistical Analysis.</u>
Dispersion, Skewness, Index Number And many other Statistical Measures depend on averages.

5) <u>Averages establish the relationship between Variables.</u>
Mathematical relationship such as the

<u>Criterion (or) Requisities or characteristics (or) Desirable properties of an average.</u>

5m

1. It should be clearly defined
2. It should be based on all the observations.
3. It should be simple to understand and easy to calculate
4. It should be suitable for further Mathematical treatment. i.e. if we are given the averages and sizes of two sets, It should be possible to find the average of the combined set.
5. It should be Least affected by flucations of sampling.
6. It should not be affected much by extreme Values.

Arithmetic Mean (1) Individual Series.

(Arithmetic Mean of a set of observati is their sum divided by the number of obsern

Indi Arithmetic Mean of 'n' observations is,

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$)

2. Discrete Distribution:

For the frequency distribution, x_i and f_i are given,

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum f_i x_i}{\sum f_i}$$

3. Grouped (or) Continuous Distribution

Class intervals and their frequencies are given

Here x_i - are the mid points of the classes

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} \rightarrow \text{Direct Method.}$$

~~Step~~-Deviation Method;

If $d_i = x_i - A$; where A is the Assumed Mean or arbitrary Value.

then $$\bar{x} = A + \left\{ \frac{\sum f_i d_i}{\sum f_i} \right\}$$

Step-deviation (or) Short Cut Method.

If $d_i = \frac{x_i - A}{c}$; where c is the width of the class

then $$\bar{x} = A + \left\{ \frac{\sum f_i d_i}{\sum f_i} \right\} \times c$$

Properties of Arithmetic Mean ($\bar{x}$)

1). The sum of the deviations of all values from their Arithmetic Mean is Zero.

For Individual Series; $\Sigma(x_i - \bar{x}) = 0$

$\Sigma(x_i - \bar{x}) = \Sigma x_i - n\bar{x}$

$= n\bar{x} - n\bar{x} = 0.$

$\left[\bar{x} = \frac{\Sigma x_i}{n} \Rightarrow n\bar{x} = \Sigma x_i\right]$

For a discrete distribution $\Sigma f_i(x_i - \bar{x}) = 0$.

Proof. $\Sigma f_i(x_i - \bar{x}) = \Sigma f_i x_i - \bar{x}.\Sigma f_i$

$= \bar{x}.\Sigma f_i - \bar{x}.\Sigma f_i$

$= 0.$

$\left[\because \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} \Rightarrow \bar{x}.\Sigma f_i\right]$

Property (2)

The sum of the squares of the deviations of the values from their Arithmetic Mean is minimum.

Proof: For a frequency distribution x_i/f_i;

Let $Z = \sum_{i=1}^{n} f_i(x_i - A)^2$ be the sum of squares of the deviations taken from any arbitrary point A.

4). If k is added to each Value, the Arithmetic Mean increases by k. If k is subtracted from each Value, the Arithmetic Mean decreases by k. If each and every value is multiplied by k, the mean becomes k times the original mea. If each value is divided by k the mean gets divided by k.

Merits of Arithmetic Mean.

1. It is rigidly defined. Mathematical formulae are available.
2. It is easy to understand and easy to Calcula
3. It is based on all the Observations.
4. It is least affected by fluctations of Sampli
5. It canbe used for furthur analysis and algebraic treatment.

If n_1, n_2 and $\bar{x}_1$ and $\bar{x}_2$ are given, then

$$\text{Combined Mean} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$$

Demerits

1. It cannot be Located by inspection graphicall
2. It cannot be used ~~by~~ for qualitative data
3. It cannot be calculated even if one Value is missing
4. It is affected very much by extreme Values
5. It cannot be calculated for open-end data

2nd Define Median.

(Median is the middle value of the variable when they are arranged in ascending or decending order of magnitude. It divides the whole distribution into two equal parts. It is also called positional average.)

For Individual Series. [x_i are given].

If 'n' is odd Median is the middle value after arranging them into ascending or decending order.

If 'n' is even, there are two middle terms. The average of the two middle terms is the Median.

Discrete Distribution:

x_i and f_i are given.

Median is the value of x which has the cumulative frequency just greater than $N/2$. N is the total frequency ($N = \Sigma f_i$)

Continuous frequency distribution.

$$\text{Median} = l + \left\{\frac{N/2 - m}{f}\right\} \times c$$

Where, l = Lower Limit of the Median Class.

f = frequency of the Median Class.

c = Width of the Median Class.

m = cumulative frequency of the class preceding the Median Class.

Median Class is the class which has the cumulative frequency Just greater than $\frac{N}{2}$

$N = \Sigma f_i$ = Total frequency

Merits:

1. It is clearly defined
2. It is easy to understand and easy to calcu calculate
3. It is not at all affected by extreme values. maximum
4. It canbe determined by inspection from a frequency curve.
5. It canbe found for distribution with open-end classes.

Demerits.

1. It is not used for Algebraic treatment
2. It is not based on all the observations.
3. It is insensitive. i.e. Any value canbe replaced by other without affecting the Median.
4. It canbe affected much by flucEations of Sampling
5. It canbe determined only after arranging the data into ascending or decending order.

MODE

Mode is the Value which occurs most frequently in a set of Values.

Discrete Distribution:

x_i and f_i are given. In discrete distribution Mode is the value of 'x' which has maximum frequency.

Mode by method of grouping.

Mode is determined by the method of grouping,

(i) If the maximum frequency is repeated.

(ii) If the maximum frequency occurs in the beginning (or) at the end of the distribution.

(iii) If there are irregularities in the distribution.

Continuous frequency distribution

$$\text{Mode} = l + \left\{\frac{f_m - f_1}{2f_m - f_1 - f_2}\right\} \times c$$

C = Width of the Modal class.

l = Lower limit of the Modal class.

f_m = frequency of the Modal Class

f_1 = frequency of the class preceding the Modal class

f_2 = frequency of the class succeeding the Modal class.

Modal class is the class which has the maximum frequency

Note:

1) For a symmetrical (or) Normal distribu
Mean, Median and Mode all are coincide.

Mean = Median = Mode.

11) If Mode is ill defined we can find it by the relation $\boxed{\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}}$ formulae

This is called empirical Mode.

Merits:

1) It is readily comprehensible and easy to calculate.
2) It canbe found by inspection.
3) It is not affected by extreme values.
4) It canbe calculated for distribution with open- end classes.

Demerits:

1. A clearly defined mode doesnot always exist. It is ill-defined.
2. It is not based on all the observations.
3. It is not capable of furthur Mathematical treatment
4. Mode is affected by fluctations of Sampling

Calculate Mode for the following distribution.

Size (x)	1	2	3	4	5	6	7	8	9	10	11	12
fre (f)	3	8	15	23	35	40	32	28	20	45	14	6

Grouping Table.

Size (x)	I	II	III	IV	V	VI
1	3	11		26		
2	8		23		46	
3	15	38				73
4	23		58	78		
5	35	75			107	
6	40		72			100
7	32	60		80		
8	28		48		93	
9	20	65				79
10	45		59	65		
11	14	20				
12	6					

I → Original frequencies are given in Column I

II → Frequencies are combined two by two

III → After leaving the first frequency, the other frequencies are combined two by

IV → The original frequencies are combined three by three.

V → After leaving the first frequency, the other frequencies are combined three by three

VI → After leaving the first two frequencies the other frequencies are combined three by three.

Next we form another table for the maximum frequency of each column and their respective sizes. The size (x) which occurs maximum number of times is the mode.

Analysis Table.

Columns	Maximum frequency	Size of items
I	45	10
II	75	5, 6
III	72	6, 7
IV	98	4, 5, 6
V	107	5, 6, 7
VI	100	6, 7, 8

Since the item 6 occurs 5 times.

∴ Mode is 6.