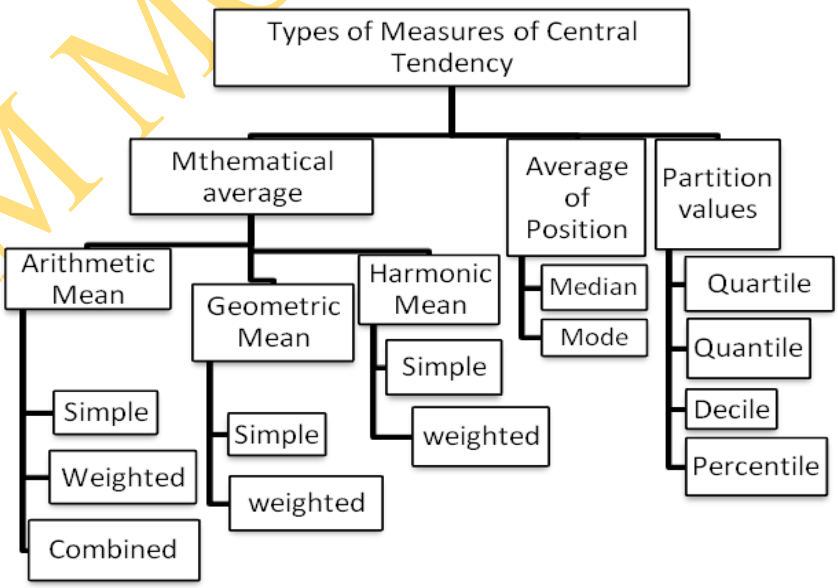
LECTURE 19

MEAN

PROF SHRUTI

Measures of Central Tendency:

It is also known as *measures of location* or *measures of averages*. These are the representatives of entire distribution. This measure is to be *ideal measure*, if it should be based on all observations, rigidly defined, easy to understand and easy to calculate, suitable for further mathematical treatment and not much affected by extreme values and fluctuations of sampling. Arithmetic mean, median, mode, geometric mean, harmonic mean and partition values are the popular measures of central tendency. Here arithmetic mean is the ideal measure of central tendency. The given diagram represents the types of measures of central tendency:



Definition

The word average or the term measures of central tendency have been defined by various authors in their own way. Some of the definitions are given below :—

Simpson and Kafka observe that "A measure of central tendency is a typical value around which other figures congregate."

Lawrence J. Kaplan has defined these terms in the following words :

"One of the most widely used set of summary figures is known as measures of location, which are often referred to as averages, central tendency or central location. The purpose for computing an average value for a set of observations is to obtain a single value which is representative of all the items and which the mind can grasp in simple manner quickly. The single value is the point of location around which the individual items cluster."

Va-Lun Chou states that "an average is a typical value in the sense that it is, sometimes, employed to represent all the individual values in a series or of a variable."

It is, thus, clear from the above definitions that an average is a single value which represents a whole series and is supposed to contain its major characteristics.

Objectives of Averaging

1. To get a single value which is representative of the characteristics of the entire mass of data : Averages give a bird's-eye view of the huge mass of statistical data which ordinarily are not easily intelligible.

They are devices to aid the human mind in grasping the true significance of large aggregates of facts and measurements. They set aside the unnecessary details of the data and put forward a concise picture of the complex phenomena under investigation because the human mind is not capable of grasping all the details of large numbers and their interrelationship. For example, it is not possible to keep in mind, the details of heights. weights. incomes and expenditures of even 200 students, what to talk of big figures. This difficulty of keeping all the details in mind necessitates the use of averages. An average is a single number representing the whole data and is useful in grasping the central theme of the data.

Why is an average a representative? The reason why an average is a valid representative of a series lies in the fact that ordinarily most of the items of a series cluster in the middle. On the extreme ends, the number of items is very little. In a population of 10,000 adults, there would hardly be any person who is 60 cms. high or whose height is above 240 cms. There will be a small range within which these values would vary, say, 150 cms. to 200 cms. Even within this range, a large number of persons would have a height between, say, 160 cms. and 180 cms. In other class intervals of height, the number of persons would be comparatively small. Under such circumstances if we conclude that the height of this particular group of persons would be represented by, say, 170 cms., we can reasonably be sure that this figure would, for all practical purposes, give us a satisfactory conclusion. This average would satisfactorily represent the whole group of figures from which it has been calculated.

2. To facilitate comparison : Since measures of central tendency or averages reduce the mass of statistical data to a single figure, they are very helpful for purposes of making comparative studies, for example, the average marks obtained by two sections of a class would give a reasonably clear picture about the level of their performance, which would not be possible if we had two full series of marks of individual students of the two sections.

However, when such a comparison is made, we have to be careful in drawing inferences, as the marks of students in one section may vary within a small range and in the other section some students may have got very high marks and others very few marks. The comparison of averages in such a case may give misleading conclusions .

Characteristics of A Good Average

1. It should be rigidly defined: If an average is left to the estimation of an observer and if it is not a definite and fixed value, it cannot be representative of a series. The bias of the investigator in such cases would considerably affect the value of the average. If the average is rigidly defined, this instability in its value would be no more, and it would always be a definite figure.

2. It should be based on all the observations of the series: If some of the items of the series are not taken into account in its calculation, the average cannot be said to be a representative one.

3. It should be capable of further algebraic treatment: If an average dose not possess this quality, its use is bound to be very limited. It will not be possible to calculate, say, the combined average of two or more series from their individual averages; further, it will not be possible to study the average relationship of various parts of a variable if it is expressed as the sum of two or more variables etc.

4. If should be easy to calculate and simple to follow: If the calculation of the average involves tedious mathematical processes, it will not be readily understood by a person of ordinary intelligence and its use will be confined only to a limited number of persons and, hence, can never be a popular measure. As such, one of the qualities of a good average is that it should not be too abstract or mathematical and should be easy to calculate.

5. It should not be affected by fluctuations of sampling: If two independent sample studies are made in any particular field, the averages thus obtained, should not materially differ from each other. No doubt, when two separate enquiries are made, there is bound to be a difference in the average values calculated but, in some cases, this difference would be great while in others comparatively less. These averages in which this difference. which is technically called "fluctuation of sampling", is less, are considered better than those in which its difference is more.

Types of Averages

Measures of central tendency or averages are usually of the following types :

1. **Mathematical Averages:**
   - (a) Arithmetic Average or Mean
   - (b) Geometric Mean
   - (c) Harmonic Mean

2. **Averages of Position (Positional averages) :**
   - (a) Median
   - (b) Mode

Of the above mentioned five important averages, Arithmetic Average, Median and Mode are the most popular ones. Geometric mean and Harmonic mean come next. We shall study them in this very order.

Arithmetic Average

Arithmetic Average or Mean of a series is the figure obtained by dividing the sum of the values of the various items by their number. If the heights of a group of eleven persons are 164, 169, 163, 160, 165, 168, 162, 167, 170, 166, 161 centimeters, then to find the arithmetic average of the heights of these persons we shall add these figures and divide the total so obtained, by the number of items which is 11. The total of the items in this case is 1815[1] cms. and if it IS divided by 11, we get the figure of 165 cms. This is the mean or arithmetic average of the series.

Calculation of the arithmetic average in a series of individual observations

$$\text{A.M.} = \frac{\text{Sum of the values}}{\text{Number of the values}} \Rightarrow \text{(AM) Number of values} = \text{Sum of the values}$$

Suppose the values of a variable are respectively X_1, X_2, X_3, ... X_n, and their arithmetic average is represented by $\overline{X}$, then

$$\overline{X} = \frac{1}{N}\left(X_1 + X_2 + X_3 \ldots\ldots X_n\right)$$

or $$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_1$$ or $$\overline{X} = \frac{\sum_{i=1}^{N} X_1}{N}$$

where, $\overline{X}$ = Arithmetic average; X_1's = values of the variable; Σ = Summation or total; N = Number of items.

The following example would illustrate this formula.

Example 1. Calculate the simple arithmetic average of the following items :

Size of items		
20	50	72
28	53	74

[1] 164 + 169 + 165 + 160 + 165 + 168 + 162 + 167 + 170 + 166 + 161 = 1815

34	54	75
39	59	78
42	64	79

Solution. Direct Method :

Computation of arithmetic average Size of items

X : 20, 28, 34, 39, 42, 50, 53, 54, 59, 64, 72, 74, 75, 78, 79

ΣX = 20+28+34+39+42+50+53+54+59+64+72+74+75+78+79 = 821

Arithmetic mean = $\frac{\sum X}{N} = \frac{821}{15} = 54.73$

Calculation of arithmetic average in a discrete series

Direct Method: In a discrete series, the values of the variable are multiplied by their respective frequencies and the products so obtained are totalled. This total is divided by the number of items which, in a discrete series, is equal to the total of the frequencies. The resulting quotient is a simple arithmetic average of the series.

Algebraically,

If, f_1, f_2, f_3, etc., stand respectively for the frequencies of the values X_1, X_2, X_3, etc.

$$\overline{X} = \frac{1}{N}\left(X_1 f_1 + X_2 f_2 + X_3 f_3 + \ldots\ldots + X_n f_n\right)$$

or $\overline{X} = \frac{\sum fX}{N}$ or $\frac{\sum fX}{\sum f}$

Short-cut method I : A short-cut method can be used in the discrete series also. In this method, the deviations of the items from an assumed mean are first found out and they are multiplied by their respective frequencies. The total of these products is divided by the total frequencies and added to the assumed mean. The resulting figure is the actual arithmetic average.

Algebraically : $$\overline{X} = A + \frac{\sum fdx}{N}$$

where, Σfdx = the total of the products of the deviations from the assumed average and the respective frequencies of the items.

Step deviation method

In step deviation method, we define $d'x = \frac{dx}{C}$, where C is some common factor in dx values and then apply the formula :

$$\overline{X} = A + \frac{\sum f\left(\frac{dx}{C}\right)}{N} \times C.$$ This is called **short-cut method II.**

Example 2. The following table gives the marks obtained by a set of students in a certain examination. Calculate the average mark per student.

Marks	Number of students	Marks	Number of students
10—20	1	60—70	12
20—30	2	70—80	16
30—40	3	80—90	10
40—50	5	90—100	4
50—60	7		

Solution. **Short-cut Method**

Computation of average marks per student

Marks (X)	Mid values (m.v.)	No. of students (f)	Deviation from assumed mean	Step deviations (10) (dx)	Total deviation (fdx)

			(55)		
10–20	15	1	– 40	– 4	– 4
20—30	25	2	– 30	– 3	– 6
30—40	35	3	– 20	– 2	– 6
40—50	45	5	– 10	– 1	– 5
50—60	55	7	– 0	0	0
60—70	65	12	+ 10	+ 1	+ 12
70—80	75	16	+ 20	+ 2	+ 32
80—90	85	10	+ 30	+ 3	+ 30
90—100	95	4	+ 40	+ 4	+ 16
		N = 60			Σfdx = + 69

Arithmetic average or $\overline{X} = A + \left(\frac{\sum fdx}{N} \times i\right) = 55 + \left(\frac{69}{60} \times 10\right)$ = 66.5 marks.

Merits of arithmetic average

The arithmetic average is the most popularly used measure of central tendency. There are many reasons for its popularity. In the beginning of this chapter, we had laid down certain characteristics which an ideal average should possess. We shall now see how far the arithmetic average fulfils these conditions:

1. The first condition that an average should be rigidly defined is fulfilled by the arithmetic average. It is rigidly defined and a biased investigator shall get the same arithmetic average from the series as an unbiased one. Its value is always definite.

2. The second characteristic that an average should be based on all the observations of a series is also found in this average. Arithmetic average cannot be calculated even if a single item of a series is left out.

3. Arithmetic average is also capable of further algebraic treatment. While discussing the algebraic properties of the arithmetic average, we have already seen in detail, how various mathematical processes can be applied to it for purposes of further analysis and interpretation of data. It is on account of this characteristic of the arithmetic average that:

   (a) It is possible to find the aggregate of items of a series if only its arithmetic average and the number of items is known.

   (b) It is possible to find the arithmetic average if only the aggregate of items and their number is known.

4. The fourth characteristic laid down for an ideal average that it should be easy to calculate and simple to follow, is also found in arithmetic average. The calculation of the arithmetic average is simple and it is very easily understandable. It does not require the arraying of data which is necessary in case of some other averages. In fact, this average is so well known that to a common mean average means an arithmetic average,

   Thus, the arithmetic average

   (a) is simple to calculate,

   (b) does not need arraying of data,

   (c) is easy to understand.

5. The last characteristic of an ideal average that it should be least affected by fluctuations of sampling is also present in arithmetic average to a certain extent. If the number of items in a series is large, the arithmetic average provides a good basis of comparison, as in such cases, the abnormalities in one direction are set off against the abnormalities in the other direction.

Summary:

Arithmetic Mean (A.M.):

It is also known as mean or average. This is most often utilized measure. Mathematically it is denoted as $\bar{x}$.

- For ungrouped data, If $x_1, x_2, \ldots.., x_n$ be n values of any variable X, then mean is,

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 \ldots \ldots \ldots \ldots \ldots \ldots + x_n}{n}$$

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

- For grouped data or in case of frequency distribution x_i/f_i, if $x_1, x_2, \ldots.., x_n$ be n values of any variable X with relating frequencies $f_1, f_2, \ldots.., f_n$, then mean is,

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + f_4x_4 \ldots \ldots \ldots \ldots \ldots \ldots + f_ix_n}{f_1 + f_2 + f_3 + f_4 + \cdots \ldots \ldots \ldots \ldots + f_n}$$

$$\bar{x} = \frac{1}{N} \sum_{1}^{n} f_ix_i$$

Where, $N = f_1 + f_2 + f_3 + f_4 + \cdots \ldots \ldots \ldots \ldots + f_n$

$$N = \sum_{1}^{n} f_i$$

- In case of class intervals suppose $m_1, m_2, \ldots.., m_n$ be the mid values of corresponding class interval, then the mean is,

$$\bar{x} = \frac{f_1m_1 + f_2m_2 + f_3m_3 + f_4m_4 \ldots \ldots \ldots \ldots \ldots \ldots + f_im_n}{f_1 + f_2 + f_3 + f_4 + \cdots \ldots \ldots \ldots \ldots + f_n}$$

$$\bar{x} = \frac{1}{N} \sum_{1}^{n} f_im_i$$

Where midpoint m is,

$$m = \frac{Lower\ limit\ of\ CI + Upper\ limit\ of\ CI}{2}$$

➢ If the value of x (or m) is large, then calculations from above formula *(direct method)* are difficult. In such situation use *shortcut method*, in this method shift the origin and change the scale. From the shortcut method the mean is,

$$\bar{x} = a + \frac{h}{N}\sum_{1}^{n} f_i d_i$$

Where

$$d_i = \frac{x_i - a}{h}$$

a is an arbitrary point and *h* is scale (common magnitude) of *CI*.

Hence, it shows the mean is not independent of change of origin and scale.

➢ If $\bar{x}_1$ and $\bar{x}_2$ be the mean of two series with n_1 and n_2 total number of observations, then the size of whole series is $n = n_1 + n_2$, and ***combined mean*** is

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

Advantages of Mean:

(i) The mean is sensitive to the exact values of all the scores in the distribution. Since you have to add the scores to calculate the mean, a change in any of the scores will cause a change in the mean.

(ii) Mean is very sensitive to extreme scores. If we add an extreme score (one that is very far from the mean), it would greatly disrupt the balance. The mean would have to shift a considerable distance to reestablish balance. The mean is more sensitive to extreme than is the median or the mode. This known as 2^{nd} property of mean.

(iii) Of the measures used for central tendency, the mean is least subject to sampling variation under most circumstances. If repeated samples are drawn from a population, the mean would vary from sample to sample. The same is true for the median and the mode. However the mean varies less than these other measures of central tendency. This is very important in inferential statistics and is a major reason why the mean is use in inferential statistics whenever possible.

(iv) It takes into account all the scores in a distribution so; mean offers a good representation of the central tendency by making use of the most information.

(v) Mean is used in many statistical formulas, making it a more widely used measure.

Limitations of Mean: Mean can be misleading if there are extreme values in the distribution, for example, if the distribution is skewed (asymmetrical) of the level of measurement is less than interval. Sometimes people are interested in misleading others by making use of 'illegitimate' statistics. The following example illustrates this point.