UNIT-I: MEASURES OF CENTRAL TENDENCY

1. The Arithmetic Mean ,
2. The Weighted Arithmetic Mean ,
3. Properties of the Arithmetic Mean ,
4. The Arithmetic Mean Computed from Grouped Data ,
5. The Median ,
6. The Mode,
7. The Empirical Relation Between the Mean, Median, and Mode,
8. The Geometric Mean G,
9. The Harmonic Mean H ,
10. The Relation Between the Arithmetic, Geometric, and Harmonic Means,
11. The Root Mean Square

INTRODUCTION:

- One number that represents a data set and gives an idea of the middle quantity of the data set.
- According to Croxton and Cowden "An average is a single value within the range of the data that is used to represent all the values in the series".
- Since an average is somewhere within the range of data, it is sometimes called a **measure of central value.**
- Since all these values in some way or the other represent the central values or average values of the data, they are referred to as Measures of Central Tendencies.
- There are three types of measures of central tendencies or Average and they are mean**, median and mode.**

1. The Arithmetic Mean:

- An average is a single value within the range of data that is used to represent all of the values in the series.
- The mean is the most popular and widely used.
- It is sometimes called **Arithmetic mean.**
- "Arithmetic mean" or simply "mean" of a given data is defined as the sum of all the values of the data divided by the total number of values.

- Mean is most useful when the data has no extreme values. Extreme values distort the value of mean due to which it is not able to represent the data correctly.

a. The **mean** of *n* data items $x_1, x_2, \ldots, x_n$, is given by the formula $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

The **mean** of *n* data items $x_1, x_2, \ldots, x_n$ whose corresponding frequencies are $f_1, f_2, \ldots, f_n$ is given by the formula $c = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}$

b. Step deviation method:

If in a frequency table the class intervals have equal width, say h than the A.M. by step deviation method is given by $\overline{\mathbf{X}} = \mathbf{A} + \mathbf{h} * \frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i}$

Where **u=(x-A)/h** ,and h is length of the interval, A is the assumed value.

c. Weighted mean

The different weight are assigned to different observations according to their relative importance and then average is calculated by considering weight as well.

This average is called Weighed Arithmetic Mean or Simple weighted mean, denoted by $\overline{x_w}$

$$\overline{x_w} = \frac{\sum_i w_i x_i}{\sum_i w_i}$$

d. Combined mean:

For k sub group of data consisting of $n_1, n_2, \ldots, n_k$ observation having respective means $\overline{x_1}, \overline{x_2}, \overline{x_3}, \ldots, \overline{x_k}$ then combined mean is given by $\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2 + n_3\overline{x}_3 + \cdots + n_k\overline{x}_k}{n_1 + n_2 + n_3 + \cdots + n_k}$

2. Median :

- The median is defined as the measure of the central term, when the given terms (i.e. values of the variate) are arranged in the ascending or descending order of magnitudes.
- The median is the value of the variable which divides the group into two equal parts one part comprising all values greater, and the other all values less than the median‖.
- **Case 1.** If n is **odd** then value of $((n+1)/2)^{th}$ term gives the median.
- **Case2**. If n is **even** then there are two central terms i.e., $(n/2)^{th}$ and $((n+1)/2)^{th}$, The mean of these two values gives the median.

- Median in continuous series **(or grouped series).** $Median = l_1 + \frac{\left(\frac{N}{2} - c.f.\right)*h}{f}$

  l_1 : lower class limit of Medial class interval

  c.f. : preceding cumulative frequency of less than type of Medial class interval

  h : width

  f : frequency of Medial class interval

3. MODE:

The value of the variable which occurs most frequently in the distribution is called the mode.'

Mode = $l_1 + \frac{(f_m - f_0)*h}{2f_m - f_0 - f_1}$

Where, l_1 : lower class limit of Modal class interval

f_m : Maximum frequency

f_0 : Preceding frequency of modal class interval

f_1 : succeeding frequency of modal class interval

h : width

4. EMPIRICAL RELATION BETWEEN MEDAIN AND MODE

For moderately asymmetrical distribution (or for asymmetrical curve), the relation

Mean – Mode = 3 (Mean - Median),

Approximately holds. In such a case, first evaluate mean and median and then mode is determined by

Mode = 3 Median – 2 Mean.

If in the asymmetrical curve the area on the left of mode is greater than area on the right then

Mean < median < mode, i. e., (M < Md < M0)

• Merits and demerits of mean, median and mode

MERITS OF A.M.	DEMERITS OF A.M.
1. The mean summarizes all the information in the data	1. It can neither be determined by inspection or by graphical location

2. It is the average of the all observation	2. Arithmetic mean cannot be computed for qualitative data like data on intelligence honesty and smoking habit etc.
3. It is the single point that can be viewed as the point where all the weight of the observation is concentrated	3. It is too much affected by extreme observations and hence it is not adequately represent data consisting of some extreme point
4. It is the center of mass of the data	4. Arithmetic mean cannot be computed when class intervals have open ends

MERITS OF MEDIAN	DEMERITS OF MEDIAN
1) There is a unique median for each data set.	1) Lack of representative character Median fails to be a representative measure in case of such series the different values of which are wide apart from each other. Also, median is of limited representative character as it is not based on all the items in the series.
2) It is not affected by extremely large or small values and is therefore a valuable measure of central tendency when such values occur.	2) Unrealistic:- When the median is located somewhere between the two middle values, it remains only an approximate measure, not a precise value.
3) It can be computed for ratio-level, interval-level, and ordinal-level data.	3) Lack of algebraic treatment: - Arithmetic mean is capable of further algebraic treatment, but median is not. For example, multiplying the median with the number of items in the series will not give us the sum total of the values of the series.
4) It can be computed for an open-ended frequency distribution if the median does not lie in an open-ended class.	

Merits of Mode:	Demerits Of Mode:
(1) Simple and popular: - Mode is very simple measure of central tendency. Sometimes, just at the series is enough to locate the model value. Because of its simplicity, it s a very popular measure of the central tendency. (2) Less effect of marginal values: - Compared top mean, mode is less affected by marginal values in the series. Mode is determined only by the value with highest frequencies. (3) Graphic presentation:- Mode can be located graphically, with the help of histogram. (4) Best representative: - Mode is that value which occurs most frequently in the series. Accordingly, mode is the best representative value of the series. (5) No need of knowing all the items or frequencies: - The calculation of mode does not require knowledge of all the items and frequencies of a distribution. In simple series, it is enough if one knows the items with highest frequencies in the distribution.	Following are the various demerits of mode: (1) Uncertain and vague: - Mode is an uncertain and vague measure of the central tendency. (2) Not capable of algebraic treatment: - Unlike mean, mode is not capable of further algebraic treatment. (3) Difficult: - With frequencies of all items are identical, it is difficult to identify the modal value. (4) Complex procedure of grouping:- Calculation of mode involves cumbersome procedure of grouping the data. If the extent of grouping changes there will be a change in the model value. (5) Ignores extreme marginal frequencies:- It ignores extreme marginal frequencies. To that extent model value is not a representative value of all the items in a series.

5. HARMONIC MEAN

The Harmonic mean of a series of values is the reciprocal of the arithmetic means of their reciprocals. Thus if x1,x2,…, xn (none of them being zero) is a series and H is its harmonic mean then

$$\frac{1}{H} = \frac{1}{N}\left(\frac{1}{x1} + \frac{1}{x2} + \cdots + \frac{1}{xn}\right)$$

If f_1, f_2, …, fn be the frequencies of x_1, x_2, … , x_n (none of them being zero) then harmonic mean H is given by H.M.= $= \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i \left(\frac{1}{x_i}\right)}$

6. GEOMETRIC MEAN

If x_1, x_2, … ,xn. are n values of the variate x, none of which is zero . Then their geometric mean G is defined by $G = (x_1, x_2, \ldots x_n)^{1/n}$

If f_1, f_2, … , fn are the frequencies of $x_1, x_2, \ldots, x_n$ respectively, then geometric mean G is given by

$G = (x_1^{f1} * x_2^{f2} \ldots * x_n^{fn})^{1/N}$

taking Log on both the side

Log (G) = 1/N ($f_1 * \log(x_1) + f_2 * \log(x_2) + \ldots + fn * \log(x_n)$)

Log (G) = 1/N ($\sum_{i=1}^{n} f_i * log(x_i)$)

7. The Relation Between the Arithmetic, Geometric, and Harmonic Means

If A.M. denotes Arithmetic mean, G.M. geometric mean and H.M. harmonic means then their relation is given by

A.M. * G.M. = H.M.^{2} **OR** **H.M. = $\sqrt{A.M. * G.M.}$**

$A.M. \geq G.M. \geq H.M.$

8. Root mean square:

The root mean square or quadratic mean of set a set of numbers $x_1, x_2, \ldots, x_n$ is sometimes denoted by $\sqrt{\overline{X^2}}$ and is defined as RMS = $\sqrt{\overline{X^2}}$