Introduction

BASIC COMPONENTS

- **Frequency:-** The number of times that a variate takes places is known as its frequency.
- **Class Limits:-** The class limits are the lowest and the highest values that can be included in the class.
- **Class Interval or Height of A Class:-** The difference between upper limit and lower limit is known as class interval, denoted by i or h **.**
- **Class Marks or Mid Point:-** It is the value exactly at the middle of the lower and upper limits of a class.

$$\text{Class marks} = \frac{\text{Upper limit} + \text{Lower limit}}{2}$$

MEASURE OF CENTRAL TENDENCY

One of the most important objectives of statistical analysis is to get one single value that describes the characteristic of entire data, such a value is called the central value.
The values of variables around which other values can be describes is known as measure of central tendency.

There are five types of measures of central tendency
- ✓ Arithmetic mean
- ✓ Geometric mean
- ✓ Harmonic mean
- ✓ Median and Quartile
- ✓ Mode

ARITHMETIC MEAN

The arithmetic mean can be obtained by adding the values and dividing by number of values.

$$\text{A. M.} = \frac{\text{Sum of observations}}{\text{Number of observations}}$$

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

There are three methods to find out mean
- ✓ Direct method
- ✓ Assumed mean method
- ✓ shortcut method (Step deviation method)

Aims Tutorial
Oriental Bank of Commerce
H.No.:69/339 VT Road Ward 27,
Mansarovar Jaipur (Rajasthan) 302020
Contact No: +91-8947920041

Direct Method:-

$$\bar{x} = \frac{\sum f x}{\sum f}$$

$Where\ \bar{x} = mean$

$\sum f x = Sum\ of\ multiplication\ of\ \ variate\ and\ corresponding\ frequency$

$\sum f\ = Sum\ of\ \ frequencies$

Merit:

- ✓ It is the simplest average to understand and easiest to complete.
- ✓ It is affected by the value of every item in the series.
- ✓ It is defined by a rigid mathematical formula with the result that everyone who compute the average gets the same answer.
- ✓ It is a calculated value and not based on position in the series.

Limitations /Demerit:

- ✓ It is tedious of compute for a large data set as every point in the data set is to be used in computations.
- ✓ It cannot be calculated for qualitative characteristics, such as beauty or intelligence.
- ✓ The arithmetic mean is not always a good measure of central tendency. The mean provide a 'characteristic ' value, in the sense of indicating where most of the value lies only when the distribution of the variable is reasonably normal (Bell shaped). In case of U – Shaped distribution the mean is not likely to serve a useful purpose.

MEDIAN

A median by definition refers to the middle value in a distribution. Median is that value of a variable which divides the series in such a manner that the number of item below it is equal to the number of items above it.
In other words, half the total numbers of observation lie below the median and half above it.

$$M = \left(\frac{n+1}{2}\right)^{th} \text{ will be median.}$$

If the number of items are even then

Aims Tutorial
Oriental Bank of Commerce
H.No.:69/339 VT Road Ward 27,
Mansarovar Jaipur (Rajasthan) 302020
Contact No: +91-8947920041

$$M = \frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2} \text{ will be median.}$$

Merit:

- ✓ It is easy to calculate and readily understood.
- ✓ Extreme values do not affect the median as strong as they do the mean.
- ✓ Where the arithmetic mean would be distorted by extreme values, the median is especially useful.
- ✓ The value of median can be graphically find out.
- ✓ Median is usually an actual figure of the series. In case of individual observation with odd numbers and discrete variable.

Limitations /Demerit:

- ✓ For calculating median it is necessary to arrange the data.
- ✓ The value of median is affected more by sampling fluctuations than the value of the arithmetic mean.
- ✓ If median and number of items of distribution are given, then total value cannot be known.
- ✓ Median cannot be used for determining the combined median of two or more groups as is possible in case of mean.

MODE

A value which occurs most frequently is known as mode.

E.g. In the series 6, 5, 3, 4, 3, 7, 8, 5, 9, 5, 4, we notice that 5 occurs most frequently. Therefore, 5 is the mode.

Mode of Grouped Data:

$$Median = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Where l = lower limit of modal class.
f_1 = frequency of modal class.
f_0 = frequency of the class preceding the modal class.
f_2 = frequency of the class succeeding the modal class.
i = class interval.

Merit:

- ✓ Mode is the most typical or representative value of a distribution.
- ✓ It is not affected by extreme values.
- ✓ The value of mode can also be determined graphically.
- ✓ Mode is easy to calculate and easy to understand.
- ✓ Mode having more practical use as compared to mean or median.

Demerits:

- ✓ It is not capable of algebraic manipulation.
- ✓ The value of mode is not based on each and every item of the series.
- ✓ The value of mode cannot always be determined.

Empirical Relationship among mean, median and mode:-

$$Mode = 3\ Median - 2\ Mean$$

GEOMETRIC MEAN

Let $x_1, x_2, x_3, \dots\dots x_n$ be n observations of a variable then its geometric mean is given by

$$G = (x_1 . x_2 . x_3 \dots\dots . x_n)^{\frac{1}{n}}$$

$$\log G = \frac{1}{n}\left[log x_1 + log x_2 + \cdots log x_n\right]$$

$$G = antilog\left[\frac{1}{n}[log x_1 + log x_2 + \cdots log x_n]\right]$$

Drawback: If any of the observation is zero or negative then G can not be calculated.

Harmonic mean: Let $x_1, x_2, x_3, \dots\dots x_n$ be n observation then its harmonic mean is given by

$$H = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots \frac{1}{x_n}\right)}$$

Requisites of Good Measure of Central Tendency (Properties of Good Measure of Central Tendency):

- ✓ It should be rigidly define.
- ✓ It should be based on the all observation.
- ✓ It should be easy to understand and easy to calculate.
- ✓ It should be suitable for further mathematical or algebraic treatments.
- ✓ It should not be much affected by extreme value.
- ✓ It should not be much affected by sampling fluctuations.

Demerit of Measures of Central Tendency:

- ✓ Arithmetic mean is rigidly defined but it is very much affected by extreme values.
- ✓ Median is not suitable for further algebraic treatments.
- ✓ It is also very much affected by sampling fluctuations.
- ✓ Mode is well defined and is not based upon all the observations.
- ✓ It is not capable of further mathematical treatment.
- ✓ G.M. and H.M. are very complicated.
- ✓ There are not easy to understand and calculate.

MEASURE OF DISPERSION

Measure of central tendency gives only a central value on this basis we can not determine about scatteredness, range of value etc. So only on the basis of measure of central tendency the analysis and interpretation of data insufficient.
A measure of dispersion be defined as the measure of the extent of the scatterdness of items around a measure of central tendency.

In other words Dispersion is a measure of the variation of the items.

Characteristics of Good Measure of Dispersion:-

- ✓ It should be based on all observations.
- ✓ It should be easily calculated.
- ✓ It should be affected a little as possible by fluctuations of sampling.
- ✓ It should be capable of further mathematical treatment.
- ✓ It should not be affected by the extreme values.

The following are the common measures of dispersion.

- ✓ The range
- ✓ The semi – interquartile range or quartile deviation.
- ✓ The mean deviation
- ✓ The standard deviation

RANGE

It is the simplest measure of dispersion.
It is given by the difference between the highest and the lowest observations.
Range = Highest value – Lowest value = $H - L$

$$\text{Coefficient of Range } = \frac{H - L}{H + L}$$

Drawback: Although the range is simplest measure of dispersion but as it depends on only two observations.
So it is not measure of dispersion.

QUARTILE DEVIATION

Let Q_1 and Q_3 be the first and third quartile then quartile deviation is given by:-

$$Quartile\ deviation = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient of } Quartile\ deviation = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Merit: Since the Q.D. depends on half of the observation so it is better measure than range.

Demerit: But is not depending on the remaining on the half of observation so it also not a good measure of dispersion.

MEAN DEVIATIONS

Let $x_1, x_2, x_3, \dots\dots x_n$ be n variable with frequency $f_1, f_2, f_3, \dots\dots f_n$ then the mean deviation about any point "*a*" is given by

$$Mean\ deviation = \frac{1}{N}\sum_{i=1}^{n} f_i|x_i - a|$$

Where "*a*" may be arithmetic mean, G.M, H.M., Median value or some other general point between the observations.

Merit: As mean deviation depends on all the observations so it is better measure than range and quartile deviation.

Demerit: But it does not consider the negative sign so it is also not a good measured of d dispersion.

ROOT MEAN SQUARE DEVIATION

Let $x_1, x_2, x_3, \dots\dots x_n$ be n variable with frequency $f_1, f_2, f_3, \dots\dots f_n$ then the root mean square deviation is given by the formula:-

$$Root\ mean\ square\ deviation = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - a)^2}$$

STANDARD DEVIATION (σ_x)

If the root mean square deviation is calculated from the actual mean then it is called standard deviation.
It is denoted by (σ_x).

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2}$$

COMBINING MEAN AND STANDARD DEVIATIONS OF TWO DISTRIBUTIONS

Combined mean:

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2}$$

Combined standard deviation:

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \overline{x_1})^2 + N_2(\bar{x} - \overline{x_2})^2}{N_1 + N_2}}$$

VARIANCE

The square of standard deviation is called "Variance". It is denoted by ${\sigma_x}^2$.

$${\sigma_x}^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$$

$$Coefficient\ of\ variance = 100 \times \frac{s.d.}{mean}$$