# Measure of Central Tendency and dispersion

# **1.** Measures of Central Tendency

It is that single value which may be taken as the most suitable representative of the data. This single value is known as the average. Averages are generally, the central part of the distribution and therefore, they are also called the measures of **Central Tendency**.

It can be divided into two groups :

## (a) MATHEMATICAL AVERAGE :

- i. Arithmetic mean or mean
- ii. Geometric mean
- iii. Harmonic mean

### (b) POSITIONAL AVERAGE :

- i. Median
- ii. Mode or positional average

## 2. Arithmetic Mean

#### 2.1 Individual observation or unclassified data :

If  $x_1, x_2,...,x_n$  be n observations, then their arithmetic mean is given by

$$\overline{\mathbf{x}} = \frac{\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n}{n}$$
 or  $\overline{\mathbf{x}} = \frac{\sum_{i=1}^{n} \mathbf{x}_i}{n}$ 

#### 2.2 Arithmetic mean of discrete frequency distribution :

Let  $x_1, x_2, \dots, x_n$  be n observation and let  $f_1, f_2, \dots, f_n$  be their corresponding frequencies, then their mean

$$\overline{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots + f_n} \text{ or } \overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

2.2.1

**Short cut method :** If the values of x or (and) f are large the calculation of arithmetic mean by the previous formula used, is quite tedious and time consuming. In such case we take the deviation from an arbitrary point A.

$$\overline{\mathbf{x}} = \mathbf{A} + \frac{\Sigma \mathbf{f}_i \mathbf{d}_i}{\Sigma \mathbf{f}_i}$$

where A = Assumed mean

$$d_i = x_i - A =$$
 deviation for each term

2.2.2

**Step deviation method :** Sometimes during the application of shortcut method of finding the mean, the deviation  $d_i$  are divisible by a common number h (say). In such case the arithmetic is reduced to a great extent taken by

u<sub>i</sub> = 
$$\frac{x_i - A}{h}$$
, i = 1, 2,.....n  
∴ mean  $\overline{x} = A + h\left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right)$ 

#### 2.3 Weighted arithmetic mean :

If  $w_1$ ,  $w_2$ ,  $w_3$ ,...., $w_n$  are the weight assigned to the values  $x_1$ ,  $x_2$ ,  $x_3$ ,...., $x_n$  respectively, then the weighted average is defined as -

Weighted A.M. = 
$$\frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n}$$
  
 $\overline{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$ 

#### 2.4 Combined mean :

If  $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$  are the mean of k series of sizes  $n_1, n_2, ..., n_k$  respectively then the mean of the composite series is given by

$$\overline{\mathbf{x}} = \frac{\mathbf{n}_1 \overline{\mathbf{x}}_1 + \mathbf{n}_2 \overline{\mathbf{x}}_2 + \dots + \mathbf{n}_k \overline{\mathbf{x}}_k}{\mathbf{n}_1 + \mathbf{n}_2 + \dots + \mathbf{n}_k}$$

#### 2.5 Properties of Arithmetic Mean :

**1.** In a statistical data, the sum of the deviation of items from A.M. is always zero.

i.e. 
$$\sum_{i=1}^{n} (x_i - \overline{x}) = 0$$
  
or 
$$\sum_{i=1}^{n} x_i - n\overline{x}$$
  
or 
$$n\overline{x} - n\overline{x} = 0$$
$$\left(\Theta \ \overline{x} = \frac{\sum x_i}{n}\right)$$

2. In a statistical data, the sum of squares of the deviation of items from A.M. is least i.e.  $\sum_{n=1}^{n} (1 - n)^{2} + 1 = 1$ 

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 \text{ is least}$$

- **3.** If each of the n given observation be doubled, then their mean is doubled.
- If x̄ is the mean of x<sub>1</sub>, x<sub>2</sub>,....,x<sub>n</sub>. The mean of ax<sub>1</sub>, ax<sub>2</sub>,....,ax<sub>n</sub> is a x̄ where a is any number different from zero.
- 5. Arithmetic mean is independent of origin i.e. it is not effected by any change in origin.

Power by: VISIONet Info Solution Pvt. Ltd		
Website : www.edubull.com	Mob no. : +91-9350679141	

# 2.6 MERITS AND DEMERITS OF ARITHMETIC MEAN:

## 2.6.1 Merits of Arithmetic Mean :

- i. It is rigidly defined.
- **ii.** It is based on all the observation taken.
- iii. It is calculated with reasonable ease and rapidity.
- iv. It is least affected by fluctuations in sampling.
- v. It is based on each observation and so it is a better representative of the data.
- vi. It is relatively reliable.
- vii. Mathematical analysis of mean is possible.

### 2.6.2 Demerits of Arithmetic Mean :

- i. It is severely affected by the extreme values.
- **ii.** It cannot be represented in the actual data since the mean does not coincide with any of the observed value.
- iii. It cannot be computed unless all the items are known.
- **iv.** It cannot be calculated for qualitative data incapable of numerical measurements.
- v. It cannot be used in the study of ratios, rates etc.

# 3. Geometric Mean

**3.1 Individual data :** If x<sub>1</sub>, x<sub>2</sub>, x<sub>3</sub>,....., x<sub>n</sub> are n values of a variate x, none of them being zero, then the geometric mean G is defined as-

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

or 
$$G = antilog\left(\frac{\log x_1 + \log x_2 + \dots + \log x_n}{n}\right)$$

or 
$$G = \operatorname{antilog}\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)$$

## 3.2 Geometric Mean of grouped data :

Let  $x_1, x_2, \dots, x_n$  be n observation and let  $f_1, f_2, \dots, f_n$  be their corresponding frequency then their Geometric Mean is

$$G = \left(x_1 f_1 x_2 f_2 \dots x_n f_n\right)^{1/N}$$
  
where N =  $\sum_{i=1}^n f_i$   
 $\therefore G = \operatorname{antilog}\left(\frac{\sum_{i=1}^n f_i \log x_i}{\sum_{i=1}^n f_i}\right)$ 

Power by: VISIONet Info Solution Pvt. Ltd Website : www.edubull.com

# 4. Harmonic Mean

Harmonic mean is reciprocal of mean of reciprocal.

## 4.1 Individual observation :

The H.M. of  $x_1, x_2, \dots, x_n$  of n observation is given by

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$
  
i.e. 
$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

### 4.2 H.M. of grouped data :

Let  $x_1, x_2, \dots, x_n$  be n observation and let  $f_1, f_2, \dots, f_n$  be their corresponding frequency then H.M. is

$$\mathbf{H} = \frac{\sum_{i=1}^{n} \mathbf{f}_{i}}{\sum_{i=1}^{n} \left(\frac{\mathbf{f}_{i}}{\mathbf{x}_{i}}\right)}$$

Relation between A.M. , G.M., and H.M.

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

Equality sign holds only when all the observations in the series are same.

# 5. Medain

Median is the middle most or the central value of the variate in a set of observations, when the observations are arranged either in ascending or in descending order of their magnitudes. It divides the arranged series in two equal parts.

### 5.1 Median of an individual series :

Let n be the number of observations-

- (i) arrange the data in ascending or descending order.
- (ii) (a) if n is odd then-

Median (M) = Value of 
$$\left(\frac{n+1}{2}\right)^{m}$$
 observation

(b) If n is even then

Median (M) = mean of 
$$\left(\frac{n}{2}\right)^{\text{th}}$$
 and  $\left(\frac{n}{2}+1\right)^{\text{th}}$  observation

i.e. M = 
$$\frac{\left(\frac{n}{2}\right)^{\text{th}} \text{observation} + \left(\frac{n}{2} + 1\right)^{\text{th}} \text{observation}}{2}$$

Mob no. : +91-9350679141

### 5.2 Median of the discrete frequency distribution:

Algorithm to find the median :

**Step - I** : Find the cumulative frequency (C. F.)

**Step-II** : Find 
$$\frac{N}{2}$$
, where  $N = \sum_{i=1}^{n} f_i$ 

Step -III : See the cumulative frequency (C.F.) just greater than  $\frac{N}{2}$  and determine the corresponding value of the variable.

Step-IV : The value obtained in step III is the median.

### 5.3 Median of grouped data or continuous series:

Let the no. of observations be n

- (i) Prepare the cumulative frequency table
- (ii) Find the median class i.e. the class in which the

$$\left(\frac{N}{2}\right)^{\text{th}}$$
 observation lies

(iii) The median value is given by the formulae

Median (M) = 
$$\lambda + \left[\frac{\left(\frac{N}{2}\right) - F}{f}\right] \times h$$

 $N = total frequency = \Sigma f_i$ 

- $\lambda =$ lower limit of median class
- f = frequency of the median class
- F = cumulative frequency of the class preceding the median class
- h = class interval (width) of the median class

#### 5.4 Properties of Median :

- (A) The sum of the absolute value of deviations of the items from median is minimum
- (B) It is a positional average and it is not influenced by the position of the items.

```
6. Mode
```

Mode is that value in a series which occurs most frequently. In a frequency distribution, mode is that variate which has the maximum frequency.

#### 6.1 Computation of Mode :

(i) Mode for individual series : In the case of individual series, the value which is repeated maximum number of times is the mode of the series.

# 6.2 Mode for grouped data (discrete frequency distribution series)

In the case of discrete frequency distribution, mode is the value of the variate corresponding to the maximum frequency.

#### 6.3 Mode for continuous frequency distribution :

- (i) First find the model class i.e. the class which has maximum frequency. The model class can be determined either by inspecting or with the help of grouping data.
- (ii) The mode is given by the formula

Mode = 
$$\lambda + \frac{f_{m} - f_{m-1}}{2f_{m} - f_{m-1} - f_{m+1}} \times h$$

where  $\lambda \rightarrow$  lower limit of the model class

 $h \rightarrow$  width of the model class

 $f_m \rightarrow$  frequency of the model class

 $f_{m-1} \rightarrow$  frequency of the class preceding model class

 $f_{m+1} \rightarrow$  frequency of the class succeeding model class

(iii) In case the model value lies in a class other than the one containing maximum frequency (model class) then we use the following formula :-

Mode : 
$$\lambda + \frac{f_{m+1}}{f_{m-1} + f_{m+1}} \times h$$

**6.4 Properties of Mode** : It is not effected by presence of extremely large or small items.

#### Relationship between mean, mode and median :

- (i) In symmetrical distribution
  - Mean = Mode = Median
- (ii) In skew (moderately symmetrical) distributionMode = 3 median 2 mean

### . Measures of Dispersion

Dispersion is the measure of the variations. The degree to which numerical data tend to spread about an average value is called the dispersion of the data.

The measures of dispersion commonly used are:

- (i) Range
- (ii) Quartile deviation or the semi- interquartile range
- (iii) Mean Deviation
- (iv) Standard Deviation

Here we will discuss the mean deviation and standard deviation.

**Mean Deviation :** Mean deviation is defined as the arithmetic mean of the absolute deviations of all the values taken about any central value.

Power by: VISIONet Info Solution Pvt. Ltd		
Website : www.edubull.com	Mob no. : +91-9350679141	

#### i. Mean deviation of individual observations:

If  $x_1, x_2, ..., x_n$  are n values of a variable x, then the mean deviation from an average A (median or AM) is given by

M.D. 
$$= \frac{1}{n} \sum_{i=1}^{n} |\mathbf{x}_i - \mathbf{A}|$$
$$= \frac{1}{n} \sum_{i=1}^{n} |\mathbf{d}_i|, \text{ where } \mathbf{d}_i = \mathbf{x}_i - \mathbf{A}$$

#### ii. Mean deviation of a discrete frequency distribution :

If  $x_1, x_2,..., x_n$  are n observation with frequencies  $f_1, f_2, ..., f_n$ , then mean deviation from an average A is given by -

Mean Deviation = 
$$\frac{1}{N} \sum f_i |x_i - A|$$
  
where N =  $\sum_{i=1}^{n} f_i$ 

# iii. Mean deviation of a grouped or continuous frequency distribution:

For calculating mean deviation of a continuous frequency distribution the procedure is same as for a discrete frequency distribution. The only difference is that here we have to obtain the mid-point of the various classes and take the deviations of these mid points from the given central value (median or mean)

#### 8. Variance and Standard Deviation

The variance of a variate x is the arithmetic mean of the squares of all deviations of x from the arithmetic mean of the observations and is denoted by var (x) or  $\sigma^2$ 

The positive square root of the variance of a variate x is known as standard deviation i.e. standard deviation =  $+\sqrt{var(x)}$ 

#### (i) Variance of Individual observations :

If  $x_1$ ,  $x_2$ ,....,  $x_n$  are n values of a variable x, then by definition

$$\operatorname{var}(\mathbf{x}) = \frac{1}{n} \left[ \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})^2 \right] = \sigma^2 \qquad \dots (i)$$

or var (x) = 
$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2$$
 ...(ii)

If the values of variable x are large, the calculation of variance from the above formulae is quite tedious and time consuming. In that case, we take deviation from an arbitrary point A (say) then

var (x) = 
$$\frac{1}{n} \sum_{i=1}^{n} d_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} d_i\right)^2$$
 ...(iii)

#### (ii) Variance of a discrete frequency distribution:

If  $x_1, x_2,..., x_n$  are n observations with frequencies

$$f_1, f_2, \dots, f_n \text{ then } \text{ var } (x) = \frac{1}{N} \left\{ \sum_{i=1}^n f_i (x_i - \overline{x})^2 \right\} \dots (i)$$

or var (x) = 
$$\frac{1}{N} \sum_{i=1}^{n} f_i x_i^2 - \overline{x}^2$$
 ...(ii)

If the values of x or f are large, we take the deviations of the values of variable x from an arbitrary point A. (say)

$$\therefore d_{i} = x_{i} - A ; i = 1, 2, ..., n$$
  
$$\therefore Var(x) = \frac{1}{N} \left( \sum_{i=1}^{n} f_{i} d_{i}^{2} \right) - \left( \frac{1}{N} \sum_{i=1}^{n} f_{i} d_{i} \right)^{2} ...(iii)$$
  
where  $N = \sum_{i=1}^{n} f_{i}$ 

Sometime  $d_i = x_i - A$  are divisible by a common number h(say)

then

$$u_i = \frac{x_i - A}{h} = \frac{d_i}{h}, i = 1, 2, ..., n$$

then

var (x) = h<sup>2</sup> 
$$\left[ \frac{1}{N} \sum_{i=1}^{n} f_{i} u_{i}^{2} - \left( \frac{1}{N} \sum_{i=1}^{n} f_{i} u_{i} \right)^{2} \right]$$
 ...(iv)

# (iii) Variance of a grouped or continuous frequency distribution :

In a grouped or continuous frequency distribution any of the formulae discussed in discrete frequency distribution can be used.

# 9. Mean and of Variance of Binomial Distribution

If the frequencies of the values 0, 1, 2,..., n of a variate are represented by the following coefficients of a binomial :

$$q^n,\,{}^n\!C_1\,q^{n\!-\!1}p,\,{}^n\!C_2q^{n\!-\!2}\,p^2,\!...,\,p^n$$

where p is the probability of the success of the experiment (variate), q is the probability of its failure and p + q = 1 i.e. distribution is a binomial distribution : then

$$\begin{split} P(x=r) &= {}^{n}C_{r} \ q^{n-r} \ p^{r} \\ mean \quad \overline{x} &= \sum p_{i}x_{i} \ = np \\ Variance \ \sigma^{2} &= npq = \ \overline{x} \ q \end{split}$$

 Power by: VISIONet Info Solution Pvt. Ltd

 Website : www.edubull.com
 Mob no. : +91-9350679141