4

Measures of Central Tendency

Q. 1 What do you mean by average ? What are the objects and requisites of a good average ?

Ans. It is essential to condense the data into a single value, because single value is suitable for comparison. Such a single value is treated as a representative of data. This value itself gives clear idea regarding the phenomenon under study. Such a value is referred to as an *average or central value or measure of central tendency*. It is desired that all the important properties of the observations in the data should be represented in the average. The word average is very commonly used in day-to-day life. For example : Average profit, average marks, average number of students in each class of the school, average salary per month of the workers etc. Thus the average is an essential quantity. It is a value around which most of the observations are clustered.

Objects of average

1. To facilitate comparison.
2. To obtain a single representative quantity for the entire data.

Requisites of a good average

1. It should be easy to understand.
2. It should be simple to compute.
3. It should depend upon each and every item of the data.
4. It should not be unduly affected by extreme observations.
5. It should be properly defined so that it has one and only one interpretation.
6. It should be capable of further statistical computations so that its utility is enhanced.
7. It should have sampling stability. This means that if we pick 10 different groups of college students and compute the average of each group, we should expect to get approximately the same value.

It does not mean, however, that there can be no difference in the values of different samples. There may be some difference but those samples in which this difference (i.e. sampling fluctuation) is less are considered better than those in which this difference is more.

Q. 2 Write a note on Arithmetic Mean.

Ans. 1. Simple arithmetic mean for individual observations : To compute the arithmetic mean for individual observations, the various values of the variable are added together and the total obtained is divided by the number of items.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{N} = \frac{\Sigma x}{N}$$

where $\bar{x}$ is arithmetic mean, Σx is the sum of all the values of the variable x i.e. $x_1, x_2, x_3, \ldots, x_n$ and N is the number of observations.

2. Arithmetic mean for ungrouped data : If $x_1, \ldots, x_n$ are n variables with frequencies $f_1, \ldots, f_n$ respectively, then arithmetic mean of the variable x is denoted by $\bar{x}$ and

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n}$$

In summation notation,

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

3. Arithmetic mean for the grouped data : If the data is grouped then class marks are taken as x_i and the frequency f_i of the C.I. is considered as the frequency of the corresponding class mark.

Short cut method or shift of origin : If the variables and frequencies are large in number then the variable is shifted to the point called assumed mean and it is given by $u_i = \frac{x_i - A}{C}$ where C is the class length. The arithmetic mean $\bar{x}$ is given by $\bar{x} = A + \bar{u}$ where $\bar{u}$ is the A.M. of the real variable and $\bar{u} = \frac{\Sigma u_i f_i}{\Sigma f_i}$.

4. Combined arithmetic mean : If $\bar{x}_1$ is A.M. of a group of n_1 items, $\bar{x}_2$ is A.M. of a group of n_2 items, then the A.M. of combined group of $n_1 + n_2$ items is given by

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

The formula can be extended to any number of groups.

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

5. Weighted mean : If n variables $x_1, \ldots, x_n$ are assigned weights $w_1, w_2, \ldots, w_n$ respectively, then weighted arithmetic mean is given by

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_n x_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum wx}{\sum w}$$

Q. 3 Explain how the median is calculated ?

Ans. (i) If the data is arranged in ascending order the value of the central term is called Median.

For example : Consider the data 1 5 4 3 4 2 5 3 4 4 2

The data can be rearranged in ascending order as follows :

1 2 2 3 3 4 4 4 4 5 5

There are 11 terms, the central term is t_6.

Here $t_6 = 4$ $\therefore$ Median = 4

(ii) Median for a grouped data :

If the central term is in the class interval $l_1 - l_2$ (known as median class) then

$$\text{Median} = l_1 + \frac{\frac{N}{2} - \text{c.f.}}{f} (l_1 - l_2)$$

where N = total frequency

c.f. = cumulative frequency of the class preceding the median class

f = frequency of the median class.

Q. 4 Explain the ' mode ' as a statistical average.

Ans. Mode : (i) Mode is the value of the variable which has a maximum frequency in the data.

∴ The data can have more than one mode. i.e. frequency curve may have more than one peaks. Such a distribution is called bimodal distribution or multimodal distribution. To find mode of such a distribution one may use the empirical relation between mean, median and mode.

The relation is given by

Mean – Mode = 3 (Mean – Median)

1. The modal class $l_1 - l_2$ is class which has a maximum frequency. f_1 is the frequency of the modal class, f_0 is the frequency of the class preceding modal class, f_2 is the frequency of the class following the modal class.

Then $d_1 = f_1 - f_0$ and $d_2 = f_1 - f_2$

$$\text{Mode} = Z = l_1 + \frac{d_1}{d_1 + d_2}(l_2 - l_1)$$

2. **Mode from Histogram**

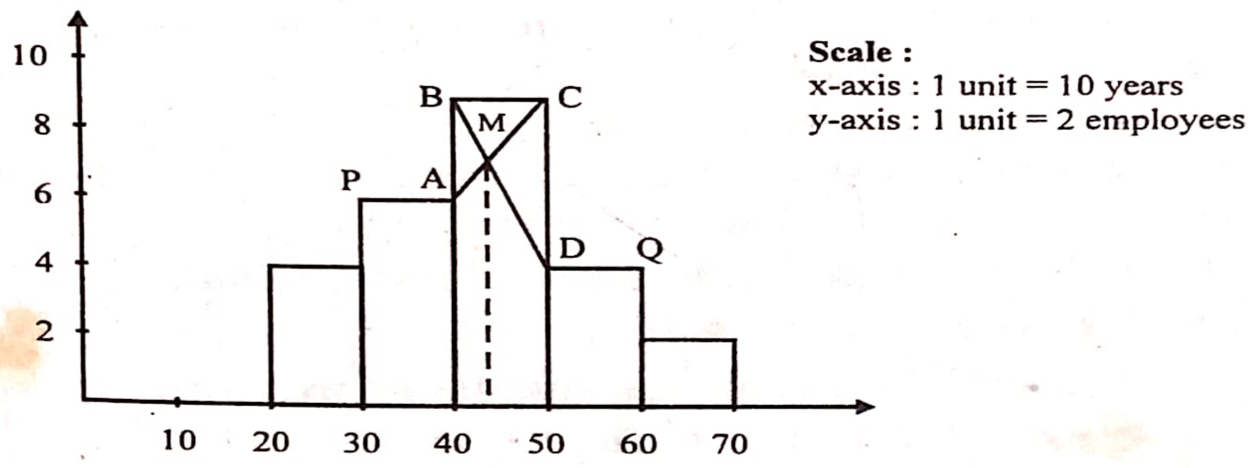
The method is illustrated by the following example.

Ex. Age distribution of 25 employees of a company is given by

Age	20–30	30–40	40–50	50–60	60–70
Employees	4	6	9	4	2

Find the made of the data.

We first draw the Histogram.



Graph

Modal Class is 40–50.

$$f_0 = 6,\ f_1 = 9,\ f_2 = 4$$

$$d_1 = 9 - 6 = 3 \text{ and } d_2 = 9 - 4 = 5$$

∴ By formula :

$$\text{Mode} = l_2 + \frac{d_1}{d_1 + d_2}(l_2 - l_1)$$

$$= 40 + \frac{3}{3+5} \times 10 = 40 + \frac{30}{8} = 43.6$$

By Histogram : Consider the rectangles corresponding to modal class, class preceding the modal class and following the modal class. PA, BC and DQ are the upper sides of these rectangles corresponding to the classes 30-40, 40-50, 50-60 respectively. Join BD and AC intersecting in M. Projection of M on x-axis is mode.

From the figure, mode = 43.6

Board

Q. 5 What is a measure of central tendency ? State the requisites of a good measure of central tendency. Compare mean and median in the light of these requisites and usefulness.

Ans. When the statistical data is orderly arranged, a large number of observations are concentrated in the central part of such arrangement. Such a characteristic of statistical data is called central tendency. The single figure which indicates the position of central tendency is generally situated near the centre or middle of the orderly arranged data and is the point of heavy concentration of observation. It is commonly known as average since it condenses a large mass of data in one single figure. It is the most representative or typical value of the data.

Since, a measure of central tendency, is a single figure used to represent a large mass of data, it is desirable that such a value possesses the following requisites or properties.

Board 2

1. It should be simple to understand and to compute.
2. It should be rigidly defined.
3. It should be amenable to mathematical treatment.
4. It should be based on all the observations.
5. It should be least affected by sampling variations.
6. It should not be unduly affected by extreme values. 2

It should posses sampling stability

Arithmetic Mean : The A. M. of N values $X_1, X_2, X_3, \ldots X_i$ of a variable X is defined as –

$$X = \frac{X_1 + X_2 + X_3 + \ldots + X_i}{N} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

It is commonly used average. It satisfies the first 5 requisites of a good measure of central tendency. However, it is unduly affected by extreme values. It cannot be computed unless all values are known. It cannot be used to represent average level of attributes. It cannot be calculated for distribution with open end class.

Median : The middle value of the orderly arranged data is called the median of the data. It is simple to understand and to calculate. It is not affected by extreme values. It can be located graphically. It can be used to ascertain the average level of certain attributes like intelligence, efficiency, honesty, beauty, etc. It is the appropriate average for markedly skew distribution like income and price distribution. The sum of the absolute deviations of values from their median is least. i.e. $\Sigma \mid X_1 - Me \mid$ is least.

It is not rigidly defined. It is not based on all values of the data. It is not capable of mathematical treatment and hence, is unsuitable for further statistical analysis. It is greatly affected by sampling variations as compared with median.

(6) **Q. 6 Discuss briefly, the relative merits and demerits of mean, median and mode.**

Ans. Merits of Mean :

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. It is based on each and every observation of the series.
4. It is capable of mathematical treatment.
5. It is affected least by fluctuations of sampling.

Demerits of Mean :

1. It is unduly affected by extreme observations.
2. In the case of open end class, it cannot be calculated without assumption about the class interval.
3. It gives equal weight to all the observations.

Merits of Median :

1. It is easy to calculate and readily understood.
2. It is not affected by extreme values.
3. It can be computed for open end classes.
4. It is the most appropriate average in dealing with qualitative data.

Demerits of Median :

1. It is necessary to arrange the data in order of magnitude for the calculation of median.
2. It is not capable of further mathematical treatment.
3. It is not based on each and every observation of the series.

Merits of Mode :

1. It is readily comprehensible and easy to calculate.
2. It is not affected by extreme values.
3. It can be calculated in case of unequal class intervals.
4. As it is not based on each and every observation, it can be calculated even in case of open classes.

Demerits of Mode :

1. It is not based on each and every observation of the series.
2. It is not capable of further mathematical treatment.
3. It is not suitable in case relative importance of the observations is to be considered.

Q. 7 What do you understand by central tendency ? Under conditions is median more suitable than other measures of central tendency ?

Ans. One of the objectives of the statistical analysis is to get one single value which describes the characteristics of the entire mass of unwieldy data. Such a value is called the central value or an average. Since an average represents the entire data, its value lies somewhere in between the two extremes, i.e., the largest and the smallest value. For this reason, it is also called as measure of central tendency. A measure of central tendency is a typical value around which other values concentrate or cluster.

Median is used most frequently when we require a measure of location which is not affected by high or low value item, and when we wish to measure the change in different sets of distribution which move in a similar direction in similar manner.

Q. 8 Define Quartiles. How are they calculated ?

Ans. The observations Q_1, Q_2, Q_3 which divide the total number of observations into four equal parts are called quartiles. These are called as the partition values. For calculation of quartiles, first of all, less than cumulative frequencies are determined. Using these cumulative frequencies, a class in which partition value lies is decided and then using the formula, the values are determined.

$$\text{First quartile } (Q_1) = l + \left[\frac{\frac{N}{4} - C.F.}{f}\right] \times h$$

$$\text{Second quartile } (Q_2) = l + \left[\frac{\frac{2N}{4} - C.F.}{f}\right] \times h$$

$$\text{Third quartile } (Q_3) = l + \left[\frac{\frac{3N}{4} - C.F.}{f}\right] \times h$$

where l = lower boundary of the class in which respective quartile lies

N = total frequency

C.F. = cumulative frequency less than type of the class just preceding the class in which respective quartile lies.

f = frequency of the class in which respective quartile lies

and h = class width.

Note that second quartile of the data is nothing but the median of that data.