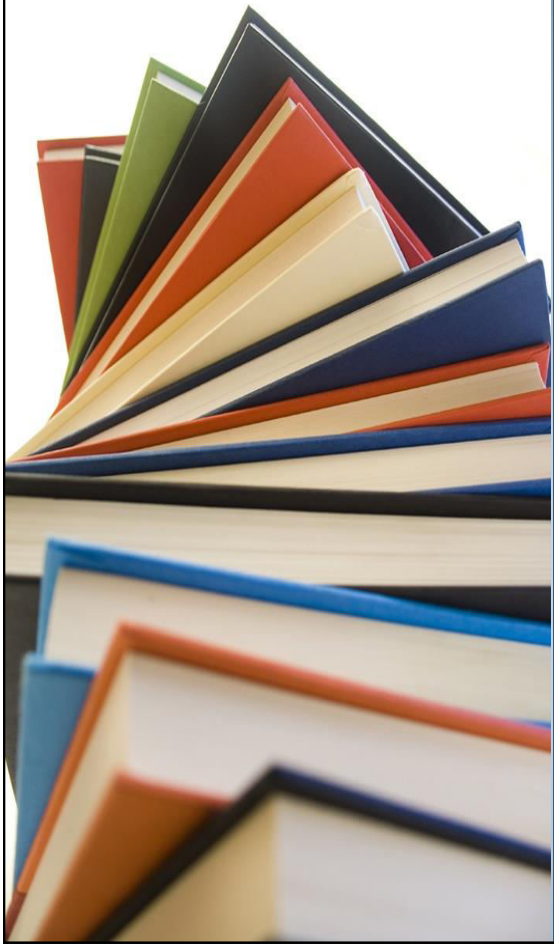

Unit-2
Measure of Central Tendency

Definition:

- When we have a large data set corresponding to a single variable, than it becomes essential to find a value which can be considered as a representative of the entire set of observation. The representatives is generally one single value which is approximately at the centre far away from both the extremes.
- A measure indicating the value to be expected of typical or a middle data point.
- There are different Measures of Central Tendency. The most important of them are the Mean, Median & Mode.3

Characteristics:

The important characteristics for an ideal Measure of Central Tendency are:

i. It should be based on all the observations of the series.

ii. It should be easy to calculate and simple to understand.

iii. It should not be affected by extreme values.

iv. It should be rigidly defined.

v. It should be capable of further mathematical treatment.

vi. It should be least affected by the fluctuations of sampling.

2.1Mean:

The mean is also called as an average. There are three types of mean namely – Arithmetic Mean, Geometric Mean, Harmonic Mean.

2.1.1Arithmetic Mean:

The Arithmetic Mean is obtained from a set of numbers by dividing the sum of those numbers by the number of observation.

If there are n observations, then their arithmetic mean is given by

$x_1, x_2, x_3, \ldots\ldots, x_n$

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

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

Computation of Mean when frequencies are provided:

- Sometimes we come across data where along with the values of observation we are also provided with the corresponding frequencies

  If x_1 occurs f_1 times

  If x_2 occurs f_2 times

  ....

  If x_n occurs f_n times

  then Arithmetic Mean is given by

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + ... + f_nx_n}{f_1 + f_2 + ... + f_n}$$

$$= \frac{\sum f_ix_i}{\sum f_i}$$

$$= \frac{\sum f_ix_i}{N} \; where N = \sum f_i$$

Though there are 3 different form of mean, yet by mean we generally refer to the Arithmetic mean.

Example:

- The following data set gives the number of children in 100 families in a certain village

No. of Children	1	2	3	4	5	6	7
No. of Families	7	9	25	22	18	11	8

There Arithmetic Mean is given by, $\bar{x} = \sum \frac{f_i x_i}{N}$

No. of Children (x_i)	No. of Families (f_i)	$f_i x_i$
1	7	7
2	9	18
3	25	75
4	22	88
5	18	90
6	11	66
7	8	56
	N=100	$\sum f_i x_i$ =400

$\therefore$ the required AM

$$\bar{x} = \frac{1}{N} \sum f_i x_i$$

$$= \frac{1}{100} \times 400$$

$$= 4 \ \textit{childern}$$

Short Method:

$$= A + \frac{\sum f_i d_i}{\sum f_i} ; \textit{where}\ d_i = x_i - A$$

$$\textit{and } A \textit{ is assumed mean}$$

Step Deviation Method:

$$= A + \frac{\sum f_i d_i}{\sum f_i} \times i \,; where\, d_i = \frac{x_i - A}{i}$$

and A is assumed mean and i is class length

2.1.2Geometric Mean:

- The geometric mean of a number is given by the root of the product of those numbers. The GM is used for the average of rates and ratios. It cannot be computed if any value is negative. If any observation is zero, GM is also zero.
- If X_1 , X_2 , X_3 ,................ X_n are n observations, then their geometric mean is given by

$$GM = \sqrt[n]{x_1 \times x_2 \times .... \times x_n}$$

$$= \sqrt[n]{\prod_{i=1}^{n} x_i}$$

$$= \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

2.1.3 Harmonic Mean:

- The reciprocal of the harmonic mean is the arithmetic mean of reciprocal number.
- If H is the Harmonic mean then

$$\frac{1}{H} = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots.. + \frac{1}{x_n}}{n}$$

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots.. + \frac{1}{x_n}}$$

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

2.2 Median:

- Median is the middle value of a set of observations. It is obtained by selecting the central value of a data set after arranging the data in ascending or descending order.

In case of odd number of observation, 2n+1(say) the n (n+1/2)th observation after arranging the data in ascending or descending order gives the median.

Example: For the following dataset
23, 19, 16, 34, 41, 7, 62
we arrange them in ascending order as follows
7, 16, 19, 23, 34, 41, 62
the value in the middle, i.e. 23 is the median of the dataset.

In case of even number of observation, 2n(say) the average of (N/2) and ((N/2)+1)observation provides the median.

For Example: For the following dataset
23, 19, 16, 34, 41, 7
we arrange them in ascending order as follows
7, 16, 19, 23, 34, 41
the average of 19 and 23 i.e. $\frac{19+23}{2} = 21$ is the median.

The Mathematical formulae for computing median is:

$$Median = l + \frac{\frac{N}{2} - C}{f} \times h$$

where l = lower limit of the median class
N=total frequency
C=cumulative frequency of the class previous to the median class
f=frequency of the median class
h=class interval of the median class

Example:

- From the following dataset, let us compute the median

Ages (in years)	25-30	30-35	35-45	45-50	50-55	55-60	60-65
No. of Employees	13	17	14	16	7	3	2

In order to obtain the median we construct the following table:

Class	f_i	Cumulative frequency
25-30	13	13
30-35	17	30
35-45	14	44
45-50	16	60
50-55	7	67
55-60	3	70
60-65	2	72
	N=72	

Now, $Median = l + \frac{\frac{N}{2} - C}{f} \times h$

Here, $\frac{N}{2} = \frac{72}{2} = 36$

from the cumulative frequency column we find that 35-45 is the median class, so we have, l = 35, C = 30, f = 14, h = 10.

Thus median $= 35 + \frac{36-30}{14} \times 10$

= 39.28 years

2.3 Mode:

- Mode is that value of variate which have the maximum frequency. The particular value of a variable which occurs maximum number of times on repetition is called as the mode.

For Example: From the following set of marks, find the mode.

Marks	Frequency
1	2
2	7
3	7
4	4
5	6
6	5

The marks 2 and 3 have the highest frequency. So, the modes are 2 and 3.

The above example also shows that a set of observations may have more than one mode.

The Mathematical formulae for computing mode is:

$$Mode = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

From the following dataset let us compute the mode:

Ages (in years)	25-30	30-35	35-45	45-50	50-55	55-60	60-65
No. of Employees	13	17	14	16	7	3	2

In order to obtain the mode we construct the following table:

Class	f_i
25-30	13
30-35	17
35-45	14
45-50	16
50-55	7
55-60	3
60-65	2

here the class 30-35 has the highest frequency, so this class is the model class. So, l = 30(lower limit of model class), f_1=17(frequency of model class), f_0= 13(frequency prior to the model class), f_2= 14(frequency following to the model class), h=5(class interval).

$$Mode = 30 + \frac{17-13}{2\times 17 - 13 - 14} \times 5$$

$$= 32.86\, years$$

Relation between Mean, Median & Mode:

- In case of symmetrical data, an empirical relation exist between the mean, median and mode. This relation enables one to find the value of mean or median or mode provided the values of the other two are given. The relation is –

  Mean – Mode = 3 (Mean - Median)

  For Example: Let us compute the mode of a data set if the mean is 40.21 and median is 39.28

  Sol^n : We know that,

  mean-mode=3(mean-median)

  $\Rightarrow$ mode = $3(40.21 - 39.81) - 40.21$

  $\Longrightarrow$ mode = 37.42

- The empirical relationship should not be considered as a regular practice for calculating the value of mean, median or mode. This is not a formula and the relation ship holds only for a moderately skewed distribution.

Uses of Central Tendency:

- The importance of the mode, mean and median in business depends on the analysis required and the business function to which the results apply. For some data, the three values are close or the same, while for other types of data, the mode or median may differ substantially from the mean. When the three calculations give different results, the key is to choose the value that will give the desired guidance. This choice is different for different business functions.

  Mean: In business, the mean is the most important value when data is scattered, without a typical pattern. Such patterns can occur in procurement, where costs vary according to external factors. The mean gives the average cost and forms a good basis for estimating future costs, as long as the external factors remain the same.

- **Median**: The median is the most important value when the data has several values that occur frequently, and several comparatively very high values. An analysis of salaries often focuses on the amounts commonly paid but ignores extremes that are probably special cases. The median salary gives a value close to the average salary commonly paid, without taking the extreme values into consideration.
  Mode: Mode is generally used in garment industries and by shoe manufacturer as by different sizes they choose would fit the maximum number of people.