LESSON THREE: MEASURES OF CENTRAL TENDENCY

3.1 Introduction

- A measure of central tendency, also called measures of location or averages, is a single value within the range of data that is used to represent all the values in the series.

3.2 Characteristics of a good average

Should be-

- Rigidly defined
- Based on all values
- Easily understood and calculated
- Least affected by the fluctuations of sampling
- Capable of further algebraic treatment
- Least affected by extreme values

3.3 Types of averages

The measures of central tendency that are generally used in business are:

a) Arithmetic mean

b) Median

c) Mode

d) Geometric mean

e) Harmonic mean

3.3.1 The Arithmetic Mean

It is obtained by summing up the values of all the items of a series and dividing this sum by the number of items.

Computation of the arithmetic mean for

Individual series :-

Direct method

$\overline{X} = \frac{\sum X}{n}$ where $\overline{X}$ = arithmetic mean , n = number of items

Indirect method

$\overline{X} = P.M. + \frac{\sum Dx}{n}$ where $P.M$ = provisional mean, Dx = Deviations from P.M, $\sum Dx$ = the sum of deviations from P.M

Grouped series

Direct method

$\overline{X} = \frac{\sum xf}{n}$ Where f = frequencies, n = number of items

Indirect method

$\overline{X} = P.M. + \frac{\sum fDx}{n}$

NB: For a grouped frequency distribution the value of X is taken as the mid point of each class.

<u>Examples</u>

1. The monthly sales of ABC stores for the period of 6 months were as follows:

37,000, 48,000, 84,000, 73,000, 35,000, 53,000.

Calculate the mean

2.Calculate the mean of the following distribution

Number of vehicles serviced (x)	0	1	2	3	4	5
Number of days (f)	2	5	11	4	4	1

The following tables gives the marks of 58 students in statistics.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of students	4	8	11	15	12	6	2

Calculate the mean mark.

Advantages of the arithmetic mean

- Can be easily understood

- Takes into account all the items of the series
- It is not necessary to arrange the data before calculating the average
- It is capable of algebraic treatment
- It is a good method of comparison
- It is not indefinite
- It is used frequently.

Disadvantages of the arithmetic mean

- It is affected by extreme values to a great extent
- It may be a figure that does not exist in a series
- It cannot be calculated if all the items of a series are not known
- It cannot be used incase of qualitative data

Properties of Arithmetic Mean

1. The product of the arithmetic mean and the number of items is equal to the sum of all the given values.
2. The algebraic sum of the deviations of the values from the arithmetic mean is equal to zero. As such the mean may be characterized as a point of balance.
3. The sum of the squares of deviations from arithmetic mean is the least.
4. As the arithmetic mean is based on all the items in a series, a change in the value of any item will lead to a change in the value of the arithmetic mean.
5. If we have the arithmetic means and number of observations of two or more than two groups, we can compute combined mean of these groups using this formula:

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

Examples

1.There are two branches of a company employing 100 and 80 employees respectively. If arithmetic means of the monthly salaries paid by the two branches are \$4570 and \$6750 respectively. Find the arithmetic mean of the salaries of the employees of the company put together.

3.3.2 The Median

- It is the middle value when data has been arranged increasing or decreasing magnitude.

Computation of the Median for Ungrouped data

- If the number of observations is odd, the median is the middle value after the observations have been arranged in some order
- If the number of observations is even, the median is the arithmetic mean of the two middle observations after the data has been arranged in some order

Computation of the median in discrete series with Frequencies

Steps

1. Construct the less than cumulative frequency distribution
2. Find $N/2$, where $N = \sum f$
3. Check the cumulative frequency just greater than $N/2$
4. The corresponding value of the variable is the median

Computation of the Median in Grouped Data.

There are two approaches

1. Graphical method - Using the cumulative frequency curve (Ogive curve)
2. Interpolation formula

Interpolation Formula

Steps

1. Construct the less than cumulative frequency distribution
2. Find $N/2$, where $N = \sum f$
3. Check the cumulative frequency just greater than $N/2$
4. The corresponding class contains the median and is called the median class.

The median has to be interpolated in the class interval containing the median using the formula:-

$$median = L + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

where L = Lower class boundary of the median class

h= Length of the classes

f = Frequency of the median class

N = Total frequency

C = cumulative frequency of the class preceding the median class.

Examples

1. Find the median of the data below:

   a) 5, 5, 4, 7, 0, 7, 8

b) 20, 15, 30, 45, 60, 10

2. Determine the median of the data below.

Grade	A	B	C	D	E
No of students (f)	10	15	67	50	21

3. Determine the median for the grouped data below.

Marks	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 – 79	80-89
No of students	2	7	15	30	20	4	1

Properties of the Median

- It is a positional average and is influenced by the position of the items in the series and not by the size of items
- The sum of the absolute values of deviations is least.

Advantages of the Median

- It is easy to calculate
- It is simple and is understood easily
- It is less affected by the value of extreme items

- It can be calculated by inspection in some cases
- It is useful in the study of phenomenon which are of qualitative nature

Disadvantages of the Median

- It is not a suitable representative of a series in most cases
- It is not suitable for further algebraic treatment
- It is not used frequently like arithmetic mean

Quartiles, deciles and percentiles

- Quartiles are the values of the items that divide the series into four equal parts.
- Deciles divide the series into 10 equal parts.
- Percentiles divide the series into 100 equal parts.
- The 2nd quartile, 5th decile and 50th percentile are equal to the median.

Computation of the Quartiles

First Quartile: $$Q_1 = L_{Q_1} + \frac{h}{f_{Q_1}}\left(\frac{N}{4} - C\right)$$

Second Quartile: $$Q_2 = L_{Q_2} + \frac{h}{f_{Q_2}}\left(\frac{2N}{4} - C\right)$$

Third Quartile: $$Q_3 = L_{Q_3} + \frac{h}{f_{Q_3}}\left(\frac{3N}{4} - C\right)$$

In general, the three quartiles can be computed for grouped data by the formula

$$Q_i = L_{Q_i} + \frac{h}{f_{Q_i}}\left(\frac{iN}{4} - C\right)$$

where L_{Q_i} = Lower class boundary of the i^{th} quartile class

h = Length of the classes

f_{Q_i} = Frequency of the i^{th} quartile class

N = Total frequency

C = Cumulative frequency of the class preceding the i^{th} quartile class.

Computation of the Deciles

$$D_i = L_{D_i} + \frac{h}{f_{D_i}}\left(\frac{iN}{10} - C\right)$$

where L_{D_i} = Lower class boundary of the i^{th} decile class

h = Length of the classes

f_{D_i} = Frequency of the i^{th} decile class

N = Total frequency

C = Cumulative frequency of the class preceding the i^{th} decile class.

Computation of the Percentiles

$$P_i = L_{P_i} + \frac{h}{f_{P_i}}\left(\frac{iN}{100} - C\right)$$

where L_{P_i} = Lower class boundary of the i^{th} percentile class

h = Length of the classes

f_{P_i} = Frequency of the i^{th} percentile class

N = Total frequency

C = Cumulative frequency of the class preceding the i^{th} percentile class.

NB: Analogous to the graphical method of estimating the median, the quartiles, deciles and percentiles of a grouped frequency distribution can be estimated using the cumulative frequency curve (ogive curve).

Examples

1. Find the 1st , 2nd and 3rd quartiles for the following data

   13, 9, 18, 15, 14, 21, 7, 10, 11, 20, 5, 18, 25, 16, 17

2. Given below is the number of families in a locality according to their monthly expenditure

Monthly expenditure	No. of families
140 - 150	17
150 - 160	29
160 - 170	42
170 - 180	72

180 - 190	84
190 – 200	107
200 – 210	49
210 – 220	34
220 – 230	31
230 – 240	16
240 – 250	12

Calculate:

i) All the quartiles

ii) 7^{th} decile

iii) 90^{th} percentile

3.3.3 The Mode

- The mode is the value, which occurs most often in the data. A distribution with one mode is called unimodal, with two modes bimodal and with many modes, multimodal distribution.
- There are two methods that can be used to estimate the mode of grouped data .
  a) Graphically, using a histogram
  b) Using an interpolation formula

Graphical determination of mode

Procedure

1. Construct a histogram for the data
2. Locate the highest cell in the histogram, join the upper class boundary of the cell with the upper boundary of the preceding cell; then join the lower class boundary of the highest cell with the lower class boundary of the succeeding cell, locate the intersection,
3. Draw a vertical line from the intersection to the horizontal.
4. The value of the vertical line on the horizontal axis is the mode.

Interpolation Formula

$$Mode = L + \frac{h(f_m - f_1)}{2f_m - f_1 - f_2}$$

$$= L + \left[\frac{D_1}{D_1 + D_2}\right] \times i$$

Where $L =$ Lower class boundary of the modal class

h = Length of the classes

$f_m =$ Frequency of the modal class

$f_1 =$ Frequency of the class preceding the modal class

$f_2 =$ Frequency of the class succeeding the modal class

$D_1 = f_1 - f_0, \quad D_2 = f_1 - f_2$

Examples

1. Find the mode for the data below

a) 1, 2, 3, 4, 5, 6; **Solution:** The mode does not exist

b) 7, 8, 3, 8, 6, 10, 8 **Solution:** Mode = 8; This is a **uni-modal** distribution

c) 29,30,60,13,30,7,2,7 **Solution:** Modes are 30 and 7; This is a **bi-modal** distribution

d)

X	4	5	6	7	8	9	10
F	2	5	21	18	9	2	1

Solution: Mode = 6; it has the highest frequency.

2. Calculate the mode for the following data

Class (marks)	No of student
0 – 10	2
10 – 20	7
20 – 30	11
30 – 40	6
40 – 50	4

Properties of the mode

- It represents the most typical value of the distribution and it should coincide with existing items
- It is not affected by the presence of extremely large or small items

Advantages of the Mode

- It is easy to understand
- Extreme items do not affect its value
- It possesses the merit of simplicity

Disadvantages of the Mode

- It is often not clearly defined
- Exact location is often uncertain
- It is unsuitable for further algebraic treatment
- It does not take into account extreme values.

Relationship between the mean, median and mode

There usually exists a relationship among the mean, median and mode for moderately asymmetrical distributions.

- If the distribution is symmetrical, the mean median and mode will have identical values.
- If the distribution is skewed (moderately) the mean, median and mode will pull apart. If the distribution tails off towards higher values, the mean and the median will be greater than the mode i.e. In case, a distribution is skewed to the right, then mean> median> mode. Generally, income distribution is skewed to the right where a large number of families have relatively low income and a small number of families have extremely high income. In such a case, the mean is pulled up by the extreme high incomes.

If it tails off towards lower values, the mode will be greater than either of the two measures i.e. When a distribution is skewed to the left, then mode> median > mean. This is because here mean is pulled down below the median by extremely low values.

In either case the median will be about one third as far away from the mean as the mode is. This means that

Mode = mean –3 (mean – mode)

= 3(median) – 2(mean)

3.3.4 Geometric Mean

Geometric Mean (GM) is the nth root of the product of n values

For ungrouped data

$$G.M = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}$$

$$\Rightarrow G.M = \left(x_1 \times x_2 \times \ldots \times x_n\right)^{\frac{1}{n}}$$

$$\Rightarrow Log\ G.M = \frac{1}{n}\left(\log x_1 + \log x_2 + \ldots + \log x_n\right)$$

$$\Rightarrow \log G.M = \frac{1}{n}\sum \log x_i$$

$$\Rightarrow G.M = Anti\log \frac{\sum Logx}{n}$$

Grouped data

$$G.M = \sqrt[N]{x_1^{f_1} \times x_2^{f_2} \times \ldots x_n^{f_n}}$$

$$\Rightarrow G.M = \left(x_1^{f_n} \times x_2^{f_2} \times \ldots \times x_n^{f_n}\right)^{\frac{1}{N}}$$

$$\Rightarrow Log\ G.M = \frac{1}{N}\left(f_1 \log x_1 + f_2 \log x_2 + \ldots + f_n \log x_n\right)$$

$$\Rightarrow \log G.M = \frac{1}{N}\sum f_i \log x_i$$

$$\Rightarrow G.M = Anti\log \frac{\sum f_i Logx_i}{N} \qquad \text{where } N = \sum f$$

Examples

1. The weekly incomes (‘000) of 10 families are given below. Find the geometric mean?

   50, 80, 45, 70, 15, 75, 85, 40, 36, 25

2. Calculate the geometric mean of the given data

X	15	20	25	30	35	40	45	50

F	2	22	29	24	7	8	6	2

Merits of the Geometric mean

- It takes into account all the items in the data and condenses them into one representative value.
- It gives more weight to smaller values than to large values.
- It is amenable to algebraic manipulations

Demerits

- It is difficult to use and compute
- It is determinate for positive values and cannot be used for negative values or zero.

3.3.5 Harmonic Mean

It is the reciprocal of the arithmetic mean of the reciprocal of a series of observations.

Ungrouped data

$$\text{H.M} = \frac{n}{\sum\left(1/x\right)} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}}$$

Grouped data

$$\text{H.M} = \frac{\sum f}{\sum\left(f/x\right)} = \frac{f_1+f_2+\ldots+f_n}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ldots+\frac{f_n}{x_n}}$$

Examples

1. Calculate the Harmonic mean of the following data

   11, 13, 15, 16, 19, 22, 13, 20

2. Calculate the Harmonic mean of the following data

X	15	20	25	30	35	40	45	50
F	2	22	29	24	7	8	6	2

Merits of the Harmonic mean

- It takes into account all the observations in the data
- It gives more weight to smaller items
- It is amenable to algebraic manipulations
- It measures the rates of change

Demerits

- It is difficult to compute when the number of items is large
- It assigns too much weight to smaller items.

3.4 Factors to consider in the choice of an average

- The purpose for which the average is being used
- The nature, characteristics and properties of the average
- The nature and characteristics of the data.

3.5 Exercise

1. What are the requirements of a good average? Compare the mean, the median and the mode in the light of these requirements.
2. Find the mean, median and mode for the following set of data
   i) 3, 5, 2, 6, 5, 9, 5, 2, 8 and 6
   ii) 51.6, 48.7, 50.3, 49.5 and 48.9
3. The following data pertain to marks obtained by 120 students in their final examination in mathematics:

Marks	Number of Students
30 -39	1
40 – 49	3
50 – 59	11
60 – 69	21
70 – 79	43
80 -89	32
90 - 99	9
Total	120

   Calculate the mode and the median.

4. Suppose we are given the following series:

Class interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	6	12	22	37	17	8	5

i) Draw the histogram and the Ogive from these data

ii) Estimate the median and the mode from the graphs in (i) above

5. The mean of marks in statistics of 100 students of a class was 72. The mean of marks of boys was 75 while their number was 70. Find out the mean mark of girls in the class.