

# PSYCHOLOGICAL STATISTICS

**B Sc. Counselling Psychology**  
2011 Admission onwards

**II SEMESTER**

**COMPLIMENTARY COURSE**



**UNIVERSITY OF CALICUT**

**SCHOOL OF DISTANCE EDUCATION**

CALICUT UNIVERSITY.P.O., MALAPPURAM, KERALA, INDIA – 673 635

**UNIVERSITY OF CALICUT**  
**SCHOOL OF DISTANCE EDUCATION**

STUDY MATERIAL

**B Sc. Counselling Psychology**

II Semester

COMPLEMENTARY COURSE  
PSYCHOLOGICAL STATISTICS

Prepared by:

*Dr. Vijaya Kumari.K.,*  
*Associate Professor,*  
*Farook Training College,*  
*Farook College (P.O)*  
*Feroke*

Edited and Scrutinised by

*Prof. (Dr.) C. Jayan*  
*Dept. of Psychology*  
*University of Calicut*

Layout & Settings  
Computer Section, SDE

©  
Reserved

## CONTENTS

MODULE I	04-15
MODULE II	16-26
MODULE III	27-34

## MODULE 1

### FREQUENCY DISTRIBUTION

#### Objectives:

The student will be

1. acquainted with the knowledge of frequency table and various diagrams and graphs.
2. capable of preparing frequency table using raw data.
3. capable of drawing pie-diagram, histogram, frequency polygon, frequency curve and ogives.

#### Introduction

Statistics deals with numerical data. The word 'data' is plural, the singular being 'datum' meaning fact. Usually in statistics, the term 'data' means evidence or facts describing a group or a situation. In practical sense, data indicates numerical facts such as measures of height, weight, scores on tests like achievement tests, intelligence tests, creativity tests or scores on any measuring instrument.

The data in its original form is known as raw data. That is, they are taken as such without any classification or organization. For example the marks obtained by ten students in an English achievement test are given as

20, 30, 35, 15, 10, 22, 36, 38, 21, 16.

These marks (scores) constitute the raw data. But the raw data will not be meaningful to the investigator or the data will be too large to manipulate. For getting a clear picture about the data and making the analysis possible, one has to classify and tabulate the information. Then raw data are grouped together or organized in such a way that the features of the data are revealed. In grouped data, the individual score has no meaning, but the characteristics of the total data is revealed. Grouping helps the investigator to know about the distribution and makes the analysis easier.

#### Classification and tabulation

The collected data after editing (avoiding incomplete or incorrect information) is classified. Classification is the process of grouping of related data into classes. It is the first step of tabulation. Classification helps to organize the data in a tabular form. This can be done in two ways; either as discrete frequency table or continuous frequency table.

#### Frequency table

A frequency table is an arrangement of the raw data revealing frequency of each score or of a class.

For preparing a discrete frequency distribution (table), first place all the possible values of the variable from lowest to highest. Then put a vertical line (tally) against the particular value for each item in the given data. Usually blocks of five bars are prepared by crossing the four bars already marked by the fifth one. Finally count the number of bars and write the frequency.

**Example**

The achievement scores of 50 students are given below.

48	42	37	35	41	27	28	30	27	31
43	41	38	34	42	28	31	37	28	35
21	27	36	28	26	31	38	28	27	30
42	37	35	41	36	34	37	29	28	40
43	41	42	36	37	41	37	28	26	27

This data can be converted into a discrete frequency table by following the steps described below.

1. First list the individual scores that are included in the data without out repetition in ascending order.
2. Then put a tally mark against each score when ever it occurs in the raw data.
3. Add the tally marks which represent the frequency of each data.
4. Write the sum of the frequencies which will be equal to the total number of scores.

Marks	Tally marks	Frequency
21	I	1
26	II	2
27	<del>III</del>	5
28	<del>III</del> II	7
29	I	1
30	II	2
31	III	3
34	II	2
35	III	3
36	III	3
37	<del>III</del> I	6
38	II	2
40	I	1
41	<del>III</del>	5
42	IIII	4
43	II	2
48	I	1
Total		50

In the case of discrete variables, one can prepare such types of frequency distribution. But, discrete frequency table again may be lengthy especially when the data contains a large number of scores and a long series. For proper handling of data, adequate organization will be needed. For this, numerical data will be grouped into some groups or classes and the frequency of each class is found out by putting tally marks against each class when an individual score belongs to that class.

The systematic steps of forming a frequency distribution are given below.

**Step 1: Calculate the range**

Range is the difference between the lowest and highest score in the set of data. That is Range = H – L where H is the highest score and L is the lowest score.

**Step 2: Determination of class interval**

The number and size of the classes are to be decided. The number of classes is decided according to the number of scores included. Usually the number of classes is limited to 20, but if number of items are small, say 50, the number of classes may be around 10. The class size or class interval denoted as ‘i’ is calculated using the formula  $i = \frac{\text{Range}}{\text{Number of classes}}$  ‘i’ should be taken as a whole number and hence nearest approximate number can be taken as the class interval.

**Step 3 : Writing the classes**

Classes are written from lowest to highest from bottom to top. The lowest class is prepared so that it contains the lowest score in the set and the last class is written so that the highest number is included in that class.

**Step 4: Marking tally and writing the frequency**

Mark tally against each class for items belonging to that class. In the next column, write the frequency, which represents the number of items in that class.

The total of the third column ‘frequency’ should be equal to the number of data. In the above example, the least score is 21 and largest is 48.

$$\begin{aligned} \text{Hence range} &= 48 - 21 \\ &= 27 \end{aligned}$$

Here the total number of data is 50 and we can classify it into approximately 10 classes. Then the class interval will be

$$\begin{aligned} i &= \frac{\text{Range}}{\text{Number of classes}} \\ &= \frac{27}{10} = 2.7 \end{aligned}$$

Class interval may be taken as a whole number, and hence ‘i’ can be approximated to 3.

Now we have to write the classes. The first class should include 21 and the last class should include 48.

Classes	Tally marks	Frequency
47 – 49	I	1
44 – 46		0
41 – 43	<del>III</del> III I	11
38 – 40	III	3
35 – 37	<del>III</del> III II	12
32 – 34	II	2
29 – 31	<del>II</del> I	6
26 – 28	<del>II</del> III IIII	14
23 – 25		0
20 – 22	I	1
		N = 50

Here 20- 22 is the first class and it is taken so that the lowest score 21 is included in it. Similarly the last class is 47-49 which include the highest value 48.

The class interval of the class 20-22 is three because 20, 21 and 22 are included in this class. The value 20 is known as the lower limit of that class and 22 as the upper limit. The mid point of the class is 21 which is the average of the upper and lower limits of the class  $[\frac{(20+22)}{2}]$

In this frequency distribution it is assumed that there is no scores between 22 and 23, 25 and 26 and so on. But if the variable measured is a continuous variable, one cannot take the classes like this and hence should convert the classes into actual classes. This is done by bridging the gap between the upper limit of a class and lower limit of the next class. For this 0.5 is added to the upper limit of the classes and 0.5 is subtracted from the lower limit of each class. The new frequency table of the above data will be,

<b>Actual classes</b>	<b>Frequency</b>
46.6 – 49.5	1
43.5 – 46.5	0
40.5 – 43.5	11
37.5 – 40.5	3
34.5 – 37.5	12
31.5 – 34.5	2
28.5 – 31.5`	6
25.5 – 28.5	14
22.5 – 25.5`	0
19.5 – 22.5	1
	-----
	50

In this frequency distribution, it is assumed that the lower limits are included in the class but the upper limits are excluded from that class. Thus, in the class 46.5 - 49.5, the value 46.5 is included in the class, but 49.5 is excluded from the class. The class interval will not change by this procedure. That is class interval 'i' of the class 46.6 – 49.5 = 49.5 – 46.5 (u.l – l.l)

$$= 3.$$

### **Diagrams and Graphs**

Statistical data may be displayed pictorially through various types of diagrams, graphs and maps. Diagrams and graphs are convincing and appealing ways in which statistical data may be presented. These presentations will be more interesting to the common man than the frequency tables.

Diagrams and graphs are important because,

- They provide a bird's eye-view of the entire data and therefore the information is easily understood. Large number of figures may be confusing, but the pictorial presentation makes the data simple to understand and interesting to the readers.
- They are attractive to the eye. When figures are over looked by the common men, pictures create greater interest.
- They have great memorising effect. The impressions created by diagrams are long lasting than that made by data in tabular form.

- They facilitate comparison of data. Quick and accurate comparison of data is possible through diagrammatic presentation.
- They bring out hidden facts and relationship and help in analytical thinking and investigation.

**Pie Diagrams**

Pie diagrams are diagrammatic presentation of data in which the data is presented through the sections of a circle. Usually they are used to show percentage break downs.

In constructing a pie diagram (pie chart) the following steps are to be followed.

**Step 1**

The data to be represented should be transposed into corresponding degrees on the circle. The total frequency is equated to 360° and then the angles corresponding to component parts are calculated. If percentage of each component data is considered, to convert it into degree measures multiply the value with 3.6. ie.,  $\left(\frac{360}{100}\right)$

**Step 2**

Draw a circle with appropriate size with a compass.

**Step 3**

Mark the points on the circle representing the size of each sector with the help of protractor. Usually the largest sector is drawn from 12'0 clock position on the circle. The other components are placed in clockwise succession in descending order of magnitude.

**Illustration**

Draw a pie diagram to represent the following data detailing the monthly expenses of an institution

Salary of the staff	:	60,000
Electricity water & Telephone bills	:	15,000
Office Stationary	:	10,000
Miscellaneous	:	15,000
		-----
		1,00,000

Here data is not given as percentage. Hence to convert the data into degree measures multiply each value with  $\frac{360}{1,00,000}$ , ie.,(360/Total).

Then the new set of data will be

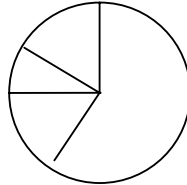
Salary	:	$60,000 \times \frac{360}{1,00,000} = 216^\circ$
Electricity	:	$15,000 \times \frac{360}{1,00,000} = 54^\circ$
Office stationary	:	$10,000 \times \frac{360}{1,00,000} = 36^\circ$
Miscellaneous	:	$15,000 \times \frac{360}{1,00,000} = 54^\circ$



Starting from the 12'O clock position in the circle sectors are marked in the descending order of magnitude. But 'miscellaneous' section is to be marked at the last without considering the magnitude.

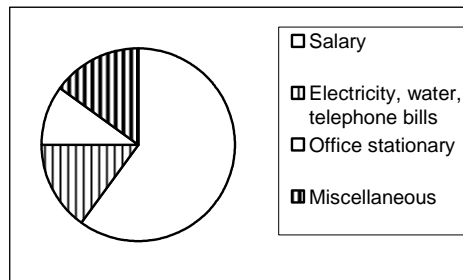
To mark the first sector put the protractor in the line and mark  $216^{\circ}$  in the clock wise direction. From this line mark  $54^{\circ}$  and so on.

The resulting figure will be like the following.



Now these sectors can be shaded properly to make it more attractive and striking.

Monthly expenses of an institution



Pie diagrams are attractive and can be effectively used to show the split up of a whole. But if there are a large numbers of components in the data or if the difference between the components is very small, pie diagram will not be much useful. Pie diagrams are less effective for comparison between different sets of data.

### Graphs of Frequency Distribution

A frequency distribution can be presented graphically in the following ways.

1. Histogram
2. Frequency polygon
3. Frequency curve
4. Ogives

#### 1. Histogram

Histogram is a set of vertical bars whose areas are proportional to the frequencies represented. It is the most popular method of presenting a frequency distribution.

#### Construction of a histogram

While constructing a histogram, the variable is always taken on the X-axis and the frequencies on the Y-axis. Each class is marked on the X-axis by taking an appropriate scale to represent the class interval. Then rectangles are erected at each class with height as the frequency of that class. The area of each rectangle is proportional to the frequency of that class and the total area of the histogram is proportional to the total frequency.

[Note: While drawing graphs, it is conventional to take the scales on X-axis and Y-axis so that the height of the graph is 75% of its width.]

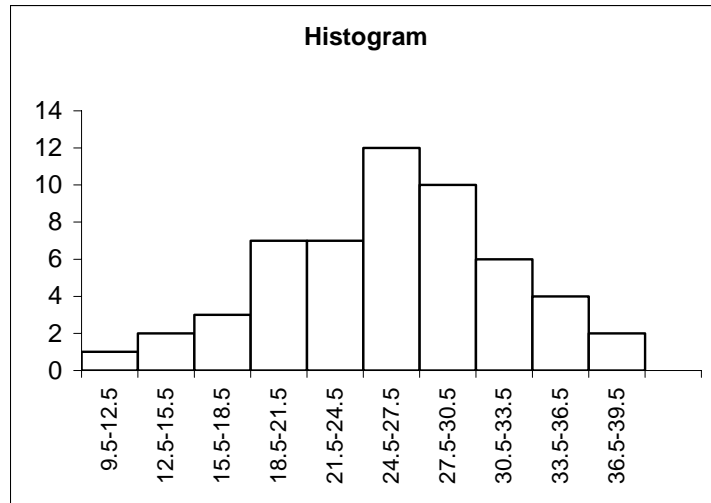
Illustration: Draw histogram for the following data.

<b>Classes</b>	<b>Frequency</b>
37-39	2
34-36	4
31-33	6
28-30	12
25-27	7
22-24	7
19-21	7
16-18	3
13-15	2
10-12	1

Here the classes are not continuous and hence have to be converted into actual classes. For this reduce 0.5 from each lower limit and add 0.5 to each upper limit.

The actual classes will be

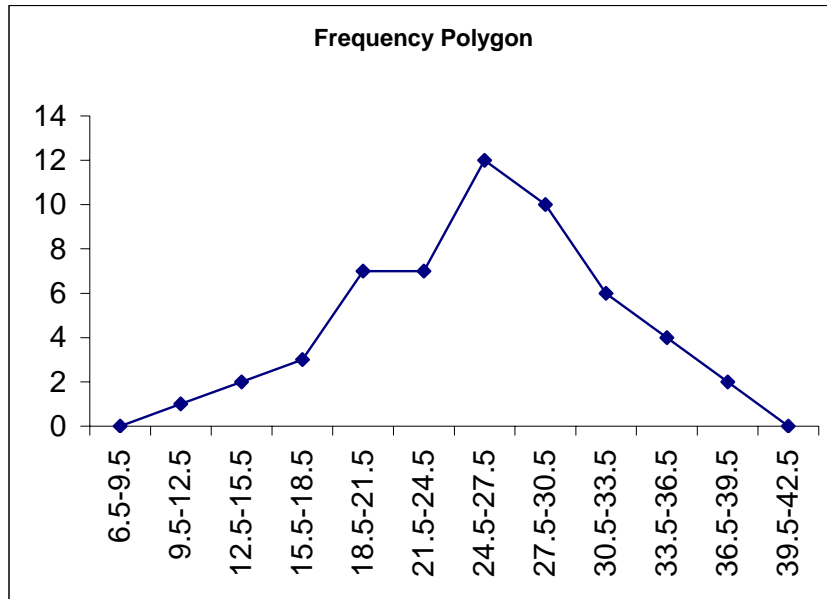
<b>Classes</b>	<b>Frequency</b>
36.5-39.5	2
33.5-36.5	4
30.5-33.5	6
27.5-30.5	10
24.5-27.5	12
21.5-24.5	7
18.5-21.5	7
15.5-18.5	3
12.5-15.5	2
9.5-12.5	1
	54



### Frequency Polygon

A frequency polygon is a graph of frequency distribution. In a histogram, if the midpoints of the upper bases of the rectangles are connected by straight lines, we will get a frequency polygon. It is assumed that the area under the polygon is proportional to the total frequency. We can construct a frequency polygon directly from the frequency distribution without drawing the histogram. For this, as a first step calculate the midpoints of the classes and mark them on the X-axis. Then plot the frequency corresponding to each point and join all these points by straight lines. Join the two ends to the X-axis, the left end to the mid point of the class before the first class and the right end to the midpoint of the class after the last class.

Illustration Classes	Frequency	Mid Point= (Upper limit + Lower limit)/2
36.5-39.5	2	38
33.5-36.5	4	35
30.5-33.5	6	32
27.5-30.5	10	29
24.5-27.5	12	26
21.5-24.5	7	23
18.5-21.5	7	20
15.5-18.5	3	17
12.5-15.5	2	14
9.5-12.5	1	11
	54	



Frequency polygon has some advantages over histogram. Some of them are

- More than one frequency distribution can be presented as frequency polygon in the same graph facilitating comparison. But histogram of different distributions are to be drawn in different graph papers.
- Frequency polygon is simpler than histogram.
- Frequency polygon gives a better idea about the nature of the distribution than histogram.

### Frequency Curve

A frequency curve can be drawn by joining the points plotted for a frequency polygon by a smooth curve. The curve is drawn freehand so that the total area under the curve is approximately the same as that under the polygon.

The smoothed frequency curve can be drawn by calculating the smoothed frequency of each class and plotting the points based on the smoothed frequencies. Smoothed frequency of a class is calculated by adding the three consecutive frequencies and dividing by three.

ie.,

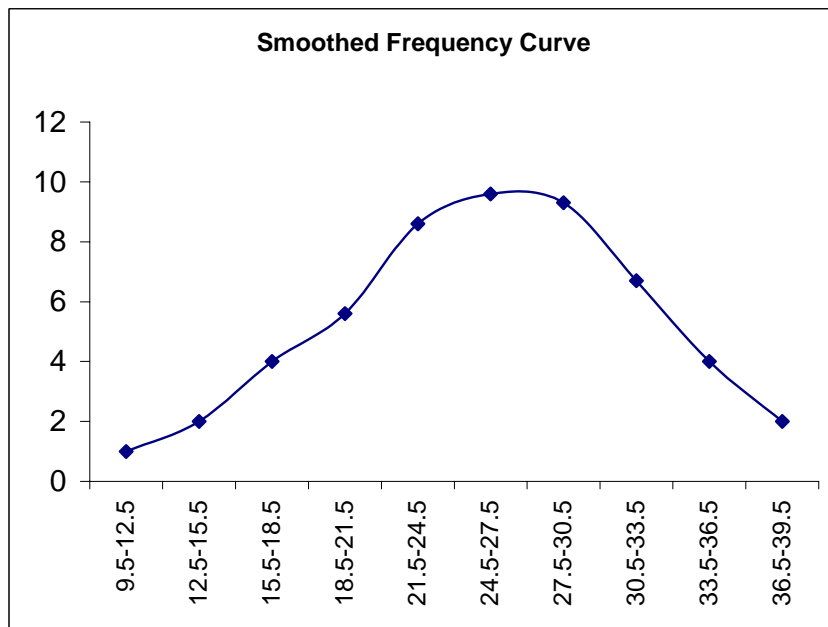
*Smoothed frequency* =

$$\frac{\text{Frequency of the given class} + \text{frequencies of the two adjacent classes}}{3}$$

Illustration

Mid pts	Frequency	Smoothed frequency
38	2	$\frac{0 + 2 + 4}{3} = \frac{6}{3} = 2$
35	4	$\frac{2 + 4 + 6}{3} = \frac{12}{3} = 4$
32	6	$\frac{4 + 6 + 10}{3} = \frac{20}{3} = 6.67$
29	10	$\frac{6 + 10 + 12}{3} = \frac{28}{3} = 9.3$

26	12	$\frac{10+12+7}{3} = \frac{29}{3} = 9.6$
23	7	$\frac{7+7+12}{3} = \frac{26}{3} = 8.6$
21	7	$\frac{7+7+3}{3} = \frac{17}{3} = 5.6$
17	3	$\frac{7+3+2}{3} = \frac{12}{3} = 4$
14	2	$\frac{3+2+1}{3} = \frac{6}{3} = 2$
11	1	$\frac{2+1+0}{3} = \frac{3}{3} = 1$



### Ogives (Cumulative Frequency Curve)

The curve obtained by plotting cumulative frequencies is called a cumulative frequency curve. There are two types of cumulative frequency -less than and greater than. To calculate the less than cumulative frequency, upper limits are taken into consideration and the number of scores less than the upper limit of each class is calculated.

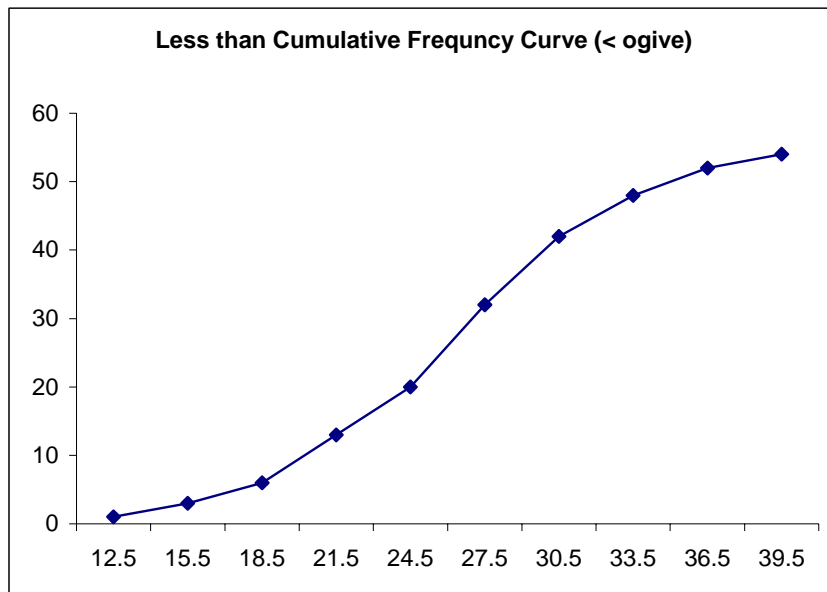
In the case of greater than cumulative frequency, number of scores greater than the lower limit of each class is calculated. The graph drawn with upper limits of the classes on X-axis and less than cumulative frequencies on Y-axis is called a less than cumulative frequency curve.

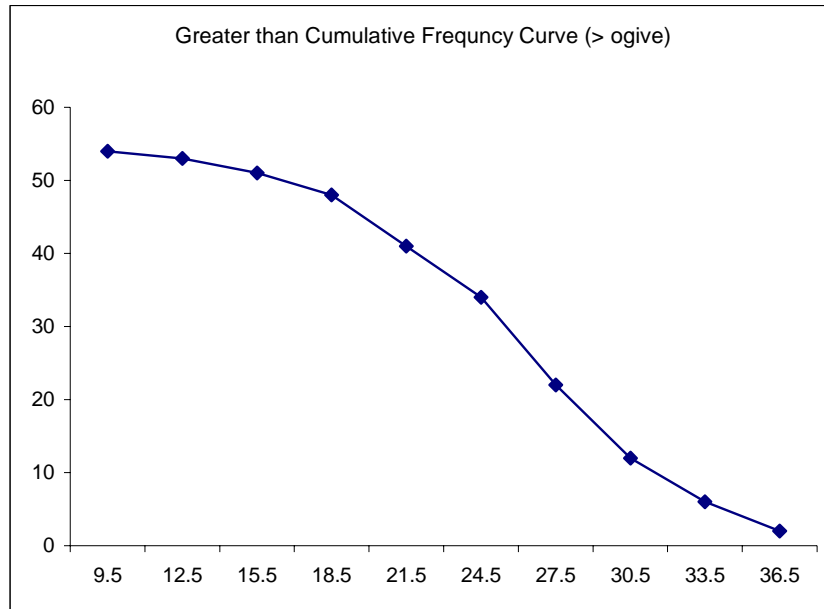
The graph drawn with lower limits of the classes on X-axis and greater than cumulative frequencies on Y-axis is called a greater than cumulative frequency curve.

These curves help the reader to determine the number of cases above or below a given value. They help to determine graphically values of median and quartiles.

Less than cumulative frequency curve and greater than cumulative frequency curve.

Upper Limit	<c.f.(less than cumulative frequency)	Lower limits	>c.f (greater than cumulative frequency)
39.5	54	36.5	2
36.5	52	33.5	6
33.5	48	30.5	12
30.5	42	27.5	22
27.5	32	24.5	34
24.5	20	21.5	41
21.5	13	18.5	48
18.5	6	15.5	51
15.5	3	12.5	53
12.5	1	9.5	54





If the cumulative frequencies are converted into the corresponding percentages the respective curves will be called as less than Ogive (read an Ojive) and greater than ogive. Ogives help the investigator to determine percentiles and deciles directly from the graph.

These curves have much uses in research but they are not simple to interpret.

### **Limitations of graphs**

Diagrams and graphs are powerful mode of presenting frequency distribution, but they can not always substitute the tabular presentation. While selecting graph or diagram for presenting the data, utmost care must be taken to select the most appropriate one for the given purpose.

Some limitations of graphs and diagrams are

- They present only approximate values.
- They are representing only a limited amount of information.
- They can be easily misinterpreted.
- They are used for explaining quantitative data to the common man, for a statistician they are not much helpful for analysis of data.

## MODULE 2

### MEASURE OF CENTRAL TENDENCY

#### Objectives

The student will be

1. acquainted with the knowledge of various measures of central tendency, their characteristics and uses.
2. able to compute various measures of central tendency.

#### Meaning:

If a large set of measures are taken, say intelligence score of a group of students, one can observe that majority of scores lie around a single value in between the extreme values. The value around which other measures cluster or tend to cluster is known as measure of central tendency or simply an average. It represents the whole data. An average can be defined as "a typical value in the sense that it is sometimes employed to represent all the individual values in a series or of a variable". (Ya-Lun-Chou).

#### Characteristics of a good measure of central tendency

Measure of central tendency is a single value representing a group of values and hence is supposed to have the following properties.

**1. Easy to understand and simple to calculate.**

A good measure of central tendency must be easy to comprehend and the procedure involved in its calculation should be simple.

**2. Based on all items**

A good average should consider all items in the series.

**3. Rigidly defined**

A measure of central tendency must be clearly and properly defined. It will be better if it is algebraically defined so that personal bias can be avoided in its calculation.

**4. Capable of further algebraic treatment**

A good average should be usable for further calculations.

**5. Not be unduly affected by extreme values**

A good average should not be unduly affected by the extreme or extra ordinary values in a series.

**6. Sampling stability**

A good measure of central tendency should not be affected by sampling fluctuations. That is, it should be stable.

The most common measures of central tendency are

- Arithmetic mean
- Median and
- Mode



Each average has its own advantages and disadvantages while representing a series of number. The details of these averages are given below.

**1. Arithmetic Mean**

The most useful and popular measure of central tendency is the arithmetic mean. It is defined as the sum of all the items divided by the number of items. Mean is usually denoted as M or  $\bar{X}$

**Mean for raw data:**

Let  $x_1, x_2, x_3, \dots, x_n$  be the n scores in a group. Then its Arithmetic mean  $\bar{X}$  is calculated as

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum x}{n}$$

Where  $\sum x$  denote the sum of 'n' items.

For example if 10, 15, 16, 9, 8, 11, 12, 17, 18, 14 are the marks obtained by ten students in a unit test, its arithmetic mean will be

$$\bar{X} = \frac{10+15+16+9+8+11+12+17+18+14}{10} = \frac{130}{10} = 13$$

That is '13' is a single score that can be used to represent the given marks of 10 students.

**Arithmetic mean for Grouped data**

If the data is given as a discrete frequency table, then Arithmetic mean  $\bar{X} = \frac{\sum f_x}{N}$  where

X - the score, f - frequency of that score and N - Total frequency.

Calculate A.M. for the following data.

Score	Frequency
14	2
21	4
23	3
28	4
35	2

Here the data reveals that 14 occur 2 times, 21 occur 4 times 23 occur 3 times and so on. When the formula for A.M is used

$$\bar{X} = \frac{14 \times 2 + 21 \times 4 + 23 \times 3 + 28 \times 4 + 35 \times 2}{2 + 4 + 3 + 4 + 2} = \frac{363}{15} = 24.2$$

Again	x	f	fx
	14	2	28
	21	4	84
	23	3	69
	28	4	112
	35	2	70
	-----	-----	
	N = 15	$\sum f_x$	363

$$\bar{X} = \frac{\sum f_x}{N} = \frac{363}{15} = 24.2$$

That is, A.M is the same when we use different methods to calculate the mean.

If the data is presented as a continuous series (classes and frequencies), then A.M is calculated by using the formula

$$\bar{X} = \frac{\sum f_x}{N} \text{ where}$$

- X - midpoint of a class
- f - frequency of that class
- N - Total frequency.

Classes	Frequency
40-50	15
30-40	20
20-30	10
10-20	15
0-10	10
	$N = 70$

To calculate mean, midpoint of each class is to be calculated. For this upper limit and lower limit of each class is added and divided by two.

Thus midpoints of the classes will be

$$X = \frac{40 + 50}{2} = 45$$

$$\frac{30 + 40}{2} = 35$$

$$\frac{20 + 30}{2} = 25$$

$$\frac{10 + 20}{2} = 15$$

$$\frac{0 + 10}{2} = 5$$

These midpoints are then multiplied by the corresponding frequencies.

ie X x f

$$45 \times 15 = 675$$

$$35 \times 20 = 700$$

$$25 \times 10 = 250$$

$$15 \times 15 = 225$$

$$5 \times 10 = 50$$

Now, sum of these  $Xf$ , i.e.  $\sum fX$ , is computed and divide  $\sum fX$  by the total frequency.

$$\text{Thus } \bar{X} = \frac{1900}{70} = 27.14$$

**Short-cut Method (Assumed mean Method)**

To make the calculation more easy, we can assume a value as Assumed mean and using the following formula. Arithmetic mean can be calculated.

$$\bar{X} = \frac{A + i \sum f d}{N} \text{ where } A - \text{Assumed mean}$$

$i$  - Class interval

$f$  - Frequency of each class

$$d = \frac{X - A}{i}$$

$X$  - Midpoint of the class

$N$  - Total frequency

**Illustration**

Classes	Frequency	Midpoint $X$	$d = \frac{X - A}{i}$	$fd$
40 - 50	15	45	$\frac{45 - 25}{10} = 2$	$2 \times 15 = 30$
30 - 40	20	35	$\frac{35 - 25}{10} = 1$	$1 \times 20 = 20$
20 - 30	10	25	0	$0 \times 10 = 0$
10 - 20	15	15	$\frac{15 - 25}{10} = -1$	$-1 \times 15 = -15$
0 - 10	10	5	$\frac{5 - 25}{10} = -2$	$-2 \times 10 = -20$
	$N=70$			$50 - 35 = 15$

Take 25 as the assumed mean and calculate 'd' using the formula  $d = \frac{X - A}{i}$ . Multiply each 'd' with the corresponding frequency. Add these values (care should be taken as there will be positive and negative values. Add the numbers with same sign and subtract the smaller from the larger and put the sign of the larger number).

$$\begin{aligned} \text{A.M } \bar{X} &= A + \frac{i \sum f d}{N} \\ &= 25 + \frac{10 \times 15}{70} \\ &= 25 + \frac{150}{70} \\ &= 25 + 2.14 \\ &= 27.14 \end{aligned}$$

### Merits

Arithmetic mean is the most widely used average. It has many advantages. Some of them are:

- It is simple to understand and easy to calculate
- It takes into account all the items of the series
- It is rigidly defined and is mathematical in nature
- It is relatively stable
- It is capable of further algebraic treatment
- Mean is the centre of gravity of the series, balancing the values on either side of it and hence is more typical

### Demerits:

As mean considers all items of the series for its calculation, the value is unduly affected by extreme values (highest or least values). For example, consider the data 10, 30, 35, 36, 34. Here the mean is  $145/5=29$ . A single value 10, reduced the A.M to 29. Similarly even a single higher value will increase the mean of the set of data. Also, when a single item is missing or if the classes are open ended, it is not possible to calculate mean. In distributions, which are highly deviating from normal distribution, mean will not be a suitable measure to represent the data.

### Merits

Arithmetic mean is the most widely used average. It has many advantages some of them are

- It is simple to understand and easy to calculate.
- It takes into account all the items of the series.
- It is rigidly defined and mathematical in nature.
- It is relatively stable.
- It is capable of further algebraic treatment
- Mean is the centre of gravity of the series, balancing the values on either side of it and hence is more typical.

### Demerits

As mean considers all items of the series, for its calculation, the value is unduly affected by extreme values (highest or least values). For example, consider the data 10, 30, 35, 36, 34. Here the mean is  $\frac{145}{5} = 29$ . A single value 10, reduced the A.M. to 29. Similarly even a single higher value will increase the mean of the set of data. Also, when a single item is missing or if the classes are open ended, it is not possible to calculate mean. In distributions, which are highly deviating from normal distribution, mean will not be a suitable measure to represent the data.

### Median

Median is the central value in a series, when the measures are arranged in the order of magnitude. One-half of the items in the distribution have value less than or equal to the median value and one-half have a value greater than or equal to median. That is median is the middle value in a distribution and it splits the set of values (observations) into two halves. Median is a positional average and not a value calculated from every items of the series. Median is the value so that equal number of items lie on either side of it.

When the number of observations in a series is odd, it will be easy to calculate median.

Arrange the items in the order of magnitude, take the  $\left(\frac{n+1}{2}\right)^{\text{th}}$  item in the series. It will be the median.

For example, consider the data

210, 121, 98, 81, 226, 260, 180, 167, 140, 138, 149.

Re arrange the values according to magnitude.

81, 98, 121, 138, 140, 149, 167, 180, 210, 226, 260.

As there are 11 observations take  $\left(\frac{11+1}{2}\right)^{\text{th}}$  item i.e. 6<sup>th</sup> item in the series which is 149. Median of the given set of measures is 149 and number of observations to the left of 149 is the same as that to the right.

But when the number of observations is even, median is the average of the two middle position values. That is, when 'n' is even, median is the average of  $\left(\frac{n}{2}\right)^{\text{th}}$  and  $\left(\frac{n}{2}+1\right)^{\text{th}}$  items after arranging in the order of magnitude.

Consider the values, 269, 247, 272, 282, 254, 266 when arranged in the order of magnitude,

247, 254, 266, 269, 272, 282 the middle values are  $\left(\frac{n}{2}\right)^{\text{th}}$  and  $\left(\frac{n}{2}+1\right)^{\text{th}}$  values i.e. 3<sup>rd</sup> and 4<sup>th</sup> values.

266 and 269 respectively median will be the average of 266 and 269 i.e.  $\frac{266+269}{2} = 267.5$

Median of a frequency distribution

Discrete series

For a discrete series, calculation of median involves the following steps.

**Step 1** Arrange the values in the order of magnitude.

**Step 2:** Write the cumulative frequency, i.e. the number of observations less than that value.

**Step 3:** Find the  $\left(\frac{N}{2}\right)^{\text{th}}$  item in the series. This can be done by looking for the item with

cumulative frequency equal to or greater than  $\frac{N}{2}$ ,

Score	Frequency
84	4
38	7
71	3
65	8
40	5
	27

Re arranging the series

Scores	Frequency	C.F	
			$\frac{N}{2} = \frac{27}{2} = 13.5$
38	7	7	'F' Take the 14 <sup>th</sup> item in the series. When cumulative frequency is observed we can see that 14 <sup>th</sup> item will be 65 as the c.f. of the previous item 40 is 12 which less than N/2
40	5	12	
65	8	20	
71	3	23	
84	4	27	
	-----		
	27		

∴ Median is 14<sup>th</sup> item = 65  
 ===

**Continuous series**

For a continuous series median is calculated using the formula

$$\text{Median} = l + \frac{i(N - m)}{f} \text{ where}$$

- l = lower limit of the median class
- i = Class interval of the median class
- N = Total frequency;
- m = Cumulative frequency of the class below the median class
- f - Frequency of the median class.

Median class is that class for which the cumulative frequency is equal to or greater than  $\frac{N}{2}$  for the first time in the distribution. It is the class in which median will fall. (The classes must be written as exact classes).

Classes	Frequency	Exact classes	c.f.
41-50	5	40.5-50.5	50
31-40	15	30.5-40.5	45
21-30	8	20.5-30.5	30
11-20	17	10.5-20.5	22
1-10	5	0.5-10.5	5
	-----		
	50		

$$\frac{N}{2} = \frac{50}{2} = 25$$

The class with c.f. 25 or above is 20.5 - 30.5 (c.f = 30) and hence it is the median class. The class just before the median class is 10.5-20.5 with c.f. 22.

$$\begin{aligned} \therefore \text{Median} &= 1 + \frac{i(N-m)}{f} && l = 20.5 \\ &= 20.5 + 10 \frac{(25-22)}{8} && i = 10 \\ &= 20.5 + 10 \times \frac{3}{8} && \frac{N}{2} = 25 \\ &= 20.5 + \frac{30}{8} && m = 22 \\ &= 20.5 + 3.75 && F = 8 \\ &= 24.25 \\ &===== \end{aligned}$$

**Merits**

- Median is not affected by extreme values. Median of the data 10, 11, 12, 13, 14 is 12 and that of 0, 10, 12, 14, 100 is also 12. Hence if we know that the distribution contains extreme values, median will be more representative than mean.
- In open ended classes and if the data is incomplete, if the relative position of the missing data is known, median can be calculated.
- Median can be calculated graphically (by drawing the ogives and taking the x-co-ordinate of the meeting points of the two curves).
- Median is simple to understand and easy to calculate.

**Demerits**

- Calculation of median need re arrangement of data and hence will be difficult if number of observations is large.
- It is only a [positional average and is not based on all other items.
- It can not be used for further algebraic treatment.
- Less stable than arithmetic mean.
- Median is not mathematically defined or rigidly defined. As the number of observations becomes odd or even, the position of median changes.

**Mode**

The mode is that value in a series of observations which occurs with the greatest frequency.

**Example**

5, 3 8, 5, 13, 9, 5, 11, 5, 8, 10, 8, 5, 6, 5

In this set of observations, 5 is repeated six times and is the most frequent item in the series taken other values. Hence mode in this case is 5.

Mode is the value which occurs most often in the data. It is the value at the point around which the items tend to be most heavily concentrated.

In a raw data, mode can be found out by counting the number of times the various values repeat themselves and finding the value occurring maximum number of times. If there are two or

more values with the highest frequency, mode is ill-defined or all the values with highest frequency are modal values. Then the distribution is bi-modal or multimodal.

If all the items of a series are not repeating or repeating the same time, one can say that there is no mode for the distribution.

For grouped data

In a discrete series, mode is the item with highest frequency.

For a continuous series, mode can be calculated by the formula.

$$\text{Mode} = L + \frac{i f_2}{f_1 + f_2} \text{ where}$$

L = Exact lower limit of the modal class (the class with highest frequency).

i - Class interval

$f_1$  - Frequency of the class just below the modal class (preceding the modal class)

$f_2$  - Frequency of the class just above the modal class (succeeding the modal class)

### Empirical formula

Mode can be estimated from the values of mean and median using the formula.

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

### Illustration

Raw data 9, 8, 6, 11, 12, 9, 9, 8, 6, 13 here 9 repeats three times and no other value repeat that much. That is 9 is the most frequent item in the series. Hence mode of these observations is 9.

If the series is

9, 8, 6, 11, 6, 9, 9, 8, 6, 13, 6 and 9 are occurring most frequently and there are two modes 6 and 9.

If the values are 9, 8, 6, 11, 13 there is no mode.

Discrete series

Marks	:	10	15	20	25	30	35
Frequency:		8	12	30	27	18	9

In this frequency distribution, 20 is the mark with maximum frequency and therefore mode is 20.

Continuous series

Class	Frequency	Exact class
41 - 50	5	40.5 - 50.5
31 - 40	15	30.5 - 40.5
21 - 30	8	20.5 - 30.5
11 - 20	17	10.5 - 20.5
1 - 10	5	0.5 - 10.5

10.5 - 20.5 is the modal class as the frequency is maximum for that class.



$$\text{Mode} = 1 + \frac{i f_2}{f_1 + f_2}$$

$l = 10.5$ ,  $f_1$  - frequency of the preceding class ie frequency of the class  $0.5 - 10.5 = 5$

$f_2$  = frequency of the succeeding class

= frequency of the class  $20.5 - 30.5$

= 8

$i = 10$

$$\begin{aligned} \therefore \text{Mode} &= 10.5 + \frac{10 \times 8}{5 + 8} = 10.5 + \frac{80}{13} = 10.5 + 6.15 \\ &= 16.65 \\ &= = = = \end{aligned}$$

Supposed in a frequency distribution Mean = 44.8 and Median = 44. Then mode can be taken as

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

$$= 3 \times 44 - 2 \times 44.8$$

$$= 132 - 89.6 = 42.4$$

### Merits

1. Mode is the simplest measure of central tendency.
2. It gives quickest measure of central tendency. Therefore when a quickest, but approximate, value to represent a group of observations is needed, mode can be used.
3. Mode is not affected by extreme values.
4. In open ended classes, mode can be calculated.
5. Even qualitative data can be described through mode. (When we say the consumer preferences of a product, modal value is used instead of median or mean which are not even meaningful).

### Demerits

1. Mode is not rigidly defined. In all cases we can not calculate a unique mode. It may be bimodal or multimodal. That is mode is ill-defined.
2. It is not capable of further algebraic calculations.
3. It is not based on all items of a series.
4. Mode is less used in quantitative data as mean and median are more representative of the distribution.

When to use Mean, Median and Mode.

Arithmetic mean is the most reliable and accurate measure of central tendency. It is more stable than median or mode and is less affected by sampling fluctuations. When we need a reliable, more accurate measure to represent the data, mean can be used. If we want to compute more statistics like standard deviation, correlation etc, Mean is recommended.

But when there are extreme values in the set of data, mean will not be a true representation. If extreme values exist, median will give more representation of data than mean. If we want to find the middle most value in the series, median is calculated. If the classes are open-ended, or some values are missing, but their relative position is known, mean cannot be calculated, where median becomes the most reliable measure of central tendency.

When a crude (rough) measure of central tendency is needed or if we want to know the most often recurring value, mode is calculated. Mode can be easily obtained from the graphs like histogram, frequency polygon or frequency curve.

It should be borne in mind that these values, mean, median or mode are values representing a group of values. That is average or measure of central tendency is a single value representing a group of values and hence it must be properly interpreted, otherwise will arrive at wrong decisions. It lies between the lowest and highest values in the series. Some times the average need not be a value in the series. This may also lead the individual to wrong interpretation. For example, the average size of a family is 4.6 is absurd as a family can not have a size of 4.6.

Two or more set of values may have the same measure of central tendency but differ in their nature. Therefore while comparing distributions measures of central tendency will not give complete picture of the distributions.

Also, if the data is not having a clear single concentration, of observations, an average will not be meaningful.

## MODULE 3

### MEASURE OF DISPERSION

#### Objectives

The module will help the learner to

- Know about various measures of dispersion like range, quartile deviation, mean deviation and standard deviation and their calculations.
- develop skill in calculating various measures of dispersion.
- decide upon the use of various measures of dispersion.
- know how to calculate and interpret variance and co-efficient of variance.

#### Measure of dispersion

As you know, measures of central tendency provide a single value to represent a set of values, but they are not adequate to describe a set of observations unless all the observations are the same. To describe a set of measures, the variability or dispersion also should be considered and hence a measure of dispersion is needed to study the important characteristics of a distribution.

Measure of dispersion can be defined as the degree to which numerical data tend to spread about an average value. More clearly, dispersion measures the extent to which the items vary from some central value.

The dispersion or scatter or variability of a set of values can be calculated mainly through four measures

1. Range
2. Quartile deviation
3. Mean deviation and
4. Standard deviation

These four measures give idea about the variability or dispersion of the values in a set of data.

1. Range: Range is the simplest measure of dispersion. It is the difference between the smallest and largest items of a distribution.

ie  $\text{Range} = H - L$  where H - the highest measure in the distribution and  
L - lowest measure in the distribution.

It is a very rough measure of dispersion, considering only the extreme values.

#### Merits

Range is simple to understand and easy to computer. When a quick rather than a very accurate picture of variability is needed, range may be computed.

#### Demerits

It is not based on each item of the series. Range is affected by sampling fluctuation. Range does not give any idea about the features of the distribution in between the extreme values.

In open ended distribution range cannot be calculated. (In a frequency distribution range is the difference between the upper limit of the highest class and the lower class).

**2. Quartile Deviation (Semi inter quartile range)**

$$\text{Quartile Deviation Q.D} = \frac{Q_3 - Q_1}{2} \quad \text{where}$$

$Q_1$  and  $Q_3$  are the 1<sup>st</sup> and 3<sup>rd</sup> quartiles of the distribution. The first quartile  $Q_1$  is that value below which 25% of the distribution fall. The third quartile  $Q_3$  is the point below which 75% cases lie.

$$Q_1 = l_1 + \frac{i \left( \frac{N}{4} - m_1 \right)}{f_1} \quad \text{where } l_1 - \text{lower limit of the first quartile class.}$$

$i$  – class interval

$N$  – total frequency

$m_1$  – c.f upto the first quartile class.

$f_1$  – frequency of the first quartile class

$$Q_3 = l_3 + \frac{i \left( \frac{3N}{4} - m_3 \right)}{f_3}$$

$m_3$  – c.f upto the third quartile class.

$f_3$  - frequency of the third quartile class.

First quartile class in the class with c.f. greater than or equal to  $\frac{N}{4}$  for the time.

Third quartile class in the class with C.f. greater than or equal to  $\frac{3N}{4}$  for the first time.

**For raw data**

20, 18, 16, 25, 28, 32, 15

Arranging according to magnitude, the series becomes.

15, 16, 18, 20, 25, 28, 30, 32

$$\frac{N}{4}^{\text{th}} \text{ item} = \frac{8}{4}^{\text{th}} \text{ item} = 2^{\text{nd}} \text{ item} = 16. \quad \therefore Q_1 = 16.$$

$$\left( \frac{3N}{4} \right)^{\text{th}} \text{ item} = 3 \times 2 = 6^{\text{th}} \text{ item} = 28. \quad \therefore Q_3 = 28.$$

$$\text{Q.D} = \frac{Q_3 - Q_1}{2} = \frac{28 - 16}{2} = \frac{12}{2} = 6$$

**For discrete Series**

Marks	:	10	20	30	40	50	60
Frequency:		4	7	8	10	6	5
c.f.	:	4	11	19	29	35	40

$$Q_1 = \left( \frac{N}{4} \right)_{\text{item}}^{\text{th}} = \frac{40}{4} = 10^{\text{th}} \text{ item} \therefore Q_1 = 20$$

$$Q_3 = \left( \frac{3N}{4} \right)_{\text{item}}^{\text{th}} = 3 \times 10 = 30^{\text{th}} \text{ item} \therefore Q_3 = 50$$

$$\therefore Q.D = \frac{Q_3 - Q_1}{2} = \frac{50 - 20}{2} = \frac{30}{2} = 15$$

**For continuous frequency distribution**

Class	Frequency	c.f
40 - 5 - 50.5	5	50
30.5 - 40.5	15	45
20.5 - 30.5	8	30
10.5 - 20.5	17	22
05 - 10.5	5	5
	-----	
	50	
	===	

$Q_1$  class is the class with c.f. greater than or equal to  $\frac{N}{4} \left( \frac{50}{4} = 12.5 \right)$ .  $Q_1$  class is 10.5 - 20.5.

III<sup>rdly</sup>  $Q_3$  class - the class with c.f.  $\geq \frac{3N}{4}$  ( $= 37.5$ ) is 30.5-40.5

$$\begin{aligned} \therefore Q_1 &= l_1 + \frac{i \left( \frac{N}{4} - m_1 \right)}{f} \\ &= 10.5 + 10 \frac{(12.5 - 5)}{17} = 10.5 + \frac{10 \times 7.5}{17} \\ &= 10.5 + \frac{75}{17} = 10.5 + 4.4 = 14.9 \end{aligned}$$

$$\begin{aligned} Q_3 &= l_3 + i \left( \frac{3N}{4} - m_3 \right) / f_3 \\ &= 30.5 + 10 \frac{(37.5 - 30)}{15} \\ &= 30.5 + \frac{10 \times 7.5}{15} \\ &= 30.5 + \frac{75}{15} = 30.5 + 5 = 35.5 \end{aligned}$$

$$\begin{aligned} Q.D &= \frac{Q_3 - Q_1}{2} \\ &= \frac{35.5 - 14.9}{2} = \frac{20.6}{2} = 10.3 \end{aligned}$$

**Merits**

1. It is superior to range as a measure of dispersion. Range considers the extreme values whereas Q.D considers the range of middle 50% of cases.
2. In open ended class Q.D can be computed.
3. Q.D. is not affected by extreme values.

**Demerits**

1. It is not capable of further mathematical calculation.
2. It is not based on all observations.
3. It is affected by sampling fluctuations.

**3. Mean Deviation**

Mean deviation can be defined as the mean of deviations of all the separate scores in the series taken from their mean. M.D is the simplest measure that rally takes into account the variation of observations from an average (measure of central tendency).

Mean deviation is also termed an average deviation and can be used for finding variation with respect to median or more instead of mean. (The sum of deviations of observations from their Median in the minimum when signs are ignored).

**Mean deviation for Raw data**

If  $x_1, x_2, x_3 \dots x_n$  are in observations of a series,      Mean Deviation.

$$M.D = \frac{1}{n} \sum |X - \bar{X}|$$

$$\text{Or } = \frac{1}{n} \sum |D| = \frac{\sum |D|}{n}$$

Where  $|D| = |X - \bar{X}|$ ,  $\bar{X}$  = A.M. of the series.

Note:  $|x| = x$  if  $x$  is positive

$-x$  if  $x$  is negative For example  $|2|=2, |-2| = 2$ .

(If mean deviation is small, the distribution is highly compact or uniform).

If 15, 21, 26, 13, 14, 18, 28, 25, 12

$$\text{Mean} = \frac{15 + 21 + 26 + 13 + 14 + 18 + 25 + 12}{8} = \frac{144}{8} = 18$$

| D | : |15-18|, |21-18|, |26-18|, |13-18|, |14-18|, |18-18|, |25-18|, |12-18|

$$\sum |D| = 3 + 3 + 8 + 5 + 4 + 0 + 7 + 6 = 36$$

$$\therefore MD = \frac{\sum |D|}{n} = \frac{36}{8} = 4.5$$

**Discrete Series**

In the case of a frequency distribution M.D can be calculated by using the formal.

$$M.D = \frac{\sum f|X - \bar{X}|}{N} \text{ or } = \frac{\sum f|D|}{N}$$

Where,  $|D| = |X - \bar{X}|$  f – frequency.

X: 10	20	30	40	50	60
F: 4	7	8	10	6	5

$$\begin{aligned} X &= \frac{\sum fX}{N} = \frac{10 \times 4 + 20 \times 7 + 30 \times 8 + 40 \times 10 + 50 \times 6 + 60 \times 5}{4 + 7 + 8 + 10 + 6 + 5} \\ &= \frac{40 + 140 + 240 + 400 + 300 + 300}{40} \\ &= \frac{1420}{40} = 35.5 \end{aligned}$$

$$|X - \bar{X}|: |10 - 35.5|, |20 - 35.5|, |30 - 35.5|, |40 - 35.5|, |50 - 35.5|, |60 - 35.5|$$

$$f | D |: 25.5 \times 4 + 5.5 \times 7 + 5.5 \times 8 + 4.5 \times 10 + 14.5 \times 6 + 24.5 \times 5$$

$$\sum f|D| : 102 + 108.5 + 44 + 45 + 87 + 122.5 = 631.5$$

$$M.D = \frac{631.5}{40} = 15.79$$

**For a Continuous Series**

$$M.D = \frac{\sum f |D|}{N} \text{ Where } |D| = |X - \bar{X}|, X \text{ -midpoint of the class.}$$

Classes	Frequency	Midpoint (X)	$d = \frac{X-A}{i}$	fd
40.5-50.5	5	45.5	2	10
30.5-40.5	15	35.5	1	15
20.5-30.5	8	25.5	0	0
10.5-20.5	17	15.5	-1	-17
0.5-10.5	5	5.5	-2	-10
	50			-2

Assumed Mean = 25.5

$$\begin{aligned} \text{Mean } X &= A + i \frac{\sum fd}{N} = 25.5 + 10 \times \frac{-2}{50} \\ &= 25.5 - \frac{2}{5} = 25.5 - 0.4 \\ &= 25.1 \quad \text{---} \end{aligned}$$

$ D  =  X - \bar{X} $	$f  D $
20.4	102
10.4	156
0.4	3.2
9.6	163.2
19.6	-----
	522.4

$$\text{M.D} = \frac{\sum f |D|}{N} = \frac{522.4}{50} = 10.45$$

**Merits**

1. It is simple to understand and easy to calculate.
2. It is based on all observations.
3. It is rigidly defined.
4. It is not unduly affected by extreme values.
5. It is statistically stable.

**Demerits**

Mean deviation is a non-algebraic measure as the formula includes the absolute deviations. It can not be used for further algebraic treatment.

Mean deviation is used as a measure of dispersion in small samples and the results are presented before the public with less statistical background. But for higher statistical purposes, mean deviation is no recommended as a measure of dispersion.

**4. Standard Deviation**

Karl Pearson introduced the concept of standard deviation in 1823. It is defined as the square root of the mean of the squared deviations from the arithmetic mean. Standard deviation is usually represented by the small Greek letter 'σ' (sigma). (Σ-the symbol to denote 'sum' is the capital Greek letters sigma).

$$\text{Thus S. D } \sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$$

Where X – individual score

$\bar{X}$  = arithmetic mean

n = number of observation.

**For the raw data**

15, 21, 26, 13, 14, 18, 25, 12

Mean = 18

Sum of squares of deviation from ----



$$\begin{aligned}\sum(X - \bar{X})^2 &= (-3)^2 + 3^2 + 8^2 + (-5)^2 + (-4)^2 + 0 + 7^2 + (-6)^2 \\ &= 9+9+64+25+16+49+36 \\ &= 208\end{aligned}$$

$$\therefore \sigma = \sqrt{\frac{208}{8}} = \sqrt{26} = 5.10$$

In a discrete series.

$$\sigma = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}}$$

Marks:            10        20        30        40        50        60

Frequency:      4        7        8        10       6        5

$$\bar{X} = 35.5$$

$$\begin{aligned}(X - \bar{X})^2 &= (-25.5)^2 \quad (15.5)^2 \quad (-5.5)^2 \quad (4.5)^2 \quad (14.5)^2 \quad (24.5)^2 \\ &: 650.25; \quad 240.25, 30.25 \quad 20.15, \quad 210.25, 60.25\end{aligned}$$

$$\begin{aligned}\sum f(X - \bar{X})^2 &: 4 \times 650.25 + 7 \times 240.25 + 8 \times 30.25 + 10 \times 20.15 + 6 \times 210.25 + 5 \times 60.25 \\ &= 2601 + 1681.75 + 242 + 205 + 1261 + 301.25 \\ &= 6292.5\end{aligned}$$

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} = \sqrt{\frac{6292.5}{40}} = \sqrt{157.3} \\ &= 12.54\end{aligned}$$

For a continuous series.

$$\sigma = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} \text{ or } \sigma = \frac{i}{N} \sqrt{N \sum fd^2 (\sum fd)^2}$$

$$\text{Where } d = \frac{X - A}{i} \text{ A - assumed mean.}$$

(Shortcut method reduces the complexity of Calculation)

Classes	Frequency	X	d = $\frac{X - A}{i}$	d <sup>2</sup>	fd	fd <sup>2</sup>
40.5-50.5	5	45.5	2	4	10	20
30.5-40.5	15	35.5	1	1	15	15
20.5-30.5	8	25.5	0	0	0	0
10.5-20.5	17	15.5	-1	1	-17	17
0.5-10.5	5	5.5	-2	4	-10	20
	-----				----	-----
	50				-2	72

Take assumed mean A = 25.5

$$\begin{aligned}\sigma &= \frac{1}{N} \sqrt{N \sum fd^2 - (\sum fd)^2} \\ &= \frac{10}{50} \sqrt{50 \times 72 - (-2)^2} \\ &= \frac{10}{50} \sqrt{3600 - 4} \\ &= \frac{1}{5} \sqrt{3596} = \frac{1}{5} \times 59.96 \\ &= 11.99\end{aligned}$$

### Merits

SD is well defined and is based on all observation. It is less influenced by sampling fluctuation. The only measure of dispersion that can be used for further mathematical treatment is standard deviation. It is the most stable, reliable measure of dispersion.

### Demerits

It is not easy to compute or simple to understand.

### Variance

Variance is the square of standard deviation. That is variance is the average of each score's squared difference from mean.

If one distribution is more spread out than another, its variance will be larger as the deviation scores will be larger.

$$\text{Variance} = \sigma^2$$

$$= \frac{\sum (X - \bar{X})^2}{n}$$

(for raw data)

$$\text{and variance} = \frac{\sum f(X - \bar{X})^2}{N}$$

(for frequency distribution)

### Coefficient of Variation:

The relative measure of Variation based on standard deviation is coefficient of variation (developed by Karl Pearson).

Coefficient of Variation

$$C.V = \frac{\sigma}{\bar{X}} \times 100$$

While comparing the variance of two or more series C.V is preferred. If CV is greater for a distribution it is less uniform or stable than the other.